LICENTIATE THESIS

Markku Mielityinen

A Study of Spatial Correlations with an Application in Paper Science

UNIVERSITY OF JYVÄSKYLÄ
JYVÄSKYLÄ 2006

ABSTRACT

Mielityinen, Markku
A Study of Spatial Correlations with an Application in Paper Science
Jyväskylä: University of Jyväskylä, 2006, 119 p.
Finnish summary

In this study our concern is to contemplate spatial correlations and their applications in a real world setting. The objective is to evaluate different approaches that are commonly used in this problem area and, as it becomes necessary, to develop improved methodology that is able to meet the needs of applications. In this respect we propose a new spatial correlation measure that qualifies and quantifies the amount of dependency between two images that are generated by a spatial stochastic process. This correlation measure provides a robust and geometrically interpretable alternative to the standard tools that are currently used in spatial statistics. For evaluation purposes, the proposed methodology is implemented on a computer platform and an example data analysis is performed on measurement images that originate from paper industry.

Keywords: spatial correlations, stochastic processes, exploratory data analysis, spatial statistics, stochastic geometry, dependencies between paper properties

Author Markku Mielityinen
Department of Mathematics and Statistics
University of Jyväskylä
Finland

Supervisor DrTech Pasi Koikkalainen
Department of Mathematics and Statistics
University of Jyväskylä
Finland

Reviewers Professor Olli Silven
Information Processing Laboratory, Department of Electrical Engineering
University of Oulu
Finland

Professor Risto Ritala
The Institute of Measurement and Information Technology, Department of Automation and Department of Electrical Engineering
Tampere University of Technology
Finland

ACKNOWLEDGMENTS

I am grateful to DrTech Pasi Koikkalainen for all the time and support that he has placed in guiding me through my postgraduate studies. His broad understanding of many scientific disciplines and his firm expertise from real world applications has helped me to adopt somewhat similar attitude towards academic research. Beyond the mere academics he has taken the time to advice me on matters of writing. I also want to thank all my current and former colleagues from the Laboratory of Data Analysis, University of Jyväskylä. Former postgraduate students PhD Anssi Lensu and PhD Jouni Raitamäki have helped me in numerous occasions especially in my early years of academic life. MSc Michael Haranen and MSc Ismo Horppu started their postgraduate studies approximately at the same time that I did and together we have made a team in which we have been able to assist each other in our daily chores.

I also wish to recognize the broader support structure from professors Antti Penttinen and Hannu Oja. Their expertise in spatial and robust statistics has played an important part in my research and they have also had a considerable role in forming my way of thinking. The seminars that were organized by the Center for Mathematical and Computational Modeling (CMCM) and by the Laboratory of Data Analysis have helped me to broaden my academic horizons beyond my own work.

The research that is presented in this thesis has received funding from Academy of Finland under grant MaDaMe/72497 and from National Technology Agency of Finland (TEKES) through a KCL Project 1131 Toner-Paper Interactions, subproject 5. From this project I wish to thank DrTech Erkki Hellén and MSc Pasi Puukko from our sometimes daily collaboration. Some traveling expenses have been compensated by the Graduate School in Computational Methods of Information Technology (ComMIT). Department of Mathematics and Statistics, University of Jyväskylä has provided additional funding for the finalization of the thesis.

Finally I want to acknowledge the support of my family and friends that have patiently supported me through the whole process. My parents Eeva and Pekka have always encouraged me in my studies and they have provided the surroundings that have enabled me to continue my studies this far. My sister Tuija has filled her part by pursuing an academic career of her own in engineering. As for my friends, which are too numerous to mention here by name, some of them have been able to assist me in my research whereas all of them have helped in keeping my spirits high.

Jyväskylä, April 10, 2006

Markku Mielityinen

CONTENTS

LIST OF SYMBOLS

Images and features		
Symbol	*Description*	*Page*
(x, y)	a pixel coordinate	14
$Z(x, y)$	a random pixel value at (x, y)	15
$z(x, y)$	a realization from $Z(x, y)$	14, 15
w_{image}	image width in pixels	14
h_{image}	image height in pixels	14
D	image domain	14
m	image	14
w_{local}	local area width in pixels	14
h_{local}	local area height in pixels	14
D_{sample}	sampling domain	14
Ω	index set for observations	14
(x_i, y_i)	observation $i \in \Omega$ pixel coordinate	14
D_{local}	local domain	14
D_i	observation $i \in \Omega$ domain	14
M_i	a random observation over D_i	15
m_i	a realization from M_i	15
ϕ	a scalar valued feature function	15
f_i	a scalar feature representation for m_i	15
$\boldsymbol{\phi}$	a vector valued feature function	15
$\mathbf{f}_i$	a vector feature representation for m_i	15
$\mathbf{F}$	a feature matrix	15
P	a set of paper properties	16
a, b, c	indexes of paper properties	16

Univariate statistics

Symbol	*Description*	*Page*
Z	a random scalar value	
z	a realization from Z	
$\mathbb{E}[Z]$	expectation of Z	
μ	mean	
$\mathbb{V}ar[Z]$	variance of Z	
σ^2	variance	
$\mathbb{C}ov[Z^1, Z^2]$	covariance between Z^1 and Z^2	
σ_{12}^2	covariance	
$\mathbb{C}or[Z^1, Z^2]$	correlation between Z^1 and Z^2	32
ρ	correlation coefficient	32
$\hat{\rho}$	classical estimator for ρ	33
$\bar{\rho}$	robust estimator for ρ	33

Multivariate statistics

Symbol	*Description*	*Page*
$\boldsymbol{Z}$	a random vector	
$\boldsymbol{z}$	a realization from $\boldsymbol{Z}$	
$\mathbb{E}[\boldsymbol{Z}]$	vector expectation	
$\boldsymbol{\mu}$	mean vector	
$\mathbb{V}ar[\boldsymbol{Z}]$	covariance between elements of $\boldsymbol{Z}$	
$\boldsymbol{\Sigma}$	covariance matrix	
$\mathbb{C}ov[\boldsymbol{Z}^1, \boldsymbol{Z}^2]$	cross-covariance between elements of $\boldsymbol{Z}^1$ and $\boldsymbol{Z}^2$	
$\boldsymbol{\Sigma}_{12}$	cross-covariance matrix	
λ_j	canonical correlation coefficient	33
$\boldsymbol{\beta}_j^1$	canonical direction vector for $\boldsymbol{Z}^1$	33
$\boldsymbol{\beta}_j^2$	canonical direction vector for $\boldsymbol{Z}^2$	33

Random fields

Symbol	*Description*	*Page*
Z	a spatial stochastic process	36
θ_Z	process parameters	40
μ	process expected value	36
$C(t)$	process covariance function	36
c_Z	process small scale discontinuity (nugget effect)	38
σ_Z^2	process variance (sill)	38
a_Z	process range (range)	38
$\gamma^{a,b}(t)$	semivariogram or cross-semivariogram between a and b	37
$\hat{\gamma}^{a,b}(t)$	classical estimator for $\gamma^{a,b}(t)$	37
$\bar{\gamma}^{a,b}(t)$	robust estimator for $\gamma^{a,b}(t)$	37
$\rho^{a,b}(t)$	correlogram or cross-correlogram between a and b	36
s	parametric Gaussian random field model scaling parameter	39

Proposed CCA method

Symbol	*Description*	*Page*
Ω	index set for observations	49
Ω_{tr}	index set for training observations	51
Ω_{val}	index set for validation observations	51
m_i^a	observation $i \in \Omega$ from a	49
m_i^b	observation $i \in \Omega$ from b	49
$\mathbf{f}_i^a$	a feature representation of m_i^a	50
$\mathbf{f}_i^b$	a feature representation of m_i^b	50
$\mathbf{F}_{tr}^a$	training data matrix for a	51
$\mathbf{F}_{tr}^b$	training data matrix for b	51
$\mathbf{F}_{val}^a$	validation data matrix for a	51
$\mathbf{F}_{val}^b$	validation data matrix for b	51
p_i^a	projected observation $i \in \Omega$ for a	55
p_i^b	projected observation $i \in \Omega$ for b	55
$\hat{\rho}^{a,b}$	classical correlation estimate between p_i^a and p_i^b	55
$\bar{\rho}^{a,b}$	rank correlation estimate between p_i^a and p_i^b	55
$\varrho^{a,b}$	proposed CCA correlation coefficient	57
r	a subimage-to-image scaling factor	50
$\varrho^{a,b}(r)$	proposed CCA correlation function	57

1 INTRODUCTION

The concern of this thesis is to develop and to evaluate a methodology for detecting and measuring spatial dependencies between real world images. These images are often generated by mutually dependent random processes. A natural source for such images is in manufacturing industry where product quality is assessed through visual observations. In this thesis we use paper measurement images as an example of how to employ the proposed methodology on a real world application.

In this chapter we give an introduction to the thesis. First we motivate the reader by presenting our assignment from the paper making industry. Then we explain how a paper measurement image is studied and how dependencies between paper measurement images are analyzed in this thesis. After that we discuss about the challenges that we are going to face in our analysis. The remainder of the chapter presents a taxonomy of the thesis and some words are dedicated to specifying the author's contribution. The chapter is concluded with a discussion on how the the thesis relates to its respective research field.

1.1 Assignment from the paper making industry

The paper making industry has decades of history in quality assessment during which many techniques have been developed for measurement of paper structure. The analysis of the measurements in terms of product quality is still based on simple methods. In fact, many measurements are inspected visually, which is mostly due to lack of a systematic analyzing methodology. This has prevented the industry from enjoying the full benefits that these measurements have to offer. Detailed surveys on paper research are available in the literature; for an introduction we refer the reader to browse (Mark, Habeger, Borch and Lyne 2002) and (Borch, Lyne, Mark and Habeger 2002).

Our experimental data set comes from an industrial partner. This data set contains two dimensional measurements of four paper sheets from smooth and rough paper types. From each sheet we have measurements of surface topography, which is measured on both sides of a paper sheet, and of mass distribution, which commonly is referred as formation in paper industry. From these measurements we are able to derive computational measurements for thickness and density distributions. The measured area is a $10cm \times 10cm$ square region that is measured with approximately $0.1mm$ resolution yielding 1024×1024 pixels images. The employed resolution is not enough to identify individual fibers but it does identify fiber clusters, which are sometimes referred as flocks in paper making. An illustration of the implemented experiment and examples of obtained measurements are presented in Figure 1. Whereas there are obvious dependencies between the measured paper properties, observable even with a naked eye, the structures in them seem complex and random to say the least. Later in the thesis we will show that this randomness follows

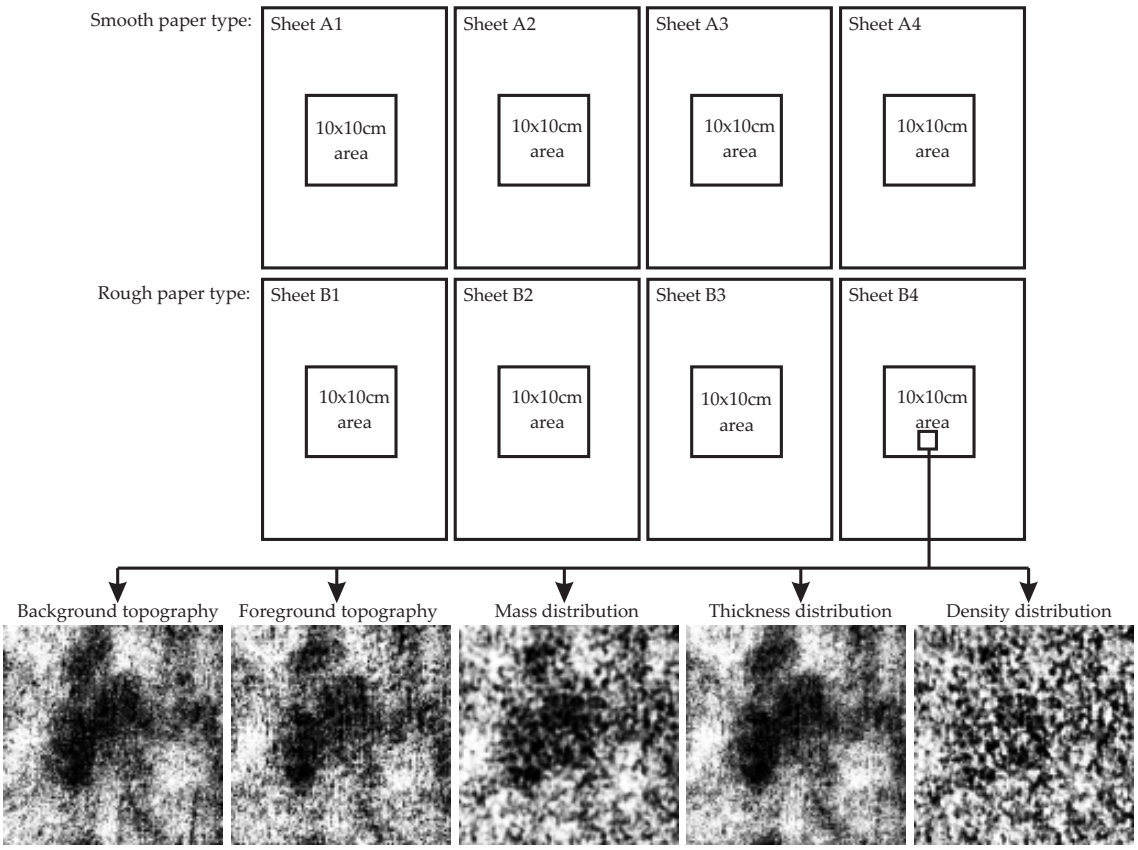


FIGURE 1: Our empirical data set. Four paper sheets of smooth and rough paper types have been measured to gain 1024×1024 pixels images from background and foreground topographies along with mass, thickness, and density distributions of the same $10cm \times 10cm$ paper area.

approximately the stochastic laws of Gaussian Random Fields (GRF).

In this thesis we use the described data set to carry out the following assignments:

- **A comparison of different measuring techniques.** Consider that we have a paper sheet, and from this sheet we have images of five different paper properties. How much do these images have in common? More explicitly what is the joint information content of these images? Are there paper properties that approximately tell the same story and hence we can just measure one of them and still get approximately the same information?

- **A comparison of different paper sheets.** Consider that we have four paper sheets from smooth and rough paper types, and from these sheets we have images of the same paper property. What are the similarities between papers of the same type and what are the distinguishing dissimilarities between papers of the two types? Are these similarities and dissimilarities consistent over the five measured paper properties or if not what are the properties that

best highlight the similarities and the dissimilarities?

Successful completion of the assignments involves design, implementation and employment of techniques that range from computer vision to spatial data analysis.

1.2 How a paper measurement image is studied in this thesis

The images that we have are two dimensional projections of paper structure: foreground and background topographies along with mass, thickness, and density distributions. Let us say that we measure the paper at some point (x, y) and thus obtain a measurement value $z(x, y)$ that we call a pixel. These pixels are quantified with $8 - 24$ bits resolution depending on the applied instrument. Sampling such pixels on a regular lattice domain

$$D = \{(x, y) : x \in \{1, \ldots, w_{image}\}, y \in \{1, \ldots, h_{image}\}\},$$

where we have width $w_{image} = 1024$ pixels and height $h_{image} = 1024$ pixels, yields us a two-dimensional image

$$m = \{z(x, y) : (x, y) \in D\},$$

where $z(x, y)$ is the value of the pixel at coordinate $(x, y) \in D$.

The phenomena that we are interested in these images are seen approximately in $0.1mm - 4.0mm$ spatial scale. In order to focus on these scales we observe the phenomena through a random sample of independently and identically distributed (iid) subimages. In our work we use subimages that have width w_{local} pixels and height h_{local} pixels. Let us define a sample domain, which is the domain area from which $w_{local} \times h_{local}$ pixels subimages can be sampled, as

$$D_{sample} = \left\{ \left\lceil \frac{w_{local}}{2} \right\rceil, \ldots, w_{image} - \left\lceil \frac{w_{local}}{2} \right\rceil \right\} \times \left\{ \left\lceil \frac{h_{local}}{2} \right\rceil, \ldots, h_{image} - \left\lceil \frac{h_{local}}{2} \right\rceil \right\},$$

where we use $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ operators for rounding up and down respectively. The subimages are sampled around points

$$(x_i, y_i) \in D_{sample},$$

where $i \in \Omega = \{1, \ldots, N\}$ are the indexes of observations. To get a desired sample of subimages, the points (x_i, y_i) are taken randomly, independently and identically distributed according to a uniform distribution over D_{sample}. These points are then used as pivot points for a local domain

$$D_{local} = \left\{ 1 - \left\lceil \frac{w_{local}}{2} \right\rceil, \ldots, w_{local} - \left\lceil \frac{w_{local}}{2} \right\rceil \right\} \times \left\{ 1 - \left\lceil \frac{h_{local}}{2} \right\rceil, \ldots, h_{local} - \left\lceil \frac{h_{local}}{2} \right\rceil \right\},$$

which defines the spatial extent of the observation, to define observation subdomains as

$$D_i = \{(x_i + x, y_i + y) : (x_i, y_i) \in D_{sample}, (x, y) \in D_{local}\}.$$

Because all the subdomains $D_i \subset D$ are proper subsets of the domain D, we can use these subdomains to define subimages

$$M_i = \{z(x, y) : (x, y) \in D_i\}$$

that give local descriptions of the paper around the pivoting points (x_i, y_i). The randomness and the independence of the points (x_i, y_i) ensure that the observed subimages M_i are also random and independent. Assuming that the studied images are stationary, the observed subimages M_i are also identically distributed according to an unknown subimage distribution that we try to evaluate empirically. Satisfying these three statistical principles ensures that we can make statistically found inference over the subimages. In empirical studies we use realizations m_i that are sampled around fixed pivot points.

The pixel representation of a subimage consumes $w_{local} \times h_{local}$ scalars of storage space. Sometimes it is possible to move to a more interpretable and computationally efficient representation by using a feature function ϕ to describe the subimage content with a single but well characterizing scalar value. Such scalar values are called features

$$\mathrm{f}_i = \phi(m_i),$$

where i indicates that we give a description of the image around point (x_i, y_i). It is common practice to use more than one feature function at the same time and we use a vector function $\boldsymbol{\phi}$ to simultaneously map a subimage to d features that are stored as elements of a d-dimensional feature vector

$$\mathbf{f}_i = \begin{pmatrix} \mathrm{f}_{i,1} \\ \vdots \\ \mathrm{f}_{i,d} \end{pmatrix} = \boldsymbol{\phi}(m_i) = \begin{pmatrix} \phi_1(m_i) \\ \vdots \\ \phi_d(m_i) \end{pmatrix}.$$

It is mathematically convenient to stack these feature vectors as rows of a feature matrix

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_1^T \\ \vdots \\ \mathbf{f}_N^T \end{pmatrix} = \begin{pmatrix} \mathrm{f}_{1,1} & \dots & \mathrm{f}_{i,d} \\ \vdots & \ddots & \vdots \\ \mathrm{f}_{N,1} & \dots & \mathrm{f}_{N,d} \end{pmatrix} = \begin{pmatrix} \phi_1(m_1) & \dots & \phi_d(m_1) \\ \vdots & \ddots & \vdots \\ \phi_1(m_N) & \dots & \phi_d(m_N) \end{pmatrix},$$

where each row represents the observations that are made from a particular subimage.

From a statistical point of view we can consider an image pixel $z(x, y)$ as a sample from a random variable $Z(x, y)$. Similarly an image $z = \{z(x, y) : (x, y) \in D\}$ is seen as a realization from a generating process $Z = \{Z(x, y) : (x, y) \in D\}$. In chapter 5 we show that preprocessed paper measurement images, at least the ones that we study in this thesis, are approximately stationary Gaussian. Thus we can assume second-order stationarity, which means that the expectation and the covariance structure of the generating process are both translation invariant over the

domain. That is, pairs of points that are some arbitrary transition (t_x, t_y) apart have

$$\mathbb{E}[Z(x+t_x, y+t_y)] = \mathbb{E}[Z(x,y)] = \mu \text{ for all } (x+t_x, y+t_y), (x,y) \in D,$$

and

$$\mathbb{C}ov[Z(x+t_x, y+t_y), Z(x,y)] = C(t_x, t_y) \text{ for all } (x+t_x, y+t_y), (x,y) \in D.$$

A stationary covariance function C, which is also known as a covariogram, depends only on the transition (t_x, t_y). In this thesis we use words image and realization from a random field interchangeably depending on the presentation context. Similarly terms generating process, stochastic process, and random field all refer to a two dimensional stationary Gaussian random field.

All the notation that is presented above has focused on studying one image at a time. In this thesis we study interconnections between multiple images and hence we prefix all variable superscripts with paper property indexes $a, b, c, \ldots \in P$ where

$$P = \{\text{background}, \text{foreground}, \text{mass}, \text{thickness}, \text{density}\}$$

for background topography, foreground topography, mass distribution, thickness distribution, and density distribution respectively. For a property a we write the measured image as $m^a = \{z^a(x,y) : (x,y) \in D\}$, the random subimages as $M_i^a = \{z^a(x,y) : (x,y) \in D_i\}$ and their realizations as m_i^a, the feature vectors as $\mathbf{f}_i^a$, the feature matrix as $\mathbf{F}^a$, and the generating process as $Z^a = \{Z^a(x,y) : (x,y) \in D\}$. The domain D, the sample domain D_{sample}, the pivot points (x_i, y_i), the local domain D_{local}, and the subdomains D_i are shared between all the paper properties and hence no superscript prefixing is needed.

1.3 How dependencies between paper measurement images are studied in this thesis

In this thesis we study three types of spatial correlation: the correlation between random point measurements $Z^a(x,y)$ and $Z^b(x,y)$, the correlation between random local area measurements M_i^a and M_i^b, and the second-order spatial correlation between random point measurements $Z^a(x+t_x, y+t_y)$ and $Z^b(x,y)$ where (t_x, t_y) is a spatial transition.

The most trivial way to evaluate dependency between images m^a and m^b is to measure correlation between random point measurements $Z^a(x,y)$ and $Z^b(x,y)$ with a correlation coefficient

$$\rho^{a,b} = \mathbb{C}or[Z^a(x,y), Z^b(x,y)] \text{ where } (x,y) \in D. \tag{1}$$

A geometrical interpretation for this correlation coefficient is obtained by plotting points $(z^a(x,y), z^b(x,y))$. Whether these points lie on a straight line formation we

have proof of strong dependency between the images, whereas the opposite case does not proof that such a dependency cannot exist as this correlation coefficient is vulnerable to different types of noise. The existing dependency may be hiding behind obscuring phenomena. An example plot of the dependency between mass distribution and foreground topography is presented in Figure 2. From this plot we can see that there is a weak connection between the two paper properties. This connection is however among the strongest between our images and hence this approach is clearly not sufficient for our purposes.

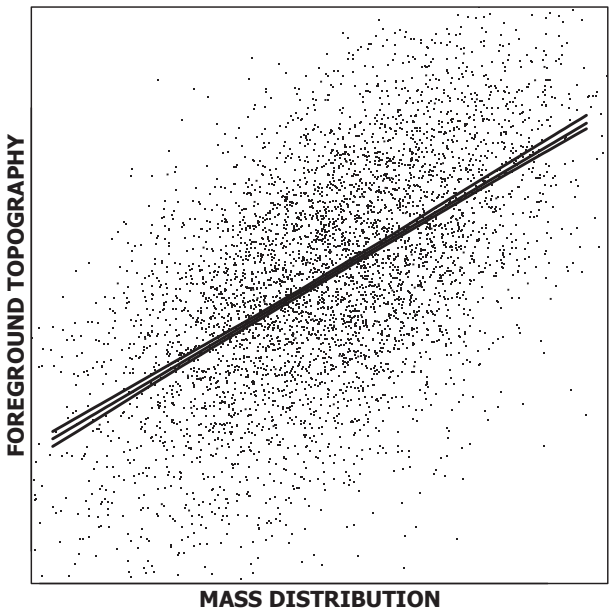


FIGURE 2: Correlation between pixels of mass distribution and foreground topography. Each point represents a point measurement of two paper properties. Whereas the presented dependency between mass distribution and foreground topography is one of the strongest in our data set, this approach is unable to provide a good explanation for the dependency between these two paper properties.

Instead of individual point measurements we can enlarge our scope of observation to measurements of local areas. The correlation between random subimages is measured with a correlation coefficient

$$\varrho^{a,b} = \mathbb{C}or[\bar{\phi}^a(\{Z^a(x,y) : (x,y) \in D_i\}), \bar{\phi}^b(\{Z^b(x,y) : (x,y) \in D_i\})],$$

where operators $\bar{\phi}^a(\cdot)$ and $\bar{\phi}^b(\cdot)$ are special types of feature functions that compress the information of a subimage into a scalar value in an optimal way. The way these feature functions are constructed require a careful discussion, for a detailed discussion on the subject see chapter 3, and is not a subject for this introduction. A geometrical visualization for this correlation coefficient is gained by plotting points that represent pairs $(\bar{\phi}^a(m_i^a), \bar{\phi}^b(m_i^b))$. The interpretation for this plot is almost the same that it is with the correlation between pixels: whether the points lie on a straight line we have proof of strong dependency between the images, whereas the

opposite case again does not proof that a dependency does not exist. An example plot of the dependency between mass distribution and foreground topography is presented in Figure 3. This is the same case that was previously examined with the correlation between pixels and thus we can see that in this case the correlation between subimages is clearly able to outperform the correlation between pixels. This is because the correlation between subimages is more resilient to different types of noise. An interesting extension to this idea is to evaluate the amount of correlation between subimages of different spatial scales. An example plot of correlation between mass distribution and mass distribution as a function of spatial scale is presented in Figure 4. According to this plot the amount of correlation seems to increase with the increasing size of the subimages.

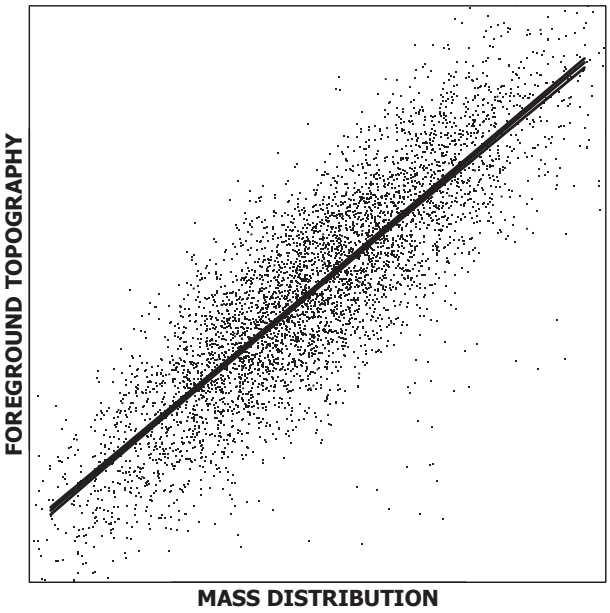


FIGURE 3: Correlation between subimages of mass distribution and foreground topography. Each point represents a local area measurement of two paper properties that are compressed into two scalar values in an optimized way. As we can see the level of observed dependency is clearly higher than we saw in Figure 2.

Because our image content is approximately stationary Gaussian, we can study second-order statistical correlations within ($a = b$) and between ($a \neq b$) images m^a and m^b with a semivariogram function

$$\gamma^{a,b}(t_x, t_y) = \frac{1}{2}\mathbb{V}ar[Z^a(x + t_x, y + t_y) - Z^b(x, y)] \text{ where } (x + t_x, y + t_y), (x, y) \in D, \tag{2}$$

which is also known as a cross-semivariogram function when $a \neq b$. The semivariogram function measures the amount of dissimilarity between point measurements $Z^a(x + t_x, y + t_y)$ and $Z^b(x, y)$, that is pixels of transition (t_x, t_y) apart. The larger values the semivariogram function gets at a given transition (t_x, t_y) the weaker dependency these pixels have. Typically the dependency gets weaker as the transition

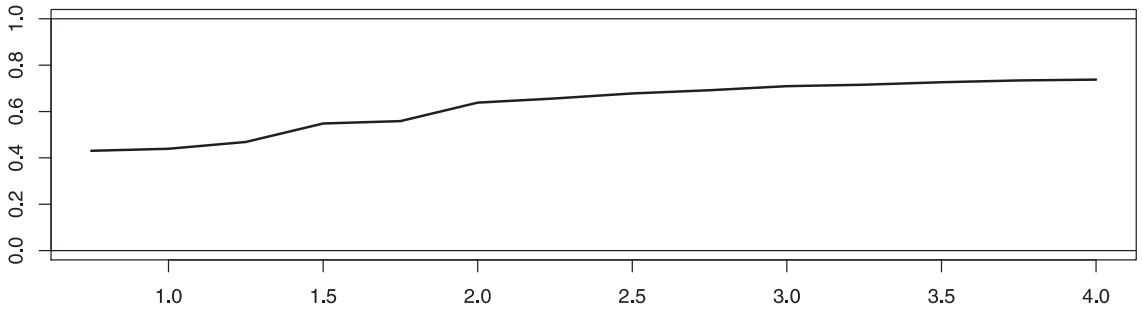


FIGURE 4: Correlation between subimages of mass distribution and foreground topography as a function of spatial scale. The horizontal axis indicates the width and the height of the subimages (in millimeters) between which correlation is measured, and the vertical axis measures the amount of correlation.

becomes larger. A semivariogram function for our example case of mass distribution and foreground topography is presented in Figure 5. From this plot we can see how the amount of dependency between the pixels decreases as a function of distance.

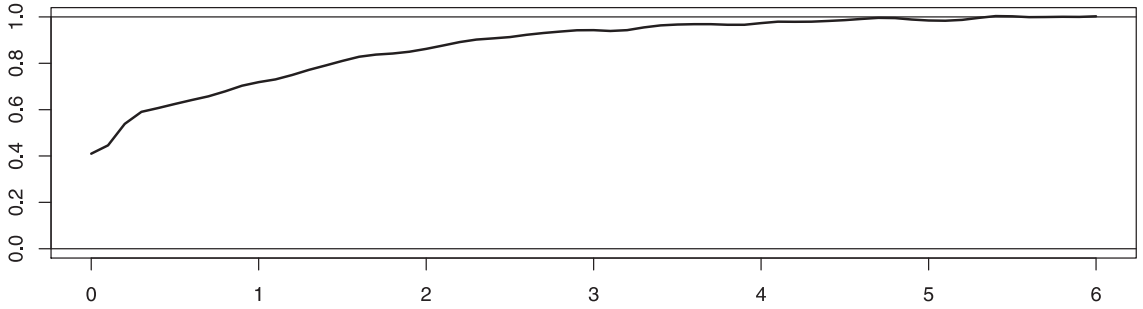


FIGURE 5: Cross-semivariogram between foreground topography and mass distribution. The horizontal axis indicates the transition in millimeters between the two points between which dissimilarity is measured, and the vertical axis indicates the amount of observed dissimilarity.

The correlation between pixels is an ad hoc solution for measuring spatial dependencies between images. In practice its performance is poor and is of interest merely as a point of reference to the other more successful methods. The correlation between subimages is a new dependency measure, which is described in chapter 3. This correlation measure is able to overcome most of the problems of the correlation between pixels. The second-order spatial correlation is a standard dependency measure in spatial statistics. This correlation measure is comparable to the correlation between subimages, approximately the same information is provided in a different form, and hence it serves as a good point of comparison. For an introduction to second-order spatial correlations see chapter 2, or (Cressie 1993, p. 40 and pp. 58-104) and (Cressie and Wikle 1998).

1.4 Foreseeable challenges

There are foreseeable challenges that we must prepare ourselves to when carrying out an analysis on the paper measurement images. Here is a list of the five most important that ones we can think of:

- **Spatial distortions.** In many cases, and in the case of our paper application, the raw images come from independent instruments and thus differ slightly in the measured paper area and in the employed measurement resolution. To make such images suitable for our methodology they need to be spatially transformed so that all the pixels that are measured from a particular paper location are placed at the same pixel coordinate. In the case of our paper application, this transformation is difficult to automate and thus requires laborious work by hand, which inevitably subjects the obtained results to human error. The present variation in the employed measurement resolutions inflict additional errors.

- **Measurement noise.** Unprocessed images tend to contain random noise from an unknown distribution. Lack of knowledge on this distribution makes corrective procedures difficult. A typical solution is to make an educated guess on the shape of the distribution, and depending on how successful the guess is, it is usually possible to restore the image very close to its correct condition. Sometimes the presence of noise is so severe that even the best correction algorithms cannot remedy the damage and the image becomes unusable. Because the measurement noise problem is so common in practical applications, there is an entire field in statistics developing robust version of the classical methodology that is able to function properly in the presence of measurement noise.

- **Erroneous and missing measurements.** Faulty pixels are the most common reason for a measurement to become unusable. Much depends on the used instrument as some errors are correctable while other render the measurement worthless. A chapter of their own are incomplete measurements in which part of the intended area is in practice not measured at all. In the case of spatial measurements, there is usually not much to do to compensate such a loss.

- **Unknown phenomena.** Sometimes measurements contain ambiguous phenomena that cannot be explained. This is especially the case in explorative data analysis where the sole purpose of the study is to reveal previously unknown information. For not to miss any relevant pieces of information we need to include all the ambiguous observations into our analysis , and as a result many peculiarities find their way into final results. In some sense this is unavoidable but in the same it is the single largest cause of concern when we are interpreting our results. It has happened more than one occasion that the entire interpretation is heavily biased due to such an event.

- **Computational load.** High resolution imaging has a way of generating massive data sets, which again has a direct bearing on computational resources needed to complete the analysis. A typical data set is likely to consume several gigabytes of storage space, an amount considered to be a lot at the time this thesis was published. Classical methodology is based on heavy algebraic operations that scale ill to large problems. Typically processing power and memory demand scale according to second or third order polynomials leaving all but the smallest problems unsolved. Whereas the computational resources have been doubling for every 18 months for a long time, the demand is still increasing in even more gigantic steps and thus the methodological development offers the only viable solution for this problem.

1.5 The contents of this thesis

Chapter 1 introduces our assignment from the paper making industry to motivate the reader for things to come. First we presented an experiment from which we have images of five paper properties from eight paper sheets. Then we discussed about how these images and dependencies between these images are studied in this thesis. Here we present the taxonomy of the remainder of the thesis and then dedicate some words to evaluate the author's contribution to the research field.

Chapter 2 introduces the methodological foundation of our work. We begin by introducing a general taxonomy for image analysis and then present some image preprocessing techniques that we later use in our work. Then we introduce statistical tools for analyzing dependencies between two univariate or multivariate random variables, and spatial dependencies between two Gaussian random fields. The remainder of the chapter reviews six studies to introduce current research from the field of paper science.

Chapter 3 presents the proposed methodology in detail. First we contemplate how to make proper observations from the images and how this affects the developed methodology. Then we present our strategy for breaking large images into smaller and more manageable subimages. These subimages are then used to build models of dependency between two images based on their mutual information. The subimages are analyzed with canonical correlation analysis (CCA) from which we obtain a set of mutually dependent visual operator pairs that can be used to detect changes in paper structure. The remainder of the chapter makes empirical experiments with the methodology with respect to the challenges that we have in our paper application.

Chapter 4 describes an example implementation of the proposed methodology. A computer implementation is necessary in order to analyze our application data and in order to empirically evaluate the proposed methodology. First we present five design goals and a general system architecture for data analysis. Then we give a brief description of an example implementation that we use in our own work and evaluate it through the five design goals that we set in the beginning of the chapter.

Chapter 5 uses our real world application to empirically evaluate the proposed

methodology. We start by introducing the measurements and then consider the type of phenomena that we can see in the available images. Each image is then reviewed and conspicuous phenomena in them are studied in detail. Then we conduct a tentative analysis of the images using standard statistical tools from geostatistics. Here it turns out that paper images, which seem visually complex, can be characterized to large extent with parametric Gaussian random field models. The analysis in continued with our own methodology to identify and explain dependencies between the five studied paper properties, and to identify dissimilarities between the two studied paper types. Throughout the analysis we contemplate the obtained results from both application and methodological perspective. This chapter presents the main results with respect to our paper application and raises a few new topics for future research.

Chapter 6 collects the main results and the most important conclusions from our work.

1.6 Contribution of the author

During the postgraduate years the author has received counseling and assistance from DrTech Pasi Koikkalainen, the work mentor, and from other members of the research group: PhD Anssi Lensu, PhD Jouni Raitamäki, MSc Michael Haranen and MSc Ismo Horppu.

The problem setting and the empirical data set, which were introduced in chapter 1, were obtained from our partner in cooperation Oy Keskuslaboratorio-Centrallaboratorium Ab (KCL). On matters regarding this data set the author has been in regular contact with KCL employees DrTech Erkki Hellén and MSc Pasi Puukko. Preprocessing of the images required a lot of manual work and hence was implemented as a collaboration between the author and KCL personnel.

The literature survey, which is presented in chapter 2, has been conducted by the author under the mentor DrTech Pasi Koikkalainen's guidance. The presentation is based on the material that were found from Jyväskylä University Library, from various online libraries, and from Internet homepages of many individual researchers.

The proposed methodology for detecting and measuring dependencies between images, which is presented in chapter 3, is a collaboration between the author and the mentor DrTech Pasi Koikkalainen. With a few aspects of the methodology the author has gained additional expertise from consultations with Prof Antti Penttinen.

The working environment for the study, which is presented in chapter 4, is based on an open source project "The R Project for Statistical Computing". For additional information about the R project the reader is referred to (The R Foundation 2005a) and (The R Foundation 2005b). The design and implementation of the required extension libraries that are written in C++, and the user interface that is written in R macro language are the sole products of the author.

The data analysis that is conducted on the paper images, which is presented in chapter 6, is performed solely by the author. The mentor DrTech Pasi Koikkalainen

has implemented some part of the analysis on his own and with an another software platform. The results from these two studies are consistent for the most part and hence boost our confidence on the validity of the obtained results.

As a final remark, the work mentor DrTech Pasi Koikkalainen has advised the author on matters of writing and hence has directly influenced the written form of the thesis.

1.7 Discussion

The assignment from the paper making industry and the proposed methodology that is used to carry out this assignment fall into a research field that has not received much attention. We are unaware of parallel studies that would answer the type of questions that we have in the current paper application. Similarly we are unaware of other studies that would employ techniques that are analogous to those that we propose for detecting and measuring spatial dependencies between images that have stochastic content. Such images are easily obtained from various industrial manufacturing processes. The studied paper images constitute just an example of the many possible applications.

2 INTRODUCTION TO METHODOLOGICAL FOUNDATION

This chapter provides a review of image analysis and correlations. The apparent richness of information in the paper images, which were introduced in the previous chapter, and the complexity of the needed analysis lead us to techniques that are developed in the fields of image analysis and statistics. First we study the principles of and then learn a few standard techniques from image analysis: geometric transformation of images, background trend removal from images, noise removal from images, and image enhancement in the frequency domain. These techniques are later used for preprocessing in our own data analysis. Then we move on to techniques for measuring correlations: tools from classical statistics, such as correlation and canonical correlation analysis, and more specific tools from spatial statistics, such as correlogram and semivariogram. These tools are later used to analyze the preprocessed content of the studied images. In the remainder of the chapter we review studies from the field of paper science to gain an understanding of the current state-of-the-art in this field. Finally we conclude the chapter with a discussion.

2.1 An introduction to image analysis

Image analysis consists of large variety of techniques that range from pixel-level operations, such as noise filtering and feature extraction, to higher-level operations that come mostly from pattern recognition, such as classification. In practical applications image analysis divides logically into three relatively independent stages. These three stages are low-level processing, intermediate-level processing, and high-level processing (see Figure 6). Although the boundaries between these stages are not clear and depend on the application context, the presented taxonomy is general enough to fit most applications of imaging. Whereas there are many detailed taxonomies that are presented in the literature, we make use of the one that is presented in (Gonzales and Woods 1992, pp. 572-573).

Low-level processing consists of image acquisition and image preprocessing, which are typically performed close to the imaging hardware. The image acquisition contemplates good techniques and settings for imaging, whereas the image preprocessing considers ways to improve the quality of images. Typical operations perform mechanical pixel-level processing that correct distortions and emphasize essential information. Characteristic to this stage is having large amount of relatively unprocessed numerical data without a good understanding of their content. Hence, the performed operations have to be selected based on some kind of prior information about the imaging system.

Intermediate-level processing consists of image segmentation and image description that are performed on preprocessed images. The objective is to extract the studied objects from the background and then to describe them through an alterna-

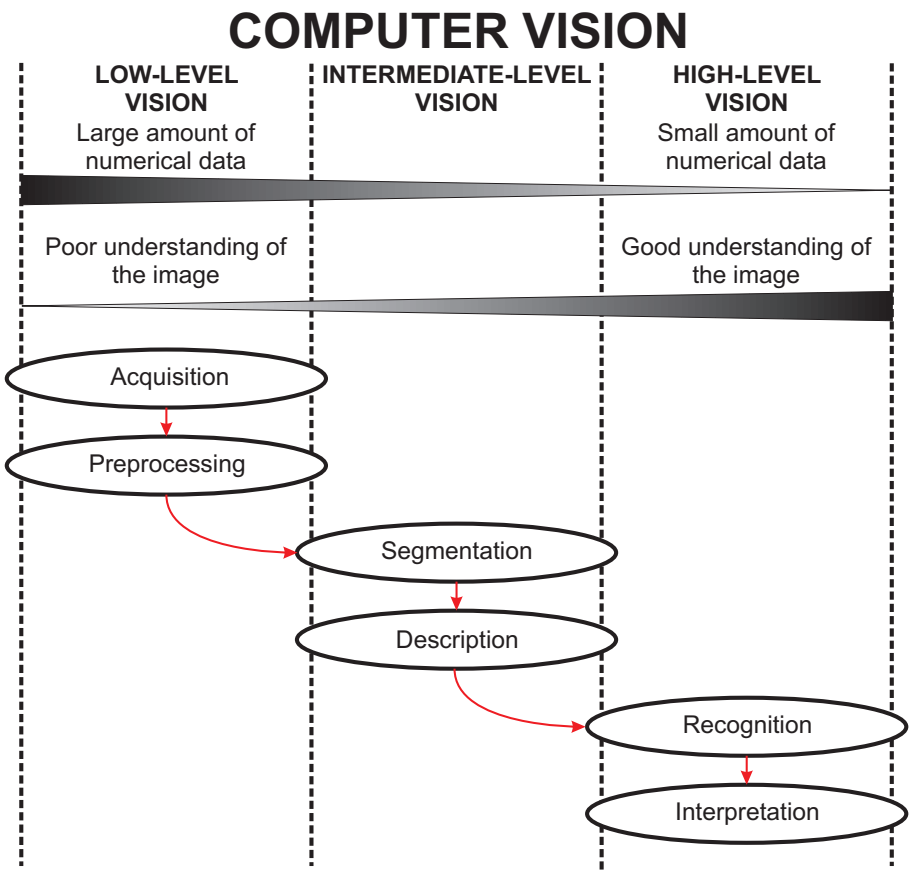


FIGURE 6: Taxonomy of image analysis.

tive representation. Whereas intermediate-level operations also work on pixel-level, their end result is a higher-level representation that characterizes the image content in a structured manner.

High-level processing involves image recognition and image interpretation based on the alternative representation that was generated at the previous stage. The two most common objectives of this stage are to classify the studied objects into predefined classes and to interpret the image content from the application perspective. The goal is to answer the questions that are posed from the application context. Typically the employed solutions are heavily application dependent and careful consideration is needed in order to obtain the best functioning system.

2.2 Preprocessing of images

In this section we learn four image preprocessing techniques. The notation that was introduced in chapter 1 is used for all equations. In addition to that, $z(x, y)$ is the unprocessed pixel intensity and $z'(x, y)$ is the preprocessed pixel intensity at a pixel coordinate $(x, y) \in D$.

2.2.1 Geometric transformation of images

Geometric transformations are standard operations in image analysis. These operations are often used to correct geometric distortions, such as rotation, scaling, and skewing. In addition they are used to solve pixel correspondence problems between images. The ideas that are presented here are commonly known and discussed in almost all books covering computer vision. From this introduction we learn the general idea according to (Sonka, Hlavac and Boyle 1999, pp. 62-68). An alternative introduction to the subject is available in (Gonzales and Woods 2002, pp. 270-275).

The general idea is to use a geometric transformation function $\mathbf{T}(x, y) = (T_x(x, y), T_y(x, y))$ to map pixel coordinates (x, y) into new pixel coordinates (x', y') such that

$$x' = T_x(x, y) \text{ and } y' = T_y(x, y). \tag{3}$$

A more detailed inspection of practical implementations shows that geometric transformation consists of two basic operations: geometric transformation of pixel coordinates, and interpolation of pixel intensities.

The pixel correspondence structure can be identified based on prior information about the correspondence. This information comes for example from our knowledge of the imaging system, from similarities between local image areas, or from special markers that are used to pinpoint pairs of pixels with known mutual correspondence. The given list is not complete but gives a rough understanding over the variety of possibilities. In the case of special markers one formulation for the geometric transformation of equation 3 is to write

$$x' = a_0 + a_1 x + a_2 y \text{ and } y' = b_0 + b_1 x + b_2 y. \tag{4}$$

The transformation coefficients a_0, a_1, a_2, and b_0, b_1, b_2 are determined as a solution to a least squares problem, which is based on N example pairs of coordinates (x_i, y_i) and (x'_i, y'_i), where $i \in \{1, \ldots, N\}$. In the case of affine transformations, such as the one that is formalized in equation 4, at least three example pairs of coordinates are needed.

In the case that was presented in equation 4, the inferred geometric transformation $\mathbf{T}$ is a continuous function. This means that mapping from integer valued coordinates (x, y) gives real valued coordinates (x', y'), which is a problem when working with the discrete rasters of digital images. A common solution is to approach the problem so that the pixel intensity at an integer valued coordinate (x', y') is determined through interpolation from intensities of pixels that are close to the real valued coordinate (x, y). More formally, a bi-cubic interpolation estimate for the pixel intensity at integer valued coordinate (x', y') is written as

$$z'(x', y') = \sum_{(s_x, s_y) \in D} K(x - s_x, y - s_y) z(s_x, s_y) \tag{5}$$

where K is a bi-cubic interpolation kernel function

$$K(t_x, t_y) = \begin{cases} 1 - 2||(t_x, t_y)||^2 + ||(t_x, t_y)||^3, & \text{for } 0 \leq ||(t_x, t_y)|| < 1 \\ 4 - 8||(t_x, t_y)|| + 5||(t_x, t_y)||^2 - ||(t_x, t_y)||^3, & \text{for } 1 \leq ||(t_x, t_y)|| < 2 \\ 0, & \text{otherwise,} \end{cases}$$

and $x = T_x^{-1}(x', y')$ and $y = T_y^{-1}(x', y')$. The bi-cubic interpolation estimate that is presented in equation 5 is not the only viable solution but this estimate has gained wide acceptance in computer vision due to its good properties.

2.2.2 Background trend removal from images

We typically are interested in phenomena that are seen in some specific spatial scale and the phenomena outside of this scale are likely to disturb our analysis. Unwanted phenomena of too large spatial scale, commonly known as image trending problem, raise from the unfortunate fact that some instruments are unable to produce a homogeneous measurement of the whole image area. One approach to trend removal is to build a global model that explains the large scale variation. A commonly used algorithm for this is median polishing, which is introduced for example in (Cressie 1993, pp. 46-48 and pp. 183-190). The founding idea behind polishing is to decompose large scale variation into average, row and column effects. More formally, there is a position independent average effect z^a, an x-coordinate dependent column effect $z^c(x)$, and a y-coordinate dependent row effect $z^r(y)$ that generate the background trend

$$z^\dagger(x, y) = z^a + z^c(x) + z^r(y) \text{ for } (x, y) \in D.$$

Whereas these effects are obviously unknown, they can be estimated with the polishing algorithm that is presented in Table 1. The algorithm begins from a real world image and iteratively sweeps medians out of rows and then out of columns of the image until the algorithm converges. The convergence is commonly measured with the amount of change between iterations and when it becomes small enough the iterating is stopped. A trend corrected image is then obtained by subtracting the estimated trend as

$$z'(x, y) = z(x, y) - z^\dagger(x, y).$$

The choice to use medians to estimate average, column and row effects is wise but not the only viable option. For a discussion about the merits of different estimates see the reference that is mentioned above.

2.2.3 Noise removal from images

In real world applications measurements contain noise, which comes from an unknown distribution that depends on the employed instrument. The most common type of noise in real world images is additive noise, which is defined as

$$z(x, y) = z_{true}(x, y) + \epsilon(x, y),$$

1. Initialize:
   (a) $z^a \leftarrow 0$
   (b) $z^c(x) \leftarrow 0$ for all $x \in \{1, \ldots, w_{image}\}$
   (c) $z^r(y) \leftarrow 0$ for all $y \in \{1, \ldots, h_{image}\}$.
2. Iterate while changes are large enough:
   (a) Sweep medians out of columns
      i. $z^{c'}(x) \leftarrow \text{median}(\{z(x', y) : (x' = x, y) \in D\})$ for all $x \in \{1, \ldots, w_{image}\}$
      ii. $z(x, y) \leftarrow z(x, y) - z^{c'}(x)$ for all $(x, y) \in D$
      iii. $z^c(x) \leftarrow z^c(x) + z^{c'}(x)$ for all $x \in \{1, \ldots, w_{image}\}$
   (b) and adjust the global average
      i. $z^{c''} \leftarrow \text{median}(\{z^c(x) : x \in \{1, \ldots, w_{image}\}\})$
      ii. $z^c(x) \leftarrow z^c(x) - z^{c''}$ for all $x \in \{1, \ldots, w_{image}\}$
      iii. $z^a \leftarrow z^a + z^{c''}$.
   (c) Sweep medians out of rows
      i. $z^{r'}(y) \leftarrow \text{median}(\{z(x, y') : (x, y' = y) \in D\})$ for all $y \in \{1, \ldots, h_{image}\}$
      ii. $z(x, y) \leftarrow z(x, y) - z^{r'}(y)$ for all $(x, y) \in D$
      iii. $z^r(y) \leftarrow z^r(y) + z^{r'}(y)$ for all $y \in \{1, \ldots, h_{image}\}$
   (d) and adjust the global average
      i. $z^{r''} \leftarrow \text{median}(\{z^r(y) : y \in \{1, \ldots, h_{image}\}\})$
      ii. $z^r(y) \leftarrow z^r(y) - z^{r''}$ for all $y \in \{1, \ldots, h_{image}\}$
      iii. $z^a \leftarrow z^a + z^{r''}$.
3. The algorithmic estimate for the large scale variation is

$$z^\dagger(x, y) = z^a + z^c(x) + z^r(y) : (x, y) \in D.$$

TABLE 1: Median polish algorithm (Cressie 1993).

where $z(x, y)$ corresponds to the measured pixel intensity, $z_{true}(x, y)$ corresponds to the true pixel intensity, and $\epsilon(x, y)$ corresponds to the added noise respectively. As (Gonzales and Woods 2002, pp. 230-242) suggests, spatial filtering is the method of choice in situations when only additive noise is present. An another introduction to the subject is in (Sonka, Hlavac and Boyle 1999, pp. 68-77).

One common type of additive noise, which is known as Gaussian noise, has $\epsilon(x, y)$ independently and identically distributed (iid) from a Gaussian distribution.

In a real world image this noise shows as an unexplained variation in the areas that should be displaying a constant pixel intensity. Because the effect of this noise to pixel intensity is symmetric and spatially uncorrelated, a filter based on smoothing over a local area of pixel intensities yields an acceptable result. The simplest and the most commonly applied of such filters is the mean filter. The mean filter replaces the measured pixel intensity with the average of local area pixel intensities. Formally the mean filter estimate for the pixel intensity at coordinate (x, y) is written as

$$z'(x,y) = \frac{1}{w_{local}h_{local}} \sum_{(s_x,s_y)\in D_{local}} z(x+s_x, y+s_y).$$

An another type of additive noise, which is known as impulse noise or salt-and-pepper noise, consists of sharp local spikes. In a real world image it shows as grain like structures having pixel intensities that clearly separate from the surrounding pixel intensities in a manner that is not consistent with the image content. Whereas the noise effectively replaces the original pixel intensities in these grains, a filter that is based on smoothing yields inadequate results. One solution is to use a filter that is based on the order-statistics of the local area pixel intensities. Of such filters, the median filter that replaces the measured pixel intensity with the local area pixel intensity that has rank order $\lfloor w_{local}h_{local}/2 \rfloor$ among the local area pixel intensities is the best known. More formally, a median filter estimate for the pixel intensity value at coordinate (x, y) is written as

$$z'(x,y) = \text{median}(\{z(x+s_x, y+s_y) : (s_x, s_y) \in D_{local}\}).$$

2.2.4 Image enhancement in frequency domain

Switching from spatial domain into frequency domain through the Fourier transformation

$$\mathfrak{F}z(u,v) = \frac{\sum_{x=1}^{w_{image}} \sum_{y=1}^{h_{image}} ((-1)^{x+y} z(x,y)) e^{-2j\pi((u-1)(x-1)/w_{image} + (v-1)(y-1)/h_{image})}}{\sqrt{w_{image}h_{image}}}$$

for $(u, v) \in D$, and back through the inverse transform

$$z(x,y) = \frac{(-1)^{x+y} \sum_{u=1}^{w_{image}} \sum_{v=1}^{h_{image}} (\mathfrak{F}z(u,v)) e^{2j\pi((u-1)(x-1)/w_{image} + (v-1)(y-1)/h_{image})}}{\sqrt{w_{image}h_{image}}}$$

for $(x, y) \in D$ are one of the most important operations in image preprocessing and analysis (Gonzales and Woods 2002, pp. 147-214 and pp. 242-253). The multiplier $(-1)^{x+y}$ is not compulsory but is used to transform the DC peak at power spectrum $|\mathfrak{F}z(u,v)|^2$ from $(1, 1)$ to $(w_{image}/2 + 1, h_{image}/2 + 1)$, which eases visual interpretation. The decomposition of an image into different frequencies makes it easy to study various periodic phenomena. From computational perspective, as performing convolution of images z and h in the spatial domain equals performing multiplication

of the same images in the frequency domain

$$\begin{aligned}(z \star h)(x,y) &= \frac{1}{w_{image}h_{image}} \sum_{s_x=1}^{w_{image}} \sum_{s_y=1}^{h_{image}} z(x,y)h(x-s_x, y-s_y) \\ &= \mathfrak{F}^{-1}(\mathfrak{F}z\mathfrak{F}h)(x,y) \text{ where } (x,y) \in D \text{ and} \\ (\mathfrak{F}z\mathfrak{F}h)(u,v) &= \mathfrak{F}z(u,v)\mathfrak{F}h(u,v) \text{ for } (u,v) \in D.\end{aligned}$$

Convolution theorem provides a fast implementation for large convolution masks. In addition it is possible to derive new types of filters for which implementations would be difficult in the spatial domain. Such are the frequency filters that are presented in the following three paragraphs. For these the filtered image is obtained via inverse transform

$$z'(x,y) = \mathfrak{F}^{-1}(\mathfrak{F}z\mathfrak{F}h)(x,y) \text{ for } (x,y) \in D,$$

where $\mathfrak{F}h$ is the employed frequency filter. For mathematical convenience we assume that $N = w_{image} = h_{image}$ is a power of two. The corresponding reject filters are obtained as $\mathfrak{F}h_R(u,v) = 1 - \mathfrak{F}h_P(u,v)$.

A lowpass filter allow frequencies that are lower than some given value f to pass the filter. For mathematical convenience a distance measure from the DC peak is defined as

$$dist(u,v) = ||(u - N/2 + 1, v - N/2 + 1)||.$$

Three of the most common lowpass filters are then formally defined as

$$\mathfrak{F}h_{IL}(u,v) = \begin{cases} 1 & \text{if } dist(u,v) < f \\ 0 & \text{otherwise} \end{cases} \quad \text{for } (u,v) \in D,$$

$$\mathfrak{F}h_{BL}(u,v|n) = \frac{1}{1 + \left(\frac{dist(u,v)}{f}\right)^{2n}} \text{ for } (u,v) \in D, \text{ and}$$

$$\mathfrak{F}h_{GL}(u,v|\sigma) = e^{-dist(u,v)^2/2\sigma} \text{ for } (u,v) \in D$$

for ideal, Butterworth, and Gaussian lowpass filters respectively. Here the filter parameters f, n and σ define the transfer function of the filter.

A bandpass filter passes frequencies that belong to a certain $[f_1, f_2]$ frequency band. Three of the most common bandpass filters are defined as

$$\mathfrak{F}h_{IB}(u,v) = \begin{cases} 1 & \text{if } f_1 \leq dist(u,v) \leq f_2 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } (u,v) \in D,$$

$$\mathfrak{F}h_{BB}(u,v|n) = 1 - \frac{1}{1 + \left(\frac{dist(u,v)(f_2-f_1)}{dist(u,v)^2-(f_1/2+f_2/2)^2}\right)^{2n}} \text{ for } (u,v) \in D, \text{ and}$$

$$\mathfrak{F}h_{GB}(u,v) = e^{-\frac{1}{2}\left(\frac{dist(u,v)^2-(f_1/2+f_2/2)^2}{dist(u,v)(f_2-f_1)}\right)^2} \text{ for } (u,v) \in D$$

for ideal, Butterworth, and Gaussian band pass filters respectively. The filter parameter n functions as previously defined.

A notch pass filter blocks frequencies that are located around some prototype frequency that is located at (u_0, v_0) in the two dimensional power spectrum. For mathematical convenience two additional distance measures, which are symmetric with respect to the DC peak, are defined as

$$dist_1(u, v|u_0, v_0) = ||(u - N/2 - u_0 + 1, v - N/2 - v_0 + 1)||$$
$$dist_2(u, v|u_0, v_0) = ||(u - N/2 + u_0 + 1, v - N/2 + v_0 + 1)||$$

along with an arbitrary threshold coefficient $size$ that indicates the size of the filter. With the help of these the three most common notch pass filters are defined as

$$\mathfrak{F}h_{IN}(u, v) = \begin{cases} 1 & \text{if } dist_1(u, v|u_0, v_0) \leq size \text{ or } dist_2(u, v|u_0, v_0) \leq size \\ 0 & \text{otherwise} \end{cases} \quad \text{for } (u, v) \in D,$$

$$\mathfrak{F}h_{BN}(u, v) = 1 - \frac{1}{1 + \left(\frac{size^2}{dist_1(u,v|u_0,v_0), dist_2(u,v|u_0,v_0)}\right)^{2n}} \text{ for } (u, v) \in D, \text{ and}$$

$$\mathfrak{F}h_{GN}(u, v) = e^{-\frac{1}{2}\left(\frac{dist_1(u,v|u_0,v_0) dist_2(u,v|u_0,v_0)}{size^2}\right)} \text{ for } (u, v) \in D$$

for ideal, Butterworth, and Gaussian notch pass filters respectively. Again the filter parameter n functions as previously defined.

The frequency filters that have been presented this far have an effect on all the pixels of the processed image. This means that those pixels that are unaffected by the studied periodic pattern are changed unnecessarily, which may lead into loss of valuable information. This is unfortunate as the filtered periodic pattern may be confined to only certain image areas leaving most of the pixels unaffected. This is the case with images showing paper structure where a phenomenon known as wire is caused by the pattern of the conveyer felt being mechanically superimposed on the otherwise stochastic pulp distribution (see chapter 5). The wire pattern consists of a few frequency spikes but is not present throughout the paper and hence a more intelligent way to remove it is needed.

One possible solution is to use adaptive filtering in which the founding idea is to use frequency filters to identify a periodic pattern, then bring it back to spatial domain, and finally remove it only from the affected image pixels. Let us assume that a frequency filter $\mathfrak{F}h(u, v)$ combines the previously presented pass filters such that taking the inverse transform of $\mathfrak{F}z\mathfrak{F}h$ yields an image of the periodic pattern

$$z^{\dagger}(x, y) = \mathfrak{F}^{-1}(\mathfrak{F}z\mathfrak{F}h)(x, y) \text{ for } (x, y) \in D.$$

The next step is to estimate a proper spatial weighting $z^{\omega}(x, y)$ that determines where this pattern is present and how much of it is present so that for each pixel it is known how much of the periodic pattern must be subtracted to properly remove

the pattern. One way to define this weighting is to minimize the local area variance of the filtered image. For this the weighting

$$z^{\omega}(x,y) = \frac{\overline{z(x,y)z^{\dagger}(x,y)} - \overline{z(x,y)}\ \overline{z^{\dagger}(x,y)}}{\overline{z^{\dagger}(x,y)z^{\dagger}(x,y)} - \overline{z^{\dagger}(x,y)}\ \overline{z^{\dagger}(x,y)}},$$

where

$$\overline{z(x,y)} = \frac{\sum_{(s_x,s_y)\in D_{local}} z(x+s_x, y+s_y)}{w_{local}h_{local}}$$

is the mean of the local area pixel intensities of image z and the other local area means are defined in an analogous manner. A filtered image is then obtained by subtracting the periodic phenomenon as

$$z^{'}(x,y) = z(x,y) - z^{\omega}(x,y)z^{\dagger}(x,y).$$

2.2.5 Other image enhancement techniques

Beyond the four presented image preprocessing techniques, there is abundance of other image enhancement techniques that are presented in the literature; for an overview see (Gonzales and Woods 2002), (Sonka, Hlavac and Boyle 1999), (Jain 1988), (Davies 2005), or (Bovik et al. 2000). The image preprocessing requirements vary heavily from application to application and depending on the employed measuring instrument. Hence it is impossible to give an all-inclusive list of the problems encountered. The presented four problems and their solutions are the most interesting from our perspective. Even in these the presented solutions are by no means the only viable options and the reader is encouraged to acquaint himself with the provided references.

2.3 Analyzing statistical dependencies between two data sets

In this section we present three techniques for analyzing statistical dependencies between two data sets. Some of these techniques are later used in the proposed methodology whereas others are used as a point of reference for a methodological comparison.

2.3.1 Correlation between two univariate random variables

The best known and the most widely used measure of linear dependency between two random variables Z^a and Z^b is the correlation coefficient

$$\rho = \mathbb{C}or[Z^a, Z^b],$$

which has range $[-1, 1]$. The interpretation of a correlation coefficient reads as follows: the sign of the coefficient determines whether the correlation is positive,

that is if one of the variables becomes larger the other one becomes larger as well, or negative, that is if one of the variables becomes larger the other one becomes smaller, whereas the absolute value of the coefficient determines how strong the dependency is. An absolute coefficient value equal to 0.5 corresponds to 25% of the variation been explained, equal to 0.7 corresponds to 49% of the variation been explained, and equal to 1 corresponds to 100% of the variation been explained respectively.

For a sample of observed values $z_1^a, \ldots, z_N^a \sim Z^a$ and $z_1^b, \ldots, z_N^b \sim Z^b$, an empirical estimate for the correlation coefficient is obtained from Pearson product moment correlation estimate that is better known as Pearson's correlation estimate

$$\hat{\rho} = \frac{N\sum_{i=1}^N z_i^a z_i^b - \sum_{i=1}^N z_i^a \sum_{i=1}^N z_i^b}{\sqrt{N\sum_{i=1}^N \left(z_i^a\right)^2 - \left(\sum_{i=1}^N z_i^a\right)^2}\sqrt{N\sum_{i=1}^N \left(z_i^b\right)^2 - \left(\sum_{i=1}^N z_i^b\right)^2}},$$

or from Spearman's rank order correlation estimate that is more commonly known as Spearman's correlation estimate

$$\bar{\rho} = 1 - \frac{6\sum_{i=1}^N \left(\text{rank}(z_i^a) - \text{rank}(z_i^b)\right)^2}{N(N^2-1)},$$

where the operator $\text{rank}(\cdot)$ returns the observation rank order with respect to its peers. Of the two the Pearson's estimate is more commonly known and used whereas the Spearman's estimate is more robust to errors.

2.3.2 Correlation between two multivariate random variables

Canonical correlation analysis (CCA), which was originally introduced in the early 20th century by Harold Hotelling in (Hotelling 1936), is a multivariate version of the univariate correlation that we learned above. For two random vectors $\boldsymbol{Z}^a$ and $\boldsymbol{Z}^b$, which are of dimensions d^a and d^b respectively, CCA produces $d = \min(d^a, d^b)$ pairs of canonical vectors

$$\boldsymbol{\beta}_i^a, \boldsymbol{\beta}_i^b \text{ for } i = 1, \ldots, d$$

and canonical correlations

$$\lambda_i = \mathbb{C}or[(\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a, (\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b] \text{ for } i = 1, \ldots, d.$$

The interpretation for CCA reads as follows: a canonical correlation coefficient λ_i is interpreted the same way as the linear correlation coefficient with the extension that now it measures the linear dependency between directions $\boldsymbol{\beta}_i^a$ and $\boldsymbol{\beta}_i^b$. The geometrical interpretation is that for realizations $\boldsymbol{z}^a \sim \boldsymbol{Z}^a$, $\boldsymbol{z}^b \sim \boldsymbol{Z}^b$ variation of $\boldsymbol{z}^a$ towards direction $\boldsymbol{\beta}_i^a$ has the strongest correlation with the variation of $\boldsymbol{z}^b$ towards direction $\boldsymbol{\beta}_i^b$.

Formally stated CCA is looking for orthogonal directions $\boldsymbol{\beta}_i^a, \boldsymbol{\beta}_i^b : i = 1, \ldots, d$ of maximal inter-data correlation $\mathbb{C}or[(\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a, (\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b]$. Without loss of generality we can assume expectations $\boldsymbol{\mu}^a = \mathbb{E}[\boldsymbol{Z}^a] = \boldsymbol{0}$ and $\boldsymbol{\mu}^b = \mathbb{E}[\boldsymbol{Z}^b] = \boldsymbol{0}$ yielding covariances $\boldsymbol{\Sigma}^{aa} = \mathbb{V}ar[\boldsymbol{Z}^a] = \mathbb{E}[\boldsymbol{Z}^a(\boldsymbol{Z}^a)^T]$, $\boldsymbol{\Sigma}^{bb} = \mathbb{V}ar[\boldsymbol{Z}^b] = \mathbb{E}[\boldsymbol{Z}^b(\boldsymbol{Z}^b)^T]$, and a joint

covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{aa} & \boldsymbol{\Sigma}^{ab} \\ \boldsymbol{\Sigma}^{ba} & \boldsymbol{\Sigma}^{bb} \end{pmatrix} = \mathbb{E}\left[\begin{pmatrix} \boldsymbol{Z}^a \\ \boldsymbol{Z}^b \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}^a \\ \boldsymbol{Z}^b \end{pmatrix}^T \right].$$

In analytical formulae, the problem is to maximize

$$\begin{aligned} \mathbb{C}or[(\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a, (\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b] &= \frac{\mathbb{E}[((\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b)((\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a)^T]}{\sqrt{\mathbb{E}[((\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b)((\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b)^T]}\sqrt{\mathbb{E}[((\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a)((\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a)^T]}} \\ &= \frac{(\boldsymbol{\beta}_i^b)^T \mathbb{E}[\boldsymbol{Z}^b (\boldsymbol{Z}^a)^T] \boldsymbol{\beta}_i^a}{\sqrt{(\boldsymbol{\beta}_i^b)^T \mathbb{E}[\boldsymbol{Z}^b (\boldsymbol{Z}^b)^T] \boldsymbol{\beta}_i^b}\sqrt{(\boldsymbol{\beta}_i^a)^T \mathbb{E}[\boldsymbol{Z}^a (\boldsymbol{Z}^a)^T] \boldsymbol{\beta}_i^a}} \\ &= \frac{(\boldsymbol{\beta}_i^b)^T \boldsymbol{\Sigma}^{ba} \boldsymbol{\beta}_i^a}{\sqrt{(\boldsymbol{\beta}_i^b)^T \boldsymbol{\Sigma}^{bb} \boldsymbol{\beta}_i^b}\sqrt{(\boldsymbol{\beta}_i^a)^T \boldsymbol{\Sigma}^{aa} \boldsymbol{\beta}_i^a}}, \end{aligned}$$

with constraints

$$\begin{aligned} (\boldsymbol{\beta}_i^a)^T \boldsymbol{\Sigma}^{aa} \boldsymbol{\beta}_i^a &= 1, \\ (\boldsymbol{\beta}_i^a)^T \boldsymbol{\beta}_j^a &= 0 \text{ for all } j < i, \\ (\boldsymbol{\beta}_i^b)^T \boldsymbol{\Sigma}^{bb} \boldsymbol{\beta}_i^b &= 1, \\ (\boldsymbol{\beta}_i^b)^T \boldsymbol{\beta}_j^b &= 0 \text{ for all } j < i. \end{aligned}$$

Solving this maximization problem yields two eigenvalue problems

$$\begin{aligned} (\boldsymbol{\Sigma}^{aa})^{-1} \boldsymbol{\Sigma}^{ab} (\boldsymbol{\Sigma}^{bb})^{-1} \boldsymbol{\Sigma}^{ba} \boldsymbol{\beta}_i^1 &= (\lambda_i)^2 \boldsymbol{\beta}_i^1 \\ (\boldsymbol{\Sigma}^{bb})^{-1} \boldsymbol{\Sigma}^{ba} (\boldsymbol{\Sigma}^{aa})^{-1} \boldsymbol{\Sigma}^{ab} \boldsymbol{\beta}_i^2 &= (\lambda_i)^2 \boldsymbol{\beta}_i^2 \end{aligned}$$

from which it is seen that the ith canonical vectors $\boldsymbol{\beta}_i^1$ and $\boldsymbol{\beta}_i^2$ are equivalent to the eigenvectors that correspond to the ith largest eigenvalues of matrixes $(\boldsymbol{\Sigma}^{aa})^{-1} \boldsymbol{\Sigma}^{ab} (\boldsymbol{\Sigma}^{bb})^{-1} \boldsymbol{\Sigma}^{ba}$ and $(\boldsymbol{\Sigma}^{bb})^{-1} \boldsymbol{\Sigma}^{ba} (\boldsymbol{\Sigma}^{aa})^{-1} \boldsymbol{\Sigma}^{ab}$ respectively. Similarly, canonical correlation λ_i is equivalent to the square roots of the the ith largest eigenvalues of both of these matrixes. Because these canonical correlations are squared in the eigenvalue problems, solving these problems may offer canonical vector pairs between which correlation is negative. Hence, it is necessary to postprocess such pairs by replacing $\boldsymbol{\beta}_i^b$ with $-\boldsymbol{\beta}_i^b$ ($\boldsymbol{\beta}_i^b \leftarrow -\boldsymbol{\beta}_i^b$), which yields a requested solution with all canonical correlations being positive. Analytically this is justified because $\mathbb{C}or[(\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a, (\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b] = -\mathbb{C}or[(\boldsymbol{\beta}_i^a)^T \boldsymbol{Z}^a, (-\boldsymbol{\beta}_i^b)^T \boldsymbol{Z}^b]$. From the problem formulation it can be seen that CCA is invariant with respect to all affine transformations of $\boldsymbol{Z}^a$ and $\boldsymbol{Z}^b$. In recent years it has been shown (Kay 1992) that performing CCA on two data sets that come from elliptically symmetric distributions is equivalent to maximizing mutual information between these two data sets.

For a sample of observed vectors $\boldsymbol{z}_1^a, \ldots, \boldsymbol{z}_N^a \sim \boldsymbol{Z}^a$ and $\boldsymbol{z}_1^b, \ldots, \boldsymbol{z}_N^b \sim \boldsymbol{Z}^b$, empirical estimates $\hat{\boldsymbol{\beta}}_i^a, \hat{\boldsymbol{\beta}}_i^b, \hat{\lambda}_i : i = 1, \ldots, d$ are obtained from postprocessed eigenvectors and square roots of eigenvalues of matrixes $(\hat{\boldsymbol{\Sigma}}^{aa})^{-1} \hat{\boldsymbol{\Sigma}}^{ab} (\hat{\boldsymbol{\Sigma}}^{bb})^{-1} \hat{\boldsymbol{\Sigma}}^{ba}$ and

$(\hat{\mathbf{\Sigma}}^{bb})^{-1}\hat{\mathbf{\Sigma}}^{ba}(\hat{\mathbf{\Sigma}}^{aa})^{-1}\hat{\mathbf{\Sigma}}^{ab}$, where

$$\hat{\mathbf{\Sigma}} = \begin{pmatrix} \hat{\mathbf{\Sigma}}^{aa} & \hat{\mathbf{\Sigma}}^{ab} \\ \hat{\mathbf{\Sigma}}^{ba} & \hat{\mathbf{\Sigma}}^{bb} \end{pmatrix} = \frac{1}{N-1} \begin{pmatrix} (\boldsymbol{z}_1^a)^T & (\boldsymbol{z}_1^b)^T \\ \vdots & \vdots \\ (\boldsymbol{z}_N^a)^T & (\boldsymbol{z}_N^b)^T \end{pmatrix}^T \begin{pmatrix} (\boldsymbol{z}_1^a)^T & (\boldsymbol{z}_1^b)^T \\ \vdots & \vdots \\ (\boldsymbol{z}_N^a)^T & (\boldsymbol{z}_N^b)^T \end{pmatrix}.$$

In the case of real world data, use of robust covariance estimators is strongly recommended. For an introduction to such see (Huber 2003), (Rousseeuw and Leroy 2003), and a recent study in (Taskinen 2003).

One common application of projection methods, in which CCA also belongs to, is to find an alternative representation through a base change where the original axes vectors for $\boldsymbol{Z}^a$ and $\boldsymbol{Z}^b$ are replaced with vectors $\boldsymbol{\beta}_i^a$ and $\boldsymbol{\beta}_i^b$ respectively. Geometrically this is equivalent to rotation so that the highest correlating directions become the main axes, which can make graphical plots easier to interpret. A closer inspection of such plots shows how an observed data set is distributed. The axes that correspond to the largest correlation coefficients are the most important ones and typically the last axes are used to explain uncorrelated noise. Selecting the subspace spanned by d^*, where $d^* < d$, largest axes will explain

$$\frac{\sum_{i=1}^{d^*} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

percent of the data variation with $d^* : d$ reduction in observed vector dimensions. This reduction is often substantial, which can mean computational and storage advantages. This technique is commonly used for noise removal as the noise can be assumed to be uncorrelated with respect to studied phenomena and the variance of the noise is assumed to be smaller than the d^* largest eigenvalue.

Using CCA for image analysis is a relatively new concept. The idea has been discussed and some implementations do exist prior to our work. During the last ten years, Swedish research groups led by Magnus Borga and Hans Knutsson have publicly philosophized about the possibility of using CCA to generate feature extracting visual operators based on empirical training sets. For an overview of their work the reader is referred to browse (Borga, Knuttson and Landenius 1997), (Knuttson and Borga 1999), (Knuttson, Andersson, Borga and Wiklund 2000), and (Borga and Knuttson 2001). They have also implemented some of their ideas, mostly in the field of biomedical imaging, but their applications are not similar to the one presented in this thesis nor does their methodology answer the type of questions that are posed in our application.

2.3.3 Correlation within and between random fields

In spatial statistics, especially in geostatistics, there is an established methodology for analyzing second-order statistical dependencies within and between random fields. For an introduction to geostatistics the reader is referred to browse (Cressie 1993), (Wackernagel 2003), (Lantuéjoul 2002),

(Banerjee, Bradley and Gelfand 2004), or (Chilès and Delfiner 1999). In the following we have an introduction to spatial stochastic processes that are known as Gaussian random fields (GRFs). We learn tools for process characteristic estimation and interpretation, discuss about parametric modeling of real world images, and comment on how the obtained models can be used to simulate new images that exhibit the same process characteristics.

A two-dimensional random field is said to be second-order stationary if all random variables $Z(x, y) : (x, y) \in D$ have a finite expected value

$$\mathbb{E}[Z(x, y)] = \mu \text{ for all } (x, y) \in D,$$

and if the covariance

$$\begin{aligned}
&\mathbb{C}ov[Z(x + t_x, y + t_y), Z(x, y)] \\
&= \mathbb{E}[Z(x + t_x, y + t_y) - \mathbb{E}[Z(x + t_x, y + t_y)]]\mathbb{E}[Z(x, y) - \mathbb{E}[Z(x, y)]] \\
&= C(t_x, t_y)
\end{aligned}$$

between all pairs of random variables $Z(x + t_x, y + t_y)$ and $Z(x, y)$ with $(x + t_x, y + t_y), (x, y) \in D$ is finite and depends only on the spatial transition (t_x, t_y). The function $C(t_x, t_y)$ is known as covariogram and is also referred as covariance function and autocovariance function. A mathematically proper covariogram satisfies these four conditions (Lantuéjoul 2002, p. 24):

1. $C(0, 0) \geq 0$.
2. $C(-t_x, -t_y) = C(t_x, t_y)$.
3. The integral of C over the function support is non-negative.
4. Function C is positive definite. That is for any finite sequence of points $(x_i, y_i)_{i=1}^N$ and for any finite sequence of real numbers $(\lambda_i)_{i=1}^N$ the equation

$$\sum_{i=1,j=1}^{N} \lambda_i \lambda_j C(x_i - x_j, y_i - y_j) \geq 0$$

holds.

A derivative measure that is known as correlogram or autocorrelation function is defined as

$$\rho(t_x, t_y) = \frac{C(t_x, t_y)}{C(0, 0)}$$

provided that $C(0, 0) > 0$. This correlation measure can be considered as a spatial extension to the other correlation measures that were previously presented in this section. In addition, a stationary random field is said to be ergodic if

$$\lim_{||(t_x, t_y)|| \to \infty} C(t_x, t_y) = 0,$$

that is the random variables that are spatially far away from each other are also uncorrelated. With these two assumptions a random field $Z = \{Z(x, y) : (x, y) \in D\}$ is said to be Gaussian if any linear combination of random variables $Z(x, y)$ for $(x, y) \in D$ follows a Gaussian distribution.

Gaussian random fields are important from both theoretical and practical perspective. First, with the relatively strong assumptions of second-order stationarity and ergodicity make analytical calculations easier, which allows rigorous formalism to be used in the description of the problem. This leads to better formulation and reduces the need for ad hoc solutions. On the other hand the central limit theorem suggest that summing large quantities of equally distributed random fields, in which the distribution may be non-Gaussian, together results in a cumulative random field that is approximately Gaussian. This means that complex real world phenomena, which are usually sums of simpler co-factors, tend to be Gaussian.

The most commonly used measure for spatial dependency within $(a = b)$ and between $(a \neq b)$ random fields $Z^a = \{Z^a(x, y) : (x, y) \in D\}$ and $Z^b = \{Z^b(x, y) : (x, y) \in D\}$ is a semivariogram

$$\gamma^{a,b}(t_x, t_y) = \frac{1}{2}\mathbb{V}ar[Z^a(x + t_x, y + t_y) - Z^b(x, y)] \text{ where } (x + t_x, y + t_y), (x, y) \in D,$$

which measures the dissimilarity between pixels $(x + t_x, y + t_y)$ and (x, y). In literature the form $a \neq b$ is often called as a cross-semivariogram but in this thesis use the name semivariogram for both forms. An empirical estimate for the semivariogram is obtained from Matheron's semivariogram estimate that is also known as the classical semivariogram estimate

$$\hat{\gamma}^{a,b}(t_x, t_y) = \frac{1}{2}\frac{\sum_{x,y:(x+t_x,y+t_y),(x,y)\in D}(z^a(x + t_x, y + t_y) - z^b(x, y))^2}{\sum_{x,y:(x+t_x,y+t_y),(x,y)\in D} 1}.$$

An another empirical estimate is Hawkins and Cressie's semivariogram estimate

$$\bar{\gamma}^{a,b}(t_x, t_y) = \frac{\sum_{x,y:(x+t_x,y+t_y),(x,y)\in D} \sqrt{|z^a(x + t_x, y + t_y) - z^b(x, y)|}}{0.914(\sum_{x,y:(x+t_x,y+t_y),(x,y)\in D}) + 0.988}.$$

Of these two the Matheron's semivariogram estimate is more commonly known and used whereas the Hawkins and Cressie's semivariogram estimate is more robust to errors in pixel intensities. In the case of second-order stationary and ergodic GRFs the semivariogram γ and the covariogram C are related through equation

$$\gamma^{a,b}(t_x, t_y) = C^{a,b}(0, 0) - C^{a,b}(t_x, t_y). \tag{6}$$

The main difference between the two is that they have different estimator properties. Whereas the semivariogram contains approximately the same information as the covariogram, the empirical estimator for the semivariogram is unbiased with less demanding assumptions and is thus preferred in data analysis.

The most common way to study the second-order statistics of an image is to plot an estimated semivariogram for visual observations; for an example of such

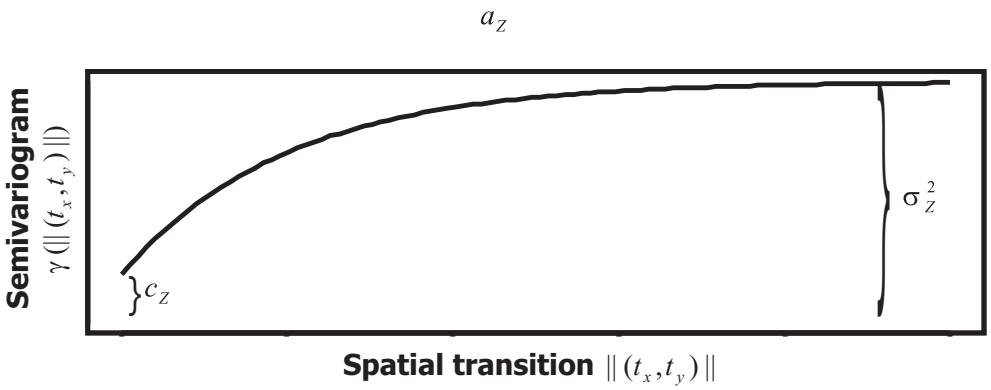


FIGURE 7: An example semivariogram with model parameters $\theta_Z = (c_Z, \sigma_Z^2, a_Z)$ where c_Z measures process nugget effect, σ_Z^2 measures process sill, and a_Z measures process range respectively.

see Figure 7. There are a few aspects of the semivariogram that an experienced data analyst notices when building an understanding of the generating process. Microscale variance, commonly referred as **nugget effect** and formally defined as

$$c_Z = \lim_{||(t_x,t_y)|| \to 0} \gamma(t_x, t_y),$$

measures the amount of dissimilarity within small transitions (t_x, t_y). From mathematical point of view all GRFs have $c_Z = 0$ but in practice real world images contain measurement errors and noise which result in $c_Z > 0$. The process variance, more commonly referred as **sill** and formally defined as

$$\sigma_Z^2 = \lim_{||(t_x,t_y)|| \to \infty} \gamma(t_x, t_y),$$

measures the amount of variation in random variables $Z(x, y)$. The process **range**, defined as the smallest value of $a_Z = ||(t_x, t_y)||$ for which

$$\gamma((1+\epsilon)(t_x, t_y)) = \sigma_Z^2 \text{ for any } \epsilon > 0,$$

measures the effective range after which there is no spatial correlation between random variables $Z(x + t_x, y + t_y)$ and $Z(x, y)$ when $(x + t_x, y + t_y), (x, y) \in D$. For a more detailed discussion on how to interpret the semivariogram, see (Cressie 1993, pp. 58-68 and pp. 127-135).

An alert reader might wonder if it is possible to compare two random fields in which pixel intensities may have entirely different units of measurement. This problem is discussed in (Cressie and Wikle 1998). In this thesis we assume that all the studied images are preprocessed through equation

$$z'(x, y) = \frac{z(x, y) - \hat{\mu}_Z}{\hat{\sigma}_Z} \text{ for } (x, y) \in D$$

to have $\mathbb{E}[Z(x, y)] = 0$ and $\mathbb{V}ar[Z(x, y)] = 1$, which effectively remedies the problem by making the obtained results easily comparable.

Model	Covariance function	Conditions
Spherical	$\left(1 - \frac{3\|\|(t_x,t_y)\|\|}{2s} + \frac{1}{2}\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)^3\right)1_{\frac{\|\|(t_x,t_y)\|\|}{s}\leq 1}$	$d \leq 3$
Exponential	$e^{-\frac{\|\|(t_x,t_y)\|\|}{s}}$	-
Stable	$e^{-(\frac{\|\|(t_x,t_y)\|\|}{s})^\alpha}$	$0 < \alpha \leq 2$
Hyperbolic	$\frac{1}{1+\frac{\|\|(t_x,t_y)\|\|}{s}}$	-
Gaussian	$e^{-(\frac{\|\|(t_x,t_y)\|\|}{s})^2}$	-
Cardinal size	$\frac{\sin\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)}{\frac{\|\|(t_x,t_y)\|\|}{s}}$	$d \leq 3$
J-Bessel	$2^\mu\Gamma(\mu+1)\frac{J_\mu\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)}{\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)^\mu}$	$\mu \geq \frac{d}{2} - 1$
K-Bessel	$\frac{\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)^\mu}{2^{\mu-1}\Gamma(\mu)}K_{-\mu}\left(\frac{\|\|(t_x,t_y)\|\|}{s}\right)$	$\mu > 0$

TABLE 2: Covariograms for theoretical parametric models of Gaussian random fields (Lantuéjoul 2002, p. 187 and p. 242). Γ, J_μ and $K_{-\mu}$ are Gamma, J-Bessel and K-Bessel functions respectively and s is a spatial scaling parameter, which is not to be confused with the process range a_Z.

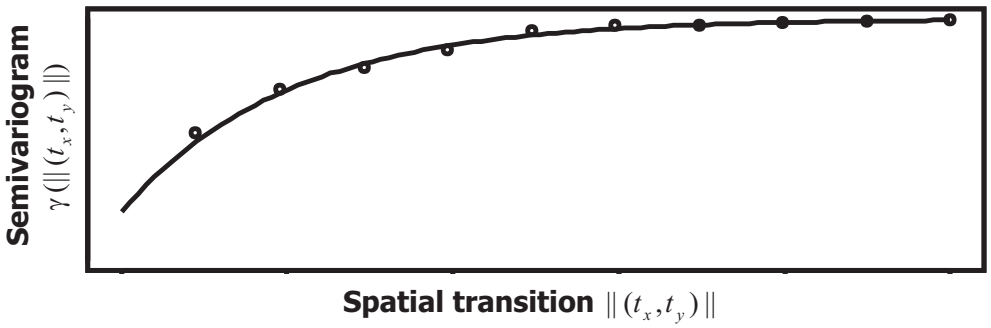


FIGURE 8: An example of parametric model fitting for Gaussian random fields.

In practice the covariograms that are estimated through empirical semivariograms are almost never mathematically proper covariograms as they do not satisfy the previously presented four conditions. The solution is to model the process by fitting a proper covariogram to the obtained empirical estimate. Some commonly known covariograms are listed in Table 2. Consider that we have obtained estimates $\hat{\gamma}(t_{x,1}, t_{y,1}), \ldots, \hat{\gamma}(t_{x,N}, t_{y,N})$ for transitions $(t_{x,1}, t_{y,1}), \ldots, (t_{x,N}, t_{y,N})$; examples of such are presented in Figure 8. One way to fit a parametric semivariogram model that has gained wide acceptance is to estimate model parameters

$\theta_Z = (c_Z, \sigma_Z^2, a_Z)$ through least squares minimization

$$\hat{\theta}_Z = \arg\min_{\theta_Z} \sum_{i=1}^{N} (\gamma(t_{x,i}, t_{y,i}|\theta_Z) - \hat{\gamma}(t_{x,i}, t_{y,i}))^2,$$

where $\gamma(t_{x,i}, t_{y,i}|\theta_Z)$ is a parametric model obtained from Table 2 through equation 6 with parameters θ_Z. For an analysis of this and other parametric model fitting techniques, see (Cressie 1993, pp. 90-104).

One obvious benefit of the process modeling is that the obtained models can be used for simulating new realizations that exhibit the same second-order process characteristics that the real world image has. A visual comparison between the real world image and such simulations gives a hands-on understanding of how well the identified parametric model is able to describe the present real world phenomena. There are about a dozen simulation algorithms of which the two preferred are the spectral method and the turning bands method. These two are probably the best known, the most widely used, and they generate computational loads that are very competitive with respect to the other known algorithms. Due to the constrained number of pages usable for this introduction we are unable to present the details of these simulation algorithms but instead advice the reader to browse the previously named references.

2.4 A survey to existing research in paper science

In this section we review existing research in the field of paper science. There are about a dozen good books covering the structure of paper. For a detailed introduction to the subject the reader is advised to browse for example (Deng and Dodson 1994), (Niskanen 1998), (Mark, Habeger, Borch and Lyne 2002), and (Borch, Lyne, Mark and Habeger 2002). In the following we are going to review some recently published studies. First we take a look at the stochastic structure of paper. Then we introduce two studies concerning fiber orientation, which is an important aspect of paper structure. This is followed by a review of two parametric modeling techniques that can be used to model and to simulate paper structure. The emphasis is on presenting general ideas whereas most of technical details are omitted for clarity. These details are available from the original publications. The reviewed studies cover only a small fraction of all the research that is conducted the field of paper science but they do outline the current state of the art in their respective application fields.

2.4.1 On the stochastic structure of paper

A study presented in (Dodson, Oba and Sampson 2001a) and continued in (Dodson, Oba and Sampson 2001b) investigates the relations between four paper properties: formation Z^m, thickness Z^t, density Z^d, and porosity Z^p. Four random

fields are derived for local formation $\tilde{Z}^m$, local thickness $\tilde{Z}^t$, local density $\tilde{Z}^d$, and local porosity $\tilde{Z}^p$ according to

$$\tilde{Z}^i = \left\{ \tilde{Z}^i(x,y) : \tilde{Z}^i(x,y) = \frac{\sum_{(s_x,s_y)\in D_{local}} Z^i(x+s_x, y+s_y)}{w_{local} h_{local}}, (x,y) \in D_{sample} \right\}$$

where $i \in \{m,t,d,p\}$. Similarly four random variables are derived for global formation $\bar{Z}^m$, global thickness $\bar{Z}^t$, global density $\bar{Z}^d$, and global porosity $\bar{Z}^p$ according to

$$\bar{Z}^i = \frac{\sum_{(s_x,s_y)\in D} Z^i(s_x, s_y)}{w_{image} h_{image}}.$$

From the physical model the four random fields are known to be linked as

$$\begin{aligned} \tilde{Z}^m &= \tilde{Z}^t \tilde{Z}^d, \\ \tilde{Z}^d &= (1 - \tilde{Z}^p)\delta, \text{ and} \\ \tilde{Z}^p &= 1 - \frac{\tilde{Z}^m}{\tilde{Z}^t \delta}, \end{aligned}$$

where variable δ is used to represent fiber density. Similarly the four random variables are linked as

$$\begin{aligned} \bar{Z}^m &= \bar{Z}^t \bar{Z}^d, \\ \bar{Z}^d &= (1 - \bar{Z}^p)\delta, \text{ and} \\ \bar{Z}^p &= 1 - \frac{\bar{Z}^m}{\bar{Z}^t \delta}. \end{aligned}$$

The first paper takes an empirical approach to determine the relation between the variation of local thickness and the variation of local density with respect to the variation of local formation. The amount of variation is measured with variance $\mathbb{V}ar[\tilde{Z}]$ and coefficient of variation $\mathbb{CV}[\tilde{Z}] = \sqrt{\mathbb{V}ar[\tilde{Z}]} \Big/ \mathbb{E}[\tilde{Z}]$. In the case of thickness, $\mathbb{C}or[\mathbb{V}ar[\tilde{Z}^t(x,y)], \mathbb{V}ar[\tilde{Z}^m(x,y)]]$ and $\mathbb{C}or[\mathbb{CV}[\tilde{Z}^t(x,y)], \mathbb{CV}[\tilde{Z}^m(x,y)]]$ over D_{sample}, the empirical study shows positive linear correlations for both measures of variation. The former correlation seems to depend on pulp type but on the other hand seems independent of sheet grammage, whereas the latter correlation seems independent of the pulp type but dependent on the sheet grammage. For a similar case of density, $\mathbb{C}or[\mathbb{V}ar[\tilde{Z}^d(x,y)], \mathbb{V}ar[\tilde{Z}^m(x,y)]]$ and $\mathbb{C}or[\mathbb{CV}[\tilde{Z}^d(x,y)], \mathbb{CV}[\tilde{Z}^m(x,y)]]$ over D_{sample}, the study shows positive linear correlations again for both measures of variation. Interestingly, there seemed to be a negative correlation between variations of thickness and density when measured with the coefficient of variation, $\mathbb{C}or[\mathbb{CV}[\tilde{Z}^t(x,y)], \mathbb{CV}[\tilde{Z}^d(x,y)]]$ over D_{sample}, whereas no correlation was found when measured with the variance $\mathbb{C}or[\mathbb{V}ar[\tilde{Z}^t(x,y)], \mathbb{V}ar[\tilde{Z}^d(x,y)]]$ over D_{sample}. In the former case the correlations seem to depend on the pulp type but on the same independent of the sheet grammage.

The second paper takes an analytic approach to modeling the dependencies between local variations based on a prior work presented in

(Dodson and Sampson 1999). In this study the processes Z^m, Z^t, Z^d, and Z^p are considered to come from planar projections of a near-planar stochastic fiber network; for an additional reading on such see (Stoyan, Kendall and Mecke 1995). The first presented equation

$$\begin{aligned}\mathbb{V}ar[\tilde{Z}^p(x,y)] &= \frac{1}{\delta^2}\left(\frac{\bar{Z}^m}{\bar{Z}^t}\right)^2\left(\frac{\mathbb{V}ar[\tilde{Z}^m(x,y)]}{(\bar{Z}^m)^2} - \frac{2\mathbb{C}ov[\tilde{Z}^m(x,y),\tilde{Z}^t(x,y)]}{\bar{Z}^m\bar{Z}^t} + \frac{\mathbb{V}ar[\tilde{Z}^t(x,y)]}{(\bar{Z}^t)^2}\right) \\ &\approx \frac{1}{\delta^2}\left(\frac{\bar{Z}^m}{\bar{Z}^t}\right)^2 |\mathbb{CV}[\tilde{Z}^m(x,y)]^2 - \mathbb{CV}[\tilde{Z}^t(x,y)]^2|\end{aligned}$$

explains the variance of local porosity with respect to local and global variation of formation and thickness. The approximate equation holds when $\bar{Z}^p \approx \tilde{Z}^p(x,y)$ for $(x,y) \in D_{sample}$. Similarly the second presented equation

$$\begin{aligned}\mathbb{V}ar[\tilde{Z}^d] &= \left(\frac{\bar{Z}^m}{\bar{Z}^t}\right)^2\left(\frac{\mathbb{V}ar[\tilde{Z}^m(x,y)]}{(\bar{Z}^m)^2} - \frac{2\mathbb{C}ov[\tilde{Z}^m(x,y),\tilde{Z}^t(x,y)]}{\bar{Z}^m\bar{Z}^t} + \frac{\mathbb{V}ar[\tilde{Z}^t(x,y)]}{(\bar{Z}^t)^2}\right) \\ &\approx \left(\frac{\bar{Z}^m}{\bar{Z}^t}\right)^2 |\mathbb{CV}[\tilde{Z}^m(x,y)]^2 - \mathbb{CV}[\tilde{Z}^t(x,y)]^2| = \delta^2\mathbb{V}ar[\tilde{Z}^p(x,y)]\end{aligned}$$

explains the variance of local density with respect to local and global variation of formation and thickness. The approximate equation holds when $\bar{Z}^d \approx \tilde{Z}^d(x,y)$ for $(x,y) \in D_{sample}$. Based on analytical and empirical experimentation, it is suggested that the relationship between averages of local density $\tilde{Z}^t(x,y)$ and local thickness $\tilde{Z}^d(x,y)$ over D_{sample} is approximately linear and is well described by the bivariate normal distribution.

2.4.2 On estimation and modeling of fiber orientation distributions

A gradient-based method, which was originally introduced in (Erkkilä 1995) and later discussed in (Erkkilä, Pakarinen and Odell 1998), estimates a non-parametric fiber orientation distribution from a grayscale image $z = \{z(x,y) : (x,y) \in D\}$. First the gradients of the image are observed as

$$\nabla z(x,y) = \left(\frac{\partial z(x,y)}{\partial x}, \frac{\partial z(x,y)}{\partial y}\right) \text{ for all } (x,y) \in D.$$

In the study these gradients are obtained through convolving the image with 5×5 pixels horizontal and vertical gradient detector masks. The founding idea of this method is that the fiber orientation angle $\alpha(x,y)$ at a given pixel coordinate (x,y) is assumed to be perpendicular to the gradient $\nabla z(x,y)$. The magnitude of the gradient $\nabla z(x,y)$ is used as an observation weight for the observed angle. An empirical estimate for the probability that a fiber has angle $\alpha \in \triangle\alpha \subset [0, 2\pi[$ is then obtained as

$$\hat{\mathbb{P}}(\triangle\alpha) = \frac{\sum_{(x,y)\in D, \alpha(x,y)\in\triangle\alpha} |\nabla z(x,y)|}{\sum_{(x,y)\in D} |\nabla z(x,y)|}.$$

If necessary, this probability measure can then be used to fit a parametric model based on empirical data. Since its publication this method has gained acceptance in the industry and has become almost the de facto approach for fiber orientation distribution estimation.

A variogram-based method, which was originally introduced in (Kärkkäinen, Penttinen, Ushakov and Ushakova 2001) and (Kärkkäinen and Jensen 2001), and later revisited in (Kärkkäinen 2003), models parametric fiber orientation distribution in a random field $Z = \{Z(x, y) : (x, y) \in \Omega \subset \mathbb{R}^2\}$ based on the expected number of crossings $P_L(\beta)$ between fibers and a transect line L_β per unit area. Here β identifies the angle of the transect line with respect to the x-axis. Whereas it may be impossible to separate individual fibers in the image, all the necessary information is shown to be present in a scaled variogram

$$V_L(d, \beta) = \frac{\mathbb{E}[|Z(x_1, y_1) - Z(x_2, y_2)|]}{d} \propto P_L(\beta)$$

in which $(x_1, y_1), (x_2, y_2) \in \Omega$ and $d = ||(x_1 - x_2, y_1 - y_2)||$. From stereological theory it is suggested that

$$P_L(\beta) = L_A \int_0^\pi |\sin(\alpha - \beta)| f(\alpha) d\alpha$$

where L_A is the mean fiber length per unit area and $f(\alpha)$ is the fiber angle probability density function. For a parametric fiber orientation distribution the study suggests elliptic density distribution function

$$f(\alpha|\theta = (\tau, \kappa)) = \frac{c}{\sqrt{1 - (\kappa \cos(\alpha - \tau))^2}} \text{ for } 0 \leq \alpha < \pi$$

in which c is a normalizing constant, $\tau \in [0, \pi)$ is the angle of the most common fibers, and $\kappa = \sqrt{1 - (b/a)^2}$ where a and b are the radii of the major and minor axes of the ellipse. This seems like a reasonable choice as there are a lot of studies suggesting that paper fibers tend to have elliptic orientation distribution. The variogram-based method has several advantages over the gradient-based method but for the time being the former has not been able to replace the latter in the industry.

2.4.3 Parametric modeling of paper structure

A study presented in (Johansson 2002) proposes three approaches for parametric modeling of paper structure: a hierarchical shot-noise model, a fractal model version of the previous shot-noise model, and a hierarchical Gibbs-Markov random field model. Of these three the first one is briefly presented.

The shot-noise model, which is a special case of Poisson germ-grain model in which germs correspond to spatial locations $\{(x_i, y_i)\}_{i=1}^N$ that are randomly placed in domain $\Omega \subset \mathbb{R}^2$ according to Poisson distribution and in which grains correspond to geometric objects $\{\Xi_i\}_{i=1}^N$ defining spatial structures that are placed at these

coordinates, constructs a random field

$$Z = \left\{ Z(x,y) : Z(x,y) = \sum_{i=1}^{N} \Psi((x,y);\Xi_i) \text{ for } (x,y) \in D \right\}$$

by summing up the cumulative effects $\Psi((x,y);\Xi_i)$ of grains Ξ_i at position (x,y). For such a random field, the expected value is

$$\mu_Z = \lambda|\Omega|\mathbb{E}[\Psi((x,y);\Xi_i)],$$

the covariogram is

$$C(t_x,t_y) = \lambda|\Omega|\mathbb{E}[\Psi((x,y);\Xi_i)\Psi((x+t_x,y+t_y);\Xi_i)],$$

and the semivariogram is

$$\gamma_Z(t_x,t_y) = \lambda|\Omega|\mathbb{E}[\Psi((x,y);\Xi_i)(\Psi((x,y);\Xi_i) - \Psi((x+t_x,y+t_y);\Xi_i))]$$

where λ is the expected number of germs per unit area of domain Ω, and $|\Omega|$ is the area of domain Ω. Hence, the total number of gains N is a random variable that is defined as $N \sim Poisson(\lambda|\Omega|)$. The founding idea of the study is to model the paper structure as a random field Z that is a sum of two independent and stationary shot-noise models: a large scale pulp clustering model Z^K and a micro-scale fiber distribution model Z^F, which conformally can be expressed as

$$Z(x,y) = Z^K(x,y) + Z^F(x,y).$$

The semivariogram for this random field has form

$$\gamma_Z(t_x,t_y) = \gamma_Z^K(t_x,t_y) + \gamma_Z^F(t_x,t_y).$$

The cluster model Z^K is used to model the flocculation of pulp due to the mechanism on which the pulp mass is sprayed on the conveyer felt in a Fourdrinier paper machine, which is the most common machine design in large paper mills. The cluster model consists of $N^K \sim Poisson(\lambda^K|\Omega|)$ round disks that are placed at coordinates $\{(x_j^K, y_j^K)\}_{j=1}^{N^K}$, have height δ^K, and radius r. For such a random field the semivariogram is

$$\gamma_Z^K(t_x,t_y) = \begin{cases} (\delta^K)^2\lambda^K\Big(\pi r^2 - \Big(2r^2\arccos\Big(\frac{||(t_x,t_y)||}{2r}\Big) \\ \qquad -||(t_x,t_y)||\sqrt{r^2 - \frac{||(t_x,t_y)||^2}{4}}\Big)\Big), & \text{for } ||(t_x,t_y)|| \le 2r \\ (\delta^K)^2\lambda^K\pi r^2, & \text{for } ||(t_x,t_y)|| > 2r. \end{cases}$$

The fiber model Z^F is used to model the micro-scale variation that results from the stochastic arrangement of individual fibers on the conveyer felt. The fiber model consists of $N^F \sim Poisson(\lambda^F|\Omega|)$ rectangles that are placed at coordinates

$\{(x_k^F, y_k^F)\}_{k=1}^{N^F}$, have height δ^F, length l, width w, and directions $\{(v_{x,k}, v_{y,k})\}_{k=1}^{N^F}$. The semivariogram of the model is

$$\gamma_Z^F(t_x, t_y) = (\delta^F)^2 \lambda^F \Big(lw - \int_{-\pi/2}^{\pi/2} (l - ||(t_x, t_y)|| \cos(\alpha - \arctan(t_y/t_x)))_+ \\ (w - ||(t_x, t_y)|| \sin(\alpha - \arctan(t_y/t_x)))_+ f(\alpha) d\alpha \Big),$$

where operator $(\cdot)_+$ returns the positive part of the parameter and $f(\alpha)$ is the probability density function for fiber angle $\alpha = \arctan(v_y/v_x)$.

These two presented models supplement each other. This is because it is computationally heavy to model large paper areas using individual fibers. In addition, the flocculation of pulp mass and the stochastic arrangement of individual fibers happen in completely different resolutions and hence are seen to be relatively independent from each other. Modeling the two independently allows a computationally feasible way to obtain large areas with near to realistic micro-scale structure.

An another study presented in (Brown, Diggle and Henderson 2003) derives a hierarchical non-Gaussian model with the objective of identifying micro-scale variation from stochastic fiber placement, medium scale variation from fiber flocculation, and large variation due to change in manufacturing conditions. The proposed model has flexible covariance structure, the model parameters have physical interpretations, and the number of hierarchical layers used for different scales is not limited to any fixed number. The estimation of model parameters is performed on a one-dimensional transect, which is a considerable advantage in many real world applications.

A hierarchical non-Gaussian model Z on a domain $\Omega \subset \mathbb{R}^2$ is built as a sum of N independent component random fields $Z^i : i = 1, \ldots, N$ as

$$Z = \left\{ Z(x, y) \middle| Z(x, y) = \sum_{i=0}^{N} Z^i(x, y) \text{ for } (x, y) \in \Omega \right\}.$$

The first component $Z^0(x, y)$ is used to model micro-scale variation due to individual fibers with a white noise process that has mean α and variance τ^2. The other components are used to model pulp flocculation at different scales of variation. These components are formally defined as

$$Z^i(x, y) = \beta^i \sum_{j=1}^{N^i} f^i(x - x_j, y - y_j) \text{ where } i = 1, \ldots, N$$

where β^i set the maximum heights of the flocks, $f^i(\cdot, \cdot)$ define the shape of the flocks as bivariate probability density functions, and $(x_{i,j}, y_{i,j}) : j = 1, \ldots, N^i$ are flock positions from homogeneous Poisson point processes with intensities λ^i. The study suggest the use of Matérn functions

$$f^i(x, y) = \frac{\nu^i 2^{1-\nu^i}}{\pi(\sigma^i)^2 \Gamma(\nu^i + 1)} \left(\frac{||(x, y)||}{\sigma^i / 2\sqrt{\nu^i}} \right)^{\nu^i} K_{\nu^i} \left(\frac{||(x, y)||}{\sigma^i / 2\sqrt{\nu^i}} \right) : i = 1, \ldots, N$$

in which $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν^i and σ^i determines the flock scale. Setting $\nu^i \to \infty$ yields a Gaussian probability distribution function, whereas smaller values of ν^i result in rougher surfaces.

For the random field Z a one-dimensional (v_x, v_y) directional transect is formally defined as

$$Z^\dagger = \{Z^\dagger(w) | Z^\dagger(w) = Z(x_0 + wv_x, y_0 + wv_y)\}$$

where (x_0, y_0) is some arbitrary point in Ω and $||(v_x, v_y)|| = 1$. The spectrum of such a transect is obtained as

$$h(\omega) = \int_{-\infty}^{\infty} e^{-2\pi\sqrt{-1}\omega t} C_{Z^\dagger}(t)dt = 2\int_{-\infty}^{\infty} \cos(2\pi\omega t) C_{Z^\dagger}(t)dt,$$

where $C_{Z^\dagger}(t) = \mathbb{C}ov[Z^\dagger(w), Z^\dagger(w+t)]$. Due to the independence and the stationarity of the component random fields Z^i, the spectrum of $Z^\dagger$ is the sum of the component spectra, that is

$$h(\omega) = \sum_{i=0}^{N} h^i(\omega).$$

The derived hierarchical model has a flexible covariance structure and the model parameters can be estimated fast, which allows a paper manufacturing process to be monitored online. A longer term goal is to understand how the spectrum changes with respect to the underlying process. As for the time being, there are no published empirical studies concerning the actual use of the presented hierarchical model and hence the full evaluation of the model is unfinished.

2.5 Discussion

In this chapter we have reviewed techniques from image analysis and analysis of correlations, which are needed in our work, along with recently published studies from paper science, which give an overview of the current state-of-the-art in this field. An existing methodology that would enable us to make spatial comparisons between different measuring techniques and different paper sheet, recall section 1.1, in the presence of technical challenges, recall section 1.4, that are typical in paper applications was not found from the literature. Applying classical canonical correlation, as described in 2.3.2, does not yield spatial understanding of the dependency whereas correlation measures between random fields, as presented in 2.3.3, do but in a limited manner because they are susceptible lacking pixel correspondence between the measured random fields. In addition neither of these methodologies is able to yield spatial descriptions of the phenomena between which the dependencies are being measured. Hence a methodology that would be able to identify and measure dependencies between spatial patterns, which depending on the application can have clear physical interpretations, would be of more value in practical applications. Inspired by this challenge a new correlation measure that aims to fulfil these needs is proposed in the next chapter.

3 PROPOSED METHODOLOGY FOR SPATIAL DEPENDENCY ANALYSIS

In this chapter we propose a methodology for detecting and measuring spatial dependencies between two images that are generated by a random process. Our images, which where introduced in chapter 1, are two dimensional measurements of five paper properties: foreground and background topographies, mass distribution, thickness distribution, and density distribution. Examples of mass distribution and foreground topography over the same paper area are shown in Figure 9. The two open questions that are studied in this chapter are: 1) how to detect and measure spatial dependency between two images and 2) how to quantify the proportional importances of spatial dependencies from different pairs of images. For this purpose, we assume that there are observable dependencies between the images. Furthermore, we assume that these dependencies can be characterized through pairs of features that are obtained from local image areas. While there are many incidental problems with the measured images, such as spatial distortions and measurement noise, these are disregarded in this chapter as they are considered a separate issue of data preprocessing.

The key benefits of the proposed methodology are:

- Compress the dependency information into a single easily interpretable scalar value.
- Explain the observed dependency in terms of the spatial structures of the studied images.
- Compare the amount of dependency in different spatial scales.
- Compare the dependencies between multiple image pairs.

This chapter is organized as follows. Section 3.1 gives an overview of the methodology. Section 3.2 considers different aspects of observing the dependency structure between two images and discusses how these aspects affect the developed methodology. Section 3.3 uses the previously made observations to build a model that explains the present dependency structure. Section 3.4 devises a robust correlation measure that we later use in our own analysis. For this three visualization techniques are presented in section 3.5. Section 3.6 evaluates the proposed methodology through empirical study of three scenarios that are important in our paper application. The chapter is concluded with a discussion about the methodology. Throughout the chapter we make use of the notation that was introduced in chapter 1.

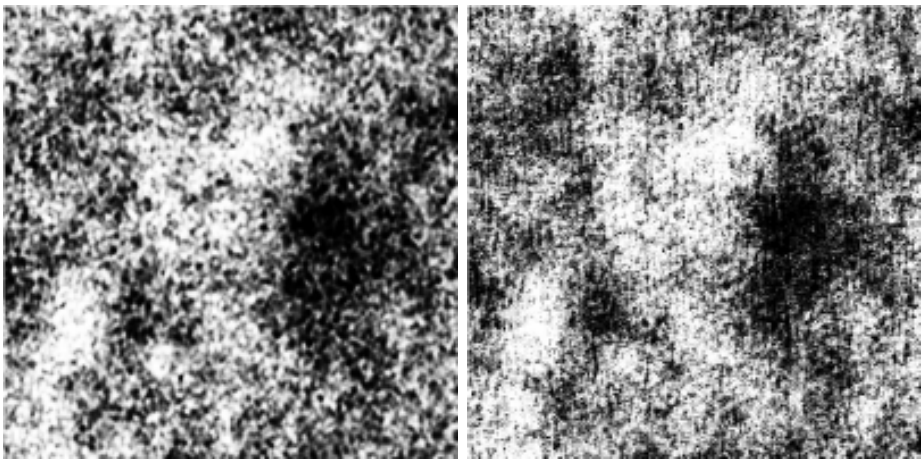


FIGURE 9: Examples of mass distribution and foreground topography from a $2cm \times 2cm$ area. The presence of dependency between these two paper properties can be seen even with a naked eye.

3.1 An overview of the methodology

In our study the dependency between images a and b is measured via a correlation coefficient $\rho^{a,b}$, which is a new spatial correlation measure that is proposed in this thesis. An overview of the algorithm that is used to estimate this correlation coefficient is presented in Table 3. This algorithm has two main phases. In the first phase we estimate the correlation between local image areas. Such estimates can have high variation due to preprocessing and due to the randomness of the sampling (see for example Figure 18). Hence in the second phase we use the estimated correlations to estimate a new robust correlation measure, which attains approximately the same values no matter how many times we repeat the estimation process.

The proposed methodology yields two-dimensional scatter plots (see Figure 12) of observations between two images. The mechanism for defining the scatter plot points is independent for the two studied paper images and hence any observed dependency cannot be an artifact of modeling but instead must be a real dependency between the two studied paper properties. From the scatter plots it is possible to make visual inference of the nature of the dependency. In order to study proportional importances of multiple paper properties, the results from multiple correlation estimates can be visualized with a box-and-whisker plot (see Figure 14).

3.2 Observing the dependency structure

In order to understand dependencies between two paper properties, we must study the images on a pixel level. We assume that all the images are preprocessed with geometric transformations as described in section 2.2.1. This is done to obtain pixel correspondence between the studied images. A tentative study shows that measuring dependencies between two images by simply measuring correlations between pixel

1. Obtain K correlation estimates $\hat{\rho}_k^{a,b}$ through the following procedure:
   (a) Collect a sample of independent and identically distributed observations (equation 7).
   (b) Preprocess the observations (equation 8).
   (c) Describe the preprocessed observations through features (equation 9).
   (d) Preprocess the features (equation 10).
   (e) Divide the preprocessed features into training and validation data.
   (f) Obtain canonically correlated feature directions and canonical correlations based on training data.
   (g) Compress the validation features into scalar projections (equation 11).
   (h) Calculate the amount of correlation between the scalar projections of validation data (equation 12 or 13).
2. Calculate an estimate for a robust correlation coefficient (equation 14).

TABLE 3: An overview of the proposed algorithm for estimating the amount of spatial correlation between two images.

values yields inadequate results, like those that are shown in Figure 2 on page 17. This is most likely due to the presence of spatial distortions and measurement noise that obscure the existing dependency. Analyzing large images as a whole is also difficult as the present structures are very complex and thus hard to comprehend. Our solution is to make observations from local image areas of images m^a and m^b through a random sample of independent and identically distributed (iid) pairs of subimages

$$m_i^a = \{z^a(x,y) : (x,y) \in D_i\} \text{ and } m_i^b = \{z^b(x,y) : (x,y) \in D_i\}, \tag{7}$$

which describe the local image areas around sample points $(x_i, y_i) : i \in \Omega = \{1, \ldots, N\}$.

From our application standpoint we are interested in spatial variations of material properties and how these variations are correlated between different material properties. Unfortunately making a good definition for the variation is problematic as a human eye is able to dismiss many types of variation if they are presented in a vantage context. In this thesis, rather than contemplating what is a good definition for the variation, we find it easier to approach this problem from the opposite

direction. For the observations that we have we define what is lack of variation (or uninteresting variation from the problem perspective) via means of subimages

$$\bar{m}_i^a = \left\{ \bar{z}^a(x,y) : \bar{z}^a(x,y) = \frac{\sum_{(s_x,s_y)\in D_i} z^a(s_x,s_y)}{w_{local} h_{local}} \text{ for } (x,y) \in D_i \right\} \text{ and}$$
$$\bar{m}_i^b = \left\{ \bar{z}^b(x,y) : \bar{z}^b(x,y) = \frac{\sum_{(s_x,s_y)\in D_i} z^b(s_x,s_y)}{w_{local} h_{local}} \text{ for } (x,y) \in D_i \right\}.$$

Because the variation of means is uninteresting, obtaining the interesting variation is a matter of subtraction. We define centered subimages as

$$\tilde{m}_i^a = m_i^a - \bar{m}_i^a \text{ and } \tilde{m}_i^b = m_i^b - \bar{m}_i^b. \tag{8}$$

For the remainder of the chapter we focus our attention on modeling and visualizing dependencies between $\tilde{m}_i^a$ and $\tilde{m}_i^b$.

Whereas it is possible to use the pixel intensities of the subimages $\tilde{m}_i^a$ and $\tilde{m}_i^b$ directly for modeling, there are at least two good reasons why switching to an alternative representation through feature functions should be considered. If we have prior knowledge about the type of phenomena that are interesting from the application perspective, we can design a set of feature functions $\phi_j : j = 1, \ldots, d$ so that they directly measure the presence of such phenomena. This also means that it should be easier to interpret the results from the application perspective. The other reason for switching to an alternative representation is to reduce the induced computational load of modeling by decreasing the amount of processed information. This is accomplished by quantifying only important aspects of an observation while disregarding all the rest. This means that the number of scalar values that are required to describe the observation is reduced. Formally we observe the change in the subimages through features

$$\mathbf{f}_i^a = \begin{pmatrix} \phi_1(\tilde{m}_i^a) \\ \vdots \\ \phi_d(\tilde{m}_i^a) \end{pmatrix} \text{ and } \mathbf{f}_i^b = \begin{pmatrix} \phi_1(\tilde{m}_i^b) \\ \vdots \\ \phi_d(\tilde{m}_i^b) \end{pmatrix}. \tag{9}$$

Without loss of generality we can assume that the observed features are preprocessed as

$$\mathbf{f'}_i^a \leftarrow \mathbf{f}_i^a - \frac{\sum_{i=1}^N \mathbf{f}_i^a}{N} \text{ and } \mathbf{f'}_i^b \leftarrow \mathbf{f}_i^b - \frac{\sum_{i=1}^N \mathbf{f}_i^b}{N}, \tag{10}$$

which essentially centers the feature vectors. These features are then stacked into matrixes

$$\mathbf{F}^a = \begin{pmatrix} (\mathbf{f'}_1^a)^T \\ \vdots \\ (\mathbf{f'}_N^a)^T \end{pmatrix} \text{ and } \mathbf{F}^b = \begin{pmatrix} (\mathbf{f'}_1^b)^T \\ \vdots \\ (\mathbf{f'}_N^b)^T \end{pmatrix}$$

for mathematical convenience.

In our application we do not have the required prior knowledge to select the most appropriate feature functions and making a wrong guess here would certainly lead into poor modeling results. One solution is to define the feature functions $\boldsymbol{\phi}^a(\cdot)$ and $\boldsymbol{\phi}^b(\cdot)$ so that they use the subimage pixel values directly as features. The upside of this approach is that it maintains a clear spatial interpretation between the features and the spatial structure that they describe. The downside is that increasing the observed spatial area also increases the amount of pixels that are involved and thus increases the dimensions of the feature vectors, which soon becomes an unbearable computational strain. Our solution is to improve this idea by fixing the number of pixels in the subimage while expanding the observed spatial area by introducing a subimage to image scaling factor r that controls the spatial extent of the subimages with respect to the measurement image. In practice this means that $w_{local}r$ pixels times $h_{local}r$ pixels image areas around observation points (x_i, y_i) are rescaled to w_{local} pixels times h_{local} pixels subimages using geometric transformations as defined in section 2.2.1. This means that we still have a clear spatial interpretation for the features but now we can also observe larger spatial areas without increasing the induced computational burden. An additional benefit of this approach is that we can easily study the amount of dependency between the images in different spatial scales.

3.3 Estimating the dependency structure

The first thing that we need to consider with respect to the observations is how to use them correctly from the statistical point of view. In statistical modeling a common approach is to divide the observations into two disjoint data sets. One of these data sets is called training data, which is used to train a model, and the other data set is called validation data, which is used to empirically evaluate the model. In our case this idea is implemented by dividing the observation indexes in set $\Omega = \{1, \ldots, N\}$ into two disjoint sets Ω_{tr} and Ω_{val} such that $\Omega = \Omega_{tr} \cup \Omega_{val}$ and $\Omega_{tr} \cap \Omega_{val} = \emptyset$. Based on this division we obtain training data matrixes $\mathbf{F}^a_{tr}$ and $\mathbf{F}^b_{tr}$, and validation data matrixes $\mathbf{F}^a_{val}$ and $\mathbf{F}^b_{val}$.

Next we devise a statistical model that qualifies and quantifies the dependency between the observations. Our solution is to use canonical correlation analysis (CCA), which was introduced in section 2.3.2, to decompose complex spatial structures in the observations into simpler and more interpretable structures, and then measure the dependency between these simpler structures. This idea is illustrated in Figure 10. Formally CCA yields pairs of d dimensional canonical vectors $\boldsymbol{\beta}^a_j, \boldsymbol{\beta}^b_j$ and canonical correlations λ_j for which $j = 1, \ldots, d$ (in our case scaled subimage pixels are used as features and hence we have $d = w_{local}h_{local}$). Because of the way that we defined CCA in section 2.3.2, the canonical correlations λ_j are always positive and in range $[0, 1]$. Each pair of canonical vectors $\boldsymbol{\beta}^a_j, \boldsymbol{\beta}^b_j$ is statistically linked so that the variation of $\mathbf{f}'^a_i$ towards direction $\boldsymbol{\beta}^a_j$ has the highest correlation with the variation of $\mathbf{f}'^b_i$ towards direction $\boldsymbol{\beta}^b_j$. The spatial interpretation for these reads as follows: Be-

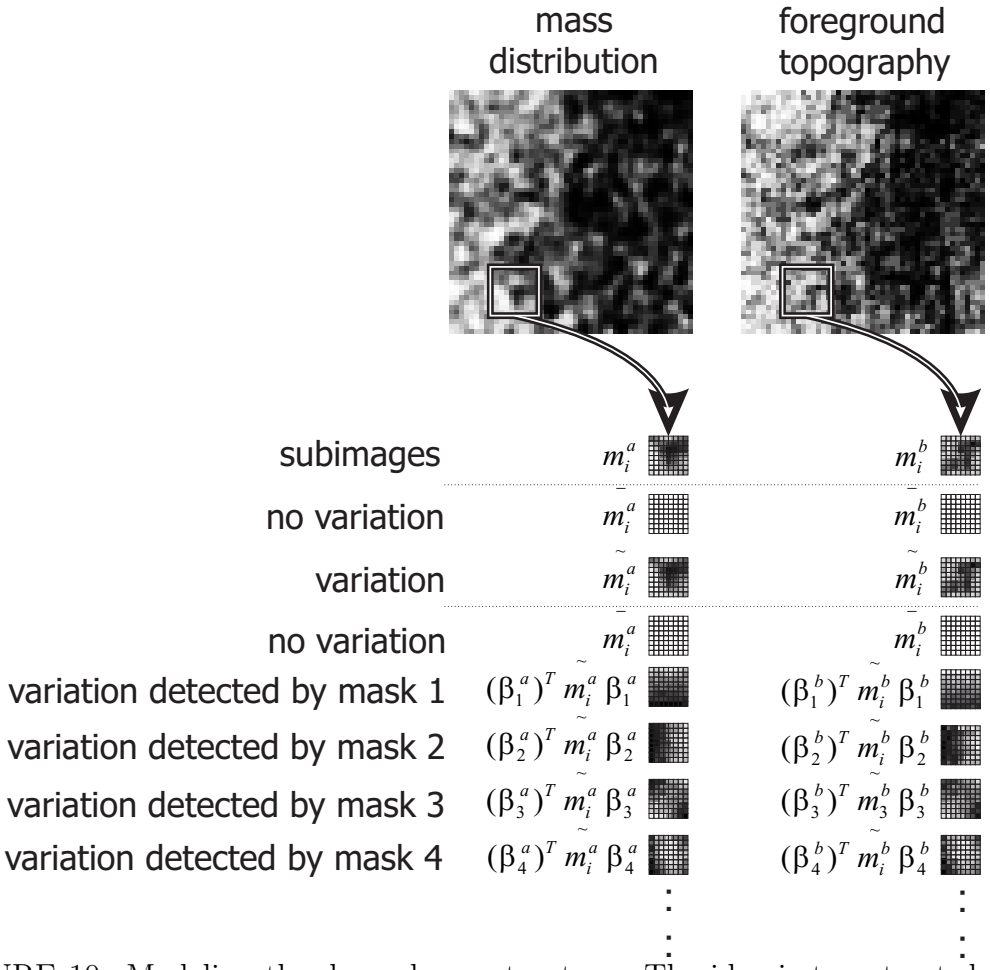
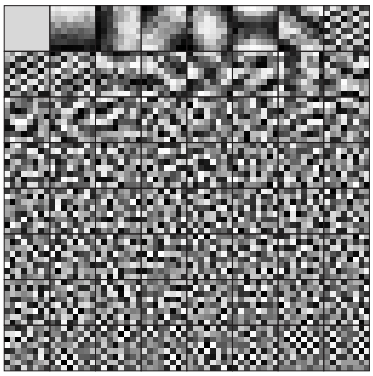


FIGURE 10: Modeling the dependency structure. The idea is to extract descriptions of the local variations, then describe them as linear combinations of simpler component variations that we call masks, and finally measure the amount of linear correlation between these linear combinations. In our work we use pixels of scaled subimages as features and hence feature vectors and canonical vectors have a clear spatial interpretation.

cause we use subimage pixel values as features, the canonical vectors $\boldsymbol{\beta}_j^a$ and $\boldsymbol{\beta}_j^b$ can be interpreted as $w_{local} \times h_{local}$ pixels masks, which have spatial interpretation that is equivalent to the subimages. Hence observing a spatial pattern of $\boldsymbol{\beta}_j^a$ in material property a indicates that there is a statistical change that a spatial pattern of $\boldsymbol{\beta}_j^b$ is present in material property b. Here the word statistical means that observing the spatial pattern of $\boldsymbol{\beta}_j^a$ in property a does not guarantee observing the spatial pattern of $\boldsymbol{\beta}_j^b$ in property b but does state that it is possible with a probability that is proportional to λ_j.

The estimation of masks $\boldsymbol{\beta}_j^a, \boldsymbol{\beta}_j^b$ and correlations λ_j is based on matrixes $\mathbf{F}_{tr}^a$ and $\mathbf{F}_{tr}^b$. First we define a joint covariance matrix as

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^{aa} & \hat{\boldsymbol{\Sigma}}^{ab} \\ \hat{\boldsymbol{\Sigma}}^{ba} & \hat{\boldsymbol{\Sigma}}^{bb} \end{pmatrix} = \frac{\begin{pmatrix} \mathbf{F}_{tr}^a & \mathbf{F}_{tr}^b \end{pmatrix}^T \begin{pmatrix} \mathbf{F}_{tr}^a & \mathbf{F}_{tr}^b \end{pmatrix}}{N-1}$$



1 NA	2 0.763	3 0.710	4 0.583	5 0.553	6 0.478	7 0.372	8 0.351
9 0.334	10 0.331	11 0.313	12 0.285	13 0.262	14 0.254	15 0.241	16 0.234
17 0.229	18 0.215	19 0.211	20 0.203	21 0.202	22 0.198	23 0.195	24 0.192
25 0.184	26 0.183	27 0.180	28 0.177	29 0.170	30 0.169	31 0.164	32 0.161
33 0.154	34 0.153	35 0.151	36 0.148	37 0.145	38 0.144	39 0.140	40 0.137
41 0.135	42 0.132	43 0.131	44 0.127	45 0.126	46 0.123	47 0.121	48 0.120
49 0.118	50 0.114	51 0.113	52 0.109	53 0.107	54 0.106	55 0.102	56 0.097
57 0.092	58 0.090	59 0.085	60 0.083	61 0.080	62 0.079	63 0.073	64 0.071

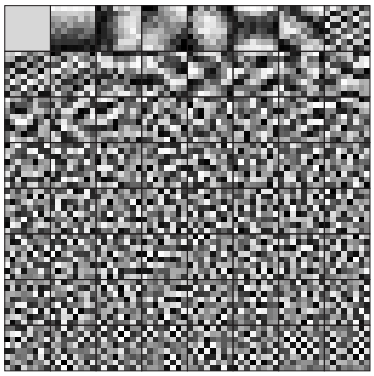


FIGURE 11: Example of masks for mass distribution (left) and foreground topography (right). The center matrix defines the component order j and the amount of linear correlation λ_j between masks $\hat{\boldsymbol{\beta}}_j^a$ and $\hat{\boldsymbol{\beta}}_j^b$. The first masks are reserved for the lack of variation (uninteresting variation) that was removed in the preprocessing step. In our work we use pixels of scaled subimages as features and hence the feature vectors and canonical vectors have a clear spatial interpretation.

and then use it to define two matrixes $(\hat{\boldsymbol{\Sigma}}^{aa})^{-1}\hat{\boldsymbol{\Sigma}}^{ab}(\hat{\boldsymbol{\Sigma}}^{bb})^{-1}\hat{\boldsymbol{\Sigma}}^{ba}$ and $(\hat{\boldsymbol{\Sigma}}^{bb})^{-1}\hat{\boldsymbol{\Sigma}}^{ba}(\hat{\boldsymbol{\Sigma}}^{aa})^{-1}\hat{\boldsymbol{\Sigma}}^{ab}$ from which we obtain $\hat{\boldsymbol{\beta}}_j^a, \hat{\boldsymbol{\beta}}_j^b$ as the postprocessed eigenvectors and $\hat{\lambda}_j$ as the square roots of the eigenvalues respectively. Postprocessing, as described in section 2.3.2, makes sure that the obtained correlations between masks $\hat{\boldsymbol{\beta}}_j^a, \hat{\boldsymbol{\beta}}_j^b$ are positive, which makes interpretations easier. Use of robust covariance estimators (Rousseeuw and Leroy 2003) may be necessary to obtain a reliable result. We have made a habit of using minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) estimates for $\boldsymbol{\Sigma}$ to avoid the need to check each analysis for outliers. An example of estimated masks $\hat{\boldsymbol{\beta}}_j^a, \hat{\boldsymbol{\beta}}_j^b$ and estimated correlations $\hat{\lambda}_j$ are presented in Figure 11. Visual examination of the masks shows that the highest correlating masks tend to represent clear patterns of variation whereas the latter ones look more like shot noise. There is a natural explanation for this: the highest correlating masks represent those geometric structures that are statistically most commonly linked between the paper properties. The remaining masks are then left to handle less frequent incidents and noise. From this heuristic, it makes sense to reduce the effect of noise by limiting any further analyzes to only those masks that have a clear geometric structure or a dependency with interpretation. Formalization of this idea means that we have a set of indexes of all possible masks $\Xi = \{1, \ldots, d\}$ from which we select a subset $\Xi^* \subset \Xi$ that consists of indices of those masks that we want to include into our analysis. As selecting the masks by hand soon becomes laborious, we have decided to make use of the fact that the first d^* masks explain approximately $(\sum_{j=1}^{d^*} \lambda_j)/(\sum_{j=1}^{d} \lambda_j)$ percent of the

variation between the training observations. This means that we can set

$$\Xi^* = \{1, \ldots, d^*\} \text{ where } d^* = \arg\min_{d'} \frac{\sum_{j=1}^{d'} \lambda_j}{\sum_{j=1}^{d} \lambda_j} > 0.9$$

to automatically select a subset of masks that is likely to explain the dependency well but does not include the less useful masks. This approach, although not rigorously justified, has proven to work well in our paper application.

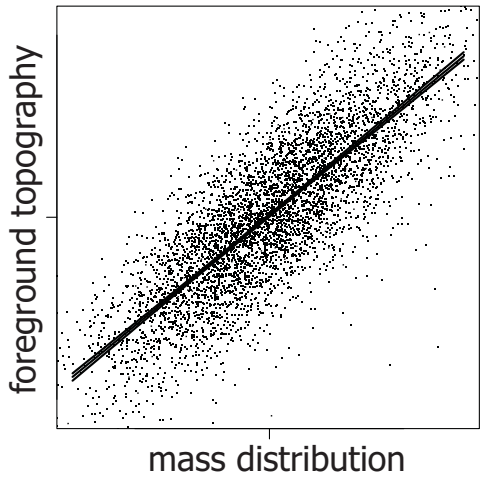


FIGURE 12: Example scatter plot of projected validation observations in CCA subspaces, which are made from mass distribution (horizontal axis) and foreground topography (vertical axis). The horizontal and vertical positions for each validation observation are obtained from equation 11. The regression line and its empirical 95% confidence intervals show the general trend between the observations.

On pixel level the masks are able to characterize the existing dependencies. Unfortunately the pixel level representation is also far too complex to give a clear understanding of what these dependencies actually mean. This is due to the fact that several masks can be jointly responsible for the detection of visually simple structures. As a result the geometric interpretation can be hidden under an unknown combination of masks. What is needed is a low dimensional illustration of the dependency that can be interpreted visually. Because the masks, which were obtained from CCA, are orthogonal and hence statistically independent, we can device a projection scheme that compresses the information detected by the masks into two scalar values. Two measures for the amount of masks $\boldsymbol{\beta}_j^a$ and $\boldsymbol{\beta}_j^b$ in observations m_i^a and m_i^b are obtained as projections

$$p_{i,j}^a = (\boldsymbol{\beta}_j^a)^T m_i^a \text{ and } p_{i,j}^b = (\boldsymbol{\beta}_j^b)^T m_i^b \text{ for all } i = 1, \ldots, N \text{ and } j = 1, \ldots, d.$$

Because of the statistical independency between $p_{i,j}^a : j = 1, \ldots, d$ and $p_{i,j}^b : j = 1, \ldots, d$ we have devised an ad hoc rule to sum these to obtain two

cumulative projections

$$p_i^a = \sum_{j \in \Xi^*} \lambda_j p_{i,j}^a \text{ and } p_i^b = \sum_{j \in \Xi^*} \lambda_j p_{i,j}^b \tag{11}$$

that use the amount of correlation between the masks to automatically adjust the proportional importance of the masks. These cumulative projections can be visualized in a scatter plot as we have done in Figure 12. Although a rigorous justification for the proposed projection rule is hard to establish, this projection seems to function well based on our empirical experiments. An another approach, which we have tried successfully, is to sum the projections of the first masks without any weighting.

From the scatter plot we can see that there is an obvious linear dependency between the cumulative projections p_i^a, p_i^b of the validation observations. Because the observations that are used to build the dependency model are are not the same that are used for visualization and because the observation mechanism is independent for the two studied paper properties, this dependency cannot be an artifact of the model but instead must be a real dependency between the observations. Each point of the plot corresponds to a pair of observations that are made from the two studied paper properties. The horizontal position of a point describes the variation in paper property a and the vertical position of the same point describes the variation in material property b. In addition we have plotted a least squares linear regression line with empirical 95% confidence intervals to illustrate the general trend and its statistical stability. Tight confidence bounds, such as the ones that we have here, indicate that the regression model is statistically reliable and can be used for example for prediction purposes. The next step is to devise a measure that is able to quantify the amount of correlation between the observations with a single scalar value. A natural choice for this is to use Pearson's correlation estimate

$$\hat{\rho}^{a,b} = \frac{N \sum_{i=1}^N p_i^a p_i^b - \sum_{i=1}^N p_i^a \sum_{i=1}^N p_i^b}{\sqrt{N \sum_{i=1}^N \left(p_i^a\right)^2 - \left(\sum_{i=1}^N p_i^a\right)^2} \sqrt{N \sum_{i=1}^N \left(p_i^b\right)^2 - \left(\sum_{i=1}^N p_i^b\right)^2}}, \tag{12}$$

or Spearman's correlation estimate

$$\bar{\rho}^{a,b} = 1 - \frac{6 \sum_{i=1}^N \left(\text{rank}(p_i^a) - \text{rank}(p_i^b)\right)^2}{N(N^2-1)}. \tag{13}$$

These estimates were introduced in section 2.3.1. In our example case, which was presented in Figure 12, the former estimate yields 0.825, which is an indication of strong linear correlation between the two studied paper properties. This concurs with our intuition from Figure 9 where the present dependency is evident even to a naked eye.

3.4 Modeling the dependency structure

Performing the observation and the estimation of the last two sections yield a scalar estimate that measures the amount of dependency between the two studied images. Repeating this observation-estimation procedure K times produces

$\hat{\rho}_k^{a,b} : k = 1, \ldots, K$ estimates that vary between repetitions. This is due to the fact that the proposed estimator contains randomness from the image preprocessing and from the stochastic placement of the local area observations. This means that the estimator, which is a random variable, can yield poor estimates with some probability and thus a single estimate cannot be trusted to give a fair assessment of the true amount of dependency. Our solution is to use all the K independent and approximately identically distributed estimates to empirically evaluate the distribution of the estimator.

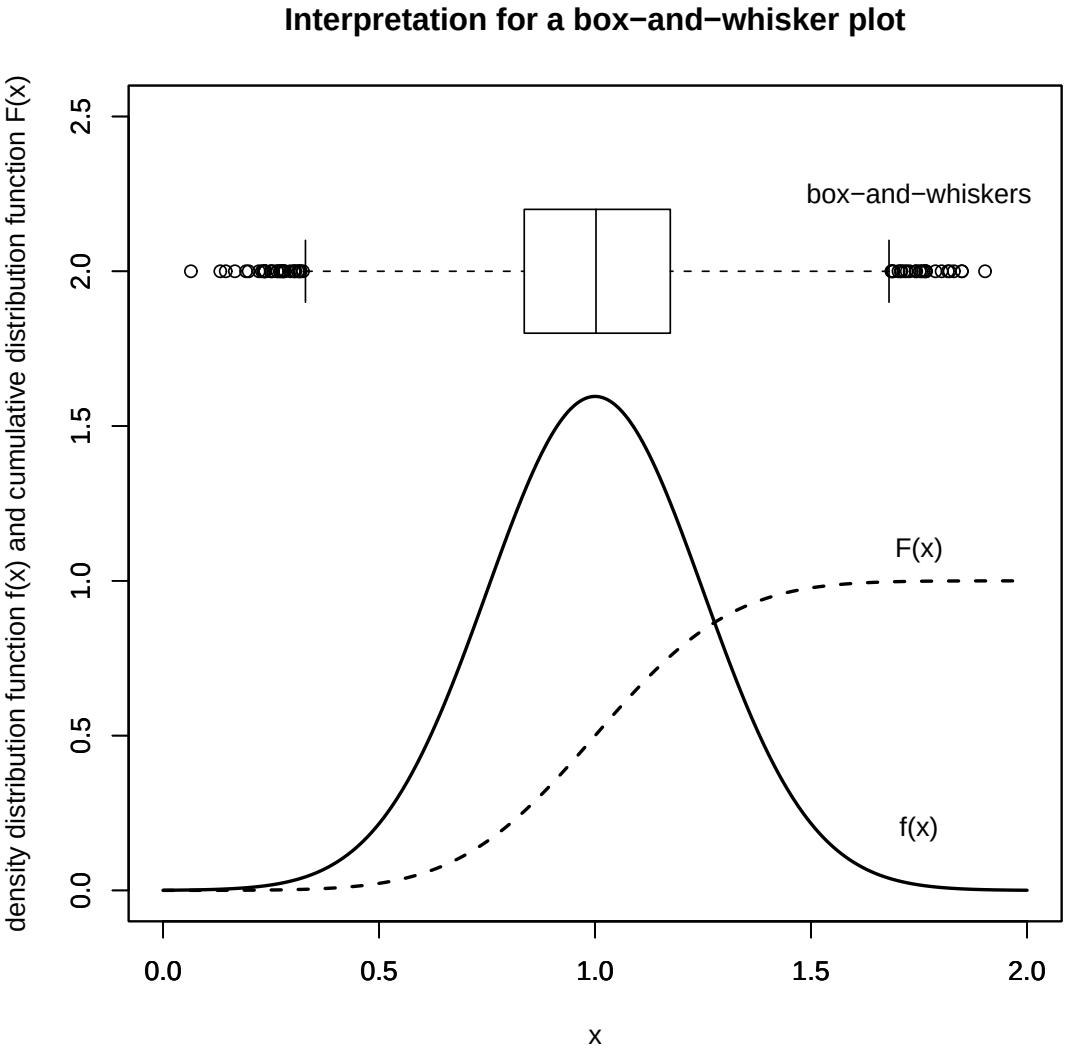


FIGURE 13: Interpretation for a box-and-whisker plot illustrating a visualization of $Normal(1, 0.25)$ distribution.

The estimator distribution can be visualized with a graphical box-and-whisker plot, which was originally proposed by Tukey (Tukey 1977, pp. 39-43) and is illustrated in Figure 13. The interpretation for the plot reads as follows: The box is drawn between 25% and 75% distribution quantiles. Inside this box a line indicates the position of the 50% distribution quantile, which is known as the median. Out-

side of the box whiskers are drawn to the farthest points that are not considered outliers (that are within 3/2 times the interquartile range of 25% and 75% distribution quantiles). The outlier points that lie outside of the range of the whiskers are drawn with a circle. From our application perspective this plotting technique includes all the relevant information in an easily interpretable form. Placing multiple plots side-by-side allows easy comparison of relations between different paper properties.

Our solution to obtain a statistically sound estimator for the true amount of dependency is to define a robust correlation coefficient

$$\varrho^{a,b} = \text{median}(\{\rho_k^{a,b} : k = 1, \ldots, K\}), \tag{14}$$

which estimates are stable enough to be used for statistical inference. Although an analytical evaluation of this estimator is difficult due to fact that large number of factors cannot be exactly defined and thus such an analysis is omitted from our study, from our empirical experimentation we can state that calculating repeated estimates of this correlation coefficient yield closely the same result. Similarly other distribution quantiles, especially those that are not too far from the median like 25% and 75% distribution quantiles, are also stable and taking pairs of such provides an easy way to estimate the variation of the area correlation coefficient. From application perspective, we are often interested in how the correlation coefficient $\varrho^{a,b}$ evolves as a function of spatial scale. An easy way to empirically study this is to estimate the correlation coefficient as a function of the subimage to image scaling factor r. In the later chapters this function, which we denote with

$$\varrho^{a,b}(r), \tag{15}$$

forms the backbone of our own analyzes.

3.5 Visualization techniques

In this section we propose a few visualization techniques that illustrate how the model that was devised in the previous sections explains the correlation between spatial variations of different paper properties. For illustration purposes we use images of mass distribution and foreground topography (see Figure 9). The analysis is based on $N = 10000$ observations that are partitioned into $N_{tr} = 5000$ training and $N_{val} = 5000$ validation observations, and on $w_{local} = 10$ pixels times $h_{local} = 10$ pixels subimages with the subimage to image scaling factor r belonging to range $[0.25, 4.00]$. In spatial terms this means that the observed physical area ranges from $0.25mm \times 0.25mm$ to $4.00mm \times 4.00mm$. In the presented result plots the horizontal axes indicate the parametrization of the studied phenomenon whereas the vertical axes always measure the amount of dependency.

Estimation of robust correlation coefficients for all pairs of paper properties allows us to draw collections of box-and-whisker plots where visual comparisons

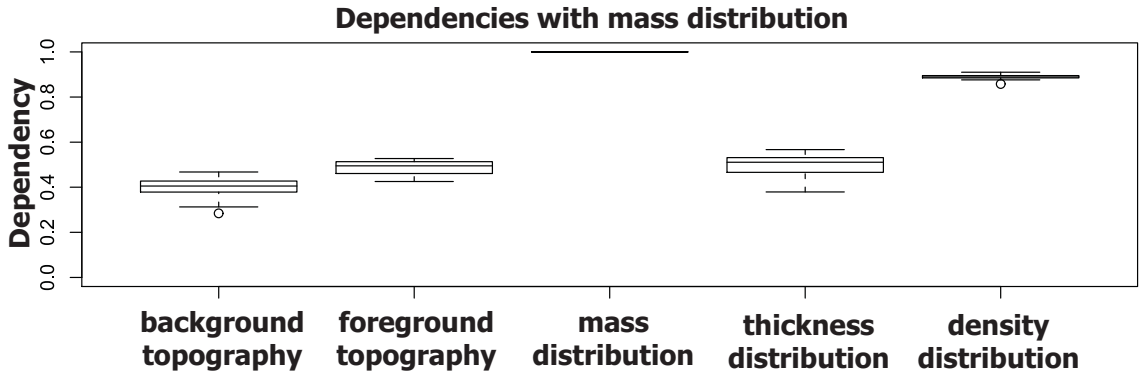


FIGURE 14: Correlation coefficients between mass distribution and the other paper properties with subimage to image scaling 1 : 1. Each box-and-whisker plot describes the distribution of a correlation coefficient $\varrho^{a,b}$ between mass distribution and the paper property whose name is indicated below the plot. The vertical axis measures the amount of dependency.

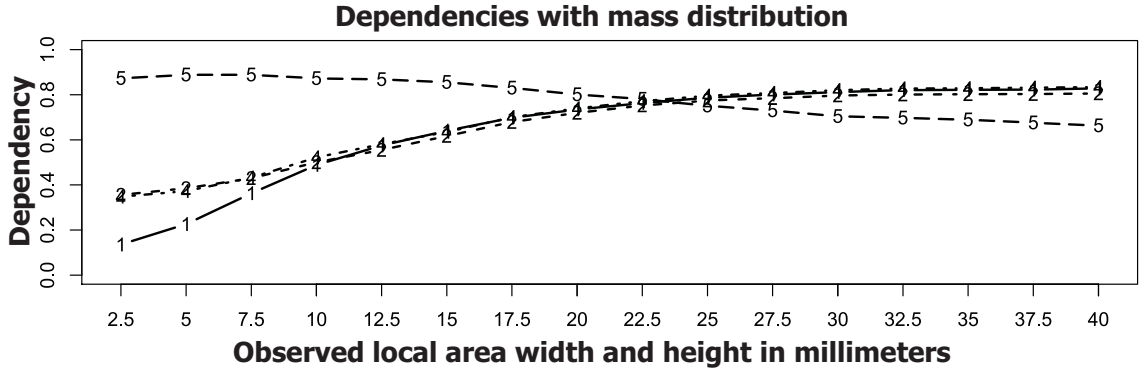


FIGURE 15: Correlation functions between mass distribution and the other paper properties. Each paper property is identified with a number: 1=background topography, 2=foreground topography, 3=mass distribution (omitted), 4=thickness distribution, and 5=density distribution. The horizontal axis indicates the size of the observed local area and the vertical axis measures the amount of dependency.

between the plots are easy. An obvious application for this is to select one paper property towards which all the other paper properties are compared. This idea is illustrated in Figure 14. The interpretation for this plot is as follows: The title line indicates the paper property to which we compare all the other properties. The horizontal axis identifies the other property as the name of the other property for each plot is drawn under the plot. The vertical axis indicates the amount of dependency and always uses the range $[0, 1]$. In this example we study the dependency between mass distribution and the other four measured paper properties. As we can see, in this case the density distribution is the highest correlating paper property whereas thickness distribution along with background and foreground topographies come behind. The conclusion is that in this case all the studied paper properties are correlated with the mass distribution, which concurs with our intuition and is

hence very easy to accept. An another good visualization is to omit the confidence bounds of the correlation coefficient and to draw the correlation coefficient $\varrho^{a,b}(r)$ with respect to different subimage to image scaling factors r as we have done in Figure 15. As we can see, the effect of the scaling factor is substantial in this case. Whereas the confidence bounds of the estimate distributions are important for problem diagnostic purposes, the amount of dependency with respect to spatial scale is usually more interesting from the application perspective.

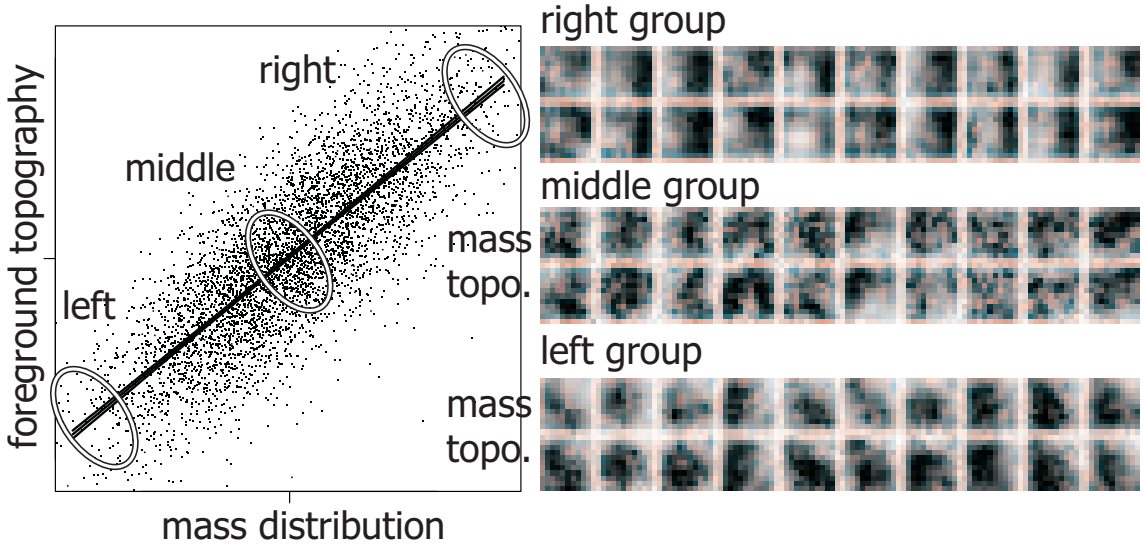


FIGURE 16: Spatial interpretation for the regression model. Visualizing observations that are well explained by a regression model yields an intuitive interpretation for this regression model.

In Figure 12 we used a linear regression model to illustrate the general trend between the observations that were made from the studied paper properties. We know that the observations that are located at the lower-left corner are somehow different from the ones that are located at the upper-right corner but how they are different has not been made clear yet. Including a spatial interpretation for the regression line answers this question. One way to obtain such an interpretation is to pick observations from the discussed corners and in addition from the center and place pairs of subimages $\tilde{m}_i^a$ and $\tilde{m}_i^b$ from these observations on three groups as we have done in Figure 16. Formally the observations in these groups are indexed with

$$\Omega_{left} = \{i : \mathrm{rank}(\mathrm{rank}(p_i^a) + \mathrm{rank}(p_i^b)) \leq M, i \in \Omega_{val}\}$$

$$\Omega_{middle} = \left\{i : \mathrm{rank}\left(\mathrm{abs}\left(\mathrm{rank}(p_i^a) - \frac{N_{val}}{2}\right) + \mathrm{abs}\left(\mathrm{rank}(p_i^b) - \frac{N_{val}}{2}\right)\right) \leq M, i \in \Omega_{val}\right\}$$

$$\Omega_{right} = \{i : \mathrm{rank}(\mathrm{rank}(p_i^a) + \mathrm{rank}(p_i^b)) > N_{val} - M, i \in \Omega_{val}\},$$

where M is the number of picked observations and the operator $\mathrm{rank}(\cdot)$ returns the rank order of the parameter among its peers and the operator $\mathrm{abs}(\cdot)$ returns the

absolute value of the parameter. It should be noted that these three equations rely on all the canonical correlations to be positive. In Figure 16 the three groups of masks are interpreted as follows: The left group of observations consists of pairs of subimages $\tilde{m}_i^a, \tilde{m}_i^b, i \in \Omega_{left}$ so that in pairs $\tilde{m}_i^a$ is placed above $\tilde{m}_i^b$. The middle and the right groups of observations are defined in an analogous manner for observations in Ω_{middle} and Ω_{right} respectively. As we can see from Figure 16, the left and the right group consists of observations that have similar spatial interpretations whereas the middle group does not. This concurs with the fact that deviation in the direction of the regression line increases the amount of dependency between the observations. As we can see, this visualization technique suffers from noise presence and hence it is recommended that all images are filtered to remedy the effects of noise prior to visualization.

3.6 Empirical experimentation of the methodology

Before we are able to conduct effective data analysis with the proposed methodology, we run some experiments to see how the methodology reacts to three scenarios that are important from our application perspective. These scenarios study how subimage size and scaling, mean filter preprocessing, and pixel correspondence affect the obtained results. In our experiments we use the same images, mass distribution and foreground topography, that we used in the previous section. All the analyzes are conducted with $K = 25$ repetitions and based on $N = 10000$ observations that are partitioned into $N_{tr} = 5000$ training and $N_{val} = 5000$ validation observations. If not otherwise indicated in the result figures, the analysis is performed using $w_{local} = 10$ pixels times $h_{local} = 10$ pixels subimages with the subimage to image scaling factor r ranging in $[0.25, 4.00]$. In the presented result plots the horizontal axes indicate the parametrization of the studied phenomenon whereas the vertical axes always measure the amount of dependency.

In the first experiment we study how increasing the size of the observed image areas affects the measured dependency. There are two ways to do this: we can increase the amount of pixels in our subimages, which means increasing w_{local} and h_{local}, and we can increase the subimage to image scaling factor r that changes the spatial extent of the subimages with respect to the images. Results from both of these approaches are presented in Figure 17. It seems that increasing the spatial extent of the observations increases the amount of observed dependency. Increasing the amount of pixels in the subimages allows the masks to capture larger structures. It seems that the material structures that have the strongest dependencies between them in the two studied images do not properly fit in the smaller observation areas and hence the observed dependency is also weaker. Increasing the subimage to image scaling factor r increases the spatial extent of the observations and leads into the same conclusion. In addition it increases the amount of implicit smoothing in the geometric transformations of the subimages. This reduces the amount of noise which results in the correlation coefficient estimator distributions becoming tighter.

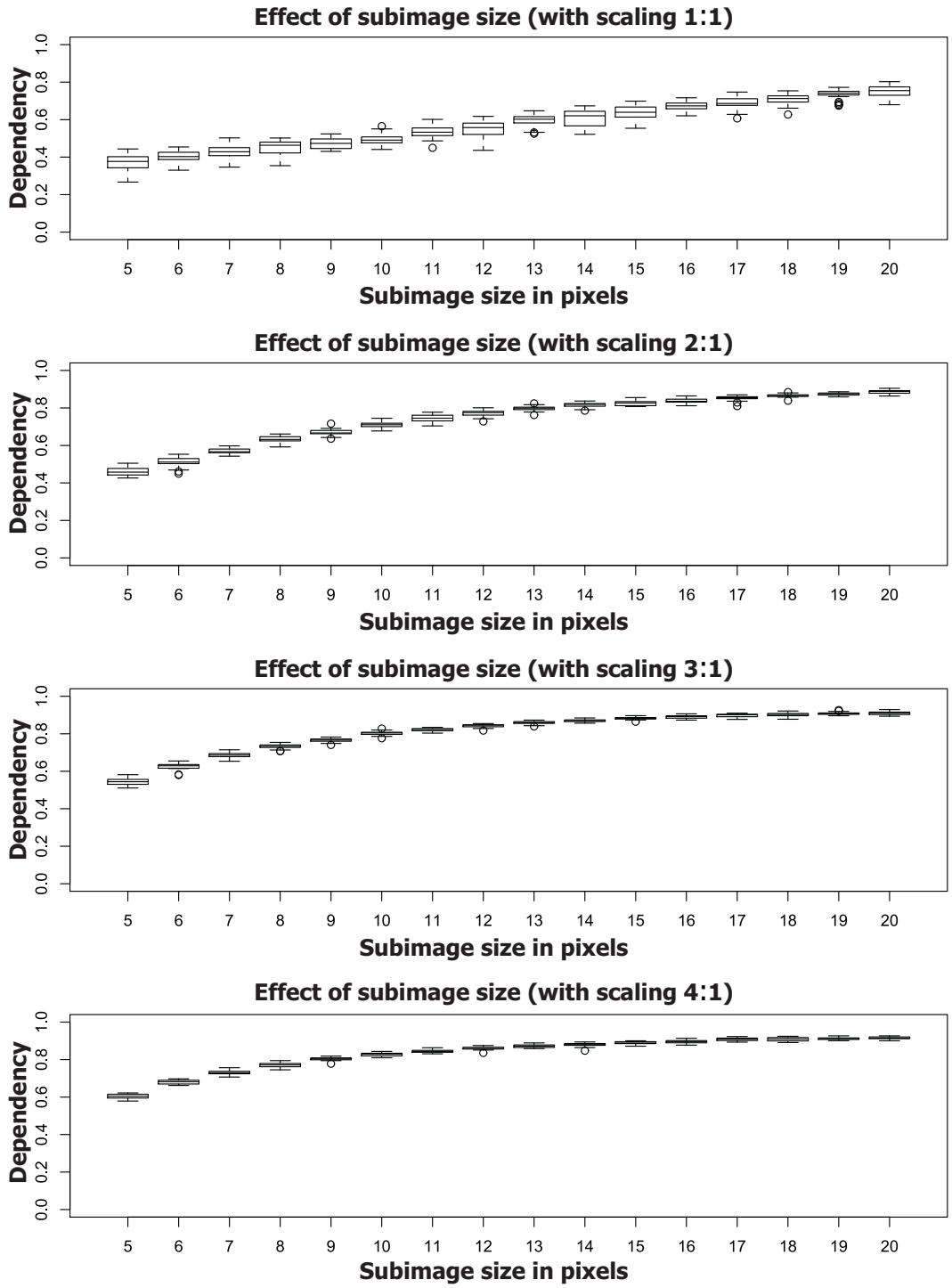


FIGURE 17: Effect of subimage size and subimage to image scaling factor on dependency between mass distribution and foreground topography. The four plots use subimage to image scaling $1:1$, $2:1$, $3:1$, and $4:1$ respectively. The horizontal axes indicate the used subimage width w_{local} and height h_{local}, whereas the vertical axes measure the amount of dependency.

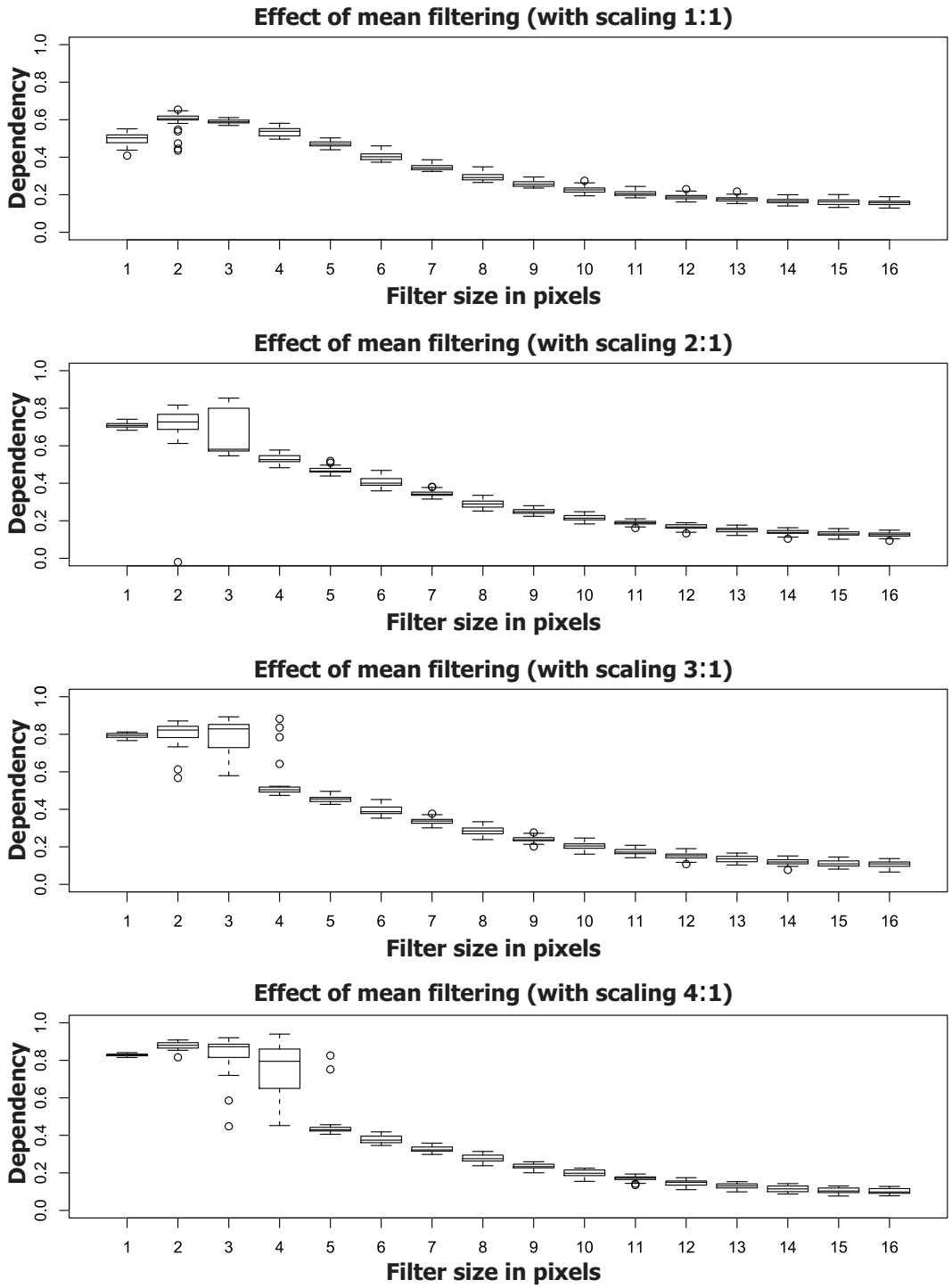


FIGURE 18: Effect of mean filter preprocessing on dependency between mass distribution and foreground topography. The four plots use subimage to image scaling $1:1$, $2:1$, $3:1$, and $4:1$ respectively. The horizontal axes indicate the used mean filter width and height, whereas the vertical axes measure the amount of dependency.

In the second experiment we study the effect of noise and how it influences our ability to make proper conclusions – especially from the available spatial rep-

resentations (see Figure 16). Whereas from theoretical point of view CCA is able to bypass random noise because it does not correlate, in practice this is not always the case. In fact the noise can correlate by chance and obscure parts of the spatial structures between which the dependencies are sought for. In case of pixel features, CCA does not understand the spatial relations between the pixels and hence filtering techniques that see the image as a two-dimensional lattice may be able to produce added value. The obtained results are presented in Figure 18. There is a small but evident benefit coming from filtering. From the results, we can see that in this case conservative filtering with 2×2 or 3×3 pixels mean filters produce the best results. Filter sizes larger than this are clearly harmful as the payoff from reduced noise diminishes rapidly while the increased smoothing wipes out ever increasing parts of valuable information. Unfortunately these results vary between different combinations of measurement images and hence one all-inclusive answer cannot be given. It is the responsibility of the data analyst to select the proper amount of filtering for each studied image.

In the third experiment we study the effect of pixel correspondence. In our paper application lacking pixel correspondence originates from imprecise detection of the markers that are used to define the pixel correspondence structure (recall section 2.2.1). These markers occupy multiple pixel regions and hence it is impossible to pinpoint all the markers with complete accuracy. In reality the markers have two types of problems: some markers may be accidentally misplaced far away from their correct locations making them outliers, and all markers may contain noise from an unknown distribution. The former problem of outlier markers can effectively be remedied with the use of robust regression techniques. The latter case of noisy markers is more difficult as the only way for improvement is to repeat the manual pinpointing of markers, which requires a lot of manual work. Hence, in this experiment we study the effects of noisy markers on our methodology. We do this by taking one image, the mass distribution image, and measure the amount of dependency between two instances of this image when the geometry of one of these instances is distorted by manipulating the markers of this instance with Gaussian noise of zero mean and small standard deviations. The obtained results are presented in Figure 19. As we can see the amount of dependency starts from 1 which is logical as here the two instances are identical and thus we have perfect pixel correspondence. As we increase the amount of distortion the dependency decreases almost linearly. It is interesting to see the effect of subimage to image scaling factor. Whereas the larger scaling factors implicitly smooth out some amount of distortion and hence operate a bit better, the loss of dependency as a function of the amount of distortion is so dramatic that the smoothing on its own cannot remedy the problem. As a conclusion we can state that the methodology is able to tolerate small scale noise but also that in larger scales the noise has a very negative effect on the methodology and hence the lack of pixel correspondence can effectively make or brake the whole analysis. Whereas some feature functions are, to some degree, less sensitive to the lack of pixel correspondence, even these cannot solve the problem entirely.

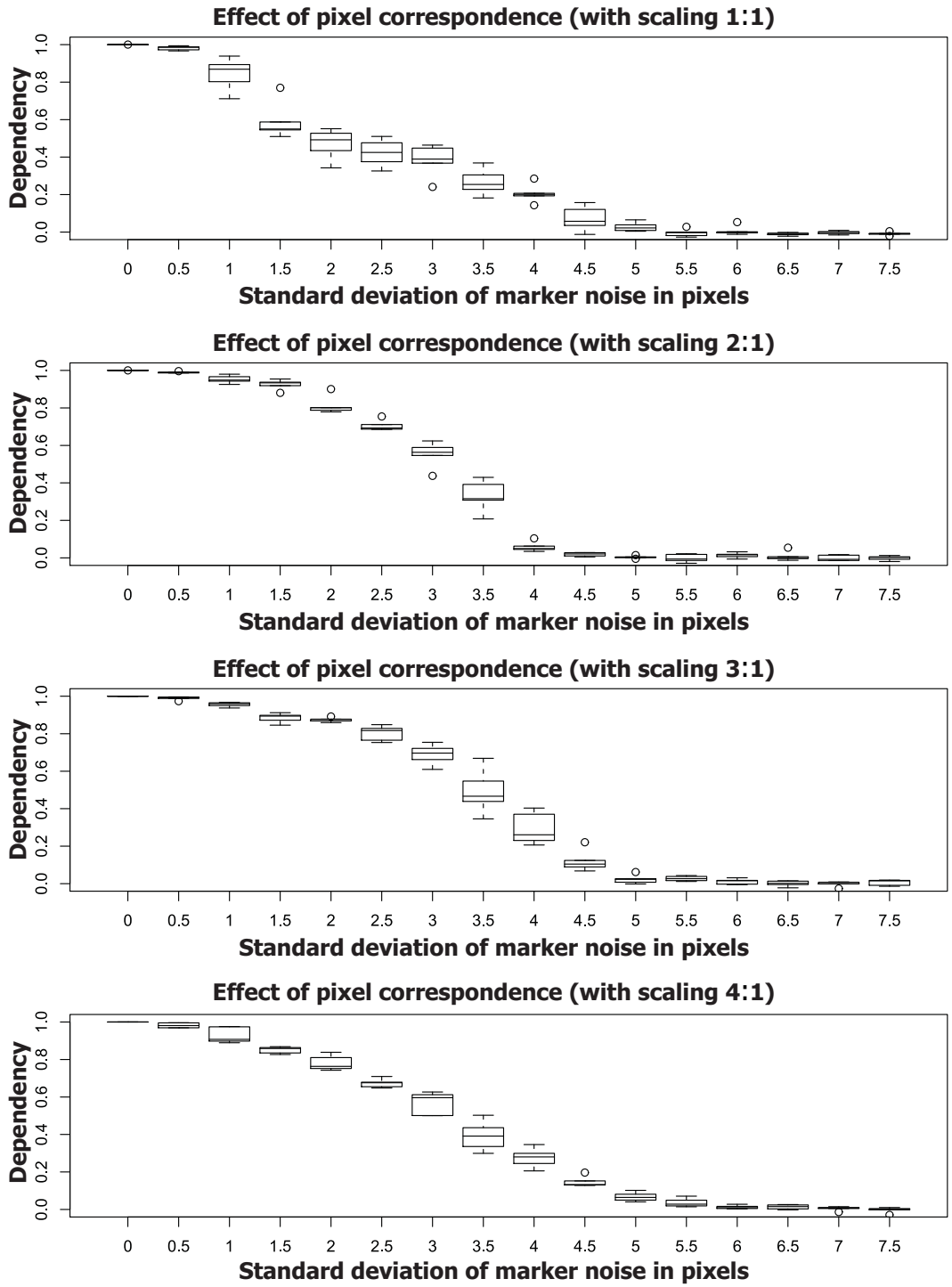


FIGURE 19: Effect of pixel correspondence on dependency between mass distribution and foreground topography. The four plots use subimage to image scaling $1:1$, $2:1$, $3:1$, and $4:1$ respectively. The horizontal axes indicate the standard deviations of the Gaussian noise that is added to the markers, whereas the vertical axes measure the amount of dependency.

3.7 Discussion

In this chapter we have proposed a methodology for detecting and measuring spatial dependencies between images. As far as we know there are no comparable methodologies in the literature that would answer the type of questions that we try to answer in our study. In chapter 5 the proposed methodology is employed on our paper images. With respect to future research we state that extending the methodology, which was presented in a simplified form in this chapter for illustration purposes, is very easy. Two extensions that are of particular interest are using linear CCA between more than two images at the same time and using non-linear CCA or even kernel CCA to observe non-linear dependencies. The necessary building blocks for both of these ideas are already available in the literature.

4 INTERACTIVE WORKING ENVIRONMENT FOR DATA ANALYSIS

In this chapter we outline an example implementation of the methodology that was proposed in chapter 3. Having an implementation is necessary for the empirical evaluation that we do in chapter 5. We begin by setting a few design goals and introducing a generic system architecture for data analysis. This system architecture is then used for an example implementation. The chapter is concluded with a brief discussion.

4.1 Design goals

In this section we set a few design goals for the implementation and discuss about what a data analysis software should, at least from our perspective, look like. Currently no implementation can completely replace the data analyst in data exploration but when cleverly constructed a computer implementation can be of great assistance and can have a dramatic boost on the data analyst's productivity. For this purpose we consider things like ease of use, interactivity, modifiability, portability, and respectability in software development. As these concepts are intricate in nature we content ourselves to presenting their main ideas in an abstract level.

Ease of use. Ease of use has become a kind of cliché in software engineering as every software developer eagerly states that their software is easy to use. Whereas this is seldom true according to the users, we try to formulate a good definition for ease of use in away that we are in practice able to accommodate. In our opinion a scientific software is easy to use when it minimizes the amount of details that the user has to remember and transparently automates some of the most laborious tasks that are involved. This does not mean that everyone should be able to use the software without bothering to browse the user manual beforehand but it does mean that when the user reaches a certain level of experience he feels comfortable working with the software. This very abstract but insightful definition will be the most important guideline in our design.

Interactivity. An important part of the usability, especially when exploring new data sets, is to enable and to encourage the user to work in an interactive manner. This means that rather than giving all instructions at once and then running the whole analysis in a batch process, the user should be able to make small decisions and be allowed to come back to change his mind along the whole analyzing chain. To enable such working habits the software should be constructed from small functional blocks that the user concatenates to create larger analyzing chains. When the user sees he has made an error in judgement he should be given the opportunity go back to make changes to his previous choices. Such an approach allows the user to easily tinker with a data set and at the same time to gain a better understanding of this data set.

Modifiability. An another important part of usability is application modifiability. From continuous use of the software the user is likely going to get new ideas that could improve the functionality of the software even further. For the added value from these ideas to realize the user must be able to make his own modifications to the functionality of the software. Most of the current software that is in use, even those that support recording macros or ship with full source code, are in effect not modifiable. Gladly some of the scientific software make an exception with their high level scripting languages that are used to implement application specific functionality. We want to see more software that allows the user to modify any part of the software functionality when he feels it necessary.

Portability. Whereas the general public tends to prefer using the de facto operating platform, which at the time of writing is based on Intel hardware and Microsoft software, the researchers constitute a more heterogeneous group of users that work with the whole range of available platforms. Some of them feel strongly about the platform that they have become accustomed to and are thus very reluctant to making any changes that would inconvenience their daily routines. Hence portability across multiple platforms must be emphasized more with scientific software than it is currently done in software engineering. Our objective is to make our implementation an example of good portability so that switching to a new platform requires recompilation of a few libraries only. To define a concrete objective we aim for Microsoft Windows, Apple MacOS, and Linux.

Respectability. No respectable scientist is going to use any piece of software unless they feel completely comfortable that the software meets certain quality requirements on proper functionality. Either this comfort raises from the fact that they have implemented the software themselves and thus know exactly how it works, or from a well established position that the software has in their respective research society. Outside the field of information technology the latter justification one is prevailing and if possible this should be taken advantage of.

4.2 A generic system architecture for data analysis

In this section we present a generic system architecture that can be used for applications in data analysis. This architecture is used for the presented example implementation. The design goals of the previous section set a high standard for the architecture. In practice achieving all these goals is difficult. Although there is always room for improvements, the system architecture that is presented in this thesis is very close to meeting all the stated design goals. The presented architecture consists of six main components: a core for computations, extension libraries, an application logic, user interfaces, a user, and a data storage (see Figure 20). Each of these components has a specific role as described in the following.

R core provides a high level macro language with operations for naming, storing, and manipulating data objects such as vectors, matrixes, and data frames (for a detailed description see (The R Foundation 2005a)). Whereas any high level com-

puting core would do, we have chosen to use R because in our eyes it has many considerable advantages: R is versatile enough to be suitable for all types of data analyzes, R is mature and portable enough for our purposes, R is distributed as a free software with full source code, and R has established a solid position in statistical community. A brief introduction to R is given in section 4.2.1.

Extension libraries are used to extend the basic functionality of the R core with new operations that are stored in dynamically loadable software libraries. These libraries are programmed using compiled languages like C and Fortran. Because the libraries are compiled to efficient system code, which can sometimes yield quantum leaps in execution speed when compared to interpreted macro language implementation, the developed operations function as effectively as the internal operations of the core. In fact the distinction between the operations of the core and the operations that are loaded from extension libraries is negligible, and operations of both types are used almost exactly the same way. An introduction to author's extension libraries, which we need for the example implementation, is given in section 4.2.2.

Application logic uses the services of the R core and of the extension libraries to perform application specific tasks as instructed by the user through one of the user interfaces. In the presented system architecture the application logic is always implemented using R macro language and only computationally heavy operations are implemented in the extension libraries using C language. From our experience of developing various scientific applications we have found this approach to be the most productive as it minimizes the burden of development. A few lines of macro code typically corresponds to hundreds of lines of C code and the induced computational overhead is often negligible. An additional benefit of this approach is that modifying and porting applications becomes much easier.

User interfaces form a bridge between the users and the application logic. The requirements for the user interfaces vary heavily between different users and different applications. Hence there is a clear need to allow the system to be used through multiple user interfaces each of which fulfils a particular need. In the presented architecture the most generic interface is the command prompt interface, which is provided by the R core. This user interface is always available and it provides a direct access to all data objects and operations in the system. Hence this user interface is good for developers and advanced users who already know what they are doing and want to do it in the most effective way. Unfortunately the command prompt interface is also far too complicated for the end users who need the software to steer them in order for them to successfully use the system. For such users a graphical interface that contains a limited set of operations that are tailored to fit the exact needs of the application is usually the best approach.

User plays an important role in the system. The user must place experiment data, which contains the results from an experiment, and metadata, which describes the contents of the experiment data, in an appropriate format to a predefined data storage, and then use the services of the system to analyze this data. The role of the user is to incorporate cognitive reasoning to otherwise mechanical data processing.

This is essential for making proper conclusions about causalities between the studied phenomena and interpreting the obtained results from the application perspective.

Data storage is used to keep experiment data, metadata, and system settings when the system is not in use. The organization of the data and the used file formats are application dependent and become fixed through the implementation.

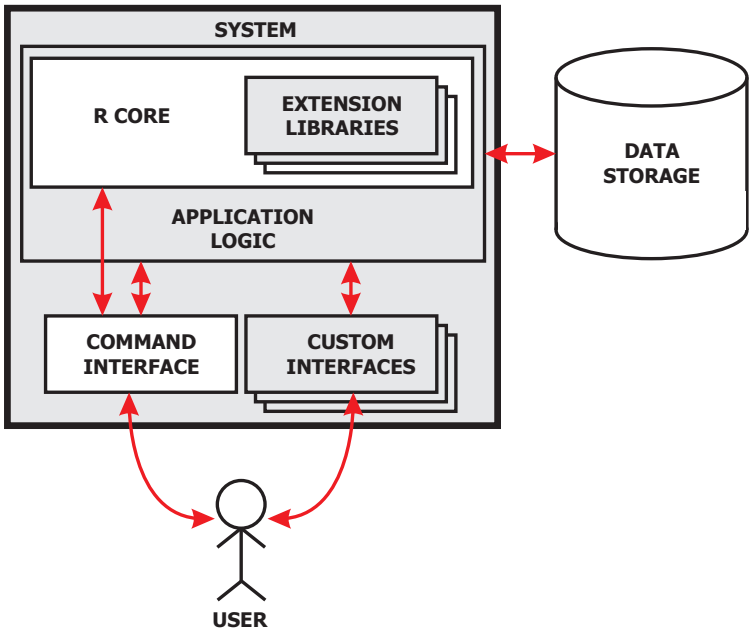


FIGURE 20: An example system architecture for data analysis.

4.2.1 Introduction to R language and environment

R is the name of a language and of an environment that are targeted for statistical computing (see `www.r-project.org`). It is a free software that is similar to commercial S language and environment, which were developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and his colleagues. Despite the fact that the two languages are almost identical, that is the macros that are written using one language tend to meet the requirements of the other language as well and hence macros are usually portable between the two environments, it is important to understand that the two environments are developed by two independent organizations and their internal designs are hence also different. The most important contributors in the R project are Douglas Bates, John Chambers, Peter Dalgaard, Robert Gentleman, Kurt Hornik, Stefano Iacus, Ross Ihaka, Friedrich Leisch, Thomas Lumley, Martin Maechler, Duncan Murdoch, Paul Murrell, Martyn Plummer, Brian Ripley, Duncan Temple Lang, Luke Tierney, and Simon Urbanek.

R is a self-sufficient statistical environment that provides tools for data analysis, modeling, statistical inference, and graphical visualizations (see Figure 21). The R macro language is designed to be a true computer language that allows users to develop new operations by defining new functions. The language syntax knows

the most common data structures, like vectors, matrixes and data frames, and is able to directly perform many algebraic operations between these structures. More complex operations are implemented as functions. The list of existing operations in the current software version include clustering, classification, linear and nonlinear modeling, and statistical testing to name a few.

Despite being a free software that depends on contributions of voluntary programmers, most of the documentation that is covering the R macro language, the R environment, and the extension libraries are of good quality. This documentation is available from R command prompt, as hypertext that is shown through a web browser, and in easily printable portable document format (PDF). An interested reader should start by browsing (The R Foundation 2005a), which gives a brief description of the R language with examples, and then continue reading (The R Foundation 2005b), which describes how new operations can be implemented using C language. Based on the author's own experience for a person that is familiar with programming principles it is possible to learn to use R in one day.

The newest version of R can be downloaded from the project website `http://www.r-project.org/` under the terms of the Free Software Foundation's GNU General Public License. The software is made available in source code and in compiled binaries. List of supported operating systems include Microsoft Windows, Apple MacOS, and most distributions of Unix with X11.

4.2.2 Introduction to author's extension libraries

When the operations of the R core are not enough to accommodate application needs, new operations can be implemented either in R macro language or in C, C++, or Fortran languages. These operations are assembled into libraries from which they can be dynamically embedded to the resource pool of the R core. The libraries that have reached a certain level of maturity are usually bundled into R packages, which are complete operational entities that contain the library, documentation, licensing information, and possible some example data sets. Sharing a package is made easy as there is a well functioning distribution network (for details see `http://cran.r-project.org/`). The author has designed and implemented a few packages that address some of the shortcomings of R. All of the libraries that are presented here are made portable to Microsoft Windows, Apple MacOS, and Fedora Linux with X11 server.

Digital image processing library complements the R core with the most common operations that are used in image analysis. R has no native support for working with images and hence without an extension library the applications of this type are subject to poor execution speeds. The author's library provides the most common image processing operations ranging from geometric transformations and filtering to feature extraction and beyond. As images are stored in R matrixes the user has a direct pixel level access from the R macro language to the images at all times, and hence mixing operations that are implemented using the macro language

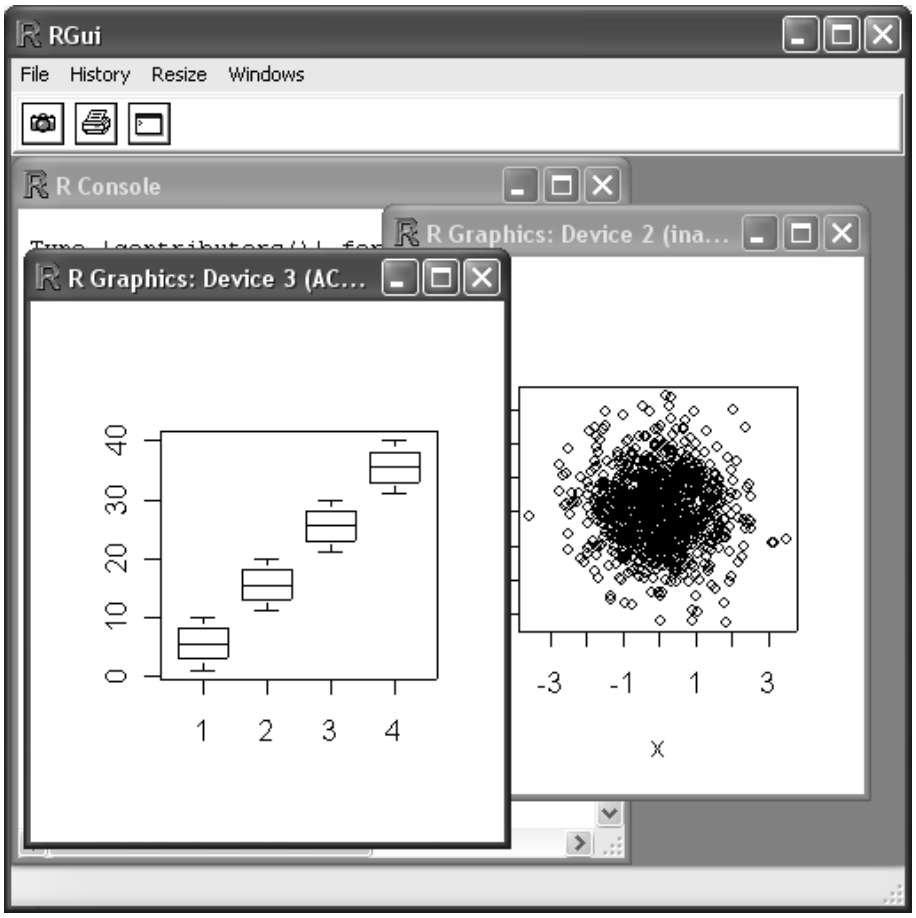


FIGURE 21: Example views to demonstrate the graphical interface of the R environment.

with those from the library is easy.

Graphical user interface library allows the R macro language to be used to build interactive graphical user interfaces (GUIs). Interactive GUIs are valuable in data analysis because effective visualization is sometimes indispensable in order to gain good understanding of the results of analyzes. The R core provides good tools for drawing simple statistical graphs such as scatter plots and histograms. However, support for bitmap graphics is somewhat lacking and there is no way to make new user interfaces that are tailored to the needs of the application. The guiding principle of the author's library is that implementing GUIs should be as easy as implementing the application logic so that the GUI can evolve in the same pace with the functionality of the application logic. In most scientific software this is not the case, actually far from it, and hence a thorough planning is required prior implementation. Whereas the author's library cannot remove the need for good planning, it manages to reduce the amount of work that is involved. A typical application tailored GUI requires only a few hundred lines of R macro language and rewriting these should be a matter of few hours of work at most.

Multithreading library provides an easy way to take advantage of the multi-

core central processing units that are becoming popular in current workstation and server designs. The author's library offers POSIX style multithreading and synchronization services to be used directly from the R command prompt. Unfortunately R is not completely thread safe, there are a few operations that cause the execution to hang in case of two or more simultaneous calls, and hence a case by case study is needed to make sure that the performed operations are working properly. This is an unfortunate shortcoming in the design of R that will hopefully be corrected in the forthcoming releases. Based on empirical experimentation that were limited to thread safe operations the automatic multithreading of bootstrapped procedures using this library offers clear advantages for computationally intensive applications.

4.3 Example implementation of the proposed methodology

In this section we review a brief description of an example implementation that uses the system architecture of the previous section to implement the methodology that was proposed in chapter 3. The implemented system provides a working environment in which a data analyst, hereafter referred as the user, can conduct analysis for detecting and measuring spatial dependencies between two or more images. Performing analyzes with this system consist of five stages that are illustrated in Figure 22.

Preceding activities are performed before the user can start using the implemented system. These activities include storing all experiment data and metadata in a system specified format. In the example case the experiment data consist of paper measurement images and metadata define pixel coordinates of the markers that are used in geometric transformations of the images, and the image areas that are rejected from the analyzes.

Image processing includes operations that take an image and return an another image as a result. Some of the implemented image processing operations were described in chapter 2. The implemented system provides operations for

i) opening measurement images from data files (storing processed images is also possible),

ii) loading metadata that restricts analyzed image areas and describes marker pixel coordinates,

iii) performing geometric transformations for obtaining pixel correspondence between the studied images (as presented in section 2.2.1),

iv) performing trend removal for equalizing measurements over different image areas (as presented in section 2.2.2),

v) image mean and median filtering for measurement noise removal (as presented in section 2.2.3),

vi) extracting stochastic and deterministic parts of measurement images for modeling purposes (this is an application specific operation that is presented in section 5.2),

vii) extracting observations and converting them into features (as presented in section 3.2),

viii) and visualizing the result after each of the previous operations (this feature is available in the graphical user interface only).

The parametrization of these operations is not fixed and the user can interactively build the best image processing chain for each studied image. All the image views can be linked together in such way that changing the visible area in one of the image views will automatically update the other image views so that the same image area is always visible in all of the image views. With this feature it is easy to focus on different paper properties of the same paper area.

Model building is an operation chain that takes observations from two images and returns an estimate of the proposed CCA correlation between these two images. The implemented system provides operations for

i) building a canonical correlation model between the observations of two images (as presented in sections 3.3 and 3.4),

ii) visualizing the canonical correlation model (as we have done in Figure 11 and in Figure 12),

iii) and interactively grouping the observations in image space (from an image) or in model space (from a scatter plot) using the pointer (this feature is available in the graphical user interface only).

Again the parametrization of these operations is not fixed and the user can interactively build multiple models to see how they describe the studied dependency. When a model is built the user can divide the observation points into groups using the pointer. In the scatter plots these groups are visualized with different point colors, and in the image views the observations are indicated with subimage borders that are drawn over the images with the group color. One example of how to use this feature is to select a set of outlier observations from a scatter plot and then see how they are spatially distributed in the images. The grouping is shared between all model and image views and hence selecting interesting observations from one model will also show how they spread in the other model spaces. Similarly the user can select and group observations from a particular image area using the pointer. This way he can see for example how visually observable anomalies project back to model space. We have found that moving back and forth between the image and the model representations gradually improves the understanding of the models and the underlying phenomena.

Analysis operations use image processing and model building to provide an answer to a specific question. A typical question contemplates how the observed spatial dependency changes when we change one of the two studied images or when we keep the two images but change one of the parameters in image processing or in model building. Examples of possible analyzes are provided in section 3.6 and more concrete analyzes are demonstrated in chapter 5.

Interpretation of the results is based on graphical visualizations and on numerical statistics that are provided by the implemented system. The actual interpretation of the results is not performed by the system but is instead based on cognitively reasoning of an application expert. In our application this is the only feasible way to merge understanding that is coming from empirical experimentation and modeling of the data to the prior knowledge and understanding of the application engineer. A hands-on example of how to interpret the obtained results in a real world case is provided in chapter 5.

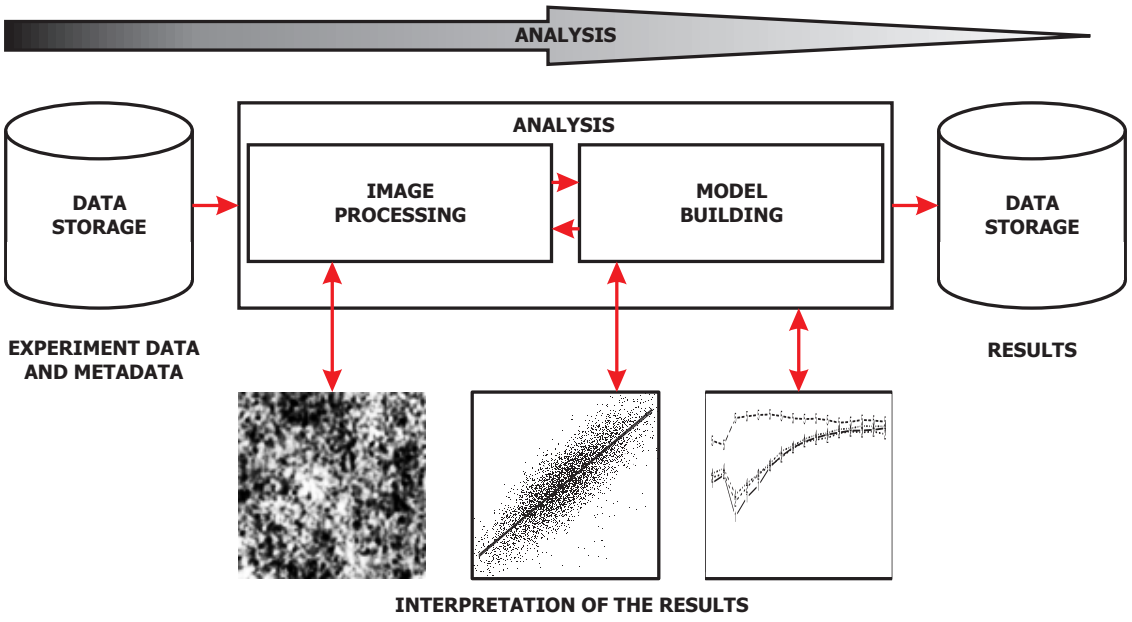


FIGURE 22: A schematic of workflow in the example implementation.

The implemented system provides two user interfaces. The first one, which is suitable for developers only, is really a package of macro files. The user must edit the macro files with an editor and then run them from the R command prompt. Naturally the level of interactivity is small but on the other hand modifying the system requires less expertise and this approach is also good for batch processing. The second one, which is more suitable for end users, is a graphical desktop on which the user can open images, build models, and execute analyzes. According to the established GUI design rules all user actions are commanded from a menu that is located on top of the desktop. Similarly on the bottom of the desktop there is a status bar through which the user is informed on the actions that he is currently using. Remainder of the desktop is reserved for a multiple document interface (MDI) where

a new window is opened for each image, model and analysis. An example workspace with multiple images and models is shown in Figure 23. The system keeps track of the user's actions and allows the current state of the workspace to be saved to and loaded from a file. This way the user can save his work and continue on working from the same state at a later date. When working this way, it should be noted that the workspace contains only references to the original measurement data that must hence also always be made available. This way the workspace structure is very light to use and does not consume large quantities of disk space.

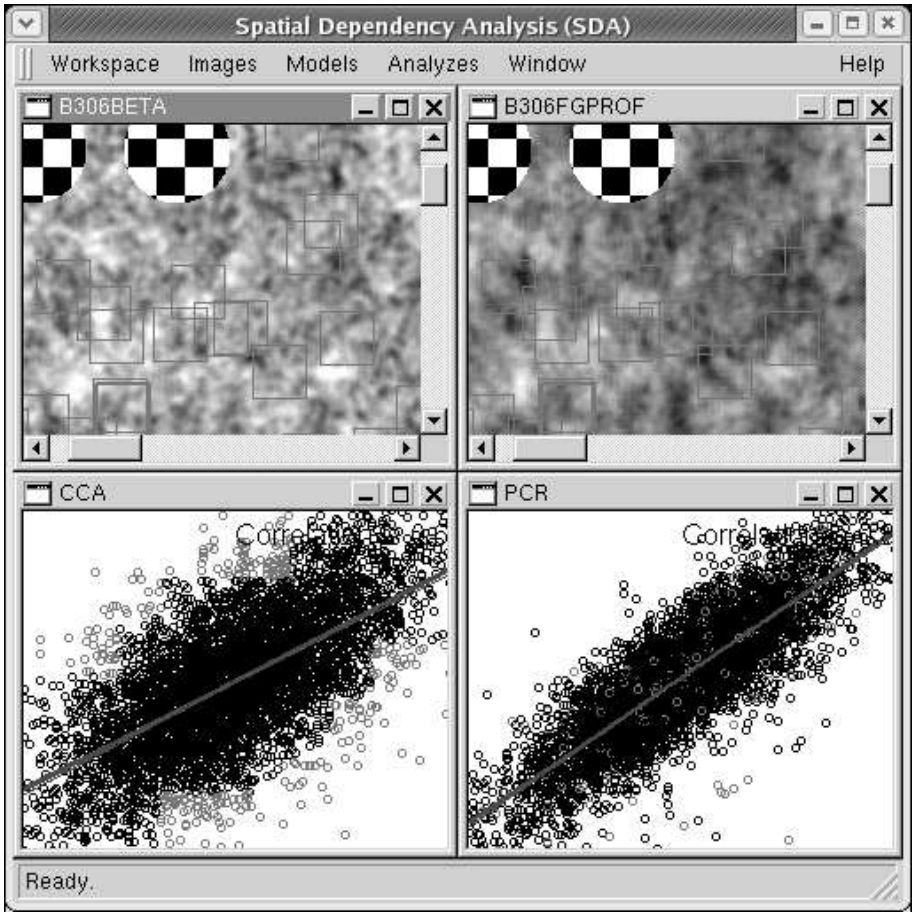


FIGURE 23: Implemented graphical user interface.

4.4 Discussion

In this chapter we have gone over some design goals, an example system architecture, and an example implementation for the methodology that was proposed in chapter 3. The presented implementation is a necessary part of our evaluation of the methodology because a computer tool is the only feasible way to use the proposed methodology for a real world application. The presented system architecture seems very generic and suitable for data analysis in the fields of statistics and computational intelligence.

5 EXPERIENCES FROM THE PAPER APPLICATION

In this chapter we present an example of data analysis on our paper measurement images. We start with a brief introduction to conducted measurements and how the images were preprocessed. The presented analysis begins with standard geostatistical methodology, which is used to gain an understanding of the content of the paper images. This understanding is important for the proper use and evaluation of the proposed CCA correlation methodology, which was proposed in chapter 3. Based on the results from CCA correlations we are able to answer most of the questions that we have from methodological and application perspective. Here is a list of the most important findings from methodological perspective:

- It is possible to understand the role of spatial scale of the structures inside an image.
- Modeling with Gaussian random fields allows a compression of all essential information from an image into a few scalar values. Simulating new images using these models provides examples of natural process variation.
- Empirical correlograms and cross-correlograms provide an easy way to make a tentative study of the spatial correlations within and between images.
- Empirical correlations from the proposed CCA methodology complement the results of the tentative study.
- Additional visualizations provide spatial descriptions of the structures that are behind the observed correlations.

This list is continued with the most important findings from the paper application perspective:

- There are clear differences between the studied smooth and rough paper types.
- The amount of observed dependency depends on spatial scale. In case of our paper sheets the highest correlations are seen in $2.5mm$ - $4.0mm$ spatial scale.
- Estimated thickness distribution explains the paper structure better than either of the surface topographies.
- Making reliable conclusions with respect to density distribution is problematic based on the conducted experiment.
- Thickness distribution is the best paper characterizing property in the studied context. Mass distribution is the best paper characterizing property that can be measured directly.

5.1 Introduction to the paper images

A paper sheet is essentially a realization of a real world stochastic process that exists in paper manufacturing. This process is challenging to model because of the difficulties involving measurements and because of the complexity of the paper structure. However, it is possible to characterize the properties of this process based on observations that we can make from an image of a paper sheet. This way the problem of understanding a paper sheet can be divided into smaller and more manageable subproblems, which we tackle in the following sections.

In order to understand the structure of a paper sheet we need to obtain two-dimensional spatial maps of various paper properties. In our study we use images that have 1024×1024 pixels to represent approximately $10cm \times 10cm$ paper area. We have images of background and foreground topography, along with mass distribution, which is commonly referred as formation, and computationally obtained thickness and density distributions (recall the exact problem setting and experiment from section 1.1). Omitting errors and noise, in these images small pixel values, where pixels are dark, correspond to small measurement values. For the topographies dark implies that the pixel is in a topographical hole or a valley. In case of the distributions dark indicates that the paper is light, that the paper is thin, and that the paper is sparse respectively. Analogously we have that large pixel values, where pixels are light, correspond to large measurement values and hence indicate that the pixel is in a topographical hill or a hill range, and that the paper is heavy, thick and dense respectively.

Conducting measurements with more than one instrument generates images that are not in the same scale and position. In our application this problem, commonly referred as the pixel correspondence problem as introduced in section 2.2.1, raises from the fact that in practice it is impossible to align the measured paper sheets exactly in the same way with all the measuring instruments. In addition the resolution of the measurements may vary between the instruments. From our methodology we have the precondition, recall section 3.2, that those images that measure different properties of the same paper sheet must have all the pixel measurements from the same point of the paper placed at the same pixel coordinate. Hence all the raw images are geometrically transformed to accommodate this requirement. Unfortunately it is not possible to obtain full pixel correspondence, which is likely to affect the obtained results. The reasons for this were discussed in section 3.6.

The most common approach to measure paper surface topographies is to use laser profilometry. This approach is discussed in (Banerjee, Bradley and Gelfand 2004, pp. 661-698) and (Borch, Lyne, Mark and Habeger 2002, pp. 429-450). In our study we have measurement images

$$z^{background} = \{z^{background}(x, y) : (x, y) \in D\} \text{ and}$$
$$z^{foreground} = \{z^{foreground}(x, y) : (x, y) \in D\}$$

for background and foreground topographies respectively. The terms background and foreground are used to indicate the intended printing side, which is considered to be the foreground. Whereas we do not study printing related issues in this work, making a clear distinction between the two sides of the paper is necessary as there are papers that do not have similar topography characteristics on both sides of the paper. An example of surface topography and its pixel intensity histogram are presented in Figure 24. From this resolution we cannot see individual fibers but we can clearly see that the surface is unevenly distributed as the surface structure contains hills and valleys. Closer inspection reveals that there is also an apparent grid structure, which may or may not be visible in different parts of the image. The pixel intensities seem to have a near Gaussian distribution as can be seen from the pixel intensity histogram. The same surface topography but now in frequency domain is presented in Figure 25. Except a few spikes that are discussed later, the power spectrum of the image seems close to circular which lead us to conclude that the surface topography is isotropic, which means that the paper is similar to all directions. In practice papers with anisotropic surface topographies are also common. We can also see sharp frequency spikes around the centered DC peak. These frequency spikes correspond to the grid pattern that we saw in the spatial domain. Although there is some literature on the subject, it is not the place of the author to speculate whether these frequency spikes represent the wire pattern or some other reflection of the manufacturing process. The phase angle of the image seems to consist mostly of random shot noise. Based on the pixel intensity distribution, the power spectrum, and the phase angle we assume that the surface topography follows approximately the stochastic laws of GRFs.

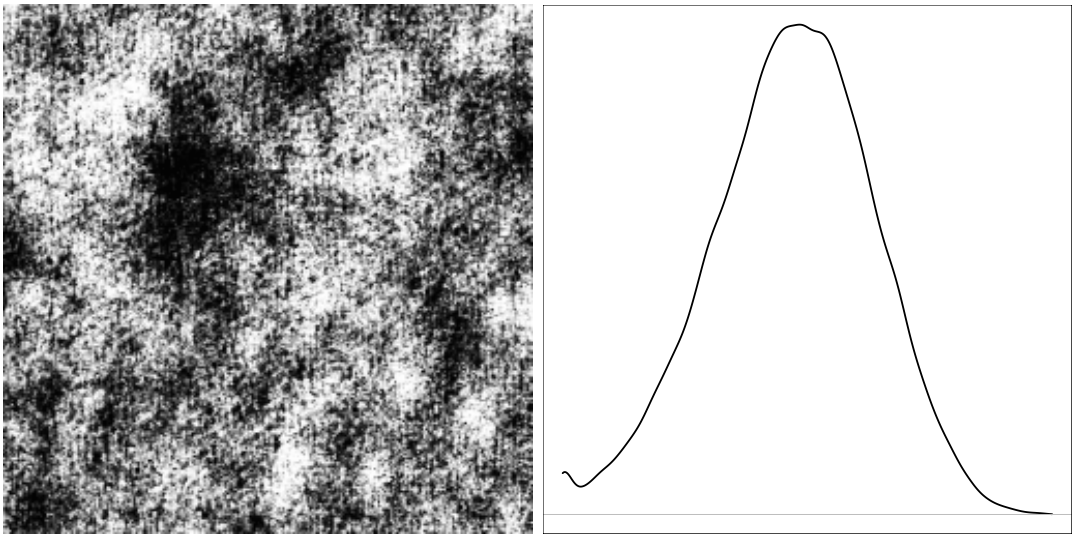


FIGURE 24: Example of surface topography from a $2.56cm \times 2.56cm$ area of rough paper type sheet B1 shown in spatial domain: measured image (left) and corresponding pixel intensity distribution (right).

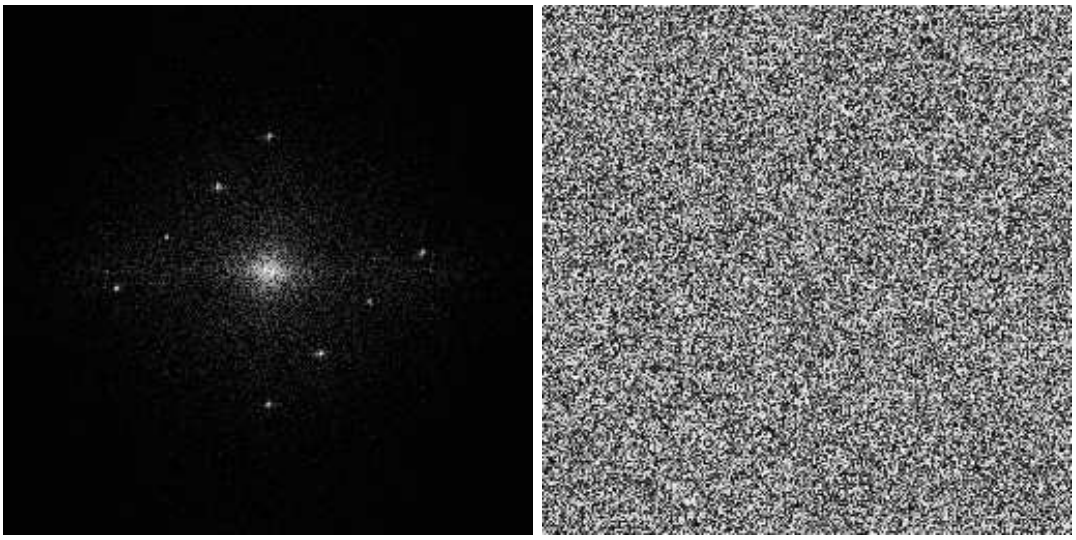


FIGURE 25: Example of surface topography from a $2.56cm \times 2.56cm$ area of rough paper type sheet B1 shown in frequency domain: power spectrum (left) and phase angle (right).

Distribution of mass in a paper sheet can be measured with β-radiography. For an introduction to use of β-radiography in paper science the reader is advised to see for example (Banerjee, Bradley and Gelfand 2004, pp. 661-698). The measurements are based, at least in our case, on a well known absorbtion law: β-rays are known to attenuate and absorb in proportion to the total mass they pass through. The composition of this mass does not affect the rays and we can model the absorbtion through formula

$$T = e^{-\mu w},$$

where T is a transmission factor, μ is an absorbtion coefficient that is constant for all known paper components, and w is the basis weight, which is the amount of mass that the beta rays have to penetrate. The wavelength of β-radiation is so large that the rays do not scatter while passing through the paper sheet. With appropriate scaling we obtain a mass distribution image

$$z^{mass} = \{z^{mass}(x, y) : (x, y) \in D\}.$$

An example of mass distribution and its pixel intensity histogram are presented in Figure 26. This resolution is again insufficient to distinguish individual fibers but the mass distribution has similar hills and valleys that we already saw in the surface topograph. Based on visual observations the surface topography and the mass distribution share similar kind of large scale structure, which concurs with our intuition. An another observation is that the grid structure, which we saw in the surface topograph, is not clearly visible in the mass distribution. The pixel intensity histogram seems to follow the shape of the Gaussian distribution. The same

example of mass distribution in frequency domain is presented in Figure 27. Again the power spectrum is approximately circular implying that the mass distribution is isotropic in this case. The grid inducing frequency spikes are also visible but with lower intensity. This indicates that, albeit we did not see the grid in the spatial domain, it is still present. The phase angle consists mostly of uninformative noise. Although the phase angle is not totally random we can argue that based on the pixel intensity histogram, the power spectrum, and the phase angle it seems that the mass distribution follow loosely the laws of GRFs.

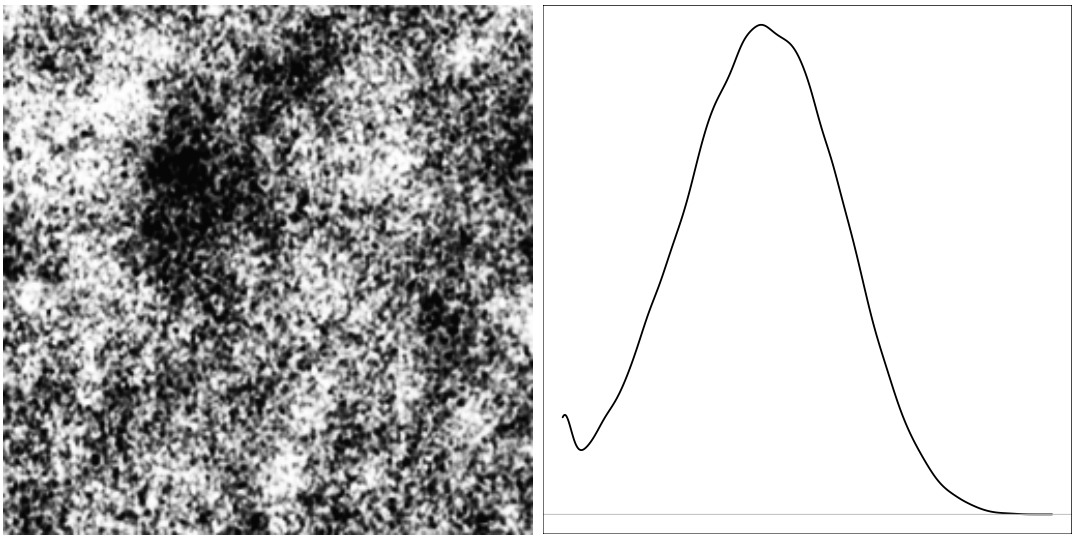


FIGURE 26: Example of mass distribution from a $2.56cm \times 2.56cm$ area of rough paper type sheet B1 shown in spatial domain: measured image (left) and corresponding pixel intensity distribution (right).

In our study we use the surface topographies and the mass distributions to generate additional computational measurements. These computational measurements give valuable insight to the paper structure but we must always understand that some findings that we make may be artifacts of the computations instead of actual physical relations that exist between the paper properties. If we know, and usually we do, the mean thickness of the measured paper sheet, we can also obtain an estimate for the paper thickness distribution by adding centered surface topographies to the mean thickness (see Figure 28). In analytical formulae, we have

$$\begin{aligned} z^{thickness}(x,y) = z^{thickness}_{mean} \\ &+ (z^{foreground}(x,y) - \bar{z}^{foreground}) \\ &+ (z^{background}(x,y) - \bar{z}^{background}) \text{ for } (x,y) \in D, \end{aligned}$$

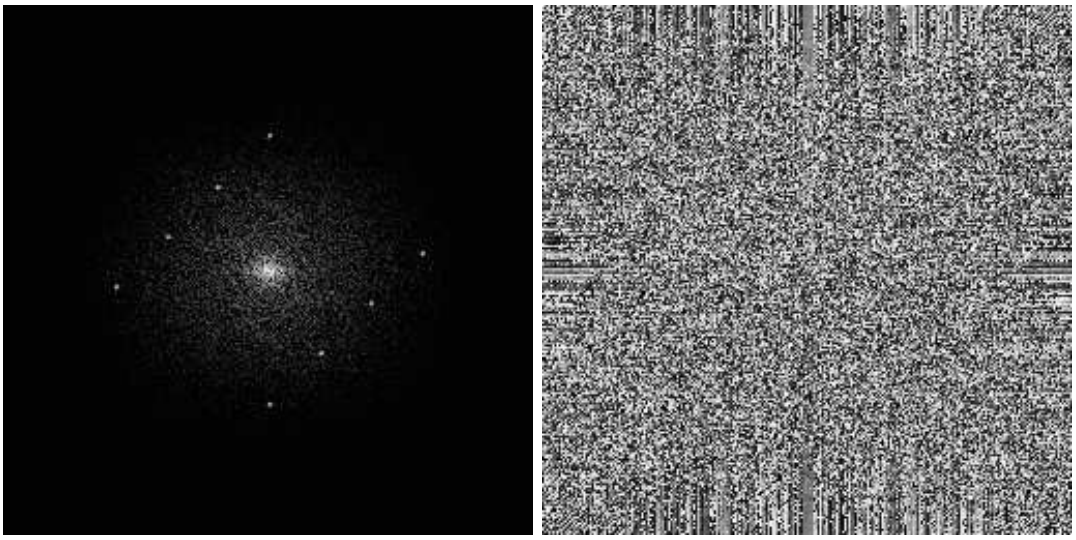


FIGURE 27: Example of mass distribution from a $2.56cm \times 2.56cm$ area of rough paper type sheet B1 shown in frequency domain: power spectrum (left) and phase angle (right).

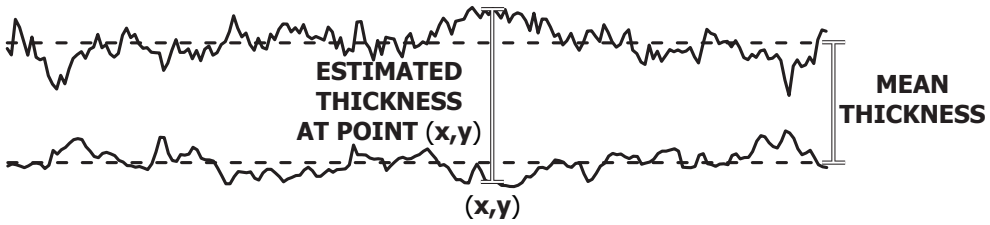


FIGURE 28: Obtaining an estimate for thickness distribution based on background and foreground topographies.

where $z_{mean}^{thickness}$ is the mean thickness of the paper and

$$\bar{z}^{foreground} = \frac{\sum_{(s_x,s_y)\in D} z^{foreground}(s_x, s_y)}{w_{image} h_{image}} \text{ and}$$
$$\bar{z}^{background} = \frac{\sum_{(s_x,s_y)\in D} z^{background}(s_x, s_y)}{w_{image} h_{image}}.$$

As there may be some ill measured pixels in the topographies and as the thickness cannot be negative, we threshold the thickness pixel values between zero and some paper type specific maximum. From the mass and the thickness distributions we can further estimate the density distribution by dividing the local mass by the local thickness. In analytical formulae, we have

$$z^{density}(x, y) = \frac{z^{mass}(x, y)}{z^{thickness}(x, y)} \text{ for } (x, y) \in D. \tag{16}$$

Again there may be some poorly measured pixels and thus we threshold the density pixel values between zero and a paper type specific maximum.

For research purposes we have obtained measurements from eight paper sheets, four paper sheets of smooth paper type (sheets A1, A2, A3, and A4), and rough paper type (sheets B1, B2, B3, and B4). During the measurements one paper sheet was damaged and thus the smooth paper type actually has only three viable specimens. The main difference between the two paper types is in the manufacturing process: the smooth paper type is calendered (roughness Bendtsen $45ml/min$) whereas the rough paper type is coated (roughness Bendtsen $180ml/min$). Example measurements of the two paper types are presented in Figure 29 and Figure 30 respectively. From these we can see how all the others, except the density, look very similar. The density distribution seems to have smaller grains and to be more independent of the other paper measurements.

All the measurements that were successfully completed, which is all but sheets A4, seem to be without significant errors. Four of the most common errors, which are not serious, are illustrated in Figure 31. The first image shows a patch that is likely a result from a paper wrinkle. This wrinkle is present in all the sheets of the smooth paper type (sheets A1, A2, and A3). The second image shows a patch which origin cannot be determined. The images of the smooth paper type all have multiple instances of such patches. The third image visualizes the type of stripes that are present in the rough paper type sheet B1. The fourth image illustrates the type of scan line measurement errors that are present in the rough paper type sheet B2. A summary of observations from all the images is presented in Table 5. As a conclusion we can state that unexplainable patches are present in all images of the smooth type (sheets A1, A2, and A3) with the exception of mass distribution images. In the rough paper type (sheets B1, B2, B3 and B4) we have clearly visible grid structure in all the foreground topographies. For the rough paper type all the mass distributions and their derivative density distributions show clear trending, which was removed with the algorithm that is presented was section 2.2.2. All of these and other visually observable errors are carefully marked and removed from further analyzes (see Table 4). Hence the presented results are not affected by these errors.

1. Obtain pixel correspondence with geometrical transformations.
2. Rescale to obtain zero mean ($\hat{\mu}_Z = 0$) and unit variance ($\hat{\sigma}_Z^2 = 1$).
3. Decompose into stochastic and deterministic components (for explanation see the beginning of section 5.2).
4. Remove trending with median polishing.
5. Reject invalid image areas from taking part to the analyzes.

TABLE 4: An overview of the image preprocessing chain.

	GRID	TREND	PATCHES	STRIPES	OTHER
SHEET A1					
Background			X		
Foreground			X		
Mass					
Thickness			X		
Density			X		
SHEET A2					
Background			X		
Foreground			X		
Mass					
Thickness			X		
Density			X		
SHEET A3					
Background			X		
Foreground			X		
Mass					
Thickness			X		
Density			X		
SHEET B1					
Background				X	
Foreground	X				
Mass		X			
Thickness					
Density		X			
SHEET B2					
Background					X
Foreground	X				
Mass		X			
Thickness					X
Density		X			X
SHEET B3					
Background					
Foreground	X				
Mass		X			
Thickness					
Density		X			
SHEET B4					
Background					
Foreground	X				
Mass		X			
Thickness					
Density		X			

TABLE 5: Four of the most common errors in the paper images.

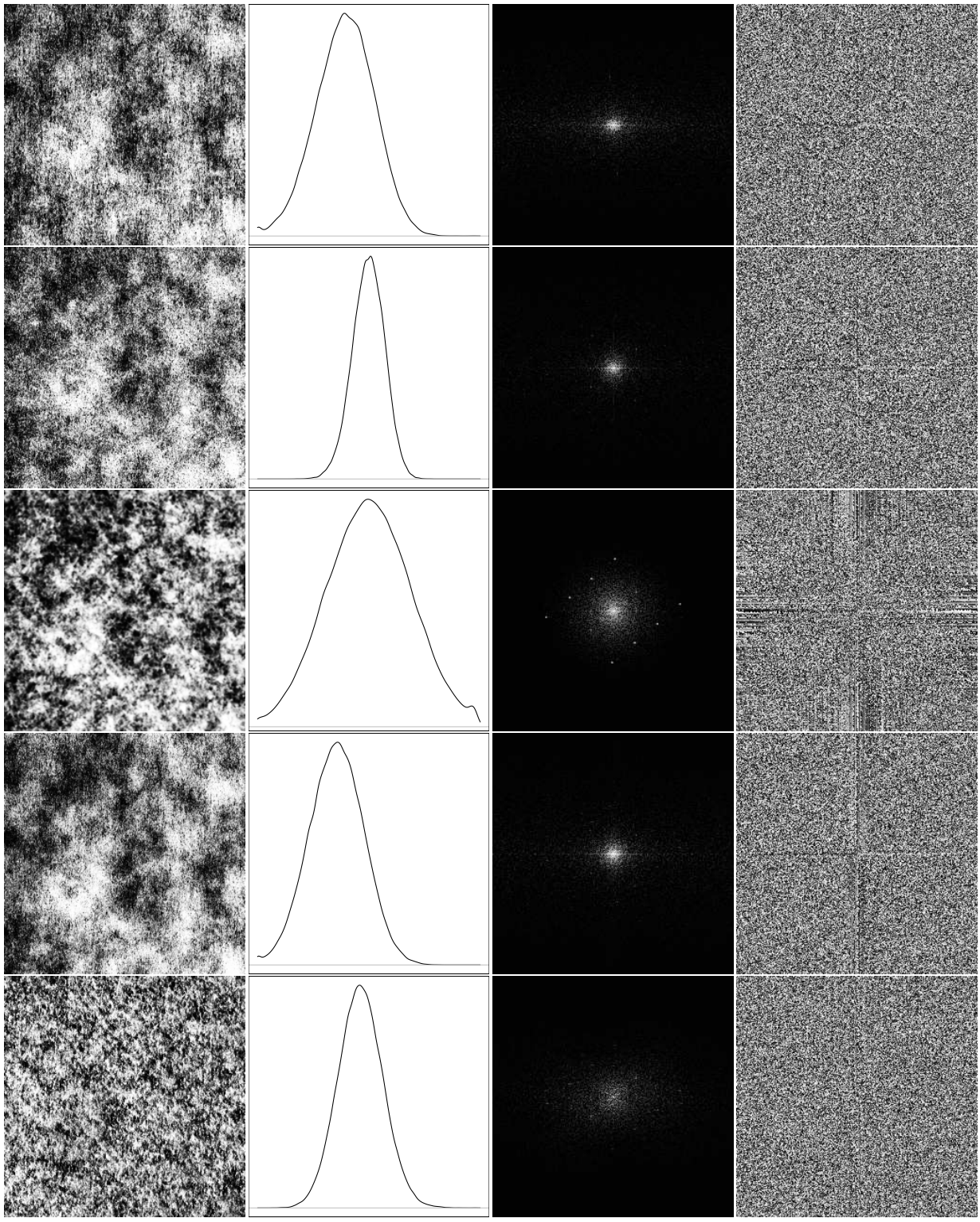


FIGURE 29: Examples of measurements from a $2.56cm \times 2.56cm$ area of smooth paper type sheet A1. Columns: measured image, pixel intensity histogram of the image, power spectrum of the image, and phase angle of the image. Rows: background topography, foreground topography, mass distribution, thickness distribution, and density distribution.

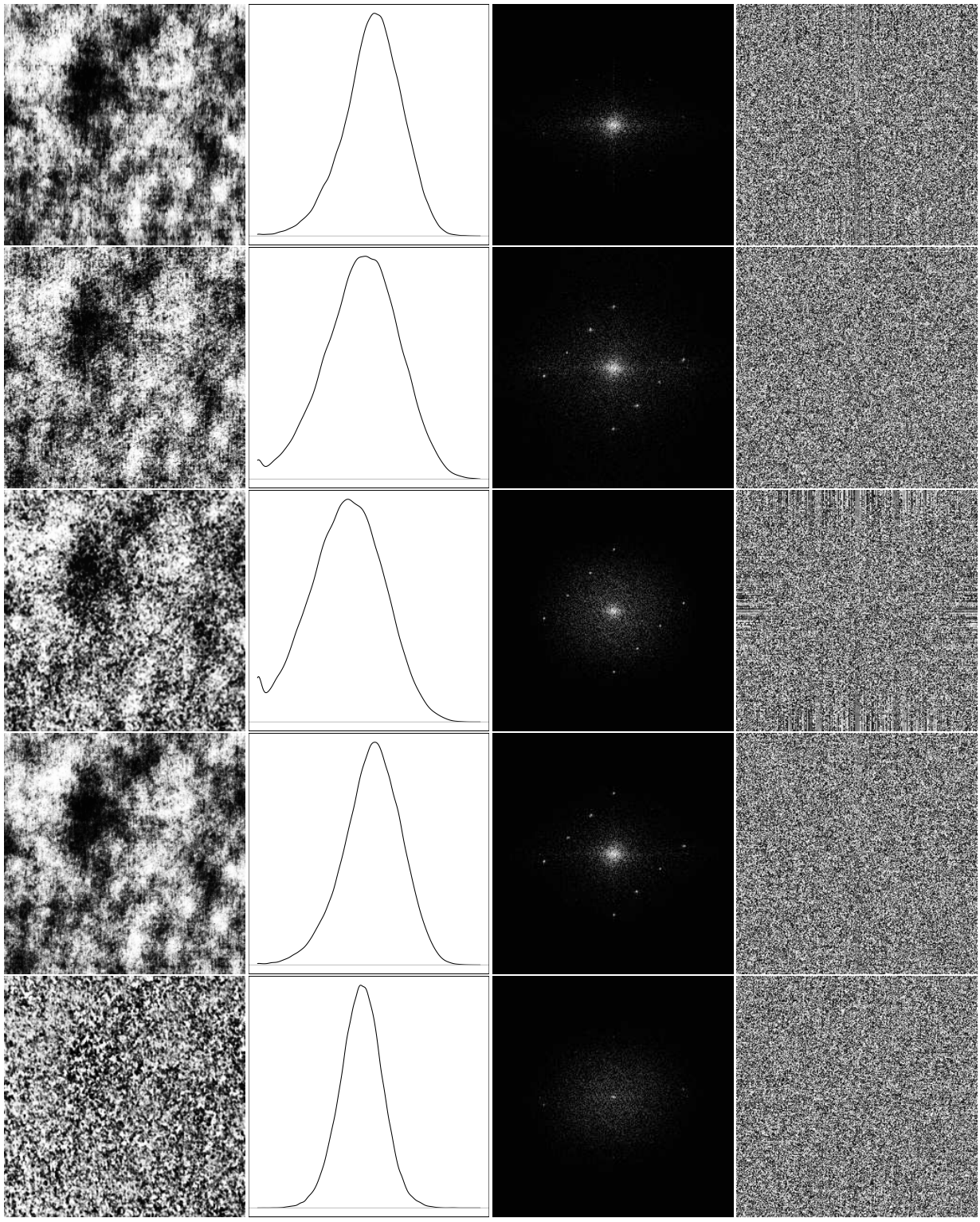


FIGURE 30: Examples of measurements from a $2.56cm \times 2.56cm$ area of rough paper type sheet B1. Columns: measured image, pixel intensity histogram of the image, power spectrum of the image, and phase angle of the image. Rows: background topography, foreground topography, mass distribution, thickness distribution, and density distribution.

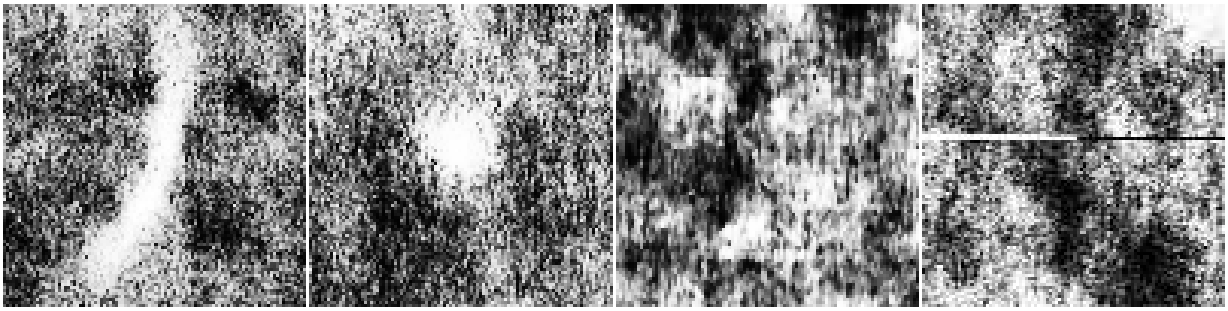


FIGURE 31: Typical errors from $1cm \times 1cm$ areas: a patch from a paper wrinkle in the foreground topography of sheet A1 (left), a patch from an unknown source in the foreground topography of sheet A1 (center left), stripes from the measuring instrument in the background topography of sheet B1 (center right), and measurement errors from the measuring instrument in the background topography of sheet B2 (right).

5.2 Parametric modeling of the paper images

In order to employ GRF modeling to our paper images we need to remove all deterministic structures as they cannot be explained by this family of models. The foreground of rough paper type sheet B1 contains a clearly visible grid structure and we use it to illustrate the decomposition process in which we separate the pulp distribution induced stochastic component and the more deterministic grid component from each other. This objective is approximately achieved with optimum notch filtering (recall adaptive filtering from section 2.2.4) on the frequency spikes around the centered DC peak (see Figure 32). First we place symmetric notch pass filters over all grid frequencies and then transform the result back to spatial domain. This yields us a spatial representation of the grid pattern. This pattern is not present in all parts of the measured image and hence we use the rule of minimum local variance to determine the amount of grid presence. As a result we obtain two images z_{stoc} and z_{det} that contain the stochastic and the deterministic components respectively. For these images we have

$$z(x, y) = z_{stoc}(x, y) + z_{det}(x, y) \text{ for } (x, y) \in D.$$

An example of sheet B1 along with extracted stochastic and deterministic components is presented in Figure 33. As we can see from the stochastic component, the grid is not visible either in the spatial or in the frequency domain. The removed grid is the only thing in the deterministic component. Hence we can consider the stochastic component as an estimate of the original surface topography due to the pulp distribution before the grid was mechanically pressed to it.

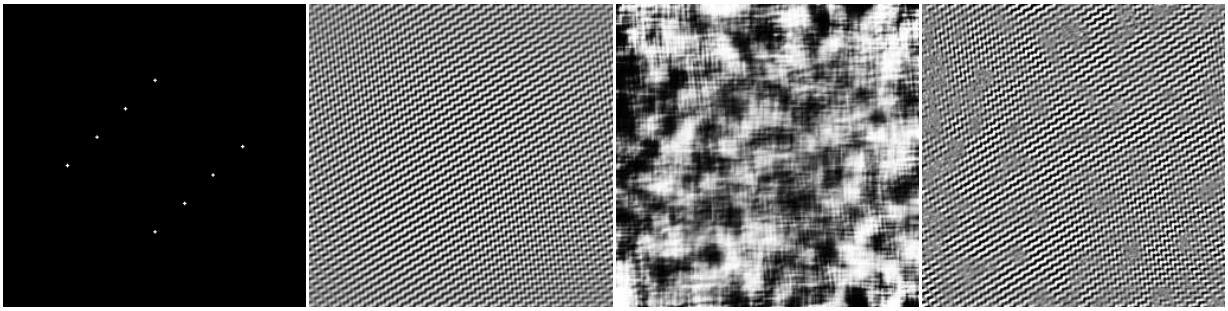


FIGURE 32: Extracting stochastic and deterministic components: applied notch pass filter (left), extracted grid pattern (center left), calculated weight (center right), and estimated grid presence (right).

Taking the stochastic component and fitting a GRF model based on an empirical semivariogram yields a powerful model interpretation for the estimated pulp distribution (recall our introduction to GRF in section 2.3.3). We assume that in all the images the nugget effect is negligible ($c_Z = 0$), and that the images are preprocessed to show process expectation $\mu_Z = 0$ and process variance, which is called process sill

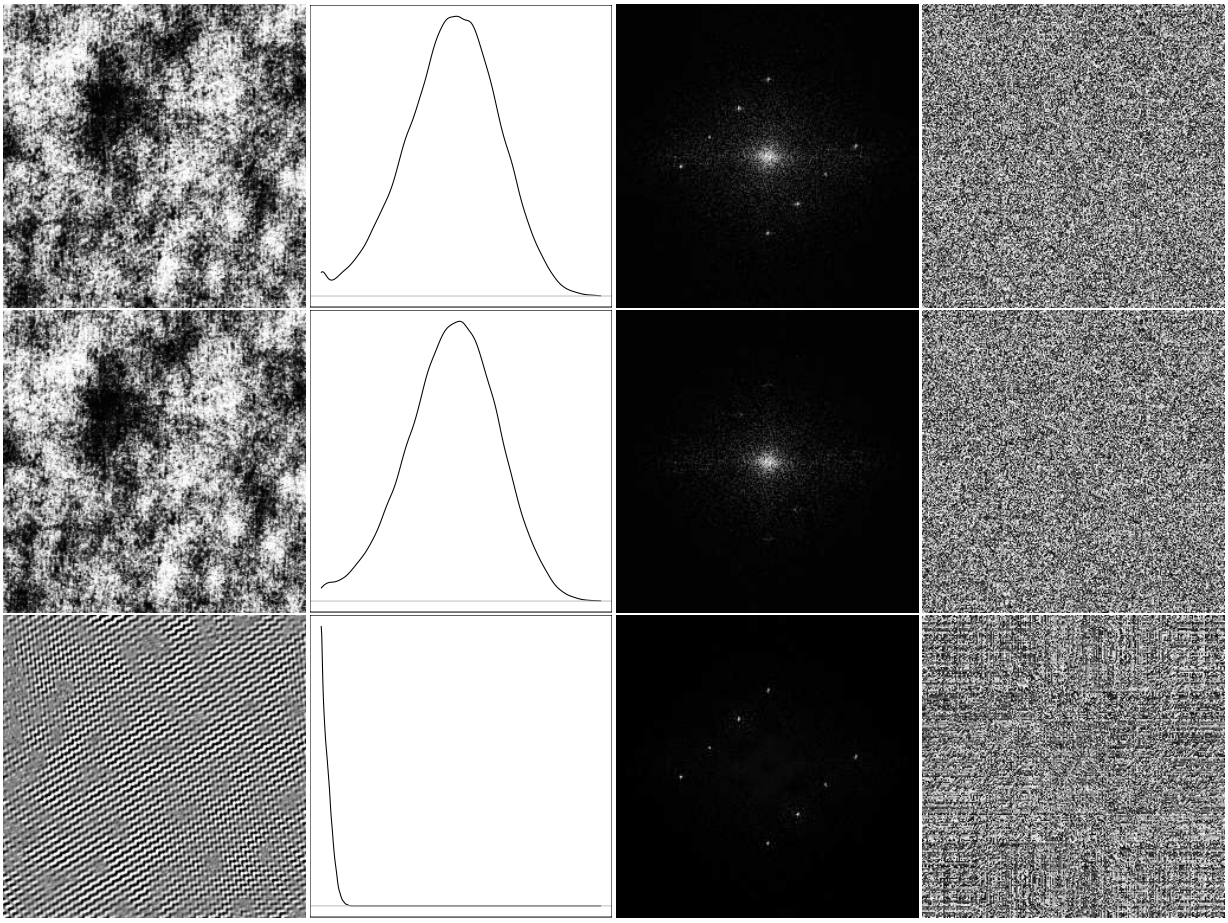


FIGURE 33: Extracted stochastic and deterministic components: the original measurement image (above), the extracted stochastic component (center), and the extracted deterministic component (below). In columns we have the image, the pixel intensity distribution of the image, the power spectrum of the image, and the phase angle of the image in this order.

from now on, $\sigma_Z^2 = 1$. As the center area of $(x, y) \in \{150, \ldots, 899\} \times \{150, \ldots, 899\}$ pixels is without errors in all the studied images we can use it for estimation. From a tentative study it was seen that it is feasible to set the maximum transition (or lag) to 60 pixels. In the estimation we benefit greatly from the fact that the properties of both paper types are approximately isotropic, as it was argued in the previous section. Although this approximate isotropy means that we can estimate the process semivariogram based on one direction only, we increase robustness by using an alternative estimator that uses horizontal and vertical directions. For mathematical convenience we introduce a new subdomain $D_{vario} = \{150, \ldots, 839\} \times \{150, \ldots, 839\}$ that is used to define the pixels that take part to the estimation. Formally the

estimate at i pixels transition is obtained as

$$\hat{\gamma}_i = \left(\frac{\sum_{(x,y)\in D_{vario}} (z(x+i,y) - z(x,y))^2 + (z(x,y+i) - z(x,y))^2}{4|D_{vario}|} \right) \text{ for } i = 0, \ldots, 60 \tag{17}$$

where $|D_{vario}|$ equals the number of pixels in the D_{vario} domain. This estimate is simple the mean of horizontal and vertical classical estimators. The results from smooth paper type sheet A1 and rough paper type sheet B1 are presented in Figures 34 and 35 respectively. A complete list of the semivariogram estimations is presented in appendix A. The presented semivariogram plots contain both empirical and theoretical semivariograms. The theoretical semivariograms are semivariograms of parametric random field models that were presented in Table 2. For a moment we focus on the empirical semivariograms only and the theoretical semivariograms are ignored for the time being. Our first conclusion is that the empirical semivariograms of smooth and rough paper types are distinctively different between the paper types. For both paper types the empirical semivariograms of background and foreground topographies look similar but the corresponding mass distributions are clearly different.

Next we evaluate empirically the process ranges, that is the maximum distances beyond which the correlation between two pixels is negligible, for the paper properties. From our methodological perspective this is interesting because from the process range we get valuable insight into the extent of spatial correlation in each image. Hence we can make more educated selection for the size of the local areas that are used for the correlation estimates. We estimate the process range in an ad hoc manner by taking the smallest spatial transition from which the five consecutive semivariogram estimates are close to process sill. Formally this is written as

$$\hat{a}_Z = \arg\min_i \left(\sum_{j=0}^{4} |1 - \hat{\gamma}_{i+j}| \right) < 0.2.$$

According to visual inspection in our case this technique is capable of estimating the correct process range.

The obtained results are presented in Table 6. It seems that all the topographies have their process range between $3.5mm$ - $4.5mm$, whereas the process ranges for the mass distributions are between $2.5mm$ - $3.5mm$, which is considerably less that the topographies have. This difference may help in explaining the forthcoming results should they favor the topographies. The process ranges of the thickness distributions range between $4.0mm$ - $4.7mm$, which is approximately the same with respect to the topographies. This concurs with intuition as the thickness is estimated from the topographies. The reasons behind the slight increase in range, which is observed in some sheets, with respect topographies are not clear to us. Our hypothesis, which we cannot substantiate at the moment, is that estimating the thickness distribution from the two topographies generates additional value, that is the thickness distribution is more informative with respect to the generating process than the topographies

are on their own, which shows as an increase in spatial correlation. The process ranges of the density distributions are approximately $1.0mm$ for the smooth paper type, and approximately $0.5mm$ for the rough paper type. These process ranges are considerably less than they are for the other paper properties. On the other hand the process range of the density distributions are the only ones that are clearly distinctively different between the two paper types. By this we mean that the obtained process ranges separate the studied images into two clearly separable classes according to the paper type. Based on the estimated process ranges we can conclude that in our case observing $3.0mm$ - $4.0mm$ local areas should make an adequate choice for further analyzes. From theoretical perspective using larger local areas than this should yield only modest improvement.

	A1	**A2**	**A3**	**B1**	**B2**	**B3**	**Distinctive**
Background	43	39	41	45	38	42	no
Foreground	44	41	40	41	40	35	no
Mass	29	27	28	34	31	27	no
Thickness	47	44	43	44	41	42	no
Density	11	10	11	6	7	6	yes

TABLE 6: Empirically estimated process ranges for the paper sheets. All ranges are given in pixels that in real world scale equal to tenths of millimeters. The distinctive column indicates whether a row of results is distinctively different between the smooth and the rough paper types.

Then we fit parametric models (recall section 2.3.3) to the empirical semivariograms that we have estimated from the stochastic components. From methodological perspective this serves us as the shape of the semivariogram gives us a geometrical understanding of the image. The model parameter s, which should not to be confused with process range a_Z (recall Table 2 on page 39), can be used as a measure of scale in the images. Larger values of s correspond to larger scale spatial structures as can be seen from Figure 36. For each model the parameter s is estimated using least squares fitting for the first twenty empirical semivariogram estimates $\hat{\gamma}_i : i = 1, \ldots, 20$, which were obtained from equation 17. The choice to use twenty estimates makes an acceptable compromise between small and large scale fitting. Formally the estimate is written as

$$\hat{s} = \arg\min_{s} \sum_{i=0}^{20} (\gamma(i|s) - \hat{\gamma}_i)^2, \tag{18}$$

where $\gamma(i|s)$ is a parametric model semivariogram, which is calculated from Table 2 with transition $||\mathbf{t}|| = i$ pixels and scaling s. According to visual observations, some of which can be made from Figures 34 and 35, this yields good model fits to our data.

The obtained results are presented in Table 7. For the topographies it seems that the calendered smooth paper (type A) obtains distinctively smaller parameter s values than the coated rough paper (type B). This is consistent with the spatial

interpretation for the parameter as given above. For the mass distributions the situation is reversed, the smoother paper type obtains larger parameter s values than the rough paper type, which is interesting as both intuition and visual observations would suggest that the topographies and the mass distributions were closely connected. We can only hypothesize that the postprocessing of the dry weight profile, which we assume is approximately similar for both paper types, is playing a role here. Calendering is known to remove water from thick areas of the dry weight profile, whereas coating adds new mass to thin areas of the dry weight profile. For the thickness distributions both paper types obtain similar parameter s values. The explanation behind this in not know to us at the moment. Only the hyperbolic model is able to make, albeit an unreliable, distinction between the two paper types. As for the density distributions, their parameter s values behave consistently with the corresponding mass distributions although approximately in four times smaller scale scale. Based on the modeling results, which show a clear distinction between the paper types, we are optimistic that also further analyzes should be able to see differences between the smooth and the rough paper types. The spatial understanding of the images, coming from the shape of the covariance structure and the scaling parameter s, can be used to justify interpretations of forthcoming results. As a sidenote, we are inclined to believe that, should there be a need for a paper type classifier and results from a larger experiment were available, this would be a fruitful platform for a Bayesian classifier.

	A1	**A2**	**A3**	**B1**	**B2**	**B3**	**Distinctive**
Background							
Spherical	16.621	16.877	19.206	28.365	23.663	23.684	yes
Exponential	6.765	6.827	7.455	12.338	9.451	9.554	yes
Hyperbolic	3.085	3.131	3.492	7.530	5.082	5.210	yes
Foreground							
Spherical	15.962	17.240	17.711	20.880	19.502	19.208	yes
Exponential	6.586	6.899	7.033	8.142	7.722	7.612	yes
Hyperbolic	2.987	3.168	3.249	4.106	3.848	3.786	yes
Mass							
Spherical	17.233	17.418	18.294	14.299	12.594	13.284	yes
Exponential	7.072	7.122	7.463	6.137	5.626	5.733	yes
Hyperbolic	3.661	3.688	3.918	2.935	2.666	2.720	yes
Thickness							
Spherical	25.754	26.097	27.028	28.374	26.846	25.638	no
Exponential	10.360	10.589	11.114	12.330	11.292	10.736	no
Hyperbolic	5.596	5.789	6.188	7.507	6.563	6.213	yes
Density							
Spherical	4.681	4.750	5.248	4.207	4.107	4.038	yes
Exponential	2.038	2.065	2.250	1.753	1.734	1.693	yes
Hyperbolic	0.837	0.834	0.923	0.715	0.700	0.674	yes

TABLE 7: Empirically estimated parametric models for the paper sheets. For each measurement image we have three parameter s values for spherical, exponential, and hyperbolic models respectively. The distinctive column indicates whether a row of results is distinctively different between the smooth and the rough paper types.

Next we estimate the shape of the covariance structure for each image based on which of the three parametric models obtained the smallest sum of squared errors

(SSE) from equation 18. The chosen covariance structures are presented in Table 8. It seems that the topographies and the thickness distributions are described quite well by the hyperbolic covariance structure, which is heavily tailed and has theoretically an infinite process range (see Figure 37). This means that spatial correlations may not be as strong in smaller scales as they are with other models but instead they do extend further into larger scales. The only exception is the background profile of sheet B2, where the estimated semivariogram is explained almost equally well by the hyperbolic model. For the mass distributions it seems that the smooth paper type is best described by exponential covariance structure, whereas the rough paper is most similar to hyperbolic covariance structure. This means that the covariance structures of mass distributions are distinctively different between the two paper types whereas this is not the case with the topographies.

The exponential covariance structure can be considered as a kind of standard covariance structure as it has an average process range (see Figure 37). This means that we have normal spatial correlations in small scales and that the maximum scale of the large scale correlations is bounded. In the image this shows as stochastic landscape with hill ranges and valleys (see Figure 36). In case of the density distributions it seems that the smooth paper type is similar to exponential covariance structure whereas the rough paper type is most similar to spherical covariance structure. This is undoubtedly a result of the differences of covariance structures of the mass distributions.

The spherical covariance structure is very short tailed (see Figure 37). This means that we have good spatial correlations in small scale whereas large scale spatial correlations are nonexistent. In the images this shows as grainy, almost noisy, structures where the lack of large scale structure is evident even to a naked eye. Based on the shapes of the covariance structures we should conclude that the difference between the two paper types is more a result of the mass distributions than it is of the topographies – unless the difference is explained solely by the removed grid structure. This conclusion is inconsistent with our prior knowledge of the papers where the aim of the postprocessing is to smoothen the topographies and the obtained results are quite different according to independently conducted roughness Bendtsen measurements ($45ml/min$ versus $180ml/min$ for smooth and rough paper types respectively). Obtaining a solution to this dilemma requires more application expertise than we have at the moment and hence the question is left open for further study.

Finally we study the paper measurements through correlograms and cross-correlograms using $0mm$ - $4mm$ transitions. These correlograms are shown in Figures 38 and 39. The former figure compares the different paper sheets from the perspective of different paper properties whereas the latter figure compares the different paper properties from the perspective of different paper sheets. While both of these images contain the same information, the former is easier to read. Therefore, we present our findings through Figure 38. In the plot matrix the diagonal plots show correlograms that measure the amount of spatial correlation within the

	A1	A2	A3	B1	B2	B3	Distinctive
Background	Hyperbolic	Hyperbolic	Hyperbolic	Exponential	Hyperbolic	Hyperbolic	no
Foreground	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	no
Mass	Exponential	Exponential	Exponential	Hyperbolic	Hyperbolic	Hyperbolic	yes
Thickness	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	no
Density	Exponential	Exponential	Exponential	Spherical	Spherical	Spherical	yes

TABLE 8: Empirically estimated shape of the covariance structure for the paper sheets. For each measurement image we choose the shape that obtains the smallest sum of squared errors (SSE) from equation 18. The distinctive column indicates whether a row of results is distinctively different between the smooth and the rough paper types.

images. These are just an alternative way of presenting the semivariograms, some of which were already shown in Figures 34 and 35, and hence lead into the same conclusions as we made above. The off-diagonal elements show cross-correlograms that measure the amount of spatial correlation between the images. These have not been analyzed as of yet and as the understanding of the cross-covariance structures is important with respect to evaluation of the proposed methodology we do it here.

It seems that the cross-correlograms between the surface topographies and the thickness distributions are as strong and extends as far as the correlograms within these images. The only noticeable difference between these correlograms and cross-correlograms is that the cross-correlograms show nugget effects (the left hand side limits of the correlograms are clearly not equal to 1). These nugget effects mean that there are small scale discontinuities in the cross-covariance structure of the images. We are confident that these nugget effects are results from not having complete pixel correspondences between the images instead of actually having such discontinuities between the paper properties. The fact that the diagonal plots do not show similar nugget effects supports this conclusion. An another observation is that these correlograms and cross-correlograms are distinctively different for the calendered smooth paper (type A) and the coated rough paper (type B).

For the mass distributions on the other hand it seems that the cross-correlograms with the surface topographies and with the thickness distributions are even stronger than the correlograms are within the mass distributions, with the exception of the nugget effect that is seen in these cross-correlograms as well. This is a somewhat unexpected result and at the moment we are unable to provide an explanation for it. We assume that the stronger cross-correlations are seen because of the larger process ranges that the surface topographies and thickness distributions have when compared to the mass distributions, recall Table 6, but on the other hand we are unable to offer any physical justification for this phenomenon. In addition it seems that based on these correlograms and cross-correlograms we are unable to make distinctions between the smooth and the rough paper types.

Based on the above conclusions we are confident that there are measurable dependencies between the images and now feel ready to start using the proposed

methodology. For the density distribution it seems that the cross-correlations are negligible except with mass distributions with which we can see correlations in small scales. It is unlikely that the proposed methodology could find strong dependencies between the density and the other paper properties. Hence we need to be conservative in our conclusions about the dependencies with the density distribution.

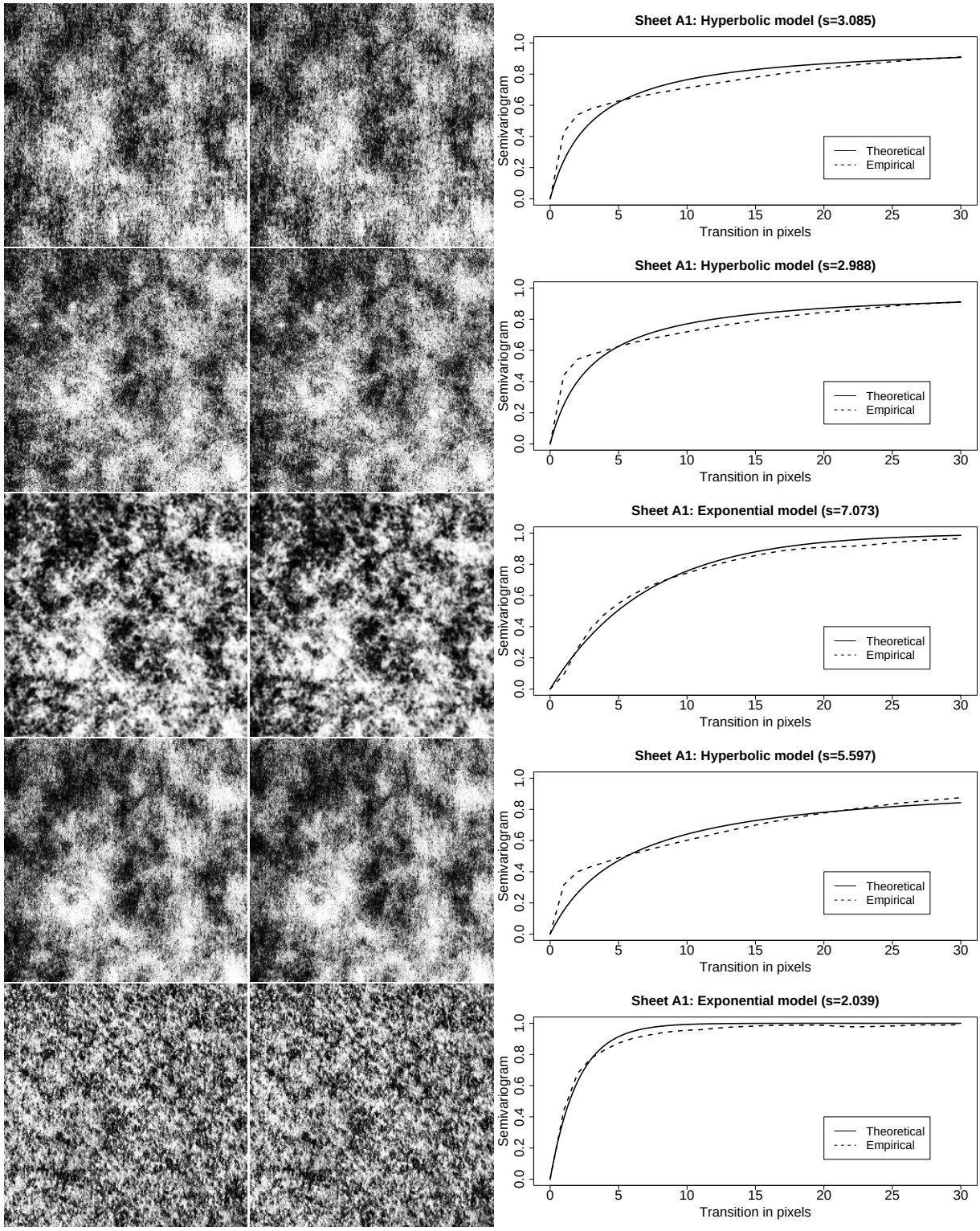


FIGURE 34: Semivariograms for calendered smooth paper type sheet A1: the original measurement image (left), the extracted stochastic component (center), and the semivariograms for the pulp component (right). Rows consist of background and foreground topographies along with mass, thickness, and density distributions in this order.

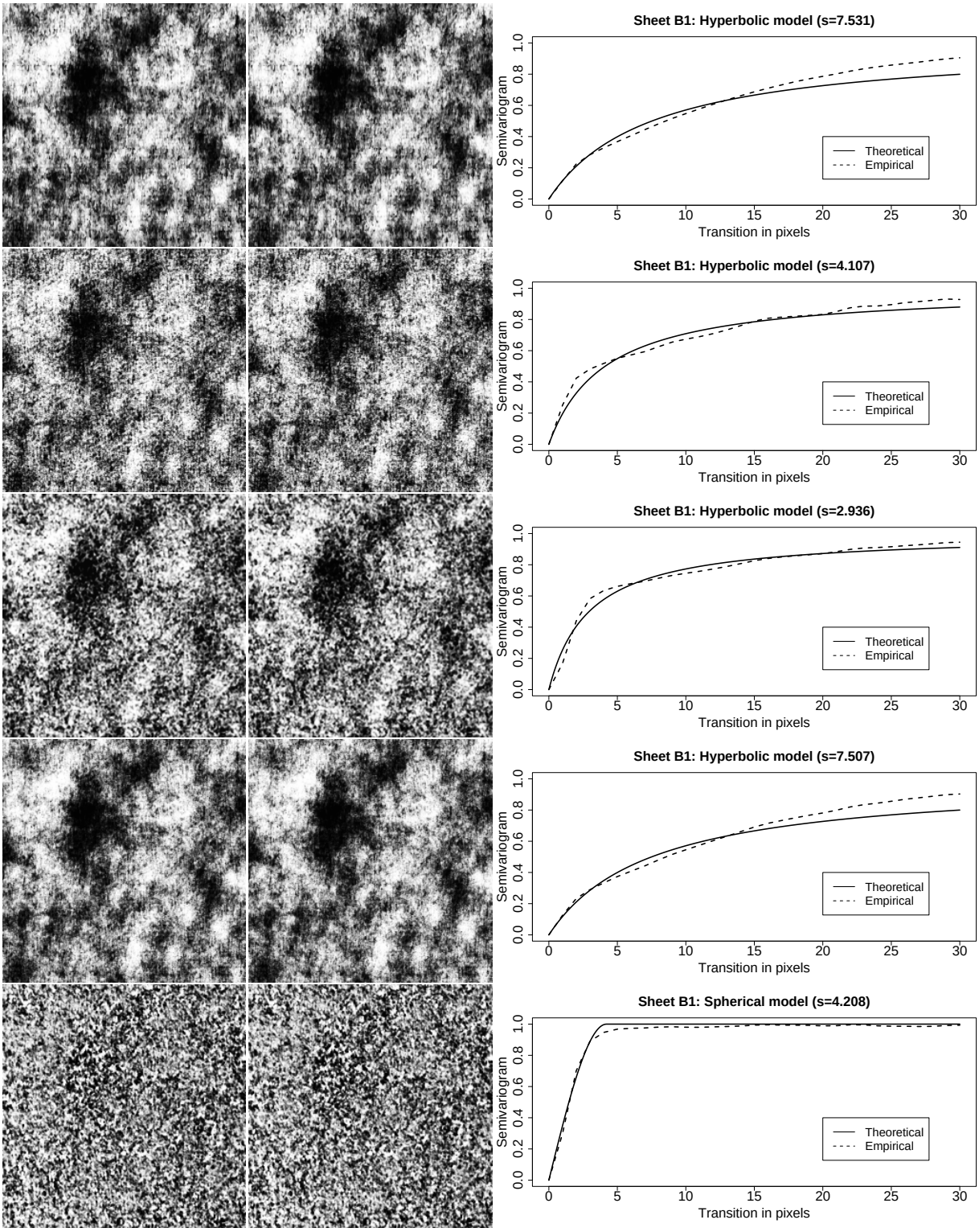


FIGURE 35: Semivariograms for coated rough paper type sheet B1: the original measurement image (left), the extracted stochastic component (center), and the semivariograms for the pulp component (right). Rows consist of background and foreground topographies along with mass, thickness, and density distributions in this order.

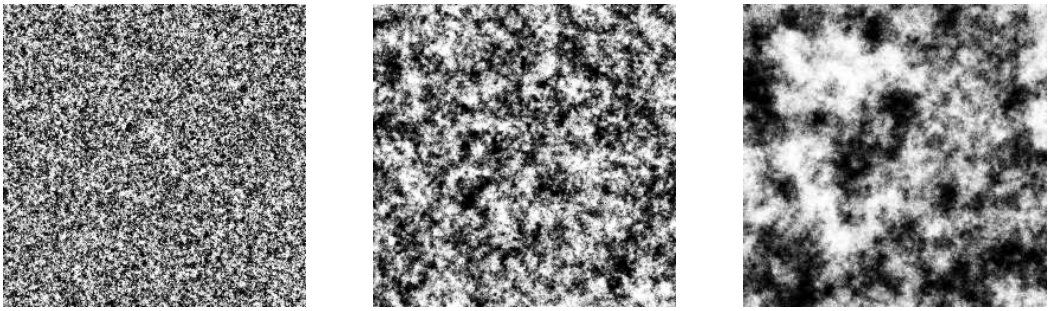


FIGURE 36: Effect of parametric model parameter s. The three images show an exponential model with parameter $s = 1$ (left), $s = 4$ (center), and $s = 16$ (right).

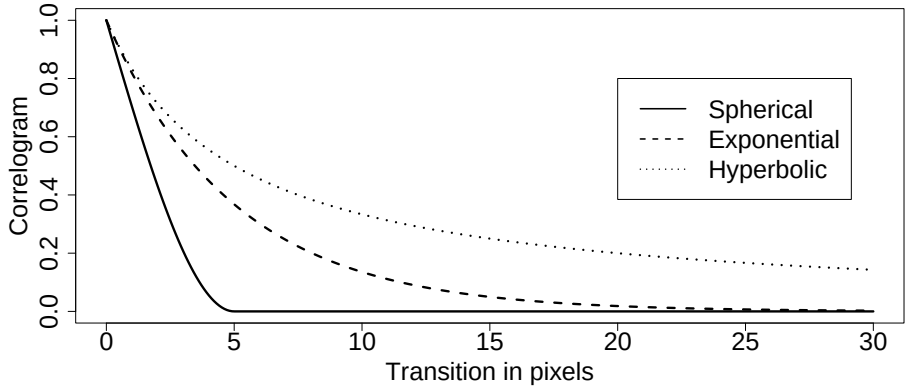


FIGURE 37: Parametric model covariance structures for spherical, exponential, and hyperbolic models. All the visualized models use scaling parameter $s = 5$.

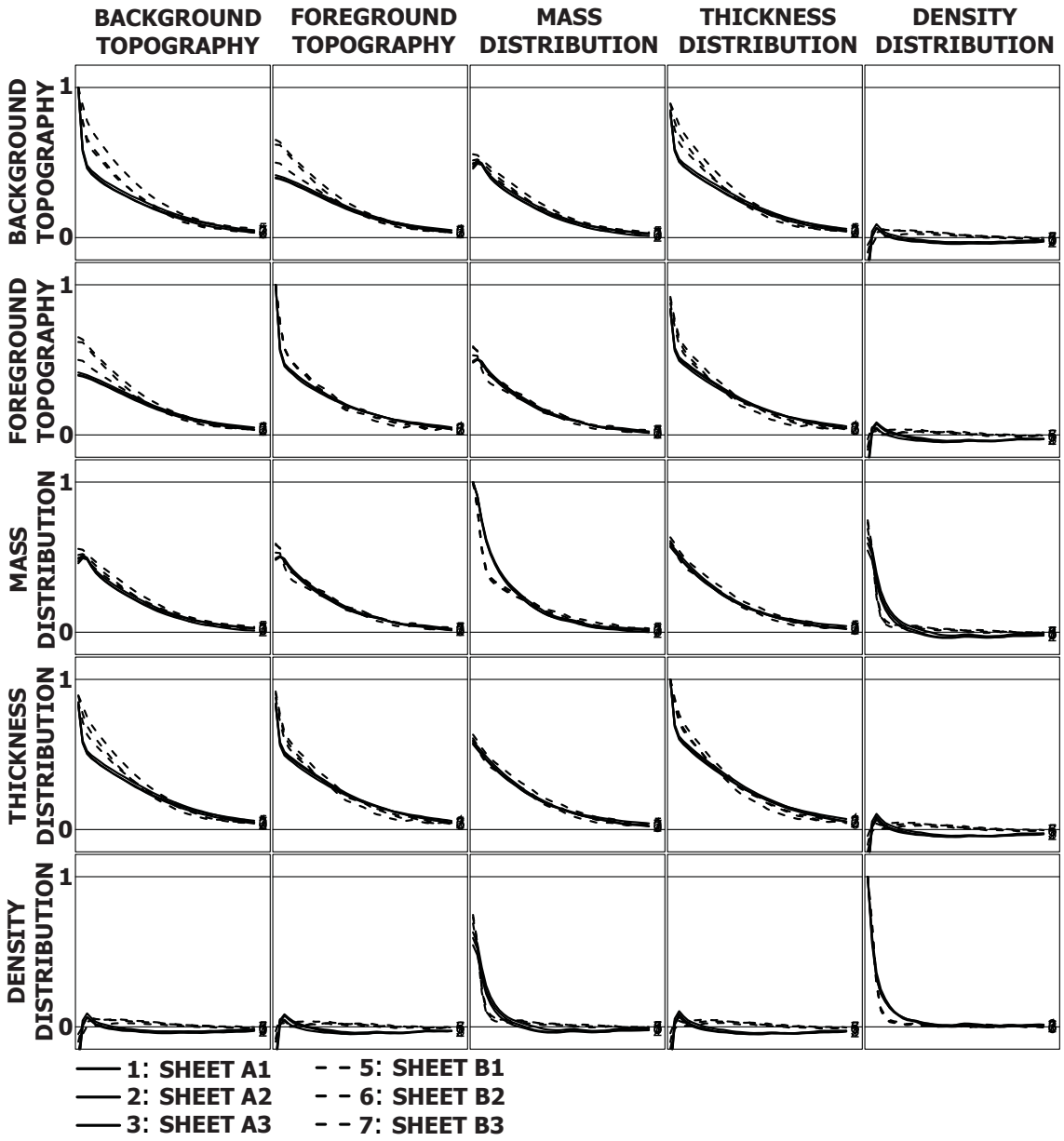


FIGURE 38: Empirical correlograms for different paper sheets from the perspective of different paper properties. Columns and rows select the two paper properties between which the dependency is measured whereas the line type indicates the paper type. The horizontal axes indicate spatial transitions from $0.0mm$ to $4.0mm$, which correspond to $0 - 40$ pixels, and the vertical axes measure the amount of observed correlation.

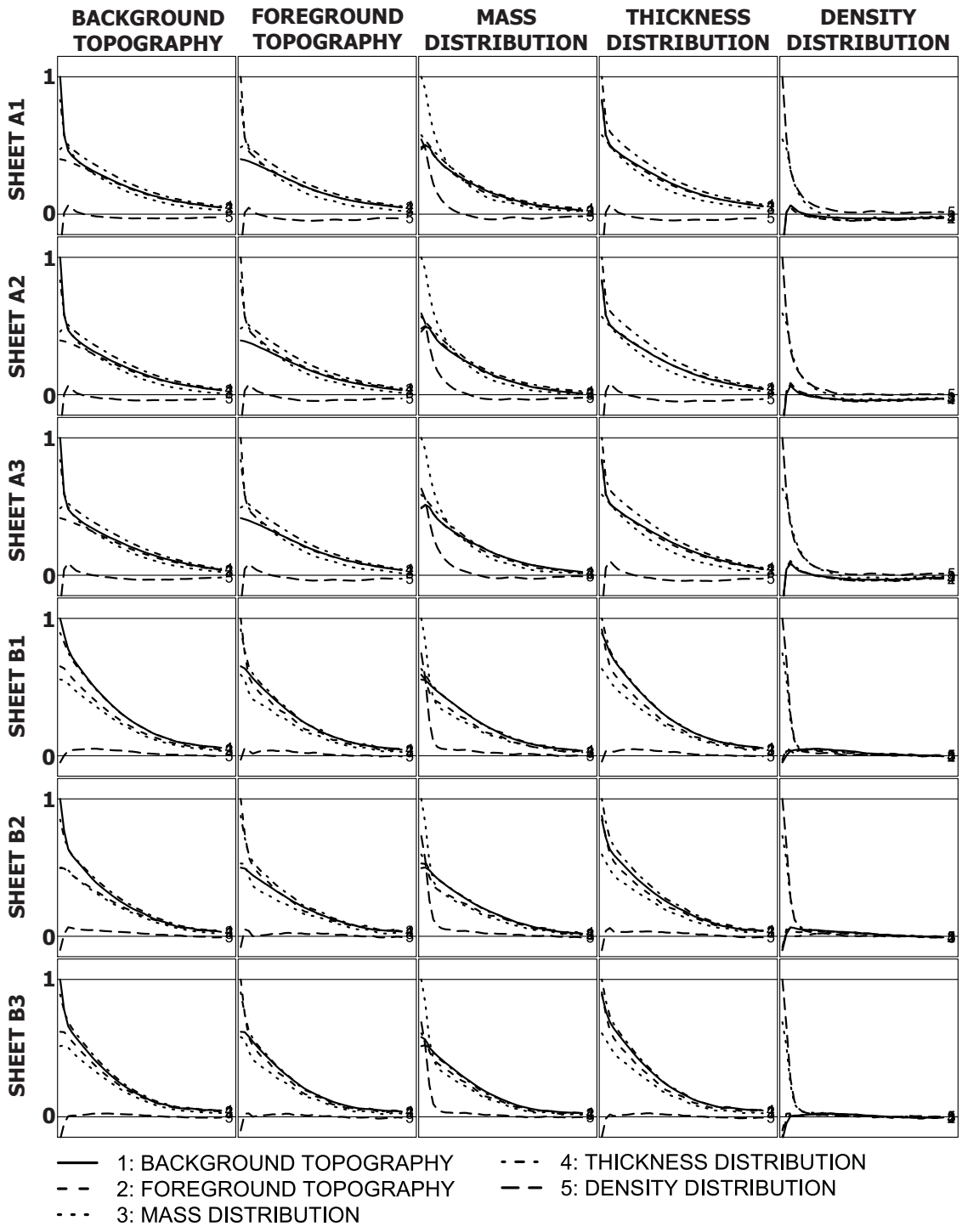


FIGURE 39: Empirical correlograms for different paper properties from the perspective of different paper sheets. Columns and line type select the two paper properties between which the dependency is measured whereas the row indicates the paper type. The horizontal axes indicate spatial transitions from $0.0mm$ to $4.0mm$, which correspond to $0 - 40$ pixels, and the vertical axes measure the amount of observed correlation.

5.3 Simulations and visual inspections

The results that have been analyzed thus far seem very consistent. In order to validate the obtained results and the made conclusions we check them visually by studying the goodness of the parametric model fits through the visual differences between the empirical and theoretical semivariograms, and the visual differences between measurement images and their simulations. As we stated in the previous section, the paper properties are only approximately Gaussian. Even further, it is unrealistic to assume that any real world data would obey a theoretic model exactly, and thus observing only the first two moments of the data distribution may not be enough to give an accurate description. We have found that using the parametric models to simulate new measurement images and making visual comparisons between the real world images and their simulations yields a good understanding of how well the model is able to describe the data and also on the phenomena that are and are not captured by the model.

Due to the large amount of simulations only four examples are presented in Figure 40. From these examples we can see that the parametric model is able to capture the geometric nature of the images. Visual observation of foreground topography of sheet A2 indicates that the measurement image contains a large scale material structure along with small scale variation, which may be a result of measurement noise. As we did not do hierarchical modeling the scale of our simulation is between those large and small scale phenomena. The simulations of mass distribution of calendered smooth paper sheet A2 seem to be very successful. If it was not known a priori which of the images were simulated, picking the original measurement image would be very difficult. The foreground topography of coated rough paper sheet B2 shows how the simulations ignore the grid pattern that was removed at the beginning of this section. Again it seems that the simulated images have scale that are between the small scale and the large scale structures of the original measurement image – even though the grid was removed. As a conclusion we can state that the simulations seem visually similar to the measurement images but when the measurement images contains phenomena of different scales the scale of the simulation images is likely not correct. An improvement would be to use hierarchical modeling, which is almost as easy to implement as the presented non-hierarchical approach. As we are satisfied with the currently obtained results, we leave the hierarchical modeling subject to further study.

From the visual observations, which we analyzed above, we re-evaluated the computationally estimated results on the shapes of the covariance structures (as presented in Table 8). The re-evaluated results are presented in Table 9. With respect to surface topographies and thickness distribution, which is their derivative, the prior made conclusions seem valid without revisions. For the calendered smooth paper (type A) the semivariogram fits are not perfect in $0.2mm$ - $0.5mm$ range, whereas for the coated rough paper (type B) the hyperbolic model is overestimating the spatial correlation above $2.0mm$ scale. The results from the mass

distributions, which have a good semivariogram fit for all scales, on the other hand seem more inconclusive as both exponential and hyperbolic models are able to produce approximately the same result. Although the computationally estimated best shapes of the covariance structures do seem better than the corresponding second best shapes, the difference is not as clear as it is with the surface topographies. In fact the difference between the selected and the second best covariance structures is so small that it approaches the level when the difference can be explained with stochastic variation. The same phenomena is naturally transferred into the density distributions for which the rough paper type (sheets B1, B2, and B3) are indecisive between the spherical and the exponential models. Combined with the unexplained inconsistency with what explains the difference between the two paper types (as discussed two paragraphs above), it seems that in order to fully understand the effect of mass distribution more research and application expertise is needed.

	A1	**A2**	**A3**	**B1**	**B2**	**B3**	**Distinctive**
Background	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	no
Foreground	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	no
Mass	Exponential Hyperbolic	Exponential Hyperbolic	Exponential Hyperbolic	Hyperbolic Exponential	Hyperbolic Exponential	Hyperbolic Exponential	inconclusive
Thickness	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	Hyperbolic	no
Density	Exponential	Exponential	Exponential	Spherical Exponential	Spherical Exponential	Spherical Exponential	inconclusive

TABLE 9: Visually inspected shape of the covariance structure for the paper sheets. For each measurement image we choose the shape that based on semivariograms and simulations seems visually the most compatible. The distinctive column indicates whether a row of results is distinctively different between the smooth and the rough paper types.

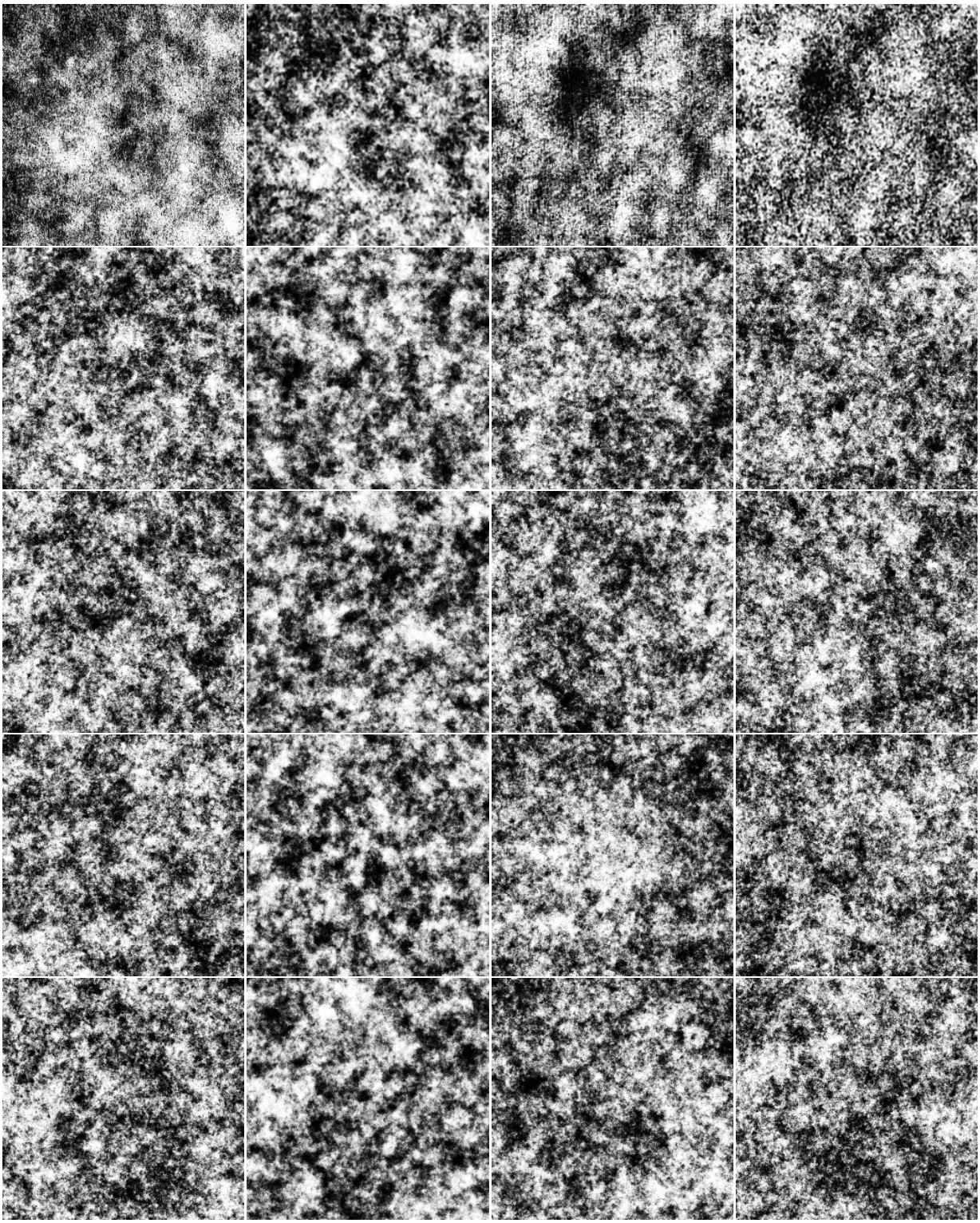


FIGURE 40: Example simulations of paper images. Columns: foreground topographies of smooth paper type sheet A2, mass distributions of smooth paper type sheet A2, foreground topographies of rough paper type sheet B2, and mass distributions of rough paper type sheet B2. Rows: original measurement images, 1st simulations, 2nd simulations, 3rd simulations, and 4th simulations.

5.4 Explanations behind the spatial correlations of the paper images

As a natural continuation to the previous section we study the dependencies between paper properties on spatial scales $0.75mm$ - $4.00mm$ using the proposed CCA method, which was introduced in chapter 3. This is an entirely different approach for analyzing spatial correlations than what was used in the previous section. The presented analysis is based on $N = 10000$ observations that are partitioned into $N_{tr} = 5000$ training and $N_{val} = 5000$ validation observations, and on $w_{local} = 10$ pixels times $h_{local} = 10$ pixels subimages with the subimage to image scaling factor r ranging in $[0.75, 4.00]$.

The estimated correlation functions, which are computed with the proposed CCA methodology, are shown in Figures 41 and 42. These figures are organized in the same way as Figures 38 and 39: Figure 41 compares the different paper sheets from the perspective of different paper properties, whereas Figure 42 compares the different paper properties from the perspective of different paper sheets. Again both of these figures contain the same information and hence the same conclusions can be made from either figure. In order to ease seeing the differences between the calendered smooth (type A) and the coated rough (type B) papers and between different spatial scales, proposed correlation coefficients from $1mm$, $2mm$, and $4mm$ spatial scales are presented in Table 10.

From the estimated process ranges we concluded that spatial scales $3.0mm$ - $4.0mm$ should have the highest correlations between most of the paper properties. Based on the presented correlation functions, see Figures 41 and 42, it seems that the highest correlations are in $2.5mm$ - $4.0mm$ spatial scales. The only exception is found from the process ranges of the density distributions, which were shown to be in $0.5mm$ - $1.2mm$. These are also the spatial scales that show the highest correlations for the density distribution. These observations support our previous conclusion: the process ranges and the highest correlating local image areas are approximately of the same spatial scale. Enlarging the observed local image areas beyond the scale of the process range seems to be harmful. The physical explanation for this seems clear: the process range defines the spatial scale of the present material structures in the images. Too small subimages are unable to represent these structures in the most efficient way, and too large subimages are impractical as the structures from the further corners of the subimages are no longer correlated.

It seems that with respect to both surface topographies and thickness distribution these three paper properties have a bit stronger correlation with the coated rough paper type (type B) than with the calendered smooth paper type (type A). This fits with our prior knowledge of how the dry weight profiles of the sheets are postprocessed. The surface topographies of the smooth paper type, which have roughness Bendtsen value $45ml/min$, experience heavy changes during the mechanical compression of the calendering and seem to some extent lose their connection to the original pulp distribution of the dry weight profile. The surface topographies

of the rough paper type, which have roughness Bendtsen value $180ml/min$, do not have to sustain such a rough treatment and hence they are able to retain more of their original distributions from the dry weight profile. The same conclusion can be made from the estimated model scale parameters (Table 7) for which the rough paper type obtains larger values, which likely mean that originally there are larger correlating material structures that are destroyed in the calendering process.

When the surface topographies and the thickness distributions are compared with the mass distributions the situation seems to be reversed: the surface topographies of the calendered smooth paper type (type A) have stronger correlations with the mass distributions than the coated rough paper type (type B) has. This phenomenon is clear below $2mm$ scales. We believe, but based on our own research are not able to substantiate, that this is because of the coating, which adds a filler substance to the holes and valleys of the surface topographies. The physical properties of this filler are usually different from those of the pulp mass of the dry weight profile. This means that the correlation between the mass distributions of the dry weight profile and the processed paper diminishes and hence the connection between the surface topographies and the mass distribution of the postprocessed paper is also weakened. This conclusion is supported by the lower model scale parameters values (Table 7), which mean that the calendered smoother paper type has spatially larger mass structures than the coated rough paper type has. Unfortunately a visual confirmation of this conclusion is difficult and all that we can say is that the mass distribution of the rough paper type looks more grainy in small scales, which is consistent with our conclusion. An another interesting observation is that there seems to be a stronger correlation between the two surface topographies themselves than between them and the mass distribution. Physically this seems unjustified as the mass lies between the two surfaces. The only plausible explanation that we can think of is that the two topographies are measured with the same measuring instrument whereas the mass distribution is measured with a different instrument and using a lower spatial resolution.

Based on our results it seems that the calculated thickness distribution outperforms the two measured surface topographies. This is especially especially below $2mm$ scales as in larger scales than this the difference is observable but considerable smaller. In smaller scales the surface topographies are relatively independent, but they both are heavily correlated with the estimated thickness distribution. We cannot reliably quantify how much of this dependency is a result of the computational equation that we used to calculate the thickness. In addition the thickness distribution seems to have higher correlation with the mass distribution. This leads us to conclude that the thickness distribution is more informative with respect to the underlying material structure than the surface profiles are on their own. As a parallel conclusion was already made from the process ranges, we are inclined to believe in this explanation.

Perhaps the most surprising result comes from the observed dependencies with respect to the density distribution. As the density distribution is estimated compu-

tationally from the other measurements, we cannot quantify reliably how much of the observed dependency comes artificially from equation 16. Hence we need to be conservative in our conclusion for the application perspective. From the methodological perspective on the other hand the obtained results seem very interesting. Based on visual observations the density distributions of both paper types look grainy and somewhat independent of the other paper properties (see Figures 34 and 35). The estimated process ranges tell us that spatial correlations exist in below $0.5mm$ - $1.2mm$ scales, and based on the estimated cross-correlograms there seemed to be next to no spatial correlations with respect to the other paper properties. Hence it is surprising that the proposed CCA method is able to see spatial correlations in all scales of the smooth paper type. The role of the smoothness is easy to understand as the smoother the surfaces are the closer the density is to the mass distribution, the scaling of the pixel values being the main difference. An absolutely smooth paper would show exactly the same amount of spatial correlation with respect to both mass and density distributions. From our methodological design, recall chapter 3, we know that the observed levels of correlation that are computed via the proposed CCA method cannot be explained by chance. Hence we must conclude that assessing the dependency through visual observations can be deceptive as the human eye seems to be unable to see certain types of dependencies.

Due to the large number of combinations of image pairs we are unable to present spatial interpretations for all of them. Instead we select only two of them based on how interesting they are from the methodological perspective. Visualizations of dependency between foreground topography and mass distribution, and foreground topography and density distribution of the coated rough paper type sheet B1 are presented in Figures 43 and 44 respectively. Of these Figure 43 illustrates a typical observed dependency. In this case the observation points are clearly lined on the diagonal and the regression model seems like a good fit to the observation points. The corresponding subimage pairs within all three groups seem to show similar spatial structures and the subimage pairs in left and right groups are also somewhat similar. With respect to our paper application we can interpret these subimage pairs so that, as the dark areas represent holes and valleys in the foreground topography and low amount of mass in the mass distribution, there are round holes of $0.25mm$ - $1.00mm$ radius in the surface that are a result of low amount of pulp and filler in those regions. This indicates that the coating procedure was unable to smooth out the roughness of the dry weight profile. As a conclusion, which is based on all the obtained results, we can say that in most cases the proposed CCA method is able to measure the amount of dependency, and to provide a clear physical interpretation for this reasons behind the dependency. To show an exception Figure 44 presents a case where the CCA method was unable to find clear dependencies, as we already saw from Figures 41 and 42. As we can see, the dependency between the observation points is almost nonexistent and the regression model is unable to explain these points. This reflects to the subimage pairs, which no longer show similar spatial structures, and hence it is also impossible to try to build any kind

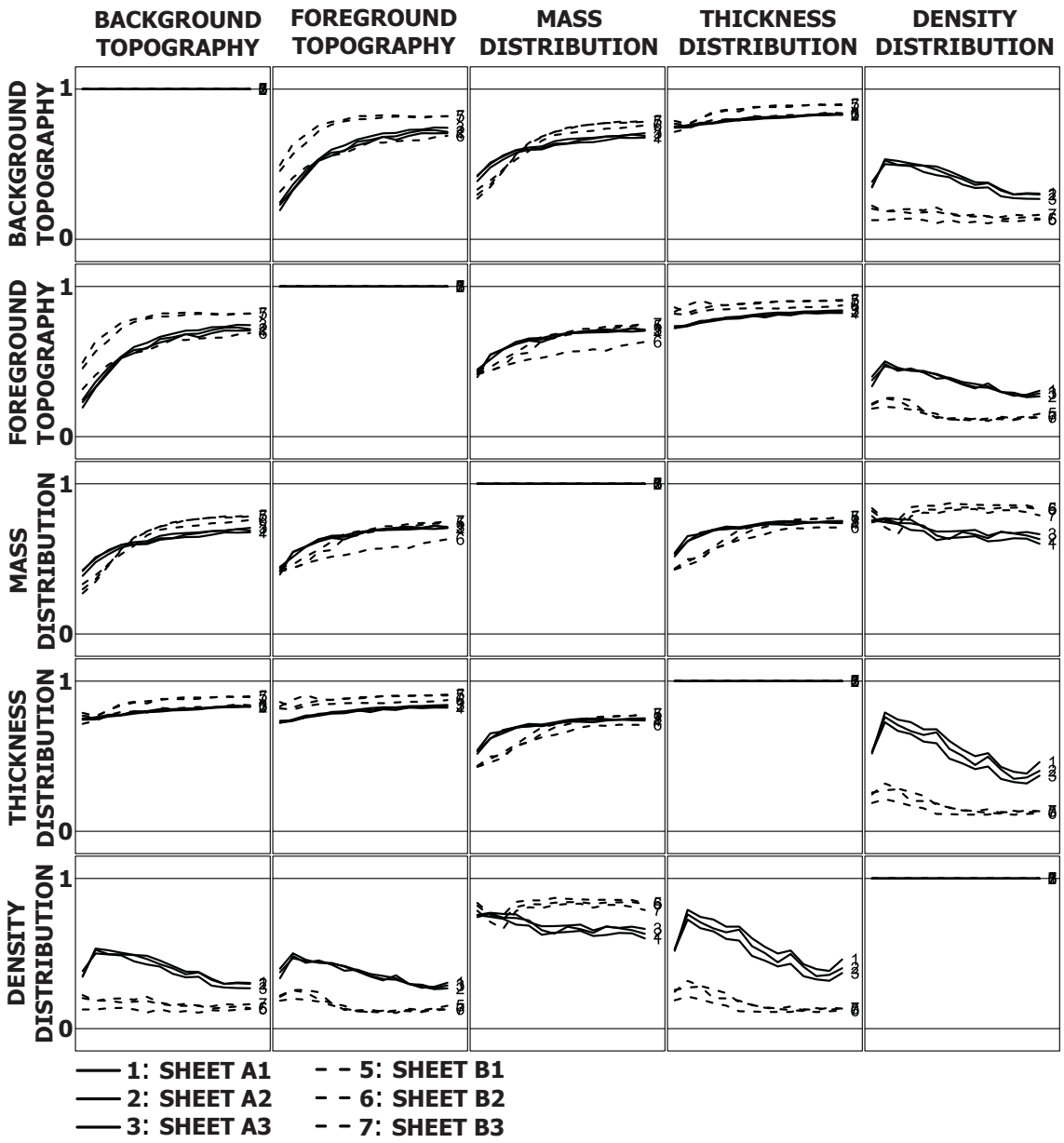


FIGURE 41: Estimated CCA correlation functions for different paper sheets from the perspective of different paper properties. Columns and rows select the two paper properties between which the dependency is measured whereas the line type indicates the paper type. The horizontal axes indicate size of the observed local areas from $0.75mm \times 0.75mm$ to $4.00mm \times 4.00mm$ and the vertical axes measure the amount of observed dependency.

of physical interpretation. In the current analysis problems like this were limited to the density distribution of the rough paper type. The explanation might be in the role of coating, or because the relation between density and the other studied properties is nonlinear. Nonlinearities are not handled well by the proposed version of the methodology and appropriate improvements are suggested in the discussion

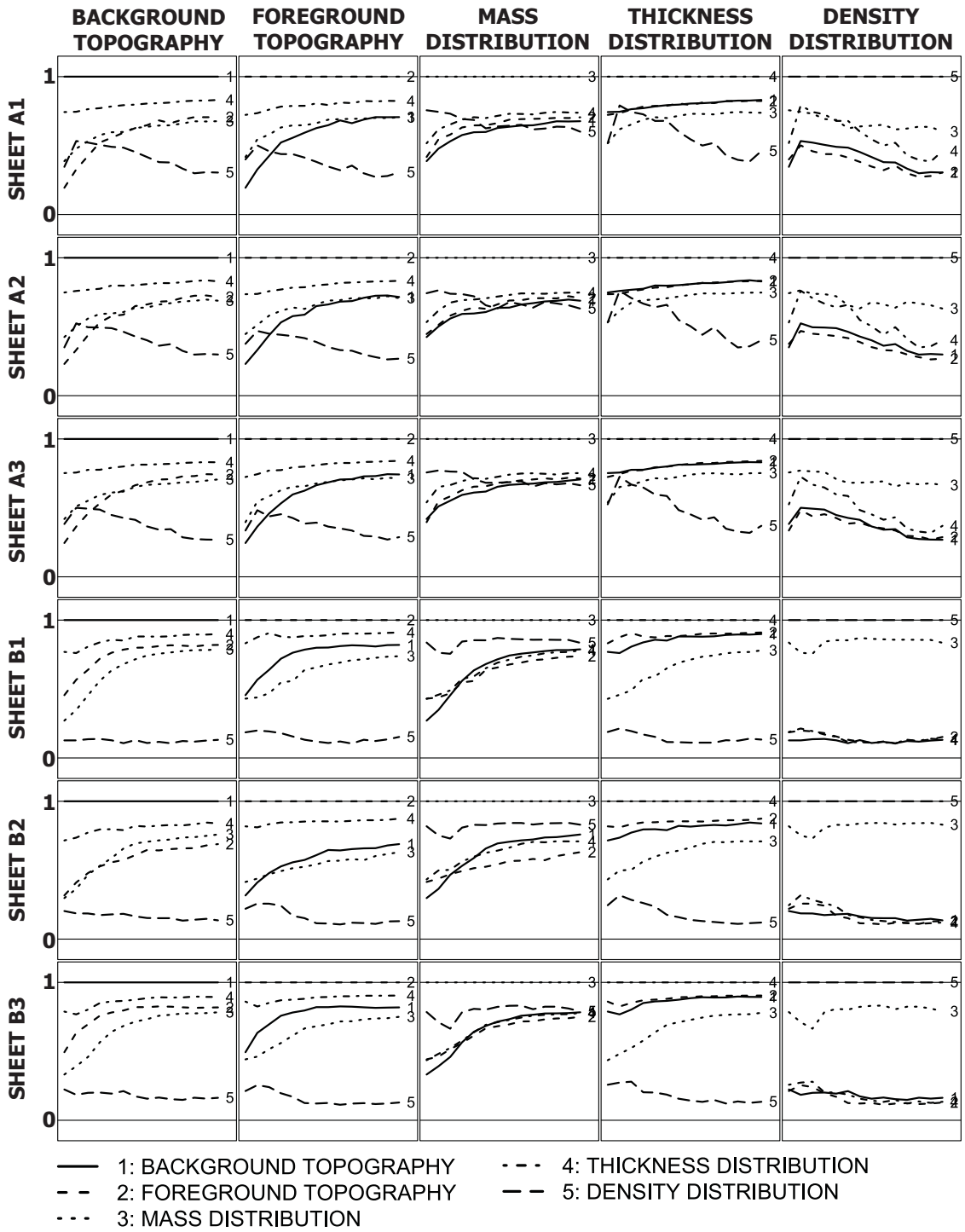


FIGURE 42: Estimated CCA correlation functions for different paper properties from the perspective of different paper sheets. Columns and line type select the two paper properties between which the dependency is measured whereas the row indicates the paper type. The horizontal axes indicate size of the observed local areas from $0.75mm \times 0.75mm$ to $4.00mm \times 4.00mm$ and the vertical axes measure the amount of observed dependency.

	A1	A2	A3	B1	B2	B3	Distinctive
Background Foreground							
$1mm$	0.327	0.331	0.364	0.567	0.411	0.634	yes
$2mm$	0.593	0.587	0.625	0.786	0.574	0.796	no
$4mm$	0.706	0.715	0.742	0.820	0.689	0.818	no
Background Mass							
$1mm$	0.478	0.501	0.510	0.348	0.364	0.392	yes
$2mm$	0.598	0.606	0.618	0.681	0.657	0.684	yes
$4mm$	0.677	0.688	0.706	0.788	0.759	0.783	yes
Background Thickness							
$1mm$	0.745	0.759	0.756	0.760	0.736	0.768	no
$2mm$	0.793	0.797	0.800	0.851	0.790	0.865	no
$4mm$	0.831	0.829	0.831	0.898	0.839	0.894	yes
Background Density							
$1mm$	0.533	0.524	0.501	0.128	0.187	0.185	yes
$2mm$	0.485	0.463	0.427	0.107	0.183	0.212	yes
$4mm$	0.304	0.298	0.269	0.132	0.136	0.163	yes
Foreground Mass							
$1mm$	0.545	0.513	0.547	0.439	0.439	0.462	yes
$2mm$	0.647	0.629	0.657	0.638	0.525	0.666	no
$4mm$	0.705	0.708	0.713	0.738	0.630	0.750	no
Foreground Thickness							
$1mm$	0.735	0.738	0.744	0.875	0.810	0.825	yes
$2mm$	0.788	0.787	0.799	0.884	0.853	0.881	yes
$4mm$	0.823	0.833	0.841	0.910	0.874	0.904	yes
Foreground Density							
$1mm$	0.501	0.469	0.483	0.199	0.258	0.255	yes
$2mm$	0.413	0.418	0.386	0.132	0.152	0.125	yes
$4mm$	0.306	0.268	0.288	0.152	0.130	0.129	yes
Mass Thickness							
$1mm$	0.618	0.617	0.651	0.459	0.498	0.482	yes
$2mm$	0.700	0.706	0.711	0.651	0.625	0.692	yes
$4mm$	0.737	0.748	0.752	0.780	0.709	0.779	no
Mass Density							
$1mm$	0.744	0.764	0.771	0.763	0.752	0.711	no
$2mm$	0.625	0.656	0.682	0.853	0.827	0.803	yes
$4mm$	0.600	0.631	0.664	0.835	0.830	0.789	yes
Thinkness Density							
$1mm$	0.791	0.760	0.726	0.214	0.318	0.274	yes
$2mm$	0.679	0.659	0.586	0.116	0.176	0.186	yes
$4mm$	0.461	0.402	0.370	0.133	0.121	0.134	yes

TABLE 10: Estimated CCA correlations in $1mm$, $2mm$, and $4mm$ scales. The distinctive column indicates whether a row of results is distinctively different between the smooth and the rough paper types.

of chapter 3.

Based on all the findings that we have made in this section, it seems that the thickness distribution is the paper property that best characterizes a paper sheet. It has a strong spatial correlation to all other paper properties, with the exception

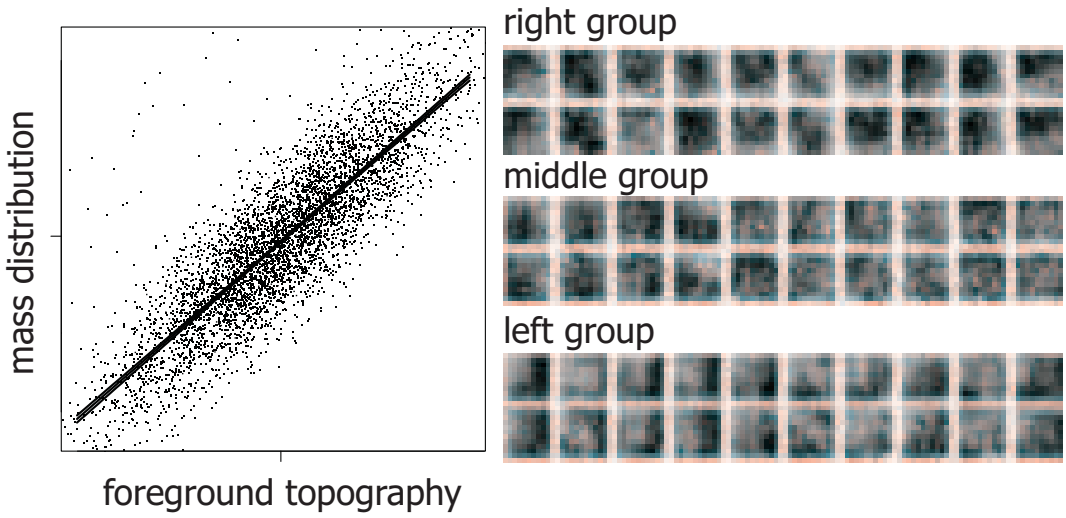


FIGURE 43: Visualization of dependency between foreground topography and mass distribution of rough paper type sheet B1 (for interpretation recall section 3.5): a scatter plot of observations with a regression model (left) and a spatial interpretation for this regression model (right). Each 10×10 pixels subimage represents a $4mm \times 4mm$ area paper measurement.

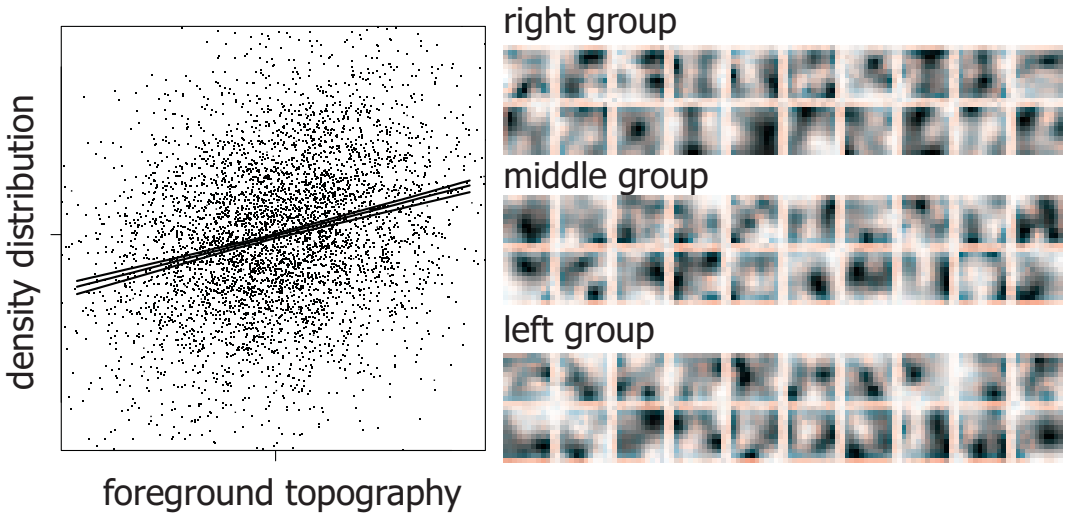


FIGURE 44: Visualization of dependency between foreground topography and density distribution of rough paper type sheet B1 (for interpretation recall section 3.5): a scatter plot of observations with a regression model (left) and a spatial interpretation for this regression model (right). Each 10×10 pixels subimage represents a $1mm \times 1mm$ area paper measurement.

of density distribution for which the reasons were already discussed above. Unfortunately measuring the thickness distribution is problematic and hence where it the case that only one measurement could be conducted, we are inclined to recommend measuring the mass distribution, which gives a more detailed description of

the sheet than either of the surface topographies are able to give on their own. As measuring the mass distribution is also the de facto approach in the paper industry, making such measurements is relatively easy and there is abundance of literature from previous studies that concentrate solely on mass distribution that can be used to evaluate newly obtained measurements.

5.5 Discussion

In this chapter we have performed an example data analysis using our methodology (chapter 3) and our example implementation (chapter 4) on our paper images (chapter 1 and 5). First we introduced our paper measurement images and how they were preprocessed for the analyzes. Then we conducted a tentative study where we modeled the images with parametric Gaussian random field models. These models where then used to simulate new images from which we were able to make a visual assessment of how well each model was able to explain the modeled image. From all this we obtained a good understanding of the contents of each image. This understanding of the images was considered important in order to learn how the proposed methodology works in a real world setting. This was one of the main goals of this chapter. In section 1.3 we presented the three spatial correlation measures that are studied in this thesis. Based on our research we can now assess the strengths and the weaknesses of these dependency measures.

The strengths of the correlation between pixel values, which we call pixel correlation, are simplicity and straightforwardness. In an ideal case it measures exactly the thing that we are interested in and hence it is easy to motivate and to understand. In a real world case the weaknesses of the pixel correlation on the other hand are plentiful: spatial distortions, measurement errors, and noise can easily render this approach useless. Whereas to some degree it is possible to remedy the effects of measurement errors in pixel values through robust correlation estimates, there is no known salvation from spatial distortions that can very effectively hide any existing dependency. In practice measurement images tend to contain all of these errors and hence the applicability of this methodology in a real world setting is questionable. In addition the obtained result does not help us to understand spatial correlations between the studied images. In this thesis the pixel correlation was considered unworthy based on our tentative study and was not even included in the actual analysis that was presented in this chapter.

The strengths of the semivariogram and the correlogram lie in their solid theoretical foundation and their well established position in the statistical community. A considerable advantage is that we have a family of theoretical parametric models that can be used to explain the image contents with only a few scalar values (the model parameters). These models can then be used to simulate new measurement images. As for the weaknesses there are at least two with substantial consequences. The semivariogram and its derivatives are susceptible to lack of pixel correspondence. The presence of any geometric distortions render the obtained estimates

theoretically unfounded and hence it is questionable to make further inference on them. This problem simply was not considered in the original problem settings that was used to derive the methodology. An another problem is with measurement noise, which affect the obtained results as the estimators do not include any type of spatial smoothing. Regardless of these shortcomings the semivariogram and the correlogram represent the state of the art in a short list of available spatial correlation measures and hence they were included in our example analysis.

The strengths of the proposed CCA correlation coefficient and CCA correlation function are in interpretability and robustness towards noise and lacking pixel correspondence. At least from the perspective of our paper application we can argue that the proposed CCA correlation measure is able to answer the posed questions unlike the other two studied correlation measures. By this we mean that the methodology yields a spatial description of the structures that are correlated between the images. The ability to measure the amount of dependency in different spatial scales is important in many real world applications. There are also a few noticeable shortcomings with the most important being the current lack of rigorous analytical analysis of the proposed methodology. In addition there are no theoretical models that could explain the content of the image and that could be used for simulations. Depending on the implementation the estimation of the proposed CCA correlation measure is also computationally heavier, which means that the results are not available as fast as they are from the other two approaches.

As a summary we can state that the pixel correlation is next to useless in a real world setting. The second-order correlation is theoretically justified but may break down in many real world situations. The proposed CCA correlation measure, although not analytically analyzed as of yet, is more interpretable and robust in real world situations. The geometrical interpretations of the latter two are different and hence rather than competing they complement each other. Giving an example of how to use these two together to build a more complete picture of the studied phenomenon was the other main goal of this chapter.

6 SUMMARY

The objective of this thesis is to study all aspects of spatial correlations between images that have stochastic content. It is assumed that there is an underlying generator process that induces all the images but that each image exhibits slightly different aspects from that process. Hence in order to study the generating process, we must find a way to combine information from all the available images and based on that try to build an understanding of the process characteristics. This is an example of classical statistical inference that is expanded to multivariate spatial data analysis.

Our work is motivated and demonstrated with a real world application that is raised from the paper making industry. The objective of the application is to design an experiment for studying dependencies between five paper properties: background and foreground topographies along with mass, thickness, and density distributions. From our partner we have obtained a data set that contains these measurements from eight paper sheets that come from two clearly different paper types. The experiment was designed so that we can study similarities that exist between the sheets of the same type and on the other hand the dissimilarities that exist between the sheets of the two paper types.

The development of the proposed methodology began from a tentative study that showed clearly that modeling dependencies between single pixel values yields inadequate results. On the other hand the stochastic content and the mere size of the images made us think that processing the images as a whole would be too complex a problem. Instead we proposed that decomposing the images into small and simple subimages would solve both the theoretical and the computational issues that are blocking the two extreme approaches. The proposed methodology takes two images and measures the amount of spatial dependency between them with a single scalar value. For this value the methodology yields confidence bounds that can be used to evaluate the obtained result. As an essential part of the methodology, four visualization techniques were also introduced.

In order to evaluate the proposed methodology and to use it to analyze the paper measurement data we needed a computer implementation for it. For this purpose we reviewed a system architecture that is based on R language and environment, and outlined an example implementation that we later used in our example data analysis. The presented example implementation provides a graphical working environment in which the user can work with multiple images and models in an interactive manner.

The presented example data analysis of paper data began from a detailed study of the available measurement images. Based on this study we were able to conclude that the studied paper images can be decomposed into a deterministic component, which contains regular patters such as wire pattern, and a stochastic component, which contains for example the stochastic distribution of pulp. It was then shown that the stochastic component can effectively be modeled with parametric Gaussian

random field models. These models were used to characterize the studied paper images with only a few scalar values. With these values we were able to quantify the stochastic nature and spatial scale of the five studied paper properties in a spatially interpretable manner. Then we studied the dependencies between the paper properties with correlograms and with the proposed CCA correlation measures. Based on the obtained results we were able to compare the two studied paper types and the five paper properties, and to make an evaluation of the three spatial correlation measures that are studied in this thesis.

The four cornerstones of the thesis: the paper application, the proposed methodology, the example implementation, and the example data analysis, constitute a versatile study of problems and their example solutions that we encountered in our own work. The employed manner of presentation is aimed not to be limited to theoretical discussion without a contact to real world nor to solving a real world application without understanding the theoretical foundations, but rather the presentation tries to merge both views in a harmonious way with the objective of attracting interest from both traditionally separate disciplines.

7 YHTEENVETO (SUMMARY IN FINNISH)

Tämän työn tavoitteena on tutkia monipuolisesti spatiaalisia korrelaatioita jotka esiintyvät stokastisia rakenteita sisältävien kuvien välillä. Näiden kuvien taustalla oletetaan olevan spatiaalinen stokastinen prosessi joka generoi kuvat siten että jokainen kuva ilmentää tämän prosessin ominaisuuksia. Tämän prosessin ymmärtämiseksi meidän täytyy löytää keinoja joilla useista kuvista saatava informaatio voidaan yhdistää prosessia parhaiten kuvaavalla tavalla. Käytännössä tämä tapahtuu laajentamalla perinteisen tilastollinen päättelyn työkaluja moniulotteiseen spatiaaliseen data-analyysiin sopiviksi.

Esitetty työ on motivoitu ja kerrottu paperiteollisuudesta saadun käytännön ongelman kautta. Tämän ongelman ratkaisun tavoitteena on suunnitella koe jonka avulla voidaan selvittää riippuvuuksia paperin ominaisuuksien välillä. Tässä työssä tutkitut ominaisuudet ovat paperin etu ja kääntöpuolen pintaprofiili sekä massa-, paksuus- ja tiheysjakaumat. Yhteistyökumppaniltamme saatu datajoukko sisältää näiden paperiominaisuuksien mittaukset kahdeksasta paperiarkista jotka edustavat kahta selvästi erilaista paperityyppiä. Koe on suunniteltu siten että sen avulla voidaan tutkia samanlaisuuksia saman paperityypin sisällä ja erilaisuuksia eri paperityyppien välillä.

Uuden menetelmän kehitys alkoi alustavan tutkimuksen osoitettua perinteisen yksittäisiin pikseleihin perustuvan menetelmän riittämättömyyden paperikuvien analyysissä. Toisaalta kuvissa esiintyvä stokastinen rakenne ja kuvien suuri koko saivat meidän epäilemään että kuvien käsittely kokonaisina olisi liian kompleksinen tehtävä. Näiden lähestymistapojen sijaan esitimme että kuvat jaettaisiin pienemmiksi ja yksinkertaisemmiksi alikuviksi jolloin edellä mainittuja lähestymistapoja vaivaavat teoreettiset ja laskennalliset ongelmat on helpompi ratkaista. Alikuvien käyttöön perustuva esitetty menetelmä mittaa kahden kuvan välillä esiintyvän riippuvuuden määrää yhdellä skalaariluvulla. Tälle luvulle saadaan menetelmästä luottamusvälit joiden avulla on mahdollista arvioida saatua tulosta. Tämän lisäksi ja oleellisena osana menetelmää ehdotimme neljää visualisointitekniikkaa.

Esitetyn menetelmän arvioimiseksi ja käytännön ongelmasta saadun mittausaineiston analysoimiseksi on menetelmä implementoitu tietokoneella. Tähän liittyen kävimme läpi järjestelmäarkkitehtuurin joka perustuu R-kieleen ja laskenta ympäristöön, luonnostelimme esimerkki-implementaation jota myöhemmin käytettiin esimerkki data-analyysissä. Esitetty esimerkki-implementaatio tarjoaa graafisen työpöydän jonka avulla käyttäjä voi työskennellä interaktiivisesti useiden kuvien ja mallien kanssa.

Esitetty esimerkki data-analyysi paperikuville alkoi hankittuihin mittauskuviin perehtymällä. Tämän perusteella saatoimme päätellä että tutkittujen kuvien rakenteet voidaan jakaa deterministiseen komponenttiin, joka sisältää säännöllisiä rakenteita kuten viiran rakenne, ja stokastiseen komponenttiin, joka sisältää esimerkiksi stokastisen paperimassan jakauman. Tämän jälkeen osoitimme että tämä

stokastinen komponentti voidaan mallintaa tehokkaasti parametristen Gaussisten satunnaiskenttien avulla. Näiden mallien avulla saatoimme määrittää tutkituissa paperin ominaisuuksissa esiintyvät stokastiset rakenteet ja näiden skaalat spatiaalisesti tulkittavassa muodossa. Analyysin seuraavassa vaiheessa tutkimme paperiominaisuuksien välisiä riippuvuuksia korrelogrammin ja esitetyn CCA korrelaatiomitan kautta. Saatujen tulosten perusteella saatoimme tehdä vertailua eri paperityyppien ja tutkittujen paperin ominaisuuksien välillä, ja samalla arvioida kolmen erilaisen tässä työssä käytetyn spatiaalisen korrelaatiomitan ominaisuuksia.

Tämän työn neljä tukijalkaa: sovellus paperin rakenteesta, esitetty menetelmä, esimerkki-implementaatio ja esimerkki data-analyysi muodostavat monipuolisen näkökulman meidän omassa työssämme esiintyneisiin ongelmiin ja niiden esimerkkiratkaisuihin. Käytetty esitystapa ei rajoitu teoreettiseen pohdintaan eikä pelkän käytännön ongelman ratkaisuun vailla riittävää menetelmällistä ymmärrystä, vaan pyrkii yhdistämään molempia näkökulmia tasapainoisesti tavoitteenaan herättää kiinnostusta molemmissa perinteisesti erillisissä koulukunnissa.

REFERENCES

Banerjee S., Bradley P.C. and Gelfand A.E.: *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC. 2004.

Borch J., Lyne M.B., Mark R.E. and Habeger C.C. Jr: *Handbook of Physical testing of Paper, Volume 2, 2nd Ed.* Marcel Dekker, Inc. 2002.

Borga M., Knutsson H. and Landenius T.: *Learning Canonical Correlations*. In proceedings of the 10th Scandinavian Conference on Image Analysis (SCIA). 1997.

Borga M. and Knutsson H.: *Canonical Correlation Analysis in Early Vision Processing*. In proceedings of the 9th European Symposium on Artificial Neural Networks (ESANN). 2001.

Bovik A. (editor): *Handbook of Image & Video Processing*. Academic Press. 2000.

Brown P., Diggle P. and Henderson R.: *A Non-Gaussian Spatial Process Model for Opacity of Flocculated Paper*. In Scandinavian Journal of Statistics 30. Pages 355-368. 2003.

Chilès J.-P. and Delfiner P.: *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, Inc. 1999.

Cressie N.A.C.: *Statistics for Spatial Data*. John Wiley & Sons, Inc. 1993.

Cressie N.A.C. and Wikle C.K.: *The Variance-Based Cross-Variogram: You can Add Apples and Oranges*. In Mathematical Geology Volume 30 No 7. pp. 789-799. 1998.

Davies E.R.: *Machine Vision - Theory Algorithms Practicalities, Third Edition*. Morgan Kaufmann, Elsevier Inc. 2005.

Deng M. and Dodson C.T.J.: *Paper: An Engineered Stochastic Structure*. Tappi Press, Atlanta. 1994.

Dodson C.T.J. and Sampson W.W.: *Spatial statistics of stochastic fibre networks*. In Journal of Statistical Physics 96. Pages 447-458. 1999.

Dodson C.T.J., Oba Y., and Sampson W.W.: *On the distributions of mass, thickness and density of paper*. In Appita Journal 54. Pages 385-389. 2001.

Dodson C.T.J., Oba Y., and Sampson W.W.: *Bivariate normal thickness-density structure in real near-planar stochastic fibre networks*. In Journal of Statistical Physics 102. Pages 345-353. 2001.

Erkkilä A.-L.: *Paperin kerroksellisen orientaation mittaaminen ja syntymekanismit*. Licentiate Thesis. Department of Physics, University of Jyväskylä. 1995.

Erkkilä A.-L., Pakarinen P. and Odell M.: *Sheet forming studies using layered orientation analysis*. In Pulp & Paper Canada Volume 99. Pages 81-85. 1998.

Gonzales R.C. and Woods E.W.: *Digital Image Processing*. Addison-Wesley Pub Company. 1992.

Gonzales R.C. and Woods E.W.: *Digital Image Processing, Second Edition*. Prentice-Hall, Inc. 2002.

Hotelling H.: *Relations between two sets of variates*. In Biometrika 28. Pages 321-327. 1936.

Huber P.J.: *Robust Statistics*. John Wiley & Sons, Inc. 2003.

Jain A.K.: *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc. 1988.

Johansson J.-O.: *Models of Surface Roughness with Applications in Paper Industry*. Doctoral Thesis. Lund Institute of Technology, Lund University. 2002.

Kay J.: *Feature discovery under contextual supervision using mutual information*. In International Joint Conference on Neural Networks (IJCNN). 1992.

Knutsson H. and Borga M.: *Learning Visual Operators from Examples: A New Paradigm in Image Processing*. In proceedings of the 10th International Conference on Image Analysis and Processing (ICIAP). 1999.

Knuttson H., Andersson M., Borga M. and Wiklund J.: *Automated Generation of Representations in Vision*. In proceedings of the 15th International Conference on Pattern Recognition (ICPR). 2000.

Kärkkäinen S. and Jensen E.B.V.: *Estimation of fibre orientation from digital images*. In Image Analysis & Stereology 20. Pages 199-202. 2001.

Kärkkäinen S., Penttinen A., Ushakov N.G. and Ushakova A.P.: *Estimation of orientation characteristic of fiborous material*. In Advances in Applied Probability 33. Pages 559-575. 2001.

Kärkkäinen S.: *Orientation Analysis of Stochastic Fibre Systems with an Application to Paper Research*. Doctoral Thesis. Department of Mathematics and Statistics, University of Jyväskylä. 2003.

Lantuéjoul C.: *Geostatistical Simulation - Models and Algorithms*. Springer-Verlag, GmbH. 2002.

Mark R.E., Habeger C.C. Jr, Borch J. and Lyne M.B.: *Handbook of Physical testing of Paper, Volume 1, 2nd Ed*. Marcel Dekker, Inc. 2002.

Niskanen K.: *Paper Physics*. Fapet Oy. 1998.

Sonka M., Hlavac V. and Boyle R.: *Image processing, Analysis, and Machine Vision*. Brooks/Cole Publishing Company. 1999.

Rousseeuw P.J. and Leroy A.M.: *Robust Regression & Outlier Detection*. John Wiley & Sons, Inc. 2003.

Stoyan D., Kendall W.S. and Mecke J.: *Stochastic Geometry and its Applications, Second Edition*. John Wiley & Sons, Inc. 1995.

Stoyan D. and Stoyan H.: *Fractals, Random Shapes and Point Fields*. John Wiley & Sons, Inc. 1994.

Taskinen S.: *On Nonparametric Tests of Independence and Robust Canonical Correlation Analysis*. Doctoral Thesis. Department of Mathematics and Statistics, University of Jyväskylä. 2003.

R Development Core Team: *An Introduction to R*. Version 2.2.1 (2005-12-20). `http://cran.r-project.org/doc/manuals/R-intro.pdf`. 2005.

R Development Core Team: *Writing R Extensions*. Version 2.2.1 (2005-12-20). `http://cran.r-project.org/doc/manuals/R-exts.pdf`. 2005.

Tukey J.W.: *Exploratory Data Analysis*. Addison-Wesley Publishing Company, Inc. 1977.

Wackernagel H.: *Multivariate Geostatistics, Third Edition*. Springer-Verlag, GmbH. 2003.

A PARAMETRIC MODELING OF PAPER IMAGES

In this appendix we review the results of fitting three parametric Gaussian random field models to our paper images. For a detailed description of the paper images, image preprocessing, and modeling procedures, read chapter 5.

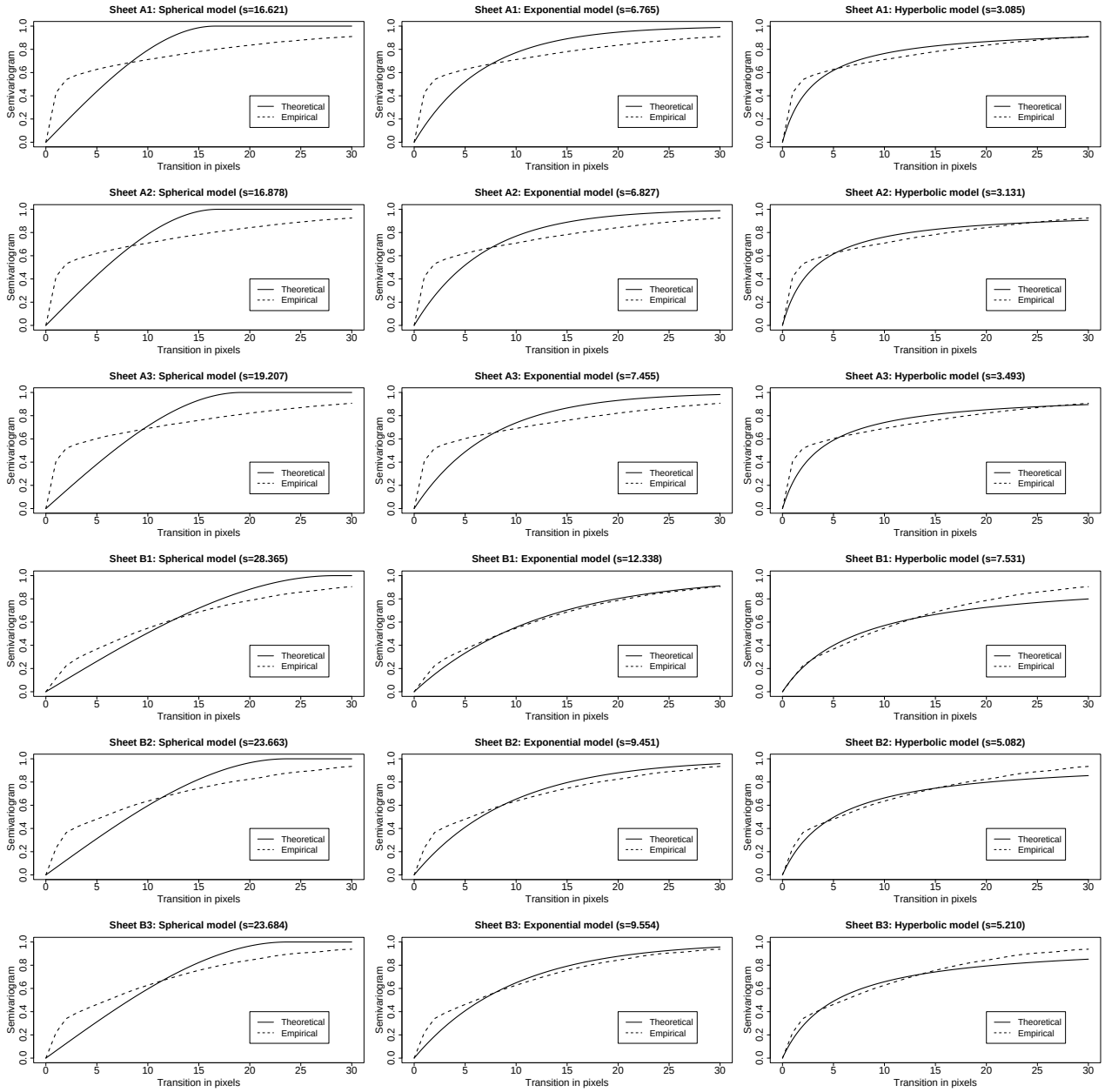


FIGURE 45: Semivariograms for background topographies. Columns indicate the employed parametric model: spherical (left), exponential (center), and hyperbolic (right). Rows indicate the studied sheet: A1 (top), A2, A3, B1, B2, and B3 (bottom). For interpretation see section 5.2.

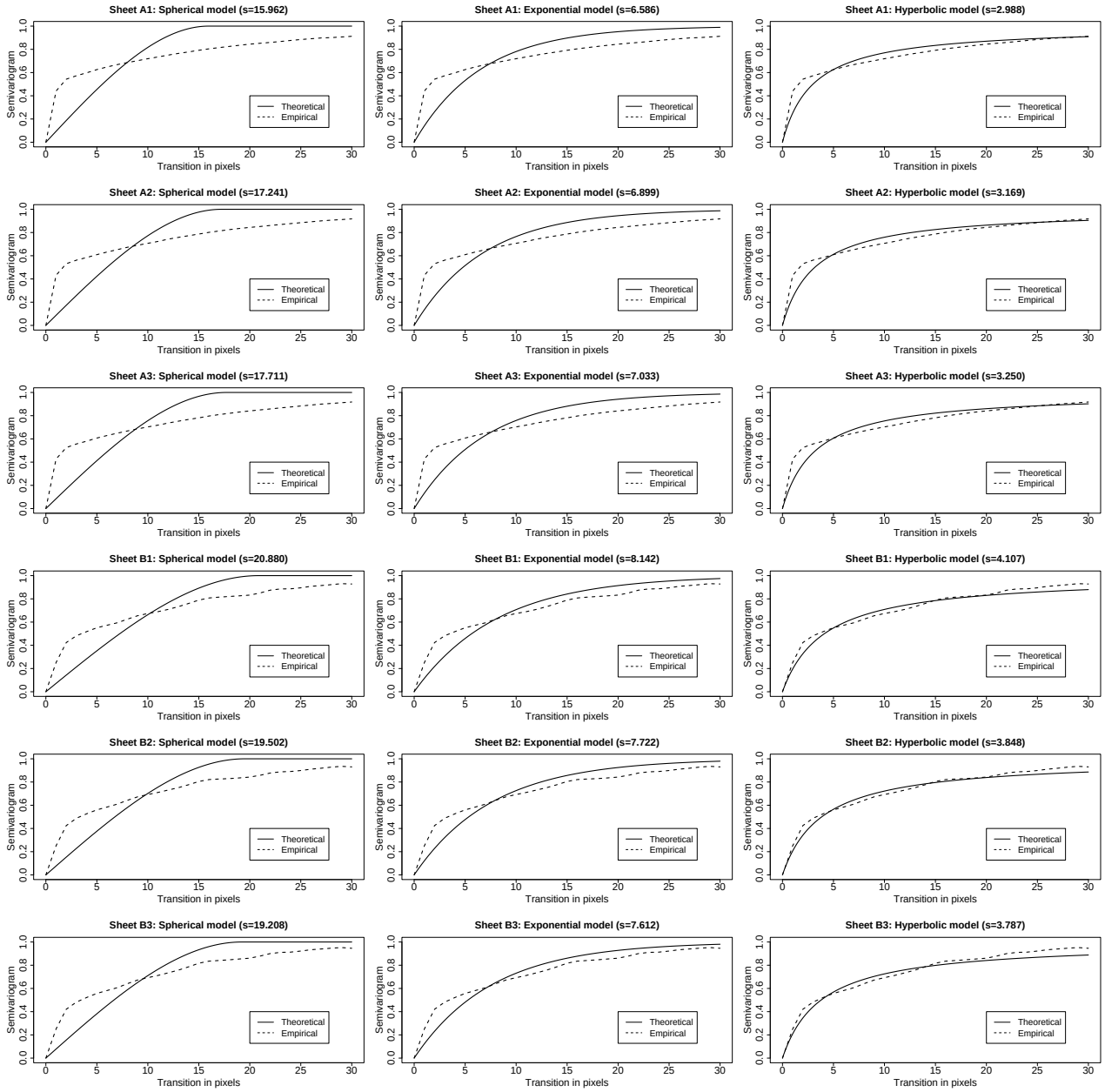


FIGURE 46: Semivariograms for foreground topographies. Columns indicate the employed parametric model: spherical (left), exponential (center), and hyperbolic (right). Rows indicate the studied sheet: A1 (top), A2, A3, B1, B2, and B3 (bottom). For interpretation see section 5.2.

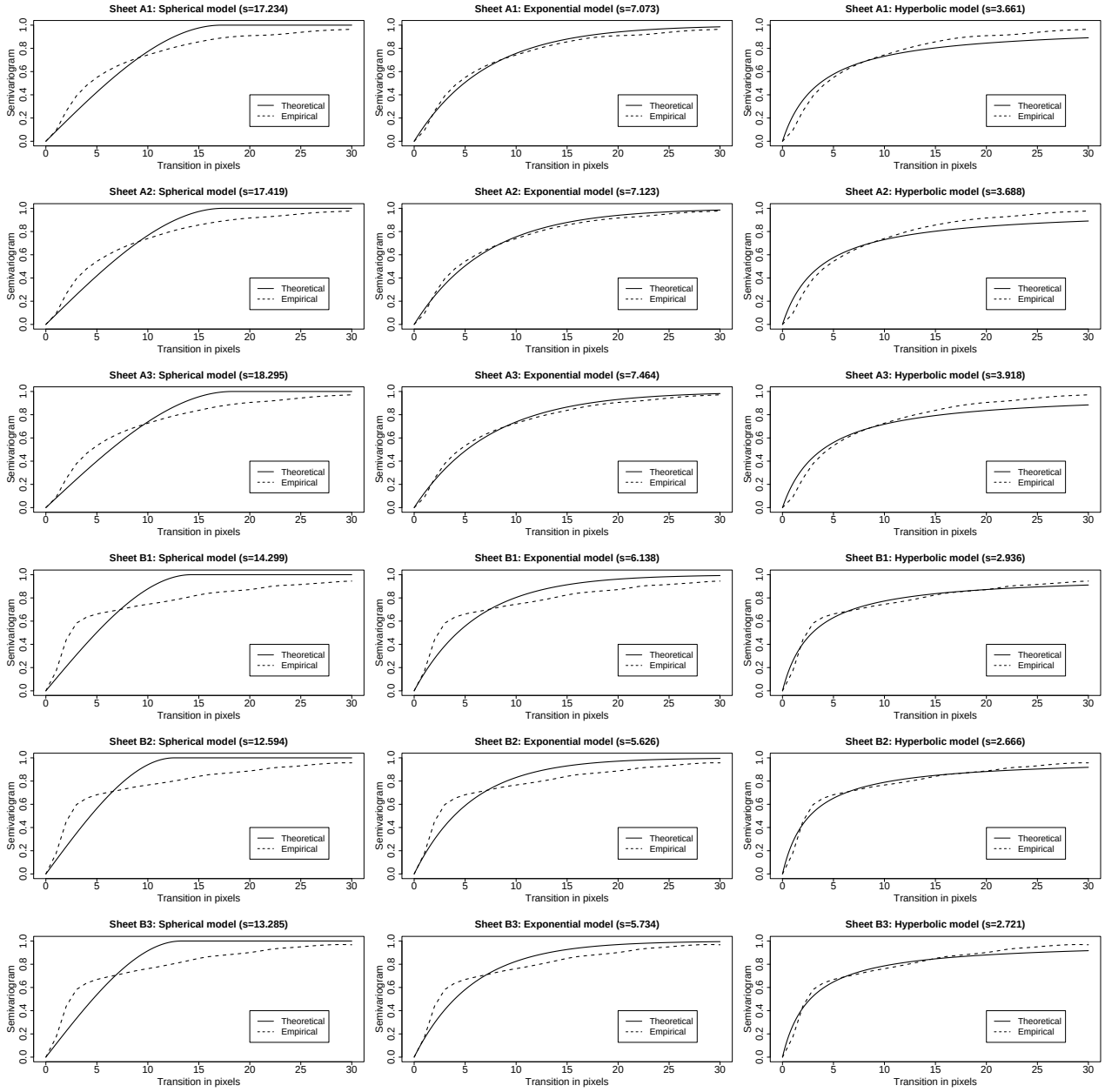


FIGURE 47: Semivariograms for mass distributions. Columns indicate the employed parametric model: spherical (left), exponential (center), and hyperbolic (right). Rows indicate the studied sheet: A1 (top), A2, A3, B1, B2, and B3 (bottom). For interpretation see section 5.2.

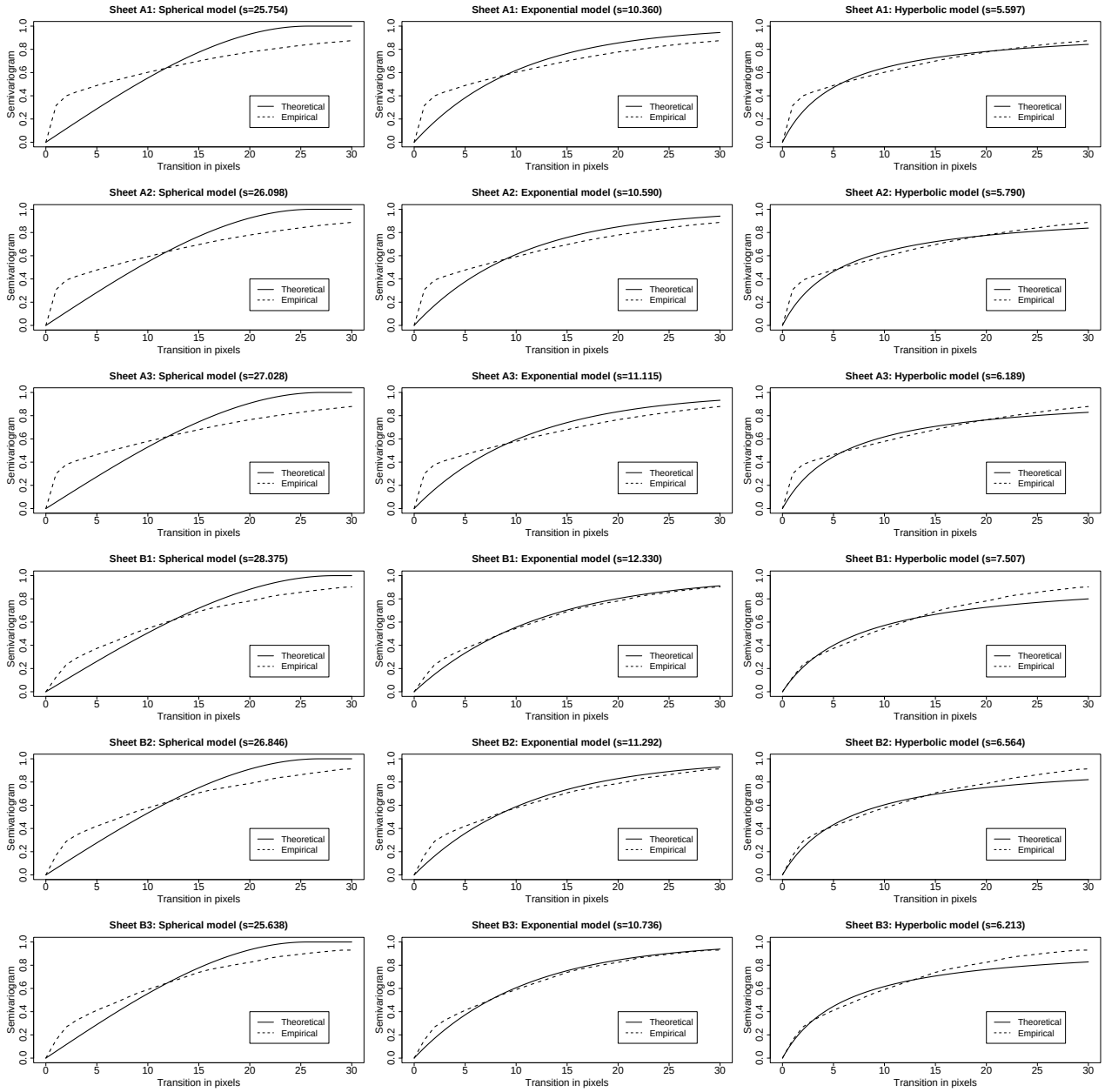


FIGURE 48: Semivariograms for thickness distributions. Columns indicate the employed parametric model: spherical (left), exponential (center), and hyperbolic (right). Rows indicate the studied sheet: A1 (top), A2, A3, B1, B2, and B3 (bottom). For interpretation see section 5.2.

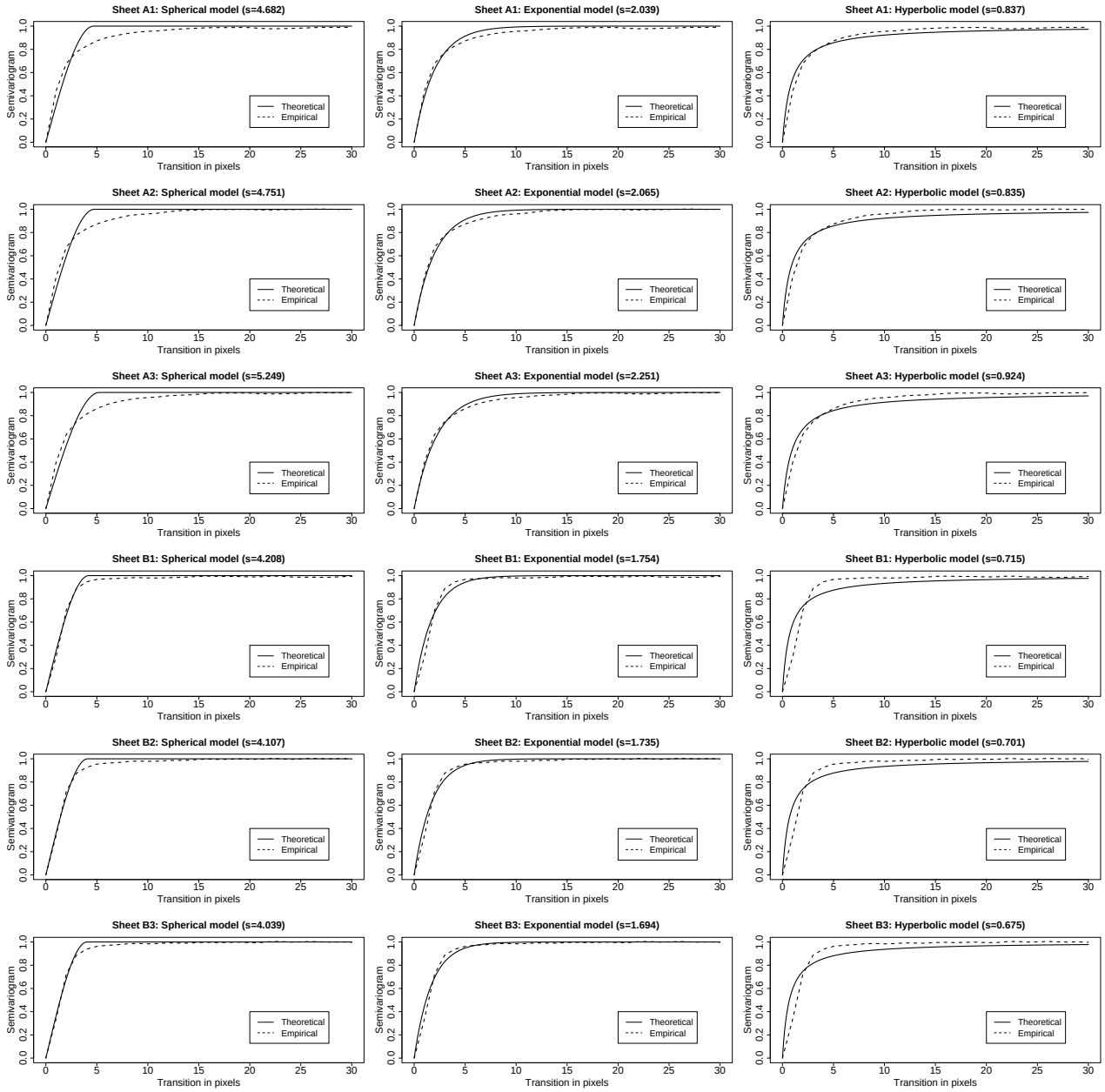


FIGURE 49: Semivariograms for density distributions. Columns indicate the employed parametric model: spherical (left), exponential (center), and hyperbolic (right). Rows indicate the studied sheet: A1 (top), A2, A3, B1, B2, and B3 (bottom). For interpretation see section 5.2.