Title

Brain Waves Analysis Via a Non-Parametric Bayesian Mixture of Autoregressive Kernels.

Authors

Granados-Garcia, Guilllermo
Fiecas, Marc
Babak, Shahbaba
et al.

Publication Date

2022-10-01

DOI

10.1016/j.csda.2021.107409

Peer reviewed

HHS Public Access

Author manuscript

Comput Stat Data Anal. Author manuscript; available in PMC 2023 October 01.

Published in final edited form as:

Comput Stat Data Anal. 2022 October ; 174: . doi:10.1016/j.csda.2021.107409.

Brain Waves Analysis Via a Non-Parametric Bayesian Mixture of Autoregressive Kernels

Guilllermo Granados-Garcia[a,*], **Marc Fiecas**[b], **Shahbaba Babak**[c], **Norbert J. Fortin**[c], **Hernando Ombao**[a]

[a]King Abdullah University of Science and Technology

[b]University of Minnesota

[c]University of California Irvine

Abstract

The standard approach to analyzing brain electrical activity is to examine the spectral density function (SDF) and identify frequency bands, defined a priori, that have the most substantial relative contributions to the overall variance of the signal. However, a limitation of this approach is that the precise frequency and bandwidth of oscillations are not uniform across different cognitive demands. Thus, these bands should not be arbitrarily set in any analysis. To overcome this limitation, the Bayesian mixture auto-regressive decomposition (BMARD) method is proposed, as a data-driven approach that identifies (i) the number of prominent spectral peaks, (ii) the frequency peak locations, and (iii) their corresponding bandwidths (or spread of power around the peaks). Using the BMARD method, the standardized SDF is represented as a Dirichlet process mixture based on a kernel derived from second-order auto-regressive processes which completely characterize the location (peak) and scale (bandwidth) parameters. A Metropolis-Hastings within the Gibbs algorithm is developed for sampling the posterior distribution of the mixture parameters. Simulations demonstrate the robust performance of the proposed method. Finally, the BMARD method is applied to analyze local field potential (LFP) activity from the hippocampus of laboratory rats across different conditions in a non-spatial sequence memory experiment, to identify the most prominent frequency bands and examine the link between specific patterns of brain oscillatory activity and trial-specific cognitive demands.

Keywords

Spectral density estimation; Bayesian nonparametrics; local field potentials; Dirichlet Process; Markov chain Monte Carlo

1. Introduction

Considerable research indicates that the hippocampus — a brain region highly conserved across mammals — plays a key role in mammals' ability to remember the order in which

*Corresponding author. Tel.: +966 567862693, guillermo.granadosgarcia@kaust.edu.sa (Guilllermo Granados-Garcia).

daily life events occur [1]. However, there is little insight into how these processes are accomplished at the neuronal level.

A major scientific goal of co-author's Fortin research (Neurobiology laboratory, UC Irvine) is to study the ability of animals to remember the specific sequence in which events occurred, and to understand how the hippocampus supports this capacity. To identify the neuronal mechanisms underlying this capacity critical to daily life function, an experiment was conducted in which neural activity, specifically the local field potentials (LFP), was recorded in the hippocampus of rats as they performed a complex, non-spatial sequence memory task (similar to paradigms used in humans (see [2]). The five rats (hereafter referred to as subjects) were trained to recognize a sequence of five different odors (A = Lemon, B = Rum, C = Anise, D = Vanilla, E = Banana). A trial (i.e., a single odor presentation within the sequence) is labeled as "in sequence" (InSeq) if the odor is presented in the correct sequence position (e.g., ABC…); otherwise, the trial is labeled as "out of sequence" (OutSeq; e.g., ABE…), as illustrated in Figure 1. The particular aim was to identify features (e.g., the most prominent oscillatory patterns) in the observed signals that are associated with specific aspects of task performance.

The standard analysis for electrophysiological signals detects the most prominent oscillatory components of the signal by estimating power in the spectral density function (SDF) in predefined frequency bands, namely, delta (0.5–4 Hertz), theta (4–8 Hertz), alpha (8–12 Hertz), beta (12–30 Hertz), and gamma (> 30 Hertz). The power at the delta band indicates the contribution of slow oscillations to the total variance of the signal; whereas the gamma power is the contribution of fast oscillations. Visually, the spectral power forms peaks along the spectral density with certain bandwidths. The SDF is a good descriptor of any stationary stochastic process because it quantifies the amount of variability in a signal (such as the LFP) that is accounted for by the different frequency bands [3, 4, 5, 6].

However, the current segmentation of the frequency range into the delta-to-gamma frequency bands is *ad hoc* and is primarily driven by pragmatic considerations [7]. In current research, many neuroscientists now consider some of these bands to be *too wide* and thus they do not possess the required level of precision in order to identify differences between different type of stimuli and between patient or treatment groups (see [8, 9]). There is an increased demand for more precise analyses with finer subdivisions within the bands, for example as "low"-alpha and "high"-alpha (see [10], [11]). Moreover, these frequency bands are predefined, according to the species under study. This is a limitation because different experimental conditions may elicit different oscillatory activity (different peak locations and different bandwidths).

The plot of the LFP on the second half of the epoch displayed in dark blue and dark red in Figure 2 reveals a highly dynamic pattern of hippocampal oscillations during task performance, reflecting the distinct cognitive demands at different moments in time. Notably, the specific frequencies and bandwidth of the observed oscillations do not map well with standard predefined frequency bands (delta, theta, alpha, beta, and gamma bands). The goal is to identify–in a data-driven manner–the frequency peaks and bandwidth, and link these with specific types of information processing.

There are two general classes of methods for estimating the SDF. One class is modeled in the time domain where a parametric model, typically autoregressive moving average (ARMA), is fitted to the data and an estimate of the SDF is obtained by plugging in the estimates of the ARMA parameters. Another class is nonparametric techniques based on kernel-smoothing or wavelet-denoising of the observed periodogram curves (or log periodogram curves). The main limitations of the standard estimation methods and the general approach where the frequency bands are *a priori* defined are: (1) the lack of direct connection between the SDF estimator parameters to the time domain properties of the signal; (2) imprecise location of the frequencies of the spectral peaks; and (3) the need to increase the number of basic components of a model to satisfy some estimation criterion (e.g., squared estimation error) which leads to an unnecessarily more complex representation (or a less parsimonious representation) of the SDF.

The primary motivation is to address the need for precise identification of oscillations in LFPs by developing the Bayesian mixture auto-regressive decomposition (BMARD) that will produce estimates of the SDF with peaks and bandwidths that are determined by the data – rather than arbitrarily defined by the standard bands. The proposed BMARD method is applied to the hippocampal (LFP) activity recorded from laboratory rats to detect task-related changes in brain signals through changes in the peak activity or shifts in the frequency content. BMARD sets a Dirichlet process prior on the standardized SDF leading to characterization of the SDF as a weighted mixture of kernels. The proposed kernel is derived from the standardized SDF of a second-order autoregressive process, which is associated with a unique oscillatory pattern. The BMARD approach leads to a representation of the signal as a weighted mixture of latent second-order autoregressive processes, where the number of components in the mixture is an unknown parameter in the model. For the BMARD estimator, it will be sufficient to have a small number of kernels in order to obtain a suitable fit. Thus, BMARD provides simple representations of the final SDF estimators, to enable easier interpretation and a more straightforward framework for addressing scientific questions related to the frequency and time domain properties of the data and perform comparisons of experimental conditions. Most importantly, BMARD is not constrained by the a priori frequency bands used in standard analyses; it fully relies on the data to identify the location of the spectral peaks and also the frequency bandwidths of these oscillations.

The remainder of this paper is organized as follows. In Section 2, the general Bayesian non-parametric framework through a Dirichlet Process prior is described and a new kernel is proposed from the standardized SDF of a second-order autoregressive (AR(2)) process. The robustness of the BMARD method to model misspecification is demonstrated through simulation studies in Section 3. A realistic simulation study is conducted, where the observed simulated LFP is a mixture of AR(2) latent processes with peaks that were actually observed in the the LFP signals recorded during the experiments in the Fortin laboratory. In Section 4, the hippocampal LFP signals of five laboratory rats are analyzed using the proposed BMARD method to address important neuroscience questions using a non-spatial sequence memory experiment. Finally, the MCMC algorithm for obtaining posterior samples of the BMARD estimator is described in the Appendix.

2. Bayesian Mixture Auto-regressive Decomposition (BMARD)

2.1. Overview of spectral analysis

To develop the specific ideas of the proposed approach, a brief overview of spectral analysis is given. Consider a process X_t that is locally weakly stationary with zero-mean and autocovariance sequence given by $\{\gamma(h) = E(X_{t+h}X_t),\ h = 0, \pm 1, \ldots\}$ that is absolutely summable, i.e., $\sum_h |\gamma(h)| < \infty$. The SDF (within this stationary epoch) is formally defined as $f(\omega) = \sum_h \gamma(h)\exp(-i2\pi\omega h)$, where $\omega \in (-0.5, 0.5)$. Consider now the observed signal within an epoch $\{X_t\}_{t=1}^{T}$, where T is even and the sample mean $\overline{X} = 0$. A nonparameteric estimate of the SDF $f(\omega)$ obtained from the observed signal are the periodogram values $I(\omega_k) = \frac{1}{T}\left|\sum_{t=1}^{T} X_t \exp(-i2\pi\omega_k t)\right|$, which are computed at the fundamental frequencies $\omega_k = \frac{k}{T}$ (where $k \in \left\{-\left(\frac{T}{2} - 1\right)\ldots, \frac{T}{2}\right\}$. The periodogram is an asymptotically unbiased estimator for the SDF. However, it is not consistent because its variance does not decay to 0, even when the length T of the observed process increases.

One way to construct a consistent estimator for the SDF $f(\omega)$ is by smoothing (or denoising) the periodogram or log periodogram. Several nonparametric methods have been proposed. Bandwidth selection methods for kernel and spline smoothing have been developed for this approach (see for example, [12], [13], and [14]). These nonparametric methods aim to find a spectral estimator that minimizes a well-defined global criterion, such as complexity-penalized deviance or integrated mean squared error. In addition, [15] proposed a Whittle likelihood-based approach. In [16], a functional mixed models approach was developed to account for the variation of the spectra for multiple processes data.

An alternative class of methods has been developed under the Bayesian framework. In [17], a Bayesian method was proposed for estimating the log-SDF which is modeled as a mixture of Gaussian distributions with frequency-specific means and logistic weights. A related approach, using the ideas in Bayesian non-parametric methods, uses kernels that are based on the Bernstein polynomial (BP). This was first proposed in [18] to estimate a probability density function. These ideas were transferred to spectral density function estimation in [19] and [20], where the estimator uses a Dirichlet process (DP) mixture model with BP kernels. The Bernstein polynomial approach was extended to the multivariate processes in [21], which gives a decomposition in the frequency domain in terms of BP-DP approximation. Furthermore, [22] generalized the BP kernels with a procedure using a prior based on B-splines, which reduces the L_1-error. A semiparametric Bayesian estimator for the SDF with a BP-DP prior, along with a nonparametric correction in the frequency domain of a parametric likelihood, was developed by [23].

The mixture models based on BPs have the advantage of providing consistent pseudo posterior estimates. However, in practice, BP mixtures can either oversmooth the peaks of the spectral estimates or require a high number of polynomials to achieve certain levels of accuracy. Another disadvantage of the BP method is their lack of association to an easily interpretable data-generating mechanism in the time domain.

The state-space models (SSM) are a recent advancement on spectral estimation (see [11] and [24]). In [11], an evolutionary SSM is proposed to offer a representation of a weakly stationary process as a mixture of second-order auto-regressive processes. The AR(2) SDF, derived from the phase and magnitude of the non-real complex-valued roots of the AR(2) polynomial function, are assumed to be constant in time. A property of this method is that the discrepancy between the true SDF and the approximate SDF derived from the AR(2) mixture vanishes by increasing the number of components. However, in practice, both the number of components and their phases are fixed. In some practical scenarios, it might be necessary to add components in order to produce better estimates. This is a consequence of the rigidity when the phases of the complex-valued roots are constrained to be fixed. However, in practice, a better solution is to adaptively identify the locations of the various peaks in the SDF, which will produce a more parsimonious model and enable a more precise identification of the spectral peaks. The proposed BMARD method accomplishes these tasks.

Nonparametric methods are flexible but, their main disadvantage is that they generally do not employ a straightforward data-generating mechanism. The parametric approach, on the other hand, is generally efficient, but may not always be directly justified from the underlying brain physiology and may suffer from model misspecification.

BMARD is a Bayesian nonparametric approach for spectral density estimation that combines the best of both time and frequency domains. BMARD decomposes the SDF of a signal as a mixture of spectra of the AR(2) processes, providing a framework that precisely describes each specific oscillatory content in the signal. More precisely, the standardized SDF is modeled as a multimodal probability density function by a DP-mixture model with kernels derived from the standardized SDF of second auto-regressive processes.

BMARD has the advantage of data-adaptively providing an estimate of the number of latent processes with unique locations and scale parameters matching a single peak of the target standardized SDF. Unlike standard methods for spectral estimation, the BMARD method can precisely estimate the frequencies that produce the highest peaks in the standardized SDF without the need to constrain the peak locations to pre-specified bands, and determines the width of each of these peaks (through a bandwidth parameter). Thus, BMARD provides the practitioner with a more informative estimate of the oscillatory activity of brain signals and provides a more direct mapping between physiological signals and animal behavior (or cognitive responses to various stimuli).

2.2. Autoregressive Kernel and the DP mixture

In the proposed model, a kernel g is proposed as the *standardized* spectral density function (SDF) of an AR(2) process, leading to a representation of the observed process as a linear mixture of multiple uncorrelated latent stochastic AR(2) processes, each with unique spectra.

A weakly stationary process Z_t is said to be autoregressive of order 2, AR(2), if it has the representation $Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} = W_t$, where W_t is a white noise process with variance σ_W^2 and the roots of the AR(2) polynomial function $\Phi(u) = 1 - \phi_1 u - \phi_2 u^2$ do not lie on the

unit circle. Furthermore, when the roots of $\Phi(u)$ have magnitudes greater than 1, the AR(2) process is causal. When the roots, denoted as u_1 and u_2, are non-real complex-valued, then they are complex-conjugates of each other, that is, $u_1 = u_2^*$ and $|u_j| > 1$. The above conditions ensures the causality of the process Z_t. When Z_t is causal with non-real complex-valued roots, then both roots have the polar representation

$$u_1 = M\exp(i2\pi\psi) \text{ and } u_2 = M\exp(-i2\pi\psi).$$

with magnitude $M > 1$. The AR(2) polynomial function above can be completely characterized by the coefficients (ϕ_1, ϕ_2), the roots (u_1, u_2), or the magnitude-phase of the roots through the following one-to-one relation between the roots and the coefficients in terms of the log-modulus $L = \log(M) > 0$:

$$\phi_1 = 2\cos(2\pi\psi)\exp(-L), \quad \phi_2 = -\exp(-2L), \quad \psi \in (-1/2, 1/2), \quad L > 0. \tag{1}$$

Due to the symmetry of the SDF $f(\omega)$ at 0, it is sufficient to specify the standardized SDF only at the frequency range $\omega \in (0, 0.50)$. In order to represent the SDF of Z_t as a valid probability density function, it must integrate to 1, which is achieved by scaling by (or dividing by) $\int_0^{1/2} f(\omega)d\omega = Var(Z_t)/2 = \frac{(1-\phi_2)\sigma_W^2}{2(1+\phi_2)\left((1-\phi_2)^2 - \phi_1^2\right)}$. Then the *standardized* SDF is defined as

$$g(\omega; \psi, L) = \frac{2\left(1 - e^{-2L}\right)\left(\left(1 + e^{-2L}\right)^2 - 4\cos^2(2\pi\psi)e^{-2L}\right)}{\left(1 + e^{-2L}\right)\left|1 - 2\cos(2\pi\psi)e^{-L}\left(e^{-i2\pi\omega}\right) + e^{-2L}\left(e^{-i4\pi\omega}\right)\right|^2}. \tag{2}$$

where $\omega, \psi \in (0, 1/2)$ and $L > 0$. Here, ψ is the location parameter of the kernel that attains a localized peak of the SDF of Z_t. The role of L is to control the spread of the kernel as a scale parameter; hence, it will be called the bandwidth parameter. Figure 3 illustrates the roles of ψ and L in producing the different kernels.

According to [3], given an arbitrary stationary time series and $\epsilon > 0$ there exists an AR(p) process such that the absolute difference of their SDFs is less than ϵ for all frequencies. Based on this result, a more formal justification for the selection of a second-order autoregressive model is based on structural process modeling (see [25], [26], [3]), where an AR(p) process is represented in such a way that is equivalent to a consecutive input-output system of simpler processes. If the set of p roots of its characteristic equation consists of p_1 real roots and $2p_2$ complex roots (here, $p = p_1 + 2p_2$), then the characteristic equation is expressed as $(\Lambda - \lambda_1)(\Lambda - \lambda_1^*)\ldots(\Lambda - \lambda_{p_2})(\Lambda - \lambda_{p_2}^*)\ldots(\Lambda - \lambda_{p_2+1})\ldots(\Lambda - \lambda_{p_1+p_2})$, where * denotes the complex conjugate. Then, the AR(p) can be represented through a $p_1 + p_2$ consecutive input-output system composed of p_1 AR(1) processes and p_2 AR(2) processes. Each pair of roots associated to an AR(2) process are constructed with the following coefficients: $\phi_1^{(i)} = \lambda_i + \bar{\lambda}_i$ and $\phi_2^{(i)} = -|\lambda|^2$ for $i \in 1, \ldots, p_2$, while the real root associated to an AR(1) has coefficient $\phi_2^{(j)} = \lambda_j$ for $j \in p_2 + 1, \ldots, p_1 + p_2$. The method proposes

approximation of the AR(1) components by AR(2) processes based on the flexibility to fit the AR(1) SDF illustrated previously when appropriate values for ψ and M are estimated. Using this representation, it follows that the SDF of a weakly stationary process can be approximated by a mixture of AR(2) as latent factors, equivalent to a parallel structural model.

In general, for an arbitrary set of uncorrelated weakly stationary processes $\{Z_t^c, c = 1, \ldots, C\}$ with corresponding identifiable autocorrelation functions $\{\rho_{Z^c}(h), c = 1, \ldots, C\}$ and identifiable SDFs $\{g_c(\omega), c = 1, \ldots, C\}$, their linear combination is still a weakly stationary process. Moreover, define $X_t = \sum_{c=1}^{C} a_c Z_t^c$ then there exist the following relations:

$$\sum_{c=1}^{C} a_c^2 g_c(\omega) = f_X(\omega) \text{ and } \sum_{c=1}^{C} a_c^2 \rho_{Z^c}(h) = \rho_X(h).$$

where $f_X(\omega)$ and $\rho_X(h)$ are the SDF and auto-correlation function of X_t, respectively. Conversely, a decomposition in the frequency domain may have this one-to-one relation only in one direction, since two processes with the same spectrum can be not correlated.

2.3. Dirichlet Process Mixture Model

Let $\{X_t, t = 1, \ldots, T\}$ be the observed time series from a zero-mean weakly stationary process with SDF $f(\omega)$. [27] proposed a quasi-likelihood of the joint distribution of the periodogram values at frequency ω_k, denoted $I(\omega_k)$, expressed as the log-likelihood up to some additive constant.

$$\ell(f \mid X_1, \ldots, X_T) = \sum_{k=1}^{\lfloor (T-1)/2 \rfloor} -\log(f(\omega_k)) - I(\omega_k)/f(\omega_k). \quad (3)$$

where $\omega_k = 2\pi k/T$, $k \in \{1, \ldots, (T/2 - 1)\}$. This quasi log-likelihood is based on the property that for sufficiently large T, then $\{I(\omega_k)\}$ are approximately jointly distributed as uncorrelated exponential random variables with $E(I(\omega_k)) \approx f(\omega_k)$ for $k = 1, \ldots \frac{T}{2} - 1$.

To formulate the model, consider the SDF $f(\omega)$ as an infinite-dimensional object modeled as a mixture of the autoregressive kernels g with respect to an unknown mixing probability distribution G defined over the support of the parameter vector $\theta = \{p_1, \ldots, p_C, \psi_1, \ldots, \psi_C, L_1, \ldots, L_C, C\}$. Then, the mixture model is specified given the vector θ, and a DP prior is set for the mixing measure G.

$$I(\omega_k) \sim \frac{1}{f(\omega_k)} \exp(-I(\omega_k)/f(\omega_k)),$$

$$f(\omega_k) \mid \theta = \sum_{c=1}^{C} p_c g(\omega; \psi_c, L_c),$$

$$\theta \mid G \sim G,$$

$$G \sim DP(G_0, \alpha).$$

The prior $DP(G_0, \alpha)$ refers to the Dirichlet process [28, 29, 30] with parameter α and probability measure G_0. Then, for any finite partition of measurable sets (S_1, S_2, …, S_k), the probabilities ($G(S_1)$, …, $G(S_k)$) have a Dirichlet prior with parameters ($\alpha G_0(S_1)$, …, $\alpha G_0(S_k)$). Due to the symmetry of the SDF, a partition over the interval (0, 0.5) is considered. The parameter α is a scale parameter of the DP that gives an indication of the number of estimated components. Low values of α lead to a posterior distribution of G that is dominated by a few components (see [31, 32]). The prior G_0 is called a base measure and is associated with the prior distributions of the components of θ. The following papers give a detailed description of the role of α in determining the estimated number of components: [33], [34], [35], [32]. From now on, C is referred to as the random number of components with a prior distribution over the positive integers truncated up to a certain high possible value, based on the so-called truncated Dirichlet process (TDP), introduced in [36].

Note that in the usual DP-based clustering model, individual observations are assigned to each cluster with some probability. In contrast, the goal of BMARD is to estimate the standardized SDF $g(\omega) = \frac{f(\omega)}{\int f(\omega)d\omega}$. Since $\int g(\omega)d\omega = 1$, $g(\omega)$ captures the shape of the SDF $f(\omega)$ and gives the proportion of variance explained by each frequency component, the standardized SDF $g(\omega)$ for each of the experiment trials is approximated using the periodogram values computed by first standardizing the trial data to have unit variance. Note that the periodogram values are not being ”clustered” – rather, they are associated with a convex combination of kernels.

The next section constructs a kernel that is based on a parametric model, where the shape of the standardized SDFs is a single peak represented in terms of a location (frequency peak) and scale (frequency bandwidth) parameter in the frequency domain. The new kernel allows us to specify the standardized SDF distribution, F, in the DP mixture model.

2.4. Specification of the prior

The definition of the parametric kernel completes the general specification of the DP mixture model stated in equation 2.3. Now, we discuss the prior distributions of the location and scale parameters of the kernel, as well as the choice of a prior probability mass function for the number of components.

In practical implementation of the MCMC algorithm, the potential label switching problem of the location parameter ψ_c observed in mixture models, see [37], is avoided by defining a random partition over the interval (0, 0.50) as $0 < \epsilon_1 < \cdots < \epsilon_C$. The first and last frequency block are fixed as $\epsilon_0 = 0$ and $\epsilon_C = 0.50$. The partition protects the identifiability of the mixture components since each location parameter ψ_c is restricted over the interval $\epsilon_{c-1} <$

$\psi_c < \epsilon_c$ for $c = 1, \ldots, C$, thus imposing the constraint that there is only one component in each frequency block.

At each iteration of the MCMC algorithm, the random cut-off points ϵ_c are key to the procedure of generating the proper number of components. First, a component label c is selected at random and is proposed to be deleted (death) or to create a new component (birth) with equal probability. When "birth" is chosen, the candidate value for the partition cut off ϵ_* is draw uniformly over the interval $(\epsilon_{c-1}, \epsilon_c)$ and a ψ_* is proposed randomly over the interval $\epsilon_{c-1} < \psi_c < \epsilon_*$. When "death" is selected, then cut off ϵ_c is eliminated and a random ψ_* is proposed.

In the BMARD implementation, a prior for the bandwidth parameter L_c is selected to penalize the full log-likelihood conditional on the DP mixture model parameters to capture sharp peaks. One example of the prior takes the form $L_c \mid C \sim L_c^{\delta}$; note here that $\delta = -2$ defines the Jeffrey's prior. Utilizing this type of prior on the full conditional likelihood has some effect on the identification of the components, as well as the smoothness of the components. In the M-H step, the resulting contribution of a new proposal L^*,— relative to a previous value L— on the log-likelihood is $\delta \log(L/L_*)$, which depends on the sign of δ. If $\delta < 0$ ($\delta > 0$) the algorithm penalizes the M-H acceptance probability if the shape of the proposed peak is sharper (broader). However, if the update steps are small (i.e., $L^* = L + \epsilon$), then this contribution will vanish due to the ratio L/L_* being close to 1.

When the M-H step involves the creation of a new component, then the contribution of a new L_{c+1} is $-\delta \log(L_{c+1})$; if the generated L_{c+1} is close to 0, then the contribution will be substantial. Then, for $\delta < 0$ the penalization is over proposing a higher number of components. The prior for the number of components is chosen from the general form $\pi(C) = \exp(\lambda C^q)$ based on [19], which suggested $\lambda = -0.50$, $q = 2.0$. This prior penalizes decreasing the number of components (in order to guarantee a more precise identification of peaks) since the contribution to the likelihood in the M-H step will be $\lambda(c^q - (c+1)^q) > 0$. This prior will have a more significant impact in cases when the observed signal is composed of low frequencies and when all possible peaks are not well separated. Then the model could collapse to the simplest representation of one component and oversmooth the peaks. Allowing the number of components to increase as necessary helps to detect and isolate peaks that are not easily distinguishable.

Given the prior definitions, the base measure G_0 of the DP Mixture model associated with the prior distribution over the parameter vector $\theta = (\psi, L, \epsilon)$ is as follows:

$$\epsilon_c | \epsilon_{-c}, C \sim U(\epsilon_c - 1, \epsilon_c + 1), \quad \psi_c | \psi_{-c}, \bar{\epsilon}, C \sim U(\epsilon_c - 1, \epsilon_c), \quad L_c \mid C, b \sim L_c^{\delta}.$$

Where b is a user-defined value to set an upper bound of the space of L_c. From Figure 3, the kernel flattens as L_c increases. Another influence on the algorithm performance are small weights that will hinder the M-H steps discard spurious components, in practice setting $b = 1$ allows the kernels to be flexible enough to fit the SDF peaks. The posterior distribution of the DP mixing weights is estimated with the "stick breaking" representation introduced

by [30]. The posterior distribution of the parameter α is sampled based on two different priors: (i) *Gamma* prior (as motivated in [38]) and updating the sample for α_i by a Gibbs sampling through the marginal $\alpha_{i+1}|V \sim \text{gamma}(M+a-1, b-\log(q_M))$, where V is the vector of "breaks" $V_j \sim \text{Beta}(1, \alpha)$, $q_M = V_M \prod_{i=1}^{M-1}(1-V_i)$ and M is the length of V, which is the level of truncation; or (ii) log-normal prior and sampling from the distribution of α using a slice sampler [39] based on the posterior distribution given by [40].

3. Simulation Study

3.1. Setting and criteria

Simulation studies were conducted to examine the relative strengths and weaknesses of the BMARD method compared to three spectral estimation methods: (i) kernel regression smoothing using the Nadaraya-Watson estimator; (ii) a cubic smoothing spline approach and (iii) the non-parametric Bayesian estimator developed by [19] based on Bernstein polynomials. Both (i) and (ii) are optimized with respect to leave-oneout cross-validation (LOOCV) for the smoothness parameters.

The criteria—Three different parametric processes were used (see descriptions below). The methods were compared under the following criteria: (A.) The local integrated absolute error criterion (local IAE). Let ω_{max} be the true value of the frequency at which the true SDF attains a peak and define $f(\omega_{max})$ to be the value of the SDF at the peak. Moreover, define a local interval around the peak to be $(\omega_{max}-\epsilon_1, \omega_{max}+\epsilon_2)$ where $\epsilon_1 > 0$ and $\epsilon_2 > 0$ satisfy $f(\omega_{max}-\epsilon_1) \approx 0.9f(\omega_{max})$ and $f(\omega_{max}-\epsilon_2) \approx 0.9f(\omega_{max})$. Then the local IAE of the estimator $\hat{f}$ around the frequency peak ω_{max} is defined as

$$\text{Local IAE}_{\omega_{max}} = \int_{\omega_{max}-\epsilon_1}^{\omega_{max}+\epsilon_2} \left|\hat{f}(\omega) - f(\omega)\right| d\omega \,. \tag{4}$$

(B.) The maximal-phase disparity criterion which is inspired on [41]. Here, the focus is on identifying the frequency at which the SDF is maximized for a specific band – as opposed to the local IAE criterion which focused on estimating the peak value of the SDF. This criterion was used in particular for the alpha frequency band motivated from cognitive studies where the alpha band is associated with learning. The maximal-phase-disparity is computed as the absolute difference $|\hat{\omega}_{max} - \omega_{max}|$ where $\hat{\omega}_{max}$ is the location of the local maximizer of an estimator $\hat{f}$ within a frequency band b, $\hat{\omega}_{max} = \arg\max_{\omega \in b}(\hat{f}(\omega))$.

The integrated absolute error (IAE) computed through the whole frequency range as a global metric, i.e., it examines the performance across the entire range of frequencies. However, the local IAE measures the performance of the methods only in a local frequency range that contains the spectral peak. On the other hand the phase-disparity helps to evaluate the performance of the estimator to properly locate the peaks of the SDF.

The simulation settings.—The first is a mixture of three AR(2) processes Z_t^c, $c = 1, 2, 3$ with peak locations chosen similar to the ones observed in the LFPs in the data analysis. The frequency peaks were located at 4, 34, and 60 Hz (assuming that the sampling rate is

1000 Hertz per second but only half-second worth of data (i.e., $T = 500$) is observed which mimics the rat LFP data analyzed in Section 4.

The peaks are associated to the following weights $p_1 = 0.1$, $p_2 = 0.6$, $p_3 = 0.3$. Moreover, setting $L_c = 0.01$, $c = 1, 2, 3$, the peaks in the SDF will be sharp. The construction simulates realistic brain signals since a common component at very low frequencies, a second peak around 30 Hz, and a third peak representing an artifact at 60 Hz were observed in a exploratory analysis. Each pair of values (ψ_c, L_c) defines a unique AR(2) process Z_t^c.

The goal of the next two settings is to test the robustness of BMARD with respect to a deliberately misspecified parametric model. The second simulation setting is an AR(12) process that was studied in [14] to test the smoothing splines estimator of the standardized SDF. Moreover, the same setting defined in Equation 5 was also studied in [19] using Bernstein polynomials to estimate the AR(12) standardized SDF:

$$X_t = 0.9X_{t-4} + 0.7X_{t-8} - 0.63X_{t-12} + \epsilon_t, \tag{5}$$

where ϵ_t is a white noise process. The setting is useful to test the robustness of the BMARD method under model misspecification since the true process is not a mixture of $AR(2)$ processes. The scenario examines how well the mixture can approximate a model with 3 main peaks at $\omega = 0$, 250, 500 Hz and two smaller peaks at $\omega = 150$, 350 Hz. As a side remark, [3] explain that higher order AR models can approximate the SDF of any arbitrary stationary linear process which motivates the importance of estimating this type of model. The third setting is a MA(4) process generated as

$$X_t = -.3\eta_{t-4} - .6\eta_{t-3} - .3\eta_{t-2} + .6\eta_{t-1} + \eta_t, \tag{6}$$

where η_t is a white noise process. Similar processes are discussed in [14], [12], [13], [42], and [43]. The standardized SDF of this moving average process is a smooth curve centered at $\omega = 250$ Hz with an extra bump around 500 Hz. This model would help to test BMARD under the scenario of fitting broad and smooth peaks, when the misspecification is not only in terms of the order but also in the structure of dependency since the MA processes have a zero correlation beyond the order (or when the absolute value of the lag exceeds the order).

To evaluate the performance of BMARD, 1000 processes were generated per setting each of $T = 500$ time points. The data settings match the window size in the LFP data analysis. The spectral spline estimation was deployed using the package connection between R and C++: Rcpp [44], [45], [46] in order to boost its efficiency, with the number of MCMC samples fixed to 100000 for six chains discarding 95000 as burn-in samples.

The following process was performed to summarize the MCMC samples containing different dimensions at different iterations across chains. First, for each parameter, the after burn-in samples are stacked in a vector and merged across chains. After making these vectors for each parameter ψ, L, W, the vectors are bound into a matrix of three columns generating a cloud of points in a 3-dimensional space. Then, a clustering algorithm is run to identify all configurations found by the MCMC. The number of components C from the after burn-in samples are summarized in their unique values to consider all possible

components found in the MCMC. These unique C values are the input for the classification algorithm for the possible clusters. The method chosen to classify the components is the mixture of Gaussians from the package mclust [47]. This clustering algorithm is optimized based on the integrated complete likelihood (see [48]). Finally, the cluster means $\bar{\psi}$, $\bar{L}$, $\overline{W}$ are used as the estimated components. To obtain the curves summary the pointwise median of the sampled curves is computed considering 5000 after burn-in samples and all MCMC chains.

The reported results are based on the implementation based on the gamma prior for parameter a, the initial number of components was randomly selected in the set {1, …, 20} for each chain to start with different initial conditions aiming to convergence to the same posterior distribution. The level of truncation for the stick breaking representation was set random per each chain in the set {20, …, 30} as a conservative rule based on [19].

Roughly BMARD takes 4 hours to finish all chains running in parallel. In comparison, the BP method can take 10 hours to run all chains in parallel and use a more considerable amount of memory when the number of components increases. However, a computational efficiency study was not carried out rigorously. It was also noticed that the number of atoms used in the DP prior for both methods was not a significant factor for the computational times.

3.2. Results

Figure 4 displays the logarithm of all the pointwise median curves per simulated processes for each of the methods. For the BMARD estimators the associated peak locations is added in green vertical lines with its linewidth determined by the mixture weights. The first row corresponds to the AR(2) mixture setting. The results demonstrate that the BMARD method retrieves better the true standardized SDF peaks locations and shapes than the other methods. It is also evident that the Bernstein polynomial method consistently smooths out the peaks. The spline and kernel estimators identify the peaks in different simulations but, in general, show higher variability capturing the main components, and more peaks across all frequencies.

The results from modeling the AR(12) process standardized SDF shows how the BMARD method out-performs the Bernstein polynomial method at retrieving the shapes of the peaks. The curves in log-scale indicate that when the SDF contains well-spaced and sharp peaks, the BMARD method produces better estimates. Regarding the algorithm sensitivity, different values were tested for δ and λ for the bandwidth prior, and the level of truncation of the DP prior. The BMARD curves across chains generally consistently converged to similar curves and parameter settings.

In the MA(4) setting, the BMARD method was able to locate the peaks around the maximum of the true standardized SDF. However, it requires several components to approximate the smooth shape of the target standardized SDF due to the convexity of the autoregressive kernel - whereas the shape of the modes of MA(4) standardized SDF is concave. This behavior of the BMARD method is pointed out as a limitation if there is an interest in the shape around the peak (not only the actual location of the peak). The cubic

spline estimator has a better performance followed by the Kernel smoother, showing as well higher variability on the curves observed in Figure 4.

Tables 1–3 are computed using a topological data analysis (TDA) filtration to detect the number of peaks of the estimated curves. The filtration process finds portions in the curve where a peak appears called the rips filtration computed through the TDA package [49]. According to [50], the rips filtration detect the peaks, but their location is missed. Therefore this process is used only to count the number of peaks. The intuition behind this evaluation is that the wiggliness of the estimated curve can lead to spurious peaks, which is a false positive of activity detected in the signal. The BMARD method was tested under the TDA approach and by the MCMC iterations estimating the number of components. The Metropolis-Hasting steps can keep some components containing minimal weights because its deletion does not affect the Whittle likelihood value. Thus, the components considered are those with weights larger than 1%. Table 1 shows that BMARD based on its MCMC summary, finds three components for 43% BMARD of the estimations, while in 27% it finds two peaks. The TDA analysis of BMARD counts that in half of the simulations were found two peaks and three peaks in 25% of the cases. In contrast, the BP method in most of the simulations oversmooth the periodogram into a single peak activity. The splines and kernel smoothing method estimate a single component around 60% of the simulations; however, these methods retrieve more than five peaks in several estimations. The BMARD method performance for the AR(12) process shown in Table 2 demonstrates to fit five components according to the TDA filtration, while the estimated number of peaks based on the MCMC summary is distributed between three to five components. The consequence of the wiggliness shown in Figure 4 for the competitors is the high number of peaks detected in the estimated curves. The BP method, in some cases, reaches to generates more than seven peaks. The kernel and spline have the worst performance since, in most cases, these smoothing methods estimate more than ten components. The filtration applied for the MA(4) SDF estimates confirms from Figure 4 the BP method has the best performance for this SDF with smooth curvature since it predominantly detects three components. The BMARD method estimates between three to five peaks which extra components are mainly located at around 250 Hz. The kernel and spline methods have mixed performance giving smooth estimates, but in some simulations reaching to estimate more than seven components.

The local error measures were compared mainly for the AR(2) mixture. The local IAE was similar for all methods, most likely because a window of only 500 points was sampled while the sampling rate is 1000 Hz, leading to compute the local IAE with only three frequencies within the band for each peak. The Bernstein method has a lower local error in the first two peaks, at 8 and 30 Hz, while the BMARD method and the classical smoothers behave similarly. For the last peak at 60 Hz, the BMARD method achieved a lower local error in several simulations. This pattern leads us to conclude that when the peaks occur in closely-placed frequencies, implying potentially indistinguishable periodogram peaks, the BMARD method tends to identify them as a single peak, which turns out in higher error. Of course this problem can be alleviated by increasing the number of observations but keeping the sampling rate fixed, i.e., by observing the data for a longer physical time. In contrast, when there is sufficient space between peaks (higher frequency resolution), the BMARD method can identify the peak activity even when the contribution to the total SDF is small.

Each of the single MCMC chain estimates of the parameters (ψ, L, p) are used to evaluate the average of the absolute for the maximal-phase disparity of the estimates compared to the true values. Only the chains that correctly estimate the true number of components are considered to compute the average difference across the six chains over the 1000 simulations. Table 4 reports the average absolute disparity and its standard deviation across the 1000 simulations for all the parameters of the individual components. The interval of the mean disparity plus two standard deviations contains 0 for the location parameters ψ_c and the bandwidth parameters L_c, $c = 1, 2, 3$, which leads to conclude the estimation of both parameters is unbiased, for all components. However, the weights errors show the estimations differ from their true values, even when the shapes in the log scale of the mean curves in Figure 4 demonstrate the estimated curves are close enough to the true SDF.

This behavior is due to having two parameters that contribute to the scale of the individual components of the mixture since the bandwidth L_c at lower values narrows the Kernel implying a higher maximum. On the other hand, the weight p_c directly shrinks or expands the scale of the c–th autoregressive kernel in the mixture. In further joint analysis of the parameters, is noted that the overall shape and contribution to the SDF estimation of the individual components is more sensitive to the bandwidth values.

To assess model fit, the convergence of the Whittle log-likelihood was investigated to ensure the stationarity of the MCMC for each of the generated datasets and each chain run. The evolution of the Whittle likelihood (not shown here) displayed an initial increase followed by a stationary behavior for all chains before reaching 50,000 iterations of the MCMC. Convergence of the Whittle log-likelihood to a stationary pattern was also observed for the Bernstein polynomial method.

An evaluation to find the location of the central peak of the misspecification scenario was performed for all methods. The argument at which the maximum value is reached within a fixed window of 10 HZ (240–260 Hz) was computed for all the estimators. The densities of the absolute disparity between the local maximum location and 250 Hz are shown in Figure 5. All methods have similar performance; however, BMARD has a higher density around 0, which is associated with a slightly better performance among the considered methods.

In summary, the simulations show that the BMARD method provides better estimates of the SDF when the true underlying process has auto-regressive (even higher order). However, it is less desirable than the nonparametric Bernstein polynomial method to fit the MA(4) SDF. Based on the AR(2) mixture simulations, the BMARD method was able to identify the SDF peaks more accurately compared with other methods as the BMARD estimated curves neither oversmooth nor overfit the periodogram. The higher accuracy of identifying peaks and bandwidth in the SDF is the central contribution of the BMARD, which directly addresses the current limitations of spectral analysis of electrophysiological signals in neuroscience.

4. Analysis of hippocampal LFPs from 5 rats

In the experiment performed in the Fortin lab, each subject is trained to identify InSeq trials by holding its nose in the port for 1.2s (when an auditory signal is delivered), and OutSeq trials by withdrawing its nose from the odor port before 1.2-seconds. The LFPs were recorded from 20 electrodes (tetrodes) positioned in the pyramidal layer of the dorsal CA1 region of the hippocampus measured at a sampling rate of 1000 Hertz (1000 time points per second).

The data was first examined in [10] where the authors analyzed the LFP activity using predefined frequency bands (4–12 Hz and 20–40 Hz). The authors found that, as animals ran toward the odor port, power was high in the 4–12 Hz band, particularly in the 7–10 Hz range. Upon odor delivery (when animals were immobile with their nose in the port), that oscillation seems to reduce in frequency (stronger in the 5–8 Hz range). Power in the 20–40 Hz range increased during odor presentations (particularly in the 19–35 Hz range), but was weak during the running period. Notably, 20–40 Hz power showed an association with session performance (higher in sessions with high performance). It also differed between InSeq and OutSeq trial types (higher on InSeq trials), although that analysis could not completely rule out the effects of uncontrolled differences in the animal's behavior. Clearly, such dynamic and frequency-specific (narrow band) patterns analysis in the LFP is necessary but cannot be derived from standard analyses using broad, predefined frequency bands.

To address this limitation, the BMARD posterior curves are used to conduct inference over all the frequencies and locate those with significant changes in power across experimental conditions. In the analysis, the first odor was omitted because, regardless of the sequence, odor A is always presented first (and hence always in correct order). In order to focus on the interaction between stimulus condition and temporal context, the trials in which a subject made the wrong response were excluded since different and more complex brain processes is expected to be present over wrong responses. The number of the trials during the experiment is displayed in Table 6.

To identify the LFP dynamics associated with the processing of the odor stimuli the analysis is focused on a single electrode aligned at the same brain location for all subjects for two time periods: a Pre-Odor baseline period (500 ms before odor presentation), and an odor period (focusing on the first 500 ms following onset of odor, during which the animal's behavior is consistent across trial types). LFPs are generally non-stationary but it is reasonable to model each of the LFP recosrds to be locally stationary and hence quasi-stationary within the very brief intervals of 500 milliseconds. Separate analysis on a single electrode using BMARD consistently retrieved the peak activity observed in the observed periodogram curves. It is also of interest to consider the joint variability in oscillatory behavior across different tetrodes. Indeed there is a keen interest in the community to study potential lead-lag relationships between tetrodes and also various types of spectral dependence between pairs of tetrodes including coherence [51], partial coherence [52, 53], partial directed coherence [54], copulas [55], and Hierarchical Bayesian models [56]. Future work will be on the generalization of BMARD for multivariate models. Under this

framework methods for inference on cross-tetrodes connectivity will be developed in further BMARD extensions.

The BMARD was used to explore the posterior curves of each periodogram running eight chains of size 100000 considering a burn-in period of 90000 samples. The initial number of components (of latent AR(2) processes) were randomly set between 10 and 20. The sampler used for the posterior distribution of α is based on a gamma prior with initial parameters $a = 0.10$, $b = 0.10$ to set a less informative prior. The DP parameter was set to a initial value of $\alpha_0 = 1.00$. The level of truncation of the stick-breaking representation was selected randomly between 20 and 30. The final estimated curves were computed as the point-wise median of 10000 after burn-in posterior curves. With respect to the model fit assessment, the log-likelihood trace plots shown an increase and convergence to a stationary behavior for all the different trials, stimulus conditions, and temporal contexts.

In the analysis, the LFPs from all trials were used and was applied the two-stage approach for estimating the SDF separately for the Inseq and Outseq conditions. In the first stage, the SDF was estimated separately for each trial; in the second stage, the information across trials within each of the Inseq and Outseq condition was combined. There are many possible ways to obtain some ”summary” across the SDFs including (a.) functional median curve [57] (the option used); (b.) point-wise median for each frequency; (c.) weighted average of all trial-specific SDF estimates where the weight is inversely proportional to the variance of the estimate (a spectral curve estimate derived from a very noisy trial should have low weight). Regardless, the summary tells about the center of the distribution of the true SDFs across all hypothetically infinitely many trials.

Some of the major questions posed in the Fortin laboratory answered by BMARD are the following: (a.) Are there differences in peak activity between stimulus conditions (Pre-Odor vs. odor) on *Inseq* trials; (b.) Are there differences in peak activity between stimulus conditions on *Outseq* trials; (c.) Is there a potential interaction between stimulus condition and temporal context, i.e., is the difference between the Pre-Odor and the odor periods the same across Inseq and Outseq trials?

The distribution of the peak frequency of each subject is summarized by considering the AR(2) component with the highest estimated contribution to the variance for each trial. More specifically, for each subject, trial, and temporal context, the component with the biggest estimated weight is selected. Usually their values were at least 65% with high concentration around 90%. The main-peak activity subject-specific distributions (derived across all trials for each combination of pre vs post-odor and Inseq vs Outseq conditions) are shown in Figure 6.

The Inseq trials (top row) for each of the 5 subjects, show first for subject 1 (S1), the distribution of the peak frequency Pre-Odor is unimodal with support over 0–25 Hertz; for post-odor the distribution of the peak frequencies is also unimodal with the same support. The main difference between the Pre-Odor and post-odor for the distribution of the peak frequencies during the Inseq condition is the mode: it is 8 Hertz for Pre-Odor while it is higher at 10 Hertz for post-odor. For subject 2 (S2), the distributions are unimodal for both

pre- and post-odor but the support is narrower with concentration on 0–18 Hertz. The mode for the peak frequency for Pre-Odor is 6 Hertz, while it is 8 Hertz for post-odor. For subject 3 (S3), the support is even narrower with concentration on 0–12 Hertz and the modes are almost identical for Pre-Odor and post-odor peak activity at approximately 6 Hertz. Subject 4 (S4) has a similar support as subject 1 (S1) but displays a unique feature because the mode for the Pre-Odor is 7 Hertz which is higher than that for the post-odor which is at 5 Hertz. For subject 5 (S5), the mode for the peak frequency at Pre-Odor is 5 Hertz vs 7 Hertz for the post-odor. Note that for the Inseq condition there is quite a variation in the brain functional response across the 5 subjects. For this reason it is not possible find a solid justification for developing a single unifying model for these 5 subjects. Thus, the individual modeling for each subject should be performed to describe similarities and differences in the results across the subjects.

A formal test was conducted for the hypothesis of equality of the distributions of Pre-Odor vs post-odor using the Kolmogorov-Smirnov method. These tests were conducted separately for each of the 5 subjects and the findings shown on each of the distribution graph in Figure 6. Under a 5% of significance level, S5 during InSeq trials is the only subject with significant difference of the pre and post distributions (pvalue=4.5×10^{-5}). Note that subjects S1, S2, and S3 clearly suggest equality of distributions while for S4 is not significant despite the bimodality of the post-sequence distribution. Note that the identification of these precise peaks were made possible because the BMARD method gives a representation of the SDF in terms of the building blocks which are the AR(2) spectra. These precise differences in the modes would not have been detected using standard approaches where the frequency bands were predetermined rather than adapted to the specific data that is being analyzed.

The distribution of peaks for the Outseq condition looks different from the Inseq condition uniformly across the 5 subjects. Some evidence of bimodality and greater visual separation between the distribution of the Pre-Odor and post-odor presentation is observed uniformly across the 5 subjects. Most notable again is S5 (fondly called ”Super-Chris” in the laboratory) whose distribution of Pre-Odor and post-odor peaks have a different support - despite the estimated modes being very similar. For Super-Chris, the distribution of the peaks for pre-odor is very tight from roughly 4–10 Hertz while the post-odor peak has more variation with the spread from approximately 4 – 16 Hertz. As in the Inseq condition, a Kolmogorov-Smirnov test for the equality of the distributions of peaks for pre-odor vs post-odor was used. These tests were conducted separately for each of the 5 subjects which results can be found on the side of each graph of the bottom row of Figure 6 where now subjects S3 and S5 display significant differences(S3: pvalue=0.035, S5: pvalue=0.015) among distributions while subject S1, S2, and S4 show similar test outcomes as in the InSeq trials.

To address questions (a.) and (b.) above, let the frequency-specific difference in the SDF within InSeq trials and within Outseq trials be, respectively,

$$\Delta^{I}(\omega) = f_{A}^{I}(\omega) - f_{B}^{I}(\omega),$$

$$\Delta^{O}(\omega) = f_A^O(\omega) - f_B^O(\omega),$$

for all $\omega \in (0, .5)$ where $f_A^I(\omega)$ and $f_B^I(\omega)$ are the SDF for Inseq trials for, respectively, the Pre-Odor and post-odor presentation. Thus, the quantity $\Delta^I(\omega)$ measures the extent of the frequency-specific change after the subject detects an odor presented under the setting of correct sequential order. The SDFs for Outseq trials the frequency-specific change $\Delta^O(\omega)$ are defined in a similar manner. The functional boxplots for the pre-odor vs odor differences on Inseq and Outseq trials ($\Delta^I(\omega)$ and $\Delta^O(\omega)$) are given in Figure 7 where is noted positive differences for higher frequencies, consistent with Figure 6 on the activation of this frequencies after the stimulus is presented to each subject. It is stronger in the InSeq trials shown on the top row. Besides, some peaks on the boxplots between 8 and 14 Hz aligned with the mode of the main peaks locations.

The next question of (c.) regarding the interaction between stimulus conditions and temporal context is now addressed. Indeed, a natural question posed by the neuroscientists is whether the change (pre vs post-odor) differs between the Inseq and Outseq conditions. In particular, the task is to identify the specific frequencies (or bands) where the changes (pre vs post-odor) are more highlighted for Outseq trials (and which are more emphasized for Inseq trials). To conduct a formal inference on the interaction, let define

$$\Delta^{I-O}(\omega) = \Delta^{I}(\omega) - \Delta^{O}(\omega).$$

For a particular frequency ω^*, when $\Delta^{I-O}(\omega^*) > 0$ then the change in pre vs post-odor for the Inseq condition is greater than that for the Outseq condition. This will be important for identifying physiological features in signals that differentiate between the two experimental conditions. In the implementation, the last MCMC 5000 posterior samples were extracted for each of the conditions before described for all trials.

The use of the proposed "difference of the change" $\Delta^{I-O}(\omega)$ can also be interpreted from another point of view. For example, when $\Delta^{I-O}(\omega) > 0$ then $\Delta_i^I(\omega) > \Delta_j^O(\omega) \Rightarrow f_A^I(\omega) - f_B^I(\omega) > f_A^O(\omega) - f_B^O(\omega)$. This can be rewritten in another form and thus leads to the interpretation

$$f_A^I(\omega) - f_A^O(\omega) > f_B^I(\omega) - f_B^O(\omega).$$

The quantity $f_A^I(\omega) - f_A^O(\omega)$ measures the difference between the spectral power for Inseq vs Outseq conditions during the Pre-Odor presentation; whereas the $f_B^I(\omega) - f_B^O(\omega)$ measure the difference between Inseq and Outseq during the post-odor presentations.

The posterior inference of the curve $\Delta^{I-O}(\omega)$, shown in Figure 8, displays the functional boxplot from the *fda* package in R set to show the 95% internal region. For subjects S1, S3 and S4, is observed a similar pattern of decay for frequencies higher to 8 Hz but only

S3 shows significant differences below 0 for frequencies between 12 and 24 Hz. subject 4 has a significant $\Delta^{I-0}(\omega)$ for 22 <= ω <= 32Hz while the significant differences for S4 appear only for $\omega < 8Hz$. A different behavior is noted for S2 who shows positive $\Delta^{I-0}(\omega)$ for $\omega > 32$Hz. Super-Chris (S5) shows quite a distinct pattern. Most notably, the change for pre-vs-post during Outseq is significantly greater than the change during Inseq at a very narrow band around 6–10 Hertz. Thus, the BMARD method produced a highly specific band which identifies the difference in the brain reaction to a correct sequence vs incorrect sequence of the odors.

The potential impact to neuroscience brought by the new findings obtained from BMARD includes are as follows: (a.) identifying the specific frequency (or narrow bands) of the most dominant neuronal oscillations that are engaged in memory; (b.) leading to new sets of hypothesis about memory and designing new experiments that test for intervention effects such as applying electrical stimulation at the identified frequency peaks.

5. Conclusion

The primary contribution of the BMARD method is to extract information from the data to provide highly specific frequency information about spectral density function. It gives the estimated number of spectral components (peaks) and provides precise identification of the dominant frequencies where the SDF attains localized peaks. It also gives the corresponding spread (bandwidth) for each of the spectral peaks that were identified. The determination of the number of components, the location of the peaks, and the spread are all data-adaptive rather than imposed *a priori* using the standard methods. The Bayesian framework facilitates inference on many subject-matter hypotheses. The construction of the SDF under the BMARD method uses a family of autoregressive kernels which, along with a Dirichlet process prior, gives rise to a Bayesian nonparametric discrete mixture model. BMARD decomposes a stationary univariate process as a linear mixture of latent AR(2) processes, where each component is associated to a unique peak on the SDF. The weights of the mixture provide an insight into the components contribution of each latent process to the total variance of the observed signal. Moreover, as demonstrated in the simulations, BMARD gives very good estimates without requiring a higher number of components, which is essential when the sampling rate is low. Thus, even with a relatively few components, the location of the frequencies corresponding to spectral peaks are well estimated because the method data-adaptively identifies the optimal placement of these peak frequencies.

The comparison of BMARD with other approaches points to a limitation when estimating a moving average SDF due to its concavity and smooth shape. Since the AR(2) kernel is a convex function, fitting BMARD to a smooth SDF results in a mixture of several kernels localized by peaks arising from the variability of the periodogram. When the SDF is shaped by sharp peaks as the autoregressive models, BMARD provides parsimonious estimator for the SDF, and outperform other methods. However, despite this limitation, the BMARD still performs very well using the metric of identifying the frequencies that produce the spectral peaks.

The LFP analysis of 5 rats shows how the signal is decomposed into different frequency components during performance of an odor sequence memory task, and how the distribution of peak activity varies across trial types. The significant difference between InSeq and OutSeq trials (in which odors were presented in the correct or incorrect order, respectively) provides compelling evidence that hippocampal LFP activity carries significant information about the sequential organization of human experiences, specifically whether or not events occurred in the expected order. This is an important finding because LFP activity reflects the summed influence of large groups of neurons near the electrode tip. In fact, hippocampal oscillations are generally viewed as playing an important role in synchronizing neural activity across neuronal ensembles and circuits, or promoting distinct information processing states (reviewed in [58]). Beyond the data analysis specific findings, the development of this model may have broader implications in neuroscience as a novel approach to extract additional trial-specific information from LFP recordings, an electrophysiological approach extensively used in the field.

The decomposition representation in the BMARD method provides a different insight for weakly stationary processes as composed of latent processes with various oscillatory behavior. The application of BMARD is broad and could extend well beyond neuroscience. It is applicable to other types of data such as weather evolution composed of natural cycles with different periodicity or financial data that exhibit economical cycles that are not necessarily sinusoidal but can be better represented by simpler stochastic processes explaining short, medium, and long term tendencies.

One of the possible extensions of the BMARD is nonstationary time series analysis. Some works with similar goals than ours in this branch of time series analysis are [59] showing how to find the optimal frequency band for nonstationary time series while [60],[61], [62], [63] and the empirical mode decomposition by [64] gives other decomposition approaches.

Acknowledgments

The authors thank Dr. Hart (see [20]) for generously sharing his computer codes. Financial support is acknowledged from the KAUST Research Fund and the NIH 1R01EB028753-01 to B. Shahbaba and N. Fortin.

Appendix: MCMC Algorithm

A Metropolis-Hastings within Gibbs is next described to sample from the posterior distribution of the parameters and is available on github through https://github.com/Cuauhtemoctzin/BMARD. The algorithm first updates the number of components with a birth-death process, which at each iteration proposes with equal probability to increase the number of components by one or decrease it by one. In the case of a birth step, the M-H ratio is.

$$q(\theta \mid \theta^*)/q(\theta^* \mid \theta) = \left(I\left(\epsilon_j^{*(i)} < \psi_j^{(i)}\right)\left(\epsilon_j^{*(i)} - \epsilon_{j-1}\right)^{-1} + I\left(\epsilon_j^{*(i)} > \psi_j^{(i)}\right)\left(\epsilon_j - \epsilon_j^{*(i)}\right)^{-1} \right)^{-1} \quad (7)$$

Where $\epsilon^{*(i)}$ represent a new generation for the partition in the interval $(\epsilon_{j-1}, \epsilon_j)$ assuming the random selection of the index j to split that subinterval and I is the indicator function. While in the the case of choosing a death step the M-H probability is computed as

$$q(\theta \mid \theta^*)/q(\theta^* \mid \theta) = I\left(\epsilon_j > \psi_j^{*(i)}\right)\left(\epsilon_j - \epsilon_{j-1}\right)^{-1} + I\left(\epsilon_j < \psi_j^{*(i)}\right)\left(\epsilon_{j+1} - \epsilon_j\right)^{-1} \quad (8)$$

where ψ^* represents the uniform draw when two component are joined by deleting the value ϵ_j from the partition. The proposal distribution to update of the location parameters is uniform in the interval $(\psi_c - \epsilon, \psi_c + \epsilon)$ with appropriate conditions to take the modulus over the subinterval defined by the partition. For the scale parameter, is used a similar uniform draw over the interval $(L_c - \epsilon, L_c + \epsilon)$.

The alpha parameter is updated based on the slice sampling to generate a random walk over the subgraph of the marginal posterior distribution of α given by:

$$\pi(\alpha \mid \boldsymbol{C}) \propto \pi(\alpha)\alpha^{C-1}(\alpha + T)\beta(\alpha + 1, T) \quad (9)$$

where $\pi(\alpha)$ is the prior over α in the algorithm is set $\pi(\alpha)$ as log-normal, T the observed process size, and $\beta(.)$ is the beta function. The next algorithm presents the steps described.

Algorithm 1:

M-H within Gibbs DP AR(2) mixture for stationary processes

Propose: $C^{(0)}, M, d$;

Initialize randomly: $\psi_1^{(0)}, \ldots, \psi_C^{(0)}, L_1^{(0)}, \ldots, L_C^{(0)}, V_1^{(0)}, \ldots, V_M^{(0)}, \epsilon_1^{(0)}, \ldots, \epsilon_{C-1}^{(0)}$;

while $i \leq MCMC\ chain\ size$ **do**

  Choose Death or Birth with probability equal to 1/2;

  if *Death* **then**

    Choose j randomly with probability $1/C^{(i)}$;

    Remove ϵ_j;

    Propose a new $\psi_j^{*(i)}$ uniformly on $(\epsilon_{j-1}, \epsilon_{j+1})$;

    Propose a new $L_j^{*(i)}$ uniformly on $(0, d)$;

    Compute the M-H probability ;

    Make a Reject-Acceptance M-H step for the new $C^{*(i)} = C^{(i-1)} - 1, \bar{psi}^*, \bar{L}^*$;

  else

    Birth: Generate a new component ;

    Choose a j randomly with probability $1/C^{(i)}$;

    Propose a new $\epsilon_j^{*(i)}$ uniformly on $(\epsilon_{j-1}, \epsilon_j)$;

    Propose a new $\psi_j^{*(i)}$ uniformly on the interval where is not yet a ψ ;

    Propose a new $L_j^{*(i)}$ uniformly on $(0, d)$;

    Compute the M-H probability ;

    Make a Reject-Acceptance M-H step for the new $C^{*(i)} = C^{(i-1)} + 1, \bar{psi}^*, \bar{L}^*$;

  end

  Sample jointly the vector $\bar{\psi}^{(i)}$;

  Sample jointly the vector $\bar{L}^{(i)}$;

  Sample jointly the vector $\bar{V}^{(i)}$;

  Sample jointly the vector $\bar{Z}^{(i)}$;

  Sample $\alpha^{(i)}$;

end

References

[1]. Eichenbaum H, Time cells in the hippocampus: a new dimension for mapping memories, Nature Reviews Neuroscience 15 (11) (2014) 732–744. [PubMed: 25269553]

[2]. Allen TA, Morris AM, Mattfeld AT, Stark CE, Fortin NJ, A sequence of events model of episodic memory shows parallels in rats and humans, Hippocampus 24 (10) (2014) 1178–1188. arXiv: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hipo.22301, doi:10.1002/hipo.22301. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hipo.22301 [PubMed: 24802767]

[3]. Shumway RH, Stoffer DS, Time series analysis and its applications: with R examples, Springer, 2017.

[4]. Kass RE, Eden UT, Brown EN, service SO), Analysis of Neural Data, Springer New York, New York, NY, 2014.

[5]. Ombao H, Lindquist M, Thompson SD, Wesley (Of University of California, J. O. U. o. C. Aston, Handbook of neuroimaging data analysis, CRC Press, Taylor & Francis Group, Boca Raton, 2017.

[6]. Prado R, West M, Time series: modeling, computation, and inference, CRC Press, 2010.

[7]. Buzsáki G, Rhythms of The Brain, 2009, pp. xiv, 448 p. doi:10.1093/acprof:oso/9780195301069.001.0001.

[8]. Klimesch W, Doppelmayr M, Russegger H, Pachinger T, Schwaiger J, Induced alpha band power changes in the human EEG and attention, Neuroscience Letters 244 (2) (1998) 73–76. doi:10.1016/S0304-3940(98)00122-0. URL https://linkinghub.elsevier.com/retrieve/pii/S0304394098001220 [PubMed: 9572588]

[9]. Doppelmayr M, Klimesch W, Pachinger T, Ripper B, Individual differences in brain dynamics: important implications for the calculation of event-related band power, Biological cybernetics 79 (1) (1998) 49–57. [PubMed: 9742677]

[10]. Allen TA, Salz DM, McKenzie S, Fortin NJ, Nonspatial sequence coding in ca1 neurons, Journal of Neuroscience 36 (5) (2016) 1547–1563. arXiv:https://www.jneurosci.org/content/36/5/1547.full.pdf, doi:10.1523/JNEUROSCI.2874-15.2016. URL https://www.jneurosci.org/content/36/5/1547 [PubMed: 26843637]

[11]. Gao X, Shen W, Shahbaba B, Fortin N, Ombao H, Evolutionary state-space model and its application to time-frequency analysis of local field potentials, arXiv preprint arXiv:1610.07271.

[12]. Lee TC, A simple span selector for periodogram smoothing, Biometrika 84 (4) (1997) 965–969.

[13]. Ombao HC, Raz JA, Strawderman RL, von Sachs R, A simple generalised crossvalidation method of span selection for periodogram smoothing, Biometrika 88 (4) (2001) 1186–1192. URL http://www.jstor.org/stable/2673711

[14]. Wahba G, Automatic smoothing of the log periodogram, Journal of the American Statistical Association 75 (369) (1980) 122–132.

[15]. Krafty R, Collinge W, Penalized multivariate whittle likelihood for power spectrum estimation, Biometrika 100 (4) (2013) 447–458.

[16]. Krafty R, Hall M, Guo W, Penalized multivariate whittle likelihood for power spectrum estimation, Biometrika 98 (3) (2011) 583–598. [PubMed: 26855437]

[17]. Cadonna A, Kottas A, Prado R, Bayesian mixture modeling for spectral density estimation, Statistics & Probability Letters 125 (2017) 189 – 195. doi:10.1016/j.spl.2017.02.008. URL http://www.sciencedirect.com/science/article/pii/S0167715217300573

[18]. Petrone S, Bayesian density estimation using bernstein polynomials, Canadian Journal of Statistics 27 (1) (1999) 105–126.

[19]. Choudhuri N, Ghosal S, Roy A, Bayesian estimation of the spectral density of a time series, Journal of the American Statistical Association 99 (468) (2004) 1050–1059. arXiv:10.1198/016214504000000557, doi:10.1198/016214504000000557. URL 10.1198/016214504000000557

[20]. Hart B, Guindani M, Malone S, Fiecas M, A nonparametric Bayesian model for estimating spectral densities of resting-state EEG twin data, Biometrics (2020) biom.13393doi:10.1111/biom.13393. URL https://onlinelibrary.wiley.com/doi/10.1111/biom.13393

[21]. Macaro C, Prado R, Spectral decompositions of multiple time series: a bayesian non-parametric approach, Psychometrika 79 (1) (2014) 105–129. [PubMed: 24154824]

[22]. Edwards MC, Meyer R, Christensen N, Bayesian nonparametric spectral density estimation using b-spline priors, Statistics and Computing 29 (1) (2019) 67–78.

[23]. Kirch C, Edwards MC, Meier A, Meyer R, Beyond whittle: Nonparametric correction of a parametric likelihood with a focus on bayesian time series analysis, Bayesian Anal. 14 (4) (2019) 1037–1073. doi: 10.1214/18-BA1126. URL 10.1214/18-BA1126

[24]. Degras D, Ting C-M, Ombao H, Markov-Switching State-Space Models with Applications to Neuroimaging, arXiv:2106.05092 [stat]ArXiv: 2106.05092. URL http://arxiv.org/abs/2106.05092

[25]. Ozaki T, Time series modeling of neuroscience data, CRC press, 2012.

[26]. Brockwell PJ, Davis RA, Time Series: Theory and Methods, Springer-Verlag, Berlin, Heidelberg, 1986.

[27]. Whittle P, Curve and periodogram smoothing, Journal of the Royal Statistical Society: Series B (Methodological) 19 (1) (1957) 38–47.

[28]. Ferguson TS, A bayesian analysis of some nonparametric problems, The annals of statistics (1973) 209–230.

[29]. Antoniak CE, Mixture of Dirichlet process with applications to Bayesian nonparametric problems, Annals of Statistics 273(5281) (1974) 1152–1174.

[30]. Sethuraman J, A constructive definition of dirichlet priors, Statistica sinica (1994) 639–650.

[31]. Müller P, Quintana FA, Jara A, Hanson T, Bayesian nonparametric data analysis, Springer, 2015.

[32]. Shahbaba B, Neal R, Nonlinear models using dirichlet process mixtures, Journal of Machine Learning Research 10 (Aug) (2009) 1829–1850.

[33]. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB, Bayesian data analysis, CRC press, 2013.

[34]. Congdon P, Bayesian statistical modelling, Vol. 704, John Wiley & Sons, 2007.

[35]. Neal RM, Markov chain sampling methods for dirichlet process mixture models, Journal of computational and graphical statistics 9 (2) (2000) 249–265.

[36]. Ishwaran H, Zarepour M, Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models, Biometrika 87 (2) (2000) 371–390.

[37]. Jasra A, Holmes CC, Stephens DA, Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling, Statistical Science (2005) 50–67.

[38]. Ishwaran H, James LF, Approximate dirichlet process computing in finite normal mixtures: smoothing and prior information, Journal of Computational and Graphical statistics 11 (3) (2002) 508–532.

[39]. Robert C, Casella G, Monte Carlo statistical methods, Springer Science & Business Media, 2013.

[40]. Escobar MD, West M, Bayesian density estimation and inference using mixtures, Journal of the american statistical association 90 (430) (1995) 577–588.

[41]. Dickinson A, DiStefano C, Senturk D, Jeste SS, Peak alpha frequency is a neural marker of cognitive function across the autism spectrum, European Journal of Neuroscience 47 (6) (2018) 643–651. [PubMed: 28700096]

[42]. Fan J, Gijbels I, Local polynomial modelling and its applications: monographs on statistics and applied probability 66, Vol. 66, CRC Press, 1996.

[43]. Pawitan Y, O'sullivan F, Nonparametric spectral density estimation using penalized whittle likelihood, Journal of the American Statistical Association 89 (426) (1994) 600–610.

[44]. Eddelbuettel D, François R, Rcpp: Seamless R and C++ integration, Journal of Statistical Software 40 (8) (2011) 1–18. doi:10.18637/jss.v040.i08. URL http://www.jstatsoft.org/v40/i08/

[45]. Eddelbuettel D, Seamless R and C++ Integration with Rcpp, Springer, New York, 2013, iSBN 978–1-4614–6867-7. doi:10.1007/978-1-4614-6868-4.

[46]. Eddelbuettel D, Balamuta JJ, Extending extitR with extitC++: A Brief Introduction to extitRcpp, PeerJ Preprints 5 (2017) e3188v1. doi:10.7287/peerj.preprints.3188v1. URL 10.7287/peerj.preprints.3188v1

[47]. Scrucca L, Fop M, Murphy TB, Raftery AE, mclust 5: clustering, classification and density estimation using Gaussian finite mixture models, The R Journal 8 (1) (2016) 289–317. URL 10.32614/RJ-2016-021 [PubMed: 27818791]

[48]. Biernacki C, Celeux G, Govaert G, Assessing a mixture model for clustering with the integrated completed likelihood, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (7) (2000) 719–725, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. doi:10.1109/34.865189.

[49]. Fasy BT, Kim J, Lecci F, Maria C, Millman DL, Rouvreau V., TDA: Statistical Tools for Topological Data Analysis, r package version 1.7.7 (2021). URL https://CRAN.R-project.org/package=TDA

[50]. Ravishanker N, Chen R, Topological Data Analysis (TDA) for Time Series, arXiv:1909.10604 [stat]ArXiv: 1909.10604. URL http://arxiv.org/abs/1909.10604

[51]. Ombao H, Van Bellegem S, et al. , Coherence analysis of nonstationary time series: a linear filtering point of view, IEEE Transactions on Signal Processing 56 (2006) 2259–2266.

[52]. Park T, Eckley IA, Ombao HC, Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes, IEEE Transactions on Signal Processing 62 (20) (2014) 5240–5250.

[53]. Wang Y, Ting C-M, Ombao H, Modeling effective connectivity in high-dimensional cortical source signals, IEEE Journal of Selected Topics in Signal Processing 10 (7) (2016) 1315–1325.

[54]. Baccala LA, Sameshima K, Partial directed coherence: a new concept in neural structure determination, Biological Cybernetics 84 (6) (2001) 463–474. doi:10.1007/PL00007990. URL http://link.springer.com/10.1007/PL00007990 [PubMed: 11417058]

[55]. Fontaine C, Frostig RD, Ombao H, Modeling non-linear spectral domain dependence using copulas with applications to rat local field potentials, Econometrics and Statistics 15 (2020) 85–103. doi:10.1016/j.ecosta.2019.06.003. URL https://linkinghub.elsevier.com/retrieve/pii/S2452306219300450

[56]. Hu L, Guindani M, Fortin NJ, Ombao H, A hierarchical bayesian model for differential connectivity in multi-trial brain signals, Econometrics and Statistics 15 (2020) 117–135. doi:10.1016/j.ecosta.2020.03.009. URL https://linkinghub.elsevier.com/retrieve/pii/S2452306220300423 [PubMed: 33163735]

[57]. Ngo D, Ombao H, Sun Y, Genton MG, Wu J, Srinivasan R, Cramer S, An exploratory data analysis of electroencephalograms using the functional boxplots approach, Frontiers in neuroscience 9 (2015) 282. [PubMed: 26347598]

[58]. Colgin LL, Rhythms of the hippocampal network, Nature Reviews Neuroscience 17 (4) (2016) 239–249. [PubMed: 26961163]

[59]. Bruce SA, Tang CY, Hall MH, Krafty RT, Empirical frequency band analysis of nonstationary time series, Journal of the American Statistical Association 115 (532) (2020) 1933–1945. [PubMed: 34108777]

[60]. Ombao H, Ringo Ho M.-h., Time-dependent frequency domain principal components analysis of multichannel non-stationary signals, Computational Statistics & Data Analysis 50 (9) (2006) 2339–2360. doi:10.1016/j.csda.2004.12.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0167947304004037

[61]. Rosen O, Wood S, Stoffer DS, AdaptSPEC: Adaptive Spectral Estimation for Nonstationary Time Series, Journal of the American Statistical Association 107 (500) (2012) 1575–1589. doi:10.1080/01621459.2012.716340. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2012.716340

[62]. Sato JR, Morettin PA, Arantes PR, Amaro E, Wavelet based time-varying vector autoregressive modelling, Computational Statistics & Data Analysis 51 (12) (2007) 5847–5866. doi:10.1016/j.csda.2006.10.027. URL https://linkinghub.elsevier.com/retrieve/pii/S0167947306004208

[63]. Song AH, Ba D, Brown EN, PLSO: A generative framework for decomposing nonstationary timeseries into piecewise stationary oscillatory components, arXiv:2010.11449 [stat]ArXiv: 2010.11449. URL http://arxiv.org/abs/2010.11449

[64]. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time

series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454 (1971) (1998) 903–995. doi:10.1098/rspa.1998.0193. URL https://royalsocietypublishing.org/doi/10.1098/rspa.1998.0193

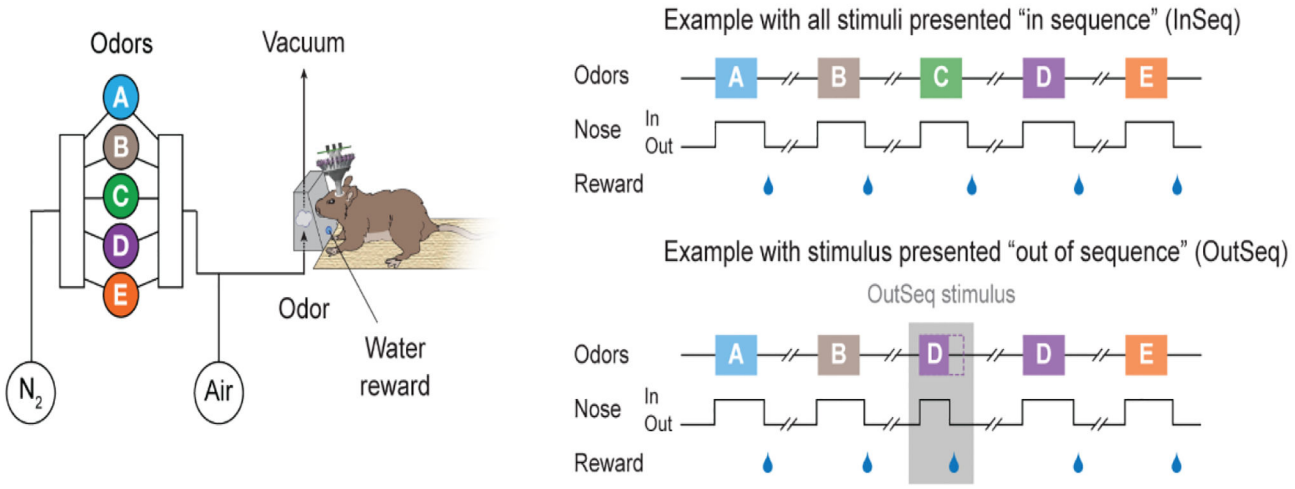

Figure 1:

Five rats received multiple sequences of five odors (left; odors A, B, C, D, E). The animals were required to correctly identify whether the odor was presented "in sequence" (top right; by holding its nose in the port for ~1.2 s, when an auditory signal is delivered) or "out of sequence" (bottom right; by withdrawing its nose before the signal) to receive a water reward.

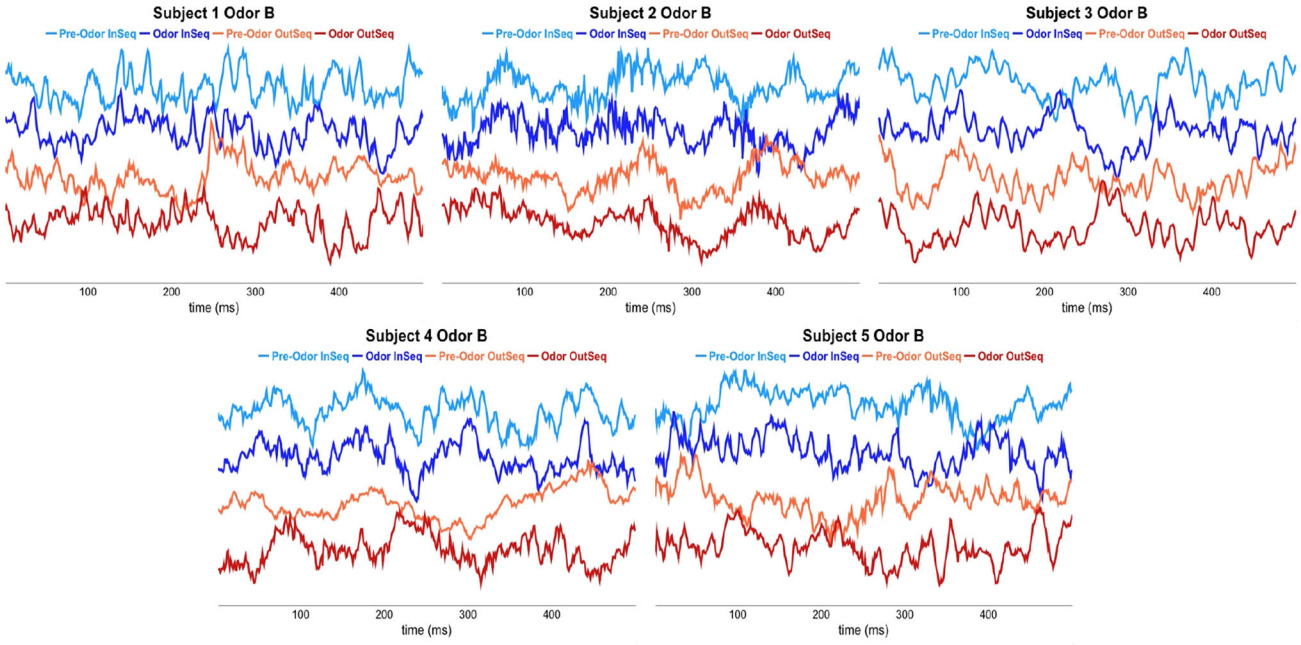

Figure 2:

Local field potential (LFP) signals recorded from five rats during an odor sequence memory experiment. Two stimulus conditions are considered: before the odor is presented (”Pre-Odor”) vs. during odor delivery (”Odor”). Two types of trials with different temporal contexts are displayed: trials in which the odor is presented in the correct order (”InSeq”) vs. trials in which the odor is presented in the incorrect order (OutSeq).

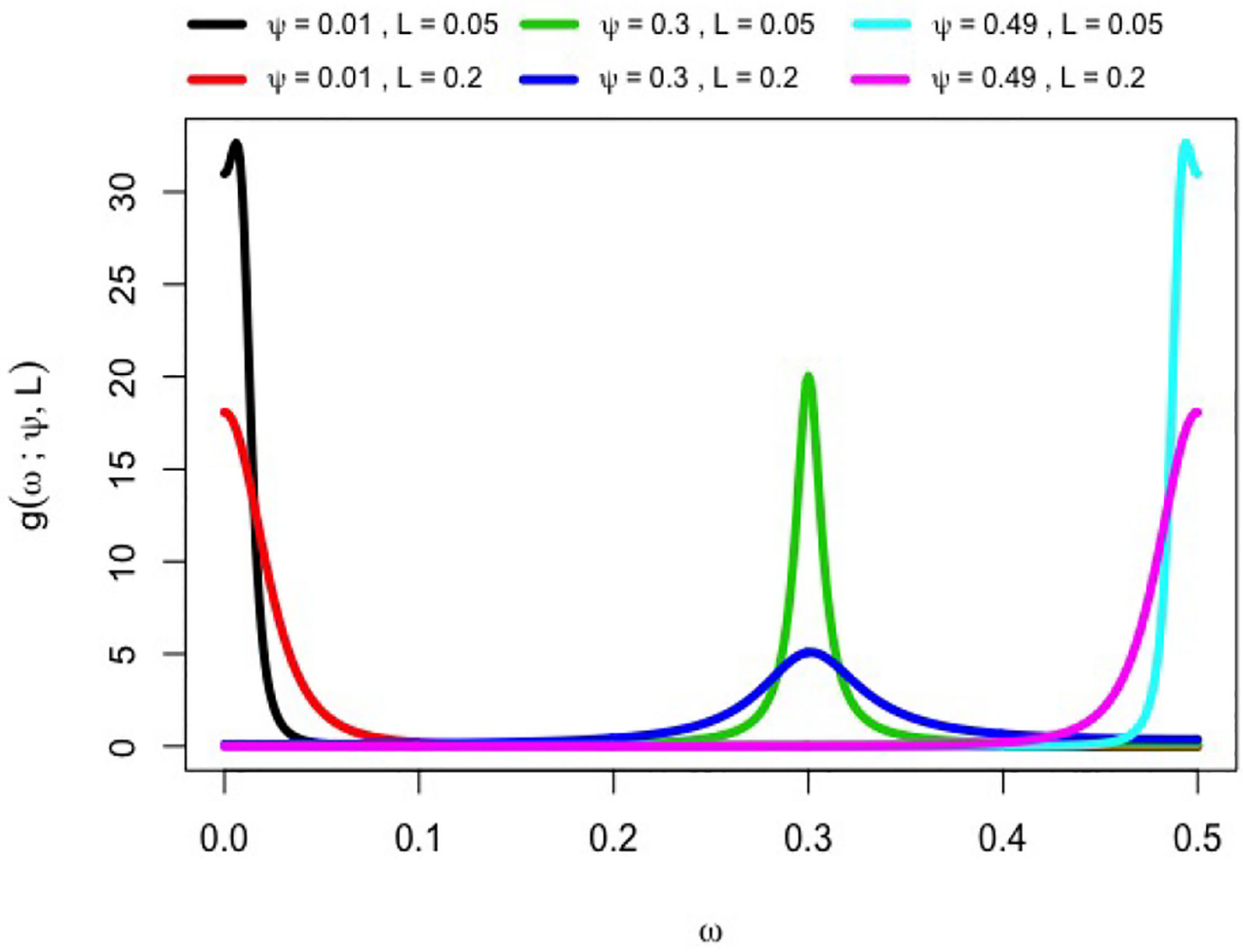

Figure 3:
Some AR(2) kernels $g(\omega; \psi, L)$ with different values of the phase parameter $\psi = 0.05, 0.30, 0.49$ as a location (frequency peak) parameter and $L = 0.05, 0.20$ as a scale (bandwidth) parameter.

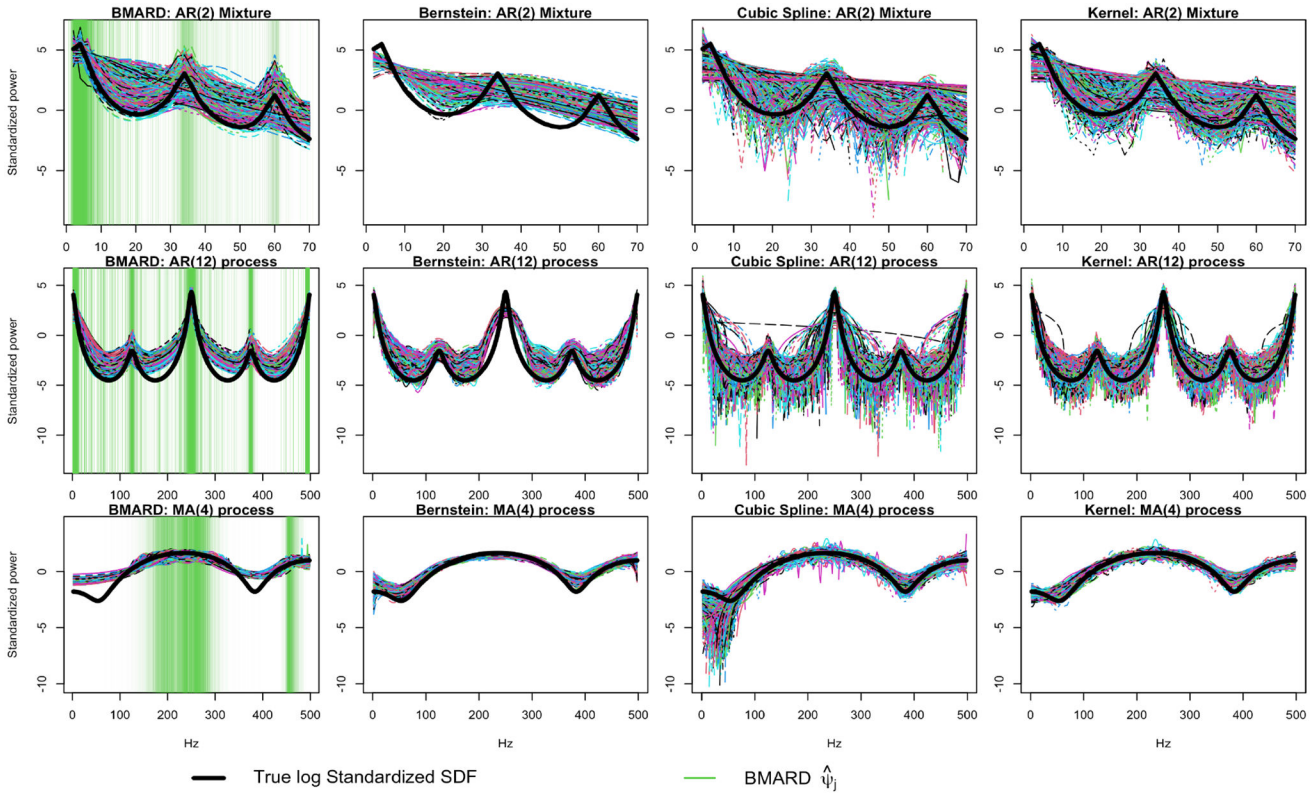

Figure 4:

Estimated standardized SDF curves in log scale for all 1000 simulations for the BMARD method. The thick black curve represents the logarithm of the true standardized SDF of each simulation setting, visualized in green are the estimation of the location parameters $\hat{\psi}_j$ obtained from post-processing MCMC samples associated to each estimated curve.

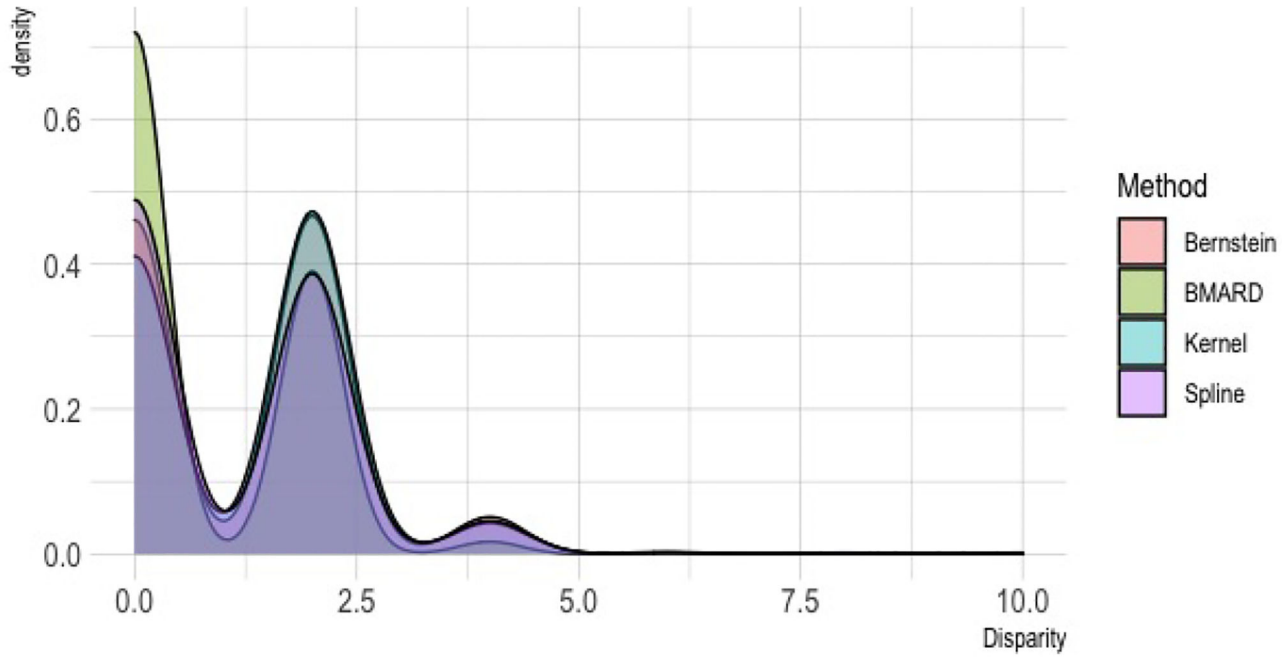

Figure 5:

Absolute disparity between the AR(12) SDF for the central peak of misspecification scenario compared to the argument at which the local maximum is reached on the range of 240–250 Hz for all methods.

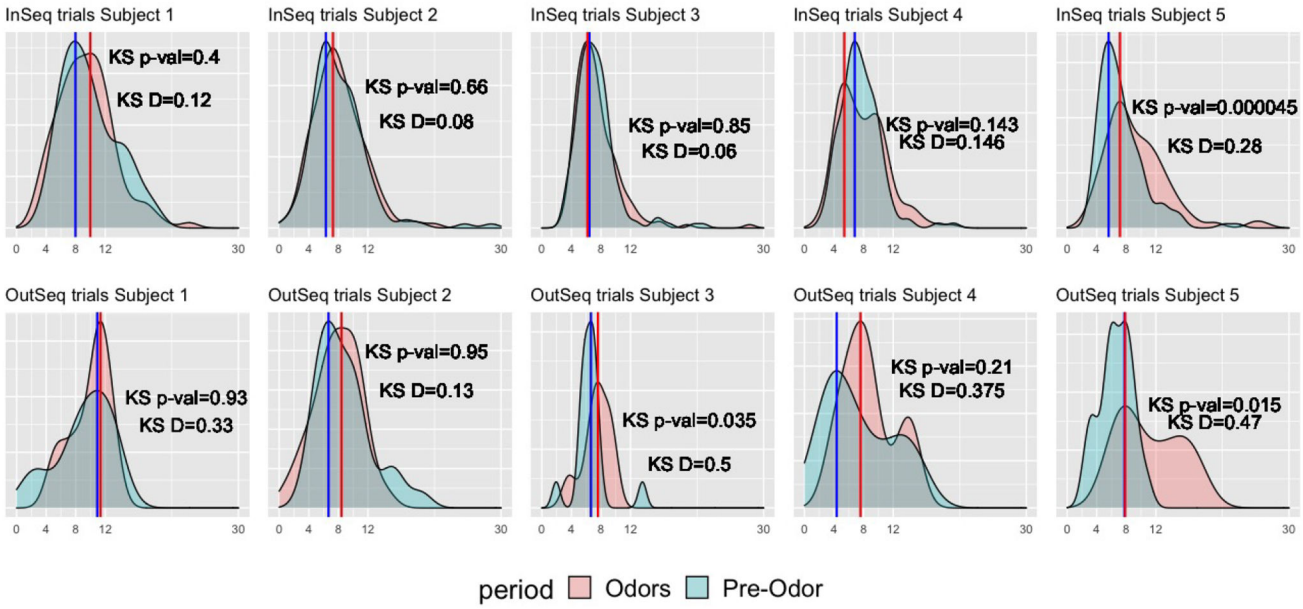

Figure 6:

Peak activity distribution for one hour session for all subjects, the vertical lines correspond to the mode of the localized peaks with the highest mixture weight by OutSeq and InSeq contexts accordingly. The right of each graph contains the p-value and test statistic D of the Kolmogorov-Smirnov test.

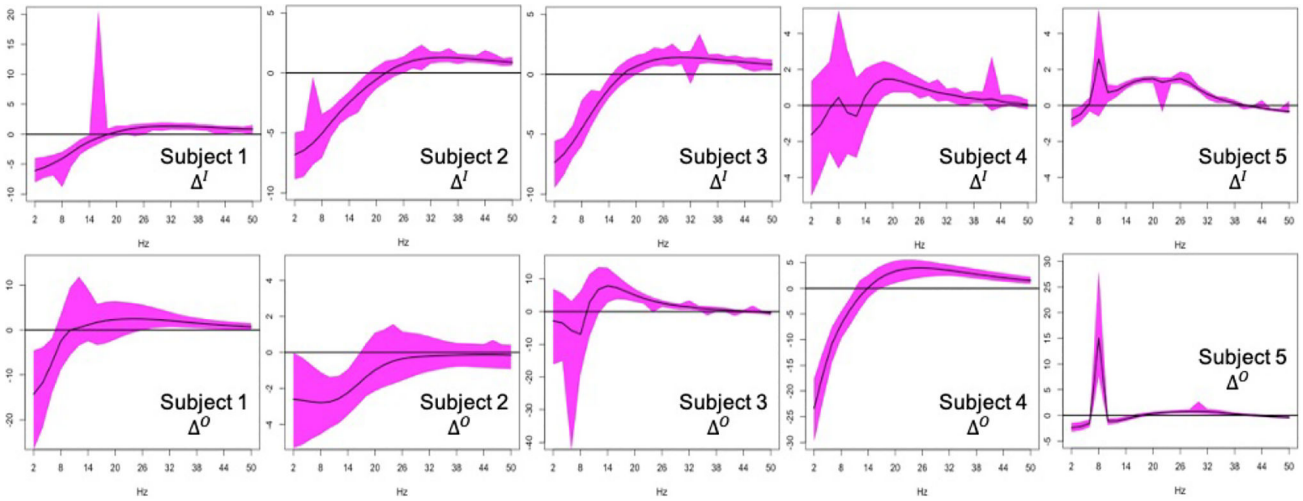

Figure 7:

(top row) Functional Boxplot with 95% of internal region computed from 100000 iterations of a resampling scheme on the difference $\Delta^I(\omega) = f_A^I(\omega) - f_B^I(\omega)$ for each subject. (bottom row) Functional Boxplot with 95% of internal region computed from 100000 iterations of a resampling scheme on the difference $\Delta^O(\omega) = f_A^O(\omega) - f_B^O(\omega)$ for each subject.

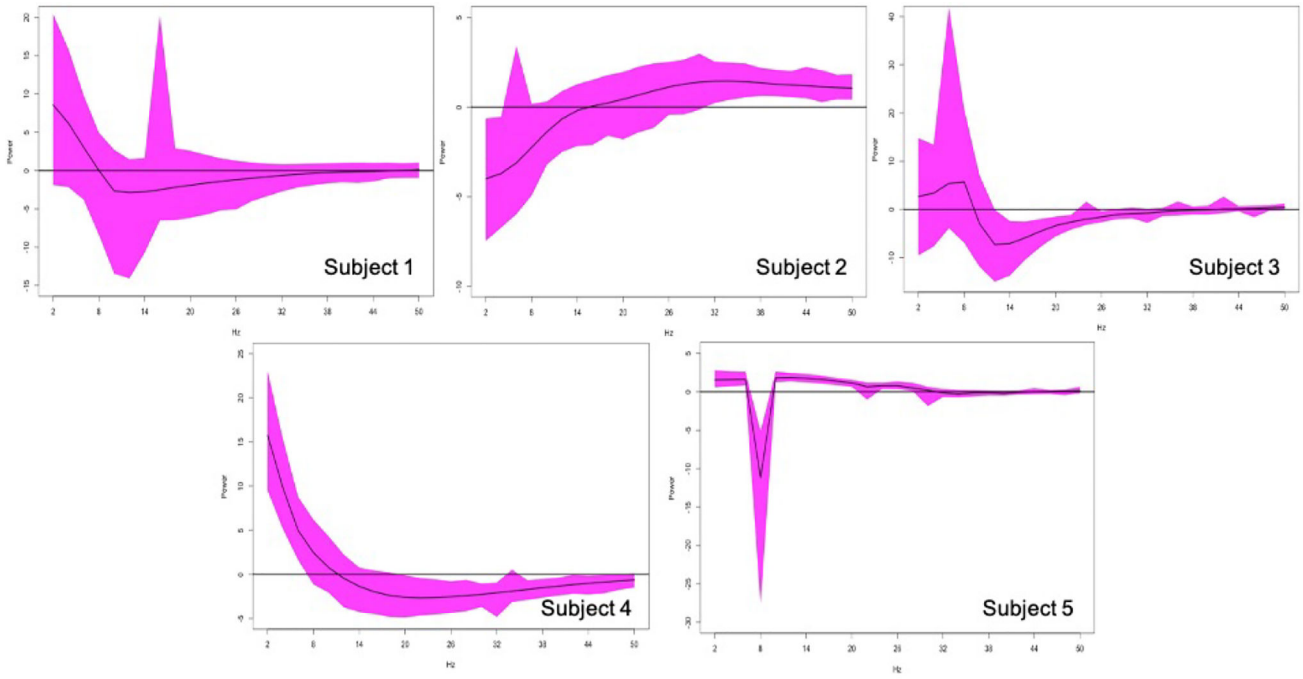

Figure 8:

Functional Boxplots with 95% of internal region for the curves $\Delta^{I-O}(\omega) = \Delta^{I}(\omega) - \Delta^{O}(\omega) = f_A^I(\omega) - f_B^I(\omega) - \left(f_A^O(\omega) - f_B^O(\omega)\right)$ showing the median curve. The curves used in the boxplot were computed using a resample scheme of 10^5 samples from the BMARD posterior SDF curves.

Table 1:

Number of peaks found by applying the rips filtration (TDA) to each median curve estimated for the AR(2) mixture. For the BMARD method is also reported the number of components C^* using a clustering algorithm to post-process the 6 MCMC chains

# of peaks	1	2	3	4	5	>5
BMARD TDA	19.7%	51.8%	25.6%	2.5%	0.4%	0%
BMARD $\widehat{p}_j > 1\%$	2%	27.6%	43.5%	24.1%	2.7%	0.1%
Bernstein TDA	99%	1%	0%	0%	0%	0%
Kernel TDA	66.7%	6.1%	15.1%	5.4%	2.9%	3.8%
Spline TDA	62.6%	9.6%	16.1%	3.9%	3.4%	4.4%

Table 2:

Number of peaks found by applying the rips filtration (TDA) to each median curve estimated for the AR(12) process. For the BMARD method is also reported the number of components C^* using a clustering algorithm to post-process the 6 MCMC chains

# of peaks	3	4	5	6	7	8	9	10	>10
BMARD TDA	0%	0.2%	59.3%	30.9%	8.3%	1.2%	0%	0.1%	0%
BMARD $\widehat{p}_j > 1\%$	17.1%	38.5%	27.2%	14.7%	2.5%	0%	0%	0%	0%
Bernstein TDA	0%	0%	0.8%	6.4%	19.2%	30.6%	29.5%	11.5%	1.8%
Kernel TDA	0%	0%	0.4%	0.4%	0.4%	1.2%	1.7%	2.3%	93.6%
Spline TDA	1%	0.1%	1.7%	0.1%	0.3%	0.2%	9.7%	4.4%	82.5%

Table 3:

Number of peaks found by applying the rips filtration (TDA) to each median curve estimated for the MA(4) process. For the BMARD method is also reported the number of components C^* using a clustering algorithm to post-process the 6 MCMC chains

# of peaks	2	3	4	5	6	7	>7
BMARD TDA	7.2%	44.1%	31.6%	13.6%	2.9%	0.6%	0%
BMARD $\widehat{p}_j > 1\%$	0%	4.9%	45.4%	46.9%	2.7%	0.1%	0%
Bernstein TDA	0.4%	88.8%	9%	1.8%	0%	0%	0%
Kernel TDA	2.9%	39.4%	26.3%	12.4%	7%	4.6%	7.4%
Spline TDA	39.6%	48.6%	5.2%	2.2%	1.3%	0.1%	3%

Table 4:

Mean and standard deviation of the absolute difference with respect to the AR(2) mixture simulation parameters. Only for generated processes with at least one chain that correctly identified the true number of components. All components were generated with $L = .03$

	$p_1 = .1$	$p_2 = .6$	$p_3 = .3$
Mean Disparity	$\psi_1 = 4$ **Hz**	$\psi_2 = 34$ **Hz**	$\psi_3 = 60$ **Hz**
$\lvert\psi_c - \hat{\psi}_c\rvert$ Hz	4.34(1.43)	7.77(5.64)	8.76(8.76)
$\lvert L_c - \hat{L}_c\rvert$	0.02(0.04)	0.02(0.04)	0.02(0.03)
$\lvert p_c - \hat{p}_c\rvert$	0.68(0.13)	0.43(0.09)	0.21(0.06)

Table 5:

Mean and standard deviation of the absolute difference with respect to the AR(12) peaks in the SDF. Only for generated processes with at least one chain that correctly identified the true number of components (5).

Mean Disparity	ψ_1 **= 0Hz**	ψ_2 **= 150Hz**	ψ_3 **= 250Hz**	ψ_4 **= 350Hz**	ψ_5 **= 500Hz**
$\lvert\psi_c - \hat{\psi}_c\rvert$ Hz	4.77(2.19)	32.67(21.36)	3.94(14.38)	31.09(19.58)	3.68(2.07)

Table 6:

Total number of trials/epochs for each **subject** (S1, S2, S3, S4, S5) for each **odor type** (B-C-D-E) and **stimulus type** (OutSeq vs Inseq).

Subject	B InSeq	B OutSeq	C InSeq	C OutSeq	D InSeq	D OutSeq	E InSeq	E OutSeq
S1	34	1	25	0	26	3	21	2
S2	38	8	26	8	42	6	29	7
S3	57	3	47	5	37	4	23	4
S4	40	2	30	4	29	8	24	2
S5	41	3	37	5	31	8	26	5