STOCHASTIC MODEL REDUCTION FOR SLOW-FAST SYSTEMS WITH MODERATE TIME-SCALE SEPARATION*

JEROEN WOUTERS† AND GEORG A. GOTTWALD‡

Abstract. We propose a stochastic model reduction strategy for deterministic and stochastic slow-fast systems with a moderate time-scale separation. The stochastic model reduction strategy improves the approximation of systems with finite time-scale separation, when compared to classical homogenization theory, by incorporating deviations from the infinite time-scale limit considered in homogenization, as described by an Edgeworth expansion in the time-scale separation parameter. To approximate these deviations from the limiting homogenized system in the reduced model, a surrogate system is constructed the parameters of which are matched to produce the same Edgeworth expansion as in the original multi-scale system. We corroborate the validity of our approach by numerical examples, showing significant improvements to classical homogenized model reduction.

Key words. multi-scale dynamics; homogenization; stochastic parametrization; Edgeworth expansion

AMS subject classifications. 60Fxx, 60Gxx

1. Introduction.

Complex systems in nature and in the engineered world often exhibit a multi-scale character with slow variables driven by fast dynamics. For example, large proteins [12] and the climate system [26] exhibit both fast, small scale fluctuations and slow, large scale transitions. The high complexity often puts the system out of reach of both analytical and numerical approaches. Typically one is, however, only interested in the dynamics of the slow variables or observables thereof. It is then a formidable challenge to distill reduced slow equations which can make the problem amenable to theoretical analysis, allowing to identify relevant physical effects, or, from a computational perspective, allow for a larger computational time step tailored to the slow time scale.

Homogenization theory [7, 28] derives reduced slow dynamics by assuming an infinitely large time-scale separation between slow and fast variables. It has been rigorously proven for multi-scale systems with stochastic [16, 17, 27] and deterministic chaotic fast dynamics [25, 8, 14] and has been applied with great success in the design of numerical algorithms for molecular dynamics [3, 15] and in stochastic climate modelling [19, 21].

Several challenges remain, however, in formulating reliable stochastic slow limit systems. Whereas homogenization is rigorously proven only for the limiting case of infinite time scale separation, this assumption is never met in the real world. Hence, homogenized stochastic systems may fail in reproducing the statistical behaviour of the underlying deterministic multi-scale system for finite time-scale separation when an intricate interplay between the fast degrees and the slow degrees of freedom is at play.

Homogenization relies on the fact that the slow dynamics experiences the integrated effect of, in the limit of infinitely fast dynamics, infinitely many fast fluctuations. Therefore, homogenization is in effect a manifestation of the central limit theorem (CLT). Finite time scale effects are then akin to finite sums of random variables. In

*Submitted to the editors DATE.
†Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark. (j.wouters@reading.ac.uk)
‡School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia, (georg.gottwald@sydney.edu.au).

the context of random variables, corrections to the CLT for sums of finite length n can be described by the Edgeworth expansion, which provides an expansion of the distributions of sums, asymptotic in $1/\sqrt{n}$ [2]. Such an expansion provides an improved approximation of the pdf of sums for large enough n. Edgeworth expansions have been developed for independent and for weakly dependent identically distributed random variables [10], continuous-time diffusions [1] and ergodic Markov chains [11]. In [30], we have derived an expression for the Edgeworth expansion of multi-scale systems, including the deterministic case. Similarly to the case of sums of random variables, we obtained an improved approximation of transition probabilities of the slow variable for a large enough time scale separation.

The Edgeworth expansion is universal in the sense that it is agnostic about the microscopic details of the fast process. Only integrals over its higher-order correlation functions appear in the analytical expressions we obtain. We will use this aspect of Edgeworth expansions to derive a reduced model by constructing a low-dimensional surrogate model with the same Edgeworth corrections as the original multi-scale model. Surrogate models have previously been used to sample from complex multi-scale systems, see for example [29]. We numerically demonstrate that this surrogate system is superior to homogenization in reproducing the statistical behaviour of the slow dynamics.

The paper is organised as follows. In Section 2 we introduce the multi-scale systems under consideration and their diffusive limits in the case of infinite time scale separation, as provided by homogenization theory. In Section 3 we establish corrections to the homogenized limit using Edgeworth expansions. These are then used in Section 4 to construct a reduced surrogate stochastic model which captures finite time-scale separation effects. We conclude in Section 5 with a discussion and an outlook.

2. Multi-scale systems.

We consider multi-scale systems of the form

$$\mathrm{d}x = \frac{1}{\varepsilon} f_0(x,y)\,\mathrm{d}t + f_1(x,y)\,\mathrm{d}t \tag{1}$$

$$\mathrm{d}y = \frac{1}{\varepsilon^2} g_0(y)\,\mathrm{d}t + \frac{1}{\varepsilon}\beta(y)\,\mathrm{d}W_t + \frac{1}{\varepsilon} g_1(x,y)\,\mathrm{d}t, \tag{2}$$

with slow variables $x \in \mathbb{R}^d$ and fast variables $y \in \mathbb{R}^N$. We assume that the fast dynamics $\mathrm{d}y = g_0\mathrm{d}t + \beta\mathrm{d}W_t$ admits a unique invariant physical measure $\nu(\mathrm{d}y)$ and the full system admits a unique invariant physical measure $\mu^{(\varepsilon)}(\mathrm{d}x, \mathrm{d}y)$ [1]. The system may be stochastic with a non-zero diffusion matrix $\beta \in \mathbb{R}^{N\times l}$ and l-dimensional Brownian motion $\mathrm{d}W_t$, or may be deterministic with $\beta \equiv 0$. In the latter case we assume that the fast dynamics is sufficiently chaotic [2].
Homogenization theory deals with the limit of infinite time-scale separation $\varepsilon \to 0$. In this limit it is well known that when the leading slow vector field averages to zero, i.e. $\langle f_0(x,y)\rangle = 0$, where $\langle A(y)\rangle := \int \nu(\mathrm{d}y)A(y)$, the slow dynamics is approximated by an Itô stochastic differential equation [16, 17, 27, 25, 9, 13] of the form

$$dX = F(X)\mathrm{d}t + \sigma(X)\,\mathrm{d}W_t\ . \tag{3}$$

[1] An ergodic measure is called physical if for a set of initial conditions of nonzero Lebesgue measure the temporal average of a typical observable converges to the spatial average over this measure.

[2] The assumptions on the chaoticity of the fast subsystem are mild. For continuous-time fast system, an associated Poincaré map needs to have a summable correlation function (irrespective of the mixing properties of the flow). Systems with such mild conditions on the chaoticity include, but go far beyond, Axiom A diffeomorphisms and flows, Hénon-like attractors and Lorenz attractors; see [22, 23, 24]

The drift coefficient is given by

$$F(x) = \langle f_1(x,y)\rangle + \int_0^\infty \mathrm{d}s \left(\langle f_0(x,y)\cdot\nabla_x f_0(x,\varphi^t y)\rangle \right.$$
$$\left. + \langle g_1(x,y)\cdot \nabla_y \left(f_0(x,\varphi^t y)\right)\rangle\right), \tag{4}$$

where φ^t denotes the flow map of the fast dynamics, and the diffusion coefficient is given by the Green-Kubo formula

$$\sigma(x)\sigma^T(x) = \int_0^\infty \mathrm{d}s \left\langle f_0(x,y)\otimes f_0(x,\varphi^t y) + f_0(x,\varphi^t y)\otimes f_0(x,y)\right\rangle, \tag{5}$$

where the outer product between two vectors is defined as $(a\otimes b)_{ij} = a_i b_j$ [3]. For details the reader is referred to [13].

3. Edgeworth approximation for dynamical systems. There are three distinct time scales in the system (1)-(2): a fast time scale of $\mathcal{O}(\varepsilon^2)$, an intermediate time-scale of $\mathcal{O}(\varepsilon)$ on which the fast dynamics has equilibrated but the slow dynamics has not yet evolved, and a long diffusive time scale of $\mathcal{O}(1)$ on which the slow variables exhibit non-trivial dynamics. It is on the intermediate time scale that we can expect corrections to the CLT: the time scale is sufficiently long for the fast dynamics to generate near-Gaussian noise but not long enough for the slow dynamics to dominate. This is also reflected in the homogenized SDE (3): displacements of the slow variable are near-Gaussian with $dX \sim \sigma(X)\,\mathrm{d}W_t$ on short time scales. We therefore focus our attention on the limit $\varepsilon \to 0$ with $t/\varepsilon = \theta$ constant, and study the transition probabilities between initial conditions x_0 into the interval $(\mathrm{x},\mathrm{x}+\mathrm{dx})$

$$\pi_\varepsilon(\mathrm{x},t,x_0) = \mathbb{P}\left(\frac{x(t)-x_0}{\sqrt{t}} \in (\mathrm{x},\mathrm{x}+\mathrm{dx})\middle| x(0)=x_0,\, y(0)\sim\mu^{(\varepsilon)}_{x_0}\right).$$

Here $\mu^{(\varepsilon)}_{x_0}$ denotes the conditional measure of $\mu^{(\varepsilon)}$ conditioned on $x = x_0$. In the limit of homogenization theory $\varepsilon\to 0$, the transition probability π_ε with t/ε constant converges to a normal distribution $\mathbf{n}_{0,\sigma^2}(\mathrm{x})$ with the covariance given by the Green-Kubo formula (5). For finite ε, the transition probability will not be Gaussian but will have correction terms of $\mathcal{O}(\sqrt{\varepsilon})$, the so called Edgeworth corrections. As we have shown in [30], the corrections to the limiting Gaussian distribution of $\hat{x}(t) := (x(t)-x_0)/\sqrt{t}$ are most readily calculated through the characteristic function $\chi_\varepsilon(\omega) = \mathbb{E}^{x_0,\mu}_\varepsilon\left[\exp(i\omega\hat{x})\right]$ where $\mathbb{E}^{x_0,\mu}_\varepsilon$ is the expectation value w.r.t. π_ε. We can expand the characteristic function and then determine the expansion of the probability distribution by inverse Fourier transform. Since $\ln\chi_\varepsilon = \sum_n c^{(n)}_\varepsilon (i\omega)^n/n!$ with the cumulants of $\hat{x}$

$$c^{(p)}_\varepsilon = m^{(p)}_\varepsilon - \sum_{j=1}^{p-1}\binom{p-1}{j-1} m^{(p-j)}_\varepsilon c^{(j)}_\varepsilon,$$

and the moments $m^{(p)}_\varepsilon = \mathbb{E}^{x_0,\mu}_\varepsilon[\hat{x}^p]$, we can expand χ_ε by seeking an asymptotic expansion

$$c^{(p)}_\varepsilon = c^{(p)}_0 + \sqrt{\varepsilon}c^{(p)}_{\frac{1}{2}} + \varepsilon c^{(p)}_1 + \mathcal{O}(\varepsilon^{\frac{3}{2}}).$$

[3] As stated here the formulae for the drift and diffusion matrix are only valid for correlation functions which are slightly more than integrable. When the autocorrelation function of the fast driving system is decaying but is only integrable, more complicated formulae apply; see [14] for details.

To this end, the expectation values appearing in the cumulants $\mathbb{E}_\varepsilon^{x_0,\mu}$ are expressed as

$$\mathbb{E}_\varepsilon^{x_0,\mu}\left[A(x(t),y(t))\right] = \int\int A(x,y)e^{\mathcal{L}_\varepsilon t}\delta_{x_0}(\mathrm{d}x)\mu(\mathrm{d}y)\,,$$

with the transfer operator $e^{\mathcal{L}_\varepsilon t}$ (also known as Frobenius-Perron operator) associated with the multi-scale system (1)-(2). This transfer operator can be expanded by successive application of the Duhamel-Dyson formula [4, 32], resulting in explicit expressions for the $c_j^{(p)}$. We find $c_0^{(1)} = c_1^{(1)} = 0$, $c_{\frac{1}{2}}^{(1)} = F(x_0)$, $c_0^{(2)} = \sigma^2$, $c_{\frac{1}{2}}^{(2)} = 0$, $c_0^{(3)} = c_1^{(3)} = 0$, $c_0^{(4)} = c_{\frac{1}{2}}^{(4)} = 0$ and $c_\varepsilon^{(p)} = \mathcal{O}(\varepsilon^{\frac{3}{2}})$ for $p > 4$, while the coefficients $c_1^{(2)}$, $c_{\frac{1}{2}}^{(3)}$ and $c_1^{(4)}$ depend non-trivially on the correlations of y (see appendix A for their expressions). Finally, by taking the inverse Fourier transform of χ_ε, we can formally expand the probability density $\pi_\varepsilon = \pi_\varepsilon^{(2)} + \mathcal{O}(\varepsilon^{\frac{3}{2}})$ with

$$\pi_\varepsilon^{(2)}(\mathrm{x}, t=\theta\varepsilon, x_0) = \mathbf{n}_{0,\sigma^2}(\mathrm{x})\left[1 + \sqrt{\varepsilon}\left(\frac{F(x_0)}{\sigma}H_1\left(\frac{\mathrm{x}}{\sigma}\right) + \frac{c_{\frac{1}{2}}^{(3)}}{3!\sigma^3}H_3\left(\frac{\mathrm{x}}{\sigma}\right)\right) + \varepsilon\left(\frac{F(x_0)^2 + c_1^{(2)}}{2\sigma^2}H_2\left(\frac{\mathrm{x}}{\sigma}\right)\right.\right.$$
$$\left.\left. + \frac{c_1^{(4)} + 4F(x_0)c_{\frac{1}{2}}^{(3)}}{4!\sigma^4}H_4\left(\frac{\mathrm{x}}{\sigma}\right) + \frac{{c_{\frac{1}{2}}^{(3)}}^2}{2(3!\sigma^3)^2}H_6\left(\frac{\mathrm{x}}{\sigma}\right)\right)\right]. \tag{6}$$

Here $H_n(\mathrm{x}) = (\mathrm{x} - \frac{\mathrm{d}}{\mathrm{dx}})^n 1$ are Hermite polynomials of degree n. It is readily seen from (6) that for $\varepsilon \to 0$, the homogenization limit $\lim_{\varepsilon\to 0}\pi_\varepsilon = \mathbf{n}_{0,\sigma^2}$ is recovered. For a derivation of the Edgeworth expansion and explicit formulae for the $c_j^{(p)}$ the reader is referred to [30]. For completeness we present in the Appendix the expressions for the Edgeworth expansion coefficients. Note that the expressions for the cumulant expansions as derived in [30] determine the form of the expansion, but are not sufficient to show that an Edgeworth expansion actually holds for a given class of dynamical systems. However, the numerical evidence presented below and in [30] suggests strongly that Edgeworth expansions hold for the model systems studied.

3.1. Numerical validation of the Edgeworth expansion. We now numerically demonstrate the validity of the Edgeworth expansion for a multi-scale system of the form (1)-(2). In particular, we consider

$$\dot{x} = \frac{1}{\varepsilon}f_0(y) + f_1(x) \tag{7}$$
$$\dot{y}_i = \frac{1}{\varepsilon^2}g_0(y) \tag{8}$$

with $y \in \mathbb{R}^N$, $f_1(x) = -\partial_x V(x)$, $V(x) = x^2(b^2x^2 - a^2)$, $g_1(x,y) = 0$, $g_0(y) = y_{i-1}(y_{i+1} - y_{i-2}) + R - y_i$ and $y_{N+i} = y_i$ for $1 \le i \le N$. The system consists of a single degree of freedom x in a symmetric double well potential V driven by a fast Lorenz '96 (L96) y-system. The L96 system was introduced to mimic atmospheric chaos in the midlatitudes [18]. The system (7)-(8) can therefore be viewed as a simple toy model of the ocean exhibiting two regimes, driven by a fast chaotic atmosphere. We take the classical parameters of Lorenz' with $N = 40$, $R = 8$ and choose $f_0(y) = \sigma_m\left(\frac{1}{5}\sum_{i=1}^{5} y_i^2 - C_0\right)$ where C_0 is chosen such that $\langle f_0\rangle = 0$. Randomness is introduced solely through a random choice of the initial condition y_0, distributed according to the physical invariant measure of the fast L96 system.

To demonstrate the validity of the Edgeworth expansion we show in Figure 1 the transition probabilities for the full multi-scale system (7)-(8) as well as those of the reduced homogenized system (3) and of the Edgeworth expansion (6). Whereas homogenization fails to approximate the transition probability (with a relative error in the skewness of 0.87), our Edgeworth approximation describes the statistics of the true system remarkably well. Note that the transition probability $\pi_\varepsilon^{(2)}$ is not a proper probability density function in the sense that it is not a non-negative function. The occurrence of negative values is due to the expansion of $\pi_\varepsilon^{(2)}$ in Hermite polynomials (cf. (6)). This implies that one cannot sample directly from the Edgeworth-approximated transition probability $\pi_\varepsilon^{(2)}$. However, as we will see in the next section, one can construct a dynamical system with expansion coefficients approximating those in $\pi_\varepsilon^{(2)}$, and this surrogate system can then be used to sample from a pdf which has the same Edgeworth expansion of the transition probability as the full multi-scale system.

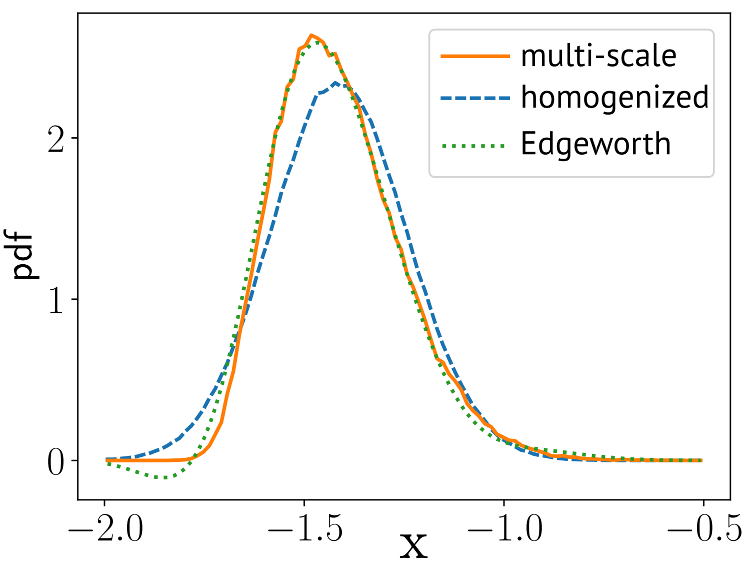

Fig. 1: Transition probability $\pi_\varepsilon(\mathrm{x}, t = 0.02, x_0 = -\sqrt{2})$ of the system (7)-(8) (labelled "multiscale") with $a = 1$, $b = 0.5$, $\varepsilon = 0.1$ and $\sigma_m = 0.1821$ (implying $\sigma = 1.25$), the Edgeworth expansion $\pi_\varepsilon^{(2)}$ (6) (labelled "Edgeworth") and the pdf of $X(t)$ in (3) (labelled "homogenized").

We now describe how the Edgeworth coefficients of Eqs. (7)-(8) are estimated numerically. For the case of the multi-scale Lorenz '96 system Eqs. (7)-(8) the formulae for the Edgeworth coefficients σ, $c_1^{(2)}$, $c_{\frac{1}{2}}^{(3)}$ and $c_1^{(4)}$ appearing in the transition probability $\pi_\varepsilon^{(2)}(x, t = \theta\varepsilon, x_0)$ (6) presented in the appendix yield

$$F = -\partial_{x_0} V(x_0)$$

$$\sigma^2 = \mu_{20}$$

$$c_1^{(2)} = -\theta\sigma^2\partial^2_{x_0}V(x_0) + \frac{1}{\theta}\,\mu_{21}$$

$$c_{\frac{1}{2}}^{(3)} = \frac{1}{\sqrt{\theta}}\,\mu_{30}$$

$$c_1^{(4)} = \frac{1}{\theta}\,\mu_{40}$$

where

$$\mu_{20} = 2\int_0^\infty C_2(\tau)\,\mathrm{d}\tau \tag{9}$$

$$\mu_{21} = -2\int_0^\infty \tau\, C_2(\tau)\,\mathrm{d}\tau \tag{10}$$

$$\mu_{30} = 6\int_0^\infty C_3(\tau_1,\tau_2)\,\mathrm{d}\tau_1\mathrm{d}\tau_2 \tag{11}$$

$$\mu_{40} = 6\mu_{20}\,\mu_{21} - 24\int_0^\infty \left(C_4(\tau_1,\tau_2,\tau_3) - C_2(\tau_1)C_2(\tau_3)\right)\,\mathrm{d}\tau_1\mathrm{d}\tau_2\mathrm{d}\tau_3 \tag{12}$$

with the two-point autocorrelation function $C_2(\tau) = \langle f_0(y)f_0(\varphi^\tau y)\rangle$, the three–point autocorrelation function $C_3(\tau_1,\tau_2) = \langle f_0(y)f_0(\varphi^{\tau_1}y)f_0(\varphi^{\tau_1+\tau_2}y)\rangle$ and the four–point autocorrelation function $C_4(\tau_1,\tau_2,\tau_3) = \langle f_0(y)f_0(\varphi^{\tau_1}y)f_0(\varphi^{\tau_1+\tau_2}y)f_0(\varphi^{\tau_1+\tau_2+\tau_3}y)\rangle$, where we recall that φ^t denotes the flow map of the fast dynamics.
The terms μ_{20}, μ_{21}, μ_{30} and μ_{40} can be calculated directly by estimating the correlation functions $C_{2,3,4}$. This, however, is computationally expensive to get accurate results. Here we estimate the terms as follows. As shown in [30], the Edgeworth coefficients appear as the coefficients of an expansion in t and ε of the cumulants of transition probabilities of the multi-scale system. If we were to set $V = 0$, the terms μ_{20}, μ_{21}, μ_{30} and μ_{40} are the leading order terms appearing in the Edgeworth expansion of the second, third and fourth cumulant. More specifically, for the system

$$\dot{\tilde{x}} = \frac{1}{\varepsilon}f_0(\tilde{y}) \tag{13}$$

$$\dot{\tilde{y}} = \frac{1}{\varepsilon^2}g_0(\tilde{y}) \tag{14}$$

with initial conditions $\tilde{x}(t=0) = \tilde{x}_0$ and $\tilde{y}(t=0) = \tilde{y}_0$, we can integrate the slow dynamics to obtain

$$\xi_\varepsilon := \frac{\tilde{x}(t=\varepsilon) - \tilde{x}_0}{\sqrt{\varepsilon}} = \sqrt{\varepsilon}z(\frac{1}{\varepsilon})$$

with $z(t) := \int_0^t f_0(y(\tau))\,\mathrm{d}\tau$. As shown in [30], the second, third and fourth cumulants of ξ_ε can be expanded in orders of $\sqrt{\varepsilon}$ as

$$\mathbb{E}_\varepsilon^{x_0,\mu}\left[\xi_\varepsilon^2\right] = \mu_{20} + \varepsilon\mu_{21} + \mathcal{O}(\varepsilon^2)$$

$$\mathbb{E}_\varepsilon^{x_0,\mu}\left[\xi_\varepsilon^3\right] = \sqrt{\varepsilon}\mu_{30} + \mathcal{O}(\varepsilon^{\frac{3}{2}})$$

$$\mathbb{E}_\varepsilon^{x_0,\mu}\left[\xi_\varepsilon^4\right] - 3\mathbb{E}_\varepsilon^{x_0,\mu}\left[\xi_\varepsilon^2\right]^2 = \varepsilon\mu_{40} + \mathcal{O}(\varepsilon^2)\,.$$

It follows by taking $t = \frac{1}{\varepsilon}$ that $\mu_2 := \mathbb{E}\left[z(t)^2\right]$, $\mu_3 := \mathbb{E}\left[z(t)^3\right]$ and $\mu_4 := \mathbb{E}\left[z(t)^4\right]$ scale with t as

$$\frac{\mu_2}{t} = \mu_{20} + \frac{\mu_{21}}{t} + \mathcal{O}(\frac{1}{t^2})$$

$$\frac{\mu_3}{t} = \mu_{30} + \mathcal{O}(\frac{1}{t})$$
$$\frac{\mu_4 - 3\mu_2^2}{t} = \mu_{40} + \mathcal{O}(\frac{1}{t})\,.$$

This suggests to perform a least squares fit of $\frac{\mu_2}{t}$, $\frac{\mu_3}{t}$ and $\frac{\mu_4-3\mu_2^2}{t}$ to a two-parameter family of functions $\ell(t) = a + b/t$. Denoting the result of the least squares fit of $\frac{\mu_2}{t}$ by $a_2^\star$ and $b_2^\star$, of $\frac{\mu_3}{t}$ by $a_3^\star$ and $b_3^\star$ and of $\frac{\mu_4-3\mu_2^2}{t}$ by $a_4^\star$ and $b_4^\star$, we can extract the leading order coefficients. From the fits we obtain $\mu_{20} = a_2^\star$ and $\mu_{21} = b_2^\star$, $\mu_{30} = a_3^\star$ and $\mu_{40} = a_4^\star$. Figure 2 shows the scaled cumulants of $z(t)$ together with their respective least squares fit of functions $\ell(t) = a + b/t$.

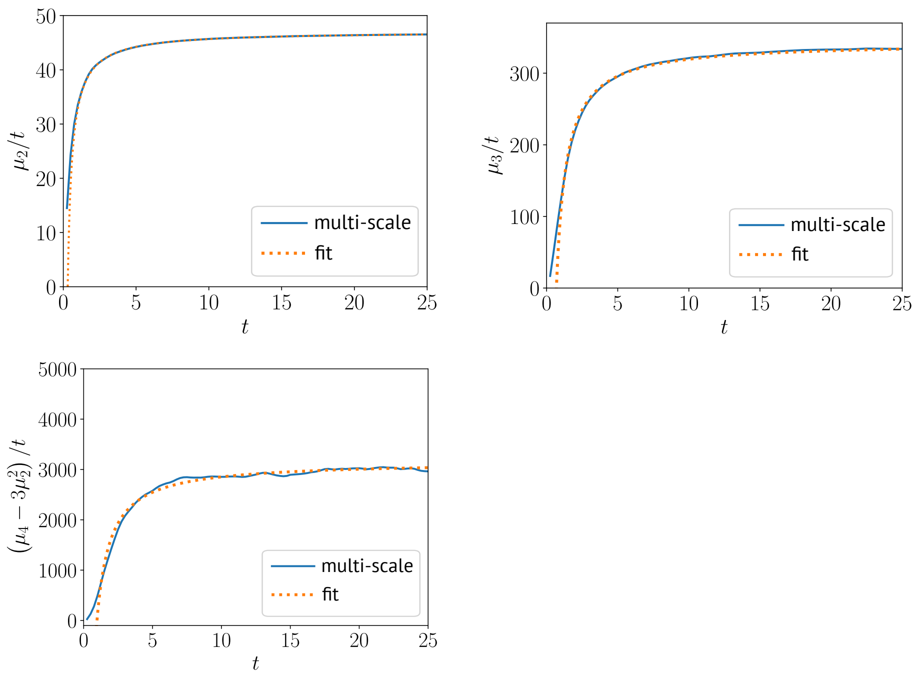

Fig. 2: Scaled cumulants of $z(t)$ for the system (13)-(14) with f_0 and g_0 as in (7)-(8). The smooth line represents a least squares fit to $\ell(t) = a + b/t$. Top left: second cumulant, top right: third cumulant, bottom: fourth cumulant.

4. The surrogate system. The Edgeworth expansion is universal in the sense that only a limited number of statistical properties of the fast system appear in the expansion. Therefore, the microscopic details of the fast y-dynamics are of no importance to the slow x-dynamics. As we have seen, one cannot sample directly form the Edgeworth expansion of the transition probability $\pi_\varepsilon^{(2)}$ since it is not a proper probability density function and involves negative values due to the expansion in Hermite polynomials (cf. (6)). However, we can construct a surrogate system such that the Edgeworth expansion of its transition probability, which we label $\pi_{\rm surr}^{(2)}$, closely approximates the expansion $\pi_\varepsilon^{(2)}$ of transition probabilities of the full multi-scale system. From the macroscopic point of view the y-dynamics can be substituted with a simpler surrogate system, as long as the statistical properties encoded in the Edgeworth expansion are preserved. This suggests a new way of performing stochastic model reduction for the slow dynamics: construct a class of simple surrogate systems $(X(t), \eta(t))$ dependent on a set of parameters $\mathfrak{p}_{\rm surr}$. Here $X \in \mathbb{R}^d$ denotes the slow variables, approximating the slow dynamics x in the multi-scale system (1)-(2), and $\eta \in \mathbb{R}^k$ with $k < N$ mimics the effect of the fast dynamics y. The functional form of the surrogate system, determining the evolution of $X(t)$ and $\eta(t)$, and the dimension k of the fast surrogate variables η are chosen sufficiently simple to allow for an explicit analytical expression of the Edgeworth expansion coefficients of the transition probability $\pi^{(2)}_{\rm surr}$ of the surrogate system. These coefficients will depend on the set of free parameters $\mathfrak{p}_{\rm surr}$ appearing in the surrogate system. Judiciously choosing the free parameters of the surrogate system $\mathfrak{p}_{\rm surr}$ allows us to match the Edgeworth corrections of the surrogate system to the observed Edgeworth corrections of the original multi-scale model we set out to model. This is achieved as follows: the transition probability of the surrogate slow variables X,

$$\pi_{\rm surr}({\rm x}, t = \theta\varepsilon, x_0) = \mathbb{P}\left(\frac{X(t) - X(0)}{\sqrt{t}} \in ({\rm x}, {\rm x} + {\rm dx}) \middle| X(0) = x_0 \right),$$

is approximated by the second order Edgeworth expansion $\pi_{\rm surr} = \pi^{(2)}_{\rm surr} + \mathcal{O}(\varepsilon^{\frac{3}{2}})$. The expression for the Edgeworth expansion of $\pi_{\rm surr}$ is the same as for π_ε given in (6). We denote the cumulant expansion coefficients for $\pi^{(2)}_{\rm surr}$ in (6) as $c_k^{(p,s)}$. The free parameters $\mathfrak{p}_{\rm surr}$ of the surrogate system are then determined by the constrained optimization, at a fixed time which we choose arbitrarily as $t = \varepsilon$,

$$\underset{\mathfrak{p}_{\rm surr}}{\arg\min} \left\| \pi^{(2)}_{\rm surr}({\rm x}, t = \varepsilon, x_0) - \pi^{(2)}_\varepsilon({\rm x}, t = \varepsilon, x_0) \right\| \tag{15}$$

of the L_2-norm with respect to x for fixed initial condition x_0 subject to the constraint of the exact matching of the leading order diffusivity σ (5) and drift F (4). A further appropriately weighted norm w.r.t. x_0 (e.g. weighted with the invariant measure restricted to x) can be taken to ensure one set of parameter values for all x_0. Since σ and F determine the limiting system (3), this constraint assures that the surrogate system and the full deterministic system have the same homogenized limit. Using the Edgeworth expansions for both $\pi^{(s)}$ and π_ε, we have, if $c_0^{(2,s)} = c_0^{(2)} = \sigma^2$ and $c_{\frac{1}{2}}^{(1,s)} = c_{\frac{1}{2}}^{(1)} = F$, that

$$\|\pi^{(2)}_{\rm surr}({\rm x}, \varepsilon, x_0) - \pi^{(2)}_\varepsilon({\rm x}, \varepsilon, x_0)\| = \varepsilon \mathcal{E}^{(1)}(x_0) + \varepsilon^2 \mathcal{E}^{(2)}(x_0)\,, \tag{16}$$

with

$$\mathcal{E}^{(1)}(x_0) = \frac{15\,\kappa_3^2}{16\sqrt{\pi}\sigma}$$

$$\mathcal{E}^{(2)}(x_0) = \frac{3\left(16\,\kappa_2^2 - 80\,\kappa_2\kappa_4 + 140\,\kappa_4^2 + 3465\,\kappa_6^2 + 140\,(2\,\kappa_2 - 9\,\kappa_4)\kappa_6\right)}{128\sqrt{\pi}\sigma}\,,$$

where the coefficients

$$\kappa_2 = \frac{c_1^{(2)} - c_1^{(2,s)}}{2\sigma^2}, \quad \kappa_3 = \frac{c_{\frac{1}{2}}^{(3)} - c_{\frac{1}{2}}^{(3,s)}}{6\sigma^3}, \quad \kappa_4 = \frac{c_1^{(4)} - c_1^{(4,s)}}{24\sigma^4}, \quad \kappa_6 = \frac{{c_{\frac{1}{2}}^{(3)}}^2 - {c_{\frac{1}{2}}^{(3,s)}}^2}{72\sigma^6}$$

are given in terms of the expansion coefficients of the original multiscale system and of the surrogate system. The expansion coefficients of the original multi-scale system $c_k^{(p)}$ are determined numerically through evaluation of their expressions for long-time numerical simulations, as described in Section 3.1. Their surrogate counterparts $c_k^{(p,s)}$ can be determined analytically as a function of the free parameters $\mathfrak{p}_{\rm surr}$. This then allows to evaluate the error terms in (16). The constrained optimization problem (15) can then be solved by varying the surrogate parameters $\mathfrak{p}_{\rm surr}$.

We consider here the following family of surrogate models for the multi-scale system (1)-(2)

$$\dot{X} = \frac{1}{\varepsilon} f_0^{(s)}(X,\eta) + F(X) + f_1^{(s)}(X,\eta) \tag{17}$$

$$d\eta = -\frac{1}{\varepsilon^2}\Gamma^{(s)}\eta\, \mathrm{d}t + \frac{\sigma^{(s)}}{\varepsilon}\mathrm{d}W_t + \frac{1}{\varepsilon} g_1^{(s)}(X,\eta)\,. \tag{18}$$

The fast process $\eta(t)$ is a k-dimensional Ornstein-Uhlenbeck process with $\Gamma_{ij}^{(s)} = \gamma_i \delta_{ij}$ and $\sigma_{ij}^{(s)} = \zeta_i \delta_{ij}$. The noise is here, different to the homogenized diffusive limits, coloured and enters the slow dynamics in an integrated way, allowing for non-trivial memory.

The vector fields $f_0^{(s)}$, $f_1^{(s)}$ and $g_1^{(s)}$ of the surrogate system are chosen to be polynomial

$$f_l^{(s)}(X,\eta) \;=\; \sum_{|\alpha|<\alpha_l, |\beta|<\beta_l} a_l^{(\alpha,\beta)} X^\alpha\, \eta^\beta \tag{19}$$

$$g_1^{(s)}(X,\eta) = \sum_{|\alpha|<\alpha_2, |\beta|<\beta_2} a_2^{(\alpha,\beta)} X^\alpha\, \eta^\beta \tag{20}$$

for $l=0,1$. The degree of the polynomials α_l and β_l, $l=0,1,2$, and the dimensionality of the surrogate process k are chosen as the smallest degree and dimension which still allow the surrogate system to capture the statistical features of the vector field $f_0(x,y)$ of the original multi-scale system (1)-(2).

4.1. Surrogate model for the Lorenz '96 driven system. To test the ability of the Edgeworth expansion-based surrogate model (17)-(18) to approximate the statistics of the slow variable x, we first consider the multi-scale system (7)-(8). Since $g_1 = 0$ in this case, we set $\alpha_2 = \beta_2 = 0$. Furthermore, we find that $k=1$, $\alpha_1 = \beta_1 = 0$, $\alpha_0 = 3$, $\beta_0 = 1$ are sufficient. The Edgeworth coefficients (9)-(12) for the surrogate model can be explicitly calculated. We obtain

$$c_0^{(2,s)} = \frac{11\, {a_0^{(3,0)}}^2 \zeta_1^6 + 4\, {a_0^{(1,0)}}^2 \gamma_1^2 \zeta_1^2 + 2\left({a_0^{(2,0)}}^2 + 6\, a_0^{(1,0)} a_0^{(3,0)}\right)\gamma_1 \zeta_1^4}{4\,\gamma_1^4} \tag{21}$$

$$c_{2,-1}^{(2,s)} = -\frac{29\, {a_0^{(3,0)}}^2 \zeta_1^6 + 12\, {a_0^{(1,0)}}^2 \gamma_1^2 \zeta_1^2 + 3\left({a_0^{(2,0)}}^2 + 12\, a_0^{(1,0)} a_0^{(3,0)}\right)\gamma_1 \zeta_1^4}{12\,\gamma_1^5} \tag{22}$$

$$c_{1,-\frac{1}{2}}^{(3,s)} = \frac{3\left(22\, a_0^{(2,0)} {a_0^{(3,0)}}^2 \zeta_1^8 + 4\, {a_0^{(1,0)}}^2 a_0^{(2,0)} \gamma_1^2 \zeta_1^4 + \left({a_0^{(2,0)}}^3 + 18\, a_0^{(1,0)} a_0^{(2,0)} a_0^{(3,0)}\right)\gamma_1 \zeta_1^6\right)}{2\,\gamma_1^6} \tag{23}$$

$$
\begin{aligned}
c_0^{(4,s)} &= 6\, c_0^{(2,s)} c_1^{(2,s)} + \Big(48\, \gamma_1^4 {a_0^{(1,0)}}^4 + 420\, \gamma_1^3 \zeta_1^2 {a_0^{(1,0)}}^2 {a_0^{(2,0)}}^2 \\
&\quad + 66\, \gamma_1^2 \zeta_1^4 {a_0^{(2,0)}}^4 + 480\, \gamma_1^3 \zeta_1^2 {a_0^{(1,0)}}^3 a_0^{(3,0)} \\
&\quad + 2268\, \gamma_1^2 \zeta_1^4 a_0^{(1,0)} {a_0^{(2,0)}}^2 a_0^{(3,0)} + 1976\, \gamma_1^2 \zeta_1^4 {a_0^{(1,0)}}^2 {a_0^{(3,0)}}^2 \\
&\quad + 3259\, \gamma_1 \zeta_1^6 {a_0^{(2,0)}}^2 {a_0^{(3,0)}}^2 \\
&\quad + 3912\, \gamma_1 \zeta_1^6 a_0^{(1,0)} {a_0^{(3,0)}}^3 + 3109\, \zeta_1^8 {a_0^{(3,0)}}^4 \Big) \frac{\zeta_1^4}{8\, \gamma_1^9} \, .
\end{aligned}
\tag{24}
$$

The parameter $a_0^{(0,0)} = -a_0^{(0,2)} \zeta_1^2/(2\gamma_1)$ is fixed by requiring the centering condition $\langle f_0^{(s)} \rangle \equiv 0$. The remaining parameters for the surrogate system are determined by constrained minimization of (15) using sequential least squares programming as implemented in the SciPy library.

Figure 3 shows the invariant measure and the third moment of the slow dynamics of the multiscale Lorenz system (7) with a moderate time scale separation $\varepsilon = 0.15$, as well as of the homogenized equation (3) and of the surrogate process (17)-(18). It is clearly seen that the stochastic model reduction based on the Edgeworth expansion captures the nontrivial non-Gaussian behaviour of the full slow dynamics very well, whereas the homogenized equation converges to a Gaussian with a zero third moment. Note that the surrogate naturally supports an invariant measure from which one can sample, unlike the expansion $\pi_{\rm surr}^{(2)}$ which was used for its construction. Figure 4 shows the second and fourth cumulants, for the full multi-scale system (7) and for the homogenized equation (3) as well as for the surrogate process (17)-(18). For the second moment we show the long-time behaviour where homogenization matches well, as well as the intermediate time evolution where the Edgeworth expansion clearly outperforms the homogenized result. For the fourth moment the classical homogenization results fail to capture the long-time and the intermediate time temporal evolution whereas the Edgeworth expansion closely follows the true evolution of the moments, capturing the non-Gaussian behaviour of the slow dynamics in the moderate timescale separation case.

4.2. Surrogate model for a triad system. We now treat a multiscale system that includes a non-zero backcoupling term g_1. In particular, we consider the triad model

$$
\mathrm{d}x = \frac{B_0}{\varepsilon} y_1 y_2 \, \mathrm{d}t \tag{25}
$$

$$
\mathrm{d}y_1 = \frac{B_1}{\varepsilon} y_2 x \, \mathrm{d}t - \frac{\gamma_1^{(t)}}{\varepsilon^2} y_1 \, \mathrm{d}t + \frac{\sigma_1^{(t)}}{\varepsilon} \, \mathrm{d}W_1 \tag{26}
$$

$$
\mathrm{d}y_2 = \frac{B_2}{\varepsilon} x y_1 \, \mathrm{d}t - \frac{\gamma_2^{(t)}}{\varepsilon^2} y_2 \, \mathrm{d}t + \frac{\sigma_2^{(t)}}{\varepsilon} \, \mathrm{d}W_2 \, . \tag{27}
$$

This model has been used as a low-dimensional toy model for fluid flows with quadratic nonlinearities [20]. The triad system allows for an explicit calculation of the homogenized system and Edgeworth coefficients (see Appendix A for the general formulae). For the zero order homogenized equations, we obtain for the drift F and diffusion coefficient σ

$$
F(X) = \Theta X \, , \tag{28}
$$

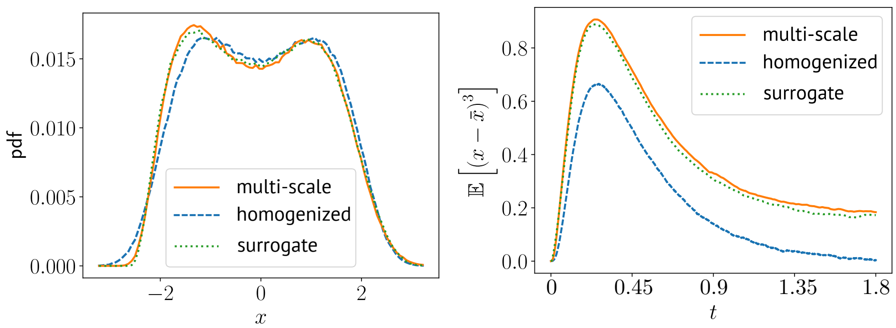

Fig. 3: The invariant measure (left) and third moment (right) for x of the multi-scale Lorenz system (7)-(8) with $\varepsilon = 0.15$, $a = 1$, $b = 2/3$ and $\sigma_m = 0.48567$ (implying $\sigma = 10/3$), the homogenized equation (3) and the surrogate process (17)-(18). The parameters of the surrogate process are obtained by the method in Section 4 as $\gamma_1 = 2.479$, $\zeta_1 = 25.793$, $a_0^{(0,3)} = -9.7467\ 10^{-3}$, $a_0^{(0,2)} = 19.72\ 10^{-2}$, $a_0^{(0,1)} = 7.1933$ and $a_0^{(0,0)} = -a_0^{(0,2)}\zeta_1^2/(2\gamma_1)$.

$$\sigma^2 = 2\frac{B_0^2\sigma_{1\infty}^2\sigma_{2\infty}^2}{\gamma_1^{(t)}+\gamma_2^{(t)}} \tag{29}$$

with $\Theta = \frac{B_0}{\left(\gamma_1^{(t)}+\gamma_2^{(t)}\right)^2}(B_1\sigma_{2\infty}^2 + B_2\sigma_{1\infty}^2)$ and $\sigma_{i\infty}^2 = \frac{\sigma_i^{(t)2}}{2\gamma_i^{(t)}}$.

For the Edgeworth coefficients up to order $\varepsilon^{3/2}$ we find

$$c_{2,-1}^{(2)} = -\frac{\sigma^2}{\gamma_1^{(t)}+\gamma_2^{(t)}}$$

$$c_{0,1}^{(2)} = \sigma^2\Theta + \Theta^2 x_0^2$$

$$c_{2,-1}^{(4)} = 6\sigma^2 c_{2,-1}^{(2)} + 6\frac{\sigma^4}{\gamma_1^{(t)}+\gamma_2^{(t)}}\left(\frac{(\gamma_1^{(t)}+\gamma_2^{(t)})^2}{\gamma_1^{(t)}\gamma_2^{(t)}}+2\right)$$

and $c_{0,\frac{1}{2}}^{(3)} = c_{1,-\frac{1}{2}}^{(3)} = c_{1,0}^{(2)} = c_{0,1}^{(4)} = c_{1,0}^{(4)} = 0$.

Since g_1 is now non-zero, we construct a surrogate system with non-zero α_2. We find that a simple surrogate system of the form

$$\mathrm{d}x = \frac{1}{\varepsilon} f_0^{(s)}(y)\,\mathrm{d}t \tag{30}$$

$$\mathrm{d}y = \frac{a_2^{(1,0)}}{\varepsilon} x\,\mathrm{d}t - \frac{\gamma_1}{\varepsilon^2} y\,\mathrm{d}t + \frac{\zeta_1}{\varepsilon}\mathrm{d}W \tag{31}$$

with

$$f_0^{(s)}(y) = a_0^{(3,0)}y^3 + a_0^{(2,0)}y^2 + a_0^{(1,0)}y + a_0^{(0,0)}$$

gives a good approximation. For the zero order homogenized equations of the surrogate, we obtain a drift $F^{(s)}(x) = \Theta^{(s)}x$ with $\Theta^{(s)} = \frac{a_2^{(1,0)}}{2\gamma_1^2}(2\gamma_1 a_0^{(1,0)} + 3a_0^{(3,0)}\zeta_1^2)$ and diffusion ${\sigma^{(s)}}^2 = \frac{11\,{a_0^{(3,0)}}^2\zeta_1^6 + 4\,{a_0^{(1,0)}}^2\gamma_1^2\zeta_1^2 + 2\left({a_0^{(2,0)}}^2 + 6\,a_0^{(1,0)}a_0^{(3,0)}\right)\gamma_1\zeta_1^4}{4\,\gamma_1^4}$. The non-zero

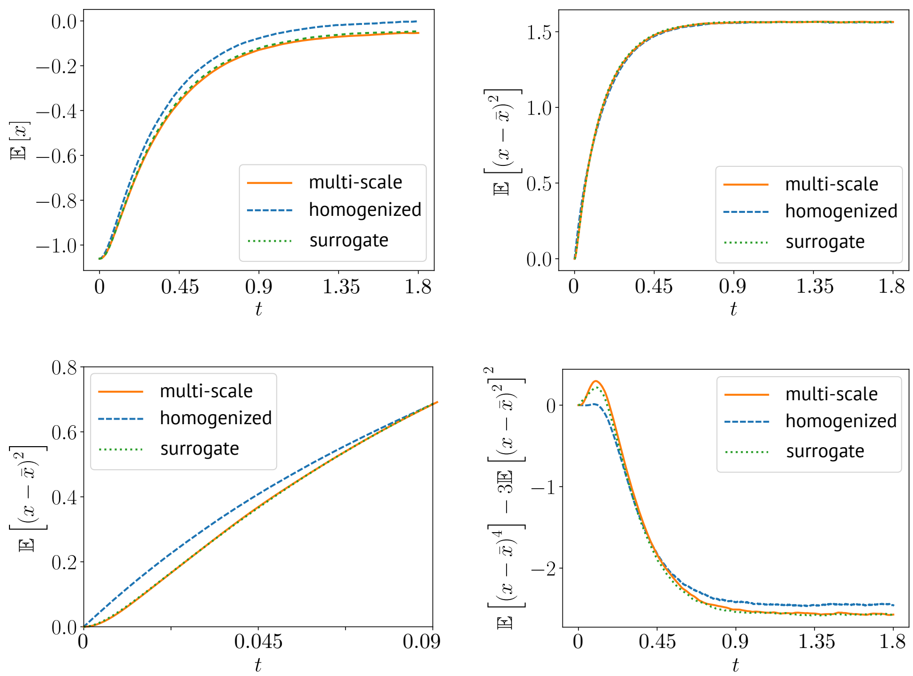

Fig. 4: The first moment (top left), second moment (over long times (top right) and over intermediate times (bottom left)) and fourth cumulant over long times (bottom right) for x of the multi-scale Lorenz system as a function of time. Parameters are $\varepsilon = 0.15$, $a = 1$, $b = 2/3$ and $\sigma_m = 0.48567$ (implying $\sigma = 10/3$). We show results for the full multi-scale system (Eqn (7)), the homogenized equation (Eqn (3)) and the surrogate process (Eqns (17)-(18)). The parameters of the surrogate process are obtained by the method in Section 3 as $\gamma_1 = 2.479$, $\zeta_1 = 25.793$, $a_0^{(0,3)} = -9.7467\ 10^{-3}$, $a_0^{(0,2)} = 19.72\ 10^{-2}$, $a_0^{(0,1)} = 7.1933$ and $a_0^{(0,0)} = -a_0^{(0,2)}\zeta_1^2/(2\gamma_1)$.

Edgeworth coefficients of the surrogate system are given by those in Eqns. (21)-(24) and

$$c_{0,1}^{(2,s)} = {\sigma^{(s)}}^2\Theta^{(s)} + {\Theta^{(s)}}^2 x_0^2$$

$$c_{1,0}^{(2,s)} = a_2^{(1,0)} x \frac{91 a_0^{(2,0)} a_0^{(3,0)} \zeta_1^4 + 42 a_0^{(1,0)} a_0^{(2,0)} \gamma_1 \zeta_1^2}{12\gamma_1^4}$$

Figure 5 shows the mean and standard deviation over time of an ensemble of realizations starting from a fixed initial condition $x_0 = -1$ for the multiscale triad system (25)-(27), the limiting homogenized equation (3) with drift (28) and diffusion (29) and the surrogate model (30)-(31). The mean and standard deviation are indistinguishable from those of the multiscale triad system, whereas the standard deviation of the homogenized equation exhibits significant deviations from that of the original triad system.

5. Discussion. We developed a new framework in which to perform stochastic model reduction of multi-scale systems with moderate time scale separation. We

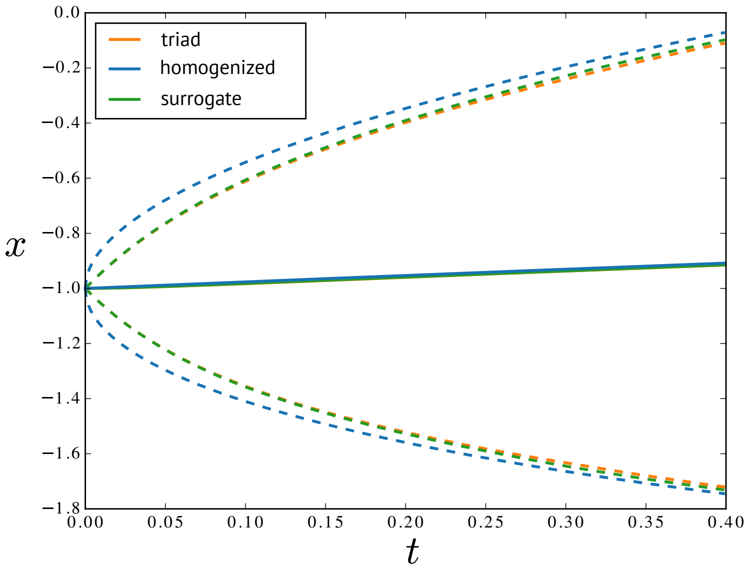

Fig. 5: Mean and standard deviation of the triad model (25)-(27), the homogenized system (3) and the surrogate system (30)-(31). The solid lines represent the mean of the sample, while the upper and lower dashed lines represent the mean plus or minus two standard deviations, respectively. The parameters of the triad model are $B_0 = -0.75$, $B_1 = -0.25$, $B_2 = 1$, $\gamma_1^{(t)} = 4/3$, $\sigma_1^{(t)} = \sqrt{8/3}$, $\gamma_2^{(t)} = 1$, $\sigma_2^{(t)} = \sqrt{2}$, $\varepsilon = 0.25$. The parameters of the surrogate model are $\gamma_1 = 2.166$, $\zeta_1 = 1.243$, $a_0^{(3,0)} = 0.786$, $a_0^{(2,0)} = -5.6\ 10^{-6}$, $a_0^{(1,0)} = 0.301$ and $a_2^{(1,0)} = -0.4569$.

showed how Edgeworth expansions can be used to construct reduced models for the slow dynamics of a chaotic deterministic multi-scale model. The surrogate system implies a non-Markovian effective slow dynamics, where the noise enters the slow dynamics in an integrated fashion. This reflects the memory effects in slow-fast systems with finite time-scale separation, where the fast dynamics has not yet sufficiently equilibrated on a slow characteristic time scale, preventing the homogenized Markovian limit. We considered a family of surrogate models where the free parameters were chosen to match the Edgeworth expansion of the original multi-scale model under consideration. The degree of the surrogate model was chosen by assuring to have the lowest possible order of the polynomials while still allowing for the surrogate system to capture the overall statistical features of the full multi-scale system. Matching the Edgeworth expansion then singles out the optimal member in the prescribed class.
We remark that the Edgeworth expansion is based on the transition probability on the intermediate time scale. In some applications, such as weather forecasting, one is interested in the transitional dynamics and their statistical modelling rather than in the long term statistical behaviour. In this situation Edgeworth expansions allow for a faithful description of the effects of finite time scale separation. The aim of the reduced model in other applications, however, may be to describe the statistical behaviour on the longer diffusive time scale, for example in climate science. We observe that in the system considered here, matching the short time transition probabilities translates into a more reliable description of the long time statistics as well. Although

this property may not hold in general, we expect it to hold in sufficiently smooth systems.

Our framework is not limited to deterministic continuous time systems. It can be extended to stochastic multi-scale systems and to discrete time maps which would allow the study of numerical integrators and their statistical limiting behaviour of resolved modes. More importantly, Edgeworth approximations can be determined from observational data; this allows for the application to systems with high complexity prohibiting an analytical estimation of the Edgeworth corrections. This opens up the door to perform mathematically sound stochastic model reductions for real-world problems. Furthermore, Edgeworth approximations are not limited to multi-scale systems. As an extension of the CLT, they can be used to study finite size effects to the thermodynamic limit of weakly coupled systems such as Kac-Zwanzig heat baths for distinguished particles [6, 31, 5].

Acknowledgments. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^oPIOF-GA-2013-626210. We thank Ben Goldys and Françoise Pène for enlightening discussions and comments.

Appendix A. Cumulant expansion for slow-fast system.

In [30] we derived expression for the expansion in ε of the cumulants of the slow variable x in the slow-fast system (1)-(2).

The first cumulant is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(1)} = \sqrt{t}\, c^{(1)}_{0,\frac{1}{2}}\,, \tag{32}$$

where

$$c^{(1)}_{0,\frac{1}{2}} = F(x_0) = \langle f_1 \rangle - \langle f_0 \mathcal{L}^{-1}_{0\perp} \partial_x f_0 \rangle - \langle (g_1 \partial_y) \mathcal{L}^{-1}_{0\perp} f_0 \rangle\,. \tag{33}$$

Upon explicit substitution of the intermediate time scaling $t = \varepsilon\theta$, with fixed θ, this becomes

$$c^{(1)} = \sqrt{\varepsilon\theta}\, c^{(1)}_{0,\frac{1}{2}}\,. \tag{34}$$

The second cumulant is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(2)} = m^{(2)} = c^{(2)}_0 + t\, c^{(2)}_{0,1} + \frac{\varepsilon^2}{t} c^{(2)}_{2,-1} + \varepsilon\, c^{(2)}_{1,0}. \tag{35}$$

Upon explicit substitution of the intermediate time scaling $t = \varepsilon\theta$, with fixed θ, this becomes

$$c^{(2)} = m^{(2)} = c^{(2)}_0 + \varepsilon c^{(2)}_1\,, \tag{36}$$

with $c^{(2)}_1 = \theta\, c^{(2)}_{0,1} + \frac{1}{\theta} c^{(2)}_{2,-1} + c^{(2)}_{1,0}$. The $\mathcal{O}(1)$ contribution is given by the homogenized Green-Kubo formula (5)

$$c^{(2)}_0 = \sigma^2 = -2\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0 \rangle$$

and higher-order contributions are given by

$$c^{(2)}_{0,1} = \frac{1}{2}\,\sigma^2 \left(\frac{\partial \sigma}{\partial x}\right)^2 + \frac{1}{2}\,\sigma^3 \frac{\partial^2 \sigma}{\partial x^2} + \sigma^2 \frac{\partial F}{\partial x} + F\sigma \frac{\partial \sigma}{\partial x} + F^2 \tag{37}$$

$$c^{(2)}_{2,-1} = -2\langle f_0 \mathcal{L}^{-2}_{0\perp} f_0\rangle \tag{38}$$

$$\begin{aligned} c^{(2)}_{1,0} = &-2\langle f_0 \mathcal{L}^{-1}_{0\perp} f_1\rangle - 2\langle f_1 \mathcal{L}^{-1}_{0\perp} f_0\rangle + 2\langle f_0 \mathcal{L}^{-1}_{0\perp}\partial_x f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \\ &+ 4\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp}\partial_x f_0\rangle + 2\langle f_0 \mathcal{L}^{-1}_{0\perp}(g_1\partial_y)\mathcal{L}^{-1}_{0\perp} f_0\rangle \\ &+ 2\langle (g_1\partial_y)\mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle. \end{aligned} \tag{39}$$

Here $\mathcal{L}^{-1}_{0\perp}$ denotes the invertible operator whose inverse is the restriction of $\mathcal{L}_0$ to the space orthogonal to the projection onto the invariant measure $\mu^{(0)}_{x_0}$

The third moment and its cumulant are given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(3)} = m^{(3)} = \sqrt{t}\, c^{(3)}_{0,\frac{1}{2}} + \frac{\varepsilon}{\sqrt{t}}\, c^{(3)}_{1,-\frac{1}{2}}\,. \tag{40}$$

Upon explicit substitution of the intermediate time scaling $t = \varepsilon\theta$, with fixed θ, this becomes

$$c^{(3)} = m^{(3)} = \sqrt{\varepsilon} c^{(3)}_{\frac{1}{2}}\,, \tag{41}$$

with

$$c^{(3)}_{\frac{1}{2}} = \sqrt{\theta}\, c^{(3)}_{0,\frac{1}{2}} + \frac{1}{\sqrt{\theta}}\, c^{(3)}_{1,-\frac{1}{2}} \tag{42}$$

$$c^{(3)}_{0,\frac{1}{2}} = 6\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \frac{\partial}{\partial x}\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \tag{43}$$

$$c^{(3)}_{1,-\frac{1}{2}} = 6\left\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0\right\rangle\,. \tag{44}$$

The fourth cumulant is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(4)} = t\, c^{(4)}_{0,1} + \varepsilon\, c^{(4)}_{1,0} + \frac{\varepsilon^2}{t}\, c^{(4)}_{2,-1}\,. \tag{45}$$

Upon explicit substitution of the intermediate time scaling $t = \varepsilon\theta$, with fixed θ, this becomes

$$c^{(4)} = \varepsilon c^{(4)}_1 \tag{46}$$

with

$$c^{(4)}_1 = \theta\, c^{(4)}_{0,1} + c^{(4)}_{1,0} + \frac{1}{\theta}\, c^{(4)}_{2,-1} \tag{47}$$

$$c^{(4)}_{0,1} = -24\,\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \left(\frac{\partial}{\partial x}\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle\right)^2 - 16\,\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle^2 \frac{\partial^2}{\partial x^2}\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \tag{48}$$

$$c^{(4)}_{1,0} = -24\,\langle \frac{\partial}{\partial x} f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle - 36\,\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \frac{\partial}{\partial x}\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle \tag{49}$$

$$c^{(4)}_{2,-1} = 24\left(\langle f_0 \mathcal{L}^{-2}_{0\perp} f_0\rangle\langle f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle - \langle f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0 \mathcal{L}^{-1}_{0\perp} f_0\rangle\right). \tag{50}$$

REFERENCES

[1] Y. Aït-Sahalia, *Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach*, Econometrica, 70 (2002), pp. 223–262, doi:10.1111/1468-0262.00274.
[2] R. N. Bhattacharya and R. R. Rao, *Normal approximation and asymptotic expansions*, vol. 64 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010.
[3] W. E and B. Engquist, *The heterogeneous multiscale methods*, Comm. Math. Sci., 1 (2003), pp. 87–132.
[4] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids*, Cambridge University Press, 2008.
[5] G. W. Ford and M. Kac, *On the quantum Langevin equation*, J. Statist. Phys., 46 (1987), pp. 803–810.
[6] G. W. Ford, M. Kac, and P. Mazur, *Statistical mechanics of assemblies of coupled oscillators*, J. Mathematical Phys., 6 (1965), pp. 504–515.
[7] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–127.
[8] 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.
[9] 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).
[10] F. Götze and C. Hipp, *Asymptotic distribution of statistics in time series*, Ann. Statist., 22 (1994), pp. 2062–2088, doi:10.1214/aos/1176325772, http://dx.doi.org/10.1214/aos/1176325772.
[11] L. Hervé and F. Pène, *The Nagaev-Guivarc'h method via the Keller-Liverani theorem*, Bull. Soc. Math. France, 138 (2010), pp. 415–489.
[12] S. C. L. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, *Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems*, Annual Review of Physical Chemistry, 62 (2011), pp. 41–64, doi:10.1146/annurev-physchem-032210-103335.
[13] D. Kelly and I. Melbourne, *Deterministic homogenization for fast-slow systems with chaotic noises*, arXiv:1409.5748 [math.PR], (2014).
[14] D. Kelly and I. Melbourne, *Deterministic homogenization for fast–slow systems with chaotic noise*, Journal of Functional Analysis, 272 (2017), pp. 4063–4102.
[15] 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.
[16] 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.
[17] T. G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, Journal of Functional Analysis, 12 (1973), pp. 55–67.
[18] E. N. Lorenz, *Predictability - a problem partly solved*, in Predictability, T. Palmer, ed., European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, UK, 1996.
[19] 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.
[20] 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.
[21] 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.
[22] I. Melbourne and M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Commun. Math. Phys., 260 (2005), pp. 131–146.
[23] 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.
[24] I. Melbourne and M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Annals of Probability, 37 (2009), pp. 478–505.
[25] I. Melbourne and A. Stuart, *A note on diffusion limits of chaotic skew-product flows*, Nonlinearity, 24 (2011), pp. 1361–1367.
[26] T. Palmer and P. Williams, eds., *Stochastic Physics and Climate Modelling*, Cambridge University Press, Cambridge, 2010.
[27] G. C. Papanicolaou, *Some probabilistic problems and methods in singular perturbations*,

Rocky Mountain Journal of Mathematics, 6 (1976), pp. 653–674.
[28] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, New York, 2008.
[29] W. F. Thompson, R. A. Kuske, and A. H. Monahan, *Reduced α-stable dynamics for multiple time scale systems forced with correlated additive and multiplicative gaussian white noise*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 27 (2017), p. 113105, doi:10.1063/1.4985675.
[30] J. Wouters and G. Gottwald, *Edgeworth expansions for slow-fast systems with finite time-scale separation*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 475 (2019), p. 20180358, doi:10.1098/rspa.2018.0358.
[31] R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Stat. Phys., 9 (1973), pp. 215–220.
[32] R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, Oxford, 2001.