Analysis of Noisy Spatio-Temporal Data

Oliver Kamps[1] and Joachim Peinke[2]

[1] Center for Nonlinear Science, Universität Münster, Germany
okamp@uni-muenster.de
[2] Institut für Physik & ForWind, Universität Oldenburg, Germany
peinke@uni-oldenburg.de
http://www.uni-oldenburg.de/twist

Abstract. In this article we discuss an extension of a method to extract Langevin equations from noisy time series to spatio-temporal data governed by stochastic partial differential equations (SPDEs). The reconstruction of the SPDEs from data is traced back to the estimation of multivariate conditional moments.

Keywords: Stochastic partial differential equations, data analysis

1 Introduction

The synergetic approach to complex systems, composed of many interacting subsystems, shows that the influence of the fast degrees of freedom on the order parameter dynamics can be described by dynamical noise. This leads to a mathematical description in terms of nonlinear Langevin equations of the form

$$\dot{y} = D^{(1)}(y) + \sqrt{D^{(2)}(y)}\Gamma \tag{1}$$

or a corresponding Fokker-Planck equation [1]. Here, $y(t)$ is the order parameter, $D^{(1)}(y)$, $D^{(2)}(y)$ are the drift and diffusion coefficients, respectively, and $\Gamma(t)$ describes a Gaussian distributed and δ-correlated stochastic force with zero mean. The two coefficients are defined according to

$$D^{(n)}(Y) = \lim_{\tau\to 0} \frac{1}{\tau}\frac{1}{n!}\langle [y(t+\tau) - y(t)]^n | y(t) = Y\rangle \tag{2}$$

which can for example be derived from the Kramers–Moyal expansion [2].

Pursuing the question of how to formulate an evolution equation for the velocity fluctuations on different scales in turbulent flows, Friedrich and Peinke [3] proposed to estimate these coefficients directly from experimental data by the approximation

$$D_*^{(n)}(Y) = \frac{1}{\tau_{\min}}\frac{1}{n!}\langle [y(t+\tau) - y(t)]^n | y(t) = Y\rangle \tag{3}$$

with $D^{(n)} \approx D_*^{(n)}$. In this context, $\tau_{\min}$ is the smallest available time scale which is still big enough to ensure that the stochastic dynamics can be described by a

A. Pelster and G. Wunner (eds.), *Selforganization in Complex Systems: The Past, Present, and Future of Synergetics*, Understanding Complex Systems,
DOI: 10.1007/978-3-319-27635-9_22

Markov process. In a large number of subsequent publications this method has been successfully applied to different scientific topics like turbulence research [3–5], finance [6], medicine [7–10] or engineering [11] to mention just a few. Besides applications to different scientific disciplines, research focused on technical questions connected to finite sampling times [12, 13] or measurement noise [14]. An recent overview is given in [15].

2 Outline of the Method

Many physical observables of complex systems do not only depend on time but also on space [1]. The dynamics of these systems is also influenced by dynamical noise [16]. In this contribution we want to show how the estimation of Langevin equations can be extended to spatio-temporal data governed by equations like

$$\partial_t y(\mathbf{z},t) = D^{(1)}[y(\mathbf{z},t)] + \sqrt{D^{(2)}[y(\mathbf{z},t)]}\,\Gamma(\mathbf{z},t) \tag{4}$$

where the drift and the diffusion coefficient depend on a spatially extended field $y(\mathbf{z},t)$. In general, the time evolution at the point $\mathbf{z}$ depends only implicitly via operators like e.g. spatial derivatives or integral expressions on the evolution at other points $\mathbf{z}'$. Due to this, we make the assumption that $D^{(1)}$ and $D^{(2)}$ are functions of the operators and $y(\mathbf{z},t)$ only and rewrite the equation as

$$\dot{y}_{\mathbf{z},t} = D^{(1)}[O_1,\ldots,O_N] + \sqrt{D^{(2)}[O_1,\ldots,O_N]}\,\Gamma_{\mathbf{z},t} \tag{5}$$

with $y_{\mathbf{z},t} := y(\mathbf{z},t)$ and the noise term $\Gamma_{\mathbf{z},t} := \Gamma(\mathbf{z},t)$. The stochastic force is defined according to

$$\langle \Gamma_{\mathbf{z},t}\rangle = 0\,, \quad \langle \Gamma_{\mathbf{z},t}\Gamma_{\mathbf{z}',t'}\rangle = \delta(t-t')C(|\mathbf{z}-\mathbf{z}'|) \tag{6}$$

where $C(|\mathbf{z}-\mathbf{z}'|)$ denotes the spatial correlation of the noise. The O_i represent the various operators. For example, in case of the reaction-diffusion equation

$$\partial_t y_{\mathbf{z},t} = \epsilon \Delta y_{\mathbf{z},t} + y_{\mathbf{z},t} - y^3_{\mathbf{z},t} + \sqrt{q(1+y^2_{\mathbf{z},t})}\,\Gamma_{\mathbf{z},t} \tag{7}$$

we would have $O_1 := y_{\mathbf{z},t}$ and $O_2 := \Delta y_{\mathbf{z},t}$. The drift and diffusion coefficients are then

$$D^{(1)}[O_1,O_2] = O_1 - O_1^3 + \epsilon O_2 \tag{8}$$

$$D^{(2)}[O_1,O_2] = q(1+O_1^2)\,. \tag{9}$$

One can show [17] that in close analogy to the normal Langevin equation (1), the drift and diffusion coefficient are defined as the multidimensional conditional averages

$$D^{(1)}[Y_1,\ldots,Y_N] = \lim_{\tau\to 0}\frac{1}{\tau}\langle y_{\mathbf{z},t+\tau} - y_{\mathbf{z},t}|O_1 = Y_1,\ldots,O_N = Y_N\rangle \tag{10}$$

$$D^{(2)}[Y_1,\ldots,Y_N] = \lim_{\tau\to 0}\frac{1}{\tau}\frac{1}{2}\langle [y_{\mathbf{z},t+\tau} - y_{\mathbf{z},t}]^2|O_1 = Y_1,\ldots,O_N = Y_N\rangle\,. \tag{11}$$

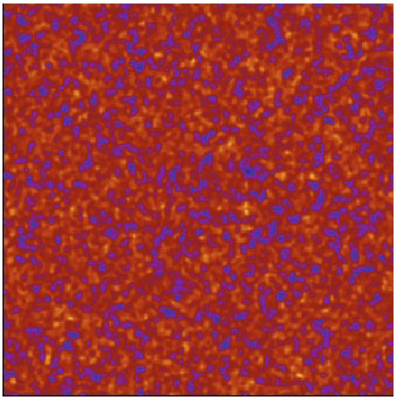

Fig. 1. Snapshot of a solution of the nonlinear SPDE (7).

It is important to note that in contrast to ordinary Langevin equations the number of conditions is not known in advance, because we do not know on how many operators the right-hand side of the SPDE depends. Therefore one has to use different combinations of operators in order to test which combination is more suitable to reproduce the data. As criterion for the correct selection of the operators, one could rearrange (4) (using the estimated coefficients $D_*^{(1,2)}$) to extract the noise, and analyze it for Gaussianity and δ-correlation. A further discussion of this question and related issues will be presented in [17].

The definitions of the drift and diffusion coefficients relate the problem of finding the structure of the SPDE to an estimation of multidimensional conditional averages or, in other words, to a multidimensional regression problem. Since kernel based methods show better convergence properties than histograms, we choose a local linear estimator [18] to determine the conditional averages.

3 Numerical Example

We now turn to a simple example illustrating the outlined procedure. By numerically integrating equation (7) we produce a time series of noisy spatial fields $y(\mathbf{z}, t)$. The data are then used to reconstruct the SPDE. The parameters used for the simulation are $\epsilon = 0.25$ and $q = 1$. The correlation function of the noise is proportional to $\exp(|\mathbf{z} - \mathbf{z}'|^2/(2l_c^2))$ with $l_c = 0.5$. The computational domain is discretized by a 256^2 grid and has a side length of $L = 100$. In Fig. 1, an example of the solution of (7) is depicted. Without noise, the equation shows moving fronts as solutions. Due to the strong noise, these structures are not visible anymore.

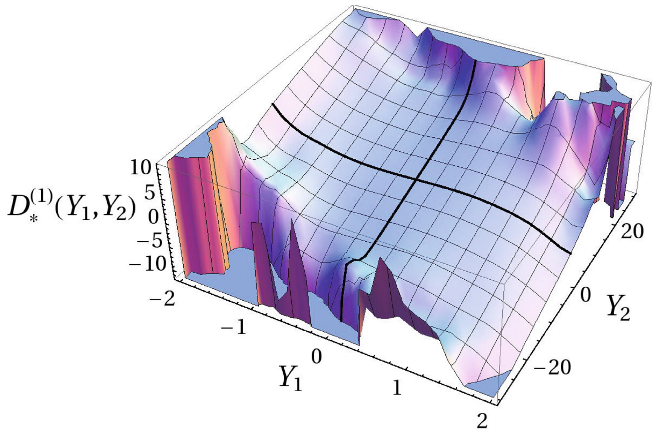

Fig. 2. Visualization of $D_*^{(1)}[Y_1, Y_2]$ estimated from the simulation data. The two thick lines highlight the cuts $Y_1 = 0$ and $Y_2 = 0$ shown in Fig. 3.

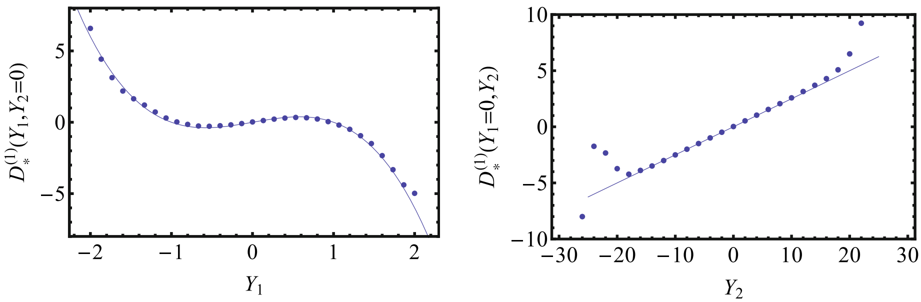

Fig. 3. The estimated drift coefficient (dots) $D_*^{(1)}[Y_1, Y_2 = 0]$ (left), $D_*^{(1)}[Y_1 = 0, Y_2]$ (right) together with the exact result (line).

Given the data from 100 time steps, the unknown coefficients are estimated via the conditional moments

$$D_*^{(1)}[Y_1, Y_2] = \frac{1}{\tau_{\min}} \langle y_{\mathbf{z},t+\tau} - y_{\mathbf{z},t} | Y_1 = y_{\mathbf{z},t}, Y_2 = \Delta y_{\mathbf{z},t} \rangle \tag{12}$$

$$D_*^{(2)}[Y_1, Y_2] = \frac{1}{\tau_{\min}} \frac{1}{2} \langle [y_{\mathbf{z},t+\tau} - y_{\mathbf{z},t}]^2 | Y_1 = y_{\mathbf{z},t}, Y_2 = \Delta y_{\mathbf{z},t} \rangle \,, \tag{13}$$

where $\tau_{\min}$ is the time difference between two subsequent time steps. In Fig. 2, a three-dimensional plot of (12) is shown. One can clearly see the cubic dependence on Y_1 and the linear dependence on Y_2, i.e. $\Delta y_{\mathbf{z}}$. At the boundaries, the estimation of $D^{(1)}$ becomes less reliable due to the small amount of data in this region, which results in larger fluctuations in the conditional moment.

For a better quantitative comparison, cuts along the Y_1-axis and the Y_2-axis are shown in Fig. 3 . In the center, the estimated conditional moment $D_*^{(1)}$

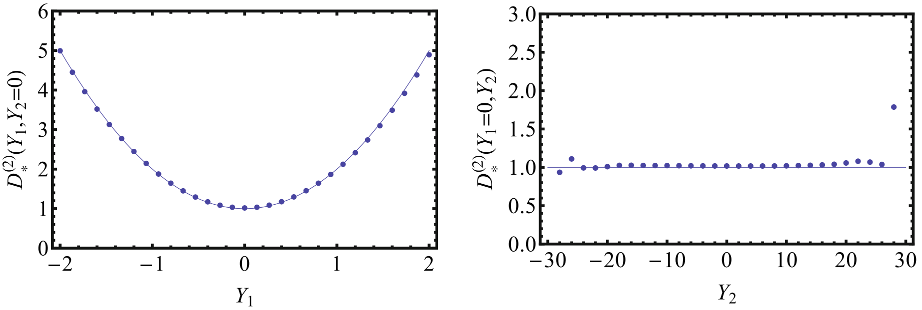

Fig. 4. The estimated diffusion coefficient (dots) $D_*^{(2)}[Y_1, Y_2 = 0]$ (left), $D_*^{(2)}[Y_1 = 0, Y_2]$ (right) together with the exact result (line).

correctly reconstructs the dependence of $D^{(1)}$ on Y_1 and Y_2. In Fig. 4, the same plot is shown for $D_*^{(2)}$: The diffusion coefficient also is estimated correctly.

4 Conlusion and Outlook

In this paper, we have shown that the method to extract drift and diffusion coefficients for Langevin equations from time-series as introduced in [3] can be extended to processes governed by SPDEs. In contrast to pure time-series, one has to estimate conditional moments depending on several conditions, each condition representing one kind of operator in the right hand side of the SPDE. Since the number and the kind of operators are not known in advance one has to test several combinations of operators, or one has to make assumptions on the general structure of the SPDE [17] to reduce the complexity of the problem. Also the methods reviewed in [15] and the extension presented here are suited to analyze experimental data, we want to point out that these methods can also be valuable tools to analyze data from large scale simulations of complex systems where a reduced description in terms of lower number of degrees of freedom is needed. An example for this approach can be found in the context of simulations of large biomolecules [19].

References

1. Haken, H.: Synergetics: Introduction and Advanced Topics. Springer (2004)
2. Risken, H.: The Fokker-Planck Equation: Methods of Solutions and Applications. Springer Series in Synergetics, 2nd edn. Springer (1996)
3. Friedrich, R., Peinke, J.: Description of a Turbulent Cascade by a Fokker-Planck Equation. Phys. Rev. Lett. **78**(5), 863–866 (1997)
4. Renner, C., Peinke, J., Friedrich, R.: Experimental indications for Markov properties of small-scale turbulence. Journal of Fluid Mechanics **433**, 383–409 (2001)
5. Renner, C., Peinke, J., Friedrich, R., Chanal, O., Chabaud, B.: Universality of small scale turbulence. Phys. Rev. Lett. **89**(12), 124502 (2002)

6. Renner, C., Peinke, J., Friedrich, R.: Evidence of Markov properties of high frequency exchange rate data. Physica A: Statistical Mechanics and its Applications **298**(3–4), 499–520 (2001)
7. Kuusela, T.: Stochastic heart-rate model can reveal pathologic cardiac dynamics. Phys. Rev. E **69**(3), 031916 (2004)
8. Prusseit, J., Lehnertz, K.: Stochastic qualifiers of epileptic brain dynamics. Phys. Rev. Lett. **98**(13), 138103 (2007)
9. Kirchner, J., Meyer, W., Elsholz, M., Hensel, B.: Nonstationary Langevin equation: Statistical properties and application to explain effects observed in cardiological time series. Phys. Rev. E **76**(2), 021110 (2007)
10. Bahraminasab, A., Ghasemi, F., Stefanovska, A., McClintock, P.V.E., Friedrich, R.: Physics of brain dynamics: Fokker-Planck analysis reveals changes in EEG $\delta-\theta$ interactions in anæsthesia. New Journal of Physics **11**(10), 12 (2009). 103051
11. Gradišek, J., Siegert, S., Friedrich, R., Grabec, I.: Analysis of time series from stochastic processes. Phys. Rev. E **62**(3), 3146–3155 (2000)
12. Kleinhans, D., Friedrich, R., Nawroth, A., Peinke, J.: An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes. Phys. Lett. A **346**(1–3), 42–46 (2005)
13. Honisch, C., Friedrich, R.: Estimation of Kramers-Moyal coefficients at low sampling rates. Phys. Rev. E **83**, 066701 (2011)
14. Böttcher, F., Peinke, J., Kleinhans, D., Friedrich, R., Lind, P.G., Haase, M.: Reconstruction of complex dynamical systems affected by strong measurement noise. Phys. Rev. Lett. **97**(9), 090603 (2006)
15. Friedrich, R., Peinke, J., Sahimi, M., Tabar, M.R.R.: Approaching complexity by stochastic methods: From biological systems to turbulence. Physics Reports **506**(5), 87–162 (2011)
16. Sagués, F., Sancho, J.M., García-Ojalvo, J.: Spatiotemporal order out of noise. Rev. Mod. Phys. **79**(3), 829–882 (2007)
17. Kamps, O., Peinke, J. (in preparation)
18. Härdle, W.K., Müller, M., Sperlich, S., Werwatz, A.: Nonparametric and Semiparametric Models. Springer (2004)
19. Hegger, R., Stock, G.: Multidimensional Langevin modeling of biomolecular dynamics. J. Chem. Phys. **130**(3), 034106 (2009)