UNIVERSITY OF TROMSØ UIT

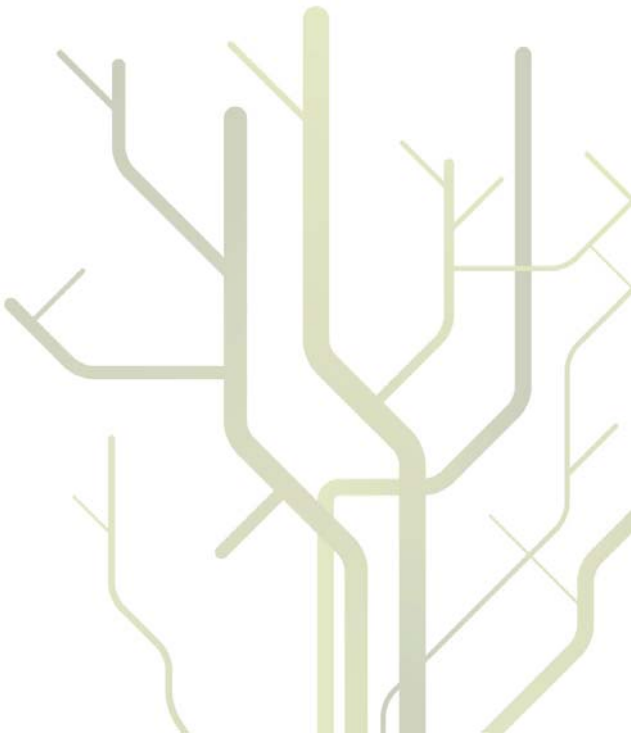
FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF MATHEMATICS AND STATISTICS

Multiscale Methods for Statistical Inference on Regular Lattice Data

Kevin Thon

A dissertation for the degree of Philosophiae Doctor

August 2013



ABSTRACT

This thesis presents methods for multiscale statistical inference on random fields on a regular two-dimensional lattice. There are two distinct concepts of scale that are used in the thesis. The first one is connected to the computer vision community's understanding of scale-space as a family of smooths of a digital image, with fine structure being revealed at low levels of smoothing and the coarser structures standing out at high levels. The second way scale enters is through use of the two-dimensional wavelet transform, and the different scales can be though of as providing information on the energy content in different frequency bands.

All the methods that are developed herein are in the form of statistical hypothesis tests. Paper I uses a Bayesian framework, and statistically significant gradient and curvature is determined through thresholding their posterior probabilities. The last two papers use standard frequentist methods, in Paper II for determining if a random field is isotropic, meaning that the second order statistical properties are independent of direction. Paper III addresses the problem of determining if two samples from a random field are realizations from the same underlying random field.

Much emphasis is on computational efficiency issues, since the proposed methods are intended to be practical for use on large images as typically attained from modern digital cameras. This is achieved at the expense of some flexibility, restricting the models to be that of stationary or intrinsically stationary random fields on a regular grid, and for all the methods developed herein there will be a region of pixels near the image borders where inference is not valid.

The methods are demonstrated in practical examples, for Paper I on dermascopic images of skin lesions where it is used to detect hairs and dots/globules, potentially important components of a system for evaluating the risk of melanoma. In Paper II the test for isotropy is applied to paper density images of handsheets, useful for determining the quality of the paper. Finally, the methodology from Paper III is applied on small projections of the cosmic microwave background (CMB) temperature map from the Planck mission, and offers an alternative test for the overall isotropy of the CMB.

ACKNOWLEDGMENTS

The writing of this thesis would not have been possible without the help of a great number of people. I would like to thank my main supervisor, Professor Fred Godtliebsen, for giving me the opportunity to pursue my research interests and for reining me in when necessary. His relentless optimism throughout the process has been an important factor for this work reaching completion. He has facilitated my attendance to international conferences, and was also instrumental in making my stay in Seattle as a visiting scholar at the University of Washington possible. At UW I initiated a collaboration with my subsequent co-supervisor, Professor Don Percival who, in addition to being an overall great guy who made me and my family feel very welcome in Seattle, had the ideas upon which papers two and three are based. He has always given valuable feedback in our biweekly Skype sessions, and for all this he has my sincerest gratitude. My other co-supervisor, Stein Olav Skrøvseth, has been the only one of my supervisors to share office building with me at NST, and as such has been privileged to listen to the more mundane day to day frustrations of a Ph.D. candidate. For this and his innumerable contributions in technical discussions, I thank him. In the early days of my Ph.D., I had the privilege of spending a week with Håvard Rue at NTNU. In this highly productive week the seeds were sown for what became the first paper in this thesis, and I am also indebted to him for allowing me access to his computational resources at NTNU. Vedad Hadziavidic was originally my co-supervisor, and he is also the one who encouraged me to apply for the Ph.D position in the first place. Although I have at times cursed him for it, I have finally reached a place from where I can give him my heartfelt thanks.

I have been lucky to have been surrounded by good colleagues these last years. Marc Geilhufe and I have shared office on two continents, and our collaboration resulted in the second paper in this thesis. Kristian Hindberg deserves thanks for good discussions and advice in general, but mostly for making me aware that the transposition operator in `Matlab`, will in fact give the complex conjugate transpose when applied to a complex matrix. If only I could get those weeks of my life back. To all the people I have shared office with, Jörn, Kajsa, and Maciel, as well as the above mentioned, thank you for your patience with the finger drumming.

The last years have not been all work, and I am grateful to my friends Bjørn, Stian, Asbjørn, Jonne, Jussi, Trond, Torgeir, and Vegar, among many others, for providing an adequate amount of distractions from the world of statistics. A high point in every possible way was the three peak hike.

Without the support of family, this thesis could never have been written. During our stay in Seattle, we were fortunate to have my aunt and uncle, Roisin and Dominic, close by in Victoria, and I cannot thank them enough for their hospitality. Having small children is a source of tremendous joy, and I thank my parents-in-law who on many occasions have been willing to relieve us of some of the joy. My own parents have been nothing but supportive, and I am forever grateful to my mother for the many times she took care of Olav so that I didn't lose touch with my research entirely when

on paternal leave. Lastly, and most importantly, I would like to thank my wife Kjersti who, as well as her love and support, has given me Halvard and Olav in the course of this Ph.D. These little boys, while perhaps not supporters of a good night's sleep, have enriched my life more ways than I can tell. Finalizing a project like this does not come without sacrifices, and I would not blame Kjersti for feeling that she has been something of a single mother these last months. In the future I promise to heed the words of Don Corleone and strive to be a real man: *A man who doesn't spend time with his family can never be a real man.* And, needless to say, I will never take sides with anyone against the family. Ever.

Kevin – Tromsø, August 30, 2013

LIST OF PUBLICATIONS

Paper I Bayesian multiscale analysis of images modeled as Gaussian Markov random fields, **Thon, K., Rue, H., Skrøvseth, S. O. and Godtliebsen, F.**, *Computational Statistics and Data Analysis*, 2012, Vol. 56, pp. 49-61

Paper II A multiscale wavelet-based test for isotropy of random fields on a regular lattice, **Thon, K., Geilhufe, M. and Percival, D. B.**, being revised for *IEEE Transactions on Image Processing*

Paper III A scale-based test for equality of Gaussian random fields with application to the cosmic microwave background radiation, **Thon, K., Percival, D. B., Geilhufe, M. and Skrøvseth, S. O.**, unpublished manuscript

CONTENTS

Part I

INTRODUCTION

1

BACKGROUND

All three papers that make up the main part of this thesis are concerned with the development of methods for multiscale statistical inference for random fields on regular two-dimensional grids. In Papers II and III scale enters through wavelet filtering operations, while the concept of scale in Paper I is closely related to the scale-space concept within the field of computer vision, where it refers to a family of smooths of a digital image. Scale-space ideas gained widespread popularity with the seminal work of Lindeberg (1994). The important insight underlying the methods is that different features will appear at different scales, and only appear as meaningful entities over certain ranges of scales. Rather than searching for the, in some sense, optimal scale, the scale-space mode of thinking demands that a wide range of scales are considered, and higher scales will reveal coarse structures whereas the lower scales will reveal details. Scale is, in this context, introduced through convolution of the image with kernels of different bandwidths, with the bandwidth of the kernel determining the scale.

An important development in scale-space methodology came, from a statistical viewpoint, with the introduction of SiZer (SIgnificant ZERo crossings of derivatives) in Chaudhuri and Marron (1999), where scale-space ideas are combined with statistical methods to find significant structures at different levels of smoothing. The SiZer methodology has later been refined and expanded upon in, e.g., Chaudhuri and Marron (2000); Hannig and Marron (2006); Hannig and Lee (2006), and in Godtliebsen et al. (2002) and Godtliebsen et al. (2004) the methodology was developed for the two-dimensional cases of, respectively, bivariate density estimation and determining significant structure in images. The above can be considered as traditional scale-space methods, with the bandwidth of a smoothing kernel defining the scale. A somewhat different approach was taken in Godtliebsen and Øigård (2005), and later improved in most respects in Øigård et al. (2006). Here, the underlying signal is given a prior distribution, and inference is done on multiple scales, where the scale is defined in terms of the parameter defining the prior through a local neighborhood structure. As in SiZer, the scale determines the level of smoothing, but the Bayesian framework allows for valid inference on much lower scales. Paper I in this thesis essentially develops the above methods for two dimensions, where the prior image model is that of a Gaussian Markov random field. The focus is on finding statistically significant gradient and curvature in the form of forward differences and, as such, has much in common with Godtliebsen and Øigård (2005). Computationally, the method is closer to the methods of Øigård et al. (2006).

Also in wavelet analysis the concept of scale is central. In the spatial context, a wavelet analysis offers a decomposition that is localized in space as well as in scale, which is related to frequency, and is often contrasted with the standard Fourier transform, which is only localized in frequency. Traditionally, the wavelet transform has therefore found use in non-stationary signal analysis. A two-dimensional wavelet transform is generally constructed by the successive application of one-dimensional differencing (wavelet) and/or averaging (scaling) filters to the rows and columns of the image (Mondal and Percival, 2012). The differencing filters are approximate band-pass filters, while the scaling filters are low-pass filters (Percival and Walden, 2000). Considering stationary random fields, this allows for an interpretation of the scale as representing particular frequency bands. Although there exist explicit connections between traditional scale-space theory and some wavelet transforms (Ferdman et al., 2007), this will not be explored in this thesis. We note informally that convolution with a Gaussian kernel represents a low-pass filtering operation, so traditional scale-space representations can also be thought of in terms of a frequency decomposition. Papers II and III in this thesis develop a multiscale wavelet-based test for isotropy and equality of (Gaussian) random fields, respectively.

The spatial models used in Papers I-III will be presented in Section 1.1. In Section 1.2 relevant background theory regarding the wavelet transform will be given, and in Section 2 all three papers will be discussed and summarized, including suggestions for further developments. Part II contains the full manuscripts of Papers I,II and III, and constitutes the main part of this thesis.

1.1 SPATIAL MODELS

The underlying image model in all three papers in this thesis is that of a stationary or intrinsically stationary Gaussian Field (GF) on a regular two-dimensional lattice $\{X_{u,v} : (u,v) \in \mathbb{Z}^2\}$. A GF is normally specified in terms of its mean function $\mu(\cdot)$ and covariance function $C(\cdot)$. Stationary fields are defined in terms of having constant mean $\mu(\cdot) = \mu$ and covariance function that is a function only of the relative displacement of the sites, $C(\boldsymbol{\kappa}) = \text{cov}(X_{u,v}, X_{u+\kappa_1,v+\kappa_2})$ for all lags $\boldsymbol{\kappa} = (\kappa_1, \kappa_2) \in \mathbb{Z}^2$ (Cressie, 1993). Following Künsch (1987), the more general class of intrinsically stationary fields of order d are defined in terms of all increment processes of order d being stationary. Intrinsically stationary fields of unit order, common in geostatistics, are specified through the *semivariogram* $\gamma(\boldsymbol{\kappa}) = \frac{1}{2}\text{var}(X_{u,v} - X_{u+\kappa_1,v+\kappa_2})$. If, in addition, the covariance function or semivariogram is a function only of the Euclidean distance $|\boldsymbol{\kappa}|$ between the sites, the fields are called isotropic. Specification of a field through the covariance function allows for a intuitive interpretation of the fields properties. These types of models are commonly used in geostatistics and spatial statistics in general (Chilès and Delfiner, 2012; Cressie, 1993).

A different modeling strategy is to use a Gaussian Markov random field (GMRF) model, which is a discretely indexed Gaussian field where the full conditionals $\pi(X_i|X_{-i})$ depend only on a set of neighbors ∂i to each site i (Rue and Held, 2005). For our purposes $i = (u,v) \in \mathbb{Z}^2$,

and X_{-i} denotes the field excluding location i. The advantages of GMRF models are mainly computational: The number of non-zero elements in the *precision* (inverse covariance) matrix $\mathbf{Q}$ will depend directly on the neighborhood structure, since $Q_{ij} \neq 0 \Leftrightarrow i \in \partial j \cup j$. Usually, the neighborhood will be relatively small, meaning that $\mathbf{Q}$ will be sparse, and this allows for the use of highly efficient sparse matrix methods for calculating most quantities of interest. GMRFs can be specified a number of ways. In *conditional autoregressions* (CAR) models, pioneered by Besag (1974), it is done implicitly through specifying the full conditionals, subject to some constraints to ensure the semi-positive definiteness of the precision matrix. Alternatively, the GMRF may be specified in terms of a normal increment process, from which the full conditionals follow. This is the approach that has been favored in Paper I, where the two image models under consideration are defined in terms of having first or second order normal increments. Due to their computationally attractive qualities, GMRF models have found use in a wide range of applications, including spatial statistics and image analysis, see, e.g., Rue and Held (2005) for an overview.

1.2 WAVELET VARIANCE ANALYSIS

In Papers II and III the tests that are developed are based on the wavelet variances of two-dimensional random fields on a regular lattice. In this section some further background regarding wavelet variances will be given, with the focus being on the interpretation of the scaling and wavelet filters as low-pass and band-pass filters, respectively. The wavelet transform described here is the maximal overlap discrete wavelet transform (MODWT), which is a modified version of the traditional discrete wavelet transform (DWT) (Percival and Walden, 2000). In contrast with the DWT, there is no output sub-sampling with the MODWT, thus making it a highly redundant and non-orthogonal wavelet transform. The main disadvantage of the MODWT compared to the DWT is an increase in computational burden. Since there exists a highly efficient pyramid algorithm, with the same computational complexity as the fast Fourier transform (FFT), the computational requirements are usually manageable. Also, the redundancy of the MODWT requires the storage of large amounts of output data. On the other hand, a circular shift of the input to the DWT can lead to a different DWT-based power spectrum, while the MODWT-based power spectrum is invariant to shifting. The partial DWT, i.e., the standard DWT stopped after J_0 repetitions, requires the length of the input to be a integer multiple of 2^{J_0}. A desirable property of the MODWT is that there are no restrictions on the size of the input input (Percival and Walden, 2000). Although we will consistently be using the MODWT, it should be noted that there exist a multitude of transforms that are similar, or practically identical, but under different names, see, e.g., (Shensa, 1992; Beylkin, 1992; Unser, 1995; Coifman and Donoho, 1995; Nason and Silverman, 1995; Pesquet et al., 1996; Bruce and Gao, 1996).

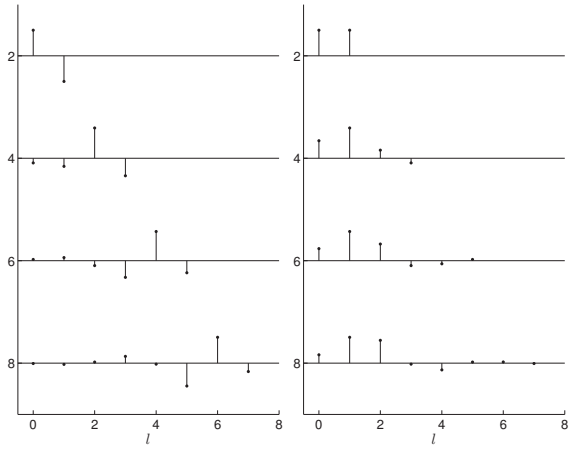


Figure 1: Wavelet (left) and scaling (right) D(L) filters for $L = 2, 4, 6, 8$.

The particular unit-level wavelet filters $\{h_{1,l}\}$ we are considering are of the *Daubechies* class. These are the filters having the maximal number of vanishing moments for a given support (Daubechies, 1992), and additionally satisfy the following criteria:

$$\sum_{l=0}^{L-1} h_{1,l} = 0, \quad \sum_{l=0}^{L-1} h_{1,l}^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=0}^{L-1} h_{1,l} h_{1,l+2n} = \sum_{l=-\infty}^{\infty} h_{1,l} h_{1,l+2n} = 0, \tag{1}$$

with L being the length of the filter, and n being any non-zero integer. The corresponding scaling filter, $\{g_{1,l}\}$, is the *quadrature mirror* filter of $\{h_l\}$,

$$g_{1,l} \equiv (-1)^{l+1} h_{1,L-1-l}.$$

Fig. 1 displays the unit-level wavelet and scaling filters for the Daubechies D(L) (see below) filters of indicated length L. Note that the factor $1/2$ in the second relation in Equation (1) follows since we are considering the MODWT. The squared gain functions $\mathcal{H}(f)$ and $\mathcal{G}(f)$ of $\{h_{1,l}\}$ and $\{g_{1,l}\}$ are defined as the squared modulus of their transfer functions $H(f)$ and $G(f)$, i.e.,

$$\mathcal{H}(f) = |H(f)|^2 \quad \text{with} \quad H(f) = \sum_{l=0}^{L-1} h_{1,l} e^{-i2\pi fl},$$
$$\mathcal{G}(f) = |G(f)|^2 \quad \text{with} \quad G(f) = \sum_{l=0}^{L-1} g_{1,l} e^{-i2\pi fl}.$$

Higher level wavelet and scaling filters are obtained from the unit-level filters through the relations,

$$H_j(f) = H(2^{j-1}f)\prod_{l=0}^{j-2} G(2^l f) \quad \text{and} \quad G_j(f) = \prod_{l=0}^{j-1} G(2^l f).$$

As shown in Percival and Walden (2000), the squared gain function of the unit-level Daubechies scaling filter is

$$\mathcal{G}(f) \equiv \cos^L(\pi f) \sum_{l=0}^{L/2-1} \binom{\frac{L}{2}-1+l}{l} \sin^{2l}(\pi f), \tag{2}$$

with

$$\binom{a}{b} \equiv \frac{a!}{b!(a-b)!},$$

where the associated squared gain function of the wavelet filter follows from the relation $\mathcal{H}(f) = \mathcal{G}(f+\frac{1}{2})$;

$$\mathcal{H}(f) \equiv \sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{\frac{L}{2}-1+l}{l} \cos^{2l}(\pi f).$$

In other words, unit-level Daubechies class scaling (and wavelet) filters differ only in their phase response $\theta^{(G)}(f)$, since

$$G(f) = |\mathcal{G}(f)|^{\frac{1}{2}} e^{i\theta^{(G)}(f)}.$$

Using different criteria for the *spectral factorization* of $\mathcal{G}(f)$ leads to different phase responses, giving the *extremal phase* D(L) and the *least asymmetric* LA(L) wavelets. The technicalities regarding the spectral factorization process for obtaining the Daubechies wavelets are intricate, and are beyond the scope of this introduction (see, e.g., Daubechies (1992) or Percival and Walden (2000) for an in depth exposition). For our purposes, the LA(L) filters are the most interesting. They are constructed such that, of all length L filters with squared gain function as in Equation (2), their phase is as close as possible to being linear. For $L \leq 6$ the D(L) and LA(L) filters are identical, also in phase. Fig. 2 displays the phase responses of the D(8) (left) and LA(8) (right) scaling filters, with a dashed line indicating the closest linear phase function, illustrating the near linearity of the LA(8) filter. Linearity in phase is highly desirable in many applications, since this corresponds to all frequency components of the input being shifted equally, thus allowing the output to be aligned with the parts of the underlying signal it represents, a property that is exploited in Paper III.

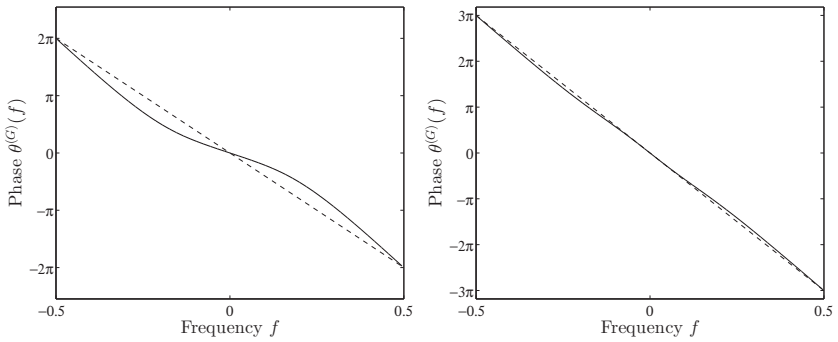


Figure 2: Phase response of D(8) (left) and LA(8) (right) scaling filters. The dashed line indicates the closest linear phase function.

The two-dimensional MODWT of a random field $\{X_{u,v}\}$ is defined in terms a tensor product of the one-dimensional wavelet and/or scaling filters, yielding the wavelet-wavelet (ww), scaling-wavelet (sw), wavelet-scaling (ws), and scaling-scaling (ss) coefficient processes,

$$W_{j,j'}(u,v) = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} h_{j,l}h_{j',l'}X_{u-l,v-l'}, \tag{3}$$

$$U_{j,j'}(u,v) = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} g_{j,l}h_{j',l'}X_{u-l,v-l'}, \tag{4}$$

$$V_{j,j'}(u,v) = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} h_{j,l}g_{j',l'}X_{u-l,v-l'}, \tag{5}$$

$$Z_{j,j'}(u,v) = \sum_{l=0}^{L_j-1}\sum_{l'=0}^{L_{j'}-1} g_{j,l}g_{j',l'}X_{u-l,v-l'}, \tag{6}$$

where $L_j = (2^j-1)(L-1)+1$ and $L_{j'}$ denote the length of the level j and j' wavelet and scaling filters, respectively. The level j wavelet filters essentially produce differences on scale $\tau_j = 2^{j-1}$, while the level j scaling filters average on scale $2\tau_j$. The wavelet variances are defined simply as the variances of the coefficient processes, i.e., with C denoting any of the coefficients from Equations (3) – (6), $\nu^2_{C,j,j'} = \text{var}\{C_{j,j'}(u,v)\}$. The wavelet variances can be used to decompose the variance of the underlying process, and Mondal and Percival (2012) derived the following relationships for stationary random fields,

$$\text{var}(X_{u,v}) = \sum_{j=1}^{\infty}\sum_{j'=1}^{\infty} \nu^2_{W,j,j'}, \tag{7}$$

$$\text{var}(X_{u,v}) = \sum_{j=1}^{\infty} \nu^2_{W,j,j} + \sum_{j=1}^{\infty} \nu^2_{U,j,j} + \sum_{j=1}^{\infty} \nu^2_{V,j,j}. \tag{8}$$

Since the level j wavelet filters are approximate band-pass filters with, assuming unit-distance between neighboring points, pass-band $(1/2^{j+1}, 1/2^j)$, while the level j scaling filters are ap-

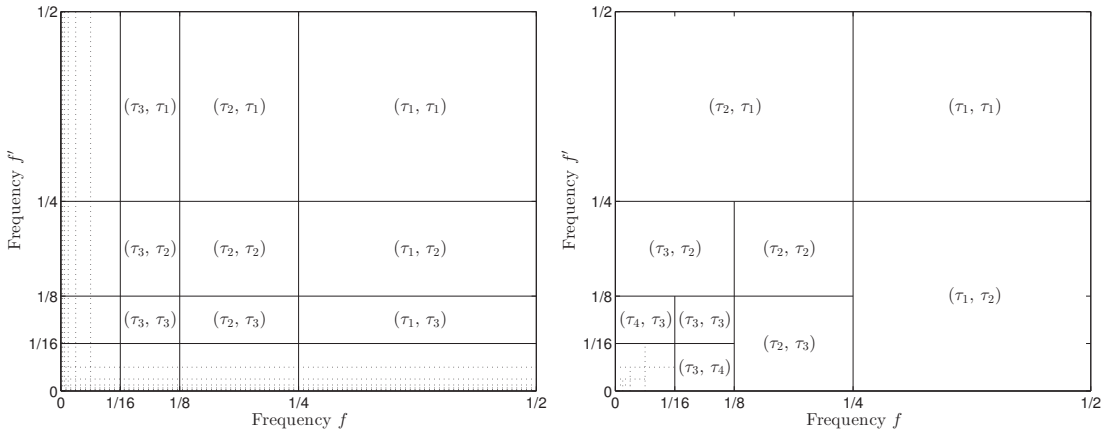


Figure 3: Partitioning of the frequency domain from Equation (7) (left) and (8) (right)

proximate low-pass filters with cutoff at $f = 1/2^{j+1}$, these decompositions of variance correspond to two different partitionings of the frequency domain, as illustrated in Fig. 3. Through the Wiener-Khinchin relation and standard filtering results, the spectral density function (SDF) of the coefficient processes in Equations (3) – (5) may be written out in terms of the transfer functions and SDF of the underlying field, i.e,

$$
\begin{aligned}
S_{W_{jj'},W_{jj'}}(f,f') &= |H_j(f)|^2 |H_{j'}(f')|^2 S_X(f,f'),\\
S_{U_{jj'},U_{jj'}}(f,f') &= |G_j(f)|^2 |H_{j'}(f')|^2 S_X(f,f'),\\
S_{V_{jj'},V_{jj'}}(f,f') &= |H_j(f)|^2 |G_{j'}(f')|^2 S_X(f,f'),
\end{aligned}
$$

which makes explicit the partitioning in Fig. 3.

An important question when doing a wavelet analysis, is what wavelet filter to use, and what constitutes an optimal filter will usually be application dependent. The MODWT is not as sensitive to the choice of wavelet as the DWT, but the choice will still affect the analysis. In the applications in this thesis, the wavelet variances are used for making scale-based inference, with scale interpreted as representing frequency as above. The validity of this interpretation relies to some extent on the properties of the wavelet used in the analysis and, in general, the approximation of the wavelet and scaling filters as band-pass and lowpass filters improves with increasing filter length, as can be seen in Fig. 4 where we have plotted the squared gain functions for the Haar (D(2)) and LA(20) wavelet filters for levels $1-4$. The dotted vertical lines in the figure indicate the approximate pass-band of the filters, so the area beneath the solid curve outside the dotted region indicates the amount of *spectral leakage* that would occur compared to the ideal highpass or band-pass filter. Spectral leakage is a well known phenomenon in the field of spectral estimation (see, e.g., Percival and Walden (1993) or Priestley (1981)), and can in the current context be thought of as a measure of the influence of frequencies outside the assumed

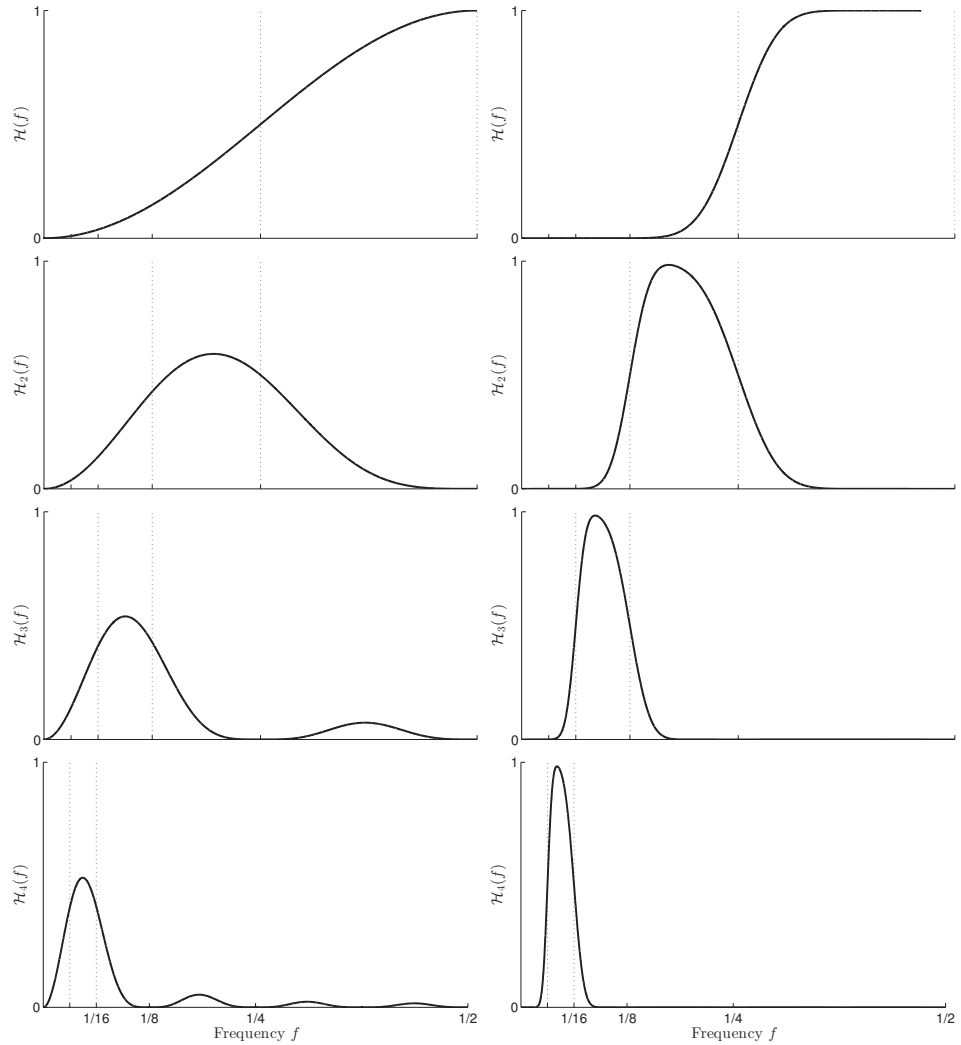


Figure 4: Squared gain function of wavelet filter for Haar (left) and LA(20) (right) wavelet filters levels $1-4$.

pass-band. In principle, it would therefore be recommended to use as long filters as possible, but practical considerations will often limit the choice of filter length. Boundary effects, e.g., will cause a band of coefficients of length L_j to be invalid for the level j coefficients, unless the data is truly circular (on a torus in the two-dimensional case). In the applications we consider in Papers II and III, this favors the shorter filters, though it is near impossible to give universal recommendations with regard to what wavelet to use in a general application.

2

RESULTS AND DISCUSSION

In this chapter a short summary of Papers I-III will be given, along with a limited discussion of the main findings. Some supplementary material connected to Paper II will be presented, and possible avenues for further research based on the work in each paper will be discussed.

2.1 PAPER I - BAYESIAN MULTISCALE ANALYSIS

This paper (Thon et al., 2012) develops tests for statistically significant gradient and curvature for noisy images modeled as two different types of GMRFs, and is essentially a two-dimensional version of the methodology in Godtliebsen and Øigård (2005) and Øigård et al. (2006). It is based on a simple image plus noise model, where it is assumed that the noise variance is known or can be estimated. The method is Bayesian in the sense that a prior image model is used, and statistical significance is determined through the posterior distribution of the gradient and curvature. The two prior image models used in the paper are defined by the first or second order normal increment processes on a local neighborhood having precision κ, and yield priors that are first-order or second-order intrinsic GMRFs, respectively. The level of smoothing is determined by the κ parameter, which can be thought of as placing different weight on the model compared to the noise, effectively determining the assumed signal to noise ratio.

The method is intended for use with large images from modern digital cameras, and much emphasis is on computational efficiency issues. Computational efficiency is achieved through restricting the scope to (intrinsically) stationary models defined on a toroidal grid, following Rue and Held (2005). This introduces correlations across opposing image borders, but allows for the use of the FFT to calculate both the posterior expectations and covariance matrix. With N being the number of pixels in the image, the computational cost of the method is $\mathcal{O}(N\log(N))$, while a straightforward implementation involving naive direct matrix inversion would be $\mathcal{O}(N^3)$. The FFT implementation is also faster than the sparse matrix methods described in Rue and Held (2005), which are typically $\mathcal{O}(N^{3/2})$. In Holmström and Pasanen (2012) a similar methodology to that of Thon et al. (2012) is developed, with the main focus being on finding significant differences between noisy images. In this formulation there is more flexibility with regard to the types of priors that can be used, and also the uncertainty in the noise variance can be built in through giving it a prior distribution reflecting the uncertainty. Also versions with correlated or heteroscedastic noise are developed. In general the more refined models rely on Gibbs sampling

for finding the posterior quantities of interest, and the computational cost vastly exceeds that of our simple model.

Our method is demonstrated on two practical problems, both related to dermascopic images of skin lesions. The first application shows the ability of the method to detect hairs, which will show up as *valleys* in the image, and the other shows how the method can be used to detect dots/globules in the skin lesion. The use of multiple scales facilitates the detection of both the finer and coarser structures.

Future work

An obvious extension of the method would be to allow for more flexibility in the choice of prior image model. While the priors used in Paper I are quite reasonable choices for many natural images, there might be applications where one wishes to incorporate specific knowledge of the spatial covariance structure of the random field that goes beyond these simple models. Recent work by Lindgren et al. (2011) has established a explicit link between the GF and GMRF modeling strategies, and shows how a GF with covariance matrix $\mathbf{\Sigma}$ can be represented by a GMRF with a local neighborhood and precision matrix $\mathbf{Q}$ such that $\mathbf{Q}^{-1}$ is close to $\mathbf{\Sigma}$ in some norm. They demonstrate how a GMRF representation can be constructed explicitly through a stochastic partial differential equation that has GFs with with Matérn covariance function as solution when driven by Gaussian white noise. This, in general, allows for the use of the full array of sparse matrix methods for doing calculations regarding the GF. Such an approach can certainly be used for constructing priors based on specific covariance models for use with the methods of Paper I. However, since the computational efficiency of our method depends on the structure of the precision/covariance matrix, rather than the sparsity, a simpler approach may be preferable. In, e.g., Dietrich and Newsam (1997); Stein (2002), a procedure called *circulant embedding* of the covariance matrix is employed for fast and exact simulations from GFs. The basic idea is to create a larger block-circulant matrix (see Paper I for definitions regarding block-circulant matrices) from the block-Toepliz covariance matrix that is defined in terms of the GFs covariance function, thus making available all the FFT-based methods for calculating, e.g., eigenvalues and inverses. This amounts to wrapping the GF on a torus, and an important issue is determining the minimal size of the embedded field for guaranteeing that the resulting block-circulant covariance matrix is non-negatively definite. To use this methodology for defining prior image models as in Paper I, would require the base of the block-circulant matrix to be the same size as the image. This would likely put limits on the range of dependencies in the model that can be approximated by such a procedure.

Another possible extension of the method would be to sacrifice some computational efficiency by avoiding the use of a toroidal grid, and rather use a more standard GMRF formulation, relying on the sparse matrix methods in Rue and Held (2005). The obvious advantage from this is that

the whole image can be considered, also the borders. To incorporate the use of priors defined through the covariance function of a GF, as outlined above, would in this case demand the use of a method such as in Lindgren et al. (2011) (for GFs with with Matérn covariance function), or some other method for representing a GF as a GMRF, see, e.g., Rue and Tjelmeland (2002).

2.2 PAPER II - WAVELET-BASED TEST FOR ISOTROPY

A common assumption when using random field models, is that the field in question is isotropic, meaning that the second order statistics of the field are independent of direction. In Paper II, a multiscale statistical test for isotropy of random fields is developed, based on the estimated wavelet variances over multiple scales. For an isotropic random field, the covariance function or semivariogram will depend only on the Euclidean distance between the locations, meaning that it will be completely symmetric around the origin when plotted as a function of lag (iso-contour lines will be perfectly circular). For a stationary field, the spectral density function will exhibit the same symmetries so, referring to Fig. 3, this means that $\nu^2_{W,j,j'} = \nu^2_{W,j',j}$ and $\nu^2_{U,j,j} = \nu^2_{V,j,j}$, where the last relation generalizes to $\nu^2_{U,j,j'} = \nu^2_{V,j',j}$. As a consequence, for an isotropic field,

$$\log\left(\frac{\nu^2_{U,j,j'}}{\nu^2_{V,j',j}}\right) = 0 \quad \text{and} \quad \log\left(\frac{\nu^2_{W,j,j'}}{\nu^2_{W,j',j}}\right) = 0, \tag{9}$$

and replacing the variances in Equation (9) with their estimates from Equations (3) to (5) gives the test statistic for the single scale test developed in Paper I. The test is further developed to include the ratios of choice, giving a multiscale test of isotropy. Large sample results are used to find the null distribution of the test statistic as a chi square with J degrees of freedom, where J denotes the number of ratios included. A simulation study shows that the asymptotic null distribution is a reasonable approximation for the empirical distribution even for quite small grids, and the test has considerable power in rejecting anisotropic fields. The test is compared against three existing methods (Lu and Zimmerman, 2005; Guan et al., 2004; Richard and Bierme, 2010). The wavelet-based test is shown to outperform the others since its rejection rate under isotropy is consistently closest to the nominal level, while having rejection rates comparable to or exceeding the existing methods. In addition, the wavelet-based test is more general in that it handles the case of intrinsically stationary fields. The method is demonstrated on paper density images of handsheets.

Supplementary material

This paragraph can only be understood with reference to the power study in Paper II, and the relevant definitions and description of the study are to be found there. In the paper, due to

space considerations, rejection rates are given only for the 5% nominal level and only for the D(4) wavelet, although the simulation study was run also with the Haar and LA(8) wavelet. To complete the story, Fig. 5 shows three receiver operating characteristic (ROC) curves. The power is calculated as the average over thirty different parametrizations of anisotropic exponential fields, each of which are based on 1000 samples. The average rejection rate (power) of the test performed on the exponential fields is plotted as a function of the size of the test, for 20×20 (left), 40×40 (middle), and 128×128 (right) grids. As apparent from the ROC curves, the power is consistently lower for the longer filters, and this is likely connected to there being fewer coefficients available for estimating the necessary variances/covariances for constructing the test. The effect is quite pronounced for the smallest 20×20 and, to a lesser extent, 40×40 fields. For the 128×128 fields the shorter wavelets still have a slightly higher rejection rate but, for practical purposes, the performance of the different wavelets is identical. This is consistent with the effect having to do with the number of valid coefficients available for calculating the necessary estimates, since as the grid size increases, the relative proportion of invalid coefficient decreases for all wavelets, but more dramatically for the longer ones.

Another choice that was made in the power study, was to consider only the test constructed from a single ratio. This was done since we wished to make a fair comparison with the other methods in the comparative study, using the same test statistic for all image sizes. In Fig. 6 the average power of the D(4) test based on unit-level, levels (1,2), and (1,2,3) diagonal sw/ws ratios for 128×128 exponential (left), spherical (middle) and fractional Brownian fields (FBFs) (right) are plotted. For the exponential and spherical fields, the average rejection rate is actually highest for the single scale test, and decreases with the addition of higher level ratios. In view of Fig. 3 this is somewhat counter intuitive, since inclusion of more levels should correspond to including more of the power spectrum in the test. The likely explanation for this effect is that the elements of $\mathbf{\Sigma}$ needed for constructing the test need to be estimated. The small and possibly correlated errors in the the estimates are sufficient to reduce the power of the test, and including more scales warrants the estimation of more quantities. The samples used in the study are also such that the field is equally anisotropic on all scales, so the benefits of including more scales are outweighed by the additional estimation needed. In real world applications, this need not be the case, and a field can conceivably be anisotropic on some scales and on others not. In such a case the multiple ratio test statistic will be relevant for evaluating the anisotropy of the field as a whole. For the FBFs, the trend is different, and the average power increases with the inclusion of more scales. This is likely an effect of the long range dependencies in the FBFs, and causes the discriminatory power to be larger for the higher scales.

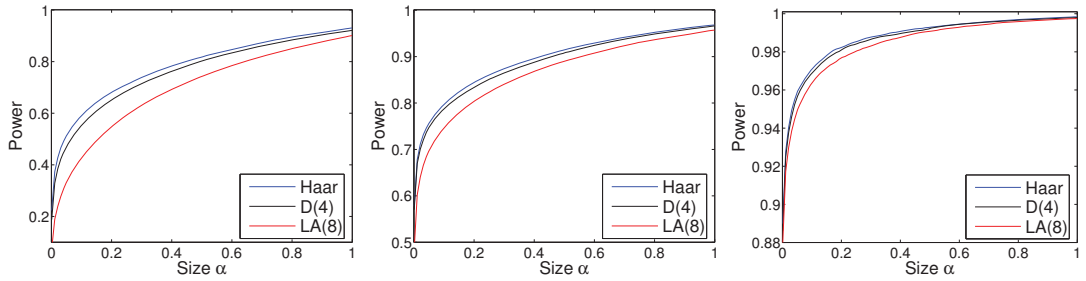


Figure 5: The average power of the Haar, D(4) and LA(8) test based on unit-level diagonal sw/ws ratio for 20×20 (left), 40×40 (middle), and 128×128 (right) grids. Note that the scale on the y-axis is different in the three ROC curves.

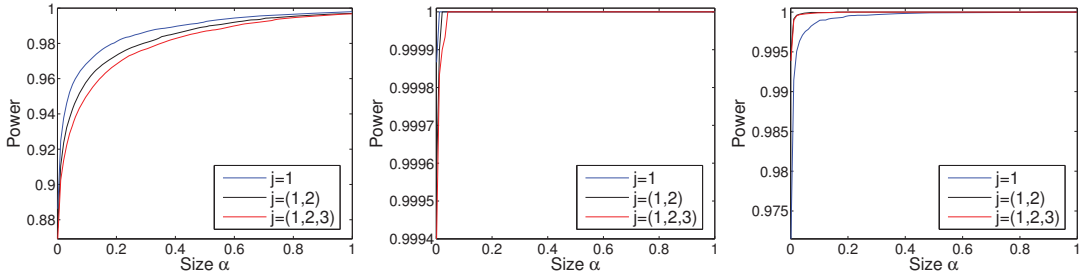


Figure 6: The average power of the D(4) test based on unit-level, levels (1,2), and (1,2,3) diagonal sw/ws ratios for 128×128 exponential (left), spherical (middle) and fractional Brownian fields (right). Note that the scale on the y-axis is different in the three ROC curves.

Future work

The estimators of the covariances needed for constructing the test rely on use of the periodogram, whose suboptimal properties are well known. It would be interesting to explore the use of other spectral estimates, although this would lead to an increase in computational burden. The two-dimensional wavelet transform has already been used successfully within the field of texture analysis and segmentation, see, e.g., Unser (1995); Do and Vetterli (2002). An interesting avenue of research would be to explore the usefulness of the test statistic in Paper II as a texture measure, i.e., use the degree of anisotropy as an image feature. The wavelet transform should then be applied over the entire image, and the coefficients shifted to compensate for the non-zero phase of wavelet filters. The image can then be divided into boxes of a chosen size, and the variances can be calculated locally. With appropriate modification of the variance estimators and the estimators of the covariance between the different variance estimators, the test applies as described in Paper II.

2.3 PAPER III - WAVELET-BASED TEST FOR EQUALITY OF RANDOM FIELDS

The test developed in Paper III is for determining if two independent samples are realizations from the same random field. It is based on the vector difference of the estimated wavelet variances of two fields over a range of scales. Samples from the same field should, up to sampling variability, have the same wavelet variances, and the test is developed using the large sample distribution of the wavelet variances for estimating the distribution of the multiple scale test statistic. The scales included in the test should reflect the frequency bands that are most interesting for the application at hand. A simulation study shows that the test statistic follows the hypothesized null distribution closely, even for moderately sized images, and that it has considerable power in differentiating between samples from different random fields.

The dominant theories within the field of cosmology predict that the cosmic microwave background (CMB) should be a realization from an isotropic Gaussian field (Dodelson, 2003). With the release of the latest CMB measurements from the Planck satellite, researchers have available CMB data of unprecedented quality and resolution (Planck Collaboration, 2013a). The high resolution of the measurements allow for making small projections to regular grids, and the methodology developed in Paper III can be applied to projections in antipodal regions, i.e., regions separated by the maximum great circle distance, since the assumption of isotropy on the sphere means that the projections should be realizations from the same Gaussian field. A simulation study involving sampling CMB realizations with the required angular spectrum (Planck Collaboration, 2013b) on a sphere demonstrates the validity of the approach under the null hypothesis of isotropy. Finally, we apply the method to the actual Planck CMB temperature map. While we do find indications of anisotropies, these have thus far not been verified in terms of robustness against systematic measurement errors, or other noise sources.

Future work

In this section some methods that are closely related to the test for equality of random fields from Paper III are outlined. Much of the definitions are from Paper III, so this section can only be fully understood with reference to the paper. The methodologies presented here, are to be considered merely an outline for future research and are as such neither complete nor backed up by extensive simulation studies or practical examples.

Test against theoretical random field

We wish to develop a test of whether a random field sample is from a particular random field with a particular parametrization. First we note that, for a stationary field $\{X_{u,v}\}$ with covariance function s_{X,κ_1,κ_2}, the wavelet variances may be written as

$$\nu^2_{W,j,j'} = \sum_{l,l'=0}^{L_j-1} \sum_{k,k'=0}^{L_{j'}-1} h_{j,l} h_{j,l'} h_{j',k} h_{j',k'} s_{X,l-l',k-k'}, \tag{10}$$

$$\nu^2_{U,j,j'} = \sum_{l,l'=0}^{L_j-1} \sum_{k,k'=0}^{L_{j'}-1} g_{j,l} h_{j,l'} g_{j',k} h_{j',k'} s_{X,l-l',k-k'}, \tag{11}$$

$$\nu^2_{V,j,j'} = \sum_{l,l'=0}^{L_j-1} \sum_{k,k'=0}^{L_{j'}-1} h_{j,l} g_{j,l'} h_{j',k} g_{j',k'} s_{X,l-l',k-k'}. \tag{12}$$

When $\{X_{u,v}\}$ is intrinsically stationary, we can express Equations (10) – (12) in terms of its semivariogram,

$$\nu^2_{W,j,j'} = -\sum_{l,l'=0}^{L_j-1} \sum_{k,k'=0}^{L_{j'}-1} h_{j,l} h_{j,l'} h_{j',k} h_{j',k'} \gamma_{X,l-l',k-k'}. \tag{13}$$

with similar forms for $\nu^2_{U,j,j'}$ and $\nu^2_{V,j,j'}$. Now, assume that the field X comes from a specific covariance model parametrized by $\boldsymbol{\theta}$. Letting $\hat{\boldsymbol{\beta}} = [\hat{\nu}^2_{W,1,1}, \ldots, \hat{\nu}^2_{W,J,J}]^T$ denote the vector of diagonal ww variance estimators, J denoting the highest included level, we know from Mondal and Percival (2012) that

$$\hat{\boldsymbol{\beta}} \xrightarrow{D} \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \operatorname{var}\{\hat{\nu}^2_{W,1,1}\} & \operatorname{cov}\{\hat{\nu}^2_{W,1,1}, \hat{\nu}^2_{W,2,2}\} & \cdots & \operatorname{cov}\{\hat{\nu}^2_{W,1,1}, \hat{\nu}^2_{W,J,J}\} \\ \operatorname{cov}\{\hat{\nu}^2_{W,2,2}, \hat{\nu}^2_{W,1,1}\} & \operatorname{var}\{\hat{\nu}^2_{W,2,2}\} & \cdots & \operatorname{cov}\{\hat{\nu}^2_{W,2,2}, \hat{\nu}^2_{W,J,J}\} \\ \vdots & & \ddots & \vdots \\ \operatorname{cov}\{\hat{\nu}^2_{W,J,J}, \hat{\nu}^2_{W,1,1}\} & \cdots & \cdots & \operatorname{var}\{\hat{\nu}^2_{W,J,J}\} \end{bmatrix}.$$

with $\boldsymbol{\beta} = [\nu^2_{W,1,1}, \ldots, \nu^2_{W,J,J}]^T$ being the vector of true wavelet variances. We wish to test if the considered sample is from a particular random field, so we by assumption know the covariance function s_{X,κ_1,κ_2} or *semivariogram* $\gamma_{X,\kappa_1,\kappa_2}$ for intrinsically stationary fields. This means that we can easily calculate the theoretical wavelet variances $\boldsymbol{\beta}_{\boldsymbol{\theta}}$ for the field in question through Equation (10) or (13). The covariances between the estimators can be calculated efficiently by

the fast Fourier transform (FFT), so assuming that the sample is from the hypothesized random field,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

A test for determining whether the sample is from the random field in question can therefore be phrased as

$$H_0 : \boldsymbol{\beta} - \boldsymbol{\beta}_{\boldsymbol{\theta}} = \mathbf{0};\ H_1 : \boldsymbol{\beta} - \boldsymbol{\beta}_{\boldsymbol{\theta}} \neq \mathbf{0}.$$

$\chi = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}})$ is under the null hypothesis distributed as χ^2_J, the chi-square distribution with J degrees of freedom. Using χ as the test statistic, we find that for significance level α, H_0 will be rejected when $\chi > q_{\chi^2_J}(\alpha)$, the upper (100α)th percentile of the χ^2_J distribution.

The test as described above, is for testing what is described in DasGupta (2008) as a simple hypothesis H_0, where the model is fully specified both in terms of covariance type and parametrization. A more realistic scenario is that you wish to test for whether a sample is from a class of covariance functions (e.g., exponential, spherical, fractional Brownian surface), and the parameters need to be estimated. This constitutes a composite hypothesis, and the methodology described above does not immediately lend itself to the task. $\boldsymbol{\beta}_{\boldsymbol{\theta}}$ would be replaced by the random variable $\boldsymbol{\beta}_{\hat{\boldsymbol{\theta}}}$, where $\hat{\boldsymbol{\theta}}$ is some estimator for $\boldsymbol{\theta}$. Even for a normally distributed estimator $\hat{\boldsymbol{\theta}}$ with known variance, finding the distribution of $\boldsymbol{\beta}_{\hat{\boldsymbol{\theta}}}$ through Equations (10) or (13) is non-trivial.

Maximum likelihood estimator of parameters of random field

The same basic idea as above can be used for estimating the unknown parameters defining the field sample in question. The likelihood is

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}) = P(\hat{\boldsymbol{\beta}} | \boldsymbol{\theta}) = P(\hat{\boldsymbol{\beta}} | \boldsymbol{\beta}_{\boldsymbol{\theta}}) = \mathcal{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right) \tag{14}$$

Finding the value of $\boldsymbol{\theta}$ that maximizes Equation (14) is equivalent to finding the $\boldsymbol{\theta}$ that minimizes

$$-2\log\left(L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}})\right) = J\log(2\pi) + \log(|\boldsymbol{\Sigma}|) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}}),$$

so

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\log(|\boldsymbol{\Sigma}|) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\boldsymbol{\theta}})\right),$$

gives the maximum likelihood estimate. The complicated relationship between $\boldsymbol{\Sigma}$ and the parametrization of the underlying field prohibits an analytical solution of the above. Since the sup-

port of the parameters of most fields is usually limited, we rather recommend a numerical procedure, where the minimization is done first on a coarse grid, and then on a finer grid around the minimum from the previous step.

Steps towards a test of correlated samples

The test developed in Paper III assumes that the samples being tested for equality are independent. An interesting application of the methodology from Paper III might be to quantify the textural difference between regions within an single image, taking into account the possible correlations between the regions. In the following, some steps along the way to developing such a test are outlined.

With X and Y denoting the respective fields under consideration, and with $\hat{\boldsymbol{\beta}}_X$ and $\hat{\boldsymbol{\beta}}_Y$ denoting the vector of wavelet variances from field X and Y, respectively, we construct the vector

$$\hat{\boldsymbol{\beta}} = [\hat{\nu}^2_{W_X,1,1}, \hat{\nu}^2_{W_Y,1,1}, \ldots, \hat{\nu}^2_{W_X,J,J}, \hat{\nu}^2_{W_Y,J,J}]^T$$

by placing pairwise the corresponding elements from $\hat{\boldsymbol{\beta}}_X$ and $\hat{\boldsymbol{\beta}}_Y$. $\hat{\nu}^2_{W_X,j,j}$ and $\hat{\nu}^2_{W_Y,j,j}$ denote the level j diagonal ww-variance estimators of the first and second field, respectively. The asymptotic normality of the estimators gives,

$$\hat{\boldsymbol{\beta}} \xrightarrow{D} \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_1),$$

where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \operatorname{var}\{\hat{\nu}^2_{W_X,1,1}\} & \operatorname{cov}\{\hat{\nu}^2_{W_X,1,1}, \hat{\nu}^2_{W_Y,1,1}\} & \operatorname{cov}\{\hat{\nu}^2_{W_X,1,1}, \hat{\nu}^2_{W_X,2,2}\} & \cdots & \operatorname{cov}\{\hat{\nu}^2_{W_X,1,1}, \hat{\nu}^2_{W_Y,J,J}\} \\ \operatorname{cov}\{\hat{\nu}^2_{W_Y,1,1}, \hat{\nu}^2_{W_X,1,1}\} & \operatorname{var}\{\hat{\nu}^2_{W_Y,1,1}\} & \operatorname{cov}\{\hat{\nu}^2_{W_Y,1,1}, \hat{\nu}^2_{W_X,2,2}\} & \cdots & \operatorname{cov}\{\hat{\nu}^2_{W_Y,1,1}, \hat{\nu}^2_{W_Y,J,J}\} \\ \operatorname{cov}\{\hat{\nu}^2_{W_X,2,2}, \hat{\nu}^2_{W_X,1,1}\} & \operatorname{cov}\{\hat{\nu}^2_{W_X,2,2}, \hat{\nu}^2_{W_Y,1,1}\} & \operatorname{var}\{\hat{\nu}^2_{W_X,2,2}\} & \cdots & \operatorname{cov}\{\hat{\nu}^2_{W_X,2,2}, \hat{\nu}^2_{W_Y,J,J}\} \\ \vdots & & & \ddots & \vdots \\ \operatorname{cov}\{\hat{\nu}^2_{W_Y,J,J}, \hat{\nu}^2_{W_X,1,1}\} & \cdots & \cdots & \cdots & \operatorname{var}\{\hat{\nu}^2_{W_Y,J,J}\} \end{bmatrix},$$

with $\boldsymbol{\beta} = [\nu^2_{W_X,1,1}, \ldots, \nu^2_{W_Y,J,J}]^T$ being the vector of theoretical wavelet variances. Introducing the differencing matrix

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & \cdots & & & & 1 & -1 \end{bmatrix},$$

results in, using standard results from multivariate theory (see, e.g., Johnson and Wichern (2007)),

$$\mathbf{B}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_X - \hat{\boldsymbol{\beta}}_Y \xrightarrow{D} \mathcal{N}(\boldsymbol{\beta}_X - \boldsymbol{\beta}_Y, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_1\mathbf{B}^T$, with elements

$$\Sigma_{s,t} = \text{cov}\{\hat{\beta}_{2s}, \hat{\beta}_{2t}\} + \text{cov}\{\hat{\beta}_{2s-1}, \hat{\beta}_{2t-1}\} - \text{cov}\{\hat{\beta}_{2s}, \hat{\beta}_{2t-1}\} - \text{cov}\{\hat{\beta}_{2s-1}, \hat{\beta}_{2t}\}.$$

A test can therefore be formulated as

$$H_0 : \boldsymbol{\beta}_X - \boldsymbol{\beta}_Y = \mathbf{0};\ H_1 : \boldsymbol{\beta}_X - \boldsymbol{\beta}_Y \neq \mathbf{0}.$$

Under the null hypothesis $\chi = (\hat{\boldsymbol{\beta}}_X - \hat{\boldsymbol{\beta}}_Y)^T\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_X - \hat{\boldsymbol{\beta}}_Y)$ is distributed as χ^2_J. Using χ as the test statistic, we find that for significance level α, H_0 will be rejected when $\chi > q_{\chi^2_j}(\alpha)$. In a practical application of the test, the elements in $\boldsymbol{\Sigma}_1$ must be replaced with their estimates. For the elements pertaining to the cross-level covariances of the ww variance estimates from the same field, the estimators from Paper III can be used, but the between field estimators need to be developed. In the following this development will be somewhat superficial and the interested reader is referred to Paper II for more detail. First we note that, assuming that the random fields are both of size $N \times M$, and with $N_j = N - L_j + 1$ (M_j, N_k and M_k are defined equivalently),

$$\begin{aligned}\text{cov}\left\{\hat{v}^2_{W_X,j,j}, \hat{v}^2_{W_Y,k,k}\right\} &= \text{cov}\left\{\frac{1}{N_jM_j}\sum_{a=L_j-1}^{N-1}\sum_{b=L_j-1}^{M-1} W_{X\,j,j,a,b}^2, \frac{1}{N_kM_k}\sum_{c=L_k-1}^{N-1}\sum_{d=L_k-1}^{M-1} W_{Y\,k,k,c,d}^2\right\} \\ &= \frac{1}{N_jM_jN_kM_k}\sum_{a=L_j-1}^{N-1}\sum_{b=L_j-1}^{M-1}\sum_{c=L_k-1}^{N-1}\sum_{d=L_k-1}^{M-1} \text{cov}\left\{W_{X\,j,j,a,b}^2, W_{Y\,k,k,c,d}^2\right\}.\end{aligned}$$

After applying Isserlis Theorem and some asymptotically justified simplifications, we find that

$$\text{cov}\left\{\hat{v}^2_{W_X,j,j}, \hat{v}^2_{W_Y,k,k}\right\} \approx \frac{2}{N^+M^+}\sum_{t=-\infty}^{\infty}\sum_{t'=-\infty}^{\infty} s^2_{W_{X,jj}W_{Y,kk}}(t,t') = \frac{2A_{j,j,k,k}}{N^+M^+},$$

where $s_{W_{X,jj}W_{Y,kk}}(t,t') = \text{cov}\left\{W_{X\,j,j,u,v}, W_{Y\,k,k,u+t,v+t'}\right\}$. In the above $N^+ = \max(N_j, N_k)$ and $M^+ = \max(M_j, M_k)$. By Parseval's theorem,

$$A_{j,j,k,k} = \sum_{t=-\infty}^{\infty}\sum_{t'=-\infty}^{\infty} s^2_{W_{X,jj}W_{Y,kk}}(t,t') = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} |S_{W_{X,jj}W_{Y,kk}}(f,f')|^2\, df df',$$

$S_{W_{X,jj}W_{Y,kk}}(f,f')$ being the cross-spectral density between $W_{X,jj}$ and $W_{Y,kk}$. We further note that

$$S_{W_{X,jj}W_{Y,kk}}(f,f') = H_j(f)^*H_j(f')^*H_k(f)H_k(f')S_{X,Y}(f,f'),$$

and

$$S_{W_{X,jj}W_{X,jj}}(f,f')S_{W_{Y,kk}W_{Y,kk}}(f,f') = |H_j(f)|^2|H_j(f')|^2|H_k(f)|^2|H_k(f')|^2 S_X(f,f')S_Y(f,f'),$$

so,

$$\frac{S_{W_{X,jj}W_{X,jj}}(f,f')S_{W_{Y,kk}W_{Y,kk}}(f,f')}{|S_{W_{X,jj}W_{Y,kk}}(f,f')|^2} = \frac{S_X(f,f')S_Y(f,f')}{|S_{X,Y}(f,f')|^2} = \frac{1}{\xi_{X,Y}(f,f')},$$

where $\xi_{X,Y}(f,f')$ denotes the magnitude squared coherence (MSC) between X and Y. Thus, the expectation of the squared modulus of the cross-periodogram can be written

$$\begin{aligned} E\left\{|\hat{S}^{(p)}_{W_{X,jj}W_{Y,kk}}(f,f')|^2\right\} &\approx S_{W_{X,jj}W_{X,jj}}(f,f')S_{W_{Y,kk}W_{Y,kk}}(f,f') + |S_{W_{X,jj}W_{Y,kk}}(f,f')|^2 \\ &= \left(\frac{S_{W_{X,jj}W_{X,jj}}(f,f')S_{W_{Y,kk}W_{Y,kk}}(f,f')}{|S_{W_{X,jj}W_{Y,kk}}(f,f')|^2} + 1\right)|S_{W_{X,jj}W_{Y,kk}}(f,f')|^2 \\ &= \left(\frac{1}{\xi_{X,Y}(f,f')} + 1\right)|S_{W_{X,jj}W_{Y,kk}}(f,f')|^2. \end{aligned}$$

Assuming that an unbiased estimator $\hat{\xi}_{X,Y}(f,f')$ can be found for the MSC (Jakobsson et al., 2007), this means that an approximately unbiased estimator for $\mathrm{cov}\left\{\hat{v}^2_{W_X,j,j}, \hat{v}^2_{W_Y,k,k}\right\}$ is given by

$$\hat{\sigma}_{W_{X\,jj}W_{Y\,kk}} = \frac{2\hat{A}_{j,j,k,k}}{N^+M^+},$$

with

$$\hat{A}_{j,j,k,k} = \sum_{n=-(N^--1)}^{N^--1}\sum_{m=-(M^--1)}^{M^--1} \frac{\hat{\xi}_{X,Y}(f_n,f'_m)}{\hat{\xi}_{X,Y}(f_n,f'_m)+1}|\hat{S}^{(p)}_{W_{X,jj}W_{Y,kk}}(f_n,f'_m)|^2$$

where $N^- = \min(N_j,N_k)$ and $M^- = \min(M_j,M_k)$, and the summation is over the Fourier frequencies. This assumes that $\hat{\xi}_{X,Y}(f,f')$ is independent of $\hat{S}^{(p)}_{W_{X,jj}W_{Y,kk}}(f,f')$. We note that when X and Y are independent, $\hat{\xi}_{X,Y}(f_n,f'_m)$ should be close to zero, making $\hat{\sigma}_{W_{X\,jj}W_{Y\,kk}}$ close to zero, in agreement with the test for independent samples. Also, if $X = Y$ the MSC equals unity, and we get the estimator from Paper II for the cross-level covariances.

Implementation and testing of the above estimators will be the subject of future research.

BIBLIOGRAPHY

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B 36 (2), 192–236.

Beylkin, G., 1992. On the representation of operators in bases of compactly supported wavelets. SIAM Journal on Numerical Analysis 29 (6), 1716–1740.

Bruce, A., Gao, H.-Y., 1996. Applied wavelet analysis with S-plus. Springer-Verlag , Inc., New York.

Chaudhuri, P., Marron, J., 1999. Sizer for exploration of structures in curves. Journal of the American Statistical Association 94 (447), 807–823.

Chaudhuri, P., Marron, J. S., 2000. Scale space view of curve estimation. Annals of Statistics 28 (2), 408–428.

Chilès, J. P., Delfiner, P., 2012. Geostatistics: Modeling spatial uncertainty, 2nd Edition. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Coifman, R. R., Donoho, D. L., 1995. Translation-invariant de-noising. Springer.

Cressie, N., 1993. Statistics for spatial data, Revised Edition. John Wiley & Sons, Inc., New York, NY, USA.

DasGupta, A., 2008. Asymptotic theory of statistics and probability. Springer.

Daubechies, I., 1992. Ten Lectures on Wavelets. Cbms-Nsf Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA.

Dietrich, C., Newsam, G. N., 1997. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. SIAM Journal on Scientific Computing 18 (4), 1088–1107.

Do, M. N., Vetterli, M., 2002. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Transactions on Image Processing 11 (2), 146–158.

Dodelson, S., 2003. Modern Cosmology. Academic Press.

Ferdman, Y., Sagiv, C., Sochen, N., 2007. Full affine wavelets are scale-space with a twist. In: Scale Space and Variational Methods in Computer Vision. Springer, pp. 1–12.

Godtliebsen, F., Marron, J. S., Chaudhuri, P., 2002. Significance in scale space for bivariate density estimation. Journal of Computational and Graphical Statistics 11 (1), 1–21.

Godtliebsen, F., Marron, J. S., Chaudhuri, P., 2004. Statistical significance of features in digital images. Image Vision Computing 22 (13), 1093–1104.

Godtliebsen, F., Øigård, T. A., February 2005. A visual display device for significant features in complicated signals. Computational Statistics & Data Analysis 48 (2), 317–343.

Guan, Y., Sherman, M., Calvin, J. A., 2004. A nonparametric test for spatial isotropy using subsampling. Journal of the American Statistical Association 99 (467), 810–821.

Hannig, J., Lee, T. C., 2006. Robust SiZer for exploration of regression structures and outlier detection. Journal of Computational and Graphical Statistics 15 (1), 101–117.

Hannig, J., Marron, J., 2006. Advanced distribution theory for SiZer. Journal of the American Statistical Association 101 (474), 484–499.

Holmström, L., Pasanen, L., 2012. Bayesian scale space analysis of differences in images. Technometrics 54 (1), 16–29.

Jakobsson, A., Alty, S. R., Benesty, J., 2007. Estimating and time-updating the 2-d coherence spectrum. Signal Processing, IEEE Transactions on 55 (5), 2350–2354.

Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis, 6th International edition. Prentice Hall.

Künsch, H. R., 1987. Intrinsic autoregressions and related models on the two-dimensional lattice. Biometrika 74 (3), 517–524.

Lindeberg, T., 1994. Scale-Space Theory in Computer Vision. Kluwer/Springer.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (4), 423–498.

Lu, N., Zimmerman, D. L., 2005. Testing for directional symmetry in spatial dependence using the periodogram. Journal of Statistical Planning and Inference 129 (1-2), 369–385.

Mondal, D., Percival, D. B., 2012. Wavelet variance analysis for random fields on a regular lattice. IEEE Transactions on Image Processing 21 (2), 537–549.

Nason, G. P., Silverman, B. W., 1995. The stationary wavelet transform and some statistical applications. In: Wavelets and statistics. Springer, pp. 281–299.

Øigård, T. A., Rue, H., Godtliebsen, F., December 2006. Bayesian multiscale analysis for time series data. Computational Statistics & Data Analysis 51 (3), 1719–1730.

Percival, D. B., Walden, A. T., 1993. Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge University Press, Cambridge, UK.

Percival, D. B., Walden, A. T., 2000. Wavelet Methods for Time Series Analysis. Cambridge University Press, New York, NY, USA.

Pesquet, J.-C., Krim, H., Carfantan, H., 1996. Time-invariant orthonormal wavelet representations. IEEE Transactions on Signal Processing 44 (8), 1964–1970.

Planck Collaboration, 2013a. Planck 2013 results. I. Overview of products and scientific results. arXiv preprint arXiv:1303.5062.

Planck Collaboration, 2013b. Planck 2013 results. XXIII. Isotropy and Statistics of the CMB. arXiv preprint arXiv:1303.5083.

Priestley, M. B., 1981. Spectral Analysis and Time Series. Academic Press.

Richard, F., Bierme, H., 2010. Statistical tests of anisotropy for fractional Brownian textures. Application to full-field digital mammography. Journal of Mathematical Imaging and Vision 36 (3), 227–240.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, London, UK.

Rue, H., Tjelmeland, H., 2002. Fitting Gaussian Markov random fields to Gaussian fields. Scandinavian Journal of Statistics 29 (1), 31–49.

Shensa, M. J., 1992. The discrete wavelet transform: Wedding the à trous and Mallat algorithms. IEEE Transactions on Signal Processing 40 (10), 2464–2482.

Stein, M. L., 2002. Fast and exact simulation of fractional Brownian surfaces. Journal of Computational and Graphical Statistics 11 (3), 587–599.

Thon, K., Rue, H., Skrøvseth, S. O., Godtliebsen, F., 2012. Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. Computational Statistics & Data Analysis 56 (1), 49–61.

Unser, M., 1995. Texture classification and segmentation using wavelet frames. IEEE Transactions on Image Processing 4 (11), 1549–1560.

Part II

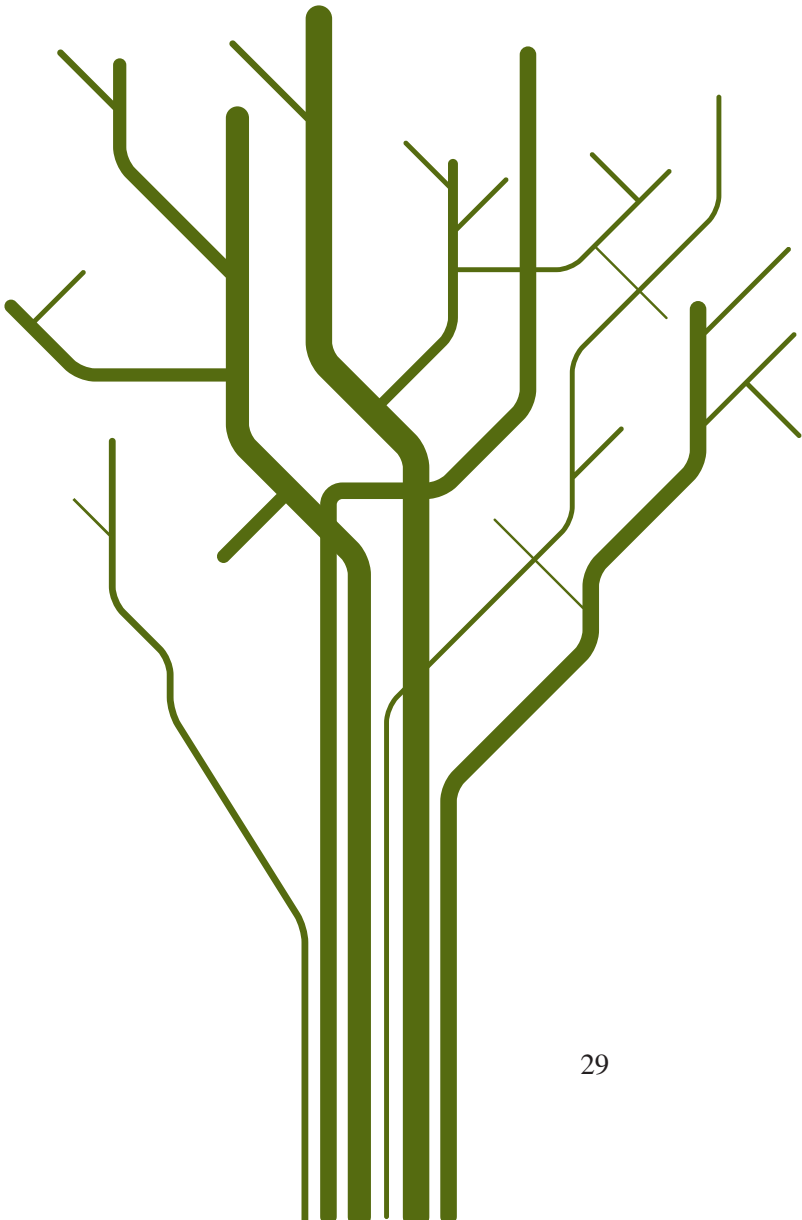
PAPERS

Paper I

Bayesian multiscale analysis of images modeled as Gaussian Markov random fields

Thon, K., Rue, H., Skrøvseth, S. O. and Godtliebsen, F.

Computational Statistics and Data Analysis, 2012, Vol. 56, pp. 49-61

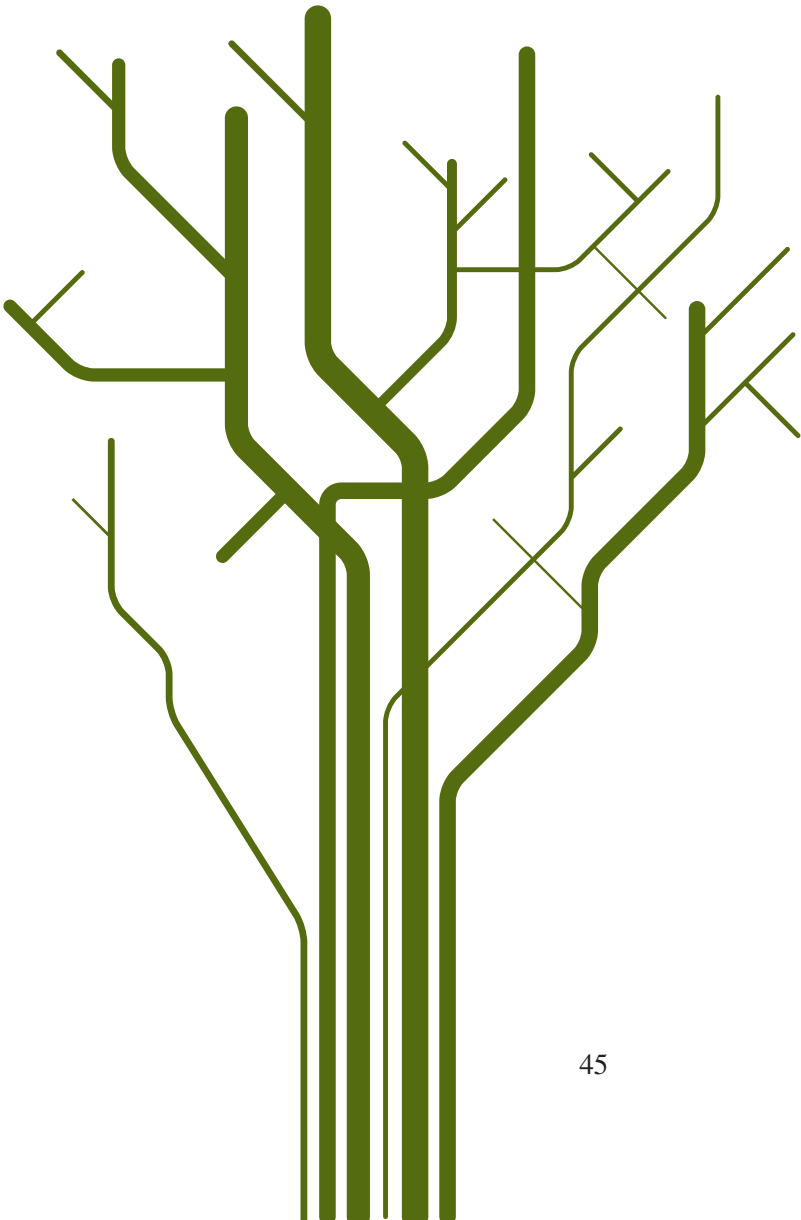


Paper II

A multiscale wavelet-based test for isotropy of random fields on a regular lattice

Thon, K., Geilhufe, M. and Percival, D. B.

being revised for *IEEE Transactions on Image Processing*



Paper III

A scale-based test for equality of Gaussian random fields with application to the cosmic microwave background radiation

Thon, K., Percival, D. B., Geilhufe, M. and Skrøvseth, S. O.

unpublished manuscript