**THE EUROPEAN
PHYSICAL JOURNAL
SPECIAL TOPICS**

Review

A tutorial on time-evolving dynamical Bayesian inference

Tomislav Stankovski[1], Andrea Duggento[2], Peter V.E. McClintock[1],
and Aneta Stefanovska[1,a]

[1] Department of Physics, Lancaster University, Lancaster, LA1 4YB, United Kingdom
[2] Medical Physics Section, Faculty of Medicine, Tor Vergata University, Rome, Italy

Abstract. In view of the current availability and variety of measured data, there is an increasing demand for powerful signal processing tools that can cope successfully with the associated problems that often arise when data are being analysed. In practice many of the data-generating systems are not only time-variable, but also influenced by neighbouring systems and subject to random fluctuations (noise) from their environments. To encompass problems of this kind, we present a tutorial about the dynamical Bayesian inference of time-evolving coupled systems in the presence of noise. It includes the necessary theoretical description and the algorithms for its implementation. For general programming purposes, a pseudocode description is also given. Examples based on coupled phase and limit-cycle oscillators illustrate the salient features of phase dynamics inference. State domain inference is illustrated with an example of coupled chaotic oscillators. The applicability of the latter example to secure communications based on the modulation of coupling functions is outlined. MatLab codes for implementation of the method, as well as for the explicit examples, accompany the tutorial.

1 Introduction

Science can be defined as the systematic observation and analysis of nature, and of how natural processes evolve in time and space. Typically, science tries to generate models based on data collected over some specified time interval – one that depends on how fast the processes occur, which may be over e.g. centuries or seconds or microseconds. The data are usually measured in streams, and they are analyzed in sequential blocks. Depending on the dynamical characteristics, different methods of data analysis may be suitable, including for example Bayesian methods, Kalman filters, particle filters and maximum likelihood estimators [1–6].

The Kalman filter represents a major milestone in the history of such algorithms. It provides a way of predicting a particle's motion, given a noisy data stream representing its position. The algorithm is named after Rudolf E. Kalman [7] who, in the sixties, brought it to the attention of the NASA researchers facing the challenges

---

[a] e-mail: `aneta@lancaster.ac.uk`

raised by the Apollo space program. Although based on linear functions, the Kalman filter was an immediate success. The method has been much improved over the last decades, including the introduction of non-linear Kalman filters [8].

Another important class of sequential data analysis methods is those based on Bayesian theory [9]. They exploit Bayes' theorem to make use of prior knowledge in estimating the current or posterior model parameters. In this way, the Bayesian methods use the evolution of acquired knowledge to accomplish improved inference. This review is focused on some recent developments and advances in this rapidly evolving area, based on Bayesian methods [1,10–12].

There are many different ways in which Bayes' theorem may be applied, and different groups of methods exist for performing so-called "Bayesian inference" [3,4,12–15]. Of particular interest among these are the Bayesian methods for dynamical inference. They enable a dynamical model in terms of ordinary or stochastic differential equations to be inferred from the observed data. Dynamical Bayesian methods provide the basis for important signal processing techniques that have been applied to e.g. physics, biology, communications, and climate [3,11,12,14].

A great advantage of these Bayesian methods is their ability to infer dynamics when the system under consideration is not isolated, but is influenced by its environment and other processes to which it may be weakly coupled [16–19]. One manifestation of such external influence occurs when the underlying dynamical systems are subject to noise, and it has already been studied in detail theoretically [1,10,12,20–22]. When dealing with data from natural systems one should allow for the possibility of the dynamics being time-varying. In such cases, the Bayesian method [1,10] published recently is of particular interest: it can identify time-varying dynamics even in the presence of noise, and it is able to follow the time-evolution of the parameters. It is the latter method that provides the main focus of this tutorial.

The method can be applied to various types of dynamical system. In what follows, however, we will focus our attention on coupled oscillatory dynamical systems. The latter frequently arise in physiology and include, for example, the cardio-respiratory or neuronal systems [23–26] which are, in fact, coupled dynamical systems with time-varying parameters that are subject to external noise. In that context, one might be interested in e.g. detecting causal interactions and the directionality of influence [27–30], or in coherence and synchronization [11,31–33]. Of special interest is the recently developed detection and description of interactions in terms of *coupling functions* [1,6,34]. Dynamical Bayesian inference has already been used to investigate how the cardiorespiratory coupling functions are affected by aging [35]. The same method has also been applied in a quite different context, namely, to facilitate the use of inter-oscillator coupling functions to improve the security of communications systems [11].

In the tutorial presented below, we discuss the practicalities of how this particular Bayesian method can be implemented, including the algorithms, programming and applications. The problems and phenomena that were originally treated by the method [1,10,12,21,22] were relatively complex and had to be presented in a rather compressed format. Given that the tutorial is intended for a wider audience, we adopt here a different presentation style using examples that are simple and an exposition that is quite detailed. The tutorial is organized as follows. First, we start by summarising the background and basics of Bayesian probability as proposed by Thomas Bayes, the originator of this inference approach. Then in Sect. 3 we introduce the theoretical terms needed for the implementation of dynamical Bayesian inference. The algorithms and the programming description are given in Sect. 4. Applications of the method to three examples are discussed in Sect. 5. The first of these uses coupled phase oscillators to present the basics of the inference of time-evolving phase dynamics in the presence of noise. The second example uses coupled limit-cycle oscillators, also

describing the reconstruction and detection of synchronization and coupling functions. The third and final example discusses the inference of coupled chaotic systems in state space, as an implementation of the secure encryption technique. Finally, in Sect. 6, we discus possible generalizations of the method, consider the implications for other areas, point out the relationships to other methods, and offer some concluding remarks.

2 The legacy of Thomas Bayes (1701–1761)

It was fortunate for Thomas Bayes' legacy that his friend Richard Price significantly edited and updated his work, and read it posthumously to the Royal Society on his behalf in 1763.

It was published in *Philosophical Transactions of the Royal Society of London* the following year. The ideas gained only limited exposure until they were independently rediscovered and further developed by Laplace, who first published their modern formulation in his 1812 *Théorie analytique des probabilités*.

The classical approach to statistics defines the probability of an event as "The number of times the event occurs over the total number of trials, in the limit of an infinite series of equiprobable repetitions". Many of the limitations inherent in this definition can be avoided, and paradoxes resolved, by taking a Bayesian stance about probabilities. Bayes defines probability as:

> "The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening."

However even Bayes himself might not have embraced the broad interpretation now referred to as Bayesian. It is difficult to assess Bayes' philosophical views on probability, because his work does not go into questions of interpretation.

Today Bayesian probability is used to describe several different, but related, interpretations of probability. To evaluate the probability of a hypothesis, Bayesian probability specifies some prior probability, which is then updated in the light of new, relevant data. "Bayesian" has been used in this sense since the rebirth of Bayes' ideas in the 20th century. Advances in computer technology have allowed scientists from many disciplines to extend the approach and to apply it in diverse fields. Sir Harold Jeffreys, who wrote the book *Theory of Probability*, which first appeared in 1939, played an important role in the revival of the Bayesian view of probability. He wrote that Bayes' theorem "is to the theory of probability what Pythagoras's theorem is to geometry".

So what exactly is Bayes' theorem? Scientific hypotheses are typically expressed through probability distributions for observable data $\mathcal{X}$ which depend on the model parameters $\mathcal{M}$. In the Bayesian framework, current knowledge about the model parameters is expressed by placing a probability distribution on the parameters, called the "prior distribution", often written as $p_{\mathrm{prior}}(\mathcal{M})$. When new data $\mathcal{X}$ become available, the information they contain regarding the model parameters is expressed in the "likelihood," which is proportional to the conditional distribution of the observed data given the model parameters $\ell(\mathcal{X}|\mathcal{M})$. This information is then combined with the prior to produce an updated probability distribution called the "posterior distribution," on which all Bayesian inference is based. Bayes' theorem, an elementary identity in probability theory, states how the update is done mathematically – the posterior is proportional to the prior times the likelihood, over the whole available parameter space:

$$p_{\mathcal{X}}(\mathcal{M}|\mathcal{X}) = \frac{\ell(\mathcal{X}|\mathcal{M})p_{\mathrm{prior}}(\mathcal{M})}{\int \ell(\mathcal{X}|\mathcal{M})\,p_{\mathrm{prior}}(\mathcal{M})d\mathcal{M}}. \tag{1}$$

From this point of view, the task to be faced in Bayesian analysis is to construct and evaluate the likelihood function, given that the data and the prior knowledge are available.

This single line of Bayes' theorem appears simple, but has a rather profound meaning. It gives one a means of reversing the problem of inference – starting from observations to get back to the nature of the causation. Even better for practical applications is that it works with probabilities, expressing "beliefs" or the level of how likely something is to happen. It is a powerful method that provides natural ways for people in many disciplines to structure their data and knowledge, and to yield direct and intuitive answers to their practical questions.

Today, Bayesian theory enjoys wide interpretation and application over most of science and experimental areas quite generally. Since Bayes first propounded his ideas, there have been huge developments in both the theory and applications. The latter span practically every aspect of science, including particle physics, astrophysics, cosmology, geophysics, communications, pharmacology and medical and biological physics [1,3,4,11–13,36,37]. There are entire societies, such as the *International Society for Bayesian Analysis* (ISBA), special conferences are organized on Bayesian analysis, and there exist numerous journals specialized in Bayesian theory and analysis. Given the current level of activity in the field, Bayesian inference promises to become even more useful and involved in solving the great scientific and everyday problems faced by humanity.

3 Dynamical Bayesian inference

In the present context, dynamical inference refers to a procedure for inferring a model in terms of differential equations based on the analysis of a time-series. The method resembles Feynman's path integral (which is itself based on a Wiener process), whose central idea is that, for the motion of a particle between two points in space, all possible connecting trajectories should be considered and a probability amplitude assigned to each one of them. This path integral gives the likelihood of observation of a dynamical trajectory for a given set of distributions of the model parameters. Once the actual dynamical trajectory is measured in the experiment, the distributions for the set of model parameters can be improved by use of Bayes' theorem. The other main feature, characteristic of the method, is that the Bayesian framework is applied to a model whose deterministic part is allowed to be time-varying.

The aim is to provide a method that can infer a model of two (or more) weakly-interacting systems subject to noise:

$$\dot{\boldsymbol{\chi}}_i = \mathbf{f}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j | \mathbf{c}) + \sqrt{\mathbf{E}}\xi_i, \tag{2}$$

where $i \neq j = 1, 2$, and $\mathbf{f}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j | \mathbf{c})$ is a vector of functions representing the deterministic part of the internal and the interacting dynamics. The vector $\mathbf{c}$ denotes the parameters of the model. The dynamical noise is assumed to be white, Gaussian, and parameterized by a noise diffusion matrix ($\mathbf{E}$ 2×2): $\langle \xi_i(t)\xi_j(\tau) \rangle = \delta(t - \tau)E_{ij}$. By $\sqrt{\mathbf{E}}$ we indicate the Cholewsky decomposition of $\mathbf{E}$. In what follows we exploit the Bayesian method presented in [1,10,12,21,22], and the readers interested in additional theoretical details are directed to those papers and to the references therein.

At this point we speak of $\boldsymbol{\chi}_i$ in general, but later we will refer separately to the phase or state domain, depending on the type of data that we are inferring. When we apply the method to analyze different systems, the base functions are the only thing that will change in the inferential framework. Here, we will use polynomial base

functions for the state domain, while for the phase domain we will decompose the dynamics into Fourier components:

$$\dot{\phi}_i = \sum_{k=-K}^{K} c_k^{(i)} \, \phi_{i,k}(\phi_i, \phi_j) + \sqrt{\mathbf{E}}\xi_i, \tag{3}$$

where $\mathbf{\Phi}_{1,0} = \mathbf{\Phi}_{2,0} = 1$, $c_0^{(i)} = \omega_i$ are the respective frequencies, and the rest of the $\mathbf{\Phi}_{i,k}$ and $c_k^{(i)}$ are the K most important Fourier components serving as base functions.

Given that $2 \times N$ time-series $\mathcal{X} = \{\boldsymbol{\chi}_n \equiv \boldsymbol{\chi}(t_n)\}$ $(t_n = nh)$ are provided, and assuming that the model base functions are known, the main task for dynamical Bayesian inference [12,21] is to infer the unknown model parameters and the noise diffusion matrix $\mathcal{M} = \{\mathbf{c}, \mathbf{E}\}$. The problem eventually reduces to maximization of the posterior conditional probability $p_{\mathcal{X}}(\mathcal{M}|\mathcal{X})$ of observing the parameters $\mathcal{M}$, given the data $\mathcal{X}$. The relationship of the posterior probability $p_{\mathcal{X}}(\mathcal{M}|\mathcal{X})$ to the prior distribution $p_{\text{prior}}(\mathcal{M})$ and the likelihood function $\ell(\mathcal{X}|\mathcal{M})$ is defined by Bayes' theorem, given above as (1). The maximization of the posterior probability is a difficult task in its own right but we will see below how, with a parameterization of the model, we can obtain a posterior probability density that is easily maximized.

The prior distribution, enclosing previous knowledge of the unknown parameters based on observations, is assumed to be known. The task is therefore to determine the likelihood functions in order to infer the final posterior result. If the sampling h is small enough, using the acquired time-series one can construct the Euler midpoint approximation of Eqs. (2):

$$\boldsymbol{\chi}_{i,n+1} = \boldsymbol{\chi}_{i,n} + h\mathbf{f}(\boldsymbol{\chi}_{i,n}^*, \boldsymbol{\chi}_{j,n}^*|\mathbf{c}) + h\sqrt{\mathbf{E}}\mathbf{z}_n, \tag{4}$$

where $\boldsymbol{\chi}_n^* = (\boldsymbol{\chi}_{n+1} + \boldsymbol{\chi}_n)/2$ and $\mathbf{z}_n$ is the stochastic integral of the noise term over time: $\mathbf{z}_n \equiv \int_{t_n}^{t_{n+1}} \mathbf{z}(t)\,dt = \sqrt{h}\,H\,\boldsymbol{\xi}_n$ for the H matrix that satisfies the Cholesky decomposition $HH^T = \mathbf{E}$. The parameters $\mathbf{c}$ act as a scale coefficients for the base functions $\mathbf{P}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)$: $\mathbf{f}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j|\mathbf{c}) = \mathbf{c}\mathbf{P}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)$. The idea here is to let $\mathbf{P}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)$ be arbitrary and non-linear with respect to the dynamical variables, while assuming that the function $\mathbf{f}$ can be linearly parameterized with respect to the vector of parameters $\mathbf{c}$. Use of the stochastic integral for noise that is white and independent leads to a likelihood function that is given by the product over n of the probability of observing $\boldsymbol{\chi}_{n+1}$ at each time. The negative log-likelihood function is then $S = -\ln \ell(\mathcal{X}|\mathcal{M})$ given as:

$$S = \frac{N}{2}\ln|\mathbf{E}| + \frac{h}{2}\sum_{n=0}^{N-1} \left(\mathbf{c}\frac{\partial \mathbf{P}(\boldsymbol{\chi}_{\cdot,n})}{\partial \boldsymbol{\chi}} + [\dot{\boldsymbol{\chi}}_n - \mathbf{c}\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*)]^T(\mathbf{E}^{-1})[\dot{\boldsymbol{\chi}}_n - \mathbf{c}\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*)]\right), \tag{5}$$

where $\dot{\boldsymbol{\chi}}_n = (\boldsymbol{\chi}_{n+1} - \boldsymbol{\chi}_n)/h$ and the dot index in $\boldsymbol{\chi}_{\cdot,n}$ represents the appropriate (i or j in this case) index. The likelihood (5) is of quadratic form and, if the prior is a multivariate normal distribution, so also will be the posterior. Given such a distribution as a prior for the parameters $\mathbf{c}$, with mean $\bar{\mathbf{c}}$, and covariance matrix Σ_{prior} (or its inverse, known as the concentration matrix $\Xi_{\text{prior}} \equiv \Sigma_{\text{prior}}^{-1}$), the final stationary point of S is calculated recursively from the following four equations:

(a) the noise matrix $\mathbf{E}$

$$\mathbf{E} = \frac{h}{N}\left(\dot{\boldsymbol{\chi}} - \mathbf{c}\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*)\right)\left(\dot{\boldsymbol{\chi}}_n - \mathbf{c}\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*)\right)^T, \tag{6}$$

(b) the concentration matrix Ξ

$$\Xi = \Xi_{\text{prior}} + h\,\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*)^T(\mathbf{E}^{-1})\,\mathbf{P}(\boldsymbol{\chi}_{\cdot,n}^*), \tag{7}$$

(c) a temporary matrix variable $\mathbf{r}$

$$\mathbf{r} = \boldsymbol{\Xi}_{\mathrm{prior}}\,\mathbf{c} + h\,\mathbf{P}(\boldsymbol{\chi}^*_{.,n})\,(\mathbf{E}^{-1})\,\dot{\boldsymbol{\chi}}_n - \frac{h}{2}\mathbf{v}(\boldsymbol{\chi}^*_{.,n}), \qquad (8)$$

where the components of matrix $\mathbf{v}$ are the partial derivatives of the base functions
$\mathbf{v}(\boldsymbol{\chi}^*_{.,n}) = \frac{\partial \mathbf{P}(\boldsymbol{\chi}^*_{.,n})}{\partial \boldsymbol{\chi}_.}$,

(d) and the final parameters $\mathbf{c}$

$$\mathbf{c} = \boldsymbol{\Xi}^{-1}\mathbf{r}, \qquad (9)$$

where summation over $n = 1, \ldots, N$ is assumed. The first initial prior can be set as a non-informative flat normal distribution, $\boldsymbol{\Xi}_{\mathrm{prior}} = 0$ and $\bar{\mathbf{c}}_{\mathrm{prior}} = 0$.

By evaluating the four equations (6)-(9) using the readout time series $\mathcal{X}$, one can calculate effectively the multivariate probability $\mathcal{N}_{\mathcal{X}}(\mathbf{c}|, \bar{\mathbf{c}}, \boldsymbol{\Xi})$ which explicitly defines the probability density of each parameter set of the dynamical system.

The inference method needs to follow the time-evolution of the parameter set $\mathbf{c}$ while separating dynamical effects from the noise. In order to achieve this, we modify the propagation procedure between the covariance of the current posterior $\boldsymbol{\Sigma}^n_{\mathrm{post}}$ and the next prior $\boldsymbol{\Sigma}^{n+1}_{\mathrm{prior}}$ [1]. The definite matrix $\boldsymbol{\Sigma}_{\mathrm{diff}}$ is introduced in order to show how much each parameter diffuses normally. Thus, the next prior probability of the parameters is the convolution of two current normal multivariate distributions, $\boldsymbol{\Sigma}_{\mathrm{post}}$ and $\boldsymbol{\Sigma}_{\mathrm{diff}}$: $\boldsymbol{\Sigma}^{n+1}_{\mathrm{prior}} = \boldsymbol{\Sigma}^n_{\mathrm{post}} + \boldsymbol{\Sigma}^n_{\mathrm{diff}}$. To avoid propagation in the assumptions about correlation between parameters, we consider $\boldsymbol{\Sigma}_{\mathrm{diff}}$ to be diagonal ($\rho_{ij} = \delta_{ij}$ in [1]). We assume each standard deviation σ_i to be a known fraction of the relevant standard deviation from the posterior covariance (or parameters) $\sigma_i = p_w(\sigma^n_{\mathrm{post}})_i$, where p_w is a constant parameter. In practice this means that $\boldsymbol{\Sigma}_{\mathrm{diff}}$ has zero values everywhere, except for the diagonal values, which are a fraction of the diagonal values of the posterior $\boldsymbol{\Sigma}_{\mathrm{post}}$.

4 Algorithms and programming

In this section we discuss the algorithmic and programming details needed for the implementation. We start by presenting what is arguably the most complicated part – the algorithm for dynamical Bayesian inference applied within a single window of readout data. The algorithm employing recursion using the Eqs. (6)–(9) can be summarized in terms of the following steps:

    i) the algorithm starts from a $\mathbf{c}_{\mathrm{prior}}$ and $\boldsymbol{\Xi}_{\mathrm{prior}}$,
    ii) noise matrix $\mathbf{E}_{\mathrm{new}}$ is calculated using Eq. (6),
    iii) $\boldsymbol{\Xi}_{\mathrm{new}}$ is calculated using Eq. (7),
    iv) $\mathbf{r}$ is calculated using Eq. (8),
    v) $\mathbf{c}_{\mathrm{new}}$ is calculated using Eq. (9),
    vi) then again to point ii) using $\mathbf{c}_{\mathrm{new}}$ as $\mathbf{c}$.

The stopping rule is that "convergence" has been reached i.e. when further iteration of the algorithm would not modify $\mathbf{c}$ and $\boldsymbol{\Xi}$ any more. For example, we used the condition: $\sum (\mathbf{c}_{old} - \mathbf{c}_{new})^2/\mathbf{c}^2_{new} < \epsilon$ where ϵ is some very small constant. Because the problem is parabolic, this convergence is very fast – typically a few cycles. The initial prior distribution is assumed to be a noninformative "flat" distribution, representing the initial limit of an infinitely large normal distribution, and obtained by setting $\Xi_{\mathrm{prior}} = 0$ and $\mathbf{c}_{\mathrm{prior}} = 0$.

For general programming purposes, we now outline an informal pseudo-code description of the main algorithms and sub-algorithms. Comments are presented in grey.

First we describe the algorithm for Bayesian inference:

Algorithm 1: **Bayesian inference**

\\calculate temporary variables beforehand
 – calculate $\mathbf{P}$
 – calculate $\mathbf{v}$

$\mathbf{c}_{\mathrm{pt}} = \mathbf{c}_{\mathrm{pr}}$
FOR lp=1:MaxLoops \\main recursive loop
 – calculate $\mathbf{E}$
 – calculate $\mathbf{c}_{\mathrm{pt}}$
 IF $\mathrm{SUM}((\mathbf{c}_{\mathrm{pr}} - \mathbf{c}_{\mathrm{pt}})^2/\mathbf{c}_{\mathrm{pt}}^2) < \mathrm{SmallError}$
  RETURN
 ENDIF
 $\mathbf{c}_{\mathrm{pr}} = \mathbf{c}_{\mathrm{pt}}$
ENDFOR

The sub-algorithms "*calculate P*" and "*calculate v*" depend on the particular base functions and their partial derivatives. These two functions depend on the specific model to be inferred, and they are the only part needing change if one selects a different model for the inference. The sub-algorithms only involve the simple evaluation of base functions in respect of the χ^* time series, and will not be discussed in detail. The other main calculations are performed within "*calculate $\mathbf{E}$*" and "*calculate $\mathbf{c}_{\mathrm{pt}}$*", which are discussed in detail below.

Algorithm 2: **calculate E** \\use of Eq. (6)

$\mathbf{E} = \mathbf{E}+(\dot{\boldsymbol{\chi}}\text{-}\mathbf{P}^*\mathbf{c})^*(\text{transpose of } (\dot{\boldsymbol{\chi}}\text{-}\mathbf{P}^*\mathbf{c}))$
$\mathbf{E} = \mathrm{h/N}^*\mathbf{E}$

Finally, the algorithm for calculation of the parameters used in the main recursive loop is expressed as:

Algorithm 3: **calculate c**
invE = inverse of $\mathbf{E}$

\\calculate $\boldsymbol{\Xi}$, Eq. (7)
FOR i=1:l
 FOR j=1:l
  $\boldsymbol{\Xi}_{\mathrm{pt}}((i-1) \cdot K + 1 \text{ TO } i \cdot K, (j-1) \cdot K + 1 \text{ TO } j \cdot K) =$
  $\boldsymbol{\Xi}_{\mathrm{pr}}((i-1) \cdot K + 1 \text{ TO } i \cdot K, (j-1) \cdot K + 1 \text{ TO } j \cdot K)+\mathrm{h}^*\mathbf{invE}(i,j)^*\mathbf{P}^*(\text{transpose of } \mathbf{P})$
 ENDFOR
ENDFOR

\\calculate $\mathbf{r}$, Eq. (8)
ED=invE$^*\dot{\boldsymbol{\chi}}$
FOR i=1:l

FOR j=1:l
   $\mathbf{r}(\text{ALL}, i) = \mathbf{r}(\text{ALL}, i) + \mathbf{\Xi}_{\text{pr}}((i-1) \cdot K + 1 \text{ TO } i \cdot K, (j-1) \cdot K + 1 \text{ TO } j \cdot K) * \mathbf{c}(\text{ALL}, j)$
ENDFOR
  $\mathbf{r}(\text{ALL}, i) = \mathbf{r}(\text{ALL}, i) + \text{h} \cdot \mathbf{P} * (\text{transpose of } \mathbf{ED}) - \text{h}/2 \cdot \text{sum}(\mathbf{v}(\text{ALL}, i))$
ENDFOR

\\calculate $\mathbf{c}$, Eq. (9)
$\mathbf{c} = (\text{inverse of } \mathbf{\Xi}_{\text{pt}}) * \mathbf{r}$

---

where l is the number of time-series provided, and $K = M/l$ with M being the total number of base function used. The operators $*$ and $\cdot$ refer to matrix multiplication and scalar multiplication, respectively. For the calculations within the matrixes, one should interpret TO as the linear span of integer indices within a column or row of a matrix, and similarly ALL as all the respective indices within a column or row of a matrix. Note that Eqs. (7) and (8) are implemented with $l = 2$ *for*-cycles here, but for general l one should include l nested *for*-cycles.

    Given in this vectorized form, the algorithms are quite efficient and the processing time needed is very short. In the implementation one should simply follow the dimensions of the vectors and matrices to reach the correct evaluation. For example, if we are given two ($l = 2$) time-series of length $N = 4000$ and we use six base functions in total ($M = 6$) with three ($K = M/l = 3$) base functions for each data time series, then the respective dimensions of the matrices are: $\mathbf{E}_{l \times l} = \mathbf{E}_{2 \times 2}$, $\mathbf{r}_{K \times l} = \mathbf{r}_{3 \times 2}$, $\mathbf{\Xi}_{M \times M} = \mathbf{\Xi}_{6 \times 6}$, $\mathbf{c}_{k \times l} = \mathbf{c}_{3 \times 2}$, $\mathbf{P}_{K \times N} = \mathbf{P}_{2 \times 4000}$ and $\mathbf{v}_{l \times K \times N} = \mathbf{v}_{2 \times 3 \times 4000}$. For matrix implementation, it makes sense for some of the variables to be broken into parts like $\mathbf{v}_{l \times K \times N} = \{\mathbf{v1}_{K \times N}, \mathbf{v2}_{K \times N}\}$, or to be filled in parts e.g. $\mathbf{\Xi}_{M \times M}$ can be filled in four $\mathbf{\Xi}_{K \times K}$ parts. Note that in this way the dimensional representation of the vectors is slightly altered from the theoretical notation in order to reach the vectorized form in the interests of faster calculations.

    The three algorithms described are applied to a single window of data. The time-series are separated into sequential blocks, and the algorithms are applied to each of them in turn. The core of dynamical Bayesian inference is that it uses informative priors i.e. the evaluation of the next block of data depends on and uses the evaluation results from the previous block. The process of information propagation, between the n posterior and the next $n + 1$ prior distribution, can be adjusted to allow the time-variability of the parameters to be followed. We used propagation depending on the concentration matrix $\mathbf{\Xi}_{\text{pt}}$ or on the vector of parameters $\mathbf{c}_{\text{pt}}$. We now describe propagation with respect to the concentration matrix $\mathbf{\Xi}_{\text{pt}}$:

Algorithm 4: **Propagation**
$\mathbf{c}_{\text{pr}}^{n+1} = \mathbf{c}_{\text{pt}}^{n}$

$\mathbf{inv\Xi}^{n} = \text{inverse of } \mathbf{\Xi}_{\text{pt}}^{n}$;
$\mathbf{invDiff}^{n} = 0$
FOR i=1:K
    $\mathbf{invDiff}^{n}(\text{i,i}) = p_w^2 * \mathbf{inv\Xi}^{n}(\text{i,i})$
ENDFOR
$\mathbf{\Xi}_{\text{pt}}^{n+1} = \text{inverse of } (\mathbf{invDiff}^{n} + \mathbf{inv\Xi}^{n})$

Once the inference has been performed, one can use the inferred parameters to detect certain dynamical and phenomenological characteristics of the interacting

systems. For example, calculating the norm of the inferred parameters together with the relevant base functions, one can detect the coupling strength and directionality between the oscillators [27,28,30,38,39]. Similarly, one can reconstruct the form of the coupling function [1,6,40–42]. One can determine whether or not the systems are synchronized [33,43,44], once the parameters have been inferred. For the case of phase synchronization this can be done, for example, with a map representation and the modified Newton root-finding method (see [10] for details), while for generalized synchronization it can be done by evaluation of the largest Lyapunov exponents [11]. In this way both phase and generalized synchronization can be detected within the same framework based on the dynamical Bayesian inference presented. It is worth pointing that, by adopting this approach, one detects intrinsic synchronization based on the effective connectivity [45], which differs from other synchronization detection methods based on statistics of the phase and state time-series [31–33,46]. This can sometimes be advantageous [47].

5 Examples

In this section we outline three examples of the reconstruction of coupled oscillators that have time-varying dynamics and which are also subject to noise. The first example illustrates the basics of the inference method on a simple phase oscillator model, while the second example involves limit-cycle oscillators and presents the detection of several characteristics and relationships that can be detected from the inferred parameters. The third example presents the inference of coupled chaotic systems in state space. MatLab codes for the examples are provided at the following link[1].

5.1 Coupled phase oscillators

In order to present in a transparent way the basics of the inference technique we first consider two coupled phase oscillators [48] subject to noise:

$$
\begin{aligned}
\dot{\phi}_1 &= \omega_1(t) + a_1 \sin(\phi_1) + a_3(t) \sin(\phi_2) + \sqrt{E_{11}}\xi_1(t) \\
\dot{\phi}_2 &= \omega_2 + a_2 \sin(\phi_1) + a_4 \sin(\phi_2) + \sqrt{E_{22}}\xi_2(t).
\end{aligned}
\tag{10}
$$

Each oscillator is described by the frequency parameters ω_1, ω_2, the parameters for their self-dynamics a_1, a_4 and the coupling parameters a_2, a_3 for the direct influence coming from the other oscillator. Two parameters are set to be periodically time-varying, the frequency $\omega_1(t) = 2 - 0.5\sin(2\pi 0.00151t)$ and the coupling parameter $a_3(t) = 0.8 - 0.3\sin(2\pi 0.0012t)$. The noises are set to be white Gaussian and mutually uncorrelated. The other parameter values are $\omega_2 = 4.53$, $a_1 = 0.8$, $a_2 = 0$, $a_4 = 0.6$, $E_{11} = 0.03$ and $E_{22} = 0.01$. For all the examples we used a sampling of $h = 0.01$. Figure 1 shows samples from the resultant time series which serve as input for the dynamical Bayesian inference.

The choice of phase oscillators is very convenient for inference, because one needs to reconstruct the phase dynamics. Therefore, the phase model is known beforehand and the deterministic terms of the *rhs* of the coupled system (10) are the actual base functions to be used for inference of the six parameters (ω_1, ω_2, a_1, a_2, a_3 and a_4). Inference results from a single block of data are presented in Table 1. The agreement between the actual (intrinsic) parameters and their inferred values is excellent, and the method evidently works to high precision. Additionally – and which is unique for this
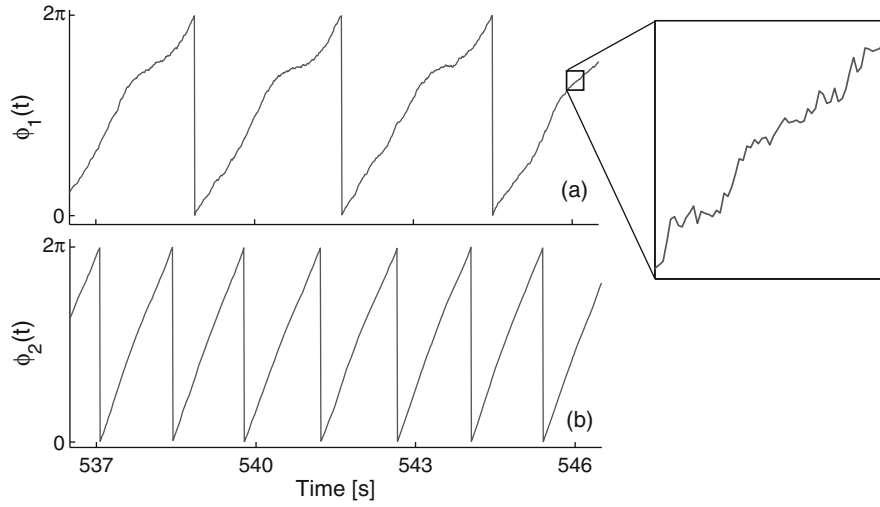
---

[1] `http://py-biomedical.lancaster.ac.uk/`

Fig. 1. The instantaneous phases generated by the model (10) of two coupled phase oscillators. (a) The phase ϕ_1 and (b) ϕ_2. For simpler presentation, the phases are "wrapped" within 2π. The enlarged panel on the right illustrates the noise perturbations of ϕ_1 on an expanded scale.

Table 1. Results from the inference of the numerically simulated system (10). The first row describes the physical meaning of the parameters, and the second and third two rows show, respectively, the actual values of the parameters and their inferred mean values. The results are presented for one window of data around $t = 1980$ s.

Parameters	ω_1	ω_2	a_1	a_2	a_3	a_4	E_{11}	E_{12}, E_{21}	E_{22}
Intrinsic values	2.032	4.530	0.800	0.000	1.013	0.600	0.030	0.000	0.010
Inferred means	2.026	4.537	0.803	-0.014	1.054	0.596	0.029	0.000	0.010

method – the intensity and the correlations of the noise are inferred very precisely. Figure 2 shows the time-variations of all the parameters inferred from sequential windows of length $w = 40$ s, with a propagation constant $p_w = 0.2$. It is clearly evident that the parameters and their time-variability are inferred precisely.

5.2 Coupled limit-cycle oscillators – phase domain inference

The second example involves a system of two coupled limit-cycle oscillators, which can serve as a model for a number of oscillatory processes that occur in nature, including electrochemical, mechanical, cardio-respiratory, and other biological systems [1,6,49,50]. The model consists of two interacting Poincaré oscillators subject to noise:

$$
\dot{x}_1 = -\left(\sqrt{x_1^2 + y_1^2} - 1\right)x_1 - \omega_1(t)y_1 + \varepsilon_1(x_2 - x_1) + \sqrt{E_{11}}\xi_1(t)
$$

$$
\dot{y}_1 = -\left(\sqrt{x_1^2 + y_1^2} - 1\right)y_1 + \omega_1(t)x_1 + \varepsilon_1(y_2 - y_1) + \sqrt{E_{22}}\xi_2(t)
$$

$$(11)$$

$$
\dot{x}_2 = -\left(\sqrt{x_2^2 + y_2^2} - 1\right)x_2 - \omega_2 y_2 + \varepsilon_2(t)(x_1 - x_2) + \sqrt{E_{33}}\xi_3(t)
$$

$$
\dot{y}_2 = -\left(\sqrt{x_2^2 + y_2^2} - 1\right)y_2 + \omega_2 x_2 + \varepsilon_2(t)(y_1 - y_2) + \sqrt{E_{44}}\xi_4(t),
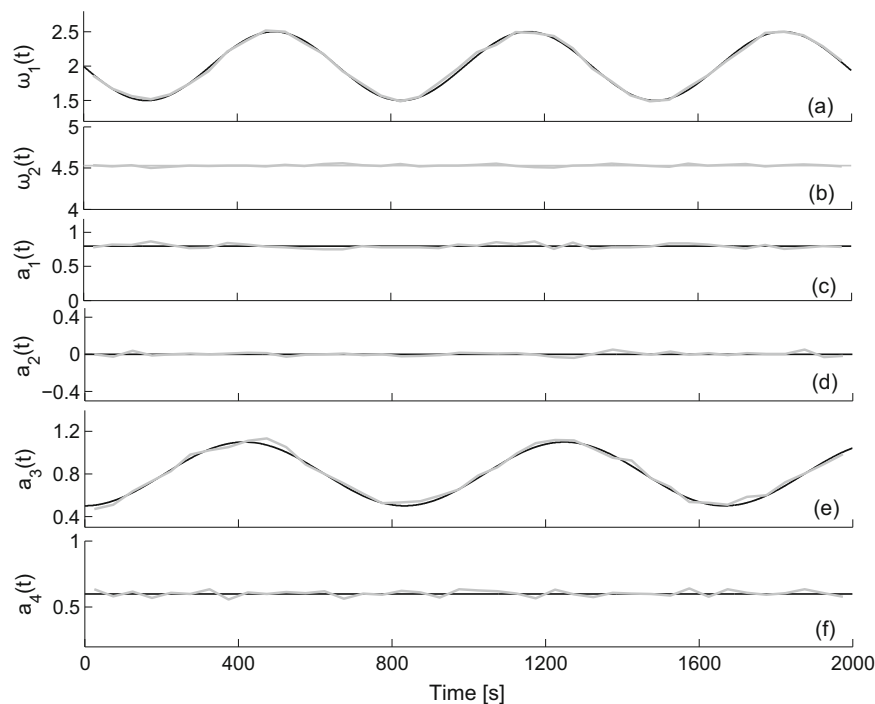$$

Fig. 2. Time-evolution of the parameters inferred from model (10). (a) and (b) Present the two frequencies, (c) and (f) are the self-dynamics parameters, and (d) and (e) present the coupling parameters. In each case, the actual values of the parameters are indicated by the black curves underlying their inferred values which are in gray.

where periodic time-variability is introduced in the frequency of the first oscillator $\omega_1(t) = 1 - 0.4\sin(2\pi 0.002t)$ and in the coupling parameter from the first to the second oscillator $\varepsilon_2(t) = 0.2 - 0.1\sin(2\pi 0.0017t)$. The noises are again white and Gaussian, with no correlations between them. The other parameters are $\omega_2 = 4.91$, $\varepsilon_1 = 0.05$, $E_{11} = E_{22} = 0.007$ and $E_{33} = E_{44} = 0.004$. The systems are simulated in state space, and the corresponding signals and phase portrait of the first system are given in Figs. 3a and b.

The phases can be estimated as $\phi_i = \arctan(y_i/x_i)$ (arctan being a four-quadrant function) from the state signals. Alternatively, one can use the Hilbert [50] or synchrosqueezed transforms [51]. The two phase time-series act as input for the dynamical Bayesian inference. The choice of base functions for the inference needs to be determined in such a way that the phase dynamics can be reconstructed effectively. Because of their oscillatory nature and periodic solutions, we decompose the phase dynamics into Fourier series. Hence Fourier series up to some order serve as base functions for the dynamical Bayesian inference. However, care must be taken to ensure that none of the functions have strong linear dependences on each other, as this can lead to imprecise and wrong separation of parameters within the inference: for example, choice of $\sin(x)$ and $\sin(-x)$, would rise to problems of precisely this kind because of their mutual linear dependence. In the example, we used only one side of the expansion, e.g. for $\sin(n\phi_1 + m\phi_2)$ we used the components $n = 1, \ldots, K$ instead of $n = -K, \ldots, K$. The reconstruction results for the time-varying parameters are presented in Figs. 3c and d. The periodic sine variations are evident both in the
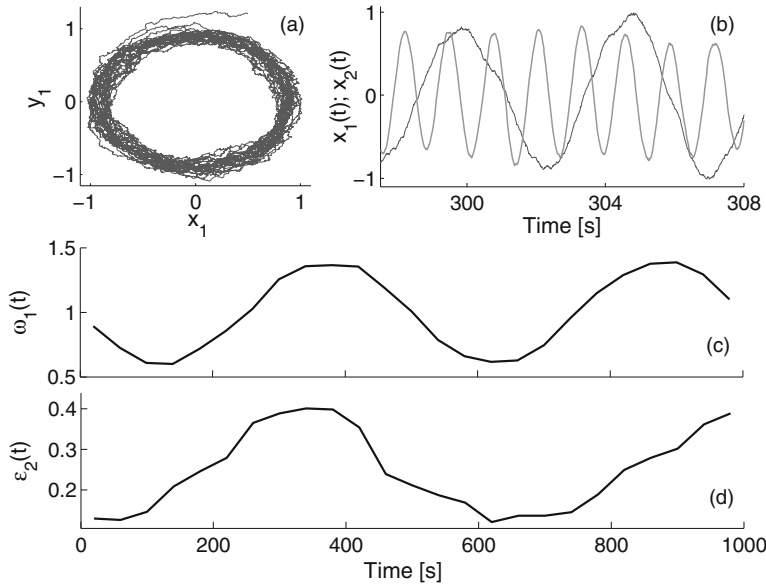
Fig. 3. Inference of time-varying parameters from the model (11) of two coupled limit-cycle oscillators. (a) Phase portrait showing the noisy state of the first oscillator. (b) The time-series of the two oscillators $x_1(t)$ and $x_2(t)$. (c) The inferred time-evolution of the frequency parameter, and (d) the inferred net coupling from the first to the second oscillator.

frequency and the coupling strength. Note that the coupling amplitude is evaluated as the norm of all the relevant inferred parameters that describe this influence.

One can use the inferred parameters, not only to evaluate the characteristics of individual oscillators, but also to determine whether the coupled system undergoes any qualitative transitions. An obvious example of the latter is the onset or disappearance of synchronization. For this reason we modify the parameters of model (11) with $\omega_1(t) = 1 - 0.4 \sin(2\pi 0.0022t)$, constant coupling parameter $\varepsilon_2 = 0.2$ and $\omega_2 = 1.4$. Such a combination of parameters takes the coupled system into and out of synchrony intermittently. The phase difference shown in Fig. 4a is bounded during the synchronized intervals [50]. Applying the procedure of return maps and the modified Newton root-finding method [10] we can identify the synchronization intervals (Fig. 4b), which correspond to the intervals of constant phase difference as expected. The map procedure is equivalent to a determination of whether or not the coupled phase oscillator model, with the inferred parameter values, is synchronized. Note also that, during the synchronized intervals, the phases remain almost identical and they do not span enough of the available space for inference (the phase difference appears as non-zero plateaus in Fig. 4a because it includes an offset corresponding to the phase difference that existed at the start of the synchronization interval considered). This can result in inferred parameters that are far from their intrinsic values. However, the whole set of parameters is again correlated as if it was coming from a synchronized system. One might ask: why do we need to undertake such a complicated procedure to detect synchronization, when something as simple as the phase difference can give a similar answer? The point is that, with the use of the intrinsic inferred parameters, one can distinguish whether or not the phase slips and synchronization transitions are noise-induced [1].

Coupling functions are arguably the most important part of the description of the inter-oscillator interactions. They can describe the functional relationship, the
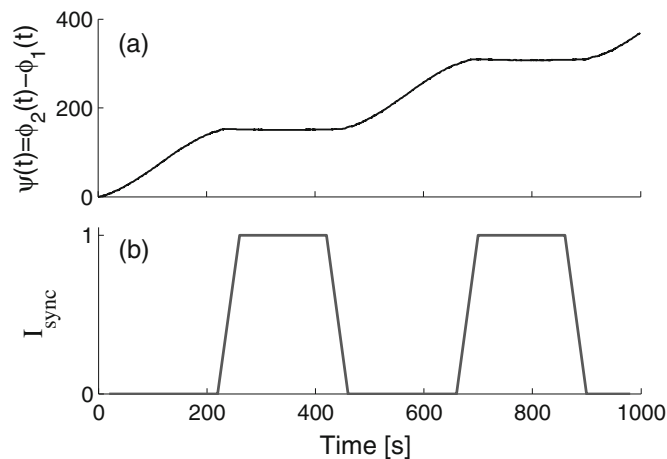
Fig. 4. Detection of intermittent synchronization in the system (11). (a) The phase difference $\psi(t)$ showing the qualitative statistical occurrence of synchronized intervals, which appear as bounded plateaus. (b) The synchronization index demonstrating the detected intrinsic synchronization intervals, evaluated from the inferred parameters. The high values $I_{\mathrm{sync}} = 1$ denote synchronization.
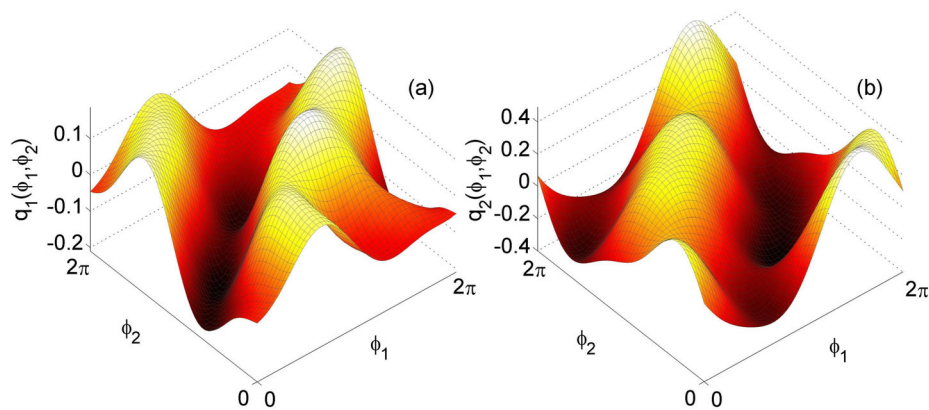

Fig. 5. Coupling functions inferred for from the model (11) of two interacting limit-cycle oscillators. (a) The functional influence $q_1(\phi_1, \phi_2)$ from the second to the first oscillator. (b) The coupling function $q_1(\phi_1, \phi_2)$ from the first to the second oscillator. The functions are inferred within one window of data.

law governing the mutual interactions, and the routes to qualitative transitions. As already mentioned above, coupling functions can be used quite generally to describe different aspects of the interactions that occur between a great diversity of oscillatory systems, whether e.g. cardio-respiratory, electrochemical or mechanical [1,6,40–42,52,53]. By representation on a 2π-phase grid evaluated for the relevant inferred parameters, we can determine and visualize the phase dynamics very effectively. The coupling functions of systems (11) inferred from a single window are shown in Fig. 5. By inferring the dynamics in a succession of windows, we can follow the time-evolution of the functions [1]. Moreover, the coupling function $q_i(\phi_i, \phi_j)$ can be further decomposed into self, direct and indirect coupling influences [35], and each of these can be studied separately.

5.3 Coupled chaotic systems – state domain inference

Up to this point, the review has focused mainly on the inference of phase dynamics. The main reason was to present a method that will be quite generally applicable to coupled oscillatory systems. However, there are some situations where the dynamics needs to be analyzed directly from the measured signals in state space. For example, the estimation of phases from chaotic systems can be problematic, while inference in the state domain is directly accessible. If state signals are to be analyzed, then the model Eq. (2) will still hold; the base functions can have e.g. a polynomial form, and all the rest of the equations and algorithms can equally be applied to the state signals. For other applications of the inference of (coupled) dynamics in state space see [11,12,21,22,54].

Coupled state space systems, especially chaotic systems, have played an important role in applications to the secure encryption of communications [55–58]. Recently, a new class of secure communication that is highly resistent to conventual attacks, was introduced [11] using the same Bayesian method presented in this tutorial. The scheme makes use of the coupling functions between interacting dynamical systems. The information signals are encrypted as the time-variations of independent coupling functions between the coupled systems. Using predetermined forms of coupling function, we can apply dynamical Bayesian inference on the receiver side to detect and separate the information signals while simultaneously eliminating the effect of external noise. The procedure results in an unbounded number of encryption key possibilities, allows the transmission/reception of more than one signal simultaneously, and is robust against external noise. The use of chaotic systems is not essential for the encryption. It does, however, bring an additional level of security associated with the scrambled random-like appearance of chaotic signals.

It was at the beginning of the chaotic communication era when Pérez and Cerdeira [59] demonstrated a way of breaking such communication schemes by extracting the messages they carried. Their methods were based on reconstruction of the attractors' properties from the transmitted signals only. This valuable work not only broke the existing schemes, but also introduced an attacking principle that all future (chaotic) attractor-based schemes must resist. The new scheme based on coupling functions and dynamical Bayesian inference is able to withstand such attacks because the information is encrypted in the weak coupling between two independent and self-sustained attractors.

To illustrate dynamical Bayesian inference in state space we consider a coupled pair of chaotic Rössler and Lorenz oscillators. This coupled system represents the coupling-function-based communication model and is duplicated on both the transmitter and receiver sides. The model is given by a Rössler system:

$$
\begin{aligned}
\dot{x}_1 &= -2y_1 - z_1 \\
\dot{y}_1 &= 2x_1 + 0.45y_1 + \sqrt{E_{11}}\xi_1(t) \\
\dot{z}_1 &= 2 + x_1 z_1 - 10 z_1,
\end{aligned}
\tag{12}
$$

driving a Lorenz system

$$
\begin{aligned}
\dot{x}_2 &= 10y_2 - 10x_2 \\
\dot{y}_2 &= 28x_2 - x_2 z_2 - y_2 \\
\dot{z}_2 &= x_2 y_2 - 2.66 z_2 + \varepsilon_1(t) y_1 + \varepsilon_2(t) x_1 z_1 + \sqrt{E_{22}}\xi_2(t),
\end{aligned}
\tag{13}
$$

where the noises are taken to be white and Gaussian with noise intensities $E_{11} = 0.05$ and $E_{22} = 0.3$. The coupling parameters $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are set to be time-varying
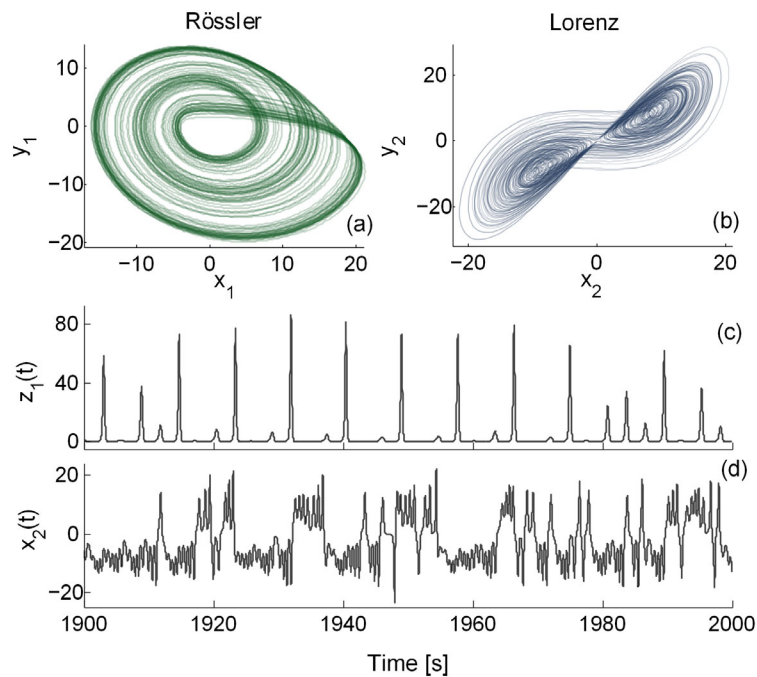
Fig. 6. The coupled chaotic Rössler and Lorenz systems with time-varying parameters and subject to noise (12), (13). The phase portraits or strange attractors are shown (a) for the Rössler and (b) for the Lorenz systems. Short time-segments of the chaotic time series are presented: (c) $z_1(t)$ from the Rössler and (d) $x_2(t)$ from the Lorenz system.

i.e. they can represent the information messages that are to be securely encrypted. The first coupling is binary $\varepsilon_1(t) = \{0, 2\}$, while the second is continuous $\varepsilon_2(t) = 3 + 0.3\sin(2\pi 0.001t)$. The attractors and signals of such chaotic systems subject to noise and time-varying couplings are shown in Fig. 6.

This particular example is very convenient for the use of dynamical Bayesian inference on the receiver side because one knows both the model and its base functions *a priori*, and has access to the time-series of all the dimensions. Additionally, the chaotic signals with their notion of strange-attractors tend to span a broad region of state space (larger than e.g. limit-cycle oscillators), which provides more information and makes the inference easier and more precise. Hence, by using the functions on the *rhs* of systems (12), (13) as base functions, one can infer the model parameters and their time-variability. All six multivariate state time-series from the two coupled systems serve as input for the Bayesian inference. Among them are the two time-varying coupling parameters, which in the communication example could convey the information messages as shown in Fig. 7. As it can be seen, both the digital (e.g. 1010010) and the continuous (e.g. sine or speech) message can be inferred with great precision.

6 Discussion and conclusion

The aim of this review has been to provide insights into the implementation of dynamical Bayesian inference in a clear and simple way. Further details of the method and its applications can be find in [1,10,12,21,22]. Needless to say, the overall framework can be much broader and a number of important generalizations are possible.
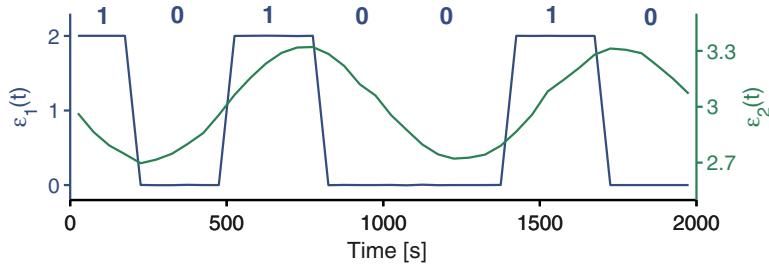
Fig. 7. The inferred time-varying coupling parameters of the model (12), (13) as an illustration of the information carriers in secure communications. The ordinate axis on the left shows the binary time-varying parameter $\varepsilon_1(t)$, which can be used e.g. to encrypt the digital message *1010010* shown along the top of the figure. One can also encrypt another signal simultaneously, as illustrated by the continuous sine message embodied as the variation of the $\varepsilon_2(t)$ parameter on the right ordinate axis.

The examples here included only a pair of coupled systems, while in general the technique can be applied to a larger group e.g. a small-scale network of oscillators [10,60]. The phase decompositions can be applied for pairwise couplings but, more importantly, joint coupling influences can also be inferred. In such cases one can detect unique states characteristic only of the joint couplings, like triplet synchronization [61]. In these kind of analysis, the effective coupling [45,62] should be distinguished by appropriate use of surrogate testing [63].

As examples, we analyzed two types of model – oscillatory phase dynamics models and a chaotic state-space model. The phase dynamics coming from the oscillatory systems was well-modeled in terms of Fourier series while, for the chaotic secure communications example, we exploited the fact that the model was known beforehand. In general, however, the model might not have been known or easily decomposable. In such cases, one would have needed to apply one or some of the known system identification techniques [64,65] prior to the inference.

The procedure of information propagation that allows time-variability to be tracked, depends on p_w which acts as a free parameter. One can further improve this procedure by making the parameter adaptive in order to follow the time-variability more closely and to infer the noise more precisely. This might be realized by determining the optimal parameter from a spectrum of values within each window. Even though this procedure would be very slow, it might prove helpful and could be necessary in certain cases.

The phase base functions are not strictly confined all to be from Fourier series, and other additional functions can be included. For example, if we have expansion up to second order $K = 2$, one can include also other components such as $\sin(7\phi_1 - 4\phi_2)$ in order to detect synchronization more precisely for 7:4 synchronization ratios. Similarly the state base functions can be extended to include a large set of functions and only the ones that intrinsically belong to the underlying model will be inferred as "non-zero" [12].

Within the framework of dynamical inference, the differential $\dot{\chi}_n$ is calculated as $\dot{\chi}_n = (\chi_{n+1} - \chi_n)/h$. Improved performance can be accomplished if one provides this as directly estimated instantaneous frequency – e.g. using the synchrosqueezed transform [51] or nonlinear mode decomposition [35]. Alternatively, certain digital filters (e.g. the Savitzky-Golay filter) can be used for smoothing the noise effect before evaluating the derivatives.

The inference of causality between interacting oscillatory systems has attracted much attention in the last decade. In addition to nonlinear dynamics methods, both

for inferring direction of coupling or the coupling function [27,34], methods based on information theory [28] and wavelet bispectral analysis [30], there is the dynamical Bayesian inference method presented above. It facilitates comprehensive reconstruction of the dynamical properties of the interacting systems – either two or a whole network – and allows every aspect of their interactions to be studied, including their synchronization, direction of coupling, and coupling function. The approach provides deep insight into the properties of the dynamical systems of interest and thus makes possible both diagnosis and prognosis of their behaviour.

Of course, dynamical Bayesian inference possesses relationships, similarities and complementarities with other methods, including the inference of deterministic or stochastic models, based on Bayesian theory, particle filters and maximum likelihood estimators [2–6]. The inference of coupling causality has huge practical applicability and methods based on Granger causality or transfer entropy [28,29,66–68] have recently become popular in this context. We note however, that such methods infer statistical effects, while the method presented here, being based on a dynamical model, can infer causal mechanisms [69]. In other words, Granger causality-like methods infer only the existence of causal effect and not the nature and mechanism of the cause itself. In neuroscience, Granger causality methods are linked to directed functional connectivity, while dynamical inference methods distinguish effective connectivity [62].

In summary, this tutorial has been intended to familiarize the reader with a technique for the Bayesian inference of time-evolving coupled dynamics in the presence of noise. A comprehensive description of the method, including the theoretical constraints, algorithms, implementation, and demonstrations of their main features in relation to a few characteristic examples has been provided. To facilitate the first steps in applying this powerful and useful method MatLab codes including examples are also being made available. We hope that the tutorial will lead the reader to new insights into dynamical phenomena from measured data, and that it will provide a useful aid for tackling a diverse range of signal processing problems.

References

1. T. Stankovski, A. Duggento, P.V.E. McClintock, A. Stefanovska, Phys. Rev. Lett. **109**, 024101 (2012)
2. D.A. Smirnov, B.P. Bezruchko, Phys. Rev. E **79**(4), 046204 (2009)
3. K.J. Friston, NeuroImage **16**(2), 513 (2002)
4. M. Sanjeev Arulampalam, S. Maskell, N. Gordon, T. Clapp, Signal Processing, IEEE Trans. **50**(2), 174 (2002)
5. H.U. Voss, J. Timmer, J. Kurths, Int. J. Bifurcat. Chaos **14**(06), 1905 (2004)
6. I.Z. Kiss, Y. Zhai, J.L. Hudson, Phys. Rev. Lett. **94**, 248301 (2005)
7. R.E. Kalman, J. Fluid. Eng. **82**(1), 35 (1960)
8. S.J. Julier, J.K. Uhlmann, *New extension of the Kalman filter to nonlinear systems*, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 3068 (1997) p. 182
9. T. Bayes, Phil. Trans. **53**, 370 (1763)
10. A. Duggento, T. Stankovski, P.V.E. McClintock, A. Stefanovska, Phys. Rev. E **86**, 061126 (2012)
11. T. Stankovski, P.V.E. McClintock, A. Stefanovska, Phys. Rev. X **4**, 011026 (2014)

12. V.N. Smelyanskiy, D.G. Luchinsky, A. Stefanovska, P.V.E. McClintock, Phys. Rev. Lett. **94**(9), 098101 (2005)
13. U. von Toussaint, Rev. Mod. Phys. **83**(3), 943 (2011)
14. V.N. Smelyanskiy, D.G. Luchinsky, D.A. Timuchin, A. Bandrivskyy, Phys. Rev. E **72**(2), 026202 (2005)
15. W.D. Penny, V. Litvak, L. Fuentemilla, E. Duzel, K. Friston, J. Neurosci. Meth. **183**(1), 19 (2009)
16. H. Haken, Rev. Mod. Phys. **47**, 67 (1975)
17. P.E. Kloeden, M. Rasmussen, *Nonautonomous Dynamical Systems* (AMS Mathematical Surveys and Monographs, New York, 2011)
18. Y.F. Suprunenko, P.T. Clemson, A. Stefanovska, Phys. Rev. Lett. **111**(2), 024101 (2013)
19. P.T. Clemson, Y.F. Suprunenko, T. Stankovski, A. Stefanovska, Phys. Rev. E **89**, 032904 (2014)
20. J. Daunizeau, K.J. Friston, S.J. Kiebel, Physica D **238**(21), 2089 (2009)
21. D.G. Luchinsky, V.N. Smelyanskiy, A. Duggento, P.V.E. McClintock, Phys. Rev. E **77**(6), 061105 (2008)
22. A. Duggento, D.G. Luchinsky, V.N. Smelyanskiy, I. Khovanov, P.V.E. McClintock, Phys. Rev. E **77**(6), 061106 (2008)
23. A. Stefanovska, H. Haken, P.V.E. McClintock, M. Hožič, F. Bajrović, S. Ribarič, Phys. Rev. Lett. **85**(22), 4831 (2000)
24. Y. Shiogai, A. Stefanovska, P.V.E. McClintock, Phys. Rep. **488**, 51 (2010)
25. D. Rudrauf, A. Douiri, C. Kovach, J.P. Lachaux, D. Cosmelli, M. Chavez, C. Adam, B. Renault, J. Martinerie, M.L. Van Quyen, Neuroimage **31**(1), 209 (2006)
26. V. Jirsa, V. Müller, Frontiers Comput. Neurosci. **7**, 78 (2013)
27. M.G. Rosenblum, A.S. Pikovsky, Phys. Rev. E. **64**(4), 045202 (2001)
28. M. Paluš, A. Stefanovska, Phys. Rev. E **67**, 055201(R) (2003)
29. A. Porta, L. Faes, V. Bari, A. Marchi, T. Bassani, G. Nollo, N.M. Perseguini, J. Milan, V. Minatel, A. Borghi-Silva et al., PloS one **9**(2), e89463 (2014)
30. J. Jamšek, M. Paluš, A. Stefanovska, Phys. Rev. E **81**(3), 036207 (2010)
31. P. Tass, M.G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, H.J. Freund, Phys. Rev. Lett. **81**(15), 3291 (1998)
32. F. Mormann, K. Lehnertz, P. David, C.E. Elger, Physica D **144**(3-4), 358 (2000)
33. N.F. Rulkov, M.M. Sushchik, L.S. Tsimring, H.D.I. Abarbanel, Phys. Rev. E **51**(2), 980 (1995)
34. B. Kralemann, M. Frühwirth, A. Pikovsky, M. Rosenblum, T. Kenner, J. Schaefer, M. Moser, Nat. Commun. **4**, 2418 (2013)
35. D. Iatsenko, A. Bernjak, T. Stankovski, Y. Shiogai, P.J. Owen-Lynch, P.B.M. Clarkson, P.V.E. McClintock, A. Stefanovska, Phil. Trans. R. Soc. Lond. A **371**(1997), 20110622 (2013)
36. I. Manolopoulou, M.P. Matheu, M.D. Cahalan, M. West, T.B. Kepler, J. Amer. Stat. Assoc. **107**(499), 855 (2012)
37. P. Lemey, A. Rambaut, A.J. Drummond, M.A. Suchard, PLOS Comp. Biol. **5**(9), e1000520 (2009)
38. M. Staniek, K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008)
39. A. Bahraminasab, F. Ghasemi, A. Stefanovska, P.V.E. McClintock, H. Kantz, Phys. Rev. Lett. **100**(8), 084101 (2008)
40. B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, R. Mrowka, Phys. Rev. E **76**(5), 055201 (2007)
41. I.T. Tokuda, S. Jain, I.Z. Kiss, J.L. Hudson, Phys. Rev. Lett. **99**, 064101 (2007)
42. J. Miyazaki, S. Kinoshita, Phys. Rev. Lett. **96**, 194101 (2006)
43. M.G. Rosenblum, A.S. Pikovsky, J. Kurths, Phys. Rev. Lett. **76**(11), 1804 (1996)
44. L. Kocarev, U. Parlitz, Phys. Rev. Lett. **76**(11), 1816 (1996)
45. H.J. Park, K. Friston, Science **342**(6158), 1238411 (2013)
46. U. Parlitz, L. Junge, W. Lauterborn, L. Kocarev, Phys. Rev. E **54**, 2115 (1996)
47. T. Stankovski, P.V.E. McClintock, A. Stefanovska, Phys. Rev. E **89**, 062909 (2014)

48. Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence* (Springer-Verlag, Berlin, 1984)
49. B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, R. Mrowka, Phys. Rev. E **77**(6, Part 2), 066205 (2008)
50. A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization – A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, 2001)
51. I. Daubechies, J. Lu, H. Wu, Appl. Comput. Harmon. Anal. **30**(2), 243 (2011)
52. R.F. Galán, G.B. Ermentrout, N.N. Urban, Phys. Rev. Lett. **94**, 158101 (2005)
53. T. Pereira, J. Eldering, M. Rasmussen, A. Veneziani, Nonlinearity **27**(3), 501 (2014)
54. V.N. Smelyanskiy, D.A. Timucin, D.G. Luchinsky, A. Stefanovska, A. Bandrivskyy, P.V.E. McClintock, in *IUTAM Symposium on Nonlinear Stochastic Dynamics*, edited by N.S. Namachchivaya, Y.K. Lin (Kluwer, Amsterdam, 2003), p. 455
55. K.M. Cuomo, A.V. Oppenheim, Phys. Rev. Lett. **71**, 65 (1993)
56. L. Kocarev, U. Parlitz, Phys. Rev. Lett. **74**(25), 5028 (1995)
57. A. Argyris, D. Syvridis, L. Larger, V. Annovazzi-Lodi, P. Colet, I. Fischer, J. Garcia-Ojalvo, C.R. Mirasso, L. Pesquera, K.A. Shore, Nature **438**(7066), 343 (2005)
58. G. Alvarez, S. Li, Intern. J. Bifurc. Chaos **16**(8), 2129 (2006)
59. G. Pérez, H.A. Cerdeira, Phys. Rev. Lett. **74**, 1970 (1995)
60. B. Kralemann, A. Pikovsky, M. Rosenblum, Chaos **21**, 025104 (2011)
61. B. Kralemann, A. Pikovsky, M. Rosenblum, Phys. Rev. E **87**, 052904 (2013)
62. K.J. Friston, Brain. Connect. **1**(1), 13 (2011)
63. T. Schreiber, A. Schmitz, Physica D **142**(3-4), 346 (2000)
64. L. Ljung, *System identification* (Springer, 1998)
65. K.P. Burnham, D.R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach* (Springer, 2002)
66. C.W. Granger, Econ: J. Econ. Soc., 424 (1969)
67. A.K. Seth, Network: Comput. Neural Syst. **16**(1), 35 (2005)
68. T. Schreiber, Phys. Rev. Lett. **85**(2), 461 (2000)
69. A.B. Barrett, L. Barnett, Front. Neuroinf. **7** (2013)