Applications of dynamical inference to the analysis of noisy biological time series with hidden dynamical variables

A. Duggento*, D. G. Luchinsky*,†, V. N. Smelyanskiy†, M. Millonas** and P. V. E. McClintock*

*Department of Physics, Lancaster University, Lancaster LA1 4YB, UK
†NASA Ames Research Center, Mail Stop 269-2, Moffett Field, CA 94035, USA
**Mission Critical Technologies Inc., 2041 Rosecrans Ave. Suite 225 El Segundo, CA 90245, USA

Abstract. We present a Bayesian framework for parameter inference in noisy, non-stationary, nonlinear, dynamical systems. The technique is implemented in two distinct ways:
(i) *Lightweight implementation*: to be used for on-line analysis, allowing multiple parameter estimation, optimal compensation for dynamical noise, and reconstruction by integration of the hidden dynamical variables, but with some limitations on how the noise appears in the dynamics ;
(ii) *Full scale implementation*: of the technique with extensive numerical simulations (MCMC), allowing for more sophisticated reconstruction of hidden dynamical trajectories and dealing better with sources of noise external to the dynamics (measurements noise).

Keywords: <Enter Keywords here>
PACS: <Replace this text with PACS numbers; choose from this list: `http://www.aip..org/pacs/index.html`>

Developing earlier works [1, 2], we consider the following an M-dimensional time-series $\mathscr{Y} = \{\mathbf{y}_n \equiv \mathbf{y}(t_n)\}$ ($t_n = nh$), representing N observations of the system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}|\mathbf{c}) + \sqrt{\mathbf{D}}\,\xi(t), \qquad \mathbf{y}(t) = \mathbf{g}(\mathbf{x}|\mathbf{b}) + \sqrt{\mathbf{M}}\,\eta(t). \tag{1}$$

The first eq.s(1) defines the L-dimensional underlying stochastic dynamics (with white uncorrelated noise source) and the second one defines the observed variable $\mathscr{Y}$ (with an extra observational noise source). Our task is to infer the unknown model parameters, their time variations, the noise intensities and $\mathscr{X}$-trajectory: $\mathscr{M} = \{\mathbf{c}(t), \mathbf{b}(t), \mathbf{D}, \mathbf{M}, \{\mathbf{x}_n\}\}$.

The form of the likelihood depends on the approximations of the theory. For an Euler approximation of the dinamics $\mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{f}(\mathbf{x}_n^*|\mathbf{c}) + \sqrt{h\mathbf{D}}\xi_n$, with $\mathbf{x}_n^* = (\mathbf{x}_{n+1} + \mathbf{x}_n)/2$, the minus log-likelihood function $S = -\ln \ell(\mathscr{Y}|\mathscr{M})$ can be written as: (see [3, 1])

$$\begin{aligned} S = {}& \frac{N}{2}\ln|\mathbf{D}| + \frac{h}{2}\sum_{n=0}^{N-1}\left\{\frac{\partial(\mathbf{f}(\mathbf{x}_n)|\mathbf{c})_k}{\partial x_k}. + [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]^{\mathrm{T}}\,\mathbf{D}^{-1}\,[\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]\right\} \\ & + \frac{N}{2}\ln|\mathbf{M}| + \frac{1}{2}\sum_{n=1}^{N}[\mathbf{y}_n - \mathbf{g}(\mathbf{x}_n|\mathbf{b})]^T\,\mathbf{M}^{-1}\,[\mathbf{y}_n - \mathbf{g}(\mathbf{y}_n, \mathbf{x}_n|\mathbf{b})] + (L+M)N\ln(2\pi h), \end{aligned} \tag{2}$$

where $\dot{\mathbf{x}}_n = \frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{h}$ and summation over k is implicit in the term $\frac{\partial(\mathbf{f}(\mathbf{x}_n)|\mathbf{c})_k}{\partial x_k}$.

Let us assume for a moment that $\mathscr{X}$ is given. In this case, by the parameterising the vector field $\mathbf{f}(\mathbf{x}_n^*|\mathbf{c}) = \mathbf{U}(\mathbf{x}_n^*)\mathbf{c} \equiv \mathbf{U}_n\mathbf{c}$, linearly in respect of its parameters, and assuming a multivariate normal prior PDF for $\mathbf{c}$, the posterior is also normal, and its mean is given by [3]:

$$\begin{aligned} \mathbf{c}_{post} &= \left(h \sum_{n=0}^{N-1} \mathbf{U}_n^{\mathrm{T}} \mathbf{D}^{-1} \mathbf{U}_n \right)^{-1} \left(h \sum_{n=0}^{N-1} \left[\mathbf{U}_n^{\mathrm{T}} \mathbf{D}^{-1} \dot{\mathbf{x}}_n - \frac{1}{2} \sum_{l=1}^{L} \frac{\partial \mathbf{U}_{lm}(\mathbf{x})}{\partial x_l} \right] \right), \\ \langle \mathbf{D} \rangle &= \frac{h}{N} \sum_{n=0}^{N-1} \left[\dot{\mathbf{x}}_n - \mathbf{U}_n \mathbf{c} \right] \left[\dot{\mathbf{x}}_n - \mathbf{U}_n \mathbf{c} \right]^{\mathrm{T}}. \end{aligned} \tag{3}$$

The obtained result holds uniquely in presence of additive noise ($\mathbf{D}$ constant). If this was not the case, then an extra parameterisation of the noise should have been employed, and heavy approximation and assumptions made in order to make the problem algebraically treatable.

When $\mathscr{X}$ is not observable, a global optimization technique should, in general, be employed. In the next two following sections two example will be discussed. In the first one, a 'lightweight' implementation will be used and the use of global optimization will be avoided. In the second example the most probable state for the dynamical space will be obtained thanks to an MCMC techinique.

LIGHTWEIGHT IMPLEMENTATION

In our first example, we decode the parameters of a system of neurons modelled by an L-dimensional system of FitzHugh-Nagumo (FHN) oscillators [5]:

$$\dot{v}_j = -v_j\left(v_j - \alpha_j\right)\left(v_j - 1\right) - q_j + \eta_j + \sqrt{D_{ij}}\,\xi_j, \tag{4}$$

$$\dot{q}_j = -\beta\, q_j + \gamma_j\, v_j; \qquad \langle \xi_j(t)\, \xi_i(t') \rangle = \delta_{ij}\delta(t-t'), \tag{5}$$

$$y_i = X_{ij}\, v_j. \qquad i,j = 1,\ldots,L. \tag{6}$$

where v_j models the membrane potentials and q_j are slow recovery variables. Parameters η_i control the potential threshold for the self-excited dynamics, controlling the firing rate, and they will be considered as time-varying parameters. We assume that neither v_j nor q_j are read directly (i.e. 'hidden' variables), but that the measurements are made through an *unknown* measurement matrix X in eq.(6). Our tasks are to: (i) reconstruct coefficients appearing in Eq.(4-5); (ii) reconstruct the mixing matrix X; (iii) reconstruct the hidden variables q_j; (iv) perform tasks (i)-(iii) taking into account that some parameters might have explicit time dependence. We assume no measurement noise in eq.(6): indeed in such systems the measurement noise is often negligible and in this way we can avoid global optimization and better estimate the performance of the Bayesian inference itself. Following [2], a convenient way to treat this problem is by integration of the slow recovery variable q_i and to substitute it into the top equation in eq.(4), and consequently in eq.(6) we obtain the explicit form for the dynamics of the readout variable:

$$\begin{aligned} \dot{y}_i &= \tilde{\eta}_i + \tilde{\alpha}_{ij} y_j + \tilde{b}_{ik_1k_2} y_{k_1} y_{k_2} + \tilde{c}_{ik_1k_2} y_{k_1} y_{k_2}^2 \\ &+ e^{-\beta t}\tilde{q}_i - \textstyle\int_0^t e^{\beta(t-\tau)} \gamma_{ij} y_j d\tau + \sqrt{\tilde{D}_{ij}}\xi_j(t), \end{aligned} \tag{7}$$

where parameters of the gtransformed dynamics in eq.(7) are function of the original parameters and the matrix X. This explicit dependence is given in [2]. Although the number of base functions N_ϕ for the mixed dynamic is much larger than the number of polynomial terms in eq.s(4-5) the inferencial algorithm exibits good performances and high speed in inferring parameters even when few of them are explicetelly time dependent. Some results are presentented on Fig. 1.

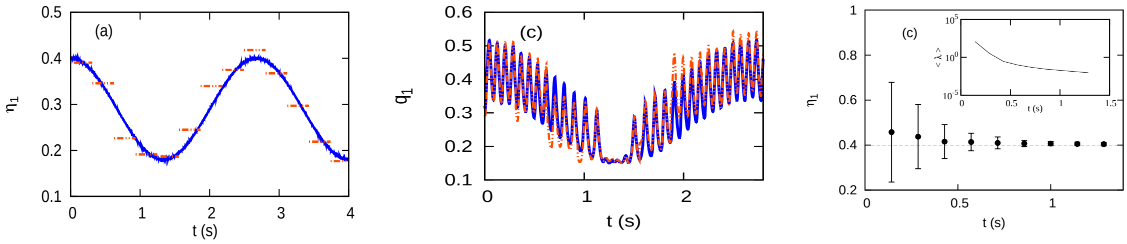

FIGURE 1. Inference of η_1 and η_2 from eq.(4) for a 2-FHN system, while smoothly varying in the presence of noise. No prior knowledge of the model parameters is assumed. (a) The inferred values of η_1 (dashed red lines) are compared with their true values (full blue lines); here the ability in detecting continuous evolution of the control parameters in the adiabatic limit is demonstrated. (b) The recontructed time-trace of the hidden coordinate $q_1(t)$. (c) Typical convergence of the control parameters η_j as functions of the measurement time t; qualitative behaviour of the biggest eigenvalues of the covariance matrix is given in the box. In [2] a detailed discussion is presented.

MCMC IN AN ECOLOGICAL SYSTEMS

In ecological problems the emphases are on the off-line recovery of hidden population dynamics. We therefore consider a general MCMC approach. We consider an example of predator-prey dynamics, also considered in [6, 7, 8], where the cycling dynamics of the vole population in Finnish Lapland can be modelled by the following equations for the fluctuating densities of rodents N and their predators P (weasels, foxes, owls, and others). Dynamical inference method cannot be applied "as is" to the model of autors of [6, 7, 8] because: (i) the noise terms are multiplicative; (ii) the predator trajectory is hidden; and (iii) the prey dynamics is measured together with some measurement noise. If the first problem can be overcome by making an *ad hoc* change of variables, the second and third problem are more complex and can be solved in different ways accordingly to the approximations that one can introduce. In this respect, some ways has been investigated in [4]. Here, the aim is to show how an MCMC techinique can be employed for the reconstruction of the hidden dynamical sample. In particular it is very useful to analyse what happen in a one-dimensional approximation. For a detailed discussion of how to reduce to this system and what are the approximations involved see [4, 8, 7]. The resulting one-dimesional systems considered has the form:

$$\dot{x_1} = r\left(1 - e_1 \sin(2\pi t + \psi_0)\right) - \tilde{r}e^{x_1} - \frac{g e^{x_1}}{e^{2x_1} + h^2} - \frac{a z^{-1}}{e^x + d} + r\sigma_n \xi_n(t), \tag{8}$$

$$z = e^{-s_1 t - \frac{s_2}{2\pi}\cos(2\pi t)}\left(c_0 + s_3 \int_{t_0}^{t} \frac{d\tau}{n(\tau)} e^{s_1 \tau + \frac{s_2}{2\pi}\cos(2\pi\tau)}\right), \tag{9}$$

$$y(t) = x_1(t) + \sigma_{obs}\eta(t). \tag{10}$$

where the only observable is $y(t)$. For the sake of simplicity we assume the noise intensities to be fixed and introduce an abbreviated vector of the unknown parameters $\tilde{\mathscr{M}} = \{\mathbf{c}, \{\mathbf{x}_k\}\}$. The MCMC algorithm can be briefly summarized as follows: (i) Take an initial guess for $\tilde{\mathscr{M}}^{(0)} = \{\mathbf{c}^{(0)}, \{\mathbf{x}_k^{(0)}\}\}$; (ii) Sample a trajectory from $p(x_k|x_{k-1}, x_{k+1}, \tilde{\mathscr{M}}, \mathbf{D}, \sigma_{obs}, y_t)$ for $k = 0, \ldots, K$ using Gibbs sampler with Metropolis-Hastings (M-H) steps; (iii) Sample model parameters from $p(\tilde{\mathscr{M}}|\{x_t\}, \mathbf{D}, \sigma_{obs}, \{y_t\})$ using M-H algorithm; (iv) Repeat steps (ii)-(iv) until convergence is achieved.
A graphical results is summarised on Fig.2.

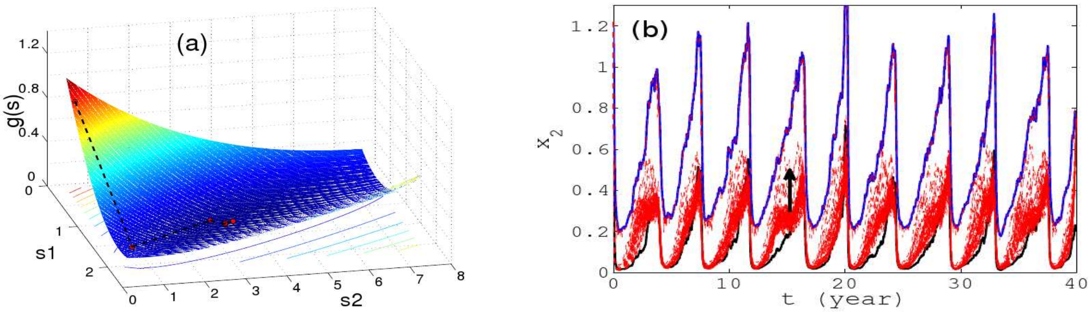

FIGURE 2. Inference parameters from eq.(8); (a) Typical evolution of the solution of the optimization problem starting from some initial values and descending the hyperplane of the cost function defined by the posterior minus-log-likelihood; (b) Results of the MCMC calculations: convergence of the unknown predator trajectories from an initial guess (solid black line at the bottom of the figure) to the actual trajectory (solid blue line at the top of the figure) is shown by dashed red lines. The arrow indicates the direction of convergence as a function of number of iterations.

CONCLUSIONS

We have considered the problem of dynamical inference in presence of noise and provided different approaches for a fast implementation (on-line applications), investigating the boundaries of the resolution for slowly varying parameters; and for more computationally demanding problem of global reconstruction with the heavy use of MCMC for discovering the latent state variables for the extreme case of missing data.

REFERENCES

1. D. G. Luchinsky, V. N. Smelyanskiy, A. Duggento, and P. V. E. McClintock, *Phys. Rev. E* **77**, 061105 (2008).
2. A. Duggento, D. G. Luchinsky, V. N. Smelyanskiy, I. Khovanov and P. V. E. McClintock, *Phys. Rev. E* **77**, 061106 (2008).
3. V. N. Smelyanskiy, D. G. Luchinsky, D. A. Timucin, and A. Bandrivskyy, *Phys. Rev. E* **72**, 026202 (2005).
4. D. G. Luchinsky, V. N. Smelyanskiy, M. Millonas, P. V. E. McClintock, *Europ. Phys. J. B* **65** pp. 369–377 (2008)
5. J. Nagumo, S. Animoto, and S. Yoshizawa, *Proc. Inst. Radio Engineers* **50**, 2061 (1962).
6. I. Hanski, H. Henttonen, E. Korpimaäki, L. Oksanen, P. Turchin, *Ecology* **82**(6), 1505 (2001)
7. O. Gilg, I. Hanski, B. Sittler, *Science* **302**, 866 (2003)
8. P. Turchin, S.P. Ellner, *Ecology* **81**(11), 3099 (2000)