Time series reconstruction of environmental proxy records.

Veerle Beelaerts

Members of the jury:

Chairman: Prof. W. Bauwens
Vice-chairman: Prof. H. Terryn
Promotoren: Prof. Rik Pintelon
Prof. Frank Dehairs
Secretary: Prof. J. Schoukens
Members: Prof. J. Vandewalle
Prof. G. Munhoven
Prof. D. Paillard
Prof. M. Elskens

Proefschrift ingediend tot het behalen van de academische graad van doctor in de ingenieurswetenschappen.
I would like to thank Fjo De Ridder and my promotors, Rik Pintelon and Frank Dehairs, for giving me the opportunity to do this research. I would especially like to thank Rik Pintelon who guided me through the mathematics, and helped me gain insight in the problems at hand.

I also would like to thank the members of the jury for the interest and the time they put into this work.

For the financial support, I would like to thank the Fund of Scientific Research (FWO Vlaanderen), the Flemish government (Methusalem Fund METH1) and the Vrije Universiteit Brussel (HOA VI/4), and the Belgian Government (Interuniversity Poles of Attraction IAP VI/4), by the Belgian Federal Science Policy Office (CALMARS II, contract SD/CS/02A) and by the Research Foundation Flanders (contract: G.0642.05) in association with the European Science Foundation (ESF) under the EUROCORES Programme EuroCLIMATE through contract ERAS-CT-2003-980409 (PaleoSalt).

For the real proxy data, I would like to thank: Nele Schmitz from the Laboratory of General Botanica and Nature Management, Vrije Universiteit Brussel, Brussels, Belgium and the Royal Museum for Central Africa, Brussels, Belgium, who provided the stem disks of Rhizophora mucronata and helped me with the vessel density measurements and photos of the Arctica islandica shell; Emma Versteegh from the Department of Analytical and Environmental Chemistry, Vrije Universiteit Brussel, Belgium, who provided the Mg/Ca data of the Mytilus trossulus specimen; Rob Witbaard from the Ecology Department at the Nederlands instituut der Zee (NIOZ) in Texel, the Netherlands, who provided the Arctica islandica specimen, made it possible to photograph this bivalve shell, prepare the samples and measure the δ¹⁸O-records at the NIOZ; Geerard Nieuwland who weighed the last δ¹⁸O-samples and Evaline Van Weerle and Michiel Kienhuis who analyzed the δ¹⁸O -samples for me at the NIOZ.

Finally I would like to thank all my colleges, my friends, my family and my husband, Nabil, for their friendship and support.
# CONTENTS

Acknowledgements ........................................................................................................... 3

1  Introduction .................................................................................................................. 9

  1.1  The need for climate reconstructions ...................................................................... 3

  1.2  Paleo proxy data ....................................................................................................... 3

      1.2.1  Ocean and lake sediments ................................................................................ 4

      1.2.2  ice cores ............................................................................................................ 4

      1.2.3  Tree rings .......................................................................................................... 5

      1.2.4  calcareous marine skeletons ............................................................................. 5

      1.2.5  Speleothem ....................................................................................................... 6

      1.2.6  Historical records .............................................................................................. 7

  1.3  From natural archive to climate data ...................................................................... 7

      1.3.1  Dating of the material ...................................................................................... 7

      1.3.2  Preprocessing the proxy data .......................................................................... 10

      1.3.3  Reconstructing past climatic conditions ............................................................ 17

  1.4  What will be discussed in the thesis ...................................................................... 19

2  Nonparametric elimination of bias averaging-errors in proxy records ....................... 21

  2.1  Introduction .............................................................................................................. 23

  2.2  Linear growth rate .................................................................................................... 24

      2.2.1  Limitations ......................................................................................................... 28

      2.2.2  Unequally spaced proxy data .......................................................................... 30

  2.3  Non-linear growth rate ............................................................................................ 32

  2.4  Simulations .............................................................................................................. 34

      2.4.1  Constant growth rate ....................................................................................... 36

      2.4.2  Non-constant growth rate ................................................................................. 45

  2.5  Applications ............................................................................................................. 52
2.5.1 Mangrove trees ................................................................. 52
2.5.2 Ice core data ........................................................................ 58
2.5.3 Remark ................................................................................ 62
2.6 Conclusion ............................................................................... 62

Appendix 2.A: Mathematical outline ............................................... 64
Appendix 2.B: Least squares solution ............................................... 66

3 Time-series reconstruction from natural archive data with the averaging effect taken into account ......................................................... 69
3.1 Introduction ................................................................................ 71
3.2 The parametric signal model ....................................................... 72
  3.2.1 Harmonic signal model .......................................................... 73
  3.2.2 Measurement model ............................................................... 77
3.3 The identification algorithm ........................................................ 78
  3.3.1 Numerical issues .................................................................... 80
  3.3.2 Generation of starting values .................................................. 82
  3.3.3 Model Selection ...................................................................... 82
  3.3.4 Cramér-Rao lower bound ....................................................... 82
  3.3.5 Comparison of the method with ($\delta = 0$) and without ($\delta \neq 0$) an averaging effect ........................................................ 83
3.4 Simulation results ...................................................................... 85
  3.4.1 Robustness to the noise on the proxy data .............................. 85
  3.4.2 Robustness to the noise on the distance axis ............................ 89
3.5 Application: Mangrove trees ..................................................... 91
  3.5.1 Equally Spaced Data .............................................................. 91
  3.5.2 Equally spaced with missing data .......................................... 94
  3.5.3 Unequally spaced ................................................................. 96
  3.5.4 Comparison .......................................................................... 98
3.6 Conclusion

Appendix 3.A: the jacobian matrix

Appendix 3.B: the uncertainty on the estimated parameters

4 Periodic time series modeling of environmental proxy records with guaranteed positive growth rate estimation

4.1 Introduction

4.2 The parametric signal model

4.2.1 Harmonic signal model

4.2.2 Measurement model

4.3 Problem statement

4.3.1 Gaussian maximum likelihood cost function

4.3.2 Nonlinear optimization with inequality constraints

4.4 2-Step optimization with linear constraint

4.4.1 The 2-step approach

4.4.2 Description of the linear equality constraints

4.4.3 Implementation of the linear equality constraints

4.4.4 Generation of starting values

4.4.5 Model selection

4.5 Comparison of the time inversion constraint method with the positive growth rate estimation method

4.5.1 $ncT < b$

4.5.2 $b < ncT$

4.6 Simulation

4.6.1 Polynomial basis functions

4.6.2 Splines as basis functions

4.7 Case study

4.7.1 Polynomials: Rhizophora mucronata
## CONTENTS

4.7.2  Splines: Mytilus trossulus .......................................................... 127

4.8  Conclusion ...................................................................................... 132

5  Case Study: *Arctica Islandica* ............................................................. 135

5.1  Introduction .................................................................................... 137

5.2  Materials and method ...................................................................... 138

5.2.1  Age determination ...................................................................... 138

5.2.2  Stable isotope analysis ................................................................. 138

5.2.3  Correction for averaging ............................................................... 139

5.2.4  Temperature reconstructions ......................................................... 140

5.3  Results and discussion .................................................................... 141

5.3.1  Age determination ...................................................................... 141

5.3.2  Stable isotopes ............................................................................ 141

5.3.3  Correction for averaging ............................................................... 143

5.3.4  Temperature reconstruction .......................................................... 149

5.4  Conclusion ...................................................................................... 151

Appendix 5.A.: The raw $\delta^{18}$O-values ........................................... 152

6  Conclusions and future perspectives .................................................. 155

6.1  Conclusions .................................................................................... 157

6.2  Future perspectives ......................................................................... 158

References ........................................................................................... 163

Publications ........................................................................................... 177

Journal papers ........................................................................................ 178

Published and accepted ........................................................................ 178

In review and in preparation .................................................................. 178

Conference papers ................................................................................ 179

Abstracts .............................................................................................. 179
1 INTRODUCTION
1.1 THE NEED FOR CLIMATE RECONSTRUCTIONS

Past climate was proven to be an evolving system. Fifty million years ago our planet was a tropical oasis, with rainforests from Canada in the North to Patagonia in the South. However, 2.5 million years ago our planet slipped into a cooler climate with the beginning of the ice age. Since then, our climate is characterized by long cold periods, called glacial periods, with big ice sheets, covering most of Europe and Northern America alternated by shorter warm periods, similar as today, called inter-glacial periods. Changes in climate are induced by internal forcings such as ocean and ocean-atmosphere coupled dynamics and tectonic movements, and external forcings such as solar radiation, volcanic activity, and, since the beginning of the industrial revolution, human activity.

Human society is reliant on knowledge about our climate. Consequently, it is important to know how the climate will evolve in the future. Therefore it is important to understand what controls Earth’s climate. This requires, an investigation into our climate’s past. Because instrumental records are very short, we have to rely on paleo-climate data to reconstruct the climate of the past.

1.2 PALEO PROXY DATA

Paleo climate data can be derived from natural climate recorders, such as ocean and lake sediments (Herbert, 1994; Huang et al., 2000; Jones et al., 2009; Weedon, 1989), ice cores (Augustin et al., 2004; Jouzel et al., 1997), tree rings (Briffa et al., 1990; Graumlich, 1993; Verheyden et al., 2005), marine organisms with a calcareous skeleton, amongst which echinoderms (Borremans et al., 2009), sclerosponges (Lazareth et al., 2000), corals (Kuhnert et al., 2002; Marshall and McCulloch, 2002; Sinclair et al., 1998; Wei et al., 2000) and bivalves (Fleitmann et al., 2004; Gillikin et al., 2005; Lazareth et al., 2003; Vander Putten et al., 2000), speleothems (Fleitmann et al., 2004; Wang et al., 2008) and historical records (Lüterbacher et al., 2002; Pfister, 1980; Wang and Zhang, 1988). Variables such as, for instance, trace elements and stable isotopes (TEI’s) incorporated in an accreting substrate may reflect and archive the temporal variation of environmental conditions. In that sense these variables can be useful as proxies of the conditions prevailing at the time the archive was constituted. These proxies allow for the reconstruction of the history of climate past. In the following sections of this chapter
CHAPTER 1: INTRODUCTION

1.2.1 OCEAN AND LAKE SEDIMENTS

Tons of sediment accumulates every year on ocean and lake floors. As such, they record our planet’s history. Scientists drill cores of sediment (Fig. 1.1) from the basin floors to reconstruct this history. These sediment cores are long term climate recorders with a low temporal resolution: Core V28-238, a Pacific deep-sea core, for example is only 16 m long and contains up to 870,000 years of environmental information (Shackleton and Opdyke, 1973). A vast number of sites have been cored all over the ocean floor. The proxies present in these marine sediment cores can be e.g.: (i) fossils of marine animals which carry information as temperature and chemical composition of the ocean (de Vernal et al., 2005; Koc et al., 1993); (ii) volcanic glass providing information on volcanic activity (McGuire et al., 1997); (iii) sediments from shelves and continent which inform on ocean currents, dust storms, submarine earthquakes, ancient coastlines (Lozano-Garcia et al., 1993). Sediments also record the magnetic history archiving changes in the magnetic orientation of the poles (Verosub and Roberts, 1995).

1.2.2 ICE CORES

Ice in mountain glaciers and ice sheets accumulated from snow fall over hundreds of thousands of years. Consequently, these layers of ice are a rich source of environmental information. Scientists drill ice cores to investigate the environmental conditions of the past (Fig. 1.2). Ice cores are the only environmental archives which can provide climate information for the highest latitudes and altitudes. They have a higher resolution than deep-sea sediment cores: the ice core recovered from Dome C in Antarctica, for example, contains 740,000 years of
climate information preserved in 3200 m of ice (Andersen et al., 2004; Augustin et al., 2004). Nevertheless, only a limited amount of ice cores are available. The variety of proxies in ice cores is greater than in other environmental archives. Proxies such as stable isotopes, radioisotopes, dust composition, snow accumulation rate, air bubbles, and volcanic ash or sulfate can provide information on air temperature, atmospheric circulation variations, precipitation amount, atmospheric composition, solar activity, and records of volcanic activity (Andersen et al., 2004; Etheridge et al., 1996; Jouzel et al., 1997; Langway et al., 1995).

1.2.3 TREE RINGS

The study which dates and analyses tree rings (Fig. 1.3) is called dendrochronology. Trees generally produce one tree-ring a year and record as such environmental information. This results in very long and detailed records. Trees can contain annually-resolved proxy information and can grow thousands of years. In Dalarna, Sweden, for example, a tree, 9,950 years old, was identified (Umea University, 2008). Proxies such as tree ring width, the number of cells, cell size, cell-wall thickness, wood density and stable isotope composition tell us something about ecological changes, geomorphologic variations, biotic effects of increasing CO₂, isotopic variations, circulation patterns, and climate change (Fletcher, 1975; Luckman et al., 1997; Poussart et al., 2004; Verheyden et al., 2005).

1.2.4 CALCAREOUS MARINE SKELETONS

Many marine organisms form a calcium carbonate skeleton in oxygen isotopic equilibrium with the surrounding water (Epstein et al., 1953; Mook and Vogel, 1968) (Fig. 1.4). These carbonates generally show periodic growth lines. In analogy with dendrochronology, the discipline which dates and analyses these growth lines is called sclerochronology. Different organisms have different characteristics. Corals, for example, have more or less constant growth rates, live hundreds of years and can be
sampled with a bimonthly to annual resolution. Bivalves, on the other hand, have varying growth rates, the short lived bivalves only live for a few years, while some long lived species can live up to 400 years (Schöne et al., 2005c). Depending on the growth rate, sampling along a growth axis achieves daily to yearly resolution. Proxies such as increment width, stable isotopes, and trace elements provide information about temperature, salinity, ocean currents, etc. (Hudson et al., 1976; Wanamaker et al., 2007; Zinke et al., 2004).

**1.2.5 SPELEOTHEM**

Speleothems (cave formations) are recorders of terrestrial climate (Fig. 1.5). They are high resolution records with well known chronologies. The temporal resolution and length of the different cave records vary. The Shanbao Cave record from China, for example, covers 240,000 years with a temporal resolution of about 100 year (Wang et al., 2008), while the records from Kahf Dufore in Oman are shorter but have a temporal resolution of one year (Fleitmann et al., 2004). Proxies encountered in speleothems can be: thickness of annual growth bands which is a proxy of the surface precipitation (Fleitmann et al., 2004); oxygen isotope ratio which is a proxy for the variation of cave temperature and rainfall properties; carbon isotope ratio which is a proxy for changes in overlying plant vegetation and vegetation density (Mickler et al., 2004); growth intervals which is a proxy to determine wetter or drier (Mugrove et al., 2001), warmer or cooler climate intervals; and trace elements

![Fig. 1.4: Cross section of an *Artica islandica* shell](image1)

![Fig. 1.5: Young speleothem from south Australia.](image2)
which are proxies for the rainfall (Ayalon et al., 1999).

1.2.6 HISTORICAL RECORDS

Historical records can contain an enormous amount of climate information (Fig. 1.6). A variety of data can be used such as agricultural statistics, records of snow fall, freezing and thawing, phonological data and sea-ice data. By means of this information climate variations of the last 500 years can be determined (Lüterbacher et al., 2002).

1.3 FROM NATURAL ARCHIVE TO CLIMATE DATA

The reconstruction of climate from environmental records requests:

i. Dating of the material
ii. Preprocessing the proxy data
iii. Reconstructing past climatic conditions

In this section, these steps will be discussed further.

1.3.1 DATING OF THE MATERIAL

Environmental archives are found in numerous places. Since it is not always known when these archives were produced, one of the most important steps in climate reconstruction from proxy records is dating the environmental archive. Many absolute dating techniques exist, the most important techniques for dating environmental archives are radiometric techniques, layering techniques and cross-dating techniques.
1.3.1.1 RADIOMETRIC DATING

In radiometric techniques the age of the environmental archive is determined from the decay of their radio-active elements. Radioactive decay follows an exponential curve with the exponent equal to the decay constant x time evolved. At half-life half of the parent atoms decayed to daughter atoms (Fig. 1.7).


The most commonly used radiometric techniques for dating environmental archives are based on the decay of ratio-isotopes such as $^{14}$C, $^{226}$Ra, $^{234}$U, $^{230}$Th and $^{210}$Pb, $^{137}$Cs, ... The half-life times, $t_{1/2}$ (i.e. the time needed for half of the original amount of the radioactive element to have decayed) of some of these radioactive isotopes are shown in Table 1.1.

<table>
<thead>
<tr>
<th>$t_{1/2}$ (years)</th>
<th>$^{14}$C</th>
<th>$^{234}$U</th>
<th>$^{230}$Th</th>
<th>$^{210}$Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>5730</td>
<td>245000</td>
<td>75000</td>
<td>22.3</td>
<td></td>
</tr>
</tbody>
</table>

Radiocarbon dating ($^{14}$C; half life = 5370 yr) is mainly used to date carbonaceous biological produced materials such as trees (Cain and Suess, 1976) and calcareous skeletons (Bard et al., 1990; Brunnberg and Possnert, 1992), but also Holocene sediments can be dated with the radiometric methods via the fossil calcareous skeletons (mainly foraminifera) they contain. Uranium-thorium dating ($^{230}$Th half-life = 75,000 y) can also be used to date massive calcareous skeletons (corals; sclerosponges) as young as 3 and as old as 600000 y (Andrews and Tedesco, 1992; Edwards et al., 2003; Edwards et al., 1988), and has been applied on speleothems (Dorale et al., 1998; Lauritzen and...
Lundberg, 1999). The $^{210}\text{Pb}$ method is the main method to date recent sediment cores (Dorale et al., 1992; Muir et al., 1996), for older cores the $^{230}\text{Th}$ method is the main method (Rosenheim et al., 2007).

1.3.1.2 LAYERING TECHNIQUES

Layering techniques are based on the seasonal variation in the accretion rate of environmental archives. Amongst these layering techniques are tree-ring counts (Fig. 1.8a) (Niklasson, 2002), growth increment counts in calcareous skeletons (Veinott and Cornett, 1996), counting of sedimentary layers (Hardy et al., 1996) and ice layers (Fig. 1.8b) (Rasmussen et al., 2006).

Fig. 1.8: Layering techniques. (a) Tree rings in the mangrove tree *Rizophora mucronata*. (b) Layers in an ice core (Blogspot [http://randomblogn.com/2008_03_01_archive.htm](http://randomblogn.com/2008_03_01_archive.htm) Accessed 10/08/2010.)

1.3.1.3 CROSS DATING

Cross-dating techniques are based on matching patterns and characteristics amongst different proxy records (Fig. 1.9). In trees and bivalves for example, ring and increment widths or ring and increment characteristics are matched to obtain long term overlapping records (Cook et al., 1995; Marchitto et al., 2000).

Fig. 1.9: Cross-dating. Archipelagos, Institute of Marine Conservation (http://web.utk.edu/~grissino/Site/images/xdate.gif 08/07/2010).
1.3.2 PREPROCESSING THE PROXY DATA

Since models are matched to proxies, proxies need to be as precise and as accurate as possible; otherwise the models, as well as the conclusions drawn, will be biased. In other words, the proxy data need to be treated in order to eliminate possible errors.

1.3.2.1 GROWTH ANOMALY

Proxies are measured on a distance scale, mostly along an axis of maximum accretion. Since different environmental archives grow at different speeds and since the growth rate may vary during its life-time, it is impossible to compare different proxy records. Unless a common axis can be constructed (for example a time axis) proxies from different records cannot be compared. Moreover, if we would calculate the annual mean of an environmental proxy, without correcting for this growth rate anomaly, the result would be biased towards the faster growing part (Wilkinson and Ivany, 2002).

The transformation from a distance to a time axis requires information about the accretion rate. Because the accretion rate is generally unknown, we have to rely on indirect information or assumptions. Periodicity is a common assumption, because many proxy records exhibit a seasonal cycle.

A number of methods which describe a periodic time series are described in literature. Following de Brauwere et al. (2009) these methods can be subdivided into two classes: the mapping methods and the periodic signal model methods. The mapping methods generate a time axis by assuming similarity between the proxy record and a reference function. Since this mostly concerns periodic time series, this reference function is often sinusoidal. The periodic signal model methods generate a time base by fitting a signal model onto the data. The periodic signal model methods use a much less stringent assumption than the mapping methods. In contrast to the mapping methods, the signal model methods do not a priori fix the reference function in advance. Instead, a model is proposed for this reference function with parameters which still need to be optimized.

We will discuss three mapping methods: the anchor point method (Paillard et al., 1996), the correlation maximization method (Lisiecki and Lisiecki, 2002; Yu and Ding, 1998) and Martinson et al.’s method (Martinson et al., 1982a; Martinson et al., 1987; Martinson et al., 1982b). Three periodic signal model methods will be discussed as well: a time
domain method (Wilkinson and Ivany, 2002), a frequency domain method (De Ridder et al., 2004), and a parametric time base distortion approach (de Brauwere et al., 2008).

In the anchor point method certain observations, called anchor points, are dated. The intermediate dates can then be estimated by linear interpolation, thus assuming a linear accretion rate (Fig. 1.10). This method is the most frequently used method (Charles et al., 1997; Felis et al., 2000). However, in the presence of stochastic noise it does not perform very well, though it does perform reasonably well in the presence of modeling errors. The method has 3 major disadvantages: the precision is limited because the real maxima and minima sometimes fall between two subsequent samples (discretization errors, see Section 1.3.2.4), the assumption of a linear growth rate is unrealistic, the number and position of anchor points are arbitrarily chosen by the user and the result is thus dependent on this choice (de Brauwere et al., 2009).

Fig. 1.10: Visualization of the Anchor point method. Horizontally: the proxy as a function of distance; vertically; the reference function as a function of time. The date of some observations is known and between these anchor points a linear accretion is assumed (dotted line). From (de Brauwere et al., 2009).

The correlation maximization (Lisiecki and Lisiecki, 2002; Yu and Ding, 1998) also works with anchor points. Though, the dates assigned to these anchor points are optimized so that the correlation between the proxy record and the reference function is maximized. In Yu and Ding (1998) the number of anchor points is equal to the number of observations. Consequently, this method is very sensitive to stochastic measurement noise. In Lisiecki and Lisiecki (2002), on the other hand, the number of anchor points are limited; this reduces the influence to the stochastic noise greatly. In the presence of modeling errors the method by Yu and Ding does not perform well, while the method by Lisiecki and Lisiecki gives reasonable results (de Brauwere et al., 2009). In Fig. 1.11 the method described in Lisiecki and Lisiecki is outlined by means of a simple example.
Fig. 1.11: The method of Lisiecki and Lisiecki is outlined by means of a simple example. The points in series A need to be matched to a subset of point in series B so as to minimize the sum of the square of their differences. (a) Series A with $n = 4$ points will be matched to a subset of the $m = 5$ points in series B. (b) Each point in A is subtracted from each point in B, these differences are squared and displayed in a table, where each column represents a point in A and each row represents a column in B. The double bordered boxes represent the alignment with the smallest sum. (c) Series A is aligned to series B according to the solution shown in the table. From Lisiecki and Lisiecki (2002).

Martinson et al.’s method (Martinson et al., 1982a; Martinson et al., 1987; Martinson et al., 1982b) models the distance-time relationship as a linear combination of some simple functions (e.g. sines or splines). As such, the accretion rate is no longer constant between two subsequent anchor points. By correlating the proxy record with a known reference function (e.g. a harmonic signal with known amplitudes, frequency and phase), the distance-time relationship, and accordingly the time axis, can be recovered (Fig. 1.12). The method is insensitive to stochastic noise and performs well in the presence of modeling errors. Nevertheless, the performance of the method depends on the choice of the reference function and the complexity of the model for the distance-time relationship (de Brauwere et al., 2009).
The time domain method by Wilkinson and Ivany (2002) is based on the assumption that the proxy record on a time axis is sinusoidal. Best-fit sine waves (amplitude, frequency and phase) are determined for subsequent subsets of the data, called windows. These windows advance by one sample each step until the entire dataset has been processed (Fig. 1.13). In De Ridder et al. (2007) an improvement to the method was suggested by taking not only the frequency, but also the phase into account. The performance of the method depends highly on the choice of the window width. The method performs well in the presence of noise, but can be sensitive to modeling errors (de Brauwere et al., 2009).

The frequency domain method (De Ridder et al., 2004) is based on a phase demodulation technique. The assumption here is that the signal is periodic on a time base and bandwidth limited. The signal on a distance scale is distorted and, consequently, non-harmonic. When a non-harmonic signal is transformed to the
frequency domain via a Discrete Fourier transform (DFT), errors, called leakage, are introduced in the spectrum. Due to leakage, peaks in the frequency spectrum are broadened. These leakage errors contain information about the distortion present in the distance series and thus about the distance-time relationship. The information can be extracted by windowing the spectrum around the first harmonic (Fig. 1.14). The distortion can then be calculated by translating the resulting window to the origin of the spectrum, applying the Inverse Discrete Fourier Transform (IDFT) and taking its phase. This method performs well in the presence of stochastic noise and in the presence of modeling errors. Nevertheless, the sensitivity to the noise and modeling errors is mainly dependent on the choice of the window size.

Fig. 1.14: Visualization of windowing the spectrum around the first harmonic. The window is represented by the full line, which covers the first harmonic. Altered from De Ridder et al. (2004).

The parametric time base distortion method (de Brauwere et al., 2008) is based on a parametric time base model as in Martinson et al.’s method. The difference here is that the reference function is not fixed, but modeled as a linear combination of sines and cosines with unknown amplitudes, fundamental frequency and number of harmonics. The parameterization of the signal model reduces the sensitivity to modeling errors. Moreover, to ensure robustness to over- and under-modeling, an automated model selection procedure is used to select the optimal number of parameters (de Brauwere et al., 2009).
1.3.2.2 AVERAGING EFFECTS

Environmental archives are generally solid substrates. Whether sampling is done by drilling a hole and collecting the drilled material, or counting a proxy per unit surface, the sample always has a certain volume, or area (the volume of the drill hole or the counting window). As a consequence, a given sample will provide the mean value of the proxy over a corresponding volume or area. When the width of the sample covers a considerable part of the variation, the signal will be averaged and the natural signal variations will be systematically underestimated. This is exemplified in a simulation (Fig. 1.15) where the averaging effects can cause a severe bias in the measurement. Without a correction for these effects, the amplitude shows an apparent decline with time, while in reality this is not the case.

Fig. 1.15: The effect of averaging for a descending growth rate. (a) Growth rate. (b) The true signal (full black line) and the measured signal (dashed gray line) as a function of time. Horizontal lines: sampled time windows.

To our knowledge, averaging was first described by Harrington (1989) and Krantz, Jones and Williams (1989). A detailed investigation into averaging problems was performed in Goodwin et al. (2004; 2003). Observations of averaging errors are abundant in literature (e.g. Kennedy et al., 2001; Kingston et al., 2008). To overcome averaging effects, researchers select for the smallest possible samples, thereby increasing spatial and temporal resolution. However, working close to the detection limit will lower the signal-to-noise ratio, and thus the accuracy of the measurement. Therefore, sampling strategies are adapted to the growth rate, choosing small samples for low growth rates and larger samples for fast growth rates (Fells et al., 2004; Schöne et al., 2005c). To our knowledge, corrections for averaging are not reported in literature.
Averaging can also occur during incorporation instead of during sampling. An example of this type of averaging occurs during amelogenesis, which is the formation of enamel on teeth, in mammal teeth. Different layers of enamel are deposited as a mineral poor matrix, that gradually accumulates minerals over an extended period of time (Hiller et al., 1975; Robinson et al., 1978). Consequently, the minerals in the teeth will be younger than the surrounding matrix. A solution to this problem presented by Passey and Cerling (2002). Here, the accumulation of minerals is modeled and with inverse modeling techniques the unaveraged signal is recovered.

A second example of averaging during incorporation is air transport in the firn ice. Firn ice is an intermediate state between snow and ice in which gas transport is still possible. Consequently, the gases in the ice can be a thousand year younger than the surrounding ice (Barnola et al., 1991; Schwander et al., 1993). A solution to this is provided by inverse modeling techniques, in which the transport of gas in the firn ice is modeled and the gas concentrations at a certain time are reconstructed (Rommelaere et al., 1997).

1.3.2.3 ONTOGENETIC EFFECTS

For environmental archives constructed by living organisms proxy incorporation may also be dependent on the physiology of the organism (Poorter, 1999; Rossi et al., 2004; Sosdian et al., 2006). Physiological effects are often translated as representing an ontogenetic trend in the proxy record. Since this effect is generally largest at a younger stage of the organism’s life, this younger stage of the environmental archive is often not measured. Nevertheless, ontogenetic trends are eliminated by detrending the proxy record (Butler et al., 2009; Cook and Kairiukstis, 1990).

1.3.2.4 DISCRETIZATION ERRORS

Some discretization errors will be present in the measured proxy record. This means that the real maxima and minima might not be measured, because only a limited number of samples are taken. This is shown in Fig. 1.16, where the ‘true’ proxy signal (back line) and the measured proxy signal (gray line) are plotted as a function of distance.

Seasonality is generally estimated from differences between maximum and minimum values in the data set (Wilkinson and Ivany, 2002). Therefore, when the seasonal range of an environmental record is calculated, discretization often leads to under-representations. Moreover, when one is interested in proxy values at a certain time
instance, it is not recommendable to perform a simple linear interpolation. In Fig. 1.16
the gray line here is linearly interpolated in between two samples, it becomes
immediately obvious that the gray line is not a good representation of the black signal in
between two samples. To eliminate these errors, there are two types of solutions.

The first solution is called band-limited interpolation. The amplitude and the phase of a
signal can be reconstructed when more than 2 samples per period are measured
(Shannon theorem). This reconstruction is performed by evaluating the proxy record at
a denser uniform grid in the frequency domain. The disadvantage of this method is that
it is very sensitive to stochastic noise. When the signal-to-noise ratio is low, or the
number of samples is low, the method does not perform well.

![Figure 1.16: Discrete sampling. The continuous proxy signal (full black line), and the
discretized proxy signal (full gray line) on a distance scale.](image)

The second approach is fitting a signal model onto the proxy record. Methods which fit a
harmonic signal model on proxy data are described in Section 1.3.2.1. These methods
are much less sensitive to the noise.

1.3.3 RECONSTRUCTING PAST CLIMATIC CONDITIONS

Proxies are indirect measures of the climate; therefore, a calibration of the different
proxies with instrumental data is necessary.

Laboratory experiments have been designed to study growth and proxy incorporation of
environmental archives under controlled conditions. As such, a relationship between the environmental archive studied (bivalves; echinoderms; corals; ...) and the corresponding environmental parameters can be established (see e.g. Freitas et al., 2008; Wanamaker et al., 2007). These experiments can be repeated in the field (natural conditions) in which the environmental conditions are closely monitored during the experiment. For slowly accumulating archives, this is obviously not feasible. In this case, relationships between environmental parameters and proxies can be obtained by comparing the proxy records with long records of instrumental data (Douglass, 1920; Schöne et al., 2004).

The relationship between proxies and the environmental parameters are not always straightforward. One proxy is often influenced by more than one environmental parameter. Moreover, one environmental parameter can influence more than one proxy. Relationships between environmental parameters and proxies are translated into models. Consequently, a multi-proxy model which combines all these proxies and environmental parameters should yield the best results. A simplification of such a multi-proxy model is shown in the following equation:

\[
\begin{align*}
P_1 &= a_1 E_1 + b_1 E_2 + \cdots + c_1 E_N \\
P_2 &= a_2 E_1 + b_2 E_2 + \cdots + c_2 E_N \\
\quad \quad \vdots \\
P_N &= a_N E_1 + b_N E_2 + \cdots + c_N E_N \\
\begin{cases}
E_1 = \alpha_1 P_1 + \beta_1 P_2 + \cdots + \gamma_1 P_N \\
E_2 = \alpha_2 P_1 + \beta_2 P_2 + \cdots + \gamma_2 P_N \\
\quad \quad \vdots \\
E_N = \alpha_N P_1 + \beta_N P_2 + \cdots + \gamma_N P_N
\end{cases}
\]

(1.1)

where \( P \) are the proxies and \( E \) are the unknown environmental parameters. If a set of proxies is influenced by a set of environmental parameters, the environmental parameters are found by solving the equation for the environmental parameters. When a model is built with more than one proxy, more robust and accurate reconstructions will be obtained (Bauwens et al., 2010a). Nevertheless, in literature, the use of multi-proxy models is not frequently mentioned. Sometimes, a second proxy is used to confirm the observations made with the first proxy (Klein et al., 1996; Schöne et al., 2006) or to eliminate the influence of a secondary environmental parameter (Bice et al., 2006; Gentry et al., 2008). To our knowledge, except for the work of Freitas et al. (2006) and Bauwens et al. (2010b), real multi-proxy models are nonexistent.
Most relationships between the environmental parameter and proxy are non-linear (Cloern et al., 1995). Nevertheless, the most commonly used technique to establish a relationship between an environmental parameter and a possible proxy is linear regression (e.g. Klein et al., 1996; Wanamaker et al., 2008). Nonlinear techniques have been less used, because they are much more difficult to develop. In some papers an exponential or an inverse exponential relationship is assumed (e.g. Freitas et al., 2005). The most used nonlinear techniques to reconstruct environmental conditions are Artificial Neural Networks (Guiot et al., 2005; Juillet-Leclerc et al., 2007; Woodhouse, 1999). However other techniques such as Support Vector Machines and Manifold Learning can also be used (Bauwens et al., 2010b).

Once the models are tuned, they can be used to reconstruct the climate. Climate reconstruction can be validated with independent evidence. In historical records, for example, natural disasters are often well documented and consequently can be used to validate the results (Linderholm and Molin, 2005; Mann et al., 1998; Stott et al., 2001). Independent proxy records, for example terrestrial is marine records, can also be used for validation purposes (Baales et al., 2002; Jones et al., 2009).

1.4 WHAT WILL BE DISCUSSED IN THE THESIS

This thesis focuses on in the preprocessing step of proxy data. Four problems are addressed: the growth anomaly, the averaging effect caused by sampling, the discretization errors, and the handling of non-uniform samples. Throughout this thesis, averaging caused by sampling will be referred to as averaging. Furthermore, the averaging effects are assumed to be in one direction only, that is, the growth direction of the axis on which the measurements are performed (Fig. 1.17). Furthermore, the sample window is assumed to be rectangular with two negligible dimensions (height and depth) and one non-negligible dimension (width). Typical examples for this type of samples are the quadrants in dendrochronology (Verheyden et al., 2005) or ice cores or sediment cores (Augustin et al., 2004). However, the conclusions of this study also hold for a wide variety of other sample shapes. In principle this method can be applied to any proxy record, measured with a sample size that is large relative to the variation that needs to be reconstructed. Furthermore, the proxy signal on a time scale is assumed to be harmonic. This assumption is justifiable, as environmental archives often exhibit a
seasonal cycle.

The outline of this thesis is as follows. A non-parametric method for the correction of averaging effects is established in Chapter 2. A parametric method which simultaneously estimates the non-linearities in the distance-time relationship (also called distance base distortions) and which takes into account the averaging effects is introduced in Chapter 3. An optimization technique with linear constraints on the signal model parameters that guarantees a positive growth rate, is described in Chapter 4. A case study with the bivalve *Artica islandica* is presented in Chapter 5. Finally, conclusions are formulated in Chapter 6.

![Averaging in one direction](image)

Fig. 1.17:: Averaging in one direction. $\delta$ is the width of the sample, i.e. the distance over which the signal is averaged. $\Delta$ is the distance between two subsequent samples. $n$ is the sample position, $n \in \{1, \ldots, N\}$, which is related to the distance $x$ as follows: $x(n) = \Delta n$. $t(n)$ is the unknown time variable.
2 NONPARAMETRIC ELIMINATION OF BIAS AVERAGING-ERRORS IN PROXY RECORDS


2.1 INTRODUCTION

Environmental archives are generally solid substrates. Whether sampling is done by drilling a hole or counting a proxy per unit surface, the sample always has a certain volume (the volume of the drill hole or the count space). As a consequence, a sample will give the mean value of the proxy over a corresponding volume. To our best knowledge, the volume of the sample has always been neglected till now. However, this is only allowed when the width of the sample is small with respect to the variation that needs to be reconstructed. Researchers choose intuitively the smallest possible samples; however, working near the detection limit will lower the signal-to-noise ratio, and thus the accuracy of the measurement. On the other hand, when the width of the sample covers a considerable part of the variation, the signal will be averaged and the variations will systematically be underestimated. This is exemplified in a simulation, shown in Fig. 2.1. As can be seen in this illustration, averaging effects can cause a severe bias in the measurement. Without a correction for such effects, the amplitude shows an apparent decline with time, while in reality this is not the case. It is clear that the problem is an underestimation of the amplitude of the signal due to averaging.

![Figure 2.1: The effect of averaging for a descending growth rate.](image)

Because models are matched on proxies, proxies need to be as precise and as accurate as possible; otherwise the models will be biased, as well as the conclusions drawn. In this chapter a non-parametric calibration method will be presented which reduces averaging errors in the measurements. In a parametric method the data is described by a number of parameters; this number of parameters is smaller than the number of data points. In a
non-parametric method the data is not described by parameters.

The outline of this chapter is as follows. In Section 2.2 a simulation with a linear growth rate will give a better insight into the problem and will lead to a correction for the averaging effect. This correction will then be tested on simulated data with a non-linear growth rate in Section 2.3, and finally its effectiveness will be illustrated on real mangrove and ice core data in Section 2.5.

## 2.2 LINEAR GROWTH RATE

In this section averaging is outlined by means of a simple example in which the simulated proxy signal is equidistantly sampled, is harmonic and has an integer number of periods within the measurement window on a time scale. Fig 2.2 shows a transformation from a distance scale (horizontal) to a time scale (vertical), where the growth rate is constant (diagonal). For a constant growth rate, the transformation of equidistant data on a distance scale results in equidistant data on a time scale. This means that a harmonic signal on a time scale will still be harmonic on a distance scale. Consequently, averaging on the time and distance scale is identical.

![Fig. 2.2: Conceptual graph showing the transformation over a linear growth rate from a distance scale to a time scale. The measured signal (dashed line), true signal (full line) and samples (horizontal/vertical lines).](image)

In Fig. 2.3a the effects of averaging in the time domain can be seen, where the true signal (full black line) and the measured signal (dashed gray line) are plotted as a function of
time. The width of the sample was 0.2 year and the simulated signal, \( y(t) \), had the following form:

\[
y(t) = \sin(2\pi f_0 t) + \frac{1}{2} \sin(4\pi f_0 t),
\]  

where \( t \) denotes the time and \( f_0 = 1.067 \text{ year}^{-1} \), the fundamental frequency.

When the experimental conditions are such as in Fig. 2.3a, the underlying ideal signal will be averaged over the width of the sample. This means that the sample value will be an average of the true signal from one end of the sample to the other. In the frequency spectrum the consequences of averaging can be seen in terms of frequency components and amplitude.

In Fig. 2.3b, averaging is shown in the frequency domain, where the true signal (full black line) and the measured signal (dashed gray line) are plotted as a function of frequency. It is clear that averaging causes amplitude reduction. This small attenuation of the lower frequency components and large attenuation of the higher frequency components is typical for a low pass filter. It will be shown that the true signal has been filtered by a low pass filter. The goal is to characterize the filter and, then recover the ideal signal by compensating for the filter effect on the measured signal.
Averaging is assumed to be in one direction only, i.e. the direction of the growth axis and the sample shape is assumed to be square. This is exemplified in Fig. 2.4.

\[ 1 \quad 2 \quad 3 \quad \ldots \quad n \]

\[ \delta \quad \Delta \]

Fig. 2.4: Averaging in one direction. \( \delta \) is the width of the sample, i.e. the distance over which the signal is averaged. \( \Delta \) is the distance between two subsequent samples. \( n \) is the sample position, \( \in \{1, \ldots, N\} \), which is related to the distance \( x \) as follows: \( x(n) = \Delta n \). \( t(n) \) is the unknown time variable.

When the sample shape is not rectangular but, for example, cylindrical (e.g. drill hole) averaging will be overestimate in the corners. This is exemplified in Fig. 2.5. The dark arrows represent averaging for a square sample where the sample value is equal to the average over the length of the sample as an equal amount of powder is taken at any point in distance. The light arrows represent the averaging in a cylindrical sample, where the sample value is equal to a weighted average over the length of the sample, as more powder is taken in the middle of the sample than at the borders.

\[ \text{Distance} \]

Fig. 2.5: Averaging in one direction with a rectangular sample (dark gray arrows) and a round sample (light gray arrows).

In Appendix 2.A averaging in one direction for a square sample shape is outlined mathematically, the relation between the true signal and the measured signal is established, and the filter is characterized. It is shown that averaging (the amplitude attenuation) is equivalent to multiplying the true frequency spectrum by a sinc-function (see Appendix 2.A and Fig. 2.5). The sinc function is only dependent on the width of the sample:
\[ \mathfrak{H}[\bar{y}(x_n)] = \operatorname{sinc}(\frac{\delta \omega}{2}) \mathfrak{H}[y(x_n)] \]  

(2.2)

where \( \bar{y}(x_n) \) is the discrete measured signal as a function of the sampling position \( x_n \), \( \omega \) is the angular spatial frequency, \( y(x_n) \) is the discrete ideal signal at position \( x_n \), \( \mathfrak{H}[y(x_n)](\omega) \) is the Fourier transform, and \( \delta \) is the width of the sample.

In Fig. 2.6 the absolute value of a sinc function (full black line) is shown. From Equation 2.2 it follows that the frequencies where the filter values are zero correspond to the frequencies where the sampling size equals one period for the first zero (arrow \( 1/\delta \)), two periods for the second zero (arrow \( 2/\delta \)), etc. Note that to reconstruct the continuous signal \( y(x) \) uniquely from its samples \( y(x_n) \), one should obey Shannon’s theorem which states that more than two samples per period of the highest frequency must be taken.

![Fig. 2.6: The absolute value of a sinc-function (full black line), the first zero corresponds to \( 1/\delta \) (arrow), the second zero to \( 2/\delta \) (arrow), etc., with \( \delta \) the sample width.](image)

After characterizing the filter, a correction (compensation) is effected to obtain the signal in the frequency domain as follows: simply multiply the measured signal by the inverse of the sinc-function.

To assess whether it is important to account for averaging errors, a comparison must be made between the size of the averaging errors and the measurement uncertainties. In Fig. 2.7 the errors due to averaging (full black line) are shown as a function of the ratio...
of the sample width versus the variation that needs to be reconstructed. The larger the sample widths in comparison with the variation, the larger the averaging-errors and, hence, the more important the correction method will be.

![Graph](image)

**Fig. 2.7**: The errors due to averaging in percentage as a function of the ratio of the sample width and the period of the signal of interest.

The smaller the sample width is, on the other hand, the larger the measurement noise will be. When a sample has a certain width which leads to a concentration that is equal to the detection limit of an analytical machine, the measurement error for that sample will be 100%. The measurement error will decrease with an increasing sample width, \( \delta \), as \( \sim \frac{1}{\delta} \). The reason for this is that when we have ten times more sample material, the signal is ten times better. This is shown in Fig. 2.7 (full gray line), where it is assumed that when the sample width/period of the sample is equal to 0.1, the concentration is equal to the detection limit.

A trade-off between the averaging errors and the measurement noise can be made in order to find the optimal sample width. At the cross section of both curves, the total error (noise and averaging error) is minimal and therefore the optimal sample width is located here. In the example in Fig. 2.7 the optimal sample width/period of the sample is located at 0.4.

---

### 2.2.1 LIMITATIONS

Here, a distinction has to be made between overlapping and non-overlapping sampling.
When sampling is non-overlapping the sampling rate can only be as fast as the inverse of the sample width. In terms of the Shannon theorem, this means that the highest frequency that can be reconstructed, the Nyquist frequency, will be one-half of this sampling frequency. When sampling is overlapping there are more samples. More samples mean a higher sampling frequency, and therefore a higher Nyquist frequency. When the Nyquist frequency exceeds the frequency that coincides with the first zero of the sinc-function (physically this is the frequency at which the sample width overlaps one period exactly), as in Fig. 2.8a, another limitation will take over, as will be explained next. In reality all measured proxies are disturbed by noise. When a noisy frequency spectrum is corrected, the signal-to-noise-ratio of the signal stays the same, since we multiply the noise as well as the noiseless signal by the same filter value. Due to the shape of the filter (Fig. 2.8b) white noise will be transformed into colored noise.

Fig. 2.8: Noiseless signal (a) noisy signal (b). The spectrum of the true signal (*full black line + markers*), the spectrum of the measured signal (*dashed dark gray*) and the filter (*sinc-function*) (*dash-dotted light gray*) as a function of frequency.

Fig. 2.8b is similar to Fig. 2.8a, but with signal-noise ratio of 5. Note that the component at the highest frequency (second peak) can hardly be distinguished from the noise (Fig. 2.8b). In terms of the correction this means that these high frequency noise components, when multiplied by the inverse of the sinc-function, will be blown out of proportion. As a consequence, frequency components which are higher than a certain threshold value cannot be reconstructed and, therefore, will be multiplied by one. This threshold value is the frequency that coincides with the first zero of the sinc-function; physically this is the frequency at which the sample width overlaps one period exactly. Intuitively, this restriction can be illustrated as follows: if the sample width is, for example, equal to one
day, night-day variations (12 h) cannot be reconstructed.

2.2.2 UNEQUALLY SPACED PROXY DATA

Most data records are unequally spaced (e.g. adaptive sampling Goodwin et al., 2003), or equally spaced but with missing data (e.g. missing data due to failed measurements). The method, as is described now, is based on the Fourier transform of the proxy signal on a distance axis. A classical Fourier transform experiences some technical problems when applied to unequally spaced data, amongst the fact that the projection is no longer orthogonal.

The Regressive Discrete Fourier Transform (RDFT), first introduced by Arruda (1992), offers a solution to this problem. In the next paragraph the Fourier series model will be generalized, following Arruda (1992).

A noiseless, band-limited periodic signal \( y(x) \) can exactly be described by

\[
y(x) = \sum_{k=-H}^{H} Y_k e^{j2\pi kx/\tau}, \tag{2.3}
\]

with \( Y_k \) the Fourier coefficients \( (Y_{-k} = \bar{Y}_k \text{ complex conjugate of } Y_k) \), \( H \) the number of harmonics and \( \tau \) the fundamental signal period. Evaluating (2.3) at a non-uniform distance grid \( x_n, n = 1, 2, ..., N \), gives

\[
y(x_n) = \sum_{k=-H}^{H} Y_k e^{j2\pi kx_n/\tau}, \tag{2.4}
\]

where the measurement window \( T = |x_N - x_1| \) is related to the signal period \( \tau \) as

\[
\tau = aT, \tag{2.5}
\]
with $\alpha \geq 1$.

Note that the non-uniform sampling in (2.4) should be dense enough such that no aliasing occurs (Martin, 1997).

In case of noisy measurements (2.4) is replaced by

$$y(x_n) = \sum_{k=-H}^{H} Y_k e^{j2\pi k x / \tau} + e(x_n), \quad (2.6)$$

with $e(x_n)$ the measurement noise.

In the case of equally spaced or unequally spaced harmonic data with an integer number of periods in the measurement window, $\alpha$ in (2.5) is equal to 1 and the Fourier coefficients $Y_k$ can be obtained from (2.6) via a linear least-squares (LLS) optimization (see Appendix 2.B) (Arruda, 1992).

In theory, signals that are periodic but do not have an integer number of periods within the measurement window and non-periodic signals can be expanded in an infinite Fourier series with period $\tau$. In practice, an infinite number of terms would be impossible to realize, and therefore, the classical Discrete Fourier transform introduces leakage errors into the frequency spectrum, when a non-periodic signal is transformed. In Fig. 2.9 the frequency spectrum of a periodic signal (a) and a non-periodic signal (b) is shown. Leakage present in Fig. 2.9b, results in an amplitude reduction and broadening of the frequency peaks. The larger the deviation from an integer number of periods is, the stronger the effect.

The RDFT described in Section 2.2.2 offers a solution to this problem. Indeed, the introduction of a tunable $\alpha$-value in (2.5) reduces the number of terms due to the ringing in the Fourier series expansion significantly. If $\alpha$ in (2.5) is known, then the Fourier coefficients $Y_k$ can be obtained from (2.6) via a linear least squares (LLS) optimization. If $\alpha$ in (2.5) is unknown, then we solve (2.6) out of a fixed number of harmonics for different $\alpha$-values, for example, $\alpha$ varies between 1 and 2 in steps of 0.01. The optimal $\alpha$-value and $Y_k$-values correspond to the minimal $\sum_n e^2(x_n)$.
In Vanherzeele et al. (2008) an optimized regressive Fourier transform (ORFT) is described in which $\alpha$ is optimized together with the parameters $Y_k$. In this work an ORFT was not applied because the solutions of the non-parametric method will be used as starting values for the parametric approach.

The relationship in (2.2), assuming that the $\alpha$ values for $\mathcal{F}_R[\tilde{y}(x_n)](\omega)$ and $\mathcal{F}_R[y(x_n)]$ are equal, is generalized to:

$$\mathcal{F}_R[\tilde{y}(x_n)](\omega) = \text{sinc}\left(\frac{\delta\omega}{2}\right) \mathcal{F}_R[y(x_n)], \quad (2.7)$$

in which the classical Fourier transform, $\mathcal{F}$, is replaced by the RDFT, $\mathcal{F}_R$.

### 2.3 NON-LINEAR GROWTH RATE

In reality growth rates or accretion rates are almost never constant. On the contrary, they often depend on environmental parameters, which are in turn not constant. In this section a proxy signal is simulated with a descending growth rate, mimicking an ontogenetic trend.

In Fig. 2.10 the transformation from a distance scale (horizontal) to a time scale (vertical) is shown. Equidistant data on a distance scale will no longer be equidistant data on a time scale after transformation with a varying growth rate. This means that the signal will be deformed during transformation. The signal on a distance scale looks like a
compressed sine and during transformation the signal is stretched to a normal sine, resulting in a periodic signal on the time scale. Not only is the signal stretched, but also are the samples. As a consequence, the averaging effect is larger near the end of the signal on both the time and distance scales. When the effects of averaging are not taken into account, it seems as if the measured signal is fading, due to some physiological or environmental factors, and thus may be misinterpreted.

Fig. 2.10: Conceptual graph showing the transformation over a non-linear growth rate. The measured signal (dashed line), the true signal (full line) and the samples (horizontal lines) as a function of distance and time.

In contrast to the case with the linear growth rate, averaging on a time and a distance scale is not identical in the case of a non-linear growth rate. Therefore, a choice between a correction on distance data or time data has to be made. The filter of interest is a sinc-function, which only depends on the sample width. The measured data have to be multiplied by the inverse of this filter to reconstruct the true value. This worked for harmonic data with uniform samples, but what about non-harmonic data (distance data) or non-uniform samples (time data)? Distance-samples (samples on a distance scale) are of identical width, and as the filter only depends on the width of the sample, one single filter can correct all the samples. The difficulty with distance data is that it is non-harmonic, even when the time data is harmonic. The difficulty with time-samples (samples on a time scale) is that they are of non-identical widths; this implies that each time-sample corresponds to a different filter in the frequency domain. In this case a time-varying filter is needed, which corrects every time-sample in the frequency domain.
with its own filter. Because non-harmonic data are easier to work with than a time-varying filter, the correction on the distance-samples will be used.

The data on a distance scale are non-periodic. Here, the RDFT can also be applied to non-periodic signals. A non-periodic signal can be expanded in a Fourier series with an infinite number of terms, and the introduction of a tunable $\alpha$-value in (2.6) reduces the number of terms due to the ringing in the Fourier series expansion significantly.

## 2.4 SIMULATIONS

In this section the method will be tested on synthetic data. First a periodic signal was created on a time scale:

$$y(t) = \sin(2\pi f t) \tag{2.8}$$

with $y(t)$ the true signal (see Fig. 2.11a, (full black line)), $f = 1.0811$ the frequency, $t$ the time. This signal was transformed onto a distance scale by means of a simulated growth rate which is either constant or non-constant. This signal represents the true signal; this is the signal as it was incorporated by the environmental archive.

From the true signal on a distance scale $y(x)$, the measured signal on a distance scale $\bar{y}(x_n)$ was calculated by averaging the true signal from one end of the sample to the other end of the sample:

$$\bar{y}(x_n) = \frac{1}{M} \sum_{k=0}^{M} y(x_n - \frac{\delta}{2} + \frac{k}{M} \delta) \tag{2.9}$$

with $M$ the number of $x$-values in the interval $[x_n - \frac{\delta}{2}, x_n + \frac{\delta}{2}]$ and $\delta$ the sample width. The measured signal on a time scale was calculated by transforming the measured signal on a distance scale by means of the growth rate. The experiment was designed such that the measured signal on a time scale is periodic and has an integer amount of
periods within the measurement window.

The method for the correction of averaging was applied on the measured data on a distance scale. The corrected signal on a time scale was calculated from the corrected signal on distance scale by means of the growth rate. After correction these signal should be more similar to the true signals.

The following cases were tested: (i) an equally spaced periodic measured signal with an integer amount of periods within the measurement window (Section 2.4.1.1), (ii) an equally spaced periodic measured signal with an integer amount of periods within the measurement window with missing data (Section 2.4.1.2), (iii) an unequally spaced periodic measured signal with an integer amount of periods within the measurement window (Section 2.4.1.3), (iv) an equally spaced non-periodic measured signal (Section 2.4.2.1), (v) an equally spaced non-periodic measured signal with missing data (Section 2.4.2.2), and (vi) an unequally spaced non-periodic measured signal (Section 2.4.2.3).

The method for the correction of averaging was tested on the noiseless measured signals. In case of missing data or unequally spaced data a second approach was tested in which a band-limited interpolation was performed prior to the correction of averaging. The Shannon-theorem states that the highest frequency that can be reconstructed, the Nyquist frequency, is half the sampling frequency. This means that the amplitudes and the phase of a signal can be reconstructed when more than 2 samples per period are measured. This reconstruction implies evaluating the simulated measured distance data at a denser uniform distance grid \( x_l = l\Delta \) (\( \Delta \) is the distance between two subsequent samples)

\[
\hat{y}(x_l) = \sum_{k=-H}^{H} \hat{Y}_k e^{j2\pi k x_l / \tau}
\]  

(2.10)

with \( \hat{Y}_k \) the linear least squares estimate of the Fourier coefficients. Equation (2.10) is a band-limited interpolation of the non-uniformly measured averaged signal. Nevertheless, a band-limited interpolation might not always give better results. The band-limited interpolation relies on the frequency spectrum (Fourier coefficients) and when the signal-to-noise ratio is low, the frequency lines from the noise can be mistaken
for frequency lines from the signal.

Therefore, the first approach (the correction for averaging) and the second approach (the band-limited interpolation plus the correction for averaging) were tested in a Monte Carlo simulation on noisy measured data. For each case, 500 measurement sets were generated. Therefore, zero mean white Gaussian noise was added to the noiseless measured signals, such that the signal-to-noise-ratio (SNR) is 11.6. This SNR was used to test the noise sensitivity of the different methods in de Brauwere et al. (2008), it is a realistic noise level for proxy records. The noise was added to the measured signal and, therefore, it represents the measurement noise and not environmental noise. These 500 noisy measurement sets were corrected for averaging. The difference between the true signal and the average of the 500 corrected noisy measurements were plotted together with 95% confidence levels for each case, to gain more insight into the robustness to the noise.

2.4.1 CONSTANT GROWTH RATE

In a first part of the simulation the growth rate was constant. Considering that the measured signal on a time scale is periodic with an integer number of period in the measurement window, the measured signal on a distance scale is also periodic with an integer number of periods in the measurement window.

The true signal (full black line) and the measured signal (dashed dark gray line) are shown on a distance scale and on a time scale in Fig. 2.11a and c. The constant growth rate, \( v = 1 \), is shown in Fig. 2.11b.

Fig. 2.11: The measured signal (dashed gray line), the true signal (full black line) and the samples (horizontal gray lines, only a number of samples is shown) as a function of distance (a) and time (c) and the constant growth rate (b).
2.4.1.1 EQUALLY SPACED

In this example, the data set was measured with equally spaced samples. The sampling was overlapping, with a sample width of \( \delta \) of 0.3 cm. A total of \( N = 100 \) samples were taken in a measurement window of \( T = 3.6630 \) cm. Since the measured signal was periodic with an integer amount of periods within the measurement window, leakage did not occur when transforming the measured signal to the frequency spectrum. Therefore, the optimal \( \alpha \)-value was equal to 1.

In Fig. 2.12 the noiseless case is shown. The measured signal (dashed dark gray line), the true signal, (full black line) and the signal which was corrected for averaging (dash-dotted light gray line) are shown as a function of distance (a) and as a function of time (b). It is clear from this figure that both the true signal and the signal which was corrected for averaging coincide completely.

![Fig. 2.12: Equally spaced. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are shown as a function of distance (a) and as a function of time (b).](image)

The result of the Monte Carlo simulation is shown in Fig. 2.13 as a function of distance (a) and time (b). It is clear from this figure that about 30% of the points fall outside the 95% confidence level. Therefore, it can be concluded that the method is not very robust to the noise in case of equally spaced data which was periodic with an integer number of periods in the measurement window.

Next, a correction of the noiseless signal, sampled with only \( N = 12 \) non-overlapping samples is performed. In Fig. 2.14 the true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are shown
on a distance scale (a) and a time scale (b). The correction for averaging puts the measured points (markers) exactly onto the true continuous signal (i.e. the signal as it was incorporated in the environmental archive).

Fig. 2.13: Difference between the true simulated signal and the average of 1000 corrected equally spaced harmonic measurement sets (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

Nevertheless, the actual minima and maxima are not measured. These are discretization errors as described in Section 1.3.2.4. In this example, there are 3 samples per period and following the Shannon theorem a reconstruction is possible. This is illustrated in Fig 2.14 where the signal is reconstructed with a band-limited interpolation (2.8) (on the distance grid from Fig. 2.12) (dashed dark gray line) prior to the correction for averaging (dash-dotted dark gray line).

Fig. 2.14: Equally spaced (12 samples). The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) as a function of distance scale (a) and time (b).

From Fig. 2.15 it follows that the reconstructed signal and the true signal coincide
completely.

Fig. 2.15: Equally spaced (12 samples), band-limited interpolation. The true signal (full black line), the band-limited interpolated signal (dashed dark gray line) and the correction of the band-limited interpolated signal (dash-dotted light gray line) as a function of distance (a) and time (b).

In Fig. 2.16 the result for the Monte-Carlo simulation is shown in case of a simple correction for averaging on a distance scale (a) and on a time scale (b). From this figure it follows that half of the points fall outside the 95% confidence level, therefore we can conclude that the method is not very robust to the noise for equally spaced harmonic data when only 12 samples are available.

Fig. 2.16: Difference between the true simulated signal and the average of 1000 corrected equally spaced harmonic measurement sets (12 samples) (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

In Fig. 2.17 the result for the Monte Carlo simulation is shown in case of a band-limited interpolation prior to the correction for averaging as a function of distance (a) and time (b). When a band-limited interpolation is performed, the noise on the neighboring
points will no longer be independent. Therefore, the difference between the true signal and the simulated signal should be evaluated at the distances and times in Fig. 2.16 (before the band-limited interpolation). From Fig. 2.17 it follows that 5 out of 12 points fall outside the confidence level. Therefore, we can conclude that a band-limited interpolation (12 samples), followed by the correction for averaging performs slightly better than a simple correction for averaging in case of an equally spaced periodic record with an integer number of periods in the measurement window.

![Fig. 2.17: Difference between the true simulated signal and the average of 1000 corrected equally spaced harmonic measurement sets (12 samples band-limited interpolation) (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).](image)

**2.4.1.2 Equally Spaced with Missing Data**

In this example the measured signal was equally spaced, though with missing data points. Therefore, the measured signal from the previous section was used and samples 4 to 6, 21, 30 to 33, 55 to 57 were omitted to create an equally spaced data set with missing points with a total of \( N = 89 \) samples. Since the measured signal was periodic with an integer amount of periods in the measurement window, leakage did not occur when transforming the measured signal to the frequency spectrum. Therefore, the optimal \( \alpha \)-value was equal to 1.

In Fig. 2.18 the noiseless case is shown. The true signal (full black line), the measured signal (dashed dark gray line) and the signal which was corrected for averaging (dash-dotted light gray line) are shown on a distance scale (a) and on a time scale (b). From this figure it follows that the reconstruction is complete. Where data is missing there is a gap. In Fig. 2.18 these gaps are linearly connected and it becomes clear that the signals
do not overlap at these locations. Therefore, it can be concluded that a linear interpolation, as is generally done to reconstruct missing points, is not justified. These are discretization errors which can be addressed with a band-limited interpolation.

Fig. 2.18: Equally spaced with missing data. The true signal (full black line), the measured signal (dashed dark gray line), and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

In a second approach a band-limited interpolation onto the distance grid from Section 2.4.1.1 is performed prior to the correction for averaging. In Fig. 2.19a and b the results for the noiseless case are shown on a distance scale and on a time scale, respectively. It is clear that the reconstructed signal and the true signal coincide completely.

Fig. 2.19: Equally spaced with missing data, band-limited interpolation. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

The result of the Monte Carlo simulation in case of a simple correction for averaging is shown in Fig. 2.20 as a function of distance (a) and time (b).
It is clear from this figure that about 40% of the points fall outside the 95% confidence boundaries. Therefore, it can be concluded that a simple correction for averaging is not very robust to the noise in case of an equally spaced periodic record with integer number of periods in the measurement window and missing sample.

In Fig. 2.21 the results of the Monte Carlo simulation are shown in case of a band-limited interpolation prior to the correction for averaging.

About 5% of the points fall outside the 95% confidence level and therefore, it can be concluded that the method for an equally spaced periodic record with integer number of periods in the measurement window and missing samples is robust to the noise, when a
band interpolation is performed prior to the correction for averaging. In this case a band-limited interpolation prior to the correction for averaging definitely performs better than a simple correction for averaging.

### 2.4.1.3 UNEQUALLY SPACED

In this example the measured data record was unequally spaced. The measured signal was constructed as in Section 2.4.1.1 and was than interpolated over a random distance axis \([0.15, 3.7796]\) \(\{x_2, \ldots, x_{N-1}\}\) sorted random numbers, \(x_1 = 0.15\) and \(x_{end} = 3.7796\), yielding a total of \(N = 94\) samples. Since the measured signal was periodic with an integer number of periods in the measurement window, leakage did not occur when transforming the measured signal to the frequency spectrum. Therefore, the optimal \(\alpha\)-value was equal to 1.

In Fig. 2.22 the true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are shown on a distance scale (a) and on a time scale (b). From this figure it follows that the signal is completely reconstructed.

![Fig. 2.22: Unequally spaced. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) as a function of distance (a) and time (b).](image)

The second approach (band-limited interpolation onto the equally spaced distance grid from Section 2.4.1.1 plus correction for averaging) was tested here as well. The results are displayed in Fig. 2.23. It is clear that the true signal and the reconstructed signal coincide completely.
Fig. 2.23: Unequally spaced, band-limited interpolation. The true signal (full black line), the band-limited interpolated signal (dashed dark gray line) and the correction of the band-limited interpolated signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

In Fig. 2.24 the results from the Monte Carlo simulation in case of a simple correction for averaging are shown as a function of distance (a) and time (b). From this figures it follows that about 50% of the points fall outside the 95% confidence levels and therefore it can be concluded that the method is not robust to the noise.

Fig. 2.24: Difference between the true simulated signal and the average of 1000 corrected unequally spaced harmonic measurement sets (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line) (based on the standard deviation of the sample mean).

The results for the Monte Carlo simulation in case of a band-limited interpolation prior to the correction for averaging are shown in Fig. 2.25. About 50% of the points fall outside of the confidence level. When a band-limited interpolation is performed on the unequally spaced harmonic data record prior to the correction for averaging, it is as least as robust to the noise at a simple correction for averaging.
2.4.2 NON-CONSTANT GROWTH RATE

In a second part of the simulation the growth rate was non-constant, meaning that a periodic signal on a time scale will no longer be periodic on a distance scale.

The true signal and the measured signal are shown on a distance scale and on a time scale in Fig. 2.26a and c, and the non-constant growth rate is shown in Fig. 2.26b.

2.4.2.1 EQUALLY SPACED

The data set in this example is equally spaced. The width of the sample was 0.15 and sampling was overlapping. A total of $N = 100$ samples were taken in a measurement window $T = 1.8315$ cm. The parameter $\alpha$ for which $\sum_n e^2(x_n)$ was minimal was 1.3260.

In Fig. 2.27 the true signal (full black line), the measured signal (dashed dark gray line) and the results for the correction for averaging (dash-dotted light gray line) are shown.
on a distance scale (a) and on a time scale (b). The results are very promising even for the low amplitudes.

Fig. 2.27: Equally spaced. The true signal (full black line), the uniformly measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

In Fig. 2.28 the result for the Monte Carlo simulation is shown on a distance scale and on a time scale. From this figure it follows that most of the points fall outside of the 95% confidence level. The larger the averaging effect, the less good the results are.

Fig. 2.28: Difference between the true simulated signal and the average of 1000 corrected equally spaced non-harmonic measurement sets (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

2.4.2.2 EQUALLY SPACED WITH MISSING DATA

The data set in this example was equally spaced with missing data points. The measured signal were constructed as in the previous section, though samples 4 to 6, 21, 30 to 33, 55 to 57 were omitted from the measured signal, giving $N = 89$ samples. The parameter
$\alpha$ for which $\sum_n e^2(x_n)$ was minimal was 1.502.

In Fig. 2.29 the noiseless case is shown. The true signal (full black line), the measured signal (dashed dark gray line) and the result for the correction for averaging are shown on a distance scale (a) and on a time scale (b). It is clear from this figure that the method still performs well even when the data is not equally spaced and not harmonic.

![Fig. 2.29: Equally spaced with missing data. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).](image)

The second approach (band-limited interpolation onto the equally spaced distance grid from Section 2.4.2.1 plus correction for averaging) was tested here as well. In Fig. 2.30, the results for this approach are shown. Here the true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dah-dotted light gray line) are shown on a distance scale (a) and on a time scale (b). At a first glance the reconstructions for both approaches seem very similar.

Therefore, a more objective comparison of the results obtained with the correction for averaging and with the band-limited interpolation plus correction for averaging with the non-averaged signal was performed by comparing the root mean square error (RMSE). From these root mean squared errors it follows that the correction for averaging (RMSE = 0.0782 (a.u.)) performs slightly better than the interpolation plus correction for averaging (RMSE = 0.0816 (a.u.)).
Fig. 2.30: Equally spaced with missing data, ban-limited interpolation. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

The results for the Monte Carlo simulation for a simple correction for averaging are shown in Fig. 2.31 on a distance scale (a) and on a time scale (b). Most of the samples fall outside the 95% confidence level. When the averaging effect is larger, the correction is less robust to the noise. Overall it can be concluded that the method is not robust to the noise.

The results of the Monte-Carlo simulation in case a band-limited interpolation prior to the correction for averaging are shown in Fig. 2.32 on a distance scale (a) and on a time scale (b). From this figure it follows that most points fall outside the confidence interval.

Fig. 2.31: Difference between the true simulated signal and the average of 1000 corrected equally spaced non-harmonic measurement sets with missing data (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

Under these conditions, a band-limited interpolation prior to the correction for averaging is less robust to the noise, compared to a simple correction for averaging.
CHAPTER 2: NONPARAMETRIC ELIMINATION OF BIAS-AVERAGING ERRORS IN PROXY RECORDS

2.4.2.3 UNEQUALLY SPACED

The data set in this example was unequally spaced. The measured signal was interpolated over the distance axis from Section 2.4.2.1 to obtain unequally spaced measurements. The parameter $\alpha$ for which $\sum_n e(x_n)^2$ was minimal was 1.94.

In Fig. 2.33 the results for the first approach (correction for averaging) are shown in the noiseless case. Here the true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are plotted on a distance scale (a) and on a time scale (b). Again the amplitudes are not always completely reconstructed. The largest errors occur where the discretization errors are the largest.

The second approach (band-limited interpolation onto the equally spaced distance grid...
(Section 2.4.2.1) plus correction for averaging) was also applied here, to bid a solution to the discretization errors. The results for the noiseless case are shown in Fig. 2.34. Here, the true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line). From this figure it follows that the second approach gives better results in the noiseless case.

Fig. 2.34: Unequally spaced, band-limited interpolation. The true signal (full black line), the measured signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) on a distance scale (a) and on a time scale (b).

The result for the Monte Carlo simulation in case of a simple correction for averaging is shown in Fig. 2.35 on a distance scale (a) and on a time scale (b). From this figure it follows that the method is not robust to the noise. Overall, it can be said that the method is not robust to the noise for unequally spaced data.

Fig. 2.35: Difference between the true simulated signal and the average of 1000 corrected unequally spaced non-harmonic measurement sets with missing data (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

The result for the Monte Carlo simulation in case of band-limited interpolation prior to
the correction for averaging is shown in Fig. 2.36 on a distance scale (a) and on a time scale (b). Again it becomes clear that the method is not robust to the noise. The larger the averaging effect, the larger the errors. When we compare the graph of the second approach with that of the first approach we can see that the errors are smaller, except for a very large averaging effect. Overall, it can be concluded that a band-limited interpolation prior to the correction for averaging yields better results.

Fig. 2.36: Difference between the true simulated signal and the average of 1000 corrected band-limited interpolated unequally spaced non-harmonic measurement sets with missing data (full black line) plotted as a function of distance (a) and time (b), with the 95% confidence interval (dash-dotted gray line).

2.4.2.4 CONCLUSIONS ON THE ROBUSTNESS TO THE NOISE

The method is not robust to the noise. When the signal is non-harmonic, the robustness to the noise increases with an increasing averaging effect.

Discretization errors are solved via a band-limited interpolation in the simulations. The combination of a band-limited interpolation with the correction for averaging is more robust to the noise than without a band-limited interpolation in all cases except for the case of the non-harmonic signal with missing data points. Overall, the errors produced by the method with the band-limited interpolation are smaller.

The location where samples are missing in an equally spaced data set with missing data is important for the robustness to the noise. Missing samples at locations where the averaging effect is large, will lead to less robust corrections.

Possibly, the robustness to the noise can be improved by estimating the $\alpha$-value in equation 2.5 with the ORFT described in Vanherzeele et al. (2008). Nevertheless, as this method is only used to generate starting values for the parametric method the
CHAPTER 2: NONPARAMETRIC ELIMINATION OF BIAS-AVERAGING ERRORS IN PROXY RECORDS

estimation of the $\alpha$ value was performed as explained in Section 2.2.2.

2.5 APPLICATIONS

2.5.1 MANGROVE TREES

Mangrove trees were chosen as a case study because the sampling procedures are non-destructive and, as a consequence, identical sections can be measured with different sample sizes, which makes it possible to check the method for different sampling strategies.

A stem disc of a mangrove tree *R. mucronata* was collected in November 1999 from Makongeni, Kenya (39.46°E, 5.7°S), located 50 km South of Mombasa. The disc is now part of the xylarium of the Royal Museum for Central Africa (RMCA) in Tervuren Belgium.

The bimodal distribution of the Kenyan climate is locally expressed by a long rainy season from April to July and a short one from October to November, with a mean annual precipitation of 1144 mm (1980-1985) (Leith et al., 1999). The temperatures range from 23.3 to 29.9 °C, with a mean annual temperature of 26.1°C (1931-1990). The vessel density in *R. mucronata* is a proxy for the rainfall in the tropics: during the rainy season earlywood with a low vessel density is produced and during the dry season latewood with a high vessel density is produced. The bimodal climate distribution is reflected in this vessel density by more than one frequency peak.

The stem discs were air dried and their transversal sections were polished (sand grain 100–1200 grit). In Fig. 2.37a an example of such a disc is shown. All measurements were performed along a radial transect from bark to pith. Two window widths were used to measure the exact, same section (300 μm and 1200 μm, the height was constant at 2092.9 μm). The vessel density was equidistantly measured with a window height of 300 μm yielding a 62 samples long record. The 300 μm window was much smaller than the variation of interest. Therefore, this dataset is the true dataset, as it was recorded by the mangrove tree: the dataset with no averaging errors $y(x_n)$ (SS= small sample). Furthermore the vessel density was measured (i) equidistant (Section 2.5.1.1), (ii) equidistant with missing data (Section 2.5.1.2), and (iii) non-equidistant (Section
2.5.1.3) with a window width of 1200 μm. The larger the sample height, the larger the averaging effect, and, consequently, an underestimation of the variability of the data measured with a 1200 μm window is expected. Therefore, these data represent the averaged data: $\bar{y}_{ES}$ (ES = equally spaced), $\bar{y}_{MD}$ (MD = equally spaced, missing data) and $\bar{y}_{US}$ (US = unequally spaced). The vessels were counted at a magnification of 12 x using image analysis software (AnalySIS 3.0) and recalculated to find the vessel density per square millimeter. Fig. 2.37b shows an example of such an image.

![Image](image.png)

**Fig. 2.37: Rhizophora mucronata.** (a) is a picture of a cross section of a mangrove tree. The vertical arrow represents the axis on which the measurements were performed. (b) is a microscopic view of stem cross section (diameter = 10 cm). The lines delimit the samples (300 μm apart), while the white dots are the vessels.

Identical sections are measured with both windows, the results should therefore be identical, except for averaging errors. When sampling procedures are at the base of diverse results, a calibration is needed; this is the role of the method presented in this chapter.

### 2.5.1.1 EQUALLY SPACED DATA

In this section, the 1200 μm window data was equidistantly measured, yielding 59 samples. The vessel density measured with the 300 μm window, $y(x_n)$, was compared in Fig. 2.38 with that measured with the 1200 μm window, $\bar{y}_{ES}(x_n)$, and the expected underestimation of the variability was observed.

Both datasets were subjected to the correction for averaging.

For the vessel density measured with a window of 300 μm, $y(x_n)$, the results showed no
significant change, so it was concluded that this was indeed the “true” dataset.

The correction of the vessel density measured with a window of 1200 \( \mu m \), \( y_{ES}(x_n) \), showed a significant improvement towards the “true” dataset, \( y(x_n) \). This can be seen in Fig. 2.38, where the true dataset (full black line), the measured dataset (dashed dark gray line), and the corrected dataset (dash-dotted light gray line) are shown on a distance scale. The optimal \( \alpha = 1.2 \) and the measurement window \( T = 12900 \mu m \).

![Figure 2.38: The correction for averaging for equidistant vessel density.](image)

Fig. 2.38: The correction for averaging for equidistant vessel density. The averaged data, \( \overline{y}_{ES}(x_n) \) (dashed dark gray line); the “true” data, \( y(x_n) \) (full black line); and corrected data, \( y_{ES}(x_n) \) (dash-dotted light gray line) plotted as a function of distance.

The results were then transformed to a time scale using the Time Base Distortion Method (TBD) (De Ridder et al., 2004); this is shown in Fig. 2.39a. It is clear that the corrected data (dash-dotted light gray line) reflect the bimodal climate in Kenya, whereas the measured data (dashed dark gray line) do not. This becomes even clearer in the frequency domain Fig. 2.39b. The measured dataset (dashed dark gray line) shows one strong frequency component and a very weak frequency component, while the corrected dataset (dash-dotted light gray line) shows two clear frequency components. The higher the frequency, the more the peaks are corrected; this can be explained by the shape of the filter (Fig. 2.6). The improvement in the second peak accounts for the sudden appearance of the bimodal climate structure.

### 2.5.1.2 EQUALLY SPACED DATA SET WITH MISSING DATA

In this section, the data set was equidistantly measured with missing data points. To achieve this, samples 4 to 7 and 24 to 27 were omitted from the data in Section 2.5.1.1.
Fig. 2.39: The correction for averaging for equidistant vessel density data. The measured vessel density data (*dashed dark gray line*) and corrected vessel density data (*dash-dotted light gray line*) in time domain (a) and in the frequency domain (b).

The length of the final record was 51 samples. The vessel density measured with the 300 μm window, \( y(x_n) \), was compared with that measured with the 1200 μm window, \( \bar{y}_{MD}(x_n) \), in Fig. 2.40 and the expected underestimation of the variability was observed. The missing points are situated around 600 μm to 800 μm and 12000 μm to 13000 μm (see Fig. 2.40).

Fig. 2.40: The correction for averaging for equidistant vessel density data with missing samples. The averaged vessel density data \( \bar{y}_{MD}(x_n) \) (*dashed dark gray line*), the non-averaged vessel density data \( y(x_n) \) (*full black line*) and the corrected vessel density record \( y_{MD}(x_n) \) (*dash-dotted light gray line*) as a function of the distance at sample position \( x_n \).

The correction of the vessel density data measured with 1200 μm, \( y_{MD}(x_n) \), is shown in Fig. 2.40, where the true data (full black line), the measured data (dashed dark gray line)
and the measured averaged signal (dash-dotted light gray line) are plotted as a function of distance. The correction resulted in a significant improvement towards the “true” signal, \( y(x_n) \). The optimal \( \alpha \) for this correction was 1.25 and the measurement window was \( T = 12900 \mu m \). The places where the samples are missing are linearly connected, as is usually done. It becomes clear that this is not justified.

A band-limited interpolation was performed onto the equidistant distance grid from Section 2.5.1.1 prior to the correction for averaging, in order to test which approach yields better results. In Fig. 2.41, the result, a refined signal on a distance scale \( \tilde{y}_{IMD}(x_n) \) (IMD = interpolated missing data) (dashed dark gray line), is shown. The band-limited interpolation introduces an unexpected minimum around 0.65. These errors are introduced by the Fourier series model on which this interpolation is based.

![Fig. 2.41: The correction for averaging for equidistant vessel density data with missing samples (band-limited interpolation). The averaged vessel density data \( \tilde{y}_{IMD}(x_n) \) (dashed dark gray line), the non-averaged vessel density data \( y(x_n) \) (full black line) and the corrected vessel density record \( y_{IMD}(x_n) \) (dash-dotted light gray line) as a function of the distance at sample position \( x_n \).](image)

The correction of this interpolated data record, \( y_{IMD}(x_n) \), is shown in Fig. 2.41 (dash-dotted light gray line), where the “true” signal (full black line), the measured averaged signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are plotted as a function of the distance at sample position \( x_n \). From this figure it follows that the corrected signal, \( y_{IMD}(x_n) \), is improved toward the “true” signal, \( y(x_n) \). The bimodal distribution is more visible, nevertheless most of the maxima are overestimated. The optimal \( \alpha \)-value was equal to 1.27 and the measurement window
In order to make sure which method performs better, the RMSE value was calculated for both approaches. The method without a band-limited interpolation performs better (RMSE = 2.1530 #/mm²) than the method with a band-limited interpolation (RMSE = 2.5206 #/mm²).

### 2.5.1.3 UNEQUALLY SPACED DATA

In this section, the 1200 µm window data was non-equidistantly measured, yielding a total of 52 samples. The vessel density measured with the 300 µm window, y, was compared with that measured with the 1200 µm window, \( \bar{y}_{US} \), in Fig. 2.42a, and the expected underestimation of the variability was observed.

![Fig. 2.42: The correction for averaging for non-equidistant vessel density data.](image)

The correction of the unequally sampled signal, \( y_{US}(x_n) \), is shown in Fig. 2.42. Here, the “true” signal (full black line), the measured averaged signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are plotted as a function of the distance at sample position \( x_n \). The corrected signal, \( y_{US}(x_n) \), leans more towards the true signal, though the first peak and the last dale are overestimated. The optimal \( \alpha \) was equal to 1.13 and the measurement window \( T = 12600 \mu m \).

In a second approach, a band-limited interpolation onto the distance grid from Section
2.5.1.1 was performed prior to the correction for averaging. The result is a refined signal on the distance scale, $y_{US}$. The correction for averaging for this data record is shown in Fig. 2.43, where the “true” signal (full black line), the measured averaged signal (dashed dark gray line) and the corrected signal (dash-dotted light gray line) are plotted as a function of the distance at sample position $x_n$. From this figure it follows that the data has improved towards the true signal. The optimal $\alpha$-value = 1.16 and the measurement window $T = 12600 \mu m$.

In order to check which method performs better, the RMSE values of both approaches were calculated. The method with band-limited interpolation (RMSE = 1.7486 #/mm$^2$) performs better than the method without a band-limited interpolation (RMSE = 2.4395 #/mm$^2$).

### 2.5.2 ICE CORE DATA

Ice cores were chosen as a case study because identical sections can be evaluated with different sample sizes.

Ice cores are core samples of ice removed from ice sheets which contain an abundance of climate information (Fig. 2.44). The data used in this section are the EPICA Dome C Ice Cores Deuterium Data (Jouzel and EPICA, 2004). This ice core provides a climate record.
for the past 740000 years, here only the last 214683 years are used. The isotopic composition of the ice, $\delta D$, is classically used as an indicator of temperature change. The length of the core samples, drilled by the EPICA team, was 0.55 m. The data used in the EPICA paper were the $\delta D$ data averaged over 3.85 m, meaning that each $\delta D$ value is the average $\delta D$ value of 7 successive cores. Averaging over successive cores improves the signal-to-noise ratio. For this experiment two sample widths were used (3.85 m and 23.1 m), the sample values for the largest sample were calculated as the average value across six successive data points. Sampling was overlapping for the 23.1 m sample width, with a distance of 3.85 m between successive samples.

The larger the sample width, the larger the averaging effect, so non-identical results are expected. The smallest sample width was much smaller than the variation of interest, so the correction is expected to be insignificant; this dataset is the ‘true dataset’, i.e. as it was recorded in the ice sheet $y(x_n)$. The data set with a sample width of 21.3 m was (i) equidistantly measured and (ii) equidistantly measured with missing data. Non-equidistant measurements were not performed here, because we were limited to a minimum distance of 3.85 m. These data sets represent the averaged data sets $y_{ES}(x_n)$ and $y_{MD}(x_n)$. As a consequence the corrected 21.3 m data record is expected to be more similar to this true data record, after correction.

### 2.5.2.1 EQUALLY SPACED DATA

The $\delta D$ record from the 21.3 m core was compared with the record from the 3.85 m core and as expected underestimation of the variation was observed in the 21.3 m data record (Fig. 2.45).

Both datasets were subjected to the correction for averaging. The $\delta D$ record from the
3.85 m core showed no significant change, as expected. The δD record from the 21.3 m core, on the other hand, showed a significant improvement towards the ‘true’ data set. This can be seen in Fig. 2.45, where the “true” δD record (full black line), the measured, averaged δD record (dashed dark gray line) and the corrected δD record (dash-dotted light gray line) are plotted as a function of depth. The maxima and minima are mostly completely reconstructed, they are however sometimes shifted with respect to the “true” record.

![Fig. 2.45: The correction for averaging for equidistant δD data. The averaged δD data \( \bar{y}_{ES}(x_n) \) (dashed dark gray line), the non-averaged δD data \( y(x_n) \) (full black line) and the corrected δD data \( y_{IUS}(x_n) \) (dash-dotted light gray line) as a function of the depth at sample position \( x_n \).](image)

2.5.2.2 EQUALLY SPACED DATA SET WITH MISSING DATA

Here an equidistant record was used with missing data. To achieve this, samples 4 to 6 and 41 to 42 were omitted from the record measured with a sample width of 21.3 m in Section 2.5.2.1. The “true” data set \( y(x_n) \) and the averaged data set \( \bar{y}_{MD}(x_n) \) were compared and the expected underestimation of the variability was observed (see Fig. 2.46).

The correction for averaging, \( y_{MD}(x_n) \), is shown in Fig. 2.46, where the true δD record (full black line), the measured averaged δD record (dashed dark gray line) and the corrected δD record (dash-dotted light gray line) are plotted as a function of depth. The correction for averaging results in a signal that is more similar to the “true” signal. Nevertheless, some overshoots are visible and the peak between 2950 m and 3000 m...
seems like a mirror image of the “true” peak. Remark here that the sampling frequency is $\frac{1}{3.85}$ as sampling is overlapping.

Fig. 2.46: The correction for averaging for equidistant $\delta D$ data with missing data. The averaged $\delta D$ data $\bar{y}_{MD}(x_n)$ (dashed dark gray line), the non-averaged $\delta D$ data $y(x_n)$ (full black line) and the corrected $\delta D$ data $y_{MD}(x_n)$ (dash-dotted light gray line) as a function of the depth at sample position $x_n$.

In a second approach, a band-limited interpolation was performed prior to the correction for averaging. In Fig. 2.47 the “true” $\delta D$-record (full black line), the measured $\delta D$-record (dashed dark gray line) and the corrected $\delta D$-record (dash-dotted light gray line) was shown as a function of depth.

Fig. 2.47: The correction for averaging for equidistant $\delta D$ data. The averaged $\delta D$ data $\bar{y}_{MD}(x_n)$ (dashed dark gray line), the non-averaged $\delta D$ data $y(x_n)$ (full black line) and the corrected $\delta D$ data $y_{MD}(x_n)$ (dash-dotted light gray line) as a function of the depth at sample position $x_n$. 

61
The correction improved the record towards the “true” signal. In order to check which approach performed better, the RMSE values were calculated. The method without a band-limited interpolation (RMSE = 0.3492 ‰) performs slightly better than the method with a band-limited interpolation (RMSE = 0.3918 ‰).

2.5.3 REMARK

The correction for averaging does not provide an estimation of the time base. In this section the time base was calculated via the procedure explained in De Ridder et al. (2004).

In Chapter 3, a parametric approach will be used to integrate the growth anomaly and the correction for the averaging effect. The use of a parametric model requires good initial values for the parameters, which can be provided by the non-parametric approach presented in this paper. The noise reduction, which is an additional advantage of a parametrical model, result in a maximum precision, while the model selection simultaneously maximizes the accuracy.

2.6 CONCLUSION

Sampling with nonzero sample sizes causes an averaging of the true signal over the volume of that sample. As a result the value of that sample will be an underestimate of the true value.

In this chapter a non-parametric method is presented to correct for these averaging errors. It was seen that the averaging behaved as a low pass filter in the frequency domain. Moreover, the correction implied that the data on the distance scale was multiplied by a sinc-function, which only depends on the sample width. Since the width of the sample is known, these artifacts can be corrected for.

In case of non-overlapping, $f_s/2$ should be smaller than the first zero of the sinc-function (Shannon theorem), the correction works well from DC to $f_s/2$.

In case of overlapping sampling $f_s/2$ can be higher than the first zero of the sinc-function, and a shortcoming of this method is that frequency components, which are higher than the first zero of the sinc, are not reconstructed due to noise problems.
However, one can wonder if these can even be detected.

Environmental proxy data are often non-harmonic or unequally spaced, therefore the Regressive Discrete Fourier Transform was applied in order to transform the data to the frequency domain.

The method was applied to synthetic data with a constant growth rate which was equidistantly sampled, equidistantly sampled with missing data and non-equidistantly sampled. Two approaches were tested: (i) correction for averaging and (ii) band-limited interpolation plus correction for averaging. When noiseless data was tested, the signal was completely reconstructed. The robustness to the noise was investigated by means of a Monte-Carlo simulation. The first approach proved to be less robust to the noise than the second approach.

Next the method was used to correct synthetic data with a non-linear growth rate. Again, equidistant, equidistant with missing data and non-equidistant sampling was tested. In general, it can be said that the method performed well in all circumstances. Nevertheless, the amplitudes were sometimes not completely reconstructed. The second approach yielded much better results in all cases except in the case of the non-periodic, equally spaced record with missing points, where the results were slightly less good. The robustness to the noise was tested for all cases. Here we saw that the method was more robust to the noise with the band-limited interpolation, except for the equally spaced record with missing points.

The vessel density in *R. mucronata* was chosen as a first case study. The record was measured equally spaced, equally spaced with missing data and unequally spaced. The results after correction, were much more similar to the ‘true’ signal and, moreover, the bimodal structure of the climate became visible. Only for the unequally spaced data record, the results were disappointing. The approach with a band-limited interpolation yielded less good results than the approach without a band-limited interpolation.

Deuterium isotopes in ice cores were chosen as a second case study. The record was measured equally spaced and equally spaced with missing data. The results after correction, were much more similar to the ‘true’ signal. The true maxima and minima were mostly completely reconstructed, though, some maxima and minima were shifted.
The combination of a band-limited interpolation and a correction for averaging gives the best results in most cases. Therefore we suggest to combine both methods for the generation of the starting values for the parametric method discussed in the next Chapter.

The method discussed in this paper is a valuable tool for calibrating proxy measurements; the method not only corrects the visible frequency components but also corrects the nearly invisible frequency components. Proceeding in this way boundaries are stretched. This conclusion also holds for other proxies and other samples shapes.

APPENDIX 2.A: MATHEMATICAL OUTLINE

The averaged or measured signal can be written as:

\[ \bar{y}(x) = \frac{1}{\delta} \int_{x_n - \frac{\delta}{2}}^{x_n + \frac{\delta}{2}} y(x') \, dx' = \frac{1}{\delta} \int_{-\infty}^{+\infty} y(x') \chi_{[x - \frac{\delta}{2}, x + \frac{\delta}{2}]}(x') \, dx' \]  

(2.12)

where \( \bar{y}(x) \) is the measured signal as a function of distance, \( x \), \( y(x') \) is the ideal continuous signal as a function of the distance, \( x' \), \( \delta \) is the width of the sample, and \( \chi_{[x - \frac{\delta}{2}, x + \frac{\delta}{2}]} \) is the characteristic function of the interval \( [x - \frac{\delta}{2}, x + \frac{\delta}{2}] \), equal to 1 on that interval and equal to 0 outside: \( \chi_{[x - \frac{\delta}{2}, x + \frac{\delta}{2}]}(x') \) is equal to one wherever

\[ x - \frac{\delta}{2} \leq x' \leq x + \frac{\delta}{2} \]  

(2.13)

which is equivalent to

\[ -\frac{\delta}{2} \leq x - x' \leq \frac{\delta}{2}. \]  

(2.14)
Hence:

\[ \chi[x - \frac{\delta}{2} + x'] = \chi[-\frac{\delta}{2}](x - x') \tag{2.15} \]

and thus:

\[ \bar{y}(x) = \frac{1}{\delta} y * \chi[-\frac{\delta}{2}](x), \tag{2.16} \]

where * denotes the convolution product. As mentioned, the Fourier transform of a convolution product is equal to the product of the Fourier transforms.

Now the Fourier transforms of the characteristic function \( \chi[-\frac{\delta}{2}] \) is

\[
\mathfrak{F} \left[ \chi \left[ -\frac{\delta}{2} \right] (x) \right] (\omega) = \int_{-\infty}^{\infty} \chi \left[ -\frac{\delta}{2} \right] (x)e^{-j\omega x}dx
\]

\[
= \int_{-\delta/2}^{\delta/2} e^{-j\omega x}dx
\]

\[
= \frac{1}{-j\omega} \left[ e^{-j\omega \frac{\delta}{2}} - e^{-j\omega \frac{-\delta}{2}} \right]
\]

\[
= \frac{e^{j\frac{\delta \omega}{2}} - e^{-j\frac{\delta \omega}{2}}}{j\omega} = \frac{2 \sin \left( \frac{\delta \omega}{2} \right)}{\omega} \tag{2.20}
\]

and

\[
\mathfrak{F}[\bar{y}(x)](\omega) = \frac{1}{\delta} \mathfrak{F} \left[ y * \chi \left[ -\frac{\delta}{2} \right] (x) \right] (\omega) \tag{2.21}
\]
\[
\frac{2\sin(\frac{\delta \omega}{2})}{\delta \omega} \Im \{y(x)\}(\omega)
\]

and defining

\[
\frac{\sin(x)}{x} = \text{sinc}(x)
\]

Equation 2.22 becomes:

\[
\Im \{\bar{y}(x)\}(\omega) = \text{sinc}\left(\frac{\delta \omega}{2}\right) \Im \{y(x)\}(\omega)
\]

APPENDIX 2.B: LEAST SQUARES SOLUTION

Fourier series:

\[
\begin{bmatrix}
\bar{y}_1 \\
y_1 \\
\vdots \\
\bar{y}_N
\end{bmatrix} = \begin{bmatrix}
1 & \cos\left(\frac{2\pi x_1}{\tau}\right) & \ldots & \cos\left(\frac{2\pi H x_1}{\tau}\right) & \sin\left(\frac{2\pi x_1}{\tau}\right) & \ldots & \sin\left(\frac{2\pi H x_1}{\tau}\right) \\
1 & \cos\left(\frac{2\pi x_2}{\tau}\right) & \ldots & \cos\left(\frac{2\pi H x_2}{\tau}\right) & \sin\left(\frac{2\pi x_2}{\tau}\right) & \ldots & \sin\left(\frac{2\pi H x_2}{\tau}\right) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cos\left(\frac{2\pi x_N}{\tau}\right) & \ldots & \cos\left(\frac{2\pi H x_N}{\tau}\right) & \sin\left(\frac{2\pi x_N}{\tau}\right) & \ldots & \sin\left(\frac{2\pi H x_N}{\tau}\right)
\end{bmatrix} \begin{bmatrix}
A_0 \\
A_H \\
A_{1+H} \\
\vdots \\
A_{2H}
\end{bmatrix}
\] (2.25)

where \(2H + 1 \leq N\).

From here on:
The Least Squares solution for $A$ is given by:

$$A = (F^T F)^{-1} F^T Y$$  \hspace{1cm} (2.29)$$

$A$ is calculated via the reduced singular value decomposition (svd) of $F$ (Golub and Van Loan, 1996), $F = U \Sigma V^T$, giving

$$A = V \Sigma^+ U^T Y.$$  \hspace{1cm} (2.30)$$
3 TIME-SERIES RECONSTRUCTION FROM NATURAL ARCHIVE DATA WITH THE AVERAGING EFFECT TAKEN INTO ACCOUNT


3.1 INTRODUCTION

Measuring proxies produces very short data records and mostly involves sampling solid substrates. Sampling is subject to some problems as mentioned in the introduction (see Section 1.3.2.1, 1.3.2.2, 1.3.2.3 and 1.3.2.4). In this chapter some of these problems will be addressed.

A first problem is that the accretion rate of natural archives is dependent on environmental and physiological factors. Each data record therefore has its unique non-constant accretion rate and thus its own nonlinear distance–time relationship. Natural archives are sampled on a distance grid along their accretion axis. Starting from these distance series, a time series needs to be constructed, as comparison of different data records is only meaningful on a time grid. In Fig. 2.10a transformation is shown from the distance scale onto the time scale over a non-constant distance–time relationship. From this figure it follows that the signal on a distance scale and on a time scale are very different, due to this non-constant growth rate. Without a correction for this growth anomaly, erroneous conclusions would be made.

A second problem, a typical example of sampling solid substrates, is drilling. As a result of the dimensions of the drill, the holes drilled will not be infinitesimally small. Consequently, samples are not taken at a point in distance, but rather over a volume in distance. This holds for most sampling methods in solid substrates. Thus, when the continuous proxy signal is sampled, it will be averaged over the volume of the sample (Goodwin et al., 2003). This is also exemplified in Fig. 2.10 on the distance axis, where the averaged signal is derived by repetitively taking the average value of the continuous signal over a length of 0.2 cm. Whether this averaging effect is significant depends on the volume of the sample and the variations of interest of the proxy signal. Starting from the measured signal, the continuous signal needs to be reconstructed in order to eliminate these averaging errors.

The existing methods (e.g. de Brauwere et al., 2008; De Ridder et al., 2004; Lisiecki and Lisiecki, 2002; Martinson et al., 1982a; Paillard et al., 1996; Yu and Ding, 1998) (see Section 1.3.2.1) do not take into account problem two. However, when solid substrates are sampled with a significantly large sample volume, the amplitudes of the signal
harmonics will be underestimated. The aim of this chapter is to provide an efficient identification algorithm for identifying a parametric model of the nonlinearities in the distance–time relationship (also called distance base distortions) that takes into account the averaging effects; this is performed in the distance domain. As such, both of the aforementioned problems will be simultaneously addressed here. The method described in Chapter 2 is a non-parametric approach to problem two and will be used for the generation of starting values and validation purposes in this chapter.

The outline of this chapter is as follows. The harmonic signal model and the measurement model are described in Section 3.2. A Gaussian maximum likelihood is constructed in Section 3.3 and its statistical properties are discussed. The following important issues are handled: the numerical stable calculation of the estimates, the generation of starting values, the model selection for short data records (the number of data points is not much larger than the number of model parameters), and the calculation of the Cramér-Rao lower bound. Finally the new modeling approach is compared with the classical method that does not account for the averaging effect. Since environmental records are short, the consistency of the estimator and the approximation of the covariance matrix by the Cramér-Rao lower bound are verified by Monte-Carlo simulations in Section 3.4. Finally, in Section 3.5 the method is exemplified by the vessel density in mangrove trees.

### 3.2 THE PARAMETRIC SIGNAL MODEL

To identify the nonlinearities in the distance-time relationship, some assumptions about the unknown signal on a time axis need to be made. The mapping methods described in the introduction match the proxy record with a temporally calibrated reference function. This reference function is completely fixed and an assumption is needed that is less stringent than this complete fixation of the time axis. Therefore, in this work, the signal on a time scale \( y(t) \) is assumed to be harmonic. This is justifiable as proxy records often exhibit a seasonal cycle. Nevertheless, this assumption restricts the method to those proxies which have a clear periodical signature. If the signal is non-periodic, then it can still be arbitrarily well approximated by a classical Fourier series (Brigham, 1988) with an infinite number of sines and cosines. To overcome the infinite series number, non-periodic signals will be approximated by a Fourier series where the period length is
chosen to be somewhat larger than the time span of the data record (Arruda, 1992; Vanherzeele et al., 2006). This will significantly reduce the required number of sines and cosines for a given approximation error (Section 2.2.2). Since the distance-time relationship is nonlinear, the signal on a distance scale will be nonlinearly distorted.

The aim is thus to estimate a nonlinearly distorted harmonic signal, in the presence of additive noise and averaging effects. Parametric models which estimate a harmonic signal in the presence of additive noise and a nonlinear distortion are intensively studied in literature (de Brauwere et al., 2008; Schoukens et al., 1997; Stenbakken and Deyst, 1998; Vandersteen et al., 2001; Verspecht, 1994). The novelty here is that the parametric signal model includes the averaging effects. Provided that the method is robust to over- and under-modeling, it has the advantage that the stochastic noise and significant variations are separated.

### 3.2.1 HARMONIC SIGNAL MODEL

A proxy record is a continuous signal to be measured. When this signal is sampled, each sample value is equal to the average of the continuous proxy signal over the width of the sample. The result is a measured data record, that is, the measured averaged signal $\bar{y}(x_n)$, which is related to the true continuous proxy record. For example, the continuous signal $y(x)$, we want to identify as

$$\bar{y}(x_n) = \frac{1}{\delta} \int_{x_n - \frac{\delta}{2}}^{x_n + \frac{\delta}{2}} y(x) dx,$$

(3.1)

where $\delta$ is the sample width over which the continuous proxy signal is averaged (see Fig. 1.17), $x_n = n\Delta$ is distance at sample position $n$, with $n \in \{1, 2, ..., N\}$, $\Delta$ the distance between two subsequent samples and $x$ the continuous distance. For non-uniformly sampled signals $\Delta$ varies from sample to sample. From this equation, it is clear that the measured averaged signal $\bar{y}(x_n)$ depends on the sample width $\delta$ over which it is averaged. Since the results should not be influenced by the sample characteristics, it is important to find the true value of the continuous proxy signal $y(x_n)$ at $x_n$, the distance
at sample position \(n\).

To model the continuous proxy signal, we start from the hypothesis that the continuous proxy signal on a time scale is a sum of sine and cosine waves

\[
y(t) = A_0 + \sum_{k=1}^{H}[A_k \cos(2\pi f_0 kt) + A_{k+H} \sin(2\pi f_0 kt)],
\]

where \(A_0\) is the offset; \(A_k\) and \(A_{k+H}\) are the amplitudes of the \(k\)th harmonic \((k \in \{1, \ldots, H\})\), \(H\) is the number of signal harmonics, \(f_0\) is the fundamental frequency, and \(t\) the time. The distance-time relationship is nonlinear and can be expressed as follows

\[
2\pi f_0 t = \psi(x),
\]

with \(2\pi f_0 t\) described in (3.2) and \(\psi(x)\) a nonlinear function of the distance \(x\). This nonlinear function \(\psi(x)\) is modeled as the sum of a linear and a nonlinear contribution

\[
\psi(x, L) = 2\pi \lambda_0 (x + g(x, B)),
\]

where \(\lambda_0\) is the spatial frequency expressed in m\(^{-1}\), \(g(x, B)\) is the nonlinear phase distortion that is modeled as

\[
g(x, B) = \sum_{l=1}^{b} B_l \Phi_l(x),
\]

with \(B = [B_1, B_2, \ldots, B_b]\) the vector of unknown phase distortion parameters, \(\Phi_l(x)\), \(l = 1, 2, \ldots, b\) a set of basis functions, and \(L = [\lambda_0, B]\). In general, the distortion can be expanded by any set of basis functions \(\Phi_l(x)\). An appropriate choice of basis functions
will limit the number of coefficients $B_i$ needed to describe the nonlinear function $g(x, B)$. The method is developed with either splines (see Fig. 3.1a) or orthogonal polynomials (see Fig. 3.1b), developed through a Gram-Schmidt orthogonalization, as basis functions. Splines are chosen as basis functions in this chapter (Dierckx, 1993), in case accretion hiatuses (growth stops under certain conditions) are present in the archives.

Collecting Equation (3.2) to (3.5), the following signal model $y(x, \theta)$ is obtained

$$y(x, \theta) = A_0 + \sum_{k=1}^{H} [A_k \cos(k\psi(x,L)) + A_{k+H} \sin(k\psi(x,L))],$$

where $\theta = [A_0, A_1, ..., A_{2H}, L]$ are the model parameters, with $A_0$ the unknown offset, $A_k$ and $A_{k+H}$ the amplitudes of the $k$th harmonic ($k \in \{1, ..., H\}$), $H$ the number of signal harmonics and $L = [\lambda_0, B]$. Remark that the signal model in (3.6) still holds for signals of the form:

$$A_0 + \sum_k A_k \cos(2\pi f_k t) + A_{k+H} \cos(2\pi f_k t)$$

where the frequencies $f_1, f_2, ..., f_H$ are integer multiples of the same fundamental
frequency \( f_0 \). This frequency \( f_0 \) is equal to the greatest common divisor of the frequencies \( f_1, f_2, \ldots, f_H \): \( f_k = n_k f_0 \) with \( n_1 < n_2 < \cdots < n_H \) integer numbers. Replacing \( k \) in (3.6) by \( n_k \) then gives the signal model corresponding to (3.7). Note that (3.7) is a special case of (3.2) where the amplitudes of the non-excited harmonics are set to zero.

When a signal is constructed with spatial frequencies \( \lambda_k = 1 \text{ cm}^{-1} \) and 1.7 cm\(^{-1} \), for example, then the fundamental frequency, which is the greatest common divisor of 1 and 1.7, will be equal to 0.1. The only excited harmonics are the 10\(^{th} \) harmonic with \( \lambda_{10} = 1 \) and the 17\(^{th} \) harmonic with \( \lambda_{17} = 1.7 \). The appropriate excited harmonics are selected in practice via a model selection criterion (see Section 3.3.3).

From (3.1) and (3.6) follows the model for the measured averaged signal \( \bar{y}(x, \theta) \)

\[
\bar{y}(x_n, \theta) = \frac{1}{\delta} \int_{x_n - \frac{\delta}{2}}^{x_n + \frac{\delta}{2}} y(x, \theta) \, dx,
\]

with \( y(x, \theta) \) the model for the continuous proxy signal, as described in (3.6).

Since, in general, \( \varphi(x, L) \) in (3.5) is a nonlinear function of \( x \), (3.8) cannot be solved analytically. Hence, the integral in (3.8) will be numerically approximated using the composite trapezoidal rule (Ralston and Rabinowitz, 1984) that divides the averaging interval \([x - \frac{\delta}{2}, x + \frac{\delta}{2}]\) into \( M \) intervals of length \( \frac{\delta}{M} \)

\[
\bar{y}(x_n, \theta) = \frac{1}{M} \left[ \frac{1}{2} y(x_n - \frac{\delta}{2}, \theta) + \sum_{k=1}^{M-1} y(x_n - \frac{\delta}{2} + \frac{k}{M} \delta, \theta) + \frac{1}{2} y(x_n + \frac{\delta}{2}, \theta) \right].
\]

By identifying the parametric model (3.8), model (3.6) is evaluated with the estimated model parameters \( \hat{\theta} \) giving an estimate \( y(x_n, \hat{\theta}) \) of the true proxy value at \( x_n \), the distance at sample position \( n \). Furthermore, to calculate the time series, the distance-time relationship (3.3) should be calibrated. For this purpose, one needs the time laps between two observations (relative time):
with $x_1$ and $x_2$ the two observations with corresponding relative times $t_1$ and $t_2$. From Equation (3.10), it follows that

$$f_0 = \frac{\psi(x_2) - \psi(x_1)}{2\pi(t_2 - t_1)},$$

(3.11)

which then allows the time instant to be calculated using (3.3). Small errors in relative times $t_1$ and $t_2$ will lead to small errors in $f_0$. Therefore, after the estimation of $t_1$ and $t_2$ these values can be fine-tuned based on some prior knowledge. For example, when we know that we want to reconstruct a yearly cycle, $t_1$ and $t_2$ can be tuned such that the frequency of one of the harmonics in our signal model is equal to exactly 1 year$^{-1}$. Nevertheless, we should be careful when we fine-tune these time instants. The model estimates the average fundamental frequency for the proxy record which has to be modeled. For short records for example, it is possible that a certain event did not happen exactly one year later than the same event the previous year and therefore deviations from a frequency of 1 year$^{-1}$ are possible. Because the calibration of the time axis is done after the parameter estimation this does not affect the parameter values.

Note that the time $t$ depends on the nonlinear function $\psi(x)$ and thus on the distance $x$ as well. This implies that $y(x, \theta)$, (3.6), is equal to $y(\psi^{-1}(2\pi f_0 t), \theta)$, (3.3) and (3.6), and thus that $y(x, \theta)$ is also a function of $t$.

### 3.2.2 Measurement Model

Since the samples are collected independently and analyzed separately, it is reasonable to postulate the following measurement model

$$\bar{y}(x_n) = y_0(x_n) + e(x_n),$$

(3.12)
with \( \bar{y}(x_n) \) the measured averaged proxy record, \( \bar{y}_0(x_n) \) the true unknown averaged proxy record, and \( e(x_n) \) zero mean, independent and identically distributed noise with variance \( \sigma^2 \).

### 3.3 THE IDENTIFICATION ALGORITHM

The signal model parameters and phase distortion parameters should be estimated starting from \( N \) measurements \( \bar{y}(x_n) \), \( n = 1,2,\ldots,N \). Assuming that the measurement noise \( e(x_n) \) is normally distributed, the Gaussian maximum likelihood (ML) estimator \( \hat{\theta} \) minimizes

\[
V(\bar{y}, \theta) = \sum_{n=1}^{N} (\bar{y}(x_n) - \bar{y}(x_n, \theta))^2,
\]

w.r.t \( \theta \) (Pintelon and Schoukens, 2001), with \( \theta \) the parameters, \( V(\bar{y}, \theta) \) the cost function, \( \bar{y}(x_n, \theta) \) described in (3.9). The ML estimate of the variance \( \sigma^2 \) of the measurement noise \( e(x_n) \) is given by

\[
\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (\bar{y}(x_n) - \bar{y}(x_n, \hat{\theta}))^2,
\]

The Gaussian ML estimator \( \hat{\theta} \) has the following asymptotic \( (N \to \infty) \) properties: it is (i) consistent (the estimate converges to the true value); (ii) asymptotically efficient (the covariance matrix asymptotically reaches the Cramér-Rao lower bound); and (iii) asymptotically normally distributed. In addition, the consistency and the asymptotic normality properties are robust to departures from the independence and normality assumptions on \( e(x_n) \). Indeed, the asymptotic properties remain valid for non-Gaussian noise \( e(x_n) \), correlated over the distance \( n \) as long as the correlation tends sufficiently fast to zero (mixing condition), the proofs of all these properties can be found in (Pintelon and Schoukens, 2001).

In case of signal modeling errors, \( \hat{\sigma}^2 \) is the sum of the noise variance and the mean of the
squared modeling errors. Indeed, using (3.12),

\[ \bar{y}(x_n, \hat{\theta}) = \bar{y}(x_n) + \Delta \bar{y}(x_n, \hat{\theta}) \]  

(3.15)

and

\[ \theta_* = \lim_{N \to \infty} \theta, \]  

(3.16)

with \( N \) the number of measurements. \( \hat{\sigma}^2 \) (3.14) becomes for \( N \) sufficiently large:

\[ \hat{\sigma}^2 \approx \frac{1}{N} \sum_{n=1}^{N} (e(x_n) + \Delta \bar{y}(x_n, \theta_*))^2 \]

\[ \approx \frac{1}{N} \sum_{n=1}^{N} e^2(x_n) + \frac{1}{N} \sum_{n=1}^{N} (\Delta \bar{y}(x_n, \theta_*))^2 \]

\[ \approx \sigma^2 + \frac{1}{N} \sum_{n=1}^{N} (\Delta \bar{y}(x_n, \theta_*))^2 \]  

(3.18)

If the modeling error \( \Delta \bar{y}(x_n, \theta_*) \) is uniformly spread over the samples \( x_n \), then the uncertainty bounds based on \( \hat{\sigma}^2 \) will not reveal the modeling errors. However, if \( \Delta \bar{y}(x_n, \theta_*) \) is concentrated in a few samples, then the residuals \( \bar{y}(x_n) - \bar{y}(x_n, \theta_*) \) at these samples, will be significantly larger than the 95% uncertainty bound based on \( \hat{\sigma}^2 \) and, hence, the modeling error will be detected.

Remark: When outliers are present in the proxy signal, e.g. Barium peaks in bivalves (Gillikin et al., 2008), the ML estimator needs to be robustified towards these outliers (Huber, 1972; Ljung, 1999). This is typically done by replacing the squared residuals in (3.12) by their absolute value if they exceed same threshold.
3.3.1 NUMERICAL ISSUES

The iterative Newton-Gauss algorithm (Fletcher, 1991) for minimizing (3.13) is of the form

\[(J^T J) \Delta \theta = J^T e,\]  \hspace{1cm} (3.19)

where \(\theta\) are the parameters, \(\Delta \theta = \theta^{(i)} - \theta^{(i-1)}\), \(i\) is the iteration number, \(e\) is an \(N \times 1\) vector with \(n\)th entry such that

\[e(x_n) = \bar{y}(x_n) - \tilde{y}(x_n, \theta^{(i)}),\]  \hspace{1cm} (3.20)

with \(\bar{y}(x_n)\) the measured values and \(\tilde{y}(x_n, \theta)\) as described in (3.9), and where \(J\) is the \(N \times (2H+b+2)\) Jacobian matrix with \((n,j)\) the entry

\[J_{[n,j]} = \frac{\partial \tilde{y}(x_n, \theta^{(i)})}{\partial \theta_{[j]}},\]  \hspace{1cm} (3.21)

(see Appendix 3.A for the explicit expressions). To solve expression (3.19) in a numerically stable way, the columns of \(J\) are first scaled by their 2-norm

\[J \rightarrow J_S = JS^{-1},\]  \hspace{1cm} (3.22)

where \(S\) is a diagonal matrix containing the 2-norm of each column of \(J\). Using (3.22), (3.19) can be rewritten as
\[(J_S^T S) \Delta \theta_S = J_S^T e, \quad (3.23)\]

where $\Delta \theta_S = S \Delta \theta$. Next, (3.23) is solved via the singular value decomposition (Golub and Van Loan, 1996) of the scaled Jacobian matrix $J_S = U \Sigma V^T$

\[
\Delta \theta_S = V \Sigma^{-1} U^T e, \quad (3.24)
\]

and the actual parameter update $\Delta \theta$ is obtained as

\[
\Delta \theta = S^{-1} \Delta \theta_S. \quad (3.25)
\]

Finally, to increase the convergence area of the iterative procedure, the Levenberg-Marquardt version of (3.24) is implemented

\[
\Delta \theta_S = V \left[ \Sigma (\Sigma^2 + \lambda^2 I_{n_0})^{-1} \right] U^T e, \quad (3.26)
\]

with $\lambda$ the Levenberg-Marquardt factor (Fletcher, 1991; Pintelon and Schoukens, 1996):

\[
\begin{cases}
\lambda_{\text{start}} = \frac{\sigma_{\text{max}}}{100} \\
\text{cost decreases: } \lambda \rightarrow 0.4 \lambda' \\
\text{cost increases: } \lambda \rightarrow 10 \lambda
\end{cases} \quad (3.27)
\]

$I_{n_0}$ the $n_0 \times n_0$ identity matrix, and $n_0$ the number of parameters.
3.3.2 GENERATION OF STARTING VALUES

The cost function will only be minimized successfully when a reasonable set of initial values are selected. In principle the method described in this chapter can be used for time series reconstruction from equally spaced data, equally spaced data with missing data points and unequally spaced data. The difficulty is finding good starting values.

Different methods can be used to generate starting values. The non-parametric method described in De Ridder et al. (2004), produces good starting values for equally spaced data sets. This method assumes that there was no averaging. The method described in Chapter 2, produces good starting values for equally spaced data with missing data points and for unequally spaced data. In both cases the averaging is taken into account.

3.3.3 MODEL SELECTION

Choosing the optimal model complexity is very important; because models which are too complex are sensitive to the noise and models which are not complex enough introduce systematic errors.

In most cases, some prior knowledge is available. This can be helpful to narrow down the choice and help set a maximum number of parameters. The actual model selection is performed in an objective manner by means of a model selection criterion. This model selection criterion is a trade-off between the residual cost function, which is a measure for the model fit, and a penalty function, which penalizes increasing model complexities. The modified minimum description length selection criterion, MDLc, based on (De Ridder et al., 2005) was used here, because it is especially developed for short data records.

If some harmonic frequencies are not exited, then the model selection is more complex. Indeed, it is not sufficient to try signal models with increasing number of harmonics because all possible signal models where some of the harmonics are not excited should also be considered.

3.3.4 CRAMÉR-RAO LOWER BOUND

Since the ML estimate \( \hat{\theta} \) is asymptotically efficient, the covariance matrix \( \text{Cov}(\hat{\theta}) \) is
approximated very well by the Cramér-Rao lower bound (CRB). Following the lines of Pintelon and Schoukens (2001); Pintelon et al. (2007), it is easily found that

$$\text{Cov}(\hat{\theta}) \approx \hat{\sigma}^2 (J^T J)^{-1},$$

(3.28)

where the Jacobian $J$ is defined in (3.21), and $\hat{\sigma}^2$ in (3.14). Together with the asymptotic normality property of the estimates, this allows the construction of uncertainty bounds on the estimates with a given confidence level.

3.3.5 COMPARISON OF THE METHOD WITH ($\delta = 0$) AND WITHOUT ($\delta \neq 0$) AN AVERAGING EFFECT

The novelty in this work is the averaging effect and, hence, necessitates a comparison between the methods, including the averaging ($\delta \neq 0$) with the method neglecting averaging ($\delta = 0$). The residual cost function is usually employed for comparison, but it will be clear from this section that both cases ($\delta \neq 0$ and $\delta = 0$) can be equally well explained by the signal model. By applying the mean value theorem for integrals (Kaplan, 1993)

$$\int_a^b f(x)dx = f(c)(b - a),$$

(3.29)

with $f(x)$ a continuous function and $c \in (b, a)$, it follows that

$$\bar{y}(x_n) = \frac{1}{\delta} \int_{x_n - \frac{\delta}{2}}^{x_n + \frac{\delta}{2}} y(x)dx = y(x_i),$$

(3.30)

where $x_i \in \left\{x_n - \frac{\delta}{2}, x_n + \frac{\delta}{2}\right\}$. This shows that the averaged data $\bar{y}(x_n)$ can also be
modeled exactly as (3.6). When the averaging effect is not taken into account ($\delta = 0$), the measured data are assumed equal to the real proxy data, and therefore the measured data will be modeled as (3.6)

$$\bar{y}_{(\delta=0)}(x_n, \theta) = A_0 + \sum_{k=1}^{H}[A_k \cos(k\psi(x_n, L)) + A_{k+H} \sin(k\psi(x_n, L))], \quad (3.31)$$

When the averaging effect is taken into account ($\delta \neq 0$), a single measured sample is the averaged value of the continuous proxy signal over the sample width and, therefore, the measured data will be modeled as (3.8)

$$\bar{y}_{(\delta \neq 0)}(x_n) = \frac{1}{\delta} \int_{x_n-\frac{\delta}{2}}^{x_n+\frac{\delta}{2}} y(x) dx, \quad (3.32)$$

Note that the cost function for the model with $\delta = 0$ (3.31) will be smaller than that for the model with $\delta \neq 0$ (3.32), because no numerical approximation (3.9) of integral (3.8) is needed.

The main difference between both cases will be the difference in the amplitudes of the signal harmonics. In Chapter 2, it was shown that the spectra of the continuous signal and measured averaged signal are related as follows

$$\mathfrak{Z}[\bar{y}(x_n)](\omega) = \text{sinc}\left(\frac{\delta\omega}{2}\right) \mathfrak{Z}[y(x_n)](\omega), \quad (3.33)$$

where $\omega$ is the angular frequency, $\mathfrak{Z}[\bar{y}(x_n)](\omega)$, $\mathfrak{Z}[y(x_n)](\omega)$ are the Fourier transforms of the measured averaged signal and the continuous signals respectively, and $\text{sinc}(x) = \sin(x)/x$. This proves that the averaging effect leads to an underestimation of the amplitudes of the signal harmonics. When $\delta = 0$, this effect will not be corrected, while for $\delta \neq 0$, this effect will be corrected, explaining the difference in the amplitudes.
3.4 SIMULATION RESULTS

3.4.1 ROBUSTNESS TO THE NOISE ON THE PROXY DATA

From the theory it is known that the estimate is consistent and that the covariance matrix calculated in (3.23) is asymptotically \((N \to \infty)\) exact. Whether these asymptotic properties are still valid for short data records, which are (i) equally spaced, (ii) equally spaced with missing data and (iii) unequally spaced, is still an open question. Therefore, and in order to assess the influence of the noise on the actual parameters, a Monte Carlo simulation is performed for each case with \(N = 100\) samples (Fishman, 1995). The harmonic content and the distortion parameters of the signal model described in Equations (3.8) and (3.6), where the nonlinearly varying phase is given by Equation (3.4) and the distortion by Equation (3.5) were simultaneously estimated by minimizing (3.13).

For each case, 500 measurement sets in the time interval \([0,2]\) year were generated with the following parameters: the fundamental spatial frequency, \(\lambda_0 = 1\) \(\text{cm}^{-1}\); signal harmonics, \(A = [1, 0.1, 0.8, 0.7, 0.6]\); and the distortion parameters, \(B = [-0.2, 0.5, -0.3, 0.4]\). The basis functions in the nonlinear distortion function \(\psi(x)\) (3.4) were splines. The length of the record was \(N = 100\) samples. The sample width was \(\delta = 0.2\) and the number of intervals which divides the averaging interval \([x_n - \frac{\delta}{2}, x_n + \frac{\delta}{2}]\) is \(M = 1000\) (3.8). In Fig. 3.2a the true signal (full black line) and the measured signal (dashed gray line) and in Fig. 3.2b time is shown as a function of distance. Zero mean white Gaussian noise was added, leading to a signal-to-noise-ratio of 11.6, which is the signal-to-noise ratio used to test the noise sensitivity of the different methods in de Brauwere et al. (2009) and is a realistic noise level for proxy records.

The simulations were also performed with polynomials as basis functions and yielded similar results.
Fig. 3.2: (a) The true signal (full black line) and the measured signal (dashed gray line) and (b) time as a function of distance.

3.4.1.1 EQUALLY SPACED DATA

In this simulation 500 equally spaced measurement sets were created. The results for this simulation can be found in Table 3.1. The CRB (3.23) was calculated and compared to the sample variances from the 500 Monte Carlo runs of the estimated parameter. For 500 Monte-Carlo runs, the ratio between the sample variance and CRB (3.23) lies between 0.91 and 1.09 with 95% confidence. From Table 3.2, it can be seen that the ratio sample variance/CRB of all the parameters fall within this bound.

Table 3.1: Summary of the uncertainties on the parameter estimated with a Monte-Carlo simulation.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0$</th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>b</td>
<td>1.0006</td>
<td>0.0032</td>
<td>1.0008</td>
<td>0.8003</td>
<td>0.6999</td>
</tr>
<tr>
<td>c</td>
<td>0.0003</td>
<td>0.0071</td>
<td>0.0158</td>
<td>0.0220</td>
<td>0.0183</td>
</tr>
<tr>
<td></td>
<td>$A_4$</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$b_4$</td>
</tr>
<tr>
<td>a</td>
<td>0.6</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>b</td>
<td>0.5988</td>
<td>-0.2002</td>
<td>0.4999</td>
<td>-0.3002</td>
<td>0.4002</td>
</tr>
<tr>
<td>c</td>
<td>0.0264</td>
<td>0.0062</td>
<td>0.0067</td>
<td>0.0069</td>
<td>0.0064</td>
</tr>
</tbody>
</table>

a: true parameter values, b: estimated parameter values and c: standard deviation of the sample mean.
Table 3.2: Comparison of the Cramer Rao lower bound to the variances of the estimated parameters.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0$</th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.3416e-7</td>
<td>5.0702e-5</td>
<td>2.4856e-4</td>
<td>4.8297e-4</td>
<td>3.393e-4</td>
</tr>
<tr>
<td>B</td>
<td>1.3223e-7</td>
<td>5.6708e-5</td>
<td>2.4311e-4</td>
<td>4.7570e-4</td>
<td>3.2308e-4</td>
</tr>
<tr>
<td>C</td>
<td>1.0146</td>
<td>0.8941</td>
<td>1.0224</td>
<td>1.0153</td>
<td>1.0337</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>0.9794</td>
<td>0.9503</td>
<td>1.1000</td>
<td>0.9640</td>
<td>1.0201</td>
</tr>
</tbody>
</table>

a: sample variances of the estimated parameter values, b: CRB, c: sample variance/CRB.

### 3.4.1.2 EQUALLY SPACED WITH MISSING DATA

In this simulation 500 equally spaced data sets were generated with missing samples. To achieve this, samples 4 to 6, 21 to 23, 30 and 55 to 59 were omitted from the record, leading to $N = 78$ samples. The results for the simulation can be found in Table 3.3. From this table it follows that all the parameters are found within the 95% confidence interval. The sample variance reaches the Cramer-Rao lower Bound (CRB) asymptotically, which means that the estimator is consistent. The ratio sample variance/CRB are shown in Table 3.4. All values in Table 3.4 lie in the 95% confidence bound [0.91,1.09].

Table 3.3: Summary of the uncertainties on the parameters estimated with a Monte Carlo simulation for equally spaced data with missing points.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0$</th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>b</td>
<td>1.0000</td>
<td>-0.0006</td>
<td>1.0005</td>
<td>0.7992</td>
<td>0.7003</td>
</tr>
<tr>
<td>c</td>
<td>0.0004</td>
<td>0.0083</td>
<td>0.0190</td>
<td>0.0293</td>
<td>0.0218</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.6</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>b</td>
<td>0.6003</td>
<td>-0.2000</td>
<td>0.4993</td>
<td>-0.2999</td>
<td>0.4003</td>
</tr>
<tr>
<td>c</td>
<td>0.0357</td>
<td>0.0083</td>
<td>0.0088</td>
<td>0.0096</td>
<td>0.0085</td>
</tr>
</tbody>
</table>

a: true parameter values, b: estimated parameter values and c: standard deviation of the sample mean.
Table 3.4: Comparison of the Cramer Rao lower bound to the variances of the estimated parameters for equally spaced data with missing points.

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>$2.0730 \times 10^{-7}$</td>
<td>$6.9231 \times 10^{-5}$</td>
<td>$3.6017 \times 10^{-4}$</td>
<td>$8.6140 \times 10^{-4}$</td>
</tr>
<tr>
<td>b</td>
<td>$2.2421 \times 10^{-7}$</td>
<td>$6.5450 \times 10^{-5}$</td>
<td>$3.7262 \times 10^{-4}$</td>
<td>$9.1287 \times 10^{-4}$</td>
</tr>
<tr>
<td>c</td>
<td>0.9246</td>
<td>1.0578</td>
<td>0.9666</td>
<td>0.9436</td>
</tr>
</tbody>
</table>

$A_4$, $b_1$, $b_2$, $b_3$, $b_4$

<table>
<thead>
<tr>
<th></th>
<th>$A_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.0013</td>
<td>6.8580</td>
<td>7.6948</td>
<td>9.1554</td>
<td>7.2972</td>
</tr>
<tr>
<td>b</td>
<td>0.0014</td>
<td>7.3086</td>
<td>7.8649</td>
<td>9.0961</td>
<td>6.8828</td>
</tr>
<tr>
<td>c</td>
<td>0.9105</td>
<td>0.9383</td>
<td>0.9784</td>
<td>1.0065</td>
<td>1.0602</td>
</tr>
</tbody>
</table>

a: sample variances of the estimated parameter values, b: CRB, c: sample variance/CRB.

3.4.1.3 UNEQUALLY SPACED DATA

In this simulation 500 measurement sets were generated with randomly chosen spacings between the different distance instants (random between 0 and 0.15), with a total of 93 samples. The results for the simulation with the unequally spaced data set can be found in Table 3.5. All parameters are found within the 95% confidence interval. Again in this example the sample variance reaches the Cramer-Rao lower Bound (CRB) asymptotically. The ratio sample variance/CRB are shown in Table 3.6. All values in Table 3.6 lie in the 95% confidence bound [0.91,1.09].

Table 3.5: Summary of the uncertainties on the parameters estimated with a Monte Carlo simulation with unequally spaced data

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>b</td>
<td>1.0000</td>
<td>-0.0003</td>
<td>1.0012</td>
<td>0.806</td>
</tr>
<tr>
<td>c</td>
<td>0.0005</td>
<td>0.0082</td>
<td>0.0207</td>
<td>0.0335</td>
</tr>
</tbody>
</table>

$A_4$, $b_1$, $b_2$, $b_3$, $b_4$

<table>
<thead>
<tr>
<th></th>
<th>$A_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.6</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>b</td>
<td>0.5978</td>
<td>-0.2001</td>
<td>0.5006</td>
<td>-0.3007</td>
<td>0.4003</td>
</tr>
<tr>
<td>c</td>
<td>0.0412</td>
<td>0.0111</td>
<td>0.0080</td>
<td>0.0077</td>
<td>0.0088</td>
</tr>
</tbody>
</table>

a: true parameter values, b: estimated parameter values and c: standard deviation of the sample mean.
Table 3.6: Comparison of the Cramer Rao lower bound to the variances of the estimated parameters for unequally spaced data.

<table>
<thead>
<tr>
<th></th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2.8941e-7</td>
<td>6.7355e-5</td>
<td>4.2749e-4</td>
<td>7.5849e-4</td>
</tr>
<tr>
<td>b</td>
<td>2.9227e-7</td>
<td>6.8449e-5</td>
<td>4.3114e-4</td>
<td>7.0583e-4</td>
</tr>
<tr>
<td>c</td>
<td>0.9902</td>
<td>0.9840</td>
<td>0.9841</td>
<td>1.0746</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.0017</td>
<td>1.2307e-4</td>
<td>6.4019e-5</td>
<td>5.9138e-5</td>
<td>7.7389e-5</td>
</tr>
<tr>
<td>b</td>
<td>0.0018</td>
<td>1.0876e-4</td>
<td>6.3052e-05</td>
<td>5.4160e-5</td>
<td>7.5395e-5</td>
</tr>
<tr>
<td>c</td>
<td>0.9212</td>
<td>1.1316</td>
<td>1.0153</td>
<td>1.0919</td>
<td>1.0264</td>
</tr>
</tbody>
</table>

a: sample variances of the estimated parameter values, b: CRB, c: sample variance/CRB.

3.4.2 ROBUSTNESS TO THE NOISE ON THE DISTANCE AXIS

Noise is not only present in the proxy data, but also in the distance data. Indeed, when the sampling device is relocated to the next sampling position, some errors can be introduced in the values on the distance axis. Errors can be introduced by the positioning system of the sampling device and by the user. When we measure the vessel density in mangrove trees for example (see Section 2.5.1) the width of the windows, and therefore, also the distance between two samples, are determined with image analysis software. The error on the distance instances is in the order of magnitude of micrometers, which means that it is negligible with respect to the sample width ($300 \mu m$ – $1200 \mu m$). When the ice cores removed from the ice sheets, they are divided in samples with a saw which is manually controlled. The error on the distance axis in this case is estimated in the order of magnitude of a few millimeters, which is small in comparison to the width of the sample (0.55 m). In case of extracting oxygen isotopes from for example bivalves, errors are introduced both by the user and the positioning system of the sampling device. Carbonate powder is drilled from the bivalve shell with a drill which moves automatically to the next sample position; this sampling position is chosen manually. The errors introduced by the positioning system of the drill are estimated in the order of magnitude of a few micrometers; while the errors introduced by the user are in the order of magnitude of 10 micrometer. These errors are small in comparison to the size of the drill bit ($300 \mu m$). Due to the positioning mechanism, the uncertainty on the positions is independently distributed.
In order to test the robustness of the method to the errors on the distance axis, a Monte Carlo simulation was performed. 500 equally spaced measurement sets were generated with parameters: the fundamental spatial frequency, $\lambda_0 = 1$ cm$^{-1}$; signal harmonics, $A = [1, 0, 1, 0.8, 0.7, 0.6]$; and the distortion parameters, $B = [-0.2, 0.5, -0.3, 0.4]$. The basis functions in the nonlinear distortion function $\psi(x)$ (3.4) were splines. The length of the record was $N = 100$ samples. The sample width was $\delta = 0.2$ and the number of intervals which divides the averaging interval $[x_n - \frac{\delta}{2}, x_n + \frac{\delta}{2}]$ is $M = 1000$ (3.8). In Fig. 3.2a the true signal (full black line) and the measured signal (dashed gray line) and in Fig. 3.2b time is shown as a function of distance. Zero mean white noise was added to the distance axis, with a signal-to-noise ratio of 5.8, this noise ratio was chosen such that the noise on the distance scale was not larger than the distance between two samples. From the examples given in the previous paragraph it is clear that this is an exaggeration. In Table 3.7 the average of the 500 estimated parameters are shown with the standard deviation of the sample mean. From this table it follows that almost all parameters fall within the 95% confidence level.

Table 3.7: Summary of the uncertainty on the parameter estimated with a Monte-Carlo simulation.

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0$</th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>b</td>
<td>1.0000</td>
<td>-0.0017</td>
<td>0.9753</td>
<td>0.7294</td>
<td>0.6806</td>
</tr>
<tr>
<td>c</td>
<td>0.0447</td>
<td>-7.470e-05</td>
<td>0.0436</td>
<td>0.0326</td>
<td>0.0304</td>
</tr>
<tr>
<td></td>
<td>$A_4$</td>
<td>$A_5$</td>
<td>$A_6$</td>
<td>$A_7$</td>
<td>$A_8$</td>
</tr>
<tr>
<td>a</td>
<td>0.6</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>b</td>
<td>0.5417</td>
<td>-0.1979</td>
<td>0.4975</td>
<td>-0.2969</td>
<td>0.3977</td>
</tr>
<tr>
<td>c</td>
<td>0.0242</td>
<td>-0.0088</td>
<td>0.0222</td>
<td>-0.0132</td>
<td>0.0178</td>
</tr>
</tbody>
</table>

a: true parameter values, b: estimated parameter values and c: standard deviation of the sample mean.

We can conclude that even for very high uncertainties on the distance position, no bias in the estimates can be detected.

This test was also performed for equidistant measurements with missing samples and for non-equidistant measurements giving similar results.
3.5 APPLICATION: MANGROVE TREES

The vessel density measured in the mangrove tree *Rhizophora mucronata* is used to illustrate the method. Sampling procedures are non-destructive. Consequently, identical sections can be measured with different sample sizes, which makes it possible to check the algorithm.

A stem disk of the *R. mucronata* was collected in 1999 from Makongeni, Kenya (39.46°E, 5.7°S), located 50 km South of Mombasa. The disk is now part of the xylarium of the Royal Museum for Central Africa (RMCA) in Tervuren, Belgium. The Kenyan climate has a bimodal distribution, with a long rainy season from April to July and a short rainy season from October to November. The vessel density is a proxy (that is, a source of climate information stored in a natural archive) for the rainfall in tropical coastal regions. During the rainy season, earlywood with a high vessel density is produced, and during the dry season, latewood with a low vessel density is produced.

Before measuring the vessel density, the stem disks are dried and their transversal sections are polished (sand grain 100-1200 grit). In Fig. 2.37a, an example of such a stem disk is shown. The dark rings correspond to low vessel density, the light rings to high vessel density.

Note the non-equidistant spacing of the growth rings, indicating a non-constant growth speed and, hence, a nonlinear distance-time relationship. This can be expected, because the growth of any environmental archive depends on physiological and environmental factors. The data was measured along a radial transect on the stem disk from pith to bark. A window height of a 1200 μm was used, the width was the same everywhere. The data was measured (i) equidistantly, (ii) equidistantly with missing data and (iii) non-equidistantly. The vessels were counted at a magnification of 12x using image analysis software (AnalySIS 3.0) and divided by the surface area of the sample to find the vessel density per square millimeter (Fig. 2.37b).

3.5.1 EQUALLY SPACED DATA

In this section, the 1200 μm window data was equidistantly measured, yielding 34 samples, covering 3 years. In Fig. 3.3a, the vessel density data on a distance scale is
shown. The optimal model complexity as described in Section 3.3.3. This resulted in a signal model (3.6) and (3.8) consisting of $H = 5$ harmonics, $A$, and a distortion model (3.5) with $b = 3$ parameters, $B$, giving a total of 14 signal parameters. After estimating these parameters, the nonlinear phase distortion (3.5), the signal model for the measured averaged vessel density (3.8), and the signal model for the continuous vessel density (3.6) can be calculated. A time base can be calculated using (3.10) and (3.11), and after fine-tuning $t_{end}$ such that one of the frequency components is 1 year$^{-1}$, we obtain a difference, $t_{end} - t_1 = 2.7715$ years (see Section 3.2.1, the next to last paragraph). In Fig. 3.3b, the vessel density data (full line) as well as the signal model for the measured averaged vessel density (dashed line) is shown on the newly calculated time axis. In Fig. 3.3c, the distance-time relationship is shown (dash-dotted line). It is clear from this figure that the distance-time relationship is nonlinear, which is due to environmental and physiological factors.

Fig. 3.3: Vessel density in mangrove trees measured with $\delta = 1200$ µm. The measured data on a distance scale (a), the measured data (full black line) and the averaged signal model $\bar{y}_{ES}(x_n, \theta)$, (3.8), (dashed gray line) on the constructed time scale (b) and the distance-time relationship (dash-dotted line) (c).

Fig. 3.4a shows the signal model for the measured averaged vessel density $\bar{y}_{ES}(x_n, \theta)$,
(3.8), (dashed line), and the signal model with no averaging effects, \( y_{ES}(x_n, \theta) \), (3.6), (full line), on the newly calculated time axis. In Fig. 3.4b, the Fourier spectra of both of these models are shown. Knowing that the fundamental frequency \( f_0 \) for \( y_{ES}(x_n, \theta) \) was 0.5 year\(^{-1}\) and that the model consists of \( H = 5 \) harmonics, one would expect a frequency peak at 0.5, 1, 1.5, 2 and 2.5 year\(^{-1}\). In Fig. 3.4b, only peaks at 1 and 2.5 year\(^{-1}\) are visible. Since the data record is not an integer multiple of the known period \( \left( \frac{1}{f_0} = 2 \text{years} \right) \), leakage effects hide possible peaks at frequencies 0.5, 1.5 and 2 year\(^{-1}\).

![Diagram](image)

Fig. 3.4: The correction for averaging in equally spaced vessel density data. The averaged signal model, \( \bar{y}_{ES}(x_n, \theta) \), (dashed light gray line) and the signal model for the ‘true’ signal, \( y_{ES}(x_n) \), (full dark gray line) in the time domain (a) and the frequency domain (b).

The non-averaged signal model shows an increase of its amplitudes and reveals some new harmonics. In order to test whether this increase of the amplitudes is significant, signal model \( y_{ES}(x_n, \theta) \) (3.6) is fitted on the raw data \( \bar{y}_{ES}(x_n) \). Only the amplitudes of the data are estimated as only the amplitudes are affected by the averaging effect (Section 3.3.3), and, hence, the distortion parameters where kept the same as in \( \bar{y}_{ES}(x_n, \theta) \). The optimal number of amplitudes was chosen with the MDLc selection criterion (Section 3.3.4). In Table 3.8, the magnitude and the standard deviation of the resulting estimated amplitudes are compared with the results obtained when averaging is considered. The covariance matrix of the parameters is approximated by (3.23), for the explicit expression see Appendix 3.B.
Table 3.8: Significance of the amplitude difference between the amplitudes for the continuous vessel density* and the amplitudes of the measured averaged vessel density**.

<table>
<thead>
<tr>
<th>n</th>
<th>(nf_0)</th>
<th>(\sqrt{A_k^2 + A_{k+H}^2} \pm \text{std}^*)</th>
<th>(\sqrt{A_k^2 + A_{k+H}^2} \pm \text{std}^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>2.03 ± 0.21</td>
<td>2.37 ± 0.18</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4.78 ± 0.25</td>
<td>3.59 ± 0.19</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>1.97 ± 0.37</td>
<td>1.56 ± 0.19</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.59 ± 0.53</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2.5</td>
<td>1.93 ± 0.88</td>
<td>-</td>
</tr>
</tbody>
</table>

Here it can be seen that all the estimated magnitudes of the amplitudes are significant. However, the differences between the magnitudes of the amplitudes are not always significant. For the first harmonics, the difference is insignificant, for the second and the third harmonics the difference is significant, and for the fourth and fifth harmonic the amplitudes of the measured averaged vessel density cannot be estimated. This is expected since at higher frequencies the effect of averaging is more prominent and pushes the amplitudes closer to the noise level, making them harder to identify. The result highlights the attainable improvement when the effect of averaging in the data is considered. Note also that the ratio standard deviation/amplitude of the continuous vessel density (Table 3.8, left column) increases with the harmonic number. This is consistent with the observation that the spatial averaging decreases the harmonic amplitudes as one over the spatial frequency, as can be seen in Equation (3.24) and Chapter 2.

3.5.2 EQUALLY SPACED WITH MISSING DATA

In this section the 1200 µm data from Section 3.5.1 was used. Samples 20 to 23 and 40 to 43 were omitted from the record. A band-limited interpolation was performed to eliminate the discretization errors caused by the missing samples. In Fig. 3.5a the vessel data (full dark gray line) and the interpolated vessel density data (dashed light gray line) on a distance scale is shown. The model complexity was chosen the same as in the previous section, because it concerns the same data measured with a different sampling strategy. This resulted in a signal model (3.6) and (3.8) consisting of \(H = 5\) harmonics, \(A\), and a distortion model (3.5) with \(b = 3\) parameters, \(B\), giving a total of 14 signal parameters. After estimating these parameters, the nonlinear phase distortion (3.5), the
signal model for the measured averaged vessel density (3.8), and the signal model for
the continuous vessel density (3.6) can be calculated. A time base can be calculated
using (3.10) and (3.11). After fine-tuning $t_{\text{end}}$, such that one of the frequencies is equal
to 1 year$^{-1}$, a difference, $t_{\text{end}} - t_1 = 2.8135$ years was obtained. This difference is not
equal to that of Section 3.6.1. The difference between the results of Sections 3.6.1 and
3.6.2, 0042 years, is negligible. In Fig. 3.5b, the interpolated vessel density data (dashed
light gray line) as well as the signal model for the measured averaged vessel density (full
dark gray line) is shown on the newly calculated time axis.

Fig. 3.5: Equally spaced vessel density with missing samples in mangrove trees
measured with $\delta = 1200 \, \mu m$. The measured data (full dark gray line) and the band-
limited interpolated measured data (dashed light gray line) on a distance scale (a), the
interpolated data (dashed dark gray line) and the averaged signal model $\bar{y}_{MD}(x_n, \theta)$,
(3.8), (full light gray line) on the constructed time scale (b).

Fig. 3.6a shows the signal model for the band-limited interpolated measured vessel
density $\bar{y}_{MD}(x_n, \theta)$, (3.8), (dashed line), and the signal model corrected for the averaging
effects, $y_{MD}(x_n, \theta)$, (3.6), (full line), on the newly calculated time axis. In Fig. 3.6b, the
Fourier spectra of both of these models are shown. Knowing that the fundamental
frequency $f_0$ for $y_{MD}(x_n, \theta)$ was 0.5 year$^{-1}$ and that the model consists of $H = 5$
harmonics, one would expect a frequency peak at 0.5, 1, 1.5, 2 and 2.5 year$^{-1}$.

In Fig. 3.6b, only peaks at 1 and 2.5 year$^{-1}$ are visible. Since the data record is not an
integer multiple of the known period ($\frac{1}{f_0} = 2\, \text{years}$), leakage effects hide possible peaks
at frequencies 0.5, 1.5 and 2 year$^{-1}$.
CHAPTER 3: TIME SERIES RECONSTRUCTION FROM NATURAL ARCHIVE DATA WITH THE AVERAGING EFFECT TAKEN INTO ACCOUNT

3.5.3 UNEQUALLY SPACED

In this section the 1200 μm data was measured unequally spaced, yielding a total of 36 samples. A band-limited interpolation was performed to eliminate the discretization errors caused by the missing samples. In Fig. 3.7a the vessel data (full dark gray line) and the interpolated vessel density data (dashed light gray line) on a distance scale is shown. The model complexity was chosen the same as in the previous section, i.e. = 5 harmonics, \( A \), and a distortion model (3.5) with \( b = 3 \) parameters, \( B \), because it concerns the same data, measured with a different sampling strategy. After estimating these parameters, the nonlinear phase distortion (3.5), the signal model for the measured averaged vessel density (3.8), and the signal model for the continuous vessel density (3.6) can be calculated. A time base can be calculated using (3.10) and (3.11). After fine-tuning \( t_{\text{end}} \) such that the frequency of one of the harmonics is equal to 1 year\(^{-1}\), the difference, \( t_{\text{end}} - t_{1} = 2.6999 \) years. This means that the record is shorter in time than the records from Section 3.5.1 and 3.5.2. The difference between the result of Section 3.6.1 and 3.6.2 can be neglected. In Fig. 3.7b, the interpolated vessel density data (dashed light gray line) as well as the signal model for the measured averaged vessel density (full dark gray line) is shown on the newly calculated time axis.

Fig. 3.8a shows the signal model for the band-limited interpolated measured averaged...
vessel density $\bar{y}_{US}(x_n, \theta)$, (3.8), (dashed line), and the signal model with no averaging effects, $y_{US}(x_n, \theta)$, (3.6), (full line), on the newly calculated time axis.

Fig. 3.7: Unequally spaced vessel density with missing samples in mangrove trees measured with $\delta = 1200 \, \mu m$. The measured data (full dark gray line) and the band-limited interpolated measured data (dashed light gray line) on a distance scale (a), the interpolated data (dashed dark gray line) and the averaged signal model $\bar{y}_{US}(x_n, \theta)$, (3.8), (full light gray line) on the constructed time scale (b).

Fig. 3.8: The correction for averaging in equally spaced vessel density data. The averaged signal model, $\bar{y}_{US}(x_n, \theta)$, (dashed light gray line) and the signal model for the 'true' signal, $y_{US}(x_n)$, (full dark gray line) in the time domain (b) and the frequency domain.

In Fig. 3.8b, the Fourier spectra of both of these models are shown. Knowing that the fundamental frequency $f_0$ for $y_{US}(x_n, \theta)$ was 0.5 year$^{-1}$ and that the model consists of $H = 5$ harmonics, one would expect a frequency peak at 0.5, 1, 1.5, 2 and 2.5 year$^{-1}$. In
Fig. 3.8b, only peaks at 1 and 2 year\(^{-1}\) are visible. Since the data record is not an integer multiple of the known period \(\frac{1}{f_0} = 2\text{years}\), leakage effects hide possible peaks at frequencies 0.5, 1.5 and 2.5 year\(^{-1}\).

### 3.5.4 COMPARISON

The method for the correction for averaging for equally spaced data, equally spaced data with missing samples, and unequally spaced data was validated in this section. Therefore, the continuous proxy record was re-sampled with a very small sample width (300 \(\mu\text{m}\)) allowing averaging to be neglected. The result is a data record in which averaging does not occur, \(y(x_n, \theta)\).

Note that the smallest possible sample is not always the most favorable sample. Small samples can lead to a problematic signal-to-noise ratio. However, big samples lead to big averaging effects. With the method presented in this paper, the influence of averaging is greatly reduced. Similarly, a time series for this method is still needed. Therefore, signal model (3.6) with the same number of harmonics and the same number of distortion parameters as for the current method was fitted onto \(y(x_n)\), giving a parametric estimation \(y(x_n, \theta)\) of \(y(x_n)\) and a time series \(t(x_n, L)\).

The signal models of the method for equally spaced data \(y_{ES}(x_n, \theta_{ES})\) (dotted light gray line), for equally spaced data with missing samples \(y_{MD}(x_n, \theta_{MD})\) (dashed dark gray line), for unequally spaced data \(y_{US}(x_n, \theta_{US})\) (dash-dotted middle gray line), and the signal model for the signal measured with a small sample \(y(x_n, \theta)\) (full black line) are plotted as a function of their corresponding distance axis (a) and time axis (b) in Fig. 3.9.

The signal models on a distance scale coincide nicely. Each time axis coincide nicely, except for the time axis of the signal model with the small sample, \(t(x_n, L)\). However, these figures do not show whether the difference is significant. Therefore, the 3 signal models \(y_{ES}(x_n, \theta_{ES})\) (light gray dotted line), \(y_{MD}(x_n, \theta_{MD})\) (dark gray dashed line) and \(y_{US}(x_n, \theta_{US})\) (dash-dotted middle gray line) were each subtracted from the signal model for the measurements with the small sample, \(y(x_n, \theta)\), and plotted in Fig. 3.10a as a function of \(x_n\) along with the 95% confidence interval (based on the standard deviation on the parameters of the method for the small sample, see Appendix 3.B) (full black line). Furthermore, the for the 3 newly calculated time bases \(t_{ES}(x_n, L_{ES})\) (light gray
Fig. 3.9: Comparison of the signal models. $y(x_n, \theta)$ (a) and $t(x_n, L)$ (b) (full black line), $y_{ES}(x_n, \theta_{ES})$ (a) and $t_{ES}(x_n, L_{ES})$ (b) (dotted light gray line), $y_{MD}(x_n, \theta_{MD})$ (a) and $t_{MD}(x_n, L_{MD})$ (b) (dashed dark gray line) and $y_{US}(x_n, \theta_{US})$ and $t_{US}(x_n, L_{US})$ (dash-dotted middle gray line) plotted as a function of distance $x_n$.

dotted line), $t_{MD}(x_n, L_{MD})$ (dark gray dashed line) and $t_{US}(x_n, L_{US})$ (dash-dotted middle gray line) were each subtracted from the time base for the measurements with the small sample, $t(x_n, L))$ and plotted in Fig. 3.10a as a function of $x_n$ along with the 95% confidence interval (based on the standard deviation on the parameters of the method for the small sample, see Appendix 3.B) (full black line).

Fig. 3.10: Difference between signal models. (a) $y_{ES}(x_n, \theta) - y(x_n, \theta)$ (light gray dotted line), $y_{MD}(x_n, \theta) - y(x_n, \theta)$ (dark gray dashed line), and $y_{US}(x_n, \theta) - y(x_n, \theta)$ (dash-dotted middle gray line and (b) $t_{ES}(x_n, L_{ES}) - t(x_n, L)$ (light gray dotted line), $t_{MD}(x_n, L_{MD}) - t(x_n, L)$ (dark gray dashed line), and $t_{US}(x_n, L_{US}) - t(x_n, L)$ (dash dotted middle gray line) plotted as a function of distance $x_n$, 95% confidence interval (full black line).

It is clear from Fig. 3.10a that more than 5% of the points fall outside the 95% confidence interval for all signal models. This can be explained by residual signal modeling errors. From Fig. 3.10b it follows that almost all the points fall in the 95% confidence level, meaning that modeling errors are mainly due to the harmonic signal
model and not due to the time base model. Overall, the results are satisfactory.

3.6 CONCLUSION

A method is presented that identifies the distance base distortions while eliminating the averaging effect. The novelty of the method is the correction for averaging, which results in an increase of the estimated signal harmonic amplitudes. The method is based on the assumption that the proxy record on a time scale is periodic. When the proxy record on a time scale is not harmonic, it can still be approximated by a Fourier series with a finite number of harmonics. Moreover, the harmonic signal model can be replaced by a non-harmonic signal model, without altering the distance-time relationship. However, the generation of starting values can be the difficulty here.

The optimal complexity of the model was chosen based on prior knowledge and the cost function, which makes the method robust to over- and under-modeling. The numbers of parameters are chosen with the MDLc, allowing the salient features to be extracted from the data record. Moreover, the method separates the stochastic noise from the significant variations. By employing a Monte Carlo simulation, the robustness to influence of the noise on the measurement was established. The estimator was proven to be an efficient estimator, as the variances of the estimates achieve the Cramér-Rao lower bound.

The vessel density in R. mucronata was chosen as a case study. The data on a time base has a fundamental frequency of 0.5 in all cases (equally spaced, equally spaced with missing samples and unequally spaced data). The corrected signal model shows a significant amplitude increase in some of the harmonics. The method for all cases was compared to the method for a small sample. Some points fell outside of the confidence interval due to residual modeling errors, but overall the method performs well. The method discussed in this work is a valuable tool for the construction of a time base and for the correction of averaging effects. The conclusions hold for other environmental archives with averaging in more than one spatial direction.

The parametric method presented here, greatly reduces the noise in comparison to the non-parametric method, resulting in a better precision; while the use of a good selection criterion results in a better accuracy.
APPENDIX 3.A: THE JACOBIAN MATRIX

The Jacobian $J$ is given by (3.10); from (3.2) this means that $J$ will be calculated as follows

$$J_{[n,j]} = \frac{1}{M} \left[ \frac{\partial y(x_n - \frac{\delta}{2}\theta)}{\partial \theta_{[ij]}} + \sum_{k=1}^{M-1} \frac{\partial y(x_n - \frac{\delta}{2} + \frac{k\delta}{M}\theta)}{\partial \theta_{[ij]}} + \frac{1}{2} \frac{\partial y(x_n + \frac{\delta}{2}\theta)}{\partial \theta_{[ij]}} \right], \quad (3.34)$$

with $n = 1, \ldots, N$

$$\frac{\partial y(x_n\theta)}{\partial A_0} = 1 \quad (3.35)$$

$$\frac{\partial y(x_n\theta)}{\partial A_k} = \begin{cases} \sin(k\psi(x,L)) & k = 1, \ldots, H \\ \cos(k\psi(x,L)) & k = H + 1, \ldots, 2H \end{cases} \quad (3.36)$$

$$\frac{\partial y(x_n\theta)}{\partial \omega} = \sum_{k=1}^{H} A_k 2\pi k (x + g(x,B)) \cos(k\psi(x,L)) - A_{k+H} 2\pi k (x + g(x,B)) \sin(k\psi(x,L)) \quad (3.37)$$

$$\frac{\partial y(x_n\theta)}{\partial B_i} = \left( \sum_{k=1}^{H} A_k k \omega \cos(k\psi(x,L)) - A_{k+H} k \omega \sin(k\psi(x,L)) \right) \frac{\partial (x + g(x,B))}{\partial B_i} \quad (3.38)$$

$$\frac{\partial (x + g(x,B))}{\partial B_i} = \phi_i(x) \quad (3.39)$$
APPENDIX 3.B: THE UNCERTAINTY ON THE ESTIMATED PARAMETERS

The uncertainty on the estimated parameters in Table 3.8 and Fig. 3.10 are calculated as follows

\[ f(\hat{\theta}) = \sqrt{A_h^2 + A_{h+H}^2} \quad \text{and} \quad f(\hat{\theta}) = y(x, \hat{\theta}). \]  
(3.40)

The linearization of \( f(\hat{\theta}) \) gives an approximation of the variance of \( f(\hat{\theta}) \):

\[ f(\hat{\theta}) \approx f(\theta_0) + \frac{\partial f(\theta_0)}{\partial \theta_0} (\hat{\theta} - \theta_0) \]  
(3.35)

\[ \text{var} \left( f(\hat{\theta}) \right) \approx \frac{\partial f(\theta_0)}{\partial \theta_0} \text{cov}(\hat{\theta}) \left( \frac{\partial f(\theta_0)}{\partial \theta_0} \right)^T \]  
(3.41)

\[ \approx \frac{\partial f(\hat{\theta})}{\partial \hat{\theta}} \hat{\sigma}^2 (f^T J)^{-1} \left( \frac{\partial f(\hat{\theta})}{\partial \hat{\theta}} \right)^T \]  
(3.42)
CHAPTER 4: PERIODIC TIME SERIES MODELING WITH GUARANTEED POSITIVE GROWTH RATE ESTIMATION

4.1 INTRODUCTION

The unconstrained identification of the harmonic signal models in de Brauwere et al. (de Brauwere et al., 2008) and in Chapter 3 chooses the best parameters irrespective of the order of the data observations. In other words, subsequent observations can be inversed on the time axis. Both methods are especially vulnerable to this artifact when the number of phase distortion parameters is relatively high and thus when the noise sensitivity is larger.

In de Brauwere et al. (2008) a method, the time inversion constraint method, is described which corrects these time inversions by including inequality constraints on the phase distortion parameters (Fletcher, 1991). The constraint is activated when two subsequent time instants are smaller than 20% of the sample period. For large violations against the order of the time instants, the number of constraints becomes larger than the number of parameters, and consequently, the method becomes unfeasible.

The aim of this chapter is to impose the positivity of the growth rate with a reduced number of linear constraints on the signal model parameters (Fletcher, 1991).

The outline of this chapter is as follows. Section 4.2 describes the parametric signal model and the measurement model. The Gaussian maximum likelihood solution with positive growth rate inequality constraint is formulated in Section 4.3. Section 4.4 presents a 2-step approximation of the exact solution with linear equality constraints. It includes the generation of starting values, and the model selection. Next, the positive growth rate estimation method is compared with the time inversion constraint method (Section 4.5). Further, the noise sensitivity is tested by means of a Monte-Carlo simulation (Section 4.6). Finally, the method is illustrated in Section 4.7 on real life examples: the vessel density in mangrove trees and the Mg/Ca ratio in bivalve shells.

4.2 THE PARAMETRIC SIGNAL MODEL

4.2.1 HARMONIC SIGNAL MODEL

Time series reconstructions as in de Brauwere et al. (2008) and Chapter 3 are based on
parametric signal models, which assume that the signal on a time scale is harmonic. Remark: This assumption is justifiable, as many proxy signals exhibit a seasonal cycle. Nevertheless, this assumption restricts the method to those proxies which have a clear periodic signature. In principle the signal can still be approximated arbitrarily well by a Fourier series, when the proxy record on a time scale is not periodic. Moreover, the harmonic signal model can be replaced by a non-harmonic signal model, without altering the distance-time relationship.

Assuming harmonicity, the signal on a time scale can be written as the sum of sine-waves:

\[ y(t) = A_0 + \sum_{k=1}^{H} [A_k \cos(2\pi f_0 t) + A_{k+H} \sin(2\pi f_0 t)] \]  \hspace{1cm} (4.1)

with \( A_0 \) the unknown offset, \( A_k \) to \( A_{k+H} \) the amplitudes of the \( k \)th harmonic \( (k \in \{1, \ldots, H\}) \), \( H \) the number of signal harmonics, \( f_0 \) the fundamental frequency, and \( t \) the time.

Growth rates depend on environmental and physiological factors and therefore, the distance-time relationship is nonlinear. It can be expressed as follows:

\[ 2\pi f_0 t = \psi(x), \]  \hspace{1cm} (4.2)

with \( 2\pi f_0 t \) described in (4.1) and \( \psi(x) \) a nonlinear function of the distance \( x \). This nonlinear function \( \psi(x) \) is modeled as the sum of a linear and a nonlinear contribution:

\[ \psi(x, L) = 2\pi \lambda_0 \left( x + g(x, B) \right), \]  \hspace{1cm} (4.3)

where \( \lambda_0 \) is the spatial frequency expressed in \( \text{m}^{-1} \), and \( g(x, B) \) is the nonlinear phase...
distortion that is modeled as

\[ g(x, B) = \sum_{i=1}^{b} B_i \phi_i(x), \]  

(4.4)

with \( B = [B_1, B_2, \ldots, B_b] \) the vector of unknown phase distortion parameters, \( \phi_i(x) \), \( l = 1,2,\ldots,b \) a set of basis functions, and \( L = [\lambda_0, B] \). In general the nonlinear phase distortion can be expanded by any set of basis functions \( \phi_i(x) \). An appropriate choice of basis functions will limit the number of coefficients \( B_i \) needed to describe the nonlinear function \( g(x, B) \) and hence the time inversions. In this work orthogonal polynomials, constructed through a Gram-Schmidt orthogonalization, and first order splines are chosen as a basis functions. The basis functions, 1st order splines (a) and polynomials (b) are shown in Fig. 4.1 as a function of the number of points.

Fig. 4.1: (a) 1st order splines and (b) polynomials as basis function (from black to light gray: 2nd order to 11th order).

Sampling solid substrates introduces averaging errors (Goodwin et al., 2004). These averaging errors can best be explained by means of a typical example of sampling a solid substrate: drilling. The drilling holes are not infinitesimally small, and consequently samples are not taken at a point in distance, but rather over a volume in distance. Therefore, the continuous proxy signal, which is the signal that is to be reconstructed, will be averaged over the volume of the sample (see Chapter 2). This leads to the following relationship between the continuous proxy signal \( y(x) \) and the measured averaged proxy signal \( \bar{y}(x_n) \):

107
where $\delta$ is the sample width over which the continuous proxy signal is averaged, $x_n = n\Delta$ is distance at sample position $n$, with $n \in \{1,2,\ldots,N\}$, $\Delta$ the distance between two subsequent samples and $x$ the continuous distance. From this equation it is clear that the measured averaged signal $\overline{y}(x_n)$ depends on the sample width $\delta$ over which it is averaged. The significance of the averaging effect, is thus dependent on the volume of the sample but also on the variations of interest of the proxy signal. The larger the relative sample volume, the greater the averaging effect.

From equations (4.1), (4.2), (4.3) and (4.4), the signal model for the continuous proxy signal on a distance axis follows:

$$y(x, \theta) = A_0 + \sum_{k=1}^{H} [A_k \cos(k\psi(x,L)) + A_{k+H} \sin(k\psi(x,L))],$$

(4.6)

where $\theta = [A_0, A_1, \ldots, A_{2H}, L]$ are the unknown model parameters, with $A_0$ the offset, $A_k$ to $A_{k+H}$ the amplitudes of the $k$th harmonic ($k \in \{1, \ldots, H\}$), $H$ the number of signal harmonics and $L = [\lambda_0, B]$.

From (4.5) and (4.6) follows the signal model for the averaged measured proxy signal $\overline{y}(x, \theta)$:

$$\overline{y}(x_n, \theta) = \frac{1}{\delta} \int_{x_n-\frac{\delta}{2}}^{x_n+\frac{\delta}{2}} y(x, \theta) \, dx.$$  

(4.7)

When the samples are very small in comparison with the variations of interest, the signal model (4.7) reduces to the signal model (4.6). Throughout this paper, very small samples are used and, hence, the signal model (4.6) is chosen to describe the measured proxy record.
The unconstrained identification of signal model (4.6) or (4.7) chooses the best parameters irrespective of the order of the data observations. When subsequent time instants are inversed, the difference between these time instants, and as a consequence also the growth rate, is smaller than zero. This is physically impossible. Therefore, the positivity of the growth rate should be imposed.

Once the parametric model (4.6) or (4.7) is identified, the time series can be calculated by calibrating the distance-time relationship (4.2). For this purpose, one needs the time laps between two observations (relative time):

\[
\begin{align*}
(2\pi f_0 t_1 &= \psi(x_1) \\
(2\pi f_0 t_2 &= \psi(x_2))'
\end{align*}
\]  

(4.8)

with \(x_1\) and \(x_2\) the two observations with corresponding relative times \(t_1\) and \(t_2\). From equation (4.8) it follows that the unknown frequency \(f_0\) can be estimated as:

\[
f_0 = \frac{\psi(x_2) - \psi(x_1)}{2\pi(t_2 - t_1)},
\]  

(4.9)

which then allows the time instant to be calculated using (4.2).

\[\textbf{4.2.2 MEASUREMENT MODEL}\]

Since the samples are collected independently and analyzed separately, it is reasonable to postulate a next measurement model:

\[
\tilde{y}(x) = \bar{y}_0(x) + e(x),
\]  

(4.10)

with \(\tilde{y}(x)\) the measured proxy record, \(\bar{y}_0(x)\) the true unknown averaged proxy record, and \(e(x)\) zero mean, independent and identically distributed noise with variance \(\sigma^2\).
4.3 PROBLEM STATEMENT

The aim is to impose the positivity of the growth rate. Hence, an expression for the growth rate \( v(x, B) \) is needed. From (4.2) to (4.4), it follows that:

\[
v(x, B) = \frac{dx}{dt(x,B)} = \frac{1}{\frac{dT(x,B)}{dx}} = \frac{f_0}{\lambda_0(1+\sum_{i=1}^{b}B_i\phi'_i(x))}, \tag{4.11}
\]

where \( f_0, t \) and \( x \) are defined in (4.2), \( \lambda_0 \) is defined in (4.3) and where the last equality uses (4.4) with \( \phi'_i(x) = \frac{d\phi_i(x)}{dx} \).

Imposing a positive growth rate \( (v(x, B) \geq 0) \) it follows from (4.11) that:

\[
h(x, B) = 1 + g'(x, B) > 0. \tag{4.12}
\]

In other words, the distortion parameters \( B_i \) may not take on values that violate inequality (4.12).

Remark: \( 1 + g'(x, B) \) is inversely proportional to \( v(x, B) \). Consequently, when \( h(x, B) \) tends to zero, \( v(x, B) \) goes to infinity. Because \( h(x, B) \) is a smooth function which tends gradually to zero before it becomes negative, it can be concluded that time inversions occur when the growth rate goes to infinity.

4.3.1 GAUSSIAN MAXIMUM LIKELIHOOD COST FUNCTION

Assuming that the measurement noise \( e(x) \) is normally distributed, the cost function \( V(y, \theta) \) that is to be minimized w.r.t. the parameters \( \theta \) is described by:

\[
V(y, \theta) = \sum_{n=1}^{N}(y(x) - y(x, \theta))^2, \tag{4.13}
\]
with $y(x, \theta)$, the signal model (4.6) or (4.7) if averaging is important for the proxy signal, and $y(x)$ the measurements.

## 4.3.2 Nonlinear Optimization with Inequality Constraints

To avoid time inversions in the identified signal model, the inequality constraint (4.12) is added to the cost function (4.13). Hence, the following nonlinear optimization problem with the inequality constraints must be solved:

$$V(y, \theta) = \sum_{n=1}^{N} (y(x) - y(x, \theta))^2 \text{ subject to } h(x, B) = 1 + g'(x, B) > 0. \tag{4.14}$$

This optimization is difficult as it comes to finding starting values (Fletcher, 1991). Therefore, an approximation of (4.14) is proposed in the next section.

## 4.4 2-Step Optimization with Linear Constraint

### 4.4.1 The 2-Step Approach

In the first step an unconstrained optimization of (4.13) is performed. In the second step of the optimization, an inequality constrained is locally imposed on the phase distortion parameters, whenever (4.12) is violated.

The time inversion constrained method described in (de Brauwere et al., 2008) uses this 2-step approach. In this constraint optimization method the constraints becomes active when the difference between two subsequent time instances is smaller than 20% of the average difference between two time instances, and it fixes these values at this minimum. This 20% is an arbitrary bound. The novelty in the method proposed here is that the linear constraints impose a positive growth rate instead of a positive difference between subsequent time instants. The required number of linear constraints is much smaller in the method presented here than in the time inversion constraint method. This is an important improvement, because when the number of constraints exceeds the number of parameters, the method becomes unfeasible. In the sequel of this section the
linear constraints on the distortion parameters are defined and the implementation of the minimization algorithm is explained.

### 4.4.2 Description of the Linear Equality Constraints

When (4.12) is violated, \( h(x, B) \) is smaller than zero in a certain area. In general the constraint will be formulated such that the minimum of \( h(x, B) \) is shifted to a threshold value above zero. Consequently, it is not necessary to shift all the values of \( h(x, B) \) which are smaller than zero, only one value per negative area will be sufficient. The choice of this value depends on the type of basis functions; this will be explained further on in this section. The point in distance corresponding to this value of \( h(x, B) \) will be denoted as \( x_c \). At \( h(x, B) = 0 \) the growth rate is infinitely large, which is physically impossible. Therefore, \( h(x_c, B) \) will be set to a value larger than 0, namely \( \frac{f_0}{\lambda_0 u_{\text{max}}} \), where \( u_{\text{max}} \) represents the maximum growth rate of the archive. Using (4.11), these linear constraints can be written as:

\[
CB = d, \quad (4.15)
\]

in which \( C \) denotes the \( n_c \times b \) matrix with columns \( \phi_i'(x_c), l = 1, 2, \ldots, b \), \( b \) the number of phase distortion parameters, and \( n_c \) the number of constraints. \( B \) is the \( n_\theta \times 1 \) vector with the phase distortion parameters, and \( d = -1 + \frac{f_0}{\lambda_0 u_{\text{max}}} \).

In Fig. 4.2a an example of \( h(x, B) = 1 + g'(x, B) \) is shown for an orthogonal polynomial basis \( \phi_i(x) \). From this figure it can be seen that \( h(x, B) \) is smaller than zero (dashed line) between \( x_a \) and \( x_b \). Shifting the minimum of \( h(x, B) \) in between \( x_a \) and \( x_b \) to a threshold value above zero, will raise the negative area to the threshold value Fig. 4.2b

The constraint for polynomials will be located at the minimum value of the negative area and can hereby be formulated as \( (x_c = x_{\text{min}}) \):

\[
\sum_{l=1}^{b} B_l \phi_l'(x_{\text{min}}) = -1 + \frac{f_0}{\lambda_0 u_{\text{max}}}. \quad (4.16)
\]
CHAPTER 4: PERIODIC TIME SERIES MODELING WITH GUARANTEED POSITIVE GROWTH RATE ESTIMATION

Fig. 4.2: $h(x,B)$ (full line) and the 0-boundary (dashed line) as a function of distance, $x$. a: $h(x,B)$ before the correction b: $h(x,B)$ after the correction.

The minima $h(x_{\text{min}},B)$ are found by calculating the roots of the linear combination of orthogonal polynomials (Rolain et al., 1995).

This approach gives good results for low order polynomials. However, when the order increases, the polynomial oscillates around the target function, due to least squares approximation, causing more time inversions. In Fig. 4.3a, a target function (full black line) and its polynomial fit (16th order) $x + g(x,B)$ (dashed gray line) is shown as a function of distance. In Fig. 4.3b the derivative of the target function (full black line) and the derivative polynomial fit $h(x,B)$ (dashed gray line) is shown as a function of distance.

Fig. 4.3: a: $x + g(x,B)$ (gray dashed line) with polynomials as basis functions and the true $x + g(x)$ (full black line). b: $h(x)$ (full black line) and $h(x,B)$ (dashed gray line).

The derivative of the polynomial fit, $h(x,B)$, is negative in many areas, while the derivative of the target function is only negative in the neighborhood of $x =1$. This
means that modeling errors introduce new time inversions. For this reason a more robust approach is created, with first order splines as basis functions.

In Fig. 4.4a, an example of $h(x, B)$ (full line) is given for first order splines $\phi_i(x)$. From this figure it can be seen that $h(x, B)$ is smaller than zero (dashed line) between $x_a$ and $x_b$. The derivative of $g(x, B)$, $g'(x, B)$ is piecewise constant. As a consequence, the negative areas of $h(x, B)$ are constant. Therefore, shifting one value of $h(x, B)$ in between $x_a$ and $x_b$ to a threshold value above zero, will raise the complete negative area to the threshold value (see Fig. 4.4b). This means that the constraint for these first order splines will be located at any value $x_c$ between $x_a$ and $x_b$, and can hereby be formulated as:

$$
\sum_{l=1}^{b} B_l \phi'_l(x_c) = -1 + \frac{f_0}{\lambda_0 \nu_{max}}. 
$$  \hspace{1cm} (4.17)

Fig. 4.4: $h(x, B)$ (full line) and the 0-boundary (dashed line) as a function of distance, $x$. a: $h(x, B)$ before the correction, b: $h(x, B)$ after the correction.

Remark: First order splines are chosen to simplify the description of the linear equality constraints. Nevertheless, the method can be expanded to higher order splines. In this case, the derivative of $g(x, B)$ would no longer be piecewise constant and the constraint should then be located at the minimum of the negative area as in case of the polynomial basis functions.
4.4.3 IMPLEMENTATION OF THE LINEAR EQUALITY CONSTRAINTS

The linear constrained optimization method can be found in Fletcher (1991). In this work, for numerical stability, (4.15) is solved via the thin singular value decomposition (Golub and Van Loan, 1996) of $C^T = U_1 \Sigma_1 V^T$. The solution of (4.15) contains two contributions, a particular solution of $CB = d$ and the solutions of the homogeneous equation $CB = 0$:

$$B = U_1 \Sigma_1^{-1} V^T d + U_1^t B_{\xi}, \quad (4.18)$$

with $U_1$ an orthogonal $n_\theta \times n_c$-matrix, $\Sigma_1$ a $n_c \times n_c$-matrix and $V$ an orthogonal $n_c \times n_c$-matrix. $U_1^t$ is the orthogonal $n_\theta \times (n_\theta - n_c)$ complement of $U_1$ ($U_1^T U_1^t = 0$), and $B_{\xi}$ are the reduced $(n_\theta - n_c) \times 1$ distortion parameters.

In practice these reduced parameters $B_{\xi}$ can be found by reducing that part of the Jacobian matrix $J$ corresponding to the phase distortion parameters $B$:

$$J_{\xi} = \begin{bmatrix} \frac{\partial y(x, \theta)}{\partial \lambda_0} ; \frac{\partial y(x, \theta)}{\partial \lambda_0} ; \frac{\partial y(x, \theta)}{\partial \lambda_k} ; \frac{\partial y(x, \theta)}{\partial B_{\xi}} U_1^t \end{bmatrix}, \quad (4.19)$$

with $J_{\xi}$ the reduced Jacobian matrix. This implies that all the other parameters follow the normal Levenberg-Marquardt iteration step.

As mentioned in the beginning of this section, the number of constraints is equal to the number of negative areas in $h(x, B)$. The constraints depend on the choice of basis functions $\Phi_i(x)$. In this work orthogonal polynomials and first order splines are chosen as basis functions, because they are easy to manipulate when it comes to defining the constraints on the parameters.
CHAPTER 4: PERIODIC TIME SERIES MODELING WITH GUARANTEED POSITIVE GROWTH RATE ESTIMATION

4.4.4 GENERATION OF STARTING VALUES

The cost function will only be minimized successfully, when a reasonable set of initial values are selected.

In the unconstrained optimization step, the starting values are generated through the non-parametric method described in (De Ridder et al., 2004).

In the constrained optimization step, the phase distortion parameters are locally constrained via (4.15). Consequently, some degrees of freedom are lost. Therefore, it is possible that a larger number of phase distortion parameters are needed for a better approximation of the unconstrained $\psi(x, B)$, in the area where no time inversion occurs.

The optimal number of parameters is chosen by solving the linear least squares problem given by (4.20), while varying the number of phase distortion parameters $b$:

$$V(g, B_c) = \sum_{n=1}^{N} \left( g(x, B_U) - \sum_{l=1}^{p} B_{C_l} \varphi_l(x) \right)^2,$$

with $V(g, B_c)$, the cost function, $B_c$, the constrained phase distortion parameters, $g(x, B_U)$, the signal model for the nonlinear phase distortion in (4.4) constructed with the unconstrained phase distortion parameters $B_U$, $\varphi(x)$ as described in (4.4). The new phase distortion parameters $B_c$ are chosen via the linear constraint in (4.16) or (4.17).

4.4.5 MODEL SELECTION

An aid in choosing the optimal model complexity is the prior knowledge about the signal. However, in order to make an objective choice for the model’s complexity, the minimum description length selection criterion or MDL$_C$ (De Ridder et al., 2005) is used.

4.5 COMPARISON OF THE TIME INVERSION CONSTRAINT METHOD WITH THE POSITIVE GROWTH RATE ESTIMATION METHOD

The novelties in this work are the guaranteed positive growth rate and the reduced number of linear constraints. In this section, the current guaranteed positive growth
rate estimation method is compared with the time inversion constraint method described in (de Brauwere et al., 2008). The comparison is based on the value of the cost function.

The number of linear constraints \( n_c \) in (4.15) is bounded by the number of phase distortion parameters \( b \). If the number of constraints \( n_c \) is larger than the number of phase distortion parameters \( b \), the method becomes unfeasible. The time inversion constraint method produces the largest number of constraints, as in this method \( n_c \) is equal to the number of points for which \( h(x,B) \) is lower than zero, while for the proposed positive growth rate estimation method \( n_c \) equals the number of areas where \( h(x,B) \) is negative.

In a first step of this comparison, the phase distortion parameters are chosen such that the \( b \) is larger than the number of constraints \( n_c \) produced by the time inversion constraint method.

In a second step of this comparison, the phase distortion parameters are chosen such that \( b \) is smaller than \( n_c \).

In all simulations, the harmonic content of the signal model (4.6), the nonlinearly varying phase (4.3), and the phase distortion (4.4), were simultaneously estimated by minimizing (4.13) subject to (4.16) or (4.17) depending on the choice of basis with no noise was added. The number of samples is \( N = 100 \), and the time interval is \([0, 2]\) year. The harmonic content in all simulations is the same, with a spatial frequency \( \lambda_0 = 1 \text{ cm}^{-1} \), and signal harmonics \( A = [0, 1, 1] \). The phase distortion parameters \( B \) determine the time inversions, consequently these parameters were chosen as a function of the simulation example. The phase distortion parameters for all simulations were chosen such that time inversions occur. When the time inversions are not corrected, the simulation parameters are found.

Different choices of basis functions lead to different distortion parameters \( B \) for the description of the same growth rate. For this reason the results for the current method with splines and the results for the current method with polynomials cannot be compared. The time inversion constraint method can be applied to any choice of basis functions.
In the sequel of the chapter subscript T indicates the time inversion constraint method, subscript P and S indicate the positive growth rate estimation method with respectively orthogonal polynomial, and spline basis functions.

\[ 4.5.1 \quad n_{ct} < b \]

### 4.5.1 POLYNOMIAL BASIS FUNCTIONS

For the orthogonal polynomial basis, the phase distortion parameters are: \( B = [0.5476, -0.4943, 0.0348, -0.0175, -0.0599, 0.0485, 0.0458, -0.0816, 0.1359] \). For this choice, the number of active constraints equal, \( n_{cp} = 1 \) and \( n_{ct} = 4 \).

The value of the cost function (see equation 4.13) for the time inversion constraint method, \( V_T \), and the positive growth rate method for polynomials, \( V_P \), are given in Table 4.1. It follows that the positive growth rate estimation method performs better than the time inversion constraint method: the cost function \( V_P \) is a factor 7.781 smaller than \( V_T \). This is confirmed by the time-axis plot in Fig. 4.5a, where the time axis without constraints (full black line), corrected with the current method for polynomials (dashed dark gray line) and corrected with the time inversion constraint method (dotted light gray line) are plotted as a function of distance. Time inversions occur, when the time-distance function is not monotonically increasing. Consequently, it can be seen that both methods correct the time inversions. The positive growth rate estimation method (the dashed dark grey line) follows the uncorrected time axis (full black line) much better than the time inversion constraint method (dotted light grey line).

### 4.5.1.2 SPLINE BASIS FUNCTIONS

For the spline basis, the phase distortion parameters are: \( B = [0.1, -0.1, 0.1, 0.05, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.05] \). For this choice, the number of active constraints equal \( n_{cs} = 1 \) and \( n_{ct} = 8 \).

The value of the cost function (see equation 4.13) of the positive growth rate method for splines, \( V_S \), and the time inversion constraint method, \( V_T \), can be found in Table 4.1. Again, the current method performs better by a factor 8.763. Fig. 4.5b shows the time axis without constraints (full black line), corrected with the current method for splines (dashed dark gray line) and corrected with the time inversion constraint method
(dotted light gray line) plotted as a function of distance. From this figure it follows that both methods correct the time inversions. However, the current method (dashed dark grey line) follows the uncorrected time axis (full black line) much better.

![Figure 4.5](image)

**Fig. 4.5:** The time axis without constraints (full black line), corrected with the current method for polynomials (a) or splines (b) (dashed dark gray line) and corrected with the time inversion constraint method (dotted light gray line) as a function of distance, $x$.

**Table 4.1:** The comparison of the values of the cost functions for the time inversion constraint method and the positive growth rate method for splines and polynomials.

<table>
<thead>
<tr>
<th></th>
<th>$V_F(x, \theta)$</th>
<th>$V_T(x, \theta)$</th>
<th>$V_S(x, \theta)$</th>
<th>$V_T(x, \theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{cT} &lt; n_B$</td>
<td>0.0736</td>
<td>0.5727</td>
<td>0.1277</td>
<td>1.1190</td>
</tr>
<tr>
<td>$n_{cT} &gt; n_B$</td>
<td>0.2351</td>
<td>-</td>
<td>0.1874</td>
<td>-</td>
</tr>
</tbody>
</table>

---

### 4.5.2 $b < n_{cT}$

### 4.5.2.1 POLYNOMIAL BASIS FUNCTIONS

The simulated phase distortion parameters are $B = [2, -0.4, -0.6, 0.4, 0.4, -0.4, -0.2]$. For the positive growth rate method this leads to $n_{cp} = 2$ constraints, while the time inversion constraint method needs $n_{cT} = 11$ constraints. Since, the number of parameters, $b$, is smaller than the number of active constraints $n_{cT}$, the time inversion constraint method becomes unfeasible, while for the current method acceptable results are found.

The value of the cost function (see equation 4.13) of the current method for splines is given in Table 4.1. Fig. 4.6a shows the time axis without constraints (full black line) and the time axis corrected with the current method for polynomials (dashed gray line) as a function of distance. It can be seen that the time axis constructed with the current
method (dashed, grey line) matches quite well the uncorrected time (full, black line) axis.

4.5.2.1.1 SPLINE BASIS FUNCTIONS

In this simulation a spline basis is chosen for the nonlinear distortion function $\psi(x)$, (4.3) with distortion parameters $B = [0.0305, -0.2041, 0.3025, 0.1027]$. For the positive growth rate method this leads to $n_{cS} = 2$ constraint while the time inversion constraint method requires $n_{cT} = 38$ constraints. The number of parameters $b$ is smaller than the number of constraints $n_{cT}$ and, therefore, the time inversion constraint method is again unfeasible.

The value of the cost function (see equation 4.13) for the current method is listed in Table 4.1. Fig. 4.6b shows the time axis without constraints (full black line) and the time axis corrected with the current method for splines (dashed gray line) as a function of distance. Note that the corrected time for the current method (dashed grey line) follows the uncorrected line (full black line) very well.

![Fig. 4.6](image)

**Fig. 4.6**: The time axis functions without constraints (full black line) and the time axis corrected with the current method for polynomials (a) and splines (b) (dashed gray line) as a function of distance, $x$.

4.6 SIMULATION

In order to assess the influence of the noise on the actual parameters, a Monte Carlo simulation is performed (Fishman, 1995).
4.6.1 POLYNOMIAL BASIS FUNCTIONS

Five hundred noisy data sets are generated with the parameters of Section 4.5.1.1. Zero mean was added, such that the signal-to-noise–ratio was 11.6, which is a realistic noise level for environmental records. The estimated time instants can only be compared to the constrained time instants of Section 4.5.1.1. The results for these simulations can be seen in Fig. 4.7. In Fig. 4.7a the difference between the noiseless signal model $y_{P_0}$ (= the signal model of Section 4.5.1.1) and the average of the 500 estimated noisy signal models $\hat{y}_P$ (dashed dark gray line) is plotted as a function of distance with the 95% confidence interval (based on the standard deviation of the sample mean) (dotted light gray line). In Fig. 4.7b the difference between the noiseless time $t_{P_0}$ (= time of Section 4.5.1.1) and the average of the 500 estimated noisy times $\hat{t}_P$ (dashed dark gray line) is plotted as a function of distance, with the 95% confidence interval (based on standard deviation of the sample mean) (dotted light gray line). It is clear from these figures that these differences fall well within the 95% confidence interval. Consequently it can be concluded that the method using a polynomial basis is robust to noise.

![Fig. 4.7: Difference between the models - polynomial method. $y_{P_0} - \hat{y}_P$ (a) or $t_{P_0} - \hat{t}_P$ (b) (dashed dark gray line), and the 95% confidence interval (light gray dotted line) as a function of distance, x.](image)

4.6.2 SPLINES AS BASIS FUNCTIONS

Five hundred noisy data sets are generated with the parameters of Section 4.5.1.2. Zero mean white Gaussian noise was added, such that the signal-to-noise–ratio was 11.6. The estimated parameters can only be compared to the constrained parameters from the
noiseless case of Section 4.5.1.2. The results for the simulations can be found in Fig. 4.8. 
In Fig. 4.8a the difference between the noiseless signal model $y_{S_0}$ (= the signal model of 
Section 4.5.1.2) and the average of the 500 realizations of the Monte Carlo simulation for 
the signal model are $\hat{y}_S$ (dashed dark gray line) is plotted as a function of distance with 
the 95% confidence interval (based on the standard deviation of the sample mean) 
(dotted light gray line). In Fig. 4.8b. the difference between the noiseless time $t_{S_0}$ (= the 
time of Section 4.5.1.2) and the average of the 500 realizations noisy time realizations $\hat{t}_S$ 
(dashed dark gray line) is plotted as a function of distance with the 95% confidence 
interval (based on the standard deviation of the sample mean) (dotted light gray line). 
From this figure it is clear that the differences fall well inside the 95% confidence 
interval. In conclusion, it can be said that the method using the spline basis is robust to 
noise.

![Fig. 4.8: Difference between the models - Splines method: $y_{S_0} - \hat{y}_S$ (a) or $t_{S_0} - \hat{t}_S$ (b) (dashed dark gray line) and the 95% confidence interval (light grey dotted line) as a function of distance, $x$.]

### 4.7 CASE STUDY

#### 4.7.1 POLYNOMIALS: RHIZOPHORA MUCRONATA

The vessel density measured in growth rings of the mangrove tree *Rhizophora mucronata* (see Fig. 2.37) is used to illustrate the polynomial method. A stem disk of the 
*R. mucronata* was collected in 1999 from Makongeni, Kenya (39.46°E, 5.7°S), located 50 
km South of Mombasa. The disk is now part of the xylarium of the Royal Museum for 
Central Africa (RMCA) in Tervuren Belgium.
The Kenyan climate has a bimodal distribution, with a long rainy season from April to July and a short rainy season from October to November. The vessel density is a proxy (i.e., a source of climate information stored in a natural archive) for the rain fall in tropical coastal regions. During the rainy season, earlywood with a high vessel density is produced, and during the dry season, latewood with a low vessel density is produced.

Before measuring the vessel density, the stem disks are dried and their transversal sections are polished (sand grain 100-1200 grit). In Fig. 2.37a, an example of such a stem disk is shown.

The vessel density was measured in adjacent windows along a radial transect on the stem disk from pith to bark. A window width of 300 µm was used, i.e., a very small sample width (300 µm) allowing averaging to be neglected. The vessels were counted at a magnification of 12x using image analysis software (AnalySIS 3.0) and recalculated to find the vessel density per square millimeter (see Fig 2.37b). The data record is 37 samples long. In Fig. 4.9 the vessel density data \( y(x) \) on a distance scale is shown.

![Vessel density in a mangrove trees measured \( y(x) \) with \( \delta = 300 \) µm on a distance scale, \( x \).](image)

The optimal model complexity was chosen as explained in Section 4.4.5. The best model is a signal model (4.6) with \( H = 5 \) harmonics, \( A \), and a distortion model (4.4) with polynomial basis functions, and \( b = 5 \) distortion parameters, \( B \), giving a total of 17 parameters. With these parameters the signal model for the vessel density \( y_{pc}(x, \theta_{pc}) \) (4.6) can be calculated (subscript \( c \) stands for constrained). The time instants are fine-
tuned, such that the frequency of one of the harmonics is equal to 1 year⁻¹. This led to a difference, \( t_{\text{end}} - t_1 = 2.81 \) year. The time, \( t_{\text{PC}}(x, L_{\text{PC}}) \), (4.2), (4.3) can be now be calculated. The number of parameters for the constrained case \( n_{\theta_c} \) is equal to the number of parameters in the unconstrained case \( n_{\theta_u} \), thus the unconstrained parameters are chosen as initial estimates. In Fig. 4.10a the measured vessel density (full black line) and the signal model \( y_{\text{pu}}(x, \theta_{\text{pu}}) \) (dashed gray line) (subscript \( u \) stands for constrained) on the unconstrained time scale \( t_{\text{pu}}(x, L_{\text{pu}}) \) are shown. Time inversions are present when the signal model goes back in time. In Fig. 4.10b the unconstrained time series \( t_{\text{pu}}(x, L_{\text{pu}}) \) is shown as a function of distance. Here the time inversion is clearly visible, as the displayed function is not monotonically increasing.

![Fig. 4.10: (a): The measured vessel density, \( y(x) \) (full black line) and the signal model, \( y_{\text{pu}}(x, \theta) \), (dashed gray line) on an unconstrained time scale, \( t_{\text{pu}}(x, L) \). (b): Time, \( t_{\text{pu}}(x, L) \), as a function of distance, \( x \).](image)

In Fig. 4.11 the results are shown for the linear constrained optimization. In Fig. 4.11a the measured vessel density record (full black line) and the signal model \( y_{\text{pc}}(x, \theta_{\text{pc}}) \) (dashed gray line) are plotted on the constrained time series \( t_{\text{pc}}(x, L_{\text{pc}}) \). This figure shows that the time inversions at the end of the record are removed. This follows also from Fig. 4.11b. were the constrained time \( t_{\text{pc}}(x, L_{\text{pc}}) \) is plotted as a function of distance. The displayed function is monotonically increasing and consequently no time inversions are present.

In Fig. 4.12 the frequency spectrum of the signal model \( y_{\text{pc}}(x, \theta_{\text{pc}}) \) is shown. Taking into account the fundamental frequency \( f_0 = 0.5 \) year⁻¹, and knowing that the number of harmonics \( H = 5 \) for this signal model, peaks are expected at 0.50, 1, 1.5, 2 and 2.5 year⁻¹. In Table 4.2 the magnitude of the amplitudes is shown and their corresponding
standard deviation. From this table it is clear that all magnitudes are significant. Due to the leakage errors of the time to frequency transform, some peaks are shifted. Only the peaks at 1 and 2 will be discussed. The peak at 1 year$^{-1}$ indicates that there is a yearly cycle; this is exactly what is expected, as a full seasonal cycle takes exactly one year. The Kenyan climate has a bimodal distribution, which explains the peak at 2 year$^{-1} = 0.5$ yearly cycle.

![Fig. 4.11](image1.png)

(a): The measured vessel density, $y(x)$, (full black line) and the signal model, $y_{pc}(x, \theta)$, (dashed grey line) on the constrained time scale, $t_{pc}(x, L)$. (b): Time, $t_{pc}(x, L)$, as a function of distance, $x$.

![Fig. 4.12](image2.png)

The frequency spectrum of the signal model for the vessel density $y_{pc}(x, \theta_{pc})$. 

**Fig. 4.11:** (a): The measured vessel density, $y(x)$, (full black line) and the signal model, $y_{pc}(x, \theta)$, (dashed grey line) on the constrained time scale, $t_{pc}(x, L)$. (b): Time, $t_{pc}(x, L)$, as a function of distance, $x$. 

**Fig. 4.12:** The frequency spectrum of the signal model for the vessel density $y_{pc}(x, \theta_{pc})$. 

CHAPTER 4: PERIODIC TIME SERIES MODELING WITH GUARANTEED POSITIVE GROWTH RATE ESTIMATION

Table 4.2: Significance of the amplitudes for the vessel density of mangrove trees.

<table>
<thead>
<tr>
<th>n</th>
<th>$n_f \text{ (year}^{-1}\text{)}$</th>
<th>$\sqrt{A_k^2 + A_{k+H}^2} \pm \text{std}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>2.10 ± 0.20</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.55 ± 0.20</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>2.53 ± 0.19</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.36 ± 0.20</td>
</tr>
<tr>
<td>5</td>
<td>2.5</td>
<td>1.33 ± 0.19</td>
</tr>
</tbody>
</table>

Finally, the linear constrained optimization method with polynomial basis functions was compared with the method with first order spline basis functions. Therefore, a signal model $y_{Su}(x, \theta_{Su})$ with the same number of harmonics $H = 5$ and distortion parameters $b = 5$ was identified with the current method. No time inversion was present in this phase distortion model, consequently the harmonic signal model was estimated without imposing linear constraints. In Fig. 4.13 the results of both signal models $y_{pc}(x, \theta_{pc})$ (dashed gray line) and $y_{Su}(x, \theta_{Su})$ (full black line) are shown as a function of their calculated time axis $t_{pc}(x, L)$ and $t_{Su}(x, L)$ respectively. From this figure it is clear that both methods coincide nicely.

Fig. 4.13: Comparison of the unconstrained splines method ($Su$) to the constrained polynomial method ($Pc$). $y_{Su}(x, \theta)$ (full black line) and $y_{pc}(x, \theta)$ (dashed gray line) plotted as a function of $t_{Su}(x, L)$ and $t_{pc}(x, L)$ respectively.

Nevertheless, this figure doesn’t show whether the difference between both signal models is significant. Therefore, the difference between the signal models $y_{pc}(x, \theta_{pc}) - y_{Su}(x, \theta_{pc})$ (full black line) and the difference between both time series $t_{Sc}(x, L_{Sc}) - t_{pu}(x, L_{pu})$ (full black line) are plotted in Fig. 4.14a and b, respectively, as a function of distance along with the 95% confidence interval (based on the standard deviation of the
linear constrained optimization method with a polynomial basis) (dashed gray line). From this figure we conclude that both models coincide well. However, more than 5% of the points fall outside the 95% confidence interval. This can be explained by residual signal model errors.

![Graph](image)

Fig. 4.14: Difference between the constrained polynomial \((Pc)\) and the unconstrained spline models \((Su)\). \(y_{Pc}(x, \theta_{Pc}) - y_{Su}(x, \theta_{Su})\) (a) or \(t_{Pc}(x, B_{Pc}) - t_{Su}(x, B_{Su})\) (b) (full black line) and 95% confidence interval (dashed gray line).

### 4.7.2 SPLINES: MYTILUS TROSSULUS

Skeletal Mg/Ca ratios of the bivalve *Mytilus trossulus* are used to illustrate the splines method. A *Mytilus trossulus* specimen was collected in 2008, from Disko Bay, Greenland. The specimen was perished at the time of collection, with the erosion on the outer shell indicating it must have remained in such a state for a few months prior to the date of collection.

Greenland has an Artic climate, with very cold winters and relatively mild summers. The Mg/Ca ratio in the shell is a proxy for the sea surface temperature (Klein et al., 1996).

Mg/Ca ratios were measured in skeletal calicite of *Mytilus trossulus*. In calcite Mg/Ca concentrations are positively correlated with temperature (Mucci, 1987), high Mg/Ca values are incorporated in the shell during the summer, and low Mg/Ca values are incorporated during the winter.

Some bivalves, such as *Mytilus trossulus*, cease shell growth or greatly reduce their growth rate during the cold winter months. Dark bands in the shell surface represent these growth rate reductions or growth cessations, and as such, annual growth bands
appear which can be counted to estimate the age of the bivalve (Versteegh et al., 2009).

Prior to the measurements, the shell was imbedded in an epoxy resin (Araldite 2020), a thick section of the shell was cut along the axis of the fastest growth, and was polished. This polished thick section was measured from the ventral margin to umbo. The shell was analyzed using LA-ICP-MS (New wave UP-193, Thermo Scientific XSERIES 2 ICP-MS). The sample size is very small and consequently, averaging can be neglected (Beelaerts et al., 2010). The data record is 217 samples long and covers approximately 13 years. In Fig. 4.15 the Mg/Ca ratio is shown on a distance scale. The optimal model complexity was chosen following Section 4.4.5.

The best model is a signal model with $H = 10$ harmonics, $A$, and a distortion model (3.3) with as polynomial basis containing $b = 9$ distortion parameters, $B$, giving a total of 31 parameters. After estimating these parameters, the time-distance model, $t_{sc}(x, L_{sc})$ (3.2), (3.3) and the signal model for the Mg/Ca ratio, $y_{sc}(x, \theta_{sc})$, (3.6) can be calculated.

The number of parameters for the constrained case $n_{Bc}$ is equal to the number of parameters in the unconstrained case $n_{Bu}$, thus the unconstrained parameters are chosen as initial estimates. In Fig. 4.16a the measured Mg/Ca ratio $y(x)$ (full black line) and the signal model $y_{su}(x, \theta_{su})$ (dashed gray line) are plotted as a function of the unconstrained time scale $t_{su}(x, L)$. Around year 10 time inversions are visible. This can also be seen in Fig. 4.16b where the unconstrained time $t_{su}$ is plotted as a function of distance $x$. Time inversions are present because this plot is not monotonically increasing.

![Fig. 4.15: Mg/Ca ratio in Mytilus trossulus measured $y(x)$ on a distance scale, $x$.]
for $85 \text{ mm} < x < 90 \text{ mm}$.

Consequently, a constrained optimization method is necessary in order to eliminate these time inversions. The signal model (4.6) was calculated, and the time instants were fine-tuned, such that one of the harmonics had a frequency of $1 \text{ year}^{-1}$ (difference, $t_{\text{end}} - t_1 = 13.3 \text{ year}$). Next the time base was calculated. The results of this constrained optimization are shown in Fig. 4.16. In Fig. 4.17a the measured Mg/Ca ratio $y(x)$ (full black line) and the signal model for this ratio $y_{\text{sc}}(x, \theta_{\text{sc}})$ (dashed gray line) are shown as a function of the constrained time series $t_{\text{sc}}(x, L_{\text{sc}})$. The time inversion problem is solved. This is also visible in Fig. 4.17b, where the constrained time axis $t_{\text{sc}}(x, L_{\text{sc}})$ is shown as a function of distance $x$. As it can be seen this is a monotonically increasing function.

Fig. 4.17: (a): The measured Mg/Ca ratio, $y(x)$, (full black line) and the signal model, $y_{\text{sc}}(x, \theta)$ (dashed gray line) on the constrained time scale, $t_{\text{sc}}(x, L)$. (b): Time, $t_{\text{sc}}(x, L)$, as a function of distance, $x$. 
In Fig. 4.18 the frequency spectrum of the signal model for the Mg/Ca ratio $y_{Sc}(x, \theta_{Sc})$ is shown. Knowing that the fundamental frequency $f_0$ for $y_{Sc}(x, \theta_{Sc})$ was 0.14 year$^{-1}$ and that the model consists of $H = 10$ harmonics, peaks are expected at 0.14, 0.29, 0.43, 0.57, 0.71, 0.86, 1, 1.14, 1.29 year$^{-1}$.

Fig. 4.18: The frequency spectrum of the signal model for the Mg/Ca ratio $y_{Sc}(x, \theta_{Sc})$.

In Table 4.3 the magnitude of the amplitudes is shown and there standard deviation.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$nf_0$ (year$^{-1}$)</th>
<th>$\sqrt{A_k^2 + A_{k+H}^2} \pm \text{std}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14</td>
<td>0.36± 0.02</td>
</tr>
<tr>
<td>2</td>
<td>0.29</td>
<td>0.14 ± 0.02</td>
</tr>
<tr>
<td>3</td>
<td>0.43</td>
<td>0.15 ± 0.02</td>
</tr>
<tr>
<td>4</td>
<td>0.57</td>
<td>0.08 ± 0.03</td>
</tr>
<tr>
<td>5</td>
<td>0.71</td>
<td>0.26 ± 0.03</td>
</tr>
<tr>
<td>6</td>
<td>0.86</td>
<td>0.23 ± 0.03</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.12 ± 0.03</td>
</tr>
<tr>
<td>8</td>
<td>1.14</td>
<td>0.12 ± 0.02</td>
</tr>
<tr>
<td>9</td>
<td>1.29</td>
<td>0.14 ± 0.02</td>
</tr>
</tbody>
</table>

From this table it is clear that all magnitudes are significant. Due to leakage problems in the time-frequency transform, most peaks are not clearly visible. Only the peaks which can be explained for now are discussed: the peak at 0.14 year$^{-1}$ and the peak at 0.98 year$^{-1}$. The peak at 0.14 year$^{-1}$ corresponds to a 7.14 year cycle, which is possibly related
to the North Atlantic Oscillation periods (da Costa and de Verdiere, 2002; Hurrell, 1995). This periodicity was also seen in *Artica islandica* shells by (Schöne et al., 2004). The peak at 0.98 year\(^{-1}\) implies a more or less yearly cycle, which is expected as one seasonal cycle has an exact duration of 1 year.

Finally, the linear constrained optimization method with splines was compared with the method using a polynomial basis. Therefore, a signal model \(y_{pu}(x, \theta_{pu})\) with the same number of harmonics \(H = 10\) as for the current method, and a distortion model with \(b = 3\) phase distortion parameters was constructed. No time inversions were present in this phase distortion model, consequently the harmonic signal model was estimated without imposing linear constraints. In Fig. 4.19, the result for both signal models \(y_{pu}(x, \theta_{pu})\) (full black line) and \(y_{sc}(x, \theta_{sc})\) (dashed gray line) were plotted as a function of their respective time axis \(t_{pu}(x, L_{pu})\) and \(t_{sc}(x, L_{sc})\). As can be seen from this figure, the signal models show great similarities. However, the time series do not coincide everywhere. This is due to modeling errors.

![Fig. 4.19: Comparison of the unconstrained polynomial method (Pu) to the constrained splines method (Sc).](image)

In order to look at this problem closer, the difference between the signal models \(y_{sc}(x, \theta_{sc}) - y_{pu}(x, \theta_{pu})\) (full black line) and the difference between both time series \(t_{sc}(x, L_{sc}) - t_{pu}(x, L_{pu})\) (full black line) as a function of distance along with the 95% confidence interval (based on the standard deviation of the linear constrained optimization method with spline as a basis function) (dashed gray line) is shown in Fig. 4.20. In this figure, it can be seen that more than 5% of the points fall outside the
confidence level. This can be explained by residual signal model errors; however, the overall results are satisfactory.

Fig. 4.20: Difference between the constrained spline ($Sc$) models and the unconstrained polynomial ($Pu$) models. $y_{Sc}(x, \theta_{Sc}) - y_{Pu}(x, \theta_{Pu})$ (a) or $t_{Sc}(x, B_{Sc}) - t_{Pu}(x, B_{Pu})$ (b) (full black line) and the 95% confidence interval (dashed gray line) as a function of distance, $x$.

### 4.8 CONCLUSION

A method is presented that eliminates the time inversions in the identification of a harmonic time series from an environmental proxy record. Via linear equality constraints the proposed method imposes locally the positivity of the growth rate.

The novelties in this method are the guaranteed positive growth rate and the reduced number of constraints. The method was developed for polynomials and first order splines basis functions for the nonlinear time-distance model.

The method was compared with the time inversion constraint method described in de (de Brauwere et al., 2008). The current methods with polynomials and with spline basis functions are shown to out-perform that of the time inversion constraint method. Moreover, the time inversion constraint method produced the largest number of constraints, and when the number of constraints is larger than the number of parameters, the method becomes unfeasible.

The noise sensitivity was tested by means of a Monte Carlo simulation and the results fell well within the 95% confidence interval.

Mangrove trees were chosen as a case study for the positive growth rate method with a
polynomials basis. The time inversions were eliminated when the parameters were constrained. Moreover, when compared with the unconstrained method with a spline basis function both methods showed great similarities.

Bivalves were chosen as a case study for the positive growth rate method with spline basis. The time inversions were eliminated when the parameters were constrained. The method was also compared with the unconstrained method with a polynomial basis and both methods showed great similarities.

The method described in this paper is a valuable tool for the elimination of time inversions, which is a commonly occurring problem in the identification of harmonic time series.
5 CASE STUDY: ARCTICA ISLANDICA
CHAPTER 5: CASE STUDY ARCTICA ISLANDICA

5.1 INTRODUCTION

The use of bivalves for the reconstruction of inter-annual and seasonal variation of environmental conditions in mid to high latitudes has become more and more popular amongst paleoclimatologists (e.g. Gillikin et al., 2005; Marsh et al., 1999; Schöne et al., 2004; Tripati et al., 2001). Bivalves are interesting to investigate, because they grow by periodic accretion which makes it easy to date the individual portions of the shell (e.g. Clark II, 1975; Pannella and MacClintock, 1968). Furthermore, bivalves usually form their shell in isotopic equilibrium with the surrounding water and as such they record environmental information in their shell (Epstein et al., 1953; Mook and Vogel, 1968; Wefer and Berger, 1991).

Arctica islandica has a great potential as climate archive because it is an extremely long-lived bivalve and consequently can provide century-long environmental records (Schöne et al., 2005a). A. islandica forms an aragonitic shell and is found at mid to high latitudes (35-70°N) in the Atlantic Ocean as well as in the North Sea and the Baltic Sea, at 6 to 200 m depth (Merrill and Ropes, 1969; Witbaard and Bergman, 2003). A. islandica forms annual growth increments with major growth lines formed during reproduction in late summer/fall and minor growth lines formed during winter growth decline or growth stop in winter (Jones, 1980; Witbaard, 1997).

Oxygen isotopic composition in marine bivalve shells is controlled by the oxygen isotopic composition of the seawater and the water temperature. Carbon isotopes have potential to reconstruct primary productivity (Schöne et al., 2005b). Carbonate samples for the determination of the oxygen or carbon isotopic composition are generally collected by milling a portion from the shell along its major growth axis, with a micro-mill. Usually the volume of these milled samples is kept constant, though shell growth rate decreases with age. At age 20 the yearly increment widths of A. islandica shells decrease to about 1 mm and even lower (Schöne et al., 2004). The smallest drill bits known to the author have a diameter of 20 μm (Schöne et al., 2005a). At a certain point, the width of the sample will represent a large fraction of the yearly increment width and therefore, averaging errors and discretization errors are likely to occur. Schöne et al. (2005c) tried to tackle this problem by using an adaptive sampling strategy in which the drill bit diameter was adapted to the growth rate, but did not take into account the intra-
annual variations in the growth rate.

In this chapter, an *Arctica islandica* shell is analyzed for oxygen isotopes. Averaging is detected and corrected with the methods described in Chapter 2 and Chapter 3. Averaging is assumed to be in one direction only (i.e. the direction of the growth) and the shape of the drilled sample is assumed rectangular, while the true drill holes are cylindrical. This implies that averaging will be overestimated at the corners of the rectangle as already stated in Section 2.2 and exemplified in Fig. 2.5.

The outline of this chapter is as follows. Section 5.2 describes the materials and methods used to perform (i) the age determination, (ii) the stable isotope analysis, (iii) the correction for averaging and (iv) the temperature reconstruction. In Section 5.3, the results for all these steps are presented and discussed.

## 5.2 MATERIALS AND METHOD

The *Arctica islandica* specimen used in this study was collected alive from the Central Oyster Grounds (53.52.64 N and 04.58.62 E; depth of 40m) in the North Sea on the 13th of March, 1991.

After removing the soft tissue, one valve was coated with an epoxy resin (Polypox THV 500) and cut along the axis of maximum growth with a diamond saw and polished.

### 5.2.1 AGE DETERMINATION

The shell was dated by means of a layering technique (Section 1.3.1.2). The annual growth patterns were identified in an acetate peel, which is a surface replica from the shell (Jones, 1980). The acetate peel was mounted on a glass plate and digitized with a camera and analyzed by image analysis software, AxioVision 4.7 Software (Fig. 5.1). The annual growth lines were counted such that calendar years could be assigned to each increment. In addition, the widths of the annual growth bands were measured to obtain an idea of the averaging effect and discretization errors.

### 5.2.2 STABLE ISOTOPE ANALYSIS

Shell powder was milled with a New wave/Mercantec micromill, using a drill bit with a
Fig. 5.1: Photograph of the acetate peel. The umbo (hinge) (right) represents the oldest part of the shell, the ventral margin (left) the most recently accreted shell material. 0.3 mm diameter. The shell was sampled from year 2 to 20, yielding a total of 130 samples. After year 19 the drill bit diameter exceeded the yearly increment width, impeding correction for averaging on from this point (see Section 2.2.1).

The samples were reacted with phosphoric acid in an automated carbonate preparation device (ThermoFinnigan Kiel IV) coupled to a dual inlet Isotope Ratio Mass Spectrometer (ThermoFinnigan Delta XL) at NIOZ (Nederlands Instituut voor Onderzoek der Zee), Texel, The Netherlands. Measurements were conducted using the NBS-19 carbonate standard which has certified values for $\delta^{18}$O and $\delta^{13}$C of -2.2 and +1.95, respectively. The precision on repeat measurements of NBS-19 were better than 0.4 ‰ and 0.04 ‰ for $\delta^{18}$O and $\delta^{13}$C, respectively. Final results are reported relative to the VPDB (Vienna PeeDee Belemnite = the common reference for carbonate samples).

5.2.3 CORRECTION FOR AVERAGING

In Chapter 2, it was shown that the frequency spectrum of the ‘true’ $\delta^{18}$O -signal ($\delta^{18}$O as it was incorporated by the shell), $\mathcal{F}\{y(x_n)\}$, and the frequency spectrum of the measured $\delta^{18}$O -signal, $\mathcal{F}\{	ilde{y}(x_n)\}$, are related as follows:

$$\mathcal{F}\{	ilde{y}(x_n)\} = \text{sinc}\left(\frac{\delta\omega}{2}\right) \mathcal{F}\{y(x_n)\}.$$  

(5.1)
with $y(x_n)$ the measured $\delta^{18}O$-signal, $y(x_n)$ the ‘true’ $\delta^{18}O$-signal, $x_n$ the distance at sample position $n$, $\delta$ the width of the sample, and $\omega$ the spatial angular frequency.

From this equation it follows that averaging is solely dependent on the width of the sample, in this case 0.350 mm. To verify whether averaging errors occur, the frequency spectrum of the measured $\delta^{18}O$ signal was plotted with the sinc-filter ($\text{sinc}\left(\frac{\delta\omega}{2}\right)$). As sampling was non-overlapping, the Nyquist frequency, which is also the highest frequency that can be corrected, was half the sampling frequency ($1/0.8 \text{ mm} = 1.25 \text{ mm}^{-1}$).

Once averaging was ascertained, the averaging errors were eliminated with the parametric method described in Chapter 3. The cost function will only be minimized successfully when a reasonable set of initial values are selected. Therefore, several options were tested using the non-parametric correction described in Chapter 2, a band-limited interpolation described in Section 1.3.2.4, the method described in De Ridder et al. (2004), the method described by de Brauwere et al. (2008), and combination of these methods. The modified minimum description length selection criterion, MDLc, based on De Ridder et al. (2005) was used to select the model complexity.

5.2.4 TEMPERATURE RECONSTRUCTIONS

An equation which captures the relationship between aragonitic bivalves and sea water temperature has been defined by Grossmann and Ku (1986) and rewritten to correct the water values to the SMOW (Standard Mean Ocean Water) scale (Schöne et al., 2005b), the standard against which both oxygen and hydrogen isotopes are reported:

$$T\delta^{18}O(\text{°C}) = 20.60 - 4.34[\delta^{18}O_{\text{aragonite}} - (\delta^{18}O_{\text{seawater}} - 0.20)]$$

(5.2)

with $T\delta^{18}O(\text{°C})$ the temperature reconstruction, $\delta^{18}O_{\text{aragonite}}$ the oxygen isotope value of the shell, and $\delta^{18}O_{\text{seawater}}$ the oxygen isotopes value of the seawater.

Sea water temperatures estimated by means of this equation correspond well ($\pm 1.2$°C)
with instrumental data (Weidman et al., 1994).

Without δ^18O_{\text{seawater}}-values, the sea temperature cannot be reconstructed. Therefore, instrumental data on salinity from 50° to 55° N and 0° to 05° E from a depth of 30 to 50 m. will be used to calculate these values. (ICES. http://www.ices.dk/ocean/data/surface/surface.htm. Accessed 10/09/2010).

In Witbaard et al. (1994) an equation was provided, to obtain δ^18O_{\text{seawater}} from salinity S.

\[
\delta^{18}O_{\text{seawater}} = 0.417S - 14.555
\]  

(5.3)

As such temperature was calculated and compared to instrumental temperature data from 50° to 55° N and 0° to 05° E from a depth of 30 to 50 m. (ICES. http://www.ices.dk/ocean/data/surface/surface.htm. Accessed 10/09/2010).

5.3 RESULTS AND DISCUSSION

5.3.1 AGE DETERMINATION

The shell was collected in March (1991) and, consequently, shows a distinct dark growth line near the ventral margin which corresponds to the spawning break and the corresponding growth stop (Schöne et al., 2005b). A total of 47 years were counted with the year closest to the umbo being 1944.

In Fig 5.2 the annual increments are plotted against ontogenetic age. At the age of 22 the annual increment width had decreased below 1 mm. Since milled samples are on average 0.4 mm apart and since a drill bit of 0.3 mm diameter was used, 2 to less than 1 sample were taken per year on from age 22.

5.3.2 STABLE ISOTOPES

The results for the δ^18O (full black line) are displayed in Fig. 5.3, with their standard deviation. The results for δ18O (full black line) are displayed in Fig. 5.3 and the standard
deviation are the dots above and under the samples. In Appendix 5.A., these results are shown in table-form. The $\delta^{18}O$ maxima reflect the lowest winter temperatures ($\delta^{18}O$ range: $+1.624$ to $+3.786 \, \text{‰}$), while minima reflect highest summer temperatures ($\delta^{18}O$ range: $+0.621$ to $+1.957 \, \text{‰}$). Near the ventral margin, the $\delta^{18}O$-signal shows less extreme values. To provide a better insight in why the $\delta^{18}O$-signal is fading, the major growth lines (gray dashed lines) and the age were plotted (gray numbers) on the $\delta^{18}O$-signal in Fig. 5.3.

It is clear from this figure that the temporal resolution near the ventral margin is very low, with 1 sample per year from age 16 to age 19. This confirms the results from Fig 5.2.

Between age 16 and age 19, a single drill hole averages a complete year. As shown in equation 5.1, averaging behaves like a sinc-function (see Section 2.2). In case a single sample averages a complete year, the corresponding frequency peak coincides with the spatial frequency indicating the first zero of the sinc-function; this is exemplified in Fig. 2.6. In case a drill hole sample averages two, three, four, etc. years, the corresponding frequencies coincide with the frequency indicating the second, third, fourth, etc. zero of the sinc-function respectively. Consequently, the amplitude of the spatial frequency peak corresponding to one year (which is the period of interest here) is multiplied by zero,
leading to sample values which are approximately equal to the average value of the signal.

![Graph showing δ¹⁸O values as a function of distance from the umbo and major growth lines.](image)

Fig. 5.3: The δ¹⁸O-values (full black line and markers) as a function of the distance from the umbo and the major growth lines (= spawning break) (dashed gray lines), the dots are the standard deviation on the measurement.

Before age 16 more than 1 sample per year were obtained. Nevertheless, discretization errors and averaging errors are probably still present in this early part of the record. The Shannon-theorem states that the amplitude and the phase of a signal can be obtained when at least 2 samples per period are measured.

### 5.3.3 CORRECTION FOR AVERAGING

In Fig. 5.4 the spectrum of the measured δ¹⁸O-signal was plotted together with the sinc-function. The Nyquist frequency of the spectrum is 1.25 mm⁻¹. Knowing that the frequency spectrum of the measured δ¹⁸O is the multiplication of this sinc-function with the spectrum of the “true” δ¹⁸O, it is clear that considerable averaging errors are present beyond 0.6 mm⁻¹.

The discretization errors were eliminated by performing a band-limited interpolation. The Regressive Discrete Fourier Transform (see Section 2.2.2) was used because the data set had 5 missing samples and was not harmonic (α = 1.6, measurement window T= 51.59). The distance grid was refined from 130 samples to 738 samples. The results for this band-limited interpolation is shown in Fig. 5.5, where the band-limited...
interpolation (dashed gray line) and the raw data (full black line) are plotted as a function of distance from umbo. Small improvements towards more extreme $\delta^{18}$O-values are visible.

![Graph](image1)

**Fig. 5.4:** The frequency spectrum of the measured $\delta^{18}$O (full black line) plotted with the averaging-filter (sinc-function) (dashed gray line).

![Graph](image2)

**Fig. 5.5:** Band-limited interpolation of the $\delta^{18}$O -signal (dashed gray line) as a function of the interpolated distance, $x_i$, and the measured $\delta^{18}$O-signal (full black line) as a function of distance at sample position, $x_n$.

As a first step, averaging errors are corrected by the non-parametric method described in Chapter 2. Following equation 5.1, the spectrum of the measured signal is corrected by multiplying it with the inverse of the sinc-function. Since the measured $\delta^{18}$O-signal is not harmonic the spectrum was calculated with the Regressive Discrete Fourier
transform (see Section 2.2.2) ($\alpha=1.6$, measurement window $T=51.59$). In Fig. 5.6, the band-limited interpolation of the measured $\delta^{18}O$- (dashed dark gray line) and the $\delta^{18}O$-signal corrected for averaging (dash-dotted light gray line) are shown on the distance scale. Again a small improvement is visible.

![Graph](image)

Fig. 5.6: Corrected for averaging $\delta^{18}O$-values *(dash-dotted light gray line)* and the band-limited interpolation of the measured $\delta^{18}O$-values *(dashed dark gray line)* versus the distance axis, $x_1$.

In order to have an idea of the total difference, the measured $\delta^{18}O$-signal (raw data) *(full black line)* and the completely corrected $\delta^{18}O$-values (band-limited interpolation + correction for averaging) *(dash-dotted gray line)* are plotted in in the distance domain (a) and in the frequency domain (b) in Fig. 5.7.

![Graph](image)

Fig. 5.7: The measured $\delta^{18}O$-values *(full black line)* as a function of the distance at sample position, $x_n$, and the completely corrected $\delta^{18}O$-values *(dashed gray line)* in the distance domain (a) and in the spatial frequency domain (b).

From the frequency plot it follows that the highest frequency components show the
largest correction. This is as expected: the higher the frequency the bigger the averaging effect; see also the shape of the filter (Fig. 2.6).

Next, the parametric method described in Chapter 3 is applied to the data. This data is only displayed from age 4, because no good fit was found for the data from age 2. When the growth rate of the shell is large, it is possible that carbonate does not precipitate under isotopic equilibrium in the shell due to kinetic effects (McConnaughey et al., 1997), although, to our knowledge, this has never been proven in *A. islandica*. It is possible that the model could not fit the first years, because the δ¹⁸O does not reflect the ambient temperature but rather the kinetics of the processes related with shell growth. The last years of the record (age 15-20) were also omitted, because with less than 2 samples per year a reconstruction of the signal is useless. The parametric method fits a harmonic signal model on the data and simultaneously estimates a time base. After fine-tuning the time instants, such that one of the harmonics has a frequency of 1 year⁻¹, the record was 10.6 years old. Several starting values were tested, but only the band-limited interpolation combined with the non-parametric correction for averaging resulted in acceptable results. Before fitting the signal model onto the data, these were detrended (3rd order) to account for slow variations within the measurement window. A signal model with 10 harmonics and 10 time base distortion parameters proved to be the best model (following the minimum description length criterion), leading to a total of 32 parameters. After estimating the parameters, the signal model for the measured data, the corrected signal model and a time base could be calculated. The trend which was subtracted before the parameter estimation was added to the signal models. In Fig. 5.8 a and b, the measured δ¹⁸O-signal (full black line) and the signal model for this measured δ¹⁸O-signal (dashed gray line) is plotted as a function of distance and as a function of the newly constructed time base, respectively. In Fig. 5.8c the distance-time relationship is shown.

The fundamental frequency was equal to 1 year⁻¹, meaning that the slowest variation is one cycle per year. Frequencies lower than 1 year⁻¹ were removed before the estimation of the parameters by detrending the data. Looking at Fig. 5.3, we would expect two years between 47 and 52 mm, though, the time base model, interpreted this as one year. Between age 14 and 15 very few samples were taken. Therefore, this is mistaken for variation within one year. This is shown in Fig. 5.8d where we see a magnification of the
signal on a time scale from age 13 to age 15. An alternative approach, in which the starting values were interpolated such that each year had the same number of samples, was tested, but did not lead to acceptable results.

![Graph showing measured $\delta^{18}O$-signal and signal model](image)

**Fig. 5.8:** The measured $\delta^{18}O$-signal (*full black line*) and the signal model for the measured $\delta^{18}O$-signal (*dashed gray line*) as a function of distance (a) and the newly constructed time base (b); the distance-time relationship (c); and a magnification of the measured $\delta^{18}O$-signal (*full black line*) and the signal model for the measured $\delta^{18}O$-signal (*dashed gray line*) on a time scale from age 13 to age 15.

The newly constructed distance-time relationship can be compared to the distance-time relationship derived from Fig. 5.3, where every major growth line represents a new year. In Fig. 5.9a the result for both distance-time relationships are shown. The black line is the result for the parametric method described in Chapter 3 and the gray markers is the result derived from Fig. 5.3; the markers on the black line represent the points in distance where the gray markers are shown. From this figure it follows that up to age 13 the markers are comparable. From age 13 up the markers diverge, this is due to the fact that the model misinterpreted the data in between age 13 and age 15, as explained in the previous paragraph. This discrepancy can be corrected by using the anchor point method for the last 2 years of the data. The results for this correction are shown in Fig. 5.9b.
5.9b. It becomes clear that the markers are similar now over the whole distance axis. Note that the anchor point method does not account for differences in growth rate within one year; this is also visible in Fig. 5.9 where the black markers are now linearly connected.

Fig. 5.9: The distance-time relationship. (a) calculated with the method described in Chapter 3 (line with bullets), (b) calculated with the method described in Chapter 3 for year 4 to year 13, and the anchor point method for year 13-15 (line with bullets); and calculated with the information from Fig. 5.3 (gray bullets).

In Fig. 5.10 the signal model for the measured $\delta^{18}O$ and the signal model which was corrected for averaging were plotted as a function of the newly constructed time base.

Fig. 5.10: The signal model for the measured $\delta^{18}O$-signal (dashed dark gray line) and the corrected signal model (dashed light gray line) as a function of the newly constructed time base.

The result is very different from the non-parametric method. A test was performed to
verify this result: the signal model for the measured $\delta^{18}$O-signal was subjected to the non-parametric correction, yielding similar results as in Fig. 5.7. Another feature which is now appearing more clearly is a bimodal structure, while temperature in the North Sea an unimodal distribution. This phenomenon is also present, though less pronounced, in the raw $\delta^{18}$O-data (see Fig. 5.3).

5.3.4 TEMPERATURE RECONSTRUCTION

For the calculation of temperature, $\delta^{18}$O$_{\text{seawater}}$ needs to be known. $\delta^{18}$O$_{\text{seawater}}$ values were calculated from the instrumental salinity data, with equation 5.3. Next, temperatures were calculated with equation (5.2). In Fig. 5.11, the instrumental temperature data, $T(\degree\text{C})$ (full light gray line) were compared with the temperature data derived from the shell $\delta^{18}$O-record, $T\delta^{18}$O$_{\text{M}}(\degree\text{C})$ (dashed dark gray line) and the temperature data derived from the corrected $\delta^{18}$O-data, $T\delta^{18}$O$_{\text{C}}(\degree\text{C})$ (full black line).

Fig. 5.11: The instrumental temperature record (full light gray line), the temperature record derived from the measured $\delta^{18}$O (dashed dark gray line) and the temperature record derived from the corrected $\delta^{18}$O (full black line) as a function of the calendar year.

First of all we see that $T\delta^{18}$O$_{\text{M}}(\degree\text{C})$ does not give the expected results ($\pm 1.2 \degree\text{C}$). An answer for this can be found in the sampling strategy that was used. Apparently a drill with a 300 $\mu$m diameter does not give the expected results. Probably discretization errors are at the base of this problem, as samples are only taken every 0.4 mm.

Some discrepancies between the measured sea temperature $T(\degree\text{C})$ and the calculated sea
temperature $T\delta^{18}O_{C}(^\circ C)$ are visible. $T\delta^{18}O_{C}(^\circ C)$, has practically the same range as the real temperature data. Some problems occur in the summer of 1952 and in the winter of 1952-1953, where $T\delta^{18}O_{C}(^\circ C)$ is lower than $T( ^\circ C)$. Moreover, the bimodal structure is also visible in these reconstructed temperature data. The cause for this difference can be subdivided in environmental noise, measurement noise and modeling errors.

Environmental noises are all factors, except temperature, which influence the uptake of $\delta^{18}O$ in the bivalve shell. Some possible environmental noise factors are discussed here. Salinity influences the $\delta^{18}O$ concentration in the water surrounding the bivalve shell and, therefore, temperature reconstructions without a correction for salinity are biased. In (5.2) salinity is accounted for, therefore, salinity is not the reason for the differences. A second possibility is described in foraminifera, where a relationship between oxygen isotopic composition and pH controlled $CO_3^{2-}$ concentration in ambient water has been described by Spero et al. (1997), with $\delta^{18}O$ signature decreasing with increasing [$CO_3^{2-}$]. It is possible that the bimodality is a result of a change in pH in the medium, but this is unlikely to have occurred. A third possibility is described in Owen et al. (2002), where the higher $\delta^{18}O$ values in shell carbonate are due to differences of pH and $CO_2$ concentration between the extrapaleal fluid (i.e. the site where the biomineralization takes place) and the surrounding water. Environmental noise is most likely the largest contributor to the difference between the measured temperature $T( ^\circ C)$ and the reconstructed temperature $T\delta^{18}O_{C}(^\circ C)$.

Measurement errors can be subdivided measurement errors in the temperature measurements and in $\delta^{18}O$ measurements. The temperature measurements can for example help to explain the bimodal structure in the data. Cold water pockets could have surrounded the bivalve shell (oral conversation, Rob Witbaard). These cold water pockets may not have been detected by the instrumental temperature data $T( ^\circ C)$, because these data range from 30 to 50 m depth. The deeper the water is, the scarcer the data becomes; this can bias the result. The instrumental temperature record was not always measured in the same manner. The people who collected the data were either personnel on a cruise ship, or scientists. The data were reduced from mooring data, bottle data or pump data. Nevertheless, it is not likely that the large differences between the measured and the reconstructed data can be explained by differences in measurement strategies. $\delta^{18}O$ measurements were performed using the NBS-19
carbonate standard, which eliminates systematic measurement errors. Furthermore, each sample is the result of an average of four measurements of the same sample and is accompanied by their standard deviation (see Appendix 5.A.). Therefore, it can be concluded that this is not the main contributor to the difference. Finally, errors can be present in the distance data. A Monte Carlo simulation was performed in Chapter 3 and the robustness to this phenomenon was seen. Therefore, these errors are not expected to be the main contributor to this phenomenon.

Modeling errors are all errors introduced by the model. Modeling errors include over- and under-modeling or the use of the MDLc model selection criterion as written in Chapter 3 for signals in which a fundamental frequency is not excited. The use of the MDLc criterion should make the method robust to over- and under-modeling. Nevertheless, it is possible that over-modeling is at the base of the bimodal structure in the data.

Overall, $T\delta^{18}O_{C}(^\circ C)$ gives significantly better results than $T\delta^{18}O_{M}(^\circ C)$.

5.4 CONCLUSION

*A. Islandica* are long lived aragonitic shells and therefore excellent subjects for climate reconstruction.

In this chapter, an *A. islandica* specimen of 47 years old was dated by means of a layering technique. It was seen that the *A. islandica* specimen has a normal growth pattern with a quite sudden drop in growth rate after the age of 22.

The bivalve shell was analyzed for oxygen isotopic composition. The drill bit had a 0.3 mm diameter, which represents a large time section in comparison to the frequency of the variation we wanted to reconstruct, namely 1 year. Averaging and discretization errors were thus present in the record. These averaging errors were eliminated with the parametric method for correcting averaging errors.

Finally, the oxygen isotopic record was converted to a temperature record by means of the modified Grossman and Ku (1986) equation (referencing to the SMOW standard). The results were compared to an instrumental temperature record. The model interpreted recurring bimodal structure in the sinusoidal shape as a double peak which
became more prominent after correction for averaging. This deviation from the
temperature was probably due to the differences in CO$_2$ and pH between the extra-
paleal fluid and the surrounding water. Nevertheless, the full range of the temperature
was reconstructed with the corrected model, while this was not possible with the raw
data.

APPENDIX 5.A.: THE RAW $\delta^{18}$O-VALUES

<table>
<thead>
<tr>
<th>Distance (mm)</th>
<th>d$^{18}$O ($)</th>
<th>std d$^{18}$O</th>
<th>Distance (mm)</th>
<th>d$^{18}$O ($)</th>
<th>std d$^{18}$O</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>2.1890</td>
<td>0.0350</td>
<td>27.30</td>
<td>2.3161</td>
<td>0.0180</td>
</tr>
<tr>
<td>0.35</td>
<td>2.2300</td>
<td>0.0160</td>
<td>27.65</td>
<td>2.8751</td>
<td>0.0230</td>
</tr>
<tr>
<td>0.70</td>
<td>2.2000</td>
<td>0.0190</td>
<td>28.00</td>
<td>3.0251</td>
<td>0.0430</td>
</tr>
<tr>
<td>1.05</td>
<td>1.9560</td>
<td>0.0270</td>
<td>28.35</td>
<td>3.2151</td>
<td>0.0190</td>
</tr>
<tr>
<td>1.40</td>
<td>1.6920</td>
<td>0.0220</td>
<td>28.70</td>
<td>3.2911</td>
<td>0.0290</td>
</tr>
<tr>
<td>1.75</td>
<td>2.6240</td>
<td>0.0320</td>
<td>29.05</td>
<td>3.3531</td>
<td>0.0210</td>
</tr>
<tr>
<td>2.80</td>
<td>2.1270</td>
<td>0.0293</td>
<td>29.40</td>
<td>1.9571</td>
<td>0.0280</td>
</tr>
<tr>
<td>4.20</td>
<td>1.4760</td>
<td>0.2830</td>
<td>29.75</td>
<td>2.1561</td>
<td>0.0280</td>
</tr>
<tr>
<td>4.90</td>
<td>1.4270</td>
<td>0.0240</td>
<td>30.10</td>
<td>2.3031</td>
<td>0.0340</td>
</tr>
<tr>
<td>5.25</td>
<td>1.6240</td>
<td>0.0170</td>
<td>30.45</td>
<td>2.6391</td>
<td>0.0410</td>
</tr>
<tr>
<td>5.60</td>
<td>1.3320</td>
<td>0.0170</td>
<td>30.80</td>
<td>2.7771</td>
<td>0.0230</td>
</tr>
<tr>
<td>5.95</td>
<td>1.7460</td>
<td>0.0230</td>
<td>31.15</td>
<td>2.9381</td>
<td>0.0200</td>
</tr>
<tr>
<td>6.30</td>
<td>1.5070</td>
<td>0.0180</td>
<td>31.50</td>
<td>2.7601</td>
<td>0.0300</td>
</tr>
<tr>
<td>6.65</td>
<td>1.8730</td>
<td>0.0180</td>
<td>31.85</td>
<td>3.1741</td>
<td>0.0120</td>
</tr>
<tr>
<td>7.00</td>
<td>1.7920</td>
<td>0.0180</td>
<td>32.20</td>
<td>3.1851</td>
<td>0.0170</td>
</tr>
<tr>
<td>7.35</td>
<td>2.0890</td>
<td>0.0210</td>
<td>32.55</td>
<td>3.3251</td>
<td>0.0120</td>
</tr>
<tr>
<td>7.70</td>
<td>0.8200</td>
<td>0.0110</td>
<td>32.90</td>
<td>2.7051</td>
<td>0.0100</td>
</tr>
<tr>
<td>8.05</td>
<td>0.8630</td>
<td>0.0210</td>
<td>33.25</td>
<td>1.6771</td>
<td>0.0360</td>
</tr>
<tr>
<td>8.40</td>
<td>2.1480</td>
<td>0.0210</td>
<td>33.60</td>
<td>2.2531</td>
<td>0.0450</td>
</tr>
<tr>
<td>8.75</td>
<td>2.1730</td>
<td>0.0190</td>
<td>33.95</td>
<td>2.5041</td>
<td>0.0220</td>
</tr>
<tr>
<td>9.10</td>
<td>2.0180</td>
<td>0.0220</td>
<td>34.30</td>
<td>2.4821</td>
<td>0.0250</td>
</tr>
<tr>
<td>9.45</td>
<td>1.9020</td>
<td>0.0170</td>
<td>34.65</td>
<td>2.6671</td>
<td>0.0310</td>
</tr>
<tr>
<td>9.80</td>
<td>2.1140</td>
<td>0.0170</td>
<td>35.00</td>
<td>2.6721</td>
<td>0.0250</td>
</tr>
<tr>
<td>10.15</td>
<td>2.1980</td>
<td>0.0250</td>
<td>35.35</td>
<td>2.8541</td>
<td>0.0170</td>
</tr>
<tr>
<td>10.50</td>
<td>2.0790</td>
<td>0.0180</td>
<td>35.70</td>
<td>3.0281</td>
<td>0.0230</td>
</tr>
<tr>
<td>10.85</td>
<td>1.9700</td>
<td>0.0180</td>
<td>36.05</td>
<td>3.1001</td>
<td>0.0270</td>
</tr>
<tr>
<td>11.20</td>
<td>2.5180</td>
<td>0.0250</td>
<td>36.75</td>
<td>2.0791</td>
<td>0.0110</td>
</tr>
<tr>
<td>11.55</td>
<td>1.3330</td>
<td>0.0210</td>
<td>37.10</td>
<td>1.5691</td>
<td>0.0230</td>
</tr>
<tr>
<td>11.90</td>
<td>0.6210</td>
<td>0.0140</td>
<td>37.45</td>
<td>1.6151</td>
<td>0.0140</td>
</tr>
<tr>
<td>12.25</td>
<td>1.6423</td>
<td>0.0430</td>
<td>37.80</td>
<td>1.6351</td>
<td>0.0150</td>
</tr>
<tr>
<td>12.60</td>
<td>1.7323</td>
<td>0.0080</td>
<td>38.15</td>
<td>1.9221</td>
<td>0.0180</td>
</tr>
<tr>
<td>12.95</td>
<td>2.1663</td>
<td>0.0220</td>
<td>38.50</td>
<td>2.1671</td>
<td>0.0310</td>
</tr>
<tr>
<td>13.30</td>
<td>2.7163</td>
<td>0.0240</td>
<td>38.85</td>
<td>2.0656</td>
<td>0.0070</td>
</tr>
<tr>
<td>13.65</td>
<td>3.0563</td>
<td>0.0150</td>
<td>39.20</td>
<td>2.5346</td>
<td>0.0260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>14.00</td>
<td>2.7943</td>
<td>0.0370</td>
<td>39.55</td>
<td>2.9146</td>
<td>0.0390</td>
</tr>
<tr>
<td>14.35</td>
<td>1.6823</td>
<td>0.0330</td>
<td>39.90</td>
<td>2.4316</td>
<td>0.0260</td>
</tr>
<tr>
<td>14.70</td>
<td>1.6743</td>
<td>0.0280</td>
<td>40.25</td>
<td>1.2756</td>
<td>0.0650</td>
</tr>
<tr>
<td>15.05</td>
<td>1.7893</td>
<td>0.0250</td>
<td>40.60</td>
<td>2.1066</td>
<td>0.0340</td>
</tr>
<tr>
<td>15.40</td>
<td>1.9683</td>
<td>0.0180</td>
<td>40.95</td>
<td>2.5196</td>
<td>0.0320</td>
</tr>
<tr>
<td>15.75</td>
<td>2.1423</td>
<td>0.0200</td>
<td>41.30</td>
<td>2.6386</td>
<td>0.0380</td>
</tr>
<tr>
<td>16.10</td>
<td>2.3893</td>
<td>0.0230</td>
<td>41.65</td>
<td>2.7856</td>
<td>0.0130</td>
</tr>
<tr>
<td>16.45</td>
<td>2.6433</td>
<td>0.0150</td>
<td>42.00</td>
<td>2.9316</td>
<td>0.0300</td>
</tr>
<tr>
<td>16.80</td>
<td>2.8083</td>
<td>0.0190</td>
<td>42.35</td>
<td>2.8896</td>
<td>0.0280</td>
</tr>
<tr>
<td>17.15</td>
<td>2.8023</td>
<td>0.0150</td>
<td>42.70</td>
<td>2.1966</td>
<td>0.0330</td>
</tr>
<tr>
<td>17.50</td>
<td>1.8103</td>
<td>0.0290</td>
<td>43.05</td>
<td>1.7076</td>
<td>0.0270</td>
</tr>
<tr>
<td>17.85</td>
<td>1.1603</td>
<td>0.0160</td>
<td>43.40</td>
<td>1.9726</td>
<td>0.0540</td>
</tr>
<tr>
<td>18.20</td>
<td>1.1183</td>
<td>0.0340</td>
<td>43.75</td>
<td>2.1976</td>
<td>0.0190</td>
</tr>
<tr>
<td>18.55</td>
<td>1.6893</td>
<td>0.0130</td>
<td>44.10</td>
<td>2.3786</td>
<td>0.0630</td>
</tr>
<tr>
<td>18.90</td>
<td>2.3953</td>
<td>0.0250</td>
<td>44.45</td>
<td>2.4636</td>
<td>0.0170</td>
</tr>
<tr>
<td>19.25</td>
<td>2.6963</td>
<td>0.0070</td>
<td>44.80</td>
<td>2.9786</td>
<td>0.0120</td>
</tr>
<tr>
<td>19.60</td>
<td>3.0333</td>
<td>0.0240</td>
<td>45.15</td>
<td>3.5616</td>
<td>0.0310</td>
</tr>
<tr>
<td>19.95</td>
<td>2.6363</td>
<td>0.0250</td>
<td>45.50</td>
<td>3.7856</td>
<td>0.0100</td>
</tr>
<tr>
<td>20.30</td>
<td>1.1813</td>
<td>0.0160</td>
<td>45.85</td>
<td>3.3926</td>
<td>0.0300</td>
</tr>
<tr>
<td>20.65</td>
<td>1.5673</td>
<td>0.0330</td>
<td>46.20</td>
<td>3.1146</td>
<td>0.0380</td>
</tr>
<tr>
<td>21.00</td>
<td>2.1453</td>
<td>0.0160</td>
<td>46.55</td>
<td>2.2926</td>
<td>0.0220</td>
</tr>
<tr>
<td>21.35</td>
<td>2.5533</td>
<td>0.0330</td>
<td>46.90</td>
<td>1.3696</td>
<td>0.0210</td>
</tr>
<tr>
<td>21.70</td>
<td>2.8363</td>
<td>0.0410</td>
<td>47.25</td>
<td>1.3806</td>
<td>0.0220</td>
</tr>
<tr>
<td>22.05</td>
<td>3.2533</td>
<td>0.0360</td>
<td>47.60</td>
<td>1.3076</td>
<td>0.0410</td>
</tr>
<tr>
<td>22.40</td>
<td>1.9573</td>
<td>0.0380</td>
<td>47.95</td>
<td>1.3656</td>
<td>0.0420</td>
</tr>
<tr>
<td>22.75</td>
<td>1.2183</td>
<td>0.0150</td>
<td>48.30</td>
<td>1.6786</td>
<td>0.0200</td>
</tr>
<tr>
<td>23.10</td>
<td>1.0313</td>
<td>0.0300</td>
<td>48.65</td>
<td>1.4676</td>
<td>0.0250</td>
</tr>
<tr>
<td>23.45</td>
<td>1.0173</td>
<td>0.0120</td>
<td>49.00</td>
<td>1.6176</td>
<td>0.0210</td>
</tr>
<tr>
<td>23.80</td>
<td>1.2313</td>
<td>0.0110</td>
<td>49.35</td>
<td>1.8846</td>
<td>0.0150</td>
</tr>
<tr>
<td>24.15</td>
<td>2.3283</td>
<td>0.0160</td>
<td>49.70</td>
<td>2.2356</td>
<td>0.0060</td>
</tr>
<tr>
<td>24.85</td>
<td>2.6183</td>
<td>0.0260</td>
<td>50.05</td>
<td>2.2886</td>
<td>0.0310</td>
</tr>
<tr>
<td>25.55</td>
<td>1.8571</td>
<td>0.0450</td>
<td>50.40</td>
<td>2.3486</td>
<td>0.0310</td>
</tr>
<tr>
<td>26.25</td>
<td>2.5841</td>
<td>0.0550</td>
<td>50.75</td>
<td>2.3336</td>
<td>0.0150</td>
</tr>
<tr>
<td>26.60</td>
<td>3.1621</td>
<td>0.0220</td>
<td>51.45</td>
<td>2.2266</td>
<td>0.0430</td>
</tr>
<tr>
<td>26.95</td>
<td>1.5181</td>
<td>0.0420</td>
<td>51.80</td>
<td>2.4906</td>
<td>0.0190</td>
</tr>
</tbody>
</table>
6 CONCLUSIONS AND FUTURE PERSPECTIVES
The aim of this study was to address three problems concerning the preprocessing step of climate reconstruction from environmental data: namely the growth anomaly, the averaging errors, and the discretization errors.

### 6.1 CONCLUSIONS

In Chapter 2, a non-parametric method is presented to correct for averaging errors.

It was seen that the averaging behaved as a low pass filter in the frequency domain. Moreover, the correction implied that the data on the distance scale was multiplied by a sinc-function, which only depends on the sample width. Since the width of the sample is known, these artifacts can be corrected for.

Environmental proxy data are often non-harmonic or unequally spaced, therefore the Optimized Regressive Fourier Transform was applied in order to transform the data to the frequency domain.

Discretization errors are also present in the records, therefore a band-limited interpolation prior to the correction for averaging was performed.

The non-parametric method discussed in this chapter is used as starting values for the parametric signal models.

In Chapter 3 a parametric method is presented that identifies the distance base distortions (growth anomalies) while eliminating the averaging effect. The novelty of the method is the correction for averaging, which results in an increase of the estimated signal harmonic amplitudes. The method is based on the assumption that the proxy record on a time scale is periodic. When the proxy record on a time scale is not harmonic, it can still be approximated arbitrarily well by a Fourier series. Moreover, the harmonic signal model can be replaced by a non-harmonic signal model, without altering the distance-time relationship. However, the generation of starting values can be the bottleneck here.

The optimal complexity of the model was chosen based on prior knowledge and the cost function, which makes the method robust to over- and under-modeling. The numbers of parameters are chosen with the MDLc, allowing the salient features to be extracted from the data record. Moreover, the method separates the stochastic noise from the
significant variations.

In Chapter 4 time inversions, which occurs in the identification of a harmonic time series from an environmental proxy record, are addressed. Via linear equality constraints the proposed method imposes locally the positivity of the growth rate.

The novelties in this method are the guaranteed positive growth rate and the reduced number of constraints in comparison to de Brauwere et al. (2008).

In Chapter 5 all chapters were combined in a case study: oxygen isotopes in *A. islandica*. After the elimination of the averaging errors, the discretization errors and the growth anomalies, the temperature was reconstructed and compared to instrumental temperature data. The temperature reconstruction from the altered data proved to be more similar to the instrumental temperature data than the temperature reconstruction from the raw data, proving the usefulness of this work.

In this thesis some valuable tools were developed for the preprocessing step in climate reconstruction from environmental records. The conclusions hold for other environmental archives and for other sample shapes.

### 6.2 FUTURE PERSPECTIVES

1) In the last Chapter, the limitations of the square sample shape became obvious. Therefore, it would be interesting to expand the method to other sample shapes.

The relationship between the averaged signal model and the ‘true’ signal is as follows for a square sample:

\[
\tilde{y}(x_n) = \frac{1}{\delta} \int_{x_n-\delta/2}^{x_n+\delta/2} y(x) \, dx, \tag{6.1}
\]

where \(\delta\) is the sample width over which the continuous proxy signal is averaged (see Fig 1.16), \(x_n = n\Delta\) is distance at sample position \(n\), with \(n \in \{1, 2, ..., N\}\), \(\Delta\) the distance between two subsequent samples and \(x\) the continuous distance. For non-
uniformly samples signals $\Delta$ varies from sample to sample.

**Hypothesis:** growth lines are parallel to the $z$-axis:

When a circular sample is implemented a weighting factor is needed to describe the sample shape (see Fig. 6.1).

![Diagram of a circle on an x-z grid](image)

Fig. 6.1: A circle on an x-z grid

The equation of the circle is as follows:

$$z^2 = r^2 - x^2$$  \hspace{1cm} (6.2)

with $z$ the $z$-coordinate, $x$ the $x$-coordinate and $r$ the radius of the circle.

This can be translated to our averaging problem as follows. The $z$-coordinate on the circle is half the weightings factor. The radius is half the sample width $\delta$. The $x$ coordinate is the distance from the center of the sample, with the center of the sample being 0.

The relationship between the averaged signal and true signal is then as follows:

$$\bar{y}(x_n) = \frac{8}{\pi \delta^2} \int_{x_n - \frac{\delta}{2}}^{x_n + \frac{\delta}{2}} \left( \frac{\delta}{2} \right)^2 - (x - x_n)^2 y(x) dx$$  \hspace{1cm} (6.3)
This can be expanded to other sample shapes as well.

2) **Hypothesis**: growth lines are concentric circles

Growth lines on shells or trees for example can be approximated by circles. In Fig. 6.2a a bivalve shell is shown, the growth lines, which are the dark lines on the exterior of the shell are approximately circular. In Fig. 6.2b a stem disk of a tree is shown, the growth lines are approximately circular.

![Fig. 6.2: A bivalve: growth rings can be approximated by circles.](image)

In Fig. 6.3 a schematic is shown of circular samples in a shell or bivalve. The relationship between the averaged and the true signal is then as follows:

\[
\bar{y}(x_n) = \frac{8}{\pi \delta^2} \int_{-\delta/2}^{\delta/2} dx \int_0^{\sqrt{\frac{1}{4} - x^2}} y(x, z) dz.
\]  
\[ (6.4) \]

Because the use of Cartesian coordinates requires information we do not possess, a transformation to Polar coordinates is performed (see Fig. 6.3):

\[
\begin{align*}
\left\{ x = r \cos \theta - x_c \right. \\
\left. z = r \sin \theta \right. \\
\end{align*}
\]

\[
\begin{align*}
dx \, dz &= r \, dr \, d\theta \\
y(x, z) &= y(r) \cdot
\end{align*}
\]  
\[ (6.5) \]
Fig. 6.3: Schematic of circular growth lines.

The small circle, which represents the sample, is also transformed into polar coordinates:

\[ x^2 + z^2 = \left( \frac{\delta}{2} \right)^2 \]  \hspace{1cm} (6.6)

\[ (r \cos \theta)^2 + (r \sin \theta)^2 = \left( \frac{\delta}{2} \right)^2 \]  \hspace{1cm} (6.7)

Solving for \( r \):

\[
\begin{align*}
& \Rightarrow \begin{cases} 
  r_1(\theta) = x_c \cos \theta - \sqrt{\left( \frac{\delta}{2} \right)^2 - x_c^2 \sin^2 \theta} \\
  r_2(\theta) = x_c \cos \theta + \sqrt{\left( \frac{\delta}{2} \right)^2 - x_c^2 \sin^2 \theta}
\end{cases} 
\end{align*}
\]  \hspace{1cm} (6.8)
Leading to:

$$\bar{y}(x_n) = \frac{8}{\pi \delta^2} \int_0^{\theta_{\text{max}}} d\theta \int_{r_1(\theta)}^{r_2(\theta)} y(r) r dr,$$

where \( \theta_{\text{max}} = \arcsin \left( \frac{\delta}{2x_c} \right) \).

3) On the web a vast database of paleoclimate data is available: [http://www.ncdc.noaa.gov/paleo/paleo.html](http://www.ncdc.noaa.gov/paleo/paleo.html). It would be interesting to test whether climate reconstructions can be improved if the data is corrected for averaging errors.

Ice cores for example are samples of ice which accumulated from snow fall over hundreds of thousands of years (see 1.2.2). Due to the weight of the ice, the layers are more and more compressed with increasing depth. Consequently, averaging errors could be present in the records.

4) In order to make the method more accessible, it would be interesting to develop a user friendly toolbox. The inputs of this toolbox would be the signal, the two dated observations (see equation 3.9) and the sample width. The output would be a signal which is corrected for averaging.


REFERENCES


REFERENCES


Freitas, P., Clarke, L.J., Kennedy, H., Richardson, C. and Abrantes, F., 2005. Mg/Ca,Sr/Ca, and stable-isotope (delta O-18 and delta C-13) ratio profiles from the fan mussel browser.


Jouzel, J. and EPICA, c.m., 2004. EPICA Dome C Ice Cores Deuterium Data, IGBP PAGES/World Data Center for Paleoclimatology, Data Contribution Series # 2004-038. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.


REFERENCES


Martinson, D., Menke, W. and Stoffa, P., 1982a. An inverse approach to signal correlation
REFERENCES


and Measurement, 45(2): 588-593.


173

REFERENCES


Witbaard, R., 1997. Tree of the sea - the use of internal growth lines in the shell of A. islandica (Bivalvia, Mollusca) for the retrospective assessment of marine environmental change, University of Groningen, 149 pp.


JOURNAL PAPERS

PUBLISHED AND ACCEPTED


IN REVIEW AND IN PREPARATION


Bauwens M., Mouchi V., Versteeg E., Poulin C., **Beelaerts V.,** Barbé K., Schoukens J. and Dehairs F., Introducing dynamics to improve the temperature reconstructions
from high resolution trace element profiles in biogenic carbonates. In preparation.

CONFERENCE PAPERS


ABSTRACTS


Bauwens, M, Poulain, C., Beelaerts, V., Servaes, S., Schoukens, J. and Dehairs, F. A non-linear multi-proxy approach for climate reconstruction based on archaeological


Bauwens, M., Ohlsson, H., Beelaerts, V., Barbé, K., Schoukens, J. and Dehairs, F. Multi
