CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Lancaster E-Prints

Published in *J. Stat. Mech.* P01025 (2009).

Inferential framework for nonstationary dynamics: theory and applications

Andrea Duggento[1], Dmitri G. Luchinsky[1,2], Vadim N. Smelyanskiy[2], Peter V. E. McClintock[1]

[1]Department of Physics, Lancaster University, Lancaster LA1 4YB, UK
[2]NASA Ames Research Center, Mail Stop 269-2, Moffett Field, CA 94035, USA

E-mail: a.duggento@lancaster.ac.uk

Abstract. An extended Bayesian inference framework is presented, aiming to infer time-varying parameters in nonstationary nonlinear stochastic dynamical systems. The convergence of the method is discussed. The performance of the technique is studied using, as an example, signal reconstruction for a system of neurons modeled by FitzHugh-Nagumo oscillators: it is applied to reconstruction of the model parameters and elements of the measurement matrix, as well as to inference of the time-varying parameters of the non-stationary system. It is shown that the proposed approach is able to reconstruct unmeasured (hidden) variables of the system, to determine the model parameters, to detect stepwise changes of control parameters for each oscillator, and to track the continuous evolution of the control parameters in the adiabatic limit.

1. Introduction

The inference of parameters from time-series, for models based on stochastic nonlinear dynamical systems, is an open field that is currently attracting much attention on account of its importance and wide applicability. Such models can reproduce a diversity of complex phenomena in technology and nature, providing the information needed for diagnosis of faults, prognosis of future conditions, or control, in e.g. reactors [1], helioseismology [2], physiology [3] and neuroscience [4].

Much effort has been devoted to the solution of the problem of parameter inference under different conditions [3, 5, 6, 7, 8, 9, 10, 11, 12, 13]. However, time variation of the control parameters in non-stationary systems, and how the inference algorithm can be adjusted to accommodate them, has not yet been widely explored. It is therefore highly desirable to extend the inferential framework in this direction, hoping to encompass almost-real-time tracking of time-varying parameters.

The algorithms developed earlier all have their own particular advantages and disadvantages. The disadvantages include e.g. the requirement of extensive numerical simulation [9, 10, 13], or the need for very large amount of data [5, 7] (cf. econometric series analysis [14]). Their adaptation for parameter-tracking in non-stationary stochastic nonlinear systems cannot be done in any obvious way. Moreover, most calculations of flows produce biased estimates due to their lack of a term related to the Jacobian of transformation from stochastic to deterministic variables. This is a very important issue, because this term gives [15] a leading-order contribution to the inference results in the presence of strong dynamical noise. In our own earlier work we introduced an analytic solution of the dynamical inference problem [15, 16] based on Bayesian statistics and a path-integral formulation of stochastic dynamics. As a direct result, fast, unbiased estimation of model parameters became possible. The technique also provides optimal compensation for dynamical noise.

In the present paper we extend the Bayesian framework, allowing us to infer information encoded in the time-varying control parameters of a nonlinear non-stationary system, almost in real time. The parameter-tracking algorithm is effectively embedded into the learning inferential framework, enabling us to reconstruct both the model parameters and also the unmeasured (hidden) variables of the system. Such an inferential framework can have a wide range of interdisciplinary applications, including nanosensors [17], physiology [18], and aerospace [19, 20].

We consider an application of the scheme to a model of physiological signalling. In particular, we demonstrate how the technique might be used for the analysis of signals from neuronal systems. Their dynamics have not yet been well understood, and the highly nonlinear and non-stationary nature of their behavior makes it difficult to apply standard techniques for the reliable inference of control parameters. Internal and measurement noises in these systems strongly affect their dynamics, and the time variation of the control parameters is directly related to information coding. We show that our approach is able to decode the time evolution of the control parameters

in a system of neurons modeled as a set of FitzHugh-Nagumo (FHN) equations [21, 22, 23, 24], including detection of large stepwise changes for either oscillator and continuous variation in the adiabatic limit. We illustrate this ability by our reconstruction of the system parameters assuming that the original parameters of the model are unknown, that only one coordinate of each oscillator is available for recording, and that the measurements are mixed by a measurement matrix.

The paper is organized as follows. Sec. 2 presents the general approach of the Bayesian inferential framework, and the main idea of the inferential framework for non-stationary dynamics. In Sec. 2.2 the theory of Bayesian inference for a system of L FHN oscillators is developed, providing the basis for physiological applications; and in Sec. 2.3 the case of nonstationary dynamics is discussed. Sec. 3 presents simulation results. In Sec. 4 results are summarized and conclusions are drawn.

2. Bayesian inferential framework for non-stationary dynamics

2.1. A general approach

The fundamental problem in dynamical inference can be defined as follows. An M-dimensional time-series of observational data $\mathcal{Y} = \{\mathbf{y}_n \equiv \mathbf{y}(t_n)\}$ ($t_n = nh$) is provided, and the time variation of the unknown model parameters and the unknown dynamical trajectory $\mathcal{M} = \left\{\mathbf{c}, \mathbf{b}, \hat{\mathbf{D}}, \hat{\mathbf{M}}, \{\mathbf{x}_n\}\right\}$ is to be inferred, under the conditions that the underlying dynamics can be described by a set of L-dimensional ($L \geq M$) stochastic differential equations in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)|\mathbf{c}) + \sqrt{\hat{\mathbf{D}}}\boldsymbol{\xi}(t), \tag{1}$$

and that the observations $\mathcal{Y}$ are related to the actual unknown dynamical variables $\mathcal{X} = \{\mathbf{x}_n \equiv \mathbf{x}(t_n)\}$ via the following measurement equation

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t)|\mathbf{b}) + \sqrt{\hat{\mathbf{M}}}\boldsymbol{\eta}(t). \tag{2}$$

Here $\mathbf{g}(\mathbf{x}|\mathbf{b})$ is a measurement function, $\boldsymbol{\xi}(t)$ and $\boldsymbol{\eta}(t)$ are L- and M-dimensional Gaussian white noises, and $\hat{\mathbf{D}}$ and $\hat{\mathbf{M}}$ are $L \times L$ and $M \times M$ dimensional dynamical and measurement diffusion matrices respectively. It is assumed that the sampling is dense enough to use the Euler mid-point discretization; in this case Eqs. (1),(2) can be discretized in the form

$$\left.\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n + h\,\mathbf{f}(\mathbf{x}_n^*|\mathbf{c}) + \sqrt{h\,\hat{\mathbf{D}}}\boldsymbol{\xi}_n, \\ \mathbf{y}_n &= \mathbf{g}(\mathbf{x}_n|\mathbf{b}) + \sqrt{\hat{\mathbf{M}}}\boldsymbol{\eta}_n, \end{aligned}\right\} \tag{3}$$

where $\mathbf{x}_n^* = (\mathbf{x}_{n+1} + \mathbf{x}_n)/2$.

In Bayesian statistics a given *prior* density $\rho_{\text{prior}}(\mathcal{M})$ that encloses expert knowledge of the unknown parameters and the *likelihood* function $\ell(\mathcal{Y}|\mathcal{M})$, the probability density to observe $\{\mathbf{y}_n(t)\}$ given choice $\mathcal{M}$ of the dynamical model, are both employed to

calculate the so-called *posterior* density $\rho_{\rm post}(\mathcal{M}|\mathcal{Y})$ of the unknown parameters $\mathcal{M}$ conditioned on observations through the use of Bayes' theorem

$$\rho_{\rm post}(\mathcal{M}|\mathcal{Y}) = \frac{\ell(\mathcal{Y}|\mathcal{M})\,\rho_{\rm prior}(\mathcal{M})}{\int \ell(\mathcal{Y}|\mathcal{M})\,\rho_{\rm prior}(\mathcal{M})d\mathcal{M}} \tag{4}$$

The construction of the likelihood is of great importance because it contains all the approximations of the theory. For independent white Gaussian noise sources, the likelihood is given by a product over n of the probability to observe $\mathbf{y}_{n+1}$ at each time, (see [25, 15]) and the minus log-likelihood function $S = S_{dyn} + S_{meas} = -\ln \ell(\mathcal{Y}|\mathcal{M})$ can be written as

$$\begin{aligned} S = & \frac{N}{2}\ln|\hat{\mathbf{D}}| + \frac{h}{2}\sum_{n=0}^{N-1}\left\{\frac{\partial(\mathbf{f}(\mathbf{x}_n)|\mathbf{c})_k}{\partial x_k}. + [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]^{\rm T}\,\hat{\mathbf{D}}^{-1}\,[\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]\right\} \\ & + \frac{N}{2}\ln|\hat{\mathbf{M}}| + \frac{1}{2}\sum_{n=1}^{N}[\mathbf{y}_n - \mathbf{g}(\mathbf{x}_n|\mathbf{b})]^{\rm T}\,\hat{\mathbf{M}}^{-1}\,[\mathbf{y}_n - \mathbf{g}(\mathbf{y}_n,\mathbf{x}_n|\mathbf{b})] + \\ & + (L+M)N\ln(2\pi h), \end{aligned} \tag{5}$$

where $\dot{\mathbf{x}}_n = \frac{\mathbf{x}_{n+1}-\mathbf{x}_n}{h}$ and summation over k is implicit in the term $\frac{\partial(\mathbf{f}(\mathbf{x}_n)|\mathbf{c})_k}{\partial x_k}$. Here S_{dyn} and S_{meas} are the dynamical (the first two terms in the first line) and measurement (next three terms) parts of the minus log-likelihood function.

An explicit calculation with an algorithm capable of minimizing Eq. (5) has been provided in [25]. Basically it consists of an iterative optimization of S in the space of dynamical paths $\{\mathbf{x}_n\}$ and in the space of parameters $\left\{\mathbf{c}, \mathbf{b}, \hat{\mathbf{D}}, \hat{\mathbf{M}}\right\}$.

Let assume for now that the hidden dynamical variables $\{\mathbf{x}_n\}$ are given and let focus our attention only on the minimization of S_{dyn}. The key point is the parameterization of the fields in respect of the parameter vector $\mathbf{c}$:

$$\mathbf{f}(\mathbf{x}|\mathbf{c}) = \hat{\mathbf{F}}(\mathbf{x})\,\mathbf{c}, \tag{6}$$

where matrix $\hat{\mathbf{F}}$ takes the form

$$\hat{\mathbf{F}} = \left[\begin{pmatrix} \phi_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \phi_1 \end{pmatrix} \dots \begin{pmatrix} \phi_F & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \phi_F \end{pmatrix}\right], \tag{7}$$

and $\{\phi_i\}$ are the F-dimensional sets of arbitrary base functions.

With this linear parameterization of $\mathbf{f}$, and given that the log-likelihood quadratic in $\mathbf{f}$ (see Eq. 5), we obtain a quadratic log-likelihood in respect of the vector parameter $\mathbf{c}$. Hence, using a multivariate normal distribution for the prior probability immediately leads to a multivariate normal distribution for the posterior. This is highly desirable for two reasons: (i) from one side a Gaussian posterior is computationally extremely convenient because there is the guarantee of a unique maximum, the mean vector and covariance matrix completely characterize the distribution and give us the most significant information; (ii) but most of all the multivariate normal posterior can be used again as a prior in presence of a new block of data and knowledge about the system can easily be updated. This last feature is essential for any real time application

because it ensures that the complexity of the algorithm does not change with the length of the input data-stream.

With a multivariate normal distribution as a prior for parameters $\mathbf{c}$, with mean $\bar{\mathbf{c}}$, and covariances $\hat{\boldsymbol{\Xi}}^{-1}_{\text{prior}}$, the stationary point of S_{dyn} is calculating recursively with the following equations (cf. [15]):

$$\langle \hat{\mathbf{D}} \rangle = \frac{h}{N} \sum_{n=0}^{N-1} \left[\dot{\mathbf{x}}_n - \hat{\mathbf{F}}_n \, \bar{\mathbf{c}} \right] \left[\dot{\mathbf{x}}_n - \hat{\mathbf{F}}_n \, \bar{\mathbf{c}} \right]^{\mathrm{T}}, \tag{8}$$

$$\langle \mathbf{c} \rangle = \hat{\boldsymbol{\Xi}}^{-1}_{\mathcal{X}}(\hat{\mathbf{D}}) \, \mathbf{w}_{\mathcal{X}}(\hat{\mathbf{D}}), \tag{9}$$

$$\mathbf{w}_{\mathcal{X}}(\hat{\mathbf{D}}) = \hat{\boldsymbol{\Xi}}^{-1}_{\text{prior}} \bar{\mathbf{c}} + h \sum_{n=0}^{N-1} \left[\hat{\mathbf{F}}^{\mathrm{T}}_n \, \mathbf{D^{-1}} \, \dot{\mathbf{x}}_{\mathbf{n}} - \frac{\mathbf{v}(\mathbf{x_n})}{\mathbf{2}} \right], \tag{10}$$

$$\hat{\boldsymbol{\Xi}}_{\mathcal{X}}(\hat{\mathbf{D}}) = \hat{\boldsymbol{\Xi}}_{\text{prior}} + h \sum_{n=0}^{N-1} \hat{\mathbf{F}}^{\mathrm{T}}_n \, \hat{\mathbf{D}}^{-1} \, \hat{\mathbf{F}}_n, \tag{11}$$

where $\hat{\mathbf{F}}_n \equiv \hat{\mathbf{F}}(\mathbf{x}_n)$, and the components of the vector $\mathbf{v}(\mathbf{x})$, are

$$v_m(\mathbf{x}) = \sum_{l=1}^{L} \frac{\partial F_{lm}(\mathbf{x})}{\partial x_l}, \quad m = 1, \ldots, F. \tag{12}$$

In absence of any prior knowledge about the system, a non informative prior can be used: the limit of an infinitely large normal distribution is set with the computational initial value of $\hat{\boldsymbol{\Xi}}_{\text{prior}} = 0$ and $\bar{\mathbf{c}}_{\text{prior}} = 0$.

At this point an optimization technique for reconstruction of the driving dynamics of $\mathcal{X}$ is needed. The problem is far too vast to be reviewed in any detail in the present work. Suggestions for reconstruction of the best path in the dynamical space range from the Markov Chain Monte Carlo (MCMC) techniques [26], to the extended Kalman filter [10], to the Langevin method of sampling the posterior [27]. To bound our discussion of how to tackle the problem in the case of non-stationay dynamics we will not dig into any optimization technique in the space of observable variables. Therefore, in the latter, a physiological example will be considered where the observational variable $\mathbf{y}$ is noiseless and has the same dimension of the dynamics-driving variable.

2.2. System of FitzHugh-Nagumo oscillators

To appreciate how non-stationary Bayesian inference could be applied, we want to provide a physiologically relevant example, and so we consider a typical situation of neuron readout. We will simulate neurons firing at the rate of ∼5-10 s^{-1} and our explicit aim will be to reconstruct the varying parameters with a correlation of ∼500-1000 ms. This means that we need a computational inference delay shorter that 500 ms, corresponding to a few periods of firing of the action potential. To model this spiking behavior we use the FitzHugh-Nagumo system in the form

$$\begin{aligned} \dot{v}_j &= -v_j \, (v_j - \alpha_j) \, (v_j - 1) - q_j + \eta_j + \sqrt{D_{ji}} \, \xi_i, \\ \dot{q}_j &= -\beta \, q_j + \gamma_j \; v_j, \end{aligned} \tag{13}$$

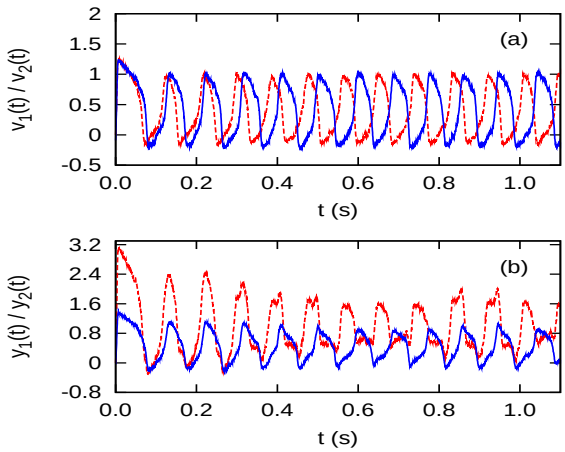

Figure 1. Time-series data generated by the model (13), (14) before and after mixing, for the parameters given in Table 1. Parameters η_1 and η_2 fluctuate between 0.35 and 0.45. The (blue) solid lines show $v_1(t)$ and $y_1(t)$, and the (red) dotted lines show $v_2(t)$ and $y_2(t)$.

$$\langle \xi_j(t)\, \xi_i(t') \rangle = \delta_{i\,j}\, \delta(t - t'), \quad j = 1 : L.$$

This system (13) represents the simplified dynamics of L non-interacting neurons [22], 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. Realistically, the membrane potential is difficult to measure from each single neuron, so we will assume that the variables v_i remain unobserved, and observable readout consists of mixed variables y_i defined as

$$y_i \;=\; X_{ij}\, v_j\,. \tag{14}$$

where X_{ij} is an unknown invertible mixing matrix whose coefficients have to be inferred. Examples of noisy signals before and after mixing are presented in Fig. 1. Often in such systems the measurement noise is negligible, and we therefore assume no noise in Eq. (14). This assumption will also allow us to avoid global optimization and to estimate better the performance of the Bayesian inference itself. Our task is to infer the set of model parameters $\mathcal{M} = \{\eta_i, \alpha_i, q_i(0), \gamma_i, D_{ij}, X_{ij}\}$, with on-line tracking of the time-varying parameters $\{\eta_i\}$ for each neuron, starting from the time series data $\{y_i\}$.

Following [32], a convenient way to treat this problem is by integration of the slow recovery variable q_i of equations in (13),

$$q_j(t) = \gamma_j \int_0^t d\tau e^{-\beta\,(t-\tau)} v_j(\tau) + e^{-\beta\, t} q_j(0). \tag{15}$$

and to substitute (15) into the top equation in (13), obtaining

$$\begin{aligned} \dot{v}_j &= -\alpha_j v_j + (1+\alpha_j) v_j^2 - v_j^3 + \eta_j \\ &- \gamma_j \int_0^t d\tau e^{-\beta\,(t-\tau)} v_j(\tau) - e^{-\beta\, t} q_j(0) + \sqrt{D_{ij}}\xi_j. \end{aligned} \tag{16}$$

with $q_j(0)$ as a set of initial coordinates for the unobservable variable $q_j(t)$, thus reducing the reconstruction of unobservable variables $q_j(t)$ to the inference of the L initial conditions $q_j(0)$.

The variables $v_j(t)$ are also not observed so, using Eq. (14) and substituting $\boldsymbol{v} = \boldsymbol{X}^{-1}\boldsymbol{y}$ into (16), we obtain in vector notation an explicit form for the dynamics of the readout variable:

$$\begin{aligned}\dot{\boldsymbol{y}} &= \boldsymbol{X}\boldsymbol{\alpha}\left(\boldsymbol{X}^{-1}\boldsymbol{y}\right) + \boldsymbol{X}(\boldsymbol{1}+\boldsymbol{\alpha})\left(\boldsymbol{X}^{-1}\boldsymbol{y}\right)^2 + \boldsymbol{X}\left(\boldsymbol{X}^{-1}\,\boldsymbol{y}\right)^3 + \\ &+ e^{-\beta t}\boldsymbol{X}\boldsymbol{q}_0 - \int_0^t e^{\beta(t-\tau)}\boldsymbol{X}\boldsymbol{\gamma}\left(\boldsymbol{X}^{-1}\boldsymbol{y}\right)d\tau + \boldsymbol{X}\boldsymbol{\eta} + \boldsymbol{X}\sqrt{\boldsymbol{D}}\boldsymbol{\xi}(t),\end{aligned} \tag{17}$$

where $q_0 = q(t=0)$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are matrices with α_i and γ_i on the respective diagonals, and

$$\left(\boldsymbol{X}^{-1}\,\boldsymbol{y}\right)^n = \begin{pmatrix} \left(\sum_{i=1}^L x_{1i}y_i\right)^n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \left(\sum_{i=1}^L x_{Li}y_i\right)^n \end{pmatrix}.$$

Here x_{ij} are elements of the inverse matrix $\boldsymbol{X}^{-1}$. The dynamics represented in Eq. (17) can be described by a set of base functions for the L dimensional readout $\boldsymbol{y}$ and their respective coefficients. The minimal set of base functions for Eq. (17) is

$$\begin{aligned}\phi(x) = \{&1, y_1, ..., y_L, y_1^2, y_1y_2, ..., y_1y_L, y_2^2, \\ &y_2y_3, ..., y_2y_L, ..., y_1^3, y_1^2y_2, ..., y_1^2y_L, y_2^3, y_2^2y_1, ..., \\ &y_2^2y_L, ..., y_L^2y_{L-1}\,, y_L^3\,, \Phi_1, ..., \Phi_L\,, e^{-\beta t}\}\end{aligned} \tag{18}$$

where Φ_i is defined as

$$\Phi_i \equiv \int_0^t y_i(\tau)e^{\beta(\tau-t)}\,d\tau.$$

Introducing a new notation for the coefficients of the transformed dynamics one can write Eq. (17) as

$$\begin{aligned}\dot{y}_i &= \tilde{\eta}_i + \tilde{\alpha}_{ij}y_j + \tilde{b}_{ik_1k_2}\,y_{k_1}y_{k_2} + \tilde{c}_{ik_1k_2}\,y_{k_1}y_{k_2}^2 \\ &+ e^{-\beta\,t}\tilde{q}_i - \int_0^t e^{\beta(t-\tau)}\gamma_{ij}\,y_j d\tau + \sqrt{\tilde{D}_{ij}}\xi_j(t),\end{aligned} \tag{19}$$

We notice that the number of base functions N_ϕ for the mixed dynamic is much larger than the number of polynomial terms in Eq. (13)

$$N_\phi = 2 + 2L + \frac{L(L+1)}{2} + L^2\ ,$$

and that increases as L^2 with the number of systems.

Also, due to the symmetries of the system that have been introduced with the substitution of the dynamics, the number of unknown coefficients of the system (17) N_c is much larger than the number of parameters in the original dynamics and $N_c = N_\phi \times L + L^2 + \frac{L(L+1)}{2}$.

From Eq. (19) it is clear that the parameter set

$$\tilde{\mathcal{M}} = \{\tilde{\eta}_i, \tilde{\alpha}_{ij}, \tilde{b}_{ijk}, \tilde{c}_{ijkl}, \tilde{\gamma}_{ij}, \tilde{q}_i(0), \tilde{D}_{ij}\}$$

obtained by inferring the coefficients for the base functions of $\boldsymbol{y}$ differs from the original

$$\mathcal{M} = \{\eta_i, \alpha_i, b_i, c_i, \gamma_i, q_i(0), D_{ij}, X_{ij}\}\,.$$

The explicit relation between the two set of parameters $\tilde{\mathcal{M}}$ and $\mathcal{M}$ can be worked out by comparing Eqs. (16), (17) and (19):

$$\tilde{\eta}_i = X_{ij}\,\eta_j, \tag{20}$$

$$\tilde{\alpha}_{ij} = X_{im}\,\alpha_m\,\left(X^{-1}\right)_{mj}, \tag{21}$$

$$\tilde{b}_{ijl} = X_{im}\,(1+\alpha_m)\,\left(X^{-1}\right)_{mj}\left(X^{-1}\right)_{ml}, \tag{22}$$

$$\tilde{c}_{jklm} = X_{ji}\left(X^{-1}\right)_{ik}\left(X^{-1}\right)_{il}\left(X^{-1}\right)_{im}, \tag{23}$$

$$\tilde{\gamma}_{i\,l} = X_{ij}\,\gamma_j\,\left(X^{-1}\right)_{jl}, \tag{24}$$

$$\tilde{q}_i = X_{ij}\,q_j, \tag{25}$$

$$\sqrt{\tilde{D}}_{ij} = X_{ik}\,\sqrt{D}_{kj}. \tag{26}$$

The inferential algorithm discussed in previous sections is able to select the optimal posterior density in respect of the parameters of the transformed dynamics $\tilde{\mathcal{M}}$. There is no bijective function from $\tilde{\mathcal{M}}$ to $\mathcal{M}$ because we are going from a space larger in dimensions to a smaller one; the selection of the best original parameter cannot be done via algebraic calculations. The problem is a constrained search for the maximum of the posterior distribution within the set of original parameters. In practice, the log-posterior i.e. a multivariate parabolic function, standard non-linear least-squares technique are particularly efficient [29]. The problem of reconstruction of the original parameters $\mathcal{M}$ can be stated in the following terms: we have the explicit transformation rule from $\mathcal{M}$ to $\tilde{\mathcal{M}}$ and we have the explicit density probability in respect of $\tilde{\mathcal{M}}$ (the inferred posterior). We have to find the maximum of the posterior subject to the constraint represented by $\mathcal{M}$. Employing a least-squares algorithm for this task means: (i) to choose an initial evaluation of $\mathcal{M}_0$; (ii) to improve the initial condition in the direction where the probability of the corresponding $\tilde{\mathcal{M}}_0$ is highest; and (iii) repeat the steps till convergence is achieved. In the simulations that we carried out, the choice of the initial $\mathcal{M}_0$ had no influence on the convergence. The initial guess of X is arbitrary set as the unitary matrix. The initial values for α_i can safely be chosen at random between one order of magnitude greater or smaller than the coefficients $\tilde{\alpha}_{ij}$. In an FHN model, the coefficients b_i have no degree of freedom, and the relation $b_i = \alpha_i + 1$ must hold, suggesting the natural initial condition for b_i. An analogous argument suggests that we initialise $c_i = -1$ accordingly to the FHN dynamics. Details of the application of this technique for the recovery of the $\mathcal{M}$ parameter set have been explained in [32]. We note that the redundancy of this approach lets one introduce unknown parameters for the couplings between the FHN systems and the reconstruction of these parameters will bis then a trivial extension.

2.3. Non-stationary dynamics: a stepwise approach

The main idea of our approach consists of considering the parameters to be stationary within reasonably short time-segments, infering parameters for this *step*. For each block of data k the posterior probability density $p_{\text{post}}(\mathbf{c})$ can be used as a prior for the next block $k+1$. Since our posterior probability is a multivariate Gaussian distribution characterized by a mean vector $\mathbf{c}_{\text{post}}$ and a covariance matrix Ξ_{post}^{-1}, it is easy to modify this last quantity to take into account that some parameters, being non-stationary, could have been drifted when considering the next block.

This approximation, where slowly-varying parameters can be assumed constant, will be regarded as an adiabatic approximation. To consider it as a valid starting point, some preliminary considerations are in order.

There are multiple timescales involved in the inferential problem and, when dealing with non-stationary dynamics, the two most significant ones are the timescale that defines the evolution of parameters τ_{param}, and the timescale at which we acquire information through the measurement τ_{meas}. The first one can be regarded as an approximate measure of the biggest change of the dynamical parameters

$$\min_i \left[c_i \left(\frac{\mathrm{d}c_i}{\mathrm{d}t} \right)^{-1} \right] \approx \tau_{\text{param}}$$

and it is an intrinsic property of the system. Conversely, the informational timescale is limited by the sampling frequency h and the number N_{step} of sampling points needed to infer the form of the vector field with the desired precision. Thus

$$h\, N_{\text{step}} \approx \tau_{\text{meas}}.$$

In order to consider parameters to be constant during one step, the timescale condition

$$\tau_{\text{param}} \gg \tau_{\text{meas}}$$

must hold. So the question can be posed in this way: how fast is the parameter convergence, and how small can N_{inf} be for the timescale condition to hold? In the following paragraphs we consider the inference of a set of FHN oscillators as a possible application of this inference in short steps. Because the convergence of parameters is crucial in relation to the first timescale, it will be kept as the central point of the discussion.

3. Simulation results

3.1. Convergence speed for parameter inference

We are mainly interested in the qualitative behavior of convergence of parameters and although what been presented so far applies to any number of FHN oscillators, we now restrict ourselves to a pair of oscillators. We have analysed the convergence of the model

parameters as a function of $T = hN$, where h is the sampling time-step and N is the number of points in a block of data. Synthetic data for the 2 FHN oscillators have been generated using the Heun scheme [31] to integrate the model (13), (14) and the parameters as been chosen as in Table 1. Only the time-series data $y_1(t)$ and $y_2(t)$ have been used as an input for the algorithm. Coefficients of the dynamics in Eq. (19) have been inferred; use of these parameters has been made to reconstruct the original parameters for the dynamics and the coefficients of the mixing matrix. Typical results of the inference of the parameters and the results of the reconstruction of the mixing matrix and the coefficients for the original dynamics are shown in Table 2. It can be seen that convergence of order of 1.5% or lower is achieved within 30000 points, which correspond to less than 1 s at a sampling rate of 35 kHz. To investigate the qualitative behavior of the information gained as a function of number of available data-points we have carried out the following test: we used a data stream 45000 points long and we divided it in 9 blocks of data with 5000 points in each block. The division of the data-stream into several blocks is meant to simulate the learning process of the inferential machine: the posterior information of each block is used as a prior for the following block. The reconstruction of parameters has been done after each block with the information available at that moment. We repeated this procedure for 1000 *runs*: in each run a random data sequence is generated from random initial conditions.It is then inferred and the means of the parameters distribution are recorded. After each block we have a statistical distribution of inferred parameters coming from a different realization of the dynamics. The result of the test is presented in Fig. 2.

Matrix $\hat{\mathbf{\Xi}}$ in Eq.(11) is the inverse of the covariance matrix for parameters in the posterior normal density and gives a measures of how sharply-peaked this distribution is about its mean value. For a qualitative comparison between the parameters' convergence and the evolution of the biggest eigenvalues λ_i of $\hat{\mathbf{\Xi}}^{-1}$ see [30, 32].

Next, we consider the efficiency of the method under non-stationary conditions.

3.2. Non-stationary dynamics

Let us assume that we have to deal with some non-stationary parameters. In particular we will take η_i to be the varying parameters, keeping all the others fixed. To show capability in dealing with this situation we perform three different tests. In the first test the η_1 and η_2 change at random and in a step-like manner, and no other information

$\alpha_1 = 0.35$	$\eta_1 = 0.4$	$d_{11} = 0.0002$	$d_{12} = 0.00007$
$\alpha_2 = 0.20$	$\eta_2 = 0.3$	$d_{22} = 0.0002$	$d_{21} = 0.00007$
$\gamma_1 = 0.0153$	$\beta = 0.0151$	$x_{11} = 1.7$	$x_{12} = 0.8$
$\gamma_2 = 0.0153$		$x_{22} = 0.2$	$x_{21} = 0.9$

Table 1. Parameter values of the model (13), (14) used to generate stationary time-series data.

parameter	real	inferred	% rel. error
$\tilde{\eta}_1$	0.9200	0.924384	0.476522
$\tilde{\eta}_2$	0.3500	0.351001	0.28600
$\tilde{b}_{222}$	1.7550	1.758011	0.171567
$\tilde{b}_{112}$	-2.1086	-2.114731	0.290762
X_{11}	1.7	1.686459	0.796526
X_{12}	0.8	0.794263	0.717092
X_{21}	0.2	0.196746	1.626811
X_{22}	0.9	0.898222	0.197610
η_1	0.4	0.406227	1.556788
η_2	0.3	0.302462	0.820660
α_1	-0.35	-0.351992	0.569082
α_2	-0.2	-0.200376	0.188228
b_1	1.35	1.357427	0.550145
b_2	1.2	1.203863	0.321885
c_1	-1.0	-0.999520	0.047957
c_2	-1.0	-0.999114	0.088582

Table 2. Inferred values of some of the mixed dynamics coefficients (first four lines), and values of some of the reconstructed parameters. The inference is based on 30000 points. The actual values (second column) are compared with the inferred values (third column). Relative errors are given in the last column.

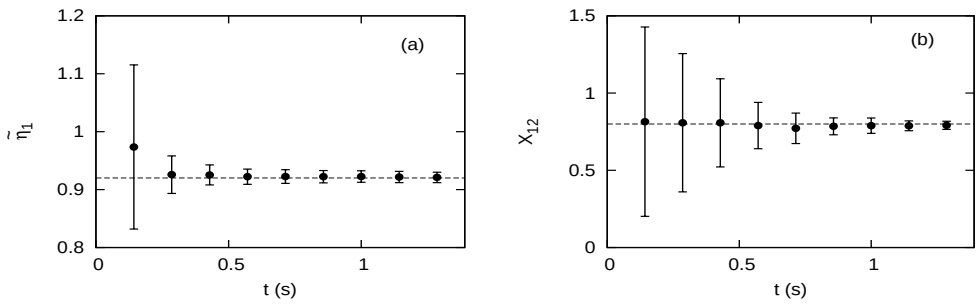

Figure 2. Typical example of parameter convergence and reconstruction of original coefficients as a function of signal length. Convergences of parameter $\tilde{\eta}_1$ (a) and of the reconstruction of an element of the mixing matrix (b) are shown. The first point corresponds to a block of 5000 data points; each successive point after that corresponds to an additional 5000 data, as discussed in the text. Vertical bars show the standard deviations of the inferred values, calculated over 1000 realizations.

is assumed; in the second test we demonstrate that inference of stepwise change of η_i are detected much faster if we assume knowledge of the other parameters; and in the third test η_1 and η_2 are allowed to change continuously and information about the other parameters is assumed to be limited. In discussing these examples we do not aim to cover all possible cases that might arise. Rather we selected these three tests in the

hope of providing an overall view of how to apply the present algorithm and of how it behaves under different circumstances.

First test Parameters η_1 and η_2 change at random in time in a step-like manner, and remain constant between steps. The time interval between steps is approximately 5 periods of firing of the action potential. It contains one block of data with 20000 points, corresponding to approximately 0.55 s. Other parameters of the model are fixed but assumed *unknown* (the actual values are given in Table 1). At each step we infer all the parameters. Their initial values are assumed to be zero and their initial dispersion to be infinity as already discussed above. Fig. 3 presents the inference results.

Second test Parameters η_1 and η_2 change at random as in the previous test, but now the other parameters of the model are fixed at *known* values. The time interval between steps is only 1000 points, corresponding to approximately 0.03 s. At each step we infer only parameters η_1 and η_2. From Fig. 4 it is clear that when other parameters are known, the time required for inferring η_i is smaller by two orders of magnitude than in the case when all parameters have to be inferred.

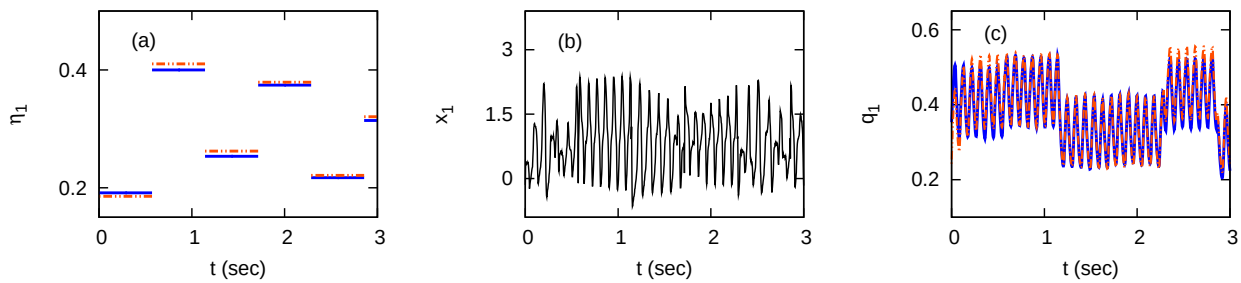

Figure 3. Inference of the parameters of two uncoupled FHN systems mixed by the measurement matrix. It is assumed that η_1 and η_2 change step-wise while all other parameters of the system are fixed and unknown. (a) The inferred values of η_1 (dashed red lines) are compared with their true values (full blue lines). (b) Measured mixed values of the coordinate $x_1(t)$. (c) Inferred values of the coordinate $q_1(t)$ (red dotted line) are compared with its true values of (blue solid line). The other parameters are fixed at the values given in Table 1. The noise amplitude was $\sqrt{d_1} = \sqrt{d_2} = .01225$.

Third test In our final test, we infer smoothly varying parameters η_1 and η_2 with added noise, without knowing any other parameters of the model. The partial information on parameters is simulated by inferring parameters from the first block (with 30000 points) of stationary dynamics; then for all other blocks of data we use acquired information to fix the model parameters constant at their inferred values, and we track in time only variations of the control parameters η_i. Each block of data (except the first one) contains 12000 points and has a time length of $t \approx 0.34$ sec. The inferred time evolution of the control parameters η_i is compared with its true variation in Fig. 5. It is evident from the figure that the method allows us to infer the unknown constant parameters

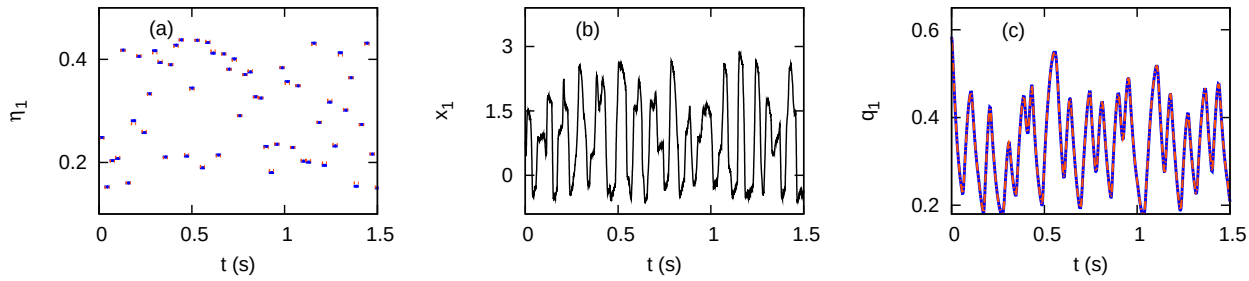

Figure 4. Inference of the model parameters of two uncoupled FHN systems mixed by the measurement matrix. It is assumed that η_1 and η_2 change step-wise while all other parameters of the system are fixed and known. (a) Inferred values of η_1 (short elements of red dashed line) are compared with their true values (short elements of full blue line) as a function of time. (b) The time-trace of the measured coordinate $x_1(t)$. (c) The time-trace of the inferred coordinate $\tilde{q}_1(t)$ (red dotted line) is compared with its true value $q_1(t)$ (blue solid line). The values of the other parameters are fixed, as given in Table 1. The noise amplitude was $\sqrt{d_1} = \sqrt{d_2} = .01225$.

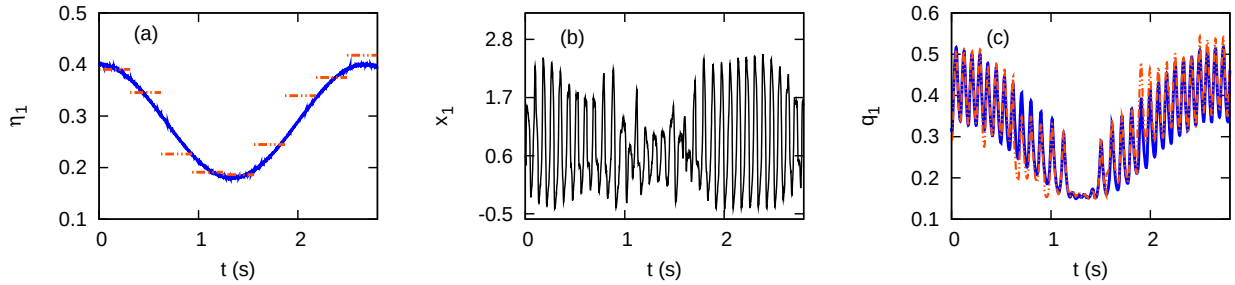

Figure 5. Inference of η_1 and η_2, while smoothly varying in the presence of noise. No prior knowledge of the model parameters is assumed. (a) The inferred values of η_1 (dashed red lines) are compared with their true values (full blue lines). (b) The measured time-trace of the mixed coordinate $x_1(t)$. (c) The inferred time-trace of the mixed coordinate $\tilde{q}_1(t)$ (dashed red line) is compared with its true value $q_1(t)$ (full blue line). The values of the other parameters are given in Table 1. The noise amplitude was $\sqrt{d_1} = \sqrt{d_2} = .01225$.

of the model, and then also to use this information to track in time the non-stationary control parameters of the system with a time resolution of the order of 0.3 sec.

4. Conclusion

The inferential framework presented above is an extended version of the that proposed in [15] . It is based on Bayesian statistics and a path-integral formulation of stochastic nonlinear dynamics, and it allows reconstruction of the parameters of the dynamical and measurement models from noise-corrupted time series data with subsequent fast tracking of time-varying control parameters. The presented algorithm is open to modification: *ad hoc* external information can always be added without changing the theoretical framework.

The convergence speed of the method is an important issue for fast-tracking of the parameters. We have discussed the convergence of the method in the parameter space which is important issue in this context. We also we provided an overview of how the reconstruction machine might be modified, depending on availability of system information.

We tested the performance of our scheme on simulations of physiological signals, modeling the action potentials of an array of neurons as a set of L noisy FHN oscillators. The readout was assumed mixed by an unknown measurement matrix. We have established that the method does indeed facilitate on-line tracking of key control parameters with a suitable time resolution. The results are achieved by embedding the fast on-line tracking of the control parameters within a Bayesian learning framework for the more slowly varying coefficients of the system. To simplify the analysis of performance, we neglected measurement noise, but suitable plug-in techniques exist for dealing with this extra noise source without any alteration of the core method. The use of our inferential framework is very practical and multidisciplinary. With *ad hoc* adjustments on a per-case base, its natural employment is within any field where inferential capabilities are needed for a non-linear dynamical system.

References

[1] H. Konno and K. Hayashi, Annals of Nuclear Energy **23**, 35 (1996).
[2] J. Christensen-Dalsgaard, Rev. Mod. Phys. **74**, 1073 (2002).
[3] V. N. Smelyanskiy, D. G. Luchinsky, A. Stefanovska, and P. V. E. McClintock, Phys. Rev. Lett. **94**, 098101 (2005).
[4] E. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* (MIT Press, Cambridge, MA, 2006).
[5] S. Siegert, R. Friedrich, and J. Peinke, Physics Letters A **243**, 275 (1998).
[6] P. E. McSharry and L. A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).
[7] R. Friedrich *et al.*, Phys. Lett. A **271**, 217 (2000).
[8] J. P. M. Heald and J. Stark, Phys. Rev. Lett. **84**, 2366 (2000).
[9] R. Meyer and N. Christensen, Phys. Rev. E **62**, 3535 (2000).
[10] R. Meyer and N. Christensen, Phys. Rev. E **65**, 016206 (2001).
[11] J.-M. Fullana and M. Rossi, Phys. Rev. E **65**, 031107 (2002).
[12] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke, Europhys. Lett. **61**, 466 (2003).
[13] F. Watanabe, H. Konno, and S. Kanemoto, Annals of Nuclear Energy **31**, 375 (2004).
[14] W. Enders, *Applied Econometric Time Series, 2nd Edition* (Wiley, Hoboken, NJ, 2004).
[15] V. N. Smelyanskiy, D. G. Luchinsky, D. A. Timucin, and A. Bandrivskyy, Phys. Rev. E **72**, 026202 (2005).
[16] V. N. Smelyanskiy, D. G. Luchinsky, A. Stefanovska, and P. V. E. McClintock, Phys. Rev. Lett. **94**, 098101 (2005).
[17] R. W. Z. Turner, K.L.; Baskaran, Decision and Control, 2003. Proceedings. 42nd IEEE Conference on **3**, 2650 (9-12 Dec. 2003).
[18] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran, Neuron **52**, 205 (2006).
[19] V. V. Osipov, D. G. Luchinsky, V. N. Smelyanskiy, and D. A. Timucin, *Proc. AIAA/ASME/SAE/ASEE Joint Propulsion Conf. and Exhibit*, *AIAA Conference Proceedings* (AIAA, Cincinnati, OH, 2007), AIAA 2007-5823 (2007).

[20] D. G. Luchinsky et al, *Proc. AIAA Infotech@Aerospace 2007 Conf. and Exhibit*, *AIAA Conference Proceedings* (AIAA, Robnert Park, CA, 2007), AIAA 2007-2829 (2007).
[21] R. FitzHugh, Biophys. J. **1**, 445 (1961).
[22] J. Nagumo, S. Animoto, and S. Yoshizawa, Proc. Inst. Radio Engineers **50**, 2061 (1962).
[23] A. T. Winfree, *The Geometry of Biological Time* (Springer-Verlag, New York, 1980).
[24] E. V. Pankratova, A. V. Polovinkin, and B. Spagnolo, Phys. Lett. A **344**, 45–50 (2005).
[25] D. G. Luchinsky, V. N. Smelyanskiy, and J. Smith, in *Unsolved Problems of Noise and Fluctuations*, Vol. 800 of *AIP Conference Proceedings*, ed. by L. Reggiani *et al.* (AIP, Melville, NY, 2005), pp. 539–545.
[26] C. Calder, M. Lavine, P. Müller, and J. S. Clark, Ecology **84**, 1395 (2003).
[27] A. Apte, M. Hairer, A. M. Stuart, and J. Voss, Physica (Utrecht) **D230**, 50 (2007).
[28] D. G. Luchinsky, V. N. Smelyanskiy, M. Millonas, and P. V. E. McClintock, in *Noise in Complex Systems and Stochastic Dynamics III*, Vol. 5845 of *(SPIE)*, ed. L. B. Kish et al (SPIE, Bellingham, WA, 2005), pp. 173–181.
[29] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2 edn. (Cambridge University Press, 1992).
[30] D. G. Luchinsky et al, Phys. Rev. E **77**, 061105 (2008).
[31] R. Mannella, Intern. J. Mod. Phys. C **13**, 1177 (2002).
[32] A. Duggento et al, Phys. Rev. E **77**, 061106 (2008).