Markov Chain Monte Carlo Estimation of Nonlinear Dynamics from Time Series

Christopher L. Bremer and Daniel T. Kaplan*
Department of Mathematics and Computer Science
Macalester College
St. Paul, MN 55105

May 17, 2001

Abstract: Much of nonlinear time series analysis is concerned with inferring unmeasured quantities — e.g., system parameters, the shape of attractors in state space — from a noisy measured time series. From a Bayesian perspective, the time series is a vector sample picked at random from a probability density. The density reflects the system dynamics and our subjective uncertainty about system parameters, the measurement function, dynamical noise and measurement noise. The conditional probability density of the system parameters given the measured data is the basis of a Bayesian estimate of the system parameters. Using illustrative chaotic systems with large-amplitude dynamical and measurement noise, we show here that it is feasible to use the Markov chain Monte Carlo (MCMC) technique to generate the Bayesian conditional probabilities. The resulting parameter estimates are markedly superior to those based on conventional least-squares methods: the MCMC-based estimates are unbiased and allow estimates of dynamical parameters on unmeasured components of the state vector. In addition, the MCMC method enables de-noised attractors to be reconstructed, not just in an embedding based on lags of measured variables but in the state space that includes unmeasured components of the dynamics' state vector. The general purpose MCMC technique effectively combines techniques of nonlinear noise reduction and nonlinear parameter estimation.

*Corresponding author. kaplan@macalester.edu

Many techniques for time series analysis are based on the idea of representing the time series $z_1, \ldots, z_N$ as a probability density. For example, the correlation and information dimensions can be written in terms of an integral over a probability density $p(z_t, z_{t-h}, \ldots, z_{t-(m-1)h})$ in a lag embedding space [1] of dimension m and lag h ; the mutual information [2] is constructed from integrals over the same density typically with $m = 2$; nonlinear noise reduction [3] often involves modeling the probability density $p(z_t, z_{t-h}, \ldots, z_{t-(m-1)h})$ locally as having support on a linear manifold of dimension $k < m$; nonlinear prediction [4] implicitly or explicitly examines a conditional probability $p(z_{t+f}|z_t, z_{t-h}, \ldots, z_{t-(m-1)h})$.

All of the above methods have in common that the dimension m of the space in which the probability density is considered is $m \ll N$, where N is the number of data points in the time series. For example, in computing correlation dimensions the embedding dimension is typically in the range $m = 2$ to 20 even for time series of length $N \gg 1000$. [5]

In this report, we explore some advantages of considering probability densities in spaces with dimension $> N$. We take a Bayesian perspective where the probability density in the high-dimensional space is generated by two factors: the measured time series and a model for the system dynamics.

We consider models in the form of discrete-time, noisy dynamical systems

$$\vec{x}_{t+1} = F(\vec{x}_t, \vec{c}) + \nu_t \quad (1)$$

$$z_t = G(\vec{x}_t) + \mu_t \quad (2)$$

where the measurement z_t at time t is related to the vector state $\vec{x}_t$ of the system by the measurement function $G(\cdot)$. The vector $\vec{c}$ are the parameters of the model. The vector-valued function F sets the deterministic part of the dynamics. ν_t is additive dynamical noise and μ_t is measurement noise which we take to be gaussian white noise of perhaps unequal variances.

Such models usually include unknown parameters $\vec{c}$ as well as the unknown, random values of ν_t and μ_t at each time t. The modeller has a notion of the probable range of the parameters and the distribution of ν_t and μ_t (perhaps summarized by their variances σ_ν^2 and σ_μ^2 whose exact values are usually unknown and therefore might be described in terms of a probability distribution) — this knowledge is termed a "prior" in Bayesian nomenclature which we can denote $p(\vec{c}, \sigma_\nu, \sigma_\mu)$.

Given the prior, forward simulation of the model is easy: use a random number generator to pick specific parameter values from $p(\vec{c}, \sigma_\nu, \sigma_\mu)$, construct a fixed sequence ν_t and μ_t, and then iterate equations (1). Each such simulation can be seen as a Monte Carlo generation of one point out of the modeller's prior probability distribution for the quantities in the model. For example by looking at the model state at a specific time, say $t = 9$, from the simulations, the distribution of the system state $\vec{x}_9$ implied by the modeller's prior can be inferred.

Such forward simulations are often used informally by modellers to see whether a model can generate outputs that look like the measured data. When the simulated data look like the measured data, the parameters are accepted as plausible candidates for the data. Systematically varying the model parameters in an attempt to match the model output to the measured data is a way of fitting the model to the data.

The informal approach of simulate $\rightarrow$ compare-with-data is a way to generate conditional probabilities. If the modeller rejects all samples from the prior probability distribution leading to a simulation output that does not resemble the data, the remaining samples will be what is called a "posterior" in the Bayesian terminology: the probability distibution of the model parameters conditional on the measured data, that is $p(\vec{c}, \sigma_\nu, \sigma_\mu | z_1, \ldots, z_N)$. By looking at a variable of interest from the non-rejected simulations, say, $\vec{x}_9$, one gets the probability distribution of that variable conditioned on the data. This is a way of inferring a model variable from the measured time series.

This logic for computing the posterior relies on the possibility of a simulation's output closely matching the time series. Unfortunately, due to the random μ_t and ν_t, a forward simulation approach is extremely unlikely to produce values that so exactly match the measured time series that any confidence can be had in the inferred values. Instead, a "backwards" simulation can be used where the output of the simulation is fixed to the observed data and the variables chosen from the modeller's prior that lead to this output. The Markov chain Monte Carlo (MCMC) technique provides a means for implementing this backwards simulation: matching the output to the data while selecting variables in a manner that corresponds to the modeller's prior distribution.

The MCMC approach was initially demonstrated in nonlinear dynmics time series analysis by Davies [6] for the purposes of model-based noise reduction, and by Heald and Stark [7] for estimating the relative size of measurement and dynamical noise. Very recently published work by Meyer and Christensen [8], building on a maximum likelihood approach from McSharry and Smith [9], has shown that the technique can be used to estimate parameters of chaotic models. Here, we show that it is a practical means for estimating parameters of models as well as noise levels, and for inferring unmeasured state variables from noisy measured data.

To fix concepts, we examine time series from the noisy Henon system

$$\begin{aligned} x_{t+1} &= c_2 - c_1 x_t^2 + c_3 x_{t-1} + \nu_t \\ z_t &= x_t + \mu_t \end{aligned} \tag{3}$$

with $c_1 = 1.3$, $c_2 = -1$ and $c_3 = 0.3$ and where the state vector $\vec{x}_t = (x_t, x_{t-1})$. Plots of successive values of $\vec{x}_t$ are shown in Figs. 1A1 and 1B1 for various amounts of dynamical noise. From time series with measurement noise such as shown in Fig. 1A2 and 1B2 we seek to compute estimates $\widehat{c_1}$, $\widehat{c_2}$, and $\widehat{c_3}$ of the dynamical parameters as well as the variances of the white noise processes ν_t and μ_t. We note that this system is somewhat special in that the time series measures all components of the state vector $\vec{x}_t$. Therefore in this special case we can write the system dynamics (Eq. 3) directly in terms of the measurement z_t, giving

$$z_{t+1} = \widehat{c_1} - \widehat{c_2} z_t^2 + \widehat{c_3} z_{t-1} + \omega_t \tag{4}$$

where ω_t is a noise term that reflects both the dynamical noise and measurement noise. A common approach to inference about dynamics from time series is to apply ordinary least squares linear regression of z_{t+1} against z_{t-1} and z_t^2 to find $\widehat{c_1}$, $\widehat{c_2}$, and $\widehat{c_3}$. With the least squares approach we can use the residuals

$$z_{t+1} - \widehat{c_1} - \widehat{c_2} z_t^2 + \widehat{c_3} z_{t-1}$$

to estimate the variance of ω_t.

The dynamical noise ν_t in Eq. 3 may affect the system's trajectory but it does not lead to bias in the estimates $\widehat{c_1}$, $\widehat{c_2}$, and $\widehat{c_3}$ when performing ordinary least squares regression. In contrast, measurement noise μ_t systematically biases the estimates (Fig. 2). This bias is explained by the fact that that ω_t in Eq. 4 is not gaussian white noise as can be seen by substituting $z_t - \mu_t$ for x_t in Eq. 3. (For example, ω_t includes a state-dependent term $c_2 z_t \mu_t$.)

Accurate estimation of system parameters requires appropriate incorporation of the measurement noise into the dynamical model. One way to do this is to estimate a measurement noise term $\widehat{\mu_t}$ for each point in the time series and transform the measured time series to an inferred, de-noised time series ξ_t using

$$\xi_t = z_t - \widehat{\mu}_t. \tag{5}$$

Kostelich and Yorke [10] and Davies [11] developed methods to compute $\widehat{\mu_t}$ from time series by minimization of $\sum \omega_t^2$ and $\sum \hat{\mu}_t$ from models such as Eq. 4 and 5. These minimization methods require some *a priori* estimate of the relative RMS amplitudes σ_ω and σ_μ of the ω_t and μ_t noise processes: by making $\sum \omega_t^2$ too small when minimizing one overcleans the time series. Consideration of this problem led Davies [6] and Heald [7] to develop Bayesian techniques for estimating $\sigma_\omega / \sigma_\mu$.

Bayesian Perspective

Rather than looking for an optimal single value of $\vec{c}$, we consider the conditional probability of the model parameters conditioned on the measured time series, $p(\vec{c}|z_1 \ldots z_N)$ which is termed the "posterior" probability since it gives the probability of the parameter vector $\vec{c}$ (containing, *inter alia* $\widehat{c_1}$, $\widehat{c_2}$, and $\widehat{c_3}$) *after* the measurement. Bayes' rule allows one to write the posterior probability as a product of a likelihood function and a prior probability

$$p(\vec{c}|z_1 \ldots z_N) \propto p(z_1 \ldots z_N|\vec{c})p(\vec{c})$$

The prior probability $p(\vec{c})$ can in principle incorporate any *a priori* knowledge of the parameters but in this work we take it to be a wide and uninformative distribution. The likelihood function $p(z_1 \ldots z_N|\vec{c})$ reflects the model of the dynamics of the system.

To make explicit the relationship between the model dynamics and the likelihood function, we incorporate an unmeasured initial condition $\vec{\xi_0}$ into the conditional probability

$$\begin{aligned} p(\vec{c}|z_1 \ldots z_N) &\propto \int_{R_{\vec{\xi}_0}} p(\vec{c}, \vec{\xi_0}|z_1 \ldots z_N) dR_{\vec{\xi}_0} \\ &\propto \int_{R_{\vec{\xi}_0}} p(z_1 \ldots z_N|\vec{c}, \vec{\xi_0}) p(\vec{c}, \vec{\xi_0}) dR_{\vec{\xi}_0} \end{aligned} \tag{6}$$

where $R_{\vec{\xi}_0}$ is the domain of $\vec{\xi_0}$.

Using the dynamical/measurement model with fixed parameters $\vec{c}$

$$\begin{aligned} \xi_{t+1} &= \widehat{c_1} - \widehat{c_3}\xi_t^2 + \widehat{c_3}\xi_{t-1} + \nu_t \\ z_t &= \xi_t + \widehat{\mu}_t \end{aligned} \tag{7}$$

guesses of $\vec{\xi_0}$ can be propagated to a description of the joint probabilities of ξ_t and z_t for all t. In order to do this we need a model of the probability distribution of ν_t and $\widehat{\mu}_t$ which we take to be gaussian white noise of unknown variances σ_ν^2 and σ_μ^2 respectively. Notationally, we will incorporate σ_ν^2 and σ_μ^2 into the parameter vector $\vec{c}$.

Continuing to expand the conditional probabilities in terms of the inferred, de-noised variables ξ_t we have

$$p(\vec{c}|z_1 \ldots z_N) \propto \tag{8}$$
$$\int_{R_{\vec{\xi}_0}} \int_{R_{\xi_1}} \ldots \int_{R_{\xi_N}} p(z_1 \ldots z_N, \xi_1 \ldots \xi_N | \vec{c}, \vec{\xi_0}) p(\vec{c}, \vec{\xi_0}) d_{R_{\vec{\xi}_0}} d_{R_{\xi_1}} \ldots d_{R_{\xi_N}}$$

The MCMC technique allows one to sample from a joint probability distribution such as in Eq. 8 by successively sampling from the conditional probability of each of the variables *holding constant* all the other variables during each sample.[12]

We implemented MCMC sampling of $p(\vec{c}|z_1 \ldots z_N)$ using the Bayesian Estimation using Gibbs Sampling software package (BUGS) [13], a standard engine for MCMC sampling, which automatically deduces the needed conditional distributions from statements that describe the relationships between sets of variables. (An example of the form of these statements is given in the Appendix). The dynamical model of Eq. 7 induces the conditional probability for each ξ_{t+1} given ξ_t, ξ_{t-1} and parameters $\vec{c}$. Expressed in terms of conditional probabilities, the essential dynamical relationships are these (using $N(a, b)$ to denote a gaussian distribution of mean a and variance b) :

$$\begin{array}{lll} \text{Measurement:} & \xi_t \sim N(x_t, \sigma_\mu^2) & \text{for } \; t=1 \;\text{ to }\; N \\ \text{Dynamics:} & \xi_{t+2} \sim N(c_1 - c_2\xi_{t+1}^2 + c_3\xi_t, \sigma_\nu^2) & \text{for } \; t=1 \;\text{ to }\; N-2. \end{array} \tag{9}$$

In addition to the conditional probability relationships expressed in above, one must provide a prior probability density for the parameters in $\vec{c}$ (including the noise variances) as well as the initial condition $\vec{\xi_0}$.

Estimation of Model Parameters using MCMC

To test the ability of the MCMC method to estimate model parameters in the face of measurement and dynamical noise, we generated time series of length $N = 1000$ from Eq. 3. For each time series, measurement and dynamical noise were generated as computer pseudo-random numbers with fixed variance σ_μ^2 and σ_ν^2 respectively. Different time series were generated with σ_ν ranging from 0 to 1.0, with the dynamical noise held at $\sigma_\mu = 0.04$.

Ordinary least squares regression was used to generate estimates of c_1, c_2, and c_3. The noise amplitude was estimated by the standard deviation of the residuals from the fit. This noise estimate was then translated in the standard way into an estimate of the variance in the estimated c_1, c_2, and c_3. These least squares estimates are plotted in Fig. 2.

For the MCMC estimation, we set the prior on $\vec{\xi_0}$ to be normal with mean and standard deviation set to that of the time series. The priors for σ_μ and σ_ν were set to be uniform with minimum 0 (the no-noise limit) and maximum equal to the standard deviation of the time series (the all-noise limit). The priors for c_1, c_2, and c_3 were set to be centered on the least squares estimates, with variance 50 times that indicated by the least squares estimates. The BUGS engine was used to generate samples of c_1, c_2, c_3, σ_μ, and σ_ν conditional on the time series z_t.

Results are presented in Fig. 2. Even for large amounts of measurement noise where linear regression is substantially biased, MCMC gives unbiased estimates of the dynamical parameters as well as accurate estimates of the variances σ_μ^2 and σ_ν^2 of the measurement and dynamical noise.

In contrast to the case for linear regression, error bars for the MCMC estimates reasonably reflect deviations from the true values. In addition, the estimates of individual inferred state values ξ_t appear to lie close to the system attractor. We draw attention to the estimation method's ability to distinguish between different amounts of dynamical noise (Figs. 1A1 vs. 1B1) when fitting identical forms of models to very noisy data. When additional first- and second-order polynomial terms are added to the model, the coefficients on these terms are statistically indistinguishable from zero.

Indirect Measurement of Dynamics

In typical experimental or field work one does not measure all components of the state vector. In this sense, the Henon map is an unrealistic model of experimental dynamics. We therefore examine two cases where the measurement captures only one component of a two-dimensional state vector: the Ikeda map and the Tinkerbell Map [14].

The Ikeda map (with a complex-valued state vector $\vec{x}_t$)

$$\begin{aligned} \vec{x}_{t+1} &= c_1 + c_2 \vec{x}_t \exp i \left(c_3 - c_4/(1 - |\vec{x}_t|^2)\right) \\ z_t &= \mathrm{Re}(\vec{x}_t) + \mu_t \end{aligned}$$

with parameters $c_1 = 1$, $c_2 = 0.7$, $c_3 = 0.4$ and $c_4 = 6.0$. The Tinkerbell map operates on state (x_t, y_t)

$$\begin{aligned} x_{t+1} &= x_t^2 - y_t^2 + c_1 x_t + c_2 y_t \\ y_{t+1} &= 2 x_t y_t + c_3 x_t + c_4 y_t \\ z_t &= x_t + \mu_t \end{aligned}$$

with parameters $c_1 = -0.3$, $c_2 = -0.6$, $c_3 = 2.0$, $c_4 = -0.27$. In addition to the inferred variable that reflects the de-noised measurement, we add to the estimation procedure another inferred variable representing the 2nd component of the state vector. The inferred variable is therefore represented as a vector $\vec{\xi}_t$ which, for the purposes of stating marginal probabilities, is tied to the time series via the measurement function $z_t = G(\vec{\xi}_t)$.

Takens' delay embedding theorem [1] and extensions [15] indicate when an attractor reconstructed from lag embedding of a time series will be topologically equivalent to the attractor in the true state space but the attractor shapes can be very different in the two spaces. (For example, compare the top and bottom rows in Fig. 3.)

A dynamical function $\mathcal{F}$ inferred from the embedded time series can have a different form from the function f in the true state space. The attractor described by the inferred state variables from MCMC estimation has the same shape as the state-space attractor for signal-to-noise ratios of as small as 8 dB (Fig. 3), although individual points $\vec{\xi}_t$ are not necessarily identical to the true state $\vec{x_t}$. Model parameters for f estimated from the inferred state variables are unbiased.

Note that in the Tinkerbell map there is a symmetry $(y, c_2, c_3) \to -(y, c_2, c_3)$. The measurement $z_t = x_t + \mu_t$ does not break this symmetry: if the priors on c_2 and c_3 range over both positive and negative values, the inferred y_t will include both branches of the symmetry. Here, we have reflected all inferred quantities onto one branch.

In many experimental situations, a measurement is made not of a state variable, but of some function of the state. The MCMC technique is able to draw valid inferences here as well. We have investigated this using the measurement function $z_t = y_t^2 + \mu_t$ in the Tinkerbell map. Since y_t takes on both positive and negative values, this measurement function introduces a genuine ambiguity when inferring from z_t to y_t. Nonetheless, despite the extra burden of inferring the sign on each y_t, the MCMC method is able to provide estimates of model parameters that cover the correct values and to reconstruct the attractor appropriately.

Preliminary work on the Lorenz equation system (modelled as an Euler finite-difference system) demonstrates that the MCMC method can make appropriate inferences with measurement functions such as $y_t^2 x_t$ or $x_t y_t z_t$.

Local Linear Functions

The use of a global model is important to the MCMC method or to any other method that seeks to construct inferred, de-noised variables since it allows each de-noised estimate ξ_t to appear in more than one conditional probability expression. In the examples considered so far, the model has played an essential role since the objective was either to estimate model parameters or to infer a model state. However, for different objectives such as estimating Lyapunov exponents from data, the specific form of the model is relatively unimportant so long as it has the flexibility to represent the dynamics that underlie the data.

Davies [6] used radial basis function models with the MCMC technique. Such models are hard to use within the BUGS framework, so we have explored the ability to make inferences from time series data using another class of general-purpose model, threshold autoregressive models [16] with fixed domains defined by $d_1, d_2, \ldots$ for each linear segment. For one-dimensional states, such models were introduced by Pijn for the analysis of EEG data [17], and can be written

$$\xi_{t+1} = \begin{cases} a_1\xi_t + b_1 & \text{if} \quad \xi_t \leq d_1 \\ a_2\xi_t + b_2 & \text{if} \quad d_1 < \xi_t \leq d_2 \\ & \vdots \end{cases} \tag{10}$$

To illustrate the technique, we model the quadratic map with noisy measurement:

$$\begin{aligned} x_{t+1} &= 4x_t(1 - x_t) \qquad (11) \\ z_t &= x_t + \mu_t \end{aligned}$$

We added gaussian white measurement noise μ_t to the dynamics of 11, collected $N = 500$ measurements and then fit these measurements to a locally linear model of the form Eq. 10 with 4 evenly spaced linear segments. The resulting piecewise linear model was then used to estimate the lyapunov exponent of the data using the mean of the logarithm of the slope of linear segment on which each data point fell.

Fig. 4 shows the estimated lyapunov exponents using MCMC estimation, ordinary least squares regression, and total least squares regression [18]. Due to the measurement-noise induced bias in least squares regression, the lyapunov exponent estimated from the least squares locally linear model is biased to be too low for large amounts of measurement noise (Fig. 4B). The MCMC technique also produces a somewhat biased estimate, but at moderate noise levels (SNR $>$ 7 dB) is systematically superior to the least squares model's estimates. (Since in this system $\sigma_\nu = 0$, total least squares can also be used for the local linear estimates, but produces segments that are consistently too steep. We believe this is due to the short length of domains.)

Discussion

We have shown that the MCMC technique offers a practical means of analyzing time series that are contaminated with measurement noise. In many cases it provides markedly superior performance to ordinary least squares regression.

Davies [6] has recently demonstrated the use of MCMC for noise reduction in data when the form of the dynamical model is unknown. Heald [7] has shown that Bayesian methods allow estimation of the relative sizes of dynamical and measurement noise when the dynamical model and its parameters are known. We extend this previous work by showing that MCMC allows estimation of the dynamical system's parameters simultaneously with the estimation of measurement and dynamical noise, and that even parameters relating to unmeasured components of the dynamical system's state can be accurately measured. Davies [6] showed that MCMC can estimate linear parameters of a general purpose model (radial-basis function) and speculated that estimation of nonlinear parameters could be made was well. Our work supports Davies' speculation: the nonlinear parameters of the Ikeda system were accurately estimated by MCMC.

An important issue in the MCMC technique is convergence of the iterates of the Markov chain to the true probability distribution. We use 1000 to 10000 iterations before commencing sampling. Depending on the particular model being used and the number of start-up iterations, the computations take approximately 1-60 minutes for time series of length 1000 on a standard 300 MHz Intel 586 computer. The required convergence time generally increases with the measurement noise level. In the trials reported here, convergence was easy to assess because we know the actual values used in the data generation. When this is not the case, one can use other diagnostics for convergence [19] or one can use simulations to estimate convergence requirements.

Use of prior probabilities that are sensibly related to the measured data — e.g., the measurement noise variance is no larger than the variance of the data, model parameters have values centered on those found from linear regression — helps to speed regression. Using the data to set priors, a practice termed "empirical Bayes"[12], is in some sense a violation of the Bayesian approach: the priors are reflecting our knowledge *before* the data have been seen. We note that the empirical priors that we set use information that is conveyed by essentially two degrees of freedom in the data, the center and width, which is typically a minute part of the overall data. In addition, experimental data has generally been scaled by amplifiers, etc. in a way that involves previous examination of a signal outside of the recorded data set. Thus, the empirical priors may reasonably be regarded as genuine priors, where the recorded data themselves serve merely to convey information about how the experimenter set up the apparatus.

When carrying out noise reduction on data, the choice of a method should depend on the use to which the noise-reduced data will be put. If the purpose is, for example, to estimate dimensions or power spectra, a model-independent noise reduction method may be appropriate such as those reviewed in [3] and [4]. However, if the de-noised data are to be fitted to a model, we suggest that the model should be used as well in the noise reduction step. The MCMC method appears to be effective in this regard.

We note that the de-noised data $\vec{\xi}_t$ are generally not the same as the true state $\vec{x}_t$ but we find that $\vec{\xi}_t$ fall close to the true attractor. (E.g., see Fig. 3.) This is perhaps to be expected; the estimate of the measurement noise $\widehat{\mu}_t$ is based on two components: the estimated parameters of the model and the redundancy in $\widehat{\mu}_t$ that stems from the participation of measurement z_t in more than one embedding vector. (In an embedding space of dimension m, z_t appears in m embedding vectors for most t.) All of the data points are participating in the estimates of the model parameters, making these fairly well known. However only m data points contribute to the estimate of each $\widehat{\mu}_t$. The ability to estimate

$\widehat{\mu_t}$ is based not so much on the power of averaging the redundant appearances of z_t in multiple embedding vectors as on the movement of $\vec{\xi_t}$ towards the functional surface defined by the model f. The movement is "towards" rather than "onto" because dynamical noise allows ξ_t to be off of the surface. The appearance of $z_t = \xi_t - \widehat{\mu_t}$ in multiple embedding vectors constrains $\widehat{\mu_t}$ to obey a set of consistency conditions

$$\begin{aligned} \xi_t &= f(\xi_{t-1}, \xi_{t-2}, \ldots) + \nu_{t-1} \\ \xi_{t+h} &= f(\xi_t, \xi_{t-1}, \ldots) + \nu_t \\ \xi_{t+2h} &= f(\xi_{t+1}, \xi_t, \ldots) + \nu_{t+1} \\ &\vdots \end{aligned}$$

These multiple conditions allow the discrimination between measurement and dynamical noise to be made, unlike the situation in Eq. 4 where both forms of noise are brought together into the single term ω_t.

The power of the MCMC technique to infer deterministic dynamics in the presence of large amounts of measurement noise invites the question of whether the technique will similarly infer determinism even when there is none and provide precise (but meaningless) fits of deterministic models to data that are purely dynamical noise. This is indeed the case. For example, in trials fitting pure noise to locally linear maps, MCMC underestimates the level of dynamical noise, accounting for much of the time series in terms of measurement noise. This occurance of this situation can be identified using the high variance in the estimated model parameters and by the cross-validation technique of comparing two models generated from the first half and second half of the data set.

It should be kept in mind that the inferences drawn by MCMC are conditioned on the assumed model. The inferences do not directly confirm or reject the correctness of the model; for this purpose it is necessary to introduce alternative models and examine the relative evidence for each of the models. Techniques for doing this are described in Carlin and Louis [12].

The ability to estimate the dynamics of components of the dynamical state that are not directly measured should facilitate the interpretation of data in terms of physical models. One can foresee applications, for example, such as using measurements of heart rate to estimate the parameters of the components of the cardio-respiratory regulatory system that are not directly measureable. Whether such estimation is practical will depend in part on the quality of the models being used. The incorporation of dynamical noise into the estimation, as reported here, allows for a certain amount of mismatch between the model and the dynamics. Further work needs to be done to understand how much mismatch is possible. The Bayesian framework is ideal for such estimation, because it allows explicit statement (using the prior probability distributions on parameters) of the confidence in different parts of the model.

Acknowledgements

CB was supported by a grant from the Hughes Summer Research Program at Macalester College. We thank James Heald for extensive and insightful comments on an early version of the manuscript, and Lenny Smith for suggestions

and insight into the relationships between models and data. We would also like to note that all of the specialized software used in the research reported here (BUGS and the R statistics language) is freely available over the Internet and we acknowledge the contribution of the many people who have worked on these projects.

References

[1] F. Takens (1981) Lecture Notes in Math, v. 898, Springer, New York

[2] A.M. Fraser and H.L. Swinney (1986) "Independent coordinates for strange attractors from mutual information" Phys. Rev. A, v.33, 1134

[3] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, T. Schreiber (1993) On noise reduction methods for chaotic data, Chaos, 3, 127

[4] H. Kantz and T. Schreiber (1997) "Nonlinear time series analysis" Cambridge Univ. Press.

[5] T. Sauer and J.A. Yorke (1993) How many delay coordinates to you need? Int. J. Bifurc. and Chaos, v.5, 737; M. Ding, C. Grebogi, E. Ott, T. Sauer, and J.A. Yorke (1993) Plateau onset for correlation dimension: when does it occur? Phys Rev. Lett. v.65, 3211

[6] M.E. Davies (1998) Nonlinear noise reduction through Monte Carlo Sampling, Chaos v.8 (4), 775-781

[7] J.P.M. Heald and J.Stark, Estimation of noise levels for models of chaotic dynamical systems, Preprint

[8] R. Meyer and N. Christensen (2000), Bayesian reconstruction of chaotic dynamical systems, PRE v.62(3), 3535-3542

[9] P.E. McSharry and L.A. Smith (1999) Better nonlinear models from noisy data: attractors with maximum likelihood, PRL v. 83(21), 4285-4288

[10] E.J. Kostelich and J.A. Yorke (1988) Noise reduction in dynamical systems, Phys. Rev. A v.38, 1649

[11] M.E. Davies (1992) Noise reduction by gradient descent, Int. J. Bifurc. and Chaos, v. 3, 113

[12] B.P. Carlin and T.A. Louis, "Bayes and empirical Bayes methods for data analysis" 2nd edition, Chapman & Hall/CRC, London (2000)

[13] W.R. Gilks, A. Thomas, A., and D.J. Spiegelhalter (1994) A language and program for complex Bayesian modelling. The Statistician, 43, 169-78. The software is available at `http://www.mrc-bsu.cam.ac.uk/bugs/`. It provides a high-level, symbolic means of specifying the conditional probabilities for MCMC. The BUGS-readable files that we created for this study are available at `http://www.math.macalester.edu/~kaplan/Papers` along with other software for replicating the results reported in this paper.

[14] K.T. Alligood, T.D. Sauer, J.A. Yorke, "Chaos: An introduction to dynamical systems" Springer-Verlag 1996

[15] T. Sauer, J. Yorke, and M. Casdagli (1991) Embedology, J. Stat. Phys., v.65, 579

[16] H. Tong (1990) Non-linear time series analysis, Oxford Univ. Press

[17] J.P.M. Pijn and F. Lopes da Silva (1993) Propagation of electrical activity: nonlinear association and time delays between EEG signals. In S. Zschocke, E.J. Speckmann (eds) Basic mechanisms of the EEG, Birkauser, Boston, 41-61

[18] W. Gander and U. von Matt (1994) Solving problems in scientific computing using Maple and Matlab, 2nd Ed., Springer

[19] N. Best, M.K. Cowles, K. Vines, "Convergence diagnosis and output analysis software for Gibbs sampling output" MRC Biostatistics Unit, Inst. of Public Health, Cambridge, UK `bugs@mrc-bsu.cam.ac.uk`

Appendix

The BUGS systems reads in the description of a model in the form of statements that describe the prior distribution and the relationships between variables. From these relationships are deduced the forms of the single-variate conditional probabilities needed in the MCMC technique. The following is the complete model specification file needed to describe the Tinkerbell model with a measurement $z_t = y_t^2 + \mu_t$.

```
# BUGS specification for Tinkerbell Map
# measuring Y^2

model tinkerbell;
const N=100;  # 100 data points

var x[N], y[N], # the state variables
    m1[N], meas1[N],         # the measurements
    dx[N], dy[N], # calculation intermediates
    atau, asigma, # measurement noise amplitude
    btau, bsigma, # dynamical noise in x
    ctau, csigma, # dynamical noise in y
    c[4];         # the parameters

data meas1 in "tinky2.dat"; # file with time series data

# files with starting values for the markov chain.
inits in "tink.ini", x in "tinkx.dat", y in "tinky.dat";

{
# Specification of the priors
x[1] ~ dnorm(0,1) I(-1,2); # priors on initial conditions
y[1] ~ dnorm(0,1) I(-1,1);
c[1] ~ dunif(-1,0);        # priors on the model parameters
c[2] ~ dunif(-1,0);
c[3] ~ dunif(0,5);
c[4] ~ dunif(-1,0)
# priors on the noise terms
atau  ~ dgamma(1.0E-3, 1.0E-3); # measurement noise
btau  ~ dgamma(1.0E-3, 1.0E-3); # dynamical noise in x
ctau  ~ dgamma(1.0E-3, 1.0E-3); # dynamical noise in y
```

```
for (i in 2:N ) {
   # the dynamics
   dx[i] <- x[i-1]*x[i-1] - y[i-1]*y[i-1] + c[1]*x[i-1] + c[2]*y[i-1];
   dy[i] <- 2*y[i-1]*x[i-1] + c[3]*x[i-1] + c[4]*y[i-1];
   # dynamical noise
   x[i] ~ dnorm(dx[i],btau) I(-1,2);
   y[i] ~ dnorm(dy[i],ctau) I(-1,1);
}

# the measurements
for (i in 1:N ){
   # the measurement function
   m1[i] <- y[i]*y[i];
   meas1[i] ~ dnorm(m1[i], atau);
}

}
```

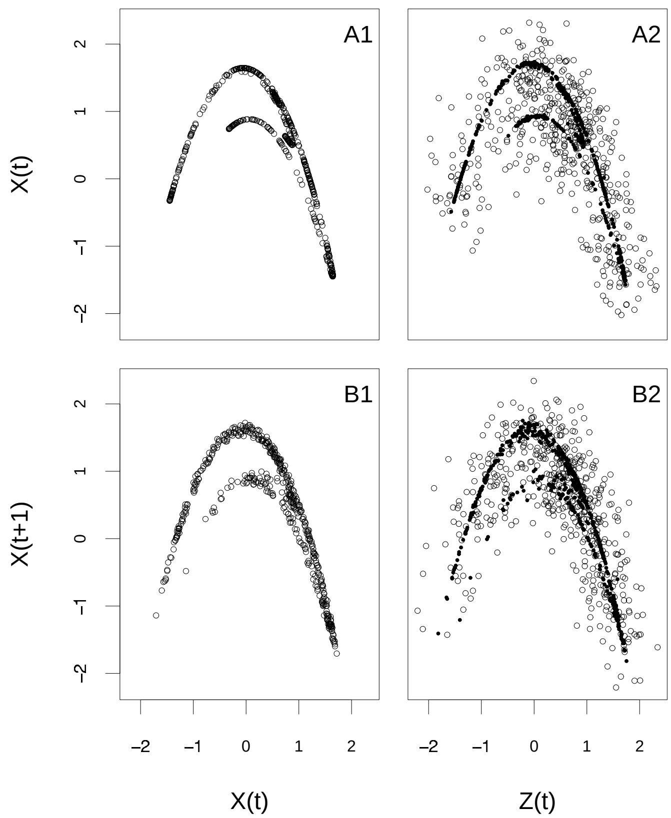

Figure 1: Henon Map data with dynamical and measurement noise. A1: True state variable $\vec{x}_t = (x_t, x_{t-1})$ with $\sigma_\nu = 0$. A2: $\circ$, measured data (z_t, z_{t-1}) with $\sigma_\mu = 0.3$; $\bullet$, inferred state ξ_t, ξ_{t-1} using the MCMC method. B1: True state variable with $\sigma_\nu = 0.04$. B2: $\circ$, measured data with $\sigma_\mu = 0.3$; $\bullet$, inferred state.

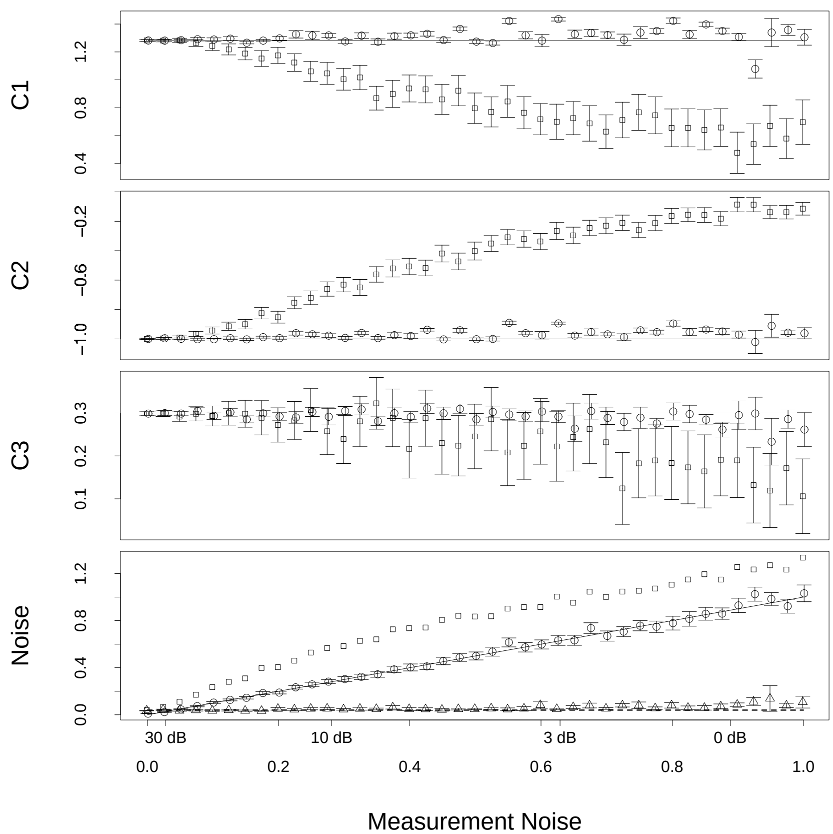

Figure 2: Parameters estimated using linear regression and MCMC from Eq. 4 versus measurement noise (measured in terms of σ_μ and also shown as a signal-to-noise ration in dB) for time series of length $N = 500$ with $\sigma_\nu = 0.04$.
$\square$ linear regression; $\circ$ MCMC. Solid lines show the true values of the parameters. For "noise", both measurement noise $\circ$ and dynamical noise $\triangle$ are shown for the MCMC method.

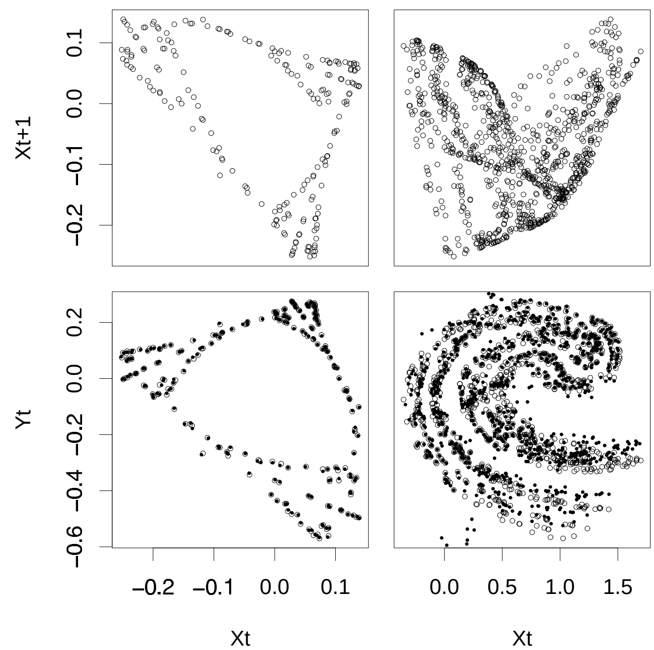

Figure 3: Top row: Lag embeddings (z_{t+1}, z_t) of time series from the Ikeda (right) and Tinkerbell (left) maps with $\sigma_\mu = 0$. Bottom row: $\circ$ the true state ($\vec{x}_t$ for the Ikeda system, (x_t, y_t) for Tinkerbell) together with the inferred state $\bullet$ from the MCMC technique. There is virtually identical placement onto the attractor with measurement noise as large as 8dB. Note that some of the de-noised points are off the attactor. In this case this stems from insufficient convergence time in the MCMC method; these points moved onto the attractor when convergence time is increased.

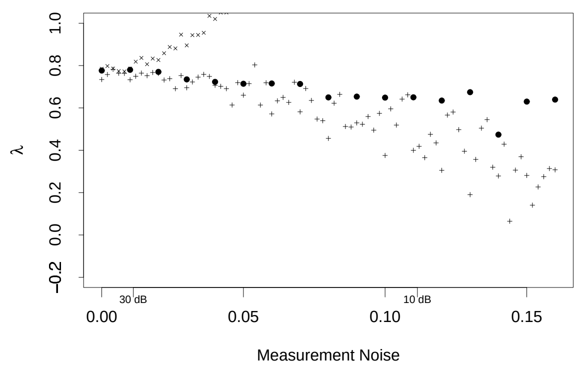

Figure 4: Lyapunov exponent λ estimated from a local linear model (Eq. 10) fit to data with measurement noise for the quadratic map (Eq. 11) versus size of the measurement noise σ_μ. $\bullet$, MCMC estimation based on ξ_t; $+$, ordinary least squares estimation based on z_t; x, total least squares estimation based on z_t.