universität wien

DIPLOMARBEIT

An investigation of the properties of dendrochronological data and dating methods

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Verfasser:	Wolf Naetar
Matrikel-Nummer:	0106267
Studienrichtung:	Mathematik
Betreuer:	Ao. Univ.-Prof. Dipl.-Ing. Dr. Hermann Schichl

Wien, am 8. September 2010

Abstract

Dendrochronology, the dating of wood based on the analysis of tree ring patterns, is of prime importance in archaeology as well as in the environmental sciences. The basis of dendrochronological dating is the correlation of climatic influence factors (such as precipitation and temperature) and tree growth, which results in a similarity of the tree ring structure (usually available in the form of so-called tree ring series, i.e., series of radially measured ring widths) in contemporary trees.

Several methods for automatized dating and subsequent estimation of the statistical significance of the observed similarity have been described in the dendrochronological literature. The primary goal of this work is to present many of the commonly used methods and to improve upon them to allow for reliable dating even if little reference wood is available and the ring width sequences to date are relatively short, as it typically is the case in an archaeological context.

First, the common preprocessing, comparison and significance testing methods (and slight variations of them) used for automatized dating are explained. Free parameters occurring in some of the methods are determined by optimizing the dating performance in a test set. Then, a new method for comparison of tree ring series based on concepts commonly found in machine learning or data mining is sketched. Lastly, the methods presented are tested for their practical performance.

Although details have to be refined and a proper mode of statistical inference remains to be devised, the test performance of the new method is promising, effectively outperforming the traditional methods and their variations.

Acknowledgments

2

Contents

1 Introduction

1.1 Tree rings

Trees are heavily influenced by their environment and readily react to a change of climate. As a tree grows, the characteristics of the newly formed wood depends on the conditions at the time of its growth.

In regions with a temperate climate, one distinguishes between **early** and **late wood**. Early wood, which is formed during the early summer, consists of cells with a short life cycle and grows rapidly. The cells formed in late summer have a much longer life cycle, which results in very slow growth and a much larger density compared with early wood. As late wood is also much darker than early wood, a cyclical pattern following the change of seasons emerges. In the cross-section the tree, this pattern appears in the form of **tree rings** composed of an early and a late wood part.

In general, both density and width (of early and late wood) are of interest for tree ring studies as they are only weakly correlated and thus contain different information (see, e.g., GRIESER [11]). In most cases, however, only total ring widths will be available as archaeological wood is often unsuitable for densitometric measurements.

The density and width of early and late wood growth in a given year depend, among other factors, strongly on the climate. In humid climate the main growth limiting (and therefore reconstructible) factor is temperature, whereas in arid climate mainly the amount of precipitation shapes the composition of the rings. In moderate climate both a play significant role.

When using tree ring data to infer about the climate parameters at the time of growth, one has to consider that climatic occurrences outside the growth period may too have an impact (e.g., frost damage or winter drought). Tree growth can also be heavily influenced by a number of non-climatic factors including infestation by insects, disease, pollution or competition between trees (see, e.g., GRIESER [11]).

The topography of the habitat of a tree can profoundly influence the relationship of climate and tree growth. A tree growing on a large body of earth may, for example, be less susceptible to a dry year than a nearby tree growing on a thin layer of earth over hard rock (cf. SCHWEINGRUBER [6]).

Common sources of tree ring data are drill cores sampled from living trees, subfossil dead wood and wooden archaeological finds. This document mainly concentrates on dendrochronological methods for the use with cedar wood. In contrast to more long living species often used for dendrochronological purposes (e.g., the Bristlecone pine), drill cores gathered from living cedars rarely go back further than 600 years into the past.

In most cases, only ring widths measured in one unknown radial direction will be available. The widths may vary significantly with the measuring direction (see Figure 1.4). A series of radially measured ring widths is called a ring width series or **tree ring series (TRS)**. Measuring is

usually performed by hand using specialized tree ring measuring tables or software-assisted by (manually) marking ring borders on a scanned image.

As mean precipitation and temperature curves, the main growth limiting factors, will typically show a relatively similar progression in a large region, geographically close trees are likely to generate TRS with similar characteristics (e.g., large bumps and drops or periods of fluctuation), which is the basis of tree ring dating.

When a tree is exposed to intense stress early in its growing season, it may form late wood for a short period of time before returning to its regular growing pattern. This temporary transition to late wood growth (which was not induced by a change of seasons) may, at least upon first sight, look identical to a regular year ring. The reverse case, a year ring missing completely, may occur, for instance, in years where the conditions are extraordinarily bad (cf. PIEGORSCH [10]). These **false year rings** and **missing year rings** can complicate subsequent data analysis significantly.
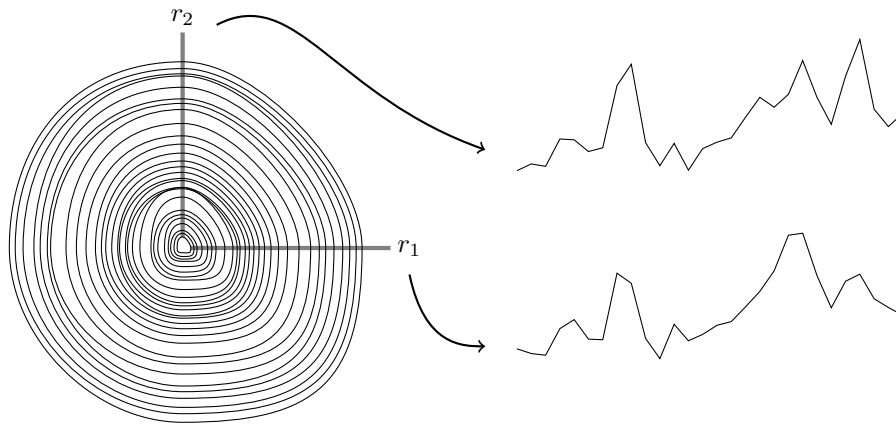
Figure 1.1: Two TRS extracted from the cross-section of a tree. The measured widths depend on the measuring direction.

1.2 Applications

Tree ring data can be used for a broad range of applications.

In **dendroclimatology**, the goal is to determine past climate conditions using tree ring widths or density as a so-called **proxy**. A proxy is a carrier of information strongly dependent on the climate at a given time. Other commonly used proxies are ice cores or sediment cores (see BRADLEY [14]). A function that maps the dendrochronological data (e.g., average tree ring densities and/or widths collected in the region of interest) to an estimate of the climate parameters of interest (e.g., mean summer temperature) is called a **transfer function**. After calibrating and cross-validating the transfer function in a period where both dendrochronological and meteorological data is available, it can then be used to reconstruct the climate parameters in periods where only the dendrochronological data is at hand. A common approach is to assume a linear relationship and fit, for example, by least squares. Other approaches include

exponential transfer functions or transfer functions generated by a neural network (for details, see HELEMA ET AL. [13]).

Dendroarchaeology aims to date wooden archaeological artifacts using dendrochronological techniques, e.g., by comparison with a standard tree ring chronology for the region the artifact is assumed have its origin.

Tree rings can be used to investigate the occurrence of **geomorphological events** such as earthquakes or landslides (cf. SCHWEINGRUBER [8]).

As tree rings can be dated exactly, they can also be used to generate **calibration curves** (which improve ^{14}C-estimated age using dendrochronological information) for **radiocarbon dating** (see SCHWEINGRUBER [9]).

1.3 Dendrochronology

The goal of dendrochronology is the reliable dating of wood based on the observed tree ring patterns, for example, to find the age of wooden artifacts or to construct a database for climatological studies. Tree ring patters are usually given in the form of radial measurements, i.e., only ring widths d_t (with t denoting the year the ring was formed) measured in an unknown radial direction are available.[1] The corresponding time series $d\colon T \to \mathbb{R}^+$ with $d = (d_t)_{t \in T}$ is called a **tree ring series (TRS)**. The index set $T \subseteq \mathbb{Z}$ contains the dating information.

We use ring width information from reference trees of known age for dating. So, given M reference series $d^i = (d_t^i)_{t \in T^i}, i = 1, \ldots, M$ with known T^i one wants to either find the unknown index set T of another series $d = (d_t)_{t \in T}$ with a high degree of confidence or recognize it as undateable with the given data.

While the index set T of d is unknown, we know its size $n = |T|$ and the yearly growths $d_{T_1}, \ldots, d_{T_n}$. Given a position $S \in \mathcal{P}$ (i.e., given a guess S for T), we can shift d into the position S, obtaining the TRS $d|^S = (d|_t^S)_{t \in S}$, where $d|_{S_i}^S := d_{T_i}$ for $i = 1, \ldots, n$. The set $\mathcal{P}$ containing all feasible (and dateable) positions of d may, e.g., be chosen by $\mathcal{P} = \{S \subseteq \mathbb{Z} \mid |S| = |T| \wedge |S \cap \mathcal{T}| \geq L\}$ with L denoting the minimal required overlap (for reliable dating) of the index sets and $\mathcal{T} = \bigcup_{i=1}^{M} T^i$ the years where ring widths of known age are available. As $\mathcal{P}$ as defined above is very large, we have to additionally require $S \in \mathcal{P}$ to consist of consecutive years (i.e., to be of the form $S = \{S_1, S_1 + 1, \ldots, S_n - 1, S_n\}$, which corresponds to the assumption that there are no missing year rings in d), or to contain only a small number of missing year rings to keep the computation time reasonable.

Given the above definitions, we then want to find T as the solution of an optimization problem of the form

$$\hat{T} = \underset{S \in \mathcal{P}}{\arg\max}\ \widetilde{Q}(d|^S, d^1, \ldots, d^M). \tag{1.1}$$

$\widetilde{Q}$, a function which rates how well $d|^S$ agrees with $d^1, \ldots, d^M$ is a function of $M+1$ TRS which

---

[1]In practice the measurements may not be perfectly radial, especially when the series is sampled from an archaeological artifact. Depending on the measuring direction, this may lead to a significant enlargement of rings close to the core.

maps into $\mathbb{R}$ and is invariant under a permutation of its last M arguments. If the maximum is below a certain threshold, i.e., if the agreement is not good enough to be considered significant, we disregard the result (ideally, this will happen only if $T \notin \mathcal{P}$ or if there is a very large geographical distance between the data sources).

As recent wood (of definitely known age, usually gathered from living trees) often does not reach far enough into the past to date a given piece of wood, one has to work one's way into the past by successively adding reliably dated non-recent TRS to the database of series of known age. The dating is based on the overlapping part of the new TRS and the already dated ones. Traditionally, dating is performed by (visual or automatic) comparison with the **mean chronology** $\bar{d}_t = 1/n_t \sum_{i=1}^{M} d_t^i$ generated by the already placed TRS $d^1, \ldots, d^M$ ($n_t \leq M$ denotes the amount of trees which formed a ring in year t). The averaging can be performed in batches or every time a series with a significant degree of similarity is found. Some tree ring chronologies generated this way go as long as 9111 years into the past (e.g., NICOLUSSI ET AL. [12]).
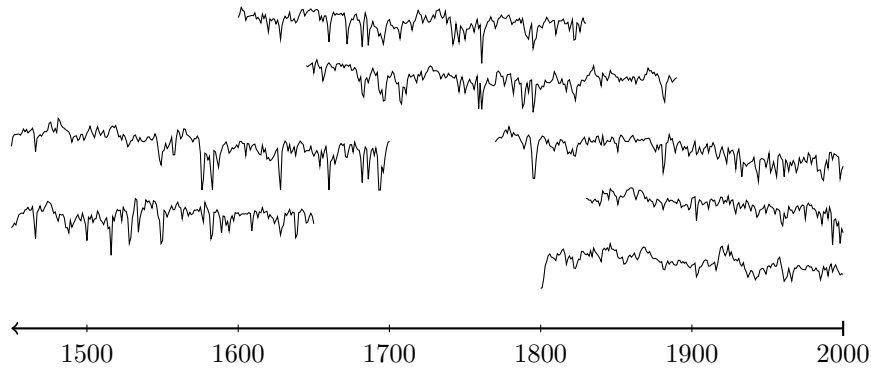
Figure 1.2: Building a tree ring database. The non-recent TRS are placed in the time line by comparison with the already dated ones. For a mean chronology, the vertically aligned widths are averaged (for every year).

In some cases, it may not be possible to date a set of TRS (which is showing a large degree of internal consistency) with the available database. This situation may occur, for instance, when no material has been found for some period of time between the present and the origin of the TRS to date, leading to a gap in the database. In this case, it is still possible to compare the TRS among themselves, leading to a so-called **floating chronology** which still allows the researcher to draw conclusions regarding the relative age of the material and can be anchored in time once the gap is filled.

1.4 Signals contained in a tree ring series

A TRS $d = (d_t)_{t \in T}$ is dependent on several unobserved subseries.

$$d_t = f(c_t, a_t, ld_t, sd_t, e_t) \tag{1.2}$$

where

c_t ... climate signal, e.g., rainfall or average temperatures

a_t ... the age trend in ring growth

ld_t ... local disturbance pulse, e.g., competition with other trees or landslides

sd_t ... stand-wide disturbance pulse, e.g., pest infestation or pollution

e_t ... noise term for other unexplained variation.

This is a slight modification of the additive conceptual model given by COOK [1]. The question of what can be considered as signal and what as noise depends on the application. For dendrochronological purposes, one wishes to extract a signal mostly dependent on c_t, which is assumed to be similar for all trees in a relatively large area. The age trend and regionally confined disturbance pulses are considered noise signals. For other purposes, one may want to accentuate other sub-signals.

1.5 Available data

The dendrochronology group of the **Vienna Institute for Archaeological science (VIAS)** kindly supplied two different sets of ring width data gathered from recent cedar trees.
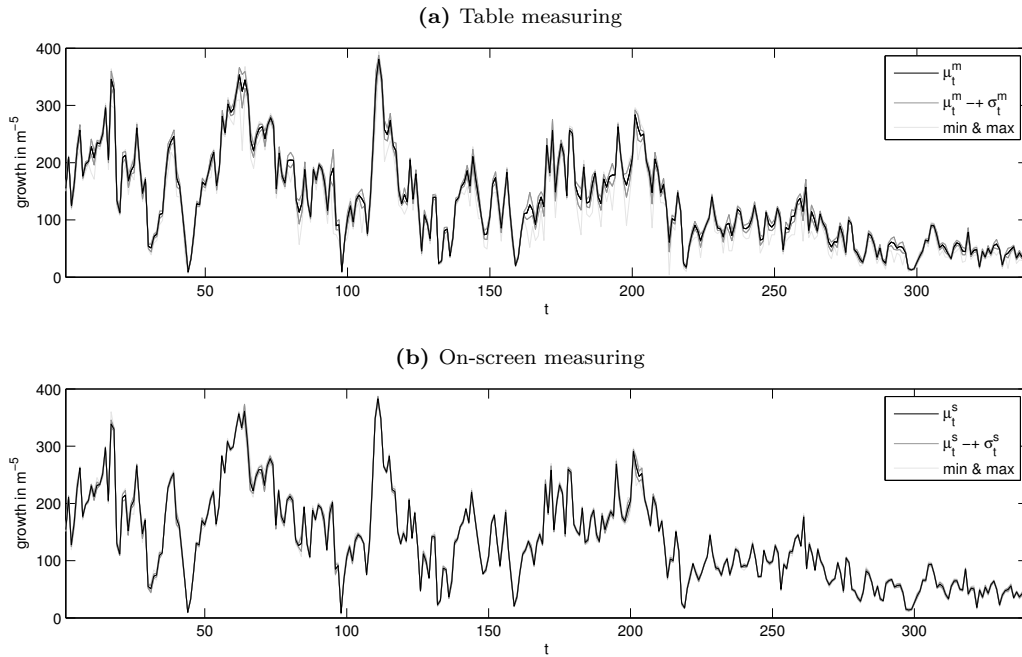
Figure 1.3: Measuring accuracy using table measuring and on-screen measuring. The charts show the average yearly measurements μ_t^m and μ_t^s, their standard deviations σ_t^m and σ_t^s and the maximal and minimal recorded widths (in the lower plot the three graphs coincide for most years). The average standard deviations are 6.89 for table and 2.91 for on-screen measuring.

The first data set is intended for research concerning the consistency of measurements and the variation of TRS within a tree. To determine measuring accuracy and consistency, the

measuring procedure was repeated 14 times using a measuring table and 12 times using on-screen measuring of image scans for a single drill core. Figure 1.3 shows that the measuring is quite accurate as the variation is small compared to the scale of the data. The plots suggest that on-screen measuring is more accurate (disregarding a possible systematic error).

To determine the amount variation of TRS within a tree, ring widths were measured in multiple radial directions on the cross-sections of 3 cedar trees. In Figure 1.4 we can see that the amount of variation clearly depends on the measuring direction. While the differences are significant, the pictures suggest that at least the general shape and prominent features (e.g., large bumps and drops, periods of fluctuation) are similar for all radii.

(a) BA11

(b) BA12

(c) BA5
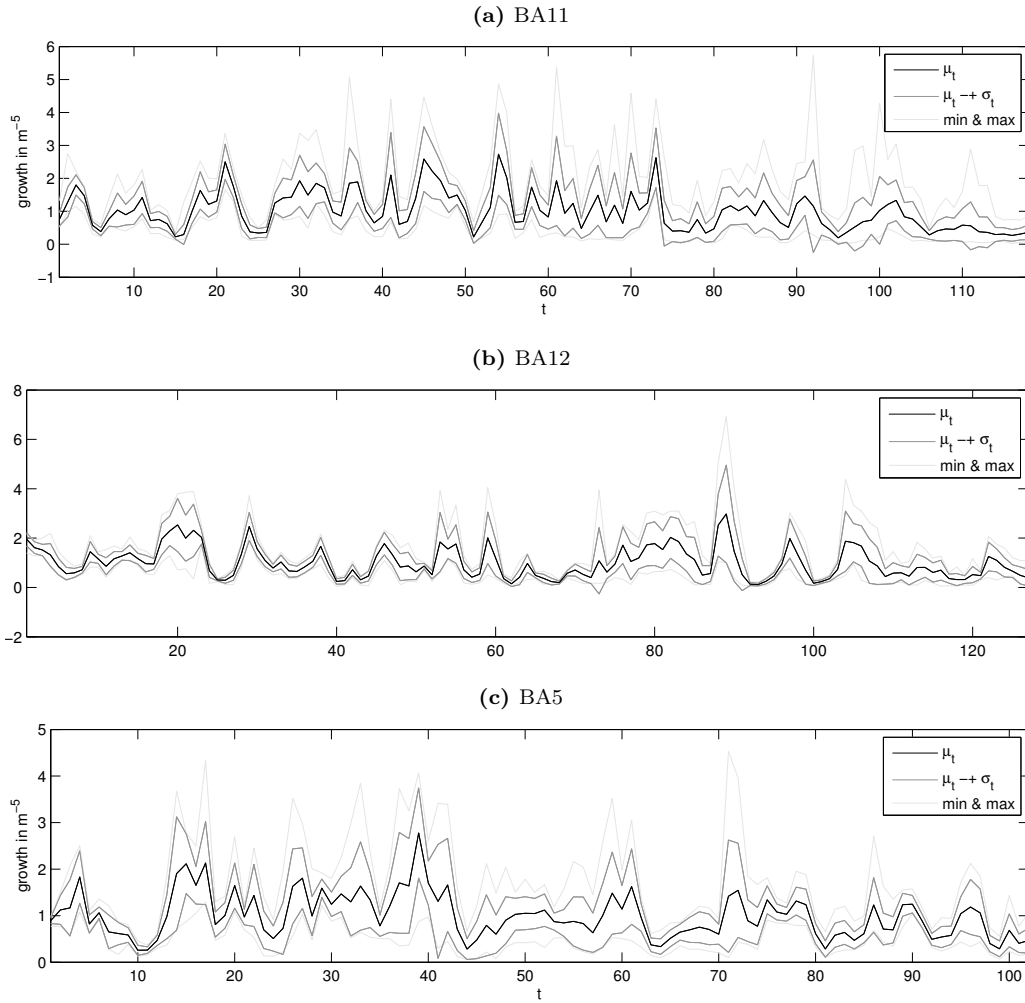
Figure 1.4: Variation of TRS within a tree. Several radii were measured using table measuring for the trees BA11 (12 directions), BA12 (6 directions) and BA5 (9 directions). The resulting TRS were normalized (i.e., their mean was set to 1) to account for a possible ellipsoidal shape of the rings. The charts show the average yearly measurements μ_t, their standard deviations σ_t and the maximal and minimal recorded widths.

The second data set is mainly intended for testing TRS dating algorithms. It contains data from 6 stands in Lebanon, namely

F1 Arz Jaj (52 TRS)

F2 Barouk (31 TRS)

F3 Besheri (58 TRS)

F4 Horsh Ehden (22 TRS)

F5 Maaser (19 TRS)

and 2 stands from Anatolia,

G1 Elmali (41 TRS)

G2 Katrandagi (26 TRS).

As customary, missing year rings which the original editors of the set manually added to the TRS (by comparison of the series) are encoded by zeros. 0.18% of all rings from Lebanese and 0.22% from Anatolian cedars have width 0.

Figure 1.5 shows that the distribution of normalized ring widths is unimodal with a strong positive skew and a few positive outliers. The log-data has a slightly smaller negative skew and a few negative outliers of roughly the same magnitude as the positive outliers of the original data.
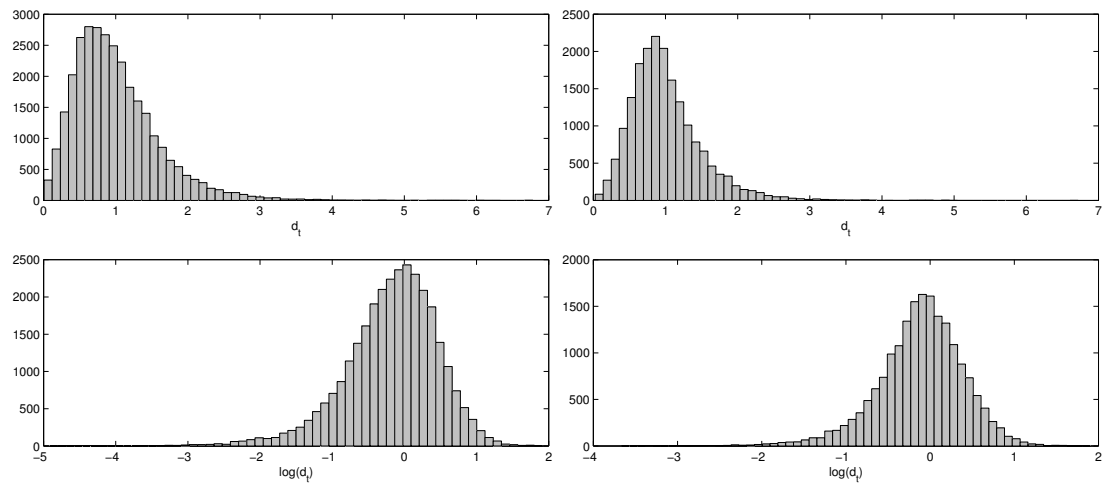
Figure 1.5: Distribution of non-zero ring widths and their natural logarithms in TRS from Lebanese (left column) and Anatolian (right column) stands.

The majority of TRS are about 100-150 years long as can be seen in Figure 1.6. Most long series are from the stands F3 and G1, with the longest of them starting as early as the 14th century. Almost all TRS end in the years 2000 or 2001.

TRS show significant autocorrelation (i.e., correlation with a time-shifted copy of itself) as we
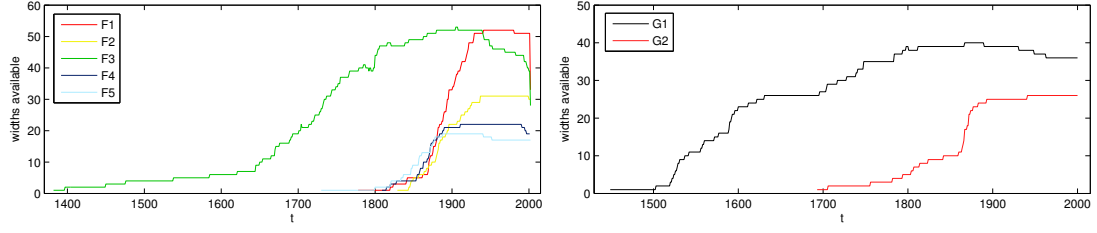
Figure 1.6: Number of ring widths available by year and stand for Lebanese (left) and Anatolian (right) stands.

can see in the figures 1.7 and 1.8. The amount of autocorrelation depends on the trees sampled and may vary. The autocorrelation is mostly caused by the **growth trend** present in the data. When difference detrending (see section 2.2.1) is performed before generating the plots, little to no autocorrelation is left.
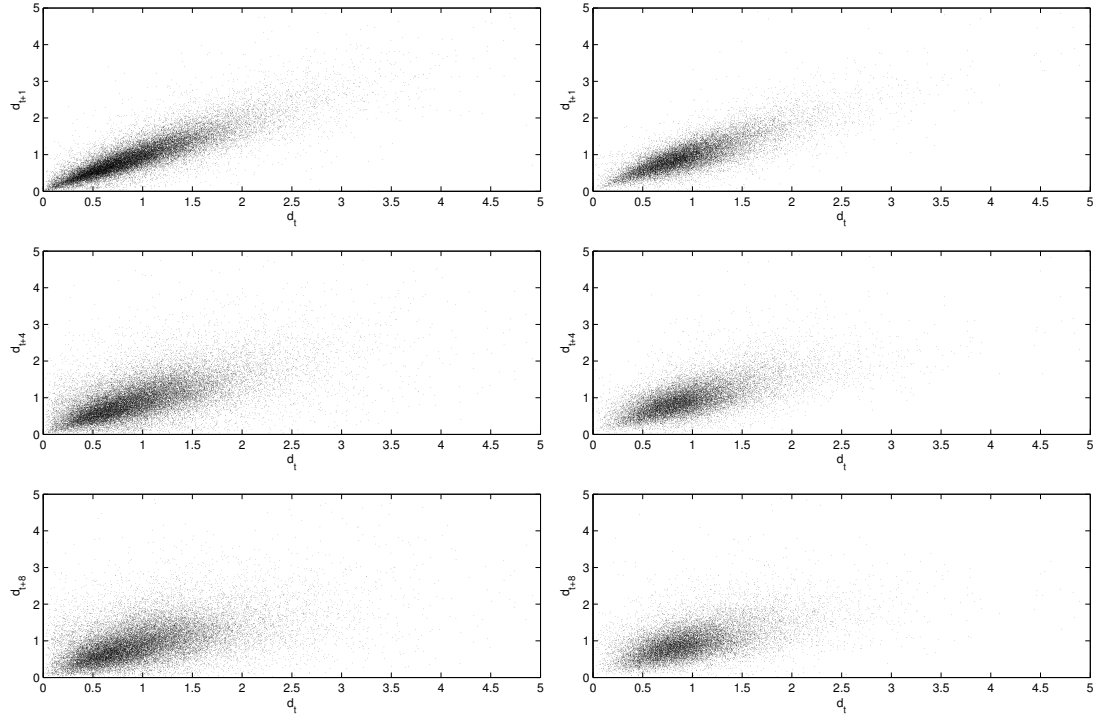
Figure 1.7: Scatter plots of d_t against d_{t+k} for $k = 1, 4, 8$ for Lebanese (left) and Anatolian (right) cedars. All available pairs in the data sets are included in the plots.
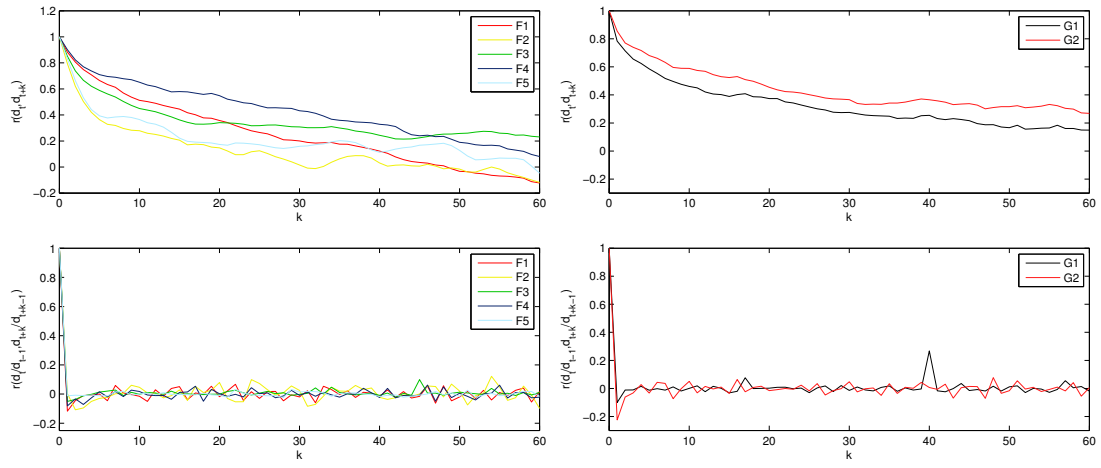
Figure 1.8: Linear correlation of d_t and d_{t+k} as a function of k for Lebanese (left) and Anatolian (right) stands using original and difference detrended ring widths. The correlation is estimated over all available pairs (d_t, d_{t+k}) for a given stand (always $> 10^3$).

2 Detrending and preprocessing

In both dendroclimatology and dendrochronology, it is common practice to **detrend**, i.e., to remove all non-climatic trend components (especially the age trend a) from the TRS, before using them. When using quantitative methods to date this can be of prime importance, as most automatic dating methods don't distinguish between long and short term variation. In fact, they may align the time series solely according to conforming age trends, which may be the dominant signal in a TRS (see Figure 2.1).
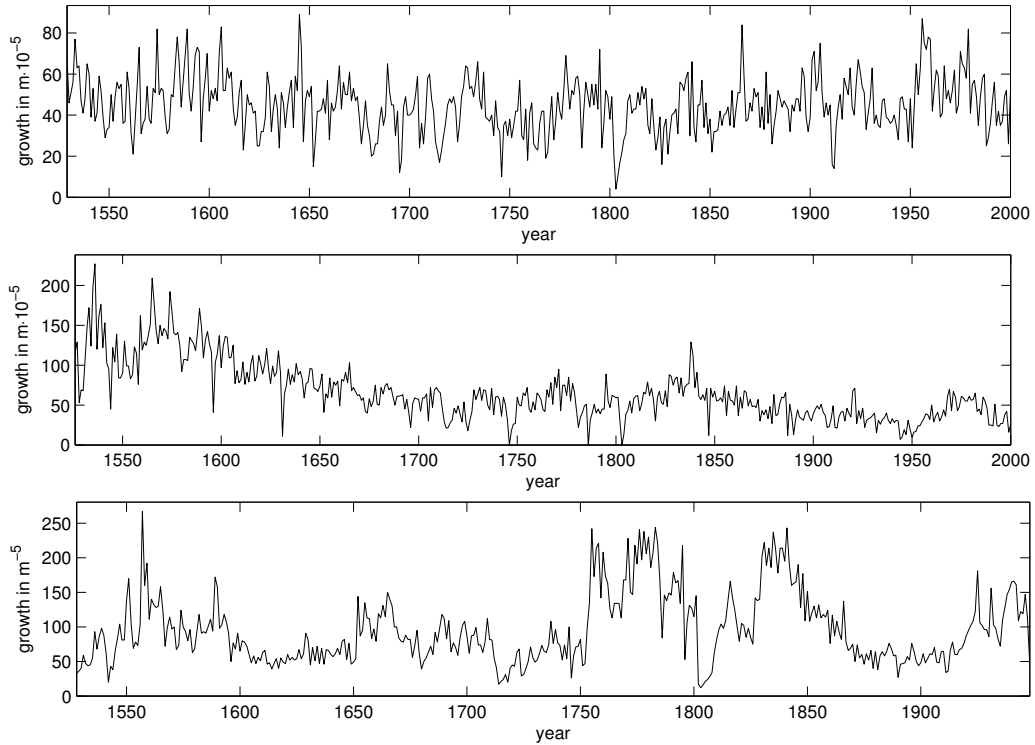
Figure 2.1: Four TRS taken from cedars in Elmali, Anatolia showing different long-term behavior: no clearly visible growth trend, a decreasing trend and erratic behavior.

We may consider the TRS $d = (d_t)_{t \in T}$ as an instance of a stochastic process $D = (D_t)_{t \in T}$ with a well-defined expectation $E(D_t)$ and variance $\text{Var}(D_t)$ for every $t \in T$. Furthermore, D is assumed to be the product of a deterministic term $g = (g_t, t \in T)$, the **growth trend**, which contains the non-climatic factors in tree growth, and a stochastic term $H = (H_t)_{t \in T}$, the **index process** (or **detrended process**), following the climate signal:

$$D_t = g_t H_t \quad \text{for all } t \in T. \tag{2.1}$$

If we make the simplifying assumption that H has constant expectation, we get $E(D_t) = E(g_t H_t) = g_t$ by setting w.l.o.g. $E(H_t) = 1$.

So, to estimate the growth trend g one has to estimate the expectation $E(D_t)$ for every year t given one concrete realization d of the stochastic process D. This can be considered a regression problem which can be solved in several different ways, some of which will be presented in this chapter. On the data level, the estimated trend $\hat{g}(d) = (\hat{g}(d)_t)_{t \in T}$ is removed by

$$h(d)_t = \frac{d_t}{\hat{g}(d)_t} \quad \text{for all } t \in T. \tag{2.2}$$

The estimated **index series** $h = (h_t)_{t \in T}$ is now considered trend-free. Another important issue to note is the fact that TRS are heteroskedastic, i.e., the variance $\text{Var}(D_t)$ may vary strongly with the year t. Empirical evidence (see Figure 2.2) suggests $\text{Var}(D_t) \approx C g_t^2$ for some constant C which may vary for different TRS, i.e., the variance is higher during periods of fast growth. This explains why a multiplicative model is chosen rather than an additive one:

$$\text{Var}(H_t) = \text{Var}\left(\frac{D_t}{g_t}\right) = \frac{\text{Var}(D_t)}{g_t^2} \approx C. \tag{2.3}$$

So after dividing by the trend, the detrended time series H is not only stationary (constant expectation), but also relatively homoskedastic (constant variance), which we can (simplifyingly) expect from the climate signal. For some purposes (e.g., linear correlation, building average series), homoskedasticity is an advantageous property as non-constant variance may lead to years with high variance being weighted more strongly than those with low variance. A constant mean and variance is also a common assumption for hypothesis testing where $\{h_t \mid t \in T\}$ is assumed to be a sample drawn from a random variable with a known distribution.
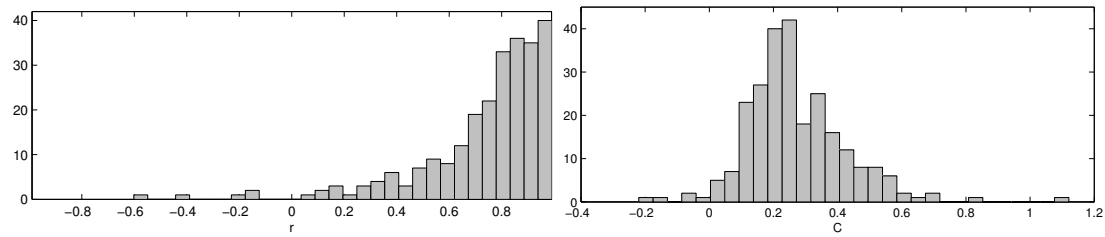
Figure 2.2: The linear correlation r (see section 3.1.1) of the growth trend (estimated by a 25-year moving average) and the standard deviation (estimated by measuring the sample standard deviation in a 25-year window) for 249 unaltered and detrended TRS from Anatolian and Lebanese cedars. On the left histogram, showing the distribution of r for the unaltered TRS, we can see that most trees have a strong correlation between mean and standard deviation. The right histogram shows distribution of the respective proportionality C (the slope of the regression line in the scatter plot mean-std).

When no detrending is performed, it may be practical to standardize TRS (set their mean to 1) by

$$\widetilde{d} = \frac{d}{1/|T| \sum_{t \in T} d_t} \tag{2.4}$$

to simplify visual comparison or to be able to average TRS.[1]

After detrending, SCHWEINGRUBER [6] suggests to perform an additional logarithmic transformation[2] to reduce the variance and remove skewness in the distribution of the index series's data points and make the data's distribution more closely resemble a normal distribution (which is useful for inference, see section 3.3). This may depend on the method of trend estimation.

By using $\log(d_t)$ instead of d_t we could also switch to an additive model where the trend is removed by subtraction (and estimated from $\log(d_t)$). This would likely lead to a more coherent theory (e.g., when applying filters in section 2.2.2 as the frequency decomposition of a time series is additive). COOK [2] states that taking the logarithm may overcompensate the heteroskedacity (as not all TRS are strongly heteroskedastic as seen in Figure 2.2) and advises further research on the matter. We will, however, follow the standard detrending procedure with multiplicative trend removal.

Some authors suggest additional preprocessing measures. In a dendroclimatological context, COOK [1] proposes the use of **autoregressive moving average (ARMA) models** to remove possible autocorrelation and thereby enhance the correlation among TRS used for a mean chronology. Due to the difficulty of implementing ARMA models, which require nonlinear optimization to fit, we will unfortunately have to refrain from experimenting with them.

Many different methods of trend estimation have been suggested, we will describe some of them in the next pages. Some of them require index consisting of consecutive years (i.e., for missing year rings may zeros have to added before detrending). A common method not contained here is the detrending of TRS using cubic splines (which results in a trend similar to a moving average estimation as described in 2.2.2) (see COOK [5]).

2.1 Parametric trend estimation

One way to estimate the trend is to assume a model for g (e.g., g is an affine linear or negative exponential function of t), and then fit the parameters of the model so that it matches the data as good as possible (in some chosen sense). In practice, this is typically done by minimizing some quantity $f(g(\theta), d)$ over θ, the set of parameters of the model.

A popular way to derive estimators of θ for a given data vector d is the **maximum likelihood principle**. If we assume a θ-dependent statistical model for the data (in our case, for instance, by equation (2.1), a model for g and additionally assuming that H_t is independently normal distributed for $t \in T$) we get a probability density function $p(d \mid \theta)$. If we consider p a function

---

[1] When the TRS are detrended, this is not necessary as all common trend estimation methods are homogenous, i.e., they satisfy $\hat{g}(a \cdot d) = a \cdot \hat{g}(d)$ for all $a \in \mathbb{R}$.

[2] In order to take the logarithm and also for other purposes such as difference detrending (see section 2.2.1), we may have to replace zero entries (which usually encode missing year rings) with $\frac{1}{n} \min\limits_{t \in T} d_t$ for a small $n \in \mathbb{N}$ as required. This will retain the general shape of the TRS while avoiding undefined errors.

of θ for fixed d, we get the **likelihood function** $l(\theta \mid d)$, which describes the likelihood the data was generated by the given model with some chosen θ. A maximum likelihood estimator of θ maximizes $l(\theta \mid d)$, i.e.

$$\theta = \arg\min_{\theta} l(\theta \mid d). \tag{2.5}$$

To ease the problem, we will simplifyingly assume that the H_t are independent and identically distributed (which may not necessarily be the case for the real climate signal).

For example, one may choose $\hat{g} = g(\hat{\theta})$ minimizing the variance of the index signal h

$$\theta = \arg\min_{\theta} \sum_{t \in T} \left(\frac{d_t}{g_t(\theta)} - 1 \right)^2. \tag{2.6}$$

If we assume that $H_t \sim \mathcal{N}(1, C)$ for all $t \in T$ we get, using equation (2.1), that

$$P(d_t \mid \theta) = \frac{1}{\sqrt{2\pi C}} \, e^{-\frac{(d_t g_t(\theta)^{-1} - 1)^2}{C^2}} \tag{2.7}$$

is the likelihood of d_t given θ, so by independence, the estimate given by (2.6) is the maximum likelihood estimation of g.

Alternatively, we may take the logarithm on both sides of equation (2.1) and fit $\log(\hat{g})$ to $\log(d)$ with the method of least squares

$$\theta = \arg\min_{\theta} \sum_{t \in T} \left(\log(g_t(\theta)) - \log(d_t) \right)^2. \tag{2.8}$$

Similarly, this is the maximum likelihood estimation of g assuming $\log(H_t) \sim \mathcal{N}(0, \log(C))$ for all $t \in T$ (which of course contradicts the normality assumption of H_t made above). The problem (2.8) may be easier to solve than (2.6) (e.g., for exponential trends, where (2.8) becomes linear).

We can also choose to ignore the heteroskedasticity of d altogether and fit $\hat{g}(\theta)$ by regular least squares fitting

$$\theta = \arg\min_{\theta} \sum_{t \in T} \left(g_t(\theta) - d_t \right)^2. \tag{2.9}$$

For trends linear in θ (e.g., linear or polynomial), (2.9) becomes a linear least squares problem which is easier to solve than the problems (2.6) and (2.8).

2.2 Non-parametric trend estimation

When, as in our case, no reliable parametric model for a regression problem is known, one may choose methods like kernel regression, local regression or spline regression to estimate the regression curve (the growth trend g). These methods express g directly as a function of d instead of estimating a set of parameters.

2.2.1 Differential detrending

Differential detrending is the easiest form of non-parametric trend estimation. Essentially, we assume that D is a **martingale**, a stochastic process which satisfies

$$E(D_t) = d_{t-1}. \tag{2.10}$$

On the data level this means we take $\hat{g}(d)_t = d_{t-1}$, leading to the transformation

$$h(d)_t = \frac{d_t}{d_{t-1}} \text{ for } t \in T_2, \dots, T_{|T|}, \tag{2.11}$$

with the index set being shortened.

2.2.2 Kernel regression

If we assume that g has a low variation locally, i.e., that $g_s \approx g_t$ for s close to t, it is plausible to estimate $g_t = E(d_t)$ by averaging the values d_t in a certain window of time $[t - s, t + s], s \in \mathbb{N}$. This is referred to as a **moving average**. More generally, we may also use a weighted mean

$$\hat{g}_t(d) = \frac{1}{\sum_{s \in T} w_{s-t}} \sum_{s \in T} d_s w_{s-t}. \tag{2.12}$$

If we set

$$\bar{d}_t = \begin{cases} d_t & \text{if } t \in T, \\ 0 & \text{if } t \notin T \end{cases} \tag{2.13}$$

and choose the weight vector w such that

$$w_t = K_\lambda(t) = k\left(\frac{t}{\lambda}\right), \tag{2.14}$$

where $\lambda \in R$ and $k \colon \mathbb{R} \to \mathbb{R}$, we get

$$\hat{g}_t(d) = \frac{1}{\sum_{s \in T} K_\lambda(s-t)} \sum_{s \in T} d_s K_\lambda(s-t) = \frac{1}{\sum_{s \in T} K_\lambda(s-t)} (\bar{d} * K_\lambda) \tag{2.15}$$

with $*$ denoting the discrete convolution. K_λ is called a **kernel**, $\lambda > 0$ the **kernel radius**.[3] For our purposes, k may be arbitrary, but it will typically be symmetrically decreasing and/or have compact support.

We may for example choose a **uniform kernel**

---

[3]The transformation (2.15) is also called a **low-pass filter**, it attenuates the high-frequency content of d while keeping the lower frequencies.

$$\mathrm{k}(x) = \mathbf{1}_{\{|x| \leq 1\}} \tag{2.16}$$

which results in uniform weights and thus a moving average estimation of g. Other possibilities include a **Gaussian kernel** (which removes the high frequencies better than a uniform kernel, resulting in a smoother trend) as recommended by COOK [3].

$$\mathrm{k}(x) = e^{\frac{-x^2}{2}}, \tag{2.17}$$

a **left-side Gaussian kernel**

$$\mathrm{k}(x) = e^{\frac{-x^2}{2}} \, \mathbf{1}_{\{x \leq 0\}}, \tag{2.18}$$

or the **Epanechnikov kernel**

$$\mathrm{k}(x) : (1 - x^2) \, \mathbf{1}_{\{|x| \leq 1\}}. \tag{2.19}$$
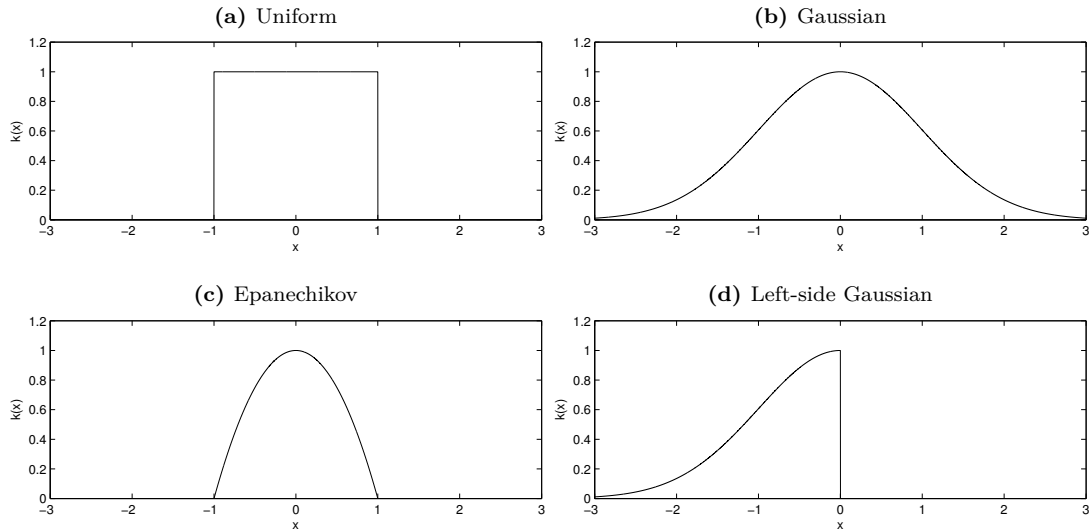
Figure 2.3: Graphs of the kernels used in equations (2.15) and (2.20).

The kernel radius λ determines the frequency range filtered by formula (2.15). The extreme cases are $\hat{g}(d) \approx d$ if $\lambda \approx 0$ (everything but the lowest frequencies is filtered in (2.2)) and $\hat{g}(d) \approx \frac{1}{n} \sum_{t \in T} d_t$ if $\lambda \gg 1$ ((2.2) merely normalizes).

In the index series h obtained by (2.2), a large λ leads to an index series where most of the low-frequency variation of d is still present while a smaller λ will remove more of the signal d and keep only the higher frequencies.

In practice, choosing λ may be difficult. If one chooses to remove most of the low-frequency variation, a lot of climatic variation may be removed along with the age trend, which can

be especially counterproductive for dendroclimatogical studies. Keeping the low-frequency variation, on the other hand, may result in a signal where the high-amplitude variation is still mostly due to non-climatic influences.

COOK [4] suggests choosing λ by maximizing the **signal-to-noise ratio** which (using his definitions) is equivalent to maximizing the average linear correlation within a training set of λ-kernel regression detrended TRS. We take a slightly different approach and try to maximize the amount of TRS with maximal correlation in the correct position when testing a fixed number of positions (see section 3.2).

2.2.3 Local regression

For local regression or **LOWESS** (locally weighted scatter plot smoothing), we assume that the growth trend g is locally well-approximated by a low-degree polynomial, i.e., that $g_s \approx p(s-t, \theta_t)$ for s close to t, where p is a low-degree polynomial with coefficients θ_t. We estimate θ_t by locally (using only data points d_s with s close to t fitting the polynomial $p(\theta)$ to the TRS d. The trend for the year t is then taken to be $p(0, \theta_t)$.

For ease of calculation, we will use weighted linear least squares fitting (which erroneously assumes homoskedacity, see 2.1). We will also restrict ourselves to quadratic polynomials $p(s, a, b, c) = as^2 + bs + c$. Then we get

$$(\hat{a}_t, \hat{b}_t, \hat{c}_t)^T = \operatorname*{arg\,min}_{a_t, b_t, c_t} \sum_{s \in T} K_\lambda(s-t)(a_t(s-t)^2 + b_t(s-t) + c_t - d_s))^2 \qquad (2.20)$$

$$\hat{g}_t(d) = p(0, \hat{a}_t, \hat{b}_t, \hat{c}_t) = \hat{c}_t. \qquad (2.21)$$

We obtained the weights from a kernel K_λ with radius λ as we did in section 2.2.2. The same kernels can be used. The radius λ again determines how tightly the growth trend matches the data.

To solve the problem (2.20) we solve the normal equations for weighted linear least squares

$$\left(A^T W A\right) \hat{\theta}_t = A^T W \vec{d}, \qquad (2.22)$$

where $\hat{\theta}_t = (a_t, b_t, c_t)^T$ are the parameters to be estimated, $\vec{d} = (d_{T_1}, \ldots, d_{T_{|T|}})^T \in \mathbb{R}^{|T|}$ the TRS d as a vector, $W \in \mathbb{R}^{|T| \times |T|}$ a diagonal matrix with $W_{ii} = K_\lambda(T_i - t)$ and $A \in \mathbb{R}^{|T| \times 3}$ with $A_{i:} = ((T_i - t)^2, (T_i - t), 1)$.

In practice d, W and A are reduced in size for faster computation by using only data points which are weighted non-zero.

3 Traditional dating methods

In this chapter, we will describe some commonly used dating methods (and slight variations of them). We start with the special case where only two TRS at a time are compared using **matching methods**. The more general case where $M > 1$ TRS of known age are available is handled by averaging the already dated TRS to a **mean chronology** which the TRS to date are compared with.[1]

3.1 Matching methods

We begin by describing methods which can compare only two TRS at a time, i.e., which estimate the unknown index set T of a TRS $d = (d_t, t \in T)$ given only one TRS $d^1 = (d_t^1, t \in T^1)$ with T^1 known and $T \cap T^1 \neq 0$.

A function $m \colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, where $N \in \mathbb{N}$ is arbitrary, is used to evaluate the likeness of two time series in a given relative position. We call m a **matching function**. An often-used example is the linear correlation coefficient (see section 3.1.1).

Using m, we can now calculate the m-score of the detrended time series $h(d|^S)$ and $h(d^1)$ with

$$\widetilde{m}(h(d|^S), h(d^1)) = m \left(\{h(d|^S)_t \mid t \in S \cap T^1\}, \{h(d^1)_t \mid t \in S \cap T^1\} \right), \tag{3.1}$$

where $\widetilde{m}$ is the function of two time series that matches their overlapping years with m, and estimate T by taking the position with maximal m-score

$$\hat{T} = \arg\max_{S \in \mathcal{P}} \widetilde{m}(h(d|^S), h(d^1)). \tag{3.2}$$

If the maximum is below some previously determined threshold (see section 3.3), we conclude $T \notin \mathcal{P}$. The following sections describe several matching methods and evaluate their efficiency and properties.

3.1.1 Linear correlation coefficient

The **linear correlation coefficient** (also called the **product-moment** or **Pearson correlation coefficient**) measures the level of linear dependency of x and y. For $x, y \in \mathbb{R}^N$ it is defined by

---

[1] Experimentation with other methods to combine the matching scores (e.g., mean/median/maximal scores) showed no significant improvement and were thus left out of the text. Due to the strong correlation of the scores it is also harder to devise a "clean" notion of statistical significance of the average/median/maximal scores.

$$r(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\, \text{Var}(y)}}, \tag{3.3}$$

where

$$\text{Var}(x) = \text{Cov}(x, x). \tag{3.4}$$

$$\text{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}), \tag{3.5}$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \ \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{3.6}$$

are the sample covariance and the respective sample means and variances (normalizing with $N-1$ obviously gives the same coefficient r).

If we fit a line through the data points (x_i, y_i) using simple linear regression

$$\hat{y}_i = a + b x_i \tag{3.7}$$

and define the **mean squared error** and **mean squared residual** by

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{3.8}$$

$$\text{MSR} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 \tag{3.9}$$

we get

$$1 - r(x, y)^2 = \frac{\text{MSE}}{\text{Var}(y)} \tag{3.10}$$

$$r(x, y)^2 = \frac{\text{MSR}}{\text{Var}(y)}. \tag{3.11}$$

For a proof, see, for instance, WOLFRAM MATHWORLD [15]. These results give us two interpretations of the correlation coefficient, equation (3.10) states that $1 - r(x, y)^2$ is equal to the average squared error in the linear regression of x on y, measured in units of $\text{Var}(y)$. Equation (3.11) shows that r^2 is the fraction of $\text{Var}(y)$ accounted for by the variance of the estimated points on the regression line.

As $\text{Cov}(x, y) = \frac{1}{N}(x - \bar{x})^T(y - \bar{y})$ we also get

$$r(x, y) = \frac{(x - \bar{x})^T(y - \bar{y})}{|x - \bar{x}||y - \bar{y}|} = \cos \angle (x - \bar{x}, y - \bar{y}), \tag{3.12}$$

i.e., the correlation coefficient is the cosine of the planar angle between the vectors $x - \bar{x}$ and $y - \bar{y}$.

Special cases are

$$r(x,y) = \operatorname{sgn} a \Leftrightarrow y = ax + b \tag{3.13}$$

$$r(x,y) = 0 \Leftrightarrow \operatorname{Cov}(x,y) = 0 \tag{3.14}$$

For two correctly aligned and detrended TRS x, y we expect $r(x,y) \gg 0$ assuming that they depend linearly on the same climate signal with little noise (local disturbance signals).

3.1.2 Non-parametric correlation

The product-moment correlation coefficient (3.3) measures the how strongly two random variables (or their data vectors) are **linearly** correlated. The **Spearman** and **Kendall rank correlation coefficients**, on the other hand, make no assumption about the particular way the random variables are correlated. Both will have maximal absolute value when the relationship is monotonic, i.e., when $y_i = F(x_i)$ for any monotonic function F. Like the linear correlation coefficient, they are expected to be around 0 for uncorrelated data. In our context, non-parametric correlation allows for trees to show different responses (linear, quadratic, etc.) to the climate signal, as long as they are are monotonically increasing the correlation will be maximal.

The Spearman rank correlation coefficient ρ computes the linear correlation of the **ranks** of $x, y \in \mathbb{R}^N$,

$$\rho(x,y) = r\left(\operatorname{rank}(x), \operatorname{rank}(y)\right), \tag{3.15}$$

where r is the product-moment correlation coefficient (3.3) and $\operatorname{rank}\colon \mathbb{R}^N \to \mathbb{N}^N$ the rank operator satisfying $\operatorname{rank}(x)_i = n$ if x_i is the nth-smallest number in x. More precisely, if $\pi \in S_n$ is the permutation that orders x by ascending magnitude, we have

$$\operatorname{rank}(x)_i = \frac{1}{|J|} \sum_{J = \{j | x_j = x_i\}} \pi^{-1}(j). \tag{3.16}$$

Tied ranks are averaged. If there are no tied ranks, one may equivalently calculate ρ with

$$\rho(x,y) = 1 - \frac{6 \sum_{i=1}^{N} \delta_i^2}{N(N^2 - 1)}, \tag{3.17}$$

with $\delta_i = \operatorname{rank}(x)_i - \operatorname{rank}(y)_i$ denoting the rank differences between x and y.

Another measure of the degree of correspondence of $\operatorname{rank}(x)$ and $\operatorname{rank}(y)$ is Kendall's rank correlation coefficient

$$\tau(x,y) = \frac{n_c - n_d}{\frac{1}{2} N(N - 1)}. \tag{3.18}$$

$n_c = |\{(i,j) \mid \operatorname{sgn} x_i - x_j = \operatorname{sgn} y_i - y_j\}|$ and $n_c = |\{(i,j) \mid \operatorname{sgn} x_i - x_j \neq \operatorname{sgn} y_i - y_j\}|$ are the numbers of concordant and discordant pairs and $\frac{1}{2} N(N-1) = \binom{N}{N-1}$ the total number of pairs, so τ measures the difference of the fractions of concordant and discordant pairs.

3.1.3 Sign test

The **sign test** G measures the fraction of years where both TRS lie on the same side of their trend, or equivalently, the fraction of years where the detrended TRS are both > 1 or < 1.

Formally, for detrended (or normalized) $x, y \in \mathbb{R}^N$, we have

$$
\mathrm{G}(x, y) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} |\operatorname{sgn}(x_i - 1) + \operatorname{sgn}(y_i - 1)|. \tag{3.19}
$$

The special case using difference detrending (see section 2.2.1) is called Gleichläufigkeit in German (see SCHWEINGRUBER [7]).

3.2 Method and parameter selection

To find out which detrending-matching (and possibly a logarithmic transformation) combination to use, we evaluate the performance of formula (3.2), detrending with

KRG Kernel regression, Gaussian kernel
KRL Kernel regression, left-side Gaussian kernel
KRU Kernel regression, uniform kernel (moving average)
KRE Kernel regression, Epanechikov kernel
LRG Local regression, Gaussian kernel
LRL Local regression, left-side Gaussian kernel
LRU Local regression, uniform kernel
LRE Local regression, Epanechikov kernel
DIF Difference detrending
NOD No detrending.

and matching with the methods from section 3.1

r   Linear correlation
r $\circ$ log Linear correlation after taking the logarithm
ρ Spearman correlation
τ Kendall correlation
G   Sign test

in a simplified scenario. We take all possible pairs of TRS in the 3 test sets (the stands the TRS were sampled from being disregarded)

L1 Lebanon 1900-1999 (8128 pairs)

L2 Lebanon 1830-1899 (1711 pairs)

A1 Anatolia 1870-1969 (1431 pairs)

and determine the percentage of correctly dated pairs PCD where the correct position $S = T$ was found.

We assume index sets consisting of consecutive years and check matching scores ± 30 years off the correct position, i.e., require a minimal overlap of 70 for **L1**, **A1** and 40 for **L2**.

Although this is an artificial scenario (TRS of equal size, large required overlap) with no significance testing, it should nevertheless give us a rough idea of how the different methods perform in practice.

For detrending methods which require a bandwidth parameter we will test a broad range of values to check how strongly the dating performance depends on λ and to determine the optimal magnitude.

The graphs in figures 3.1 (**r**), 3.2 (**r** $\circ$ log), 3.3 (ρ), 3.4 (τ) and 3.5 (**G**) show the measured PCD values in the 3 test sets as a function of λ for each of the 8 λ-dependent detrending methods and the given matching method. The tables below show the maximal PCD values and their arguments, i.e., the λ values the maxima were obtained with.

Figure 3.6 shows the PCD values obtained in all 3 test sets for all possible combinations of the detrending and matching methods above in a cross-validation setup. Here, for every detrending/matching combination, λ was determined by maximizing PCD on one of the 3 test sets while the evaluation was performed on one of the others, giving a more realistic estimation of the method's performance in practice.

(a) KRG

(b) LRG

(c) KRL

(d) LRG

(e) KRU

(f) LRU

(g) KRE

(h) LRE

$\max_\lambda$ PCD
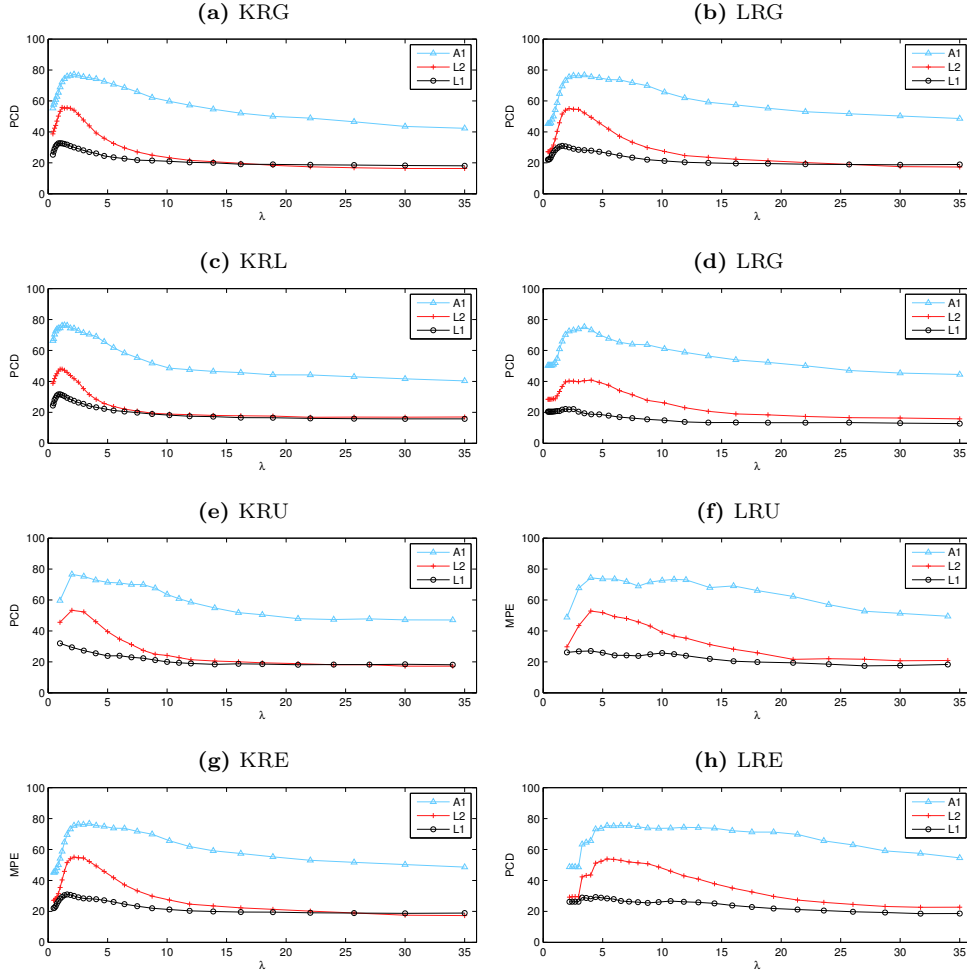
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**
L1	32.7	31.7	32	32.5	30.9	22.1	27	29.3
L2	55.8	47.9	53.4	55.1	55.1	40.8	52.8	53.9
A1	77.1	76.3	76.6	76.9	76.7	75.3	74.4	75.5

$\arg\max_\lambda$ PCD

	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**
L1	1.01	1.01	1	2.09	1.6	1.87	4	4.4
L2	1.18	1.18	2	2.67	2.18	4.04	4	5.37
A1	2.18	1.37	2	3.85	3.46	3.46	4	6.54

Figure 3.1: Position finding results using the correlation coefficient r with kernel and local regression detrending and various kernels. The charts show PCD values in dependence of the kernel bandwidth parameter λ. The tables show the extremal values in the charts with 3 digits of accuracy.
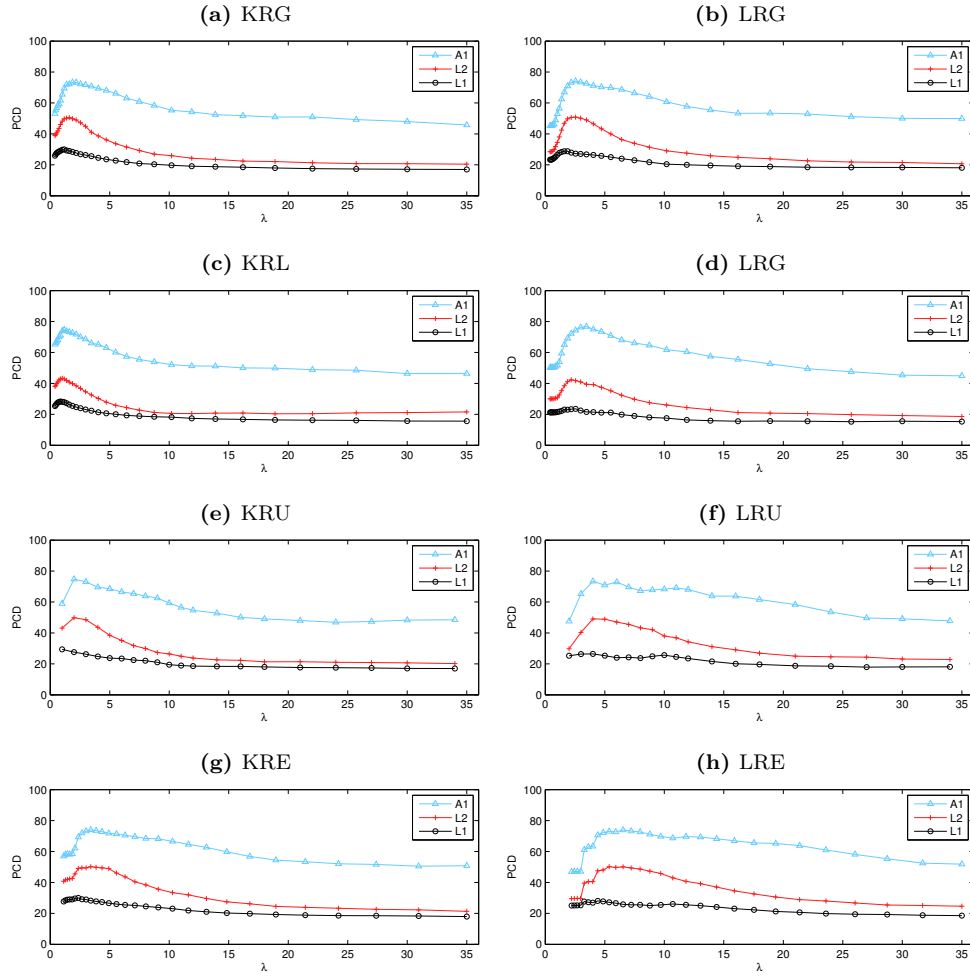
			$\max_\lambda$ PCD					
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**
L1	29.9	28.3	29.4	30	28.8	23.6	26.5	28.1
L2	50.5	43.3	49.9	50.2	50.8	42.3	49.1	50.2
A1	73.3	74.7	74.8	74	74.1	76.7	73.4	73.9

			$\arg\max_\lambda$ PCD					
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**
L1	1.18	0.865	1	2.36	1.6	2.54	4	4.4
L2	1.6	1.01	2	3.41	2.54	2.18	4	5.37
A1	1.87	1.18	2	3.41	2.54	3.46	4	6.54

Figure 3.2: Position finding results using the correlation coefficient r with kernel and local regression detrending and various kernels. The logarithm was taken after detrending. The charts show PCD values in dependence of the bandwidth parameter λ. The tables show the extremal values in the charts with 3 digits of accuracy.

(a) KRG   **(b)** LRG

(c) KRL   **(d)** LRG

(e) KRU   **(f)** LRU

(g) KRE   **(h)** LRE

$\max_\lambda$ PCD
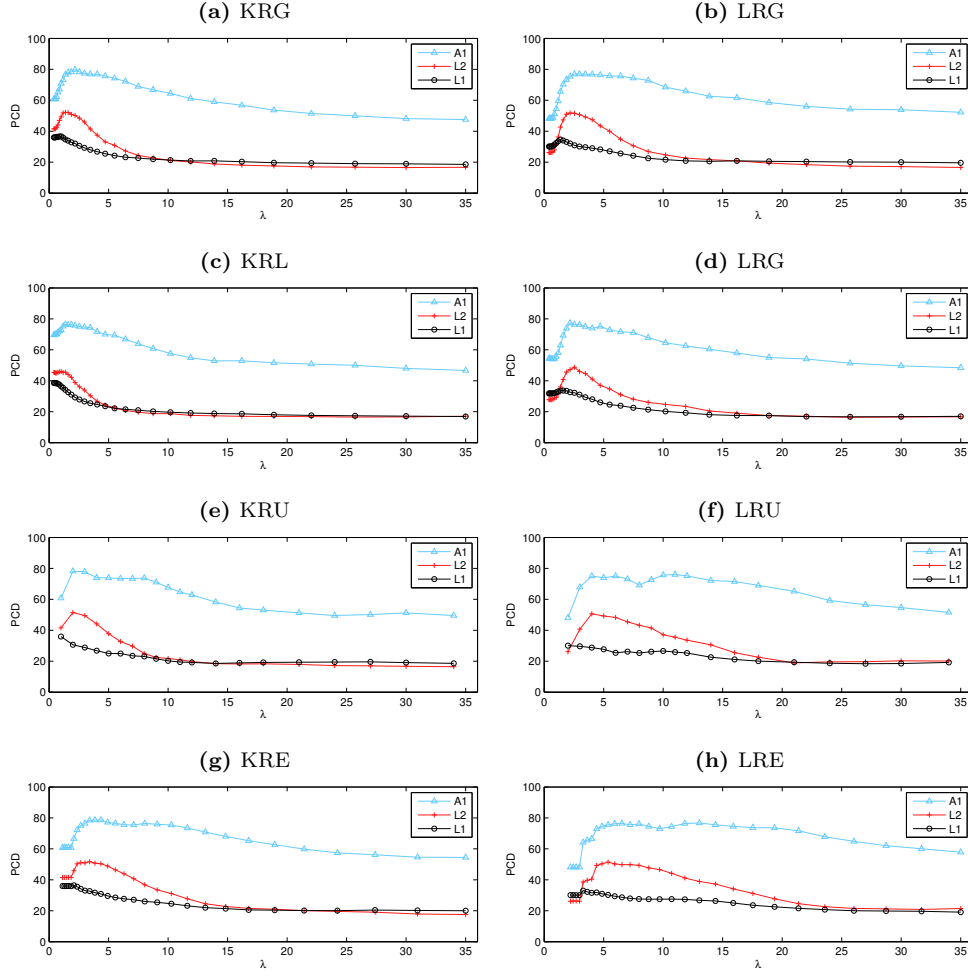
	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	36.7	38.6	35.9	36.4	34.6	33.7	30.1	33
L2	52.2	46	51.5	51.7	51.8	48.7	50.6	51.4
A1	79.8	76.5	78.2	78.6	77	77.3	76	76.7

$\arg\max_\lambda$ PCD

	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	0.865	0.467	1	2.09	1.37	1.37	2	3.28
L2	1.37	1.01	2	3.41	2.18	2.54	4	5.37
A1	2.18	1.37	2	4.35	2.97	2.18	11	13

Figure 3.3: Position finding results using the Spearman correlation coefficient ρ with kernel and local regression detrending and various kernels. The charts show PCD values in dependence of the kernel bandwidth parameter λ. The tables show the extremal values in the charts with 3 digits of accuracy.
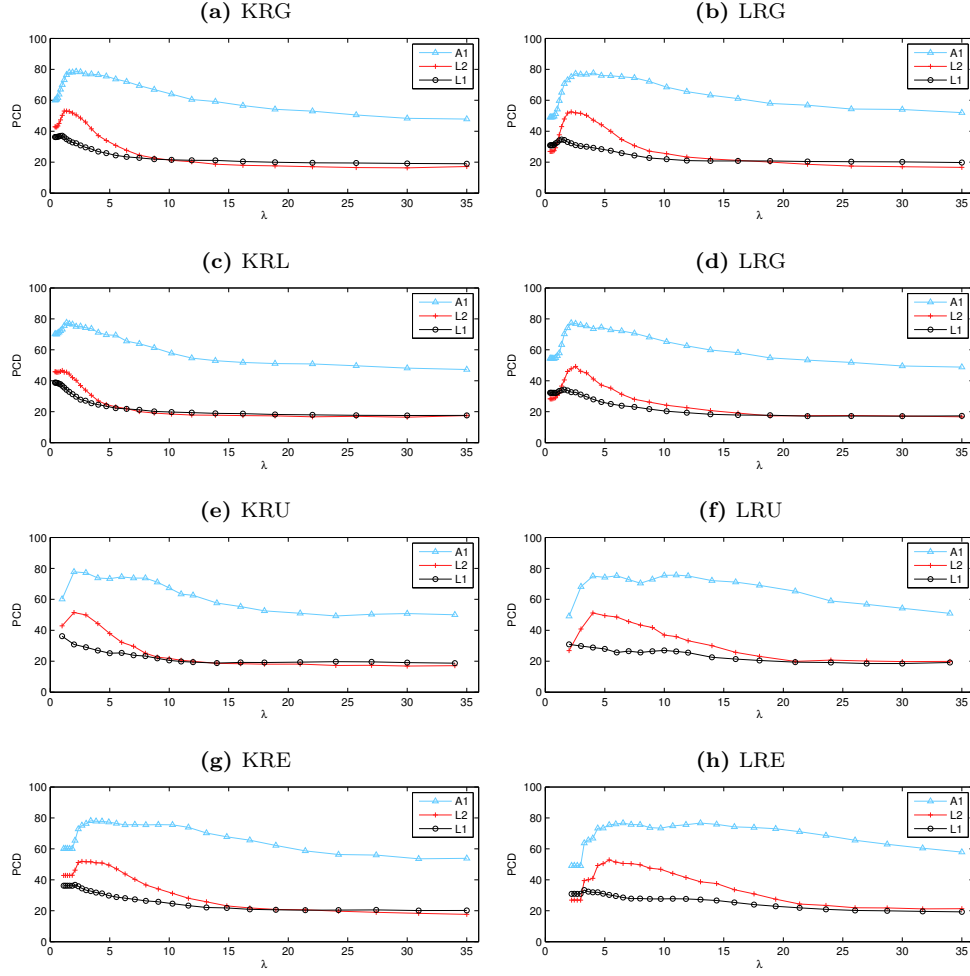
(a) KRG **(b)** LRG
(c) KRL **(d)** LRG
(e) KRU **(f)** LRU
(g) KRE **(h)** LRE

$\max_\lambda$ PCD

	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	37.1	38.8	36.2	36.7	34.6	34.3	30.9	33.4
L2	53.1	46.6	51.5	51.9	52.5	49.3	51.2	52.8
A1	78.5	77.6	77.8	78.1	77.6	77.4	75.7	76.7

$\arg\max_\lambda$ PCD

	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	1.01	0.467	1	2.09	1.37	1.6	2	3.28
L2	1.37	1.01	2	2.67	2.18	2.54	4	5.37
A1	2.18	1.37	2	3.41	4.04	2.18	11	13

Figure 3.4: Position finding results using the Kendall correlation coefficient τ with kernel and local regression detrending and various kernels. The charts show PCD values in dependence of the kernel bandwidth parameter λ. The tables show the extremal values in the charts with 3 digits of accuracy.
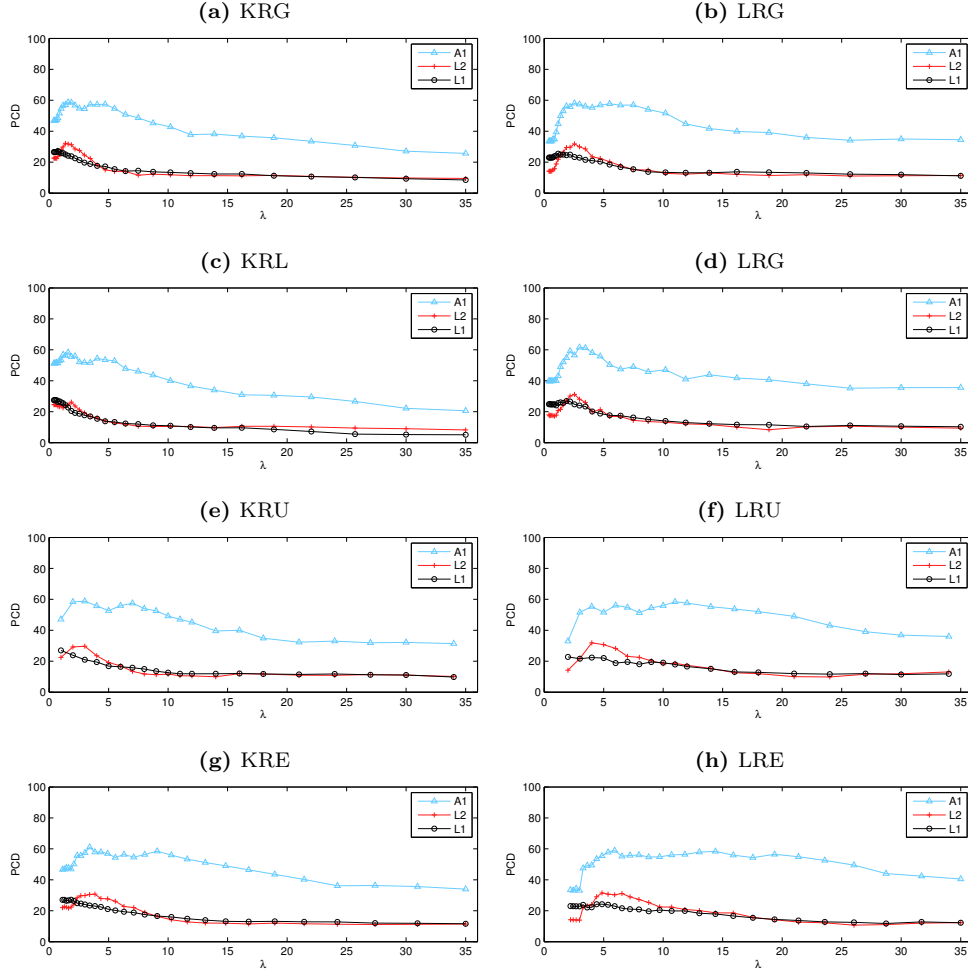
(a) KRG   **(b)** LRG

(c) KRL   **(d)** LRG

(e) KRU   **(f)** LRU

(g) KRE   **(h)** LRE

$\max_\lambda$ PCD

	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	27.2	27.5	27	27.1	25.5	26.9	22.8	24.2
L2	32.1	26.2	29.7	30.7	31.8	31.2	31.9	31.5
A1	58.8	58.3	58.8	61.1	58.1	61.6	58.4	58.9

$\arg\max_\lambda$ PCD

	KRG	KRL	KRU	KRE	LRG	LRL	LRU	LRE
L1	0.741	0.544	1	1.85	1.18	1.87	2	4.86
L2	1.37	1.87	3	3.85	2.54	2.54	4	4.86
A1	1.6	1.6	3	3.41	2.54	2.97	11	5.92

Figure 3.5: Position finding results using the sign test G with kernel and local regression detrending and various kernels. The charts show PCD values in dependence of the kernel bandwidth parameter λ. The tables show the extremal values in the charts with 3 digits of accuracy.

r									
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**	**DIF**	**NOD**
L1	32.6	31.7	29.4	31.6	30	18.7	27	28.4	28.4	19.5
L2	54	47.2	53.4	54.3	52.3	40.4	52.8	53	46.1	18.5
A1	69	74.5	59.7	65.3	69.6	70.2	74.4	73.2	57	33.8

$\mathbf{r} \circ \log$									
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**	**DIF**	**NOD**
L1	29	28.1	27.6	28.3	27.2	23.3	26.5	27.1	32.1	18.8
L2	49.8	42.8	49.9	50.2	50.8	39.5	49.1	50.1	47.2	24.2
A1	69.4	71	58.9	69.3	67	74.4	73.4	70.6	70.9	37.1

ρ									
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**	**DIF**	**NOD**
L1	34.7	36.4	30.7	32.7	32.1	32.2	28.8	30.3	38.6	20
L2	50.2	45.6	51.5	50.7	50.7	47.3	35.5	38.9	45.4	18.1
A1	67.2	69.7	60.8	66.7	65.7	62.9	48.1	64.2	69.5	38.3

τ									
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**	**DIF**	**NOD**
L1	34.9	36.8	30.9	34.4	32.2	32.5	28.9	30.2	38.7	20.5
L2	50.5	45.6	51.5	51.7	47.2	47.7	35.9	50.6	45.7	18.8
A1	70.1	70.2	60.2	65.4	65.1	70.2	49.1	63.7	70	38.7

G									
	KRG	**KRL**	**KRU**	**KRE**	**LRG**	**LRL**	**LRU**	**LRE**	**DIF**	**NOD**
L1	25	20.5	21	23	23.3	24.8	22.3	24.2	28.5	5.57
L2	31.7	24.4	29.7	30.5	31.8	28.1	18.7	30.3	25.6	7.95
A1	48.6	51.3	47.1	46.9	44.7	54.9	33	55.4	52.8	9.85

Figure 3.6: Position finding results using several matching methods in a cyclical cross-validation setup. The tables show PCD values on every test set with λ chosen to be optimal (individually for every matching-detrending combination) for one of the other data sets.

The figures 3.1-3.5 show mostly unimodal curves with a pronounced maximum for relatively small λ (in a range where the trend fits to the data tightly). The optimal choice of λ appears to depend strongly on the test set and, to a lesser degree, on the matching function (and, of course, on the choice of detrending, which was to be expected).

Furthermore, one immediately notices that the percentage of correctly dated pairs strongly depends on the test set. Partially, this is due to the different composition of the test sets (see section 1.5), L1 contains TRS from five different source locations (for which the climate or other factors influencing tree growth may vary strongly), and thus mostly TRS from different stands are compared, while in L2 most of the TRS are actually from one stand. A1 is composed of TRS from two different stands.

We can also see that for practically all detrending/matching combinations, the optimal choice of λ is lower for L1 than for L2. In Figure 3.7 we can see that for the TRS from L1 cross-stand matching apparently requires smaller λ than matching within a stand, which explains the difference of optimal λ between L1 and L2. One reason for this effect could be that stronger detrending tends to eliminate more of the regional disturbances. The test set A1, however, appears to generally favor bigger values of λ and the equivalent of Figure 3.7 shows little difference between the λ-dependence of cross-stand and within-stand matching.
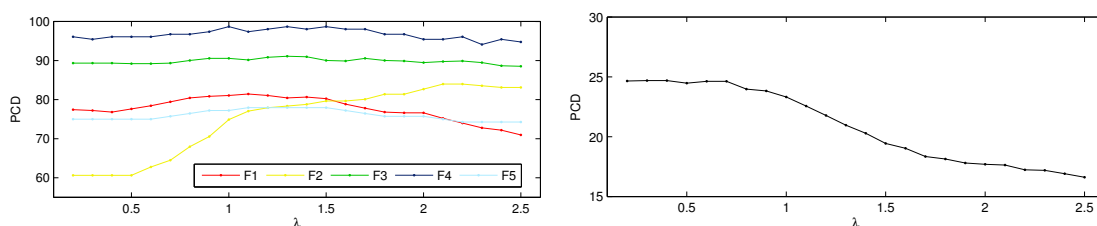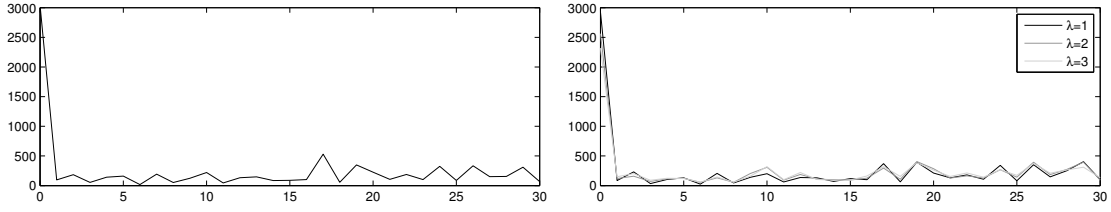

Figure 3.7: Position finding results using Kernel regression with a Gaussian kernel and the Kendall correlation coefficient τ in the test set L1. A minimal overlap of 70 years was required, all pairs of TRS were used. The charts show PCD values in dependence of the kernel bandwidth parameter λ in the critical range. In the left chart the pairs where both TRS from the same source location (within-stand) were compared for the stands F1-F5. The right chart shows the dating results for the pairs where the TRS are from different locations (cross-stand).

In general, the difference between the different kernel choices appears to be mostly negligible, with kernel regression giving slightly better results than local regression. The choice of matching function, on the other hand, appears to be quite significant. Non-parametric correlation shows better results than linear correlation, independent of the choice of detrending. The question whether taking the logarithm prior to matching improves the linear correlation results remains inconclusive as it appears to depend on the detrending function. Compared to the other matching methods, the sign test gives uniformly worse results.

Figure 3.8 suggests that the positional errors are distributed relatively even, so we can assume that none of the incorrect positions are favored.

In light of Figure 3.6 and figures 3.1-3.5 the Kendall correlation coefficient τ combined with either Gaussian kernel regression detrending with λ between 0.8 and 1.5 or the parameter-free difference detrending appear like the most solid choices in our setup.

Figure 3.8: Frequency of positional errors $|\hat{T} - T|$ for pairwise matching of TRS from L1, using the Kendall correlation coefficient τ, difference detrending (left chart) and Kernel regression with a Gaussian kernel with $\lambda = 1, 2, 3$ (right chart). A minimal overlap of 70 years was required, all pairs were used.

Considering the matching score's strong dependence on λ, it appears reasonable to maximize the score over λ as well, i.e., estimate $\hat{T}$ by

$$\hat{T} = \underset{S \in \mathcal{P}, \lambda_1 \in \mathbb{R}^+, \lambda_2 \in \mathbb{R}^+}{\arg\max} \widetilde{m}(h_{\lambda_1}(d|^S), h_{\lambda_2}(d^1)). \tag{3.20}$$

In a quick test using Kendall's τ and kernel regression with a Gaussian kernel (and the maximum roughly estimated by maximizing over a grid with step-size 0.1 in the critical range 0.8-1.5), however, the results in our test setup were similar to Figure 3.6 (35.23 for L1, 52.07 for L2 and 75.12 for A1), so this is probably not be worth the additional effort.

3.3 Significance testing

In practice, given a TRS $d = (d_t, t \in T)$, it is usually unknown whether T is actually one of the feasible positions $\mathcal{P}$. Given a position $S \in \mathcal{P}$ and the corresponding matching score $\widetilde{m}(h(d|^S), h(d^1))$ with another TRS d_1, we want to be able to decide if $d|^S$ is correctly or wrongly positioned.

For the rest of the section, let $x = \{h(d|^S)_t \mid t \in S \cap T^1\}$ and $y = \{h(d^1)_t \mid t \in S \cap T^1\}$ denote the overlapping parts of the two detrended TRS and assume they are realizations of random vectors X and Y.

In the classical frequentist approach to hypothesis testing we want to determine if $m = m(x, y)$ is significantly better than the expected matching score of TRS in wrong positions.

The assumption $T \neq S$ is our **null hypothesis** H_0 which we want to disprove by showing that the observed matching score is very unlikely to have occurred if H_0 held. If $F_{m,N}(y) = P(M \leq y \mid H_0) = P(m(X, Y) \leq y \mid H_0)$ is the CDF of the distribution of matching scores under the null hypothesis (it is presumed to depend only on the overlap size $N = S \cap T^1$ and m) and $F_{m,N}(m(x, y)) \geq 1 - \alpha$ for some significance level α, we reject the null hypothesis and assume $S = T$. Note that this is a **right-tailed** test, i.e., one that tests for positive correlation only.

We can estimate it by matching random samples of real data, or, assuming a model for the data, theoretically. For the latter the approach depends on the matching function. In general, we have to assume that $\{X_i \mid i = 1, \ldots, N\}$ and $\{Y_i \mid i = 1, \ldots, N\}$ are independent and identically distributed (i.i.d.) for $i = 1, \ldots, N$, or, in other words, that x_i and y_i are independent samples of some random variable. This is a significant assumption which is unlikely (also depending on

the detrending method) to be fulfilled by real TRS. Under this assumption, F_{m_N} is called the **sampling distribution** of m under H_0 based on a size N sample.

If $m = r$, we assume that X_i and Y_i are uncorrelated if H_0 holds and that (X_i, Y_i) is bivariate normal for all $i = 1, \ldots, N$, then

$$t = r(X, Y)\sqrt{\frac{N - 2}{1 - r(X, Y)^2}} \tag{3.21}$$

approximately follows a t-distribution with $N - 2$ degrees of freedom (cf. MONTGOMERY [16]). The above transformation $t\colon [-1, 1] \to \mathbb{R}$ is bijective, so the percentiles of t are the t-transforms of those of M.[2]

Alternatively, we may use a **Fisher transformation**. z defined by

$$z_r = \frac{\sqrt{N - 3}}{2} \log \frac{1 + r(X, Y)}{1 - r(X, Y)}. \tag{3.22}$$

approximately follows a standard normal distribution $\mathcal{N}(0, 1)$ (cf. MONTGOMERY [16]).

For $m = \tau$ the sampling distribution is discrete as τ depends only on the ranks of the data. If we assume that the ranks are independent, i.e., that all rankings of y_i are equally likely given x_i, the scores are approximately $\mathcal{N}(0, {}^{4N+10}/_{9N(N-1)})$ distributed (cf. KENDALL [17]) so

$$z_\tau = \frac{\tau(X, Y)}{\sqrt{{}^{4N+10}/_{9N(N-1)}}} \tag{3.23}$$

is approximately standard normal distributed.

For the sign test G and $i = 1, \ldots, N$ we have

$$
\begin{aligned}
&P(\tfrac{1}{2}|\operatorname{sgn}(X_i - 1) + \operatorname{sgn}(Y_i - 1)| = 1) \\
&= P(X_i \geq 1, Y_i \geq 1) + P(X_i < 1, Y_i < 1) \\
&= P(X_i \geq 1)P(Y_i \geq 1) + P(X_i < 1)P(Y_i < 1) \\
&= p
\end{aligned} \tag{3.24}
$$

and similarly

$$P(\tfrac{1}{2}|\operatorname{sgn}(X_i - 1) + \operatorname{sgn}(Y_i - 1)| = 0) = 1 - p \tag{3.25}$$

if X_i is assumed to be independent of Y_i and $\operatorname{sgn}(0) := 1$. Since we assumed that X_i and Y_i are i.i.d. we get

---

[2] The commonly used **Baillie-Pilcher** (5-year moving average detrending) and **Holstein** (difference detrending) tests use a two-tailed version of this test statistic (with decision if the TRS are positioned correctly based on $t(|m|)$ being over a certain threshold rather than the right-tailed conditional probability being small).

$$F(\tfrac{k}{N}) = P(G(X,Y) \le \tfrac{k}{N})$$
$$= \sum_{i=1}^{k} P(\tfrac{1}{N} \sum_{i=1}^{N} \tfrac{1}{2} | \operatorname{sgn}(X_i) + \operatorname{sgn}(Y_i) | = \tfrac{k}{N}) \tag{3.26}$$
$$= \sum_{i=1}^{k} \binom{N}{k} p^k (1-p)^k$$

as $N \cdot G(X,Y)$ is a sum of independent Bernoulli trials and thus $\mathcal{B}(N,p)$ distributed. For large N the distribution of $G(X,Y)$ is well approximated by $\mathcal{N}(p, {}^{p(1-p)}/_N)$ due to the central limit theorem.

If both X and Y are symmetrically distributed around 1 we get $p = 1 - p = \tfrac{1}{2}$. p can also be estimated as the mean of a random-simulated empirical distribution (i.e., as the average G-score of random off-positioned TRS with overlap N). For Gleichläufigkeit $p = 0.48$ gives a good fit (see Figure 3.9).

Despite the strong assumptions made for deriving the theoretical distributions, the theoretical and simulated distributions agree well (the critical zones are only slightly underestimated), as we can see in Figure 3.9.

Using $F_{m,N}$ we can now also choose to estimate T by

$$\hat{T} = \arg\max_{S \in \mathcal{P}} F_{m,N}(\widetilde{m}(h(d|^S), h(d^1))), \tag{3.27}$$

i.e., by taking $\hat{T}$ whose matching score is least likely to stem from a random position rather than maximizing the effect size (independent of N) with formula (3.2). Equivalently, one can maximize the z-scores obtained from equations (3.23) or (3.23) since their transformation to significance levels is bijective. The results of the problems (3.27) and (3.2) should be the same if a large enough minimal overlap is required.

Unfortunately, determining significance by accepting TRS matching over some chosen significance level α may still lead to a considerable number of false positives when testing a large number of positions as we only assessed $P(M \le m(x,y) \mid H_0)$, the conditional probability of a single score being less than $m(x,y)$ under the null hypothesis, without taking the multiple testing that occurred into account.

Ideally, the posterior probability $P(H_0 \mid M = m(x,y))$ can be found with **Bayesian inference** using Bayes theorem

$$P(H_0 \mid M = m(x,y)) = \frac{P(M = m(x,y) \mid H_0)P(H_0)}{P(M = m(x,y))} \tag{3.28}$$
$$= \frac{P(M = m(x,y) \mid H_0)P(H_0)}{P(M = m(x,y) \mid H_0)P(H_0) + P(M = m(x,y) \mid \neg H_0)P(\neg H_0)}$$
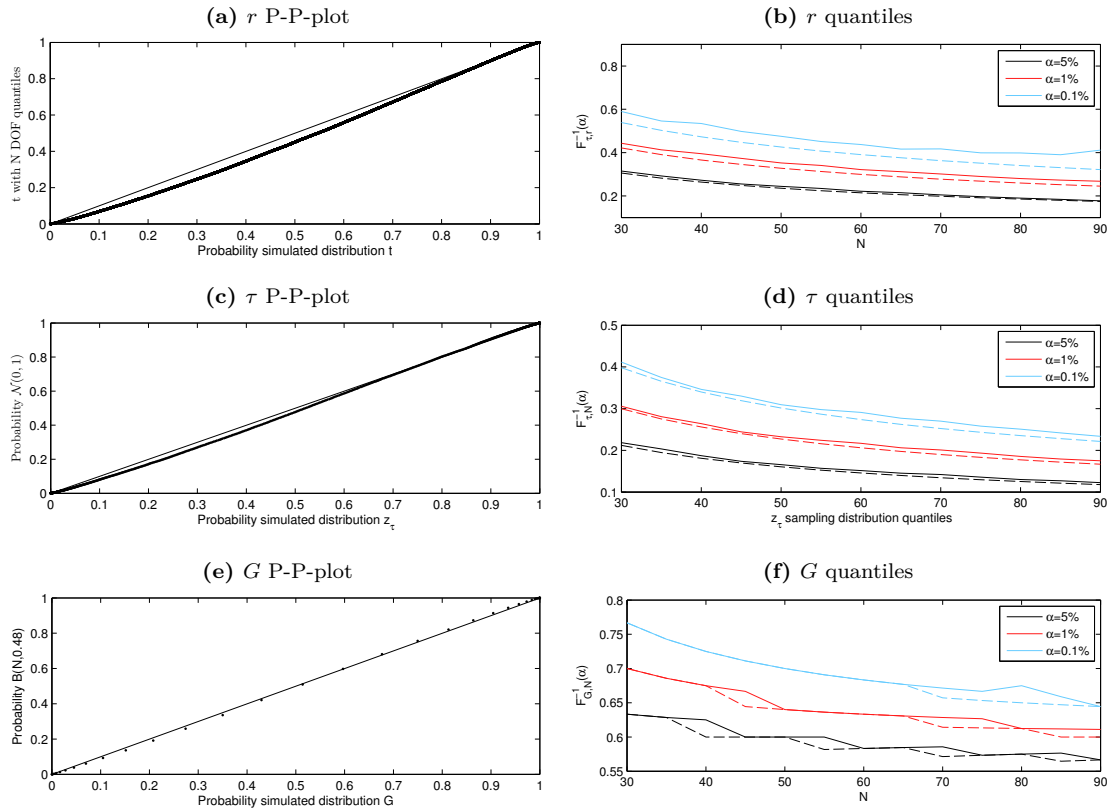
as H_0 and $\neg H_0$ are mutually exclusive events.

Using the likelihood ratio $q = {}^{P(M=m(x,y)|H_0)}/_{P(M=m(x,y)|\neg H_0)}$ (with P denoting the likelihood function) (3.28) can be expressed as

$$P(H_0 \mid M = m(x,y)) = \frac{qP(H_0)}{qP(H_0) + P(\neg H_0)}. \tag{3.29}$$

We see that the posterior probability $P(H_0 \mid M = m(x,y))$ depends on the prior probability $P(H_0)$. Without any further information (e.g., from radiocarbon dating) a reasonable choice of $P(\neg H_0)$, the prior probability that the position S is correct, would be

$$P(\neg H_0) = \frac{C}{|\mathcal{P}|}. \tag{3.30}$$

(a) r P-P-plot **(b)** r quantiles

(c) τ P-P-plot **(d)** τ quantiles

(e) G P-P-plot **(f)** G quantiles

Figure 3.9: Comparison of theoretical (dashed lines) and simulated (solid lines) distributions of r, τ and G. The figures on the left are P-P plots of the matching scores obtained by matching random TRS which fulfill H_0 against their theoretically obtained sampling distributions. For 48000 randomly off-positioned pairs of difference detrended and logarithmized TRS from Lebanon with overlap $N = 30, 35, \ldots, 90$ matching scores $m_{i,N}, i = 1, \ldots, 48000$ were calculated. The P-P plots are scatter plots of $\bar{F}(m_{i,80})$ with $\bar{F}$ denoting the empirical CDF of the m_i against $F(m_{i,80})$ as given by t, normal or binomial distribution. The plots on the right show the $\alpha = 5, 1, 0.1\%$ quantiles of r, tau and G calculated with $\bar{F}$ (solid lines) and the theoretical sampling distributions (dashed lines) as a function of N.

$0 \leq C \leq 1$ denotes the (subjective) amount of trust one has that the given TRS fits somewhere in the database.

With the prior chosen as above, $P(H_0 \mid M = m(x, y))$ may be significantly bigger than $P(M = m(x, y) \mid H_0)$. This effect is especially pronounced when $|\mathcal{P}|$ is big or when the TRS is very unlikely to fit in $\mathcal{P}$ (this could be seen as an example of the prosecutor's fallacy).

Unfortunately, it's hard to estimate $P(M = m(x, y) \mid \neg H_0)$ as it strongly depends on the data composition (see Figure 3.10). Live simulation is tricky due to the dependence of the distribution on S and the fact that in practice only very few TRS fully overlapping with S may be available.
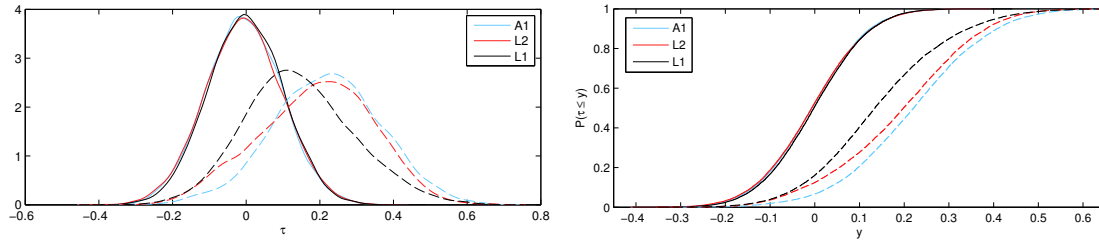
Figure 3.10: Distributions of correctly positioned (dashed lines) and off-positioned (solid lines) TRS matching scores. For 18000 random correctly and off-positioned pairs with overlap $N = 45$ of difference detrended TRS from the test sets **L1**, **L2** and **A1** Kendall correlation scores $\tau_i, i = 1, \ldots, 18000$ were calculated. The plot on the left shows kernel density estimations (generated using the MATLAB statistics toolbox routine *ksdensity*) of the distributions of the τ_i, on the right we can see the empirical CDFs.

From a frequentist perspective we can recognize that when comparing a large number of positions the chance of a random position getting a score in the top α-percentile is actually much higher than estimated above as, under the null hypothesis $S \neq T$ for all $S \in \mathcal{P}$, $\widetilde{m}(h(d|^{\hat{T}}), h(d^1))$ with $\hat{T}$ obtained by formula (3.1) is distributed like the maximum of $|\mathcal{P}|$ t-, normal or binomially distributed random variables M_S with overlaps $N_S, S \in \mathcal{P}$. If we assume them to be independent (which appears reasonable) we get

$$\bar{F}(y) = P(\bar{M} \leq y \mid H_0) = P(\max_{S \in \mathcal{P}}(M_S) \leq y \mid H_0) \tag{3.31}$$
$$= P(M_{S_1} \leq y, \ldots, M_{S_{|\mathcal{P}|}} \leq y \mid H_0) = \prod_{S \in \mathcal{P}} P(M_S \leq y \mid H_0) = \prod_{S \in \mathcal{P}} F_{m, N_S}(y).$$

For (3.27) a similar formula can be derived. So, if $\min_{S \in \mathcal{P}}(N_S) = N$, we actually have to require a conditional probability larger than $\sqrt[|\mathcal{P}|]{1 - \alpha}$ to ensure that only α of TRS fulfilling the null hypothesis have a position with a maximal matching score in the significant range. This probability depends on the number of positions checked $|\mathcal{P}|$.

3.4 Mean chronologies

In most of the commonly used TRS dating methods, the TRS to be dated is compared with a **mean chronology** of the region it is suspected to originate from. The general idea is that the mean chronology averages out the various noise signals of the individual TRS it is composed of and strongly resembles the climate signal of the area the samples were taken from. Comparison may happen by hand (visually), or automatically, utilizing one of the methods from section 3.1. In the setup preceding equation (1.1) we have two possibilities for the latter. Either we average the detrended TRS and get $\hat{T}$ by

$$\hat{T} = \underset{S \in \mathcal{P}}{\arg\max}\ \widetilde{m}((d|^S), \bar{h}), \tag{3.32}$$

where $\bar{h} = (\bar{h}_t)_{t \in \mathcal{T}}$ with $\bar{h}_t = \frac{1}{n_t} \sum_{i=1}^{M} h(d^i)_t$ and $\mathcal{T} = \bigcup_{i=1}^{M} T^i$, or we average the normalized TRS and detrend only to match positions, giving

$$\hat{T} = \underset{S \in \mathcal{P}}{\arg\max}\ \widetilde{m}(h(d|^S), h(\bar{d})), \tag{3.33}$$

where $\bar{d} = (\bar{d}_t)_{t \in \mathcal{T}}$ with $\bar{d}_t = \frac{1}{n_t} \sum_{i=1}^{M} d_t^i$ and $\mathcal{T} = \bigcup_{i=1}^{M} T^i$. n_t denotes the number of (detrended) TRS averaged over in the year t, i.e., $|\{h(d^i)_t, i = 1, \ldots, M\}|$ or $|\{d_t^i, i = 1, \ldots, M\}|$.

COOK [2] states that using formula (3.33) is inferior, as it averages possibly different distributed quantities. Also note that in years t where a new TRS d^i is introduced into the mean series, the shape of the $\bar{d}$ close to the year t may get distorted if $d_t^i \gg \bar{d}_t$ or $d_t^i \ll \bar{d}_t$, i.e., if the growth trend of d^i is vastly different from the average growth trend, especially if the average is composed of a low number of TRS. Due to heteroskedasticity the yearly fluctuations of TRS with a large growth trend will also be weighted much heavier than those of TRS with a small growth trend. We will use (3.32) for the rest of this section.

Recognizing the presence of outliers, it is also possible to use more robust measures such as the median or trimmed means rather than the arithmetic mean.

It is common among dendrochronologists to **preselect** the TRS introduced into the mean chronology, which means to take only those TRS which agree well with each other (even if their age is known). Given a set of series of known age and their mean, TRS are removed from the mean series by reverse order of their *matching score with the mean of the others* until all remaining TRS score in the significant range for some chosen level α. The idea behind this practice is to reduce the amount of random noise in the mean chronology.

For hypothesis testing, all methods from section 3.3 can be used as we assume all properties required of TRS to hold for mean TRS as well.

4 Odds ratio test using covariance models

4.1 Time windows

In this chapter, we take a different approach suggested by Arnold Neumaier. Most concepts and notation are taken from his lecture course on data analysis held in 2008.

First of all, rather than comparing whole TRS with each other, we will compare parts of them. Given a **window size** $l \in \mathbb{N}$, a TRS $d = (d_t)_{t \in T}$ and some detrending function h, the TRS d in the l**-sized window at the time** t is defined by

$$w(d)_t = \begin{pmatrix} h(d)_t \\ \vdots \\ h(d)_{t+l-1} \end{pmatrix} \quad \text{for all } t \in \widetilde{T} = \{T_1, \ldots, T_{|T|-l+1}\}, \tag{4.1}$$

which we normalize by setting the Euclidian norm to 1, giving

$$\widetilde{w}(d)_t = \frac{w(d)_t}{\|w(d)_t\|_2}. \tag{4.2}$$

We can also use several different representations of the data in the window at t (log-data, detrended, ranks, etc.) at the same time by merging them into a column vector $\widetilde{w}(d)_t \in \mathbb{R}^L$ (as long as they are sufficiently linear independent). Due to the piecewise normalization, detrending can be expected to be less vital than for methods of comparison which compare whole series at once.

4.2 Covariance potentials and sample statistics

Until now, we have assumed that $h(d)_t, t \in T$ are independent samples drawn from a random variable. The approach here is much more universal. We start from the setup of equation (1.1). Assuming that $d^1, \ldots, d^M$ are drawn from a stochastic process D, the general idea is that we want to utilize the (l-dimensional) probability distribution of $\widetilde{w}(D)_t$ (estimated using the sample $W_t = \{\widetilde{w}(d^1)_t, \ldots, \widetilde{w}(d^M)_t\}$) to measure the likelihood of $\widetilde{w}(d|^S)_t$, the TRS to date shifted into the position $S \in \mathcal{P}$ in the window at t, occurring.

Unfortunately, modeling a multi-variate probability distribution with high precision is impossible if the number of variables is large, as the amount of data points required for a solid estimate of the distribution grows rapidly as the number of variables (here l) increases (explained in detail by NEUMAIER [18]). As data is scarce and l typically large, we will have to be content with a crude estimate. Given n data points $y_1, \ldots, y_n \in \mathbb{R}^l$ we can describe the rough shape of their probability distribution using **ellipsoid potentials** of the form

$$V(x) = (x - \mu)^T C^{-1} (x - \mu) + c, \tag{4.3}$$

with $\mu \in \mathbb{R}^l$ and $C \in \mathbb{R}^{l \times l}$ positive definite. The level sets $V(x) = r^2$ are ellipsoids centered at μ with principal axes $a_i = r\sqrt{\lambda_i} u_i$, where $\lambda_i \in \sigma(C)$ are the eigenvalues and u_i the corresponding normalzed eigenvectors of C (cf. NEUMAIER [18]).

If we take, for example, $\mu = \mu(\{y_1, \ldots, y_n\}) = \frac{1}{n} \sum_{i=1}^{n} y_i$, the sample mean, $C = C(\{y_1, \ldots, y_n\})$ such that $C_{ik} = Cov(y_i, y_k)$, the sample **covariance matrix** and $c(\{y_1, \ldots, y_n\}) = \log \det C$, we get a good fit to the data, as with μ and C as just defined

$$\rho(x) = \frac{1}{\sqrt{2\pi \log \det C}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)} \tag{4.4}$$

is the probability density with maximal likelihood when assuming normality and $V(x) = -2\log(\rho(x))$ its so-called **negative log-likelihood**, which is small for likely and large for unlikely values of x (see NEUMAIER [18]). The resulting potential is called a **covariance potential**. Covariance potentials are meaningful even in the non-Gaussian case as they describe the rough shape of the data cloud.

In practice, we use **sample statistics** to estimate mean and covariance matrix (and thus the potential). The **feature vector**

$$F(y) = \begin{pmatrix} 1 \\ y \\ yy^T \end{pmatrix} \in \mathbb{R} \times \mathbb{R}^l \times \mathbb{R}^{l \times l} \tag{4.5}$$

contains all information the data point y contributes to the potential. We call a weighted sum of feature vectors

$$S(\{y_1, \ldots, y_n\}) = \sum_{i=1}^{n} \alpha_i F(y_i) \tag{4.6}$$

the statistics of the sample $\{y_1, \ldots, y_n\}$ with respect to the feature mapping F. The weights determine how much every data point contributes to the shape and position of the potential. Given sample statistics

$$S = \begin{pmatrix} \alpha(S) \\ a(S) \\ A(S) \end{pmatrix} \tag{4.7}$$

we get the necessary parameters for the corresponding potential $V(S)$ by

$$\mu(S) = \frac{a(S)}{\alpha(S)} \tag{4.8}$$

$$C(S) = \frac{A(S)}{\alpha(S)} - \mu(S)\mu(S)^T. \tag{4.9}$$

For uniform weights $\alpha_i = 1$, $\mu(S)$ is the mean and $C(S)$ the covariance matrix.

4.3 Building potentials, likelihood ratios and cumulative distribution functions

For dating purposes, taking uniform weights, we have year-wise statistics

$$S_t = S(W_t) = \sum_{i=1}^{M} F(\widetilde{w}(d^i)_t) \quad \text{for all } t \in \mathcal{T} = \bigcup_{i=1}^{M} \widetilde{T}^i \tag{4.10}$$

and a time-independent statistic

$$S = \sum_{t \in \mathcal{T}} S_t. \tag{4.11}$$

If $n_t = |W_t|$ is the number of TRS $d^1, \ldots, d^M$ with data in the window at year t and

$$n_t < l, \tag{4.12}$$

$C(S_t)$ is singular (very badly conditioned) as it is a sum of n_t matrices with rank 1 and thus

$$\text{rank}(C(S_t)) \leq n_t < l. \tag{4.13}$$

To **regularize**, i.e., to avoid the singularity, we use

$$\widetilde{S}_t = S_t + \delta S \quad \text{for all } t \in \mathcal{T} \tag{4.14}$$

to estimate the potential $V_t = V(\widetilde{S}_t)$ which approximates the distribution of the data points in W_t. We take $\delta = \frac{\gamma}{\alpha(S)}$, γ determines the influence of the regularizing term. If $n_t \gg 1$, $\delta\alpha(S) \ll n_t = \alpha(S_t)$, so the regularization term is negligible for years where a lot of data is available.

We also compute potentials $U_t = W(S - S_t)$ from $S - S_t$ which approximate the distribution of the TRS $d^1, \ldots, d^M$ in all off-position windows $s \neq t$. The log-likelihood-ratio

$$q_t(y) = V_t(y) - U_t(y) \quad \text{for all } t \in \mathcal{T} \tag{4.15}$$

is a quadratic function of y which, in the window at t, separates the on-position (where it is expected to be small) and off-position data (where it is expected to be large). For multivariate normal data it does so optimally.

If $C_{V,t}$, $C_{U,t}$ and $\mu_{V,t}$, $\mu_{U,t}$ denote the covariance matrices and means corresponding to the potentials V_t and U_t, we get

$$q_t(y) = y^T(C_{V,t}^{-1} - C_{U,t}^{-1})\, y + 2y^T(C_{U,t}^{-1}\, \mu_{U,t} - C_{V,t}^{-1}\, \mu_{V,t}) + c \tag{4.16}$$

$$c = \mu_{V,t}^T\, C_{V,t}^{-1}\, \mu_{V,t} - \mu_{U,t}^T\, C_{U,t}^{-1}\, \mu_{U,t} + \det(C_{V,t}) - \det(C_{U,t}), \tag{4.17}$$

which is slightly faster than (4.15), by expanding and regrouping.
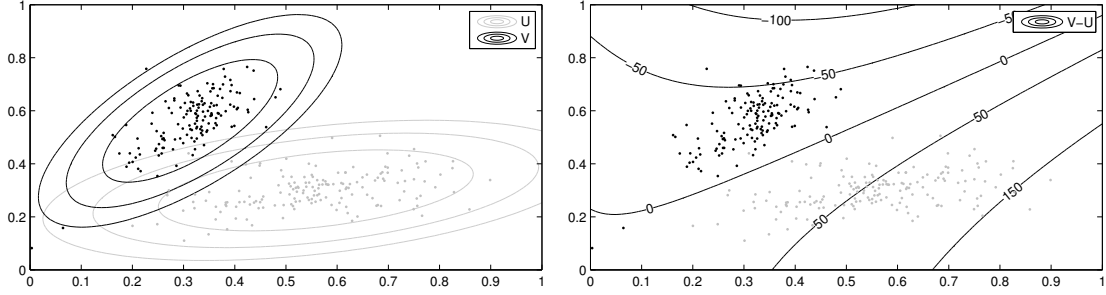
Figure 4.1: The left Figure shows two clusters of (artificially generated) data and a contour plot of the corresponding potentials V and U. The right Figure is a contour plot of $q = V - U$.

We then calculate empirical cumulative distribution functions

$$v_t(x) = \frac{|\{y \in W_t \ | \ q_t(y) < x\}|}{n_t + \eta} \tag{4.18}$$

$$u_t(x) = \frac{|\{y \in W \backslash W_t \ | \ q_t(y) < x\}| + \eta}{|W \backslash W_t| + \eta} \tag{4.19}$$

where $W = \bigcup\limits_{s \in \mathcal{T}} W_s$ and $\eta > 0$ a regularization parameter to avoid indeterminate forms.

v_t is an estimate of the CDF of the q_t-score of (all) TRS in the window at t given $S = T$, (correctly positioned TRS), u_t of the CDF of the reverse case $S \neq T$ (off-positioned TRS).

4.4 Dating and significance testing

For $q = q_t(\widetilde{w}(d|^S)_t)$, the **odds** in the window at t

$$o_t(q) = \frac{1 - v_t(q)}{u_t(q)} \tag{4.20}$$

determine how well $d|^S$ fits in the window at t. Due to the partial scoring, the detection of missing year rings is simplified (the search space of feasible positions can be extended with relatively little additional effort) but we will not go into further detail here.

The window odds o_t are strongly dependent so we can not simply calculate the joint odds/probabilities by building their product. We can, however combine them by building their geometric mean (which for odds is a more natural choice than the arithmetic mean) for a total score

$$o(d|^S) = \prod_{t \in \widetilde{S} \cap \mathcal{T}} o_t(q_t(w(d|^S)_t)^{1/|\widetilde{S} \cap \mathcal{T}|}. \tag{4.21}$$

As the next step, we want to develop some kind of criterion when to consider $o(d|^S)$ good enough to be accepted as dated.

While it would be natural to use the geometric mean of $1 - u_t(q_t(\widetilde{w}(d|^S)_t)$ to determine a significance threshold for the total odds (using it as a measure of the probability of the null hypothesis $S \neq T$ similar to section 3.3), it almost never reaches the $1 - \alpha$ level in correct positions $S = T$ (even for reasonably small α and large M) as the $u_t(q)$ become very small for some windows t (see Figure 4.2), so it can't be used like a total conditional probability of a score as good as $\widetilde{w}(d|^S)$ occurring under $S \neq T$.
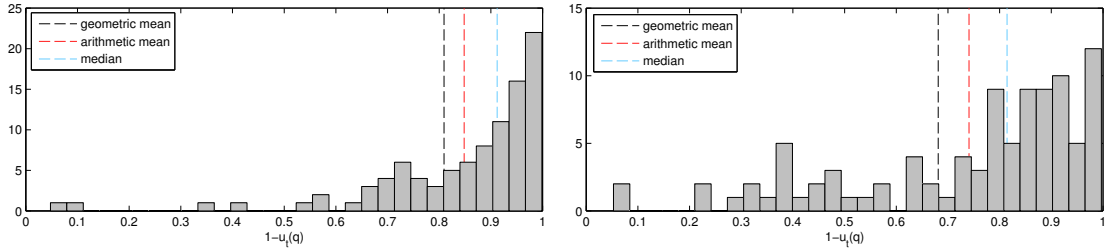

Figure 4.2: Two examples of the distribution of the $1 - u_t(q)$ for $t \in \widetilde{S} = \{S_1, \ldots, S_{|S|-l+1}\}$ and $S = T$ (correctly positioned) with window size $l = 8$ and $|W_t| = 80$ for all t. The data was difference detrended and logarithmized.

Another possibility is the visual assessment of a plot of position (or the first year S_1 when assuming index sets consisting of consecutive years) against total odds (the equivalent using the correlation coefficient r is called a cross-correlation plot). If the highest peak is distinct enough to be unlikely to stem from the same distribution as the other scores, we consider the TRS dated. For this plot it is practical to use a logarithmic scale for the total odds for improved visibility. A few examples can be seen in Figure 4.3.

Some of the plots in Figure 4.3 show a pronounced trend distorting the results. By slightly altering the algorithm (e.g., by adding weights at the appropriate point), it may be possible to alleviate this issue, further improving the method's performance.

Varying l, γ and η independently in different setups with a small number of samples showed $l = 8$, $\gamma = 5$ (i.e $\delta = \frac{5}{\alpha(S)}$) and $\eta = 2$ to be solid choices. The observed differences in dating performance (number of correctly dated TRS, score difference between correctly and wrongly positioned data) were small but appeared significant.

Similar tests suggested the use of difference detrended and logarithmized data.
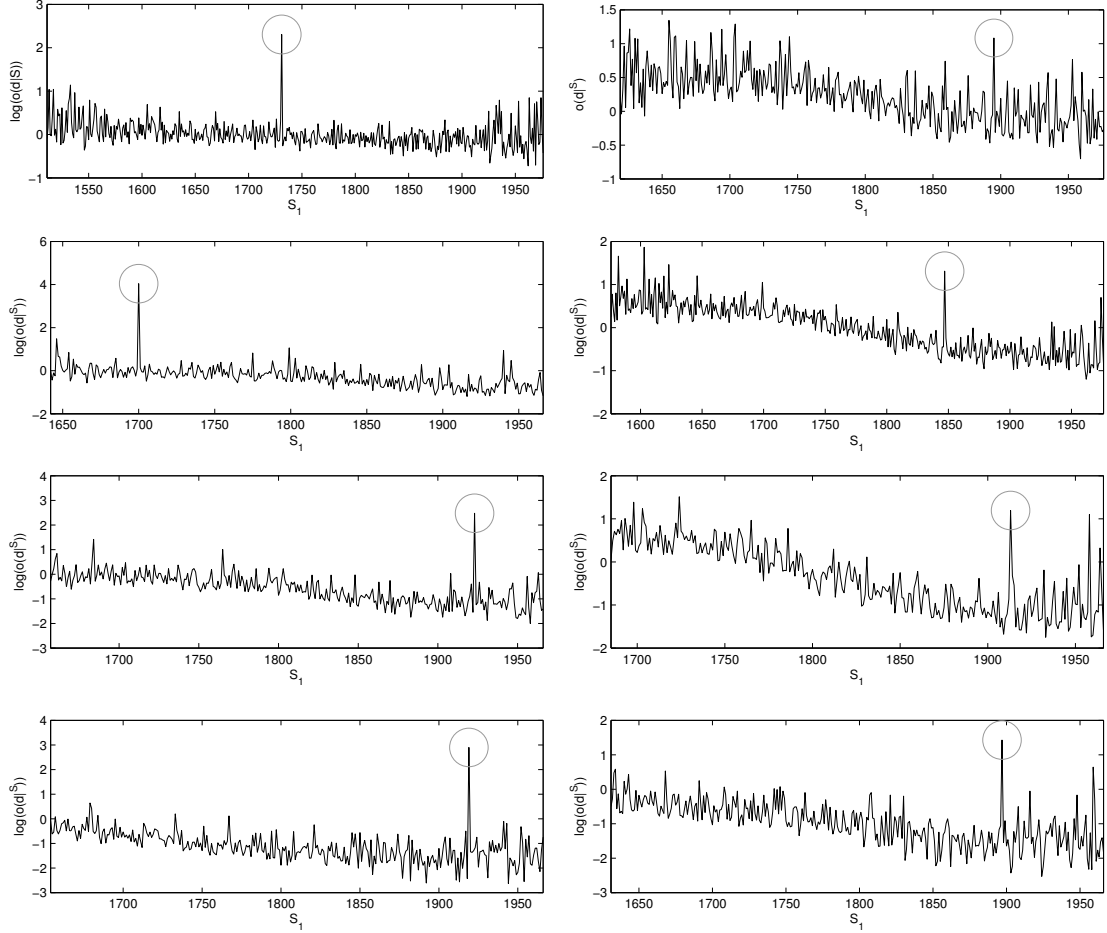
Figure 4.3: Plot of (the logarithm of the) total odds against position S_1. The correct position is marked with a circle. This type of plot can be used to infer if the highest peak stems from a correct position. The plots in the left column show clearly isolated peaks in the correct position (some of them less pronounced than others). In the right column all positions should be considered ambiguous as the peaks are not very dominant compared to their background. The trend visible in some of the graphs (i.e., that the average score appears to be higher for small t) is likely due to the number of TRS available decreasing steadily for $t < 1900$ (and that fact that off-position scores tend to be lower on average when more data is available as the border between on- and off-position data becomes more well-determined).

5 Evaluation

In this section, we want to test the practical performance of the dating algorithms suggested in the sections 3.4 and 4. The test setup is chosen to be as realistic as possible while still allowing for control of the relevant parameters (like the size of the database or the length of the series).

Using the sets

L1 Lebanon 1880-1999 (96 TRS)

L2 Lebanon 1382-1880 (55 TRS)

we generate 200 size-M random subsets (scenarios) $L_i \subset L$ for several $M \in \mathbb{N}$. For each of the TRS in $L1 \setminus L_i$, a random N-year subsequence is selected and dated (using L_i) requiring full overlap (i.e., testing the positions with $S_1 = 1880, \ldots, 1999 - N + 1$) and the maximal score tested for significance. We distinguish the four cases

Positive (P) correctly dated, score deemed significant

False positive (FP) incorrectly dated, score deemed significant

Negative (N) incorrectly dated, score deemed insignificant

False negative (FN) correctly dated, score deemed insignificant

and record their percentages as measured over all scenarios L_i.

When matching using mean chronologies, we use the sampling distributions of the matching functions under the null hypothesis as given in section 3.3 at a 99.98% confidence level to obtain a threshold (for the correlation coefficient the t-test is used). The practice of preselection, i.e., the reduction of the mean series to only those TRS which match significantly with the mean (of the others), is also tested.

We use

r Linear correlation coefficient

τ Kendall correlation coefficient

G Sign test

for matching and

d Difference detrending (with the logarithm taken after detrending)

g Kernel regression with a Gaussian kernel and $\lambda = 1$

to detrend.

For the odds-ratio test we use a simple heuristic to determine thresholds. The idea is to have a simple criterion which gives decisions similar to the visual assessment of score-versus-position

charts as in Figure 4.3. A score should be considered significant if the peak is large enough. So, if $\hat{T}$ is the position with maximal score and μ and σ are mean and standard deviation of $\{o(d|^S) \mid S \in \mathcal{P} \land S \neq \hat{T}\}$, the total odds in the other positions, we consider $o(d|^{\hat{T}})$ significant if

$$o(d|^{\hat{T}}) \geq \mu + \kappa\sigma \tag{5.1}$$

for a fixed $\kappa \in \mathbb{N}$.

We use $l = 8$, $\gamma = 5$ and $\eta = 2$ as suggested in chapter 4. We test two representations of the data

d Difference detrended (with the logarithm taken after detrending)

o Original, unaltered data.

For use in practice, i.e., when n_t varies and a trend like in 4.3 appears, the above test may also be used, but μ and σ should be the local mean and standard deviation (obtained by weighting positions S close to T more heavily than those far apart, for example, with Gaussian weights) to estimate how distinct the peak is from its background.

The charts in the following figures show the test results, i.e., the measured values of P, N, FP and FN as a function of M for several detrending and, for the traditional dating methods, matching methods. In Figure 5.1, traditional dating methods without preselection were used, whereas in Figure 5.2 the TRS averaged into the mean chronology were preselected. Figure 5.3 shows the results for the odds ratio test.
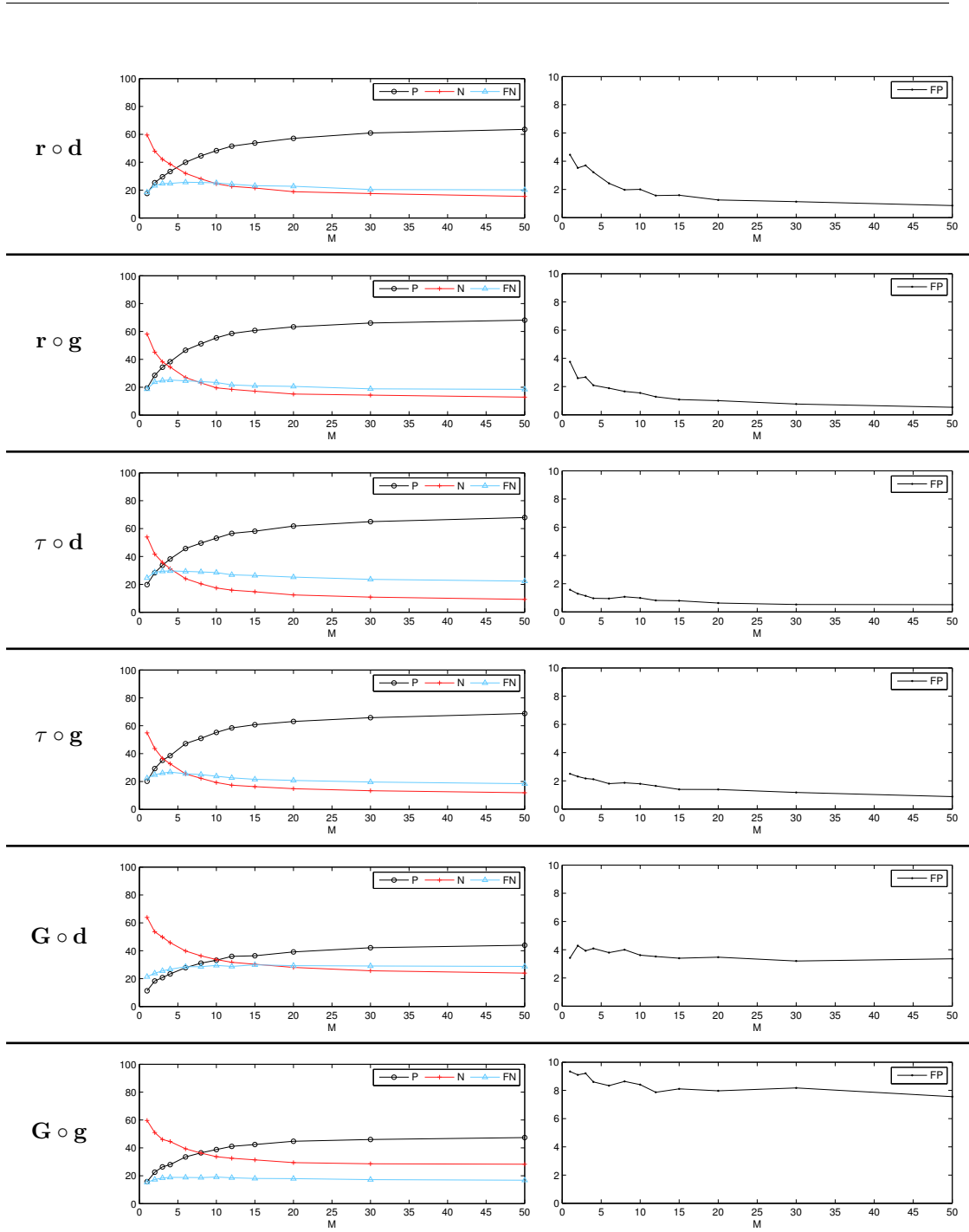
Figure 5.1: Evaluation results using the traditional methods described in chapter 3. The mean chronologies were built without preselecting. The left column shows the percentages of positives, negatives, false negatives and the right column the percentage of false positives for varying M. The rows represent different combinations of matching and detrending methods. We chose $M \in \{1, 2, 3, 4, 6, 8, 10, 12, 15, 20, 30, 50\}$ and $N = 70$.
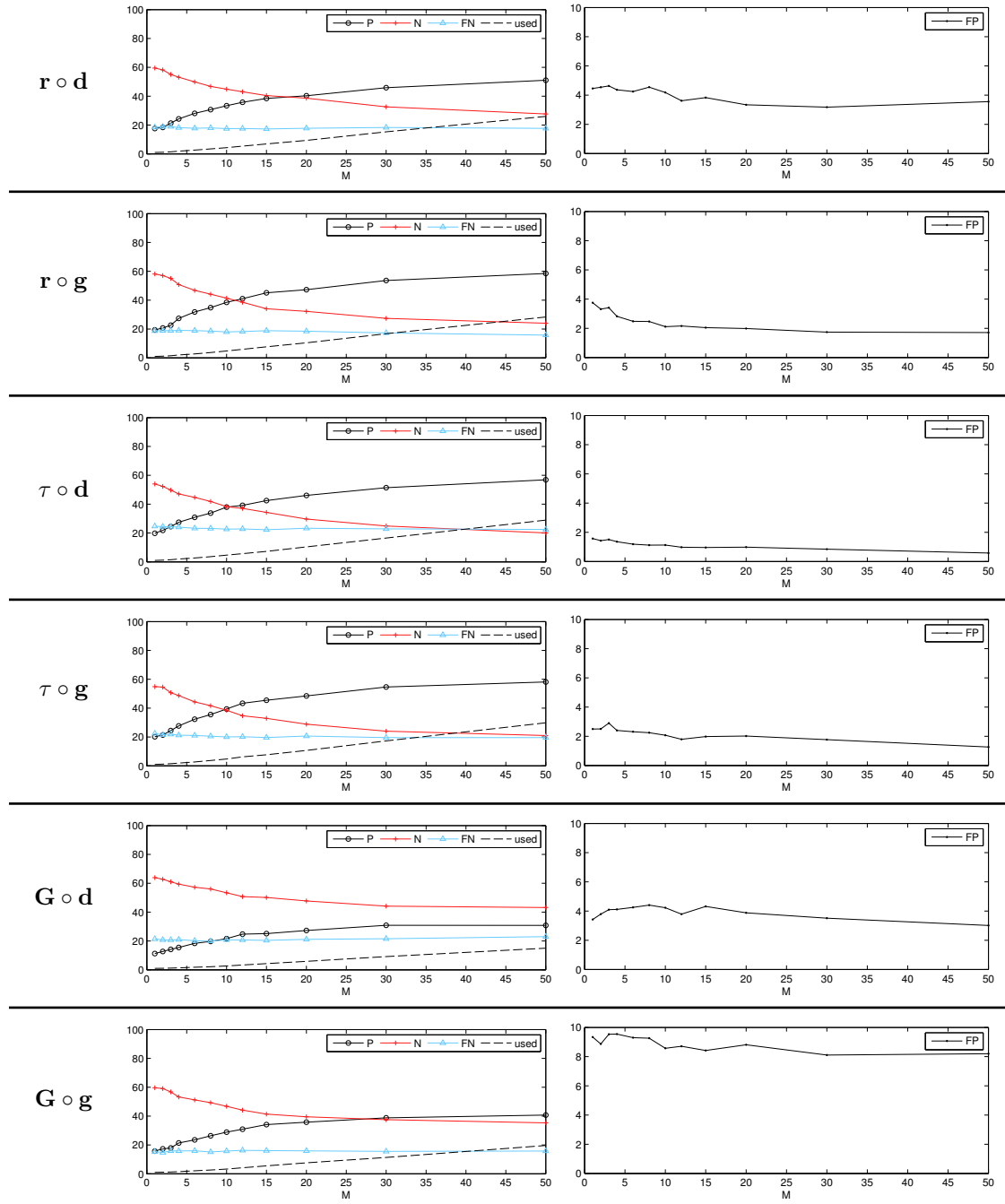
Figure 5.2: Evaluation results using the traditional methods described in chapter 3. Only TRS agreeing well were preselected to form the mean chronologies (see section 3.4). The dashes lines show the average number of TRS used to build the chronology. The left column shows the percentages of positives, negatives, false negatives and the right column the percentage of false positives for varying M. The rows represent different combinations of matching and detrending methods. We chose $M \in \{1, 2, 3, 4, 6, 8, 10, 12, 15, 20, 30, 50\}$ and $N = 70$.
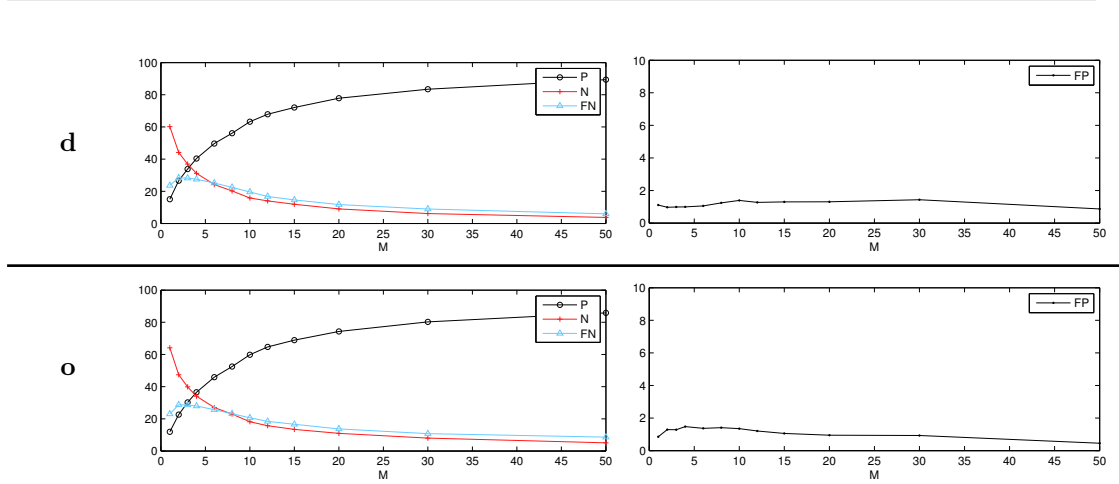
Figure 5.3: Evaluation results using the odds ratio test described in chapter 4. The dashes lines show the average number of TRS used to build the chronology. The left column shows the percentages of positives, negatives, false negatives and the right column the percentage of false positives for varying M. For the upper row difference detrended and logarithmized data was used, for the lower row the original data. We chose $M \in \{1, 2, 3, 4, 6, 8, 10, 12, 15, 20, 30, 50\}$, $N = 70$ and $\kappa = 8$ (the latter by raising its value until the number of false positives was roughly in the 1% range).

As expected, raising the amount of TRS available for comparison improves the dating performance for all methods. In the figures 5.1, 5.2 and 5.3 this is reflected by P rising drastically (and N dropping sharply) as M gets bigger.

For the **traditional dating methods** without preselecting, FN, the percentage of TRS which were dated correctly with a score not deemed significant, slightly rises until $M \approx 10$ (as more TRS get dated correctly with rising M) followed by a very slow decline (as the score of the correctly dated TRS rises). The figures 5.4 and 5.5 suggest that when matching with a mean chronology scores increase only barely as more TRS are introduced into a mean chronology after a certain M, which explains why FN remains almost constant for $M \geq 10$. With preselection, FN appears to be relatively constant. In the evaluation graphs of the odds ratio test, FN drops continuously after a small rise for small M (as here the scores apparently more often keep increasing when more data is used for comparison even for larger M).

Due to equation (3.31) we would have expected FP, the percentage of false positives, to be less than $0.9998^50 \approx 1\%$ (with some randomly high off-position scores dominated by correct-position scores) for the traditional methods. The charts, however, show much bigger values for most methods. They tend to drop for bigger M (as more TRS get dated correctly). For the odds ratio test, FP is roughly at the 1% mark by choice of κ and slightly less for very large M.

Of the traditional methods, $\tau \circ \mathbf{d}$ the combination of the Kendall correlation coefficient τ with difference detrending appears to work best. It shows by far the lowest number of false positives on average (less than 1% for most M) and only slightly less positives than $r \circ \mathbf{d}$ and $\tau \circ \mathbf{g}$. In the charts with preselection, P grows much slower with M than in those without while FP is about the same, so at least in the test setup using all available TRS for the mean chronology appears to be worthwhile.

The **odds ratio test** shows promising evaluation results, especially when using difference

detrended data. For $M \geq 4$, it is able to date more TRS than the traditional methods, for very large M drastically so. The odds ratio test also works very well with unaltered data. As the age trend and longer-lasting non-climatic signals are also removed by formula (4.2) (normalizing in the individual time windows), this is not entirely surprising.
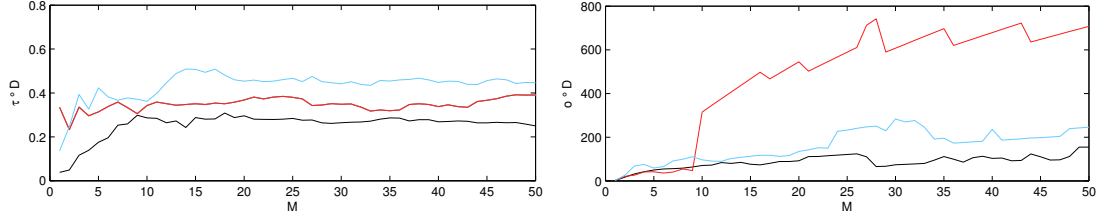

Figure 5.4: Three examples of the development of the score of a correctly aligned TRS as the number of TRS available for comparison increases. In the left Figure we compare 3 (correctly positioned) 70-year TRS with the mean chronology of M others using difference detrending and the Kendall correlation coefficient. In the right Figure the odds ratio test with difference detrended data, $l = 8$, $\gamma = 5$ and $\eta = 2$ is applied to the same data.
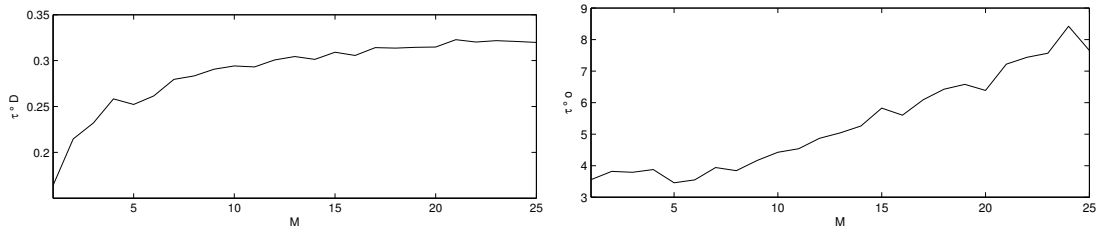

Figure 5.5: Average score as a function of M. For $M = 1, \ldots, 25$ we compare 1400 randomly selected correctly aligned TRS with a database of M others using traditional dating with difference detrending and the Kendall correlation coefficient (left figure) and the odds ratio test with difference detrended data, $l = 8$, $\gamma = 5$ and $\eta = 2$ (right figure).

In the figures 5.6 and 5.7 we can see that $\tau \circ \mathbf{d}$ and $r \circ \mathbf{g}$ give pretty similar information, though they differ enough that one of them may lie in the significant range while the other does not. Gleichläufigkeit and the odds ratio test give information quite different from $\tau \circ \mathbf{d}$. Figure 5.6 suggests that for large M, odds ratio scores vary more (i.e., are more likely to be either very small or or very large) than those where M is small. Based on Figure 5.7 alone, no difference in separation sharpness can be made out. Visually, all scores distinguish between correctly and wrongly positioned TRS.
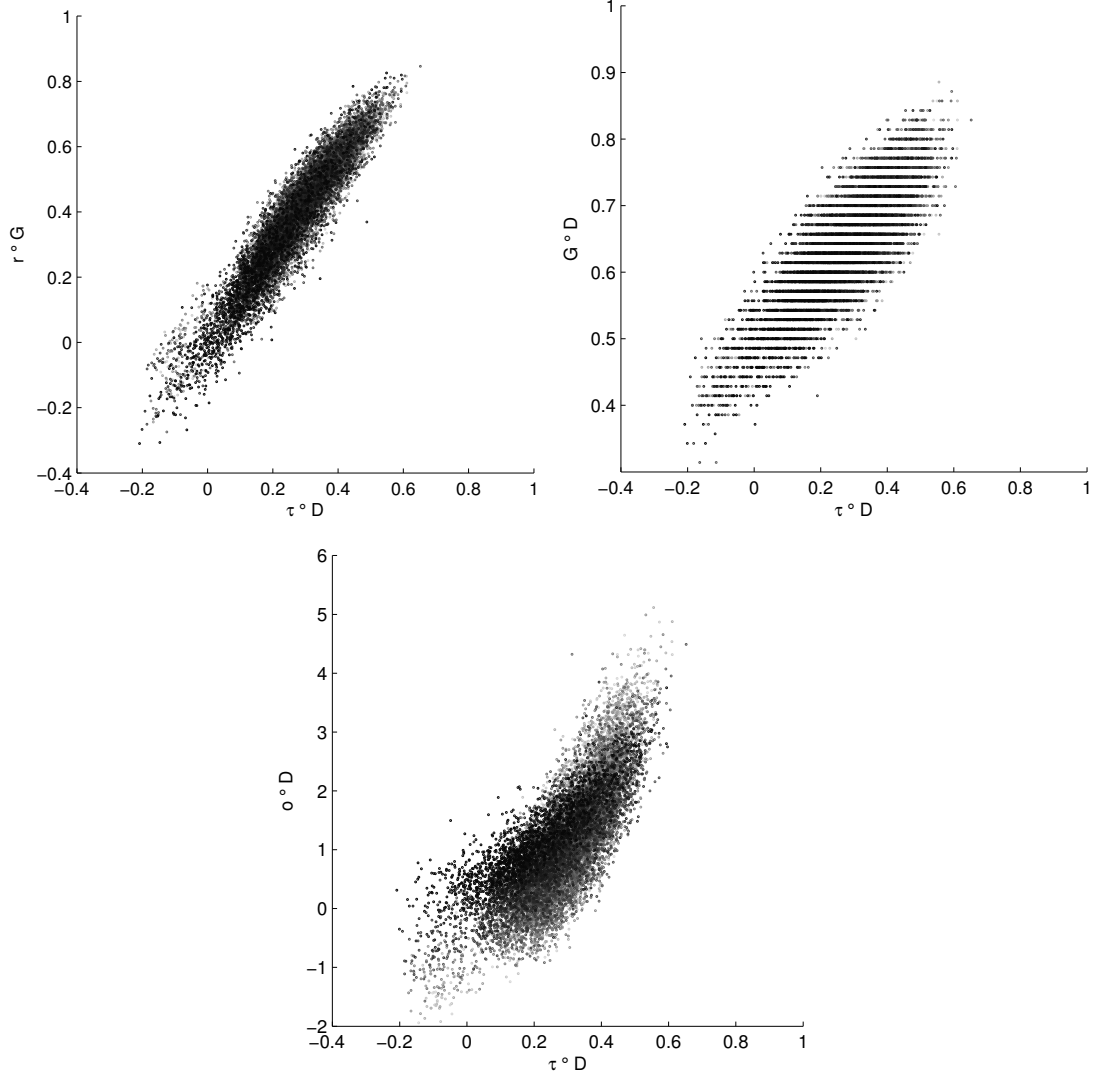
Figure 5.6: Correlation of the scores generated by different dating methods. The figures show the scores obtained by traditional dating with with difference detrending and the Kendall correlation coefficient ($\tau \circ \mathbf{d}$) plotted against the scores received by using Gaussian detrending with $\lambda = 1$ and linear correlation ($r \circ \mathbf{g}$), Gleichläufigkeit ($G \circ \mathbf{d}$) and the odds ratio test with difference detrended data, $l = 8$, $\gamma = 5$ and $\eta = 2$ ($o \circ \mathbf{d}$). All scores stem from correctly positioned (randomly chosen) 70-year TRS. Those obtained from scenarios with large M are bright, those where M was small are dark. M ranges from 1 to 25.
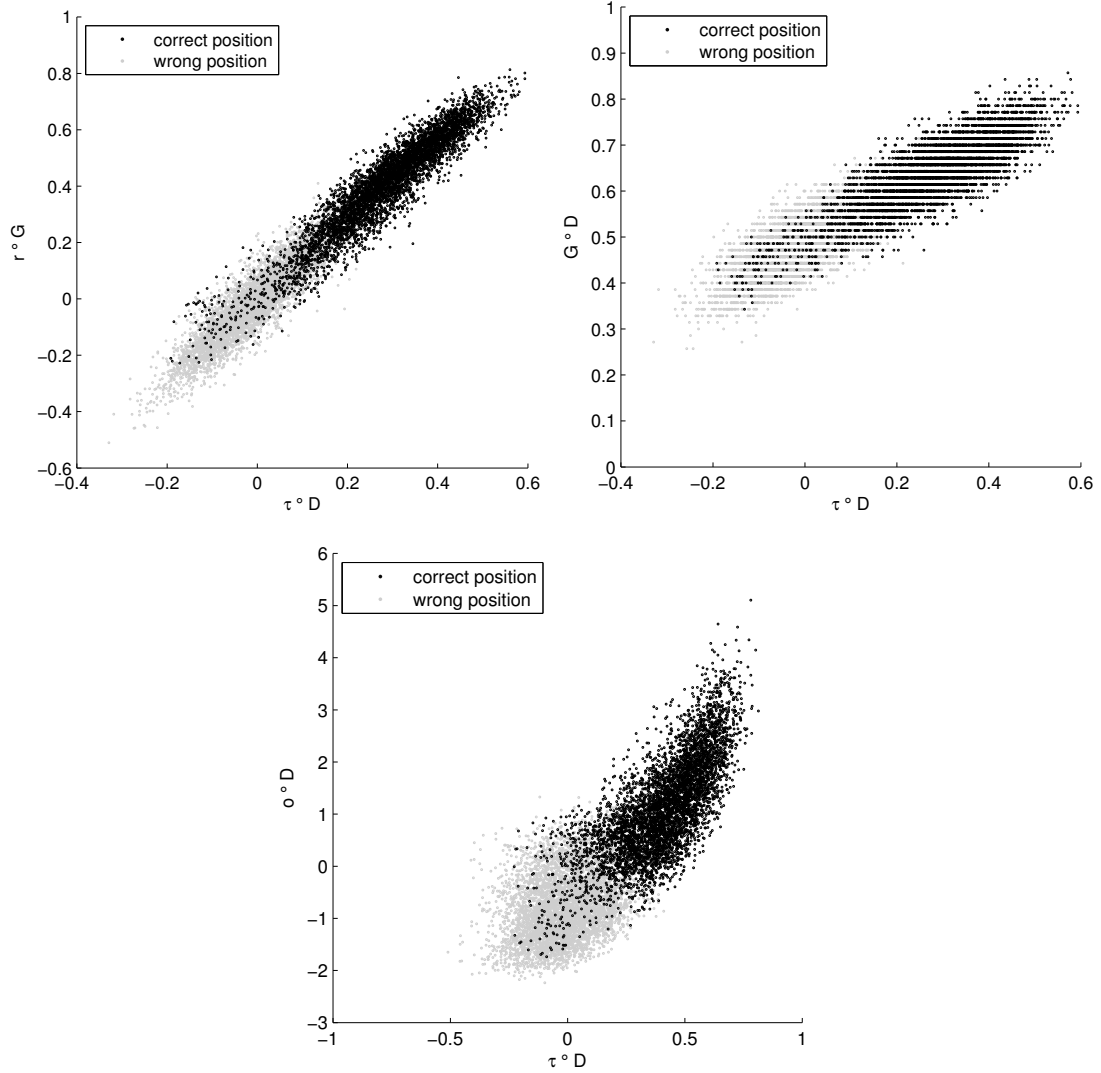
Figure 5.7: Correlation of the scores generated by different dating methods. The figures show the scores obtained by traditional dating with with difference detrending and the Kendall correlation coefficient ($\tau \circ \mathbf{d}$) plotted against the scores received by using Gaussian detrending with $\lambda = 1$ and linear correlation ($r \circ \mathbf{g}$), Gleichläufigkeit ($G \circ \mathbf{d}$) and the odds ratio test with difference detrended data, $l = 8$, $\gamma = 5$ and $\eta = 2$ ($o \circ \mathbf{d}$). All scores stem from correctly positioned (randomly chosen) 70-year TRS. Scores of wrongly positioned (randomly shifted) TRS are drawn in gray, the scores from correctly positioned TRS are plotted in black. M ranges from 1 to 25.

6 Conclusion

In chapter 5, the dating methods presented in the previous chapters were tested systematically. Although some of the methods appear to work relatively well, the results were not completely satisfactory.

While for most TRS the correct position produced the highest score (at least when a large amount of data was available for comparison), the distinction between correctly dated TRS and random high scores remains problematic. Due to the multiple testing (of all possible positions) and the severity of false positives, a very high degree of confidence is required. For the traditional dating methods, the amount of false positives was much higher than predicted by theory and for the odds ratio test a parameter (κ) was hand-fit to ensure that the percentage of false positives is at a certain level, which is unsuitable for practical purposes (as it is questionable whether the same κ will give similar results for shorter TRS or differently composed data). With the methods presented here, the likelihood that a dated TRS is a false positive cannot be estimated reliably.

The detrending methods suggested in chapter 2 are mostly based on heuristics such as equation (2.1). At present, non-climatic signals can only be removed by building differences or high-pass filtering. A good model of the climate-tree interaction might lead to an improved signal extraction and thus better dating results. To get such a model, more detailed tree data (containing, for example, stand information like soil quality or altitude of the stand) and climate data may be necessary.

Another problem out of the scope of this document is the possibility of error propagation when building a chronology. False positives in the database could lead to a distortion of the scores and thus increase the likelihood of further false positives. When building very long chronologies composed of short sequences this may pose a serious problem, so further investigations should examine this issue as well.

A related problem is the proper maintenance of the internal consistency of a database, i.e., whether it is advisable to remove or re-date TRS from the database if the addition of other (high-scoring) series lowers their score below significance. The results in chapter 5 (more precisely, the experiments with preselection) suggest that, at least when starting a new database, requiring internal consistency (by removing TRS which do not fit) does not immediately improve the dating.

As all proposed dating methods work with lower reliability in the one-to-one case where only a single TRS is available for comparison, building a floating chronology/database (see section 1.3) might require different methods to get good results. A possibility would be allowing already positioned TRS to be moved as more series are added or the simultaneous determination of all positions at once, for instance, by solving a problem of the type

$$\hat{T}_1, \ldots, \hat{T}_n = \underset{S_1 \in \mathcal{P}, \ldots, S_n \in \mathcal{P}}{\arg\max} \ \widetilde{Q}(d^1|^{S_1}, \ldots, d^n|^{S_n}) \tag{6.1}$$

where $d^1, \ldots, d^n$ are the TRS to compose the floating chronology, $\hat{T}_1, \ldots, \hat{T}_n$ their estimated dates and $\widetilde{Q}$ a scoring function of M TRS (which should be invariant under a permutation of its arguments). As the search space grows very quickly as n grows, solving such a problem would require an intelligent traversal of the search space (perhaps using discrete optimization techniques).

Bibliography

[1] Cook, E.R. 1987. *The decomposition of tree-ring series for environmental studies.* Tree-Ring Bulletin **47**. pp. 37-59.

[2] Cook, E.R., Kairiukstis, L.A. 1990. *Methods of Dendrochronology. Applications in the Environmental Sciences.* Kluwer Academic Publishers. Dordrecht, Boston, London. p. 106

[3] Cook, E.R., Kairiukstis, L.A. 1990. *Methods of Dendrochronology. Applications in the Environmental Sciences.* Kluwer Academic Publishers. Dordrecht, Boston, London. p. 110

[4] Cook, E.R., Kairiukstis, L.A. 1990. *Methods of Dendrochronology. Applications in the Environmental Sciences.* Kluwer Academic Publishers. Dordrecht, Boston, London. pp. 112-113

[5] Cook, E.R. 1981. *The smoothing spline: A new approach to standardizing forest interior tree-ring width series for dendroclimatic studies.* Tree-Ring Bulletin **41**. pp. 45-53.

[6] Schweingruber, F.H. 1983. *Der Jahrring. Standort, Methodik, Zeit und Klima in der Dendrochronologie.* Paul Haupt, Bern. pp. 24-25

[7] Schweingruber, F.H. 1983. *Der Jahrring. Standort, Methodik, Zeit und Klima in der Dendrochronologie.* Paul Haupt, Bern. p. 93

[8] Schweingruber, F.H. 1983. *Der Jahrring. Standort, Methodik, Zeit und Klima in der Dendrochronologie.* Paul Haupt, Bern. pp. 174-177

[9] Schweingruber, F.H. 1983. *Der Jahrring. Standort, Methodik, Zeit und Klima in der Dendrochronologie.* Paul Haupt, Bern. pp. 210-213

[10] El-Shaarawi, A.H, Piegorsch, W.W. 2002. *Encyclopedia of environmetrics. Volume 1.* John Wiley & Sons, New York. p. 491

[11] Grieser, J. 1997. *Analyse und Interpretation dendrochronologischer Daten zur Klimageschichte.* Preprint available at `http://www.juergen-grieser.de/germanoldcentury/pdfs/Dendrolog.pdf`, pp. 7-9

[12] Nicolussi, K. et al. 2009. *A 9111 year long conifer tree-ring chronology for the European Alps: a base for environmental and climatic investigations.* The Holocene **19**. pp. 909-920

[13] Helama, S. et al. 2009. *Dendroclimatic transfer functions revisited: Little Ice Age and Medieval Warm Period summer temperatures reconstructed using artificial neural networks and linear algorithms.* Ann. Geophys. **27**. pp. 1097-1111

[14] Bradley, R. 1999. *Paleoclimatology. Reconstructing Climates of the Quaternary. Second Edition.* Academic Press, London. pp. 4-5

[15] Weisstein, Eric W. *Correlation Coefficient.* MathWorld - A Wolfram Web Resource. `http://mathworld.wolfram.com/CorrelationCoefficient.html`

[16] Montgomery, D.C., Runger, G.C. 2003. *Applied Statistics and Probability for Engineers. Third Edition.* John Wiley & Sons, New York. pp. 402-403

[17] Kendall, M.G. 1938. *A New Measure of Rank Correlation* Biometrika **30** pp. 81-93

[18] Neumaier, A. 2008. *Data Analysis.* Lecture notes.

curriculum vitæ

Wolf Naetar

Geboren am	27. April 1982
Geburtsort	Wien
Staatsbürgerschaft	Österreich
Familienstand	Ledig

Sprachen

Deutsch (Muttersprache)

Englisch (verhandlungssicher)

Spanisch (gute Kenntnisse)

Schulische Ausbildung

1992-1996	BRG Laaerbergstrasse, Wien
1996-2001	HTL Spengergasse für EDV und Organisation, Wien

Studium und Weiterbildung

2001-	Studium der Mathematik an der Universität Wien
2006	Auslandssemester im Rahmen eines Erasmus Stipendiums, Madrid