Wavelet methods for the statistical analysis of image texture

Sarah L. Taylor, BSc. (Hons.)

Department of Mathematics and Statistics

Lancaster University

Submitted for the degree of Doctor of Philosophy at Lancaster University.

July 2013

Sarah L. Taylor, BSc. (Hons.)

Submitted for the degree of Doctor of Philosophy at Lancaster University, April 2013.

Abstract

This thesis considers the application of locally stationary wavelet-based stochastic models to the analysis of image texture. In the first part we propose a test of stationarity for spatial data on a regular grid. This test is then incorporated into a segmentation framework in order to determine the number of textures contained within an image, a key feature to many texture segmentation approaches. These novel methods are subsequently applied to various texture analysis problems arising from work with an industrial collaborator. The second part of this thesis considers the modelling of the spectral structure of a non-stationary *multivariate* image, i.e. an image containing different colour channels. We propose a multivariate locally stationary wavelet-based modelling framework which permits a measure of dependence between pairs of channels. The performance of this modelling approach is then assessed using various colour texture examples encountered by an industrial collaborator.

Acknowledgements

Firstly I would like to thank my supervisor (aka "Boss") Idris Eckley, for his continuing support, guidance and patience throughout the last few years. I have really enjoyed sharing this journey with you.

I am very grateful to the EPSRC and Unilever Research for funding me. I've also had the pleasure of extra advice, enthusiasm and hospitality on visits to Unilever from Eric Mahers, Rob Treloar and Brian Newby.

I'd like to thank the Mathematics and Statistics department at Lancaster University for all the support it has given me over many years. In particular, I would like to thank the B57 legends Jenny, Shell and Mark for providing such a banterous office and making coming into work such a fun experience. In addition thanks to Aimée, Jak and my other department fellows for all the pep talks, advice and most importantly giggles! An extra special thanks to Becki Killick for her patience in proof reading.

To everyone in Lancaster, past and present who has made the last four years such a fun experience, whether it be my fellow PhD goers, Graduate College peeps or the lovely Lancaster locals, I have had a blast!

To friends from other walks of life, the Pendle Posse, Maths Geeks and Richmond Ladies for always being there and helping me stay true to myself. Special thanks to Ellie, Jo, Cat and Andy for the endless phone dates and keeping my spirits up especially in the last few months.

Penultimately I would like to thank my family, especially Mum, Dad and my sister Rachel. It is impossible to express how much I am grateful for your continuing encouragement, guidance and support. In particular to my parents for always going the extra mile, "You're simply the best!"

Finally to Matt, for being an amazing collaborator, housemate and friend. Thank you for everything. I dedicate this to you!

Declaration

I declare that this thesis is my own work and has not been submitted in any form for the award of a higher degree elsewhere.

Sarah L. Taylor

List of Papers

This thesis has the following papers:

- Taylor S. L., Eckley I. A., Nunes M. A., (2013), *A test of stationarity for textured images*, Submitted.
- Taylor S. L., Nunes M. A., Eckley I. A., (2013), *A multiscale approach to determining the number of textures within an image*, Submitted.

The papers are contained in Chapters 3 and 4 respectively. Please note that in Chapter 3 Section 3.2 includes additional detail to that presented in Taylor et al. (2013). This new, background material, is included to ensure the completeness and accessibility of the work presented in this thesis. In addition one of the examples provided in Section 3.4.1 differs from the paper. Each paper contains its own motivation and review of the appropriate literature. The relevant appendices can be found at the end of each chapter, whilst the bibliography is listed at the end of the thesis. Chapter 4 contains a paper based on a problem of broad interest to the image processing community, thus the style of the paper reflects this.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since their inception wavelets have proven popular in a number of statistical applications, such as signal denoising, nonparametric curve estimation and the modelling, analysis and forecasting of time series. Wavelets offer locality in both time and scale, a property which traditional Fourier or time domain methods do not. This proves to be a significant advantage when analysing data with locally changing behaviour (Dahlhaus, 2012).

In particular, in recent years wavelets have been used to develop models for the locally stationary covariance structure of both time series and spatial data defined on a regular lattice. The seminal paper of Nason et al. (2000) introduces a model for locally stationary wavelet processes (LSW) applicable in a time series setting. This modelling paradigm was then extended to two dimensions by Eckley et al. (2010), introducing 2D locally stationary wavelet processes (LS2W) on a regular grid. This has proved useful for analysing textured images. Both the LSW and LS2W models permit a location-scale decomposition of the covariance structure of a process, allowing stationarity within local regions. This has significant advantages over traditional Fourier-based spatial models which assume stationarity over the entire data. A review of the key properties of wavelets and their associated transforms is given in Chapter 2, together with a brief introduction to the main application focus of this thesis: texture analysis.

The first part of the thesis focuses on problems arising with univariate, (i.e.) greyscale texture images. In many problems within texture analysis it is useful to be able to detect non-stationary structure, for example to identify whether there is an area of uneven wear

within a sample of material. This can be challenging especially if the changes are visually subtle. To avoid the subjectivity of human inspection of materials it is thus desirable to develop an automatic detection method for uneven wear. Chapter 3 addresses this problem by proposing a test of stationarity for random fields on a regular lattice.

Existing approaches to texture analysis often require the number of distinct textures in an image to be known a priori. In practice this information is rarely available and has to be estimated along with the stationary regions. Chapter 4 embeds the stationarity test within a quadtree framework in order to distinguish the number of textures in an image and highlight areas of subtle local texture change. We conclude the above chapters by demonstrating the proposed methodology on both simulated and real examples. In particular, showing how the proposed approach in Chapter 4 can be used (unsupervised) to identify the number of textures in a multi-texture image. The software that implements the methods from Chapters 3 and 4 is available in the R package LS2Wstat, details of which are contained in Appendix B.

The second part of the thesis considers a problem arising from colour texture analysis. Chapter 5 develops a novel framework for the modelling of non-stationary multivariate images, together with an associated estimation scheme. In particular, an extension to multivariate images requires consideration of the additional cross-covariance structure (dependency measure) between image channels which we denote coherence. The theoretical properties of the model estimates are also considered in Chapter 5, whilst a case study exploring the applicability of the approach to the discrimination and classification of colour texture images is presented in Chapter 6.

Chapter 2

Literature Review

Over recent years there has been an increase in wavelet techniques in the scientific literature across various fields such as engineering, geology, medicine and signal processing. This explosion in the literature is due to wavelets offering specific advantages over Fourier approaches, such as their multiscale nature and ability to decompose a signal in both the time and scale domain simultaneously.

The first part of this chapter provides a brief introduction to wavelets and their various transforms. We begin by giving a short summary of Fourier theory, highlighting the shortcomings of this approach with respect to key features which we encounter in everyday signals and images. Following this, Section 2.1 provides a detailed description of the fundamental properties of wavelets by considering *multiresolution analysis*. In Section 2.2 the discrete wavelet transform is introduced. Section 2.3 then provides accounts of other wavelet transforms including the non-decimated wavelet transform. The use of wavelets in statistics is introduced in Section 2.4, paying particular attention to locally stationary (1D) wavelet processes. Having reviewed the use of wavelets in one-dimension, the final part of the chapter is devoted to two-dimensional wavelets and their transforms (Section 2.5). We conclude in Section 2.6 by briefly reviewing statistical approaches to texture analysis, which provides the main application area for the novel methodology developed in this thesis.

2.1 Introduction to wavelets and their transforms

2.1.1 Fourier Analysis

We begin our review of the literature pertinent to this thesis by revisiting some basic results from Fourier theory. This will be useful for some of the work we will discuss later in this chapter, namely the Fourier domain properties of wavelets.

Let $f \in L^1(\mathbb{R})$ be an absolutely integrable function. The Fourier transform can be used to analyse f for its frequency properties.

Definition 1 *Let $f(x)$ be a function such that $\int_{-\infty}^{\infty} |f(x)|dx < \infty$, then its Fourier transform exists and is given by*

$$F(\omega) = \int_{-\infty}^{\infty} f(x)\ exp(-i\omega x)dx. \tag{2.1}$$

We can also invert the transform to recover the original function using the inverse Fourier transform:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega x)d\omega. \tag{2.2}$$

For *periodic* square integrable functions, i.e $f \in L^2([-\pi, \pi])$, there exists the related function representation, Fourier series. The function f can be represented in terms of the Fourier basis $\{\exp(inx)\}_{n=-\infty}^{\infty}$ as follows

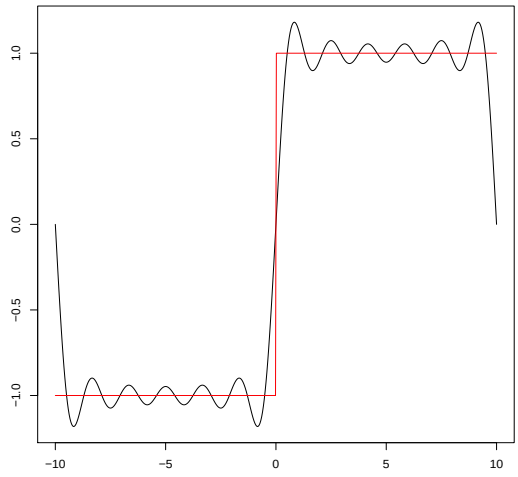
$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp(inx), \tag{2.3}$$

where the Fourier coefficients are given by

$$c_n = (2\pi)^{-1} \int_{-\pi}^{\pi} f(x) \exp(inx)dx. \tag{2.4}$$

Since $\exp(inx) = \cos(nx) + i\sin(nx)$, the Fourier series in equation (2.3) can be regarded as an expansion of f in terms of sine and cosine functions. In particular, the Fourier expansion is equivalent to using the basis function set $\{1, \cos(nx), \sin(nx)\}_{n\in\mathbb{Z}}$.

The main shortcoming of the Fourier series is that it cannot represent jump discontinuities efficiently (i.e. a large number of terms are required). Even then the representation can be poor and *Gibbs effects* can occur near the discontinuities. This is essentially because sines and cosines are smooth so cannot adapt to discontinuities in the signal. An example of this phenomenon can be seen in Figure 2.1, where the approximation has large oscillations near the jump.



(a)

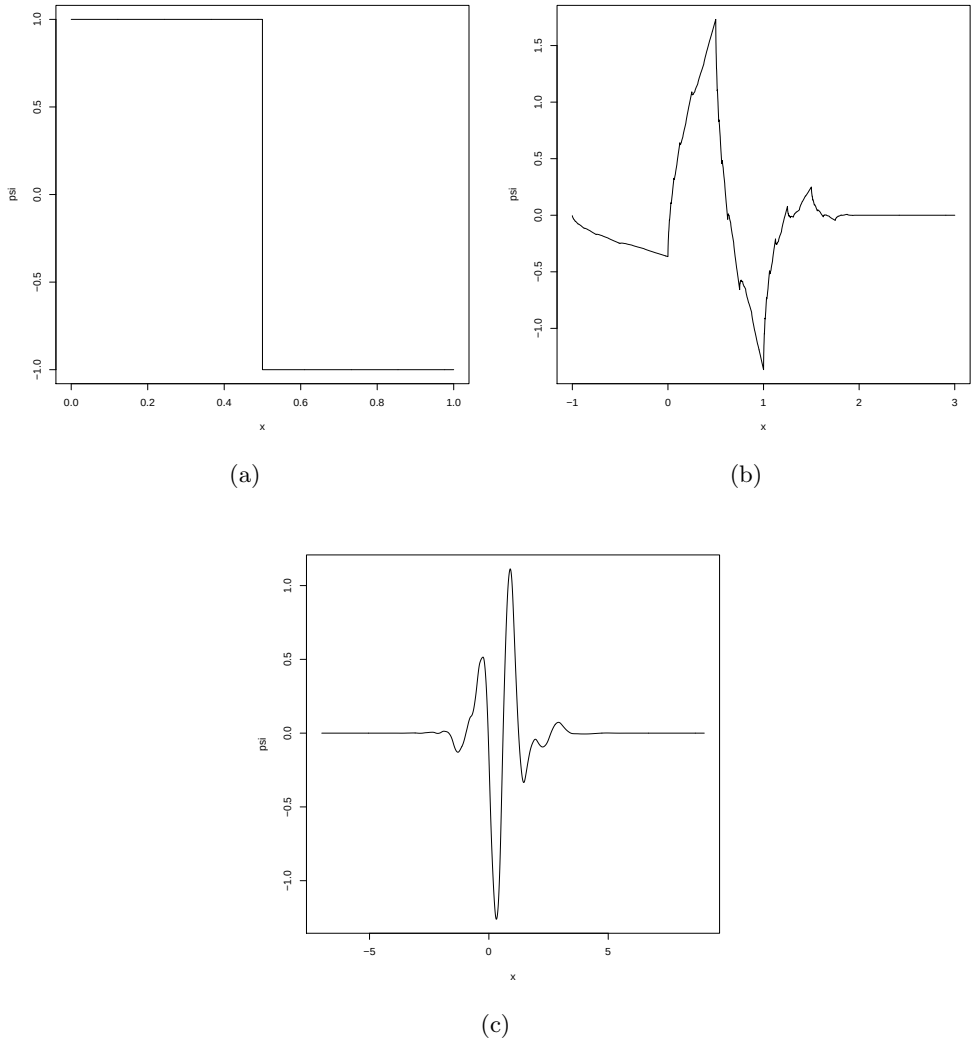
Figure 2.1: Example of the Gibbs effects which can be seen at the estimate of the discontinuity of the true function (shown in red).

A second drawback of the Fourier approach is it is not possible to obtain information on the *local* frequency behaviour of a signal. To obtain information about a particular period, we have to integrate over the whole domain of f. This is due to the fact that basis functions are localized in frequency but not in time. A more detailed discussion of Fourier series and their properties is contained within Walker (1986).

These drawbacks motivate a more efficient representation of a series for certain function types such as those with discontinuities or sharp spikes. In order to represent such local structure efficiently we require basis functions which have compact support in time. In the next section, we introduce wavelets, which possess these desired properties. Another possibility is the windowed Fourier transform, however there are certain constraints to consider such as the window size and the overlap between windows.

2.1.2 What is a wavelet?

Wavelets are functions that oscillate but decay quickly to zero. They are capable of both efficiently modelling frequency components for a range of signals, including those that contain discontinuities, and also capturing the smooth structure of a given signal or series. Unlike Fourier analysis which only has one basis shape (sinusoidal) to represent all functions, wavelets have many different forms. Examples of some wavelets can be seen in Figure 2.2, including the well-known Haar wavelet and Daubechies wavelets.



(a) (b)

(c)

Figure 2.2: Examples of Mother wavelets. (a) Haar Wavelet (Daub ExPhase N=1), (b) Daubechies Extremal Phase wavelet (N=2) and (c) Daubechies Least Asymmetric wavelet (N=7).

Daubechies (1988) proposes two commonly used wavelet families. These families are frequently referred to as "extremal phase" and "least asymmetric". Throughout this thesis we consider Haar and Daubechies wavelets, although a number of other wavelet families exist. See Daubechies (1992), Vidakovic (1999), Percival and Walden (2006) and Nason (2008) for an introduction to these wavelets.

Our starting point to analyse signals using wavelet theory is a basic function called the *mother wavelet*, ψ. A wavelet basis is constructed by translating and dilating the mother wavelet. Following Meyer (1992) we formally define a mother wavelet as follows:

Definition 2 *Let $N \in \mathbb{N}$. Then for $x \in \mathbb{R}$, a function $\psi(x)$ is called a* mother wavelet *of order N if the following properties hold:*

W1 *If $N = 0, \psi(x) \in L^{\infty}(\mathbb{R})$. If $N \geqslant 1$, then $\psi(x)$ and all its derivatives up to order N belong to $L^{\infty}(\mathbb{R})$.*

W2 *$\psi(x)$ and all its derivatives up to order N decrease rapidly as $x \to \pm\infty$.*

W3 *For each $k \in \{0, ..., N\}$,*

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0. \tag{2.5}$$

W4 *The collection $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$, the $\psi_{j,k}$ being constructed from the mother wavelet using the identity*

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k). \tag{2.6}$$

Here j relates to the dilation, known as the scale, whilst k relates to the translation, known as location.

The first three conditions describe smoothness and localisation of the wavelet ψ. W3 is often referred to as the *vanishing moments property*. This important property is used so wavelets can produce sparse representations of functions. In other words the wavelet coefficients (see equation 2.8) of any polynomial up to order N will be zero. A mother wavelet is said to be of order N if equation (2.5) holds true for $k = 0, 1, \ldots, N$. As Figure 2.2 shows as N increases

the wavelets are smoother in nature. Condition W4 generates orthonormal wavelet bases by dilation and translation operations, where the index j dilates the function by a factor of 2^j and k translates the wavelet along the axis by $2^{-j}k$ for all $j, k \in \mathbb{Z}$.

A function $f \in L^2(\mathbb{R})$ can be represented as a linear combination of wavelet bases using condition W4, i.e.

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(x). \tag{2.7}$$

Due to the orthogonality of wavelets, the wavelet (detail) coefficients are then given by

$$d_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx =< f, \psi_{j,k} > \quad \text{for all } j, k \in \mathbb{Z}, \tag{2.8}$$

where $< \cdot, \cdot >$ denotes the inner product. Note in particular that the function consists of localised scale information where j represents scale and k represents the location. Hence the detail $d_{j,k}$ coefficients tell us about the *local* oscillatory behaviour. In the next section we introduce multiresolution analysis (MRA), which plays an important role in the construction of wavelets and wavelet bases.

2.1.3 Multiresolution Analysis

Multiresolution analysis (MRA) was introduced by Meyer (1985) and Mallat (1989a). The aim of multiresolution analysis (MRA) is to examine a function $f \in L^2(\mathbb{R})$ at a particular resolution $j \in \mathbb{Z}$. This is achieved by a (linear) projection of f onto an approximation subspace $V_j \subset L^2(\mathbb{R})$. Following Mallat (1989a) we define a multiresolution analysis as follows:

Definition 3 *A multiresolution analysis (MRA) is a nested sequence of closed subspaces, $V_j \subset L^2(\mathbb{R})$ for $j \in \mathbb{Z}$,*

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots \tag{2.9}$$

such that the following conditions hold:

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

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}\,,\ \overline{\bigcup}_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R}). \tag{2.10}$$

2. *The following two-scale relation exists:*

$$f(x) \in V_j \iff f(2x) \in V_{j+1},\ \forall j \in \mathbb{Z}. \tag{2.11}$$

3. *The V-spaces are self similar in time:*

$$f(x) \in V_0 \iff f(x-k) \in V_0,\ \forall k \in \mathbb{Z}. \tag{2.12}$$

4. *And, finally, there exists a scaling function,* $\phi \in V_0$*, whose integer translations* $\{\phi_{0,k} : k \in \mathbb{Z}\}$ *constitute an orthonormal basis of* V_0*.*

Equation (2.9) shows that the spaces $\{V_j\}_{j\in\mathbb{Z}}$ form a ladder or hierarchy. As Nason (2008) explains this means each space of functions V_j contains detail up to some finest scale resolution. In other words MRA provides a framework for examining functions at different scales: zooming in to see the fine detail and moving out to obtain a broader view. We note that equations (2.11) and (2.12) imply that $\{\phi_{j,k} : k \in \mathbb{Z}\}$, constitutes an orthonormal basis of V_j. If $\{V_j\}_{j\in\mathbb{Z}}$ and $\phi \in V_0$ satisfy the above definition they are said to form a multiresolution approximation of $L^2(\mathbb{R})$.

With Definition 3 in place we introduce the concept of a multiresolution projection operator (Daubechies, 1988). Put simply, we denote by P_j the orthogonal projection operator which projects a function onto the space V_j described above. Since $\{\phi_{j,k}(x)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_j, the projection can be written as

$$f_j(x) = \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x) = P_j f. \tag{2.13}$$

The $\{c_{j,k}\}$ are denoted *father wavelet* coefficients and the orthogonality of the basis $\{\phi_{j,k}(x)\}$ means that they can be computed in a similar way to the details coefficients in equation

(2.8):

$$c_{j,k} = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)dx =< f, \phi_{j,k} > . \tag{2.14}$$

Since V_0 is a subspace of V_1, $\{\phi_{1,k}(x)\}$ is a orthonormal basis for V_1. Given $\phi(x) \in V_0$, we can write

$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k\phi_{1,k}(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\phi(2x-k), \tag{2.15}$$

where $h = \{h_k\}_{k\in\mathbb{Z}}$ is referred to as a low-pass (averaging) filter and is given by $h_k =< \phi, \phi_{1,k} >$. Equation (2.15) is called the dilation equation or scaling function equation. It controls how the scaling functions $\phi_j(x)$ relate to each other for two consecutive scales. Thus it is a very important result in wavelet theory as its solution enables one to begin building a general MRA.

As we shall see in Section 2.1.4 the coefficients $\{h_k\}_{k\in\mathbb{Z}}$ play an extremely important role in the construction of wavelets. Using the orthonormality of $\phi_{0,k}$, Daubechies (1988) identified some important properties of the wavelet filters, h, associated with an orthogonal MRA, namely:

1. $\sum_n \mid h_n \mid^2 = 1$;
2. $\sum_n h_n = \sqrt{2}$;
3. $\sum_n h_n h_{n-2j} = \delta_j, \forall j \neq 0$.

For proofs of these results refer to Vidakovic (1999).

As we move from one approximation space V_{j+1} to another V_j we lose some information. As Mallat (1989a) notes, the representation of this information loss is a key feature in wavelet transforms (see Section 2.2). We capture this lost information in the *difference space*, which we denote W_j.

Definition 4 *Let W_j be a subspace containing the 'detail' at level j. We have,*

$$V_{j+1} = V_j \oplus W_j. \tag{2.16}$$

In other words, W_j is the orthogonal complement of V_j in V_{j+1}. When the sequence of subspaces V_j satisfy the properties of a multiresolution analysis then there exists a wavelet

$\psi(x)$, where $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0 and orthogonal to all functions in V_j. This enables information at scale $j+1$ to be split into detail W_j and coarse V_j information at scale j. Moreover because of the MRA relations, each W_j is a scaled version of W_0 in that

$$f(x) \in W_0 \iff f(2^j x) \in W_j, \tag{2.17}$$

consequently $\psi_{j,k}$ is an orthonormal basis for W_j.

From the construction in Definition 4 we know that $W_j \subset V_{j+1}$ and therefore we have $\psi_{j,k}(x) \in V_{j+1}$. Hence $\psi_{j,k}(x)$ can be represented by a linear combination of the basis functions for V_{j+1}. We can consequently construct a wavelet function from a given scaling function as follows:

$$\psi(x) = \sum_{k\in\mathbb{Z}} g_k \phi_{1,k}(x), \tag{2.18}$$

where $\mathbf{g} = \{g_k\}_{k\in\mathbb{Z}}$ is the high-pass filter associated to ψ.

2.1.4 Fourier properties of wavelets

Having introduced the wavelet function $\psi(x)$ and scaling function $\phi(x)$, we now introduce Fourier domain representations of these functions. These are useful since they allow us to further explore the properties of multiresolution analysis.

Following Vidakovic (1999), the Fourier transform of the scaling equation $\phi(x)$ as defined in equation (2.15) is given by

$$\hat{\Phi}(\omega) = m_0(\omega/2)\hat{\Phi}(\omega/2), \tag{2.19}$$

where $m_0(\omega) = 2^{-1/2}\sum_k h_k \exp(-i\omega k)$. Similarly the Fourier transform of the wavelet function given in equation (2.18) is given by,

$$\hat{\Psi}(\omega) = m_1(\omega/2)\hat{\Phi}(\omega/2), \tag{2.20}$$

where $m_1(\omega) = 2^{-1/2}\sum_k g_k \exp(-i\omega k)$.

Daubechies (1992, Chapter 5) shows that the orthonormality of the scaling functions leads to the condition that

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 1. \tag{2.21}$$

From Mallat (1989a) we also have that $|m_0(0)| = 1$ and therefore from equation (2.21), $|m_0(\pi)| = 0$. It can also be shown that $|m_1(\omega)| = |m_0(\omega + \pi)|$ and hence equation (2.21) can be written as,

$$|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1. \tag{2.22}$$

See Vidakovic (1999, pg 58–59) for further details of the derivation.

Using equation (2.22) and the representation of $m_1(\omega)$, the coefficients $\mathbf{g}$ and $\mathbf{h}$ can be related as follows:

$$g_n = (-1)^n h_{1-n}. \tag{2.23}$$

This relationship between $\mathbf{g}$ and $\mathbf{h}$ is referred to as the **quadrature mirror filter** relation and means that we have the mutual orthogonality relation $\sum_n h_n g_{n+2j} = 0, \forall j \in \mathbb{Z}$. The quadrature mirror relation in equation (2.23) gives us the ability to take any scaling function $\phi_{j,k}(x)$ which satisfies the MRA properties from Definition 3 and use it to derive a wavelet function using equation (2.18).

Having established representations for elements of both V_j and W_j we can now rewrite equation (2.13) as a wavelet representation of a function $f(x)$. Specifically given a wavelet basis, the representation of a function, f, is given by the sum of the detail coefficients and an approximation at the coarser level. We have:

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x), \tag{2.24}$$

where the coefficients $c_{j_0,k}$ and $d_{j,k}$ can be found efficiently using Mallat's cascade algorithm which we introduce in the next section.

2.2 Wavelet transforms for discrete data

In data analysis, instead of the continuous setting described above in equation (2.24), we typically need to make inference on a set of observations recorded at a finite number of discrete time points. We now consider various wavelet transforms for discrete data starting with the discrete wavelet transform, discussing practical issues such as different basis choices and boundary conditions.

2.2.1 Discrete wavelet transform (DWT)

The discrete wavelet transform (DWT), proposed by Mallat (1989a,b), provides an efficient method for performing a wavelet-based transformation of discrete data. It consists of splitting a sequence of data of length 2^J for some $J \in \mathbb{N}$ into several "smooth" $\{c_{j,k}\}_{j,k\in\mathbb{Z}}$ and "detail" $\{d_{j,k}\}_{j,k\in\mathbb{Z}}$ series. Prior to describing the algorithm we first introduce some basic theory.

Recall from the previous section that by a subsitution of indices in equations (2.15) and (2.18) we obtain:

$$\phi_{j-1,k}(x) = \sum_{l\in\mathbb{Z}} h_{l-2k}\phi_{j,k}(x) \tag{2.25}$$

$$\psi_{j-1,k}(x) = \sum_{l\in\mathbb{Z}} g_{l-2k}\phi_{j,k}(x). \tag{2.26}$$

We assume there exist functions $v_j \in V_j$ and $w_j \in W_j$, such that,

$$v_j(x) = \sum_k c_{j,k}\phi_{j,k}(x) \tag{2.27}$$

$$w_j(x) = \sum_k d_{j,k}\psi_{j,k}(x), \tag{2.28}$$

where the $\{c_{j,k}\}$ are the smooth coefficients and the detail coefficients $\{d_{j,k}\}$ provide information about a local variation on a given scale of the original series. These are the key theoretical properties used in the discrete wavelet transform algorithm.

Consider the father wavelet, $\phi(x)$. If this satisfies the attributes of MRA, then the above properties can be used to efficiently calculate $c_{j-1,k}$ and $d_{j-1,k}$ from $c_{j,k}$, where $c_{j-1,k}$ and

$d_{j-1,k}$ represent the coefficients at the next coarsest scale. More precisely,

$$\begin{aligned} c_{j-1,l} &= < v_j, \phi_{j-1,l} > \\ &= < v_j, \sum_k h_{k-2l}\phi_{j,k} > \text{ by (2.25)} \\ &= \sum_k h_{k-2l} < v_j, \phi_{j,k} > \text{ by additivity of inner products} \\ &= \sum_k h_{k-2l} c_{j,k} \text{ by (2.27).} \end{aligned} \tag{2.29}$$

Similarly,

$$d_{j-1,l} = \sum_k g_{k-2l} c_{j,k}. \tag{2.30}$$

Now we have provided the constituent parts to the DWT we consider its implementation.

Performing the DWT: Consider a sequence c_J of length 2^J where c_J is the original data. This is filtered using the equations (2.29) and (2.30) where $\mathbf{h}$ and $\mathbf{g}$ are the filters from equations (2.15) and (2.18) respectively. A set of both smooth ($c_{J-1,k}$) and detail ($d_{J-1,k}$) coefficients are obtained for the finest scale by this filtering process. The set of 'smooth' coefficients are then filtered again using the equations (2.29) and (2.30) and coefficients at the next coarsest scale are obtained. This process is continually repeated until the coarsest scale is reached. Note that due to the filtering in equations (2.29) and (2.30) at each level only even filtered coefficients are retained, so there are half as many coefficients as at the previous scale. The final set of coefficients are then given by $(\mathbf{c}_0, \mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{J-1})$ where $\mathbf{c}_0$ is the coarsest father coefficient and the $\mathbf{d}_j = \{d_{j,k}\}$ represent the detail coefficients. This scheme is commonly referred to as Mallat's pyramid algorithm. Figure 2.3 shows a schematic diagram of the algorithm.

Inverse DWT: Due to the orthogonality of the $\psi_{j,k}$, the DWT algorithm is invertible. Mallat (1989b) shows that the inversion is given by

$$c_{j,k} = \sum_l h_{k-2l} c_{j-1,l} + \sum_l g_{k-2l} d_{j-1,l} \tag{2.31}$$

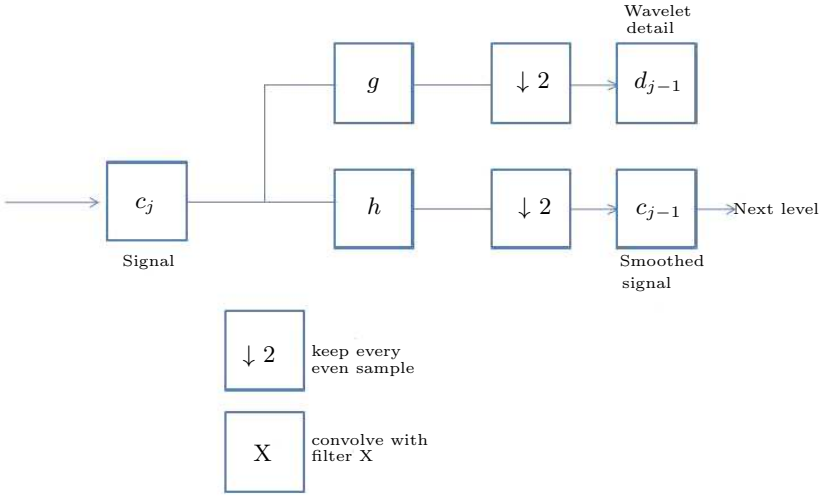


Figure 2.3: The discrete wavelet transform of a data sequence into smooth and detail components using quadrature mirror filters, **h** and **g**.

where **h** and **g** are the filters as defined previously. Specifically, from the detail and smooth coefficients at the current scale, $j-1$, it is possible to obtain the smooth coefficients at level j. The original sequence can then be obtained from repeated applications of this inversion step. For further details of the DWT please refer to Vidakovic (1999), Percival and Walden (2006) and Nason (2008).

2.2.2 An operator description of the DWT

An alternative way of describing the DWT is to consider the transform in terms of operators. Following Nason and Silverman (1995), let $\mathcal{H}$ and $\mathcal{G}$ represent the low and high-pass convolution operators such that,

$$(\mathcal{H}x)_k = \sum_n h_{n-k}x_n \tag{2.32}$$

$$(\mathcal{G}x)_k = \sum_n g_{n-k}x_n \tag{2.33}$$

where $\{x_n\}_{n\in\mathbb{Z}}$ is some series. As discussed previously we subsample (or dyadically decimate) the data at each step to maintain orthogonality of the transform. This step can be

represented in operator notation by the dyadic decimation operator, $\mathcal{D}_0$, given by,

$$(\mathcal{D}_0 x)_l = x_{2l}, \ \forall j \in \mathbb{Z}, \tag{2.34}$$

i.e. $\mathcal{D}_0$ picks out the even indexed elements.

With the above notation established we can perform a DWT on a data series $\{c_{J,k}\}_{k=0,\ldots,2^{J-1}}$ of length 2^J using the equations (2.29) and (2.30), which can also be written as,

$$c_{j-1} = \mathcal{D}_0 \mathcal{H} c_j \tag{2.35}$$

$$d_{j-1} = \mathcal{D}_0 \mathcal{G} c_j, \tag{2.36}$$

assuming the original data is defined to be c_J and $j = J, \ldots, 1$.

We can represent the inverse discrete wavelet transform in this notation as follows:

$$\mathbf{c}_j = \mathcal{H}\mathcal{D}_0^{-1} c_{j-1} + \mathcal{G}\mathcal{D}_0^{-1} d_{j-1} \tag{2.37}$$

where we let the inverse of $\mathcal{D}_0$ be denoted by $\mathcal{D}_0^{-1}$ where this operator has the effect of inputting a zero between each sample.

From Nason and Silverman (1995) the whole set of DWT detail and father wavelet coefficients at level j (in terms of the original data) can be expressed as

$$d_j = \mathcal{D}_0 \mathcal{G} (\mathcal{D}_0 \mathcal{H})^{J-j-1} c_J \tag{2.38}$$

$$c_j = (\mathcal{D}_0 \mathcal{H})^{J-j} c_J \tag{2.39}$$

for $j = 0, \ldots, J-1$.

2.2.3 $\epsilon -$ decimated wavelet transform

The established convention is to use D_0 and perform an *even* decimation in the DWT. However as Nason and Silverman (1995) suggest, replacing $\mathcal{D}_0$ by $\mathcal{D}_1$ in the wavelet transform where $(\mathcal{D}_1 x)_k = x_{2k+1}$ (i.e. choose every odd member of a sequence) is merely a selection of a different orthogonal basis. In other words, it doesn't matter whether we use an odd

or even decimation we still end up with an orthogonal transform. Furthermore, we could switch between odd and even decimations at each stage. So a particular orthogonal basis can be represented by the J- digit binary number $\boldsymbol{\epsilon} = \epsilon_{J-1}\epsilon_{J-2}\ldots\epsilon_0$ where ϵ_j is one if $\mathcal{D}_1$ is used to produce level j and zero if $\mathcal{D}_0$ is used. Such a transform is termed the $\epsilon-$ *decimated wavelet transform* by Nason and Silverman (1995). As the orthogonality of the transform is retained we can perform the inverse transform of the decomposition in order to obtain the original signal. The choice of odd or even decimation is important for the understanding of the non-decimated wavelet transform discussed in Section 2.3.1.

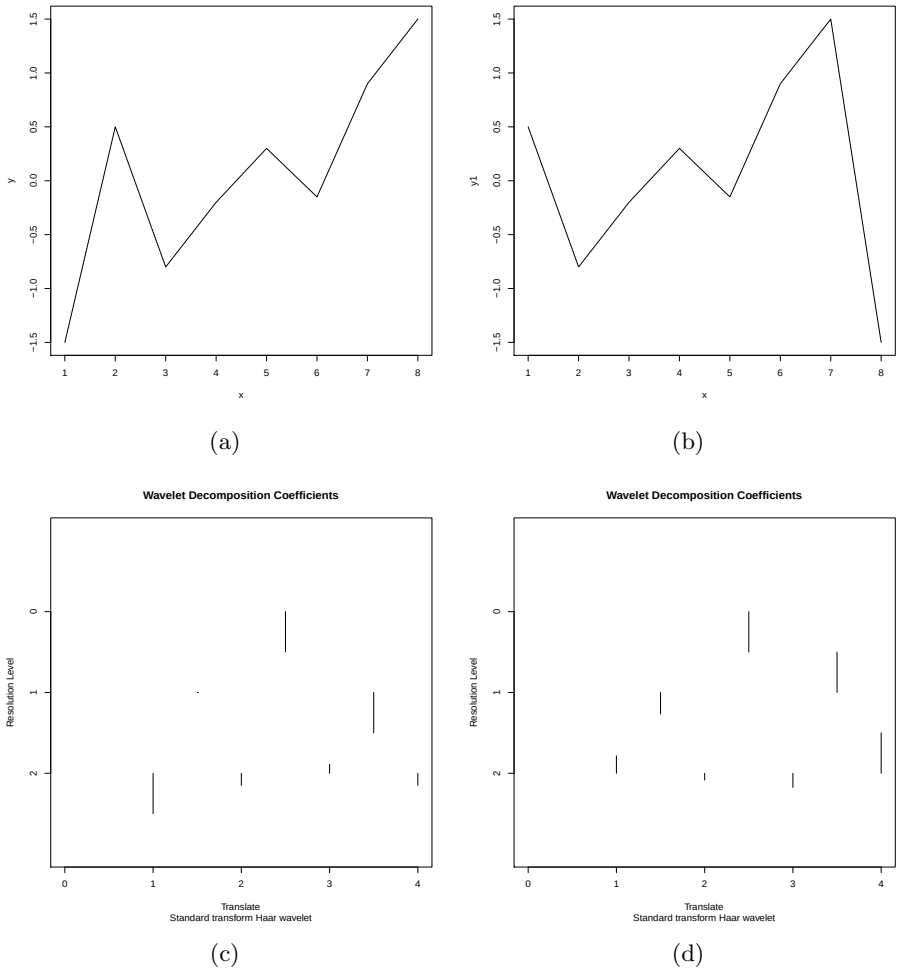
2.2.4 Boundary conditions

When analysing data using equations (2.29) and (2.30) the filters can sometimes extend beyond the range of the data, creating boundary issues. A fortunate feature of Haar wavelets is that computing coefficients near 'boundaries' is straight forward as no boundary conditions are needed. The Haar transform filters the dyadic series in pairs to produce another dyadic series, which can then be processed in the same way. For more general Daubechies wavelets, boundaries have to be treated with much more care as the filters are not dyadic in length, i.e. 2^J.

As Nason (2008) describes, an obvious way to practically resolve this potential boundary problem is to artificially extend the boundary in some way and then convolve the filter with the extended data in the usual manner. Two types of boundary extension are commonly used in practice: *periodic* and *symmetric*. With a periodic boundary condition the sequence wraps around itself, so after the last term in the sequence we begin the sequence again. For a symmetric case we reflect the sequence in its boundary so after the last term of the sequence, we begin the sequence again in reverse. Another possibility is to modify the wavelet so that it always remains on the domain of the original data. Such a modification is the basis of the procedure known as 'wavelets on the interval' proposed by Cohen et al. (1993). Wavelet coefficients are produced at coarser scales but no information is borrowed from periodisation or reflection. The lifting scheme, which we review in Section 2.3.4, can also be used to overcome this practical issue. In this thesis we focus on Haar wavelets so there are no boundary conditions to consider.

2.2.5 Weakness of the DWT

We now turn to consider one of the main weaknesses of the DWT: its lack of translation invariance. One small shift in the data can change the wavelet coefficients completely. To explore this further we consider an example.



(a) (b)

(c) (d)

Figure 2.4: Example of the DWT and the DWT on shifted data. (a) shows the original data with (c) being the corresponding DWT; (b) shows the data shifted along by one with (d) being the corresponding DWT.

Figure 2.4 displays a test data set $\boldsymbol{x}$ along with the corresponding wavelet coefficients obtained from the DWT. It also shows a shifted version of $\boldsymbol{x}$ and applying the DWT to this series. In this case the data has been rotated by one, with the first observation now being the last. The coefficients $d_{j,k}$ are plotted with the finest-scale coefficients at the bottom of the plot and coarsest at the top. In this case we have 8 points in each series and it can be

seen that at each stage of the transform the number of coefficients halves. The magnitude of the coefficients is represented by the length of the vertical mark.

In Figure 2.4 we can clearly see how the sharp spike causes larger wavelet coefficients. This example highlights how the discrete wavelet transform is not translation invariant and shifting the data has changed the wavelet coefficients completely. This example was performed using the R package `wavethresh` (Nason, 2013). It would be desirable that a shift in the data results in a shift of the coefficients otherwise our interpretation of the data alters radically. In the next section we introduce the non-decimated wavelet transform which overcomes this.

2.3 Alternative wavelet transforms

The DWT is only one possible (discrete) wavelet transform in the literature. In this section we briefly review other transforms, including the non-decimated wavelet transform highlighting its translation invariance properties. We then consider the wavelet packet transform, complex wavelet transform and lifting schemes which are included for completeness although will not be used in the remainder of the thesis.

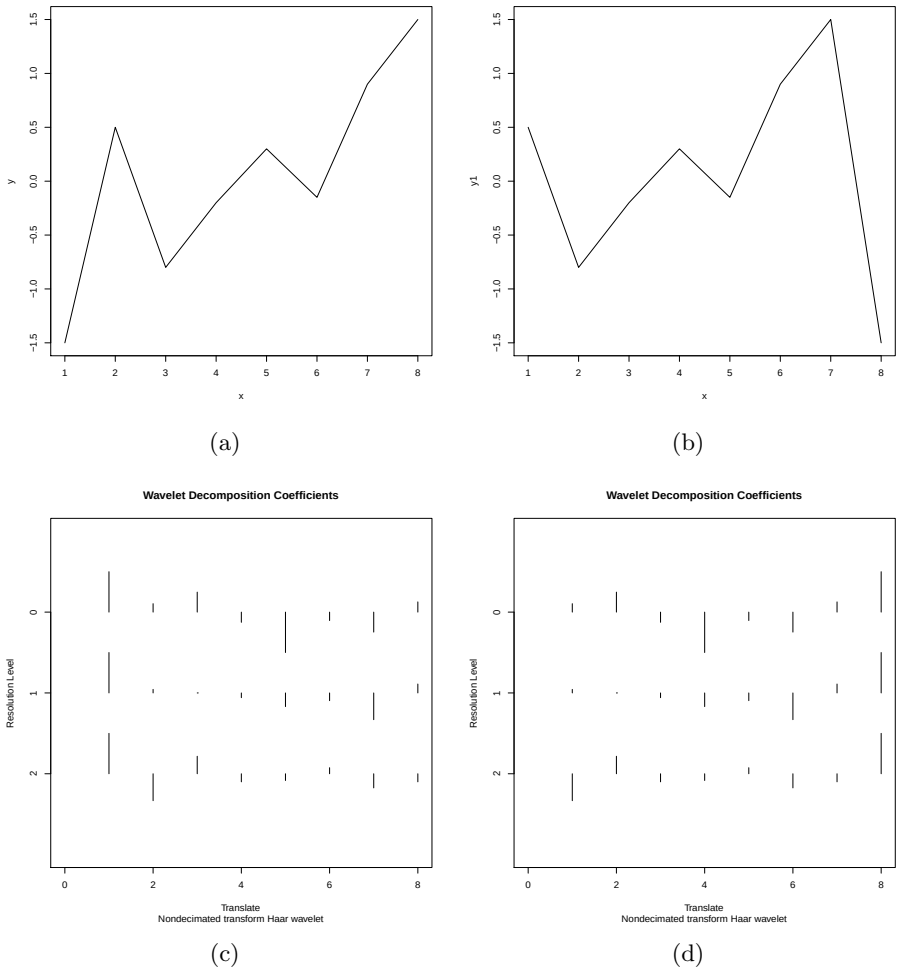
2.3.1 Non-decimated wavelet transform (NDWT)

We now introduce the non-decimated wavelet transform (NDWT). This transform was proposed by Nason and Silverman (1995) to overcome the translation invariance issues of the DWT highlighted in Section 2.2.5.

The basic form of the NDWT is to apply high and low pass filters, $\mathbf{g}$ and $\mathbf{h}$ respectively, to the data at each level to produce two series at the next level. The two new series each have the same length as the original series as no decimation step occurs. Instead, the filters are modified at each level by padding them out with an appropriate number of zeros, in other words adding zeros between each element of the filter.

An alternative view of the NDWT is to perform the $\epsilon - \textit{decimated wavelet transform}$ with every choice of ϵ as defined in Section 2.2.3 (i.e. every combination of $\mathcal{D}_0$ and $\mathcal{D}_1$). This is equivalent to the set of DWT coefficients from all shifts of the data.

Whilst the NDWT provides more information than the DWT at any given scale, using both odd and even decimated coefficents destroys the orthogonal structure. However, one of the distinct advantages of the NDWT, is its translation equivariance, i.e. a systematic rotation in the data results in a shift in the coefficients. As an example of this, consider Figure 2.5, which plots the NDWT of the original and shifted data used for the example in Section 2.2.5. Unlike the results for the DWT in Figure 2.4, here we have the same coefficients obtained for (a) and (c), but (c) has been shifted along one as in the data itself.



(a) (b)

(c) (d)

Figure 2.5: Example of the NDWT and the NDWT on shifted data. (a) shows the original data with (c) being the corresponding NDWT, (b) shows the data shifted along by one with (d) being the corresponding NDWT.

A related transform to the NDWT is the Maximal Overlap Discrete Wavelet Transform

(MODWT). This provides the same advantages as the NDWT but allows a time series to be of any length. See Percival and Walden (2006) for a full description of the MODWT.

2.3.2 Wavelet packet transforms

One particular aspect of the DWT and NDWT is that the features focus on decomposing the low frequency structure in a sequence. In other words the DWT and NDWT only carry the 'smooth' coefficients c forward to the next level. One might equivalently focus on the high frequency content. This is the main idea behind the wavelet packet transform (WPT). Both the 'smooth' c and 'detail' d coefficients are carried forward to the next level and then filtered to produce coarser level coefficients. In this way the wavelet packet transform analyses a series using a diverse collection of wavelet packets (frequency objects) at different scales and locations. If a complete decomposition of smooths and details is used then this leads to a redundant transform, however an orthogonal basis can be obtained by careful packet selection. See Coifman and Wickerhauser (1992) for an introduction to best basis selection for choosing packets to form an orthogonal transform. For a more detailed treatment of this transform, see Vidakovic (1999).

2.3.3 Complex wavelets

Although the DWT provides us with a computationally efficient and sparse representation of real-world signals, it suffers from some fundamental disadvantages. In particular, these include shift sensitivity (see Section 2.2.5), poor directionality (in a two-dimensional setting) and lack of phase information. However, as Kingsbury (1999) note, the Fourier transform does not suffer from the above shortcomings as it is based on complex-valued oscillating sinusoids. It has therefore been suggested that the above problems can be solved by effectively computing a complex wavelet transform ($\mathbb{C}$WT) (Kingsbury, 1999).

The initial versions of complex wavelet transforms were introduced by Lawton (1993) and Lina (1997). Lina and Mayrand (1995) and Barber and Nason (2004) also discuss complex valued wavelets. The structure of the $\mathbb{C}$WT is similar to that of the DWT, except that the $\mathbb{C}$WT filters have complex coefficients and generate complex output samples. However this extension of the DWT to a complex setting does not allow for perfect reconstruction of signals and exhibits 2^d redundancy compared to the DWT, where d is the dimension of the signal being transformed. This motivated the work of Kingsbury et al. (2005) who developed the m-dimensional dual tree discrete wavelet transform (DT-DWT), which allows for perfect reconstruction as well as mitgating the disadvantages highlighted previously. However the design of the quadrature wavelet filters pairs can be quite complicated. See Selesnick (2001, 2002); Kingsbury et al. (2005) for further details on complex wavelets and the dual tree complex wavelet transform. Complex wavelet transforms have been used in various areas of statistics, see for example Whitcher and Craigmile (2004) and Whitcher et al. (2005).

2.3.4 The lifting scheme

Lifting schemes were introduced by Sweldens (1996, 1997) to provide a way of decomposing signals with data arising from non-standard situations such as signals observed on an irregular grid or with missing data.

With the lifting scheme we can build wavelet-like basis functions $\psi_{j,k}(x)$ with desirable properties. These properties, such as smoothness, can be chosen to be dependent on the signal under analysis. Possible applications of lifting schemes in statistics include non-

parametric regression, using adaptive lifting (Nunes et al., 2006; Knight and Nason, 2009) and time series analysis (Knight et al., 2012).

The lifting scheme can be broken down into 3 steps. Let $\{x_i\}$ be the points at which the data $\{f_i\}$ is observed. Initially, we partition the range of $\{x_i\}$ into intervals associated to each x_i. The scheme then proceeds:

1. Split: Split the data $\{f_i\}$ into two subsets f_I and f_J.

2. Predict: Use f_J to predict f_I (using e.g. linear regression). The lifted coefficients are the regression residuals: $d_I = f_I - \hat{f}_I$, in other words the difference between function values and predictions. The coefficients d_I represent local signal changes, in a similar manner to detail coefficients produced from traditional wavelet transforms.

3. Update: Remove x_I and redistribute lost signal content to the values f_J using the lifted coefficients d_I. Repeat 1-3 until one gridpoint x_i is left.

This flexibility to handle non-standard data comes at a cost: lifting schemes are non-orthogonal which complicates issues such as thresholding (see Section 2.4.1) and inference in the wavelet domain since norms of coefficient sequences are not preserved during the transform. In other words it is difficult to carry out efficient analysis of variance and covariance properties in the wavelet domain. See Jansen and Oonincx (2005) for more details on lifting transforms.

2.4 Wavelets in Statistics

Whilst the focus of this thesis is on modelling the locally stationary covariance structure of 2D (image) data, there has been considerable work on using wavelets more generally in statistics during the last 20 years. This section provides a brief introduction to this area, for further more comprehensive reviews we refer the reader to Vidakovic (1999), Abramovich et al. (2000), Percival and Walden (2006) or Nason (2008).

2.4.1 Wavelet Smoothing

One of the original applications of wavelets in statistics was in signal denoising, commonly known as smoothing or non-parametric regression. In this setting we have a signal corrupted by noise and we wish to remove the noise from the data. More formally we consider the setting where the aim is to estimate the unknown vector $\mathbf{g} = (g(x_1), \ldots, g(x_n))$ using the data $\{y_i\}_{i=1,\ldots,n}$ where $\{y_i\}$ is a collection of observed data points. We model the data as,

$$y_i = g(x_i) + e_i, \tag{2.40}$$

where $e_i \sim N(0, \sigma^2)$ is independent noise. The general wavelet method for estimating $\mathbf{g}$ is to, (i) apply a wavelet transform to the data, (ii) modify the wavelet coefficients in some way to reduce the noise and then (iii) apply the inverse of the transform to the new coefficients to obtain an estimate of $\mathbf{g}$ (Donoho and Johnstone, 1994).

The motivation for this procedure is as follows. Since the DWT is orthogonal, the wavelet coefficients of $\{y_i\}$ can be written as:

$$d^* = d + \epsilon. \tag{2.41}$$

where d^* denotes the wavelet coefficients from the noisy $\{y_i\}$, d represents the true (uncorrupted) wavelet coefficients and the ϵ are independent Gaussian noise obtained from the DWT of e_i. They all are a vectors of length n. By performing the DWT, the detail coefficient sequence is sparse due to the orthogonality property of the DWT. Thus to obtain an estimate of d a shrinkage rule is applied, where coefficients d^* that are smaller than some threshold are removed as they are assumed to be insignificant signal information (i.e. noise).

Two of the most commonly used shrinkage rules were proposed by Donoho and Johnstone (1994): hard and soft thresholding. These are defined as follows:

- Hard - The coefficients are compared to a threshold(s). If the coefficient is smaller in magnitude than the threshold it is removed, otherwise it is kept unaltered. i.e.

$$\hat{d} = \eta_H(d^*, \lambda) = d^* \mathbb{I}\{|d^*| > \lambda\}. \tag{2.42}$$

- Soft - All coefficients are modified by the formula below. In other words all coefficients greater than the threshold are shrunk towards zero. i.e.

$$\hat{d} = \eta_S\left(d^*, \lambda\right) = sgn\left(d^*\right)\left(|d^*| - \lambda\right)\mathbb{I}\left\{|d^*| > \lambda\right\}. \tag{2.43}$$

In equations (2.42) and (2.43), $\mathbb{I}$ is the indicator function, λ is the threshold value to be chosen and d^* are the empirical coefficients (i.e. noise corrupted) to be thresholded. Naturally a key question for these thresholding methods is, how best to choose such thresholds. Assuming Gaussian noise Donoho and Johnstone (1994) introduced the universal threshold which can be estimated from the data and achieves almost ideal risk. It can be expressed as

$$\lambda^u = \sigma\sqrt{2\log n}, \tag{2.44}$$

where n is the number of data points and σ is the noise level. In real problems we estimate σ by $\hat{\sigma}$, some estimate of the common standard deviation of the noise ϵ_i. Donoho and Johnstone (1994) suggest estimating σ by computing the median absolute deviation (MAD) of the finest-scale wavelet coefficients.

A number of other threshold choices exist. Donoho and Johnstone (1995) introduced the SureShrink approach. This is based on the minimization of Stein's unbiased risk estimator (SURE). It uses both the universal and SURE thresholds and selection of the threshold is dependent on the sparsity of the wavelet representation at a given level. Specifically if the wavelet representation is sparse, then the universal threshold is selected, otherwise the level-dependent SURE threshold is used.

Johnstone and Silverman (2004) and Johnstone and Silverman (2005) proposed another threshold, based on an empirical Bayes approach to modelling wavelet coefficients. This method places a prior distribution on the true wavelet coefficients from which, given the observed coefficients, the posterior distribution may be determined. The suggested prior takes the form

$$d^*_{j,k} \sim (1 - \pi)\delta_0 + \pi\gamma, \tag{2.45}$$

where γ is the density of the wavelet coefficient, conditional on it being non-zero. Hence π

represents the (prior) probability of the wavelet coefficient being non-zero. Many choices for γ have been proposed in the literature, see Chipman et al. (1997), Abramovich et al. (1998), Clyde et al. (1998) and Reményi and Vidakovic (2013). There are many other threshold policies, an overview of which can be found in Vidakovic (1999, Chapter 6).

2.4.2 Locally stationary wavelet processes in time series

We now turn to consider the application of wavelets to modelling locally stationary time series. To this end we assume a time series $\{X_t\}_{t\in\mathbb{Z}}$ is observed at evenly spaced time points with a unit sampling interval. However, prior to this we provide a brief reminder of established time series theory.

Traditional approaches to time series analysis assume (second-order) stationarity, i.e. that the autocovariance properties remain constant over time. Such series can also have a spectral representation. In particular, Priestley (1965) defines the second-order stationary stochastic processes X_t for $t \in \mathbb{Z}$ as

$$X_t = \int_{-\pi}^{\pi} A(\omega) exp(i\omega t) d\zeta(\omega), \tag{2.46}$$

where $d\zeta(\omega)$ is an orthonormal increment process. This model can be used in the estimation of the Fourier spectrum, which provides a frequency decomposition of the process variance. See Priestley (1981), Brillinger (2001) and Chatfield (2004) for more details on spectral analysis.

Most time series encountered in practice have an autocovariance structure which changes over time, i.e. they exhibit some time-varying second-order structure. Hence the focus of recent research has been on modelling non-stationary time series, specifically time series that can be considered locally stationary. The locally stationary approach is one method to dealing with non-stationarity. Locally stationary time series have a second-order structure that is approximately stationary around a local point but which is non-stationary across the entire series. For an overview of locally stationary processes, see Dahlhaus (2012).

A locally stationary extension to the stationary process representation considered in

equation (2.46) was introduced by Dahlhaus (1997). This takes the form:

$$X_{t,T} = \int_{\pi}^{\pi} exp(i\omega t) A_{t,T}(\omega) d\xi(\omega), \tag{2.47}$$

where $A_{t,T}(\omega)$ is a transfer function which varies smoothly over time, allowing a locally stationary representation. An alternative, proposed by Nason et al. (2000), is to concentrate on processes whose second-order structure changes slowly over time by replacing the set of harmonics $\{\exp(i\omega t)|\omega \in [-\pi, \pi]\}$ in equation (2.47) by a set of *discrete non-decimated wavelets*. In particular they introduce a novel process representation, given in Definition 5 below, which permits a location-scale decomposition of the covariance structure of a time series.

Definition 5 *A locally stationary wavelet process (LSW) is a stochastic process* $\{X_{t;T}\}_{t=1,\ldots,T}$ *for* $T = 2^J$ *represented as,*

$$X_{t;T} = \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k;T} \psi_{j,k}(t) \xi_{j,k}, \tag{2.48}$$

where $\xi_{j,k}$ *is a random orthonormal increment sequence,* $\{\psi_{j,k}(t)\}$ *is a discrete non-decimated family of wavelets based on the mother wavelet* ψ *and* $w_{j,k;T}$ *is a set of amplitudes for* $j = 1, 2, \ldots, k = 0, \ldots, n-1$.

In order to estimate the covariance structure of an LSW process, there are various assumptions which must be satisfied, these can be found in Nason et al. (2000). Briefly they assume each LSW process has zero mean and the amplitudes $w_{j,k;T}$ are slowly varying. Full details of the assumptions can be found in Chapter 3 in a two-dimensional setting.

The LSW model permits a time series to be stationary within localized regions, although its structure may evolve from one region to another. Various estimates of statistical properties may be obtained by collecting sufficient information in a locally stationary region. Nason et al. (2000) define the *evolutionary wavelet spectrum* (EWS) which quantifies how the power of a process varies locally over time and scale. Furthermore they show how this quantity can be estimated by a smoothed mean-corrected wavelet periodogram which can, in turn, be used to provide an estimable time-localized autocovariance. The process

of estimating the wavelet spectrum will be explained in more detail in Chapter 3 in the two-dimensional setting, which is the focus of this thesis.

In recent years the LSW framework has been applied in a number of different areas. For example, Fryźlewicz et al. (2003) use the framework to forecast locally stationary time series whilst Fryźlewicz and Ombao (2009) consider classification of time series using LSW processes. Cho and Fryźlewicz (2012) and Killick et al. (2013) use the approach to detect changepoints within piecewise second-order stationary time series and Knight et al. (2012) consider spectral estimation of non-stationary time series with missing observations and provide an extension to the LSW model.

In other work Nason and Cardinali (2010) introduce a method using the LSW model where given two (or more) locally stationary time series as defined in Definition 5, a time-varying, linear combination of them that are stationary (costationary series) can be obtained. Chapter 3 develops methodology for testing for stationarity within 2D data based on theory in Nason and Cardinali (2010), which we extend to 2D. In addition Nason and Cardinali (2010) develop a measure for the cross-periodogram and cross-covariance of locally stationary wavelet processes using similar methods to that in Nason et al. (2000). In Chapter 5 we consider an extension of these quantities to a multivariate two-dimensional setting.

Finally, Eckley et al. (2010) extend the LSW modelling framework to two dimensions. This extension is discussed in detail in Chapter 3 since it forms the basis of the test of stationarity we develop.

2.5 Extension to two dimensions

Recall that the focus of this thesis is the analysis of spatial data, specifically data on a regular lattice such as an image. For this we need to consider two-dimensional wavelets and their transforms. Several different approaches to the two-dimensional transform exist. In this thesis we focus our attention on the widely adopted separable multiresolution approximations of $L^2\left(\mathbb{R}^2\right)$ considered by Daubechies (1988) and Mallat (1989b). However other non-separable approaches exist, see for example Daubechies (1992). Whilst 2D non-separable wavelet approaches are more flexible, for the application under consideration, developing a rigorous

estimation scheme would prove challenging as extensions to basic wavelet theory would be required. We begin this section by extending multiresolution analysis to a 2D setting before giving a detailed overview of the separable 2D wavelet transform in Section 2.5.2.

2.5.1 Multiresolution Analysis in two dimensions

Recall that the multiresolution analysis framework described previously in Section 2.1.3 is suitable for functions on the real line. We now consider a two-dimensional setting by considering the wavelet $\psi \in L^2(\mathbb{R}^2)$. In this case, as Mallat (1989b) demonstrated, we construct a multidimensional MRA by using the tensor product of one-dimensional multiresolution analyses. Essentially through the properties of tensor products, the two-dimensional space inherits the orthogonal structure of the original MRA. Following Vidakovic (1999) we formally define a two-dimensional multiresolution analysis as follows:

Definition 6 *A two-dimensional multiresolution analysis (MRA) can be defined as a sequence of subspaces,* $\mathbf{V}_j \subset L^2(\mathbb{R}^2)$ *for* $j \in \mathbb{Z}$,

$$\cdots \subset \mathbf{V}_{-2} \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \ldots \tag{2.49}$$

where $\mathbf{V}_j = V_{j,(1)} \otimes V_{j,(2)}$ *and each* V_j *represents a one-dimensional subspace. These subspaces satisfy* $\bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{\mathbf{0}\}$ *and* $\overline{\bigcup}_{j\in\mathbb{Z}} \mathbf{V}_j = L^2(\mathbb{R}^2)$.

As in the one dimensional setting we can associate a scaling function with this two-dimensional multiresolution analysis. Specifically we have,

$$\phi(x, y) = \phi(x)\phi(y), \tag{2.50}$$

where the $\phi(x)$ and $\phi(y)$ correspond to one dimensional scaling functions. Dilated and translated versions of this scaling function are obtained, as follows:

$$\begin{aligned} \phi_{j,\mathbf{k}}(\mathbf{x}) &= \phi_{j,k_1}(x)\phi_{j,k_2}(y), & (2.51)\\ &= 2^j \phi(2^j x - k_1)\phi(2^j y - k_2), & (2.52) \end{aligned}$$

where $\mathbf{k} = (k_1, k_2)$ with $k_1, k_2 \in \mathbb{Z}$ and $\mathbf{x} = (x, y) \in \mathbb{Z}^2$. It follows that $\{\phi_{j,\mathbf{k}}(\mathbf{x})\}$ constitutes an orthonormal basis of $\mathbf{V_j}$. Recalling the multiresolution analysis construction in one dimension, for each $j \in \mathbb{Z}$ let $\mathbf{W}_j$ be the orthogonal (difference) complement of $\mathbf{V}_j$ in $\mathbf{V}_{j+1}$. Following Mallat (1999, Section 7.7) we obtain,

$$\begin{aligned}
\mathbf{V}_{j+1} &= (V_{j,(1)} \oplus W_{j,(1)}) \otimes (V_{j,(2)} \oplus W_{j,(2)}) && (2.53)\\
&= (V_{j,(1)} \otimes V_{j,(2)}) \oplus (V_{j,(1)} \otimes W_{j,(2)}) \oplus (V_{j,(1)} \otimes V_{j,(2)}) \oplus (W_{j,(1)} \otimes W_{j,(2)}) && (2.54)\\
&= \mathbf{V}_j \oplus \mathbf{W}_j^{(h)} \oplus \mathbf{W}_j^{(v)} \oplus \mathbf{W}_j^{(d)}. && (2.55)
\end{aligned}$$

The superscripts $\{h, v, d\}$ in the above stand for horizontal, vertical and diagonal directions, since the coefficients in these spaces describe the horizontal, vertical and diagonal features of the image. The spaces $\mathbf{W}_j^l$ where l corresponds to the direction are spanned by wavelets functions as follows:

$$\begin{aligned}
\psi^h_{j,\mathbf{k}}(x, y) &= \phi_{j,k_1}(x)\psi_{j,k_2}(y)\\
\psi^v_{j,\mathbf{k}}(x, y) &= \psi_{j,k_1}(x)\phi_{j,k_2}(y)\\
\psi^d_{j,\mathbf{k}}(x, y) &= \psi_{j,k_1}(x)\psi_{j,k_2}(y).
\end{aligned}$$

Hence an image, $f(x, y)$, defined in the space $L^2(\mathbb{R}^2)$ can be expressed as follows;

$$f(x, y) = \sum_{\mathbf{k}} c_{j_0,\mathbf{k}}\phi_{j_0,\mathbf{k}}(x, y) + \sum_l \sum_j \sum_{\mathbf{k}} d^{(l)}_{j,\mathbf{k}}\psi^{(l)}_{j,\mathbf{k}}(x, y), \tag{2.56}$$

where the sum over l is the sum over the three directions (horizontal, vertical and diagonal), $\mathbf{k} = (k_1, k_2)$ for $k_1, k_2 \in \mathbb{Z}$ and j relates to scale. In order to obtain the coefficients $c_{j_0,\mathbf{k}}$ and $d^{(l)}_{j,\mathbf{k}}$ we introduce the two-dimensional discrete wavelet transform using an extension of the cascade algorithm introduced in Section 2.2.1.

2.5.2 The two-dimensional discrete wavelet transform

Recall the one-dimensional DWT from Section 2.2.1. In a similar fashion the two-dimensional discrete wavelet transform proceeds as follows, under the assumption an image has a dyadic

dimensions. Suppose one has an n x n matrix c_J where $n = 2^J$ for $J \in \mathbb{Z}$. The formulae for calculating the smooth $c_{j_0,\mathbf{u}}$ and detail $d^{(l)}_{j,\mathbf{u}}$ coefficients for $j = J, ..., 1$ at a given location $\mathbf{u} = (u, v)$ in the image are given by,

$$c_{j-1,\mathbf{u}} = \sum_m \sum_n h_{m-2u} h_{n-2v} c_{j,(m,n)}, \tag{2.57}$$

$$d^h_{j-1,\mathbf{u}} = \sum_m \sum_n g_{m-2u} h_{n-2v} c_{j,(m,n)}, \tag{2.58}$$

$$d^v_{j-1,\mathbf{u}} = \sum_m \sum_n h_{m-2u} g_{n-2v} c_{j,(m,n)}, \tag{2.59}$$

$$d^d_{j-1,\mathbf{u}} = \sum_m \sum_n g_{m-2u} g_{n-2v} c_{j,(m,n)}. \tag{2.60}$$

A full wavelet decomposition is achieved by decomposing c_J using the above equations at each scale. The final set of coefficients are $(\mathbf{c}_0, \mathbf{d}^h_0, \mathbf{d}^v_0, \mathbf{d}^d_0, \ldots, \mathbf{d}^h_{J-1}, \mathbf{d}^v_{J-1}, \mathbf{d}^d_{J-1})$ where c_0 is the coarsest father coefficient and the $\mathbf{d}^l_j = \{d^l_{j,k}\}$.

The filtering equations $(2.56) - (2.59)$ can be written in terms of the operator notation introduced in Section 2.2.2. In particular they can be described as follows. Firstly, both the $\mathcal{D}_0\mathcal{H}$ and $\mathcal{D}_0\mathcal{G}$ operators are applied to the *rows* of c_J. This results in two n x $(n/2)$ matrices which can be labelled as H and G. Both operators are subsequently applied to the *columns* of H and G. This results in four matrices which we label HH, GH, HG, GG, each of dimension $(n/2)$ x $(n/2)$. The matrix HH is the result of applying the 'low-pass' operator $\mathcal{D}_0\mathcal{H}$ to both the rows and columns of c_J. The other matrices GH, HG and GG create the finest-scale wavelet detail in the horizontal, vertical and diagonal directions respectively. This algorithmic step is then repeated by applying the same filtering operations to HH, generating new matrices of dimension $(n/4)$ x $(n/4)$. These new matrices are HH, GH, HG and GG at the next finest scale and then the step is repeated by application to the new HH and so on. A schematic diagram of this can be seen in Figure 2.6. As one might expect the 2D-DWT is orthogonal and so we can invert the above algorithm. The inversion process is similar to the one dimensional case discussed previously. See Mallat (1989b) and Nason and Silverman (1994) for further details.

Figure 2.7 shows an example image and its two-dimensional DWT decomposition. With the exception of the lower left quadrant, the outer level represents the finest resolution level

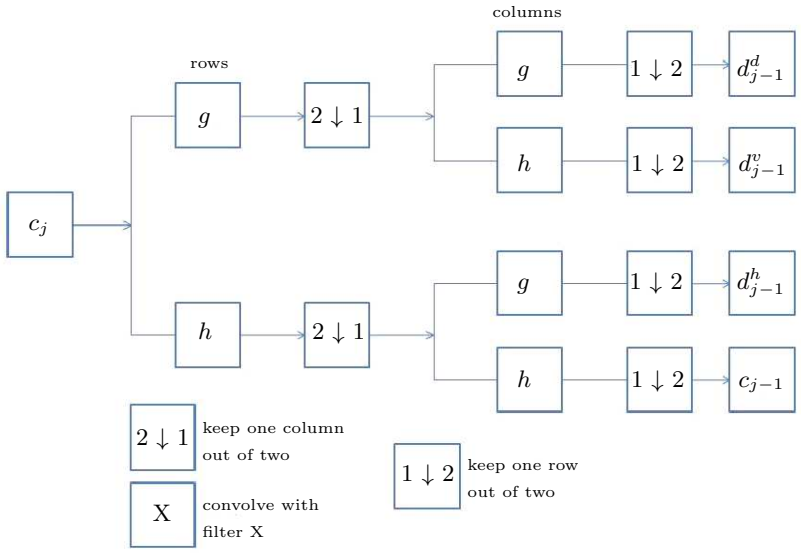


Figure 2.6: 2D DWT: decomposition of an image into smooth and detail components using quadrature mirror filters, **h** and **g**.

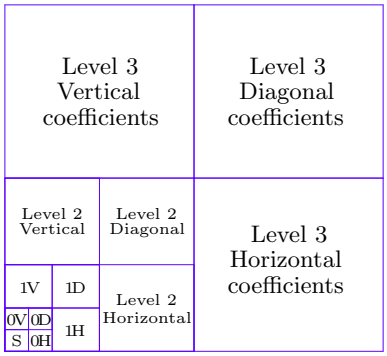
for each direction. The lower quadrant is segmented into four, where with the exception of the lower left again the others represent the detail of each direction at the next finest resolution level. This process repeats itself until we reach the coarsest scale smooth coefficient represented by S. The smooth coefficients appear as a coarser representation of the original image.

Recall in Section 2.3.1 we introduced the NDWT. This also has a natural extension to two dimensions in that we use zero-padded filters to convolve with the rows and columns of our data. This results in detail images at each level/direction which are the same dimension as the original.

2.6 Texture analysis

The main application area in this thesis is texture analysis, considering both greyscale and colour images. Here we give a very brief introduction to the area though further review material will be provided in Chapters 3 and 4, with a review of colour texture analysis postponed until Chapter 6. Specifically in this section we introduce a definition of texture and discuss various approaches to the three main texture analysis activities, namely

(a) (b)



(c)

Figure 2.7: (a) Original image, (b) DWT of image and (c) DWT hierarchy.

discrimination, classification and segmentation.

Whilst there is no formal mathematical definition of texture we adopt the definition of Petrou and Sevilla (2006) who state that texture is the variation of data at scales smaller than the scales of interest. In addition, we can consider that a region in an image has a *constant texture* if a set of local statistics or other local properties of the image function are constant, slowly varying or approximately periodic (Tuceryan and Jain, 1998). In particular in subsequent chapters we focus on the variance and covariance properties of an image. Figure 2.8 provides examples of textured images taken on a standard digital camera. The images in Figure 2.8 are visually very different and easy to distinguish. Such patterns are

(a) (b) (c)

Figure 2.8: (a) Brick Wall, (b) Grass, (c) Pebbles. All images were provided by A. Gott.

induced by physical properties, for example roughness and edges.

One of the most important properties of texture is its scale dependence as structure in an image frequently exists on many different scales. For example, Figure 2.8(a) shows an image of a brick wall displaying different textured elements at the coarsest scale. We can see longer bricks in the middle and smaller bricks around the edges. Texture analysis involves using summary approaches to extract features in order to characterise texture. Such features could include:

i) Edges - these are discontinuities, for example the edge of each brick in Figure 2.8(a).

ii) Non-stationary structure, for example:

   1. Changes in mean, e.g. the change in shade between pebbles in Figure 2.8(c).
   2. Changes in covariance structure, e.g. the grass as a whole rather than looking at each blade individually in Figure 2.8(b).

iii) Scale structure, for example, in Figure 2.8(a) if we are interested in the wall as a whole, then each brick is considered as a texture. However if we are interested in identifying an individual brick, each brick is a non-textured object at this scale of interest, as there are hardly any details inside it.

The above examples are reasonably straightforward to differentiate. However the examples encountered by our industrial collaborator such as the images in Figure 2.9 are more challenging for the eye to identify differences, as they exhibit complex and subtle structure. For example Figure 2.9 shows images of hair which are similar so it takes longer to identify

(a) (b)

Figure 2.9: Hair images obtained from an industrial collaborator.

dissimilarities, such as shade and caliber. By eye we can see how the image on the right of Figure 2.9 appears to show more of a wave in the central part of the image.

From a statistical perspective Petrou and Sevilla (2006) argue that there are two key classes of textured images: stationary and non-stationary. A texture is said to be stationary if it contains a single type of texture, i.e. the same texture fills up the whole image and so its local statistical properties are the same everywhere. It is classified as non-stationary if it contains more than one type of texture.

We now briefly outline various texture analysis approaches. As we shall see these are predominantly based on statistical thinking rather than formal modelling. For a more comprehensive review of this area we refer the reader to Arivazhagan and Ganesan (2003).

Traditional statistical approaches to texture analysis include: co-occurence matrices (Chen and Pavlidis (1979)), second-order statistics (Haralick (1979)), Gaussian Markov random fields (Cross and Jain (1983)) and local linear transforms (Unser (1986)). However these methods are restricted to the analysis of spatial interactions over small neighbourhoods viewed on a single scale. This is not ideal for texture-based applications as textures possess structure on many different scales. In addition several researchers have highlighted that the human and mammalian visual systems process images in a multiscale manner, preserving both local and global information (Field, 1999). The multiscale behaviour of texture (Jolion and Rosenfeld, 1994; Petrou and Sevilla, 2006) would then suggest that a multiscale approach would perform better than traditional single resolution techniques and a number of wavelet-based approaches have been considered.

Comprehensive reviews of wavelet-based statistical approaches to texture analysis are

given by Scheunders et al. (1998) and Sonka et al. (1999). We give a brief discussion here of some of the existing wavelet methods. Briefly, wavelet-based approaches to image segmentation include Mao and Jain (1992); Lu et al. (1997); Zhang and Oe (1998); Arivazhagan and Ganesan (2003). These methods take a wavelet decomposition of the original image and texture features are calculated from the wavelet coefficients. Such features include energy measures at different scales and the mean and variance detection for each filtered image.

Unser (1995) demonstrated the advantage of using the wavelet transform within a segmentation framework due to its many properties such as multiresolution representation and orthogonality. Unser (1995) concluded that second-order statistics such as the variance and covariance may be best for segmentation of microtextures due to their ability to represent the inherent structure of texture efficiently.

Possible wavelet approaches to texture classification include Chang and Kuo (1993) which is based on the wavelet packet transform and Unser (1995), based on a form of non-decimated wavelet. In both cases, statistical thinking rather than formal modelling is used to create a feature set for classification purposes. Eckley et al. (2010) also use a NDWT, but embed the wavelet transform within a statistical modelling framework, reviewed in Chapter 3. The advantages of this framework are shown in their results, with the modelling approach showing higher classification successes than using the DWT or NDWT.

Throughout the remainder of this thesis we consider texture and its properties in further detail. An automatic detection method to establish stationarity properties is discussed in Chapter 3. Chapter 4 proposes a statistical-based approach to count the number of unique texture regions within an image. This can then be used to segment an image. Chapter 5 introduces a multivariate extension of the two-dimensional locally stationary wavelet model and we conclude in Chapter 6 by applying this to a discrimination and classification problem relating to colour texture images.

Chapter 3

A test of stationarity for textured images

Abstract

This article proposes a test of stationarity for random fields on a regular lattice motivated by a problem arising from texture analysis. Our approach is founded on the locally stationary two-dimensional wavelet (LS2W) process model for lattice processes which has previously been used for standard texture analysis tasks such as texture discrimination and classification. The proposed test can be performed on a single realisation – a feature of particular practical importance within texture analysis. We illustrate our approach with pilled fabric data, demonstrating that the test is capable of identifying visually-subtle changes in stationarity.

3.1 Introduction

This article is motivated by an application emerging from the analysis of textured images. When one thinks about *texture*, a typical example that comes to mind is that of a woven material, straw or a brick wall. More formally, *image texture* is the visual property of an image region with some degree of regularity or pattern: it describes the variation in the data at smaller scales than the current perspective (Petrou and Sevilla, 2006). Texture structure can be thus considered to exist on several different *scales* of an image. Moreover it is well-

documented that image processing within the mammalian visual system is performed in such a manner to preserve local and global information, see for example Daugman (1990), Jolion and Rosenfeld (1994) or Field (1999). It is therefore desirable that texture analysis tools reflect these two important properties of texture, namely that (i) it has a location-dependent structure and (ii) it is *multiscale* in nature.

The image set which we consider (Figure 3.1) arises from work with an industrial collaborator. It comprises six fabrics buffed to varying degrees in an attempt to simulate different levels of garment wear. The effect of this abrasive process is to induce pilling – a build up of fibrous clusters on the surface of the material. By its very nature pill is a localised (short memory) phenomenon with the amount of pill in any region dependent on the amount of wear in that particular area (see Chan and Pang (2000) and Palmer et al. (2011) for further details). Pilling can therefore induce non-stationary behaviour across the fabric. Assessing the level of pilling is useful for product evaluation. For example highlighting the efficiency of fabric detergents in order to improve the wear of the material, e.g. by slowing the rate of pilling. Incorrectly judging the non-stationary behaviour of a texture can lead to a false perception of wear and efficiency of a product.

Within the field of texture analysis researchers are typically interested in three broad activities, namely texture discrimination, classification and segmentation. As such, appropriate and efficient modelling of the second-order properties of an image can often be an important consideration. Many established techniques for texture analysis have an underlying assumption of (second-order) stationarity, see for example Gonzalez and Woods (2001). In other words, the process has a constant mean, but the covariance between two spatial locations is a function of the vector difference between them: $\text{Cov}(X_{\mathbf{r_1}}, X_{\mathbf{r_2}}) = \gamma(\mathbf{r_1} - \mathbf{r_2})$.

Conversely, to account for the inherent multiscale structure of such images, others have proposed the use of wavelet-based approaches, see for example Laine and Fan (1993); Unser (1995); Eckley et al. (2010) or Mondal and Percival (2012). In particular the approach proposed by Eckley et al. (2010) provides a mechanism for modelling and estimating the non-stationary structure of textured images.

Wavelets are a form of localised basis function which provide a scale based decomposition of an image's structure (Vidakovic, 1999; Nason, 2008). Recent research by Eckley et al.

(2010) indicated that when a textured image appeared visually to be stationary then, as one might anticipate, Fourier-based classification approaches provide improved classification performance when compared against their wavelet counterparts. Conversely when an image is non-stationary, a wavelet based approach is more appropriate, due to their multiscale nature and location dependent decomposition.

In practice one will not generally know *a priori* whether or not the images being analysed are (second-order) stationary. It is therefore difficult to identify whether a multiscale or stationary approach should be adopted to analyse a given set of images. One way of resolving this issue would be to develop a test of (second-order) stationarity for such short memory processes based on a single realisation of each image. This is the question which we address in this paper, developing a new test of stationarity for random fields, highlighting its application with textured images.

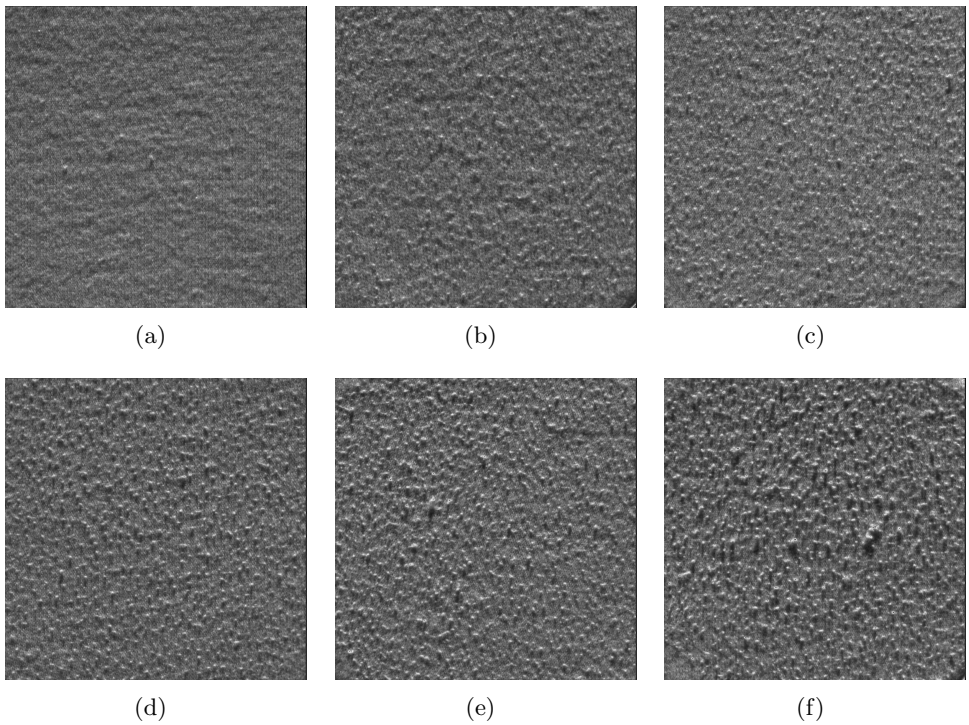


Figure 3.1: A sequence of six pilled fabric images. The amount of pilling increases across the images from (a) (lowest pill) to (f) (highest pill). The first five images clearly show a high degree of stationarity across the pills; due to the increased pilling, the sixth image shows small areas of uneven bobbling.

Existing tests of spatial stationarity. Several tests of stationarity for spatial processes have been proposed in the literature in recent years, with contributions coming from statistics, geology and the environmental sciences. We now consider the suitability of these to the texture analysis application described above. In particular we seek to identify an approach capable of using single process realisations and, ideally, which takes into account the scale-based nature of images.

Ephraty and collaborators have suggested a number of tests of spatial stationarity. Initial work, described in Ephraty et al. (1996), proposed a test statistic for stationarity calculated using the l_2 norm of the off-diagonal elements of the second-order cumulant spectrum (the Fourier transform of the image cross-correlation function). Under the null hypothesis of stationarity, only the diagonal elements of the matrix should be zero. A relatively large sample size is needed to ensure the accuracy of the method. Further work introduced by Ephraty et al. (2001) introduced a likelihood-based test and also a test using spectral estimation methods under the assumption of a low degree of stationarity. However, unfortunately this approach also requires multiple realisations of the process – a requirement that is often not possible in texture analysis and remote sensing settings.

Bose and Steinhardt (1996) formulate a hypothesis test using the generalised likelihood ratio statistic under the assumption of a centrosymmetric form of the spatial covariance. The covariance of stationary processes is known to have this property. This in turn means that subspaces spanned by particular eigenvectors can be inspected for orthogonality under the null hypothesis. However, the test also assumes multiple realisations of the spatial process and, as noted by Fuentes (2005), the proposed test is likely to be sensitive to the form of covariance.

A hierarchical Bayesian approach is proposed by Fuentes (2005) who models the continuous spatial process using a parametric form for the covariance structure and estimating the spectral density of a process via weight functions evaluated on windows. The author tests for stationarity by extending the 'analysis of variance' approach of Priestley and Rao (1969) to spatial processes. The dependence of the approach on the choice of window (weight function bandwidth) is an obvious computational drawback. Additionally we seek a *discrete-spatial* approach which can encapsulate the scale-based nature inherent within textured images.

Finally, Blanc et al. (2008) investigate the large sample behaviour of the empirical mean and variance statistics over a class of spatial processes for which the theoretical asymptotic behaviour is known. The rates of convergence of the statistics for an observed spatial process are estimated by image subsampling and then fitting a nonparametric estimator. Non-stationarity of the mean or variance is detected by looking for 'anomalies' in the behaviour of the empirical statistics compared with theoretical rates. Evaluating the large sample trend of the statistics is obviously computationally intensive. In addition, there is no clear measure of a sufficiently convergent trend and thus automatic implementation of the method is not considered. Hence, detection of non-stationarity is achieved through visual inspection. This approach is therefore not suitable for use with textured images where typically one may have a large number of candidate textures which need to be (automatically) analysed to identify whether they are stationary.

A wavelet-based approach? Each of the above methods suffer various disadvantages for the application under consideration. The test of stationarity which we propose in Section 3.3 adopts the recently proposed wavelet-based model of Eckley et al. (2010). Specifically, the locally stationary framework of Eckley et al. (2010) provides the flexibility to accommodate realistic non-stationary behaviour whilst also being able to model the inherent multiscale structure of texture. It is therefore natural to consider whether this framework can be used to develop a test of stationarity for short memory random fields, such as the pill images, particularly since the wavelet-based framework permits estimation of the local spectrum with a single realisation of the spatial process. As such our proposed test does not suffer from the limitation of requiring multiple process realisations. Moreover we find that it is able to detect quite subtle locally non-stationary behaviour of the spatial process and demonstrate how this can be applied in the texture context.

The article is organised as follows. We begin, in Section 3.2, by providing a brief introduction to wavelets and two-dimensional locally stationary wavelet process. We then propose our test of stationarity in Section 3.3, providing assessments of the performance of the test through simulation in Section 3.4. Our approach is then applied to several texture examples provided by an industrial collaborator (Sections 3.4.2 and 3.4.3).

3.2 Wavelets and 2D locally stationary wavelet (LS2W) processes

Briefly, wavelets are oscillatory basis functions which provide efficient (sparse) multiscale representations of signals. For example, for a function $f \in L^2(\mathbb{R})$, we have the expression $f(x) = \sum_{k\in\mathbb{Z}} c_{0,k}\phi_{0,k}(x) + \sum_{j\leq J}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x)$, where the wavelet $\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$ is a dilated and translated version of a (mother) wavelet ψ and similarly for the father wavelet ϕ. The coefficients $d_{j,k}$ at location k and scale j represent the oscillatory behaviour of the signal f at a particular frequency, whereas the coefficients $c_{j,k}$ give information about the mean behaviour of the signal at different scales j.

Note: In the remainder of this thesis we consider j to be positive.

Wavelets have received considerable attention within the statistics community during the last 20 years, not least because of their ability to provide an efficient location-scale decomposition of signals (see Vidakovic (1999), Percival and Walden (2006) or Nason (2008) for accessible introductions to this area). Below we provide a brief overview of the pertinent theory associated with locally stationary two-dimensional wavelet processes.

3.2.1 Discrete wavelets

We begin by providing a formal definition of the key building blocks within the LS2W framework, namely discrete wavelets.

Let ψ be a (compactly supported) wavelet, such as those introduced by Daubechies (1992), and denote by $\{h_k, g_k\}$ the low- / high-pass filter pair associated with ψ. Furthermore, let N_h denote the number of non-zero coefficients of $\{h_k\}$, and define $L_j = (2^j - 1)(N_h - 1) + 1$, where L_j represents the discrete wavelet length.

As we are following the separable approach and considering the tensor products of two one dimensional filters. We begin by introducing one-dimensional discrete wavelets. The one-dimensional *discrete wavelets* at a given scale $j \in \mathbb{Z}^+$ as introduced by Nason et al. (2000) are defined to be the vectors $\psi_j = \left(\psi_{j,0}, \ldots, \psi_{j,L_j-1}\right)$, with $\psi_{-1n} = \sum_k g_{n-2k}\delta_{0k} = g_n$ and $\psi_{(j-1)n} = \sum_k h_{n-2k}\psi_{jk}$ for $n = 0, \ldots, L_{j-1} - 1$, where δ_{0k} is the Kronecker delta. The discrete father wavelet is defined similarly using the associated low-pass filter $\{h_k\}$. As

Eckley et al. (2010) note, this construction can easily be extended to two dimensions as follows:

Definition 7 *Let* $\mathbf{k} = (k_1, k_2)$ *where* $k_1, k_2 \in \mathbb{Z}$. *The 2D discrete wavelet filters,* $\{\psi_j^l\}$, *are defined as the finite square matrices (of dimension* $L_j \times L_j$*):*

$$\psi_j^l = \begin{bmatrix} \psi^l_{j,(0,0)} & \cdots & \psi^l_{j,(0,L_{j-1})} \\ \vdots & \vdots & \vdots \\ \psi^l_{j,(L_{j-1},0)} & \cdots & \psi^l_{j,(L_{j-1},L_{j-1})} \end{bmatrix},$$

for directions $l = h, v$ *or* d, *i.e. horizontal, vertical and diagonal, where the elements of the wavelets are the tensor products of the corresponding 1D discrete wavelets:* $\psi^h_{j,\mathbf{k}} = \phi_{j,k_1}\psi_{j,k_2}$; $\psi^v_{j,\mathbf{k}} = \psi_{j,k_1}\phi_{j,k_2}$ *and* $\psi^d_{j,\mathbf{k}} = \psi_{j,k_1}\psi_{j,k_2}$.

Discrete father wavelets $\phi_{j,\mathbf{k}}$ can be defined similarly in two dimensions by taking the tensor product of 1D discrete father wavelets.

Finally, a family of *nondecimated* discrete wavelets is formed via translations in $\mathbb{Z}^2$ as $\psi^l_{j,\mathbf{u}}(\mathbf{r}) = \psi^l_{j,\mathbf{u}-\mathbf{r}}$ for every scale j, direction l and locations $\mathbf{u}, \mathbf{r} \in \mathbb{Z}^2$. It is these which we use in the spatial process model introduced in the next section.

3.2.2 The LS2W spatial model

The test of stationarity which we introduce in Section 3.3 extends ideas presented in a time series context by Cardinali and Nason (2011) to a spatial setting. To achieve this, we adopt the locally stationary spatial modelling framework introduced by Eckley et al. (2010). This introduced a new class of multiscale lattice processes with a location-dependent second-order structure. Instead of assuming a stationary process behaviour, these processes are assumed to have a *locally stationary* character. In other words, the covariance is assumed to vary across (pixel) locations of an image, as typically seen in many everyday examples of texture, e.g. wear on a garment made from woven fabric. Eckley et al. (2010) refer to spatial processes constructed under such a model as *locally stationary wavelet fields* (LS2W). We now provide an introduction to the main elements of the LS2W modelling approach.

The locally stationary two-dimensional wavelet process model introduced by Eckley et al.

(2010) is defined as

$$X_{\boldsymbol{r};\boldsymbol{R}} = \sum_{l} \sum_{j=1}^{\infty} \sum_{\boldsymbol{u}} w^{l}_{j,\boldsymbol{u};\mathbf{R}} \psi^{l}_{j,\boldsymbol{u}}(\boldsymbol{r}) \xi^{l}_{j,\boldsymbol{u}}, \tag{3.1}$$

for directions $l = h, v$ or d and spatial locations $\boldsymbol{r} = (r, s) \in \{0, \ldots, R-1\} \times \{0, \ldots, S-1\} = \mathcal{R}$, where $R = 2^m$, $S = 2^n \geq 1$, with $n, m \in \mathbb{N}$. We denote the lowest common scale as $J(R, S) \equiv \log_2\{\min(R, S)\}$. In equation (3.1),

(i) $\{\xi^{l}_{j,\boldsymbol{u}}\}$ is a zero-mean random orthonormal increment sequence;

(ii) $\{\psi^{l}_{j,\boldsymbol{u}}\}$ are a set of discrete nondecimated wavelets (see Definition 7);

(iii) $\{w^{l}_{j,\boldsymbol{u}}\}$ are a collection of amplitudes which quantify the contribution made to the process at location $\mathbf{u}$.

The LS2W model requires the following assumptions:

1. $\mathbb{E}[\xi^{l}_{j,\mathbf{u}}] = 0$ hence $\mathbb{E}(X_{\mathbf{r}}) = 0$ for all l, j and $\mathbf{u}$.
2. The increment sequence should be uncorrelated, $\mathbb{E}\left(\xi^{l}_{j,\mathbf{k}} \xi^{p}_{m,\mathbf{n}}\right) = \delta_{j,m}\delta_{\mathbf{k},\mathbf{n}}\delta_{l,p}$.
3. For each decomposition directions, l and scale $j \geq 1$, there exists a Lipschitz continuous function, $W^{l}_{j}(\mathbf{z})$ such that:
   (i) $\sum_l \sum_{j=1}^{\infty} |W^{l}_{j}(\mathbf{z})|^2 < \infty$ uniformly in $\mathbf{z} \in (0, 1)^2$.
   (ii) The Lipschitz constants L^{l}_{j} of W^{l}_{j} are uniformly bounded in j, l and

$$\sum_{l} \sum_{j=1}^{\infty} 2^{2j} L^{l}_{j} < \infty.$$

   (iii) Also there exists a sequence of constants C^{l}_{j} such that for each lattice $\mathbf{R}$

$$\sup_{\mathbf{u}} |w^{l}_{j,\mathbf{u};\mathbf{R}} - W^{l}_{j}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)| \leqslant \frac{C^{l}_{j}}{max\{R, S\}} \tag{3.2}$$

   where $\{C^{l}_{j}\}$ fulfills $\sum_l \sum_{j=1}^{\infty} C^{l}_{j} < \infty$.

Note that the locally stationary wavelet process has a dependence on the dimension of the image, $\boldsymbol{R} = (R, S)$. However, for notational convenience we drop this explicit dependence and denote such a process as $X_{\boldsymbol{r}}$, though the dependence is always assumed.

3.2.2.1 Example of an LS2W process

We now introduce an example of an LS2W process, namely a unit variance diagonal 2D Haar MA process of order 2. We assess the stationarity of this process for various image sizes (See S3 in Section 3.4.1) in order to show the reliability of our testing procedure (see Section 3.3). A Haar MA field of order 2, in direction d, is defined to be the LS2W process $\mathbf{X}_{\mathbf{r}}^{2,d}$ generated by Haar 2D nondecimated discrete wavelets with the following amplitudes:

$$w_{j,\boldsymbol{u}}^{l} = \begin{cases} 1 & \text{for} \quad j = 2, l = d \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

and assuming an orthonormal increment sequence. Substituting these conditions into the representation of the model in equation (3.1) gives

$$\mathbf{X}_{\mathbf{r}} = \sum_{\mathbf{u}} w_{2,\mathbf{u}}^{d} \psi_{2,\mathbf{u}}^{d}(\mathbf{r}) \xi_{2,\mathbf{u}}^{d}.$$

3.2.3 The local wavelet spectrum

Analogous to Fourier-based spectral theory, one can define the *local wavelet spectrum* (LWS) associated with an LS2W process.

Definition 8 *The LWS for a given location $\boldsymbol{z} \in (0,1)^2$, at scale j in direction l of a LS2W process $\{\mathbf{X}_{\mathbf{r}}\}$ is*

$$S_j^l(\mathbf{z}) = |W_j^l(\mathbf{z})|^2 \tag{3.4}$$

where $W_j^l(\mathbf{z})$ can be defined loosely as

$$W_j^l(\mathbf{z}) \approx w_{j,\mathbf{u}}^{l}, \tag{3.5}$$

due to equation (3.2).

As such the LWS quantifies the contribution to the process variance at *rescaled* spatial locations $\boldsymbol{z} \in (0,1)^2$, decomposition directions l, and scales, j. It should be noted that we use rescaled time, i.e. $\mathbf{z} = (\boldsymbol{u}/\boldsymbol{R})$ where $\boldsymbol{u}/\boldsymbol{R} := (u/R, v/S)$ since this allows us to obtain more information about the local structure of $W_j^l(\mathbf{z})$ as R, S tends to ∞.

The LWS also inherits Lipschitz properties. Following Eckley et al. (2009), if we set $\mathbf{z} = (z_1, z_2)$ and $\boldsymbol{\tau} = (\tau_1, \tau_2)$, then

$$\left| S_j^l \left(\frac{z_1 + \tau_1}{R}, \frac{z_2 + \tau_2}{S} \right) - S_j^l(z_1, z_2) \right| = O\left(L_j^l \left(\frac{|\tau_1|}{R} + \frac{|\tau_2|}{S} \right) \right). \tag{3.6}$$

Thus,

$$S_j^l \left(\frac{z_1 + \tau_1}{R}, \frac{z_2 + \tau_2}{S} \right) = S_j^l(z_1, z_2) + O\left(L_j^l \left(\frac{|\tau_1| + |\tau_2|}{min\{R, S\}} \right) \right). \tag{3.7}$$

This is useful in proving the asymptotic properties of the spectrum estimator which is discussed below.

To assess the stationarity, or otherwise, of a textured image, the approach proposed in Section 3.3 uses an estimate of the LWS. Drawing parallels with estimation theory associated with the Fourier spectral density, Eckley et al. (2010) propose the following estimator for the LWS:

$$I_{j,\boldsymbol{u}}^l = |d_{j,\boldsymbol{u}}^l|^2 = \left(\sum_{\boldsymbol{r}} X_{\boldsymbol{r}} \psi_{j,\boldsymbol{u}}^l(\boldsymbol{r}) \right)^2, \tag{3.8}$$

where $d_{j,\boldsymbol{u}}^l = \sum_{\boldsymbol{r}} X_{\boldsymbol{r}} \psi_{j,\boldsymbol{u}}^l(\boldsymbol{r})$ denotes the empirical wavelet coefficients of the LS2W process, $X_{\boldsymbol{r}}$, at a particular location, scale and direction. Defining $\boldsymbol{z} = \boldsymbol{u}/\boldsymbol{R}$ as above, the array $\boldsymbol{I}(\boldsymbol{z}) = \{I_{j,\boldsymbol{u}}^l\}$ for $j = 1, \ldots, J$, $l \in \{h, v, d\}$, and locations $\boldsymbol{u} \in \mathcal{R}$ in equation (3.8) is referred to as the raw *local wavelet periodogram* (LWP). As in Fourier theory, the raw LWP is biased as an estimator for the LWS. However Eckley et al. (2010) established that the periodogram estimator can be bias-corrected using the inverse of the inner product matrix of two-dimensional discrete autocorrelation wavelets (A_J), i.e.

$$\hat{S}(\boldsymbol{z}) = \boldsymbol{L}(\boldsymbol{z}) = A_J^{-1} \boldsymbol{I}(\boldsymbol{z}), \tag{3.9}$$

(see Section 3.2.4 for further details). It is this bias-corrected version of the LWP which we incorporate within our test of stationarity, introduced in Section 3.3.

3.2.4 Autocorrelation wavelets and local covariance

We begin by introducing the autocorrelation functions of discrete wavelets. These functions are required to obtain a measure of the local autocovariance structure of LS2W processes. Following Nason et al. (2000), in the one-dimensional setting autocorrelation wavelets are defined as follows:

$$\Psi_j(\tau) = \sum_k \psi_{j,k}(0)\psi_{j,k}(\tau), \tag{3.10}$$

for $j > 0, \tau \in \mathbb{Z}$. The autocorrelation father wavelet $\Phi_j(\tau)$ is defined analogously by replacing ψ by ϕ in equation (3.10). The two-dimensional extension for 2D discrete wavelets is defined as follows.

Definition 9 *The 2D autocorrelation wavelet at scale $j \in \mathbb{N}$, direction $l \in \{v, h, d\}$ and lag $\tau \in \mathbb{Z}^2$ is defined as,*

$$\Psi_j^l(\tau) = \sum_{\mathbf{v}\in\mathbb{Z}^2} \psi_{j,\mathbf{v}}^l(\mathbf{0})\psi_{j,\mathbf{v}}^l(\boldsymbol{\tau}). \tag{3.11}$$

The 2D autocorrelation wavelets are separable in each direction and can be represented in terms of the 1D autocorrelation wavelets given in equation (3.10). We have $\Psi_j^h(\boldsymbol{\tau}) = \Phi_j(\tau_1)\Psi_j(\tau_2), \Psi_j^v(\boldsymbol{\tau}) = \Psi_j(\tau_1)\Phi_j(\tau_2), \Psi_j^d(\boldsymbol{\tau}) = \Psi_j(\tau_1)\Psi_j(\tau_2)$ where $\boldsymbol{\tau} = (\tau_1, \tau_2)$. The 2D discrete autocorrelation father wavelet can be expressed similarly.

Eckley et al. (2010) show that the LWS as defined in the previous section is unique for a given LS2W process. In order to show this the inner product matrix of discrete autocorrelation wavelets $A = (A_{\eta,\nu})_{\eta,\nu\geq 1}$ is required, where

$$(A_{\eta,\nu}) = (< \Psi_\eta, \Psi_\nu >) = \sum_\tau \Psi_\eta(\tau)\,\Psi_\nu(\tau)\,. \tag{3.12}$$

The matrix A can also be shown to be invertible (see Eckley et al. (2010) for details) and is used in equation (3.9) to produce an unbiased estimator of the spectrum. Within this definition we have introduced an alternative form of indexing. Instead of having two separate indices representing scale and direction (i.e j and l), a combination of both provides a single index, η, each value of which represents a particular decomposition scale in a given direction. We have $\eta(j, l) = j + g(l)$ for all $j = 1, \ldots, J$ where $g(l) = 0, J, 2J$ and $l = v, h, d$

respectively. Hence the first J entries of η correspond to vertical wavelets, the next J to horizontal and the last J to diagonal. To simplify this notation we will omit the dependency on j and l and simply refer to η.

Recall from classical time series analysis that the covariance of a stationary process may be defined in terms of its spectral representation. With a spectral representation of the LS2W process established it is perhaps not surprising that in the same way the covariance of a LS2W process can be expressed in terms of its spectrum. Given Definition 9 we may now define the local autocovariance function for a given LS2W process.

Definition 10 *The local autocovariance (LCV), $C(\mathbf{z}, \boldsymbol{\tau})$, of a LS2W process with LWS $\{S_\eta(\mathbf{z})\}$, is defined as*

$$C(\mathbf{z}, \boldsymbol{\tau}) = \sum_{\eta} S_\eta(\mathbf{z}) \Psi_\eta(\boldsymbol{\tau}), \tag{3.13}$$

where $\boldsymbol{\tau} \in \mathbb{Z}^2$ and $\mathbf{z} \in (0,1)^2$.

In addition we can represent the spectrum in terms of the local autocovariance.

In the next section we use the LS2W modelling framework in order to design a hypothesis-based test of stationarity.

3.3 Testing for stationarity in LS2W processes

We now introduce our test of stationarity for textured images. Our approach extends earlier work by Cardinali and Nason (2011) from a time series to a two-dimensional setting. The test takes the form of a hypothesis test for which a particular statistic is used to measure the degree of non-stationarity under the null hypothesis of stationarity.

To formalise the hypothesis test we note that a (spatial) process is stationary if and only if its spectrum is constant across locations for all scale-direction pairs. In other words, we can test a process spectrum for constancy in order to determine whether the process is stationary. Thus given an observed process, $X_{\mathbf{r}}$, with associated wavelet spectrum $\boldsymbol{S} = \{S_j^l(\boldsymbol{z})\}_{j,l}$, our hypothesis test is

$$H_0\text{: } S_j^l(\boldsymbol{z}) \text{ is a constant function of } \boldsymbol{z} \text{ for all } j > 0 \text{ and } l \in \{h, v, d\}$$

H_A: $S_j^l(\boldsymbol{z})$ is not a constant function of $\boldsymbol{z}$ for some scale j and direction l.

In view of the above observation, the null hypothesis corresponds to an assumption of stationarity of a spatial process. We therefore look for departures from constancy within each scale and direction of the local wavelet spectrum to indicate non-stationarity. This variation in the wavelet spectrum can be quantified using the following measure (one of many which could be adopted):

$$T\{S_j^l(\boldsymbol{z})\} = (3J)^{-1} \sum_l \sum_{j=1}^{J} \int \{S_j^l(\boldsymbol{z}) - \bar{S}_j^l\}^2 \, d\boldsymbol{z} \quad \text{with } \boldsymbol{z} \in (0,1)^2. \tag{3.14}$$

Here $\bar{S}_j^l = \int S_j^l(\boldsymbol{z}) \, d\boldsymbol{z}$ for a particular direction $l \in \{h, v, d\}$ and scale level $j = 1, \ldots, J$. Note that the quantity in (3.14) is zero if, and only if, the spectrum $S_j^l(\boldsymbol{z})$ is constant across locations for each scale-direction pair.

Since the spectrum $\boldsymbol{S}(\boldsymbol{z})$ is unknown, in practice it is replaced by an estimator, $\hat{\boldsymbol{S}}(\boldsymbol{z})$, for example the corrected LWS $\boldsymbol{L}(\boldsymbol{z})$ described in Section 3.2.3. Note that, for a given realisation, $\hat{S}_j^l(\boldsymbol{z})$ can be denoted $\hat{S}_{j,\boldsymbol{u}}^l$. In addition, the integral $\bar{S}_j^l$ can be estimated by

$$\widetilde{S}_j^l = (RS)^{-1} \sum_{\boldsymbol{u} \in \mathcal{R}} \hat{S}_{j,\boldsymbol{u}}^l. \tag{3.15}$$

Taking $\hat{S}_{j,\boldsymbol{u}}^l = L_{j,\boldsymbol{u}}^l$ (the corrected LWP) in (3.15) results in a consistent estimator for the spectrum under the assumption of stationarity (see Appendix 3.6 for a proof which establishes this property). In a similar fashion, we can estimate the integral in equation (3.14) using the empirical variance (across locations) for a fixed scale-direction spectrum pair. Hence our test statistic can be calculated as follows

$$T\{\hat{S}_j^l(\boldsymbol{z})\} = (3J)^{-1} \sum_l \sum_{j=1}^{J} \mathrm{Var}_{\boldsymbol{u}}(\hat{S}_{j,\boldsymbol{u}}^l), \tag{3.16}$$

where the variance is taken over the lattice $\mathcal{R}$. The test statistic in equation (3.16) can be seen as the mean empirical variance of the spectrum estimate, where the average is taken over all scale-direction pairs.

3.3.1 The test procedure

To perform the above hypothesis test we require knowledge of the distribution of the test statistic $T\{\hat{S}_j^l(\mathbf{z})\}$, under the null hypothesis. Unfortunately, in most practical cases this distribution will in general be unknown. However, the spectrum $\boldsymbol{S}$ determines the distribution of the test statistic (assuming stationarity). We therefore perform (parametric) bootstrap resampling on the model innovations $\{\xi_{j,\boldsymbol{u}}^l\}$ together with the assumed stationary spectral structure to establish the distribution of the test statistic under H_0.

As noted in Cardinali and Nason (2011), the estimator $\hat{S}_j^l(\boldsymbol{z})$ in Section 3.3 needs to be a consistent estimator for the stationary spectrum (under the null hypothesis) for the parametric bootstrap implementation to be valid. To this end, we establish the consistency of our estimator $\bar{\boldsymbol{L}} = (RS)^{-1}\sum_{\boldsymbol{u}\in\mathcal{R}}\boldsymbol{L}_{\boldsymbol{u}}$ in Appendix 3.6.

The bootstrap test works by first computing the spectral estimate $\boldsymbol{L}$ for the observed image and then calculating the test statistic T^{obs} (equation (3.16)). We then obtain simulated realisations by the following procedure (Algorithm 3.1). In essence we simply simulate the underlying innovations $\{\xi_{j,\boldsymbol{u}}^l\}$ (see Remark 1) and feed these into our process model under the null hypothesis, inverting to obtain a realisation in the spatial domain. For each simulated stationary process, we compute the test statistic (3.16). The significance of the test statistic for the observed image can then be assessed by appealing to Monte Carlo arguments (see Davison et al. (1999) for more details). Our bootstrap approach, which we call $\texttt{Bootstat}_{\texttt{LS2W}}$, is summarised in Algorithm 3.1.

The $\texttt{Bootstat}_{\texttt{LS2W}}$ test can be interpreted as evaluating how unlikely the value of T^{obs} is compared to realistic (bootstrap) values of T assuming a *stationary* spectral structure based on the observed process. Thus the p-value of the test can be seen as a measure of how non-stationary the observed process is.

Remark 1. Within the bootstrap procedure described above, the realisations of the LS2W process under the null hypothesis are simulated by assuming the innovations of the process take a given distributional form. In particular we have assumed that the innovations are Gaussian; however, other distributions can be used in the model representation (3.1).

`Bootstat`$_{\texttt{LS2W}}$`:`

1. Compute the estimate of the LWS for the observed image, $\hat{S}_j^l(\boldsymbol{z})$.
2. Evaluate T (equation (3.16)) on the observed image, call this value T^{obs}.
3. Compute the pixel average stationary spectrum $\widetilde{S}_j^l$ by taking the average of spectrum values for each scale and direction.
4. **Iterate** for i in 1 to B bootstraps
   (a) Simulate $X_{\boldsymbol{r}}^{(i)}$ from the stationary LS2W model using squared amplitudes given by $\widetilde{S}_j^l$ and Gaussian process innovations.
   (b) Compute the test statistic T on the simulated realisation, call this value $T^{(i)}$.
5. Compute the p-value for the test as $p = \frac{1+\#\{\, T^{obs} \leq T^{(i)} \,\}}{B+1}$.

Algorithm 3.1: The bootstrap algorithm for testing the stationarity of locally stationary images.

3.4 Examples

We now consider the performance of the `Bootstat`$_{\texttt{LS2W}}$ test of stationarity proposed in Section 3.3.1, focussing in particular on data generated from an industrial application. The analyses were performed using the LS2W spectral estimation implementation in the R add-on package *LS2W* (Eckley and Nason, 2011a). For simulating the Gaussian random fields in the study below, we used the *RandomFields* R package (Schlather, 2012); the spatial moving average processes were simulated with the R package *spdep* (Bivand et al., 2011) using modifications to code featured in Anselin (2005).

3.4.1 Simulated performance of the LS2W stationarity test

In order to investigate the performance of our test of stationarity from Section 3.3.1, we performed a simulation experiment focussing in particular on the size and power properties of the test.

Size assessment. To explore the empirical size of the test, we chose a number of different process types. Each process represents a different form of second-order *stationary* structure. S1 is a two-dimensional white noise process, i.e $X_{\boldsymbol{r}} \sim N(0, 1)$. S2 is a spatial moving average

process with parameter $\rho = 0.9$, i.e.

$$X(\boldsymbol{u}) = 0.9W\varepsilon + \varepsilon,$$

where ε is a random component with variance σ^2, W is a spatial weight matrix associated to a chosen neighbour structure. A realisation of this process is shown in Figure 3.2(b).

We also consider a unit variance diagonal 2D Haar MA process of order 2 (as discussed in Section 3.2.2.1). We denote this process by S3 (see Figure 3.2(c)). The fourth stationary process, S4, is an exponential covariance with range parameter $\phi = 2$:

$$C(\boldsymbol{u}, \boldsymbol{v}) = \sigma^2 \exp\left\{\frac{-\|\boldsymbol{u} - \boldsymbol{v}\|}{2}\right\}. \tag{3.17}$$

This process is similar to that considered for continuous spatial processes in Fuentes (2005). A realisation of this process is shown in Figure 3.2(d).

For the processes considered, we simulated $K = 1000$ spatial realisations for different square image sizes $R = S = 2^J$, with $J = 6, \ldots, 9$. We then examined each realisation with the test of stationarity as follows. We evaluated the Monte Carlo significance test in Section 3.3.1 using $B = 250$ bootstrap simulations, each time treating the realisation as observed. In other words, we perform the $\texttt{Bootstat}_{\texttt{LS2W}}$ hypothesis test and record whether the realisation is stationary or not at a 5% significance level. For each process, we then note the number of simulated realisations which resulted in rejecting the null hypothesis of stationarity. We then compare the proportion of those rejecting the null with the nominal size.

Table 3.1 explores the size properties of the $\texttt{Bootstat}_{\texttt{LS2W}}$ test on the stationary specifications expressed as a percentage of the $K = 1000$ images rejecting stationarity. For all four stationary processes, the percentage of images rejecting the null hypothesis of stationarity (i.e. judged as non-stationary) was below the nominal size of 5%. These results mirror those obtained by Cardinali and Nason (2011) for the time series setting – namely that this test approach is conservative.

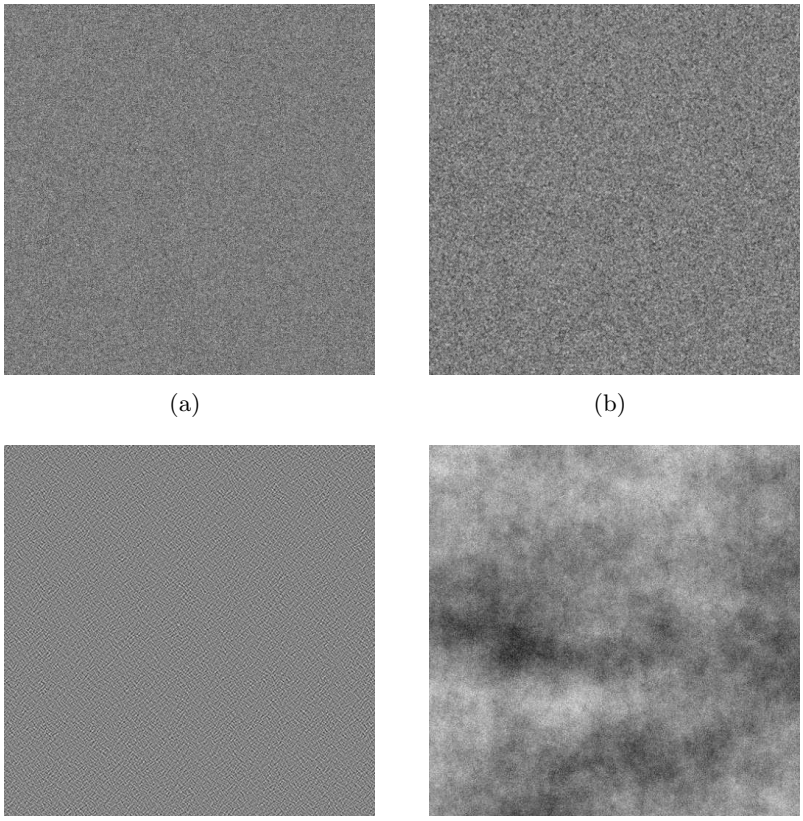


Figure 3.2: Process realisations from the stationary models for the simulations in Section 3.4.1 with $\boldsymbol{R} = 2^9 \times 2^9$. (a) S1. A two-dimensional Gaussian field; (b) S2. A moving average process; (c) S3. The process spectrum has power at a fixed coarse scale (level 2) in the diagonal detail direction; (d) S4. An exponential Gaussian random field with covariance (3.17).

Next we explore the power of our test to identify whether power is lost or maintained as a result of this conservativeness.

Power assessment. To evaluate the power of the LS2W stationarity test, we consider three contrasting processes which exhibit different non-stationary behaviour. In particular, we consider the following forms. NS1 describes a piecewise white noise spatial process, in which the left half-plane has unit variance, whereas the second (vertical) half-plane has variance σ^2. In other words, we simulate K images, each constructed from $2^{(n-1)} \times 2^n$ i.i.d. $N(0,1)$ samples for $\boldsymbol{z} \in (0, 1/2) \times (0, 1)$ concatenated with $2^{(n-1)} \times 2^n$ values sampled independently from $N(0, \sigma^2)$ for $\boldsymbol{z} \in (1/2, 1) \times (0, 1)$. Similarly we also consider a process in which the left half-plane is a unit variance white noise process, with the second half being

Table 3.1: Results from a simulation experiment to assess the empirical size of the LS2W test of stationarity. The table indicates the size (rounded to the nearest 1%) using a 2D white noise stationary process (S1); a spatial moving average process (S2); a diagonal 2D Haar MA process of order 2 (S3); an exponential Gaussian random field (S4). See text for details of the process used.

	S1				S2				S3				S4			
	Image dimension				Image dimension				Image dimension				Image dimension			
σ	64	128	256	512	64	128	256	512	64	128	256	512	64	128	256	512
0.50	0	0	0	0	0	1	0	2	0	0	0	0	1	0	2	0
1.0	0	0	0	0	1	1	0	2	0	0	0	0	0	0	0	0
1.4	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
1.8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0
2.4	0	0	0	0	2	1	0	1	0	0	0	0	1	0	0	2

an isotropic Gaussian random Field with a Matérn covariance (Matérn, 1960) with shape parameter $\nu = 1$ (NS2). NS3 represents an image which is a *montage* of four stationary processes in the four quadrants of the image (see Figure 3.3(b)); this is an LS2W process with spectral structure given by

$$w^d_{j,[2^J\mathbf{z}]} = \begin{cases} \sigma & \text{if } j = 1 \text{ and } \mathbf{z} \in (0, 1/2)\,\mathrm{x}\,(0, 1/2); \\ \sigma & \text{if } j = 2 \text{ and } \mathbf{z} \in (1/2, 1)\,\mathrm{x}\,(0, 1/2); \\ \sigma & \text{if } j = 3 \text{ and } \mathbf{z} \in (0, 1/2)\,\mathrm{x}\,(1/2, 1); \\ \sigma & \text{if } j = 4 \text{ and } \mathbf{z} \in (1/2, 1)\,\mathrm{x}\,(1/2, 1); \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

In particular, the process is a montage of four *diagonal Haar moving average* processes with different orders (Eckley et al., 2010).

The specification of σ in realisations of NS1 and NS3 controls the degree of non-stationarity in the simulated images: for low values of σ, the processes describe behaviour which is approximately stationary (i.e. the values are closer to unit variance across the whole image); for higher variance values the boundaries are more marked. For this study, we investigate the performance of the stationarity test with values in the range $\sigma = \{1.2, \ldots, 1.6\}$.

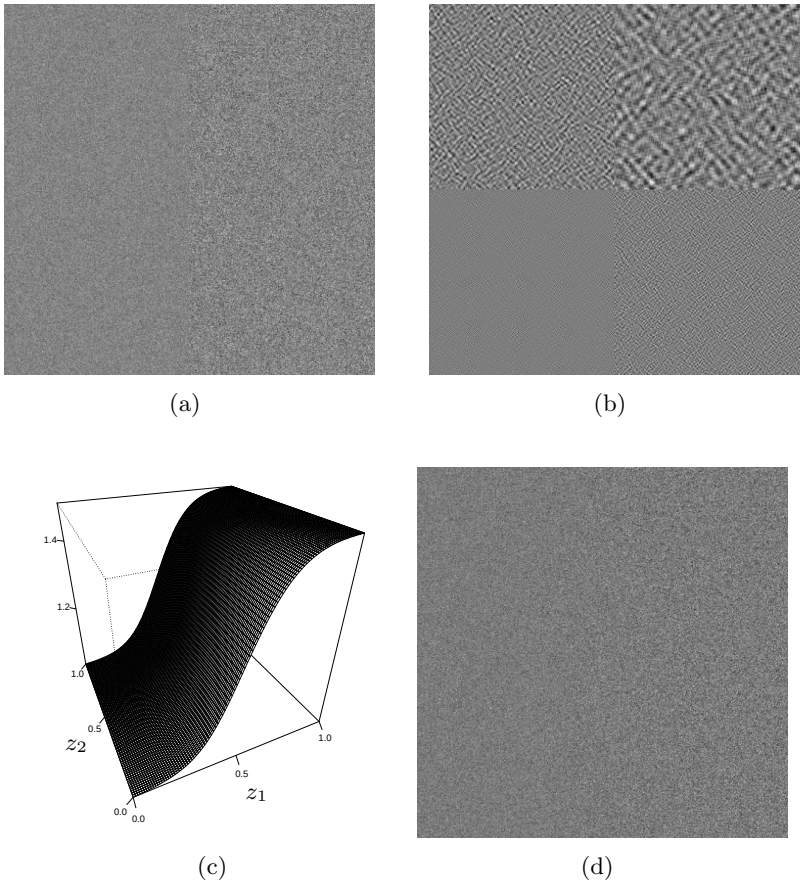


Figure 3.3: Process realisations from the non-stationary models for the simulations in Section 3.4.1 with $\boldsymbol{R} = 2^9 \times 2^9$. (a) NS1. The image consists of two Gaussian half planes; (b) NS3 (equation (3.18)). Each quadrant of the image is a $2^8 \times 2^8$ subimage sampled from a different diagonal Haar MA process. The texture ranges from finest detail (bottom-left) to coarsest (top-right); (c) A two-dimensional representation of the second order structure of the process NS4 specified in the spatial domain; (d) A realisation of the model NS4 (see equation (3.19)). The power displayed in the image varies smoothly across the horizontal plane according to Figure 3.3(c).

To investigate more subtle changes in second order structure, we also define a process specified in the spatial domain by $X_{\boldsymbol{r}}(\boldsymbol{z}) \sim N(0, \sigma^2_{a;\tau;\delta}(\boldsymbol{z}))$, where $\sigma_{a;\tau;\delta}(\boldsymbol{z})$ is constructed such that it varies spatially across an image. For the simulations below we use a particular choice of the standard deviation function which changes across the horizontal coordinate of the image according to the parameters a, τ and δ in the following way:

$$\sigma_{a;\tau;\delta}(\boldsymbol{z}) = \tau + \frac{\delta - \tau}{1 + \exp(-10a(2z_1 - 1))} \quad \text{for } z_1 \in (0, 1). \tag{3.19}$$

In other words, the standard deviation of the stochastic process varies smoothly from τ to

δ (see Figure 3.3(c)). Changing the parameter a has the effect of changing the shape of the deviation curve with lower values of a giving a more subtle behaviour across the x-axis. We note that if $\tau = 1$ in equation (3.19), as a increases the second order structure of NS4 process realisations will resemble that of realisations from model NS1. A realisation of NS4 can be seen in Figure 3.3(d).

A similar analysis to the size assessment was performed using 1000 realisations from the non-stationary processes NS1 – NS4 described above. In each case, the 1000 realisations were assessed for stationarity and the number which rejected the null recorded.

Table 3.2 shows the statistical power results (expressed as a percentage) for the test when applied to simulated images from processes NS1 – NS3. The results illustrate that the test is unable to distinguish between the two noise variances in NS1 for the lowest values of σ when the image dimension is small. However the results improve dramatically for larger image sizes. This is not particularly surprising since, for low values of σ, the non-stationary behaviour is more difficult to detect because the boundary in the simulated images will appear blurred. However, for intermediate and high values of σ and moderate image sizes, the test has good performance, rejecting the null hypothesis of stationarity for the majority of $K = 1000$ images (Table 3.2). The uniformity across σ values for processes NS2 and NS3 is reassuring, indicating that the test is insensitive to the severity of boundary non-stationarities, even for small variances.

For the smallest image dimension, the power results for NS4 indicate that the test fails to detect the non-stationary behaviour of the images. However, as observed for NS1, the percentage of correctly classified images improves dramatically as the size of the image under analysis increases for all scenarios, even for the quite subtle change in structure for NS4 described by low a values in $\sigma_{a;\tau;\delta}(\boldsymbol{z})$ (see Table 3.3). These results are consistent with the findings of Table 3.2 for NS1 and $\sigma = 1.5$, which can be seen as an extreme case of the process NS4.

These simulations suggest that for the non-stationary processes considered, the $\texttt{Bootstat}_{\texttt{LS2W}}$ test achieves good statistical power for image dimensions greater than $n = 128$. It is reassuring that power is maintained despite the conservative nature of the test observed in our exploration of empirical size. These results are consistent with those of other recently

Table 3.2: Results from a simulation experiment to assess the empirical power of the LS2W test of stationarity. The table indicates the power (rounded to the nearest 1%) under different non-stationary spectral specifications: piecewise Normal plane (NS1); Gaussian-Matérn process (NS2); Haar Montage (NS3). See text for details of the processes used.

		Image dimension			
Model		64	128	256	512
NS1	σ=1.2	1	24	99	100
	σ=1.3	20	98	100	100
	σ=1.5	97	100	100	100
	σ=1.6	100	100	100	100
NS2	σ=1.2	50	100	100	100
	σ=1.3	100	100	100	100
	σ=1.5	100	100	100	100
	σ=1.6	100	100	100	100
NS3	σ=1.2	100	100	100	100
	σ=1.5	100	100	100	100

Table 3.3: Empirical power assessment of the LS2W test of stationarity for process NS4. The table indicates the power (rounded to the nearest 1%) under different location-dependent variance structures given by equation (3.19).

	Image dimension			
(a, τ, δ)	64	128	256	512
(0.25,1,1.5)	8	77	100	100
(0.5,1,1.5)	58	100	100	100
(1,1,1.5)	86	100	100	100

published tests of stationarity in the time series literature. In particular the recent contributions by Cardinali and Nason (2011) and Dwivedi and Subba Rao (2011) each report a similarly conservative nature, yet retain power. In other words, for a stationary process, each of these procedures (including our own) will not reject the null hypothesis rather than report a false positive (reject for non-stationarity). However for a non-stationary process they would reject in favour of the alternative.

3.4.2 Analysis of pilled fabric images

We now apply the test of stationarity proposed in Section 3.3.1 to a number of real examples of textured images. In particular, we use our test of stationarity on a series of images of garment material which have been artificially buffed until the fabric displayed increasing degrees of pilling (build up of clumped fibres). A similar set of images, taken under slightly

different lighting conditions, were previously analysed by Eckley et al. (2010) in the context of texture classification. The fabrics are illustrated in Figure 3.1. Note that visually, all the images appear to be spatially stationary. However, it is arguable that the most heavily-pilled image (Figure 3.1(f)) displays a degree of non-stationarity due to more increased and irregular "bobbling" of the fabric.

As described in Section 3.1, our interest is to examine whether the images are indeed assessed as stationary by the test introduced in Section 3.3.1. Prior to performing the test, as outlined in Section 3.3, the six images were examined to verify that their spatial structure did not have a significant amount of long-range dependence. In addition, a median polish (Tukey, 1977) was applied to each image to remove any non-zero trend (so that the images met the zero-mean assumption of an LS2W process). The $\texttt{Bootstat}_{\texttt{LS2W}}$ stationarity test was then applied to each of the six individual images. In each case $B = 250$ bootstrap simulations were made. The p-value associated with each (bootstrap) test was then recorded for each fabric image. Note that the size of images analysed in this industrial application is typically large, and so we expect the test to be reliable in view of Section 3.4.1.

As one would perhaps anticipate, the $\texttt{Bootstat}_{\texttt{LS2W}}$ test judges the first five images as stationary (with p-values in the range (0.21,0.95)). This concurs with our visual perception of the images. The sixth image is judged as non-stationary, with a p-value of 0.03. This reflects what we might visually assess, namely that the sixth fabric has less regular texture structure than the other fabrics. Indeed, one could argue that this image contains regions of heavier and lighter pill.

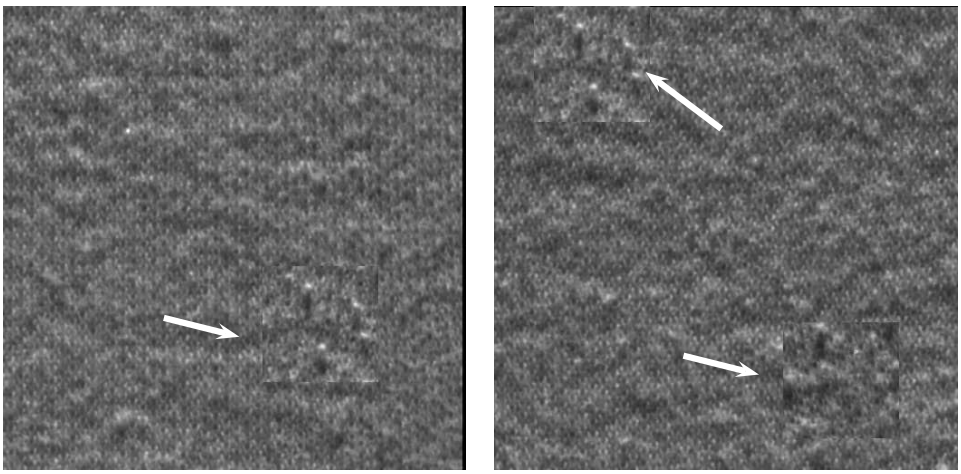
The test has also been applied to other textured images, such as a sequence of images of differing hair types and examples taken from reference texture libraries. In all cases we obtained similar statistical results – namely that mono-textures appear to be stationary. For reasons of brevity we do not report these results here.

3.4.3 Analysis of texture mosaics

We now turn to consider the application of our test of stationarity for a more realistic situation in fabric analysis. In many settings it is useful to be able to detect differing fabric structure, for example to identify whether there is an area of uneven wear within a sample

of material. To avoid the subjectivity of human inspection of materials it is thus desirable to develop an automatic detection method for uneven wear.

We begin by noting that regions which contain uneven wear consist of multiple texture types – the majority of one pill level, with some patches of another pill level. In other words, an image of a fabric which contains uneven wear could be considered to be non-stationary. Consequently a test of stationarity could be used as an automated proxy to detect whether such an area exists within a fabric sample. To examine how our proposed testing method handles this type of textural non-stationarity, we performed an experiment to mimic the situation described above. More specifically, we constructed some pilled fabric mosaics for analysis with the bootstrap test. A texture mosaic comprises two or more different texture subimages combined into one image for analysis. Two such mosaics were constructed by inserting sections of a texture image from Section 3.4.2 into another texture. Firstly a subimage of Pilled fabric 3 was inserted as the central part of Pilled fabric 1 (Mosaic A); Mosaic B contains a section of Pilled fabric 2 as well as Pilled fabric 3 within the lightest pilled fabric sample 1. The mosaics can be seen in Figure 3.4. The inserted subimages are indicated within the figures.



(a) (b)

Figure 3.4: Examples of texture mosaics of different pilled fabrics. (a) Mosaic A: a portion of Pill 3 inside Pill 1; (b) Mosaic B: a portion of Pills 2 and 3 inside Pill 1. The images display localised changes in texture. They can therefore be considered to be non-stationary, though this structure can be difficult to detect visually. On each image, the arrows show the subimages which were inserted to create the multitextured mosaics.

In both cases, application of the LS2W stationarity test on the texture mosaics indicated non-stationarity. This is to be expected since the images display different textural properties across the images, but is reassuring nevertheless since the texture boundaries are difficult to pick out visually. We also note that test was also able to detect non-stationarity when applied to examples of reference texture mosaics taken from several texture libraries.

3.5 Concluding remarks

As discussed in Section 3.1, recent work by Eckley et al. (2010) has indicated that, as a precursor to conducting a texture analytic task such as classification, it advisable to verify the stationarity (or otherwise) of candidate images. The work presented in this paper has addressed this issue by developing a bootstrap-based test of stationarity using the LS2W framework, extending the work of Cardinali and Nason (2011) to two dimensions. A benefit of this approach is that it permits the testing of the hypothesis using a single realisation.

When analysing simulated images, the proposed approach demonstrated good size and power performance, highlighting insensitivity of the test to image size and localised image variance, including slowly-varying second order structure. We also applied the test procedures to images encountered by an industrial collaborator. The results from these tests are consistent with our visual assessment of the images, namely that these textures are stationary, with heavy pilling introducing non-stationarity. In addition when analysing quite subtle texture mosaics, the test was correctly able to detect non-stationarity. As such, this approach could be used to detect, for example, regions of uneven wear within a material.

It is perhaps a little surprising that no *local Fourier* equivalent of this spatial test currently exists. Hence, we note that the development of a spatial analogue of local Fourier time series tests, such as Dwivedi and Subba Rao (2011) would be an interesting avenue for future research. The implementation of such a test could be similar to the $\texttt{Bootstat}_{\texttt{LS2W}}$ test, using bootstrapping in the Fourier domain.

3.6 Appendix: Properties of the estimator of the stationary spectrum

In this section we establish the asymptotic consistency of our estimator of $S_j^l(\boldsymbol{z})$ under the null hypothesis of stationarity, to confirm the validity of the bootstrap approach to the $\texttt{Bootstat}_{\texttt{LS2W}}$ test. The estimator takes the form:

$$\bar{\boldsymbol{L}} = (RS)^{-1} \sum_{\boldsymbol{u} \in \mathcal{R}} \boldsymbol{L}_{\boldsymbol{u}} = (RS)^{-1} \sum_{\boldsymbol{u} \in \mathcal{R}} A_J^{-1} \boldsymbol{I}_{\boldsymbol{u}}. \tag{3.20}$$

Our result extends that of Cardinali and Nason (2011) to a spatial setting.

Under the assumption of stationarity, the spectrum will be constant across all locations for each scale-direction pair. We denote this spectrum by $\boldsymbol{S}$, which is a length $3J$ vector of coefficients. In what follows, we use the shorthand notation $\eta = \eta(j, l)$ to represent an index running over all scale-direction pairs.

Let $\boldsymbol{\beta} = \{\beta_\eta\}_{\eta=1}^{3J} = A_J \boldsymbol{S}$. From (3.20), we consider the "uncorrected" periodogram estimator $\hat{\boldsymbol{\beta}}$ as an estimator of $\boldsymbol{\beta}$, with $\hat{\boldsymbol{\beta}} := A_J \bar{\boldsymbol{L}} = \bar{\boldsymbol{I}}$ and where $\bar{\boldsymbol{I}} = \frac{1}{RS} \sum_{\boldsymbol{u} \in \mathcal{R}} I^l_{j,\boldsymbol{u}}$ denotes the $3J$-vector of average periodogram values. From Theorem 2 of Eckley et al. (2010), we have that $\mathbb{E}(I_{\eta,[\boldsymbol{zR}]}) = A_J S(\boldsymbol{z}) + \mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)$. Thus under the null hypothesis of stationarity, $\mathbb{E}(\boldsymbol{I}_{\boldsymbol{u}}) = A_J \boldsymbol{S} + \mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)$ for all $\boldsymbol{u} \in \mathcal{R}$. Hence $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \frac{1}{RS} \sum_{\boldsymbol{u}} \mathbb{E}(\boldsymbol{I}_{\boldsymbol{u}}) = A_J \boldsymbol{S} + \mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)$, i.e. $\hat{\boldsymbol{\beta}}$ is an (asymptotically) unbiased estimator for $\boldsymbol{\beta}$.

For consistency, we need to show that $\mathrm{Var}(\hat{\beta}_\eta) \to 0$ as $R, S \to \infty$ for each η. We first note that under stationarity, equation (18) of Eckley et al. (2010) means that the periodogram variance can be expressed as

$$\mathrm{Var}(I_{\eta,\boldsymbol{u}}) = 2 \left(\sum_{\eta_1} A_{\eta_1 \eta} S_\eta \right)^2 + \mathcal{O}\left(\frac{2^{j(\eta)}}{\min\{R,S\}} \right), \tag{3.21}$$

for all locations $\boldsymbol{u} \in \mathcal{R}$. In other words, the variance is asymptotically constant, with the

constant, $\kappa(\eta)$, say, only dependent on η. For a fixed η,

$$\mathrm{Var}(\hat{\beta}_\eta) = \mathrm{Var}\left(\frac{1}{RS}\sum_{\boldsymbol{u}\in\mathcal{R}} I_{\eta,\boldsymbol{u}}\right) = \frac{1}{(RS)^2}\sum_{\boldsymbol{u}\in\mathcal{R}}\mathrm{Var}(I_{\eta,\boldsymbol{u}}) + \frac{2}{(RS)^2}\sum_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{A}(\mathcal{R}^2)}\mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}}), \tag{3.22}$$

where $I_{\eta,\boldsymbol{u}}$ denotes the ηth element of $\boldsymbol{I}_{\boldsymbol{u}}$ and the covariance sum is taken over the set $\mathcal{A}(\mathcal{R}^2)$ of all unique pairs of points $(\boldsymbol{u},\boldsymbol{v})$ in $\mathcal{R}\times\mathcal{R}$.

Consider each term in (3.22) separately. The first term, using the asymptotically constant expression from (3.21), can be written as

$$\frac{R\cdot S}{(RS)^2}\left(\kappa(\eta) + \mathcal{O}\left(\frac{2^{j(\eta)}}{\min\{R,S\}}\right)\right) \leq \frac{1}{\min\{R,S\}^2}\left(\kappa(\eta) + \mathcal{O}\left(\frac{2^{j(\eta)}}{\min\{R,S\}}\right)\right).$$

Hence the first quantity of (3.22) is asymptotically zero as $\min\{R,S\}\to\infty$ (or equivalently as $R,S\to\infty$).

Next we consider the second term in equation (3.22). Theorem 3 of Eckley et al. (2010) establishes that

$$\begin{aligned}\mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}}) &= 2\left\{\sum_{\eta_0}\sum_{\boldsymbol{u}_0}(w_{\eta_0,\boldsymbol{u}_0})^2\,\alpha_{\eta\eta_0}(\boldsymbol{u},\boldsymbol{u}_0)\alpha_{\eta\eta_0}(\boldsymbol{v},\boldsymbol{u}_0)\right\}^2 \\ &= 2\left\{\sum_{\eta_0}\sum_{\boldsymbol{u}_0}(w_{\eta_0,\boldsymbol{u}_0})^2\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,\boldsymbol{u}}(\boldsymbol{r}_1)\psi_{\eta_0,\boldsymbol{u}_0}(\boldsymbol{r}_1)\right)\left(\sum_{\boldsymbol{r}_2}\psi_{\eta,\boldsymbol{v}}(\boldsymbol{r}_2)\psi_{\eta_0,\boldsymbol{u}_0}(\boldsymbol{r}_2)\right)\right\}^2,\end{aligned}$$

where we abuse notation slightly and define $\alpha_{\eta_1\eta_2}(\boldsymbol{u}_1,\boldsymbol{u}_2) = \sum_{\boldsymbol{r}}\psi_{\eta_1,\boldsymbol{u}_1}(\boldsymbol{r})\psi_{\eta_2,\boldsymbol{u}_2}(\boldsymbol{r})$. Using the definition of the nondecimated wavelet vector $\psi_{\eta,\boldsymbol{u}}(\boldsymbol{r})$ and performing the substitutions $\boldsymbol{u}_0 = \boldsymbol{x}+\boldsymbol{u}$ and $\boldsymbol{s} = \boldsymbol{r}_2 - \boldsymbol{r}_1$, we write the covariance as

$$\begin{aligned}\mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}}) &= 2\left\{\sum_{\eta_0}\sum_{\boldsymbol{u}_0}(w_{\eta_0,\boldsymbol{u}_0})^2\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,\boldsymbol{u}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{u}_0-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{r}_2}\psi_{\eta,\boldsymbol{v}-\boldsymbol{r}_2}\psi_{\eta_0,\boldsymbol{u}_0-\boldsymbol{r}_2}\right)\right\}^2 \\ &= 2\left\{\sum_{\eta_0}\sum_{\boldsymbol{x}}(w_{\eta_0,\boldsymbol{x}+\boldsymbol{u}})^2\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,\boldsymbol{u}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}+\boldsymbol{u}-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{s}}\psi_{\eta,\boldsymbol{u}-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}+\boldsymbol{u}-\boldsymbol{s}-\boldsymbol{r}_1}\right)\right\}^2 \\ &= 2\left\{\sum_{\eta_0}\sum_{\boldsymbol{x}}(w_{\eta_0,\boldsymbol{x}+\boldsymbol{u}})^2\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{s}}\psi_{\eta,-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{s}-\boldsymbol{r}_1}\right)\right\}^2, \end{aligned} \tag{3.23}$$

where $\boldsymbol{k} = \boldsymbol{u}-\boldsymbol{v}$. The last simplification of the sums over $\boldsymbol{r}_1$ and $\boldsymbol{s}$ follow from treating $\boldsymbol{u}$ as a constant with respect to $\boldsymbol{r}_1$ and $\boldsymbol{s}$. From the definition of the spectrum, $(w_{\eta_0,\boldsymbol{x}+\boldsymbol{u}})^2 =$

$S_{\eta_0}\left(\frac{\boldsymbol{x}+\boldsymbol{u}}{\boldsymbol{R}}\right)+\mathcal{O}\left(\frac{C_{\eta_0}}{\max\{R,S\}}\right)$ (see Eckley et al. (2009)). Substituting this expression within (3.23), the right-hand side becomes

$$2\left\{\sum_{\eta_0}\sum_{\boldsymbol{x}}\left(S_{\eta_0}\left(\frac{\boldsymbol{x}+\boldsymbol{u}}{\boldsymbol{R}}\right)+\mathcal{O}\left(\frac{C_{\eta_0}}{\max\{R,S\}}\right)\right)\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{s}}\psi_{\eta,-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{s}-\boldsymbol{r}_1}\right)\right\}^2.$$

The two product sums in the expression above are finite and bounded for fixed η due to the compact support of the discrete wavelets. In addition, the constants $\{C_{\eta_0}\}$ satisfy $\sum_{\eta_0} C_{\eta_0} < \infty$ due to the smoothness constraints on the spectrum given in Section 3.2.2. Hence the order term can be taken outside the sums, so that the right-hand side of the equation (3.23) above further becomes

$$2\left\{\sum_{\eta_0}\sum_{\boldsymbol{x}}S_{\eta_0}\left(\frac{\boldsymbol{x}+\boldsymbol{u}}{\boldsymbol{R}}\right)\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{s}}\psi_{\eta,-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{s}-\boldsymbol{r}_1}\right)+\mathcal{O}\left(\frac{C_{\eta_0}}{\max\{R,S\}}\right)\right\}^2.$$

Using the Lipschitz properties of the spectrum as given in equation (3.7) and using the assumed stationary of the spectrum, the above equation simplifies to

$$2\left\{\sum_{\eta_0}S_{\eta_0}\sum_{\boldsymbol{x}}\left(\sum_{\boldsymbol{r}_1}\psi_{\eta,-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{r}_1}\right)\left(\sum_{\boldsymbol{s}}\psi_{\eta,-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{s}-\boldsymbol{r}_1}\right)+\mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)\right\}^2.$$

Rearranging the summations, we have

$$\begin{aligned}
\mathrm{Cov}(I_{\eta,\boldsymbol{u}},I_{\eta,\boldsymbol{v}}) &= 2\left\{\sum_{\eta_0}S_{\eta_0}\sum_{\boldsymbol{r}_1}\sum_{\boldsymbol{s}}\psi_{\eta,-\boldsymbol{k}-\boldsymbol{s}-\boldsymbol{r}_1}\psi_{\eta,-\boldsymbol{r}_1}\sum_{\boldsymbol{x}}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{r}_1}\psi_{\eta_0,\boldsymbol{x}-\boldsymbol{s}-\boldsymbol{r}_1}+\mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)\right\}^2 \\
&= 2\left\{\sum_{\eta_0}S_{\eta_0}\sum_{\boldsymbol{s}}\Psi_{\eta_0}(\boldsymbol{s})\Psi_{\eta}(\boldsymbol{k}+\boldsymbol{s})+\mathcal{O}\left(\frac{1}{\min\{R,S\}}\right)\right\}^2 \\
&= 2\left\{\sum_{\boldsymbol{s}}C(\boldsymbol{s})\Psi_{\eta}(\boldsymbol{s}+\boldsymbol{k})\right\}^2+\mathcal{O}\left(\frac{1}{\min\{R,S\}}\right). \qquad (3.24)
\end{aligned}$$

Here $C(\boldsymbol{s})$ denotes the stationary local covariance (see Definition 10).

The autocorrelation wavelets Ψ are compactly supported and bounded by one (Eckley and Nason, 2005). Hence $\sum_{\tau} C(\boldsymbol{\tau}) < \infty$, and

$$0 \leq 2\left\{\sum_{\boldsymbol{s}}C(\boldsymbol{s})\Psi_{\eta}(\boldsymbol{s}+\boldsymbol{k})\right\}^2 \leq 2\left\{\sum_{\boldsymbol{s}}C(\boldsymbol{s})\right\}^2 < \infty. \qquad (3.25)$$

This finite bound means that the covariance term $\mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}})$ in equation (3.24) is asymptotically constant due to the finite bound in (3.25).

Theorem 3 of Eckley et al. (2010) notes that the covariance between periodogram values is zero when $\|\boldsymbol{k}\| = \|\boldsymbol{u} - \boldsymbol{v}\|$ exceeds the support of the wavelet ψ_η. Thus the covariance sum in (3.22) can be written as

$$\sum_{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{A}(\mathcal{R}^2)} \mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}}) \quad = \sum_{\{(\boldsymbol{u},\boldsymbol{v})\in\mathcal{A}(\mathcal{R}^2)|\boldsymbol{k}\in\mathcal{A}_{\psi_\eta}\}} \mathrm{Cov}(I_{\eta,\boldsymbol{u}}, I_{\eta,\boldsymbol{v}}),$$

for $\boldsymbol{k}$ in some finite subset $\mathcal{A}_{\psi_\eta} \subset \mathcal{R}$, where the set $\mathcal{A}_{\psi_\eta}$ is independent of R and S. Since the expression is a finite sum of asymptotically constant terms, it is itself asymptotically constant. Hence it follows that the limit of the second term in (3.22) is zero as $R, S \to \infty$.

Since both terms of (3.22) are asymptotically zero, $\hat{\beta}_\eta$ is a consistent estimator of β_η for all η. Finally, applying A_J^{-1} to $\hat{\boldsymbol{\beta}}$ thus establishes that $\bar{\boldsymbol{L}}$ is an unbiased and consistent estimator of $\boldsymbol{S}$.

Chapter 4

A multiscale approach to determining the number of textures within an image

Abstract

This article proposes a method to distinguish the number of distinct textures present within a textured image. Our approach employs a recent test of stationarity for locally stationary random fields which permits the comparison of candidate textures. By embedding this test within a quadtree image segmentation procedure we are able to count the number of distinct texture regions within an image. The application of this method is demonstrated on both simulated examples and an example related to material pill.

4.1 Introduction

In this article we propose a method to count the number of textures within an image. The ability to identify the number of textures in an unsupervised manner is useful in several different areas. In particular, the feasibility and efficacy of many segmentation techniques is reliant on the *a priori* knowledge of the number of textures in a given image (Pal and Pal, 1993; Salari and Ling, 1995; He and Chen, 2000). In what follows, we motivate our

technique using data which arises from work with an industrial collaborator. The textured images we consider represent garment materials which may contain regions of different wear, i.e. *pilling* (see Figure 4.1). In some cases there can be multi-pilling across an image. It is of interest to automatically assess the number of different pilled textures present in a given image, particularly since it can be challenging to achieve this by human inspection. Applying this procedure within a segmentation framework allows isolation of regions of local change which is useful for further analysis. It gives an indication of the appearance of the material in question, indicating areas of damage, which is useful in quality control.

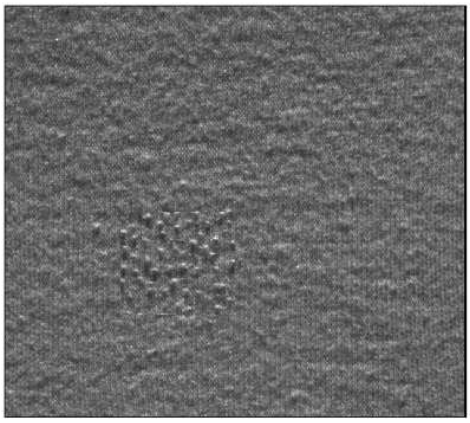


Figure 4.1: A pilled texture mosaic representing fabric of uneven wear, created from textures arising from an industrial application.

We are by no means the first to consider finding the number of textures in an image. Existing work in this area includes techniques based on iterative cluster validation (Coggins and Jain, 1985; Lu et al., 1997); penalised image segmentation methods (Zhang and Modestino, 1990; Bouman and Liu, 1991) and association rules between textures (Rushing et al., 2001). Each of the above methods suffer disadvantages. In particular, the performance of some techniques, such as Coggins and Jain (1985) or Rushing et al. (2001) depend on a subjective threshold or parameter choices. Cost function minimisation methods (Zhang and Modestino, 1990; Bouman and Liu, 1991) are known to be computationally intensive, and can be difficult to minimise in practice due to undesired multiple local minima (Bouman

and Liu, 1991).

The approach which we introduce in this paper uses a recently proposed hypothesis test of image stationarity (spatial homogeneity), by Taylor et al. (2013). We apply this test to *montages* of textures to assess whether two candidate textures are the same. In this context we define a montage to be a concatenation of two images. If the montage is assessed to be stationary, we conclude it comprises of a single texture type. Conversely if the test assesses the montage as being non-stationary, we say that the two textures are distinct. By embedding this test within an image segmentation implementation, our proposed technique can be used to distinguish between textures as they are identified during the segmentation process and thus count the number of textures in an image.

The novelty of our approach to the texture count question lies in the fact that we adopt a formal hypothesis testing framework to distinguish between image textures. Furthermore, it can be performed independently of, or during an image segmentation technique, without affecting the resulting image decomposition.

This article is structured as follows. An overview of the bootstrap test of stationarity of Taylor et al. (2013) is given in Section 4.2. We then discuss our proposed approach to determining the number of textures within an image in Section 4.3. In particular, we first describe how to employ the $\texttt{Bootstat}_{\texttt{LS2W}}$ test to distinguish between textures. This is followed by details of our procedure to produce distinct textures from an image by embedding the $\texttt{Bootstat}_{\texttt{LS2W}}$ test within a quadtree algorithm in Section 4.3.1. Simulated and real examples from an industrial application are given in Section 4.4. Some concluding remarks are made in Section 4.5.

4.2 Establishing stationarity of spatial processes

Before introducing our approach to counting textures within images (Section 4.3), we introduce the test of second-order stationarity proposed by Taylor et al. (2013) which underpins the method we introduce in this article.

We begin by providing a brief review of discrete non-decimated wavelets and locally stationary two-dimensional wavelet processes. Those readers interested in further details

are referred to Vidakovic (1999) or Nason (2008) for comprehensive introductions to wavelet methods.

4.2.1 Discrete wavelets and the LS2W model

Definition 11 *Let ψ be a compactly supported wavelet with corresponding scaling function ϕ, and let $\mathbf{k} = (k_1, k_2)$ be a spatial location, with $k_1, k_2 \in \mathbb{Z}$. Then the two-dimensional discrete wavelets are defined as*

$$\psi_j^\ell = \begin{bmatrix} \psi^\ell_{j,(0,0)} & \cdots & \psi^\ell_{j,(0,L_{j-1})} \\ \vdots & \vdots & \vdots \\ \psi^\ell_{j,(L_{j-1},0)} & \cdots & \psi^\ell_{j,(L_{j-1},L_{j-1})} \end{bmatrix}, \tag{4.1}$$

for scales j and $\ell = h, v$ or d corresponding to horizontal, vertical and diagonal wavelet decomposition directions (Eckley et al., 2010). The elements of the matrices are formed from the products:

$$\begin{aligned} \psi^h_{j,\mathbf{k}} &= \phi_{j,k_1}\psi_{j,k_2} \\ \psi^v_{j,\mathbf{k}} &= \psi_{j,k_1}\phi_{j,k_2} \\ \psi^d_{j,\mathbf{k}} &= \psi_{j,k_1}\psi_{j,k_2}, \end{aligned}$$

where the vectors $\psi_j = \big(\psi_{j,0} \ldots, \psi_{j,L_j-1}\big)$, comprise of entries $\psi_{1,n} = \sum_k g_{n-2k}\delta_{0k} = g_n$ and $\psi_{(j+1),n} = \sum_k h_{n-2k}\psi_{j,k}$ for $n = 0, \ldots, L_{j-1} - 1$. In this notation δ_{0k} is the Kronecker delta which takes the value 1 if $k = 0$ and 0 otherwise. The quantity L_{j-1} is defined according to the wavelet filter associated to ψ as $L_j = (2^j - 1)(N_h - 1) + 1$, where N_h is the number of non-zero coefficients. Similarly the discrete scaling function $\phi_{j,\mathbf{k}} = \phi_{j,k_1}\phi_{j,k_2}$ can also be defined. See Eckley and Nason (2005) for more details on these vectors.

With the building blocks in place we now introduce Eckley et al. (2010)'s definition of a *locally stationary two-dimensional wavelet* (LS2W) process.

Definition 12 *A locally stationary two-dimensional wavelet process, $X_{\boldsymbol{r},\mathcal{R}}$, is defined to be*

$$X_{\boldsymbol{r},\mathcal{R}} = \sum_{\ell} \sum_{j=1}^{\infty} \sum_{\boldsymbol{u}} w^{\ell}_{j,\boldsymbol{u};\mathcal{R}} \psi^{\ell}_{j,\boldsymbol{u}}(\boldsymbol{r}) \xi^{\ell}_{j,\boldsymbol{u}}, \tag{4.2}$$

for spatial locations $\boldsymbol{r} = (r, s) \in \mathcal{R} = \{0, \ldots, R-1\} \times \{0, \ldots, R-1\}$. The two-dimensional discrete non-decimated wavelets $\psi^{\ell}_{j,\boldsymbol{u}}(\boldsymbol{r})$ are defined for scales j, locations $\boldsymbol{u}$ and directions $\ell = h, v$ or d. In equation (4.2), the random variables $\{\xi^{\ell}_{j,\boldsymbol{u}}\}$ are a collection of zero-mean random orthonormal increments and $\{w^{\ell}_{j,\boldsymbol{u};\mathcal{R}}\}$ are process amplitudes bounded by Lipschitz continuous functions W^{ℓ}_{j}, whose behaviour is constrained to vary slowly across locations $\boldsymbol{r}$. It is the control on the amplitudes which enables second-order non-stationary processes to be modelled effectively. For notational convenience, for the remainder of this article we drop the explicit dependence of the model parameters on the spatial lattice $\mathcal{R}$.

In order to analyse the second-order structure of a LS2W process, Eckley et al. (2010) introduced the *local wavelet spectrum* (LWS). The spectrum allows a scale-direction-location decomposition of the variance of a LS2W process. The LWS for rescaled location $\boldsymbol{z} \in (0, 1)^2$, at scale j in direction ℓ is given by $S^{\ell}_{j}(\boldsymbol{z}) = |W^{\ell}_{j}(\boldsymbol{z})|^2$.

Since the local wavelet spectrum is unknown, it has to be estimated. Eckley et al. (2010) propose an estimator for the LWS using a bias-corrected periodogram. To assess the stationarity, or otherwise, of a textured image, the approach proposed in Taylor et al. (2013) uses this periodogram as we describe in the next section.

For more details on the LS2W model and the estimator of the wavelet spectrum, please refer to Eckley et al. (2010).

4.2.2 Testing the hypothesis of stationarity for LS2W processes

With the notation of the LS2W model established we now summarise the $\texttt{Bootstat}_{\texttt{LS2W}}$ test of stationarity for textured images proposed by Taylor et al. (2013). The test forms the basis of both the approach to texture comparison proposed in Section 4.3 and our suggested image splitting technique in Section 4.3.1.

The motivation behind the $\texttt{Bootstat}_{\texttt{LS2W}}$ test stems from the observation that the LS2W process described in equation (4.2) will be second-order stationary if, and only if, its local

wavelet spectrum is constant across all locations for each scale-direction pair. Hence the test for stationarity is formed by inspecting the spectrum of a process for constancy.

More formally, suppose we have a LS2W process $X_{\mathbf{r}}$, with associated wavelet spectrum, $\boldsymbol{S} = \{S_j^\ell(\boldsymbol{z})\}_{j,\ell}$ for locations $\mathbf{z} \in (0,1)^2$. Then the hypothesis for the test

$$H_0 \colon X_{\mathbf{r}} \text{ is stationary}$$
$$H_A \colon X_{\mathbf{r}} \text{ is non-stationary}$$

is equivalent to the test

$$H_0 \colon S_j^\ell(\mathbf{z}) \text{ is constant across } \mathbf{z} \text{ for all } j \text{ and } \ell$$
$$H_A \colon S_j^\ell(\mathbf{z}) \text{ is not constant across } \mathbf{z} \text{ for some } j \text{ and } \ell.$$

This hypothesis motivates the use of a test statistic which measures departures from constancy within each scale-direction pair of the local wavelet spectrum to signify evidence for nonstationarity. Taylor et al. (2013) thus propose using the test statistic

$$T\{S_j^\ell(\boldsymbol{z})\} = (3J)^{-1} \sum_\ell \sum_{j=1}^{J} \int \{S_j^\ell(\boldsymbol{z}) - \bar{S}_j^\ell\}^2 \, d\boldsymbol{z},$$

with $\boldsymbol{z} \in (0,1)^2$ where $\bar{S}_j^\ell = \int S_j^\ell(\boldsymbol{z}) \, d\boldsymbol{z}$ for a particular scale-direction pair. The test statistic is zero if, and only if, the spectrum $S_j^\ell(\boldsymbol{z})$ is constant. In practice, the measure above is based on an estimate of the wavelet spectrum, $\hat{\boldsymbol{S}}(\boldsymbol{z})$. Specifically Taylor et al. (2013) compute

$$T\{\hat{S}_j^\ell(\boldsymbol{z})\} = (3J)^{-1} \sum_\ell \sum_{j=1}^{J} \mathrm{Var}_{\boldsymbol{u}}(\hat{S}_{j,\boldsymbol{u}}^\ell), \tag{4.3}$$

where $\hat{S}_j^\ell$ is given by the bias-corrected periodogram, an estimator of the LWS (see Taylor et al. (2013)).

Typically the distribution of the test statistic (4.3) under the null hypothesis will be unknown. However, since the spectrum characterises the stationarity properties of the observed random field, Taylor et al. (2013) suggest determining the p-value of the hypothesis

$\texttt{Bootstat}_{\texttt{LS2W}}$:

1. Compute the estimate of the LWS for the observed image, $\hat{S}_j^{\ell}(\boldsymbol{z})$.
2. Evaluate T (equation (4.3)) on the observed image, call this value T^{obs}.
3. Compute the pixel average stationary spectrum $\widetilde{S}_j^{\ell}$ by taking the average of spectrum values for each scale and direction.
4. **Iterate** for i in 1 to B bootstraps
   (a) Simulate $X_{\boldsymbol{r}}^{(i)}$ from the stationary LS2W model using squared amplitudes given by $\widetilde{S}_j^{\ell}$ and Gaussian process innovations.
   (b) Compute the test statistic T on the simulated realisation, call this value $T^{(i)}$.
5. Compute the p-value for the test as $p = \frac{1+\#\{\, T^{obs} \leq T^{(i)} \,\}}{B+1}$.

Algorithm 4.1: The bootstrap algorithm for testing the stationarity of locally stationary images.

test by using a parametric bootstrap test, a form of simulation-based testing. This equates to sampling LS2W processes assuming stationarity under the null hypothesis, and comparing the observed test statistic to that of the simulated LS2W processes under stationarity. The stationarity test is called $\texttt{Bootstat}_{\texttt{LS2W}}$. For pseudo-code of the test see Algorithm 4.1.

In the next section, we will show how $\texttt{Bootstat}_{\texttt{LS2W}}$ can be used to count the number of textures within an image.

4.3 Determining the number of textures within an image

In this section we describe our approach to determine the number of textures within an image. Segmentation methods, such as split-and-merge algorithms or the quadtree implementation described in Section 4.3.1, identify statistically homogeneous areas of an image representing textures. Our approach involves using the $\texttt{Bootstat}_{\texttt{LS2W}}$ testing framework on such a set of textured regions to identify whether the regions are statistically different or similar. More specifically, in order to decide whether two textures are similar or not, we propose the following testing procedure:

1. Create a (square) *montage* of the two images.

2. Use the stationarity test of Taylor et al. (2013) to establish whether the montage is

stationary.

3. If the montage is stationary, we then conclude that the two constituent parts represent the same texture type; if the montage is non-stationary then the two textures are considered of different types.

By recursively applying this hypothesis test on montages created from pairs of images from a set of textures, we can partition the set into a number of groups representing distinct textures. Note that it is not necessary to perform the texture discrimination test for *all* possible ${}^{k}C_{2}$ pairs of images in the set; by only comparing an unclassified texture with one candidate from each labelled texture group, all images can be classified with a texture label in a "greedy" fashion. We denote this recursive labelling of textures as $\texttt{TextCount}_{\texttt{LS2W}}$.

By embedding our proposed texture counting procedure $\texttt{TextCount}_{\texttt{LS2W}}$ within an image segmentation method, we are able to progressively segment the image into textures whilst simultaneously count the number of textures within an image. In the next section we give further details of one such segmentation method based on the $\texttt{Bootstat}_{\texttt{LS2W}}$ test.

4.3.1 A quadtree image segmentation implementation using the $Bootstat_{LS2W}$ test

We now outline how to use the $\texttt{Bootstat}_{\texttt{LS2W}}$ test within a quadtree decomposition algorithm in order to partition an image into regions of spatial stationarity (homogeneous second-order structure). This approach is similar in spirit to methods described in Sonka et al. (1999, Chapter 5) or Freixenet et al. (2002). Since texture is loosely defined as a region of an image which is in some sense regular (Petrou and Sevilla, 2006), it is natural to employ statistical measures of homogeneity in this way to capture textural information.

A quadtree algorithm iteratively splits an image into smaller regions based on some statistical criterion. In our case this splitting decision is made according to the stationarity properties of a subimage under analysis: at each stage, a subimage is further divided into four quadrants if it is judged as non-stationary by the $\texttt{Bootstat}_{\texttt{LS2W}}$ test. In practice, the quadtree implementation proposed here continues until all subregions are considered to be stationary, or until the subregions reach a particular minimal dimension. The motivation for

$\texttt{QuadTree}_{\texttt{LS2W}}$:
For an input image X:

- Use the $\texttt{Bootstat}_{\texttt{LS2W}}$ test to assess whether X is second-order stationary. If X is stationary, stop.

If not,

1. Divide the image into four quadrants.
2. For each quadrant, assess its stationarity with the $\texttt{Bootstat}_{\texttt{LS2W}}$ test.
3. For each quadrant assessed as non-stationary, recursively repeat steps 1–2, until the minimum testing region is reached or until all subimages are judged to be stationary.

Algorithm 4.2: The quadtree algorithm for segmenting an image into textured regions using the stationarity test.

this stopping criterion is that the notion of a texture is not statistically or visually meaningful for very small image regions (Chen and Pavlidis, 1979). The quadtree implementation we use for the examples in Section 4.4, $\texttt{QuadTree}_{\texttt{LS2W}}$, is summarised in Algorithm 4.2.

Our technique for determining the number of textures within an image can thus be summarised in two steps:

1. Perform the quadtree partitioning of the image using the test of stationarity described in Section 4.3.1.

2. Use the recursive procedure proposed in Section 4.3 ($\texttt{TextCount}_{\texttt{LS2W}}$) to separate the textured subimages produced by step 1 into texture groups.

Note that Step 2 in the procedure above can be performed either after the quadtree decomposition or as stationary regions are identified *during* the decomposition.

We stress here that the quadtree decomposition algorithm described in this section is not the only decomposition approach one can use, but is included since using the $\texttt{Bootstat}_{\texttt{LS2W}}$ test in an image segmentation technique in this manner has a natural pairing with our proposed $\texttt{TextCount}_{\texttt{LS2W}}$ method outlined in Section 4.3. Indeed, quadtree segmentation methods are merely one example of an image subdivision scheme, and other more sophisticated segmentation methods could be employed (with the $\texttt{Bootstat}_{\texttt{LS2W}}$ stationarity test or otherwise) in combination with our $\texttt{TextCount}_{\texttt{LS2W}}$ method for counting textures.

4.4 Examples

In this section we apply our $\texttt{TextCount}_{\texttt{LS2W}}$ method outlined in Section 4.3 to simulated and real examples displaying a number of different but visually subtle, textures.

4.4.1 Simulated performance of the texture count method

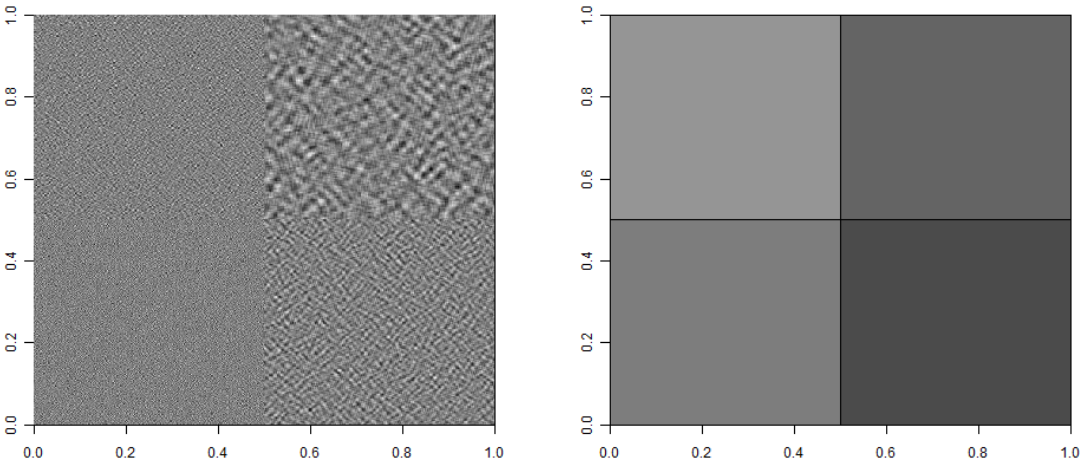
As our first simulated example of a spatial process, we take a LS2W *Haar montage* process introduced in Eckley et al. (2010) and analysed in Taylor et al. (2013). The process can be described as a LS2W process (equation 4.2) with amplitudes given by

$$w^d_{j,[2^J\mathbf{z}]} = \begin{cases} \sigma & \text{if } j = 1 \text{ and } \mathbf{z} \in (0, 1/2)\,\mathrm{x}\,(0, 1/2); \\ \sigma & \text{if } j = 2 \text{ and } \mathbf{z} \in (1/2, 1)\,\mathrm{x}\,(0, 1/2); \\ \sigma & \text{if } j = 3 \text{ and } \mathbf{z} \in (0, 1/2)\,\mathrm{x}\,(1/2, 1); \\ \sigma & \text{if } j = 4 \text{ and } \mathbf{z} \in (1/2, 1)\,\mathrm{x}\,(1/2, 1); \\ 0 & \text{otherwise.} \end{cases}$$

In other words each quadrant of the image has structure only in the diagonal decomposition direction, but with different coarseness in each of the four quadrants (see Figure 4.3(a)).

The second simulated process we consider is a white noise process with a subregion containing another noise process with larger standard deviation. More specifically, the middle portion of the top-left quadrant of the white noise process is assigned a different standard deviation $\sigma = \sigma_1 > 1$. We consider an image of dimension 512×512 with $\sigma_1 = 1.6$ in a subregion of dimension 128×128 (see Figure 4.4(a)). Note that the difference in texture is visually difficult to identify in this example.

Figure 4.3(b) shows the partition of the Haar montage process from the quadtree implementation (Section 4.3.1) where each shade of the partitioning represents a different texture. Our technique correctly identifies that the process has four distinct textures and separates the texture regions exactly. For Process 2, the algorithm is also able to establish the true number of textures present in the image (i.e. two textures represented by the two shades in Figure 4.4(b)). The quadtree partition also locates the subtexture accurately despite being difficult to identify by eye.



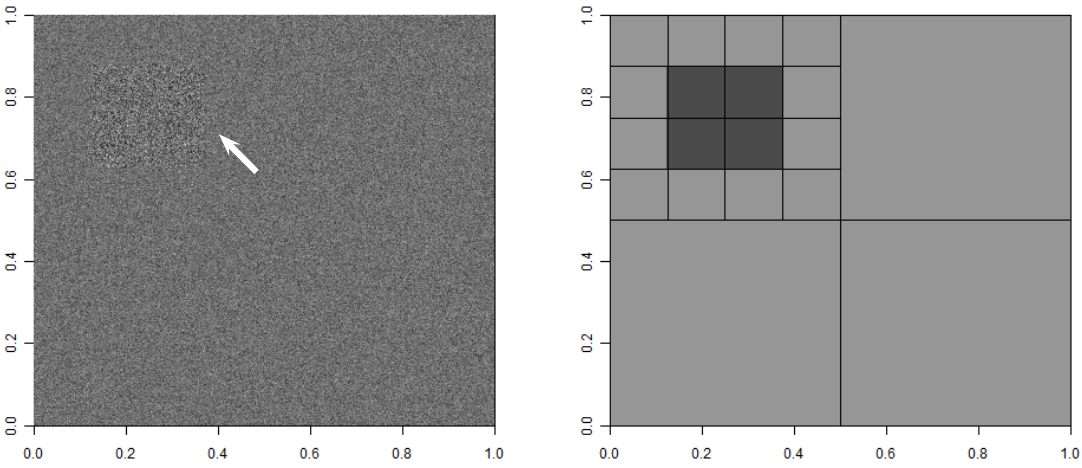
(a) (b)

Figure 4.3: Left: a *Haar montage* LS2W process (Process 1) containing four different textures; right: the quadtree decomposition of the process. Different texture labels assigned to each of the regions of the image are represented by different shades. The subtextures are located and identified as four different textures correctly by our $TextCount_{LS2W}$ hypothesis testing approach.

4.4.2 Fabric Pill example

Our algorithm is perhaps most useful in deciding the number of distinct textures within images where the change is difficult to locate visually. An automatic procedure of this kind is advantageous within industrial settings as it would be more reliable than a human observer, whose image inspection decisions could be affected by variables such as fatigue and lighting.

The example which we consider arises from fabric analysis. *Fabric pilling* occurs when material fibres, through the repeated wearing of a garment, clump together to form "pills" (see e.g. Palmer et al. (2011)). It is often of interest to analyse such fabric images displaying pilling, for example for quality control purposes (Chan and Pang, 2000; Abouelela et al., 2005). In order to show how our method would be useful in these practical settings, we firstly apply our $\texttt{TextCount}_{\texttt{LS2W}}$ method to a collection of pilled fabric images. We then apply the $\texttt{QuadTree}_{\texttt{LS2W}}$ procedure to some fabric mosaics in order to find the number of textured regions within an image.



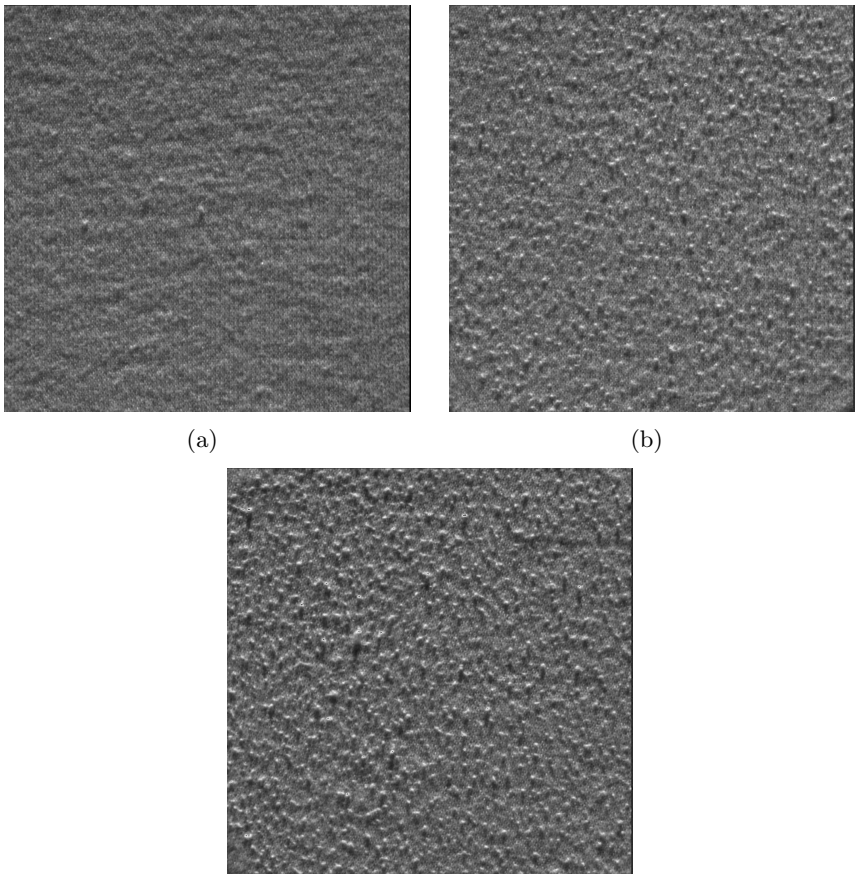
(a) (b)

Figure 4.4: A realisation of Process 2 of size 512×512 with a subregion of size 128×128 of standard deviation 1.6 (indicated by arrow), together with the decomposition of the realisation with the quadtree implementation. Our $TextCount_{LS2W}$ procedure is able to separate the two textures successfully.

To test our $\texttt{TextCount}_{\texttt{LS2W}}$ discrimination procedure, we consider three pilled images, each of varying levels of degradation. These can be seen in Figure 4.5. The images are visually difficult to discriminate between especially the latter two, but the $\texttt{TextCount}_{\texttt{LS2W}}$ method was able to correctly distinguish the three textures using images of dimension $n = 64, 128, 256, 512$.

Next, we consider a more complex but industrially relevant task: identifying pill regions in a multi-pill image. To this end, we construct texture mosaics. A texture mosaic comprises of two or more different texture subimages superimposed into one textural image for analysis. Our texture mosaics were constructed by inserting sections of a pilled fabric texture image into another to represent material with uneven wear. Our goal is to identify the number of textures present within these pilled fabric mosaics.

Our pilled fabric mosaics were constructed as follows. Firstly a subimage of medium pill was inserted into an image of light pilled fabric, i.e. it contains two different textures (Mosaic A, see Figure 4.6(a)). Mosaic B consists of three textures, namely sections of medium *and* heavy pill in lightly pilled material (Figure 4.7(a)).



(a) (b)

(c)

Figure 4.5: Real textured images tested using the $TextCount_{LS2W}$ texture discrimination technique. The textures represent pilled fabric images of increasing degradation.

The $\texttt{QuadTree}_{\texttt{LS2W}}$ and $\texttt{TextCount}_{\texttt{LS2W}}$ procedures were applied to both Mosaic A and Mosaic B. In both cases, the $\texttt{TextCount}_{\texttt{LS2W}}$ procedure is able to judge the correct number of pilled segments in the mosaics, indicated by the differing shades in Figure 4.6(b) and Figure 4.7(b). More specifically, two textures were identified in Mosaic A whereas Mosaic B was judged to contain three different textures. The textured regions are located accurately by the quadtree homogeneity implementation for Mosaic A; the segmentation of Mosaic B provides a reasonable indication of regions of textural change within the image. The approximate nature of the identified regions is attributed to the often simplistic image segmentations provided by the quadtree embedding. Indeed, we reiterate that more complicated region-based segmentation methods could better locate the visually-subtle textures. However, in this article the quadtree method provides a simple and computationally efficient way to

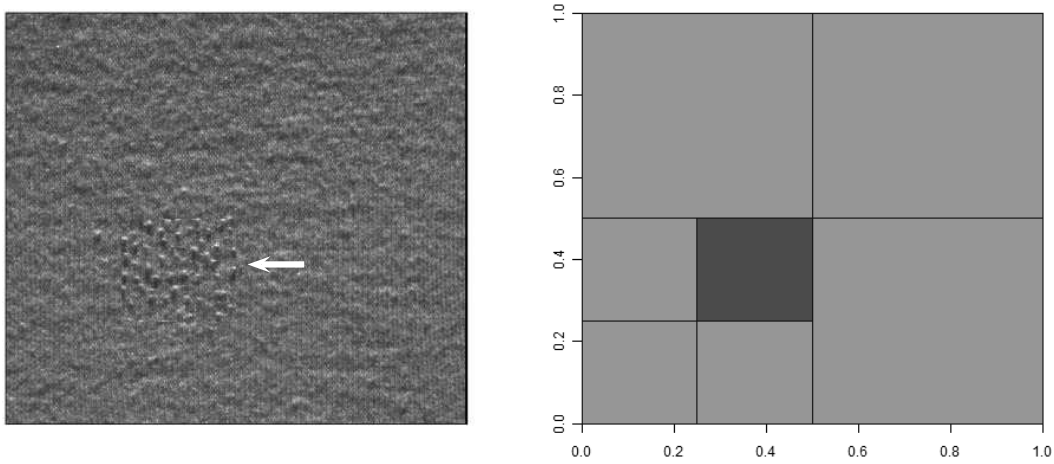


Figure 4.6: Left: Texture Mosaic A consisting of two pilled fabric textures; right: Corresponding quadtree decomposition of the texture mosaic in Figure 4.6(a). The textured subregion is distinguished by the quadtree implementation. The $TextCount_{LS2W}$ procedure correctly determines the two distinct pilling textures present in the mosaic.

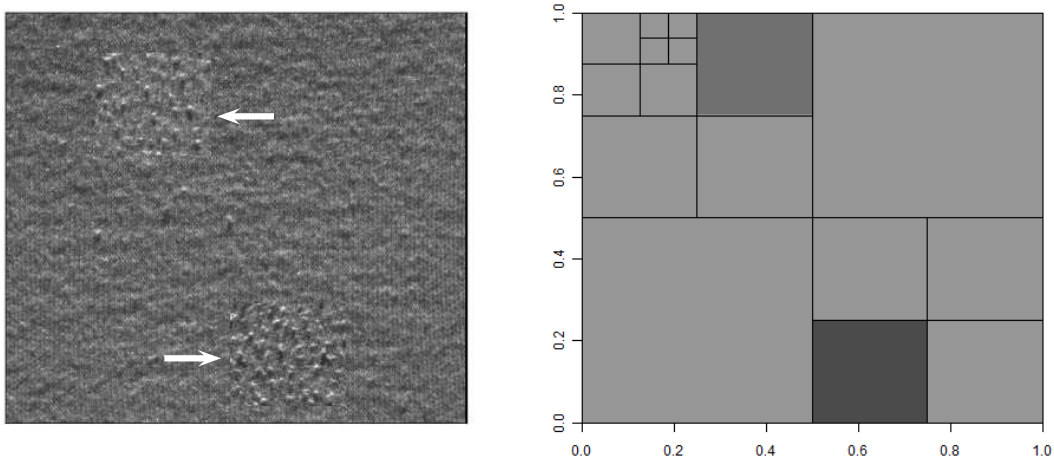


Figure 4.7: Left: Texture Mosaic B, containing three fabric textures, namely light pill with two subregions of medium and heavy pilling (indicated by arrows). Right: Quadtree decomposition of texture Mosaic B. Shades represent different textures identified by our technique. The number of textures is correctly determined; the quadtree algorithm also approximately identifies the location of the textured subregions.

decompose a textured image.

4.5 Concluding remarks

The work presented in this paper addresses the problem of counting the number of textures in an image, considering the specific case where the textures present in the images are visually difficult to differentiate. Our technique adopts a formal statistical modelling approach, employing the stationarity test of Taylor et al. (2013) on *montages* of textures to judge whether two textures are the same. This approach can then be applied recursively to count the number of distinct textures in a set of images, such as regions from a quadtree decomposition. The proposed discrimination technique has been shown to perform well for both simulated and real examples.

We have also proposed using the stationarity test of Taylor et al. (2013) algorithm as a homogeneity measure within region-based image segmentation methods, such as quadtree implementations. Whilst this is not the focus of this article, an attractive by-product of using our texture counting technique in this manner is that it can isolate areas of textural change in images with visually subtle textured regions, such as the pilled texture mosaics considered in this paper.

Chapter 5

Multivariate locally stationary 2D wavelet processes

In this chapter we introduce a statistical model for multivariate two-dimensional locally stationary processes on a regular grid of points. Our motivation for developing this framework arises from a problem in colour texture analysis, specifically the discrimination and classification of colour texture. In the colour texture setting data comes in the form of multivariate images with three channels representing the red, green and blue colour planes. Traditionally such images were analysed by experienced research scientists. However, as suggested by Liang et al. (2012), manual methods of image inspection are subjectively dependent on human vision. In particular they can be skill-orientated, unobjective and inconsistent. The work presented in this chapter therefore provides an automatic mechanism for modelling and estimating the non-stationary structure of multivariate images.

The image set which motivates this work arises from work with an industrial collaborator. It comprises of various hair images with different colourants and preparation processes applied, giving a variety of colour texture properties with subtle changes in structure. It is of considerable interest to be able to accurately classify and discriminate between these hair images.

The approach which we introduce extends the time series work of Sanderson et al. (2010) and Cho and Fryźlewicz (2013) to a 2D locally stationary wavelet model which we

call LS2Wmv. Each channel of this multivariate process takes the form of an LS2W process (see Section 3.2.2), i.e. individual channels can have their own non-stationary behaviour. Our LS2Wmv model seeks to not only represent individual channels as locally stationary, but also represent the local *coherence* between them. Mathematically, coherence is defined as the normalized cross-power spectrum, and it is computed between two simultaneously recorded signals. In our approach the coherence will be decomposed into contributions from different scales and directions for each pair of channels. This measure of dependence, known as the *LS2W coherence structure* is discussed in detail in Section 5.2.

This chapter is structured as follows. We begin by giving an overview of existing multivariate spatial models in Section 5.1. Our proposed modelling framework and definition of LS2W coherence (our measure of cross-channel dependency) is outlined in Section 5.2. A local measure of power, called the local wavelet cross-spectrum, is introduced in Section 5.3. The estimation scheme for the local wavelet cross-spectrum is discussed in Section 5.4. All proofs of results stated in this chapter may be found in Appendix A. Later in Chapter 6, we will apply our multivariate model to the analysis of problems related to images of hair.

5.1 Introduction

In this section we review existing multivariate two-dimensional models. An important statistical property to consider in a multivariate setting is the cross-covariance function. This measures the similarity of two processes and can also be referred to as coherence. The modelling approaches specific to colour texture are discussed in Chapter 6, so are not included in this review. We give a brief consideration to approaches that create classes of stationary spatial processes but our main focus is on current methods that allow for non-stationary behaviour.

Early work in multivariate two-dimensional models focusses on an assumption of stationarity. This includes the work of Mardia (1988), Ver Hoef and Barry (1998), Gaspari and Cohn (1999), Gelfand and Vounatsou (2003), Majumdar and Gelfand (2007) and Gneiting et al. (2010). Whilst each of these approaches has its own merits, fundamentally they were not designed to cope with the non-stationary structure which can exist in images such as

those which occur in environmental science (Kleiber and Nychka, 2012) or texture analysis (Petrou and Sevilla, 2006). Hence these stationary approaches will tend to overlook (or rather average out) the local changes in structure.

Next we review some existing non-stationary multivariate approaches. Gelfand et al. (2004) gave one of the first non-stationary multivariate models, extending the linear model of coregionalization, which is defined in Wackermagel (2003) to a non-stationary setting. Coregionalization is the mutual spatial behaviour between two or more variables within a certain region. Although this is model-based, it only considers dependencies on each plane individually, rather than seeking to describe cross-plane dependence. Also whilst the local nature of spatial neighbourhoods captures the non-stationary behaviour, this model does not give the multiscale decomposition so often required for textured images.

Majumdar et al. (2010) generalize the stationary convolution model for correlated Gaussian processes proposed by Majumdar and Gelfand (2007) into a non-stationary setting. This method is suggested to outperform Gelfand et al. (2004). However, it is computationally expensive since it requires Monto Carlo simulation and the output can be difficult to interpret as it is dependent on the prior. Also like Gelfand et al. (2004), Majumdar et al. (2010) do not consider a multiscale approach.

An approach to modelling non-stationary multivariate processes using Matérn covariances is considered by Kleiber and Nychka (2012). The model includes spatially varying correlation coefficient functions, and allows the strength of between-channel relationships to vary across space. However a restrictive assumption in this model is that the cross-covariance model is symmetric, which is often unrealistic in practice.

Whilst the above papers provide multivariate modelling approaches allowing for non-stationarity, they are all based on multivariate geostatistical data (i.e. data with continuous spatial index). We note that the bulk of published work in this area deals with this general spatial setting. See Wackernagel (2003) for a good overview of this literature.

In this Chapter we consider a multivariate modelling process which allows for changes in covariance structure for data on a regular lattice. Literature on modelling multivariate lattice data is sparse. In recent years Jin et al. (2005), Sain and Cressie (2007), Greco and Trivisano (2009) and various others explored multivariate spatial models for lattice data,

adopting a Bayesian framework for inference of process quantities. The main disadvantage of this framework is in choosing the prior distribution for the model parameters. The approaches above are based on subjective priors which can introduce bias in the estimation procedure. Sain (2009) developed the maximum likelihood estimation procedure for the multivariate Gaussian conditional autoregressive (CAR) model of Sain and Cressie (2007). However, this method can be computationally intensive as the lattice models are embedded in a multinomial model.

In the next section we consider a non-stationary multivariate spatial model with locally varying parameters and consider the local wavelet coherence.

5.2 The multivariate locally stationary wavelet model

We now introduce the multivariate locally stationary wavelet process model; a multiscale approach for modelling non-stationary images. This section extends the work of Eckley et al. (2010) to a multivariate image setting, drawing on recent time series ideas presented by Sanderson et al. (2010) and Cho and Fryźlewicz (2013). The work also introduces various scale-direction-location measures which describe the spectral and cross-spectral behaviour of non-stationary images.

In what follows we use the simplified notation η for a scale-direction pair as introduced in Chapter 3. We start by considering a m-dimensional spatial process, $\mathbf{X}_{\mathbf{r};\mathbf{R}} = [X^{(1)}_{\mathbf{r},\mathbf{R}}; X^{(2)}_{\mathbf{r},\mathbf{R}}; \ldots; X^{(m)}_{\mathbf{r},\mathbf{R}}]'$, where each element is an individual channel (i.e. spatial plane) of the multivariate image. Each individual channel can be defined as a stochastic process defined on a regular lattice. The definition of the LS2Wmv process model is given as follows.

Definition 13 *The multivariate two-dimensional locally stationary wavelet process model (LS2Wmv)* $\{\mathbf{X}_{\mathbf{r};\mathbf{R}}\}$ *, is defined as follows*

$$\begin{aligned} X^{(1)}_{\mathbf{r},\mathbf{R}} &= \sum_{\eta}\sum_{\mathbf{u}} W^{(1)}_{\eta}(\mathbf{u}/\mathbf{R})\psi_{\eta,\mathbf{u}-\mathbf{r}}\xi^{(1)}_{\eta,\mathbf{u}} \\ &\vdots \\ X^{(m)}_{\mathbf{r},\mathbf{R}} &= \sum_{\eta}\sum_{\mathbf{u}} W^{(m)}_{\eta}(\mathbf{u}/\mathbf{R})\psi_{\eta,\mathbf{u}-\mathbf{r}}\xi^{(m)}_{\eta,\mathbf{u}}. \end{aligned} \tag{5.1}$$

Here $\mathbf{r} = (r,s) \in \{0,\ldots,R-1\}\times\{0,\ldots,R-1\}$ *and* $R = 2^k$ *for some* $k \in \mathbb{N}$*. The* $\{\psi_{\eta,\mathbf{u}}\}$ *are a collection of discrete, real valued, compactly supported, non-decimated 2D wavelets as defined in Chapter 3. The* $\{W^{(i)}_{\eta}(\mathbf{u}/\mathbf{R})\}$ *can be thought of as scale-direction-location dependent transfer functions in rescaled space where the index* i *represents a particular channel of the multivariate image. Furthermore the* $\{\xi^{(i)}_{\eta,\mathbf{u}}\}$ *are assumed to be zero mean unit-variance random orthonormal increment sequences. To simplify notation in what follows the explicit dependence on* $\mathbf{R}$ *will be dropped although it is still assumed.*

Modelling assumptions for the LS2Wmv model: In order to develop a principled estimation theory for LS2Wmv processes, we require the following modelling assumptions which help control the degree of non-stationarity of the process. Specifically we assume that each channel of an LS2Wmv process satisfies the assumptions of an LS2W process as given in Chapter 3. Each channel has zero mean and we can adopt the rescaled location approach similar to Dahlhaus (1997), which allows us to collect increasing information and make inferences about the local structure of X^i as $R \to \infty$. In addition we also assume that:

1. For each channel i, the transfer function $W^i_{\eta}(\mathbf{z})$, is Lipschitz continuous with constants $L^{(i)}_{\eta}$ which are uniformly bounded in η and

$$\sum_{\eta} 2^{2j(\eta)} L^{(i)}_{\eta} < \infty. \tag{5.2}$$

2. In addition the $\{\xi^{(i)}_{\eta,\mathbf{u}}\}$ have the following properties,

$$\begin{aligned} \mathrm{Cov}[\xi^{i}_{\eta,\mathbf{k}}\xi^{i}_{\eta_1,\boldsymbol{m}}] &= \delta_{\eta,\eta_1}\delta_{\mathbf{k},\boldsymbol{m}}, & (5.3)\\ \mathrm{Cov}[\xi^{p}_{\eta,\mathbf{k}}\xi^{q}_{\eta_1,\boldsymbol{m}}] &= \delta_{\eta,\eta_1}\delta_{\mathbf{k},\boldsymbol{m}}\rho^{p,q}_{\eta}(\mathbf{k}/\mathbf{R}), & (5.4) \end{aligned}$$

where δ_{η,η_1} is the Kronecker delta.

As a consequence of extending to a multivariate setting, we also have to consider the possible coherence structure between each pair (p, q), of image channels where we assume the images are in-phase. We define this measure as the LS2W coherence (the spatial equivalent of correlation) given by $\rho^{p,q}_{\eta}(\mathbf{u}/\mathbf{R})$ (see Definition 14). The quantity $\rho^{p,q}_{\eta}(\mathbf{z})$ is a direct measure of the linear dependence between the innovation sequences of two channels at scale-direction η (as shown above in the second modelling assumption). In a colour setting this quantity explores the relationship between the colour planes. The LS2W coherence of a pair of channels p and q can be represented in terms of the locally stationary wavelet cross-spectrum (LWCS) of two channels (see Section 5.3.1) from an LS2Wmv process. In order to estimate this relationship between channels we use a pairwise approach, this is discussed further in Section 5.4.

Definition 14 *The local wavelet coherence (LWC) for two channels p and q of an LS2Wmv process is defined to be:*

$$\rho^{p,q}_{\eta}(\mathbf{z}) = \frac{S^{p,q}_{\eta}(\mathbf{z})}{\sqrt{S^{(p)}_{\eta}(\mathbf{z})S^{(q)}_{\eta}(\mathbf{z})}}, \qquad (5.5)$$

where the individual local wavelet spectrum (LWS) of each channel $S^{(p)}_{\eta}(\mathbf{z})$ (see Chapter 3) and the local wavelet cross-spectrum (LWCS) $S^{(p)}_{\eta}(\mathbf{z})S^{(q)}_{\eta}(\mathbf{z})$ (see Section 5.3.1) together provide a normalised measure of the relationship between two channels.

Modelling assumptions for the LS2W coherence:

1. The LS2W coherence of each scale-direction pair is also assumed to be Lipschitz continuous with Lipschitz constants, $R^{(p,q)}_{\eta}$ satisfying

$$\sum_{\eta} 2^{2j(\eta)} R^{(p,q)}_{\eta} < \infty. \qquad (5.6)$$

2. The value of the coherence determines the level of dependence between two channels with +1 and -1 indicating a postive and negative dependence respectively and a value of zero showing no dependence at a given scale-direction η and rescaled location $\mathbf{z}$.

5.3 Measuring local power

Recall from Section 3.2.3 that the spectral structure of a signal or image can be used to describe its second-order structure. For the situation which we consider, with multiple channels, this leads to the consideration of spatially localised wavelet spectra (i.e. the structure within a single channel) and cross-spectra (i.e. the structure across channels). The local wavelet spectra (LWS) provides a measure of the local contribution to the variance of each channel and the local wavelet cross-spectra the cross-covariance between channels at a particular scale, direction and location. Below we formally define each of these quantities.

5.3.1 The local wavelet cross-spectrum

We begin by defining the local wavelet cross-spectrum which provides a measure of dependence between two channels of a LS2Wmv process at a specific (rescaled) location $\mathbf{z}$, and scale-direction η.

Definition 15 *Let $X_{\mathbf{r}}^{(p)}$ and $X_{\mathbf{r}}^{(q)}$ be two channels of a LS2Wmv process with amplitude functions $W_{\eta}^{(p)}(\mathbf{z})$ and $W_{\eta}^{(q)}(\mathbf{z})$ respectively. Then the local wavelet cross-spectrum (LWCS) of the two channels $X_{\mathbf{r}}^{(p)}$ and $X_{\mathbf{r}}^{(q)}$ is then given by*

$$S_{\eta}^{p,q}(\mathbf{z}) = W_{\eta}^{(p)}(\mathbf{z})W_{\eta}^{(q)}(\mathbf{z})\rho_{\eta}^{p,q}(\mathbf{z}), \tag{5.7}$$

for $\mathbf{z} \in (0,1)^2$ and scale-direction η.

In other words, the LWCS is a scale-location-direction decomposition of the covariance structure between channels p and q. In the case where $p = q$ we obtain the auto-spectra as defined in equation (5.8). Since $\rho_{\eta}^{p,p}(\mathbf{z}) = 1$ we have for all $\mathbf{z} \in (0,1)^2$ and scale-direction η,

$$S_{\eta}^{p,p}(\mathbf{z}) = |W_{\eta}^{(p)}(\mathbf{z})|^2 \quad . \tag{5.8}$$

Having considered the Lipschitz properties of $W_\eta^i(\mathbf{z})$ and $\rho_\eta^{p,q}(\mathbf{z})$ in the previous section we now introduce the following lemma which provides a result of the properties of products of Lipschitz continuous variables. This is useful in establishing the asymptotic properties of the LWCS estimator given in Section 5.4.

Lemma 1 *Suppose $W_\eta^i(\mathbf{z})$ and $\rho_\eta^{p,q}(\mathbf{z})$ are Lipschitz with constants $L_\eta^{(i)}$ and $R_\eta^{(p,q)}$ respectively. Furthermore, assume that there exists a positive constant, C^*, such that for all η and $\mathbf{z}$, $|W_\eta^i(\mathbf{z})| \leq C^*$. Then by the property of products of Lipschitz continuous variables, denoting $B_\eta^{(p,q)} = max(L_\eta^{(p)}, L_\eta^{(q)}, R_\eta^{(p,q)})$, we have*

$$\left|W_\eta^{(p)}(\mathbf{u}/\mathbf{R})W_\eta^{(q)}(\mathbf{u}/\mathbf{R})\rho_\eta^{p,q}(\mathbf{u}/\mathbf{R}) - W_\eta^{(p)}(\mathbf{r}/\mathbf{R})W_\eta^{(q)}(\mathbf{r}/\mathbf{R})\rho_\eta^{p,q}(\mathbf{r}/\mathbf{R})\right| \leq (R^{-1})CB_\eta^{(p,q)}||\mathbf{u}-\mathbf{r}||. \tag{5.9}$$

In other words,

$$\left|S_\eta^{(p,q)}(\mathbf{u}/\mathbf{R}) - S_\eta^{(p,q)}(\mathbf{r}/\mathbf{R})\right| \leq (R^{-1})CB_\eta^{(p,q)}||\mathbf{u}-\mathbf{r}||, \tag{5.10}$$

the difference in the cross-spectra between two spatial points is bounded by the distance between the points.

Proof. See Appendix A. ∎

As the LWCS for each pair of channels is based upon the redundant NDWT (see Section 2.3.1), it is important that we establish its uniqueness given the corresponding LS2Wmv process.

Theorem 1 *The LWCS for each p and q is uniquely defined given the corresponding LS2Wmv process.*

Proof. See Appendix A. ∎

In the next section we introduce examples of LS2Wmv processes with given coherence and spectral structure as defined in equations (5.5) and (5.8).

5.3.2 Examples of LS2Wmv processes

To help demonstrate our multivariate model and LS2W coherence measure we present two examples of data simulated using LS2Wmv processes which exhibit differing second-order structure. We can construct stationary and non-stationary processes which possess features at multiple scales and directions. Throughout this simulation study we use Haar wavelets in the model (Definition 13) though other Daubechies wavelets can naturally be used. This is an interesting avenue for future work.

Example 1

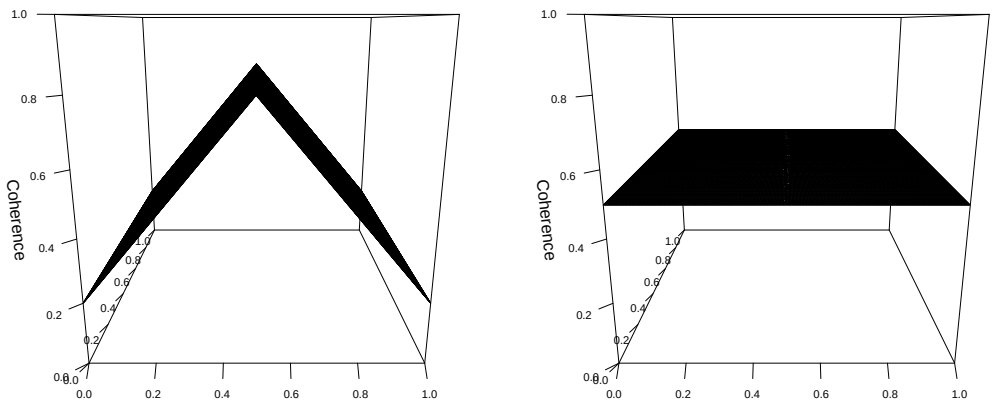
We begin by extending a bivariate time series example introduced by Sanderson et al. (2010) to the image setting. The data is simulated from a multivariate image of dimension 2^9 x 2^9 with two channels and a known coherence structure (equation (5.5)), which varies between the different scales. For even scales each channel has a stationary structure with $\rho_{\eta}^{p,q}(\mathbf{u}/\mathbf{R}) = 0.5$, however for odd scales we assume a non-stationary structure that forms an "inverted v" along the horizontal axis as shown in Figure 5.1(a). More explicitly, the coherence is described by the equation:

$$\rho_{\eta}^{p,q}(\mathbf{u}/\mathbf{R}) = \begin{cases} \tau + \frac{(\delta-\tau)}{a} u_1 & u_1 \in (0, a) \\ \delta - \frac{(\delta-\tau)}{1-a}(u_1 - a) & u_1 \in (a, 1), \end{cases} \tag{5.11}$$

for $\mathbf{u} = (u_1, u_2) \in (0, 1)^2$. Here τ is the minimum value of the coherence, δ is the maximum value and a describes the proportion of the u_1 axis covered by the first (increasing) linear function. Higher values of a mean a more gradual increase to the maximum value. In this example we have $\tau = 0.2$, $\delta = 0.8$ and $a = 0.5$. Finally, using equation (5.8) the LWS is set to be $S_{\eta}^{p,p}(\mathbf{u}/\mathbf{R}) = 2^{-j}$ where $j = 1, 2, \ldots, J$ for all directions and locations for each channel.

Example 2

Our second example demonstrates the potential of our method on a trivariate case, reminicent of the colour image setting. Suppose that for all three channels in this example the spec-



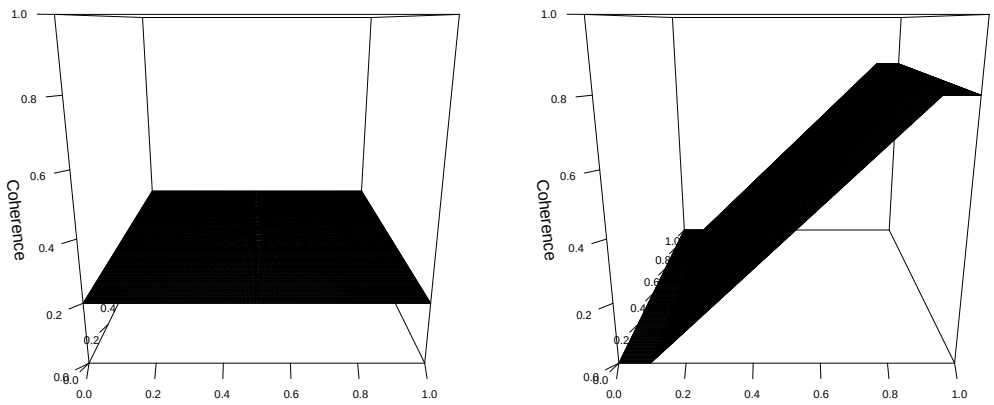
(a) (b)

Figure 5.1: True coherence for Example 1 (J=9). (a) Represents the true coherence for odd scales, (b) even scales.

trum is constant at all scales and directions, i.e. $S_\eta^{1,1}(\mathbf{u}/\mathbf{R}) = S_\eta^{2,2}(\mathbf{u}/\mathbf{R}) = S_\eta^{3,3}(\mathbf{u}/\mathbf{R}) = 2$. Further assume that the coherence between the first and second channel is $\rho_\eta^{p,q}(\mathbf{u}/\mathbf{R}) = 0.2$ and the coherence between the second and third channel increases linearly along the horizontal axis as demonstrated in Figure 5.2. The equation below gives a mathematical representation of this coherence:

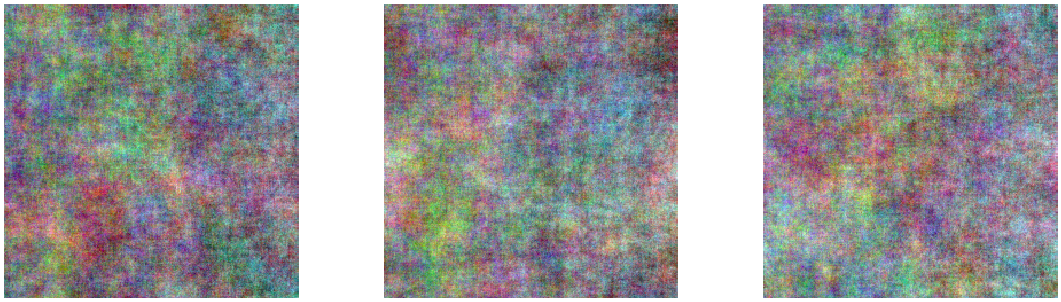
$$\rho_\eta^{p,q}(\mathbf{u}/\mathbf{R}) = \left\{ \tau + \frac{(\delta-\tau)}{1+\exp(-10a(2u_1-1))} \quad u_1 \in (0,1), \right. \tag{5.12}$$

for $\mathbf{u} = (u_1, u_2) \in (0,1)^2$. Equation (5.12) generates a non-stationary covariance structure where τ is the minimum value of the coherence, δ is the maximum value and a changes the shape of the u_1 axis. In other words the coherence is constructed such that it varies spatially across an image. In this example we have $\tau = 0$, $\delta = 0.8$ and $a = 0.10$. Again we simulate each channel to be of dimension 2^9 x 2^9 and realisations of this process in the RGB representation are given in Figure 5.3. In the next section we consider how, given a set of channels like these, we can estimate the cross-spectrum and coherence for each pair.



(a) (b)

Figure 5.2: True coherence of Example 2 (same for all scales). (a) flat coherence between channels 1 and 2, (b) coherence between channels 2 and 3.



(a) (b) (c)

Figure 5.3: Three realisations of a trivariate LS2Wmv process (in RGB colour space): Example 2.

5.4 Estimation of spectral properties for LS2Wmv processes

Recall from Section 5.2 that in order to analyse the properties of a LS2Wmv process and find the LS2W coherence for each pair of channels we must be able to estimate the auto- and cross-spectra. The following sections consider the estimation and consistency of the cross-spectra for each pair of channels p and q. We also consider how to estimate the coherence together with practical problems that can occur in using this estimate. We then return to

our simulated examples in order to show the accuracy in our estimation procedure.

5.4.1 Estimating the local wavelet cross-spectra

Recall from standard Fourier theory that the estimate of the spectrum is the square of the transformed process coefficients. In a similar fashion we define the (raw) cross-spectra as the product of empirical wavelet coefficients from individual channels. To formalise this the following extends the definition of wavelet coefficients from Eckley et al. (2010) to LS2Wmv processes.

Definition 16 *For each channel $X_{\mathbf{r}}^{(i)}$ of a LS2Wmv process the empirical wavelet coefficients are defined to be*

$$d_{\eta,\mathbf{u}}^{(i)} = \sum_{\mathbf{r}} X_{\mathbf{r}}^{(i)} \psi_{\eta,\mathbf{u}-\mathbf{r}}. \tag{5.13}$$

Hence using these coefficients we can define an estimator of the cross-spectrum of each pair of channels as the localized wavelet cross-periodogram, $I_{\eta,\mathbf{u}}^{(p,q)}$.

Definition 17 *The local (raw) wavelet cross-periodogram for two channels $X_{\mathbf{r}}^{(p)}$ and $X_{\mathbf{r}}^{(q)}$ of a LS2Wmv process is given by $I_{\eta,\mathbf{u}}^{(p,q)} = d_{\eta,\mathbf{u}}^{(p)} d_{\eta,\mathbf{u}}^{(q)}$.*

Next we consider the statistical properties of the wavelet cross-periodogram, as an estimator of the LWCS. To begin we consider the expectation and variance of the (raw) cross-periodogram.

Theorem 2 *Let $X_{\mathbf{r}}^{(p)}$ and $X_{\mathbf{r}}^{(q)}$ be two channels, p and q, of an LS2Wmv process. Then asymptotically, the expectation of the (raw) cross-periodogram between these two channels, $I_{\eta,\mathbf{s}}^{(p,q)}$, is given by*

$$\begin{aligned} \mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) &= \sum_{\eta_1} W_{\eta_1}^{(p)}(\mathbf{s}/\mathbf{R}) W_{\eta_1}^{(q)}(\mathbf{s}/\mathbf{R}) \rho_{\eta_1}^{p,q}(\mathbf{s}/\mathbf{R}) A_{\eta_1,\eta} + O\left(\frac{1}{R}\right), & (5.14) \\ &= \sum_{\eta_1} S_{\eta_1}^{p,q}(\mathbf{s}/\mathbf{R}) A_{\eta_1,\eta} + O\left(\frac{1}{R}\right), & (5.15) \end{aligned}$$

where $\mathbf{R} = (R, R)$ *and* $R = 2^k$ *for* $k \in \mathbb{N}$. *Similarly, the variance is given by*

$$Var(I_{\eta,\mathbf{s}}^{(p,q)}) = \sum_{\eta_1} S_{\eta_1}^{(p)}(\mathbf{s}/\mathbf{R}) A_{\eta_1,\eta} \sum_{\eta_1} S_{\eta_1}^{(q)}(\mathbf{s}/\mathbf{R}) A_{\eta_1,\eta} + \left(\sum_{\eta_1} S_{\eta_1}^{p,q}(\mathbf{s}/\mathbf{R}) A_{\eta_1,\eta} \right)^2 + O\left(\frac{2^{2j(\eta_1)}}{R} \right). \tag{5.16}$$

Here $j(\eta_1)$ *simply refers to scale for each direction.*

Proof. See Appendix. ∎

Remark. Note that if $p = q$ we have the standard expectation and variance of an LS2W periodogram (See Eckley et al. (2010)).

As demonstrated in Theorem 2, the local raw cross-periodogram of each pair of channels is a biased estimator of the cross-spectrum. The bias is described by the inner product matrix A_J as defined in equation (3.12). A transformation of the spectra by A_J^{-1} results in an asymptotically unbiased estimate of each wavelet cross-spectrum. However, as in other settings the variance of the raw estimator does not decrease as the dimension of the channel tends to infinity, i.e. it is not a consistent estimator. Therefore as with the other spectral estimators we need to smooth to achieve consistency (Nason et al., 2000; Eckley et al., 2010). We choose to use the the well-known Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964) for smoothing the cross-spectra, which is given by the weighted average

$$\widetilde{I}_{\eta}^{(p,q)}(\mathbf{s}) = \sum_{u \in \mathcal{R}} w_{\mathbf{u}} I_{\eta}^{(p,q)}(\mathbf{u}), \tag{5.17}$$

where the lattice weights are given by $w_{\mathbf{u}} = \frac{K_h(||\mathbf{s}-\mathbf{u}||}{\sum_{\mathbf{u}} K_h(||\mathbf{s}-\mathbf{u}||)}$. In this expression $K_h(\cdot)$ is a (two-dimensional) bounded kernel function on $\mathbb{R}^2$ with bandwidth h. In other words, $K : \mathbb{R}^2 \to \mathbb{R}$ satisfies (i) $\int K(\mathbf{x}) d\mathbf{x} < \infty$ and (ii) $|K(\mathbf{x})| < K_M < \infty$ $\forall \mathbf{x}$.

In order to find the expectation of the smoothed-corrected estimator, we first provide the intermediate result for the smoothed cross-periodogram. As discussed in Nason et al. (2000) in the time series context and Eckley et al. (2010) in the spatial context, assuming the innovations $\xi_{\eta,\mathbf{u}}$ are Gaussian, squaring each element of the wavelet periodogram gives a χ^2-distribution. In our case, finding the product between elements of each channel would also

result in a χ^2-distribution. However the bias correction of the cross-periodogram estimate using A_J^{-1} leads to a complex correlated distribution for the local wavelet cross-periodogram. Consequently calculating the asymptotic variance behaviour is difficult. Thus we suggest firstly smoothing the χ^2-distributed variables prior to correction by A_J^{-1}. The asymptotic expectation of the resulting smoothed cross-periodogram, $\tilde{I}$ is given by Theorem 3. The result holds for any pair of channels p and q of the multivariate image.

Theorem 3 *The (asymptotic) expectation of* $\widetilde{I}_{\eta,\mathbf{s}}^{(p,q)}$ *is given by,*

$$\mathbb{E}(\widetilde{I}_{\eta,\mathbf{s}}^{(p,q)}) = \sum_{\eta_1} S_{\eta_1}^{p,q}(\mathbf{s}/\mathbf{R})\, A_{\eta_1,\eta} + \left(\frac{1}{2\lfloor h \rfloor + 1}\right) O\left(\frac{1}{R}\right). \tag{5.18}$$

Proof. See Appendix A. ■

Defining the smoothed corrected cross-periodogram $\widehat{I}$ as follows:

$$\widehat{I}_{\eta}^{(p,q)}(\mathbf{s}) = \sum_{\eta_1} A_{\eta,\eta_1}^{-1} \widetilde{I}_{\eta_1}^{(p,q)}(\mathbf{s}), \tag{5.19}$$

we obtain an asymptotically unbiased estimator of the LWS:

$$\mathbb{E}(\widehat{I}_{\eta,\mathbf{s}}^{(p,q)}) = S_{\eta_2}^{p,q}(\mathbf{s}/\mathbf{R}) + \left(\frac{1}{2\lfloor h \rfloor + 1}\right) O\left(\frac{1}{R}\right). \tag{5.20}$$

This is a straightforward consequence of Proposition 4 of Nason et al. (2000). Now we have an unbiased estimator we wish to show that its variance vanishes asymptotically.

Proposition 1 *The variance of the smoothed cross-periodogram is asymptotically consistent:*

$$Var(\widetilde{I}_{\eta,\mathbf{s}}^{(p,q)}) \to 0, \tag{5.21}$$

as $h, R \to \infty$ *with* $(h/R) \to 0$.

Proof. See Appendix A. ■

The corrected and smoothed periodogram thus provides a consistent and asymptotically unbiased estimator of the LWCS for each pair of channels.

5.4.2 Estimation of the coherence

In order to estimate the coherence we consider the estimated spectral quantities as discussed in detail in Section 5.4.1, namely the local wavelet cross-spectrum (LWCS), $\widehat{S}_{\eta}^{p,q}(\mathbf{s}/\mathbf{R})$, and local wavelet spectra LWS, $\widehat{S}_{\eta}^{(i)}(\mathbf{s}/\mathbf{R})$. We define the estimator of the locally stationary wavelet coherence as:

$$\widehat{\rho}_{\eta}^{(p,q)}(\mathbf{s}/\mathbf{R}) = \frac{\widehat{S}_{\eta}^{(p,q)}(\mathbf{s}/\mathbf{R})}{(\widehat{S}_{\eta}^{(p)}(\mathbf{s}/\mathbf{R})\hat{S}_{\eta}^{(q)}(\mathbf{s}/\mathbf{R}))^{1/2}}. \tag{5.22}$$

Using Slutsky's theorem (Slutsky, 1925) it can be shown that this estimator converges in probability to equation (5.5).

In practice one can encounter problems when estimating the LS2W coherence (equation (5.22)) for each pair of channels, in that values of the auto-spectra and cross-spectra close to or below zero, lead to instabilities in the estimator. This problem was also encountered by Sanderson et al. (2010) in the bivariate time series setting. Below we propose an approach to resolve this issue.

5.4.3 Practical considerations: Regularisation

On some occasions after smoothing and correcting the periodogram, the resulting spectral array may contain negative elements. In practice this is due to the bias correction step, which involves multiplying the estimator by A_J^{-1} (equation (3.12)) which itself can contain negative values. This can result in an estimated coherence structure which will not be valid as some coherence values will not be in the interval [-1,1]. One approach to resolve this is to regularise the spectral matrix to ensure it is positive semi-definite. Various different regularisation approaches exist such as those proposed by Andrey Tikhonov (Tikhonov and Arsenin, 1977; Tikhonov et al., 1987). Of these we adopt the method of Higham (1988), which we outline briefly below. The advantage of this approach is that it only modifies the spectral estimates if necessary, therefore making the regularisation procedure computationally efficient.

In order to regularise a spectral estimate $\hat{I}_{\eta}(\mathbf{s})$ we take its eigendecomposition at fixed

η and $\mathbf{s}$. In other words we write,

$$\widehat{I}_{\eta}(\mathbf{s}) = \mathbf{C}_{\eta,\mathbf{s}}\mathbf{M}_{\eta,\mathbf{s}}\mathbf{C}'_{\eta,\mathbf{s}}, \tag{5.23}$$

where $\mathbf{C}_{\eta,\mathbf{s}}$ is the matrix whose columns are the set of eigenvectors of $\widehat{I}_{\eta}(\mathbf{s})$ and $\mathbf{M}_{\eta,\mathbf{s}}$ is the diagonal matrix whose elements are the set of eigenvalues of $\widehat{I}_{\eta}(\mathbf{s})$. To ensure $\widehat{I}_{\eta}(\mathbf{s})$ has no negative eigenvalues we define the matrix $\mathbf{M}^{+}_{\eta,\mathbf{s}}$ which is the diagonal matrix $\mathbf{M}_{\eta,\mathbf{s}}$ with any negative eigenvalues set to zero. The regularised spectral matrix $\widehat{I}^{+}_{\eta}(\mathbf{s})$ is then calculated as,

$$\widehat{I}^{+}_{\eta}(\mathbf{s}) = \mathbf{C}_{\eta,\mathbf{s}}\mathbf{M}^{+}_{\eta,\mathbf{s}}\mathbf{C}'_{\eta,\mathbf{s}}. \tag{5.24}$$

This procedure essentially finds a positive semi-definite spectral matrix closest to $\widehat{I}_{\eta}(\mathbf{s})$ (in Frobenius norm), see Higham (1988) for more details. To regularise the spectral array $\widehat{I}$, we repeat this process for all η and $\mathbf{s}$.

We now return to the simulated examples in Section 5.3.2, applying the estimation method discussed in this section to estimate the LS2W coherence for all pairs of channels.

5.5 Examples of LS2Wmv processes: Estimating the coherence

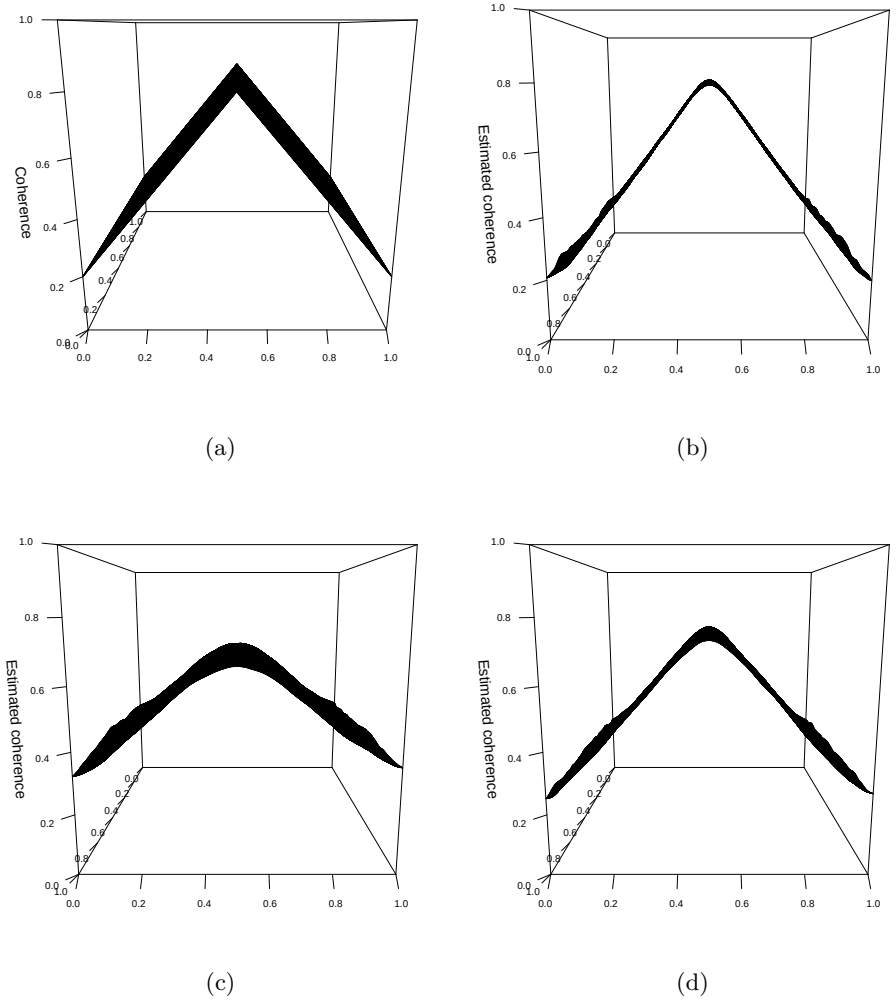
Example 1

Recall that our first example of Section 5.3.2 considers a bivariate case of a LS2Wmv process of lattice dimension 512×512, with a constant spectrum at each scale and direction. Further we assumed a stationary coherence for the even scales of each channel and a non-stationary coherence for the odd scales of each channel.

The estimated coherence structure from an average of 100 realisations of the process is shown in Figure 5.4 at the finest scale ($j = 1$) in each direction. The "inverted" v shape is clearly captured in all directions. However, in the vertical and diagonal cases the minimum and maximum values of the coherence are more accurately estimated than in the horizontal case. Figure 5.5 shows the average estimates at $j = 2$. Whilst these are, on the whole, very

good there is some slight deviation from the true flat structure. This suggests that there has been some power leakage despite the correction.

The mean square error (MSE) of the estimate across all scales, directions and locations is only 0.0363.



(a) (b)

(c) (d)

Figure 5.4: Example 1: Coherence at the finest scale (j=1), (a) represents the true coherence whilst (b), (c) and (d) are the estimated coherences representing the mean of 100 simulations in the vertical, horizontal and diagonal directions respectively.

Example 2

Our second example from Section 5.3.2 considers a trivariate case of lattice dimension 512×512, with a constant spectrum for all scales and directions, a stationary coherence between

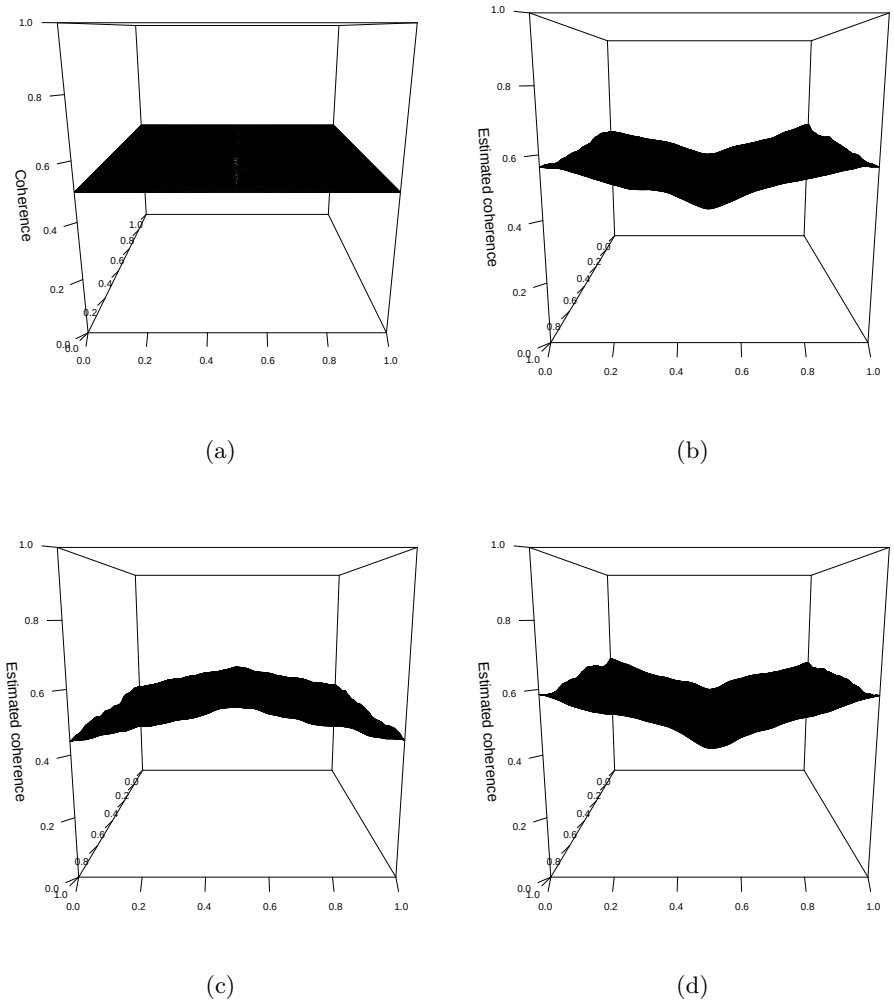
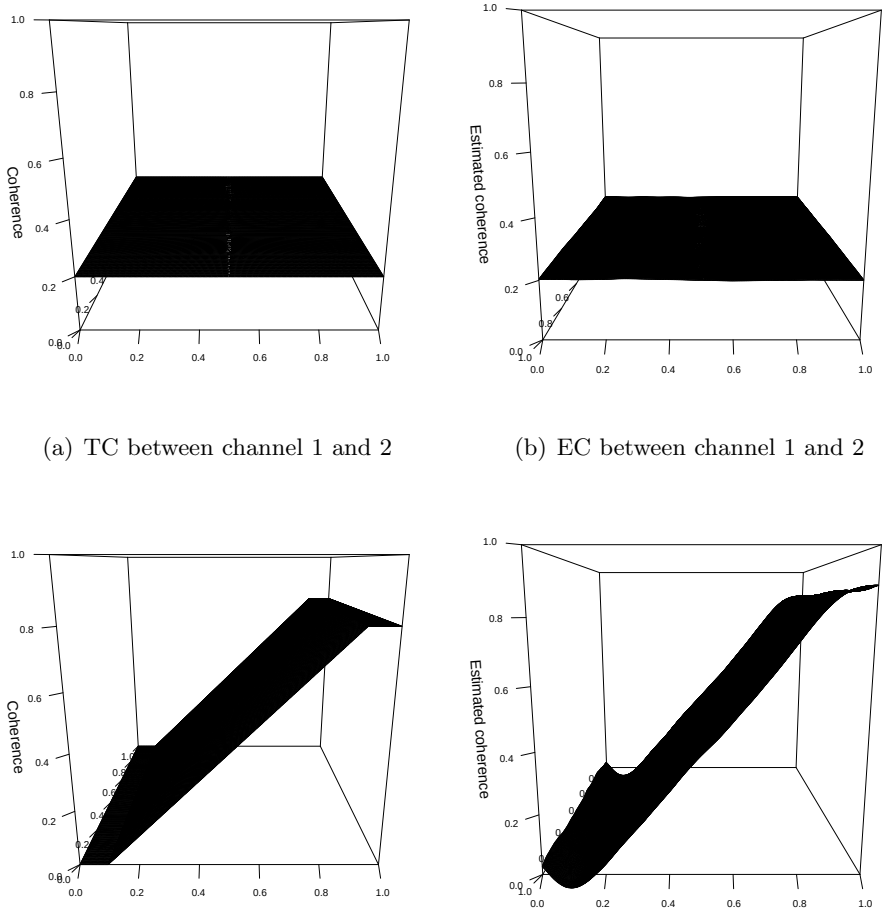


Figure 5.5: Example 1: Coherence at the second finest scale (j=2), (a) represents the true coherence whilst (b), (c) and (d) are the estimated coherences representing the mean of 100 simulations in the vertical, horizontal and diagonal directions respectively.

the first two channels and a non-stationary coherence between the second and third channels.

Again we consider the average estimated structure of 100 simulations. Figure 5.6 shows the coherence at the finest scale in the vertical direction; we observe similar results for the other directions. As we can see the estimate for the coherences between the first two channels is very good. We also obtain a reasonable estimate for the more difficult coherence structure between the second and third channel, the estimate capturing the general structure. The MSE of the estimate for this example is 0.0748. In this example the flat coherence is more accurately estimated. A reason for this could be that in this case the coherence for a fixed

pair of channels is the same across all scales whereas in example 1 it varies between scales, making it more difficult to estimate. Both these examples show that our method is able



(a) TC between channel 1 and 2

(b) EC between channel 1 and 2

(c) TC between channel 2 and 3

(d) EC between channel 2 and 3

Figure 5.6: Results comparing the true (TC) and estimated (EC) coherence at the finest scale: Subplot a) and b) corresponds to the coherence between the first and second channel, c) and d) the second and third channel.

to reasonably capture the changes in coherence structure between multiple channels. We note here that whilst we examine the average of 100 realisations, the behaviour of a single realisation is consistent with the average.

5.6 Capturing the cross-covariance structure

Previous work on locally stationary processes (Dahlhaus, 1997; Nason et al., 2000; Eckley et al., 2010), has shown a link between the local spectrum and the local covariance. Therefore we conclude the methodological development in this chapter by introducing the local cross-covariance and consider whether there is a similar link between this and the local cross-spectrum defined in Section 5.3.1.

We begin by introducing the local cross-covariance (LCCV) highlighting its various properties.

Definition 18 *Let $c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau})$ denote the local cross-covariance between channels p and q from a LS2Wmv process at lag $\boldsymbol{\tau} \in \mathbb{Z}^2$. We define this function in terms of the local cross-spectrum by,*

$$c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) = \sum_{\eta} S_{\eta}^{p,q}(\mathbf{z}) \Psi_{\eta}(\boldsymbol{\tau}), \tag{5.25}$$

for $\boldsymbol{\tau} \in \mathbb{Z}^2$. Here $\Psi_{\eta}(\boldsymbol{\tau}) = \sum_{\mathbf{v} \in \mathbb{Z}^2} \psi_{\eta,\mathbf{v}}(\mathbf{0}) \psi_{\eta,\mathbf{v}}(\boldsymbol{\tau})$ are the two-dimensional autocorrelation wavelets as defined in Section 3.2.4.

In order to estimate this quantity we require the estimate of the LWCS. With the above definition in place the result below demonstrates that the local cross-covariance of each pair of channels asymptotically tends to the true LS2Wmv process cross-covariance given by $c_{\mathbf{R}}^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) = \text{Cov}\left(X_{[\mathbf{zR}]}^{(p)}, X_{[\mathbf{zR}+\boldsymbol{\tau}]}^{(q)}\right)$.

Proposition 2 *Suppose we have two channels of an LS2Wmv process that satisfy Definition 18. We have that $\left|c_{\mathbf{R}}^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) - c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau})\right| = O\left(R^{-1}\right)$ for $\boldsymbol{\tau} \in \mathbb{Z}^2$ and $\mathbf{z} \in (0,1)^2$ and as $R \to \infty$.*

Proof. See Appendix A. ∎

As one might anticipate from previous work in the LS2W area, due to the invertibility of the inner product matrix A_J we can also establish that the LWCS can be considered an 'inverse transform' of the local cross-covariance for each pair of channels. This results in the following lemma.

Lemma 2 *The relationship between the LWCS and the local cross-covariance is invertible, with the inverse relationship given by*

$$S_{\eta}^{p,q}(\mathbf{z}, \boldsymbol{\tau}) = \sum_{\eta_1} A_{\eta,\eta_1}^{-1} \sum_{\boldsymbol{\tau}} c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) \Psi_{\eta_1}(\boldsymbol{\tau}), \tag{5.26}$$

where A_{η,η_1} is the autocorrelation wavelet inner product matrix defined in Section 3.2.4.

Proof. See Appendix A. ■

5.7 Concluding remarks

In this chapter we have introduced a multivariate locally stationary wavelet process model and proposed an unbiased and consistent measure of the dependence between two locally stationary channels, namely the locally stationary wavelet coherence. Following this we discussed the local wavelet cross-spectrum highlighting its uniqueness and detailing a full estimation procedure. Finally, we introduced a measure of the local cross-covariance structure and demonstrated that the local wavelet cross-spectra are a form of inverse transform of the local cross-covariance.

Chapter 6

Multivariate locally stationary 2D wavelet processes: an application to colour texture

In Chapter 5 we introduced a two-dimensional multivariate locally stationary wavelet model (LS2Wmv) and associated estimation theory for the locally stationary wavelet coherence between pairs of channels. We now consider a practical application of this modelling approach, in particular looking at an application related to the discrimination and classification of colour texture in hair analysis.

The automatic classification of visually similar textured images has been considered for many industrial product types including cereamic tiles, leather and fabric (Bianconi et al., 2012). We consider a different and potentially more subtle image set, namely that of hair. Historically there has been little work on this problem in the literature, especially for colour texture analysis. Instead industrial researchers have tended to use human classification via panellist trials. This motivates the need to develop an automated approach to the discrimination and classification of colour texture. This is the problem which we consider in this chapter.

The work presented is structured as follows. Section 6.1 considers the definition of colour and colour texture. We then review both traditional and wavelet-based approaches

to colour texture classification. Of these we select three methods with which to compare our approach, and discuss them in detail in Section 6.1.1. Following this we then consider forming a feature vector using the LS2Wmv approach and then applying these four methods to two texture analysis problems: discrimination and classification of test texture datasets. Finally in Section 6.5 we turn to the application of hair analysis.

6.1 Colour texture

Colour is commonly defined to be the by-product of the spectrum of light, reflected or absorbed, as received by the human eye and processed by the human brain. As discussed in Sengur (2008), the colour of a pixel is typically represented by the RGB tristimulus values. These correspond to the red, green and blue frequency bands of the visible light spectrum. When red, green and blue light are added together in equal quantities they produce white, whilst the absence of these colours creates black. This is by no means the only colour representation. Alternatives include HSV and Lab (Hunt, 1998). For our modelling approach we need not distinguish between these representations but knowledge of the colour framework is naturally useful for interpretation.

The visual perception of an image is a combination of colour, texture and scale where the change in scales will lead to change in textural appearence, as well as change in colour perception. Colour and texture are two of the most important properties, especially when one is dealing with real world images (Van de Wouwer et al., 1999). A colour texture can be regarded as a pattern described by the relationship between its *chromatic* and *structural* distribution. Specifically as Drimbarean and Whelan (2001) note, two images consisting of either (a) the same colour but different texture patterns or (b) the same texture pattern but different colours are said to be two *different* colour textures.

Many texture descriptors, initially defined for greyscale images, have been extended to colour spaces and used to classify colour textures. Such texture descriptors include statistics derived from Markov random fields (Panjwani and Healey, 1995; Hernandez et al., 2005), wavelet transforms (Van de Wouwer et al., 1999; Arivazhagan et al., 2005; Sengur, 2008), co-occurence matrices (Palm, 2004; Akhloufi et al., 2007), colour histograms and moments

(Pietikainen et al., 2002) and Gabor filtering (Drimbarean and Whelan, 2001). The conclusion common to all of these papers is that the introduction of colour information significantly improves colour texture classification. Since we consider cross-channel dependencies as fundamental to colour texture classificiation, below we restrict our review to those papers which take into account the correlation of texture measures between the different colour channels.

In one of the earlier papers Paschos (1998) proposed using directional histograms in order to obtain a set of discriminative chromatic correlation features. Akhloufi et al. (2007) considered isotropic co-occurence matrices in a colour texture classification framework. The co-occurence matrix was used as a feature extraction technique for each colour band. Although co-occurrence matrices provide good results in practice, the approach is computationally expensive. In addition both the above approaches lack a stochastic model-based justification and are based on heuristics, hence we do not consider them further.

Next we turn to consider spatially-based approaches in the literature. Panjwani and Healey (1995) proposes a Gaussian Markov Random Field model, where texture is characterized using pixel intensity on local spatial neighbourhoods. In addition Panjwani and Healey (1995) consider the interactions across the different planes. A disadvantage to this approach is the large number of coefficients in the model. Bennett and Khotanzad (1998) propose a multi-spectral random field model to analyse colour texture using maximum likelihood methods for parameter estimation. They represent spatial interactions in multiband images. Both Bennett and Khotanzad (1998) and Panjwani and Healey (1995) require a method to select the neighbour sets that define the interactions which is difficult in practice. Selecting these neighbouring sets is important as large numbers of parameters are used to define the interactions between and within colour planes. This has the effect of increasing the complexity of the approach. Finally, Suen and Healey (2000) introduce a spatial correlation model. This measures the within- and cross-band correlations to determine basis textures for each texture class. Whilst these types of models work well for certain applications, they fail to capture the multiscale nature of texture.

Van Heel et al. (1982) and Saxton and Baumeister (1982) introduce a Fourier measure namely the *Fourier-ring correlation* function. This calculates the correlation coefficients per frequency component. See Section 6.1.1 for further details. The main disadvantage of global

Fourier-based methods is they implicitly assume stationarity. This is unlikely in most real world images.

The literature also contains several wavelet-based approaches to dealing with this problem. Early work in this area is proposed by Van de Wouwer et al. (1999) who use wavelet 'energy-correlation signatures' for colour texture classification. Wavelet correlation signatures capture the characteristics of textures in terms of the energies of each colour plane, together with the cross correlation between different planes. The authors build feature vectors by linear transformation of these signatures to a different colour space. For each different colour space transformation and corresponding feature set considered, a k-nearest neighbour classifier is designed. This approach however, is highly dependent on the choice of the transformation (i.e. RGB, HSV etc) and has no formal stochastic modelling framework. Hence the potential to understand which features permit classification can be limited. Furthermore, the approach is based on the discrete wavelet transform so there could be key information missing, which would permit improved classification performance (see Eckley et al. (2010) for more details on this subtle point).

Ding et al. (2005) use a wavelet domain hidden Markov model for colour texture analysis. The proposed approach is used to model the dependencies between colour planes as well as interactions across scales. The wavelet coefficients at the same location, scale and direction, but different colour planes are grouped into one vector and a multivariate Gaussian mixture model is employed. This is used to approximate the marginal distribution of the wavelet coefficient vectors at a particular scale. A disadvantage to this approach is each texture is represented by a corresponding wavelet hidden Markov tree (WHMT) model. Therefore each WHMT model has to be trained with a single texture image. The training step in fitting the model can be computationally intensive due to the large number of model parameters; the model is often simplified drastically to cope with this computational cost, resulting in potentially unrealistic image representation and inference.

More recently Sengur (2008) proposed a method using the wavelet transform followed by an adaptive neuro-fuzzy inference system classifier. Wavelet entropies and wavelet energies of each colour plane at different scales are used for forming the feature vector of each colour texture since both quantities are widely used in the image processing literature. A

disadvantage to using neural networks is that they are known to be computationally intensive in the training step of the algorithm. More importantly for our application understanding, the features which permit the classification is important, however such knowledge is not possible from a neural set.

In order to show the full potential of our method we choose to compare it against three approaches: Van de Wouwer et al. (1999), Sengur (2008) and a Fourier-based approach using the ideas of Van Heel et al. (1982) and Saxton and Baumeister (1982). We choose to compare against the method of Van de Wouwer et al. (1999) as it is a popular approach, using a fast algorithm which specifically focusses on cross-correlation measures. Sengur (2008) is included since the author reports better classification results than Van de Wouwer et al. (1999) for their chosen examples. As Sengur (2008) is a measure which does not account for the dependencies between channels, it will be interesting to see whether accounting for the coherence improves classification performance. We also compare against an alternative method based on the Fourier spectrum and coherence (Van Heel et al., 1982; Saxton and Baumeister, 1982), due to practitioners continued familarity with Fourier methods.

In the following section we describe these three benchmark approaches and create the feature vectors to be used in the discrimination and classification algorithm. Here on in we shall refer to Sengur (2008) as 'Sengur', Van de Wouwer et al. (1999) as 'VdW' and the Fourier approach as 'Fourier'.

6.1.1 Existing multiscale colour texture methods

The basis on which much texture analysis is performed is to describe an image as a set of summary statistics or features. In the context of wavelet-based texture analysis, the features are generated from the wavelet coefficients of the decomposition subimages. We start by considering the approach by Van de Wouwer et al. (1999) (VdW).

VdW use wavelet energy correlation signatures to form a feature vector of each colour texture. These are given by:

$$C_{\eta}^{(p,q)} = \frac{1}{R^2} \sum_{\mathbf{u}} d_{\eta,\mathbf{u}}^{(p)} d_{\eta,\mathbf{u}}^{(q)}. \tag{6.1}$$

Here $d_{\eta,\mathbf{u}}^{(i)}$ is the wavelet coefficient at scale-direction η, location $\mathbf{u}$ and colour channel i

where $i \in \{1,2,3\}$. The set $\{C_\eta^{(p,q)} \mid p,q = 1,2,3; p \leq q, \forall\ \eta\}$ are called the wavelet energy signatures. For $p = q$ these capture the energy distribution of the wavelet coefficients over the scale-direction and colour space whereas the remainder ($p \neq q$) represent the covariance between colour spaces. For example in the case $j = 1$, the finest scale, we have six energy signatures for each of the horizontal, vertical and diagonal direction: i.e. $C^{(1,1)}$, $C^{(2,2)}$, $C^{(3,3)}, C^{(1,2)}, C^{(1,3)}$ and $C^{(2,3)}$. In other words this gives a feature vector of length 18.

The second method, Sengur, uses wavelet entropies and wavelet energies to form a feature vector of each colour texture. The energy vector is given as follows:

$$C_\eta^{(p,p)} = \frac{1}{R^2} \sum_{\mathbf{u}} (d_{\eta,\mathbf{u}}^{(p)})^2. \tag{6.2}$$

In set $\{C_\eta^{(p,p)} | p = 1,2,3\}$ gives the averaged l_2-norm as a measure of energy contained in the image. Following Sengur (2008) we take $j = 1$ considering only the finest scale, so we have 9 energy values, one for each scale-direction subimage for each of the three colours planes. These energy values are then used to form the feature vector of each colour texture. In this case we consider norm entropy with exponent 1.5 as suggested Sengur (2008). The entropy vector is given by:

$$H_\eta^{(p)} = \sum_{\mathbf{u}} |d_{\eta,\mathbf{u}}^{(p)}|^{1.5}. \tag{6.3}$$

In other words, the set $\{H_\eta^{(p)} | p = 1,2,3, \forall\ \eta\}$ gives the set of entropy values summed over location for each colour plane and each scale-direction subimage within that plane. Thus in the case of $j = 1$, we have 9 entropy values for each of the three colour planes. Combining the energy and entropy vectors together gives a feature vector of length 18.

Finally, as discussed previously we consider a Fourier approach, using the auto-spectra and Fourier coherence (as defined in Van Heel et al. (1982); Saxton and Baumeister (1982)) as a feature vector. The classical Fourier 2D coherence is given by

$$K(\omega) = \frac{|\sum_{\boldsymbol{\tau}} C_X^{(p,q)}(\boldsymbol{\tau}) e^{-p\boldsymbol{\omega}\boldsymbol{\tau}}|}{\sqrt{\sum_{\boldsymbol{\tau}} C_X^{p}(\boldsymbol{\tau}) cos(\boldsymbol{\omega}\boldsymbol{\tau})} \sqrt{\sum_{\boldsymbol{\tau}} C_X^{q}(\boldsymbol{\tau}) cos(\boldsymbol{\omega}\boldsymbol{\tau})}}, \tag{6.4}$$

and the auto- and cross-spectra are given by d^p x $\bar{d}^q$ where $p,q = 1,2,3$ and $p \leq q$ (similar

to the cross-periodogram). Here p and q denote the multiple channels and d^p represents the Fourier coefficients at channel p. $\sum_{\boldsymbol{\tau}} C_X^{(i,j)}(\boldsymbol{\tau}) e^{-i\boldsymbol{\omega}\boldsymbol{\tau}}$ represents the cross-spectra and $\sum_{\boldsymbol{\tau}} C_X^{j}(\boldsymbol{\tau}) cos(\boldsymbol{\omega}\boldsymbol{\tau})$ give the auto-spectra over lags $\boldsymbol{\tau}$. We construct feature vectors using the auto-spectra and coherences summed over all locations. Therefore we have 3 values for the auto-spectra and 3 values for the coherence giving a feature vector of length 6. In the following section we discuss how we construct a feature vector for the LS2Wmv modelling approach discussed in Chapter 5.

6.2 The LS2Wmv modelling approach as a colour texture analysis tool

In this section we introduce a feature vector based on the LS2Wmv model (Chapter 5). The key to our approach is finding the coherence between the colour channels as we believe the additional information provided by the coherence will allow more subtle differentiation between visually similar images. The feature vector we suggest considers the average auto and cross-spectral structure and the average local wavelet coherence at each scale-direction pair.

Algorithm 6.1 below describes the method which we use to obtain the feature vector of length $9 \times 3 \times J^*$ (for some $J^* \leq J$) for an LS2Wmv process with three channels.

6.3 The discrimination and classification procedure

The first part of our analysis investigates the ability of the four feature-generating methods Fourier, Sengur, VdW and LS2Wmv to discriminate between a set of images. These images are discussed in detail in Sections 6.4 and 6.5. To discriminate we use Linear Discriminant Analysis (LDA) (Fisher, 1936) which chooses a feature projection in order to maximise the separability of the classes. This is a popular approach amongst practitioners as it is parameter-free and computationally efficient. Whilst other discrimination techniques may also be appropriate we apply LDA since the purpose of the this chapter is to assess the reliability of the feature vector and not the choice of discrimination algorithm.

`LS2WmvFV:`

1. Given a colour textured image we denote each colour channel to be $X_{\mathbf{u}}^{(p)}$ for $p = 1, 2, 3$.
2. Estimate the wavelet auto- and cross-spectra of the image using equation (5.19) and then regularise these estimates (see Section 5.4.3) for scales $j = 1, \ldots, J$, directions $l = h, v, d$ and $p, q = 1, 2, 3$ where $p \leq q$.
3. Compute an estimate of the coherences $\rho_{\eta}^{(p,q)}$ (equation (5.22)) for all directions, scales and unique pairs of image channels (p, q).
4. Compute auto- and cross-spectral features using the location averages: $\bar{S}_{\eta}^{(p,q)} = \frac{1}{R^2} \sum_{\mathbf{u}} S_{\eta,\mathbf{u}}^{(p,q)}$ for $p, q = 1, 2, 3$ where $p \leq q$, $l \in \{h, v, d\}$ and scales $j = 1, \ldots, J^*$. J^* is the number of scale-direction pairs we wish to use in the feature vector.
5. Compute the average coherence features for the chosen scales as: $\bar{\rho}_{\eta}^{(p,q)} = \frac{1}{R^2} \sum_{\mathbf{u}} \rho_{\eta,\mathbf{u}}^{p,q}$ for all scales, directions and where $p, q = 1, 2, 3$ and $p < q$.
6. The feature vector is then the set of all spectral and coherence features: $F_{\text{LS2Wmv}} = \{\bar{S}_{\eta}^{(p,q)}, \bar{\rho}_{\eta}^{(p,q)} \mid j = 1, \ldots, J^*, l \in \{h, v, d\}$ and p, q as defined above for each separate feature$\}$.

Algorithm 6.1: Finding the feature vector for an LS2Wmv process with three channels.

We sample fifty sub-images of dimension m x m for each method from the upper half of each image to form the set used for discrimination. A feature vector using each method is then evaluated for each sub-image and this is the data we use within the LDA approach.

In the second part of our analysis we attempt to classify a test set of sub-images for the colour texture images in question. We use the upper half plane subimages as our training set, whilst the sub-images for testing are taken from the lower half of the image. The LDA analysis is then used in our classification routine. This is but one of many classification approaches which could be used. Specifically we assign a subimage candidate to a texture class based on the following procedure:

1. Perform LDA on the training set of all texture classes.
2. For each test sub-image, calculate the LDA-transformed feature vector and assign each sub-image to the class whose mean is closest in Euclidean distance.

We assess classification performance using an approach used by Van de Wouwer et al. (1999). The mean error rate $\hat{\epsilon}$ is given as

$$\hat{\epsilon} = \text{Total number of correctly classified test samples/N}, \tag{6.5}$$

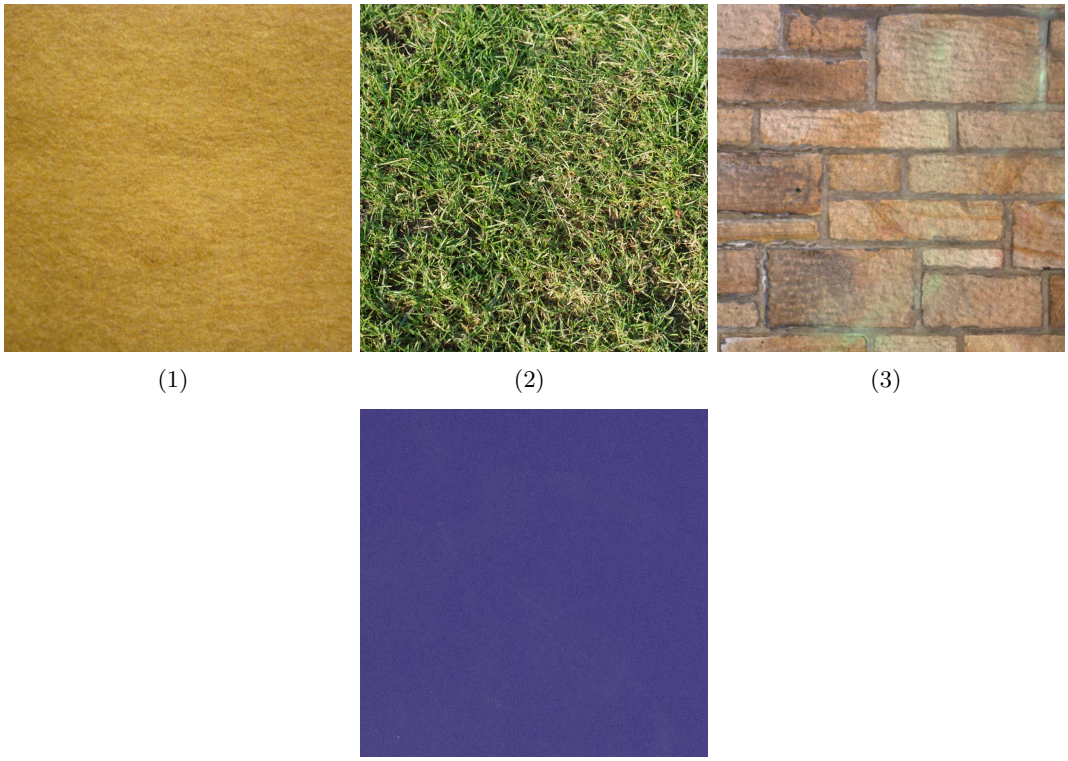
where N is the total number of samples. This number estimates the percentage of test samples classified correctly and is used as a measure of performance of the classifier.

Now that we have described the setup of our analysis, we consider the potential of each feature set in the next sections. Section 6.4 focusses on a typical test set of images commonly used in the literature whilst Section 6.5 considers an application to various texture analysis problems encountered by an industrial collaborator.

6.4 Classifying typical textures

We begin by focussing on a set of four textures shown in Figure 6.2. This set of images were chosen as they represent both stationary and non-stationary examples and also exhibit a variety of different colours and thus represent a selection of colour texture features. Each textured image we consider has dimension 1024 x 1024. To the eye, these textures are very different so we would hope all methods are able to discriminate effectively. We follow the discrimination and classification approach as outlined in Section 6.3, sampling fifty sub-images from the upper half of an image as training data and sampling from the lower half of an image as the test set.

Figure 6.3 shows that the three wavelet-based approaches discriminate the different texture types quite easily, with each texture class being reasonably localised within the discriminant plane. In contrast, the Fourier approach has less compact clusters. This is perhaps to be expected as some of these images seem to exhibit a non-stationary structure. We would therefore not expect a Fourier approach to discriminate as efficiently due to the lack of spatial localisation. In particular, images (1) and (4) from Figure 6.2 appear to be more stationary than images (2) and (3). This corresponds to what we see in Figure 6.3(a) where the stationary images correspond to the tighter clusters. Next we turn to the classification performance of the four approaches with this image set. Table 6.1 shows the classification rate of all the methods. All methods achieve very high classification results. This is not surprising as these textures are easy to classify by eye. The next section considers a more subtle texture classification problem.



(1) (2) (3)

(4)

Figure 6.2: Typical textures: (1) Thick yellow cloth, (2) Grass, (3) Brick wall and (4) Blue fabric. Images (2) and (3) were provided by A. Gott. Images (1) and (4) were downloaded from an open source image database http://www.imageafter.com.

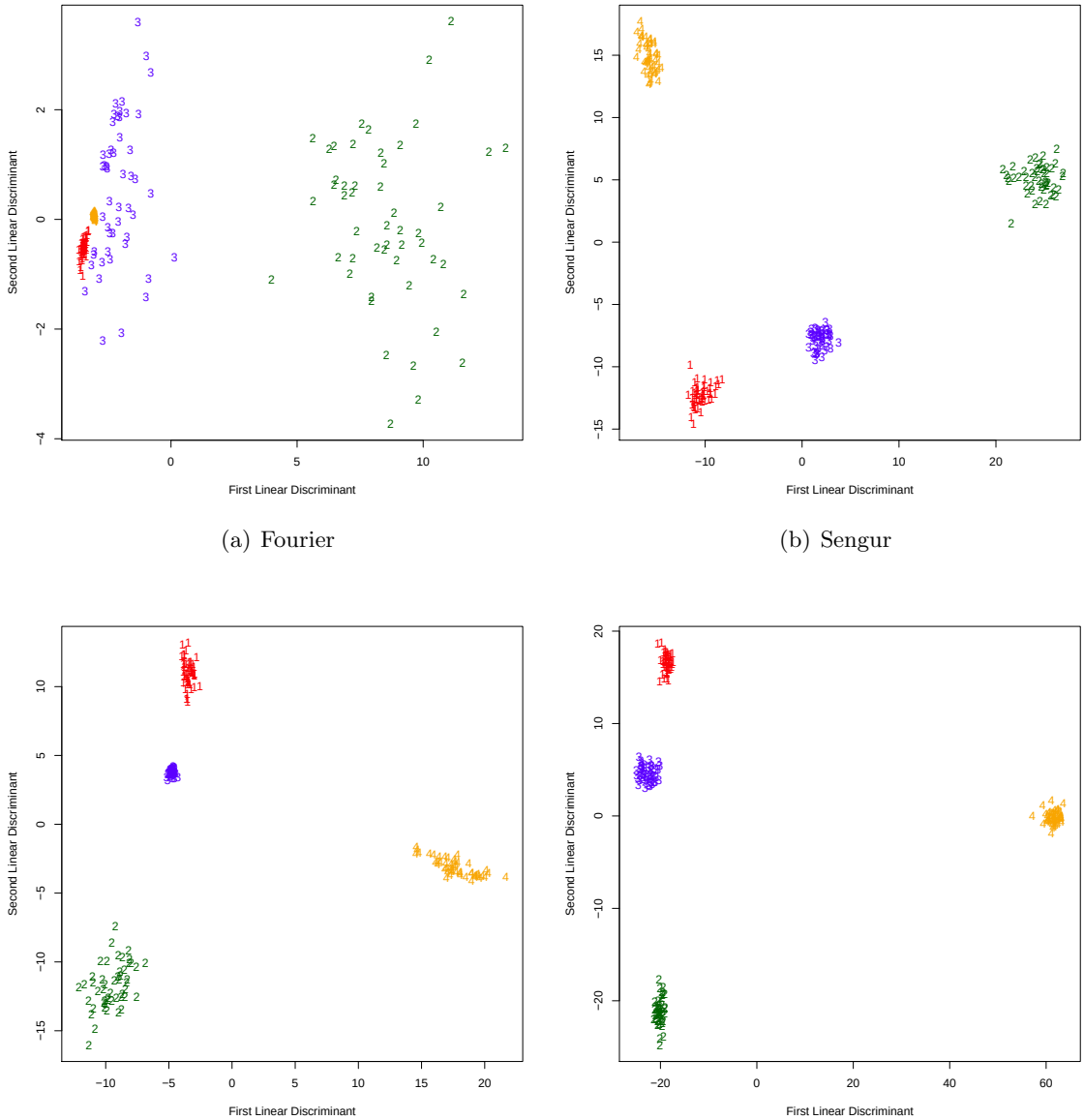
Method	Fourier	Sengur	VdW	LS2Wmv
% class correctly	94.5	100	100	99

Table 6.1: Percentage of textures classified correctly using the different methods for the four textures displayed in Figure 6.2.

6.5 Analysis of hair images

The following texture analysis example arises from work with an industrial collaborator. We have one original colour texture, a real hair sample, depicted in Figure 6.4. The original hair type was 50% dark brown European and 50% white, with a base hair colour of mixed brown and grey. Such hair swatches are used regularly for tests of new hair products.

In the following we consider the application of our modelling framework to three different colour texture experiments using the hair sample above. In the first example, we analyse a set of hair images showing a change in colour between the images but the *same texture*



(a) Fourier (b) Sengur

(c) VdW (d) LS2Wmv

Figure 6.3: Plots of the first two discriminant axes for each method based on sub-samples of size 64 x 64 of the images in Figure 6.2. Each subplot represents the different methods: (a) Fourier, (b) Sengur, (c) VdW and (d) LS2Wmv.

(Section 6.5.1). In this case three colourants have been applied to the original hair sample in Figure 6.4. The second and third set of images are consistent in colour, but their textural properties differ since the images represent hair that have had different preparation processes applied (Section 6.5.2). In both settings we wish to discriminate and classify between the

Figure 6.4: Original hair sample.

textures in each of these examples using the method outlined in Section 6.3. In a practical setting this is useful to determine features such as, for example, the state of a product by giving an indication of it's age or intensity of variation under different conditions. Such a task can be challenging even to the human eye, thus it is of interest to see which of the four methods performs well in these differing cases.

6.5.1 Hair treatments: Different colourants

In our first example we wish to distinguish between three different hair colourants A, B and C as shown in Figure 6.5. Each image is 256 x 256 in size and the only difference in these images is in the colour itself. In other words, each image has the same texture but different colour.

(1) (2) (3)

Figure 6.5: The different colourants, A,B and C.

We again follow the discrimination and classification approach as outlined in Section 6.3,

sampling fifty sub-images from the upper half of each image as training data and sampling fifty sub-images from the lower half of each image as the test set. Although our original image sizes are smaller than those in Section 6.4, we choose our sub-samples to have dimension 64 x 64 since spectral estimation methods can be less reliable for smaller image sizes due to the asymptotics associated with the estimation theory.

Figure 6.6 displays plots of the first two discriminant variables based on the sub-samples for each method. They are all reasonably well localised in the plane but the LS2Wmv approach shows more compact clusters. Table 6.2 again shows all methods have high classification rates with LS2Wmv achieving the highest correct classification. In this case the physical texture is the same across all images, however the colour changes. Hence, as we would anticipate, the two methods which take coherence into consideration produced the best results, namely Fourier and LS2Wmv.

Method	Fourier	Sengur	VdW	LS2Wmv
% class correctly	96	87.33	79.33	96.67

Table 6.2: Percentage of textures classified correctly using the different methods for the three textures displayed in Figure 6.5.

6.5.2 Hair colourants: Different preparation processes

Our next example takes the original image from Figure 6.4 and applies colourant B to three different physical texture processes (Figure 6.7). In other words, after colourant B has been applied to the original image, three different processes are undergone independently. We then repeat the experiment for colourant C as shown in Figure 6.8. As colourant A considers an image of black hair, we decided not to test it as this particular study considers colour texture images. From Figures 6.7 and 6.8 these samples are very difficult to discriminate visually.

Similar to our previous examples, the first part of our analysis is to discriminate between these images, for each colourant independently. We sample fifty sub-images from each image of dimension 64 x 64. A feature vector is then evaluated for each sub-image. Figure 6.9 displays the first two discriminant variables for the different methods for colourant

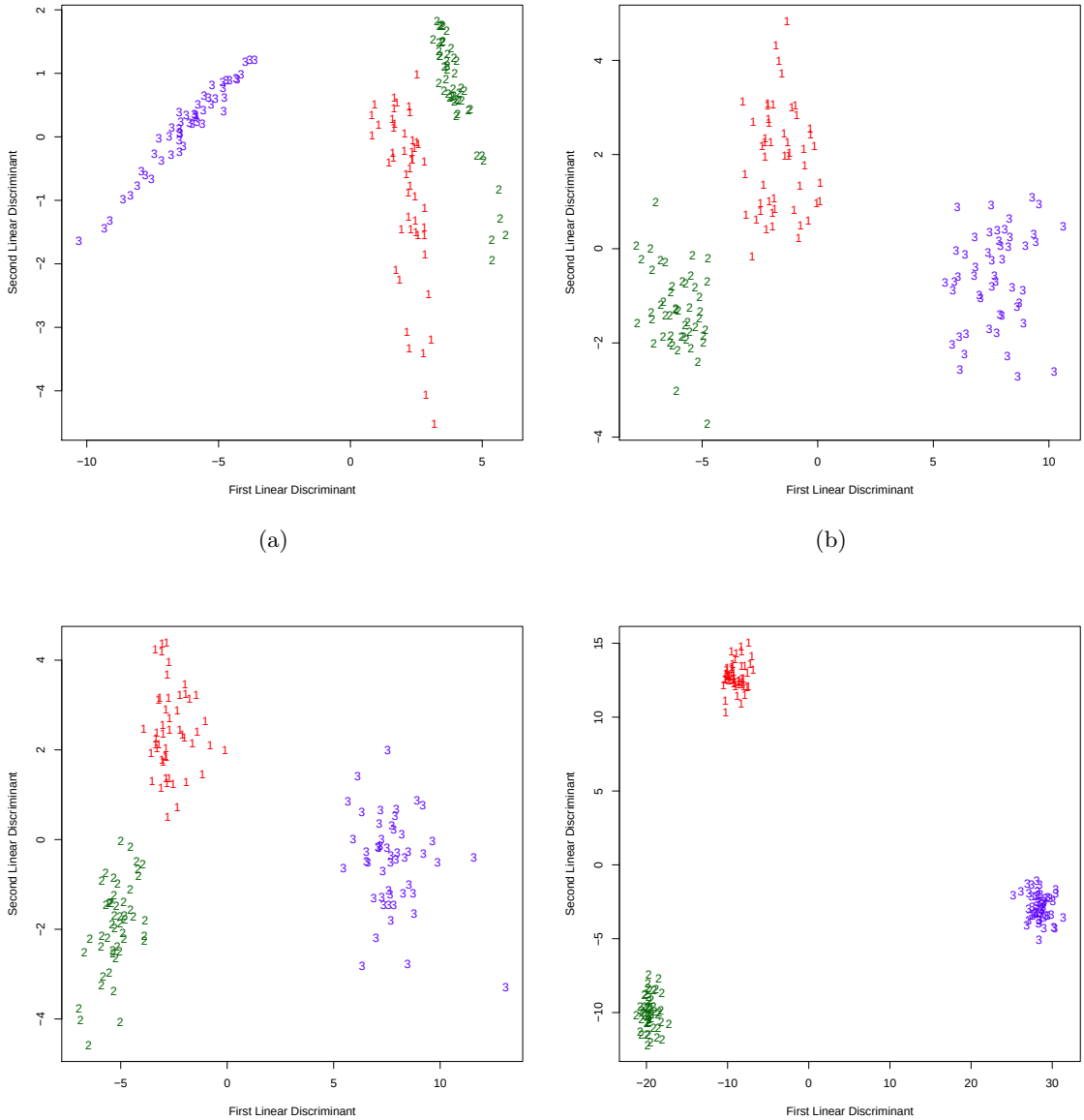


Figure 6.6: Plots of the first two discriminant axes for each method based on sub-samples of size 64 x 64 of the images in Figure 6.5. Each subplot represents the different methods: (a) Fourier, (b) Sengur, (c) VdW and (d) LS2Wmv.

B. The different texture types are still reasonably well localised for Fourier, Sengur and VdW, but the LS2Wmv approach again gives more compact clusters. Table 6.3 shows the classification results for colourant B where again the LS2Wmv method achieves the highest classification rate. As expected VdW is comparable due to the feature vector containing

(1) (2) (3)

Figure 6.7: Hair images of three preparation processes (i), (ii) and (iii) applied to a sample of colourant B.

(1) (2) (3)

Figure 6.8: Hair images of three preparation processes (i), (ii) and (iii) applied to a sample of colourant C.

wavelet correlation signatures taking into account the textural change across processes. As all the images are for colourant B they have less variation across the colour planes so the Fourier approach does not fare well as the images show change in their textural properties.

Method	Fourier	Sengur	VdW	LS2Wmv
% class correctly	77.33	68	90	91.33

Table 6.3: Percentage of textures classified correctly using the different methods for the three textures displayed in Figure 6.7.

Figure 6.10 displays plots of the first two discriminant variables from textures displayed in Figure 6.8 (Colourant C). The different texture types are well localised in the plane for the LS2Wmv approach. However the other three methods struggle to disguish between

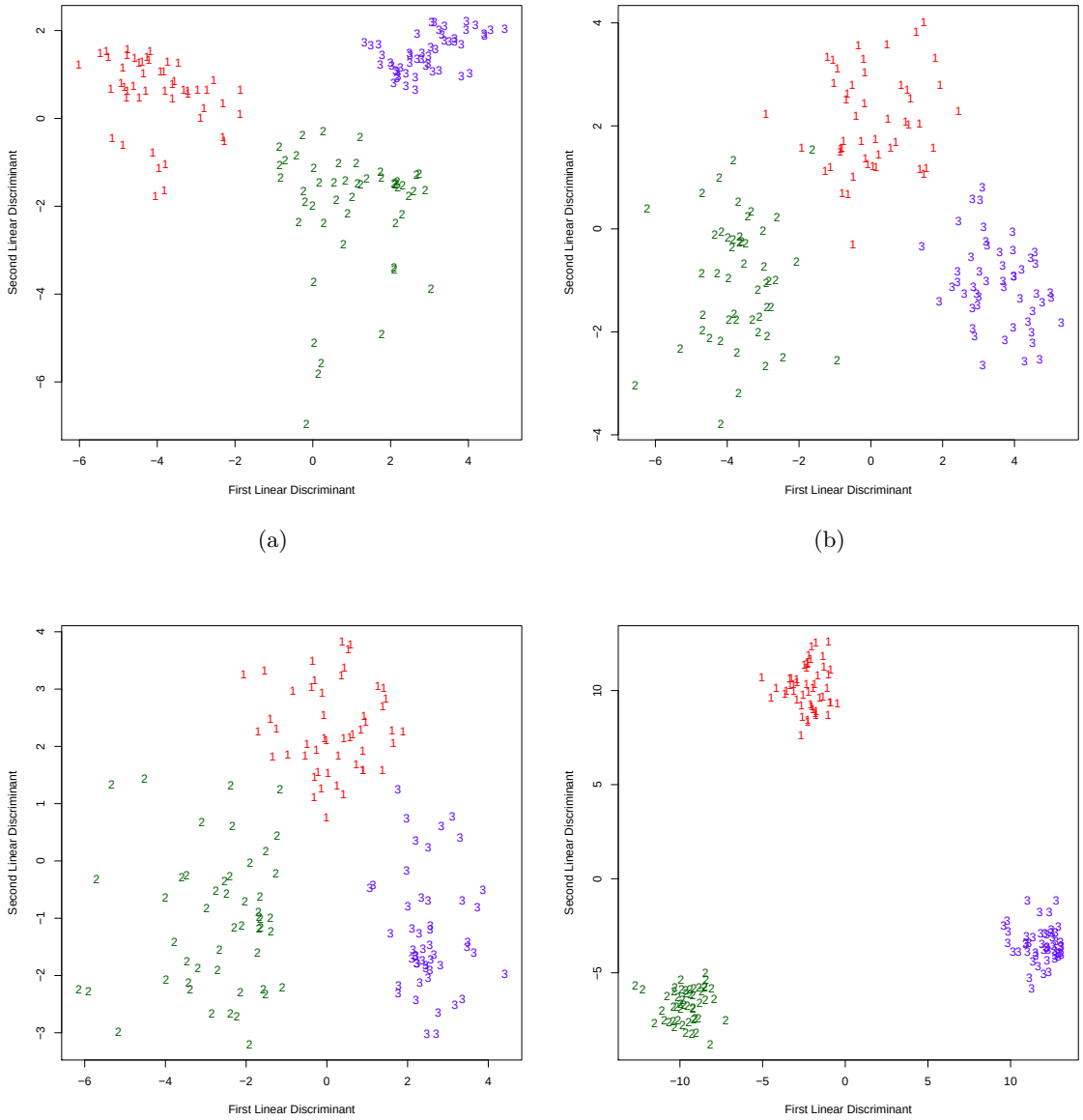
them. Depending on the colour of the image, certain colour planes may be redundant when it comes to LDA classification. This is highlighted in this second example where the classification percentages are lower. Table 6.4 shows the classification results for colourant C, highlighting LS2Wmv again gains the higher classification results.

Method	Fourier	Sengur	VdW	LS2Wmv
% class correctly	56.67	57.33	68.67	74.67

Table 6.4: Percentage of textures classified correctly using the different methods for the three textures displayed in Figure 6.8.

6.6 Concluding remarks

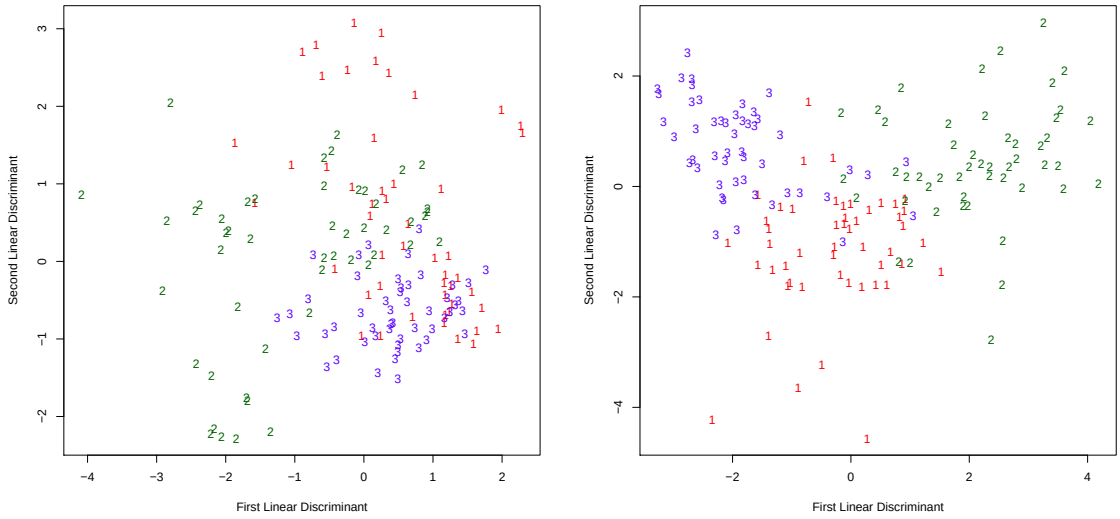
In this chapter we have discussed an application of the LS2Wmv modelling approach applying LS2W coherence in a colour texture analysis setting. We draw similar conclusions to Eckley et al. (2010) in that a wavelet approach produces higher correct classification in non-stationary images. The examples show that if texture images differ in colour, then Fourier and LS2Wmv are preferable, whereas if texture images differ on physical texture, then VdW and LS2Wmv are more favourable. Fourier and Sengur were generally found to be less robust methods in this case due to their inability to capture the multiscale nature of texture and the subtle changes between the colour texture images. Hence, on this basis we would generally prefer LS2Wmv over all other approaches since it can deal well with both colour and texture feature changes. In this case the coherence contributes to the higher classification rate and thus underlines the importance of coherence in a colour texture setting. In conclusion, the results suggest that the LS2Wmv method is more robust to different texture types.



(a) (b)

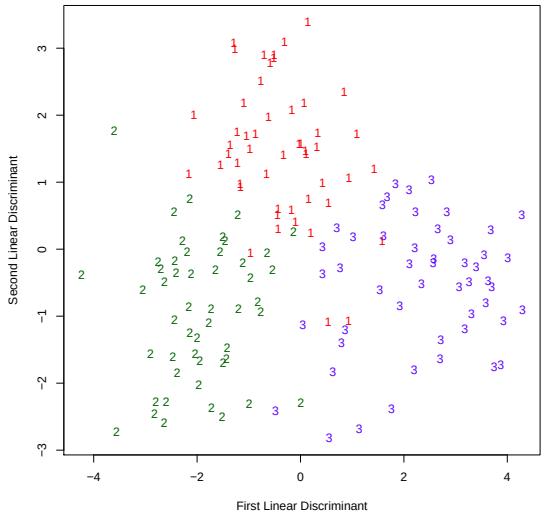
(c) (d)

Figure 6.9: Plots of the first two discriminant axes for each method based on sub-samples of size 64 x 64 of the images in Figure 6.7. Each subplot represents the different methods: (a) Fourier, (b) Sengur, (c) VdW and (d) LS2Wmv.

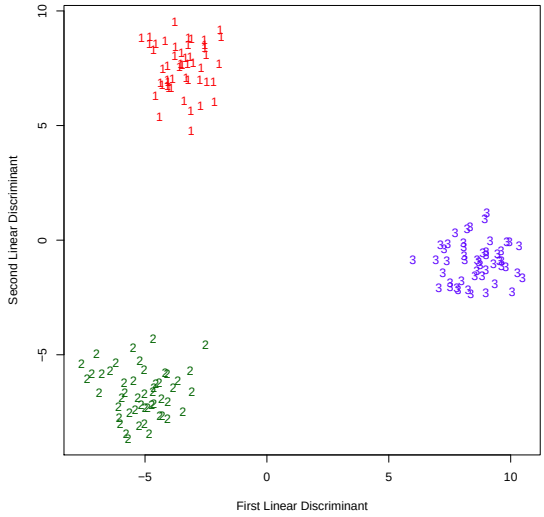


(a) Fourier

(b) Sengur



(c) VdW



(d) LS2Wmv

Figure 6.10: Plots of the first two discriminant axes for each method based on sub-samples of size 64 x 64 of the images in Figure 6.8. Each subplot represents the different methods: (a) Fourier, (b) Sengur, (c) VdW and (d) LS2Wmv.

Chapter 7

Conclusions and future directions

This thesis has focussed on developing statistical methodology for modelling image texture. Prior to analysing an image it is useful to know its stationarity properties in order to determine the most effective modelling approach to undertake. To this end, in Chapter 3 we developed a bootstrap-based test of stationarity, using the LS2W framework, extending the work of Cardinali and Nason (2011) to two dimensions. This test of stationarity is then used in Chapter 4 to address the problem of counting the number of textures in an image set. The testing procedure is embedded within an image segmentation method using the $\texttt{Bootstat}_{\texttt{LS2W}}$ test of Taylor et al. (2013) algorithm as a homogeneity measure. This enables us to not only obtain the number of textures in an image but also isolate areas of textural change. Details of the software used to implement the above methodology can be found in Appendix B.

In Chapter 5 we developed a rigorous wavelet-based modelling framework for non-stationary multivariate images, namely the LS2Wmv approach. Each multivariate image is assumed to have k channels and the local wavelet coherence between these channels was defined and estimated. In order to calculate this estimate we considered the scale-direction-location based spectral structure measure of both the individual channels and across channels. We provided a robust estimation procedure considering all asympotic properties, proving our measures to be unbiased, consistent and unique. Finally in Chapter 6 we explored the use of this model in a colour texture analysis problem, namely the discrimination and classification of hair images. Comparing our results with other approaches we found that

the LS2Wmv approach gives the best result in both changing colour and changing texture circumstances.

We conclude this thesis by considering a number of avenues for future research. As suggested in Chapter 3 it is surprising that no local Fourier equivalent of our spatial test currently exists. Hence, we note that the development of a spatial test for stationarity using a Fourier approach could be considered. It would be natural to consider a SLEX (Smooth Localised Complex Exponentials) based approach (Ombao et al., 2002, 2005) to model texture and therefore perform a local spatial stationarity test. This could then be applied as a homogeneity measure within an image segmentation framework as in Chapter 4.

Future work arising from Chapter 4 might include an investigation of other segmentation techniques which could provide more efficient or accurate ways to detect texture region boundaries. The limitation in our current work is due to the nature of the quad tree segmentation approach. In order to determine more accurate boundaries of textured regions, further segmentation is required.

Further, since the work from Chapters 3 and 4 focuses on univariate LS2W processes. It would also be interesting to extend these to a multivariate setting, especially with the additional coherence condition between channels to consider. Naturally, this would require careful consideration of what stationarity means in a colour setting.

Finally, the LS2W and LS2Wmv models underpinning the work presented in this thesis have assumed data on a regular grid of size 2^nx2^n. In practical applications this isn't always the case, for example in Chapter 4 we may have mosaics of different shapes. It would be of interest to investigate the extension of these models to a more general structure, i.e. non-regular grids. Lifting transforms can handle multidimensional irregularly spaced data so would be a natural consideration in extending these approaches, drawing on ideas from a recently proposed time series approach by Knight et al. (2012).

Appendix A

Proofs for Chapter 5

Proof of Lemma 1: Properties of products of Lipschitz continuous variables.

We begin by proving the product of the transfer functions $(W_\eta^{(p)}(\mathbf{u}/\mathbf{R})W_\eta^{(q)}(\mathbf{u}/\mathbf{R}))$ is Lipschitz. We have,

$$\begin{aligned}
\left|W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) - W_\eta^{(p)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| &\leq \left|W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) - W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| \\
&+ \left|W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right) - W_\eta^{(p)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| \\
&= \left|W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)\right|\left|W_\eta^{(q)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) - W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| \\
&+ \left|W_\eta^{(q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right|\left|W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) - W_\eta^{(p)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| \\
&\leq \frac{C^*}{R}L_\eta^q||\mathbf{u}-\mathbf{r}|| + \frac{C^*}{R}L_\eta^p||\mathbf{u}-\mathbf{r}|| \\
&= \frac{C^*}{R}(L_\eta^{(p)} + L_\eta^{(q)})||\mathbf{u}-\mathbf{r}||,
\end{aligned}$$

since $|W_\eta^{(p)}(\mathbf{u}/\mathbf{R})| \leq C^*$ and using the Lipschitz properties of $W_\eta^{(p)}$ and $W_\eta^{(q)}$. Therefore $(W_\eta^{(p)}(\mathbf{u}/\mathbf{R})W_\eta^{(q)}(\mathbf{u}/\mathbf{R}))$ is Lipschitz with constant $C^*(L_\eta^{(p)} + L_\eta^{(q)})$. By similar arguments, the product of $(W_\eta^{(p)}(\mathbf{u}/\mathbf{R})W_\eta^{(q)}(\mathbf{u}/\mathbf{R}))$ and $\rho_\eta^{(p,q)}(\mathbf{u}/\mathbf{R})$ is Lipschitz. Using equation (5.7) we have,

$$\begin{aligned}
\left|S_\eta^{(p,q)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) - S_\eta^{(p,q)}\left(\frac{\mathbf{r}}{\mathbf{R}}\right)\right| &\leq 1 \cdot \left[\frac{C^*}{R}(L_\eta^{(p)} + L_\eta^{(q)}) + \frac{R_\eta^{(p,q)}}{R}\right]||\mathbf{u}-\mathbf{r}|| \\
&\leq \frac{CB_\eta^{(p,q)}}{R}||\mathbf{u}-\mathbf{r}||,
\end{aligned}$$

since $|\rho_{\eta}^{(p,q)}(\mathbf{u}/\mathbf{R})| \leq 1$ and defining $C = \frac{\max(C^*,1)}{3}$ and $B_{\eta}^{(p,q)} = \max(L_{\eta}^{(p)}, L_{\eta}^{(q)}, R_{\eta}^{(p,q)})$. ■

Proof of Theorem 1: Uniqueness of representation.

The structure of this proof is very similar to the two-dimensional case, considered by Eckley et al. (2010). In order to prove the uniqueness of the multivariate spectral representation, it suffices to consider the properties of the cross-spectrum, since the uniqueness of the auto-spectra has been established in Eckley et al. (2010).

Suppose, by way of contradiction, that there exist two cross-spectral representations of the same LS2Wmv process which also possess the same cross-covariance structure. In other words

$$c^{p,q}(\mathbf{z}, \boldsymbol{\tau}) = \sum_{\eta} S_{\eta}^{(1)(p,q)}(\mathbf{z}) \Psi_{\eta}(\boldsymbol{\tau}) = \sum_{\eta} S_{\eta}^{(2)(p,q)}(\mathbf{z}) \Psi_{\eta}(\boldsymbol{\tau})$$

where $c^{p,q}(\mathbf{z}, \boldsymbol{\tau})$ is defined in equation (5.25) and $S_{\eta}^{(i)(p,q)}(\mathbf{z}) = W_{\eta}^{p}(\mathbf{z}) W_{\eta}^{q}(\mathbf{z}) \rho_{\eta}^{(p,q)}(\mathbf{z})$ for $i = 1, 2$ and $p, q \in \{1, \ldots m\}$ with $p \neq q$. Let $\Delta_{\eta}^{(p,q)}(\mathbf{z}) \equiv S_{\eta}^{(1)(p,q)}(\mathbf{z}) - S_{\eta}^{(2)(p,q)}(\mathbf{z})$ be the difference between the two representations. To establish uniqueness we must show that

$$\begin{aligned} 0 &= \sum_{\eta} \Delta_{\eta}^{(p,q)}(\mathbf{z}) \Psi_{\eta}(\boldsymbol{\tau}), \quad \forall\ \mathbf{z} \in (0,1)^2,\ \forall\ \boldsymbol{\tau} \in \mathbb{Z}^2 \\ \Rightarrow 0 &= \Delta_{\eta}^{(p,q)}(\mathbf{z}), \quad \forall\ \eta = 1, \ldots, 3J,\ \forall\ \mathbf{z} \in (0,1)^2. \end{aligned}$$

What we actually show is an equivalent implication:

$$0 = \sum_{\eta} \tilde{\Delta}_{\eta}^{(p,q)}(\mathbf{z}) \Psi_{\eta}(\boldsymbol{\tau}), \quad \forall\ \mathbf{z} \in (0,1)^2,\ \forall\ \boldsymbol{\tau} \in \mathbb{Z}^2 \tag{A.1}$$

$$\Rightarrow 0 = \tilde{\Delta}_{\eta}^{(p,q)}(\mathbf{z}), \quad \forall\ \eta = 1, \ldots, 3J,\ \forall\ \mathbf{z} \in (0,1)^2, \tag{A.2}$$

where $\tilde{\Delta}_{\eta}^{(p,q)}(\mathbf{z}) = 2^{-2j(\eta)} \Delta_{\eta}^{(p,q)}(\mathbf{z})$. We begin by recalling some key properties of the autocorrelation wavelets established in Eckley et al. (2010). Specifically using Parseval's relation,

the discrete autocorrelation wavelet inner product matrix can be expressed as

$$\begin{aligned} A_{\eta,\nu} &= \sum_{\boldsymbol{\tau}} \Psi_\eta(\boldsymbol{\tau})\Psi_\nu(\boldsymbol{\tau}) \\ &= \left(\frac{1}{2\pi}\right)^2 \int\int \hat{\Psi}_\eta(\boldsymbol{\omega})\hat{\Psi}_\nu(\boldsymbol{\omega})d\boldsymbol{\omega}, \end{aligned} \tag{A.3}$$

where $\hat{\Psi}_\eta(\boldsymbol{\omega}) = |\hat{\psi}_\eta(\boldsymbol{\omega})|^2$ and $\hat{\Psi}_\eta(\boldsymbol{\omega})$ represents the Fourier Transform of $\Psi(\boldsymbol{\tau})$. Thus assuming (A.1), for a given pair of channels p and q, we find that

$$\begin{aligned} 0 &= \left(\sum_\eta \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\Psi_\eta(\boldsymbol{\tau})\right)\left(\sum_\nu \tilde{\Delta}_\nu^{(p,q)}(\mathbf{z})\Psi_\nu(\boldsymbol{\tau})\right) \\ &= \sum_{\boldsymbol{\tau}}\left(\sum_\eta \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\Psi_\eta(\boldsymbol{\tau})\right)\left(\sum_\nu \tilde{\Delta}_\nu^{(p,q)}(\mathbf{z})\Psi_\nu(\boldsymbol{\tau})\right) \\ &= \sum_\eta\sum_\nu \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\tilde{\Delta}_\nu^{(p,q)}(\mathbf{z})\sum_{\boldsymbol{\tau}}\Psi_\eta(\boldsymbol{\tau})\Psi_\nu(\boldsymbol{\tau}). \end{aligned}$$

Applying equation (A.3), this becomes,

$$\begin{aligned} 0 &= \sum_\eta\sum_{\boldsymbol{\nu}} \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\tilde{\Delta}_\nu^{(p,q)}(\mathbf{z})\left(\frac{1}{2\pi}\right)^2\int\int \hat{\Psi}_\eta(\boldsymbol{\omega})\hat{\Psi}_\nu(\boldsymbol{\omega})d\boldsymbol{\omega}, \\ &= \int\int d\boldsymbol{\omega}\left(\sum_\eta \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\hat{\Psi}_\eta(\boldsymbol{\omega})\right)^2. \end{aligned} \tag{A.4}$$

Since $\sum_\eta S_\eta^{(p,q)}(\mathbf{z}) < \infty$ uniformly in $\mathbf{z}$, we have $\sum_\eta |\Delta_\eta^{(p,q)}(\mathbf{z})| < \infty$ and hence $\sum_\eta 2^{2j(\eta)}|\tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})| < \infty$. For any channel pair (p,q), $\hat{\Psi}_\eta(\boldsymbol{\omega})$ is continuous on $[-\pi,\pi]^2$ and in turn $\sum_\eta \tilde{\Delta}_\eta^{(p,q)}\hat{\Psi}_\eta(\boldsymbol{\omega})$ is continuous on $[-\pi,\pi]^2$ as a function of $\boldsymbol{\omega}$. Hence equation (A.4) implies that,

$$0 = \sum_\eta \tilde{\Delta}_\eta^{(p,q)}(\mathbf{z})\hat{\Psi}_\eta(\boldsymbol{\omega}), \quad \forall\ \boldsymbol{\omega} \in [-\pi,\pi]^2,\ \forall\ \mathbf{z} \in (0,1)^2. \tag{A.5}$$

In order to complete this proof we reconsider the Fourier properties of wavelets as

discussed in Chapter 2. For $\boldsymbol{\omega} = (\omega_1, \omega_2)$,

$$\left.\begin{aligned}
\left|\hat{\Psi}^v_\eta(\boldsymbol{\omega})\right|^2 &= 2^{2j}\left|m_1(2^{j-1}\omega_1)^2\right|\left|m_0(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 \\
\left|\hat{\Psi}^h_\eta(\boldsymbol{\omega})\right|^2 &= 2^{2j}\left|m_0(2^{j-1}\omega_1)^2\right|\left|m_1(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 \\
\left|\hat{\Psi}^d_\eta(\boldsymbol{\omega})\right|^2 &= 2^{2j}\left|m_1(2^{j-1}\omega_1)^2\right|\left|m_1(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 .
\end{aligned}\right\} \tag{A.6}$$

Let $\tilde{\Delta}^{(p,q)}_\eta = \tilde{\Delta}^{(p,q)}_\eta(\mathbf{z})$ at some fixed point $\mathbf{z} \in (0,1)^2$. Then using equation (A.6) we can rewrite equation (A.5) as follows:

$$\begin{aligned}
0 = &\sum_{\eta=1}^{J}\tilde{\Delta}^{(p,q)}_\eta 2^{2j}\left|m_1(2^{j-1}\omega_1)^2\right|\left|m_0(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 \\
&+\sum_{\eta=J+1}^{2J}\tilde{\Delta}^{(p,q)}_\eta 2^{2j}\left|m_0(2^{j-1}\omega_1)^2\right|\left|m_1(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 \\
&+\sum_{\eta=2J+1}^{3J}\tilde{\Delta}^{(p,q)}_\eta 2^{2j}\left|m_1(2^{j-1}\omega_1)^2\right|\left|m_1(2^{j-1}\omega_2)^2\right|\prod_{s=0}^{j-2}\left|m_0(2^s\omega_1)m_0(2^s\omega_2)\right|^2 .
\end{aligned} \tag{A.7}$$

The RHS of equation (A.7) is a continuous function of $\boldsymbol{\omega}$ and so must vanish for all $\boldsymbol{\omega} \in [-\pi, \pi]^2$. In order to show that $\tilde{\Delta}_1$ is zero, we insert $\boldsymbol{\omega} = (\pi, 0)$ into equation (A.7). This gives,

$$\begin{aligned}
0 &= \tilde{\Delta}^{(p,q)}_1 4|m_1(\pi)|^2|m_0(0)|^2 + \tilde{\Delta}^{(p,q)}_{2J+1}4|m_1(\pi)|^2|m_1(0)|^2 \\
&= \tilde{\Delta}^{(p,q)}_1 4\cdot 1\cdot 1 + \tilde{\Delta}^{(p,q)}_{2J+1}4\cdot 1\cdot 0 &\text{(A.8)}\\
&= \tilde{\Delta}^{(p,q)}_1, &\text{(A.9)}
\end{aligned}$$

since $|m_1(\pi)|^2$=1, $|m_0(0)|^2$=1 and therefore $|m_1(2\pi n)|^2$=0 (recalling properties of m_0 and m_1 from Section 2.1.4).

To show that $\tilde{\Delta}^{(p,q)}_{2J+1} = 0$ we reconsider equation (A.7). Taking $\boldsymbol{\omega} = (\pi, \pi)$ and since $\tilde{\Delta}^{(p,q)}_1 = 0$, we have,

$$\begin{aligned}
0 &= \tilde{\Delta}^{(p,q)}_{2J+1}4|m_1(\pi)|^2|m_1(\pi)|^2 \\
&= \tilde{\Delta}^{(p,q)}_{2J+1}4\cdot 1\cdot 1 &\text{(A.10)}\\
&= \tilde{\Delta}^{(p,q)}_{2J+1}. &\text{(A.11)}
\end{aligned}$$

Finally we insert $\boldsymbol{\omega} = (0, \pi)$ into equation (A.7). Given that $\tilde{\Delta}_1^{(p,q)} = 0$ and $\tilde{\Delta}_{2J+1}^{(p,q)} = 0$ we can also derive that $\tilde{\Delta}_{J+1}^{(p,q)} = 0$.

We next show that $\tilde{\Delta}_2^{(p,q)} = 0$ by setting $\boldsymbol{\omega} = (\pi/2, 0)$:

$$\begin{aligned} 0 &= \tilde{\Delta}_2^{(p,q)} 16 |m_1(\pi)|^2 |m_0(0)|^2 + \tilde{\Delta}_{2J+2}^{(p,q)} 16 |m_1(\pi)|^2 |m_1(0)|^2 \\ &= \tilde{\Delta}_2^{(p,q)} 16 \cdot 1 \cdot 1 + \tilde{\Delta}_{2J+2}^{(p,q)} 16 \cdot 1 \cdot 0 \qquad \text{(A.12)} \\ &= \tilde{\Delta}_2^{(p,q)}. \qquad \text{(A.13)} \end{aligned}$$

Similarly we can show that $\tilde{\Delta}_{2J+2}^{(p,q)} = 0$ and $\tilde{\Delta}_{J+2}^{(p,q)} = 0$. Continuing with this process recursively setting $\omega = \pi/2^{(j-1)}$ we can show that

$$\tilde{\Delta}_\eta^{(p,q)}(\mathbf{z}) = 0, \quad \forall \eta, \forall \mathbf{z} \in (0,1)^2. \qquad \text{(A.14)}$$

Hence as the difference between the two cross-spectra is zero, we assume they give the same covariance structure and so the cross-spectral representations are uniquely defined given the corresponding LS2Wmv process. ■

Proof of Theorem 2: The expectation and variance of the raw cross-periodogram between two channels, p and q of an LS2Wmv process.

Expectation: The expectation of the local wavelet raw cross-periodogram is given by,

$$\begin{aligned} \mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) &= \mathbb{E}[(d_{\eta,\mathbf{s}}^p)(d_{\eta,\mathbf{s}}^q)], \\ &= \mathbb{E}\left[\sum_{\mathbf{r_1}} \sum_{\mathbf{r_2}} X_{\mathbf{r_1}}^{(p)} X_{\mathbf{r_2}}^{(q)} \psi_{\eta,\mathbf{s}}(\mathbf{r_1}) \psi_{\eta,\mathbf{s}}(\mathbf{r_2}) \right] \\ &= \mathbb{E}\Bigg[\sum_{\mathbf{r_1},\mathbf{r_2}} \sum_{\eta_1,\eta_2} \sum_{\mathbf{u_1},\mathbf{u_2}} W_{\eta_1,\mathbf{u_1}}^{(p)} \psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1}) \xi_{\eta_1,\mathbf{u_1}}^{(p)} \times \\ &\qquad W_{\eta_2,\mathbf{u_2}}^{(q)} \psi_{\eta_2,\mathbf{u_2}}(\mathbf{r_2}) \xi_{\eta_2,\mathbf{u_2}}^{(q)} \psi_{\eta,\mathbf{s}}(\mathbf{r_1}) \psi_{\eta,\mathbf{s}}(\mathbf{r_2}) \Bigg], \end{aligned}$$

using the LS2Wmv model Definition 13. Next we recall that by definition,

$$\begin{aligned}\mathbb{E}\left(\xi^{(p)}_{\eta_1,\mathbf{u_1}}\xi^{(q)}_{\eta_2,\mathbf{u_2}}\right) &= \delta_{\eta_1,\eta_2}\delta_{\mathbf{u_1},\mathbf{u_2}}\rho^{(p,q)}_{\eta_1,\mathbf{u_1}},\\ &= \begin{cases}\rho^{(p,q)}_{\eta_1,\mathbf{u_1}} & \eta_1=\eta_2,\mathbf{u_1}=\mathbf{u_2}\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

The expectation thus reduces as follows:

$$\begin{aligned}\mathbb{E}(I^{(p,q)}_{\eta,\mathbf{s}}) &= \sum_{\mathbf{r_1},\mathbf{r_2}}\sum_{\eta_1}\sum_{\mathbf{u_1}} W^{(p)}_{\eta_1,\mathbf{u_1}}W^{(q)}_{\eta_1,\mathbf{u_1}}\rho^{(p,q)}_{\eta_1,\mathbf{u_1}}\psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1})\psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_2})\psi_{\eta,\mathbf{s}}(\mathbf{r_1})\psi_{\eta,\mathbf{s}}(\mathbf{r_2})\\ &= \sum_{\eta_1}\sum_{\mathbf{u_1}} S^{(p,q)}_{\eta_1,\mathbf{u_1}}\sum_{\mathbf{r_1}}\psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1})\psi_{\eta,\mathbf{s}}(\mathbf{r_1})\sum_{\mathbf{r_2}}\psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_2})\psi_{\eta,\mathbf{s}}(\mathbf{r_2}). \qquad (A.15)\end{aligned}$$

Substituting $\mathbf{u_1}=\mathbf{x}+\mathbf{s}$ into the equation above we obtain,

$$\begin{aligned}\mathbb{E}(I^{(p,q)}_{\eta,\mathbf{s}}) &= \sum_{\eta_1}\sum_{\mathbf{x}} S^{(p,q)}_{\eta_1,\mathbf{x}+\mathbf{s}}\left\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}+\mathbf{s}}(\mathbf{r})\psi_{\eta,\mathbf{s}}(\mathbf{r})\right\}^2\\ &= \sum_{\eta_1}\sum_{\mathbf{x}} S^{(p,q)}_{\eta_1,\mathbf{x}+\mathbf{s}}\left\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}+\mathbf{s}-\mathbf{r}}\psi_{\eta,\mathbf{s}-\mathbf{r}}\right\}^2 \qquad (A.16)\\ &= \sum_{\eta_1}\sum_{\mathbf{x}} S^{(p,q)}_{\eta_1,\mathbf{x}+\mathbf{s}}\left\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}-\mathbf{r}}\psi_{\eta,-\mathbf{r}}\right\}^2.\end{aligned}$$

By the assumptions of the model given in Definition 13, W^p_η and $\rho^{p,q}_\eta$ are Lipschitz continuous with constants $L^{(p)}_\eta$ and $R^{(p,q)}_\eta$ respectively. Therefore it follows from Lemma 1 that, $S^{p,q}_\eta$ is also Lipschitz continuous with constant $CB^{(p,q)}_\eta$. Using this property it can be seen that,

$$\left|S^{(p,q)}_\eta\left(\frac{\mathbf{x}+\mathbf{s}}{\mathbf{R}}\right)-S^{(p,q)}_\eta\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\right|\leq\frac{CB^{(p,q)}_\eta||\mathbf{x}||}{R}, \qquad (A.17)$$

for some constant $C\in\mathbb{R}$. In other words,

$$S^{(p,q)}_\eta\left(\frac{\mathbf{x}+\mathbf{s}}{\mathbf{R}}\right)=S^{(p,q)}_\eta\left(\frac{\mathbf{s}}{\mathbf{R}}\right)+O\left(\frac{B^{(p,q)}_\eta||\mathbf{x}||}{R}\right). \qquad (A.18)$$

Incorporating this Lipschitz property of the $\{S_{\eta}^{(p,q)}\}$ into equation (A.16), we obtain,

$$\begin{aligned}\mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) &= \sum_{\eta_1}\sum_{\mathbf{x}}\left(S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)+O\left(\frac{B_{\eta_1}^{(p,q)}||\mathbf{x}||}{R}\right)\right)\left\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}-\mathbf{r}}\psi_{\eta,-\mathbf{r}}\right\}^2 \\ &= \sum_{\eta_1}\sum_{\mathbf{x}}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\left\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}-\mathbf{r}}\psi_{\eta,-\mathbf{r}}\right\}^2+O\left(\frac{1}{R}\right),\end{aligned}$$

since $\sum_{\eta}2^{2j(\eta)}B_{\eta}^{(p,q)}<\infty$ and $\{\sum_{\mathbf{r}}\psi_{\eta_1,\mathbf{x}-\mathbf{r}}\psi_{\eta,-\mathbf{r}}\}^2$ is finite. Expanding the square then gives,

$$\mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) = \sum_{\eta_1}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\left\{\sum_{\mathbf{x}}\sum_{\mathbf{r_1}}\sum_{\mathbf{r_2}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_1}}\psi_{\eta,-\mathbf{r_1}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_2}}\psi_{\eta,-\mathbf{r_2}}\right\}+O\left(\frac{1}{R}\right).$$

Substituting $\mathbf{r_0}=\mathbf{r_2}-\mathbf{r_1}$ gives,

$$\begin{aligned}\mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) &= \sum_{\eta_1}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\sum_{\mathbf{x}}\left\{\sum_{\mathbf{r_1}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_1}}\psi_{\eta,-\mathbf{r_1}}\sum_{\mathbf{r_0}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_1}-\mathbf{r_0}}\psi_{\eta,-\mathbf{r_1}-\mathbf{r_0}}\right\}+O\left(\frac{1}{R}\right) \\ &= \sum_{\eta_1}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\left\{\sum_{\mathbf{r_1}}\sum_{\mathbf{r_0}}\psi_{\eta,-\mathbf{r_1}}\psi_{\eta,-\mathbf{r_1}-\mathbf{r_0}}\sum_{\mathbf{x}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_1}}\psi_{\eta_1,\mathbf{x}-\mathbf{r_1}-\mathbf{r_0}}\right\}+O\left(\frac{1}{R}\right).\end{aligned}$$

Finally we note that each of the summations within the brackets is simply an autocorrelation wavelet (see equation (3.11)). Hence

$$\begin{aligned}\mathbb{E}(I_{\eta,\mathbf{s}}^{(p,q)}) &= \sum_{\eta_1}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)\sum_{\mathbf{r_0}}\Psi_{\eta_1}(\mathbf{r_0})\Psi_{\eta}(\mathbf{r_0})+O\left(\frac{1}{R}\right) \\ &= \sum_{\eta_1}S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta}+O\left(\frac{1}{R}\right),\end{aligned}$$

where A is the inner product matrix given in equation (3.12). Therefore by correcting the cross-periodogram with the inverse of the A matrix it can be an unbiased estimator of the spectrum.

Variance: We now consider the proof of the variance of the cross-periodogram. We begin by recalling the standard definition of the variance, $\mathrm{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$. We have

already established the form of $\mathbb{E}[X]$. Hence we begin by focussing on $\mathbb{E}[X^2]$.

$$\begin{aligned}
\mathbb{E}\left[\left(I_{\eta,\mathbf{s}}^{(p,q)}\right)^2\right] &= \mathbb{E}\left[\left(\left(d_{\eta,\mathbf{s}}^{(p)}\right)\left(d_{\eta,\mathbf{s}}^{(q)}\right)\right)^2\right], \\
&= \mathbb{E}\Bigg[\Bigg(\sum_{\mathbf{r_1}}\sum_{\eta_1}\sum_{\mathbf{u_1}} W_{\eta_1,\mathbf{u_1}}^{(p)}\psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1})\xi_{\eta_1,\mathbf{u_1}}^{(p)}\psi_{\eta,\mathbf{s}}(\mathbf{r_1})\times \\
&\quad \sum_{\mathbf{r'_1}}\sum_{\eta'_1}\sum_{\mathbf{u'_1}} W_{\eta'_1,\mathbf{u'_1}}^{(q)}\psi_{\eta'_1,\mathbf{u'_1}}(\mathbf{r'_1})\xi_{\eta'_1,\mathbf{u'_1}}^{(q)}\psi_{\eta,\mathbf{s}}(\mathbf{r'_1})\Bigg)^2\Bigg], \\
&= \prod_{i=1}^{4}\sum_{\eta_i}\sum_{\mathbf{u_i}}\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)W_{\eta_i}^{M_i}(\mathbf{u_i}/\mathbf{R})\sum_{\mathbf{r_i}}\psi_{\eta_i,\mathbf{u_i}}(\mathbf{r_i})\psi_{\eta,\mathbf{s}}(\mathbf{r_i}),
\end{aligned}$$

where M_i is the following indicator function:

$$M_i = \begin{cases} p & \text{if } i\text{=1,2;} \\ q & \text{if } i\text{=3,4.} \end{cases}$$

Next we consider the term $\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)$. Using a result due to Isserlis (1918), we have,

$$\begin{aligned}
\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right) &= \mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_2,\mathbf{u_2}}^{(p)}\right)\mathbb{E}\left(\xi_{\eta_3,\mathbf{u_3}}^{(q)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)+ \\
&\quad \mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\right)\mathbb{E}\left(\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)+\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)\mathbb{E}\left(\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\right).
\end{aligned}$$

Substituting this decomposition into our expression for $\mathbb{E}\left[\left(I_{\eta,\mathbf{s}}^{(p,q)}\right)^2\right]$ we obtain,

$$\begin{aligned}
\mathbb{E}\left[\left(I_{\eta,\mathbf{s}}^{(p,q)}\right)^2\right] &= \prod_{i=1}^{4}\sum_{\mathbf{r_i}}\sum_{\eta_i}\sum_{\mathbf{u_i}} W_{\eta_i}^{M_i}(\mathbf{u_i}/\mathbf{R})\psi_{\eta_i,\mathbf{u_i}}(\mathbf{r_i})\psi_{\eta,\mathbf{s}}(\mathbf{r_i})\times \\
&\quad \Bigg\{\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_2,\mathbf{u_2}}^{(p)}\right)\mathbb{E}\left(\xi_{\eta_3,\mathbf{u_3}}^{(q)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)+\mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\right)\mathbb{E}\left(\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)+ \\
&\quad \mathbb{E}\left(\xi_{\eta_1,\mathbf{u_1}}^{(p)}\xi_{\eta_4,\mathbf{u_4}}^{(q)}\right)\mathbb{E}\left(\xi_{\eta_2,\mathbf{u_2}}^{(p)}\xi_{\eta_3,\mathbf{u_3}}^{(q)}\right)\Bigg\}, \\
&= I_1 + I_2 + I_3,
\end{aligned}$$

where for example,

$$I_1 = \prod_{i=1}^{4} \sum_{\mathbf{r_i}} \sum_{\eta_i} \sum_{\mathbf{u_i}} \mathbb{E}\left(\xi^{(p)}_{\eta_1,\mathbf{u_1}} \xi^{(p)}_{\eta_2,\mathbf{u_2}}\right) \mathbb{E}\left(\xi^{(q)}_{\eta_3,\mathbf{u_3}} \xi^{(q)}_{\eta_4,\mathbf{u_4}}\right) W^{I_{\eta_i}}_{\eta_i}(\mathbf{u_i}/\mathbf{R}) \psi_{\eta_i,\mathbf{u_i}}(\mathbf{r_i}) \psi_{\eta,\mathbf{s}}(\mathbf{r_i}). \quad \text{(A.19)}$$

Since, $\mathbb{E}\left(\xi^{(p)}_{\eta_1,\mathbf{u_1}} \xi^{(p)}_{\eta_2,\mathbf{u_2}}\right) = \delta_{\eta_1,\eta_2}\delta_{\mathbf{u_1},\mathbf{u_2}}\delta_{p,p} = 1$ and $\mathbb{E}\left(\xi^{(q)}_{\eta_3,\mathbf{u_3}} \xi^{(q)}_{\eta_4,\mathbf{u_4}}\right) = 1$, (A.19) simplifies to,

$$\begin{aligned} I_1 = &\left(\sum_{\eta_1} \sum_{\mathbf{u_1}} \left(W^{(p)}_{\eta_1}\left(\frac{\mathbf{u_1}}{\mathbf{R}}\right)\right)^2 \sum_{\mathbf{r_1}} \psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1}) \psi_{\eta,\mathbf{s}}(\mathbf{r_1}) \sum_{\mathbf{r_2}} \psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_2}) \psi_{\eta,\mathbf{s}}(\mathbf{r_2}) \right) \times \\ &\left(\sum_{\eta_3} \sum_{\mathbf{u_3}} \left(W^{(q)}_{\eta_3}\left(\frac{\mathbf{u_3}}{\mathbf{R}}\right)\right)^2 \sum_{\mathbf{r_3}} \psi_{\eta_3,\mathbf{u_3}}(\mathbf{r_3}) \psi_{\eta,\mathbf{s}}(\mathbf{r_3}) \sum_{\mathbf{r_4}} \psi_{\eta_3,\mathbf{u_3}}(\mathbf{r_4}) \psi_{\eta,\mathbf{s}}(\mathbf{r_4}) \right). \end{aligned}$$

From the proof of the expectation of the cross-periodogram (equation(A.15)) in the case where $p = q$ this is simply equal to $\mathbb{E}\left[I^{(p)}_{\eta,\mathbf{s}}\right] \mathbb{E}\left[I^{(q)}_{\eta,\mathbf{s}}\right]$. Applying the same step with I_2 we obtain,

$$\begin{aligned} I_2 = &\left(\sum_{\eta_1} \sum_{\mathbf{u_1}} \left(S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{u_1}}{\mathbf{R}}\right)\right) \sum_{\mathbf{r_1}} \psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_1}) \psi_{\eta,\mathbf{s}}(\mathbf{r_1}) \sum_{\mathbf{r_3}} \psi_{\eta_1,\mathbf{u_1}}(\mathbf{r_3}) \psi_{\eta,\mathbf{s}}(\mathbf{r_3}) \right) \times \\ &\left(\sum_{\eta_2} \sum_{\mathbf{u_2}} \left(S^{(p,q)}_{\eta_2}\left(\frac{\mathbf{u_2}}{\mathbf{R}}\right)\right) \sum_{\mathbf{r_2}} \psi_{\eta_2,\mathbf{u_2}}(\mathbf{r_2}) \psi_{\eta,\mathbf{s}}(\mathbf{r_2}) \sum_{\mathbf{r_4}} \psi_{\eta_2,\mathbf{u_2}}(\mathbf{r_4}) \psi_{\eta,\mathbf{s}}(\mathbf{r_4}) \right) \\ = &\, \mathbb{E}\left[I^{(p,q)}_{\eta,\mathbf{s}}\right]^2. \end{aligned}$$

I_3 also gives this result. Putting this all together gives,

$$\mathbb{E}\left[\left(I^{(p,q)}_{\eta,\mathbf{s}}\right)^2\right] = \mathbb{E}\left[I^{(p)}_{\eta,\mathbf{s}}\right] \mathbb{E}\left[I^{(q)}_{\eta,\mathbf{s}}\right] + 2\mathbb{E}\left[I^{(p,q)}_{\eta,\mathbf{s}}\right]^2.$$

From here we calculate the variance as follows,

$$\begin{aligned} \mathrm{Var}(I^{(p,q)}_{\eta,\mathbf{s}}) &= \mathbb{E}\left[I^{(p)}_{\eta,\mathbf{s}}\right] \mathbb{E}\left[I^{(q)}_{\eta,\mathbf{s}}\right] + 2\mathbb{E}\left[I^{(p,q)}_{\eta,\mathbf{s}}\right]^2 - \mathbb{E}\left[I^{(p,q)}_{\eta,\mathbf{s}}\right]^2, \\ &= \mathbb{E}\left[I^{(p)}_{\eta,\mathbf{s}}\right] \mathbb{E}\left[I^{(q)}_{\eta,\mathbf{s}}\right] + \mathbb{E}\left[I^{(p,q)}_{\eta,\mathbf{s}}\right]^2, \\ &= \left(\sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right)\left(\sum_{\eta_1} S^{(q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right) + \\ &\quad \left(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right)^2. \end{aligned}$$

From the work of Nason et al. (2000), it is known that $\Psi_\eta(\boldsymbol{\tau}) = O(1)$ uniformly in $\boldsymbol{\tau}$. Hence it follows that,

$$A_{\eta_1,\eta} = \sum_{\boldsymbol{\tau}} \Psi_\eta(\boldsymbol{\tau})\Psi_{\eta_1}(\boldsymbol{\tau}) = O\left(2^{2j(\eta_1)}\right). \tag{A.20}$$

Thus, expanding the square gives,

$$\begin{aligned}\mathrm{Var}(I^{(p,q)}_{\eta,\mathbf{s}}) = &\sum_{\eta_1} S^{(p)}_{\eta_1}\Big(\frac{\mathbf{s}}{\mathbf{R}}\Big)A_{\eta_1,\eta} \sum_{\eta_1} S^{(q)}_{\eta_1}\Big(\frac{\mathbf{s}}{\mathbf{R}}\Big)A_{\eta_1,\eta} + \\ &\left(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\Big(\frac{\mathbf{s}}{\mathbf{R}}\Big)A_{\eta_1,\eta}\right)^2 + O\left(\frac{2^{2j(\eta_1)}}{R}\right).\end{aligned}$$

■

Proof of Theorem 3: The expectation of the smoothed cross-periodogram.

Recall that we are using a Nadaraya-Watson kernel smoother to smooth the (raw) cross-periodogram. Hence we focus on the asymptotic properties of:

$$\begin{aligned}\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \mathbb{E}\left(\sum_{\mathbf{u}} w_{\mathbf{u}} I^{(p,q)}_{\eta,\mathbf{u}}\right) \\ &= \sum_{\mathbf{u}} w_{\mathbf{u}} \mathbb{E}(I^{(p,q)}_{\eta,\mathbf{u}}),\end{aligned}$$

where $w_{\mathbf{u}} = \frac{K_h(\mathbf{s}-\mathbf{u})}{\sum_{\mathbf{u}} K_h(\mathbf{s}-\mathbf{u})}$ and define $K_h(\mathbf{s}-\mathbf{u}) = K(\frac{\mathbf{s}-\mathbf{u}}{h})$ for bandwidth $h > 0$. Defining $\lambda = \sum_{\mathbf{u}} K_h(\mathbf{s}-\mathbf{u})$ and recalling Theorem 2, we find that

$$\begin{aligned}\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{\lambda}\sum_{\mathbf{u}} K_h(\mathbf{s}-\mathbf{u})\ \mathbb{E}(I^{(p,q)}_{\eta,\mathbf{u}}) \\ &= \frac{1}{\lambda}\sum_{\mathbf{u}} K_h(\mathbf{s}-\mathbf{u})\left[\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right].\end{aligned}$$

Setting $\mathbf{u} = \mathbf{s} + \boldsymbol{\tau}$ we obtain:

$$\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) = \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\left[\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}+\boldsymbol{\tau}}{\mathbf{R}}\right)A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right].$$

As discussed in the proof of Theorem 2, $S_\eta^{(p,q)}$ is Lipschitz continuous with respect to the L_1 norm, with constant $CB_\eta^{(p,q)}$. Using this property, the cross-spectral term can be simplified to obtain

$$S_\eta^{(p,q)}\left(\frac{\mathbf{s}+\boldsymbol{\tau}}{\mathbf{R}}\right) = S_\eta^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) + O\left(\frac{B_\eta^{(p,q)}||\boldsymbol{\tau}||}{R}\right). \tag{A.21}$$

Hence

$$\begin{aligned}
\mathbb{E}(\tilde{I}_{\eta,\mathbf{s}}^{(p,q)}) &= \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\left[\sum_{\eta_1}\left(S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) + O\left(\frac{B_{\eta_1}^{(p,q)}||\boldsymbol{\tau}||}{R}\right)\right)A_{\eta_1,\eta} + O\left(\frac{1}{R}\right)\right] \\
&= \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\left[\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta} + \sum_{\eta_1} B_{\eta_1}^{(p,q)}2^{2j(\eta_1)}O\left(\frac{||\boldsymbol{\tau}||}{R}\right) + O\left(\frac{1}{R}\right)\right],
\end{aligned}$$

where $A_{\eta_1,\eta} = O(2^{2j(\eta_1)})$. We then have,

$$\mathbb{E}(\tilde{I}_{\eta,\mathbf{s}}^{(p,q)}) = \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\left[\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta} + ||\boldsymbol{\tau}||O\left(\frac{1}{R}\right) + O\left(\frac{1}{R}\right)\right],$$

since $\sum_{\eta_1} B_{\eta_1}^{(p,q)}2^{2j(\eta_1)} < \infty$. Therefore,

$$\begin{aligned}
\mathbb{E}(\tilde{I}_{\eta,\mathbf{s}}^{(p,q)}) &= \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\left[\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta} + (||\boldsymbol{\tau}|| + 1)O\left(\frac{1}{R}\right)\right] \\
&= \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\Big(\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta}\Big) + \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\Big((||\boldsymbol{\tau}|| + 1)O\left(\frac{1}{R}\right)\Big) \\
&= \tilde{I}_1 + \tilde{I}_2,
\end{aligned} \tag{A.22}$$

where $\tilde{I}_1 = \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\Big(\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta}\Big)$ and $\tilde{I}_2 = \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\Big((||\boldsymbol{\tau}|| + 1)O\left(\frac{1}{R}\right)\Big)$. First we focus on $\tilde{I}_1$:

$$\begin{aligned}
\tilde{I}_1 &= \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})\Big(\sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta}\Big) \\
&= \sum_{\eta_1} S_{\eta_1}^{(p,q)}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A_{\eta_1,\eta},
\end{aligned} \tag{A.23}$$

since $\sum_{\mathbf{u}} w_{\mathbf{u}} = \frac{1}{\lambda}\sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau}) = 1$. In order to evaluate the second term in equation (A.22), we consider the maximum number of lattice points, $(2\lfloor h \rfloor + 1)^2$, in the support of a kernel of bandwidth h. More specifically we divide both the numerator and denominator by this quantity. In so doing we obtain

$$\tilde{I}_2 = \left(\frac{1}{(2\lfloor h \rfloor + 1)^2} \sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau}) \left[(||\boldsymbol{\tau}|| + 1) O\left(\frac{1}{R}\right) \right] \right) \Big/ \left(\frac{\lambda}{(2\lfloor h \rfloor + 1)^2} \right). \tag{A.24}$$

We now analyse the numerator and denominator of $\tilde{I}_2$ separately, denoting these by $\tilde{I}_{2,N}$ and $\tilde{I}_{2,D}$ respectively. The numerator is given by,

$$\begin{aligned}
\tilde{I}_{2,N} &= \frac{1}{(2\lfloor h \rfloor + 1)^2} \sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau}) \Big((||\boldsymbol{\tau}|| + 1) O\left(\frac{1}{R}\right) \Big) \\
&\leq \frac{\sum_{||\boldsymbol{\tau}||_1 \leq \lfloor h \rfloor} K_h(\boldsymbol{\tau}) \Big((\lfloor h \rfloor + 1) O\left(\frac{1}{R}\right) \Big)}{(2\lfloor h \rfloor + 1)^2} \qquad \text{since} ||\boldsymbol{\tau}||_1 = |\tau_1| + |\tau_2| \leq \lfloor h \rfloor \\
&= \frac{\Big((\lfloor h \rfloor + 1) O\left(\frac{1}{R}\right) \Big)}{(2\lfloor h \rfloor + 1)^2} \\
&= \left(\frac{1}{2\lfloor h \rfloor + 1} \right) O\left(\frac{1}{R}\right),
\end{aligned} \tag{A.25}$$

since

$$0 < \sum_{||\boldsymbol{\tau}||_1 \leq \lfloor h \rfloor} K_h(\boldsymbol{\tau}) < \int_{||\boldsymbol{x}||_2 < h} K_h(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

Remark: The kernel is non-zero only for $||\boldsymbol{\tau}||_1 \leq \lfloor h \rfloor$, where h is the bandwidth of the kernel discussed in Section 5.4.

The denominator of $\tilde{I}_2$ is given by,

$$\tilde{I}_{2,D} = \frac{\lambda}{(2\lfloor h \rfloor + 1)^2},$$

where $\lambda = \sum_{\boldsymbol{\tau}} K_h(\boldsymbol{\tau})$. We have,

$$\begin{aligned}
\tilde{I}_{2,D} &< \tfrac{1}{(2\lfloor h \rfloor + 1)^2} \times (2\lfloor h \rfloor + 1)^2 K_M \\
&= O(1),
\end{aligned} \tag{A.26}$$

since $K(x) \leq K_M < \infty$ where K_M is the maximum point of the kernel. Therefore the expectation of the smoothed cross-periodogram is

$$\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) = \sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right) A_{\eta_1,\eta} + \left(\frac{1}{2\lfloor h \rfloor + 1}\right) O\left(\frac{1}{R}\right). \tag{A.27}$$

■

Proof of Proposition 1: The variance of the smoothed cross-periodogram is asymptotically vanishing.

We begin by giving the following definition, which will be useful for the proof of this proposition.

Definition 19 *We define* $A^{\boldsymbol{\tau}}_{\eta_1,\eta} = \sum_{\mathbf{k}} \Psi_{\eta_1,\eta}(\mathbf{k})\Psi_{\eta_1,\eta}(\mathbf{k}+\boldsymbol{\tau})$*, where* $\Psi_{\eta_1,\eta}(\mathbf{k}) = \sum_{\mathbf{t}} \psi_{\eta_1,\mathbf{t}}\psi_{\eta,\mathbf{t}+\mathbf{k}}$ *following the one-dimensional expression given by Fryźlewicz and Nason (2006).*

The variance of the smoothed periodogram is given by:

$$\mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) = \mathbb{E}\left[(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})^2\right] - \left[\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})\right]^2. \tag{A.28}$$

Since $\mathbb{E}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})$ was established in Theorem 3 and is given in equation (A.27), we will focus on $\mathbb{E}\left[(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})^2\right]$.

$$\begin{aligned}
\mathbb{E}\left[(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})^2\right] &= \mathbb{E}\left(\left[\frac{1}{\lambda}\sum_{\mathbf{u}} K_h(\mathbf{s}-\mathbf{u}) I^{(p,q)}_{\eta,\mathbf{u}}\right]^2\right) \\
&= \mathbb{E}\left(\frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v}) I^{(p,q)}_{\eta,\mathbf{u}} I^{(p,q)}_{\eta,\mathbf{v}}\right) \\
&= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left(I^{(p,q)}_{\eta,\mathbf{u}} I^{(p,q)}_{\eta,\mathbf{v}}\right).
\end{aligned}$$

Using the definition of the raw cross-periodogram and arguments based on Isserlis' Theorem similar to those in the proof of the variance of the raw cross-periodogram, this can be

simplified as follows:

$$\begin{aligned}\mathbb{E}\left[(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}})^2\right] &= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}d^{(q)}_{\eta,\mathbf{v}}\right] \\ &= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{u}}\right]\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{v}}d^{(q)}_{\eta,\mathbf{v}}\right] + \\ &\quad \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}\right]\mathbb{E}\left[d^{(q)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{v}}\right] + \\ &\quad \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{v}}\right]\mathbb{E}\left[d^{(q)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}\right].\end{aligned}$$

The first term is simply equal to $\mathbb{E}\left[\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}\right]^2$. Hence the variance of the smoothed periodogram becomes,

$$\begin{aligned}\mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}\right]\mathbb{E}\left[d^{(q)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{v}}\right] + \\ &\quad \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{v}}\right]\mathbb{E}\left[d^{(q)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}\right].\end{aligned}$$

Following the same method as for the expectation of the raw cross-periodogram the expectations can be expressed as,

$$\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(q)}_{\eta,\mathbf{v}}\right] = \sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta} + O\left(\frac{1}{R}\right) \tag{A.29}$$

$$\mathbb{E}\left[d^{(p)}_{\eta,\mathbf{u}}d^{(p)}_{\eta,\mathbf{v}}\right] = \sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta} + O\left(\frac{1}{R}\right), \tag{A.30}$$

where $\boldsymbol{\tau} = \mathbf{u}-\mathbf{v}$ and $A^{\boldsymbol{\tau}}_{\eta_1,\eta}$ is as defined in Definition 19. We also note that using similar arguments to Eckley et al. (2009), $A^{\boldsymbol{\tau}}_{\eta_1,\eta} = O\left(2^{2j(\eta_1)}\right)$.

Remark The terms in the expectations are very similar to the raw cross-periodogram except that there is a lag between the terms.

The variance can thus be expressed as

$$\begin{aligned}\mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\Bigg[\Bigg(\sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\left(\frac{1}{R}\right)\Bigg)\times \\ &\quad \Bigg(\sum_{\eta_1} S^{(q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\left(\frac{1}{R}\right)\Bigg)+\Bigg(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\left(\frac{1}{R}\right)\Bigg)^2\Bigg] \\ &= \frac{1}{\lambda^2}\sum_{\mathbf{u}}\sum_{\mathbf{v}} K_h(\mathbf{s}-\mathbf{u})K_h(\mathbf{s}-\mathbf{v})\Bigg[\sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\sum_{\eta_1} S^{(q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+ \\ &\quad \Bigg(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{u}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)^2+O\left(\frac{2^{2j(\eta_1)}}{R}\right)\Bigg].\end{aligned}$$

Now let $\mathbf{u}=\mathbf{s}+\boldsymbol{\tau'}$, $\mathbf{v}=\mathbf{s}+\boldsymbol{\tau_1}$ so $\boldsymbol{\tau}=\mathbf{u}-\mathbf{v}=\boldsymbol{\tau'}-\boldsymbol{\tau_1}$. It therefore follows that

$$\begin{aligned}\mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{\lambda^2}\sum_{\boldsymbol{\tau'}}\sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'})K_h(\boldsymbol{\tau_1})\Bigg[\Bigg(\sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{s}+\boldsymbol{\tau'}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)\times \\ &\quad \Bigg(\sum_{\eta_1} S^{(q)}_{\eta_1}\left(\frac{\mathbf{s}+\boldsymbol{\tau'}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)+\Bigg(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}+\boldsymbol{\tau'}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)^2+O\left(\frac{2^{2j(\eta_1)}}{R}\right)\Bigg].\end{aligned}$$

Using the Lipschitz properties of $S^{(p,q)}$, we obtain,

$$\begin{aligned}\mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{\lambda^2}\sum_{\boldsymbol{\tau'}}\sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'})K_h(\boldsymbol{\tau_1})\Bigg[\Bigg(\sum_{\eta_1}\left(S^{(p)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)+O\Big(\frac{L^{(p)}_{\eta_1}||\boldsymbol{\tau'}||}{R}\Big)\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)\times \\ &\quad \Bigg(\sum_{\eta_1}\left(S^{(q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)+O\Big(\frac{L^{(q)}_{\eta_1}||\boldsymbol{\tau'}||}{R}\Big)\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)+ \\ &\quad \Bigg(\sum_{\eta_1}\left(S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)+O\Big(\frac{B^{(p,q)}_{\eta_1}||\boldsymbol{\tau'}||}{R}\Big)\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}\Bigg)^2+O\left(\frac{2^{2j(\eta_1)}}{R}\right)\Bigg] \\ &= \frac{1}{\lambda^2}\sum_{\boldsymbol{\tau'}}\sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'})K_h(\boldsymbol{\tau_1})\Bigg[\Bigg(\sum_{\eta_1} S^{(p)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\Big(\frac{||\boldsymbol{\tau'}||}{R}\Big)\Bigg)\times \\ &\quad \Bigg(\sum_{\eta_1} S^{(q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\Big(\frac{||\boldsymbol{\tau'}||}{R}\Big)\Bigg)+\Bigg(\sum_{\eta_1} S^{(p,q)}_{\eta_1}\left(\frac{\mathbf{s}}{\mathbf{R}}\right)A^{\boldsymbol{\tau}}_{\eta_1,\eta}+O\Big(\frac{||\boldsymbol{\tau'}||}{R}\Big)\Bigg)^2+ \\ &\quad O\left(\frac{2^{2j(\eta_1)}}{R}\right)\Bigg],\end{aligned}$$

since $\sum_{\eta_1} L_{\eta_1}^{(p)} 2^{2j(\eta_1)} < \infty$ and $\sum_{\eta_1} B_{\eta_1}^{(p,q)} 2^{2j(\eta_1)} < \infty$. Expanding this equation we find that

$$\begin{aligned}
\mathrm{Var}(\tilde{I}_{\eta,\mathbf{s}}^{(p,q)}) &= \frac{1}{\lambda^2} \sum_{\boldsymbol{\tau'}} \sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'}) K_h(\boldsymbol{\tau_1}) \Bigg[\sum_{\eta_1} S_{\eta_1}^{(p)} \left(\frac{\mathbf{s}}{\mathbf{R}} \right) A_{\eta_1,\eta}^{\boldsymbol{\tau}} \sum_{\eta_1} S_{\eta_1}^{(q)} \left(\frac{\mathbf{s}}{\mathbf{R}} \right) A_{\eta_1,\eta}^{\boldsymbol{\tau}} + \\
&\quad \left(\sum_{\eta_1} S_{\eta_1}^{(p,q)} \left(\frac{\mathbf{s}}{\mathbf{R}} \right) A_{\eta_1,\eta}^{\boldsymbol{\tau}} \right)^2 + O\left(\frac{2^{2j(\eta_1)} ||\boldsymbol{\tau'}||}{R} \right) + O\left(\frac{||\boldsymbol{\tau'}||^2}{R^2} \right) + O\left(\frac{2^{2j(\eta_1)}}{R} \right) \Bigg] \\
&= \frac{1}{\lambda^2} \sum_{\boldsymbol{\tau'}} \sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'}) K_h(\boldsymbol{\tau_1}) \Bigg[O\left(2^{4j(\eta_1)} \right) + O\left(2^{4j(\eta_1)} \right) + (||\boldsymbol{\tau'}|| + 1) O\left(\frac{2^{2j(\eta_1)}}{R} \right) + \\
&\quad (||\boldsymbol{\tau'}||^2) O\left(\frac{1}{R^2} \right) \Bigg],
\end{aligned}$$

since $\sum_{\eta_1} S_{\eta_1}^{(p)}(\mathbf{s}/\mathbf{R}) < \infty$ and $\sum_{\eta_1} S_{\eta_1}^{(p,q)}(\mathbf{s}/\mathbf{R}) < \infty$ and $A_{\eta_1,\eta}^{\boldsymbol{\tau}} = O(2^{2j(\eta_1)})$. In order to show the variance is asympotically vanishing, similarly to the expectation we consider the maximum number of lattice points in the support of the kernel product. We have,

$$\begin{aligned}
\mathrm{Var}(\tilde{I}_{\eta,\mathbf{s}}^{(p,q)}) = \Bigg(\frac{1}{(2\lfloor h \rfloor + 1)^4} \sum_{\boldsymbol{\tau'}} \sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau'}) K_h(\boldsymbol{\tau_1}) \Bigg[O\left(2^{4j(\eta_1)} \right) + O\left(2^{4j(\eta_1)} \right) + \\
(||\boldsymbol{\tau'}|| + 1) O\left(\frac{2^{2j(\eta_1)}}{R} \right) + (||\boldsymbol{\tau'}||^2) O\left(\frac{1}{R^2} \right) \Bigg] \Bigg) \Bigg/ \left(\frac{\lambda^2}{(2\lfloor h \rfloor + 1)^4} \right).
\end{aligned}$$

In order to show consistency of $\tilde{I}$ we consider the numerator and denominator separately denoted as $\tilde{I}_N$ and $\tilde{I}_D$ respectively. The numerator is given by,

$$\begin{aligned}
\tilde{I}_N &\leq \frac{1}{(2\lfloor h \rfloor + 1)^4} \sum_{||\boldsymbol{\tau'}||_1 \leq \lfloor h \rfloor} \sum_{||\boldsymbol{\tau_1}||_1 \leq \lfloor h \rfloor} K_h(\boldsymbol{\tau'}) K_h(\boldsymbol{\tau_1}) \Bigg[O\left(2^{4j(\eta_1)} \right) + O\left(2^{4j(\eta_1)} \right) + \\
&\quad (\lfloor h \rfloor + 1) O\left(\frac{2^{2j(\eta_1)}}{R} \right) + (\lfloor h \rfloor^2) O\left(\frac{1}{R^2} \right) \Bigg] \Bigg) \\
&= \frac{1}{(2\lfloor h \rfloor + 1)^4} \cdot 1 \cdot 1 \cdot \Bigg[O\left(2^{4j(\eta_1)} \right) + O\left(2^{4j(\eta_1)} \right) + (\lfloor h \rfloor + 1) O\left(\frac{2^{2j(\eta_1)}}{R} \right) + (\lfloor h \rfloor^2) O\left(\frac{1}{R^2} \right) \Bigg] \Bigg),
\end{aligned}$$

since,

$$0 < \sum_{||\boldsymbol{\tau}||_1 \leq \lfloor h \rfloor} K_h(\boldsymbol{\tau}) < \int_{||\boldsymbol{x}||_2 < h} K_h(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

The denominator is given by,

$$\tilde{I}_D = \frac{\lambda^2}{(2\lfloor h \rfloor + 1)^4}$$

where $\lambda^2 = \sum_{\boldsymbol{\tau}'} K_h(\boldsymbol{\tau}') \sum_{\boldsymbol{\tau_1}} K_h(\boldsymbol{\tau_1})$. We thus find that:

$$\begin{aligned} \tilde{I}_D &< \tfrac{1}{(2\lfloor h \rfloor+1)^4} \times (2\lfloor h \rfloor + 1)^4 K_M \\ &= K_M \qquad (A.31) \\ &= O(1), \qquad (A.32) \end{aligned}$$

since $K(x) \leq K_M < \infty$. Therefore the variance of the smoothed wavelet cross-periodogram is given by

$$\begin{aligned} \mathrm{Var}(\tilde{I}^{(p,q)}_{\eta,\mathbf{s}}) &= \frac{1}{(2\lfloor h \rfloor + 1)^4} \cdot 1 \cdot 1 \cdot \left[O\left(2^{4j(\eta_1)}\right) + O\left(2^{4j(\eta_1)}\right) + (\lfloor h \rfloor + 1) O\left(\frac{2^{2j(\eta_1)}}{R}\right) + (\lfloor h \rfloor^2) O\left(\frac{1}{R^2}\right) \right] \Bigg) \\ &= \frac{1}{(2\lfloor h \rfloor + 1)^4} \left[O\left(2^{4j(\eta_1)}\right) + (\lfloor h \rfloor + 1) O\left(\frac{2^{2j(\eta_1)}}{R}\right) + (\lfloor h \rfloor^2) O\left(\frac{1}{R^2}\right) \right] \Bigg). \end{aligned}$$

Finally, as $h \to \infty, R \to \infty$ with $h/R \to 0$, $\mathrm{Var}(\tilde{I}) \to 0$. Thus the smoothed wavelet cross-periodogram is asympotically consistent. ■

Proof of Proposition 2: Aymptotics of the local cross-covariance.

$$\begin{aligned} c^{(p,q)}_{\mathbf{R}}(\mathbf{z}, \boldsymbol{\tau}) &= \mathrm{Cov}\left(X^{(p)}_{\mathbf{r}}, X^{(q)}_{\mathbf{r}+\boldsymbol{\tau}}\right) \qquad (A.33) \\ &= \mathbb{E}\left(\left(X^{(p)}_{\mathbf{r}} - \mu^{(p)}_{\mathbf{r}}\right)\left(X^{(q)}_{\mathbf{r}+\boldsymbol{\tau}} - \mu^{(q)}_{\mathbf{r}+\boldsymbol{\tau}}\right)\right). \qquad (A.34) \end{aligned}$$

By the modelling assumptions of LS2Wmv, $\mathbb{E}(X_{\mathbf{r}}) = 0$ for all $\mathbf{r}$. Hence,

$$\begin{aligned} c^{(p,q)}_{\mathbf{R}}(\mathbf{z}, \boldsymbol{\tau}) &= \mathbb{E}\left(X^{(p)}_{\mathbf{r}}, X^{(q)}_{\mathbf{r}+\boldsymbol{\tau}}\right), \\ &= \mathbb{E}\left(\sum_{\eta}\sum_{\mathbf{u}} W^{(p)}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \xi^{(p)}_{\eta,\mathbf{u}} \sum_{\eta}\sum_{\mathbf{u}} W^{(q)}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} \xi^{(q)}_{\eta,\mathbf{u}}\right), \\ &= \sum_{\eta}\sum_{\mathbf{u}} W^{(p)}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) W^{(q)}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} \mathbb{E}\left(\xi^{(p)}_{\eta,\mathbf{u}} \xi^{(q)}_{\eta,\mathbf{u}}\right). \end{aligned}$$

From equation (5.4) we have,

$$\begin{aligned} \mathbb{E}\left(\xi^{(p)}_{\eta,\mathbf{u}} \xi^{(q)}_{\eta,\mathbf{u}}\right) &= \delta_{\eta,\eta}\delta_{\mathbf{u},\mathbf{u}} \rho^{p,q}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right), \\ &= \rho^{p,q}_{\eta}\left(\frac{\mathbf{u}}{\mathbf{R}}\right). \end{aligned}$$

If $p = q$ then $\rho_\eta^{p,q}(\mathbf{u}/\mathbf{R}) = 1$ and so the proof follows as in Eckley et al. (2009). However for the case where $p \neq q$ we have,

$$\begin{aligned} c_{\mathbf{R}}^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) &= \sum_\eta \sum_{\mathbf{u}} W_\eta^{(p)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) W_\eta^{(q)}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \rho_\eta^{p,q}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}}, \\ &= \sum_\eta \sum_{\mathbf{u}} S_\eta^{p,q}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}}. \end{aligned}$$

Finally we consider the absolute difference between the cross-covariance and the local cross-covariance.

$$\begin{aligned} \left| c_{\mathbf{R}}^{(p,q)}(\mathbf{z_1}, \boldsymbol{\tau}) - c^{(p,q)}(\mathbf{z_2}, \boldsymbol{\tau}) \right| &= \left| \sum_\eta \sum_{\mathbf{u}} S_\eta^{p,q}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} - \sum_\eta S_\eta^{p,q}\left(\frac{\mathbf{r}}{\mathbf{R}}\right) \Psi_\eta(\boldsymbol{\tau}) \right|, \\ &= \left| \sum_\eta \sum_{\mathbf{u}} S_\eta^{p,q}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} - \sum_\eta S_\eta^{p,q}\left(\frac{\mathbf{r}}{\mathbf{R}}\right) \sum_{\mathbf{u}} \psi_{\eta,\mathbf{u}} \psi_{\eta,\mathbf{u}-\boldsymbol{\tau}} \right|, \\ &= \left| \sum_\eta \sum_{\mathbf{u}} S_\eta^{p,q}\left(\frac{\mathbf{u}}{\mathbf{R}}\right) \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} - \sum_\eta S_\eta^{p,q}\left(\frac{\mathbf{r}}{\mathbf{R}}\right) \sum_{\mathbf{u}} \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} \right|. \end{aligned}$$

Using the result of Lemma 1 we obtain

$$\left| c_{\mathbf{R}}^{(p,q)}(\mathbf{z_1}, \boldsymbol{\tau}) - c^{(p,q)}(\mathbf{z_2}, \boldsymbol{\tau}) \right| \leq \left| \left(\frac{1}{R}\right) \sum_\eta \sum_{\mathbf{u}} C B_\eta ||\mathbf{u} - \mathbf{r}|| \psi_{\eta,\mathbf{u}-\mathbf{r}} \psi_{\eta,\mathbf{u}-\mathbf{r}-\boldsymbol{\tau}} \right|.$$

Since $\Psi_\eta(\boldsymbol{\tau}) = O(1)$ uniformly in $\boldsymbol{\tau}$ and the support of $\Psi_\eta(\boldsymbol{\tau})$ is bounded by $K 2^{2j(\eta)}$, the distance $||\mathbf{u} - \mathbf{r}||$ is bounded by this amount too. Finally we obtain

$$\begin{aligned} \left| c_{\mathbf{R}}^{(p,q)}(\mathbf{z_1}, \boldsymbol{\tau}) - c^{(p,q)}(\mathbf{z_2}, \boldsymbol{\tau}) \right| &= \left| \left(\frac{1}{R}\right) CK \sum_\eta B_\eta 2^{2j(\eta)} \right|, \\ &= O\left(\frac{1}{R}\right), \end{aligned}$$

since the Lipschitz constants B_η are uniformly bounded in η with $\sum_\eta B_\eta 2^{2j(\eta)} < \infty$ (as stated in Definition 13). ■

Proof of Lemma 2: Invertibility of the local wavelet cross-spectrum.

This proof is identical to the two-dimensional case, considered in Eckley et al. (2010). We begin by considering the term: $\sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\boldsymbol{\tau}} c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) \Psi_{\eta_1}(\boldsymbol{\tau})$. By definition $c^{(p,q)}(\mathbf{z}, \boldsymbol{\tau}) =$

$\sum_\nu S_\nu^{p,q}(\mathbf{z})\Psi_\nu(\boldsymbol{\tau})$. Hence

$$\begin{aligned}\sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\boldsymbol{\tau}} c^{(p,q)}(\mathbf{z},\boldsymbol{\tau})\Psi_{\eta_1}(\boldsymbol{\tau}) &= \sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\boldsymbol{\tau}} \left\{ \sum_{\nu} S_\nu^{p,q}(\mathbf{z})\Psi_\nu(\boldsymbol{\tau}) \right\} \Psi_{\eta_1}(\boldsymbol{\tau}). \\ &= \sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\nu} S_\nu^{p,q}(\mathbf{z}) \sum_{\boldsymbol{\tau}} \Psi_\nu(\boldsymbol{\tau})\Psi_{\eta_1}(\boldsymbol{\tau}).\end{aligned} \tag{A.35}$$

Since $\sum_\eta S_\eta^{p,q}(\mathbf{z}) < \infty \; \forall \mathbf{z}$ and the sum of $\boldsymbol{\tau}$ is finite, the order of the summations in equation (A.35) may be rearranged. Hence,

$$\begin{aligned}\sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\boldsymbol{\tau}} c^{(p,q)}(\mathbf{z},\boldsymbol{\tau})\Psi_{\eta_1}(\boldsymbol{\tau}) &= \sum_{\eta_1} A^{-1}_{\eta,\eta_1} \sum_{\nu} S_\nu^{p,q}(\mathbf{z}) A_{\nu,\eta_1} \\ &= \sum_{\nu} S_\nu^{p,q}(\mathbf{z}) \sum_{\eta_1} A^{-1}_{\eta,\eta_1} A_{\eta_1,\nu} \\ &= \sum_{\nu} S_\nu^{p,q}(\mathbf{z}) \delta_{\eta,\nu} \\ &= S_\eta^{p,q}(\mathbf{z}). \qquad \blacksquare\end{aligned} \tag{A.36}$$

Appendix B

LS2Wstat Software Suite

In many applications in the field of image processing, such as texture discrimination, classification and segmentation, assumptions are made relating to the second-order structure of an image. Many methods require stationarity while those that allow non-stationarity do not tend to perform as well on stationary images as traditional Fourier based methods (Eckley et al., 2010). It is thus important to test this assumption of stationarity before performing further image analysis.

In Chapter 3 a test of spatial stationarity was proposed based on the *locally stationary two-dimensional wavelet* model (LS2W) of Eckley et al. (2010). The test, denoted $\texttt{Bootstat}_{\texttt{LS2W}}$ employs bootstrap resampling under the null hypothesis assumption of stationarity to assess its significance. By performing the $\texttt{Bootstat}_{\texttt{LS2W}}$ test on a *montage* of two image textures, the test enables texture discrimination and it can be used to identify regions in an image of spatial homogeneity by recursively applying the technique within a quadtree image decomposition (Chapter 4). These two elements thus allow the number of textures in an image to be determined even those which contain visually-subtle textures.

In this appendix we describe the package `LS2Wstat` which illustrates the software implementation of work from Chapters 3 and 4, which can currently be downloaded from `http://www.maths.lancs.ac.uk/~taylors8/software.html`.
The package has been developed in R and makes use of several functions within the LS2W package (Eckley and Nason, 2011a,b). Below we provide brief descriptions of the main functions within the `LS2Wstat` package.

- `Sim`: Simulation function for LS2W processes.
- `TOS2D.TS`: A test statistic for spatial stationarity.
- `TOS2D`: Performs the bootstrap stationarity test for images.
- `CompareIm`: Assesses whether two textured images have the same covariance structure.
- `TextCount`: Groups a list of (stationary) images into texture classes based on their covariance properties.
- `imageQT`: Performs an image quadtree decomposition.
- `plot.imageQT`: A plot function for the quadtree decomposition and optionally the textured region classification output from `TextCount`.

Below we illustrate the use of the `LS2Wstat` package by way of a simulated example, based on a non-stationary texture. This is generated by the simulation function `Sim` described in Section B.1. Section B.2 describes the key functions used to test an image for stationarity (Chapter 3), whilst Section B.3 describes the functions associated with finding the number of textures in an image (Chapter 4).

B.1 Simulating LS2W processes

Sim

Simulation function for LS2W processes.

Description

Several different spatially stationary and non-stationary random fields can be generated with the `Sim` function.
The stationary processes which can be simulated are:

`S1:` a random normal process of specified standard deviation, `sd`;

`S2:` a spatial moving average process with parameter `rho`;

`S3:` an isotropic random field with a Matérn covariance with shape parameter `nu`;

`S4:` a Gaussian random field with exponential covariance, range parameter 2;

`S5:` a diagonal Haar moving average process of a specified order `order` and standard deviation `sd` (see the `Haar2MA.diag` function in the `LS2W` package for more detail).

We can also generate several non-stationary processes:

`NS1:` a random process with unit standard deviation on the first half-plane, concatenated with a random normal half-plane of standard deviation `sd`;

`NS2:` a white noise half-plane concatenated with a Matérn stationary process;

`NS3:` a *Haar Montage* of specified standard deviation `sd` (see the `LS2W` `HaarMontage` function for more details);

`NS4:` a process with a slowly-varying covariance structure across the horizontal axis of the image; this structure is specified by a function `fn` with arguments `start`, `end`, and `a`, describing the start, end and "gradient" of the change from `start` to `end`. Current variance functions included in the `LS2Wstat` package for the `fn` argument are `scurve`

and `lincurve` representing an "s"-shaped curve and linear function respectively. See Chapter 3 for examples of this covariance structure.

`NS5`: a white noise process with a central subregion of random Normal deviates with non-unit standard deviation `sd` and subregion size `n/prop`;

`NS6`: a white noise process with a subregion (of size `n/prop`) of random Normal deviates with non-unit standard deviation `sd` in the middle section of the top left quadrant;

`NS7`: a process similar to `NS5`, except that there is an additional texture in a subregion of the image. In other words, the image is a montage of three two-dimensional Normal processes with differing standard deviations. The base texture is again of unit variance, whereas the other two textures have standard deviations `sd` and `sd2`. The sizes of the subregions (expressed as a proportion of the original image dimension) can be specified with the `prop` and `prop2` arguments to the `Sim` function. Furthermore, the positions of the two inserted textures can be given. Possible options are "a", "b", "c", "d", "e" which corresponds to (a) top-right, (b) bottom-right, (c) top-left, (d) bottom-left and (e) centred. A more exact location may be specified by inputting pos=c(x,y), which represents the position in pixels from the top-left of the image (i.e. c(x,y) puts Image A x pixels down and y pixels across from the top-left corner of Image B.)

Usage

```
Sim(n, sd = 1, K = 150, imtype = "S1", ...)
```

Required arguments

`n`

The dimension of the image to be generated.

`sd`

The standard deviation of base texture or of the increments of the LS2W process to be generated (see Section 3.4.1).

`K=150`

The number of images to generate.

`imtype="S1"`

The type of image(s) to create. Must be one of `"S1"`, `"S2"`, `"S3"`, `"S4"`, `"S5"`, `"NS1"`, `"NS2"`, `"NS3"`, `"NS4"`, `"NS5"`, `"NS6"` or `"NS7"`.

Optional arguments

The other optional arguments for `Sim` are as follows:

`type`

The type of neighbourhood dependence for the random field, either "queen" or "rook" (see the `cell2nb` function documentation in the `spdep` package for more details).

`rho`

Moving average parameter for the process `S2`.

`nu`

Shape parameter for the Matérn covariance for process `S3`.

`order`

Haar moving average order for `S5`.

`fn`

`scurve` or `lincurve` for `NS4`.

`start`

Start value for `NS4` (passed into `scurve` or `lincurve`).

`end`

End value for `NS4` (passed into `scurve` or `lincurve`).

`a`

"Gradient" for `NS4` (passed into `scurve` or `lincurve`).

`prop`

Proportion of inserted subimage for `NS5`, `NS6` and the first subimage in `NS7`.

`sd2`

Standard deviation of second inserted subimage for `NS7`.

`prop2`

Proportion of second inserted subimage for `NS7`.

`pos1`

Position of first inserted subimage for `NS7`.

`pos2`

Position of second inserted subimage for `NS7`.

Value

`images`

A list of length `K`, with each list entry being an image of dimension `n x n` with the chosen spectral structure.

Examples

```
> X <- Sim(512, K=1, imtype="NS4", sd=1.6, prop=0.25)[[1]]
> image(plotmtx(X), col=grey(255:0/256))
```

The simulated image `X` is shown in Figure B.1. It displays a realisation of `NS4`: a white noise process with a subregion of random Normal deviates in the center of the process with a standard deviation of 1.6. In particular, we consider an image of dimension 512×512 with a subregion of dimension 128×128. Throughout the remainder of this appendix we shall apply our methods to this image.

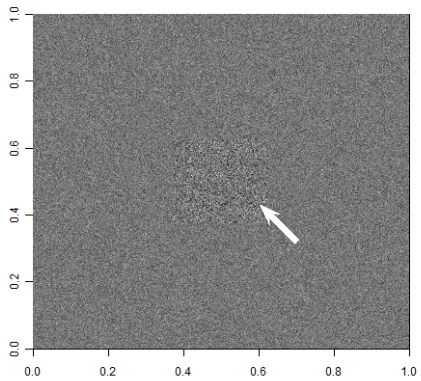


Figure B.1: An example of a textured image (`NS4`) simulated with the `Sim` function.

B.2 Testing the spatial stationarity of images

In this section we describe the implementation of the $\texttt{Bootstat}_{\texttt{LS2W}}$ test of stationarity. Throughout this section let us assume that we have an image $X_{\mathbf{r}}$ we wish to test for spatial stationarity. Since the model in equation (3.1) assumes the process is zero mean, if necessary the image should be detrended, for example by estimating the spatial mean and retaining the residual image for analysis. This can be achieved in R by using the core `stats` package function `medpolish`, which implements Tukey's median polish technique (Tukey, 1977).

In practice the test statistic for stationarity as introduced in equation (3.16) in Chapter 3 is computed on an (unbiased) estimate of the local wavelet spectrum, produced by the LS2W function `cddews`. For a (square) image `X`, the test statistic is calculated using the function `TOS2D.TS`.

TOS2D.TS

A test statistic for spatial stationarity.

Description

The test statistic given in equation (3.16) for a test of stationarity can be seen as the mean empirical variance of the spectrum estimate, where the average is taken over all scale-direction pairs. It is computed for use in the bootstrap testing procedure (`TOS2D`).

Usage

```
TOS2D.TS(spectrum)
```

Required arguments

`spectrum`

A local wavelet spectrum estimate, i.e. a `cddews` object.

Value

`statistic`

The value of the test statistic for the given spectrum.

Examples

```
# Generate a cddews object of image X and find the value of the test statistic.
> TSvalue<-TOS2D.TS(cddews(X,smooth=FALSE))
> TSvalue
  [1] 0.205782
```

TOS2D

Performs the bootstrap stationarity test ($\texttt{Bootstat}_{\texttt{LS2W}}$) for images.

Description

This function first crops an image (if necessary) to have dyadic dimensions. The test statistic (`theTS`), which should be based upon the local wavelet spectrum, is calculated for this original image as shown in the example of `TOS2D.TS`. The local wavelet spectrum under the null hypothesis of stationarity is calculated, so as to be able to simulate realisations under the null hypothesis. `bsims` images are simulated and test statistics are found for each. The function returns all the test statistic values, which may be passed to `TOS2D.pval` in order to find a p-value for the test. For full details on this testing procedure, see Chapter 3.

Usage

```
TOS2D(image, bsims = 100, smooth = FALSE, verbose = TRUE, theTS = TOS2D.TS,
levels, zeromean = FALSE)
```

Required arguments

`image`

The image you want to analyse.

`bsims`

The number of bootstrap simulations to carry out. By default this takes the value 100.

`smooth`

Whether or not to carry out wavelet periodogram smoothing.

`verbose`

If TRUE, informative messages are printed.

`theTS`

This specifies the test statistic function to be used within the testing procedure to measure non-stationarity. The test statistic should be based on the local wavelet spectrum and by default is the function `TOS2D.TS` representing the statistic in equation (3.16) in Chapter 3.

`levels`

This (optionally) specifies the wavelet transform levels which are smoothed when estimating the wavelet spectrum of the image.

`zeromean`

This specifies whether to use Tukey's median polish method (Tukey, 1977) to remove the image trend.

Value

`TOS2D`

This is a vector of length `bsims+1` containing each of the test statistics calculated in the bootstrap test. The first entry is the value of the test statistic computed on the original image while the remaining entries are test statistic values for the simulated images.

Examples

```
# Run test of stationarity on image X.
> Xbstest <- TOS2D(X, bsims=100, smooth=FALSE)
# The object Xbstest is thus a vector of length 101.
```

To compute a p-value using the outputted bootstrap test statistic values, we use the function `TOS2D.pval`.

TOS2D.pval

Computes a p-value for the output of the test for stationarity.

Description

This function returns the parametric bootstrap p-value for the test, from the test statistics provided by counting those test statistic values less than the test statistic value for the original image (see Davison et al. (1999) for more details).

Usage

```
TOS2D.pval(BS2D, alpha=0.05, verbose=TRUE)
```

Required arguments

`BS2D`

The vector of test statistics, such as the output from `TOS2D`. The first value must be the value of the test statistic for the original image.

`alpha`

A size for the test, by default this is `0.05`.

`verbose`

If TRUE then the p-value is printed and a sentence declaring "stationary" or "not stationary" is printed (default is `TRUE`).

Value

`p`

The p-value of the test (See Algorithm 3.1).

Examples

```
# We can find the p-value for the test relating to the image X using
# the object Xbstest created previously. We have
> pval <- TOS2D.pval(Xbstest)
```

```
#Realized Bootstrap is  0.205782
#p-value is  0
#Image was NOT stationary
#
#In this case we print out the outcome of the test and use the default
#test size of alpha=0.05.
#The results of this test for X indicates that the image was assessed
#to be non-stationary as expected.
```

The function `TOS2D.bin` performs the test of stationarity as in `TOS2D` but outputs a binary value indicating whether the image is stationary or not. This function is useful in finding the number of textures in an image (Section B.3).

TOS2D.bin

A function to assess the stationarity of an input image.

Description

The function performs the LS2W test of stationarity contained in the function `TOS2D`. A binary value indicating the outcome of the test is returned.

Usage

```
TOS2D.bin(x, bs = 250, alpha = 0.05, ...)
```

Required arguments

`x`

The image to be tested for stationarity.

`bs`

The number of bootstrap samples to use in the `TOS2D` test.

`alpha`

The significance level of the hypothesis test.

`...`

Any other optional arguments to `TOS2D`.

Value

`test.out`

The outcome of the test: `TRUE` indicates stationary; `FALSE` indicates non-stationary.

Examples

```
# We perform the test of stationarity on image X.
> X.test<-TOS2D.bin(X, bs=100, smooth=FALSE)
> X.test
[1] FALSE
# This indicates the image is non-stationary as expected.
```

B.3 Determining the number of textures in an image using the bootstrap test of stationarity

In this section we describe a method for distinguishing unique textures in a number of (stationary) images. This procedure, in combination with the quadtree algorithm discussed below counts the number of textures within an image.

A quadtree algorithm implementation

In essence, a region splitting algorithm recursively subdivides an input image into smaller regions, with the subdivision decisions being based on some statistical criterion. More specifically in a *quadtree representation*, at each stage, a (sub)image is divided into its four subquadrants if the criterion is not satisfied. In our case the statistical criterion we use is (lack of) homogeneity, that is, a quadrant is further divided if it is considered as non-stationary by the $\texttt{Bootstat}_{\texttt{LS2W}}$ test using `TOS2D.bin`. This procedure segments an image into regions of spatial stationarity.

For the first subdivison each subimage is of size n/2 x n/2. The sizes of the regions halve in size at each progressive division but increase in number. The R function in `LS2Wstat` which creates the quadtree structure is `imageQT`.

imageQT

Performs an image quadtree decomposition.

Description

This function works by assessing an image for spatial homogeneity. If it is not homogeneous, the image is split into its four subquadrants. Each of these is then tested for homogeneity. The heterogeneous subimages are then again subdivided and tested again. This procedure is repeated until either all subimages are deemed stationary or until the minimum testing size `minsize` is reached.

This particular way of splitting an image has a convenient indexing representation to identify

the position of subregions within an image. If a subimage is subdivided into four quadrants, we assign it a base 4 label as follows: 0 - top-left quadrant; 1 - bottom-left quadrant; 2 - top-right quadrant; 3 - bottom-right quadrant. By continuing in this manner, we can assign an index to each tested subregion, with the number of digits in the index indicating how many times it has been subdivided from the "root" of the tree (the original image). This indexing system is illustrated for a quadtree decomposition in Figure B.2.

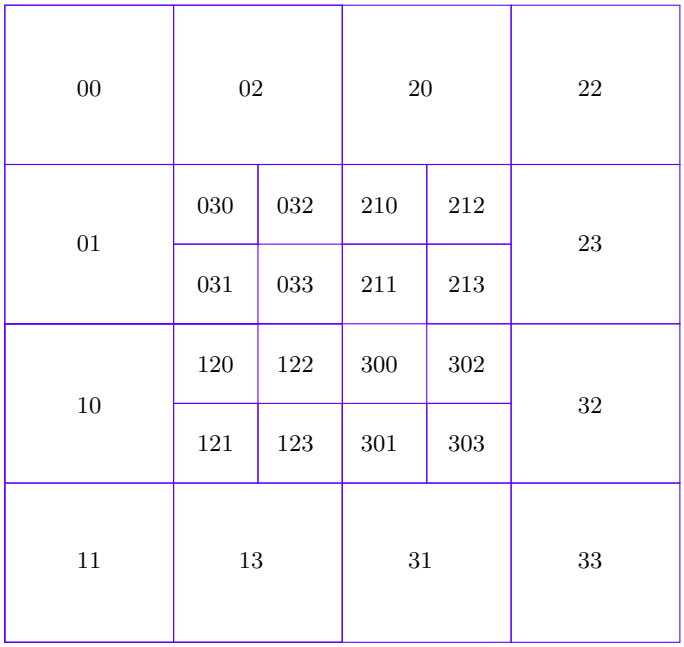


Figure B.2: An example of a quadtree decomposition. The location of the subimages in the decomposition are described by the indexing system described in the text.

Usage

```
imageQT(image, binfun = TOS2D.bin, minsize = 16, ...)
```

Required arguments

`image`

An image to be decomposed.

`binfun`

A function to assess regions for spatial homogeneity, for example `TOS2D.bin`. This function should return a boolean value.

`minsize`

The smallest region size to test for homogeneity.

`...`

Any other optional arguments to `binfun`.

Value

`indl`

The index representation of the non-stationary images in the quadtree decomposition.

`resl`

The results of the stationarity testing (from `binfun`) during the quadtree decomposition. The results giving FALSE correspond to those contained in the `indl` component and the results giving TRUE correspond to those contained in the `indS` component.

`imsize`

The original image dimension.

`imS`

The stationary subimages in the quadtree decomposition.

`indS`

The index representation of the stationary images in the quadtree decomposition.

Examples

```
# Consider the code below to decompose the (non-stationary) input image
# X given in Figure B.1:
> QTdecX<-imageQT(X,binfun=TOS2D.bin, smooth=FALSE, bs=100)
#This will implement a quadtree decomposition, stopping after
#testing regions of 16 x 16 pixels.
> QTdecX$indl
[[1]]
```

```
[1] "0" "1" "2" "3"
[[2]]
[1] "03" "12" "21" "30"
[[3]]
character(0)
#These are the non-stationary sub-images.
> QTdecX$resl
[1] 0
[[2]]
[1] FALSE FALSE FALSE FALSE
[[3]]
 [1] TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
[13] FALSE  TRUE  TRUE  TRUE
[[4]]
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[16] TRUE
#The results giving FALSE correspond to the non-stationary sub-images
#and the results giving TRUE correspond to the stationary ones.
#The final output of this quadtree decomposition is illustrated in Figure B.2,
#which displays all the stationary sub-images.
```

Discrimination between textures

To assess the similarity between two textures using the `LS2Wstat` package, we can use the `CompareIm` function.

CompareIm

Assesses whether two textured images are the same texture.

Description

An image *montage* of two images is created, and the (binary) homogeneity measure `TOS2D.bin` is used to assess stationarity of the montage. If the image is assessed as stationary, the two images are considered as the same texture.

Usage

```
CompareIm(Im1, Im2, testsize = min(nrow(Im1), nrow(Im2)), ...)
```

Required arguments

`Im1`

The first image to be compared.

`Im2`

The second image to be compared.

`testsize`

The size of the combined image montage to be tested for stationarity.

`...`

Any other optional arguments to `TOS2D.bin`.

Value

`montageres` A boolean value indicating whether the montage of `Im1` and `Im2` is stationary.

Example

```
# We simulate realisations of the two textures contained within Figure B.1:
> X1<-Sim(64,K=1,imtype="S1")[[1]]
> X2<-Sim(64,K=1,imtype="S1", sd=1.6)[[1]]
# Use the test to compare them:
```

```
> test<-CompareIm(X1,X2,bs=100, smooth=FALSE)
> test
[1] FALSE
# They are considered different as expected.
```

By repeating this procedure for montages formed from sequential pairs of images in a set, all images can be assigned a texture label. In order to assign texture labels to a number of textures in an image list, the `LS2Wstat` package contains the function `TextCount`.

TextCount

Groups a list of (stationary) images into texture classes.

Description

The function takes a *list* of images to which we wish to assign texture labels. The function optionally detrends the input images (using the `medpolish` function from the **stats** package) by specifying the `medpol` argument. Note that if the candidate textures were, for example, output from the $\texttt{Bootstat}_{\texttt{LS2W}}$ stationarity test within the quadtree algorithm, or had initially been detrended, we would want to set `medpol = FALSE`. The procedure recursively uses the function `CompareIm` to decide whether two images are of the same texture or not. More specifically, the first image is sequentially tested with all others in the list, assigning the images the label "1" if assessed as the same texture as the first image. All other (unclassified) images are then similarly compared with candidates from different texture classes, until all images have been assigned a group label. Testing recursively in this way, there are at most `choose(length(Imgs),2)` comparisons performed, but in reality the number could be a lot fewer.

Usage

```
TextCount(Imgs, medpol = TRUE, ...)
```

Required arguments

`Imgs`

A list of images to classify into textures.

`medpol`

A boolean value indicating whether to zero mean the images (with Tukey's median polish) prior to classification.

`...`

Any other optional arguments to the discrimination function `CompareIm`.

Value

`Iclass`

A vector (of length `length(Imgs)`) of texture labels corresponding to each image in `Imgs`.

Example

```
# We now have a list of candidate textures from the Quad tree procedure
# which we wish to classify. To find out which images represent the same
# texture, one would use the code:
> texclass<-TextCount(QTdecX$imS, medpol = FALSE, smooth=FALSE)
> texclass
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1
#In this example, there were two textures to classify into texture groups.
#The TextCount function automatically determines that there are two unique
#textures in the list.
```

Plotting image segmentation results

Suppose we have performed the quadtree decomposition and the texture discrimination techniques described previously. The `LS2Wstat` package includes a function to plot quadtree decompositions and optionally the textured region classification output from `TextCount`.

plot.imageQT

A plot function for quadtree decompositions.

Description

The function plots the chosen quadtree decomposition, and optionally the textured region classification output from `TextCount`.

Usage

```
plot(x, cires, unclassval = 0, class = TRUE, QT = FALSE, return = FALSE, qtl
= 4, ...)
```

Required arguments

`x`

A quadtree decomposition object, such as output from `imageQT`.

`cires`

Results of `TextCount` for the classification of subimages produced by the quadtree decomposition.

`unclassval`

A value for unclassified values in a quadtree decomposition.

`class`

A boolean value indicating whether to plot the results from `TextCount`. If the classification output is plotted (`class=TRUE`), each textured region is uniquely coloured according to its texture group.

`QT`

A boolean value indicating whether to overlay the full quadtree decomposition onto the figure.

`return`

A boolean value indicating whether to return the matrix associated to the plotted image.

`qtl`

Colour specification for the lines drawn in the image segmentation (for `QT=TRUE`).

`...`

Any other arguments passed to the core R plotting function `image`.

Value

`immat`

The matrix associated to the plotted image. This only occurs if return=TRUE.

Examples

```
#The simulated textured image $X$ is shown in Figure B.1.
#The quadtree and texture classification techniques were then performed.
#The quadtree plotting function plot.imageQT can be used as follows:
>plot(QTdecX, texclass, class=T, QT=TRUE)
>plot(QTdecX, texclass, class=T, QT=FALSE)
```

The quadtree decomposition from this example is shown in Figure B.3; the same decomposition is shown together with the texture classification from `TextCount` in Figure B.4.

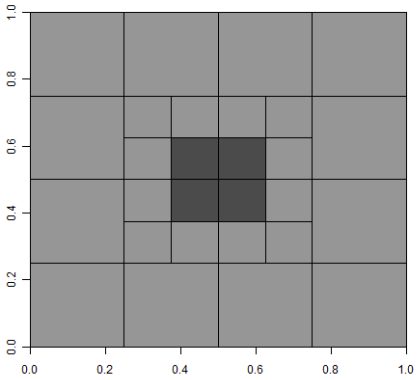


Figure B.3: An example of a quad-tree decomposition using `imageQT`.

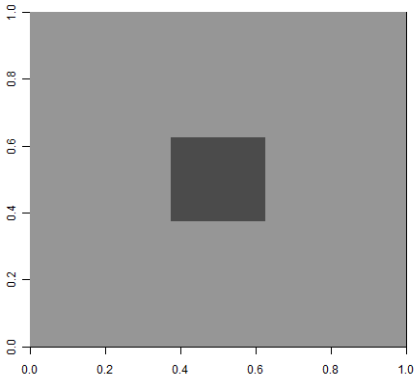


Figure B.4: An example of a quad-tree decomposition with subimage texture classification using `TextCount`.

Bibliography

Abouelela, A., Abbas, H. M., Eldeeb, H., Wahdan, A. A., and Nassar, S. M. (2005). Automated vision system for localizing structural defects in textile fabrics. *Pattern Recognition Letters*, 26(10):1435–1443.

Abramovich, F., Bailey, T., and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *Journal of the Royal Statistical Society (Series D)*, 49(1):1–29.

Abramovich, F., Sapatinas, T., and Silverman, B. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society (Series B)*, 60(4):725–749.

Akhloufi, M. A., Ben Larbi, W., and Maldague, X. (2007). Framework for color-texture classification in machine vision inspection of industrial products. In *IEEE International Conference on Systems, Man and Cybernetics, ISIC.*, pages 1067–1071. IEEE.

Anselin, L. (2005). Spatial Regression Analysis in R–A workbook. *Urbana*, 51:61801.

Arivazhagan, S. and Ganesan, L. (2003). Texture segmentation using wavelet transform. *Pattern Recognition Letters*, 24(16):3197–3203.

Arivazhagan, S., Ganesan, L., and Angayarkanni, V. (2005). Color texture classification using wavelet transform. In *Sixth International Conference on Computational Intelligence and Multimedia Applications*, pages 315–320. IEEE.

Barber, S. and Nason, G. (2004). Real nonparametric regression using complex wavelets. *Journal of the Royal Statistical Society (Series B)*, 66(4):927–939.

Bennett, J. and Khotanzad, A. (1998). Multispectral random field models for synthesis and

analysis of color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):327–332.

Bianconi, F., González, E., Fernández, A., and Saetta, S. (2012). Automatic classification of granite tiles through colour and texture features. *Expert Systems with Applications*, 39:11212–11218.

Bivand, R., Anselin, L., Berke, O., Bernat, A., Carvalho, M., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Lewin-Koh, N., et al. (2011). spdep: Spatial dependence: weighting schemes, statistics and models.

Blanc, R., Da Costa, J. P., Stitou, Y., Baylou, P., and Germain, C. (2008). Assessment of texture stationarity using the asymptotic behavior of the empirical mean and variance. *IEEE Transactions on Image Processing*, 17(9):1481–1490.

Bose, S. and Steinhardt, A. O. (1996). Invariant tests for spatial stationarity using covariance structure. *IEEE Transactions on Signal Processing*, 44(6):1523–1533.

Bouman, C. and Liu, B. (1991). Multiple Resolution Segmentation of Textured Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):99–113.

Brillinger, D. (2001). *Time series: data analysis and theory*, volume 36. Society for Industrial and Applied Mathematics.

Cardinali, A. and Nason, G. P. (2011). Costationarity of locally stationary time series. *Journal of Time Series Econometrics*, 2(2).

Chan, C. and Pang, G. K. (2000). Fabric defect detection by Fourier analysis. *IEEE Transactions on Industry Applications*, 36(5):1267–1276.

Chang, T. and Kuo, C. C. J. (1993). Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing*, 2(4):429–441.

Chatfield, C. (2004). *The Analysis of Time Series*. Chapman & Hall, Boca Raton.

Chen, P. C. and Pavlidis, T. (1979). Segmentation by Texture using a Co-occurrence Matrix and a Split-and-Merge Algorithm. *Computer Graphics and Image Processing*, 10(2):172–182.

Chipman, H., Kolaczyk, E., and McCulloch, R. (1997). Adaptive Bayesian Wavelet Shrinkage. *Journal of the American Statistical Association*, 92(440):1413–1421.

Cho, H. and Fryźlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22:207–229.

Cho, H. and Fryźlewicz, P. (2013). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. In submission.

Clyde, M., Parmigiani, B., and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85(2):391–401.

Coggins, J. M. and Jain, A. K. (1985). A spatial filtering approach to texture analysis. *Pattern Recognition Letters*, 3(3):195–203.

Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81.

Coifman, R. R. and Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–719.

Cross, G. and Jain, A. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1):25–39.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. (2012). Locally Stationary Processes. In Subba Rao, T., Subba Rao, S., and Rao, C. R., editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 351–413. Elsevier.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–966.

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

Daugman, J. G. (1990). An information-theoretic view of analog representation in striate cortex. *Computational neuroscience*, 2:403–424.

Davison, A., Hinkley, D., and Canty, A. J. (1999). *Bootstrap Methods and their Application*. Cambridge University Press.

Ding, S., Yang, J., and Xu, Q. (2005). Color texture analysis using the wavelet-based hidden Markov model. *Pattern Recognition Letters*, 26(11):1710–1719.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Drimbarean, A. and Whelan, P. (2001). Experiments in colour texture analysis. *Pattern Recognition Letters*, 22(10):1161–1167.

Dwivedi, Y. and Subba Rao, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis*, 32(1):68–91.

Eckley, I. A. and Nason, G. P. (2005). Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Statistics and Computing*, 15(2):83–92.

Eckley, I. A. and Nason, G. P. (2011a). LS2W: Implementing the locally stationary 2D wavelet process approach in R. *Journal of Statistical Software*, 43(3):1–23.

Eckley, I. A. and Nason, G. P. (2011b). LS2W: Software implementation of locally stationary wavelet fields. R package version 1.1-2.

Eckley, I. A., Nason, G. P., and Treloar, R. L. (2009). Technical appendix to locally stationary wavelet fields with application to the modelling and analysis of image texture. Technical Report 09:12, University of Bristol.

Eckley, I. A., Nason, G. P., and Treloar, R. L. (2010). Locally stationary wavelet fields with application to the modelling and analysis of image texture. *Journal of the Royal Statistical Society (Series C)*, 59(4):595–616.

Ephraty, A., Tabrikian, J., and Messer, H. (1996). A test for spatial stationarity and applications. In *8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pages 412–415. IEEE Computer Society.

Ephraty, A., Tabrikian, J., and Messer, H. (2001). Underwater Source Detection using a Spatial Stationarity test. *The Journal of the Acoustical Society of America*, 109:1053–1063.

Field, D. J. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Society of London (Series A)*, 357(1760):2527–2542.

Fisher, R. (1936). The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188.

Freixenet, J., Muñoz, X., Raba, D., Martí, J., and Cufí, X. (2002). Yet another survey on image segmentation: Region and boundary information integration. *Computer Vision–ECCV 2002*, pages 408–422.

Fryźlewicz, P. and Nason, G. P. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society (Series B)*, 68:611–634.

Fryźlewicz, P. and Ombao, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *Journal of the American Statistical Association*, 104(485):299–312.

Fryźlewicz, P., Van Bellegem, S., and Von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55(4):737–764.

Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, 96(1):30–54.

Gaspari, G. and Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

Gelfand, A., Schmidt, A., Banerjee, S., and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312.

Gelfand, A. and Vounatsou, P. (2003). Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis. *Biostatistics*, 4(1):11–25.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167–1177.

Gonzalez, R. C. and Woods, R. E. (2001). *Digital Image Processing*. Prentice Hall, 2nd edition.

Greco, F. P. and Trivisano, C. (2009). A multivariate car model for improving the estimation of relative risks. *Statistics in medicine*, 28(12):1707–1724.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

He, H. and Chen, Y. Q. (2000). Unsupervised texture segmentation using resonance algorithm for natural scenes. *Pattern Recognition Letters*, 21(8):741–757.

Hernandez, O., Cook, J., Griffin, M., Rama, C. D., and McGovern, M. (2005). Classification of color textures with random field models and neural networks. *Journal of Computer Science and Technology*, 5:150–157.

Higham, N. (1988). *Matrix nearness problems and applications*. University of Manchester Department of Mathematics.

Hunt, R. (1998). *Measuring Colour*. Fountain Press, England, 3rd edition.

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(1/2):134–139.

Jansen, M. and Oonincx, P. (2005). *Second generation wavelets and applications*. Springer Verlag.

Jin, X., Carlin, B., and Banerjee, S. (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics*, 61(4):950–961.

Johnstone, I. and Silverman, B. (2004). Needles and straw in haystacks: Empirical bayes estimates of possibly sparse sequences. *The Annals of Statistics*, 32(4):1594–1649.

Johnstone, I. and Silverman, B. (2005). Empirical Bayes selection of wavelet thresholds. *The Annals of Statistics*, 33:1700–1752.

Jolion, J. M. and Rosenfeld, A. (1994). *A pyramid framework for early vision: multiresolutional computer vision*. Kluwer Academic Publishers.

Killick, R., Eckley, I., and Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. In submission.

Kingsbury, N. (1999). Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London (Series A)*, 357(1760):2543–2560.

Kingsbury, N. C., Baraniuk, R. G., and Selesnick, I. W. (2005). The Dual-Tree Complex Wavelet Transform. *IEEE Signal Processing Magazine*, 22(6):123–151.

Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis*, 112:76–91.

Knight, M. and Nason, G. (2009). A nondecimated lifting transform. *Statistics and Computing*, 19:1–16.

Knight, M., Nunes, M., and Nason, G. (2012). Spectral estimation for locally stationary time series with missing observations. *Statistics and Computing*, 22(4):877–895.

Laine, A. and Fan, J. (1993). Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1191.

Lawton, W. (1993). Applications of Complex Valued Wavelet Transforms to Subband Decomposition. *IEEE Transactions on Signal Processing*, 41(12):3566–3568.

Liang, Z., Bingang, X., Chi, Z., and Feng, D. (2012). Intelligent characterization and evaluation of yarn surface appearance using saliency map analysis, wavelet transform and fuzzy artmap neural network. *Expert Systems with Applications*, 39(4):4201–4212.

Lina, J. and Mayrand, M. (1995). Complex Daubechies Wavelets. *Applied and Computational Harmonic Analysis*, 2(3):219–229.

Lina, J. M. (1997). Image Processing with Complex Daubechies Wavelets. *Journal of Mathematical Imaging and Vision*, 7(3):211–223.

Lu, C., Chung, P., and Chen, C. F. (1997). Unsupervised texture segmentation via wavelet transform. *Pattern Recognition*, 30(5):729–742.

Majumdar, A. and Gelfand, A. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology*, 39(2):225–245.

Majumdar, A., Paul, D., and Bautista, D. (2010). A generalized convolution model for multivariate nonstationary spatial processes. *Statistica Sinica*, 20(2):675–695.

Mallat, S. G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87.

Mallat, S. G. (1989b). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mallat, S. G. (1999). *A wavelet tour of signal processing*. Academic Press.

Mao, J. and Jain, A. K. (1992). Texture classification and segmentation using multiresolution simultaneous autogressive models. *Pattern Recognition*, 25(2):173–188.

Mardia, K. (1988). Multi-dimensional Multivariate Gaussian Markov Random Fields with Application to Image Processing. *Journal of Multivariate Analysis*, 24(2):265–284.

Matérn, B. (1960). Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden fran statens Skogsforskningsinstitut*, 49(5):144.

Meyer, Y. (1985). Principe d'incertitude, bases Hilbertiennes et algèbres d'opératuers. *Sèminaire Bourbaki*, 28:209–223.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.

Mondal, D. and Percival, D. B. (2012). Wavelet variance analysis for random fields. *IEEE Transactions on Image Processing*, 21(2):537–549.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, 9(1):141–142.

Nason, G. (2013). *wavethresh: Wavelets statistics and transforms.* R package version 4.6.2.

Nason, G. and Cardinali, A. (2010). *costat: Time series costationarity determination and tests of stationarity.* R package version 1.1-1.

Nason, G. P. (2008). *Wavelet methods in statistics with R*. Springer Verlag.

Nason, G. P. and Silverman, B. W. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, 3(2):163–191.

Nason, G. P. and Silverman, B. W. (1995). The stationary wavelet transform and some applications. In Antoniadis, A. and Oppeheim, G., editors, *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistic*, pages 281–300. New York: Springer-Verlag.

Nason, G. P., Von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society (Series B)*, 62(2):271–292.

Nunes, M., Knight, M., and Nason, G. (2006). Adaptive lifting for nonparametric regression. *Statistics and Computing*, 16(2):143–159.

Ombao, H., Raz, J., von Sachs, R., and Guo, W. (2002). The SLEX model of a non-stationary random process. *Annals of the Institute of Statistical Mathematics*, 54(1):171–200.

Ombao, H., Sachs, R. V., and Guo, W. (2005). SLEX analysis of Multivariate Nonstationary Time Series. *Journal of the American Statistical Association*, 100(470):519–530.

Pal, N. R. and Pal, S. K. (1993). A review of Image Segmentation techniques. *Pattern Recognition*, 26(9):1277–1294.

Palm, C. (2004). Color texture classification by integrative co-occurrence matrices. *Pattern Recognition*, 37(5):965–976.

Palmer, S., Zhang, J., and Wang, X. (2011). New methods for objective evaluation of fabric pilling by frequency domain image processing. *Research Journal of Textile and Apparel*, 13(1):11–23.

Panjwani, D. and Healey, G. (1995). Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 17(10):939–954.

Paschos, G. (1998). Chromatic correlation features for texture recognition. *Pattern Recognition Letters*, 19(8):643–650.

Percival, D. B. and Walden, A. T. (2006). *Wavelet methods for time series analysis*, volume 4. Cambridge University Press.

Petrou, M. and Sevilla, P. G. (2006). *Image processing: dealing with texture*. John Wiley & Sons Inc.

Pietikainen, M., Maenpaa, T., and Viertola, J. (2002). Color texture classification with color histograms and local binary patterns. In *Workshop on Texture Analysis in Machine Vision*, pages 109–112.

Priestley, M. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society (Series B)*, 27:204–237.

Priestley, M. B. and Rao, T. S. (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(1):140–149.

Reményi, N. and Vidakovic, B. (2013). Bayesian Wavelet Shrinkage Strategies: A Review. In Shen, X. and Zayed, A. I., editors, *Multiscale Signal Analysis and Modeling*, pages 317–346. Springer.

Rushing, J. A., Ranganath, H. S., Hinke, T. H., and Graves, S. J. (2001). Using association rules as texture features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):845–858.

Sain, S. (2009). Parameter estimation for multivariate spatial lattice models. *Biometrics*.

Sain, S. and Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259.

Salari, E. and Ling, Z. (1995). Texture segmentation using hierarchical wavelet decomposition. *Pattern Recognition*, 28(12):1819–1824.

Sanderson, J., Fryźlewicz, P., and Jones, M. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, 97(2):435–446.

Saxton, W. O. and Baumeister, W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. *Journal of Microscopy*, 127(2):127–138.

Scheunders, P., Livens, S., Van de Wouwer, G., Vautrot, P., and Van Dyck, D. (1998). Wavelet-based texture analysis. *International Journal on Computer Science and Information Management*, 1(2):22–34.

Schlather, M. (2012). Randomfields: Simulation and Analysis of Random Fields. R package version 2.0.0.

Selesnick, I. (2001). Hilbert transform pairs of wavelet bases. *IEEE Signal Processing Letters*, 8(6):170–173.

Selesnick, I. (2002). The design of Hilbert transform pairs of wavelet bases. *IEEE Transactions on Signal Processing*.

Sengur, A. (2008). Wavelet transform and adaptive neuro-fuzzy inference system for color texture classification. *Expert Systems with Applications*, 34(3):2120–2128.

Slutsky, E. (1925). Über stochastische asymptoten und grenzwerte. *Metron*, 5:3–89.

Sonka, M., Boyle, R., and Hlavac, V. (1999). *Image processing, Analysis, and Machine Vision*. PWS Publishing, 2nd edition.

Suen, P. and Healey, G. (2000). The analysis and recognition of real-world textures in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):491–503.

Sweldens, W. (1996). Wavelets and the lifting scheme: a 5 minute tour. *ZAMM-Zeitschrift fur Angewandte Mathematik und Mechanik*, 76(2):41–44.

Sweldens, W. (1997). The lifting scheme: a construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29:511–546.

Taylor, S. L., Eckley, I. A., and Nunes, M. A. (2013). A test of stationarity for textured images. Under revision.

Tikhonov, A. and Arsenin, V. (1977). *Solutions of Ill-Posed Problems*. Winston.

Tikhonov, A., Goncharski, A., Stepanov, V., and Kochikov, I. (1987). Ill-posed image processing problems. *Soviet Physics - Doklady*, 32:456–458.

Tuceryan, M. and Jain, A. K. (1998). Texture analysis. In Chen, C. H., Pau, L. F., and Wang, P. S. P., editors, *Handbook of Pattern Recognition and Computer Vision*, pages 207–248. World Scientific Publishing, 2nd edition.

Tukey, J. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Unser, M. (1986). Local linear transforms for texture measurements. *Signal Processing*, 11(1):61–79.

Unser, M. (1995). Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 4(11):1549–1560.

Van de Wouwer, G., Scheunders, P., Livens, S., and Van Dyck, D. (1999). Wavelet correlation signatures for color texture characterization. *Pattern Recognition*, 32(3):443–451.

Van Heel, M., Keegstra, W., Schutter, W., and van Bruggen, E. J. F. (1982). Arthropod hemocyanin structures studied by image analysis. *Life Chemistry Reports, Supplement*, 1:69–73.

Ver Hoef, J. and Barry, R. (1998). Constructing and Fitting Models for Cokriging and Multivariable Spatial Prediction. *Journal of Statistical Planning and Inference*, 69(2):275–294.

Vidakovic, B. (1999). *Statistical Modelling by Wavelets*. Wiley: New York.

Wackermagel, H. (2003). *Multivariate geostatistics: An introduction with applications*. Springer, 2nd edition.

Wackernagel, H. (2003). *Multivariate geostatistics*. Springer Verlag, 3rd edition.

Walker, P. (1986). *The Theory of Fourier Series and Integrals*. Wiley: Chichester.

Watson, G. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26:359–372.

Whitcher, B. and Craigmile, P. (2004). Multivariate spectral analysis using hilbert wavelet pairs. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(04):567–587.

Whitcher, B., Craigmile, P., and Brown, P. (2005). Time-varying spectral analysis in neurophysiological time series using hilbert wavelet pairs. *Signal Processing*, 85(11):2065–2081.

Zhang, J. and Modestino, J. W. (1990). A model-fitting approach to cluster validation with application to stochastic model-based image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):1009–1017.

Zhang, J. and Oe, S. (1998). Texture image segmentation method based on wavelet transform and neural networks. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 5, pages 4595–4600. IEEE.