Correlation integral likelihood for stochastic differential equations

Heikki Haario, Janne Hakkarainen, Ramona Maraia, Sebastian Springer

Abstract A new approach was recently introduced for the task of estimation of parameters of chaotic dynamical systems. Here we apply the method for stochastic differential equation (SDE) systems. It turns out that the basic version of the approach does not identify such systems. However, a modification is presented that enables efficient parameter estimation of SDE models. We test the approach with basic SDE examples, compare the results to those obtained by usual state-space filtering methods, and apply it to more complex cases where the more traditional methods are no more available.

1 Introduction

The difficulty of estimating parameters of chaotic dynamical models is related to the fact that a fixed model parameter does not correspond to a unique model integration, but to a set of quite different solutions as obtained by setting slightly different initial values, selecting numerical solvers used to integrate the system, or tolerances specified for a given solver. But while all such trajectories are different,

Heikki Haario
School of Engineering Science, Lappeenranta University of Technology, e-mail: heikki.haario@lut.fi

Janne Hakkarainen
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland
Earth Observation, Finnish Meteorological Institute, e-mail: janne.hakkarainen@fmi.fi

Ramona Maraia
School of Engineering Science, Lappeenranta University of Technology, e-mail: ramona.maraia@lut.fi

Sebastian Springer
School of Engineering Science, Lappeenranta University of Technology, e-mail: sebastian.springer@lut.fi

they approximate the same underlying attractor and should be considered in this sense equivalent. In [3] we introduced a distance concept for chaotic systems based on this insight. Modifying one of the fractal dimension definitions, the correlation dimension, we calculate samples from the phase space of the system and map these points onto a stochastic vector. The vector turns out to be Gaussian, providing a natural likelihood concept that quantifies the chaotic variability of points of a chaotic system within a given setting of observations.

Stochastic differential equation (SDE) systems behave partly in a similar way: each integration of a given system with fixed model parameters produces a different realization. This calls for methods that can quantify the variability of the realizations. On the other hand, the stochastic nature of a SDE system is clearly different from the chaotic variability of a deterministic chaotic system. Consequently, the phase space behavior of each type of systems is different as well. The aim of this work is to study to which extent the parameter estimation approach originally developed for chaotic systems can be applied to SDE models.

The rest of the paper is organized as follows. In the Background section we recall the correlation integral likelihood concept and outline the results obtained for chaotic systems.In Numerical experiments we exhibit the performance of the method for the Ornstein-Uhlenbeck model and extensions of it, together with comparisons to more standard, Kalman filter based methods.

2 Background

The standard way of estimating parameters of dynamical systems is based on the residuals between the data and the model responses, both given at the time points of the measurements. Supposing the statistics of the measurement error is known, a well defined likelihood function can be written. The maximum likelihood point is typically considered as the best point estimator, and it coincides with the usual least squares fit in the case of Gaussian noise. The full posterior distribution of parameters can be sampled by Markov chain Monte Carlo (MCMC) methods. The approach has become routine for the parameter estimation of deterministic models in Bayesian inference.

The estimation of the parameters of stochastic models is not so straightforward. A given model parameter does not correspond to a fixed solution, but a whole range of possible realizations. Several methods have been proposed to overcome this difficulty. State-based approaches estimate the joint distribution of the state vector and the parameters. The likelihood for the parameter is obtained as a marginal distribution, effectively by 'integrating out' the state space. This approach is routine in the context of linear time series modeling, and implemented by the likelihood obtained by application of the Kalman filter formulas, see [2, 7, 11].

Here we study a different way of characterizing the stochastic variability of the state space. Supposing that a sufficient amount of data is available, we create a mapping from it onto a feature vector. The mapping is based on averaging, and

the vector turns out to be asymptotically Gaussian. From real data, the mean and covariance of this Gaussian distribution can be empirically estimated. Thus we have a likelihood available, both for maximum likelihood parameter estimation and for MCMC sampling of the parameter posterior. The idea is the same as that earlier used for estimating parameters of chaotic models in [3] , but certain modifications are needed for SDE systems. We discuss the basic setting of the approach below, as well as the reasons behind the modifications needed.

2.1 Likelihood via Filtering

A standard way of estimating the parameters with stochastic models is to use filtering methods for constructing the likelihood (see, e.g., [2, 7, 11] for basic references and implementation, or [8] for recent variant). By using the Kalman filter, the idea is to build the marginal filter likelihood from the prediction residual $\mathbf{r}_k$ and its error covariance matrix $\mathbf{C}_k^{\mathbf{r}}$ at each filtering time step k.

The basic linear Kalman filter is written as a pair

$$\mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \boldsymbol{\xi}_k, \tag{1}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\varepsilon}_k, \tag{2}$$

where $\mathbf{x}_k$ is the state and $\mathbf{y}_k$ is the measurement vector. Matrix $\mathbf{M}_k$ is the linear state-space model, and matrix $\mathbf{H}_k$ is the observation operator that maps from the state space to the observation space. The error terms $\boldsymbol{\xi}_k$ and $\boldsymbol{\varepsilon}_k$ are typically assumed zero mean and Gaussian: $\boldsymbol{\xi}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$ and $\boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \mathbf{R}_k)$. This dynamical system is solved using Kalman filter formulas (see, e.g., [11]).

Given a set of observation $\mathbf{y}_{1:K}$ and the parameter vector $\boldsymbol{\theta}$, the marginal filter likelihood is written as

$$p(\mathbf{y}_{1:K}|\boldsymbol{\theta}) = \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \left[\mathbf{r}_k^{\mathrm{T}} (\mathbf{C}_k^{\mathbf{r}})^{-1} \mathbf{r}_k + \log |\mathbf{C}_k^{\mathbf{r}}| \right] \right), \tag{3}$$

where $|\cdot|$ denotes the matrix determinant. Here the prediction residual and its error covariance matrix are calculated by the formulas

$$\mathbf{r}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^{\mathrm{prior}}, \tag{4}$$

$$\mathbf{C}_k^{\mathbf{r}} = \mathbf{H}_k \mathbf{C}_k^{\mathrm{prior}} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k, \tag{5}$$

where $\mathbf{x}_k^{\mathrm{prior}}$ is the prior estimate computed from the previous state $\mathbf{x}_k^{\mathrm{prior}} = \mathbf{M}_k \mathbf{x}_{k-1}^{\mathrm{est}}$, and $\mathbf{C}_k^{\mathrm{prior}} = \mathbf{M}_k \mathbf{C}_{k-1}^{\mathrm{est}} \mathbf{M}_k^{\mathrm{T}} + \mathbf{Q}_k$ is the respective error covariance matrix. Note that the normalizing "constant" $|\mathbf{C}_k^{\mathbf{r}}|$ has to be included, since it depends on the parameters via the prediction model.

This approach is well established in the framework of linear time series or linear SDE systems, where the additive model noise is known or may be estimated, as one of the unknowns in the vector $\boldsymbol{\theta}$. In case the drift part of the system (1) is nonlinear, one still may use the approach in the extended Kalman filter (EKF) form, based on the approximation by linearization. Often the EKF approach is also applied to filtering of deterministic systems. In that setting the model error term is rather postulated and interpreted as a measure of bias. The covariances $\mathbf{Q}$ and $\mathbf{R}$ represent then our trust on the model and data, respectively, previous work [5], motivated by closure parameter estimation in climate research, is an example of this approach. A related option is to employ ensemble filtering. In [12] this approach was employed in order to tune the ensemble prediction system parameters. It was observed, however, that the method resulted in a highly stochastic cost function that prevented a successful application of parameter optimization algorithms. Moreover, the tuning parameters of the filter itself may bias the model parameter estimation, see [6]. Recently, some additional criticism toward using the filtering for estimating the parameters in real-world applications (other than finance) has been presented see [10].

Next, we present the method developed in [3] for deterministic chaotic systems. While computationally more demanding, it is free of the pitfalls listed above, and can be applied to stochastic systems more general than the class of additive noise given by (1).

2.2 Correlation Integral Likelihood.

In this section we briefly summarize the correlation integral likelihood method used for creating a likelihood for complex patterns [3].

Let us use the notation $\mathbf{s} = \mathbf{s}(\boldsymbol{\theta}, \mathbf{x})$ for a state vector $\mathbf{s}$ that depends on parameters $\boldsymbol{\theta}$ and other inputs $\mathbf{x}$ such as, e.g., the initial values of a dynamical system. We consider two different trajectories, $\mathbf{s} = \mathbf{s}(\boldsymbol{\theta}, \mathbf{x})$ and $\tilde{\mathbf{s}} = \mathbf{s}\left(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{x}}\right)$, evaluated at $N \in \mathbb{N}$ time points $t_i,\ i = 1:N$, with explicit dependency on the respective initial and parameter values. For $R \in \mathbb{R}$, the *modified correlation sum* is defined as

$$C(R, N, \boldsymbol{\theta}, \mathbf{x}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{x}}) = \frac{1}{N^2} \sum_{i,j} \#\left(\left\|\mathbf{s}_i - \tilde{\mathbf{s}}_j\right\| < R\right). \tag{6}$$

In the case $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$ and $\tilde{\mathbf{x}} = \mathbf{x}$ the formula reduces to the well known definition of *correlation sum*, the *Correlation Integral* is then defined as the limit $C(R) = \lim_{N\to\infty} C(R, N)$, and the *Correlation Dimension* ν as the limit

$$\nu = \lim_{R\to 0} \log C(R) / \log(R).$$

In numerical practice, the limit $R \to 0$ is approximated by the small scale values of the ratio above, by the log-log plot obtained by computing $\log C(R)$ at various values of $\log R$.

However, we do not focus on the small-scale limit as in the above definition, but rather use the expression (6) at all relevant scales R to characterize the distance between two trajectories. For this purpose, a finite set of decreasing radii $R = (R_k),\ k = 1,...,M$, is chosen. The radii values R_k are selected so as to involve both small and large scale properties of the trajectory samples. Typically, the radii are chosen as $R_k = b^{-k}R_0$, with $R_0 = \max_{i,j} \left\| s_i - \tilde{s}_j \right\|$ or somewhat larger to ensure that all the values are inside the largest radius. The values of M and b should be chosen in a way that R_M is small enough. For more details see [3].

Consider now the case with given data $\mathbf{s}_i$, which corresponds to the case of a fixed but unknown model parameter vector, $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} = \boldsymbol{\theta}_0$. We select two subsets $\mathbf{s}$ and $\tilde{\mathbf{s}}$ of size N from the data (see more details below). If we fix the radii values $R = (R_k), k = 1,...,M$ the expression (6) defines a M dimensional vector with components $y_k = C(R_k, \boldsymbol{\theta}_0, \mathbf{x})$. A training set of these vectors is created by repeatedly selecting the subsets $\mathbf{s}$ and $\tilde{\mathbf{s}}$. The statistics of this vector can then be estimated in a straightforward way.

Indeed, the expression (6) is an average of distances, so by the Central Limit Theorem it might be expected to get Gaussian. More exactly, each expression $\mathbf{y} = (\mathbf{y}_k)$ gives the empirical cumulative distribution function of the respective set of distances. The basic form of the Donsker's theorem tells that empirical distribution functions asymptotically tend to a Brownian bridge. In a more general setting, close to what we employ here, the Gaussianity was established by Borovkova et.al. [1].

At a pseudo code level the procedure can be summarized as follow:

- Using the measured data, create a training set of the vectors $\mathbf{y}$ for fixed radii values (R_k) by sampling data at measurement times (t_i).
- Create the empirical statistical distribution of the training set $\mathbf{y}$ as a Gaussian likelihood, by computing the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ of the training set vectors.
- Find the maximum likelihood model parameter $\boldsymbol{\theta}_0$ of the distribution

$$P_{\theta_0}(\theta, x) \sim \exp -\frac{1}{2}(\mu - y(\theta, x))^T \Sigma^{-1} (\mu - y(\theta, x))$$

- Sample the likelihood to find those model parameters $\boldsymbol{\theta}$ for which the vector $\mathbf{y} = C(\boldsymbol{\theta_0}; \mathbf{x}; \boldsymbol{\theta}; \tilde{\mathbf{x}})$ belongs to the distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The first step will be discussed more in detail in the examples below. Note that in [3] we assumed a parameter value $\boldsymbol{\theta}_0$ given and created the training data by model simulations, while here we start with given data, create the training set from subsets of data, and proceed to estimate a maximum likelihood parameter value $\boldsymbol{\theta}_0$.

Remark. In all the cases the prior distribution is assumed to be flat uniform.

3 Numerical experiments

The main objective of this section is to modify the Correlation integral likelihood (CIL) method for identifying SDE system parameters. The new version of the method is compared with the filter likelihood results. After this validation the approach is applied to a more complex case.

3.1 Ornstein-Uhlenbeck with modification for dynamics.

We start with a basic SDE example, the Ornstein-Uhlenbeck (OU) process model. We use it as a benchmark to verify that the CIL method is able to produce results comparable to standard filter likelihood methods in a setting where these classical methods perform perfectly well. The OU process equation is given by

$$dX_t = -\theta X_t dt + \sigma dW_t. \tag{7}$$

In the numerical simulations, we use $\theta = 10$ and $\sigma = 1.5$ as the 'true' values. For simplicity, the mean value of the process is set to zero (but all the results and conclusions are valid for a non-zero mean as well). We create a data signal of 3000 points on the time interval $[0, 30]$, with initial value $X = 0$.

Figure 1 exhibits the signal used as data, obtained by integration of (7) using the Euler-Maryama method, with a time step $dt = 0.01$ and using a fixed Gaussian $N(0, \sigma^2)$ as the diffusion part. The figure presents three different realizations. Note that essentially the same results as those given below were obtained by any realizations used.

Let us first apply the CIL method in the basic form. To create the sample sets $\mathbf{s}_i$ we randomly select 1500 of the data points of the signal in Fig. 1 and use the rest of the points as $\mathbf{s}_j$ to get the set of distances needed in (6). This process is repeated around 2000 times to get a representative set of the feature vectors $\mathbf{y}$. The likelihood is then obtained by computing the mean and covariance of the training vectors $\mathbf{y}$, and the Normality of the vectors can be verified by the usual χ^2 test.

Next, we find the distribution of the model parameters θ, σ that follows this distribution by creating a MCMC chain of length 20000 using adaptive Metropolis [4]. The result in Fig. 2 shows, however, that the model parameters are not identified by this likelihood. This situation is different from those reported in [3], and several unpublished cases, for chaotic systems, where the same likelihood construction is able to identify the model parameters.

We conclude that too much information is lost in the mapping from data to the feature vectors $\mathbf{y}$. Indeed, this is not surprising in view of the fact that only the distances between randomized data points is considered, while the order or differences between consecutive points is lost. A trivial example is given by any vector or random points: sorting it in increasing order gives a definitely different signal, but with just the same set of points and distances between them.

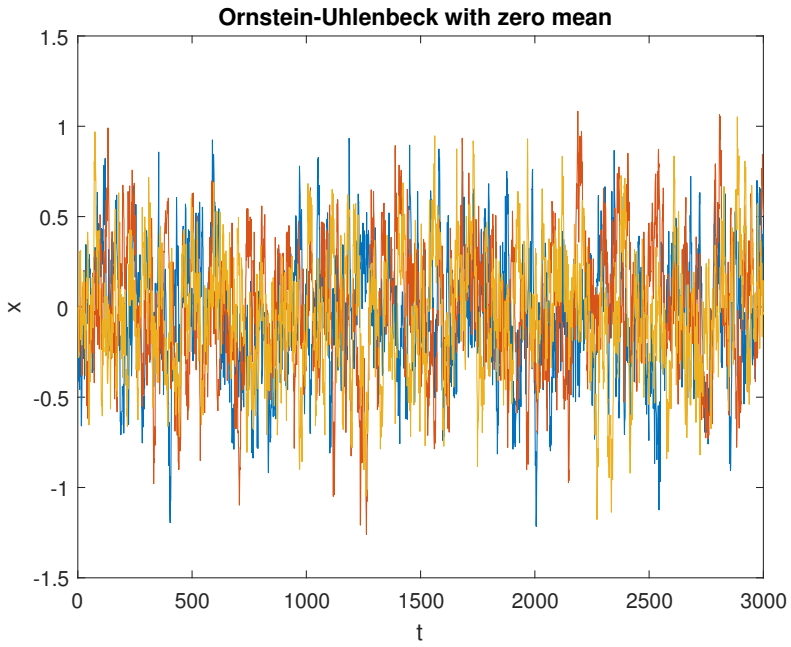

Fig. 1: Ornstein-Uhlenbeck signal used for the experiments.

Intuitively, the mean reverting dynamics is lost here, so some additional modification of the method is needed. The large posterior in Fig. 2 exhibits only what it is programmed to do: signals whose distance distributions remain close, which in this case does not characterize the signals. The feature vector can be modified in various ways. Here we present the impact of extending it in the obvious way: we include the differences between consecutive points. We create the feature vectors separately for the signal and for the differences. The final feature vector is created by concatenating the curves, and the Gaussianity of the combined vector can be tested by the χ^2 test. Figure 2 illustrates the posterior obtained using three different levels of information: only the data signal, only difference between consecutive points, and both together. We see how the first two are not enough, while the posterior of the extended case, practically the intersection of the two other posteriors, significantly improves the identification.

Next, we compare the Correlation Integral Likelihood results with that obtained by filter likelihood estimation based on Kalman filtering. We use the same data signal as above, using all the points $X_k, k = 1,...,3000$ as exact measurements (no noise added) of the state vectors, and create MCMC samples of the likelihood given by the expression (3). The comparison presented in Fig. 3. As expected, the filtering method is more accurate with this amount of data (we use every Euler-Maryama integration step as data for filtering), but the results by CIL are comparable.

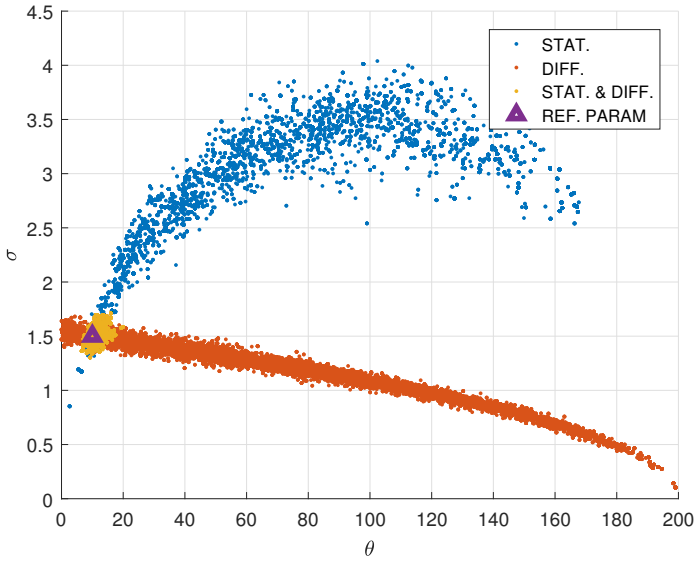

Fig. 2: The use of both state and difference information leads to a posterior (yellow) that is located around the intersection of the posterior generated by the state information only (blue) and the one generated using the difference only (orange).

Remarks. In the above examples we have used the known value of $\boldsymbol{\theta}_0$ as the starting point for the MCMC sampling. However, as the likelihood is created by the data signal, we can equally well use it as the cost function to estimate $\boldsymbol{\theta}_0$ first. We omit here the details of this step.

Note that there is a difference in computational times of the two methods, in this particular case they are approximately 20min for CIL and around 6min for KF. The difference is basically due to the additional computation of the distances needed for CIL.

Note that using a larger time step between data points would decrease the accuracy of the KF estimate. However, it does not impact the CIL estimate, as it is based on independent samples X_i in random order, not on predicting X_{i+1} by X_i.

Finally, we note that the use of the present modification, including the system 'dynamics' by signal differences, is not limited to the OU example. Rather, it can be used generally to improve the model parameter identification of both SDE and deterministic chaotic systems. However, a more detailed discussion is outside the scope of this work.

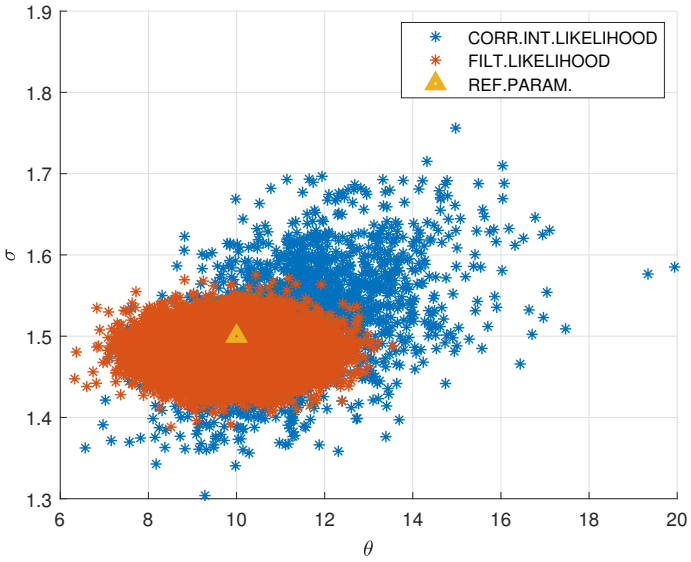

Fig. 3: Illustration of the results obtained by comparing CIL with the Filter likelihood method in parameter estimation for a zero mean Ornstein-Uhlenbeck.

3.2 Stochastic Chaos

Here we study the CIL approach for chaotic dynamics, extended with stochastic perturbations. Now the stochasticity is no more of the additive form (1) but is contained in the model equations in a nonlinear way. The specific forms of of the perturbations discussed here come from meteorology. In the so called Ensemble Prediction Systems (EPS) an ensemble of weather predictions, with carefully perturbed initial values, is launched together with the main prediction. The motive is to create probabilistic estimates for the uncertainty of the prediction. However, it is difficult to create a spread of the ensemble predictions that would match the observed uncertainty; the spread of the model simulations tends to bee too narrow. To increase the spread the so called stochastic physics is employed: the right hand side of the model differential equation is multiplied by a random factor (close to one) at every integration step. More recently, so called stochastic parametrization is used in addition: certain model parameters are randomized likewise at every integration step of the system. For more details of these methods see [9].

As a case study for the parameter estimation with stochastic physic and stochastic parametrization a classical chaotic attractor, the Rossler system, is chosen. We give the Rossler system in the form where the stochastic physics is introduced by the multiplicative factors $1+c_k\varepsilon$, and the model parameters α,β,γ are likewise re-

placed by perturbed terms $\alpha + c_k\varepsilon$, etc., $k = 1:6$, $\varepsilon \sim N(0,1)$. The system reads as

$$\begin{cases} \dot{X} = (1+c_1\varepsilon_1)(-Y-Z) \\ \dot{Y} = (1+c_2\varepsilon_2)(X+(\alpha+c_3\varepsilon_3)Y) \\ \dot{Z} = (1+c_4\varepsilon_4)((\beta+c_5\varepsilon_5)+Z(X-(\gamma+c_6\varepsilon_6))) \end{cases} \tag{8}$$

with 'true' parameters $\alpha = \beta = 0.2$ and $\gamma = 5.7$. The magnitudes c_k were chosen so that the maximum relative error would not exceed 40% in any of the cases.

Figure 4 shows the time evolutions of one of the components, the values of X for different combinations of added stochasticity. Each plot consists of 80 runs with slightly perturbed initial values. We see that the interval of predictable behavior shrinks to almost one half of that of deterministic chaos when both types of perturbations are added.

The task of parameter estimation is now to try to find the distribution of the mean value of each of the perturbed parameters. The construction of the likelihood is performed via the standard procedure: from a long enough data signal (here, produced by simulating (8)) we sample subsets to calculate the distances, and repeat this for a number of times to be able to empirically determine the statistics of the feature vectors. Again, the Gaussianity of the statistics can be verified. Both a maximum likelihood parameter estimate, and the subsequent MCMC sampling for the posterior can then be performed.

For the examples we create the data by simulating (8) over a total time interval $[0, 120000]$ and select data points at frequency shown in Fig. 4 with the green circles. To get one feature vector y we select two disjoint sets of 2000 consecutive data points. To create the statistics for y we repeat this procedure for around 1800 times. The number of radius values used was 10.

The results of the runs for different setting of the perturbations are given in Fig. 5. We can conclude that the approach performs as expected: the more stochasticity in the model, the wider are the parameter posteriors. However, in all cases we get bounded posteriors, and the algorithm performs without any technical issues.

4 Conclusions

In this work we have applied the recently developed Correlation Integral Likelihood method to estimate parameters of stochastic differential equation systems. Certain modifications are needed to get satisfactory results, comparable to those achieved by standard filter likelihood methods for basic SDE systems. But the main focus is on situations where the standard methods are not available, such as the stochastic physics and parametrizations employed in meteorology for uncertainty quantification. Several extensions of the approach are left for future work.

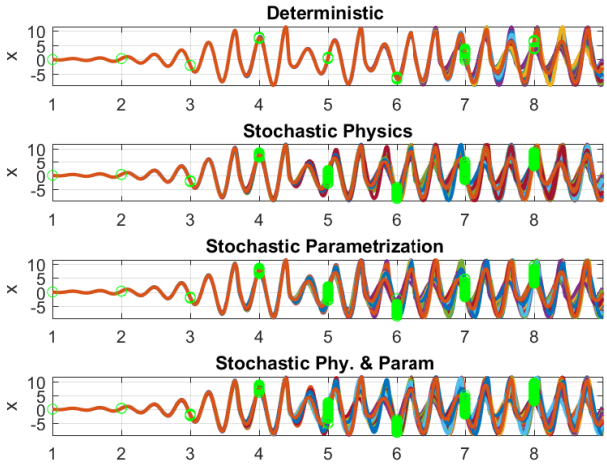

Fig. 4: The X component of the Rossler model with four different options for stochasticity.

Acknowledgements This work was supported by the Centre of Excellence of Inverse Problems, Academy of Finland.

References

1. Borovkova, S., Burton, R., Dehling, H.: Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. Transactions of the American Mathematical Society **353**(11), 4261–4318 (2001). DOI 10.1090/S0002-9947-01-02819-7. URL https://doi.org/10.1090/S0002-9947-01-02819-7
2. Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods. Oxford University Press (2012)
3. Haario, H., Kalachev, L., Hakkarainen, J.: Generalized correlation integral vectors: A distance concept for chaotic dynamical systems. Chaos: An Interdisciplinary Journal of Nonlinear Science **25**(6), 063,102 (2015). DOI 10.1063/1.4921939. URL http://dx.doi.org/10.1063/1.4921939
4. Haario, H., Laine, M., Mira, A., Saksman, E.: DRAM: Efficient adaptive MCMC. Statistics and Computing **16**(4), 339–354 (2006). DOI 10.1007/s11222-006-9438-0. URL https://doi.org/10.1007/s11222-006-9438-0
5. Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., Järvinen, H.: On closure parameter estimation in chaotic systems. Nonlinear Processes in Geophysics **19**(1), 127–143 (2012). DOI 10.5194/npg-19-127-2012. URL http://dx.doi.org/10.5194/npg-19-127-2012
6. Hakkarainen, J., Solonen, A., Ilin, A., Susiluoto, J., Laine, M., Haario, H., Järvinen, H.: A dilemma of the uniqueness of weather and climate model closure parame-

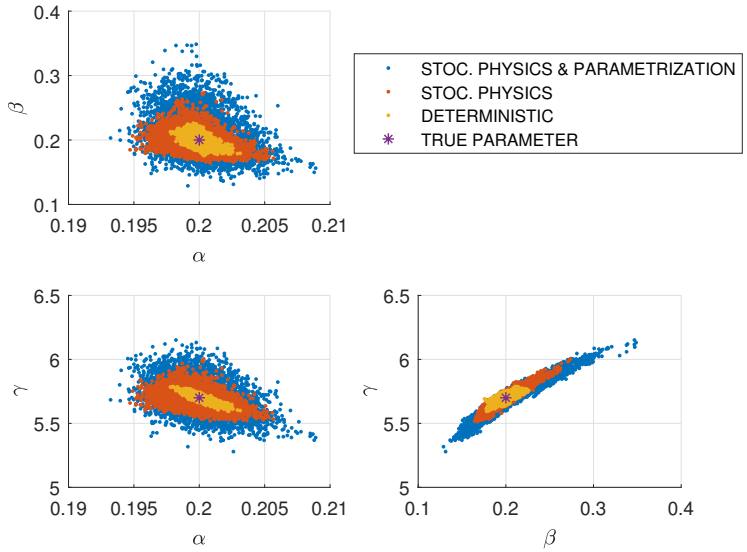

Fig. 5: Parameter posteriors for three different options of stochasticity for the Rossler model

ters. Tellus A: Dynamic Meteorology and Oceanography **65**(1), 20,147 (2013). DOI 10.3402/tellusa.v65i0.20147. URL http://dx.doi.org/10.3402/tellusa.v65i0.20147

7. Laine, M., Latva-Pukkila, N., Kyrölä, E.: Analysing time-varying trends in stratospheric ozone time series using the state space approach. Atmospheric Chemistry and Physics **14**(18), 9707–9725 (2014). DOI 10.5194/acp-14-9707-2014. URL https://www.atmos-chem-phys.net/14/9707/2014/
8. Mbalawata, I.S., Särkkä, S., Haario, H.: Parameter estimation in stochastic differential equations with markov chain monte carlo and non-linear kalman filtering. Computational Statistics **28**(3), 1195–1223 (2013). DOI 10.1007/s00180-012-0352-y. URL https://doi.org/10.1007/s00180-012-0352-y
9. Ollinaho, P., Lock, S.J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R.M., Haiden, T., Hogan, R.J., Sandu, I.: Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ecmwf ensemble. Quarterly Journal of the Royal Meteorological Society **143**(702), 408–422 (2017). DOI 10.1002/qj.2931. URL http://dx.doi.org/10.1002/qj.2931
10. Rougier, J.: 'Intractable and unsolved': some thoughts on statistical data assimilation with uncertain static parameters. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **371**(1991) (2013). DOI 10.1098/rsta.2012.0297. URL http://rsta.royalsocietypublishing.org/content/371/1991/20120297
11. Särkkä, S.: Bayesian Filtering and Smoothing. Cambridge University Press (2013)
12. Solonen, A., Järvinen, H.: An approach for tuning ensemble prediction systems. Tellus A: Dynamic Meteorology and Oceanography **65**(1), 20,594 (2013). DOI 10.3402/tellusa.v65i0.20594. URL http://dx.doi.org/10.3402/tellusa.v65i0.20594