*Submitted to the Annals of Applied Statistics*

# MULTISCALE SPECTRAL MODELLING FOR NONSTATIONARY TIME SERIES WITHIN AN ORDERED MULTIPLE-TRIAL EXPERIMENT

BY JONATHAN EMBLETON[1,*], MARINA I. KNIGHT[1,†] AND HERNANDO OMBAO[2]

[1]*Department of Mathematics, University of York, UK,* [*]*je687@york.ac.uk;* [†]*marina.knight@york.ac.uk*

[2]*King Abdullah University of Science and Technology (KAUST), Saudi Arabia, hernando.ombao@kaust.edu.sa*

Within the neurosciences, it is natural to observe variability across time in the dynamics of an underlying brain process. Wavelets are essential in analysing brain signals because, even within a single trial, brain signals exhibit nonstationary behaviour. However, neurological signals generated within an experiment may also potentially exhibit evolution across trials (replicates) even for identical stimuli. As neurologists consider localised spectra of brain signals to be most informative, we propose the MULtiple-Trials Locally Stationary Wavelet process (MULT-LSW) that fills the gap in the literature by directly giving a stochastic wavelet representation of the time series of ordered replicates itself. MULT-LSW yields a natural desired time- and trial-localisation of the process dynamics, capturing nonstationary behaviour both within and across trials. While current techniques are restricted by the assumption of uncorrelated replicates, here we account for between-trial correlation. We rigorously develop the associated wavelet spectral estimation framework along with its asymptotic properties. By means of thorough simulation studies, we demonstrate the theoretical estimator properties hold in practice. A real data investigation into the evolutionary dynamics of the hippocampus and nucleus accumbens during an associative learning experiment demonstrates the applicability of our proposed methodology, as well as the new insights it provides. Our model is general and facilitates wider experimental data analysis than the current literature allows.

**1. Introduction** In an experimental setting consisting of repeated trials, inference is typically carried out on the average dynamics of the underlying process over all trials. However, a recent study on neurological signals (Fiecas and Ombao, 2016) suggests that this approach is naive due to its failure to account for the possibility of a change in the process dynamics, even for identical stimuli, over the course of the experiment. Their data example focusses on the hippocampus (Hc) and the nucleus accumbens (NAc), both known to play important roles in cognitive processing as they are individually associated with memory recall and the processing of reward, respectively. Recordings of electrical activity (at approximately 1000Hz) using local field potentials (LFPs) were obtained from the Hc and NAc of an awake behaving macaque during an associative learning experiment. For each trial, the macaque was presented with one of four pictures and was then tasked with associating this picture with one of the four doors appearing on the screen. Upon making a correct association, the macaque was rewarded with a small quantity of juice. Plots of the LFPs obtained from trials in which a correct association was made are shown in Figures 1 and 2, where the experimental timeline was ensured to match from trial to trial. Our focus here is to dynamically explore and characterise the brain activity both within a trial and across ordered trials, as manifest at the level of Hc and NAc and quantified by the spectral representations of these recorded complex time series signals. Variability in neuronal activity within both brain regions has also been observed

[*]Supported by EPSRC EP/N509802/1.

[†]Corresponding author: marina.knight@york.ac.uk

*Keywords and phrases:* neuroscience, wavelet-based spectra, cross-trial dependence.

over the trials of a learning experiment in other recent studies, including Seger and Cincotta (2006); Gorrostieta *et al.* (2012); Abela *et al.* (2015). Such traits present the challenge of modelling time series that display potential nonstationary behaviour not only across time, but also across the timeline of the *temporally sequentially ordered* trials. Additionally, analyses undertaken so far on such data have been carried out under the simplifying assumption of trial uncorrelation, an assumption that we absorb as a particular case of the proposed model. Our aim is to identify the existence of activation windows through the experimental trials and to directly connect these to the within-trial tasks, thus achieving an insight into the cognitive learning process manifest during the experiment. In order to achieve a comparative analysis into the working memory and reward mechanisms, the MULT-LSW framework departs from previous approaches such as that of Fiecas and Ombao (2016) and separately consider the sets of ordered correct and incorrect trials.

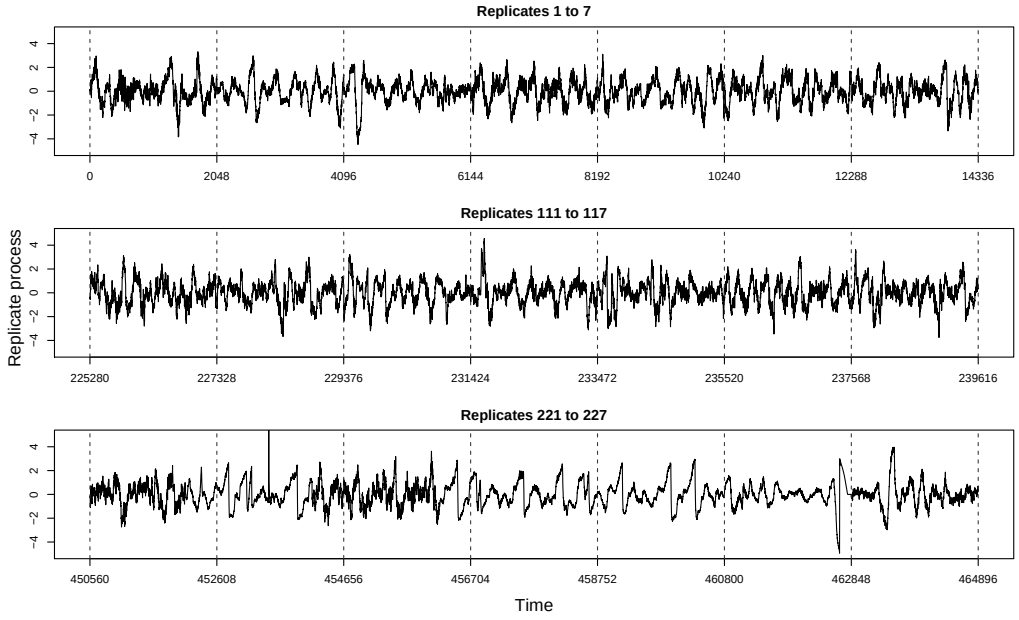


Fig 1: Concatenated series of the hippocampus (Hc) data in the correct response trials (replicates). (Vertical lines denote breaks between trials. Concatenation is only used for meta-process visualisation.)

In the specific context of brain signals, the usefulness of time-scale decompositions that are typical of wavelet constructions has already been established in the literature (Sanderson *et al.*, 2010; Park *et al.*, 2014). For the purpose of quantifying meta-process variation across times and trials, the ability of wavelet decompositions to extract localised information across dyadic scales conveying 'zoomed-in' or '-out' information is crucial. Via the process spectrum, this in turn will allow for an explicit split of the process power across scales, times and trials. Our aim is therefore to develop a *wavelet*-based model that directly gives a *stochastic representation to the meta-process of ordered trials* and simultaneously captures, in a scale-dependent manner, the *evolutionary dynamics of the underlying brain process across time within each trial and across the trials of the experiment*. The *multiscale* insight afforded by the wavelet decomposition offers an additional layer of information and contributes to the understanding of the *within-trial* evolution. As already pointed out, an important assumption typically undertaken is that of independent/ uncorrelated trials. However, some studies document evidence of correlation across trials in an experiment (Arieli *et al.*, 1996; Huk *et al.*, 2018).

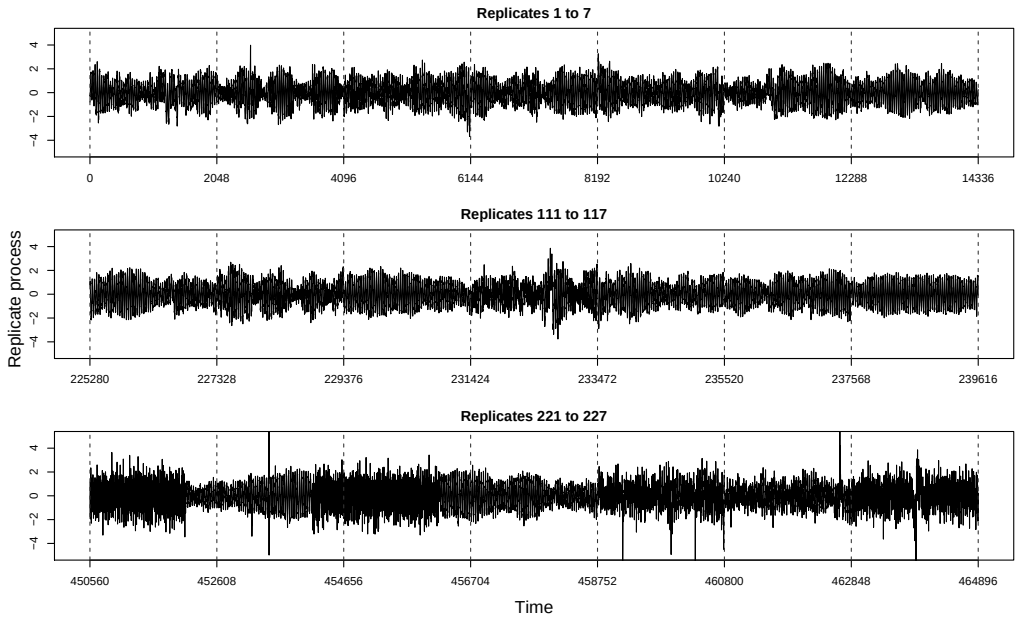


Fig 2: Concatenated series of the nucleus accumbens (NAc) data in the correct response trials (replicates). (Note each trial is a time series and concatenation is only used to aid meta-process visualisation.)

Thus, one important advantage of our model is that it *accounts for the dependence between trials* by means of a coherence quantity that acts as a measure for cross-trial dependence. Nevertheless, should trial uncorrelation be a reasonable assumption, our model is able to flexibly incorporate it.

As time series data encountered in practice are often of a *nonstationary* nature and much recent research has been concerned with developing statistical models that capture this behaviour within the series, we limit our (brief) review to developments that can handle nonstationary data. Priestley (1965) was first to study in detail nonstationary processes with time varying spectral characteristics. Further models using Fourier waveforms have been developed by Dahlhaus (1997) who introduced the methodology of locally stationary processes which allowed for asymptotics to be established, and Ombao *et al.* (2002) whom utilise the SLEX (smoothed localised complex exponentials) library of waveforms which give localisation in both time and frequency. Similar representations have been developed using wavelets. Approaches that use wavelet thresholding for smoothing the spectra of locally stationary time series have been considered by von Sachs and Schneider (1996) and Neumann and von Sachs (1997). While wavelets have long been used for nonparametric function estimation, they also serve as building blocks for stochastic representation of nonstationary processes, starting with the locally stationary wavelet framework developed by Nason *et al.* (2000). Their approach used a set of discrete non-decimated wavelets to replace the Fourier exponentials used in Dahlhaus (1997), and this in turn offered localisation in both time and scale. Extensions to the multivariate setting appear in the work of Ombao *et al.* (2005), Sanderson *et al.* (2010) and Park *et al.* (2014), where a coherence structure between channels is embedded in the model.

What is not well accounted for in the literature, is the statistical modelling of second order nonstationarity across a collection of (constant mean) time series arising from the same stimulus presented in a designed experiment. For stationary time series, Diggle and Al Wasel (1997) accounted for the stochastic variation that arises across replicated time series via a subject-specific realisation of a random process. They proposed a generalised linear mixed-effects model for periodograms to estimate the population spectrum where the subject-specific

random term in the model captures the cross-subject variation. Saavedra *et al.* (2000) extend this work through the proposal of a kernel-based estimator and Krafty *et al.* (2011) introduce a mixed effects Cramér representation, with both fixed and random effect components contained within the replicate-specific transfer functions, both still under the confines of stationarity. Extension to the nonstationary case has been addressed by Qin *et al.* (2009) whom, through the locally stationary processes framework, developed a time-frequency functional model, where the time-varying log-spectra determine the evolution of the stochastic variation. More recently, Fiecas and Ombao (2016) proposed a model that captures the evolution in the Fourier spectral characteristics (in particular, autospectra and cross-channel coherence) of dynamic brain processes across a collection of trials in an associative learning experiment. The authors refer to trials as replicates, a term we will also interchangeably borrow in our nomenclature, while emphasizing that individual replicates (trials) are not to be understood as identical realisations from the same process, i.e. the realisations from the same stimulus can still vary from one trial to another.

Modelling replicate time series data is often framed as a functional regression problem (see e.g. Morris (2015) for a review). In contrast to our approach that proposes to model the meta-process itself, functional regression approaches often deal with the replicate data by projecting it into the Fourier or wavelet domain, where the spectral representations become subject to modelling (Martinez *et al.*, 2013) and thus potentially embedding within- and between-trial variability. A similar ethos is adopted by Gott *et al.* (2015), where the authors formulate a random effects model for the wavelet spectrum. A possible avenue for future research might be to indeed fuse the two lines of modelling in order to augment estimation. Additionally, Chau and von Sachs (2016) stress the importance of allowing for correlation among replicates and, in the spirit of Morris (2015) highlight the danger of arriving to misleading inferences stemming from inefficiencies of estimators obtained under a blanket assumption of replicate independence.

Our approach acknowledges the work of Fiecas and Ombao (2016), who developed their time–replicate model using Fourier waveforms, but the limitation of their work is the assumption of uncorrelated replicates. This may be unrealistic since these signals are being recorded from the same individual (human, animal or plant, as dictated by the experimental field). Moreover, ignoring autocorrelation might produce misleading results from statistical inference, e.g. through the use of incorrect confidence bands. Our work aims to fill this gap and to the best of our knowledge, our proposed methodology is the first to directly give a stochastic wavelet representation of the multiple-trial process itself while simultaneously allowing for nonstationary behaviour and for cross-trial dependence. Nevertheless, despite the attractive properties of wavelets, this work seeks to promote wavelet process modelling approaches for time series that display nonstationary characteristics as complementary to Fourier-based tools.

Major advantages of our model are that (i) it offers the superior time-localisation typical of wavelet constructions, and (ii) it takes into account the correlation of brain signals across trials. Subsequent benefits amount to finer spectral characterisations of within- and between-trial evolution. We propose to model the meta-process within a locally stationary wavelet process paradigm that builds upon the framework introduced by Nason *et al.* (2000) for a single process (here, replicate or trial). Our proposed general model allows for between-trial correlation, and the constraint of uncorrelated replicates is tackled as a particular case. This amounts to developing novel evolutionary wavelet quantities and associated estimation theory that encompass variation across time *and* within- and between-trial. To obtain well-behaved, consistent spectral estimates we propose to perform local smoothing of the raw wavelet periodograms across trials, as opposed to employing the smoothing over time typically undertaken in the locally stationary process context. Trial-coherence estimation theory is also treated and is shown to provide novel, useful information about the process trial evolution.

The article proceeds as follows. Section 2 (i) introduces our proposed meta-process model that allows for correlation between trials while accounting for intra- and cross- trial nonstationarity, and (ii) develops its associated estimation theory. The generality of this new model allows us to treat the absence of cross-trial dependence, a fundamental assumption underpinning previous work, as its particular manifestation. Section 3 details simulation studies that showcase the practical behaviour of the proposed methodology and demonstrates its advantages. Section 4 details the application of the proposed methodologies to a real data study within neuroscience, where we crucially contrast the learning associations as evidenced at the level of the correct and incorrect trials. Section 5 concludes the paper.

## 2. The proposed multiple-trials nonstationary model

*Brief introduction of locally stationary wavelet processes* Before we describe our proposed framework, we recall some of the defining features of the *locally stationary wavelet* (LSW) framework of Nason *et al.* (2000). The LSW model provides a *time-scale* representation of nonstationary time series with time-varying second order structure, where the building blocks are the discrete non-decimated wavelets (see Vidakovic (1999) or Nason (2008) for an extensive introduction to wavelets). For $T = 2^{J(T)}$, a sequence of stochastic processes $\{X_{t;T}\}_{t=0,\ldots,T-1}$ is a LSW process if it admits the representation

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \omega_{j,k;T} \psi_{j,k}(t) \xi_{j,k},$$

where for scale $j$ and location $k$, $\omega_{j,k;T}$ is the amplitude corresponding to the discrete non-decimated wavelet $\psi_{j,k}(t)$ and $\{\xi_{j,k}\}$ are a set of orthonormal random variables. Modelling under the concept of local stationarity means that the variation of the amplitudes $\{\omega_{j,k;T}\}_k$, happens slowly over time and this is controlled by a smoothly varying continuous Lipschitz function $W_j(k/T)$, that can be thought of as a scale $(j)$ and time $(k)$ dependent transfer function (Fryzlewicz and Nason, 2006). Nason *et al.* (2000) further propose the *evolutionary wavelet spectrum* as a means to quantify the contribution to the overall process variance at scale $j = 1, 2, \ldots$ (corresponding to finest scale, second finest and so on) and rescaled time $z = k/T \in (0, 1)$ and define this as $S_j(z) = |W_j(z)|^2$. The raw wavelet periodogram is used for estimation of the spectrum and is defined as $I_{j,k;T} = |d_{j,k;T}|^2$ where $d_{j,k;T}$ denotes the wavelet coefficient at scale $j$ and time $k$ associated to a discrete non-decimated family of wavelets as defined in Nason *et al.* (2000).

The original LSW formulation is for a single-replicate. Its multivariate extension, the MV-LSW process of Park *et al.* (2014) can deal with a finite number of replicates but does not require, nor can it account for, the ordering of replicates. Thus, these modelling tools cannot capture the dynamics of time series data recorded for several trials over the course of an entire experiment, nor can they account for the potential dependence across trials. Our setting here presents additional challenges, notably the fact that these signals behave in a way that is nonstationary at multiple scales, (i) within the signal in each trial, and (ii) across trials over the course of the entire experiment.

### 2.1. *MULtiple-Trials Locally Stationary Wavelet (MULT-LSW) process*

A serious limitation of previous work is the assumption that the replicate time series are uncorrelated. We now develop the theory to allow for cross-trial dependence by means of a *between-trial* coherence structure. This is a major innovation of this work, as to the best of our knowledge this is the first paper that simultaneously accounts for correlation across trials while also embedding potential nonstationarity in multiple time scales: locally within a trial and globally across trials in the entire experiment. However, the lack of trial dependence can also be naturally incorporated by our proposed framework, as a particular case.

DEFINITION 1. We define a sequence of stochastic processes $\{X_{t;T}^{r;R}\}$, with time $t = 0, \ldots, T-1$ where $T = 2^{J(T)}$ and trial $r = 0, \ldots, R-1$ where $R = 2^{J(R)}$ to be a *multiple-trials locally stationary wavelet* (MULT-LSW) process if it admits the following representation

$$X_{t;T}^{r;R} = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \omega_{j,k;T}^{r;R} \psi_{j,k}(t) \xi_{j,k}^{r}. \tag{1}$$

For each trial (or replicate) $r$ and within-trial time $k$, $\omega_{j,k;T}^{r;R}$ are the amplitudes associated to the discrete non-decimated wavelets $\psi_{j,k}(t)$ at scale $j \geq 1$. Assume that within each replicate $r$, the innovations $\{\xi_{j,k}^{r}\}_{j,k}$ are a set of orthonormal random variables. Letting $\nu = r/R$ denote rescaled trial and $z = k/T$ denote rescaled within-trial time, the quantities in (1) possess the following properties:

1. For all $j$, $k$ and $r$, $\mathbf{E}[\xi_{j,k}^{r}] = 0$ ($\Rightarrow \mathbf{E}[X_{t;T}^{r;R}] = 0$).
2. In addition to $\{\xi_{j,k}^{r}\}_{j,k}$ being orthonormal within trial $r$, we have $\mathbf{E}[\xi_{j,k}^{r} \xi_{j',k'}^{r'}] = \delta_{j,j'} \delta_{k,k'} \rho_{j,k;T}^{r,r';R}$, where $\{\rho_{j,k;T}^{r,r';R}\}_k$ determine the dependence structure between trials $r$ and $r'$, at each scale $j$. Note that the within-trial orthonormality induces $\left|\rho_{j,k;T}^{r,r';R}\right| \leq 1$ for all $j$, $k$,$r$, $r'$, with equality when $r = r'$. The assumption of uncorrelated trials amounts to $\rho_{j,k;T}^{r,r';R} = \delta_{r,r'}$ for all $j$, $k$.
3. For each scale $j \geq 1$, there exists a Lipschitz continuous function in both rescaled time ($z$) and rescaled trial ($\nu$), denoted by $\widetilde{W}_j(z,\nu)$ with the following properties
   a)
   $$\sum_{j=1}^{\infty} \left|\widetilde{W}_j(z,\nu)\widetilde{W}_j(z,\nu')\right| < \infty \text{ uniformly in } z \in (0,1), \nu, \nu' \in (0,1). \tag{2}$$
   b) Let $L_j^{\nu}$ denote the bounded Lipschitz constant corresponding to the time dimension at a particular rescaled trial $\nu$ and scale $j$. Similarly, denote by $N_j^{z}$ the bounded Lipschitz constant corresponding to the trial dimension at a particular rescaled time $z$ and scale $j$. Denote $L_j = \sup_{\nu \in (0,1)} L_j^{\nu}$ and $N_j = \sup_{z \in (0,1)} N_j^{z}$ and assume they are uniformly bounded in $j$. Further assume that
   $$\sum_{j=1}^{\infty} 2^j L_j < \infty \text{ and } \sum_{j=1}^{\infty} 2^j N_j < \infty. \tag{3}$$
   c) There exist sequences of bounded trial-specific constants $\{C_j^{r}\}_r$ and time-specific constants $\{D_j^{k}\}_k$, such that for each $T$ and $R$ respectively, the amplitudes are forced to vary slowly across time within a trial and across trials, in the sense that
   $$\sup_{k=0:T-1} \left|\omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right)\right| \leq \frac{C_j^{r}}{T}, \quad \forall j, r, \tag{4}$$
   $$\sup_{r=0:R-1} \left|\omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right)\right| \leq \frac{D_j^{k}}{R}, \quad \forall j, k. \tag{5}$$
   Denote $C_j = \sup_r C_j^{r}$ and $D_j = \sup_k D_j^{k}$ and assume the sequences $\{C_j\}$, $\{D_j\}$ fulfill $\sum_{j=1}^{\infty} 2^j C_j < \infty$ and $\sum_{j=1}^{\infty} 2^j D_j < \infty$.
4. For each scale $j \geq 1$, there exists a Lipschitz continuous function in rescaled time $z$ and rescaled trial arguments $\nu$ and $\nu'$, denoted by $\rho_j(z,\nu,\nu')$, which constrains the innovation covariance structure and fulfills the assumptions below, as follows

a) Let $Q_j^{\nu,\nu'}$ denote the bounded Lipschitz constant corresponding to the time dimension at particular (rescaled) trials $\nu$ and $\nu'$, at scale $j$. Similarly, denote by $P_j^z$ the bounded Lipschitz constant corresponding to the trial dimension at a particular (rescaled) time ($z$), at scale $j$. Denote $Q_j = \sup_{\nu,\nu' \in (0,1)} Q_j^{\nu,\nu'}$, $P_j = \sup_{z \in (0,1)} P_j^z$ and assume they are uniformly bounded in $j$. Further assume that

$$\sum_{j=1}^{\infty} 2^j Q_j < \infty \text{ and } \sum_{j=1}^{\infty} 2^j P_j < \infty. \tag{6}$$

b) There exist sequences of bounded trial-specific constants $\{\tilde{C}_j^{r,r'}\}_{r,r'}$ and time-specific constants $\{\tilde{D}_j^k\}_k$, such that for each $T$ and $R$ respectively, the covariances are forced to vary slowly, in the sense that

$$\sup_{k=0:T-1} \left| \rho_{j,k;T}^{r,r';R} - \rho_j\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) \right| \leq \frac{\tilde{C}_j^{r,r'}}{T}, \quad \forall j, r, r' \tag{7}$$

$$\sup_{r,r'=0:R-1} \left| \rho_{j,k;T}^{r,r';R} - \rho_j\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) \right| \leq \frac{\tilde{D}_j^k}{R}, \quad \forall j, k. \tag{8}$$

Denote $\tilde{C}_j = \sup_{r,r'} \tilde{C}_j^{r,r'}$ and $\tilde{D}_j = \sup_k \tilde{D}_j^k$ and assume the sequences $\{\tilde{C}_j\}$, $\{\tilde{D}_j\}$ fulfill $\sum_{j=1}^{\infty} 2^j \tilde{C}_j < \infty$ and $\sum_{j=1}^{\infty} 2^j \tilde{D}_j < \infty$.

REMARK 1 (**rescaled time and trial evolution**). Within each scale $j$, the transfer function $\widetilde{W}_j(z, \nu)$ controls the (non-stochastic) evolution of the amplitudes, forcing them to vary slowly over *both* rescaled time ($z$) and trial ($\nu$) dimensions. At the heart of it, this assumption is enabled by the ordered nature of the experiment, not only through time within each trial, but also across trials. The evolution of the amplitudes over time within each trial (replicate) happens in a smooth manner. The evolution across trials is such that while the spectral properties of different replicates may also be different, however it is reasonable to assume that the spectral properties of neighboring trials are similar. Nevertheless, further apart replicates may display different traits. Such a meta-process evolution appears later in Figure 4 (Section 3.1).

REMARK 2 (**rescaled trial dependence**). Note that from equations (7) and (8) we directly obtain for each $T$ and $R$

$$\sup_{r,r'=0:R-1} \sup_{k=0:T-1} \left| \rho_{j,k;T}^{r,r';R} - \rho_j\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) \right| = \mathcal{O}(\tilde{C}_j T^{-1}) + \mathcal{O}(\tilde{D}_j R^{-1}).$$

Hence for some rescaled time $z$ and rescaled trials $\nu$ and $\nu'$, we have in the limit

$$\rho_j(z, \nu, \nu') = \lim_{\substack{T \to \infty \\ R \to \infty}} \left( \rho_{j,\lfloor zT \rfloor;T}^{\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor;R} \right),$$

where $\lfloor zT \rfloor$ and $\lfloor \nu R \rfloor$ denote the largest integer less than or equal to $zT$ and $\nu R$, respectively.

For a scale $j$ and (rescaled) time $z$, the quantity $\rho_j(z, \nu, \nu')$ thus gives a measure of the dependence between (rescaled) trials $\nu$ and $\nu'$, and it is zero for uncorrelated trials. We remind the reader that previous methodology was developed under the working assumption of uncorrelated trials.

2.2. *Trial evolutionary wavelet spectrum and coherence* As is common in spectral domain analysis (both Fourier and wavelet-based), we do not work directly with the time- and trial-specific multiscale transfer functions $\{\widetilde{W}_j(\cdot,\cdot)\}_j$, but instead we define a scale-dependent measure for the time and (cross-)trial contribution to the overall process variance.

As noted, the current LSW model quantities are not capable of capturing the multiscale evolution of brain signals along trials. Next, we develop a novel evolutionary wavelet spectrum capable to extract nonstationarity both in time and within- and between- trials.

DEFINITION 2. For a MULT-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 1, we define its trial evolutionary wavelet spectrum as follows.

1. The *within-trial evolutionary wavelet spectrum* at scale $j$, rescaled trial $\nu$, rescaled within-trial time $z$ is given by

$$S_j(z,\nu) = \left|\widetilde{W}_j(z,\nu)\right|^2 = \lim_{\substack{T\to\infty \\ R\to\infty}} \left( \left| \omega_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;R} \right|^2 \right). \tag{9}$$

2. The *between-trial evolutionary wavelet spectrum* defined at scale $j$, rescaled time $z$ within rescaled trials $\nu$ and $\nu'$ is given by

$$S_j(z,\nu,\nu') = \widetilde{W}_j(z,\nu)\widetilde{W}_j(z,\nu')\rho_j(z,\nu,\nu').$$

Note in the above notation that the spectrum corresponding to any rescaled trials $\nu = \nu'$ is the within-trial spectrum, i.e. $S_j(z,\nu,\nu) = S_j(z,\nu)$. (Alternatively, $|\tilde{W}_j(z,\nu)| = (S_j(z,\nu,\nu))^{1/2}$.) Also note that under the assumption of uncorrelated trials of Fiecas and Ombao (2016), the between-trial spectrum is zero, i.e. $S_j(z,\nu,\nu') = 0, \forall\, \nu' \neq \nu$.

From equations (4) and (5) we directly obtain that for each $T$ and $R$ we have

$$\sup_{r=0:R-1} \sup_{k=0:T-1} \left| \omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right) \right| = \mathcal{O}(C_j T^{-1}) + \mathcal{O}(D_j R^{-1}), \tag{10}$$

hence the right-hand equality in equation (9).

In the MULT-LSW setup, we thus quantify the between-trial dependence by means of

$$\rho_j(z,\nu,\nu') = \frac{S_j(z,\nu,\nu')}{\{S_j(z,\nu)S_j(z,\nu')\}^{1/2}}, \tag{11}$$

and we shall refer to it as the *locally stationary between-trial coherence*, with values ranging from $-1$, indicating an absolute negative correlation, to 1 indicating an absolute positive correlation. Uncorrelation across trials amounts to $\rho_j(z,\nu,\nu') = 0, \forall\, \nu' \neq \nu$.

REMARK 3 (MULT-LSW versus LSW processes). An innovation of the proposed MULT-LSW model is to impose within each scale not only a smooth spectral behaviour across each (replicate) time series, but also to constrain the 'meta'-spectral evolution across replicates to happen in a smooth manner, as detailed by the conditions in Definition 1. Note that a MULT-LSW process is thus *not* to be understood only as a collection of locally stationary wavelet (LSW) processes that happen to be observed across several replicates, as this would limit its capacity to represent multiscale behaviour across trials.

For completeness and in order to aid our theoretical developments (and proofs), below we also introduce the local covariance functions associated to a MULT-LSW process, but note that as brain signals are evaluated in the spectral domain, in practice we will not extensively pursue these quantities here.

DEFINITION 3. For a MULT-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 1, we define its *local autocovariance functions* as follows.

1. The *within-trial local covariance* at some rescaled time $z \in (0,1)$ within rescaled trial $\nu \in (0,1)$, at time-lag $\tau \in \mathbb{Z}$ is given by
$$c(z,\nu;\tau) = \sum_{j=1}^{\infty} S_j(z,\nu)\Psi_j(\tau),$$
where $\Psi_j(\tau) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{j,k}(\tau)$ denotes the scale $j$ autocorrelation wavelet.
2. The *between-trial local covariance* at rescaled time $z \in (0,1)$ within rescaled trials $\nu$ and $\nu'$, both in $(0,1)$, at time-lag $\tau \in \mathbb{Z}$ is given by
$$c(z,\nu,\nu';\tau) = \sum_{j=1}^{\infty} S_j(z,\nu,\nu')\Psi_j(\tau).$$
Observe that $c(z,\nu,\nu;\tau) = c(z,\nu;\tau), \forall\,\nu$, and under the assumption of uncorrelated trials, $c(z,\nu,\nu';\tau) = 0, \forall\,\nu' \neq \nu$.

Note that $|c(z,\nu,\nu';\tau)| < \infty$ follows directly from the coherence range between $-1$ and 1, and from the uniform bounds in lag ($\tau$) and rescaled replicates ($\nu$, $\nu'$) for both the limiting amplitudes and the autocorrelation wavelets (see equation (2)).

The local covariance defined above can be shown to be an approximation of the process covariance corresponding to particular rescaled replicate(s), as follows.

PROPOSITION 1. *For a MULT-LSW process* $\{X_{t;T}^{r;R}\}$ *with properties as in Definition 1,* $\left|\mathrm{cov}(X_{\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;T}, X_{\lfloor zT\rfloor+\tau;T}^{\lfloor \nu' R\rfloor;T}) - c(z,\nu,\nu';\tau)\right| = \mathcal{O}(T^{-1}) + \mathcal{O}(R^{-1})$, *uniformly in* $\tau$ *at (rescaled) time* $z$ *and trials* $\nu$, $\nu'$.

PROOF. The proof appears in Section G of the Supplementary Material (Embleton *et al.*, 2021) and uses the approximation properties in the definition of the MULT-LSW process. □

2.3. *Estimation theory* We start our proposed estimation procedure for the spectral quantities, by first computing the raw wavelet periodogram and exploring its asymptotic properties as an estimator for the true, unknown spectrum. We note here that a well-behaved spectral estimator can then be used to construct an estimator for the meta-process local autocovariance, both within- and between-trials, by directly replacing the unknown spectrum in Definition 3. However, the development and investigation of the local covariance, potentially along with the partial covariance and their estimators for MULT-LSW processes are beyond the current interest of this work and are left as further research (see Killick *et al.* (2020) for the local partial autocorrelation function in LSW processes).

DEFINITION 4. For a scale $j$ and time $k$, we define the *raw wavelet periodogram between trials* $r$ and $r'$ of a MULT-LSW process to be
$$I_{j,k;T}^{(r,r');R} = d_{j,k;T}^{r;R} d_{j,k;T}^{r';R},$$
where $d_{j,k;T}^{r;R} = \sum_{t=0}^{T-1} X_{t;T}^{r;R}\psi_{j,k}(t)$ are the process empirical wavelet coefficients constructed using a family of discrete non-decimated wavelets, $\{\psi_{j,k}(t)\}_{j,k}$.

Note that when $r = r'$, the above becomes the *within-trial raw wavelet periodogram* and we simplify notation as

$$I^{r;R}_{j,k;T} = \left| d^{r;R}_{j,k;T} \right|^2 .$$

On a practical note, issues occur during the construction of the wavelet coefficients, $d^{r;R}_{j,k;T}$, when the support of the wavelet filter extends beyond the sequence boundary. Periodic bounding and reflection at the boundary are common options to tackle this problem (see Nason and Silverman (1994)), and we utilise the former in this work.

For reasons that will become obvious next, we also define a transformed spectral quantity $\beta_j(z, \nu, \nu') = \sum_{l=1}^{\infty} A_{j,l} S_l(z, \nu, \nu')$, where $A_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau \in \mathbb{Z}} \Psi_j(\tau)\Psi_l(\tau)$ is the inner product matrix of the autocorrelation wavelets. When the (rescaled) replicates $\nu$ and $\nu'$ coincide, we simplify the notation and we equivalently refer to $\beta_j(z, \nu) := \beta_j(z, \nu, \nu)$. The invertibility of the matrix $A$ and boundedness of its inverse norm (Nason *et al.*, 2000) ensure that finding a well-behaved estimator of the meta-process wavelet spectrum $S$ is equivalent to finding a well-behaved estimator for the spectral quantity $\beta$. Hence we next focus on estimating $\beta$ and note that the theoretical results below are derived under the Gaussianity assumption.

REMARK 4 (**Normality assumption**). Constrained by theoretical arguments, the results below hold under the assumption that the innovations follow a Gaussian distribution, a commonplace assumption in time series analysis in general and in LSW modelling in particular, e.g. Oh *et al.* (2003); Van Bellegem and von Sachs (2008). The previous Fourier-based meta-process modelling under the restriction of uncorrelated trials (Fiecas and Ombao, 2016) also relies on the Gaussianity assumption applied to the neurological data. Gaussianity is shown to hold for data arising from other fields, e.g. for experimental circadian data (Hargreaves *et al.*, 2019). In Appendix B and Supplementary Material (Embleton *et al.*, 2021) Section B we discuss the robustness of our approach to departures from normality and show that the normality assumption is mostly tenable for our neurological experimental data pertaining to the correct trials, while the incorrect trials display some departures from these assumptions. As we shall see, these findings are in-line with the stronger evidence of learning displayed across the correct trials.

PROPOSITION 2. *For a MULT-LSW process $\{X^{r;R}_{t;T}\}$ as in Definition 1, the trial raw wavelet periodogram has the following asymptotic properties for any fixed scale $j$ and rescaled time $z$, within rescaled trials $\nu$ and $\nu'$:*

*Expectation*

$$\mathbf{E}\left[ I^{(\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor);R}_{j,\lfloor zT \rfloor;T} \right] = \beta_j\left(z, \nu, \nu'\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}), \tag{12}$$

*Variance*

$$var\left( I^{(\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor);R}_{j,\lfloor zT \rfloor;T} \right) = \beta_j\left(z, \nu, \nu\right) \beta_j\left(z, \nu', \nu'\right) + \beta_j^2\left(z, \nu, \nu'\right) + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}).$$

PROOF. The proof appears in Section E.1 of the Supplementary Material (Embleton *et al.*, 2021). □

From Proposition 2, we see that the raw periodogram is asymptotically unbiased for $\beta$, but inconsistent due its asymptotically non-vanishing variance. Thus we next propose to smooth the raw periodogram in order to obtain consistency, and then we will correct for bias to obtain an asymptotically unbiased estimator for $S$.

DEFINITION 5. We define a *trial-smoothed estimator* for the rephrased spectral quantity $\beta_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})$ to be

$$\tilde{I}_{j,k;T}^{(r,r');R} = (2M+1)^{-1} \sum_{s=-M}^{M} I_{j,k;T}^{(r+s,r'+s);R}, \tag{13}$$

where $(2M+1)$ is the length of the smoothing window and $M$ is an integer such that as $T, R \to \infty$, we have that $M \to \infty$ and $M/R \to 0$.

REMARK 5 (**smoothing across trials**). Unlike for the usual locally stationary processes where the periodogram is smoothed over time (and over frequency for the classical Fourier-based models) in order to achieve consistency, here we propose a smoothing procedure that operates over trials by locally averaging the spectral estimates across a window of neighbouring replicates. This approach is indeed theoretically justified by the assumption of spectral smoothness across the trial-dimension. In practice, these assumptions will have to be verified in order to determine some empirically guided choice of $M$, as seen in the simulation study (Section 3.1).

PROPOSITION 3. *Under the properties of Definition 1 and the additional assumption* $\sup_{z,\nu\in(0,1)} \sum_{\eta\in\mathbb{Z}} \left|c(z,\nu,\nu+\frac{\eta}{R};\tau)\right| = \mathcal{O}(1)$ *for any time lag* $\tau$*, the trial-smoothed wavelet periodogram in equation* (13) *has the following asymptotic properties for any fixed scale* $j$ *and rescaled time* $z$ *within rescaled trials* $\nu$, $\nu'$*:*

*Expectation*

$$\mathbf{E}\left[\tilde{I}_{j,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}\right] = \beta_j\left(z,\nu,\nu'\right) + \mathcal{O}(MR^{-1}) + \mathcal{O}(2^j T^{-1}),$$

*Variance*

$$var\left(\tilde{I}_{j,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}\right) = \mathcal{O}(2^{2j}M^{-1}) + \mathcal{O}(2^j R^{-1}) + \mathcal{O}(M^2R^{-2}).$$

*Additionally, under the assumption of uncorrelated trials, the within-trial variance satisfies*

$$var\left(\tilde{I}_{j,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu R\rfloor);R}\right) = \mathcal{O}(2^{2j}M^{-1}) + \mathcal{O}(2^j R^{-1}) + \mathcal{O}(MR^{-2}).$$

PROOF. Section E.2 of the Supplementary Material (Embleton *et al.*, 2021) contains the proof which manipulates the amplitude properties across trials as opposed to those across time, both in the absence and in the presence of cross-trial dependence. □

Note that as $T$, $R$ and $M \to \infty$ and using the condition $M/R \to 0$, the bias of the smoothed periodogram becomes asymptotically negligible, while its variance tends to zero for any fixed fine enough scale $j$ (with $2^j = o\left(\min\left\{T, R, (2M+1)^{1/2}\right\}\right)$). The usual bias–variance trade-off here is manifest through the increase of $M$ resulting in a decrease of the variance at the price of an increase in the bias. As the trial-smoothed periodogram proposed above is an asymptotically unbiased and consistent estimator for the true $\beta$, the relationship between the true spectral quantities $\beta$ and $S$ suggests a natural way of constructing a well-behaved spectral estimator for the unknown $S$.

PROPOSITION 4. *Under the assumptions of Proposition 3, the following is an asymptotically unbiased and consistent estimator for the unknown wavelet spectrum for each fixed scale* $j$ *and rescaled time* $z$ *within rescaled trials* $\nu$, $\nu'$

$$\hat{S}_j(z,\nu,\nu') = \sum_{l=1}^{J} A_{j,l}^{-1} \tilde{I}_{l,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}, \tag{14}$$

*where $A^{-1}_{j,l}$ is the $(j,l)$ entry of the inverse of the inner product matrix $A$ of the autocorrelation wavelets and $J = \lfloor \alpha J(T) \rfloor$ with $\alpha \in (0,1)$, provided that $M/R \to 0$ as $T, R$ and $M \to \infty$.*

PROOF. Section E.3 of the Supplementary Material (Embleton *et al.*, 2021) contains the proof which hinges on the properties of the trial-smoothed periodogram shown in Proposition 3 above. □

This paves the way towards proposing the between-trial coherence estimator

$$\hat{\rho}_j(z,\nu,\nu') = \frac{\hat{S}_j(z,\nu,\nu')}{\left\{\hat{S}_j(z,\nu)\hat{S}_j(z,\nu')\right\}^{1/2}}, \tag{15}$$

where the involved spectral quantities are consistently estimated as proposed in Proposition 4 and the use of the same smoothing windows guarantees that the values of the resulting coherence estimator are indeed quantities between $-1$ and 1 (see proof in Section H of the Supplementary Material (Embleton *et al.*, 2021)). The following proposition shows that the step of examining $\hat{\rho}_j(z,\nu,\nu')$ is theoretically justified.

PROPOSITION 5. *Under the assumptions of Proposition 3, the coherence estimator in (15) is asymptotically consistent for the true coherence $\rho_j(z,\nu,\nu')$ at each fixed scale $j$ and rescaled time $z$ within rescaled trials $\nu$, $\nu'$.*

PROOF. See Section E.4 of the Supplementary Material (Embleton *et al.*, 2021). □

REMARK 6 (**smoothing across trials; and within-trial**). The results in Proposition 3 highlight the small sample dependence of the bias and variance of the smoothed periodogram on the number of trials $R$, on the time series length $T$ and on the smoothing window $(2M+1)$, as well as well as on the ratio of (replicate) smoothing window to the total number of trials. While still having a bias–variance tradeoff, the variance can be further improved by additionally smoothing across the time-dimension.

Specifically, using a within-trial (time-)smoothing window of length $(2M_T+1)$ such that $M_T \to \infty$ and $M_T/T \to 0$ (the reader may also refer to Park *et al.* (2014)) and chosen as usual under LSW modelling (see e.g. Nason (2013)), and preserving the previous notation of $(2M+1)$ for the trial-smoothing window, we define the trial- and time-smoothed periodogram

$$\tilde{\tilde{I}}^{(r,r');R}_{j,k;T} = (2M+1)^{-1}(2M_T+1)^{-1} \sum_{s=-M}^{M} \sum_{t=-M_T}^{M_T} I^{(r+s,r'+s);R}_{j,k+t;T} \tag{16}$$

to act as an estimator for the transformed spectral quantity $\beta_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})$. We next show that this estimator has desirable asymptotic properties, leading to faster convergence than its counterpart involving only trial-smoothing.

PROPOSITION 6. *For a MULT-LSW process as in Definition 1 and satisfying the additional assumption of autocovariance summability, $\sup_{z,\nu\in(0,1)} \sum_{n\in\mathbb{Z}} \sum_{\eta\in\mathbb{Z}} \left|c(z,\nu,\nu+\frac{\eta}{R};n)\right| = \mathcal{O}(1)$, the smoothed trial- and time-specific wavelet periodogram defined in equation (16) has the following asymptotic properties for any fixed scale $j$ and rescaled time $z$ within rescaled trials $\nu$, $\nu'$:*

*Expectation*

$$\mathbf{E}\left[\tilde{\tilde{I}}^{(\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor);R}_{j,\lfloor zT \rfloor;T}\right] = \beta_j(z,\nu,\nu') + \mathcal{O}(M_T T^{-1}) + \mathcal{O}(MR^{-1}) + \mathcal{O}(2^j T^{-1}),$$

*Variance*

$$var\left[\tilde{\tilde{I}}^{(\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor);R}_{j,\lfloor zT \rfloor;T}\right] = \mathcal{O}(2^{2j}(M_T M)^{-1}) + \mathcal{O}(2^{2j} M_T^{-1} M^2 R^{-2}).$$

*Additionally, under the assumption of uncorrelated trials, the within-trial variance satisfies*

$$var\left[\tilde{\tilde{I}}^{(\lfloor \nu R \rfloor, \lfloor \nu R \rfloor);R}_{j,\lfloor zT \rfloor;T}\right] = \mathcal{O}(2^{2j}(M_T M)^{-1}) + \mathcal{O}(2^{2j} M_T^{-1} M R^{-2}).$$

PROOF. Section E.5 of the Supplementary Material (Embleton *et al.*, 2021) contains the proof which makes use of the smoothing in both directions. □

The trial- and time- smoothed periodogram can then be used to further build a well-behaved estimator of the unknown trial wavelet spectrum $S$ by means of

$$\hat{\tilde{S}}_j(z, \nu, \nu') = \sum_{l=1}^{J} A^{-1}_{j,l} \tilde{\tilde{I}}^{(\lfloor \nu R \rfloor, \lfloor \nu' R \rfloor);R}_{l,\lfloor zT \rfloor;T}.$$

It is straightforward to show that this is also asymptotically unbiased and consistent for $S_j(z, \nu, \nu')$, in the same manner as in the proof of Proposition 4.

**3. Simulation Study** This section explores through simulation the behaviour of our proposed methodology, first in the setting where the time series across trials are uncorrelated, and second where the dependence is taken into account.

3.1. *Simulation study under the assumption of uncorrelated trials* Here we aim to assess the behaviour of our proposed MULT-LSW methodology as well as compare it to a classical approach involving the LSW methodology (Nason *et al.*, 2000). Specifically, we evaluate (i) the classical approach where one would independently estimate the spectrum for each replicate using a localised time smoother and then average over all trials ('LSW'), (ii) our proposed methodology involving localised trial smoothing ('MULT-LSW$_1$'), and (iii) our proposed methodology involving localised trial and time smoothing ('MULT-LSW$_2$'). In order to match the current practice for LSW estimation, e.g. Nason *et al.* (2000); Park *et al.* (2014), we have set $J = J(T)$ (corresponding to $\alpha = 1$), although in a bivariate spectral estimation context Sanderson *et al.* (2010) set a similar measure to $\alpha = 0.7$ and remark on its improved results when compared to $\alpha = 1$. We carry out simulations over $N = 100$ runs and explore performance across a range of time series lengths $T$ from 128 to 1024, number of trials $R$ from 64 to 512 and smoothing windows $(2M + 1)$ from 9 to 25. We explore spectral structures pertaining to high- and low- frequency behaviour as well as combinations across scales (see also Embleton *et al.* (2020)). We report the mean squared errors (MSE) and squared bias results, and note these also implicitly infer the variance across simulations.

Overall, based on our findings, we recommend the use of trial and time smoothing methodology (MULT-LSW$_2$) with a window length of $(2M + 1)$ guided by the choice of $M = \frac{3}{4}\sqrt{R}$, thus ensuring that $M/R \to 0$.

**Illustrative example**. We choose to present here the behaviour of our proposed methodology on a process with a challenging spectral structure, as shown in Figure 3 and mathematically defined in Section C.1 of the Supplementary Material (Embleton *et al.*, 2021). Further simulation setups are detailed in Section C.2 of the Supplementary Material (Embleton *et al.*, 2021), including MSE and squared bias tables.

For $R = 256$ trials each of length $T = 256$ ($J(T) = 8$), the process places spectral content at level $j := J(T) - 3 = 5$, manifest through a decreasing amplitude of the cosine across the last 192 trials within the third finest scale, and at level $j := J(T) - 2 = 6$, where the periodicity of the cosine increases across the first 128 trials within the second finest scale.

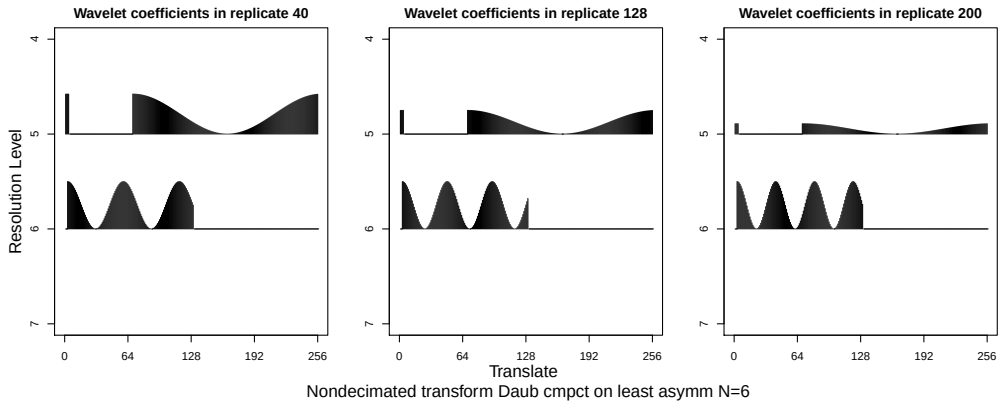


Fig 3: True evolutionary wavelet spectra plots for trials (replicates) 40, 128 and 200, depicting within-trial nonstationarity and evolution in the statistical characteristics across the trials. The plots use the resolution level notation $j := J(T) - j$ for the $j$th finest scale.

A concatenated realisation of this process is shown in Figure 4. Note however that this is an abuse of representation, since each trial is a time series of its own, and the sole purpose of this visualisation is to highlight the evolution of the meta-process. Furthermore this process departs somewhat from the requirement that the amplitudes evolve slowly over both rescaled time $(z)$ and replicate $(\nu)$ dimensions, however we show that despite this the methodology still performs well.

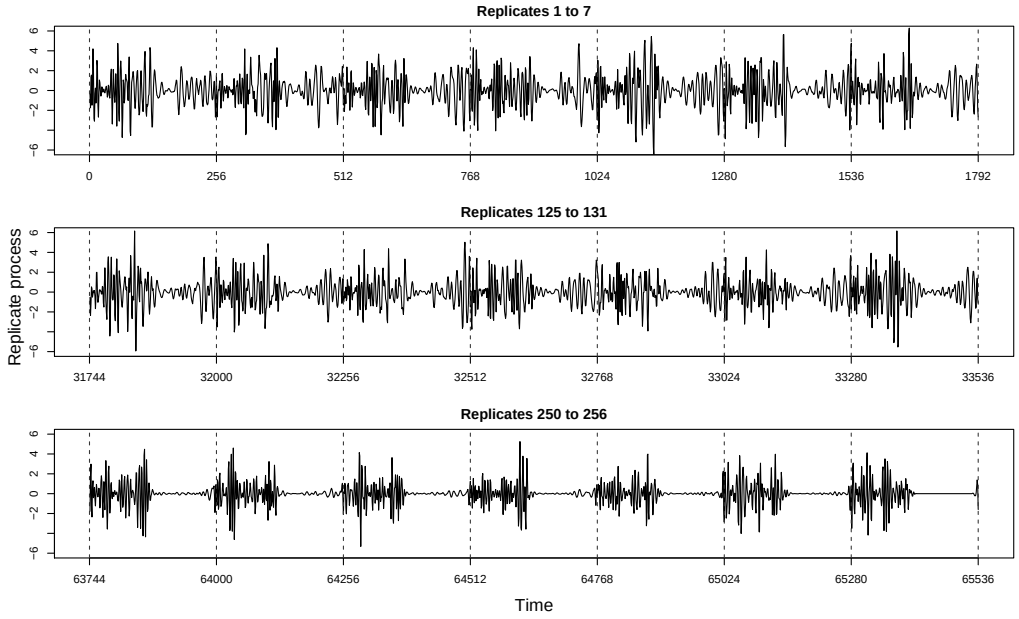


Fig 4: Realisation of a MULT-LSW process with spectra as in Figure 3. (Vertical lines denote breaks between trials. Concatenation is only used for meta-process visualisation.)

Spectral estimates have been computed using discrete non-decimated wavelets built by means of Daubechies Least Asymmetric family with 6 vanishing moments (Daubechies, 1992). For the MULT-LSW method, local averaging involved windows of 9 trials corresponding to $M = 4$ and we note that numerical MSE results in Appendix A.1 highlight that we chose

to visually present here some of our least performant results. The LSW and MULT-LSW($_1$) spectral estimates appear in Figure 5, along with the truth.

From the figures we get a visual clarification that the MULT-LSW model is doing a good job at capturing the evolving characteristics of the spectra across trials and the leakage across the neighbouring levels $j = 5$ and 6 is minor. They also highlight that when neglecting the possibility of evolutionary behaviour over replicates, when it is in fact present as seen for levels 5 and 6 in the top row plots of the true spectrum, the LSW model struggles to reflect this and either under or over-estimates, as seen in the middle row plots. The bottom row plots show that the MULT-LSW$_1$ estimates do indeed pick up the evolution over trials. Figure 6 provides 'single-shot' estimates of the spectrum to give evidence for the potential level of noise a user might experience in practice, while Figures 8 and 9 in Section C.1 of the Supplementary Material (Embleton *et al.*, 2021) further support the evidence for evolutionary behaviour of the spectral quantities both across time and trials.

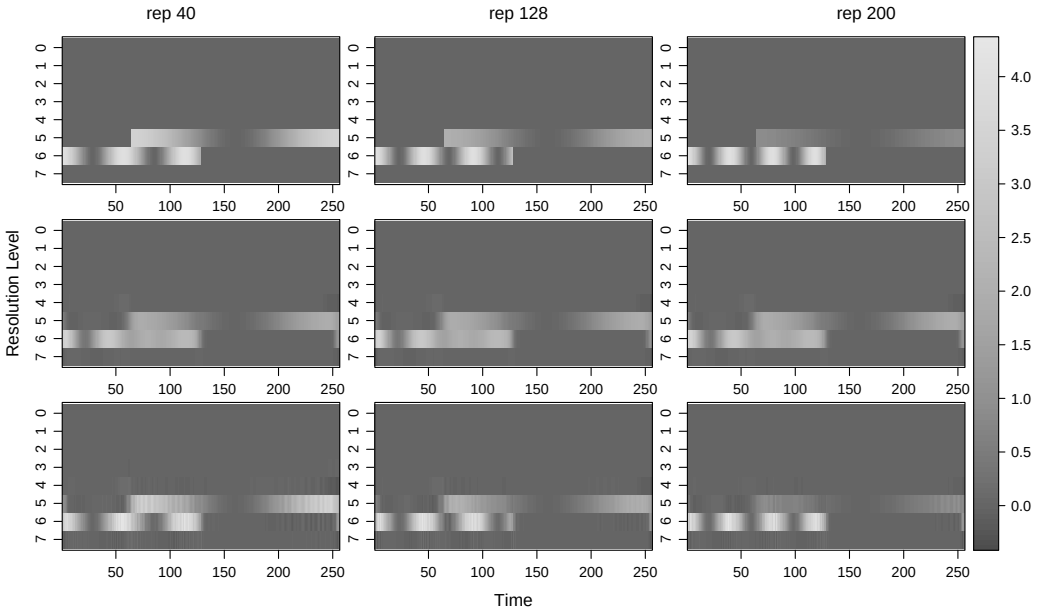


Fig 5: Time-scale plots for trials (replicates) 40, 128 and 200, respectively the first, second and third columns. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimates from the LSW method averaged over all replicates; *Bottom*: estimates using MULT-LSW($_1$). Plots depict multiscale within-trial nonstationarity and evolution in the statistical characteristics across the trials.

Histograms of the MSEs over the 100 simulations are shown in Section C.1 of the Supplementary Material (Embleton *et al.*, 2021) and highlight not only how the increase in $M$ improves performance but also how the increase in $R$ and $T$ reduces the MSEs, thus demonstrating the expected asymptotic behaviour of our smoothed estimator. To numerically strengthen our visual inference, we examine the MSEs and squared bias results in Table 1 (Appendix A.1). Note these results implicitly also provide evidence for the variance across the simulation runs. The best results in terms of lowest MSEs are obtained by MULT-LSW$_2$, despite MULT-LSW$_2$ incurring a somewhat higher bias than MULT-LSW$_1$. Both our methods have a substantially lower bias than the LSW. The higher bias of the LSW estimates is unsurprisingly resulting from averaging over all the replicates and thus failing to account for the evolutionary behaviour through trials. The benefit of taking a local smoothing approach over both time and trials is that it always results in spectral estimates with lower bias and MSE when

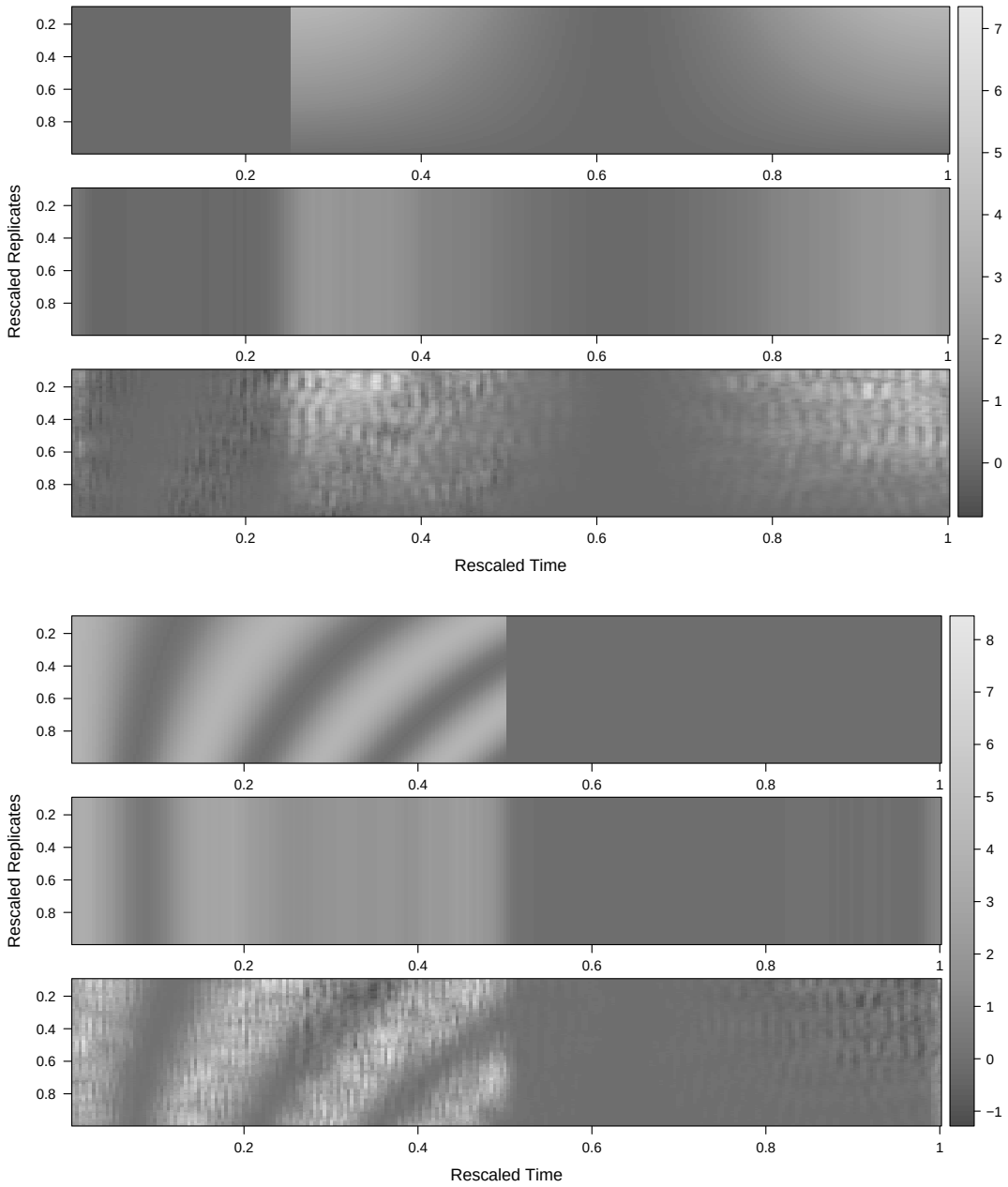


Fig 6: Time-replicate spectral plots with replicate smoothing window of 25 ($M = 12$) for levels 5 (top plot) and 6 (bottom plot). Estimates are for a single realisation. Within each plot, *top*: true spectra; *middle*: estimated spectra from the LSW method averaged over all replicates; *bottom*: estimates using MULT-LSW($_1$).

compared to LSW, although it is worth pointing out that taking a local smoothing approach over trials only, while yielding lower bias, might increase the MSE for inappropriately small windows.

Asymptotically, the MSEs associated to our methods decay much faster than for the LSW and as we increase the local averaging window length ($2M + 1$), the performance of our the MULT-LSW methodology improves. A replicate rectangular window with length determined by the choice $M = \frac{3}{4}\sqrt{R}$ appears to work well across all our investigations. Killick

*et al.* (2020) illustrate the robustness of LSW estimation to window width choices and form. However, we note the discussion in (Cryer and Chan, 2008, §14.2) and suggest that for a deeper understanding of the spectral characteristics, a user might wish to obtain estimates over a range of $M$, e.g. $M = \frac{1}{2}\sqrt{R}, \frac{3}{4}\sqrt{R}, \sqrt{R}$.

3.2. *Coherence illustration via simulation* We shall now investigate through simulation the performance of our proposed methodology for coherence estimation. We use the mean squared error (MSE) and squared bias of the estimates $\hat{\rho}$, averaged over all time-scale points and trials, as detailed in Appendix A.2. Note these also allow us to infer the variance across simulations. We display the behaviour of our estimators on a simulated example below, and provide a further simulation study in Section D of the Supplementary Material (Embleton *et al.*, 2021).

**Illustrative example**. We simulate a MULT-LSW process with $R = 256$ trials that feature dependence, measured at $T = 512 = 2^9$ time points. The within-trial wavelet spectra are defined by a sine wave whose periodicity and magnitude evolve slowly over the trials in such a way that the spectral characteristics of neighbouring replicates do not look too dissimilar whilst there is a noticeable difference between replicates further apart (for their mathematical expression, see Simulation 1, Section C.2 of the Supplementary Material (Embleton *et al.*, 2021)). Here we have $J(T) = 9$ (in short, $J$) and the spectral characteristics are placed in level $j = J(T) - 4 = 5$. In addition to the within-trial spectral characteristics, we also define a challenging between-trial spectral structure by means of defining their (true) coherence at each level $j$ and location $k$. For level $j = 5$ and time $k = 1, \ldots, 256$ we define the non-zero trial coherence matrices as follows: the first 128 trials have a strong positive coherence (0.99) with one another, however this coherence becomes negative (-0.71) with the last 128 trials. A (weaker) positive coherence (0.5) also exists between the last 128 trials. Over the last 256 locations, we set the coherence to be zero. The expressions of the non-zero coherence matrices appear in Section C.2 of the Supplementary Material (Embleton *et al.*, 2021) and the illustrative true coherence structures for replicates 50 (top row) and 200 (bottom row) can be visualised in Figure 7 (left panels).

Coherence estimates obtained using the methodology proposed in Section 2.3 are represented in Figure 7 (right panels) for trials 50 (top row) and 200 (bottom row). Non-decimated discrete wavelets built using Daubechies least asymmetric family with 10 vanishing moments and local averaging over a window of 9 trials were employed.

In terms of correctly estimating the coherence structure switch over times and trials, as well as identifying the positive or negative character of the coherence, the proposed estimation procedure does a good job. We do however note that the estimated coherence intensity does exhibit some bias, which we may attribute to the smoothing performed in order to address the practical computation considerations (see the remark in Appendix A.2). Nevertheless, we could argue that the model does give a good indication for the degree of the positiveness of the coherence, approximately 0.99 and 0.5 for replicates 50 and 200 respectively (right panels of Figure 7). Recall that this dependence structure would otherwise be invisible to previous approaches (Fiecas and Ombao, 2016), with consequent implications of decreased spectral estimation precision when compared to estimation that takes into account cross-trial dependence.

Tables 2 (Appendix A.2) and 4 (Section D in the Supplementary Material (Embleton *et al.*, 2021)) illustrate MSE results for two smoothing approaches: the first involves smoothing only over a window of trials; the second involves local averaging through trials and time. For both simulations, the results paint the same picture. As we increase the trial smoothing window (such that $M/R \to 0$) the performance of our models improves in terms of MSEs, and the double smoothing over trial and time further reduces the MSEs. The price to pay for double

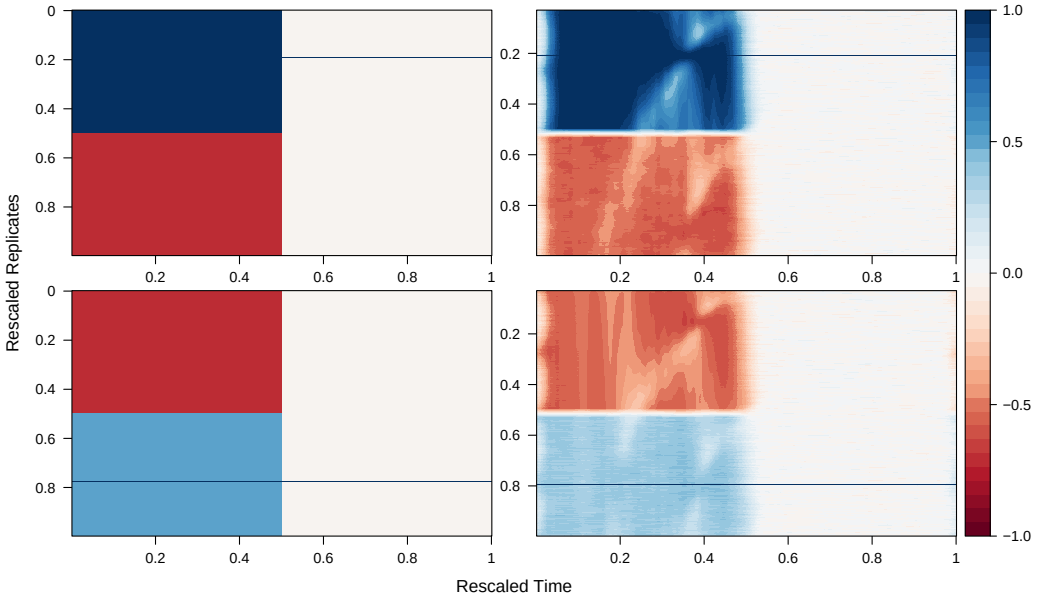


Fig 7: Coherence plots for replicates 50 (*top row*) and 200 (*bottom row*) over rescaled time and replicates in level 5. *Left*: true coherence; *Right*: coherence estimates averaged over 100 simulations. The horizontal line indicates an autocoherence value of 1.

smoothing as usual is a slightly higher bias (than when using averaging over trials only). In order to ensure that our spectral estimates are positive, our correction procedure uses the correction matrix $A^{-1}$ truncated at zero. Inevitably this introduces bias, evident through the increasing MSEs as $R$ and $T$ increase. In their work on bivariate channel coherence estimation, Sanderson *et al.* (2010) reported better results when additionally employing smoothing over scales. We conjecture that this is also applicable for our work, but leave the further numerical treatment for future research.

**4. Analysis of Macaque Local Field Potentials** We perform our MULT-LSW spectral analysis on the correct and incorrect trial datasets of local field potentials (LFPs) from the hippocampus (Hc) and nucleus accumbens (NAc) over the course of an associative learning experiment involving a single macaque monkey in a neurobiology laboratory at the Massachusetts General Hospital. Due to their roles in the consolidation of memory information and the processing of rewarding stimuli, the Hc and NAc have been studied in relation to learning tasks for monkeys, rats and humans (Wirth *et al.*, 2003; Abela *et al.*, 2015; Seger and Cincotta, 2006). Our analysis yields new evidence for the different learning associations exhibited across the correct versus incorrect trials, thus offering not only confirmations to the results of previous studies, but also providing additional insights to the understanding of the dynamics of the LFPs. This is achieved by capturing the evolutionary characteristics of brain processes *within* and *across* the trials of the experiment, in a *scale-dependent* manner, through the use of the wavelet transform. Notable benefits of our model in contrast to previous Fourier-based methodology (Fiecas and Ombao, 2016) are (i) its flexibility to represent potential *between-trial* dependence, (ii) its superior *time-localisation* and (iii) further understanding of the *within-trial* variation attained through the multiscale decomposition in the (within-trial) time direction, as we will see next.

Each trial consists of $T = 2048$ time points, corresponding to approximately 2 seconds of data. The design splits each trial into four time blocks of 512 milliseconds each and it is in the final time block that the macaque was tasked with associating one of four doors (appearing on

a screen) with the picture visual presented in the second time block. The macaque had to learn the associations through repeated trials and the macaque is given the message that the selected association was correct through a juice reward. Further details are given in Section A of the Supplementary Material (Embleton *et al.*, 2021). We carry out the analysis on $R = 256$ trials (replicates). The data has been grouped into sets of "correct" and "incorrect" responses. This approach departs from the global trial approaches of Chau and von Sachs (2016); Fiecas and Ombao (2016) in order to additionally investigate how the contributions of the Hc and NAc to the learning process differ between these groups (Gorrostieta *et al.*, 2012). A more subtle implication is the fact that local averaging is carried out in this context across like-for-like trials, rather than both correct and incorrect trials potentially contributing to the final spectral estimate.

To ensure comparability of the wavelet spectra across trials, each trial is standardised to have mean zero and unit variance. Plots for the correct responses appeared in Figures 1 and 2 (Section 1) and for the incorrect responses in Figures 1 and 2 (Supplementary Material (Embleton *et al.*, 2021) Section A). Appendix B and Supplementary Material (Embleton *et al.*, 2021) Section B provide evidence that the Gaussian assumption is mostly tenable for the macaque correct trials and presents occasional departures for the incorrect trials, while also empirically discussing robustness to departures from normality.

REMARK 7 (**bootstrapped confidence intervals.**). As means to quantify the validity of our results, we construct pointwise in time (for each level $j$) confidence intervals for the wavelet spectrum and coherence via bootstrap sampling. In this context, it is vital to allow for the potential cross-trial dependence in order to avoid subsequent inference using erroneous confidence bands, as also strongly argued by Morris (2015); Chau and von Sachs (2016) in a functional context. In brief, the procedure uses our proposed MULT-LSW($_2$)-spectral estimates as proxy for the true, unknown spectral structure that embeds cross-trial dependence. These estimates are used for simulating $B$ bootstrap sampled MULT-LSW processes. Using our spectral estimation procedure for each of the $B$ samples, we then obtain a bootstrapped spectral estimate. This yields a bootstrapped distribution, from which pointwise confidence intervals for the spectrum and coherence are constructed. We expect that on average (over the bootstraps) the true spectral evolutionary characteristics are captured and that these characteristics are displayed in the lower and upper confidence bootstrap bounds, thus validating the identified spectral patterns of process evolution. Further details are given in Appendix C.

4.1. *Results* *Hippocampus*. The proposed spectral estimates for the correct and incorrect sets of trials appear in Figures 8 (below) and 3 (Section A, Supplementary Material (Embleton *et al.*, 2021)), respectively. The bottom row plots show the MULT-LSW($_2$)-spectral estimates averaged across 30 replicates to illustrate the process behaviour in the beginning, middle and end of the experiment. These demonstrate (i) the sequenced activation of within-trial time blocks and (ii) the evolutionary behaviour of the wavelet spectrum along the course of the experiment. These features are also captured by the lower and upper 95% bootstrap confidence bounds for the spectral characteristics, as illustrated by Figure 14 in Appendix C, thus rendering further confidence in the results that follow. Note that crucially these bounds incorporate the between-trial dependence, and thus differ from those obtained under the assumption of uncorrelated trials.

For both correct and incorrect datasets, the "activity" is primarily captured within the coarser levels 3 and 4 of the wavelet periodograms, approximately corresponding to frequencies 2-8Hz. Of these, the theta band frequencies 4-8Hz are typical of slow activity, known for their association to hippocampal activity in mammals and to promote memory (Buzsaki, 2006). Fiecas and Ombao (2016) report the low frequency range 1-12Hz to account for most

variability in the Hc data. Our analysis offers a finer characterisation that does not fully support activation of low delta waves (under 2Hz), known to be typical of deep sleep, and shows weak alpha band (8-12Hz) and low beta band (12-16Hz) alertness at certain time blocks within each trial. Due to its construction, the MULT-LSW model was able to capture the process evolutionary behaviour across trials, and thus overcomes the limitations of analyses using the standard LSW model. The LSW-based estimates do capture the evolution of the statistical properties within a trial but cannot characterise these changes across trials in the entire experiment. The MULT-LSW estimates have the capacity to highlight the individual time blocks as they activate through the course of the experiment, an insight invisible to LSW and weakly represented in the Fourier approach of Fiecas and Ombao (2016). Note that this activation could also be visualised by plotting spectral differences instead.

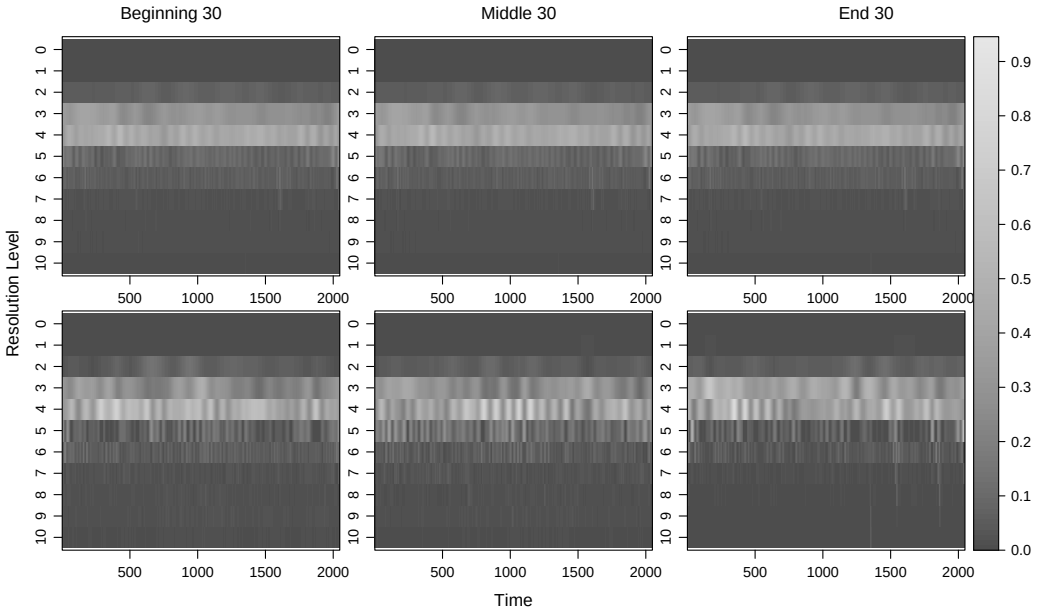


Fig 8: Time-scale hippocampus (Hc) plots for the correct trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: MULT-LSW method with smoothing over time and replicates.

For the correct Hc trials, the MULT-LSW model captures the bulk of "activity" in frequencies 2-8Hz. In early replicates this is fairly even through time, while for middle replicates the bursts of activity shift centrally within time, thus coinciding with the second block of the macaque being shown the visual stimulus and possibly with the expectation of the picture to continue being shown. For the final trials, the activity is clearly localised around time-point 500 (corresponding to the visual exposure) and towards the final quarter of time (corresponding to the time when the macaque made the correct association). When compared to a Fourier approach, our wavelet-based analysis thus brings to the fore novel information that links the experimental time blocks to Hc activation. Specifically, as the correct trials progress, the activity in the Hc is evident at the visual cue time and also at the selection task time, thus suggesting learning of the picture associations.

The incorrect trials still display evidence of evolutionary behaviour. However, as the experiment progresses, there is evidence of less spectral activity across these trials, with a brief Hc activation in the visual exposure block for the middling trials, and a burst of Hc activity

localised in the last time block, when the task is carried out, for the end trials. The spectrum suggests that whereas the Hc displays prolonged activity in the second time block for the correct trials (corresponding to the picture being presented), this feature is not as marked in the incorrect trials and thus the macaque is not making the association between the picture presented and the selection task. Scientific literature has shown (Seger and Cincotta, 2006) that during a learning experiment, activity in the Hc decreases as associations/rules are learned but would spike upon their application. The capacity of our model to extract localised information in time and within/ between-trial, thus highlights novel traits that suggest that the macaque in this experiment has not yet fully learned the associations, but evidence of learning is indeed present.

Figure 9 further illustrates that spectral evolution through time in scales 3 and 4 is captured by both the LSW and MULT-LSW models. As the experiment progresses, the MULT-LSW model identifies that across the correct trials the Hc displays (within-trial) time dependent peaks of activity that gradually span the course of the experiment. As we previously noted, the final trials display more Hc activity towards the visual exposure (time-point 500) and task times of the experiment, than the starting trials do, again indicating a learning process. In contrast, the incorrect trials display a much less structured behaviour, with time-dependent activity distributed more evenly across the replicates.

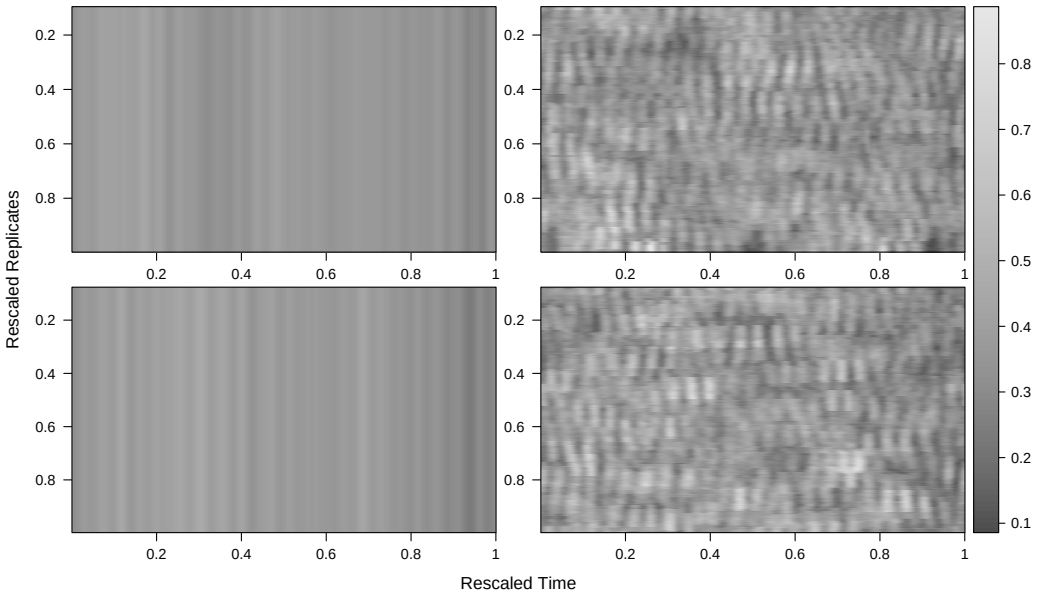


Fig 9: Time-replicate plots of the hippocampus (Hc) average spectra across 2-8Hz for the correct (*Top*) and incorrect (*Bottom*) trials. *Left*: estimates from the LSW method averaged over all replicates; *Right*: estimates from the MULT-LSW method with smoothing over time and replicates.

*Nucleus accumbens*. The resulting (MULT-)LSW spectral estimates appear in Figures 10 (below) and 4 (Section A, Supplementary Material (Embleton *et al.*, 2021)) for the correct and incorrect sets of trials, respectively. These plots are to be understood in the same manner as those for the hippocampus, with activation scales now 6 and 7. Reassuringly, the estimated spectral characteristics are also rendered by the lower and upper 95% bootstrap confidence bound plots, here omitted for brevity. Fiecas and Ombao (2016) find that the bulk of variability in the NAc is accounted for by (high) beta band frequencies (20-30Hz), while for both the correct and incorrect trials we place this in the wider range of beta band waves 16-30Hz,

associated to focussed activity. Our analysis also offers evidence for low gamma frequency waves (31-60Hz), typical of working memory activation (Iaccarino *et al.*, 2016). Additional to Fourier analysis, the MULT-LSW model also shows that nonstationarity across time is clearly present, as well as some spectral evolution across the trials. Although not as obvious as for the Hc, for the beginning and middle replicates of the correct group, NAc activity is manifest towards the trial start and end, while for the final replicates activity is captured in the final quarter of time. Also note that the NAc activity decreases in intensity from the beginning to end replicates for both groups. However, the patterns of behaviour in the beginning incorrect and correct trials are similar, but then gradually diverge as the experiment progresses.

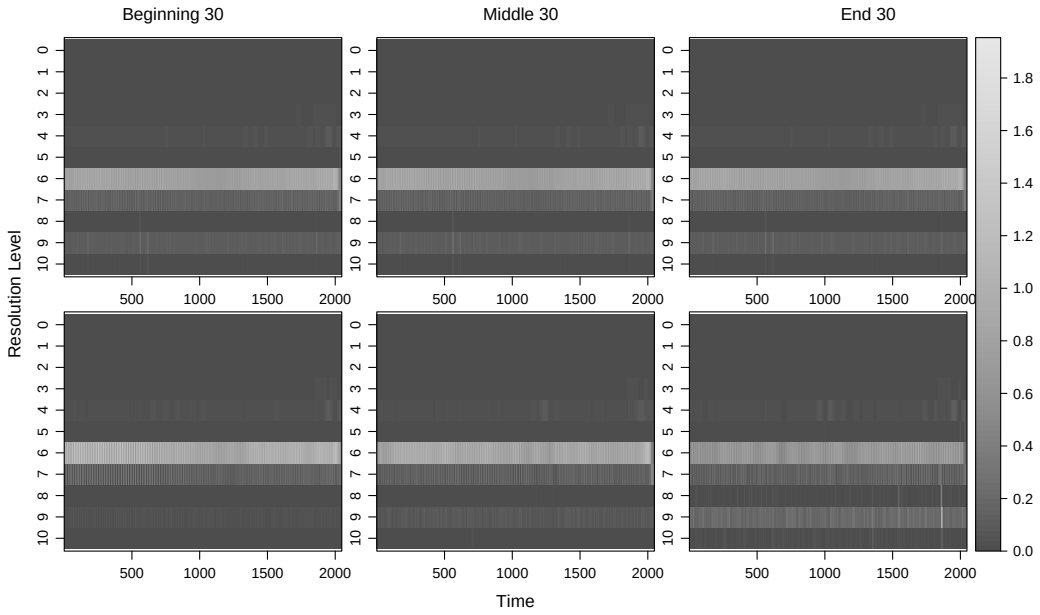


Fig 10: Time-scale nucleus accumbens (NAc) plots for the correct trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: MULT-LSW method with smoothing over time and replicates.

The NAc is part of the ventral striatum and plays a role in the processing of rewarding stimuli. The activity seen in the final 512 milliseconds can be attributed to the macaque expecting and receiving the juice reward in the correct trials, or expectation of reward in the incorrect trials. The impact of reward expectation (Schultz *et al.*, 1992; Hollerman and Schultz, 1998; Mulder *et al.*, 2005) could also explain the activity we see at the trial start for the beginning replicates and its observed periodicity across the experiment. Upon receiving no reward in an incorrect trial, the NAc activity decreases and with it the reward expectation falls for the next trial. Our analysis reflects the results of other studies on learning experiments (Hollerman and Schultz, 1998; Fiecas and Ombao, 2016) that highlight that the activity in the ventral striatum decreases as the stimuli are learned.

Figure 11 also reinforces our previous comments. Evolution in the spectra across time in scales 6 and 7 is captured by both models, with the NAc activity displaying periodic patterns, while the MULT-LSW estimation highlights a decrease in NAc activity along both the correct and incorrect trials of the experiment, with a more pronounced character for the incorrect trials.

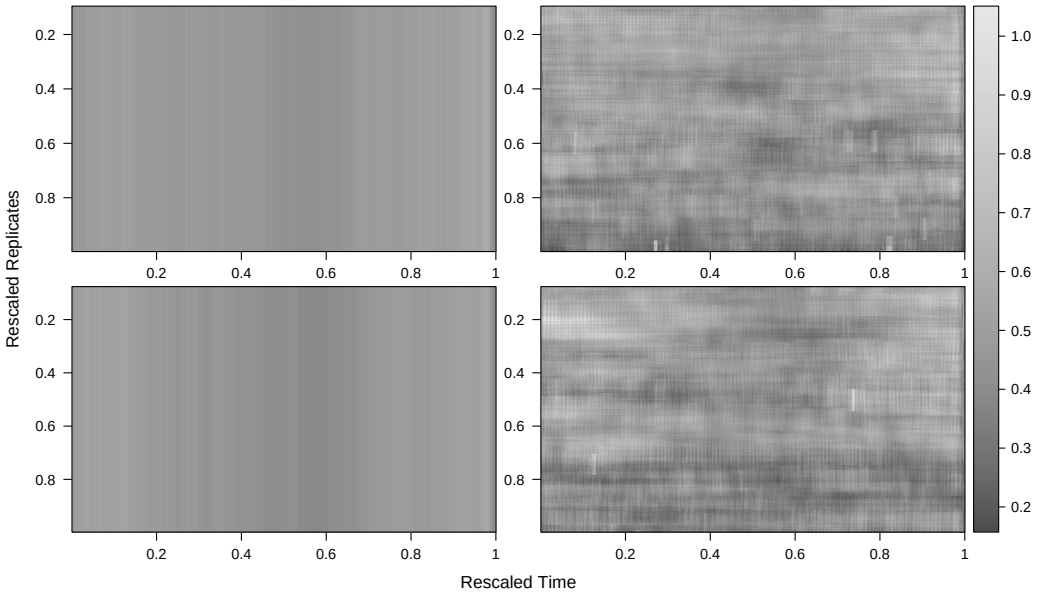


Fig 11: Time-replicate plots of the nucleus accumbens (NAc) average spectra across 16-60Hz for the correct (*Top*) and incorrect (*Bottom*) trials. *Left*: estimates from the LSW method averaged over all replicates; *Right*: estimates from the MULT-LSW method with smoothing over time and replicates.

REMARK 8. Our analysis demonstrates how the simplifying assumption of trials that are identical realisations of the same process, leading one to draw conclusions solely based on averaging across all replicates, could cause an important understanding in the process evolution through the experiment to be missed. Our proposed MULT-LSW methodology captures the spectral *time* and *within- and between-trial* evolutionary behaviour, thus yielding new scale-based results and advancing the findings of Fiecas and Ombao (2016) in the Fourier domain. We next explore the existence and strength of between-trial dependencies, and demonstrate the untenable position of the trial uncorrelation assumption.

4.2. *Capturing between-trial dependence* Let us now investigate between-trial dependencies, which under the MULT-LSW framework are readily quantified by the locally stationary between-trial coherence, a quantity defined in (11).

For the Hc activity, the MULT-LSW 95% bootstrap confidence bounds indicate that the presence of coherence across trials cannot be discounted, for both correct and incorrect groups. Moreover, our analysis finds evidence of a moderate dependence across neighbouring trials in the NAc at the beta band frequencies (16-30Hz), known to be responsible for brain activity related to reward feedback mechanisms. The estimated NAc trial coherence (absolute value) is shown in Figure 12 at level 6 for trials 20, 100 and 200 in the incorrect group and for correct trial 200, depicting typical behaviour. Some burst areas are present, indicating a moderate neighbouring trial coherence, with most meaningful values either side of 0.4 and a few above 0.5. For the beginning and middling incorrect trials, this is apparent in the time periods leading up to and inclusive of the trial task phase, upon which the macaque would receive a juice reward if the task was done correctly. The 95% bootstrap confidence bounds for the trial coherence characteristics (exemplified in Figure 5, Section A of the Supplementary Material (Embleton *et al.*, 2021)) indeed reinforce the presence of between-trial dependence.

REMARK 9. For this particular data, Chau and von Sachs (2016) illustrate the presence of between-trial correlation, with a stronger structure identified between neighbouring replicates

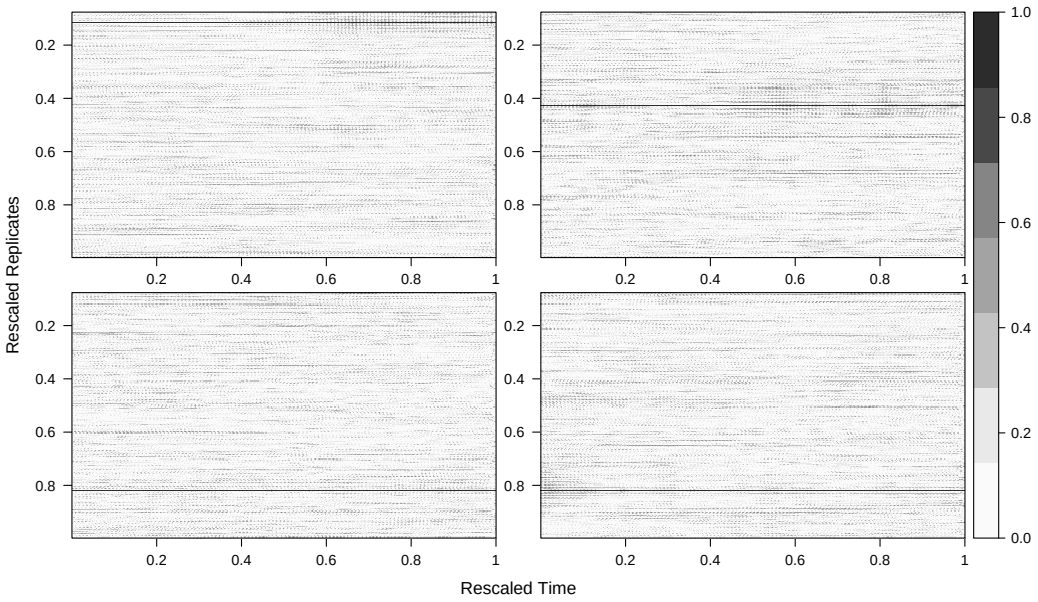


Fig 12: Level 6 absolute value coherence for NAc incorrect trials: 20 (*top left*), 100 (*top right*) and 200 (*bottom left*); and for NAc correct trial 200 (*bottom right*). The horizontal line indicates an autocoherence value of 1.

in the beginning and end of the experiment, but amassing individual trial timelines and frequencies of interest. Our analysis provides additional novel evidence in the *temporal and scale* (frequency)-dimensions that mild to moderate dependence is exhibited in both the correct and incorrect trials. This is primarily evident in the final correct trials, potentially as the manifest result of learning, and at the onset of the incorrect trials as the likely result to the expectation of reward. This finding, entirely invisible through the work of Fiecas and Ombao (2016), impacts our understanding of the experiment, and should not be ignored or assumed inexistent from the onset of the analysis. Indeed, being able to identify and quantify the extent of changes across trials in the entire experiment, not only has a qualitative impact on our understanding of the meta-process evolution, but also carries a substantive quantitative impact through e.g. producing incorrect spectral estimates and confidence bands (Chau and von Sachs, 2016), and thus potentially yielding misleading results (Morris, 2015).

**5. Concluding remarks** Our work proposed a novel wavelet-based methodology that successfully captures nonstationary process characteristics for time series collected across trials of a designed experiment. Its desirable properties were evidenced by simulation studies and through a real data application from the neurosciences. However, the methodology itself is not restricted to use within this field, and the authors envisage its utility in other experimental areas where wavelet spectral analysis has proved to be ideally suited, e.g. circadian biology (Hargreaves *et al.*, 2019). Nevertheless, we point out that the practitioner should exhibit care in their approach and ensure that indeed a natural trial ordering is present in their experiment. Should this not be the case, clustering and then averaging within each cluster might still be the preferred, appropriate approach (Hargreaves *et al.*, 2018; Ting *et al.*, 2018).

This work demonstrates the dangers of approaching experimental trial series as identical process realisations and the misleading results this can yield when studying the process dynamics along the course of the experiment. Notably, the proposed statistical model and associated estimation theory encompass meta-processes that feature dependence between

trials, and the assumption of trial uncorrelation, if appropriate, is dealt with by the proposed methodology as a particular case. A next natural step would be to investigate the MULT-LSW local (partial) autocorrelation function (for the LSW setup, see Killick *et al.* (2020)). A different avenue for research would be to investigate a flexible MULT-LSW construction hinged upon a trial-adaptive wavelet construction, which would in turn further tune the spectral and coherence estimates.

## APPENDIX A: SUPPORTING EVIDENCE FOR SIMULATION STUDIES

**A.1. MSE and squared bias tables for Section 3.1** As means to quantify the performance of the models, we employ the mean squared error (MSE) and squared bias, calculated as the average over all time-scale points and trials as follows

$$MSE(\hat{S}) = (RJT)^{-1} \sum_{r,j,k} \left[ \frac{1}{N} \sum_{n=1}^{N} \left( \hat{S}_j^{(n)} \left( \frac{k}{T}, \frac{r}{R} \right) - S_j \left( \frac{k}{T}, \frac{r}{R} \right) \right)^2 \right],$$

$$Bias^2(\hat{S}) = (RJT)^{-1} \sum_{r,j,k} \left[ \frac{1}{N} \sum_{n=1}^{N} \hat{S}_j^{(n)} \left( \frac{k}{T}, \frac{r}{R} \right) - S_j \left( \frac{k}{T}, \frac{r}{R} \right) \right]^2.$$

Note that although not explicitly calculated, the variance in estimates across simulations can thus be implicitly obtained from the above measures.

*Remarks on estimates at the boundaries*. We do not assess edges that involve local averaging over the first and last $(M-1)$ replicates. This has also been accounted for when calculating the MSE and squared bias. As a result, the reported measures whose values correspond to modelling via LSW will appear to change (in a very minor way), when in fact they should be the same for all choices of $M$ at fixed $R$ and $T$.

TABLE 1

*MSE and squared bias results, averaged over all time-scale points and trials for the simulation in Section 3.1 and $N = 100$ runs. 'LSW' denotes the classical approach of averaging over the replicates. 'MULT-LSW$_1$' denotes our proposed approach using localised trial smoothing. 'MULT-LSW$_2$' denotes our proposed approach using localised trial and time smoothing. Our proposed methods 'MULT-LSW$_1$'and 'MULT-LSW$_2$' use a trial smoothing window of length $(2M+1)$, while the time-smoothing window for 'LSW' and 'MULT-LSW$_2$' is automatically chosen.*

Mean squared errors ($\times 100$)

| | | | LSW | | MULT-LSW$_1$ | | MULT-LSW$_2$ | |
|---|---|---|---|---|---|---|---|---|
| R | T | M | mse | bias$^2$ | mse | bias$^2$ | mse | bias$^2$ |
| 256 | 128 | 4 | 17.55 | 17.27 | 22.31 | 2.53 | 11.52 | 3.88 |
| | | 12 | 16.74 | 16.46 | 9.45 | 2.40 | 6.54 | 3.82 |
| | 256 | 4 | 14.21 | 13.95 | 19.62 | 1.04 | 8.46 | 1.29 |
| | | 12 | 13.48 | 13.22 | 7.58 | 0.92 | 3.83 | 1.23 |
| | 512 | 4 | 12.25 | 12.01 | 17.59 | 0.53 | 7.15 | 0.52 |
| | | 12 | 11.59 | 11.35 | 6.51 | 0.42 | 2.85 | 0.48 |

Mean squared errors ($\times 100$)

| | | | LSW | | MULT-LSW$_1$ | | MULT-LSW$_2$ | |
|---|---|---|---|---|---|---|---|---|
| R | T | M | mse | bias$^2$ | mse | bias$^2$ | mse | bias$^2$ |
| 512 | 256 | 4 | 14.29 | 14.14 | 19.68 | 1.05 | 8.49 | 1.28 |
| | | 12 | 13.89 | 13.76 | 7.62 | 0.92 | 3.81 | 1.22 |
| | 512 | 4 | 12.30 | 12.18 | 17.61 | 0.53 | 7.15 | 0.52 |
| | | 12 | 11.96 | 11.84 | 6.54 | 0.42 | 2.85 | 0.48 |
| | 1024 | 4 | 10.92 | 10.81 | 15.83 | 0.32 | 6.29 | 0.25 |
| | | 12 | 10.61 | 10.50 | 5.78 | 0.22 | 2.38 | 0.21 |

**A.2. MSE and squared bias tables for Section 3.2** As numerical tools that quantify the performance of an estimate $\hat{\rho}_j(z,\nu)$ across all time points $z = k/T$ with $k = 0,\dots,T-1$, trials (replicates) $\nu = r/R$ with $r = 0,\dots,R-1$ and scales $j = 1,\dots,J$, we will use the mean squared error and squared bias, defined in this context as

$$MSE(\hat{\rho}) = (R^* J T)^{-1} \sum_{r,r^*,j,k} \left[ \frac{1}{N} \sum_{n=1}^{N} \left( \hat{\rho}_j^{(n)} \left( \frac{k}{T}, \frac{r}{R}, \frac{r^*}{R} \right) - \rho_j \left( \frac{k}{T}, \frac{r}{R}, \frac{r^*}{R} \right) \right)^2 \right],$$

$$Bias^2(\hat{\rho}) = (R^* J T)^{-1} \sum_{r,r^*,j,k} \left( \frac{1}{N} \sum_{n=1}^{N} \hat{\rho}_j^{(n)} \left( \frac{k}{T}, \frac{r}{R}, \frac{r^*}{R} \right) - \rho_j \left( \frac{k}{T}, \frac{r}{R}, \frac{r^*}{R} \right) \right)^2,$$

where due to the symmetry of the coherence matrix, we have used $R^* = R(R+1)/2$ and $r^* = r' \geq r$, and $N$ denotes the number of simulation runs. As in Section 3.1, we also adopt here $J = J(T)$.

*Remarks on implementation*. From the theoretical model construction, the within-trial spectra are positive quantities. However, our spectral estimates may take values that are negative or close to zero after correction, and this in turn can cause problems when normalising for coherence estimation. In order to bypass this issue, we choose to correct our raw wavelet periodogram estimates before smoothing. The theoretical properties of the coherence estimator show that using trial-smoothing does yield an estimator with good properties, albeit its rate of convergence is dependent on the smoothing window width $(2M+1)$. A local averaging window over time for smoothing each replicate before applying smoothing over replicates could also be employed, just as proposed for spectral estimation in Section 2.3. A possible further step, to be done after smoothing through time, is to smooth over scales as proposed by Sanderson *et al.* (2010) for estimating the linear dependence between bivariate LSW time series. However, we do not pursue this approach here.

TABLE 2

*MSE and squared bias* $(\times 100)$*, averaged over all time-scale point and trials for the simulation in Section 3.2 and* $N = 100$ *runs. Subscripts 1 and 2 denote the models with smoothing over trials only and trial-time smoothing, respectively.*

| R | T | M | $\text{mse}_1$ | $\text{bias}^2{}_1$ | $\text{mse}_2$ | $\text{bias}^2{}_2$ |
|---|---|---|---|---|---|---|
| 128 | 256 | 7 | 18.10 | 9.02 | 15.50 | 10.71 |
| | | 12 | 14.53 | 8.22 | 13.05 | 9.64 |
| 128 | 512 | 7 | 18.91 | 10.56 | 16.72 | 12.37 |
| | | 12 | 15.10 | 9.40 | 13.95 | 10.92 |
| 256 | 512 | 7 | 20.82 | 11.58 | 18.40 | 13.60 |
| | | 12 | 17.78 | 10.93 | 16.43 | 12.81 |

## APPENDIX B: SUPPORTING EVIDENCE FOR NORMALITY OF MACAQUE DATA

In what follows we explore the tenability of the Gaussian assumption, as suggested by Fryzlewicz (2005) and also demonstrated by Hargreaves *et al.* (2019). For each trial, we propose to standardise the (zero-mean) process using a localised estimate of the standard deviation. The

estimate was obtained by means of taking the square root of the estimated lag zero localised autocovariances (Nason, 2013). Further, we have found similar results when obtaining the estimate by means of a localised Gaussian kernel with bandwidth chosen using the methods of Fryzlewicz (2005). For the correct and incorrect trials respectively, in Figures 13 and 6 (Section B, Supplementary Material (Embleton *et al.*, 2021)), we report Q–Q plots of standardised series against the normal quantiles. These correspond to trials displaying typical behaviour for Hc and NAc series, both from the correct and incorrect sets. The plots show that the normality assumption mostly holds, with some slight departures for NAc records.

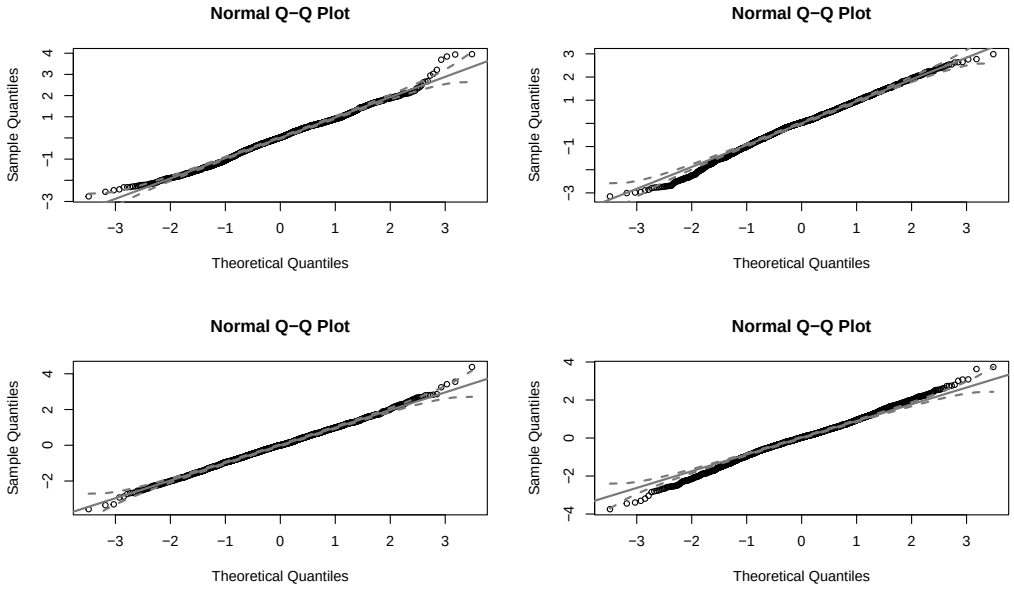


Fig 13: Q-Q plots of the Hc (*top row*) and NAc (*bottom row*) data in the correct trials (replicates). *Left column*: replicate 121; *Right column*: replicate 231. 95% confidence bounds are given by the dashed grey lines.

## APPENDIX C: CONSTRUCTION OF THE BOOTSTRAPPED CONFIDENCE INTERVALS

To construct pointwise confidence intervals, we use the MULT-LSW process Definition 1 and take the square root of our MULT-LSW$(_2)$-spectral estimates, $\hat{\hat{S}}_j(\frac{k}{T}, \frac{r}{R})$ to replace the within-trial amplitudes $\omega^{r;R}_{j,k;T}$ of the non-decimated wavelets $\psi_{j,k}(t)$ (see equation (9)) for each trial $r$, within-trial time $k$ and scale $j$. Furthermore, to allow for cross-trial dependence (see property 2 of Definition 1) we obtain estimates, $\hat{\rho}_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})$, of the innovation dependence structure between trials $r$ and $r'$.

Then for each $b = 1, \ldots, B$ bootstraps, we simulate a bootstrapped sample MULT-LSW process $\{X^{(b);r;R}_{t;T}\}$ following Definition 1

$$X^{(b);r;R}_{t;T} = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \hat{\hat{S}}_j \left(\frac{k}{T}, \frac{r}{R}\right)^{\frac{1}{2}} \psi_{j,k}(t) \tilde{\xi}^{(b);r}_{j,k},$$

with $r = 0, \ldots, R-1$ trials across time $t = 0, ..., T-1$. Within each trial $r$, the innovations $\{\tilde{\xi}_{j,k}^{(b);r}\}_{j,k}$ are assumed to be a set of orthonormal random variables, while the between-trial innovation dependence is established through the estimated coherence $\hat{\rho}_j$ above.

For each bootstrap $b = 1, \ldots, B$, we obtain a bootstrapped MULT-LSW$(_2)$ estimate, $\{\hat{\tilde{S}}_j^{(b)}(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})\}$, of the spectral quantities underlying the simulated process $\{X_{t;T}^{(b);r;R}\}$. Repeating this for $B$ bootstraps will then yield a bootstrap distribution for the smoothed wavelet periodograms, from which pointwise confidence intervals for the spectrum can be constructed in the usual manner by taking the respective quantiles, as illustrated below in Figure 14.

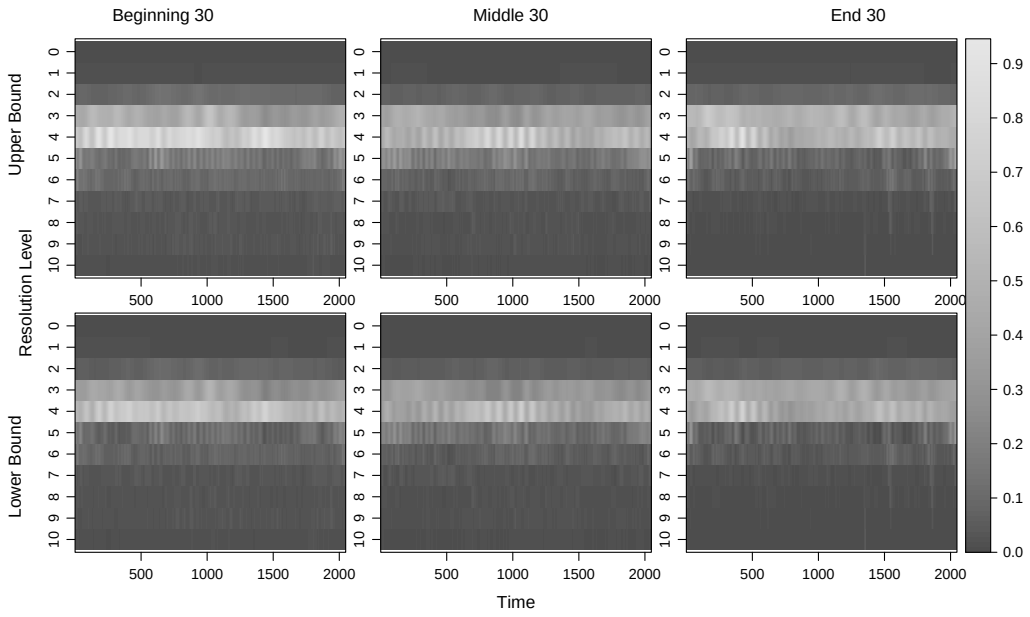


Fig 14: Time-scale 95% bootstrap confidence bounds for the hippocampus (Hc) spectrum for the correct trials. Spectral estimate bounds are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: upper bound; *Bottom*: lower bound. Estimation is via the MULT-LSW method with smoothing over time and replicates, and takes into account across-trial correlation.

## SUPPLEMENTARY MATERIAL

**Supplementary material for "Multiscale spectral modelling for nonstationary time series within an ordered multiple-trial experiment"**

The supplementary material contains a description of the experimental data, further details and figures for the macaque data study, additional simulation studies and all proofs. (pdf file) An associated R-package will be released on the Comprehensive R Archive Network (CRAN) in due course. The real data can be made available upon request.

## REFERENCES

Abela, A., Duan, Y. and Chudasama, Y. (2015). Hippocampal interplay with the nucleus accumbens is critical for decisions about time. *Eur. J. Neuroscience*, 42, 2224–2233.

Arieli, A., Sterkin, A., Grinvald, A. and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273, 1868–1871.

Buzsaki, G. (2006). *Rhythms of the brain.* Oxford University Press.

Chau, J. and von Sachs, R. (2016). Functional mixed effects wavelet estimation for spectra of replicated time series. *Electron. J. Statist.*, 10, 2461—2510.

Cryer, J. D. and Chan, K-S. (2008). *Time Series Analysis with Applications in R*. Springer.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.*, 25, 1–37.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM.

Diggle, P. and Al Wasel, I. (1997). Spectral analysis of replicated biomedical time series. *J. R. Statist. Soc. C*, 46, 31–71.

Embleton, J., Knight, M. I., and Ombao, H. (2020). Multiscale modelling of replicated nonstationary time series. *arXiv:2005.09440*.

Embleton, J., Knight, M.I. and Ombao, H. (2021). Supplementary material for "Multiscale spectral modelling for nonstationary time series within an ordered multiple-trial experiment."

Fiecas, M. and Ombao, H. (2016). Modelling the evolution of dynamic brain processes during an associative learning experiment. *J. Am. Statist. Ass.*, 111, 1440–1453.

Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *J. Appl. Stat.*, 32, 503–528.

Fryzlewicz, P. and Nason, G. P. (2006). Haar - Fisz estimation of evolutionary wavelet spectra. *J. R. Statist. Soc. B*, 68, 611–634.

Gott, A.N., Eckley, I.A. and Aston, J.A.D. (2015). Estimating the population local wavelet spectrum with application to non-stationary functional magnetic resonance imaging time series. *Statist. Med.*, 34, 3901–3915.

Gorrostieta, C., Ombao, H., Prado, R., Patel, S. and Eskandar, E. (2012). Exploring dependence between brain signals in a monkey during learning. *J. Time Ser. Anal.*, 33, 771–778.

Hargreaves, J.K., Knight, M.I., Pitchford, J.W., Oakenfull, R. and Davis, S.J. (2018). Clustering nonstationary circadian plant rhythms using locally stationary wavelet representations. *Multiscale Model. Simul.*, 16, 184—214.

Hargreaves, J.K., Knight, M.I., Pitchford, J.W., Oakenfull, R., Chawla, S., Munns, J. and Davis, S.J. (2019). Wavelet spectral testing: Application to nonstationary circadian rhythms. *Ann. Appl. Stat.*, 13, 1817–1846.

Hollerman, J. and Schultz, W. (1998). Dopamine neurons report an error in temporal prediction of reward during learning. *Nature Neuroscience*, 1, 304–309.

Huk, A., Bonnen, K. and He, B. J. (2018). Beyond trial-based paradigms: continuous behaviour, ongoing neural activity, and natural stimuli. *J. Neuroscience*, 38, 7551–7558.

Iaccarino, H., Singer, A.C., Martorell, A.J., Rudenko, A., Gao,F., Gillingham, T.Z., Mathys, H., Seo, J., Kritskiy, O., Abdurrob, F., Adaikkan, C., Canter, R.G., Rueda, R., Brown, E.N., Boyden, E.S. and Tsai, L.-H.(2016). Gamma frequency entrainment attenuates amyloid load and modifies microglia. *Nature*, 540, 230–235.

Killick, R., Knight, M.I., Nason, G. P. and Eckley, I.A. (2020). The local partial autocorrelation function and some applications. *Electron. J. Statist.*, 14, 3268–3314.

Krafty, R.T., Hall, M. and Guo, W. (2011). Functional mixed effects spectral analysis. *Biometrika*, 98, 583—598.

Martinez, J.G., Bohn, K.M., Carroll, R.J. and Morris, J.S. (2013). A study of Mexican free-tailed bat chirp syllables: Bayesian functional mixed models for nonstationary acoustic time series. *J. Am. Statist. Ass.*, 108, 514–526.

Morris, J.S. (2015). Functional Regression. *Ann. Rev. Stat. Appl*, 2, 321–359.

Mulder, A., Shibata, R., Trullier, O. and Wiener, S. (2005). Spatially selective reward site responses in tonically active neurons of the nucleus accumbens in behaving rats. *Exp. Brain Res.*, 163, 32–43.

Nason, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer.

Nason, G. P. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary times series. *J. R. Statist. Soc. B*, 75, 879–904.

Nason, G. P. and Silverman, B. W. (1994). The discrete wavelet transform in S. *J. Comput. Graph. Statist.*, 3, 163–191.

Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. B*, 62, 271–292.

Neumann, M. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, 25, 38–76.

Oh, H.-S., Ammann, C. M., Naveau, P., Nychka, D. and Otto-Bliesner, B. L. (2003). Multi-resolution time series analysis applied to solar irradiance and climate reconstructions. *J. Atmos. Sol.-Terr. Phys.*, 65, 191–201.

Ombao, H,. Raz, J., von Sachs, R. and Guo, W. (2002). The slex model of a non-stationary random process. *Ann. Inst. Statist. Math.*, 54, 171–200.

Ombao, H., von Sachs, R. and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Ass.*, 100, 519–531.

Park, T., Eckley, I. and Ombao, H. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Trans. Signal Process.*, 62, 5240–5250.

Priestley, M. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B*, 27, 204–237.

Qin, L., Guo, W. and Litt, B. (2009). A time-frequency functional model for locally stationary time series data. *J. Comput. Graph. Stat.*, 18, 675–693.

Saavedra, P., Hernandez, C.N. and Artiles, J. (2000). Spectral analysis with replicated time series. *Commun. Statist. - Theory Meth.*, 29, 2343–2362.

Sanderson, J., Fryzlewicz, P. and Jones, W. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, 97, 435–446.

Schultz, W., Apicella, P., Scarnati, E. and Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neuroscience*, 12, 4595–4610.

Seger, C. and Cincotta, C. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, 16, 1546–1555.

Ting, C.M., Ombao, H., Samdin, S.B. and Salleh, S.-H. (2018). Estimating Dynamic Connectivity States in fMRI Using Regime-Switching Factor Models. *IEEE Trans. Medical Imaging*, 37, 1011–1023.

Van Bellegem, S. and von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist.*, 36, 1879–1924.

Vidakovic, B. (1999). *Statistical Modelling by Wavelets*. John Wiley & Sons, Inc.

von Sachs, R. and Schneider, K. (1996). Wavelet smoothing of evolutionary spectra by nonlinear thresholding. *Appl. Comput. Harm. Anal.*, 3, 268–282.

Wirth, S., Yanike, M., Frank, L., Smith, A., Brown, E. and Suzuki, W. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science* , 300, 1578–1581.