# Extension of Maximum Autocorrelation Factorization

With application to imaging mass spectrometry data

Daníel Freyr Hjartarson

**TU**Delft
Delft
University of
Technology

Delft Center for Systems and Control

# Extension of Maximum Autocorrelation Factorization

### With application to imaging mass spectrometry data

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Systems and Control at Delft University of Technology

Daníel Freyr Hjartarson

Wednesday 9$^{\text{th}}$ October, 2019

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

TUDelft Delft University of Technology

DCSC

# Abstract

Multivariate images are built up by measuring multiple features or variables simultaneously while recording a measurement's location. An example of such images is Imaging Mass Spectrometry (IMS) data. IMS is a technique for recording the mass-over-charge ratio of molecules in (biological) samples while also recording the molecules' spatial location. High dimensionality in multivariate images (e.g. many recorded features per pixel) often makes direct human interpretation infeasible and computational analysis impractical. For this reason, unsupervised and data-driven factorization is often applied prior to any exploration of the data, with the goal of reducing its dimensionality. However, one of the more promising factorization methods for multivariate images, Maximum Autocorrelation Factorization (MAF), still depends on some input from the user.

Unlike most factorization methods, that focus solely on the spectral content of the observations, MAF also utilizes the spatial structure of the input data to generate matrix factors that try to capture both spatial and spectral patterns in the data. The factors of MAF represent the content of a matrix, ranked according to spatial autocorrelation, i.e. how rapidly they vary spatially. The idea for application of MAF in a bioimaging context is that naturally occurring, signals tend to form larger, more uniform areas and therefore change more slowly spatially, compared to noisy, non-biological measurement patterns. Since MAF factors are ordered according to autocorrelation, noisy measurement (with low autocorrelation) tend to get demoted in the order of factors, effectively separating them from biological data components (with positive autocorrelation). The goal of this thesis is to build upon the MAF algorithm and remove the current need for user input, making an extension of MAF.

This novel factorization method is named Extended Maximum Autocorrelation Factorization (EMAF). Similarly to MAF, EMAF is invariant to linear transformations, utilizes spatial and spectral information of the dataset to determine its factors, and produces uncorrelated factors at all distances under certain conditions. The EMAF algorithm is fully unsupervised and produces factors ranked according to spatial autocorrelation. Unlike MAF, EMAF does not unnecessarily promote spatial artifacts oriented in one particular direction over other directions. The exact formulation of EMAF, derivations of the mentioned traits, and practical examples of EMAF are found in the following thesis. Preliminary experiments show that EMAF returns factors of improved quality compared to MAF.

# Table of Contents

# List of Figures

# List of Tables

# Preface

This thesis and the whole master studies have been an interesting journey. I have learnt a whole lot both technically and about myself. For that I wish to thank a few people without whom, my graduation probably would not have been a reality.

First I wish to thank my supervisor, Raf Van de Plas, for helping me through this thesis project. Thank you for your good guidance, valuable lessions and help. You backed me up even when I lost hope in the whole project.

Secondly I wish to thank all my friends in Delft. You made my stay in Delft enjoyable and fun. Introducing me to different views on the world, different cultures and habits. You taught me valuable lessons, in your own way and helped me develop as a person. Thank you.

Lastly I wish to thank my family for always supporting me and being by my side. Ég er ykkur öllum ævinlega þakklátur, takk fyrir allt saman.

I have fond memories of this period in my life and will think back to it with a smile. I am grateful for all of you. Thank you!

Delft, University of Technology                                      Daníel Freyr Hjartarson
Wednesday 9$^{\text{th}}$ October, 2019

"Sá einn veit
er víða ratar
ok hefr fjölð of farit,
hverju geði
stýrir gumna hverr,
sá er vitandi er vits."

— *Gestaþáttr, Hávamál*
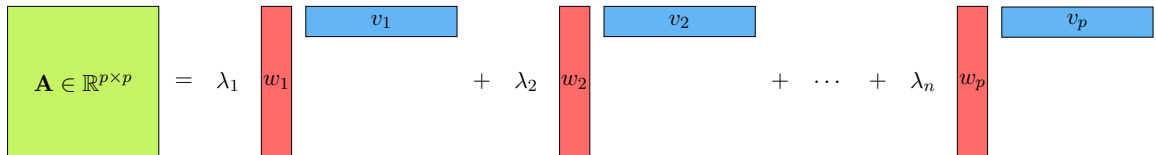
# Chapter 1

# Introduction

Multivariate, spatially structured data or multivariate images are the result of measuring multiple features or variables simultaneously along with recording their measurement location. An example of this type of data is Imaging Mass Spectrometry (IMS) data [1], the technique of recording ions in multiple mass-over-charge ranges simultaneously and their locations. This technique is currently being used to study changes in the proteome due to diseases or tumors [2], to achieve a better understanding of brain structure [1], and to better understand fundamental biological concepts [3]. A common problem with these types of datasets is that the high dimensionality makes the datasets impractical for direct human interpretation. For example in [4] a dataset with 2611 images or peaks is being analysed, each image being $\approx 20.000$ pixels. This can also cause impractically long computation or processing times when the data is being processed [5, 6].

Matrix factorization or matrix decomposition [7] is the operation of decomposing a matrix into matrix products. This procedure is illustrated in Figure 1-1. Matrix factorization plays a role in different types of problems, for example outlier removal [8] and dimensionality reduction [9].



**Figure 1-1:** An example of matrix factorization. Using eigenvalue decomposition a $p \times p$ matrix, $\mathbf{A}$, can be broken up into a weighted sum of vector products, decomposing the original matrix into $p$ factors. Note that $w_i \in \mathbb{R}^{p \times 1}$, $v_i \in \mathbb{R}^{1 \times p}$, and $\lambda_i \in \mathbb{C}$ for all $i \in [1, 2, \ldots, p]$.

Various factorization methods have been applied to IMS data, one of the more common ones being Principal Component Analysis (PCA) [9]. PCA is applied to IMS data in [10, 11, 12, 13] for example. Generally speaking, dimensionality reduction is the problem of finding a reduced set of variables which can be used to represent the original dataset with minimal loss of

information. PCA can be used for dimensionality reduction, where the problem is to create a reduced set of variables that is able to capture the variance present in the original variables [9].

It has previously been demonstrated that utilizing spatial information available in data can reveal patterns that were previously hidden [14, 15, 16, 17]. Maximum Autocorrelation Factorization (MAF) [18] is a factorization method that takes spatial structure of the data into account. MAF incorporates the spatial information by creating the matrix factors based on the spatial covariance matrix. As MAF utilizes the spatial structure of the data, it is a strong candidate for dimensionality reduction of IMS data.

The idea behind utilizing the spatial information available in the data is to encourage discovery of biological patterns, or simply patterns containing a signal rather than noise. Spatially noisy patterns will for example often be single pixel artifacts, like salt-and-pepper noise. Signals, especially biological, on the other hand will tend to form larger areas with uniform values. This uniformity can be measured and quantified by spatial autocorrelation. This is exactly what MAF tries to make use of.

One problem that arises with MAF is the fact that the algorithm is not always invariant of its input parameter, the so called shift parameter. In this thesis, the main goal is to address the tuning of the shift parameter. To alter the MAF algorithm in a way that makes the algorithm independent of both the length and direction of the shift parameter, hopefully returning factors which are at least of the same, if not better, quality than the standard MAF algorithm. This alteration would make the MAF algorithm fully unsupervised, as now it is dependent on an input from the user. This altered version of MAF will be called Extended Maximum Autocorrelation Factorization (EMAF). The goal is to derive EMAF in a way which does not compromise any of the attractive traits of the original MAF. EMAF should also be invariant to linear transformations and produce fully uncorrelated features in the same special cases as MAF.

Along with developing an extended version of MAF, the following research questions will examined:

- How do MAF factors vary with different lenghts of the shift parameter?

- How do MAF factors vary with different orientations of the shift parameter?

- How does the shift parameter influence results when MAF is applied on a practical dataset, e.g. with dimensionality reduction in mind?

These research questions will illustrate the current disadvantages of MAF and thereby give guidance on how to extend the algorithm to remove the need for user input.

In this thesis, we will start by covering some fundamental concepts. We will start with the basics of IMS, since IMS will be used as a case study for the altered algorithm. We will then move on to geostatistics as MAF is based upon ideas from the field of geostatistics. Finally, we finish the fundamentals chapter with an overview of factorization methods, building up to and finishing with MAF.

The main matter of the thesis will be making alterations to the MAF algorithm, and deriving the EMAF algorithm, a method which takes more than a single shift parameter into account

and that is fully unsupervised. The performance of EMAF will then be assessed against PCA and standard MAF in several experiments, which will illustrate both the advantages and disadvantages of each of these factorization methods.

In the final chapter, a few concluding remarks are given about the work carried out in this thesis and we propose ideas on future work and improvements.

# Chapter 2

# Fundamentals

## 2-1 Imaging Mass Spectrometry

Imaging Mass Spectrometry (IMS) is a technique for recording the spatial location of electrically charged molecules or ions in tissue specimens while measuring their mass-over-charge ratio. Combining the chemical specificity and simultaneously detection of multiple analytes from mass spectrometry with imaging capabilities, IMS enables scientists to examine the content of a tissue sample without the requirement of any prior labels or assumptions. With improvements in sensitivity and availability, IMS has become a popular technique for biochemically analyzing tissue samples [1].

IMS measurements can be carried out using different preparation techniques and instruments. To get a general overview of the measurement techniques, relevant for the topic of this review, the different procedures involved will be examined briefly. First, we look at how the spatial information is retrieved. Second, we examine how the mass-over-charge measurements are obtained. Finally, we discuss the resulting data and how it is generally handled.

### 2-1-1 Spatial Measurements

There are two main approaches to obtaining spatial information in IMS: microscope mode and microprobe mode. Both methods result in a 2D pixel grid of measurements. The methods differ in terms of what type of sensor is used for registering the spatial location of the ions, also differing in attainable spatial resolution. Both methods are illustrated in Figure 2-1.

In the microprobe setting, a focused ionizing laserbeam is used to analyze a relatively small area of a sample. The location of the laserbeam is registered when fired, which is used as the spatial location of the molecule(s) measured. To cover the whole sample, the laser is shifted by a set distance after each measurement, following a grid pattern. This means that the shift of the laserbeam determines the spatial resolution of the dataset. A downside of this method is that all spatial information within the spatial location where the laserbeam hits is lost, such that higher resolution than the laserbeam size can not be obtained.

In the microscope setting, the ions are released from the sample, using a laser, and land on a position sensitive sensor. This makes the measurements invariant of the ionization beam size. The pixel sizes are determined by the resolution of the sensor. This method can conserve spatial information within the ionizing beam, depending on the sensor resolution [1].



**(a)** Microprobe mode. The position of the laser-beam is registered, after each measurement the laserbeam is shifted by a set distance, forming a grid pattern covering the whole specimen.

**(b)** Microscope mode. The relative position of ions is maintained until they are recorded by the position sensitive sensor, registering the ions spatial locations.
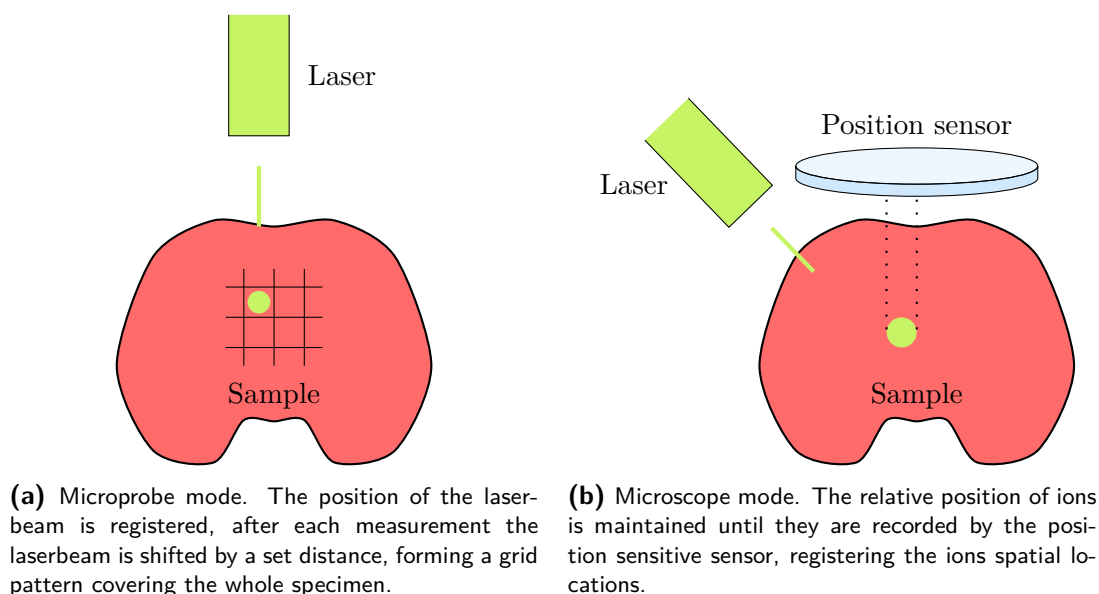
**Figure 2-1:** A comparison of the microprobe- and microscope measurement configurations. Simplified illustrations showing the two different methods used in IMS to obtain spatial measurements.

### 2-1-2   Mass-over-charge Measurements

The mass-over-charge measurements in IMS experiments are usually obtained by ionizing and releasing the ions from a sample or specimen, which for example can be biological tissue such as, brain matter, or a section of a spinal cord. The two most common techniques for ionization are Matrix-Assisted Laser Desorption/Ionization (MALDI) and Secondary Ion Mass Spectrometry (SIMS) [19].

In SIMS measurements, a narrow ion beam ($50\,\text{nm}$ spot size beams are commercially available) is fired at a sample to extract and ionize molecules. SIMS measurements usually give a relatively high spatial resolution but have a much lower ion mass-over-charge range ($m/z <$ $500$) than MALDI. With MALDI measurements, a sample section is coated with a chemical matrix. This matrix embeds the molecules of the sample, forming crystals. The coating or deposition is usually done by spraying the matrix across the specimen as evenly as possible. A laserbeam ($\approx 10\,\mu\text{m}$ spot size are commercially available) is fired at the sample, ejecting ions at a certain binding energy, controlled by the pulse length of the laser [1]. The data used in this thesis was collected using the MALDI process.

The MALDI procedure is shown in Figure 2-2. First, the sample is placed on a piece of glass. The sample is then coated with a chemical matrix, for example sinaptic acid or 2,5-dihydroxyacetphenone. Finally, a laserbeam is fired at the sample, releasing and ionizing the embedded molecules. After these steps, the mass-over-charge measurements take place.
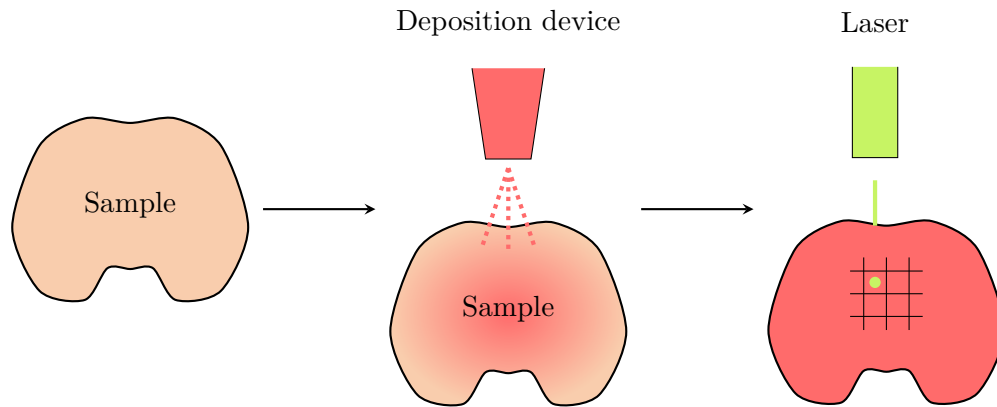
**Figure 2-2:** The MALDI specimen preparation procedure. An illustration of a sample being coated with a chemical matrix, as evenly as possible. Then a laserbeam is used to desorp and ionize molecules from the sample.

The mass-over-charge value of the extracted ions is usually measured using electro-magnetic fields. Two common techniques are Time of Flight (ToF) [1] and Fourier Transform Ion Cyclotron Resonance (FT-ICR) analysers [20]. In ToF measurements, ions are accelerated by a known electro-magnetic field through a field-free region. This field-free region covers a set distance $d$ and the ions are timed traveling this distance. A simple relation between the measured time and the mass-over-charge value can now be found by looking at the energy equilibrium of the ion. We have

$$E_p = E_k. \tag{2-1}$$

$$qU = \frac{1}{2}mv^2, \tag{2-2}$$

$$qU = \frac{1}{2}m\left(\frac{d}{t}\right)^2, \tag{2-3}$$

$$\implies t = \frac{d}{\sqrt{2U}}\sqrt{\frac{m}{q}}, \tag{2-4}$$

where $t$ is the measured time it takes the ion to travel the distance $d$, $U$ is the electric potential of the electro-magnetic field, and $\frac{m}{q}$ is the mass-over-charge of the measured ion. The ion charges can also be expressed as a multiple of the elementary charge, the electric charge of a single proton, that multiple is denoted with a $z$. It should be noted that this is a simple example of how ToF operates, neglecting for example the initial potential energy of the ion and aerodynamic drag of the ion. Now, to produce a spectrum, as shown in Figure 2-3, the ions that fall within a certain $\frac{m}{z}$-range, or $\frac{m}{z}$-bin, are counted [1, 21].

Another technique to measure the mass-over-charge value of the extracted ions is FT-ICR. In an FT-ICR setup, the ions are trapped inside an electro-magnetic field, spinning around in a circular orbit. An FT-ICR instrument will consist of three pairs of electrodes, two pairs to generate electro-magnetic fields and one pair to detect ions. An example of the setup of these electrode pairs is shown in Figure 2-4.

One of the electrode pairs functions as an ion trap, generating an electro-magnetic field that accelerates the ions into a circular orbit. The angular velocity of the ions can be related to
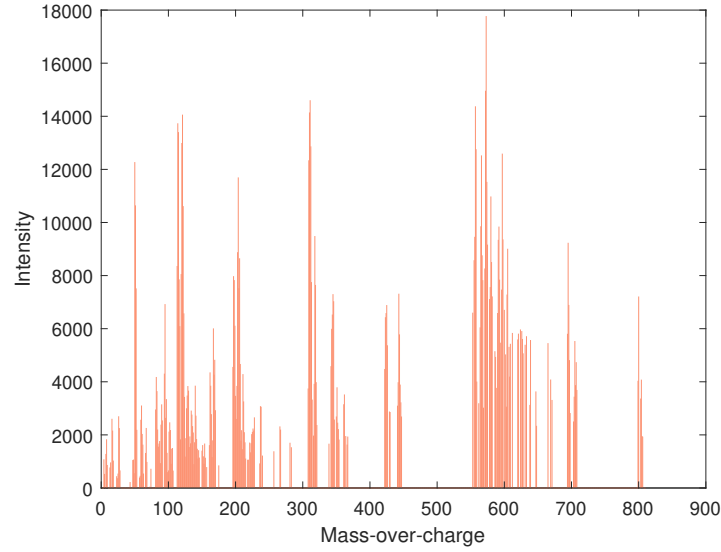
**Figure 2-3:** An example of a pixel specific mass spectrum. The intensity or ion count is plotted on the y-axis and the mass-over-charge value of ions is on the x-axis.

their mass-over-charge value with

$$\omega = B\frac{q}{m}, \tag{2-5}$$

where $\omega$ is the cyclotron frequency of the ion, $B$ is the strength of the electro-magnetic field, and $q/m$ is the inverse mass-over-charge of the ion [20]. The ions will now orbit in circles with radius,

$$r = \frac{1}{qB}\sqrt{2mkT}, \tag{2-6}$$

where $q$ is the ion's charge, $m$ is the ion's mass, $k$ is the Boltzman constant, $T$ is the environmental temperature, and $B$ is the strength of the applied electro-magnetic field. This orbital radius will generally be small enough that the detecting electrodes will not be able to detect the ions without the ions being excited first.

The second electrode pair is used to generate another electro-magnetic field, perpendicular to the ion trap field. This excitation field is used to increase the kinetic energy of the trapped ions, or excite them, to achieve three things: (1) increasing their orbital radius to an extent that they can be detected; (2) increasing their orbital radius to eject them from the instrument; and (3) cause ion and/or ion-molecule dissociation. This electro-magnetic field is set to promote a certain cyclotron frequency that excites only ions within a certain $m/z$-range. This way a chosen pack of ions can be excited to be detected or ejected [20].

The last pair of electrodes is used to indirectly count the ions that are detectable in the instrument. This is done by using a relation between the amplitude of the current and the number of ions detectable in the trap. The current induced by the orbiting ions is recorded by inducing a current in the detection electrode pair.
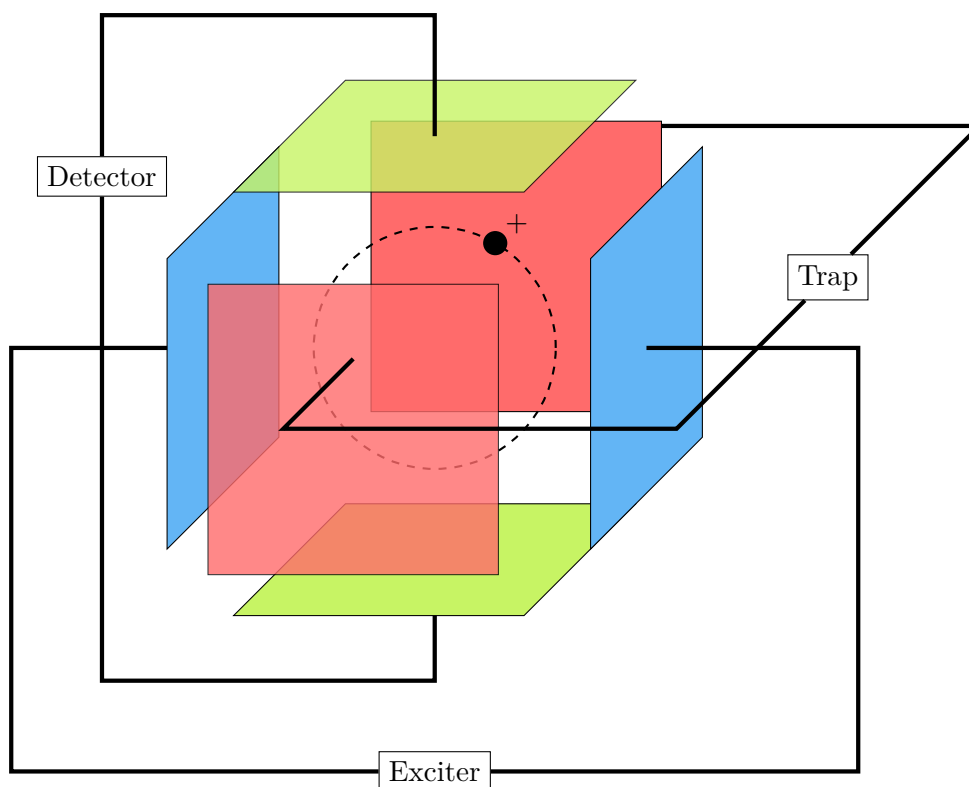
**Figure 2-4:** An example electrode setup of an FT-ICR instrument. One pair is used to trap the ions in a circular motion. Another pair, the exciter, is used to excite the ions, increasing their orbital radius or ejecting the ions out of the instrument. The last electrode pair, the detector, is used to measure the current amplitude induced by the orbiting ions.

Instead of exciting the ions at only a single frequency at a time, the excitation field can be swept across all frequencies. The time domain measurements of the current are then Fourier transformed, yielding the frequency or the angular velocity of the ions. These angular velocities can then be related to the mass-over-charge of the ions using equation 2-5 which then results in the mass(-over-charge) spectrum [22].

### 2-1-3 Dataset

IMS datasets are typically represented as a matrix, with one dimension storing the spatial locations and the other dimension encoding the mass-over-charge values ($m/z$-dimension). For interpretation, it is common to reorder the spatial dimension into two separate spatial dimensions ($n_x$ and $n_y$), forming a 3-dimensional tensor as shown in Figure 2-5. Examining this 3-dimensional tensor at a set value along the $m/z$-dimension results in an ion image [6].

Ion images are false color visualizations of the spatial distribution and abundance for a given $m/z$-value, showing how ions of a certain mass-over-charge ratio are distributed across the sample. Figure 2-5b shows an example of an ion image [6]. These images are obtained by looking at a single column of the data matrix.

The size of IMS datasets will often contain high pixel counts as well as a large mass-over-charge range. For a sample of $1\,cm^2$ and a spatial resolution of $10\,\mu m$, we get an image of
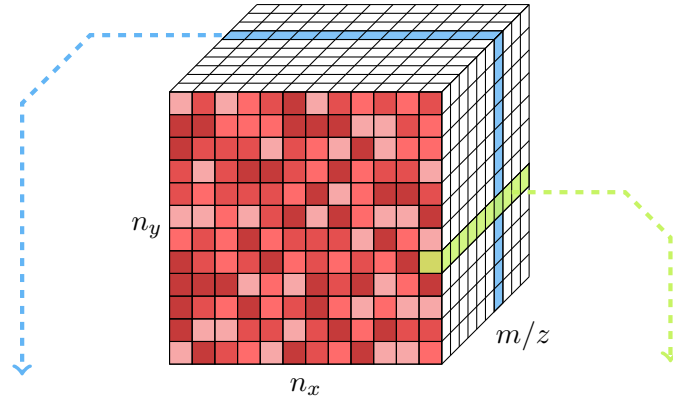
**(a)** An example of an IMS data-tensor.
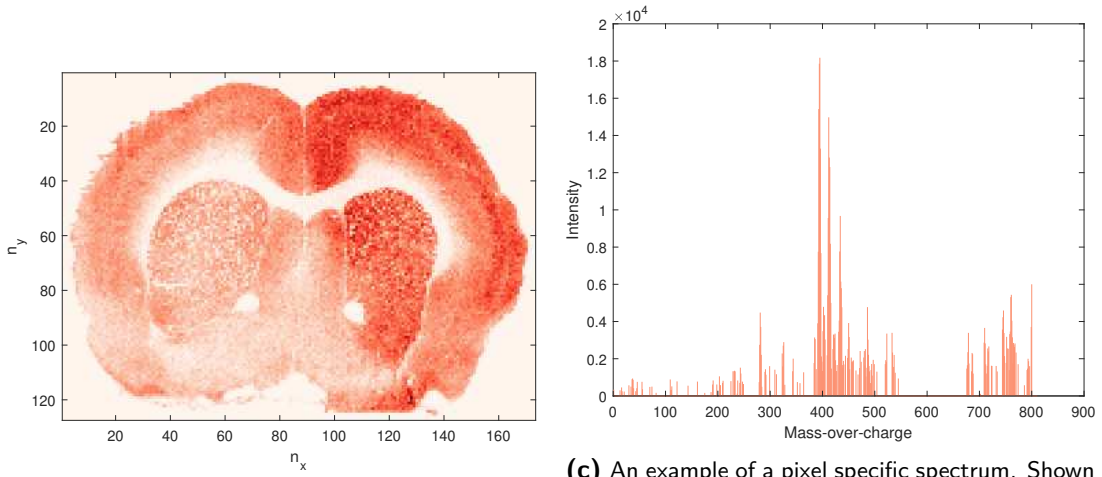


**(b)** An example ion image. Shown in blue, where the tensor is examined for a single $m/z$-value.



**(c)** An example of a pixel specific spectrum. Shown in green, where a single pixel is examined over all $m/z$-values.

**Figure 2-5:** An example of a datatensor show in Figure 2-5a. If the tensor is sliced long the $m/z$-axis (marked in blue) the resulting data is an ion image, shown in Figure 2-5b. If a single pixel is examined (marked in green), the result is a pixel specific spectrum, shown in Figure 2-5c.[22]

$10000 \times 10000$ pixels, often with large mass-over-charge value ranges. This results in datasets which are impractical for direct human interpretation due to dimensionality. In [4] for example a dataset with 2611 images, or peaks, is analysed, each image being $\approx 20.000$ pixels. This is one of the main reasons why feature extraction and factorization of IMS data is increasingly studied [23, 19, 1, 24, 2, 25].

## 2-2    Geostatistics

Representing IMS data mathematically, a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ with $n$ observations of $p$ variables, that is $n$ pixels and intensity values for $p$ (variables) mass-over-charge ratios, can

be constructed. This can be written as

$$\mathbf{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \ldots & X_{1,p-1} & X_{1,p} \\ X_{2,1} & X_{2,2} & \ldots & X_{2,p-1} & X_{2,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n-1,1} & X_{n-1,2} & \ldots & X_{n-1,p-1} & X_{n-1,p} \\ X_{n,1} & X_{n,2} & \ldots & X_{n,p-1} & X_{n,p} \end{bmatrix} \tag{2-7}$$

$$= \begin{bmatrix} X_1 & X_2 & \ldots & X_p \end{bmatrix}, \tag{2-8}$$

where $X_i \in \mathbb{R}^{n \times 1}$ for $i \in [1, 2, \ldots, p]$. The data matrix is now centered around zero. This is done by subtracting the mean of each observation, arriving at another data matrix $\mathbf{Z}$, which is centered around zero. $\mathbf{Z}$ is defined as

$$\mathbf{Z} = \begin{bmatrix} X_{1,1} - \mathrm{E}\,[X_1] & X_{1,2} - \mathrm{E}\,[X_2] & \ldots & X_{1,p-1} - \mathrm{E}\,[X_{p-1}] & X_{1,p} - \mathrm{E}\,[X_p] \\ X_{2,1} - \mathrm{E}\,[X_1] & X_{2,2} - \mathrm{E}\,[X_2] & \ldots & X_{2,p-1} - \mathrm{E}\,[X_{p-1}] & X_{2,p} - \mathrm{E}\,[X_p] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n-1,1} - \mathrm{E}\,[X_1] & X_{n-1,2} - \mathrm{E}\,[X_2] & \ldots & X_{n-1,p-1} - \mathrm{E}\,[X_{p-1}] & X_{n-1,p} - \mathrm{E}\,[X_p] \\ X_{n,1} - \mathrm{E}\,[X_1] & X_{n,2} - \mathrm{E}\,[X_2] & \ldots & X_{n,p-1} - \mathrm{E}\,[X_{p-1}] & X_{n,p} - \mathrm{E}\,[X_p] \end{bmatrix} \tag{2-9}$$

$$= \begin{bmatrix} Z_1 & Z_2 & \ldots & Z_p \end{bmatrix}, \tag{2-10}$$

where $\mathbf{Z} \in \mathbb{R}^{n \times p}$ and $Z_i \in \mathbb{R}^{n \times 1}$ for $i \in [1, 2, \ldots, p]$. This way,

$$\mathrm{E}\,[Z_i] = 0 \tag{2-11}$$

for all $i \in [1, 2, \ldots, p]$.

### 2-2-1   Covariance

Covariance is a measure of variance between two variables. For two given random variables, $Z_i$ and $Z_j$, the theoretical covariance between them is defined as

$$\mathrm{Cov}\,[Z_i, Z_j] = \mathrm{E}\,[(Z_i - \mathrm{E}\,[Z_i])(Z_j - \mathrm{E}\,[Z_j])] \tag{2-12}$$

which is

$$\mathrm{Cov}\,[Z_i, Z_j] = \mathrm{E}\,[Z_i Z_j], \tag{2-13}$$

if the variables have been centered. as defined in [26]. Inspecting equation 2-13, two numbers are being multiplied, these two numbers are the variable values with the mean subtracted. The results of this subtraction are called residuals. The residual is a measure of how the variable spreads or disperses around its mean. So the covariance is a joint measure of how variables spread around their mean. If both variables have a large spread around their mean, the covariance will be large. Similarly, if both variables have a small spread around their mean, that will result in a small covariance.

To get the covariance from sample data, an experimental covariance can be defined as

$$\text{Cov}\left[Z_i, Z_j\right] = \frac{1}{n}\sum_{\alpha=1}^{n}\left(Z_{\alpha,i} - \text{E}\left[Z_i\right]\right)\left(Z_{\alpha,j} - \text{E}\left[Z_j\right]\right) = \frac{1}{n}\sum_{\alpha=1}^{n}Z_{\alpha,i}Z_{\alpha,j}. \qquad (2\text{-}14)$$

Equation 2-14 can be rewritten to a matrix form, the experimental covariance matrix $\text{Cov}\left[\mathbf{Z}\right] \in \mathbb{R}^{p\times p}$, defined as

$$\text{Cov}\left[\mathbf{Z}\right] = \frac{1}{n-1}\mathbf{Z}^T\mathbf{Z}. \qquad (2\text{-}15)$$

It is worth pointing out that the diagonal of the covariance matrix is the covariance of a variable with itself, $\text{Cov}\left[Z_i, Z_i\right]$, which is the variance of said variable, that is $\text{Cov}\left[Z_i, Z_i\right] = \text{Var}\left[Z_i\right]$. This is the reason the covariance matrix is often referred to as a variance-covariance matrix [27].

Covariance is a dissimilarity measure, assigning a numerical value to how two variables change jointly, or as described above, how they spread around their mean. It is often desirable to compare variables with different units of measure and scale, age, and salary for example. Due to the different units and scales, it can be necessary to standardize measurements onto an arbitrary unit of measure. Looking back at our example, a small fluctuation in salary might have much higher influence on the values in the covariance matrix than the age due to the different scales of the two. That is the fluctuation of a salary might be on the scale of hundreds to thousands while age might only fluctuate by ones to tens. The covariance matrix of this example dataset would predominantly describe the fluctuations of the salary, and the age would barely be noticed. Standardization of the measurements would standardize the covariance and return correlation coefficients, which is a standardized dissimilarity measure between variables. But when working with measurements of the same unit and scale, standardizing measurements is an unnecessary and often undesirable step. Standardizing measurements of the same scale would make interpretation harder, since all measurements would now be on an arbitrary, unitless scale with no physical meaning [9]. For that reason, when working with IMS data, we will stick to examining the covariance instead of the correlation coefficients.

### 2-2-2   Spatial Covariance Function

The idea of covariance can be extended by adding spatial information to it, resulting in spatial covariance. Spatial covariance is a dissimilarity measure between two variables, spatially located somewhere relative to each other. This spatial covariance is often a function of the two locations of the variables being examined. Let us start by looking at some notation and terminology before defining spatial covariance further.

Let $\mathbf{Z}(\mathbf{x})$ denote the spatial location $\mathbf{x} \in \mathcal{D} \subset \mathbb{Z}^2$ of the data in $\mathbf{Z}$, where $\mathcal{D}$ is a finite region of allowed positions. Note that $\mathcal{D}$ is commonly referred to as the lag space. In the case of IMS data, these allowed locations are the pixels. $\mathcal{D}$ can be seen as a sampled description of the specimen being measured. This is shown in Figure 2-8.

The spatial covariance function is defined as

$$C_{i,j}(\mathbf{h}) = \text{Cov}\left[\left(Z_i(\mathbf{x}) - \text{E}\left[Z_i(\mathbf{x})\right]\right), \left(Z_j(\mathbf{x}+\mathbf{h}) - \text{E}\left[Z_j(\mathbf{x}+\mathbf{h})\right]\right)\right] \qquad (2\text{-}16)$$

$$= \text{Cov}\left[Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h})\right], \qquad (2\text{-}17)$$
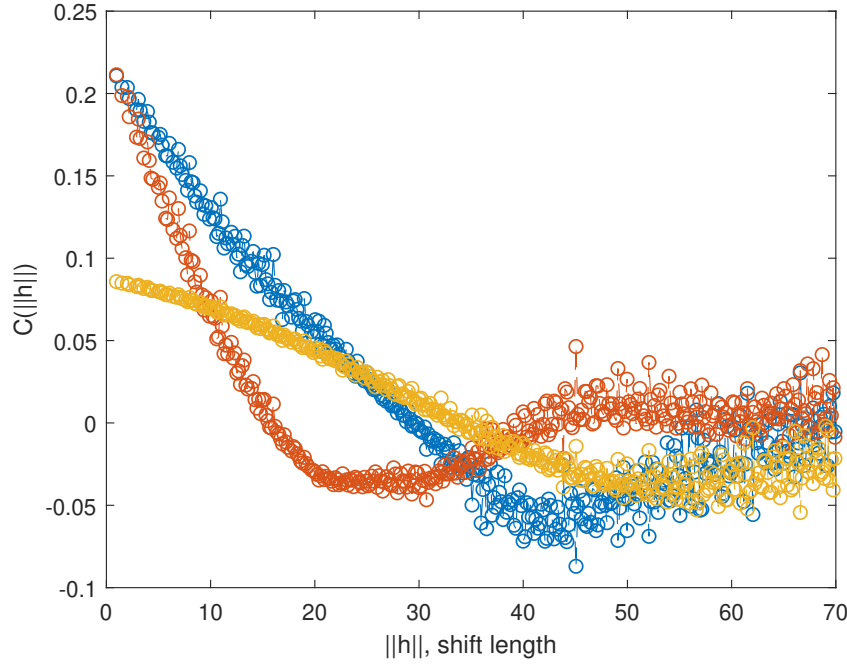
**Figure 2-6:** An example 2D visualization of a spatial covariance function. In this example the spatial covariance of three different variables is shown on the same plot. Generally the 2D visualization is generated by averaging across the direction dimension, $\theta$, of the shift parameter. This explains the spread in the points as the length of the shift parameter increases.

where $C_{i,j}(\mathbf{h}) \in \mathbb{R}$ for all values $i,j \in [1,2,\ldots,p]$, and $\mathbf{h} \in \mathbb{Z}^2$. The covariance between two variables $Z_i$ and $Z_j$ is being examined, more specifically this is called the spatial cross-covariance function. If a variable is being compared to itself, $C_{i,i}$, the resulting spatial covariance function is called a spatial auto-covariance function, spatial direct-covariance function, or simply a spatial spatial covariance function. Like the more traditional covariance, the spatial covariance function is a measure of how two variables vary together, only spatially. That is, examining the covariance between two observations at two different spatial locations [28].

It is worth stating at this point that the spatial shift $\mathbf{h}$ is a vector. It is a vector linking two spatial points within $\mathcal{D}$ together. The vector $\mathbf{h}$ will have both a length or magnitude and direction tied to it.

The definition of the covariance function can be used to define the spatial correlation coefficient function. The spatial correlation coefficient is defined as

$$\rho_{i,j}(\mathbf{h}) = \frac{C_{i,j}(\mathbf{h})}{C_{i,j}(\mathbf{0})}, \tag{2-18}$$

where $\rho_{i,j}(\mathbf{h}) \in [-1,1] \subset \mathbb{R}$ for all values $i,j \in [1,2,\ldots,p]$, and $\mathbf{h} \in \mathbb{Z}^2$.

These definitions can now be assembled into matrix forms, covering the cases dealing with multiple variables. This is achieved by ordering the newly defined functions into matrices according to indices. This way, we end up with 2 matrices $\mathbf{C}(\mathbf{h}) \in \mathbb{R}^{p \times p}$ and $\rho(\mathbf{h}) \in [-1,1]^{p \times p} \subset \mathbb{R}^{p \times p}$, for all values $\mathbf{h} \in \mathbb{Z}^2$, which are defined as follows
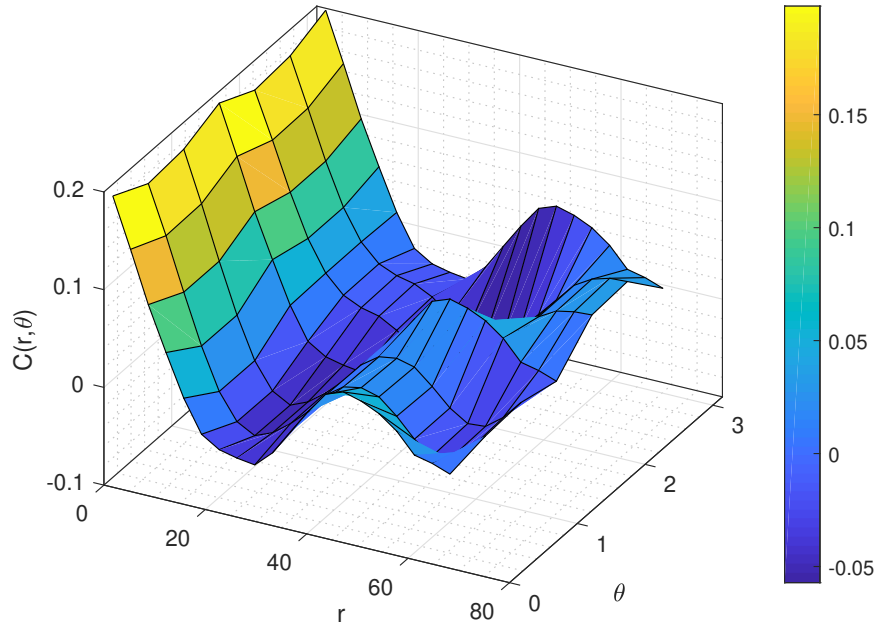
**Figure 2-7:** An example 3D visualization of a spatial covariance function. This example shows the spatial covariance function of a single variable. Notice how the covariance changes both with the direction, $\theta$, and length of the shift parameter $\mathbf{h}$, this means that the variable is anisotropic.
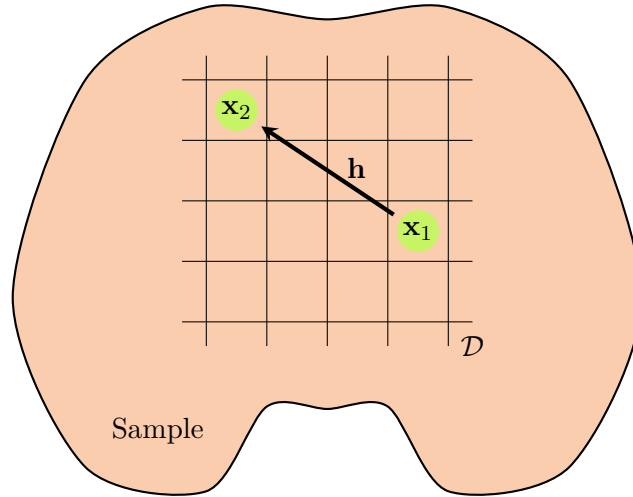


**Figure 2-8:** A measured sample being shifted by a vector $\mathbf{h}$. The sample is divided into a grid of pixels $\mathcal{D}$ when measured. A location $\mathbf{x}_1$ is shown and how the shift parameter $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$ changes the location of focus. This is analogous to spatially translating the image by the vector $\mathbf{h}$.

$$\mathbf{C}(\mathbf{h}) = \begin{bmatrix} C_{1,1}(\mathbf{h}) & C_{1,2}(\mathbf{h}) & \ldots & C_{1,p}(\mathbf{h}) \\ C_{2,1}(\mathbf{h}) & C_{2,2}(\mathbf{h}) & \ldots & C_{2,p}(\mathbf{h}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p,1}(\mathbf{h}) & C_{p,2}(\mathbf{h}) & \ldots & C_{p,p}(\mathbf{h}), \end{bmatrix} \tag{2-19}$$

$$\rho(\mathbf{h}) = \begin{bmatrix} \rho_{1,1}(\mathbf{h}) & \rho_{1,2}(\mathbf{h}) & \dots & \rho_{1,p}(\mathbf{h}) \\ \rho_{2,1}(\mathbf{h}) & \rho_{2,2}(\mathbf{h}) & \dots & \rho_{2,p}(\mathbf{h}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p,1}(\mathbf{h}) & \rho_{p,2}(\mathbf{h}) & \dots & \rho_{p,p}(\mathbf{h}). \end{bmatrix} \tag{2-20}$$

By using these multivariate extensions along with equation 2-18 arriving at the relation

$$\mathbf{C}(\mathbf{h}) = \mathbf{C}(\mathbf{0})\rho(\mathbf{h}), \tag{2-21}$$

where $\mathbf{C}(\mathbf{0})$ is the covariance matrix of $\mathbf{Z}(\mathbf{0})$ [29].

### 2-2-3 Variogram

The spatial covariance function measures the spatial (dis)similarity between two variables at two spatial locations. The dissimilarity caused by this spatial shift can be examined in isolation by subtracting the value of each variable at the initial point. This results in what is called a variogram. A variogram is a measure of dissimilarity between two variables across a distance. The variogram is defined as

$$\gamma_{i,j}(\mathbf{h}) = \frac{1}{2} \operatorname{Cov}\left[ (Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})), (Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})) \right], \tag{2-22}$$

where $\gamma_{i,j}(\mathbf{h}) \in \mathbb{R}$ for all values of the indices $i, j \in [1, 2, \dots, p]$ and $\mathbf{h} \in \mathbb{Z}^2$. This particular definition is the cross-variogram, but $\gamma_{i,i}$ is called the auto-variogram. The variogram, $\gamma_{i,j}(\mathbf{h}) \in \mathbb{R}$, is an even function, that is

$$\gamma_{i,j}(\mathbf{h}) = \gamma_{i,j}(-\mathbf{h}) \tag{2-23}$$

and the values of the variogram are non-negative, that is

$$\gamma_{i,j}(\mathbf{h}) \geq 0. \tag{2-24}$$

Under certain circumstances the variogram can be related to a covariance function, this will be discussed later [30].

Similarly to the spatial covariance function and the spatial correlation function the variogram can be assembled to a matrix structure. That gives the matrix

$$\gamma(\mathbf{h}) = \begin{bmatrix} \gamma_{1,1}(\mathbf{h}) & \gamma_{1,2}(\mathbf{h}) & \dots & \gamma_{1,p}(\mathbf{h}) \\ \gamma_{2,1}(\mathbf{h}) & \gamma_{2,2}(\mathbf{h}) & \dots & \gamma_{2,p}(\mathbf{h}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p,1}(\mathbf{h}) & \gamma_{p,2}(\mathbf{h}) & \dots & \gamma_{p,p}(\mathbf{h}), \end{bmatrix} \tag{2-25}$$

where $\gamma(\mathbf{h}) \in \mathbb{R}^{p \times p}$ for all values $\mathbf{h} \in \mathbb{Z}^2$. Now since variogram values are all non-negative, the variogram matrix, $\gamma(\mathbf{h})$, positive semi-definite.

In Equation 2-22, the definition of a variogram, note the $\frac{1}{2}$ in front of the covariance. That is there due to convention. The $\frac{1}{2}$ is likely there to get a mathematically neater relationship between the variogram and the spatial covariance function, but it could also be a historical
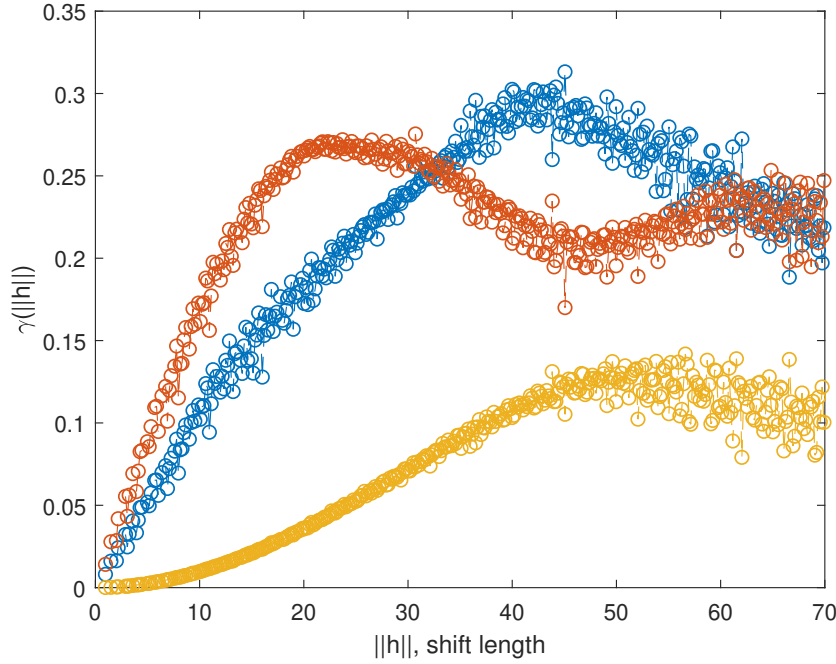
**Figure 2-9:** An example 2D visualization of a variogram. In this example the variogram of three different variables are shown on the same plot. Generally the 2D visualization is generated by averaging across the directional dimension, $\theta$, of the shift parameter. This explains the spread in the points as the length of the shift parameter increases.

confusion. Due to the $\frac{1}{2}$, the variogram is also often called the semi-variogram in literature [31]. In this thesis the name variogram will be used.

In practice, the variogram is calculated by dividing the shift parameter $\mathbf{h}$ into $k$ bins, denoted $\mathfrak{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots \mathbf{h}_k\}$, segmenting the field of available shifts into bins of different shift parameter lengths and angles. This segmentation is shown in Figure 2-11. The variogram is then averaged across each bin or segment, arriving at an experimental estimate of the variogram. This can be written as

$$\gamma_{i,j}^*(\mathbf{h}) = \frac{1}{2N} \sum_{\alpha=1}^{N} \left(Z_i(\mathbf{x}_\alpha + \mathbf{h}) - Z_i(\mathbf{x}_\alpha)\right) \left(Z_j(\mathbf{x}_\alpha + \mathbf{h}) - Z_j(\mathbf{x}_\alpha)\right) \quad \text{with } \mathbf{h} \in \mathfrak{H} \qquad (2\text{-}26)$$

where $N$ is the number of shift parameters which end up in each bin in $\mathfrak{H}$ [32].

If $\gamma_{i,i}^*(\mathbf{h}) = \gamma_{i,i}^*(\|\mathbf{h}\|)$ then $Z_i$ is said to be isotropic, otherwise $Z_i$ is said to be anisotropic. If the variogram does not vary with the angle of the shift parameter $\mathbf{h}$, but only the length the data set said to be isotropic. There are different types of anisotropy and different methods to try and correct the effects of it, as mentioned in [33], [34], and [35].

Looking at a two-dimensional representation of a variogram, the variogram has a few different features. These features are called the range, sill, and nugget. The nugget, also sometimes called the nugget-effect, is the initial value an experimental variogram takes, that means that there is a spatial change in the data at a smaller scale than the resolution of the images allow to be analyzed. The name nugget comes from the fact that scientist used the variogram to
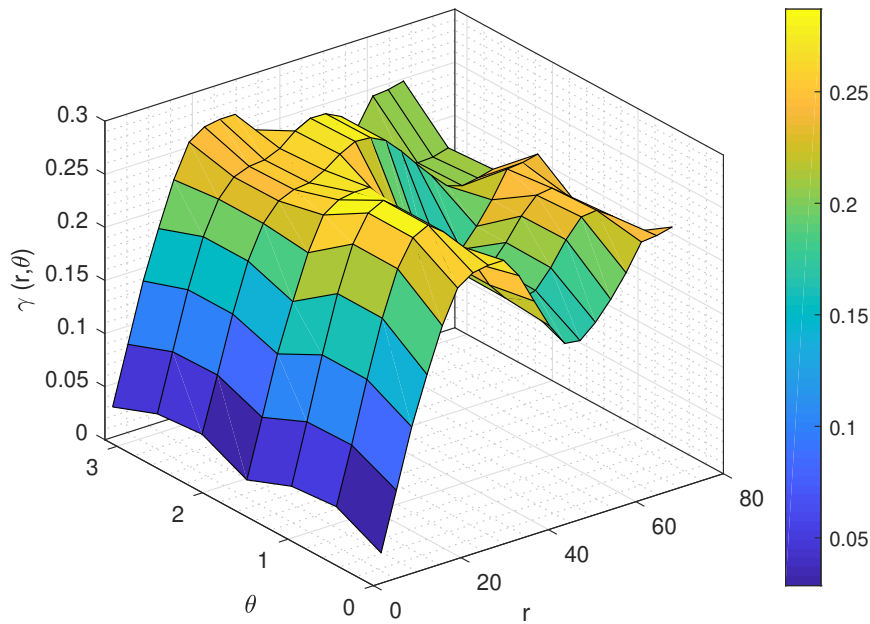
**Figure 2-10:** An example 3D visualization of a variogram. This example shows the variogram of a single variable. Notice how the variogram changes both with the direction, $\theta$, and length of the shift parameter $\mathbf{h}$, this means that the variable is anisotropic.

look for gold in soil, so this non-zero inital value of the variogram would imply that there might be goldnuggets in the sample.

The lowest upper bound of a variogram is called the sill and the distance at which the sill is reached is called the range. Only when a variogram has a sill a covariance function can be related to it. These three terms are visualized in Figure 2-13 [30].

### 2-2-4    Stationarity of First Two Moments

In the following chapter spatial stationarity will be introduced. This is a spatial equivalent to the weak-sense stationarity or wide-sense stationary seen in signal processing. Looking at a dataset $\mathbf{Z}(\mathbf{x}) = \begin{bmatrix} Z_1(\mathbf{x}) & Z_2(\mathbf{x}) & \dots & Z_p(\mathbf{x}) \end{bmatrix}$, the variable $Z_i(\mathbf{x})$ is said to be second order stationary if for all $\mathbf{h} \in \mathcal{D}$

1. The mean $\mathrm{E}\left[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})\right] = m_i(\mathbf{h}) = 0$ is constant.

2. The covariance $C(\mathbf{h})$ has a finite value and is a function of $\mathbf{h}$ and invariant of $\mathbf{x}$.

Another version of stationarity is the intrinsic stationarity of the first two moments (often called intrinsic stationarity), which applies to the variogram rather than the spatial covariance function. The variable $Z_i(\mathbf{x})$ is said to be intrinsic (second order) stationary if for all $\mathbf{h} \in \mathcal{D}$

1. The mean $\mathrm{E}\left[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})\right] = m_i(\mathbf{h}) = 0$ is constant.
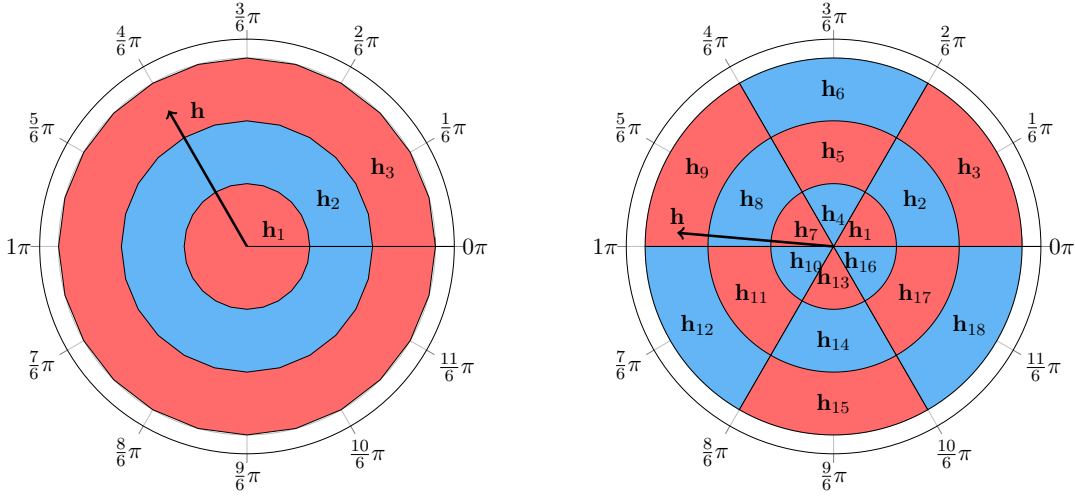
**(a)** An example segmentation of the shift parameter field or lag space. In this case, each segment spans a region of $2\pi$ in terms of the angle and a radial length of 1. An example vector $\mathbf{h}$ is shown, that falls within region $\mathbf{h}_3$. This particular segmentation could be represented by 2D visualization of a variogram.

**(b)** An example segmentation of the shift parameter field. In this case, each segment spans a region of $\frac{2}{6}\pi$ in term of the angle parameter and a radial length of 1. An example vector $\mathbf{h}$ is shown, that falls within region $\mathbf{h}_9$. This segmentation would result in a 3D visualization of a variogram.

**Figure 2-11:** Two example segmentations of the lag space. All values of the variogram calculated using shift parameters within the same region (red and blue) will be averaged, calculating the experimental variogram.

2. The variogram $\gamma_{i,i}(\mathbf{h})$ has a finite value and is a function of $\mathbf{h}$ and invariant with of $\mathbf{x}$.

If a variable is intrinsic stationary, a covariance function can be deduced from the variogram with the relationship

$$\gamma_{i,j}(\mathbf{h}) = C_{i,j}(\mathbf{0}) - C_{i,j}(\mathbf{h}). \tag{2-27}$$

This can also be written as

$$\gamma_{i,j}(\mathbf{h}) = C_{i,j}(\mathbf{0}) - \frac{1}{2}\left(C_{i,j}(-\mathbf{h}) + C_{i,j}(\mathbf{h})\right), \tag{2-28}$$

If the variogram does not have a finite value, it will not have a covariance function related to it. [32, 30]

## 2-2-5   Linear Model of Coregionalization

The measured variables can be considered to be composed of different elementary signals and noises. That can be reflected in the structure of the covariance matrix by writing it as a linear combination of the elementary components. This decomposition is called a linear model of coregionalization (LMC) [36].
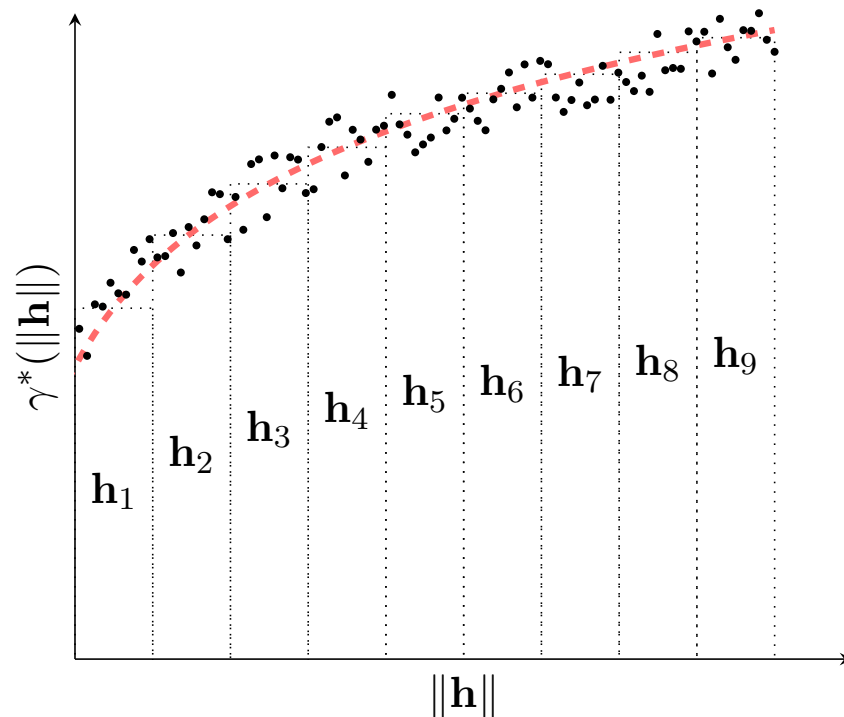
**Figure 2-12:** An illustration of how the experimental variogram is calculated. The variogram can be calculated experimentally by dividing the shift parameters up into bins and averaging out the resulting datapoints. In this example, the shift parameters are divided up into bins only according to their length, similar to what is shown in Figure 2-11a, averaging across all angles for a certain length of shift parameter.
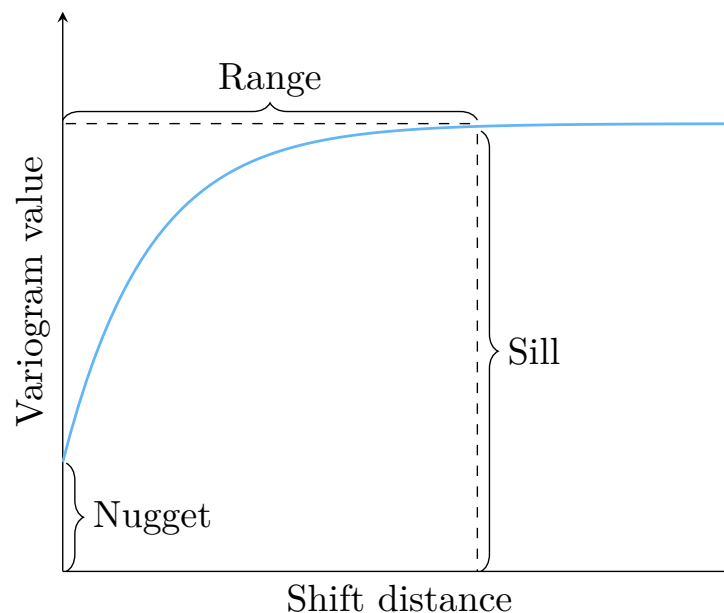


**Figure 2-13:** The definitions of nugget, range and sill illustrated.

Looking at the dataset $\mathbf{Z}(\mathbf{x}) = \begin{bmatrix} Z_1(\mathbf{x}) & Z_2(\mathbf{x}) & \ldots & Z_p(\mathbf{x}) \end{bmatrix}$ from earlier that is assumed to be second-order stationary, each variable $Z_i(\mathbf{x})$ can be decomposed as a sum of $S$ spatially uncorrelated components

$$Z_i(\mathbf{x}) = \sum_{s=1}^{S} Z_i^s(\mathbf{x}), \tag{2-29}$$

where $Z_i(\mathbf{x}), Z_i^s(\mathbf{x}) \in \mathbb{R}^{n \times 1}$. Note that here $s$ is an index, and not a power. For all values of the index $i \in [1, 2, \ldots, p]$ and $s, r \in [1, 2, \ldots, S]$ with $s \neq r$, we require

$$\text{Cov}\left[ Z_i^s(\mathbf{x}), Z_i^r(\mathbf{x} + \mathbf{h}) \right] = 0. \tag{2-30}$$

Comparing the uncorrelated components between two variables we define

$$\text{Cov}\left[ Z_i^s(\mathbf{x}), Z_j^s(\mathbf{x} + \mathbf{h}) \right] = C_{i,j}^s(\mathbf{h}). \tag{2-31}$$

This means spatial cross-covariance can be written as

$$C_{i,j}(\mathbf{h}) = \sum_{s=1}^{S} C_{i,j}^s(\mathbf{h}) = \sum_{s=1}^{S} b_{i,j}^s \rho_{i,j}^s(\mathbf{h}), \tag{2-32}$$

where $b_{i,j}^s \in \mathbb{R}$ and $\rho_{i,j}^s(h) \in \mathbb{R}$ for all $s \in [1, 2, \ldots, S]$. This can be assembled into the matrix form as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}^s \rho^s(\mathbf{h}), \tag{2-33}$$

where $\mathbf{B}^s \in \mathbb{R}^{p \times p}$ and $\rho^s(h) \in \mathbb{R}$ for all $s \in [1, 2, \ldots, S]$.

Similarly the variogram can now be calculated as

$$\text{Cov}\left[ Z_i^s(\mathbf{x} + \mathbf{h}) - Z_i^s(\mathbf{x}), Z_j^s(\mathbf{x} + \mathbf{h}) - Z_j^s(\mathbf{x}) \right] = \gamma_{i,j}^s(\mathbf{h}). \tag{2-34}$$

So we can write

$$\gamma_{i,j}(\mathbf{h}) = \sum_{s=1}^{S} g_{i,j}^s \gamma_{i,j}^s(\mathbf{h}), \tag{2-35}$$

where $g_{i,j}^s \in \mathbb{R}$ and $\gamma_{i,j}^s(\mathbf{h}) \in \mathbb{R}$. Which in the multivariate format becomes

$$\gamma(\mathbf{h}) = \sum_{s=1}^{S} \mathbf{G}^s \gamma^s(\mathbf{h}), \tag{2-36}$$

where $\mathbf{G}^s \in \mathbb{R}^{p \times p}$ and $\gamma^s(\mathbf{h}) \in \mathbb{R}$ for all $s \in [1, 2, \ldots, S]$. All matrices $\mathbf{G}^s$ are required to be positive semi-definite, this is since all variogram values are non-negative.

In practice, this means that every variable of the dataset has to be described using the same set of $S$ variogram or covariance structures. A total of $S$ structures are picked and if one or more of those structures do not apply to a certain variable the corresponding scale coefficients, $b_{i,j}^s$ or $g_{i,j}^s$ depending on whether the covariance or the variogram is being modeled, are set to zero [36, 37].

## 2-2-6 Moran's I

Moran's I, introduced by P. A. P. Moran [38], is a method of estimating whether there is evidence of spatial correlation. It is a measure of the spatial correlation of a two, or higher, dimensional space. In this thesis, Moran's I will only be used in a two-dimensional setting, so that will be the focus of this section.

For a random, two-dimensional, stationary variable $Z \in \mathbb{R}^{n \times 1}$ with $\mathrm{E}\left[Z_i\right] = 0$, Moran's I is defined as

$$I = \frac{n}{\sum_{i=0}^{n} \sum_{j=0}^{n} w_{i,j}} \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} w_{i,j} z_i z_j}{\sum_{i=0}^{n} z_i^2}, \tag{2-37}$$

where $z_i \in \mathbb{R}$ is the $i$-th element of the variable $Z$ and $w_{i,j} \in \mathbb{R}$ is a weighting factor between observations $i$ and $j$. This means that the value of Moran's I is dependent on a weighting matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$. This weighting matrix is used to describe the influence between observations. Often the weighting matrix is setup such that Moran's I only takes the $k$ nearest neighboring observations into account. Note that the weight matrix is a symmetric $n \times n$ matrix, so each column or row represent the influence of one observation to the $n-1$ other observations. An example of such a weighting matrix is shown in Figure 2-14 along with a rearranged column of the same weighting matrix.



**(a)** 1 nearest neighboring weight matrix.

**(b)** A rearranged row or column of a 1-nearest-neighboring weight matrix.
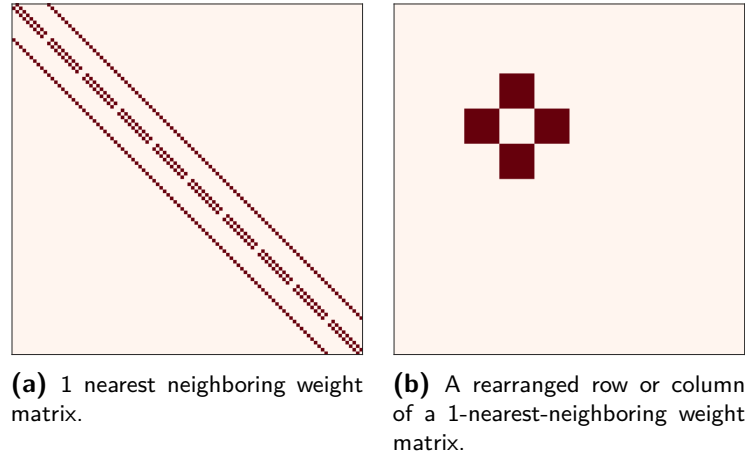
**Figure 2-14:** An example of a 1-nearest-neighboring weight matrix. The figure also shows a rearranged version of one of the rows/columns of the same weight matrix.

Moran's I returns a real value $I$ such that,

$$-1 \leq I \leq 1. \tag{2-38}$$

The values of the $I$ is a measure of the spatial structure or the spatial autocorrelation in the variable, for a given weight matrix. An illustration of what the extreme values of the Moran's I imply is shown in Figure 2-15. In practice, Moran's I can be used to analyze what the spatial autocorrelation is at a certain distance, by using different weight matrices. The weight matrices are picked as $k$ nearest neighboring matrices with $k \in [1, \ldots, k_{max}]$ where

$[1, \ldots, k_{max}]$ relates to the distance range being analyzed. The word 'relates' is used here since $k$ in this context is pixels or observations, but they relate to a physical distance based on the resolution of the image being analyzed.



**(a)** Moran's I of value -1. Perfect spatial dispersion of values.

**(b)** Moran's I of value 0. Values show random structure.

**(c)** Moran's I of value 1. Perfect spatial separation of values.
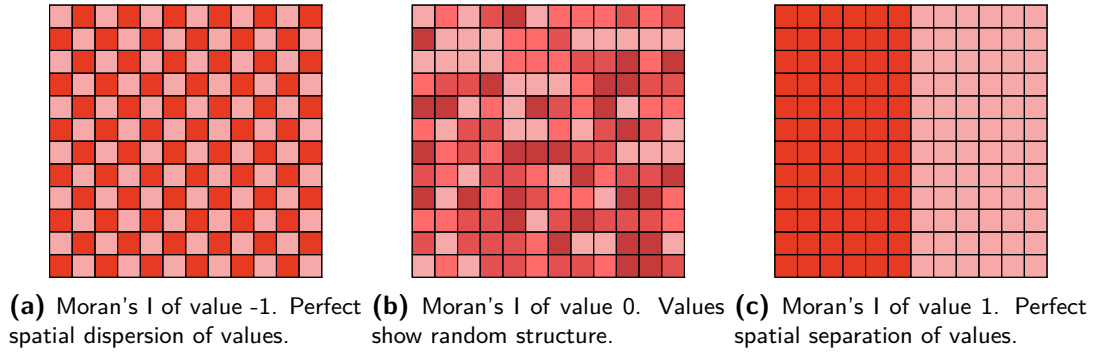
**Figure 2-15:** A visualization of the extreme values of Moran's I. Three examples illustrating different values of Moran's I and how the corresponding spatial structure looks.

## 2-3   Factorization Methods

Matrix factorization or decomposition is the technique of expressing a matrix in parts with structures that are more useful in some way, perhaps more practical for computation or human interpretation [7]. Factorization is commonly used to obtain insights into trends or patterns in data by extracting a lower-dimensional representation [39, 6]. Factorization can also be used as a method for separating noise from signal, hence giving the analyst a clearer view of the data [9].

When working with IMS data, the dimensionality of the datasets make them impractical for humans to interpret directly [5, 6]. Various factorization methods have been applied to IMS data in order to extract useful information, the most common being Principal Component Analysis (PCA) [6, 40, 41, 42, 43, 44], Independent Component Analysis (ICA) [43, 44], Non-Negative Matrix Factorization (NNMF) [43], and Maximum Autocorrelation Factorization (MAF) [45, 11, 46, 47, 48, 49].

In the following chapter an overview of the workings of PCA will be put forth. Expanding on the principals of PCA by incorporating spatial information into the algorithm, will lead us to MAF. Using the dataset $\mathbf{Z}(\mathbf{x}) = \begin{bmatrix} Z_1(\mathbf{x}) & Z_2(\mathbf{x}) & \ldots & Z_p(\mathbf{x}) \end{bmatrix}$, and assuming the data matrix $\mathbf{Z}$ has zero mean, without loss of generality, the data matrix has a covariance matrix, $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$, which is estimated to be

$$\mathbf{\Sigma} = \mathrm{Cov}\left[\mathbf{Z}\right] = \frac{1}{n-1}\mathbf{Z}^T\mathbf{Z}, \qquad (2\text{-}39)$$

where $\mathbf{Z} \in \mathbb{R}^{n \times p}$.

### 2-3-1  Principal Component Analysis

PCA is a factorization method which generates matrix products from a dataset, where these products are such that they maximize the accumulated variance accounted for. These resulting matrices are often referred to as the loading matrix and the score matrix.

We are now looking for a decomposition of $\mathbf{Z}$ that maximizes the accumulated variance accounted for. So, we are seeking a column vector $w_i \in \mathbb{R}^{p \times 1}$ that maximizes

$$\text{Var} \left[ \mathbf{Z} w_i \right] = w_i^T \mathbf{\Sigma} w_i. \tag{2-40}$$

Maximizing the variance without constraints on $w_i$ would result in a non-finite value of $w_i$. So a normalization constraint on $w_i$ must be included. The most common constraint to pick at this point $w_i^T w_i = 1$, that is the sum of squares of $w_i$ must equal 1 [9]. There are other available choices for this constraint, which will be discussed in-depth later.

Using a Lagrange multiplier, $\lambda_i$, the maximization problem can then be formulated as

$$\max w_i w_i^T \mathbf{\Sigma} w_i - \lambda_i (w_i^T w_i - 1). \tag{2-41}$$

Taking the derivative with respect to $w_i$ and finding the equilibrium point yields

$$\mathbf{\Sigma} w_i - \lambda_i w_i = 0, \tag{2-42}$$
$$(\mathbf{\Sigma} - \lambda_i \mathbf{I}_p) w_i = 0, \tag{2-43}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix. This reveals that $\lambda_i$ is an eigenvalue of $\mathbf{\Sigma}$ and $w_i$ is the corresponding right eigenvector.

Looking back at the variance to maximize, applying Equation 2-42 gives

$$\text{Var} \left[ \mathbf{Z} w_i \right] = w_i^T \mathbf{\Sigma} w_i = w_i^T \lambda_i w_i = \lambda_i w_i^T w_i = \lambda_i. \tag{2-44}$$

This means that the largest eigenvalue, $\lambda_i$, and its corresponding eigenvector $w_i$ will account for the maximum variance. This requires the eigenvalues and corresponding eigenvectors to be ordered according to the size of the eigenvalues. Let the eigenvalues be ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \cdots \geq \lambda_p$, their corresponding eigenvectors will share the same subscript [9].
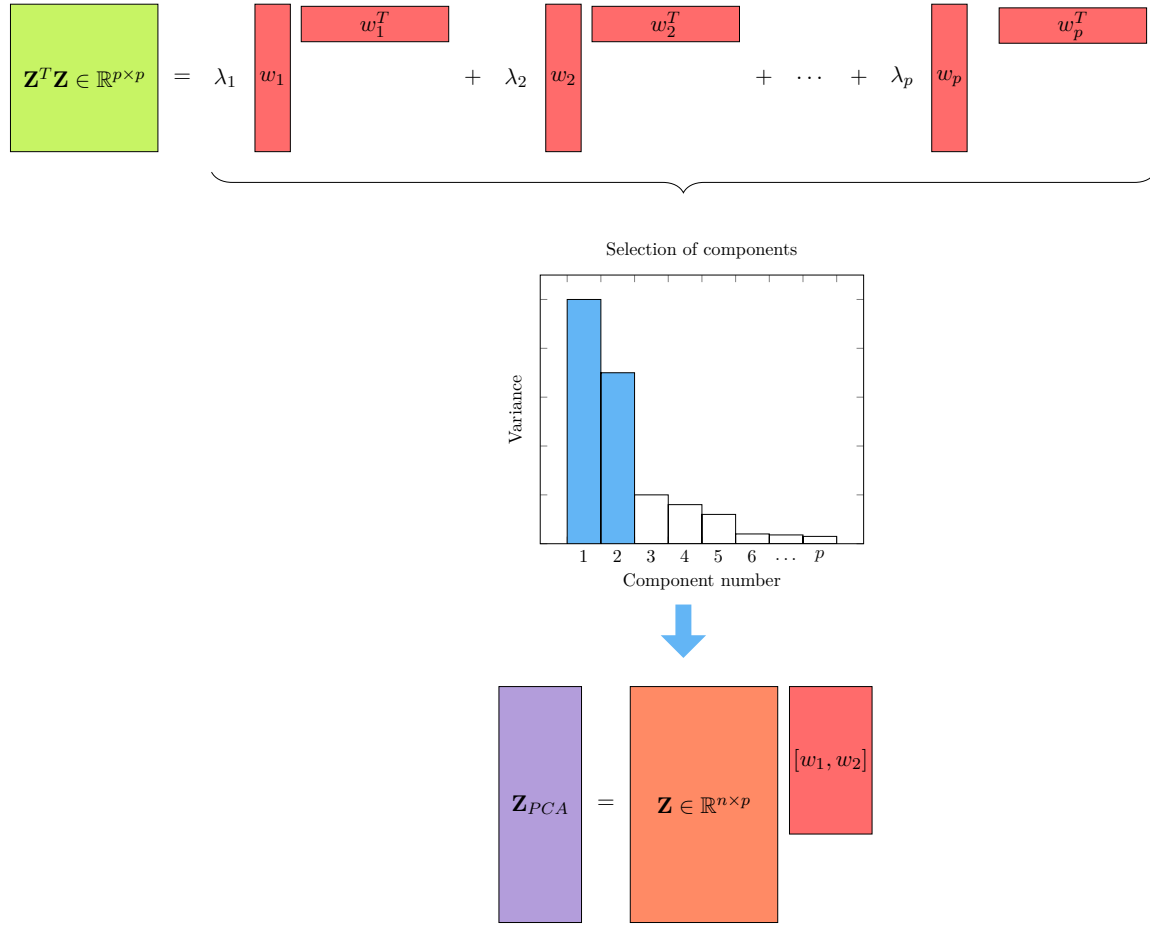
**Figure 2-16:** Dimensionality reduction with PCA. The covariance matrix of the dataset is factorized using eigenvalue decomposition. Looking at the eigenvalues (variance) a few components can be chosen and used to create a lower-dimensional representation of the dataset. Note that in this example $w_i \in \mathbb{R}^{p \times 1}$, $v_i \in \mathbb{R}^{1 \times p}$, and $\lambda_i \in \mathbb{R}$ for all $i \in [1, 2, \ldots, p]$, $\mathbf{Z} \in \mathbb{R}^{n \times p}$, and $\mathbf{Z}_{PCA} \in \mathbb{R}^{n \times 2}$.

The data matrix can now be broken up into factors, using the eigenvectors of the covariance matrix as follows

$$\mathbf{Z}_{PCA} = \mathbf{Z}\left[w_1, w_2, \ldots, w_q\right] = \mathbf{Z}\mathbf{L}_{PCA}, \tag{2-45}$$

where $[w_1, w_2, \ldots, w_q] = \mathbf{L}_{PCA} \in \mathbb{R}^{p \times q}$ with $q \in [1, p]$. It is to be noted that $q \in [1, p]$ can be chosen according to how many components are desired in this lower-dimensional representation. The matrix $\mathbf{Z}_{PCA}$ is called the score matrix and $\mathbf{L}_{PCA}$ is called the loading matrix. The original dataset can be reconstructed with [9]

$$\mathbf{Z} \approx \mathbf{Z}_{PCA}\mathbf{L}_{PCA}^{\dagger}, \tag{2-46}$$

where $\mathbf{L}_{PCA}^{\dagger}$ is the pseudo-inverse in case $\mathbf{L}_{PCA}$ is not a square matrix. This procedure is visualized in Figure 2-17. The data is being transformed onto another coordinate system, where most of the variance is explained by fewer axes than in the original coordinate system.

In Figure 2-17, a two-dimensional dataset is shown being transformed onto another two-dimensional coordinate system, where the new $x'$ axis explains most of the variance of the data.

With the normalization constraint used in this section, $w_i^T w_i = 1$, and if the eigenvalues are all distinct, the matrix of eigenvectors will be orthonormal. That allows the following reconstruction of the dataset.

$$\mathbf{Z} \approx \mathbf{Z}_{PCA}\mathbf{L}_{PCA}^T \qquad (2\text{-}47)$$

Be aware that this only holds given the above conditions.
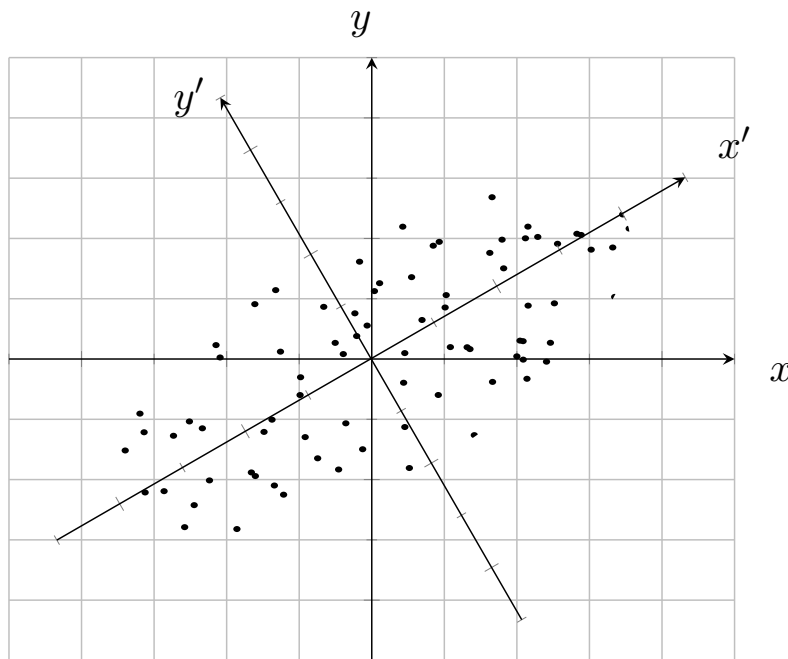


**Figure 2-17:** Two-dimensional illustration of PCA. Each original point is mapped onto a new pair of axes $(x' - y')$ by applying a transformation. These axes are such that they explain the variance of the data as best as possible.

## Principal Component Rotations

It has become common practice to rotate the principal components, most commonly to ease interpretability. The main idea behind principal component rotations is that the component loadings are post-multiplied with a transformation, $A$, arriving at the rotated loadings. Rotations can either be orthogonal or oblique. That is, either the principal components are rotated onto a new pair of axes which are orthogonal or non-orthogonal. A non-exhaustive list of available rotations is given in [50]. Different rotations can be applied to achieve different results, for example maximizing the dispersion of the loadings as is done with a rotation called Varimax.

In his paper, Richman, mentions four disadvantages unrotated principal components exhibit and argues that due to these disadvantages, scientists should consider rotations. These disadvantages all have to do with interpretability of the principal components [50].

The four disadvantages Richman mentions are the following:

1. Domain shape dependence, that is the unrotated components are determined by the geometry of the domain they are described in, posing a limit on the components. In this case domain refers to the coordinate system being used to describe the data. For example, different covariance matrices have been analysed in the same geometrically shaped coordinate system and they returned similar patterns.

2. Subdomain stability, the principal components are not invariant as sections of the domain are examined. The domain being the coordinate system describing the data. So, if a section of the whole domain is analyzed, that might show patterns which were not present if the whole domain was analyzed.

3. Sampling errors, if the populations is sampled in a way that the eigenvalues lie close to each other (this resembles sampling white noise) the principal components may shift order.

4. Richman finally mentions that rotated principal components sometimes simply yield results which are easier to interpret.

There have been discussions on when principal components should be rotated and when they should not. Jolliffe and Richman have for example been replying to each other discussing this in [51] and [52]. A further discussion on this topic can also be found in [9] and [53].

**Loading Normalization**

In order to get finite eigenvectors when maximizing the variance of the transformed data, a normalization constraint must be applied to the eigenvectors. There are a number of normalization constraints available, but three of them are mainly used when carrying out PCA. The constraints all give rise to different traits of the eigenvectors or the principal components.

The constraint that has already been mentioned is

$$w_i^T w_i = 1. \tag{2-48}$$

This constraint rises naturally when deriving the principal components. This normalization will preserve distances from the original data into the components.

A second constraint is

$$w_i^T w_i = \lambda_i, \tag{2-49}$$

where $\lambda_i$ is the eigenvalue corresponding to the eigenvector $w_i$, $\lambda_i$ is also the variance of the $i$-th principal component when $w_i^T w_i = 1$. This constraint enables the eigenvector $w_i$ to be interpreted as the correlation between the $i$-th principal component and the original variables. This only applies if the original variables have been standardized to unit variance.

The third and last constraint examined in this review is

$$w_i^T w_i = \lambda_i^{-1}, \tag{2-50}$$

where $\lambda_i$ is the eigenvalue corresponding to the eigenvector $w_i$. This normalization will give all the principal components the same, unit variance. This means that the principal components will all have equal importance, which can for example be useful for outlier detection [54].

With component rotations in mind, it is worth noting that the choice of normalization will have an impact on the rotation results. As Jolliffe outlines in [54], applying rotations with certain normalization constraints applied will lead to correlation between principal components, loss of orthogonality between components or even both. Mestas-Nuñez *et al.* also give a good overview of how different normalization constraints affect orthogonality properties in rotations [55].

### 2-3-2 Maximum Auto-correlation Factorization

MAF was introduced in 1984 by Paul Switzer and A. A. Green as a method of isolating noise components from signal components in multi-channel remotely sensed data. MAF differs from PCA mainly in that it makes use of global spatial statistics of the observed data as well as being invariant to linear transformations [18].

MAF seeks to find a transformation maximizing the autocorrelation between observations or measurements. The idea is that the signal component of the observation will exhibit high autocorrelation while the noise components will have low autocorrelation, and hence be suppressed [18]. Most factorization methods used with IMS data treat observations as if they were spatially independent, that is they do not take any spatial relationships of the observations into account. However, physical and biological phenomena are generally larger than a single or few pixels in size and should therefore, in theory, exhibit higher spatial autocorrelation than noisy substructures [13, 56].

An appealing trait of MAF is that no prior knowledge about the underlying signals is needed, that is MAF is an unsupervised factorization method. In [57], the authors point out that by assuming the statistical distribution of the noise the resulting factors could be improved, and they carry out experiments with that in mind. The authors also point out that MAF seems to perform poorly if the data has certain local spatial structures, like edges. Another drawback of MAF is, similar to PCA, negative peaks values in the resulting factors, as mentioned in [43].

Recall the following definitions of covariance matrices

$$\mathbf{C}(\mathbf{h}) = \mathrm{Cov}\left[\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x}+\mathbf{h})\right], \tag{2-51}$$

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{Cov}\left[\mathbf{Z}(\mathbf{x}+\mathbf{h}) - \mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x}+\mathbf{h}) - \mathbf{Z}(\mathbf{x})\right]. \tag{2-52}$$

$$\tag{2-53}$$

Recall from Section 2-2-3 that if the dataset is intrinsic stationary, the following applies

$$\gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h}) \tag{2-54}$$

$$\implies \mathbf{C}(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \gamma(\mathbf{h}). \tag{2-55}$$

The covariance between a linear combination of the data matrix and a linear combination of

the shifted data matrix, is

$$\text{Cov}\left[\mathbf{Z}(\mathbf{x})w_i, \mathbf{Z}(\mathbf{x}+\mathbf{h})w_i\right] = w_i^T\mathbf{C}(\mathbf{h})w_i \tag{2-56}$$

$$= w_i^T(\mathbf{C}(\mathbf{0}) - \gamma(\mathbf{h}))w_i \tag{2-57}$$

$$= w_i^T\mathbf{C}(\mathbf{0})w_i - w_i^T\gamma(\mathbf{h})w_i \tag{2-58}$$

$$\implies \frac{w_i^T\mathbf{C}(\mathbf{h})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i} = 1 - \frac{w_i^T\gamma(\mathbf{h})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i}, \tag{2-59}$$

where $w_i \in \mathbb{R}^{p\times 1}$. The term on the right in equation 2-59 is the autocorrelation of the linear combination. Now to maximize the autocorrelation, the following quotient needs to be minimized:

$$\frac{w_i^T\gamma(\mathbf{h})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i}. \tag{2-60}$$

A quotient with the format $\dfrac{x^T\mathbf{A}x}{x^T\mathbf{B}x}$, where $\mathbf{A}$ and $\mathbf{B}$ are real symmetric matrices and $x$ is a real vector, is called a generalized Rayleigh quotient, so we are minimizing a generalized Rayleigh quotient. As shown in the A-2 Section of the Appendix, the minimization of the generalized Rayleigh quotient is equivalent to the generalized eigenvalue problem, which gives

$$\gamma(\mathbf{h})w_i = \lambda_i\mathbf{C}(\mathbf{0})w_i, \tag{2-61}$$

where $w_i$ is the eigenvector corresponding to the eigenvalue $\lambda_i$.

At this point it is worth pointing out that this eigenvalue decomposition can also be written as

$$\gamma(\mathbf{h})\mathbf{C}(\mathbf{0})^{-1}w_i = \lambda_i w_i. \tag{2-62}$$

That is, the eigenvectors, $w_i$, of the matrix $\gamma(\mathbf{h})\mathbf{C}(\mathbf{0})^{-1}$ can be used to transform the original dataset in order to obtain factors which vary minimally across the distance $\mathbf{h}$. Varying minimally across $\mathbf{h}$ is similar to having high spatial autocorrelation. The result can be achieved by first whitening, a process described in detail in the A-3 Section of the Appendix, with $\mathbf{C}(\mathbf{0})$, such that $\mathbf{C}(\mathbf{0}) = \mathbf{I}_p$, and then whitening with respect to the $\gamma(\mathbf{h})$. This is exactly the practical implementation suggested in [18].

Now, an eigenvalue problem has been formulated, which gives a way of transforming the original dataset into factors, but it still needs to be ensured that the first resulting factor will have the highest spatial autocorrelation. Let us start by addressing that problem before showing exactly how the original dataset can be transformed. The Rayleigh quotient can be re-written as

$$\frac{w_i^T\gamma(\mathbf{h})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i} = \frac{w_i^T\lambda_i\mathbf{C}(\mathbf{0})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i} \tag{2-63}$$

$$= \lambda_i\frac{w_i^T\mathbf{C}(\mathbf{0})w_i}{w_i^T\mathbf{C}(\mathbf{0})w_i} \tag{2-64}$$

$$= \lambda_i. \tag{2-65}$$

Replacing the Rayleigh quotient in equation 2-59 with the solution to the generalized eigen-value problem yields

$$\frac{w_i^T \mathbf{C}(\mathbf{h}) w_i}{w_i^T \mathbf{C}(0) w_i} = 1 - \frac{w_i^T \gamma(\mathbf{h}) w_i}{w_i^T \mathbf{C}(0) w_i} = 1 - \lambda_i. \tag{2-66}$$

So $1 - \lambda_i$ is the spatial autocorrelation of the $i$-th MAF factor. This means that to ensure that the first MAF factor has the highest spatial autocorrelation the eigenvalues must be ordered such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$. This way the noisy substructures, which exhibit low spatial autocorrelation, will get demoted to the later factors. The first factors, however, will have high spatial autocorrelation, resembling mostly biological phenomena [18, 45]. This will ensure that when MAF is used for dimensionality reduction, the first few factors can be used as a lower-dimensional representation of the original dataset, while still retaining most of the biological information. In a way, the factors (with low spatial autocorrelation) that are removed will mostly contain noise.

Taking the first $q$ eigenvectors, $w_1, w_2, \ldots, w_q$, corresponding to the mentioned ordering, the original dataset can be transformed as follows

$$\mathbf{Z}_{MAF} = \mathbf{Z} \left[ w_1, w_2, \ldots, w_q \right] = \mathbf{Z} \mathbf{L}_{MAF}, \tag{2-67}$$

where $\mathbf{L}_{MAF} \in \mathbb{R}^{p \times q}$. $\mathbf{Z}_{MAF}$ is a matrix with the MAF scores and $\mathbf{L}_{MAF}$ is a matrix containing the MAF loadings, with $i \in [1, p]$, which means any number of factors can be included in the factorization. The original dataset can then be reconstructed using

$$\mathbf{Z} \approx \mathbf{Z}_{MAF} \mathbf{L}_{MAF}^\dagger, \tag{2-68}$$

where $\mathbf{L}_{MAF}^\dagger$ is the pseudo inverse in case $\mathbf{L}_{MAF}$ is not a square matrix.

### Nested Substructures

Examining the dataset $\mathbf{Z}(\mathbf{x})$ further, assuming each variable has a covariance structure with a known number of substructures. This is rarely the case in practice, but is however done here to analyze and understand the behavior of MAF in each case. In the following chapter the cases of one, two, and three underlying covariance structures will be examined.

**One Nested Substructure**    In the case of a single nested substructure the covariance function of $\mathbf{Z}(\mathbf{x})$ can be written as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}), \tag{2-69}$$

where $c_1(\mathbf{h}) \in \mathbb{R}$ for all $\mathbf{h} \in \mathbb{Z}^2$ is a covariance function of the single substructure and $\mathbf{B}_1 \in \mathbb{R}^{p \times p}$ is a weight matrix for that substructure. The covariance function at lag $\mathbf{h} = (0, 0)$ can be made identity by applying PCA to the dataset. This procedure is called whitening and is further explained in the A-3 Section of the Appendix. This gives uncorrelated factor at lag $\mathbf{h} = (0, 0)$, that is

$$\mathbf{C}(0) = \mathbf{B}_1 = \mathbf{I}_p. \tag{2-70}$$

Now only factors at lag $\mathbf{h} \neq (0,0)$ need to be converted into uncorrelated factors. The variogram at this point is

$$\gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h}) \tag{2-71}$$

$$= (1 - c_1(\mathbf{h}))\mathbf{I}_p. \tag{2-72}$$

Since the variogram is now proportional to the identity matrix, hence diagonal and invariant of $\mathbf{h}$, when MAF is carried that will result in uncorrelated factors at all lag distances $\mathbf{h}$ [58].

**Two Nested Substructures**   If the dataset has two nested substructures, then

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}) + \mathbf{B}_2 c_2(\mathbf{h}), \tag{2-73}$$

where $c_1(\mathbf{h}) \in \mathbb{R}$ and $c_2(\mathbf{h}) \in \mathbb{R}$ for all $\mathbf{h} \in \mathbb{Z}^2$ are covariance functions of the underlying substructures and $\mathbf{B}_1 \in \mathbb{R}^{p \times p}$ and $\mathbf{B}_2 \in \mathbb{R}^{p \times p}$ are weight matrices for those substructures. Again, the dataset can be made to have unit covariance at lag distance $\mathbf{h} = (0,0)$, that gives

$$\mathbf{C}(\mathbf{0}) = \mathbf{B}_1 + \mathbf{B}_2 = \mathbf{I}_p \tag{2-74}$$

$$\implies \mathbf{B}_2 = \mathbf{I}_p - \mathbf{B}_1. \tag{2-75}$$

From this the covariance function can be rewritten as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}) + (\mathbf{I}_p - \mathbf{B}_1) c_2(\mathbf{h}), \tag{2-76}$$

with this in mind the variogram can be constructed as follows

$$\gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h}) \tag{2-77}$$

$$= \mathbf{I}_p - \mathbf{B}_1 c_1(\mathbf{h}) - (\mathbf{I}_p - \mathbf{B}_1) c_2(\mathbf{h}) \tag{2-78}$$

$$= \mathbf{I}_p (1 - c_2(\mathbf{h})) - \mathbf{B}_1 (c_1(\mathbf{h}) - c_2(\mathbf{h})). \tag{2-79}$$

Eigenvalue decomposition of the variogram gives

$$\gamma(\mathbf{h}) = \mathbf{I}_p (1 - c_2(\mathbf{h})) - \mathbf{B}_1 (c_1(\mathbf{h}) - c_2(\mathbf{h})) = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T, \tag{2-80}$$

where $\mathbf{W} \in \mathbb{R}^{p \times p}$ is the eigenvector matrix and $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$ is the eigenvalue matrix. Note that since the variogram matrix is a real symmetric matrix, $\mathbf{W}$ is orthogonal. It is further assumed that $\mathbf{W}$ has been normalized such that $\mathbf{W}^T \mathbf{W} = \mathbf{I}_p$. Using this eigenvector matrix to whiten the variogram results in

$$\gamma_{MAF}(\mathbf{h}) = \mathbf{W}^T \gamma(\mathbf{h}) \mathbf{W} = \mathbf{W}^T \left( \mathbf{I}_p (1 - c_2(\mathbf{h})) - \mathbf{B}_1 (c_1(\mathbf{h}) - c_2(\mathbf{h})) \right) \mathbf{W} \tag{2-81}$$

$$= \mathbf{W}^T \mathbf{I}_p \mathbf{W} (1 - c_2(\mathbf{h})) - \mathbf{W}^T \mathbf{B}_1 \mathbf{W} (c_1(\mathbf{h}) - c_2(\mathbf{h})) \tag{2-82}$$

$$= \mathbf{I}_p (1 - c_2(\mathbf{h})) - \mathbf{W}^T \mathbf{B}_1 \mathbf{W} (c_1(\mathbf{h}) - c_2(\mathbf{h})). \tag{2-83}$$

The matrix product $\mathbf{W}^T \mathbf{B}_1 \mathbf{W}$ is diagonal. Since $\mathbf{I}$ and $\mathbf{B}_1$ commute, that is $\mathbf{I}\mathbf{B}_1 = \mathbf{B}_1\mathbf{I}$, they share the same invariance space and this transformation will therefore diagonalize $\mathbf{B}_1$, according to [59]. This means that the variogram is at this point diagonal, resulting in uncorrelated factors. Examining equation 2-80, it is worth noting that two arbitrary matrices $\mathbf{A}$ and $a\mathbf{I} + b\mathbf{A}$ $(a, b \in \mathbb{R}, b \neq 0)$ do not have the same eigenvalues, but they do have the same

eigenvectors. A proof of this statement can be found in the A-4 Section of the Appendix. This means that $\gamma(\mathbf{h})$ and $\mathbf{B}_1$ will have the same eigenvectors, which also means that the eigenvectors are invariant of the lag distance $\mathbf{h}$. Since the MAF factors are derived using eigenvectors which are, in this case, invariant of the shift parameter, the MAF factors will be uncorrelated at all lag distances $\mathbf{h}$ [60, 58].

In [58], the authors carry out MAF on two nested structures, similar to what is shown here. They conclude that in this context MAF is theoretically a perfect orthogonalization of the nested structures. This statement is proven in [18].

**Three Nested Structures**  If the dataset has three nested substructures, the covariance function can be written as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}) + \mathbf{B}_2 c_2(\mathbf{h}) + \mathbf{B}_3 c_3(\mathbf{h}), \tag{2-84}$$

where $c_1(\mathbf{h}) \in \mathbb{R}$, $c_2(\mathbf{h}) \in \mathbb{R}$, and $c_3(\mathbf{h}) \in \mathbb{R}$ for all $\mathbf{h} \in \mathbb{Z}^2$ are covariance functions of the underlying substructures and $\mathbf{B}_1 \in \mathbb{R}^{p \times p}$, $\mathbf{B}_2 \in \mathbb{R}^{p \times p}$, and $\mathbf{B}_3 \in \mathbb{R}^{p \times p}$ are weight matrices for those substructures. Whitening the covariance matrix at a lag distance $\mathbf{h} = 0$ gives

$$\mathbf{C}(\mathbf{0}) = \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3 = \mathbf{I}_p \tag{2-85}$$

$$\implies \mathbf{B}_3 = \mathbf{I}_p - \mathbf{B}_1 - \mathbf{B}_2, \tag{2-86}$$

yielding the variogram

$$\gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h}) = \mathbf{I}_p - \mathbf{B}_1 c_1(\mathbf{h}) + \mathbf{B}_2 c_2(\mathbf{h}) + (\mathbf{I}_p - \mathbf{B}_1 - \mathbf{B}_2) c_3(\mathbf{h}) \tag{2-87}$$

$$= (1 - c_3(\mathbf{h})) \mathbf{I}_p - \mathbf{B}_1 (c_1(\mathbf{h}) - c_3(\mathbf{h})) + \mathbf{B}_2 (c_2(\mathbf{h}) - c_3(\mathbf{h})). \tag{2-88}$$

Now the variogram can not be diagonalized for $\mathbf{B}_1$ and $\mathbf{B}_2$ simultaneously, unless they commute [59], that is if $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{B}_2 \mathbf{B}_1$. If they indeed do commute, the factorization can carry on and all factors will be uncorrelated for all lag distances $\mathbf{h}$, similarly to the two nested structure case.

If the matrices do not commute, one of the two matrices can be diagonalized while scaling the other. No literature was found on this problem. This same problem arises in the case of more than three nested substructures, except then three or more matrices must commute to produce uncorrelated factors.

A similar effect is seen in the case where the dataset is anisotropic. If the data is anisotropic and two or more nested substructures are present, the resulting factors will only be uncorrelated if certain matrices commute, similar to what has been demonstrated on the cases above. This is described in [58].

This study of nested substructures raises the question of how to choose the shift parameter $\mathbf{h}$, since orthogonal factors cannot be ensured for all distances unless certain criteria are met. When MAF was originally proposed, Switzer and Green suggested a unit horizontal shift and a unit vertical shift [18]. This procedure has been used since and seems to be standard practice in most implementations of MAF, see for example [48] and [46]. The question of how to choose the shift parameter has been raised before, for example in [61].

This analysis above means that in the case where there are only two isotropic covariance structures in the dataset, the choice of the shift parameter does not matter, showing that

uncorrelated factors are ensured at all lag distances. However, the case where there are more than two isotropic covariance structures present needs to be addressed. How the shift parameter, $\mathbf{h}$, affects the resulting MAF factors in this specific case seems to be yet unexplored. Whether the originally proposed diagonal shift is the universally optimal choice, whether this diagonal shift promotes spatial structures which are aligned in the direction of the shift, or whether there might be a different way of choosing the shift parameter remain open questions.

**Loading Signs and Normalization**

Normalization constraints set on the loadings can have a large effect on the properties of the resulting factors, as seen with PCA. When solving the generalized eigenvalue problem, as seen in equation 2-61, the eigenvectors ar constrained such that

$$w_i^T \mathbf{C}(\mathbf{0}) w_i = \mathbf{I}_p. \tag{2-89}$$

After that the eigenvalue problem $\gamma(\mathbf{h}) w_i = \lambda_i w_i$ is solved. This implies that there are no choices to be made regarding loading normalization in MAF. That is, in the MAF algorithm the eigenvectors have already been constrained such that their values will remain finite after minimizing the variogram. If this constraint would not have been applied, the values of the eigenvectors would have been infinite. This is similar to what was discussed with PCA in Equation 2-40.

Another known problem when dealing with loadings is that the scale and sign are arbitrary. Having a look at equation 2-61 for example, it is clear that if the loadings $-w_i$ would be used, they would still satisfy the equation, but the resulting factors would be different. When eigenvalue decomposition is carried out, the signs of the loadings have no meaningful interpretation and can be seen as assigned randomly [62].

In [62], Bro *et al.* suggested a solution by making the loadings have the same sign as the majority of the data-vectors being decomposed. That is, the loadings will point in the same direction as the majority of the data. This will result in a decomposition without side effects in terms of the eigenvector sign, so the same input it will always give the same output.

**Data Transformations**

When carrying out component analysis, it is common to transform the original dataset to correct for sensor gains as seen in for example [9]. This raises the question of how MAF factors are affected by transformations. A transformed dataset is defined as

$$\mathbf{Y} = \mathbf{Z}\mathbf{A}, \tag{2-90}$$

where $\mathbf{A} \in \mathbb{R}^{p \times p}$ is any real, non-singular transformation matrix. Let the dataset $Z(\mathbf{x})$ have the covariance function $\mathbf{C}(\mathbf{h})$ and the variogram $\gamma(\mathbf{h})$. The transformed dataset will then have the following covariance function and variogram

$$\mathrm{Cov}\left[\mathbf{Y}(\mathbf{x}), \mathbf{Y}(\mathbf{x})\right] = \mathrm{Cov}\left[\mathbf{Z}(\mathbf{x})\mathbf{A}, \mathbf{Z}(\mathbf{x})\mathbf{A}\right] = \mathbf{A}^T \mathbf{C}(\mathbf{0})\mathbf{A}, \tag{2-91}$$

$$\mathrm{Cov}\left[\mathbf{Y}(\mathbf{x}+\mathbf{h}) - \mathbf{Y}(\mathbf{x}), \mathbf{Y}(\mathbf{x}+\mathbf{h}) - \mathbf{Y}(\mathbf{x})\right] = \mathrm{Cov}\left[\mathbf{Z}(\mathbf{x}+\mathbf{h})\mathbf{A} - \mathbf{Z}(\mathbf{x})\mathbf{A}, \mathbf{Z}(\mathbf{x}+\mathbf{h})\mathbf{A} - \mathbf{Z}(\mathbf{x})\mathbf{A}\right] \tag{2-92}$$

$$= \mathbf{A}^T \gamma(\mathbf{h})\mathbf{A}. \tag{2-93}$$

Looking at the derivation of the MAF factors, the spatial autocorrelation can now be written as

$$\frac{v_i^T \mathbf{C}(\mathbf{h}) v_i}{v_i^T \mathbf{C}(\mathbf{0}) v_i} = 1 - \frac{v_i^T \mathbf{A}^T \gamma(\mathbf{h}) \mathbf{A} v_i}{v_i^T \mathbf{A}^T \mathbf{C}(\mathbf{0}) \mathbf{A} v_i}, \tag{2-94}$$

now since $\mathbf{A}^T \gamma(\mathbf{h}) \mathbf{A}$ and $\mathbf{A}^T \mathbf{C}(\mathbf{0}) \mathbf{A}$ are both real symmetric matrices, this is called a Generalized Rayleigh quotient. The Generalized Rayleigh quotient can now be minimized as shown in the A-2 Section of the Appendix in order to maximize the spatial autocorrelation. This is similar to was done in Equation 2-61. This gives the solution

$$\mathbf{A}^T \gamma(\mathbf{h}) \mathbf{A} v_i = \lambda_i \mathbf{A}^T \mathbf{C}(\mathbf{0}) \mathbf{A} v_i \tag{2-95}$$

$$\gamma(\mathbf{h})(\mathbf{A} v_i) = \lambda_i \mathbf{C}(\mathbf{0})(\mathbf{A} v_i). \tag{2-96}$$

The transformation does not change the resulting loadings, that is $\mathbf{A} v_i = w_i$. This results in the following MAF factors

$$\mathbf{Y}[v_1, v_2, \ldots, v_q] = \mathbf{Y}\mathbf{V} = \mathbf{Z}\mathbf{A}\mathbf{V} = \mathbf{Z}\mathbf{L}_{MAF}, \tag{2-97}$$

where $\mathbf{L}_{MAF} \in \mathbb{R}^{p \times q}$ is the matrix of the $q$ first eigenvectors $[w_1, w_2, \ldots, w_q]$. This means that the Maximum Autocorrelation Factorization is invariant of linear transformations of the original dataset. This feature of MAF makes data scaling or instrumentation gain irrelevant [18, 63].

# Chapter 3

# Methods

It has been demonstrated that utilizing the spatial information available in data can reveal patterns that were otherwise hidden [14, 15, 16, 17]. Maximum Autocorrelation Factorization (MAF) [18] is a factorization method that takes both spectral and spatial structure of data into account by constructing factors based upon the spatial covariance function. As well as utilizing spatial information in the data, MAF is an unsupervised factorization method, that is no prior information about the data is provided by the user. Therefore, MAF is considered a strong candidate for factorization and unsupervised exploratory analysis of Imaging Mass Spectrometry (IMS) and other spatially structured, multivariate data.

However, as demonstrated in Chapter 4, MAF is dependent on a spatial shift parameter, which needs to be tuned and set by the user before analyzing the data. This shift parameter has typically been set to a default value since MAF was introduced, namely a diagonal shift (to capture variation along the two axes of a two-dimensional image). In this work, the goal is to extend MAF to become independent of user-input, optimizing the shift parameter without human intervention.

Similarly to Chapter 2-2, we have a dataset

$$\mathbf{Z} = \begin{bmatrix} Z_1 & Z_2 & \ldots & Z_p \end{bmatrix}, \tag{3-1}$$

with $\mathbf{Z} \in \mathbb{R}^{n \times p}$ and $Z_i \in \mathbb{R}^{n \times 1}$ for $i \in [1, 2, \ldots, p]$. That is, the dataset contains $n$ observations (usually pixels) of $p$ variables (features measured per pixel). It is assumed, without loss of generality, that all variables are zero mean, that is

$$\mathrm{E}\left[Z_i\right] = 0 \tag{3-2}$$

for all $i \in [1, 2, \ldots, p]$. It is further assumed that this data is intrinsic stationary, that is the mean does not change if the data is shifted spatially and the variogram is only a function of the shift parameter. This assumption is discussed in detail in Chapter 2-2-4.

We propose an altered version of the original MAF algorithm, will be used for the Extended Maximum Autocorrelation Factorization (EMAF) algorithm, which is a single-step approach

to achieving the eigenvalues and eigenvectors needed to transform the dataset. This single-step approach, unlike the originally proposed algorithm, does not involve sphering the data. The reason why sphering is avoided is discussed in detail in Chapter 3-5. A description of sphering can be found in Chapter A-3. Note that this alternative approach is the same approach as introduced in the Chapter 2-3-2. The same approached is used in for example [61, 64].

So, for comparison, the originally proposed algorithm is:

1. *Sphering:* Transform the data, $\mathbf{Z}$, using a transformation matrix $\mathbf{S} \in \mathbb{R}^{p \times p}$, such that $\text{Cov}\,[\mathbf{ZS}] = \mathbf{I}_{p \times p}$,

2. *The eigenvalue problem:* solve $\mathbf{C}(\mathbf{h})w_i = \lambda_i w_i$ to get eigenvalues $\lambda_i$, and eigenvectors $w_i$, where $\mathbf{C}(\mathbf{h}) \in \mathbb{R}^{p \times p}$ is the spatial covariance function of the dataset for a specific shift vector $\mathbf{h}$,

where $\mathbf{I}_{p \times p} \in \mathbb{R}^{p \times p}$, $w_i \in \mathbb{R}^{p \times 1}$, and $\lambda_i \in \mathbb{R}$. A more detailed discussion about the original MAF algorithm can be found in [18]. The alternative to this two-step algorithm is referred to as the single-step method, which is solving the eigenvalue problem

$$\mathbf{C}(\mathbf{h})w_i = \lambda_i \mathbf{C}(\mathbf{0})w_i, \tag{3-3}$$

directly.

The shift parameter will be described in Cartesian coordinates, that is $\mathbf{h} = (x, y)$. This choice is made for practical reasons. When implementing the algorithm the shifts will end up being described as $x$ pixels horizontally and $y$ pixels vertically. It is to be noted that the images can only be shifted an integer number of pixels, that is $x, y \in \mathbb{Z}$.

In this chapter the focus will be on altering the eigenvalue problem shown in Equation 3-3. To start off, an intuitive understanding of how this eigenvalue problem works will be established and then the eigenvalue problem will be altered. The hope is to alter the eigenvalue problem in such a way that it becomes independent of a shift parameter input from the user. Later in the chapter, the sphering step in the original MAF algorithm will be discussed and what its effects on the resulting factors are. Then, a robust version of the EMAF algorithm will be discussed briefly. The last section in this chapter will summarise the traits of the newly developed EMAF algorithm.

## 3-1   The MAF eigenvalue problem

The MAF eigenvalue problem is generally formulated as

$$\mathbf{C}(\mathbf{h})w_i = \lambda_i \mathbf{C}(\mathbf{0})w_i, \tag{3-4}$$

where $\mathbf{C}(\mathbf{h}) \in \mathbb{R}^{p \times p}$ is the spatial covariance function of the dataset, $\mathbf{C}(\mathbf{0}) \in \mathbb{R}^{p \times p}$ is the spatial covariance function of the dataset evaluated at location $\mathbf{h} = \mathbf{0}$ (so without spatial shift), $w_i \in \mathbb{R}^{p \times 1}$ are the eigenvectors and $\lambda_i \in \mathbb{R}$ are the eigenvalues.

Generally speaking, when solving an eigenvalue problem we are looking for vectors that are only scaled by the transformation in question, the spatial correlation function in this case.

This change in scale is quantified by the eigenvalues and the vectors being scaled are the eigenvectors. The eigenvectors can also be though of as the axes of the transformation, for example the axes which this linear transformation is rotating, flipping, skewing, etc. around.

In the case where the transformation matrix has a physical meaning tied to it, the eigenvalues can have a physical meaning, as well. In the case of MAF, the eigenvalues are related to the spatial autocorrelation of the resulting MAF factors [61]. The goal here is to find a linear transformation of the dataset which will maximize this spatial autocorrelation of the resulting factors. How such a linear transformation can be found is shown later in this thesis. To get factors with maximal spatial autocorrelation, the factors can simply be ordered according to their eigenvalues. One problem with MAF is, looking at Equation 3-4, that the eigenvalues are dependent on the user input $\mathbf{h}$. How should users choose this parameter?

The user's goal is to extract information from the dataset and hopefully be able to separate out noise. In the case of IMS data, biological patterns are generally more likely to exhibit high spatial autocorrelation compared to noise. Single pixel artifacts for example are most likely not biological information, with single pixel artifacts having close to zero spatial autocorrelation. Before being able to apply MAF the user has to know in which direction and distance the autocorrelation is greatest, this direction and distance will assemble the shift parameter. The idea behind EMAF is to relief the user from having to analyze the data and pick this parameter. This can be achieved by taking more than one shift into account, making the algorithm invariant of the shift parameter, in terms of both its length and direction.

In order to take multiple shifts into account, the lag space is segmented and the values of the spatial correlation function are averaged across each segment. A similar idea can be used to calculate MAF factors. This way the MAF algorithm will not promote spatial artifacts oriented in one direction over any other direction, it will rather seek to maximize the average spatial autocorrelation across all the shift parameters taken into account.

One of the goals here is to alter the MAF algorithm such that it will not promote spatial distributions oriented in one direction over any other. This is achieved by calculating an average of all shift parameter angles in a certain Region of Interest (ROI) of the lag space. There will be different angles available depending on the size of this ROI. If, for example, the ROI is a circle of radius $r = 1$, there will be four angles, $\theta$, available, that is $\theta \in [0, \frac{\pi}{2}, \pi, \frac{3\pi}{4}]$. The number of available angles increases as $r$ increases. But the idea is that all angles are utilized, whatever the ROI radius is.

Formulating this averaging idea mathematically, the spatial correlation function is first assumed to be a continuous function; the results will then later be discretized for practical applications.

First the ROI is formulized, that is the part of the lag space that is being averaged across. Since all angles should be treated with equal weight, the region should to be circular. So the ROI is a circle in the lag space, centered at $(0, 0)$ with radius $r$, described by the equation $x^2 + y^2 = r^2$, where the shift parameter is $\mathbf{h} = (x, y)$. There are two two different options that arise. One option is to take all shift parameters that fall within this ROI into account. This would result in a relatively accurately approximated spatial correlation over the ROI, albeit slowing down calculation times, since more shifts are taken into account. This method of taking the average across the whole ROI will be refered to as the *disc implementation*. An alternative to this, which might prove more practical due to lower calculation time, is to only

look at the rim of the ROI. This method of taking only shift parameters which land on the rim of the ROI into account will be referred to as the *rim implementation*.



**(a)** The ROI for the *disc implementations*.        **(b)** The ROI for the *rim implementations*.
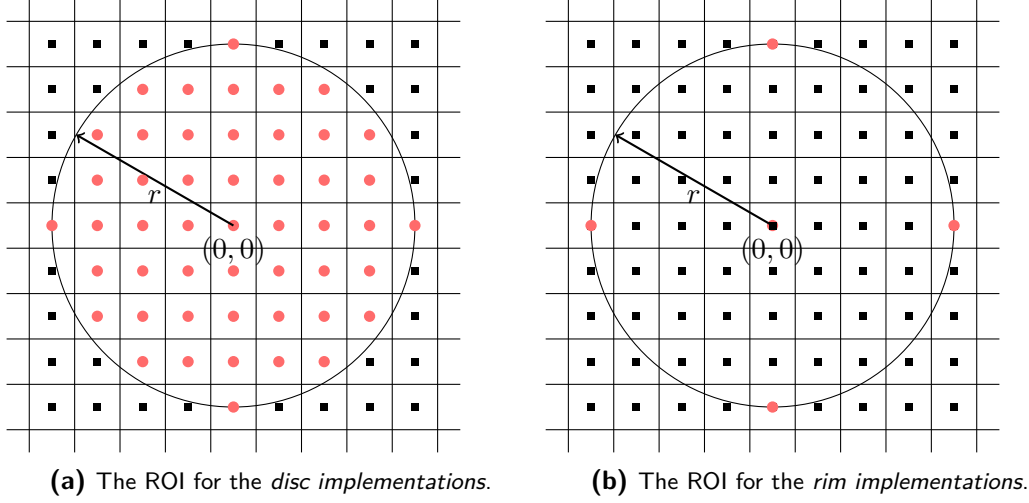
**Figure 3-1:** The ROI for both *disc-* and *rim implementations*. The circular region with radius $r$ that determines what values in the lag space will be included in the average correlation calculations. Shifts not included in the average correlation are indicated with black squares and the included shifts are indicated with red circles.

## 3-2   MAF disc extension

This section focuses on extending the MAF algorithm with the disc implementation, taking all shifts that fall within the ROI into account. These shifts can be described as all shifts $\mathbf{h} = (x, y)$ where $-r \leq x \leq r$ and $-\sqrt{r^2 - x^2} \leq y \leq \sqrt{r^2 - x^2}$. Figure 3-1a shows an example of what shifts or lags fall within this region. The region has an area of $A = r^2\pi$, the calculations for this surface area are shown in the A-1-1 Section of the Appendix. This is also done to show that this parameterization of the region is valid.

First the average spatial correlation will be calculated, which will then be used to formulate an optimization problem. The average spatial covariance can be found by integrating the spatial correlation function across some ROI and dividing the integral by the area of this ROI. For now, all shifts are assumed to be continuous, the results will be discretized later.

A general formula for the average spatial covariance across the disc is, $\overline{\mathrm{cov}}_{disc}$ is

$$\overline{\mathrm{cov}}_{disc} = \frac{1}{A} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \mathbf{C}(x,y) dy dx, \tag{3-5}$$

where the area of the ROI is $A$. Since the ROI is a circle, we have $A = r^2\pi$, so

$$\overline{\mathrm{cov}}_{disc} = \frac{1}{A} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \mathbf{C}(x,y) dy dx \tag{3-6}$$

$$= \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \mathbf{C}(x,y) dy dx. \tag{3-7}$$

The data is assumed to be intrinsically stationary, so

$$\gamma(x,y) = \mathbf{C}(0,0) - \mathbf{C}(x,y) \tag{3-8}$$

$$\implies \mathbf{C}(x,y) = \mathbf{C}(0,0) - \gamma(x,y). \tag{3-9}$$

Equation 3-7 can now be written with the variogram instead of the spatial correlation function, getting

$$\overline{\text{cov}}_{disc} = \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \mathbf{C}(x,y)dydx \tag{3-10}$$

$$= \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \left(\mathbf{C}(0,0) - \gamma(x,y)\right) dydx \tag{3-11}$$

$$= \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \mathbf{C}(0,0)dydx - \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx \tag{3-12}$$

$$= \mathbf{C}(0,0)\frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dydx - \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx \tag{3-13}$$

$$= \mathbf{C}(0,0)\frac{1}{r^2\pi}r^2\pi - \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx \tag{3-14}$$

$$= \mathbf{C}(0,0) - \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx, \tag{3-15}$$

by using the results shown in the A-1-1 Section of the Appendix, to calculate

$$\int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dydx. \tag{3-16}$$

This average spatial correlation can now be maximized using a linear transformation, using a similar idea as for Principal Component Analysis (PCA). First, the average spatial covariance $\overline{\text{cov}}_{disc}$ is multiplied by a weight vector, $w_i \in \mathbb{R}^{p \times 1}$, the the average spatial correlation can be formulated and maximized. That is

$$w_i^T \overline{\text{cov}}_{disc} w_i = w_i^T \left( \mathbf{C}(0,0) - \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx \right) w_i \tag{3-17}$$

$$= w_i^T \mathbf{C}(0,0)w_i - w_i^T \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx w_i \tag{3-18}$$

$$\implies \frac{w_i^T \overline{\text{cov}}_{disc} w_i}{w_i^T \mathbf{C}(0,0)w_i} = \mathbf{I}_{p \times p} - \frac{w_i^T \frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx w_i}{w_i^T \mathbf{C}(0,0)w_i} \tag{3-19}$$

It has until now been assumed the the lag space is continuous, but it is discrete. The variogram for example is not a continuous function, but a discrete function. Because of this, integrating the function is not an option in practice. The integral can be estimated by using a Riemann sum. This approximation gives

$$\frac{1}{r^2\pi} \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)dydx \approx \frac{1}{r^2\pi} \sum_{x=-r}^{r} \sum_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y)\Delta y \Delta x, \tag{3-20}$$

where $x, y \in \mathbb{Z}$, $\Delta x$ and $\Delta y$ are the spatial intervals between the points where the variogram function is evaluated. The variogram will be evaluated at every pixel in the ROI. Since there is now a single pixel shift between every evaluation of the variogram meaning that

$$\Delta x = 1, \tag{3-21}$$

$$\Delta y = 1. \tag{3-22}$$

This approximation is visualized in Figure 3-1, all the available values of the variogram which fall within the circular ROI are being summed up. This results in a weighted average correlation of

$$\frac{w_i^T \overline{\text{cov}}_{disc} w_i}{w_i^T \mathbf{C}(0,0) w_i} = \mathbf{I}_{p \times p} - \frac{w_i^T \frac{1}{r^2 \pi} \sum_{x=-r}^{r} \sum_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y) w_i}{w_i^T \mathbf{C}(0,0) w_i}. \tag{3-23}$$

This weighted average correlation can now be maximized, yielding

$$\max_{w_i} \frac{w_i^T \overline{\text{cov}}_{disc} w_i}{w_i^T \mathbf{C}(0,0) w_i} \approx \min_{w_i} \frac{w_i^T \frac{1}{r^2 \pi} \sum_{x=-r}^{r} \sum_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y) w_i}{w_i^T \mathbf{C}(0,0) w_i}. \tag{3-24}$$

This minimization problem is a Generalized Rayleigh Quotient. The minimization of a Generalized Rayleigh Quotient is shown in Section A-2 of the Appendix. Applying the results from the Appendix yields the following eigenvalues and eigenvectors

$$\left( \frac{1}{r^2 \pi} \sum_{x=-r}^{r} \sum_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \gamma(x,y) \right) w_i = \lambda_i \mathbf{C}(0,0) w_i, \tag{3-25}$$

where $\lambda_i \in \mathbb{R}$ is the $i$-th eigenvalue, $w_i \in \mathbb{R}^{p \times 1}$ is the corresponding, $i$-th eigenvector for all $i \in [1, 2, \ldots, p]$. Since the sum of variogram values over the ROI is being minimized, the eigenvalues are ordered such that $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$. This is calculated using the results shown in the A-2 Section of the Appendix.

Unlike when solving a standard eigenvalue problem, in this case there is no need to pick a constraint on the eigenvectors. When solving a standard eigenvalue problem the eigenvectors have to be contrained in order to get finite valued eigenvectors. The problem being that there are a few constraints to choose from, each with different traits. This is discussed for example in [9, 54]. In this case, the eigenvectors constraint

$$w_i^T \mathbf{C}(\mathbf{0}) w_i = \mathbf{I}_p, \tag{3-26}$$

rises naturally, as discussed in Chapter 2-3-2.

In this case the eigenvalues, $\lambda_i$ will represent the average variogram divided by the covariance, $\mathbf{C}(\mathbf{0})$, across the ROI. Looking at Equation 3-23, this eigenvalue can be related to the average spatial autocorrelation by the equation $1 - \lambda_i$. This means that the produced factors should have a high average spatial autocorrelation, when shifted within the ROI. That way promoting factors with larger spatial artifacts of uniform values.

Taking the first $q$ eigenvectors, $w_1, w_2, \ldots, w_q$, corresponding to the mentioned ordering, the dataset $\mathbf{Z} \in \mathbb{R}^{n \times p}$ can now be transformed using the EMAF eigenvectors to obtain the EMAF factors as

$$\mathbf{Z}_{EMAF} = \mathbf{Z}\left[w_1, w_2, \ldots, w_q\right] = \mathbf{Z}\mathbf{L}_{EMAF}, \tag{3-27}$$

where $\mathbf{L}_{EMAF} \in \mathbb{R}^{p \times q}$. $\mathbf{Z}_{EMAF}$ is a matrix with the EMAF scores and $\mathbf{L}_{EMAF}$ is a matrix containing the EMAF loadings. With $i \in [1, p]$, which means any number of factors can be included in the factorization. The original dataset can then be reconstructed using

$$\mathbf{Z} \approx \mathbf{Z}_{EMAF}\mathbf{L}_{EMAF}^{\dagger}, \tag{3-28}$$

where $\mathbf{L}_{EMAF}^{\dagger}$ is the pseudo inverse in case $\mathbf{L}_{EMAF}$ is not a square matrix.

This is now a directionally independent version of MAF which results in factors with maximal average spati This method is described in Algorithm 1.

---

**Algorithm 1** Disc version of EMAF.

---

**Require:** $\mathbf{Z} \in \mathbb{R}^{n \times p}$, $r \in \mathbf{R}_{>0}$
  **procedure** $EMAF_{disc}(\mathbf{Z}, r)$
    $\mathbf{\Sigma} = \operatorname{cov}\left[\mathbf{Z}(0,0)\right]$
    $(x, y) = \{x, y \in \mathbb{Z} \mid x^2 + y^2 \leq r^2\}$
    **variogram_sum** $= \mathbf{0}_{p \times p}$
    **for all** $(x, y)$ **do**
      $\gamma = \frac{1}{2}\operatorname{cov}\left[\mathbf{Z(x, y)} - \mathbf{Z(0, 0)}\right]$
      **variogram_sum** = **variogram_sum** $+ \gamma$
    **end for**
    **variogram_average** $= \frac{1}{r^2\pi}$**variogram_sum**
    $\mathbf{W}, \mathbf{\Lambda} = \operatorname{eig}(\textbf{variogram\_average}, \mathbf{\Sigma})$
  **return** $\mathbf{ZW} \in \mathbb{R}^{n \times p}$, $\mathbf{W} \in \mathbb{R}^{p \times p}$, $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$
  **end procedure**
**Require:** $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{B} \in \mathbb{R}^{p \times p}$
  **procedure** EIG$(\mathbf{A}, \mathbf{B})$
    $\mathbf{AW} = \mathbf{BW\Lambda}$
  **return** $\mathbf{W} \in \mathbb{R}^{p \times p}$, $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$
  **end procedure**

---

## 3-3   Rim implementation

In this section, the MAF algorithm will be extended using the *rim implementation*. Only shifts that fall exactly on the rim or boundary of the ROI are taken into account in this case. Recall that the ROI is a circular region with radius $r$, centered at $(0, 0)$ in the lag space. These shifts can be described as all shifts $\mathbf{h} = (r_{shift}, \theta)$ where $0 \leq \theta \leq 2\pi$ and $r_{shift} = r$, that is the shifts are noted in polar coordinates. Figure 3-1b shows an example of what shifts or lags fall within this region. This average can be carried out as a single dimension integral in polar coordinates, that is the radial dimension $\theta$. The length of the rim is $L = 2r\pi$, the calculations for this are shown in A-1-2 Section of the Appendix.

First the average spatial correlation is calculated, which will then later be used to form an optimization problem. The average spatial correlation is found by integrating the spatial correlation function across the boundary of the ROI and dividing the integral by the circumference. It is assumed for now that the shifts are continuous and the results will be discretized later.

A general formula for the rim average spatial correlation, $\overline{\text{cov}}_{rim}$ is

$$\overline{\text{cov}}_{rim} = \frac{1}{L} \int_{\theta=0}^{2\pi} \mathbf{C}(r,\theta) r d\theta, \tag{3-29}$$

where $L$ is the circumference of the ROI. As stated previously $L = 2\pi r$, and using the relation between the variogram and the spatial covariance function derived in the previous section, Equation 3-9, we get

$$\overline{\text{cov}}_{rim} = \frac{1}{L} \int_{\theta=0}^{2\pi} \mathbf{C}(r,\theta) r d\theta \tag{3-30}$$

$$= \frac{1}{2\pi r} \int_{\theta=0}^{2\pi} \mathbf{C}(r,\theta) r d\theta \tag{3-31}$$

$$= \frac{1}{2\pi r} \int_{\theta=0}^{2\pi} \left( \mathbf{C}(0,0) - \gamma(r,\theta) \right) r d\theta \tag{3-32}$$

$$= \frac{1}{2\pi r} \int_{\theta=0}^{2\pi} \left( \mathbf{C}(0,0) - \gamma(r,\theta) \right) r d\theta \tag{3-33}$$

$$= \frac{1}{2\pi r} \left( \int_{\theta=0}^{2\pi} \mathbf{C}(0,0) r d\theta - \int_{\theta=0}^{2\pi} \gamma(r,\theta) r d\theta \right) \tag{3-34}$$

$$= \frac{1}{2\pi r} \left( \mathbf{C}(0,0) \int_{\theta=0}^{2\pi} r d\theta - \int_{\theta=0}^{2\pi} \gamma(r,\theta) r d\theta \right) \tag{3-35}$$

$$= \frac{1}{2\pi r} \left( \mathbf{C}(0,0) 2\pi r - \int_{\theta=0}^{2\pi} \gamma(r,\theta) r d\theta \right) \tag{3-36}$$

$$= \mathbf{C}(0,0) - r \frac{1}{2\pi r} \int_{\theta=0}^{2\pi} \gamma(r,\theta) d\theta \tag{3-37}$$

$$= \mathbf{C}(0,0) - \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \gamma(r,\theta) d\theta, \tag{3-38}$$

using the results shown in the A-1-2 Section of the Appendix and the fact that $r$ is constant.

Similarly to the disc case, it was assumed that the shift parameter was continuous, but it needs to be discretized. Although in this case it was further assumed that the shift parameter was given in polar coordinates so the shift parameter needs to be transformed into Cartesian coordinates as well. The shift parameter can be cast into Cartesian coordinates, such that $\mathbf{h} = (x,y)$, using

$$\theta = \arctan\left(\frac{y}{x}\right). \tag{3-39}$$

Then, discretizing the average spatial correlation gives

$$\frac{1}{2\pi} \int_{\theta=0}^{2\pi} \gamma(r,\theta) d\theta \approx \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \gamma(r,\theta) \Delta\theta \tag{3-40}$$

In this case the angle does not change constantly, so the evaluations of the variograms get weighted depending on the length of the rim they cover. This length can be calculated using

$$\Delta\theta = -\frac{y}{x^2 + y^2}\Delta x + \frac{x}{x^2 + y^2}\Delta y \tag{3-41}$$

where $\Delta x$ and $\Delta y$ are the changes in the $x$ and $y$ coordinates between two shift parameter values, respectively.

The final goal is to find a linear transformation that will maximize the average spatial correlation of the resulting EMAF factors. First, this linear transformation is formulated, then the maximization is carried out later. Using vectors $w_i \in \mathbb{R}^{p \times 1}$ to transform the average spatial covariance gives

$$w_i^T \overline{\text{cov}}_{rim} w_i = w_i^T \left( \mathbf{C}(0,0) - \frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(r,\theta)\Delta\theta \right) w_i \tag{3-42}$$

$$= w_i^T \mathbf{C}(0,0)w_i - w_i^T \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\gamma(r,\theta)d\theta w_i \tag{3-43}$$

$$\implies \frac{w_i^T \overline{\text{cov}}_{rim} w_i}{w_i^T \mathbf{C}(0,0)w_i} = \mathbf{I}_{p \times p} - \frac{w_i^T \frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(r,\theta)\Delta\theta w_i}{w_i^T \mathbf{C}(0,0)w_i} \tag{3-44}$$

Maximizing this linear transformations gives

$$\max_{w_i} \frac{w_i^T \overline{\text{cov}}_{rim} w_i}{w_i^T \mathbf{C}(0,0)w_i} \approx \min_{w_i} \frac{w_i^T \frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(r,\theta)\Delta\theta w_i}{w_i^T \mathbf{C}(0,0)w_i} \tag{3-45}$$

This minimization problem is a Generalized Rayleigh Quotient. The minimization of Generalized Rayleigh Quotients is covered in Section A-2 of the Appendix. Similarly to the disc implementation, the vectors $w_i$ need to be constrained. Again the constraint $w_i^T \mathbf{C}(\mathbf{0})w_i = 1$ rises naturally, as discussed in Section 2-3-2. This leaves the eigenvalue problem

$$\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(r,\theta)\Delta\theta w_i = \lambda_i \mathbf{C}(0,0)w_i, \tag{3-46}$$

where $\lambda_i \in \mathbb{R}$ is the $i$-th eigenvalue, $w_i \in \mathbb{R}^{p \times 1}$ is the corresponding, $i$-th eigenvector for all $i \in [1, 2, \ldots, p]$. Since the sum of variogram values over the boundary of the ROI is being minimizing, the eigenvalues are ordered such that $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$.

In this case, the eigenvalues will represent the average variogram across the ROI, and so $1 - \lambda_i$ will be the average spatial autocorrelation of the $i$-th factor if it is shifted to the rim of the ROI. This relation is derived from Equation 3-40. This means that the produced factors should have a high average spatial autocorrelation when shifted to the rim of the ROI, that way promoting factors with larger spatial artifacts of uniform values.

Taking the first $q$ eigenvectors, $w_1, w_2, \ldots, w_q$, corresponding to the mentioned ordering, the dataset $\mathbf{Z} \in \mathbb{R}^{n \times p}$ can now be transformed using the EMAF eigenvectors to obtain the EMAF factors as

$$\mathbf{Z}_{EMAF} = \mathbf{Z}[w_1, w_2, \ldots, w_q] = \mathbf{Z}\mathbf{L}_{EMAF}, \tag{3-47}$$

where $\mathbf{L}_{EMAF} \in \mathbb{R}^{p \times q}$. $\mathbf{Z}_{EMAF}$ is a matrix with the EMAF scores and $\mathbf{L}_{EMAF}$ is a matrix containing the EMAF loadings. With $i \in [1, p]$, which means any number of factors can be included in the factorization. The original dataset can then be reconstructed using

$$\mathbf{Z} \approx \mathbf{Z}_{EMAF} \mathbf{L}_{EMAF}^{\dagger}, \tag{3-48}$$

where $\mathbf{L}_{EMAF}^{\dagger}$ is the pseudo inverse in case $\mathbf{L}_{EMAF}$ is not a square matrix.

This results in a version of MAF which produces factors with maximal average spatial autocorrelation, calculated from the boundary of a set ROI with radius $r$. This method is described in Algorithm 2.

---

**Algorithm 2** Rim version of EMAF

---

**Require:** $\mathbf{Z} \in \mathbb{R}^{n \times p}$, $r \in \mathbf{R}_{>0}$
  **procedure** $EMAF_{disc}(\mathbf{Z}, r)$
    $\mathbf{\Sigma} = \mathrm{cov}\,[\mathbf{Z}(0, 0)]$
    $(x, y) = \{x, y \in \mathbb{Z} \,|\, x^2 + y^2 = r^2\}$
    **variogram_sum** $= \mathbf{0}_{p \times p}$
    **for all** $(x, y)$ **do**
      $\gamma = \frac{1}{2}\mathrm{cov}\,[\mathbf{Z}(\mathbf{x}, \mathbf{y}) - \mathbf{Z}(\mathbf{0}, \mathbf{0})]$
      $\Delta\theta = -\dfrac{y}{x^2 + y^2}\Delta x + \dfrac{x}{x^2 + y^2}\Delta y$
      **variogram_sum** = **variogram_sum** $+ \gamma\Delta\theta$
    **end for**
    **variogram_average** $= \frac{1}{2\pi}$**variogram_sum**
    $\mathbf{W}, \mathbf{\Lambda} = \mathrm{eig}(\textbf{variogram\_average}, \mathbf{\Sigma})$
  **return** $\mathbf{ZW} \in \mathbb{R}^{n \times p}$, $\mathbf{W} \in \mathbb{R}^{p \times p}$, $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$
  **end procedure**
**Require:** $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{B} \in \mathbb{R}^{p \times p}$
  **procedure** EIG$(\mathbf{A}, \mathbf{B})$
    $\mathbf{AW} = \mathbf{BW\Lambda}$
  **return** $\mathbf{W} \in \mathbb{R}^{p \times p}$, $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$
  **end procedure**

---

## 3-4   Region of interest radius

The question of how to choose the radius of this ROI remains. How to choose a radius for the ROI is not as straightforward as dealing with the angle of the shift parameter. The radius of the ROI is bounded by the size of the image being analyzed. Two constraints to keep in mind is that calculating a variogram value can be time consuming and spatial covariance generally decreases with increasing shift distance (in real-world data that is). For these reasons, the radius of the ROI should be kept relatively small. However, the accuracy of the approximation of the average spatial correlation increases with increasing radius, as more data points are used to calculate the variogram with a larger radius.

A good balance between a large and a small radius of the ROI should be found. We need to find a radius such that it gives a good balance between accuracy and computational time,

**Table 3-1:** The radius estimation methods as well as control methods applied. Note that $r_{roi}^2 = x^2 + y^2$.

| Method name | Abbreviation | Mathematical description |
|---|---|---|
| Maximum spatial autocorrelation | Max AC | $r = \underset{r_{roi}}{\arg\max}\, I(Z(x,y))$ |
| Minimal spatial autocorrelation | Min AC | $r = \underset{r_{roi}}{\arg\min}\, I(Z(x,y))$ |
| Unit circle | UC | $r = 1$ |

while still maintaining a high spatial autocorrelation of the resulting factors. Since there is no obvious choice of radius at this point, a few different methods to select a radius of the ROI will be proposed. Some of these methods rely on having an estimate of spatial correlation at certain distances. To keep calculation times low, the spatial correlation should be estimated using other methods than the variogram directly. In this thesis, Moran's I will be used to estimate the spatial autocorrelation, due to its ease of implementation and how computationally inexpensive it is.

The methods proposed are found and described in Table 3-1. In Section 5-3, these methods will be compared both to each other as well as to PCA and MAF, using both the *rim-* and *disc implementation* of EMAF.

The reasons for looking at the Unit Circle (UC) is that spatial autocorrelation generally decreases with distance, so similarly to the original MAF algorithm a unit shift might work best. Another method is to find the radius with the maximum spatial autocorrelation, Maximum Autocorrelation (Max AC). The idea is that if the maximum spatial autocorrelation does not occur at a unit length shift, this method will at least get close to evaluating the variogram where the maximum spatial autocorrelation occurs. The third method is to find the radius where the minimum spatial autocorrelation occurs, Minimum Autocorrelation (Min AC). Setting the radius of the ROI to the radius where the minimum spatial autocorrelation occurs might enable the data to be transformed and increase the spatial autocorrelation by the largest amount.

## 3-5  The problem with data sphering

As mentioned at the beginning of this chapter, this extension of MAF deviates from the originally proposed MAF algorithm. The original algorithm suggests sphering the data before solving the eigenvalue problem. The problem with that is that the sphering method has to be affine equivariant. A sphering function $S(\mathbf{X})$ is said to be affine equivariant if for any non-singular matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$

$$S(\mathbf{X}\mathbf{A}) = S(\mathbf{X})\mathbf{A}. \tag{3-49}$$

If the sphering method is not affine equivariant, the MAF algorithm loses its invariance towards linear transformations. This is similar to what is stated in Section 2-3-2, where the

original data is sphered using a sphering algorithm $S(\mathbf{X})$. When the sphered data is then transformed by a linear transformation $\mathbf{A} \in \mathbb{R}^{p \times p}$, we get

$$\mathrm{Cov}\left[S(\mathbf{ZA}), S(\mathbf{ZA})\right] = \left(S(\mathbf{ZA})\right)^T S(\mathbf{ZA}). \tag{3-50}$$

If the sphering method is affine equivariant,

$$\mathrm{Cov}\left[S(\mathbf{ZA}), S(\mathbf{ZA})\right] = \left(S(\mathbf{ZA})\right)^T S(\mathbf{ZA}) \tag{3-51}$$
$$= \mathbf{A}^T S(\mathbf{Z})^T S(\mathbf{Z}) \mathbf{A} \tag{3-52}$$
$$= \mathbf{A}^T \mathbf{I}_{p \times p} \mathbf{A}, \tag{3-53}$$

where $\mathbf{I}_{p \times p} \in \mathbb{R}^{p \times p}$ is the covariance matrix of the sphered dataset.

This statement is only true if both the sphering algorithm and the covariance estimator are affine equivariant. A covariance estimator, $\hat{\Sigma}(\mathbf{Z})$, is said to be affine equivariant if

$$\hat{\Sigma}(\mathbf{ZA} + \mathbf{b}) = \mathbf{A}^T \hat{\Sigma}(\mathbf{Z}) \mathbf{A}, \tag{3-54}$$

where $\mathbf{Z} \in \mathbb{R}^{n \times p}$ is the data matrix being analyzed, $\mathbf{A} \in \mathbb{R}^{p \times p}$ is any non-singular matrix and $\mathbf{b} \in \mathbb{R}^{n \times p}$ is a bias matrix [65].

By skipping the sphering step, finding a sphering method which is equivariant is not a problem. In his paper on sphering and its properties, Jian Zhang presents three different sphering methods [66]. None of these sphering methods are fully equivariant, namely the Square Root Decomposition (SRD) algorithm is orthogonally equivariant, that is the transformation matrix is limited to being orthogonal instead of only non-singular. In [66], Jian Zhang points out that the three methods covered in his paper are the three methods found in literature. So whether an affine equivariant sphering method exists, is unanswered. For these reasons the EMAF factors are calculated using the single-step method, avoiding the sphering.

## 3-6  Robustness

Geir Storvik suggests, in his paper on data reduction by separation of signal and noise components for multivariate spatial images [67], to use a robust estimator to estimate the covariance matrix for the MAF algorithm. Storvik does not specifically recommend one single robust covariance estimator, but instead compares the performance of three different estimators. It is concluded that using robust estimators when implementing MAF yields factors containing more information, especially if the underlying classes in the data do not mix too much.

Robust estimators of the covariance matrix are estimators which are resistant towards outlying measurements. The EMAF algorithm is used to factorize and preprocess data, it is highly likely that the data contains a large proportion of noisy measurements better. Those measurements will, if not handled correctly, skew the estimated covariance matrix. A robust estimator handles these noisy measurements. The performance of these robust estimators is often measured by a metric called the breakdown point. The breakdown point is the signal-to-noise ratio of the data the estimator can handle without loosing performance. An estimator might have a breakdown point of 10%, meaning that it can handle data where up to 10% of the measurements are noisy or outliers.

Since Storvik published his paper, a new high breakdown, affine equivariant covariance estimator has been introduced. The algorithm in question is the Minimum Covariance Determinant (MCD) [68, 69]. As discussed in the previous section, the covariance estimator needs to be affine equivariant for MAF, otherwise MAF will not be invariant to linear transformations. Furthermore, since MAF operates by transforming the data linearly, if the covariance estimator was not affine equivariant the transformed data would not have the same covariance estimate as the original data. This is discussed in [65].

MCD is considered a good candidate covariance estimator to use in EMAF, first of all because it is an affine equivariant covariance estimator. Secondly, MCD has the highest possible breakdown point for affine equivariant estimators. Its breakdown point is $\left\lfloor \frac{n-p+2}{2} \right\rfloor$, where $\lfloor \cdot \rfloor$ is the floor function, when estimating the covariance of a matrix with $p$ variables and $n$ observations [68]. Thirdly, the MCD has been implemented using an algorithm which is considered faster than comparable algorithms [70]. When concluding about the robust covariance estimation, Storvik mentions that the robust covariance estimators he tried were all computationally expensive. A downside to using MCD is that it is recommended to have a ratio of $\frac{n}{p} < 5$ between observations and variables [68]. It has also been pointed out that MCD seems to have poor performance in high dimensions [65]. That problem has been adressed with a version of MCD which is designed specifically for high dimensions [71], but that estimator is not affine equivariant, hence not suitable for EMAF.

For some applications the high dimensional performance might be a limit to keep in mind, although for IMS datasets are often peak picked before any processing which in some cases might keep the datasets within reasonable limits for application of MCD. For those reasons, robust estimation of the covariance matrix will be kept optional in the implementation of EMAF. In the case of the IMS-RBDS dataset for example, the MCD estimation is applicable since we end up with a ratio of about 27 observations per variable. Although if the dataset is analysed before peak picking, this ratio is down to about 3 obersations per variable, which is not enough for the MCD. For this reason, it is still considered impractical to apply robust estimation of the covariance matrix to IMS data in general, although it is a promising future step. On the other hand, for datasets that generally contain a low number of variables compared to obersevations, for example RBG images, this method is considered a good addition to the EMAF algorithm.

In Figure 3-2, the resulting factors are shown for when EMAF is carried out on the Imaging Mass Spectrometry - Rat Brain Dataset (IMS-RBDS) dataset in a robust setting using a unit circle ROI. The resulting factors seem very similar to the equivalent non-robust factors, seen in Figure 5-23. The factors are so similar that comparing them visually is not useful. In order to assess the increase in quality, an artificial dataset will be needed.


## 3-7   Properties of EMAF

In this section an overview of the properties of EMAF the extended version of MAF, is given. This is mainly done to underline that the attractive traits of MAF, such as invariance towars linear transformations, are still present in EMAF.

It has previously been shown that the standard MAF algorithm is invariant to linear transformations. This is an important feature, as preprocessing steps like corrections for instrument
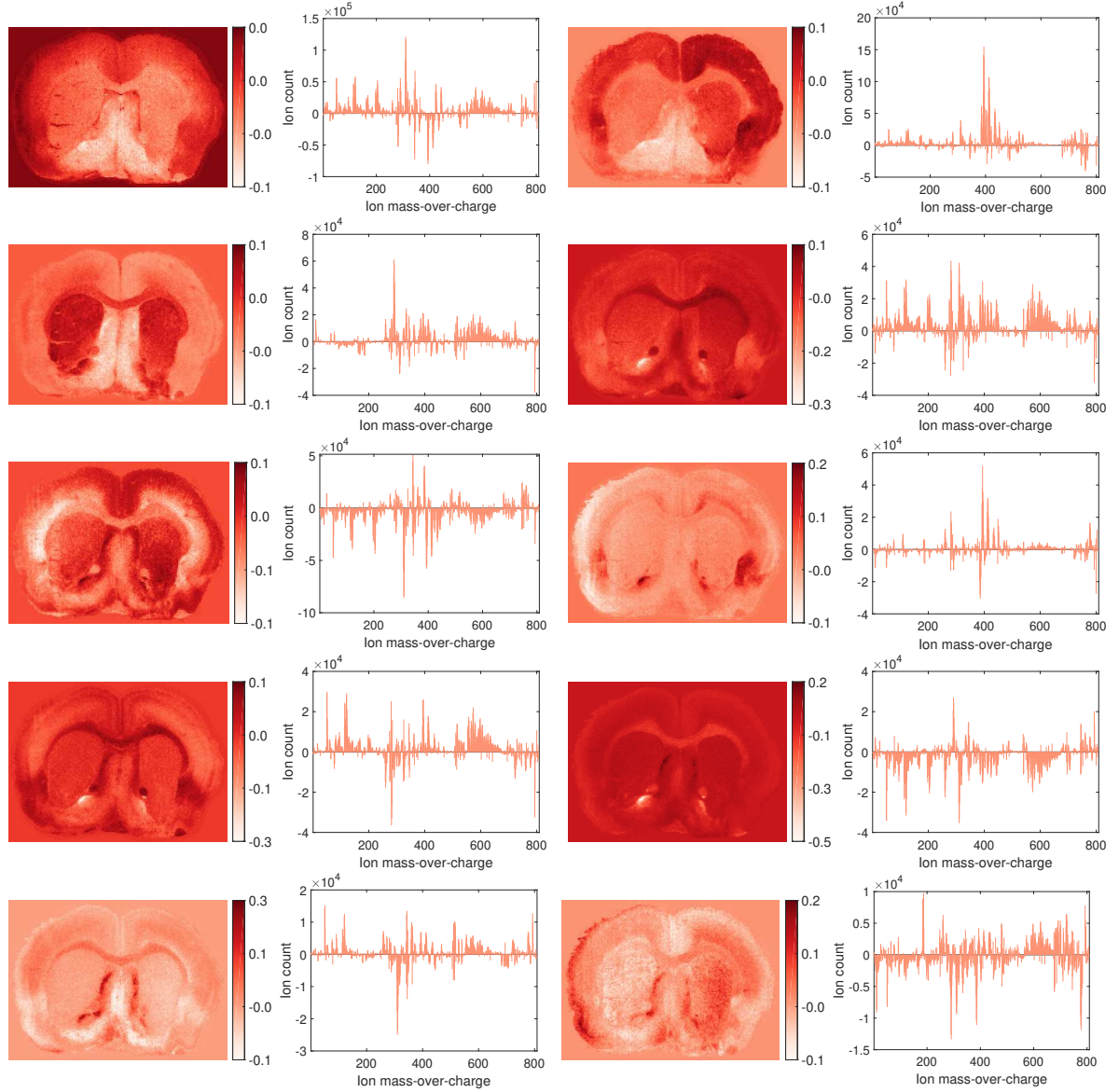
**Figure 3-2:** EMAF robust unit circle factors 1 - 10. Evaluating the quality of these factors is not practical through visual inspection, since these factors are very similar to the EMAF UC factors.

gain on different bands would otherwise alter the resulting factors. To show that EMAF is also invariant of linear transformations, EMAF is carried out on a transformed dataset

$$\mathbf{Y} = \mathbf{ZA}, \tag{3-55}$$

where $\mathbf{Z} \in \mathbb{R}^{n \times p}$ is the dataset before transformation and $\mathbf{A} \in \mathbb{R}^{p \times p}$ is any non-singular transformation matrix. Calculating the covariance matrix and variogram for this transformed

dataset gives

$$\text{Cov}\left[\mathbf{Y}(\mathbf{x}), \mathbf{Y}(\mathbf{x})\right] = \text{Cov}\left[\mathbf{Z}(\mathbf{x})\mathbf{A}, \mathbf{Z}(\mathbf{x})\mathbf{A}\right] = \mathbf{A}^T\mathbf{C}(\mathbf{0})\mathbf{A}, \qquad (3\text{-}56)$$

$$\text{Cov}\left[\mathbf{Y}(\mathbf{x}+\mathbf{h}) - \mathbf{Y}(\mathbf{x}), \mathbf{Y}(\mathbf{x}+\mathbf{h}) - \mathbf{Y}(\mathbf{x})\right] = \text{Cov}\left[\mathbf{Z}(\mathbf{x}+\mathbf{h})\mathbf{A} - \mathbf{Z}(\mathbf{x})\mathbf{A}, \mathbf{Z}(\mathbf{x}+\mathbf{h})\mathbf{A} - \mathbf{Z}(\mathbf{x})\mathbf{A}\right]$$
$$(3\text{-}57)$$

$$= \mathbf{A}^T\gamma(\mathbf{h})\mathbf{A}. \qquad (3\text{-}58)$$

If this is applied to the eigenvalue problem for the *rim implementation* of EMAF, that gives

$$\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\mathbf{A}^T\gamma(\mathbf{h})\mathbf{A}\Delta\theta v_i = \lambda_i\mathbf{A}^T\mathbf{C}(\mathbf{0})\mathbf{A}v_i \qquad (3\text{-}59)$$

$$\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(\mathbf{h})\Delta\theta\mathbf{A}v_i = \lambda_i\mathbf{C}(\mathbf{0})\mathbf{A}v_i, \qquad (3\text{-}60)$$

where $v_i \in \mathbb{R}^{p\times 1}$ are the eigenvectors for the transformed dataset. This works since $\mathbf{A}$ is invariant of $\mathbf{h}$ and can therefore be taken outside of the sum. Defining $w_i = \mathbf{A}v_i$, gives

$$\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(\mathbf{h})\Delta\theta\mathbf{A}v_i = \lambda_i\mathbf{C}(\mathbf{0})\mathbf{A}v_i \qquad (3\text{-}61)$$

$$\frac{1}{2\pi}\sum_{\theta=0}^{2\pi}\gamma(\mathbf{h})\Delta\theta w_i = \lambda_i\mathbf{C}(\mathbf{0})w_i. \qquad (3\text{-}62)$$

This is the exact same result as for the non-transformed data, which means that the factors remain the same. The same result applies to the *disc implementation* of EMAF.

Another feature of MAF, shown for example in [58], is that the MAF factors are invariant of the choice of shift parameter when there are only two or fewer isotropic structures present in the data being factorized. The same if true for EMAF, this is shown in the A-5 Section of the Appendix. This result applies both to the *disc-* and *rim implementation* of EMAF.

In addition to these features, EMAF is fully unsupervised, in the sense that no prior knowledge is required to run the algorithm. It is particularly worth emphasizing that the EMAF factors are not only calculated from a single shift parameter, but from an average of all shift parameters with the same length. This is a similar idea as when the experimental variogram is calculated, as demonstrated in Section 2-2-3. In the EMAF setting this gives a more realistic value to the factors instead of including a single shift.

# Chapter 4

# Experiments

In order to motivate and evaluate the extended version of Maximum Autocorrelation Factorization (MAF) presented in Chapter 3, a set of experiments need to be conducted and for that, datasets are needed. In some cases, artificial datasets exhitibing certain features have to be constructed. In other cases, real-world datasets will be used. This chapter starts out by giving an overview of the experiments conducted. Furthermore, we motivate why said experiments are carried out and what the expected result from each of them is. The datasets used throughout this thesis will be described, as well as how the artificial datasets are constructed, where the real-world dataset comes from, and what the datasets will be used for. A tabular overview of the datsets can be found in Table 4-1.

**Table 4-1:** An overview of the datasets used in this thesis.

| Name | Description | Chapter |
|------|-------------|---------|
| ArtDirDS | Artificially created dataset with different sized spatial artifacts oriented in various directions. | 4-1-2 |
| ArtLenDS | Artificially created dataset with spatial artifacts that have different sized circles. That is spatial artifacts that have the same length in all directions. | 4-1-3 |
| IMS-RBDS | A real-world Imaging Mass Spectrometry (IMS) dataset obtained from a coronal rat brain section as part of a study on Parkinson's disease [4]. | 4-1-4 |

The goal is to motivate and evaluate the alterations made to MAF. In order to do so, three experiments will be carried out using the same dataset, but different factorization methods. Each experiment is given a name, which references the method in question. A tabular overview of these experiments can be found in Table 4-2. For example, an experiment analysing if MAF is dependent on the direction of the shift parameter would be called DirDep - MAF.

**Table 4-2:** An overview of the experiments conducted in this thesis.

| Name | Description | Dataset |
|------|-------------|---------|
| DirDep - *[method]* | Analyze whether [Method] factors are dependent on the spatial direction of the shift parameter. | ArtDirDS |
| LenDep - *[method]* | Analyze whether [Method] factors are dependent on the spatial length of the shift parameter. | ArtLenDS |
| MEval - *[method]* | Evaluation of the performance of the various alterations proposed. | IMS-RBDS |

To motivate the alterations made to the MAF algorithm, the DirDep and LenDep experiments are carried out. With the DirDep - *[method]* experiment a certain factorization method is applied to the ArtDirDS dataset, passing as arguments shift parameters with varying direction. This dataset has spatial artifacts which have different lengths in different directions. If the method in question is dependent on the direction of the shift parameter, it will probably promote spatial artifacts in the dataset of a certain direction. If the method does not return the same factors, for the different input directions, it is an indication that the method in question is dependent on the direction of the shift parameter.

The LenDep - *[method]* experiment is carried out in order to analyze if a particular method is dependent on the physical length of the shift parameter, by applying the same method with shift parameters of different lengths as arguments. This is done on the ArtLenDS dataset that consists of spatial artifacts with different lengths, independent of direction. If the method in question is dependent on the length of the shift parameter, it will return different factors for different shifts for different input lengths.

The MEval experiment is carried out to compare the different alterations made to the MAF algorithm. More specifically, comparing the *rim-* and *disc implementations* with the three different distance determination methods listed in Table 3-1. For this experiment, all versions of the Extended Maximum Autocorrelation Factorization (EMAF) algorithm are applied, as well as applying Principal Component Analysis (PCA) and MAF for comparison. All methods will be carried out on the IMS-RBDS dataset. Spatial autocorrelation and other metrics will be used to evaluate and analyze the resulting factors. The metrics used will be discussed in detail in Section 5-3.

## 4-1   Datasets

### 4-1-1   Construction of artificial datasets

In the following section a number of experiments will be carried out on artificial datasets. To construct artificial multivariate images, a set of spatial distributions and spectral distributions (or spectra) are generated. Being an artificial dataset, neither of the distributions can be tied

to any specific physical meaning. However, to give some idea about the function of each of these distributions, the spatial distribution corresponds to ion images in the case of IMS and the spectra correspond to mass-over-charge spectra in IMS. In the following section, when the datasets are introduced, these distributions will be shown.

To construct the dataset, the spatial and spectral distributions are multiplied together, which assigns spectral intensity to the spatial distributions. Each peak on the spectral distribution will indicate how strong or intense the spatial distribution is at the spectral location. To be able to multiply the two distributions together, the spatial distributions first need to be vectorized, that is reshaped from being an $n_x \times n_y$ matrix to being an $n_x \cdot n_y \times 1 = n \times 1$ vector. So the vectorized, $n \times 1$ spatial distribution and the $1 \times p$ spectral distribution are multiplied together forming an $n \times p$ matrix. The set of all these $n \times p$ matrices are then all added up, resulting in the final dataset. This procedure is shown in Figure 4-1. When carrying out factorization, this procedure is essentially reversed using the constraints and specific goals of the factorization algorithm.
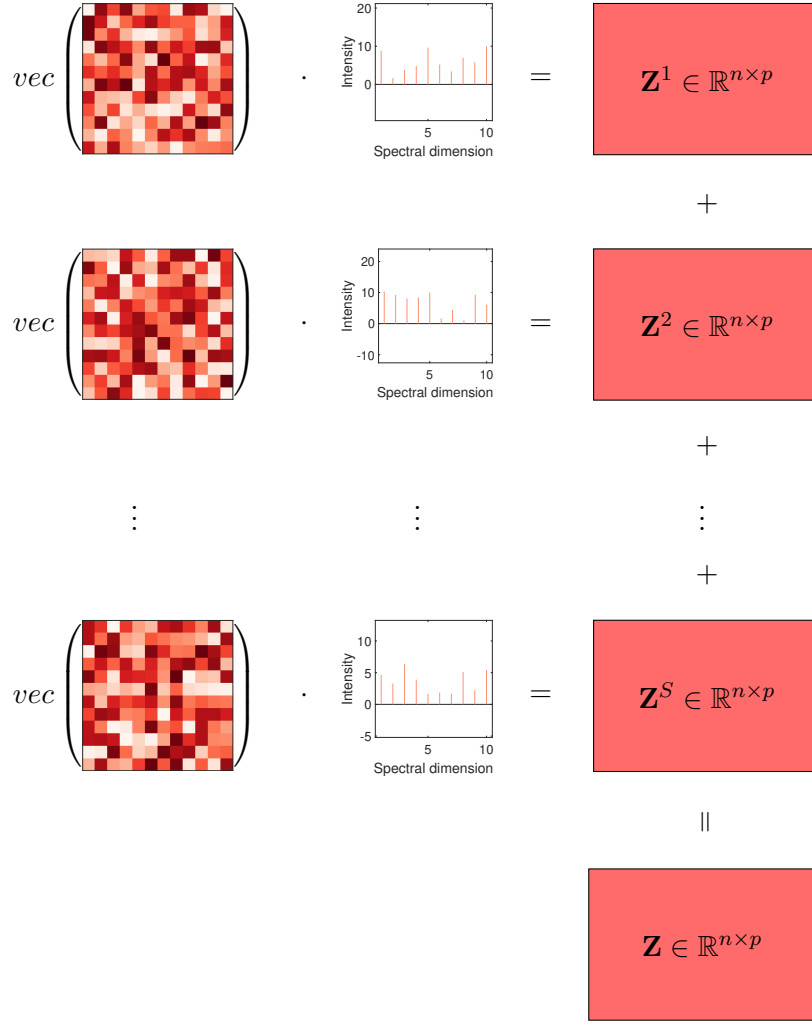
**Figure 4-1:** An illustration of the procedure of generating artificial datasets. In this example, the dataset $\mathbf{Z} \in \mathbb{R}^{n \times p}$ is constructed from $S$ different spatial and spectral distributions, such that $\mathbf{Z} = \sum_{s=1}^{S} \mathbf{Z}^s$, where $\mathbf{Z}^s \in \mathbb{R}^{n \times p}$ for all $s \in [1, 2, \ldots, S]$. The spatial and spectral distributions may be constructed to suit the experiment. In this case the procedure is illustrated with randomly generated spatial and spectral distributions. The vectorized or flattened spatial distributions will form $n \times 1$ vectors and the spectral distribution form $1 \times p$ vectors, which are multiplied together forming the $n \times p$ matrices. Note that $s$ in this case is an index and not a power. The $vec(\cdot)$ operator flattens an $n_x \times n_y$ matrix into an $(n_x \cdot n_y) \times 1$ matrix.

## 4-1-2    The ArtDirDS Dataset

The artificial dataset used to examine whether a certain method is dependent on the direction of the shift parameter is shown in Figure 4-2. This dataset is refered to as ArtDirDS. The dataset is made up of five different spatial distributions, each with a shape oriented in a certain direction. That enables us to see if a method with a given shift parameter promotes spatial artifacts oriented in a certain direction over any other. The spectral distributions used are simply unit impulses at different spectral locations. That means that the factors are spectrally separated. The factors are constructed of a unit vector loading and score, containing a shape which is mostly oriented in a certain direction.



**(a)** Factor 1- Score    **(b)** Factor 1- Loading    **(c)** Factor 2- Score    **(d)** Factor 2- Loading

**(e)** Factor 3- Score    **(f)** Factor 3- Loading    **(g)** Factor 4- Score    **(h)** Factor 4- Loading

**(i)** Factor 5- Score    **(j)** Factor 5- Loading
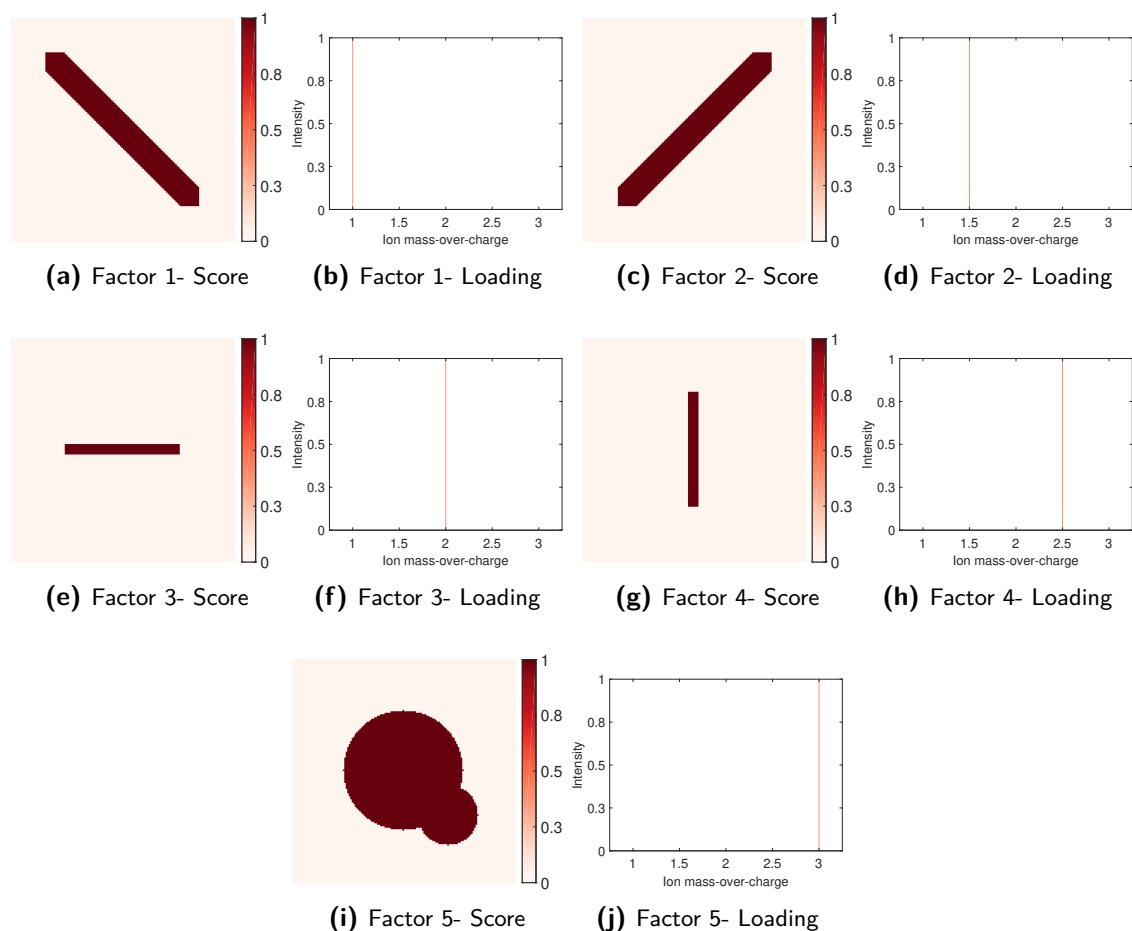
**Figure 4-2:** An unmixed, artificial dataset used to examine the dependence of MAF factors to the shift parameter direction. The dataset is constructed of five factors, each a multiple of a unit vector loading and a score containing a shape that is oriented mostly in a certain direction.

## 4-1-3 The ArtLenDS Dataset

In order to examine how the shift parameter length affects a given method, another artificial dataset introduced, shown in Figure 4-3. The spatial distributions used to construct this dataset are circles of different diameter. Using circles ensures that the length of the shapes is independent of which direction the images are shifted in. This decreases the effect of the shift parameter length on the results. For the spectral distributions, unit vectors at separated spectral locations are used. Each factor is constructed from a unit vector loading and a score matrix with shapes of different sizes.



**(a)** Factor 1- Score        **(b)** Factor 1- Loading

**(c)** Factor 2- Score        **(d)** Factor 2- Loading

**(e)** Factor 3- Score        **(f)** Factor 3- Loading
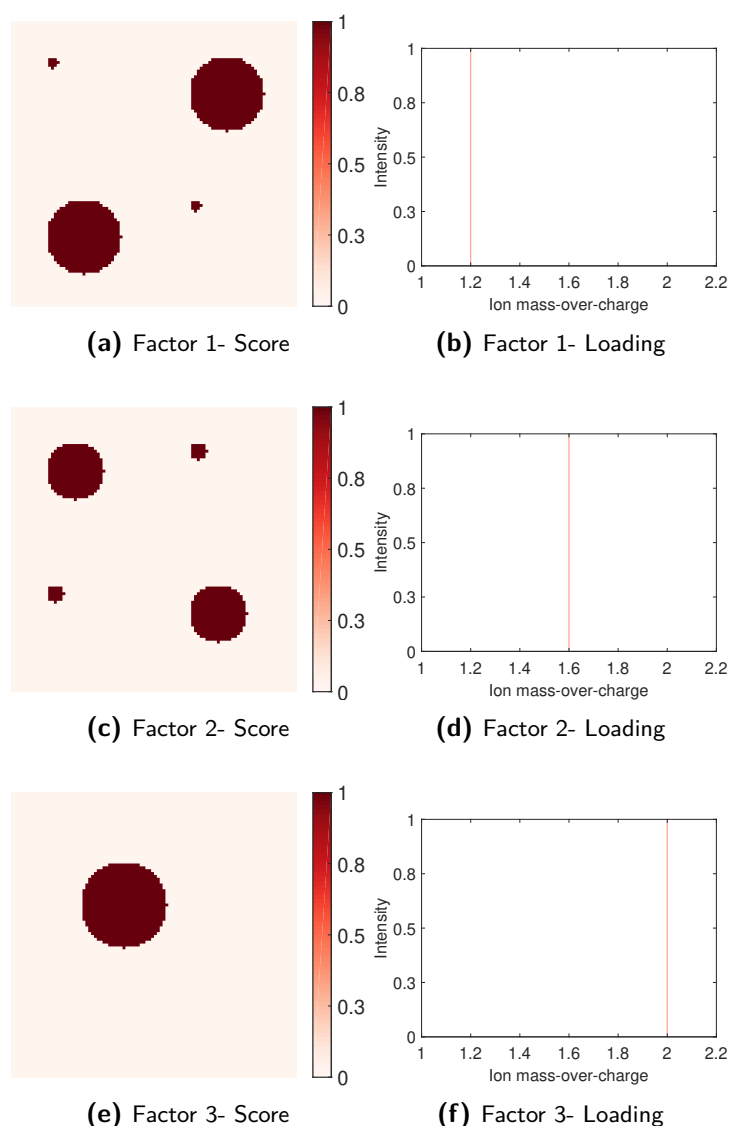
**Figure 4-3:** An unmixed, artificial dataset used to examine the dependence of MAF factors to the shift parameter length. The dataset is constructed of three factors, each a multiple of a unit vector loading and a score containing a shape that will result in varying variogram ranges.

### 4-1-4    The IMS-RBDS Dataset

The real-world dataset used in this thesis describes a section of a rat brain where scientists have simulated Parkinson's disease by compromising the rat's dopamine receptors in one hemisphere of the brain. The other hemisphere of the brain was left untouched as a control. Note that all in-house animal experiments were performed with approval by the Vanderbilt Institutional Animal Care and Use Committee. The rat brain was frozen, prepared, and sectioned coronally into $10\mu m$ sections. Each section was then prepared for Matrix-Assisted Laser Desorption/Ionization (MALDI) measurements and finally measured in a 15T Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer with a spatial sampling resolution of $75\mu m$ and a mass resolving power of $50000\frac{m}{FWHM}$ at $m/z$ 5000. The molecular images lie on a $m/z$-range from 1300 to 24000 with a total of about 21000 pixels. After measuring, the data was peak picked, resulting in a total of 809 $m/z$-peaks. This procedure is described in full detail in [4]. Sample ion images from this dataset are shown in Figure 4-4 and sample mass-over-charge spectra are shown in Figure 4-5.



**(a)** $m/z$ 5629.0            **(b)** $m/z$ 5028.5            **(c)** $m/z$ 2663.2
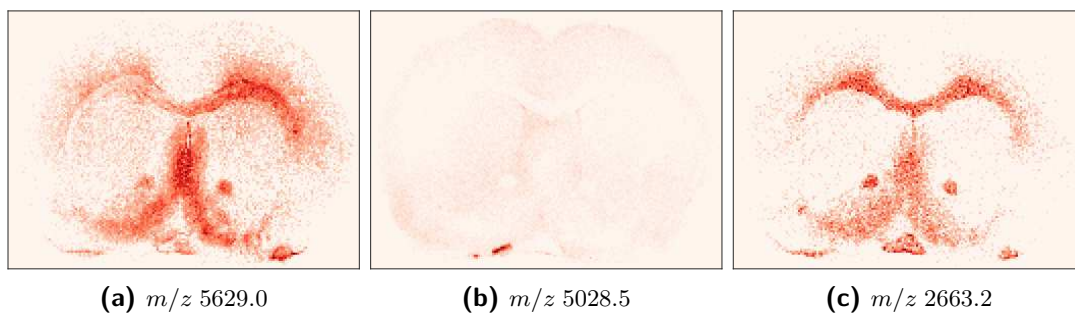
**Figure 4-4:** Three random samples of ion images from the rat brain dataset.



**(a)** pixel number 12488.        **(b)** pixel number 8368.        **(c)** pixel number 20006.
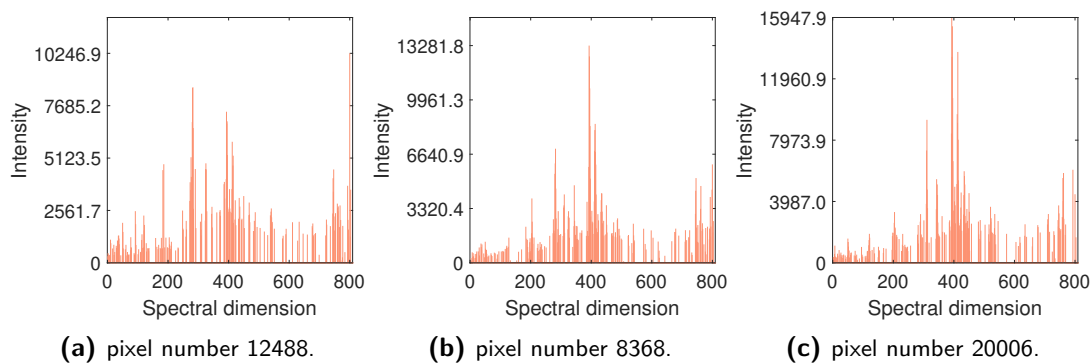
**Figure 4-5:** Three random samples of mass spectra from the rat brain dataset.

# Chapter 5

# Discussions and results

## 5-1 Dependency on shift parameter direction

### 5-1-1 MAF

In past literature, it has been pointed out that under certain conditions Maximum Autocorrelation Factorization (MAF) factors are invariant of the shift parameter $\mathbf{h}$ [58, 18]. One of the goals of this experiment is to explore cases where these conditions are not met, analyzing if there are potential improvements that could be made to MAF. specifically we are interested in changes to address the cases where MAF factors are invariant of the shift parameter, where the variable(s) being factorized do(es) not contain more than two isotropic covariance structures. The ArtDirDS is designed to test these features of a factorization method, so the ArtDirDS dataset will be used for this experiment. This means that a variable for example can contain two isotropic covariance structures or contain a single anisotropic covariance structure, which would be equivalent to two isotropic covariance structures. The conditions for invariance are outlined both in Section 2-3-2 and in [58].

In order to analyze if the conditions where MAF is invariant of the shift parameter are met or not, the two- and three-dimensional representations of each variable's variogram analysed. The two-dimensional representation is shown in Figure 5-1 and the three-dimensional representations are shown in Figure 5-2.

In Figure 5-1 the variogram is shown as a function of shift distance in pixels. The shift distance in pixels is the same as looking at the length of the shift parameter, i.e. $\|\mathbf{h}\|$. In Figure 5-2 the three-dimensional variogram is shown as a function of the shift parameter $\mathbf{h} = (r, \theta)$. In this case, the shift parameter is described in polar coordinates, but the shift parameter is still limited to values $\mathbf{h} = (x, y)$ with $x, y \in \mathbb{Z}$.

Note that in Figure 5-2 all of the variables are at least anisotropic, i.e. the variograms change with both the direction and length of the shift parameter. In addition to all the variograms being anisotropic, they also all have a nugget-effect. That is, the value of the variogram is not estimated to be zero at a shift distance of zero. The nugget-effect is a

separate covariance structure [72], so in this case there are at least two covariance structures, one of which is anisotropic. This means that the resulting MAF factors will vary with different shift parameters.

To gain some intuition on how MAF works under different angles, the ArtDirDS dataset is factorized using MAF with different shift parameters as input and the resulting factors interpreted. In order to limit the effect of the length of the shift parameter, only shifts lying close to the unit circle will be used, with the limitation that the shift coordinates are limited to integers. The three shift parameters used in this experiment are: $\mathbf{h} = [0, 1]$, $\mathbf{h} = [1, 0]$ and $\mathbf{h} = [1, 1]$.
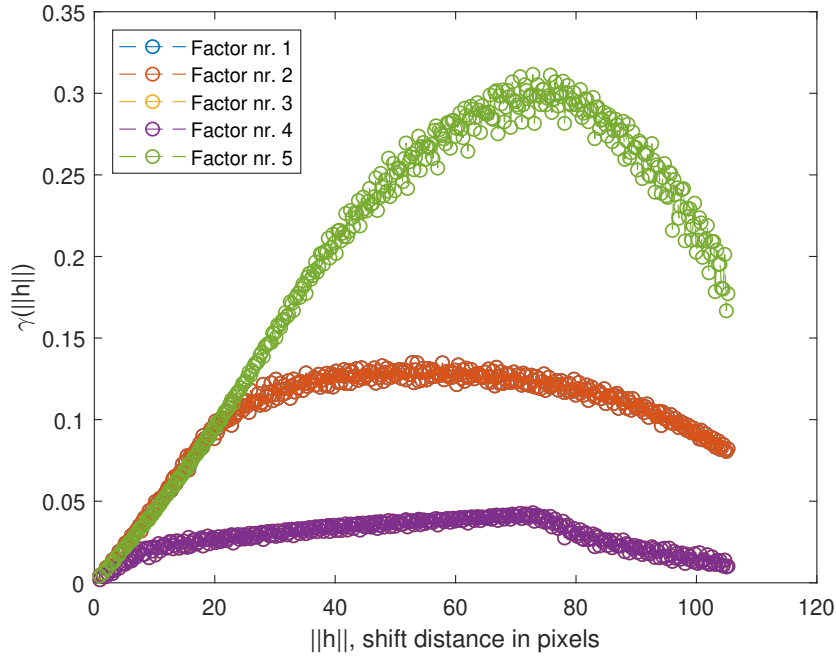


**Figure 5-1:** The isotropic version of the variograms of all the five factors in the Artificial, Directionally Dependent Dataset (ArtDirDS) dataset. In these 2D variogram plots, the value is averaged over all directions of shift vectors of a certain length. There is a high similarity between this 2D plot and the 3D plots in 5-2. Note that factors 1 and 2 share variograms as do factors 3 and 4.

Figure 5-3 shows that the algorithm promotes horizontal lines and demotes vertical lines. This is because a copy of the dataset is shifted by $\mathbf{h} = [0, 1]$ and the difference between this shifted dataset and the original, non-shifted dataset determines the order of the factors. Looking at the cases of the horizontal and vertical lines, if the horizontal line is shifted by $\mathbf{h} = [0, 1]$, that is one pixel vertically. That will result in a difference of a single pixel times the length of the line. The vertical line on the other hand will result in a difference of a single pixel times the line width. Since the line length is larger than the line width, the horizontal line gets promoted.

The opposite case is seen with the shift parameter $\mathbf{h} = [1, 0]$, as shown in Figure 5-4. There the vertical line gets promoted and the horizontal line gets demoted. The same reasoning still applies, the shift parameter has been rotated by $\pi/2$ rad and therefore also the results. It is worth noting that scores 3 and 4 remain similar, so the algorithm has found some factors

**(a)** Factor 1- Variogram

**(b)** Factor 2- Variogram

**(c)** Factor 3- Variogram

**(d)** Factor 4- Variogram

**(e)** Factor 5- Variogram
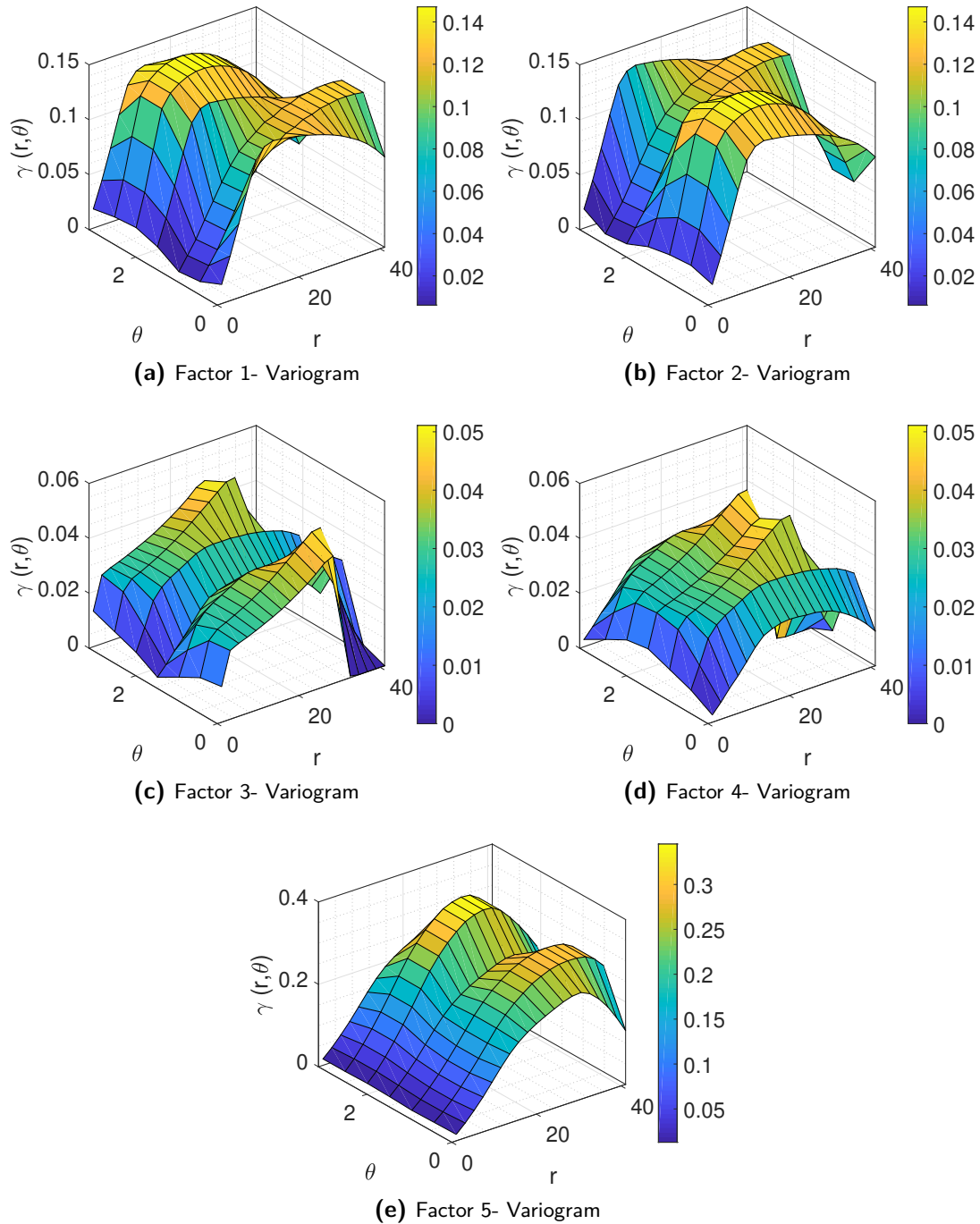
**Figure 5-2:** The variogram of each individual variable of the ArtDirDS dataset. Notice that each factor's variogram changes both with distance and angle of the shift parameter, that is all the factors are anisotropic.

which result in relatively high autocorrelation for both $\mathbf{h} = [0, 1]$ and $\mathbf{h} = [1, 0]$.

In the last case MAF is carried out with the shift parameter $\mathbf{h} = [1, 1]$, the resulting factors are shown in Figure 5-5. This shift does not lie exactly on the unit circle, but the effects of the shift parameters length should not affect the results. This is concluded since the

vertical ordering of the variograms, in Figure 5-1, does not change between a shift distance of 1 and $\sqrt{2}$, which are the shift distances used in this experiment. For this shift parameter, the algorithm seems to promote spatial artifacts that are oriented diagonally. Note that the horizontal and vertical lines do not appear on their own anymore, but only in combination.

So in all the cases examined so far the resulting factors are dependent on the direction of the shift parameter. The variograms that are shown in Figure 5-2, could have been used to deduce which variables would rank highly for each of the shift parameters. The main takeaway from this experiment is that the MAF factors do depend on the direction of the shift parameter when more than two covariance structures are present.

When MAF was first introduced by P. Switzer and A. A. Green in [18], they proposed using the shift parameter $\mathbf{h} = [1,1]$. However, as shown in Figure 5-5, that shift parameter on its own will promote scores with a certain orientation. When dealing with natural phenomena there might not be any reason to promote autocorrelation in one direction rather than any other. In the general use case the first MAF factors should contain spatially large shapes with high loading values, regardless of the directional orientation of the shapes of the score. This is one reason the shift parameter needs to be determined differently.



**(a)** Score 1     **(b)** Loading 1     **(c)** Score 2     **(d)** Loading 2

**(e)** Score 3     **(f)** Loading 3     **(g)** Score 4     **(h)** Loading 4

**(i)** Score 5     **(j)** Loading 5
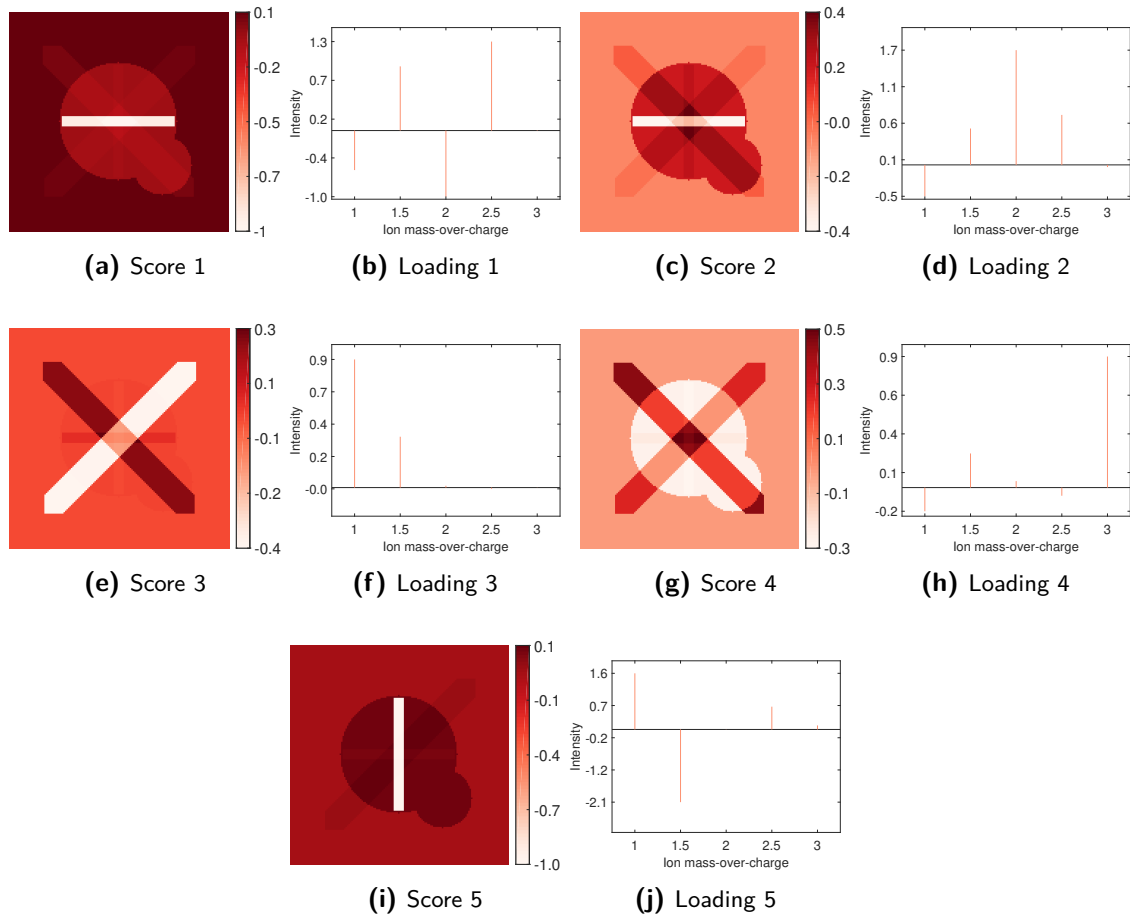
**Figure 5-3:** The resulting factors and loadings when MAF is applied to the ArtDirDS dataset with the shift parameter $\mathbf{h} = [0,1]$.

**(a)** Score 1

**(b)** Loading 1

**(c)** Score 2

**(d)** Loading 2

**(e)** Score 3

**(f)** Loading 3

**(g)** Score 4

**(h)** Loading 4

**(i)** Score 5
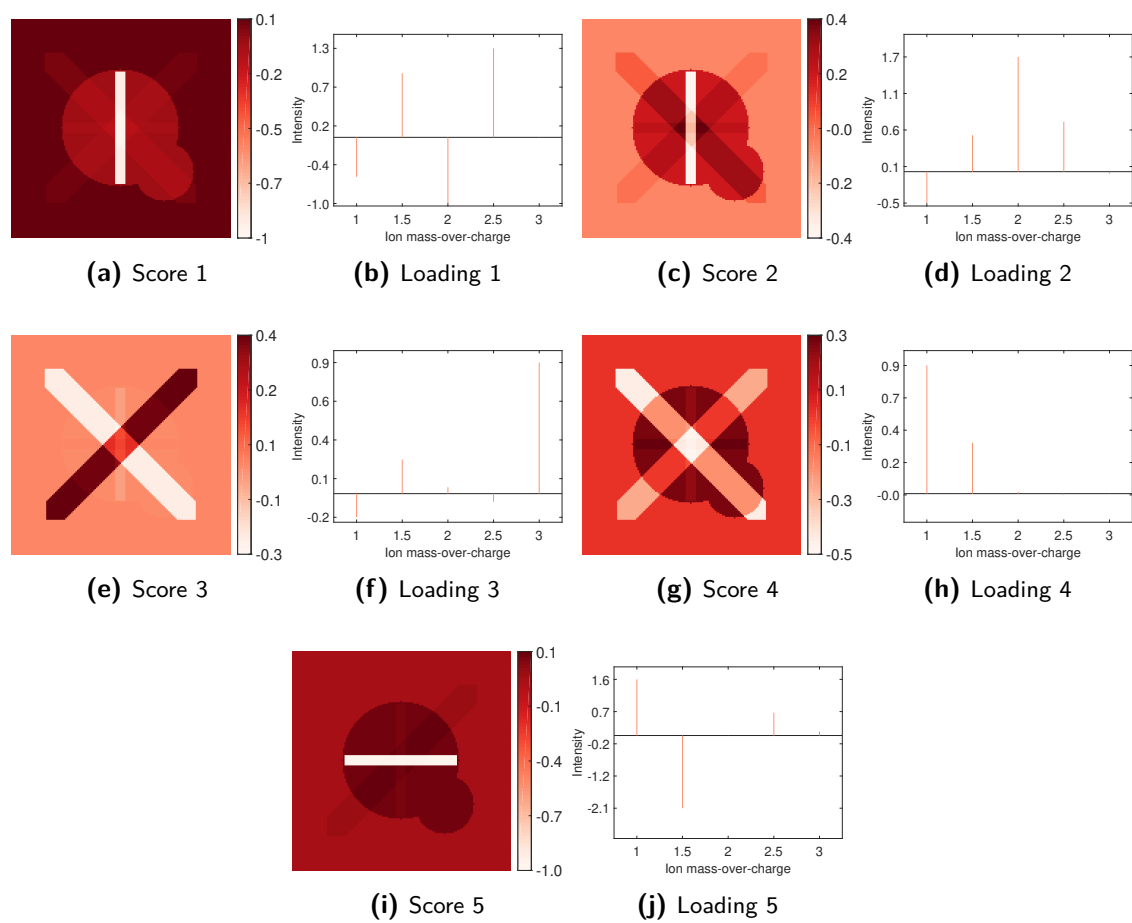
**(j)** Loading 5

**Figure 5-4:** The resulting factors and loadings when MAF is applied to the ArtDirDS dataset with the shift parameter $\mathbf{h} = [1, 0]$.

**(a)** Score 1    **(b)** Loading 1    **(c)** Score 2    **(d)** Loading 2

**(e)** Score 3    **(f)** Loading 3    **(g)** Score 4    **(h)** Loading 4

**(i)** Score 5    **(j)** Loading 5
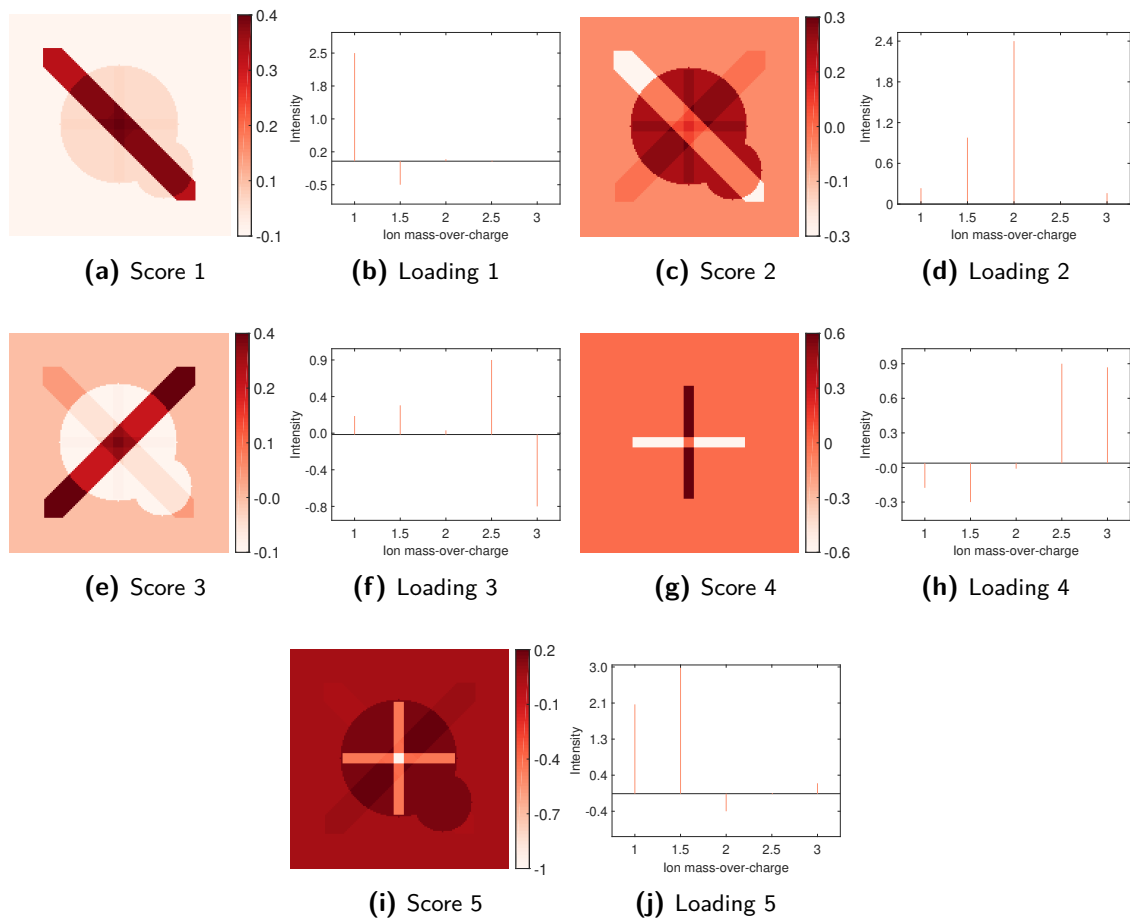
**Figure 5-5:** The resulting factors and loadings when MAF is applied to the ArtDirDS dataset with the shift parameter $\mathbf{h} = [1, 1]$.

## 5-1-2 EMAF

To enable a comparison of Extended Maximum Autocorrelation Factorization (EMAF) and MAF, the DirDep experiment is also carried out using EMAF. This is also done to ensure that EMAF does not favour spatial artifacts oriented in a certain direction. That is, EMAF should return factor which are not oriented in any particular direction, at least not on this dataset as the spatial distributions pointing in opposite directions have the same size or surface area. To make the comparison fair, the EMAF UC method is used, since only shifts lying on the unit circle were used when conducting the experiment with MAF. The resulting factors are shown in Figure 5-6.

Looking at the EMAF factors in Figure 5-6, we notice that the algorithm does not seem to favour spatial distributions oriented in any direction. In all the factors the spatial artifacts that are oriented in a certain direction (e.g. the horizontal and vertical lines) always appear in pairs, meaning that the EMAF algorithm is not favouring one direction over another.



**(a)** Score 1    **(b)** Loading 1    **(c)** Score 2    **(d)** Loading 2

**(e)** Score 3    **(f)** Loading 3    **(g)** Score 4    **(h)** Loading 4

**(i)** Score 5    **(j)** Loading 5
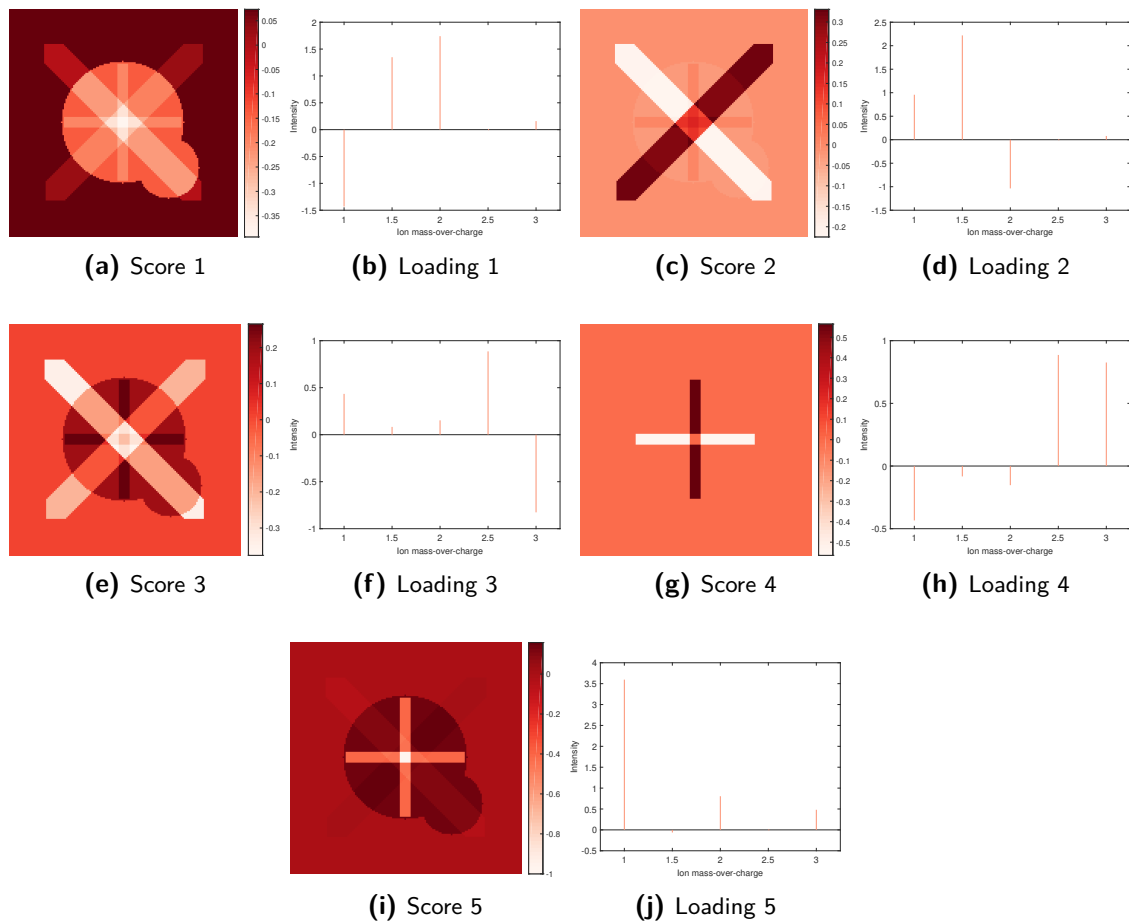
**Figure 5-6:** The resulting factors and loadings when EMAF is applied to the ArtDirDS dataset.

## 5-2   Dependency on shift parameter length

### 5-2-1   MAF

We start by analyzing if the dataset meets the criteria where MAF is invariant of the shift parameter. In order to do that, the variogram of the ArtLenDS dataset is analyzed. Figure 5-7 shows the two-dimensional representation of the variogram of the three factors of the dataset. The variogram is shown as a function of shift distance in pixels. The shift distance in pixels is the same as looking at the length of the shift parameter, i.e. $\|\mathbf{h}\|$. Notice how the different ranges of the factors correlate with the shapes seen in the respective scores of the factors. Note for example, the peak on the variogram for factor 2 at a shift value of length 20, the larger circles in the spatial distribution of factor 2 have a radius of 20.

Since the variograms show a large spread, the three-dimensional representation of the variogram for each variable should be analyzed, those variograms are shown in Figure 5-8. The three-dimensional variogram is shown as a function of the shift parameter $\mathbf{h} = (r, \theta)$. Note that in this case the shift parameter is described in polar coordinates, but the shift parameter is still limited to values $\mathbf{h} = (x, y)$ with $x, y \in \mathbb{Z}$. Looking at the three-dimensional variograms, we notice that the variables are indeed anisotropic. All the variograms exhibit somewhat of a nugget-effect and a nugget-effect is considered to be a separate covariance structure [72]. That means each variable has at least two covariance structures, one of which is anisotropic, so the resulting MAF factors should change with different shift parameters.

In order to analyze the effect of the shift parameter length on MAF factors, MAF will now be carried out with three different shift parameters, oriented in the same direction but with different lengths. In this case, the lengths chosen were $\|\mathbf{h}\| \in \{15, 26, 40\}$. These lengths were chosen with the variogram in mind. The behavior of the factors will be analysed in relation to the variogram. So for that reason, the three shift lengths are chosen at a location where the variograms are separated, but in a different order compared to the other shift lengths. The shift parameter lengths are shown as vertical lines on the variogram in Figure 5-9.

The resulting factors for the shift $\|\mathbf{h}\| = 15$ are shown in Figure 5-10, the factors for the shift $\|\mathbf{h}\| = 26$ in Figure 5-11, and the factors the shift $\|\mathbf{h}\| = 40$ in Figure 5-12. One thing that we do note is that the order of the resulting factors seems to change according to the order of the variogram. No two shifts return the same order of factors, although the factors seem relatively similar with regard to the spatial distribution.
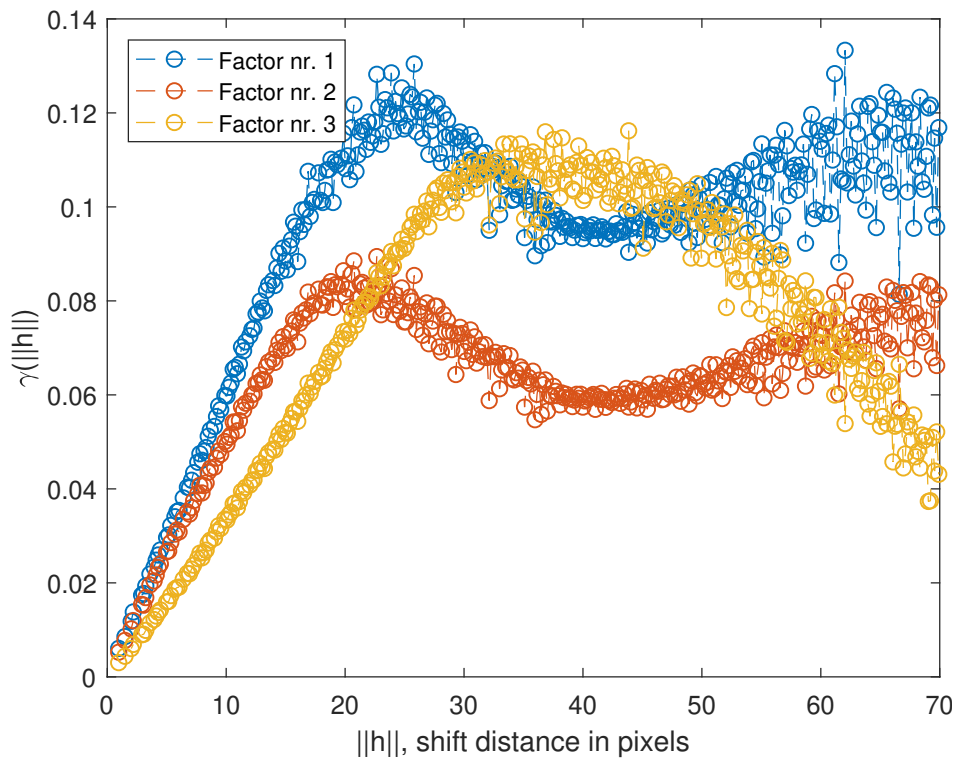
**Figure 5-7:** Variogram of the factors of the Artificial, Length Dependent Dataset (ArtLenDS) dataset. The range of the three factors changes with the length of the shift parameter. It is worth mentioning that these factor are also anisotropic, but they were not used for the directional dependence experiment since it is harder to intuitively argue why one factor is dominant rather than another for a certain direction of shift.

**(a)** Factor 1- Variogram



**(b)** Factor 2- Variogram



**(c)** Factor 3- Variogram
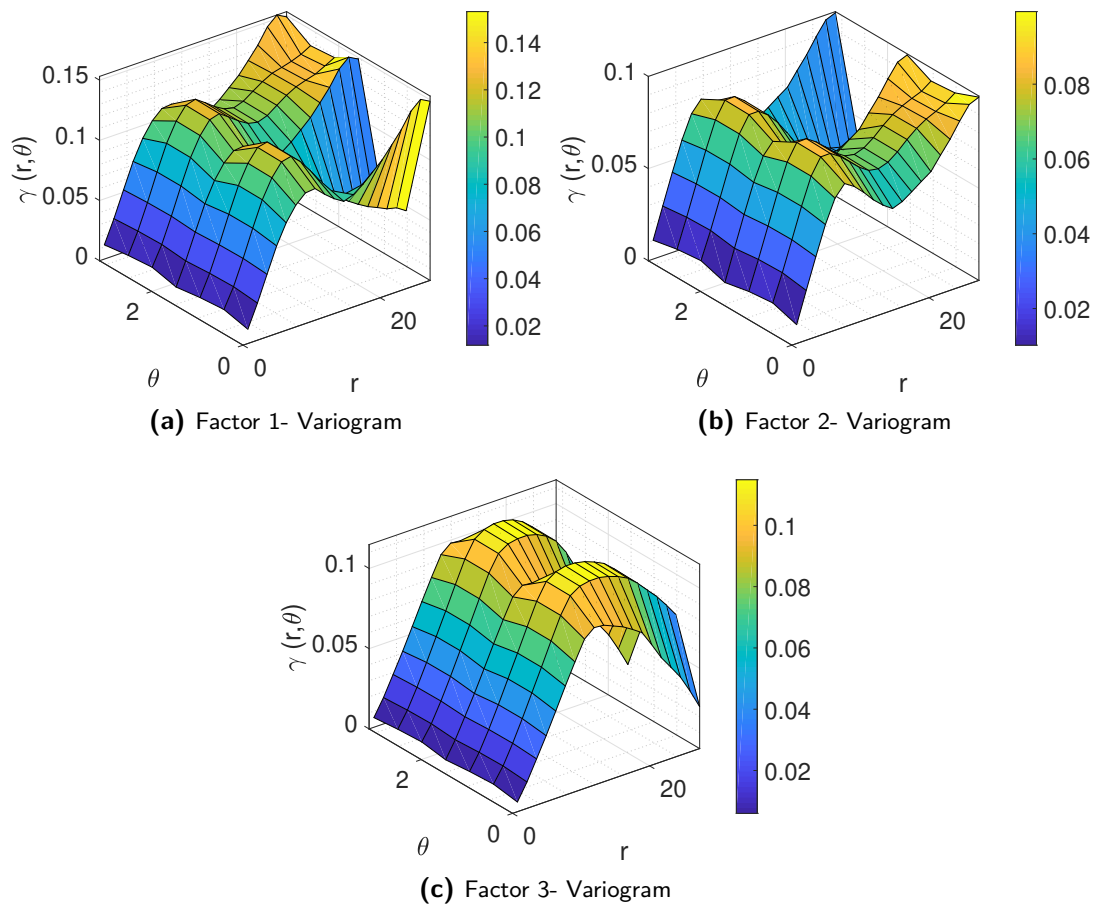
**Figure 5-8:** The variogram of each individual variable of the ArtLenDS dataset. Notice how each factors variogram changes both with length and angle of the shift parameter, that is all the factors are anisotropic.
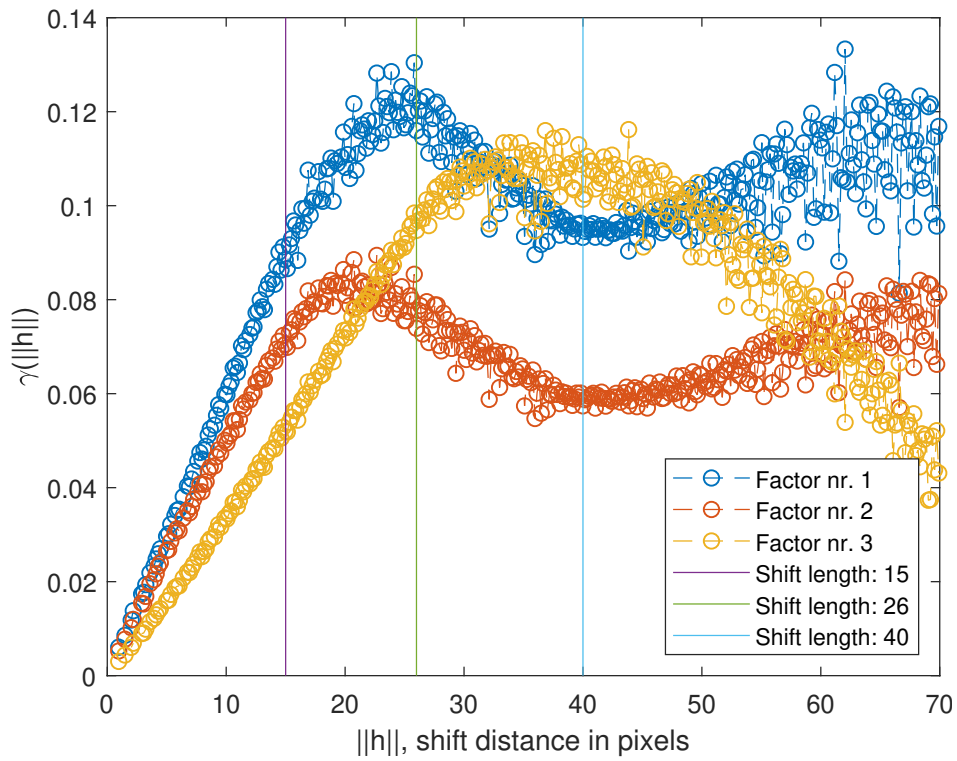
**Figure 5-9:** Variogram of the factors of the ArtLenDS dataset, with vertical lines indicating the shifts used in the Length Dependence (LenDep) experiment. These shifts are chosen since there is a change in magnitude of the variograms between the shifts. For example, in the first shift at $\|\mathbf{h}\| = 15$, variable 1 has a higher value compared to the other two. At the shift $\|\mathbf{h}\| = 40$, variable 3 has a higher value compared to the other two.



**(a)** Score 1          **(b)** Loading 1          **(c)** Score 2          **(d)** Loading 2



**(e)** Score 3          **(f)** Loading 3
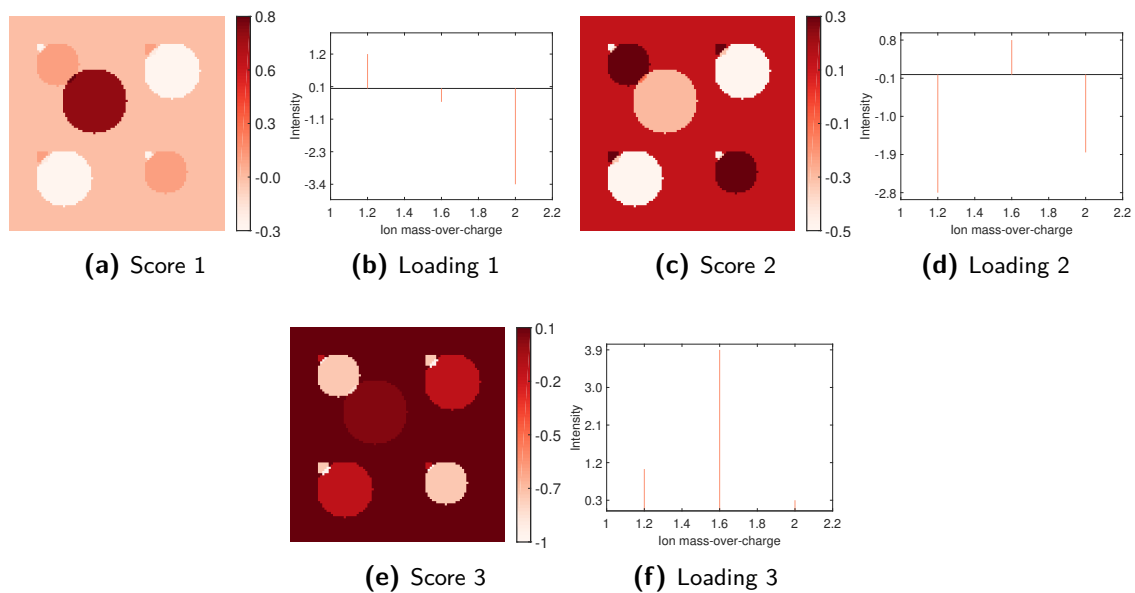
**Figure 5-10:** The resulting factors and loadings when MAF is applied to the ArtLenDS dataset with a shift parameter $\mathbf{h} = [15, 0]$.

**(a)** Score 1    **(b)** Loading 1    **(c)** Score 2    **(d)** Loading 2



**(e)** Score 3    **(f)** Loading 3
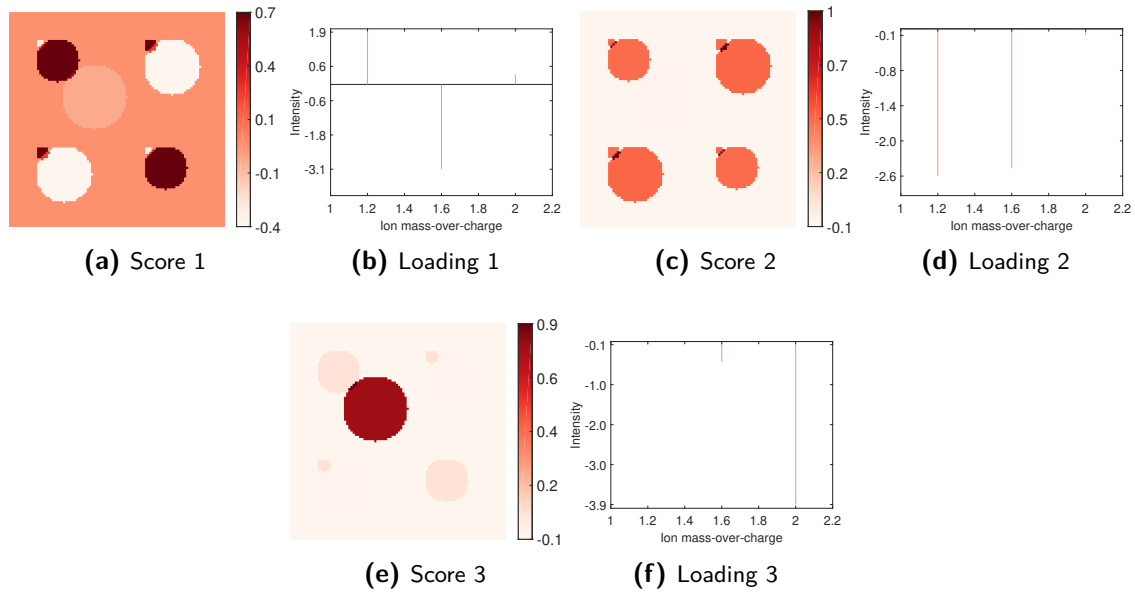
**Figure 5-11:** The resulting factors and loadings when MAF is applied to the ArtLenDS dataset with a shift parameter $\mathbf{h} = [26, 0]$.



**(a)** Score 1    **(b)** Loading 1    **(c)** Score 2    **(d)** Loading 2



**(e)** Score 3    **(f)** Loading 3
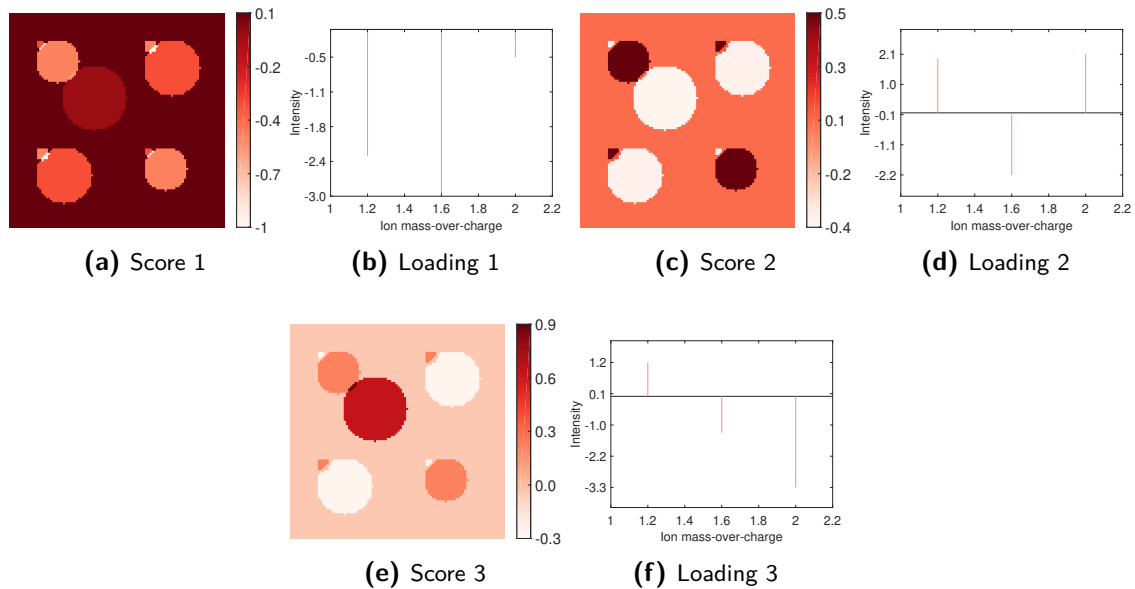
**Figure 5-12:** The resulting factors and loadings when MAF is applied to the ArtLenDS dataset with a shift parameter $\mathbf{h} = [40, 0]$.

## 5-2-2   EMAF

In order to analyze and assess the alterations made to the MAF algorithm, EMAF is also applied to the ArtLenDS dataset. Although the EMAF algorithm does not specifically offer for the radius of the Region of Interest (ROI) to be set, the experiment was carried out using the same distances as in the LenDep MAF experiment. This is done to make the comparison more fair. The shift distances used, along with the variogram of the dataset can be seen in Figure 5-9.

The resulting EMAF factors for a shift distance of 15 can be seen in Figure 5-13, the factors for a shift distance of 26 are shown in Figure 5-14, and the factor for a shift distance of 40 are shown in Figure 5-15. It is worth noting that, although the factors do change between shift distances, there are some similarities. Comparing the order of the factors to the order each variable appears on the variogram, the changes in the order of the EMAF factors seem relatively reluctant compared to the variogram.



**(a)** Factor 1- Score    **(b)** Factor 1- Loading    **(c)** Factor 2- Score    **(d)** Factor 2- Loading



**(e)** Factor 3- Score    **(f)** Factor 3- Loading
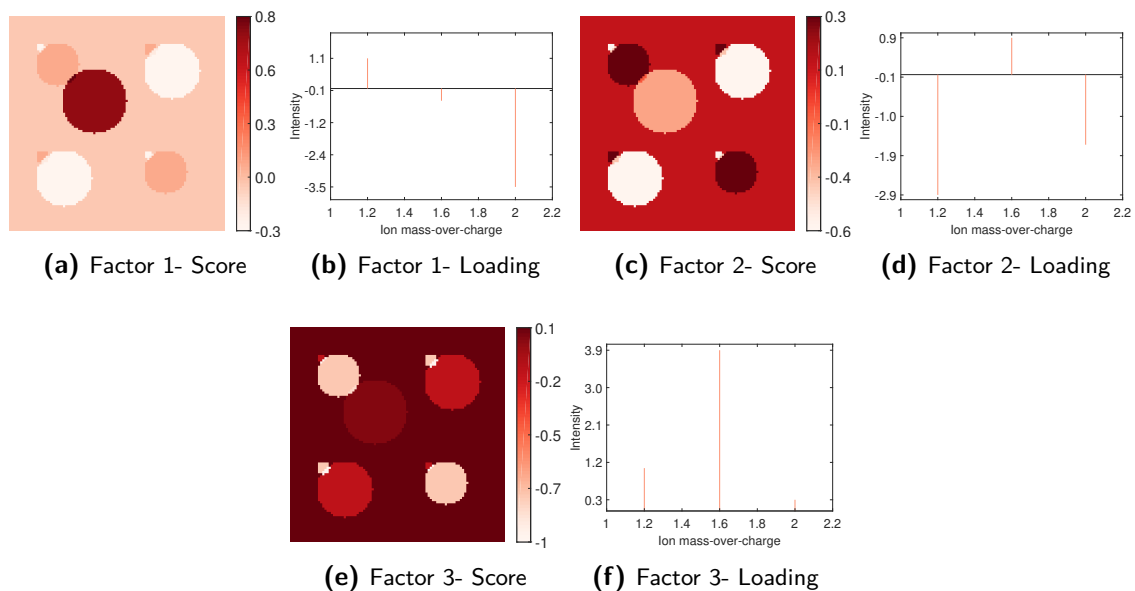
**Figure 5-13:** Resulting factors when EMAF is applied to the ArtLenDS dataset, using a ROI with radius 15.

**(a)** Factor 1- Score  **(b)** Factor 1- Loading  **(c)** Factor 2- Score  **(d)** Factor 2- Loading



**(e)** Factor 3- Score  **(f)** Factor 3- Loading
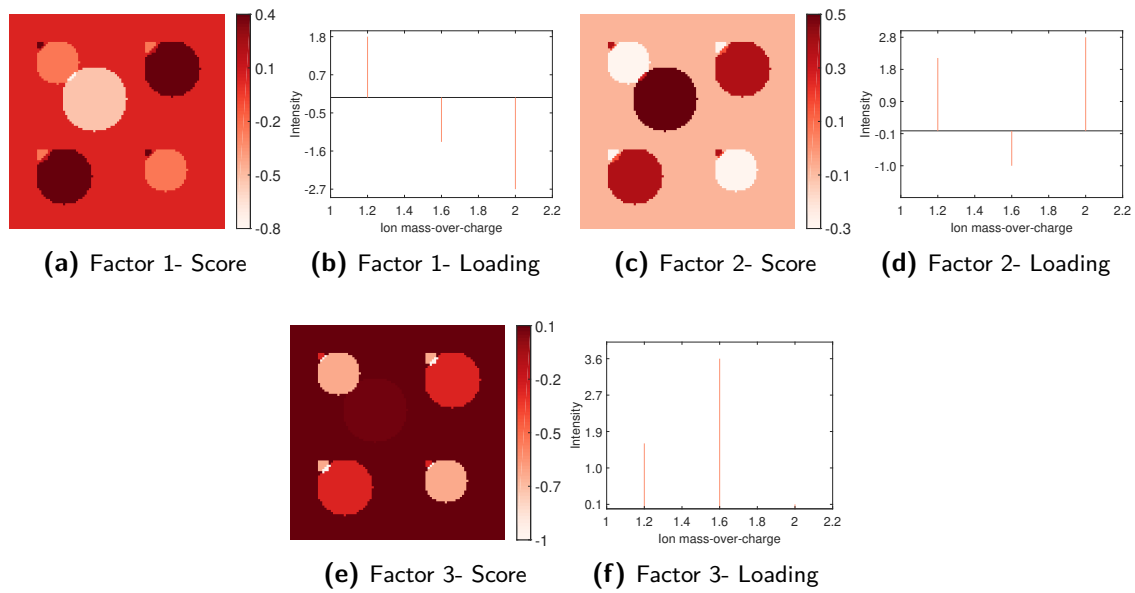
**Figure 5-14:** Resulting factors when EMAF is applied to the ArtLenDS dataset, using a ROI with radius 26.



**(a)** Factor 1- Score  **(b)** Factor 1- Loading  **(c)** Factor 2- Score  **(d)** Factor 2- Loading



**(e)** Factor 3- Score  **(f)** Factor 3- Loading
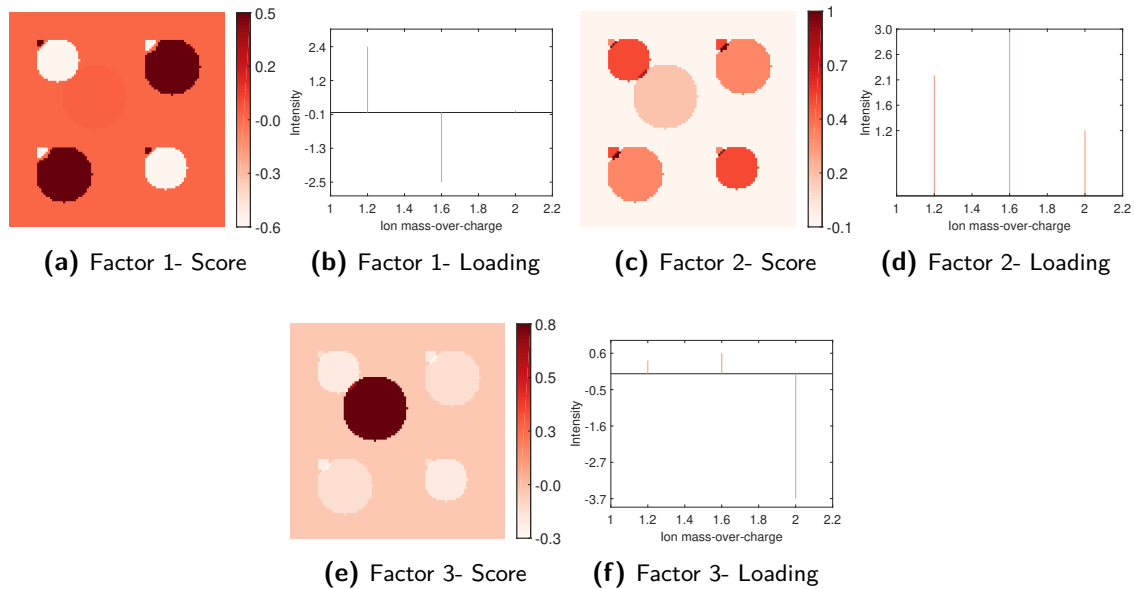
**Figure 5-15:** Resulting factors when EMAF is applied to the ArtLenDS dataset, using a ROI with radius 40.

## 5-3   Method evaluation

As discussed in Chapter 3, there are a few possible variations of EMAF that need to compared. These methods are the *rim-* and *disc implementations* of EMAF and the three different methods of picking the radius of the ROI, which are described in Table 3-1. The goal of this section is to compare and evaluate the performance of these different methods.

All combinations of the EMAF configuration are tried out in this chapter. That is, for the *rim implementation* of EMAF the Maximum Autocorrelation (Max AC), Minimum Autocorrelation (Min AC), and Unit Circle (UC) radius estimation methods are used and for the *disc implementation* of EMAF the Max AC, Min AC, and UC radius estimation methods are also used. It is noted that there is no difference between the UC radius estimation method for the *rim-* and *disc implementation* of EMAF, so these methods will simply be referred to as EMAF UC. In order to evaluate the performance of EMAF the resulting factors are compared to the results of Principal Component Analysis (PCA) and MAF.

Comparing factorization methods can be hard, since each method has different goals and objectives. Therefore, finding a single performance metric and comparing the methods using only that would be unfair. For example, if variance accounted for is used to compare the results, PCA would turn out to be the best choice of algorithm. With that in mind, this experiment should illustrate what might be gained from applying EMAF instead of any comparable factorization method. The goal is also to evaluate if there are any performance gains in particular from using the *disc implementation* of EMAF compared to the *rim implementation* of EMAF for example. The *disc implementation* is more computationally expensive, since it takes more variogram matrices into account. The evaluation will both be based on performance metrics as well as visual inspection of the resulting factors.

First, the different algorithms are compared using three performance metrics, namely spatial autocorrelation, Spearman correlation, and variance accounted for. In the following subsections, the first few component scores and loadings of each algorithm are shown and analyzed briefly.

The EMAF algorithm estimated the shift distance with the maximum spatial autocorrelation to be at a distance of 1, and the minimum spatial autocorrelation to be at a distance of 6. The maximum spatial autocorrelation distance does not come as a surpise, as spatial autocorrelation generally decreases with distance. The distance at which the minimum spatial autocorrelation appears is suprising however, one would have expected that to decrease further with longer distance.

Figure 5-16 shows the estimated autocorrelation for the first 50 factors, across a range of shift distances for each factorization algorithm. What we notice is that all algorithms show a similar structure in the spatial autocorrelation. It seems that MAF and EMAF show a slightly higher concentration of high spatial autocorrelation for the first few factors. Notice that EMAF Min AC appears to show slightly worse performance than the other methods it is being compared to.

Figure 5-17 shows the Spearman correlation of the spatial autocorrelation of each factor across a distance for each algorithm. The Spearman correlation is a measure of whether the spatial autocorrelation of the factors decreases. So a Spearman correlation of 1 means that the spatial autocorrelation is decreasing across the factors, i.e. if factor 1 has a spatial autocorrelation

**(a)** PCA    **(b)** MAF    **(c)** EMAF UC

**(d)** EMAF Max AC Disc    **(e)** EMAF Min AC Disc    **(f)** EMAF Max AC Rim

**(g)** EMAF Min AC Rim
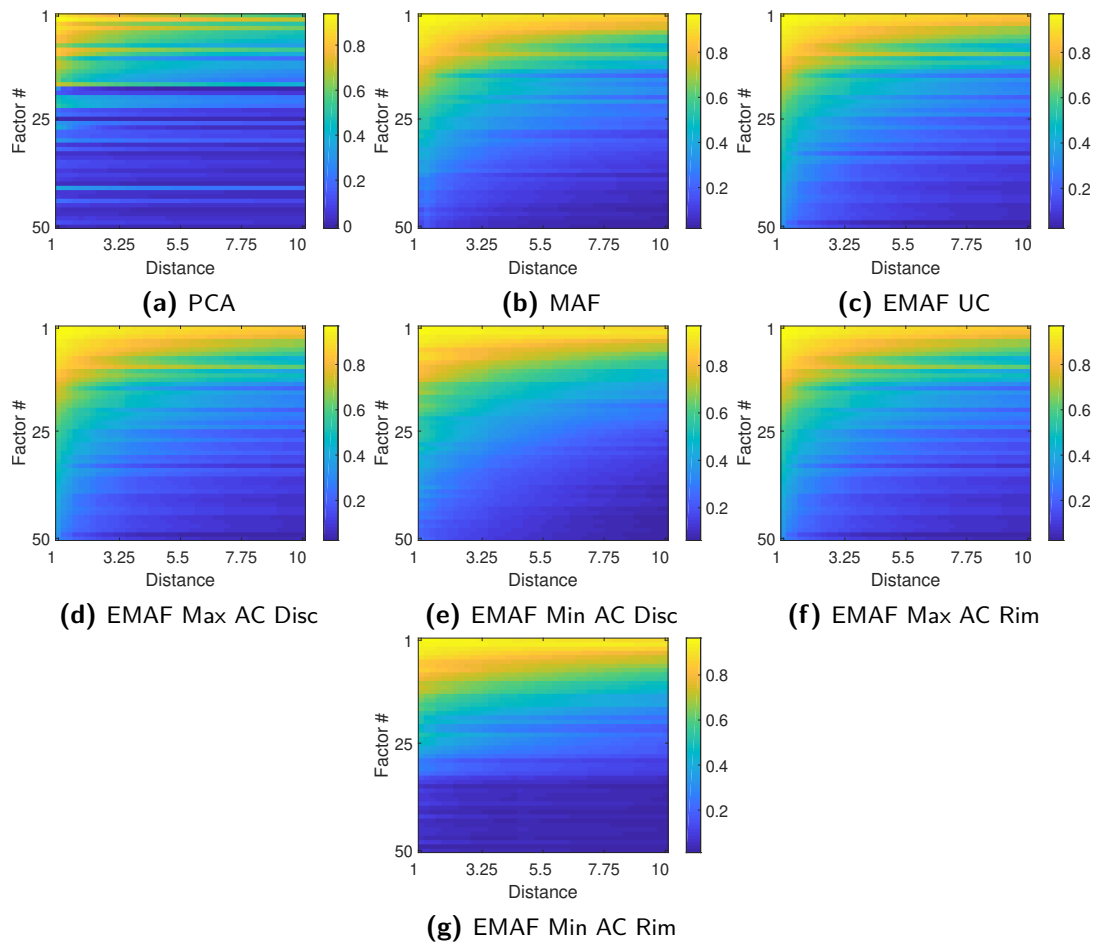
**Figure 5-16:** Comparison of estimated spatial autocorrelation.

of $I_1$, then factor two has a spatial autocorrelation of $I_2 \leq I_1$ and so on. If factors with high spatial autocorrelation appear in higher number factors that will show in the Spearman correlation. The higher the Spearman correlation, the better, since factors with high spatial autocorrelation should appear in the first factors, i.e. the low number factors.

Looking at Figure 5-17, notice that MAF already has about 67% higher Spearman correlation at shorter shift distances and about 27% higher Spearman correlation at longer shift distances. This confirms that MAF is actually returning factors with decreasing spatial autocorrelation and behaving as intended. Comparing MAF to EMAF, notice that the Spearman correlation is very similar for both algorithms, apart from the Min AC radius determination method. Another thing to note is the jump in Spearman correlation in the case of MAF at short distance, which is not present in most of the EMAF algorithms.

Figure 5-18 shows the variance accounted for, for the different algorithms. That is, the factors are ordered according to how much of the total variance of the dataset they explain, showing how much each factor contributes to that variance. PCA seems to have far better performance in this case than the other two algorithms. First of all, neither the MAF nor EMAF factors are in any sort of order, contrary to the PCA factors. Secondly, the variance accounted for by either MAF or EMAF only reaches about 10% with 10 factors combined,

**(a)** EMAF unit disc     **(b)** EMAF Max AC Disc     **(c)** EMAF Min AC Disc

**(d)** EMAF Max AC Rim     **(e)** EMAF Min AC Rim
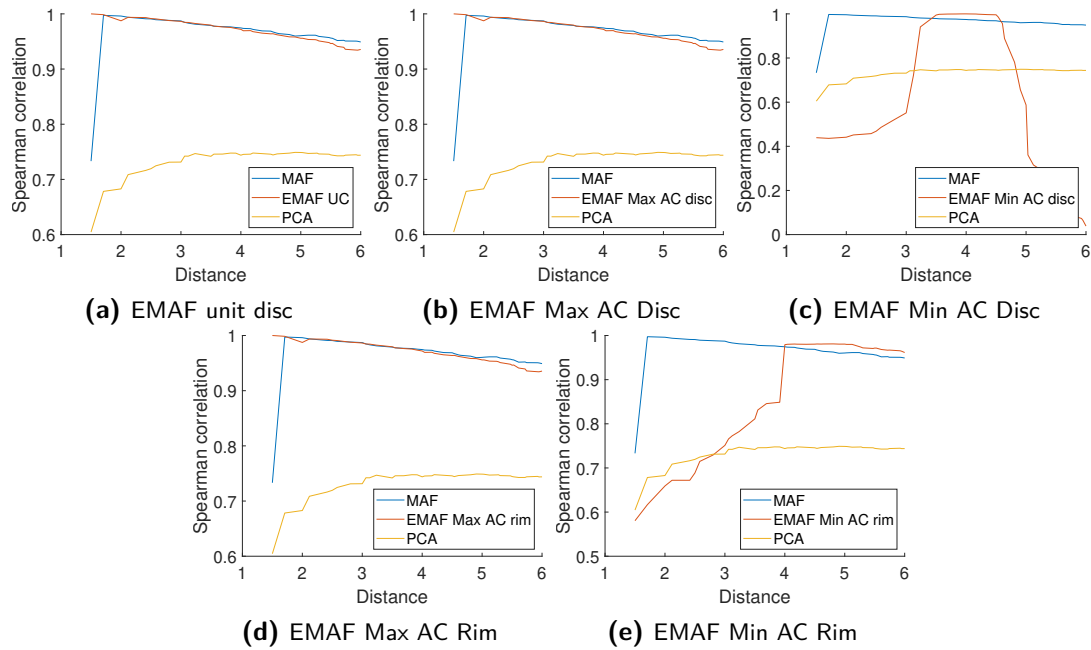
**Figure 5-17:** Comparison of Spearman auto-correlation.

compared to around 70% for PCA. This is to be expected, since the goal of PCA is to maximize the variance accounted for. It is questionable, however, if variance accounted for is a good metric to be looking at when using factorization methods for dimensionality reduction purposes. Noisy measurements will often exhibit high variance, so by including these noisy measurements in the reduced dimensional representation of the data, the variance accounted for will increase.

**(a)** PCA        **(b)** MAF        **(c)** EMAF unit Disc

**(d)** EMAF Max AC Disc     **(e)** EMAF Min AC Disc     **(f)** EMAF Max AC Rim

**(g)** EMAF Min AC Rim
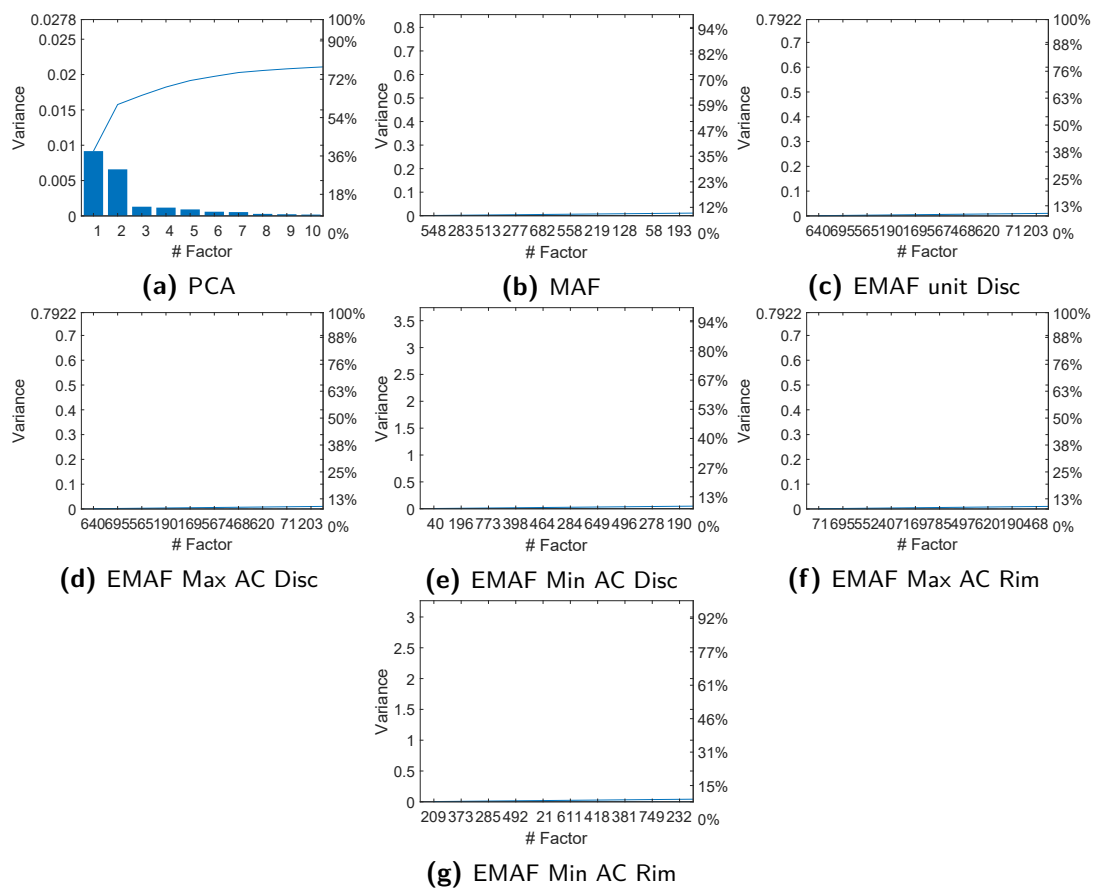
**Figure 5-18:** Comparison of variance accounted for.

## 5-3-1   PCA

Figure 5-19 shows the first 10 PCA factors and Figure 5-20 shows factors 11 to 20. The scores look clear and mostly free from noise, although some single pixel artifacts are already visible in the first score. At around score 8, noisy artifacts start to appear more dominantly. From score 8 to around 15 the noise mostly consists of speckles and after score 16, there is barely any resemblence with the original rat brain anymore. The last few scores are highly contaminated with single pixel artifacts and striped noise patterns.

The high contamination of noise seen in the PCA scores is explained by the fact that PCA generates and orders the factors according to variance. The noise patterns are likely high variance noise patterns, since they are promoted by the PCA algorithm. Score 18 is an example of this behaviour.

Examining the loadings, the first thing to notice is that the loadings are negative, which is a problem all the methods being compared in this experiment have. Since the original measurements are ion counts, all the loadings would preferably be non-negative. There are some distinct peaks and bundles of peaks in the first few loadings. As higher number loadings are examined, they seem to become more and more contaminated with noise, containing mostly single peaks, for example in loadings 16, 18, and 19. These single peaks are an indication of a high concentration of a certain mass-over-charge, i.e. a strong signal.

**(a)** Factor 1

**(b)** Loading 1

**(c)** Factor 2

**(d)** Loading 2

**(e)** Factor 3

**(f)** Loading 3

**(g)** Factor 4

**(h)** Loading 4

**(i)** Factor 5

**(j)** Loading 5

**(k)** Factor 6

**(l)** Loading 6

**(m)** Factor 7

**(n)** Loading 7

**(o)** Factor 8

**(p)** Loading 8

**(q)** Factor 9

**(r)** Loading 9

**(s)** Factor 10

**(t)** Loading 10
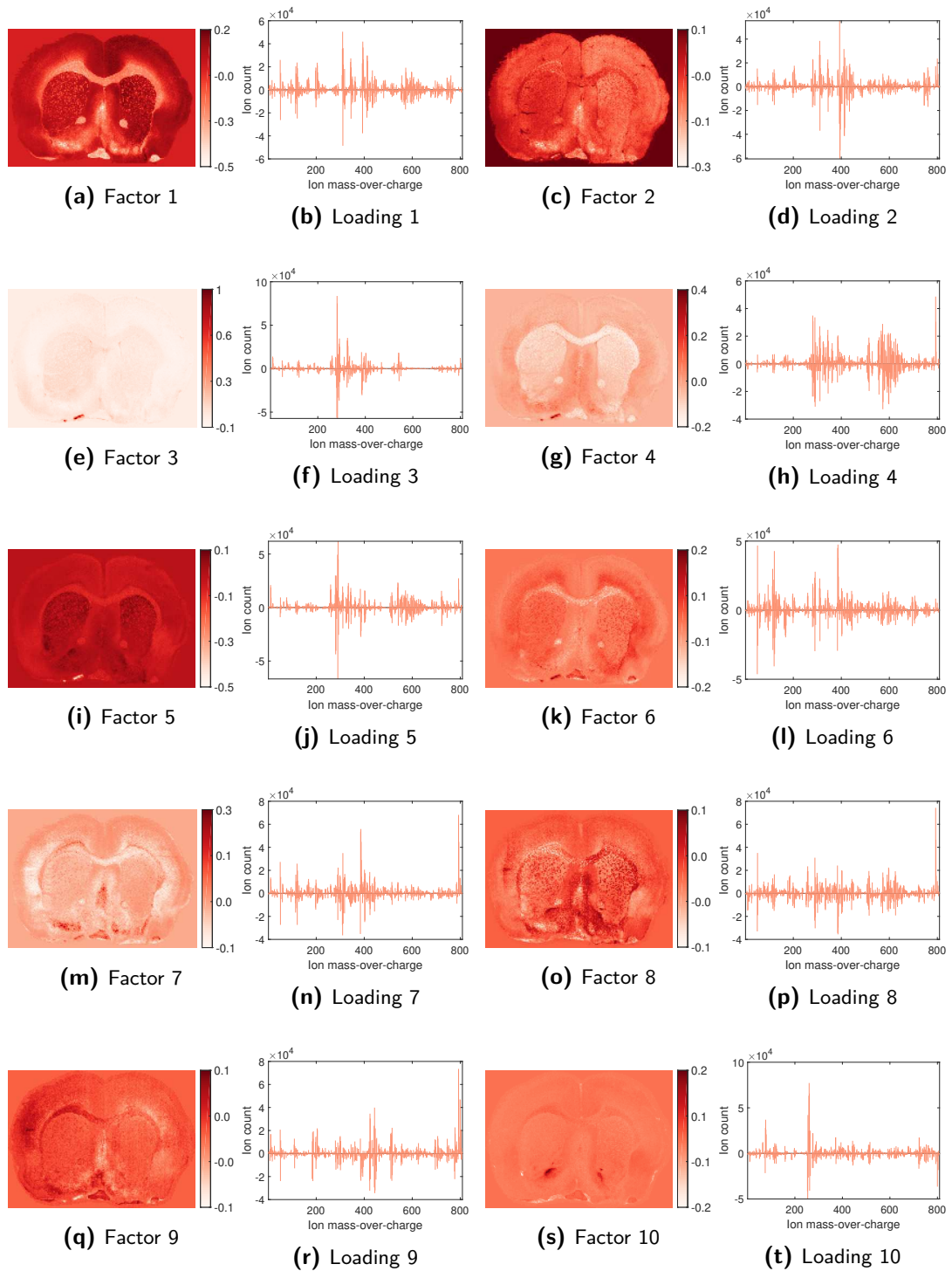
**Figure 5-19:** PCA factors 1 - 10.
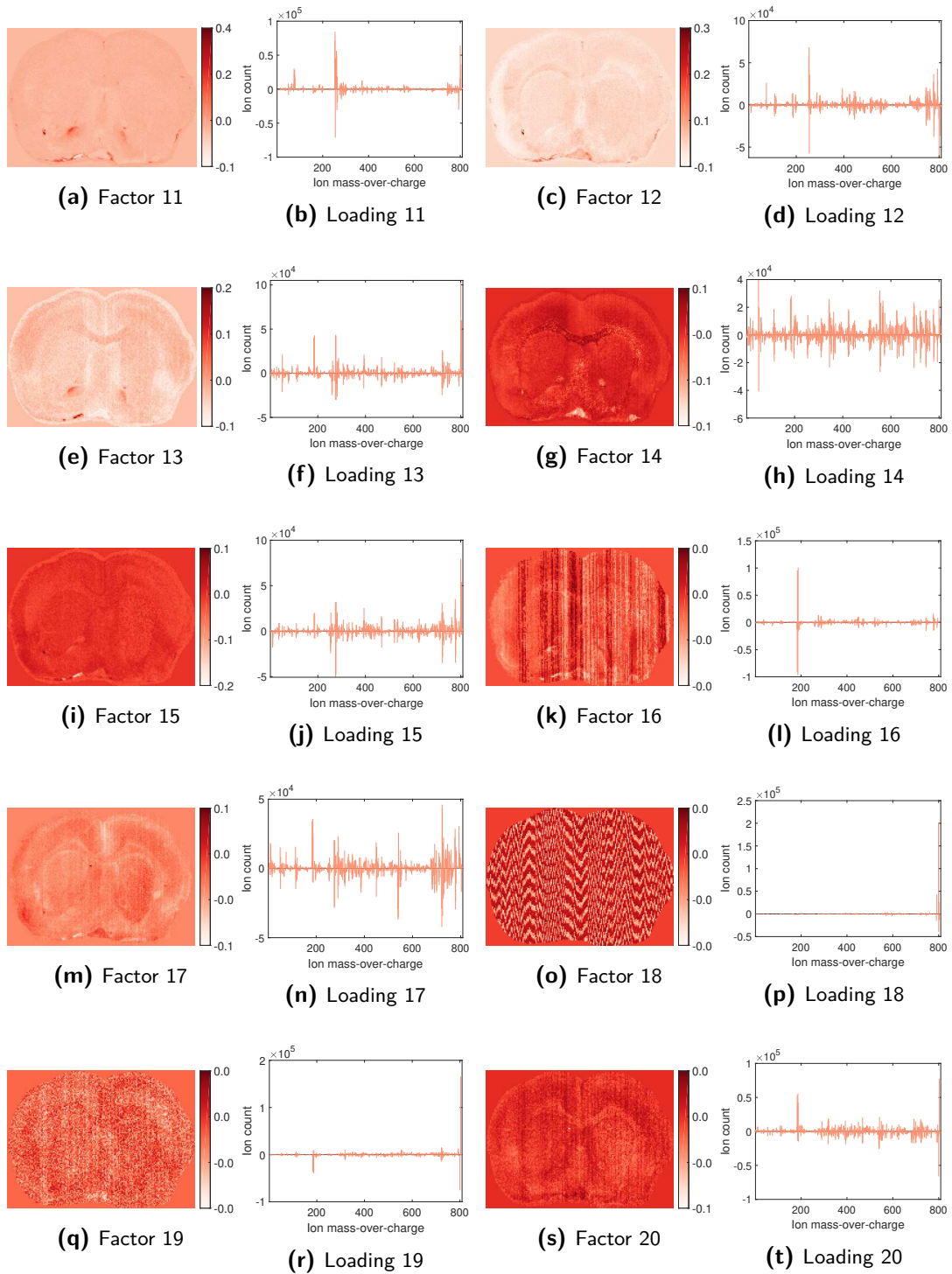
**Figure 5-20:** PCA factors 11 - 20.

## 5-3-2  MAF

Figure 5-21 shows the first 10 MAF factors and in Figure 5-22 factors 11 to 20 are shown. The first scores seem clear and free from single pixel artifacts. The single pixel artifacts do not start appearing until around factor 10. Each score seems to highlight a particular part of the sample, which was not obvious when looking at the PCA factors. Another thing to notice is that only the last score is contaminated by speckled spatial noise. The striped patterned noise seen in the PCA scores is not visible in the first 20 MAF factors. Note how blurred the MAF factors seem compared to the PCA factors. This is to be expected due to the fact that a shifted version of the dataset is used to calculate the MAF factors.

Looking at the loadings of the MAF factors, they seem comparable to the ones in the PCA case. One difference is that the values on the loadings are higher in many cases for the MAF loadings. This might be an indication of a higher signal to noise ratio but could also be a matter of scaling. This would need to be tested on a dataset where the ground truth is available. In the case of MAF, the first two loadings reach close to $15 \cdot 10^4$ while the PCA loadings are only around $4 \cdot 10^4$.
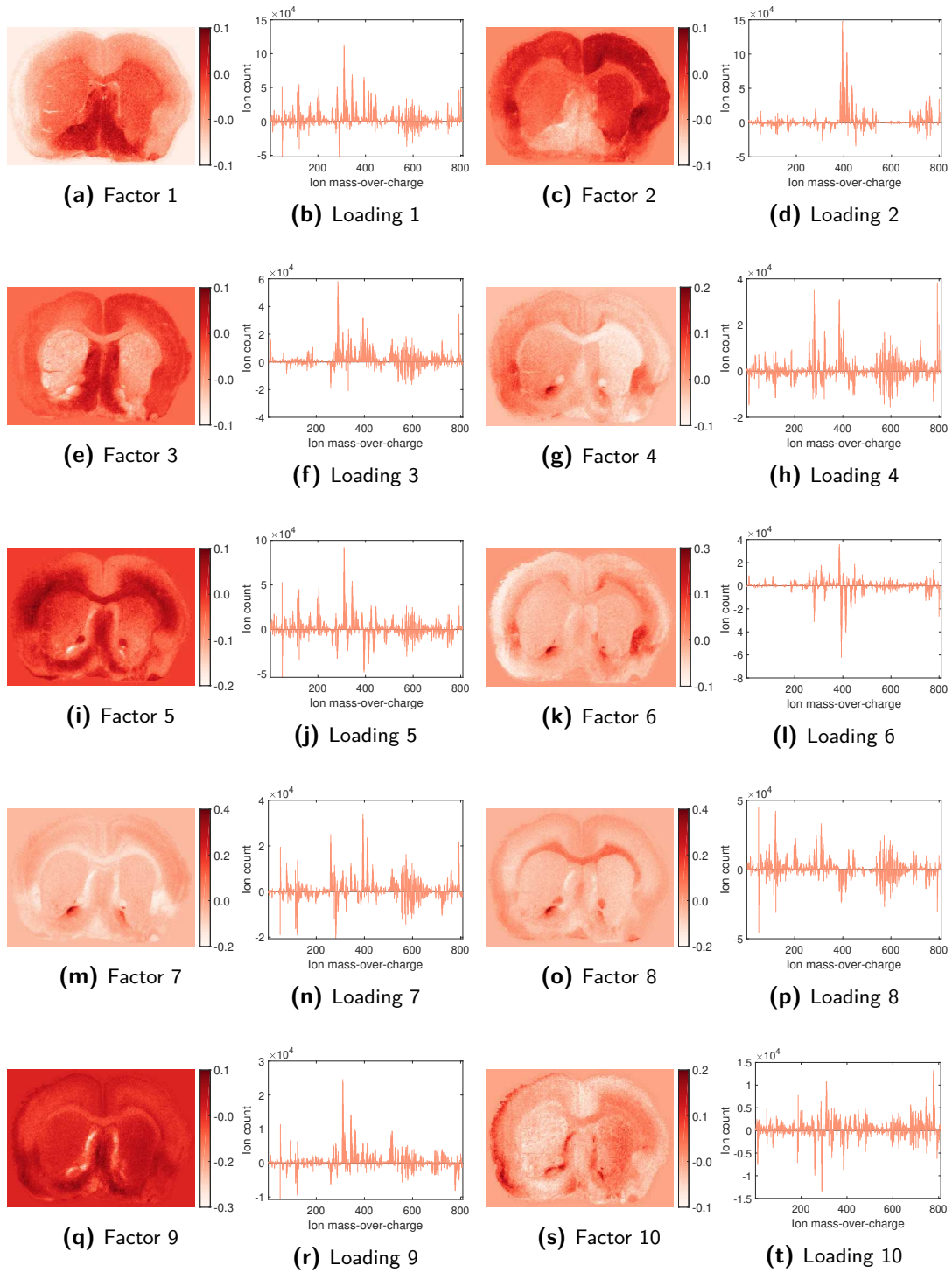
**Figure 5-21:** MAF factors 1 - 10.

(a) Factor 11

(b) Loading 11

(c) Factor 12

(d) Loading 12

(e) Factor 13

(f) Loading 13

(g) Factor 14

(h) Loading 14

(i) Factor 15

(j) Loading 15

(k) Factor 16

(l) Loading 16

(m) Factor 17

(n) Loading 17

(o) Factor 18

(p) Loading 18

(q) Factor 19

(r) Loading 19

(s) Factor 20

(t) Loading 20
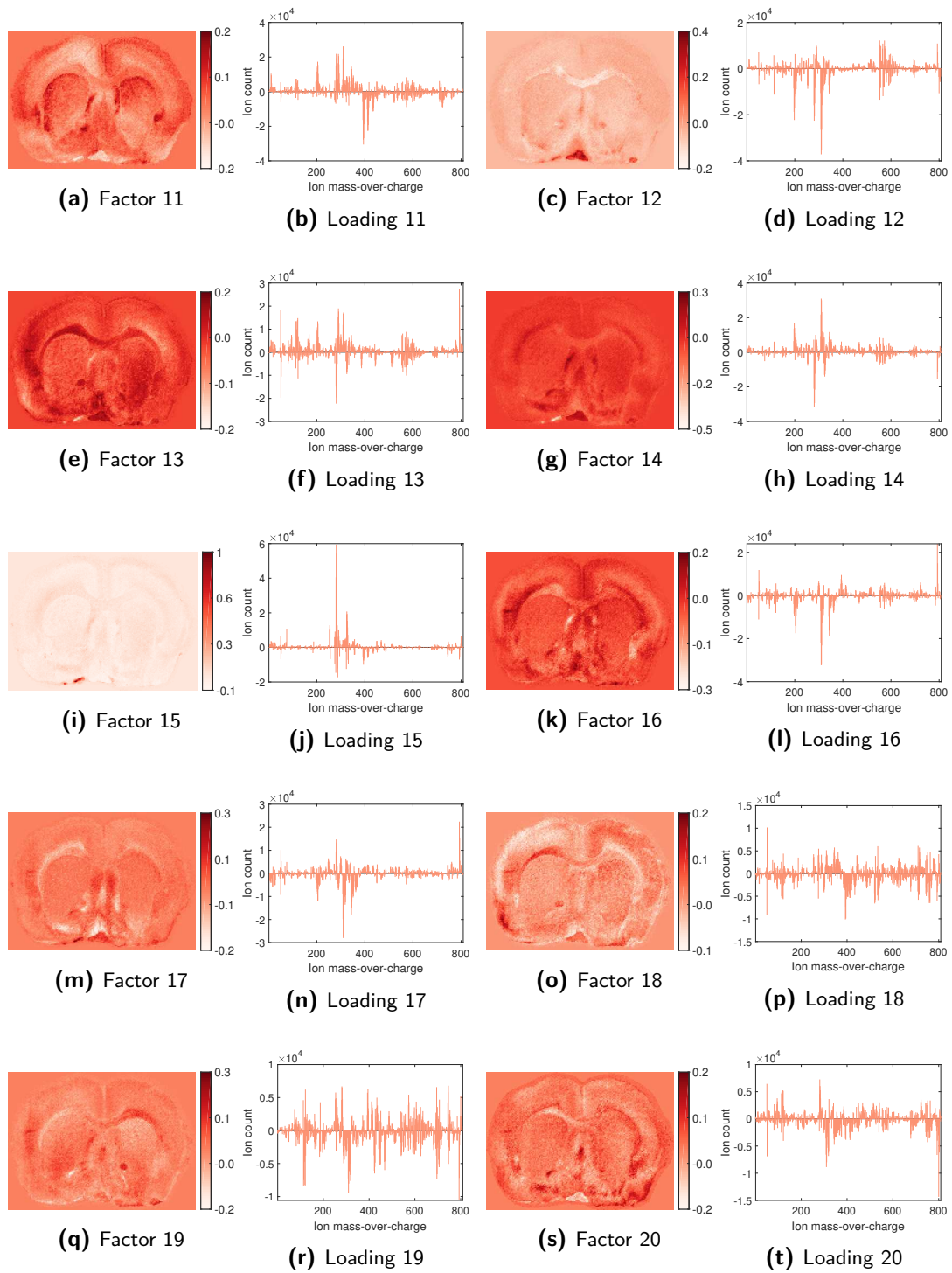
**Figure 5-22:** MAF factors 11 - 20.

### 5-3-3 EMAF

**EMAF - Unit circle**

Figure 5-23 shows the first 10 EMAF UC factors and Figure 5-24 shows factors 11 to 20. Similar to the MAF scores, the EMAF UC scores seem clear and highlight certain areas of the sample. Noisy structures in the scores do no appear until around factor number 10. Whether the speckled noise is more or less than in the MAF case is hard to state, that would have to quantified further in order to compare it. These EMAF factors are, like the MAF factors, slightly blurred compared to the PCA factors. This is again due to the shifted nature of the matrix used to calculate the factors.

The loadings show some high peaks. Loading 2 for example goes up to $20 \cdot 10^4$ and seems to contain low noise levels. This might be an indication of a higher signal to noise ratio, but could also be a matter of scaling. This would need to be tested on a dataset where the ground truth is available. However, in other aspects, the loadings seem similar to the ones seen in the MAF results.

(a) Factor 1

(b) Loading 1

(c) Factor 2

(d) Loading 2

(e) Factor 3

(f) Loading 3

(g) Factor 4

(h) Loading 4

(i) Factor 5

(j) Loading 5

(k) Factor 6

(l) Loading 6

(m) Factor 7

(n) Loading 7

(o) Factor 8

(p) Loading 8

(q) Factor 9

(r) Loading 9

(s) Factor 10

(t) Loading 10
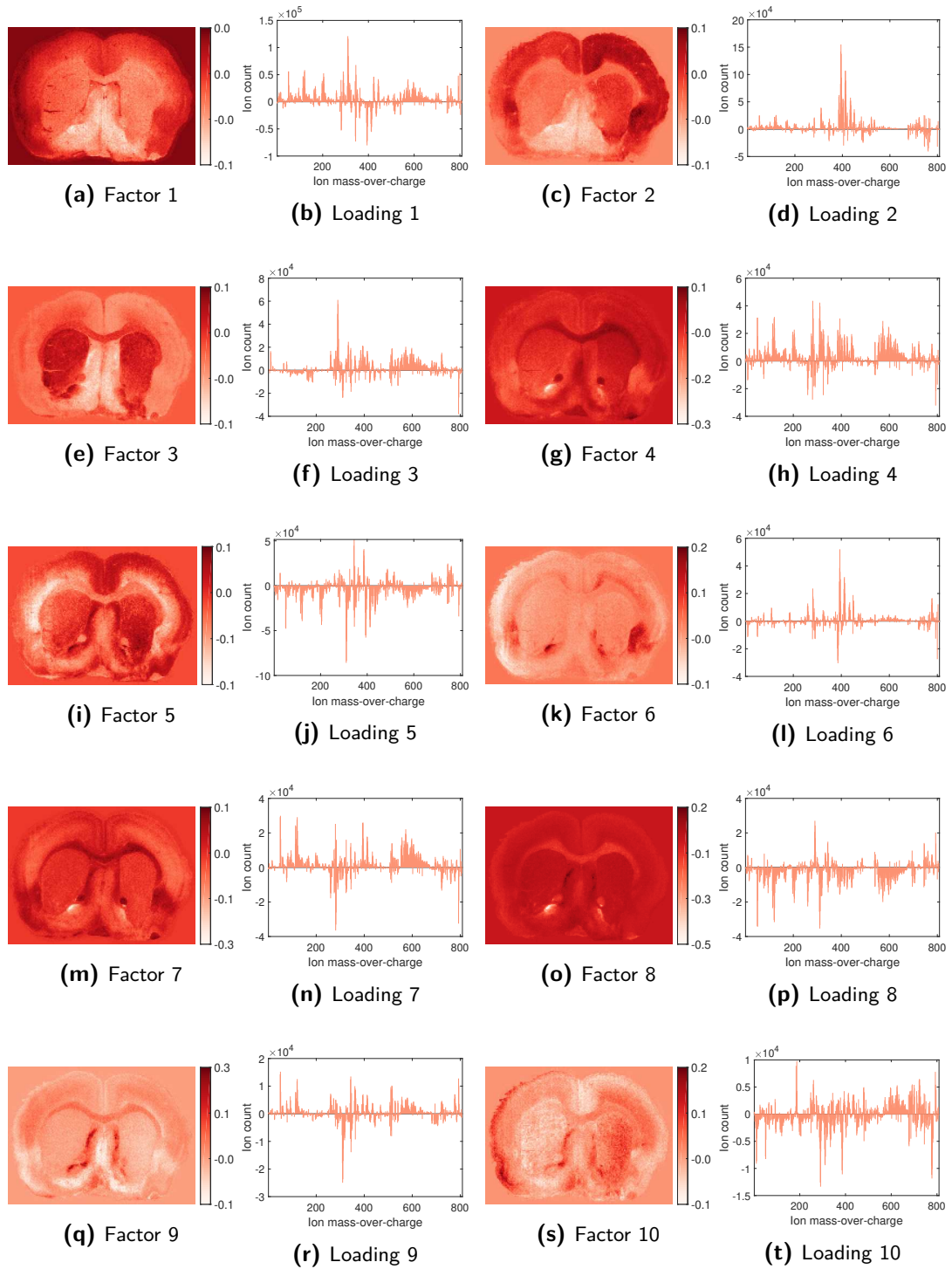
**Figure 5-23:** EMAF UC factors 1 - 10.

(a) Factor 11

(b) Loading 11

(c) Factor 12

(d) Loading 12

(e) Factor 13

(f) Loading 13

(g) Factor 14

(h) Loading 14

(i) Factor 15

(j) Loading 15

(k) Factor 16

(l) Loading 16

(m) Factor 17

(n) Loading 17

(o) Factor 18

(p) Loading 18

(q) Factor 19

(r) Loading 19

(s) Factor 20

(t) Loading 20
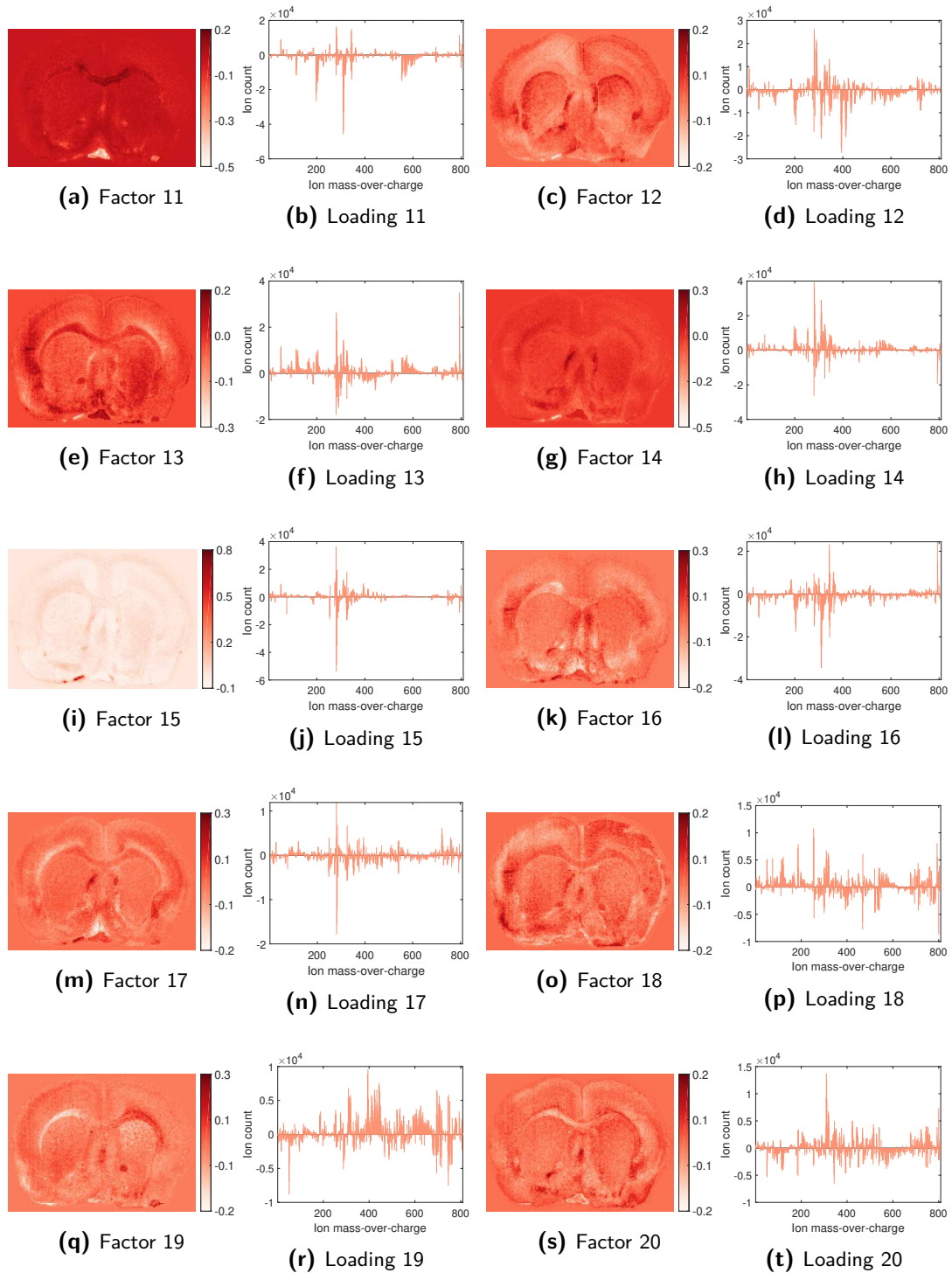
**Figure 5-24:** EMAF UC factors 11 - 20.

**EMAF - Rim**

**Maximum estimated spatial autocorrelation region of interest**
Figure 5-25 shows the first 10 EMAF *Rim* Max AC factors and Figure 5-26 shows factors 11-20. There is no difference to be seen between the EMAF UC and EMAF Max AC factors; this is because the distance used in the EMAF Max AC turned out to be 1, so in fact they are the same. This is also seen by comparing the loadings and scores of the two resulting factors. They are indentical.

**(a)** Factor 1

**(b)** Loading 1

**(c)** Factor 2

**(d)** Loading 2

**(e)** Factor 3

**(f)** Loading 3

**(g)** Factor 4

**(h)** Loading 4

**(i)** Factor 5

**(j)** Loading 5

**(k)** Factor 6

**(l)** Loading 6

**(m)** Factor 7

**(n)** Loading 7

**(o)** Factor 8

**(p)** Loading 8

**(q)** Factor 9

**(r)** Loading 9

**(s)** Factor 10

**(t)** Loading 10
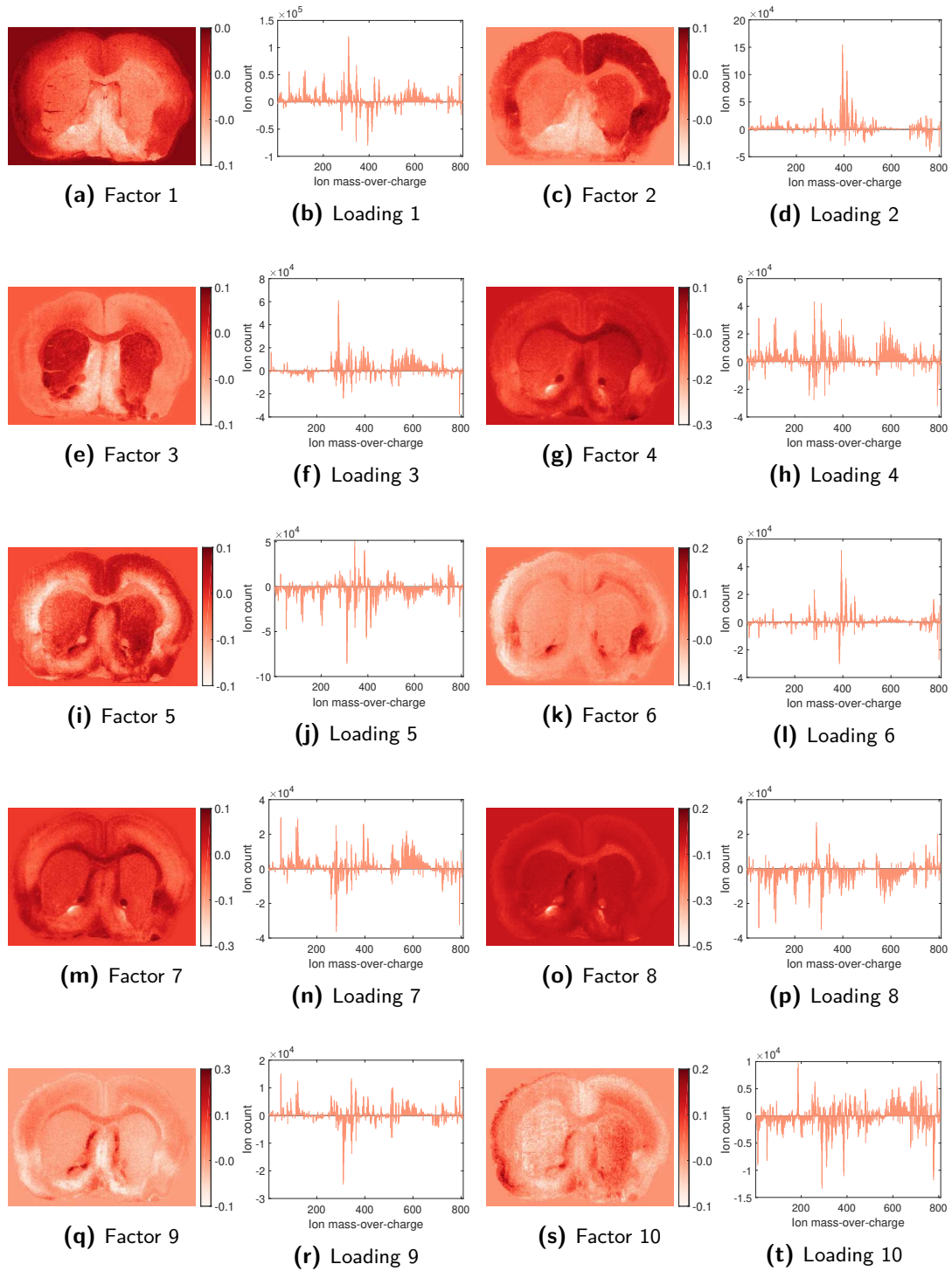
**Figure 5-25:** EMAF *Rim* Max AC rim factors 1 - 10.

(a) Factor 11

(b) Loading 11

(c) Factor 12

(d) Loading 12

(e) Factor 13

(f) Loading 13

(g) Factor 14

(h) Loading 14

(i) Factor 15

(j) Loading 15

(k) Factor 16

(l) Loading 16

(m) Factor 17

(n) Loading 17

(o) Factor 18

(p) Loading 18

(q) Factor 19

(r) Loading 19

(s) Factor 20

(t) Loading 20

**Figure 5-26:** EMAF *Rim* Max AC rim factors 11 - 20.

**Minimum estimated spatial autocorrelation region of interest**
Figure 5-27 shows the first 10 EMAF *Rim* Min AC factors and Figure 5-28 shows factors 11-20. Observing the scores of these factors, they seem to contain more noise than the other EMAF scores that have been examined so far. The striped noise patterns, observed in the scores of the PCA patterns, start appearing at around factor 15. The loadings seem similar to other EMAF loadings.

**(a)** Factor 1

**(b)** Loading 1

**(c)** Factor 2

**(d)** Loading 2

**(e)** Factor 3

**(f)** Loading 3

**(g)** Factor 4

**(h)** Loading 4

**(i)** Factor 5

**(j)** Loading 5

**(k)** Factor 6

**(l)** Loading 6

**(m)** Factor 7

**(n)** Loading 7

**(o)** Factor 8

**(p)** Loading 8

**(q)** Factor 9

**(r)** Loading 9

**(s)** Factor 10

**(t)** Loading 10
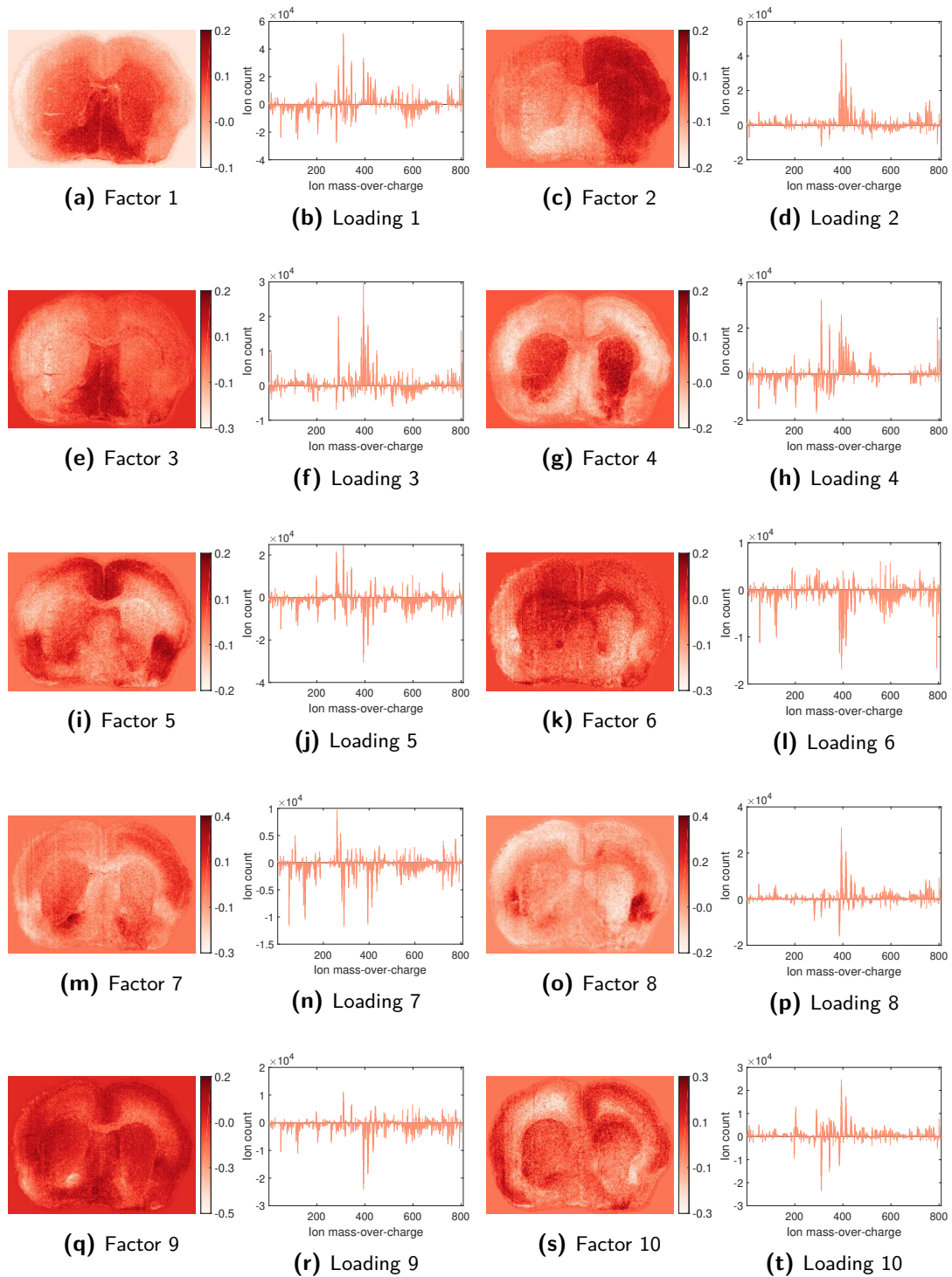
**Figure 5-27:** EMAF *Rim* Min AC rim factors 1 - 10.
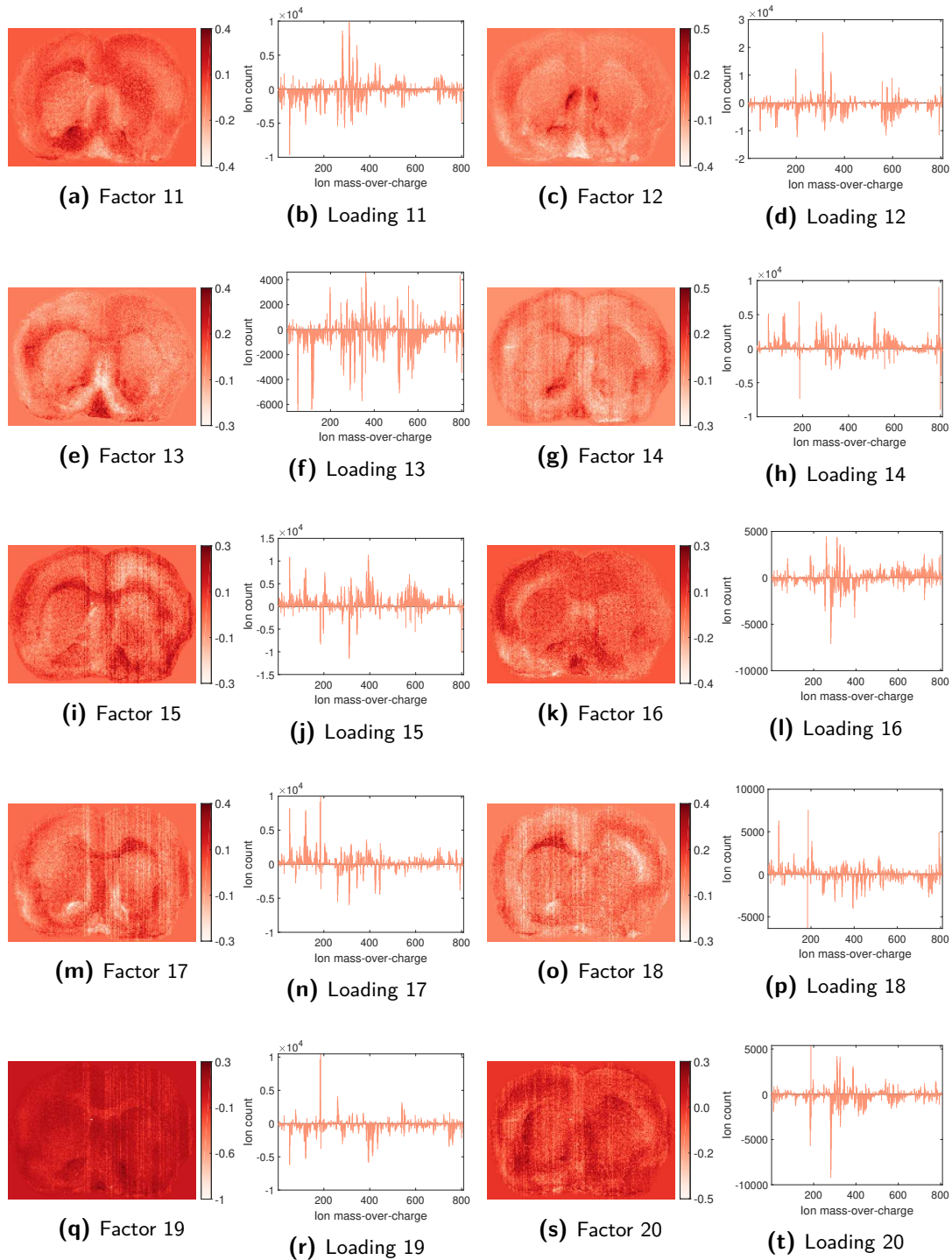
**Figure 5-28:** EMAF *Rim* Min AC rim factors 11 - 20.

**EMAF - Disc**

**Maximum estimated spatial autocorrelation region of interest**
Figure 5-29 shows the first 10 EMAF *Disc* Max AC factors and Figure 5-30 shows factors 11-20. There is no difference to be seen between the EMAF UC and EMAF Max AC factors; this is because the distance used in the EMAF Max AC turned out to be 1, so in fact they are the same. This is also seen by comparing the loadings and scores of the two resulting factors. They are indentical.

**(a)** Factor 1 **(b)** Loading 1 **(c)** Factor 2 **(d)** Loading 2

**(e)** Factor 3 **(f)** Loading 3 **(g)** Factor 4 **(h)** Loading 4

**(i)** Factor 5 **(j)** Loading 5 **(k)** Factor 6 **(l)** Loading 6

**(m)** Factor 7 **(n)** Loading 7 **(o)** Factor 8 **(p)** Loading 8

**(q)** Factor 9 **(r)** Loading 9 **(s)** Factor 10 **(t)** Loading 10
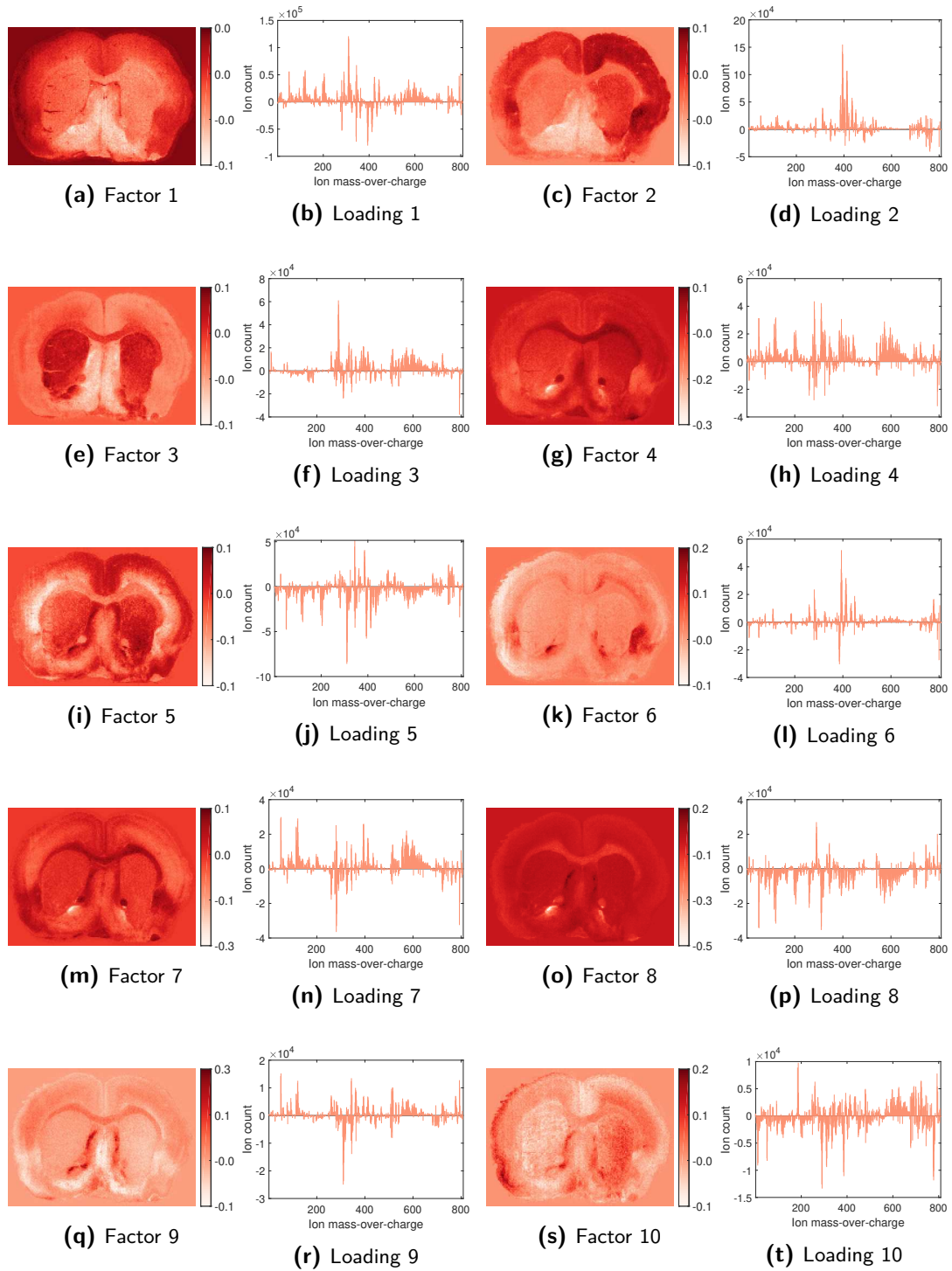
**Figure 5-29:** EMAF *Disc* Max AC disc factors 1 - 10.
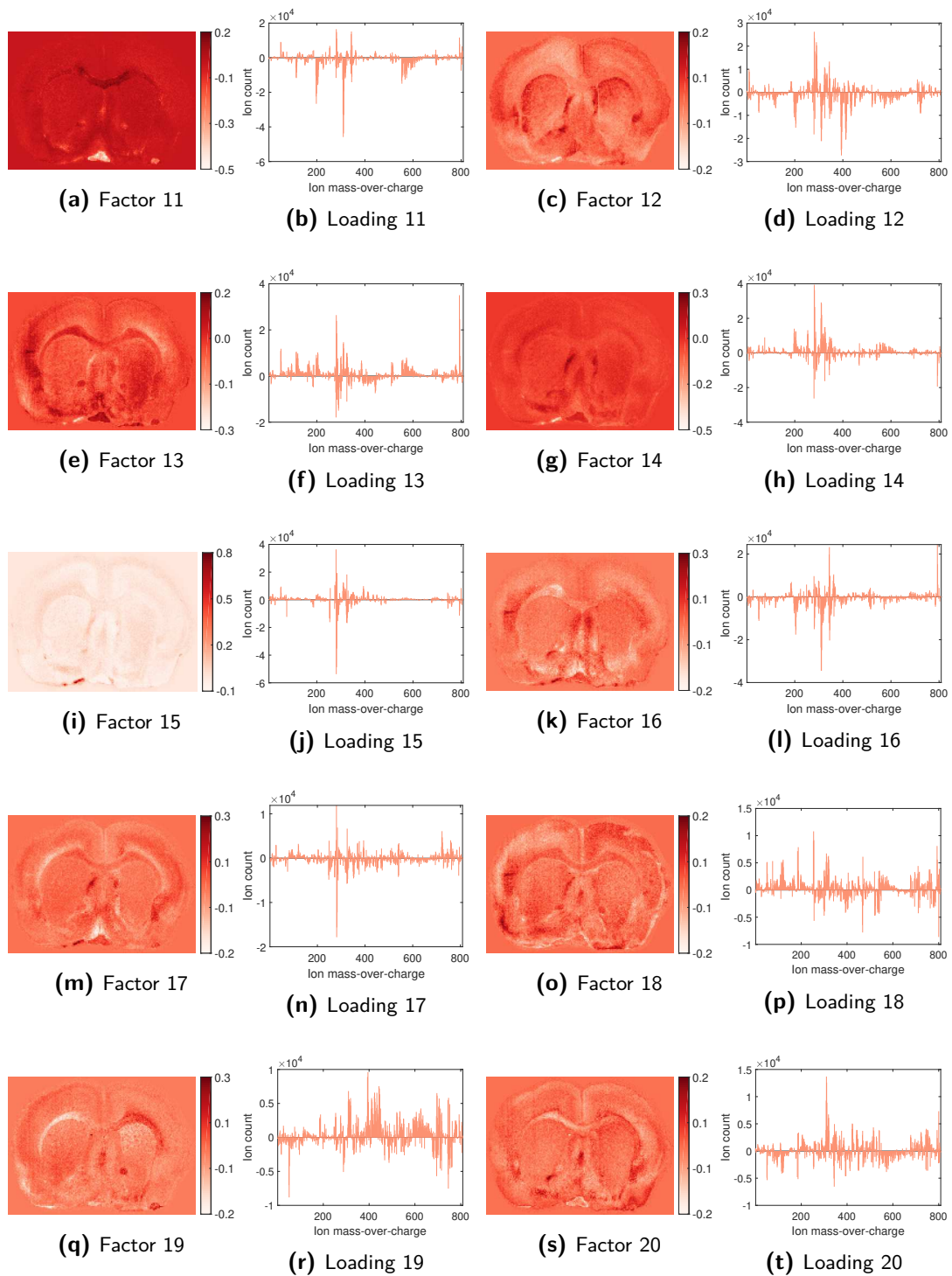
**Figure 5-30:** EMAF *Disc* Max AC disc factors 11 - 20.

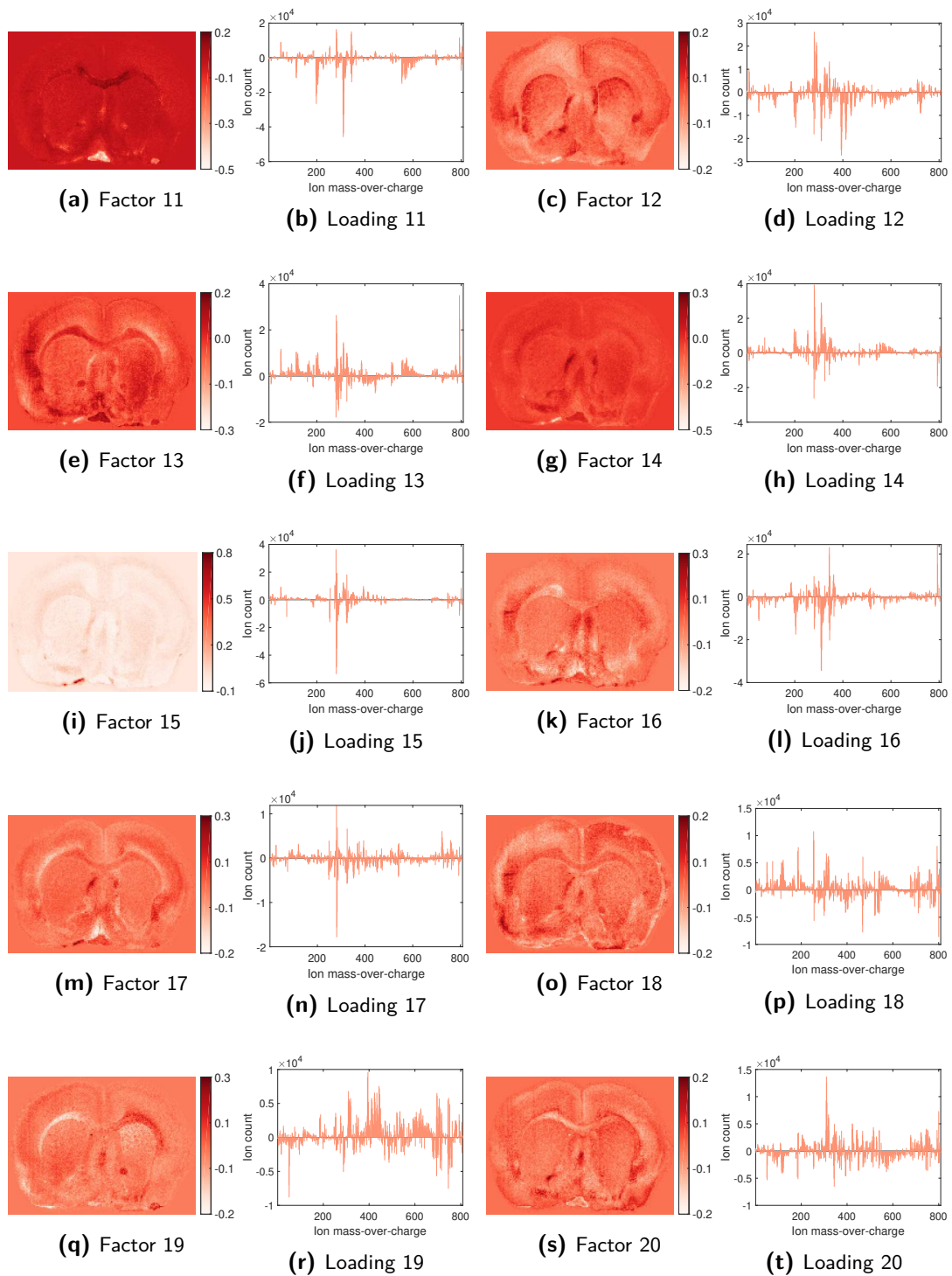**Minimum estimated spatial autocorrelation region of interest**
Figure 5-27 shows the first 10 EMAF *Disc* Max AC factors and Figure 5-28 shows factors 11-20. What is most interesting to compare here is the difference between the Min AC *rim*- and *disc implementations*. Unlike the *rim implementation*, in this case there are no longer any striped noise patterns present in the scores, albeit the scores are still relatively contaminated with noise and unclear. Looking at the loadings, they seem similar to other loadings that have been examined so far.

**(a)** Factor 1

**(b)** Loading 1

**(c)** Factor 2

**(d)** Loading 2

**(e)** Factor 3

**(f)** Loading 3

**(g)** Factor 4

**(h)** Loading 4

**(i)** Factor 5

**(j)** Loading 5

**(k)** Factor 6

**(l)** Loading 6

**(m)** Factor 7

**(n)** Loading 7

**(o)** Factor 8

**(p)** Loading 8

**(q)** Factor 9

**(r)** Loading 9

**(s)** Factor 10

**(t)** Loading 10
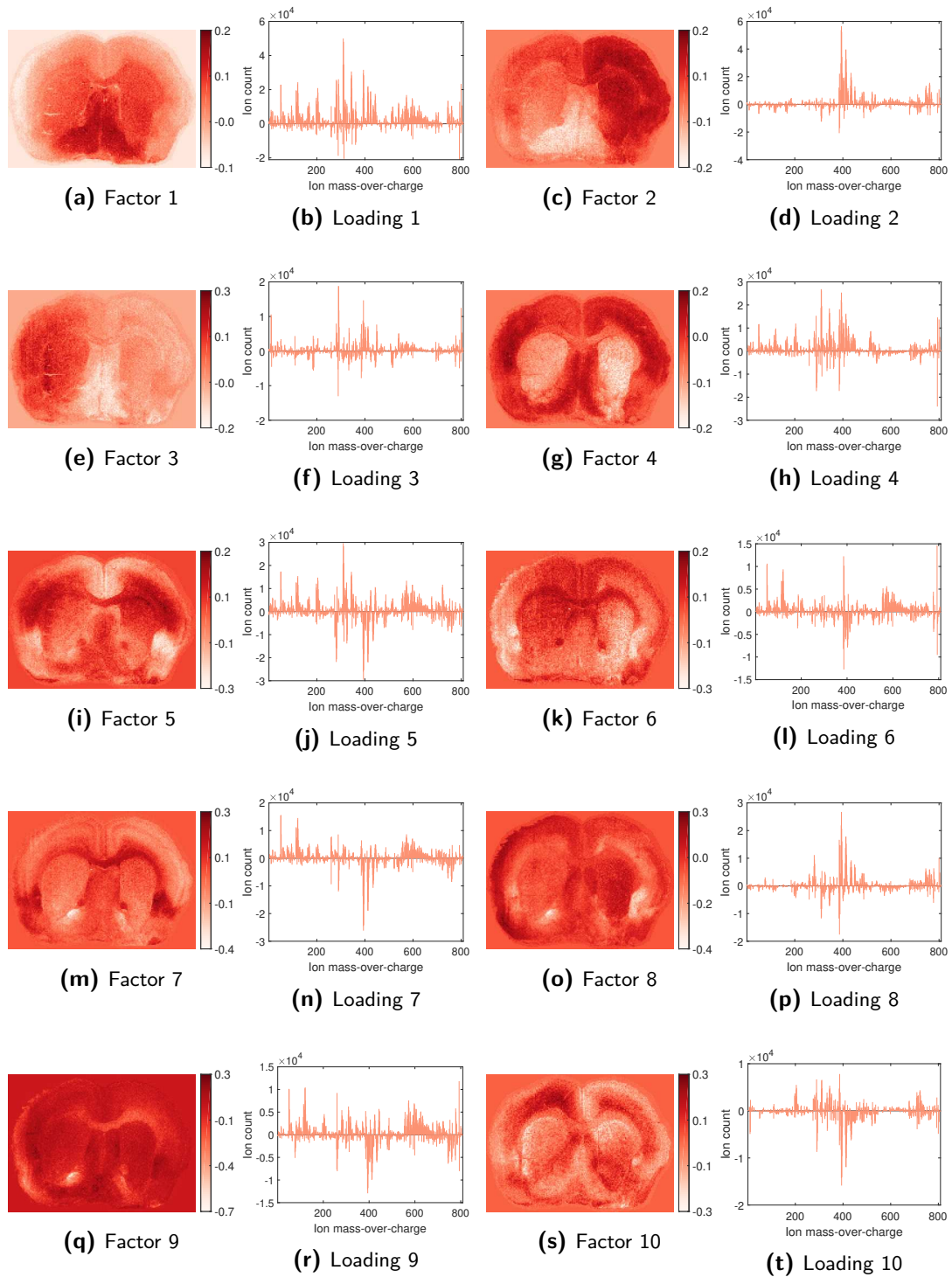
**Figure 5-31:** EMAF *Disc* Min AC disc factors 1 - 10.

(a) Factor 11

(b) Loading 11

(c) Factor 12

(d) Loading 12

(e) Factor 13

(f) Loading 13

(g) Factor 14

(h) Loading 14

(i) Factor 15

(j) Loading 15

(k) Factor 16

(l) Loading 16

(m) Factor 17

(n) Loading 17

(o) Factor 18

(p) Loading 18

(q) Factor 19

(r) Loading 19

(s) Factor 20

(t) Loading 20
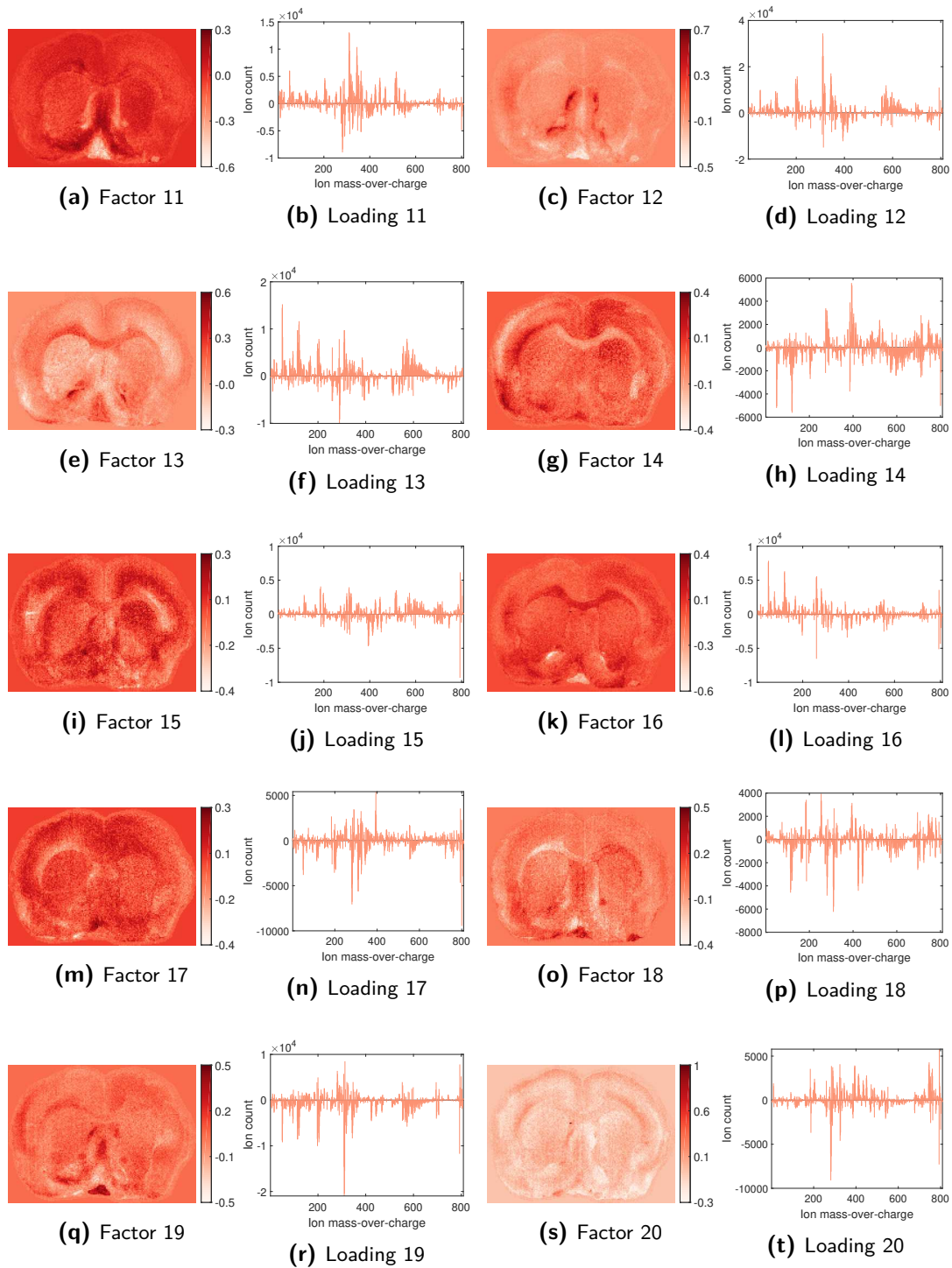
**Figure 5-32:** EMAF *Disc* Min AC disc factors 11 - 20.

# Chapter 6

# Case study: An application to Imaging Mass Spectrometry (IMS) data

To illustrate and show how useful the resulting Extended Maximum Autocorrelation Factorization (EMAF) factors potentially are for a user, a case study is carried out. To do so, the IMS-RBDS dataset is used, this is the same dataset introduced in Chapter 4-1-4. In this case study, we focus on the dataset with less stringent peak picking, which means that the dataset contains 6638 $m/z$ peaks instead of the 809 meantioned before.

The dataset is a $10\mu m$ thick coronal section of a rat brain, in which Parkinson's disease has been simulated by compromising the dopamine receptors in one of the halves of the brain. This section was then prepared for MALDI measurements and finally measured in a 15T Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer with a spatial sampling resolution of $75\mu m$ and a mass resolving power of $50000\frac{m}{FWHM}$ at $m/z$ 5000. The molecular images lie on a $m/z$-range from 1300 to 24000 with a total of about 21000 pixels. The preparation of the section is described in more detail in [4].

In a practical setting, factorization methods are often used for dimensionality reduction or noise removal. In this case study, three different factorization methods are used to carry out noise removal on the IMS data and the results are compared. The algorithms used are Principal Component Analysis (PCA), Maximum Autocorrelation Factorization (MAF) with two different shift parameters, and EMAF. In all cases the same procedure is applied:

1. The dataset is transformed using the eigenvectors that each factorization method produces.

2. A set number of components or factors of this transformed dataset is chosen.

3. The reduced set of components or factors are transformed back into the original (measurement) space.

4. The resulting, cleaned dataset is normalized to have values on the range $[0, 1]$.

In the second step, a number of components or factors is chosen. How many factors to choose remains an open question. Generally, when doing dimensionality reduction with PCA, a certain percentage of the variance accounted for is chosen and that determines the number of factors used. That way the reduced dataset explains a certain percentage, say 95% or 99%, of the original variance. A more detailed discussion on the topic of how many factors to select is found in [9]. In this case study, the first 30 components are used, which results in around 75% of the variance accounted for. This is done since IMS data is particularly noisy and the first thirty components revealed a seemingly sound result, evaluated visually. The number of factors is used for all the factorization methods. Unlike for PCA, there are no specific guidelines for how many factors to choose for MAF.

The goal of this case study is to illustrate the potential complications an analyst faces trying to remove noise from a dataset. We have already skimped on one of the decisions a user has to make, that is how to choose the number of components. In the case of MAF the user would also have to choose a shift parameter. For this reason, two different shifts are applied, where these shifts are perpendicular to each other. One is the originally proposed diagonal shift $\mathbf{h} = [1, 1]$ and the other is a shift perpendicular to that, or $\mathbf{h} = [-1, 1]$. The original shift is chosen for historical reasons and the other is chosen to illustrate what effect the shift parameter choice has on the resulting factors.

EMAF was applied with the Max AC distance estimation method and using the *disc implementation*. The estimated distance where the maximum spatial autocorrelation occurs turned out to be 1. This does not come as a surprise, as spatial autocorrelation generally decreases with distance.

Since the dataset contains 6638 $m/z$ peaks, it is infeasible to show all ion images in this thesis. Instead, a few hand picked peaks are shown along with the equivalent peak in each of the cleaned datasets. These peaks were chosen to illustrate certain points about the performance of the different factorization methods. One peak shows an example of a highly noisy measurement.

Figure 6-1 shows that the original data contains a relatively clear measurement, where the ions seem concentrated in the right hemisphere. The PCA ion image reconstructs the original image well, the image is clear and shows a difference in concentration between the two hemispheres. The concentration spots also match the original ion image. The ion images cleaned with other factorization methods all have in common that they are blurred.The EMAF ion image is less blurred that the MAF images. It is also worth noting that the MAF images are blurred in the direction of the respective shift parameters used in the algorithm.

Figure 6-2 shows an example of a highly noisy measurement. Note that PCA was not able to remove the stripe patterned noise from the data. This is likely because this noise has high variance and is included in the first 30 principal components because of that. The other methods, however, show a reconstruction of the measurement without the striped pattern, albeit not a very clear reconstruction. It is questionable where this reconstruction comes from, this might an explaination for why MAF factors do not correlate well with the original dataset compared to PCA [73].
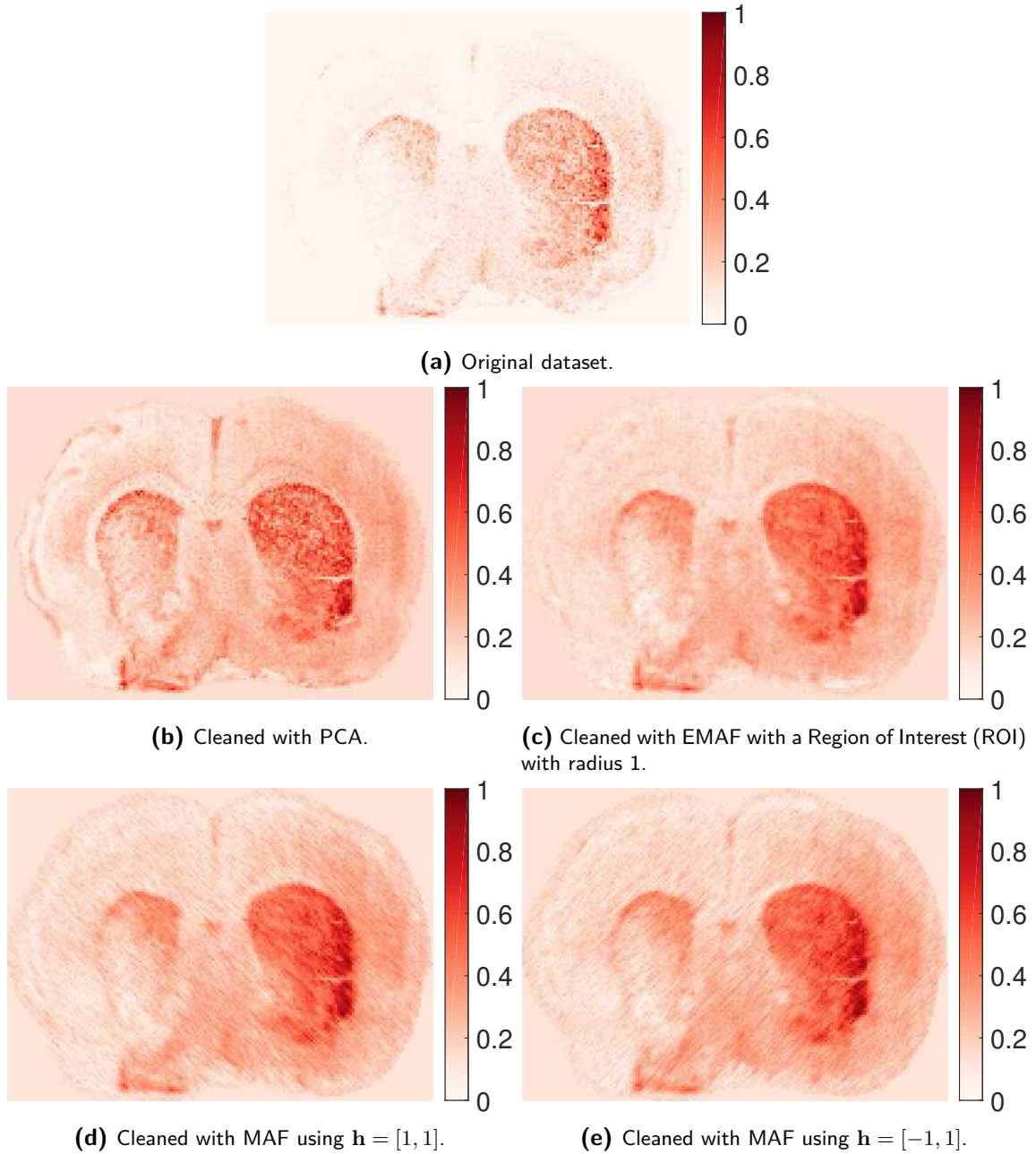
Figure 6-3 shows a measurement of a distinct biological pattern, although maybe contaminated with noise. The ion image cleaned with PCA seems to be an exact copy of the original ion image. The PCA ion image appears clear and seems to capture even the spatially sparse

measurements, which show as speckled patterns. Both the MAF and EMAF ion images have in common that they are slightly blurry compared to the PCA ion image. It is interesting to note that the MAF are blurred in the orientation of their respective shifts. It also seems that the EMAF image is less blurred that the MAF images.

This case study illustrates some of the things an analyst needs to keep in mind when carrying out noise reduction using these factorization methods. The figures shown in this chapter illustrate the difference the shift parameter setting makes in a practical setting. Since all the cleaned data is of decent quality, the number of component or factors included is not likely to be skewing the results. The visually assessed quality of the MAF cleaned data seems to be blurred compared to EMAF and PCA. To this day there are no guidelines for scientist on how to choose this shift parameter, so that is one more thing scientist have to keep in mind, if they decide to use MAF as a factorization method.

However, these examples clearly demonstrate that although PCA based decomposition of multivariate imag

EMAF, on the other hand, is able to avoid such high variance noise and delivers a result set that is enriched

**(a)** Original dataset.

**(b)** Cleaned with PCA.

**(c)** Cleaned with EMAF with a Region of Interest (ROI) with radius 1.

**(d)** Cleaned with MAF using $\mathbf{h} = [1, 1]$.

**(e)** Cleaned with MAF using $\mathbf{h} = [-1, 1]$.

**Figure 6-1:** Original and cleaned versions of the ion image of $m/z$ value $1755$. The dataset was cleaned by using 30 components to reconstruct the original dataset. Each image has been scaled to the range $[0, 1]$. The ion image cleaned with PCA is clear and resembles the original image well. The MAF-cleaned ion images are blurry in the direction of the shift used in the algorithm.

**(a)** Original dataset.



**(b)** Cleaned with PCA.



**(c)** Cleaned with EMAF with a ROI with radius 1.



**(d)** Cleaned with MAF using $\mathbf{h} = [1, 1]$.



**(e)** Cleaned with MAF using $\mathbf{h} = [-1, 1]$.
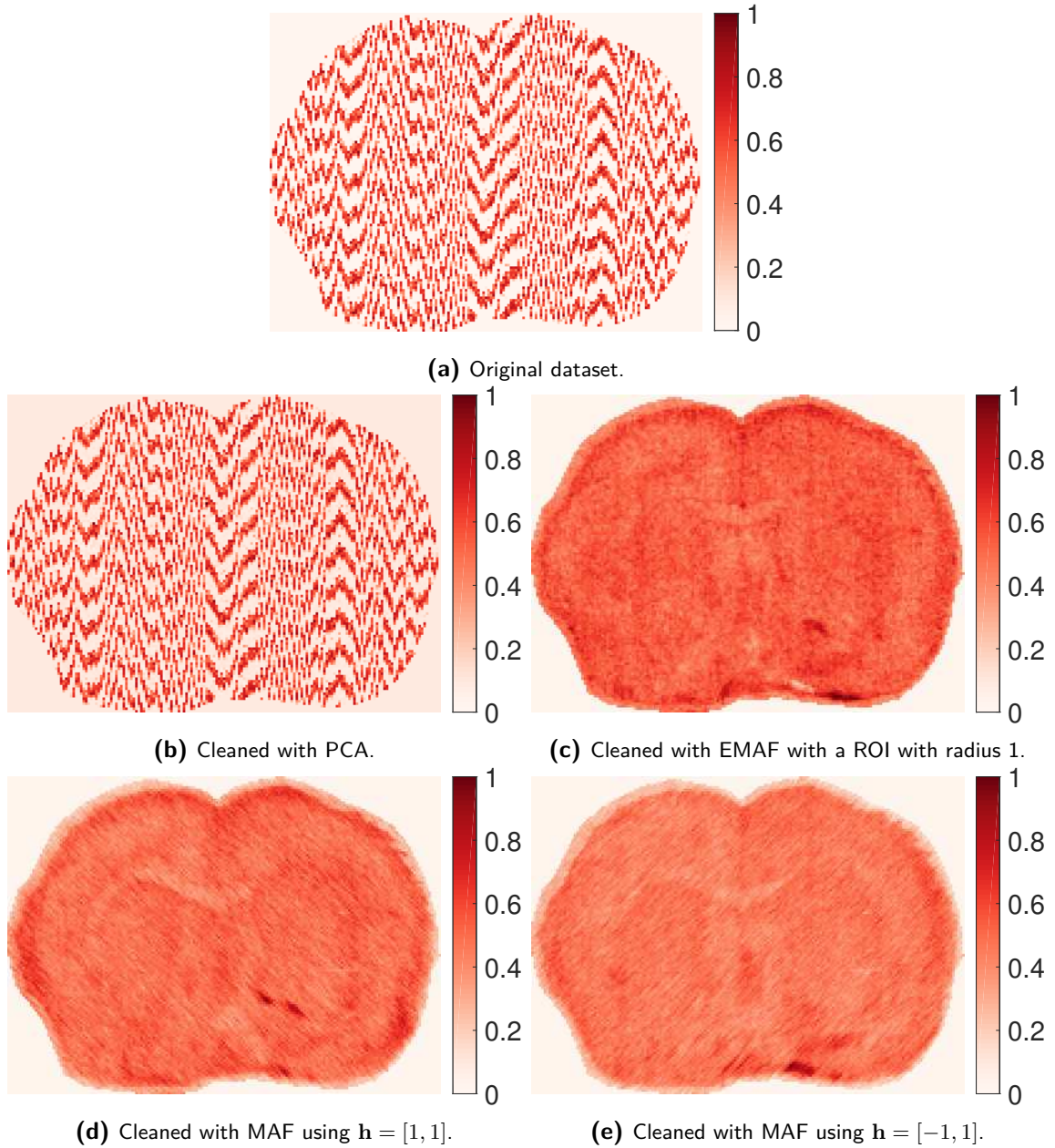
**Figure 6-2:** Original and cleaned versions of the ion image of $m/z$ value $9979$. The dataset was cleaned by using 30 components to reconstruct the original dataset. Each image has been scaled to the range $[0, 1]$. MAF and EMAF cleaned images have managed to reject the striped (and non-biological) pattern found in the original image, unlike the PCA cleaned image. The MAF based methods simply seem to approximate a homogeneous tissue pattern, rejecting the dominant stripe pattern that dominates the PCA results and the original data.

**(a)** Original dataset.



**(b)** Cleaned with PCA.



**(c)** Cleaned with EMAF with a ROI with radius 1.



**(d)** Cleaned with MAF using $\mathbf{h} = [1, 1]$.
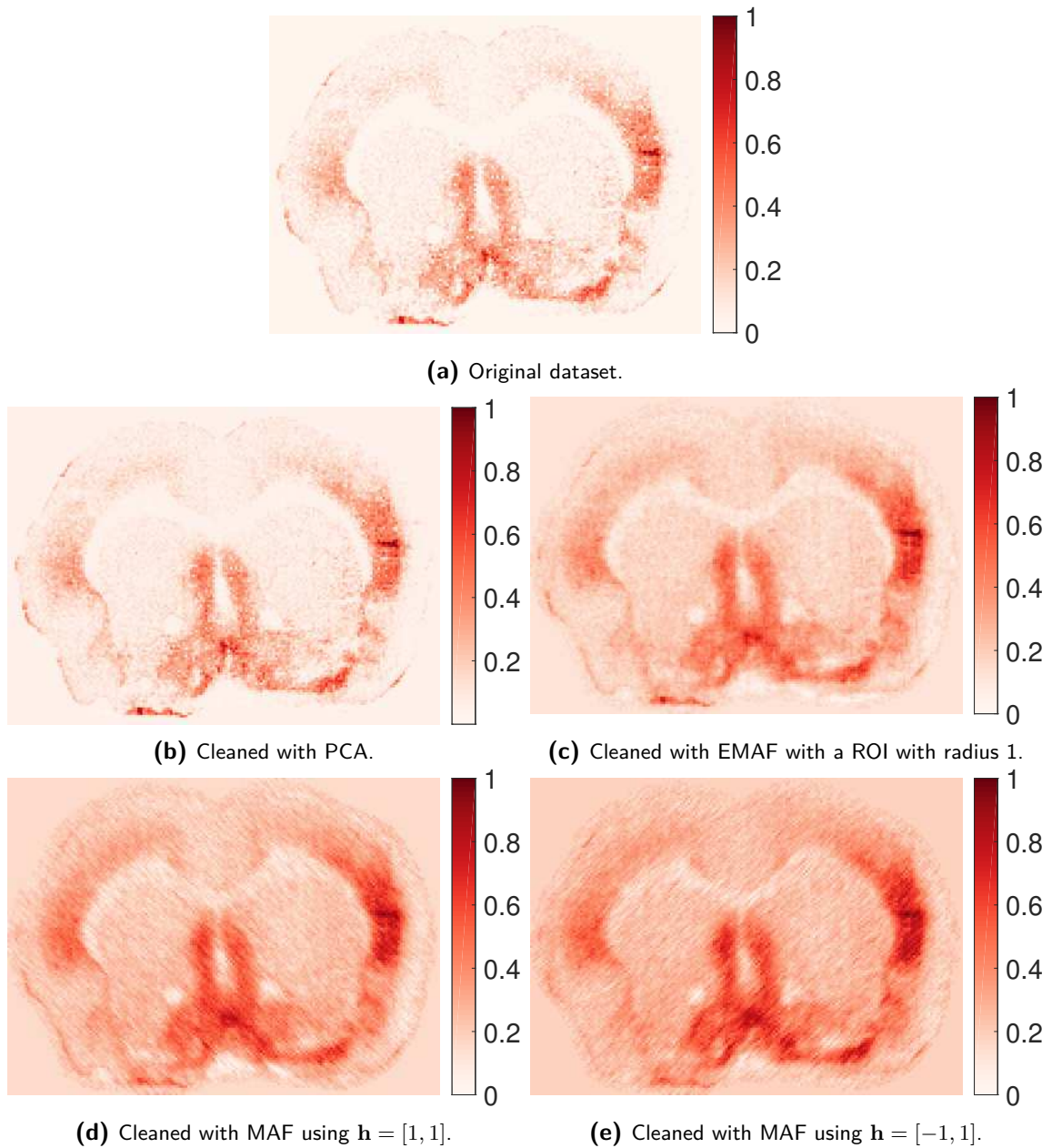


**(e)** Cleaned with MAF using $\mathbf{h} = [-1, 1]$.

**Figure 6-3:** Original and cleaned versions of the ion image of $m/z$ value 9598. The dataset was cleaned by using 30 components to reconstruct the original dataset. Each image has been scaled to the range $[0, 1]$. The MAF cleaned images are blurred in the direction of the shift used. It is also worth noting that the intensities vary slighly between the MAFs, especially in the lower part of the images. The PCA images captures sparse, speckled measurements well and results in a clear image. The EMAF image is slightly blurred, but manages to capture the original measurement the best.

# Chapter 7

# Conclusions

When evaluating the results of the experiments carried out in this thesis, the DirDep and LenDep experiments demonstrated that Maximum Autocorrelation Factorization (MAF) factors are affected by the shift parameter, more specifically in the case where more than two isotropic covariance structures, or an equivalent thereof, are present in the data. In real world data, there are often multiple isotropic and anisotropic covariance structures in the variables of a dataset. As an example of this, a variogram of real world data is shown in Figure 7-1. The variogram in Figure 7-1 would have to be modeled with at least one anisotropic covariance structure and a nugget-effect structure.

The conclusion that MAF factors will in some cases depend on the shift parameter raises the question of how this shift parameter should be tuned. The goal in this thesis was to alter the MAF algorithm in a way that makes the algorithm fully unsupervised, thereby not relying on any inputs from the user. This was to be done without losing any of the attractive features of the original MAF algorithm.

The DirDep experiment showed that the resulting MAF factors favor spatial artifacts oriented in the same direction as the shift parameter or perpendicular to it. This same behavior was also seen in the case study, in Chapter 6. To mitigate this behavior, more shifts were included in the calculations of the factors. More specifically an average across a circular Region of Interest (ROI) was used to calculate the eigenvectors, which should in theory result in an algorithm which does not favor spatial artifacts that are oriented in one direction over any other. To make sure this produces expected results, the DirDep experiment was conducted with Extended Maximum Autocorrelation Factorization (EMAF) as well. The difference between the MAF and EMAF factors in the DirDep experiment indicated that the EMAF algorithm is at least radially invariant of the shifts.

Examining the results of the LenDep experiment is not as straight forward as the DirDep experiment. MAF was applied to the ArtLenDS dataset using different shift parameters with three different lengths, but the same orientation. The resulting MAF factors were different for each of the three shift parameters, showing that MAF is dependent on the length of the shift parameter. This means that the MAF algorithm is reordering and changing which factors are
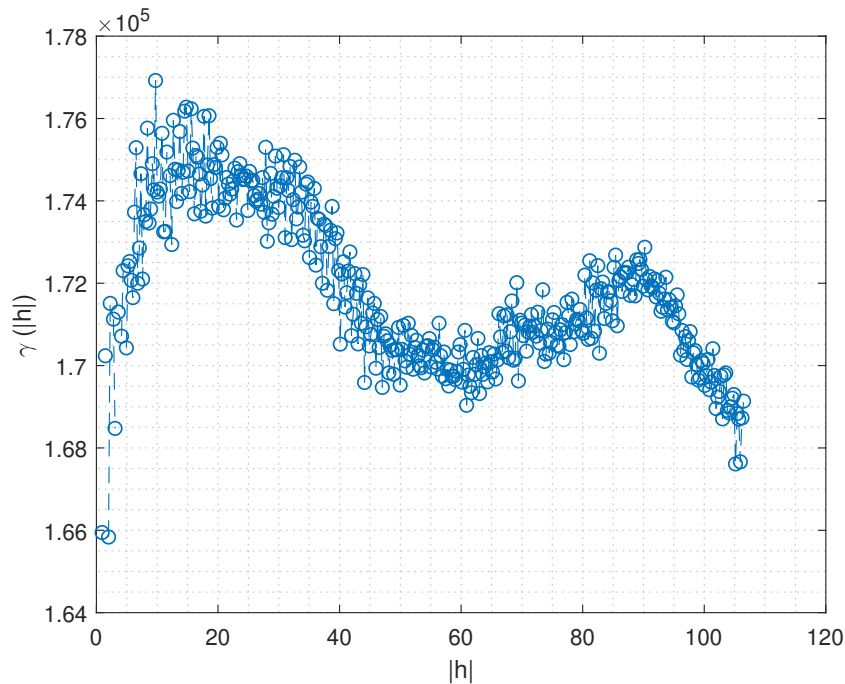
**Figure 7-1:** A real world variogram. This particular variogram would most likely be modeled with a single anisotropic covariance structure as well as a nugget-effect structure.

mostly signal and which are mostly noise. How to extend MAF and address that dependence is not as obvious as with the directional dependence.

MAF operates by trying to separate signal from noise, by ordering the resulting factors from "contains mostly signal" to "contains mostly noise". This ordering is done by observing how rapidly all the measurements in the dataset vary spatially. MAF compares the autocorrelation of the dataset at its initial spatial position with the autocorrelation of the dataset shifted by the shift parameter, this difference in autocorrelation determines where in the signal-to-noise ranking each factor ends up. The shift parameter length should therefor be set to a distance where the difference between spatial change of signal and spatial change of noise is the greatest. Intuitively this difference is greatest where the maximum autocorrelation occurs.

Knowing that spatial autocorrelation generally decreases with distance, it seems logical to use the smallest shift available for the purpose of MAF. Since the shifts can only be integers, that would be a unit length shift. This was done in the original version of MAF [18]. Although, this strategy will give the best results in most cases, in this thesis three strategies of choosing the shift parameter length were proposed. The performance of these strategies will be discussed in detail later.

The LenDep experiment was also carried out using the EMAF algorithm and three different shift lengths. EMAF also changes the order of the factors, similarly to what was seen with MAF, but EMAF seemed more reluctant to change the factors or the order of the factors. This is likely because instead of taking the values of the variogram at a single instance like the MAF algorithm does, the EMAF algorithm looks at the area under the variogram and generates the factors according to that area. This results in the "smoother" change in factors

noticed in the LenDep experiment.

These experiments answer two of the research questions put forth in this thesis. The questions on how the shift parameter length and direction affect the MAF factors. The direction of the shift parameter will make MAF promote spatial artifacts oriented in that particular direction. The length of the shift parameter affects the MAF factors by changing their order and alter the weights assigned to each of the underlying structures, which make up the factors. The underlying structures are a combination of the covariance structures found in the dataset.

Three different methods of picking a radius of the ROI were proposed in Chapter 3. These methods were then compared in the MEval experiment, in Chapter 5. One of these methods was the naive method of choosing a shift lying on the unit circle and the other two were data driven methods finding the distance at which the minimum and maximum spatial autocorrelation occur. As was expected, based on these experiments, it is concluded that, although in most cases the Unit Circle (UC) will probably work well in most cases, the Maximum Autocorrelation (Max AC) methods covers some additional special cases, where the maximal spatial autocorrelation does not occur at a distance of 1.

The logic behind using the Max AC method is that generally spatial autocorrelation will decrease with an increase in shift distance. This means that, naively, a good strategy of picking a shift distance is using the shortest shift available, which is a shift distance of 1 in this case. Although, in special cases where the maximal spatial autocorrelation, i.e. the greatest difference in spatial autocorrelation between signal and noise does not occur at a distance of 1, the Max AC method will pick up on that at relatively low computational expense. For this reason, the Max AC is used for EMAF instead of the UC radius estimation strategy.

Quantifying and assessing the quality of factors is hard when working with unsupervised methods and unlabeled data. In order to do so, an artificial dataset would have to be created and different types of noise added to it. That way, the ground truth is known. Creating an artificial dataset like that was considered out of the scope of this thesis. As discussed earlier, optimally the distance should be set to a value at which the spatial change between the signal and the noise contained in the dataset is the greatest. From the comparison carried out in this thesis, between the radius estimation methods, it is concluded that using the Max AC method is best suited for the EMAF algorithm. But this should preferably be confirmed using an artificial dataset.

Two different implementations of the EMAF algorithm were derived, one which included all shift vectors within the ROI in the lag space, the so called *disc implementation* and another that only included the shifts lying on the rim of the ROI, the so called *rim implementation*. The *disc implementation* seemed to give slightly higher quality factors, although at an expense of increased computation time. Since EMAF is a preprocessing method intended to be used on high dimensional data, this increase in computation time is not thought to be worth the effort in most practical cases. For this reason, the *rim implementation* of EMAF is recommended. It is to be noted that these two methods produce the same results in the case where the radius of the ROI is 1.

In the case study chapter, the factorization methods of interest were applied to real world data, EMAF seemed to perform better than MAF when reducing noise in the dataset, judging from the visual quality of the factors. The MAF-cleaned data seemed to be blurred in the direction

of the shift parameter used. On the other hand, the EMAF cleaned data was less blurred, and not directionally. Judging from visual quality, the EMAF cleaned data was competing with the Principal Component Analysis (PCA)-cleaned data. The PCA cleaned data did pick up some of the sparse measurements which the EMAF regarded as noise. However, this performance gain from using EMAF instead of MAF should be quantified and confirmed using an artificial dataset where the ground truth is known. From a practical standpoint, using EMAF is less troublesome to apply, compared to MAF, since there is no need to select or tune the shift parameter.

The examples shown in the case study clearly demonstrate that although PCA based decomposition of multivariate images is less blurry than EMAF results, it has a non-trivial risk of being seriously skewed by non-biological noise patterns with high variance. This behavior of PCA might explain why PCA-cleaned data often correlates better with the original dataset, compared to MAF-cleaned data. Since both the PCA-cleaned data and original dataset are contaminated with these high variance, non-biological noise patterns. This skewing or contamination was not seen with any of the MAF based methods.

The case study combined with the experiments in this thesis cast light on the last research question, how the shift parameter influences MAF factors in practice. MAF factors change with both length and direction of the shift parameter. However, since only a single direction is used to calculate the MAF factors, when MAF is used for dimensionality reduction that results in a directional blur in the reduced dataset. This effect was shown in the case study, Chapter 6. This blurring effect was reduced by including more shifts in the EMAF algorithm, at all available directions, when calculating the factors.

To conclude, the initial goal was making a version of MAF which was not dependent on the shift parameter, both in terms of the length and direction of the shift parameter. A new algorithm, EMAF, was derived for this purpose. It has been shown that EMAF is independent of linear transformations to the dataset being processed, similarly to MAF. It was also shown that EMAF is independent on the shift parameter in the case where there are two or fewer covariance structures (or an equivalent thereof) present in the data. The performance of this new algorithm was compared, visually, to MAF and it seemed to produce clearer scores and similar loadings as MAF when applied to the same dataset. An additional perk of the EMAF algorithm is that it does not depend on any input from the user.

## 7-1 Future work

In this section, a few ideas of future work will be presented. These ideas could be carried out to further improve the development of unsupervised factorization methods.

- Test EMAF in a classification pipeline and evaluate the performance of the algorithm in such a setting.

- Analyze the performance of EMAF further. Evaluating what type of noise it can handle and in what quantities, compared to PCA or MAF, especially looking at EMAF in a robust setting.

- Optimize the EMAF algorithm to become less computationally expensive, making it more feasible as an industry standard for noise reduction, evaluating its runtime and making sure it scales well with both increased spectral and spatial dimensions.

- Explore how EMAF factors perform when used in a classification setting, compared to PCA components or MAF factors.

- Research if there are any ways of choosing the number of EMAF factors to use in a practical setting, with the ultimate goal of at least publishing guidelines for practical use of the algorithm.

- Look into what possibilities there are of applying rotations to EMAF factors. Could rotations potentially be used to improve the interpretability of factors, similar to what is done with PCA components? This could be a way of constraining the EMAF loadings to be non-negative for example, which would greatly improve interpretability for Imaging Mass Spectrometry (IMS) applications.

# Appendix

## A-1 Region of Interest (ROI)

### A-1-1 Disc

To calculate the area of the ROI mention in Section 3-2 and shown in Figure 3-1a the following integral needs to be evaluated

$$\int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dxdy. \qquad \text{(A-1)}$$

This area is a circle, centered at the origin of the lag space with a radius $r \in \mathbb{R}$. This integral is illustrated in Figure A-1, the area is segmented into small steps along the $x$-axis and the $y$-values constrained to lie on the range $[-\sqrt{r^2-x^2}, \sqrt{r^2-x^2}]$. The area $A$ can now be evaluated as
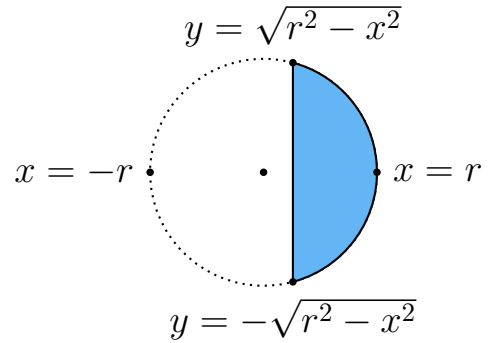


**Figure A-1:** An illustration of the disc being integrated across.

$$A = \int_{x=-r}^{r} \int_{y=-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dydx \qquad \text{(A-2)}$$

$$= \int_{x=-r}^{r} \left( \sqrt{r^2-x^2} - \left(-\sqrt{r^2-x^2}\right) \right) dx \qquad \text{(A-3)}$$

$$= 2 \int_{x=-r}^{r} \sqrt{r^2-x^2} dx. \qquad \text{(A-4)}$$

Now defining $x = r\sin(u)$, getting $dx = r\cos(u)du$. These values can then be substituted in the integral, getting

$$A = 2\int_{x=-r}^{r} \sqrt{r^2 - x^2}dx \tag{A-5}$$

$$= 2\int_{u=\arcsin(-r)}^{\arcsin(r)} \sqrt{r^2 - r^2\sin^2(u)}r\cos(u)du \tag{A-6}$$

$$= 2\int_{u=\arcsin(-r)}^{\arcsin(r)} r\sqrt{1 - \sin^2(u)}r\cos(u)du \tag{A-7}$$

$$= 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \sqrt{1 - \sin^2(u)}\cos(u)du. \tag{A-8}$$

Using the identity

$$\sin^2(u) + \cos^2(u) = 1 \tag{A-9}$$

$$\implies \cos^2(u) = 1 - \sin^2(u), \tag{A-10}$$

we can now get

$$A = 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \sqrt{1 - \sin^2(u)}\cos(u)du \tag{A-11}$$

$$= 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \sqrt{\cos^2(u)}\cos(u)du \tag{A-12}$$

$$= 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \cos(u)\cos(u)du \tag{A-13}$$

$$= 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \cos^2(u)du. \tag{A-14}$$

Using the identity

$$\cos^2(u) = \frac{1 + \cos(2u)}{2}, \tag{A-15}$$

we can now substitute and get

$$A = 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \cos^2(u)du \tag{A-16}$$

$$= 2r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} \frac{1 + \cos(2u)}{2}du \tag{A-17}$$

$$= r^2\int_{u=\arcsin(-r)}^{\arcsin(r)} (1 + \cos(2u))\,du \tag{A-18}$$

$$= r^2\left[u + \frac{1}{2}\sin(2u)\right]_{u=\arcsin(-r)}^{\arcsin(r)}. \tag{A-19}$$

Substituting back to the original variable $x$, by using $u = \arcsin\left(\dfrac{x}{r}\right)$, we get

$$A = r^2\left[u + \frac{1}{2}\sin(2u)\right]_{u=\arcsin(-r)}^{\arcsin(r)} \tag{A-20}$$

$$= r^2\left[\arcsin\left(\frac{x}{r}\right) + \frac{1}{2}\sin\left(2\arcsin\left(\frac{x}{r}\right)\right)\right]_{x=-r}^{r} \tag{A-21}$$

$$= r^2\left[\arcsin\left(\frac{r}{r}\right) + \frac{1}{2}\sin\left(2\arcsin\left(\frac{r}{r}\right)\right) - \left(\arcsin\left(\frac{-r}{r}\right) + \frac{1}{2}\sin\left(2\arcsin\left(\frac{-r}{r}\right)\right)\right)\right] \tag{A-22}$$

$$= r^2\left[\frac{\pi}{2} + \frac{1}{2}\sin\left(2\frac{\pi}{2}\right) - \left(-\frac{\pi}{2} + \frac{1}{2}\sin\left(-2\frac{\pi}{2}\right)\right)\right] \tag{A-23}$$

$$= r^2\left[\frac{\pi}{2} + \frac{1}{2}\sin\left(\pi\right) + \frac{\pi}{2} - \frac{1}{2}\sin\left(-\pi\right)\right] \tag{A-24}$$

$$= r^2\left[\pi + \frac{1}{2}\sin\left(\pi\right) - \frac{1}{2}\sin\left(-\pi\right)\right] \tag{A-25}$$

$$= r^2\pi, \tag{A-26}$$

since $\sin(\pi) = \sin(-\pi) = 0$. The ROI is a circle so this is the expected result. But to show that this parameterization of the region is valid and to use this result in the thesis we carry out these calculations here.

### A-1-2   Rim

To calculate the circumference of the ROI mention in section 3-3 and shown in Figure 3-1b we need to integrate across the boundary, that is evaluating the integral

$$\int_{\theta=0}^{2\pi} r d\theta. \tag{A-27}$$

This integral is the circumference of a circle with radius $r$, centered at the origin, so evaluating the integral we get

$$L = \int_{\theta=0}^{2\pi} r d\theta = r\int_{\theta=0}^{2\pi} d\theta \tag{A-28}$$

$$= r\left[\theta\right]_0^{2\pi} = r(2\pi - 0) \tag{A-29}$$

$$= r2\pi. \tag{A-30}$$

## A-2   Minimizing the Generalized Rayleigh Quotient

The generalized Rayleigh quotient is defined as

$$\frac{x^*\mathbf{A}x}{x^*\mathbf{B}x}, \tag{A-31}$$

where **A** and **B** are both symmetric matrices. In the case where the matrices **A** and **Y** and the vectors $x$ are all real, the definition reduces to

$$\frac{x^T \mathbf{A} x}{x^T \mathbf{B} x}. \tag{A-32}$$

Factorizing **B** such that $\mathbf{B} = \mathbf{C}^T \mathbf{C}$ and then substituting $y = \mathbf{C} x$, the generalized Rayleigh quotient can then be rewritten to the form

$$\frac{y^T \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1} y}{y^T y}. \tag{A-33}$$

For formatting purposes let $\mathbf{D} = \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1}$, arriving at

$$\frac{y^T \mathbf{D} y}{y^T y}, \tag{A-34}$$

this format is known as the Rayleigh quotient. It is to be noted that **D** is a symmetric matrix. The objective function of the minimization problem can now be formatted using a Lagrange multiplier, where the vector $y$ gets normalized according to the equation $y^T y = 1$, as follows

$$\min_y y^T \mathbf{D} y - \lambda (y^T y - 1). \tag{A-35}$$

This gives the Lagrangian

$$L(y) = y^T \mathbf{D} y - \lambda (y^T y - 1), \tag{A-36}$$

with the derivatives

$$\frac{\partial L}{\partial y} = y^T (\mathbf{D} + \mathbf{D}^T) + 2\lambda y^T, \tag{A-37}$$

$$\frac{\partial L}{\partial \lambda} = y^T y - 1. \tag{A-38}$$

Setting $\frac{\partial L}{\partial y} = 0$ gives

$$y^T (\mathbf{D} + \mathbf{D}^T) + 2\lambda y^T = 0, \tag{A-39}$$

$$\implies \mathbf{D} y = -\lambda y, \tag{A-40}$$

since **D** is symmetric. From this it is clear that the value minimizing the Rayleigh quotient is the smallest eigenvalue of **D**. Substituting back the definition of $y$ and **D** to get the original vectors and matrices gives

$$\mathbf{D} y = -\lambda y \tag{A-41}$$

$$\implies \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1} \mathbf{C} x = -\lambda \mathbf{C} x, \tag{A-42}$$

$$\mathbf{C}^{-T} \mathbf{A} x = -\lambda \mathbf{C} x, \tag{A-43}$$

$$\mathbf{A} x = -\lambda \mathbf{C}^T \mathbf{C} x, \tag{A-44}$$

$$\mathbf{A} x = -\lambda \mathbf{B} x, \tag{A-45}$$

according to the definition of **C**. Translating this result, the minimum of the generalized Rayleigh quotient is the minimum eigenvalue of the generalized eigenvalue problem.

## A-3  Data Sphering

Data sphering is a transformation to remove the covariance information contained within data, often performed to help highlight structures in the data beyond linear correlation. Sphering a dataset involves transforming it such that the data gets an identity covariance matrix. There are various ways a dataset can be be sphered, in this section we will list one of these methods as an example, but more methods can be found for example in [66]. The following sphering method is sometimes called Square Root Decomposition (SRD) or SRD for short.

Given a zero-mean variable $\mathbf{Z} \in \mathbb{R}^{n \times p}$, such that

$$\mathrm{E}\left[\mathbf{Z}\right] = 0, \tag{A-46}$$

$$\mathrm{Var}\left[\mathbf{Z}\right] = \mathbf{Z}^T \mathbf{Z} = \mathbf{\Sigma}. \tag{A-47}$$

Note that $\Sigma$ is an estimation of the covariance. In this example a simple estimation technique will be used, although more robust methods could be applied to estimate the covariance. The only condition is that the covariance estimator has to be affine equivariant, that is for any covariance estimator $\hat{\Sigma}(\mathbf{X})$, we have

$$\hat{\Sigma}(\mathbf{X}\mathbf{A} + \mathbf{b}) = \mathbf{A}^T \hat{\Sigma}(\mathbf{X})\mathbf{A}, \tag{A-48}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the data matrix being analyzed, $\mathbf{A} \in \mathbb{R}^{p \times p}$ is any non-singular matrix and $\mathbf{b} \in \mathbb{R}^{n \times p}$ is a bias matrix.

The covariance estimate can now be decomposed as

$$\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{\Sigma}^{-\frac{1}{2}}, \tag{A-49}$$

A variable $\mathbf{Y}$, which contains the sphered data from $\mathbf{Z}$, i.e. with unit covariance, can be defined as

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Sigma}^{-\frac{1}{2}} \tag{A-50}$$

The covariance of $\mathbf{Y}$ can now be estimated as

$$\mathrm{Var}\left[\mathbf{Y}\right] = \mathbf{Y}^T\mathbf{Y} = \left(\mathbf{Z}\mathbf{\Sigma}^{-\frac{1}{2}}\right)^T \mathbf{Z}\mathbf{\Sigma}^{-\frac{1}{2}} \tag{A-51}$$

$$= \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Z}^T\mathbf{Z}\mathbf{\Sigma}^{-\frac{1}{2}} \tag{A-52}$$

$$= \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{\Sigma}^{-\frac{1}{2}} \tag{A-53}$$

$$= \mathbf{I}, \tag{A-54}$$

since the covariance matrix is symmteric. This is the result that was expected. Sphering is visulized in Figure A-2.

## A-4  Equal Eigenvectors

Proving the statement that the two, arbitrary matrices $\mathbf{A}$ and $a\mathbf{I} + b\mathbf{A}$ have the same eigenvalues, where $\mathbf{I}$ is the identity matrix with the same size as $\mathbf{A}$ and $a, b \in \mathbb{R}$, $b \neq 0$. Let the eigenvalue decomposition of $\mathbf{A}$ be

$$\mathbf{A} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1}, \tag{A-55}$$

**(a)** Original data                    **(b)** Sphered data
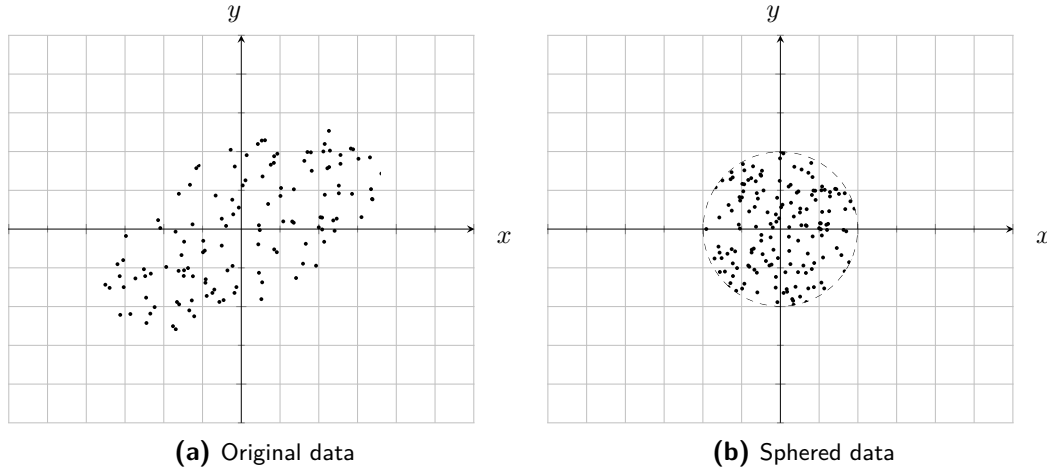
**Figure A-2:** Two-dimensional illustration of sphering. The original dataset is rotated and scaled to have close to unit covariance in all dimensions. In this case the data has been centered around the origin, without loss of generality.

where $\mathbf{W}$ is a matrix with the eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}$. Note that this requires the matrix of eigenvectors to be normalized such that

$$\mathbf{W}^{-1}\mathbf{W} = \mathbf{I}. \tag{A-56}$$

Then

$$a\mathbf{I} + b\mathbf{A} = a\mathbf{W}\mathbf{I}\mathbf{W}^{-1} + b\mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1} = \mathbf{W}\left(a\mathbf{I} + b\mathbf{\Lambda}\right)\mathbf{W}^{-1}, \tag{A-57}$$

which means that $a\mathbf{I} + b\mathbf{A}$ and $\mathbf{A}$ do have the same eigenvectors $\mathbf{W}$. The two matrices do not have the same eigenvalues, $\mathbf{A}$ will have the eigenvalues $\mathbf{\Lambda}$, while $a\mathbf{I} + b\mathbf{A}$ has eigenvalues $a\mathbf{I} + b\mathbf{\Lambda}$ as demonstrated.

## A-5   Proof: Special case of Extended Maximum Autocorrelation Factorization (EMAF) invariance to the shift parameter

In this section the goal is to proof that similarly to Maximum Autocorrelation Factorization (MAF) factors, EMAF factors will return the same factors regardless of how the shift parameter is chosen, for a special case of covariance structures on the data being factorized. This special case is when each variable only contains two or fewer isotropic covariance structures or the equivalent thereof. In the case of MAF this is both shown in Section 2-3-2 and discussed in [58].

To prove this the EMAF factors will be calculated in the case of one and two covariance structures, which will show that the spatial autocorrelation function can be diagnolized in both cases, regarless of how the shift parameter is chosen. This will, however, only be shown for the *rim implementation*, as for the *disc implementation* the exact same logic and argumentation applies.

## A-5-1 One covariance structure

In the case of a single nested substructure the covariance function of $\mathbf{Z}(\mathbf{x})$ can be written as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}), \tag{A-58}$$

where $c_1(\mathbf{h}) \in \mathbb{R}$ for all $\mathbf{h} \in \mathbb{Z}^2$ is a covariance function of the single substructure and $\mathbf{B}_1 \in \mathbb{R}^{p \times p}$ is a weight matrix for that substructure. This means that $\mathbf{C}(0,0) = \mathbf{B}_1$. In this case it is noted that the shift parameter is annotated in polar coordinates, that is $\mathbf{h} = (r, \theta)$. Then the minimization problem becomes

$$\frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \gamma(r, \theta) \Delta\theta w_i}{w_i^T \mathbf{C}(0,0) w_i} = \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} (\mathbf{C}(0,0) - \mathbf{C}(r,\theta)) \Delta\theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-59}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} (\mathbf{B}_1 - \mathbf{B}_1 c_1(\mathbf{h})) \Delta\theta w_i}{w_i^T \mathbf{B}_1 w_i} \tag{A-60}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} (\mathbf{B}_1 (1 - c_1(\mathbf{h}))) \Delta\theta w_i}{w_i^T \mathbf{B}_1 w_i} \tag{A-61}$$

$$= \frac{w_i^T \mathbf{B}_1 w_i}{w_i^T \mathbf{B}_1 w_i} \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} (1 - c_1(\mathbf{h})) \Delta\theta \tag{A-62}$$

$$= \mathbf{I}_{p \times p} \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} (1 - c_1(\mathbf{h})) \Delta\theta, \tag{A-63}$$

since $\mathbf{B}_1$ is invariant of the shift parameter. This means that since the weighted average correlation is similar to the identity matrix, it can be diagnolized, regardless of how the shift parameter is chosen. This in turns means that EMAF will, in this special setting, produce spatially uncorrelated factors at all distances regardless of how the shift parameter is set.

## A-5-2 Two covariance structures

In the case of two nested substructures in the covariance function of $\mathbf{Z}(\mathbf{x})$ can be written as

$$\mathbf{C}(\mathbf{h}) = \mathbf{B}_1 c_1(\mathbf{h}) + \mathbf{B}_2 c_2(\mathbf{h}), \tag{A-64}$$

where $c_1(\mathbf{h}), c_2(\mathbf{h}) \in \mathbb{R}$ for all $\mathbf{h} \in \mathbb{Z}^2$ are covariance functions of the covariance structures and $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{p \times p}$ are weight matrices for those structure. This means that $\mathbf{C}(0,0) = \mathbf{B}_1 + \mathbf{B}_2 \implies \mathbf{B}_2 = \mathbf{C}(0,0) - \mathbf{B}_1$.

Similarly to the single covariance structure case, the average correlation is analysed. In this

case it is noted that the shift parameter is annotated in polar coordinates, that is $\mathbf{h} = (r, \theta)$.

$$\frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \gamma(r, \theta) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-65}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( \mathbf{C}(0,0) - \mathbf{C}(r,\theta) \right) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-66}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( \mathbf{B}_1 + \mathbf{B}_2 - \mathbf{B}_1 c_1(\mathbf{h}) - \mathbf{B}_2 c_2(\mathbf{h}) \right) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-67}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( \mathbf{C}(0,0) - \mathbf{B}_1 c_1(\mathbf{h}) - \left( \mathbf{C}(0,0) - \mathbf{B}_1 \right) c_2(\mathbf{h}) \right) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-68}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( \mathbf{C}(0,0) \left( 1 - c_2(\mathbf{h}) \right) - \mathbf{B}_1 \left( c_1(\mathbf{h}) - c_2(\mathbf{h}) \right) \right) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-69}$$

$$= \frac{w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \mathbf{C}(0,0) \left( 1 - c_2(\mathbf{h}) \right) \Delta \theta w_i - w_i^T \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \mathbf{B}_1 \left( c_1(\mathbf{h}) - c_2(\mathbf{h}) \right) \Delta \theta w_i}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-70}$$

$$= \frac{w_i^T \mathbf{C}(0,0) w_i \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( 1 - c_2(\mathbf{h}) \right) \Delta \theta - w_i^T \mathbf{B}_1 w_i \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( c_1(\mathbf{h}) - c_2(\mathbf{h}) \right) \Delta \theta}{w_i^T \mathbf{C}(0,0) w_i} \tag{A-71}$$

$$= \mathbf{I}_{p \times p} \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( 1 - c_2(\mathbf{h}) \right) \Delta \theta - \frac{w_i^T \mathbf{B}_1 w_i}{w_i^T \mathbf{C}(0,0) w_i} \frac{1}{2\pi} \sum_{\theta=0}^{2\pi} \left( c_1(\mathbf{h}) - c_2(\mathbf{h}) \right) \Delta \theta. \tag{A-72}$$

The Generalized Rayleigh Quotient, $\dfrac{w_i^T \mathbf{B}_1 w_i}{w_i^T \mathbf{C}(0,0) w_i}$ is diagonalizable [59] and independent on the included shift parameters. The matrix $\mathbf{I}_{p \times p}$ is also diagonal and independent on the included shift parameters. That means that the resulting EMAF factors will be spatially uncorrelated at all distances invariant of the included shifts, in this special case.

# Bibliography

[1] L. A. McDonnell and R. M. Heeren, "Imaging mass spectrometry," *Mass Spectrometry Reviews*, vol. 26, pp. 606–643, Apr. 2007.

[2] P. Chaurand, S. A. Schwartz, M. L. Reyzer, and R. M. Caprioli, "Imaging Mass Spectrometry: Principles and Potentials," *Toxicologic Pathology*, vol. 33, pp. 92–101, Jan. 2005.

[3] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann, "Large-Scale Proteomic Analysis of the Human Spliceosome," *Genome Research*, vol. 12, pp. 1231–1245, Aug. 2002.

[4] N. Verbeeck, J. M. Spraggins, M. J. M. Murphy, H.-d. Wang, A. Y. Deutch, R. M. Caprioli, and R. Van de Plas, "Connecting imaging mass spectrometry and magnetic resonance imaging-based anatomical atlases for automated anatomical interpretation and differential analysis," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1865, pp. 967–977, July 2017.

[5] P. W. Siy, R. A. Moffitt, R. M. Parry, Y. Chen, Y. Liu, M. C. Sullards, A. H. Merrill, and M. D. Wang, "Matrix factorization techniques for analysis of imaging mass spectrometry data," in *2008 8th IEEE International Conference on BioInformatics and BioEngineering*, pp. 1–6, Oct. 2008.

[6] R. Van de Plas, F. Ojeda, M. Dewil, L. Van Den Bosch, B. De Moor, and E. Waelkens, "Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 458–469, 2007.

[7] D. C. Lay, *Linear Algebra and Its Applications*. Pearson, 4th edition ed., 2012.

[8] K. A. Hoo, K. J. Tvarlapati, M. J. Piovoso, and R. Hajare, "A method of robust multivariate outlier replacement," *Computers & Chemical Engineering*, vol. 26, pp. 17–39, Jan. 2002.

[9] I. T. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, New York: Springer-Verlag, 2 ed., 2002.

[10] R. V. de Plas, B. D. Moor, and E. Waelkens, "Imaging mass spectrometry based exploration of biochemical tissue composition using peak intensity weighted PCA," in *2007 IEEE/NIH Life Science Systems and Applications Workshop*, pp. 209–212, Nov. 2007.

[11] J.-W. Park, H. Min, Y.-P. Kim, H. K. Shon, J. Kim, D. W. Moon, and T. G. Lee, "Multivariate analysis of ToF-SIMS data for biological applications," *Surface and Interface Analysis*, vol. 41, pp. 694–703, Aug. 2009.

[12] B. Balluff, C. K. Frese, S. K. Maier, C. Schöne, B. Kuster, M. Schmitt, M. Aubele, H. Höfler, A. M. Deelder, A. Heck, P. C. W. Hogendoorn, J. Morreau, A. F. Maarten Altelaar, A. Walch, and L. A. McDonnell, "De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry," *The Journal of Pathology*, vol. 235, pp. 3–13, Jan. 2015.

[13] E. A. Jones, S.-O. Deininger, P. C. W. Hogendoorn, A. M. Deelder, and L. A. McDonnell, "Imaging mass spectrometry statistical analysis," *Journal of Proteomics*, vol. 75, pp. 4962–4989, Aug. 2012.

[14] W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O'Shea, "Global analysis of protein localization in budding yeast," *Nature*, vol. 425, p. 686, Oct. 2003.

[15] M. Stoeckli, T. B. Farmer, and R. M. Caprioli, "Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument," *Journal of the American Society for Mass Spectrometry*, vol. 10, pp. 67–71, Jan. 1999.

[16] H. Nygren, P. Malmberg, C. Kriegeskotte, and H. F. Arlinghaus, "Bioimaging TOF-SIMS: Localization of cholesterol in rat kidney sections," *FEBS Letters*, vol. 566, no. 1-3, pp. 291–293, 2004.

[17] R. M. A. Heeren, "Proteome imaging: A closer look at life's organization," *PROTEOMICS*, vol. 5, no. 17, pp. 4316–4326, 2005.

[18] P. Switzer and A. A. Green, "Min/Max Autocorrelation Factors for Multivariate Spatial Imagery | Department of Statistics," tech. rep., Stanford University, Stanford, California, Apr. 1984.

[19] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, pp. 198–207, Mar. 2003.

[20] A. G. Marshall, C. L. Hendrickson, and G. S. Jackson, "Fourier transform ion cyclotron resonance mass spectrometry: A primer," *Mass Spectrometry Reviews*, vol. 17, no. 1, pp. 1–35, 1998.

[21] W. C. Wiley and I. H. McLaren, "Time-of-Flight Mass Spectrometer with Improved Resolution," *Review of Scientific Instruments*, vol. 26, pp. 1150–1157, Dec. 1955.

[22] A. G. Marshall and C. L. Hendrickson, "Fourier transform ion cyclotron resonance detection: Principles and experimental configurations," *International Journal of Mass Spectrometry*, vol. 215, pp. 59–75, Apr. 2002.

[23] B. Domon and R. Aebersold, "Mass Spectrometry and Protein Analysis," *Science*, vol. 312, pp. 212–217, Apr. 2006.

[24] P. J. Trim, M.-C. Djidja, T. Muharib, L. M. Cole, B. Flinders, V. A. Carolan, S. Francese, and M. R. Clench, "Instrumentation and software for mass spectrometry imaging—Making the most of what you've got," *Journal of Proteomics*, vol. 75, pp. 4931–4940, Aug. 2012.

[25] T. Alexandrov, "MALDI imaging mass spectrometry: Statistical data analysis and current computational challenges," *BMC Bioinformatics*, vol. 13, p. S11, Nov. 2012.

[26] H. Wackernagel, "Mean, Variance, Covariance," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 9–14, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[27] H. Wackernagel, "Linear Regression and Simple Kriging," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 15–26, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[28] H. Wackernagel, "Direct and Cross Covariances," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 145–150, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[29] H. Wackernagel, "Intrinsic Multivariate Correlation," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 154–157, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[30] H. Wackernagel, "Variogram and Covariance Function," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 50–56, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[31] M. Bachmaier and M. Backes, "Variogram or semivariogram? Understanding the variances in a variogram," *Precision Agriculture*, vol. 9, pp. 173–175, June 2008.

[32] H. Wackernagel, "Variogram Cloud," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 45–49, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[33] H. Wackernagel, "Anisotropy," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 62–65, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[34] D. L. Zimmerman, "Another look at anisotropy in geostatistics," *Mathematical Geology*, vol. 25, pp. 453–470, May 1993.

[35] M. Eriksson and P. P. Siska, "Understanding Anisotropy Computations," *Mathematical Geology*, vol. 32, pp. 683–700, Aug. 2000.

[36] H. Wackernagel, "Multivariate Nested Variogram," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 175–182, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[37] "Geostatistics: Modeling Spatial Uncertainty, 2nd Edition." https://www.wiley.com/en-us/Geostatistics%3A+Modeling+Spatial+Uncertainty%2C+2nd+Edition-p-9780470183151.

[38] P. A. P. Moran, "Notes on Continuous Stochastic Phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

[39] M. Hanselmann, M. Kirchner, B. Y. Renard, E. R. Amstalden, K. Glunde, R. M. A. Heeren, and F. A. Hamprecht, "Concise Representation of Mass Spectrometry Images by Probabilistic Latent Semantic Analysis," *Analytical Chemistry*, vol. 80, pp. 9649–9658, Dec. 2008.

[40] L. A. Klerk, A. Broersen, I. W. Fletcher, R. van Liere, and R. M. A. Heeren, "Extended data analysis strategies for high resolution imaging MS: New methods to deal with extremely large image hyperspectral datasets," *International Journal of Mass Spectrometry*, vol. 260, pp. 222–236, Feb. 2007.

[41] T. A. Blake, J. F. Kelly, N. B. Gallagher, P. L. Gassman, and T. J. Johnson, "Passive standoff detection of RDX residues on metal surfaces via infrared hyperspectral imaging," *Analytical and Bioanalytical Chemistry*, vol. 395, pp. 337–348, Sept. 2009.

[42] G. McCombie, D. Staab, M. Stoeckli, and R. Knochenmuss, "Spatial and Spectral Correlations in MALDI Mass Spectrometry Images by Clustering and Multivariate Analysis," *Analytical Chemistry*, vol. 77, pp. 6118–6124, Oct. 2005.

[43] Y. Gut, M. Boiret, L. Bultel, T. Renaud, A. Chetouani, A. Hafiane, Y.-M. Ginot, and R. Jennane, "Application of chemometric algorithms to MALDI mass spectrometry imaging of pharmaceutical tablets," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 105, pp. 91–100, Feb. 2015.

[44] N. Verbeeck, *Datamining of Imaging Mass Spectrometry Data for Biomedical Tissue Exploration.* PhD thesis, KU Leuven, 2014.

[45] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 65–74, Jan. 1988.

[46] B. J. Tyler, G. Rayal, and D. G. Castner, "Multivariate analysis strategies for processing ToF-SIMS images of biomaterials," *Biomaterials*, vol. 28, pp. 2412–2423, May 2007.

[47] J. Hanrieder, O. Karlsson, E. Brittebo, P. Malmberg, and A. G. Ewing, "Probing the lipid chemistry of neurotoxin-induced hippocampal lesions using multimodal imaging mass spectrometry," *Surface and interface analysis: SIA*, vol. 46, pp. 375–378, Nov. 2014.

[48] E. A. Jones, A. van Remoortere, R. J. M. van Zeijl, P. C. W. Hogendoorn, J. V. M. G. Bovée, A. M. Deelder, and L. A. McDonnell, "Multiple Statistical Analysis Techniques Corroborate Intratumor Heterogeneity in Imaging Mass Spectrometry Datasets of Myxofibrosarcoma," *PLOS ONE*, vol. 6, p. e24913, Sept. 2011.

[49] A. Henderson, J. S. Fletcher, and J. C. Vickerman, "A comparison of PCA and MAF for ToF-SIMS image interpretation," *Surface and Interface Analysis*, vol. 41, pp. 666–674, Aug. 2009.

[50] M. B. Richman, "Rotation of principal components," *Journal of Climatology*, vol. 6, no. 3, pp. 293–335, 1986.

[51] I. T. Jolliffe, "Rotation of principal components: Some comments," *Journal of Climatology*, vol. 7, pp. 507–510, Sept. 1987.

[52] M. L. B. Richman, "Rotation of principal components: A reply," *Journal of Climatology*, vol. 7, pp. 511–520, Sept. 1987.

[53] I. T. Jolliffe, "Rotation of Ill-Defined Principal Components," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 38, no. 1, pp. 139–147, 1989.

[54] I. T. Jolliffe, "Rotation of principal components: Choice of normalization constraints," *Journal of Applied Statistics*, vol. 22, pp. 29–35, Jan. 1995.

[55] A. M. Mestas-Nuñez, "Orthogonality properties of rotated empirical modes," *International Journal of Climatology*, vol. 20, pp. 1509–1516, Oct. 2000.

[56] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies," *Remote Sensing of Environment*, vol. 64, pp. 1–19, Apr. 1998.

[57] M. R. Keenan and V. S. Smentkowski, "Simple statistically based alternatives to MAF for ToF-SIMS spectral image analysis," *Surface and Interface Analysis*, vol. 43, pp. 1616–1626, Dec. 2011.

[58] J. A. Vargas-Guzmán and R. Dimitrakopoulos, "Computational properties of min/max autocorrelation factors," *Computers & Geosciences*, vol. 29, pp. 715–723, July 2003.

[59] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 2nd ed., 2012.

[60] A. J. Desbarats and R. Dimitrakopoulos, "Geostatistical Simulation of Regionalized Pore-Size Distributions Using Min/Max Autocorrelation Factors," *Mathematical Geology*, vol. 32, pp. 919–942, Nov. 2000.

[61] R. Larsen, "Decomposition using maximum autocorrelation factors," *Journal of Chemometrics*, vol. 16, pp. 427–435, Aug. 2002.

[62] R. Bro, E. Acar, and T. G. Kolda, "Resolving the sign ambiguity in the singular value decomposition," *Journal of Chemometrics*, vol. 22, no. 2, pp. 135–140, 2008.

[63] A. A. N. Nielsen, *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-Temporal Data*. PhD Dissertation Inf., Technical University of Denmark, Denmark, 1994.

[64] A. A. Nielsen, "Kernel Maximum Autocorrelation Factor and Minimum Noise Fraction Transformations," *IEEE Transactions on Image Processing*, vol. 20, pp. 612–624, Mar. 2011.

[65] R. A. Maronna, D. R. Martin, and V. J. Yohai, "Multivariate Analysis: Theory and methods," in *Robust Statistics*, pp. 175–228, John Wiley & Sons, Ltd, 2006.

[66] G. Li and J. Zhang, "Sphering and Its Properties," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 60, no. 1, pp. 119–133, 1998.

[67] G. Storvik, "Data reduction by separation of signal and noise components for multivariate spatial images," *Journal of Applied Statistics*, vol. 20, pp. 127–136, Jan. 1993.

[68] M. Hubert and M. Debruyne, "Minimum covariance determinant," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 36–43, 2010.

[69] M. Hubert, M. Debruyne, and P. J. Rousseeuw, "Minimum Covariance Determinant and Extensions," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 10, p. e1421, May 2018.

[70] P. J. Rousseeuw and K. V. Driessen, "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, vol. 41, pp. 212–223, Aug. 1999.

[71] K. Boudt, P. J. Rousseeuw, S. Vanduffel, and T. Verdonck, "The minimum regularized covariance determinant estimator," *Statistics and Computing*, Apr. 2019.

[72] H. Wackernagel, "Examples of Covariance Functions," in *Multivariate Geostatistics: An Introduction with Applications* (H. Wackernagel, ed.), pp. 57–61, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[73] B. J. Tyler, "Multivariate statistical image processing for molecular specific imaging in organic and bio-systems," *Applied Surface Science*, vol. 252, pp. 6875–6882, July 2006.

# Glossary

## List of Acronyms

**MAF**        Maximum Autocorrelation Factorization

**EMAF**       Extended Maximum Autocorrelation Factorization

**PCA**        Principal Component Analysis

**IMS**        Imaging Mass Spectrometry

**MALDI**      Matrix-Assisted Laser Desorption/Ionization

**SIMS**       Secondary Ion Mass Spectrometry

**ToF**        Time of Flight

**FT-ICR**     Fourier Transform Ion Cyclotron Resonance

**ICA**        Independent Component Analysis

**NNMF**       Non-Negative Matrix Factorization

**ROI**        Region of Interest

**Min AC**     Minimum Autocorrelation

**Max AC**     Maximum Autocorrelation

**UC**         Unit Circle

**SRD**        Square Root Decomposition

**MCD**        Minimum Covariance Determinant

**ArtDirDS**   Artificial, Directionally Dependent Dataset

**ArtLenDS**   Artificial, Length Dependent Dataset

**IMS-RBDS**   Imaging Mass Spectrometry - Rat Brain Dataset

**DirDep**       Directional Dependence

**LenDep**       Length Dependence

**MEval**        Method Evaluation