Online bayesian inference in some time-frequency representations of non-stationary processes

Article

Accepted Version

Everitt, R. G., Andrieu, C. and Davy, M. (2013) Online bayesian inference in some time-frequency representations of non-stationary processes. IEEE Transactions on Signal Processing, 61 (22). pp. 5755-5766. ISSN 1053-587X doi: https://doi.org/10.1109/TSP.2013.2280128 Available at https://centaur.reading.ac.uk/33926/

Online Bayesian Inference in Some Time-Frequency Representations of Non-Stationary Processes

Richard G. Everitt*, Christophe Andrieu, and Manuel Davy

***Abstract*—The use of Bayesian inference in the inference of time-frequency representations has, thus far, been limited to offline analysis of signals, using a smoothing spline based model of the time-frequency plane. In this paper we introduce a new framework that allows the routine use of Bayesian inference for online estimation of the time-varying spectral density of a locally stationary Gaussian process. The core of our approach is the use of a likelihood inspired by a local Whittle approximation. This choice, along with the use of a recursive algorithm for non-parametric estimation of the local spectral density, permits the use of a particle filter for estimating the time-varying spectral density online. We provide demonstrations of the algorithm through tracking chirps and the analysis of musical data.**

***Index Terms*—Signal processing algorithms, particle filters, spectrogram, Bayesian methods, frequency domain analysis. EDICS Categories: DSP-TFSR, MLR-BAYL, MLR-MUSI, SSP-NSSP, SSP-TRAC.**

I. Introduction

A. Background

Time-frequency representations (TFRs) are celebrated signal processing tools, for they turn time domain signals into images, representing the time-frequency decomposition of the signal, whose interpretation is intuitive. Such images are thus often used as early analysis tools, to be used when faced with non-stationary signals for which the concept of frequency is relevant. Typical applications range from audio signals (speech, music, animal cries) [1], biomedical time series (EEG, ECG, EMG) [2], accelerometer signals [3] collected on dynamic mechanical systems, etc.

Aside their ability to provide easy-to-understand images, TFRs can be used in automated decision applications. [4] showed that some Cohen's group TFRs can be used to implement optimal linear detection, by providing an equivalent implementation of the time domain matched filter in the time frequency domain. This seminal work was followed by a number of studies about time-frequency detection / classification of signals, see e.g., [5]–[8]. All these approaches rely on the idea that the noise is spread all over the time-frequency plane, thus increasing the local signal-to-noise ratio. Based on the same idea, several algorithms have been proposed to estimate signal parameters directly from TFRs, most of which being devoted to linear chirps, see e.g., [9]. Others are devoted to more general time-varying spectrum estimation. In particular, many studies have focused on estimating narrow band time-frequency trajectories (also termed *components*), see [10], [11] among others. Most of these techniques implement Kalman filtering where the observations are extracted from peaks of TFRs (sequential approach), or fit curves onto the TFR (batch approach). As shown by these many previous studies, using TFRs for estimation or decision purposes leads to powerful algorithms. Many practical problems have received a convincing solution through the use of these methods [12], [13]. For a general overview of the time-frequency analysis, see [14].

In parallel to the development of time-frequency techniques, statistical signal processing tools have developed considerably over the last few years. In particular, up-to-date Bayesian approaches enable to estimate parameters in situations where the model is highly non-linear and/or non-Gaussian, thanks to Monte Carlo methods. Surprisingly, few works have been devoted to apply these techniques to TFRs. Previous work can be summarised as follows:

1) Monte Carlo Markov Chain (MCMC) for estimating the coefficients of Gabor representations [15], [16].
2) Tracking the pole parameterisation of time-varying autoregression (TVAR) models using sequential Monte Carlo [17].
3) Tracking trajectories in spectrograms by sequential Monte Carlo (SMC) [18].
4) Using reversible-jump MCMC to infer the parameters of a model for the time-varying spectrum of a locally stationary process [19] and [20]. The model in these papers partitions the process into segments, using a mixture of smoothing splines to model the log spectrum of each segment and time varying mixing weights to introduce non-stationarity across the segments.

Despite these advances, there exists no principled approach to Bayesian inference that uses the full TFR as a raw observation and relates it to some statistical model through a likelihood function. Devising such a methodology is the aim of this paper.

R. G. Everitt* is with the Department of Mathematics and Statistics, University of Reading, UK e-mail: richard.g.everitt@gmail.com.

C. Andrieu is with the Department of Mathematics, University of Bristol, UK.

M. Davy is with the LAGIS Laboratory, CNRS, Lille, France.

Manuscript received xxx xx, xxxx; revised xxx xx, xxxx.

B. Overview of the approach

This paper brings together TFRs and the Bayesian paradigm for statistical inference. More precisely, we aim to take advantage of the numerous conceptual benefits of TFRs in order to motivate and develop models for non-stationary processes while at the same time exploiting the Bayesian paradigm for both incorporating prior information and quantifying uncertainty in the inference procedure. This leads to a natural description of the process to be analysed in terms of a state-space model which is well suited to the sequential processing of data (often referred to as filtering) and the incorporation of prior temporal information (e.g. smoothness, abrupt changes, etc). Sequential processing might be desirable in scenarios where an update of the information contained in the signal is required as soon as a new data is available, but can also present a significant computational and storage advantage when large datasets are involved. SMC methods (aka particle filters) are well established statistical techniques that allow one to effectively and efficiently carry out sequential inference for such models, in the presence of nonlinearity and non-Gaussianity.

The main difficulty inherent to achieve the goals set above was to find a principled way of relating TFRs to the classical statistical parametric inference framework and in particular define a likelihood function for such objects. Our approach consists of exploiting Whittle type approximations to the likelihood of Gaussian processes. These approximate likelihoods have the advantage that they can be shown to depend exclusively on the spectral properties of the data and the candidate statistical parametric model. Although originally developed in the context of stationary processes, recent theoretical advances have allowed for the rigorous generalisation of this framework to some types of non-stationary processes; a review of such approximations is provided in section II. This is the route we follow in this paper. In section III we show how the standard Bayesian state-space modelling framework, in conjunction with Whittle type likelihoods, lends itself naturally to the modelling of non-stationary processes in the spectral domain. A generic particle filter to perform sequential inference is also briefly reviewed.

At the core of our approach to sequential estimation is to use efficient recursive estimators of time-frequency images from data. A novel solution to this problem is described in section III-C, which relies on an unusual interpretation of the power spectral density (PSD) of stationary processes and its non-stationary counterparts. In section IV we apply our methodology to tracking chirps whose signals overlap in time-frequency space and in section V to a simple problem in the analysis of musical data.

II. From the Gaussian likelihood to the spectrogram

A. The Whittle approximation

When looking at TFRs, it is tempting to try to fit a parametric model in order to reduce dimensionality and improve interpretability, including interpretability by a computer. However the choice of a likelihood in such situations is not always obvious. In this section we develop such a likelihood which is motivated by Whittle's approximant of the likelihood for stationary processes and inspired by its extension suggested in [21] for locally stationary processes.

Consider a zero mean real valued stationary Gaussian process $\{y_t | t = 1, \ldots, T\}$ for some $T \geq 1$ with nonzero PSD $f_\theta : [-\pi, \pi] \to (0, \infty)$ dependent on a parameter $\theta \in \Theta$. In [22] it was shown that the log-likelihood of a real valued realisation of such a stationary Gaussian process can for a general class of processes be approximated for T large enough by

$$\mathcal{L}_\theta(y_{1:T}) := C - \frac{T}{4\pi} \int_{-\pi}^{\pi} [\log(2\pi f_\theta(\omega)) + I_T(\omega)/f_\theta(\omega)] d\omega, \tag{1}$$

for some constant C, where $I_T(\omega)$ is the periodogram of the data, defined for $\omega \in [-\pi, \pi]$ as

$$I_T(\omega) := \frac{1}{T} \left| \sum_{t=1}^{T} y_t \exp(-\imath \omega t) \right|^2 . \tag{2}$$

This approximation can easily be extended to accommodate vector valued time series and rates of convergence of this approximation can also be derived (see [22] and [23]). The interest of this approximation is that it relates the spectral properties of the data (the periodogram) to the statistical model for the data (the PSD f_θ) in a principled manner and presents the advantage of allowing modelling of this class of processes directly in the spectral domain.

There is a large literature on the use of this approximation for maximum likelihood inference of the parameter θ ([23] is a good starting point). However, the approach of [24], where the Whittle approximation to the likelihood as a part of a Bayesian model, is of more relevance to this paper. In this paper a numerical approximation to the integral in equation (1) is used, evaluating the integrand at the Fourier frequencies $\{\omega_k = 2\pi k/T | \, 0 \leq k \leq \lfloor T/2 \rfloor\}$ to obtain

$$\hat{\mathcal{L}}_\theta(y_{1:T}) = C - \frac{1}{2} \sum_k [\log(2\pi f_\theta(\omega_k)) + I_T(\omega_k)/f_\theta(\omega_k)]. \tag{3}$$

In the Bayesian model in [24] this likelihood is further approximated by a mixture of Gaussians (for computational reasons), and MCMC is used to infer a smoothing spline representation of the log PSD. From our perspective, the most important characteristic of this work is that the use of the Whittle approximation provides a principled approach to Bayesian inference of any appropriate statistical model of the PSD, using the periodogram directly as an observation. This flexibility in the choice of model for the PSD is exploited in [25] and [26], amongst others.

B. Locally stationary processes

For our analogous methodology for the Bayesian estimation of TFRs, a natural question is that of the existence of such approximations for Gaussian non-stationary processes that would allow us to relate a TFR of the observed process and a parametric model through a likelihood, or an approximation of such a likelihood. A significant step in this direction was achieved by [21] who extended Whittle's approximant to a particular class of non-stationary processes called "locally stationary processes" [21], [27]–[30]. This class of processes can be thought of as being a time varying generalisation of stationary harmonisable processes and are assumed to have a representation of the form (assuming for simplicity that $\mathbb{E}(Y_{t,T}) = 0$ for $t = 1, \ldots, T$)

$$Y_{t,T} = \int_{-\pi}^{\pi} \exp(\imath \omega t) \Lambda_T(t,\omega) d\xi(\omega), \tag{4}$$

where $\{\xi(\omega)\}$ is a complex valued Gaussian process on $[-\pi, \pi]$ with additional statistical properties detailed in the appendix and $\{\Lambda_T(t,\omega) : \{1,\ldots,T\} \times [-\pi,\pi] \to \mathbb{C}, T \in \mathbb{N}\}$ is a family of complex valued functions which can be approximated by a function $\Lambda : [0,1] \times [-\pi,\pi] \to \mathbb{C}$, $\Lambda(u,\omega)^* = \Lambda(u,-\omega)$ such that there exists K satisfying

$$\sup_{(t,\omega)\in\{1,\ldots,T\}\times[-\pi,\pi]} |\Lambda_T(t,\omega) - \Lambda\left(t/T,\omega\right)| \leq KT^{-1}. \tag{5}$$

This class of processes is especially useful, since it defines a unique evolutionary spectrum for many commonly used processes, including time-varying ARMA or GARCH processes, for example.

In the context of parametric estimation it is natural to consider a family of potential candidates $\{\Lambda_\theta, \theta \in \Theta\}$ in order to explain the data. The quantity $f_\theta(u,\omega) := |\Lambda_\theta(u,\omega)|^2 \geq 0$, the *evolutionary spectrum*, can be thought of as being a parametric TFR for the process.

In close analogy with Whittle's likelihood approximation for Gaussian stationary processes, and for the purpose of statistical inference, In [21] the following approximation of the log likelihood for T observations of a Gaussian locally stationary process is derived

$$\begin{aligned}\mathcal{L}_\theta(y_{1:T}) \quad := \quad & C + \sum_{t=1}^{T} -\frac{1}{4\pi} \int_{-\pi}^{\pi} [\log(2\pi f_\theta(t,\omega)) + \\ & \tilde{I}_T(t,\omega)/f_\theta(t,\omega)] d\omega,\end{aligned} \tag{6}$$

where C is a constant and $\tilde{I}_T(t,\omega)$ is the *pre-periodogram*

$$\begin{aligned}\tilde{I}_T(t,\omega) \quad = \quad & \sum_{\{k\in\mathbb{N}:1\leq t+1/2\pm k/2\leq T\}} y_{\lfloor t+1/2+k/2 \rfloor} \\ & y^*_{\lfloor t+1/2-k/2 \rfloor} \exp(-\imath\omega k),\end{aligned} \tag{7}$$

with * denoting complex conjugation and $\lfloor\rfloor$ the largest integer smaller than the argument (the pre-periodogram is a particular discretisation of the Wigner-Ville distribution, and is essentially an estimator of the evolutionary spectrum). It is noted in [21] that the likelihood approximation may be factorised as $\mathcal{L}_\theta = \sum_{t=1}^{T} l_{\theta,t}$, where each factor can be thought of as the "local log likelihood" at time t. By analogy with the stationary case, it is suggested by [21] to maximise this approximate likelihood over values of θ, approximating a maximum likelihood estimator. Theoretical results are provided concerning the asymptotic consistency of the maximum approximate likelihood estimator derived by maximising $\mathcal{L}_\theta(y_{1:T})$, as well as a central limit theorem.

Bayesian estimation of the evolutionary spectrum of a locally stationary process has previously been considered in [19] and [20]. In common with this paper, the choice of likelihood in these papers is also motivated by [30]. However, they differ from our work in that the likelihood used for the segmented process in their model (see section I-A) is simply the Whittle approximation for the likelihood of the data in each segment - inference of the parameters of their model is based on the periodogram of the data in each segment. In addition, our goal is to infer the evolutionary spectrum online, whereas these papers take a batch approach.

C. Approach

In the present paper we develop a methodology to perform sequential Bayesian inference for processes modelled in the spectral domain that uses time-frequency estimates as input. Our model for the observed data $y_{1:T}$ is parameterised by the time-varying parameter $\{\theta_t\}$, which we treat as a random variable. The aim of the Bayesian approach is to infer a sequence of posterior distributions $\{p(\theta_t|Y_{1:t} = y_{1:t})\}$ thus describing the uncertainty present in inferring $\{\theta_t\}$, in contrast to maximum likelihood estimators which seek only a point estimate. The Bayesian formulation consists of specifying: a prior $p(\theta_{1:T})$ on the time-varying parameter; and a likelihood $g(Y_{1:T}|\theta_{1:T})$, modelling how a random process $Y_{1:T}$ arises given the underlying parameter $\theta_{1:T}$. The posterior $p(\theta_{1:T}|Y_{1:T} = y_{1:T})$ is then found via Bayes' theorem: $p(\theta_{1:T}|Y_{1:T} = y_{1:T}) \propto p(\theta_{1:T}) g(Y_{1:T} = y_{1:T}|\theta_{1:T})$.

The likelihood we use is inspired by Dahlhaus' developments for locally stationary processes, but differs in that we assume the evolutionary spectrum to consist of a succession of local spectra f_{θ_t} for the varying values $\{\theta_t\}$ and we chose the prior such that $p(\theta_{1:T})$ forms a Markov chain: $p(\theta_{1:T}) = p(\theta_1) \prod_{t=1}^{T-1} p(\theta_{t+1}|\theta_t)$. This flexible approach allows us to easily incorporate information about the shapes for the possible "local" spectra as well as smoothness or jump in the sequence $\{\theta_t\}$. This, together with the sequential constraint naturally leads us to formulate the related inference about the process as that of a filtering problem in a dynamical system framework. Due to the non-linearity involved, one needs to resort to a particle filter in in order to estimate the sequence of posterior distributions $\{p(\theta_t|y_{1:t})\}$. A crucial point in order to carry out this inference sequentially in time is the ability to recursively estimate the evolving spectrum; this is detailed in section III.

III. State-space formulation and particle filter

A. Statistical model

We assume that the data $\{y_t\} \subset \mathsf{Y}^{\mathbb{N}}$ is generated by the following state-space model

$$\begin{aligned}\log(g(y_t|\theta_t, y_{1:t-1})) \quad = \quad & C - \frac{1}{4\pi} \int_{-\pi}^{\pi} [\log(2\pi f_{\theta_t}(\omega)) + \\ & \hat{I}(t,\omega)/f_{\theta_t}(\omega)] d\omega\end{aligned} \tag{8}$$

for some constant C, a time-frequency estimate $\hat{I}(t,\omega)$ of the the evolving spectrum and a parametric model for the evolving spectra $f_{\theta_t} : [-\pi,\pi] \to [0,+\infty)$ dependent on a parameter $\theta_t \in \Theta$. The sequence $\{\theta_t\}$ is a Markov chain with initial distribution $\theta_0 \sim \mu$ and $\theta_t|\theta_{1:t-1} \sim f(\cdot|\theta_{t-1})$. Note that there are numerous other possibilities for the choice of the log-likelihood, such as $-\int_{-\pi}^{\pi} |f_{\theta_t}(\omega) - \hat{I}(t,\omega)|^2 d\omega$, but that their choice might be more difficult to justify statistically than ours (see section III-E).

Example 1. *Assume that we are interested in tracking chirps, but that we are not given enough information concerning the evolution the instantaneous frequency of the process. In such situations one for example may choose that for* $\theta = (a, \omega^0, s) \in \Theta = [0,+\infty) \times [0,\pi) \times [0,+\infty)$

$$f_\theta(\omega) = a \exp(-(\omega - \omega^0)^2/s)\mathbb{I}(0 \le \omega \le \pi) \tag{9}$$

or

$$f_\theta(\omega) = \frac{a}{1 + (\omega - \omega^0)^2/s}\mathbb{I}(0 \le \omega \le \pi). \tag{10}$$

Then one can suggest the following a priori evolution in time of the parameter of the evolving spectrum, where $\mathcal{N}(.;,\sigma^2)$ *denotes a normal distribution with variance* σ^2*:*

$$a_{t+1}|(a_t, \omega_t^0, s_t) \sim \mathcal{N}(a_{t+1}; a_t, \sigma_a^2)\mathbb{I}(a_{t+1} \ge 0) \tag{11}$$

$$\omega_{t+1}^0|(a_t, \omega_t^0, s_t) \sim \mathcal{N}(\omega_{t+1}^0; \omega_t^0, \sigma_\omega^2)\mathbb{I}(0 \le \omega_{t+1}^0 \le \pi) \tag{12}$$

$$s_{t+1}|(a_t, \omega_t^0, s_t) \sim \mathcal{N}(s_{t+1}; s_t, \sigma_s^2)\mathbb{I}(s_{t+1} \ge 0). \tag{13}$$

Using a model such as that defined above, it is possible to use static inference methods such as MCMC for the inference of the parameters $\{\theta_t\}$. However, our aim is to estimate $\{\theta_t\}$ recursively in time as the data $\{y_t\}$ become available. In our framework all information about $\{\theta_t\}$ is given by the so-called filtering distributions $\{p(\theta_t|y_{1:t})\}$ which cannot be computed analytically and recursively in time due to the intractability of the likelihoods involved. Hence we suggest here to resort to a particle filter algorithm.

B. Particle filters

Particle filters fall in the category of Monte Carlo algorithms, whose principle consists of replacing the difficult to use algebraic representation of a probability density with a non-parametric representation in terms of (dependent) samples from the underlying distribution. The concentration of samples in a particular region of the space is representative of the probability distribution that we are trying to approximate. This turns out to be a powerful principle which has the major advantage of circumventing analytical intractability in complex systems. We now briefly describe how such methods can be used in the present context in order to perform sequential inference.

Assume that at time $t-1$, a collection of N ($N \gg 1$) random samples $\{\theta_{1:t-1}^i | i = 1, \ldots, N\}$, called particles, distributed approximately according to $p\left(d\theta_{1:t-1}\right| y_{1:t-1})$ is available. The empirical distribution

$$\hat{p}^N\left(d\theta_{1:t-1}\right| y_{1:t-1}) = \frac{1}{N}\sum_{i=1}^N \delta_{\theta_{1:t-1}^i}\left(d\theta_{1:t-1}\right) \tag{14}$$

is an approximation of $p\left(d\theta_{1:t-1}\right| y_{1:t-1})$, where $\delta_{\theta_0}\left(d\theta_0\right)$ represents the delta Dirac mass function located in θ_0. Now at time t, we wish to produce N particles which will define an approximation $\hat{p}^N\left(d\theta_{1:t}\right| y_{1:t})$ of $p\left(d\theta_{1:t}\right| y_{1:t})$. The simplest method to achieve this consists of sampling $\theta_t^i \sim f\left(\cdot\right| \theta_{t-1}^i)$. The resulting empirical distribution of the particles $\{\theta_{1:t}^i\}$ is an approximation of the joint density $p\left(d\theta_{1:t-1}\right| y_{1:t-1}) f\left(\theta_t\right| \theta_{t-1})$. We correct for the discrepancy between this density and the target $p\left(\theta_{1:t}\right| y_{1:t})$ using importance sampling. This yields the following approximation of $p\left(d\theta_{1:t}\right| y_{1:t})$

$$p^N\left(d\theta_{1:t}\right| y_{1:t}) = \sum_{i=1}^N W_t^i \delta_{\theta_{1:t}^i}\left(d\theta_{1:t}\right), \tag{15}$$

where

$$W_t^i \propto g(y_t|\theta_t^i, y_{1:t-1}) \text{ and } \sum_{i=1}^N W_t^i = 1. \tag{16}$$

To obtain an unweighted approximation of $p\left(d\theta_{1:t}\right| y_{1:t})$ of the form (14), we resample particles $\{\theta_{1:t}^i\}$ according to probabilities proportional to their weights $\{W_t^i\}$. The underlying idea is to get rid of particles with small weights and multiply particles which are in the regions with high probability masses. Many such resampling schemes have been proposed in the literature; see [31]. The resampling step is crucial for the method to work in practice.

This brief description only covers the simplest particle filtering algorithms since we use nothing more complex than this in our application. For more difficult problems, more sophisticated algorithms will be required (see [32] for a recent review).

C. Computation of the likelihood

There are two issues that require resolution in order to implement the particle filter in practice. Firstly, each evaluation of the likelihood (equation (8)) requires the calculation of an intractable integral. Thus in practice our choice of likelihood is analogous to the numerical approximation used in the stationary case by [24]. In particular, we use a likelihood based on fixed, evenly spaced grid of M frequencies $\omega_k = \pi k/M$ over $[0,\pi]$:

$$\log(\hat{g}(y_t|\theta_t, y_{1:t-1})) = C - \frac{1}{2}\sum_{k=1}^M [\log(2\pi f_{\theta_t}(\omega_k)) + \hat{I}(t,\omega_k)/f_{\theta_t}(\omega_k)] \tag{17}$$

The grid of frequencies must be chosen such that it is fine enough to capture the important features of the modelled and observed spectra. However, M should not be larger than necessary since the sum in equation (17) dominates the computation time of our particle filtering algorithm. Moving further from the statistical justification for our choice of likelihood, one can envisage that for some applications it may be useful to choose the set of frequencies in a different way. For example, in the analysis of long memory processes one may follow [33], omitting higher frequency terms.

The second problem we face in the evaluation of the likelihood is the calculation of the time-frequency estimate $\hat{I}(t,\omega)$ at every iteration. For the method to be computationally feasible, it is required to estimate the evolving spectrum recursively. The method used here is as follows. Consider the PSD of a stationary process (with $\{\bar{Y}_t = Y_t - \mathbb{E}(Y_t)\}$)

$$\begin{aligned} 2\pi f(\omega) &= \mathbb{E}(\bar{Y}_t^2) + 2\sum_{\tau=1}^{+\infty} \mathbb{E}\left[\bar{Y}_t \bar{Y}_{t-\tau}\right] \\ &\qquad \exp(-\imath\omega\tau) \quad (18) \\ &= \mathbb{E}(\bar{Y}_t^2) + 2\mathbb{E}[\bar{Y}_t \sum_{\tau=1}^{+\infty} \bar{Y}_{t-\tau} \\ &\qquad \exp(-\imath\omega\tau)], \quad (19) \end{aligned}$$

where $\mathbb{E}$ is the expectation with respect to the probability distribution of the process. The recursive estimation of the first term from realisations of $\{Y_t\}$ is standard (see the recursion for $\{\Sigma_t\}$ below), but this is far less standard for the second term. However one can notice that the second term can be interpreted as being the product of two random variables, $\bar{Y}_t$ and $B = \sum_{\tau=1}^{+\infty} \bar{Y}_{t-\tau} \exp(-\imath\omega\tau)$ (whose distributions are by assumption independent of the time index t). One can generate random variables asymptotically distributed according to the distribution of B using the following recursion (assuming here for simplicity that $\mathbb{E}(Y_t) = 0$), initialised with $B_1 = Y_1 \exp(-\imath\omega)$ and defined for $t \geq 1$ as

$$B_{t+1} = \exp(-\imath\omega)(B_t + Y_{t+1}). \quad (20)$$

Naturally these random variables are dependent by construction, but under rather mild ergodicity assumptions on $\{Y_t\}$ one can construct the following estimator for $t \geq 2$

$$A_t = \frac{1}{t} \sum_{j=1}^{t-1} Y_{j+1} B_j. \quad (21)$$

This sum can itself be recursively updated as follows

$$A_{t+1} = (1 - \frac{1}{t+1})A_t + \frac{1}{t+1} Y_{t+1} B_t, \quad (22)$$

and the PSD easily estimated by combination with an estimator of the variance.

In the non-stationary case $\{A_t\}$ can be thought of as being "instantaneous" estimates of the cross term in the expression for the PSD. These estimates are likely to have a high variance and hence averaging over several neighbouring time instants might be needed, resulting in the introduction of the classical bias/variance dilemma. An example of such a smoothing procedure is described below, where the time dependent and decreasing stepsize $1/(t+1)$ in the recursion above has been replaced with a time invariant factor $\rho \in [0,1]$, whose role is to discard estimates far in the past. Parameter ρ will control the bias/variance tradeoff. The recursions are initialised with $\mu_1 = y_1$, $A_1 = y_2 y_1 \exp(-\imath\omega)$, $B_1 = y_1 \exp(-\imath\omega)$ and given for $i \geq 1$ by

$$\begin{aligned} \mu_{t+1} &= (1-\rho)\mu_t + \rho y_{t+1} && (23) \\ \bar{y}_{t+1} &= y_{t+1} - \mu_{t+1} && (24) \\ A_{t+1} &= (1-\rho)A_t + \rho \bar{y}_{t+1} B_t && (25) \\ B_{t+1} &= \exp(-\imath\omega)(B_t + \bar{y}_{t+1}) && (26) \\ \Sigma_{t+1} &= (1-\rho)\Sigma_t + \rho \bar{y}_{t+1}^2, && (27) \end{aligned}$$

from which one can evaluate the estimate $\hat{I}(t,\omega)$:

$$\hat{I}(t,\omega) = \frac{1}{2\pi}\left(\Sigma_t + 2A_t\right). \quad (28)$$

Generalisations to multivariate time series, and a time-scale variant, are also possible. A common modification is the introduction of a "lag window" λ into the definition of $B = \sum_{\tau=1}^{+\infty} \bar{y}_{t-\tau} \lambda(\tau) \exp(-\imath\omega\tau)$ to reduce the variance of the PSD estimator (see [34], for example). In this case, only equation (26) changes. For example, if $\lambda(\tau) = \nu^\tau$, we have

$$B_{t+1} = \nu \exp(-\imath\omega)(B_t + \bar{y}_{t+1}). \quad (29)$$

D. *Overall algorithm*

An example of the method described in this section is given in algorithm 1. Here the SIR particle filter [35] is used, with exponential windows in lag and time for the time-frequency estimate. Note that this particular algorithm is presented only for expository purposes - for many applications more sophisticated algorithms may be used.

E. *Discussion of the choice of likelihood*

Our favoured model for the data, in equation (17), is derived through taking several different approximations to the exact likelihood of a locally stationary Gaussian process. Specifically, these approximations are:

1) the use of a Whittle/Dahlhaus approximation, as in equation (6);
2) the use of a numerical approximation to the integral in equation (6), analagous to the one in 3 in the stationary case;
3) substituting the pre-periodogram $\tilde{I}_T(t,\omega)$ in equation (6) for our own recursive estimator in equation (28).

The effect of each of these approximations has been investigated in previous work, most of which focusses on the use of the approximations in the maximum likelihood setting, thus the theoretical results about the approximations only concern the large data limit. In this limit it has been shown in [21] that when the first two approximations are used, the approximate likelihood tends to the exact likelihood. [30] proves an equivalent result when local periodograms are used as an alternative to the pre-periodogram, but no equivalent result yet exists when our recursive estimator is used.

There is a further difference between the approach in [21] and our chosen likelihood, which is that we use a different parameterisation of the plane. In [21] the time-frequency

Algorithm 1 The SIR particle filter applied to sequential inference of a TFR of a process.

Input: A realisation of a time series, $\{y_t\}_{t=1}^{T}$, a set of frequencies $\{\omega_j\}_{j=1}^{M}$, a number of particles P, parameters of the recursive estimator ρ, ν.
Output: A weighted sample $(\theta_{t+1}^{(i)}, w_{t+1}^{(i)})_{i=1}^{P}$ from the posterior $p(\theta_{t+1} \mid y_{1:t+1})$ on receipt of each data point.

$\mu_1 = y_1$;
$\bar{y}_1 = 0$;
$\Sigma_1 = 0$;
for $j = 1 : M$
  $B_1^{\omega_j} = 0$;
  $A_1^{\omega_j} = 0$;
end
for $i = 1 : P$
  Simulate $\theta_1^{(i)} \sim p(\theta_1)$;
  Let $w_1^{(i)} = 1/P$;
end
for $t = 1 : T-1$
  $\mu_{t+1} = (1-\rho)y_{t+1} + \rho\mu_t$;
  $\bar{y}_{t+1} = y_{t+1} - \mu_{t+1}$;
  $\Sigma_{t+1} = (1-\rho)\bar{y}_{t+1}(\bar{y}_{t+1})^T + \rho\Sigma_t$;
  for $j = 1 : M$
    $A_{t+1}^{\omega_j} = (1-\rho)\bar{y}_{t+1}(B_t^{\omega_j})^T + \rho A_t^{\omega_j}$;
    $B_{t+1}^{\omega_j} = \nu \exp(-i\omega_j)(\bar{y}_{t+1} + B_t^{\omega_j})$;
    $\hat{I}(t+1, \omega_j) = \frac{1}{2\pi}(\Sigma_{t+1} + 2A_{t+1}^{\omega_j})$;
  end
  for $i = 1 : P$
    Simulate $\theta_{t+1}^{(i)} \sim p(. \mid \theta_{1:t}^{(i)}, y_{1:t+1})$;
    Reweight $w_{t+1}^{(i)} = w_t^{(i)} \exp\Big(-\frac{1}{2}\sum_{j=1}^{M}\Big\{\log\Big[(2\pi)^{2d}$ $\det f_{\theta_{t+1}^{(i)}}(t+1,\omega_j)\Big] + \mathrm{tr}\Big[f_{\theta_{t+1}^{(i)}}(t+1,\omega_j)^{-1}\hat{I}(t+1,\omega_j)\Big]\Big\}\Big)$;
  end
  Resample $(\theta_{t+1}^{(i)}, w_{t+1}^{(i)})_{i=1}^{P}$;
end

plane is parameterised by a single parameter θ, whereas in our work we use a vector of parameters $\theta_{1:T}$, one for each data point. This is significant since [21] studies the behaviour of the likelihood in the large data limit, with the dimension of the parameter unchanged in this limit. In our approach the size of the parameter increases with the size of the data, thus the theory of [21] does not directly apply to our likelihood. The sense in which our chosen model approximates the exact likelihood of a locally stationary Gaussian process is a topic for future work that would provide statistical justification for our approach.

IV. Application: tracking multiple components

A. Data description and spectrogram

In this section we apply our method to the analysis of two signals, both over the domain $0 \leq t \leq 2$: the first (with a sampling frequency of 10^3Hz) consists of two chirps, one linear and one quadratic, and the second (with a sampling frequency of 10^4Hz) consists of three components, two of which exhibit a frequency modulation. In this section we will apply our methodology to sequentially infer the fundamental frequency of the two chirps.

Spectrograms of these signals are shown in Fig. 1 (a) and Fig. 3 (a) respectively. For the first signal, since the chirps overlap in time-frequency space, separately inferring their fundamental frequencies is not completely straightforward directly from the spectrogram (or other traditional TFRs). However, similar problems to this are often encountered in target tracking and are routinely solved through the use of parametric models. The framework developed in this paper allows us to apply the same approach here. We note that this problem is relatively simple - our reason for including these analyses is expository: firstly as a simple example of the utility of a parametric approach; and secondly since the existence of a ground truth enables a quantitative analysis of the effect of different parameters of the algorithm.

B. Bayesian model

To model the components we parameterise the time-frequency plane so that at each time at which the signal is observed, frequency space is modelled using a mixture of K kernels plus a constant. Specifically we use the model:

$$f_\theta(\omega) = \sum_{k=1}^{K} \frac{a^{(k)}}{1 + \left(\omega - \mu^{(k)}\right)^2 / v^{(k)}} + c \tag{30}$$

The parameter $\mu^{(k)}$ represents the positions of the two components. To allow us to distinguish between the components when they have the same location, we also include the derivative, $\mu'^{(k)}$, of the component position in our model. In this model $\theta = \left(\left\{\mu^{(1)}, \dot{\mu}^{(1)}, v^{(1)}, a^{(1)}\right\}_{k=1}^{K}, c\right) \in \Theta$, where $\mu^{(k)} \in [0, \pi)$, $\dot{\mu}^{(k)} \in (-\infty, +\infty)$, $v^{(k)}, a^{(k)}, c \in [0, +\infty)$.

Our *a priori* model for the evolution of each of these parameters is a random walk, except for the location parameters for which we use a constant velocity (CV) model. We have

$$c_{t+1}|\theta_t \sim \mathcal{N}(c_{t+1}; c_t, \sigma_c^2\tau)\mathbb{I}(c_{t+1} \geq 0) \tag{31}$$

and

$$v_{t+1}^{(k)}|\theta_t \sim \mathcal{N}(v_{t+1}^{(k)}; v_t^{(k)}, \sigma_v^2\tau)\mathbb{I}(v_{t+1}^{(k)} \geq 0) \tag{32}$$

$$a_{t+1}^{(k)}|\theta_t \sim \mathcal{N}(a_{t+1}^{(k)}; a_t^{(k)}, \sigma_a^2\tau)\mathbb{I}(a_{t+1}^{(k)} \geq 0) \tag{33}$$

$$\begin{pmatrix} \mu_{t+1}^{(k)} \\ \dot{\mu}_{t+1}^{(k)} \end{pmatrix} |\theta_t \sim \mathcal{MVN}\left(\begin{pmatrix} \mu_{t+1}^{(k)} \\ \dot{\mu}_{t+1}^{(k)} \end{pmatrix}; A\begin{pmatrix} \mu_t^{(k)} \\ \dot{\mu}_t^{(k)} \end{pmatrix}, Q\right), \tag{34}$$

where

$$A = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \tag{35}$$

and

$$Q = \begin{pmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{pmatrix} \sigma_\mu^2, \tag{36}$$

with $\tau = 0.001$ for the first signal and $\tau = 0.0001$ for the second. We used algorithm 1, and our default parameter settings are to use 500 particles, to take $\nu = 0.97$ and $\rho = 0.97$, and to use $\sigma_c^2 = 10^{-5}$, $\sigma_v^2 = 10^{-3}$, $\sigma_a^2 = 10^{-3}$ and $\sigma_\mu^2 = 1$. These prior parameters are not optimal. They were determined through a consideration of how much the parameters might evolve in one time step and tuned through pilot runs (we note that more accurate results may be obtained by using more particles and appropriately tuned priors). Below we examine the sensitivity of our results to the choice of priors.

C. Results

We first consider the analysis of the first signal. Fig. 1 (b), (c) and (d) display the results of a single run of the algorithm under the default settings. The TFR obtained by the recursive non-parametric estimator, $\hat{I}(t, \omega)$ given by equation (28) and estimated expected reconstruction from the particle filter, i.e. $\widehat{\mathbb{E}}_{\theta_t|y_{1:t}}[f_{\theta_t}(\omega)]$ for each time t are shown in Fig. 1 (b) and (c) respectively. The non-parametric estimator gives comparable results to that of the spectrogram in Fig. 1 (a), and we observe that the components in the Bayesian reconstruction of the TFR follow the chirps closely.

However, the strengths of our approach are not fully illustrated through simply examining the TFRs that are inferred. More detailed information about the signal under study can be obtained through examining the posterior distribution of θ_t itself. In particular, the posterior distribution on the $\mu^{(k)}$ allows us to retrieve estimates of the fundamental frequencies of the chirps over time. Fig. 1 (d) shows the estimated expected fundamental frequencies, $\widehat{\mathbb{E}}_{\mu_t^{(k)}|y_{1:t}}[\mu^{(k)}]$ of the chirps estimated from the particle filter (compared to the true values). We observe that the Bayesian approach has successfully tracked the two overlapping components.

We now explore the dependence of our method on the chosen prior parameters through examining the mean squared error of the posterior expectation of the fundamental frequency of each component over multiple runs of the particle filter. Note that this method is not necessarily the most appropriate way to evaluate a Bayesian technique since the aim of such an approach is not usually to obtain estimators with good frequentist properties, but it does provide a quantitative approach to describing the sensitivity of our method to the choice of prior. Fig. 2 shows the log mean squared error over time of several choices for the different parameters compared to the default choice. In general we observe some sensitivity to prior choice: changing the prior standard deviation by an order of magnitude for any of the parameters can have a large effect on the results, although finer tuning of the parameters was not found to be necessary. These observations are congruent with the situation that is encountered in many other target tracking problems. The effect of altering the prior on the σ_μ^2 parameter is particularly clear. For very small values of the parameter the prior informs the posterior more than the likelihood, and thus the posterior very closely follows the constant velocity model with the result that it completely loses track of the quadratic chirp relatively quickly (by 0.3s). Whereas, for large values of the parameter, large deviations from the constant velocity model are possible, thus the ability of the algorithm to distinguish between the two separate components is diminished, also resulting in a large error.

Finally we consider the analysis of the second signal. Fig. 3 (b) shows the estimated expected fundamental frequencies of the components from the output of the particle filter. We observe that the three components are successfully tracked, along with the frequency modulation that is observed in the spectrogram for the two lower frequency components.

V. Application: flute

A. Pitch transcription

TFRs are a natural tool for the analysis of musical data (for example, [15], [36], [37]). Further, as described in [38], such data is particularly amenable to a Bayesian analysis: often accurate physical models for the notes produced by different instruments are known, thus the use of prior knowledge in parameterising the time-frequency plane is natural. In this section we consider the problem of pitch transcription: estimating the pitch, onset time and duration of notes in a music signal. In the frequency domain a note in a music signal consists of a *fundamental frequency*, determining the pitch of the note, and *partials* or *harmonics* at approximately integer multiples of the fundamental frequency, whose amplitudes dictate the note's timbre. In the polyphonic case, the resultant complex structure in the time-frequency plane makes pitch transcription challenging [16]. Here we consider the monophonic case as a simple illustration of the methodology introduced in the paper. We use a prior model for the time-frequency plane similar to that in [39] and the likelihood described in section III.

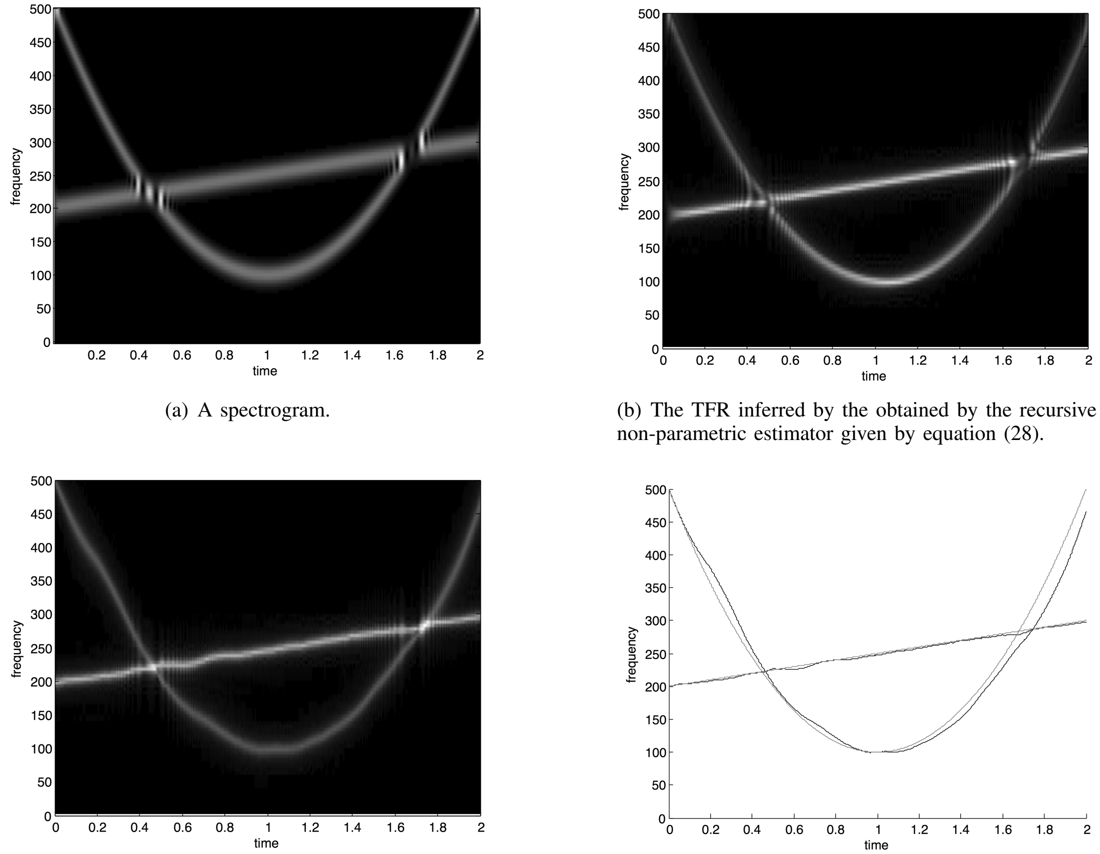
B. Transcription of monophonic flute data

We use our methodology for the transcription of a flute playing the opening twelve seconds of Debussy's Syrinx, available from the first authors's webpage at `http://www.personal.reading.ac.uk/~gt904211/flute.wav`. In frequency space, we choose the following harmonic model for a flute note:

$$f_\theta(\omega) = c + \sum_{k=1}^{K} a^{(k)} / \left(1 + ((\omega - (k + \delta^{(k)})\omega^0)^2 / S)\right). \tag{37}$$

For the data we analyse taking $K = 3$ (so that two partials are modelled) is sufficient to account for the most important parts of the signal (although note that, as expected, higher partials are observed in the data). In this case, $\theta = (\omega^0, S, a^{(1)}, a^{(2)}, a^{(3)}, \delta^{(1)}, \delta^{(2)}, \delta^{(3)}, c) \in \Theta$, with fundamental frequency $\omega_0 \in [0, \pi)$, peak width $S \in [0, +\infty)$, peak amplitudes $a^{(k)} \in [0, +\infty)$, detuning parameters [38] $\delta^{(k)} \in (-\infty, +\infty)$ and constant $c \in [0, +\infty)$.

In passing we note that, as in [38], polyphonic data can be modelled simply by using a sum over several such models (one term per note). Our *a priori* model for the evolutionary



(a) A spectrogram.

(b) The TFR inferred by the obtained by the recursive non-parametric estimator given by equation (28).

(c) The TFR given by the posterior expectation from the particle filter.

(d) The posterior expectations of $\mu^{(1)}$ and $\mu^{(2)}$ estimated from the particle filter output (black) and the true fundamental frequencies (grey).

Fig. 1: Analysis of the first signal.

spectrum of a note is that the parameters in θ each evolve with random walks (some truncated):

$$\omega^0_{t+1}|\theta_t \sim \mathcal{N}(\omega^0_{t+1}; \omega^0_t, \sigma^2_{\omega^0})\mathbb{I}(0 \leq \omega^0_{t+1} \leq \pi) \tag{38}$$

$$S_{t+1}|\theta_t \sim \mathcal{N}(S_{t+1}; S_t, \sigma^2_S)\mathbb{I}(S_{t+1} \geq 0) \tag{39}$$

$$a^{(k)}_{t+1}|\theta_t \sim \mathcal{N}(a^{(k)}_{t+1}; a^{(k)}_t, \sigma^2_a)\mathbb{I}(a^{(k)}_{t+1} \geq 0) \tag{40}$$

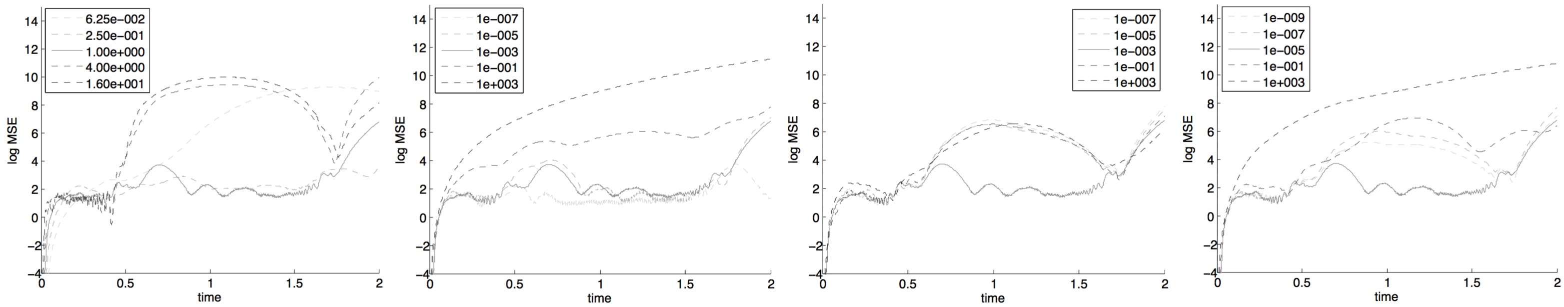
$$\delta^{(k)}_{t+1}|\theta_t \sim \mathcal{N}(\delta^{(k)}_{t+1}; \delta^{(k)}_t, \sigma^2_\delta) \tag{41}$$

$$c_{t+1}|\theta_t \sim \mathcal{N}(c_{t+1}; c_t, \sigma^2_c) \tag{42}$$

The parameters of these priors were chosen in such a way as to allow the note to change quickly enough to describe normal musical data, but not so much as to take account of unrealistic changes (in which case a large number of particles would be needed to obtain accurate results). Specifically, we chose $\sigma^2_{\omega^0} = 10^{-4}$, $\sigma^2_S = 10^{-12}$, $\sigma^2_a = 10^{-8}$, $\sigma^2_\delta = 10^{-16}$ and $\sigma^2_c = 10^{-10}$: these choices are sufficiently small to impose the desired smoothness on the inferred time-frequency plane, whilst still large enough to allow the changes in note that we expect in the flute data. Pilot runs of the algorithm suggest that the sensitivity of our results to these priors is not dissimilar from our observations in section IV, in that intricate tuning is not required in order to obtain adequate results, but pilot runs to check the order of magnitude of the initially specified priors was important. For example, we found that setting $\sigma^2_{\omega^0} = 10^{-2}$ results in the filter losing resolution of the finer features (such as the frequency modulation) of the evolution of the note (although we observed that a change of a similar magnitude to another of the parameters does not have such a dramatic effect). We used algorithm 1 with 100 particles, taking $\nu = \rho = 0.999$. The data was downsampled to 22050 Hz.

The $\log$ of the TFR obtained by the recursive non-parametric estimator (given by equation (28)) and $\log$ of the reconstruction from the expected parameter values found by the particle filter are shown in Fig. 5. In both representations we observe the basic structure of the signal: the rising and falling notes played by the flute; the "steps" representing the individual notes that are played; and the decomposition of the signal into a fundamental frequency and partials (in the non-parametric estimator the $\log$ scale also shows higher order partials that are not included in the model).

The Bayesian reconstruction of the time-frequency plane is smoother than the non-parametric version (as is promoted by our choice of prior), but not so much as to obscure the discrete changes between the notes. Again, we emphasise that a more direct use of the posterior distribution of our

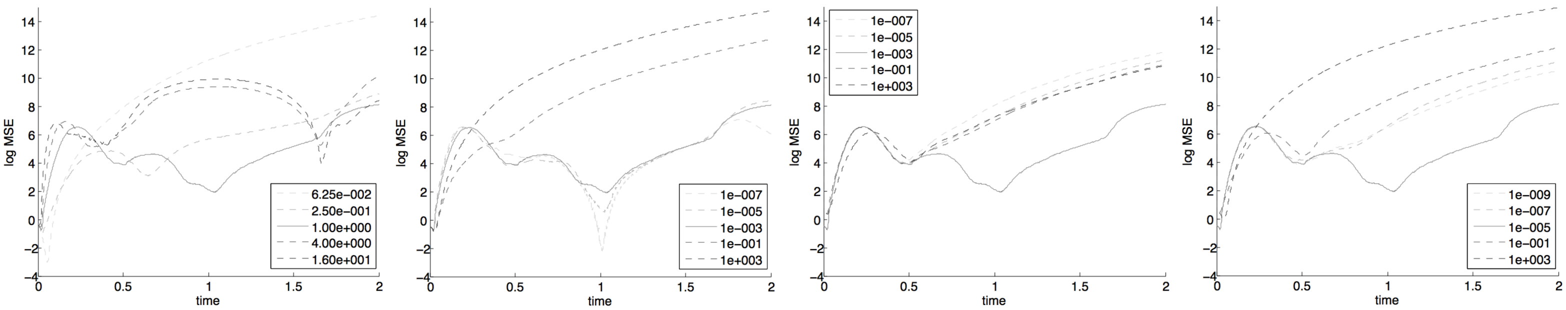


(a) log MSE of the estimate of $\mu_t^{(1)}$ for different values of σ_μ^2.

(b) log MSE of the estimate of $\mu_t^{(1)}$ for different values of σ_v^2.

(c) log MSE of the posterior expectation of $\mu_t^{(1)}$ for different values of σ_a^2.

(d) log MSE of the posterior expectation of $\mu_t^{(1)}$ for different values of σ_c^2.



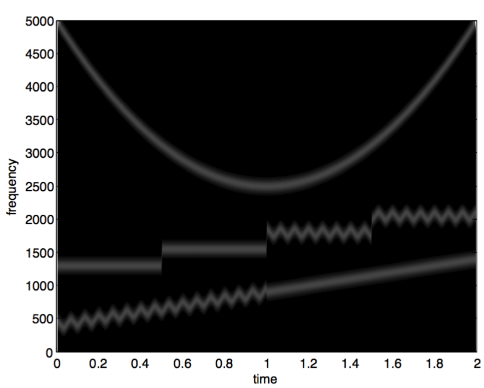
(e) log MSE of the estimate of $\mu_t^{(2)}$ for different values of σ_μ^2.

(f) log MSE of the estimate of $\mu_t^{(2)}$ for different values of σ_v^2.

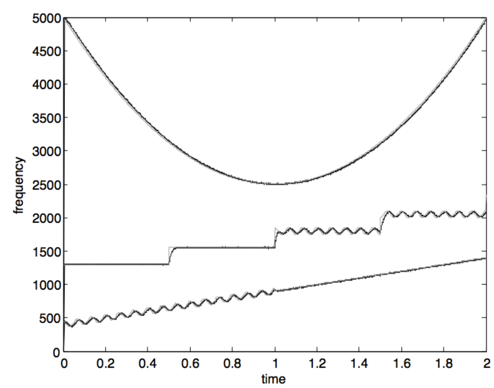
(g) log MSE of the posterior expectation of $\mu_t^{(2)}$ for different values of σ_a^2.

(h) log MSE of the posterior expectation of $\mu_t^{(2)}$ for different values of σ_c^2.

Fig. 2: The $\log$ mean-squared error of the posterior expectations of the fundamental frequency of the linear chirp (top) and the quadratic chirp (bottom) over 50 runs of the particle filter for different prior parameters. In each case the result for the default parameters uses a solid line.



(a) A spectrogram.



(b) The posterior expectations of $\mu^{(1)}, \mu^{(2)}, \mu^{(3)}$ estimated from the particle filter output (black) and the true fundamental frequencies (grey).

Fig. 3: Analysis of the second signal.

parametrisation of the time-frequency plane can reveal a much richer and subtle structure underpinning the signal being analysed. Examination of the posterior distribution of our parameterisation of the time-frequency plane can tell us more about the data. For example, Fig. 6 (a) shows the estimated expected note played by the flute at each time step, $\widehat{\mathbb{E}}_{\omega_t^0|y_{1:t}}[\text{note}(\omega^0)]$, found by the sample mean of note(.) evaluated at the fundamental frequency of each particle, where

$$\text{note}(\omega^0) = 69 + \frac{12}{\log(2)}\left(\log\left(\frac{22050\omega^0}{2\pi}\right) - \log(440)\right) \tag{43}$$

is the function that converts the fundamental frequency of the note into its MIDI pitch number. We observe that our method has successfully tracked the short changes in the note played by the flute, but also see the superimposed modulation that is heard in the audio file towards the end of the longer notes; a feature that is not immediately obvious from the TFRs themselves. Further, our Bayesian approach also allows us to assess the uncertainty associated with such summaries of the data. For example, Fig. 6 (b) shows the posterior uncertainty over the note played by the flute at each time, as represented by the transformation of the fundamental frequency of each particle into its pitch number. Fig. 6 (c) shows the standard deviation of the expected note estimated from 50 runs of the particle filter - we see that this error is small compared to the posterior uncertainty. We find 100 particles to be sufficient for this application - using 1000 particles does not dramatically alter the results.

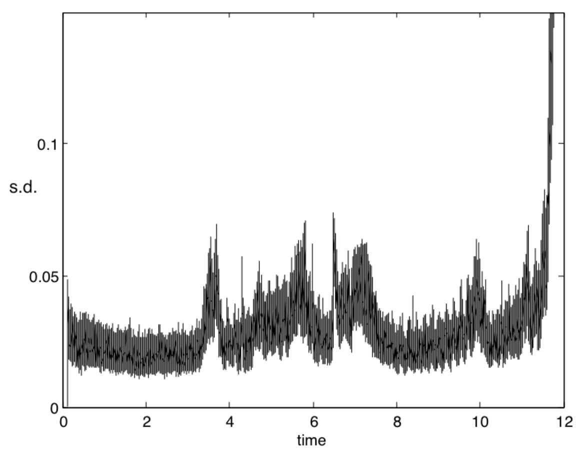


Fig. 4: The standard deviation of the estimated expected note found from 50 runs of the particle filter.

VI. Conclusions

This paper presents a new methodology for the sequential Bayesian estimation of a TFR of a time series, providing a connection between Bayesian estimation and the literature on TFRs. The key underpinning ideas in the work are the use of a likelihood motivated by the extension of the Whittle approximation in [21] and, within this, the use of a recursive estimate of a time-frequency estimate. Our approach has the following features:

1) The Bayesian approach is used, which allows the use of prior knowledge in the modelling of a TFR.
2) Signals are modelled directly in the time-frequency domain in a flexible manner.
3) Inference of the evolutionary spectrum is performed sequentially, which is particularly useful for some applications.

The work in this paper paves the way for a wider use of Bayesian methods in time-frequency analysis. Our framework permits the freedom to use the full power of Bayesian methodology in time-frequency estimation: this should make a significant impact in applications where advanced models (for example, trans-dimensional and/or semi-parametric models) are appropriate.

There are several opportunities for future research arising from this work. We have already mentioned that the computational cost of the algorithm is an important consideration, and we expect parallel implementations to make a significant contribution here. We note that this paper only considers a relatively low-dimensional example, but we also expect the methods introduced here to be of use in high dimensional settings. In such cases, since particle filters can become challenging to implement for high-dimensional problems, the careful design of the particle filtering algorithm becomes more important, but the basic framework remains the same. We anticipate that in such cases, the simple SIR filter will not be appropriate, and the users of the method introduced here will need to draw on the large literature (e.g. [32]) devoted to the design of particle filters in such settings. There is also a clear opportunity for the study of the theoretical properties of our approach. Whilst our local Whittle likelihood is inspired by that in [21], our approach contains several differences and thus the theoretical results in that paper do not apply. We note that in order to formalise the notion of our approach being used to infer a TFR, some constraints on the choice of prior may be necessary; for example, it may be important to represent the well documented Uncertainty Principle between time and frequency [40].

Appendix

Definition of the evolutionary spectrum

The framework is similar to that for stationary processes. Suppose that $\{X_{t,N} \mid t = 1, ..., N\}$ is a Dahlhaus locally stationary process with transfer function matrix Λ^0 and mean function vector μ. That is, directly from Dahlhaus [21], there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp\{\imath\omega t\}\Lambda^0_{t,T}(\omega)d\xi(\nu) \qquad (44)$$

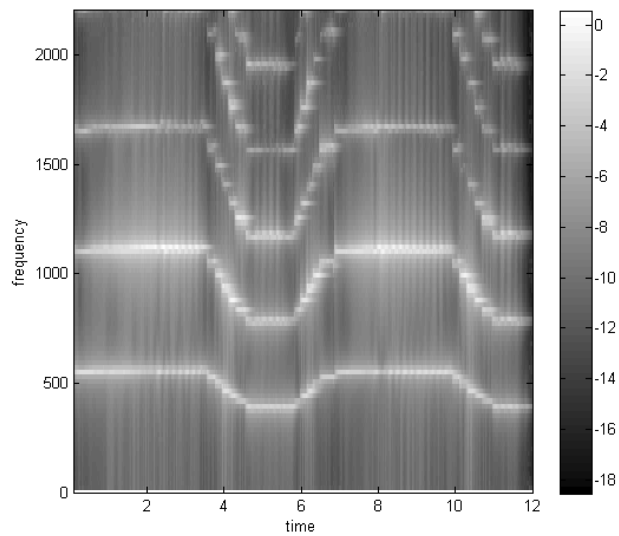
with the following properties:

1) $\xi(\omega)$is a complex valued Gaussian vector process on $[-\pi, \pi]$ with $\overline{\xi_a(\omega)} = \xi_a(-\omega)$, $\mathbb{E}[\xi_a(\omega)] = 0$ and
$$\mathbb{E}[\mathrm{d}\xi_a(\omega_1)\ \mathrm{d}\xi_b(\omega_2)] = \delta_{ab}\eta(\omega_1 + \omega_2)\ \mathrm{d}\omega_1\ \mathrm{d}\omega_2, \quad (45)$$
where $\eta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function.
2) There exists a constant K and a 2π-periodic matrix valued function $\Lambda : [0,1]\times\mathbb{R} \to \mathbb{C}^{d\times d}$ with $\overline{\Lambda(u,\omega)} = \Lambda(u, -\omega)$ and
$$\sup_{t,\omega}\left|\Lambda^0_{t,T}(\omega)_{ab} - \Lambda\left(\frac{t}{T}, \omega\right)\right| \leq KT^{-1} \qquad (46)$$
for all $a, b = 1, ..., d$ and $T \in \mathbb{N}$. $\Lambda(u, \omega)$ and $\mu(u)$ are assumed to be continuous in u.

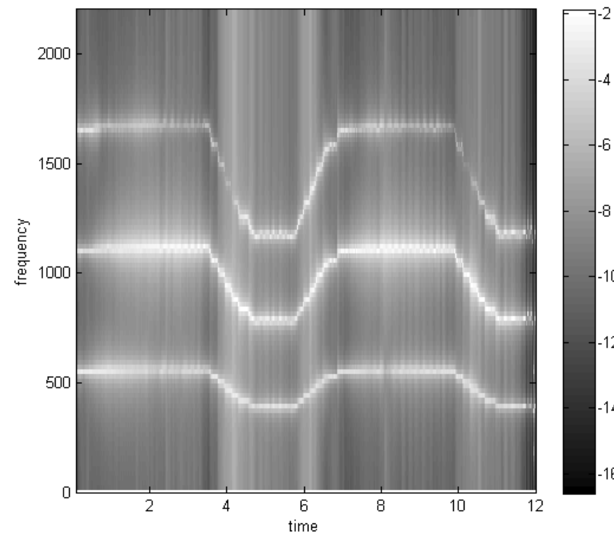
We define that $f(u, \omega) = \Lambda(u, \omega)\overline{\Lambda(u, \omega)}^T$ is the *time-varying spectral density* matrix or *evolutionary spectrum* of the process.

References

[1] W. J. Pielemeier, G. H. Wakefield, and M. H. Simoni, "Time-frequency analysis of musical signals," *Proceedings of the IEEE*, vol. 84, no. 9, pp. 1216–1230, 1996.

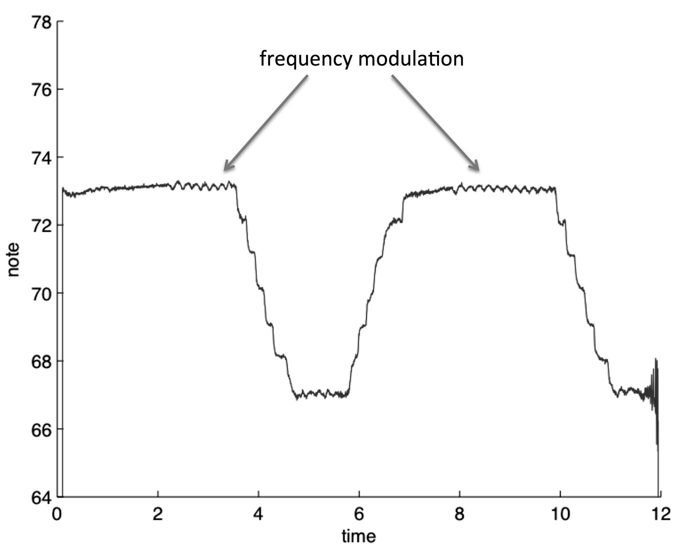


(a) The TFR inferred by the obtained by the recursive non-parametric estimator introduced in section III-C.

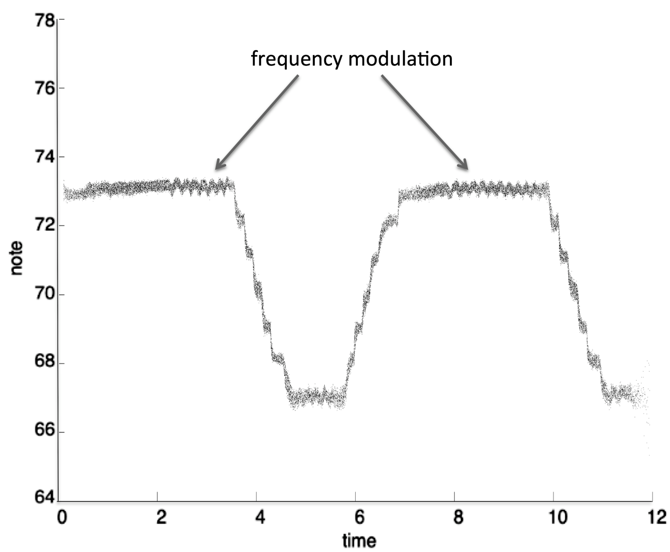


(b) The TFR obtained by taking posterior expectations from the particle filter output.

Fig. 5: TFRs inferred from the flute data.



(a) The expected note played by the flute.



(b) The posterior uncertainty about the note played by the flute, as represented by the output of the particle filter.

Fig. 6: The note played by the flute, estimated from the output of the particle filter.

[2] J. Dripps, “An introduction to time-frequency methods applied to biomedical signals,” in *Time-Frequency Analysis of Biomedical Signals (Digest No. 1997/006), IEE Colloquium on*. IET, 1997, pp. 1–1.

[3] I. Spulber, P. Georgiou, A. Eftekhar, C. Toumazou, L. Duffell, J. Bergmann, A. McGregor, T. Mehta, M. Hernandez, and A. Burdett, “Frequency analysis of wireless accelerometer and emg sensors data: Towards discrimination of normal and asymmetric walking pattern,” in *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2645–2648.

[4] P. Flandrin, “A time-frequency formulation of optimum detection,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1377–1384, 1988.

[5] M. Davy, A. Gretton, A. Doucet, and P. J. W. Rayner, “Optimized Support Vector Machines for Nonstationary Signal Classification,” *IEEE Signal Processing Letters*, vol. 9, no. 12, 2002.

[6] C. Richard, “Time-frequency-based detection using discrete-time discrete-frequency Wigner distributions,” *IEEE Transactions on Signal Processing*, vol. 50, no. 9, 2002.

[7] B. W. Gillespie and L. E. Atlas, “Optimizing time-frequency kernels for classification,” *IEEE Transactions on Signal Processing*, vol. 49, no. 3, 2001.

[8] E. Chassande-Mottin, P. Flandrin, and F. Auger, “On the Statistics of Spectrogram Reassignment Vectors,” *Multidimensional Systems and Signal Processing*, vol. 9, no. 4, 1998.

[9] S. Barbarossa, “Analysis of multicomponent LFM signals by a

combined Wigner-Hough transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, 1995.

[10] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," *Proceedings of the IEEE*, vol. 80, no. 4, 1992.

[11] O. Michel, P. Flandrin, and A. O. Hero III, "Automatic extraction of time-frequency skeletons with minimal spanning trees," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000.

[12] M. Davy and C. Doncarli, "A New Nonstationary Test Procedure for Improved Loudspeaker Fault Detection," *Journal of the Audio Engineering Society*, vol. 50, no. 6, 2002.

[13] P. Bonato, S. H. Roy, M. Knaflitz, and C. J. De Luca, "Time-Frequency Parameters of the Surface Myoelectric Signal for Assessing Muscle Fatigue during Cyclic Dynamic Contractions," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 7, 2001.

[14] F. Hlawatsch and F. Auger, *Time-frequency analysis: concepts and methods*. Wiley, 2008.

[15] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, "Bayesian Variable Selection and Regularization for Time-Frequency Surface Estimation," *Journal of the Royal Statistical Society Series B*, vol. 66, no. 3, pp. 575–589, 2004.

[16] M. Davy, S. J. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2498–2517, 2006.

[17] C. Andrieu, M. Davy, and A. Doucet, "Efficient Particle Filtering for Jump Markov Systems. Application to Time-Varying Autoregressions," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1762–1769, 2003.

[18] C. Dubois, M. Davy, and J. Idier, "Tracking of time-frequency components using particle filtering," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2005.

[19] O. Rosen, D. S. Stoffer, and S. Wood, "Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines," *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 249–262, 2009.

[20] S. Wood, O. Rosen, and D. Stoffer, "Adaptspec: Adaptive spectral estimation for nonstationary time series," University of Pittsburgh, Tech. Rep., 2011.

[21] R. Dahlhaus, "A likelihood approximation for locally stationary processes," *Annals of Statistics*, vol. 28, pp. 1762–1794, 2000.

[22] P. Whittle, "The analysis of multiple stationary time series," *Journal of the Royal Statistical Society, Series B*, vol. 15, no. 1, pp. 125–139, 1953.

[23] K. Dzhaparidze, *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*. Springer, 1986.

[24] C. Carter and R. Kohn, "Semiparametric Bayesian inference for time series with mixed spectra," *Journal of the Royal Statistical Society, Series B*, vol. 59, no. 1, pp. 255–268, 1997.

[25] G. Petris and M. West, "Bayesian time series modelling and prediction with long range dependence," Duke University, Tech. Rep., 1998.

[26] N. Choudhuri, S. Ghosal, and A. Roy, "Bayesian Estimation of the Spectral Density of a Time Series," *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 1050–1059, 2004.

[27] R. Dahlhaus, "On the Kullback-Leibler information divergence of locally stationary processes," *Stochastic Processes and their Applications*, vol. 62, pp. 139–168, 1996.

[28] ——, "Asymptotic statistical inference for nonstationary processes with evolutionary spectra," Universität Heidelberg, Tech. Rep., 1996.

[29] M. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra," *Annals of Statistics*, vol. 25, pp. 38–76, 1997.

[30] R. Dahlhaus, "Fitting time series models to non-stationary processes," *Annals of Statistics*, vol. 24, pp. 1–37, 1997.

[31] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo in Practice*. Springer-Verlag, 2001.

[32] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009.

[33] H. R. Künsch, "Statistical aspects of self-similar processes," in *Proceedings of the First World Congress of the Bernoulli Society*, Y. Prohorov and V. V. Sazanov, Eds., vol. 1. Utrecht: VNU Science Press, 1987, pp. 67–74.

[34] M. Priestley, *Spectral Analysis and Time Series*. Academic Press, 1981.

[35] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings on Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.

[36] A. Klapuri, "Pitch estimation using multiple independent time-frequency windows," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999, pp. 115–118.

[37] A. Klapuri and M. Davy, *Signal processing methods for music transcription*. Springer-Verlag, 2006.

[38] M. Davy and S. J. Godsill, "Bayesian harmonic models for musical signal analysis," in *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, Eds. Oxford University Press, 2003, pp. 105–124.

[39] C. Dubois and M. Davy, "Joint detection and tracking of time-varying harmonic components: a flexible Bayesian approach," *IEEE Transactions on Audio, Speech and Language Processing*, 2008.

[40] P. Flandrin, *Time-Frequency / Time-Scale Analysis*. Academic Press, 1999, Translated from the French by Joachim Stöckler.

Richard G. Everitt Richard G. Everitt received a BSc in Mathematics from the University of Warwick in 2000 and a PhD in Statistics from the University of Bristol in 2008. From 2008-2010 he was a Brunel Fellow in Statistics at the University of Bristol, then from 2010-2012 was a postdoctoral researcher working on statistics and pathogen genomics at the University of Oxford. He is currently a Lecturer in Statistics at the University of Reading, with research interests in Bayesian modelling, Monte Carlo methods, statistical genetics, pathogen genomics, epidemiology, social network analysis and signal and image processing.

Christophe Andrieu Christophe Andrieu graduated from Telecom Paris-Sud and obtained his PhD in statistical signal processing from University Paris XV. He was a research associate in the signal processing group at the University of Cambridge 1998-2001. Since 2001 he has been a faculty member of the Statistics group at the University of Bristol where he is now a professor in Statistics. His research interests are in the area of computational methods for statistics and cognate areas. He has been on the editorial board of JRSS-B and is currently on the editorial board of the Annals of Statistics.

Manuel Davy Manuel Davy graduated in Engineering in 1996 from the Ecole Centrale de Nantes and he obtained his Ph.D. in Signal Processing at the University of Nantes in 2000. He was a research associate at the University of Cambridge from 2000 to 2002, and became a full time researcher at the French National Research Center (CNRS) in the LAGIS laboratory located in Lille, France, from 2002 to 2012. In 2008, he created the start-up company Vekia SAS, an editor of softwares to handle stock and manpower in shops by using forecast sales. The company employs 35 persons. His research interests are Signal Processing, Bayesian statistics, Machine Learning, times series forecasting and audio signals.