DISSERTATION

zur Erlangung des Doktorgrades der
Fakultät für Angewandte Wissenschaften
der Albert-Ludwigs-Universität Freiburg
im Breisgau

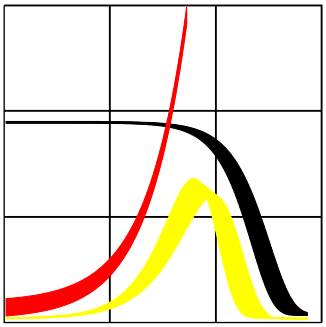
A Semi-Automatic Procedure for the Evaluation of Tooth Cementum Annulation Images for Paleodemographic Studies

von

Katy Klauenberg

SUPPORTED BY:

Max Planck Institute for Demographic Research in Rostock



DEKAN der Fakultät für Angewandte Wissenschaften der Albert-Ludwigs-Universität Freiburg: Prof. Dr. Bernhard Nebel

GUTACHTER der Dissertation: Prof. Dr.-Ing. Hans Burkhardt, Prof. Francesco Lagona, Prof. Dr. Ursula Wittwer-Backofen

DATUM der Promotion: 12.03.2007

Acknowledgements

The topic of this dissertation arose from the tooth laboratory of the Max Planck Institute for Demographic Research. It was of joint interest to anthropologists as well as to demographers. The Albert-Ludwigs-Universität Freiburg and especially the Chair of Pattern Recognition and Image Processing, Prof. Dr.-Ing. Burkhardt, kindly agreed to accept the supervision of this external dissertation.

I am therefore indebted to the Max Planck Institute for Demographic Research as well as to the University of Freiburg for their friendly cooperation and for this excellent opportunity. The development of this dissertation certainly marks a significant step in my career as a researcher.

Extraordinary thanks are devoted to Francesco Lagona. His excellent supervision at the Max Planck Institute during the entire process of my doctorate included open-ended instructions, engaged discussions, innumerable helpful comments, countless reviews and not least of all, friendship. This perfect mixture contributed a great deal to the final product.

Likewise, to my Doktorvater, Prof. Dr.-Ing. Burkhardt, belongs thanks for the repeated opportunities for discussion and advice. Invitations to the Chair of Pattern Recognition and Image Processing always provided fruitful feedback by its members.

I would also like to express my gratitude to the two referees of this dissertation as well as to all other members of the examination committee for their willingness and effort to deal with this topic and for their helpful comments.

The Max Planck Institute did not solely provide a Ph.D. position to me, but also an international and interdisciplinary working atmosphere with excellent facilities. Every single member (and alumni) contributed to this stimulating and at the same time hospitable environment. I acknowledge the director of the Max Planck Institute, Jim Vaupel, for establishing and fostering this atmosphere. I thank my boss, Jutta Gampe, for her guidance

which allowed for great latitude, as well as for flexibility during and after my own demographic transitions. I also appreciated the constant support by the tooth lab and its members, especially Alexander Fabig, Ursula Wittwer-Backofen and Uta Cleven. Special thanks go to my colleagues and friends Matthias Mauch, Sabine Schnabel and Francesco Cottone, as well as other StatLab members, alumni and associates for their support and discussions.

Jenae Tharaldson improved the English language in this thesis through her patient editing.

Uttermost gratitude goes to my family for their support and endurance related to this project, especially for their understanding during the last exhausting months and weeks. I am grateful to my son for his distractions, and to my husband and parents for limiting the distractions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Tooth Cementum Annulation (TCA) images are microscopic images from the root of human teeth. These digital images display annual incremental lines which can be used for age estimation. Until now, the incremental lines (or tooth rings) have been counted manually from under the microscope or on the (digital) TCA image. Since research using the manual observations led to contradictory results, algorithms to automatically evaluate TCA images are considered a crucial step towards computer-assisted TCA age estimation.

TCA images can be evaluated based on measuring their features, but until now no statistical model was developed. In this work, TCA images are modeled as hidden Markov random fields, because these models can incorporate prior knowledge about tooth rings and are thereby able to imitate human vision. In particular, the Markov random field is specified by the FRAME model which incorporates filter responses to the label image into the Gibbsian distribution and is thus able to take into account long-range dependencies among the observed values and periodicity in the placement of tooth rings. The estimation of model parameters is rendered possible via an EM algorithm. This coherent approach is developed step-by-step and tested extensively throughout this work.

The current chapter serves three purposes. It will first describe the TCA method in Section 1.1, starting with the tooth in general, its cementum band and a brief history as well as the aim of TCA analysis. Section 1.2 will then describe the acquisition procedure and the selection procedure for the TCA images, specifically for one sample set, the so-called series of spital-field St. Johann Basel. The content of the remaining work will be introduced in Section 1.3. This section will outline the important points of each chapter

by briefly explaining the methods which will be studied to evaluate TCA images, including the methods' main advantages and drawbacks.

1.1 Tooth Cementum Annulation

Human teeth consist of three kinds of hard tissue: enamel, dentin and cementum. The cementum surrounds the dentin at the root of the teeth, anchors teeth and undergoes appositional growth during the whole life span. Figure 1.1 displays a diagram of a vertical cross section of a tooth as well as an image of a horizontal cross section of the root. As cementum is the tissue of interest here, a short discription is given in this paragraph. (For details, see Kagerer and Grupe [2001]). Cellular cementum surrounds the tip of the root of teeth, and acellular cementum is thickest in the middle third and decreases towards the tip and crown of teeth. Acellular cementum is approximately 20 to 250 μm thick and displays as layers of alternating dark and light bands. A pair of dark and light bands is called an incremental line; a dark band is simply called a tooth ring here, and the formation process is called tooth cementum annulation.

The biological and biochemical basis for the appearance of the alternating dark and light bands is not yet fully understood, but is suspected to stem from the changes in living conditions caused by seasonal rhythms, which trigger via complex mechanisms, for example, fluctuating calcium levels. It is not known exactly when the formation process of incremental lines begins, but it is usually set to the time of a tooth's eruption. (Wittwer-Backofen et al. [2004])

Since the 1950s, wildlife biologists have estimated the age of a wide range of mammals from tooth ring structures. In the 1980s Stott, Sis and Levy were the first to test the applicability of TCA for age estimation on human teeth. In their study (Stott et al. [1982]), tooth rings in cross sections of ten teeth from three individuals were counted, and the tooth-specific eruption age was added in order to estimate age. Ever since, a variety of studies have been undertaken that support or reject the TCA method for age estimation on different samples. Because different samples, different sample sizes as well as different preparation and evaluation techniques were used, these studies are not comparable. See, for example, Stott et al. [1982], Miller et al. [1988], Großkopf [1990], Kagerer and Grupe [2001] and Wittwer-Backofen et al. [2004].

a) b)

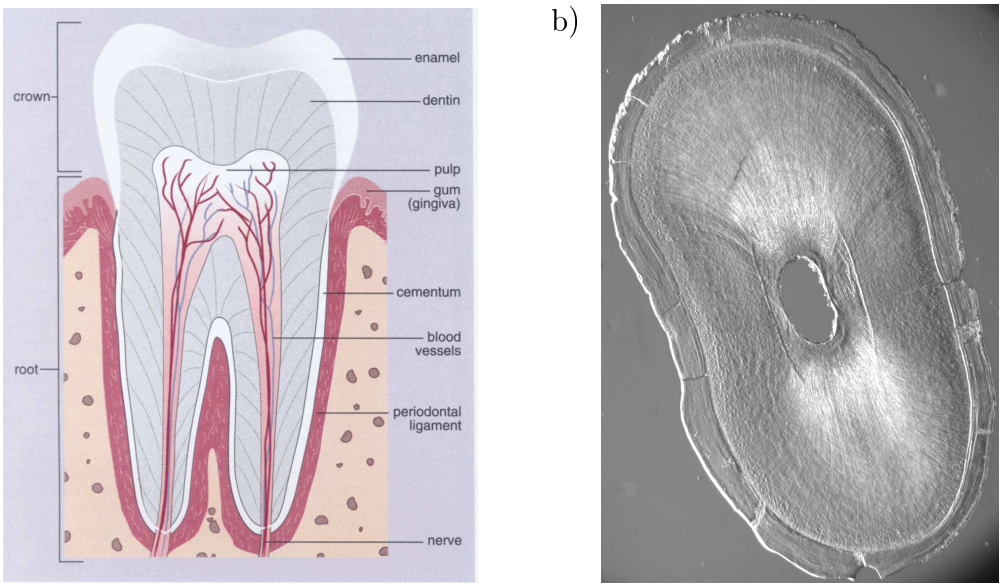


Figure 1.1: (a) Diagram of a vertical cross section of a tooth (*Britannica Concise Encyclopedia* [2003]). (b) Image of a horizontal cross section of a root of a tooth (TCA database of the MPIDR).

It is important to note that no standardized TCA method exists. Different laboratories (and studies) use different preparation and evaluation techniques, which likely influences results. Future research could aim at finding the best standard, as well as identifying the biological mechanisms and influencing factors of TCA. These along with the help of advanced technology (such as increased resolution of computer tomography), would surely improve the TCA method. A semi-automatic procedure to evaluate TCA images will further advance this research.

The Max Planck Institute for Demographic Research (MPIDR), among other things, is interested in reconstructing mortality profiles of past populations. For this purpose it is desirable to apply an accurate age estimation method that produces estimates with small variation even at the highest ages. (Wittwer-Backofen et al. [2004]) Most morphological age estimation methods (Kemkes-Grottenthaler [2002]) produce age ranges of substantial length, especially at old ages, because the individual variability of age-dependent changes in the skeleton increases with age. (Großkopf [1990]) Applying TCA

for human age estimation to past populations could have at least two advantages:

- The incremental lines are suspected to form with increasing chronological age and are not remodeled during life. The number of tooth rings therefore does not measure the biological age of a skeleton (depending on working and living conditions), as morphological age estimation methods do. TCA measures the chronological age and is thus independent from a reference population, except for the tooth eruption age which needs to be added. (Kagerer and Grupe [2001])

- Teeth are usually better preserved over time than bones and are also expected to be found in a larger number. This increases the probability of being able to apply the TCA method when finding only skeletal fragments. (Großkopf [1990])

1.2 The Basel-Project

In the city of Basel, Switzerland, the spital-field St. Johann is located. Hotz [2006] briefly describes this site as well as some past and future research on it. In the graveyard 2,561 skeletons from the preindustrial period from 1845 to 1868 were buried. Of these, 1061 skeletons could be uncovered and for a majority of them the name, origin, occupation as well as age and cause of death are known from the public record office Basel-city. The identified skeletons offer an invaluable source of information to evaluate, and more importantly, to compare existing anthropological age estimation methods. Ten of these methods were applied to 100 selected skeletons of the spital-field St. Johann. Every other skeleton was studied twice. The skeletons and the study, respectively, will here simply be referred to as the Basel series or the Basel-Project.

Tooth cementum annulation is one of the age estimation methods; it was applied by three different laboratories. The tooth laboratory of the MPIDR co-organized the Basel-Project and prepared 495 cross sections (so-called slices, see Figure 1.1 (b)) from 103 teeth which led to 2,120 digital TCA images in total. Table 1.1 lists some of the preparation and evaluation details; and Figure 1.2 displays a flowchart of the selection procedure for this sample. The TCA images are 1016x1300 pixels in size, containing 2^{16} gray values with the exposure time for the digital camera being the only

Preparation and Evaluation Details
tooth types: single-rooted
fixation of crowns for cutting: embedding in epoxy resin
location of slices: middle third of the root
microtome: annular saw Leica SP1600
grinding and polishing: by Lapping and Polishing Machine Logitech PM5
grain size: 10μm for grinding and 1μm for polishing
thickness of slices: 90-110μm
decalcified, etched, stained: no
slices cleaned/dehydrated with: distilled water, ultrasonic bath, alcohol-series, xylen
slices mounted: with Eukitt and coverglas
microscope: Leica DMRXA (bright field transmitted light)
magnification: 20 or 40 times
digital camera: Leica DC350F
TCA image: 1016x1300 pixels in size, 2^{16} gray levels
manual counting procedure: mark on the monitor one point per tooth ring
age estimate: add the tooth eruption age (according to Adler [1967]) to the aggregated counts per slice (observer-specific)

Table 1.1: Preparation and evaluation techniques applied at the MPIDR to acquire TCA images from the Basel series of spital-field St. Johann. (Based on an internal lab protocol. Details will be published in Fabig et al. [2007].)

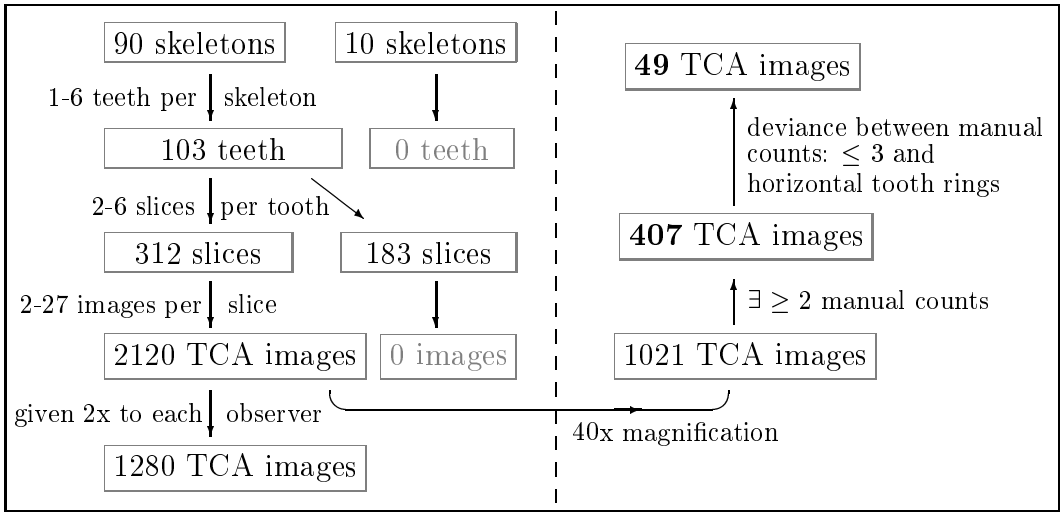


Figure 1.2: Flowchart of acquiring TCA images from 100 skeletons of the Basel series of spital-field St. Johann (on the left) and of the sample selection procedure to evaluate proposed semi-automatic algorithms (on the right).

regulation automatically applied before storing the images. TCA images were acquired, such that tooth rings run in approximately horizontal direction in most images. The dark parts of the annual lines are empirically 1 to 3 μm thick and roughly result in thin lines of 5 to 20 pixels under 40-fold magnification.

Figure 1.3 displays a typical TCA image at 40-fold magnification. It is an image of intermediate quality and was acquired from an unpolished slice which was extracted from a person aged 41. This individual belongs to a pilot study, but not to the Basel-Project itself. Subtracting the tooth eruption age, it is expected that 34 horizontal tooth rings will be found in the marked cementum band. Additionally, the image contains diagonal saw cuts and artifacts (for example, on the right). The TCA image in Figure 1.3 will be the standard exemplary TCA image used throughout this work to illustrate various approaches to TCA image analysis. Some methods will be tested on the whole cementum band, some on the marked square of 500x500 pixels in size. Other methods will be tested on the smaller, inner rectangle of this TCA image or only on the one marked column.

TCA images acquired at the MPIDR were evaluated by two different observers and 1,280 out of the 2,120 TCA images were given twice to each observer. The observers knew which TCA image belonged to each slice and tooth, but did not know the true age of the individuals. The (up to four) observations are independent. This study design allowed, for the first time, an assessment of the inter- as well as intra-observer variance of TCA age estimates; these were presented at the Paleodemography Workshop – Basel Project [2006] and might be published in a monograph Fabig et al. [2007] together with results from the other age estimation methods. The study design also permits an extended comparison of the number of tooth rings estimated by a semi-automatic procedure to the theoretical number of lines (subtracting the tooth eruption age from the known age), as well as to the number of tooth rings from up to four manual observations. For this comparison the estimated number of tooth rings for each TCA image were chosen (and not the aggregated estimates of each slice).

Of the 2,120 TCA images which were acquired, 1,021 had a 40-fold magnification, while the remaining images had a 20-fold magnification. This work will focus on TCA images with 40-fold magnification, because potentially more information on tooth rings is stored in them. Two subsets of these 1,021 TCA images will be selected in order to analyze methods proposed in this work. (Compare to Figure 1.2.) For the first subset, the 407 TCA

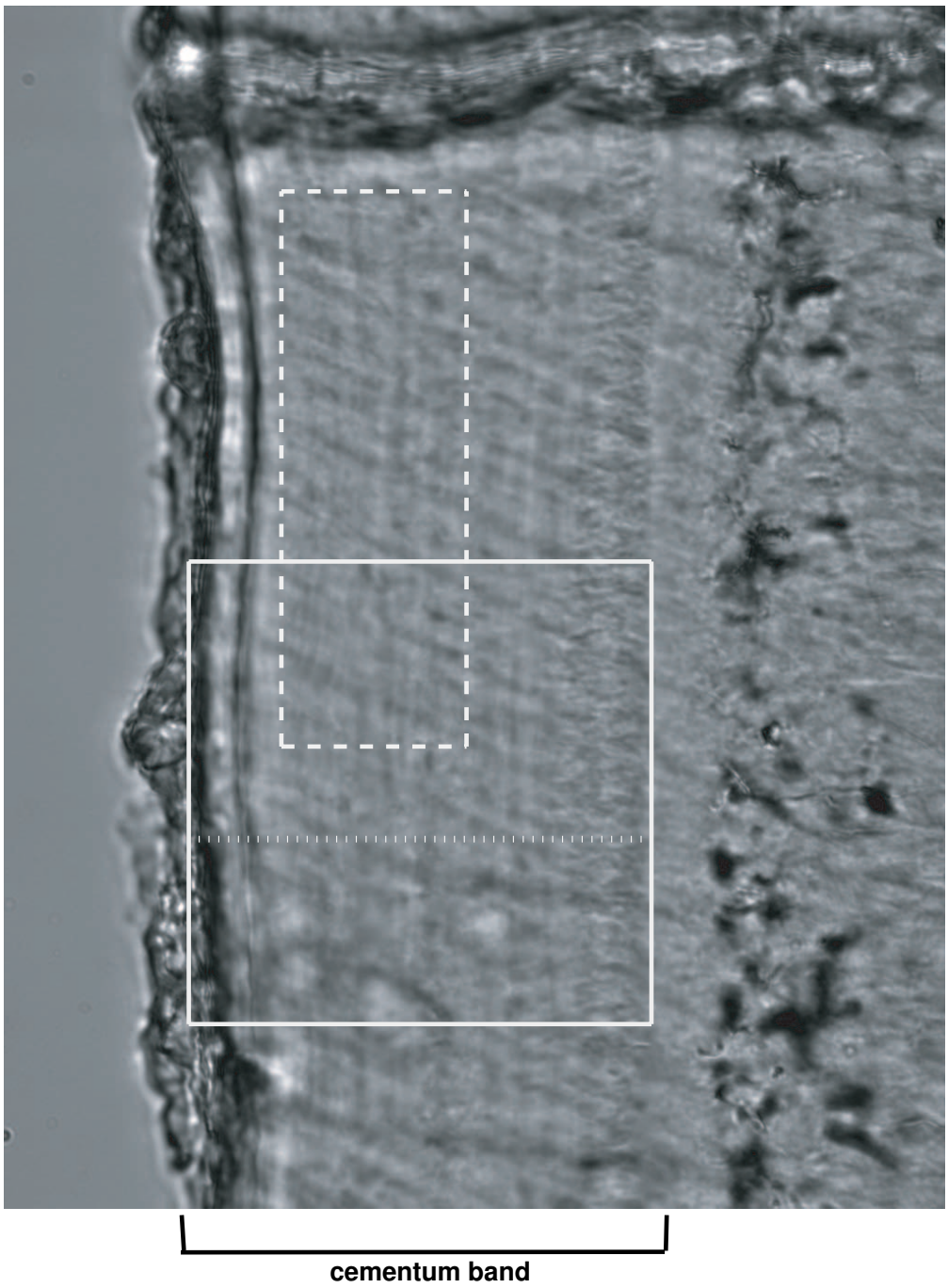


Figure 1.3: A typical unpolished TCA image of intermediate quality (IS-0000666 from the TCA database of the MPIDR). Marked areas highlight regions used for testing.

images were chosen where at least two observations from the observers of the MPIDR resulted in estimates for the number of tooth rings. These are the 'better' TCA images from the observers' perspective. In Chapter 6.1 more details will be explained about the selection of this sample as well as the reference for evaluating a semi-automatic procedure. For an initial evaluation of methods, a smaller subset containing 49 of the 407 TCA images was chosen. For these TCA images, at least two independent manual observations exist with a maximum difference of three years, and tooth rings run horizontally (while tooth rings of the larger subset are allowed to fluctuate more). This small subset will be the basis for comparing the different proposed methods.

1.3 Methods

The present work develops the first flexible model-based procedure for evaluating TCA images. A hidden Markov random field model with a large dependency structure is successfully fitted to large images and evaluated on a (relatively) large set of TCA images.

As an initial step in **Chapter 2**, one standard method from each of the fields of statistics, calculus and algebra is tested for its applicability to TCA image analysis. In particular, local regression, Fourier transform and singular value decomposition are examined. These methods are able to measure certain features of an image, but are unable to bridge missing or falsified information, which frequently occurs in TCA images. They also require the application of heuristics to estimate the number of lines and are too stringent in many cases. Still, the **feature measuring methods** serve to illustrate characteristics of TCA images and to accentuate the distinction between two aspects of TCA image analysis: recognizing tooth rings and estimating their number.

Chapters 3 to 5 develop statistical models tailored to TCA image analysis. In model-based analysis one gains knowledge by describing the observed image by a set of parameters and by generating synthetic images that are similar in certain characteristics. For this purpose, the problem of evaluating TCA images is restated as labeling problem: The observed image is assumed to be a random field and will be defined through the probability distribution of an underlying unknown random field – the label image. The labeling problem is the task of estimating the unknown label image from the

noise-corrupted observed image. Three different kinds of models attempt to solve this problem, one in each of the next few chapters.

The first model in **Chapter 3** is the **mixture model**. The idea behind mixture modeling is that observations arise from separate groups (black and white tooth rings), but neither the individual membership nor the proportions are known. In order to fit a Gaussian mixture model with a known number of groups, the EM algorithm is used to maximize the likelihood. Then the RJ MCMC algorithm is applied to fit the Gaussian mixture model for an unknown number of components in a fully Bayesian framework. Both methods reveal a most likely labeling that does not resemble tooth rings. One reason might have to do with problems of identifiability: There is not necessarily a one-to-one correspondence between the fitted mixture components and the desired groups in an application. Mixture modeling is thus not a well-posed problem for TCA images: The gray value of each pixel alone does not separate the mixtures in the required way. Mixture models are histogram based models, that is, they rely on the assumption of independent observations and do not take into account the relationship between neighboring pixels. But in images, the gray level of one pixel is highly dependent on gray levels of pixels nearby.

Chapters 4 and 5 build the main focus of this work. Two statistical models are introduced that are extensions of the mixture model. They incorporate knowledge about the relation of spatially nearby pixels and thus attempt to imitate the mechanism of bridging information the way manual observers do by knowing that tooth rings all run roughly in the same direction.

The recommended approach to TCA image analysis is established in **Chapter 4**. The independence assumption of mixture modeling is relaxed by the concept of **Hidden Markov Random Field** (HMRF) models, incorporating two-dimensional correlations between gray values. An HMRF consists of two levels, the observable field and the hidden field. These two random fields are linked by the conditional probability of the observed image, given the label image, that is specified by independent normal distributions. The Markov Random Field (MRF) models the hidden field and involves the application-driven contextual constraints. It is first specified by a simple auto-logistic model (Chapter 4.2), involving only pairwise contextual constraints and serving as a starting point for the selection of a more complex MRF model. Simulation results obtained by the Gibbs sampler show that the auto-logistic model can only fit micro-textures, but it does not express the prior convictions about TCA images. More regular textures can be sim-

ulated by extending the dependency structure. Specifically, the FRAME model (detailed in Chapter 4.3) incorporates macro-features of the TCA image into the potential function with the help of filters, and thus accounts for long-range dependence among the observed gray values. The FRAME outperforms the computationally inefficient setting of the auto-logistic model and explains its parameters (filters) more intuitively. Simulated images illustrate the capacity of the FRAME approach and show that this model can describe the periodic placement of tooth rings well by selecting the filter family from the Gaborcosine functions. An EM algorithm that exploits the mean field approximation of the hidden field distribution coherently estimates the parameters of this Gaussian hidden FRAME model and thus avoids time-intensive MCMC methods. Estimates for confidence intervals of these parameters provide assistance in judging the quality of the fit of the model to the TCA image. A good estimate and reasonable confidence interval for the number of tooth rings are obtained by fitting the Gaussian hidden FRAME model to the standard exemplary TCA image. Still, the reconstruction of tooth rings is not satisfactory due to the global assumptions for the hidden field as well as the observable field. To circumvent this shortcoming, an alternative model is examined in Chapter 5. Several alternative specifications of the FRAME model and the algorithmic set-up are examined in Chapter 6.

Chapter 5 sets up a **Coupled Hidden Markov Model** (CHMM) to take into account directed two-dimensional interactions. For one-dimensional data, hidden Markov models may solve the labeling problem by assuming that the hidden label vector has arisen from a Markov chain. Coupling neighboring chains makes this approach suitable for two-dimensional data and leads to the coupled Markov model. Specifying the hidden field of the CHMM by a FRAME model using halved filters will be called FRAME chain. This is a novel model to describe a past relevance structure for macro-textures, while distinguishing between the two different dimensions. Because of the affinity between the Gaussian hidden FRAME chain and the Gaussian hidden FRAME model, the same EM algorithm and mean field approximation are exploited for model fitting and for estimating the number of tooth rings. Simulations for the FRAME chain give similar results as for the FRAME model, and show that this model can also describe the periodic placement of tooth rings well by selecting the filter family from the halved Gaborcosine functions. Due to the thinner filter, the Gaussian hidden FRAME chain produces a more flexible reconstruction along the direction of the tooth rings.

But similar to in Chapter 4, the global assumptions, such as a constant ring width and a fixed orientation throughout the whole label image, lead to a bad detection of the actual tooth rings.

Chapter 6 mainly tests the Gaussian hidden FRAME model and its estimation procedure. The model is successfully fitted to a set of 407 TCA images for the Basel series of spital-field St. Johann and evaluated against the theoretical number of lines as well as compared to the manual observations (Chapter 6.1). Chapter 6.3 and 6.4 test and discuss different specifications of the Gaussian hidden FRAME model and the EM algorithm. They will demonstrate their flexibility and capabilities, as well as compare results to the original model and algorithm. Attempts are made to localize the FRAME model as well as the observable random field, and a diverse set of different issues connected to the EM algorithm will also be discussed. In Chapter 6.2 the smaller subset of 49 TCA images from the Basel series will serve to evaluate the Gaussian hidden FRAME chain model. Both models, the Gaussian hidden FRAME chain as well as the Gaussian hidden FRAME model, achieve competitive – partially even superior – results compared to the estimates of the manual observers on the tested subsets. Chapter 6.5 shows that the Gaussian hidden FRAME model is also applicable to related problems, especially to the yearly rings in cross sections of trees.

Chapter 7 will conclude this work. In general, Chapters 4 to 6 lead to the statement that the hidden FRAME model and also its relative the hidden FRAME chain are flexible tools for TCA image evaluation. Both models achieve age estimates that are (at least) comparable to the manual observations. The hidden FRAME model is also a powerful tool for similar layered structures as well as for other large images containing large neighborhood structures. Only the EM algorithm incorporating mean field approximation enables fitting the model to these applications in a reasonable amount of time.

Besides the statistical methodology, this work relies heavily on computer programmed algorithms. Each of the introduced methods is implemented in Matlab and was run on the compute server Hydra[1] of MPIDR. Neither the chosen programming language nor the computer is the fastest that exists nowadays, but they are convenient choices for fast experimental programming and for extended time-consuming tests respectively.

[1] Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

Chapter 2

Feature Measuring Methods

Throughout this chapter, several well-known image processing methods will be introduced and tested for their applicability to TCA image analysis. In particular, local regression, Fourier transform and singular value decomposition will be examined in this order. These methods are called feature measuring methods here because certain characteristics (features) of an image can be assessed (measured). This chapter is designed to illustrate these and other features of TCA images, and in the context of TCA image analysis the shortcomings of the introduced methods. We will therefore refrain from describing and testing every detail of any possible method. Instead, in the Sections 2.1, 2.2 and 2.3 each method will first be defined and then applied to a part of a TCA image. The Fourier transform and singular value decomposition will in addition be tested on a set of TCA images. This chapter will close with a discussion regarding the tested methods, feature measuring methods in general and characteristics of TCA images.

2.1 Local Regression

Local regression is used to describe the relation between a predictor variable and a response variable, whose functional form is not known in advance. It is a smoothing method with a history of more than 100 years and is applied in many branches of science today. Here, local regression will be applied to TCA images to remove noise by approximating one data point by a weighted average of its neighboring data points. The location's coordinates will serve as the predictor for the response variable TCA image. Section 2.1.1 describes

the method of local regression formally. The details for an application to TCA images will be specified in Section 2.1.2, followed by a discussion on local regression for the analysis of TCA images.

2.1.1 Theory of Local Regression

Local regression is well-described in the literature. Let us follow Loader [1999] closely as a means of providing a foundation of local regression in this subsection.

Let Y be the response and x the predictor variable, consisting of N pairs of observations $(x_1, Y_1), (x_2, Y_2), \ldots, (x_N, Y_N)$. Then local regression assumes the model

$$Y_n = \mu(x_n) + \epsilon_n;$$

where $\mu(x)$ is an unknown function and ϵ_n an error term representing random noise, that is here assumed to be independent, identically distributed with zero mean $E(\epsilon_n) = 0$ and finite variance $E(\epsilon_n^2) < \infty$.

Since the functional form of $\mu(u)$ is unknown, let us approximate it by the polynomial

$$p_x(u) = a_0 + a_1(u - x) + \frac{1}{2}(u - x)^2 + \ldots$$

for the fitting point x and within a smoothing window: $u \in]x-h(x), x+h(x)[$, where $h(x)$ defines the bandwidth. The observations within this window are weighted according to a weight function $W(u)$ by

$$w_n(x) = W\left(\frac{x_n - x}{h(x)}\right),$$

and the coefficients $a_0, a_1, \ldots$ of the polynomial can be estimated using least squares.

Definition 2.1.1. *The* local regression estimate *is the estimate of the first component of the polynomial p at $u = x$:*

$$\widehat{\mu}(x) = \widehat{a_0}.$$

The fitted curve $\hat{\mu}(x)$ describes the data Y through the variable x.

For applying local regression, the following components need to be specified:

(1) the bandwidth $h(x)$

The bandwidth has a critical effect on the local regression fit, known as bias-variance trade-off. A large bandwidth may produce a bad fit within the smoothing window and important features of the mean function $\mu(x)$ may be distorted (large bias). A small smoothing window, on the other hand, may result in a noisy fit (large variance).

The choice might vary from a different bandwidth for each fitting point to a constant bandwidth for all x.

(2) the degree of the polynomial p

Similar to the bandwidth, the polynomial degree used to approximate the functional relationship between response and predictor variable also affects the bias-variance trade-off: A high order polynomial will in general lead to an estimate with less bias and more variability, while a low polynomial degree provides a smoother regression curve with more bias.

The most common choices for the polynomial degree are 1 or 2, resulting in local linear and local quadratic regression respectively.

(3) the weight function W

The weight function has less effect on the bias-variance trade-off. It is usually chosen to be continuous, symmetric, peaked at 0 and supported on $[-1, 1]$. The most common choice is the tricube weight function

$$W(u) = \left(1 - |u|^3\right)^3 .$$

A more detailed description, as well as extensions and diagnostic tools for the local regression method, can be found, for example, in Loader [1999].

2.1.2 Application and Discussion

Let us now apply local regression smoothing to TCA images in order to smooth gray values dependent on their location. In order to estimate the number of lines in a TCA image, one can, for example, smooth perpendicularly to tooth rings to obtain a fitted curve with clearly identifiable minima, where minima correspond to points on a black tooth ring.

For simplicity, let the responses be the columns of the TCA image $Y(x, y_m)$ $(m = 1, \ldots, M)$ and let the predictor variable be their rows x_n $(n = 1, \ldots, N)$. Now the components (1)-(3) of local regression (see page 15) need to be specified for the application to TCA images.

TCA images contain periodic wavelike structures that are approximately vertical in direction. Let us therefore use a local quadratic regression with the tricube weight function. During the introduction, it was mentioned that tooth rings are approximately 5 to 20 pixels thick; let us therefore choose the smoothing window to be of a similar size. In principle, the bandwidth could be chosen flexibly depending on the data (according to some criterion). But for illustration purposes, constant bandwidths $h(x) \equiv 6$ and $h(x) \equiv 7$ are used for now and depicted in Figure 2.1 for one column $(y = 400)$ of the standard TCA image from Figure 1.3.

The example clearly shows that the choice of bandwidth directly influences the number of recognized minima. A deviance of only one or two pixels in the bandwidth has a fundamental impact on the estimated number of lines, which causes two principal problems. In the first place, the optimal bandwidth might not be an integer; but discrete images allow only for integer bandwidths. So the critical decision of the bandwidth might always lead to wrong estimates and therefore to an unstable procedure. Secondly, the choice of the optimal bandwidth leads directly to the question of which features of the TCA image are 'real'. This question is intrinsically related to the problem that the best fit of the regression curve does not solely depend on the data, but also on the question of interest. Since the quantity tooth ring is not clearly defined, no diagnostic tool can provide an answer to the question of the optimal bandwidth and whether a specific minimum belongs to noise or to a tooth ring.

One could extend regression smoothing to the two dimensions of an image, leading to a multivariate local regression. But the two principal problems above will remain and are the reason why we refrain from local regression for smoothing TCA images.

2.2 Discrete Fourier Transform

In the introduction it was indicated that TCA images contain periodic structures in a roughly vertical direction. In order to emphasize periodic structures belonging to tooth rings while neglecting the ones belonging to noise,

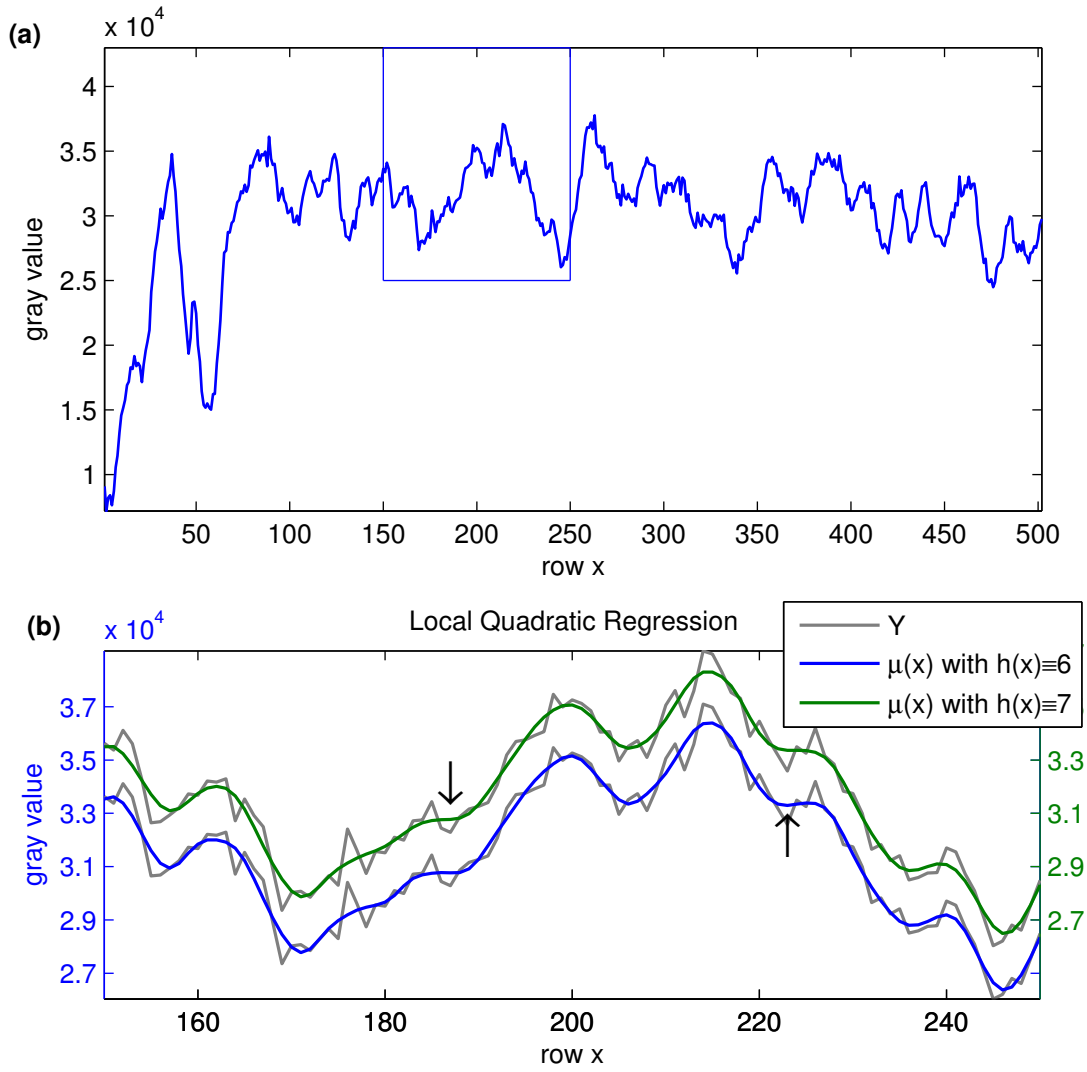


Figure 2.1: Graph (a) shows one column of the standard exemplary TCA image marked in Figure 1.3. The local quadratic regression estimate with bandwidth $h = 6$ (in blue) and $h = 7$ (in green) for the rows 150 to 250 are depicted in (b) simultaneously, showing a difference of two minima.

the Fourier transform can be used. The Fourier transform is equipped with an elegant theory and is a widespread analytical tool. It maps functions from time to frequency domain and therefore allows us to dismantle periodic components of interest.

Because of the discreteness of digital images, one is forced to use the discrete approximation of the Fourier transform, the DFT. Its theory in one dimension will be described in Section 2.2.1, the generalization to two dimensions in Section 2.2.2. Section 2.2.3 will illustrate the application of DFT for the enhancement of a TCA image and summarize results from applying DFT for estimating the number of lines in a series of TCA images. A discussion of these results will bring the section to an end.

2.2.1 Theory of the Discrete Fourier Transform

The discrete as well as the continuous Fourier transform are detailed in many textbooks. To explain the theory of the one-dimensional Discrete Fourier Transform (DFT), Jain [1989] will be followed.

Let Y be the sequence of interest, then one can define its Fourier transform as follows:

Definition 2.2.1. *The* N-point DFT *of a vector* $Y = (Y_1, Y_2, \ldots, Y_N)$ *is*

$$\mathbb{F}_{k+1} = \sum_{n=0}^{N-1} Y_{n+1} \left(e^{\frac{-2\pi i}{N}} \right)^{nk},$$

with frequencies $k+1 = 1, 2 \ldots, N$ *and amplitudes* $\frac{|\mathbb{F}_{k+1}|}{N}$.

The DFT has the following central properties:

(1) *Inversion*: The transformation

$$Y_{n+1} = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{F}_{k+1} \left(e^{\frac{2\pi i}{N}} \right)^{nk}$$

for $n = 0, 1, \ldots, N-1$ is called N-point inverse DFT of the vector $\mathbb{F} = (\mathbb{F}_1, \mathbb{F}_2, \ldots, \mathbb{F}_N)$.

(2) *Periodicity*: $\mathbb{F}_k = \mathbb{F}_{k+N}$ for all integers $k \in \mathbb{N}$

(3) *Conjugate symmetry*: if $Y \in \mathbb{R}^N$ then $\overline{\mathbb{F}}_{N-k+2} = \mathbb{F}_k$
The frequencies $\lceil \frac{N}{2} \rceil + 1, \lceil \frac{N}{2} \rceil + 2, \ldots, N$ are hence called negative frequencies.

(4) *Fast* implementation: There exists an implementation of the DFT, called Fast Fourier Transform (FFT), which requires only $O(N \log_2 N)$ operations.

The DFT (and also the inverse DFT) can be expressed with the help of sine and cosine functions

$$\mathbb{F}_{k+1} = \sum_{n=0}^{N-1} Y_{n+1} \left(\cos\left(\frac{-2\pi nk}{N}\right) + i \sin\left(\frac{-2\pi nk}{N}\right) \right),$$

by using Euler's identity $e^{\frac{2\pi il}{N}} = \cos\left(\frac{2\pi l}{N}\right) + i \sin\left(\frac{2\pi l}{N}\right)$. One can thus interpret the DFT as a conversion of a vector depending on time into a set of amplitudes belonging to different frequencies (of sine and cosine waves).

In order to enhance periodic components of interest in the vector Y, the amplitudes of certain frequencies of $\mathbb{F}$ can be emphasized while others can be dampened; followed by an inverse DFT. This procedure will be detailed for the two-dimensional DFT and the application to images in the subsequent sections.

2.2.2 Theory of the 2-D Discrete Fourier Transform

The one-dimensional discrete Fourier transform is naturally ill-suited for two-dimensional data such as images. But it can be elegantly generalized to two dimensions by repeatedly applying it. The two-dimensional discrete Fourier transform (2-D DFT) is well-known and in image processing especially used for enhancement and fast convolutions. It can be found in many textbooks and will be explained here, again drawing from Jain [1989].

Let Y be the matrix of interest, then one can define its 2-D Fourier transform as follows:

Definition 2.2.2. *The* 2-D N-by-M DFT *of a matrix* $Y = (Y_{n,m})$ *with* $n = 1, \ldots, N$ *and* $m = 1, \ldots, M$ *is defined by*

$$\mathbb{F}_{k+1,l+1} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y_{n+1,m+1} \left(e^{-2\pi i \left(\frac{nk}{N} + \frac{ml}{M}\right)} \right),$$

with vertical and horizontal frequencies $k+1 = 1, 2 \ldots, N$ *respectively* $l+1 = 1, 2 \ldots, M$ *and amplitudes* $\frac{|\mathbb{F}_{k+1,l+1}|}{NM}$.

The 2-D DFT has the following appealing properties:

- *Separability*: The 2-D N-by-M DFT can be rewritten as

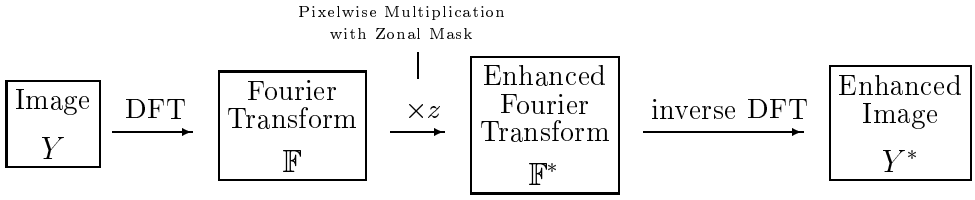
$$\mathbb{F}_{k+1,l+1} = \sum_{m=0}^{M-1} \left(\sum_{n=0}^{N-1} Y_{n+1,m+1} \left(e^{\frac{-2\pi i}{N}} \right)^{nk} \right) \left(e^{\frac{-2\pi i}{M}} \right)^{ml}.$$

  One can therefore apply the one-dimensional DFT first on the columns of Y, and then apply the one-dimensional DFT again on the resulting rows of the Fourier transform, giving the 2-D DFT $\mathbb{F}$.

- The properties (1) to (4) of the one-dimensional DFT above hold analogously for the 2-D DFT.

Using Euler's identity, the 2-D DFT can be interpreted as a conversion of a matrix depending on space into a set of amplitudes belonging to different frequencies in the two dimensions.

In order to enhance periodic components of interest, one can apply a so-called zonal mask z on the Fourier transform using pointwise multiplication; followed by the inverse Fourier transform:



This procedure is called generalized linear filtering. An exemplary scheme of a zonal mask for low-pass, band-pass and high-pass filtering is depicted in Figure 2.2; while low-pass filtering selects low frequencies, high-pass filtering selects high frequencies, and band-pass filtering selects a certain area of frequencies. A specific zonal mask for application to TCA images will be selected in the upcoming section.

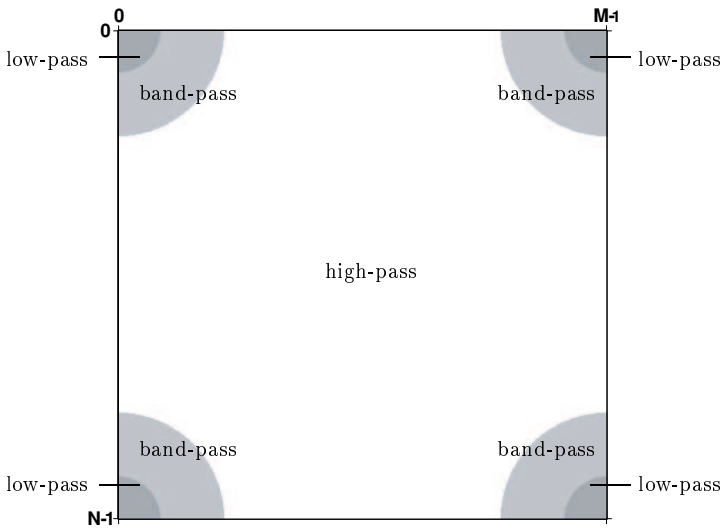


Figure 2.2: Exemplary scheme of a zonal mask z for Fourier transforms (from Jain [1989]).

2.2.3 Application

In this section, TCA images will be enhanced by emphasizing periodic structures belonging to tooth rings, while simultaneously neglecting structures belonging to noise. This can be done, using generalized linear filtering based on the two-dimensional Fourier transform, as described in the previous section.

To apply generalized linear filtering, one needs to define a zonal mask z for the Fourier transform. The selection of this mask is empirical and depends on the application at hand. For TCA image analysis, a band-pass filter will be chosen such that frequencies corresponding to the possible number of rings are selected. In particular, during the introduction it was mentioned that tooth rings are approximately 5 to 20 pixels thick. Let us therefore assume that the vertical frequency $\frac{N}{40} + \left(\frac{N}{10} - \frac{N}{40}\right)/2 = \frac{5N}{80}$ is the likeliest, where N is the width of the cementum band. The zonal mask z should select similar frequencies in vertical direction as well as similar orientations. It will

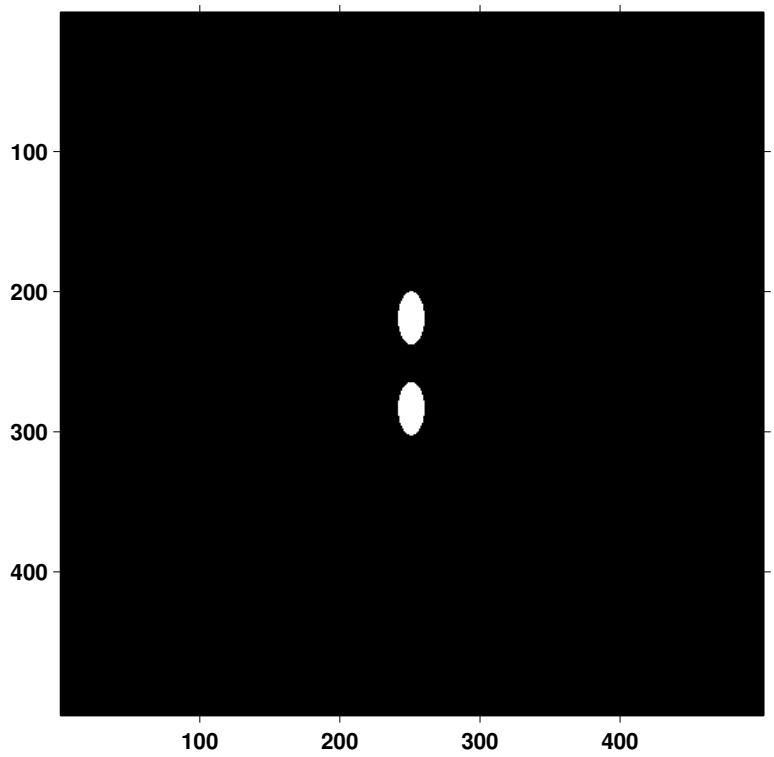


Figure 2.3: Shifted zonal mask for the Fourier transform of part of the standard exemplary TCA image from Figure 1.3.

thus be defined as a solid ellipse centered around the likeliest frequency with diameters $\left(\frac{N}{10} - \frac{N}{40}\right)$ and $\frac{1}{2}\left(\frac{N}{10} - \frac{N}{40}\right)$:

$$z(k+1, l+1) = \begin{cases} 1 & \text{if } \frac{\left(k-\frac{5N}{80}\right)^2}{\frac{1}{4}\left(\frac{N}{10}-\frac{N}{40}\right)^2} + \frac{\left(l-\frac{5N}{80}\right)^2}{\frac{1}{16}\left(\frac{N}{10}-\frac{N}{40}\right)^2} \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

for the upper quarter $k+1 = 1, 2, \ldots, \frac{N}{2}$, $l+1 = 1, 2, \ldots, \frac{M}{2}$ and analogously for other frequencies using the conjugate symmetry property (3). The zonal mask for the DFT of the square part of the standard exemplary TCA image marked in Figure 1.3 is depicted in Figure 2.3. For better visualization, the shifted version of this mask is shown, meaning that the zero frequency is shifted to the center of the spectrum.

The log-amplitudes of the Fourier transform for the standard exemplary TCA image are depicted in Figure 2.4. One can clearly see the very low

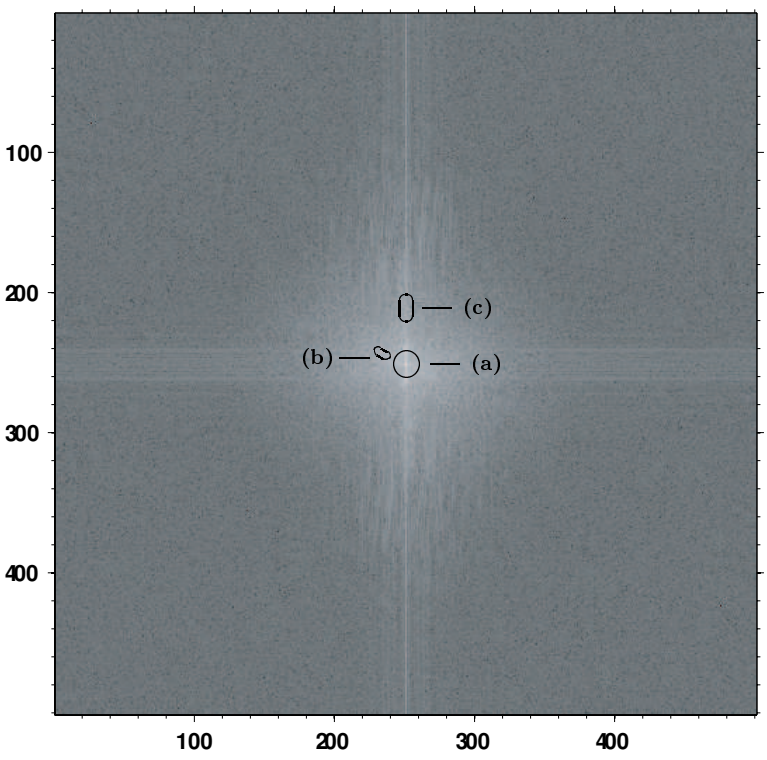


Figure 2.4: Shifted Log-amplitudes of the Fourier transform of the square marked in the standard exemplary TCA image of Figure 1.3. (a) indicates very low frequencies, (b) diagonal frequencies of saw cuts and (c) low frequencies belonging to tooth rings.

frequencies (a) corresponding to, for example, background illumination, the diagonal frequencies (b) corresponding to the saw cuts and vaguely distinguish the frequencies of interest (c) belonging to tooth rings (due to the low resolution of this print). High frequencies are generally attributed to noise.

Multiplying the zonal mask z pointwise with the Fourier transform of the squared part of the exemplary TCA image and inverse transforming the result gives the enhanced (or smoothed) TCA image shown in Figure 2.5.

A natural method for evaluating this smoothed TCA image is the search for the most frequent number of minima over the columns, where minima correspond to points on a black ring. A histogram plot of the number of minima for our example is shown in Figure 2.6, estimating 28 tooth rings.

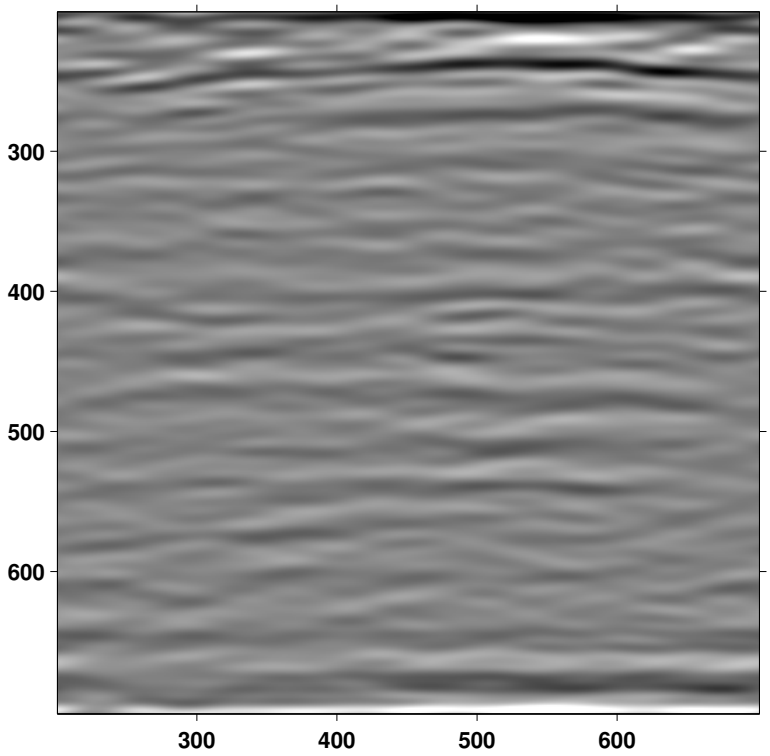


Figure 2.5: Smoothed TCA image after applying the zonal mask z from Figure 2.3 pointwise to the Fourier transform of a square part of the exemplary TCA image (see Figure 2.4) and inverse transforming the result.

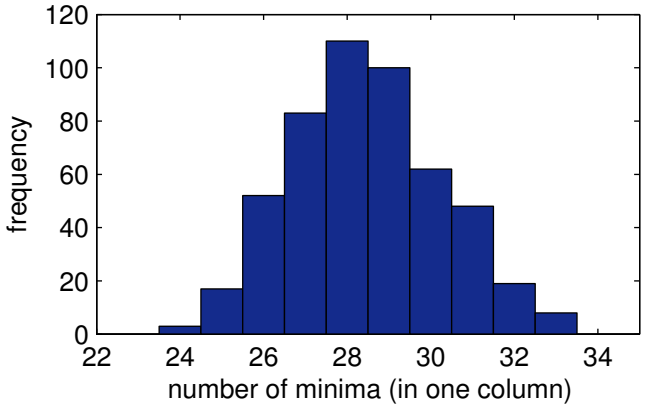


Figure 2.6: Histogram of the number of minima in the columns of the smoothed exemplary TCA image from Figure 2.5. The maximum gives the estimated number of tooth rings: 28.

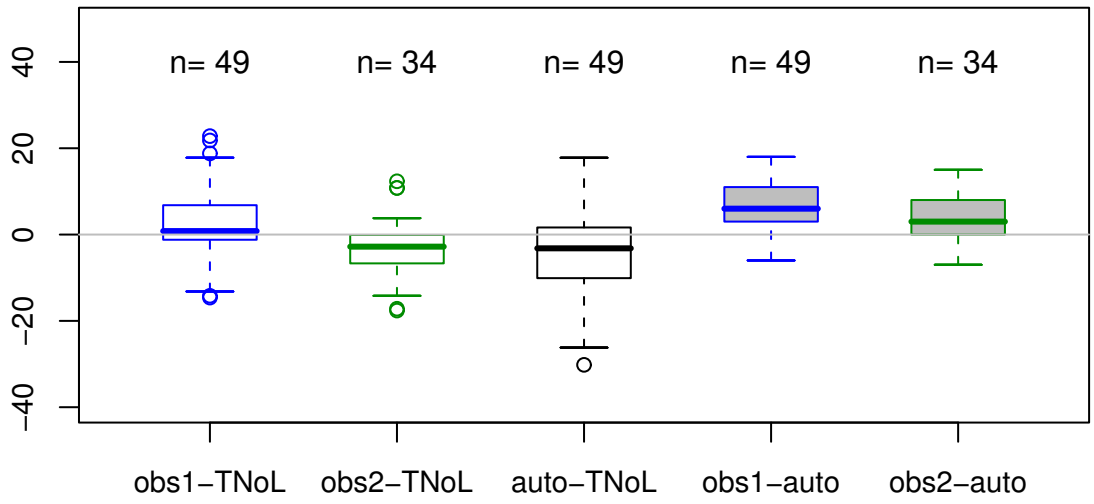


Figure 2.7: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines (TNoL) for the Fourier-based procedure ('auto') and for the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') on 49 TCA images of good quality from the Basel series.

A very similar approach for evaluating TCA images with the help of the Fourier transform was suggested independently in Czermak et al. [2006].

This procedure of generalized linear filtering TCA images and estimating the number of tooth rings by the most frequent number of minima over the columns, was applied to a series of 49 TCA images from the Basel series. For these TCA images, at least two independent manual observations exist with a maximum difference of three years and the tooth rings run roughly horizontal. The 49 TCA images therefore present a subset of the 407 TCA images of good quality from the manual observers' point of view. Figure 2.7 shows box-and-whiskers plots of the difference between the estimates for the number of tooth rings of the first (of possibly two) observation(s) of observer one, observer two and the Fourier-based procedure compared to the theoretical number of lines (TNoL), as well as the difference between the observers and the Fourier-based procedure.

On average the number of lines is underestimated by 3.8 years for the Fourier-based procedure, while observer one overestimates by 2.5 and observer two underestimates by 2.9 years for this sample. The Fourier-based procedure also shows a larger variance in the difference between its estimates

and the theoretical number of lines compared to the estimates of the manual observers.

2.2.4 Discussion

The aim of enhancing TCA images using the discrete Fourier transform was to emphasize frequencies corresponding to tooth rings. For this purpose, the assumption that tooth rings are approximately 5 to 20 pixels thick was used. Frequencies corresponding to these possible numbers of rings in approximately vertical direction were selected from the Fourier transformer and back-transformed into the space domain. In order to estimate the number of lines in a part of a Fourier smoothed TCA image, the most frequent number of minima over all columns was selected.

This procedure seems to work for the selected TCA images from the Basel series, but does not give results that are comparable to the manual TCA age estimates. This might be due to several critical assumptions and steps in the procedure:

- The selected part of the TCA image must be of optimal quality.

  It must contain all tooth rings and no large disturbances like artifacts. The shape of this part is limited to being a rectangle, and in the current application it is even limited to being a square (due to the fixed ratio between the radii of the ellipse in the zonal mask).

- The procedure heavily relies on the assumption that tooth rings are approximately 5 to 20 pixels thick.

  If the number of tooth rings is outside the range $\frac{5N}{80} \pm \frac{1}{2}\left(\frac{N}{10} - \frac{N}{40}\right)$ (where N is the number of rows in the part of interest of the TCA image), the estimate for the number of lines will be biased.

- The orientation of tooth rings must be horizontal.

  The elliptical zonal mask covers frequencies deviating at most by 15-20 degrees from being perfectly horizontal. Tooth rings with a steeper orientation will therefore be oversmoothed and valuable information will be lost. Thin tooth rings are allowed to fluctuate even less. The approach for a column-wise search of minima in the smoothed TCA image will be correct only for strictly horizontal tooth rings.

- For the histogram-based approach of searching the final estimate for the number of lines among the minima in each column, the distribution of the number of minima must be decent.

  Theoretically, this approach works only if all tooth rings are present in most of the columns in the smoothed TCA image. A counter-example would be a multi-modal distribution resulting, for instance, from tooth rings disappearing in a half of the part of the TCA image.

Generally it can be concluded that the method presented here suffers from heuristics, such as the choice of the zonal mask or the final estimation of the number of lines, whose specifications are essential for applying the Fourier transform to TCA images. But even if one succeeds in circumventing these heuristics (for example, by adaptively choosing the zonal mask and number of lines), deficiencies that are intrinsic to the DFT will remain. These are:

- Fourier smoothing is a global technique.

  Cutting out a frequency removes rings of a certain size from the whole image. Locally sensible smoothing is not possible. Thin rings cannot be kept in one part of the image while removed in another part. This is most likely the reason for underestimating the number of tooth rings in the above application, removing frequencies that are too high for some parts of the image.

  It is also impossible to Fourier smooth whole TCA images, including parts where no rings exist like the dentin or widely disturbed parts like large artifacts.

- Fourier transform is a directional method.

  Because of the globality of Fourier smoothing, cutting out an orientation removes it from the whole image.

- Fourier frequencies cannot be translated to the age of a person.

  Because of superposition of sine and cosine waves, one frequency of the Fourier transform cannot be directly interpreted into one ring width or a certain number of rings. Selected frequencies therefore do not have meaning for TCA image analysis.

- A more sophisticated procedure is needed to evaluate the Fourier smoothed image.

The Fourier transform can be used to smooth the image, but can not directly be translated into an estimate for the number of tooth rings (see above). It is not obvious how the smoothed image can be evaluated without introducing additional heuristics.

- Fourier coefficients depend on the size of the image.

  The size of the image translates frequencies into wavelengths. Because of the discreteness of the Fourier transform, rings that do not correspond exactly to an integer-valued frequency might spread out into neighboring frequencies (they "smear"). Especially for large wavelengths in high resolution images, portions of the rings might be removed as noise. On the other hand, low resolution images do not contain enough information for each of the rings. (This can be visualized by quantizing a sine wave by three points.)

We therefore conclude that the discrete Fourier transform is not (directly) suitable for the analysis of all types of TCA images. It could still be incorporated to quickly preprocess TCA images (for example, to remove large-scale noise) or to quickly receive a rough first estimate of the number of rings.

2.3 Singular Value Decomposition

Singular Value Decomposition (SVD) is a fundamental algebraic transform of matrices that is part of many different algorithms. A digital image is a matrix of gray values; SVD can therefore easily be applied in digital image processing. It usually utilizes image restoration or data compression.

The theory of singular value decomposition will be introduced in Section 2.3.1. Section 2.3.2 attempts to interpret the ingredients of the decomposition in terms of TCA image features and illustrates an application of the method to TCA images, as well as summarizes results from applying SVD for estimating the number of lines in a series of TCA images. A discussion of these results will conclude the section.

2.3.1 Theory of SVD

Singular value decomposition is detailed in many textbooks. For more theoretical aspects Trefethen and Bau [1997] will be used here, while Jain [1989] provides an overview tailored to image processing.

Let Y be the matrix of interest, then its singular value decomposition can be defined and its existence proven:

Theorem 2.3.1. *For integers N and M, and any matrix $Y \in \mathbb{R}^{NxM}$ there exists a factorization*

$$Y = UDV^T$$

called singular value decomposition, where

- *$U \in \mathbb{R}^{NxN}$ and $V \in \mathbb{R}^{MxM}$ are orthogonal matrices containing the unit length left respectively right singular vectors of Y, and*
- *$D \in \mathbb{R}^{NxM}$ is a pseudo diagonal matrix containing the singular values $d_j \geq 0$ as entries for $j = 1, \ldots, \min\{N, M\}$.*

A proof of the above theorem and definitions for algebraic terms like orthogonality can be found in any linear algebra textbook, for example, in Trefethen and Bau [1997].

The SVD of matrix Y is a separable transform that diagonalizes the matrix. Matrix Y can then be written as the linear combination

$$Y = \sum_{j=1}^{\text{rank}(Y)} d_j u_j v_j^T,$$

where u_j and v_j are the columns of U respectively V.

For convenience, let the singular values d_j of matrix Y (and its corresponding singular vectors) be arranged in non-increasing order: $d_1 \geq d_2 \geq \ldots$ Then the SVD of Y has the following central property:

- The matrix Y_k generated by the partial sum

$$Y_k = \sum_{j=1}^{k} d_j u_j v_j^T, \text{ for some } k \leq \text{rank}(Y) \tag{2.1}$$

  is the best least squares approximation of Y with rank k, making an error of $\epsilon_k^2 = \sum_{j=k+1}^{\text{rank}(Y)} d_j^2$ (using the squared Frobenius norm).

This means, that among all linear combinations of rank-1 matrices,

$$\sum_{j=1}^{k} w_j a_j b_j^T$$

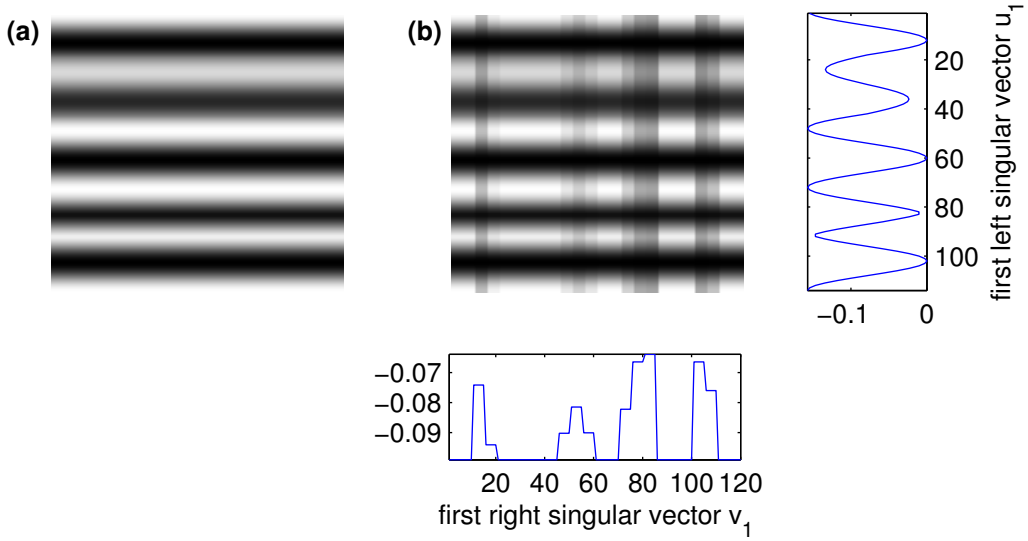


Figure 2.8: Ideal TCA images of rank one. Image (a) shows perfectly horizontal tooth rings, each line with its own intensity; (b) displays the same ideal TCA image with vertical noise (each column with its own intensity). The graphs show the first left and right singular vectors.

where a_j and b_j are sequences of unit length orthogonal vectors; the SVD in (2.1) is the one maximizing the energy in the coefficients w_j for $j = 1, \ldots, k$.

This property can be exploited to approximate images, as will be illustrated for TCA image analysis in the next Section.

2.3.2 Interpretation and Application

Throughout this section, TCA images will be approximated by the best rank-k matrix. First, it will be illustrated that the approximation of rank one theoretically carries a substantial amount of information on tooth rings. Then additional ranks are considered.

Let us imagine the ideal TCA image of perfectly horizontal tooth rings, each ring with its own intensity profile, as depicted in Figure 2.8 (a). This is a rank-1 matrix, where every column is identical to any other column. The rank-1 approximation

$$Y_1 = d_1 u_1 v_1^T$$

obtained by SVD is thus identical to the original matrix. The assumption of identical columns can even be relaxed, such that each column is a multiple of any other column, as depicted in Figure 2.8 (b). This is still a rank-1 matrix that is identical to the partial sum Y_1.

These ideal TCA images illustrate well the interpretation of the first left and right singular vector, respectively, for TCA image analysis:

- The first left singular vector u_1 is a function of the rows where the ith element approximates the ith row of Y best in the least squares sense (up to the multiple $d_1 v_{i1}$). In TCA images, this means that the first left singular vector averages out the variance along the horizontal tooth rings, and is thus a smoothed cross section, where minima correspond to tooth rings.

- The first right singular vector v_1 is a function of the columns where the jth element approximates the jth column of Y best in the least squares sense (up to the multiple $d_1 u_{j1}$). It averages vertically, such that the first right singular vector of TCA images approximates the variance within (horizontal) tooth rings. For example, vertical saw cuts can be captured by this vector.

In general, it can be said that the partial sum of rank-1 Y_1 approximates a matrix depending on the two dimensions of space by two independent functions, depending on one dimension each. Horizontal structures result in peaks in the first left, whereas vertical structures result in peaks in the first right singular vector. The singular values indicate the relative error committed by the rank-1 approximation:

$$\frac{\epsilon_1}{\epsilon_0} = \sqrt{\frac{\sum_{j=2}^{\operatorname{rank}(Y)} d_j^2}{\sum_{j=1}^{\operatorname{rank}(Y)} d_j^2}}$$

The graphs in Figure 2.8 depict the first singular vectors of the ideal TCA image from Figure 2.8 (b). The left singular vector captures the simulated tooth rings while the right singular vector portrays the simulated saw cuts.

Figure 2.9 shows the rank-1 approximation of the squared part of the standard exemplary TCA image marked in Figure 1.3 together with its decomposition into first singular vectors. The number of minima of the first

left singular vector is an estimate of the number of tooth rings; in this example, the number of minima is 28. This rank-1 approximation commits a relative error of 14% and therefore supports our hypothesis that the rank-1 approximation of TCA images captures a substantial amount of information on tooth rings.

But the underestimation of the number of lines indicates that a rank-1 approximation smoothes too intensely in the horizontal direction and is therefore not sufficient. The selection of a higher rank to approximate Y is empirical and depends on the application at hand. For illustration purposes, let us limit ranks from two to ten. The frequencies of the number of minima in each column for these rank-k approximations are depicted in Figure 2.10 for the squared part of the standard exemplary TCA image marked in Figure 1.3. The rank-2 approximation (in green) shows two clear maxima in the distribution of the number of minima at 29 and 37. The higher rank approximations rather emphasize a peak at 30 in this application. The rank-k approximations of the exemplary TCA image thus do not give a unified result for the number of tooth rings. For robustness, the sum of the frequencies over the rank-2 to rank-10 approximations could be used, as depicted in red in Figure 2.10, estimating 30 lines.

This procedure of approximating parts of TCA images by rank-2 to rank-10 matrices and estimating the number of tooth rings by the most frequent number of minima over the columns in these approximations, was applied to a series of 49 TCA images from the Basel series. For these TCA images at least two independent manual observations exist with a maximum difference of three years and the tooth rings run roughly horizontal. The 49 TCA images therefore present a subset of the 407 TCA images of good quality from the manual observers' point of view. Figure 2.11 shows box-and-whiskers plots of the difference between the estimates for the number of tooth rings of the first (of possibly two) observation(s) of observer one, observer two and the SVD-based procedure compared to the theoretical number of lines (TNoL), as well as the difference between the observers and the SVD-based procedure.

On average, the number of lines is slightly underestimated by 0.6 years for the SVD-based procedure, while observer one overestimates by 2.5 and observer two underestimates by 2.9 years for this sample. But the SVD-based procedure produces a larger variance and a larger range in the difference between its estimates and the theoretical number of lines compared to estimates of the manual observers. This variance and range is even larger than for the Fourier-based procedure.

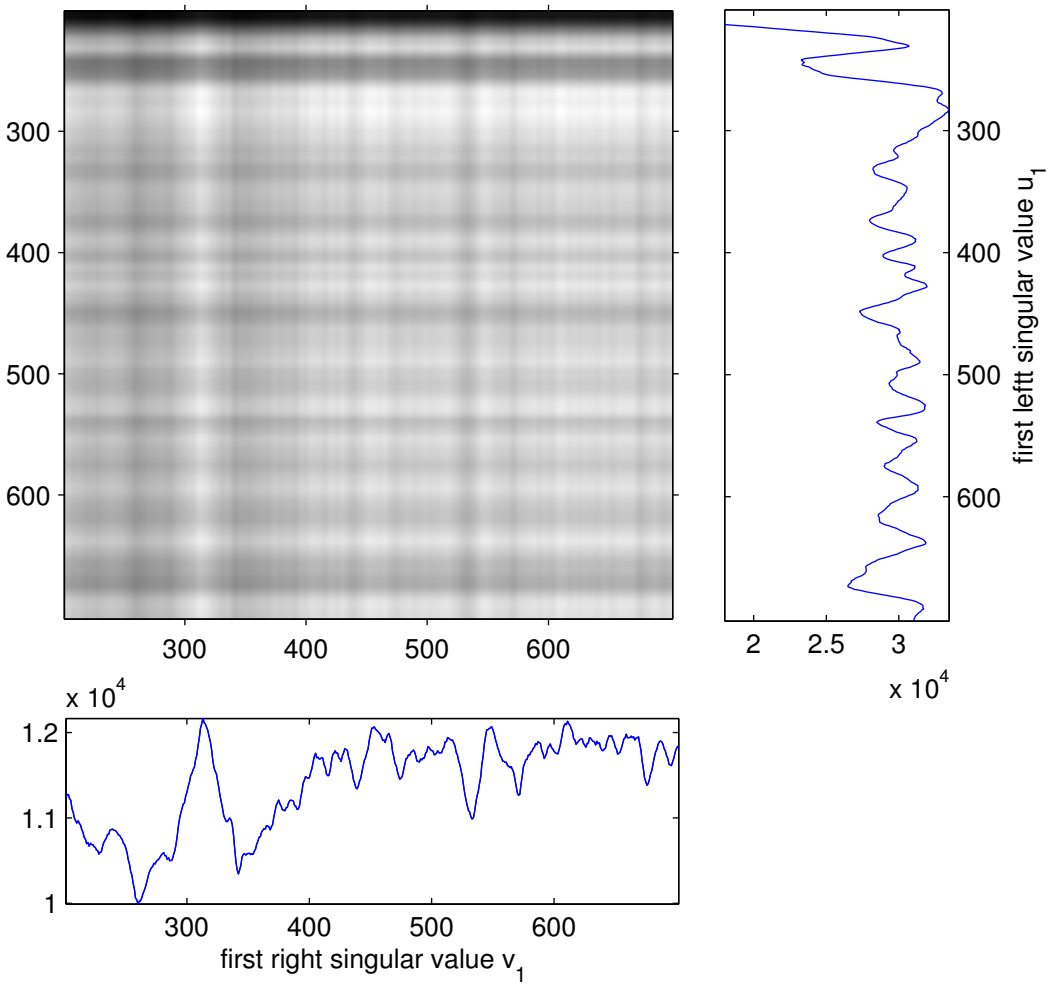


Figure 2.9: Rank-1 approximation of the squared part of the standard exemplary TCA image marked in Figure 1.3. The graphs show the first left and right singular vectors.

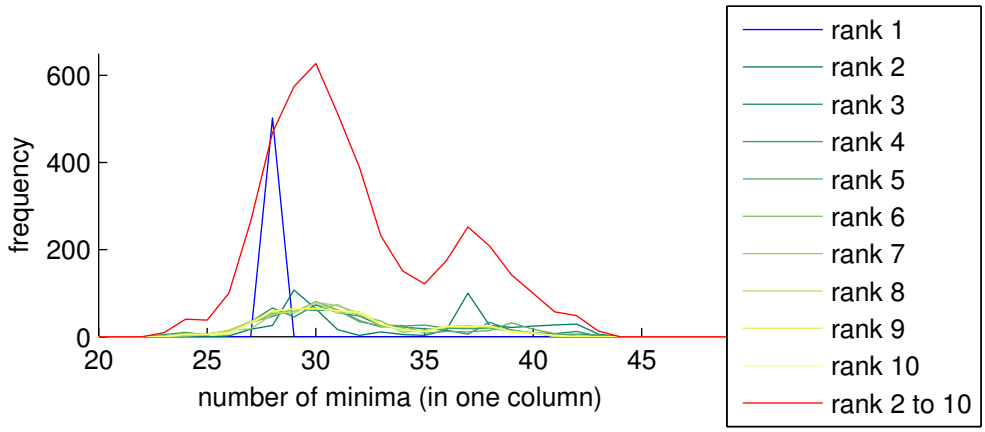


Figure 2.10: Histograms of the number of minima in the columns of the smoothed exemplary TCA image of rank one (blue) to ten (yellow). The maximum of the sum over the rank-2 to rank-10 approximations (in red) gives the estimated number of tooth rings: 30.

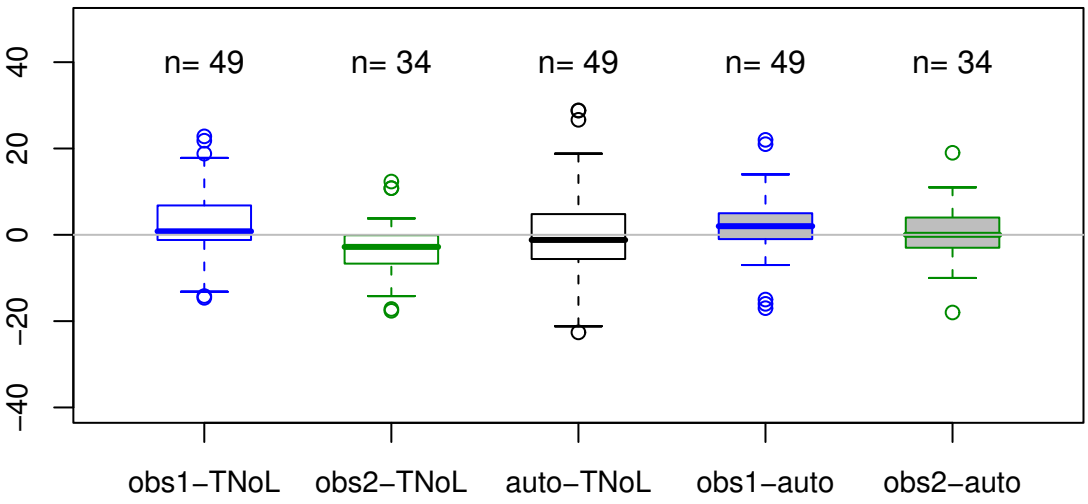


Figure 2.11: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines (TNoL) for the SVD-based procedure ('auto') and for the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') on a set of 49 TCA images of good quality from the Basel series.

2.3.3 Discussion

The aim of this section has been to approximate TCA images by one or several matrices of low rank, and in turn to enhance tooth rings and estimate their number. The rank-1 approximation of TCA images and its first left singular vector theoretically capture horizontal tooth rings. But the application showed that it smoothes too intensely in this direction. Higher rank approximations of TCA images have been evaluated by averaging the frequencies of the number of minima in the columns over the approximations of ranks two to ten.

One can certainly improve this final estimation of the number of minima in TCA images. But similar to the Fourier transform for TCA images, the method suffers from heuristics, such as the choice of the rank of the approximation or final estimation of the number of lines, whose specification is essential for applying SVD to TCA images. But even if one succeeds in circumventing these heuristics (for example, by adaptively choosing the rank and the number of lines), deficiencies that are intrinsic to the SVD will remain. These are:

- Singular value decomposition is a global technique.

  It smoothes globally over all columns and all rows. Locally sensible smoothing is not possible. Thin rings cannot be kept in one part of the image and removed in another part. This is most likely the reason for the high variance in the estimated number of tooth rings in the above application, i.e., erasing thin tooth rings that should be kept or keeping thin rings that should be erased for some parts of the image.

- Singular value decomposition is a directional method.

  Information contained in orientations other than strictly horizontal or strictly vertical will be split into different ranks. Non-horizontal tooth rings are therefore more difficult to recognize and the rank-1 approximation even gives a wrong estimate. This could be regulated by the width of Y, the part of interest in the TCA image. The more columns Y contains, the higher is the possibility of non-horizontal tooth rings. But on the other hand, the fewer columns Y contains, fewer tooth rings are likely to be present because they may have disappeared in that very part.

- The rank of a matrix cannot be translated to the age of a person.

Despite the plastic interpretation of the rank-1 approximation and the information about the committed error of the rank-k approximation to a TCA image, the meaning of the rank in terms of number of lines or of the error in terms of extinguished tooth rings of certain width is not clear.

- A more sophisticated procedure is needed to evaluate the SVD smoothed image.

  The singular value decomposition method is used to smooth the image, but (except for the rank-1 approximation) it cannot directly be translated into an estimate for the number of tooth rings. For the histogram-based approach of searching the final estimate for the number of lines among the minima in each column, tooth rings must be strictly horizontal and the distribution of the number of minima must be well-behaved. It is not obvious how one smoothed image, or even several smoothed images of different ranks, could be evaluated without introducing additional heuristics. (See also the discussion for the Fourier smoothed TCA image on page 27.)

Because of these properties and because of the mixed results for estimating the number of lines in TCA images, we conclude that singular value decomposition is not (directly) suitable for the analysis of all types of TCA images. Altogether, SVD is a very rigid approach that might not even be useful for preprocessing TCA images (removing small scale noise) or for giving a rough first estimate of the number of rings. Compared to the Fourier transform, it cannot adapt sensitively enough to different kinds of TCA images because of the rank methodology and SVD is considerably slower.

This section has also shown the distinction between two aspects of evaluating TCA images: recognizing tooth rings (by the rank-1 approximation) and estimating their number (by summarizing estimates from approximations of different ranks). The problem of estimating the number of tooth rings might be simpler and may serve as a starting point for the recognition of tooth rings.

2.4 Discussion – Feature Measuring Methods

In the course of this chapter, one standard method from each of the areas of statistics, calculus and algebra was selected and applied to TCA images.

Hereby general problems of the applied methods, features of TCA images, as well as of TCA image evaluation emerged.

First, local regression was applied to TCA images to remove noise by approximating one data point by a weighted average of its neighboring data points. The functional form of this relationship between gray values and their location is unknown and was thus approximated. In the application a principal problem of local regression smoothing appeared: One essential specification of applying local regression, the width of the smoothing window, directly influences the number of recognized tooth rings. The impact of misspecifying the bandwidth by one pixel leads to an unstable procedure in the first place. Furthermore, a question inherent to the application arises: The best fit of a regression curve does not solely depend on the data, but also on the question of interest and thus on a clear definition of the object tooth ring, something that is not available.

Secondly, the two-dimensional Fourier transform was applied to enhance TCA images by emphasizing periodic structures belonging to tooth rings while neglecting structures belonging to noise. This was realized by generalized linear filtering using an elliptical zonal mask around the heuristically most-likely frequency. And thirdly, TCA images were enhanced by singular value decomposition through approximation with matrices of low rank. The TCA age estimates by the Fourier and the SVD smoothing of a series of TCA images were not comparable to estimates by manual observers. Both methods displayed similar deficiencies in the context of TCA image analysis: They are global techniques, such that thin rings cannot be kept in one part of the image while removed in another part. Fourier transform and especially SVD are directional methods, and thus too inflexible. The Fourier frequencies as well as the matrix ranks in SVD cannot be translated to the age of a person. For this reason, a final evaluation procedure for the Fourier and the SVD smoothed TCA images needs to be specified; and it would of course be best not to introduce additional heuristics.

These first attempts to evaluate TCA images also emphasized the distinction between two aspects of it: recognizing tooth rings and estimating their number. The latter problem may well be simpler and serve as a starting point for the recognition of tooth rings. With the knowledge of the number of tooth rings, other methods, for example, object recognition methods such as the snake (Kass et al. [1988]), could reveal their virtue.

Throughout this section certain characteristics of TCA images emerged that might interfere with general feature measuring methods. Theoretically,

each tooth ring should be present throughout the whole width of the TCA image. Practically speaking, however parts of tooth rings are often missing due to biological, microscopical or imaging peculiarities. Although the orientation of tooth rings is roughly horizontal, fluctuations that might cause problems occur frequently. Disturbances in TCA images, for instance artifacts and large blurry regions, might distort the features measured by a method.

Feature measuring methods cannot bridge this missing or falsified information. They can only assess certain characteristics of the image. In the course of the next three chapters, methods that are more tailored to TCA image analysis will therefore be developed, and an attempt will be made to imitate the mechanism of bridging information as the manual observers do, because they know that tooth rings should appear in the whole TCA image and that they should all run roughly parallel in the same direction.

Chapter 3

Gaussian Mixture Model

In the course of the next few chapters we will set up a model, i.e., a mathematical process that is capable of producing and describing observations. In image analysis, a statistical model generally serves two needs: It can describe the observed image with a set of parameters, and it can generate synthetic images, which are similar in terms of certain characteristics, from these model parameters (Cross and Jain [1983]). So far, methods like local regression, FFT and SVD smoothing were applied to TCA images by measuring texture features without incorporating them into a model. That is, we did not have an ideal or representative in mind. But now the advantages of model-based texture analysis will be exploited with its possibility of synthesizing texture and, more importantly, of incorporating prior knowledge in order to gain a deeper understanding by retrieving physically meaningful parameters.

For this purpose, the problem of evaluating TCA images has to be restated as a labeling problem (see, for example, Li [2001]). The observed image is assumed to be a random field and will be defined through the probability distribution of an underlying unknown random field, the label image: Let $\mathcal{S}$ be the set of pixels forming a rectangular lattice of size $N \times M$:

$$\mathcal{S} = \{1, \ldots, NM\}.$$

For each pixel i gray value $Y(i)$ is observed:

$$\mathcal{Y} : \mathcal{S} \mapsto \mathbb{R}^{N \times M} \text{ such that } i \mapsto Y(i).$$

The whole observed image $Y(\mathcal{S})$ will simply be denoted by Y.

Definition 3.0.1. *A* labeling *(or coloring)* Λ *is assigning an event* $\lambda(i)$ *to site* i*:*

$$\Lambda : \mathcal{S} \mapsto \mathcal{G}^{N \times M} \text{ such that } i \mapsto \lambda(i)$$

where a discrete label set $\mathcal{G} = \{0, 1, \ldots, G\}$ *is assumed. The whole label image* $\lambda(\mathcal{S})$ *will be abbreviated by* λ*.*

The labeling problem is the task of estimating the unknown label image from the noise-corrupted observed image. This problem will be solved for three different kinds of models in the upcoming chapters. Chapter 3 will define a plain Gaussian mixture model and fit it by the EM and by the RJ MCMC algorithm to TCA images. Mixture models are histogram based models, that is, they rely on the assumption of independent observations and do not take into account the relationship between nearby pixels. Chapter 4 will relax this assumption to incorporate two-dimensional correlations between gray values by hidden Markov random field modeling. Chapter 5 will take into account directed two-dimensional interactions by coupled hidden Markov modeling.

3.1 Introduction to Finite Mixture Models

Mixture models are an important and frequently applied statistical approach to model distributions in which observations arise from separate groups, but individual membership is not known (Webb [1999]). In TCA images one could assume that the observed gray values arise from one of two classes: either black or white rings. But neither the proportions nor the classification of each pixel are known. Figure 3.1 illustrates this assumption. It shows the relative frequencies of gray values conditional on the membership of each pixel in a white or a black ring. This membership arose from a rough manual segmentation on a good quality TCA image.

More generally, mixture models can be applied to explore data for such group structure (McLachlan and Peel [2000]). They have provided a flexible, mathematically simple, semiparametric method for modeling unknown distributional shapes for more than a century (McLachlan and Peel [2000]). Besides the references named above, there is a wide variety of literature on theoretical issues, applications and new developments for mixture models.

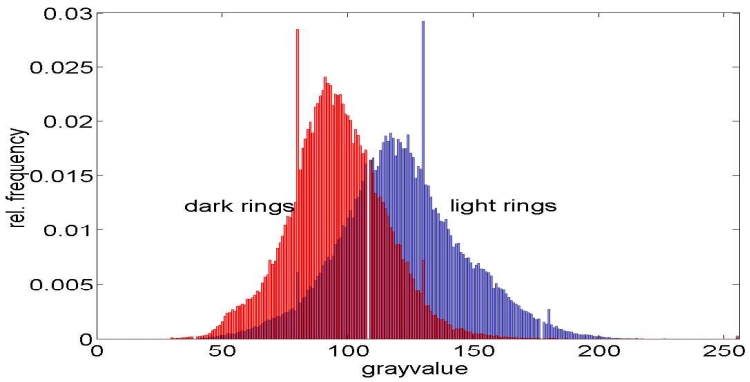


Figure 3.1: Relative frequency of gray values given the pixel belongs to a white ring (in blue) or respectively to a black ring (in red). Frequency counts were taken from a good quality TCA image, where the label of each pixel was manually chosen to form a rough segmentation into black and white lines.

Under a finite mixture model, the marginal distribution $f(Y_i)$ of a random variable $\mathcal{Y}_i$ is defined as

$$f(Y_i) = \sum_{g=0}^{G} w_g f(Y_i|\boldsymbol{\theta}_g),$$

where $G+1$ is the number of mixture components and $f(Y_i|\boldsymbol{\theta}_g)$ is the conditional density of component $g = 0, 1, \ldots, G$ depending on parameters $\boldsymbol{\theta}_g$. The probability distribution of the labels $P(\lambda_i = g)$ is denoted by w_g and can be thought of as nonnegative mixing proportions or weights with the property $\sum_{g=0}^{G} w_g = 1$.

The joint distribution of NM observations $Y = (Y_1, \ldots, Y_{NM})$ can thus be expressed by the distribution of the observed gray values conditioned on the labels weighted by the mixing proportions:

$$f(Y|\boldsymbol{\theta}, w) = \prod_{i \in \mathcal{S}} \sum_{g=0}^{G} w_g f(Y_i|\boldsymbol{\theta}_g).$$

Given the observations, the likelihood becomes a function of the parameters $(\boldsymbol{\theta}, w)$:

$$L(\boldsymbol{\theta}, w|Y) = f(Y|\boldsymbol{\theta}, w).$$

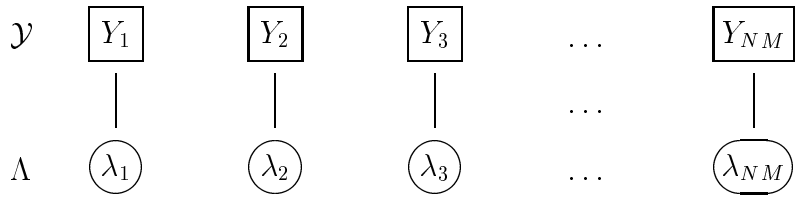


Figure 3.2: Conditional independence graph of a mixture model. Boxes represent observed or fixed quantities; circles show unknowns.

Figure 3.2 displays the conditional independence graph of a mixture model, where the existence of an edge between two vertices expresses that the quantities assigned to these vertices are conditionally dependent given all other quantities. That is, for any pair of pixels i and j

$$\begin{aligned} f(Y_i, Y_j|\lambda, Y_{\mathcal{S}\setminus\{i,j\}}) &= f(Y_i|\lambda_i)f(Y_j|\lambda_j) \\ P(\lambda_i, \lambda_j|\lambda_{\mathcal{S}\setminus\{i,j\}}, Y) &= P(\lambda_i)P(\lambda_j) \\ f(Y_i, \lambda_i|\lambda_{\mathcal{S}\setminus\{i,j\}}, Y_{\mathcal{S}\setminus\{i,j\}}) &\neq f(Y_i)P(\lambda_i) \end{aligned}$$

in the present model. This context-blind setting will be expanded in the course of the next few chapters.

One of the most popular mixture models, which is also widely used in image segmentation, is the normal mixture model (Zhang et al. [2001]). For univariate normal mixture models, each component density has the form of a normal distribution with parameters $\boldsymbol{\theta}_g = (\mu_g, \sigma_g^2)$:

$$f(Y_i|\boldsymbol{\theta}_g) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(Y_i-\mu_g)^2}{2\sigma_g^2}}.$$

In order to fit a finite mixture model to observations, there are two sets of parameters to estimate: the component parameters $\boldsymbol{\theta}_g$ and the mixing proportions $w_g = P(\lambda_i = g)$. In this application, there is also the need to estimate group membership (the label), which will be denoted by λ. Following the arguments in Titterington et al. [1985], one can see that the set of all finite univariate normal mixtures

$$\left\{ f(Y_i) = \sum_{g=0}^{G} w_g f(Y_i|\boldsymbol{\theta}_g) | G = 0, 1, \ldots \right\}$$

is identifiable, i.e., for any member there exists a unique solution for G, w_0, $w_1, \ldots$ and $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots$ (up to permutations). Several methods exist to solve for these unknowns; the best known is maximum likelihood estimation. But explicit formulas for estimates $\hat{w}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\lambda}$ are typically not available, neither for maximum likelihood nor for other approaches. But ever since the availability of modern computer technology, using iterative procedures has been possible for parameter estimation (McLachlan and Peel [2000]). The EM algorithm has been one of the favorite methods of maximizing a likelihood in this setting since Dempster et al. [1977].

In the following section, the reader will therefore be introduced to the EM algorithm in general (Section 3.2.1) and then specifically as it pertains to image segmentation (Section 3.2.2). The results therein will demonstrate that the choice of the number of components $G + 1$ is a central issue that needs to be addressed during the fitting procedure. Since the number of components changes the dimension of the parameter space, maximum likelihood is unable to simultaneously infer the labels, the mixing proportions, the density parameters and the number of components from the data. We chose the fully Bayesian framework, specifically the RJ MCMC algorithm, to resolve this difficult problem in Section 3.3. Lastly, in Section 3.4, conclusions will be drawn and problems of applying finite normal mixture models to TCA images stated.

3.2 Parameter Estimation via EM

If the estimation of parameters $(\boldsymbol{\theta}, w)$ requires optimization of a function, such as the likelihood $L(\boldsymbol{\theta}, w)$, one can use several methods. When direct optimization is intractable, one needs to rely on numerical or statistical methods, such as Newton's method or Fisher scoring. But both of these involve drawbacks, especially for estimating a large number of parameters, since both algorithms need to invert large matrices.

The EM algorithm is a simple and numerically stable algorithm which replaces the complex optimization problem by a sequence of simpler ones. For this purpose it assumes the existence of additional but missing parameters, labels λ_i in our case; and iterates the maximization of a so-called complete likelihood, instead of maximizing $L(\boldsymbol{\theta}, w)$.

Articles by Becker et al. [1997] and Bilmes [1997] respectively provide the foundation for the subsequent description of theory in Section 3.2.1 and for application in Section 3.2.2.

3.2.1 Theory

The pair (Y, λ) will be called complete data (opposed to the incomplete data Y) and assumed to have the joint density

$$P(Y, \lambda|\boldsymbol{\theta}, w) = f(Y|\lambda, \boldsymbol{\theta}, w)P(\lambda|\boldsymbol{\theta}, w).$$

From this a new likelihood function called complete-data likelihood can be specified

$$L^{full}(\boldsymbol{\theta}, w|Y, \lambda) = P\left(Y, \lambda|\boldsymbol{\theta}, w\right). \tag{3.1}$$

The EM algorithm first calculates the expectation (E-step) of the complete-data log-likelihood with respect to the unknown data λ, given current parameter estimates $(\boldsymbol{\theta}, w)^{(t-1)}$ and the observed data Y

$$\begin{aligned} & E\left[\log P(Y, \lambda|\boldsymbol{\theta}, w) \,\middle|\, Y, \boldsymbol{\theta}^{(t-1)}, w^{(t-1)}\right] \\ & \quad = \int\limits_{\lambda \in \mathcal{G}^{N \times M}} \log\left(P(Y, \lambda|\boldsymbol{\theta}, w)\right) P\left(\lambda \,\middle|\, Y, \boldsymbol{\theta}^{(t-1)}, w^{(t-1)}\right) d\lambda \end{aligned} \tag{3.2}$$

where Y and $(\boldsymbol{\theta}, w)^{(t-1)}$ are constant, $P\left(\lambda|Y, \boldsymbol{\theta}^{(t-1)}, w^{(t-1)}\right)$ is the predictive distribution of the unobserved data and $\mathcal{G}^{N \times M}$ is the space of values λ can take on.

The second step (M-step) of the EM algorithm involves maximizing this expectation to obtain a better estimate of $(\boldsymbol{\theta}, w)$:

$$(\boldsymbol{\theta}, w)^{(t)} = \underset{(\boldsymbol{\theta}, w)}{\operatorname{argmax}} E\left[\log P(Y, \lambda|\boldsymbol{\theta}, w) \,\middle|\, Y, \boldsymbol{\theta}^{(t-1)}, w^{(t-1)}\right] \tag{3.3}$$

Starting with preliminary estimates $(\boldsymbol{\theta}, w)^{(0)}$, these two steps are iterated until convergence. Since each iteration is guaranteed to increase the (incomplete) log-likelihood under mild assumptions, the EM algorithm will converge to a local maximum (McLachlan and Krishnan [1997]).

3.2.2 EM-Classification Using Normal Mixture Modeling

In the case of a normal mixture model, the maximization of the EM algorithm in (3.3) is simplifying because the complete-data log-likelihood (3.1) resolves into summands such that the parameters θ and w separate. According to Bilmes [1997] the EM algorithm is hereby reduced to the three updating formulas

$$w_g = \frac{1}{NM} \sum_{i \in \mathbf{S}} P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right) \tag{3.4}$$

$$\mu_g = \frac{\sum\limits_{i \in \mathcal{S}} Y_i P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right)}{\sum\limits_{i \in \mathcal{S}} P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right)} \tag{3.5}$$

$$\sigma_g^2 = \frac{\sum\limits_{i \in \mathcal{S}} P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right) (Y_i - \mu_g)^2}{\sum\limits_{i \in \mathcal{S}} P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right)} \tag{3.6}$$

which are then repeated. The posterior probability of component g given the data can be computed as follows

$$P\left(\lambda_i = g | Y_i, \boldsymbol{\theta}^{(t-1)}\right) = \frac{w_g^{(t-1)} f\left(Y_i | \boldsymbol{\theta}_g^{(t-1)}\right)}{\sum\limits_{g=0}^{G} w_g^{(t-1)} f\left(Y_i | \boldsymbol{\theta}_g^{(t-1)}\right)}.$$

Reasonable guesses for starting values could be means, variances and proportions of components obtained by equally partitioning the set of existing gray values. Random starting values (most likely) give different results because of the non-concavity of the likelihood. When no other prior information is available, experiments have shown that the above choice of deterministically partitioning the observations is sensible and certainly not the worst. This is underscored by the application of an MCMC method in the next section.

In order to apply the algorithm and for reasons of identifiability (see page 42), the number of components $G+1$ has to be inferred. As mentioned above, the EM algorithm and ML estimation in general are unable to address this

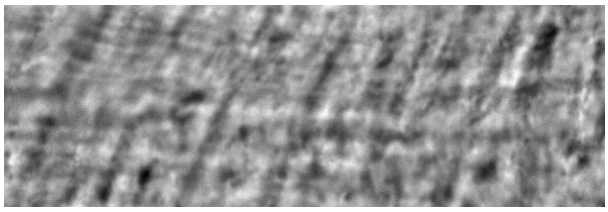


Figure 3.3: A small 200x600-pixel part of the standard exemplary TCA image IS_0000666 used as an example.

issue, because a varying number of components leads to varying dimensions of the parameter space. The user therefore needs to guess the number of components in our first application. A natural, suggested assumption would be two components: black ($g = 0$) and white rings ($g = 1$). Section 3.3 will demonstrate an approach for coherently assessing the number of components $G + 1$, the labels λ, the mixing proportions w and the component density parameters $\boldsymbol{\theta}$.

Figure 3.3 shows an image which contains 200x600 pixels, a small part of the standard exemplary TCA image in Figure 1.3. Figure 3.4 illustrates the EM algorithm (mentioned above) as applied to this example image. The estimated most likely labeling, as well as traces of the log-likelihood, the mean and a histogram of the observed gray values, the initial and the fitted distributions are shown there. The algorithm was executed on Hydra[1], and (a not-optimized version) needed about one hour depending on the number of gray levels. A graph of the trace of one exemplary parameter and the incomplete-data log-likelihood (depicted in Figure 3.4 (b) and (c)) leads one to suppose that the algorithm has converged. The parameter estimates $\hat{\mu}_0 = 29942$, $\hat{\mu}_1 = 32212$, $\hat{\sigma}_0 = 3205$ and $\hat{\sigma}_1 = 2782$ suggest that the observations are a mixture of two overlapping normal distributions, quite similar to the initial normal distributions (see Figure 3.4 (d)). But the labels do not reveal the expected results of a segmentation into white and black rings. Instead, the labeling emphasizes the vertical direction of the saw cuts rather than the horizontal one of tooth rings. This is a typical result that was reproduced for other TCA images.

[1] Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

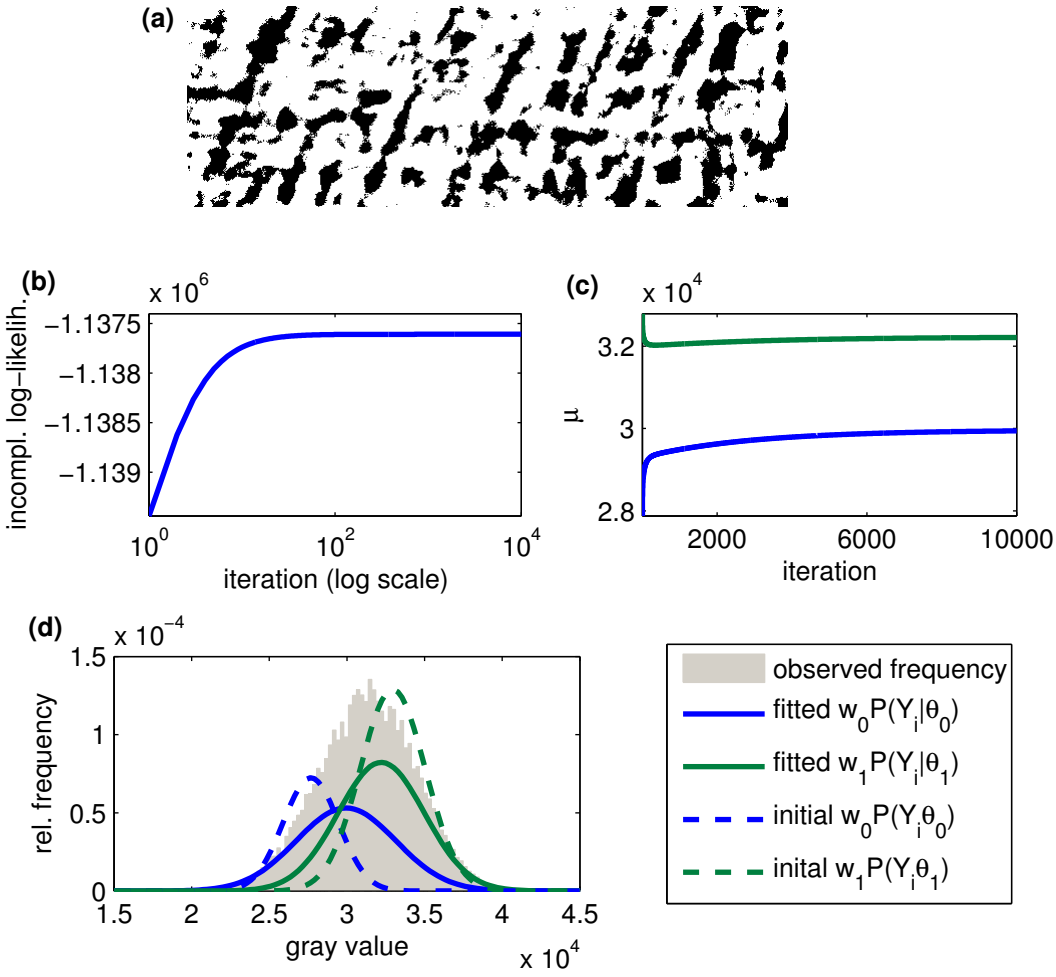


Figure 3.4: Estimated most likely labeling of the part of a TCA image shown in Figure 3.3 assuming a mixture of two normal distributions (a). Graphs for traces of the log-likelihood (b) and the means (c). The bottom graph (d) shows the fitted and initial normal distributions as well as the binned observed frequencies.

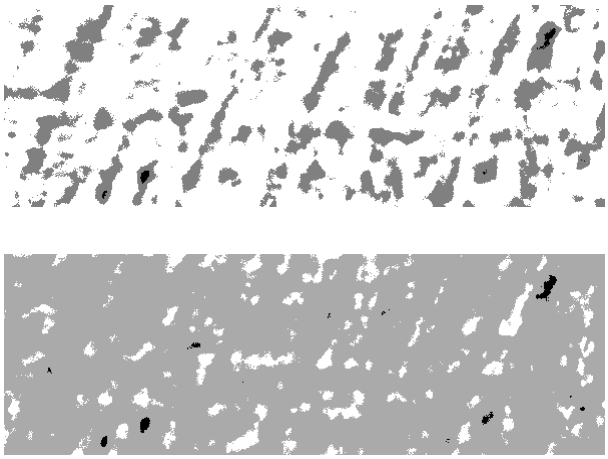


Figure 3.5: Estimated most likely labeling of image 3.3 assuming a mixture of three (top) and four (bottom) normal distributions.

One reason for the failure of fitting a normal mixture model to parts of TCA images might be the assumption of two components. To explore this deficiency, the observer can see in Figure 3.5 the results of fitting a mixture of three and four normal distributions to the same image (using more iterations). The next section will examine the possibility of estimating the number of components while fitting the mixture model.

3.3 Parameter Estimation via RJ MCMC

The issue of identifiability, stated on page 42 and the applications of the EM algorithm to TCA images above, show that the choice of the number of components is a key problem for fitting a finite mixture model to observations.

Most approaches to this difficult issue separate the problem of assessing G, from fitting the mixture model with fixed G. Two examples are penalized log-likelihood measures called information criteria (for example, Akaike's Information Criterion) and testing the hypothesis $G = G_0$ against $G = G_1$. But simultaneous estimation of the number of components, the parameters of the component densities, mixture proportions and the most likely classification of each observation can be done in a fully Bayesian framework. The Reversible Jump Markov Chain Monte Carlo (abbreviated by RJ MCMC)

method, promoted by Richardson and Green [1997], is one of few choices to fulfill this criterion. The current section will closely follow Richardson and Green [1997] to introduce this method of mixing over the fixed G cases. Section 3.3.1 will present the hierarchical Bayesian model, Section 3.3.2 will derive the proper MCMC method and Section 3.3.3 will explain the results of applying the RJ MCMC algorithm to TCA images.

3.3.1 Model Specification

The joint distribution of the number of components $G+1$ and the labels λ, the mixing proportions w, the component density parameters $\boldsymbol{\theta}$ as well as the observations Y can generally be rewritten as

$$P(G, w, \lambda, \boldsymbol{\theta}, Y) = P(G)P(w|G)P(\lambda|w, G)P(\boldsymbol{\theta}|\lambda, w, G)f(Y|\boldsymbol{\theta}, \lambda, w, G).$$

In the case of normal mixture models, the following (conditional) independencies and distributions can be assumed:

$$\begin{aligned} P(\boldsymbol{\theta}|\lambda, w, G) &= P(\boldsymbol{\theta}|G) \\ f(Y|\boldsymbol{\theta}, \lambda, w, G) &= f(Y|\boldsymbol{\theta}, \lambda) = \prod_{i \in \mathcal{S}} f(Y_i|\boldsymbol{\theta}_{\lambda_i}, \lambda_i) = \prod_{i \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_g^2}} \\ P(\lambda|w, G) &= \prod_{i \in \mathcal{S}} P(\lambda_i|w, G) = \prod_{i \in \mathcal{S}} w_{\lambda_i}. \end{aligned}$$

The Bayesian framework now considers G, w and $\boldsymbol{\theta}$ as unknowns and models them according to prior distributions. These priors are dependent on hyperparameters G_{max}, δ and η, which in turn are drawn from independent hyperpriors. The joint distribution of all variables (observations, labels, so-called unknowns and hyperparameters) is then given by

$$\begin{aligned} &P(G_{max}, \delta, \eta, G, w, \lambda, \boldsymbol{\theta}, Y) = \\ &P(G_{max})P(\delta)P(\eta)P(G|G_{max})P(w|G, \delta)P(\lambda|w, G)P(\boldsymbol{\theta}|G, \eta)f(Y|\boldsymbol{\theta}, \lambda). \end{aligned}$$

For univariate normal mixture models, the prior distributions can be specified by

$$\begin{aligned} \mu_g &\sim N(\xi, \kappa^{-1}), \\ \sigma_g^2 &\sim \Gamma(\alpha, \beta), \\ w &\sim D(\delta, \delta, \ldots, \delta) \text{ and} \\ G &\sim Uniform(1, G_{max}) \end{aligned}$$

where G_{max} is a prespecified integer, $\eta = (\xi, \kappa, \alpha, \beta)$ are hyperparameters, $\Gamma(\alpha, \beta)$ is the Gamma distribution with mean $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$ and D the symmetric Dirichlet distribution. For complete specification of the model, Richardson and Green [1997] chose the hyperparameters to be weakly informative and defined, for example,

$$G_{max} = 30,\ \alpha = 2 \text{ and } \delta = 1 \text{ to be constant,}$$

$$\xi = \frac{\max(Y_i) + \min(Y_i)}{2} \text{ and } \kappa = \frac{1}{R^2} \text{ to be range-dependent and}$$

$$\beta \sim \Gamma\left(0.2, \frac{10}{R^2}\right) \text{ to be random,}$$

where $R = \max(Y_i) - \min(Y_i)$ is the range of the observations. See Richardson and Green [1997] for heuristic, and references therein for some theoretic argumentation regarding these choices.

Figure 3.6 summarizes the hierarchy of the involved model variables. It displays a directed acyclic graph for the joint distribution, as well as a conditional independence graph of this model.

So far, the whole model described above is invariant to permutations of the labels $g = 0, 1, \ldots, G$. For reasons of identifiability we assume a unique labeling with $\mu_0 < \mu_1 < \ldots < \mu_G$.

3.3.2 The Algorithm: Sampling and Jumping

The reversible jump MCMC algorithm is a random sweep Metropolis-Hastings method and in the case of a hierarchical normal mixture model will repeat the following six move types:

(1) update the weights w

(2) update the parameters (μ, σ)

(3) update the labels λ

(4) update the hyperparameter β

(5) split one mixture component into two or combine two into one

(6) birth or death of an empty component

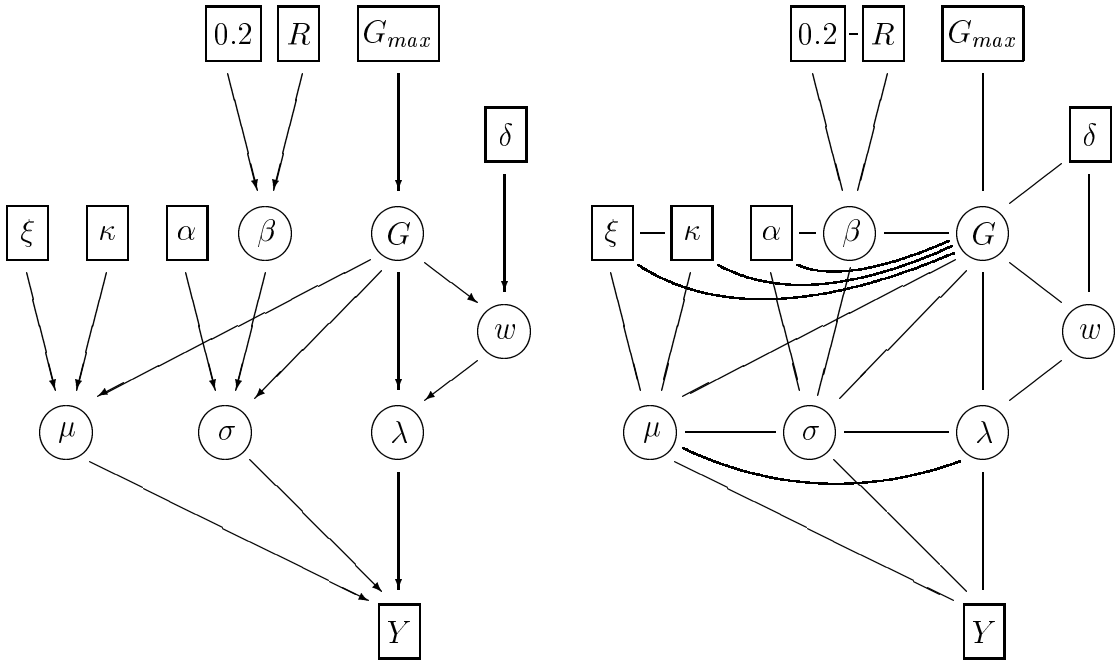


Figure 3.6: Directed acyclic graph (left) representing the joint distribution of the hierarchical normal mixtures model. Marrying parents and dropping arrows gives the corresponding conditional independence graph (right). Boxes show observed or fixed quantities, circles show unknowns. (from Richardson and Green [1997])

Move types 1 to 4 do not alter the dimension of the vector of unknowns $(\beta, \mu, \sigma, G, w, \lambda)$ and are drawn from the following full conditional distributions (also displayed in Figure 3.6):

$$P(w|\lambda, \sigma, G) = D\left(\delta + n_0, \ldots, \delta + n_G\right)$$

$$P(\mu_k|Y, \lambda, \sigma, G, \xi, \kappa) = N\left(\frac{\sigma_g^{-2} \sum_{i:\lambda_i=g} Y_i + \kappa\xi}{\sigma_g^{-2} n_g + \kappa}, \left(\sigma_g^{-2} n_g + \kappa\right)^{-1}\right)$$

$$P(\sigma_g^{-2}|Y, \lambda, \mu, G, \alpha, \beta) = \Gamma\left(\alpha + \frac{1}{2} n_g, \beta + \frac{1}{2} \sum_{i:\lambda_i=g} (Y_i - \mu_g)^2\right)$$

$$P(\beta|\sigma, G, \alpha, R) = \Gamma\left(0.2 + G\alpha, R + \sum_{g=0}^{G} \sigma_g^{-2}\right)$$

where $n_g = |\{i \in \mathcal{S} \mid \lambda_i = g\}|$ is the number of pixels with label g. The allocation variables are drawn from

$$P(\lambda_i = g|Y, \sigma, w, G) \propto \frac{w_g}{\sigma_g} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_g^2}}.$$

Thus w can be updated by drawing independent random Gamma variables and scaling them to one; the drawn μ_k are just proposals and accepted only when the ordering is unchanged.

Moves 5 and 6 can increment or decrement the number of components. The vector $(\mu, \sigma, w, \lambda)$ thus needs to be updated in a way that can be reversed.

Move 5 includes a random choice between trying to split or to combine with probabilities b_{G+1} and $d_{G+1} = 1 - b_{G+1}$, where $b = (1, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, 0)$ and $d = (0, \frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, 1)$. To combine a component, $g \leq G$ is chosen at random and G is decremented by merging components g and $g + 1$ into the new component labeled g^*. For this purpose observations Y_i with $\lambda_i = g$ or $\lambda_i = g + 1$ are reallocated by setting $\lambda_i = g^*$ and $(w_{k^*}, \mu_{k^*}, \sigma_{k^*})$ are chosen to satisfy

$$w_{g^*} = w_g + w_{g+1} \tag{3.7}$$

$$w_{g^*}\mu_{g^*} = w_g\mu_g + w_{g+1}\mu_{g+1} \tag{3.8}$$

$$w_{g^*}(\mu_{g^*}^2 + \sigma_{g^*}^2) = w_g(\mu_g^2 + \sigma_g^2) + w_{g+1}(\mu_{g+1}^2 + \sigma_{g+1}^2). \tag{3.9}$$

For the reverse split proposal, a component g^* is chosen at random and split into two components labeled g and $g+1$. The new weights, means and variances have to be chosen in correspondence to equations (3.7) to (3.9). This results in three degrees of freedom generated by $u_1 \sim be(2,2)$, $u_2 \sim be(2,2)$ and $u_3 \sim be(1,1)$. Then

$$\begin{aligned} w_g = w_{g^*} u_1 \quad &, \quad w_{g+1} = w_{g^*}(1-u_1), \\ \mu_g = \mu_{g^*} - u_2 \sigma_{g^*} \sqrt{\frac{w_{g+1}}{w_g}} \quad &, \quad \mu_{g+1} = \mu_{g^*} + u_2 \sigma_{g^*} \sqrt{\frac{w_g}{w_{g+1}}}, \\ \sigma_g^2 = u_3 \left(1-u_2^2\right) \sigma_{g^*}^2 \frac{w_{g^*}}{w_g} \quad &, \quad \sigma_{g+1}^2 = (1-u_3)\left(1-u_2^2\right) \sigma_{g^*}^2 \frac{w_{g^*}}{w_{g+1}}. \end{aligned}$$

After confirming a correctly ordered μ (and rejecting the move otherwise), the reallocation of those Y_i with $\lambda_i = g^*$ is performed analogously to the standard allocation. The proposal distributions for move 5 were chosen according to Richardson and Green [1997]. Their aim was to create a reasonable probability of acceptance under the restriction of matching the dimension. The acceptance probability for splitting component g^* is then $\min(1, A)$ where

$$\begin{aligned} A \;=\; & \prod_{i \in \mathcal{S}} \frac{f(Y_i|\boldsymbol{\theta}_{\lambda_i}^{new})}{f(Y_i|\boldsymbol{\theta}_{\lambda_i}^{old})} \frac{P(G+2)}{P(G+1)} (G+2) \frac{w_g^{\delta-1+l_1} w_{g+1}^{\delta-1+l_2}}{w_{g^*}^{\delta-1+l_1+l_2} be(\delta, (G+1)\delta)} \\ & \cdot \sqrt{\frac{\kappa}{2\pi}} e^{-\frac{1}{2}\kappa\left((\mu_g-\xi)^2+(\mu_{g+1}-\xi)^2-(\mu_{g^*}-\xi)^2\right)} \\ & \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_g^2 \sigma_{g+1}^2}{\sigma_{g^*}^2}\right)^{-\alpha-1} e^{-\beta\left(\sigma_g^{-2}+\sigma_{g+1}^{-2}-\sigma_{g^*}^{-2}\right)} \\ & \cdot \frac{d_{G+2}}{b_{G+1} P_{alloc}} \left(be(u_1|2,2) be(u_2|2,2) be(u_3|1,1)\right)^{-1} \\ & \cdot \frac{w_{g^*}|\mu_g - \mu_{g+1}| \sigma_g^2 \sigma_{g+1}^2}{u_2(1-u_2^2) u_3 (1-u_3) \sigma_{g^*}^2} \end{aligned}$$

with $G+1$ being the number of components before the split, l_1 and l_2 the number of observations proposed to be assigned to component g respectively $g+1$, and P_{alloc} is the probability that this particular allocation is made. This lengthy product results from the generalized Metropolis-Hastings ratio where lines ones through three form the ratio of the probability of being in state $x' = (w_g, \mu_g, \sigma_g^2, w_{g+1}, \mu_{g+1}, \sigma_{g+1}^2)$ from higher dimensional space, to the probability of being in the lower dimensional state $x = (w_{g^*}, \mu_{g^*}, \sigma_{g^*}^2)$, given

all observations Y. The factor $(G+2)$ results from the ratio $\frac{(G+2)!}{(G+1)!}$ of ordering the components. The fourth line represents the ratio of the probability of choosing this type of move being in state x' to the probability of this move in state x times the density of u. The last line forms the Jacobian of the transformation from (x, u) to x'.

Analogously the acceptance probability of combining components g and $g+1$ is $\min(1, A^{-1})$ with some obvious substitutions.

Move 6 is also very similar. Birth or death is randomly selected using b_{G+1} and d_{G+1}. For the birth of a new empty component, we propose $w_{g^*} \sim be(1, G+1)$, scale the weights to sum to one, $\mu_{g^*} \sim N(\xi, \kappa^{-1})$ and $\sigma_{g^*}^{-2} \sim \Gamma(\alpha, \beta)$. For a death we randomly choose any existing empty component which is deleted and the weights are scaled to one. The acceptance probabilities for birth and death are then $\min(1, A)$ respectively $\min(1, A^{-1})$ with the simplified

$$\begin{aligned} A &= \frac{P(G+2)}{P(G+1)}(G+2)\frac{w_{g^*}^{\delta-1}(1-w_{g^*})^{n+(G+1)\delta-G}}{be((G+1)\delta, \delta)} \\ &\cdot \frac{d_{G+2}}{b_{G+1}(G_0+2)}\frac{(1-w_{g^*})^{g-1}}{be(w_{g^*}|1, G)} \end{aligned}$$

where $G+1$ is the number of components and G_0+1 is the number of empty components before the birth.

The moves 1 to 6 together are called a sweep and need to be repeated sufficiently often in order to create a Markov chain.

The RJ MCMC algorithm as stated above was implemented in Matlab with the help of Fortran procedures offered on Peter J. Green's homepage: Green [2004].

3.3.3 RJ MCMC-Classification Using Normal Mixture Modeling

TCA images contain a large number of observations (5000 times more than the sample data set Richardson and Green [1997] used). We therefore chose a small portion of the standard exemplary TCA image, 200x600 pixels in size, in order to test the RJ MCMC algorithm. Additionally the number of sweeps

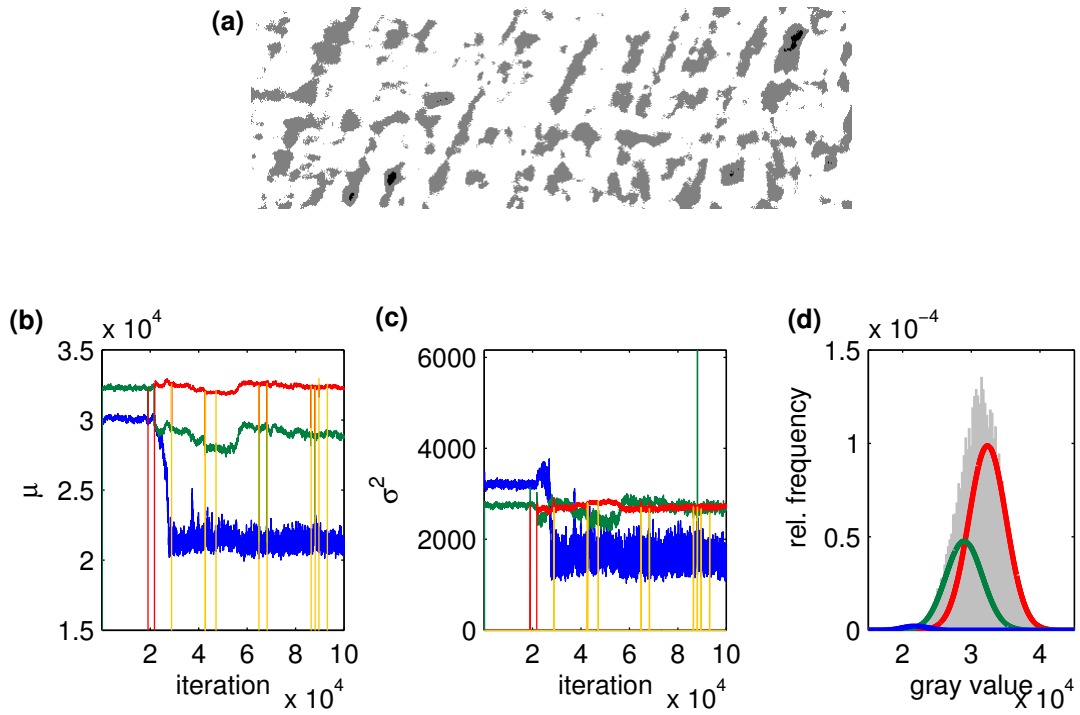


Figure 3.7: Bayes' classification of the part of a TCA image shown in Figure 3.3, assuming a mixture of normal distributions (a) and graphs for traces of the means (b), the variances (c) and the fitted normal distributions as well as the binned observed frequency (d).

was set to 100,000 including an initial burn-in period of 10,000 sweeps. These settings required a computation time of seven days on Hydra[2].

Figure 3.7 shows the result of applying RJ MCMC to the part of a TCA image shown in Figure 3.3 for fitting a normal mixture model. The most frequent (among the 90,000) colorings contained three components ($P(G = 2) = 0.78$) and the most likely coloring among those is depicted in the top of Figure 3.7. This coloring will be called a Bayes' classification. Also visible in Figure 3.7 are traces of the parameters $\boldsymbol{\theta}$, where the entrance or exit of one color indicates the birth or death, respectively, of a new component.

[2]Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

The predicted most likely distribution of each component together with the binned observed frequencies are depicted in the rightmost graph.

Comparing the results from fitting a mixture of three normal distributions by the EM algorithm on one side, to fitting a mixture of normal distributions by the RJ MCMC algorithm where the most likely number of components is three, the reader will not note any difference between Figures 3.4 and 3.7. The parameters as well as the labeling match very closely. A shorter run of (for example) 10,000 iterations favored a mixture of two normal distributions that was also very close to the result of the EM algorithm.

Figure 3.7 also demonstrates that the RJ MCMC algorithm recognizes saw cuts instead of tooth rings, just as the EM algorithm did. Similar results are obtained when applying this method to different TCA images.

Another (unexpected) drawback of the RJ MCMC is the computational inefficiency. This method possibly requires many iterations and therefore could not be applied to large data sets (like a complete TCA image) in everyday practice.

3.4 Discussion – Mixture Modeling

Fitting a normal mixture model to TCA images reveals a most likely coloring that does not resemble tooth rings. This problem appeared when using the EM algorithm to fit the model for a fixed number of labels $G + 1$, as well as when using the RJ MCMC algorithm to fit the model while mixing over G. There might be several reasons for this problem.

The most probable reason is connected to the identifiability issue that has heretofore been mentioned. For example McLachlan and Peel [2000] note that any continuous distribution can be approximated arbitrarily well by a finite number of normal densities. There is thus not necessarily a one-to-one correspondence between the fitted mixture components and the desired groups in an application. This problem occurs especially when data are skewed or when an outlier exists. The applications in Figures 3.4, 3.7 and Figure 3.8 highlight this issue and demonstrate that fitting a mixture model to TCA images is not a well-posed problem. The gray value of each pixel alone does not separate the mixture in the required way. For example, it is unlikely to separate the distribution of gray values in a good quality TCA image as shown in Figure 3.8 into the two displayed groups, solely based on the histogram. In images the gray level of one pixel is highly dependent

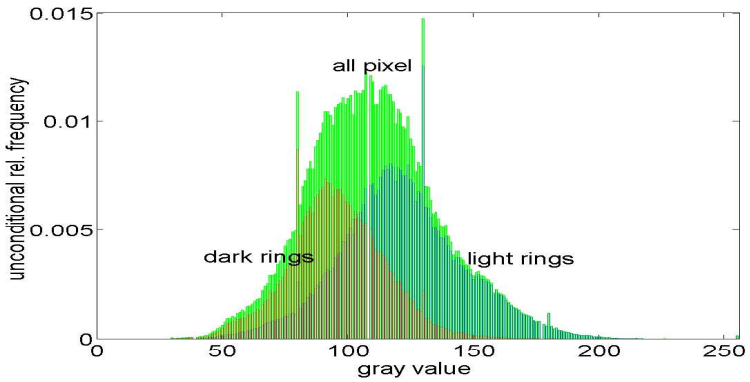


Figure 3.8: Relative unconditional frequency of gray values (green) in a manually segmented good quality TCA image (from Figure 3.1). The blue respectively red bars give the conditional frequencies multiplied by the proportion of pixels belonging to a light or dark ring.

on gray levels of neighboring pixels (Cross and Jain [1983]). And thus a histogram based model (as a mixture model) that cannot include spatial information, is intrinsically limited to well-defined images with low levels of noise (Zhang et al. [2001]). This situation is obviously not present in most TCA images.

Since each observation can influence the model fit, it might also be the case that another, more carefully chosen part of TCA images may result in a better segmentation. But smaller TCA image sections are not representative. Larger sections, like the squared one marked in Figure 1.3, have a high chance of including artifacts and very dark or bright tooth rings (for example, the outermost and innermost rings) that overlay the 'true' mixture of distributions.

Another possible reason for the unsatisfactory results of fitting a normal mixture model to TCA images might be the assumption of normality. The EM as well as the RJ MCMC algorithm can also be applied to different distributions, such as the Poisson. This latter possible source of failure of mixture modeling due to the normality assumption, will not be addressed here because the generalization of including spatial dependencies seems more promising.

The next chapters will introduce statistical models to incorporate knowledge about the relation of spatially nearby pixels.

Chapter 4

Hidden Markov Random Field Models

Hidden Markov Random Field (HMRF) models provide a natural framework for the labeling problem of images. These kinds of models incorporate contextual constraints: While using the same framework as for mixture modeling, the distribution of the labels will be generalized to include dependencies among the pixels. We will follow Zhang et al. [2001] to introduce HMRF models.

HMRF models consist of two levels: a Markov random field Λ that cannot be observed directly, and an observable random field $\mathcal{Y}$. These two random fields are linked by the conditional probability of the observed image, given the true one

$$f(Y|\lambda) = \prod_{i \in \mathcal{S}} f(Y_i|\lambda_i),$$

whose components Y_i are assumed to be conditionally independent. The distribution $f(Y|\lambda)$ is often called cost function or emission probability.

The most popular choice for a cost function is the Gaussian distribution

$$\begin{aligned} f(Y|\lambda, \boldsymbol{\theta}) &= \prod_{i \in \mathcal{S}} f(Y_i|\lambda_i, \boldsymbol{\theta}) \\ &= \prod_{i \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_g^2}} \end{aligned} \tag{4.1}$$

centered at mean μ_g for label $\lambda_i = g$ and with variance σ_g^2; or in other words

$$Y_i = \mu_{\lambda_i} + \epsilon(i) \text{ with } \epsilon(i) \sim N(0, \sigma_{\lambda_i}^2),$$

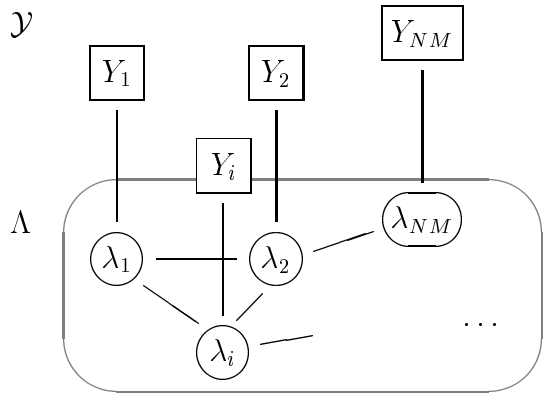


Figure 4.1: Conditional independence graph of a generic hidden Markov random field model. Boxes represent observed or fixed quantities, circles show unknowns.

where the parameters $\boldsymbol{\theta} = \{\mu_g, \sigma_g^2 \mid g \in \mathcal{G}\}$ and the labels λ_i are unknown. This setting is the same as for the normal mixture model in Chapter 3. But the former mixing proportions $P(\lambda)$ will during this chapter not be independent and may hereby contain spatial information. Figure 4.1 shows a conditional independence graph for a generic hidden Markov random field. In this sense, HMRF models are extensions of mixture models.

After explaining the general concept of MRF models in Section 4.1 (following Li [2001] and Winkler [1995]), attention is directed toward specific MRFs. Section 4.2 will describe the auto-logistic model and Section 4.3 the FRAME model. The latter section will also outline how to fit the model, that is, how to estimate the involved parameters together with their confidence intervals, and how to estimate the underlying labels. Applications to TCA images will be included as well.

4.1 General Markov Random Field Models

MRFs have a long history, beginning in physics with Ising's thesis in 1925 (Ising [1925]). Besag extended the so-called Ising model (Besag [1974]) and popularized the approach in image analysis (Besag [1986]). Ever since, MRF models are popular choices to model contextual constraints, as they describe the probability distribution of the whole image by means of local dependencies on a neighborhood structure.

Definition 4.1.1. *The* neighboring relationship *is a binary relationship on a lattice $\mathcal{S}$, such that for any two pixels i and j from $\mathcal{S}$, the following two conditions hold:*

(1) Pixel i is not a neighbor to itself (anti-reflexivity)

(2) If i is a neighbor of j, then j is also a neighbor of i (symmetry)

The neighboring relationship can also be represented by a simple undirected graph (without loops) where $\mathcal{S}$ is the set of vertices, and edges display a neighboring relation. Therefore two neighboring pixels are not necessarily geographically close to each other, but in most applications they are.

Definition 4.1.2.

- *The* neighborhood $N(i)$ *of pixel i is defined as the set*

$$N(i) = \{j \in \mathcal{S} \mid j \text{ neighbor of } i\}.$$

- *Let the* neighborhood system $\mathcal{N}$ *then be the set of all neighborhoods*

$$\mathcal{N} = \{N(i) \mid i \in \mathcal{S}\}.$$

- *A subset of sites C is called a* clique *when each pair of distinct sites are neighbors.*

The graph-theoretical representative of the set of cliques of pixel i is the complete subgraph containing vertex i. Also in graph-theory this is sometimes called a clique. By definition the empty set and the single site are cliques.

The simplest neighborhood system that is also used for mixture modeling is composed of $N(i) = \emptyset$. Only empty cliques exist and the sites are independent. A simple (not degenerated) choice for a neighborhood system on a regular lattice is the first order neighborhood system, also called a four-neighborhood system. Each (interior) site has four nearest neighbors as depicted in Figure 4.2. This neighborhood system has two clique types: single site cliques containing pixel i, and pair-site cliques containing i and one neighbor j, as depicted also in Figure 4.2.

Definition 4.1.3. *The random field Λ is said to be an* MRF *with respect to the neighborhood system $\mathcal{N}$, if for all $\lambda \in \mathcal{G}^{N \times M}$*

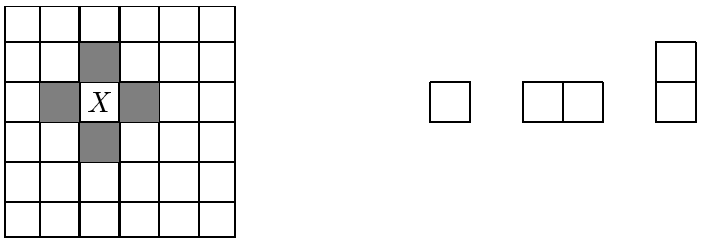


Figure 4.2: A first order neighborhood showing pixel X and its four nearest neighbors (in gray) on the left. The right-hand side displays the clique types: a single-site, a horizontal and a vertical pair-site clique.

(1) $P(\lambda) > 0$ *(positivity)*

(2) $P\left(\lambda_i|\lambda_{\mathcal{S}\setminus i}\right) = P(\lambda_i|\lambda_{N(i)})$ *(Markovianity)*

In an MRF only neighboring pixels may have direct interactions with each other. The conditional probabilities $P(\lambda_i|\lambda_{N(i)})$ of an MRF define these local characteristics of the image. One example for the first order neighborhood system will be described in the next section. First (or second) order neighborhood systems are widely used because of their simple form and their (relatively) low computational cost.

The Hammersley-Clifford theorem (Besag [1974]) establishes the equivalence between the local definition by a Markov random field and the global definition by a Gibbs random field. It provides a possibility for specifying the joint probability of a random field by means of conditional probabilities. This equivalence will be defined and proven subsequently by following Brémaud [1999].

Definition 4.1.4. *A random field* Λ *follows a* Gibbs distribution *if for all instances* $\lambda \in \mathcal{G}^{N\times M}$

$$P(\lambda) = \frac{1}{Z} e^{\sum\limits_{C\in\mathcal{C}} V_C(\lambda)} \tag{4.2}$$

holds, where Z *is a normalizing constant and for the collection of functions* $V_C : \Lambda^{\mathcal{G}^{N\times M}} \mapsto \mathbb{R} \cup \infty$ *two conditions are fulfilled:*

(1) $V_C = 0$ *if* C *is not a clique*

(2) for all λ, $\lambda' \in \mathcal{G}^{N\times M}$ *and all* $C \subset \mathcal{S}$

$$\text{if } \lambda(C) = \lambda'(C) \Rightarrow V_C(\lambda) = V_C(\lambda').$$

According to the Hammersley-Clifford theorem, an MRF is equivalent to a Gibbs Random Field (GRF):

Theorem 4.1.5 (Hammersley-Clifford). *The random field Λ is a Gibbs random field of the form* (4.1.4) *with potentials $\{V_C\}_{C\in\mathcal{S}}$ relative to the neighborhood system $\mathcal{N}$ if and only if Λ is a Markov random field with respect to the same neighborhood system.*

Proof. The proof of the Hammersley-Clifford theorem will largely follow Brémaud [1999]. First it will be shown that a Gibbs random field is always a Markov random field. Then the proof will proceed in the opposite direction.
'$\Rightarrow$' Let Λ be a Gibbs random field, then

$$\begin{aligned} P(\lambda_i|\lambda_{\mathcal{S}\setminus i}) &= \frac{P(\lambda)}{\sum_{\lambda_i'\in\mathcal{G}} P(\lambda')} = \frac{e^{\sum_{C\in\mathcal{C}} V_C(\lambda)}}{\sum_{\lambda_i'\in\mathcal{G}} e^{\sum_{C\in\mathcal{C}} V_C(\lambda')}} \\ &= \frac{e^{\sum_{C\in\mathcal{C},i\in C} V_C(\lambda)}\, e^{\sum_{C\in\mathcal{C},i\notin C} V_C(\lambda')}}{\left(\sum_{\lambda_i'\in\mathcal{G}} e^{\sum_{C\in\mathcal{C},i\in C} V_C(\lambda)}\right) e^{\sum_{C\in\mathcal{C},i\notin C} V_C(\lambda)}} \\ &\overset{(2)}{=} \frac{e^{\sum_{C\in\mathcal{C},i\in C} V_C(\lambda)}}{\sum_{\lambda_i'\in\mathcal{G}} e^{\sum_{C\in\mathcal{C},i\in C} V_C(\lambda)}}. \end{aligned}$$

The local conditional probabilities $P(\lambda_i|\lambda_{\mathcal{S}\setminus i})$ hereby only depend on cliques that contain pixel i, and hence depend only on the neighborhood $N(i)$.
'$\Leftarrow$' Let Λ be a Markov random field.

Additionally, let λ be a configuration in $\mathcal{G}^{N\times M}$ and A a subset of $\mathcal{S}$. Let λ^A denote the configuration coinciding with λ on A and with 0 outside of A. Then we define for A and λ

$$V_A(\lambda) = -\sum_{B\subset A} (-1)^{|A\setminus B|} \ln \frac{P(0)}{P(\lambda^B)}. \tag{4.3}$$

From Möbius' formula (see [Brémaud, 1999, p262]),

$$\ln \frac{P(0)}{P(\lambda)} \overset{A=\mathcal{S}}{=} -\sum_{B\subset\mathcal{S}} V_B(\lambda),$$

that is

$$P(\lambda) = P(0)e^{\sum_{A\subset\mathcal{S}} V_A(\lambda)}.$$

It remains to be shown that V_A fulfills conditions (1) and (2).

(2) V_A depends only on the values in A:
Let λ and λ' be such that $\lambda(A) = \lambda'(A)$, then for any $B \subset A$ holds $\lambda^B = \lambda'^B$ and by (4.3) $V_A(\lambda) = V_A(\lambda')$.

(1) $V_A = 0$ if A is not a clique of $\mathcal{S}$:
Let $j \in A$ be an arbitrary site, then

$$\begin{aligned} V_A(\lambda) &= -\sum_{B\subset A, j\notin B} (-1)^{|A\setminus B|} \ln \frac{P(0)}{P(\lambda^B)} - \sum_{B\subset A, j\in B} (-1)^{|A\setminus B|} \ln \frac{P(0)}{P(\lambda^B)} \\ &= \sum_{B\subset A\setminus j} (-1)^{|A\setminus B|} \left(\ln \frac{P(0)}{P(\lambda^{B+j})} - \ln \frac{P(0)}{P(\lambda^B)} \right) \\ &= \sum_{B\subset A\setminus j} (-1)^{|A\setminus B|} \ln \frac{P(\lambda^B)}{P(\lambda^{B+j})}. \end{aligned}$$

If $j \notin B$, then

$$\frac{P(\lambda^B)}{P(\lambda^{B+j})} \overset{\text{Markov}}{=} \frac{P(\lambda_j|\lambda^B_{N(j)})P(\lambda^B_{\mathcal{S}\setminus j})}{P(\lambda_j|\lambda^{B+j}_{N(j)})P(\lambda^{B+j}_{\mathcal{S}\setminus j})} = \frac{P(\lambda_j|\lambda^B_{N(j)})}{P(\lambda_j|\lambda^{B+j}_{N(j)})}$$

and therefore

$$V_A(\lambda) = \sum_{B\subset A\setminus j} (-1)^{|A\setminus B|} \ln \frac{P(\lambda_j|\lambda^B_{N(j)})}{P(\lambda_j|\lambda^{B+j}_{N(j)})}.$$

Analogously

$$\begin{aligned} V_A(\lambda) &= -\sum_{B\subset A} (-1)^{|A\setminus B|} \ln \frac{P(0)}{P(\lambda_j|\lambda^B_{N(j)})} \\ &= \sum_{B\subset A} (-1)^{|A\setminus B|} \ln P(\lambda_j|\lambda^B_{N(j)}) - \sum_{B\subset A} (-1)^{|A\setminus B|} \ln P(0) \\ &= \sum_{B\subset A} (-1)^{|A\setminus B|} \ln P(\lambda_j|\lambda^B_{N(j)}). \end{aligned}$$

Let us assume A is not a clique. Since $j \in A$, a site $i \in A$ exists, such that $i \notin N(j)$. Fixing such an i, it follows

$$\begin{aligned} V_A(\lambda) &= \sum_{B \subset A} (-1)^{|A \setminus B|} \ln P(\lambda_j | \lambda^B_{N(j)}) \\ &= \sum_{B \subset A \setminus i \setminus j} (-1)^{|A \setminus B|} \ln P(\lambda_j | \lambda^B_{N(j)}) + \sum_{B \subset A \setminus i \setminus j} (-1)^{|A \setminus B \setminus j|} \ln P(\lambda_j | \lambda^{B+j}_{N(j)}) \\ &\quad + \sum_{B \subset A \setminus i \setminus j} (-1)^{|A \setminus B \setminus i|} \ln P(\lambda_j | \lambda^{B+i}_{N(j)}) + \sum_{B \subset A \setminus i \setminus j} (-1)^{|A \setminus B \setminus i \setminus j|} \ln P(\lambda_j | \lambda^{B+i+j}_{N(j)}) \\ &= \sum_{B \subset A \setminus i \setminus j} (-1)^{|A \setminus B|} \ln \frac{P(\lambda_j | \lambda^B_{N(j)}) P(\lambda_j | \lambda^{B+i+j}_{N(j)})}{P(\lambda_j | \lambda^{B+i}_{N(j)}) P(\lambda_j | \lambda^{B+i+j}_{N(j)})}. \end{aligned}$$

Since $i \notin N(j)$, we have $P(\lambda_j | \lambda^B_{N(j)}) = P(\lambda_j | \lambda^{B+i}_{N(j)})$ and $P(\lambda_j | \lambda^B_{N(j)}) = P(\lambda_j | \lambda^{B+i+j}_{N(j)})$ and therefore $V_A(\lambda) = 0$.

□

Two specific forms of Markov or alternatively Gibbs random field models will be analyzed for TCA images during the next Sections, 4.2 and 4.3.

4.2 Auto-Logistic Model

Auto-models are a particular subclass of MRF models that involve only pairwise contextual constraints. If the labels take on one of two states: $\mathcal{G} = \{0, 1\}$ (or $\mathcal{G} = \{-1, 1\}$), they are termed auto-logistic models. A particular auto-logistic model is one that involves only a first order neighborhood system. This model is hereby a simple model among the MRF models and will be introduced now to illustrate the idea of MRFs and to obtain a starting point for the selection of a more complex MRF model.

After indicating the set-up of the auto-logistic model above, it will be properly defined in the first part of this section. Section 4.2.2 will introduce the Gibbs sampler and state the algorithm for the auto-logistic model. The results of these simulations together with a specification of the parameters of the auto-logistic model for TCA images can be found in Section 4.2.3. The last part of this section will consist of a discussion on extensions of the auto-logistic model and thereby lead us to the ensuing section on the FRAME model.

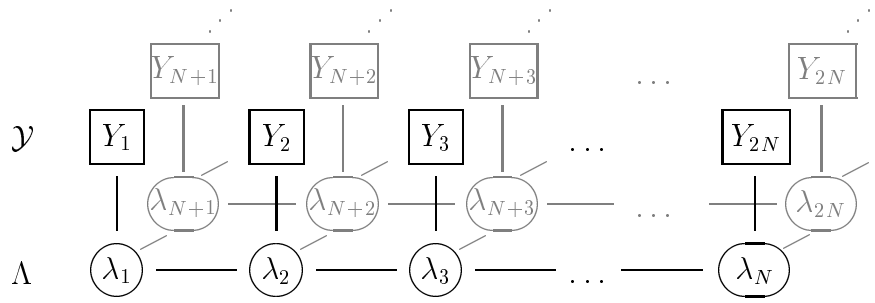


Figure 4.3: Conditional independence graph of the auto-logistic model. Boxes represent observed or fixed quantities, circles show unknowns.

4.2.1 Model Definition

The auto-logistic model (or variants of it) is explained in many probabilistic image analysis textbooks and articles. Following, it will be defined based on the book Winkler [1995], from which an interested reader can find a more embedded description.

The auto-logistic model assumes a first order neighborhood system with clique types as described on page 61 and in Figure 4.2. The conditional independence graph of the auto-logistic model is depicted in Figure 4.3. In the current application, the label of each site may take on two states: $\mathcal{G} = \{0, 1\}$. The simplest Gibbs potential (4.2) for the cost function is then defined by

$$U(\lambda) = \alpha \sum_{i \in \mathcal{S}} \lambda_i + \beta \sum_{i \in \mathcal{S}} \sum_{j \in N(i)} \lambda_i \lambda_j.$$

The first term influences the proportions of the two labels, and the second term the propensity of neighboring pixels to have the same label. This cost function U represents a homogenous, isotropic field. An auto-logistic field is called homogeneous when its parameters are independent of the pixel; i.e. $\alpha = \alpha_i$, $\beta = \beta_i$, ... for all $i \in \mathcal{S}$. An auto-logistic field is said to be isotropic at order l, if there is a common parameter assigned to the cliques belonging to the l-th neighborhood system (but not to smaller neighborhood systems). (Cross and Jain [1983])

The cost function above can therefore be generalized to form a nonhomogeneous field by varying α and β from site to site, or to model an anisotropic (directional) image by varying β between the horizontal and the vertical clique type. Since TCA images are clearly directional, the anisotropic gen-

eralization will be explored here. In the case of the auto-logistic model, it takes the following form:

$$U(\lambda) = \alpha \sum_{i \in \mathcal{S}} \lambda_i + \beta_v \sum_{(i,j) \in C_v} \lambda_i \lambda_j + \beta_h \sum_{(i,j) \in C_h} \lambda_i \lambda_j \tag{4.4}$$

where C_v and C_h comprise the set of all vertical and horizontal pair-site cliques, respectively. Applying the Hammersley-Clifford Theorem (4.1.5) to this cost function leads to the conditional probabilities:

$$P(\lambda_i | \lambda_{N(i)}) = \frac{e^{-\left(\alpha\lambda_i + \beta_v \sum\limits_{j:(i,j)\in\mathcal{C}_v} \lambda_i\lambda_j + \beta_h \sum\limits_{j:(i,j)\in\mathcal{C}_h} \lambda_i\lambda_j\right)}}{1 + e^{-\left(\alpha + \beta_v \sum\limits_{j:(i,j)\in\mathcal{C}_v} \lambda_j + \beta_h \sum\limits_{j:(i,j)\in\mathcal{C}_h} \lambda_j\right)}}$$

of the random field λ at sites i.

The auto-logistic model above with label set $\mathcal{G} = \{0, 1\}$ is equivalent to the Ising model using the label set $\mathcal{G} = \{-1, 1\}$. The first model can be transformed into the latter one by inserting

$$\lambda_i = \frac{1}{2}(\lambda_i' + 1) \qquad \text{and} \tag{4.5}$$

$$\alpha = 2\alpha' - 4\beta_h' - 4\beta_v', \qquad \beta_h = 4\beta_h',\ \beta_v = 4\beta_v'. \tag{4.6}$$

(See, for example, Guyon [1995].) The traditional model is the Ising model. In the literature, the auto-logistic model with first order neighborhood system and label set $\mathcal{G} = \{0, 1\}$ is sometimes also termed an Ising model.

4.2.2 Gibbs Simulation

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where a Markov chain $\lambda^{(t)}$ is simulated such that its stationary distribution is the distribution of interest $P(\lambda)$. (See Gilks et al. [1996] for an overview and details on the Gibbs sampler.)

The single site Gibbs sampler, for example, initializes $\lambda^{(0)}$ at time $t = 0$ and then updates each pixel by repeatedly sampling a candidate $\lambda_i^{(t+1)}$ from the full conditional, which reduces to

$$P\left(\lambda_i^{(t+1)} | \lambda_{\mathcal{S}\setminus i}^{(t)}\right) = P\left(\lambda_i^{(t+1)} | \lambda_{N(i)}^{(t)}\right)$$

under the Markov property. The transition probabilities $P(\lambda^{(t)}|\lambda^{(t-1)})$ are then guaranteed to converge to the stationary distribution $P(\lambda)$. (See Gilks et al. [1996] and references therein.)

When choosing a random initial image and a random updating order of the pixels, the Gibbs sampling algorithm can be stated as follows for the auto-logistic model:

Gibbs sampling algorithm for the auto-logistic model

---

1. input
   - initial white noise image $\lambda^{(0)}$
   - parameters of the auto-logistic model: α, β_v, β_h
2. **repeat** sufficiently often
3. **repeat** NM times ($NM = |\mathcal{S}|$ size of the image)
4. randomly select site $i \in \mathcal{S}$
5. set $\lambda_j^{(t+1)} = \lambda_j^{(t)}$ for all $j \neq i$
6. **for** each value g of component $\lambda_i^{(t+1)}$
7. calculate the new potential

$$U_i\left(\lambda_i^{(t+1)}\right) = \alpha\lambda_i^{(t+1)} + \beta_v\lambda_i^{(t+1)}\left(\sum_{j:(i,j)\in\mathcal{C}_h}\lambda_j^{(t+1)}\right) + \beta_h\lambda_i^{(t+1)}\left(\sum_{j:(i,j)\in\mathcal{C}_v}\lambda_j^{(t+1)}\right)$$

8. **for** each value g of component $\lambda_i^{(t+1)}$
   calculate the conditional probability
9. $$P\left(\lambda_i^{(t+1)} = g|\lambda_{N(i)}^{(t+1)}\right) = \frac{e^{-U_i\left(\lambda_i^{(t+1)}\right)}}{\sum_{g'=1}^{G} e^{-U_i(g')}}$$

13. set $\lambda_i^{(t+1)} = g$ with probability $P\left(\lambda_i^{(t+1)} = g|\lambda_{N(i)}^{(t+1)}\right)$

For 1,000 iterations, 128×128 pixels and two gray values, the algorithm above needs about one hour on Hydra[1]. An exemplary result of applying the Gibbs sampler for simulating a typical (high probability) image using the potential function (4.4) will be presented below.

[1] Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

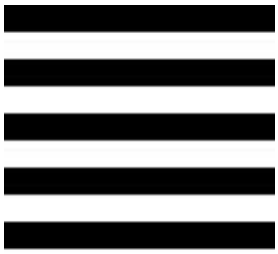


Figure 4.4: Example of an ideal TCA image

If the size of the image NM is large, it seems faster to use a coding method instead of a random updating order as pursued above. Since the academic examples in the following contain only $NM = 128 \times 128$ pixels, this possibility was not exploited in the current setting.

4.2.3 Application

In order to model tooth rings in TCA images, the auto-logistic model should be capable of generating a stack of black and white lines as depicted in Figure 4.4. The Lemma below helps selecting the parameters α, β_v and β_h in order to synthesize this ideal TCA image.

Lemma 4.2.1. *Let λ^0 be a random field of size $N \times M$ consisting of as many black as white lines of the same height $h > 3$ (like depicted in Figure 4.4). Then the joint probability $P(\lambda^0)$ using the cost function (4.4) can only be maximized among all random fields of the same size if*

(1) $\beta_v = 0$,

(2) $\beta_h < 0$ *and*

(3) $\alpha = -(\beta_h + \beta_v)$.

Proof. The parameters α, β_h and β_v need to be chosen such that for any image $\lambda \in \mathcal{G}^{N \times M}$

$$P(\lambda) \leq P(\lambda^0)$$

holds. This is the case if and only if

$$\forall \lambda \in \mathcal{G}^{N\times M} \qquad \alpha \sum_{i\in\mathcal{S}} \lambda_i^0 + \beta_h \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i^0\lambda_j^0 + \beta_v \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i^0\lambda_j^0 \leq \alpha \sum_{i\in\mathcal{S}} \lambda_i + \beta_h \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i\lambda_j + \beta_v \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i\lambda_j$$

$$\Leftrightarrow \forall \lambda \in \mathcal{G}^{N\times M} \qquad \frac{1}{2}\alpha|\mathcal{S}| + \beta_h(N-1)M + \beta_v\frac{h-1}{h}MN \leq \alpha \sum_{i\in\mathcal{S}} \lambda_i + \beta_h \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i\lambda_j + \beta_v \sum_{j:(i,j)\in\mathcal{C}_h} \lambda_i\lambda_j.$$

In particular, $P(\lambda) \leq P(\lambda^0)$ should hold for the following instances of λ

(1) Let $\lambda \in \mathcal{G}^{N\times M}$ be a random field consisting of as many black as white lines with height $h' \neq h$. (Image λ contains thicker or thinner lines than λ^0.)

$$\Rightarrow \frac{1}{2}\alpha|\mathcal{S}| + \beta_h(N-1)M + \beta_v\frac{h-1}{h}MN \leq \frac{1}{2}\alpha|\mathcal{S}| + \beta_h(N-1)M + \beta_v\frac{h'-1}{h'}MN$$

$$\Rightarrow \beta_v\frac{h-1}{h} \leq \beta_v\frac{h'-1}{h'} \text{ for any } h'$$

This is only true for $\beta_v = 0$. (With $\beta_v < 0$ thicker rings would be favored ($h > h'$).)

(2) Let us set $\lambda = \lambda^0$ and change one column in λ: switch black and white labels.

$$\Rightarrow \frac{1}{2}\alpha|\mathcal{S}| + \beta_h(N-1)M + \beta_v\frac{h-1}{h}MN \leq \frac{1}{2}\alpha|\mathcal{S}| + \beta_h(N-3)M + \beta_v\frac{h-1}{h}MN$$

$$\Rightarrow \beta_h(N-1) \leq \beta_h(N-3)$$

This is true for any $\beta_h \leq 0$.

(3) Let us consider the labeling $\mathcal{G} = \{-1, 1\}$ now. Then the labels λ and the parameters α, β_h, β_v can be transformed into the labels λ' and the parameters α', β_h', β_v' according to Equations (4.5) and (4.6).
Set the new label image $\lambda' = \lambda^{0'}$ and add to λ' one white or respectively one black row at a transition from a white to a black line.

$$\Rightarrow \quad 2\beta_h'(N-1)M + 2\beta_v'\frac{h-1}{h}MN \leq \pm\alpha' N + 2\beta_h'(N-1)M + 2\beta_v'\frac{h-1}{h}MN$$

$$\Rightarrow \quad 0 \leq \pm\alpha' N$$

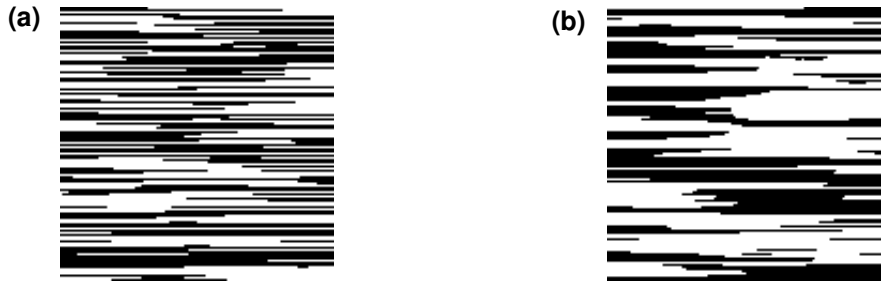


Figure 4.5: Typical images simulated by the Gibbs sampler using the anisotropic auto-logistic potential function in Equation (4.4) with $\beta_h = -10$, $\beta_v = 0$ and $\alpha = 10$ in image (a) and $\beta_v = -0.1$ and $\alpha = 10.1$ in image (b).

This is only true for $\alpha' = 0$ and hereby for $\alpha = -(\beta_h + \beta_v)$.

□

For a 128x128-pixel image, the parameters $\beta_h = -10$, $\beta_v = 0$ and $\alpha = 10$ were chosen to simulate an image close to the ideal TCA image in Figure 4.4. The typical (high probability) image generated by the Gibbs sampler (see page 68) is depicted in Figure 4.5 (a).

The images in Figure 4.4 and 4.5 (a) do not resemble each other. There might be two options to resolve this discrepancy within the setting of the auto-logistic model: Choosing a smaller β_h or a negative β_v. The first approach of choosing a smaller horizontal clique potential would reinforce the propensity to create all black or all white rows. The latter approach of choosing a smaller vertical clique potential would strengthen the attachment of pixels on top of each other. Practically, this would simulate thicker black and white lines than in the image of Figure 4.5 (a). Theoretically, this will converge to an all white or an all black image (see proof of Lemma 4.2.1, Point (1); so results as depicted in Figure 4.5 (b) probably did not converge yet. In any case, the anisotropic auto-logistic model keeps the blob-like structure that seems to be intrinsic to this model, but is not characteristic to TCA images.

4.2.4 Discussion – Auto-Logistic Models

The simulations presented above show that the auto-logistic model, the simplest among the Markov random field models cannot express the prior con-

(a)

(b)

-0.1	-0.1	-0.1
-10	10.3	

Figure 4.6: Image (a) is a typical image simulated by the Gibbs sampler using the anisotropic auto-logistic potential function in Equation (4.4) with parameters as depicted in (b).

victions we have about TCA images. But the results also indicate that the hidden Markov random field in general – if extended – might be an appropriate model for TCA images.

Two generalizations are of potential interest. The first generalization concerns the incorporation of more than two gray levels. For an unordered label set, this would result in a so-called Potts or multilevel logistic model. For an ordered label set, this could be the auto-binomial model. But neither approach will be able to express the prior conviction of (nearly) parallel running lines, because the blob-like structure of the auto-logistic model will remain.

The more promising extension regards the order of the neighborhood system. As Cross and Jain [1983] mention, the auto-logistic model can only fit micro-textures well. The local structure that is present in TCA images though, is clearly larger than two pixels. The clique size therefore needs to be increased in order to capture macro texture features and to model more regular structures.

An example of a simulation for the auto-logistic model using a second order neighborhood system is depicted in Figure 4.6 (a), with the parameter specification in Figure 4.6 (b). The reader can clearly see the trend toward a more regular structure, compared to the results of the auto-logistic model using the first order neighborhood system. But generalizing the auto-logistic model even further, to finally express the prior conviction depicted in the ideal TCA image 4.4, will result in a vast number of parameters that need to be estimated in order to fit the model to TCA images. Additionally the

simulation of a typical image will slow down considerably, because each iteration of the Gibbs sampler requires the computation of the full conditional $P\left(\lambda_i^{(t+1)} = g|\lambda_{N(i)}^{(t+1)}\right)$ for all gray values g. Therefore, this approach of extending the neighborhood system in the framework of the auto-logistic model will not be pursued further.

Instead, filters will be introduced subsequently to capture the local relationship between pixels of a clique.

4.3 FRAME Model

FRAME is the abbreviation for **F**ilters, **R**andom Fields **a**nd **M**aximum **E**ntropy. It is a Markov random field model whose energy potential incorporates filter features and thereby elegantly combines two important areas of texture analysis: MRF modeling and filtering theory. The FRAME model outperforms the computationally inefficient setting of the auto (-logistic) or the Ising model and explains its parameters (filters) more intuitively. It can model large-scale texture and can thus take into account long-range autocorrelations between labels and the periodicity of the placement of tooth rings in TCA images.

The FRAME model was mainly developed in Zhu et al. [1997], Zhu et al. [1998] and Zhu and Mumford [1997]. In Section 4.3.1, the Gibbs potential of the FRAME model will be defined and its properties explained. An EM-type procedure to coherently estimate the parameters of this prior distribution and the parameters of the observable random field as well as the underlying labels will be given in Section 4.3.2. Section 4.3.3 describes how to estimate confidence intervals of the parameters and hereby provides a measure of the goodness of fit of the model to TCA images. Section 4.3.4 will give application-driven choices of the FRAME ingredients. A typical image that illustrates the capacity of the FRAME approach is displayed in Section 4.3.5 together with the algorithm to generate it. One result of applying the FRAME in the context of HMRFs to a TCA image will be presented in Section 4.3.6, while Chapter 6.1 gives more results. Chapter 6 will also explain extensions and different specifications of the Gaussian hidden FRAME model and the EM algorithm.

The notation in some parts of the current section will slightly differ from the rest of this document. The reason is the intrinsic two-dimensional structure of image filtering that will be preserved. The site i will therefore be

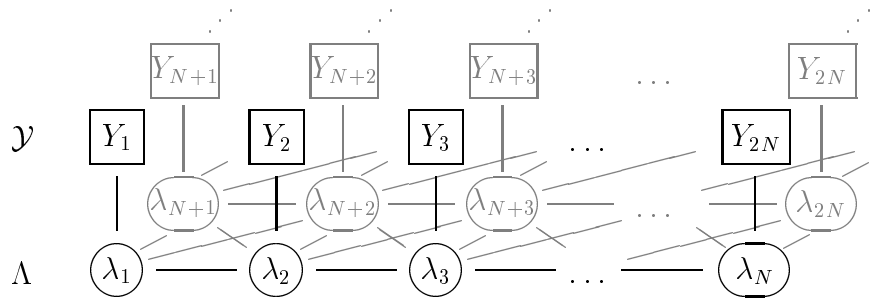


Figure 4.7: Conditional independence graph of the FRAME model (simplified). Boxes represent observed or fixed quantities, circles show unknowns.

named (x, y) in Section 4.3.4 and some other parts of this section; where x specifies the row and y the column.

4.3.1 Model Definition and Properties

It is assumed a priori that the underlying label image λ is distributed according to the Gibbs distribution

$$P(\lambda) = \frac{1}{Z} e^{\sum\limits_{i \in \mathcal{S}} \phi[(F_T * \lambda)(i)]}, \tag{4.7}$$

where Z is the normalizing constant. Moreover, the energy function

$$U(\lambda) = \sum_{i \in \mathcal{S}} \phi\left[\left(F_T * \lambda\right)(i)\right]$$

involves one filter F_T, that is known up to the parameter T and whose filter responses $(F_T * \lambda)(i)$ to λ at pixels i are evaluated pixelwise by the potential function ϕ.

There are two appealing properties of the FRAME model:

(1) The FRAME model is a generalization of traditional MRF models such as the auto-logistic model explained in Section 4.2. The Hammersley-Clifford theorem, mentioned in Section 4.1, proves in general that $P(\lambda)$ from equation (4.7) is distributed according to an MRF with

$$P(\lambda_i | \lambda_{N(i)}) = \frac{e^{\sum\limits_{C \in \mathcal{C}, i \in C} \phi[(F_T * \lambda)(C)]}}{\sum\limits_{\lambda'_i \in \mathcal{G}} e^{\sum\limits_{C \in \mathcal{C}, i \in C} \phi[(F_T (F * \lambda')(C)]}}.$$

For this purpose the neighborhood $N(x, y)$ of pixel $i = (x, y)$ in Definition 4.1.2 needs to be defined as

$$N(x, y) = \{(x', y') \in \mathcal{S} \mid |x' - x| < |F_T| \text{ and } |y' - y| < |F_T|\},$$

where $|F_T|$ denotes the number of rows and columns respectively of filter F_T. The clique $C(x, y)$ of filter F_T containing pixel $i = (x, y)$ is defined similarly as the set

$$C(x, y) = \{(x', y') \in \mathcal{S} \mid F_T(x' - x, y' - y) \neq 0\}.$$

For these cliques the Gibbs potential is defined by $V_C(\lambda) = \phi[(F_T * \lambda)(C)]$, where $(F_T * \lambda)(C)$ denotes the convolution positioned at clique C and all other potentials are set $V_C(\lambda) = 0$.

(2) The filter F_T and the potential function ϕ completely determine $P(\lambda)$. In order to fit the FRAME model (embedded in an HMRF) to a given image, only the filter and the potential function need to be specified.

The (simplified) conditional independence graph of the FRAME model is depicted in Figure 4.7.

According to Property (1), the FRAME model elegantly combines two important areas of texture analysis: MRF modeling and filtering theory. Furthermore, it can even be applied to large-scale texture. The interaction of filters and potential functions may in general create a wide variety of texture patterns.

The choice of the parametric family F_T and the function ϕ is driven by the application. In Section 4.3.4, our choice will be explained. Section 4.3.5 will illustrate the interplay between filter and potential function. But before, Section 4.3.2 will describe the compound estimation procedure for an HMRF, comprising any FRAME model (4.7) for the unknown label field.

4.3.2 Parameter Estimation and Segmentation

The underlying label field λ can be estimated by Bayes' labeling when both of the following are known: the hidden random field and the observed random field generating a pattern (Li [2001]). This labeling $\hat{\lambda}$ (often called maximum a posterior [MAP] solution) is the one maximizing the posterior probability of the labels given the data

$$P(\lambda|Y) \propto P(\lambda) f(Y|\lambda).$$

In this setting $P(\lambda)$ is often called prior distribution.

Since the labeling Λ is Gibbs distributed according to the FRAME model in (4.7) and the random variables $\mathcal{Y}_i$ are conditionally independent, this can be rewritten as:

$$\begin{aligned} P(\lambda)f(Y|\lambda) &= P(\lambda)\prod_{i\in\mathcal{S}} f\left(Y_i|\lambda_i\right) \\ &= \frac{1}{Z}e^{\left(\sum\limits_{i\in\mathcal{S}}\phi[(F_T*\lambda)(i)]\right)}\frac{1}{Z'}e^{-\left(\sum\limits_{i\in\mathcal{S}}\frac{\left(Y_i-\mu_{\lambda_i}\right)^2}{2\sigma^2_{\lambda_i}}\right)}. \end{aligned}$$

If one defines

$$U(\lambda) = \sum_{i\in\mathcal{S}}\left[\left(F_T * \lambda\right)(i)\right] \quad \text{and} \quad U(Y|\lambda) = -\sum_{i\in\mathcal{S}}\frac{(Y_i-\mu_{\lambda_i})^2}{2\sigma^2_{\lambda_i}},$$

the MAP solution $\hat{\lambda}$ is the one maximizing $U(\lambda)$:

$$\hat{\lambda} = \arg\max_{\lambda\in\mathcal{G}^{N\times M}}\{U(\lambda)+U(Y|\lambda)\}.$$

However, the hidden random field and the observed random field are not completely known; their parameters $\boldsymbol{\theta}$ and T have to be estimated as well. In order to estimate $\boldsymbol{\theta}$ and T, the maximum likelihood estimates (MLE) $\hat{\boldsymbol{\theta}}$ and $\hat{T}$ can, in principle, be found by maximizing the likelihood function

$$L(\boldsymbol{\theta},T|Y) = \sum_{\lambda\in\mathcal{G}^{N\times M}} P(\lambda,Y|\boldsymbol{\theta},T).$$

The maximization of this likelihood is intractable because of the size of the label space $\mathcal{G}^{N\times M}$.

The EM algorithm, introduced for mixture modeling in Chapter 3.2, is a widely used technique for solving this kind of problem. We will mostly follow Zhang et al. [2001] to describe this algorithm for Gaussian hidden Markov random fields.

In the case of a Gaussian random field, the EM algorithm is reduced to the two updating formulas

$$\begin{aligned}
\mu_g^{(t)} &= \frac{\sum_{i\in\mathcal{S}} Y_i P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i\in\mathcal{S}} P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)} \\
\left(\sigma_g^{(t)}\right)^2 &= \frac{\sum_{i\in\mathcal{S}} \left(Y_i - \mu_g^{(t)}\right)^2 P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i\in\mathcal{S}} P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}.
\end{aligned}$$

This was indicated in Chapter 3.2. Additionally T has to be updated by

$$T^{(t)} = \arg\max_{\{T\}} \sum_{i\in\mathcal{S}} \sum_{g=0}^{G} P\Big(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\Big) \log P\big(\lambda_i = g|\lambda_{N(i)}, T\big).$$

For a nonhomogeneous random field, one can update $T = (T_1, T_2, \ldots, T_{NM})$ for each pixel $i \in \mathcal{S}$ by:

$$\begin{aligned}
T_i^{(t)} = \arg\max_{\{T_i\}} \sum_{g=0}^{G} & P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \\
& \log P\left(\lambda_i = g|\lambda_{N(i)}, T_i, T_{N(i)}^{(t-1)}\right). \qquad (4.8)
\end{aligned}$$

But the conditional probabilities $P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ in the above formulas are not available in closed form (as in Equations (3.8) and (3.9)). They could be evaluated by simulation (i.e., the MCMC algorithm), but this is again not practical computationally because of the size of the neighborhood $N(i)$ and the size of the images in the TCA application. The alternative approach we suggest is based on the approximation

$$P(\lambda) \approx \prod_{i\in\mathcal{S}} P\left(\lambda_i|\tilde{\lambda}_{N(i)}\right) \qquad (4.9)$$

by some fixed configuration $\tilde{\lambda}_{N(i)}$. If (4.9) is a valid probability distribution, the E-step in Equation (3.2) changes to

$$\begin{aligned} & E\left[\log P(\lambda, Y|\boldsymbol{\theta}, T)|Y, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right] \\ & \quad \approx \sum_{i\in\mathcal{S}}\sum_{g=0}^{G} \log\left(P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, \boldsymbol{\theta}, T\right) f(Y_i|\lambda_i, \boldsymbol{\theta}, T)\right) \\ & \qquad \cdot P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \\ & \quad = \sum_{i\in\mathcal{S}}\sum_{g=0}^{G} \left(\log P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T\right) + \log f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}\right)\right) \\ & \qquad \cdot \frac{P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}\right) f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}\right)}{\sum_{g=0}^{G} P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}\right) f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}\right)} \end{aligned}$$

The parameter estimates can therefore be updated by

$$\mu_g^{(t)} = \frac{\sum_{i\in\mathcal{S}} Y_i P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i\in\mathcal{S}} P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)} \tag{4.10}$$

$$\left(\sigma_g^{(t)}\right)^2 = \frac{\sum_{i\in\mathcal{S}} \left(Y_i - \mu_g^{(t)}\right)^2 P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i\in\mathcal{S}} P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)} \tag{4.11}$$

$$T^{(t)} = \arg\max_{\{T\}} \sum_{i\in\mathcal{S}}\sum_{g=0}^{G} P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \cdot \log P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T\right) \tag{4.12}$$

In this application the configuration $\tilde{\lambda}$ in Equation (4.9) is chosen according to the mean field approximation theory (Celeux et al. [2003]). The mean field approximation sets $\tilde{\lambda}$ to the expected values of the label image:

$$\tilde{\lambda}_j = E[\lambda_j] \text{ for all } j \in N(i).$$

The product $\prod_{i \in \mathcal{S}} P(\lambda_i | E[\lambda_{N(i)}])$ is then a valid probability distribution and minimizes the Kullback-Leibler divergence to the true prior distribution $P(\lambda)$ among all products of this kind.

The expected values are computed iteratively to approximate $P(\lambda)$ and the resulting EM algorithm for a homogeneous field then takes the following form:

EM algorithm using mean field approximation for fitting a hidden FRAME model to TCA images

---

1. input TCA image Y

Initialization
initialize label configuration $\lambda^{(0)}$ by thresholding:
2. **for** each site i
3. $\lambda_i^{(0)} = g$ **if** $Y_i \leq a_{g+1}$
   for some values $a = (a_1, \ldots, a_G, \infty)$
initialize parameters by
4. **for** $g = 0 : G$
5. $\mu_g^{(0)} = \frac{1}{n_g} \sum_{\lambda_i^{(0)} = g} Y_i$
6. $\left(\sigma_g^{(0)}\right)^2 = \frac{1}{n_g} \sum_{\lambda_i^{(0)} = g} \left(Y_i - \mu_g^{(0)}\right)^2$
   with $n_g = |\{i \in \mathcal{S} \mid \lambda_i^{(0)} = g\}|$

Updating
7. **for** $t = 1 : t_{max}$
   update label image $\lambda^{(t)}$ by
8. $\langle \lambda \rangle = \lambda^{(t-1)}$
9. **for** each site i (randomly permuted)
10. **for** $g = 0 : G$
11. calculate the conditional probability

$$f\left(Y_i | \lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}\right) \propto e^{-\frac{\left(Y_i - \mu_g^{(t-1)}\right)^2}{2\left(\sigma_g^{(t-1)}\right)^2}}$$

12. approximate the prior energy and probability by

$$U\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right) \approx \sum_{j:\, i\in C(j)} \phi\left[\left(F_{T^{(t-1)}} * \lambda\right)(j)\right]$$

$$P\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right) = \frac{e^{U\left(\lambda_i=g|\lambda_{N(i)},T^{(t-1)}\right)}}{\sum_{g=0}^{G} e^{U\left(\lambda_i=g|\lambda_{N(i)},T^{(t-1)}\right)}}$$

10. calculate the posterior probability

$$P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) = \frac{f\left(Y_i|\lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}\right) P\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right)}{\sum_{g=0}^{G} f\left(Y_i|\lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}\right) P\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right)}$$

11. calculate the expected label

$$\langle\lambda_i\rangle = \frac{\sum_{g=0}^{G} g \cdot P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{g=0}^{G} P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}$$

12. set $\lambda^{(t)} = \langle\lambda\rangle$

update parameters

13. **for** $g = 0 : G$
14. update μ_g according to Equation (4.10)
15. update $\left(\sigma_g^{(t)}\right)^2$ according to Equation (4.11)
16. update $T \in \{T^{(t-1)} - 1, T^{(t-1)}, T^{(t-1)} + 1\}$ according to Equation (4.12)

The initialization of λ and the sequential updating of the labels were chosen according to the recommendations in Celeux et al. [2003]. The number of iterations t_{max} was selected such that the absolute value of the relative change in the log-likelihood

$$L(\boldsymbol{\theta}^{(t-1)}, T^{(t-1)}|Y) = \sum_{i\in\mathcal{S}} \log \sum_{g=0}^{G} P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)$$

over the last five iterations is less than 10^{-6}. Segmentation can finally be carried out by exploiting the posterior probability of the labels given the data $P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}\right)$, or equivalently, the mean field $E[\lambda]$ of the last iteration:

$$\lambda_i = \arg\max_{g \in \mathcal{G}} P\left(\lambda_i = g|Y_i, \lambda_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}\right), \text{ or}$$

$$\lambda_i = \arg\min_{g \in \mathcal{G}} |E[\lambda_i] - g|.$$

The EM algorithm generally converges to a local maximum likelihood estimate, as mentioned in Chapter 3.2. If the likelihood function is not concave, this local maximum might not coincide with the global maximum, depending on the starting values. As most applications do not involve concave maximization functions, this is a well-known problem of the EM algorithm that can be circumvented by MCMC type steps or algorithms. As mentioned previously, these methods are computationally too time-consuming in TCA applications. But experiments performed in Chapters 6.1 and 6.4 indicate that the chosen starting values lead to reasonable results. Chapter 6.4.1 discusses additional convergence issues connected to the EM algorithm and TCA image analysis.

4.3.3 Standard Error Calculation

To statistically specify the uncertainty involved with estimates obtained from an image, the distribution of parameter estimates could be assessed by assuming that the experiment is repeated under identical conditions; i.e., repeatedly drawing the TCA image. Subsequently it will be shown how measures of this hypothetical distribution can be derived theoretically and practically; and the extended applications in Chapter 6.1 will show how applicable statements can be derived.

An invaluable tool for assessing the distribution of estimates is the normal approximation for the distribution of Maximum Likelihood Estimates (MLE). This is detailed, for example, in Guyon [1995]. Under regularity conditions it can be proven that the MLE $(\hat{\boldsymbol{\theta}}, \hat{T})$ is asymptotically normal, saying that the distribution of $\left((\hat{\boldsymbol{\theta}}, \hat{T}) - (\boldsymbol{\theta}, T)\right)\sqrt{n}$ weakly converges to a multivariate normal distribution with mean zero as the sample size n increases. This means that if the sample size is large enough, all information concerning

the parameters $(\boldsymbol{\theta}, T)$ are summarized in the MLE $(\hat{\boldsymbol{\theta}}, \hat{T})$ and the covariance matrix, which indicates the precision with which the MLE can be determined. The covariance matrix will therefore play a central role in this section and McLachlan and Peel [2000] will be followed to approximate it.

The covariance matrix of maximum likelihood estimates is asymptotically equal to the inverse of the Fisher expected information matrix

$$\mathcal{I}(\boldsymbol{\theta}, T) = E_{(\boldsymbol{\theta},T)} \left(\frac{\partial \log L(\boldsymbol{\theta}, T|Y)}{\partial \boldsymbol{\theta}, T} \left(\frac{\partial \log L(\boldsymbol{\theta}, T|Y)}{(\partial(\boldsymbol{\theta}, T)} \right)^T \right),$$

which in turn can be estimated by the observed information matrix

$$I(\hat{\boldsymbol{\theta}}, \hat{T}, Y) = - \left. \frac{\partial^2 \log L(\boldsymbol{\theta}, T|Y)}{\partial(\boldsymbol{\theta}, T)\partial(\boldsymbol{\theta}, T)^T} \right|_{\hat{\boldsymbol{\theta}}, \hat{T}}. \tag{4.13}$$

The observed information matrix is a consistent estimator of the Fisher expected information matrix and is also convenient computationally, because the calculation of expected values is time-consuming. In the context of TCA image analysis, the estimation of expected values is computationally prohibitive, because of the size of the label space and the size of the neighborhood.

The computation of the Hessian of the (incomplete-data) log-likelihood for the observed information matrix in (4.13) involves the conditional probabilities $P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$. As explained in Section 4.3.2, these probabilities are not available in closed form, but can be approximated using mean field theory (see Equation (4.9) and thereafter). For coherence, the same approximations will be reused to realize the computation of the covariance matrix.

In the case of the Gaussian hidden FRAME model

$$\begin{aligned} f(Y) &= \sum_{\lambda \in \mathcal{G}^{N \times M}} P(Y, \lambda|\theta, T) \\ &= \sum_{\lambda \in \mathcal{G}^{N \times M}} \left(\frac{1}{Z} e^{\left(\sum_{i \in \mathcal{S}} \phi[(F_T * \lambda)(i)] \right)} \prod_{i \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma_{\lambda_i}} e^{-\frac{\left(Y_i - \mu_{\lambda_i}\right)^2}{2\sigma^2_{\lambda_i}}} \right), \end{aligned}$$

let us use two labels (say $\mathcal{G} = \{0, 1\}$) and a common variance σ^2, such that the parameter vector is $(\boldsymbol{\theta}, T) = (\mu_0, \mu_1, \sigma^2, T)$. Then the observed information matrix contains the following elements on the main diagonal:

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{11} &= -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta}, T|Y\right)}{\left(\partial\mu_0\right)^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx -\sum_{i\in\mathcal{S}} \frac{\left(\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right)^2 - \frac{1}{\hat{\sigma}^2}\right) P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)} \\
&\quad + \sum_{i\in\mathcal{S}} \left(\frac{\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right) P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)} \right)^2 ,
\end{aligned}
$$

$$
I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{22} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta}, T|Y\right)}{\left(\partial\mu_1\right)^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \text{ analogously ,}
$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{33} &= -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta}, T|Y\right)}{\left(\partial\sigma^2\right)^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} \left[\frac{3}{4\hat{\sigma}^4} - \frac{3\left(Y_i-\hat{\mu}_{\lambda_i}\right)^2}{2\hat{\sigma}^6} + \frac{\left(Y_i-\hat{\mu}_{\lambda_i}\right)^4}{4\hat{\sigma}^8}\right] P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)} \\
&\quad + \sum_{i\in\mathcal{S}} \left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}} \left(-\frac{1}{2\hat{\sigma}^2} + \frac{\left(Y_i-\hat{\mu}_{\lambda_i}\right)^2}{2\hat{\sigma}^4}\right) P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, \hat{T}\right)} \right)^2 ,
\end{aligned}
$$

and

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{44} = & -\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\left(\partial T\right)^2}\Big|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
\approx & -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}}\left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right)^2 \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right) \\
& -\sum_{i\in\mathcal{S}}\left(\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right)^2 \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2,
\end{aligned}
$$

as well as the following mixed derivatives:

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{12} &= I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{21} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_0\partial\mu_1}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx \sum_{i\in\mathcal{S}} \frac{\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right) P\left(Y_i|\lambda_i=0,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i=0|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
&\qquad \cdot \frac{\left(\frac{Y_i-\mu_1}{\sigma^2}\right) P\left(Y_i|\lambda_i=1,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i=1|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)},
\end{aligned}
$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{13} &= I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{31} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_0\partial\sigma^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx -\sum_{i\in\mathcal{S}} \frac{\left[\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_0)^2}{2\hat{\sigma}^4}\right)-\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^4}\right] P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
&\quad + \sum_{i\in\mathcal{S}} \frac{\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\left(\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2} \\
&\qquad \cdot \frac{\sum\limits_{\lambda_i\in\mathcal{G}}\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4}\right) P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{},
\end{aligned}
$$

$$
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{23} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{32} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_1\partial\sigma^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \text{ analogously ,}
$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{14} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{41} &= -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_0\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx \sum_{i\in\mathcal{S}}\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right) \\
&\quad\cdot\left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\left(\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2}-\frac{\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right), \\
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{34} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{43} &= -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\sigma^2\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx \sum_{i\in\mathcal{S}}\frac{\sum\limits_{\lambda_i\in\mathcal{G}}\left[\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right]}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
&\quad-\sum_{i\in\mathcal{S}}\left[\frac{\sum\limits_{\lambda_i\in\mathcal{G}}\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right. \\
&\quad\left.\cdot\frac{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right],
\end{aligned}
$$

and

$$I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{24} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{42} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_1\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}}$$

analogously to $I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{14}$.

Details of the above derivations can be found in Appendix A, where derivatives for the Gaborcosine filter family and the absolute valued potential function are also given.

A central measure for the behavior of a single parameter estimate is the standard deviation of its distribution. It is often referred to as standard error (SE) and is derived from the covariance matrix as follows. The standard error of the rth parameter estimate $(\hat{\boldsymbol{\theta}},\hat{T})_r$ is (asymptotically) equal to the square root of the rth element on the main diagonal of the inverse observed information matrix:

$$SE\left(\left(\hat{\boldsymbol{\theta}},\hat{T}\right)_r\right) = \sqrt{\left(I^{-1}(\hat{\boldsymbol{\theta}},\hat{T})\right)_{rr}}.$$

Let $g(\boldsymbol{\theta},T)$ now be a one-to-one transformation of the parameters $(\boldsymbol{\theta},T)$, like the number of lines in TCA images in the application in Section 4.3.6, which is a multiple of $\frac{1}{T}$. Then the standard error of the new parameter can be approximated by the delta method:

$$SE\left(g\left(\hat{\boldsymbol{\theta}},\hat{T}\right)\right) \approx \sqrt{\left(\left.\frac{dg(\boldsymbol{\theta},T)}{d(\boldsymbol{\theta},T)}\right|_{(\hat{\boldsymbol{\theta}},\hat{T})}\right)^T I^{-1}(\hat{\boldsymbol{\theta}},\hat{T})\left(\left.\frac{dg(\boldsymbol{\theta},T)}{d(\boldsymbol{\theta},T)}\right|_{(\hat{\boldsymbol{\theta}},\hat{T})}\right)}.$$

See, for example, Kalbfleisch [1985] for the above and subsequent formulas.

A useful measure to give upper and lower bounds of estimates are confidence intervals. The $100p\%$ confidence interval of the estimate $g(\boldsymbol{\theta},T)$ expresses the following: if an image is drawn repeatedly with $(\boldsymbol{\theta},T)$ fixed, then the $100p\%$ confidence interval would include the true parameter in $100p\%$ of the cases. Using the normal approximation and the approximations for SEs above, the $100p\%$ confidence interval of an estimate $g(\hat{\boldsymbol{\theta}},\hat{T})$ can be approximated by

$$g\left(\hat{\boldsymbol{\theta}},\hat{T}\right) \pm z_{\frac{1-p}{2}} \cdot SE\left(g\left(\hat{\boldsymbol{\theta}},\hat{T}\right)\right),$$

where $z_{\cdot}$ denotes the normal quantile.

Subsequently, the 95% confidence interval of the estimated number of tooth tings of a TCA image will be used in order to express the randomness within the image and to assess the goodness of fit of the model to this image. In general, one can say that confidence intervals do not express the uncertainty involved with the TCA method. In order to estimate this uncertainty, one needs to assess the variance of estimates retrospectively for a specific sample of TCA images. More detailed arguments and experiments will be presented in the applications in Chapter 6.1.

4.3.4 Filter Family and Potential Function Specification

Filtering theory has been well recognized in texture analysis at least since 1991, when Jain and Farrokhnia [1991] was published. It is based on psychophysical findings in human vision first proposed in Campbell and Robson [1968]. Marčelja [1980] has shown that two-dimensional Gabor functions

$$G_{T,\alpha}(x,y) = c_1 \cdot e^{\frac{-\left(rx'^2+y'^2\right)}{2T^2}} \cdot e^{-i\frac{2\pi}{T}x'} \tag{4.14}$$

closely conform to the receptive field profiles of simple cells in the striate cortex.

We define the filter $F_{T,\alpha}$ on the basis of the real valued, even-symmetric Gabor function:

$$Gcos_{T,\alpha}(x,y) = c_1 \cdot e^{\frac{-\left(rx'^2+y'^2\right)}{2T^2}} \cos\left(\frac{2\pi}{T}x'\right); \tag{4.15}$$

with $x' = x\cos\alpha + y\sin\alpha$, $y' = -x\sin\alpha + y\cos\alpha$, $r = 4$ being the aspect ratio and c_1 being a normalizing factor. The Gaborcosine function above is an elongated Gaussian bell multiplied by a cosine wave, where parameter T changes the wavelength and α determines the orientation of the cosine wave.

The filter $F_{T,\alpha}$ will be generated from the Gaborcosine function (4.15) by sampling at discrete integer values $x, y \in \{-\frac{4}{5}T, \ldots, \frac{4}{5}T\}$ and choosing the constant c_1 such that:

$$\sum_{(x,y):Gcos_{T,\theta}(x,y)>0} Gcos_{T,\theta}(x,y) = 1.$$

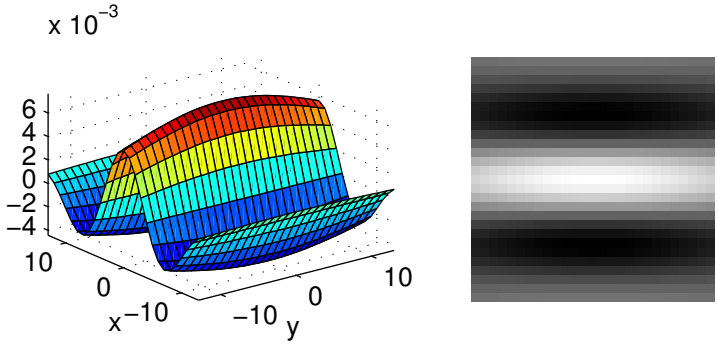


Figure 4.8: 3-D surface and image of a Gaborcosine function with parameters $T = 16$ and $\alpha = 0$.

Because of the finite filter size and discrete sampling we additionally choose a constant c_2 such that

$$\sum_{(x,y)} (Gcos_{T,\theta}(x,y) - c_2) = 0$$

and for $x, y \in \{-\lceil \frac{4}{5}T \rceil, \ldots, \lceil \frac{4}{5}T \rceil\}$ set

$$F_{T,\alpha}(x,y) = Gcos_{T,\theta}(x,y) - c_2.$$

For example, Figure 4.8 shows the Gaborcosine filter for $T = 16$, $\alpha = 0$ and in the range of $x, y \in [-13, 13]$. This filter can capture waves or lines of width 16 and an orientation of 0°.

Definition 4.3.1. *The discrete two-dimensional* convolution *of the input λ with a filter F at location (x, y) is defined by*

$$(F * \lambda)(x,y) = \sum_{x'=-m}^{m} \sum_{y'=-n}^{n} F(x',y')\lambda(x+x', y+y')$$

where $2m+1$ and $2n+1$ are the width respectively height of the filter.

The convolution hereby expresses the amount of overlap between the two functions F and λ. (Jain [1989])

In the application for TCA images the orientation is fixed to $\alpha = 0$ (unless stated otherwise), which is the main direction of tooth rings. In

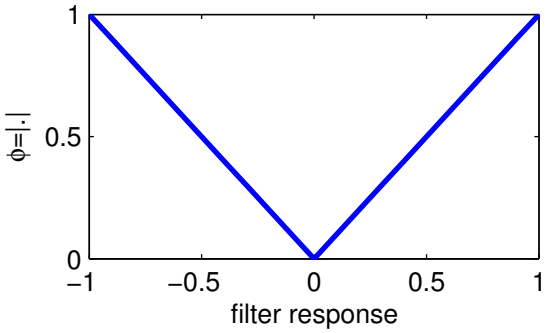


Figure 4.9: One choice for the potential function: $\phi = |.|$ in the range of $[-1, 1]$.

order to cover the range of possible tooth ring widths $T \in \{2, 3, 4, \ldots, 20\}$ was chosen. We remark that our approach is different from that in Zhu et al. [1997], because we are interested in reconstructing tooth rings that resemble only one feature of interest. We do not want to synthesize perceptional equivalent images, including noise, and therefore restrict the FRAME model to incorporate just a single filter. Besides this simplification of the FRAME model, the potential function ϕ that evaluates the filter response is assumed to be known and chosen to be the simplest among the upright curves, namely the absolute value $\phi = |.|$ as depicted in Figure 4.9.

Zhu et al. [1997], on the contrary, suggested selecting iteratively from a bank one important filter after the other, with each one maximizing the entropy decrease over the remaining filters. But this strategy seems highly unstable, especially if the number of filters in the bank is large, the image is noisy, and the form of the potential functions is unknown. Zhu et al. [1997] also suggested a more general energy function allowing for a varying evaluation of filter responses. They empirically divide the potential functions into two classes:

- upright curves that propagate extreme filter responses (filter-like patterns) and
- inverted curves that punish filter typical features and create noise-like patterns.

The potential functions could be uniquely calculated by a stochastic gradient algorithm, as suggested, once again, in Zhu et al. [1997]. But this approach

is computationally expensive. Using the Gibbs sampler, the method counteracts itself because of the need of a large image to find a good observed statistic (and hereby a good potential function) and the time-consumption for Gibbs sampling a large image with large neighborhoods.

Section 4.3.5 will illustrate the properties of the cost function (4.7) while choosing F_T from the family of Gaborcosine filters and evaluating the filter responses by the absolute valued potential function.

4.3.5 Gibbs Simulation

This part of the FRAME section will demonstrate the capacity of the FRAME model (4.7) by a Gibbs simulation, using one Gaborcosine filter and the absolute valued potential function.

In order to create a typical (high probability) image drawn from the prior distribution $P(\lambda)$ in (4.7), the algorithm of choice for statisticians is the Gibbs sampler, as introduced in Chapter 4.2. For the FRAME model the algorithm consists of the following steps:

Gibbs sampling algorithm for the FRAME model

---

1. input
   - initial white noise image $\lambda^{(0)}$
   - filter F
   - potential function ϕ
2. precompute the filter response $F * \lambda^{(0)}$
3. **repeat** sufficiently often
4. **repeat** $N \cdot M$ times ($N \cdot M = |\mathcal{S}|$ size of the image)
5. randomly select site (x, y)
6. **for** all $(x', y') \neq (x, y)$
7. set $\lambda^{(t+1)}_{(x',y')} = \lambda^{(t)}_{(x',y')}$
8. **for** each gray value g of label $\lambda^{(t+1)}_{(x,y)}$
9. **for** all $(x', y') \in C(x, y)$
10. calculate the new filter responses

$$\left(F * \lambda^{(t+1)}\right)(x', y') = \left(F * \lambda^{(t)}\right)(x', y') + F(x - x', y - y')\left(\lambda^{(t+1)}_{(x,y)} - \lambda^{(t)}_{(x,y)}\right)$$

11. **for** each gray value g of label $\lambda_{(x,y)}^{(t+1)}$
12. set $\lambda_{(x,y)}^{(t+1)} = g$ with (conditional) probability

$$P\left(\lambda_{(x,y)}^{(t+1)} = g | \lambda_{N(x,y)}^{(t)}\right) = \frac{e^{\left(\sum_{(x',y')\in C(x,y)} \left|\left(F*\lambda^{(t+1)}\right)(x',y')\right|\right)}}{\sum_{g=0}^{G} e^{\left(\sum_{(x',y')\in C(x,y)} \left|\left(F*\lambda^{(t+1)}\right)(x',y')\right|\right)}}$$

13. update the filter response $F * \lambda^{(t+1)}$

If the computer precision is not sufficient to calculate the conditional probability $P\left(\lambda_{x,y}^{(t+1)} = g | \lambda_{N(x,y)}^{(t)}\right)$ in step 12, one can easily insert a nourishing one.

To detect convergence, the Gelman-Rubin multivariate convergence statistic R (Brooks and Gelman [1998]) is used on every (20x20)th pixel of the image. The Gibbs sampler stops iterating when $R < 1.2$. The algorithm above needs about $O(|F|^2 \cdot NM \cdot G \cdot S)$ operations where $|F|^2$ is the area covered by the filter, NM the size of the image, G the number of gray values and S is the number of sweeps of the Gibbs sampler. A typical number for a small image would be $27^2 \cdot 128^2 \cdot 8 \cdot 1100 \approx 10^{11}$ in our calculations, which is equivalent to 16.8 hours of computation on Hydra[2].

Figure 4.10 presents a typical result using the Gaborcosine filter with parameters $T = 16$ and $\alpha = 0$ (see Figure 4.8), the absolute valued potential function (see Figure 4.9) and 8 gray levels. Figure 4.11 depicts two more examples for the parameter $T = 12$, $\alpha = 0$ and $T = 16$, $\alpha = 30°$, the absolute valued potential function and two gray levels. These images come very close to the ideal TCA image depicted in Figure 4.4. Orientation and width of these lines are determined by both parameters of the Gaborcosine filter.

Figures 4.10 and 4.11 contain small images simulated with the help of large neighborhood structures. The generated texture could therefore be merely a result of label specifications outside the bounds of the image. In the above application, the boundary values seem to have an effect only at the upper and lower edges. For example, specifying zeros outside the bounds

[2]Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

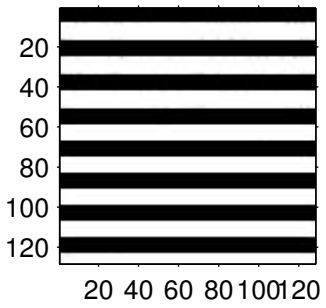


Figure 4.10: A typical image of 128x128 pixels in size simulated by the Gibbs sampler using the FRAME model with filter and potential function from Figures 4.8 and 4.9 with 8 gray levels.

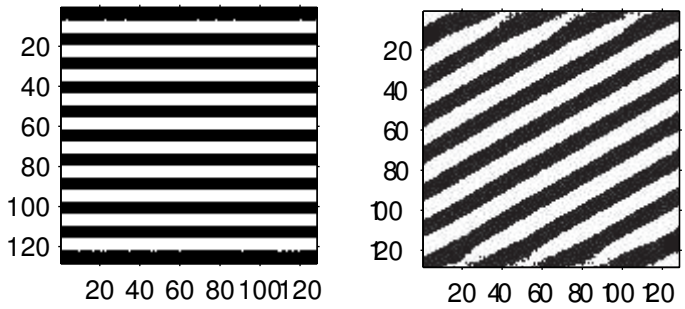


Figure 4.11: Two typical images of 128x128 pixels in size simulated by the Gibbs sampler using the FRAME model with the potential function from Figure 4.9, Gaborcosine filters with parameter $T = 12$, $\alpha = 0$ and $T = 16$, $\alpha = 30°$ and 2 gray levels.

leads to a white first and last ring, as expected. Figures 4.10 and 4.11, on the contrary, were generated with intermediate label values outside the bounds ($\lambda_i = 3.5$). Experiments with halved filters and thinner filters in Chapter 5 and simulations with larger images as well as smaller neighborhoods indicate that the left and right boundaries do not have any effect on the simulated image. In the literature (for example, Cross and Jain [1983]), it is usually assumed that an image is wrapped around a torus. For large neighborhood structures creating periodic structures (as in vertical direction of the present case), this could force artificial patterns at the boundaries if the image size is not a multiple of this period. Similar problems occur with mirroring boundary values. Randomly generated values outside the bounds (fixed over all Gibbs iterations) might also give rise to spurious results. To the author's knowledge, no other alternative exists that would circumvent boundary problems generally.

Instead of using the Gibbs sampling algorithm, Zhu and Mumford [1997] use a gradient ascent algorithm to generate typical patterns from the FRAME model. They state:

> "It seems that the leopard blobs and zebra stripes are among the most canonical patterns which can be generated with easy choices of filters and parameters."

Watching the trace of images generated by the Gibbs sampler before convergence indicates that zebra stripes (containing merging lines) are intermediate states and therefore only local maxima of the Gibbs distribution. Only a stack of lines like in Figure 4.10 can be a global maximum and therefore a typical pattern of the FRAME.

4.3.6 Application

The aim of modeling TCA images, is to uncover the black and white labeling ($\mathcal{G} = \{0, 1\}$), in order to estimate the number of tooth rings. For this purpose, the cementum band in the standard TCA image example in Figure 1.3 is marked manually by an experienced observer and in the current chapter, a Gaussian hidden Markov random field is fitted to this cementum band. The MRF is specified by the FRAME model (4.7). The parameters μ_0, μ_1 and a common variance σ^2 as well as the filter parameter T are estimated by the EM algorithm stated in Section 4.3.2. Label image λ is obtained from the mean field at the last iteration.

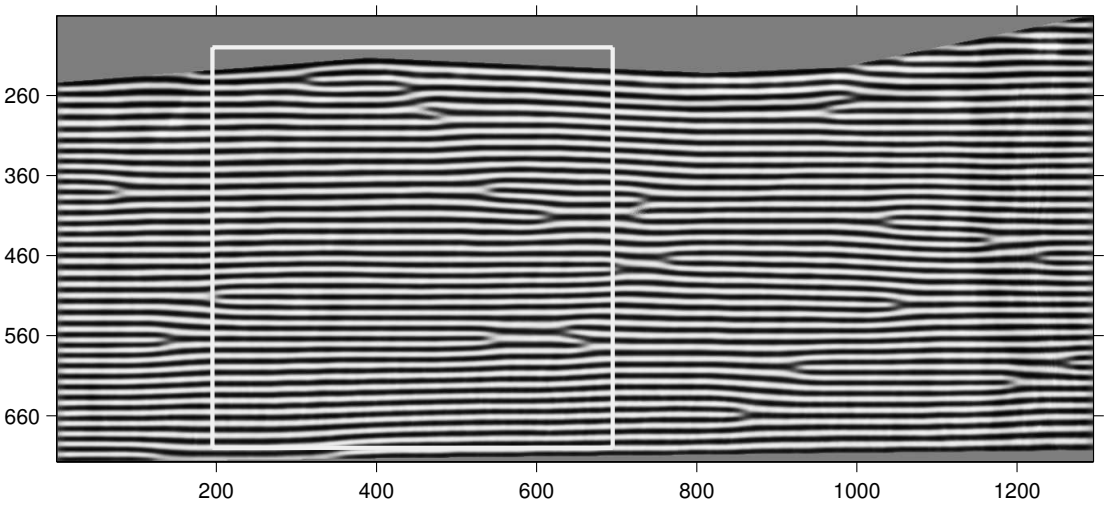


Figure 4.12: The mean field approximation at the last iteration for the cementum band of TCA image 1.3.

Figure 4.12 shows the predictive probability $P(\lambda_i|\tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T})$ of the pixels in the cementum band of Figure 1.3, where $\hat{\boldsymbol{\theta}}$ and $\hat{T}$ are the estimates of the last iteration. The parameter estimates are the means $\hat{\mu}_0 = 28565(\pm 19.8)$, $\hat{\mu}_1 = 29519(\pm 20.2)$, the common variance $\hat{\sigma}^2 = 4.8 \cdot 10^7(\pm 0.009 \cdot 10^7)$ and the ring width $\hat{T} = 14(\pm 0.57)$, where the standard errors are given in parentheses. This compound estimation procedure took a total of eight hours on Hydra[3] for this image.

For the purpose of illustration, a smaller part (marked in Figure 1.3) of the mean field approximation is thresholded ($\lambda_i = 0$ if $P(\lambda_i|\tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}) < 0.5$ and $\lambda_i = 1$ otherwise). The middle lines of the black rings are then superimposed on the original image (see Figure 4.13).

From the known age, we expect 33.61 tooth rings in the image presented in Figure 1.3. Dividing the average width of the cementum band by the estimated average ring width $\hat{T}$ equals 35 rings with a 95% confidence interval of $[32.26, 37.88]$.

The results of this application, together with additional experiments on a whole set of TCA images (see Chapter 6), will be discussed in the next subsection.

[3]Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

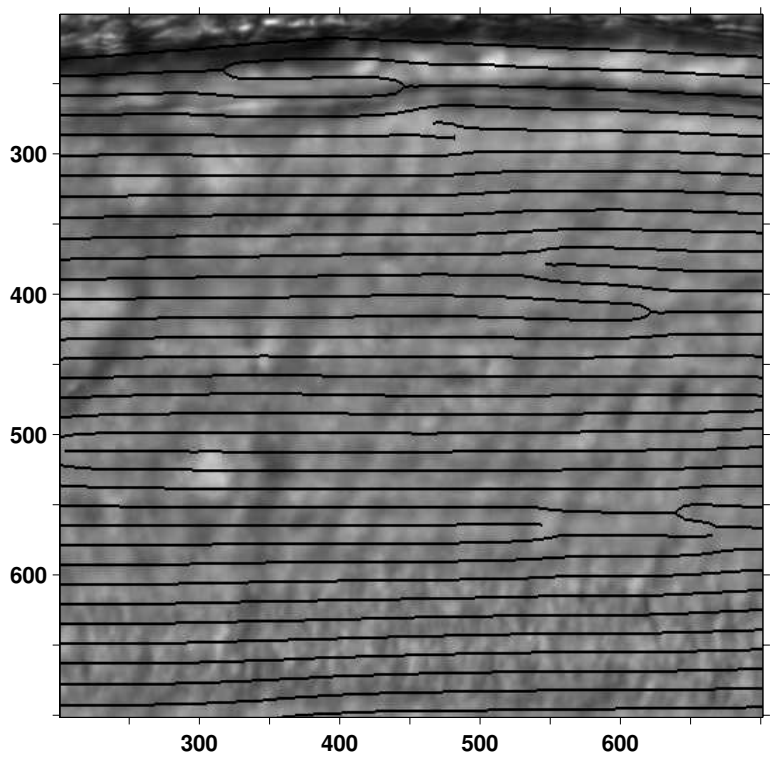


Figure 4.13: The black rings from the marked part of the mean field approximation in Figure 4.12 overlaid onto the original TCA image from Figure 1.3.

4.3.7 Conclusions – Hidden FRAME Modeling

For the evaluation of TCA images, a hidden Markov random field model was set up that proved capable of describing large-scale texture in large images in general and periodic placement of tooth rings in TCA images in particular. Namely, this is the Gaussian hidden FRAME model incorporating a wavelike Gaborcosine filter. It was fitted by exploiting the EM algorithm, which required the approximation of the posterior probabilities $P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ and the final segmentation λ. The Gibbs sampler proved to be infeasible in both cases except for in small images. For example, the simulation of the predictive distribution in Figure 4.10 took more than 16 hours on Hydra and programmed in Matlab. Therefore, mean field approximation was chosen to estimate the posterior probabilities of the labels given the data. Mean field approximation was also applied to realize the approximation of confidence intervals in order to assess the uncertainty within an image and of the model fit. Depending on the size of the cementum band, the estimated average ring width T and the number of iterations, this compound estimation procedure takes on average a total of 11 hours for TCA images on Hydra.

The current chapter has shown that the proposed model and estimation procedure gives a good estimate and a reasonable confidence interval for the number of tooth rings in one exemplary TCA image. Chapter 6.1 will fit the Gaussian hidden FRAME model to a set of 407 TCA images and will show that competetive, and partially even better, estimates for the number of tooth rings are achieved, compared to the observers' estimates and on the basis of the theoretical number of lines. The estimated confidence intervals have to be interpreted with caution, though, as Chapter 6.1 will detail.

Despite the good estimates for the number of tooth rings in TCA images, it is obvious to any reader that most rings are not well met in Figure 4.13, which is due to the global assumptions that were made for the Gaussian hidden FRAME model. These will be listed below, together with attempts to resolve each assumption, which will be made in the following chapters:

- The proposed FRAME model represents a very rigid prior distribution; especially the shape of the single filter heavily influences the reconstruction of TCA images. The model in this form can only take into account strong local changes of tooth rings. Two possible solutions will be examined:

  - Chapter 5 will propose a different prior distribution, the coupled Markov chains, which can be set up very similar to the FRAME model. Only the filter will be halved, such that the size of the neighborhood reduces, and hereby the flexibility of the prior model increases.
  - Chapter 6.3.1 will discuss a localization of the FRAME model by estimating location-dependent filters. This nonhomogeneous FRAME model estimates the filter parameter T depending on each pixel.

- The proposed Gaussian conditional distribution relies on global assumptions as well. Two of these will be relaxed:
  - Chapter 6.3.2 will examine location-dependent Gaussian distributions by specifically exploring the estimation of pixel-dependent variance parameters.
  - Chapter 6.3.3 will evaluate the use of heteroscedastic Gaussian distributions, as theoretically introduced in Sections 4.3.1 and 4.3.2 (but not pursued in the applications). The label-dependent variance parameters will be estimated for the exemplary TCA image.

- A more direct approach to localizing the Gaussian hidden FRAME model for the application to TCA images will be followed in Chapter 6.4.3. Selected TCA images will be partitioned in order to better realize the assumption of constant tooth ring widths and orientations throughout the image.

Besides the above approaches, Chapters 6.4.1 and 6.4.2 also suggest two algorithmic settings which could lead to improved results in the context of TCA image analysis: In Chapter 6.4.1 motivations for a refined grid search over the central filter parameter T are collected. Chapter 6.4.2 examines the mode field method (instead of the mean field method) to realize the posterior distribution approximation, which is necessary for the EM algorithm.

Most of the approaches discussed in Chapter 6 lead to theoretical or algorithmic problems, or to inferior results compared to the Gaussian hidden FRAME model and the EM algorithm proposed during this chapter. Few of the approaches lead to similar results. The Coupled hidden Markov Model that will be introduced in Chapter 5 leads to slightly superior results on the tested subset of TCA images.

We can thus conclude that the Gaussian hidden FRAME model fitted by an EM algorithm is one of the most promising approaches for evaluating TCA images that is examined in this work. The hidden FRAME model also contains great potential for adapting to different problems and images. It could be particularly useful for large images containing large neighborhood structures.

4.4 Discussion – HMRF Modeling

In the current chapter we introduced the hidden Markov random field model for the purpose of evaluating TCA images. This kind of mixture model incorporates spatial dependencies between nearby pixels, as opposed to assuming their independence as in Chapter 3. The cost function of the HMRF, that is, the conditional probability of the observed image given the label image, was specified by conditionally independent Gaussian distributions. For the unobserved label image, two kinds of MRF models were examined: the traditional auto-logistic model and the flexible FRAME model. Simulations for the auto-logistic model have shown that this simple MRF model does not describe macro-textures well. It can therefore not express the prior convictions we have about TCA images. At the same time it provided us with insights as to how to extend the setting to the FRAME model. This second MRF model allows for an increased size of neighborhoods by incorporating filter features in its Gibbs potential, while avoiding computational difficulties, as in the generalized setting of auto-logistic models with a vastly increasing number of parameters. Simulations of the FRAME model showed that it can describe long-range autocorrelation and periodic placement of tooth rings well by selecting the filter family from the wavelike Gaborcosine functions. The Gaussian hidden FRAME model is thereby able to imitate human vision.

To fit a hidden Markov random field model with large neighborhoods to large images, the EM algorithm can be exploited. This procedure requires the approximation of the posterior probabilities $P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ and the final segmentation λ. The Gibbs sampler proved to be infeasible in both cases except for in small images. We therefore chose mean field approximation to estimate the posterior probabilities of the labels given the data. The estimate of the average ring width $\hat{T}$ led to the total number of lines in TCA images and confidence intervals of this measure were derived. This compound estimation procedure is computationally feasible and yields reasonable results

for the TCA application. An extended experiment on 407 TCA images will show that comparable, and partially even better, estimates for the number of tooth rings are achieved by the proposed model, compared to the observer's estimates and on the basis of the theoretical number of lines.

Despite the good estimates for the number of tooth rings in TCA images, the reconstruction of tooth rings leaves something to be desired. Many different approaches for redesigning the model and the algorithm will be suggested in Chapter 6, but the ones pursued will not lead to convincing results. Experiments in this work did not pursue the suggestions of how to avoid saddle points (as maximum likelihood estimates) and to refine the grid search of the central filter parameter in order to improve the quality of confidence intervals.

Another improvement may possibly be achieved with a second parameter in the prior distribution that estimates the orientation of tooth rings. This is only of academic interest now, but it may become practicable with continued improvements in the speed of computers.

Instead of specifying an MRF model for the unobserved label image, Chapter 5 will introduce the coupled Markov chain, which can be set up very similar to the FRAME model. Chapter 6.2 shows that this so-called Gaussian hidden FRAME chain may yield slightly better results; but more extended experiments have yet to confirm this.

Many textures do not follow Gaussian distributions. Also the 'noise' in TCA images (for example, saw cuts and artifacts) is known to be non-normal. Since the application does not suggest an alternative, in the current work the simple and convenient assumption of independent and identical normally distributed gray values, given the label image, was made. This could be relaxed to any distribution of the exponential family within the framework of the EM algorithm and should be tested in future research.

Chapter 5

Coupled Hidden Markov Models

A natural framework for the labeling problem of one-dimensional data (for example, signals depending on time), is provided by Hidden Markov Models (HMM). While using the same framework as for mixture modeling, in HMMs the distribution of labels incorporates directed contextual constraints, as opposed to undirected constraints of hidden Markov random fields or no constraints in classical mixture modeling. These directed constraints are expressed by the a priori assumption that the label vector is a Markov chain. Interchangeably, one speaks about a Markov model for the labels.

Hidden Markov models were introduced in the 1960s at the latest, but were popularized only in 1972 by Baum's powerful learning rule (Rabiner [1989]). Ever since, they have enjoyed a wide range of applications and are sometimes quoted as the most favored model (presently) in speech and vision (Brand et al. [1997]).

But hidden Markov models are ill-suited for data in space that contain interactions in two directions (such as images do). One can extend the HMM approach from 1-D to 2-D by imagining that the data arose from many correlated HMMs. By coupling neighboring Markov chains, the so-called Coupled Hidden Markov Model (CHMM) is formed. Depending on the specific form of the Markov chains, on directional versus nondirectional coupling and also depending on the literature, this type of model is also termed: hidden causal 2-D Markov chain, hidden Markov mesh, hidden Nonsymmetric Half-Plane (NSHP) Markov chain, correlated hidden Markov model, linked hidden Markov model, unilateral hidden MRF, etc. (See, for example, Jeng and Woods [1987], Brand et al. [1997], Brémaud [1999].)

Analogously to HMRF models, CHMMs consist of two levels: a collection of Markov chains to model the label distribution, which is not directly observable, and an observable random field $\mathcal{Y}$. The components of the observable random field Y_i are assumed to be conditionally independent such that

$$P(Y) = \sum_{\lambda \in \mathcal{G}^{N \times M}} P(Y) \prod_{i \in S} f(Y_i|\lambda_i).$$

Following the setting of HMRF models, a Gaussian distribution with parameters $\boldsymbol{\theta} = \{\mu_g, \sigma_g^2|; g \in \mathcal{G}\}$ is chosen again for TCA application, yielding an emission probability

$$f(Y_i|\lambda_i, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_g^2}}.$$

In this chapter, the former MRF prior $P(\lambda)$ will be specified by a coupled Markov model to describe a past relevance structure and to pursue the idea of horizontally running tooth rings that depend on neighboring rings. The reader will see how one can use the experience of hidden FRAME modeling to achieve this, even for large-scale texture with long-range autocorrelations.

The notation in the current chapter will differ from that in the rest of this document. The site i will be named (x, y); where x specifies the row and y the column. The reason is the intrinsic two-dimensional structure of CMMs and the different treatment of the horizontal versus vertical directions.

Section 5.1 explains the general theory for coupled Markov models and then focuses on one group of CMMs: the (non)symmetric half-plane ([N]SHP) Markov chains. Section 5.2 shows how to slightly change the FRAME model to obtain a SHP Markov chain with a FRAME distribution, which is what we term a FRAME chain. Results from fitting the hidden FRAME chain to TCA images will be shown in Section 5.3, followed by a discussion of these results and a comparison with HMRF modeling in Section 5.4.

5.1 General Coupled Markov Models

To set up the theory of coupled Markov models we first define Markov chains and then generalize them to pth-order Markov chains and coupled Markov chains. Because the group of CMMs is complex (and not very structured in the literature), we will quickly turn to defining one specific CMM and deriving its important properties.

Markov chains have a long history in mathematics and physics beginning in 1906 with Andrey Markov. Its definition, as the one below, can be found in standard textbooks (for example, in Brémaud [1999]):

Definition 5.1.1. *A random sequence* $\{\Lambda_i\}_{i\geq 0}$ *with countable state space* G *is called a* Markov chain*, if for all integers* i *and all instances* $\{\lambda_i\}_{i\geq 0}$

$$P(\lambda_i|\lambda_0,\ldots,\lambda_{i-1}) = P(\lambda_i|\lambda_{i-1}).$$

This is a simple and mathematically tractable relaxation of the independence assumption of nearby labels. Markov chains assume that the present depends only on the most recent past and herewith assumes a directional, one-dimensional dependence structure in the data, as opposed to an undirected, two-dimensional dependence structure with MRF.

For dependencies of longer range one can define the pth-order Markov chain (see, for example, MacDonald and Zucchini [1997]):

Definition 5.1.2. *A random sequence* $\{\Lambda_i\}_{i\geq 0}$ *with countable state space* G *is called* Markov-p *(pth-order Markov), if for all integers* i *and all instances* $\{\lambda_i\}_{i\geq 0}$

$$P(\lambda_i|\lambda_0,\ldots,\lambda_{i-1}) = P(\lambda_i|\lambda_{i-1},\ldots,\lambda_{i-p}).$$

Because of the diversity of different coupled (hidden) Markov models that exist, there are also a variety of definitions. Therefore, a general definition for coupled Markov models is not very precise and is chosen here roughly following Preuß [1975] and Nefian [2002]. A more specific definition of one kind of coupled Markov model will be given subsequently.

Definition 5.1.3. *Let* $\mathcal{S}$ *be ordered in time such that for each site* i*, the subset* $A_i \subseteq \mathcal{S}$ *defines the 'full past' of* i*:* $A_i = \{i-1, i-2, \ldots, 1\}$*. Let the frontier of* A_i *be the set of all pixels in* A_i *adjacent to* i*. Then a random field* Λ *are called* coupled Markov chains*, if for all instances* $\lambda \in \mathcal{G}^{|\mathcal{S}|}$ *and all sites* $i \in \mathcal{S}$

$$P(\lambda_i|\lambda_{A_i}) = P(\lambda_i|\lambda_{B_i})$$

holds, for some subset $B_i \subseteq A_i$ *containing the frontier of* A_i*.*

Figure 5.1 displays exemplary subsets A_i and B_i of $\mathcal{S}$ for one site i.

For applications of CMMs, there are two problems: First of all, no natural order (or direction) exists in two dimensions. We therefore need to define the 'past' of a coupled Markov model according to the application at hand.

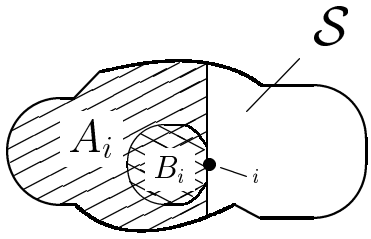


Figure 5.1: Exemplary dependence structure of coupled Markov chains. Set A_i defines the past and B_i the most recent past of site i at the frontier of A_i.

For TCA images it is assumed that tooth rings are oriented horizontally and are directional (from left to right for a first assumption). Each tooth ring will be made up of several Markov chains and each chain is correlated nondirectionally to its neighboring chains. The kind of CMM we have in mind is a direct relative of the NSHP Markov chains which were introduced by Preuß [1975]. We will call our CMM, the SHP Markov chain and follow Jeng and Woods [1987] to define it:

Definition 5.1.4. *A random field* Λ *is called* SHP Markov chain*, if for all instances* $\lambda \in \mathcal{G}^{N\times M}$ *and all sites* $i = (x, y) \in \mathcal{S}$

$$P(\lambda_{x,y}|\lambda_{x',y'}, (x', y') \in \mathcal{S}, y' < y) = P(\lambda_{x,y}|\lambda_{x',y'}, (x', y') \in C_{x,y})$$

holds, for the set $C_{x,y} = \{(x', y') \in \mathcal{S} \mid 0 < y - y' < q, |x - x'| < p)\}$ *and some integers* p *and* q*.*

Figure 5.2 shows a simplified conditional independence graph for a generic SHP Markov model. The regions of interest and dependence in the past for this (N)SHP Markov models are depicted in Figure 5.3. This model arose from the traditional NSHP Markov chain by ignoring the recent past of the column of interest y (i.e. $x' < x$) and thereby obtaining a nondirectional (noncausal) coupling of neighboring chains. We prefer this model to the NSHP Markov chains because TCA images do not have a natural vertical direction, which is biologically motivated, for example.

The second problem is the intractability of an exact solution to CHMMs in general. Since CHMMs do not fulfill the Markov property, there is no simple decomposition of the prior probability that might lead to simple parameter estimation procedures (Brand et al. [1997]). Exact solutions to CHMMs

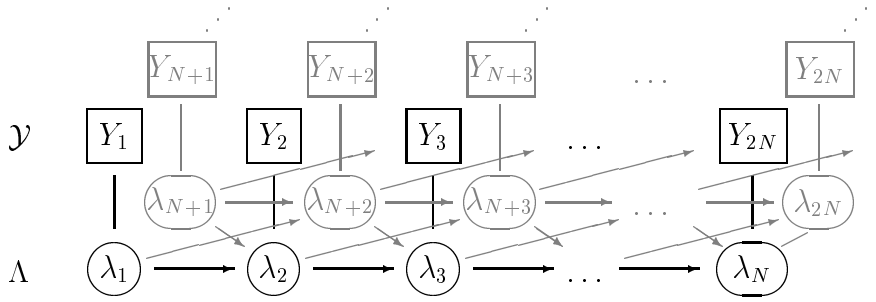


Figure 5.2: Conditional independence graph of a symmetric half-plane Markov model (simplified). Boxes represent observed or fixed quantities, circles show unknowns, arrows display conditional dependence.

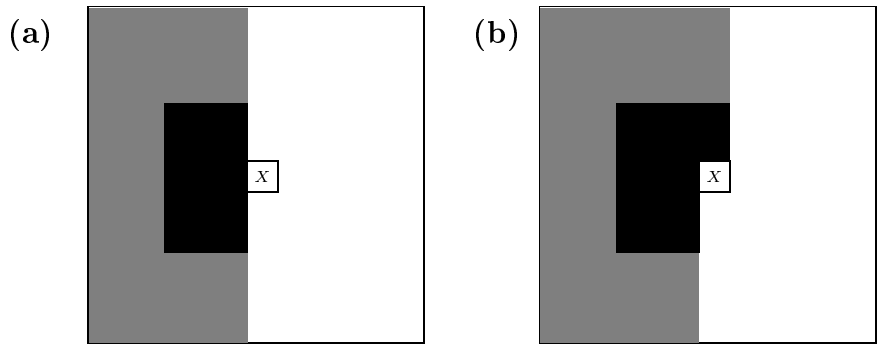


Figure 5.3: The regions of interest (gray and black) and dependence (black) in the past for a generic symmetric half-plane Markov model in (a) and for a nonsymmetric half-plane Markov model in (b).

would therefore require the same time complexity as the Cartesian product of its HMMs, leading to the same problems as in Chapter 4.3 with hidden FRAME modeling of large neighborhoods and large images. But one can also try to take advantage of the same algorithms as for hidden FRAME modeling. In order to do so, one needs to transform the (N)SHP Markov chains into Markov and Gibbs random fields (see Brémaud [1999]):

Theorem 5.1.5. *If a random field Λ consists of (N)SHP Markov chains, locally depending on $C_{x,y}$; then Λ is a Gibbs random field with potentials*

$$V_C(\lambda) = \ln P(\lambda_{x,y} | \lambda_{x',y'}, (x',y') \in C_{x,y})$$

on the collection of cliques

$$\mathcal{C} = \{C = \{C_{x,y} \cup (x,y)\} \mid (x,y) \in \mathcal{S}\},$$

and Λ a Markov random field on neighborhoods

$$N(x,y) = \left(\bigcup_{(x',y'),(x,y) \in C_{x',y'}} \{C_{x',y'} \cup (x',y')\} \backslash (x,y) \right) \cup C_{x,y}.$$

Proof. We will first prove the Gibbs property and then turn to the Markov property of SHP Markov chains. For NSHP Markov chains the same proof holds by redefining $C_{x,y}$.

From Definition 5.1.4, the joint probability $P(\lambda)$ for a SHP Markov chain can be derived by applying Bayes' rule and scanning the labels from left to right:

$$\begin{aligned} P(\lambda) &= P(\lambda_{.1}) \prod_{(x,y) \in \mathcal{S}, y>1} P(\lambda_{x,y} | (x',y') \in \mathcal{S}, y' < y) \\ &= \prod_{(x,y) \in \mathcal{S}} P(\lambda_{x,y} | \lambda_{x',y'}, (x',y') \in C_{x,y}) \\ &= e^{\sum_{C \in \mathcal{C}} V_C(\lambda)} \end{aligned}$$

where $P(\lambda_{.1})$ denotes the probability of the first column and

$$V_C(\lambda) = \ln P(\lambda_{x,y} | \lambda_{x',y'}, (x',y') \in C_{x,y})$$

clearly defines a Gibbs potential on the collection of cliques

$$\mathcal{C} = \{C = \{C_{x,y} \cup (x,y)\} \mid (x,y) \in \mathcal{S}\}.$$

Therefore Λ is a Gibbs random field with the potentials $\{V_C\}_{C \in \mathcal{C}}$ defined above.

From the Hammersley-Clifford Theorem 4.1.5, one can now easily derive the local characteristics

$$\begin{aligned} P(\lambda_{x,y} | \lambda_{\mathcal{S}\setminus(x,y)}) &= \frac{e^{\sum_{C \in \mathcal{C}, (x,y) \in C} V_C(\lambda)}}{\sum_{\lambda'_{x,y} \in \mathcal{G}} e^{\sum_{C \in \mathcal{C}, (x,y) \in C} V_C(\lambda')}} \\ &= P(\lambda_{x,y} | \lambda_{N(x,y)}) \end{aligned}$$

of the random field Λ, where

$$N(x,y) = \left(\bigcup_{(x',y'),(x,y) \in C_{x',y'}} \{C_{x',y'} \cup (x',y')\} \setminus (x,y) \right) \cup C_{x,y}$$

is the neighborhood of site (x,y). Therefore Λ is a Markov random field on the neighborhood system $\mathcal{N} = \{N(x,y) \mid (x,y) \in \mathcal{S}\}$.

A generic clique and neighborhood of the (N)SHP Markov model from Definition 5.1.4 (with $p = q = 3$) is depicted in Figure 5.4. □

Theorem 5.1.5 provides us with the possibility of defining (N)SHP Markov chains locally with the help of their past relevance. From this, the joint distribution as well as the local Markov random field distribution can be derived. One can therefore exploit the knowledge of the previous chapter on HMRFs for coupled hidden Markov chain modeling. That is, the Gibbs sampler can be used for drawing a typical image, the same EM algorithm for parameter estimation, and a slightly altered FRAME model can be incorporated. The results of HMRF and CHMM modeling will be easily comparable.

Within the EM framework, an alternative approximate procedure to MFA for fitting CHMMs is the structured mean field approximation (Ghahraman and Jordan [1997]). Instead of factoring the prior probability distribution completely (as in MFA), it factors only the coupled HMMs, thereby preserving some structure of the CMM. Although structured mean field approximation should be tested for TCA image application in the future, we will focus here on pure mean field approximation in order to obtain comparable results for HMRF and CHMM modeling.

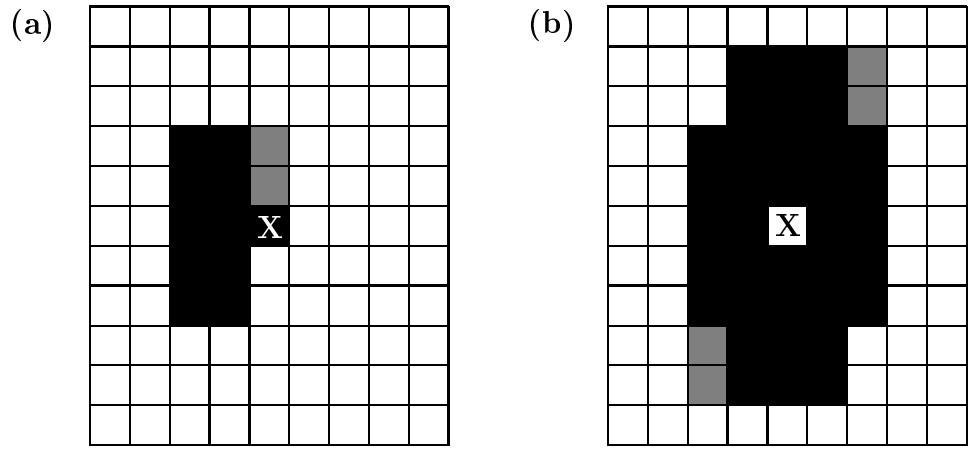


Figure 5.4: A generic clique C (a) and its neighborhood $N(x, y)$ (b) of a symmetric half-plane Markov model in black and of a nonsymmetric half-plane Markov model in black and gray.

5.2 FRAME Chain

After defining the general SHP Markov model and deriving its properties, its specific probability distribution can now be defined, depending on the application at hand. The aim of modeling remains: efficiently describing large-scale texture and taking into account long-range dependencies between labels and the periodic placement of tooth rings in TCA images. Borrowing from the FRAME model, we will incorporate filter features into the CMM in order to do so. The resulting model will be called a FRAME chain; its ingredients will be specified and one of its typical images presented.

To refresh the FRAME methodology, the reader is referred to Chapter 4.3. Specifically, one can easily turn the FRAME model (4.7) into a CMM by cutting the filter F_T into half:

Definition 5.2.1. *For any site (x, y), let F_T be a filter of size $(2p+1)\times(q+1)$ supported on the area of the clique $C = \{(x, y) \cup (x', y') \in \mathcal{S} \mid 0 < y - y' < q, |x - x'| < p)\}$ of site (x, y) and 'centered' in the middle right site*

(x, y). *Then a SHP Markov model is called* FRAME chain *if for all instances* $\lambda \in \mathcal{G}^{N\times M}$ *and all sites* $(x, y) \in \mathcal{S}$

$$P(\lambda_{(x,y)}|\lambda_{x',y'}, (x', y') \in C_{x,y}) = \frac{e^{\phi[(F_T * \lambda)(x,y)]}}{\sum\limits_{\lambda'_{(x,y)}\in\mathcal{G}} e^{\phi[(F_T * \lambda')(x,y)]}}.$$

Example (nonsymmetric filter)*:* For illustration purposes, let

$$F = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } \lambda = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Then } F * \lambda = \begin{pmatrix} 0 & -1 & -1 \\ 1 & 2 & 2 \\ 0 & -1 & -1 \end{pmatrix}$$

and $F * \lambda$ emphasized lines (ignoring the left border). □

For the FRAME chain model in Definition 5.2.1, it is assumed a priori that the label image is Gibbs distributed according to

$$P(\lambda) = \frac{1}{Z} e^{\sum\limits_{i\in\mathcal{S}} \phi[(F_T * \lambda)(x,y)]} \tag{5.1}$$

(see Theorem 5.1.5). This is very similar to Definition 4.7 of the FRAME model, with the only difference being the support of the filter. The local characteristics of the Markov random field are given by

$$P(\lambda_{(x,y)}|\lambda_{N(x,y)}) = \frac{e^{\sum\limits_{C\in\mathcal{C},(x,y)\in C} e^{\phi[(F*\lambda)(C)]}}}{\sum\limits_{\lambda'_{(x,y)}\in\mathcal{G}} e^{\sum\limits_{C\in\mathcal{C},(x,y)\in C} e^{\phi[(F*\lambda')(C)]}}}.$$

The FRAME chain hereby elegantly combines filtering theory and Markov modeling; and the EM algorithm using mean field approximation, as stated on page 79, can be applied to estimate its parameters and compute the final segmentation of labels. The FRAME chain is completely specified by the filter family and the potential function, which will be given below.

Following the reasoning of the Gaussian hidden FRAME model in Chapter 4.3.4, the filter F_T will be specified from the family of Gaborcosine functions

$$Gcos_{T,\alpha}(x, y) = c_1 \cdot e^{\frac{-\left(rx'^2+y'^2\right)}{2T^2}} \cos\left(\frac{2\pi}{T}x'\right);$$

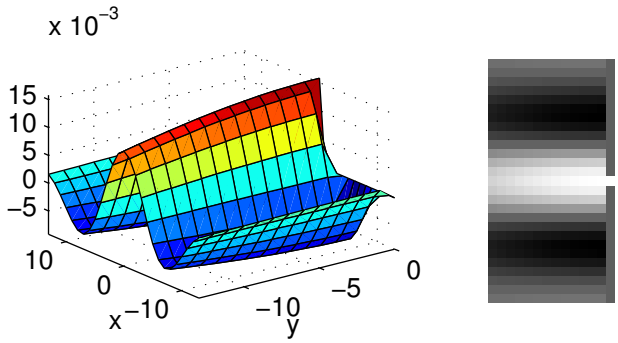


Figure 5.5: 3-D surface and image of a Gaborcosine function with $T = 16$ and $\alpha = 0$, supported on $x \in [-13, 13], y \in [-13, 0]$.

with $x' = x\cos\alpha + y\sin\alpha$, $y' = -x\sin\alpha + y\cos\alpha$, $r = 4$ being the aspect ratio and c a normalizing factor such that

$$\sum_{(x,y):Gcos_{T,\theta}(x,y)>0} Gcos_{T,\theta}(x,y) = 1 \text{ and } \sum_{(x,y)} (Gcos_{T,\theta}(x,y) - c_2) = 0.$$

The main difference is the choice of the range

$$x \in \left\{ -\left\lceil \frac{4}{5}T \right\rceil, \ldots, \left\lceil \frac{4}{5}T \right\rceil \right\} \text{ and } y \in \left\{ -\left\lceil \frac{4}{5}T \right\rceil, \ldots, -1 \right\} \text{ or } (x,y) = (0,0).$$

The Gaborcosine function above is an elongated half-normal multiplied by a cosine wave, where parameter T changes the wavelength and α determines the orientation of the cosine wave. For example, Figure 5.5 shows the Gaborcosine function for $T = 16$, $\alpha = 0$ and in the range of $x \in [-13, 13]$ and $y \in [-13, 0]$. This filter can capture waves or lines of width 16 and orientation 0°.

In the application for TCA images, the main direction of tooth rings $\alpha = 0$ will be fixed. In order to cover the range of possible tooth ring widths, $T \in \{2, 3, 4, \ldots, 20\}$ is chosen. The potential function ϕ that evaluates the filter responses pixelwise is assumed to be known and chosen to be the simplest among the upright curves, namely the absolute value $\phi = |.|$. These are the same choices as for hidden FRAME modeling in Chapter 4.3.4.

To demonstrate the capacity of the FRAME chain model and to compare it to the FRAME model (4.7), one can run a Gibbs simulation using one

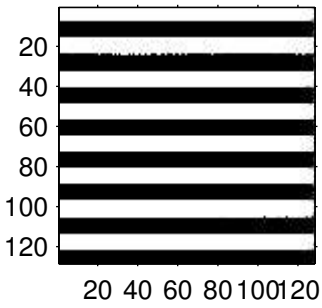


Figure 5.6: A typical image simulated by the Gibbs sampler using the FRAME chain model with the filter from Figure 5.5, potential function $\phi = |.|$ and 8 gray levels.

Gaborcosine filter and the absolute valued potential function. A typical (high probability) image is drawn from the prior distribution $P(\lambda)$ in (5.1), using the Gibbs sampler as introduced in Section 4.3.5. Figure 5.6 displays such an image for the filter parameters $T = 16$ and $\alpha = 0$. Orientation and width of these lines are determined by both parameters of the Gaborcosine filter. This image comes very close to the ideal TCA image depicted in Figure 4.4 and the typical image drawn from the FRAME model (Figure 4.10). One can see that the right hand border is fuzzier because the last column is not relevant to the past of any pixel and is therefore drawn at random. The hope that the FRAME chain is more flexible than the FRAME model (because it is more local), is not supported by this simulation example.

5.3 Application

The aim of modeling TCA images is to uncover the black and white labeling ($\mathcal{G} = \{0, 1\}$), in order to estimate the number of tooth rings. For this purpose, the cementum band in the standard TCA image example in Figure 1.3 was marked manually by an experienced observer and in the current chapter a Gaussian coupled hidden Markov model is then fitted to this cementum band. The CMM is specified by the FRAME chain as in (5.1). The parameters μ_0, μ_1 and a common variance σ^2 as well as the filter parameter T are estimated by the EM algorithm stated in Chapter 4.3.2. Label image λ is obtained from the mean field at the last iteration, just as in hidden FRAME modeling.

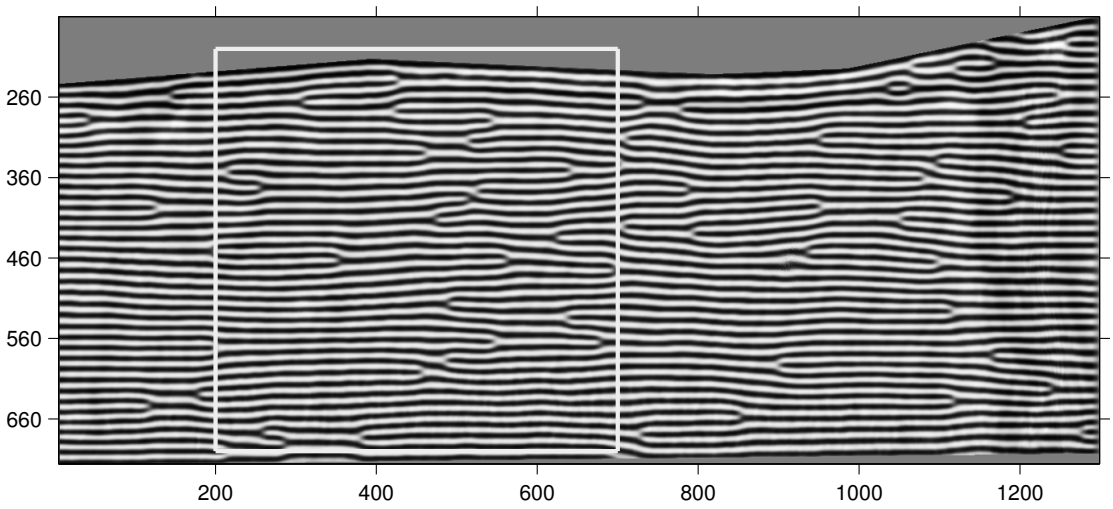


Figure 5.7: The mean field approximation at the last iteration for the cementum band of TCA image 1.3.

Figure 5.7 shows the predictive probability $P(\lambda_i|\tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T})$ of the pixels in the cementum band of Figure 1.3, where $\hat{\boldsymbol{\theta}}$ and $\hat{T}$ are the estimates of the last iteration. The parameter estimates are the means $\hat{\mu}_0 = 28395$, $\hat{\mu}_1 = 29692$, the common variance $\hat{\sigma}^2 = 4.8 \cdot 10^7$ and the ring width $\hat{T} = 14$. This compound estimation procedure requires a total of five hours on Hydra[1] for this image, which is faster than with the hidden FRAME model because the size of the filter, and thereby the size of the neighborhood, is approximately halved.

Comparing to the mean field approximation of HMRF modeling in Figure 4.12, one can see that the hidden FRAME chain model is more flexible because more bifurcations occur. To emphasize this flexibility even more, experiments with the hidden FRAME chain model and a thinner filter (a horizontal past relevance structure of one pixel) were additionally carried out. Figure 5.8 displays the mean field for such a model, clearly showing more flexibility in the reconstruction of tooth rings.

For the purpose of illustration, a smaller part (marked in Figure 1.3) of the mean field approximation in Figure 5.7 is thresholded ($\lambda_i = 0$ if

[1] Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

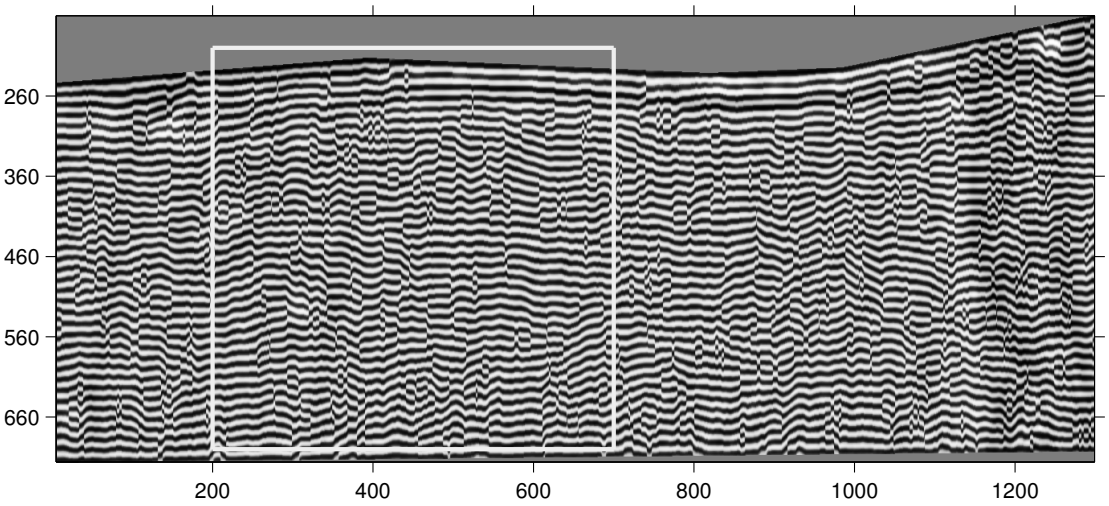


Figure 5.8: The mean field approximation at the last iteration for the cementum band of TCA image 1.3 generated by the hidden FRAME chain implementing a thin filter consisting of one column plus one pixel.

$P(\lambda_i|\tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}) < 0.5$ and $\lambda_i = 1$ otherwise). The middle lines of the black rings are then superimposed on the original image in Figure 5.9.

From the known age, 33.61 tooth rings are expected in the image presented in Figure 1.3. Dividing the average cementum band width by the estimated average ring width $\hat{T}$ gives 35 rings for both filter widths.

The results of this application, together with additional experiments on various TCA images (see Chapter 6), will be discussed in the next section.

5.4 Discussion – CHM Modeling

In the current chapter we introduced the coupled hidden Markov model for evaluating TCA images. This kind of mixture model incorporates dependencies between nearby pixels while distinguishing between different dimensions. Following Chapters 3 and 4, the conditional probability of the observed image, given the label image, was specified by conditionally independent Gaussian distributions. For the unobserved label image, a symmetric half-plane Markov model was assumed, describing directional horizontal and nondirectional vertical interactions between neighboring pixels. Specifically, this chapter developed the novel FRAME chain for the label distribution. The

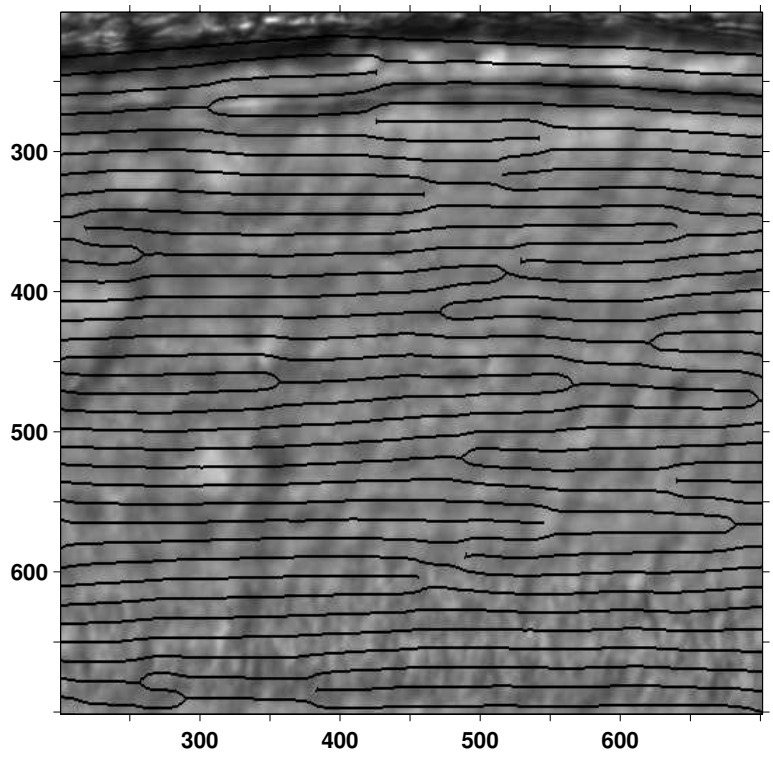


Figure 5.9: The black rings from the marked part of the mean field approximation in Figure 5.7 overlaid onto the original TCA image from Figure 1.3.

FRAME chain is a close relative of the FRAME model in Chapter 4.3 and because of this affinity, the same EM algorithm and mean field approximation were exploited for model fitting and for segmenting the image. Simulations for the FRAME chain gave similar results as for the FRAME model, and showed that this model can also describe the long-range autocorrelation and periodic placement of tooth rings well by selecting the filter family from the halved Gaborcosine functions.

The results of Chapters 4.3.6 and 5.3 in general show analogies. Particularly, overall similar results were produced for the parameter estimates as well as the reconstructions of the exemplary TCA image from fitting the Gaussian hidden FRAME model and the Gaussian hidden FRAME chain model with two different filter widths (see Figures 4.12, 5.7 and 5.8). The experiments on the selected 49 TCA images (of good quality) given in Chapter 6.1 also display similar estimates for the number of tooth rings in the Gaussian hidden FRAME model and the Gaussian hidden FRAME chain implementing the standard halved Gaborcosine filter. The difference becomes visible for the Gaussian hidden FRAME chain implementing the thin filter, which leads to slightly better results on the studied subset of TCA images. This is probably due to the extra flexibility in horizontal direction caused by the reduced size of the neighborhood. The thinner the filter in the model is, the more flexible the label reconstruction is for tooth rings that deviate from the horizontal direction. But the nearly constant ring width over the whole label image, which leads to an imperfect detection of the actual tooth rings (as mentioned in Chapter 4.3.7), also remained for the hidden FRAME chain models. In general, it is even suspected that the more flexible a model is, the poorer it will perform for bad quality TCA images; but this (or the improved estimated number of tooth rings alternatively) has to be verified on an extended set of TCA images.

The estimation procedure for fitting the Gaussian hidden FRAME chain to TCA images takes on average about ten hours for the standard halved filter and about seven hours for the thin filter. This is faster than for the Gaussian hidden FRAME model and thus may be a reason to pursue this approach in future research.

The current chapter assumed a coupled hidden Markov model with directed contextual constraints in the horizontal direction but undirected constraints in the vertical direction. Since there is no application-driven reason for this, future research should also examine the effect of the reverse assump-

tion (i.e. undirected constraints in the horizontal direction and directed ones in the vertical direction).

Chapter 6

HMRFs and CHMMs in Action

The Gaussian hidden FRAME model proposed in Chapter 4.3 seems to be one of the most promising approaches for evaluating TCA images. This chapter will mainly test this model and its estimation procedure.

Section 6.1 will evaluate the semi-automatic procedure based on the Gaussian hidden FRAME model on a whole set of TCA images from the Basel series of spital-field St. Johann (see Chapter 1.2). For this purpose, a proper reference needs to be selected to judge the quality of the method. Also, it will be described how the set of 407 TCA images was selected. Then the estimated number of lines as well as confidence intervals can be assessed.

Section 6.2 will evaluate the Gaussian hidden FRAME chain model on a smaller set of TCA images from the Basel series and compare it to the Gaussian hidden FRAME model.

In Sections 6.3 and 6.4, different specifications of the Gaussian hidden FRAME model and the EM algorithm will be tested and discussed. These specifications will demonstrate the models flexibility and capabilities, and their results will be compared with the original model and algorithm. Specifically, Section 6.3 will be devoted to localizing the FRAME model as well as the observable random field. Section 6.4 will discuss a diverse set of issues connected to the EM algorithm for the Gaussian hidden FRAME model and in the context of TCA image analysis. Convergence issues, prior probability approximation and partitioning TCA images to fit the model separately to each part are all matters that will be examined.

Section 6.5 will apply the Gaussian hidden FRAME model to cross sections of trees to recognize their rings. It will also list a divers set of layered

structures to demonstrate the range of possible applications of the hidden FRAME model.

6.1 Evaluating the HMRF on the Basel Series

The Gaussian hidden FRAME model was fitted to one exemplary TCA image in Chapter 4.3.6 giving a very good estimate for the number of tooth rings. This section will now evaluate the proposed semi-automatic procedure on a whole set of TCA images. In Section 6.1.1, possible references for judging the quality of this method will be examined. Section 6.1.2 describes the procedure of selecting a set of 407 TCA images from the Basel series of spital-field St. Johann. The estimated number of lines and confidence intervals for this sample of TCA images will be assessed in Section 6.1.3, followed by a discussion in Section 6.1.4.

6.1.1 Reference Selection

In order to evaluate the Gaussian hidden FRAME model for TCA age estimation, a proper reference needs to be established. There are two possible measures that can serve as references: the theoretical number of lines (the known age subtracting the tooth eruption age) and the manual counts of the number of tooth rings.

The reference for the quality of an age estimate is ideally the known age at death. For the TCA method, the theoretical number of lines (TNoL) is such a reference. But TCA images might contain only erroneous information about the theoretical number of lines due to the following reasons:

- biological processes (for example, anomalies in ring formation or tooth eruption age),
- the skeleton being buried,
- the preparation procedure (for example, the slicing position or angle),
- the microscopy (for example, the focus level) and
- the imaging procedure (for example, the brightness level).

In short, TCA images are corrupted by shortcomings caused by the TCA method. Since the basis of an image analysis procedure is the image, the theoretical number of lines is not a perfect reference for a semi-automatic counting procedure of TCA images.

The content of an image can be described by human perception. The manual counts of the number of tooth rings are therefore a second possible reference for a semi-automatic counting procedure of TCA images. But each manual age estimate might be erroneous due to the following reasons:

- the observer (for example, their experience),
- the observation (for example, their motivation) and
- prior beliefs (for example, the knowledge that several samples belong to the same individual)

Specifically for TCA age estimation on the Basel series, the manual observers knew which TCA image belonged to each slice and tooth. Tooth ring counts were thus not independent.

Let us therefore evaluate the semi-automatic counting procedure against both (imperfect) reference measures: the theoretical number of lines and the manual counts for the number of tooth rings. But let us in return exclude TCA images where the second reference measure collapses: images for which there exist less than two, of the possibly four, manual counts from the two observers of the MPIDR. (See Chapter 1.2 or Section 6.1.2 below for the sample design.) These are the TCA images of 'very bad' quality from the observers' perspective.

One additional confounding factor exists that has not yet been mentioned. Due to the semi-automatic nature of the procedure, the cementum band of each TCA image is marked by an observer and the orientation of tooth rings – if not horizontal – is specified. This manual task is subject to similar confounding factors as in each manual age estimate (see above). Although errors arising from the manual part of a semi-automatic procedure should not be attributed to the automatic part of the procedure, it is not obvious how to separate these sources of errors in the current study.

6.1.2 Sample Selection

Let us choose TCA images from the Basel series described in the introduction to evaluate the semi-automatic counting procedure. The skeletons from

the Basel series offer a unique possibility for evaluating anthropological age estimation methods by comparing them with each other as well as with the known age.

Also for the TCA method, an extended study was performed in which the tooth laboratory of the Max Planck Institute for Demographic Research participated with two experienced observers. From the 100 selected individuals of the Basel series, the MPIDR acquired 2,120 TCA images. Of these TCA images 1,021 were taken with 40-fold, while the other images were taken with 20-fold magnification. Let us focus on TCA images with 40-fold magnification because potentially more information on tooth rings is stored in them. The 407 TCA images were selected based on the condition that at least two (independent) observations from the MPIDR resulted in estimates for the number of tooth rings. These are the 'better' TCA images from the observers' perspective, whereas for the remaining TCA images on the other hand, one of our reference measures for evaluating the semi-automatic counting procedure does not contain enough information. (See Section 6.1.1 above.)

6.1.3 Results

For each of the 407 selected TCA images from the Basel series, one of the MPIDR observers marked the cementum band and the author specified the main orientation of the tooth rings. Then the Gaussian hidden FRAME model specified in Chapter 4.3 was fitted, and estimates for both the number of tooth rings and confidence intervals were calculated.

Let us compare this semi-automatic counting procedure to the two proximate references that were studied in Section 6.1.1: the theoretical number of lines (TNoL) and the manual observations. For simplicity, at most one of the possibly two observations per observer is considered in this comparison: that is, the first existing observation. Figure 6.1 shows box-and-whisker plots of the difference between the estimates for the number of tooth rings of the first (of possibly two) observations of observer one, observer two and the semi-automatic procedure compared to the theoretical number of lines for the 407 selected TCA images. The summary statistics are listed in Table 6.1. The average as well as the median deviance of the semi-automatic estimates to the TNoL is smaller than the observers' estimates to the TNoL. The variance of this deviance is a little bit larger for the semi-automatic procedure.

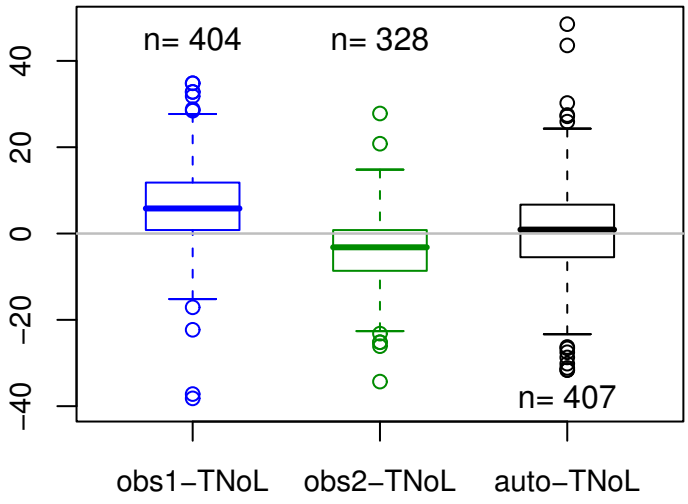


Figure 6.1: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines for the semi-automatic procedure (right) and for the first (of possibly two) observation(s) from the two observers (left and middle) in 407 selected TCA images from the Basel series.

	obs1 – TNoL	obs2 – TNoL	auto – TNoL
Max	34.81	27.81	48.51
3rd Quart.	11.80	0.80	6.69
Mean	6.33	-3.95	0.05
Median	5.81	-3.19	0.90
1st Quart.	0.80	-8.64	-5.49
Min	-38.20	-34.31	-31.65
St. Dev.	9.45	8.00	10.78

Table 6.1: Summary statistics of the difference between the estimated and the theoretical number of lines for the semi-automatic procedure and for the first (of possibly two) observation(s) from the two observers in 407 selected TCA images from the Basel series.

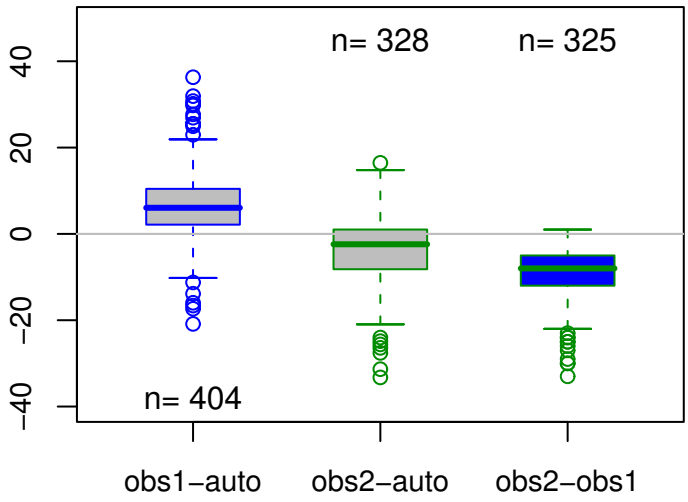


Figure 6.2: Box-and-whiskers plots of the difference between the estimated number of tooth rings for the semi-automatic procedure ('auto') and the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') in 407 selected TCA images from the Basel series.

Comparing the semi-automatic estimates directly to the observers' estimates produces the results shown in Figure 6.2. The box-and-whisker plots show that the deviance between the semi-automatic procedure and the manual observations is smaller than, and the distribution of it not as skewed as, the deviance between the two observers. The variance of these deviances is once again slightly larger for the semi-automatic procedure.

In Section 6.1.1, it was mentioned that the estimates of the observers are not independent due to the observers' knowledge about the membership of TCA images to a particular slice and tooth. The results shown in Figures 6.1 and 6.2 are thus not a fair comparison between the observers' age estimate and the semi-automatic age estimates. Let us therefore compare the results of the 407 TCA images from the Basel series at the level of 72 teeth by displaying the mean (subtracting the TNoL) and the range of estimates within each tooth. Figure 6.3 displays these measures, where each line indicates one tooth: the black lines represent the difference of the semi-automatic estimates to the TNoL, the green lines the difference of the estimates of observer one to the TNoL and the blue lines the difference of the estimates of observer two to the TNoL. The distribution of the intra tooth mean for the

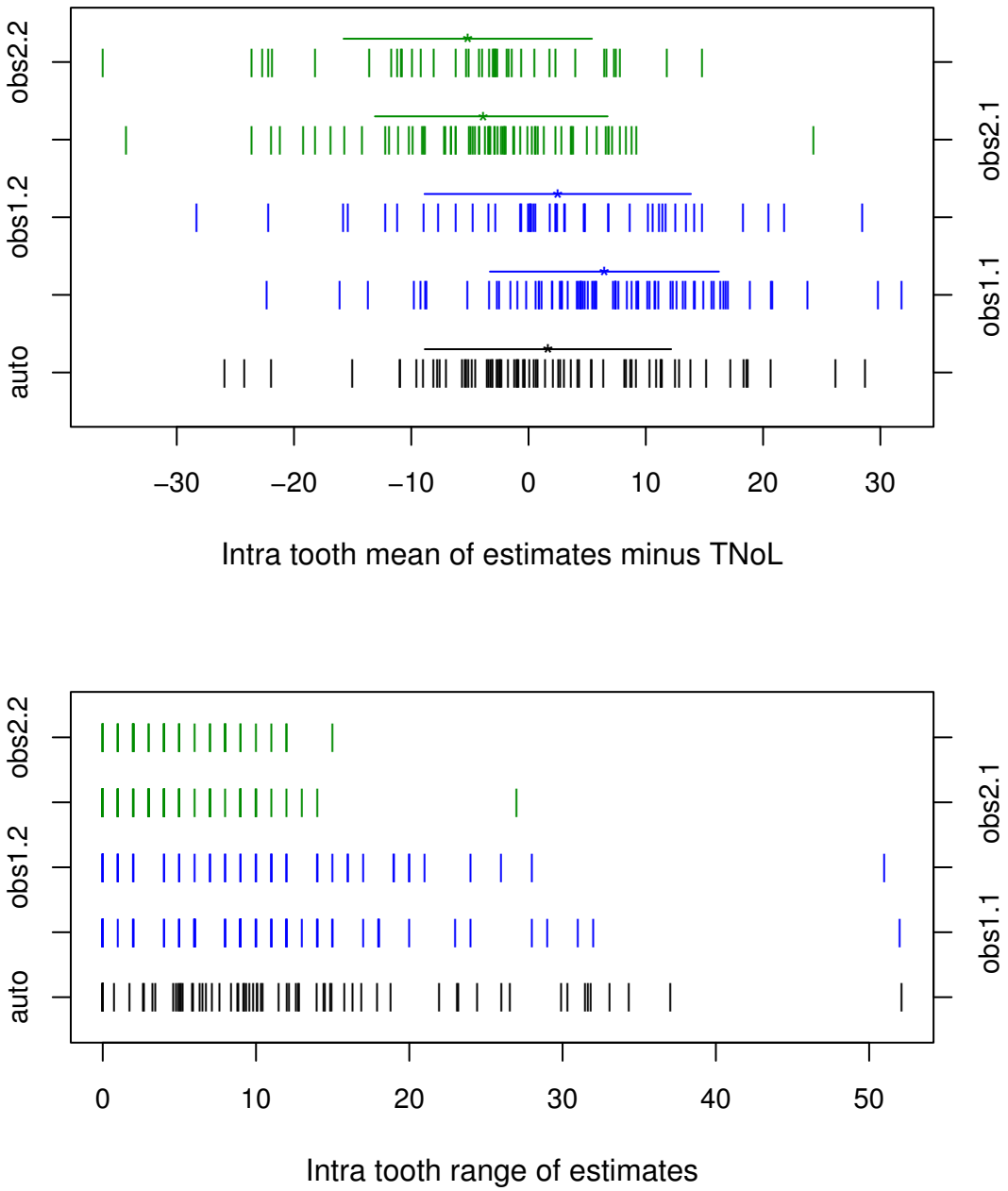


Figure 6.3: The intra tooth mean subtracting the TNoL (top) and range of estimates (bottom) by the semi-automatic counting procedure (in black) and the two observers (green and blue respectively, two observations each) for 72 selected teeth of the Basel series. Each line represents one tooth; the stars with bars indicate the overall means and standard deviations.

semi-automatic procedure is more centered around the TNoL than those of the manual observers (as their means, displayed with stars indicate). The variances of these distributions (indicated by the bars around the mean) are all very similar. It can therefore be concluded that the semi-automatic procedure is superior at the tooth level. The median as a measure of intra tooth estimates gave very similar results. For the graph of the intra tooth ranges, one must note that the manual observers estimate only integer values for the number of lines. Many teeth therefore collapse into a single point and cannot be distinguished in this graph. The intra tooth range is larger for the semi-automatic procedure than for the observers, especially compared to observer two. This was expected, because the estimates of the observers are not independent. Figure 6.3 herewith indicates that, when evaluating age estimates at the tooth level, the independent estimates of the semi-automatic procedure give more variable, but less biased, results than the a priori dependent estimates of manual observers.

Let us now turn to the calculated confidence intervals for the set of 407 selected TCA images of the Basel series. For 16 of these images (less than 4%) it was not possible to calculate confidence intervals for the estimated number of tooth rings, because the EM algorithm converged to a saddle point of the likelihood landscape. Section 6.4.1 explains details and a possible solution to this problem. During further analysis these images will be omitted. It also has to be mentioned, that one TCA image displayed an unusually large confidence interval of 109 years. This behavior remains unable to be adequately explained. Two halves of the respective TCA image were fitted separately, but both led to estimates that are saddle points. The TCA image was thus not included in the analysis of confidence intervals below. Many of the remaining 390 TCA images led to high first derivatives of the likelihood with respect to the filter parameter T, evaluated at the MLEs. This indicates that the EM algorithm did not converge to a maximum and is probably due to the discrete grid search for the parameter estimate of T, as will be explained in Section 6.4.1. Although there is the technical possibility of refining the grid during the EM algorithm, this was not realized during this first test on TCA images. In the affected TCA images, the estimates for the number of tooth rings are not drastically wrong, but the approximations of confidence intervals are meaningless. Let us therefore restrict the analysis to the 169

TCA images (43% of the TCA images where confidence intervals exist) with small derivatives of the log-likelihood with respect to the filter parameter T:

$$\left.\frac{\partial \log L\left(\boldsymbol{\theta}, T|Y\right)}{\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} < 100.$$

For the analysis of confidence intervals of these 169 remaining TCA images, the same problem as for the evaluation of the estimated number of tooth rings occurs: there is no perfect reference. The length of the confidence intervals was correlated to the difference between the estimated and the theoretical number of lines, the difference between or within the observers and the difference between the observers and the automatic or theoretical number of lines. Even the best correlation of 0.48 for the length of confidence intervals and the difference between observers is not convincing enough to draw inferences from confidence intervals on the quality of the TCA age estimate.

One could be tempted to interpret the confidence intervals for the estimated number of tooth rings in the following sense:

> If 100 TCA images from the same tooth are acquired, then for 95 TCA images, the theoretical number of lines lies in the confidence interval of the estimated number of lines.

But this statement turns out to be wrong in most TCA applications because the assumption of identical conditions when repeating the experiment is obviously violated by:

- the intra tooth variance of tooth rings,
- the intra slice variance of tooth rings and
- even the intra image variance of tooth rings (as will be demonstrated in Figure 6.17).

Additionally, the following issues may interfere with the desirable interpretation of the quality of an age estimate:

- the bias due to the TCA method, which was mentioned on page 118
- the integer-valued estimate of the filter parameter T

  Even though TCA images were limited to those with relatively low derivatives of the log-likelihood with respect to the parameter T, the

calculated confidence intervals might still be erroneous due to the remaining distance between the true, real-valued and the estimated, integer-valued $\hat{T}$.

- the assumption of a constant ring width within one TCA image

  The confidence interval around the filter parameter estimate $\hat{T}$ may thus express the variance of the tooth ring width within the TCA image. It may represent the deviance of the image from the assumption, so to speak.

Instead, the confidence intervals of the 169 TCA images have to be interpreted as a measure for the goodness of fit of the Gaussian hidden FRAME model. Confidence intervals are therefore important to state the quality of the estimate in a purely statistical sense. The statement about the quality of the TCA age estimate has to be derived retrospectively from the variance of the estimated, compared to the theoretical, number of lines in a sample (see Table 6.1).

To enable a comparison with the methods introduced in Chapters 2 as well as 5, Figure 6.4 displays the estimated number of lines compared to the theoretical number of lines for the same subset of 49 TCA images for the Gaussian hidden FRAME model. For this subset, at least two independent manual observations exist with a maximum difference of three years and tooth rings need to run roughly horizontally. They therefore present a subset of good quality from the manual observers' point of view.

6.1.4 Discussion – HMRF Modeling for TCA Images

For evaluating the fit of the Gaussian hidden FRAME model (as introduced in Chapter 4.3) to a series of TCA images, two imperfect reference measures were selected: the known age at death subtracting the tooth eruption age (the theoretical number of lines), as well as the manual counts for the number of tooth rings. These measures are corrupted by shortcomings due to the TCA method and due to subjectivity respectively, but they are the only available quantities. From the 2,120 TCA images acquired from 100 skeletons of the Basel series by the Max Planck Institute, the 407 TCA images were selected based on the criteria that a 40-fold magnification was used and at least two manual estimates for the number of tooth rings exist. Only for these TCA images the reference measure of manual counts contains adequate

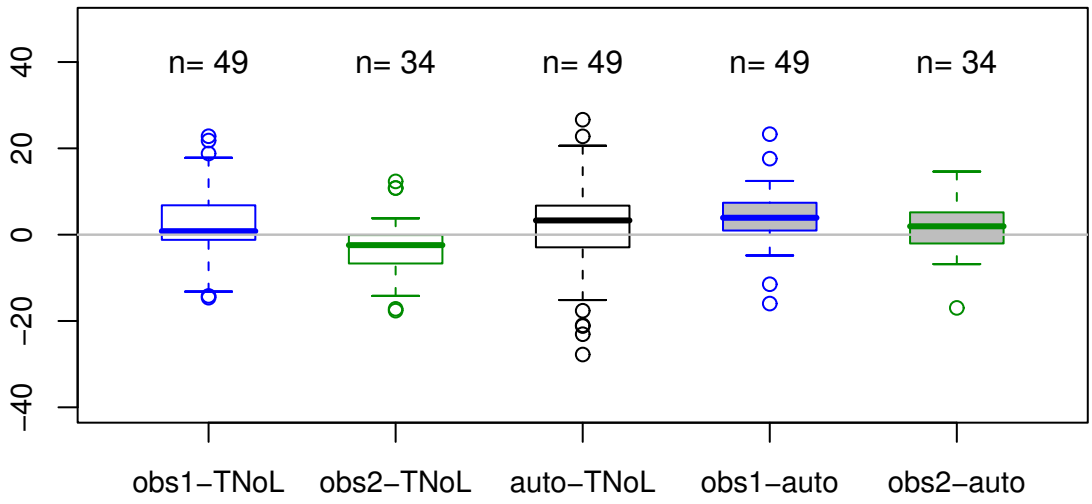


Figure 6.4: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines (TNoL) for the semi-automatic procedure ('auto') and for the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') on a set of 49 TCA images of good quality from the Basel series.

information. This subset represents the qualitatively top 40% of the Basel TCA images with 40-fold magnification from the observers' perspective.

The estimate for the number of tooth rings attained from the semi-automatic procedure is on average closer to the theoretical number of lines than the observers' estimates. The variance of the difference between the estimated and the theoretical number of lines is slightly larger for the semi-automatic procedure than for the observers. The algorithm proposed in the work is thus competitive with the manual estimates on the image level. Comparing results on the tooth level, because the manual observations of each tooth are known to be dependent, even shows that the semi-automatic procedure is superior if the estimates of one tooth are summarized by the mean or the median. The confidence intervals for the estimated number of tooth rings of TCA images have to be interpreted with caution. They measure the goodness of fit of the Gaussian hidden FRAME model to the image. A statement about the quality of the TCA age estimates has to be derived retrospectively from the variance of the estimated compared to the theoretical number of lines in a sample.

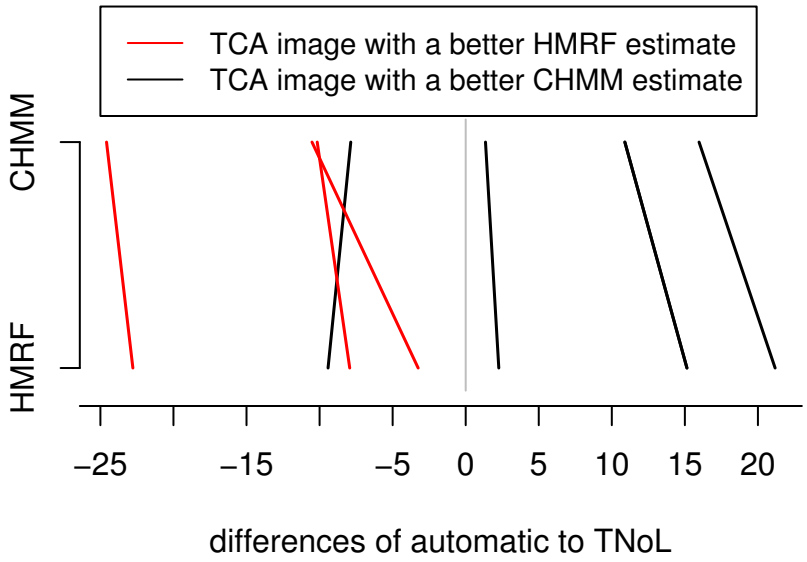


Figure 6.5: Comparing the Gaussian hidden FRAME chain with the Gaussian hidden FRAME model for the set of 49 TCA images of good quality from the Basel series. Each line depicts one of the seven TCA images resulting in a different estimate. Red lines represent images leading to a better estimate, black ones to a worse estimate, for the Gaussian hidden FRAME model compared to the theoretical number of lines (TNoL).

6.2 Evaluating the CHMM

Analogously to the previous section, where the Gaussian hidden FRAME model was evaluated, this Section will evaluate the Gaussian hidden FRAME chain model (as introduced in Chapter 5) on a subset of TCA images from the Basel series. The same two (imperfect) references will be used: the theoretical number of lines (TNoL) and the observers' estimates for the number of tooth rings. Because of the tremendous additional amount of computing time required to fit the Gaussian hidden FRAME chain to 407 TCA images, this study will restrict its attention to a subset of 49 TCA images. This subset of good quality TCA images (from the observers' perspective) was introduced in Chapter 1.1 and has already been used for evaluating the DFT and SVD based methods.

Only 7 of the 49 TCA images resulted in a different estimate for the number of tooth rings for the Gaussian hidden FRAME chain compared to

the Gaussian hidden FRAME model. Figure 6.5 depicts these seven TCA images and their difference. Red lines represent TCA images that lead to a better estimate for the Gaussian hidden FRAME model compared to the theoretical number of lines; black lines represent images with a better estimate for the Gaussian hidden FRAME chain. In general, one can say that both models, the hidden Markov random field model and the coupled hidden Markov model, lead to very similar results for this subset of 49 TCA images of good quality.

Next, will be a discussion of experiments with a thin filter consisting of one column plus one pixel, which were carried out in Chapter 5. This model will be called thin CHMM here for convenience, and will also be compared to the HMRF model for the subset of 49 TCA images. A substantial number of 28 TCA images led to different results. Figure 6.6 displays box-and-whisker plots of the difference between the estimates for the number of tooth rings of the first (of possibly two) observations of observer one, observer two and the thin CHMM procedure compared to the theoretical number of lines, as well as the difference between the observers and the thin CHMM procedure for the 49 selected TCA images. This figure can be compared to Figure 6.4 (the Gaussian hidden FRAME model) and also to Figures 2.7 (the DFT-based method) and 2.11 (the SVD-based method). The estimated number of tooth rings for the thin CHMM lead to slightly better results than the HRMF model: the median deviance to the theoretical number of lines is smaller and this deviance also has a smaller standard deviation, but it is slightly larger for the mean compared to the HMRF model.

In general, one can say that the CHMM with a standard size halved Gaborcosine filter and a thin filter lead to similar results compared to the HMRF model. The CHMM using the thin filter yields slightly better results for the tested subset of 49 TCA images of good quality. To ultimately judge the difference in the estimated number of tooth rings between the methods, a larger set of TCA images probably has to be examined. It is also possible that implementing a thin filter in the CHMM leads to worse results for TCA images of poorer quality because it is more flexible in horizontal direction than the HMRF model.

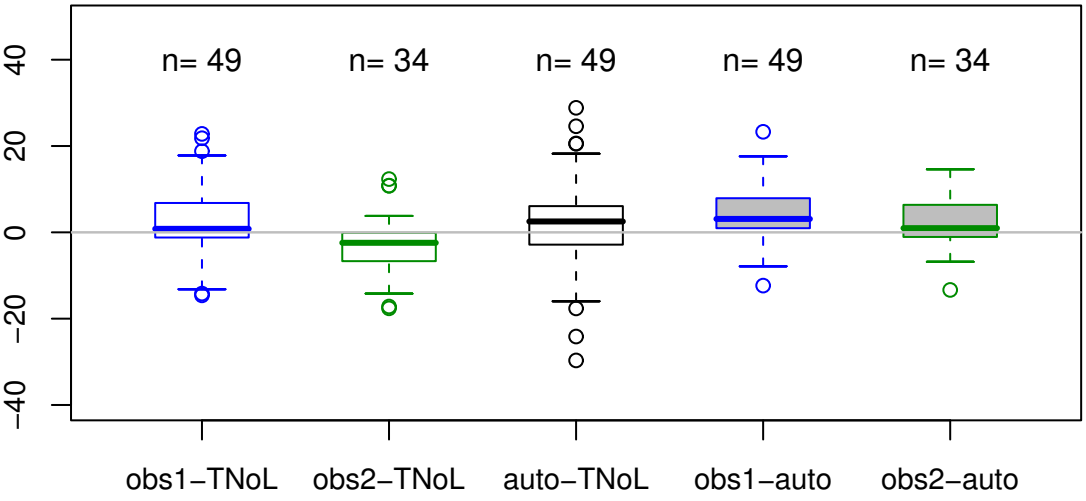


Figure 6.6: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines (TNoL) for the thin CHMM procedure ('auto') and for the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') on 49 TCA images of good quality from the Basel series.

6.3 Localization of the Hidden Markov Random Field Model

Despite the good TCA age estimate when applying the Gaussian hidden FRAME model stated in Chapter 4.3, the reader notices a bad reconstruction: most tooth rings are not well met. (See especially Figure 4.13.) This is due to the global assumptions that were made for the hidden field in the FRAME model as well as for the observable field. The reconstruction of TCA images is, on the one hand, heavily influenced by the shape of the single filter estimated for the FRAME model. The FRAME model in this form can only take into account strong local changes of tooth rings. On the other hand, it is influenced by the homoscedastic Gaussian distributions for the conditional distributions.

Chapter 5 attempted to resolve the global property of the hidden field by using a thinner filter F_T in the FRAME model and thereby reducing the size of the neighborhood in the horizontal direction. The hidden FRAME chain was one particular case of doing so, leading to horizontally more flexible tooth

rings. The reconstruction does not fit better to tooth rings than the hidden FRAME model (see Chapter 5.3) and the number of estimated tooth rings is generally similar (see Section 6.2). Section 6.3.1 will pursue another approach to the same problem: location-dependent filters. Thus, a nonhomogeneous FRAME model will be studied. Sections 6.3.2 and 6.3.3 will then try to resolve the global assumptions of the observable random field by assuming location-dependent, respectively heteroscedastic, variance parameters in the Gaussian distributions.

6.3.1 Nonhomogeneous FRAME Model

In order to make the hidden FRAME model more flexible one can choose to assume that the Markov random field parameter T is location-dependent. For this purpose one would estimate the filter parameter pixelwise as indicated in Equation (4.8). This could overcome the global property of the homogeneous hidden field that was mentioned in the TCA application in Section 4.3.6.

For the hidden nonhomogeneous FRAME model, the same parameter estimation procedure as stated in Chapter 4.3.2 can be employed, except that $T = (T_1, T_2, \ldots, T_{NM})$ should be updated by

$$\begin{aligned} T_i^{(t)} &= \arg\max_{\{T_i\}} \sum_{g=0}^{G} P\left(\lambda_i = g | Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \cdot \\ &\qquad \log P\left(\lambda_i = g | \tilde{\lambda}_{N(i)}, T_i, T_{C(i)}^{(t-1)}\right), \end{aligned}$$

instead of formula (4.12). But the following discussion shows the biased nature of estimates under this model and for the purpose of tooth ring detection:

Advice: *The parameter T of the ideal (noise free) TCA image cannot be sensibly estimated under the assumption of a nonhomogeneous FRAME model.*

Discussion. Let us assume that one can estimate the true parameter T and support the above advice by counter example.

Let $i \in \mathcal{S}$ be an arbitrary site. Then the maximum likelihood estimate of T at this site is

$$\hat{T}_i = \arg\max_{T_i} P(\lambda_i | \lambda_{N(i)}, T)$$

for any image λ. Let $\tilde{\lambda}$ coincide with λ except in i. Then the conditional probability

$$P(\lambda_i|\lambda_{N(i)}, T_i, T_{C(i)}) = \frac{e^{\sum_{j\in C(i)} |(F_{T_j}*\lambda)(j)|}}{\sum_{g\in\{0,1\}} e^{\sum_{j\in C(i)} |(F_{T_j}*\tilde{\lambda})(j)|}}$$

needs to be maximized at this site for maximum likelihood estimation. Thus

$$\begin{aligned}&\sum_{j\in C(i)} |(F_{T_j} * \tilde{\lambda})(j)| - \sum_{j\in C(i)} |(F_{T_j} * \lambda)(j)| \\ &\quad = \sum_{j\in C(i)\setminus i} \Big(|(F_{T_j} * \tilde{\lambda})(j)| - |(F_{T_j} * \lambda)(j)| \Big) + |(F_{T_i} * \tilde{\lambda})(i)| - |(F_{T_i} * \lambda)(i)|\end{aligned}$$

needs to be minimized over all T_j, which is equivalent to minimizing

$$|(F_{T_i} * \tilde{\lambda})(i)| - |(F_{T_i} * \lambda)(i)|. \tag{6.1}$$

Let us now assume λ is the ideal TCA image of Figure 4.4 with ring width $T = 12$ everywhere. Then the difference of filter responses between λ and $\tilde{\lambda}$ in (6.1) can be calculated for each column and a finite set of filters F_T, which was here chosen to be $T \in \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$. The results are depicted in Figure 6.7. For this particular image, the parameter T is estimated incorrectly everywhere, i.e. $\hat{T} \in \{2, 4, 6\}$. This is confirmed by an iterative maximum likelihood estimation $\hat{T}$ in the ideal TCA image λ by comparing the likelihood of the current T with the likelihood of $T - 2$ and $T + 2$. The result $\hat{T}$ is shown in Figure 6.8. □

As indicated above, modeling TCA images by a hidden nonhomogeneous FRAME model leads to biased results. We therefore refrain from pursuing this approach and turn to localization of the observable random field in the next two subsections.

6.3.2 Pixel-Dependent Normal Distributions

Instead of localizing the prior distribution as indicated in Section 6.3.1, one can try to localize the conditional probabilities. In Chapter 4.3 the cost function was specified by a product over identical normal distributions. This subsection will explore the possibility of specifying normal distributions whose parameters (specifically the variance parameter) differ for each pixel. This

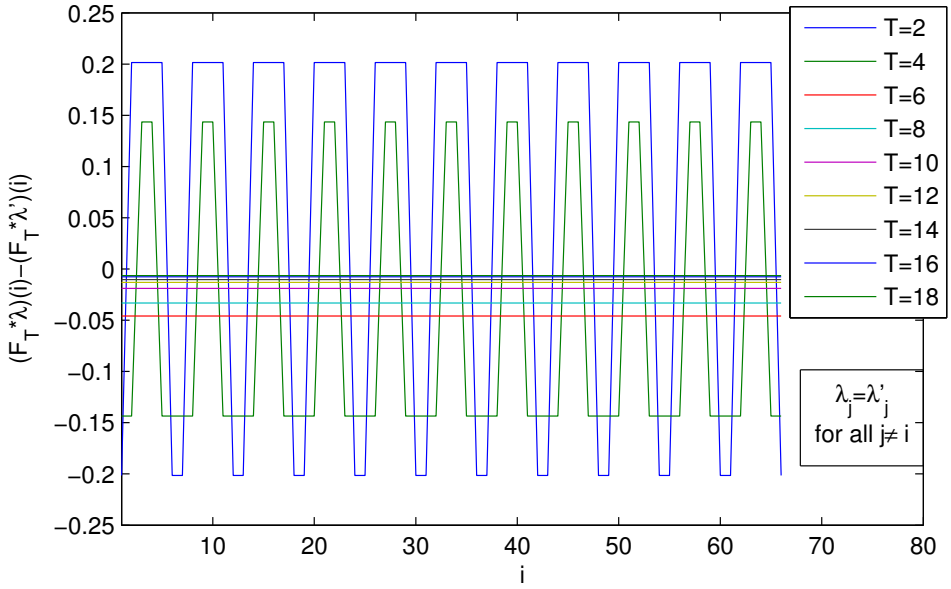


Figure 6.7: The difference between filter responses in each column of the ideal TCA image and the ideal TCA image changed at pixel i. This difference is minimized for $T \in \{2, 4, 6\}$.

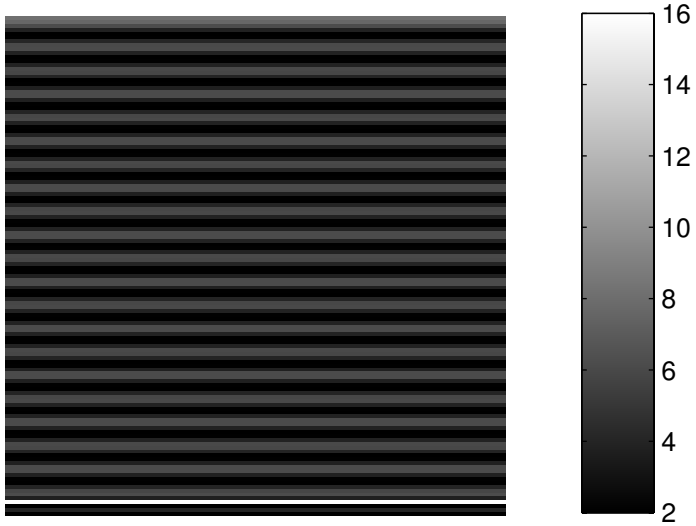


Figure 6.8: For an ideal TCA image with $T = 12$, the image shows the pixelwise estimated maximum likelihood estimate $\hat{T}$.

generalization of the model could be used to better separate regions of 'blur' and noise from regular areas. Here, this will be called a nonhomogeneous Gaussian hidden FRAME model (not to be mistaken for the hidden nonhomogeneous FRAME model of the previous subsection).

The Gaussian hidden FRAME model with a conditional distribution (4.1) and a prior distribution (4.7) will be changed to assume the following cost function:

$$\begin{aligned} f(Y|\lambda, \boldsymbol{\theta}) &= \prod_{i \in \mathcal{S}} f(Y_i|\lambda_i, \boldsymbol{\theta}_i) \\ &= \prod_{i \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_i^2}}. \end{aligned}$$

where σ_i^2 specifies the variance at pixel i. This changes the updating formula for the variance in the EM algorithm to

$$\left(\sigma_i^{(t)}\right)^2 = \frac{\left(Y_i - \mu_g^{(t)}\right)^2 P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}$$

(compare Equation (4.11)).

The nonhomogeneous Gaussian hidden FRAME model was applied to the marked cementum band of the standard exemplary TCA image in Figure 1.3. Figure 6.9 depicts the variance estimate at each pixel, clearly showing high levels of noise for saw cuts, blurry areas and parts of the large artifact. While similar mean parameters of the normal distributions are estimated, the central filter parameter of the prior distribution is underestimated: $\hat{T} = 11$. This leads to a heavy overestimation of the TCA age as well as a bad reconstruction of the tooth rings.

A possible explanation for this behavior lies in the algorithm: The nonhomogeneous model gathers more energy in the conditional probabilities compared to the traditional Gaussian hidden FRAME model. Scanning only integer values to maximize T, there is not enough energy in the likelihood left to make a long jump (of length 1) to change this parameter. The algorithm therefore stops close to the starting value of $T = 10$ and underestimates the prior parameter.

Despite the uselessness of the nonhomogeneous Gaussian hidden FRAME model for age estimation, it could still be applied to model regions of high

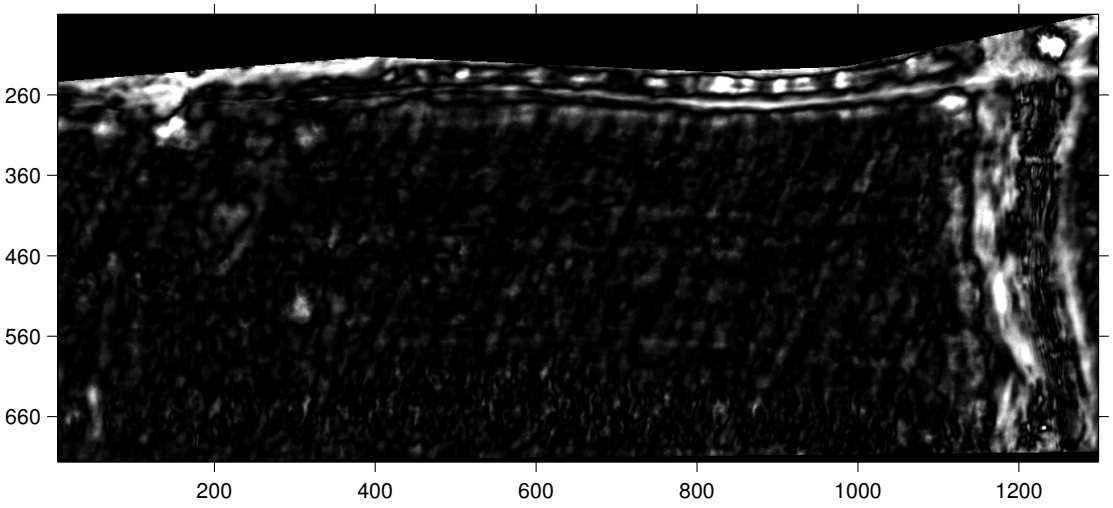


Figure 6.9: Pixelwise variance estimate $\hat{\sigma}^2$ in a Gaussian hidden FRAME model fitted to the cementum band of TCA image 1.3.

noise in TCA image analysis. But it is computationally time-consuming to do so.

6.3.3 Label-Dependent Normal Distributions

Instead of localizing the Gaussian hidden FRAME model by choosing pixel-dependent parameters as explored in the last subsections, one can try to choose label-dependent parameters. In Chapter 4.3 the cost function was specified by homoscedastic normal distributions for the application to TCA images. This subsection will explore the possibility of specifying normal distributions with label-dependent variance parameters, as theoretically stated previously in Chapter 4.3. Because of the heteroscedastic variance, this model with prior distribution (4.7) and conditional distribution (4.1) will be called a heteroscedastic Gaussian hidden FRAME model here.

Let us study the cementum band of the standard exemplary TCA image in Figure 1.3 and apply the heteroscedastic Gaussian hidden FRAME model. The reconstructed TCA image (the mean field at the last iteration) is depicted in Figure 6.10, indicating a misfit. The fitted normal distributions are shown in Figure 6.11, illustrating the problem: One normal distribution is fitted to the very dark parts of the TCA image (among others, to artifacts),

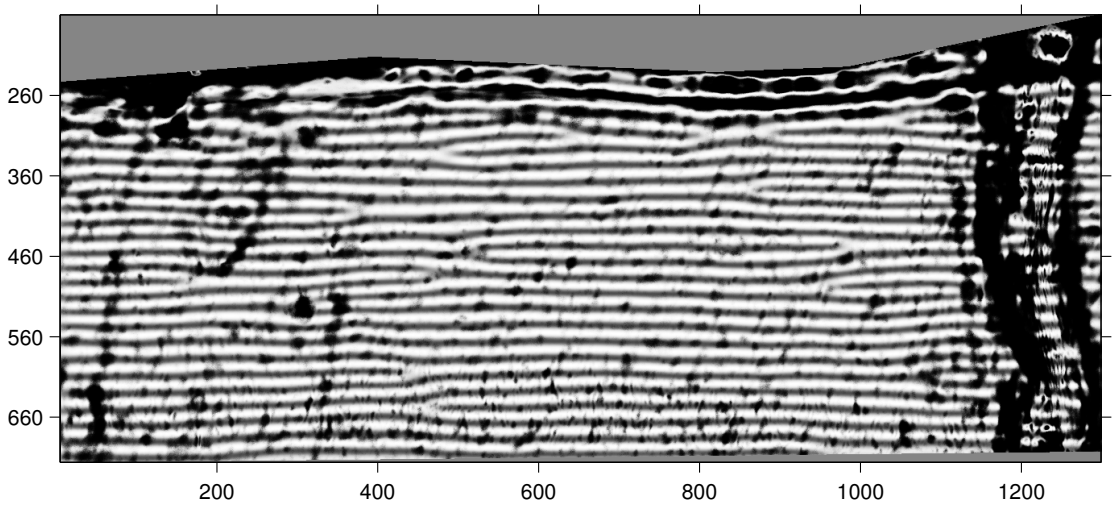


Figure 6.10: Mean field approximation at the last iteration for the cementum band of TCA image 1.3 under a heteroscedastic Gaussian hidden FRAME model.

while the second normal distribution is fitted to the rest of the image, that is, to dark as well as to bright tooth rings. This misfit is related to the identifiability problem mentioned for the traditional Gaussian mixture model in Chapter 3: the fitted groups do not resemble the grouping desirable in the application. This effect is probably amplified by the skewed distribution of the gray values.

We therefore conclude, that the heteroscedastic Gaussian hidden FRAME model is too flexible to be fitted to TCA images in the current context and we recommend applying the homoscedastic Gaussian hidden FRAME model as implemented for the experiments in Section 6.1.

6.3.4 Discussion – Localization of the HMRF Model

In the course of this section, the reader was introduced to small changes in the Gaussian hidden FRAME model with the aim of designing a more flexible (local) model, such that both the final reconstruction of the TCA image fits better to the original TCA image and that tooth rings are recognized.

Two nonhomogeneous versions of the Gaussian hidden FRAME model were introduced here. First the hidden field was allowed to incorporate a

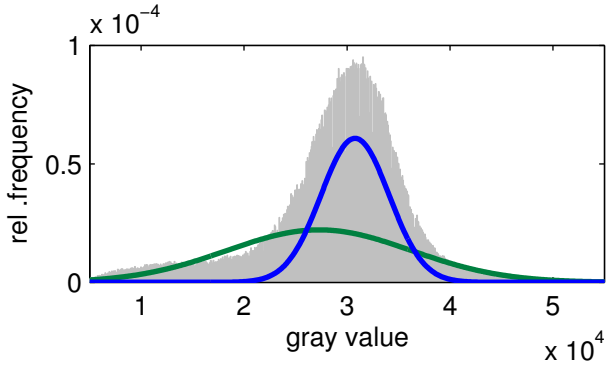


Figure 6.11: Fitted normal distributions for the cementum band of TCA image 1.3 under the heteroscedastic Gaussian hidden FRAME model (in green and blue) together with the binned observed frequency of gray values (in gray).

location-dependent filter parameter. This model destroyed the neighborhood relationship and is thus inherently not recommendable. Secondly, the observable field was allowed to incorporate a location-dependent variance parameter. Algorithmic problems appeared and thus the method was abandoned for TCA age estimation. Last, but not least, a heteroscedastic Gaussian hidden FRAME model was tested on a TCA image, resulting in a misfit. This model seems too flexible and thus does not recognize tooth rings.

These model changes were without success; consequently the next section will turn to more algorithmic and technical changes in order to improve the TCA age estimate and the reconstruction of tooth rings.

6.4 Different Algorithmic Issues and Settings

This section will discuss a diverse set of issues connected to the EM algorithm and TCA image analysis. Section 6.4.1 first discusses convergence issues of the EM algorithm in the context of TCA image analysis, including saddle points, local maxima as well as grid search problems. Then various algorithmic settings will be tested to improve the TCA age estimate and the reconstruction of tooth rings. Approximating the prior distribution with the help of a mode field (instead of a mean field) will be studied in Section 6.4.2. Partitioning the TCA image before fitting the Gaussian hidden

FRAME model to each part separately will be examined in Section 6.4.3. Finally, a discussion will close this section.

6.4.1 Convergence Issues of the EM Algorithm for TCA Images

The EM algorithm is a maximization method that (under mild conditions) converges monotonically to a stationary value of the incomplete-data log-likelihood. (See McLachlan and Peel [2000].) This subsection will discuss some of the convergence problems that occur theoretically and practically when applying the EM algorithm to TCA images. The stationary value the EM algorithm converges to does not need to be the desired global maximum, but it could be a saddle point or a local maximum. Issues connected with saddle points and local maxima will be discussed first. Then, this subsection will examine convergence issues of the EM algorithm arising from the discrete grid search for updating the estimate of the filter parameter T in the prior distribution.

A stationary point is called a saddle point (or inflection point) if the Hessian matrix of its function evaluated at the stationary point is indefinite. That is, the Hessian contains positive as well as negative eigenvalues. If the EM algorithm converges to a saddle point, a small perturbation in the direction of the eigenvector, belonging to the positive eigenvalue of the Hessian matrix of the log-likelihood, will cause the algorithm to diverge from the saddle point in the direction of a maximum. Experiments on TCA images in Section 6.1 showed that in less than 4% of the studied images, confidence intervals could not be calculated because one element on the main diagonal of the inverse observed information matrix is negative. The Hessian is thus indefinite, originating from a saddle point of the log-likelihood.

The second issue is the convergence of optimization algorithms to local maxima: If the log-likelihood has several maxima, the starting value will determine which local maximum the EM algorithm converges to. Local maxima could be circumvented by MCMC type steps or methods. But these algorithms are computationally too time-consuming in TCA applications, as mentioned before. It is therefore not obvious how local maxima can be avoided generally. The four-dimensional parameter space and the time-consuming fitting procedure also prohibit a grid study for the parameters of one exemplary TCA image. Let us therefore restrict ourselves to a grid scan-

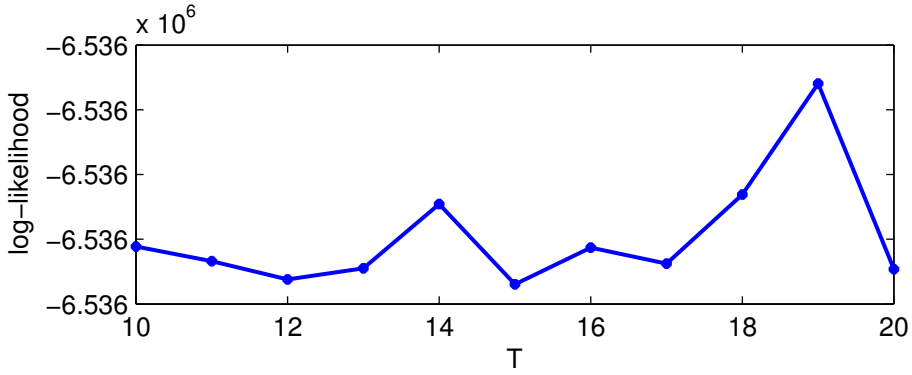


Figure 6.12: Incomplete-data log-likelihood for fitting the model to the cementum band of the standard exemplary TCA image from Figure 1.3 for fixed filter parameters T.

ning over the important filter parameter T for values $T \in \{10, 11, \ldots, 20\}$. For these fixed parameter values, Figure 6.12 shows the incomplete-data log-likelihood after convergence of the EM algorithm for the cementum band of the standard exemplary TCA image from Figure 1.3. The reader can identify at least three potential local maxima of the log-likelihood landscape. A starting value of $T = 10$ allows the EM algorithm to converge to the local maximum at $\hat{T} = 14$, as mentioned in the application in Chapter 4.3.6. In general, the small starting value of $T = 10$ leads to a smaller estimate for the filter parameter and thus generally to an overestimation of the TCA age. But the experiments in Chapter 6.1 do not show such a systematic tendency. Ignoring problems of all other parameter dimensions, the TCA experiments indicate that either the local maxima in the vicinity of $T = 10$ give the correct estimates for this application, or that in most cases, the EM algorithm converges to a global maximum for TCA images. The small starting value of $T = 10$ is also computationally recommendable, because large starting values for the filter parameter would involve large neighborhood structures and thus slow down the EM algorithm.

Let us thirdly discuss problems involved with a discrete grid search for the maximum likelihood estimate $\hat{T}$ of the filter parameter. On the one hand, the true maximum likelihood estimate might not be an integer value. The EM algorithm using a discrete grid (as stated in Chapters 4.3.2 and 4.3.6) will converge to the integer close to the true estimate and the derivative of the log-likelihood at the parameter estimates indicates the distance to the true

estimate. Since the distance between the true MLE and the integer-valued estimate is less than one pixel, the influence on the estimated number of tooth rings is very limited. But the calculation of confidence intervals is pointless for TCA images with a large distance between the true and the integer-valued estimate. Future research should choose a finer grid for the filter parameter T after a certain number of iterations in the EM algorithm in order to achieve meaningful confidence intervals. But the study in Section 6.1 restricted the analysis of confidence intervals to TCA images where derivatives of the log-likelihood with respect to the filter parameter T are less than 100. On the other hand, the grid of filter parameter values was limited to the range of 2 to 20 pixels for the average ring width in TCA images (see Chapter 4.3.4). This decision was based on the increasing computational complexity for increasing filter parameters and the empirical width of incremental lines, which ranges roughly from 10 to 40 pixels. Out of the 407 TCA images, the parameter estimate $\hat{T} = 20$ lies at the border of this range in 69 of these images for the experiments in Section 6.1. The EM algorithm might not have converged to a maximum, but to a boundary value for these images. Experiments were carried out with a maximum filter parameter of $T = 35$, which covers the range of possible tooth ring widths better. The aforementioned 69 TCA images with an estimate of $\hat{T} = 20$ were refitted to the Gaussian hidden FRAME model with the new bound. Most of these images overestimated the average ring width and therefore underestimated the number of tooth rings. Compared with using a maximum boundary value of $T = 20$, worse results were produced with the new bound.

6.4.2 Mode Field Approximation

In order to be able to fit the Gaussian hidden FRAME model containing large neighborhoods to large TCA images, the EM algorithm was suggested and the prior probability was approximated using the mean field theory in Chapter 4.3.2. The mean field approximation is not the only possibility for realizing the approximation (4.9). Celeux et al. [2003] also mention mode field approximation, which will be tested for quality and speed in the case of TCA image analysis in this subsection.

For mode field approximation, the fixed configuration $\tilde{\lambda}_{N(i)}$ in the approximation

$$P(\lambda) \approx \prod_{i \in \mathcal{S}} P\left(\lambda_i | \tilde{\lambda}_{N(i)}\right)$$

is set to the most frequent label value:

$$\tilde{\lambda}_j = \arg\max_{g \in \mathcal{G}} P\left(\lambda_j = g | \tilde{\lambda}_{N(j)}, T\right) \quad \text{for all} \quad j \in N(i).$$

The EM algorithm stated on page 79 can then be implemented analogously using mode field approximation. The computational expenditure decreases because the prior (and conditional) distribution in the neighborhood of each site (see steps 11 and 12 in the EM algorithm on page 79) do not have to be calculated for each label value $g \in \mathcal{G}$ and the fixed label $\tilde{\lambda}_i$, but only for the $|\mathcal{G}| - 1 = G$ new candidate label values. For two possible labels $\mathcal{G} = \{0, 1\}$, the time expenditure should be limited to roughly half of the time required for the mean field approximation.

At first the Gaussian hidden FRAME model using mode field approximation is fitted to the cementum band of the standard exemplary TCA image from Figure 1.3. The mode field approximation of the final iteration is depicted in Figure 6.13 and can be compared to the mean field approximation in Figure 4.12. The reconstruction using mode field approximation looks very similar to the reconstruction using the mean field approximation and thresholding subsequently. The positions of some bifurcations differ in the two reconstructions, which can probably be attributed to the random updating order in the proposed EM algorithm. The parameter estimates for this TCA image applying mean or mode field approximation are also very similar: the filter parameter estimate $\hat{T}$ (and thus the estimated number of tooth rings) is the same, the variance $\hat{\sigma}^2$ of the normal distributions is similar, while the means of the normal distributions $\hat{\mu}_g$ are slightly further apart and the log-likelihood value slightly increased for the mode field approximation. The computational time required to fit the Gaussian hidden FRAME model using mode field approximation to this exemplary TCA image is about 6.8 hours on Hydra[1], compared to about eight hours for the mean field approximation.

Secondly the Gaussian hidden FRAME model using mode field approximation will be evaluated for a whole set of TCA images from the Basel series. For this first study, let us restrict our attention again to the subset of 49 TCA images of good quality (from the observers' perspective) which was already used for the DFT and SVD based methods as well as for the CHMM. For 32 of these TCA images, the mode field based algorithm yields different

[1] Hydra is a HP ProLiant DL 580 computer with Windows 2000 Server operation system, four Intel Xeon MP 1.60 GHz processors and 2 GB memory.

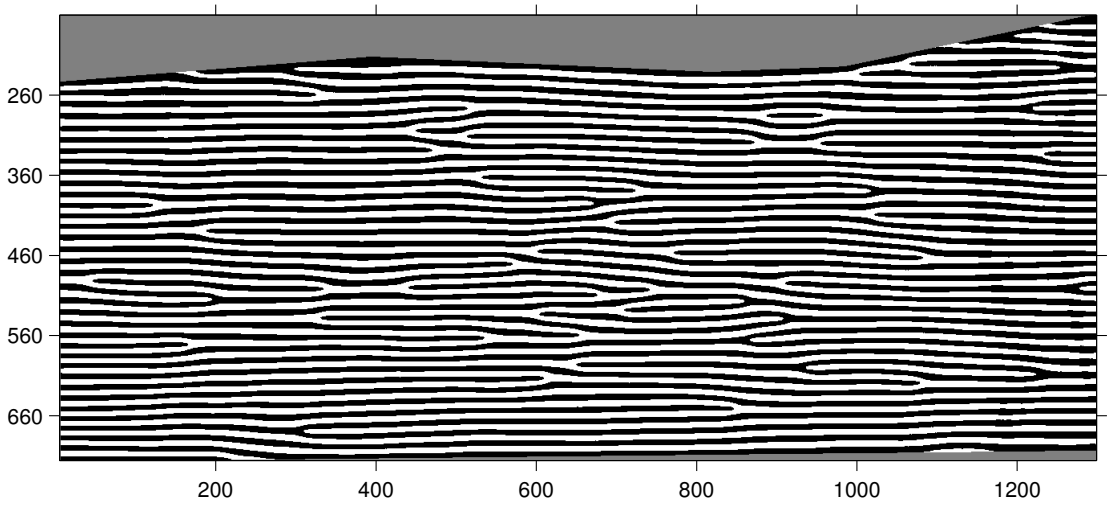


Figure 6.13: The mode field approximation at the last iteration for the cementum band of TCA image 1.3.

results than the mean field based algorithm. Fifteen of these 32 TCA images coincide with the images giving different results for the thin CHMM. Figure 6.14 displays box-and-whisker plots of the difference between the estimates for the number of tooth rings of the first (of possibly two) observations of observer one, observer two and the mode field based procedure compared to the theoretical number of lines (TNoL), as well as the difference between the observers and the mode field based procedure for these selected TCA images. This figure can be compared to Figures 6.4 (the Gaussian hidden FRAME model) and 6.6 (the thin Gaussian hidden FRAME chain), as well as to Figures 2.7 (the DFT-based method) and 2.11 (the SVD-based method). The estimated number of tooth rings for the mode field approximation lead to slightly worse results than for the mean field approximation: the mean deviance to the TNoL is larger and this deviance also has a slightly larger standard deviation. The median deviance to the TNoL is slightly smaller compared to the mean field approximation. But one 'outlier' heavily influenced this comparison: the TCA image leading to a deviance of -38.5 years to the theoretical number of lines. The average time required for the mode field based procedure is 8.8 hours, compared to the 11 hours of the mean field based procedure.

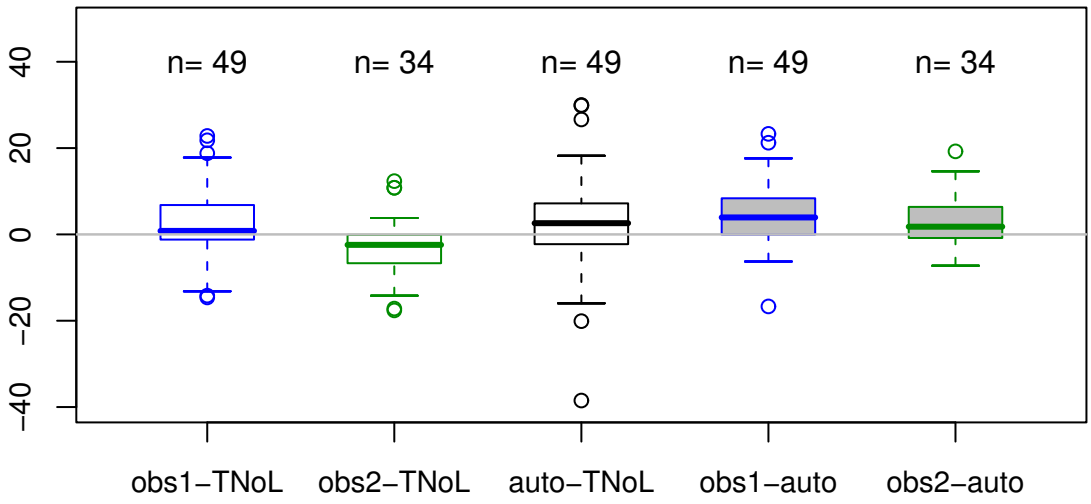


Figure 6.14: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines (TNoL) for the Gaussian hidden FRAME model applying mode field approximation ('auto') and for the first (of possibly two) observation(s) from the two observers ('obs1', 'obs2') on 49 TCA images of good quality from the Basel series.

In general, one can say that the Gaussian hidden FRAME model applying mode field approximation lead to similar results compared to the mean field based model suggested in Chapter 4.3. The mode field based procedure yields slightly worse results for the tested subset of 49 TCA images of good quality (from the observers' perspective). But to ultimately judge the difference in the estimated number of tooth rings between the methods, a larger set of TCA images probably has to be examined. The benefits of the procedure proposed in this section would clearly lie in the average time required to fit the model to one TCA image. The Gaussian hidden FRAME model using mode field approximation is about 20% faster than the procedure using mean field approximation. This is not the time gain that was expected theoretically, but it is still substantial.

6.4.3 Partitioning the Image

The Gaussian hidden FRAME model as introduced in Chapter 4.3 and applied to TCA images in Chapter 6.1, assumes a constant width as well as a

constant and fixed orientation for tooth rings throughout a TCA image. The orientation is expected to be horizontal or can be fixed manually otherwise. This assumption of constant width and orientation is clearly not satisfied in most TCA images and can be examined critically, for example, in the standard TCA image from Figure 1.3. Model-based approaches to circumvent this problem by a non-homogeneous FRAME model or thinner filters in a FRAME chain failed in Section 6.3.1 and Chapter 5 above. This subsection will now attempt to quantify the effect of the assumption of a constant ring width and orientation by partitioning TCA images manually and fitting separate Gaussian hidden FRAME models to each part.

Let us separate the problems by studying two exemplary TCA images: one with a prevailing non-constant tooth ring width as depicted in Figure 6.15 and one TCA image with a non-constant orientation as depicted in Figure 6.16. Both figures also show the partition of each image as suggested by the author. The Gaussian hidden FRAME model was fitted to each part of these two images separately, assuming a horizontal orientation in Figure 6.15 and three different orientations for the tooth rings in Figure 6.16. For both TCA images, the filter parameter T was allowed to vary across the image. Figures 6.17 and 6.18 display the mean field approximation (clearly reflecting the assumed partition), as well as the estimated filter parameters.

In Figure 6.17, a number of very different filter parameters T are estimated, leading to very distinct reconstructions. For the middle left part, $\hat{T} = 19$ was estimated and a poor quality mean field was approximated. The original TCA image contains very little information on tooth rings in this part, such that the filter parameter is probably mainly inferred from 'noise'. The lower right as well as the upper left part of the original TCA image contain artifacts which probably heavily influence the fitting procedure in these parts. Averaging the estimates for the number of lines for each part of the TCA image gives an estimate of 63.36 tooth rings, compared to 62.53 rings for fitting the whole TCA image and 57.2 for the theoretical number of lines. In general, one can say that the difference between separate estimates for several parts of the image to one estimate for the whole image is small for this example. Estimating even smaller parts of the TCA image should give theoretically better results because the width of tooth rings may vary quickly over the image. But the fit of a small part is heavily influenced by its quality (for example, artifacts or 'general noise') and information is lost at the boundaries. Partitioning TCA images into smaller parts is also impractical due to the extra time for selecting meaningful partitions.

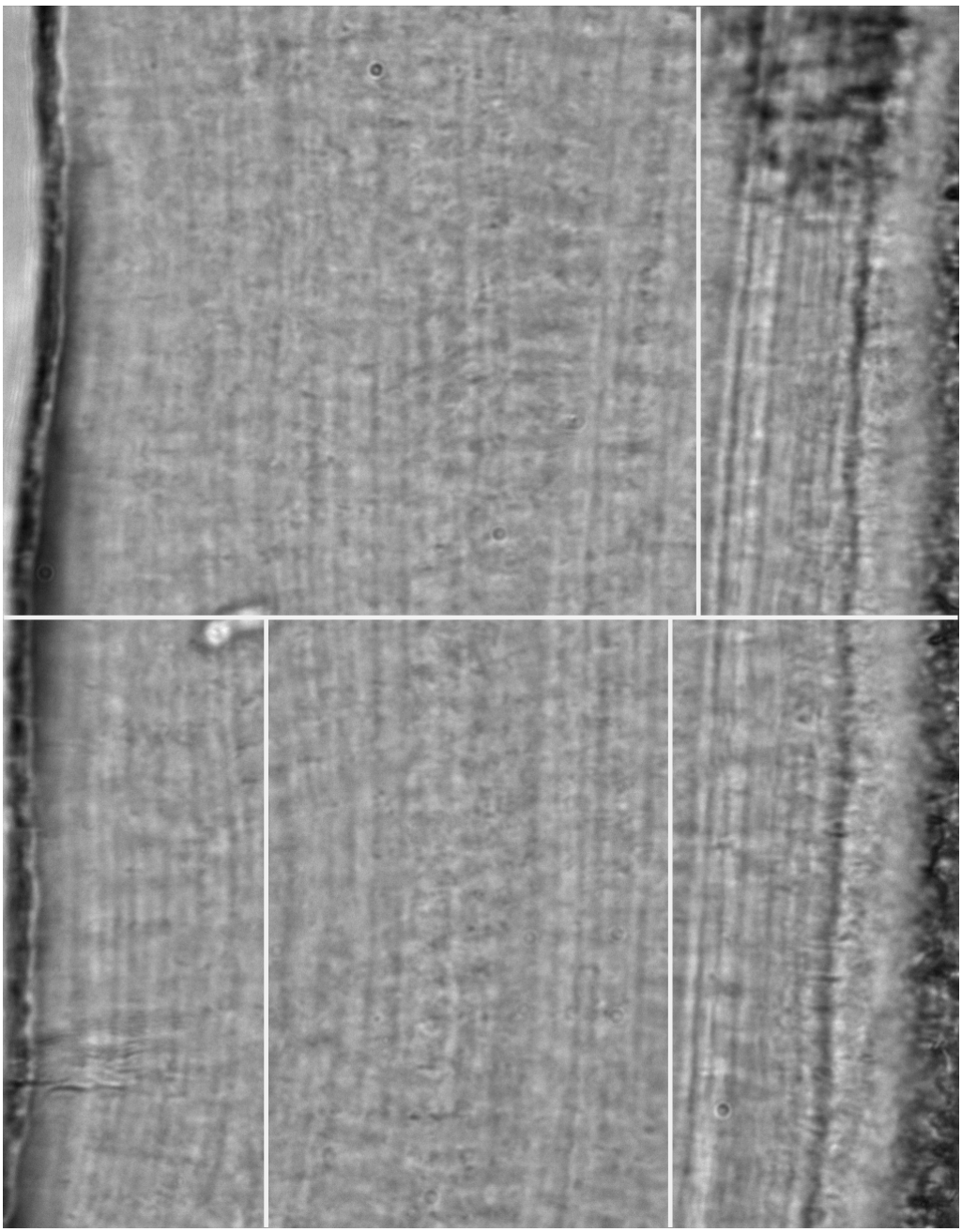


Figure 6.15: One TCA image with a non-constant tooth ring width (IS-0003429 from the TCA database of the MPIDR).

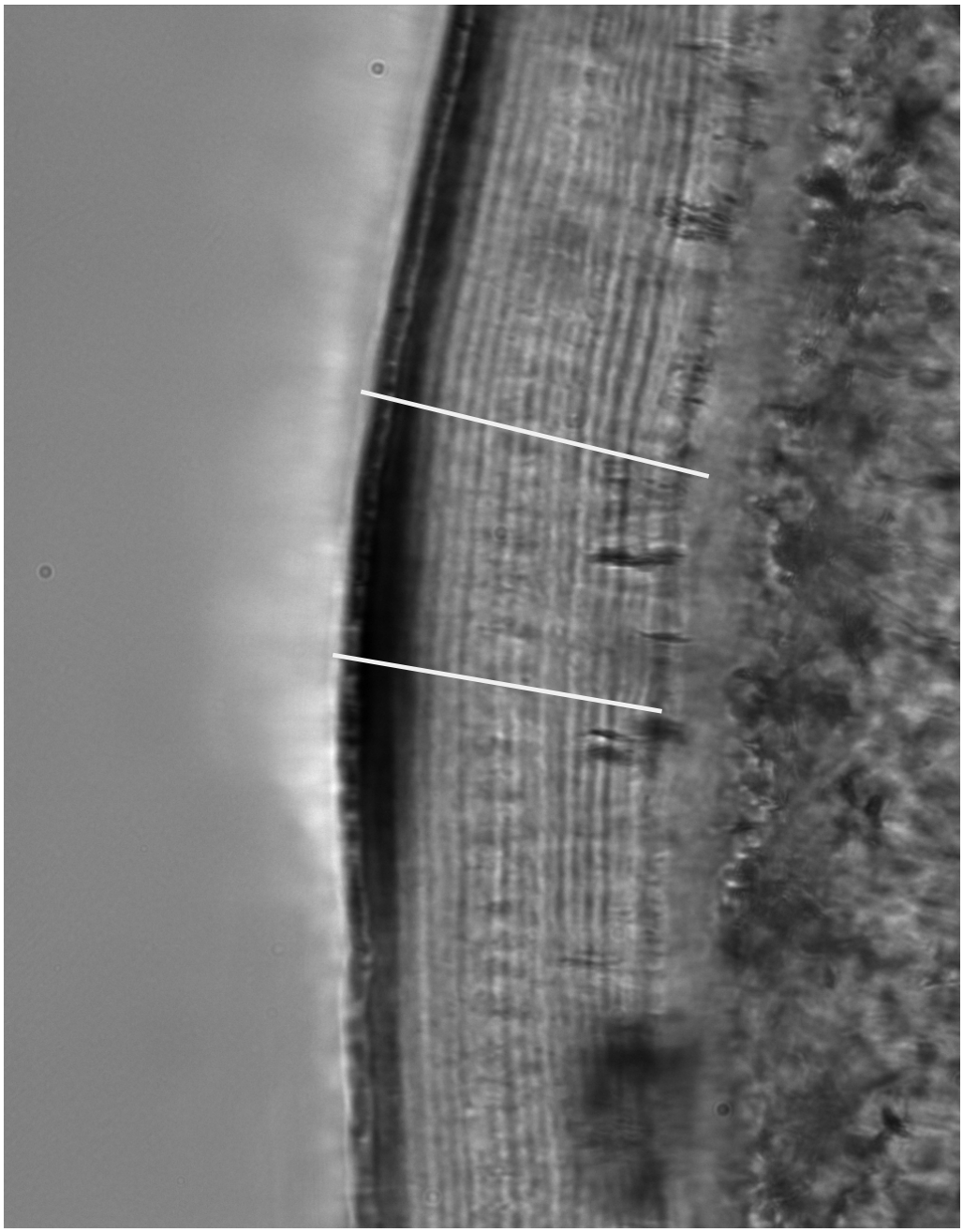


Figure 6.16: One TCA image with a non-constant orientation of tooth rings (IS-0002852 from the TCA database of the MPIDR).

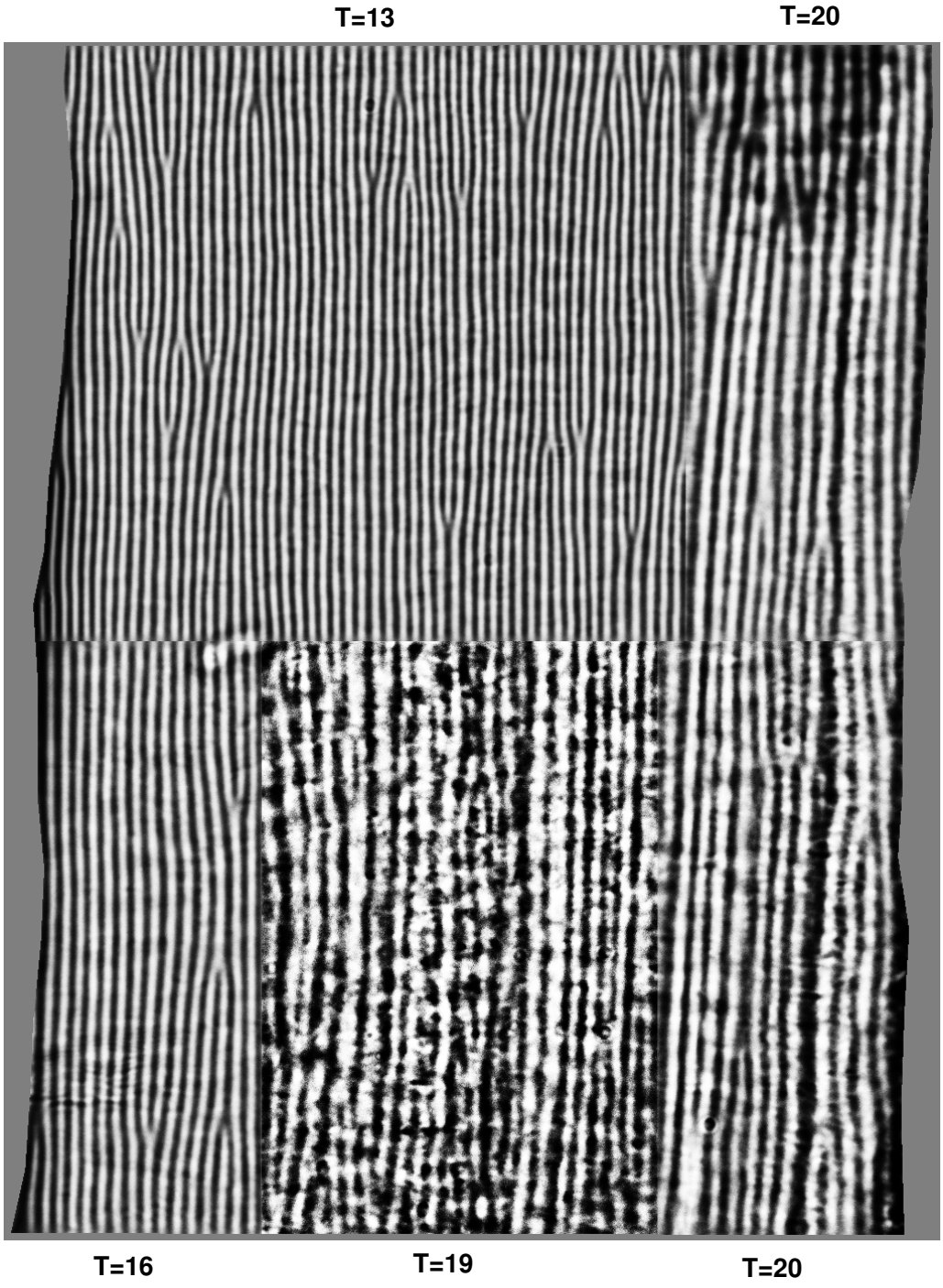


Figure 6.17: The mean field approximation at the last iteration for the partitioned TCA image from Figure 6.15 together with the estimated filter parameters $\hat{T}$.

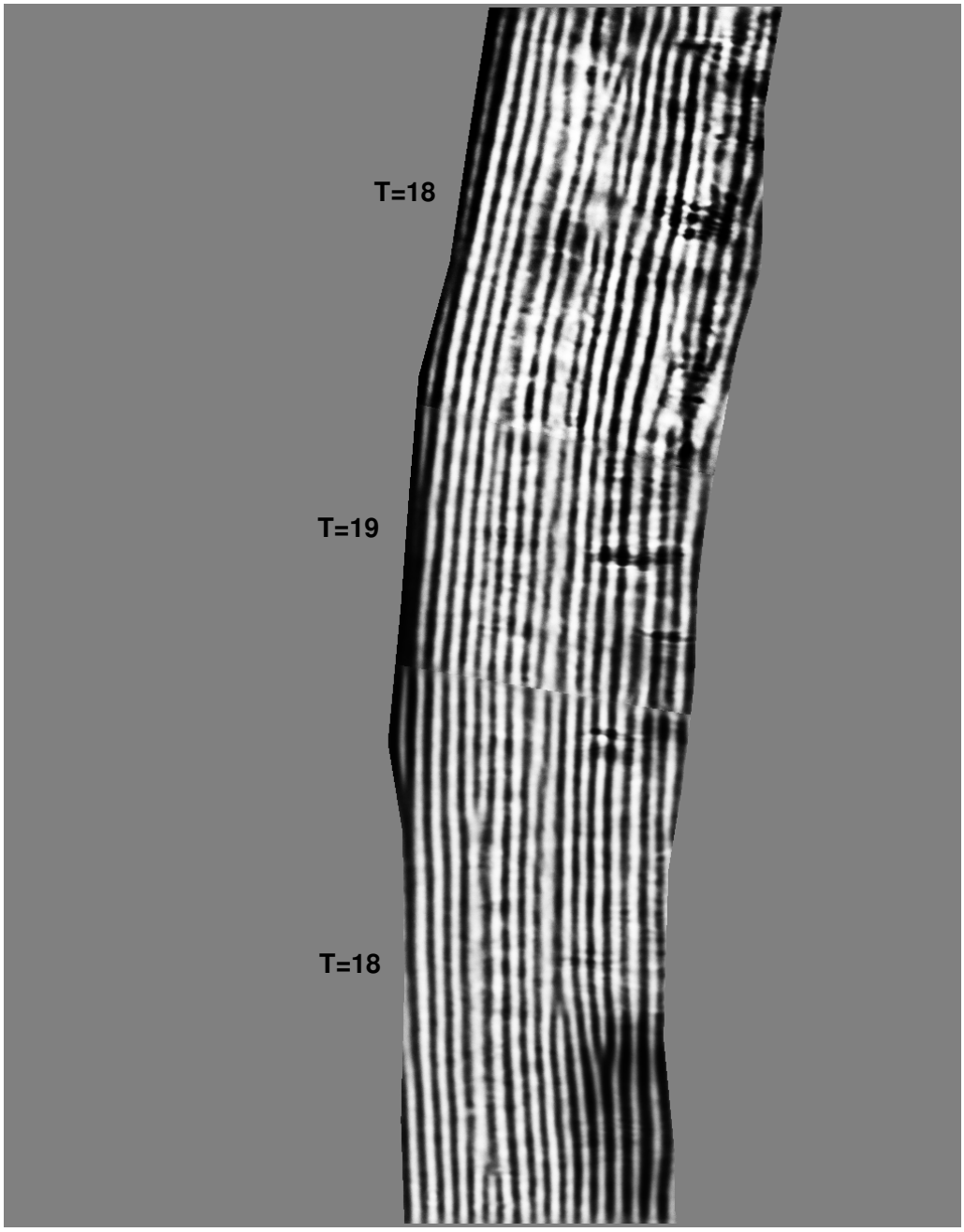


Figure 6.18: The mean field approximation at the last iteration for the partitioned TCA image from Figure 6.16 together with the estimated filter parameters $\hat{T}$.

Figure 6.18 shows the mean field approximation using the orientations 0°, -5° and -10°. It clearly fits the orientation of tooth rings better than a single orientation of 0°. The estimated filter parameter differs only for the middle part of the image, in which a relatively large artifact exists. The number of tooth rings is estimated to be 14.93 for the partitioned image and 17 for the whole image compared to 21.19 for the actual number of theoretical lines. Although partitioning the image has only a small effect on the estimation of the filter parameter T in the current example, the (negative) effect for the estimated number of lines is noticeable due to the thin cementum band.

It can be concluded that partitioning TCA images raises the dilemma of how to optimally select the parts while taking into account the behavior of tooth rings as well as the existence of artifacts and noise. In addition to showing little effect for the observed TCA image, partitioning images because of varying ring widths is impracticable because of the quickly varying tooth ring widths. Instead, partitioning images because of varying tooth ring orientations seems more promising for achieving a better reconstruction of the TCA image, and it should also result in better estimates for the number of lines if the filter parameter T is held constant over the whole image. But more extended tests should be performed in the future in order to verify this.

The increasing speed of new computers could also help to fit a Gaussian hidden FRAME model with two filter parameters to TCA images. The second filter parameter could estimate the optimal (constant) orientation of tooth rings, instead of manually guessing at it.

6.4.4 Discussion – Algorithmic Issues and Settings

This section has been devoted to discussing a diverse set of issues connected to the EM algorithm and TCA image analysis.

First, convergence issues of the EM algorithm in the context of TCA image analysis were addressed. Attention was focused on saddle points that the EM algorithm might converge to. A criterion for recognizing saddle points and a possibility of diverging from them was pointed out. The existence of local maxima, which the EM algorithm might converge to, was admitted, but no solution could be found. Only experimental evidence that the proposed compound procedure gives reasonable results for the current application can be given. The discrete grid search for the maximum likelihood estimate $\hat{T}$ of the filter parameter posed a third possible source of error. The effect on the estimated number of tooth rings is limited. At the same time, calculating

confidence intervals is pointless; but this can easily be reworked through choosing a finer grid for the filter parameter T after a certain number of iterations in the EM algorithm. Limiting the grid of filter parameter values to the range of 2 to 20 pixels for the average ring width in TCA images is computationally recommendable and also proved experimentally sensible for TCA applications.

This section considered mode field approximation instead of fitting the Gaussian hidden FRAME model to TCA images, with the help of approximating the prior probability by applying mean field theory. The two procedures lead to similar results for the tested subset of 49 TCA images; but to ultimately judge the difference in the estimated number of tooth rings, a larger set of TCA images has to be examined. The benefits of the mode field-based procedure would clearly be the decreased average time required to fit the model to one TCA image (about 20% less compared with using mean field approximation).

The Gaussian hidden FRAME model, as introduced in this work, assumes a constant width as well as a constant and fixed orientation for tooth rings throughout a TCA image. This assumption is clearly not satisfied in most TCA images. Since model-based approaches (from Section 6.3.1 and Chapter 5) to circumvent this problem failed, this subsection has attempted to quantify the effect of the assumption of constant ring width and orientation by partitioning TCA images manually and fitting separate Gaussian hidden FRAME models to each part. In general, partitioning TCA images raises the dilemma of how to optimally select the parts, taking into account the behavior of tooth rings as well as the existence of artifacts and noise. Partitioning images because of varying ring widths proved to be impracticable, while partitioning images because of varying tooth ring orientations seems more promising for the reconstruction of TCA images. For the estimated number of tooth rings, more extended tests should be performed in order to quantify the effect of partitioning TCA images.

Finally, in general, it is worth stating that faster computers could avoid or circumvent some of the problems discussed in this section.

6.5 Application of the HMRF Model on Related Problems

This section will be a short digression into layered structures in nature, other than in human tooth cementum. The examples that will be given demonstrate the range of possible applications for the hidden FRAME model.

The first subsection will focus on tree-rings. It will illustrate how the same model used for TCA image evaluation can be applied to cross sections of trees by slightly altering the fitting procedure. Section 6.5.2 will give an overview of other layered structures in nature.

6.5.1 Tree-Rings

Dendrochronology is the method of scientific dating and interpreting past events based on the analysis of tree rings. During the annual growth season, most trees produce growth rings, whose patters are mainly determined by the amount of available precipitation or by prevailing temperatures (*The New Encyclopedia Britannica: Micropï¿½ï¿½ia - Ready Reference* [1998]). By comparing, lining up and interlinking the patterns of tree-rings, so-called tree-ring calendars can be reconstructed, which can date back wood as long as 8,000 or more years ago. These tree-ring calendars are of particular interest to archeologists and climatologists. (*Brockhaus: Die Enzyklopädie* [1996])

Tree rings extend from the bark to the pith at the center of trees and can be examined by destructive or nondestructive methods. Nondestructive methods such as computer tomography acquire reconstructions of cross sections of live trees, which can be used, for example, in lumber milling or on wooden columns of existing buildings. Destructive methods include drilling cores and cutting cross sections. (Onoe et al. [1984])

Because such cross sections from trees are fairly similar to cross sections of human teeth, methods for TCA image analysis could also be applied to photographs of cross sections of trees. Figure 6.19 displays an image of a cross section of a Douglas-fir tree from the Zuni Mountains (New Mexico, USA).

Cross sections of trees usually display tree-rings at all orientations. Methods from TCA image analysis thus need to be adapted for this application. Specifically, the Gaussian hidden FRAME model, introduced in Chapter 4.3, will be used to include different, but fixed, filter orientations. In order to

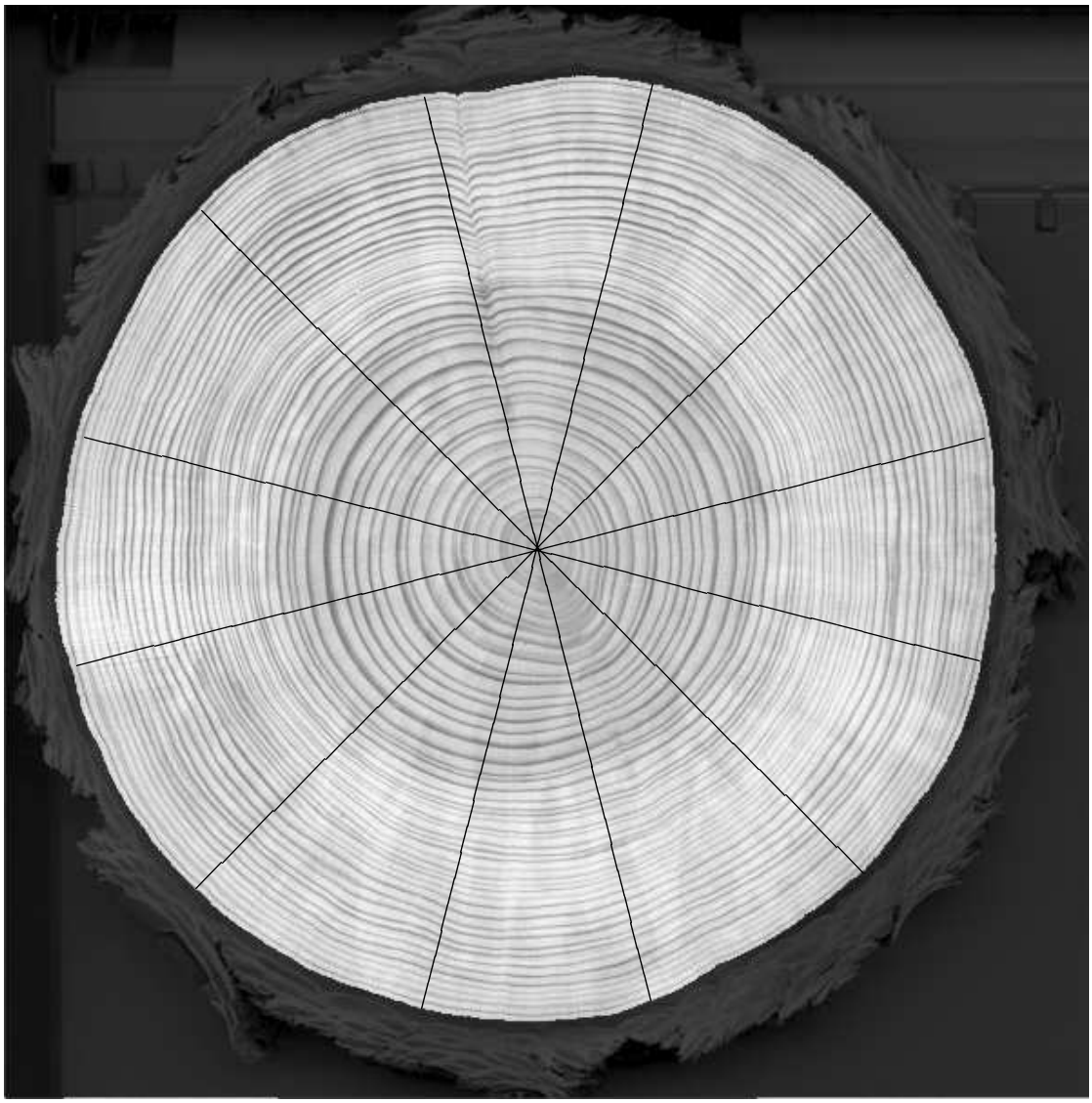


Figure 6.19: Cross section of a Douglas-fir tree from the Zuni Mountains (New Mexico, USA), divided into 12 sectors.
Source: Grissino-Mayer [2006]

do so, the author manually marked the center as well as the outermost tree-ring in cross sections of trees. This region of interest was then automatically divided into 12 sectors. (See Figure 6.19 for an examplary cross section.) Following, the estimation procedure stated on page 79 can be applied to coherently fit the Gaussian hidden FRAME model to the entire cross section, assigning one fixed orientation to each pair of opposing sectors.

Figure 6.20 displays the reconstruction of the cross section from the Douglas-fir tree of Figure 6.19. One can recognize the inner dark rings and the outer light rings, as well as some disturbances in the tree-ring formation. Some transitions from one to the neighboring sector of the cross section (caused by applying different filter orientations) are also visible in the reconstruction. These uneven transitions indicate that a finer division of the cross section into sectors could give improved results. The more regular transitions, which are not visible in the reconstruction, demonstrate that the Gaussian hidden FRAME model tolerates deviations of up to 15° in the orientation of clear ring structures. The parameters were estimated to be the means $\hat{\mu}_0 = 169.7$, $\hat{\mu}_1 = 203.5$, the common variance $\hat{\sigma}^2 = 243.8$ and the average tree-ring width $\hat{T} = 5$. The number of tree-rings of this Douglas-fir tree was thus estimated to be 51.4. Experts estimate the age of this tree to be 50.4, excluding the most inner pith.

This exemplary application of the Gaussian hidden FRAME model to estimate the age of trees from cross sections produced very good results. At the same time, it demonstrated the flexibility and strengths of this model, and proved that it is not solely applicable to TCA images. Remarkably, no changes to the model had to be made, only the fitting procedure had to be altered slightly in order to generalize from tooth rings to tree-rings.

For this particular application to tree-rings, there are certainly faster, simpler and more direct methods for estimating the number of tree-rings and their predominating patterns. The Gaussian hidden FRAME model probably cannot compete with these methods; but similar applications become conceivable where this model can confirm its value. Some of these applications will be mentioned during the next subsection.

6.5.2 Further Layered Structures

Layered structures frequently appear in nature. If there is a possibility of acquiring digital images of these structures, they could also be analyzed by the Gaussian hidden FRAME model (or a modified version of this model).

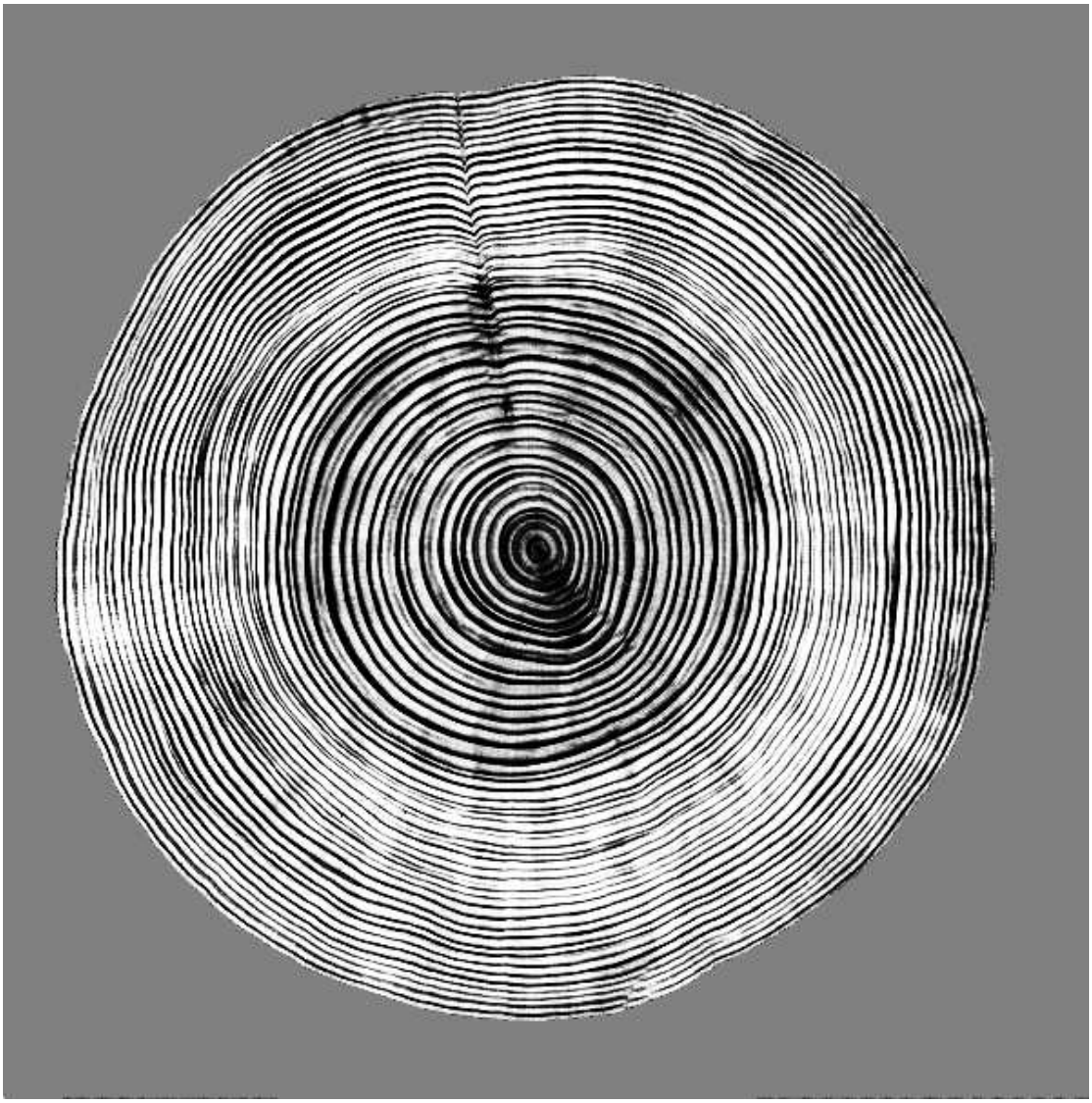


Figure 6.20: The mean field approximation of the cross section from the Douglas-fir tree from Figure 6.19.

This subsection will provide a list of examples of layered structures. Some examples, express daily patterns, some seasonal patterns and some provide other kinds of information. This list is by no means complete, but serves to indicate the possible range of conceivable applications of the Gaussian hidden FRAME model. Most of the layered structures mentioned below can be researched in subject-specific encyclopedia or combined in the online encyclopedia *Wikipedia, The Free Encyclopedia* [2006].

Human teeth do not only form rings in the cementum, but also in the enamel of the crown. These incremental lines are known as cross-striations, which form daily, and as Retzius lines, which display long-period structures. Both structures are, among other things, used for evolutionary research.

One does not need to focus on humans. Tooth cementum annulation is also found in a wide range of other mammals and applied for age estimation, as mentioned in the introduction. But other parts of animals also show annular patterns. For example, most whales grow earwax in layers each year, birds display thin rings in their bones and stress bands in their feathers. Some fish show seasonal rings in their otolith (an ear structure) and in their scales. Annual banding can also be found in coral colonies and is used analogously to dendrochronology there.

Not only in biology, but also in geology, similar layered structures can be found. For example, ice cores and varves display annual layers of snow or sediment, respectively. Information from both structures is of interest in climatology.

More generally, one-dimensional wave-like patterns exist as well, and the hidden FRAME model could be adapted to these applications. But this topic extends into the research area of time series analysis and will thus not be discussed here.

Chapter 7

Conclusions

Tooth Cementum Annulation (TCA) images are microscopic images from the root of human teeth. These digital images display annual incremental lines which can be used for age estimation. Until now, the incremental lines were counted manually under the microscope or from the (digital) TCA images. This work has developed the first model-based method for evaluating TCA images. It therefore provides an invaluable tool for further anthropological research on the TCA method.

This final chapter will first summarize the work by briefly mentioning the methods which where examined for TCA image evaluation. Subsequently, results which were obtained through using these methods will be reported. The last section will emphasize the value of this work and give an outlook on the approaches in order to improve the recommended methods and to improve the TCA age estimate with the help of a semi-automatic counting procedure.

7.1 Summary on Methods

In order to develop a semi-automatic procedure for the evaluation of TCA images, initially two approaches were examined that though not successful, provided valuable insights. Three feature measuring methods were studied; but they could not bridge missing or falsified information, which frequently occur in TCA images. Still, these methods served to illustrate characteristics of TCA images and to emphasize the distinction between two aspects of TCA image analysis: recognizing tooth rings and estimating their number.

The second approach relied on mixture models, which did not reveal reconstructions that resemble tooth rings. The reason lies in the independence assumption made about the observations. It is known that the gray level of one pixel in an image is highly dependent on gray levels of neighboring pixels. Mixture modeling therefore led to the conclusion that a useful model for TCA images needs to take into account the relationship between nearby pixels when estimating the unknown label image from the noise-corrupted observed image.

For this purpose, two models were introduced: the hidden Markov random field and the coupled hidden Markov model. Both models are capable of taking into account knowledge about the relation of spatially nearby pixels; the HMRF incorporates undirected two-dimensional interactions while the CHMM distinguishes between the two dimensions. The HMRF and the CHMM consist of two levels, the observable random field $\mathcal{Y}$ (with instances Y) and the hidden random field Λ (with instances λ):

$$f(Y) = \sum_{\lambda \in \mathcal{G}^{N \times M}} P(\lambda) f(Y|\lambda).$$

These two random fields are linked by the conditional probability $f(Y|\lambda)$ of the observed image, given the label image, which was specified by independent normal distributions.

In the context of HMRF modeling, the hidden field is assumed to be distributed as an MRF that has the central property

$$P\left(\lambda_i | \lambda_{\mathcal{S}\setminus i}\right) = P(\lambda_i | \lambda_{N(i)}),$$

where i is arbitrarily chosen among the set of pixels $\mathcal{S}$ and $N(i)$ defines the neighborhood of i. Specifying the MRF by an auto-logistic model, it became evident that a more complex MRF needs to be chosen, which incorporates macro-features of images. The FRAME model is such an MRF which accounts for long-range dependence among the observed gray values by including filter responses to the label image $(F_T * \lambda)$ into the potential function:

$$P(\lambda|T) = \frac{1}{Z} e^{\sum\limits_{i \in \mathcal{S}} \phi[(F_T * \lambda)(i)]}.$$

Implementing a halved filter F_T, leads to a coupled Markov model (a so-called FRAME chain model) – the hidden field of the CHMM. Similar to the MRF, it provides a means to model macro-texture; but the CMM is

suited for directed contextual constraints and seems more flexible along the direction of tooth rings. The FRAME, as well as the FRAME chain, both allow for an increased size in neighborhoods while avoiding computational difficulties of a growing number of parameters. Furthermore, they explain their parameters (the filters) more intuitively. The filter family was selected from the Gaborcosine functions, such that the FRAME approach is capable of describing the periodic placement of tooth rings.

Coherent estimation of the parameters of the Gaussian hidden FRAME model and the Gaussian hidden FRAME chain was rendered possible by an EM algorithm, which exploits the mean field approximation of the label distribution $P(\lambda)$. Also, mode field approximation is feasible. This means that time-intensive MCMC methods could thus be avoided. The mean field of the last iteration approximates the reconstruction of TCA images, and the parameter T of the MRF, which in our application determines the wave length of the Gaborcosine filter, expresses the average tooth ring width in TCA images. Estimation of this filter parameter therefore leads directly to an estimate for the number of tooth rings. For one TCA image, the complete estimation procedure requires on average about eleven hours for the Gaussian hidden FRAME model, seven hours for the Gaussian hidden FRAME chain using a one-column filter and about nine hours for the hidden FRAME model using mode field approximation. Confidence intervals of parameters are computable by again applying mean field approximation for the label distribution.

7.2 Summary of Results

TCA images from the Basel series were used to evaluate the abovementioned methods for estimating the number of tooth rings. The age of individuals from the Basel series is known and the TCA images were independently counted twice by two observers from the MPIDR, such that up to four manual observations exist. The manual counts as well as the theoretical number of lines (the known age subtracting the tooth eruption age) are imperfect measures for evaluating a semi-automatic procedure for image analysis. These measures are corrupted by shortcomings due to the TCA method and due to subjectivity, but are nevertheless the only available quantities.

The Gaussian hidden FRAME model applying mode field approximation was tested on a subset of 407 TCA image from the Basel series, which

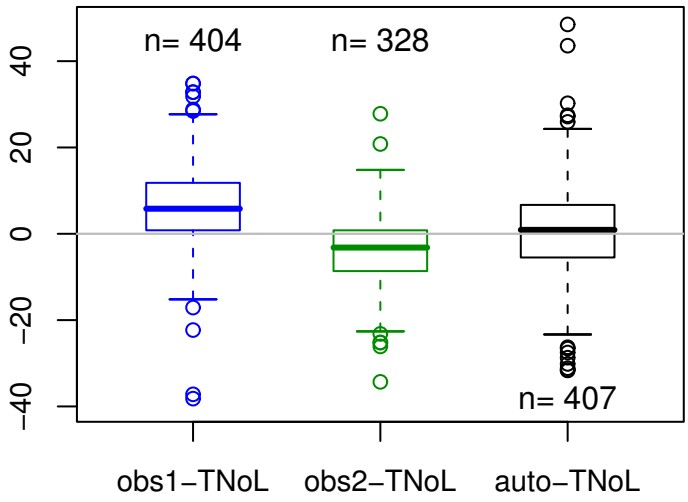


Figure 7.1: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines for the Gaussian hidden FRAME model (right) and for the first (of possibly two) observation(s) from the two observers (left and middle) in 407 selected TCA images from the Basel series.

represent the qualitatively top 40% of the Basel TCA images with 40-fold magnification (from the observers' perspective). The estimate for the number of tooth rings attained from fitting this model is on average closer to the theoretical number of lines than the observers' estimates are. The variance of the difference between the estimated and the theoretical number of lines is slightly larger for the semi-automatic procedure than for the observers. (Compare in Figure 7.1.) The Gaussian hidden FRAME model is thus competitive with the manual estimates at the level of the image. Comparing results on the tooth level, because the manual observations of each tooth are known to be dependent, even shows that the Gaussian hidden FRAME model is superior if the estimates of one tooth are summarized by the mean or the median. The confidence intervals for the estimated number of tooth rings in TCA images have to be interpreted with caution. They measure the goodness of fit of the Gaussian hidden FRAME model to the image. A statement about the quality of the TCA age estimates has to be derived retrospectively from the variance of the estimated, compared to the theoretical, number of lines in a sample.

All other methods introduced in this work where tested against a smaller subset of 49 TCA images, which represent a subset of 'better' quality from the observers' perspective. On this smaller subset, the singular value decomposition and discrete Fourier transform-based procedures from the feature measuring methods were ruled out because of their high variability. The ranges of errors for the middle 50% as well as for most of the TCA images are inappropriate, where errors represent the theoretical minus the estimated number of tooth rings. On this smaller subset, the following models – the Gaussian hidden FRAME model, the Gaussian hidden FRAME chain using a halved and a one-column filter as well as the hidden FRAME model using mode field approximation – all performed very similarly. The Gaussian hidden FRAME chain implementing a halved filter yielded almost the same results as the Gaussian hidden FRAME model and will thus not be discussed in further detail. For the remaining three methods, Figure 7.2 shows comparative box-plots of the difference between the estimated and the theoretical number of lines for the 49 TCA images. The mode field-based procedure is slightly inferior to the other models concerning the average as well as the variance and range of errors. The hidden FRAME model is superior with respect to the mean and the range of errors, and the hidden FRAME chain is superior concerning the range for the middle 50% and the variance of errors. Again, error is defined here as the difference between the theoretical and the estimated number of tooth rings. The computing time of the mode field-based procedure and the hidden FRAME chain model is, respectively, 20% and 40% less than for the hidden FRAME model.

It should be mentioned that the above comparison of alternative methods is based on a small subsample of TCA images that does not represent the quality of the larger sample of 407 TCA images. The 49 TCA images contain images of better quality, from the observers' perspective. Besides, most of the previously mentioned advantages for one or the other of the three methods, are relatively small and probably not significant (as the notches in Figure 7.2 indicate). Conclusions regarding the overall performance of these methods should therefore be treated with caution. The recommended method for TCA image evaluation in this work is the Gaussian hidden FRAME model, both because it produces very good results and because it was evaluated extensively.

It was also proven that the Gaussian hidden FRAME model is not solely applicable to TCA images, but also to other layered structures. Especially the application to cross section of trees produced very good results and demon-

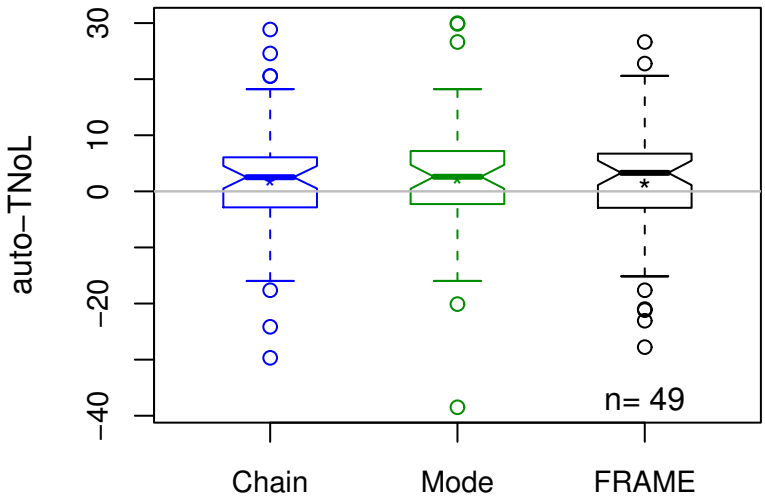


Figure 7.2: Box-and-whiskers plots of the difference between the estimated and the theoretical number of lines for three different methods: the Gaussian hidden FRAME chain using a one-column filter (left), the Gaussian hidden FRAME model using mode field approximation (middle) and the Gaussian hidden FRAME model (right) in 49 selected TCA images from the Basel series. Stars represent the mean for each method.

strated the flexibility and strengths of this model as well as the range of possible applications.

Altogether, it can be concluded that the hidden FRAME model and its relative, the hidden FRAME chain, are flexible tools for TCA image evaluation. Both models achieved comparable, even partially superior, age estimates compared to those of the manual observers and are thus the most promising approaches examined in this work.

The hidden FRAME model is also a powerful tool for similarly layered structures as well as for other large images containing large neighborhood structures. Only the EM algorithm incorporating mean field approximation enables fitting the model to these applications in a reasonable timeframe.

7.3 Outlook

The hidden FRAME model led to a very good evaluation of TCA images. The experiments performed on TCA images therefore provide a basis for further research to enhance TCA age estimation. First of all, the independence of estimates for the number of tooth rings for each TCA image provides the possibility of ascertaining how to combine estimates to one sound age estimate per individual (possibly including the range of estimates). Secondly, further statistical and biological analysis of the independent estimates for TCA images – for example, of one spot with different focus levels, of one slice, one tooth or one individual – could provide valuable insights into the biological mechanisms of ring formation. This would certainly facilitate further research. Thirdly, the semi-automatic evaluation of TCA images presents an objective measure for evaluating and thereby improving the different preparation and evaluation techniques of the TCA method. This could lead to a standardized TCA method. This work therefore provides a valuable tool for further anthropological as well as paleodemographic research on and with the TCA method.

The proposed Gaussian hidden FRAME model was evaluated on a large set of 407 TCA images from the Basel series. Other methods were tested on a subset – of 49 TCA images – that does not represent the whole range of qualities of TCA images. On this smaller subset, the following methods all yielded similar results: the Gaussian hidden FRAME model, the Gaussian hidden FRAME chain using a halved and a one-column filter respectively, as well as the hidden FRAME model using mode field approximation. Future

research should therefore test the latter three methods on a larger subset containing a wider range of TCA images in order to ultimately judge the difference between the estimated number of tooth rings. It is suspected that the more flexible a model is, the poorer it will perform for bad quality TCA images; but this (or, alternatively, the improved estimated number of tooth rings) has to be verified for the hidden FRAME chain models. The hidden FRAME chain, as well as the hidden FRAME model using mode field approximation could be more attractive because they are substantially faster than the hidden FRAME model.

In general, future research should evaluate a semi-automatic counting procedure on a different, even larger sample, than the 407 Basel series TCA images that were tested here. The first reason for this is to reproduce good (or even better) results for a different sample and secondly, to examine how well the algorithm can deal with TCA images that humans were unable to count.

Despite the good estimates for the number of tooth rings in TCA images, the reconstruction of tooth rings leaves something to be desired for the Gaussian hidden FRAME model and the Gaussian hidden FRAME chain models. Various approaches for redesigning the model were suggested, but the ones pursued led to theoretical or algorithmic problems, or to inferior results. The focus of evaluation of the hidden FRAME and hidden FRAME chain models was thus aimed solely at estimating the number of tooth rings; this can be viewed as a first step towards recognizing tooth rings. With knowledge about the number of tooth rings, different methods – for example, object recognition methods such as the snake (Kass et al. [1988]) – could reveal their virtue for detecting tooth rings. These methods were abandoned at first because without knowing the number of objects, the task seemed overwhelming, like a "chicken-and-egg" conundrum. But caution is advised because no reference is known for the position of tooth rings. The human observers marked only a single point of each tooth ring. Marking whole tooth rings is considered too time-consuming and impossible.

The current work assumed a coupled hidden Markov model with directed contextual constraints in the horizontal direction and undirected constraints in the vertical direction. Since no application-driven reason exists for this, future research should also examine the effect of a reverse assumption.

Many textures do not follow Gaussian distributions; also the 'noise' in TCA images (for example, saw cuts and artifacts) is known to be non-normal. Since TCA application does not suggest an alternative distribution, the sim-

ple and convenient assumption of independent, identical and normally distributed gray values, given the label image, was made for all models in the work. This assumption can be relaxed for any distribution of the exponential family within the framework of the EM algorithm and should be tested in future research.

The algorithms suggested in this work were implemented in Matlab and tested on the compute server Hydra of the MPIDR. Neither the chosen programming language nor the computer is the fastest nowadays, but they were convenient choices, respectively, for fast experimental programming and for extended, time-consuming tests. For computing times on up-to-date personal computers, the computing times given in this work could well be divided by two, such that fitting the Gaussian hidden FRAME model to an average TCA image costs only about 5.6 hours. Implementing vital parts of the Matlab program in C++, for example, could additionally speed up the computing time.

Fitting the Gaussian hidden FRAME model to TCA images of 20-fold magnification (instead of 40-fold magnification) may lead to faster results as well, because of a thinner cementum band and the reduced neighborhood size involved with thinner tooth ring sizes. Still, the quality of estimates has to be evaluated critically.

Additional improvements to the proposed models and algorithm are possible. Saddle points are undesirable stationary points that the EM algorithm converges to in part. A criterion to discover and diverge from these estimates was discussed during the work, but not pursued for the first performed experiments. Avoidance of saddle points could improve the estimated number of tooth rings. A refined grid search for the central filter parameter in the course of the EM algorithm was also not pursued. Both suggestions could also improve the quality of confidence intervals.

It is possible that extra improvements could be achieved by adding another parameter to the prior distribution, which estimates the orientation of tooth rings. This is only of academic interest now, but with the continued rapid development in the speed of computers, this may become practicable. Also, a three-dimensional approach incorporating several horizontally neighboring TCA images could lead to improvements in the future.

In the course of this work, the Gaussian hidden FRAME model was extensively tested. Also an application to tree rings was shown to be successful and additional related problems were examined. Different variants of the Gaussian hidden FRAME model that did not lead to success for TCA ap-

plication, still demonstrated the flexibility and capabilities of the hidden FRAME model. Some of these variants are conceivable for different applications.

Appendix A

Derivation of the Observed Information Matrix

As indicated in Chapter 4.3.3, can standard errors be calculated from the inverse Fisher expected information matrix, which can be approximated by the observed information matrix $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)$. For a Gaussian hidden FRAME model with two labels and a common variance, approximations for the elements of this observed information matrix were given on pages 83 to 86.

In this appendix the incomplete-data log-likelihood and its first derivatives will be specified. Some elements of the observed information matrix will follow readily from this as indicated in Chapter 4.3.3, while the other elements will be detailed below. Also derivatives for the filter while specifying it from the family of Gaborcosine functions, as well as simplifications of the observed information matrix when specifying the potential function as the absolute value will be given in this Appendix.

Let us first repeat the incomplete-data log-likelihood of a Gaussian hidden FRAME model

$$\begin{aligned}
\log L\left(\boldsymbol{\theta}, T|Y\right) &= \log P\left(Y|\boldsymbol{\theta}, T\right) = \log \sum_{\lambda \in \mathcal{G}^{N\times M}} P\left(Y, \lambda|\boldsymbol{\theta}, T\right) \\
&\approx \log \sum_{\lambda \in \mathcal{G}^{N\times M}} \prod_{i\in\mathcal{S}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right) P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) \\
&= \sum_{i\in\mathcal{S}} \log \sum_{\lambda_i \in \mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right) P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) \\
&= \sum_{i\in\mathcal{S}} \log \sum_{\lambda_i \in \mathcal{G}} \frac{e^{\sum\limits_{C\in\mathcal{C}, i\in C} \phi[(F_T * \lambda)(C)]}}{\sum\limits_{\lambda_i \in \mathcal{G}} e^{\sum\limits_{C\in\mathcal{C}, i\in C} \phi[(F_T * \lambda)(C)]}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_i - \mu_{\lambda_i})^2}{2\sigma^2}},
\end{aligned}$$

and now differentiate it with respect to its four parameters $(\theta, T) = (\mu_0, \mu_1, \sigma^2, T)$:

$$\frac{\partial \log L(\boldsymbol{\theta}, T|Y)}{\partial \mu_j} = \sum_{i\in\mathcal{S}} \frac{\left(\frac{Y_i - \mu_j}{\sigma^2}\right) P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right)}{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right)}, \text{ for } j = 1, 2$$

$$\frac{\partial \log L(\boldsymbol{\theta}, T|Y)}{\partial \sigma^2} = \sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i \in \mathcal{G}} \left(-\frac{1}{2\sigma^2} + \frac{\left(Y_i - \mu_{\lambda_i}\right)^2}{2\sigma^4}\right) P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right)}{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i|\lambda_i, \boldsymbol{\theta}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right)},$$

and

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta}, T|Y)}{\partial T} =& \sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i \in \mathcal{G}} P(Y_i|\lambda_i, \boldsymbol{\theta}) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right) \left(\sum\limits_{C\in\mathcal{C}, i\in C} \left(\frac{\partial F_T}{\partial T} * \lambda\right)(C) \frac{\partial \phi(\xi)}{\partial \xi}\right)}{\sum\limits_{\lambda_i \in \mathcal{G}} P(Y_i|\lambda_i, \boldsymbol{\theta}) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right)} \\
&- \sum_{i\in\mathcal{S}} \sum_{\lambda_i \in \mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle, T\right) \left(\sum_{C\in\mathcal{C}, i\in C} \left(\frac{\partial F_T}{\partial T} * \lambda\right)(C) \frac{\partial \phi(\xi)}{\partial \xi}\right).
\end{aligned}$$

The elements $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{11}$, $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{22}$, $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{12}$, $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{21}$, $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{13}$, $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{31}$ and $I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{33}$ of the observed information matrix follow easily as indicated in Chapter 4.3.3, while the other elements can be derived as follows:

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{14} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{41} &= -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_0\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx \sum_{i\in\mathcal{S}}\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right) \\
&\quad\cdot\left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}-\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}\right)}{\left(\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2}\right. \\
&\quad\left.-\frac{\left(-\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}+\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right) \\
&= \sum_{i\in\mathcal{S}}\left(\frac{Y_i-\hat{\mu}_0}{\hat{\sigma}^2}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right) \\
&\quad\cdot\left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\left(\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2}-\frac{\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right),
\end{aligned}
$$

$$I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{24} = I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{42} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\mu_1\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \text{ analogously,}$$

$$\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{34} &= I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{43} = -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\partial\sigma^2\partial T}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
&\approx \sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}}\left[\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}-\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right]}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
&\quad + \sum_{i\in\mathcal{S}}\left[\frac{\sum\limits_{\lambda_i\in\mathcal{G}}\left(-\frac{1}{2\hat{\sigma}^2}+\frac{(Y_i-\hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4}\right)P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right. \\
&\quad \left.\cdot\frac{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}-\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}}P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right)P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right]
\end{aligned}$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{34} \approx & \sum_{i \in \mathcal{S}} \frac{\sum\limits_{\lambda_i \in \mathcal{G}} \left[\left(-\frac{1}{2\hat{\sigma}^2} + \frac{(Y_i - \hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4} \right) P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum\limits_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right) \right]}{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)} \\
& - \sum_{i \in \mathcal{S}} \left[\frac{\sum\limits_{\lambda_i \in \mathcal{G}} \left(-\frac{1}{2\hat{\sigma}^2} + \frac{(Y_i - \hat{\mu}_{\lambda_i})^2}{2\hat{\sigma}^4} \right) P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)}{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)} \right. \\
& \left. \cdot \frac{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum\limits_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right)}{\sum\limits_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)} \right],
\end{aligned}
$$

and with $Z(i) = \sum\limits_{\lambda_i \in \mathcal{G}} e^{\sum\limits_{C \in \mathcal{C}, i \in C} \phi[(F_T * \lambda)(C)]}$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{44} = & -\left.\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\left(\partial T\right)^2}\right|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
\approx & -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle\,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\frac{\partial\phi(\xi)}{\partial\xi}}{\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}}\left[\frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\left(\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2}\right. \\
& \left.\cdot\sum_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(-\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}+\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right] \quad \ldots
\end{aligned}
$$

$$
\begin{aligned}
\cdots \quad & + \sum_{i \in \mathcal{S}} \sum_{\lambda_i \in \mathcal{G}} P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum_{C \in \mathcal{C}, i \in C} \left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} + \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \frac{\partial \phi(\xi)}{\partial \xi}}{\partial T} \right) \\
& - \sum_{i \in \mathcal{S}} \sum_{\lambda_i \in \mathcal{G}} P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right) \left(\frac{1}{Z(i)} \frac{\partial Z(i)}{\partial T} \right) \\
& + \sum_{i \in \mathcal{S}} \sum_{\lambda_i \in \mathcal{G}} P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right)^2
\end{aligned}
$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}}, \hat{T}, Y\right)_{44} &= - \left. \frac{\partial^2 \log L\left(\boldsymbol{\theta}, T | Y\right)}{\left(\partial T\right)^2} \right|_{\hat{\boldsymbol{\theta}}, \hat{T}} \\
&\approx - \sum_{i \in \mathcal{S}} \frac{\sum_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right)^2}{\sum_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)} \\
&\quad + \sum_{i \in \mathcal{S}} \frac{\sum_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right) \left(\sum_{C \in \mathcal{C}, i \in C} \left(\frac{\partial F_{\hat{T}}}{\partial T} * \lambda \right)(C) \frac{\partial \phi(\xi)}{\partial \xi} \right) \left(\frac{1}{Z(i)} \frac{\partial Z(i)}{\partial T} \right)}{\sum_{\lambda_i \in \mathcal{G}} P\left(Y_i | \lambda_i, \hat{\boldsymbol{\theta}}\right) P\left(\lambda_i | \langle \lambda_{N(i)} \rangle, \hat{T}\right)} \quad \cdots
\end{aligned}
$$

$$
\begin{aligned}
\cdots \quad & -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}} \frac{\left(\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right)^2}{\left(\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\right)^2} \\
& -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\left(\frac{1}{Z(i)}\frac{\partial Z(i)}{\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right) \\
& -\sum_{i\in\mathcal{S}}\left(\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right)^2 \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
I\left(\hat{\boldsymbol{\theta}},\hat{T},Y\right)_{44} = & -\frac{\partial^2 \log L\left(\boldsymbol{\theta},T|Y\right)}{\left(\partial T\right)^2}\Big|_{\hat{\boldsymbol{\theta}},\hat{T}} \\
\approx & -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& -\sum_{i\in\mathcal{S}} \frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)} \\
& +\sum_{i\in\mathcal{S}}\left(\frac{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum\limits_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)}{\sum\limits_{\lambda_i\in\mathcal{G}} P\left(Y_i|\lambda_i,\hat{\boldsymbol{\theta}}\right) P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)}\right)^2 \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}+\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial^2\phi(\xi)}{\partial\xi\partial T}\right) \\
& -\sum_{i\in\mathcal{S}}\left(\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)\right)^2 \\
& +\sum_{i\in\mathcal{S}}\sum_{\lambda_i\in\mathcal{G}} P\left(\lambda_i|\langle\lambda_{N(i)}\rangle,\hat{T}\right)\left(\sum_{C\in\mathcal{C},i\in C}\left(\frac{\partial F_{\hat{T}}}{\partial T}*\lambda\right)(C)\frac{\partial\phi(\xi)}{\partial\xi}\right)^2.
\end{aligned}
$$

The derivatives $\frac{\partial F_{\hat{T}}}{\partial T}$, $\frac{\partial^2 F_{\hat{T}}}{(\partial T)^2}$ as well as $\frac{\partial \phi(\xi)}{\partial \xi}$ and $\frac{\partial^2 \phi(\xi)}{\partial \xi \partial T}$ depend on the specification of the filter and the potential family respectively. If the filter is chosen from the family of Gaborcosine functions, such that

$$\begin{aligned} F_T(x,y) &= \frac{1}{c_2}(Gcos_{T,0}(x,y) - c_1) \text{ with} \\ Gcos_{T,\alpha}(x,y) &= e^{\frac{-(rx'^2+y'^2)}{2T^2}} \cos\left(\frac{2\pi}{T}x'\right) \text{ and} \end{aligned}$$

$$\sum_{(x,y)} (Gcos_{T,\alpha}(x,y) - c_1) = 0, \quad \sum_{(x,y):Gcos_{T,\alpha}(x,y)-c_1)>0} (Gcos_{T,\alpha}(x,y) - c_1)\frac{1}{c_2} = 1$$

$$x' = x\cos\alpha + y\sin\alpha,\ y' = -x\sin\alpha + y\cos\alpha \text{ and } r = 4,$$

then the first derivative with respect to T is

$$\begin{aligned} \frac{\partial F_T(x,y)}{\partial T} &= \frac{1}{c_2(T)}\left(\frac{\partial Gcos_{T,0}(x,y)}{\partial T} - \frac{\partial c_1(T)}{\partial T}\right) \\ &\quad -\frac{1}{c_2^2(T)}\frac{\partial c_2(T)}{\partial T}(Gcos_{T,0}(x,y) - c_1) \end{aligned}$$

with

$$\frac{\partial Gcos_{T,\alpha}(x,y)}{\partial T} = e^{\frac{-\left(rx'^2+y'^2\right)}{2T^2}}\left(\frac{(rx'^2+y'^2)}{T^3}\cos\left(\frac{2\pi}{T}x'\right) + \sin\left(\frac{2\pi}{T}x'\right)\frac{2\pi}{T^2}x'\right)$$

$$\begin{aligned} \frac{\partial c_1(T)}{\partial T} &= \frac{1}{|F_T|}\sum_{(x,y)}\frac{\partial Gcos_{T,\alpha}(x,y)}{\partial T} - \frac{1}{|F_T|^2}\frac{\partial |F_T|}{\partial T}\sum_{(x,y)} Gcos_{T,\alpha}(x,y) \\ \frac{\partial c_2(T)}{\partial T} &= \sum_{(x,y):Gcos_{T,\alpha}(x,y)-c_1)>0}\left(\frac{\partial Gcos_{T,\alpha}}{\partial T} - \frac{\partial c_1(T)}{\partial T}\right) \end{aligned}$$

where the size of the filter $|F_T| = \left(2 * \left[\frac{4T}{5}\right] + 1\right)^2$ (with [.] denoting the nearest integer) is a step function and therefore not differentiable. But one can approximate this function $|F_T| \approx \frac{(8T+5)^2}{25}$ in order to approximate its derivative by $\frac{\partial |F_T|}{\partial T} \approx \frac{16(8T+5)}{25}$.
The second derivative of F_T is

$$\begin{aligned}\frac{\partial^2 F_T(x,y)}{(\partial T)^2} &= \frac{1}{c_2(T)}\left(\frac{\partial^2 Gcos_{T,0}(x,y)}{(\partial T)^2} - \frac{\partial^2 c_1(T)}{(\partial T)^2}\right) \\ &\quad -\frac{2}{c_2^2(T)}\left(\frac{\partial Gcos_{T,0}(x,y)}{\partial T} - \frac{\partial c_1(T)}{\partial T}\right)\frac{\partial c_2(T)}{\partial T} \\ &\quad +\frac{(Gcos_{T,\alpha}(x,y)-c_1)}{c_2^2(T)}\left(2\frac{\left(\frac{\partial c_2(T)}{\partial T}\right)^2}{c_2(T)} - \frac{\partial^2 c_2(T)}{(\partial T)^2}\right)\end{aligned}$$

with

$$\begin{aligned}\frac{\partial^2 Gcos_{T,\alpha}(x,y)}{(\partial T)^2} &= e^{\frac{-(rx'^2+y'^2)}{2T^2}}\cos\left(\frac{2\pi}{T}x'\right)\left(\left(\frac{(rx'^2+y'^2)}{T^3}\right)^2 - \frac{3(rx'^2+y'^2)}{T^4}\right) \\ &\quad +2e^{\frac{-(rx'^2+y'^2)}{2T^2}}\frac{(rx'^2+y'^2)}{T^3}\sin\left(\frac{2\pi}{T}x'\right)\frac{2\pi}{T^2}x' \\ &\quad +e^{\frac{-(rx'^2+y'^2)}{2T^2}}\left(\cos\left(\frac{2\pi}{T}x'\right)\left(\frac{2\pi}{T^2}x'\right)^2 - \sin\left(\frac{2\pi}{T}x'\right)\frac{4\pi}{T^3}x'\right)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 c_1(T)}{(\partial T)^2} &= \frac{1}{|F_T|}\sum_{(x,y)}\frac{\partial^2 Gcos_{T,\alpha}(x,y)}{(\partial T)^2} - \frac{2}{|F_T|^2}\frac{\partial |F_T|}{\partial T}\sum_{(x,y)}\frac{\partial Gcos_{T,\alpha}(x,y)}{\partial T} \\ &\quad -\frac{1}{|F_T|^2}\frac{\partial^2 |F_T|}{(\partial T)^2}\sum_{(x,y)}Gcos_{T,\alpha}(x,y) + \frac{1}{|F_T|^4}\left(\frac{\partial |F_T|}{\partial T}\sum_{(x,y)}Gcos_{T,\alpha}(x,y)\right)^2 \\ \frac{\partial^2 c_2(T)}{(\partial T)^2} &= \sum_{(x,y):Gcos_{T,\alpha}(x,y)-c_1)>0}\left(\frac{\partial^2 Gcos_{T,\alpha}(x,y)}{(\partial T)^2} - \frac{\partial c_1^2(T)}{(\partial T)^2}\right)\end{aligned}$$

where the second derivative of the filter size is approximated by $\frac{\partial^2 |F_T|}{(\partial T)^2} \approx \frac{128}{25}$.

In Chapter 4.3.4 the potential function ϕ was specified by the absolute value $\phi = |.|$. In order to differentiate this continuous but not everywhere continuously differentiable function, we assume that filter responses are never exactly equal to zeros

$$\xi := (F_T * \lambda)(i) \neq 0.$$

With this assumption the derivatives become

$$\frac{\partial \phi(\xi)}{\partial \xi} = sign(\xi) \text{ and } \frac{\partial^2 \phi(\xi)}{\partial \xi \partial T} = 0,$$

such that the second derivatives of ϕ in the fourth element on the main diagonal of the observed information matrix drop out and therefore further simplify this matrix element (see pages 86 and 171 to 175).

Notation Index

Throughout this work an effort was made to use a unified notation. Because of the various methods from different fields that were used, this was not achieved everywhere. Some notation is used universally throughout this document, like Y for the part of interest of the TCA image. While other symbols may be allocated to different meanings that are bound to the chapter, like α or R. The list below gives an overview of the most important symbols, their exemplary instances and meanings in alphabetical order.

notation	instances	explanation
$\langle . \rangle$	$\langle \lambda_i \rangle$	estimated expected value
$\lceil x \rceil$		smallest integer $\geq x$
$*$	$(F * \lambda)(i)$ or $(F * \lambda)(C)$	convolution (at pixel i or positioned at clique C)
a	a_0	polynomial coefficients
a	a_g	gray values for thresholding
A		(sub) set in Ch. 4
A		acceptance probability in RJ MCMC in Ch. 3
α	$\alpha = 0$	filter orientation in the FRAME model in Ch. 4.3
α	$\alpha = 10$	parameter of the auto-logistic model in Ch. 4.2
α		parameter of a Gamma distribution in Ch. 3
b		probability of birth in RJ MCMC in Ch. 3
B		(sub) set in Ch. 4
be		beta distribution
β		parameter of a Gamma distribution in Ch. 3
β	β_h, β_v	parameter of the auto-logistic model (for horizontal or vertical pair-site cliques) in Ch. 4.2
$\mathcal{C}$	C	set of cliques
$\mathcal{C}_h$		set of all horizontal pair-site cliques
$\mathcal{C}_v$		set of all vertical pair-site cliques

notation	instances	explanation
d		probability of death in RJ MCMC in Ch. 3
D		Dirichlet distribution in Ch. 3
D	d_j	(pseudo) diagonal matrix of eigenvalues in Ch. 2
δ		parameter of a Dirichlet distribution in Ch. 3
E		expectation
ϵ		error
η	$(\xi, \kappa, \alpha, \beta)$	set of hyperparameters in Ch. 3
f		density distribution
F	$F_T, F(x, y)$	filter (with parameter T, one element of a filter)
$\mathbb{F}$	$\mathbb{F}_{k+1}$, $\mathbb{F}_{k+1,l+1}$	Fourier transform of a vector or matrix
$\mathcal{G}$	$\{0, \ldots, G\}, g$	label set (label)
Γ		Gamma distribution
κ		parameter of a normal distribution in Ch. 3
l	l_1 or l_2	number of observations (for label g or $g+1$)
L		likelihood
λ	λ_i or $\lambda_{x,y}$	label image
Λ	λ	random label field
M	m	width of the lattice
μ	μ_g or	mean (at label g) in Ch.s 3 to 5 or
	$\mu(x)$	functional relation in local regression in Ch. 2
n_g		$\sharp\{i \mid \lambda_i = g\}$
N	n	height of the lattice or length of a vector
N		normal distribution
$\mathcal{N}$	$N(i)$	neighborhood system (neighborhood of pixel i)
p		polynomial
ϕ	$\phi = \lvert . \rvert$	potential function
R		Gelman-Ruben convergence stat. in Ch.s 4 and 5
R		range of observations in RJ MCMC in Ch. 3
$\mathcal{S}$	i or (x, y)	lattice of coordinates/pixel
σ^2	σ_g^2 or σ_i^2	variance (at label g or location i)
t	$t_{\max}$	iteration
T	$T = 16$	filter parameter
$\boldsymbol{\theta}$	μ, σ^2	parameters of the cost function
u	u_i	beta distributed random variable in Ch. 3
U	$U(\lambda)$	Gibbs potential
U	u_j	matrix of left singular vectors in Ch. 2

notation	instances	explanation
V	$V_C(\lambda)$	function (on clique C) in Ch. 4
V	v_j	matrix of right singular vectors in Ch. 2
w		weight
	$w_n(x)$	for local regression estimate in Ch. 2
	$P(\lambda_i = g) = w_g$	prob. distr. of labels in Ch. 3
W		weight function in local regression
x		state vector in RJ MCMC in Ch. 3
ξ		parameter of a normal distribution in Ch. 3
Y	Y_i, $Y_{x,y}$ or $Y(x,y)$	observed image (part of TCA image)
$\mathcal{Y}$	Y	observed random field
z		zonal mask

Bibliography

Adler, P. (1967). Die Chronologie der Gebißentwicklung. *Zahn-, Mund- und Kieferheilkunde im Kindesalter*, pages 38–79.

Becker, M. P., Yang, I., and Lange, K. (1997). EM Algorithms Without Missing Data. *Statistical Methods in Medical Research*, 6:38–54.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236.

Besag, J. (1986). On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(3):259–302.

Bilmes, J. A. (1997). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report ICSI-TR-97-021, International Computer Science Institute, Berkeley CA, 94704.

Brand, M., Oliver, N., and Pentland, A. (1997). Coupled hidden Markov models for complex action recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 994–999.

Brémaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, volume 31 of *Texts in applied mathematics*. Springer, New York.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Campbell, F. W. and Robson, J. G. (1968). Application of Fourier Analysis to the Visibility of Gratings. *Journal of Physiology*, 197:551–566.

Celeux, G., Forbes, F., and Peyrard, N. (2003). EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36(1):131–144.

Cross, G. R. and Jain, A. K. (1983). Markov Random Field Texture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):25–39.

Czermak, A., Czermak, A., Ernst, H., and Grupe, G. (2006). A new method for the automated age-at-death evaluation by tooth-cementum annulation (TCA). *Anthropologischer Anzeiger*, 64:25–40.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, B(39):1–38.

Britannica Concise Encyclopedia (2003). Encyclopædia Britannica, Chicago. 'Tooth':1878.

Brockhaus: Die Enzyklopädie (1996). volume 5: CRO – DUC. F. A. Brockhaus, Leipzig, Mannheim, 20 edition. 'Dendrochronology':218.

The New Encyclopedia Britannica: Micropï¿½ia - Ready Reference (1998). volume 4: Delusion – Frenssen. Encyclopædia Britannica, Chicago, 15th edition. 'Dendrochronology':11.

Wikipedia, The Free Encyclopedia (2006). Accessed Online October, 2006. http://en.wikipedia.org/wiki/Main_Page.

Fabig, A., Gampe, J., Hotz, G., and Vaupel, J. W., editors (2007). *Lifespans from Skeletons: How Good Are the Methods? – working title*. Springer. (in progress).

Ghahraman, Z. and Jordan, M. I. (1997). Factorial Hidden Markov Models. *Machine Learning*, 29(2-3):245–273.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, London.

Green, P. J. (2004). http://www2.stats.bris.ac.uk/ peter/nmix. online.

Grissino-Mayer, H. D. (2006). Henri d. grissino-mayer's ultimate tree-ring web pages. http://web.utk.edu/ grissino/. online.

Großkopf, B. (1990). Individualaltersbestimmung mit Hilfe von Zuwachsringen im Zement bodengelagerter menschlicher Zähne. *Zeitschrift für Rechtsmedizin*, 103:351359.

Guyon, X. (1995). *Random Fields on a Network*. Probability and its Applications. Springer, New York.

Hotz, G. (2006). *Das Basel Projekt*, pages 12–13. Jahresbericht 2005. Naturhistorisches Museum Basel - Archive des Lebens, Augustinergasse 2, CH 4001 Basel. http://www.nmb.bs.ch/jahresbericht_05.pdf.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fuer Physik*, 31:253–258.

Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Prentice Hall, NJ.

Jain, A. K. and Farrokhnia, F. (1991). Unsupervised Texture Segmentation Using Gabor Filters. *Pattern Recognition*, 24(12):1167–1186.

Jeng, F.-C. and Woods, J. W. (1987). On the relationship of the Markov mesh to the NSHP Markov chain. *Pattern recognition letters*, 5(4):273–279.

Kagerer, P. and Grupe, G. (2001). Age-at-death diagnosis and determination of life-history parameters by incremental lines in human dental cementum as an identification aid. *Forensic Science International*, 118(1):75–82.

Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*, volume 2: Statistical inference of *Springer Texts in Statistics*. Springer, New York, 2nd edition.

Kass, M., Witken, A., and Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision*, pages 321–331.

Kemkes-Grottenthaler, A. (2002). *Paleodemography: Age distributions from skeletal samples*, chapter Aging through the ages: historical perspectives on ageindicator methods, pages 48–72. Cambridge studies in biological and evolutionary anthropology. Cambridge University Press, Cambridge.

Li, S. Z. (2001). *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. Springer, Tokyo [et al.].

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York [et al.].

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. Monographs on Statistics and Applied Probability ; 70. Chapman & Hall, London.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70(11):297–300.

McLachlan, G. J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.

Miller, C. S., Dove, S. B., and Cottone, J. A. (1988). Failure of Use of Cemental Annulations in Teeth to Determine the Age of Humans. *Journal of Forensic Sciences*, 33(1):137–143.

Nefian, A. V. (2002). Embedded Bayesian Networks for Face Recognition. *IEEE International Conference on Multimedia and Expo*, 2:133–136.

Onoe, M., Tsao, J. W., Yamada, H., Nakamura, H., Kogure, J., Kawamura, H., and Yoshimatsu, M. (1984). Computed tomography for measuring annual rings of a live tree. *Nuclear Instruments and Methods in Physics Research*, 221(1):213–220.

Paleodemography Workshop – Basel Project (2006). Max Planck Institute for Demographic Research and Naturhistorisches Museum Basel, organizers. Oral presentations on Tooth Cementum Annulation. (held in Baden, Switzerland).

Preuß, D. (1975). Twodimensional Facsimile Source Encoding Based on a Markov Model. *Nachrichtentechnische Zeitschrift - NTZ-Aufsätze*, 28:358–363.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Richardson, S. and Green, P. J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society*, B(59):731–792.

Stott, G. G., Sis, R. F., and Levy, B. M. (1982). Cemental Annulation as an Age Criterion in Forensic Dentistry. *Journal of Dental Research*, 61(6):814–817.

Titterington, D. M., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

Trefethen, L. N. and Bau, D. (1997). *Numerical Linear Algebra*. SIAM, Philadelphia.

Webb, A. (1999). *Statistical pattern recognition*. Arnold, London [et al.].

Winkler, G. (1995). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Instruction*, volume 27 of *Applications of mathematics*. Springer, Berlin.

Wittwer-Backofen, U., Gampe, J., and Vaupel, J. W. (2004). Tooth Cementum Annulation for Age Estimation: Results from a Large Known-Age Validation Study. *American Journal of Physical Anthropology*, 123(2):119–129.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

Zhu, S. C. and Mumford, D. B. (1997). Prior Learning and Gibbs Reaction-Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1236–1250.

Zhu, S. C., Wu, Y., and Mumford, D. B. (1998). Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling. *International Journal of Computer Vision Archive*, 27(2):107 – 126.

Zhu, S. C., Wu, Y. N., and Mumford, D. B. (1997). Minimax Entropy Principle and Its Application to Texture Modeling. *Neural Computation*, 9(8):1627–1660.

Zusammenfassung

Zahnzementannulation (TCA[1]) ist der Prozess der geschichteten Bildung von Zahnzement an der Wurzel von Zähnen und kann bei Menschen zur Schätzung des Individualalters genutzt werden. Hierzu werden mikroskopische Aufnahmen von dünnen Querschnitten der menschlichen Zahnwurzel angefertigt (sogenannte TCA-Aufnahmen), in denen die vermutlich jährlich gebildeten Zuwachslinien als dunkle und helle Bänder erscheinen. (Kagerer and Grupe [2001]) Bisher wurden die dunklen Zuwachslinien (Zahnringe) optisch am Mikroskop oder am (digitalen) Bild ausgezählt, um eine Altersschätzung abzuleiten. Ein Algorithmus zur automatischen Auswertung der TCA-Aufnahmen ist von großer Bedeutung, da Forschungsarbeiten, die manuelle Zählungen verwendeten, zu widersprüchlichen Resultaten führten.

In der vorliegenden Arbeit wurde das erste Modell-basierte Verfahren zur Auswertung von TCA-Aufnahmen entwickelt. Hierzu wurden TCA-Aufnahmen als sogenanntes ‚hidden Markov random field' modelliert, welches weitreichende Abhängigkeiten der beobachteten Grauwerte mithilfe eines Filter-basierten Modells einbezieht. Dieses Modell kann menschliches Sehen nachahmen, indem es die periodische Platzierung der Zahnringe nachbildet. Die Anwendung auf TCA-Aufnahmen erzielte konkurrenzfähige, zum Teil sogar bessere, Ergebnisse verglichen mit den manuellen Schätzungen.

Diese Arbeit stellt daher ein wertvolles Hilfsmittel dar für weitere anthropologische sowie paläodemografische Forschungsarbeiten an und mit der TCA-Methode.

Diese Zusammenfassung soll einen kurzen Überblick über den Inhalt sowie die Bedeutung dieser Arbeit geben. Dazu wird zuerst einleitend die Themenstellung erörtert. Anschließend werden die Verfahren zur Auswertung von TCA-Aufnahmen, die im Verlauf dieser Arbeit entwickelt wurden, gemeinsam mit ihren Vor- und Nachteilen zusammengefasst. Dann wird der

[1] TCA ist die Abkürzung für tooth cementum annulation

verwendete Datensatz erläutert und Resultate der Auswertung der TCA-Aufnahmen werden präsentiert. Abschließend werden die Ergebnisse dieser Arbeit diskutiert.

Ausführliche Herleitungen und Erläuterungen befinden sich im englischsprachigen Teil dieser Arbeit.

Einleitung

Diese Arbeit befasst sich mit der Auswertung von TCA-Aufnahmen für paläodemografische Untersuchungen mithilfe eines halbautomatischen Verfahrens. Hierzu sollen einleitend die folgenden Begriffe mit Bezug zum Thema erläutert werden: Paläodemografie, Auswertung, TCA-Aufnahme (und deren Charakteristika) sowie halbautomatisches Verfahren.

Eine Anwendung der TCA-Methode liegt im Bereich der Paläodemografie. Mithilfe von Altersschätzungen an menschlichen Skelettfunden können Mortalitätsprofile (prä-)historischer Populationen rekonstruiert werden. Diese stellen einen wesentlichen Schritt dar, um demografische Prozesse im Laufe der menschlichen Geschichte zu verstehen und repräsentieren ein Forschungsgebiet des Max-Planck-Instituts für demografische Forschung (MPIDR).

Die Auswertung von TCA-Aufnahmen kann in zwei wesentliche Aspekte untergliedert werden: die Erkennung von Zahnringen und die Schätzung ihrer Anzahl. Der Schwerpunkt dieser Arbeit lag in der Schätzung der Anzahl der Zahnringe, da diese Größe entscheidend für die Altersbestimmung und damit für die Rekonstruktion von Mortalitätsprofilen ist.

TCA-Aufnahmen werden üblicherweise mit 20- oder 40-facher Vergrößerung am Durchlichtmikroskop aufgenommen, sind zweidimensional, 1016x1300 Pixel groß und enthalten 2^{16} Graustufen. Zahnringe sind empirisch etwa 1 bis 3 μm breite Strukturen, welche in der digitalen Aufnahme unter 40-facher Vergrößerung und bei einer Pixelgröße von 6.7 μm als ungefähr 5 bis 20 Pixel breite dunkle Bänder erscheinen. Abbildung A.1 zeigt eine typische TCA-Aufnahme, in der 34 Zahnringe erwartet werden. Diese sogenannte theoretische Anzahl der Linien ergibt sich aus dem bekannten Alter (41 Jahre für dieses Individuum), von dem das zahn- und geschlechtsspezifische Durchbruchsalter (sieben Jahre in diesem Fall) subtrahiert wird. Die TCA-Aufnahme in Abbildung A.1 wurde während der gesamten Arbeit als Standardbeispielbild zum Testen von Bildverarbeitungsmethoden benutzt

und dient hier zur Illustration folgender allgemeiner Charakteristika der verwendeten TCA-Aufnahmen:

(1) Zahnringe verlaufen annähernd parallel und hauptsächlich horizontal.

(2) TCA-Aufnahmen enthalten einen hohen Anteil an Rauschen. Informationen zu Zahnringen sind häufig verfälscht oder sogar nicht vorhanden.

(3) TCA-Aufnahmen enthalten weitreichende Abhängigkeiten zwischen den beobachteten Grauwerten. Im Allgemeinen ist der Grauwert eines Pixels in strukturierten Bildern mit hoher Wahrscheinlichkeit ähnlich zu den Grauwerten seiner benachbarten Pixel. Die hohe Reichweite dieser Nachbarschaftsbeziehung zwischen Pixeln ist charakteristisch für TCA-Aufnahmen und entspricht in vertikaler Richtung der Größenordnung der Zahnringbreite.

Ein halbautomatisches Verfahren zur Auswertung von TCA-Aufnahmen beinhaltet die manuelle Markierung des Zahnzementes (siehe Abbildung A.1) und anschließend einen vollautomatisierten Algorithmus, der eine Schätzung für der Anzahl der Zahnringe ausgibt.

Dieses automatisierte Verfahren besteht aus einem EM-Algorithmus, der die Schätzung eines ausgewählten statistischen Modells, des sogenannten ‚Gauß'schen hidden FRAME-Modells', ermöglicht. Die Entwicklung dieses Modells und seines Schätzverfahrens werden im folgenden Abschnitt näher erläutert. Der darauf folgende Abschnitt wird dann die Anwendung des halbautomatischen Verfahrens auf 407 TCA-Aufnahmen illustrieren.

Methoden

Jedes der Kapitel 2 bis 5 dieser Arbeit untersuchte eine Methode – beziehungsweise eine Art von Methoden – auf ihre Anwendbarkeit zur Auswertung von TCA-Aufnahmen. Diese vier Methoden werden im Folgenden zusammen mit ihren Vor- und Nachteilen beschrieben. Zuerst wurden zwei Ansätze verfolgt, die nicht zum Erfolg führten. Trotzdem lieferten sie wertvolle Einsichten zur Entwicklung des entgültigen Verfahrens. Anschließend wurden zwei statistische Modelle untersucht, die methodisch unterschiedliche Ansätze darstellen, jedoch ähnlich spezifiziert werden können und vergleichbar gute Resultate liefern.

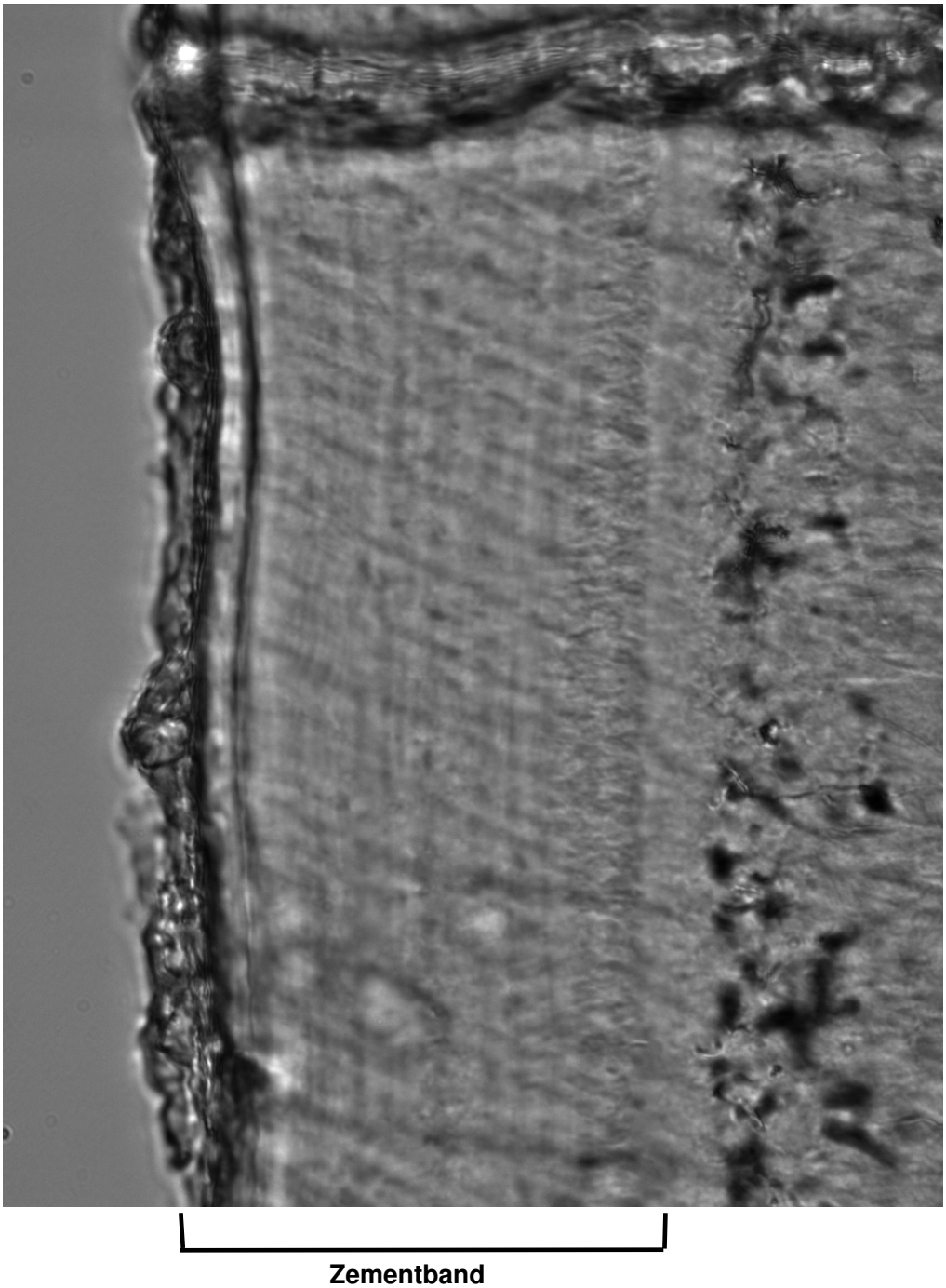


Abbildung A.1: Eine typische TCA-Aufnahme mittlerer Qualität (IS-0000666 aus der TCA-Datenbank des MPIDR).

Als erster Ansatz wurden Standardmethoden aus den drei Bereichen Statistik, Analysis und Algebra auf ihre Anwendbarkeit zur Auswertung von TCA-Aufnahmen untersucht. Diese drei **Verfahren zur Messung von Bildmerkmalen** basieren auf lokaler Regression, Fouriertransformation beziehungsweise Singulärwertzerlegung und konnten die verfälschten oder nichtvorhandenen Informationen in TCA-Aufnahmen nicht wiederherstellen (siehe Punkt (2) der Charakteristika von TCA-Aufnahmen). Trotzdem illustrieren diese Methoden Merkmale von TCA-Aufnahmen und betonen den Unterschied zwischen zwei Aspekten der Auswertung von TCA-Aufnahmen: die Erkennung von Zahnringen und die Schätzung ihrer Anzahl.

Der zweite Ansatz beruhte auf der Annahme von **Mischverteilungen**[2], die voraussetzt, dass Beobachtungen aus verschiedenen Gruppen hervorgegangen sind (McLachlan and Peel [2000]). Die individuelle Gruppenzugehörigkeit (zum Beispiel zum hellen oder dunklen Zahnring) ist jedoch nicht bekannt und wurde mithilfe des EM-Algorithmus (McLachlan and Krishnan [1997])und des RJ MCMC (Richardson and Green [1997]) geschätzt. Beide Algorithmen lieferten Rekonstruktionen der TCA-Aufnahmen, die keine Zahnringe darstellten. Der Grund dafür liegt in der vereinfachenden Annahme, dass Beobachtungen unabhängig sind. Diese Annahme ist für TCA-Aufnahmen jedoch klar verletzt. (Siehe Punkt (3) der Charakteristika von TCA-Aufnahmen.) Dieser Ansatz hat zu der Erkenntnis geführt, dass Modelle für TCA-Aufnahmen die Beziehungen zwischen räumlich nahen Pixeln berücksichtigen sollten, um das unbekannte ‚wahre' Bild aus dem verrauschten beobachteten Bild schätzen zu können.

Eine Möglichkeit, dies zu realisieren, ist das sogenannte ‚**hidden Markov random field**' (HMRF)[3]. Dieses Modell besteht aus zwei Stufen – dem beobachtbaren Zufallsfeld $\mathcal{Y}$ mit Instanzen Y und dem latenten Zufallsfeld Λ mit den Instanzen λ:

$$f(Y) = \sum_{\lambda} P(\lambda) f(Y|\lambda).$$

Für Anwendungen in dieser Arbeit repräsentiert Y eine TCA-Aufnahme und λ deren Rekonstruktion (Labelbild).

Diese beiden Zufallsfelder $\mathcal{Y}$ und Λ sind verknüpft durch die bedingte Wahrscheinlichkeit des beobachteten Bildes unter dem gegebenen Labelbild;

[2] sogenanntes ‚mixture modeling'

[3] ein latentes Markoff-Zufallsfeld, zum Beispiel in Zhang et al. [2001] erklärt

welche durch unabhängige Normalverteilungen spezifiziert wurde. Desweiteren wird angenommen, dass das latente Zufallsfeld wie ein ‚Markov random field' (MRF) verteilt ist, das folgende zentrale Eigenschaft besitzt:

$$P\left(\lambda_i|\lambda_{\mathcal{S}\setminus i}\right) = P(\lambda_i|\lambda_{N(i)}),$$

wobei i aus der Menge der Pixel $\mathcal{S}$ beliebig gewählt ist und $N(i)$ die Nachbarschaft von i definiert. MRFs können mithilfe dieser Nachbarschaft räumliche Abhängigkeiten modellieren.

Zunächst wurde das MRF durch das auto-logistische Modell spezifiziert (Winkler [1995]), welches paarweise Beziehungen zwischen Pixeln berücksichtigt. Dieses einfache MRF modelliert somit nur Mikrotexturen und konnte nicht die parallelen Zahnringe der TCA-Aufnahmen nachbilden. Regulärere Strukturen können durch die Erweiterung der Nachbarschaft N realisiert werden, beispielsweise mithilfe des FRAME[4]-Modells (Zhu et al. [1997]). Dieses Modell kann durch eine Faltung $(F_T * \lambda)$ des Labelbildes innerhalb der Gibbs-Potentialfunktion weitreichende Autokorrelationen erklären:

$$P(\lambda|T) = \frac{1}{Z} e^{\sum\limits_{i\in\mathcal{S}} \phi[(F_T*\lambda)(i)]}.$$

Durch Simulationen dieses MRFs wurden die Eigenschaften des Modells illustriert und gezeigt, dass die periodische Platzierung von Zahnringen mithilfe der Familie der Gaborkosinusfilter beschrieben werden kann.

Eine kohärente Schätzung der Parameter dieses Gauß'schen hidden FRAME-Modells wurde durch einen EM-Algorithmus ermöglicht, der die Mean-Field-Theorie nutzt, um die Verteilung des latenten Zufallsfeldes zu approximieren (Celeux et al. [2003]). Auch eine Mode-Field-Approximation ist möglich. Diese Algorithmen vermeiden rechenintensive Markov-chain-Monte-Carlo-Methoden. Die Mean-Field-Approximation der letzten Iteration des EM-Algorithmus' nähert die Rekonstruktion der TCA-Aufnahme an, und der Parameter T des MRFs, der in unserer Anwendung die Wellenlänge des Gaborkosinusfilters bestimmt, gibt die durchschnittliche Zahnringbreite in TCA-Aufnahmen wieder. Aus der Schätzung dieses Parameters folgt daher direkt ein Schätzwert für die Anzahl der Zahnringe. Konfidenzintervalle der Parameter bieten die Möglichkeit, die Anpassungsgüte des Modells an TCA-Aufnahmen zu beurteilen, und sie konnten ebenfalls mithilfe der Mean-Field-Approximation geschätzt werden.

[4] FRAME ist die Abkürzung für **F**ilters, **R**andom Fields **a**nd **M**aximum **E**ntropy

Desweiteren wurde in dieser Arbeit ein **Coupled Hidden Markov Model** (CHMM)[5] erforscht, mit dem Ziel, gerichtete zweidimensionale Abhängigkeiten zu berücksichtigen. Im Gegensatz dazu stehen die ungerichteten Abhängigkeiten des HMRF, die weniger Flexibilität bieten. Der Ansatz des CHMM beruht auf der Verbindung von benachbarten Markoff-Ketten (Preuß [1975]) und wurde analog zum hidden FRAME-Modell spezifiziert: das beobachtbare Zufallsfeld ist verteilt wie ein Produkt aus Normalverteilungen und das latente Zufallsfeld wie ein FRAME-Modell, wobei der Filter F_T halbiert wird. Dieses Modell wurde hier Gauß'sche hidden FRAME-Kette genannt und ist eine neuartige Methode, Makrotexturen zu beschreiben und zwischen den zwei Dimensionen des Raumes unterscheiden zu können. Durch Simulationen wurde gezeigt, dass auch dieses Modell die periodische Platzierung von Zahnringen beschreiben kann, indem der halbierte Filter aus der Familie der Gaborkosinusfilter gewählt wird. Der selbe EM-Algorithmus wie für das HMRF konnte zum Anpassen des Modells verwendet werden und lieferte ähnliche Ergebnisse. Die Rekonstruktion von TCA-Aufnahmen ist flexibler in der Richtung der Zahnringe für dieses Modell.

Die vorgestellten Modelle und Algorithmen zur Auswertung von TCA-Aufnahmen wurden ausführlich evaluiert. Dazu werden im nächsten Abschnitt die TCA-Aufnahmen der Baselserie vorgestellt, an denen diese Auswertung vorgenommen wurde. Anschließend werden die Resultate präsentiert.

Daten der Baselserie

In der Schweizer Stadt Basel befindet sich der Spitalfriedhof St. Johann, dessen Skelette aus der frühindustriellen Zeit weitgehend altersbekannt sind. Zehn Methoden zur Altersschätzung wurden an 100 ausgewählten Skeletten dieses Friedhofs angewendet. Dabei wurde die Hälfte der Skelette von jedem Beobachter (ohne deren Wissen) zweimal untersucht. (In Hotz [2006] werden Details zu dieser sogenannten Serie Spitalfriedhof St. Johann Basel erklärt.) Auch die TCA-Methode wurde von drei verschiedenen Laboren auf diese Baselserie angewendet. Das Zahnlabor des Max-Planck-Institutes für demografische Forschung fertigte 2120 digitale TCA-Aufnahmen an, die verblindet von zwei unabhängigen Beobachtern ausgewertet wurden. 1280 dieser Aufnahmen wurden jedem Zähler zweimal gegeben, sodass bis zu vier

[5] ein gekoppeltes latentes Markoff-Modell

Zählergebnisse pro Aufnahme existieren. Dieser Aufbau der Studie erlaubt weitreichende Vergleiche der Zählergebnissen eines halbautomatischen Verfahrens sowohl zu der theoretischen Anzahl der Linien als auch zu den manuellen Zählergebnissen.

Aus den TCA-Aufnahmen der Baselserie wurden diejenigen mit 40-facher Vergrößerung ausgewählt, zu denen mindestens zwei (der möglichen vier) manuelle Zählergebnisse existieren. Für diese 407 TCA-Aufnahmen können Zählergebnisse einer halbautomatischen Prozedur gegen zwei unvollkommene Referenzmaße evaluiert werden: die theoretische Anzahl der Linien (TNoL[6]) und die manuellen Zählergebnisse. Die TNoL ist kein ideales Maß für die Anzahl der Zahnringe in TCA-Aufnahmen, da zum Beispiel Anomalien der biologischen Prozesse, die Präparationsbedingungen oder die Mikroskopie zu Störungen führen. Die manuellen Zählergebnisse sind durch ihre Subjektivität gekennzeichnet. Diese beiden Maße, TNoL und manuelle Zählergebnisse, sind jedoch die einzigen zum Vergleich verfügbaren Größen. Zusätzlich wurde eine Teilmenge der 407 TCA-Aufnahmen ausgewählt, bei der die manuellen Zählergebnisse einer Aufnahme maximal drei Jahre voneinander abweichen. Diese Teilmenge umfasst 49 TCA-Aufnahmen und wurde für erste Untersuchungen der vorgestellten Verfahren verwendet. Diese Aufnahmen repräsentieren eine Untermenge besserer Qualität, gemessen an den manuellen Zählungen.

Resultate

An jede der 407 TCA-Aufnahmen der Baselserie wurde das Gauß'sche hidden FRAME-Modell angepasst. Die Ergebnisse sind in der Abbildung A.2 in einem Boxplot zusammengefasst. Die geschätzte Anzahl der Zahnringe ist im Mittel näher an der theoretischen Anzahl der Linien als die manuellen Zählwerte. Die Varianz der Differenz zwischen den geschätzten und der theoretischen Anzahl der Linien ist etwas größer für das Gauß'sche hidden FRAME-Modell als für die manuellen Zähler. Das entwickelte halbautomatische Verfahren ist daher konkurrenzfähig zu den manuellen Zählern auf der Ebene der TCA-Aufnahmen. Vergleicht man die Ergebnisse auf Zahnebene (da den manuellen Zählern die Zugehörigkeit der TCA-Aufnahmen zum Zahn bekannt war), ist das hidden FRAME-Modell sogar etwas überlegen, wenn

[6] TNoL ist die Abkürzung für ‚theoretical number of lines'

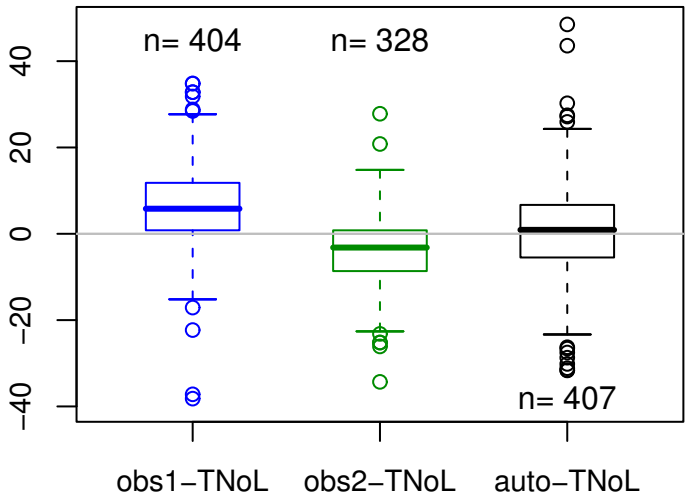


Abbildung A.2: Box-Whiskers-Plots der Differenz zwischen der geschätzten und der theoretischen Anzahl (TNoL) der Linien für das Gauß'sche hidden FRAME-Modell (rechts) und für die erste (von möglicherweise zwei) Zählung(en) der zwei Beobachter (links und Mitte) für 407 ausgewählte TCA-Aufnahmen der Baselserie.

man die Schätzungen eines Zahnes mithilfe des Durchschnittes oder des Medians zusammenfasst. Die berechneten Konfidenzintervalle für die Schätzungen der Anzahl der Zahnringe müssen mit Vorsicht interpretiert werden. Sie messen die Anpassungsgüte des Gauß'sche hidden FRAME-Modells an jede TCA-Aufnahme. Aussagen über die Qualität der TCA-Altersschätzung müssen jedoch retrospektiv aus der Varianz der geschätzten zur theoretischen Anzahl der Linien einer Stichprobe berechnet werden.

Die verbleibenden vorgestellten Verfahren zur Auswertung von TCA-Aufnahmen (sowie das oben diskutierte Verfahren) wurden an der Teilmenge von 49 TCA-Aufnahmen getestet. Die Verfahren basierend auf Singulärwertzerlegungen und Fouriertransformationen wurden aufgrund der hohen Variabilität der Schätzungen von vornherein ausgeschlossen. Außerdem wurden vier weitere Modelle an die TCA-Aufnahmen angepasst und führten zu ähnlichen Resultaten: das Gauß'sche hidden FRAME-Modell, die Gauß'sche hidden FRAME-Kette mit einem halbierten Filter und einem einspaltigen Filter, sowie das Gauß'sche hidden FRAME-Modell basierend auf der Mode-Field-Approximation. Speziell das Gauß'sche hidden FRAME-Modell und

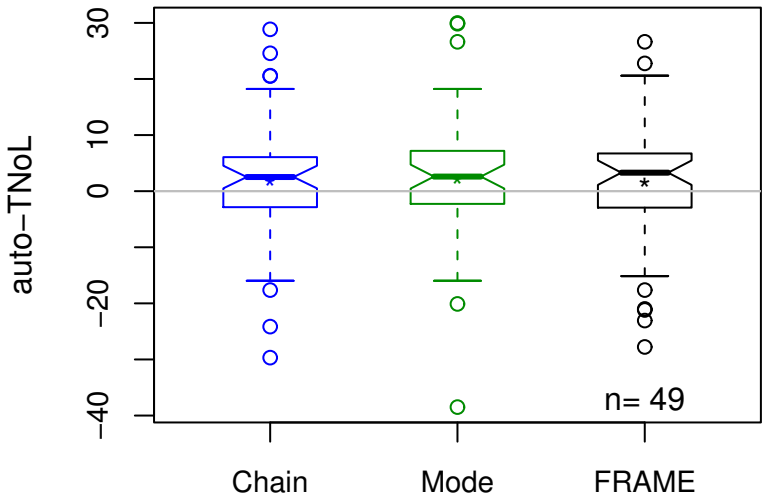


Abbildung A.3: Box-Whiskers-Plots der Differenz zwischen der geschätzten (auto) und der theoretischen Anzahl (TNoL) der Linien für drei verschiedene Verfahren: die Gauß'sche hidden FRAME-Kette mit einem einspaltigen Filter (links), das Gauß'sche hidden FRAME-Modell basierend auf Mode-Field-Approximation (Mitte) und das Gauß'sche hidden FRAME-Modell (rechts) für 49 ausgewählte TCA-Aufnahmen der Baselserie. Die Sternchen markieren den Mittelwert jeder Methode.

die Gauß'sche hidden FRAME-Kette mit einem halbierten Filter lieferten fast identische Ergebnisse, sodass das letztere Modell nachfolgend nicht separat diskutiert wird. Die Ergebnisse der anderen drei Modelle sind zusammenfassend in Abbildung A.3 dargestellt. Im Vergleich zu der theoretischen Anzahl der Zahnringe liefert die geschätzte Anzahl der Zahnringe für den Mode-Field-basierten Algorithmus etwas schlechtere Ergebnisse. Die verbleibenden zwei Methoden liefern für das eine oder das andere aggregierte Maß etwas bessere Resultate.

Diskussion

In dieser Arbeit wurde gezeigt, dass das Gauß'sche hidden FRAME-Modell und die Gauß'sche hidden FRAME-Kette vielversprechende Verfahren zur Auswertung von TCA-Aufnahmen sind. Beide Methoden erzielten vergleich-

bare, teilweise sogar bessere, Schätzwerte für die Anzahl der Zahnringe als die manuellen Zähler. Das hidden FRAME-Modell ist zudem ein leistungsfähiger Ansatz für andere Schichtstrukturen, welches am Beispiel der Jahresringe von Bäumen demonstriert wurde. Andere Anwendungen auf große Bilder, die weitreichende Nachbarschaften erfordern, sind ebenfalls denkbar. Erst der EM-Algorithmus zusammen mit der Mean-Field-Approximation ermöglichte die Anpassung des Modells an solche Anwendungen innerhalb einer angemessenen Zeit.

Trotz der guten Schätzung der Anzahl der Zahnringe lässt die Rekonstruktion der TCA-Aufnahmen noch Verbesserungen offen, sowohl für das Gauß'sche hidden FRAME-Modell als auch für die Gauß'sche hidden FRAME-Kette. Durch das letztgenannte Modell wurde zwar ein Labelbild rekonstruiert, welches flexibler in horizontaler Richtung ist, aber die gleiche, weitgehend konstante Ringbreite im gesamten Labelbild aufweist. Das Gauß'sche hidden FRAME-Modell und der EM-Algorithmus können zusätzlich durch folgende Ansätze umgestalten werden: durch ein nichthomogenes FRAME-Modell, ein pixel- oder labelabhängiges beobachtbares Zufallsfeld, außerdem durch die Mode-Field-Approximation der Verteilung des latenten Zufallsfeldes und durch eine Partitionierung von TCA-Aufnahmen. Diese Ansätze ergaben theoretische oder algorithmische Probleme, beziehungsweise lieferten schlechtere Resultate. Hauptaugenmerk der Evaluierung der Verfahren lag deshalb auf der Schätzung der Anzahl der Zahnringe. Dieses kann als erster Schritt hin zur Erkennung von Zahnringen angesehen werden, da mit der Kenntnis der Anzahl der Ringe andere Methoden, zum Beispiel aus der Objekterkennung, zum Einsatz kommen können. Vorsicht ist jedoch geboten, da für die Position der Zahnringe keine Referenz existiert. Die manuellen Zähler markieren lediglich einen Punkt jedes Zahnringes. Die Markierung ganzer Zahnringe wird als zu aufwändig und als unmöglich eingeschätzt.

Der erforderliche Rechenaufwand für die Anpassung von Modellen mit weitreichenden Nachbarschaften an die großen TCA-Aufnahmen war eine wiederkehrende Thematik dieser Arbeit. Der Rechenaufwand für das Gauß'sche hidden FRAME-Modell beträgt im Durchschnitt ungefähr elf Stunden für eine TCA-Aufnahme, etwa sieben Stunden für die Gauß'sche hidden FRAME-Kette mit einem schmalen Filter und ungefähr neun Stunden für das Gauß'sche hidden FRAME-Modell basierend auf der Mode-Field-Approximation. Diese Rechenzeiten beziehen sich auf den Computeserver

Hydra[7] des MPIDR. Die Gauß'sche hidden FRAME-Kette und das Verfahren basierend auf der Mode-Field-Approximation verkürzten somit die Rechenzeit um etwa 40%, beziehungsweise um 20%, gegenüber des Gauß'schen hidden FRAME-Modells. Weitere Verbesserungen der Geschwindigkeit könnten durch die Verwendung von TCA-Aufnahmen mit 20-facher Vergrößerung erzielt werden, da diese dünnere Zementbänder enthalten und da zusätzlich dünnere Zahnringe zu kleineren Nachbarschaften führen. Auch die Verwendung schnellerer Computer ist denkbar. Dies würde schon heutzutage die Rechenzeit halbieren.

Weitere mögliche algorithmische oder methodische Verbesserungsmöglichkeiten liegen in den folgenden Ansätzen:

- Ein verfeinertes Gitternetz zur Suche nach dem zentralen Filterparameter T könnte insbesondere eine verbesserte Schätzung von Konfidenzintervallen für einige TCA-Aufnahmen ermöglichen.

- Die Verwendung eines zusätzlichen Parameters in der Verteilung des latenten Zufallsfeldes könnte die Orientierung von Zahnringen mitschätzen. Um den Rechenaufwand zu limitieren, wurde bisher angenommen, dass Zahnringe horizontal verlaufen; anderenfalls wurde manuell deren Orientierung geschätzt.

- Ein dreidimensionaler Ansatz könnte gleichzeitig auch benachbarte TCA-Aufnahmen in die Analyse einbeziehen. Dies könnte zusätzliche Informationen in das Schätzverfahren einbringen, ist jedoch bisher zeitlich nicht realisierbar gewesen.

Im Allgemeinen sollten halbautomatische Verfahren zur Auswertung von TCA-Aufnahmen in künftigen Forschungsarbeiten an einer größeren, beziehungsweise einer anderen, Stichprobe getestet werden als die 407 TCA-Aufnahmen der Baselserie. Erstens, um die guten Resultate zu reproduzieren und zweitens, um zu untersuchen, wie sich das Verfahren bei TCA-Aufnahmen verhält, die von den manuellen Zählern abgelehnt wurden. Die Teilmenge der 49 TCA-Aufnahmen, an denen die meisten der vorgestellten Verfahren evaluiert wurden, repräsentiert nicht die volle Bandbreite der Qualität von TCA-Aufnahmen. Da zudem vermutet wird, dass die Gauß'sche hidden FRAME-Kette wegen ihrer Flexibilität dem Gauß'schen hidden FRAME-

[7] Hydra ist ein HP ProLiant DL 580 Computer mit dem Server-Betriebssystem Windows 2000, vier Intel Xeon MP 1.60 GHz Prozessoren und zwei GB Speicher.

Modell für TCA-Aufnahmen von schlechterer Qualität unterlegen ist, sollte diese Methode in Zukunft weitergehend getestet werden.

In der Fachliteratur existiert keine standardisierte TCA-Methode. Verschiedene Labore verwenden unterschiedliche Präparations- und Evaluationstechniken, die vermutlich die Forschungsergebnisse beeinflussen und zu widersprüchlichen Resultaten führen. Mithilfe des vorgeschlagenen halbautomatischen Verfahrens in dieser Arbeit eröffnen sich Forschungsansätze zur Verbesserung der TCA-Altersschätzung. Die objektive Auswertung der TCA-Aufnahmen ermöglicht eine bessere Beurteilung der Präparations- und Evaluationstechniken der TCA-Methode und könnte somit zu deren Vereinheitlichung beitragen. Zweitens liefern die unabhängigen Zählergebnisse für Zahnringe jeder TCA-Aufnahme die Möglichkeit nach einer Methode zu forschen, die Schätzwerte eines Individuums bestmöglich zu kombinieren. Als Drittes könnte die statistische und biologische Analyse von unabhängigen Schätzwerten für TCA-Aufnahmen eines Areals, aufgenommen mit verschiedenen Fokusebenen, oder eines Querschnittes, eines Zahnes oder sogar eines Individuums neue Erkenntnisse zu den biologischen Mechanismen der Zahnringbildung liefern und somit zukünftige Forschung unterstützen.

Die vorliegende Arbeit stellt daher ein wertvolles Hilfsmittel dar für weitere anthropologische sowie paläodemografische Forschungsarbeiten an und mit der TCA-Methode.