Dynamical Systems with Time Scale Separation: Averaging, Stochastic Modelling, and Central Limit Theorems

Christian Rödenbeck*, Christian Beck**, and Holger Kantz*

Abstract. Considering dynamical systems involving processes on both slow and fast time scales, we deal with two methods to obtain reduced evolution equations for the slow variables alone: While *averaging* yields effective models for prediction, issues like variablity might profit from *stochastic modelling*. Rigorous results are available only in the limit of an infinite ratio between the two time scales. In a numerical case study, we show that reduced models obtained by averaging may possess good predictive skill even far from the region of applicability of the Averaging Theorem (time scale ratio only around 10). Stochastic modelling of the same numerical example does not predict better, but it additionally provides information on the prediction error. For a practical implementation of a stochastic model, approximation of the fast variables by Gaussian white noise is desirable. We review some recent rigorous results on Central Limit Theorems and the way how deterministic chaotic dynamical systems can approach a (stochastic) Langevin process in the limit of infinitely separated time scales. Again, a numerical example indicates their relevance also under less idealized conditions.

1. Introduction

Atmospheric dynamics involves processes on an enormous range of time and length scales. In climate research, interest focuses on the largest of them, but all the small-scale dynamics ('weather') exerts essential influence on the large-scale evolution as well.

As a prototype of such a situation, consider a dynamical system with two distinct time scales: its state vector consists of 'slow' variables (observables) $\mathbf{x}_t = \mathbf{x}(t)$ and 'fast' variables $\mathbf{y}_t = \mathbf{y}(t)$:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x},\mathbf{y}), \\ \dot{\mathbf{y}} &= \frac{1}{\varepsilon}\mathbf{g}(\mathbf{x},\mathbf{y}). \end{aligned} \tag{1}$$

Here, the right-hand sides $\mathbf{f}$ and $\mathbf{g}$ are assumed to be of similar order of magnitude, such that ε determines the ratio between the 'fast' and the 'slow' time scales. Since

we envisage a situation where the variables $\mathbf{y}$ cannot be resolved, we seek to obtain an evolution equation of the slow variables $\mathbf{x}$ alone.

As the first possibility considered here, this can be achieved by *averaging* (Section 2). As a rigorous result, the Averaging Theorem [1, 2] bounds the prediction error of such a reduced model in the limit $\varepsilon \to 0$ where the ratio between the two time scales becomes infinite. Since such an extreme time scale separation cannot be expected in practice, we report results of a numerical case study indicating that averaging might remain a valuable tool even for only moderately separated time scales.

Especially in the context of variability of atmospheric behaviour, a more appropriate description is expected from *stochastic modelling* [3, 4, 5]. Stochasticity naturally arises from the uncertainty in the unresolved degrees of freedom, e.g., from drawing the (unresolved) initial conditions $\mathbf{y}_0$ of the fast variables randomly (according to their invariant measure). While we will see in Section 3 that, in our numerical example, such a stochastic model does not outperform the averaged model for prediction, it indeed gives information about the system variability. We describe the emergence of stochastic terms ('noise sources') in a reduced stochastic equation of motion of the observables.

Not surprisingly, the stochastic terms are much more complicated than the usual additive Gaussian white noise. Nevertheless, in Section 4 we will review some recent rigorous results showing how generalizations of the Central Limit Theorem indeed guarantee the equivalence of certain suitably rescaled deterministic chaotic processes and Gaussian white noise, again in the limit $\varepsilon \to 0$. One can also get statements about the probability densities of fluctuations of the slow variables. The relevance in practice is demonstrated by the numerical example system.

2. Average Skill of an Averaged Model

2.1. Averaging

As a possibility to remove the unresolved fast degrees of freedom $\mathbf{y}$ from the model (1), define an averaged right-hand side,

$$\bar{\mathbf{f}}(\mathbf{x}) := \langle \mathbf{f}(\mathbf{x}, \mathbf{y}) \rangle_{\mathbf{y}|\mathbf{x}} \equiv \int \mathbf{f}(\mathbf{x}, \mathbf{y}) \mathrm{d}\mu(\mathbf{y}|\mathbf{x}), \tag{2}$$

where $\mu(\mathbf{y}|\mathbf{x})$ denotes the invariant measure[1] of the fast variables, conditioned on the slow ones, and the integral runs over the entire space of the fast variables. Then

$$\dot{\bar{\mathbf{x}}} = \bar{\mathbf{f}}(\bar{\mathbf{x}}) \tag{3}$$

[1] If an invariant probability density exists (which, however, cannot be assumed in general for chaotic systems), it is related to the measure by $\mathrm{d}\mu(\mathbf{y}|\mathbf{x}) = \varrho(\mathbf{y}|\mathbf{x})\mathrm{d}\mathbf{y}$ where $\mathrm{d}\mathbf{y}$ denotes the *volume* differential in phase space.

constitutes a reduced system of (1). We are interested in the mean prediction error

$$s_t := \langle ||\bar{\mathbf{x}}_t - \mathbf{x}_t|| \rangle_{\mathbf{x}_0,\mathbf{y}_0} \tag{4}$$

of this reduced model with respect to the true dynamics (1). Here, $\bar{\mathbf{x}}_t$ and $\mathbf{x}_t$ denote trajectories of the respective models starting from the same initial state $\bar{\mathbf{x}}_0 = \mathbf{x}_0$ in the observables. The average is taken over the full initial condition $(\mathbf{x}_0, \mathbf{y}_0)$ drawn according to the invariant measure $\mu(\mathbf{x}, \mathbf{y})$ of the full system (1).

For widely separated time scales, the *Averaging Theorem* [1, 2] gives bounds on the prediction error s_t. Under certain assumptions on $\mathbf{f}$ and $\mathbf{g}$ (essentially Lipschitz continuity and ergodicity, i.e. the equivalence of $\bar{\mathbf{f}}$ with a time average), the true trajectory $\mathbf{x}_t$ converges, for any *finite* time t, to the trajectory $\bar{\mathbf{x}}_t$ of the averaged model in the limit $\varepsilon \to 0$ (for almost all initial conditions $\mathbf{y}_0$). This means that, in this limit, the fast subsystem enters the evolution of the observables only through the conditional measure $\mu(\mathbf{y}|\mathbf{x})$, and the initial condition $\mathbf{y}_0$ becomes irrelevant. For ε small but larger than zero, $\bar{\mathbf{x}}_t$ can be viewed as a zeroth-order approximation to the true trajectory.

While the statements of the Averaging Theorem apply to the case $\varepsilon \ll 1$, such a clear separation of time scales may not not be very realistic in practice. Thus, we numerically study the predictive skill of an example system involving two only moderately separated time scales.

2.2. The Numerical Model

As our example system, we use a model originally introduced by Lorenz to study the effect of different temporal and spatial scales on the predictability of atmospheric flow [6]. Each of the slow variables x_k belongs to one of the K large sectors of a periodic domain (similar to a latitude circle). Every sector is, in turn, divided into J subsectors, to which the fast variables $y_{j,k}$ are assigned. The equations of motion for both sets of variables include a quadratic anisotropic nonlinearity ('advection'), linear 'dissipation', constant forcing, and linear coupling to the other set of variables within the same spatial sector:

$$\begin{aligned} \dot{x}_k &= -x_{k-1}(x_{k-2} - x_{k+1}) - x_k + F_x + h_x \frac{1}{J}\sum_{j=1}^{J} y_{j,k} \\ \dot{y}_{j,k} &= \frac{1}{\varepsilon}\Big(-y_{j+1,k}(y_{j+2,k} - y_{j-1,k}) - y_{j,k} + F_y + h_y x_k\Big) \end{aligned} \tag{5}$$

The indices are understood as being periodic for both x and y, i. e., $x_{k\pm K} = x_k$, and $y_{j,k\pm K} = y_{j,k}$, $y_{j\pm J,k} = y_{j,k\pm 1}$. Adopting the parameter values investigated by Lorenz in [6], we chose a moderate time scale ratio of $\varepsilon = 0.1$. The forcing $F_x = 10$ ensures chaotic behaviour of the observables. They are coupled to the fast variables with $h_x = -0.8$. The y's are not forced directly ($F_y = 0$) but via strong coupling ($h_y = 1$) to the x's. The system size was chosen to be $K = 9, J = 8$.

To get an impression of the model behaviour, Fig. 1 (left panel) shows a typical profile of both sets of variables at a given moment in time, obtained from a numerical integration. The system domain is seen to correspond approximately

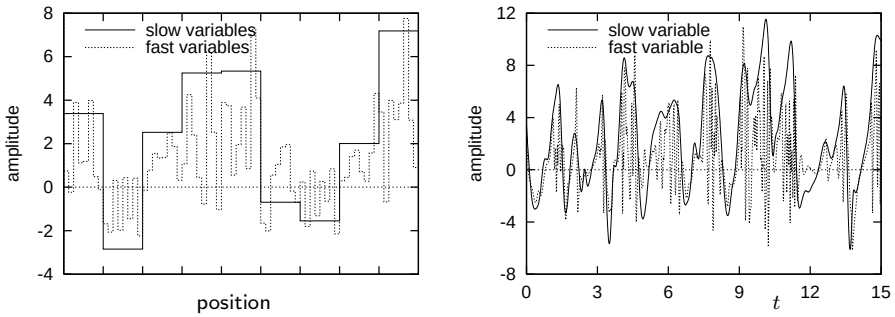

FIGURE 1. Typical profiles x_k, $y_{j,k}$ (left), and typical pieces of trajectory $x_{1;t}$, $y_{4,1;t}$ (right) of the example system (5).

to two 'wavelength' of a roughly periodic spatial structure in the observables. As time proceeds, this structure propagates, under distortion, slowly to the 'left'. The profile of the fast degrees of freedom also possesses propagating hills and valleys, whose mean height follows quickly the amplitude of the observables. This situation is confirmed by the small piece of trajectory of a selected slow and a fast variable in the right panel of Fig. 1. The trajectories illustrate the different typical time scales.

Applying the averaging procedure sketched above to the model (5) gives the following evolution equation of a reduced model:

$$\dot{\bar{x}}_k = -\bar{x}_{k-1}(\bar{x}_{k-2} - \bar{x}_{k+1}) - \bar{x}_k + F_x + h_x \mathrm{Y}_k(\bar{\mathbf{x}}) \tag{6}$$

with $\mathrm{Y}_k(\mathbf{x}) = (1/J)\sum_{j=1}^{J} \langle y_{j,k} \rangle_{\mathbf{y}|\mathbf{x}}$. The term which couples the fast degrees of freedom into the equation of motion of the observables is thus replaced by a function of the slow variables themselves. This has the form of a subgrid-scale parametrization. Details of our numerical implementation are given in the Appendix.

2.3. Prediction Experiments

The prediction error s_t was calculated according to Eq. (4), using the maximum norm $||\bar{\mathbf{x}} - \mathbf{x}||_{\max} = \max_k |\bar{x}_k - x_k|$. The result for the reduced system is plotted in Fig. 2 (solid line). The comparison with the persistence 'forecast' ($\bar{\mathbf{x}}_t = \mathbf{x}_0 \ \forall t$, dashed line) shows that the prediction range is considerably longer than the temporal range of the autocorrelations. For further comparison, the dotted line reflects the intrinsic instability of the system: It gives the error of the *full* model if the initial conditions are given with a precision of only 16 digits which is near the numerical resolution. As shown by the dash-dotted line, the prediction error of the reduced model roughly corresponds to 2 digits precise initial conditions. We conclude that reduced models constructed by averaging may possess considerable predictive skill even for only moderately separated time scales.

There remains the question whether the calculated error s_t truely reflects the properties of the averaging procedure, not being enlarged by insufficiencies of

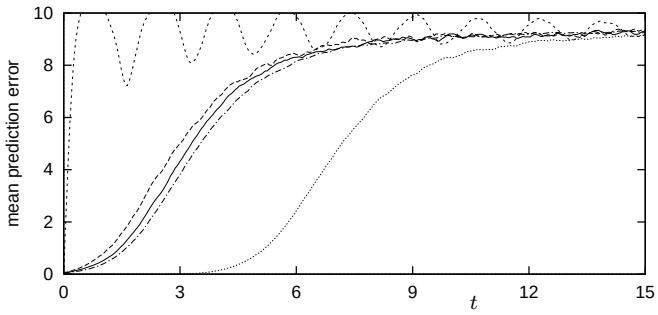

FIGURE 2. Average prediction error vs. prediction time for the reduced system (——), persistence (- - - -), the full system with disturbed initial conditions (····· : only 16 digits numerical precision, -·-·- : only 2 digits), and a reduced system with only linear parametrization (– – –). [In any case, only the error in the observables $\mathbf{x}$ was measured, cf. Eq. (4)]

the numerical implementation. Good confidence is provided in the next section. For the moment, the long-dashed line indicates the predictive skill of a reduced model with a poorer parametrization function: rather than $\mathrm{Y}_k(\mathbf{x})$ obtained in the way described in the Appendix, we used only a linear function of the form $\mathrm{Y}_k^{\mathrm{lin}}(\mathbf{x}) = ax_k + b$ with some fitted constants a and b. Obviously, using the more sophisticated parametrization is essential, but nevertheless the performance seems not to be too sensitive about inaccuracies of $\mathrm{Y}_k(\mathbf{x})$ (this was confirmed by further comparison studies).

As expected, for decreasing ε the predictive skill further improves. Details will be published elsewhere [7]. As a further 'quality check' for the averaging procedure besides the prediction experiments, it can be shown that a number of long-time properties of the original system (5), such as Lyapunov spectrum and autocorrelation function, are quite faithfully reproduced by the reduced model (6). When comparing the Lyapunov spectrum of (5) with its conditional Lyapunov exponents, we found a remarkable coincidence, making it possible to almost unambiguously assign the individual Lyapunov exponents to either of the to subsystems. Such a parting might be a system property which favours its predictability by an averaged model [7].

3. Stochastic modelling

Rather than removing the fast variables completely, one may assume the initial condition $\mathbf{y}_0$ to be drawn *randomly* according to the conditional invariant measure $\mu(\mathbf{y}|\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}$: This conditional probability measure represents all the possible information on the (unresolved) values of $\mathbf{y}_0$ if only the $\mathbf{x}_0$ are known. We will

consider three questions important for the application of this concept: First, what information may be expected from such a stochastic model, and how is it related to the information gained by an averaged model? Second, in which way does this concept lead to a reduced equation of motion for the observables $\mathbf{x}$ (which obviously will contain some stochastic terms)? Third, under which conditions could these stochastic terms be approximately modelled by white noise (which would be desirable to render the approach managable in a practical implementation)?

3.1. Predictive skill and variability

A trajectory $(\tilde{\mathbf{x}}_t, \tilde{\mathbf{y}}_t)$ starting from given $\tilde{\mathbf{x}}_0 = \mathbf{x}_0$ but random $\tilde{\mathbf{y}}_0$ obviously represents a stochastic process. We are particularly interested in the average

$$\mathbf{a}_t := \langle \tilde{\mathbf{x}}_t \rangle_{\tilde{\mathbf{y}}_0|\mathbf{x}_0} \tag{7}$$

and the dispersion

$$\sigma_t := \langle ||\tilde{\mathbf{x}}_t - \mathbf{a}_t|| \rangle_{\tilde{\mathbf{y}}_0|\mathbf{x}_0} \tag{8}$$

of the observables. When using $\mathbf{a}_t$ for prediction, the resulting mean prediction error $\langle ||\mathbf{a}_t - \mathbf{x}_t|| \rangle_{\mathbf{x}_0,\mathbf{y}_0}$ is obviously given by $\langle \sigma_t \rangle_{\mathbf{x}_0,\mathbf{y}_0}$ because the initial condition $\mathbf{y}_0$ of the true trajectory has, by construction, the same statistical properties as that of the stochastic trajectory.

In the limit $\varepsilon \to 0$, the Averaging Theory outlined in Section 2.1 implies that $\mathbf{a}_t$ converges to $\bar{\mathbf{x}}_t$ (and hence to $\mathbf{x}_t$), and σ_t converges to zero for finite time t. Note that the *scaled-up* deviation

$$\boldsymbol{\xi}_t := \frac{1}{\sqrt{\varepsilon}} (\tilde{\mathbf{x}}_t - \bar{\mathbf{x}}_t) \tag{9}$$

may show non-trivial statistical properties even for $\varepsilon \to 0$, which will be considered in more detail in Section 4. Note further that on *infinite* times t a different behaviour is possible, which is the subject of large-deviation theory [2].

To get an impression about the relation between $\mathbf{a}_t$, $\bar{\mathbf{x}}_t$, and $\mathbf{x}_t$ for ε small but finite, we again used the numerical example system of the previous section. For a given initial condition $\mathbf{x}_0$ of the observables, we prepared an ensemble $(\tilde{\mathbf{x}}_0^{(m)}, \tilde{\mathbf{y}}_0^{(m)})$ of numerical realizations of the random initial condition with $\tilde{\mathbf{x}}_0^{(m)} = \mathbf{x}_0$ as described in the Appendix. In a prediction experiment, we then evolved the true trajectory $(\mathbf{x}_t, \mathbf{y}_t)$, the ensemble of trajectories $(\tilde{\mathbf{x}}_t^{(m)}, \tilde{\mathbf{y}}_t^{(m)})$, and, as before, a trajectory of the reduced system $\bar{\mathbf{x}}_t$, forward in time. From this, we calculated the prediction error of the stochastic model, $||\mathbf{a}_t - \mathbf{x}_t||$, the prediction error of the reduced system, $||\bar{\mathbf{x}}_t - \mathbf{x}_t||$ as before, the dispersion σ_t of the stochastic trajectory, and the distance between the trajectories of both reduced systems, $||\mathbf{a}_t - \bar{\mathbf{x}}_t||$. (In every case, we again used the maximum norm $||\cdot|| = ||\cdot||_{\max}$ in the observables.) The result for a particular initial state $(\mathbf{x}_0, \mathbf{y}_0)$ is drawn in the left panel of Fig. 3. For small times t where the prediction errors are still small, the distance between the trajectories of the averaged and the stochastic models is seen to be much smaller than the distances to the true trajectory, and also much smaller than the ensemble dispersion. This shows that the trajectories forecasted by both reduced

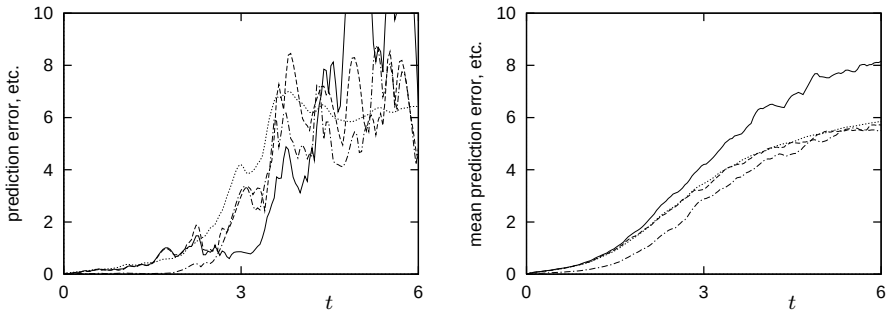

FIGURE 3. Prediction error $||\mathbf{a}_t - \mathbf{x}_t||$ of the stochastic model (- - - -), prediction error $||\bar{\mathbf{x}}_t - \mathbf{x}_t||$ of the averaged model (——), ensemble dispersion σ_t of the stochastic model ($\cdots\cdots$), and distance $||\mathbf{a}_t - \bar{\mathbf{x}}_t||$ between the trajectories of stochastic and averaged models (- · - · -). While the curves in the left panel refer to a particular initial condition $(\mathbf{x}_0, \mathbf{y}_0)$, the right panel gives the mean behaviour $(\langle\cdot\rangle_{\mathbf{x}_0,\mathbf{y}_0})$.

models are approximately identical. The extent to which this is true depends on the initial condition $(\mathbf{x}_0, \mathbf{y}_0)$, but the mean behaviour plotted in the right panel of Fig. 3 confirms this as a general result.

In a further run, we calculated the mean predictive skill of just a single stochastic trajectory. The resulting mean prediction error $\langle||\bar{\mathbf{x}}_t - \mathbf{x}_t||\rangle_{\mathbf{x}_0,\mathbf{y}_0}$ turns out to coincide with s_t of the averaged model (not shown). (We take this also as an indication that the used implementations of the reduced models are sufficiently accurate.)

As far as prediction errors are concerned, one obviously does not profit from an inclusion of stochasticity for the example studied here (for large t, where correlation to the initial condition is lost, the error of the averaged model clearly exceeds that of the stochastic model, but this is only due to the trivial fact that $\mathbf{a}_t$ is sure to approach the long-term mean $\int \mathbf{x} \mathrm{d}\mu(\mathbf{x})$ while $\bar{\mathbf{x}}_t$ keeps visiting the whole attractor). However, the stochastic model provides more information: From the ensemble dispersion we are able to estimate the prediction error without knowing the true trajectory. Also, for other quantities of interest (e.g. probability densities) stochastic modelling is often essential.

Let us note that there are plenty of other applications where noise turns out to be useful. In Ref. [8], e.g., noise has been employed in a very low dimensional model of turbulence to adjust the frequency of passing a homoclinic loop: The stochasticity helps the system to overcome the 'sticky' vicinity of the saddle point.

3.2. The emergence of stochastic terms

So far, the stochastic model was a strictly deterministic dynamics, starting from randomly chosen initial conditions. It would be desirable to replace the deterministic evolution of the fast variables by a stochastic evolution, resulting in a reduced model which is deterministic in $\mathbf{x}$ with random perturbations at every moment, representing $\mathbf{y}$.

The emergence of 'noise sources' in reduced systems has been generally considered in the literature [9, 10]. We shall be illustrating this approach using, as an introductive example, the general linear system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}, \tag{10}$$

$$\dot{\mathbf{y}} = \frac{1}{\varepsilon}(\mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{y}), \tag{11}$$

($\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ constant matrices). Let us first derive the conditional invariant measure $\mu(\mathbf{y}|\mathbf{x})$. It corresponds to a multivariate Gaussian density because this is the only probability distribution with finite variance which does not change its shape under linear transformations. From its invariance under the evolution (11), one obtains that its mean is given by

$$\langle \mathbf{y} \rangle_{\mathbf{y}|\mathbf{x}} = -\mathbf{D}^{-1}\mathbf{C}\mathbf{x} \tag{12}$$

and that its covariance matrix $\mathbf{Q} = \langle (\mathbf{y} + \mathbf{D}^{-1}\mathbf{C}\mathbf{x})(\mathbf{y} + \mathbf{D}^{-1}\mathbf{C}\mathbf{x})^{\mathrm{T}} \rangle_{\mathbf{y}|\mathbf{x}} = \mathbf{Q}^{\mathrm{T}}$ has to be a solution of

$$\mathbf{Q}\mathbf{D}^{\mathrm{T}} = -\mathbf{D}\mathbf{Q}. \tag{13}$$

We assume that a non-singular solution exists, which should be the case if the matrix $\mathbf{D}$ describes a non-dissipative dynamics[2].

Having determined the conditional invariant measure, application of the *averaging* procedure as considered in the last section would yield the reduced system

$$\dot{\bar{\mathbf{x}}}_t = \bar{\mathbf{A}}\bar{\mathbf{x}}_t \qquad \text{with} \qquad \bar{\mathbf{A}} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}. \tag{14}$$

To obtain now a *stochastic* model, we insert the solution $\mathbf{y}_t$ of Eq. (11) into Eq. (10),

$$\dot{\mathbf{x}}_t = \mathbf{A}\mathbf{x}_t + \mathbf{B}e^{\mathbf{D}t/\varepsilon}\mathbf{y}_0 + \int_0^t \mathbf{B}e^{\mathbf{D}(t-t')/\varepsilon}\mathbf{C}\mathbf{x}_{t'}\mathrm{d}t'/\varepsilon. \tag{15}$$

[2] One may, e.g., think of a harmonic Hamiltonian dynamics, defined by the Hamiltonian function $H(\mathbf{x},\mathbf{y}) = \frac{1}{2}(\mathbf{y} + \mathbf{D}^{-1}\mathbf{C}\mathbf{x})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{y} + \mathbf{D}^{-1}\mathbf{C}\mathbf{x})$. From the general equation of motion $\dot{\mathbf{y}} = \mathbf{S}\nabla_{\mathbf{y}}H(\mathbf{x},\mathbf{y})$ (the matrix $\mathbf{S} = -\mathbf{S}^{\mathrm{T}}$ being the symplectic form) one then has $\dot{\mathbf{y}} = \mathbf{S}\mathbf{Q}^{-1}\mathbf{D}^{-1}\mathbf{C}\mathbf{x} + \mathbf{S}\mathbf{Q}^{-1}\mathbf{y}$. This coincides with Eq. (11) iff $\mathbf{D} = \mathbf{S}\mathbf{Q}^{-1}$; the inversion $\mathbf{Q} = \mathbf{D}^{-1}\mathbf{S}$ clearly satisfies Eq. (13) identically (remember $\mathbf{Q} = \mathbf{Q}^{\mathrm{T}}$). As a check for consistency, note that from $\varrho(\mathbf{y}|\mathbf{x}) = \exp\{-H(\mathbf{x},\mathbf{y})\}$ one obtains exactly the Gaussian probability distribution derived above.

Finally, note that $\mathbf{D}$ determines $\mathbf{Q}$ only up to a set of arbitrary constants, which here correspond to the 'excitation' of the individual eigenfrequences of the Hamiltonian system.

Integrating by parts leads to the form [10, 11]

$$\dot{\mathbf{x}}_t = \bar{\mathbf{A}}\mathbf{x}_t + \mathbf{r}_t - \int_0^t \mathbf{K}_\tau \dot{\mathbf{x}}_{t-\tau} \mathrm{d}\tau \tag{16}$$

with $\bar{\mathbf{A}}$ again from Eq. (14) and

$$\mathbf{r}_t := \mathbf{B}e^{\mathbf{D}t/\varepsilon}(\mathbf{y}_0 + \mathbf{D}^{-1}\mathbf{C}\mathbf{x}_0), \tag{17}$$

$$\mathbf{K}_\tau := -\mathbf{B}e^{\mathbf{D}\tau/\varepsilon}\mathbf{D}^{-1}\mathbf{C}. \tag{18}$$

This is still an exact equation of motion. If now $\mathbf{y}_0$ is considered as a stochastic variable, the term $\mathbf{r}_t$ becomes stochastic as well. Since we draw $\mathbf{y}_0$ according to $\mu(\mathbf{y}|\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}$ determined above, one easily shows that the mean and the autocorrelation matrix of this noise term are

$$\langle \mathbf{r}_t \rangle_{\mathbf{y}_0|\mathbf{x}_0} = \mathbf{0} \qquad \text{and} \qquad \langle \mathbf{r}_t \mathbf{r}_{t+\tau}^{\mathrm{T}} \rangle_{\mathbf{y}_0|\mathbf{x}_0} = -\mathbf{B}e^{\mathbf{D}\tau/\varepsilon}\mathbf{Q}\mathbf{B}^{\mathrm{T}}, \tag{19}$$

which do not dependent on time (in the derivation, use of Eq. (13) is essential). Thus, $\mathbf{r}_t$ represents a stationary zero-mean Gaussian stochastic process. (Further, comparing Eqs. (18) and (19) one sees that, provided additional relations involving the coupling matrices are fulfilled, a generalized 'fluctuation-dissipation relation' $\langle \mathbf{r}_t \mathbf{r}_{t+\tau}^{\mathrm{T}} \rangle \propto \mathbf{K}_\tau$ holds between the stochastic term $\mathbf{r}_t$ and the deterministic memory kernel $\mathbf{K}_\tau$ [10, 11].)

In Ref. [11], such a derivation of stochastic terms is carried out also for a nonlinear example system from the atmospheric context (barotropic flow on a β-plane). The equations of motion are solved approximately by a perturbative expansion. In the reduced (stochastic) model, one then again has modified terms for the interaction among the slow variables (analogously to the term $\bar{\mathbf{A}}\mathbf{x}$ in the linear example), as well as a series of pairs of terms, each consisting of a stochastic and a deterministic term which are connected by formal 'fluctuation-dissipation relations'.

3.3. Decay of correlations

For any practical implementation of a stochastic model, the statistical properties of the random terms should be as simple as possible. In particular, one generally expects a noise source to have a finite range of correlations, if not even to be delta-correlated ('white noise').

Considering again the linear example system (10, 11), it is clear from Eq. (19) that the 'noise' $\mathbf{r}_t$ actually is a (quasi)periodic function, once the initial value $\mathbf{r}_0$ has been fixed. While for τ being raised from zero the autocorrelation $\langle \mathbf{r}_t \mathbf{r}_{t+\tau}^{\mathrm{T}} \rangle$ may drop down, there will be values of the time difference $\tau > 0$ where $\mathbf{r}_{t+\tau}$ is arbitrarily close to $\mathbf{r}_t$ again, i.e., where $\langle \mathbf{r}_t \mathbf{r}_{t+\tau}^{\mathrm{T}} \rangle \approx \langle \mathbf{r}_t \mathbf{r}_t^{\mathrm{T}} \rangle$. However, if the number of fast variables is large, these recurrence times may be expected to be very large. In fact, it has been shown rigorously that, in the limit of an infinite number of components of $\mathbf{y}$, the process $\mathbf{r}_t$ approaches an Ornstein-Uhlenbeck process, characterized by an

exponentially decaying autocorrelation $\langle \mathbf{r}_t \mathbf{r}_{t+\tau}^{\mathrm{T}} \rangle \propto e^{-\gamma\tau/\varepsilon}$ (with a *scalar* $\gamma > 0$)[3]. It is now easy to see that in the limit $\varepsilon \to 0$ the correlation time vanishes: The noise becomes white.

Thus, we identified two mechanisms removing (undesired) correlations from the stochastic terms: First, a *large number of fast variables* moves recurrences to the far future, while, second, a *strong separation of time scales* leads to fast initial decay of autocorrelations. These two mechanisms can be expected to work equally in general non-linear systems. There, the situation even becomes better: As a third mechanism, exponentially decaying correlations can arise from *chaotic dynamics* due to their mixing properties. This situation is considered in more detail in the next section.

4. Central Limit Theorems and their limits

In the following, we review some recent rigorous results on how Gaussian white noise and Langevin dynamics can arise from deterministic chaotic dynamics, involving generalizations of the famous Central Limit Theorem. Like the Averaging Theorem, these results apply exactly in the limit $\varepsilon \to 0$. To show their relevance in practice, we again illustrate the approach to the limit by a numerical example.

4.1. Central Limit Theorems

The central limit theorem in its simplest version is well known. It just says that the probability distribution $p(x)$ of a rescaled sum x_N of N independent identically distributed random variables y_n with zero mean,

$$x_N = \sqrt{\frac{1}{N}} \sum_{n=1}^{N} y_n, \tag{20}$$

converges to a Gaussian distribution in the limit $N \to \infty$. The rescaling of the sum by the factor $\sqrt{1/N}$ guarantees that x_N has the same variance as the variables y_n, independently of N.

What is much less known is the fact that there are generalizations of the central limit theorem for weakly dependent events. This means that the 'random variables' y_n need not to be independent, they can be weakly dependent. For example, they can be the deterministic iterates of a sufficiently mixing map T. If we have in a deterministic way

$$y_{n+1} = T(y_n) \tag{21}$$

[3]Interest in this limit has arisen in the literature already a long time ago because systems like (10,11) can be used to model a heat bath: It can be shown that a particle coupled to a large number of classical harmonic oscillators behaves, in the thermodynamic limit, like a Brownian particle driven stochastically, cf., e.g., [9, 10]. The limit is also useful here because, in an atmospheric model, the fast processes should be expected to involve a large number of active degrees of freedom.

then, regarding the initial value y_0 as a random variable distributed according to the natural invariant measure of the map, the probability distribution of x_N still becomes Gaussian in the limit $N \to \infty$. The first person who proved a theorem in this direction was probably Marc Kac in 1946 [12]. In the mean time many generalization have been worked out and many different versions of central limit theorems for deterministic and random dynamical systems T have been proved (e.g. [13]—[24]).

Using the limit theorems, (in principle) it is possible to prove that a deterministic chaotic process ('the weather'), evolving on a fast time scale, in some appropriate sense converges to Gaussian white noise under rescaling of time. Such a rescaling arises since we want to use the fast process as a driving force of a slow one ('the climate'). We then look at this fast process on the slow time scale.

Let us denote the fast time scale by τ (in the sense of Eq. (1), $\tau = \varepsilon t$ where t is the time of a typical variation in the observables). From the sequence of deterministic chaotic iterates y_n of Eq. (21), we construct the process

$$L_\tau(t) = \tau^{1/2} \sum_{n=1}^{\infty} y_n \delta(t - n\tau) \tag{22}$$

consisting of a sequence of weighted Dirac spikes at time intervals τ. Regarding the initial value y_0 as a random variable, $L_\tau(t)$ represents a stochastic process: The map (21) generates just one realization for each initial value. For $\tau > 0$, it is, of course, by no means a simple Gaussian process, but a complicated chaotic process depending on the map T. One possibility to characterize this process completely is to calculate the set of all higher-order correlation functions, or, equivalently, the characteristic functional. The entire hierarchy of all n-point correlations fully characterizes this chaotic process. Typically the n-point functions have a very complicated structure [24]. However, if T is a so-called φ-mixing map[4], one can show that $L_\tau(t)$ reduces to Gaussian white noise in the limit $\tau \to 0$, in the sense that the integrated process

$$W_\tau(t) = \int_0^t L_\tau(t')dt' = \tau^{1/2} \sum_{n=1}^{\lfloor t/\tau \rfloor} y_n, \tag{23}$$

converges to the Wiener process $W(t)$ (=Brownian motion). For convenience, we had rescaled the chaotic process $L_\tau(t)$ by the factor $\sqrt{\tau}$, which is now seen to correspond to the factor $\sqrt{1/N}$ in the central limit theorem (20).

The assumption of discrete time in eq. (21) is not very restrictive. We can always regard the y_n as stroboscopic variables of a continuous dynamics. Also note that, in contrast to the linear example in the last section, a high-dimensional phase space is not necessary for the white-noise limit, already a 1-dimensional

[4] *Mixing* is an abundent property of chaotically deterministic systems, essentially signifying the statistical independence of sufficiently distant parts of the trajectory. Depending on the system, this property can be fulfilled in a differently strict sense; as an example, *φ-mixing* always implies *strong mixing*.

phase space can do. An example of such a φ-mixing map is $T(y) = 1 - 2y^2$ on $[-1, 1]$. Further, the probability distribution of initial values y_0 is irrelevant due to the mixing properties (as long as it is absolute continuous with respect to Lebesgue measure, the iterates y_n for large n approach their invariant distribution in any case).

We can generalize to vector-valued processes $\mathbf{L}_\tau(t)$ generated by higher-dimensional maps $\mathbf{T}$ and couple this process to an ODE (or PDE) representing the slowly varying part of the dynamics,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{L}_\tau(t). \tag{24}$$

This, together with Eqs. (21) and (22), can be regarded as the deterministic chaotic analogue of a nonlinear Langevin equation. Discretizing time into steps of length τ (the fast time scale), we may equivalently formulate the dynamics as the recurrence relation

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{F}(\tau, \mathbf{x}_n) + \tau^{1/2}\mathbf{y}_{n+1}, \\ \mathbf{y}_{n+1} &= \mathbf{T}(\mathbf{y}_n), \end{aligned} \tag{25}$$

where the map $\mathbf{F}(\tau, \mathbf{x}_n)$ represents the slow dynamics $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ integrated over a singe short time step τ, i.e., $(\mathrm{d}/\mathrm{d}\tau)\mathbf{F}(\tau, \mathbf{x}) = \mathbf{f}(\mathbf{x})$, $\mathbf{F}(0, \mathbf{x}) = \mathbf{x}$. In the form (25), the considered dynamics is recognized as an example of our general system (1). (The time scale separation expresses itself by the fact that the map $\mathbf{F}$ is the near-identity.) Mathematicians call dynamical systems of the form (25) *skew products*. For a review, see [25]. For $\tau \to 0$, $t = n\tau$ finite, the limit theorems mentioned before guarantee the equivalence of (25) and the stochastic differential equation

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{L}(t), \tag{26}$$

$\mathbf{L}(t)$ now being higher-dimensional Gaussian white noise. This can be rigorously proved if each component of $\mathbf{y}$ evolves in a φ-mixing way [25].

This concept looks promising, pointing out a direction how to tackle Hasselmann's program. In practice, however, several difficulties occur, some of which we want to mention in the following.

4.1.1. Higher dimensions. Most limit theorems that are known in the mathematical literature deal with 1-dimensional maps T or trivial (uncoupled) higher-dimensional extensions of those. Only for very simple examples of chaotic maps the φ-mixing property can be proved. However, since the weather apparently is determined by a complicated high-dimensional dynamical system, we would need central limit theorems for non-trivial dynamical systems $\mathbf{T}$ in higher dimensions. Not too much is known here. Generally, these types of problems with higher dimensions also occur for the rigorous justification of non-equilibrium thermodynamics and statistical mechanics. Also here one wants deterministic chaotic forces to be described by Gaussian white noise after rescaling. Although some progress has been made by projection techniques, we are still far away from a rigorous theory [26, 5, 27].

4.1.2. Nonhyperbolic systems, local Lyapunov exponents and singular test functions. A more general chaotic process can be defined by looking at functions $\mathbf{h}(\mathbf{y}_n)$ of the iterates $\mathbf{y}_n$, i.e. the process

$$\mathbf{W}_\tau(t) = \tau^{1/2} \sum_{n=1}^{\lfloor t/\tau \rfloor} \mathbf{h}(\mathbf{y}_n) \tag{27}$$

$$\mathbf{y}_{n+1} = \mathbf{T}(\mathbf{y}_n). \tag{28}$$

Here again $\mathbf{T} : Y \to Y$ is a chaotic map on some (possibly high-dimensional) phase space Y, and $\mathbf{h}$ is a function projecting from Y onto the phase space of the ODE. In principle all types of test functions $\mathbf{h}$ can be considered, and not too much is known on which types of functions guarantee the existence of a central limit theorem, except for the 1-dimensional case. Even in the 1-dimensional case, some surprises occur. For example, one may be interested in local Lyapunov exponents, corresponding to test functions $h(y) = \log|T'(y)|$. Prasad and Ramaswamy [28] recently looked at probability distributions $p(\lambda)$ of local Lyapunov exponents λ_N defined by

$$\lambda_N = \frac{1}{N} \sum_{j=N(k-1)+1}^{kN} \log|T'(y_j)| \tag{29}$$

for finite N. These quantities depend on the initial conditions, they only become independent of the initial conditions for $N \to \infty$. For the logistic map $T(y) = ay(1-y)$, the distributions turn out to often show strong deviations from Gaussian behaviour and thus departures from the central limit theorem are evident. This is in particular observed for parameter values a with intermittent behaviour, but (surprisingly) also for the fully developed chaotic case $a = 4$, whereas 'typical' chaotic parameter values $a < 4$ usually lead to almost Gaussian behaviour. A theoretical understanding of these numerical results is still lacking. Since the iterates y_n of the logistic map for $a = 4$ are φ-mixing and hence do satisfy a central limit theorem, these numerical findings must have to do with the singularity of the test function $h(y) = \log|T'(y)|$ at $y = \frac{1}{2}$. Needless to say that all distributions $p(\lambda)$, for the above rescaling, converge to $\delta(\lambda - \bar{\lambda})$ for $N \to \infty$, where $\bar{\lambda}$ denotes the Lyapunov exponent of the map.

4.1.3. Back-coupling of the slow variables. So far we assumed that the average of the chaotic iterates $\mathbf{y}_n$ was zero. If it is not zero, we may just substract it and re-define our system such that the chaotic variables have zero average again. However, as in our general formulation Eq. (1), the average (in fact all the statistical properties) of the fast chaotic variables will in turn depend on the slow variables of the system. This back-coupling destroys the simple skew product structure of eq. (25), and rigorous analytical results are then much more difficult to prove.

4.2. Distribution of the observables and finite time scale corrections

The central limit theorems also provide information on the statistical properties of the observables $\tilde{\mathbf{x}}_t$ of the stochastic model. Under certain conditions, it can be shown that the rescaled fluctuations $\boldsymbol{\xi}$ of the stochastic trajectory around the zeroth-order approximation by the averaged model (cf. Eq. (9)) converge to a Gaussian process in the limit $\varepsilon \to 0$ [2].

It is clear that the weather evolves chaotically on a small but *finite* time scale compared to the climate. Now what kind of corrections should one expect due to a small but non-zero time scale? Apparently there must be leading-order corrections in ε or $\sqrt{\varepsilon}$ that describe the approach to a Gaussian distribution if a central limit theorem is valid for $\varepsilon \to 0$.

An answer has recently been given for a simple toy model system, namely a 1-dimensional linear Langevin-like equation, driven not by Gaussian white noise but by determinsitic chaotic forces $L_\tau(t)$ where the iterates evolve according to Tchebyscheff maps T_N of N-th order. We consider the deterministic chaotic dynamics

$$\dot{x} = -x + L_\tau(t) \tag{30}$$

with $L_\tau(t)$ given by eq. (22) and $y_{n+1} = T_N(y_n)$. The first few Tchebyscheff maps are

$$\begin{aligned} T_2(y) &= 2y^2 - 1, \\ T_3(y) &= 4y^3 - 3y, \\ T_4(y) &= 8y^4 - 8y^2 + 1, \\ T_5(y) &= 16y^5 - 20y^3 + 5y \\ \ldots &\quad \ldots \end{aligned} \tag{31}$$

An analytic expression for the invariant probability density $p(x)$ of x can be derived by making a perturbative expansion in the time constant $\tau \propto \varepsilon$. Two different techniques have been described in the literature [24, 29]. After a lengthy calculation one obtains the final result that for $N = 2$

$$\begin{aligned} p^{(2)}(x) &= \sqrt{\frac{2}{\pi}}\Bigg(1 + \tau^{\frac{1}{2}}(-2x + \frac{8}{3}x^3) \\ &\quad + \tau(\frac{32}{9}x^6 - \frac{31}{3}x^4 + \frac{15}{2}x^2 - \frac{37}{48})\Bigg)e^{-2x^2}, \end{aligned} \tag{32}$$

for $N = 3$

$$p^{(3)}(x) = \sqrt{\frac{2}{\pi}}\left(1 + \tau(\frac{1}{3}x^4 + \frac{3}{2}x^2 - \frac{7}{16})\right)e^{-2x^2}, \tag{33}$$

and for all maps with $N \geq 4$

$$p^{(N\geq 4)}(x) = \sqrt{\frac{2}{\pi}}\left(1 + \tau(-x^4 + \frac{7}{2}x^2 - \frac{11}{16})\right)e^{-2x^2}, \tag{34}$$

neglecting terms of $O(\tau^{3/2})$ [30]. These densities apply to the special case of scaled-up fluctuations ξ with $\bar{x} = 0$. The scaling-up has already been done implicitly by introducing the factor $\tau^{1/2}$ instead of τ in the definition of L_τ, Eq. (22). In the limit $\tau \to 0$ all invariant densities approach the Gaussian $p_0(x) = (2/\pi)^{1/2} e^{-2x^2}$ for all $N = 2, 3, \ldots$. Note that the first correction to the Gaussian is of order $\sqrt{\tau}$ for $T_2(y)$ but of order τ for the higher Tchebyscheff maps.

Although these analytical results have just been proved for chaotic forces generated by Tchebyscheff maps, one numerically observes universal behaviour, i.e. several other chaotic maps lead to corrections of the Gaussian limit density that are quite similar to those described by eq. (32)—(34) (for details, see [30]). Hence, although the weather is certainly a much more complicated dynamical system than a Tchebyscheff map, for universality reasons it would be interesting to investigate if meteorological data resemble some properties of the above densities. For fully developed hydrodynamic turbulence, analogous non-Gaussian density corrections obtained in a cascade-like extension of eq. (30) have indeed been observed in wind channel experiments [31].

4.3. Application to spatial averages

The two considered subsystems, the 'climate' $\mathbf{x}$ und the 'weather' $\mathbf{y}$, differ not only in the characteristic time scales of their evolution but also in the scales of the associated spatial structures (climate models, e.g., can do with a much coarser spatial grid than weather models). When the fast variables couple into the equations of motion of the slow ones, one therefore not only encounters kind of a temporal averaging as considered so far, but also some spatial averaging. This view was already incorporated by Lorenz into his toy atmospheric model Eq. (5) in an obvious way.

The concepts introduced in this section for temporal averages apply equally to spatial averaging: A φ-mixing property can be formally defined by replacing time by space. Of course, rigorous results are just as hard to derive. Thus, we illustrate the central limit theorem again by a numerical example. We use the same model as before; this time, however, with only one 'layer' [6]:

$$\dot{y}_j = -y_{j-1}(y_{j-2} - y_{j+1}) - y_j + F \tag{35}$$

Periodic boundary conditions $y_{j\pm J} = y_j$ apply. The force is again $F = 10$, and we chose $J = 100$ variables y_j. Also this system develops a roughly periodic spatial profile y_j vs. j (cf. Fig. 1).

From the numerical integration we calculated spatial averages

$$u_N = \frac{1}{N} \sum_{j=1}^{N} y_j \tag{36}$$

for different sizes N of the 'averaging area'. Fig. 4 shows the resulting histograms (solid lines), taken from a long trajectory. For comparison, we also plotted the corresponding Gaussians with the same means and variances as dotted lines. The histogram of the single variables ($N = 1$) shows pronounced non-Gaussian shape, which, because of spatial correlations, even increases for the mean of two nearest

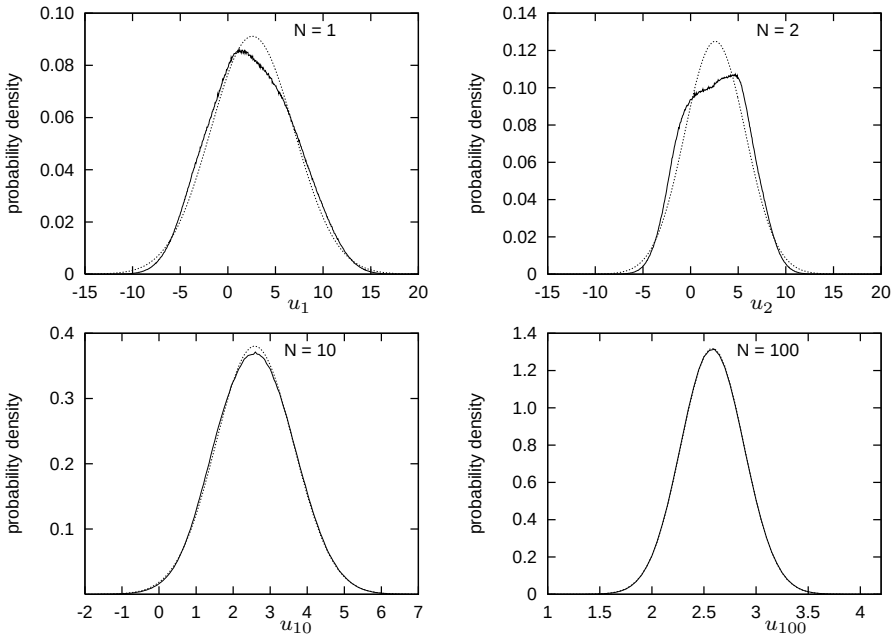

FIGURE 4. Histograms of spatial averages u_N over the variables of example system (35) for different sizes of the averaging area ($N = 1, 2, 10$, and 100)

neighbors ($N = 2$). If, however, the averaging area increases, comprising several 'wavelength' of the spatial structures, the histograms are seen to nicely approach Gaussianity ($N = 10$ and $N = 100$).

A similar behaviour could be expected for the 'deterministic noise'

$$\eta_k = \frac{1}{J} \sum_{j=1}^{J} y_{j,k} - \mathrm{Y}_k(\mathbf{x}) \tag{37}$$

in the two-layer model (5) for increasing J. Nevertheless, the long-term histogram in Fig. 5 (left panel) seems to clearly contradict this supposition. The density looks much more similar to typical non-Gaussian densities observed for velocity fluctuations in fully developed turbulence (see, e.g., [31, 32]).

A typical piece of trajectory (right panel, solid line) reveals the strong dependence of the statistics of η_k on the slow variables x_k (shown as dotted line). Sampling of *conditional* histograms turned out, for a number of (but not all) conditions $\mathbf{x}$, to be much better approximated by a Gaussian. As a cautionary remark, we note that Gaussianity might, in experimental or numerical data, be hidden behind such a mixture of different probability distributions.

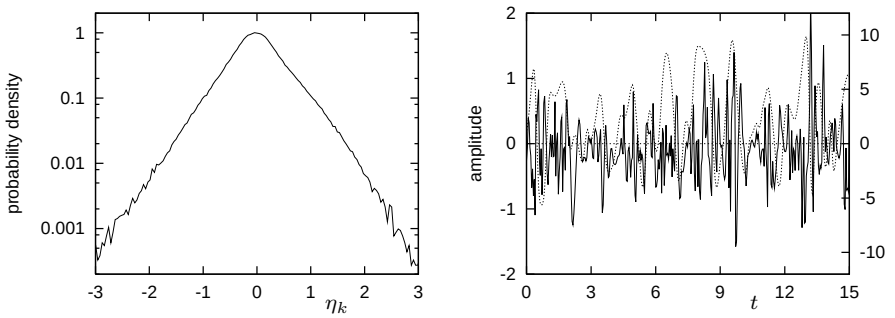

FIGURE 5. Histogram of the 'deterministic noise' η_k (left panel, note log scale) of model (5); and typical trajectory of η_1 (right panel, solid line, scale at left) along with its corresponding slow variable x_1 (dotted line, scale at right).

Conclusion

Aspects of two methods to remove unresolved fast degrees of freedom from a dynamical model were considered. If one aims to predict the actual trajectory of the (slow) observables, a reduced model can be obtained by averaging the evolution equation. For questions such as variability, a stochastic model might be more appropriate, where the randomness enters through the uncertainty of the unresolved variables.

In the limit where the ratio between the time scales of the observables and the unresolved degrees of freedom becomes infinite, mathematically rigorous statements are available. The Averaging Theorem allows to bound the prediction error of an averaged model over finite times. Generalizations of the Central Limit Theorem guarantee the equivalence of certain deterministic chaotic processes and Gaussian white noise. Using a simple example system introduced by Lorenz, there is numerical indication that the statements of these theorems are still relevant for only moderate time scale ratios.

Acknowledgement

Stimulating discussions with Dr. Frank Kwasniok are gratefully acknowledged.

Appendix: Numerical implementation details

The averaged model: Theoretically, the parametrization function Y_k depends on the entire vector $\mathbf{x}$. For the practical implementation we assumed that only the neighbouring sectors have to be considered, $\mathrm{Y}_k = \mathrm{Y}_k(x_{k-1}, x_k, x_{k+1})$. (Fig. 1 suggests that this is a good approximation, which was further justified by numerical tests.) Thus, we numerically created sample values of Y_k over a grid of 17^3 points

(x_{k-1}, x_k, x_{k+1}), and used tri-linear interpolation to calculate $\mathrm{Y}_k(x_{k-1}, x_k, x_{k+1})$ in the integration of (6). (Because of the translational symmetry of the model, the functional dependence $\mathrm{Y}_k = \mathrm{Y}_k(\mathbf{x})$ is, of course, the same for any cyclic permutation of the indices k.)

The stochastic model: Since we are not able to really sample the conditional invariant probability density $\varrho(\mathbf{y}|\mathbf{x})$, we used the following scheme to obtain random initial conditions $\tilde{\mathbf{y}}_0$: As already in the numerical sampling of the parametrization function $\mathrm{Y}(\mathbf{x})$, we assumed that $\varrho(\mathbf{y}|\mathbf{x})$ can be factorized into joint probability densities of the fast variables within one and the same spatial sector only, depending on the observable of this and both its neighbouring sectors only, i.e., $\varrho(\mathbf{y}|\mathbf{x}) = \prod_k \varrho(y_{1,k}, \ldots, y_{J,k}|x_{k-1}, x_k, x_{k+1})$. For a given initial condition $\mathbf{x}_0$ of the observables, we scanned a (very) long independent trajectory of the original model for states $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ where one of the triples $(\hat{x}_{k'-1}, \hat{x}_{k'}, \hat{x}_{k'+1})$ in the vector $\hat{\mathbf{x}}$ coincides with one of the triples (x_{k-1}, x_k, x_{k+1}) in the vector $\mathbf{x}_0$ within the accuracy $\Delta x = 0.05$. For each such 'recurrence', we took the values $\hat{y}_{j,k'}$ in sector k' as a possible configuration of $y_{j,k}$ in sector k of the (unresolved) initial condition $\mathbf{y}_0$. Subsequently, we formed a statistical ensemble of $M = 1000$ initial conditions $(\tilde{\mathbf{x}}_0^{(m)}, \tilde{\mathbf{y}}_0^{(m)})$ where $\tilde{\mathbf{x}}_0^{(m)} = \mathbf{x}_0$ for all members m, while the vectors $\tilde{\mathbf{y}}_0^{(m)}$ were assembled randomly from the previously stored configurations of the $y_{j,k}$ in the individual spatial sectors k. We assume that this ensemble reflects the conditional density $\varrho(\mathbf{y}|\mathbf{x})$ with sufficient precision (for justification, we may also invoke that the fast variables adjust, thanks to their fast time scale, rapidly to the current forcing by the observables, i.e., they quickly reach their 'conditional attractor'). The mean and the variance of the stochastic trajectory $\tilde{\mathbf{x}}_t$, respectively, were then calculated as the ensemble mean

$$\langle \tilde{\mathbf{x}}_t \rangle_{\mathbf{y}_0|\mathbf{x}_0} = \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{x}}_t^{(m)} \tag{38}$$

and the ensemble dispersion

$$\sigma_t = \frac{1}{M} \sum_{m=1}^{M} ||\tilde{\mathbf{x}}_t^{(m)} - \langle \tilde{\mathbf{x}}_t \rangle_{\mathbf{y}_0|\mathbf{x}_0}||. \tag{39}$$

References

[1] L. Arnold, these proceedings.

[2] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd ed., Springer, New York (1998), 212–282.

[3] L. Arnold, *Random Dynamical Systems*, Springer, New York (1998)

[4] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam (1981)

[5] R. F. Fox, Phys. Rep. **48C**, 179 (1978).

[6] E. N. Lorenz, *Predictability: a problem partly solved*, in: Seminar Proceedings "Predictability", European Centre for Medium-Range Weather Forecasts (1995), 1–18.
[7] C. Rödenbeck and H. Kantz, in preparation.
[8] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, coherent structures, dynamical systems, and symmetry*, Cambridge Univ. Press (1996).
[9] G. W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6** (1965), 504–515.
[10] R. Zwanzig, J. Stat. Phys. **9** (1973), 215–220.
[11] K. Lindenberg and B. J. West, J. Atmos. Sci. **41** (1984), 3021–3031.
[12] M. Kac, Ann. Math. **47**, 33 (1946)
[13] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York (1968)
[14] N.I. Chernov, Probab. Theory Relat. Fields **101**, 321 (1995)
[15] G. Keller, C.R. Acad. Sci. Paris **291**, 155 (1980)
[16] M. Denker, W. Philipp, Ergod. Theory Dyn. Syst. **4**, 541 (1984)
[17] R. Burton and M. Denker, Trans. Am. Soc. **302**, 715 (1987)
[18] I.A. Ibragimov, Theory Probab. Appl. **7**, 349 (1962)
[19] G. Keller, Z. Wahrsch. Verw. Gebiete **69**, 461 (1985)
[20] Ya.G. Sinai, Soviet Math. Dokl. **1**, 983 (1960)
[21] Y. Kifer, Trans. Am. Math. Soc. **350**, 1481 (1998)
[22] L.A. Bunimovich, Theor. Probab. Appl. **19**, 65 (1974)
[23] C. Beck, G. Roepstorff, Physica **145A**, 1 (1987)
[24] C. Beck, Nonlinearity **4**, 1131 (1991)
[25] C. Beck, Physica **233A**, 419 (1996)
[26] M.C. Mackey, *Time's Arrow: The Origns of Thermodynamic Behavior*, Springer, New York (1992)
[27] C. Beck, F. Schlögl, *Thermodynamics of Chaotic Systems*, Cambridge University Press, Cambridge (1993)
[28] A. Prasad, R. Ramaswamy, Phys. Rev. **60E**, 2761 (1999)
[29] C. Beck, J. Stat. Phys. **79**, 875 (1995)
[30] A. Hilgers, C. Beck, Phys. Rev. E **60**, 5385 (1999)
[31] C. Beck, Phys. Rev. **49E**, 3641 (1994)
[32] R. Friedrich and J. Peinke, Phys. Rev. Lett. **78**, 863 (1997); Physica D **102**, 147 (1997).

* Max Planck Institute for Physics of Complex Systems, Nöthnitzer Str. 38, D-01187 Dresden, Germany
** School of Mathematical Sciences, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, UK
E-mail address: roeden@mpipks-dresden.mpg.de; C.Beck@qmw.ac.uk; kantz@mpipks-dresden.mpg.de