Time-series analysis and predictability estimates by empirical SDE modelling

Naoto Nakano$^{1,2}$, Masaru Inatsu$^2$, Seiichiro Kusuoka$^3$ and Yoshitaka Saiki$^4$

$^1$ JST PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
$^2$ Faculty of Science, Hokkaido University, Kita10 Nishi8, Kita-ku, Sapporo 060-0810, Japan
$^3$ Grad. Sch. of Natural Science and Technology, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama, 700-8530, Japan
$^4$ Grad. Sch. of Commerce and Management, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601, Japan

E-mail: n_nakano@math.sci.hokudai.ac.jp

Abstract

The empirical evaluating formulae of the stochastic differential equation (SDE) are considered in this study, in order to reconstruct the information of the dynamics by the use of time-series itself. In this framework, we regard the time-series as a realisation of a solution of the SDE. By using the solution-coefficient relation studied in the probability theory, we can evaluate the drift vector and the diffusion matrix of the SDE, which are calculated by the first and the second order variations per unit time. The drift vector represents the deterministic behaviour of the motion and the diffusion matrix play the role of stochastic effect. Utilising the statistically evaluated SDE, one can obtain the predictability estimates in terms of the ensemble variance. The results indicate that the deterministic components are significant for the short-term prediction. Moreover, we need the stochastic components to obtain better predictability estimates by the use of the numerical simulation of the SDE.

1 Introduction

The purpose of this study is to develop and validate a method of data analysis based on the framework of stochastic differential equations (SDEs) for projected time-series. The SDE framework is an empirical modelling procedure to reconstruct the dynamics from the observed data [2, 9, 1, 8]. In this framework, the time-series is regarded as discrete samples of a realisation of a solution of the SDE. The existing studies mentioned above basically used the numerical technique of extracting the coefficients of the SDE by simulated data sets, which was introduced by Sigert et al. [7]. The estimating formulae have been derived by Sigert et al. [7] using the Kramers–Moyal expansion for the probability density function of the corresponding Fokker–Planck equation [6]. The knowledge of the scale separation is not needed in advance to use this stochastic modelling procedure, and the deterministic part of motion of the time-series is characterised by the drift vector and high-frequency fluctuation can be treated as a noise in the SDE.

This methodology works well for extracting dynamics from time-series obtained by the dimensionality reduction procedure. For such a time-series, the diffusion components compensate for the loss of information of the data due to the reduction. For example, Inatsu et al. [3, 4] considered the ensemble predictability by the use of an empirically estimated SDE from winter-time atmospheric datasets projected onto a phase space spanned by the leading loading vectors of the principal component analysis (PCA). They found a strong correlation between the spreads of ensemble members of solutions of the SDE and the operational ensemble spread of Japan Meteorological Agency. These studies implied that the dynamical reconstruction by the use of the SDE framework performed well and the model could reproduce the orbital information.

In this study, unlike the datasets having quite a large degree of freedom used in the previous studies by Inatsu et al. [3, 4], we shall focus on a low-dimensional dynamical system, a planar projection of the Lorenz system, to distinguish the effect between the drift and diffusion caused by the projection through the stochastic modelling procedure. As a technical issue, utilising the mathematical theorem on probability theory, we improve the empirically evaluating formulae for the SDE coefficients and construct discretely a numerical SDE which emulates dynamics of the original time-series on the projected 2-dimensional plane. To assess the availability of the obtained SDE, we compare the result of ensemble variance of solutions as the predictability estimate generated by the numerical SDE to that of the trajectories of the projected orbit of the time-series.
Fig. 1: A trajectory of a numerical solution of the Lorenz system (1) with $\sigma = 10$, $r = 28$ and $b = 8/3$. The trajectory is calculated by the 4th-order Runge-Kutta method with time step $1.0 \times 10^{-3}$ under the initial condition $(x(0), y(0), z(0)) = (1, 1, 1)$ and is plotted for $100 < t \leq 200$.

Fig. 2: A 2-dimensionally projected orbit of the numerical solution of (1) shown in Figure 1, generated by the 2-dimensional projection (2), namely the projection $(x, y, z) \mapsto (u, v)$ onto a plane spanned by the 2-leading loading vectors $q_1$ and $q_2$ constructed by the PCA. The horizontal and the vertical axes represent $u$ and $v$, respectively.

2 Data

The dataset used here is a planar-projected data of a numerical solution of the Lorenz system

$$
\begin{align*}
    x' &= p(y - x), \\
    y' &= x(r - z) - y, \\
    z' &= xy - bz,
\end{align*}
$$

(1)

where $p$, $r$ and $b$ are system parameters and taken as the classical ones, i.e., $p = 10$, $r = 28$, $b = 8/3$ [5]. We integrate (1) numerically by using the 4th order Runge-Kutta method with the step size $\Delta t := 1.0 \times 10^{-3}$ and the initial condition $(x(0), y(0), z(0)) = (1, 1, 1)$. We cut off the first $10^5$ steps ($t \leq t_0 := 100$) to remove transient states, and the sequence is sampled for $100 < t \leq 10100$ for a total of $10^7$ data: \(\{(x_{tn}, y_{tn}, z_{tn})\}_{t_n \in T}\), \(T = \{t_n = t_0 + n\Delta t | n = 1, 2, \ldots, 10^7\}\). The trajectory of the numerical solution exhibits the butterfly-like chaotic attractor (Fig. 1).

Then, the planar time-series \(\{X_{tn} = (u_{tn}, v_{tn})\}_{t_n \in T}\) is calculated by the use of 2-dimensional projection $(x, y, z) \mapsto (u, v)$ onto a plane spanned by the 2-leading loading vector constructed by PCA, namely

$$
\left\{ 
    \begin{align*}
    u_{tn} &= (x_{tn} - \langle x \rangle, y_{tn} - \langle y \rangle, z_{tn} - \langle z \rangle) \cdot q_1, \\
    v_{tn} &= (x_{tn} - \langle x \rangle, y_{tn} - \langle y \rangle, z_{tn} - \langle z \rangle) \cdot q_2,
    \end{align*}
\right.
$$

(2)

where $q_1 \approx (0.65390, 0.75656, -0.00630)$ and $q_2 \approx (0.00349, 0.00532, 0.99998)$ are the first and the second loading vectors of PCA, respectively, and $\langle x \rangle$, $\langle y \rangle$ and $\langle z \rangle$ denote the mean values of $x_{tn}$, $y_{tn}$ and $z_{tn}$, respectively, for example $(x) = \sum_{t_n \in T} x_{tn}/|T|$.

From Fig. 2, we can easily identify two projected unstable fixed points near $(12.07, 3.52)$ and $(-11.87, 3.37)$, since the trajectory steers around them. We can also find another projected unstable fixed point near $(0.27, -23.55)$ that the orbits recede from it. We should also remark that the trajectories go regularly in the outer side of the projected attractor, whereas they hold self-intersections around the centre region of the attractor due to the projection. Hence, it can not be considered as a trajectory of a solution of an autonomous system on the plane, and the orbits in the self-intersection region can be regarded as “non-deterministic” in the sense of a 2-dimensional autonomous system.

3 Method

In the time-series analysis based on the SDE framework, one regards time-series under consideration as a orbit of one realization of a solution of an SDE

$$
dX_t = A(X_t)dt + S(X_t)dW_t.
$$

(3)

Here, $X_t = (X^1_t, X^2_t)$ is the solution of the SDE in the 2-dimensional phase space, $A(x) = (A^1(x), A^2(x))$ the
The estimated drift vector field (red arrows) obtained by the use of formula (6) with the estimate interval (top) $\tau = 0.01$, (centre) $0.15$ and (bottom) $0.30$, displayed over the projected orbit (green) shown in Fig. 2. The estimated drift vectors are plotted for the bins centred at the even lattice holding not less than 100 data. The arrow size is reduced by $1/50$ from the original size.

![Fig. 4](image1)

The probability measure induced by a solution of the SDE (3) with the initial condition $x$ [6]. For discrete time-series $\{x_n\}_{n \in \mathbb{T}}$, utilising (4) and (5), we develop inference evaluation formulae for the coefficients in the following form. Dividing the phase space into finite-size cells $\{ \tilde{x}_m \}_{m \geq 1}$, one can estimate

$$A^i(\tilde{x}_m; \tau) = \langle X^i_{t+\tau} - X^i_t \rangle_{\tilde{x}_m}/\tau, \quad (6)$$

$$B^{ij}(\tilde{x}_m; \tau) = \{ \langle (X^i_{t+\tau} - X^i_t)(X^j_{t+\tau} - X^j_t) \rangle_{\tilde{x}_m} \} / 2\tau(7)$$

in each cell. Here, $\langle \cdot \rangle_{\tilde{x}_m}$ denotes the ensemble mean over the data visiting the cell $\tilde{x}_m$. Note that the estimated drift vector $A$ and diffusion matrix $B$ are calculated as the ensemble mean and covariance of variation of the state vector per estimating time interval $\tau$, respectively.

Hereafter, we adopt the bins dividing the phase space as $1 \times 1$-square centred at the 2-dimensional integer lattice $x_m = \{(u, v) \mid -1/2 \leq u - m_1 < 1/2, -1/2 \leq v - m_2 \leq -1/2 \}

![Fig. 5](image2)
4 Results

4.1 The drift vector

We plot the drift vector fields that are estimated by using (6) in Fig. 4. Each vector shown in Fig. 4(top) with estimating time interval $\tau = 0.01$ is given as a tangent vector of the planarly projected orbit except the neighbourhood of the vertical axis $\{ u = 0 \}$ where the orbit holds self-intersections. Hence, in the area without self-intersection, the drift vector can represent the averaged motion of orbit behaviour for sufficiently small estimating time scales. On the other hand, the vectors for the cases $\tau = 0.15$ and $0.30$ does not show the same behaviour as the tangent vectors, due to the long estimating time interval. In these case, the drift vectors are regarded as the transition maps in the corresponding time scale.

In the area where the self-intersection appears, the obtained drift vectors for all the cases in Fig. 4 are parallel to $v$-axis negative direction, since the self-intersections are composed of almost fifty-fifty of the orbits which go towards $u$-axis positive or negative direction, and they cancel each other out. This means that, near $\{ u = 0 \}$, the drift vector does not contain deterministic information about the motion along with $u$-variable and only holds information of the deterministic dynamics that every orbit proceeds to the negative direction in $v$ near $\{ u = 0 \}$. Of course, it is not complete that the estimated drift vector can represent the dynamics of the time-series by itself, however, the diffusion matrix may supplement the defect in the sense.
Fig. 8: Comparisons between (upper left) the projected orbits and (upper right, lower left and lower right) the numerical solution of the estimated SDE. (upper left): 20 randomly chosen trajectories, passing the 1 x 1-size bin centred at \((u, v) = (10, -15)\), after the elapse of time 1.5 are shown. (upper right, lower left and lower right): The orbits generated by the numerical simulation of the SDE constructed with \(\tau = (upper right) 0.01\), (lower left) 0.15 and (lower right) 0.30 are plotted. The initial conditions are given uniformly in the bin centred at \((u, v) = (10, -15)\). The cross and circle are marked at the start and end points, respectively, in the both projected trajectories and the numerical simulation curves.

of dynamics as we shall see below. Moreover, the fixed points can be easily recognised by examining the magnitude of the drift vector (Fig. 5). This information is also useful to understand the behaviour of the orbit and to presume the background dynamics which the time-series data follow. The drift vector field for \(\tau = 0.01\) shows the repellers at \((u, v)\) near \((12.07, 3.52)\) and \((-11.87, 3.37)\). However, the estimated drift vectors for \(\tau = 0.15\) and 0.30 (Figs. 4(centre) and (bottom)) indicate inconsistent results with characteristics of the original Lorenz system. It is easy to see that the estimated vector field looks like it has attractors near those projected fixed points, but those two points are projection of the two unstable fixed points of the original 3-dimensional Lorenz system. This caused by choosing large \(\tau\) to calculate \(\tilde{A}\), since the drift vector is obtained as the ensemble mean of moving distance per unit of time, \(i.e.\ \tau\), of the projected time-series. Accordingly, the drift vector obtained by \((6)\) with small \(\tau\) can be regarded as the deterministic motion of the time-series.

4.2 The diffusion matrix

We show the norm of the diffusion matrix \(\text{tr} \tilde{B}\) obtained by formula \((7)\) in Fig. 6, which the norm represents the magnitude of the variance of the orbits. In these cases, \(\text{tr} \tilde{B}\) is large in the self-intersection region, and otherwise the value does not vary significantly. It is therefore useful to see \(\text{tr} \tilde{B}\) so that we can extract self-intersections. Hence, the equilibrium-like points detected in Fig. 5 is not stagnation points occurred by the projection but the deterministic fixed points.

Moreover, in Fig. 6(bottom), we can also find the effect of the unstable fixed point \((0.27, -23.55)\), which is the origin in the original dynamical system. This implies that the trajectories of the dataset split into two groups going to the left and the right parts of the projected attractor. However, \(\text{tr} \tilde{B}\) tends to be large where the magnitude of the drift vector is large, for example, in the brim of the projected attractor in Fig. 6. This means that high moving velocity of the orbit can also be a part of the cause of orbit uncertainty. Hence, it is also important to see the magnitude of the diffusion normalised by the square of the norm of the drift vector in order to distinguish the stochastic effect from the deterministic effect. The planar plots of the common logarithm of \(\text{tr} \tilde{B}/|\tilde{A}|^2\) are showed in Fig. 7. This value is the ratio of the diffusion effect of the variation to the deterministic component. The figures show that it is
small except the self-intersection region and three projected unstable equilibria, and in such a regular region, the results are consistent with the behaviour of orbit trajectories that appear almost parallel. The origin as an unstable fixed point is also observed in Fig. 7 and \( \text{tr} \bar{\mathbf{A}}/|\mathbf{A}|^2 \) near \((0.27, -23.55)\) in the projected plane is clearly larger than that near the two projected fixed points \((12.07, 3.52)\) and \((-11.87, 3.37)\), which are projected points of repellers in the original Lorenz system. It is obvious that the ratio takes large in the area where it is near the saddle point or the self-intersection, thus this ratio can capture the magnitude of the orbit spreading, namely the predictability of the orbital dynamics. Consequently, examining the values of \( \bar{\mathbf{A}}, \text{tr} \bar{\mathbf{A}}/|\mathbf{A}|^2 \), we can describe the dynamics of the observed time-series as the characteristics of the SDE in each cell in the phase space.

4.3 Numerical SDE simulation

Using the estimated drift vector and diffusion matrix with some estimating interval \( \tau \) obtained above, we can perform the numerical SDE simulation and verify the calculation whether the numerical solution can trace the orbit observed preliminarily. We use the Euler–Maruyama method for calculating numerical solutions of the SDE. Here, we adopt the time step size \( \Delta t \) for the numerical calculation same as \( \tau \). In fact, the SDE coefficients obtained above are only obtained in the coordinate grid, thus we use the bilinear interpolation to obtain the value at the any points in the phase space.

We take 100 ensemble members as initial conditions in a \( 1 \times 1 \)-size cell and integrate solutions until \( t = 1.5 \) in each simulation, and the results are shown in Figs. 8(upper right), 8(lower left) and 8(lower right). The results for \( \tau = 0.01 \) do not transit from the right to the left part of the attractor. The reason is thought that the magnitude of estimated stochastic components is not sufficiently large to cross the intersection. Thus, we can say that the coefficients estimation for the time-scale of \( \tau = 0.01 \) captures only the deterministic effect.

On the other hand, interestingly, the trajectories of the numerical solution for large \( \tau \) can cross from the region of \( u \)-axis positive to that of negative. This feature cannot be obtained by computation of a 2-dimensional autonomous system, and it is obtained because of the appropriate time scale \( \tau \) and the diffusion effect.

Though the pathwise coincidence between simulation and observation depends on the initial conditions or the regions where the trajectories pass, it does not mean that the estimated SDE model is irrelevant to the original dynamics. It is important to examine variance of calculated trajectories to estimate the degree of orbital instability, \( \text{viz.} \) orbital predictability. We illustrate distribution of the variance of the simulation and observed trajectories in Fig. 9. The colour map at each point in the phase space indicates the logarithm of the variance of the trajectories that start at the same point.

Until the elapse of 0.6, the distribution of the ensemble variance calculated by the SDE simulation with the coefficients estimated with \( \tau = 0.01 \) shows a quantitative agreement with that calculated by the observed trajectories. Since the evaluated SDE for \( \tau = 0.01 \) represents only the deterministic behaviour, this may imply that the orbit behaviour are not influenced by the stochastic effect for such a time period.

On the other hand, the situation is changed for the ensemble variance for the solution of the long-term simulations the result of the SDE estimated by \( \tau = 0.01 \) does not show good agreement. For example, after the elapse of \( t = 0.9 \), while the magnitude of the ensemble variance of the projected time-series is small in the outer area of the projected attractor, the magnitude of the ensemble variance obtained by the simulation with \( \tau = 0.01 \) take high values. Even though the range of value is not same as that of the projected orbits, the result of the case of \( \tau = 0.15 \) shows a streak-like pattern of the ensemble variance and it consistent with that of the projected orbits qualitatively. This implies that one needs stochastic terms with sufficiently large magnitude in order to simulate the dynamics for a long term.

5 Conclusion

We constructed an SDE whose drift vector and diffusion matrix were empirically evaluated by the use of the solution-coefficient formulae arising from the probability theory. The drift vector and the diffusion matrix were estimated as the ensemble mean of the first and the second order variations of the state vector per unit time, respectively.

Then we applied this data-driven stochastic modelling method to the 2-dimensional time-series generated by the planar projection of a numerical solution of the Lorenz system with the classical parameters. We found that the statistically estimated SDE contained the useful dynamical information of the projected time-series in the phase space. Since the obtained drift was parallel to the orbit outside of the region which the trajectory of the data had self-intersections, it can be characterised as the deterministic behaviour of the time-series. The diffusion matrix implied that stochastic characteristics of the data compensated the self-intersection information which the drift does not contain. Moreover, the ratio of the magnitude of the diffusion to the square of the drift vector can detect not only the orbit self-intersection region but also the three unstable fixed points of the Lorenz system projected into the 2-dimensional subspace. Hence we can reconstruct the dynamical information by the use of the drift vector and the diffusion matrix which can be statistically estimated by the time-series.

Utilising the reconstructed SDE, we performed the numerical simulation to consider the orbital traceability in terms of the ensemble variance. For a short-term simulation \( t \leq 0.6 \), the ensemble variance of numerical
Fig. 9: Comparisons between the ensemble variance for the projected time-series (left) and the simulated data (centre and right). Results for the elapse time (a, b, c) 0.3, (d, e, f) 0.6, (g, h, i) 0.9, (j, k, l) 1.2 and (m, n, o) 1.5 are shown. The utilised SDE is the same as that with (center) $\tau = 0.01$ and (right) $\tau = 0.15$ in Fig. 8.
solutions of the SDE constructed with $\tau = 0.01$ showed a good agreement with that calculated by the projected orbits. On the other hand, the situation was changed for the result obtained by the estimating interval $\tau = 0.15$. The result for the case of $\tau = 0.15$ showed better estimates of the ensemble variance. Therefore, these results implied that the stochastic components plays a important role in order to emulate the behaviour of the orbits and this SDE modelling scheme can provide good performance of the predictability of the projected dynamics even for the long-term prediction.

6 Acknowledgments

This study was supported by PRESTO of Japan Science and Technology Agency (JST), and also partly supported by Research Program on Climate Change Adaptation, and by Grants-in-Aid for Scientific Research 25610028 and 26310201 of the Ministry of Education, Culture, Sports, Science, and Technology of Japan. This research partly used computational resources under Collaborative Research Program for Young Scientists provided by Academic Center for Computing and Media Studies, Kyoto University and the MEXT Joint Usage / Research Center "Center for Mathematical Modeling and Applications", Meiji University, Meiji Institute for Advanced Study of Mathematical Sciences (MIMS).

References


