A wavelet approach to modelling the evolutionary dynamics across ordered replicate time series

Jonathan Embleton

PhD

University of York
Mathematics

August 2021

Abstract

Experimental time series data collected across a sequence of ordered replicates often crop up in many fields, from neuroscience to circadian biology. In practice, it is natural to observe variability across time in the dynamics of the underlying process within a single replicate and wavelets are essential in analysing nonstationary behaviour. Additionally, signals generated within an experiment may also exhibit evolution across replicates even for identical stimuli.

We propose the Replicate-Evolving Locally Stationary Wavelet process (REv-LSW) which gives a stochastic wavelet representation of the replicate time series. REv-LSW yields a natural desired time- and replicate-localisation of the process dynamics, capturing nonstationary behaviour both within and across replicates, while accounting for between-replicate correlation. Firstly, we rigorously develop the associated wavelet spectral estimation framework along with its asymptotic properties for the particular case that replicates are uncorrelated. Next, we crucially develop the framework to allow for dependence between replicates. By means of thorough simulation studies, we demonstrate the theoretical estimator properties hold in practice.

Finally, it is unreasonable to make the typical assumption that all replicates stem from the same process if a replicate spectral evolution exists. Thus, we propose two novel tests that assess whether a significant replicate-effect is manifest across the replicate time series. Our modelling framework uses wavelet multiscale constructions that mitigate against the potential nonstationarities, across both times and replicates. Thorough simulation studies prove both tests to be flexible tools and allow the analyst to accordingly tune their subsequent analysis.

Throughout this thesis, our work is motivated by an investigation into the evolutionary dynamics of brain processes during an associative learning experiment. The neuroscience data analysis illustrates the utility of our proposed methodologies and demonstrates the wider experimental data analysis achievable that is also of benefit to other experimental fields, e.g. circadian biology, and not just the neurosciences.

Contents

List of Tables

List of Figures

Introduction

In an experimental setting, time series data are often collected across a sequence of repeated trials (also interchangeably referred to as replicates) with a natural ordering. A typical approach to analysing the underlying process would make inference on the average dynamics over all trials of the experiment. However, this approach naively assumes that all trials stem from the same process and fails to account for the possibility of a change in the process dynamics over the course of the experiment, even for identical stimuli. Such process behaviour was investigated in a recent study on neurological signals by Fiecas and Ombao (2016). 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 monkey 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. Variability in neuronal activity within both brain regions has also been observed over the trials of a learning experiment in other recent studies for humans, monkeys and rats (Seger and Cincotta, 2006; Gorrostieta *et al.*, 2011; Abela *et al.*, 2015). Such data experimental traits are not restricted to the neurosciences but can also be found in the analysis of, e.g. an organism's 'biological clock' in circadian biology (McClung, 2006), with such processes containing replicate nonstationary information (Hargreaves *et al.*, 2018, 2019). The process dynamics arising within experimental areas present the challenge of modelling time series that display potential evolutionary behaviour not only across time, but also across ordered trials.

Under certain statistical properties, such as process stationarity (the assumption that its mean and variance are constant over time), there exist many methods of statistical analysis of time series within the time domain, for example ARMA models, and within the spectral domain, such as classical Fourier representations of the periodogram (for an introduction, see Priestley (1982); Brockwell and Davis (1991); Box *et al.* (2008); Shumway and Stoffer (2017)). These methods are well-established in the analysis of time series arising from a variety of fields, for example, circadian biology, climatology, economics and neuroscience. 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 variation within the series, the developments within this thesis are primarily built to handle nonstationary data. Specifically, this thesis is concerned with the statistical modelling of an ordered collection of (constant mean) second-order nonstationary time series arising from the same stimulus, which we term a *replicate time series* or *meta-process*.

Wavelets can be thought of as localised 'little wave' functions with the ability to provide sparse multiscale representations of many signals. In this thesis, wavelets provide the fundamental 'building blocks' for our nonstationary process model, enabling us to capture localised replicate-, time- and scale- variability.

This thesis is motivated by the (briefly introduced above) specific neuroscience applica-

tion that records the brain process dynamics of a macaque monkey during an associative learning experiment. Our approach acknowledges the work of Fiecas and Ombao (2016), who developed their time–replicate model using Fourier waveforms, however we propose to take advantage of the multiscale and time localisation properties of wavelets. 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). 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.* Furthermore, considering that the signals are being recorded from the same subject (human, animal or plant, as dictated by the experimental field), it would not be unreasonable to assume that a dependence between trials could exist. For instance, some studies (Arieli *et al.*, 1996; Huk *et al.*, 2018) have documented evidence of correlation across trials in an experiment. Thus, to assume uncorrelation may be unrealistic and consequently produce misleading results from statistical inference. This highlights an important limitation of the work of Fiecas and Ombao (2016), namely that their methodology is developed under the assumption of uncorrelated replicates. Therefore, we develop our general model with the fundamental advantage 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 a particular case. Additionally, we also propose two tests for checking whether a replicate-effect indeed occurs over the experiment, thus justifying (or not) the adoption of the proposed replicate time series model.

This thesis is structured as follows: Chapter 1 gives an overview of the statistical concepts within the literature that are fundamental to our proposed methodology (developed in Chapters 2, 3 and 5). Wavelets require particular attention and we will give an overview of the basic concepts from wavelet theory, including multiresolution analysis which provides the framework around which wavelets are built, and the discrete wavelet transform (DWT). Next, we provide an overview of time series analysis in two parts. Firstly, we introduce stationary time series, a topic underpinned by constancy of the process mean and variance over time, and document classical approaches to their analysis in both the time and spectral domain. Secondly, we explore approaches in the spectral domain developed to handle nonstationary time series data. Specifically we introduce the locally stationary Fourier processes (LSF) framework of Dahlhaus (1997) and then the developments of this framework in the wavelet domain through the locally stationary wavelet (LSW) model of Nason *et al.* (2000).

Chapter 2 addresses the crucial problem of approaching replicate time series as identical process realisations and demonstrates the misleading results this approach can yield when studying the process dynamics over the replicate domain. Here, we develop a new wavelet-based *replicate-evolving locally stationary wavelet* (REv-LSW) model, capable of capturing evolutionary process characteristics across time within a single replicate and across all replicates. We propose to model the meta-process within a locally stationary wavelet (LSW) process paradigm that builds upon the framework introduced by Nason *et al.*

(2000) for a single process (here, replicate or trial). This amounts to developing novel evolutionary wavelet quantities and associated estimation theory that encompass variation across time within a replicate and across all replicates. To obtain well-behaved, consistent spectral estimates we propose to perform local smoothing of the raw wavelet periodograms across replicates, as opposed to employing the smoothing over time typically undertaken in the locally stationary process context. A fundamental assumption we make for the methodology in this chapter, is that replicates are uncorrelated as also assumed in Fiecas and Ombao (2016). Through an extensive simulation study, we demonstrate the new information captured by the REv-LSW model in comparison to the classical LSW approach that assumes replicates are identical process realisations.

In Chapter 3, we propose a general model that allows for between-replicate correlation, and the constraint of uncorrelated replicates, from Chapter 2, is treated as a particular case. The associated estimation theory additionally encompasses between-replicate variation, enabling us to estimate the replicate-coherence which is shown to provide novel, useful information about the process replicate evolution. By means of a thorough simulation study, our proposed model is shown to successfully capture between-replicate dependence, giving a good indication of the features of the replicate-coherence.

Chapter 4 applies the methodology developed in Chapters 2 and 3 to the real dataset on the macaque monkey associative learning experiment. The results obtained here highlight the major advantages of our model, which 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. These advantages have enabled us to infer evidence of learning throughout the experiment and furthermore detect between-trial dependence within the data obtained from the nucleus accumbens.

It is often of interest to perform tests of stationarity on a time series to investigate the presence of varying dynamics. Thus, in Chapter 5 we propose two novel hypothesis tests to assess whether a significant replicate-effect is manifest across the meta-process (replicate time series). We propose *location-specific* and *global* tests, both underpinned by the methodological framework developed in Chapter 2. Through both a thorough simulation study and an application to the macaque data, we demonstrate the tests successfully identify the presence of a replicate-effect and provide further information on where the replicate-effect is located across the trials, as well as across time and scale localisations. The results of testing on the macaque data corroborate the results obtained in Chapter 4, thus establishing the tests as favourable tools to be used as a first step to analyse a replicate time series.

Chapter 6 concludes with a summary of the proposed methodologies within the thesis and highlights potential avenues for future work.

Acknowledgements

My PhD journey would not have been as enjoyable or successful without the amazing supervision of Marina Knight. Marina, I cannot thank you enough for the support and inspiration you have given me, both academically and personally, over the past four years. Thank you for seeing my potential and always having my development at heart. Forget supervisor, you have been a superhero but more importantly a good friend.

I would like to thank Jon Pitchford and Ben Powell for the most chilled and informative TAP meetings a student could ask for. Jon, your thought-provoking (occasionally sneaky) questions always kept me on my toes and throughout my work I have been guided by your words to "tell the audience the story". Ben, your interesting thoughts influenced me to think about my work from different perspectives and our catch-up chats in Co-op were always appreciated. Additionally, it was great fun to assist you with teaching seminars for Stats 2. One final time, sorry for the length of my TAP presentations!

To my family and friends who have believed in me and given me motivation when times were challenging, thank you. A few words can make a big difference, even if they are followed with "so when are you going to get a job?"

Finally, special thanks to Sophie who has put up with me in constant work-mode, often at unreasonable times, such as my late night/early morning bursts of inspiration. She has celebrated my success and has shared the burden of tough times. Thank you for being there every step of the way.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. Under the supervision of Dr. Marina Knight, the literature research, developed methodology, programming and analysis are a result of my own work which has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

The work in this thesis was supported by the EPSRC studentship EP/N509802/1.

Macaque dataset

The macaque dataset used throughout this thesis as an application for the developed methodology to real data is courtesy of Prof. Hernando Ombao, at Statistics Program, King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

The following chapters have been submitted for publication, where I am the first author on both papers, but have received feedback and corrections from my co-authors.

Chapters 2, 3 and 4

The work in these chapters has been (partially) packaged as:
Embleton, J., Knight, M. I., and Ombao, H. (2021). Multiscale spectral modelling for nonstationary time series within a multiple-trial experiment. This has been invited for revised submission at Annals of Applied Statistics. A preprint version can be found at arXiv:2005.09440.

Chapter 5

The work in this chapter has been (partially) packaged as:
Embleton, J., Knight, M. I., and Ombao, H. (2021). Wavelet testing for a replicate-effect within a multiple-trial experiment. This has been submitted for publication for the special issue on Time Series at Computational Statistics and Data Analysis.

1 Literature review

1.1 Wavelet theory

Spectral analysis is broadly concerned with decomposing a function as a linear combination of basis functions. Various bases for well-behaved function spaces can be formed using, e.g. trigonometric functions, albeit, the methodological developments of this thesis rest on wavelets.

Wavelets are a major modern development on the concepts that form Fourier theory, thus it is only fitting to first review some of the relevant concepts of Fourier analysis. The earliest connection to proposing a trigonometric series (Definition 1.1.1 below) to approximate functions dates back to Daniel Bernoulli (1753) (see Larsen (1974)) in relation to a problem on the vibration of strings. New techniques underpinned by the sine and cosine trigonometric functions were investigated by Jean Baptiste Joseph Fourier, who asserted there exists a trigonometric series representation for any periodic function (see Bachman *et al.* (2000) for a historical introduction). This however was not the case, as some signals are too 'wild' to be represented efficiently by the Fourier sines and cosines, thus motivating new techniques, namely wavelets.

Our review of Fourier analysis will take structure from the descriptions given in Priestley (1982) with further reference to Bachman *et al.* (2000).

1.1.1 Fourier analysis

Fourier analysis is primarily concerned with using sine and cosine waves to form bases for $L^2(\mathbb{R})$ functions, where $f \in L^2(\mathbb{R})$, the space of *square integrable functions*. This consists of functions $f(\cdot)$ that fulfill $\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$. Therefore we start by defining the Fourier series decomposition of a function denoted by f.

Definition 1.1.1. Let f be a periodic function with period 2π (i.e. $f(x) = f(x + 2\pi)$) and square integrable over the interval $[-\pi, \pi)$. Then the *Fourier series* representation of f in trigonometric form is

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos(nx) + b_n \sin(nx)\right),$$

where a_n and b_n are the nth *Fourier coefficients*, calculated from

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx, \qquad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

We note that the Fourier series can be expressed in exponential form (e.g. Bachman *et al.* (2000)).

Each term within the Fourier series decomposition has a magnitude, given by the Fourier coefficients a_n and b_n, that exposes the frequency structure of the signal. Thus, various periodic functions can be constructed from a mixture of the Fourier functions, such that $\{\cos(nx), \sin(nx)\}_{n \in \mathbb{N}}$ forms an orthonormal basis. Convergence of the series is often overlooked in practice, for instance in the case of functions with discontinuities the Fourier

approximation can overshoot, referred to in the literature as the 'Gibbs phenomenon' (see (Bachman *et al.*, 2000, §4.8)). For periodic functions in $L^2([-\pi, \pi))$, convergence can be achieved in the mean-square sense i.e.,

$$\int_{-\pi}^{\pi} |f(x) - f_n(x)|^2 dx \to 0, \quad \text{as } n \to \infty,$$

where $f_n(x)$ is the nth partial sum. Additionally, if a 2π-periodic function f is absolutely integrable over $[-\pi, \pi)$, and has a finite number of extrema and finite number of discontinuities in the interval $[-\pi, \pi)$, then it is said to satisfy the *Dirichlet* conditions and its Fourier series converges point-wise. In the presence of a finite discontinuity, x_d, the Fourier series converges to $\frac{1}{2}\lim_{\ell \to 0}[f(x_d - \ell) + f(x_d + \ell)]$.

Generally, most functions are not periodic. Suppose a function f is non-periodic and defined for all real x. We can construct a Fourier series representation for $f \in L^1(\mathbb{R})$, the space of *absolutely integrable functions* (i.e. $\int_{-\infty}^{\infty} |f(x)|dx < \infty$), through application of the following transform.

Definition 1.1.2. For a function $f \in L^1(\mathbb{R})$, the *Fourier transform* of f is given by

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx,$$

where $\hat{f} \in L^1(\mathbb{R})$ and ω is the angular frequency. We seek to recover f through the *inverse Fourier transform* of $\hat{f}$, given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(\omega) e^{i\omega x} d\omega. \tag{1}$$

We note here that the angular frequency (measured in radians per second) and the Hertz frequency (denoted here by v) have the relationship $\omega = 2\pi v$. When particular conditions for convergence are satisfied, equation (1) gives a representation for f by means of a 'sum' of sine and cosine functions. Note that in contrast with the Fourier representation for periodic functions which involved a discrete range of frequencies, the integral representation here provides access to frequencies on a continuous scale.

In practice, many statistical applications are concerned with discrete-time series. Although theoretically the Fourier transform can be applied to discrete-time signals (for example the discrete-time Fourier transform), the integral may be unrealistic to evaluate. By sampling the frequencies on the continuous scale as to obtain a discrete range of frequencies, the signal f can be recovered by means of inversion of the discrete Fourier transform.

Definition 1.1.3. For a signal f defined on the finite set $\{0, 1, ..., N-1\}$, such that $f(t) = f(t + kN)$, the *discrete Fourier transform* of f is given by

$$D\hat{f}_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} f(t) e^{-2\pi i t j / N},$$

for $j = 0, 1, ..., N-1$, where j/N is the jth multiple of the fundamental frequency.

Here, the $1/\sqrt{N}$ factor is chosen for symmetry with the inverse transform, the same

reason for choosing $1/\sqrt{2\pi}$ in the Fourier transform.

As already discussed, Fourier representations for series with sharp changes may require many combinations of the sine and cosine basis functions in order to approximate at 'jump' points. However, a sparse representation, meaning few non-zero basis coefficients, can lead to a better understanding of the signal structure. Since Fourier functions are only localised in frequency and not in time, their applicability to represent signals is better suited for smoother signals which exhibit periodicity, than for signals with a more 'turbulent' behaviour with features manifest locally. Thus, basis functions that can capture local features would be desirable. Wavelets can do exactly that; many are constructed to have compact support which gives them localisation in time and we introduce them next.

1.1.2 Wavelets

Wavelets, as their name suggests can be thought of as 'little wave' functions, which possess the ability to capture local characteristics due to their localisation in time as well as frequency. The addition of locality in time, gives wavelets a natural advantage over the Fourier functions previously mentioned, to represent discontinuous functions. Motivated by the monographs of Vidakovic (1999), Daubechies (1992) and Nason (2008), we now proceed to introduce some theory on wavelets.

To begin, meet the *mother wavelet*, $\psi \in L^2(\mathbb{R})$, a square integrable function capable of generating a family of functions that can form bases for various spaces of functions. For dilation and translation parameters, $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$, respectively, a *family of wavelet functions* can be defined as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right), \tag{2}$$

where the normalisation parameter $|a|^{-1/2}$, ensures the L^2-norm of ψ is maintained: $||\psi_{a,b}||_2 = ||\psi||_2$. We understand this to mean that wavelet families are constructed through re-scales and translations of the mother wavelet ψ, where a and b are allowed to vary continuously.

Wavelets are attractive due to their localisation in both frequency and time. These properties are ensured through the following condition that the wavelet function is assumed to satisfy.

Definition 1.1.4. A wavelet, $\psi \in L^2(\mathbb{R})$, is any square integrable function which satisfies the *admissibility condition*

$$C_\psi = \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{3}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(x)$ (see Definition 1.1.2). The admissibility condition implies (Vidakovic, 1999) that at the zero frequency $\omega = 0$

$$\hat{\psi}(0) = \int_{-\infty}^{\infty} \psi(x)dx = 0,$$

which ensures the wavelet oscillates.

Daubechies (1992) characterised wavelets to have compact support, ensuring a fast decay to zero and helping to provide time localisation.

An important property of wavelet functions, $\psi \in L^2(\mathbb{R})$, is that they can be constructed with different degrees of smoothness. At one extreme is the discontinuous Haar wavelet function and at the other is the Shannon wavelet function. If ψ satisfies

$$\int_{-\infty}^{\infty} x^{\ell}\psi(x)dx = 0, \tag{4}$$

for $\ell = 0, ..., n-1$, then it is said to have n *vanishing moments*. Thus the wavelet coefficients are zero for any polynomial of degree $(n-1)$ or less. This property is advantageous as it allows for a sparse representation of a function, such that for a reasonably smooth function only few coefficients need to be estimated. The more vanishing moments ψ has, the smoother the function it will be.

Let us gain a feel for the characteristics of wavelets through introducing the *Haar wavelet* (Haar, 1910), often regarded as the simplest example of a wavelet system. The Haar mother wavelet is defined by

$$\psi^H(x) = \begin{cases} 1, & \text{for } x \in [0, \frac{1}{2}) \\ -1, & \text{for } x \in [\frac{1}{2}, 1] \\ 0, & \text{otherwise.} \end{cases}$$

Through the dilations and translations of the Haar mother wavelet using equation (2), the associated wavelet functions are given by

$$\psi^H_{a,b}(x) = \begin{cases} \frac{1}{\sqrt{a}}, & \text{for } x \in [b, \frac{a}{2} + b) \\ \frac{-1}{\sqrt{a}}, & \text{for } x \in [\frac{a}{2} + b, a + b] \\ 0, & \text{otherwise,} \end{cases}$$

where $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$. The Haar mother wavelet, $\psi^H(x) = \psi^H_{1,0}(x)$, is displayed in Figure 1 alongside an associated translated and dilated Haar wavelet.

Examples of wavelets

Daubechies' compactly supported wavelets

Daubechies wavelets are an important historical step in the development of wavelet theory. Daubechies (1992) discovered that a basis for $L^2(\mathbb{R})$ can be formed from a family of orthonormal wavelets with compact support and varying degrees of smoothness. Smoothness is determined by the number of vanishing moments (recall equation (4)) and each member of a Daubechies wavelet family is indexed by an associated number of vanishing moments, N, and will have support of size larger than or equal to $2N-1$ with minimum compact support on the interval $[-N+1, N]$ (Mallat, 2009). Note, some authors index Daubechies wavelets by the length of the compact support. The Daubechies Extremal Phase wavelet family is constructed to have the minimal support possible for N vanishing moments. Haar wavelets are included in this family, indexed by $N = 1$ with support on the interval $[0, 1]$. The Daubechies Least Asymmetric wavelet family are designed to minimise

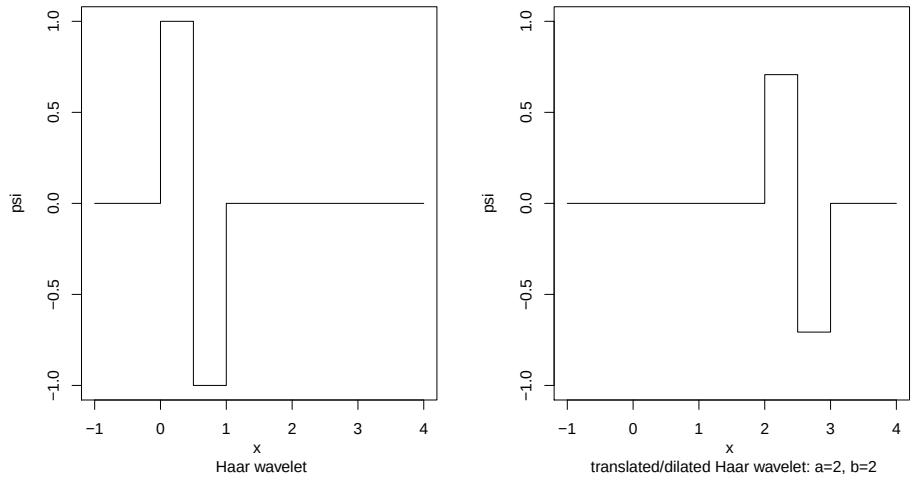


Figure 1: *Left*: Haar mother wavelet $\psi^H_{1,0}(x)$; *Right*: translated/dilated Haar mother wavelet $\psi^H_{2,2}(x)$.

the degree of asymmetry within the wavelet. For an exhaustive review of Daubechies wavelets see Daubechies (1992), alternatively see (Percival and Walden, 2000, §4.8) for a discussion the construction of Daubechies wavelet filters. Examples of Daubechies Extremal Phase and Least Asymmetric wavelets for a selection of vanishing moments, N, are displayed in Figure 2.

Shannon wavelet

We have seen that Haar wavelets are compactly supported on $[0, 1]$ and are localised in time. Shannon wavelets are the time-scale mirror image, such that they are localised in frequency and have compact support on $[-2\pi, -\pi) \cup (\pi, 2\pi]$. The Shannon mother wavelet, as defined in (Chui, 1997, §4.2), is given to be

$$\psi(x) = \frac{\sin(2\pi x) - \cos(\pi x)}{\pi(x - 1/2)},$$

with support over $\mathbb{R}$ and its Fourier transform is given by

$$\hat{\psi}(\omega) = -e^{i\omega/2}\mathbb{I}_{[-2\pi,-\pi)\cup(\pi,2\pi]}(\omega),$$

where $\mathbb{I}$ denotes the indicator function. With decay like $|x|^{-1}$ in the time domain, the Shannon wavelet is rarely used in practice but could aid theoretical understanding of wavelets with a very high number of vanishing moments. Chui (1997) remarks that all high-order orthonormal wavelets 'imitate' the Shannon wavelet. Thus, Shannon wavelets can be thought of as the limiting wavelet for Daubechies wavelets of order N as $N \to \infty$.

The continuous wavelet transform (CWT)

Statistical applications using wavelet analysis are mainly interested in discrete transformations however it is only natural to introduce wavelet theory by starting with the continuous wavelet transformation.

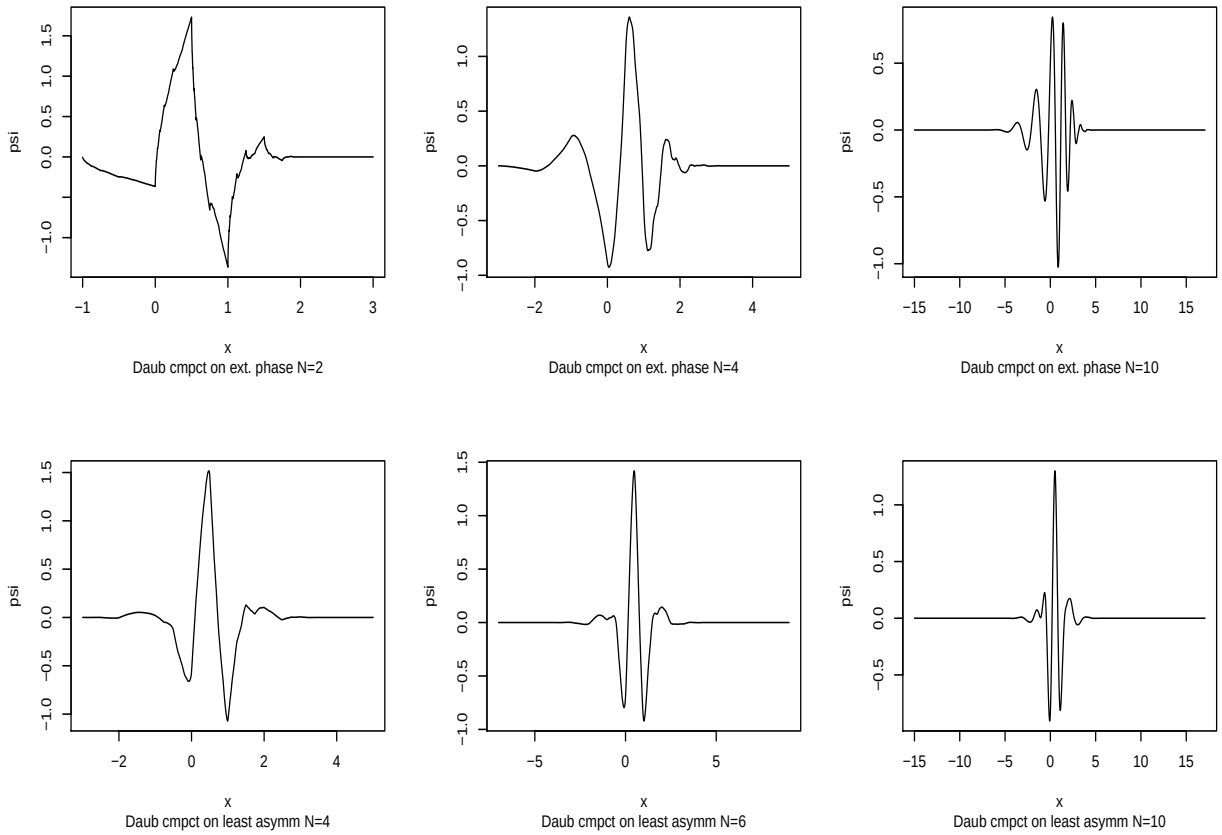


Figure 2: Daubechies wavelets of varying degrees of vanishing moments N. *Top*: Extremal Phase; *Bottom*: Least Asymmetric. Plotted using `Wavethresh` (Nason, 2008).

Definition 1.1.5. For a function $f \in L^2(\mathbb{R})$, the *continuous wavelet transform* (CWT) is defined as

$$CWT_f(a,b) = \langle f, \psi_{a,b} \rangle = \int f(x)\overline{\psi_{a,b}(x)}dx, \tag{5}$$

where $\overline{\psi_{a,b}(x)}$ is the complex conjugate of $\psi_{a,b}(x)$ and $a \in \mathbb{R}\backslash\{0\}$, $b \in \mathbb{R}$. Assuming the admissibility condition (equation (3)) is satisfied, we can find the inverse transform through the *resolution of identity* relation (Vidakovic, 1999)

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} CWT_f(a,b)\psi_{a,b}(x)da\ db, \tag{6}$$

thus providing a way to recover f by means of superimposing wavelet basis functions.

In comparison with the Fourier transform representation given in Definition 1.1.2, it can be seen that the coefficients, $CWT_f(a,b)$, in the wavelet decomposition given in equation (5) have the advantage of providing information on the amplitude at a given location (through b) as well as a given scale (through a). The coefficients in the Fourier transform can only provide information about the amplitude at a given frequency (equivalent to wavelet scale). The Haar wavelet is one of many that can be used here, for further examples, such as the 'mexican hat' wavelet, see (Vidakovic, 1999, §3.1.2). A comprehensive review of CWT can be found in Daubechies (1992).

Discretisation of the CWT

The work in this thesis mainly focuses on constructions of discrete wavelet representations. One drawback of the continuous wavelet transform is redundancy, such that the information captured in the coefficient $CWT_f(a, b)$ is also contained within neighbouring coefficients. To answer this issue, sampling the values of a and b can yield less redundant transformations. Discretisation of the continuous wavelet transform involves selecting discrete values of a and b such that we still have an invertible transformation. Most often, the choices of dyadic translations $a = 2^{-j}$ and $b = 2^{-j}k$ (as noted by Antoniadis and Gijbels (2002)) are adopted, known as *critical sampling*. This is the coarsest choice that will admit a function to be uniquely recovered from its wavelet transform. Plugging this sampling choice into equation (2), a *family of discrete (decimated) wavelet functions* can be defined as

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}, \tag{7}$$

where j and k are the resolution level (also referred to as scale) and location respectively. Note that in 'j, k' notation the mother wavelet, ψ, is equivalent to $\psi_{0,0}$ and this is the form we refer to from now on. For suitable choices of ψ, a and b, the above wavelet family can form an orthonormal basis for $f \in L^2(\mathbb{R})$ (see Daubechies (1988, 1992)). Technically, information about the function being decomposed will be contained within each scale 2^j at locations $2^{-j}k$. Through finer choices of sampling, for instance greater values of j, the transformation becomes more redundant, whereas too coarse sampling will lead to not being able to recover f uniquely. Antoniadis and Gijbels (2002) introduce the choice of $a = 2^{-j}$ and $b = k$ which leads to the construction of the *non-decimated wavelet family*

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j(x - k)), \quad j, k \in \mathbb{Z}. \tag{8}$$

Unlike the wavelet family in (7) that exists across dyadic intervals, here a wavelet exists at all integer locations. Note, this highlights the time translation invariance of transforms using wavelets from (8) and lack of translation invariance from wavelets in (7). More on this will be delved into later.

In similar fashion to the inverse continuous transform (equation (6)), any function, $f \in L^2(\mathbb{R})$, can be recovered by means of a linear combination of wavelet basis functions at different scales and locations. Specifically, we define the following representation for f.

Definition 1.1.6. For suitable choices of $\psi \in L^2(\mathbb{R})$, such that the decimated wavelet family $\{\psi_{j,k}(x)\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$, the function $f \in L^2(\mathbb{R})$ has a *wavelet representation* given by

$$f(x) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\langle f, \psi_{j,k}\rangle\psi_{j,k}(x),$$

where $\langle\cdot,\cdot\rangle$ is the L^2-inner product and due to the orthogonality of wavelets,

$$d_{j,k} = \langle f, \psi_{j,k}\rangle = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx.$$

The $\{d_{j,k}\}_{j,k\in\mathbb{Z}}$ are called the *wavelet coefficients* within a scale j for a location k. As mentioned previously for the continuous wavelet transform, the wavelet coefficients provide

information on the amplitude at a given location (through k) and scale (through j), unlike the Fourier coefficients that can only provide the latter (for frequency).

1.1.3 Multiresolution analysis (MRA)

Let us now take a slight detour away from wavelets explicitly to introduce the framework upon which wavelets are built, multiresolution analysis, introduced by Mallat (1989a,b) and Meyer (1993). We can think of a multiresolution analysis (MRA) as if painting a portrait of a function. We start with a low-resolution approximation to a function $f \in L^2(\mathbb{R})$, then by 'zooming in' on our subject (f) we can progressively add finer details from higher resolution approximations. Thus our portrait (representation) of $f(x)$ is much more detailed. Comprehensive accounts of MRA can be found in Mallat (1989a,b) and Daubechies (1988, 1992). Next, we introduce the important characteristics of MRA and the part it plays in constructing wavelets. Our discussion follows the work presented in Vidakovic (1999), Debnath (2002) and Nason (2008).

Definition 1.1.7. A *multiresolution analysis* of $L^2(\mathbb{R})$ consists of a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$,

$$... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ...,$$

thought of as successive approximation spaces, such that as j gets larger the space V_j contains functions of finer scales. The following conditions must be satisfied:

1. The spaces have trivial intersection and their union is dense in $L^2(\mathbb{R})$

$$\bigcap_{j\in\mathbb{Z}} V_j = \{\mathbf{0}\}, \quad \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}),$$

   where the overline denotes closure of the set and $\mathbf{0}$ denotes the zero function.

2. The spaces have the following interscale linkage relations

$$\begin{aligned} &f(x) \in V_j \Leftrightarrow f(2^m x) \in V_{j+m}, \quad \forall x \in \mathbb{R}, m \in \mathbb{Z}, \\ &f(x) \in V_0 \Leftrightarrow f(x-k) \in V_0, \quad \forall x \in \mathbb{R}, k \in \mathbb{Z}. \end{aligned} \tag{9}$$

3. There exists a *scaling function* $\phi \in V_0$ with $\int_{-\infty}^{\infty} \phi(x)dx = 1$, for which the set of integer translates $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ form an orthonormal basis of V_0.

Condition 2 demonstrates how successive spaces contain finer details, for instance the detail added to V_{j+1} is twice as fine as the detail in V_j. Additionally, any shift in $f(x)$ by an amount k does not change the level of resolution j that we are adding to.

By equation (9) of condition 2 along with condition 3, for a fixed j, the family $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ forms an orthonormal basis of V_j, where

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \quad j, k \in \mathbb{Z}.$$

This prompts us to define the *scaling function* as

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x-k), \quad h_k, k \in \mathbb{Z}, \tag{10}$$

which tells us that we can represent any function $\phi(x) \in V_0$ as a linear combination of functions from V_1, due to $V_0 \subset V_1$. The set of coefficients $\{h_k\}$ is referred to as a *low-pass filter*, for reasons that will soon become clear. This relation is also referred to as the *dilation equation*, and it is fundamental in the construction of wavelets and building a general MRA.

Daubechies (1988) introduced the operator P_j that projects a function into the space V_j. As such, for a resolution level j, an approximation for a function f, can be given by

$$P_j f = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j,k}\rangle \phi_{j,k}(x), \tag{11}$$

since $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ is a basis for V_j. As j increases, our approximation improves and P_j converges to f as $j \to \infty$ Daubechies (1992). Next, we will see how wavelets can 'explain' the detail that is lost when moving from an approximation space, V_{j+1}, to the next coarser space, V_j.

Constructing wavelets

The principle of multiresolution analysis (Daubechies, 1992) is that for a sequence of closed subspaces satisfying a MRA, there exists a family $\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k); j, k \in \mathbb{Z}\}$ that forms an orthonormal wavelet basis of $L^2(\mathbb{R})$, such that

$$P_{j+1} f = P_j f + \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x), \tag{12}$$

for all $f \in L^2(\mathbb{R})$, where P_j is the projection operator that projects a function into the space V_j and $\langle \cdot, \cdot \rangle$ denotes the L^2-inner product. Recall that $\psi_{j,k}$ is a translated and dilated member of the wavelet family derived from the mother wavelet $\psi = \psi_{0,0}$. It can be seen from equation (12) that the difference between projections $(P_{j+1} - P_j)f$ is a linear combination of wavelets. Thus, the detail that is lost from moving to a coarser projection, P_{j+1} to P_j, can be characterised by the orthonormal basis of wavelets $\{\psi_{j,k}(x)\}_{k\in\mathbb{Z}}$. To demonstrate this, define the detail space W_j to be the orthogonal complement of V_j in V_{j+1} such that

$$V_{j+1} = V_j \oplus W_j, \quad \forall j \in \mathbb{Z}, \tag{13}$$

where $\oplus$ denotes the direct sum. Successive application of equation (13) then gives that

$$V_{j+1} = V_0 \oplus \bigoplus_{i=0}^{j} W_i.$$

Then due to condition 1 of a MRA (Definition 1.1.7), as $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$ we obtain

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j.$$

This tells us that the basis functions that characterise the detail spaces $\{W_j\}_{j\in\mathbb{Z}}$, can provide a representation for an L^2-function. Additionally, the scaling properties given in condition 2 are inherited by the detail spaces, such that

$$f(x) \in W_j \Leftrightarrow f(2^m x) \in W_{j+m} \quad \text{and} \quad f(x) \in W_0 \Leftrightarrow f(x-k) \in W_0.$$

Putting this construction together, if we can form an orthonormal basis for W_0 through integer translations of a function $\psi(x)$, then through further translations and dyadic dilations, an orthonormal basis for W_j can be formed from $\{\psi_{j,k}(x)\}_{k\in\mathbb{Z}}$. Linking back to the discussion on equation (12), the detail that is lost when moving from P_{j+1} to P_j is characterised in W_j. So intuitively, we can think of the next finest approximation of a function f, given by $P_{j+1}f$ to be the detail already described by V_j 'plus' the finer details characterised by W_j. This concept is illustrated in Figure 3. Starting at V_5, as we 'add' the detail captured within successive resolution levels $j = 5, ..., 9$, finer approximations of f are obtained. Thus, it can be seen that the family $\{\psi_{j,k}(x); j, k \in \mathbb{Z}\}$ forms a basis of $L^2(\mathbb{R})$.

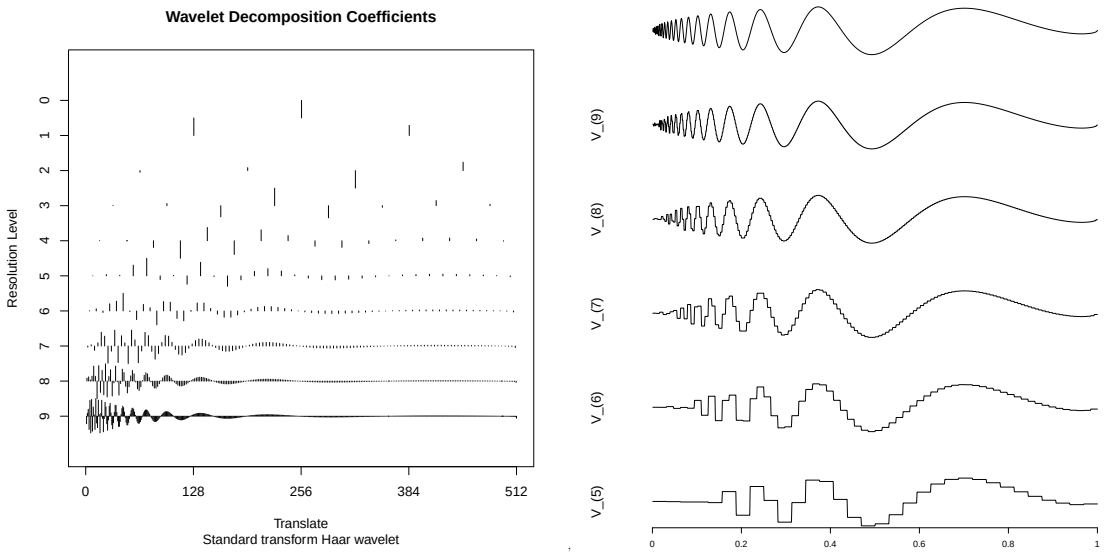


Figure 3: Multiresolution analysis of the Doppler function (topmost signal in right plot). *Left*: Wavelet coefficients characterising the detail contained in W_j; *Right*: Successive approximations of the Doppler function projected into the spaces V_j where $V_{j+1} = V_j \oplus W_j$. This figure is reproduced following Figure 3.7 in Vidakovic (1999).

As $W_0 \subset V_1$, then the wavelet function $\psi(x) \in V_1$, which can be constructed from the scaling function by

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \sqrt{2}\phi(2x-k), \quad g_k, k \in \mathbb{Z}, \tag{14}$$

where the set of coefficients $\{g_k\}$ is called a *high-pass filter*, for reasons that will soon become clear. Daubechies (1992) showed that a possible solution to equation (14), that yields a desirable mother wavelet is to set

$$g_k = (-1)^k h_{1-k},$$

which is known as the *quadrature mirror relation*. Recall that $\{h_k\}$ is the low-pass filter, that suppresses the higher frequencies whilst preserving the low ones, and with opposite

characteristics is the high-pass filter, $\{g_k\}$. Under the above relation, these coefficients are referred to as the *quadrature mirror filters*. For an extensive proof of this relation, readers can consult either Vidakovic (1999) or Daubechies (1992).

1.1.4 The discrete wavelet transform (DWT)

In a manner similar to discrete Fourier transforms, *discrete wavelet transforms* (DWT) are applied to discrete sets of data, returning outputs that are also discrete. Mallat (1989a,b) was first to connect multiresolution analysis and cascade algorithms with wavelets, in order to provide a nested structure to recursively compute the scaling and wavelet (detail) coefficients, this scheme is referred to as *Mallat's Pyramid Algorithm*. Essentially, the scheme uses the low- and high-pass filters, $\{h_k\}$ and $\{g_k\}$, associated with the scaling equations (10) and (14), to provide a relation between the wavelet coefficients at different levels in the transformation, ergo, bypassing the need to evaluate the inner products $\langle f, \phi_{j,k} \rangle$ and $\langle f, \psi_{j,k} \rangle$ in equation (12).

Recalling that $\phi = \phi_{0,0}$ and $\psi = \psi_{0,0}$, through the substitution of indices in the scaling equations (10) and (14), the following *refinement relations* can be obtained

$$\phi_{j-1,k}(x) = \sum_{l \in \mathbb{Z}} h_{l-2k} \phi_{j,l}(x), \quad \text{and} \quad \psi_{j-1,k}(x) = \sum_{l \in \mathbb{Z}} g_{l-2k} \phi_{j,l}(x). \tag{15}$$

As customary in the wavelet literature, we denote the scaling coefficients and wavelet coefficients to be, $c_{j,k} = \langle f, \phi_{j,k} \rangle$ and $d_{j,k} = \langle f, \psi_{j,k} \rangle$ respectively, then using the refinement relations (15) allows us to construct a recursive relationship structure between coefficients at different levels:

$$\begin{aligned} c_{j-1,k} &= \langle f, \phi_{j-1,k} \rangle \\ &= \left\langle f, \sum_{l \in \mathbb{Z}} h_{l-2k} \phi_{j,l} \right\rangle \\ &= \sum_{l \in \mathbb{Z}} h_{l-2k} \langle f, \phi_{j,l} \rangle \\ &= \sum_{l \in \mathbb{Z}} h_{l-2k} c_{j,l}. \end{aligned} \tag{16}$$

In the same manner, for the wavelet coefficients we obtain

$$d_{j-1,k} = \sum_{l \in \mathbb{Z}} g_{l-2k} c_{j,l}. \tag{17}$$

Formulas (16) and (17) are the mechanics referred to when computing the DWT. For each level j, they provide a recursive way to compute the scaling and wavelet coefficient vectors, $\mathbf{c_j} = \{c_{j,k}\}_{k \in \mathbb{Z}}$ and $\mathbf{d_j} = \{d_{j,k}\}_{k \in \mathbb{Z}}$, at successive coarser levels. Thus, linking this with the level j multiresolution decomposition of a function $f \in L^2(\mathbb{R})$ (see equations (11) and (12)), it is possible to represent a function in $L^2(\mathbb{R})$ by

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x),$$

where j_0 denotes the coarse level. The first part of the above sum is equivalent to $P_{j_0} f$ and captures the 'smooth' underlying global characteristics of f at level j_0. The second part of the sum represents the finer details of f that are lost when moving to coarser scales.

As with the continuous case, the DWT described by formulas (16) and (17) which is also referred to as the forward transform, has an inverse. The inverse allows us to recursively reverse the direction of the cascade algorithm such that we can obtain the next finer level coefficients starting at the coarsest level scaling and mother coefficients. The *inverse discrete wavelet transform* (IDWT) is given by

$$c_{j,l} = \sum_{k \in \mathbb{Z}} c_{j-1,k} h_{l-2k} + \sum_{k \in \mathbb{Z}} d_{j-1,k} g_{l-2k}, \tag{18}$$

where $\{h_l\}$ and $\{g_l\}$ are the quadrature mirror filters that share the same structure as the filters computed for the forward transform.

DWT in the presence of discrete data

In most practical situations, a function f is a collection of discrete values observed at $N = 2^J$ equally spaced locations $\{x_i, 0 = 1, ..., N-1\}$. The DWT is initiated from the 'finest scale' scaling coefficient vector $\mathbf{c_J}$, however as we are operating on discrete values of f we cannot directly calculate the scaling coefficients at the finest scale. Thus, we need a way to approximate $\mathbf{c_J}$ (see (Nason, 2008, §2.7.3)). An often convenient approach is to approximate the finest scale scaling coefficients by the original function samples, such that

$$c_{J,i} = f(x_i), \quad \text{for } i = 0, ..., N-1.$$

However, this approach can incur some error and is referred to as committing the 'wavelet crime' by Strang and Nguyen (1996); the authors suggest to first pre-filter the original data samples. Thus, an approximation of the observed function f can be constructed at the finest level J, by

$$\tilde{f}(x) = \sum_{k \in \mathbb{Z}} c_{J,k} \phi_{J,k}(x).$$

It is then possible to construct the coarser scaling and wavelet coefficients for $j < J$. Note that as a consequence, although we will maintain the notation $c_{j,k}$ and $d_{j,k}$, the coefficients of $\tilde{f}$ will be approximations of $c_{j,k} = \langle f, \phi_{j,k} \rangle$ and $d_{j,k} = \langle f, \psi_{j,k} \rangle$. On a further note, the empirical coefficients are approximately proportional to their continuous counterparts by a factor $1/\sqrt{N}$ (see Abramovich *et al.* (2000)).

Having chosen a particular wavelet family, and starting with the finest scaling coefficients in $\mathbf{c_J}$, the DWT progresses by applying the formulas (16) and (17) to obtain the next coarser level coefficients, $\mathbf{c_{J-1}} = \{c_{J-1,k}\}_{k \in \mathbb{Z}}$ and $\mathbf{d_{J-1}} = \{d_{J-1,k}\}_{k \in \mathbb{Z}}$. The length of each coefficient vector at scale $J-1$ is equal to half the length of the data N, so after one application $N/2$ scaling and wavelet coefficients are obtained. This halving continues with each application to the next coarsest scale and is a result of the decimation step in the DWT captured through the '$2k$' term in formulas (16) and (17). The next step then uses the output $\mathbf{c_{J-1}}$ as the input for the next application of the algorithm and this process repeats until the desired coarse level decomposition, denoted j_0, has been reached. A visual interpretation of this algorithm is given in Figure 4. The result of the DWT is therefore a

transformation of an initial vector in the time domain to the following vector in the wavelet domain:

$$DWT(f) = (\mathbf{c_{j_0}}, \mathbf{d_{j_0}}, ..., \mathbf{d_{J-2}}, \mathbf{d_{J-1}}),$$

where j_0 denotes the desired coarse level. Note, if j_0 is set to equal 0 so that a full decomposition of the series is constructed, then $\mathbf{c_0} = c_{0,0}$ and $\mathbf{d_0} = d_{0,0}$.

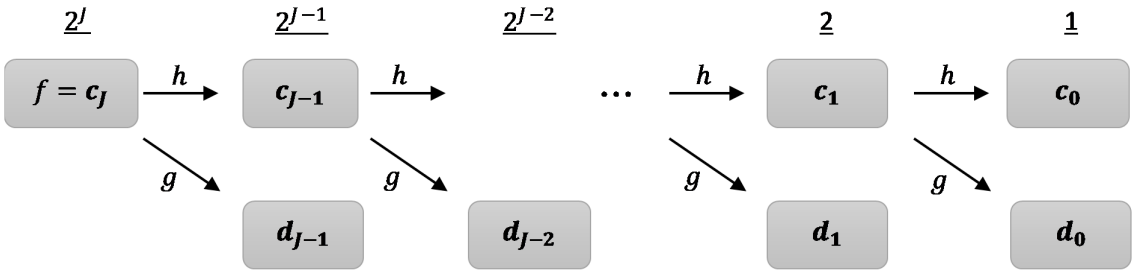


Figure 4: Flow diagram depicting the DWT on the finest scale scaling coefficients $\mathbf{c_J} = f$. The high- and low-pass filters are denoted by g and h respectively. Each step in the decomposition is headed by the number of coefficients produced for each of the scaling and wavelet coefficients.

A 'Haar'-tening example of the DWT

To demonstrate the DWT in practice, we present here a simple yet informative example that uses the Haar wavelet basis to form a wavelet decomposition for discrete data. The low- and high-pass filters for the Haar wavelet are given by

$$h_k = \begin{cases} \frac{1}{\sqrt{2}}, & \text{for } k = 0 \\ \frac{1}{\sqrt{2}}, & \text{for } k = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad g_k = \begin{cases} \frac{1}{\sqrt{2}}, & \text{for } k = 0 \\ \frac{-1}{\sqrt{2}}, & \text{for } k = 1 \\ 0, & \text{otherwise} \end{cases}. \tag{19}$$

For the discrete data sequence $\mathbf{y} = (y_0, ..., y_{N-1})$, we start the DWT by first approximating the finest scale scaling coefficients by the original data such that

$$\mathbf{c_J} = (5, 3, 1, 9, 2, 2, 8, 4) = \mathbf{y}.$$

As $\mathbf{y}$ is of length $N = 8 = 2^J$, the finest level is $J = 3$ and thus $\mathbf{c_J} = \mathbf{c_3}$. Using formulas (16) and (17), we can obtain the next coarsest level scaling coefficients $\mathbf{c_2} = (c_{2,0}, c_{2,1}, c_{2,2}, c_{2,3})$ and wavelet (detail) coefficients $\mathbf{d_2} = (d_{2,0}, d_{2,1}, d_{2,2}, d_{2,3})$ in the following manner:

$$\begin{aligned} c_{2,0} &= \sum_l h_l c_{3,l} &&= \frac{1}{\sqrt{2}} \cdot 5 + \frac{1}{\sqrt{2}} \cdot 3 = 4\sqrt{2} \\ c_{2,1} &= \sum_l h_{l-2} c_{3,l} &&= \frac{1}{\sqrt{2}} \cdot 1 + \frac{1}{\sqrt{2}} \cdot 9 = 5\sqrt{2} \\ c_{2,2} &= \sum_l h_{l-4} c_{3,l} &&= \frac{1}{\sqrt{2}} \cdot 2 + \frac{1}{\sqrt{2}} \cdot 2 = 2\sqrt{2} \\ c_{2,3} &= \sum_l h_{l-6} c_{3,l} &&= \frac{1}{\sqrt{2}} \cdot 8 + \frac{1}{\sqrt{2}} \cdot 4 = 6\sqrt{2} \end{aligned}$$

$$
\begin{aligned}
d_{2,0} &= \sum_l g_l c_{3,l} \quad = \frac{1}{\sqrt{2}} \cdot 5 - \frac{1}{\sqrt{2}} \cdot 3 = \sqrt{2} \\
d_{2,1} &= \sum_l g_{l-2} c_{3,l} = \frac{1}{\sqrt{2}} \cdot 1 - \frac{1}{\sqrt{2}} \cdot 9 = -4\sqrt{2} \\
d_{2,2} &= \sum_l g_{l-4} c_{3,l} = \frac{1}{\sqrt{2}} \cdot 2 - \frac{1}{\sqrt{2}} \cdot 2 = 0 \\
d_{2,3} &= \sum_l g_{l-6} c_{3,l} = \frac{1}{\sqrt{2}} \cdot 8 - \frac{1}{\sqrt{2}} \cdot 4 = 2\sqrt{2}.
\end{aligned}
$$

If we then continue to repeat this process to find the coefficients for $(\mathbf{c_1}, \mathbf{d_1})$ and lastly $(\mathbf{c_0}, \mathbf{d_0})$, we obtain the following output for the DWT:

$$DWT(\mathbf{y}) = (\frac{17}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -1, -4, \sqrt{2}, -4\sqrt{2}, 0, 2\sqrt{2}).$$

Figure 5 gives a visual representation of our example, where it can be seen that each successive step to the next coarsest level results in obtaining half the number of coefficients. To recover the original data sequence $\mathbf{y}$, we can perform the inverse DWT given in equation (18) on the set of wavelet coefficients obtained for $DWT(\mathbf{y})$ and the coarsest level scaling coefficient, $c_{0,0}$, which provides the coarsest characterisation of the original signal.

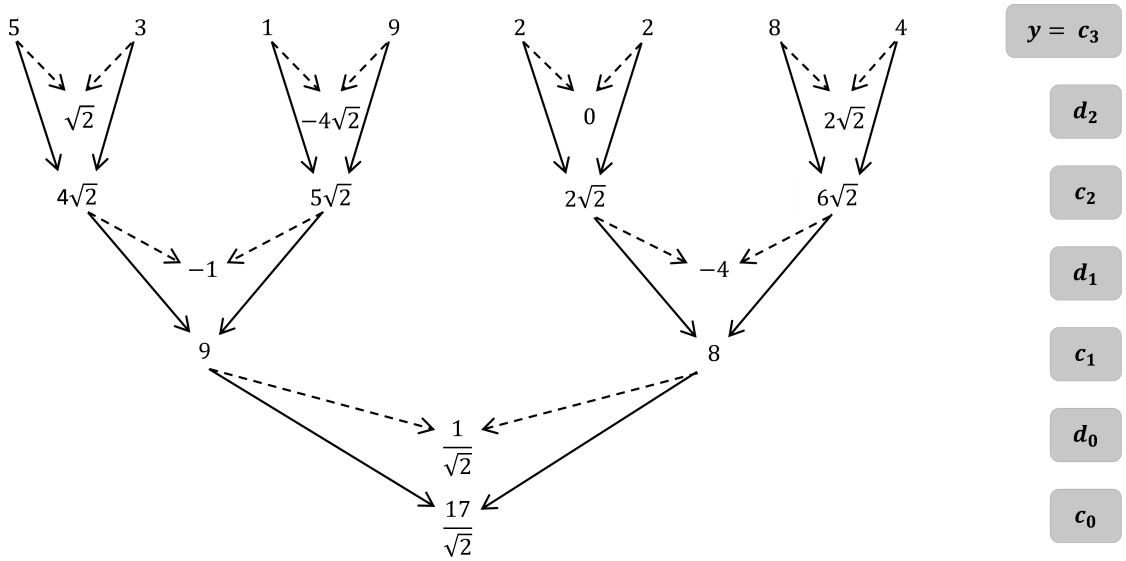


Figure 5: Graphical representation of the DWT for the Haar wavelet example. *Solid arrows*: represent application of the low-pass filter $\mathcal{H}$; *Dotted arrows*: represent application of the high-pass filter $\mathcal{G}$. This figure follows Figure 2.2 in Nason (2008).

A note on matrix representation

An alternative way to formulate the DWT, although not as computationally fast as the pyramid algorithm (Nason, 2008), is through the following matrix multiplication

$$DWT(\mathbf{y}) = W\mathbf{y},$$

where W is the orthogonal matrix (i.e. $W^TW = I_N$, where I_N is the identity matrix of order $N = 2^J$) associated with the wavelet used in the DWT process. The orthogonality of W follows from the orthonormality of the wavelet bases of the spaces V_j and W_j described

in Section 1.1.3. Since at each step of the DWT we are representing the input signal on two different bases, matrix multiplication can represent the change of basis as we move to the next coarser level (space V_{j-1}). Additionally since W is orthogonal, it follows that

$$||\mathbf{d}||^2 = \mathbf{d}^T\mathbf{d} = (W\mathbf{y})^T W\mathbf{y} = \mathbf{y}^T(W^TW)\mathbf{y} = \mathbf{y}^T\mathbf{y} = ||\mathbf{y}||^2,$$

where we have let $\mathbf{d} = DWT(\mathbf{y})$ for brevity. Thus, the length of the input vector $\mathbf{d}$ is the same as that of the input $\mathbf{y}$, which we have seen is true through our application of the pyramid algorithm. It also follows that the inverse DWT can be constructed through the matrix multiplication given by

$$\mathbf{y} = W^T DWT(\mathbf{y}),$$

noting that since W is orthogonal, $W^{-1} = W^T$.

A note on boundary conditions

Haar wavelets are constructed on the smallest support, such that the Haar filters work in pairs to transform one dyadic sequence to another. Thus we see no issues when approaching the 'boundaries' of the sequence. However, issues occur when the support of the wavelet filter extends beyond the length of the sequence. For example, consider we have four detail filters, $\{g_k\}_{k=0}^3$, associated with a wavelet (e.g. Daubechies D4 compactly supported wavelets) and we wish to transform a dyadic vector $\mathbf{y} = (y_0, ..., y_{N-1})$. The transform begins by filtering the first four elements of $\mathbf{y}$ with the corresponding filter coefficients and then skips to the right by two elements. This process continues for the remainder of $\mathbf{y}$ and can be viewed as a moving window of operations on four elements, skipping two elements at a time in order to account for dyadic decimation. However, the resulting transformed sequence is one element short of being dyadic. This may only be one element of detail missing but the issue becomes more serious for wavelets with greater support. Various approaches have been considered to tackle this issue, with a general idea being to extend the boundary, i.e., to obtain elements y_{-2} and y_{-1} in the example just given. Periodic bounding and reflection at the boundary are options discussed by Nason and Silverman (1994). A further option by Cohen *et al.* (1993), constructs coarser scale coefficients without borrowing detail from periodisation or reflection, this procedure is known as 'wavelets on the interval'.

1.1.5 The non-decimated wavelet transform (NDWT)

A defining feature of the DWT discussed in the previous section was that the transform to the next coarser scale resulted in a dyadic decimation of the number of scaling and wavelet coefficients. For instance, in our Haar wavelet example of the DWT on the series $\mathbf{y} = (y_0, ..., y_7)$, the wavelet (detail) coefficients for level $j = 2$ $(= \log_2 8 - 1)$ were calculated by

$$d_{2,k} = \frac{y_{2k} - y_{2k+1}}{\sqrt{2}}, \quad \text{for } k = 0, ..., 3.$$

So for this step of the transform, the 'information' contained in the wavelet coefficients capture the differences between the pairs $(y_0, y_1), (y_2, y_3), (y_4, y_5)$ and (y_6, y_7). It is clear that starting the DWT from a different element of the series, could lead to a different

decomposition of the data. Hence, the DWT is not translation invariant. Specifically, if we were to perform the DWT on a shifted version of the series, the resulting decomposition would not be equal to a shift of the decomposition obtained from transforming the original series (e.g. see Figure 6). Antoniadis and Gijbels (2002) highlight that the lack of translation invariance of the DWT leads to non-uniformity in the localisation of a particular detail (e.g. a discontinuity) across the scales. So this motivates the question, are we 'missing information' that might be captured by performing the DWT on shifted pairs, such as (y_1, y_2) and (y_3, y_4)?

We now focus our attention on the work of Nason and Silverman (1995) who propose a modified DWT that no longer requires decimation and results in a 'stationary' wavelet transform that is not dependent on the origin, we refer to this procedure as the *non-decimated wavelet transform* (NDWT) and recall the non-decimated wavelet family introduced in equation (8).

In the manner of Nason and Silverman (1995), let us start by introducing some notation. For a sequence $\mathbf{x} = \{x_i\}$ (assuming boundary issues are treated), the action of the low-pass filter on $\mathbf{x}$ is defined by

$$(\mathcal{H}\mathbf{x})_k = \sum_n h_{n-k} x_n, \tag{20}$$

where $\mathcal{H} = \{h_n\}_{n\in\mathbb{Z}}$. The action of the high-pass filter $\mathcal{G}$ can be defined similarly. In order to 'fill the gaps', we define the following *dyadic decimation operators* that effectively allows us to capture the information contained in the differences between all overlapping pairs, for example (x_{i-1}, x_i) and (x_i, x_{i+1}). We define the *even* and *odd dyadic decimation operators*, respectively $\mathcal{D}_0$ and $\mathcal{D}_1$, to be

$$(\mathcal{D}_0\mathbf{x})_l = x_{2l}, \quad \text{and} \quad (\mathcal{D}_1\mathbf{x})_l = x_{2l+1}.$$

Simply put, the role of these operators is to select every even/odd member of a sequence. In fact, the DWT is an application of the even dyadic decimation operator, for instance, formulas (16) and (17) can be expressed as

$$\mathbf{c_{j-1}} = \mathcal{D}_0\mathcal{H}\mathbf{c_j}, \quad \text{and} \quad \mathbf{d_{j-1}} = \mathcal{D}_0\mathcal{G}\mathbf{c_j}, \tag{21}$$

where $\mathbf{c_j} = \{c_{j,k}\}_{k\in\mathbb{Z}}$ and $\mathbf{d_j} = \{d_{j,k}\}_{k\in\mathbb{Z}}$. In other words, to obtain the next coarse scale scaling coefficients, we filter each overlapping pair of elements in $\mathbf{c_j}$ using equation (20) and then select only the evenly indexed elements of the transformed sequence to be contained in $\mathbf{c_{j-1}}$, hence the dyadic decimation of the DWT. Replacing $\mathcal{D}_0$ with $\mathcal{D}_1$ in equations (21), would correspond to selecting only the oddly indexed elements of the transformed sequence. This is equivalent to performing the DWT on shifted pairs of the original input sequence and accounts for the 'missing' information, as mentioned previously. So, by applying and retaining both sets of even and odd indexed decimated transformed coefficients at each scale, more information about the original signal could be obtained. This is the basic idea of the NDWT.

The result is a redundant, non-decimated transform, such that at each scale both sets of scaling and wavelet coefficients are the same length as the original sequence, say $N = 2^J$. For instance, the application of $\mathcal{D}_0\mathcal{G}\mathbf{c_J}$ and $\mathcal{D}_1\mathcal{G}\mathbf{c_J}$ results in two sets of wavelet coefficients

at the finest scale, $J-1$, each of length $N/2$. For the next coarse scale, $J-2$, we apply both $\mathcal{D}_0\mathcal{G}$ and $\mathcal{D}_1\mathcal{G}$ to both of $\mathcal{D}_0\mathcal{G}\mathbf{c_{J-1}}$ and $\mathcal{D}_1\mathcal{G}\mathbf{c_{J-1}}$. This results in four sets of coefficients each of length $N/4$. Therefore, it can be seen that by repeating this procedure, at each successive coarse scale $\{J-l, l=1,...,J\}$ there will contain 2^l sets of coefficients each of length $2^{-l}N$, such that the total number of wavelet coefficients will equal $2^{-l}N \times 2^l$ $(=N)$ at each scale. A similar approach is taken for the scaling coefficients using the low-pass filter $\mathcal{H}$. Computationally, this procedure works by 'padding out' the filters, $\mathcal{H}$ and $\mathcal{G}$, with zeroes at each scale. More specifically, if $\mathcal{H}^{[j]}$ denotes the set of scale j filter coefficients, the next coarse scale filter $\mathcal{H}^{[j-1]}$ is obtained by inserting a zero between every adjacent pair of filter coefficients of $\mathcal{H}^{[j]}$ (see Nason and Silverman (1995) for further information).

An example of the NDWT, 'Haar' we go!

We return to our example of the DWT using the Haar wavelet basis, where the finest scale scaling coefficients $\mathbf{c_3}$ were approximated by the original data $\mathbf{y} = (y_0, ..., y_{N-1}) = (5, 3, 1, 9, 2, 2, 8, 4)$. The coefficients computed for the DWT using even dyadic decimation, form the evenly indexed elements of the $\mathbf{c_2}$ for the NDWT. So,

$$\mathbf{c_2} = (4\sqrt{2}, c_{2,1}, 5\sqrt{2}, c_{2,3}, 2\sqrt{2}, c_{2,5}, 6\sqrt{2}, c_{2,7}).$$

Then to fill in the gaps, the NDWT (assuming periodic bounding, i.e. $y_N = y_0$), additionally computes the coefficients for the overlapping pairs $(y_1, y_2), (y_3, y_4), (y_5, y_6)$ and (y_7, y_0). So, recalling the Haar wavelet filters given by equations (19) and using equation (20), the oddly indexed coefficients of $\mathbf{c_2}$ can be found to be

$$\begin{aligned}
c_{2,1} &= (\mathcal{H}\mathbf{c_3})_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \bullet (c_{3,1}, c_{3,2}) = 2\sqrt{2} \\
c_{2,3} &= (\mathcal{H}\mathbf{c_3})_3 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \bullet (c_{3,3}, c_{3,4}) = \frac{11}{\sqrt{2}} \\
c_{2,5} &= (\mathcal{H}\mathbf{c_3})_5 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \bullet (c_{3,5}, c_{3,6}) = 5\sqrt{2} \\
c_{2,7} &= (\mathcal{H}\mathbf{c_3})_7 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \bullet (c_{3,7}, c_{3,0}) = \frac{9}{\sqrt{2}},
\end{aligned}$$

where $\bullet$ refers to the vector dot product. This completes the set for the NDWT scaling coefficients at scale 2. The whole set (even and odd indexed elements) of scale 2 wavelet coefficients $\mathbf{d_2}$ can be constructed similarly, but replacing the low-pass filter with the high-pass filter $\mathcal{G} = (1/\sqrt{2}, -1/\sqrt{2})$. The next step is to retain the sets of even (subscripted by 0) and odd (subscripted by 1) indexed coefficients of $\mathbf{c_2}$ such that

$$\mathbf{c_{2,0}} = \mathcal{D}_0\mathcal{H}\mathbf{c_3} = (c_{2,0}, c_{2,2}, c_{2,4}, c_{2,6}), \quad \text{and} \quad \mathbf{c_{2,1}} = \mathcal{D}_1\mathcal{H}\mathbf{c_3} = (c_{2,1}, c_{2,3}, c_{2,5}, c_{2,7}). \quad (22)$$

Similar can be written for the wavelet coefficients. Then in the same manner as above, for scale 1 we apply both $\mathcal{D}_0\mathcal{H}$ and $\mathcal{D}_1\mathcal{H}$ to $\mathbf{c_{2,0}}$ which results in the sets of coefficients $\mathbf{c_{1,00}}$ and $\mathbf{c_{1,01}}$, and to $\mathbf{c_{2,1}}$ which results in the sets of coefficients $\mathbf{c_{1,10}}$ and $\mathbf{c_{1,11}}$. Note, we have added an additional 0 or 1 to the second subscript to denote the scale 1 even and odd indexed coefficients of the sets $\mathbf{c_{1,0}}$ and $\mathbf{c_{1,1}}$, respectively. Each of these sets are of length 2. Note, if we maintain this even and odd dyadic decimation structure, such that

we construct sets of coefficients at each coarse scale, we will construct a NDWT with a *packet*-ordered transform. This ordering is not of relevance to this thesis, so for further understanding the reader can see Nason (2008).

For time series applications, it is useful to have a *time*-ordered transform. We still maintain the idea of performing the wavelet transform on the even and odd decimation sets of coefficients, e.g. equations (22), but instead of physically constructing the sets (or packets) of scaling and wavelet coefficients, we instead 'pad-out' the filters with zeroes at each coarser scale in the manner we have previously mentioned. Therefore, the scale 1 scaling coefficients, $\mathbf{c_1}$, can be obtained from $\mathbf{c_2} = (4\sqrt{2}, 2\sqrt{2}, 5\sqrt{2}, 11/\sqrt{2}, 2\sqrt{2}, 5\sqrt{2}, 6\sqrt{2}, 9/\sqrt{2})$ by

$$
\begin{aligned}
c_{1,0} &= (\mathcal{H}\mathbf{c_2})_0 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,0}, c_{2,1}, c_{2,2}) = 9 \\
c_{1,1} &= (\mathcal{H}\mathbf{c_2})_1 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,1}, c_{2,2}, c_{2,3}) = \frac{15}{2} \\
c_{1,2} &= (\mathcal{H}\mathbf{c_2})_2 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,2}, c_{2,3}, c_{2,4}) = 7 \\
c_{1,3} &= (\mathcal{H}\mathbf{c_2})_3 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,3}, c_{2,4}, c_{2,5}) = \frac{21}{2} \\
c_{1,4} &= (\mathcal{H}\mathbf{c_2})_4 = (\frac{1}{\sqrt{2}}, 0.\frac{1}{\sqrt{2}}) \bullet (c_{2,4}, c_{2,5}, c_{2,6}) = 8 \\
c_{1,5} &= (\mathcal{H}\mathbf{c_2})_5 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,5}, c_{2,6}, c_{2,7}) = \frac{19}{2} \\
c_{1,6} &= (\mathcal{H}\mathbf{c_2})_6 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,6}, c_{2,7}, c_{2,0}) = 10 \\
c_{1,7} &= (\mathcal{H}\mathbf{c_2})_7 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) \bullet (c_{2,7}, c_{2,0}, c_{2,1}) = \frac{13}{2},
\end{aligned}
$$

where we have inserted a zero between the adjacent elements of the wavelet filter. The scale 1 wavelet coefficients $\mathbf{d_1}$ can be obtained similarly, but with use of the high-pass filter $\mathcal{G} = (1/\sqrt{2}, 0, -1/\sqrt{2})$. To obtain scale 0 coefficients, more zeroes would be inserted into the filters between each two elements, such that the Haar wavelet filters would be given by $(\cdot, 0, 0, 0, \cdot)$. Noticeably, unlike the DWT, the NDWT produces sequences of scaling and wavelet coefficients of length equal to the original data. This is evident in Figure 6 which displays the wavelet coefficients from performing the DWT and NDWT on the original data $\mathbf{y} = (5, 3, 1, 9, 2, 2, 8, 4)$ and on shifted data $\mathbf{y_s} = (4, 5, 3, 1, 9, 2, 2, 8)$. The application of both transforms on the shifted data, demonstrates the translation invariance of the NDWT and lack of translation invariance for the DWT. Specifically, it can be seen that the wavelet coefficients for the NDWT on the shifted data, correspond to a unit shift of the NDWT wavelet coefficients for the original data but this is not the case for the DWT.

The NDWT, as we will see later on, is useful for studying nonstationary time series that exhibit local features. Additionally, an application of the NDWT that produces 'packets' of the wavelet coefficients was introduced by Coifman and Donoho (1695). Their proposed *translation-invariant denoising* procedure was designed to remove noise from signals.

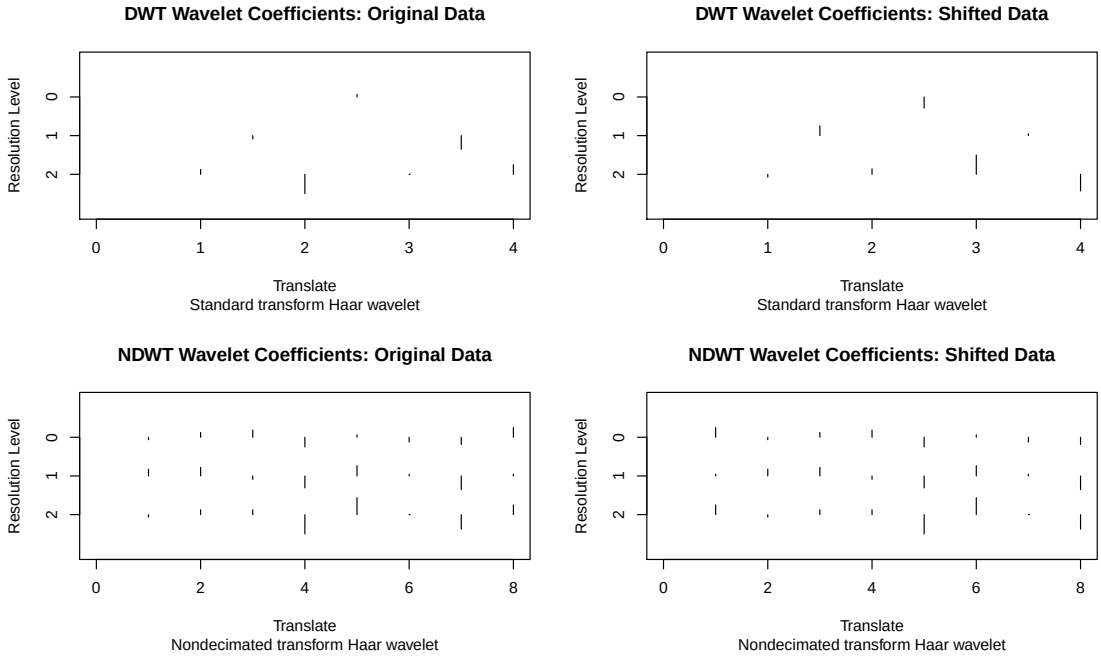


Figure 6: Haar wavelet coefficients computed using DWT and NDWT on the original data $\mathbf{y} = (5, 3, 1, 9, 2, 2, 8, 4)$ and on the shifted data $\mathbf{y_s} = (4, 5, 3, 1, 9, 2, 2, 8)$. Observe that the NDWT coefficients for the shifted data are a unit shift of the NDWT coefficients for the original data, thus demonstrating translation invariance. This is not the case for the DWT (top row).

1.2 Stationary time series analysis

Time series occur in a variety of diverse fields, such as economics (e.g. quarterly gross domestic product (GDP) or monthly unemployment rates); finance (e.g. daily closing prices of shares); meteorology (e.g. monthly rainfall measurements or temperature) and circadian biology (e.g. measurements of circadian rhythms in plants). In this thesis, our methodology will have a particular application to a problem arising in the neuroscience, where the time series are observations of electrical activity recorded in regions of the brain of a macaque monkey during an experiment. A time series given by $\{x_t\}$ can be thought of as a realisation of a stochastic process, $\{X_t\}$, observed sequentially through time. The analysis of time series has been under investigation throughout the 1900s and has split off into two main ways of approach, namely *time domain* and *spectral domain* analysis. We will briefly review both approaches. As time series analysis is a very wide subject there exist numerous reference books, a selection include Brockwell and Davis (1991), Percival and Walden (2000), Brillinger (2001), Chatfield (2003) and Shumway and Stoffer (2017).

A time series is said to be *stationary* if there is no systematic change in its statistical properties over time. In other words, its mean and variance are assumed to be constant over time. Some time series can be assumed to be stationary, however many time series exhibit seasonal variation and (long term) trend. Thus generally, the first step of time series analysis is to identify any components of variations and remove them (e.g. detrending and deseasonalising) in order to achieve stationarity.

Definition 1.2.1. A time series $\{X_t, t \in \mathbb{Z}\}$ is said to be *strictly stationary* if the joint distributions of $\{X_{t_1}, ..., X_{t_k}\}$ and $\{X_{t_{1+\tau}}, ..., X_{t_{k+\tau}}\}$ are the same for all $k \in \mathbb{Z}^+$ and for all $t_1, ..., t_k, \tau \in \mathbb{Z}$.

Intuitively, this means that the statistical properties of the time series are unaffected by a change of the time origin. In practice, stationarity is often defined in a less restrictive way.

Definition 1.2.2. A time series $\{X_t, t \in \mathbb{Z}\}$ is said to be *weakly stationary* or *second-order stationary* if its mean is constant through time and its autocovariance depends only on time-lags, such that

$$E[X_t] = \mu, \quad \text{and} \quad \text{cov}(X_t, X_{t+\tau}) = \gamma(\tau), \quad \forall t, \tau.$$

Under weak stationarity, the restrictions are placed only on the first and second order moments. The function $\gamma(\tau)$ is referred to as the autocovariance function of $\{X_t\}$ at lag τ.

A simple example of a stationary process is the purely random process defined as follows.

Definition 1.2.3. The process $\{Z_t, t \in \mathbb{Z}\}$ is said to be a *purely random process* (or *white noise*) if the random variables, $\{Z_t\}$, are uncorrelated with zero mean and variance σ_Z^2. Additionally, the random variables $\{Z_t\}$ are often assumed to be independently and identically distributed. The process is stationary with constant mean and under the assumption of independence, $\text{cov}(Z_t, Z_{t+\tau}) = \sigma_Z^2 \delta_{0,\tau}$ where δ is the Kronecker delta (Chatfield, 2003).

1.2.1 Analysis in the time domain

Key to the analysis of time series in the time domain, are the autocorrelation coefficients that describe the dependence structure between observations at different distances (lags) apart. The autocorrelation function defined by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \text{corr}(X_t, X_{t+\tau}),$$

measures the correlation between X_t and $X_{t+\tau}$. After estimation, the (sample) autocorrelation coefficients and the (sample) partial autocorrelation coefficients (which measure the amount of autocorrelation at each lag that is not accounted for by the autocorrelation at previous lags), are an essential aid to determine the order of the stochastic process chosen to model the time series.

A useful and most commonly chosen class of models for time series is formed by combining two processes, an *autoregressive* (AR) model in which X_t is regressed on past observations of X_t and a *moving-average* (MA) model that uses past errors.

Definition 1.2.4. (The ARMA(p, q) process.) The stationary process $\{X_t\}$ is said to be an *autoregressive moving-average* process of order (p, q) if for every t

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \tag{23}$$

where $\{Z_t\}$ is a purely random process.

The constants, $\{\phi_i\}$ and $\{\theta_j\}$, describe the internal process dependence of X_t. We can write equation (23) in the form

$$\Phi(B)X_t = \Theta(B)Z_t, \tag{24}$$

where $\Phi(B)$, $\Theta(B)$ are the polynomials

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \text{and} \quad \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q, \tag{25}$$

and B denotes the backward shift operator defined by $B^l X_t = X_{t-l}$. Thus, for an ARMA(p, q) process to be stationary and invertible, the latter ensuring the process is unique for a given autocorrelation, then it is necessary that the roots of polynomials

$$\Phi(z) = 0, \quad \text{and} \quad \Theta(z) = 0,$$

lie outside the unit circle, respectively for stationarity and invertibility (Shumway and Stoffer, 2017).

1.2.2 Analysis in the spectral domain

Recalling Section 1.1.1, we could consider decomposing a real-valued stationary process $\{X_t\}$ into a linear combination of sine and cosine components with uncorrelated coefficients. For instance, a discrete time series $\{X_t\}$ satisfying weak stationarity with mean zero, can be expressed by the Cramér representation (see (Brillinger, 2001, §4.6))

$$X_t = \int_{-\pi}^{\pi} A(\omega) e^{i\omega t} dZ(\omega), \tag{26}$$

where $A(\omega)$ is the *amplitude* of the process $\{X_t\}$ and $\{Z(\omega)\}$ is a random process with orthonormal increments $dZ(\omega)$ (i.e. $dZ(\omega) = Z(\omega + d\omega) - Z(\omega)$, $\mathbf{E}(dZ(\omega)) = 0$, $\mathbf{E}(dZ(\omega), \overline{dZ(\omega')}) = d\omega\delta_0(\omega - \omega')$, where δ is the Kronecker delta). Note, the amplitude is also referred to as the transfer function, $A(\omega) = \sum_{j=-\infty}^{\infty} a(j)e^{-i\omega j}$, where $a(j)$ are unknown function coefficients (Brillinger, 2001, §4.6). In comparison with the representation in equation (1) for a deterministic series, the integral in equation (26) is a stochastic integral. In other words, for each frequency, ω, each realisation of $Z(\omega)$ is allowed to vary, thus, for each realisation of X_t there will be a corresponding realisation of $Z(\omega)$. Note, the amplitude, $A(\omega)$, is deterministic.

In the time domain, we emphasised the use of the autocovariance (or autocorrelation) as a means to study the dynamics of a process over time. Here, in the spectral domain, we consider a function complementary to studying the autocovariance, the spectral density function (or spectrum).

Definition 1.2.5. The *spectral density function* or *spectrum*, is given as

$$f(\omega) = |A(\omega)|^2,$$

such that, $f(\omega)$ quantifies the contribution to the process variance of X_t, contributed by the components at frequency ω. Note, in engineering terms this is equivalent to the 'energy'

spectral density (Semmlow, 2012, §4.2).

The autocovariance function of $\{X_t\}$ can be expressed by the Fourier representation

$$c_X(\tau) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega\tau} d\omega, \tag{27}$$

where $f(\omega)$ is the spectrum. Thus, the spectrum and autocovariance are Fourier pairs.

Consider the ARMA(p, q) process in the form given by equation (24). The spectral density of a stationary ARMA(p, q) process X_t is given by (Brockwell and Davis, 1991)

$$f_X(\omega) = \frac{\sigma_Z^2}{2\pi} \left| \frac{\Theta(e^{-i\omega})}{\Phi(e^{-i\omega})} \right|^2, \tag{28}$$

where the polynomials in (25) give, $\Phi(e^{-i\omega}) = \sum_{j=0}^{p} \phi_j e^{-i\omega j}$ and similarly for $\Theta(e^{-i\omega})$ of order q. Note that the polynomials are of the same form as the transfer function $A(\omega)$ and the spectral density of an ARMA processes is often referred to as a rational spectral density. Note that in the form of equation (28), $f_X(\omega)$ is a 'power' spectral density, where σ_Z^2 is the 'energy' spectral density of $Z(\omega)$ and the $(2\pi)^{-1}$ factor is due to power being measured in terms of 'energy per unit time' (Priestley, 1982).

The periodogram is a quantity commonly used as an estimator for the spectrum. It is defined as the squared modulus of the discrete Fourier transform of the process $\{X_t\}$, and we introduce it below.

Definition 1.2.6. For a discrete time series $\{X_t\}_{t=0,\ldots,N-1}$, the *periodogram* is defined to be

$$I(\omega_j) = |D\hat{f}_j|^2$$

where $\omega_j = 2\pi j/N$ and the discrete Fourier transform of $\{X_t\}$ is given by $D\hat{f}_j = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i\omega_j t}$ (see Definition 1.1.3).

For each ω_j, the periodogram, in its 'raw' state defined above, is an asymptotically unbiased estimate of the spectrum $f(\omega)$, but despite this desirable property, the estimate is poor as it is not consistent (i.e. $\text{var}(I(\omega_j)) \not\to 0$ as $N \to \infty$). Smoothing the periodogram is the common way to achieve consistency and there are various approaches to smoothing, a simple way being to locally average adjacent periodogram ordinates. For further details on periodogram smoothing and Fourier-based spectral domain analysis, Priestley (1982) and Brillinger (2001) offer a comprehensive review.

Wavelet-based spectral analysis

Our discussion so far has been focused on Fourier methods which are generally regarded as canonical for spectral domain analysis of stationary time series. The Fourier spectrum can prove to be more informative than a wavelet-based spectrum due to the latter having lower frequency resolution (Nason, 2008). However, wavelets do have a useful part to play for analysing time series that contain information on a scale-by-scale basis (recall our earlier review of wavelet theory in Section 1.1.2). We only briefly introduce wavelet scale analysis here, as much of our interest is concerned with the analysis of nonstationary time series. Nason and von Sachs (1999) and Nason (2008) can be consulted for an introduction to this

topic, the former giving a review of relevant applications such as spectral density estimation using wavelet shrinkage denoising (Neumann, 1996). For a comprehensive review, Percival and Walden (2000) give a detailed account of the field and Chiann and Morettin (1998) introduce a wavelet periodogram for wavelet-based spectral analysis.

The *wavelet variance*, introduced by Percival and Guttorp (1994), provides a means to decompose the variance of a process across wavelet scales and hence it is a useful tool to study an event with variations occurring over a range of scales (e.g. applications in geophysics, see Kumar and Foufoula-Georgiou (1997)). The *Allan variance* (Allan, 1966), a measure of frequency stability in clocks and oscillators, can be understood as a simple example of the wavelet variance using Haar wavelets (see Nason and von Sachs (1999)), and other wavelets can be used for the general wavelet variance. The variance can be estimated based on decimated (DWT) coefficients or non-decimated (NDWT) coefficients, with the latter displaying more favourable results, which can be attributed to capturing additional information from overlapping pairs of a sequence (recall our discussion in Section 1.1.5).

Replicate time series

Traditionally, time series analysis has mostly been concerned with developing methodologies to study the dynamics of one (long) process realisation. What is not well accounted for in the literature, is the statistical modelling of a collection of (constant mean) time series, $\{x_{rt}, t = 1, ..., T, r = 1, ..., R\}$, arising from the same stimulus presented within an experimental setting. Among the first to consider modelling a replicate time series where the inferential focus is on the stochastic variation across replicates was Diggle and Al Wasel (1997). Motivated by an application to biomedical time series where the need was to analyse a collection of blood samples taken from subjects (replicates), the authors proposed a generalised linear mixed-effects model for periodograms to estimate the subject-specific population spectra.

If Y_{rj}, denotes the periodogram for the rth subject at frequency j, then for the array $\{Y_{rj}, r = 1, ..., R, j = 1, ..., m\}$, the general model is given as

$$Y_{rj} = f_r(\omega_j) Z_r(\omega_j) U_{rj}, \tag{29}$$

where for known explanatory variables $\{d_{rjk}\}$ and unknown parameters $\{\beta_k\}$,

$$f_r(\omega_j) = \exp\left\{\sum_{k=1}^{p} d_{rjk}\beta_k\right\},$$

denotes the common population spectrum corresponding to the rth subject and $\omega_j = 2\pi j/T$ denotes the jth Fourier frequency. In equation (29), U_{rj} are mutually independent, unit-mean exponential variates, and $Z_r(\omega)$ are independent copies of a stochastic process detailing the perturbation from the population spectrum of the rth subject. For further details and an exploration of parameter estimation, see Diggle and Al Wasel (1997).

Crucially note, stochastic variation was only assumed between subjects (or replicates). However, within the context of this section, each individual replicate series, $\{x_{rt}\}_t$, is assumed to be weakly stationary. In other words, the within-replicate statistical properties are assumed constant. A more general model that incorporates within-replicate

nonstationarity as well as variation across replicates will be explored in the next section.

1.3 Nonstationary time series analysis

In practice, time series are often nonstationary. Many time series exhibit a long term trend and seasonal effects that cause the properties of the time series to vary with time, albeit slowly in most cases. For example, the gross domestic product (GDP) will display an upwards or downwards trend reflecting the current performance of the economy; energy consumption differs throughout the year with changing weather patterns. Later on, our application to the macaque learning experiment will display nonstationarity as the macaque's behaviour, excitability towards reward and neuronal rhythms can contribute to the process variation over each replicate as well as over the entire experiment.

Previously for stationary processes we mentioned transforming nonstationary data to achieve stationarity, through for example differencing. The *autoregressive integrated moving-average* (ARIMA) process embeds a differencing procedure. The ARIMA(p, d, q) models is of the form given in equation (23) but we replace X_t with $W_t = \nabla^d X_t = (1 - B)^d X_t$, where B is the backwards shift operator and d is the number of differences of X_t. Thus, the process reduces to an ARMA(p, q) model for W_t. A generalisation of ARIMA, the seasonal (S)ARIMA model, includes further differencing in order to deal with seasonality (see Box *et al.* (2008) for a comprehensive review). These models approach nonstationarity by removing variation in order to apply stationary process theory. However, if the non-stationary features are of interest or the nonstationarity is manifest in the second-order moments, then it may be beneficial to model the time series in a manner that explicitly takes into account these characteristics. As this thesis is concerned with spectral domain analysis, we will now consider some approaches that generalise the spectral representation to handle nonstationary behaviour.

Short-time Fourier transform

We have seen that Fourier representations are well established in the spectral analysis of stationary processes where the statistical properties are invariant over time. The spectral analysis of nonstationary processes require time localisation in order to capture time varying features. A natural extension to Fourier analysis, which estimates the spectrum locally over time, was introduced by Gabor (1946). The *short-time Fourier transform*, also referred to as *Gabor transform* or *windowed Fourier transform*, is given by

$$STFT(t, \omega) = \int_{-\infty}^{\infty} x(\tau) g(\tau - t) e^{-i\omega\tau} d\tau,$$

where $g(\tau)$ is the time-localisation window function and $g, X \in L^2(\mathbb{R})$. The premise behind this transform is that the data within each window is assumed stationary. Thus, the window function is shifted over the time points of the signal in order to localise the Fourier transform. The contribution to the process variance of $x(t)$ at a frequency ω within the neighbourhood of t is given by the spectral density function,

$$f(t, \omega) = |STFT(t, \omega)|^2.$$

One problem with this approach concerns the choice of window width of $g(\tau)$, which remains unchanged for all frequencies. There is a trade-off between time and frequency resolution, such that better time resolution (and equivalently time localisation) comes at a cost of worsened frequency resolution, and vice versa. So intuitively, we would like to use a narrower window for finer time resolution at higher frequencies and a wider window to capture more information at lower frequencies. For a further understanding of this approach see Chui (1997) or Debnath (2002).

Evolutionary spectra

Priestley (1965) introduced the theory of *evolutionary spectra*, such that the Fourier spectrum describes the contribution to the process variance locally in the neighbourhood of the time instant t, whereas the spectrum for stationary processes (recall Section 1.2.2) described the process variation distributed over frequency over all time. Recalling the Cramér representation in equation (26), the amplitude function, $A(\omega)$ can be replaced with a time-varying version given by $A_t(\omega)$ $(= A(t, \omega))$. Consequently, a process $\{X_t\}$ is said to be an *oscillatory process* if it has representation in the form of equation (26) and $\{A_t(\omega)e^{i\omega t}\}$ are a family of oscillatory functions. The amplitude function, $A_t(\omega)$ (as a function of t) is constructed to 'vary smoothly', such that the absolute maximum of its Fourier transform is obtained at the zero frequency (i.e. A_t changes slowly as a function of time). Thus, instead of describing a function over all frequencies, in 'oscillatory' form a process, $\{X_t\}$, could be described by a sine wave with 'natural' frequency ω_0, modulated by the smoothly varying amplitude function $A_t(\omega_0)$. The natural frequency, ω_0, is the frequency at which a signal will tend to oscillate without any manipulation, or in other words, the Fourier transform of a process will be concentrated around $(\pm)\omega_0$.

1.3.1 Locally stationary processes (LSP)

The idea of having approximately stationary local processes stems from the evolutionary spectrum theory of Priestley (1965) and has led to a class of stochastic processes called *locally stationary processes*. Dahlhaus (1997) notes that due to the unpredictable nature of nonstationary processes, future observations may not give any information on the behaviour of the process at present, thus it is difficult to make considerations on the process asymptotics in the ordinary sense. To overcome this, the author introduced the concept of modelling in *rescaled time*, $u = t/T \in (0, 1)$, and thus developed a framework for asymptotic considerations within the locally stationary setting, such that the efficiency of local estimators could be evaluated. This infill asymptotic approach is most suitable for processes with slowly varying statistical characteristics but at each time point appear close to stationary. Such processes are said to be locally stationary.

Time domain methods from the stationary setting can be extended to the locally stationary case for time-varying processes through inference on segments. For instance, the coefficients of a time-varying autoregressive (tvAR) process can be estimated locally through classical Yule-Walker estimation on segments that are close to stationary (see Dahlhaus (1997)). The autoregressive conditional heteroscedasticity (ARCH) model, introduced by Engle (1982) to explicitly model a stochastic conditional variance, has been generalised to include time-varying parameters by Dahlhaus and Subba Rao (2006). The authors

introduce the tvARCH process which can be locally approximated by stationary ARCH processes. Kernel-based estimation of a time-varying regression function has been proposed by Vogt (2012) for nonparametric models. Here, the regressors are locally stationary and the regression function changes smoothly over time. The problem of prediction in locally stationary time series has been investigated by Dette and Wu (2020).

Within the spectral domain, frameworks have been developed for both locally stationary Fourier (LSF; (Dahlhaus, 1997)) and wavelet (LSW; Nason *et al.* (2000)) processes to provide time-frequency and time-scale representations. We will consider these approaches in more detail below. For a comprehensive review of inference on locally stationary processes in both the time and spectral domain, see Dahlhaus (2012).

The locally stationary Fourier framework embeds the concept of rescaled time, $u = t/T \in (0, 1)$, which would then allow the time-varying amplitude transfer function $A_t(\omega)$ to be replaced by an asymptotic version, $A(u, \omega)$. Constructing the amplitudes as functions of rescaled time means that as more observations on a time series are obtained, there is a greater availability of information to describes the local structure of $A(u, \omega)$, which in turn allows the asymptotic estimation. The behaviour of a process is ensured to be locally stationary through the smoothness of $A(u, \omega)$.

Definition 1.3.1. A stochastic process, $\{X_t\}$, modelled as a triangular stochastic array $\{X_{t;T}\}_{t=0,\ldots,T-1}$, belongs to the class of *locally stationary Fourier* (LSF) processes, if there exists a representation

$$X_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\omega) e^{i\omega t} d\xi(\omega),$$

and there exists a constant K, such that

$$\sup_{t,\omega} |A^0_{t,T}(\omega) - A(\frac{t}{T}, \omega)| \leq \frac{K}{T}, \quad \forall T, \tag{30}$$

where $\{\xi(\omega)\}$ is a random stochastic process with properties given in Dahlhaus (1997).

Then from equation (30), it can be seen how for LSF processes, the asymptotic transfer function $A(u, \omega)$ controls the slow evolution of the time dependant amplitudes $A^0_{t,T}$.

Naturally, there exists an associated evolutionary (or time-varying) spectrum for LSF processes, given by

$$f(u, \omega) = |A(u, \omega)|^2,$$

defined in terms of rescaled time $u = t/T$.

As a parallel to Dahlhaus (1997), Ombao *et al.* (2002) utilise the SLEX (smoothed localised complex exponentials) library of waveforms, which give localisation in both time and frequency, to provide a time-dependent 'Cramér-like' representation for nonstationary processes. The SLEX functions used can be thought of as localised versions of the Fourier exponential functions. Ombao *et al.* (2005) have extended this work to the multivariate setting.

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).

LSF processes for replicate time series

The extension of time-varying Fourier methods has been considered for the scenario of replicate time series where there is potential for evolutionary (replicate-varying) dynamics across the replicates. Qin *et al.* (2009) introduce a covariate-indexed locally stationary time series, a slightly modified version of LSF processes that indexes a representation for each replicate. Through the locally stationary processes framework, the authors developed a time-frequency functional model, where the time-varying log-spectra determines the evolution of the stochastic variation. The authors note that their methodology extends the theory of functional data analysis (FDA) to the locally stationary time series setting. A review of FDA is soon to follow.

More recently, Fiecas and Ombao (2016) propose a new time series model which they name the *slowly evolving locally stationary process* (SEv-LSP). The model adopts the use of the Fourier sine and cosine functions whilst allowing for a time varying spectral density matrix within and across replicates. Note, the authors refer to trials as replicates over an experiment. We again encounter the concept of rescaled time, $u = t/T \in (0, 1)$, within replicates and similarly they define *rescaled replicate-time*, $\nu = r/R \in (0, 1)$, across all replicates. Their work is developed in the multivariate setting under the assumption that replicates are uncorrelated.

Definition 1.3.2. For replicates $r = 1, ..., R$ and within-replicate time $t = 1, ..., T$, a locally stationary P-variate time series, $\boldsymbol{X}_{t,r}$ follows a SEv-LSP if it admits the representation

$$\boldsymbol{X}_{t,r} = \int_{-0.5}^{0.5} \boldsymbol{A}(\frac{t}{T}, \frac{r}{R}, \omega) e^{i2\pi\omega t} d\boldsymbol{Z}_r(\omega),$$

where $\boldsymbol{A}(u, \nu, \omega)$ denotes the $P \times P$ complex-valued transfer function matrix and $\{d\boldsymbol{Z}_r(\omega)\}$ is a P-variate orthogonal increment process with properties given in Fiecas and Ombao (2016).

To capture the evolutionary nature of the process, $\boldsymbol{A}(u, \nu, \omega)$ is allowed to vary slowly over time within a replicate and furthermore over all replicates. The 'smoothly' changing dynamics of $\boldsymbol{A}$ over replicates means that the stochastic properties across neighbouring replicates appear similar but different for replicates further apart. The authors additionally introduce the $P \times P$ *evolving evolutionary spectral density matrix*, which at frequency $\omega \in (-0.5, 0.5)$, it is defined as

$$\mathbf{f}(u, \nu, \omega) = \boldsymbol{A}(u, \nu, \omega)\boldsymbol{A}(u, \nu, \omega)^*, \quad \forall u, \nu \in (0, 1),$$

where $(^*)$ denotes the conjugate transpose. The spectral density matrix naturally inherits the smooth evolutionary properties of $\boldsymbol{A}$, such that the characteristics of the underlying SEv-LSP may appear similar for neighbouring replicates but replicates much further apart may appear as completely different processes. As we are in a multivariate setting, a further quantity of interest that measures the linear dependence between a two sets of replicate time series is the cross-channel coherence. In the context of SEv-LSPs, the *evolving evolutionary*

coherence between dimensions p and q is given by

$$\rho^2_{pq}(u,\nu,\omega) = \left| \frac{f(u,\nu,\omega)_{pq}}{\sqrt{f(u,\nu,\omega)_{pp} f(u,\nu,\omega)_{qq}}} \right|,$$

where $f(u,\nu,\omega)_{pq}$ is the (p,q)th element of $\mathbf{f}(u,\nu,\omega)$.

The above work of Fiecas and Ombao (2016) on SEv-LSPs, provides motivation for the methodology developed in this thesis, as we will bypass the limiting assumption of replicate uncorrelation while additionally providing a *multiscale* analysis. Note, 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.

1.3.2 Locally stationary wavelet (LSW) model

As previously mentioned, framework for the modelling of locally stationary stochastic processes has been developed in the wavelet domain. Nason *et al.* (2000) introduced the class of *locally stationary wavelet* (LSW) processes, whose framework underpins the methodology developed within this thesis.

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, which replace the Fourier exponentials $\{\exp(i\omega t), \omega \in (-\pi,\pi)\}$. As described in Nason *et al.* (2000), a set of compactly supported *discrete wavelets* $\boldsymbol{\psi}_j = \{\psi_0, ..., \psi_{j,N_j-1}\}_{j\in\mathbb{Z}^+}$ of length N_j can be constructed for consecutively finer scales j using the following formulae:

$$\begin{aligned}
\psi_{1,n} &= \sum_k g_{n-2k}\delta_{0,k} = g_n, \quad \text{for} \quad n = 0, ..., N_1 - 1, \\
\psi_{j+1,n} &= \sum_k h_{n-2k}\psi_{j,k}, \qquad \text{for} \quad n = 0, ..., N_{j+1} - 1 \\
N_j &= (2^j - 1)(N_h - 1) + 1,
\end{aligned}$$

where $\{h_k\}$ and $\{g_k\}$ are the low- and high-pass quadrature mirror filters as defined in Section 1.1.3, $\delta_{0,k}$ is the Kronecker delta and N_h is the number of non-zero elements of $\{h_k\}$.

To obtain the set of *non-decimated* discrete wavelets, $\{\psi_{j,k}(t)\}_{t=0,...,T-1}$, we can shift $\boldsymbol{\psi}_j$ to all integer locations k

$$\psi_{j,k}(t) = \psi_{j,k-t},$$

where $\psi_{j,k}$ is the kth element of the vector $\boldsymbol{\psi}_j$. Although this results in an overcomplete collection of shifted vectors, the NDWT is ensured to be translation invariant.

Definition 1.3.3. For $T = 2^{J(T)}$, a sequence of stochastic processes $\{X_{t;T}\}_{t=0,...,T-1}$ is a

locally stationary wavelet (LSW) process if it admits the representation

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \omega_{j,k;T} \psi_{j,k}(t) \xi_{j,k}, \tag{31}$$

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 identically distributed random variables.

The representation of $\{X_{t;T}\}$ can be thought of as the multiscale equivalent of the previous Fourier-based representations we have seen, such that the process is built through a linear combinations of oscillatory wavelet functions $\{\psi_{j,k}\}$ with random amplitudes $\{\omega_{j,k;T}\xi_{j,k}\}$. The quantities in representation (31) have the following properties:

1. For all j and k, $\mathbf{E}[\xi_{j,k}] = 0$ ($\Rightarrow \mathbf{E}[X_{t;T}] = 0$).
2. $\mathbf{E}[\xi_{j,k}\xi_{j',k'}] = \delta_{j,j'}\delta_{k,k'}$ ($= \text{cov}(\xi_{j,k}, \xi_{j',k'})$).
3. For each scale $j \geq 1$, there exists a Lipschitz continuous function in rescaled time ($z = k/T$) denoted by $W_j(z)$ with the following properties

   (a)
   $$\sum_{j=1}^{\infty} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1).$$

   (b) The Lipschitz constants L_j are bounded in j and
   $$\sum_{j=1}^{\infty} 2^j L_j < \infty.$$

   (c) There exist a sequence of constants C_j such that, for each T the amplitudes are forced to vary slowly across time, in the sense that
   $$\sup_{k=0:T-1} \left| \omega_{j,k;T} - W_j\left(\frac{k}{T}\right) \right| \leq \frac{C_j}{T}, \quad \forall j,$$
   where $\{C_j\}$ fulfills $\sum_{j=1}^{\infty} C_j < \infty$.

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 $W_j(k/T)$ (property 3), that can be thought of as a scale (j) and time (k) dependent transfer function (Fryzlewicz and Nason, 2006). Nason *et al.* (2000) propose the *evolutionary wavelet spectrum* (EWS) as a means to quantify the contribution to the overall process variance at a scale j and rescaled time $z = k/T$, formally defined as:

$$S_j(z) = |W_j(z)|^2,$$

for all $j = 1, ..., J$. (Note that the spectra are non-negative).

For stationary processes, the autocovariance and spectrum are Fourier pairs (equation (27)). A similar link can be made here for a process $\{X_{t;T}\}_{t=0,...,T-1}$ with the above

defined EWS and autocovariance $c_T(z,\tau) = \text{cov}(X_{\lfloor zT \rfloor;T}, X_{\lfloor zT \rfloor+\tau;T})$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Nason *et al.* (2000) show that c_T tends asymptotically to a *local autocovariance*, defined as

$$c(z,\tau) = \sum_{j=1}^{J} S_j(z)\Psi_j(\tau),$$

where τ is an integer lag, $z \in (0,1)$ and

$$\Psi_j(\tau) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{j,k}(\tau),$$

is the scale j *autocorrelation wavelet* for all $j = 1, ..., J$.

The *raw wavelet periodogram* is used for estimation of the EWS, $\{S_j(z)\}_j$, and is defined as

$$I_{j,k;T} = |d_{j,k;T}|^2 \tag{32}$$

where

$$d_{j,k;T} = \sum_{t=0}^{T-1} X_{t,T}\psi_{j,k}(t)$$

are the *empirical wavelet coefficients* at scale j and time k associated to a discrete non-decimated family of wavelets.

The estimator $I_{j,k:T}$, is shown to be biased for the spectrum (Nason *et al.*, 2000, Proposition 4)

$$E[I_{j,\lfloor zT \rfloor;T}] = \sum_{l=1}^{J} A_{j,l}S_l(z) + O(T^{-1}), \quad \forall z \in (0,1).$$

The asymptotic bias arises as a result of the aforementioned overcompleteness of the set of non-decimated wavelets $\{\psi_{j,k}(t)\}$ (Van Bellegem and von Sachs, 2008). Correction for bias can be attained, and we first need to define $A_{j,l}$ to be the inner product for the j and l autocorrelation wavelets given by

$$A_{j,l} = \langle \Psi_j, \Psi_l \rangle = \sum_{\tau} \Psi_j(\tau)\Psi_l(\tau),$$

which then form the elements of the $J \times J$ *autocorrelation wavelet inner product matrix* defined as $A_J = (A_{j,l})_{j,l}$.

By introducing vector notation such that, $\mathbf{I}(z) = (I_{j,\lfloor zT \rfloor;T})_{j=1,...,J}$ and $\mathbf{S}(z) = (S_j(z))_{j=1,...,J}$, a vector of *corrected* wavelet periodograms gives an asymptotically unbiased estimator of $S_j(z)$ in the form of

$$\mathbf{L}(z) = A_J^{-1}\mathbf{I}(z), \quad \forall z \in (0,1), \tag{33}$$

where $\mathbf{L}(z) = (L_{j,\lfloor zT \rfloor,T})_{j=1,...,J}$.

Furthermore, Proposition 4 of Nason *et al.* (2000) shows that the wavelet periodogram estimator, $I_{j,k:T}$, is inconsistent due to having asymptotically non-vanishing variance. Thus, to obtain consistency, the raw periodogram is first smoothed before applying the above correction procedure. Various approaches to smoothing exist and a detailed account is given by Nason (2008) with particular attention to wavelet shrinkage techniques originally

introduced by the seminal works of Donoho and Johnstone (1994); Donoho *et al.* (1995). Further considerations that naturally follow on from the LSW framework include: Fryzlewicz and Nason (2006) who propose a variance stabilisation approach based on the Haar-Fisz transform, and Van Bellegem and von Sachs (2008) who, through introducing a pointwise adaptive estimator of the evolutionary spectrum, extend the class of LSW processes to contain processes that may exhibit sudden changes in their dynamics.

A short LSW example

In practice we will start with a real time series but for now let us specify a time series that exhibits some time-varying characteristics of interest. This example is akin to Chapter 5 of Nason (2008). For a chosen $T = 2^J = 1024$, we specify the 'true' EWS to be

$$S_j(z) = \begin{cases} 1, & \text{for } j = 9, z \in (\frac{400}{1024}, \frac{500}{1024}), \\ \cos^2(4\pi z), & \text{for } j = 4, \in (0, 1), \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

In resolution level $j = 4$ we place a squared cosine behaviour and a 'burst' of value 1 appears in the finest level $j = 9$. Figure 7 provides plots for this example. A visualisation of the true spectrum can be seen in the top left plot. Additionally, the top right plot presents a simulated realisation of a process displaying the characteristics defined by the EWS. In practice we do not know the true spectrum and we will try to estimate it using the NDWT under the LSW framework. We attempt to estimate the spectrum via the raw (uncorrected) periodogram defined in equation (32) and also through the corrected periodogram defined in equation (33). We repeat the processes of simulating a realisation and estimating the spectrum 100 times in order to obtain 'on average' estimates of the wavelet periodograms. Thus, the bottom left and right plots of Figure 7 show the mean of 100 raw and corrected periodogram estimates of the true spectrum, respectively. Clearly the estimates from the corrected periodogram are much better. Bias correction has the effect of scaling up finer scale coefficients, hence the burst in the finest level is much more apparent. Additionally, bias correction 'smooths out' spectral leakage into neighbouring levels, for instance, see how the corrected spectral estimates in level 3 are closer to the truth than the raw estimates.

Extensions of the LSW framework to the multivariate setting appear in the work of Sanderson *et al.* (2010) and Park *et al.* (2014), where a coherence structure between channels is embedded in the model.

A note on tests of stationarity

In practice, the above short LSW analysis could have begun with a test of stationarity. Carrying out such a test could inform us as to whether a model allowing for nonstationary behaviour is necessary. Additionally, if a process does exhibit second-order nonstationarity, a test could give some indication of where significant nonstationarities are located. Specifically, consider the Haar wavelet test of stationarity (HWTOS) developed by Nason (2013). If we implement the HWTOS test on the realisation given in the top right plot of Figure 7, we are informed that the realisation provides enough evidence to reject the null hypothesis

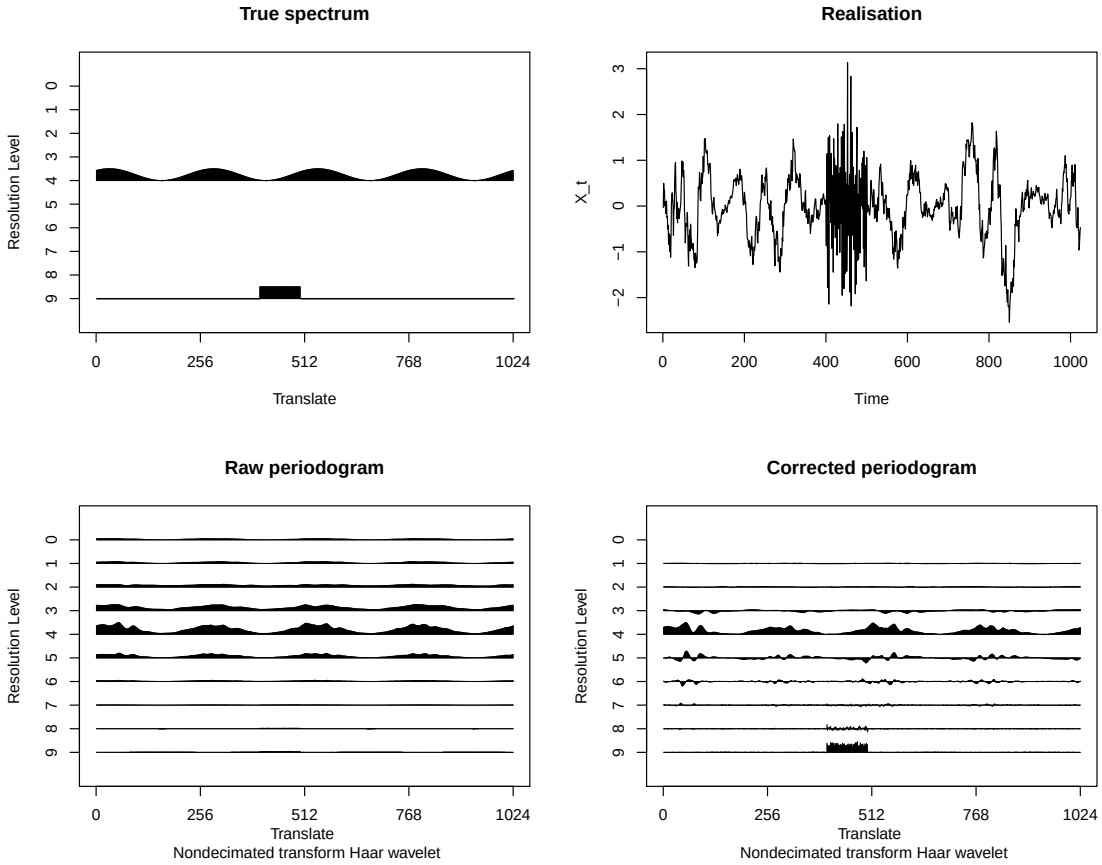


Figure 7: LSW example plots. *Top left*: true spectrum defined in equation 34; *Top right*: realisation simulated from the true spectrum; *Bottom left*: raw periodogram estimates on average over 100 realisations from the true spectrum; *Bottom right*: corrected periodogram estimates on average over 100 realisation from the true spectrum. Plotted using `Wavethresh` (Nason, 2008).

of stationarity. The HWTOS test also indicates where the nonstationarities are located in the series and this information is summarised nicely in Figure 8. The double headed arrows indicate the time locations for which nonstationarities were detected. As the test is based on wavelet transforms, the nonstationarities are also localised across scales j, given by the right-hand axis. The test detects 12 departures from stationarity. The time and scale locations of the nonstationarities appear to coincide with the true spectrum defined in equation (34) that characterises the realisation we have performed the HWTOS test on.

In Chapter 5 we propose two tests to assess whether a significant replicate-effect is manifest across a replicate time series. An in-depth review of the relevant literature on tests of stationarity will be given there.

1.4 Functional regression

Modelling replicate time series data is often framed as a functional regression problem (see e.g. Morris (2015) for a review). In contrast to the approach within this thesis 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*

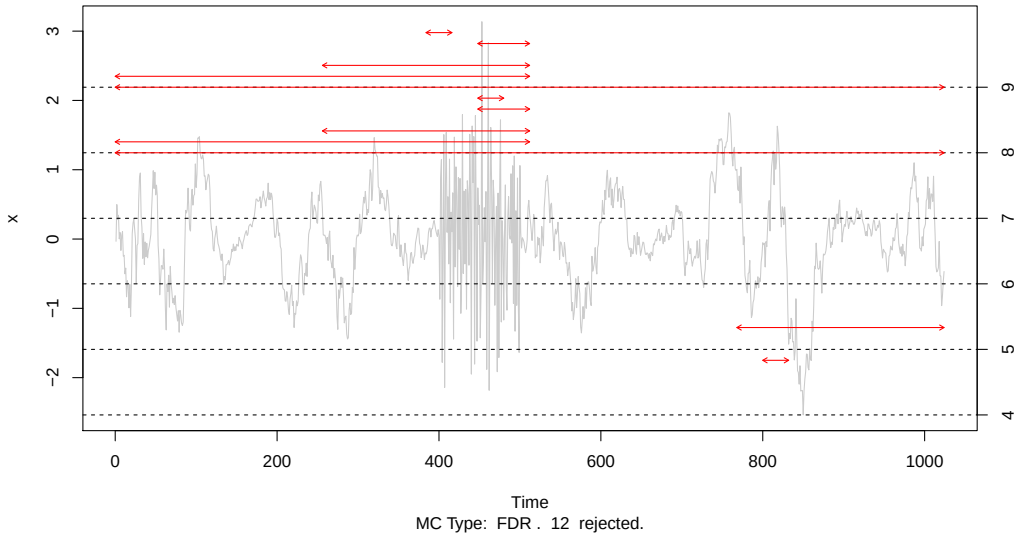


Figure 8: Test of stationarity plot of a realisation simulated from the true spectrum defined by equation (34). The location for which a nonstationarity was detected is indicated by a double headed arrow. The right-hand side axis indicates the scale j of the wavelet periodogram where the nonstationarity was detected. Plotted using `locits` (Nason, 2013).

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. For now we give a brief review of the framework for the functional regression approach.

Functional regression is a branch of *functional data analysis* (FDA) which concerns analysing discrete data expressed in the form a function. In general, FDA involves a large number of 'replicates' of the data, such that the analysis is then formed on a sample of functions taken from the population of a subject. Thus, in part, longitudinal data studies which collect data through repeated measurements per subject over an extended period of time, have largely motivated the creation of this field. The aim of these studies is often to understand change over time or time-dynamics such as growth. See Diggle *et al.* (2002) for a comprehensive review on longitudinal data analysis. From a functional perspective, principal components and regression methods (for example Fan and Zhang (2000); Yao *et al.* (2005a,b); Greven *et al.* (2010)) are key tools for FDA on longitudinal data and an excellent review is given by Müller (2005). For functional regression, most developments in the field stem from the general framework presented in the seminal textbook of Ramsay and Silverman (2005) which should be initially consulted for a further understanding on FDA. The review of Morris (2015) gives a comprehensive account of functional regression alongside numerous references to historical and current developments in the area. There exist three methods of functional regression analysis: scalar-on-function regression, function-on-scalar regression and function-on-function regression. We will limit this brief review to the case of function-on-scalar regression, which involves the regression of a functional response on a set of scalar independent variables. This approach is quite common and also forms the basis framework for models concerned with spectral domain analysis.

In order to provide a functional representation of the discrete data we use a linear combination of basis functions, and it should come as no surprise that Fourier series and wavelets are candidates for that role, with splines and principal components being further options. The key philosophy is to think of each sampled function as a single self-contained object that gives rise to the observed discrete data. Morris (2015) highlights that two key factors are at working during functional regression, *replication* and *regularisation*. Replication is the concept of combining information across replicate sampled functions in order to investigate a relationship structure between them and improve inference on their parent populations. Within each function, regularisation (smoothing) exploits the assumption of global smoothness of the functions and 'borrows strength' across neighbouring observations, which in turn leads to improved efficiency and better estimation. Regularisation is typically determined through the choice of basis functions (e.g. for wavelets see: wavelet shrinkage introduced by Donoho and Johnstone (1994); Donoho *et al.* (1995), and Nason (2008) for a review of further approaches).

Definition 1.4.1. For replicates $i = 1, ..., N$, a sample of functional responses $Y_i(t_j)$ with p scalar independent variables X_{ia}, can be expressed as a *functional response regression* model given by

$$Y_i(t_j) = \sum_{a=1}^{p} X_{ia} B_a(t_j) + E_i(t_j),$$

where $t_j, j = 1, ..., T_i$ denotes the continuum in time over which the functional data are recorded (Morris, 2015).

In the above model, $X = \{X_{ia}, i = 1, ...N; a = 1, ..., p\}$ is the $N \times p$ design matrix associated with the set of functional fixed effects $B(t_j) = \{B_a(t_j), a = 1, ..., p\}$. The set $E(t_j) = \{E_i, i = 1, ..., N\}$ are the residual error functions whose covariance structure describes the within-function covariance. The above model, which assumes an independent sample of functions, can be extended to allow for correlations between functions through including random effect terms.

Definition 1.4.2. For replicates $i = 1, ..., N$, a sample of functional responses $Y_i(t_j), j = 1, ..., T_i$ with $(p + q)$ scalar independent variables X_{ia} and Z_{ib}, can be expressed as a *functional mixed effects* model (FMM) given by

$$Y_i(t_j) = \sum_{a=1}^{p} X_{ia} B_a(t_j) + \sum_{b=1}^{q} Z_{ib} U_b(t_j) + E_i(t_j),$$

where $Z = \{Z_{ib}, i = 1, ...N; b = 1, ..., q\}$ is the $N \times q$ design matrix associated with the set of functional random effects $U(t_j) = \{U_b(t_j), b = 1, ..., q\}$.

The above FFM model was introduced by Morris and Carroll (2006) and as a result of their modelling assumptions, namely that $U(t_j)$ and $E(t_j)$ are multivariate Gaussian processes, the model allows for correlation between random effect functions through the construction of their design matrices. Additionally, the model can be further generalised by partitioning the random effects to allow for multiple levels of random effect functions. Note, as reflected in our above models, most sample functions are defined on a 1D Euclidean

domain but functions defined on higher domains are possible, for instance (Martinez *et al.*, 2013) model spectrograms as a 2D image defined for frequency and time.

Methods of functional response regression have been used for point estimation, constructing confidences intervals and various approaches to hypothesis testing, with much attention on growth curve analysis (i.e. estimation of the mean of sample functions). In Chapter 5, two hypothesis tests that assess the constancy of the spectrum are proposed. Within the spectral domain, various tests to compare the differences in mean curves have been constructed underpinned by the functional regression framework (see Chapter 5 for a brief overview of some of the developments in this area).

2 Methodology under the assumption of uncorrelated replicates

The original LSW formulation (Nason *et al.*, 2000) discussed in Chapter 1 (Section 1.3.2) is for a single-replicate. Thus, it cannot capture the dynamics of time series data recorded for several replicates (e.g. trials within an experiment), nor can it account for the potential dependence across replicates. 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 replicate, and (ii) across replicates over the course of the entire experiment.

The chapter proceeds as follows. Section 2.1 (i) introduces our proposed meta-process model under the assumption of uncorrelated replicates while accounting for intra- and cross-replicate nonstationarity, and (ii) develops its associated estimation theory. Section 2.2 details simulation studies that showcase the behaviour of the proposed methodology and demonstrates the advantage of our proposed work, both for across time and within-replicate behaviour characterisation. We conclude the chapter in Section 2.3. The potential dependence across replicates will be tackled in Chapter 3.

2.1 The proposed replicate-evolving nonstationary model

2.1.1 Replicate-Evolving Locally Stationary Wavelet (REv-LSW) process

Definition 2.1.1. 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 replicate $r = 0, \ldots, R-1$ where $R = 2^{J(R)}$ to be a *replicate-evolving locally stationary wavelet* (REv-LSW) process if it admits the following representation

$$X_{t;T}^{r;R} = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \omega_{j,k;T}^{r;R} \psi_{j,k}(t) \xi_{j,k}^{r}. \tag{35}$$

For each replicate (or replicate) r and within-replicate 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 replicate and $z = k/T$ denote rescaled within-replicate time, the quantities in (35) 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. $\mathbf{E}[\xi_{j,k}^{r} \xi_{j',k'}^{r'}] = \delta_{j,j'} \delta_{k,k'} \delta_{r,r'}$ ($= \operatorname{cov}(\xi_{j,k}^{r}, \xi_{j',k'}^{r'})$). This amounts to assuming uncorrelated replicates.
3. For each scale $j \geq 1$, there exists a Lipschitz continuous function in both rescaled time (z) and rescaled replicate (ν), denoted by $\widetilde{W}_j(z, \nu)$ with the following properties

   (a)
   $$\sum_{j=1}^{\infty} \left| \widetilde{W}_j(z, \nu) \right|^2 < \infty \text{ uniformly in } z \in (0,1), \nu \in (0,1). \tag{36}$$

   (b) Let L_j^{ν} denote the bounded Lipschitz constant corresponding to the time dimension at a particular rescaled replicate ν and scale j. Similarly, denote by

N_j^z the bounded Lipschitz constant corresponding to the replicate 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

$$\sum_{j=1}^{\infty} 2^j L_j < \infty \text{ and } \sum_{j=1}^{\infty} 2^j N_j < \infty. \tag{37}$$

(c) There exist sequences of bounded replicate-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 replicate and across replicates, in the sense that

$$\sup_{k=0:T-1} \left| \omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right) \right| \leq \frac{C_j^r}{T}, \quad \forall j, r, \tag{38}$$

$$\sup_{r=0:R-1} \left| \omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right) \right| \leq \frac{D_j^k}{R}, \quad \forall j, k. \tag{39}$$

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 $\sum_{j=1}^{\infty} 2^j D_j < \infty$.

Remark 2.1.2 (rescaled time and replicate evolution). Within each scale j, the transfer function $\widetilde{W}_j(z, \nu)$ controls the evolution of the amplitudes, forcing them to vary slowly over *both* rescaled time (z) and replicate (ν) dimensions. The evolution of the amplitudes over time within each replicate (replicate) happens in a smooth manner. The evolution across replicates 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 neighbouring replicates are similar. Nevertheless, further apart replicates may display different traits. Such a meta-process evolution is presented in Figure 9.

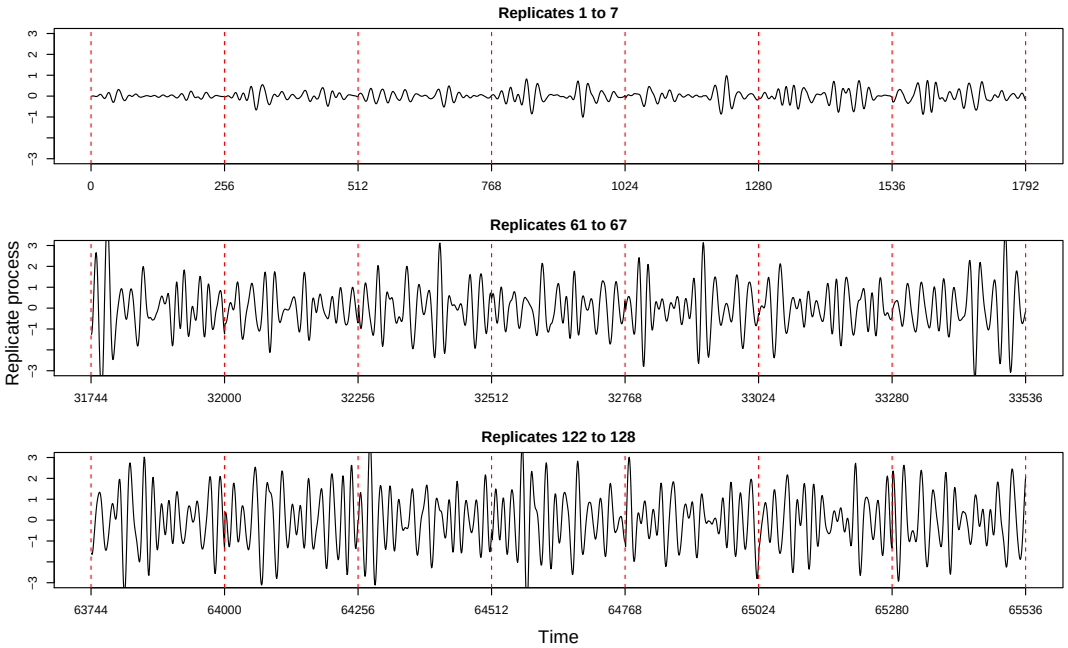


Figure 9: Realisation of a REv-LSW process demonstrating evolution across both time and replicate dimensions. (Vertical lines denote breaks between replicates. Concatenation is only used for meta-process visualisation.)

2.1.2 Replicate evolutionary wavelet spectrum

As is common in spectral domain analysis (both Fourier and wavelet-based), we do not work directly with the time- and replicate-specific multiscale transfer functions $\{\widetilde{W}_j(\cdot\,,\cdot)\}_j$, but instead we define a scale-dependent measure for the time and replicate 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 replicates. Next, we develop a novel evolutionary wavelet spectrum capable to extract nonstationarity within-replicates.

Definition 2.1.3. For a REv-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 2.1.1, the *within-replicate evolutionary wavelet spectrum* at scale j, rescaled replicate ν, rescaled within-replicate time z is given by

$$S_j(z,\nu) = \left|\widetilde{W}_j(z,\nu)\right|^2 = \lim_{\substack{T\to\infty \\ R\to\infty}} \left(\left| \omega_{j,\lfloor zT \rfloor;T}^{\lfloor \nu R \rfloor;R} \right|^2 \right). \tag{40}$$

From equations (38) and (39) we directly obtain that for each T and R we have

$$\sup_{r=0:R-1} \sup_{k=0:T-1} \left| \omega_{j,k;T}^{r;R} - \widetilde{W}_j\left(\frac{k}{T}, \frac{r}{R}\right) \right| = \mathcal{O}(C_j T^{-1}) + \mathcal{O}(D_j R^{-1}), \tag{41}$$

hence the right-hand equality in equation (40).

Remark 2.1.4 (**REv-LSW versus LSW processes).** An innovation of the proposed REv-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 2.1.1. Note that a replicate-evolving locally stationary wavelet (REv-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 replicates.

Remark 2.1.5 (**bounded variation jumps).** Our theoretical development could of course be extended to encompass bounded variation jumps, but this is outside the scope of this work. Nevertheless, we show through simulation that such behaviour is well handled by the proposed methodology.

For completeness and in order to aid our theoretical developments (and proofs), below we also introduce the local autocovariance function associated to a REv-LSW process, but note that we will not extensively pursue this quantity as our focus is on analysis within the spectral domain.

Definition 2.1.6. For a REv-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 2.1.1, the *within-replicate local covariance* at some rescaled time $z \in (0,1)$ within rescaled replicate $\nu \in (0,1)$, 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.

Note that $|c(z,\nu;\tau)| < \infty$ follows directly from the coherence range between -1 and 1, and from the uniform bounds in lag (τ) and rescaled replicates (ν) for both the limiting amplitudes and the autocorrelation wavelets (see equation (36)).

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 2.1.7. *For a REv-LSW process $\{X^{r;R}_{t;T}\}$ with properties as in Definition 2.1.1,*

$$\left|\mathrm{cov}(X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor;T}, X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor+\tau;T}) - c(z,\nu;\tau)\right| = \mathcal{O}(T^{-1}) + \mathcal{O}(R^{-1}),$$

uniformly in τ at (rescaled) time z and replicates ν.

Proof. The proof appears in Appendix A.3.1 and uses the approximation properties in Definition 2.1.1 of the REv-LSW process. □

2.1.3 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 within-replicate local autocovariance, by directly replacing the unknown spectrum in Definition 2.1.6. However, the development and investigation of the local covariance, potentially along with the partial covariance and their estimators for REv-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 2.1.8. For a scale j and time k, we define the *within-replicate raw wavelet periodogram*of a REv-LSW process to be

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

where $d^{r;R}_{j,k;T} = \sum_{t=0}^{T-1} X^{r;R}_{t;T}\psi_{j,k}(t)$ are the process empirical wavelet coefficients constructed using a family of discrete non-decimated wavelets, $\{\psi_{j,k}(t)\}_{j,k}$.

We note that unlike the Fourier periodogram, the wavelet-based raw periodogram is typically not an unbiased estimator of the wavelet spectrum, and this will also turn out to be the case here.

For reasons that will become obvious next, we also define a transformed spectral quantity $\beta_j(z,\nu) = \sum_{l=1}^{\infty} A_{j,l}S_l(z,\nu)$, 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. 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 β. Hence we next focus on estimating β and note that the theoretical results below are derived under the Gaussianity assumption.

Remark 2.1.9 (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). Nason (2013) illustrate the non-limiting character of this assumption for practical applications, which indeed holds for data arising from other fields, e.g. for experimental circadian data (Hargreaves *et al.*, 2019).

However, an empirical investigation (see Remark 2.2.2) into the quality of our estimation in the presence of heavy-tailed innovations, namely Student distributed with 5, 7 and 10 degrees of freedom, demonstrates robustness against departures from normality.

Proposition 2.1.10. *For a REv-LSW process* $\{X_{t;T}^{r;R}\}$ *as in Definition 2.1.1, the replicate raw wavelet periodogram has the following asymptotic properties for any fixed scale* j *and rescaled time* z*, within rescaled replicates* ν*:*

Expectation

$$\mathbf{E}\left[I_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;R}\right] = \beta_j(z,\nu) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}), \tag{42}$$

Variance

$$var\left(I_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;R}\right) = 2\beta_j^2(z,\nu) + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}).$$

Proof. The proof appears in Appendix A.1.1. □

From Proposition 2.1.10, we see that the raw periodogram is asymptotically unbiased for β, 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 2.1.11. We define a *replicate-smoothed estimator* for the rephrased spectral quantity $\beta_j(\frac{k}{T}, \frac{r}{R})$ to be

$$\tilde{I}_{j,k;T}^{r;R} = (2M+1)^{-1} \sum_{s=-M}^{M} I_{j,k;T}^{r+s;R}, \tag{43}$$

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 2.1.12 (**smoothing across replicates).** 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 replicates 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 replicate-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 2.2).

Proposition 2.1.13. *Under the properties of Definition 2.1.1, the replicate-smoothed wavelet periodogram in equation* (43) *has the following asymptotic properties for any fixed scale* j *and rescaled time* z *within rescaled replicates* ν*:*

Expectation

$$\mathbf{E}\left[\tilde{I}_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;R}\right] = \beta_j(z,\nu) + \mathcal{O}(MR^{-1}) + \mathcal{O}(2^j T^{-1}),$$

Variance

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

Proof. Appendix A.1.2 contains the proof which manipulates the amplitude properties across replicates as opposed to those across time in the absence of cross-replicate 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 replicate-smoothed periodogram proposed above is an asymptotically unbiased and consistent estimator for the true β, the relationship between the true spectral quantities β and S suggests a natural way of constructing a well-behaved spectral estimator for the unknown S.

Proposition 2.1.14. *Under the assumptions of Proposition 2.1.13, 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 replicates ν*

$$\hat{S}_j(z,\nu) = \sum_{l=1}^{J} A^{-1}_{j,l}\tilde{I}^{\lfloor \nu R \rfloor;R}_{l,\lfloor zT \rfloor;T}, \tag{44}$$

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. Appendix A.1.3 contains the proof which hinges on the properties of the replicate-smoothed periodogram shown in Proposition 2.1.13 above. □

Remark 2.1.15 (replicate and time smoothing). The results in Proposition 2.1.13 highlight the small sample dependence of the bias and variance of the smoothed periodogram on the number of replicates 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 replicates. While still having a bias–variance trade-off, the variance can be further improved by additionally smoothing across the time-dimension.

Specifically, using a 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 replicate-smoothing window, we define the replicate- and time-smoothed periodogram

$$\tilde{\tilde{I}}^{r;R}_{j,k;T} = (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T} I^{r+s;R}_{j,k+t;T}, \tag{45}$$

to act as an estimator for the transformed spectral quantity $\beta_j(\frac{k}{T}, \frac{r}{R})$.

The time-smoothing window parameter M_T is chosen automatically using the method proposed by Nason (2013). The replicate-smoothing window $(2M+1)$ with a choice of $M = \frac{3}{4}\sqrt{R}$ appears to work well in the following simulation studies. Killick *et al.* (2020)

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

We next show that this estimator has desirable asymptotic properties, leading to faster convergence than its counterpart involving only replicate-smoothing.

Proposition 2.1.16. *For a REv-LSW process as in Definition 2.1.1 and satisfying the additional assumption of autocovariance summability,* $\sup_{z,\nu\in(0,1)} \sum_{n\in\mathbb{Z}} |c(z,\nu;n)| = \mathcal{O}(1)$, *the smoothed replicate- and time-specific wavelet periodogram defined in equation* (45) *has the following asymptotic properties for any fixed scale* j *and rescaled time* z *within rescaled replicates* ν*:*

Expectation

$$\mathbf{E}\left[\tilde{\tilde{I}}^{\lfloor \nu R \rfloor;R}_{j,\lfloor zT \rfloor;T}\right] = \beta_j(z,\nu) + \mathcal{O}(M_T T^{-1}) + \mathcal{O}(MR^{-1}) + \mathcal{O}(2^j T^{-1}),$$

Variance

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

Proof. Appendix A.1.5 contains the proof which makes use of the smoothing in both directions. □

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

$$\hat{\tilde{S}}_j(z,\nu) = \sum_{l=1}^{J} A^{-1}_{j,l} \tilde{\tilde{I}}^{\lfloor \nu R \rfloor;R}_{l,\lfloor zT \rfloor;T}.$$

It is straightforward to show that this is also asymptotically unbiased and consistent for $S_j(z,\nu)$, in the same manner as in the proof of Proposition 2.1.14.

2.2 Simulation study under the assumption of uncorrelated replicates

Here we aim to assess the behaviour of our proposed REv-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 replicates ('LSW'), (ii) our proposed methodology involving localised replicate smoothing ('REv-LSW$_1$'), and (iii) our proposed methodology involving localised replicate and time smoothing ('REv-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 replicates R from 64 to 512 and smoothing windows $(2M + 1)$ from 9 to 25. Overall, based on our findings, we recommend the use of replicate

and time smoothing methodology (REv-LSW$_2$) with a window length of $(2M+1)$ guided by the choice of $M = \frac{3}{4}\sqrt{R}$ as a rule of thumb.

To quantify the performance of the model, we report the mean squared errors (MSE) and squared bias results, and note these also implicitly infer the variance across simulations. These measures are calculated as the average over all time-scale points and replicates as follows

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

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

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

Remark 2.2.1 (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.

Simulation 1

We simulate $N = 100$ realisations of a REv-LSW process consisting of $R = 128$ replicates, each of length $T = 256 = 2^8$ and whose wavelet spectrum, illustrated in Figure 10, evolves slowly over both rescaled time and replicates, as follows

$$S_j(z,\nu) = \begin{cases} 4\nu \sin^2\left(2\pi z(1+2\nu)\right), & \text{for } j = J(T) - 4, z \in (0,1), \nu \in (0,1) \\ 0, & \text{otherwise,} \end{cases} \tag{46}$$

recalling that $z = k/T$ and $\nu = r/R$ for $k = 0, \ldots, T-1$ and $r = 0, \ldots, R-1$. The spectral characteristics thus appear at scale $j = 8 - 4 = 4$.

The periodicity and magnitude of the sine wave evolve slowly over the replicates 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. One concatenated realisation of the meta-process with the specified spectral structure in equation (46), viewed as a series of length RT, can be seen in Figure 11. Note however that this is an abuse of representation, since each replicate is a time series of its own, and the sole purpose of this visualisation is to highlight the evolution of the meta-process.

We display in Figure 12 the true spectra and the average spectral estimates for replicates 20, 64 and 108. The non-decimated wavelet transform was computed using discrete wavelets built by means of Daubechies Least Asymmetric family with 10 vanishing moments and the local averaging for our REv-LSW$_1$ method was carried out using $M = 4$, corresponding to a window of 9 replicates (numerical MSE results in Table 1 highlight that we chose to visually present some of our least performant results). Figure 12 displays the danger of neglecting the possibility of an existing evolutionary behaviour over replicates (see e.g. level

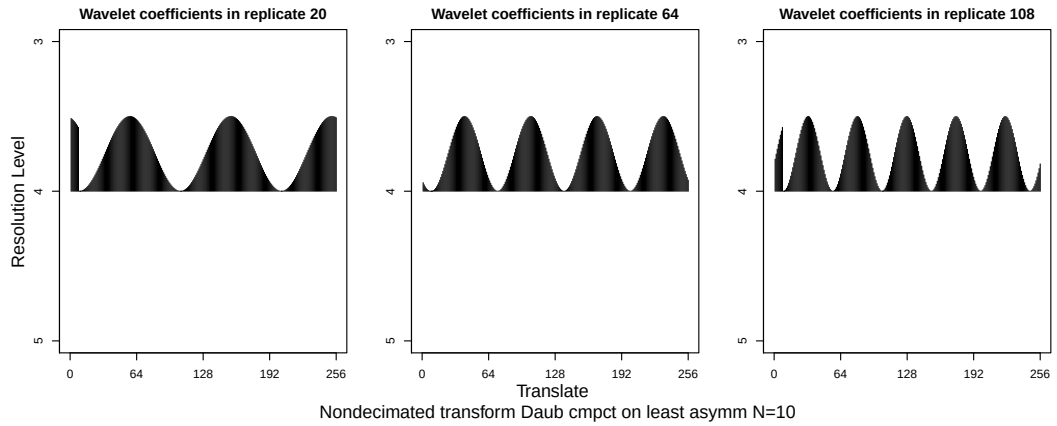


Figure 10: True wavelet spectra for replicates (trials) 20, 64 and 108 of Simulation 1.

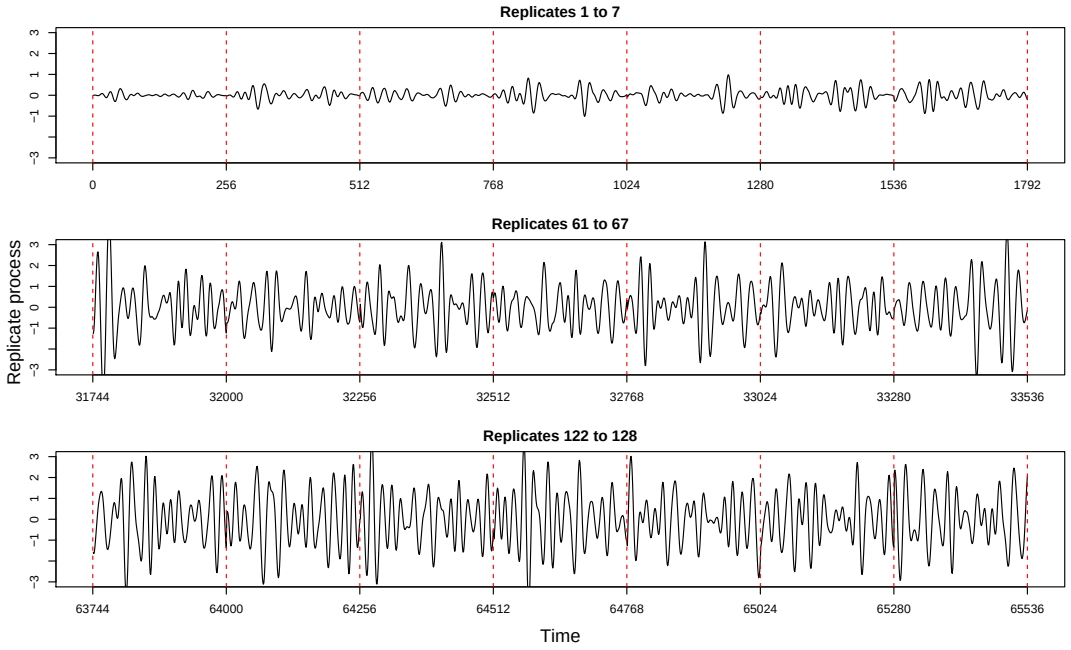


Figure 11: Realisation of a REv-LSW process with spectra defined in equation (46) for Simulation 1.

4 in the top row plots of the true spectrum), conducive to either under or over-estimation (see the middle row plots). The bottom row plots show that the REv-LSW$_1$ estimates do reflect the process evolution over replicates. To further support this, Figure 13 takes a closer look at the evolutionary behaviour of the spectral quantities over the time and replicates in level 4.

The MSE and squared bias results (recalling that we do not assess replicates at the edges) in Table 1, highlight that for this example, which adheres well to the REv-LSW assumptions, the MSEs for the LSW model are higher than those computed for the REv-LSW model. Note these results implicitly also provide evidence for the variance across the simulation runs. 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 replicates. The benefit of taking a local smoothing approach over both time and replicates is that it always results in spectral estimates with lower bias and MSE when compared to LSW, although it is worth pointing out that taking a local smoothing approach over replicates only, while yielding lower bias, might increase the MSE for inappropriately

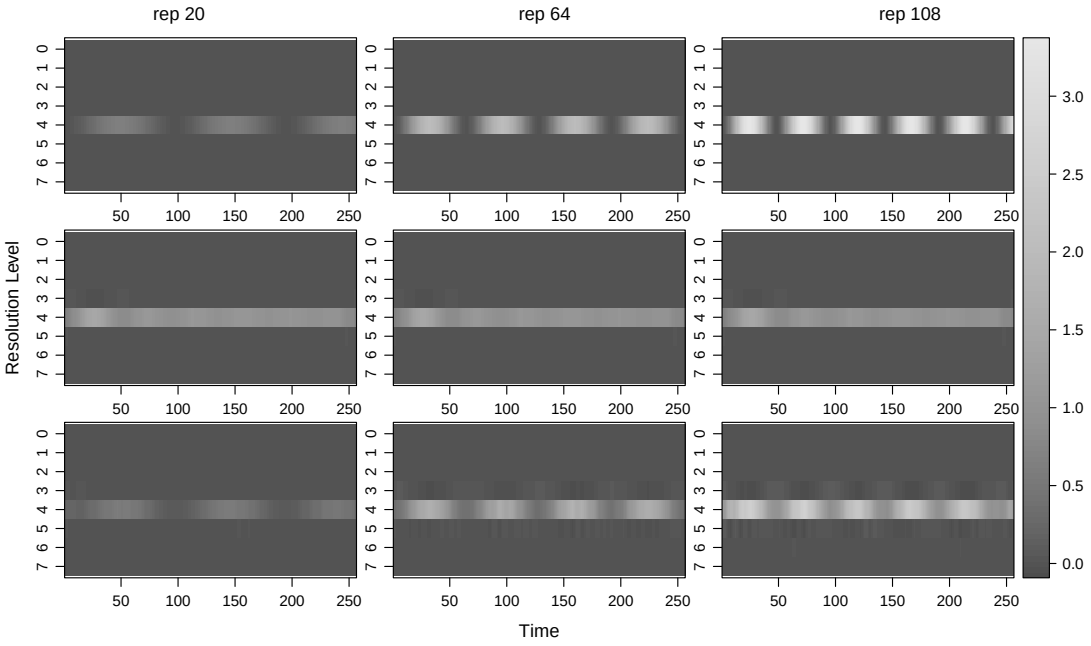


Figure 12: Simulation 1 time-scale plots for replicates (trials) 20, 64 and 108, respectively the first, second and third columns. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimates from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

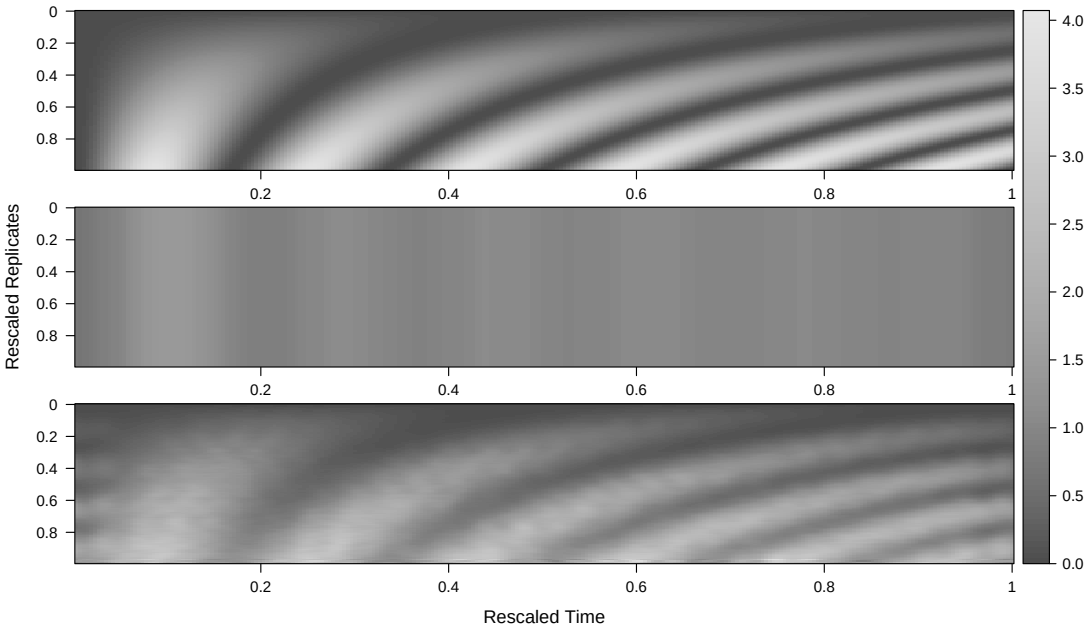


Figure 13: Simulation 1 time-replicate spectral plots in level 4. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimated spectra from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

small windows. Furthermore, we notice the REv-LSW methodology performance improves with increases in the replicate local averaging window length $(2M + 1)$. However, recalling that we require $M/R \to 0$ as $R, M \to \infty$, too large a choice of M could pull information from too many replicates and lead to poorer estimates.

Figure 14 provides a visualisation on how the REv-LSW model performed over the 100 simulations via histograms of the simulation-specific MSE. The histograms 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.

Mean squared errors (×100)

			LSW		REv-LSW$_1$		REv-LSW$_2$	
R	T	M	mse	bias^2	mse	bias^2	mse	bias^2
64	128	4	13.21	12.68	12.92	7.91	11.59	7.97
		7	12.01	11.47	10.74	7.81	9.99	7.87
		10	11.01	10.47	9.88	7.81	9.36	7.87
		12	10.48	9.94	9.51	7.79	9.08	7.84
	256	4	11.10	10.60	7.64	2.79	6.32	2.89
		7	10.06	9.56	6.54	3.71	5.82	3.82
		10	9.21	8.71	6.83	4.84	6.35	4.94
		12	8.77	8.27	7.09	5.43	6.70	5.53
128	128	4	13.67	13.40	13.11	7.92	11.70	7.99
		7	13.02	12.76	10.81	7.76	10.01	7.83
		10	12.40	12.14	9.80	7.67	9.26	7.75
		12	12.00	11.74	9.40	7.64	8.97	7.72
	256	4	11.49	11.25	7.34	2.41	6.01	2.52
		7	10.95	10.70	5.47	2.58	4.74	2.70
		10	10.41	10.17	4.94	2.91	4.47	3.03
		12	10.07	9.82	4.90	3.20	4.52	3.33

Mean squared errors (×100)

			LSW		REv-LSW$_1$		REv-LSW$_2$	
R	T	M	mse	bias^2	mse	bias^2	mse	bias^2
256	512	4	10.16	10.03	5.66	0.39	4.13	0.40
		7	9.91	9.78	3.54	0.40	2.64	0.43
		10	9.67	9.54	2.67	0.46	2.05	0.49
		12	9.51	9.38	2.36	0.52	1.86	0.55
	1024	4	9.11	8.98	5.28	0.12	3.77	0.11
		7	8.88	8.75	3.17	0.12	2.28	0.12
		10	8.66	8.53	2.30	0.15	1.68	0.15
		12	8.52	8.39	1.99	0.19	1.47	0.20
512	512	4	10.25	10.18	5.72	0.37	4.17	0.39
		7	10.12	10.05	3.55	0.36	2.64	0.38
		10	9.99	9.93	2.62	0.36	1.99	0.39
		12	9.91	9.84	2.26	0.37	1.73	0.39
	1024	4	9.18	9.12	5.29	0.11	3.77	0.10
		7	9.06	9.00	3.19	0.09	2.28	0.09
		10	8.95	8.89	2.29	0.09	1.65	0.09
		12	8.87	8.81	1.93	0.09	1.40	0.09

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

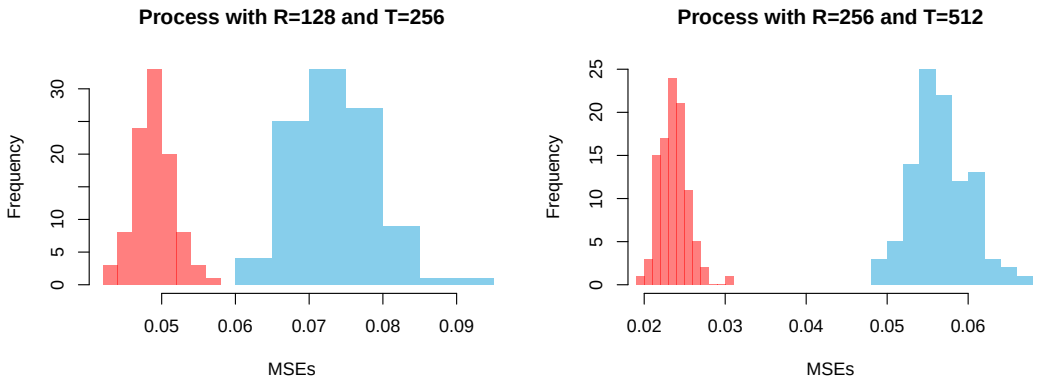


Figure 14: Histograms of the MSEs on the estimates from the REv-LSW model over 100 runs for Simulation 1. Smoothing over replicates with $M = 12$ (*red*); with $M = 4$ (*blue*).

Simulation 2

We consider $N = 100$ realisations of a REv-LSW process consisting of $R = 128$ replicates, each of length $T = 256 = 2^8$, and driven by the following wavelet spectrum

$$S_j(z,\nu) = \begin{cases} \sin^2(2\pi z + 10\nu), & \text{for } j = J(T) - 1, z \in (0,1), \nu \in (0,1) \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

Our spectra in this example are characterised by a squared sine wave in the finest level that experiences a shift from replicate to replicate. This behaviour is displayed in Figure 15 and then a concatenated realisation of the meta-process appears in Figure 16. Note that visually the meta-process behaviour does not offer any indication of transitioning through replicates, despite this actually happening.

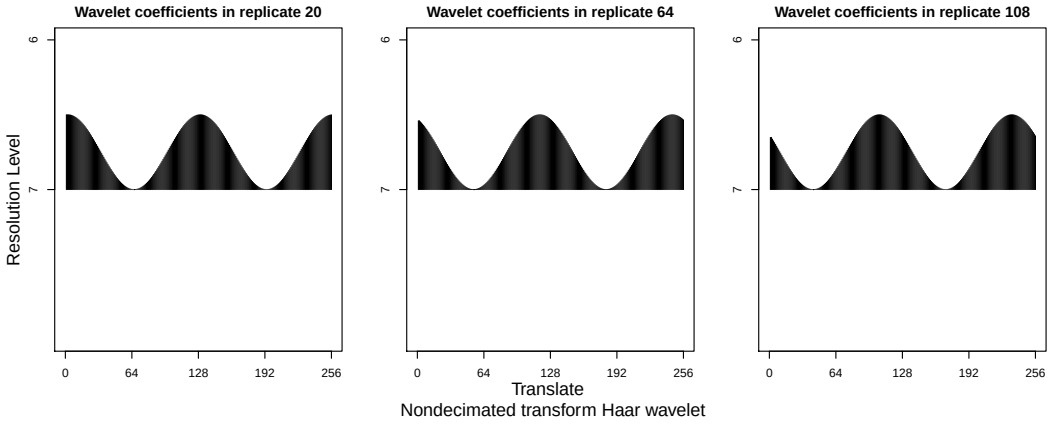


Figure 15: True wavelet spectra for replicates (trials) 20, 64 and 108 of Simulation 2.

We obtain estimates for both the LSW model averaged over all replicates and for the REv-LSW model which adopts the local averaging procedure (over replicate, and over replicate and time dimensions). On visually examining the concatenated process in Figure 16, one may not question the existence of evolutionary behaviour across replicates. Inspection of our REv-LSW estimates of the spectral characteristics of this process tells us otherwise. Figure 17 highlights that the REv-LSW$(_1)$ method (bottom row) manages to capture the evolution of the spectra in the finest level. On the other hand, the LSW

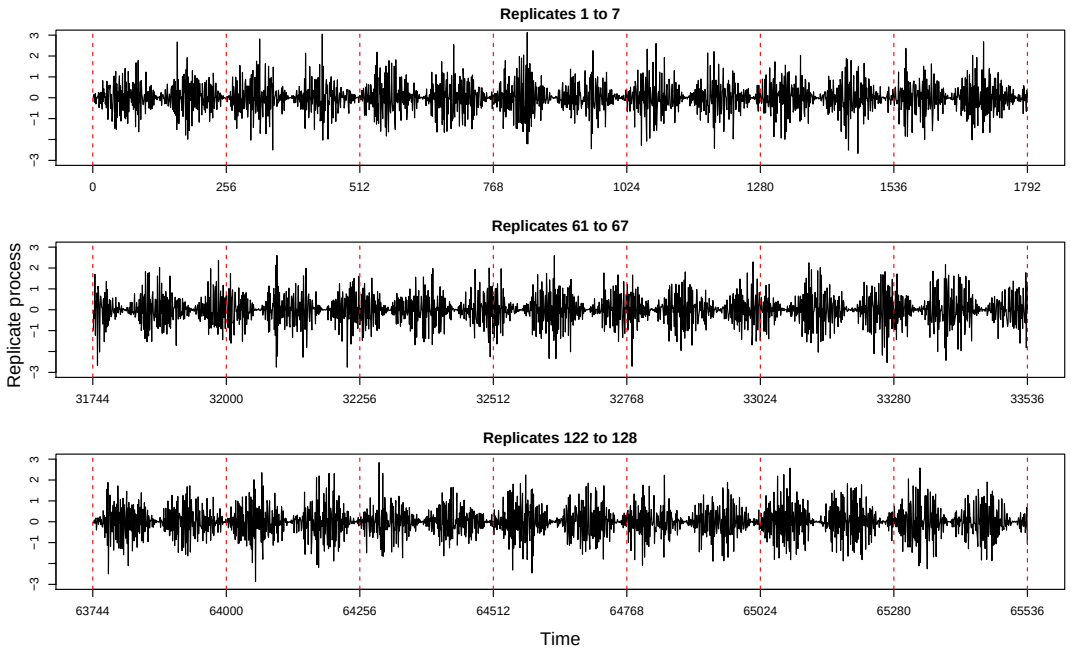


Figure 16: Realisation of a REv-LSW process with spectra defined in equation (47) for Simulation 2.

method (middle row) fails to capture this behaviour, and this is further demonstrated in Figure 18 which shows the spectral estimates across rescaled time and replicate in the finest level.

Mean squared errors and squared bias results for Simulation 2 are given in Table 2. When comparing models, we notice that for low R, performing local averaging over the replicates only (REv-LSW$_1$) appears to yield low bias estimates that nevertheless have poorer MSE results than the blanket LSW involving local time smoothing and then averaging over all replicates (despite its high bias). The small squared bias for the REv-LSW estimates implies that much of the MSE can be attributed to the variance. A possible explanation for this is the spectral leakage across neighbouring scales, a known artefact in the locally stationary spectral estimation context: the narrower choices of window to smooth over replicates in the REv-LSW model were not sufficient enough to remove the effects of the 'leaked' characteristics. These simulations provide a stronger highlight of the impact of the ratio M/R when choosing the replicate window width $(2M+1)$, with the guideline choice of $M = \frac{3}{4}\sqrt{R}$ again appearing to yield competitive results. As R and T increase, and our choice of M becomes larger with larger R (increasing the smoothing window), the MSEs improve and REv-LSW$_2$ estimation performs dramatically better than LSW. This is again in line with our asymptotic results. Also, we note here that the spectra in Simulation 2 do not behave in quite the slowly evolving manner across replicates as detailed in Definition 2.1.1, nevertheless the REv-LSW methodology still performs well at estimating the spectra and capturing the evolutionary behaviour across replicates.

As with Simulation 1, we also provide histograms for the MSEs over 100 simulations involving $R = 128$ replicates of length $T = 256$. Figure 19 highlights how increasing the smoothing window improves the performance of our REv-LSW method and leads to better estimates. In this setting, Table 2 shows how the LSW model performs better in comparison to REv-LSW$(_1)$ when $M = 4$ but as M increased, the REv-LSW$(_1)$ model soon

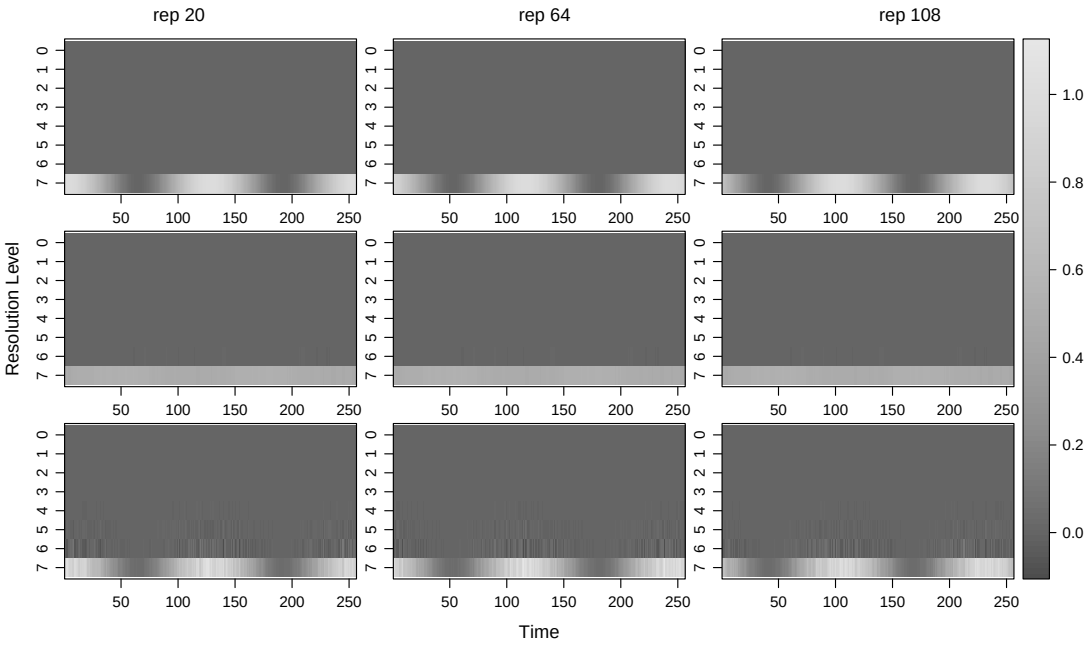


Figure 17: Simulation 2 time-scale plots for replicates (trials) 20, 60 and 108, respectively the first, second and third columns. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimates from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

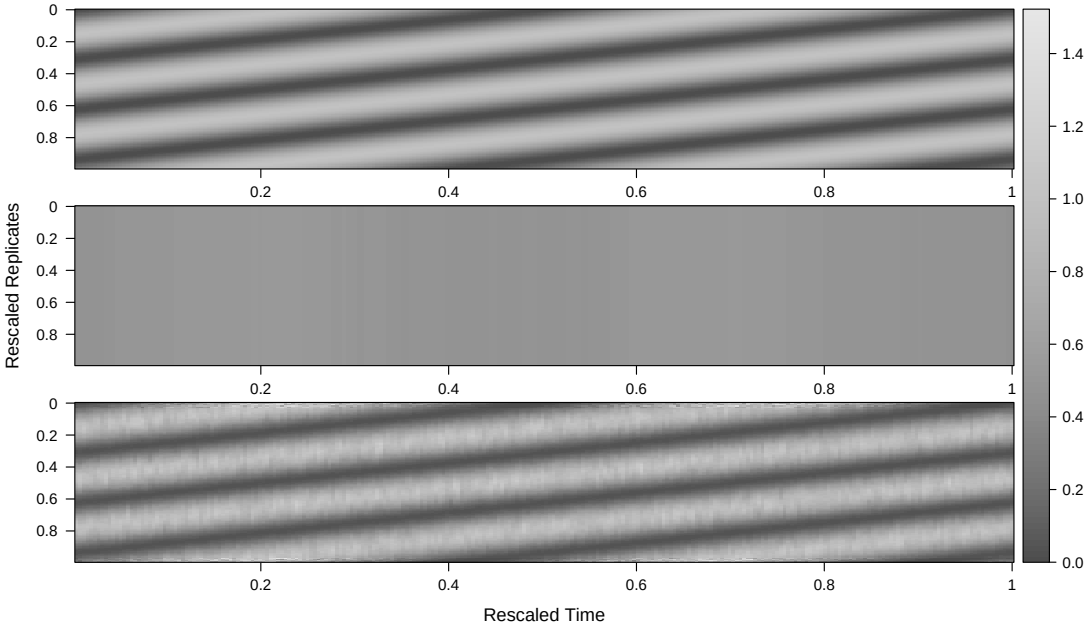


Figure 18: Simulation 2 time-replicate spectral plots in level 7. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimated spectra from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

outperformed LSW. This can also be visualised nicely through the histograms in Figure 20, where we can see the switch in performance of the models and the MSE improvement with the increase of M.

Mean squared errors (×1000)

			LSW		REv-LSW$_1$		REv-LSW$_2$	
R	T	M	mse	bias^2	mse	bias^2	mse	bias^2
64	128	4	19.37	18.09	33.91	1.97	10.69	1.85
		7	19.49	18.21	28.04	8.85	14.10	8.79
		10	19.39	18.11	33.34	19.57	23.24	19.44
		12	19.20	17.92	36.49	24.87	27.88	24.69
	256	4	16.87	15.81	29.63	1.66	9.10	1.47
		7	16.97	15.91	24.48	7.71	12.17	7.60
		10	16.89	15.83	29.11	17.13	20.29	17.03
		12	16.74	15.68	31.86	21.79	24.43	21.69
128	128	4	18.55	17.94	32.12	0.51	9.10	0.37
		7	18.65	18.04	20.06	1.09	6.32	1.06
		10	18.74	18.13	16.51	2.97	6.75	2.99
		12	18.78	18.17	16.50	5.12	8.31	5.15
	256	4	16.23	15.69	28.27	0.39	7.85	0.21
		7	16.32	15.77	17.61	0.90	5.39	0.80
		10	16.39	15.84	14.48	2.54	5.77	2.48
		12	16.43	15.88	14.46	4.42	7.14	4.38

Mean squared errors (×1000)

			LSW		REv-LSW$_1$		REv-LSW$_2$	
R	T	M	mse	bias^2	mse	bias^2	mse	bias^2
256	512	4	14.13	13.89	25.00	0.25	6.85	0.08
		7	14.16	13.93	15.04	0.19	4.15	0.09
		10	14.20	13.97	10.87	0.27	3.10	0.20
		12	14.23	13.99	9.30	0.40	2.78	0.35
	1024	4	12.72	12.50	22.58	0.23	6.15	0.07
		7	12.75	12.54	13.58	0.17	3.73	0.08
		10	12.78	12.57	9.82	0.24	2.79	0.17
		12	12.81	12.59	8.41	0.37	2.50	0.31
512	512	4	13.99	13.87	25.06	0.26	6.84	0.08
		7	14.01	13.89	15.03	0.16	4.10	0.05
		10	14.02	13.90	10.74	0.12	2.94	0.05
		12	14.04	13.91	9.03	0.12	2.49	0.05
	1024	4	12.59	12.48	22.58	0.23	6.14	0.06
		7	12.60	12.50	13.55	0.14	3.69	0.04
		10	12.62	12.51	9.69	0.11	2.64	0.04
		12	12.63	12.52	8.15	0.10	2.23	0.04

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

Simulation 3

For this simulation we present the behaviour of our proposed methodology on a process with a challenging spectral structure, as shown in Figure 21. We generate $N = 100$ realisations

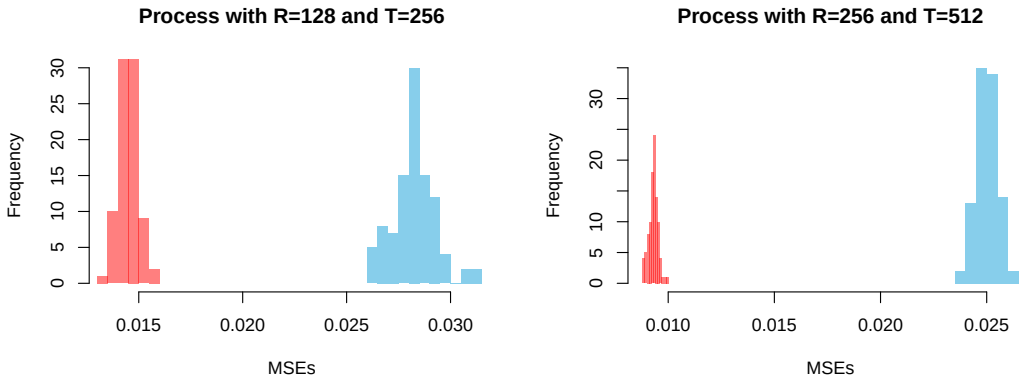


Figure 19: Histograms of the MSEs on the estimates from the REv-LSW model over 100 runs for Simulation 2. Smoothing over replicates with $M = 12$ (*red*); with $M = 4$ (*blue*).

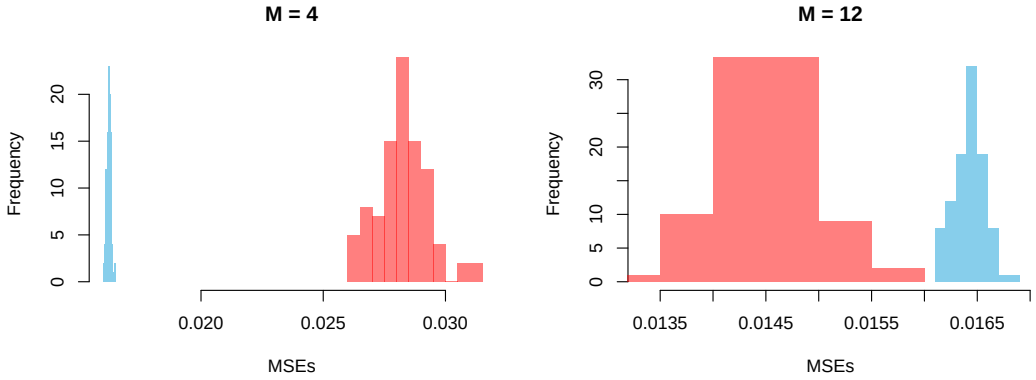


Figure 20: Histograms of the MSEs obtained over 100 runs for Simulation 2 with R = 128 and T = 256. Smoothing over replicates with $M = 4$ (*left*); with $M = 12$ (*right*). *Red:* MSEs of REv-LSW$(_1)$ estimates; *Blue:* MSEs of LSW estimates.

of a REv-LSW process consisting of $R = 256$ replicates, each of length $T = 256 = 2^8$ and with spectral structure at scales 5 and 6, defined as

$$S_j(z,\nu) = \begin{cases} 4(1-\nu)\cos^2\left(\frac{-1}{3}\pi + \frac{4}{3}\pi z\right), & \text{for } j = J(T)-3, z \in (65/256, 1), \nu \in (0,1) \\ 4\cos^2\left((4\pi + 10\nu)z\right), & \text{for } j = J(T)-2, z \in (0, 128/256), \nu \in (0,1) \\ 0, & \text{otherwise,} \end{cases} \tag{48}$$

where $z = k/T$ and $\nu = r/R$ for $k = 0, \ldots, T-1$ and $r = 0, \ldots, R-1$. The process places spectral content at scale $j = 5$, manifest through a decreasing amplitude of the cosine across the last 192 replicates, and at scale $j = 6$, where the periodicity of the cosine increases across the first 128 replicates.

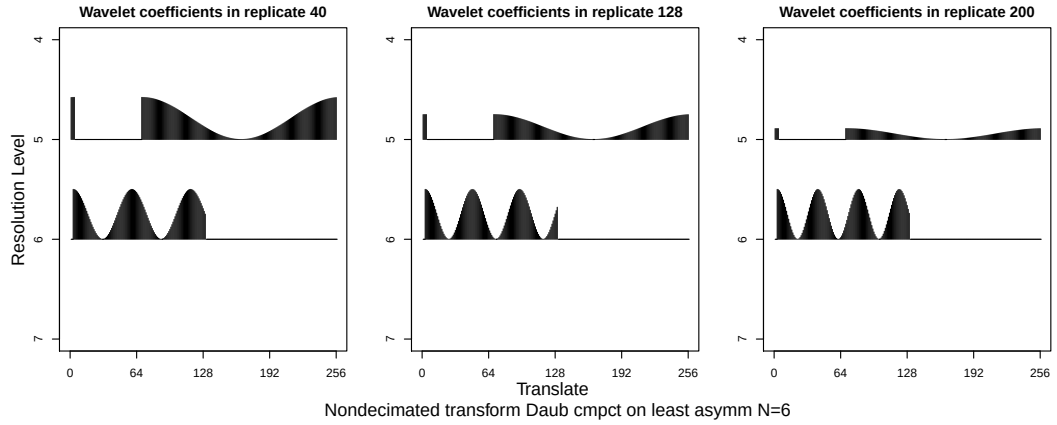


Figure 21: True evolutionary wavelet spectra for replicates (trials) 40, 128 and 200 of Simulation 3.

A concatenated realisation of this process is shown in Figure 22. We note that this process departs somewhat from the requirement that the amplitudes evolve slowly over both rescaled time (z) and replicate (ν) dimensions, however we show that despite this the methodology still performs well.

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 REv-LSW method, local averaging involved windows of 9 replicates corresponding to $M = 4$ and we note that numerical MSE results in Table 3 highlight that we chose to visually present here some of our least performant results. The LSW and REv-LSW($_1$) spectral estimates appear in Figure 23, along with the truth.

From the figures we get a visual clarification that the REv-LSW model is doing a good job at capturing the evolving characteristics of the spectra across replicates 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 REv-LSW$_1$ estimates do indeed pick up the evolution over replicates. Figures 24 and 25 further support the evidence for evolutionary behaviour of the spectral quantities both across time and replicates.

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

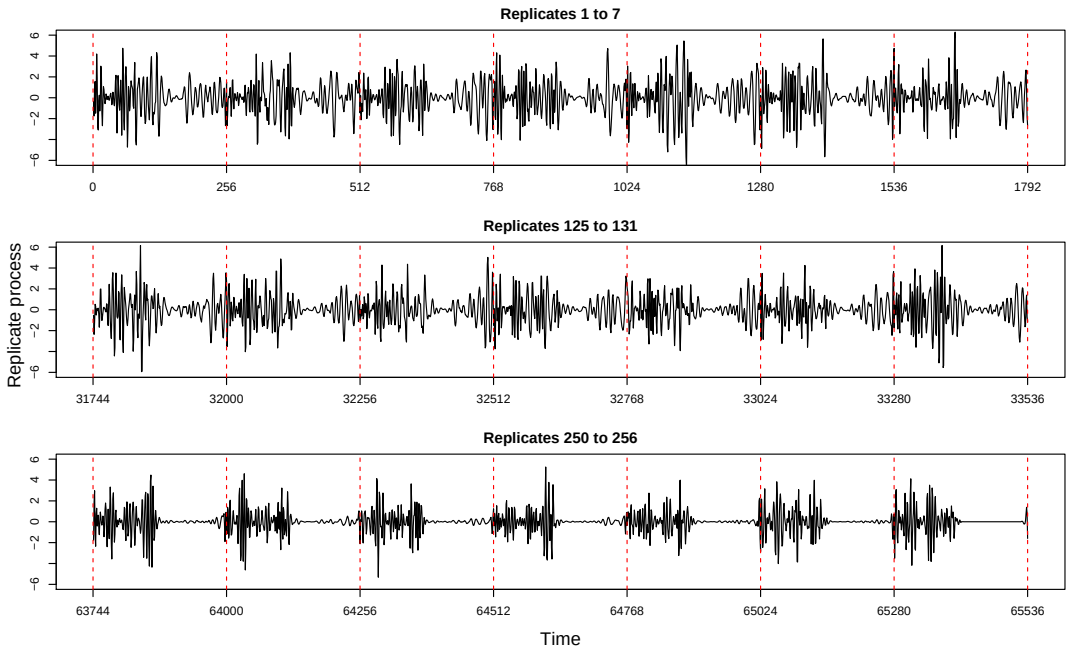


Figure 22: Realisation of a REv-LSW process with spectra defined in equation (48) for Simulation 3.

of our the REv-LSW methodology improves. A replicate window length of $(2M + 1)$ determined by the choice $M = \frac{3}{4}\sqrt{R}$ appears to work well across all our investigations.

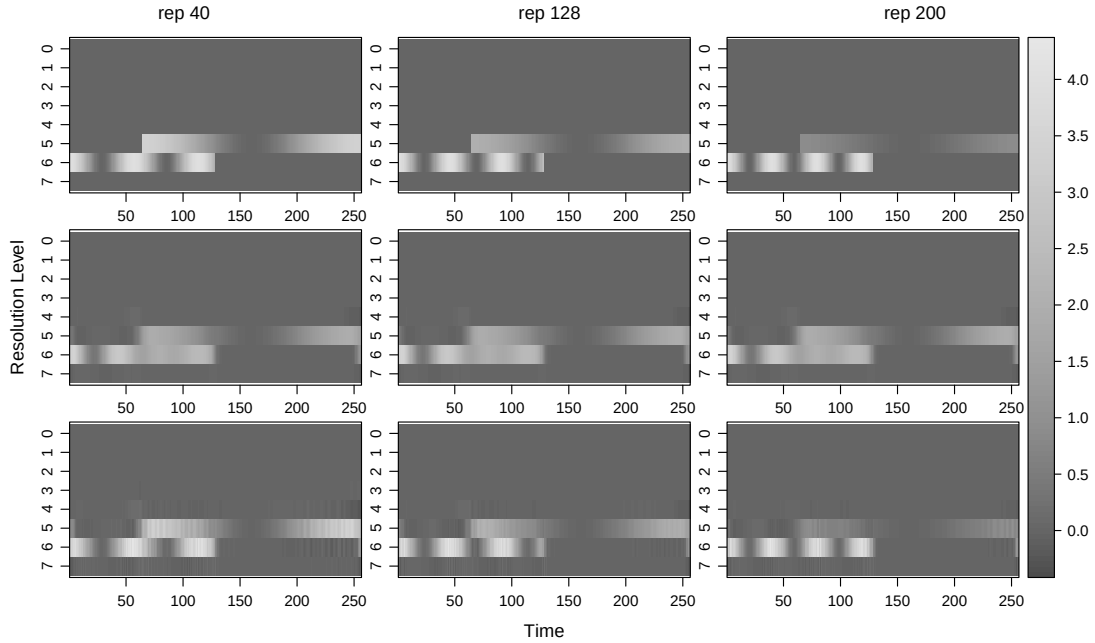


Figure 23: Simulation 3 time-scale plots for replicates (trials) 40, 128 and 200, respectively the first, second and third columns. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimates from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW$(_1)$.

Histograms of the MSEs over the 100 simulations are shown in Figure 26 and again give evidence of the improved performance as R, T and M are increased. To numerically strengthen our visual inference, we examine the MSEs and squared bias results in Table 3 (and thus implicitly the variance in the estimates across the $N = 100$ runs). The best results in terms of lowest MSEs are obtained by REv-LSW$_2$, despite REv-LSW$_2$ incurring

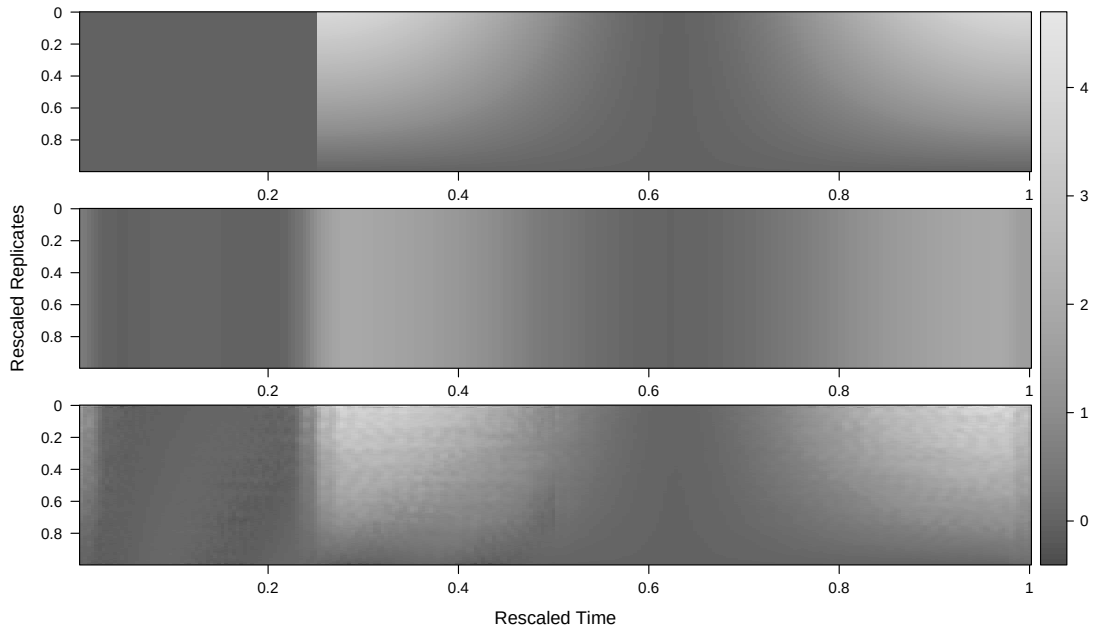


Figure 24: Simulation 3 time-replicate spectral plots for level 5. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimated spectra from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

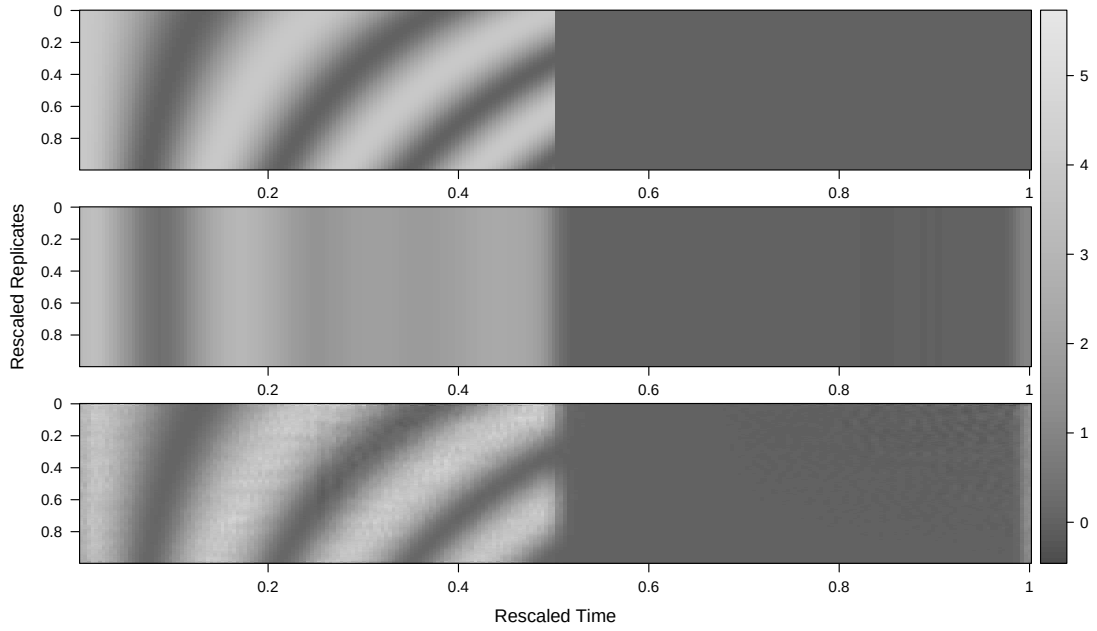


Figure 25: Simulation 3 time-replicate spectral plots for level 6. Estimates are averaged over 100 realisations. *Top*: true spectra; *Middle*: estimated spectra from the LSW method averaged over all replicates; *Bottom*: estimates using REv-LSW($_1$).

a somewhat higher bias than REv-LSW$_1$. Both our methods have a substantially lower bias than the LSW.

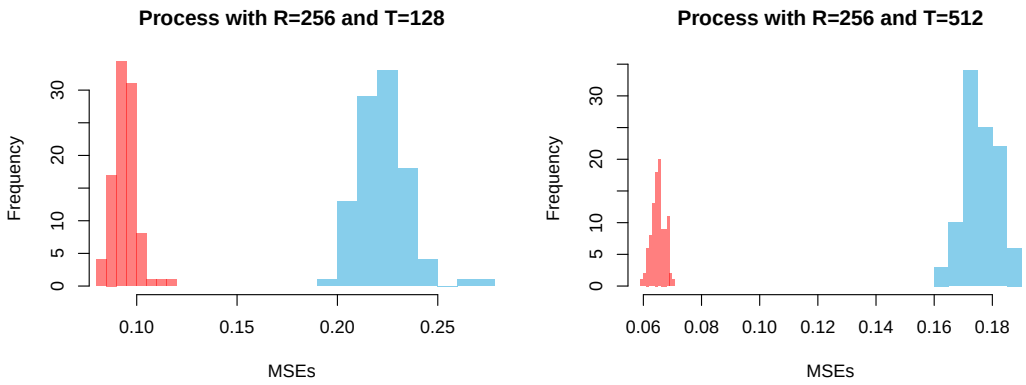


Figure 26: Histograms of the MSEs on the estimates from the REv-LSW model over 100 runs for the simulation 3. *Red:* smoothing over replicates with $M = 12$; *Blue:* smoothing over replicates with $M = 4$.

Mean squared errors (×100)

			LSW		REv-LSW$_1$		REv-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 (×100)

			LSW		REv-LSW$_1$		REv-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

Table 3: MSE and squared bias results, averaged over all time-scale points and replicates for Simulation 3 and $N = 100$ runs. 'LSW' denotes the classical approach of averaging over the replicates. 'REv-LSW$_1$' denotes our proposed approach using localised replicate smoothing. 'REv-LSW$_2$' denotes our proposed approach using localised replicate and time smoothing. Our proposed methods 'REv-LSW$_1$'and 'REv-LSW$_2$' use a replicate smoothing window of length $(2M + 1)$, while the time-smoothing window for 'LSW' and 'REv-LSW$_2$' is automatically chosen.

Remark 2.2.2 (Empirical investigation into departures from normality). In Remark 2.1.9 we pointed out that the estimation results hold under the assumption of Gaussian innovations. An empirical investigation using innovations that follow a Student distribution with 5, 7 and 10 degrees of freedom, demonstrates the robustness of the estimation theory developed in Section 2.1.3. Table 4 below shows that while the heavy tails unsurprisingly impact the quality of estimation, they do not overturn the conclusions upheld under the assumption of Gaussian innovations.

Table 4: MSEs ($\times 100$) averaged over all time-scale point and trials with true spectra as defined in Simulation 3 (Section 2.2). For $N = 100$ runs, simulations are obtained from Student's t-distribution with degrees of freedom 5, 7 and 10. 'LSW' denotes the classical approach of averaging over the replicates. 'REv-LSW$_2$' denotes our proposed approach using localised replicate and time smoothing.

			DF = 5		DF = 7		DF = 10	
R	T	M	LSW	REv-LSW$_2$	LSW	REv-LSW$_2$	LSW	REv-LSW$_2$
128	256	4	19.17	12.68	13.65	11.37	12.34	8.28
		7	16.86	14.54	13.11	8.48	11.80	6.14
		10	16.34	12.82	12.58	7.50	11.27	5.52
		12	16.00	12.34	12.24	7.32	10.93	5.46

2.3 Concluding remarks

In this chapter we proposed a novel wavelet-based methodology that successfully captures nonstationary process characteristics for time series collected across replicates. The desirable properties of the REv-LSW model were evidenced by simulation studies, demonstrating the improved estimation in comparison to the classical LSW approach that would dismiss the experimental timeline and simply average over all the replicates. This work has demonstrated the dangers of approaching replicate time series as identical process realisations and the misleading results this can yield when studying the process dynamics across the replicate domain. Crucial to the proposed statistical model was the assumption of replicate uncorrelation, a limiting assumption we drop in the proposed methodology of the following chapter.

Furthermore, recall that the methodology is developed to analyse ordered replicate time series and thus smoothness between successive replicates is assumed. We point out that the practitioner should exhibit care in their approach and ensure that indeed a natural replicate ordering is present. Should this not be the case, clustering and then averaging within each cluster could provide a preferred, appropriate approach (Hargreaves *et al.*, 2018; Ting *et al.*, 2018). Further work beyond this thesis could investigate when the smoothness assumption between replicates breaks. In this scenario, it may be of benefit to include a measure of 'stochastic closeness' when performing local averaging over the replicates, and if appropriate, weights could be included in the smoothing process.

A next natural step would be to investigate the REv-LSW local (partial) autocorrelation function (for the LSW setup, see Killick *et al.* (2020)). An extension to this work would be to develop the REv-LSW methodology in a multivariate setting (Sanderson *et al.* (2010); Park *et al.* (2014)).

3 Methodology incorporating the potential for replicate coherence

So far, 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-replicate dependence by means of a *between-replicate* coherence structure. This is a major innovation of this work, as to the best of our knowledge this methodology is the first that simultaneously accounts for correlation across replicates while also embedding potential nonstationarity in multiple time scales: locally within a replicate and globally across replicates in the entire experiment. Additionally, the lack of between-replicate dependence can also be naturally incorporated by our proposed framework, as a particular case.

The chapter proceeds as follows. Section 3.1 (i) introduces our proposed meta-process model that allows for correlation between replicates while accounting for intra- and cross-replicate nonstationarity, and (ii) develops its associated estimation theory. The generality of this new model allows us to treat the absence of cross-replicate dependence, a fundamental assumption underpinning Chapter 2, as its particular manifestation. Section 3.2 details simulation studies that showcase the behaviour of the proposed methodology and demonstrates the advantage of accounting for between- replicate behaviour characterisation, as well as across time and within- replicate characteristics. We conclude the chapter in Section 3.3.

3.1 REv-LSW model embedding replicate coherence

Definition 3.1.1. 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 replicate $r = 0, \ldots, R-1$ where $R = 2^{J(R)}$ to be a *replicate-evolving locally stationary wavelet* (REv-LSW) as in Definition 2.1.1 with the properties given there and the following amendments:

2. (*replacing property* 2) Additional to $\{\xi_{j,k}^r\}_{j,k}$ being orthonormal within replicate 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 replicates r and r', at each scale j. Note that the within-replicate orthonormality induces $\left|\rho_{j,k;T}^{r,r';R}\right| \leq 1$ for all j, k and r, r', with equality when $r = r'$. The assumption of uncorrelated replicates amounts to $\rho_{j,k;T}^{r,r';R} = \delta_{r,r'}$ for all j, k.

3. (*replacing property* 3(a) equation (36))

$$\sum_{j=1}^{\infty} \left|\widetilde{W}_j(z,\nu)\widetilde{W}_j(z,\nu')\right| < \infty \text{ uniformly in } z \in (0,1), \nu, \nu' \in (0,1). \tag{49}$$

4. (*additional property*) For each scale $j \geq 1$, there exists a Lipschitz continuous function in rescaled time z and rescaled replicate arguments ν and ν', 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) replicates ν and ν', at scale j. Similarly, denote

by P_j^z the bounded Lipschitz constant corresponding to the replicate 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

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

(b) There exist sequences of bounded replicate-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 across time within a replicate and across replicates, in the sense that

$$\sup_{k=0:T-1} \left| \rho_{j,k;T}^{r,r';R} - \rho_j \left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R} \right) \right| \leq \frac{\tilde{C}_j^{r,r'}}{T}, \quad \forall j, r, r' \tag{50}$$

$$\sup_{r,r'=0:R-1} \left| \rho_{j,k;T}^{r,r';R} - \rho_j \left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R} \right) \right| \leq \frac{\tilde{D}_j^k}{R}, \quad \forall j, k. \tag{51}$$

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 3.1.2 (**rescaled replicate dependence**). Note that from equations (50) and (51) 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 replicates ν and ν', 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 νR, 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) replicates ν and ν', and it is zero for uncorrelated replicates. We remind the reader that previous methodology was developed under the working assumption of uncorrelated replicates, now therefore viewed as a particular case.

3.1.1 Replicate evolutionary wavelet spectrum and coherence

As we do not work directly with the time- and trial-specific multiscale transfer functions $\{\widetilde{W}_j(\cdot, \cdot)\}_j$ (as is common in both Fourier and wavelet-based spectral domain analysis), we proceed (in same vein as Chapter 2) by defining a scale-dependent measure for the time and replicate contribution to the overall process variance. Additionally, this measure now incorporates the potential for a cross-replicate contribution to the overall process variance through the existence of between-replicate dependence.

Next, we develop a novel evolutionary wavelet spectrum capable to extract nonstationarity both in time and within- and between- replicates, a feat the current LSW model quantities are not designed to cope with.

Definition 3.1.3. For a REv-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 3.1.1, we define its replicate evolutionary wavelet spectrum as follows.

1. The *within-replicate evolutionary wavelet spectrum* at scale j, rescaled replicate ν, rescaled within-replicate time z agrees with that previously defined in equation (40), given by
$$S_j(z,\nu) = \left|\widetilde{W}_j(z,\nu)\right|^2.$$

2. The *between-replicate evolutionary wavelet spectrum* defined at scale j, rescaled time z within rescaled replicates ν and ν' 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 replicates $\nu = \nu'$ is the within-replicate 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 replicates (trials) of Fiecas and Ombao (2016), the between-replicate spectrum is zero, i.e. $S_j(z,\nu,\nu') = 0, \forall\, \nu' \neq \nu$.

In the REv-LSW setup, we thus quantify the between-replicate dependence by means of
$$\rho_j(z,\nu,\nu') = \frac{S_j(z,\nu,\nu')}{\{S_j(z,\nu)S_j(z,\nu')\}^{1/2}}, \tag{52}$$
and we shall refer to it as the *locally stationary between-replicate coherence*, with values ranging from -1, indicating an absolute negative correlation, to 1 indicating an absolute positive correlation. Uncorrelation across replicates amounts to $\rho_j(z,\nu,\nu') = 0, \forall\, \nu' \neq \nu$.

In addition to the *within-replicate local covariance* as defined in Definition 2.1.6, below we introduce the local covariance function associated to a REv-LSW process that captures between-replicate dependence. We remind the reader that this quantity is defined for completeness and in order to aid our theoretical developments (and proofs) but we will not extensively pursue this quantity as our focus is on analysis within the spectral domain.

Definition 3.1.4. For a REv-LSW process $\{X_{t;T}^{r;R}\}$ as in Definition 3.1.1, the *between-replicate local covariance* at rescaled time $z \in (0,1)$ within rescaled replicates ν and ν', both in $(0,1)$, at time-lag $\tau \in \mathbb{Z}$ is given by
$$c(z,\nu,\nu';\tau) = \sum_{j=1}^{\infty} S_j(z,\nu,\nu')\Psi_j(\tau).$$

where $\Psi_j(\tau) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{j,k}(\tau)$ denotes the scale j autocorrelation wavelet. Observe that $c(z,\nu,\nu;\tau) = c(z,\nu;\tau), \forall\, \nu$, and under the previous assumption of uncorrelated replicates, $c(z,\nu,\nu';\tau) = 0, \forall\, \nu' \neq \nu$.

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 (τ) and rescaled replicates (ν, ν') for both the limiting amplitudes and the autocorrelation wavelets (see equation (49)).

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 3.1.5. *For a REv-LSW process* $\{X_{t;T}^{r;R}\}$ *with properties as in Definition 3.1.1,*

$$\left|\mathrm{cov}(X_{\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;T}, X_{\lfloor zT\rfloor+\tau;T}^{\lfloor \nu' R\rfloor;T}) - c(z,\nu,\nu';\tau)\right| = \mathcal{O}(T^{-1}) + \mathcal{O}(R^{-1}),$$

uniformly in τ *at (rescaled) time* z *and replicates* ν, ν'.

Proof. The proof appears in Appendix A.3.1 and uses the approximation properties in Definition 3.1.1 of the REv-LSW process. □

3.1.2 Estimation theory

In the same manner as our framework built under the assumption of uncorrelated replicates, our estimation procedure begins by computing the raw wavelet periodogram and then builds upon this through smoothing methods in order to establish a within- and between-replicate spectral estimator with desirable asymptotic properties. We again note that an estimator for the meta-process local autocovariance, both within- and between-replicates, can be constructed by replacing the unknown spectrum in Definition 3.1.4 with a well-behaved spectral estimator. The development and investigation of the local covariance is beyond the current interest of this work and is left for further research, with potential for investigation into the partial covariance (see Killick *et al.* (2020)).

Definition 3.1.6. For a scale j and time k, we define the *raw wavelet periodogram between replicates* r and r' of a REv-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 previously defined *within-replicate raw wavelet periodogram* given by

$$I_{j,k;T}^{r;R} = \left|d_{j,k;T}^{r;R}\right|^2.$$

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 ν and ν' 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 β. Hence we next focus on estimating β and note that the theoretical results below are derived under the Gaussianity assumption (see Remark 2.1.9 in Chapter 2).

Proposition 3.1.7. *For a REv-LSW process* $\{X_{t;T}^{r;R}\}$ *as in Definition 3.1.1, the replicate raw wavelet periodogram has the following asymptotic properties for any fixed scale* j *and rescaled time* z*, within rescaled replicates* ν *and* ν'*:*

Expectation

$$\mathbf{E}\left[I_{j,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}\right] = \beta_j\left(z,\nu,\nu'\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}), \tag{53}$$

Variance

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

Proof. The proof appears in Appendix A.1.1. □

From Proposition 3.1.7 and in similar fashion to their non-coherence counterparts, the raw periodogram is asymptotically unbiased for β, but inconsistent due to its asymptotically non-vanishing variance. We now introduce smoothing to obtain consistency and then correct for bias.

Definition 3.1.8. We define a *replicate-smoothed estimator* for the rephrased spectral quantity $\beta_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})$ to be

$$\tilde{I}_{j,k;T}^{(r,r');R} = (2M+1)^{-1}\sum_{s=-M}^{M} I_{j,k;T}^{(r+s,r'+s);R}, \tag{54}$$

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 $M/R \to 0$.

Proposition 3.1.9. *Under the properties of Definition 3.1.1 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* τ*, the replicate-smoothed wavelet periodogram in equation* (54) *has the following asymptotic properties for any fixed scale* j *and rescaled time* z *within rescaled replicates* ν, ν'*:*

Expectation

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

Variance

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

Proof. Appendix A.1.2 contains the proof which manipulates the amplitude properties across replicates as opposed to those across time in the presence of cross-replicate dependence. □

The bias of the smoothed periodogram becomes asymptotically negligible, while its variance tends to zero for any fixed fine enough scale j (with $2^j = o\left(\min\left\{T, R, (2M+1)^{1/2}\right\}\right)$) as T, R and $M \to \infty$ and $M/R \to 0$. Then correcting for the bias will yield a desirable spectral estimator, as follows.

Proposition 3.1.10. *Under the assumptions of Proposition 3.1.9, 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 replicates* ν, ν'

$$\hat{S}_j(z,\nu,\nu') = \sum_{l=1}^{J} A_{j,l}^{-1}\tilde{I}_{l,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}, \tag{55}$$

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. Appendix A.1.3 contains the proof which hinges on the properties of the replicate-smoothed periodogram shown in Proposition 3.1.9 above. □

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

$$\hat{\rho}_j(z,\nu,\nu') = \frac{\hat{S}_j(z,\nu,\nu')}{\left\{\hat{S}_j(z,\nu)\hat{S}_j(z,\nu')\right\}^{1/2}}, \tag{56}$$

where the involved spectral quantities are consistently estimated as proposed in Proposition 3.1.10 and the use of the same smoothing windows guarantees that the values of the resulting coherence estimator are indeed quantities between -1 and 1.

Proof. See proof in Appendix A.3.2. □

The following proposition shows that the step of examining $\hat{\rho}_j(z,\nu,\nu')$ is theoretically justified.

Proposition 3.1.11. *Under the assumptions of Proposition 3.1.9, the coherence estimator in* (56) *is asymptotically consistent for the true coherence* $\rho_j(z,\nu,\nu')$ *at each fixed scale* j *and rescaled time* z *within rescaled replicates* ν*,* ν'*.*

Proof. See Appendix A.1.4. □

Remark 3.1.12 (replicate and time smoothing). As discussed in Remark 2.1.15 for our modelling with uncorrelated replicates, the addition of smoothing across the time dimension can improve the variance results given in Proposition 3.1.9 leading to faster convergence. Specifically, for a replicate-smoothing window of length $(2M+1)$, and a 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)), we define the replicate- and time-smoothed periodogram

$$\tilde{\tilde{I}}^{(r,r');R}_{j,k;T} = (2M+1)^{-1}(2M_T+1)^{-1} \sum_{s=-M}^{M} \sum_{t=-M_T}^{M_T} I^{(r+s,r'+s);R}_{j,k+t;T} \tag{57}$$

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 replicate-smoothing.

Proposition 3.1.13. *For a REv-LSW process as in Definition 3.1.1 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 replicate- and time-specific wavelet periodogram defined in equation* (57) *has the following asymptotic properties for any fixed scale* j *and rescaled time* z *within rescaled replicates* ν*,* ν'*:*

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

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

Proof. Appendix A.1.5 contains the proof which makes use of the smoothing in both directions. □

The replicate- and time- smoothed periodogram can then be used to further build a well-behaved estimator of the unknown replicate 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 3.1.10.

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 for $\hat{\rho}$, averaged over all time-scale points and replicates, as a measure of accuracy and in this context are defined 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 2.2, we also adopt here $J = J(T)$. Note these measures also allow us to infer the variance across simulations. We display the behaviour of our estimators on the following simulated examples.

Simulation 1

We simulate $N = 100$ realisations of a REv-LSW process with $R = 256$ replicates that feature dependence, measured at $T = 512 = 2^9$ time points. The within-replicate wavelet spectra are defined by a sine wave whose periodicity and magnitude evolve slowly over the replicates 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 2.2 of Chapter 2). Here we have $J(T) = 9$ and the spectral characteristics are placed in level $j = J(T) - 4 = 5$. In

addition to these characteristics, we also define the between-replicate spectral structure by means of defining their (true) coherence at each level j and location k. For level $j = 5$, we choose a coherence of 0.7 between all replicates over the first 256 locations and zero (no) coherence over the last 256 locations. All other levels have no coherence between replicates. Visual representations of this dependence structure appear in Figure 27 (left panels). For $j = 5$ and $k = 1, \ldots, 256$, the non-zero coherence matrices are defined as follows

$$\left(\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})\right)_{r,r'} = \begin{bmatrix} \rho_{5,k}^{1,1} & \rho_{5,k}^{1,2} & \cdots & \rho_{5,k}^{1,R} \\ \rho_{5,k}^{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{5,k}^{R-1,R} \\ \rho_{5,k}^{R,1} & \cdots & \rho_{5,k}^{R,R-1} & \rho_{5,k}^{R,R} \end{bmatrix} = \begin{bmatrix} 1 & 0.7 & \cdots & 0.7 \\ 0.7 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0.7 \\ 0.7 & \cdots & 0.7 & 1 \end{bmatrix}$$

where we have (abusively) let $\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}) = \rho_{j,k}^{r,r'}$ to ease notation.

Using the spectral estimation methodology proposed in Section 3.1.2, we obtain coherence estimates visually represented in Figure 27 (right panels) for replicate 50 (top row) and replicate 200 (bottom row). The non-decimated wavelet transform was computed using discrete wavelets built by means of Daubechies least asymmetric family with 10 vanishing moments and a window of 9 replicates ($M = 4$) was chosen for the local replicate-smoothing.

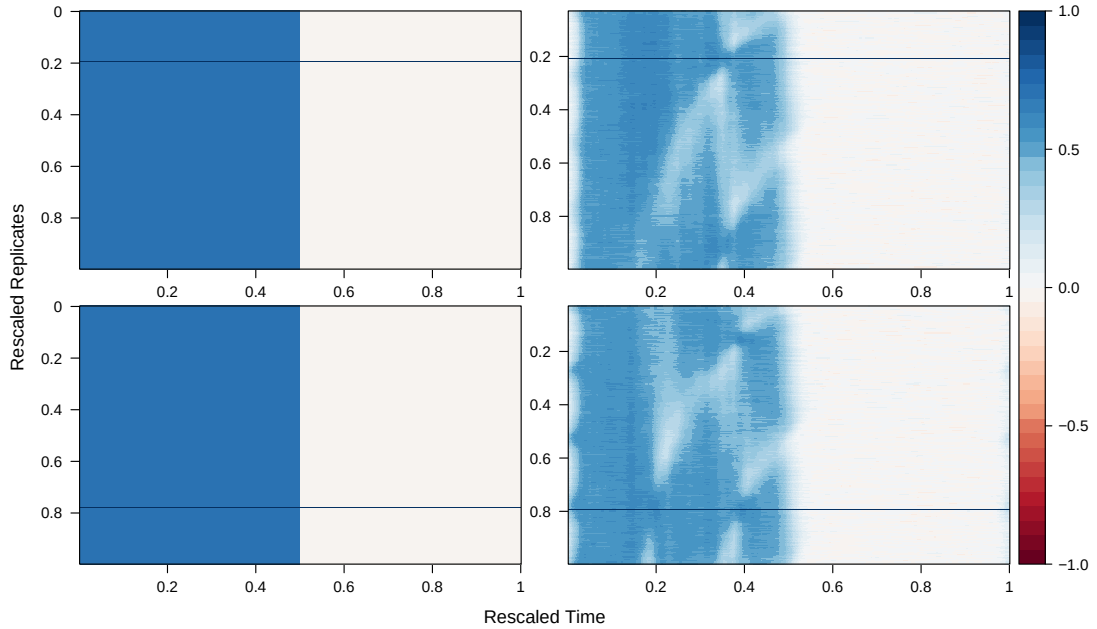


Figure 27: Simulation 1 coherence plots for replicates 50 (*top row*) and 200 (*bottom row*) over rescaled time and replicates in level 5. *Left*: true coherence; *Right*: coherence estimates averaged over 100 simulations.

It is apparent that the coherence structure is being picked up by the modelling framework in Section 3.1.2 in terms of the locations and the positiveness of the defined true coherence. We do however note that the intensity of the estimated coherence is not quite as strong as the true coherence. Table 5 reports the numerical MSE and squared bias results for two smoothing approaches: the first involves smoothing only over a window of replicates; the second involves local averaging through replicates and time. These results highlight that our correction procedure that aims to ensure positive spectral estimates has the undesired effect of increasing the MSEs and introducing bias as R and T increase, just as

reported in a bivariate coherence estimation framework by Sanderson *et al.* (2010). The improved performance via replicate- and time-smoothing is consistently shown throughout the results, with further decreases in the MSE when the smoothing window is increased from 15 replicates ($M = 7$) to 25 replicates ($M = 12$).

R	T	M	mse_1	$\text{bias}^2{}_1$	mse_2	$\text{bias}^2{}_2$
128	256	7	17.29	7.41	15.16	9.66
		12	15.06	7.42	14.04	9.66
128	512	7	17.64	8.11	16.01	10.66
		12	15.29	8.01	14.73	10.52
256	512	7	18.06	8.27	16.34	10.85
		12	15.85	8.18	15.18	10.75

Table 5: Simulation 1: MSE and squared bias ($\times 100$), averaged over all time-scale points and replicates and $N = 100$ runs. Subscripts 1 and 2 denote the models with smoothing over replicates only and replicate-time smoothing, respectively.

Simulation 2

For $N = 100$ runs, we simulate a REv-LSW process with $R = 256$ replicates that feature dependence, measured at $T = 512 = 2^9$ time points with $J(T) = 9$. The within-replicate locally stationary wavelet spectra are as defined in Simulation 1 above. In addition to the within-replicate spectral characteristics, we also define a challenging between-replicate spectral structure by means of defining their (true) coherence at each level j and location k. For level $j = J(T) - 4 = 5$ and time $k = 1, \ldots, 256$ we define the non-zero replicate coherence matrices as follows: the first 128 replicates have a strong positive coherence (0.99) with one another, however this coherence becomes negative (-0.71) with the last 128 replicates. A (weaker) positive coherence (0.5) also exists between the last 128 replicates. Over the last 256 locations, we set the coherence to be zero. The expressions of the non-zero coherence matrices are given as

$$\left(\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})\right)_{r,r'} = \begin{bmatrix} \rho_{5,k}^{1,1} & \cdots & \rho_{5,k}^{1,128} & \rho_{5,k}^{1,129} & \cdots & \rho_{5,k}^{1,R} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho_{5,k}^{128,1} & \cdots & \rho_{5,k}^{128,128} & \vdots & \cdots & \rho_{5,k}^{128,R} \\ \rho_{5,k}^{129,1} & \cdots & \cdots & \rho_{5,k}^{129,129} & \cdots & \rho_{5,k}^{129,R} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho_{5,k}^{R,1} & \cdots & \rho_{5,k}^{R,128} & \rho_{5,k}^{R,129} & \cdots & \rho_{5,k}^{R,R} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.99 & \cdots & 0.99 & -0.71 & \cdots & \cdots & -0.71 \\ 0.99 & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0.99 & \vdots & \ddots & \ddots & \vdots \\ 0.99 & \cdots & 0.99 & 1 & -0.71 & \cdots & \cdots & -0.71 \\ -0.71 & \cdots & \cdots & -0.71 & 1 & 0.5 & \cdots & 0.5 \\ \vdots & \ddots & \ddots & \vdots & 0.5 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & 0.5 \\ -0.71 & \cdots & \cdots & -0.71 & 0.5 & \cdots & 0.5 & 1 \end{bmatrix},$$

where we have (abusively) let $\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}) = \rho_{j,k}^{r,r'}$ to ease notation. The illustrative true coherence structures for replicates 50 (top row) and 200 (bottom row) can be visualised in Figure 28 (left panels).

Coherence estimates obtained using the methodology proposed in Section 3.1.2 are represented in Figure 28 (right panels) for replicates 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 replicates ($M = 4$) were employed.

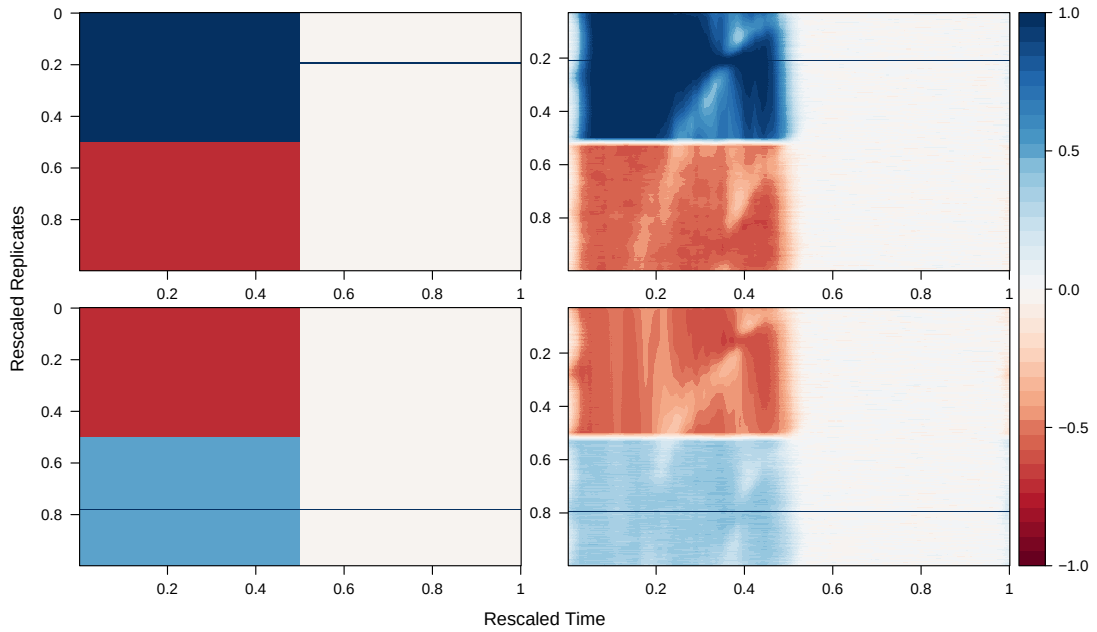


Figure 28: Simulation 2 coherence plots for replicates 50 (*top row*) and 200 (*bottom row*) over rescaled time and replicates in level 5. *Left*: true coherence; *Right*: coherence estimates averaged over 100 simulations.

In terms of correctly estimating the coherence structure switch over times and replicates, 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 Remark 3.2.1). 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 28).

As with Simulation 1, we report Simulation 2 numerical MSE and squared bias results in Table 6. For both simulations, the results paint the same picture. As we increase the replicate smoothing window (such that $M/R \to 0$) the performance of our models improves in terms of MSEs, and the double smoothing over replicate and time further reduces the MSEs. The price to pay for double smoothing as usual is a slightly higher bias (than when using averaging over replicates 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.

R	T	M	mse_1	$\text{bias}^2{}_1$	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

Table 6: Simulation 2: MSE and squared bias (×100), averaged over all time-scale points and replicates and $N = 100$ runs. Subscripts 1 and 2 denote the models with smoothing over replicates only and replicate-time smoothing, respectively.

Simulation 3

For this final simulation, we consider a coherence structure closer to the experimental data we will analyse in Chapter 4. We simulate for $N = 100$ runs, a REv-LSW process with $R = 256$ replicates that feature dependence, measured at $T = 512 = 2^9$ time points with $J(T) = 9$. The within-replicate locally stationary wavelet spectra are as defined in Simulation 1 above. At each level j and location k, we define the true coherence structure to exhibit a local between-replicate dependence across a moving window of neighbouring replicates. Specifically, for level $j = J(T) - 4 = 5$ and all time points $k = 1, \ldots, 256$ we define the non-zero replicate coherence matrices as follows. For replicates $r = 31, ..., (R-30)$, over a moving window of 61 replicates, the immediate 10 neighbouring replicates either side of replicate r have a moderate positive coherence (0.8) and the following 20 (40 total) neighbouring replicates will have a mild positive coherence (0.4). For all other replicates we set the coherence to be zero. For replicates towards the edge, $r < 31$ or $r > (R - 30)$, the window of replicates with the above defined coherence decreases correspondingly to the domain edge being approached. An expression for the non-zero coherence matrices are given as follows.

For replicates $r = 31, ..., (R - 30)$, the rows and columns of the non-zero coherence matrices can be expressed as

$$\begin{aligned}
&\left(\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})\right)_{r,r'} \\
&= \begin{bmatrix} \rho_{5,k}^{r,1} & \cdots & \rho_{5,k}^{r,r-30} & \cdots & \rho_{5,k}^{r,r-10} & \cdots & \rho_{5,k}^{r,r} & \cdots & \rho_{5,k}^{r,r+10} & \cdots & \rho_{5,k}^{r,r+30} & \cdots & \rho_{5,k}^{r,R} \end{bmatrix} \\
&= \begin{bmatrix} 0 & \cdots & 0 & 0.4 & \cdots & 0.4 & 0.8 & \cdots & 0.8 & 1 & 0.8 & \cdots & 0.8 & 0.4 & \cdots & 0.4 & 0 & \cdots & 0 \end{bmatrix}
\end{aligned}$$

where we have (abusively) let $\rho_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}) = \rho_{j,k}^{r,r'}$ to ease notation. Figure 29 (left panels) illustrates the true coherence structures for replicates 50 (top row) and 200 (bottom row).

Figure 29 (right panels) illustrate the coherence estimates obtained via the REv-LSW model using discrete wavelets built from Daubechies least asymmetric family with 10 vanishing moments for the non-decimated wavelet transform. Local replicate-smoothing was computed over a window of 9 replicates ($M = 4$). It is clear to see that the neighbouring between-replicate dependence has been captured by the estimation, with a noticeable distinction between both the mild (0.4) and moderate (0.8) positive coherence values. However, as noticed within the estimates for Simulation 2 (at rescaled replicates 0.5 in

Figure 28), there is a slight loss of coherence structure within the estimates, which occurs naturally as a result of the local replicate-smoothing.

In Table 7 we report the numerical MSEs and squared bias results for Simulation 3. Our numerical results are much more favourable than our previous results for Simulations 1 and 2, with MSEs falling faster as both R and $T \to \infty$. Note that this is likely due to the much reduced coherence structure (previous simulations had between-replicate dependence defined for all replicates), although we do not investigate this here. Estimation via replicate- and time-smoothing again demonstrates improved performance alongside increasing the replicate smoothing window $(2M + 1)$ for $M = 7, 12$.

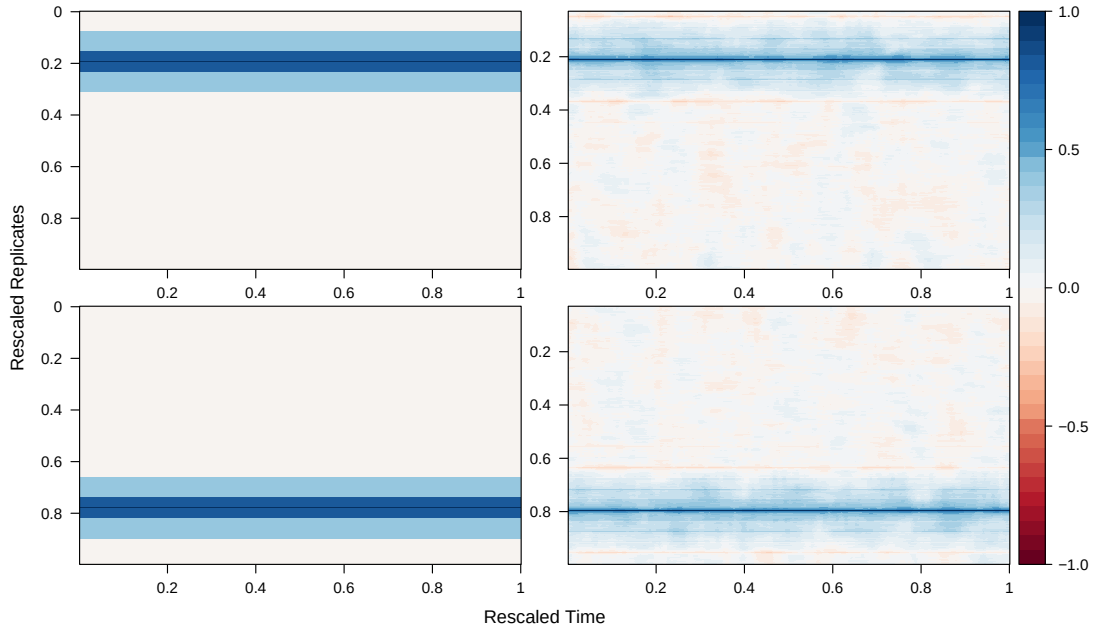


Figure 29: Simulation 3 coherence plots for replicates 50 (*top row*) and 200 (*bottom row*) over rescaled time and replicates in level 5. *Left*: true coherence; *Right*: coherence estimates averaged over 100 simulations.

R	T	M	mse_1	$\text{bias}^2{}_1$	mse_2	$\text{bias}^2{}_2$
128	256	7	19.34	3.48	13.77	4.08
		12	15.45	3.92	11.58	4.53
128	512	7	18.99	3.46	13.63	4.09
		12	15.12	3.87	11.44	4.51
256	512	7	18.42	1.85	12.39	2.15
		12	14.39	1.97	9.95	2.27

Table 7: Simulation 3: MSE and squared bias ($\times 100$), averaged over all time-scale points and replicates and $N = 100$ runs. Subscripts 1 and 2 denote the models with smoothing over replicates only and replicate-time smoothing, respectively.

Remark 3.2.1 (practical implementation). From the theoretical model construction, the within-replicate 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 replicate-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 3.1.2. Another avenue to ensure that spectral estimates are positive, would be to truncate the correction matrix A^{-1} at zero, although this approach does introduce bias. 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 as a further step, to be done after smoothing through time. However, we do not pursue this approach here and leave the further numerical treatment for future research.

3.3 Concluding remarks

This chapter extends the work of Chapter 2 through developing our proposed novel wavelet-based methodology to allow for between-replicate dependence. Through simulation studies, we have demonstrated that the estimation theory for the locally stationary between-replicate coherence yields desirable results. This work strengthens the message that approaching replicate time series as identical process realisations (e.g. classical LSW) can lead to misleading results when studying the process dynamics across the replicate domain. Notably, the proposed statistical model and associated estimation theory encompass meta-processes that feature dependence between replicates, and the assumption of replicate uncorrelation, if appropriate, is dealt with by the proposed methodology as a particular case. A next natural step would be to investigate the REv-LSW local (partial) autocorrelation function (for the LSW setup, see Killick *et al.* (2020)). A further extension would be to develop the REv-LSW methodology in a multivariate setting (Sanderson *et al.* (2010); Park *et al.* (2014)) and to additionally define and investigate the variate replicate-coherence, as seen for the bivariate case of the hippocampus and nucleus accumbens data in Fiecas and Ombao (2016) and Gorrostieta *et al.* (2011).

4 Analysis of Macaque Local Field Potentials

We now perform our REv-LSW spectral analysis on the dataset of local field potentials (LFPs) (recorded at approximately 1000Hz) 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 (dataset courtesy of Prof. H. Ombao, see Declaration). 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; Gorrostieta *et al.*, 2011; Fiecas and Ombao, 2016; Abela *et al.*, 2015; Seger and Cincotta, 2006). Our analysis offers not only confirmations to the results of previous studies, but also provides additional insights to the understanding of the dynamics of the LFPs through 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 embed potential between-trial dependence, and (ii) its superior time-localisation, as we will see next.

4.1 Experimental data description and overview of implemented methodology

Each trial (replicate) consists of $T = 2048$ time points, corresponding to approximately 2 seconds of data. The design of the experiment rigorously splits each trial into four time blocks of 512 milliseconds each, ensuring that the timeline matched from trial to trial, as follows. For the first block the macaque fixated on a screen; a picture (one of four) was then presented on the screen for the next time block; this was followed by an empty screen for the next interval; for the last 512 milliseconds the macaque was presented with a picture of four doors, one of which associated with the picture visual from the second time block. The macaque's task was to select the correct door using a joystick. Correct and incorrect selections were signified via a visual cue and a juice reward was given each time a correct selection was made. The macaque had to learn the associations through repeated trials. The data has been grouped into sets of 'correct' and 'incorrect' responses, in order to investigate how the contributions of the Hc and NAc to the learning process differ between groups (Gorrostieta *et al.*, 2011). The groups containing the correct and incorrect responses consist of 241 and 264 trials, respectively. We carry out the analysis on $R = 256$ trials (replicates). To ensure comparability of the wavelet spectra across trials, each trial is standardised to have mean zero and unit variance. Concatenated plots of the LFPs for the correct responses appear in Figures 30 and 31, and for the incorrect responses in Figures 32 and 33 below.

For both the correct and incorrect sets of the hippocampus (Hc) trial data, we compute the wavelet periodograms using non-decimated discrete wavelets built by means of Daubechies Least Asymmetric wavelet family with 10 vanishing moments. Similarly, for both sets of the nucleus accumbens (NAc) trial data we again choose Daubechies Least Asymmetric wavelet family, but however we now opt for a coarser choice of wavelet with 6 vanishing moments to reflect the behaviour of the signal. In accordance with our

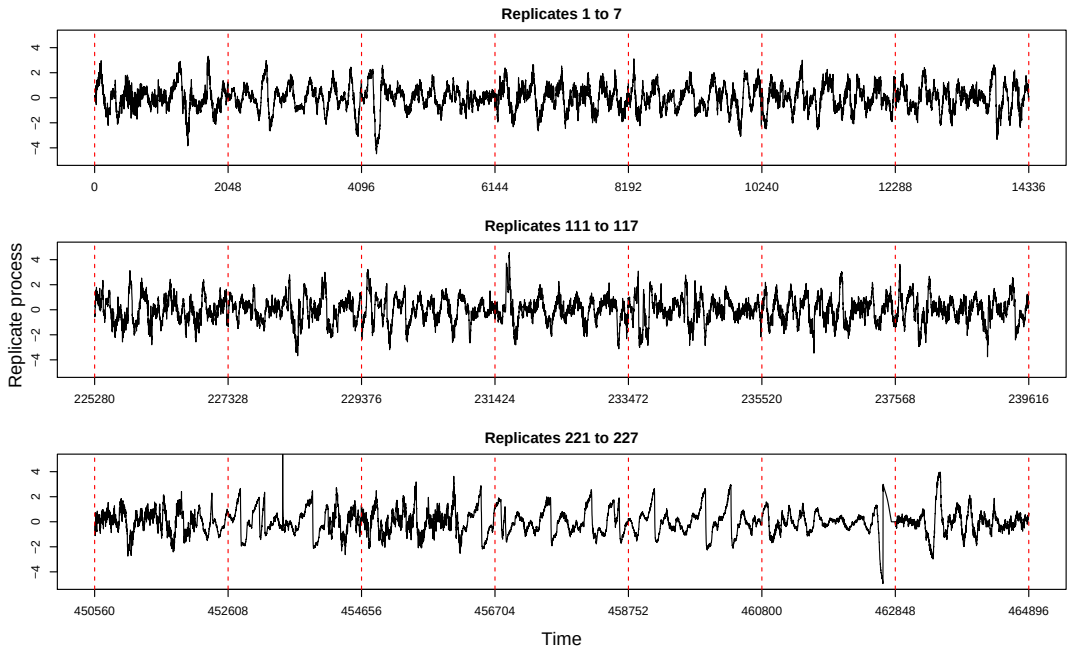


Figure 30: Concatenated series of the hippocampus (Hc) data in the correct response trials (replicates).

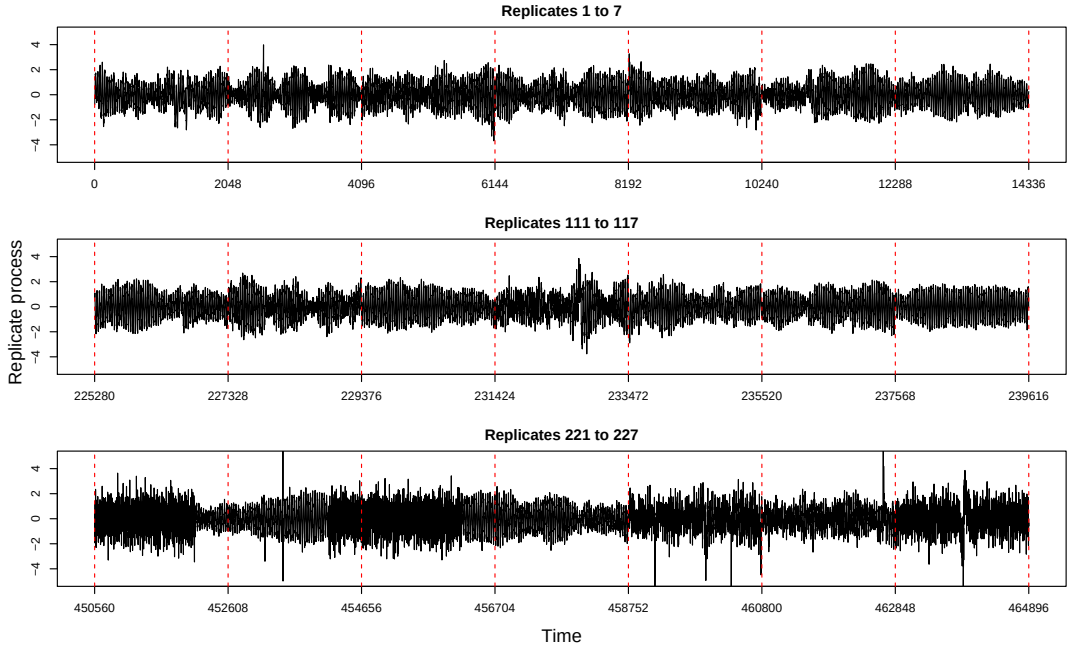


Figure 31: Concatenated series of the nucleus accumbens (NAc) data in the correct response trials (replicates).

simulation study findings, to obtain an asymptotically unbiased and consistent estimator for the trial evolutionary wavelet spectrum, we smooth the wavelet periodograms using a local averaging window over 21 trials ($M = 10$ neighbouring trials) and then correct for bias. For completeness, we also run the analysis to include a time-smoothing step before locally averaging across trials, as this was shown to lead to better performance (see the simulation study in Section 2.2). Confidence intervals using $B = 100$ bootstraps are also reported. To obtain a dyadic number of replicates necessary for estimation, here, 256,

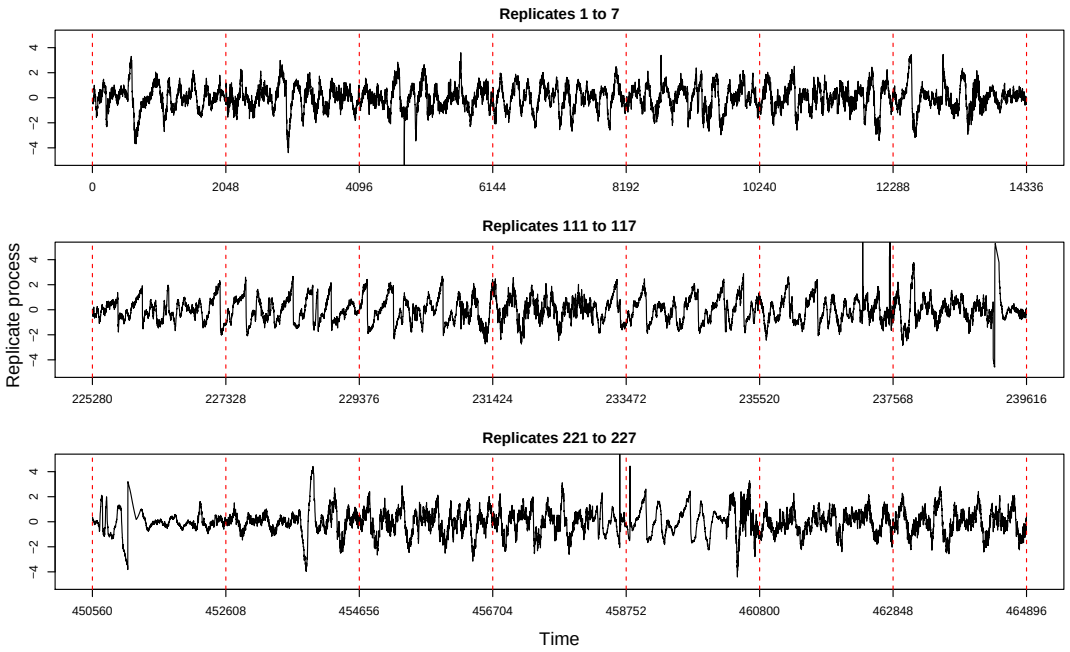


Figure 32: Concatenated series of the hippocampus (Hc) data in the incorrect response trials (replicates).

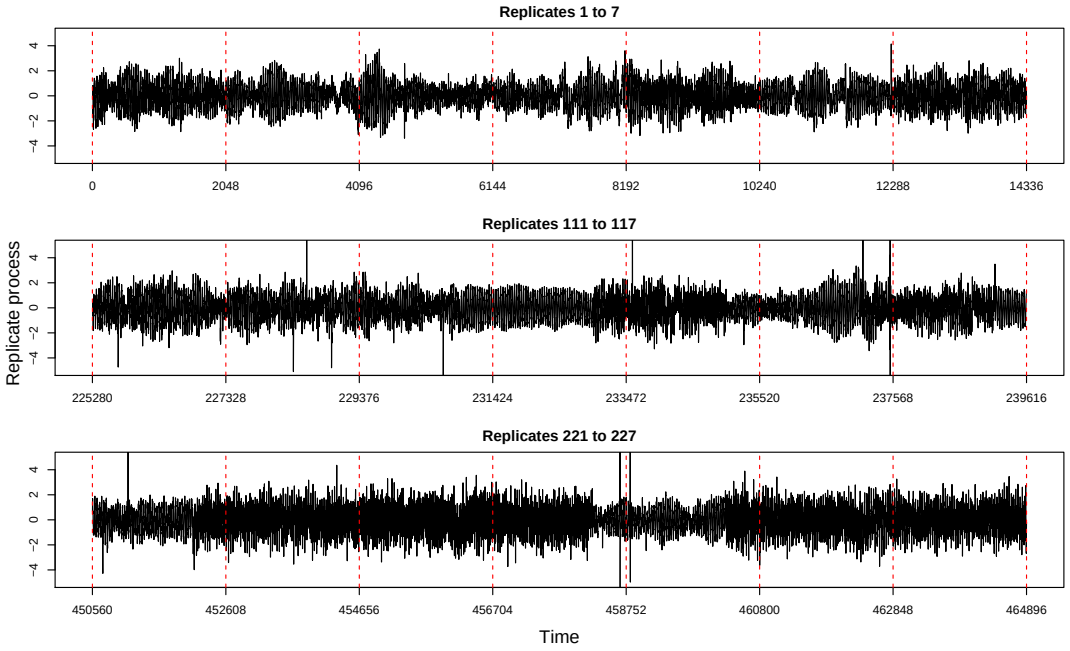


Figure 33: Concatenated series of the nucleus accumbens (NAc) data in the incorrect response trials (replicates).

for the correct response groups, we mirror the last 15 trials. This is for computational purposes only, and we naturally discard the corresponding estimates from our discussions and plots. For comparison, we additionally report the LSW estimator embedding averaging over all replicates. (Note that averaging over all replicates here refers to the averaging over the first 241 and 256 correct and incorrect response trials, respectively.)

In order to explore the consistency of the results, the analysis has also been repeated using wavelets with different vanishing moments and varying smoothing windows across

the replicates, which yielded extremely similar results to those reported here.

In the following remarks, we provide evidence that the Gaussian assumption is tenable for both the correct and incorrect trials of the macaque data, and detail the sampling procedure to construct confidence intervals used to validate our results.

Remark 4.1.1 (**evidence for normality of macaque data).** We now 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 34 and 35, 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 holds, with only some slight departures for NAc records.

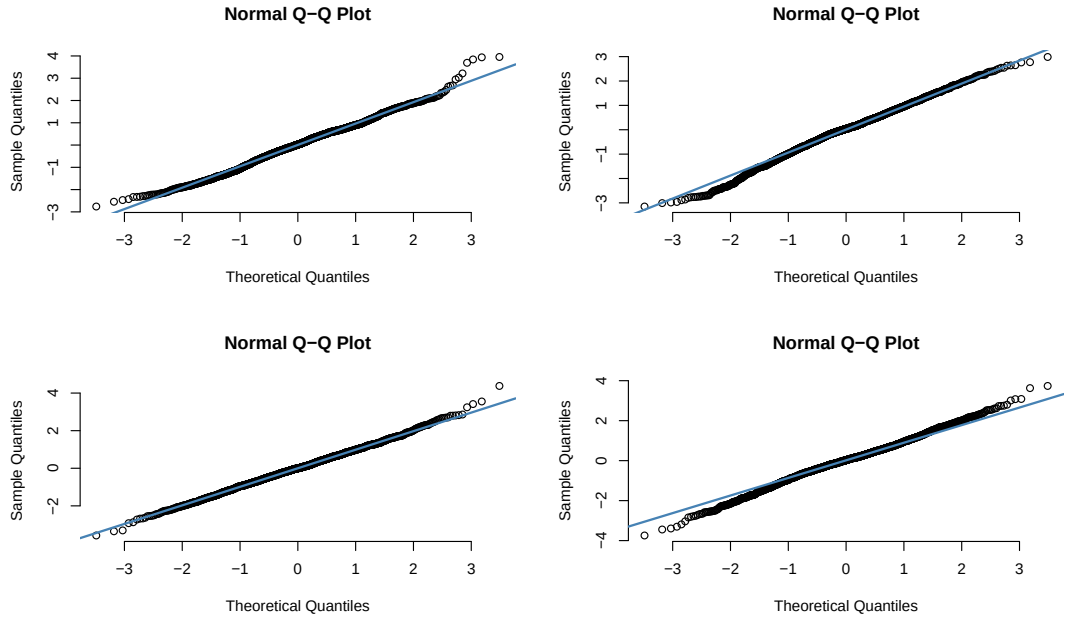


Figure 34: Q-Q plots of the Hc (*top row*) and NAc (*bottom row*) data in the correct trials (replicates). *Left column*: replicate 121; *Right column*: replicate 231.

Remark 4.1.2 (**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. As also argued by Morris (2015); Chau and von Sachs (2016) (although in a functional context), to avoid subsequent inference using erroneous confidence bands it is crucial to allow for potential cross-trial dependence. Thus, to construct pointwise confidence intervals, we use the REv-LSW process Definition 3.1.1 that embeds cross-trial dependence. We take the square root of our REv-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 (40)) for each trial r, within-trial

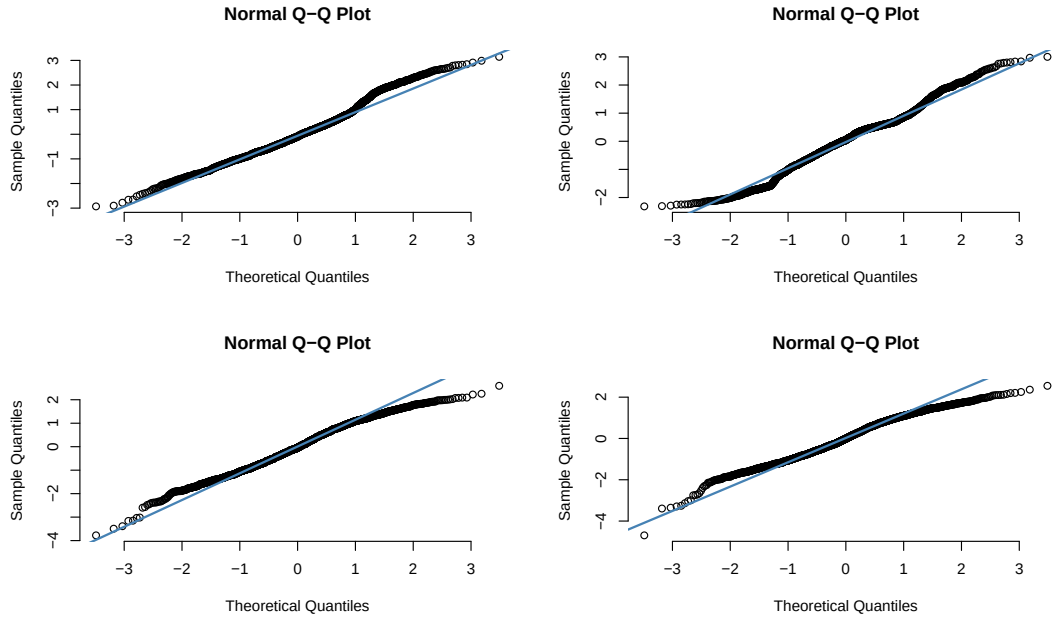


Figure 35: Q-Q plots of the Hc (*top row*) and NAc (*bottom row*) data in the incorrect trials (replicates). *Left column*: replicate 11; *Right column*: replicate 231.

time k and scale j. Furthermore, to allow for cross-trial dependence (see property 2 of Definition 3.1.1) 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 REv-LSW process $\{X_{t;T}^{(b);r;R}\}$ following Definition 3.1.1

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

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. Recall that under the assumption of uncorrelated trials, the dependence structure amounts to $\rho_{j,k;T}^{r,r';R} = \delta_{r,r'}$ for all j, k.

For each bootstrap $b = 1, \ldots, B$, we obtain a bootstrapped REv-LSW$(_2)$ estimate, $\{\hat{\hat{S}}_j^{(b)}(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R})\}$, 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. 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.

4.2 Capturing the within- and between- trial evolutionary dynamics

In the following results, we demonstrate the versatility of the REv-LSW model in capturing time- and replicate-localisations of evolving spectral characteristics within the correct and

incorrect trials of the hippocampus (Hc) and nucleus accumbens (NAc) throughout the associative learning experiment.

4.2.1 Results for the hippocampus

The proposed spectral estimates for the correct and incorrect sets of trials appear below in Figures 36 and 37, respectively. The bottom row plots show the REv-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 Figures 38 and 39, 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 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 REv-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 REv-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).

For the correct Hc trials, the REv-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.

Although we cannot compare the correct and incorrect trials like-for-like, we are still able to see evidence of evolutionary behaviour across the incorrect trials. As the experiment progresses, there is evidence of less spectral activity in the incorrect trials, with a brief Hc

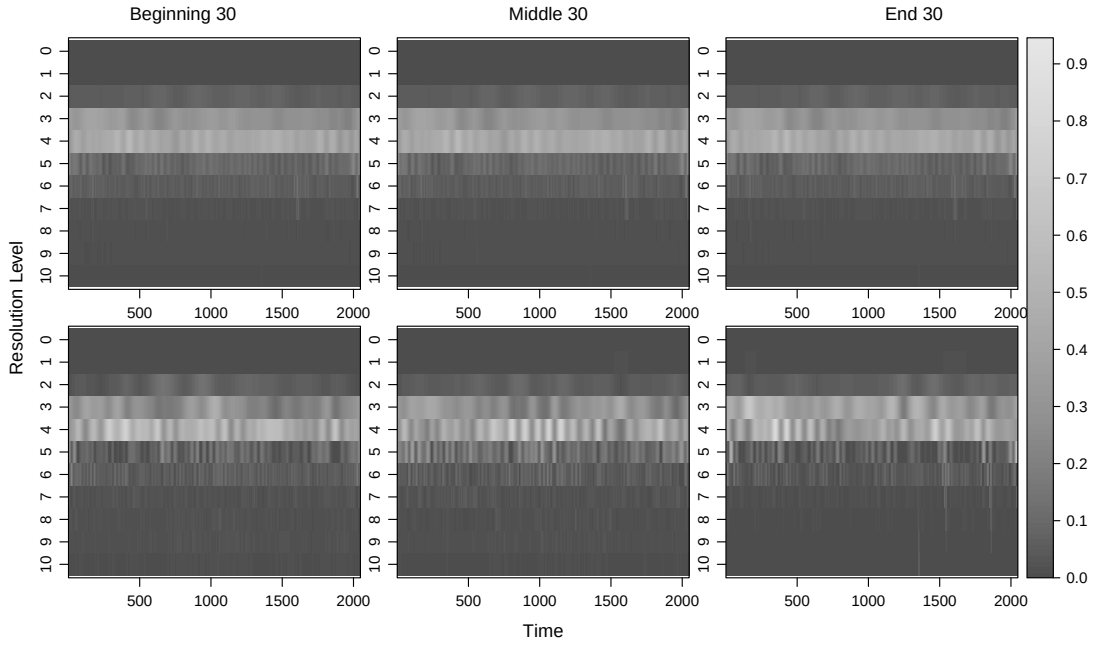


Figure 36: Time-scale hippocampus (Hc) plots for the correct trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: REv-LSW method with smoothing over time and replicates.

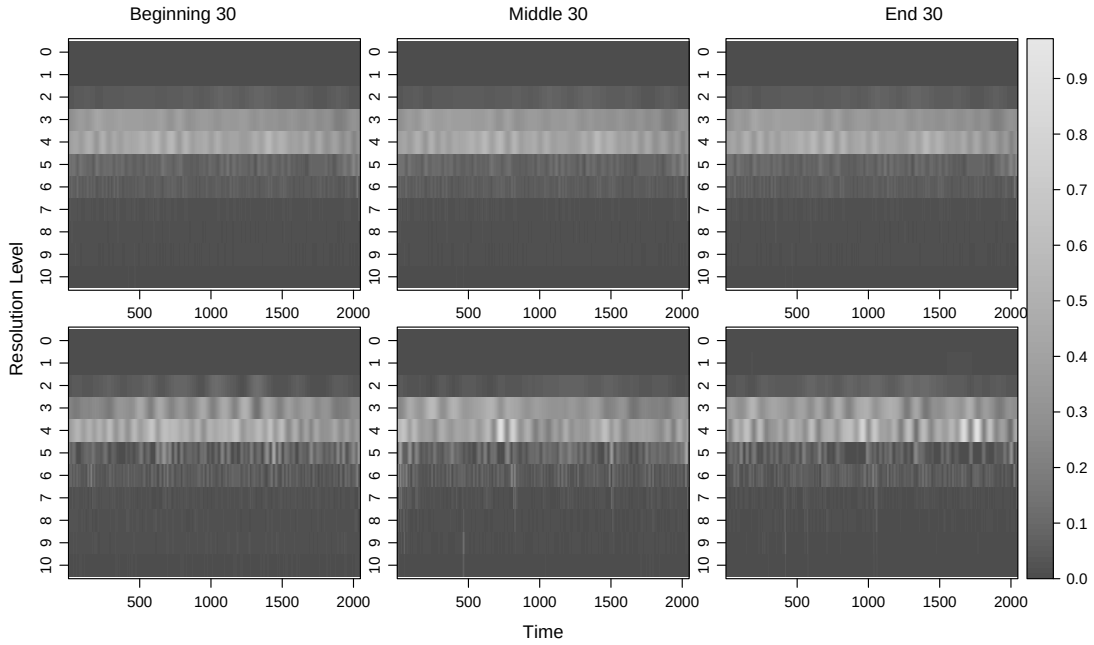


Figure 37: Time-scale hippocampus (Hc) plots for the incorrect trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: REv-LSW method with smoothing over time and replicates.

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

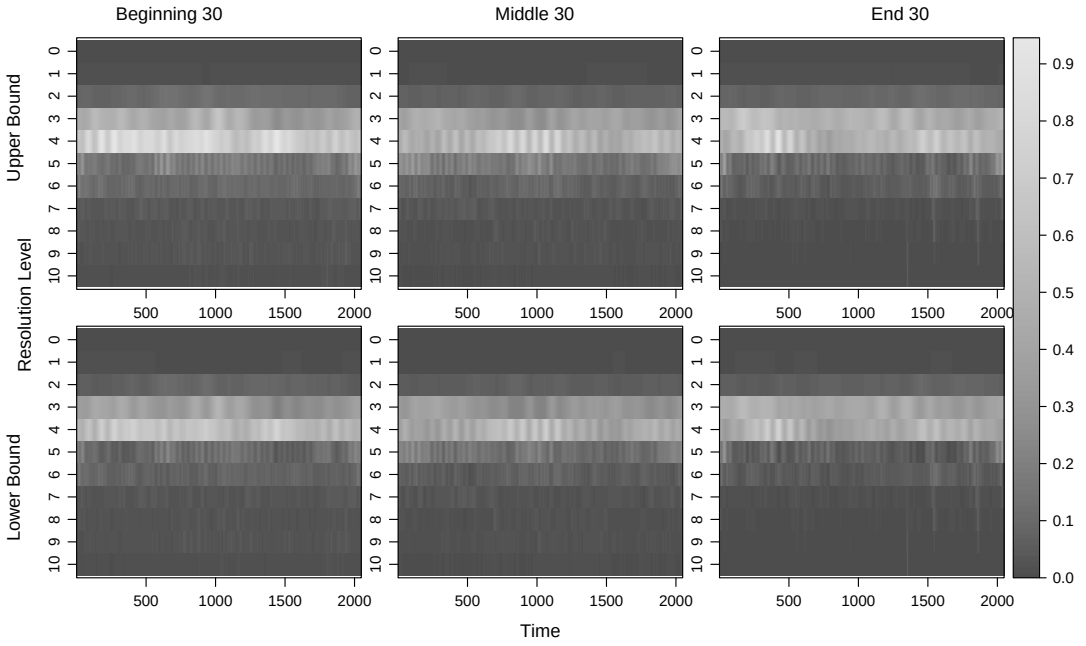


Figure 38: Time-scale 95% bootstrap confidence bounds for the hippocampus (Hc) spectrum for the correct trials. Spectral estimate bounds are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: upper bound; *Bottom*: lower bound. Estimation is via the REv-LSW method with smoothing over time and replicates, and takes into account across-trial correlation.

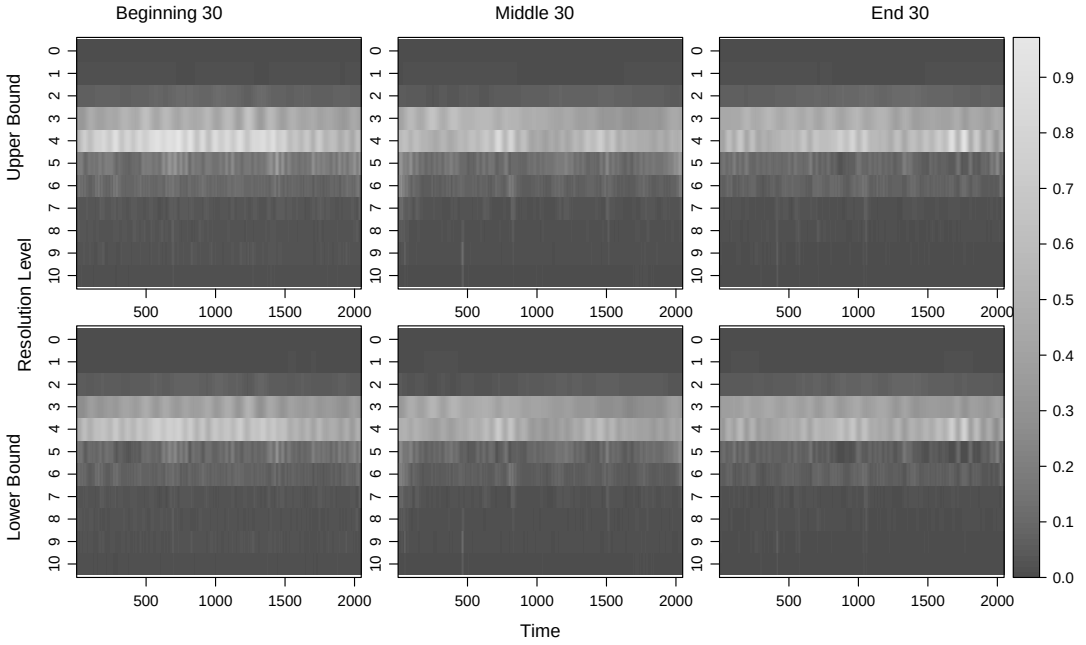


Figure 39: Time-scale 95% bootstrap confidence bounds for the hippocampus (Hc) spectrum for the incorrect trials. Spectral estimate bounds are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: upper bound; *Bottom*: lower bound. Estimation is via the REv-LSW method with smoothing over time and replicates, and takes into account across-trial correlation.

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 40 further illustrates that spectral evolution through time is captured by both the LSW and REv-LSW models. As the experiment progresses, the REv-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.

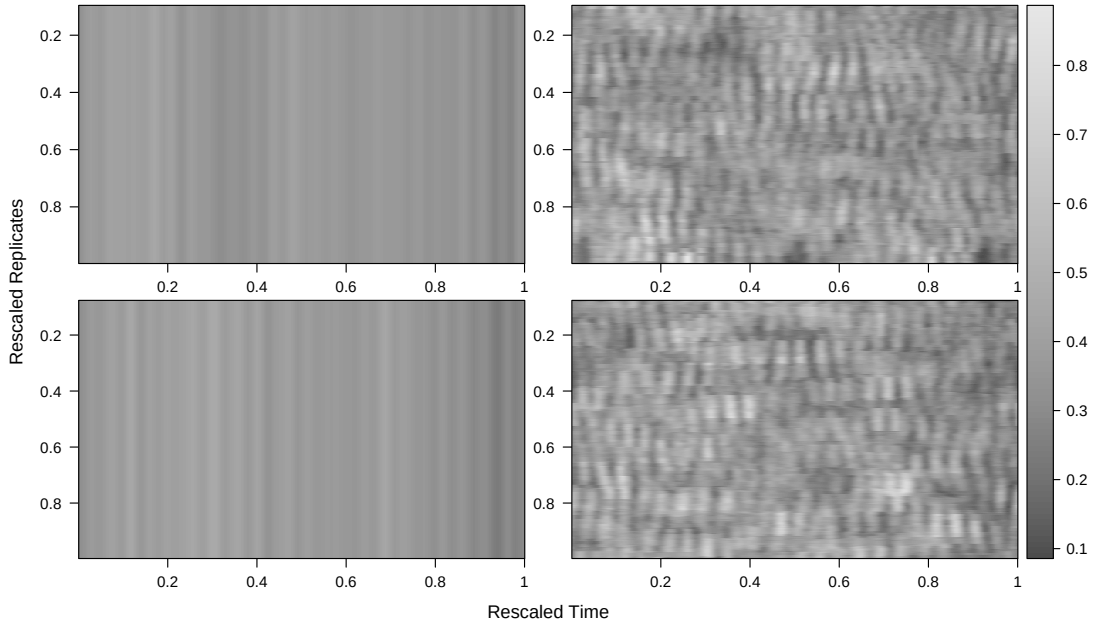


Figure 40: Time-replicate plots of the hippocampus (Hc) average spectra across 2-8Hz for the correct (*Top*) and incorrect (*Bottom*) trials. *Left*: estimates from the LSW method averaged over all replicates; *Right*: estimates from the REv-LSW method with smoothing over time and replicates.

4.2.2 Results for the nucleus accumbens

The resulting (REv-)LSW spectral estimates appear below in Figures 41 and 42 for the correct and incorrect sets of trials, respectively. These plots are to be understood in the same manner as those for the hippocampus. Reassuringly, the estimated spectral characteristics are also rendered by the lower and upper 95% bootstrap confidence bound plots, as illustrated by Figures 43 and 44. Fiecas and Ombao (2016) find that the bulk of variability in the NAc is accounted for by (high) beta band frequencies (20-30Hz), while 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 REv-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. A similar pattern of behaviour is displayed by the incorrect group of trials. Also note that the NAc activity decreases in intensity from the beginning to end replicates for both groups.

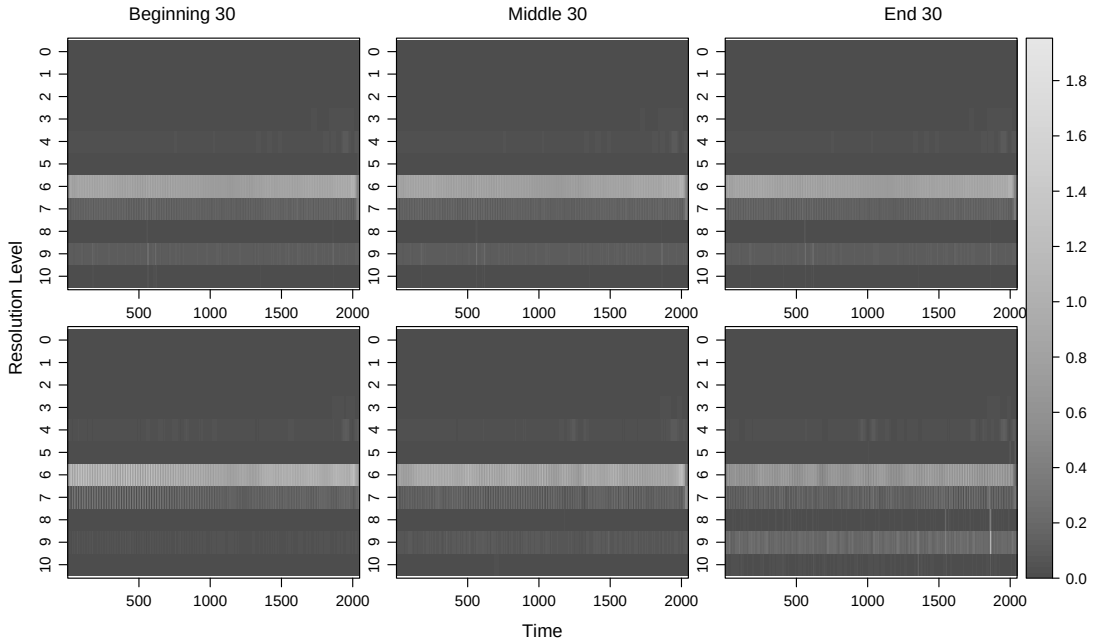


Figure 41: Time-scale nucleus accumbens (NAc) plots for the correct trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: REv-LSW method with smoothing over time and replicates.

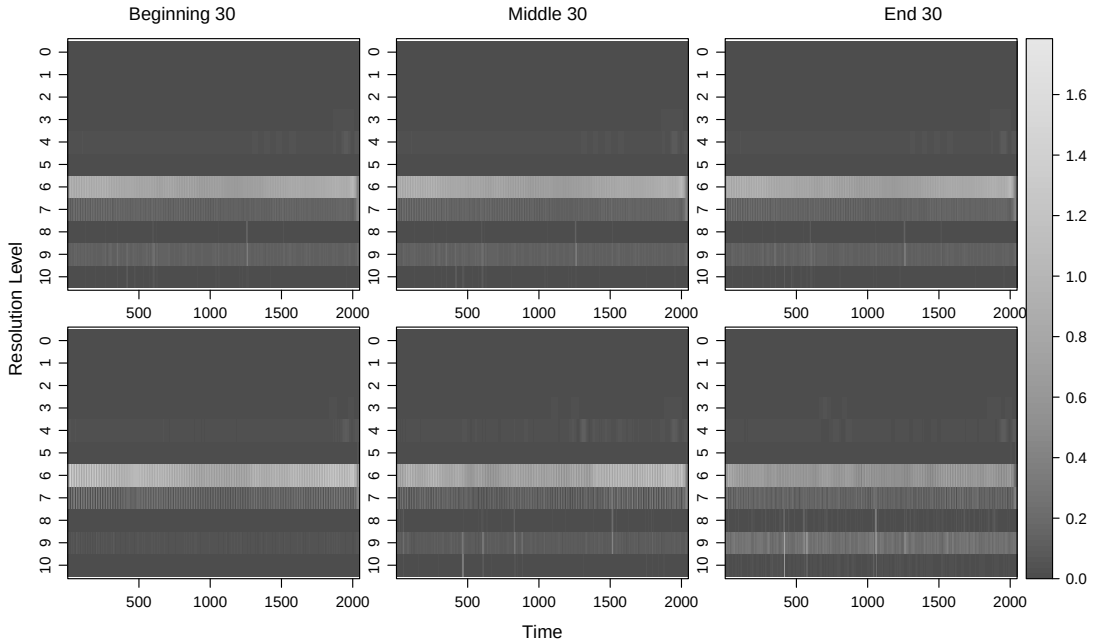


Figure 42: Time-scale nucleus accumbens (NAc) plots for the incorrect trials. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: estimates from the LSW method averaged over all replicates; *Bottom*: REv-LSW method with smoothing 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

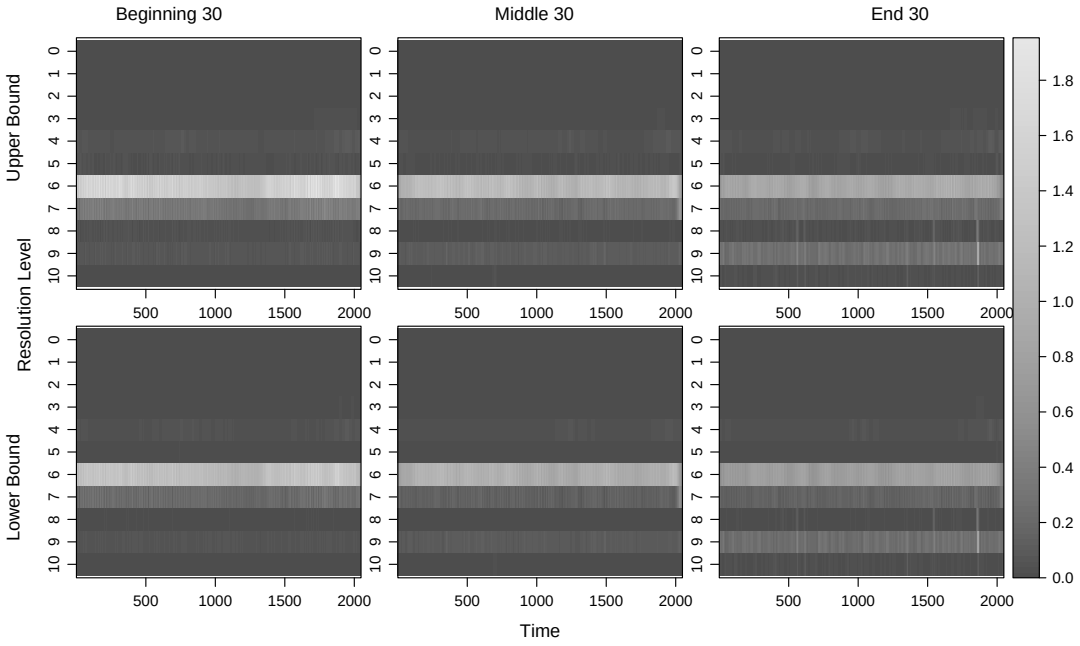


Figure 43: Time-scale 95% bootstrap confidence bounds for the nucleus accumbens (NAc) spectrum for the correct trials. Spectral estimate bounds are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: upper bound; *Bottom*: lower bound. Estimation is via the REv-LSW method with smoothing over time and replicates, and takes into account across-trial correlation.

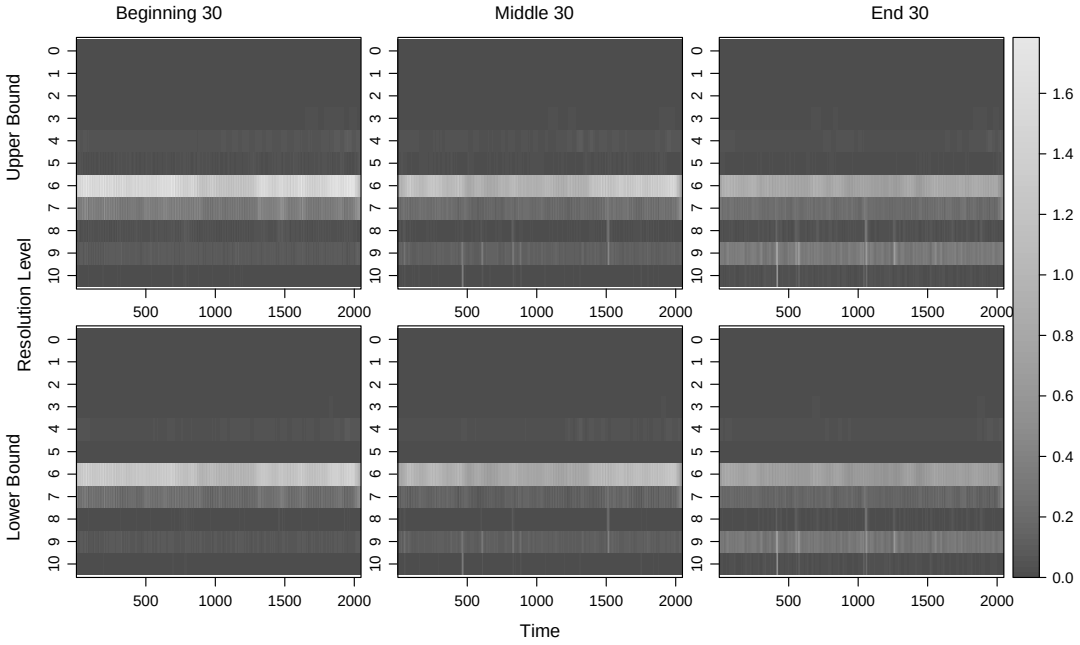


Figure 44: Time-scale 95% bootstrap confidence bounds for the nucleus accumbens (NAc) spectrum for the incorrect trials. Spectral estimate bounds are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: upper bound; *Bottom*: lower bound. Estimation is via the REv-LSW method with smoothing over time and replicates, and takes into account across-trial correlation.

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 45 also reinforces our previous comments. Evolution in the spectra across time is captured by both models, with the NAc activity displaying periodic patterns, while the REv-LSW estimation highlights a decrease in NAc activity along the trials of the experiment.

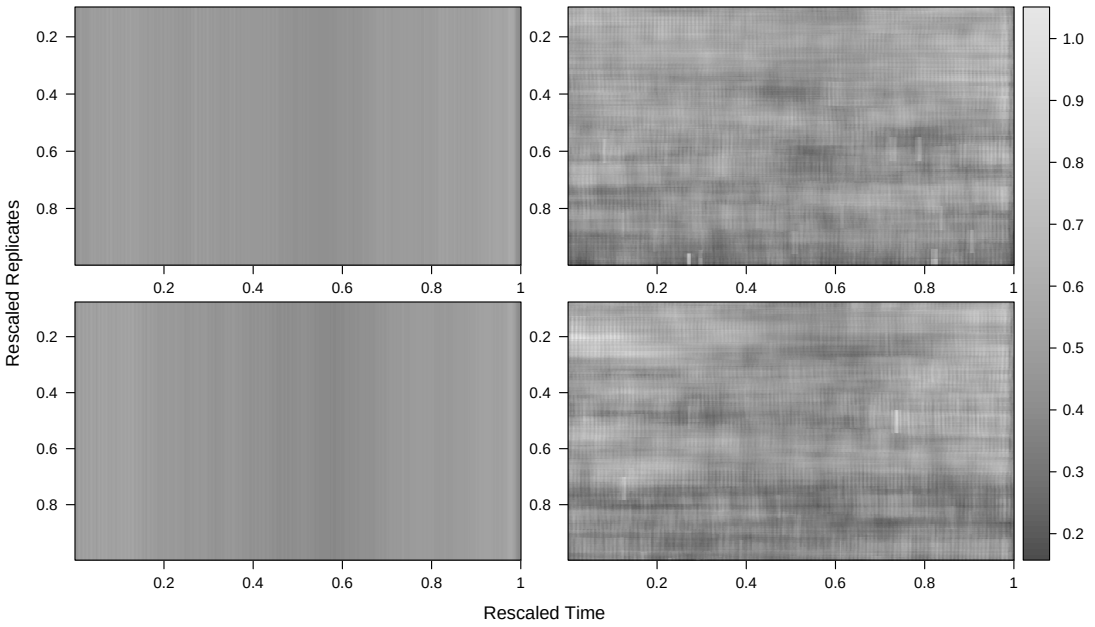


Figure 45: Time-replicate plots of the nucleus accumbens (NAc) average spectra across 16-60Hz for the correct (*Top*) and incorrect (*Bottom*) trials. *Left*: estimates from the LSW method averaged over all replicates; *Right*: estimates from the REv-LSW method with smoothing over time and replicates.

Remark 4.2.1. 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 REv-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.

4.3 Capturing between-trial dependence

Let us now investigate between-trial dependencies, which under the REv-LSW framework are readily quantified by the locally stationary between-trial coherence, a quantity defined in (52).

4.3.1 Results

Hippocampus.

For the Hc activity, the REv-LSW 95% bootstrap confidence bounds that allow for cross-trial dependence indicate that the presence of coherence across trials cannot be discounted, for both correct and incorrect groups. When investigating the estimated Hc trial coherence we have not found any substantial evidence of between-trial dependence. Given that the Hc is associated with memory recall, it is not unrealistic to expect to find some evidence of between-trial dependence. However as discussed previously, our analysis shows evidence of learning but suggests that the macaque has not fully learned the associations. This could explain the lack of between-trial dependence captured in our analysis for the Hc. For completeness, in Figure 46 we display the estimated Hc trial coherence (absolute value) at level 4 (4-8Hz) for trials 20 and 200 in both the correct and incorrect groups. Upon inspection of the plots, there is no behaviour to suggest a between-trial dependence exists and furthermore there are no clear differences between the estimated coherence structures of a beginning trial (trial 20) and end trials (trial 200).

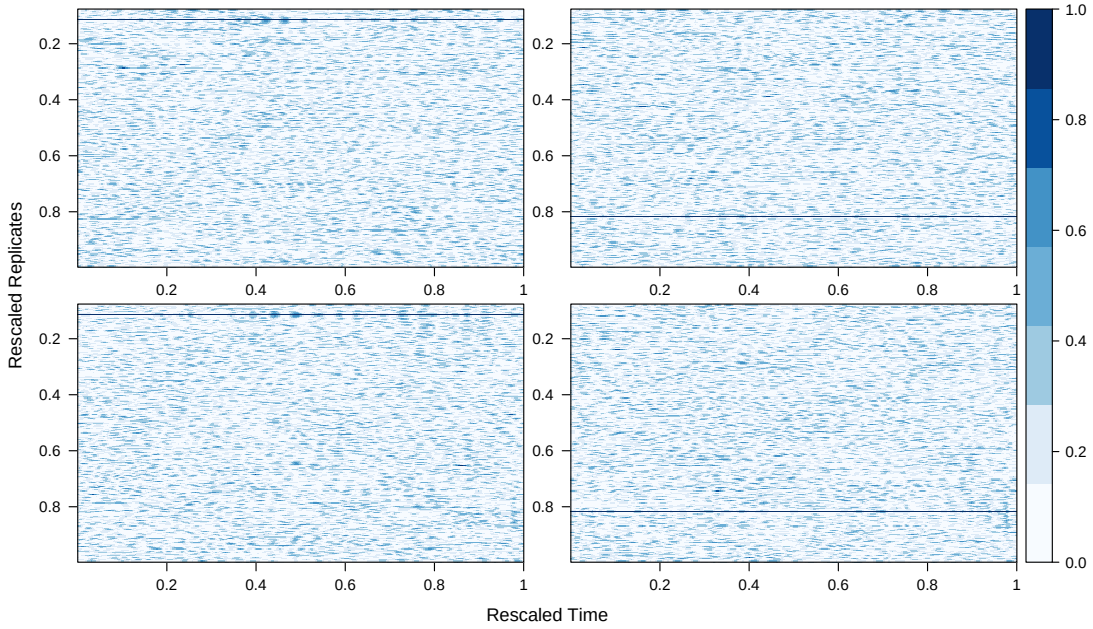


Figure 46: Level 4 absolute value coherence for Hc correct (*top*) and incorrect (*bottom*) trials. *Left*: trial 20; *right*: trial 200.

Nucleus accumbens.

For the NAc, our analysis finds evidence of a moderate dependence across neighbouring trials 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 47 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 48) indeed

reinforce the presence of between-trial dependence.

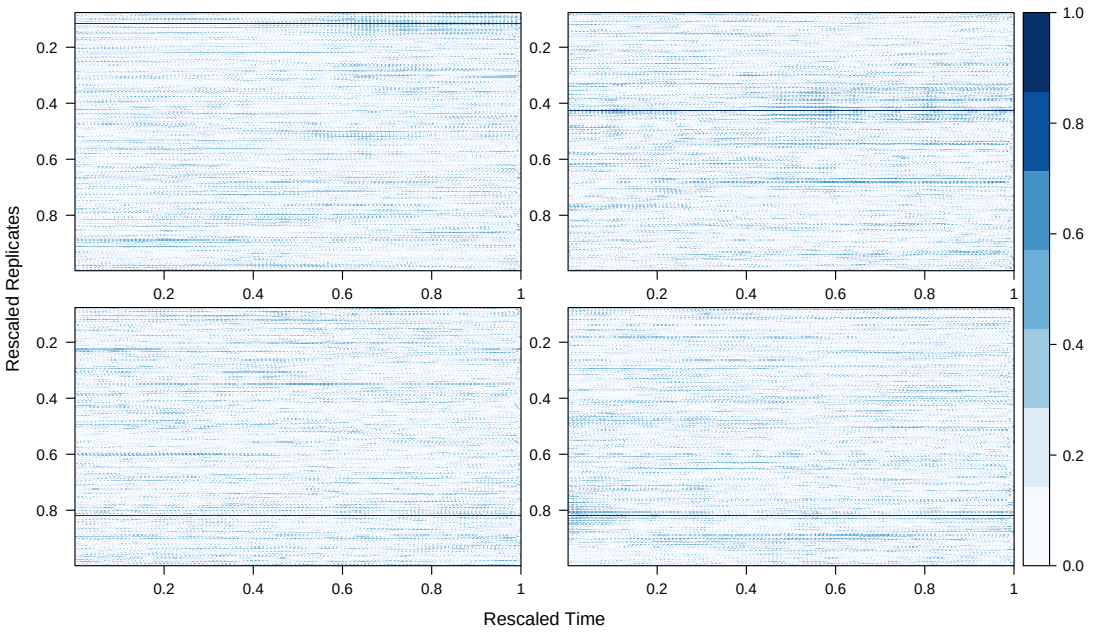


Figure 47: Level 6 absolute value coherence for NAc incorrect trials: 20 (*top left*), 100 (*top right*) and 200 (*bottom left*); and for NAc correct trial 200 (*bottom right*).

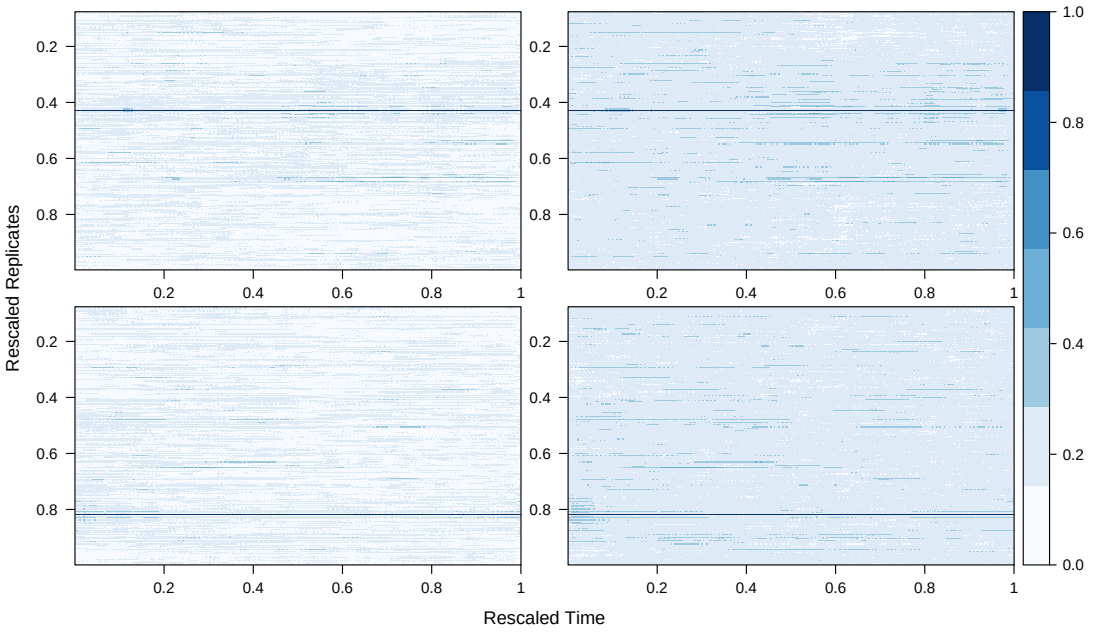


Figure 48: Lower (*left*) and upper (*right*) 95% bootstrap confidence bounds for the level 6 absolute value coherence for NAc incorrect trial 100 (*top*) and correct trial 200 (*bottom*).

Remark 4.3.1. Our analysis provides 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 impacts our understanding of the experiment, and should not be ignored or assumed nonexistent from the onset of the analysis. In comparison with our findings, for this particular data, Chau and von Sachs (2016) also illustrate the presence of between-trial correlation. The authors identified a stronger structure between neighbouring

replicates in the beginning and end of the experiment, although for an accumulation of individual trial timelines and frequencies of interest. 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 confidence bands (Chau and von Sachs, 2016), and thus potentially yielding misleading results (Morris, 2015).

Remark 4.3.2. Within the neuroscience literature (e.g. Gorrostieta *et al.* (2011)), dependence between brain regions is also of interest, with coherence measures setup between channels of interest (here, Hc and NAc). Such measures have not formed the scope of our work here, but in relating our replicate-coherence results to the reported evolutionary coherence between the Hc and NAc, the dependence we observed at the beginning of NAc correct trial 200 (approximately rescaled replicate 0.8) is reminiscent of the dependence between Hc and NAc captured at rescaled replicate-time 0.8 by Fiecas and Ombao (2016). This could also offer an explanation to the lack of evidence found for a between-trial dependence within the Hc. For this experiment, the actions of the macaque could depend on the expectation of the juice reward, in which case a coherence between the Hc and NAc may exist. However, between-trial dependence within the Hc may only come into existence once the macaque has learned the associations without the need of a reward stimulus.

4.4 Concluding remarks

The application to real data from the neurosciences carried out in this chapter, has given substantial evidence for the desirable properties achievable through our proposed novel wavelet-based methodologies detailed in Chapters 2 and 3. Unlike the approach of treating replicate time series as identical process realisations, the REv-LSW model has successfully captured nonstationary process characteristics for the hippocampus (Hc) and nucleus accumbens (NAc) across both trial and time dimensions. For both the Hc and NAc, we have extracted localised information in time within- and between-trials, suggesting that the macaque monkey) learns through the experiment. Furthermore, for the NAc we have captured evidence reflecting that over the course of a learning experiment activity in the ventral striatum (thus NAc) decreases (see also Hollerman and Schultz (1998); Fiecas and Ombao (2016)). A crucial development within our proposed methodology is accounting for between-trial dependence. When assessing for a potential coherence structure between trials within the Hc, our results did not find strong evidence for between-trial dependence, while still not being able to completely discount it. For the NAc, evidence of a moderate between-trial dependence across neighbouring trials was found, highlighting how the actions of the macaque could be influenced by expectation (e.g. of a juice reward). Thus, the results of our neuroscience data analysis give a thorough demonstration for the performance and advantages of the REv-LSW model. However, the methodology itself is not restricted to use within this field, and we envisage its utility in other experimental areas where wavelet spectral analysis has proved to be ideally suited, e.g. circadian biology (Hargreaves *et al.*, 2018, 2019). As previously stated, development of the REv-LSW methodology in a multivariate setting (Sanderson *et al.* (2010); Park *et al.* (2014)) with the addition of defining a variate trial-coherence, would allow for a multivariate analysis between channels,

for example in the bivariate case of the hippocampus and nucleus accumbens data.

5 Testing for variability in the spectra across the replicate dimension

So far we have established how taking into consideration the potential for an evolution in the spectral characteristics over replicates, improves the estimation and identifies additional information about the process dynamics of a replicated time series over the replicate dimension. In a practical sense, before investigating a replicate time series for evolutionary behaviour, potentially through visualising plots of estimated spectra for various replicates, it would be of value to have an indication of whether a spectral evolution across the replicates exists. We refer to the meta-process evolution along the ordered replicates as a *replicate-effect*. Thus, in this chapter we will propose two tests to establish and identify a replicate-effect, a *location-specific* test and a *global* test.

5.1 Overview of existing spectral domain hypothesis testing

The idea of testing for the presence of time-varying dynamics of a process within the spectral domain has existed for some time. One of the earliest tests for overall second order nonstationarity of a time series, is the 'PSR' test proposed by Priestley and Subba Rao (1969), which evaluates the log of their evolutionary spectra estimates through a two-factor analysis of variance at different moments of time. von Sachs and Neumann (2000) constructed a stationarity test which estimates the evolutionary (Fourier) spectrum through localised periodograms. They then make use of the capability of Haar wavelets to capture discontinuities across the spectral density, thus identifying any deviation from covariance stationarity. Expanding on the work of Priestley and Subba Rao (1969), Ahamada and Boutahar (2002) attain an alternative test statistic and demonstrate their test on an application to exchange rates. Paparoditis (2009, 2010) compute sample spectral densities (local periodograms) on a moving window of observations and compare these estimates via an L_2 measure of deviation to a spectral density estimate obtained for the overall time series. A testing procedure to identify whether two locally stationary time series are costationary was proposed by Cardinali and Nason (2010). They utilise a parametric bootstrap (see Davison and Hinkley (1997)) resampling procedure to obtained pseudo-test statistic values from processes under the null hypothesis that the estimated spectrum is constant as a function of time. Several of the Fourier tests of stationarity are based on selecting segments of the time series for comparison and Dwivedi and Subba Rao (2011) draw attention to the sensitivity of the test to the choice of segment length and how the number of segments can impact the rate of convergence to the asymptotic distribution. They thus propose to overcome these problems through a test of stationarity over the entire time series which aims to assess whether the power across differing Fourier frequencies is asymptotically uncorrelated. Whilst many of the above mentioned tests were constructed with Fourier spectra in mind, Nason (2013) proposed a test of similar nature to von Sachs and Neumann (2000) that utilises Haar wavelet coefficients in order to investigate the constancy over scales of the wavelet spectrum instead of Fourier frequencies. Taylor et al. (2014) extend the work of Cardinali and Nason (2010) developing a test of stationarity for spatial processes and apply their test to a texture analysis problem. Cho (2016) uses an unsystematic sub-sampling approach to assess localised second-order nonstationarity over

pairs of random intervals and propose a test statistic that maximises the difference in the spectral characteristics over the interval pairs. The unsystematic approach evades the need for dyadic constraints on the sub-samples, as is the case for the Haar wavelet coefficients in Nason (2013).

To the best of our knowledge, within the current literature there exist very limited testing procedures for the existence of evolutionary spectral characteristics across multiple, or replicate, time series. Perhaps the closest methods that could be framed to compare the (non)stationary characteristics between multiple time series lie within spectral comparison for classification. Fryzlewicz and Ombao (2009) provide a general procedure to classify processes into groups through comparing the estimated wavelet spectrum with the spectral characteristics associated to each group. This procedure was improved upon by Krzemieniewska et al. (2014) who introduce a variance correction to account for potential within-class variation of the spectra for each pair of wavelet scale j and location k. Hargreaves *et al.* (2019) were first to develop a hypothesis test which aims to establish whether the associated evolutionary wavelet spectra of two (classification) groups are significantly different. Of the little literature that exists specifically for the testing of spectral evolution over the replicate domain, Fiecas and Ombao (2016) propose a Fourier-based resampling procedure under the framework of log-linear models. Their procedure echoes the functional regression problem of testing the equality of the curves in the time domain (for example see Dette and Neumeyer (2001)). Functional regression approaches often deal with replicate time series data by projecting it into the Fourier or wavelet domain where the spectral representations become subject to modelling. For an introduction to functional regression and more specifically functional linear models see Ramsay and Silverman (2005), additionally Morris (2015) provides an updated review on current methods in the field. When testing for differences between curves in the spectral domain, techniques are centred around functional ANOVA (see Ramsay and Silverman (2005)). Shumway (1988) (or more recently Shumway and Stoffer (2017)) computed individual test statistics for each given frequency (in the Fourier domain) to detect differences in the Fourier spectra of the mean curves. The individual test statistics were combined in Fan and Lin (1998) to form an overall test based on the adaptive Neyman test to identify differences between two groups of curves, which they then generalise to compare multiple groups of curves through an adaptive high-dimensional ANOVA. Developments of functional (F)ANOVA methods have been established in the wavelet domain (WANOVA) by Raz and Turetsky (1999); Vidakovic (2001) and McKay *et al.* (2013), with the second highlighting that the decorrelation properties of wavelet transformations (Chang and Stein (2013)) are beneficial for regularisation. Atkinson et al. (2017) note the favourable properties of WANOVA to smooth, reduce dimensionality and decorrelate time series data and thus develop a WANOVA-based model validation process. In the setting of locally stationary processes (Dahlhaus, 1997), Guo et al. (2003) propose a smoothing spline (SS-)ANOVA model fitted to time-varying log-periodograms constructed using SLEX basis functions (see Ombao *et al.* (2002)). Martinez *et al.* (2013) study differences in regional bat mating chirps through a Bayesian functional mixed model and employ the approach described in Morris et al. (2011) of mapping the localised Fourier power spectra of the chirps (the functional responses) into the wavelet domain.

This chapter proceeds as follows. Section 5.2 (i) introduces the proposed location-specific test of replicate-effect under the assumption that replicates are uncorrelated, and (ii) simulation studies demonstrate the performance of the test and the advantages it brings to the analysis. In Section 5.3, (i) we give the details of the proposed global test of replicate-effect constructed under the assumption that replicates are uncorrelated, and (ii) the test performance is assessed through simulation studies. Section 5.4 recalls the neuroscience real data study and we apply our proposed tests of replicate-effect to the macaque dynamic brain processes data. We conclude the chapter in Section 5.5.

5.2 Location-specific test for replicate-effect

Recalling Definition 2.1.1 in Chapter 2, our REv-LSW process is defined to allow for a slow evolution (to allow for meaningful estimation and asymptotics) of the process amplitudes over replicates, controlled by the Lipschitz continuous transfer function $\widetilde{W}_j(z,\nu)$. Thus, we propose the *location-specific* test that is capable of identifying both the replicates and time locations within replicates where spectral evolution is present. To test whether this evolution is indeed manifest within the spectral quantities, we formulate the null hypothesis that the wavelet spectrum is constant over rescaled replicate-time $\nu \in (0,1)$ and test this for each rescaled time $z = k/T$.

Thus we propose to test that

$$\begin{aligned} H_0^k &: S_j\left(\frac{k}{T},\nu\right) \text{ is constant as a function of } \nu, \forall j \quad \text{versus} \\ H_A^k &: \exists j^\star \text{ such that } S_{j^\star}\left(\frac{k}{T},\nu\right) \text{ is non-constant over } \nu \in (0,1). \end{aligned}$$

For subsequent ease, we shall make use of the spectral quantity $\beta_j\left(\frac{k}{T},\nu\right) = \sum_l A_{j,l} S_l\left(\frac{k}{T},\nu\right)$, a quantity analogous to that introduced by Fryzlewicz and Nason (2006) and previously introduced in Chapter 2. We recall, that finding a well-behaved estimator for S is equivalent to finding a well-behaved estimator for β, due to the invertibility of the matrix A and boundedness of its inverse norm. As such, the within-replicate raw periodogram $I^{r;R}_{j,k;T}$ is an asymptotically unbiased estimator for β (see Proposition 2.1.10). For a given time k, it can be immediately verified that our testing problem above can be re-framed as the pointwise testing problem

$$\begin{aligned} H_0^k &: \beta_j\left(\frac{k}{T},\nu\right) \text{ is constant as a function of } \nu, \forall j \quad \text{versus} \\ H_A^k &: \exists j^\star \text{ such that } \beta_{j^\star}\left(\frac{k}{T},\nu\right) \text{ is non-constant over } \nu \in (0,1). \end{aligned}$$

While smoother wavelets may be preferred over Haar wavelets for spectral estimation due to their potential superior convergence rates, conversely, Haar wavelets are a suitable tool for detecting departures from constancy (von Sachs and Neumann, 2000). In the spirit of von Sachs and Neumann (2000); Nason (2013), we propose to assess the constancy of $\beta_j(\frac{k}{T},\cdot)$ by assessing the departures of its Haar wavelet coefficients from 0.

We therefore proceed to define the Haar wavelet coefficients of the spectral function

$\beta_j(\frac{k}{T}, \cdot)$ at scale $j = 1, \ldots, J$ and time $k = 0, \ldots, T-1$ as

$$\eta_{i,p}^{(j,k)} = \int_0^1 \beta_j \left(\frac{k}{T}, \nu \right) \psi_{i,p}^H(\nu) d\nu, \tag{58}$$

where $\psi_{i,p}^H(\nu)$ denote the Haar wavelet at (Haar) scale $i = 1, ..., J'$ and (Haar) location $p = 1, \ldots, 2^i - 1$. Crucially, note that the Haar wavelet is operating over the (rescaled) *replicate* dimension, as opposed to the time dimension, thus $J'(R) = \log_2(R)$. From the Haar wavelet property $\int_0^1 \psi^H(\nu) d\nu = 0$, it directly follows that the Haar wavelet coefficients in equation (58) are zero under the null hypothesis H_0^k.

Our null hypothesis now amounts to $H_0^k : \eta_{i,p}^{(j,k)} = 0$ for all j, i, and p. If for any k there exists a scale $j^\star$ and Haar scale and location $(i^\star, p^\star)$ such that $\eta_{i^\star,p^\star}^{(j^\star,k)} \neq 0$, then we conclude that $\beta_j\left(\frac{k}{T}, \cdot\right)$ is not constant and there exists a replicate-effect, i.e., the spectral quantities are indeed evolving over replicates and an REv-LSW approach to modelling the data is to be preferred over a blanket-approach across all replicates.

Before proposing a test statistic, let us first make the following assumptions that follow in a similar vein to Nason (2013).

Assumption 1. We consider Haar wavelet coefficients $\eta_{i,p}^{(j,k)}$ at coarser scales $i = \mathcal{O}(\log_2(R))$ for the asymptotic control of bias.

Assumption 2.
- (a) For a fixed time k, $\sup_\nu \left| \widetilde{W}_j(\frac{k}{T}, \nu) \right| < \infty$, at all scales j.
- (b) For a fixed time k, $\inf_\nu \left| \widetilde{W}_j(\frac{k}{T}, \nu) \right| \geq \epsilon$ for some $\epsilon > 0$ and all scales j.
- (c)
  - i. For a fixed time k, the transfer function $\widetilde{W}_j(\frac{k}{T}, \nu)$ has at most uniformly bounded variation across the replicate dimension (ν), at all scales j.
  - ii. By Definition 2.1.1 of an REv-LSW process, for each scale $j \geq 1$, there exists a Lipschitz continuous transfer function $\widetilde{W}_j^*(z, \nu)$ which forces the amplitudes to evolve slowly over rescaled time (z) and replicate (ν) dimensions. For a fixed time k, $\left| \widetilde{W}_j^*(\frac{k}{T}, \nu) \right| \leq \left| \widetilde{W}_j(\frac{k}{T}, \nu) \right|$, thus ensuring that Assumption 2(c)i holds.

In order to estimate the Haar wavelet coefficients in equation (58), we replace the spectral quantity $\beta_j\left(\frac{k}{T}, \nu\right)$ by means of its asymptotically unbiased estimator (see Proposition 2.1.10), i.e., the corresponding raw periodogram $I_{j,k;T}^{\lfloor \nu R \rfloor;R}$. For a given time k, we thus proceed to use its associated wavelet periodograms at all scales j and over all replicates $r = \lfloor \nu R \rfloor$. For notational simplicity, we denote these quantities as $I_{j,k}^r$.

Hence at time k and scale j, the estimates of the Haar wavelet coefficients are given by

$$\hat{\eta}_{i,p}^{(j,k)} = 2^{-i/2} \left(\sum_{l=0}^{2^{i-1}-1} I_{j,k}^{2^i p - l} - \sum_{q=2^{i-1}}^{2^i - 1} I_{j,k}^{2^i p - q} \right) \tag{59}$$

where again $i = 1, ..., J'$ and $p = 1, ..., 2^i - 1$ and we point out that the Haar wavelet transform is taken over replicates.

Proposition 5.2.1. *Under Assumptions 1 and 2, $\hat{\eta}_{i,p}^{(j,k)}$ is asymptotically normal and has the following properties for all j, i and p, and for a fixed k*

1. $\mathbf{E}\left[\hat{\eta}_{i,p}^{(j,k)}\right] = \eta_{i,p}^{(j,k)} + \mathcal{O}(2^j T^{-1} R^{1/2}) + \mathcal{O}(R^{-1/2})$,

2. $var\left(\hat{\eta}_{i,p}^{(j,k)}\right) = 2R^{-1}\int_0^1 \beta_j^2\left(\frac{k}{T},\nu\right)\left(\psi_{i,p}^H(\nu)\right)^2 d\nu + \mathcal{O}(2^{2j}(RT)^{-1}) + \mathcal{O}(2^j R^{-2})$.

Proof. The proof follows the steps presented in Proposition 1 of Nason (2013) and the results in our Proposition 2.1.10 in Chapter 2. □

Proposed test statistic

Under the null hypothesis H_0^k, von Sachs and Neumann (2000, Theorem 3.1) guarantees that the interval coverage rates achieved when replacing $(\sigma_{i,p}^{(j,k)})^2 = \text{var}\left(\hat{\eta}_{i,p}^{(j,k)}\right)$ by an estimator $(\hat{\sigma}_{i,p}^{(j,k)})^2$ are asymptotically equivalent to those attained by means of a normal distribution. Nason (2013) also use this result in their development.

Recalling that for a fixed time k, under the null hypothesis that $\beta_j\left(\frac{k}{T},\cdot\right)$ is constant over the rescaled replicate dimension, implies that $\eta_{i,p}^{(j,k)} = 0$ for all j, i, and p. This leads us to propose testing for H_0^k via the test statistics

$$\boldsymbol{T}_{i,p}^{(j,k)} = \frac{\hat{\eta}_{i,p}^{(j,k)}}{\hat{\sigma}_{i,p}^{(j,k)}}, \text{ at all } j, i, p,$$

where $\left(\hat{\sigma}_{i,p}^{(j,k)}\right)^2$ is an estimator for $(\sigma_{i,p}^{(j,k)})^2 = \text{var}\left(\hat{\eta}_{i,p}^{(j,k)}\right)$. The test statistics for all original scales j, Haar scales i and Haar locations p are then compared with a critical value obtained from the normal distribution.

One way to estimate $(\sigma_{i,p}^{(j,k)})^2$ is by taking

$$(\hat{\sigma}_{i,p}^{(j,k)})^2 = 2^{-i}\left[\sum_{l=0}^{2^{i-1}-1}\text{var}(I_{j,k}^{2^i p-l}) + \sum_{q=2^{i-1}}^{2^i-1}\text{var}(I_{j,k}^{2^i p-q})\right], \tag{60}$$

where we make use of the fact that the replicates are uncorrelated, hence $\text{cov}(I_{j,k}^r, I_{j,k}^{r'}) = 0$ for $r \neq r'$.

Under the null hypothesis, we could also estimate $(\sigma_{i,p}^{(j,k)})^2$ by replacing the unknown $\beta_j^2(k/T,\nu)$ in property 2 of Proposition 5.2.1 by an average of the squared raw wavelet periodograms $\left(I_{j,k}^r\right)^2$ across all replicates r (denote it as $\left(\overline{I}_{j,k}\right)^2$). This gives

$$(\hat{\sigma}_{i,p}^{(j,k)})^2 = 2R^{-1}\left(\overline{I}_{j,k}\right)^2\int_0^1 \psi_{i,p}^H(\nu)^2 d\nu = 2R^{-1}\left(\overline{I}_{j,k}\right)^2, \tag{61}$$

where we have used the unit norm property of Haar wavelets at all i and p, and we define $\left(\overline{I}_{j,k}\right)^2 = R^{-1}\sum_r\left(I_{j,k}^r\right)^2$.

As we proposed to test many hypotheses, we deal with the multiple hypothesis testing by taking a false discovery rate approach (Benjamini and Hochberg, 1995) or, for a stricter procedure, the Bonferonni correction (Nason, 2013). Additionally we note that the number of hypothesis test carried out per location is dependent on how many original wavelet

scales j and Haar wavelet scales j' we choose to test over, such that

$$\#\{tests\ per\ location\} = \#\{original\ j\ scales\} * \#\{Haar\ j'\ scales\}, \tag{62}$$

and we recall that a choice of coarser Haar wavelet scales are selected due to Assumption 1.

5.2.1 Simulation study for the location-specific testing methodology

Here we present some simulated examples to assess the capability of our test of constancy. Each simulation follows a REv-LSW process as in Definition 2.1.1 for various R and T. From our test, we aim to identify the locations k where the null hypothesis $H_0^k : \eta_{i,p}^{(j,k)} = 0$ is rejected for some j, i, and p and employ the false discovery rate with 5% nominal size to control the number of false rejections. For each of the simulations, we choose to test over the finest $(J-3)$ original scales and Haar scales $j' = 1, ..., J^*$, where $J^* = (J' - \lceil J'/2 \rceil + 1)$ and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Thus, the number of hypothesis tests per location (equation (62)) is determined by $(J-3) \times J^*$, recalling that $J = \log_2(T)$ and $J' = \log_2(R)$. For each time location, we obtain empirical power/size estimates through counting the number of times the test correctly/incorrectly identifies a breach in spectral constancy. In practice we may only be concerned with whether the test successfully identifies an evolution in the spectral quantities over the replicates of a REv-LSW process. As such, to quantify global test performance, we provide statistical measures of a binary classification test on average over 100 simulations. The precise formulas for these measures are given in Table 8, where a 'positive' is understood to be a location that is identified by the test as rejecting the null hypothesis of constancy across replicates, and a 'negative' is understood to be a location that fails to reject the null hypothesis. The reported results were obtained through estimating the variance (equation (60)), which was found to give the most favourable estimates for the empirical power and size values.

Binary Classification rates

Rate	Formula
True Positive Rate (TPR)	$\frac{tp}{tp+fn}$
True Negative Rate (TNR)	$\frac{tn}{tn+fp}$
False Discovery Rate (FDR)	$\frac{fp}{tp+fp}$
False Omission Rate (FOR)	$\frac{fn}{tn+fn}$
Fowlkes-Mallows Index (F-M)	$\sqrt{\frac{tp}{tp+fp} \cdot \mathrm{TPR}}$
Accuracy (ACC)	$\frac{tp+tn}{tp+tn+fp+fn}$
Prevalence Threshold (PT)	$\frac{\sqrt{\mathrm{TPR}(-\mathrm{TNR}+1)}+\mathrm{TNR}-1}{\mathrm{TPR}+\mathrm{TNR}-1}$

Table 8: List of binary classification rate formulas where 'tp' and 'fp' denote the true and false positives, and 'tn' and 'fn' denote the true and false negatives.

Simulation 1

The following variations are constructed using the same underlying squared sine spectral characteristics and different burst values are introduced to assess how well the location-specific test identifies locations that exhibit a spectral evolution across the (rescaled) replicates. For rescaled time $z = k/T$ and replicates $\nu = r/R$, variations of simulation 1 are defined as follows:

1A. Firstly, we aim to evaluate the test size. Specifically, we want to get a sense of how the test performs in the absence of any evolution in the true spectra across the replicates, hence we define the true spectra to be

$$S_j(z,\nu) = \begin{cases} \sin^2(2\pi z), & \text{for } j = J(T) - 4, z \in (0,1), r \in (0,1) \\ 0, & \text{otherwise.} \end{cases}$$

1B. We now introduce a burst value for the spectra which is contained in the first $T/2$ locations and defined for the second half of the replicates. We mathematically define the true spectra as a function of R and T for simulation 1B as follows

$$S_j(z,\nu) = \begin{cases} \sin^2(2\pi\nu), & \text{for } j = J(T) - 4, z \in (0,1), \nu \in (0, \frac{1}{2}) \\ & \quad \text{and } z \in (\frac{1}{2} + \frac{1}{T}, 1), \nu \in (\frac{1}{2} + \frac{1}{R}, 1) \\ \text{'burst'}, & \text{for } j = J(T) - 4, z \in (0, \frac{1}{2}), \nu \in (\frac{1}{2} + \frac{1}{R}, 1) \\ 0, & \text{otherwise,} \end{cases} \tag{63}$$

with the 'burst' taking values of 1, 2 and 5. A visual representation of the true spectra with a burst value of 2 can be seen in the left plot of Figure 49. For a simulation with $R = 256$ and $T = 512$, the first 128 replicates each have a sine squared behaviour through all 512 locations, but the second 128 replicates have a burst in the first 256 locations and picks up the sine squared behaviour in the last 256 locations.

1C. For this simulation we reproduce the spectra defined in 1B but reduce the locations containing a burst to the first $T/4$, such that

$$S_j(z,\nu) = \begin{cases} \sin^2(2\pi z), & \text{for } j = J(T) - 4, z \in (0,1), \nu \in (0, \frac{1}{2}) \\ & \quad \text{and } z \in (\frac{1}{4} + \frac{1}{T}, 1), \nu \in (\frac{1}{2} + \frac{1}{R}, 1) \\ 2, & \text{for } j = J(T) - 4, z \in (0, \frac{1}{4}), r \in (\frac{1}{2} + \frac{1}{R}, 1), \\ 0, & \text{otherwise.} \end{cases} \tag{64}$$

A visualisation of the true spectra appears in Figure 49 (right) for a simulation with $R = 256$ and $T = 512$.

We will next investigate the power and size of the proposed test within this simulated setup.

Results **1A.** For $R = 256$ and $T = 512$, we display the empirical size estimates in plot (a) of Figure 50 where we identify the locations expected to break the null hypothesis of

constancy by a horizontal line at 1 and the locations that are constant across replicates (and thus should not reject the null) are identified by a horizontal line at 0. In addition to plot (a), the binary classification measures given in Table 9 demonstrate that the test is extremely capable of correctly identifying all the locations that do not breach the null hypothesis of no replicate-effect, we show that this is true for choices of $T = 256$ to 512 and $R = 128$ to 256. In Table 9, the true negative rates (TNR) measure the proportion of negatives correctly identified by the test and the accuracy (ACC) gives the proportion of the locations correctly identified as either positive or negative. Note that here, we report fewer measures due to the absence of true positives (locations with differing spectra across replicates) and for the same reason we observe that the TNR=ACC.

Results **1B.** Recall that our aim is to verify whether the spectra for the first 256 locations are identified as not constant across the replicate dimension. For burst values of 1 and 2, we present the empirical power and size estimates for each time location in Figure 50, plots (b) and (c) respectively. Upon inspection of the plots, we notice a dip in the power level around replicate 128, which results from the spectra of the squared sine and the burst becoming closer in value. For a burst of 1, the test clearly breaks down as the difference between spectra approaches zero. Next, it is evident that the majority of false rejections for the simulation with burst value of 2 are observed at the true rejection region boundaries. This can be attributed to the artificial power leakage introduced during estimation of the spectra. Additionally, we also investigated a burst value of 5, which resulted in the empirical power estimates all lying on the horizontal line at 1. On the whole, the individual location power and size estimates suggest that the test performs very well over the 100 simulations when the difference between the spectra of the squared sine and burst characteristics is larger than 1, so it can be identified.

Binary classification measures, as defined in Table 8, are reported for this simulation in Table 10 for the different burst values, $T = 256$ to 512 and $R = 128$ to 256. Considering each burst individually, it is apparent in each case that the test performance improves as R and T increases, which is of no surprise given the improved spectral estimation as R and $T \to \infty$ (see Proposition 2.1.13 in Section 2.1). This improvement is seen in the true positive rates (TPR), which we recall measure the proportion of actual positives correctly identified by the test. Furthermore the false discovery rates (FDR) and false omission rates (FOR), which respectively measure the proportion of identified positives and identified negatives that are false, are also decreasing with R and T increasing. The Fowlkes-Mallows index (F-M) measures the similarity between the TPR and the proportion of identified positives that are true (the positive predictive rate), where greater similarity is indicated by a higher index value. The accuracy (ACC) measure of the test gives the proportion of the locations correctly identified as either positive or negative. Note however that the ACC alone can be a misleading tool since it is not explicit whether the positives or negatives are contributing more to the accuracy. When comparing across the bursts, the measures show that the test gets better at identifying the true positives as the burst increases (greater difference between spectra) at a cost of a slight worsening of the true negative rate (TNR). This, as mentioned previously, is due to the power leakage across locations during estimation. Also, the measures give further evidence to the test struggling to identify a difference between the defined burst (of 1) and sine squared spectra of less than 1, while there is significant

improvement in the measures as the burst is increasing (e.g. 2). Finally, in Figure 51 we visually supplement the above results with a receiver operator characteristic (ROC) plot, which plots the TPR against the false positive rate (FPR), with the latter measuring the proportion of actual negatives incorrectly identified as positive, i.e., (1 - TNR). This plot summarises how an increase in R and T, and furthermore how the difference in value of the spectra between replicates (greater burst) lead to an improvement in the performance of the test amounting to correctly identifying the locations that indeed present evolutionary spectral characteristics over the replicates.

Results **1C.** For this simulation with a burst value of 2, power and size estimates can be visualised in plot (d) of Figure 50 and in Table 11 we report statistical measures on the overall test. The results for this test are very favourable, with the majority of true positive and negative locations being identified. In comparison to simulation 1B (with a burst value of 2), here the spectral leakage across locations at the location rejection boundary has less impact due to a smaller difference between spectra across replicates for these locations.

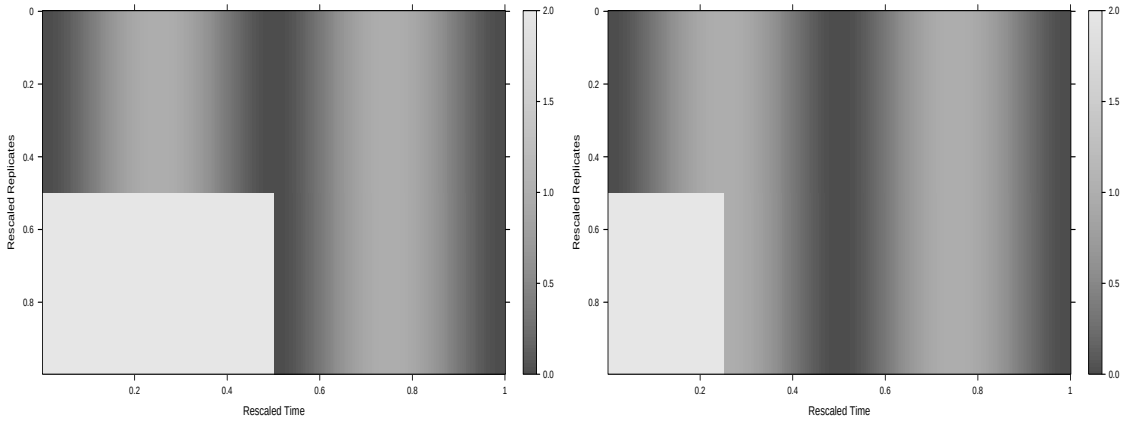
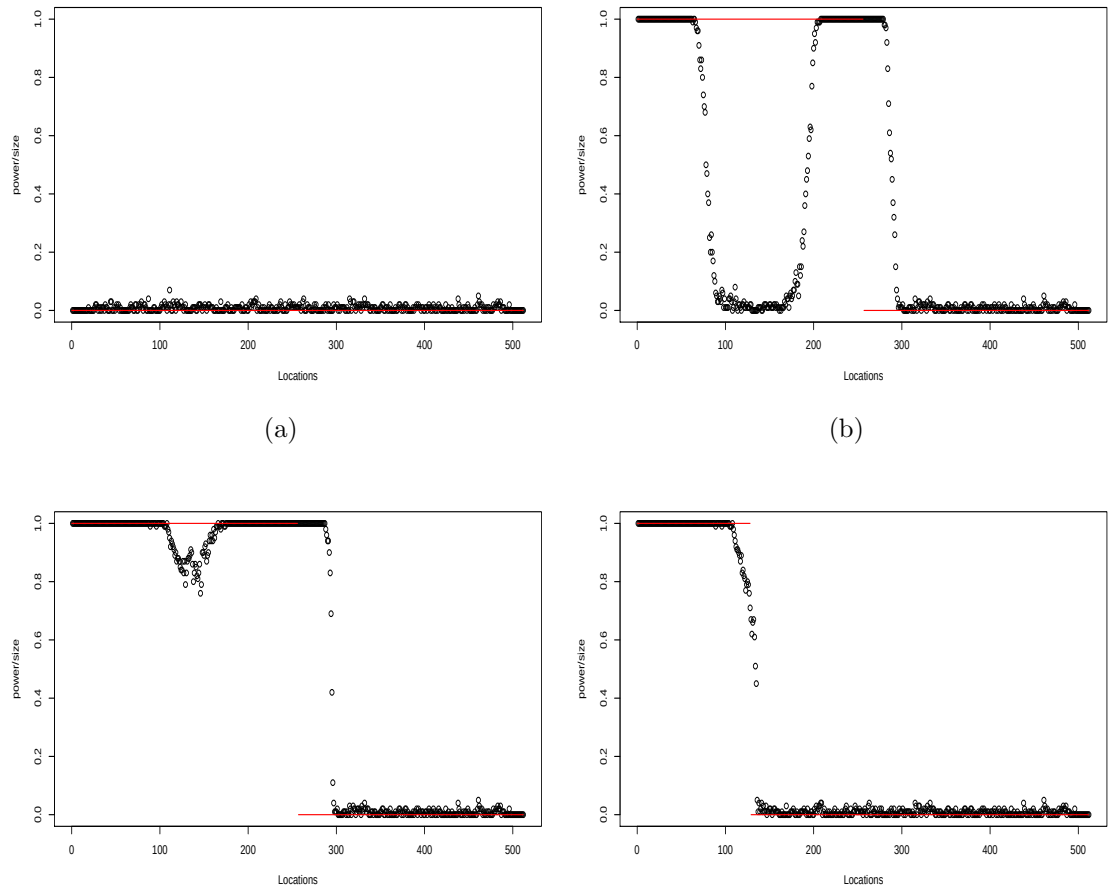


Figure 49: Time-replicate plot of the true spectra in level 5 with 'burst'= 2 for simulation 1B (*left*) and simulation 1C (*right*).

Binary classification measures

R	T	TNR	sd_{TNR}	ACC
128	256	0.9918	0.0078	0.9918
256	256	0.9941	0.0058	0.9941
256	512	0.9921	0.0049	0.9921

Table 9: Binary classification measures averaged over 100 runs for simulation 1A. Note that TNR=ACC



(a) (b)

(c) (d)

Figure 50: Simulation 1 empirical power and size estimates for each location over 100 runs. *(a)*: simulation 1A; *(b)*: simulation 1B with 'burst'= 1; *(c)*: simulation 1B with 'burst'= 2; *(d)*: simulation 1C with 'burst'= 2.

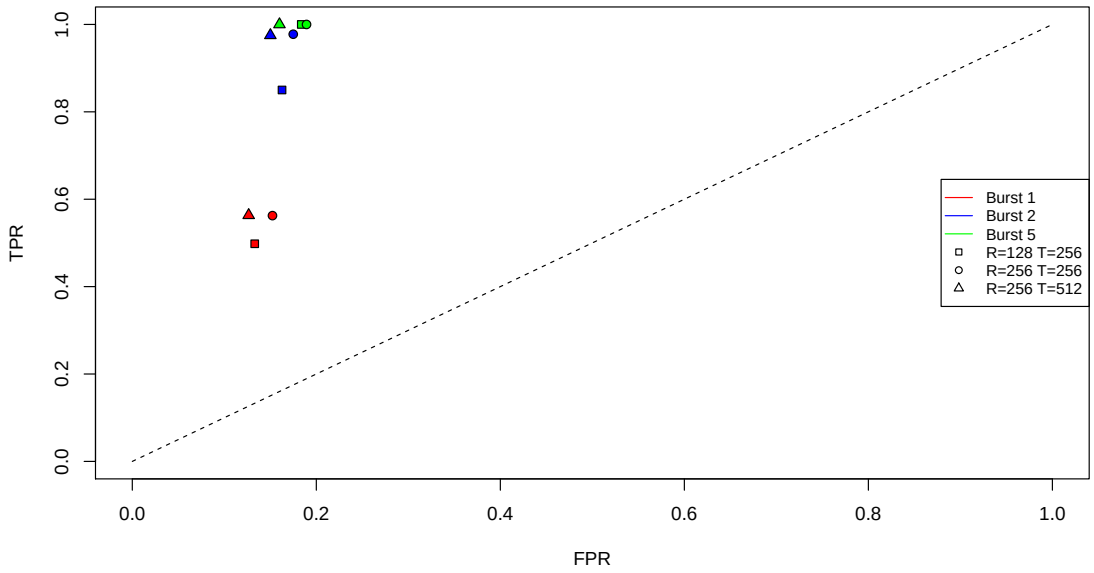


Figure 51: ROC plot for simulation 1B. It is desirable to see the improved TPR as R and T increase whilst also highlighting improved test performance on greater burst values. Furthermore the minimal impact on FPR is also a desirable find.

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.4980	0.0327	0.8591	0.0185	0.2107	0.3685	0.6267	0.6786	0.3467
256	256	0.5625	0.0282	0.8398	0.0167	0.2130	0.3422	0.6652	0.7012	0.3476
256	512	0.5634	0.0223	0.8696	0.0139	0.1833	0.3341	0.6782	0.7165	0.3244

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.85	0.0571	0.8295	0.0177	0.1610	0.1503	0.8441	0.8397	0.3092
256	256	0.9777	0.0218	0.8173	0.0102	0.1518	0.0260	0.9106	0.8975	0.3017
256	512	0.9752	0.0197	0.8460	0.0095	0.1333	0.0280	0.9193	0.9106	0.2842

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	1	0	0.8085	0.0117	0.1551	0	0.9192	0.9043	0.3042
256	256	1	0	0.8028	0.0088	0.1592	0	0.9170	0.9014	0.3074
256	512	1	0	0.8361	0.0070	0.1379	0	0.9285	0.9181	0.2881

Table 10: Binary classification measures averaged over 100 runs for simulation 1B. *Top*: 'burst'= 1; *middle*: 'burst'= 2; *bottom*: 'burst'= 5. Prevalence = 0.5.

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.9013	0.0540	0.9775	0.0140	0.0522	0.0322	0.9235	0.9584	0.1297
256	256	0.9842	0.0222	0.9688	0.0124	0.0722	0.0053	0.9553	0.9726	0.1477
256	512	0.9769	0.0230	0.9799	0.0078	0.0505	0.0077	0.9629	0.9791	0.1227

Table 11: Binary classification measures averaged over 100 runs for simulation 1C with 'burst'= 2. Prevalence = 0.25.

Simulation 2

We now introduce a more demanding spectral structure of the meta-process, where the different behaviour is manifest in the middling replicates and time locations. The simulated true spectra are defined by

$$S_j(z,\nu) = \begin{cases} \cos^2(2\pi z), & \text{for } j = J(T)-4, z \in (0,1), \nu \in (0,\frac{3}{8}) \cup (\frac{5}{8}+\frac{1}{R},1) \\ & \quad \text{and } z \in (0,\frac{3}{8}) \cup (\frac{5}{8}+\frac{1}{T},1), \nu \in (\frac{3}{8}+\frac{1}{R},\frac{5}{8}) \\ \text{'burst'}, & \text{for } j = J(T)-4, z \in (\frac{3}{8}+\frac{1}{T},\frac{5}{8}), \nu \in (\frac{3}{8}+\frac{1}{R},\frac{5}{8}) \\ 0, & \text{otherwise,} \end{cases} \tag{65}$$

where the burst is now defined for fewer replicates, and in rescaled arguments $z = k/T$ and $\nu = r/R$. For a simulation with $R = 256$ and $T = 512$, we can visualise the spectral characteristics in Figure 52 featuring a squared cosine behaviour in all replicates except for the middle $R/4 = 64$ (here, replicates 97 to 160), which have a burst across the middle $T/4 = 128$ locations (here, locations 193 to 320). It is for these locations that we would expect to see rejections of the null hypothesis of constancy across the replicate dimension.

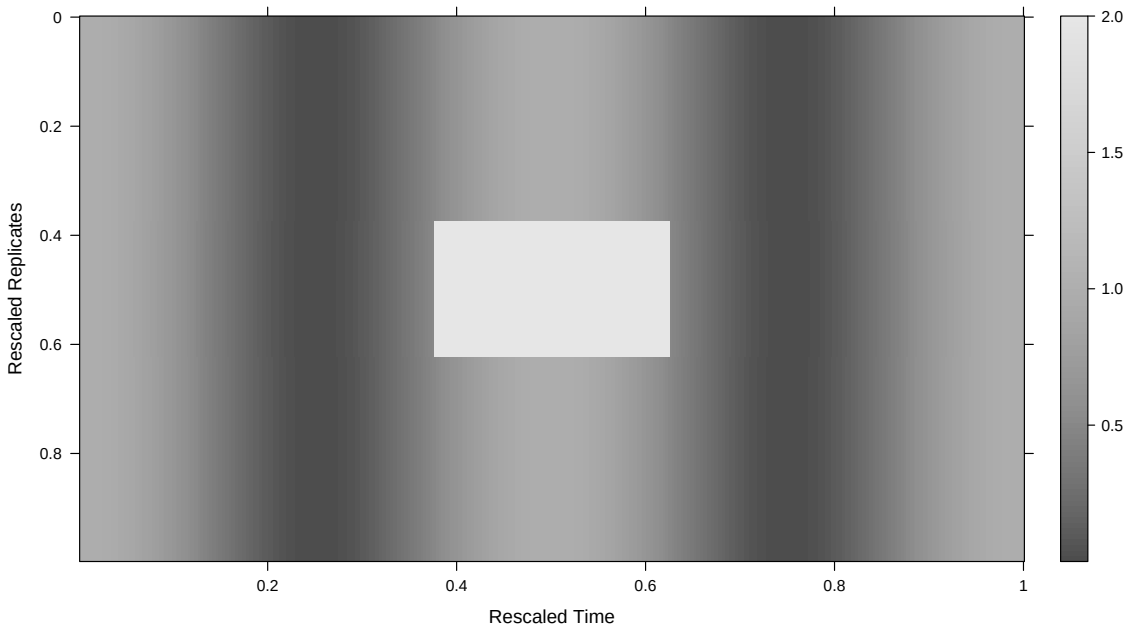


Figure 52: Time-replicate plot of the true spectra in level 5 for simulation 2.

From Figure 53, where the horizontal lines at 1/0 identify the locations expected to be rejected/not rejected, it can be seen how again the test breaks down when the burst is 1 (and thus the difference between spectra over replicates is at most 1). For burst values of 2 and then 5 we see much improvement in the number or true positives being correctly identified (rejected) but as previously noted, higher bursts correspond to more negatives incorrectly identified by the test at the location rejection boundaries.

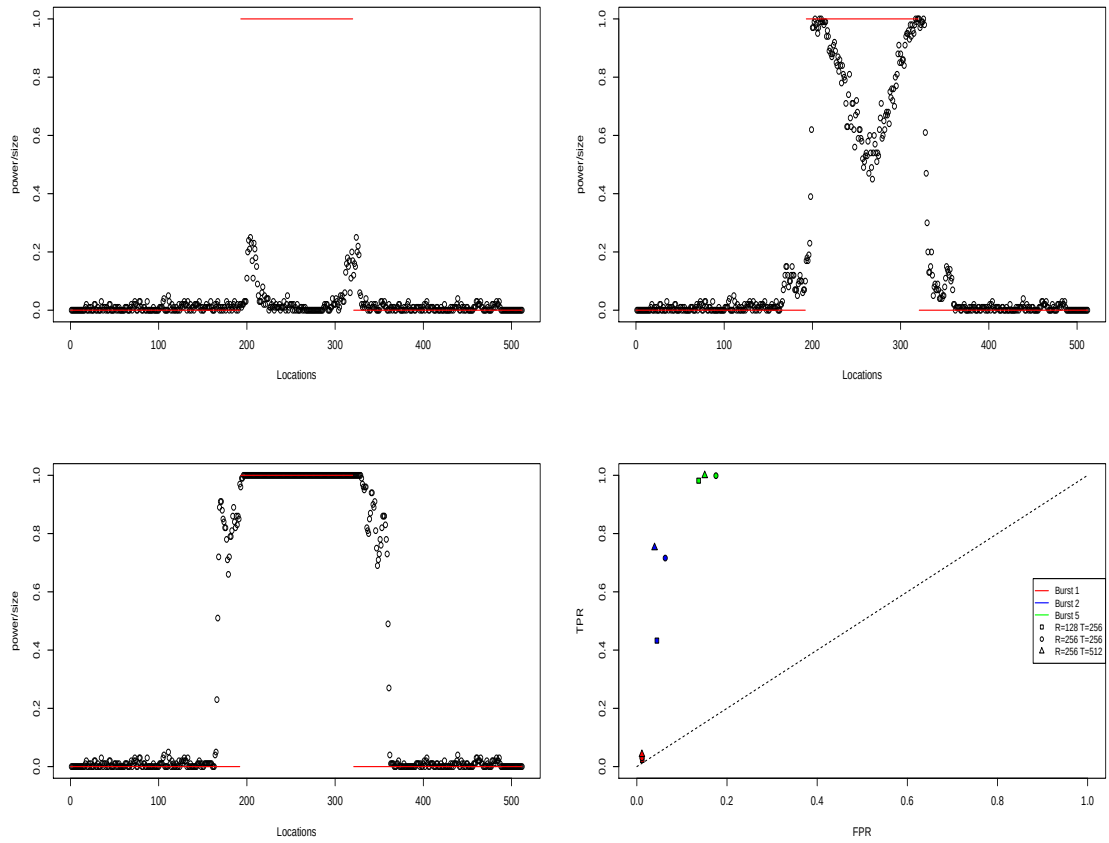


Figure 53: Simulation 2 empirical power and size estimates for each location over 100 runs. *Top left*: 'burst'= 1; *top right*: 'burst'= 2; *bottom left*: 'burst'= 5. *Bottom right*: ROC plot for simulation 2.

The statistical measures in Table 12 confirm the impact of increasing the burst, with TPR increasing and TNR decreasing. Whilst other measures remain favourable, we do note the increase in the FDR from a burst of 2 to 5, such that approximately 30% of identified positives are false. This increase moves with the decrease in TNR. However, keeping in mind the goal of the test, the jump up in TPR suggests all true positives are identified at a burst of 5 and despite the increase in FDR, the higher F-M index values indicate there is greater similarity between the identified positives and true positives clusters. Note that the presence of 'na' values is a result of the test failing to make any positive identifications on some runs of the simulation (see how the formulas in Table 8 can only be computed if $tp = fp = 0$). The ROC plot is displayed in Figure 53 (bottom right) and gives a visual comparison of the TPR and FPR for simulations with differing R, T and burst intensities. It is evident from the plot that the test performance improves (TPR increases) as R, T and the burst intensities increase. We note that this is at the expense of a minor increase in the FPR, which can be attributed to the impact of spectral leakage from the greater burst intensities.

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.0230	0.0277	0.9828	0.0145	na	0.2489	na	0.7429	na
256	256	0.03	0.0315	0.9835	0.0127	na	0.2474	na	0.7451	na
256	512	0.0420	0.0238	0.9861	0.0075	0.4443	0.2446	0.1492	0.7500	na
512	512	0.0946	0.0382	0.9817	0.0093	0.3279	0.2351	0.2477	0.7599	na

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.4325	0.1160	0.9503	0.0258	0.2318	0.1651	0.5721	0.8209	0.2485
256	256	0.7159	0.0979	0.9318	0.0233	0.2071	0.0914	0.7519	0.8779	0.2334
256	512	0.7520	0.0667	0.9581	0.0139	0.1341	0.0790	0.8060	0.9066	0.1884
512	512	0.9627	0.0222	0.9228	0.0233	0.1857	0.0132	0.8849	0.9328	0.2177

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.9816	0.0233	0.8577	0.0322	0.2919	0.0071	0.8331	0.8887	0.2736
256	256	0.9988	0.0057	0.8193	0.0117	0.3448	0.0005	0.8089	0.8642	0.2982
256	512	0.9995	0.0019	0.8468	0.0176	0.3104	0.0002	0.8301	0.8850	0.2808

Table 12: Binary classification measures averaged over 100 runs for simulation 2. *Top*: 'burst'= 1; *middle*: 'burst'= 2; *bottom*: 'burst'= 5. Prevalence = 0.25. The 'na' values are a result of the test failing to make any positive identifications (such that $tp = fp = 0$ in Table 8) on some runs of the simulation.

Simulation 3

For our third simulation we dramatically reduce the number of replicates and locations that we define to have changing spectral characteristics. The spectra are defined as follows

$$S_j(z,\nu) = \begin{cases} \cos^2(2\pi z), & \text{for } j = J(T)-4, z \in (0,1), \nu \in (0,\frac{3}{8}) \cup (\frac{4}{8}+\frac{1}{R},1) \\ & \quad \text{and } z \in (0, z^* - \frac{1}{16}) \cup (z^* + \frac{1}{16} + \frac{1}{T}, 1), \nu \in (\frac{3}{8}+\frac{1}{R}, \frac{4}{8}) \\ \text{'burst'}, & \text{for } j = J(T)-4, z \in (z^* - \frac{1}{16} + \frac{1}{T}, z^* + \frac{1}{16}), \nu \in (\frac{3}{8}+\frac{1}{R}, \frac{4}{8}) \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

where $z^* = \lfloor 0.6T \rfloor / T$ and in rescaled time $z = k/T$ and $\nu = r/R$. For a simulation with $R = 256$ and $T = 512$ we again have a squared cosine behaviour in all the replicates except for $R/8 = 32$ (here, replicates 97 to 128) where we have defined a burst of 1 in the spectra over $T/8 = 64$ locations (here, 276 to 339). Figure 54 gives a visual representation of the spectra defined for both burst intensities of 1 and 5.

From the empirical power and size estimates shown in Figure 55, it is evident that the closeness in spectra of the squared cosine and the burst value of 1 is again causing the test to struggle to reject the null hypothesis at the locations across which an evolution over the replicates exists. For burst values of 2 and 5, the test sufficiently rejects the correct locations, but present again is the effect at the location rejection boundaries.

As with the previous simulations, the measures shown in Table 13 improve for higher values of R and T, and furthermore for an increasing burst. Recalling that we have defined an evolution in the spectra over a much smaller window of replicates and locations, it is

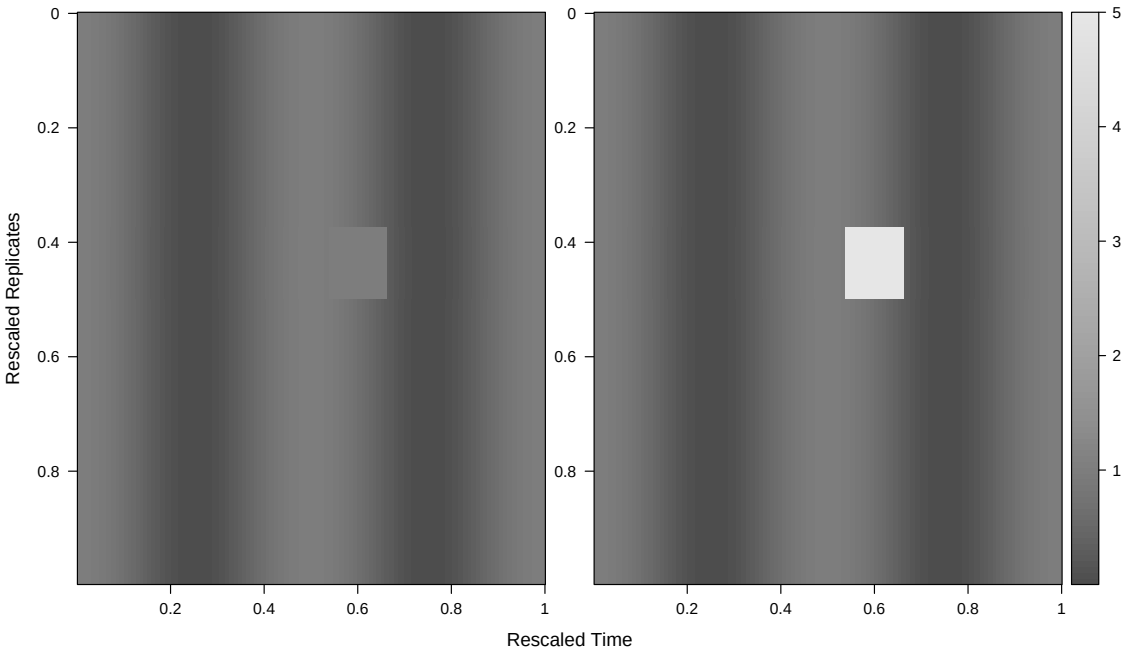


Figure 54: Time-replicate plots of the true spectra in level 5 for simulation 3. *left*: 'burst'= 1; *right*: 'burst'= 5.

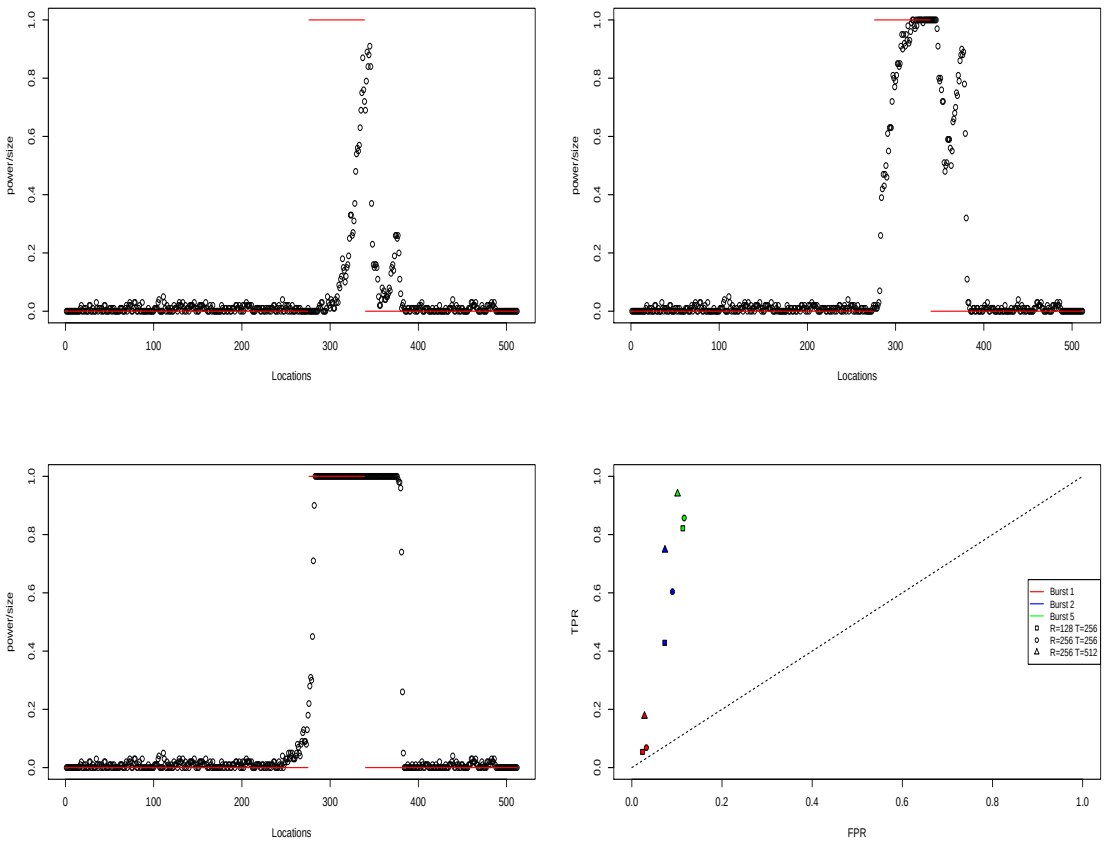


Figure 55: Simulation 3 empirical power and size estimates for each location over 100 runs. *Top left*: 'burst'= 1; *top right*: 'burst'= 2; *bottom left*: 'burst'= 5. *Bottom right*: ROC plot for simulation 3.

understandable that the measures for bursts 2 and 5 are slightly lower than in previous simulations, nevertheless the test still performs well despite the more challenging structure.

We display the corresponding ROC plot in Figure 55 (bottom right) which again highlights the improved performance from increases in R and T and furthermore the impact of greater burst intensity.

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.0538	0.0665	0.972	0.0185	na	0.1220	na	0.8572	na
256	256	0.0684	0.0632	0.9633	0.0177	0.7636	0.1213	0.1229	0.8515	na
256	512	0.1769	0.0664	0.9697	0.0130	0.5180	0.1081	0.2879	0.8706	0.2943

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.4284	0.1293	0.9226	0.0235	0.5398	0.0810	0.4408	0.8609	0.2997
256	256	0.6041	0.0901	0.9054	0.0186	0.5077	0.0587	0.5442	0.8677	0.2833
256	512	0.7473	0.0703	0.9243	0.0152	0.4045	0.0375	0.6662	0.9022	0.2405

Binary classification measures

R	T	TPR	sd_{TPR}	TNR	sd_{TNR}	FDR	FOR	F-M	ACC	PT
128	256	0.8216	0.0629	0.8825	0.0148	0.4894	0.0280	0.6472	0.8748	0.2742
256	256	0.8569	0.0543	0.8795	0.0093	0.4864	0.0227	0.6632	0.8767	0.2727
256	512	0.9402	0.0294	0.8964	0.0086	0.4295	0.0094	0.7322	0.9019	0.2490

Table 13: Binary classification measures averaged over 100 runs for simulation 3. *Top*: 'burst'= 1; *middle*: 'burst'= 2: *bottom*: 'burst'= 5. Prevalence = 0.125. The 'na' values are a result of the test failing to make any positive identifications (hence $tp = fp = 0$ in Table 8) on some runs of the simulation.

Simulation 4

We now evaluate the power of the test by means of a simulation which does not provide a burst but instead displays slowly evolving spectral characteristics over both (rescaled) time and replicates of a REv-LSW process. We define the spectra as follows

$$S_j(z,\nu) = \begin{cases} 4\nu \sin^2\left(2\pi z(1+2\nu)\right), & \text{for } j = J(T) - 4, z \in (0,1), \nu \in (0,1) \\ 0, & \text{otherwise,} \end{cases} \tag{67}$$

recalling that $z = k/T$ and $\nu = r/R$ in rescaled arguments. Figure 56 displays the true spectra for a simulation with $R = 256$ and $T = 512$. For this simulation all locations have a degree of varying spectra over the replicates and therefore we would like the test to reject all locations. This is the opposite scenario to Simulation 1A.

Figure 57 displays the empirical power estimates for the simulation with different R and T. It is clear that the test struggles for the simulation with $R = 128$ but performs well when the number of replicates increases to $R = 256$. This, as noted previously, is due to improved estimation of the spectra as $R, T \to \infty$. For both $R = 128$ and $R = 256$, note that the power estimates at locations preceding location $0.2T$ are poor indicating the test is having difficulty to identify a breach in the null hypothesis of constancy of the spectra over replicates. Comparing these locations where the test struggles to the true spectra across replicates around a rescaled time of 0.15 in Figure 56, the results suggest that the spectra evolve too slowly over the replicates to be picked up by the test. Recall

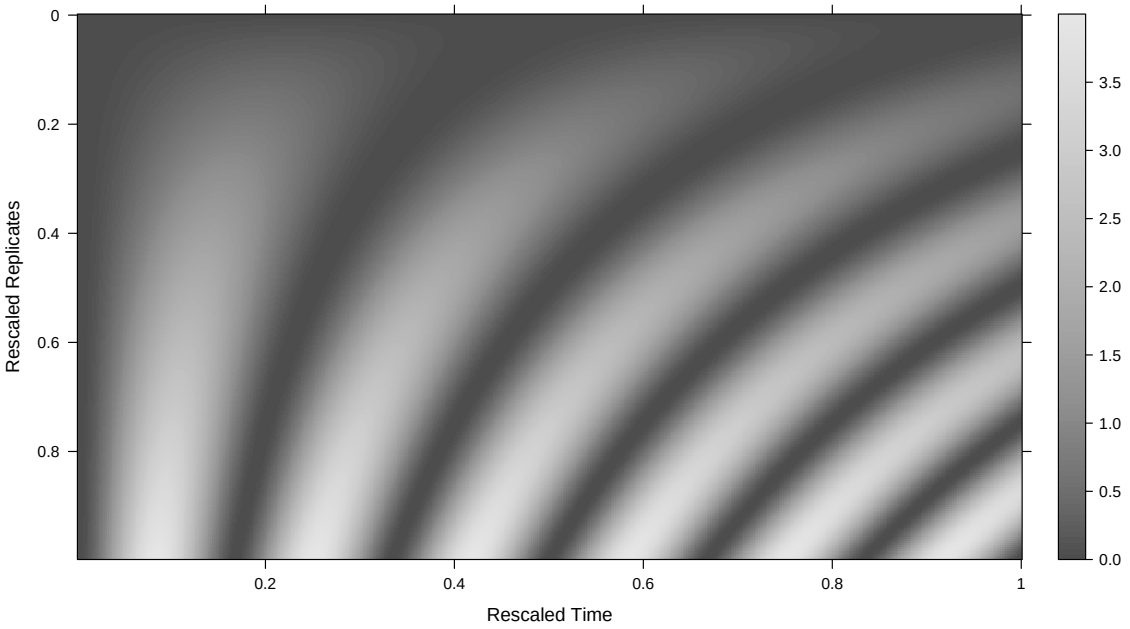


Figure 56: Time-replicate plot of the true spectra in level 5 for simulation 4.

that the test struggled in previous simulations when the difference between the burst and sine/cosine spectra was of 1 or less.

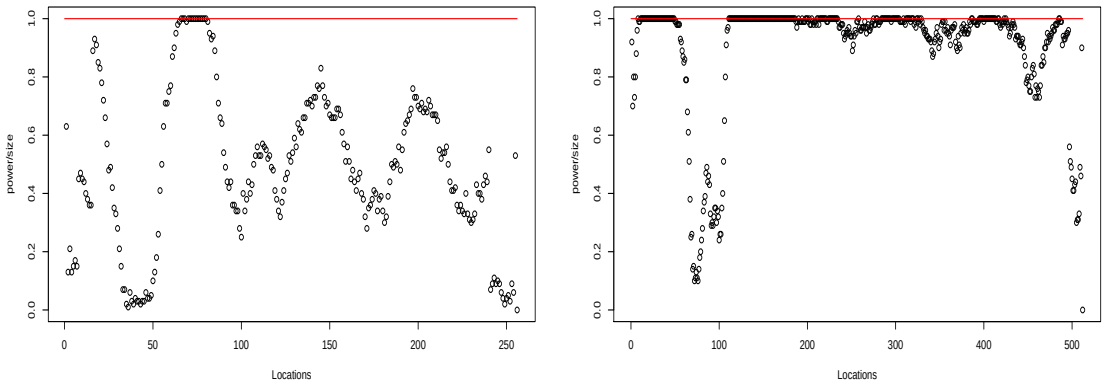


Figure 57: Simulation 4 empirical power estimates for each location over 100 runs. *left*: R=128, T=256; *right*: R=256, T=512.

In Table 14 we report the statistical measures on the test as a whole. Note that there are fewer rates reported due to the absence of true negatives (locations with constant spectra across replicates). Also note how the increase in replicates to $R = 256$ much improves the TPR and the F-M index, which we recall gives a measure of the similarity between the identified positives and true positives.

Simulation 5

Stepping up the challenge given by simulation 4, we now introduce an evolving spectral structure for two neighbouring wavelet scales and further investigate the power of the proposed test. The spectra are defined as follows

Binary classification measures

R	T	TPR	sd_{TPR}	F-M	ACC
128	256	0.4944	0.0716	0.7012	0.4944
256	256	0.8470	0.0375	0.9201	0.8470
256	512	0.8954	0.0226	0.9462	0.8954

Table 14: Binary classification measures averaged over 100 runs for simulation 4.

$$S_j(z,\nu) = \begin{cases} 4(1-\nu)\cos^2\left(\frac{-1}{3}\pi + \frac{4}{3}\pi z\right), & \text{for } j = J(T)-3, z \in (\frac{1}{4}+\frac{1}{T}, 1), \nu \in (0,1) \\ 4\cos^2\left((4\pi + 10\nu)z\right), & \text{for } j = J(T)-2, z \in (0, \frac{1}{2}), r \in (0,1) \\ 0, & \text{otherwise,} \end{cases} \tag{68}$$

recalling that $z = k/T$ and $\nu = r/R$ in rescaled arguments. The true spectra for a simulation with $R = 256$ and $T = 512$ can be seen in Figure 58. Once again we expect the test to make rejections for all locations. For different R and T, the empirical power estimates are displayed in Figure 59. The results here follow in similar vein to simulation 4, such that as estimation improves asymptotically with $R, T \to \infty$, the performance of the test improves. Furthermore we again have evidence that the test does not always correctly reject when the evolution of the spectra over the replicates is too slow, specifically between rescaled time points (approximately) between 0 and 0.05 and between 0.15 and 0.2 in Figure 58, comparable with the same rescaled time points in Figure 59. Statistical measures on the global test performance are given in Table 15 where again fewer rates are reported due to the absence of any true negatives. Both the TPR and F-M demonstrate the test performs well, with improvements obtained with increases in R and T.

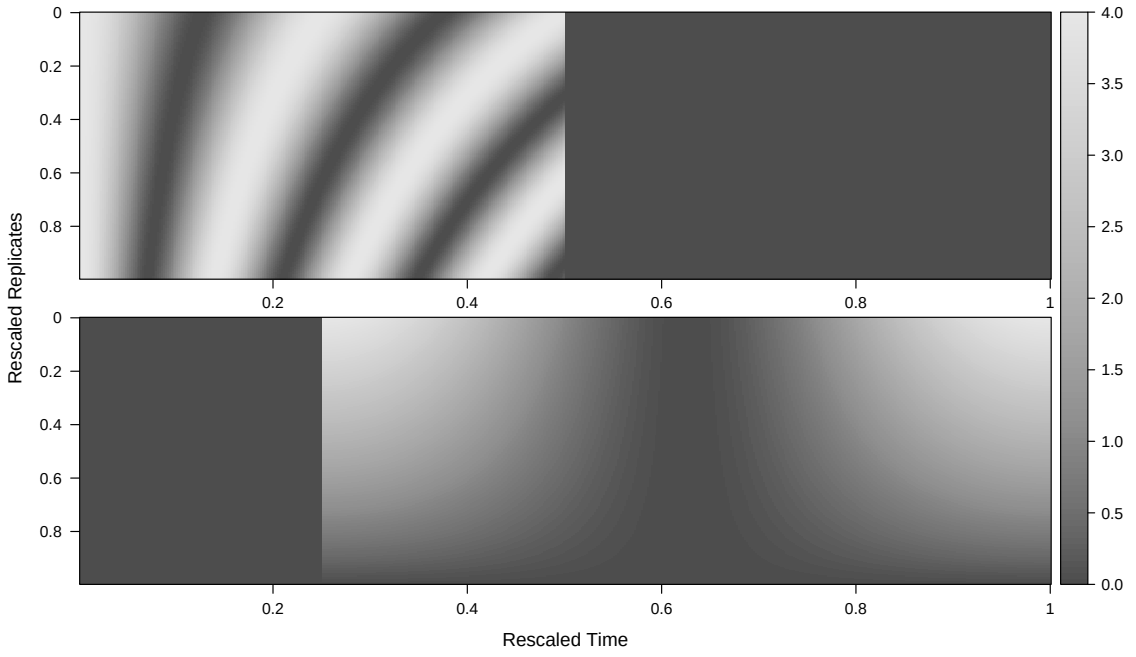


Figure 58: Time-replicate plots of the true spectra in levels 7 (*top*) and 6 (*bottom*) for simulation 5.

Binary classification measures

R	T	TPR	sd_{TPR}	F-M	ACC
128	256	0.4916	0.0568	0.6999	0.4916
256	256	0.7980	0.0207	0.8933	0.7980
256	512	0.8042	0.0162	0.8967	0.8042

Table 15: Binary classification measures averaged over 100 runs for simulation 5.

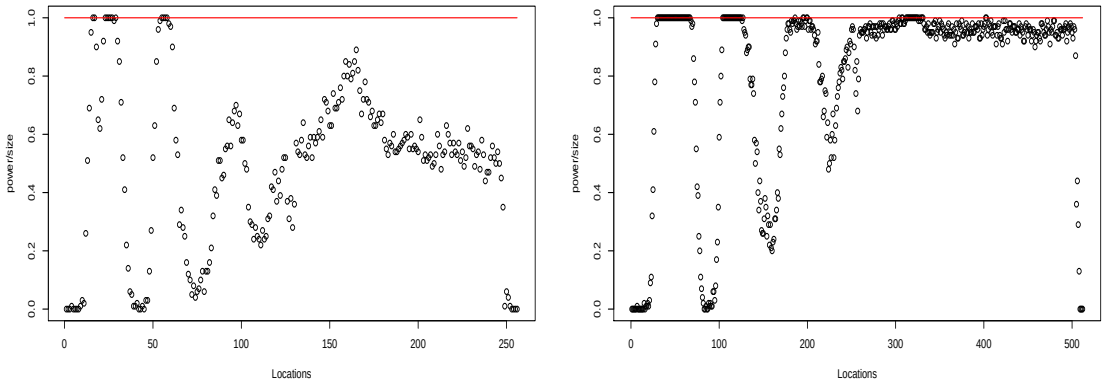


Figure 59: Simulation 5 empirical power estimates for each location over 100 runs. *left*: R=128, T=256; *right*: R=256, T=512.

Concluding remarks

Overall, the simulation study has shown that the location-specific test can correctly identify whether there exists spectral evolutionary behaviour over the replicates. When there is in fact spectral constancy across the replicates, the test successfully identifies the lack of replicate-effect. We recall that the test falsely identified spectral evolution over the replicates at the location rejection boundaries due to power leakage during estimation. However despite this, we note that in order for this to occur, there must exist an evolution in the spectra over the replicates for some locations. We recommend the use of this test if the analysis aim is to gain an indication as to whether there is any evolution in the spectral characteristics over replicates. However, this test goes further and is also able to give a good identification of the times and replicates at which an evolutionary behaviour over the spectra exists. Whilst the location-specific test is most informative, the next section proposes a global test for replicate-effect, offering the user a direct diagnostic as to whether a replicate-effect exists as opposed to assessing the individual locations.

5.3 A global approach to testing for replicate-effect

Recall that so far we aimed to identify those times k for which the null hypothesis, $H_0^k : S_j\left(\frac{k}{T}, \nu\right)$ is constant as a function of $\nu, \forall j$, holds. However, depending on the context, one might find it more useful to test the overall null hypothesis $H_0 : S_j\left(\frac{k}{T}, \nu\right)$ is constant as a function of ν, $\forall j$, k. In this setting, we are again modelling under the assumption that replicates are uncorrelated.

In order to propose an appropriate test statistic for this over-arching hypothesis, let us

first consider the following measure for the spectral departure from constancy through the replicate-dimension for a particular time k (rescaled time $z = \frac{k}{T}$) and within each scale j,

$$\mathbf{T}\left(S_j(z,\cdot)\right) = \int_0^1 \left(S_j\left(z,\nu\right) - \bar{S}_j\left(z\right)\right)^2 d\nu,$$

where we denote by $\bar{S}_j\left(z\right) = \int_0^1 S_j\left(z,\nu\right) d\nu$ (see also Cardinali and Nason (2010) for an alternative context where such a measure proved successful).

The *average spectral departure from constancy across all scales* may then be measured by

$$\mathbf{T}_{ave}\left(\left\{S_j\left(z,\cdot\right)\right\}_j\right) = J^{-1}\sum_{j=1}^{J}\mathbf{T}\left(S_j\left(z,\cdot\right)\right),$$

where we recall that $J(T) = \log_2(T)$. Furthermore, one could quantify the *overall spectral departure* not only across all scales but also through time, which for brevity we denote $\mathbf{T}_{ave}\left(S\right)$, as follows

$$\begin{aligned}\mathbf{T}_{ave}\left(S\right) &= \int_0^1 \mathbf{T}_{ave}\left(\left\{S_j\left(z,\cdot\right)\right\}_j\right) dz, \\ &= J^{-1}\sum_{j=1}^{J}\int_0^1 \mathbf{T}\left(S_j\left(z,\cdot\right)\right) dz.\end{aligned} \tag{69}$$

Under the null hypothesis (H_0) defined above, note that all measures above are 0. Conversely, since the spectral quantities are positive it is straightforward to show that if the aggregated measure $\mathbf{T}_{ave}\left(S\right)$ in (69) is 0, then the null hypothesis of spectral constancy through replicates at all times also holds. Hence we shall treat significant departures from 0 as indicative of departures from the null hypothesis.

As the true replicate wavelet spectra are unknown, we estimate the measures of spectral departure from constancy above by means of their *sample* equivalents, built upon a well-behaved spectral estimator (see for instance Proposition 2.1.16 in Section 2.1). Hence we obtain

$$\begin{aligned}\mathbf{T}\left(\hat{S}_{j,k}\right) &= \mathrm{var}_\nu\left(\hat{S}_j\left(\frac{k}{T},\nu\right)\right), \text{ as an estimate for } \mathbf{T}\left(S_j\left(\frac{k}{T},\cdot\right)\right), \\ \mathbf{T}_{ave}\left(\left\{\hat{S}_{j,k}\right\}_j\right) &= J^{-1}\sum_{j=1}^{J}\mathrm{var}_\nu\left(\hat{S}_j\left(\frac{k}{T},\nu\right)\right), \text{ for } \mathbf{T}_{ave}\left(\left\{S_j\left(\frac{k}{T},\cdot\right)\right\}_j\right), \\ \mathbf{T}_{ave}\left(\left\{\hat{S}_{j,k}\right\}_{j,k}\right) &= (JT)^{-1}\sum_{k=0}^{T-1}\sum_{j=1}^{J}\mathrm{var}_\nu\left(\hat{S}_j\left(\frac{k}{T},\nu\right)\right), \text{ for } \mathbf{T}_{ave}\left(S\right),\end{aligned}$$

where in the above $\hat{S}_j\left(\frac{k}{T},\nu\right)$ is our proposed smoothed and corrected wavelet periodogram at rescaled replicate ν and var_ν denotes the usual empirical variance, here taken over replicates within each scale j and at each time k.

Although under the null hypothesis the distribution of the test statistic $\mathbf{T}_{ave}\left(\hat{S}\right)$ is unknown, since we have shown that for all replicates ν, $\left\{\hat{S}_j\left(\frac{k}{T},\nu\right)\right\}_{j,k}$ is a consistent estimator for the true spectrum $\left\{S_j\left(\frac{k}{T},\nu\right)\right\}_{j,k}$, the bootstrap approach of Davison and

Hinkley (1997) is a valid alternative which we propose to use here (see also Cardinali and Nason (2010)). In order to allow for parametric resampling, we carry out the bootstrap simulations under a Gaussian innovations assumption, which we note that is a non-limiting assumption as discussed in Chapter 2 (Remark 2.1.9). The statistical significance of the observed test statistic is then established by means of a Monte Carlo approach used to generate pseudo-test statistics values from a process with properties akin to those of the original process under the null hypothesis. The resulting p-value is then simply computed using a count of these pseudo-test statistics that exceed the value of the observed test statistic.

The proposed algorithm to test whether a replicate-effect is present, appears detailed in Algorithm 1 below.

`Proposed BootReplicateTest algorithm:`
Assume we observe a REv-LSW process as in Definition 2.1.1 and we want to assess whether a replicate-effect exists.

0. Obtain a well-behaved spectral estimator $\hat{S}_j(k/T, r/R),\ \forall j,\ k,\ r$.
1. For each scale j and time k, under the null hypothesis of constancy through replicates, compute the average scale- and time- specific periodogram,
   $\bar{\hat{S}}_j(\frac{k}{T}) = R^{-1}\sum_{r=0}^{R-1}\hat{S}_j(k/T, r/R)$.
2. Compute the test statistic $\mathbf{T}_{ave}\left(\left\{\hat{S}_j(\frac{k}{T},\cdot)\right\}_j\right)\left(=\mathbf{T}_{ave}^{obs;k}\right)$ across all times k, and aggregate these into
   $\mathbf{T}_{ave}^{obs} = ((R-1)JT)^{-1}\sum_{j=1}^{J}\sum_{k=0}^{T-1}\sum_{r=0}^{R-1}\left(\hat{S}_j(k/T, r/R) - \bar{\hat{S}}_j(\frac{k}{T})\right)^2$.
3. *Iterate* for $b = 1$ to B bootstraps:
   - for each scale j, simulate a REv-LSW process $\{X_{t;T}^{r;R}\}^{(b)}$ with squared amplitudes given by $\bar{\hat{S}}_j(\frac{k}{T})$ for all r,
   - compute the corresponding test statistic $\mathbf{T}_{ave}^{(b)}$ corresponding to the simulated data.
4. Compute the test $p - value = (1 + \#\{\mathbf{T}_{ave}^{(b)} \geq \mathbf{T}_{ave}^{obs}\})/(B+1)$.

Algorithm 1: Proposed bootstrap test for assessing the existence of a replicate evolution across and REv-LSW process.

5.3.1 Simulation study for the global testing methodology

Our attention now turns to investigating the performance of our global testing approach. For various combinations of $R = 64$ to 256 and $T = 128$ to 512, we carry out the proposed testing procedure with smoothing performed over both the replicate and time domains, thus yielding the most accurate and consistent estimate of the wavelet spectrum as documented in Remark 2.1.15. We do this for $N = 100$ runs with $B = 100$ bootstraps within each run. Additionally it has been shown that the length of the replicate smoothing window, $(2M+1)$, and thus our choice of M, can lead to improved estimation, however for the following simulations we maintain a value of $M = 4$ but keep in mind that our test performance could improve with a different choice for M. Furthermore, unlike for the location-specific test where Haar wavelets alone were suitable for estimation, here it is preferable to choose the wavelet family most reflective of the process smoothness as this will yield the best

spectral estimators. For each simulation we report empirical size and power estimates obtained through counting the number of times the test rejects the null hypothesis of constancy (over replicates) at a significance level of 5%.

5.3.1.1 Simulations to investigate size

We first aim to assess how the test performs in the absence of any spectral evolution across the replicates. To do so we consider the following setups. We purposely include processes that we know depart from the REv-LSW assumptions, in order to probe the test performance outside of this framework.

s1. Simulated processes as detailed in Simulation 1A of Section 5.2.1.

s2. (a) An autoregressive AR(1) process, $X_t^{r;R} = \gamma X_{t-1}^{r;R} + \epsilon_t^r$, with AR parameter $\gamma = 0.9$, independent standard normal innovations.

(b) As above but with AR parameter $\gamma = -0.3$.

One concatenated realisation of the meta-process, viewed as a series of length RT, can be seen in Figure 60.

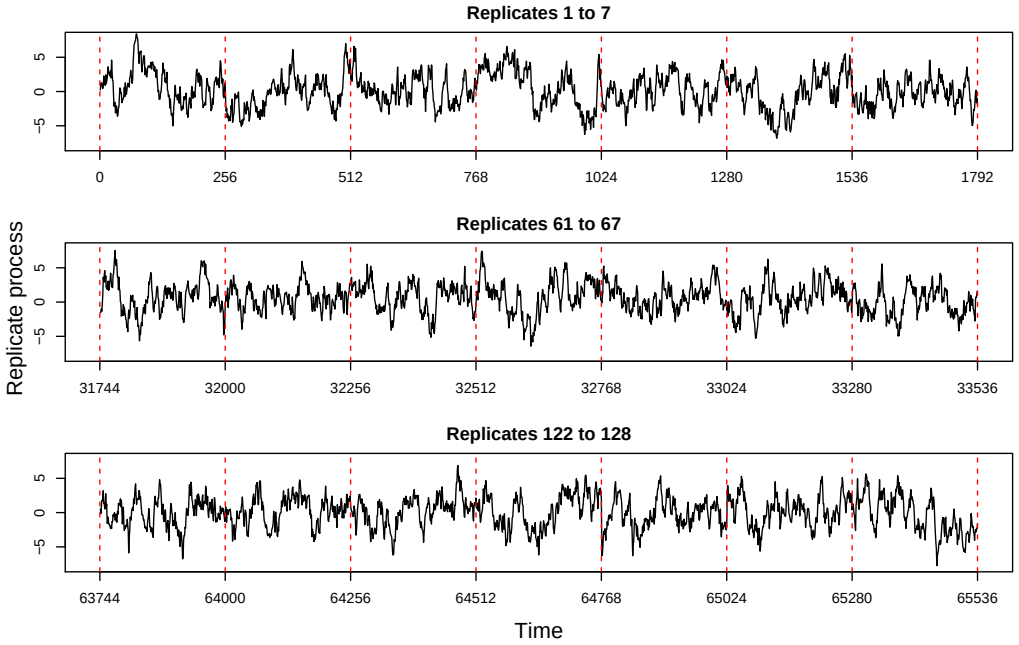


Figure 60: Concatenated series of a replicate process simulated from s2(a).

s3. (a) A time-varying autoregressive AR(1) process, $X_t^{r;R} = \gamma_t X_{t-1}^{r;R} + \epsilon_t^r$, with AR parameter γ_t evolving linearly within each replicate from 0.9 to -0.9, with independent standard normal innovations.

(b) As above but with AR parameter γ_t evolving constantly from 0.3 to -0.3.

One concatenated realisation of the meta-process, viewed as a series of length RT, can be seen in Figure 61.

Results. We compute our spectral estimates using the discrete non-decimated wavelets built by means of Daubechies Least asymmetric wavelet family with 10 vanishing moments

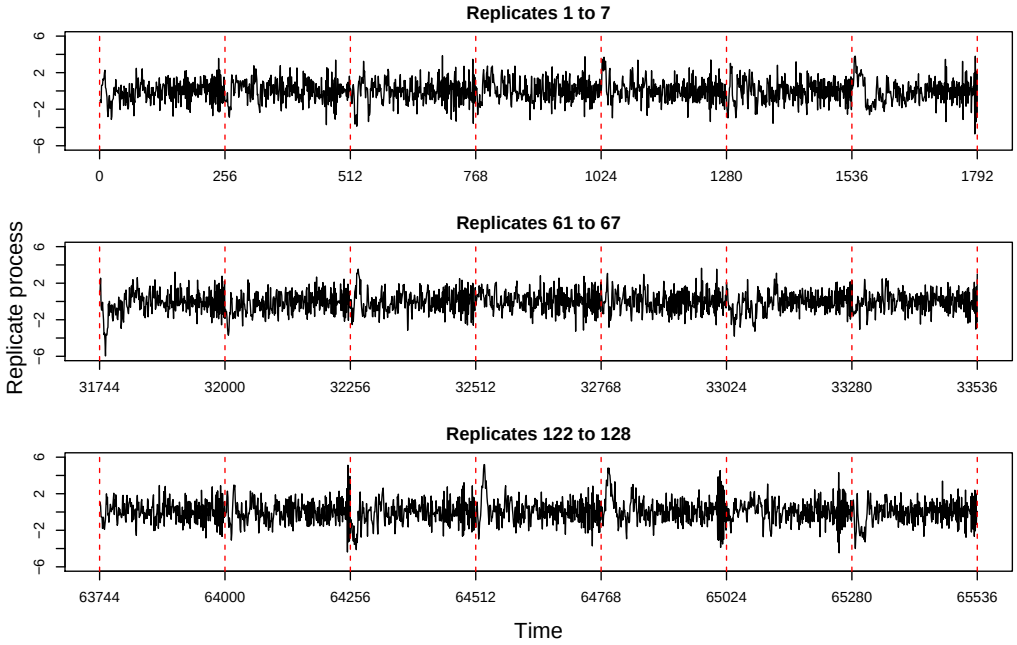


Figure 61: Concatenated series of a replicate process simulated from s3(a).

(DLA10) (Daubechies, 1992) for s1, whilst Haar wavelets were chosen for s2(a,b) and s3(a,b). These choices of wavelet were made in order to reflect the behaviour of the process, for instance DLA10 may be chosen for smooth processes whereas Haar wavelets may be preferable for choppier processes. In Table 16 we report the empirical size estimates. For the majority of the simulations, the size estimates are less than the 5% nominal size and thus indicate that the test does a good job of not incorrectly rejecting the hypothesis that spectra are constant across the replicate dimension. The one exception is simulation s3(a), for which the size values range from 27% to 60%.

		Size				
R	T	s1	$s2_a$	$s2_b$	$s3_a$	$s3_b$
64	128	0	0.01	0.03	0.27	0.05
128	128	0	0.03	0.04	0.33	0.08
128	256	0	0	0.02	0.4	0
256	256	0	0.01	0.01	0.6	0.01
256	512	0	0.01	0	0.6	0

Table 16: Empirical size values computed over 100 runs for simulations s1 - s3(a,b). Spectral estimation was via the REv-LSW model with smoothing over replicates and time.

It is not obviously clear why the global test is struggling with simulation s3(a) however a similar scenario occurs for the simulations using the Haar wavelet-based test of stationarity proposed by Nason (2013) on a single AR(1) process with an AR parameter of -0.9. Nason (2012) investigates this scenario and draws attention to the possible 'volatility clustering' exhibited by the process which is comparable to behaviour seen in financial time series and GARCH models (e.g. see Chapter 3 of Tsay (2010)). For a series of length $T = 512$, Nason (2012) finds that the empirical size value of approximately 20% decreases with decreasing the AR parameter to -0.8, down to an empirical size of approximately 4%.

For simulation s3(a), Figure 61 displays the gradual volatility clustering occurring

over time within each replicate, with the clustering becoming more apparent as the AR parameter approaches -0.9. As with Nason (2012), an improvement in the empirical size values can be obtained through decreasing the negativeness of the AR parameter. When the AR parameter γ_t is defined to evolve linearly from 0.9 to -0.9, for $R = 128$ and $T = 256$ the empirical size in Table 16 is shown to be 40%. Setting the AR parameter to evolve linearly from 0.9 to -0.8 yields an empirical size value of 18%, further still reducing if the parameter is set to evolve linearly from 0.9 to -0.7, down to an empirical size of 12%. If we consider $R = 256$ and $T = 512$, the empirical size is shown to be 60% when the AR parameter evolves linearly from 0.9 to -0.9. When set to evolve to -0.8, we obtain an empirical size of 28%, while evolution to -0.7 yields an empirical size of 16%. The additional estimated size values have not fallen below 5%, however they do demonstrate that the test dramatically improves when we set the AR parameter to evolve to a lesser extreme than -0.9, analogous to the results found in Nason (2012).

5.3.1.2 Simulations to investigate power

To examine how well our global test identifies a breach in constancy of the spectral estimates over replicates, or in other words whether there exists spectral evolution over the replicates, we consider the following setups, including process that both adhere to (p1-p3) and fall outside of (p4-p5) REv-LSW assumptions.

p1. Simulated processes as detailed in Simulation 1B of Section 5.2.1 with burst values of 1 (b1) and 2 (b2) .

p2. Simulated processes as detailed in Simulation 4 of Section 5.2.1.

p3. Simulated processes as detailed in Simulation 5 of Section 5.2.1.

p4. (a) A 'replicate-varying' autoregressive AR(1) process, $X_t^{r;R} = \gamma^r X_{t-1}^{r;R} + \epsilon_t^r$, with AR parameter γ^r evolving linearly from 0.9 to -0.9, with independent standard normal innovations.

(b) As above but with AR parameter γ^r evolving linearly from 0.3 to -0.3.

One concatenated realisation of the meta-process, viewed as a series of length RT, can be seen in Figure 62.

p5. (a) A 'time-replicate-varying' autoregressive AR(1) process, $X_t^{r;R} = \gamma_t^r X_{t-1}^{r;R} + \epsilon_t^r$, with independent standard normal innovations and the AR parameter γ_t^r evolves linearly across both time and replicate dimensions, visually represented as follows

$$\gamma_{t;T}^{r;R} = \begin{bmatrix} 0.9 & \cdots & \gamma_t^1 & \cdots & -0.9 \\ \vdots & & \vdots & & \vdots \\ \gamma_1^r & \cdots & \vdots & \cdots & \gamma_T^r \\ \vdots & & \vdots & & \vdots \\ -0.9 & \cdots & \gamma_t^R & \cdots & 0.9 \end{bmatrix},$$

with $t = 0, ..., T-1$ and $r = 0, ..., R-1$. For each row, the AR parameter evolves in equal increments of $|\gamma_1^r - \gamma_T^r|/T$ across the columns. For each column t, the AR parameter evolves in equal increments of $|\gamma_t^1 - \gamma_t^R|/R$ across the rows.

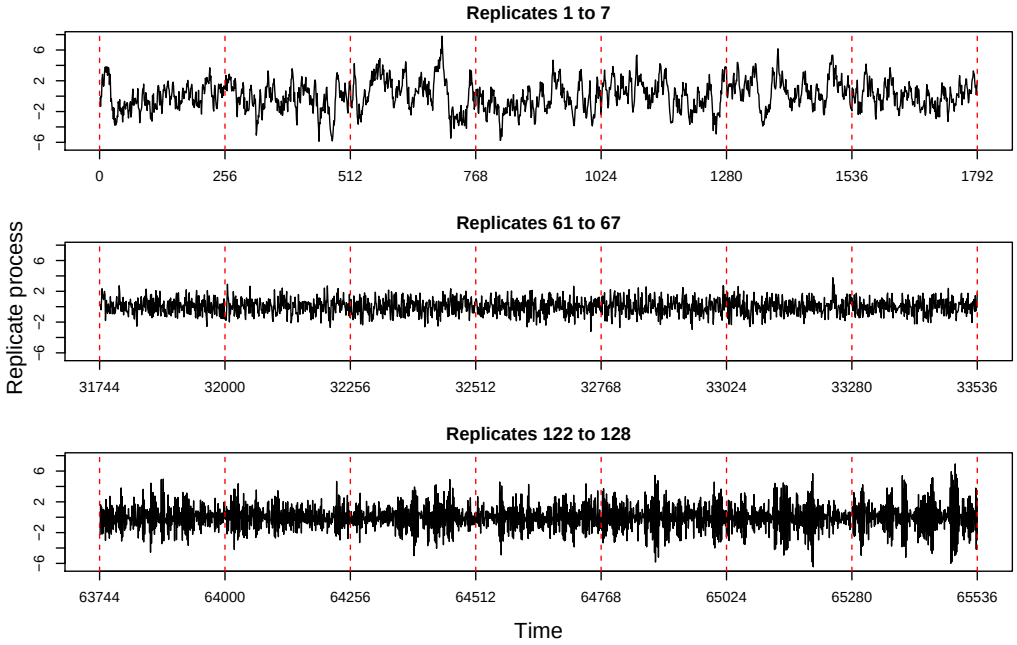


Figure 62: Concatenated series of a replicate process simulated from p4(a).

(b) As above but with matrix elements 0.9 and -0.9 being replaced with 0.3 and -0.3 respectively.

One concatenated realisation of the meta-process, viewed as a series of length RT, can be seen in Figure 63.

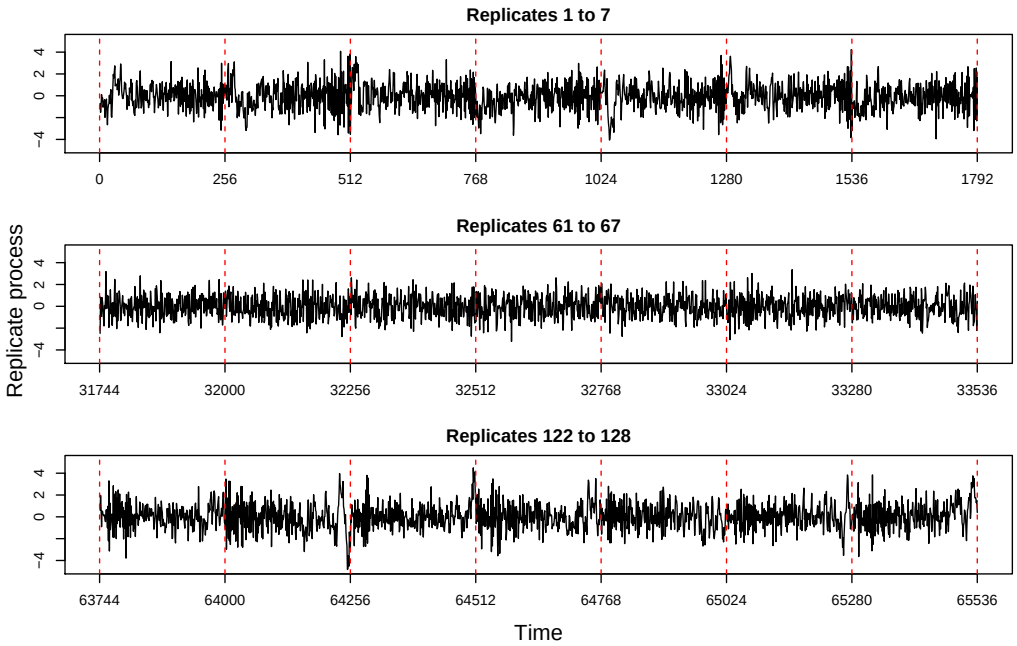


Figure 63: Concatenated series of a replicate process simulated from p5(a).

Results. We compute our spectral estimates using discrete non-decimated wavelets chosen to best reflect the behaviour of the process. The discrete non-decimated wavelets were built by means of Daubechies Least asymmetric wavelet family with 10 vanishing moments for p2 and 6 vanishing moments for p3, whereas Haar wavelets were chosen for

		Power							
R	T	$p1_{b1}$	$p1_{b2}$	p2	p3	$p4_a$	$p4_b$	$p5_a$	$p5_b$
64	128	0.05	0.38	0.97	1	1	0.97	1	0.3
128	128	0.08	0.97	1	1	1	1	1	0.76
128	256	0.23	1	1	1	1	1	1	0.97
256	256	0.79	1	1	1	1	1	1	1
256	512	0.99	1	1	1	1	1	1	1

Table 17: Empirical power values computed over 100 runs for simulations p1 - p5(a,b). Spectral estimation was via the REv-LSW model with smoothing over replicates and time.

the more volatile processes p1(b1,b2), p4(a,b) and p5(a,b). In Table 17 we report the empirical power estimates, which suggest the test performs well for most setups. The values for simulation p1(b1) is the only concern when R is low, however this was not unexpected considering the previous results of the location-specific test for the same setup (see Simulation 1B, Section 5.2.1). We recall that there the test broke down when the difference between spectra across the replicates assumed values less than 1. However, the power values do improve as our sample size increases (R, $T \to \infty$).

Concluding remarks

The simulation study has demonstrated that the global test, through encapsulating the spectral characteristics of the whole REv-LSW process, is capable of answering the general question of 'does a replicate-effect exist'. However, the unfavourable empirical size values for simulation s3(a) have demonstrated that the test can fail when the process veers too far from the smooth evolution assumptions of an REv-LSW process (see Remark 2.1.2). As the global test relies on obtaining consistent estimates of the spectra, we apply both time and replicate smoothing, which can introduce edge effects at the time and replicate domain edges, respectively. Our methodology so far does not address how to handle replicate edge effects and so even though the global test performs well already, there is room for improvement. For instance, we could consider a method that internally reflects the replicate time series at the edges in the replicate domain and then cut these artificial replicates after estimation. Alternatively, as in Fryzlewicz and Ombao (2009) we might perform the test on a suitably selected subset of the wavelet scales j, locations k and replicates r. Both the global and location-specific tests have proven to be successful within the simulation studies and the next section will illustrate the tests when applied to real data.

5.4 Analysis of Macaque Local Field Potentials: Test of replicate-effect

Our simulation studies have given a thorough demonstration on the performance of both the location-specific and global tests aiming to identify the existence of replicate spectral evolution across the experimental timeline. Under the assumption of uncorrelated replicates, we now illustrate both tests on the macaque brain processes data introduced and analysed in Chapter 4. We recall that our real data analysis was carried out over sets of trials grouped chronologically for correct and incorrect responses of the macaque. Each trial was of length $T = 2048$ (hence $J = 11$) and each group (correct and incorrect) consisted of $R = 256$ (hence $J' = 8$) trials, thus forming our replicated time series that we model using the REv-LSW framework developed in Chapter 2.

5.4.1 Correct trials for the hippocampus

In Chapter 4 we estimated the spectral characteristics of the Hc correct trials by means of non-decimated discrete wavelets from Daubechies Least Asymmetric family with 10 vanishing moments (DLA10), applied smoothing over both time and replicate domains with a replicate local averaging window over 21 trials ($M = 10$ neighbouring trials), and finally we corrected for bias. The process evolutionary behaviour was captured visually in Figure 37 (Chapter 4). The question we are aiming to answer is whether or not there is sufficient evidence to deem that spectral evolution across the trials (replicates) is indeed manifest and thus the macaque's memory recall evolves through the experiment. We will apply both the location-specific and global tests, and if appropriate, identify the times and trials for which the memory recall becomes manifest.

Location-specific test

For our location-specific test, recalling its construction, we compute our spectral estimates using Haar wavelets instead of DLA10 and additionally as smoothing is not required for this testing procedure we note that the estimated spectra are the raw wavelet periodograms. For each location there are a total of 40 hypothesis tests over 8 original scales and 5 Haar scales (recall equation (62)) and we are in the field of multiple hypothesis testing which we deal with by taking an FDR approach.

In plot (a) of Figure 64 we present a visualisation of the number of locations identified by the test as rejecting the null hypothesis of spectral constancy over the trials. The total number of locations rejected was 1153. Here, a location is classed as rejected if any one of the hypothesis tests for that location is rejected. This approach is a little naive and at most tells us that our test 'successfully' identifies numerous locations across the experiment that exhibit spectral evolution over the trials. What qualifies as a 'successful' identification is open to interpretation, for instance whether 1 rejection out of 40 hypothesis tests is satisfactory to classify a location as having a significant replicate-effect. Thus, plot (b) of Figure 64 gives a more informative visualisation through displaying the percentage of the 40 hypothesis tests that were rejected for each location, while in plot (c) we threshold the percent values at 25%. By considering the percentage of tests rejected, we can identify the locations that give a strong indication of a replicate-effect which could then be the focus for further analysis. Comparing plot (b) with the spectral estimates in Figure 37

corroborates the evolutionary patterns we see there. Interestingly, the test highlights locations around timepoint 512, which we recall corresponds to the visual exposure time block, and in the final quarter time block which we recall corresponds to the macaque making correct associations. Thus the test appears to confirm the experimental design, identifying the locations where we would expect a replicate-effect to be present. However, we reinforce our previous point on the flexibility of choosing a threshold that would decide whether a location should be deemed as 'successfully' rejecting the null hypothesis and currently this is left to the discretion of the user. It is not unreasonable to assume that if a significant replicate-effect exists for one location, then the spectral characteristics of neighbouring locations may echo a similar replicate-effect. As such, in plots (d), (e) and (f) of Figure 64 we choose to only display the locations rejected that are in the vicinity of four (two each side) rejected locations. This process of eliminating the weakly rejected locations (displayed throughout plots (a)-(f)) gives a way to identify the locations with the strongest evidence for the existence of a replicate-effect.

As well as identifying the locations where a replicate-effect exists, the location-specific test is also capable of indicating the trials for which the replicate-effect was detected. In Figure 65, for specific time locations 560 (left) and 1610 (right), indicative of the experimental blocks corresponding to picture exposure and exercised choice respectively, we plot the observed data as a time series across the replicate domain. The double headed arrows indicate the replicates for which evolutionary spectra were detected, with the width determined by the support of the underlying Haar wavelet. The right-hand axis gives the wavelet scale j of the spectral estimates tested and the vertical position of the arrow within each wavelet scale indicates the Haar scale i with the topmost arrow corresponding to the finest of the Haar scales tested. Both locations, 560 and 1610, have over 50% of hypothesis tests rejected under the FDR control, with the majority of rejections occurring over the final 128 replicates, indicating memory recall activation towards the end of the experiment. Many rejections occur for a narrower window of 16 replicates ($256/2^4$, where 4 is the finest Haar scale tested) at the end of the replicates for location 560 and around replicate 135 for location 1610.

Global test

For our global test, we want to obtain the best estimates for the spectral characteristics and thus we use the estimation procedure inclusive of smoothing and correction, as described above. To test for a global replicate-effect we proceed as follows. For $B = 100$ bootstraps, we bootstrap the process under the null hypothesis that assumes spectral constancy over the replicates, $\bar{\hat{S}}_j(\frac{k}{T})$, and estimate the corresponding process spectra. Next, we then compute the test p-value by means of step 4 in `BootReplicateTest` (of algorithm 1). For the correct trials of the Hc data we obtain a p-value of < 0.01 which gives a strong indication that a replicate-effect exists, in agreement to the findings of the location-specific tests. Figure 66 displays the histograms of the bootstrap test statistics benchmarked against the observed test statistics.

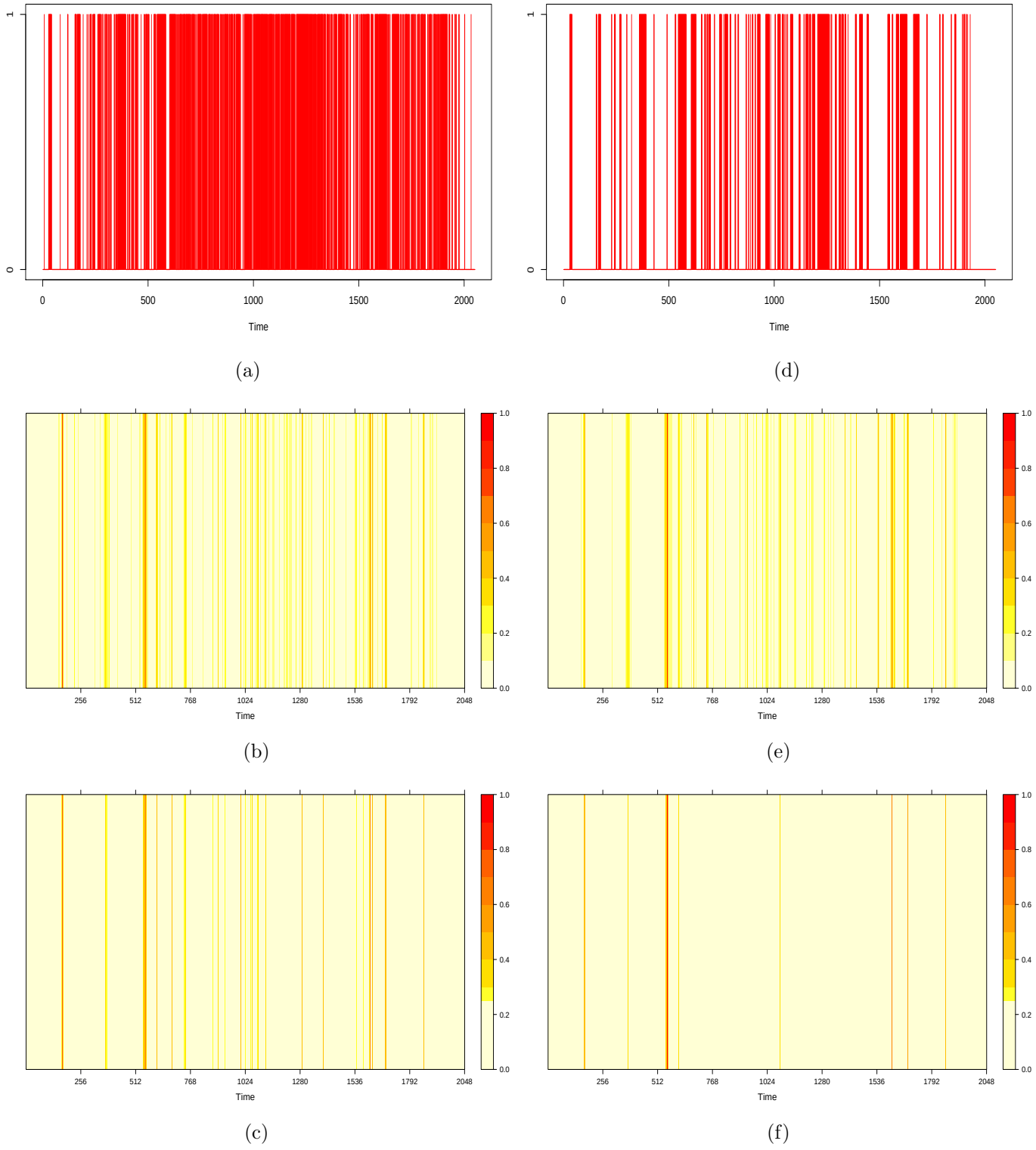


Figure 64: Location-specific rejection plots for the replicate-effect test carried out on the correct trials of the hippocampus (Hc) dataset. *(a)*: binary plot where a vertical line to 1 indicates that the test identified the location as rejecting the null hypothesis of constancy; *(b)*: percentage of hypothesis tests rejected for each location; *(c)*: percentage of hypothesis tests rejected for each location thresholded at 25%; *(d)*: as in (a) but only for rejected locations with two rejected neighbour locations each side; *(e)*: as in (b) but only for rejected locations with two rejected neighbour locations each side; *(f)*: as in (c) but only for rejected locations with two rejected neighbour locations each side. The scale for plots (b), (c), (e) and (f), indicates the percentage of rejections out of 40 hypothesis tests per location.

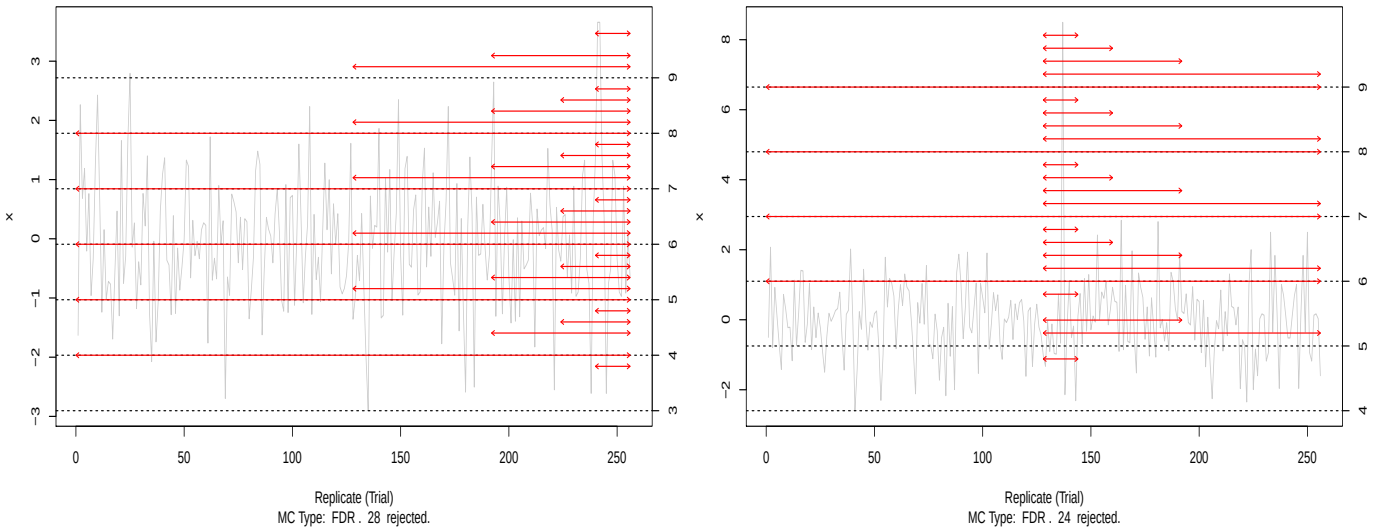


Figure 65: Replicate domain plots for correct hippocampus trials at times 560 (*left*) and 1610 (*right*). Trials where replicate-effect detections were made are indicated by the double headed arrows, and the corresponding wavelet scales $j = 3, ..., 9$ are indicated on the right axis. Within each wavelet scale, the vertical position of the arrow indicates in ascending order the Haar scales $i = 0, ..., 4$.

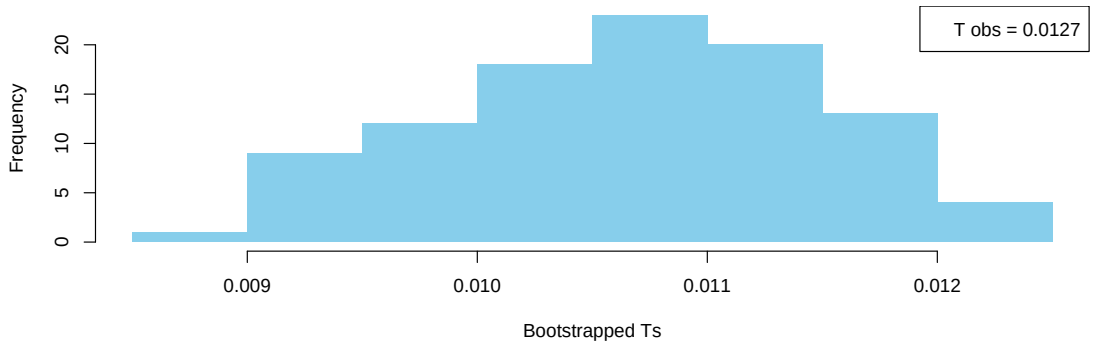


Figure 66: Histogram of the bootstrap test statistics for the global replicate-effect test carried out on the correct trials of the hippocampus (Hc) dataset.

5.4.2 Correct trials for the nucleus accumbens

Previously in Chapter 4, we estimated the spectral characteristics of the NAc correct trials by means of non-decimated discrete wavelets from Daubechies Least Asymmetric family with 6 vanishing moments (DLA6), applied both time and replicate ($M = 10$) smoothing, and finally corrected for bias. The evolutionary behaviour was captured visually in Figure 41 (Chapter 4). We are aiming to assess whether or not the macaque has learned to process the reward as the experiment progresses. Moreover, if the macaque has indeed learned, we would like to identify the specific trial(s) over which the evidence of learning is manifest.

Location-specific test

For our location-specific test, we compute unsmoothed spectral estimates from the raw periodograms using Haar wavelets instead of DLA6. Similar to the location-specific tests for the hippocampus, for each location there are a total of 40 hypothesis tests and we tackle the multiple-hypothesis testing by using the FDR procedure. The test identified 344 locations as breaking the null hypothesis of spectral constancy over the replicates and we plot these locations in plot (a) of Figure 67. Once again we learn that several locations were identified by the test and whilst this is important, further information is gained by considering the percentage of the 40 hypothesis tests per location visualised in plot(b). Evidently, there are far fewer locations with a higher number of rejections than 1 in 40. It appears that for the NAc, whilst the test gives evidence of a replicate-effect for some locations, the potential evolution of the spectra across the replicates is not as substantial as was identified for the Hc. We observe that locations around timepoint 512, the start of the second experimental time block, are also identified by the test for the NAc, which is interesting as this observation is not as clear in the spectral estimates of Figure 41, top row. However, when inspecting the top right plot of Figure 45 (Chapter 4), a small spike in activity can be observed towards the end of the rescaled replicates at around timepoint 512 (0.25 in rescaled time). This demonstrates how our location-specific test can be utilised as an essential tool to screen the locations that are 'successfully' (decided through interpretation of the percentage of hypothesis tests rejected) rejected by the test as displaying a potential replicate-effect. We then carry out further analysis on the spectral characteristics at these locations, instead of simply plotting and comparing the within-trial evolutionary wavelet spectra. To gain further clarity on which rejected locations give the strongest evidence for the existence of spectral evolution over the replicates, we eliminate weakly rejected locations and display the results through plots (a)-(f) in Figure 67. Quite clearly, the impact of thresholding the rejected locations drastically narrows down the locations of most interest. In fact plot (f) indicates that there was no local window of (5) neighbouring locations with 25% or more significant Haar wavelet coefficients (rejected hypothesis tests).

Plots in Figure 68 give an indication of the replicates for which an evolution in the wavelet spectra was detected for specific time locations 565 (left) and 1865 (right), indicative of the visual exposure and task times of the experiment, respectively. A description for the plots' construction was previously given in the discussion for the hippocampus. Both locations 565 and 1865 have under 25% of hypothesis tests rejected under the FDR control.

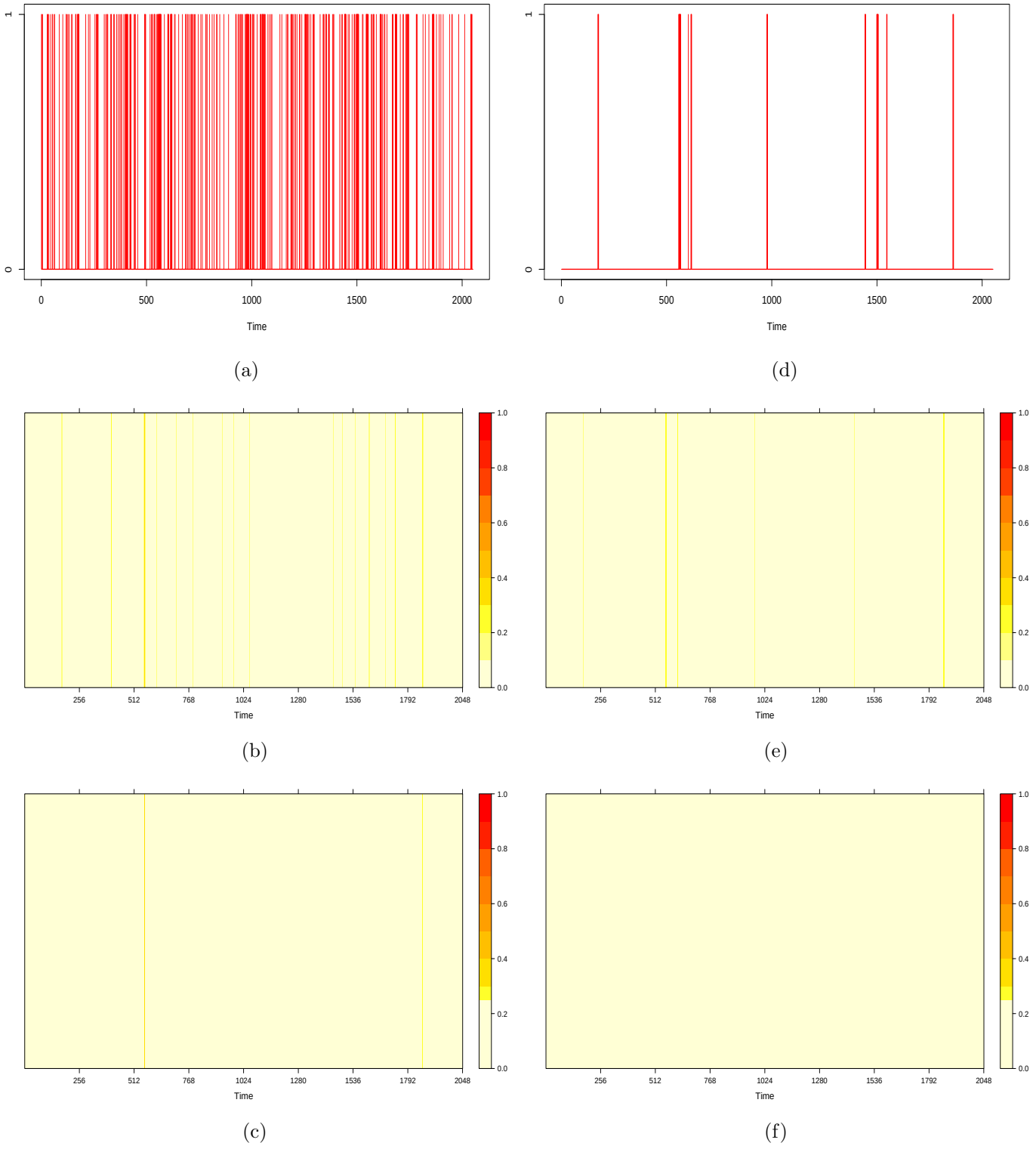


Figure 67: Location-specific rejection plots for the replicate-effect test carried out on the correct trials of the nucleus accumbens (NAc) dataset. *(a)*: binary plot where a vertical line to 1 indicates that the test identified the location as rejecting the null hypothesis of constancy; *(b)*: percentage of hypothesis tests rejected for each location; *(c)*: percentage of hypothesis tests rejected for each location thresholded at 25%; *(d)*: as in (a) but only for rejected locations with two rejected neighbour locations each side; *(e)*: as in (b) but only for rejected locations with two rejected neighbour locations each side; *(f)*: as in (c) but only for rejected locations with two rejected neighbour locations each side. The scale for plots (b), (c), (e) and (f), indicates the percentage of rejections out of 40 hypothesis tests per location.

Of these, most rejections occur over the final 128 replicates, indicating response to reward expectation towards the end of the experiment. Few rejections occur for a narrow window of 16 replicates ($256/2^4$, where 4 is the finest Haar scale tested) at the end of the replicates for location 565 and around replicate 220 for location 1865.

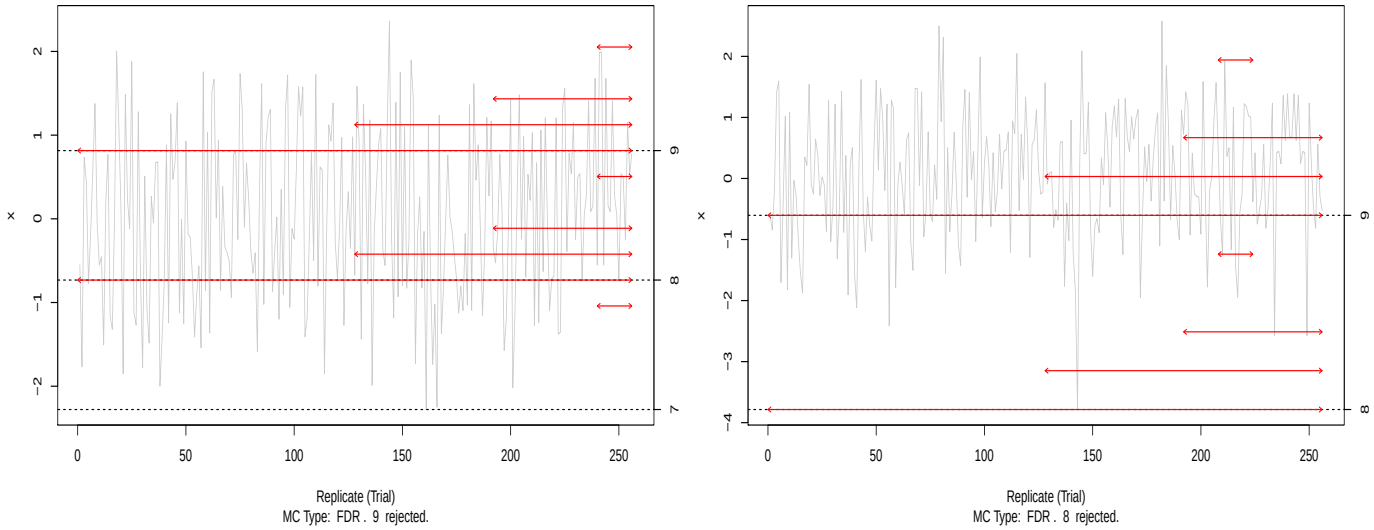


Figure 68: Replicate domain plots for correct nucleus accumbens trials at times 565 (*left*) and 1865 (*right*). Trials where replicate-effect detections were made are indicated by the double headed arrows, and the corresponding wavelet scales $j = 7, 8, 9$ are indicated on the right axis. Within each wavelet scale, the vertical position of the arrow indicates in ascending order the Haar scales $i = 0, ..., 4$.

Global test

For the global test, we obtain our spectral estimates with smoothing and correction applied. We carry out the global test on the NAc dataset and following algorithm 1 we obtain a p-value of 1, hence not enough evidence is present to deem a significant replicate-effect exists along the macaque's NAc response across the experiment. We plot the histogram of the bootstrap test statistics in Figure 69, highlighting these are larger than the observed test statistics. This indicates a greater variability in the spectral estimates under the null hypothesis of no replicate-effect than in the original dataset.

To attempt to understand why the test does not identify a global replicate-effect, in Figure 70 we examine the spectral estimates of the correct NAc trials, the averaged (over replicates) NAc spectral estimates $\bar{\hat{S}}_j(\frac{k}{T})$, and the spectral estimates computed for one bootstrapped process. The test statistics that we compute give a measure for the difference between the averaged spectral estimates (2nd row) and; (for $\mathbf{T}_{ave}^{obs}$) the spectral estimates (top row); (for $\mathbf{T}_{ave}^{(b)}$) the bootstrap spectral estimates (3rd row). Visually it is clear how $\mathbf{T}_{ave}^{(b)}$ computed for the estimated spectra of one bootstrap process (displayed in Figure 70) is larger than $\mathbf{T}_{ave}^{obs}$. This observation thus supports the outcome of not enough evidence being present to suggest a global replicate-effect exists for the correct trials of the NAc. Note, the averaged bootstrap spectral estimates (simulated under the null hypothesis) are similar both to their 'truth' and original NAc estimates (see Figure 70, bottom row versus 2nd and top rows), thus leading us towards two points to discuss.

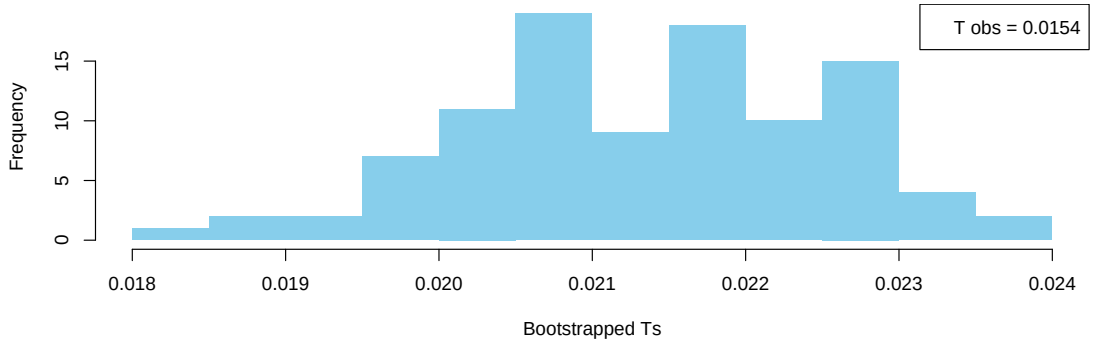


Figure 69: Histogram of the bootstrap test statistics for the global replicate-effect test carried out on the correct trials of the nucleus accumbens dataset.

Firstly, recall that the simulation study for both location-specific and global tests highlighted a weaker test performance when the spectral evolution across the replicates is very slow (e.g. simulation 1B in Section 5.2.1 and simulation p1 (b1) in Section 5.3.1). Similarly, the Haar wavelet based test of stationarity of Nason (2012, 2013) struggled to identify nonstationarity for their models P2 and P3 which both contained wavelet spectra defined to evolve very slowly over time. This is likely the situation here, as displayed in Figure 70 (top row). Secondly, recall that both the localised and global tests, are developed under the assumption of uncorrelated replicates. However, the NAc analysis in Chapter 4 (Section 4.3) uncovered the existence of a potential coherence along the replicate dimension, thus this dataset is likely evading this underpinning assumption.

5.4.3 Concluding remarks on the real data analysis

We have applied both the location-specific and global tests on the correct trials of the macaque hippocampus (Hc) and nucleus accumbens (NAc) datasets. Both test are capable of identifying the existence of a replicate-effect across the spectral estimates for the Hc, with the addition of the location-specific test identifying both the replicates (trials) and time locations within a replicate where activity was captured by the REv-LSW model. For the NAc, the location-specific test identified locations that rejected the null hypothesis of spectral constancy over the replicate domain, but the global test did not find evidence for a significant effect to be present. Our further investigation into this highlights the sensitivity of the test to the slowness of the spectral evolution over the replicates. Additionally, in Chapter 4 (Section 4.3) we allowed for the possibility of correlation between trials in the models for the real data, and there we estimated the coherence structure. We recall that there was little evidence found to suggest correlation between trials existed within the Hc data however we did identify correlation for the NAc data. We note that both our tests were constructed under the uncorrelated trials assumption, which appears reasonable for the Hc data based on our analysis of the coherence, but less so for the NAc data which was found to show evidence of a replicate coherence. For the global-test, inclusion of an estimated coherence structure between trials could improve the bootstrap simulation and estimation, therefore an important next step in the methodology would be to develop novel tests that incorporate between-trial correlation.

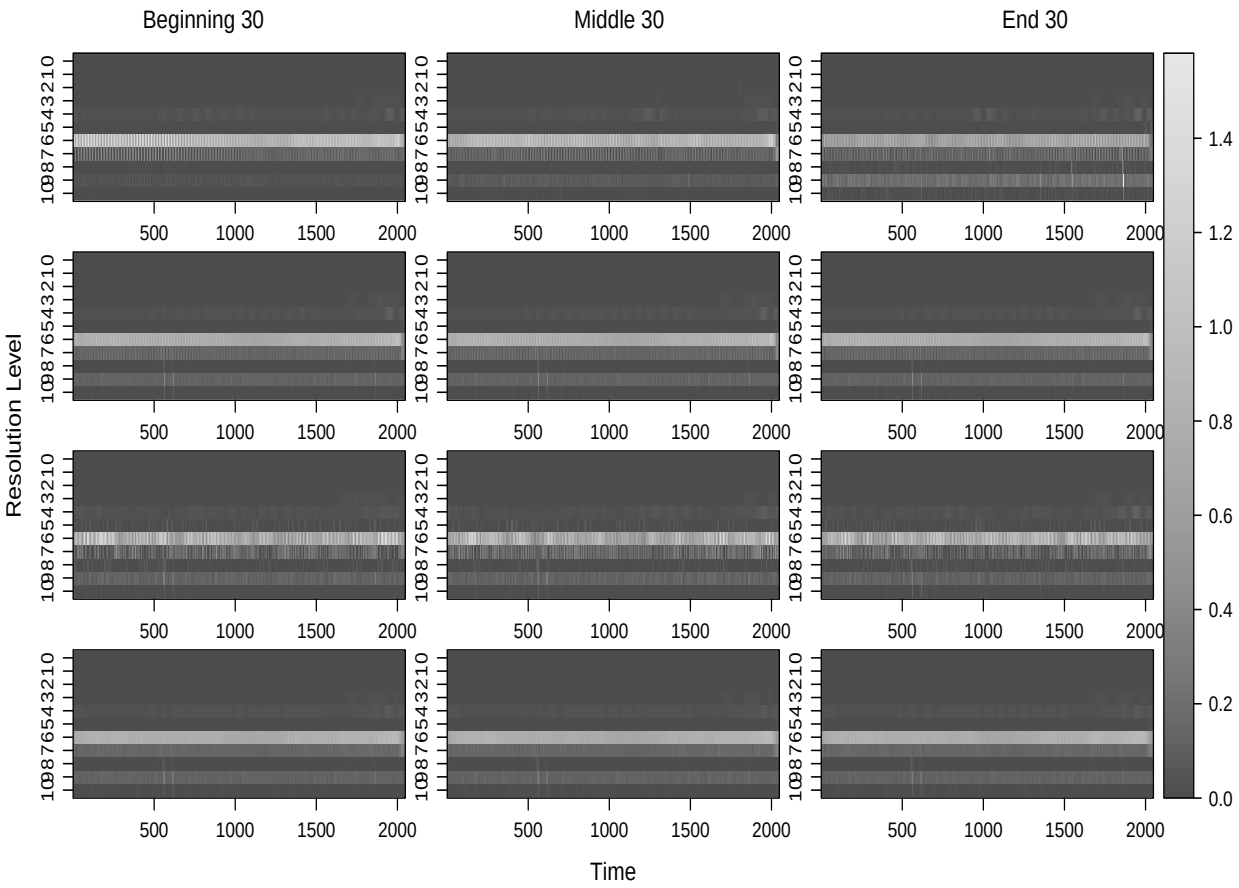


Figure 70: Time-scale nucleus accumbens (NAc) plots for the correct trials computed for the global test of replicate-effect. Spectral estimates are shown for the average over 30 replicates in the beginning, middle and end of the experiment. *Top*: spectral estimates of the correct NAc trials; *2nd*: averaged spectral estimates across the experiment, under the assumption of spectral constancy over the replicates; *3rd*: spectral estimates corresponding to one bootstrap process; *bottom*: spectral estimates averaged over 100 bootstrap realisations.

5.5 Concluding remarks

In this chapter we have proposed two wavelet-based tests for the existence of a replicate-effect along a replicate time series consisting of individual time series (replicates) that feature nonstationarities across both time and replicate-dimensions. The replicate-effect is to be understood as a measure of the departure from constancy of the process spectral characteristics across the replicates, framed here by means of a REv-LSW meta-process representation. Through simulation studies and an application to real data from the neuroscience, both tests have been shown to perform successfully, with weaker performance identified for processes with very slowly evolving amplitudes.

So, which test should we use? From the results, a natural order presents itself, with the global test offering a way to determine whether over all wavelet scales and times, a significant replicate-effect exists. Hence our recommendation is that the global test could be used as a first step in an analysis of the characteristics of a replicate time series modelled as a REv-LSW process. Its purpose could simply be a clarification tool. The location-specific test would then provide a more in-depth assessment through carrying out multiple hypothesis tests of replicate-effect for each location. Informed by the application context, the user can determine which locations are significantly rejected by the test

through assessing the percentage of the multiple hypothesis test per location that reject the null hypothesis of spectral constancy over the replicates.

Crucial to the proposed tests of replicate-effect is the assumption of uncorrelated replicates, hence naturally a next step would be to develop the tests for replicate-effect that incorporate the potential for dependence between replicates. A different direction would be to gain a further understanding into how many significant Haar wavelet coefficients (multiple hypothesis test rejections) would be sufficient to deem a location as significantly rejected by the location-specific test.

6 Conclusions and further work

This thesis has proposed a novel wavelet-based model that provides a stochastic representation to a meta-process of ordered replicates and simultaneously captures evolutionary dynamics of a nonstationary process within individual replicates and across all replicates. Additionally, we have proposed two tests for deciding whether a replicate-effect is present over the meta-process. The methodology developed throughout this thesis has been motivated by the specific neuroscience application that records the brain process dynamics of a macaque monkey during an associative learning experiment. The results obtained for the analysis of the macaque data have demonstrated the major advantages of the model proposed in this thesis, such that (i) the model captures evolutionary brain process characteristics localised across replicates and locally in time within each replicate, and (ii) it takes into account the correlation of brain signals across all trials of the experiment. We now give a summary of the main contributions developed in Chapters 2, 3 and 5, alongside a discussion of their application to the macaque learning experiment in Chapters 4 and 5. We also highlight potential avenues for future work, where some suggestions are a natural next step theoretically and others have been brought to light through the real data analysis.

REv-LSW model under the assumption of uncorrelated replicates

In Chapter 2 we proposed the *replicate-evolving locally stationary wavelet* (REv-LSW) model that captures the evolutionary process characteristics across time within an individual replicate and across all replicates. The methodology in this chapter is developed under the assumption that replicates are uncorrelated, an assumption also made in Fiecas and Ombao (2016). The REv-LSW model provides a wavelet-based solution to overcome the misleading results obtained from approaching replicate time series as identical process realisations. Underpinned by the LSW framework of Nason *et al.* (2000) and adopting the concept of rescaled time (Dahlhaus, 1997) in the replicate domain, we developed novel evolutionary wavelet quantities and associated estimation theory. Additionally, to obtain well-behaved, consistent spectral estimates we employed local smoothing solely across replicates and further considered smoothing across both replicates and time (within each replicate).

Through simulation studies we demonstrated the advantage of the proposed REv-LSW model against the classical approach of estimating each replicate spectrum independently and then studying the average spectrum over all replicates. Visually, the REv-LSW model successfully captured evolution in the spectra across the replicates, identifying new features that were unattainable through the classical approach. Additionally, we reported quantifiable measures, such as the mean squared errors (MSE), to assess the performance of our model. The MSEs proved the better performance of the REv-LSW model and furthermore gave evidence of improved spectral estimates via smoothing over both replicates and time. The MSEs also validated our asymptotic estimation theory, showing improved results as $T, R \to \infty$.

REv-LSW model embedding cross-replicate dependence

A serious limitation of the methodology developed in Chapter 2 and also in the Fourier-based approach of Fiecas and Ombao (2016), is the assumption of uncorrelated replicates.

Thus, in Chapter 3 we developed the theory to allow for cross-replicate dependence by means of a *between-replicate* coherence structure. Naturally, the proposed general model incorporates the scenario of uncorrelated replicates as a particular case. The associated estimation theory encompasses between-replicate variation through estimation of the *between-replicate evolutionary wavelet spectrum*, which in turn allows us to estimate the replicate-coherence structure.

The simulation studies proved that the between-replicate coherence can be well estimated and in doing so, provides novel, useful information about the replicate evolution. Thus, allowing for replicate-coherence in the REv-LSW model further strengthens the argument against approaching replicate time series as identical realisations (e.g. classical LSW). The MSEs obtained for the between-replicate coherence estimates again demonstrated the improved performance via replicate- and time-smoothing with further decreases in the MSE when the replicate-smoothing window $(2M + 1)$ is increased. Additionally, the MSEs highlighted that the correction procedure, necessary to ensure spectral estimates are positive, has the undesired effect of increasing the MSEs and introducing bias as $T, R \to \infty$. However, Sanderson *et al.* (2010) reported similar effects for their bivariate coherence estimation framework and so it is unsurprising to see these effects echoed in our results.

In Chapter 4 we applied the wavelet-based REv-LSW methodology developed in Chapters 2 and 3 to the hippocampus (Hc) and nucleus accumbens (NAc) brain process data recorded during the macaque associative learning experiment. The results provided substantial evidence for the existence of evolutionary spectra in both the Hc and the NAc. Crucially, the proposed REv-LSW model has been developed to account for between-trial dependence and we were able to identify a moderate between-trial dependence across neighbouring trials for the NAc. Furthermore, validation of our results was given through bootstrapped confidence intervals for the wavelet spectrum and between-trial coherence. In relation to the experiment, our analysis has extracted localised information in time within- and between-trials, suggesting that the macaque is learning over the trials of the experiment. Additionally, the dampening of NAc activity over the experiment suggests that the macaque has learned to process reward as the experiment progresses. The moderate between-trial dependence captured within the NAc suggests that expectations (e.g. of a juice reward) can influence the actions of the macaque across neighbouring trials. Note, such inference would not have been attainable through performing the analysis on the macaque data using the classical LSW approach. The application to a neuroscience problem is just one example of the utility of the methodology developed in this thesis. We foresee the methodology to be advantageous within other experimental areas where wavelet spectral analysis has proved to be ideally suited, e.g. circadian biology (Hargreaves *et al.*, 2018, 2019), .

A next step for further research would be to investigate the REv-LSW local (partial) autocorrelation function as introduced for the LSW setup by Killick *et al.* (2020), which could present additional information on the underlying structure of the replicate time series.

A natural extension of the REv-LSW model would be to develop the methodology for multivariate replicate time series and the review of Fiecas and Ombao (2016) in Chapter 1 gives a flavour of how this could be achieved using Fourier methods. Wavelet-based multivariate methods have been developed by Sanderson *et al.* (2010) and Park *et al.*

(2014). A multivariate approach would allow us to additionally define and investigate the variate replicate-coherence, providing an analysis between channels as seen for the bivariate case of the hippocampus and nucleus accumbens data in Fiecas and Ombao (2016) and Gorrostieta *et al.* (2011).

During the simulation studies and macaque data analysis, we commented on the choice of M that determines the length of the replicate-smoothing window. We recommended a window length of $(2M+1)$ with $M = \frac{3}{4}\sqrt{R}$ based on our findings. Future work could see the development of an automatic replicate-smoothing bandwidth selector, analogous to the `AutoBestBW` function in the `locits` package (Nason, 2013) developed for use in R.

Another avenue for future research stems from the discussion on practical implementation in Remark 3.2.1 (Chapter 3). We noted that one possibility to ensure that spectral estimates are positive, would be to truncate the correction matrix A^{-1} at zero. However, as seen in the simulation studies and macaque data analysis, this approach introduced bias. Sanderson (2010) and Sanderson *et al.* (2010) investigated other approaches and reported better results when additionally employing smoothing over scales. We conjecture that this is also applicable for our work as a further step, to be done after smoothing through time, and assessment of the improved estimation this approach could bring to our methodology makes for interesting future research.

Ordered replicates and thus smoothness assumptions underpin the local averaging approach proposed in this thesis. Further work could investigate estimation when the smoothness assumptions falter. However, in the case that the smoothness between replicates breaks and there is no need for an ordering, clustering methods (Hargreaves *et al.*, 2018; Ting *et al.*, 2018) could provide a preferred and appropriate analysis.

A further avenue for research would be to investigate a flexible REv-LSW model hinged upon a replicate-adaptive wavelet construction, which would in turn further tune the spectral and coherence estimates.

Tests of replicate-effect

Chapter 5 introduces two novel hypothesis tests to assess whether a significant replicate-effect is manifest across replicates of the meta-process, with both tests underpinned by the methodology developed in Chapter 2. The *location-specific* test provides a detailed investigation for the presence of a replicate-effect, identifying the replicates, r, for which evolutionary spectra are detected and furthermore identifying the specific times, k, for which a replicate-effect is manifest. The *global* test serves the purpose to give an overall diagnostic as to whether a replicate-effect exists as opposed to assessing the individual time locations.

For well-behaved processes, the simulation studies for both tests demonstrated their applicability and their favourable performance was evidenced through relevant quantifiable measures. For both tests we identified a test sensitivity when the process exhibits very slowly varying amplitudes across replicates, however Nason (2012, 2013) reported similar findings for the wavelet-based test of stationarity developed within.

Application of both tests to the correct trials of the macaque hippocampus (Hc) and nucleus accumbens (NAc) datasets corroborated the results on the real data obtained in Chapter 4. The macaque data analysis in Chapter 5 demonstrated the utility of the tests

as tools to first assess a replicate time series for the presence of a replicate-effect. For the Hc, the global test gave a strong indication (p-value < 0.01) that a replicate-effect exists. The location-specific test gave a further informative assessment, identifying the trials for which the replicate-effect was detected and additionally at which times within a trial the replicate-effect is manifest. Interestingly, the time points identified by the test coincided with the experimental time blocks corresponding to picture exposure and choice making. For the NAc, the location-specific test detected the presence of a replicate-effect for some locations but the global test did not find evidence for the existence of a significant effect. Recalling that the tests were developed under the assumption of uncorrelated replicates, we made note that this assumption appears reasonable for the Hc based on finding little evidence for correlation between trials of the Hc data in Chapter 4. However, for the NAc evidence of a replicate coherence was found.

Both proposed tests of replicate-effect have been developed under the crucial assumption that replicates are uncorrelated. Naturally, a next step for future research would be to develop the tests of replicate-effect that incorporate the potential for dependence between replicates, underpinned by the methodology developed in Chapter 3. It would then be interesting to assess how well the tests that incorporate replicate-dependence are able to find evidence for the existence of a replicate-effect in the NAc.

The macaque data analysis also highlighted that future work could be to investigate the number of significant Haar wavelet coefficients (multiple hypothesis test rejections) that would be sufficient to deem a location as significantly rejected by the location-specific test.

A Appendix: Proofs

A.1 Proofs of results on the asymptotic behaviour of proposed estimators

In this chapter, we give details of the proofs in Chapters 2 and 3 using the notation described therein. In order to ease repetition between proofs, we have chosen to present the proofs that would allow for the coherence/non-coherence counterpart to easily be obtained following similar steps.

For ease we recall here that the auto- and cross-correlation wavelets are defined for $\tau \in \mathbb{Z}$ as $\Psi_j(\tau) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{j,k}(\tau)$ and $\Psi_{j,l}(\tau) = \sum_{k\in\mathbb{Z}} \psi_{j,k}(0)\psi_{l,k}(\tau)$ respectively, where in general $\psi_{j,k}(\tau) = \psi_{j,k-\tau}$ are compactly supported discrete wavelets as defined in Nason *et al.* (2000). Note that from their construction, $\Psi_j(-\tau) = \Psi_j(\tau)$ and $\Psi_{j,l}(-\tau) = \Psi_{l,j}(\tau)$, and both have compact support with length of order 2^j, $2^j + 2^l$ respectively (Nason *et al.*, 2000; Sanderson *et al.*, 2010).

In what follows, wherever the summation domain is not specified, it is to be understood as $\mathbb{Z}$ for time indices (e.g. k, n, τ) and as $\mathbb{Z}_+^\star$ (strictly positive integers) for scale indices (e.g. j, j', l).

We also recall the autocorrelation wavelet inner product matrices, defined as

$$\begin{aligned} A_{j,l} &= \sum_{\tau} \Psi_j(\tau)\Psi_l(\tau) = \sum_{\tau} |\Psi_{j,l}(\tau)|^2, \\ A_{j,l}^{\tau} &= \sum_{n} \Psi_j(n)\Psi_l(n+\tau) = \sum_{n} \Psi_{j,l}(n)\Psi_{j,l}(n+\tau), \end{aligned}$$

with properties $A_{j,l} \geq 0$, $A_{j,l} = A_{l,j}$, $\sum_j 2^{-j} A_{j,l} = 1$ (Fryzlewicz *et al.*, 2003).

In the proofs that follow we make use of the results in the following lemmas, whose proofs appear in Appendix A.2. Note, these results are obtained in the presence of cross-replicate dependence but also hold for their non-coherence counterpart as described in the proofs.

Lemma A.1.1. *Under the assumptions of Definition 3.1.1, we have at each scale j and for every rescaled time and replicates, z and ν, ν' respectively, $\sum_{l=1}^{\infty} A_{j,l} S_l(z, \nu, \nu') = \mathcal{O}(2^j)$.*

Lemma A.1.2. *Under the assumptions of Definition 3.1.1, we have at each scale j and for every rescaled time and replicates, z and ν, ν' respectively, and time-lag τ, $\sum_{l=1}^{\infty} A_{j,l}^{\tau} S_l(z, \nu, \nu') = \mathcal{O}(2^j)$.*

Lemma A.1.3. *Under the assumptions of Definition 3.1.1, we have at scales j, j' and for every rescaled time and replicates, z and ν, ν' respectively, and time-lag τ,*

$$\sum_{l=1}^{\infty} \left| \sum_{n\in\mathbb{Z}} \Psi_{j,j'}(n)\Psi_l(n+\tau) S_l\left(z, \nu, \nu'\right) \right| = \mathcal{O}(2^{(j+j')/2}).$$

Additionally, for the results embedding replicate dependence, we make use of the following lemmas.

Lemma A.1.4. *Under the assumptions of Definition 3.1.1, we have a sequence $\{B_j\}$ of uniformly bounded Lipschitz constants in j with $\sum_j 2^j B_j < \infty$ such that for any replicates*

r, r' *and times* k, n

$$\left|\widetilde{W}_j\left(\frac{k+n}{T},\frac{r}{R}\right)\widetilde{W}_j\left(\frac{k+n}{T},\frac{r'}{R}\right)\rho_j\left(\frac{k+n}{T},\frac{r}{R},\frac{r'}{R}\right)\right.$$
$$\left.-\widetilde{W}_j\left(\frac{k}{T},\frac{r}{R}\right)\widetilde{W}_j\left(\frac{k}{T},\frac{r'}{R}\right)\rho_j\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)\right| \leq |n|B_jT^{-1}. \tag{70}$$

Note that the above result means that the replicate spectrum $S_j(\cdot\,,\nu,\nu')$ is Lipschitz continuous in the rescaled time argument within rescaled replicates ν, ν'.

Lemma A.1.5. *Under the assumptions of Definition 3.1.1, we have a sequence* $\{B'_j\}$ *of uniformly bounded Lipschitz constants in* j *with* $\sum_j 2^j B'_j < \infty$ *such that for any replicates* r, r' *and times* k, n

$$\left|\widetilde{W}_j\left(\frac{k}{T},\frac{r+s}{R}\right)\widetilde{W}_j\left(\frac{k}{T},\frac{r'+s}{R}\right)\rho_j\left(\frac{k}{T},\frac{r+s}{R},\frac{r'+s}{R}\right)\right.$$
$$\left.-\widetilde{W}_j\left(\frac{k}{T},\frac{r}{R}\right)\widetilde{W}_j\left(\frac{k}{T},\frac{r'}{R}\right)\rho_j\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)\right| \leq |s|B'_jR^{-1}. \tag{71}$$

Note that the above result effectively states that the replicate spectrum $S_j(z,\nu+\cdot\,,\nu'+\cdot\,)$ is Lipschitz continuous in the rescaled replicate arguments.

A.1.1 Proofs of Propositions 2.1.10 and 3.1.7

Proof of Proposition 3.1.7 (Expectation) ***in the presence of cross-replicate dependence***

Using Definition 3.1.1 of the REv-LSW process and its replicate raw periodogram (equation (53)), we obtain

$$\begin{aligned}
\mathbf{E}[I^{(r,r');R}_{j,k;T}] &= \mathbf{E}[d^{r;R}_{j,k;T}d^{r';R}_{j,k;T}] \\
&= \mathbf{E}\left[\left\{\sum_t X^{r;R}_{t;T}\psi_{j,k}(t)\right\}\left\{\sum_t X^{r';R}_{t';T}\psi_{j,k}(t')\right\}\right] \\
&= \mathbf{E}\left[\left\{\sum_t\sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\psi_{l,m}(t)\xi^r_{l,m}\psi_{j,k}(t)\right\}\right. \\
&\qquad\left.\times\left\{\sum_{t'}\sum_{l'=1}^{\infty}\sum_{m'\in\mathbb{Z}}\omega^{r';R}_{l',m';T}\psi_{l',m'}(t')\xi^{r'}_{l',m'}\psi_{j,k}(t')\right\}\right] \\
&= \sum_t\sum_{t'}\sum_{l=1}^{\infty}\sum_{l'=1}^{\infty}\sum_{m\in\mathbb{Z}}\sum_{m'\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\omega^{r';R}_{l',m';T}\psi_{l,m}(t)\psi_{l',m'}(t')\psi_{j,k}(t)\psi_{j,k}(t')\mathbf{E}\left[\xi^r_{l,m}\xi^{r'}_{l',m'}\right] \\
&= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\left\{\sum_t\psi_{l,m}(t)\psi_{j,k}(t)\right\}^2,
\end{aligned}$$

since $\mathbf{E}[\xi^{r}_{l,m}\xi^{r'}_{l',m'}] = \delta_{l,l'}\delta_{m,m'}\rho^{r,r';R}_{l,m;T}$ and so $l = l'$ and $m = m'$. Letting $n := m - k$ yields

$$\begin{aligned}\mathbf{E}[I^{(r,r');R}_{j,k;T}] &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}} \omega^{r;R}_{l,n+k;T}\omega^{r';R}_{l,n+k;T}\rho^{r,r';R}_{l,n+k;T}\left\{\sum_{t}\psi_{l,n+k-t}\psi_{j,k-t}\right\}^2, \\ &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left\{S_l\left(\frac{n+k}{T},\frac{r}{R},\frac{r'}{R}\right) + \mathcal{O}(C'_l T^{-1}) + \mathcal{O}(D'_l R^{-1})\right\}(\Psi_{j,l}(-n))^2,\end{aligned}$$

where above we also used the definition of the cross-correlation wavelets, as well as the result in Proposition 3.1.5 proof, recalled below

$$\left|\omega^{\lfloor \nu R\rfloor;R}_{j,\lfloor zT\rfloor;T}\omega^{\lfloor \nu' R\rfloor;R}_{j,\lfloor zT\rfloor;T}\rho^{\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor;R}_{j,\lfloor zT\rfloor;T} - S_j(z,\nu,\nu')\right| = \mathcal{O}(C'_j T^{-1}) + \mathcal{O}(D'_j R^{-1}), \tag{72}$$

where C'_j and D'_j can be shown to also fulfill equations of the type in Definition 3.1.1, 3c and 4b.

Then, by Lemma A.1.4 and using a substitution $n := -n$ it follows that

$$\begin{aligned}\mathbf{E}[I^{(r,r');R}_{j,k;T}] &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left(S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right) + \mathcal{O}(|n|B_l T^{-1}) + \mathcal{O}(C'_j T^{-1}) + \mathcal{O}(D'_j R^{-1})\right)(\Psi_{j,l}(n))^2, \\ &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left(S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)(\Psi_{j,l}(n))^2\right) \\ &+ \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left\{\mathcal{O}(|n|B_l T^{-1}) + \mathcal{O}(C'_l T^{-1}) + \mathcal{O}(D'_l R^{-1})\right\}(\Psi_{j,l}(n))^2, \text{ thus} \\ &= \sum_{l=1}^{\infty} A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right) + \mathcal{O}(T^{-1})\sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}} B_l|n|\,(\Psi_{j,l}(n))^2 \\ &+ \mathcal{O}(T^{-1})\sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}} C'_l\,(\Psi_{j,l}(n))^2 + \mathcal{O}(R^{-1})\sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}} D'_l\,(\Psi_{j,l}(n))^2.\end{aligned}$$

The order terms above can be established as follows. Since the number of terms in the wavelet cross-correlation is finite and bounded as a function of n (Nason *et al.*, 2000) and the length of the compact support of $\Psi^2_{j,l}$ is bounded by $K(2^j + 2^l)$ for some constant K (Sanderson *et al.*, 2010), we have

$$\begin{aligned}\sum_l\sum_n B_l|n|\mathcal{O}(T^{-1})\Psi^2_{j,l}(n) &\le \mathcal{O}(T^{-1})\sum_l(2^j + 2^l)B_l A_{j,l}, \\ &\le \mathcal{O}(2^j T^{-1})\sum_l B_l A_{j,l} + \mathcal{O}(T^{-1})\sum_l 2^l B_l A_{j,l}, \\ &= \mathcal{O}(2^j T^{-1}),\end{aligned} \tag{73}$$

where we used $\sum_l B_l A_{j,l} = \sum_l 2^l B_l 2^{-l} A_{j,l} \le \sum_{l'} 2^{l'} B_{l'} \sum_l 2^{-l} A_{j,l} < \infty$ since $\sum_l 2^l B_l < \infty$

and $\sum_l 2^{-l} A_{j,l} = 1$ (Fryzlewicz *et al.*, 2003), and

$$\begin{aligned} T^{-1} \sum_l 2^l B_l A_{j,l} &= 2^j T^{-1} \sum_l 2^l B_l 2^{-j} A_{j,l}, \\ &\leq 2^j T^{-1} \sum_l \left(2^l B_l \sum_{j'} 2^{-j'} A_{j',l} \right), \\ &= \mathcal{O}(2^j T^{-1}), \end{aligned}$$

again as $\sum_l 2^l B_l < \infty$ and $\sum_j 2^{-j} A_{j,l} = 1$.

Using the definition of the A matrix, $A_{j,l} = \sum_n \Psi^2_{j,l}(n)$, in the next order term, we obtain

$$\begin{aligned} \sum_l \sum_n D'_l \mathcal{O}(R^{-1}) \Psi^2_{j,l}(n) &= \mathcal{O}(R^{-1}) \sum_l 2^l D'_l 2^{-l} A_{j,l}, \\ &\leq \mathcal{O}(R^{-1}) \sum_{l'} 2^{l'} D'_{l'} \sum_l 2^{-l} A_{j,l}, \\ &= \mathcal{O}(R^{-1}) \end{aligned}$$

where we used $\sum_l 2^l D'_l < \infty$ and $\sum_l 2^{-l} A_{j,l} = 1$ (Fryzlewicz *et al.*, 2003). Using the same set of arguments and the condition $\sum_l 2^l C'_l < \infty$, we also have $\sum_l \sum_n C'_l \mathcal{O}(T^{-1}) \Psi^2_{j,l}(n) = \mathcal{O}(T^{-1})$.

Hence, retaining the maximum order terms, we have

$$\mathbf{E}[I^{(r,r');R}_{j,k;T}] = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R} \right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}),$$

which concludes the expectation part of the proof. □

Proof of Proposition 3.1.7 (Variance) ***in the presence of cross-replicate dependence***

For ease of notation let $X^{r;R}_{t;T} = X^r_t$ and $d^{r;R}_{j,k;T} = d^r_{j,k}$.

$$\begin{aligned} \mathbf{E}\left[I^{(r,r');R}_{j,k;T} I^{(r,r');R}_{j',k';T} \right] &= \mathbf{E}\left[d^r_{j,k} d^{r'}_{j,k} d^r_{j',k'} d^{r'}_{j',k'} \right] \\ &= \mathbf{E}\left[\sum_t X^r_t \psi_{j,k}(t) \sum_{t'} X^{r'}_{t'} \psi_{j,k}(t') \right. \\ &\qquad \left. \times \sum_h X^r_h \psi_{j',k'}(h) \sum_{h'} X^{r'}_{h'} \psi_{j',k'}(h') \right] \\ &= \sum_t \sum_{t'} \sum_h \sum_{h'} \psi_{j,k}(t) \psi_{j,k}(t') \psi_{j',k'}(h) \psi_{j',k'}(h') \mathbf{E}\left[X^{r;R}_{t;T} X^{r';R}_{t';T} X^{r;R}_{h;T} X^{r';R}_{h';T} \right]. \end{aligned} \tag{74}$$

Using the result of Isserlis (1918) for zero-mean Gaussian random variables, we have

$$\mathbf{E}\left[X^r_t X^{r'}_{t'} X^r_h X^{r'}_{h'} \right] = \mathbf{E}\left[X^r_t X^{r'}_{t'} \right] \mathbf{E}\left[X^r_h X^{r'}_{h'} \right] + \mathbf{E}\left[X^r_t X^r_h \right] \mathbf{E}\left[X^{r'}_{t'} X^{r'}_{h'} \right] + \mathbf{E}\left[X^r_t X^{r'}_{h'} \right] \mathbf{E}\left[X^{r'}_{t'} X^r_h \right], \tag{75}$$

and substituting (75) into (74) gives that $\mathbf{E}\left[I^{(r,r');R}_{j,k;T} I^{(r,r');R}_{j',k';T}\right] = \alpha + \beta + \gamma$.

Let us first consider

$$\begin{aligned}\mathbf{E}\left[X^r_t X^{r'}_{t'}\right] &= \mathbf{E}\left[\left\{\sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\psi_{l,m}(t)\xi^r_{l,m}\right\}\left\{\sum_{l'=1}^{\infty}\sum_{m'\in\mathbb{Z}} \omega^{r';R}_{l',m';T}\psi_{l',m'}(t')\xi^{r'}_{l',m'}\right\}\right] \\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\psi_{l,m}(t)\sum_{l'=1}^{\infty}\sum_{m'\in\mathbb{Z}} \omega^{r';R}_{l',m';T}\psi_{l',m'}(t')\mathbf{E}\left[\xi^r_{l,m}\xi^{r'}_{l',m'}\right] \\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\psi_{l,m}(t)\psi_{l,m}(t'), \qquad (76)\end{aligned}$$

since $\mathbf{E}\left[\xi^r_{l,m}\xi^{r'}_{l',m'}\right] = \delta_{l,l'}\delta_{m,m'}\rho^{r,r';R}_{l,m;T}$ and so (76) is obtained for $l = l'$ and $m = m'$.

Then the first term in equation (75) becomes

$$\begin{aligned}\mathbf{E}\left[X^r_t X^{r'}_{t'}\right]\mathbf{E}\left[X^r_h X^{r'}_{h'}\right] = \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} &\omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\psi_{l,m}(t)\psi_{l,m}(t') \\ &\times \sum_{e=1}^{\infty}\sum_{f\in\mathbb{Z}} \omega^{r;R}_{e,f;T}\omega^{r';R}_{e,f;T}\rho^{r,r';R}_{e,f;T}\psi_{e,f}(h)\psi_{e,f}(h').\end{aligned}$$

Thus the first term in equation (74), α, can be written as

$$\begin{aligned}\alpha &= \sum_t\sum_{t'}\sum_h\sum_{h'} \psi_{j,k}(t)\psi_{j,k}(t')\psi_{j',k'}(h)\psi_{j',k'}(h')\mathbf{E}\left[X^r_t X^{r'}_{t'}\right]\mathbf{E}\left[X^r_h X^{r'}_{h'}\right] \\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\sum_t \psi_{l,m}(t)\psi_{j,k}(t)\sum_{t'}\psi_{l,m}(t')\psi_{j,k}(t') \\ &\qquad \times \sum_{e=1}^{\infty}\sum_{f\in\mathbb{Z}} \omega^{r;R}_{e,f;T}\omega^{r';R}_{e,f;T}\rho^{r,r';R}_{e,f;T}\sum_h \psi_{e,f}(h)\psi_{j',k'}(h)\sum_{h'}\psi_{e,f}(h')\psi_{j',k'}(h') \\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\left\{\sum_t \psi_{l,m}(t)\psi_{j,k}(t)\right\}^2 \\ &\qquad \times \sum_{e=1}^{\infty}\sum_{f\in\mathbb{Z}} \omega^{r;R}_{e,f;T}\omega^{r';R}_{e,f;T}\rho^{r,r';R}_{e,f;T}\left\{\sum_h \psi_{e,f}(h)\psi_{j',k'}(h)\right\}^2.\end{aligned}$$

Using the result from the expectation part of the proof, namely

$$\sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}} \omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\left\{\sum_t \psi_{l,m}(t)\psi_{j,k}(t)\right\}^2 = \sum_{l=1}^{\infty} A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}),$$

we can re-write

$$\alpha = \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1})\right)$$
$$\times \left(\sum_{e=1}^{\infty} A_{j',e} S_e\left(\frac{k'}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(2^{j'} T^{-1}) + \mathcal{O}(R^{-1})\right),$$
$$= \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right)\right)\left(\sum_{e=1}^{\infty} A_{j',e} S_e\left(\frac{k'}{T}, \frac{r}{R}, \frac{r'}{R}\right)\right) + \mathcal{O}(2^{j+j'} T^{-1}) + \mathcal{O}(2^{\max\{j,j'\}} R^{-1}),$$

where the last equality follows from Lemma A.1.1.

In the same manner as for (76), we can show that

$$\mathbf{E}\left[X_t^r X_h^r\right] \mathbf{E}\left[X_{t'}^{r'} X_{h'}^{r'}\right] = \sum_{l=1}^{\infty} \sum_{m\in\mathbb{Z}} \omega_{l,m;T}^{r;R} \omega_{l,m;T}^{r;R} \rho_{l,m;T}^{r,r;R} \psi_{l,m}(t)\psi_{l,m}(h)$$
$$\times \sum_{l'=1}^{\infty} \sum_{m'\in\mathbb{Z}} \omega_{l',m';T}^{r';R} \omega_{l',m';T}^{r';R} \rho_{l',m';T}^{r',r';R} \psi_{l',m'}(t')\psi_{l',m'}(h'),$$
$$\mathbf{E}\left[X_t^r X_{h'}^{r'}\right] \mathbf{E}\left[X_{t'}^{r'} X_h^r\right] = \sum_{l=1}^{\infty} \sum_{m\in\mathbb{Z}} \omega_{l,m;T}^{r;R} \omega_{l,m;T}^{r';R} \rho_{l,m;T}^{r,r';R} \psi_{l,m}(t)\psi_{l,m}(h')$$
$$\times \sum_{l'=1}^{\infty} \sum_{m'\in\mathbb{Z}} \omega_{l',m';T}^{r;R} \omega_{l',m';T}^{r';R} \rho_{l',m';T}^{r,r';R} \psi_{l',m'}(t')\psi_{l',m'}(h).$$

Using the general property $\rho_{l,m;T}^{r,r;R} = 1$ for any l, m, r, the second term in (74), β, can be written

$$\beta = \sum_t \sum_{t'} \sum_h \sum_{h'} \psi_{j,k}(t)\psi_{j,k}(t')\psi_{j',k'}(h)\psi_{j',k'}(h') \mathbf{E}\left[X_t^r X_h^r\right] \mathbf{E}\left[X_{t'}^{r'} X_{h'}^{r'}\right]$$
$$= \sum_{l=1}^{\infty} \sum_{m\in\mathbb{Z}} \left(\omega_{l,m;T}^{r;R}\right)^2 \sum_t \psi_{l,m}(t)\psi_{j,k}(t) \sum_h \psi_{l,m}(h)\psi_{j',k'}(h)$$
$$\times \sum_{l'=1}^{\infty} \sum_{m'\in\mathbb{Z}} \left(\omega_{l',m';T}^{r';R}\right)^2 \sum_{t'} \psi_{l',m'}(t')\psi_{j,k}(t') \sum_{h'} \psi_{l',m'}(h')\psi_{j',k'}(h'). \quad (77)$$

Making the substitution $m = n + k$ and using the approximation (72), we can write first term in the product (77) as follows

$$\beta I = \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} (\omega_{l,n+k;T}^{r;R})^2 \sum_t \psi_{l,n+k-t}\psi_{j,k-t} \sum_h \psi_{l,n+k-h}\psi_{j',k'-h}$$
$$= \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} \left\{S_l\left(\frac{n+k}{T}, \frac{r}{R}\right) + \mathcal{O}(D_l' R^{-1}) + \mathcal{O}(C_l' T^{-1})\right\} \sum_t \psi_{l,n-t}\psi_{j,-t} \sum_h \psi_{l,n+k-h}\psi_{j',k'-h}$$
$$= \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} \left\{S_l\left(\frac{k}{T}, \frac{r}{R}\right) + \mathcal{O}(|n| L_l' T^{-1}) + \mathcal{O}(D_l' R^{-1}) + \mathcal{O}(C_l' T^{-1})\right\}$$
$$\times \sum_t \psi_{l,n-t}\psi_{j,-t} \sum_h \psi_{l,n+k-h}\psi_{j',k'-h},$$

where in the last equality we have used the Lipschitz continuity of $S_l(\cdot, r/R, r/R)$.

Making the substitution $u = h - t$, gives

$$\sum_{n\in\mathbb{Z}}\sum_{t}\psi_{l,n-t}\psi_{j,-t}\sum_{h}\psi_{l,n+k-h}\psi_{j',k'-h} = \sum_{u}\Psi_{j,j'}(u)\Psi_{l}(u-k+k'),$$

where $\Psi_{j,j'}(\tau) = \sum_k \psi_{j,k}\psi_{j',k-\tau}$ defines the cross-scale autocorrelation wavelets. The sums of wavelet products above were manipulated as follows,

$$\begin{aligned}
\sum_{n\in\mathbb{Z}}\sum_{t}\psi_{l,n-t}\psi_{j,-t}\sum_{h}\psi_{l,n+k-h}\psi_{j',k'-h} &= \sum_{n\in\mathbb{Z}}\sum_{t}\psi_{l,n-t}\psi_{j,-t}\sum_{u}\psi_{l,n+k-u-t}\psi_{j',k'-u-t} \\
&= \sum_{u}\sum_{t}\psi_{j,-t}\psi_{j',k'-u-t}\sum_{n\in\mathbb{Z}}\psi_{l,n-t}\psi_{l,n+k-u-t} \\
&= \sum_{u}\Psi_{j,j'}(u-k')\Psi_{l}(u-k), \\
&= \sum_{u}\Psi_{j,j'}(u)\Psi_{l}(u-k+k'),
\end{aligned}$$

where for the last equality we used a substitution $u := u - k'$. Equivalently, using the first result in the proof of Lemma A.1.3, the above could have been directly written as $\sum_u \Psi_{j,l}(u)\Psi_{j',l}(u-k+k')$.

Hence the first term in the product in equation (77) can be written as

$$\beta I = \sum_{l=1}^{\infty}\sum_{u}\Psi_{j,j'}(u)\Psi_{l}(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R}\right) + \mathcal{O}(2^{(j+j')/2}T^{-1}) + \mathcal{O}(R^{-1}), \text{ since}$$

$$\begin{aligned}
\beta I &= \sum_{l=1}^{\infty}S_l\left(\frac{k}{T},\frac{r}{R}\right)\sum_{n\in\mathbb{Z}}\sum_{t}\psi_{l,n-t}\psi_{j,-t}\sum_{h}\psi_{l,n-h}\psi_{j',-h} \\
&\quad + \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left\{\mathcal{O}(D_l'R^{-1}) + \mathcal{O}(C_l'T^{-1}) + \mathcal{O}(|n|L_l'T^{-1})\right\}\sum_{t}\psi_{l,n-t}\psi_{j,-t}\sum_{h}\psi_{l,n+k-h}\psi_{j',k'-h}, \\
&= \sum_{l=1}^{\infty}\sum_{u\in\mathbb{Z}}\Psi_{j,j'}(u)\Psi_{l}(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R}\right) + I + II + III.
\end{aligned}$$

The term I can be bounded as follows

$$\begin{aligned}
|I| &= \left|\sum_l \mathcal{O}(D_l R^{-1}) \sum_u \Psi_{j,l}(u)\Psi_{j',l}(u-k+k')\right|, \\
&\leq \mathcal{O}(R^{-1}) \sum_l \left(D_l \sum_u \left|\Psi_{j,l}(u)\Psi_{j',l}(u-k+k')\right| \right), \text{then using Cauchy Schwarz inequality} \\
&\leq \mathcal{O}(R^{-1}) \sum_l D'_l \left(\sum_u |\Psi_{j,l}(u)|^2\right)^{1/2} \left(\sum_u \left|\Psi_{j',l}(u-k+k')\right|^2\right)^{1/2}, \\
&= \mathcal{O}(R^{-1}) \sum_l D'_l (A_{j,l})^{1/2} (A_{j',l})^{1/2}, \text{from the definition of } A_{j,l}, \\
&= \mathcal{O}(R^{-1}) \sum_l (D'_l A_{j,l})^{1/2} (D'_l A_{j',l})^{1/2}, \\
&\leq \mathcal{O}(R^{-1}) \left(\sum_l D'_l A_{j,l}\right)^{1/2} \left(\sum_l D'_l A_{j',l}\right)^{1/2}, \text{ using again Cauchy Schwarz inequality} \\
&\leq \mathcal{O}(R^{-1}), \text{ since we saw in the expectation proof part that } \sum_l D'_l A_{j,l} = \mathcal{O}(1).
\end{aligned}$$

Similarly, $|II| = \left|\sum_l \mathcal{O}(C'_l T^{-1}) \sum_u \Psi_{j,l}(u)\Psi_{j',l}(u-k+k')\right| = \mathcal{O}(T^{-1})$.

We now bound the term III,

$$\begin{aligned}
|III| &\leq \sum_l \mathcal{O}(L_l T^{-1}) \left(2^l + 2^{\min\{j,j'\}}\right) \sum_u \left|\Psi_{j,l}(u)\Psi_{j',l}(u-k+k')\right|, \\
&\leq \mathcal{O}(T^{-1}) \sum_l L_l \left(2^l + 2^{\min\{j,j'\}}\right) (A_{j,l})^{1/2} (A_{j',l})^{1/2}, \text{ as above} \\
&= \mathcal{O}(T^{-1}) \sum_l L_l 2^l (A_{j,l})^{1/2} (A_{j',l})^{1/2} \\
&+ \mathcal{O}(T^{-1}) \sum_l L_l 2^{\min\{j,j'\}} (A_{j,l})^{1/2} (A_{j',l})^{1/2}, \\
&\leq \mathcal{O}(T^{-1}) 2^{(j+j')/2} \sum_l L_l 2^l (2^{-j/2} A_{j,l}^{1/2})(2^{-j'/2} A_{j',l}^{1/2}) \\
&+ \mathcal{O}(T^{-1}) 2^{(j+j')/2} \sum_l L_l (A_{j,l})^{1/2} (A_{j',l})^{1/2}, \text{ as } 2^{\min\{j,j'\}} \leq 2^{(j+j')/2}.
\end{aligned}$$

The term $\sum_l L_l 2^l (2^{-j/2} A_{j,l}^{1/2})(2^{-j'/2} A_{j',l}^{1/2}) \leq \left(\sum_l L_l 2^l 2^{-j} A_{j,l}\right)^{1/2} \left(\sum_l L_l 2^l 2^{-j'} A_{j',l}\right)^{1/2}$ from the Cauchy-Schwarz inequality, and using $\sum_j 2^{-j} A_{j,l} = 1$ and $\sum_l 2^l L_l < \infty$, we obtain its $\mathcal{O}(1)$ bound. The term $\sum_l L_l (A_{j,l})^{1/2} (A_{j',l})^{1/2} \leq \left(\sum_l L_l A_{j,l}\right)^{1/2} \left(\sum_l L_l A_{j',l}\right)^{1/2} = \mathcal{O}(1)$ since we already established that $\sum_l L_l A_{j,l} < \infty$ in the expectation part of the proof. Thus term III is bounded by $\mathcal{O}(2^{(j+j')/2} T^{-1})$.

The same arguments can be employed for the second product term in equation (77),

hence we can express β as

$$\begin{aligned}\beta &= \left(\sum_{l=1}^{\infty}\sum_{u}\Psi_{j,j'}(u)\Psi_l(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R}\right)+\mathcal{O}(2^{(j+j')/2}T^{-1})+\mathcal{O}(R^{-1})\right)\\ &\quad\times\left(\sum_{l'=1}^{\infty}\sum_{u'}\Psi_{j,j'}(u')\Psi_{l'}(u-k+k')S_{l'}\left(\frac{k'}{T},\frac{r'}{R}\right)+\mathcal{O}(2^{(j+j')/2}T^{-1})+\mathcal{O}(R^{-1})\right),\\ &= \left(\sum_{l=1}^{\infty}\sum_{u}\Psi_{j,j'}(u)\Psi_l(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R}\right)\right)\left(\sum_{l'=1}^{\infty}\sum_{u'}\Psi_{j,j'}(u')\Psi_{l'}(u-k+k')S_{l'}\left(\frac{k'}{T},\frac{r'}{R}\right)\right)\\ &\quad+\mathcal{O}(2^{(j+j')/2+(j+j')/2}T^{-1})+\mathcal{O}(2^{(j+j')/2}R^{-1}),\end{aligned}$$

where to establish the last equality we used Lemma A.1.3.

The third term in (74), γ, can be established using a similar argument as follows.

$$\begin{aligned}\gamma &= \sum_{t}\sum_{t'}\sum_{h}\sum_{h'}\psi_{j,k}(t)\psi_{j,k}(t')\psi_{j',k'}(h)\psi_{j',k'}(h')\mathbf{E}\left[X_t^r X_{h'}^{r'}\right]\mathbf{E}\left[X_{t'}^{r'}X_h^r\right]\\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega_{l,m;T}^{r;R}\omega_{l,m;T}^{r';R}\rho_{l,m;T}^{r,r';R}\sum_{t}\psi_{l,m}(t)\psi_{j,k}(t)\sum_{h'}\psi_{l,m}(h')\psi_{j',k'}(h')\\ &\qquad\times\sum_{l'=1}^{\infty}\sum_{m'\in\mathbb{Z}}\omega_{l',m';T}^{r;R}\omega_{l',m';T}^{r';R}\rho_{l',m';T}^{r,r';R}\sum_{t'}\psi_{l',m'}(t')\psi_{j,k}(t')\sum_{h}\psi_{l',m'}(h)\psi_{j',k'}(h)\\ &= \left(\sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega_{l,m;T}^{r;R}\omega_{l,m;T}^{r';R}\rho_{l,m;T}^{r,r';R}\sum_{t}\psi_{l,m}(t)\psi_{j,k}(t)\sum_{h}\psi_{l,m}(h)\psi_{j',k'}(h)\right)^2.\end{aligned}$$

Using the approximation in (72) at time $m=n+k$ and from the spectrum Lipschitz continuity around time k, the wavelet substitution then yields

$$\begin{aligned}\gamma &= \left(\sum_{l=1}^{\infty}\sum_{u}\Psi_{j,j'}(u)\Psi_l(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)+\mathcal{O}(2^{(j+j')/2}T^{-1})+\mathcal{O}(R^{-1})\right)^2,\ \text{hence}\\ &= \left(\sum_{l=1}^{\infty}\sum_{u}\Psi_{j,j'}(u)\Psi_l(u-k+k')S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)\right)^2+\mathcal{O}(2^{(j+j')/2+(j+j')/2}T^{-1})+\mathcal{O}(2^{(j+j')/2}R^{-1}),\end{aligned}$$

where the order terms are obtained using the same arguments as for the term βI above and Lemma A.1.3.

Putting these results together, we have that

$$\mathbf{E}\left[I_{j,k;T}^{(r,r');R}I_{j',k';T}^{(r,r');R}\right]=\alpha+\beta+\gamma,$$

and for $j=j'$ and $k=k'$, using the result in Lemma A.1.1 and keeping the highest order

terms, the variance is shown to be

$$\begin{aligned}
\operatorname{var}\left(I_{j,k;T}^{(r,r');R}\right) &= \mathbf{E}\left[I_{j,k;T}^{(r,r');R} I_{j,k;T}^{(r,r');R}\right] - \mathbf{E}\left[I_{j,k;T}^{(r,r');R}\right]\mathbf{E}\left[I_{j,k;T}^{(r,r');R}\right], \\
&= \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1})\right)^2 \\
&\quad + \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}\right)\right)\left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}\right)\right) + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}), \\
&\quad + \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right)\right)^2 + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}) \\
&\quad - \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right)\right)^2 + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}), \\
&= \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}\right)\right)\left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}\right)\right) \\
&\quad + \left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right)\right)^2 + \mathcal{O}(2^{2j} T^{-1}) + \mathcal{O}(2^j R^{-1}).
\end{aligned}$$

□

Proof of Proposition 2.1.10 (Expectation) ***under the uncorrelated replicates assumption***

The proof follows similar steps to the expectation proof of its coherence counterpart, using the properties given in Definition 2.1.1 of the REv-LSW process and the quantities defined under the assumption of uncorrelated replicates. Note that $I_{j,k;T}^{(r,r);R} = I_{j,k;T}^{r;R}$.

Proof of Proposition 2.1.10 (Variance) ***under the uncorrelated replicates assumption***

The proof follows similar steps to the variance proof of its coherence counterpart, using the properties given in Definition 2.1.1 of the REv-LSW process and the quantities defined under the assumption of uncorrelated replicates.

A.1.2 Proofs of Propositions 2.1.13 and 3.1.9

Proof of Proposition 2.1.13 (Expectation) ***under the uncorrelated replicates assumption***

From the definition of the replicate-smoothed periodogram in (43), we have

$$\mathbf{E}\left[\tilde{I}_{j,k;T}^{r;R}\right] = (2M+1)^{-1} \sum_{s=-M}^{M} \mathbf{E}\left[I_{j,k;T}^{r+s;R}\right]$$

and substituting the asymptotic result for the expectation $\mathbf{E}\left[I_{j,k;T}^{r+s;R}\right]$ (see equation (42)), we further obtain

$$\mathbf{E}\left[\tilde{I}_{j,k;T}^{r;R}\right] = (2M+1)^{-1}\sum_{s=-M}^{M}\left\{\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r+s}{R}\right)+\mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1})\right\}.$$

Using the Lipschitz continuity of the spectrum $S_j(k/T,\cdot)$ in replicate-dimension, we obtain

$$\begin{aligned}\mathbf{E}\left[\tilde{I}_{j,k;T}^{r;R}\right] &= (2M+1)^{-1}\sum_{s=-M}^{M}\left\{\sum_{l=1}^{\infty}A_{j,l}\left(S_l\left(\frac{k}{T},\frac{r}{R}\right)+\mathcal{O}(|s|N_lR^{-1})\right)+\mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1})\right\}\\ &= (2M+1)^{-1}\sum_{s=-M}^{M}\left\{\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R}\right)+\mathcal{O}(|s|R^{-1})\sum_{l}N_lA_{j,l}+\mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1})\right\}\\ &= \sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R}\right)+\mathcal{O}(MR^{-1})+\mathcal{O}(2^jT^{-1}),\end{aligned}$$

where we have used that $\sum_l N_lA_{j,l}<\infty$ as $\sum_l 2^lN_l<\infty$ and $\sum_l 2^{-l}A_{j,l}=1$.

□

Proof of Proposition 2.1.13 (Variance) ***under the uncorrelated replicates assumption***

The definition of the replicate-smoothed periodogram in (43) coupled with the variance result for the raw periodogram in Proposition 2.1.10, yields

$$\begin{aligned}\operatorname{var}\left(\tilde{I}_{j,k;T}^{r;R}\right) &= (2M+1)^{-2}\sum_{s=-M}^{M}\operatorname{var}\left(I_{j,k;T}^{r+s;R}\right)\\ &= (2M+1)^{-2}\sum_{s=-M}^{M}\left\{2\left(\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r+s}{R}\right)\right)^2+\mathcal{O}(2^{2j}T^{-1})+\mathcal{O}(2^jR^{-1})\right\},\ \text{thus}\end{aligned}$$

$$\begin{aligned}\operatorname{var}\left(\tilde{I}_{j,k;T}^{r;R}\right) &= (2M+1)^{-2}\sum_{s=-M}^{M}\left\{2\left(\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R}\right)+\mathcal{O}(|s|R^{-1})\sum_{l}N_lA_{j,l}\right)^2\right\}\\ &\quad+(2M+1)^{-2}\sum_{s=-M}^{M}\left\{\mathcal{O}(2^{2j}T^{-1})+\mathcal{O}(2^jR^{-1})\right\}\\ &= \mathcal{O}(M^{-1})\left(\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R}\right)\right)^2+(2M+1)^{-2}\sum_{s=-M}^{M}\left\{\mathcal{O}(2^j|s|R^{-1})+\mathcal{O}(|s|^2R^{-2})\right\}\\ &\quad+\mathcal{O}(2^{2j}(MT)^{-1})+\mathcal{O}(2^j(MR)^{-1}),\ \text{from Lemma A.1.1 and as}\ \sum_{l}N_lA_{j,l}<\infty.\end{aligned}$$

Retaining the largest order terms, it then follows that

$$\begin{aligned}\operatorname{var}\left(\tilde{I}_{j,k;T}^{r;R}\right) &= \mathcal{O}(2^{2j}M^{-1})+\mathcal{O}(2^jR^{-1})+\mathcal{O}(MR^{-2})+\mathcal{O}(2^{2j}(MT)^{-1})+\mathcal{O}(2^j(MR)^{-1}),\\ &= \mathcal{O}(2^{2j}M^{-1})+\mathcal{O}(2^jR^{-1})+\mathcal{O}(MR^{-2}).\end{aligned}$$

The expectation and variance results show that for fixed coarse enough scales j (to guard against asymptotic bias and non-vanishing variance), the proposed replicate-smoothed periodogram is an asymptotically consistent estimator for the spectral quantity β, as it is asymptotically unbiased and its variance converges to zero as $T \to \infty, R \to \infty$, $M \to \infty$ and $M/R \to 0$. □

Proof of Proposition 3.1.9 (Expectation) ***in the presence of cross-replicate dependence***

The proof follows similar steps to the expectation proof of its non-coherence counterpart, using the replicate-smoothed periodogram in (54) and the asymptotic result for the expectation in Proposition 3.1.7.

Proof of Proposition 3.1.9 (Variance) ***in the presence of cross-replicate dependence***

Using the replicate-smoothed periodogram definition under equation (54), we obtain

$$\begin{aligned}\operatorname{var}\left(\tilde{I}_{j,k;T}^{(r,r');R}\right) &= (2M+1)^{-2}\sum_{s=-M}^{M}\sum_{s'=-M}^{M}\operatorname{cov}\left(I_{j,k;T}^{(r+s,r'+s);R}, I_{j,k;T}^{(r+s',r'+s');R}\right)\\ &= (2M+1)^{-2}\sum_{s=-M}^{M}\sum_{\eta=-M-s}^{M-s}\operatorname{cov}\left(I_{j,k;T}^{(r+s,r'+s);R}, I_{j,k;T}^{(r+s+\eta,r'+s+\eta);R}\right) \qquad (78)\end{aligned}$$

where we have let $\eta = s' - s$.

Let us now take

$$\begin{aligned}\operatorname{cov}\left(I_{j,k;T}^{(r+s,r'+s);R}, I_{j,k;T}^{(r+s+\eta,r'+s+\eta);R}\right) &= \mathbf{E}\left[d_{j,k}^{r+s}d_{j,k}^{r'+s}d_{j,k}^{r+s+\eta}d_{j,k}^{r'+s+\eta}\right]\\ &\quad - \mathbf{E}\left[d_{j,k}^{r+s}d_{j,k}^{r'+s}\right]\mathbf{E}\left[d_{j,k}^{r+s+\eta}d_{j,k}^{r'+s+\eta}\right]\\ &= \mathbf{E}\left[d_{j,k}^{r+s}d_{j,k}^{r+s+\eta}\right]\mathbf{E}\left[d_{j,k}^{r'+s}d_{j,k}^{r'+s+\eta}\right]\\ &\quad + \mathbf{E}\left[d_{j,k}^{r+s}d_{j,k}^{r'+s+\eta}\right]\mathbf{E}\left[d_{j,k}^{r'+s}d_{j,k}^{r+s+\eta}\right], \qquad (79)\end{aligned}$$

where we have use Isserlis' theorem (Isserlis, 1918) in the last equality.

Using the expectation of the replicate wavelet periodogram in Proposition 3.1.7 and the result in Lemma A.1.5, we can rewrite the terms in equation (79) as follows.

$$\begin{aligned}\mathbf{E}\left[d_{j,k}^{r+s}d_{j,k}^{r+s+\eta}\right] &= \mathbf{E}\left[I_{j,k;T}^{(r+s,r+s+\eta);R}\right]\\ &= \sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r+s}{R},\frac{r+s+\eta}{R}\right)+\mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1}),\\ &= \sum_{l=1}^{\infty}\left\{A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r+\eta}{R}\right)+\mathcal{O}(|s|B_l'R^{-1})\right\}+\mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1}),\\ &= \sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r+\eta}{R}\right)+\mathcal{O}(2^jT^{-1})+\mathcal{O}(|s|R^{-1}),\end{aligned}$$

where we have also used the Lipschitz constants' property $\sum_l 2^l B'_l < \infty$ that yields $\sum_l B'_l A_{j,l} < \infty$.

Using the same steps, we also have

$$\begin{aligned}\mathbf{E}\left[d_{j,k}^{r'+s} d_{j,k}^{r'+s+\eta}\right] &= \mathbf{E}\left[I_{j,k;T}^{(r'+s,r'+s+\eta);R}\right] \\ &= \sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1}).\end{aligned}$$

Similarly, from the expectation of the wavelet periodogram in Proposition 3.1.7 and the spectrum Lipschitz continuity in rescaled replicate, it can be shown that

$$\begin{aligned}\mathbf{E}\left[d_{j,k}^{r+s} d_{j,k}^{r'+s+\eta}\right] &= \mathbf{E}\left[I_{j,k;T}^{(r+s,r'+s+\eta);R}\right] \\ &= \sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r+s}{R}, \frac{r'+s+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(R^{-1}), \\ &= \sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r+s}{R}, \frac{r+s+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|r-r'|R^{-1}),\end{aligned}$$

and using the same arguments as above, one obtains

$$\mathbf{E}\left[d_{j,k}^{r+s} d_{j,k}^{r'+s+\eta}\right] = \sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1}),$$

under the condition that the replicates r, r' are such that $|r - r'| < \infty$.

Similarly,

$$\mathbf{E}\left[d_{j,k}^{r'+s} d_{j,k}^{r+s+\eta}\right] = \sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1}).$$

Thus we can write the covariance in equation (79) as follows

$$\begin{aligned}&\operatorname{cov}\left(I_{j,k;T}^{(r+s,r'+s);R}, I_{j,k;T}^{(r+s+\eta,r'+s+\eta);R}\right) \\ &\quad= \mathbf{E}\left[d_{j,k}^{r+s} d_{j,k}^{r+s+\eta}\right]\mathbf{E}\left[d_{j,k}^{r'+s} d_{j,k}^{r'+s+\eta}\right] + \mathbf{E}\left[d_{j,k}^{r+s} d_{j,k}^{r'+s+\eta}\right]\mathbf{E}\left[d_{j,k}^{r'+s} d_{j,k}^{r+s+\eta}\right] \\ &\quad= 2\left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1})\right) \\ &\qquad\times\left(\sum_{l=1}^{\infty} A_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1})\right). \qquad (80)\end{aligned}$$

Using this expression in equation (78), the variance of the replicate-smoothed wavelet periodogram is

$$\operatorname{var}\left(\tilde{I}_{j,k;T}^{(r,r');R}\right) = \mathcal{O}(M^{-2}) \sum_{s=-M}^{M} \sum_{\eta=-M-s}^{M-s} [\text{term I} + \text{term II} + \text{term III} + \text{term IV} + \text{term V} + \text{term VI}], \qquad (81)$$

where the individual terms are

$$\text{`term I'} = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R} \right),$$

$$\text{`term II'} = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \mathcal{O}(2^j T^{-1}),$$

$$\text{`term III'} = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \mathcal{O}(|s| R^{-1}),$$

$$\text{`term IV'} = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R} \right) \mathcal{O}(2^j T^{-1}),$$

$$\text{`term V'} = \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R} \right) \mathcal{O}(|s| R^{-1}),$$

and

$$\text{`term VI'} = \mathcal{O}(|s|^2 R^{-2}) + \mathcal{O}(2^j |s| (TR)^{-1}) + \mathcal{O}(2^{2j} T^{-2}).$$

In order to assess the order of the individual terms in the variance sum, first note that under the 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 τ, we obtain for time k, replicate r and replicate-lag η that $\sum_{\eta \in \mathbb{Z}} \left| \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \right| = \mathcal{O}(2^j)$, since using the definition of the A matrix and of the local covariance we have

$$\begin{aligned}
\sum_{\eta \in \mathbb{Z}} \left| \sum_{l=1}^{\infty} A_{j,l} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \right| &= \sum_{\eta \in \mathbb{Z}} \left| \sum_{\tau \in \mathbb{Z}} \left(\sum_{l=1}^{\infty} S_l \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R} \right) \Psi_l(\tau) \right) \Psi_j(\tau) \right|, \\
&= \sum_{\eta} \left| \sum_{\tau} c \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}; \tau \right) \Psi_j(\tau) \right|, \\
&\leq \sum_{\tau} \left(|\Psi_j(\tau)| \sum_{\eta} \left| c \left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}; \tau \right) \right| \right), \\
&= \mathcal{O}(1) \sum_{\tau} |\Psi_j(\tau)| = \mathcal{O}(2^j),
\end{aligned}$$

where we used the triangle inequality and the autocorrelation wavelet property $\sum_{\tau} |\Psi_j(\tau)| = \mathcal{O}(2^j)$.

Using Lemma A.1.1 and the property above, we readily obtain that term I in the variance equation is of order $\mathcal{O}(2^{2j} M^{-1})$; terms II and IV, are both of order $\mathcal{O}(2^{2j} (MT)^{-1})$; and terms III and V are both of order $\mathcal{O}(2^j R^{-1})$.

Now considering the final order term VI, we have

$$\mathcal{O}(M^2 R^{-2}) + \mathcal{O}(2^j M (TR)^{-1}) + \mathcal{O}(2^{2j} T^{-2}).$$

Putting these results together in the variance equation (81), we obtain

$$\begin{aligned}\operatorname{var}\left(\tilde{I}_{j,k;T}^{(r,r');R}\right) &= \mathcal{O}(2^{2j}M^{-1}) + \mathcal{O}(2^{2j}(MT)^{-1}) + \mathcal{O}(2^{j}R^{-1}) \\ &\quad + \mathcal{O}(M^2R^{-2}) + \mathcal{O}(2^{j}M(TR)^{-1}) + \mathcal{O}(2^{2j}T^{-2}), \\ &= \mathcal{O}(2^{2j}M^{-1}) + \mathcal{O}(2^{j}R^{-1}) + \mathcal{O}(M^2R^{-2}).\end{aligned}$$

□

A.1.3 Proofs of Propositions 2.1.14 and 3.1.10

Proof of Proposition 3.1.10 ***in the presence of cross-replicate dependence***

As $M, T \to \infty$, for each j, z and ν, ν', the consistency result $\hat{S}_j(z,\nu,\nu') \xrightarrow{P} S_j(z,\nu,\nu')$ follows from the consistency results $\tilde{I}_{l,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R} \xrightarrow{P} \beta_l(z,\nu,\nu')$ for all fine enough scales l (as shown in Proposition 3.1.9) and then using the continuous mapping theorem (Billingsley, 1999) for the continuous function $g(x_1,\ldots,x_J) = \sum_{l=1}^{J} A_{j,l}^{-1}x_l$ that defines their linear combination with coefficients given by the matrix A^{-1} entries.

Additionally, using the properties of the matrix A, we obtain the estimator asymptotic unbiasedness from the linearity of the expectation operator and from the asymptotic unbiasedness of the corrected periodogram, as follows

$$\begin{aligned}\mathbf{E}(\hat{S}_j(z,\nu,\nu')) &= \mathbf{E}\left(\sum_{l=1}^{J} A_{j,l}^{-1}\tilde{I}_{l,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}\right), \\ &= \sum_{l=1}^{J} A_{j,l}^{-1}\mathbf{E}\left(\tilde{I}_{l,\lfloor zT\rfloor;T}^{(\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor);R}\right), \text{ then from the expectation part of Proposition 3.1.9} \\ &= \sum_{l=1}^{J} A_{j,l}^{-1}\left(\beta_l(z,\nu,\nu') + \mathcal{O}(2^{l}T^{-1}) + \mathcal{O}(MR^{-1})\right), \text{ and using the definition of } \beta \\ &= \sum_{l=1}^{J} A_{j,l}^{-1}\sum_{l'} A_{l,l'}S_{l'}(z,\nu,\nu') + \sum_{l=1}^{J} A_{j,l}^{-1}\left(\mathcal{O}(2^{l}T^{-1}) + \mathcal{O}(MR^{-1})\right), \\ &= \sum_{l'}\left(\sum_{l=1}^{J} A_{j,l}^{-1}A_{l,l'}\right)S_{l'}(z,\nu,\nu') + \sum_{l=1}^{J} A_{j,l}^{-1}\left(\mathcal{O}(2^{l}T^{-1}) + \mathcal{O}(MR^{-1})\right), \\ &= \sum_{l'}\left(A^{-1}A\right)_{j,l'}S_{l'}(z,\nu,\nu') + \sum_{l=1}^{J} A_{j,l}^{-1}\left(\mathcal{O}(2^{l}T^{-1}) + \mathcal{O}(MR^{-1})\right), \\ &= S_j(z,\nu,\nu') + \mathcal{O}(T^{\alpha-1}) + \mathcal{O}(MR^{-1}),\end{aligned}$$

where we used the boundedness of $A_{j,l}^{-1}$ and $\sum_{l=1}^{J} 2^{l} = \mathcal{O}(T^{\alpha})$.

In fact, it can be shown that for Haar wavelets, the above approximation rate is $\mathcal{O}(T^{-1/2}) + \mathcal{O}(MR^{-1})$ since $A_{j,l}^{-1} = \mathcal{O}(2^{-(j+l)/2})$ (Nason *et al.*, 2000). □

Proof of Proposition 2.1.14 ***under the uncorrelated replicates assumption***

The proof follows similar steps to the proof of its coherence counterpart, using the quantities defined under the assumption of uncorrelated replicates.

A.1.4 Proof of Proposition 3.1.11

It follows directly from the continuous mapping theorem (Billingsley, 1999) and the consistency results for the corrected replicate-smoothed periodogram in (55), i.e. $\hat{S}_j(z,\nu,\nu') \xrightarrow{P} S_j(z,\nu,\nu')$ as $M,T,R \to \infty$ and $M/R \to 0$. □

A.1.5 Proofs of Propositions 2.1.16 and 3.1.13

Proof of Proposition 3.1.13 (Expectation) ***in the presence of cross-replicate dependence***

From the definition of the replicate- and time- smoothed periodogram in (57), we have

$$\mathbf{E}\left[\tilde{\tilde{I}}_{j,k;T}^{(r,r');R}\right] = (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\mathbf{E}\left[I_{j,k+t;T}^{(r+s,r'+s);R}\right]$$

and substituting the asymptotic result for the expectation $\mathbf{E}\left[I_{j,k+t;T}^{(r+s,r'+s);R}\right]$ (see for instance equation (53)), we further obtain

$$\mathbf{E}\left[\tilde{\tilde{I}}_{j,k;T}^{(r,r');R}\right] = (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\left\{\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k+t}{T},\frac{r+s}{R},\frac{r'+s}{R}\right) + \mathcal{O}(2^jT^{-1}) + \mathcal{O}(R^{-1})\right\}.$$

Now using the results in Lemmas A.1.4 and A.1.5, we have

$$\begin{aligned}
\mathbf{E}\left[\tilde{\tilde{I}}_{j,k;T}^{(r,r');R}\right] &= (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\left\{\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)\right\}\\
&+ (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\sum_{l=1}^{\infty}A_{j,l}\left\{\mathcal{O}(|t|B_lT^{-1})+\mathcal{O}(|s|B_l'R^{-1})\right\}\\
&+ \mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1}),\\
&= (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)\\
&+ (2M+1)^{-1}(2M_T+1)^{-1}\sum_{s=-M}^{M}\sum_{t=-M_T}^{M_T}\left\{\mathcal{O}(|t|T^{-1})\sum_l B_lA_{j,l}+\mathcal{O}(|s|R^{-1})\sum_l B_l'A_{j,l}\right\}\\
&+ \mathcal{O}(2^jT^{-1})+\mathcal{O}(R^{-1}),\\
&= \sum_{l=1}^{\infty}A_{j,l}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r'}{R}\right)+\mathcal{O}(M_TT^{-1})+\mathcal{O}(MR^{-1})+\mathcal{O}(2^jT^{-1}),
\end{aligned}$$

where we have used that $\sum_l B_lA_{j,l}<\infty$ and $\sum_l B_l'A_{j,l}<\infty$ as previously shown, and as usual we retained the highest order terms. □

Proof of Proposition 3.1.13 (Variance) ***in the presence of cross-replicate dependence***

Let us now take the variance of the replicate- and time- smoothed periodogram in equation (57),

$$\begin{aligned}\mathrm{var}\left(\tilde{\tilde{I}}^{(r,r');R}_{j,k;T}\right) &= (2M_T+1)^{-2}\sum_{t=-M_T}^{M_T}\sum_{\delta=-M_T-t}^{M_T-t}\mathrm{cov}\left(\tilde{I}^{(r,r');R}_{j,k+t;T},\tilde{I}^{(r,r');R}_{j,k+t+\delta;T}\right),\\ &= (2M_T+1)^{-2}(2M+1)^{-2}\\ &\quad\times\sum_{t=-M_T}^{M_T}\sum_{\delta=-M_T-t}^{M_T-t}\sum_{s=-M}^{M}\sum_{\eta=-M-s}^{M-s}\mathrm{cov}\left(I^{(r+s,r'+s);R}_{j,k+t;T},I^{(r+s+\eta,r'+s+\eta);R}_{j,k+t+\delta;T}\right),\end{aligned} \tag{82}$$

where we also used the definition of the replicate-smoothed periodogram in equation (54) and we recall that in general $I^{(r,r');R}_{j,k;T} = d^{r;R}_{j,k;T}d^{r';R}_{j,k;T}$.

Using Definition 3.1.1 of the REv-LSW process and its corresponding replicate raw periodogram, let us derive in general

$$\begin{aligned}\mathbf{E}[d^{r;R}_{j,k;T}d^{r';R}_{j,k+\delta;T}] &= \mathbf{E}\left[\left\{\sum_t X^{r;R}_{t;T}\psi_{j,k}(t)\right\}\left\{\sum_t X^{r';R}_{t';T}\psi_{j,k+\delta}(t')\right\}\right]\\ &= \mathbf{E}\left[\left\{\sum_t\sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\psi_{l,m}(t)\xi^{r}_{l,m}\psi_{j,k}(t)\right\}\right.\\ &\qquad\left.\times\left\{\sum_{t'}\sum_{l'=1}^{\infty}\sum_{m'\in\mathbb{Z}}\omega^{r';R}_{l',m';T}\psi_{l',m'}(t')\xi^{r'}_{l',m'}\psi_{j,k+\delta}(t')\right\}\right],\ \text{hence,}\end{aligned}$$

$$\begin{aligned}\mathbf{E}[d^{r;R}_{j,k;T}d^{r';R}_{j,k+\delta;T}] &= \sum_t\sum_{t'}\sum_{l=1}^{\infty}\sum_{l'=1}^{\infty}\sum_{m\in\mathbb{Z}}\sum_{m'\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\omega^{r';R}_{l',m';T}\psi_{l,m}(t)\psi_{l',m'}(t')\psi_{j,k}(t)\psi_{j,k+\delta}(t')\mathbf{E}\left[\xi^{r}_{l,m}\xi^{r'}_{l',m'}\right]\\ &= \sum_{l=1}^{\infty}\sum_{m\in\mathbb{Z}}\omega^{r;R}_{l,m;T}\omega^{r';R}_{l,m;T}\rho^{r,r';R}_{l,m;T}\left\{\sum_t\psi_{l,m}(t)\psi_{j,k}(t)\right\}\left\{\sum_{t'}\psi_{l,m}(t')\psi_{j,k+\delta}(t')\right\},\end{aligned}$$

since $\mathbf{E}[\xi^{r}_{l,m}\xi^{r'}_{l',m'}] = \delta_{l,l'}\delta_{m,m'}\rho^{r,r';R}_{l,m;T}$ and so $l=l'$ and $m=m'$. Letting $n := m-k$ yields

$$\begin{aligned}\mathbf{E}[d^{r;R}_{j,k;T}d^{r';R}_{j,k+\delta;T}] &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\omega^{r;R}_{l,n+k;T}\omega^{r';R}_{l,n+k;T}\rho^{r,r';R}_{l,n+k;T}\left\{\sum_t\psi_{l,n+k-t}\psi_{j,k-t}\right\}\left\{\sum_{t'}\psi_{l,n+k-t'}\psi_{j,k-t'+\delta}\right\},\\ &= \sum_{l=1}^{\infty}\sum_{n\in\mathbb{Z}}\left\{S_l\left(\frac{n+k}{T},\frac{r}{R},\frac{r'}{R}\right)+\mathcal{O}(C'_lT^{-1})+\mathcal{O}(D'_lR^{-1})\right\}\left(\Psi_{j,l}(-n)\Psi_{j,l}(\delta-n)\right),\end{aligned}$$

where above we also used the definition of the cross-correlation wavelets, as well as the result in (72).

Then, by Lemma A.1.4 and using a substitution $n := -n$ it follows that

$$\begin{aligned}
\mathbf{E}[d^{r;R}_{j,k;T} d^{r';R}_{j,k+\delta;T}] &= \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} \left(S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(|n|B_l T^{-1}) + \mathcal{O}(C'_l T^{-1}) + \mathcal{O}(D'_l R^{-1}) \right) \\
&\quad \times \Psi_{j,l}(n)\Psi_{j,l}(n+\delta), \\
&= \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} \left(S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) \right) \\
&\quad + \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} \left\{ \mathcal{O}(|n|B_l T^{-1}) + \mathcal{O}(C'_l T^{-1}) + \mathcal{O}(D'_l R^{-1}) \right\} \Psi_{j,l}(n)\Psi_{j,l}(n+\delta), \\
&= \sum_{l=1}^{\infty} A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(T^{-1}) \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} B_l |n| \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) \\
&\quad + \mathcal{O}(T^{-1}) \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} C'_l \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) + \mathcal{O}(R^{-1}) \sum_{l=1}^{\infty} \sum_{n\in\mathbb{Z}} D'_l \Psi_{j,l}(n)\Psi_{j,l}(n+\delta).
\end{aligned}$$

The order terms in the sum above can be established as follows. Since the number of terms in the wavelet cross-correlation is finite and bounded as a function of n (Nason *et al.*, 2000) and the compact support length of $\Psi^2_{j,l}$ is bounded by $K(2^j + 2^l)$ for some constant K (Sanderson *et al.*, 2010), we have using first the triangle inequality and then the Cauchy-Schwarz inequality

$$\begin{aligned}
&\left| \sum_l \sum_n B_l |n| \mathcal{O}(T^{-1}) \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) \right| \\
&\qquad \leq \mathcal{O}(T^{-1}) \sum_l (2^j + 2^l) B_l \sum_n |\Psi_{j,l}(n)\Psi_{j,l}(n+\delta)|, \\
&\qquad \leq \mathcal{O}(T^{-1}) \sum_l (2^j + 2^l) B_l \left(\sum_n \Psi^2_{j,l}(n) \right)^{1/2} \left(\sum_n \Psi^2_{j,l}(n+\delta) \right)^{1/2}, \\
&\qquad = \mathcal{O}(T^{-1}) \sum_l (2^j + 2^l) B_l A^{1/2}_{j,l} A^{1/2}_{j,l}, \text{ from the definition of } A \\
&\qquad = \mathcal{O}(2^j T^{-1}),
\end{aligned}$$

using the same argument as for equation (73).

Using the definition of the A^{δ} matrix, we obtain

$$\sum_l \sum_n C'_l \mathcal{O}(T^{-1}) \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) = \mathcal{O}(T^{-1}) \sum_l C'_l A^{\delta}_{j,l},$$

where $\{C'_l\}$ satisfy $\sum_l 2^l C'_l < \infty$. Similarly, $\sum_l \sum_n D'_l \mathcal{O}(R^{-1}) \Psi_{j,l}(n)\Psi_{j,l}(n+\delta) = \mathcal{O}(R^{-1}) \sum_l D'_l A^{\delta}_{j,l}$, where $\{D'_l\}$ satisfy the condition $\sum_l 2^l D'_l < \infty$.

Hence, taking the highest order terms, we obtain in general

$$\mathbf{E}[d^{r;R}_{j,k;T} d^{r';R}_{j,k+\delta;T}] = \sum_{l=1}^{\infty} A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}\right) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(T^{-1} \sum_l C'_l A^{\delta}_{j,l}) + \mathcal{O}(R^{-1} \sum_l D'_l A^{\delta}_{j,l}). \tag{83}$$

In what follows we consider the individual covariance terms in equation (82), and using

Isserlis' result (Isserlis, 1918) and the expectation terms (83) with $k := k + t$, we have

$$\begin{aligned}
&\operatorname{cov}\left(I_{j,k+t;T}^{(r+s,r'+s);R}, I_{j,k+t+\delta;T}^{(r+s+\eta,r'+s+\eta);R}\right) \\
&\quad = \mathbf{E}\left[d_{j,k+t}^{r+s} d_{j,k+t+\delta}^{r+s+\eta}\right]\mathbf{E}\left[d_{j,k+t}^{r'+s} d_{j,k+t+\delta}^{r'+s+\eta}\right] + \mathbf{E}\left[d_{j,k+t}^{r+s} d_{j,k+t+\delta}^{r'+s+\eta}\right]\mathbf{E}\left[d_{j,k+t}^{r'+s} d_{j,k+t+\delta}^{r+s+\eta}\right], \\
&\quad = 2\left(\sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right) + \mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta}) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta})\right) \\
&\qquad \times \left(\sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right) + \mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta}) + \mathcal{O}(2^j T^{-1}) + \mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta})\right),
\end{aligned} \tag{84}$$

where for the last equality we used the results in Lemmas A.1.4 and A.1.5.

Hence the variance of the replicate- and time- smoothed wavelet periodogram in equation (82) is composed of terms of the following form

$$\text{'term I'} = \sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right),$$

$$\begin{aligned}
\text{'term II'} = &\sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta}) + \sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\mathcal{O}(2^j T^{-1}) \\
&+ \sum_{l=1}^{\infty} A_{j,l}^{\delta} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta}),
\end{aligned}$$

$$\begin{aligned}
\text{'term III'} = &\mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta})\mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta}) + \mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta})\mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta}) \\
&+ \mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta})\mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta}),
\end{aligned}$$

$$\text{'term IV'} = \mathcal{O}(|t|T^{-1}\sum_l B_l A_{j,l}^{\delta})\mathcal{O}(2^j T^{-1}) + \mathcal{O}(2^j T^{-1})\mathcal{O}(2^j T^{-1}) + \mathcal{O}(2^j T^{-1})\mathcal{O}(|s|R^{-1}\sum_l D_l' A_{j,l}^{\delta}).$$

In order to determine the order of the terms above, let us first establish some useful results in the following lemmas, whose proofs appear in Appendix A.2 .

Lemma A.1.6. *(a) For any sequence $\{B_l\}$ with $\sum_l 2^l B_l < \infty$, terms of the form below fulfil*

$$\sum_{\delta\in\mathbb{Z}}\left|\sum_{l=1}^{\infty} B_l A_{j,l}^{\delta}\right| = \mathcal{O}(2^j).$$

(b) For any sequences $\{B_l\}$ and $\{B_l'\}$ with $\sum_l 2^l B_l < \infty$, $\sum_l 2^l B_l' < \infty$, we have

$$\sum_{\delta\in\mathbb{Z}}\left|\sum_{l=1}^{\infty} B_l A_{j,l}^{\delta}\right|\left|\sum_{l=1}^{\infty} B_l' A_{j,l}^{\delta}\right| = \mathcal{O}(2^{2j}).$$

Lemma A.1.7. *Under the assumption* $\sup_{z,\nu\in(0,1)}\sum_{n\in\mathbb{Z}}\sum_{\eta\in\mathbb{Z}}\left|c(z,\nu,\nu+\frac{\eta}{R};n)\right| = \mathcal{O}(1)$,

we have the following properties.

(a) For any time k and replicate r, we have $\sum_{\delta\in\mathbb{Z}}\sum_{\eta\in\mathbb{Z}}\left|\sum_{l=1}^{\infty}A_{j,l}^{\delta}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r+\eta}{R}\right)\right| = \mathcal{O}(2^j)$.

(b) For any time k and replicates r and r', we have

$$\sum_{\delta\in\mathbb{Z}}\sum_{\eta\in\mathbb{Z}}\left|\sum_{l=1}^{\infty}A_{j,l}^{\delta}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r+\eta}{R}\right)\right|\left|\sum_{l=1}^{\infty}A_{j,l}^{\delta}S_l\left(\frac{k}{T},\frac{r'}{R},\frac{r'+\eta}{R}\right)\right| = \mathcal{O}(2^{2j}).$$

(c) For any time k, replicate r and for any sequence $\{B_l\}$ satisfying $\sum_l 2^l B_l < \infty$, we have

$$\sum_{\delta\in\mathbb{Z}}\sum_{\eta\in\mathbb{Z}}\left|\sum_{l=1}^{\infty}A_{j,l}^{\delta}S_l\left(\frac{k}{T},\frac{r}{R},\frac{r+\eta}{R}\right)\right|\left|\sum_{l=1}^{\infty}B_l A_{j,l}^{\delta}\right| = \mathcal{O}(2^{2j}).$$

From Lemma A.1.7, using property (b) we obtain that 'term I' in the variance sum in equation (82) is $\mathcal{O}(2^{2j}(M_T M)^{-1})$. Further, the order of 'term II' can be obtained from properties (a) and (c), as

$$(2M_T+1)^{-2}(2M+1)^{-2}\mathcal{O}(2^{2j})\sum_{t=-M_T}^{M_T}\sum_{s=-M}^{M}\left\{\mathcal{O}(|t|T^{-1})+\mathcal{O}(T^{-1})+\mathcal{O}(|s|R^{-1})\right\}$$

and its order is $\mathcal{O}(2^{2j})(\mathcal{O}((MT)^{-1})+\mathcal{O}((M_T MT)^{-1})+\mathcal{O}((M_T R)^{-1}))$.

To determine the order of 'term III' in the variance sum (82), using Lemma A.1.6, property (b) leads to the third term in the variance sum to be rephrased as

$$(2M_T+1)^{-2}(2M+1)^{-2}\mathcal{O}(2^{2j})\sum_{t=-M_T}^{M_T}\sum_{s=-M}^{M}\sum_{\eta=-M-s}^{M-s}\left\{\mathcal{O}(|t|^2T^{-2})+\mathcal{O}(|t||s|(TR)^{-1})+\mathcal{O}(|s|^2R^{-2})\right\},$$

thus yielding $\mathcal{O}(2^{2j})(\mathcal{O}(M_T T^{-2})+\mathcal{O}(M(TR)^{-1})+\mathcal{O}(M_T^{-1}M^2R^{-2}))$. Further, for 'term IV' in the variance, property (a) can be employed to yield

$$(2M_T+1)^{-2}(2M+1)^{-2}\mathcal{O}(2^{2j})\sum_{t=-M_T}^{M_T}\sum_{s=-M}^{M}\sum_{\eta=-M-s}^{M-s}\left\{\mathcal{O}(|t|T^{-2})+\mathcal{O}(M_T T^{-2})+\mathcal{O}(|s|(TR)^{-1})\right\}$$

thus resulting in $\mathcal{O}(2^{2j})(\mathcal{O}(T^{-2})+\mathcal{O}(M(M_T TR)^{-1}))$.

Hence

$$\begin{aligned}\operatorname{var}\left(\tilde{\tilde{I}}_{j,k;T}^{(r,r');R}\right) &= \mathcal{O}(2^{2j}(M_T M)^{-1})\\ &\quad+\mathcal{O}(2^{2j})(\mathcal{O}((MT)^{-1})+\mathcal{O}((M_T MT)^{-1})+\mathcal{O}((M_T R)^{-1}))\\ &\quad+\mathcal{O}(2^{2j})(\mathcal{O}(M_T T^{-2})+\mathcal{O}(M(TR)^{-1})+\mathcal{O}(M_T^{-1}M^2R^{-2}))\\ &\quad+\mathcal{O}(2^{2j})(\mathcal{O}(T^{-2})+\mathcal{O}(M(M_T TR)^{-1}))\\ &= \mathcal{O}(2^{2j}(M_T M)^{-1})+\mathcal{O}(2^{2j}M_T^{-1}M^2R^{-2}).\end{aligned}$$

□

Proof of Proposition 2.1.16 (Expectation) ***under the uncorrelated replicates assumption***

The proof follows similar steps to the expectation proof of its coherence counterpart, using the replicate- and time- smoothed periodogram in (45) and the asymptotic result for the expectation in Proposition 2.1.10.

Proof of Proposition 2.1.16 (Variance) ***under the uncorrelated replicates assumption***

Using both definitions of the replicate- time- smoothed periodogram (equation (45)) and the replicate-smoothed periodogram (equation (43)), we have

$$\begin{aligned}
\operatorname{var}\left(\tilde{\tilde{I}}^{r;R}_{j,k;T}\right) &= \operatorname{var}\left((2M_T+1)^{-1}\sum_{t=-M_T}^{M_T}\tilde{I}^{r;R}_{j,k+t;T}\right), \\
&= (2M_T+1)^{-2}\sum_{t=-M_T}^{M_T}\sum_{\delta=-M_T-t}^{M_T-t}\operatorname{cov}\left(\tilde{I}^{r;R}_{j,k+t;T},\tilde{I}^{r;R}_{j,k+t+\delta;T}\right), \\
&= (2M_T+1)^{-2}(2M+1)^{-2}\sum_{t=-M_T}^{M_T}\sum_{\delta=-M_T-t}^{M_T-t}\sum_{s=-M}^{M}\operatorname{cov}\left(I^{r+s;R}_{j,k+t;T},I^{r+s;R}_{j,k+t+\delta;T}\right).
\end{aligned}$$

Further building on results in equation (84) with $\eta = 0$ and Lemmas (A.1.6) and (A.1.7), following the same arguments as in the general variance proof, we obtain

$$\begin{aligned}
\operatorname{var}\left(\tilde{\tilde{I}}^{r;R}_{j,k;T}\right) &= \mathcal{O}(2^{2j}(M_TM)^{-1}) \\
&\quad + \mathcal{O}(2^{2j})(\mathcal{O}((MT)^{-1}) + \mathcal{O}((M_TMT)^{-1}) + \mathcal{O}((M_TR)^{-1})) \\
&\quad + \mathcal{O}(2^{2j})(\mathcal{O}(M_TM^{-1}T^{-2}) + \mathcal{O}((TR)^{-1}) + \mathcal{O}(M_T^{-1}MR^{-2})) \\
&\quad + \mathcal{O}(2^{2j})(\mathcal{O}(M^{-1}T^{-2}) + \mathcal{O}((M_TTR)^{-1})) \\
&= \mathcal{O}(2^{2j}(M_TM)^{-1}) + \mathcal{O}(2^{2j}M_T^{-1}MR^{-2}).
\end{aligned}$$

□

A.2 Proofs for lemmas of Appendix A.1

A.2.1 Proof of Lemma A.1.1

$$\begin{aligned}
\left|\sum_l A_{j,l} S_l\left(z,\nu,\nu'\right)\right| &= \left|\sum_l \sum_\tau \Psi_j(\tau)\Psi_l(\tau) S_l\left(z,\nu,\nu'\right)\right| \text{ from the definition of the matrix } A \\
&= \left|\sum_\tau \left(\sum_l S_l\left(z,\nu,\nu'\right)\Psi_l(\tau)\right)\Psi_j(\tau)\right| \\
&= \left|\sum_\tau c\left(z,\nu,\nu';\tau\right)\Psi_j(\tau)\right| \text{ from the local autocovariance definition} \\
&\leq \sum_\tau \left|c\left(z,\nu,\nu';\tau\right)\right| |\Psi_j(\tau)| \text{ using the triangle inequality} \\
&= \mathcal{O}(2^j),
\end{aligned}$$

where we used $|c\left(z,\nu,\nu';\tau\right)| < \infty$ for all ν, ν', τ and $\sum_\tau |\Psi_j(\tau)| = \mathcal{O}(2^j)$ (Nason *et al.*, 2000). □

A.2.2 Proof of Lemma A.1.2

$$\begin{aligned}
\left|\sum_l A^\tau_{j,l} S_l\left(z,\nu,\nu'\right)\right| &= \left|\sum_l \sum_n \Psi_j(n)\Psi_l(n+\tau) S_l\left(z,\nu,\nu'\right)\right| \text{ from the definition of the matrix } A^\tau \\
&= \left|\sum_n \left(\sum_l S_l\left(z,\nu,\nu'\right)\Psi_l(n+\tau)\right)\Psi_j(n)\right| \\
&= \left|\sum_n c\left(z,\nu,\nu';n+\tau\right)\Psi_j(n)\right| \text{ from the local autocovariance definition} \\
&\leq \sum_n \left|c\left(z,\nu,\nu';n+\tau\right)\right| |\Psi_j(n)| \text{ using the triangle inequality} \\
&= \mathcal{O}(2^j),
\end{aligned}$$

where we used $|c\left(z,\nu,\nu';n+\tau\right)| < \infty$ for all τ, ν, ν', n and $\sum_n |\Psi_j(n)| = \mathcal{O}(2^j)$. □

A.2.3 Proof of Lemma A.1.3

Let us first show that for any τ and scales j, j', l, we have

$$\sum_{n\in\mathbb{Z}} \Psi_{j,j'}(n)\Psi_l(n+\tau) = \sum_{u\in\mathbb{Z}} \Psi_{j,l}(u)\Psi_{j',l}(u+\tau).$$

This can be seen by re-expressing the right hand of the equality above by means of the

definition of the cross-correlation wavelets, hence obtaining

$$\begin{aligned}
\sum_u \Psi_{j,l}(u)\Psi_{j',l}(u+\tau) &= \sum_u \left(\sum_p \psi_{j,p}\psi_{l,p-u}\right)\left(\sum_{p'} \psi_{j',p'}\psi_{l,p'-u-\tau}\right), \\
&= \sum_p \sum_{p'} \psi_{j,p}\psi_{j',p'}\left(\sum_u \psi_{l,p-u}\psi_{l,p'-u-\tau}\right), \\
&= \sum_p \sum_{p'} \psi_{j,p}\psi_{j',p'}\Psi_l(p-p'+\tau), \\
&= \sum_n \left(\sum_{p'} \psi_{j,n+p'}\psi_{j',p'}\right)\Psi_l(n+\tau), \text{ where } n := p-p', \\
&= \sum_n \Psi_{j,j'}(n)\Psi_l(n+\tau).
\end{aligned}$$

Using the above equality, we use the fact that the spectrum is positive and now take the triangle inequality

$$\begin{aligned}
\sum_l S_l\left(z,\nu,\nu'\right)\left|\sum_n \Psi_{j,l}(n)\Psi_{j',l}(n+\tau)\right| &\leq \sum_l S_l\left(z,\nu,\nu'\right)\sum_n \left|\Psi_{j,l}(n)\Psi_{j',l}(n+\tau)\right|, \\
&\leq \sum_l S_l\left(z,\nu,\nu'\right)\left(\sum_n \Psi^2_{j,l}(n)\right)^{1/2}\left(\sum_n \Psi^2_{j',l}(n+\tau)\right)^{1/2}, \\
&= \sum_l S_l\left(z,\nu,\nu'\right)A^{1/2}_{j,l}A^{1/2}_{j',l}, \text{ from the definition of matrix } A \\
&\leq \left(\sum_l A_{j,l}S_l\left(z,\nu,\nu'\right)\right)^{1/2}\left(\sum_{l'} A_{j',l'}S_{l'}\left(z,\nu,\nu'\right)\right)^{1/2}, \\
&= \mathcal{O}(2^{(j+j')/2}),
\end{aligned}$$

where for the last two inequalities above we used the Cauchy-Schwarz inequality and for the last equality we used the result in Lemma A.1.1. □

A.2.4 Proof of Lemma A.1.4

Under the assumptions of Definition 3.1.1, the functions $\widetilde{W}_j(\cdot\,,\nu)$, $\widetilde{W}_j(\cdot\,,\nu')$ and $\rho_j(\cdot\,,\nu,\nu')$ are Lipschitz continuous in rescaled time, with Lipschitz constants L_j^{ν}, $L_j^{\nu'}$ and $Q_j^{\nu,\nu'}$ respectively. Equivalently, this can be written as $\left|\widetilde{W}_j((k+n)/T,\nu) - \widetilde{W}_j(k/T,\nu)\right| \leq |n|L_j^{\nu}T^{-1}$, $\left|\widetilde{W}_j((k+n)/T,\nu') - \widetilde{W}_j(k/T,\nu')\right| \leq |n|L_j^{\nu'}T^{-1}$ and $|\rho_j((k+n)/T,\nu,\nu') - \rho_j(k/T,\nu,\nu')| \leq |n|Q_j^{\nu,\nu'}T^{-1}$.

From standard Lipschitz function theory, the product of the Lipschitz continuous functions defined on a compact interval is also Lipschitz continuous with Lipschitz constant the maximum of the individual constants. Hence working across replicates and denoting $B_j = \sup_{\nu,\nu'} \max\{L_j^{\nu}, L_j^{\nu'}, Q_j^{(\nu,\nu')}\}$, we readily obtain that $S_j(\cdot\,,\nu,\nu') = \widetilde{W}_j(\cdot\,,\nu)\widetilde{W}_j(\cdot\,,\nu')\rho_j(\cdot\,,\nu,\nu')$ is Lipschitz continuous (the result in equation (70)) with constants $\{B_j\}$. The properties of the $\{B_j\}$ follow immediately from the similar properties of the individual Lipschitz constants $L_j = \sup_{\nu} L_j^{\nu}$ and $Q_j = \sup_{\nu,\nu'} Q_j^{\nu,\nu'}$. □

A.2.5 Proof of Lemma A.1.5

Under the assumptions of Definition 3.1.1, for any rescaled time z and trials (replicates) respectively ν', ν, the limiting coherence functions $\rho_j(z,\cdot\,,\nu')$ and $\rho_j(z,\nu,\cdot\,)$ are Lipschitz continuous with Lipschitz constants $\{P_j^z\}$ and are defined on a compact interval. Then it immediately follows that there exists a bounded constant C such that $|\rho_j(z,\nu+a,\nu'+a)-\rho_j(z,\nu,\nu')| \le C\,P_j^z|a|$ for any a such that $\nu+a,\nu'+a\in(0,1)$.

Specifically, the Lipschitz continuity assumption in rescaled trial-dimension implies that for each z we have $\left|\widetilde{W}_j(z,(r+s)/R)-\widetilde{W}_j(z,r/R)\right| \le |s|N_j^zR^{-1}$, $\left|\widetilde{W}_j(z,(r'+s)/R)-\widetilde{W}_j(z,r'/R)\right| \le |s|N_j^zR^{-1}$ and $|\rho_j(z,(r+s)/R,(r'+s)/R)-\rho_j(z,r/R,r'/R)| \le |s|P_j^zR^{-1}$ at any replicates r, r' and s.

Then, from standard Lipschitz function theory it follows that their product $S_j(z,\nu+\cdot\,,\nu'+\cdot\,) = \widetilde{W}_j(z,\nu+\cdot\,)\widetilde{W}_j(z,\nu'+\cdot\,)\rho_j(z,\nu+\cdot\,,\nu'+\cdot\,)$ is also Lipschitz continuous (the desired result in equation (71)) with constants $B_j' = \sup_z \max\{N_j^z,P_j^z\}$. The properties of the $\{B_j'\}$ follow immediately from the similar properties of the individual Lipschitz constants $N_j=\sup_z N_j^z$ and $P_j=\sup_z P_j^z$. □

A.2.6 Proof of Lemma A.1.6

(a) The desired result follows since

$$\begin{aligned}
\sum_\delta\left|\sum_l B_lA_{j,l}^\delta\right| &= \sum_\delta\left|\sum_l B_l\sum_n \Psi_j(n)\Psi_l(n+\delta)\right| \text{ from the definition of the matrix } A^\delta\\
&\le \sum_l B_l\sum_\delta\sum_n|\Psi_j(n)\Psi_l(n+\delta)| \text{ using the triangle inequality}\\
&= \sum_l B_l\sum_n\left(|\Psi_j(n)|\sum_\delta|\Psi_l(n+\delta)|\right)\\
&\le K\sum_l 2^lB_l\sum_n|\Psi_j(n)|\\
&= \mathcal{O}(2^j),
\end{aligned}$$

where we used (in order) that $\sum_\delta|\Psi_l(n+\delta)| = \mathcal{O}(2^l)$, $\sum_n|\Psi_j(n)| = \mathcal{O}(2^j)$ and $\sum_l 2^lB_l<\infty$.

(b) This follows directly from the above.

□

We also note here that $\left|\sum_{l=1}^\infty 2^{-l}A_{j,l}^\delta\right| = \mathcal{O}(1)$ for any integer δ, which can be obtained

by taking

$$\begin{aligned}\left|\sum_l 2^{-l} A^{\delta}_{j,l}\right| &= \left|\sum_l 2^{-l} \sum_n \Psi_j(n)\Psi_l(n+\delta)\right|, \\ &= \left|\sum_n \Psi_j(n) \sum_l 2^{-l}\Psi_l(n+\delta)\right|, \\ &= \left|\sum_n \Psi_j(n)\delta_{0,n+\delta}\right|, \\ &= |\Psi_j(-\delta)| = \mathcal{O}(1)\end{aligned}$$

where we used $\sum_l 2^{-l}\Psi_l(n+\delta) = \delta_{0,n+\delta}$ as shown in Fryzlewicz *et al.* (2003).

A.2.7 Proof of Lemma A.1.7

(a) Using the definitions of the A^{δ} matrix and of the local covariance we have

$$\begin{aligned}\sum_\delta \sum_\eta \left|\sum_l A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\right| &= \sum_\delta \sum_\eta \left|\sum_\tau \left(\sum_l S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\Psi_l(\tau+\delta)\right)\Psi_j(\tau)\right|, \\ &= \sum_\delta \sum_\eta \left|\sum_\tau c\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}; \tau+\delta\right)\Psi_j(\tau)\right|, \\ &\leq \sum_\tau \left(|\Psi_j(\tau)| \sum_\delta \sum_\eta \left|c\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}; \tau+\delta\right)\right|\right), \\ &= \mathcal{O}(1)\sum_\tau |\Psi_j(\tau)| = \mathcal{O}(2^j),\end{aligned}$$

where we used the triangle inequality and the autocorrelation wavelet property $\sum_\tau |\Psi_j(\tau)| = \mathcal{O}(2^j)$.

(b) Using

$$\begin{aligned}&\sum_\delta \sum_\eta \left|\sum_l A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\right| \left|\sum_l A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right)\right| \\ &\quad\leq \left(\sum_\delta \sum_\eta \left|\sum_l A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r}{R}, \frac{r+\eta}{R}\right)\right|\right)\left(\sum_\delta \sum_\eta \left|\sum_l A^{\delta}_{j,l} S_l\left(\frac{k}{T}, \frac{r'}{R}, \frac{r'+\eta}{R}\right)\right|\right), \\ &\quad= \mathcal{O}(2^{2j}),\end{aligned}$$

where for the last equality we used the property (a).

(c) It follows directly from property (a) and equation (85).

□

A.3 Further proofs

A.3.1 Proof of Propositions 2.1.7 and 3.1.5

Proof of Proposition 2.1.7 ***under the uncorrelated replicates assumption***

As the REv-LSW process is defined to have zero-mean, we have $\operatorname{cov}\left(X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor;T}, X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor+\tau;T}\right) = \mathbf{E}\left(X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor;T} X^{\lfloor \nu R \rfloor;T}_{\lfloor zT \rfloor+\tau;T}\right)$ and using the REv-LSW process definition in equation (35), we obtain

$$\begin{aligned}
\operatorname{cov}\left(X^{\lfloor \nu R \rfloor;R}_{\lfloor zT \rfloor;T}, X^{\lfloor \nu R \rfloor;R}_{\lfloor zT \rfloor+\tau;T}\right) &= \sum_{j=1}^{\infty}\sum_{k\in\mathbb{Z}} \left(\omega^{\lfloor \nu R \rfloor;R}_{j,k;T}\right)^2 \psi_{j,k}(\lfloor zT \rfloor)\psi_{j,k}(\lfloor zT \rfloor + \tau), \text{ take } k := n + \lfloor zT \rfloor, \\
&= \sum_{j=1}^{\infty}\sum_{n\in\mathbb{Z}} \left(\omega^{\lfloor \nu R \rfloor;R}_{j,n+\lfloor zT \rfloor;T}\right)^2 \psi_{j,n+\lfloor zT \rfloor}(\lfloor zT \rfloor)\psi_{j,n+\lfloor zT \rfloor}(\lfloor zT \rfloor + \tau), \\
&= \sum_{j=1}^{\infty}\sum_{n\in\mathbb{Z}} \left(\omega^{\lfloor \nu R \rfloor;R}_{j,n+\lfloor zT \rfloor;T}\right)^2 \psi_{j,n}(0)\psi_{j,n}(\tau).
\end{aligned}$$

Denoting $A = \left|\operatorname{cov}\left(X^{\lfloor \nu R \rfloor;R}_{\lfloor zT \rfloor;T}, X^{\lfloor \nu R \rfloor;R}_{\lfloor zT \rfloor+\tau;T}\right) - c(z,\nu;\tau)\right|$ and using the local covariance definition as well as the amplitude approximations in equation (41), we obtain

$$A = \left|\sum_{j=1}^{\infty}\sum_{n\in\mathbb{Z}} \left(\omega^{\lfloor \nu R \rfloor;R}_{j,n+\lfloor zT \rfloor;T}\right)^2 \psi_{j,n}(0)\psi_{j,n}(\tau) - c(z,\nu;\tau)\right|,$$

a quantity bounded by
$\left|\sum_j \sum_n \left(S_j(z+\frac{n}{T},\nu) + (D_j R^{-1}) + (C_j T^{-1})\right)\psi_{j,n}(0)\psi_{j,n}(\tau) - \sum_j S_j(z,\nu)\Psi_j(\tau)\right|$. Using the spectrum Lipschitz continuity in time, we further bound A by
$\left|\sum_j \sum_n \left(S_j(z,\nu) + (D_j R^{-1}) + (C_j T^{-1}) + (L_j|n|T^{-1})\right)\psi_{j,n}(0)\psi_{j,n}(\tau) - \sum_j S_j(z,\nu)\Psi_j(\tau)\right|$.

Hence

$$\begin{aligned}
A &\leq \left|\sum_j\sum_n \left((D_j R^{-1}) + (C_j T^{-1}) + (L_j|n|T^{-1})\right)\psi_{j,n}(0)\psi_{j,n}(\tau)\right| \text{ as } \Psi_j(\tau) = \sum_k \psi_{j,k}(0)\psi_{j,k}(\tau), \\
&\leq \left|\sum_j\sum_n (D_j R^{-1})\psi_{j,n}(0)\psi_{j,n}(\tau)\right| \\
&+ \left|\sum_j\sum_n (C_j T^{-1}\psi_{j,n}(0)\psi_{j,n}(\tau)\right| \\
&+ \left|\sum_j\sum_n (L_j|n|T^{-1})\psi_{j,n}(0)\psi_{j,n}(\tau)\right|, \\
&= \mathcal{O}(R^{-1}) + \mathcal{O}(T^{-1}),
\end{aligned}$$

where the last equality follows as the terms $\sum_j\sum_n (D_j R^{-1})\psi_{j,n}(0)\psi_{j,n}(\tau) = R^{-1}\sum_j D_j\Psi_j(\tau) = \mathcal{O}(R^{-1})$ since $\Psi_j(\tau) = \mathcal{O}(1)$ (Nason *et al.*, 2000) and we assumed $\sum_j D_j < \infty$, and similarly $\sum_j\sum_n (C_j T^{-1})\psi_{j,n}(0)\psi_{j,n}(\tau) = \mathcal{O}(T^{-1})$ using $\sum_j C_j < \infty$; the last term $\left|\sum_j\sum_n (L_j|n|T^{-1})\psi_{j,n}(0)\psi_{j,n}(\tau)\right| \leq \mathcal{O}(T^{-1})\sum_j 2^j L_j$, since there are at most order 2^j non-zero terms in the wavelet product, $\Psi_j(\tau) = \mathcal{O}(1)$ and we assumed $\sum_j 2^j L_j < \infty$ (see

also Nason *et al.* (2000)). □

Proof of Proposition 3.1.5 ***in the presence of cross-replicate dependence***

For the general case, the proof follows immediately from the amplitude and coherence approximation properties in Definition 3.1.1 of the REv-LSW process, which yield

$$\left|\omega_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor;R}\,\omega_{j,\lfloor zT\rfloor;T}^{\lfloor \nu' R\rfloor;R}\,\rho_{j,\lfloor zT\rfloor;T}^{\lfloor \nu R\rfloor,\lfloor \nu' R\rfloor;R} - S_j(z,\nu,\nu')\right| = \mathcal{O}(C_j' T^{-1}) + \mathcal{O}(D_j' R^{-1}),$$

where C_j' and D_j' can be shown to also fulfill equations of the type in Definition 3.1.1, 3c and 4b.

As the limiting amplitudes ($\widetilde{W}_j(\cdot\,,\nu)$, $\widetilde{W}_j(\cdot\,,\nu')$) and coherence ($\rho_j(\cdot\,,\nu,\nu')$) are Lipschitz continuous functions in rescaled time for rescaled trials ν, ν' and are defined on a compact interval, it follows from standard Lipschitz function theory that their product ($S_j(\cdot\,,\nu,\nu')$) is also Lipschitz continuous in rescaled time, as shown in Lemma A.1.4. The desired conclusion then follows using the same type of arguments as employed for uncorrelated trials above. □

A.3.2 Proof of estimated coherence limits for equation (56)

Let us first show that $\left|\tilde{I}_{l,k;T}^{(r,r');R}\right| \leq \sqrt{\left|\tilde{I}_{l,k;T}^{(r,r);R}\right|}\sqrt{\left|\tilde{I}_{l,k;T}^{(r',r');R}\right|}$ for any scale l, time k and trials r, r'. Equivalently, we want to show that $\left(\tilde{I}_{l,k;T}^{(r,r');R}\right)^2 \leq \left(\tilde{I}_{l,k;T}^{(r,r);R}\right)\left(\tilde{I}_{l,k;T}^{(r',r');R}\right)$.

Recalling that we use the same smoothing window and using the definitions of the trial-smoothed periodogram in (54) and of the raw periodogram in (53), the above is equivalent to having to show that

$$\left(\sum_s d_{l,k;T}^{r+s;R} d_{l,k;T}^{r'+s;R}\right)^2 \leq \left(\sum_s \left(d_{l,k;T}^{r+s;R}\right)^2\right)\left(\sum_{s'} \left(d_{l,k;T}^{r'+s';R}\right)^2\right),$$

which indeed follows from the Cauchy-Schwarz inequality.

Since the components in the estimated between-trial coherence equation (56) are of the form $\hat{S}_j(\frac{k}{T}, \frac{r}{R}, \frac{r'}{R}) = \sum_l A_{j,l}^{-1} \tilde{I}_{l,k;T}^{(r,r');R}$ (also see equation (55)), next let us take

$$\begin{aligned}
\left|\sum_l A_{j,l}^{-1} \tilde{I}_{l,k;T}^{(r,r');R}\right|^2 &\leq \left(\sum_l \left|A_{j,l}^{-1}\right| \left|\tilde{I}_{l,k;T}^{(r,r');R}\right|\right)^2, \text{ and using the relationship above} \\
&\leq \left(\sum_l \left(\sqrt{\left|A_{j,l}^{-1}\right|}\right)^2 \sqrt{\left|\tilde{I}_{l,k;T}^{(r,r);R}\right|}\sqrt{\left|\tilde{I}_{l,k;T}^{(r',r');R}\right|}\right)^2, \\
&= \left(\sum_l \left(\sqrt{\left|A_{j,l}^{-1}\right|}\sqrt{\left|\tilde{I}_{l,k;T}^{(r,r);R}\right|}\right)\left(\sqrt{\left|A_{j,l}^{-1}\right|}\sqrt{\left|\tilde{I}_{l,k;T}^{(r',r');R}\right|}\right)\right)^2, \\
&\leq \left(\sum_l A_{j,l}^{-1} \tilde{I}_{l,k;T}^{(r,r);R}\right)\left(\sum_{l'} A_{j,l'}^{-1} \tilde{I}_{l',k;T}^{(r',r');R}\right),
\end{aligned}$$

where the last inequality resulted from the Cauchy-Schwarz inequality. Hence the values of the squared estimated between-trial coherence have the property $\left|\hat{\rho}_{j,k;T}^{(r,r');R}\right| \leq 1$. □

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.

Abramovich, F., Bailey, T. and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician* , 49, 1–29.

Ahamada, I. and Boutahar, M. (2002). Tests for covariance stationarity and white noise, with an application to Euro/US dollar exchange rate: an approach based on the evolutionary spectral density. *Econ. Lett.*, 77, 177–186.

Allan, D. W. (1966). Statistics of atomic frequency clocks. In: *Proceedings of the IEEE*, 54, 221–230.

Antoniadis, A. and Gijbels, I. (2002). Detecting abrupt changes by wavelet methods. *J. Nonparam. Statist.*, 14, 7–29.

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.

Atkinson, A. D., Hill, R. R., Pignatiello, J. J. Jr., Vining, G. G., White, E. D., and Chicken, E. (2017). Wavelet ANOVA approach to model validation. *Simul. Model. Pract. Theory*, 78, 18–27.

Bachman, G., Narici, L. and Beckenstein, E. (2000). *Fourier and Wavelet Analysis.* Springer.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, 57, 289–300.

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008). *Time Series Analysis: Forecasting and Control.* John Wiley & Sons Inc.

Brillinger, D.R. (2001). *Time Series: Data Analysis and Theory.* SIAM.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* Springer.

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

Cardinali, A. and Nason, G. P. (2010) Costationarity of locally stationary time series. *J. Time Ser. Econom.*, 2, doi:10.2202/1941-1928.1074.

Chang, X. and Stein, M. L. (2013). Decorrelation property of discrete wavelet transform under fixed-domain asymptotics. *IEEE Trans. Inf. Theory*, 59, 8001–8013.

Chatfield, C. (2003). *The Analysis of Time Series.* Chapman and Hall.

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

Chiann, C. and Morettin, P. (1998). A wavelet analysis for time series. *J. Nonparam. Statist.*, 10, 1–46.

Cho, H. (2016). A test for second-order stationarity of time series based on unsystematic sub-samples. *Stat*, 5, 262–277.

Chui, C. K. (1997). *Wavelets: a Mathematical Tool for Signal Analysis*. SIAM.

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* , 1, 54–81.

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. In: Antoniadis A., Oppenheim G. (eds) *Wavelets and Statistics*, Lecture Notes in Statistics, vol 3, 125–150, Springer

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. *The Annals of Statistics*, 25, 1–37.

Dahlhaus, R. (2012). Locally stationary processes. In: Subba Rao T., Subba Rao S., Rao C. (eds) *Handbook of Statistics*, vol 30, 351–413, Elsevier

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for locally stationary ARCH models. *The Annals of Statistics*, 34, 1075–1114.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comms. Pure Appl. Math.*, 41, 909–996.

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

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

Debnath, L. (2002). *Wavelets Transforms and Their Applications*. Birkhäuser.

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Annals of Applied Statistics*, 29, 1361–1400.

Dette, H. and Wu, W. (2020). Prediction in locally stationary time series. *J. Bus. Econ. Stat.*, 1–12, doi 10.1080/07350015.2020.1819296.

Diggle, P J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data.* Oxford University Press.

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

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425–455.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). *J. R. Statist. Soc. B* , 57, 301–369.

Dwivedi, Y. and Subba Rao, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Ser. Anal.*, 32, 68–91.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.

Fan, J. and Lin, S-K. (1998). Test of significance when data are curves. *J. Am. Statist. Ass.*, 93, 1007–1021.

Fan, J. and Zhang, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Statist. Soc. B*, 62, 303–322.

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

Fryzlewicz, P. and Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *J. Am. Statist. Ass.*, 104, 299–312.

Fryzlewicz, P., Van Bellegem, S. and von Sachs, R. (2003) Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Statist. Math.*, 55, 737–764.

Gabor, D. (1946). Theory of communications. *J. Inst. Electr. Eng.*, 93, 429–457.

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. (2011). Exploring dependence between brain signals in a monkey during learning. *J. Time Ser. Anal.*, 33, doi:10.1111/j.1467-9892.2011.00767.x.

Greven, S., Crainiceanu, C., Caffo B. and Reich, D. (2010). Longitudinal functional principal components analysis. *Electron. J. Stat.*, 4, 1022–1054.

Guo, W., Dai, M. Ombao, H. and von Sachs, R. (2003). Smoothing spline ANOVA for time-dependent spectral analysis. *J. Am. Statist. Ass.*, 98, 643–652.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Math. Ann.*, 69, 331–371.

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. *Annals of Applied Statistics*, 13, 1817–1846, doi 10.1214/19-AOAS1246.

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.

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.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12, 134–139.

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.

Krzemieniewska, K., Eckley, I. A. and Fearnhead, P. (2014). Classification of non-stationary time series. *Stat*, 3, 144–157.

Kumar, P. and Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of Geophysics*, 35, 385–412.

Larsen, R. (1974). Vibrating strings, Fourier series, and some aspects of mathematical analysis. *Nordisk Matematisk Tidskrift*, 22, 85–102.

Mallat, S. G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Am. Math. Soc.*, 315, 69–87.

Mallat, S. G. (1989b). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 11, 674–693.

Mallat, S. G. (2009). *A Wavelet Tour of Signal Processing (Third Edition)*. Academic Press.

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.

McClung, C. R. (2006). Plant circadian rhythms. *The Plant Cell*, 18, 792–803.

McKay, J. L., Welch, T. D. J., Vidakovic, B. and Ting, L. H. (2013). Statistically significant contrasts between EMG waveforms revealed using wavelet-based functional ANOVA. *J. Neurophysiol.*, 109, 591–602.

Meyer, Y. (1993). *Wavelets and Operators*. Cambridge University Press.

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

Morris, J. S., Baladandayuthapni, V., Herrick, R. C., Sanna, P., and Gutstein, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Annals of Applied Statistics*, 5, 894–923.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *J. R. Statist. Soc. B*, 68, 179–199.

Müller, H. G. (2005). Functional modelling and classification of longitudinal data. *Scand. J. Statist.*, 32, 223–240.

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. (2012). Simulation study comparing two tests of second-order stationarity and confidence intervals for localized autocovariance. *Technical Report 12:02*. Statistics Group, University of Bristol, Bristol. *arXiv:1603.06415*.

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. and Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. In: Antoniadis A., Oppenheim G. (eds) *Wavelets and Statistics*, Lecture Notes in Statistics, vol 3, 281–299, Springer

Nason, G. P. and von Sachs, R. (1999). Wavelets in time series analysis. *Phil. Trans. R. Soc. Lond. A*, 357, 2511–2526.

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. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.*, 17, 601–633.

Neumann, M. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, 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.

Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli*, 15, 1190–1221.

Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Am. Statist. Ass.*, 105, 839–851.

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.

Percival, D. B. and Guttorp, P. (1994). Long memory processes, the Allan variance and wavelets. In: Foufoula-Georgiou E., Kumar P. (eds) *Wavelets in Geophysics*, vol 4, 325–344, Academic Press

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time series Analysis*. Cambridge University Press.

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

Priestley, M. B. (1982). *Spectral Analysis and Time Series*. Academic Press.

Priestley, M. B. and Subba Rao, T. (1969). A test for non-stationarity of time-series. *J. R. Statist. Soc. B*, 31, 140–149.

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

Ramsay, O. J. and Silverman, B. M. (2005). *Functional Data Analysis*. Springer.

Raz, J. and Turetsky, B. (1999). Wavelet ANOVA and fMRI. In: *Proceedings of the SPIE: Wavelet applications in signal and image processing VII*, 3813.

Sanderson, J. (2010). Wavelet methods for time series with bivariate observations and irregular sampling grids. *PhD Thesis*, University of Bristol.

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.

Semmlow, J. L. (2012). *Signals and Systems for Bioengineers: A MATLAB-based Introduction*. Academic Press.

Shumway, R. H. (1988). *Applied Statistical Time Series Analysis*. Prentice-Hall.

Shumway, R. H. and Stoffer, D. S (2017). *Time Series Analysis and Its Applications: fourth edition*. Springer.

Strang, G. and Nguyen, T. (1996). *Wavelets and Filter Banks*. Wellesley-Cambridge Press.

Taylor, S. L., Eckley, I. A. and Nunes, M. A. (2014). A test of stationarity for textured images. *Technometrics*, 56, 291–301.

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.

Tsay, R. S. (2010). *Analysis of Financial Time Series, Third Edition*. John Wiley & Sons, Inc.

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

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

Vidakovic, B. (2001). Wavelet-based functional data analysis: Theory, applications and ramifications. In: *Proceedings of PSFVIP-3*, Maui, Hawaii.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40, 2601–2633.

von Sachs, R. and Neumann, M. (2000). A wavelet-based test for stationarity. *J. Time Ser. Anal.*, 21, 597–613.

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.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, 100, 577–591.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33, 2873–2903.