Supplementary material for

Reconstructing Langevin systems from high and low-resolution time series using Euler and Hermite reconstructions

A brief on the appendices

Appendix A discusses the concept of the ‘relaxation’ time scale, a valuable metric for approximating data resolution before fitting a Langevin model. This metric aids in categorizing data into ’low’ and ’high’ resolutions, providing essential insights into the feasibility of data reconstruction and the choice of suitable reconstruction algorithms.

Appendix B delves into two additional data requirements: data stationarity and data Marcovicity, offering guidance on addressing violations of these prerequisites.

Appendix C provides a concise overview of applying maximum likelihood estimation (MLE) for model parameter estimation within Langevin systems.

Appendix D outlines a quasi-MLE approach we term the ‘Euler reconstruction’ cautioned against for datasets with low resolution.

Appendix E introduces a much more accurate but also more expensive quasi-MLE method pioneered by Aït-Sahalia, we term ’Hermite reconstruction’ recommended for datasets with lower resolutions.

Appendix F presents a refined version of Aït-Sahalia's approach in Appendix E, enhancing its applicability to general diffusion models at a slight computational cost.

Appendix G highlights the advantages of ‘spline’ modeling over traditional ‘parametric’ models, emphasizing superior speed, accuracy, and convenience in using them.

Appendix H elaborates on the optimization process for MLE for both Euler and Hermite reconstructions.

For a swift analysis of univariate datasets, we recommend starting with Appendices A and B for foundational insights. Following this, running our MATLAB code 'AllFigures.m' will provide in-depth analyses of the five examples outlined in the main text, accompanied by detailed commentary and explanations. If you have the luxury of time and a desire to delve deeper into our package, we encourage you to explore our tutorial. By executing the code lines provided, you can explore numerous additional examples, enriching your understanding of our methodology and its applications. Please find all the codes and the package in our GitHub repository in the following link:

Appendix A. Investigating the relaxation time scale of data in advance

It is important to have some knowledge or at least a feeling about the level of data resolution before embarking on the analysis. This can roughly let us know to which regime (possible or impossible to reconstruct. See the main text for the details) our dataset belongs to. And if the answer is yes (i.e., dataset resolution let us to fit a diffusion model), it also helps to find to which resolution category (low, medium, high) our dataset belongs to. In the univariate case, data resolution can be assessed by investigating the autocorrelation of data and then estimating the relaxation (or correlation) time of data: one fits the exponential $\exp(-ct)$ to some 'first lags' of the data autocorrelation function, estimate c and finally the relaxation time is $\tau_R = 1/c$.

If the data sampling time Δ is much smaller than the relaxation time(s) τ_R (the regime $\Delta \ll \tau_R$), then our dataset has a high-resolution and we can safely use simple reconstruction schemes like Langevin approach (see the references (Siegert, Friedrich & Peinke 1998; Rinn *et al.* 2016) for a detailed explanation of this reconstruction scheme but refer to (Arani *et al.* 2021) for a brief overview) or Euler scheme (see Appendix D). When the sampling time is of the same order of magnitude with relation time (the regime $\Delta \lesssim \tau_R$) there is still hope to infer the system from the data. However, when the sampling time exceeds the relaxation time (the regime $\Delta > \tau_R$), it signals that there is very little or no hope to infer the underlying system properly, and all reconstruction techniques might fail. Knowing the relaxation time prior to performing reconstruction is important to select the right reconstruction algorithm. In the first regime, we can safely perform a Euler or Langevin reconstruction, as mentioned. In the second regime, we have to use a more accurate reconstruction procedure like the Hermite reconstruction

(Appendices E and F) and use a rather large value of the parameter K (this is a key parameter in Hermite reconstruction). Furthermore, if Δ is between the first and second regime (i.e., data resolution is medium), then we can choose a smaller value of K. Since, Hermite reconstruction with a large K can be time-consuming, having prior knowledge of the data resolution is helpful for selecting the right reconstruction algorithm with proper parameters to strike a balance between accuracy and speed.

Appendix B. Two more data requirements prior to performing reconstruction

Data should be stationary at least in a weak sense. Loosely speaking, stationarity implies that that the statistical properties of the system (and hence the dataset at hand) should remain constant over time. Weak stationarity over a fixed time-window requires that the mean and variance of data to remain unchanged, and the autocorrelation function depends only on the time lag rather than the initial and final times within the window. However, even if stationarity is violated across the entire dataset, one can consider shorter (overlapping) time windows over which data remain stationary. Reconstruction can then be performed on each window separately, and the results can be interpolated between the windows. Stationarity of the data can be tested using the augmented Dickey-Fuller test (Dickey & Fuller 1979), and this was done for our real datasets (simulated data are, indeed, stationary).

Another requirement is that the noise source in a Langevin model should be uncorrelated or white. This implies that the dataset under study should be Markovian, meaning that the future state should depend only on the current state, independent of past histories. However, in this paper, using Hermite reconstruction, we can relax this assumption: if our main dataset is not Markovian, we can find a coarser time scale called the Markov-Einstein (ME) time scale (Friedrich *et al.* 2011), where Markovicity holds, and perform reconstruction on a sample of data whose time scale is the ME time scale or even coarser.

Appendix C. Maximum likelihood estimation (MLE) of Langevin models

Consider a Langevin model (the discussion here works for any stochastic Markovian system)

$$dx = \mu(x;\theta)dt + \sigma(x;\theta)dW \tag{1}$$

Where θ is the vector of unknown parameters. Assume that we have an observation $\{X_0, X_1, \dots, X_N\}$ of this model at discrete times $t = 0,1, \dots, N$ for a particular set of parameter values which we aim to estimate and let Δ be the sampling time. Assume also that $P_X(\Delta, x|x_0; \theta)$ represents the conditional density of $x := X_{t+\Delta}$ given $x_0 := X_t$. The likelihood function for this set of observations is the multi-point density $P_X(X_0, X_1, \dots, X; \theta)$. Due to the Markovian nature of the Langevin model (1) and using the Bayes rule in probability theory we can greatly simplify this multi-point density by the products of the conditional densities as below

$$P_X(X_0, X_1, \dots, X_N; \theta) = P_X(X_0; \theta) \prod_{i=1}^{N} P_X\big(\Delta, X_{i\Delta}|X_{(i-1)\Delta}; \theta\big).$$

Based on the principle of MLE the true parameters of the Langevin model (1) correspond with the global maximum of the likelihood function. It is more convenient to work with log-likelihood function

$$\ell = \sum_{i=1}^{N} \ln\big\{P_X\big(\Delta, X_{i\Delta}|X_{(i-1)\Delta}; \theta\big)\big\}, \tag{2}$$

where the first unconditional term $\ln\big(P_X(X_0)\big)$ is ignored as it has a negligible effect when $N \gg 1$ (see, for instance, the first paragraph in (Aït-Sahalia 2002b)). In this work, we actually consider the negative log-likelihood function $-\ell$ and consider solving a minimization problem, instead. Only for a limited number of Langevin models we know the analytical form of the conditional density in (2). Therefore, our focus lies on quasi-MLE procedures, where we employ approximations of the conditional density in (2) in Appendices D, E, and F.

Appendix D. Euler reconstruction

The Euler reconstruction corresponds with a maximum likelihood estimation of the Langevin model (1) when the sampling time is infinitesimally small, i.e., $\Delta \to 0$. Under this limiting case we can replace the Langevin model (1) with a difference equation based on the Euler-Maruyama discretization

$$X_{t+\Delta} = X_t + \mu(X_t; \theta)\Delta + \sqrt{\Delta}\, \sigma(X_t; \theta)\, \varepsilon_t$$

Where ε_t follows a standard normal distribution. As a result, the conditional density $P_X(\Delta, x|x_0; \theta)$ can be approximated by a normal distribution with mean $x_0 + \mu(x_0; \theta)\Delta$ and variance $\Delta\sigma^2(X_0; \theta)$, that is

$$P_X(\Delta, x|x_0;\theta) \sim P_X^{\text{Euler}}(\Delta, x|x_0;\theta) = \left(2\pi\Delta\sigma^2(x_0;\theta)\right)^{-1/2} \exp\left(-(x - x_0 - \mu(x_0;\theta)\Delta)^2/\left(2\Delta\sigma^2(x_0;\theta)\right)\right)$$

To construct a quasi-MLE procedure one replaces the conditional density in (2) by the approximation P_X^{Euler}. To proceed with parameter estimation this quasi-MLE should be solved. The big advantage of using the Euler approach is due to the fact that its implementation is very fast. Fortunately, based on our experience, this approach also shows a satisfactory performance for datasets with medium resolution.

Appendix E. Hermite reconstruction

Here, we present an advanced reconstruction procedure developed by Aït-Sahalia (Aït-Sahalia 2002a). For the sake of clarity and convenience we keep exactly the same notation as Aït-Sahalia. As discussed in Appendix C, the process of a quasi-MLE requires the estimation of conditional density $P_X(\Delta, x|x_0;\theta)$. The approach of Aït-Sahalia is more accurate than the Euler approach in Appendix D and is based on establishing a convergent Hermite series expansion for the conditional density $P_X(\Delta, x|x_0;\theta)$ and hence a convergent expansion for the log-likelihood function ℓ in (2) using Hermite polynomials. Therefore, we call this approach 'Hermite reconstruction'. This is achieved through two transformations, $X \to Y$ and $Y \to Z$, where the conditional distributions in the transformed space Z will be as close as to the standard normal distribution. Aït-Sahalia then constructs a convergent Hermite series expansion for the transformed variable Z and finally back-transforms to the original variable X. The first transformation $X \to Y$, also known as the Lamperti transform, is as follows

$$Y = \gamma(X;\theta) = \int^X \frac{du}{\sigma(u;\theta)}. \tag{3}$$

By the Itô's lemma, the corresponding Langevin equation (1) for the transformed variable Y has a unit diffusion, i.e., $dY = \mu_Y(Y;\theta)dt + dW$ and

$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta), \tag{4}$$

where γ^{-1} is the inverse function of γ. The conditional density of the variable Y is now closer to normal distribution due to its unit diffusion. Since the distribution of Y can get peaked when the sampling time

Δ is very small Aït-Sahalia performs a second transformation $Y \to Z$ to overcome this problem. Stated other way, he standardizes Y as

$$Z = \frac{Y - y_0}{\sqrt{\Delta}},$$

where y_0 is the corresponding value of x_0 following the transformation Y. Aït-Sahalia shows that Z is close enough to standard normal distribution and then expands the density of Z, P_Z, about standard normal distribution $\phi(z)$. Therefore, the conditional density $P_Z(\Delta, z|y_0;\theta)$ of Z can be expanded about the standard normal distribution, by the Hermite polynomials up to the J^{th} term as below

$$P_Z(\Delta, z|y_0;\theta) \sim P_Z^{(\mathrm{J})}(\Delta, z|y_0;\theta) = \phi(z) \sum_{j=0}^{\mathrm{J}} \eta_Z^{(j)}(\Delta, y_0;\theta)\, H_j(z), \tag{5}$$

where $H_j(z)$ are the classical Hermite polynomials being defined as $H_j(z) = e^{z^2/2} \frac{d^j}{dz^j}\left(e^{-z^2/2}\right)$, $j \geq 0$ so the first few Hermite polynomials read $H_0(z) = 1, H_1(z) = -z, H_2(z) = z^2 - 1, H_3(z) = -z^3 + 3z, \ldots$. Since $H_0(z) = 1$ we have that $\eta_Z^{(0)}(\Delta, y_0;\theta) = 1$. For $j > 1$, the coefficients $\eta_Z^{(J)}(\Delta, y_0;\theta)$ can be found by the orthogonal properties of Hermite polynomials as bellow

$$\eta_Z^{(j)}(\Delta, y_0;\theta) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) P_Z\,(\Delta, z|y_0;\theta)\, dz,$$

which is a conditional expectation. Each of the Hermite coefficients $\eta_Z^{(j)}(\Delta, y_0;\theta)$ can be approximated using Taylor expansion in Δ up to the K^{th} term as bellow (see also formula 4.21 in (Jeisman 2006))

$$\eta_Z^{(j)}(\Delta, y_0;\theta) \sim \eta_Z^{(j,K)}(\Delta, y_0;\theta) = \frac{1}{j!} \sum_{k=0}^{\mathrm{K}} \lim_{y \to y_0} \mathcal{A}_\theta^k\left(H_j(z)\right) \frac{\Delta^k}{k!} \tag{6}$$

Where $\mathcal{A}_\theta$ is the '*infinitesimal generator*' of the process Y (for an overview on the infinitesimal generator and its properties see, for instance, page 18 in (Jeisman 2006)) which is defined as

$$\mathcal{A}_\theta(.) = \mu_Y(y;\theta) \frac{\partial}{\partial y}(.) + \frac{1}{2} \frac{\partial^2}{\partial y^2}(.)$$

Note that the notation $\mathcal{A}_\theta^k$ in (6) is the k^{th} self-composition of the operator $\mathcal{A}_\theta$ (e.g., $\mathcal{A}_\theta^2 = \mathcal{A}_\theta(\mathcal{A}_\theta)$). Finally, we arrive at the following approximation $P_X^{(J,K)}(\Delta, x|x_0;\theta)$ for our quantity of interest, i.e.,

$P_X(\Delta, x|x_0;\theta)$, using the Jacobian formula and by specifying how many Hermite terms (J) in (5) and how many temporal terms (K) in (6) we wish to include

$$P_X(\Delta, x|x_0;\theta) \sim P_X^{(J,K)}(\Delta, x|x_0;\theta) = \frac{\phi(z)}{\sqrt{\Delta}\,\sigma(x;\theta)} \sum_{j=0}^{\mathrm{J}} \eta_Z^{(j,K)}(\Delta, y_0;\theta)\; H_j(z). \tag{7}$$

As Aït-Sahalia argues (and we experienced) the expansion in (5) can be well approximated by including at most the first three terms (so, $\mathrm{J} = 1,2,3$ are often enough). It is important to note that using (6) all coefficients $\eta_Z^{(j,K)}$ of the Hermite expansion of conditional density in (7) can be expressed by μ_Y and its higher derivatives at y_0 up to order 2K-2, i.e., $\mu_Y(y_0)$, $\mu_Y{}'(y_0)$, …, $\mu_Y{}^{[2\mathrm{K}-2]}(y_0)$. We will discuss on this extremely important point in more details in Appendix G.

Appendix F. A refinement to Hermite reconstruction

As discussed in Appendix E, the approach by Aït-Sahalia necessitates the availability of the function γ in (3), which maps the process X to the process Y, in an analytical form. Additionally, even if γ is analytically available, the method requires the inverse function of γ,i.e., γ^{-1}, to be known analytically as well. Consequently, this procedure imposes no restrictions on the drift function, and possible restrictions are related to the diffusion function. While this might not pose a significant limitation for most stochastic models encountered in ecology and biology, including additive Langevin models, there may be cases where such analytical requirements are not feasible (see (Bakshi & Ju 2005)). Moreover, even for models without such restrictions, it remains advantageous to consider the methodology presented in this appendix. This is because the approach here differes slightly from that of Aït-Sahalia, offering enhanced accuracy albeit with slightly longer computational time. Here, we provide a brief overview of the method, omitting detailed explanations. The refinement keeps the transformation $X \to Y$, but considers the following standardization of Y by its true mean and variance, in contrast to unlike Aït-Sahalia's approach, which he refers to as 'pseudo-normalization'(Aït-Sahalia 2002a))

$$\hat{Z} = \frac{Y-\bar{Y}}{\sqrt{\bar{\Sigma}}} = \rho(Z - \bar{Z}),$$

where $\bar{Y}$ and Σ are, respectively, the true mean and variance of Y, $\bar{Z} = (\bar{Y} - y_0)/\sqrt{\Delta}$ is the mean of $Z = (Y - y_0)/\sqrt{\Delta}$ and $\rho = \sqrt{\Delta/\Sigma} = (E(Z^2) - \bar{Z}^2)^{-1/2}$. The variable Z can be rewritten in the following form so that it would be a direct function of X and x_0 rather than Y and y_0

$$Z = \frac{Y - y_0}{\sqrt{\Delta}} = \frac{\int^{X} \frac{du}{\sigma(u;\theta)} - \int^{x_0} \frac{du}{\sigma(u;\theta)}}{\sqrt{\Delta}} = \frac{\int_{x_0}^{X} \frac{du}{\sigma(u;\theta)}}{\sqrt{\Delta}}.$$

Note that in this approach there is no need to know the integral $\int_{x_0}^{X} \frac{du}{\sigma(u;\theta)}$ analytically and only a numerical approximation is enough. The way this refinement standardizes Y and rewrites Z allows to do all the calculations directly in terms of the original drift function $\mu(x)$ and diffusion function $\sigma(x)$ and avoids expressing the results in terms of $\mu_Y(y)$ in (4), i.e., the drift function of the process Y. A Hermite expansion approximation up to the order J for the conditional density $P_X(\Delta, x|x_0;\theta)$, using the Jacobian formula applied to the density of $\hat{Z}$, reads

$$P_X(\Delta, x|x_0;\theta) \sim P_X^{(J)}(\Delta, x|x_0;\theta) = \frac{\rho}{\sqrt{\Delta}\,\sigma(x;\theta)} \phi(\rho(Z - \bar{Z})) \sum_{j=0}^{\mathrm{J}} \eta_Z^{(j)}(\Delta, x_0;\theta)\ H_j(\rho(Z - \bar{Z}))\,, \qquad (8)$$

in which the Hermite coefficients are obtained as

$$\eta_Z^{(j)}(\Delta, x_0;\theta) = \frac{1}{j!} E\left(H_j(\hat{Z})\right), \qquad (9)$$

where H_j are classical Hermite polynomials of order $j \geq 0$ (note that in (8) the first and second coefficients $\eta_Z^{(1)}$and $\eta_Z^{(2)}$ can be shown to be 0 (Bakshi & Ju 2005). Therefore, based on this algorithm j actually starts from 3). Stated other way, the Hermite expansion coefficients $\eta_j(\Delta, x_0;\theta)$ in (9) are expressed in terms of moments of the powers of $\hat{Z}$, where the later can be expressed by the moments of the powers of Z, i.e., $E(Z^i)$, $i \geq 0$. $E(Z^i)$ can be approximated by a Taylor expansion in Δ up to order K as

$$E(Z^i) \sim \sum_{k=0}^{\mathrm{K}} \lim_{x \to x_0} \mathcal{B}_\theta^k(Z^i) \frac{\Delta^k}{k!},$$

where $\mathcal{B}_\theta$ is the '*infinitesimal generator'* of the process X which is defined as

$$\mathcal{B}_\theta(.) = \mu(x;\theta) \frac{\partial}{\partial x}(.) + \frac{1}{2}\sigma^2(x) \frac{\partial^2}{\partial x^2}(.).$$

Clearly, this leads us to an approximation in terms of the parameters J and K, say $P_X^{(J,K)}(\Delta, x|x_0;\theta)$, for the conditional density of X in (8) where the Hermite coefficients $\eta_Z^{(j)}$ are approximated by $\eta_Z^{(j,K)}$. It is important to note that all Hermite expansion coefficients $\eta_Z^{(j,K)}$ can be expressed by $\mu(x)$ and its higher order derivatives at x_0 up to order 2K-2 and by $\sigma(x)$ and its higher derivatives at x_0 up to order 2K-1. We will discuss on this extremely important point in more details in Appendix G.

Appendix G. Several reasons for the superiority of spline modeling in Hermite reconstruction

In Appendix E we argued that using the approach by Aït-Sahalia Hermite expansion coefficients can be expressed by the drift function of the process Y, i.e., μ_Y, and its higher order derivatives up to order 2K-2, i.e., μ_Y, μ_Y', …, $\mu_Y{}^{[2K-2]}$. Likewise, we also argued in Appendix F that using the refined approach the corresponding Hermite expansion coefficients can be expressed by the drift and diffusion functions of the process X up to orders 2K-2 and 2K-1, respectively, i.e., $\mu, \mu', \dots, \mu^{[2K-2]}, \sigma, \sigma', \dots, \sigma^{[2K-1]}$. Therefore, using both algorithms, as the differentiation order increases, it generally leads to significant computational complexities, especially as K increases. The computational complexities also increase as the number of Hermite terms (J) in the expansion of conditional density increases. However, in practice, the first or first two Hermite terms are sufficient in both algorithms (Aït-Sahalia 2002b) (i.e., in the first algorithm we at most need J=1,2 and in the second one we at most need J=3,4). Consider, for instance, a stochastic version of the overgrazed model of May (May 1977) with drift part $\mu(x) = rx\left(1-\frac{x}{K}\right) - \frac{\gamma x^2}{a^2+x^2}$. As the differentiation order of $\mu(x)$ increases, the algebraic expressions become longer, imposing a substantial computational burden. In contrast, higher order derivatives of a cubic polynomial (which approximately fits the May model) result in shorter expressions that eventually vanish at order = 4. This suggest using polynomial forms for $\mu_Y(y)$ in the first algorithm and polynomial forms for $\mu(x)$ and $\sigma(x)$ in the second algorithm.

While polynomials are convenient candidate models, they may not be ideal due to the need for higher-order polynomials to adequately represent the nonlinearities of complex models. Instead, piecewise

polynomials, or splines, are ideal candidate models due to their flexibility and simplicity. Splines offer great accuracy, comfort, and speed in the reconstruction process. They are 'flexible' enough to capture unknown linearities in the data-generating process and have a 'simple' polynomial form in their building blocks, which reduces computational burden significantly. Since cubic splines typically suffice in practice, they are more attractive than polynomials from a computational standpoint. If $\{x_1, x_2, \dots, x_M\}$ denotes a 'knot sequence' across the state space then the model parameters in spline reconstruction are the unknown values of drift and diffusion functions at these knots. It worth mentioning that the reconstruction approach using spline forms is parametric. However, the modeling feels 'non-parametric' as if there is no explicit model involved. Therefore, we call it 'spline modeling' to distinguish it from 'typical' parametric models.

Here, we highlight additional reasons for the superiority of spline modeling beyond speed. The second reason for the superiority of splines is that splines are linear functions in terms of parameters. This significantly enhance the optimization process as it often leads to an optimization problem with one or very few local minima. Third, in general we have no idea how to choose a proper model to try, unless there is robust empirical or theoretical justification. Even in cases where a parametric model is preferred, it is recommended to consider fitting a spline model to the data first to gain insights into a proper parametric model form to try. The fourth reason is that in the process of estimating model parameters using MLE, the optimization algorithm needs to search within a sufficiently large parameter space containing the true but yet unknown parameters. However, it is generally challenging to define a proper bounded parameter region around the true solution when considering a typical parametric model. In contrast, spline modeling offers a convenient way to establish a proper bounded parameter space for the algorithm to search within. The reason is simple: in spline modeling, our parameters hold a special meaning as they represent the values of drift and diffusion functions at knots. The fifth reason is that, when using splines, our models will not exhibit global sensitivity to parameters. A change in the value of a parameter at a single knot will not propagate across the entire knot sequence. Instead, the effect is localized over an evenly spaced knot sequence, which we always choose (De Boor & De Boor 1978). A

model with sharp sensitivity to parameters places significant pressure on optimization procedure in terms of both accuracy and speed, potentially leading to failure.

Quadratic splines: A more efficient approach for Hermite reconstruction than cubic splines

Although the use of cubic spline forms leads to a huge reduction in the computational complexity, we might still need a further reduction of computational burden. Quadratic splines have a slightly less desirable smoothing property in comparison with cubic splines but the use of quadratic splines relative to cubic splines greatly reduces the computations as parameters J and/or K increase. Let's assume we want to follow the methodology in Appendix ~~Appendix E~~ (the same logic works if we follow the methodology in Appendix ~~Appendix F~~). Using cubic spline forms for drift and diffusion functions one does not need to express the Hermite coefficients in terms of higher derivatives of drift and diffusion functions up to orders 2K-2 and 2K-1, respectively. Instead, the Hermite expansion coefficients can be expressed by the derivatives of the drift and diffusion functions up to order 3 only, i.e., $\mu, \mu', \mu'', \mu''', \sigma, \sigma', \sigma'', \sigma'''$, and we get rid of higher order derivatives. Clearly, this leads us to a huge simplification of conditional density in (8) especially when J and/or K is big. On the other hand, a further level of simplification is achievable by using quadratic spline forms since we can get rid of third order derivatives and express the Hermite coefficients by $\mu, \mu', \mu'', \sigma, \sigma', \sigma''$ only. In additive models (in either case of cubic and quadratic splines) an even higher level of simplicity is attainable since $\sigma' = \sigma'' = 0$. Therefore, the reduction of computational burden from cubic splines to quadratic splines will not be felt sharply for additive models. However, in multiplicative models, which we often desire, such a reduction makes a big difference. Because of this huge reduction of computational time, we could often tackle Hermite reconstruction using quadratic splines rather than cubic splines.

A note on data standardization for spline modeling

When the range of data encompasses big numbers there is a risk for numerical instabilities. In such cases, it is better to standardize the data first, perform the analysis on the standardized data and, at the end back transform the results to the original scale of data. This is especially handy for linear models such as spline models. Consider the standardization of a state variable x, i.e., $z = (x - m_{\text{data}})/s_{\text{data}}$ in

which m_{data} and s_{data} are the mean and standard deviation for a dataset of the process x. Now, assume that the Langevin model $dz = \mu_z(z)dt + \sigma_z(z)dW$ describes the dynamics of the transformed process z. Then the corresponding Langevin model for the process x has the following drift and diffusion functions

$$\mu_x(x) = s\,\mu_z(z), \sigma_x(x) = s\,\sigma_z(z).$$

Assume that $\{x_1, x_2, \dots, x_M\}$ and $\{z_1, z_2, \dots, z_M\}$ are the knot sequences for the processes x and z. Assume further that the estimated drift and diffusion parameters of a spline model for the corresponding standardized data are $A = \{\mu_{1z}, \mu_{2z}, \dots, \mu_{Mz}\}$ and $B = \{\sigma_{1z}, \sigma_{2z} \dots, \sigma_{Mz}\}$, respectively (these are the estimated values of the drift and diffusion functions of the z process at the knot sequence $\{z_1, z_2, \dots, z_M\}$). Now, to back transform everything to the original scale (i.e., the x process) all we need to do is to multiply the elements of A and B by the data standard deviation s_{data} to find the estimated parameters of the process x over the knot sequence $\{x_1, x_2, \dots, x_M\}$.

Appendix H. Gradient descent and grey wolf optimizer algorithms used to solve the MLE

In order to estimate the parameters of the Langevin model (1) using the MLE framework we should find the global minimum of negative log-likelihood function $-\ell$ where ℓ is defined in (2) in Appendix C. The approach to tackling the MLE problem significantly differs between Euler and Hermite reconstructions. Euler reconstruction, the objective function is fully defined across the entire parameter space. However, in Hermite reconstruction, the objective function may become undefined for certain parameter values as they deviate from the true global minimum, a discrepancy that becomes more pronounced as data resolution decreases. This is because as the data resolution declines the Hermite expansion of conditional density in (8), or in (7), might converge to a density which may not be entirely positive across the state space (although it always integrates to 1), leading to a partially defined objective function. Typically, within a small enough neighborhood around the optimum parameter values, the Hermite series converges to a positive density. Yet, as this neighborhood expands, the series may not maintain a positive density for some parameter values, which we term ‘illegitimate’ solutions (as opposed to ‘legitimate’ solutions). In practice, the size and geometry of the region where objective

values are defined, i.e., the legitimate region, have a complex dependency on factors such as data resolution, model complexity, and the level of approximation adopted in selecting the K and J values in the Hermite reconstruction algorithm. Therefore, in the case of Hermite reconstruction, we face not only a non-smooth objective function but also a partially defined one, where the density of undefined (i.e., illegitimate) objectives increases the further we move from the global minimum. To handle this constraint, we employ a 'death penalty' technique, which assigns a very high positive number to illegitimate objective values. Consequently, classical optimization routines based on the gradient descent algorithm cannot straightforwardly solve this complex optimization problem.

To enable the use of gradient descent algorithms in our partially-defined optimization problem, we employ a two-phase algorithm. The initial phase involves addressing the Euler reconstruction to obtain the parameter estimate $\hat{\theta}_{Euler}$. For this optimization, we adopt a strategy of employing multiple starting points. This approach helps us to navigate around the potential issue of being trapped in a local minimum. However, based on our extensive experience, when the model is linear with respect to its parameters, we typically encounter an optimization landscape characterized by a unique local minimum that also serves as the global minimum or relatively few local minima. This observation underscores the practical benefits of working with such linear models, particularly spline models, which tend to simplify the optimization process significantly.

In the second phase, we focus on exploring the vicinity of $\hat{\theta}_{Euler}$, where we anticipate a higher density of legitimate solutions. Our aim here is to identify several legitimate solutions that can serve as a basis for constructing a surrogate model. The feasibility and effectiveness of this surrogate model hinge on the concentration of legitimate solutions near $\hat{\theta}_{Euler}$. Should we successfully establish a surrogate model, it often becomes possible to identify its global minimum $\hat{\theta}_{Surr}$, which usually lies in close proximity to the true global minimum. This surrogate solution becomes a valuable starting point for deploying a gradient descent algorithm on the original optimization problem. In instances where a surrogate model proves unattainable, the legitimate solutions already identified are utilized as initial points for a gradient descent algorithm on the original optimization problem. Regardless of the presence of a surrogate model, these starting points, which are hoped to have a high density of legitimate points

near them, can pave the way for the application of the gradient descent algorithm. Our aim is to improve $\hat{\theta}_{Euler}$ through this process, aspiring to discover a new solution $\hat{\theta}_{Hermite}$, which either improves upon $\hat{\theta}_{Euler}$ or, in the worst case, matches it. Should the complexity of the chosen model prevent any enhancement of $\hat{\theta}_{Euler}$, we consider simplifying the model and iterating the process to achieve a superior solution. Our extensive experience, supported by numerous case studies, demonstrates that this algorithm frequently yields superior outcomes when spline modeling is employed, as opposed to traditional parametric models. This is extensively documented through various examples in section 11 of our tutorial.

Accessing the uncertainty of the parameters

To calculate the variance of the estimated parameter vector, say $\hat{\theta}$, the following Fisher information (FI) matrix is needed

$$\mathcal{F} = -\mathrm{E}\big(\mathcal{H}(\theta)\big)$$

where $\mathcal{H}(\theta) = \partial^2\ell/\partial\theta\partial\theta^T$ is the Hessian matrix, i.e., the second order partial derivatives of the log-likelihood function with respect to θ (in case of minimization, as is our case, $-\ell$ is used as objective function so that the $\mathcal{F}$ matrix does not include the negative sign). The 'observed' Fisher information approximates the $\mathcal{F}$ matrix as $\mathcal{F} \sim -\mathcal{H}\big(\hat{\theta}\big)$. Consequently, the variance-covariance matrix, $\Sigma = \mathcal{F}^{-1}$, i.e., the inverse of the Fisher information matrix, is estimated by $\Sigma \sim \big[-\mathcal{H}\big(\hat{\theta}\big)\big]^{-1}$. The standard error of the i$^{\text{th}}$ parameter, say σ_{θ_i}, which is the square root of the i$^{\text{th}}$ diagonal element in Σ, is thus estimated from the square root of the i$^{\text{th}}$ diagonal element of $\big[-\mathcal{H}\big(\hat{\theta}\big)\big]^{-1}$. To assess the uncertainties in the case of Hermite reconstruction, extra caution is necessary when estimating the $\mathcal{F}$ matrix. The crux of challenge lies in ensuring that the objective values used for estimating the $\mathcal{F}$ matrix are legitimate, not just at the estimated parameters $\hat{\theta}$, which is known to be legitimate, but also at neighboring parameter values in close proximity to $\hat{\theta}$, which we are not sure. To tackle this complexity, if an illegitimate objective value is encountered, we apply an infinitesimal perturbation to the parameter in question, ensuring the perturbed parameter yields a legitimate objective.

An interesting point concerning the uncertainty of drift and diffusion parameters

In our analysis across numerous examples, we have observed a higher accuracy in the estimation of the noise parameters compared to that of drift parameters. This phenomenon is not purely empirical but is supported by theoretical underpinnings. For readers interested in a deeper mathematical explanation, we direct them to references (Sorensen 2007; Tang & Chen 2009; Chang & Chen 2011), where detailed justifications are provided.

References

Aït-Sahalia, Y. (2002a) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica,* **70,** 223-262.

Aït-Sahalia, Y.J.E. (2002b) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. **70,** 223-262.

Arani, B.M., Carpenter, S.R., Lahti, L., Van Nes, E.H. & Scheffer, M.J.S. (2021) Exit time as a measure of ecological resilience. **372,** eaay4895.

Bakshi, G. & Ju, N.J.T.J.o.B. (2005) A Refinement to Aït-Sahalia's (2002)"Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach". **78,** 2037-2052.

Chang, J. & Chen, S.X. (2011) On the approximate maximum likelihood estimation for diffusion processes.

De Boor, C. & De Boor, C. (1978) *A practical guide to splines*. springer-verlag New York.

Dickey, D.A. & Fuller, W.A.J.J.o.t.A.s.a. (1979) Distribution of the estimators for autoregressive time series with a unit root. **74,** 427-431.

Friedrich, R., Peinke, J., Sahimi, M. & Tabar, M.R.R.J.P.R. (2011) Approaching complexity by stochastic methods: From biological systems to turbulence. **506,** 87-162.

Jeisman, J.I. (2006) Estimation of the parameters of stochastic differential equations. Queensland University of Technology.

Lehle, B.J.J.o.S.P. (2013) Stochastic time series with strong, correlated measurement noise: Markov analysis in n dimensions. **152,** 1145-1169.

May, R.M. (1977) Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature,* **269,** 471-477.

Rinn, P., Lind, P.G., Wächter, M. & Peinke, J.J.a.p.a. (2016) The Langevin Approach: An R Package for Modeling Markov Processes.

Sorensen, M. (2007) Efficient estimation for ergodic diffusions sampled at high frequency. *CREATES Research Paper*.

Tabar, R. (2019) *Analysis and data-based reconstruction of complex nonlinear dynamical systems*. Springer.

Tang, C.Y. & Chen, S.X. (2009) Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics,* **149,** 65-81.