A NOTE ON STATISTICAL CONSISTENCY OF NUMERICAL INTEGRATORS FOR MULTI-SCALE DYNAMICS*

JASON FRANK† AND GEORG A. GOTTWALD‡

Abstract. A minimal requirement for simulating multi-scale systems is to reproduce the statistical behavior of the slow variables. In particular, a good numerical method should accurately appropximate the probability density function of the continuous-time slow variables. In this note we use results from homogenization and from backward error analysis to quantify how errors of time integrators affect the mean behavior of trajectories. We show that numerical simulations converge, not to the exact probability density function (pdf) of the homogenized multi-scale system, but rather to that of the homogenized modified equations following from backward error analysis. Using homogenization theory we find that the observed statistical bias is exacerbated for multi-scale systems driven by fast chaotic dynamics that decorrelate insufficiently rapidly. This suggests that to resolve the statistical behavior of trajectories in certain multi-scale systems solvers of sufficiently high order are necessary. Alternatively, backward error analysis suggests the form of an amended vector field that corrects the lowest order bias in Euler's method. The resulting scheme, a second order Taylor method, avoids any statistical drift bias. We corroborate our analysis with a numerical example.

Key words. multi-scale dynamics; homogenization; stochastic parametrization; backward error analysis

AMS subject classifications. 37Mxx, 65Pxx, 60Gxx

1. Introduction.

When simulating complex multi-scale dynamics one is often interested in the accurate description, not of the full system with all its degrees of freedom, but of only some distinct relevant variables, for example the slow variables. Numerical weather forecasting provides a good example, where we are interested in the dynamics of the large-scale high and low pressure fields which evolve on time scales of days rather than in fast buoyancy oscillations of the atmosphere's stratification surfaces. Another example is decadal climate prediction where we are not interested in the actual dynamics of the large-scale atmospheric weather but rather in their effect on the slowly evolving oceanic patterns such as El-Niño. Whereas in the example of numerical weather prediction we desire accurate time evolution of the slow relevant variables, in climate science we are often more interested in statistical properties such as mean global temperature or the frequency of extreme events. Reproducing such mean statistical behavior of the slow variables is a minimal requirement for any simulation of complex multi-scale systems. It has long been recognized that the numerical discretization scheme employed to simulate a dynamical system profoundly affects the numerically observed statistical behavior [2, 7, 8]. This note is concerned with the problem of numerically integrating multi-scale systems with the aim to reliably recover their statistical properties.

We consider here deterministic multi-scale systems of the form

$$\dot{x} = \frac{1}{\varepsilon} h(x) f_0(y) + f(x,y), \quad x(0) = \xi \tag{1}$$

$$\dot{y} = \frac{1}{\varepsilon^2} g(y), \quad y(0) = \eta, \tag{2}$$

*Submitted to the editors DATE.

†Mathematical Institute, Utrecht University, P.O. Box 80010, 3508 TA Utrecht, the Netherlands (J.E.Frank@uu.nl).

‡School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia, (georg.gottwald@sydney.edu.au).

with $x \in \mathbb{R}^d$, $y \in \mathbb{R}^\ell$. The parameter $\varepsilon \ll 1$ characterizes the degree of time scale separation. Here the slow dynamics evolves on a characteristic time of order 1 and the fast dynamics on a characteristic time of ε^2. We assume that the vector fields $f_0 : \mathbb{R}^\ell \to \mathbb{R}^m$, $h : \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R}^m)$, $f : \mathbb{R}^d \times \mathbb{R}^\ell \to \mathbb{R}^d$ and $g : \mathbb{R}^\ell \to \mathbb{R}^\ell$ satisfy certain regularity conditions and that the fast y-dynamics is sufficiently chaotic with compact chaotic attractor $\Lambda \subset \mathbb{R}^\ell$ and ergodic invariant probability measure μ. We consider the case when $\int_\Lambda f_0 \, d\mu = 0$, i.e. when classical averaging would yield trivial constant-in-time dynamics. In this situation the slow dynamics exhibits stochastic dynamics on the slow time scale $\mathcal{O}(1)$ [12, 33].

Numerical simulation of the multi-scale system (1)–(2) is challenging: To capture the slow dynamics of interest, for any fixed value of the time scale separation parameter ε, we obtain convergence in the limit $\Delta t \to 0$, but for ε small, the time step Δt used to propagate the slow variables must be chosen of the order of ε^2 to resolve the fast dynamics and meet stability restrictions, making direct numerical simulations computationally impractical. A minimal requirement for a numerical integrator is that it should reproduce the statistical behavior of the slow variables of interest. Ideally we would like to employ $\Delta t \sim \mathcal{O}(1)$[1]. However, we will see that depending on the statistical behavior of the fast dynamics, in particular on the decay of the correlation function of $f_0(y)$, a time step $\Delta t \sim \mathcal{O}(\varepsilon^2)$ may not be sufficient to recover even the statistical behavior of the slow dynamics and one will need time steps such that $\kappa = \Delta t/\varepsilon^2 \to 0$ as $\varepsilon \to 0$. (Note that for $\kappa > 0$ solutions of the fast integrator do not converge in the limit $\varepsilon \to 0$ to the exact solution of (2).) In other words, it is insufficient to simply resolve the fast motions as $\varepsilon \to 0$, one must in fact *accurately approximate them* in this limit, even when the goal is to determine the mean behavior of the slow variables. This inability of numerical time steppers of order p to reproduce the statistical behavior of the slow dynamics will be linked to the persistence of $\mathcal{O}(\Delta t^p)$-terms in the backward error analysis; furthermore these error terms have a quantifiable influence on the long-time statistics as they will be shown to correspond to drift corrections in the homogenized diffusive limit equations of the numerical discrete time maps.

For multi-scale systems of the form (1)–(2) the statistical behavior of the slow dynamics, in the limit of infinite time scale separation $\varepsilon \to 0$, is described by a stochastic differential equation (SDE) which can be explicitly stated. The mathematical tool to describe the long-time stochastic behavior of slow dynamics is known as *homogenization* [12, 33]. Homogenization describes the integrated effect of the fast (either stochastic or chaotic) dynamics on the slow variables as noise. Initially developed for stochastic multi-scale systems [20, 21, 32], homogenization has been extended recently to deterministic multi-scale systems. In the deterministic case the theory is restricted to the skew-product case (1)–(2) in which the slow dynamics does not couple back to the fast dynamics. The fully coupled case poses the potential problem that the invariant measure of the fast dynamics may not vary smoothly with the slow variable; in this instance the averaged vector fields may not even be Lipshitz and uniqueness and existence of the homogenized equation may not be guaranteed. For the deterministic skew product case (1)–(2), it was shown rigorously that for sufficiently chaotic fast dynamics the emergent stochastic long-time behavior of the slow dynamics is given by stochastic differential equations driven by Brownian motion [30, 15, 18]. The as-

[1] Special multi-scale methods have been devised to do this (see, e.g. [9, 10, 11, 19]).

sumed mild conditions on the chaoticity of the fast y-dynamics are satisfied by a large class of maps and flows. For maps, the convergence to Brownian motion holds when the correlation function is summable. For flows, it suffices that there is a Poincaré map with these properties (irrespective of the mixing properties of the flow). These include, but go far beyond, Axiom A diffeomorphisms and flows, Hénon-like attractors and Lorenz attractors. Precise statements about the validity can be found in [27, 28, 29]. We remark that for weakly chaotic dynamics when the correlations are not summable, the noise is not Brownian anymore but rather α-stable [15][2]. Homogenization has been used as a framework for stochastic parametrizations in the context of numerical weather forecasting and climate science [24, 25, 23, 26, 31, 6, 13] and is at the core of the design of several efficient numerical multi-scale integrators such as the heterogeneous multi-scale method [9, 10] and equation-free projection [11, 19].

Depending on the underlying deterministic dynamical multi-scale system, the noise appearing in the limiting homogenized SDE can be either additive or multiplicative. It is well known that the solution of an SDE is sensitive to the approximation of the Brownian motion. This sensitivity gives rise to the different interpretations of the noise such as Itô versus Stratonovich interpretations (see the insightful discussion in [17]). In [15] it was shown that in the case when the slow dynamics is one-dimensional the stochastic differential equation describing the diffusive behavior of the slow dynamics is to be interpreted in the Stratonovich sense. The intuitive argument for this result is that the noisy SDE is a rough approximation of a smooth dynamical system, hence in the limiting process of infinite time scale separation classical calculus should prevail which necessitates the Stratonovich interpretation[3]. The limiting SDE for deterministic discrete-time maps, however, was shown to be neither of Stratonovich nor of Itô type. The noise is Itô only if the fast dynamics is δ-correlated.

This immediately points to a problem when numerically simulating a continuous-time multi-scale system: The long-term statistics of a dynamical multi-scale system, be it continuous-time or discrete time, is described by its homogenized limiting SDE. However, the limiting stochastic differential equation describing the long-time statistical behavior of the discretized slow dynamics, that is of the numerical integrator, might be different from that of the continuous-time system it is designed to model. Using backward error analysis, we show that the leading-order term responsible for the difference is the limiting second order contribution of the modified equation corresponding to the numerical map. The main contribution of our work is to show that the local errors of a time stepper generate a long-time error of the mean behavior which is recovered by homogenization theory. These error terms are of the order $\mathcal{O}(\Delta t^p)$ for a pth order integrator. This result allows us to draw an important practical conclusion: In order for a numerical discretization scheme to reproduce the long-time statistical behavior of the slow dynamics it may be necessary to employ a sufficiently high order time-stepping method. In particular, the Euler scheme can lead to massively different statistical behavior with strong bias. This is the case when, as we will see, the fast chaotic dynamics does not decay sufficiently quickly and its statistical behavior is far from being close to independent identically distributed (*i.i.d.*) random

[2]We use the terminology strongly and weakly chaotic here in a manner different from the usual distinction between exponential and algebraic decay of correlations; cf.[14].

[3]This does not hold for higher-dimensional slow sub-spaces where the noise is neither Stratonovich nor Itô [18] and the conditions for the Wong-Zakai theorem are not satisfied.

variables. In contrast, first order schemes *are* sufficient to capture the long-time statistical behavior for multi-scale systems with chaotic fast dynamics exhibiting rapid decay of correlation. As we will see, discretization-induced biases can be expressed using homogenization theory. This allows us to explicitly subtract the bias from the slow vector field of the deterministic equation (1), resulting in a remarkably accurate explicit time stepper.

The paper is organized as follows. In Section 2 we introduce the diffusive limit of the deterministic multi-scale system (1)–(2) and of its associated Euler scheme. The diffusive limits of the original continuous-time deterministic multi-scale system and its Euler discretized version are shown to differ in the drift term. In Section 3 we present the backward error analysis of Euler's method and Heun's method and the homogenized limit of the lowest order modified equation for each, describing how its respective long-time statistics differs from that of (1)–(2). Section 4 presents numerical simulations corroborating our analytical results. We conclude with a summary and an outlook in Section 5.

2. The diffusive limit of the multi-scale system and its Euler scheme. Using fairly weak conditions on the chaoticity of the fast y dynamics, it was recently proved in [30, 15, 18] that the long-term behavior of deterministic multi-scale systems (1)–(2) is stochastic and is described on times of order $\mathcal{O}(1)$ by the following homogenized stochastic differential equation

$$dX = F(X)\,dt + \sigma h(X) \circ dW_t, \quad X(0) = \xi. \tag{3}$$

For simplicity of exposition and ease of computation, we choose in the following $d = m = 1$. The drift term is given by $F(X) = \int_\Lambda f(X,y)\,d\mu$, W_t is unit 1-dimensional Brownian motion with the variance given by a Green-Kubo formula with

$$\frac{1}{2}\sigma^2 = \int_0^\infty \mathcal{C}[f_0(y)](t)\,dt\,, \tag{4}$$

where $\mathcal{C}[f_0(y)](t) = \mathbb{E}[f_0(y) f_0(\varphi^t y)]$ denotes the autocorrelation function of f_0 with φ^t denoting the flow of the vector field $g(y)$ (in particular, φ^t is independent of ε), and the expectation

$$\mathbb{E}[A] = \int_\Lambda A(y) d\mu$$

is taken with respect to the fast invariant measure μ. As discussed in the Introduction, the noise is of Stratonovich type because the smooth dynamical system (1)–(2) is approximated by a rough SDE (3), and hence classical calculus has to be valid throughout the limiting procedure of homogenization. For the precise statements we refer the interested reader to [15].

When the multi-scale system (1) is discretized with time step Δt by a numerical integration method, the slow dynamics is given by a map. For instance, the first order forward Euler method gives

$$x_{n+1} = x_n + \Delta t \frac{1}{\varepsilon} h(x_n) f_0(y_n) + \Delta t\, f(x_n, y_n), \quad x_0 = \xi, \tag{5}$$

where $y_n \approx y(n\Delta t)$ is also obtained via a map $y_{n+1} = \Phi(y_n)$ that approximates φ^t on time Δt. In this paper, we compute $\Phi(y)$ using multiple time stepping, through the

K-fold application of the same numerical integrator as used for the slow dynamics,

$$y_{n,k+1} = y_{n,k} + \delta t\, \varepsilon^{-2} g(y_{n,k}), \quad k = 0, \ldots, K-1, \tag{6}$$

with initial condition $y_{n,0} = y_n$ and time step $\delta t = \Delta t/K$[4]. We set $y_{n+1} = y_{n,K}$ to define the map $y_{n+1} = \Phi(y_n)$. In the limit $\varepsilon \to 0$ we choose the scaling $\Delta t = \kappa\, \varepsilon^2$, where $\kappa > 0$ is a small but finite constant (i.e. we solve the slow equation on the fast time scale). This implies that the effective stepsize of the fast motion in (6) is $\delta t\, \varepsilon^{-2} = \kappa/K$ and the map Φ is independent of ε. Consequently, the fast motion (6) does not converge in the limit $\varepsilon \to 0$ to the exact solution of (2). Instead, the constant K is chosen such that the fast motion is well-resolved for all ε. We also assume that the discrete dynamics (6) possesses a chaotic attractor that satisfies the conditions needed for the existence of the SDE limit as discussed below.

For the map (5) it was rigorously proven in [15] that the long-time statistics on times of order $\mathcal{O}(1/\varepsilon^2)$ is governed by the following SDE

$$dX = \Big(\kappa\, F(X) - \frac{1}{2}\kappa^2 h(X) h'(X)\, \mathbb{E}[f_0^2]\Big)\, dt + \kappa\, \hat{\sigma} h(X) \circ dW_t\,, \quad X(0) = \xi, \tag{7}$$

where $F(X)$ is the same as before for the continuous-time system, W_t is again unit 1-dimensional Brownian motion and the variance is given by a Green-Kubo formula

$$\hat{\sigma}^2 = \mathbb{E}[f_0^2] + \sum_{n=1}^{\infty} \mathbb{E}[f_0(\Phi^n\, y) f_0(y)] + \sum_{n=1}^{\infty} \mathbb{E}[f_0(y) f_0(\Phi^n\, y)], \tag{8}$$

where Φ^n denotes the n-fold application of the discrete map Φ. Evaluating the time integral in (4) as a Riemann sum, comparison with (8) shows that $\hat{\sigma}^2\, \kappa \to \sigma^2$ for $\kappa \to 0$. We remark that for non-zero κ the diffusion coefficient $\hat{\sigma}^2$ may differ from the diffusion coefficient σ^2 of the continuous system. Rescaling time to be measured in units of the discretization "time step" κ, (7) can be rewritten as

$$dX = \Big(F(X) - \frac{1}{2}\kappa\, h(X) h'(X)\, \mathbb{E}[f_0^2]\Big)\, dt + \sqrt{\kappa}\, \hat{\sigma} h(X) \circ dW_t\,, \quad X(0) = \xi, \tag{9}$$

where W_t is unit 1-dimensional Brownian motion on the rescaled time.

Comparing the limiting SDE of the discretized map (9) and the limiting SDE for its associated continuous-time system (3), we see that they differ by an extra drift term in (9)

$$E = -\frac{1}{2}\kappa\, h(X) h'(X)\, \mathbb{E}[f_0^2]. \tag{10}$$

Note that the additional drift term prohibits a Stratonovich interpretation of the noise and hence for finite Δt the statistics of the map is different from the statistics of the original continuous-time system (1)–(2), which we identify with the diffusive limit system (3). A discrepancy of this form was noted in [15] in the case of general maps where the time step Δt (or rescaled time step κ) was implied. In the case where the fast dynamics of the discrete Euler map (5) is *i.i.d.*, i.e. $\hat{\sigma}^2 = \mathbb{E}[f_0^2]$, the noise in the limiting SDE of the discretized map (9) is of the Itô type. This can be heuristically understood by realizing that if the time step $\kappa \gg \tau_{\text{corr}}$ where τ_{corr} is the decorrelation time of the fast continuous-time y-dynamics, the map is already as rough as the

[4] An alternative strategy would be to use one step for the fast system with times step $\delta t = \kappa\varepsilon^2$ followed by one step of the slow system with time step $\Delta t/K$ and repeat this K times [34, 3].

discrete approximation of the noise.

Although the additional drift term (10) is formally of order $\mathcal{O}(\Delta t)$, depending on the dynamical system under consideration, the extra term E can be large and distort the statistical behavior leading to a marked difference between the numerically observed statistical behavior of the slow variable X and the statistical behavior of the slow variable x of the given continuous-time multi-scale system (1)–(2) to be modelled. In Section 4 we provide such an example and show how an Euler discretization may produce erroneous statistical information. In the next section we develop a relationship between the extra drift term (10) obtained in homogenization and backward error analysis. In particular, we will show that the extra drift term (10) generated by a first order numerical time-stepper is present in the backward error analysis and would be absent if the dynamics had been integrated with a higher order scheme instead (however, other terms are typically present in this case). This will show how the first order errors of an Euler-method directly translate into errors of the mean behavior.

3. Backward error analysis. In this section we provide a backward error analysis to explain the presence of the extra term (10) in the homogenized discrete model (9) compared to the homogenized continuous-time model (3). We will see that the extra term arises from the use of the forward Euler scheme for constructing the discrete model (5). Although the extra term (10) is of order $\mathcal{O}(\Delta t)$ and hence disappears in the small step size limit, we stress that in the context of multi-scale problems one is often interested in step size regimes that are insensitive to fast dynamics and $\mathcal{O}(1)$ with respect to the slow dynamics.

Backward error analysis [16, 22] has been successfully employed to understand finite time step effects observed in numerical simulations. The truncation error of a numerical discretization of an ordinary differential equation (ODE) can be expanded as an asymptotic series in the step size Δt with terms involving successively higher derivatives of the vector field. In backward error analysis, the terms of the truncated series are interpreted as a higher order approximation to another, perturbed vector field.

3.1. Lowest order modified equations. We consider a generic differential equation

$$\dot{z} = v(z), \tag{11}$$

the solution of which is to be approximated using a numerical method.

To understand the qualitative behavior of the numerical solution for finite step size Δt, one constructs a modified vector field as an asymptotic expansion

$$\dot{z} = \tilde{v}(z) = v(z) + \Delta t\, v_1(z) + \Delta t^2\, v_2(z) + \cdots, \tag{12}$$

where the terms v_1, v_2, etc. are to be determined such that the solution to (12) matches the expansion of a numerical method applied to (11) to a higher order of accuracy. The continuous-time solutions to the truncated modified differential equation (12) approximate to higher order the numerical output than do those of the original differential equation (11), allowing the modified equation to be used to interpret finite time step effects observed in the numerical time series.

The solution to (12) is expanded in a Taylor series about $z(t)$ to give

$$z(t+\Delta t) = z(t) + \Delta t\, \tilde{v} + \frac{\Delta t^2}{2}\tilde{v}'\tilde{v} + \frac{\Delta t^3}{6}\left[\tilde{v}''(\tilde{v},\tilde{v}) + \tilde{v}'\tilde{v}'\tilde{v}\right] + \mathcal{O}(\Delta t^4), \tag{13}$$

where all terms on the right are evaluated at $z(t)$, and where $\tilde{v}'$ denotes the Jacobian matrix of $\tilde{v}$, $\tilde{v}''$ denotes its (symmetric) three-tensor of second partial derivatives, and $\tilde{v}''(\cdot,\cdot)$ denotes the contraction of this tensor with the two vector arguments. Substituting the expansion (12) into the above and gathering terms of like order yields

$$z(t+\Delta t) = z(t) + \Delta t\, v + \Delta t^2 \left[v_1 + \frac{1}{2}v'v\right] + \Delta t^3 \left[v_2 + \frac{1}{2}(v'v_1 + v_1'v) + \frac{1}{6}(v''(v,v) + v'v'v)\right] + \mathcal{O}(\Delta t^4). \tag{14}$$

Next, one determines the functions v_1, v_2, etc. to match the expansion of a numerical integrator to higher order.

Euler's method is given by $z_{n+1} = z_n + \Delta t\, v(z_n)$. This formula is consistent with (14) up to terms of $\mathcal{O}(\Delta t^2)$, and is consequently a first order approximation to (11). However by choosing

$$v_1 = -\frac{1}{2}v'v$$

in (12), one finds that Euler's method agrees with (14) up to terms of $\mathcal{O}(\Delta t^3)$. Consequently, while Euler's method is a first order approximation of (11), it is a *second order approximation* to the modified differential equation

$$\dot{z} = v - \frac{\Delta t}{2}v'v. \tag{15}$$

This process may be repeated to derive higher order corrections (v_2, v_3, etc.) in the modified equation. The asymptotic expansion generally does not converge for fixed Δt, but may be optimally truncated [16]. Although for general systems there is no guarantee that the lowest order corrections will have the most significant impact on the observed statistics, in our numerical experiments this does appear to be the case.

Next consider the second order Runge-Kutta method (i.e. Heun's method)

$$z_{n+1} = z_n + \frac{\Delta t}{2}\left[v(z_n) + v(z_n + \Delta t v(z_n))\right]. \tag{16}$$

Expanding the right-hand side about z_n gives

$$z_{n+1} = z_n + \frac{\Delta t}{2}\left[2v(z_n) + \Delta t v'(z_n)v(z_n) + \frac{\Delta t^2}{2}v''(v,v)(z_n) + \mathcal{O}(\Delta t^3)\right].$$

By choosing

$$v_1 = 0, \qquad v_2 = \frac{1}{12}v''(v,v) - \frac{1}{6}v'v'v$$

in (12) we make this formula agree with (14) up to terms of $\mathcal{O}(\Delta t^4)$ so the modified equation associated with the second order Runge-Kutta method (16) is (up to terms of $\mathcal{O}(\Delta t^2)$)

$$\dot{z} = v(z) + \frac{\Delta t^2}{12}(v''(v,v) - 2v'v'v), \tag{17}$$

and the second order Runge-Kutta method (16) applied to the ODE (11) is a third order approximation to (17).

Note that the modified equation (15) suggests a correction to Euler's method to eliminate the second order term in (14). Specifically, one can apply Euler's method to the differential equation with corrected vector field

$$\dot{z} = v(z) + \frac{\Delta t}{2} v'(z)v(z).$$

Doing so yields the second order Taylor method

$$z_{n+1} = z_n + \Delta t\, v(z_n) + \frac{\Delta t^2}{2} v'(z_n)v(z_n), \tag{18}$$

which can be efficiently implemented using a finite difference approximation in the last term:

$$v'(z_n)v(z_n) \approx \frac{1}{\tau}\left(v(z_n + \tau v(z_n)) - v(z_n)\right), \quad \tau = \sqrt{\epsilon_m},$$

with machine precision ϵ_m. Matching (18) with (14) shows the Taylor method has modified equation expansion (up to terms of $\mathcal{O}(\Delta t^2)$)

$$\dot{z} = v(z) - \frac{\Delta t^2}{6}(v''(v,v) + v'v'v), \tag{19}$$

and is hence is a second order scheme for the original ODE (11). This will turn out to be advantageous for problems of the form considered here.

3.2. Homogenization of modified equations. In the limit $\varepsilon \to 0$, the time step scales as $\Delta t = \kappa\,\varepsilon^2$, where $\kappa > 0$ is fixed and small. As we will take this limit to homogenize the modified equation, we use $\kappa = \Delta t/\varepsilon^2$ as our expansion parameter for the backward error analysis.

The modified vector fields v_1, v_2, etc. each in turn can be expressed as expansions in ε. Upon homogenization, the lowest order term of $\mathcal{O}(1/\varepsilon)$ contributes to the diffusion and the $\mathcal{O}(1)$ term contributes to the drift. Terms of higher order in ε vanish in the homogenization limit $\varepsilon \to 0$.

Substituting the vector fields of the deterministic multi-scale system (1)–(2) into the modified equation for a first order Euler discretization (15) with $z = (x, y)$ we obtain

$$\begin{aligned}\dot{x} =& \frac{1}{\varepsilon} h(x) f_0(y) + f(x,y) \\ &- \frac{\kappa}{2}\left(\frac{1}{\varepsilon} h(x) f_0'(y)\,\tilde{g}(y) + h(x)h'(x)f_0^2(y) + \partial_y f(x,y)\tilde{g}(y) + \mathcal{O}(\varepsilon)\right) + \mathcal{O}(\kappa^2)\end{aligned} \tag{20}$$

$$\dot{y} = \frac{1}{\varepsilon^2}\tilde{g}(y), \qquad \tilde{g} = g(y) + \frac{\kappa}{2K} g'(y)g(y) + \mathcal{O}((\kappa/K)^2). \tag{21}$$

In the remainder of the Section we substitute the vector field $g(y)$ for the vector field $\tilde{g}(y)$ of the fast modified equation, and similarly substitute φ^t for $\tilde{\varphi}^t$, the flow of $\tilde{g}$, which is again independent of ε since κ and K are fixed. This is admissible if the numerical scheme for the fast dynamics is sufficiently accurate to resolve its statistical behaviour and in particular, the autocorrelation function; given the relation $\delta t = \Delta t/K$ and the scaling $\Delta t = \kappa\,\varepsilon^2$, it is clear that κ/K should be chosen small enough to accurately approximate autocorrelation functions of the unscaled chaotic system with vector field $g(y)$, to allow for the substitution of φ^t for $\tilde{\varphi}^t$.

We now show that the associated stochastic limit system of the continuous-time modified equation of the Euler method (15) is the same as that of the discrete Euler

discretization (5). The homogenized dynamics, approximating the dynamics of (20) on time scales of order $\mathcal{O}(1)$, is given up to $\mathcal{O}(\kappa)$ by

$$dX = F(X)\,dt + \sigma h(X) \circ dW_t, \quad X(0) = \xi. \tag{22}$$

where $F(X)$ is the expectation with respect to $y \sim \mu$ of the terms of $O(1)$ in ε in (20). Noting that $g(y) = \varepsilon^2 \dot{y}$, we observe that the terms $f_0'(y)g(y)$ and $\partial_y f(x,y)g(y)$ in the slow modified equation (20) can be written as a total derivative (taking x constant up to terms of $\mathcal{O}(\varepsilon)$ on the homogenization time scale). Consequently, these terms vanish in expectation as in, for example,

$$\mathbb{E}\left[\partial_y f(X,y)g(y)\right] = \mathbb{E}\left[\frac{d}{dt} f(X, y(t))\big|_{X\,\text{fixed}}\right] = 0, \tag{23}$$

and we are left with the drift term

$$\begin{aligned} F(X) &= \mathbb{E}[f(X,y)] - \frac{\kappa}{2} h(X)h'(X)\,\mathbb{E}[f_0^2] \\ &= \mathbb{E}[f(X,y)] + E. \end{aligned} \tag{24}$$

The diffusion coefficient is given by

$$\begin{aligned} \frac{1}{2}\sigma^2 = &\int_0^\infty \mathbb{E}\left[f_0(y) f_0(\tilde{\varphi}^t y)\right] dt \\ &+ \frac{\kappa}{2} h(X) \int_0^\infty \left(\mathbb{E}\left[f_0(y)\,(\frac{d}{dt} f_0)(\tilde{\varphi}^t y)\right] + \mathbb{E}\left[f_0(\tilde{\varphi}^t y)\,(\frac{d}{dt} f_0)(y)\right] \right) dt\,. \end{aligned} \tag{25}$$

Using ergodicity of the fast dynamics the spatial average in the second integral can be expressed as a time-average; partial integration then can be used to show that the second integral sums to zero. The Green-Kubo formula (8) in the limit of $\varepsilon \to 0$ is recovered provided we substitute φ^t for $\tilde{\varphi}^t$ in (25), which we argued above is admissible provided $\kappa/K \ll 1$. Hence in the limit $\varepsilon \to 0$ we recover the homogenized equation (9) for the forward Euler map (5), and the long term statistics of the Euler discretization captures well the statistics of its associated continuous-time modified equation.

For a second order method, such as the Runge-Kutta method (16) or second order Taylor method (18), the additional drift term E is absent from the modified equation. Nevertheless there are terms of $\mathcal{O}(\kappa^2)$ that could potentially influence the homogenized limit. When the Runge-Kutta method (16) is applied to the deterministic multi-scale system (1)–(2), the lowest order modified equation (17) yields, for the slow variable:

$$\begin{aligned} \dot{x} = &\frac{1}{\varepsilon} h(x) f_0(y) + f(x,y) \\ &+ \frac{\kappa^2}{12} \left(\frac{1}{\varepsilon} \left[h(x) \left(f_0''(y) g^2(y) - 2 f_0'(y) g'(y) g(y) \right) \right] \right. \\ &\left. + \left[\partial_{yy} f(x,y) g^2(y) - 2 \partial_y f(x,y) g'(y) g(y) \right] + \mathcal{O}(\varepsilon) \right) + \mathcal{O}(\kappa^3). \end{aligned} \tag{26}$$

The homogenized drift term becomes

$$F(X) = \mathbb{E}\left[f(X,y) + \frac{\kappa^2}{12} \left(\partial_{yy} f(X,y) g^2(y) - 2 \partial_y f(X,y) g'(y) g(y) \right) \right] + \mathcal{O}(\kappa^3),$$

which implies an $\mathcal{O}(\kappa^2)$ bias in the drift of the slow variables in the limit $\varepsilon \to 0$. The homogenized diffusion parameter becomes

$$\frac{1}{2}\sigma^2 = h(X)\int_0^\infty \mathcal{C}\left[f_0(y) + \frac{\kappa^2}{12}\left(f_0''(y)g^2(y) - 2f_0'(y)g'(y)g(y)\right)\right](t)\,dt + \mathcal{O}(\kappa^3).$$

Here, too, an $\mathcal{O}(\kappa^2)$ bias occurs as $\varepsilon \to 0$.

Finally, when the Taylor method (18) is applied to the deterministic multi-scale system (1)–(2), the lowest order modified equation (19) yields, for the slow variable:

$$\dot{x} = \frac{1}{\varepsilon}\,h(x)f_0(y) + f(x,y) + \frac{\kappa^2}{6}\left(\frac{1}{\varepsilon}v_1^{\text{diff}}(x,y) + v_1^{\text{drift}}(x,y) + \mathcal{O}(\varepsilon)\right) + \mathcal{O}(\kappa^3), \tag{27}$$

where

$$v_1^{\text{diff}}(x,y) = h(x)f_0''(y)g^2(y) + h(x)f_0'(y)g'(y)g(y),$$

and

$$v_1^{\text{drift}}(x,y) = 3h(x)'h(x)f_0'(y)f_0(y)g(y) + \partial_{yy}f(x,y)g^2(y) + \partial_y f(x,y)g'(y)g(y).$$

Noting that $g(y) = \varepsilon^2\dot{y}$, we observe that all terms in the drift perturbation can be written as total derivatives with x fixed and vanish in expectation (cf. (23)):

$$\mathbb{E}[v_1^{\text{drift}}(X,y)] = \mathbb{E}\left[3h'(X)h(X)\frac{d}{dt}\left(\frac{f_0^2(y)}{2}\right) + \frac{d}{dt}\left(\partial_y f(X,y)g(y)\right)_{X\,\text{fixed}}\right] = 0.$$

Consequently, the Taylor method has *no bias in drift to* $\mathcal{O}(\Delta t^3)$, and we expect the drift to be simply given by

$$F(X) = \mathbb{E}\left[f(X,y)\right]. \tag{28}$$

The diffusion term is also a total derivative:

$$v_1^{\text{diff}}(x,y) = \frac{1}{\varepsilon}h(X)\frac{d}{dt}\left(f_0'(y)g(y)\right),$$

and the diffusion parameter σ is of the form

$$\frac{1}{2}\sigma^2 = h(X)\int_0^\infty \left(\mathcal{C}[f_0](t) - \frac{\kappa^2}{3}\mathcal{C}\left[\frac{d}{dt}f_0\right](t)\right)\,dt + \mathcal{O}(\kappa^3).$$

The term does not vanish; this is a correlation function, not just an expectation.

In summary, we note that the additional drift term E of order $\mathcal{O}(\kappa)$ in (24) is absent in the modified equation of a numerical method which is at least second order, such as the Runge-Kutta method (16) or the second order Taylor method (18). For a second order time-stepping method the homogenized modified equation therefore agrees with the homogenized equation of the full multi-scale system (3) up to $\mathcal{O}(\kappa^2)$. However, a second order scheme will generally also have additional corrections to the drift and diffusion which might be of the same magnitude for finite κ as those corresponding to the continuous-time multi-scale system under consideration. Remarkably, the second order Taylor method (18) does not have bias of $\mathcal{O}(\kappa^2)$ in the drift. This can be traced to the fact that the scheme exactly agrees with the second order Taylor expansion of the error, and consequently its higher order terms are exact differentials.

4. Numerical demonstration. We now demonstrate that the additional terms in the backward error analysis may lead to significant bias in the probability density estimation for both the first order forward Euler scheme and, to a lesser degree, the second order Heun's method. In particular we show that the numerical methods converge for $\varepsilon \to 0$ with $\Delta t = \kappa\,\varepsilon^2$ to the homogenized limits of their respective modified equations, which are different from the long-time statistical limit of the original deterministic multi-scale system.

We consider the deterministic multi-scale system (1)–(2) with

$$f_0(y) = ay, \quad h(x) = \sqrt{x}, \quad f(x,y) = b(c-x)y^2,$$

so the slow dynamics is described by the continuous-time system

$$\dot{x} = \frac{1}{\varepsilon} a\sqrt{x}y + b(c-x)y^2\,. \tag{29}$$

We choose here $a = 0.1$, $b = 0.005$ and $c = 0.75$. The slow dynamics is driven by $y = \zeta_2 + \zeta_3$ generated by a fast chaotic Rössler system

$$\varepsilon^2\dot{\zeta}_1 = -\zeta_2 - \zeta_3, \tag{30}$$

$$\varepsilon^2\dot{\zeta}_2 = \zeta_1 + r\zeta_2, \tag{31}$$

$$\varepsilon^2\dot{\zeta}_3 = s + (\zeta_1 - u)\zeta_3\,, \tag{32}$$

with $r = s = 0.25$ and $u = 7$. We will compare the results of a numerical integration of this deterministic multi-scale system using first and second order discretization methods to results from the associated limiting homogenized SDE of this system, describing the long-time statistical behavior.

The diffusive limiting equation of the multi-scale dynamical system (29)–(32) can be obtained via the homogenization techniques presented in Section 2 and is given by the Cox-Ingersoll-Ross (CIR) model [4, 5]

$$dX = \sigma a\sqrt{X}\,dW_t + 2\alpha b(\beta - X)\,dt, \tag{33}$$

where W_t is unit 1-dimensional Brownian motion. The parameters are:

$$\alpha = \frac{1}{2}\,\mathbb{E}[y^2], \tag{34}$$

$$\sigma^2 = 2\int_0^\infty \mathbb{E}[(\varphi^t y)y]\,dt, \tag{35}$$

$$\beta = c + \frac{\sigma^2 a^2}{8\alpha b}. \tag{36}$$

To approximate α, an ensemble simulation of the (unscaled) Rössler system was carried out using a 1000-member ensemble on a time interval $t \in [0, 3.2\times 10^4]$ with initial conditions drawn approximately from μ (see below). We obtain $\alpha = 28.4 \pm 0.1$. To estimate σ^2 we solve $w_{n+1} = w_n + \varepsilon\,\Delta t\, y_n$ where $\Delta t = \kappa\,\varepsilon^2$ and $\kappa = 0.5$. Then for a time series of length $N = \lfloor 1/\varepsilon^2 \rfloor$, $w_N \sim \frac{\Delta t}{\sqrt{N}}\sum_{j=1}^{N-1} y_j$ is approximate Brownian motion with variance $\mathbb{V}[w_N] = \sigma^2\,N\Delta t$. In this way the diffusivity is estimated as $\sigma^2 \approx 0.140 \pm 0.002$.

The Cox-Ingersoll-Ross (CIR) model (33) has the closed form solution

$$\frac{X(t)}{c(t)} \sim H(t), \quad c(t) = \frac{\sigma^2 a^2}{8\alpha b}(1 - e^{-2\alpha b t}), \tag{37}$$

where $H(t)$ is a noncentral χ-squared distribution with $8\alpha\beta b/(a^2\sigma^2)$ degrees of freedom and noncentrality parameter $c(t)^{-1}e^{-2\alpha bt}X(0)$.

To numerically integrate the two-scale system (29)–(32) we use a multiple time-stepping approach [22], with a step size $\Delta t = \kappa\,\varepsilon^2$ for (29) and step size $\delta t = \Delta t/K$ for the fast subsystem (30)–(32). For our illustration we choose successively $\varepsilon \in \{0.05, 0.025, 0.0125, 0.00625\}$ and integrate over the interval $t \in [0, 2.5]$ using $\kappa = 0.5$ and $K = 50$. For this scaling of time step the fast dynamics (30)–(32) is well resolved but is not solved with increasing accuracy in the limit $\varepsilon \to 0$. The probability density function (pdf) of $x(t)$ is estimated using an ensemble with 160000 members. Each member starts from $x(0) = 1$ but observes a distinct time series $y(t)$. Each member initial condition $y(0)$ is drawn from the invariant measure μ by letting a randomly drawn initial condition relax onto the attractor over a transient time of length 25. We initially compare two different numerical discretizations of the multi-scale system (29)–(32). Applied to the generic differential equation

$$\dot{x}(t) = v(x(t), y(t)),$$

these methods are: the forward Euler method

$$x_{n+1} = x_n + \Delta t\, v(x_n, y_n), \tag{38}$$

and the second order Runge-Kutta method (Heun's method)

$$x_{n+1} = x_n + \frac{\Delta t}{2}\left[v(x_n, y_n) + v\left(x_n + \Delta t v(x_n, y_n), y_{n+1}\right)\right]. \tag{39}$$

In the above equations y_n denotes the approximation to $y(t_n)$ obtained from nK steps of size δt.
When the slow dynamics (29) is discretized using a forward Euler method (38) with time step Δt we obtain the map

$$x_{n+1} = x_n + \Delta t \frac{1}{\varepsilon}\, a\sqrt{x_n} y_n + \Delta t\, b(c - x_n) y_n^2. \tag{40}$$

We compare the following probability density functions at time $t = 2.5$:

- [MS1] The empirical pdf of the multi-scale system (29)–(32) computed using the forward Euler scheme (38).
- [MS2] The empirical pdf of the multi-scale system (29)–(32) computed using the second order Runge-Kutta scheme (39).
- [HMC] The exact pdf (37) of the limiting homogenized stochastic CIR model (33) with parameters α given by (34) and σ and β given by (35)–(36) associated with the continuous-time model (29)–(32).

Empirical pdfs of the slow variable x are computed using Matlab's histogram counter with bin size $\Delta x = 0.005$.

The rigorous homogenization results presented in Section 2 assert that the long-time statistics of the full deterministic multi-scale system is described by the pdf [HMC]. Figure 1 shows convergence of the empirical pdfs of the numerically approximated multi-scale system, obtained with the forward Euler ([MS1], left frame) and Runge-Kutta ([MS2], right frame) methods for decreasing ε (dashed lines, with

$\varepsilon = 0.05$ in blue, $\varepsilon = 0.025$ in red, $\varepsilon = 0.0125$ in yellow, and $\varepsilon = 0.00625$ in purple). The exact pdf [HMC] is also indicated in each frame (solid black line). Both numerically computed pdfs appear to converge, in the limit $\varepsilon \to 0$, $\Delta t = \kappa\,\varepsilon^2$, to a density with the wrong mean. The pdf [MS2] of the RK method is significantly closer to [HMC] than is [MS1], but bias is nevertheless clearly present.

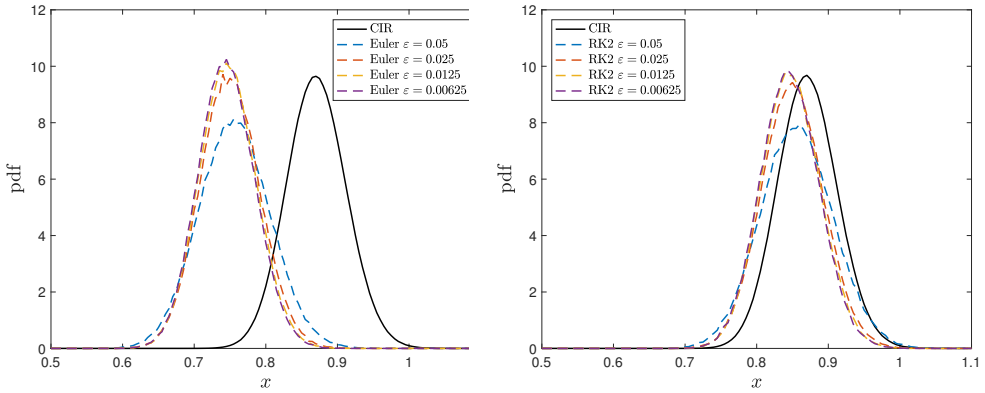

FIG. 1. *Comparison of the pdfs of the numerically approximated multi-scale system (29)–(32) with the exact density of the CIR model (37) at time* $t = 2.5$ *for the forward Euler scheme (38) (left) and the second order Runge-Kutta scheme (39) (right). In each plot the solid black line indicates the exact pdf associated with the homogenized CIR model (33) for the actual time-continuous multi-scale system (29)–(32) (i.e. with parameters (34)–(36)). The dashed lines indicate the empirical probability density function for the numerical simulations of the continuous-time multi-scale system (29)–(32) for* $\varepsilon = 0.05$ *(blue),* 0.025 *(red),* 0.0125 *(yellow), and* 0.00625 *(purple).*

For the first order forward Euler discretization (40) the homogenized SDE (22) describing the long-time behavior of the slow motion is also given by the CIR model (33), but now with parameters

$$\alpha = \frac{1}{2}\,\mathbb{E}[y^2], \tag{41}$$

$$\hat{\sigma}^2 = \mathbb{E}[y^2] + 2\sum_{n=1}^{\infty}\mathbb{E}[(\Phi^n\, y)y] = \lim_{n\to\infty} n^{-1}\mathbb{E}[(\sum_{j=0}^{n-1}\Phi^j\, y)^2], \tag{42}$$

$$\beta = c + \frac{\hat{\sigma}^2\Delta t\, a^2}{8\alpha b} - \frac{\Delta t a^2}{4b}, \tag{43}$$

and $\hat{\sigma}^2\Delta t \approx \sigma^2$. Note that the only difference in the homogenized CIR systems associated with the continuous-time multi-scale system (1)–(2) and the discrete Euler map (40) is in the parameter β (cf. (36) and (43)). For $a^2/b \gg 1$ and $\sigma^2/4\alpha \ll 1$ this difference is large and may cause significant discrepancy between the statistics of the continuous-time multi-scale system (29)–(32) and its first order Euler discretization (40). The latter condition, $\sigma^2/4\alpha = \int_0^\infty \mathbb{E}[(\varphi^t y)y]\,dt/\mathbb{E}[y^2] \ll 1$ puts a requirement on the decay of the fast dynamics and states that the fast dynamics should be far from *i.i.d.* with $\sigma^2 = 2\,\mathbb{E}[y^2]$. This requirement is satisfied for the Rössler system (30)–(32) with the parameters $r = s = 0.25$ and $u = 7$. For these parameters the autocorrelation function has a slow decay and $\mathbb{E}[y^2] \gg \sigma^2/2$. For comparison, we introduce an additional pdf:

- [HMD] The exact pdf (37) of the limiting homogenized stochastic CIR model (33) with parameters α given by (34) and σ and β given by (42)–(43) associated with the discrete Euler model (40).

The backward error analysis presented in Section 3 predicts that the empirical pdf [MS1] of the Euler method may be better approximated by the pdf [HMD], derived by homogenizing the modified equation of the Euler method. Figure 2 confirms this prediction, showing that as $\varepsilon \to 0$, the statistical behavior of the discrete Euler scheme is well described by the pdf of its associated homogenized stochastic CIR model. The extra drift term E in the homogenized discrete model leads to an error of 16% in the mean of the pdf [HMD] with respect to the mean of the pdf of the original continuous time multi-scale system (29)–(32) to be modelled [HMC].

We remark that for fast chaotic dynamics with rapidly decaying autocorrelation function such as the Lorenz 63 system with the classical parameters, we have $\mathbb{E}[y^2] \approx \sigma^2/2$. The homogenized equation of the full multi-scale dynamics and its first order Euler discretization will be close (cf. (36) and (43)), and a first order discretization would be sufficient to capture the long-time statistics of the slow dynamics.

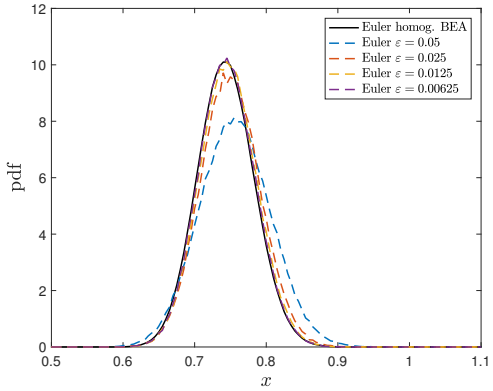

FIG. 2. *Comparison of the pdfs of the numerically approximated multi-scale system (29)–(32) with the exact density of the homogenized modified equation model at time* $t = 2.5$ *for the forward Euler scheme (38). The solid black line indicates the exact pdf associated with the homogenized modified equation (33) of the Euler scheme (i.e. with parameters (41)-(43)). The dashed lines indicate the empirical probability density function computed with the forward Euler approximation (40) applied to the continuous-time multi-scale system (29)–(32) for* $\varepsilon = 0.05$ *(blue),* 0.025 *(red),* 0.0125 *(yellow), and* 0.00625 *(purple).*

Finally, the backward error analysis of the second order Taylor method (18) indicates there is no error in the drift (28) to $\mathcal{O}(\kappa^3)$ for this method. Indeed, Figure 3 confirms this result, illustrating that the empirical pdf of the Taylor method closely matches that of the pdf [HMC] as $\varepsilon \to 0$. In fact both the mean and the variance of the distribution closely match that of the pdf [HMC], suggesting that errors in the diffusion parameter are also small for this parameter regime. The excellent approximation of the drift makes the Taylor method an attractive alternative for multi-scale problems with stochastic limit behavior.

5. Summary. To summarize, using backward error analysis we have demonstrated that the extraneous drift term (10) that arises in homogenization of the discrete map (5) compared to homogenization of the flow (1)–(2) can be traced to the relation between the map (5) and a forward Euler discretization of (1)–(2) with large step size $\Delta t = 1$. In particular, we relate this drift term which appears in the literature of homogenization for discrete time systems [15] and which neither corresponds to an Itô nor to a Stratonovich interpretation of the SDE to discretization errors of first order schemes using backward error analysis. We have shown that the local first

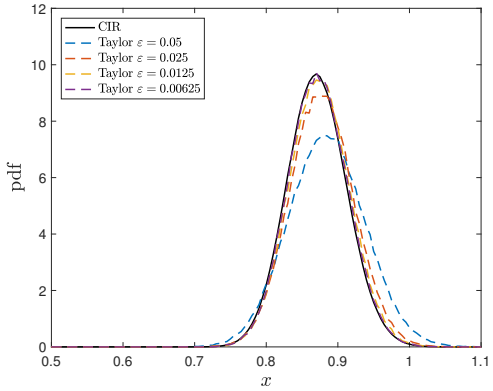

FIG. 3. *Comparison of the pdfs of the numerically approximated multi-scale system (29)–(32) with the exact density of the CIR model (37) at time* $t = 2.5$ *for the second order Taylor method (18). The solid black line indicates the exact pdf associated with the homogenized CIR model (33) for the actual time-continuous multi-scale system (29)–(32) (i.e. with parameters (34)–(36)). The dashed lines indicate the empirical probability density function computed with the second order Taylor method (18) applied to the continuous-time multi-scale system (29)–(32) for* $\varepsilon = 0.05$ *(blue),* 0.025 *(red),* 0.0125 *(yellow), and* 0.00625 *(purple).*

order errors contribute to a well-defined drift error, leading to potentially strong bias, in the long-time statistical behavior. The accumulated local error, as quantified by the backward error analysis, was shown to account for the long-time statistical error of the discretization scheme as provided by homogenization theory.

We further quantified the requirement for a dynamical system such that its Euler scheme discretization reliably recovers its long-time statistical behavior. In particular, we found that for sufficiently rapidly decaying fast dynamics an Euler scheme is sufficient. On the contrary, the failure of first order discretization methods to capture the statistics of the full continuous-time multi-scale system was shown to be exacerbated if the fast dynamics exhibits slow decay of correlations. We remark that the slow decay of correlation is not hampering the validity of the homogenized limit system and the validity of the underlying functional central limit theorem which is assured solely by requiring $\varepsilon \ll 1$. The difference is entirely given by the failure to match the limiting homogenized SDE of the first order discretization with the limiting homogenized SDE associated with the original time-continuous multi-scale system.

Here we discussed deterministic skew product systems of the form (1)–(2). In order to obtain a stochastic homogenized equation for the slow dynamics, the fast dynamics is required to support an ergodic invariant measure and generate an integrable autocorrelation function of f_0 (cf. the Green-Kubo formula (4)). Hence the conclusions drawn here for the deterministic setting remain valid in the case when the fast dynamics is stochastic. For a large class of stochastic ordinary differential equations stochastic integrators were constructed which accurately approximate the invariant measure [1]. Their construction also uses the framework of modified equations and the analysis is not restricted to the multi-scale setting. It would be interesting to compare the higher order methods developed there in the multi-scale setting considered here. This is a topic for further research.

In this article, we have examined the limit $\varepsilon \to 0$, $\Delta t \sim \mathcal{O}(\varepsilon^2)$ in multi-scale systems (1)–(2) which approach a rigorous SDE limit under homogenization. In this

limit, with constant $\kappa = \Delta t/\varepsilon^2 > 0$, the fast motion of the system remains resolved, but *does not converge* as $\Delta t \to 0$. The limit is achieved by choosing K to be sufficiently large and κ/K to be sufficiently small. Our numerical experiments demonstrate significant bias in the pdf of the slow variables in the limit $\Delta t \to 0$ under this scaling. This bias may be mitigated by reducing the ratio $\kappa = \Delta t/\varepsilon^2$ in the numerical experiments as the bias is multiplied by κ (cf. 10). However, full convergence requires $\kappa = \Delta t/\varepsilon^2 \to 0$ as $\varepsilon \to 0$.

We assumed throughout that the fast dynamics is numerically sufficiently resolved such that the statistical properties, e.g. the auto-correlation structure, is sufficiently reproduced. If this were not the case, errors arising from the flow map associated with the fast modified equation enter the Green-Kubo formulae (see, for example, (25) for the Euler method) implying errors in the diffusion coefficient in addition to the bias error.

We remark that special numerical methods are often specifically tailored to multi-scale problems to accommodate time steps that are large with respect to the fast time scale. It is precisely in this regime that statistical bias may occur. The implication of this for numerical integration of multi-scale systems is that, to avoid statistical bias, it may be important to use a higher order method for the slow variables. The second order Taylor method (18) offers an interesting alternative here, as it may be efficiently implemented and is unbiased with respect to the drift up to $\mathcal{O}(\Delta t^3)$. To avoid statistical bias altogether, one might want to solve the actual limiting SDE instead of the deterministic multi-scale system provided that ϵ is sufficiently small to allow for the central limit theorem to hold.

Acknowledgments. We gratefully acknowledge funding through the University of Sydney – University of Utrecht Partnership Collaboration Award.

REFERENCES

[1] A. Abdulle, G. Vilmart, and K. C. Zygalakis, *High order numerical approximation of the invariant measure of ergodic sdes*, SIAM Journal on Numerical Analysis, 52 (2014), pp. 1600–1622, doi:10.1137/130935616, https://doi.org/10.1137/130935616, arXiv:https://doi.org/10.1137/130935616.

[2] R. V. Abramov, G. Kovačič, and A. J. Majda, *Hamiltonian structure and statistically relevant conserved quantities for the truncated burgers-hopf equation*, Communications on pure and applied mathematics, 56 (2003), pp. 1–46.

[3] G. Ariel, B. Engquist, S. Kim, Y. Lee, and R. Tsai, *A multiscale method for highly oscillatory dynamical systems using a Poincaré map type technique*, J. Sci. Comput., 54 (2013), pp. 247–268, doi:10.1007/s10915-012-9656-x.

[4] J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross, *An intertemporal general equilibrium model of asset prices*, Econometrica, 53 (1985), pp. 363–384, doi:10.2307/1911241, http://dx.doi.org/10.2307/1911241.

[5] J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross, *A theory of the term structure of interest rates*, Econometrica, 53 (1985), pp. 385–407, doi:10.2307/1911242, http://dx.doi.org/10.2307/1911242.

[6] J. Culina, S. Kravtsov, and A. H. Monahan, *Stochastic Parameterization Schemes for Use in Realistic Climate Models*, Journal of the Atmospheric Sciences, 68 (2011), pp. 284–299, doi:10.1175/2010JAS3509.1.

[7] S. Dubinkina and J. Frank, *Statistical mechanics of Arakawa's discretizations*, Journal of Computational Physics, 227 (2007), pp. 1286 – 1305.

[8] S. Dubinkina and J. Frank, *Statistical relevance of vorticity conservation in the Hamiltonian particle-mesh method*, Journal of Computational Physics, 229 (2010), pp. 2634 – 2648.

[9] W. E, *Analysis of the heterogeneous multiscale method for ordinary differential equations*, Comm. Math. Sci., 1 (2003), pp. 423–436.

[10] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *Heterogeneous multiscale methods: A review*, Comm. Comp. Phys., 2 (2007), pp. 367–450.

[11] C. Gear and I. Kevrekidis, *Projective methods for differential equations*, SIAM J. Sci. Comp., 24 (2003), pp. 1091–1106.

[12] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–127.

[13] G. Gottwald, D. Crommelin, and C. Franzke, *Ensemble-based Atmospheric Data Assimilation*, in Nonlinear and Stochastic Climate Dynamics, C. L. E. Franzke and T. J. O'Kane, eds., Cambridge University Press, Cambridge, 2017, pp. 209–240.

[14] G. A. Gottwald and I. Melbourne, *A Huygens principle for diffusion and anomalous diffusion in spatially extended systems*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 8411–8416.

[15] G. A. Gottwald and I. Melbourne, *Homogenization for deterministic maps and multiplicative noise*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 469 (2013).

[16] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, vol. 31, Springer, 2006.

[17] W. W. Horsthemke and R. Lefever, *Noise-induced transitions : theory and applications in physics, chemistry, and biology*, Springer series in synergetics, Springer-Verlag, Berlin, New York, 1984, http://opac.inria.fr/record=b1091778.

[18] D. Kelly and I. Melbourne, *Deterministic homogenization for fast–slow systems with chaotic noise*, Journal of Functional Analysis, 272 (2017), pp. 4063 – 4102.

[19] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, G. K. Panagiotis, O. Runborg, and C. Theodoropoulos, *Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis*, Comm. Math. Sci., 1 (2003), pp. 715–762.

[20] R. Z. Khasminsky, *On stochastic processes defined by differential equations with a small parameter*, Theory of Probability and its Applications, 11 (1966), pp. 211–228.

[21] T. G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, Journal of Functional Analysis, 12 (1973), pp. 55–67.

[22] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press, Cambridge, 2005.

[23] A. Majda, I. Timofeyev, and E. Vanden-Eijnden, *A priori tests of a stochastic mode reduction strategy*, Phys. D, 170 (2002), pp. 206–252, doi:10.1016/S0167-2789(02)00578-X, http://dx.doi.org/10.1016/S0167-2789(02)00578-X.

[24] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *Models for stochastic climate prediction*, Proceedings of the National Academy of Sciences, 96 (1999), pp. 14687–14691.

[25] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, Communications on Pure and Applied Mathematics, 54 (2001), pp. 891–974.

[26] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, *Systematic strategies for stochastic mode reduction in climate*, Journal of the Atmospheric Sciences, 60 (2003), pp. 1705–1722.

[27] I. Melbourne and M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Commun. Math. Phys., 260 (2005), pp. 131–146.

[28] I. Melbourne and M. Nicol, *Large deviations for nonuniformly hyperbolic systems*, Trans. Amer. Math. Soc., 360 (2008), pp. 6661–6676, doi:10.1090/S0002-9947-08-04520-0, http://dx.doi.org/10.1090/S0002-9947-08-04520-0.

[29] I. Melbourne and M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Annals of Probability, 37 (2009), pp. 478–505.

[30] I. Melbourne and A. Stuart, *A note on diffusion limits of chaotic skew-product flows*, Nonlinearity, 24 (2011), pp. 1361–1367.

[31] A. H. Monahan and J. Culina, *Stochastic Averaging of Idealized Climate Models*, Journal of Climate, 24 (2011), pp. 3068–3088, doi:10.1175/2011JCLI3641.1.

[32] G. C. Papanicolaou, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain Journal of Mathematics, 6 (1976), pp. 653–674.

[33] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, New York, 2008.

[34] M. Tao, H. Owhadi, and J. E. Marsden, *Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs, and Hamiltonian systems with hidden slow dynamics via flow averaging*, Multiscale Model. Simul., 8 (2010), pp. 1269–1324, doi:10.1137/090771648.