Messfehlermodelle für die Survey-Statistik und die Wirtschaftsarchäologie

Inaugural-Dissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von Marcus Groß, Master of Science

aus Schwerin

Berlin 2016

Dekan: Prof. Dr. Dr. Andreas Löffler

Erstgutachter: Prof. Dr. Ulrich Rendtel

Zweitgutachter: Prof. Dr. Timo Schmid

Drittgutachterin: Dr. Eva Rosenstock

Tag der Disputation: 12. Mai 2016

Publikationsliste und Erklärung über die Koautorenschaft

Aus den 5 Kapiteln dieser Arbeit sind die folgenden Publikationen, welche mit den jeweils aufgeführten Koautoren erstellt wurden, entstanden:

1. Groß, M., U. Rendtel, T. Schmid, S. Schmon, and N. Tzavidis (2016). **Estimating the density of ethnic minorities and aged people in Berlin: Multivariate kernel density estimation applied to sensitive georeferenced administrative data protected via measurement error.** *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, forthcoming.

2. Groß, M. and U. Rendtel (2016). **Kernel density estimation for heaped data.** *Journal of Survey Statistics and Methodology*, accepted.

3. Rosenstock, E., M. Groß, A. Hujić, and A. Scheibner (2016). **Back to good shape:biological standard of living in the copper and bronze ages and the possible role of food.** In J. Kneisel, W. Kierleis, N. Taylor, M. dal Corso, and V. Tiedtke (Eds.), *Setting the Bronze Age Table: Production, Subsistence, Diet and Their Implications for European Landscapes.* Proceedings of the International Workshop 'Socio-environmental dynamics over the last 12.000 years: the creation of landscapes III (5th - 18th April 2013)', Kiel., Bonn. Habelt.

4. Groß, M (2016). **Modeling body height in prehistory using a spatio-temporal bayesian errors-in variables model.** *AStA Advances in Statistical Analysis*, forthcoming.

5. Groß, M., E. Rosenstock, and A. Hujić (2016). **Reconstruction of body height from long bones for prehistoric individuals: New methodological concepts.** *American Journal of Physical Anthropology*, submitted.

Inhaltsverzeichnis

Einleitung

Die vorliegende Arbeit befasst sich mit sogenannten Messfehlermodellen in der angewandten Statistik. Dabei wurden Daten aus zwei sehr verschiedenen Fachgebieten analysiert und verarbeitet. Zum einen Umfrage- und Registerdaten, welche in der Survey-Statistik Anwendung finden und zum anderen anthropologische Daten zu prähistorischen Skeletten. Beiden gemeinsam ist, dass einige Variablen nicht hinreichend genau erfasst werden können. Dies kann etwa aus Datenschutzgründen beabsichtigt sein oder auf (Mess-) Ungenauigkeiten beruhen. Diesen Umstand kann man unter den Oberbegriffen Messfehler oder Fehler-in-den-Variablen zusammenfassen. Diese Messfehler können fatale Auswirkungen in der statistischen Analyse haben. Carroll et al. (2006) spricht in diesem Zusammenhang von einem dreifachen Fluch ("triple whammy"). Zunächst führen Messfehler bei vielen statistischen Verfahren zu erheblichen Verzerrungen in den Parameterschätzungen. Diese Verzerrungen verschwinden nicht mit steigender Fallzahl sondern können – wie in Teil I dargelegt – bei einer nichtparametrischen Dichteschätzung etwa sogar erheblich wachsen. Ein weiteres prominentes Beispiel hierfür ist der sogenannte "attenuation bias" in der linearen Einfachregression (Frost and Thompson, 2000), in welcher der Steigungsparameter bei klassischem Messfehler in Richtung Null verzerrt ist. Zusätzlich ist ein Verlust von statistischer Effizienz bzw. Power eine Folge. Als dritten Fluch von Messfehlern kann man die stark erschwerte grafische Analyse, welche für das Entdecken von Strukturen und Zusammenhängen aber auch Ungereimtheiten in den Daten so wichtig ist, bezeichnen, da Messfehler vorhandene Zusammenhänge maskieren oder unkenntlich machen. Trotz dieser folgenschweren Auswirkungen werden Messfehler in statistischen Analysen in der Anwendung fast immer ignoriert. Diese Arbeit entwickelt daher für bekannte statistische Verfahren wie (multivariate) Kerndichteschätzung und nichtparametrische Regression eine Korrektur anhand konkreter Anwendungen.

Seien $\boldsymbol{X} = (X_1, .., X_n)$ die wahren, unbeobachteten Werte, welche als latente Variablen angesehen werden können, und $\boldsymbol{W} = (W_1, .., W_n)$ die beobachteten bzw. gemessenen Variablenwerte. Messfehler haben verschiedene Ursachen und müssen daher unterschiedlich modelliert werden. Die beiden bekanntesten Modellierungsansätze sind der klassische Messfehler mit

$$W_i = X_i + U_i, \text{ mit } U_i \text{ u.i.v und unabhängig von } X_i,\ i = 1, .., n$$

und der Berkson-Fehler (Berkson, 1950):

$$X_i = W_i + U_i, \text{ mit } U_i \text{ u.i.v und unabhängig von } W_i.$$

Beide Modelle scheinen sich auf den ersten Blick sehr zu ähneln. Um sich den Unterschied klar zu machen, kann es hilfreich sein sich die Größe der Varianzen von X_i und W_i zu verdeutlichen. Da beim klassischen Messfehlermodell der wahre Wert X_i und der Messfehler U_i unabhängig voneinander sind, folgt, dass $Var(W_i) = Var(X_i) + Var(U_i) > Var(X_i)$. Umgekehrt resultiert beim Berkson-Modell, dass $Var(X_i) = Var(W_i) + Var(U_i) > Var(W_i)$. Der klassische Messfehler findet z.B. Anwendung bei der Modellierung von fehlerhaften Messungen durch ungenaue Messinstrumente, während der Berkson-Fehler für die Modellierung der Ausprägung eines Individuums bei ausschließlicher Kenntnis des Populationsmittelwertes geeignet ist. Missklassifikation kann in diesem Zusammenhang ebenfalls als Messfehler angesehen werden. Mithilfe einer Übergangsmatrix $\mathbf{\Pi}$ lassen sich die Missklassifikationswahrscheinlichkeiten ausdrücken, wobei die Einträge π_{wx} bzw. π_{xw} für $P(W_i = w|X_i = x)$ – klassischer Messfehler – bzw. $P(X_i = x|W_i = w)$ – Berkson-Messfehler – stehen. Natürlich gibt es daneben noch viele andere Formen von Messfehlern. Hervorzuheben sind dabei multiplikative Messfehler, Mischungen aus klassischem und Berkson-Messfehler sowie Rundung oder Häufung bzw. "Heaping". Kapitel 1 und 2 befassen sich intensiv mit den beiden letztgenannten Phänomenen.

In der Regel sind die Messfehlermodelle abhängig von Parametern wie etwa der Varianz σ_u^2 des Messfehlers U_i im klassischen Messfehlermodell. Generell sind diese nur in wenigen Anwendungen beziehungsweise unter bestimmten Voraussetzungen identifizierbar. Dies beinhaltet Mehrfachmessungen, wie bei dem Geschlecht der prähistorischen Skelette in Kapitel 4 oder den Fall, dass nur ein Teil der Daten messfehlerbehaftet sind oder dass eine zweite, kleinere Studie mit zusätzlichen Evaluierungsdaten durchgeführt wurde. Andernfalls muss der Anwender die entsprechenden Parameter fix vorgeben. Eine Ausnahme stellt das Heaping (Kapitel 2) dar, bei welchem die verschiedenen Rundungswahrscheinlichkeiten nur mithilfe des Endziffernmusters der angegebenen Werte identifizierbar sind. In Kapitel 4 konnte dagegen die Ungenauigkeit der chronologischen Einordnung der Skelette aufgrund von archäologischem Expertenwissen quantifiziert werden, während in Kapitel 1 die Größe des Messfehlers durch dessen künstliche Erzeugung bereits bekannt war.

Viele Techniken zur Korrektur auf Messfehler sind nur für relativ einfache Messfehlermodelle und statistische Verfahren wie die lineare Regression realisierbar. Um komplexere Modelle zu verwirklichen, ist es daher sinnvoll ein komplettes Modell inklusive Likelihood aufzustellen. Sei $\boldsymbol{\theta}$ im Folgenden als der Parametervektor des Modells, welches man bei Kenntnis der wahren Werte $\boldsymbol{X}$ wählen würde, definiert. Die direkte Maximierung der Likelihood $L(\boldsymbol{\theta}|\boldsymbol{W})$ ist außerhalb von einfachen Spezialfällen kaum möglich, da die Berechnung der marginalen Likelihood

$$L(\boldsymbol{\theta}|\boldsymbol{W}) = \pi(\boldsymbol{W}|\boldsymbol{\theta}) = \int_{\boldsymbol{X}} \pi(\boldsymbol{W}, \boldsymbol{X}|\boldsymbol{\theta})\, d\boldsymbol{X}$$

nötig ist und hochdimensionale Integration erfordert. Um das Problem zu vereinfachen, lässt sich aber ausnutzen, dass die Schätzung der Parameter $\boldsymbol{\theta}$ mit üblichen Methoden leicht möglich wäre, falls die wahren Werte $\boldsymbol{X}$ bekannt wären. Wenn ein passendes Messfehlermodell $\pi(\boldsymbol{W}, \boldsymbol{X})$ gefunden werden kann, ist $\pi(\boldsymbol{W}, \boldsymbol{X}|\boldsymbol{\theta})$ dann proportional zu $\pi(\boldsymbol{W}, \boldsymbol{X}) \cdot \pi(\boldsymbol{X}|\boldsymbol{\theta})$. Eine geeignete Vorgehensweise ist es nun in einem iterativen Verfahren zunächst aus der konditionalen Verteilung von $\boldsymbol{X}$ gegeben $\boldsymbol{W}$ und $\boldsymbol{\theta}$ zu ziehen und im nächsten Schritt $\boldsymbol{\theta}$ mithilfe der imputierten $\boldsymbol{X}$-Werte zu schätzen. Diese Strategie ist auch unter dem Stichwort "data augmentation" bekannt. Somit wurde ein einzelnes, schwierig zu lösendes Problem auf zwei relativ einfache Probleme reduziert. Dieser Algorithmus lässt sich im Kontext eines stochastischen Expectation-Maximization-Algorithmus (SEM, Kapitel 1 und 2; Celeux et al. 1996) oder eines voll-Bayesianischen Markov-Chain-Monte-Carlo-Verfahrens (MCMC, Kapitel 3 und 4) umsetzen.

Entsprechend der sehr unterschiedlichen Anwendungsgebiete für Messfehlermodelle wurde die Arbeit in zwei Teile gegliedert. Teil I behandelt zunächst zwei Fragestellungen aus der Survey-Statistik. In Kapitel 1 wurden über einen Rundungsfehler anonymisierte Geokoordinaten der Wohnsitze von Menschen bestimmter Bevölkerungsgruppen in Berlin analysiert. Um eine sinnvolle nichtparametrische Kerndichteschätzung der Populationsverteilung zu erhalten wurde der Rundungsprozess mittels eines stochastischen Expectation-Maximization-Algorithmus umgekehrt. Der entsprechende Artikel erscheint demnächst im "Journal of the Royal Statistical Society" (Serie A). In Kapitel 2 wurde dieser Algorithmus stark erweitert, um die Verteilung von Antworten in Survey-Daten zu modellieren. Die dabei üblicherweise auftretende Häufung von bestimmten Werten wird dabei über eine Rundung mit unbekannter Genauigkeit als Zufallsvariable modelliert. Nach bestem Wissen des Autors ist dies der erste generell anwendbare Ansatz Verteilungen nichtparametrisch im Zusammenhang mit gehäuften Daten zu schätzen. Eine Veröffentlichung ist im "Journal of Survey Statistics and Methodology" geplant. Das Manuskript wurde bereits von dieser Zeitschrift bereits akzeptiert. Beide Methoden wurden auch in Form eines Paketes namens *Kernelheaping* (Groß, 2016a) für die populäre statistische Software *R* veröffentlicht. Dieses umfasst inzwischen auch eine Erweiterung der Methode aus Kapitel 1, bei der auch aggregierte Daten aus beliebig geformten Flächenstücken, wie z.B. den sogenannten Lebensweltlich orientierten Räumen ("LOR") für Berlin, zur Kerndichteschätzung verwendet werden können.

Teil II der Arbeit befasst sich mit den Ergebnissen aus dem Emmy-Noether-Projekt "Lebensbedingungen und biologischer Lebensstandard in der Vorgeschichte" – LiVES. Ein Hauptbestandteil des Projekts war die Zusammenführung von drei Datenbanken prähistorischer Skelette zu einer modernen, web-basierten MySQL-Datenbank. Genauer handelt es sich dabei um die "Mainzer Datenbank", einer in der zweiten Hälfte des vorigen Jahrhunderts entstandenen Lochkartenbank (Perscheid, 1974), einer Microsoft-Access-Datenbank aus einem Vorläuferprojekt von Frau Dr. Rosenstock an der Universität Tübingen sowie der sogenannten ADAM-Datenbank der Universität Genf (Desideri, 2015). Nach Entwicklung eines neuen Datenbankdesigns gemeinsam mit Dr. Eva Rosenstock und Julia Ebert wurden die Daten aller drei Datenbanken in Zusammenar-

beit mit Martin Badicke von der INWT Statistics GmbH in die neue MySQL-Datenbank übertragen. Nachdem alle Einträge auf Fehler (z.B. Dopplungen oder falsche Werte) geprüft und die chronologische Einordnung auf den neuesten Stand gebracht worden sind, soll die Datenbank im Jahre 2018 der Öffentlichkeit zugänglich gemacht werden. Zusätzlich wird die Datenbank laufend um neue Einträge erweitert. In Kapitel 3 und 4 wurden die bereits korrigierten Daten für eine Vorabanalyse genutzt. Hierbei sollte die Forschungsfrage beantwortet werden, wie sich die Körperhöhe als Proxy für den Lebensstandard in der Vorgeschichte entwickelt hat. Die Körperhöhe wird dabei aus den vorhandenen Langknochenmaßen rekonstruiert. Hierbei soll erwähnt werden, dass anthropologische Fragestellungen in der Anfangszeit der modernen Statistik eine große Rolle spielten. Den in der heutigen Statistik allgegenwärtigen Begriff der Regression hat Francis Galton mit einer Untersuchung geprägt, in welcher die Körperhöhe von Eltern und deren Kindern verglichen wurden (Galton, 1886). Ein anderer Gründervater der modernen Statistik, Karl Pearson, hat außerdem bereits eine lineare Regressionsgleichung für die Körperhöhenrekonstruktion entwickelt (Pearson, 1899), die bis heute (u.a. auch in dieser Arbeit) breite Verwendung findet. Der Autor hat in diesem Zusammenhang ein voll-Bayesianisches additives gemischtes Messfehlermodell entwickelt, welches die räumlich-zeitliche Entwicklung der Körperhöhe modelliert. Dabei wurde insbesondere die Unsicherheit bzw. der Messfehler in der chronologischen Einordnung der Skelette als auch die Unsicherheit über das Geschlecht jeweils über ein Berkson-Fehler-Modell berücksichtigt. Das statistische Modell wird im Detail in Kapitel 4 vorgestellt, und wird in Kürze in der Zeitschrift AStA ("Advances in Statistical Analysis") erscheinen. Kapitel 3 befasst sich dagegen mit den technischen und kulturellen Innovationen, welche zu einem verbesserten Lebensstandard respektive Größenwachstum geführt haben sowie den ernährungsphysiologischen Hintergründen. Dieses Kapitel ist als Aufsatz im Berichtsband zum internationalen Workshop "Socio-environmental dynamics over the last 12.000 years: the creation of landscapes III", welcher vom 5. bis 18. April 2013 in Kiel stattfand, erschienen. Kapitel 5 befasst sich mit der Körperhöhenschätzung und der Frage wie sich diese aus den vorhandendenen Langknochen der prähistorischen Skelette optimal schätzen lässt. Ein Messfehler in der abhängigen Variable eines Regressionsmodells – also der Körperhöhe – führt in der Regel zwar zu keiner Verzerrung der Parameterschätzer, aber zu einer verringerten Effizienz. Dies ist in Anbetracht der sehr aufwändigen Datenakquirierung ein durchaus relevantes Thema. Dabei wird die Frage diskutiert, wie man aus der Vielzahl von vorhandenen Formeln zur Körperhöhenrekonstruktion verschiedener Publikationen geeignete Formeln auswählen und gegebenenfalls kombinieren kann. In diesem Zusammenhang werden Körperhöhenschätzungen aus verschiedenen Langknochen mittels AIC-Kriterium optimal gewichtet sowie neue Universalformeln mithilfe einer statistischen Metaanalyse hergeleitet. Diese Strategien sollen in Zukunft für die Schätzung der Körperhöhe der Datenbankeinträge verwendet werden. Diese Arbeit wurde im "American Journal of Physical Anthropology" eingereicht.

Teil I

Messfehlermodelle für die Survey-Statistik

Kapitel 1

Estimating the Density of Ethnic Minorities and Aged People in Berlin: Multivariate Kernel Density Estimation Applied to Sensitive Geo-Referenced Administrative Data Protected via Measurement Error

1.1 Introduction

Modern systems of official statistics require the estimation of area-specific densities of sub-populations. In large cities researchers may be interested in identifying areas with high density of ethnic minorities or areas with high density of aged people. The focus can be even more specific for example, on density estimates of school age children of ethnic minority background. In this work the term ethnic minority will be used to define the part of the population with migration background. Estimates of this type can be used by researchers in Government Departments and other organisations for designing and implementing targeted policies.

To motivate the methodology we propose in this work, we start by presenting two maps in Figure 1.1. The left map presents an estimate of the density of the population of ethnic minority background in Berlin. The right map presents an estimate of the density of the population aged 60 or over in Berlin. The blue points superimposed on the left map show the spatial distribution of advisory centres in Berlin. These are centres that provide assistance for migrants in Berlin. The blue points superimposed on the right map show the spatial distribution of care homes in Berlin. Both kernel density estimation plots in Figure 1.1 have been produced by using real data from the Berlin

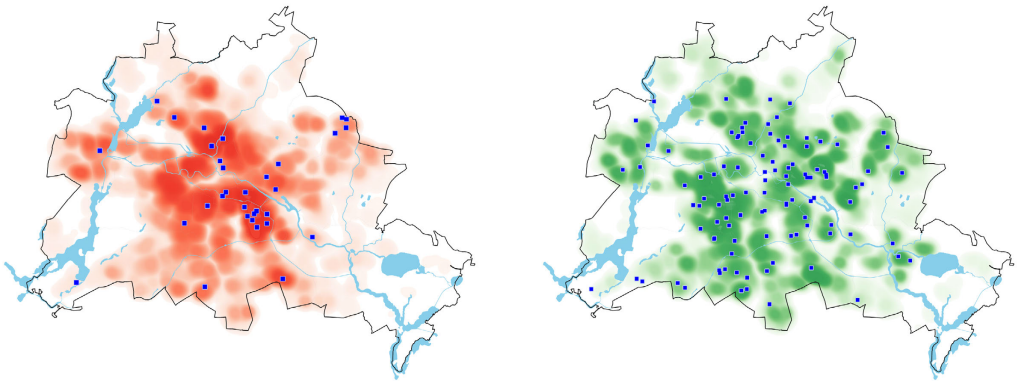

Figure 1.1: Density estimates of the population of ethnic minority background (left map) and of the population aged 60 or above in Berlin (right map). The blue points (left map) show the spatial distribution of advisory centres for migrants. The blue points (right map) show the spatial distribution of care homes.

register, which is a register of residents in all Berlin household addresses that contains exact geo-coded coordinates. At this point we must mention that the register data is available only to the data host in a safe environment. Hence, for producing Figure 1.1 we had to rely on collaborating with staff at the Berlin-Brandenburg Statistics Office who monitored the in-house use of the data. Maps such as those we presented in Figure 1.1 can be very useful for planning purposes. For example, city councils can use the density estimation plots to decide where new advisory centres for migrants are mostly needed or for deciding in which areas to offer planning permissions for opening new care homes. Register databases are updated on a frequent basis and hence their timeliness is better than that of alternative sources of data for example, Census data.

The statistical problem we face in this work is created by the fact that the register with the exact coordinates used for producing the maps in Figure 1.1 is not publicly available. Access to such data is impeded by confidentiality constraints (VanWey et al., 2005) and this holds true also for the Berlin register data. It is easy to see why confidentiality constraints are in place. The availability of precise geo-coding alongside information on demographic characteristics can increase the disclosure risk in particular for sensitive sub-groups of the population such as ethnic minorities. Restricted access to sensitive data may not only apply to users working outside the data host but also to researchers working for the data host or for related organisations for example, Government Departments. As we will see in this work, in the case of the Berlin register data specific procedures are used to ensure confidentiality of the sensitive data. Nevertheless, policies that govern access to sensitive data are country-specific. Other countries that have a long tradition of maintaining register geo-coded data are the Scandinavian ones for example, Norway and Finland. However, access to and use of such data is restricted and these restrictions are decided by the data host in each country.

The host of the data can offer access, possibly in a safe setting, to geo-coded data whilst ensuring confidentiality. One way to achieve this is by introducing measurement error to longitudes and latitudes (Armstrong et al. 1999; Ozonoff et al. 2007 or Rushton et al. 2007). However, this raises the following question. Can we derive precise density estimates of the sub-groups of interest by using data that has been subjected to disclosure control via the introduction of measurement error in the geographic coordinates? The present work proposes non-parametric multivariate density estimation in the presence of measurement error in the geographic coordinates. The aim is to investigate how the precision of density estimates produced by using coarsened data and the use of a non-parametric statistical methodology for reversing the measurement error process compares to density estimates produced by using the exact geo-referenced data. At this point we should make clear that the work does not discuss whether the released geo-referenced information makes identification possible. Instead, we assume that the parameters of the disclosure control process are decided by the data provider. For a discussion on the effectiveness of anonymisation techniques, we refer the reader to Kwan et al. (2004).

Scott and Sheather (1985) used *Naive* density estimation methods that disregard the presence of rounding. To account for rounding Härdle and Scott (1992) introduced a kernel-type estimator based on weighted averages of rounded data points and Minnotte (1998) developed an approach of histogram smoothing. An iterative estimation scheme presented by Blower and Kelsall (2002) ensures non-negative estimates and can potentially be applied to multivariate data as well. A recent publication of Xu (2014) extends this approach to asymmetric kernels. However, the bandwidth selection which is crucial in kernel density estimation is done with a rather ad-hoc approach on the binned data. Wang and Wertelecki (2013) proposed a parametric and a non-parametric kernel density estimator for rounded data but considered only the univariate case. Wang and Wertelecki (2013) showed that using a *Naive* kernel density estimator to rounded data with standard bandwidth selection may lead to poor results for large rounding intervals and large sample sizes.

An alternative idea, explored in this work, is to interpret rounding as a measurement error process and to formulate the problem by using measurement error models (Carroll et al., 2006; Fuller, 2009). For classical additive error models the problem can be regarded as density deconvolution and can be solved using Fourier methods (Stefanski and Carroll, 1990; Zhang, 1990). The topic of density deconvolution has been extensively studied and literature has focused on optimal convergence rates (Fan, 1991), different error distributions such as Gaussian or uniform distributions (Feuerverger et al., 2008) and choice of an optimal bandwidth (Delaigle and Gijbels, 2004). Moreover, the case of additive Berkson errors (Berkson, 1950) in the context of non-parametric density estimation has been investigated. Delaigle (2007, 2014) proposed a density estimator which does not require any bandwidth choice and converges at a parametric rate but with the drawback of producing spiky estimates with high variance when the measurement error is rather low. A recent paper by Long et al. (2014) empirically compares

the estimator of Delaigle (2007, 2014) to two novel approaches for multivariate kernel density estimation contaminated with Gaussian Berkson error and states that one of them shows superior performance. However, rounding error can neither be classified as classical nor Berkson additive error structure as the error is neither independent of the true coordinate nor the rounded one. Nevertheless, a Berkson model with uniform error distribution can be used as an approximation (Wang and Wertelecki, 2013). In this case the estimator by Delaigle (2007) is a bivariate histogram type estimator. When the rounding error, which governs the binwidth, is high the estimator proposed by Delaigle (2007) can be biased. Therefore, in this work we develop a method that correctly specifies the measurement error model under rounding.

From a methodological perspective the present article proposes a novel approach to multivariate non-parametric kernel density estimation in the presence of rounding errors used to ensure data confidentiality. The main advantage of the proposed methodology, compared to alternative methodologies, is that under our approach the bandwidth is derived as part of the estimation process. Moreover, our method is very easy to implement and works regardless of the dimension, the kernel and the bandwidth selection method.

In this work we assume only the availability of register geo-coded data with measurement error in the geographic coordinates. Hence, conventional estimation methods that combine Census/register data with survey data are not applicable in this case. In this work we use the Berlin register data, a complete enumeration of the entire Berlin population in private households, for illustrating how to derive precise density estimates of sensitive groups in the presence of measurement error in two applications.

The first application aims at estimating the density of the Berlin population that is of ethnic minority background. The focus on this application is motivated by the debate on integration/segregation of migrants. Residential segregation describes the phenomenon of a separation of residents according to certain characteristics such as ethnicity. Recent literature suggests that higher levels of segregation are linked with higher crime rates and lower health and educational outcomes (Peterson et al., 2008; Card and Rothstein, 2007; Acevedo-Garcia et al., 2003). To prevent the segregation of ethnic minorities it is necessary to assist these groups with integration programmes offered by advisory centres. Programmes of this kind should be established in areas with high density of ethnic minorities. For the purposes of this application we study the current location of advisory centres in relation to density estimates and identify areas where more support is potentially needed.

The second application relates to the provision of social services for the elderly and urban planing in the context of changing demographics. Longer life expectancy and declining birth rates lead to an ageing population, which needs to be accounted for in urban and social planning. For example, the German National Statistical Institute (Destatis, 2009) predicts the ratio of people over 65 to rise from 20% in 2008 to 34% in 2060. This is a common issue for other industrialised countries too. To ensure the wellbeing of the elderly and to secure adequate and affordable support for this

group it is necessary to analyze where the elderly live. Gorr et al. (2001) used the density of the elderly population as a basis for a spatial decision support system for home-delivered services (meals on wheels). Further challenges arise in urban planing, where an ageing population requires easy access to buildings, public services and public transportation. Shortcomings in urban development can be analyzed by comparing the density of the elderly population against those characteristics (Federation of Finnish Learned Societies' Open Journal Systems, 2014). In addition, many elderly people decide to live in a retirement home. To secure adequate and affordable support for the elderly population it is necessary to establish services where needed. The methodology we propose in this work is also used for providing precise density estimates of the elderly population in the Berlin area. For both applications the sensitivity of density estimation to the severity of the rounding error process is studied and the proposed methodology is contrasted to a *Naive* kernel density estimator which disregards the presence of measurement error.

The structure of the work is as follows. In Section 1.2 we describe the Berlin register data. In Section 1.3 we review multivariate kernel density estimation in the presence of measurement error. A multivariate kernel density estimator is proposed and the computational details of the proposed method are described. In Section 1.4 we present the results of the two applications by using the Berlin register data. In Section 1.5 we empirically evaluate the performance of the proposed methodology under different assumptions for the rounding error process with data generated from known bivariate densities. The precision of the density estimates provided by the proposed methodology is contrasted to the precision of the estimates derived by (a) using a *Naive* kernel density estimator that disregards the presence of rounding error and (b) alternative approaches that have been proposed in the literature. Finally, in Section 1.6 we conclude the work with some final remarks.

1.2 The Berlin Register Data

The statistical problem we face in this work is motivated by the Berlin register of residents dataset, which comprises all Berlin household addresses and contains exact geo-coded coordinates. Such a comprehensive data set is gathered because of German legislation. In particular, registration at the local residents' office is compulsory in Germany and is carried out by the federal state authorities. In the federal city state of Berlin registration is regulated by the Berlin registration law. This law requires every person who moves into a new residential unit in Berlin to be registered in person within one week.

This register is not publicly available because of the detailed geo-coded information it contains. However, a version of the register data is publicly available as part of the Open Data initiative in Berlin (`http://daten.berlin.de`), an initiative that aims at using data for improving urban development. The open dataset includes aggregates for the 447 lowest urban planning areas, the so-called LORs ('Lebensweltlich orientierte Räume'), with coordinates given by the centroid of these areas. This is a discrete and

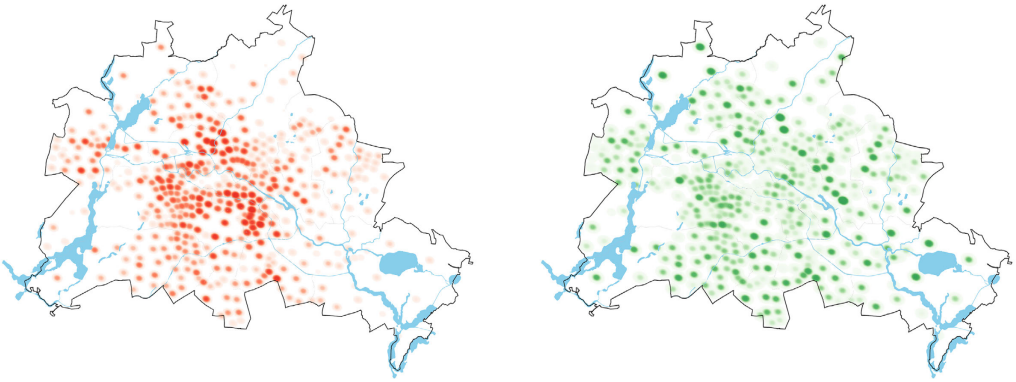

Figure 1.2: Density estimates of the population with ethnic minority background in Berlin (left map) and of the population aged 60 or above in Berlin (right map) based on the publicly available data.

possibly arbitrary demarcation. The discreteness of the demarcation is apparent in Figure 1.2, which shows kernel density estimates of the population of ethnic minorities (left map) and of the population aged 60 or over (right map) in Berlin by using the publicly available data. A main aim of the present work is to derive precise density estimates of population groups by using a more flexible definition of geographic demarcation. This in turn may provide more useful information to local authorities than the currently available LOR demarcation.

An alternative to the currently available data, and one explored by the data host, is to generate a grid-based version of the data that is independent from the somewhat arbitrary geometry of the LORs. In this case the grid-aggregates can be interpreted as the result of rounding geo-coded data for ensuring data confidentiality. Here each point of the grid defines a square-shaped area around the grid point with a longitude and latitude increment equal to the grid length. Then the value of the variable of interest is the aggregate of the values with exact geo-coordinates over the area surrounding the grid point. In fact, the LOR demarcation in Berlin can be thought of as the process of rounding the geo-referenced data by using grids of average size 2000 meters by 2000 meters. The methodology we propose in this work attempts to reverse the rounding process for deriving estimates that are more precise than density estimates that ignore the measurement error process and relate to a more flexible definition of geographic demarcation.

The data that we have access to in this work contains all 308,754 Berlin household addresses on the 31st of December 2012 with the exact geo-coded coordinates subject to different degrees of rounding error. One of the scenarios we explore is rounding by using grids of size 2000 meters by 2000 meters that approximately correspond to the LOR demarcation. The location is measured by (Soldner)-coordinates in meters. The original (without rounding error) data includes the total number of residents (Berlin

Table 1.1: Summary statistics of the number of residents living at a household address.

	Sum	Min.	1st Qu.	Median	Mean	3rd Qu.
Berlin Total	3,469,619	1	2	4	11.24	15
Migration	949,184	0	0	0	3.07	3
Migration Vietnam	21,637	0	0	0	0.07	0
Migration Turkey	176,738	0	0	0	0.57	0
Age over 60	859,170	0	0	1	2.78	3

Total) at their principal residence and the number of persons according to some key demographic characteristics. The first demographic variable is the migration background (Migration) of individuals defined by the number of people that are of (a) non-German nationality, (b) German nationality but born abroad and (c) non-German nationality who changed their nationality into German at the coordinates of the principal household address. The definition of this variable is further refined by the number of individuals of migration background from Turkey (Migration Turkey) or Vietnam (Migration Vietnam). The second demographic variable is age (age over 60) defined by the number of individuals who are older than 60 years old. The density estimates of the subgroups of interest that are produced by using the proposed methodology are contrasted to maps of the corresponding densities produced by using the data with the exact geo-coded coordinates. The use of these maps has been approved by the data host, the Berlin-Brandenburg Statistics Office.

Table 1.1 presents summary statistics of the number of residents living at a household address of the key variables based on the exact geo-coded data. Due to confidentiality restrictions we are not allowed to publish the maximum number of residents living at a household address. The average of individuals living at a household address in Berlin is 11.24 leading to a total population of 3,469,619 (registered) inhabitants. Note that a household address in the data is defined for example, as an entire block of apartments. Around 27% of the total population are of migration background and around 24.8% of the population are older than 60 years. The average number of residents of migration background is 3.07 with a median of 0, whereas the average number of individuals above 60 years of age is 2.78 with a median of 1. This gives a first indication that inhabitants with migration background are more clustered compared to older people in Berlin.

1.3 Multivariate Kernel Density Estimation in the Presence of Measurement Error

In this section we propose an approach to non-parametric multivariate density estimation in the presence of measurement error in particular, rounding of the geographical coordinates used for disclosure control of sensitive data. Multivariate kernel density estimation is introduced in Section 1.3.1. In Section 1.3.2 we investigate kernel density estimation in the presence of measurement error and in Section 1.3.3 we present a model that corrects for measurement error in multivariate kernel density estimation.

Estimation and the computational details of the algorithm we use for implementing the proposed model are described in Section 1.3.4.

1.3.1 Multivariate kernel density estimation

Kernel density estimation as a non-parametric approach is an important tool in exploratory data analysis. Multivariate kernel density estimation attempts to estimate the joint probability distribution for two or more continuous variables. This method has the advantage of producing smooth density estimates compared to a histogram whose appearance heavily depends on the bin's breakpoints. Let $\boldsymbol{X} = \{\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}\}$ denote a sample of size n from a multivariate random variable with unknown density $f(\boldsymbol{x})$. In the following, we only consider the two-dimensional case without loss of generality such that $\boldsymbol{x} = (x_1, x_2)$. Thus, $\boldsymbol{X}_i$, $i = 1, \ldots, n$ is given by (X_{i1}, X_{i2}), where – in our application – X_{i1} and X_{i2} denote longitude- and latitude- coordinates, respectively.

The multivariate kernel density estimator at point $\boldsymbol{x}$ is given by

$$\hat{f}_{\boldsymbol{H}}(\boldsymbol{x}) = \frac{1}{n|\boldsymbol{H}|^{\frac{1}{2}}} \sum_{i=1}^{n} K\left(\boldsymbol{H}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{X_i})\right), \tag{1.1}$$

where $K(\cdot)$ is a multivariate kernel function, $\boldsymbol{H}$ denotes a symmetric positive definite bandwidth matrix and $|\cdot|$ denotes the determinant. A standard choice for $K(\cdot)$, used throughout this work, is the multivariate Gaussian kernel. The choice of bandwidth $\boldsymbol{H}$ is crucial for the performance of a kernel density estimator. Approaches for bandwidth selection have been widely discussed in the literature. A popular strategy is to choose $\boldsymbol{H}$ by minimizing the asymptotic mean integrated squared error (AMISE) through plug-in or cross-validation methods (Izenman, 1991 or Silverman, 1986). In the univariate case we refer the reader to Marron (1987) or Jones et al. (1996). Wand and Jones (1994) discussed the choice of the bandwidth in the multivariate case by using a plug-in estimator. The approach by Wand and Jones (1994) is the one we use for bandwidth selection in this work.

1.3.2 Rounding and kernel density estimation

By introducing rounding for achieving anonymisation of sensitive data the true values $\boldsymbol{X} = \{\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}\}$, the exact geographical coordinates, are lost. Instead, only the rounded (contaminated by measurement error) values, denoted by $\boldsymbol{W} = \{\boldsymbol{W_1}, \boldsymbol{W_2}, \ldots, \boldsymbol{W_n}\}$, are available. As a consequence the data is concentrated on a grid of points. Using a *Naive* kernel density estimator that ignores the rounding process by replacing the true values $\boldsymbol{X_i}$ by the rounded values $\boldsymbol{W_i}$ in (1.1) may lead to a spiky density that is not close to the density of the uncontaminated (true) data. This effect becomes more pronounced with increasing sample size. In particular, as the bandwidth determinant $|\boldsymbol{H}|$ is decreasing with higher sample size this causes higher density estimates on the grid points and lower in between the grid points.

The process of rounding means that the true, unknown, values $\boldsymbol{X}_i = (X_{i1}, X_{i2})$

given the rounded values $\boldsymbol{W_i} = (W_{i1}, W_{i2})$ are distributed in a rectangle with $\boldsymbol{W_i}$ in its center,

$$\left[W_{i1} - \frac{1}{2}r, W_{i1} + \frac{1}{2}r\right] \times \left[W_{i2} - \frac{1}{2}r, W_{i2} + \frac{1}{2}r\right]. \tag{1.2}$$

The value r denotes the rounding parameter. For instance, the data is rounded to the next integer for $r = 1$.

1.3.3 The model

A model for the density $f(\boldsymbol{x})$ could be formulated parametrically, for example by a multivariate Gaussian distribution, or non-parametrically either by a mixture of parametric distributions (Escobar and West, 1995; Gelfand et al., 2005) or by using multivariate kernel density estimation as introduced in Section 1.3.1. As discussed in Section 1.3.2, the true values $\boldsymbol{X_i}$ are lost because of the rounding process and only the rounded values $\boldsymbol{W_i}$ are observed. However, we still aim to estimate the density $f(\boldsymbol{x})$ – from which our sample $\boldsymbol{X}$ is drawn – only by using the rounded values $\boldsymbol{W_i}$. Under the assumption that the rounding/anonymisation process of the $\boldsymbol{X_i}$ is known, we are able to formulate a measurement error model $\pi(\boldsymbol{W}|\boldsymbol{X})$ for $\boldsymbol{W}$. In particular, the measurement error model $\pi(\boldsymbol{W}|\boldsymbol{X})$ for rounding is defined by a product of Dirac distributions, $\pi(\boldsymbol{W}|\boldsymbol{X}) = \prod_{i=1}^{n} \pi(\boldsymbol{W_i}|\boldsymbol{X_i})$, with

$$\pi(\boldsymbol{W_i}|\boldsymbol{X_i}) = \begin{cases} 1 & \text{for } \boldsymbol{X_i} \in [W_{i1} - \frac{1}{2}r, W_{i1} + \frac{1}{2}r] \times [W_{i2} - \frac{1}{2}r, W_{i2} + \frac{1}{2}r] \\ 0 & \text{else.} \end{cases} \tag{1.3}$$

From the Bayes theorem it follows that $\pi(\boldsymbol{X}|\boldsymbol{W}) \propto \pi(\boldsymbol{W}|\boldsymbol{X})\pi(\boldsymbol{X})$. Utilizing this formulation we can draw pseudo samples (imputations) of the $\boldsymbol{X_i}$ from $\pi(\boldsymbol{X_i}|\boldsymbol{W_i})$ which enables us to estimate $f(\boldsymbol{x})$. As $\pi(\boldsymbol{X}) = \prod_{i=1}^{n} f(\boldsymbol{X_i})$ is initially unknown we propose an iterative procedure, which uses an initial estimate of $f(\boldsymbol{x})$ based on the $\boldsymbol{W_i}$ followed by alternating simulations of $\boldsymbol{X}$ from $\pi(\boldsymbol{X}|\boldsymbol{W})$ and re-estimation of $\pi(\boldsymbol{X})$ until convergence. The following subsection gives further details about the exact implementation of the algorithm and discusses how this can be viewed as a variant of the Expectation-Maximization (EM) algorithm (Dempster et al., 1977).

1.3.4 Estimation and computational details

As discussed in the previous subsection, for fitting the model we need to draw pseudo samples of the $\boldsymbol{X_i}$. The conditional distribution of the $\boldsymbol{X_i}$ given the rounded values $\boldsymbol{W_i}$ is the following:

$$\pi(\boldsymbol{X_i}|\boldsymbol{W_i}) \propto I(W_{i1} - \frac{1}{2}r \leq X_{i1} \leq W_{i1} + \frac{1}{2}r) \times I(W_{i2} - \frac{1}{2}r \leq X_{i2} \leq W_{i2} + \frac{1}{2}r) \times f(\boldsymbol{X_i}), \tag{1.4}$$

where $I(\cdot)$ denotes the indicator function. The conditional distribution of $\boldsymbol{X_i}$ is the product of a uniform distribution on the square with side length r around $\boldsymbol{W_i}$ and

density $f(\boldsymbol{x})$. As the density $f(\boldsymbol{x})$ is unknown it is replaced by an estimate, which is the multivariate kernel density estimator $\hat{f}_{\boldsymbol{H}}(\boldsymbol{x})$ defined in (1.1). In particular, $\boldsymbol{X_i}$ is repeatedly drawn from the square of side length r around $\boldsymbol{W_i}$ using the current density estimate $\hat{f}_{\boldsymbol{H}}(\boldsymbol{x})$ as a sampling weight. The steps of the algorithm are described below.

1. Get a pilot estimate of $f(\boldsymbol{x})$ by setting $\boldsymbol{H}$ to $\begin{pmatrix} l & 0 \\ 0 & l \end{pmatrix}$, where l is a sufficiently *large* value such that no rounding spikes occur.

2. Evaluate the density estimate $\hat{f}_{\boldsymbol{H}(\boldsymbol{x})}$ on an equally-spaced fine grid $\boldsymbol{G} = \boldsymbol{z_1} \times \boldsymbol{z_2}$ (with $\boldsymbol{G} = \{g_1, \ldots, g_m\}$, gridwidth δ_g and
$\boldsymbol{z_1} = \left\{ \min_i(W_{i1}) - \frac{1}{2}r, \min_i(W_{i1}) - \frac{1}{2}r + \delta_g, \ldots, \max_i(W_{i1}) + \frac{1}{2}r \right\}$,
$\boldsymbol{z_2} = \left\{ \min_i(W_{i2}) - \frac{1}{2}r, \min_i(W_{i2}) - \frac{1}{2}r + \delta_g, \ldots, \max_i(W_{i2}) + \frac{1}{2}r \right\}$ $(i = 1, \ldots, n))$,
where r denotes the rounding parameter introduced in Section 1.3.2.

3. Sample from $\pi(\boldsymbol{X_i}|\boldsymbol{W_i})$ by drawing a sample ${\boldsymbol{X_i}}^S = \left(X_{1i}^S, X_{2i}^S\right)$ randomly from $\left(\boldsymbol{z_1} \in [W_{i1} - \frac{1}{2}r, W_{i1} + \frac{1}{2}r]\right) \times \left(\boldsymbol{z_2} \in [W_{i2} - \frac{1}{2}r, W_{i2} + \frac{1}{2}r]\right)$ with sampling weight $\hat{f}_{\boldsymbol{H}({\boldsymbol{X_i}}^S)}$, $i = 1, 2, \ldots, n$.

4. Estimate the bandwidth matrix $\boldsymbol{H}$ by the multivariate plug-in estimator of Wand and Jones (1994) and recompute $\hat{f}_{\boldsymbol{H}}(\boldsymbol{x})$. Here we should mention that other bandwidth selectors are applicable.

5. Repeat steps 2-4 B (burn-in iterations) $+ N$ (additional iterations) times.

6. Discard the B burn-in density estimates and get the final density estimate of $f(x)$ by averaging the remaining N density estimates $\hat{f}_{\boldsymbol{H}}(\boldsymbol{x})$ on the evaluation grid $\boldsymbol{G}$.

The prospective reader may ask how the algorithm fits into existing estimation frameworks. Generally, a popular fitting algorithm for models that depend on latent, unobserved data (the $\boldsymbol{X_i}$ values in our case) is the Expectation-Maximization (EM) algorithm. The proposed algorithm is a variant of the classical EM algorithm, namely the Stochastic Expectation-Maximization (SEM) algorithm (Celeux et al., 1996). The SEM algorithm works by drawing samples from the conditional distribution $\pi(\boldsymbol{X_i}|\boldsymbol{W_i})$ creating a pseudo sample of $\boldsymbol{X}$ in each iteration as a replacement of the E-step in the classical EM algorithm where the conditional expectation of $\boldsymbol{X_i}$ given $\boldsymbol{W_i}$ is computed analytically. The classical EM approach would clearly not work for kernel density estimation with rounded data because all the observations within the rectangle around $\boldsymbol{W_i}$ would still be concentrated at a single point, namely the expectation of the conditional distribution of $\boldsymbol{X_i}$ given $\boldsymbol{W_i}$, computed in the E-Step, leading to spiky estimates of the density. In its original form both the EM and SEM algorithm are used for maximum likelihood estimation in the presence of unobserved variables. However, kernel density estimation is a non-parametric method. We therefore utilize a generalization of the SEM algorithm for the use of surrogates of the likelihood in the M-step (McLachlan

and Krishnan, 2007) such that the objective of maximization, i.e. the likelihood, is replaced by the minimization of the AMISE of the kernel density estimator in our case.

The estimator we propose in this work – hereafter referred to as $GRSST$ estimator – allows for estimating the bandwidth matrix $\boldsymbol{H}$ simultaneously with the density. In contrast, for the algorithm proposed by Blower and Kelsall (2002) it is not immediately clear how to estimate $\boldsymbol{H}$. Blower and Kelsall (2002) suggest using an initial estimate based on the rounded data. Another advantage is that with the proposed algorithm we can get an estimate of the variance induced by the rounding process. This is obtained as a byproduct of the Monte-Carlo process. In particular, standard errors for the density estimates at some arbitrary point can be computed by using the $\hat{f}_{\boldsymbol{H}}(\boldsymbol{x})$ produced in each iteration of the algorithm. The algorithm we propose in this work is also linked to the one proposed by Wang and Wertelecki (2013) in the univariate case. Apart from being derived only for the univariate case, the approach by Wang and Wertelecki (2013) corresponds (in the univariate case) to the method we propose in this work with $B = 0$ burn-in iterations and $N = 1$ or more sampling steps. However, without a burn-in period no convergence is achieved and final estimates can heavily depend on the pilot estimate. The influence of the burn-in iterations and the sampling steps on the quality of density estimation is evaluated in a simulation study the results of which are included as part of the supporting information. The algorithm is implemented by using function *dbivr* in the **Kernelheaping** R package (Groß, 2016a), which is available on CRAN. Additionally, the proposed approach allows for the use of an adaptive bandwidth selection method proposed by Davies et al. (2011) and is implemented in the **sparr** package.

1.4 Analysis of the Berlin Register of Residents

The benefits of using the proposed multivariate kernel density estimator that accounts for measurement error are illustrated in two applications both of which use the Berlin register data we described in Section 1.2. The first application aims at estimating the density of the population with migration background in Berlin. The density estimates are compared to the current geographical distribution of advisory centres for migrants in Berlin. The second application aims at estimating the density of the population aged 60 and above in the Berlin area. The density estimates are compared to the current geographical distribution of care homes in the Berlin area.

The analysis is carried out by using the two variables (a) Migration and (b) Age over 60. The setup of the analysis is as follows: To start with, we impose grids on the geographical space of the Berlin data set with respective grid sizes of 250, 500, 1250, 2000 and 2500 meters. The grid sizes correspond to different degrees of measurement error used for anonymisation purposes. Note that the use of the 2000m by 2000m grids is because these are of similar size to the currently used urban planning areas in Berlin a level at which data is publicly available. Subsequently, we estimate the density of the target population by using the $Naive$ and the proposed $GRSST$ density estimators for each of the grid sizes. We use $B = 5$ and $N = 20$ iterations for the proposed

Table 1.2: Berlin register data: RMISE for *Naive* and *GRSST* multivariate kernel density estimators for different grid sizes (results in units of 10^{-8})

	$r = 250m$		$r = 500m$		$r = 1250m$		$r = 2000m$		$r = 2500m$	
Variable	*Naive*	*GRSST*	*Naive*	*GRSST*	*Naive*	*GRSST*	*Naive*	*GRSST*	*Naive*	*GRSST*
Age above 60	0.66	0.67	1.32	1.27	4.52	2.46	14.08	4.06	23.34	4.66
Migration	0.98	0.97	1.98	1.84	7.33	3.43	22.07	6.12	36.94	6.31

GRSST method in the algorithm presented in Section 1.3.4. The sensitivity of the density estimators to the size of the dataset, (n) and the effect of the burn-in size, (B) and sample steps (N) is assessed in the supporting information.

The performance of a generic density estimator $\hat{f}(\boldsymbol{x})$, for example the *Naive* or the *GRSST*, is typically evaluated by the root mean integrated squared error (RMISE), which is approximated by a Riemann sum over an equally-spaced fine grid,

$$\text{RMISE}(\hat{f}(\boldsymbol{x})) = \sqrt{E\left(\int (f(\boldsymbol{x}) - \hat{f}(\boldsymbol{x}))^2 d\boldsymbol{x}\right)} \approx \sqrt{\frac{1}{m}\sum_{j=1}^{m}(f(\boldsymbol{g_j}) - \hat{f}(\boldsymbol{g_j}))^2\delta_g^2}, \quad (1.5)$$

where m is the number of grid points $\boldsymbol{g_j}$ and δ_g is the gridwidth. For computing the *Naive* estimator and the *GRSST* estimator (using the algorithm in Section 1.3.4) we use a bivariate Gaussian kernel and the plug-in bandwidth selector of Wand and Jones (1994). This is implemented by using the R functions *kde* (kernel density estimation) and *Hpi* (bandwidth selector) provided by the **ks** package (Duong, 2014). The unobserved *true* density $f(\boldsymbol{x})$ is substituted by the kernel density estimator (1.1) that uses the original data without rounding with bivariate Gaussian kernel and the plug-in bandwidth selector. This is treated as a benchmark because it is not affected by rounding error. At this point we must mention that the original data is available only to the data host. Hence, for implementing the code with the original data we had to collaborate with staff at the Berlin-Brandenburg Statistics Office. Table 1.2 shows the goodness of fit in terms of RMISE for the *Naive* and the proposed density estimators and for different grid sizes. Figures 1.3 and 1.4 present kernel density estimation plots for selected grid sizes for Age over 60 and Migration respectively. To start, we note that the proposed estimator outperforms the *Naive* estimator especially for large grid sizes ($\geq$ 1250m). For grid sizes larger or equal to 1250m the *Naive* estimator produces small spikes at the location of the grid points since in this case the probability mass is mostly attributed to the center points of the grid. In contrast, the proposed estimator preserves the fundamental structure of the underlying density. For the largest grid size (2500m), which implies strongly anonymised data, the general shape produced with the proposed estimator is clearly visible. This is not the case with the *Naive* estimator.

Having assessed the performance of both estimators, we now discuss the results of the density estimates in the context of two applications.

Advisory services for population with migration background: Around 950,000 people with migration background from around 190 countries live in the 12 districts in Berlin. The four largest communities consist of approximately 200,000 people with

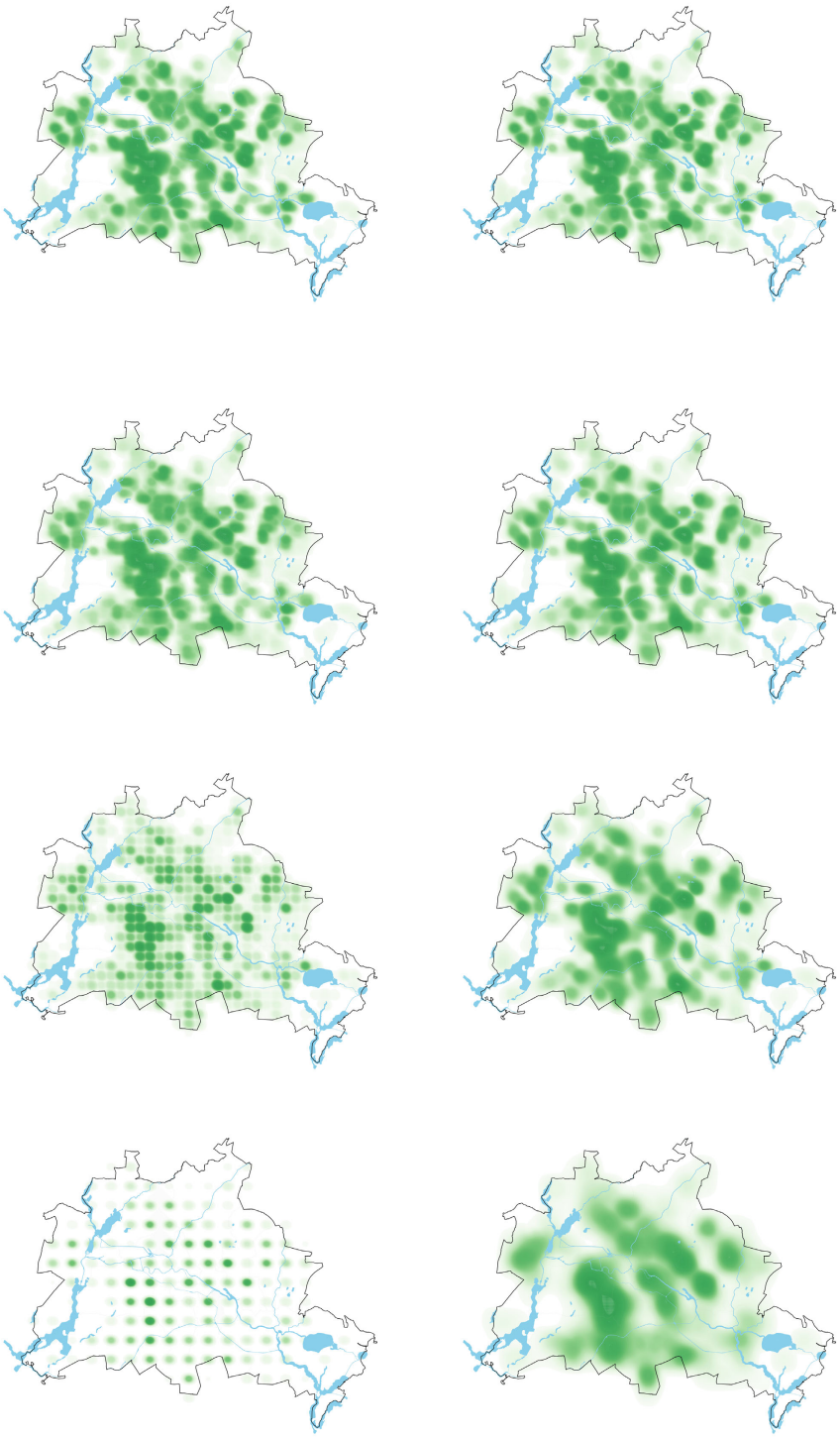

Figure 1.3: Density estimates of population aged 60 and above: *Naive* (left panel) and *GRSST* estimators (right panel) with rounding step sizes of 0 (original data), 500, 1250 and 2500 m (top down).

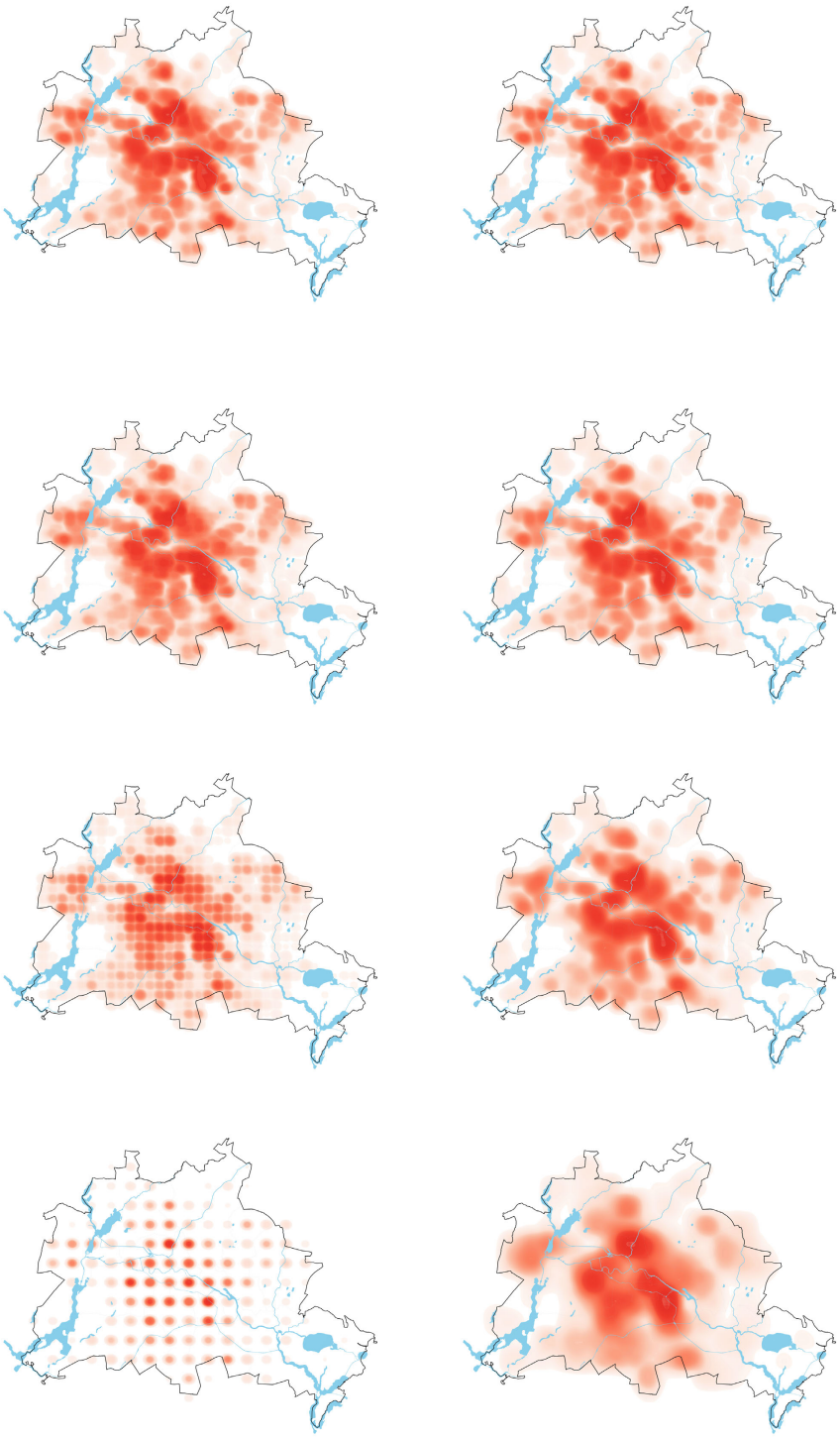

Figure 1.4: Density estimates of population with migration background: *Naive* (left panel) and *GRSST* estimators (right panel) with rounding step sizes of 0 (original data), 500, 1250 and 2500 m (top down).

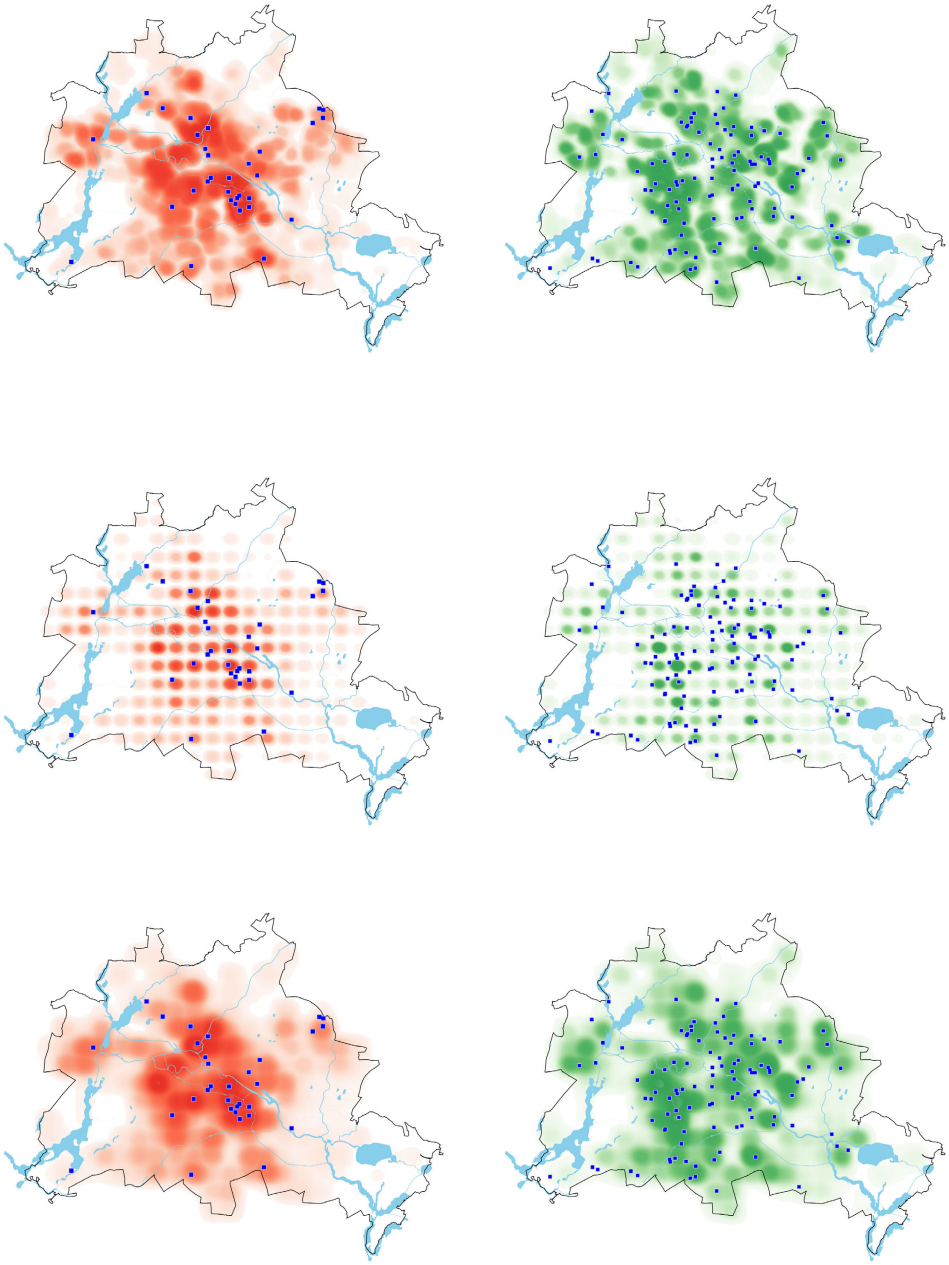

Figure 1.5: Migration background (left panel) and Age above 60 (right panel) for the original data, *Naive* method and *GRSST* method (top down) for rounding step size of 2000 m including points of interest. Blue points indicate migrant advisory centers and retirement houses respectively.

Turkish migration background, around 100,000 people from Russia or from the former Soviet Union and its successor states, approximately 60,000 people of migration background from the successor states in the former Yugoslavia and around 45,000 people of Polish migration background. The history of many migrants started in former West Berlin in the mid-sixties with the recruitment of guest workers. Workers were recruited mainly from Mediterranean countries like Greece, Italy, Yugoslavia or Turkey. In the former East Berlin workers were employed by inter-state agreements from countries like Angola, Poland or Vietnam. From the very beginning Berlin offered advisory services for migrants. For instance, Berlin has a commissioner for integration and migration. This office was established in 1981 and was the first of its kind in Germany. Nowadays, there are specialized advisory service centres that assist people with migration background. The youth migration services provide advice to young adults and teenagers of migration background. In addition, Berlin has in total 32 advisory service centres for adults. In these centres migrants can receive support and personal consultation directly that will assist with their integration. For example, people receive support with finding appropriate child care facilities. To secure an appropriate level of support it is important to establish advisory centres where mostly needed. The left panel of Figure 1.5 shows the estimated densities of the population with migration background in Berlin. The blue points represent the 32 advisory service centres for adults. The plot on the top panel shows the density estimates produced by using the original data and the exact address coordinates, which are not publicly available. The plots in the middle and at the bottom present density estimates produced by using the *Naive* and the *GRSST* density estimators with a rounding step size of 2000m. The choice of 2000m times 2000m grids is because these are of similar size to the currently used urban planning areas in Berlin. The estimates based on the original data in Figure 1.5 show that the density of populations with migration background varies by Berlin districts. The estimated density is particularly high in the former West-Berlin districts of Wedding (in the north), Neukölln (in the south-east), Kreuzberg (in the center to south) and Schöneberg (in the south). Friedrichshain and Prenzlauer Berg (in the north-east), show a lower estimated density of population with migration backgrounds.

The spatial distribution of advisory centres cover populations in the centre and north of Berlin quite well. However, there are some hotspots for example, in the western and south-west parts (Charlottenburg or Moabit) or in the very northern parts (Märkisches Viertel) of Berlin, with a high density estimate of ethnic minority populations but without any advisory service centres. The commentary on the first map above depends on precise geo-coded addresses which are not publicly available. The second and third maps show the density estimates based on the rounded data. The density plot obtained by using the *Naive* estimator (plot in the middle in Figure 1.5) produces spikes at the center of the grids. In contrast, the proposed estimator produces a map (plot at the bottom in Figure 1.5) that is able to preserve the fundamental density structure of the original data. Hence, the commentary we produced by looking at the map of the original data holds also true for the map of density estimates pro-

duced by using the proposed multivariate kernel density estimator that accounts for measurement error. In addition, the proposed density estimator produces more precise density estimates than the *Naive* one (see Table 1.2). Local authorities should prefer the density estimates produced by the proposed estimator, to the one produced by the *Naive* estimator, for making informed decisions.

Care for the elderly: Life expectancy in Germany has improved due to advances in medical research. This leads to a change in the demographic structure with an increasing number of old-aged people. Approximately 860,000 individuals aged 60 and above live in Berlin. It is projected that by 2030 the average age of Berlin's population will increase from 42.5 years (in 2007) to 45.3 years and roughly every third citizen of Berlin will be 60 years or older. With increasing age the prevalence of diseases and functional restraints, which are often chronic and irreversible, rises as well (Saß et al., 2009). In 2012, 58.3% of German women and 55.3% of German men suffered from at least one chronic disease (Robert Koch Institut, 2014). According to the World Health Organization (2005), the prevalence and incidence of various chronic diseases, such as cardiovascular diseases, cancer, diabetes mellitus, dementia or respiratory problems, is predicted to increase in the next years. For this reason older people are more likely to need help in their daily life and will increasingly depend on care. According to the nursing care insurance in 2011 there were roughly 117,500 care-dependent people in Berlin. In order to support the increasing elderly population it is necessary to offer high-quality medical and social community structures of care that are close to the people's place of residence. This is important because elderly people tend to feel connected to their neighbourhood. These structures consist of:

- Neighborhood centers: These are combinations of accessible living, residential care homes and social/cultural centres with neighbourhood cafes, which are suitable for senior citizens. Such structures offer elderly people with or without care dependency the opportunity to live actively within the community until old age.

- Foster ambulatory care: These are home care nursing services that enable care-dependent people to live at home.

- Networked care: The different forms of care systems (e.g., ambulatory care, semi-residential care or impatient care) need to be more strongly interconnected than they currently are. This will offer more choices for older people for example, live at home with ambulatory care but have the opportunity to change to semi-residential or impatient care near to the place they live.

In order to improve such services for the city of Berlin it is necessary to have an accurate picture about the distribution of the elderly population in Berlin. The right panel of Figure 1.5 shows the density estimates for the population aged 60 years or above. The blue points represent 108 retirement homes in Berlin. The location of these points was extracted by using Google Maps. The plot on the top panel indicates the density estimates based on the original data with the exact address coordinates, which are not publicly available. The plots in the middle and at the bottom present

the density estimates by using the *Naive* and the proposed density estimators with a rounding step size of 2000m. The supply of retirement houses is particularly good in the center of Berlin. However, locations for future expansion of retirement houses and other support structures can be identified. For instance, there are some hotspot areas in the north (Reinickendorf and especially Märkisches Viertel) or in the south-east (Gropiusstadt) with a high density estimate of the population over 60 but without retirement homes. As in the first application,the proposed estimator (plot at the bottom in Figure 1.5) preserves the structure of the density of the population over 60 years despite the presence of measurement error in the available data and offers more precise estimates. Hence, the use of the proposed estimator may enable local authorities and other organisations to make sound strategic decisions regarding the best places for investigating in creating infrastructure for social care without requiring access to exact geo-referenced data. A more refined analysis of the Berlin register data could consider the use of local bandwidths as opposed to a global bandwidth. This is possible by using the R package that has been written for implementing the methodology we propose in this work. Nevertheless, use of local bandwidths can increase significantly the computational time.

1.5 Simulation Study

In this section we present results from a Monte-Carlo simulation study that was conducted for evaluating the performance of the proposed multivariate kernel density estimator we presented in Section 1.3. The objective of this simulation study is to investigate the ability of the proposed methodology to account for measurement error, under different scenarios for the intensity of the measurement error process, and hence provide more precise estimates than *Naive* kernel density estimation that disregards measurement error. The proposed estimator is further compared to the estimator proposed by Delaigle (2007). Finally, the sensitivity of the proposed method in relation to the size of the data (n), to the burn-in size (B) and sample steps (N) used in the *GRSST* algorithm is evaluated and the results are provided as part of the supporting information.

The simulation data is generated under different bivariate normal distributions. Three scenarios, denoted by A, B and C, are considered. Under Scenario A data is generated by using a bivariate standard normal distribution,

$$f_A(\boldsymbol{x}) = \phi(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\phi(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate normal density with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ given by,

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ , \ \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Under Scenario B data is generated by using a mixture of three uncorrelated bivariate

normal distributions,

$$f_B(\boldsymbol{x}) = \frac{1}{3}\phi(\boldsymbol{x}|\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}) + \frac{1}{3}\phi(\boldsymbol{x}|\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2}) + \frac{1}{3}\phi(\boldsymbol{x}|\boldsymbol{\mu_3}, \boldsymbol{\Sigma_3}),$$

with

$$\boldsymbol{\mu_1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\mu_2} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \boldsymbol{\mu_3} = \begin{pmatrix} -4 \\ 1 \end{pmatrix}, \boldsymbol{\Sigma_1} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \boldsymbol{\Sigma_2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{\Sigma_3} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Finally, under Scenario C data is generated by using a mixture of three correlated normal distributions with

$$\boldsymbol{\mu_1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\mu_2} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \boldsymbol{\mu_3} = \begin{pmatrix} -4 \\ 1 \end{pmatrix}, \boldsymbol{\Sigma_1} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \boldsymbol{\Sigma_2} = \begin{pmatrix} 3 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \boldsymbol{\Sigma_3} = \begin{pmatrix} 5 & 4 \\ 4 & 6 \end{pmatrix}.$$

The corresponding density contours under the three scenarios are shown in Figure 1.6. The use of bivariate distributions is motivated by the fact that our application data in Section 1.4 is bivariate. The use of Gaussian distributions for generating the simulation data follows Zhang et al. (2006) and Zougab et al. (2014).

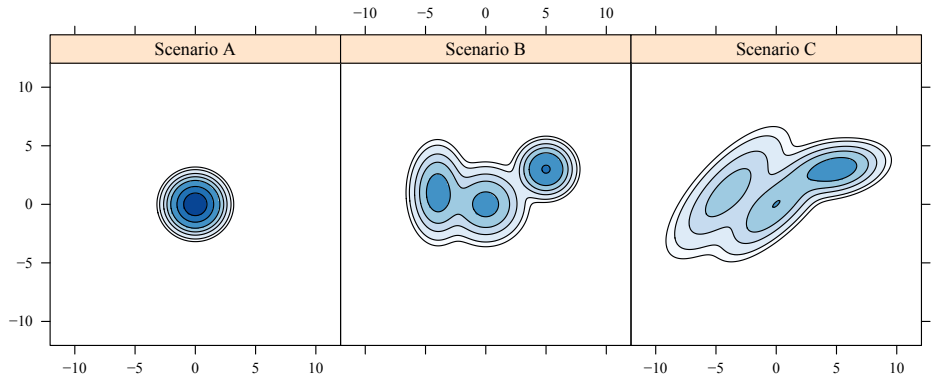

Figure 1.6: Contour plots of the simulated data under the three simulation scenarios.

For each scenario we generate a dataset S_0 of size $n = 500$ from the corresponding distribution $f_A(\boldsymbol{x}), f_B(\boldsymbol{x})$ or $f_C(\boldsymbol{x})$. The dataset S_0 includes the exact x- and y-coordinates. For introducing measurement error via rounding of the coordinates, we define a grid for the x- and y-coordinates ranging from -10 to 10 with gridwidth according to rounding values $r = 0.75, 1.5$ and 2.25. For a formal definition of r and the rounding process we refer to Section 1.3.2. We denote the dataset after rounding by S_r. Figure 1.7 shows the different scenarios for the rounding process for a specific dataset under Scenario B. The size of the points represents the number of points at a specific rounding tick.

By using S_r, we estimate the density with three methods: a) *Naive*: a standard kernel density estimator that ignores measurement error, b) *GRSST*: This is the proposed SEM estimator with $B = 5$ burn-in and $N = 20$ sample steps and c) Delaigle:

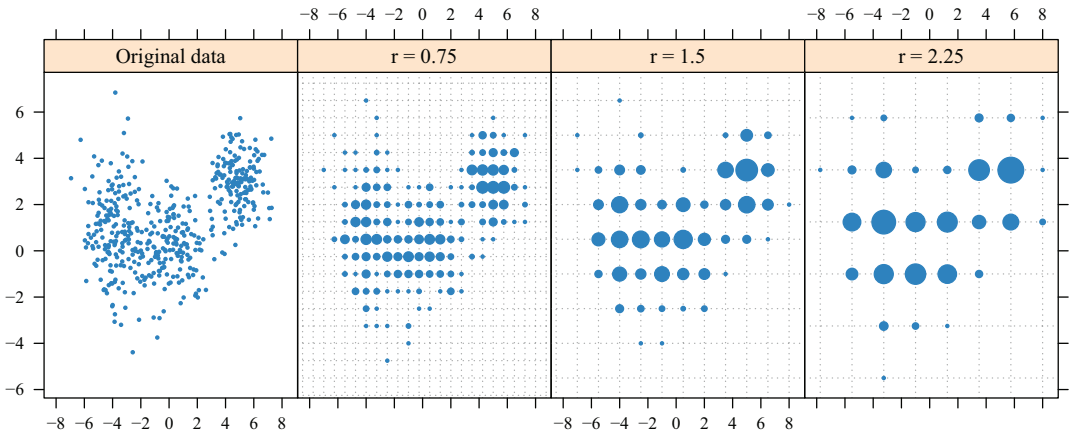

Figure 1.7: Scenario B: Rounding procedure for a specific dataset.

Table 1.3: Mean RMISE for different grid sizes (r) and scenarios. Corresponding standard errors of the RMISE in parentheses.

	$r = 0$	$r = 0.75$			$r = 1.5$			$r = 2.25$		
	Original	*Naive*	*GRSST*	Delaigle	*Naive*	*GRSST*	Delaigle	*Naive*	*GRSST*	Delaigle
Scenario A	0.205	0.238	0.239	0.301	3.952	0.242	0.887	4.917	0.568	1.113
	(0.026)	(0.029)	(0.031)	(0.027)	(0.301)	(0.030)	(0.034)	(0.248)	(0.045)	(0.056)
Scenario B	0.162	0.172	0.170	0.328	0.380	0.183	0.272	0.679	0.256	0.390
	(0.016)	(0.017)	(0.016)	(0.019)	(0.033)	(0.018)	(0.013)	(0.043)	(0.016)	(0.014)
Scenario C	0.119	0.125	0.121	0.268	0.147	0.131	0.181	0.351	0.152	0.172
	(0.012)	(0.013)	(0.012)	(0.015)	(0.013)	(0.013)	(0.009)	(0.034)	(0.014)	(0.012)

this is the estimator presented in Delaigle (2007). As in Section 1.4, for computing the *Naive* and *GRSST* estimators we use a bivariate Gaussian kernel and a plug-in bandwidth selector. The density of the original data S_0 ($r = 0$ in Table 1.3) is estimated by using function *kde* (kernel density estimation) with a bivariate Gaussian kernel and a plug-in bandwidth selector. The density estimates of the original data are treated as a *benchmark* because S_0 is not affected by rounding error. The simulation steps (generation of a dataset, rounding of the coordinates and the density estimation) are independently repeated 500 times for each scenario.

In Table 1.3 we compare the performance of the *Naive*, the *GRSST* and the Delaigle density estimators in the three scenarios. The first column of Table 1.3 shows the means and the standard deviations of the RMISE over 500 Monte-Carlo replications of the *benchmark* case i.e. in the absence of rounding error ($r = 0$). Note that in the definition of the RMISE in (1.5) $f(\boldsymbol{x})$ denotes now the underlying true density, $f_A(\boldsymbol{x}), f_B(\boldsymbol{x})$ or $f_C(\boldsymbol{x})$ respectively.

For the scenarios with small rounding errors ($r = 0.75$) we observe that the *Naive* and the *GRSST* density estimators perform similarly and both methods have RMISE which is comparable to the RMISE under the *benchmark* scenario. The Delaigle estimator reveals a higher RMISE compared to the two other approaches. Data providers may be keen, however, to introduce more severe measurement error to the data for ensuring

confidentiality. For such scenarios ($r = 1.5$ and $r = 2.25$) the *GRSST* density estimator clearly outperforms the *Naive* estimator. The Delaigle estimator performs better than the *Naive* estimator but worse compared to the *GRSST* estimator. It is notable that the *Naive* estimator performs very poorly especially for $r = 1.5$ and $r = 2.25$ in the case of a bivariate standard normal distribution (Scenario A). Presumably this is due to the small variance of the underlying density we are trying to estimate in Scenario A such that discretizing for given rounding values has a much more pronounced effect. For this reason we also tested a bivariate normal distribution with a larger variance. The results for the *Naive* method become more stable but the *GRSST* estimator still performs better. Figure 1.8 shows contour plots of a particular simulation run under Scenario B for the *Naive* and *GRSST* estimators. It appears that, unlike the *Naive*, the *GRSST* density estimator is able the maintain the underlying structure of the density for different rounding levels. Contour plots under Scenarios A and C (provided as part of the supporting information) confirm this finding. The anisotropic pattern for the *Naive* estimator ($r = 1.5$ and $r = 2.25$) is caused by a larger bandwidth in x-direction than in y-direction. This bandwidth is chosen by the plug-in bandwidth matrix selector of (Wand and Jones (1994).

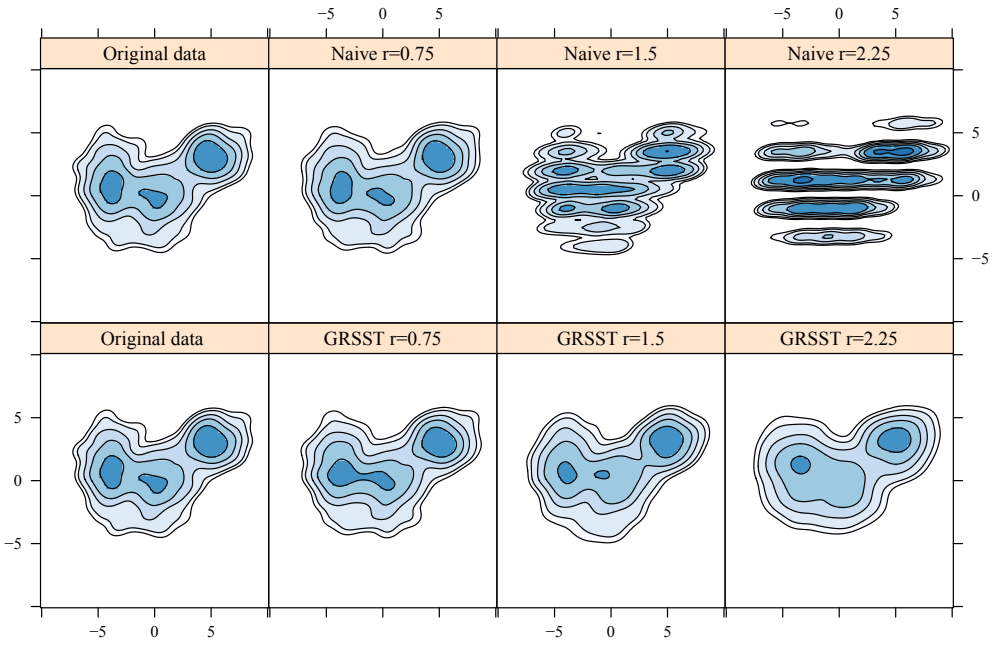

Figure 1.8: Scenario B: Contour plots of *Naive* estimator (upper panel) and *GRSST* estimator (lower panel), for grid size $r = 0.75, 1.5, 2.25$ (left to right). The original data scenario ($r = 0$) is used as the benchmark.

1.6 Discussion

Precise geo-coded data is rarely available due to confidentiality constraints. The work proposes methodology for deriving density estimates of populations of interest in the presence of rounding in the geographical coordinates used for disclosure control. The proposed methodology is motivated by reversing the measurement error process by combining a measurement error model with kernel density estimation. The method is straightforward to implement and works for different dimensions, symmetric as well as asymmetric kernel types and bandwidth selection methods. The use of the proposed methodology is facilitated by the availability of function *dbivr* in the R package **Kernelheaping** available on CRAN (Groß, 2016a). As we demonstrated with the analysis of the Berlin register data the proposed method can offer considerably deeper insights, compared to a *Naive* estimator that disregards the measurement error process, to data analysts about the density of target populations within an area of interest. The structure preserving property of the proposed method is particularly attractive when working with data that has been subjected to disclosure control via the introduction of measurement error. In addition, the work provides some first indications on how to set the grid-lengths for geo-coding in the Berlin register of residents such that a data analyst is able to derive precise density estimates. At the same time working with the data host for deciding the grid-lengths is crucial for ensuring confidentiality.

Further work could extend the proposed approach to different geographical masking or anonymisation methods including for example the use of Gaussian errors added to the original geographic coordinates. With minor adaptions to the algorithm direct use of arbitrary demarcation shapes like the LORs instead of the grid-structure induced by rounding is possible for obtaining smooth density estimates as well. The proposed method can be further generalized for application to data with varying degree of rounding (*heaping*) occurring, for example, in self-reported survey data (Pudney, 2008). Finally, one idea for further work is to explore the application of the proposed methodology for generating synthetic geo-coded data based on anonymised data sets with rounding errors.

Acknowledgments

The authors gratefully acknowledge the data access and support of the Research Data Centre of the Statistical Office for Berlin-Brandenburg and are indebted to the Editor, Associate Editor and two referees for comments that significantly improved the paper.

1.7 Supplementary Material: Additional Empirical Evaluations

As part of the supporting information we evaluate the sensitivity of the proposed method in relation to the size of the data (n), the burn-in size (B) and the sample

Table 1.4: Scenario B: Mean RMISE for different sizes of datasets. Corresponding standard errors of the RMISE in parentheses.

	$r = 0$	$r = 0.75$			$r = 1.5$			$r = 2.25$		
	Original	*Naive*	*GRSST*	Delaigle	*Naive*	*GRSST*	Delaigle	*Naive*	*GRSST*	Delaigle
$n = 100$	0.273	0.280	0.278	0.682	0.304	0.294	0.400	0.434	0.332	0.382
	(0.028)	(0.028)	(0.028)	(0.055)	(0.028)	(0.029)	(0.033)	(0.039)	(0.027)	(0.029)
$n = 500$	0.162	0.172	0.170	0.328	0.380	0.183	0.272	0.679	0.256	0.390
	(0.016)	(0.017)	(0.016)	(0.019)	(0.033)	(0.018)	(0.013)	(0.043)	(0.016)	(0.014)
$n = 1000$	0.128	0.139	0.141	0.247	0.550	0.148	0.255	0.872	0.235	0.332
	(0.013)	(0.013)	(0.013)	(0.013)	(0.036)	(0.014)	(0.011)	(0.053)	(0.016)	(0.013)
$n = 2000$	0.100	0.112	0.105	0.189	0.727	0.123	0.241	1.187	0.221	0.331
	(0.010)	(0.011)	(0.011)	(0.012)	(0.037)	(0.011)	(0.010)	(0.064)	(0.015)	(0.014)
$n = 5000$	0.072	0.165	0.102	0.148	1.033	0.107	0.228	1.803	0.210	0.304
	(0.012)	(0.022)	(0.016)	(0.012)	(0.081)	(0.016)	(0.015)	(0.137)	(0.024)	(0.020)

steps (N) used in the algorithm for implementing $GRSST$ estimator. This set of simulation results complements the results that we have included in the work. In addition we present contour plots under Scenario A and C in Figure 1.9 and 1.10 respectively. The remaining features of the simulation set up remain the same as in Section 1.5.

For evaluating the impact of the size of the dataset on the estimators, in Table 1.4 we report the means and the standard deviations of the RMISE under scenario B for $n = 100, 500, 1000, 2000$ and 5000. First, we observe that the results of the *benchmark* estimator ($r = 0$) improve as the size increases. This is expected because there is no rounding error in the data and hence the larger the size of the data, the more precise the estimates of the underlying density are. The advantage of using the proposed $GRSST$ estimator increases with the size of the dataset. For $n = 100$, the benefit from using the $GRSST$ estimator is relatively low. The small data size means that the chosen bandwidth is large. However, for larger datasets the bandwidth determinant $|H|$ gets smaller. In this case the spikes of the density estimates obtained by using the $Naive$ estimator get more pronounced, which leads to an increasing RMISE for the $Naive$ method. In contrast, the $GRSST$ and the Delaigle estimators benefit from an increasing data size. In the presence of rounding information is irreversibly lost leading to an increased RMISE for all estimators that utilise the rounded data. However, the original data is not available. The proposed $GRSST$ estimator, that accounts for the rounding process, is able to provide a density that is close to the density of the original data and appears to be more efficient than competitor estimators.

For assessing the effect of the burn-in size (B) and the sample steps (N) on the proposed method we implement the $GRSST$ algorithm for scenario B by using different combinations of burn-in sizes ($B = 0, 1, 5, 10, 20$) and sample steps ($N = 1, 2, 20, 50, 100$). Table 1.5 shows the means and standard deviations of the RMISE over 500 Monte-Carlo replications. We observe that larger B and N values improve the results in particular as the rounding error increases. Thus, the approach by Wang and Wertelecki (2013) which is identical to the proposed method with burn-in iterations $B = 0$ and $N = 1$ or more sampling steps is less efficient than the proposed $GRSST$ estimator. The improvement is only marginal, however, for B and N larger than 5 and 20 respectively.

For investigating the computing time needed for implementing the proposed esti-

Table 1.5: Scenario B: Mean RMISE for different burn-in (B), sample steps (N). Corresponding standard errors of the RMISE in parentheses.

Estimators	$r = 0.75$	$r = 1.5$	$r = 2.25$
Naive	0.172 (0.017)	0.380 (0.033)	0.679 (0.043)
GRSST (B=0, N=1)	0.176 (0.017)	0.216 (0.019)	0.300 (0.020)
GRSST (B=1, N=2)	0.172 (0.017)	0.193 (0.019)	0.274 (0.020)
GRSST (B=5, N=20)	0.170 (0.016)	0.183 (0.018)	0.256 (0.016)
GRSST (B=10, N=50)	0.170 (0.017)	0.181 (0.019)	0.254 (0.017)
GRSST (B=20, N=100)	0.170 (0.017)	0.181 (0.018)	0.254 (0.017)

Table 1.6: Scenario B: Average computing times (in seconds) for different burn-in (B), sample steps (N) and sizes of the data (n).

Estimators	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
Naive	0.741	1.447	2.817	6.957
GRSST ($B = 0$, $N = 1$)	0.856	1.576	2.973	7.196
GRSST ($B = 1$, $N = 2$)	3.955	4.877	7.026	13.391
GRSST ($B = 5$, $N = 20$)	26.608	30.262	36.626	61.687
GRSST ($B = 10$, $N = 50$)	62.143	70.392	86.821	143.661
GRSST ($B = 20$, $N = 100$)	124.862	136.824	166.030	275.633

mator, in Table 1.6 we present average computing times for different burn-in iterations (B), sample steps (N) and data size (n) under Scenario B. (Operating machine: 64-bit windows system with an Intel Core i7-2600 CPU 3.40 GHz with 16 GB RAM, parallel computing was not used). The computation time mainly depends on the number of iterations $B + N$. For instance, the computing time for $B + N = 25$ iterations is around 30 seconds for $n = 1000$ and around 60 seconds for $n = 5000$. In contrast, the computing time increases to 136 seconds for $n = 1000$ and to 275 seconds for $n = 5000$ when $B + N = 120$ iterations. The setting $B = 5$ and $N = 20$ appears to be offering a good compromise between RMISE and computing time.

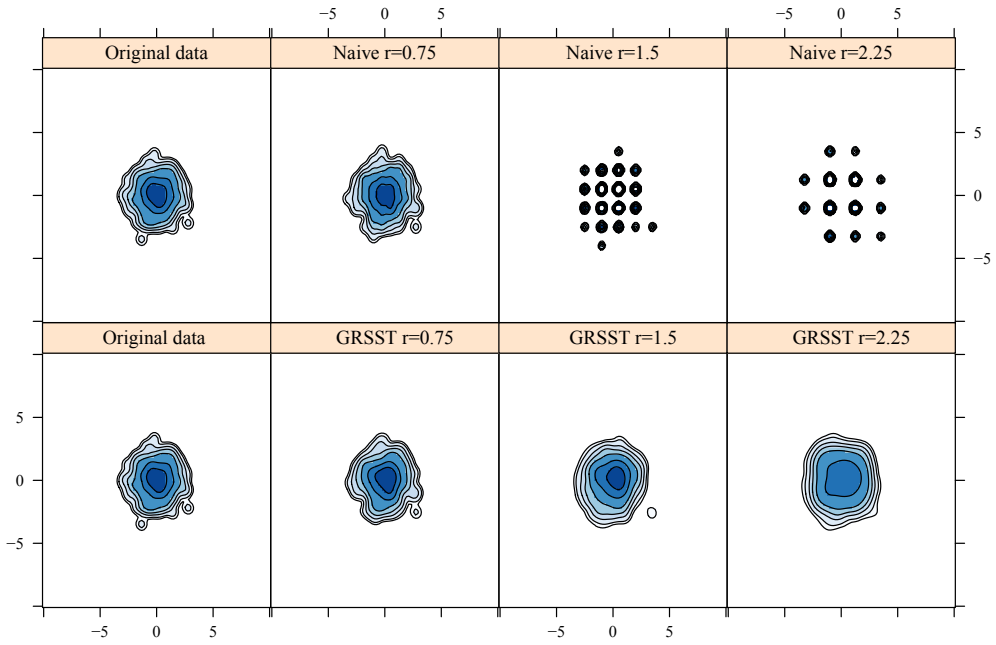

Figure 1.9: Scenario A: Contour plots of *Naive* estimator (upper panel) and *GRSST* estimator (lower panel), for grid size $r = 0.75, 1.5, 2.25$ (left to right). The original data scenario ($r = 0$) is used as the benchmark.

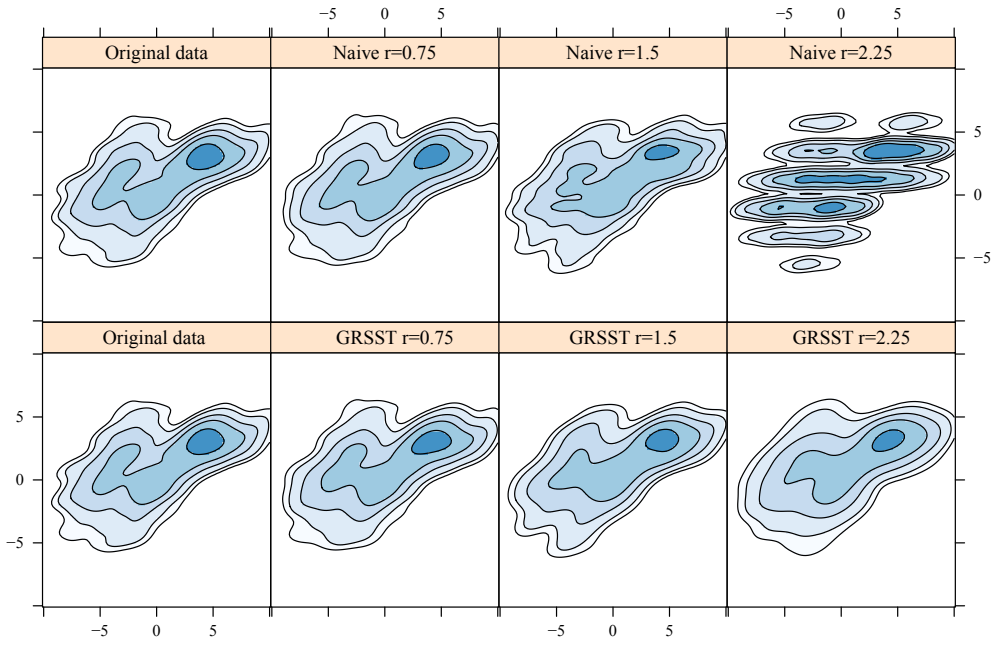

Figure 1.10: Scenario C: Contour plots of *Naive* estimator (upper panel) and *GRSST* estimator (lower panel), for grid size $r = 0.75, 1.5, 2.25$ (left to right). The original data scenario ($r = 0$) is used as the benchmark.

Kapitel 2

Kernel Density Estimation for Heaped Data

2.1 Introduction

In survey data the researcher often encounters rounded values when the participants are asked to state quantitative variables such as income (Hanisch 2007; Czajka and Denmead 2008), household expenditures (Pudney 2008), body weight and height (Taylor et al. 2006), blood pressure (De Lusignan et al. 2004) or working hours (Otterbach and Sousa-Poza 2010). The rounding behaviour of self-reported data is usually mixed, i.e. participants may round to multiples of 1, 2, 5, 10, 20, 50, 100..., or may report only two leading digits (Hanisch 2007). This type of measurement error – when data are collected with various degrees of coarseness – is called heaping. Heaping cannot be ignored because it is a well known fact (Heitjan and Rubin 1991; Schneeweiß and Komlos 2009), that if we naively use the self-reported values in the estimation of a distribution, the estimates are biased. This is especially the case in (non-parametric) kernel density estimation where we observe bumps und spikes at the multiples of the rounding values. The standard methods of choosing the bandwidth are also not very useful in this setting. The Sheather-Jones estimate (Sheather and Jones 1991), which is mostly recommended in literature, often produces completely useless density estimates from self-reported data. This is because a pilot estimate of the integral of the second derivative is employed to estimate the bandwidth. Due to the extremely multimodal nature of the heaped data, this plug-in estimate of the integrated second derivative is very large, leading to very small bandwidths. Silverman's rule of thumb behaves better because it implicitly assumes a normal distribution for bandwidth selection, but still gives not very satisfying results. Figure 2.1 shows two examples from a household survey, the German Socio-Economic Panel –'SOEP'– (Wagner et al. 2007) wave BA (2010): body weight of the female participants and monthly food and drink expenditures outside home.

Increasing the bandwidth until the density estimate is sufficiently smooth leads to oversmoothing: the tails of the distribution get too heavy and important features of the distribution may be lost. Additionally, participants may be more prone to round up or

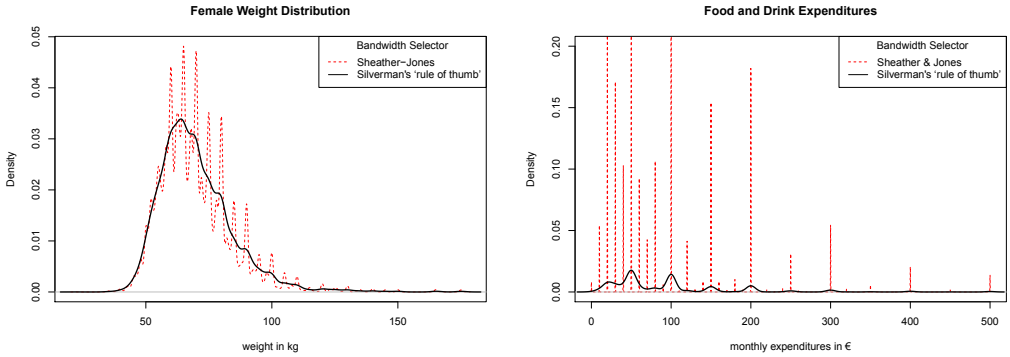

Figure 2.1: Kernel density estimator applied to self-reported female weight (left) and food and drink expenditures (right) taken from the Socio-Economic Panel 2010. The two popular bandwidth selectors ('Sheather-Jones' and Silverman's 'rule of thumb') show more or less severe spikes at the multiples of the rounding values.

down due to social desirability. For self-reported weight measurements with validation data, for example, respondents typically underreport their weight, a finding that can be partially explained by their tendency to round off (Rowland 1990; Shields et al. 2008; Merrill and Richardson 2009). This work proposes a method of non-parametric density estimation for self-reported measures in the presence of heaping. The primary goal is to provide a method that reduces the bias in kernel density estimation and estimates the parameters of the heaping process as well. To the authors best knowledge, this is the first general attempt to use a measurement error model to solve this type of problem. In particular, a latent variable model is estimated by a Stochastic Expectation-Maximization (SEM, Celeux et al. 1996) algorithm. We show that under certain assumptions it is possible to identify and estimate a rounding direction bias (unequal probability of rounding up and down). The work is organized as follows: Section 2.2.1 provides a literature overview of existing modeling approaches for heaped data. In Section 2.2.2 introduces a model for the heaping behaviour of the respondents. After a short introduction to kernel density estimation the measurement error model and its implementation are presented in Section 2.3. Section 2.4 provides a simulation study and Section 2.5 demonstrates an application to self-reported data from the SOEP. Further model extensions to overcome shortcomings of the proposed methods are discussed in Section 2.6. A short summary concludes the article.

2.2 Modeling Heaping in Self-Reported Data

2.2.1 Heaping models in applications

Heaping occurs frequently in a variety of applications in quite different fields. Almost all analytical approaches a parametric probability model for the variable subject to heaping. Heitjan and Rubin (1990) modeled the heaping process as rounding with different interval length and used a complex imputation model to estimate the age distribution of Tanzanian children. A similar approach was followed by Battistin et al.

(2003) in household food expenditures. Wang and Heitjan (2008) proposed a model for heaped cigarette counts which was subsequently extended in Wang et al. (2012) to account for 'recall' error. Crawford et al. (2015) formulated a general model for count data involving birth-death processes and applied this to the self-reported counts of the number of sex-partners. Bar and Lillard (2012) developed an approach for event time data by modeling the density by a mixture of two parametric distributions. In a very recent publication of self-reported data from the SOEP (Marcus et al. 2013), a model of the heaping process was discarded on the grounds that it was based on arbitrary assumptions and a parametric density was fitted without any correction for the heaping process. We disagree with this assessment, because although we might not be able to reproduce the heaping pattern perfectly by the heaping model assumptions, the bias in the parameter estimates may be greatly reduced.

The existing approaches to heaping correction rely primarily on smoothing techniques, but do not impose a model for the heaping process and thus lack interpretability. Camarda et al. (2008) deals with estimating age-at-death as well as body weight by assuming a smooth underlying density function modeled by B-splines. However, it was assumed that the true unobserved value was the reported value itself or one of the two immediate neighbouring integers, which is clearly not suitable for other data such as monthly income. Golyandina et al. (2012) smooths the nonparametric density so that it has less than a predefined number of modes. In contrast, this article aims to develop a model for the heaping behaviour and pursues a general strategy applicable to a wide variety of data.

2.2.2 A model for heaping

In this article the heaping process is modeled as follows: Let W_i $(i = 1, .., n)$ be the reported values which are rounded by a value $R_i \in \{r_1, r_2, .., r_m\}$, where $\boldsymbol{r} = (r_1, r_2, .., r_m)$ denotes a vector of rounding parameters. We assume that the rounding is done correctly such that the true value X_i lies within the interval $(W_i - 1/2R_i, W_i + 1/2R_i)$. As R_i is not uniform over the individuals, we have a heteroscedastic measurement error here. The choice of rounding parameters $\boldsymbol{r}$ is crucial and has to be made beforehand by taking descriptive statistics or theoretical considerations into account. When looking at the SOEP female body weight example, for instance, we observe that the most frequent end digit was **0** with 22.1% of the reported cases followed by **5** with 16.5%. Moreover, the respondents seem to prefer even over odd numbers. The end digits **2,4,6,8** are reported in 35.5% of the total cases while the end digits **1,3,7,9** only sum up to 26.0% (see Table 2.1 for details).

Table 2.1: End digits of SOEP female body weight in kg.

end digit	0	1	2	3	4	5	6	7	8	9
count	1929	399	912	712	664	1440	576	577	953	574
%	22.1	4.6	10.5	8.2	7.6	16.5	6.5	6.6	10.9	6.6

Therefore, we may choose the rounding values $\boldsymbol{r} = (1, 2, 5, 10)$. In general, suitable potential rounding parameters are $\boldsymbol{r} = (.., 0.5, 1, 2, 5, 10, 20, ..)$ for variables with decimal numeral system (e.g. blood pressure, body weight,..), $\boldsymbol{r} = (1, 2, 3, 6, 12)$ for variables with duodecimal system (e.g. time in months, length in inches,..) and $\boldsymbol{r} = (1, 5, 10, 15, 30, 60)$ for the sexagesimal system (e.g. time in minutes). A probability vector $\boldsymbol{p} = (p_1, .., p_m)$ is assigned to the rounding values $\boldsymbol{r}$ denoting the probability of the respondent to round by a certain value $R_i \in \{r_1, r_2, .., r_m\}$. In some applications one may observe unrounded values or values that are given with a very high precision. Therefore, we allow for $R_i = 0$ with rounding value $r_1 = 0$ as well [1]. For the moment $\boldsymbol{p}$ is assumed equal for all respondents and independent from the true, unobserved value X_i. This is a key assumption which is not always met and will be relaxed later on. The model implies that each combination of X_i and R_i is feasible while this is not true for the all the combinations of W_i and R_i. In the upper example a respondent with end digit **0** may have rounded by the values $R_i = 1, 2, 5$ or 10, while for end digit **3** only $R_i = 1$ is feasible. However, this makes it possible to identify the parameter vector $\boldsymbol{p}$ in the first place.

The model for the heaping process described above may not fit very well to all kinds of data. Thus, we consider two extensions. As already mentioned the respondents may more likely round down than round up or vice versa. A first suggestion is to define a parameter $a \in (0, 1)$ allocating the probability of rounding down. However, when imposing the restriction $X_i \in (W_i - 1/2R_i, W_i + 1/2R_i)$ (rounding mathematically correct) it is not possible to choose the rounding direction independently from R_i and X_i. Consider the true value $X_i = 77.8$, rounding values $\boldsymbol{r} = (1, 10)$ and assume mathematically correct rounding behaviour the respondent has to round up in any case regardless of his chosen rounding value R_i. We therefore introduce an alternative concept. We extend R_i such that it includes the rounding direction: $R_i \in \{-r_1, .., -r_m, +r_1, .., +r_m\}$ whereby negative values indicate a rounding up and positive values a rounding down. The rounding probabilities $\boldsymbol{p}$ are multiplied by a when rounding down ($R_i > 0$) and by $(1 - a)$ when rounding up ($R_i < 0$) if the combination of R_i and X_i is compliant with the assumption of correct rounding (i.e. $X_i \in (W_i, W_i + 1/2R_i)$ for $R_i > 0$ and $X_i \in (W_i + 1/2R_i, W_i)$ for $R_i < 0$) and are set to 0 else. They are scaled afterwards such that the probabilities for all R_i sum up to 1. We give a numerical example how the conditional probability distribution $\pi(R_i|X_i, \boldsymbol{p}, a)$ denoted as $\pi(R_i|\cdot)$ is modeled:

- Let $\boldsymbol{r} = (1, 2, 5, 10)$, $\boldsymbol{p} = (0.4, 0.3, 0.2, 0.1)$ and $a = 0.15$. For $X_i = 23.4$, possible reported values are $W_i = 23$ (rounding down by $R_i = 1$), $W_i = 24$ (rounding up by $R_i = 2$), $W_i = 25$ (rounding up by $R_i = 5$) and $W_i = 20$ (rounding down by $R_i = 10$). The conditional probabilities $(\pi(R_i = -1|\cdot), \pi(R_i = -2|\cdot), \pi(R_i = -5|\cdot), \pi(R_i = -10|\cdot), \pi(R_i = 1|\cdot), \pi(R_i = 2|\cdot), \pi(R_i = 5|\cdot), \pi(R_i = 10|\cdot))$ are proportional to $(0, 0.3 \cdot (1 - 0.15), 0.2 \cdot (1 - 0.15), 0, 0.4 \cdot 0.15, 0, 0, 0.1 \cdot 0.15) =$

[1] Truly continuous variables are always rounded to some degree, at least to machine precision. However, there are many examples for quasi-continuous variables such as monthly income and therefore we define an unrounded or exact value as a value which is not dividable by the smallest rounding value larger than 0.

$(0, 0.255, 0.17, 0, 0.06, 0, 0, 0.015)$. Thus, $P(W_i = 23|\cdot) = 0.12$, $P(W_i = 24|\cdot) = 0.51$, $P(W_i = 25|\cdot) = 0.34$ and $P(W_i = 20|\cdot) = 0.03$.

In general, with direction parameter $a \in (0, 1)$ the conditional probability distribution of R_i given X_i, $\boldsymbol{p}$ and a is proportional to the following expression:

$$\pi(R_i = \pm r_j | X_i, \boldsymbol{p}, a) \propto a^{I(R_i>0)} \times (1-a)^{I(R_i<0)} \times p_1^{I(R_i=-r_1)} \times .. \times p_m^{I(R_i=+r_m)} \\ \times I[sgn\{X_i \text{ mod } (|R_i|) - \frac{1}{2}|R_i|\} = -sgn(R_i)]$$

The second line serves as a check whether the combination of X_i and R_i is compatible with the restriction of mathematically correct rounding.

The value a can be interpreted as the tendency to round down ($a > 0.5$) or to round up ($a < 0.5$). The reason to restrict to mathematically correct rounding is that it allows us identify the rounding direction parameter a solely by the end digit pattern. In the simple example of a flat density, $a > 0.5$ and rounding values $\boldsymbol{r} = (1, 10)$ one would observe the end digits **1** to **4** less often than **6** to **9** (or the other way around for $a < 0.5$). This is because the respondent is only able to round down by $R_i = 10$ if X_i mod $10 \in (0, 5)$ and round up by $r = 10$ if X_i mod $10 \in (5, 10)$ with the result that for $a > 0.5$ most reported values W_i with end digit **0** correspond to a true value X_i with X_i mod $10 \in (0, 5)$. In the SOEP female body weight example, the left neighbours (**9**, **4**) of end digits **0** and **5**; show significantly higher counts (574 to 399 and 664 to 576) than their right counterparts (**1**, **6**), indicating a tendency to round off.

A second extension allows for non-constant rounding probabilities. For example, the probability of a respondent with a true income of $X_i = 1600$ to choose $R_i = 1000$ (and round up to $W_i = 2000$) might be much lower than for someone earning 8600 (and report 9000). A natural choice would be to implement an ordered probit (or logit) model for the rounding probabilities $\boldsymbol{p}$ (as already done in Heitjan and Rubin 1990) with the logarithm of the true value as independent variable:

$$g_i = log(X_i)\beta + \epsilon_i \ , \ \epsilon_i \sim N(0, 1)$$

$\mathbf{g}$ denotes the latent continuous variable and we define $\boldsymbol{\tau} = (\tau_0, \tau_1, .., \tau_m)$ as threshold parameters with $\tau_0 = -\infty$ and $\tau_m = +\infty$. The rounding probability for rounding value r_j $(j = 1, .., m)$ for respondent i is then defined as:

$$\begin{aligned} p_{ij} &= P(\tau_{j-1} < g_i \leq \tau_j) \\ &= \Phi(\tau_j - log(X_i)\beta) - \Phi(\tau_{j-1} - log(X_i)\beta) \end{aligned}$$

The rounding probabilities $\boldsymbol{p}$ may also depend on other characteristics of the respondents which can be integrated into the ordered probit regression formula as well. For $a = 0.5$ and $\beta = 0$ the extended model reduces to the standard rounding model.

The presented model considers heaping as rounding to different degrees of coarseness. However, when evaluation data is available the discrepancies between true and reported values are typically not exclusively explainable by rounding. Wang et al.

(2012) and Crawford et al. (2015) argue that a so-called 'recall' error is involved in the heaping process as well. In Section 2.6 we discuss how the model can be further extended to incorporate this and other extensions.

2.3 Methods

2.3.1 Kernel density estimation

Kernel density estimation as a non-parametric approach for density estimation is an important tool in exploratory data analysis. Let $\boldsymbol{X} = (X_1, X_2, .., X_n)$ denote a sample of size n from a random variable with density f. The univariate kernel density estimate at point x is given by:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{2.1}$$

where $K(\cdot)$ is kernel function and h denotes a bandwidth, which governs the smoothness of the density estimate. The kernel $K(\cdot)$ satisfies regularity conditions such as (a) $\int K(x)dx = 1$, (b) $\int xK(x)dx = 0$ and (c) $\int x^2K(x)dx < \infty$ (Scott 2009). The performance of a kernel density estimator is mainly affected by the choice of h (cf. Izenman 1991). Popular strategies to choose h are by minimizing the AMISE (Asymptotic Mean Integrated Squared Error) through plug-in or cross-validation methods (cf. Izenman 1991 or Silverman 1986). Sheather (2004) gives a short overview in kernel density estimation, kernels and bandwidth choice methods. For self-reported data from surveys one might consider two modifications of the standard kernel density estimator. First, a sampling weight adjustment should be carried out. Second, most self-reported data are non-negative and therefore a boundary correction is favourable. Both modifications are shortly described in Jann (2007). Unfortunately, the utilization of kernel density estimation methods with heaped data leads to severely biased estimates, as already demonstrated in the introduction.

2.3.2 Model

As discussed in Section 2.2.2, the true values $\boldsymbol{X} = (X_1, X_2, .., X_n)$ are assumed unobservable and only the reported heaped values $\boldsymbol{W} = (W_1, W_2, .., W_n)$ are available. However, we still aim to estimate the density f from which our sample $\boldsymbol{X}$ is drawn, by using the heaped values W_i. One approach to measurement error problems is to treat the unknown true values X_i as latent variables (Carroll et al. 2006). Then the likelihood can be split into two parts. We specify the following models: First, a measurement error model and second a model for the latent variables. The distribution of $\boldsymbol{X}$ can be modeled by a parametric distribution (e.g. by a Gaussian with parameter vector $\boldsymbol{\theta} = (\mu, \sigma)$) or non-parametrically either by a mixture of parametric distributions (Escobar and West 1995) or by using kernel density estimation with $\theta = h$ as introduced in the previous subsection. Applying the heaping process of Section 2.2.2 with rounding values $\boldsymbol{R} = (R_1, R_2, .., R_n)$, we formulate a measurement error model for $\boldsymbol{W}$ as well. We start with the heaping model without extensions. As W_i does only

depend on X_i and R_i, the distribution $\pi(\boldsymbol{W}|\boldsymbol{X},\boldsymbol{R})$ is defined by a product of Dirac distributions, $\pi(\boldsymbol{W}|\boldsymbol{X},\boldsymbol{R}) = \prod_{i=1}^{n} \pi(W_i|X_i,R_i)$, with

$$\pi(W_i|X_i,R_i) = \begin{cases} 1 & \text{for } X_i \in (W_i - \frac{1}{2}R_i, W_i + \frac{1}{2}R_i) \\ 0 & \text{else.} \end{cases},$$

By the definition of our heaping model without extensions in Section 2.2.2, the R_i do only depend on $\boldsymbol{p}$, and $\pi(\boldsymbol{R}|\boldsymbol{p})$ follows a multinomial distribution. Now a likelihood for $\boldsymbol{W}$ can be formulated:

$$L(\boldsymbol{W}|\boldsymbol{X},\boldsymbol{R},\boldsymbol{p},\boldsymbol{\theta}) = \underbrace{\pi(\boldsymbol{W}|\boldsymbol{X},\boldsymbol{R}) \times \pi(\boldsymbol{R},\boldsymbol{p})}_{\text{Measurement error model}} \times \underbrace{\pi(\boldsymbol{X}|\boldsymbol{\theta})}_{\text{Observation model}} \tag{2.2}$$

$$= \prod_{i=1}^{n} \pi(W_i|X_i,R_i) \times \pi(R_i|\boldsymbol{p}) \times f(X_i|\boldsymbol{\theta}) \tag{2.3}$$

In order to implement the two extensions proposed in Section 2.2.2 we have to introduce the parameters a, $\boldsymbol{\tau}$ (as threshold values for $\boldsymbol{p}$) as well as β into our likelihood. Again, the measurement error model consists of two parts, with

$$\pi(W_i|X_i,R_i) = \begin{cases} 1 & \text{for } R_i > 0 \text{ and } X_i \in [W_i, W_i + \frac{1}{2}R_i) \\ 1 & \text{for } R_i < 0 \text{ and } X_i \in (W_i + \frac{1}{2}R_i, W_i) \\ 0 & \text{else} \end{cases},$$

and (cf. Section 2.2.2):

$$\begin{aligned} \pi(R_i = \pm r_j|X_i,\boldsymbol{\tau},a,\beta) \propto\ & a^{I(R_i<0)} \times (1-a)^{I(R_i>0)} \\ & \times [\Phi\{\tau_1 - log(X_i)\beta\} - \Phi\{\tau_0 - log(X_i)\beta\}]^{I(R_i=-r_1)} \\ & \times .. \\ & \times [\Phi\{\tau_m - log(X_i)\beta\} - \Phi\{\tau_{m-1} - log(X_i)\beta\}]^{I(R_i=+r_m)} \\ & \times I[sgn\{X_i \bmod |R_i| - \frac{1}{2}|R_i|\} = -sgn(R_i)] \end{aligned}$$

The likelihood for the extended model is therefore:

$$L(\boldsymbol{W}|\boldsymbol{X},\boldsymbol{R},\boldsymbol{\tau},a,\beta,\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(W_i|X_i,R_i) \times \pi(R_i|X_i,\boldsymbol{\tau},a,\beta) \times f(X_i|\boldsymbol{\theta}) \tag{2.4}$$

The Expectation-Maximization (EM) algorithm (Dempster et al. 1977) is a natural algorithm for maximum-likelihood estimation for models with latent data. The E step of this iterative algorithm, the latent values – the X_i – and potentially other parameters not exclusively associated with the latent values such as $\boldsymbol{R}$ and $\boldsymbol{p}$), are replaced by their conditional expectations given the observed data and the current estimate of $\boldsymbol{\theta}$. The M step re-estimates $\boldsymbol{\theta}$ through maximization of the observed likelihood based on the imputed values of the latent values. However, this approach would clearly not

work for kernel density estimation with heaped data observed, because all the reported observations on a heaping value would still be concentrated at a single point (namely the expectation of the conditional distribution of X_i computed in the E-Step) in each iteration and thus not preventing spiky estimates of the density. A way out of this dilemma is to use the so-called Stochastic Expectation-Maximization (SEM) algorithm (Celeux et al. 1996), which draws samples of the conditional distribution of $\boldsymbol{X}$ – called S-step – in replacement of the E-Step to create a pseudo sample of $\boldsymbol{X}$ in each iteration. This algorithm is strongly related to the Gibbs-sampler but usually converges much faster (Diebolt and Ip 1996). As a consequence, the proposed estimator is a partly Bayesian method in the sense that the X_i as well as $\boldsymbol{p}$, a, β and $\boldsymbol{\tau}$ are treated as random variables but not $\boldsymbol{\theta}$. As this work deals with kernel density estimation, f is specified by $\hat{f}_h$, the kernel density 'estimator' presented in (2.1) [2], with $\theta = h$. Unfortunately, with direct maximization of $\pi(\boldsymbol{X}|h) = \prod_{i=1}^{n} \hat{f}_h(x)$ with respect to h, the bandwidth h would yield a degenerate solution. We therefore utilize a generalization of the (S)EM algorithm for the use of surrogates of the likelihood in the M-step (McLachlan and Krishnan 2007) such that the objective of maximization is replaced by e.g. the minimization of the asymptotic mean integrated square error (AMISE) of the kernel density estimate. In the context of non-parametric kernel density estimation, this approach enables us to use any bandwidth selection method from the rich variety available in literature, e.g. Silverman's rule of thumb or the Sheather-Jones selector. As discussed in the next section, Gibbs-sampler and Metropolis-Hastings steps are introduced into the S-step of the algorithm (cf. Diebolt and Ip 1996). The proposed algorithm can be seen as an extension to the method introduced in Groß et al. (2016) dealing with rounded data in context of bivariate kernel density estimation. Further details about the exact implementation of the algorithm are given in the next subsection.

2.3.3 Computational details

We first consider the case without extensions for the distribution of X_i and R_i given the rounded values W_i, the rounding parameters $\boldsymbol{p}$, and bandwidth h:

$$\pi\left(X_i, R_i | W_i, h, \boldsymbol{p}\right) \propto I(W_i - \frac{1}{2}R_i \leq X_i \leq W_i + \frac{1}{2}R_i) \times p_j \times \hat{f}_h(X_i),$$

The full conditional distribution of (X_i, R_i) is the product of a uniform distribution on the interval with length R_i around W_i, the probability p_j of rounding to a certain degree of coarseness r_j and the kernel density 'estimator' $\hat{f}_h(X_i)$ (2.1). The conditional distribution of $\boldsymbol{p}$ given R is the Dirichet distribution $Dir(\alpha)$:

$$\pi(\boldsymbol{p}|\boldsymbol{R}) \sim Dir(\#(\boldsymbol{R} = r_1), .., \#(\boldsymbol{R} = r_m))$$

Next we consider the case of the two extensions of the heaping model. We were able

[2] Note that the expression 'kernel density estimator' is ambiguous here as in this context it should be merely called 'kernel density'. However, as we think that a second definition of a kernel density f_h which would be equal to $\hat{f}_h$ could be even more confusing we quote the word 'estimator' when actually referring to a 'kernel density'.

to find a modified expression for the conditional distribution of (X_i,R_i) given W_i but no established distribution was found for the conditional distribution of $(\boldsymbol{\tau},a,\beta)$ given $\boldsymbol{X}$ and $\boldsymbol{R}$. A Metropolis-Hastings step turned out to be computational cumbersome because of very slow convergence with the result that a Laplace normal approximation of the joint full conditional distribution $\pi(\boldsymbol{\tau}, a^*, \beta|\cdot)$ was utilized instead, where the parametrization $a^* = \Phi^{-1}(a)$ was used for the reason of computational convenience.

As a consequence a generalized SEM algorithm is proposed, sampling iteratively from the full conditional distributions of (X_i, R_i) as well as from (an approximation of) the full conditional distributions of $(\boldsymbol{\tau}, a, \beta)$ in the S-step (which replaces the E-step in the EM-algorithm) and computing $\hat{f}_h$ combined with re-estimation of h in the M-step. Our simulations show that the proposed algorithm works very well in terms of RMSE (Root Mean Square Error) and coverage intervals. The steps of the algorithm are described below:

1. Get an initial estimate $\hat{f}_h$ of f using the heaped data $\boldsymbol{W}$ and setting h to a sufficiently *large* value such that no rounding spikes occur (e.g. $h = 2 \cdot max(\boldsymbol{r})$). Set starting values for $\boldsymbol{\tau}$ to $\Phi^{-1}(0, 1/m, 2/m, .., (m-1)/m, 1)$ and for a^*, β to 0.

2. Evaluate and save density estimate $\hat{f}_h$ on an equally-spaced fine grid G with grid width $\delta_G = \frac{min(\boldsymbol{r})}{k}$, whereby $1 < k \in \mathbb{N}$. In particular,
$G = \left\{\min(W_i) - \frac{1}{2}r_m, \min(W_i) - \frac{1}{2}r_m + \delta_G, .., \max(W_i) + \frac{1}{2}r_m\right\};\ i = 1, .., n.$

3. Sample from $\pi(X_i, R_i|\cdot)$ by computing it for every combination of R_i and values $X_i \in G;\ i = 1, .., n$.

4. Sample from $\pi(\boldsymbol{p}|\boldsymbol{R})$ in case of the model without extensions or the joint full conditional $\pi(\boldsymbol{\tau}, a^*, \beta|\boldsymbol{X}, \boldsymbol{R})$ using a Laplace normal approximation (model with extensions).

5. Estimate the bandwidth h by Silverman's rule of thumb (or another bandwidth selection method) and recompute $\hat{f}_h$.

6. Repeat steps 2 - 5 B (burn-in iterations) + N (additional iterations) times.

7. Discard the burn-in samples and get final estimate of f by averaging over the remaining samples. The samples of the measurement error parameters $\boldsymbol{p}$ or $\boldsymbol{\tau}$, a^* and β can be used to compute a point estimate by averaging as well as uncertainty intervals.

2.3.4 Computational implementation in R

All computations were performed with R version 3.1.2 (R Core Team 2014). A package called *Kernelheaping* (Groß 2016a) was made available on CRAN by the authors. It includes the full functionality as presented in this article and an additional example dataset concerning the hours per week of learning reported by students (taken from Utts and Heckard 2014). Kernel density and bandwidth estimation is done via the *density*

function coming with the default installation of R. For non-negative data the boundary correction method introduced in Jones (1993) is utilized which is implemented in the *evmix* package (Scarrott and Hu 2014). Sampling weights in kernel density estimation are not supported by the *evmix* package yet. Hence, a kernel density estimation with design weights is currently only performed when no boundary correction is required by the user. For a sample size of $n = 5000$ and 1000 iterations the computation of the extended model with four rounding parameters takes about 5 minutes (15 minutes with boundary correction) on an Intel Core i7 3.4 Ghz system with 16 GB ram. The package also provides functions to perform convergence diagnostics and other convenience functions as well as functions to perform Monte-Carlo simulation studies.

2.4 Simulation Study

In this section we present results from a Monte-Carlo simulation study to evaluate the performance of the proposed kernel density estimator for heaped data. The properties of the estimator are investigated and its performance is compared to a simple $Naive$ kernel density estimator, which ignores the heaping process. The data are generated under different univariate distributions. Four scenarios, denoted by A-D, are considered. The sample size is always $n = 1000$. Under Scenario A we consider the heaping model without extensions. The data are generated as a normal distribution,

$$X_A \sim N(0, 100),$$

with rounding values $\boldsymbol{r} = (1, 10, 100)$ and rounding probabilities $\boldsymbol{p} = (0.3, 0.4, 0.3)$. This scenario also deals with rounding of negative numbers, which is rare but can occur, as for example with self-reported yearly profits from stocks of private investors.

In Scenario B we introduce a rounding bias with $a = 0.8$. Following the motivation example of a weight distribution, the data are generated by a gamma distribution with shape α and scale θ with offset:

$$X_B \sim Ga(\alpha = 4, \theta = 8) + 45$$

The rounding values are $\boldsymbol{r} = (1, 2, 5, 10)$ with arbitrarily chosen corresponding probabilities $\boldsymbol{p} = (0.1, 0.15, 0.4, 0.35)$.

In the third scenario the data follow a log-normal distribution with unequal rounding probabilities ($\beta = -1$) to model an income-like distribution,

$$X_C \sim logN(7, 0.6),$$

with rounding values $\boldsymbol{r} = (10, 20, 50, 100, 200, 500, 1000)$ and threshold values $\boldsymbol{\tau} = (-\infty, 6.33, 6.66, 7, 7.33, 7.66, 8, \infty)$. These threshold values coincide for rounding probabilities of $\boldsymbol{p} = (0.28, 0.12, 0.13, 0.13, 0.11, 0.09, 0.14)$ for $x = 1000$ or $\boldsymbol{p} = (0.01, 0.02, 0.03, 0.05, 0.08, 0.11, 0.70)$ for $x = 5000$.

A bimodal mixture of two normal distributions is considered in scenario D. With

$$X_{D_1} \sim N(40, 4) \text{ and } X_{D_2} \sim N(55, 6),$$

and mixture probabilities 0.4 (X_{D_1}) and 0.6 (X_{D_2}), an underlying heaping model with rounding bias $a = 0.2$ and unequal rounding probabilities ($\beta = -0.5$) with threshold values $\boldsymbol{\tau} = (-\infty, 1.84, 2.64, 3.05, \infty)$ is utilized in this case.

For each scenario we performed $n_{Sim} = 500$ simulation runs with B=100 burn-in iterations and N=500 additional iterations. We compare the following three estimators:

a) The *Naive* estimator, which naively applies the kernel density estimator to the heaped data

b) The *Corrected* estimator, that uses the algorithm presented in Section 2.3.4 for kernel density estimation for heaped data

c) The *Oracle* estimator, that uses the original data (which are only available in simulations) for density estimation.

Silverman's rule of thumb was used for bandwidth selection in each case. Figure 2.2 shows these three kernel density estimators as well as the true density from which the data are generated for a single simulation run of each scenario.

While the *Naive* estimator is very spiky and shows large deviations from the true density at the heaping points, the proposed *Corrected* density estimator is very close to the oracle estimator and represents the true density pretty well. In Scenario D, we are able to recover the bimodal structure of the distribution, whereas with the *Naive* estimator this feature of the data gets lost.

Tables 2 shows the RMISE (Root Mean Integrated Square Error) of of the three estimators for each scenario. While the *Naive* estimator exhibits a rather poor performance with a RMISE up to almost 3 times as high as with the non-feasible *Oracle* estimator, the *Corrected* estimator leads to a negligible loss of some percent in RMISE. This slightly worse performance of the proposed estimator can be most likely assigned to the information loss induced by rounding.

Table 2.2: Root Mean Integrated Square Error (RMISE) for scenarios A-D for each estimator. Standard errors are given in parenthesis.

Scenario	RMISE		
	Naive	*Corrected*	*Oracle*
A	0.0089 (0.0008)	0.0033 (0.0009)	0.0032 (0.0009)
B	0.0132 (0.0018)	0.0098 (0.0024)	0.0093 (0.0023)
C	0.0036 (0.0008)	0.0020 (0.0006)	0.0018 (0.0005)
D	0.0452 (0.0040)	0.0169 (0.0031)	0.0155 (0.0026)

Besides trying to recover the true distribution one might be also interested in estimating the rounding parameters. We investigate some (frequentist) properties, namely the bias, standard deviation, RMSE and the coverage rate of the 90% uncertainty in-

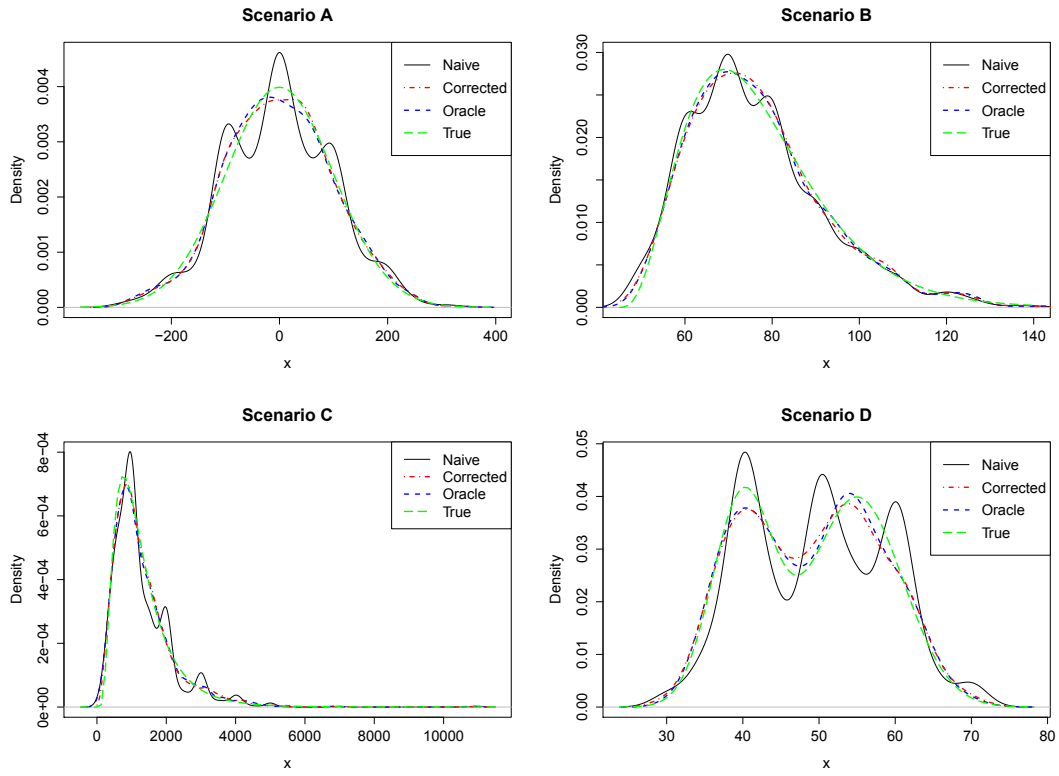

Figure 2.2: Graphical presentation of single simulation runs of scenarios A-D. The plots show kernel density estimators applied to heaped data (*Naive*, black solid line), applied to rounded data with correction algorithm (*Corrected*, red point-dotted line), applied to original data (*Oracle*, blue short-dashed line) and the true density function (*True*, green long-dotted line)

tervals of the estimates computed by the introduced algorithm. The results are shown in Tables 3 - 6.

Table 2.3: Scenario A: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter		
	p_1	p_2	p_3
True value	0.3	0.4	0.3
Bias	-0.0007	0.0013	-0.0006
SD	0.0156	0.0184	0.0156
RMSE	0.0157	0.0185	0.0156
Coverage in %	90.0	92.6	89.8

Table 2.4: Scenario B: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter				
	p_1	p_2	p_3	p_4	a
True value	0.1	0.15	0.4	0.35	0.8
Bias	0.0019	-0.0008	0.0010	-0.0020	-0.0193
SD	0.0177	0.0187	0.0338	0.0342	0.0572
RMSE	0.0178	0.0187	0.0339	0.0343	0.0603
Coverage in %	87.2	88.8	84.8	90.2	85.8

Table 2.5: Scenario C: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter						
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	β
True value	6.33	6.66	7	7.33	7.66	8	-1
Bias	-0.0117	0.0279	0.0167	0.0502	-0.0411	0.0200	-0.0360
SD	0.6289	0.6576	0.6993	0.7603	0.7084	0.6743	0.1876
RMSE	0.6290	0.6581	0.6994	0.7619	0.7095	0.6745	0.1841
Coverage in %	88.4	82.8	87.4	90.2	90.0	85.4	87.2

Apparently, the algorithm is able to identify the rounding parameters very well and the coverage rates of the 90% uncertainty intervals are near to the nominal value. One may note that the threshold values have a rather large standard deviation, but this is due to the high correlation with β. The resulting rounding probabilities are pretty stable, though. The results for the Sheather-Jones bandwidth selector were essentially equivalent for the proposed algorithm. As expected from the examples in the introduction, the RMISE for the $Naive$ estimator is much higher. The corresponding figures and tables can be found in the supplementary material.

In general, the algorithm was very stable for the proposed starting values and exhibited a very good and fast convergence behaviour. Depending on the application and heaping model $B = 5$ to $B = 500$ burn-in iterations were sufficient, but one should

Table 2.6: Scenario D: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter				
	τ_1	τ_2	τ_3	a	β
True value	1.84	2.64	3.05	0.2	-0.5
Bias	-0.0674	-0.0643	0.0681	-0.0039	-0.0052
SD	1.2912	1.2957	1.301	0.0769	0.1518
RMSE	1.2929	1.2972	1.3028	0.0770	0.1519
Coverage in %	86.2	86.2	87.6	84.6	86.8

always consider trace plots of the MCMC-chains to ensure convergence. Trace plots for both application examples can be found in supplementary material.

2.5 Application

We examine the two self-reported data examples of the SOEP 2010 already presented in the introduction. The first example is body weight data of $n = 8727$ German women. The weighted sample mean is 69.99 kg and the weighted standard deviation is 13.97 kg. We expect different probabilities for the rounding values depending on the actual weight. In particular, 48.6% of the respondents with reported weight above 90 kg report an end-digit of 0 or 5 while this is only the case for 36.3% of the group with reported weight lower than 90 kg. To investigate possible rounding bias, the heaping model with both extensions is utilized. For bandwidth estimation we used the Sheather-Jones estimate as well as Silverman's rule of thumb. The SOEP survey weights were utilized to adjust for unequal sampling probabilities for both the *Naive* and the *Corrected* method. The algorithm was executed with $B = 500$ burn-in samples and $N = 2000$ additional samples and with rounding values $\boldsymbol{r} = (1, 2, 5, 10)$. The resulting densities of both the *Corrected* and the *Naive* estimator are shown in Figure 2.3. For the Sheather-Jones bandwidth selector as well as for Silverman's rule of thumb, the algorithm produces smoother and more realistic density estimates than the *Naive* method.

Table 2.7 shows rounding parameter estimates. The threshold values $\boldsymbol{\tau}$ and the slope parameter of the ordered probit β suggest rounding probabilities of
$\boldsymbol{p} = (0.733, 0.080, 0.167, 0.020)$ for the rounding values $\boldsymbol{r} = (1, 2, 5, 10)$ at the sample mean. The point estimate of the rounding bias a is 0.83, indicating as expected that the survey respondents are much more likely to round off than to round up. As a consequence, the mean of the imputed weights X_i is more than 200 g higher now (cf. Table 2.8 for comparison of *Naive* and *Corrected* sample means and standard deviations). The lower border of the 95% uncertainty interval for a is considerably above 0.5. However, to further validate this result we ran the algorithm on a different survey data sample including a self-reported body weight variable, namely the German General Social Survey 2008 ('ALLBUS', Wasmer et al. 2007). In this survey $n = 1451$ women reported their body weight and the rounding bias was estimated to $a = 0.694$

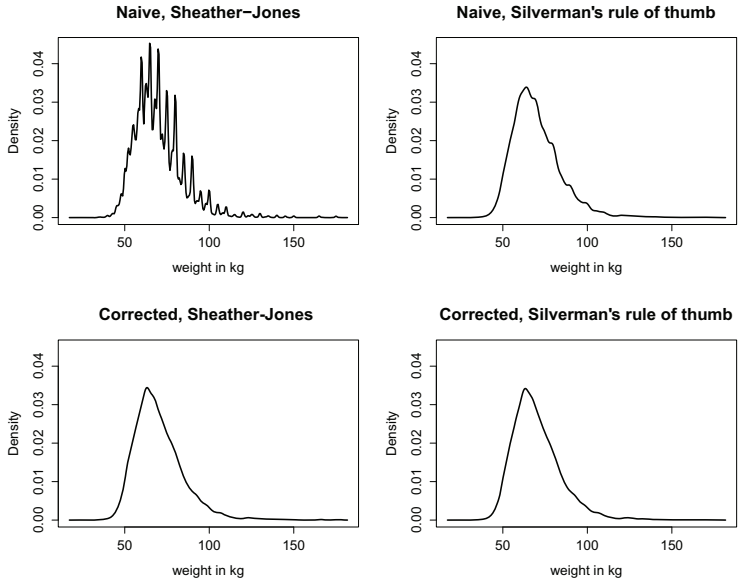

Figure 2.3: Kernel density estimation of self-reported female body weight for *Naive* (upper panel) and *Corrected* (lower panel) method for different bandwidth choices.

(with 95% uncertainty interval [0.503,0.853]), which conforms with the estimate on the SOEP data. Men, as a remark, were less prone to biased rounding with point estimation values of $a = 0.591$ for SOEP and $a = 0.569$ for ALLBUS.

Table 2.7: SOEP female body weight: Mean, standard deviation and 95% coverage intervals for rounding parameters. Bandwidth selector: Silverman's rule of thumb.

Parameter	Mean	SD	95% uncertainty interval
τ_1	10.485	0.594	[8.952, 12.273]
τ_2	10.751	0.588	[9.236, 12.494]
τ_3	11.910	0.585	[10.398, 13.624]
a	0.834	0.026	[0.773, 0.918]
β	-2.322	0.134	[-2.704, -1.981]

In the second SOEP example, households were asked to state their monthly food and drink expenditure outside home. The $n = 6096$ respondents stated a mean expenditure of 92.42€ with a standard deviation of 78.07€ . The algorithm was applied with rounding values $\boldsymbol{r} = (1, 2, 5, 10, 20, 50, 100)$. The heaping model with the ordinal probit model extension for non-constant rounding probabilities was utilized here, as the data suggest strong dependence of rounding behaviour on the magnitude of the expenditures. All reported values above 180€ are divisible by 10, while at least 6.7% of the reported values below 100€ are not. The rounding direction bias extension was spared here, because the authors believe that a biased response behaviour is rather unlikely in this application (the estimate of a in the full model was very close to 0.5 anyways). Contrary to the body weight example we could not account for sampling weights, because a boundary-corrected kernel density estimator was needed here (cf. Section 2.3.4). Figure

Table 2.8: Comparison of *Naive* and *Corrected* sample means and standard deviations.

Application 1: Body weight	Mean	SD
Naive	69.99	13.97
Corrected	70.22	14.04
Application 2: Expenditures outside home		
Naive	92.42	78.07
Corrected	92.07	77.22

2.4 displays the resulting density estimates for different bandwidth choices. Again, for the Sheather-Jones bandwidth selector, the algorithm produces a markedly improved, though still quite rough, density estimate. For Silverman's rule of thumb, the estimate is smooth but shows a bimodal structure that may not apply to the underlying true expenditures. To produce a sufficiently smooth estimate, the authors suggest manually tuning the bandwidth. A bandwidth of 1.5 times the rule of thumb generates a smooth unimodal density estimate, while the *Naive* approach is still very spiky. Further increasing the bandwidth to 4 times the rule of thumb was necessary to create a comparably smooth estimate for the *Naive* method. However, the resulting density was considerably flatter and quite different for values less than 100 € (cf. Figure 2.4).

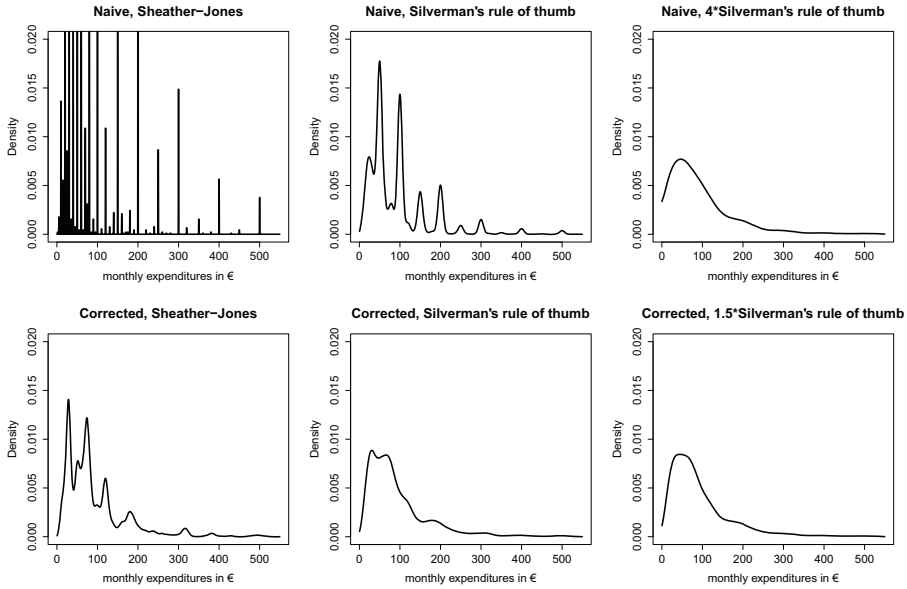

Figure 2.4: Kernel density estimation of food and drink expenditures outside home in € for *Naive* (upper panel) and *Corrected* (lower panel) method for the following bandwidth selection methods: Sheather-Jones (left), Silverman's rule of thumb (middle), 4 (*Naive*) and 1.5 (*Corrected*) times Silverman's rule of thumb (right).

The summary statistics for the rounding parameters $\boldsymbol{\tau}$ and β can be found in Table 2.9. The negative value of β indicates that higher rounding values ($\boldsymbol{r} = (1, 2, 5, 10, 20, 50, 100)$) are utilized for higher monthly expenditures. Specifically, for expenditures of 25€ the model suggests rounding probabilities $\boldsymbol{p} = (0.8, 0.4, 19.4, 48.4, 11.9, 18.9, 0.1)$, while $\boldsymbol{p}$ equals $(0.0, 0.0, 0.2, 5.6, 5.9, 73.8, 14.5)$ for monthly expenditures of 150€.

Table 2.9: Food and drink expenditures outside home: Mean, standard deviation and 95% coverage intervals for rounding parameters. Bandwidth selector: Silverman's rule of thumb.

Parameter	Mean	SD	95% uncertainty interval
τ_1	1.322	0.183	[0.951, 1.662]
τ_2	1.479	0.129	[1.223, 1.738]
τ_3	2.896	0.124	[2.656, 3.138]
τ_4	4.213	0.137	[3.952, 4.484]
τ_5	4.592	0.135	[4.333, 4.854]
τ_6	6.840	0.172	[6.508, 7.189]
β	-1.154	0.0316	[-1.215, -1.093]

The algorithm converged to the same parameter values under multiple runs and different starting values for both examples (and the simulation scenarios). Trace plots for rounding parameters of the SOEP data are shown in the supplementary material. Convergence is achieved after a burn-in period of about 50 iterations. The density estimates and the rounding parameters a and β were relatively robust to different choices of rounding values (for example $\boldsymbol{r} = (1, 5, 10)$ or $\boldsymbol{r} = (1, 2, 5, 10, 20)$ in the body weight example). However, in general, for rounding values which are not or very weakly supported by the data, the estimates (especially the threshold values as well as β) can be unstable. The user should always consult the trace plots and eliminate the questionable rounding values if necessary.

2.6 Further Model Extensions

The heaping model introduced in Section 2.2 is interpretable and identifiable without evaluation data. However, although the developed algorithm can greatly reduce the bias in kernel density estimation for self-reported data, the second application indicates that the modeling assumptions may not be completely fulfilled in real-world data. As Crawford et al. (2015) remarks, the assumption that, for example, a reported value of $W_i = 100$ with rounding value $R_i = 10$ means that the true unobserved value X_i lies inside the interval (95,105) is rather strong. Wang et al. (2012), for example, found that given exact evaluation data fewer than half of the observations are explainable by rounding. A possible solution would be to decompose the reporting process into an recall error (i.e. the person does not know its true value exactly) and a rounding error as already mentioned at the end of Section 2.2.2. This could be achieved by introducing a latent intermediate variable V_i $(i = 1, .., n)$, which can be interpreted as the value the participant is recalling. A classical Gaussian error model for V_i given X_i, the recall error model, could be imposed:

$$\pi(V_i|X_i) = N(X_i + \mu_V, \sigma_V^2)$$

It is not clear how to estimate the recall error model parameters without validation data, which are rarely ever available. However, one can presume that they are associated with

the parameters of the rounding model. The authors suggest that μ_V could be chosen such that $P(V_i - X_i < 0) = a$ and $\sigma_V = \nu R_i + \eta \sum_{j=1}^{m} p_{ij} R_j$. For $\nu, \eta = 1/3$, a quite similar pattern as in Wang et al. (2012) – with recall error slightly larger than rounding error – is produced and the model likelihood of (2.4) becomes:

$$L(\boldsymbol{W}|\boldsymbol{V}, \boldsymbol{X}, \boldsymbol{R}, \boldsymbol{\tau}, a, \beta, \boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(W_i|V_i, R_i) \times \pi(V_i|X_i, R_i, \boldsymbol{\tau}, a, \beta) \times \\ \pi(R_i|X_i, \boldsymbol{\tau}, a, \beta) \times \pi(X_i|\boldsymbol{\theta})$$

The accordingly modified algorithm produces considerably smoother – but not flatter – density estimates. However, although the authors choice of the recall error parameters may appear reasonable, this non-data driven proceeding is of rather speculative nature and the parameters may vary for different applications as well. Additional research on applications with evaluation data is advisable to find appropriate recall error models and associated parameters, but this is future work.

Another issue is that the preference for some heaped values may not be fully captured by the model, i.e. some numbers are more popular than predicted by the model. A recent publication of Zinn and Wuerbach (2015) relies on user-specified heaping values detected by empirical analysis. An alternative approach followed by the authors is to introduce a random effect γ_l in the ordered probit model for the rounding probabilities for each observed value. Specifically, a grouping structure which assigns every X_i to the nearest possible rounded value is introduced (represented by design matrix $\boldsymbol{U}$ with rows u_i):

$$g_i = log(X_i)\beta + u_i'\boldsymbol{\gamma} + \epsilon_i \ , \ \epsilon_i \sim N(0,1), \boldsymbol{\gamma} \sim N(0, \boldsymbol{\tau})$$

This extension produces smooth density estimates regardless of the chosen bandwidth selector for the price of higher computational cost, slower convergence (a Metropolis-Hastings step is necessary) and somewhat reduced numerical stability. Both extensions are straightforward to implement and are integrated into the R-package. For both extensions, density estimates for the second application example can be found in the supplementary material. A further extension could introduce a non-constant rounding bias as well. Respondents with overweight, for example, are possibly more inclined to round off than normal or underweight surveyed persons. Additionally, the estimation of parametric distributions is straightforward to integrate into this approach and with some minor modifications of the algorithm density estimation for classified data should be possible as well.

2.7 Conclusion

In this work, a novel approach for kernel density estimation for heaped data was introduced. A Stochastic Expectation-Maximization algorithm was presented, that generates smoother and more realistic non-parametric density estimates and gives additional

insights into the rounding process. More specifically, the rounding probabilities as well as a rounding bias is estimated within the proposed algorithm. This can be very helpful in assessing and validating self-reported data. In the presented example of self-reported body weight the approach was able to discover a biased response behaviour without validation data solely on the basis of reported values. Concerning the improved but still spiky density estimates under the Sheather-Jones bandwidth selector, the authors recommend to use the Silverman's rule of thumb instead and tune the bandwidth manually if necessary, or cautiously apply the extensions presented in the previous subsection. The algorithm is easy to implement and is provided by the authors in a R-package. The algorithm exhibited very good statistical properties in the simulations. In sum, the algorithm presented in this work delivers a powerful and easy to use tool for users concerned with heaped data.

2.8 Supplementary Material

Figures and Tables for Simulations with Sheather-Jones bandwidth selector

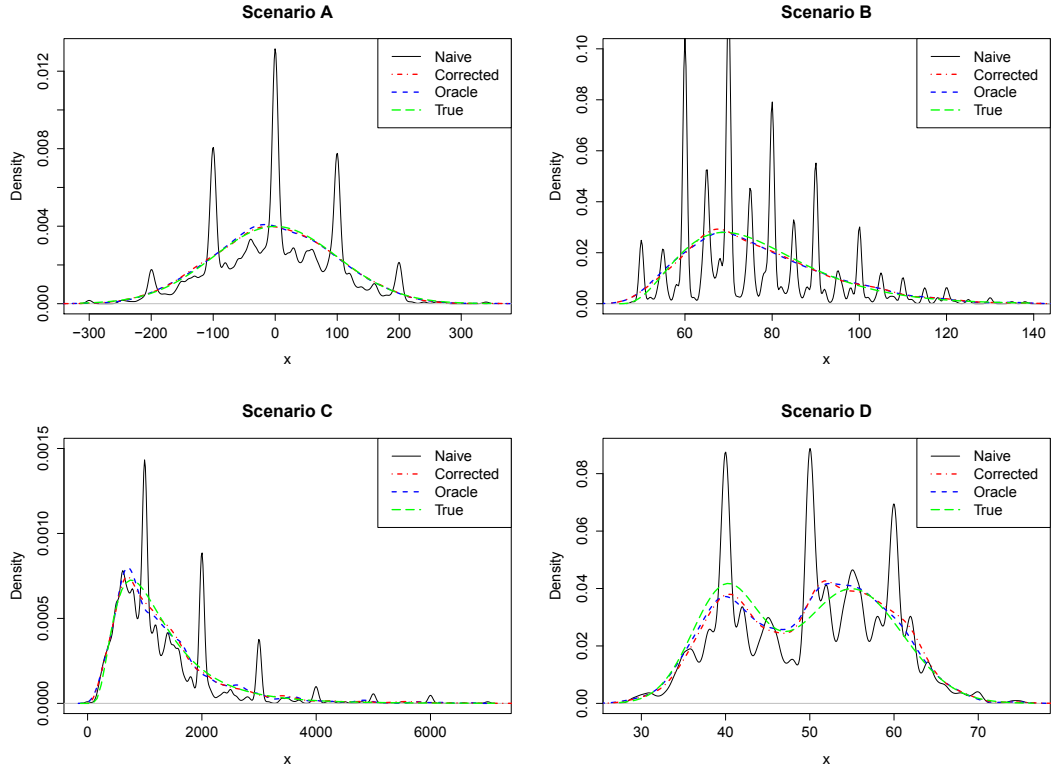

Figure 2.5: Graphical presentation of single simulation runs of scenarios A-D. The plots show kernel density estimators applied to heaped data (*Naive*, black solid line), applied to rounded data with correction algorithm (*Corrected*, red point-dotted line), applied to original data (*Oracle*, blue short-dashed line) and the true density function (*True*, green long-dotted line)

Table 2.10: Root Mean Integrated Square Error (RMISE) for scenarios A-D for each estimator. Standard errors are given in parenthesis.

Scenario	RMISE		
	Naive	*Corrected*	*Oracle*
A	0.0330 (0.0041)	0.0032 (0.0010)	0.0031 (0.0010)
B	0.1622 (0.0726)	0.0101 (0.0032)	0.0100 (0.0026)
C	0.0133 (0.0022)	0.0018 (0.0006)	0.0017 (0.0006)
D	0.2196 (0.0085)	0.0159 (0.0029)	0.0145 (0.0023)

Table 2.11: Scenario A: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter		
	p_1	p_2	p_3
True value	0.3	0.4	0.3
Bias	-0.0030	0.0018	0.0013
SD	0.0143	0.0168	0.0144
RMSE	0.0147	0.0169	0.0144
Coverage in %	88.6	87.0	93.8

Table 2.12: Scenario B: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter				
	p_1	p_2	p_3	p_4	a
True value	0.1	0.15	0.4	0.35	0.8
Bias	0.0012	0.0029	-0.0046	-0.0005	-0.0079
SD	0.0164	0.0188	0.0344	0.0339	0.0501
RMSE	0.0165	0.0191	0.0348	0.0339	0.0507
Coverage in %	89.0	86.2	93.4	91.2	89.2

Table 2.13: Scenario C: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter						
	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	β
True value	6.33	6.66	7	7.33	7.66	8	-1
Bias	-0.0493	-0.0327	0.0165	0.0415	0.0410	-0.0307	-0.0353
SD	0.7564	0.8167	0.9243	0.8342	0.5915	0.6281	0.2412
RMSE	0.7580	0.8174	0.9244	0.8352	0.5929	0.6288	0.2431
Coverage in %	87.8	85.6	92.2	94.6	90.4	87.2	88.6

Table 2.14: Scenario D: Bias, standard deviation, Root Mean Square Error and coverage rate of 90% uncertainty intervals for rounding parameters.

	Parameter				
	τ_1	τ_2	τ_3	a	β
True value	1.84	2.64	3.05	0.2	-0.5
Bias	-0.0545	-0.0561	0.0562	-0.0059	-0.0093
SD	1.1705	1.1532	1.1601	0.0524	0.1323
RMSE	1.1717	1.1545	1.1615	0.0527	0.1326
Coverage in %	89.0	92.8	87.0	91.4	90.8

Trace plots for application examples

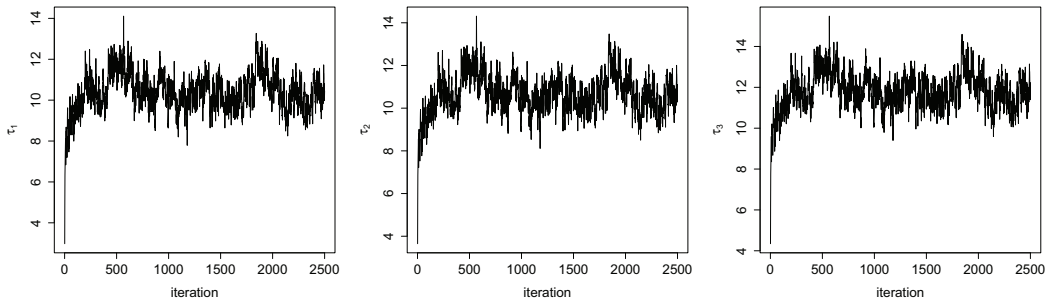

Figure 2.6: Trace plots for $\boldsymbol{\tau}$ (SOEP female body weight).

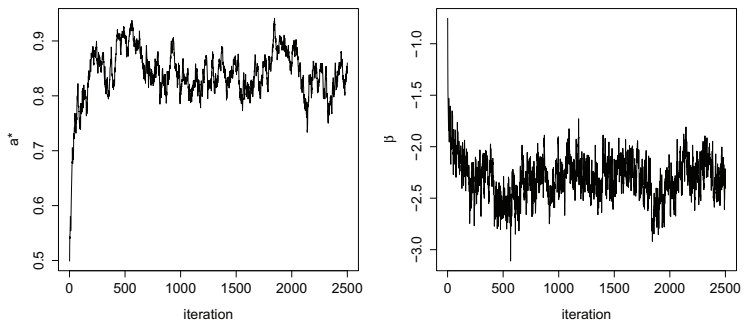

Figure 2.7: Trace plots for a and β (SOEP female body weight).

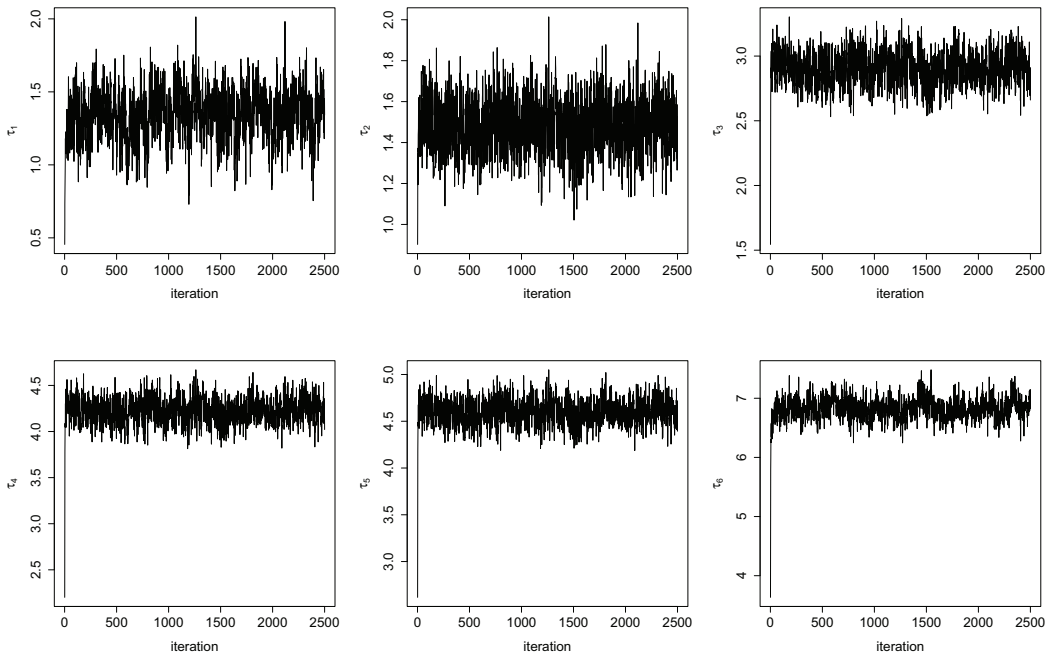

Figure 2.8: Trace plots for $\boldsymbol{\tau}$ (SOEP monthly expenditures outside home).

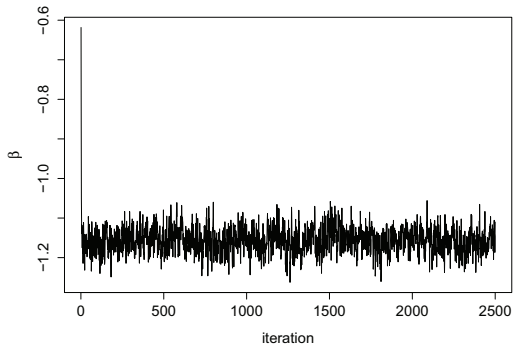

Figure 2.9: Trace plot for β (SOEP monthly expenditures outside home).

Density estimates for monthly expenditures outside home example

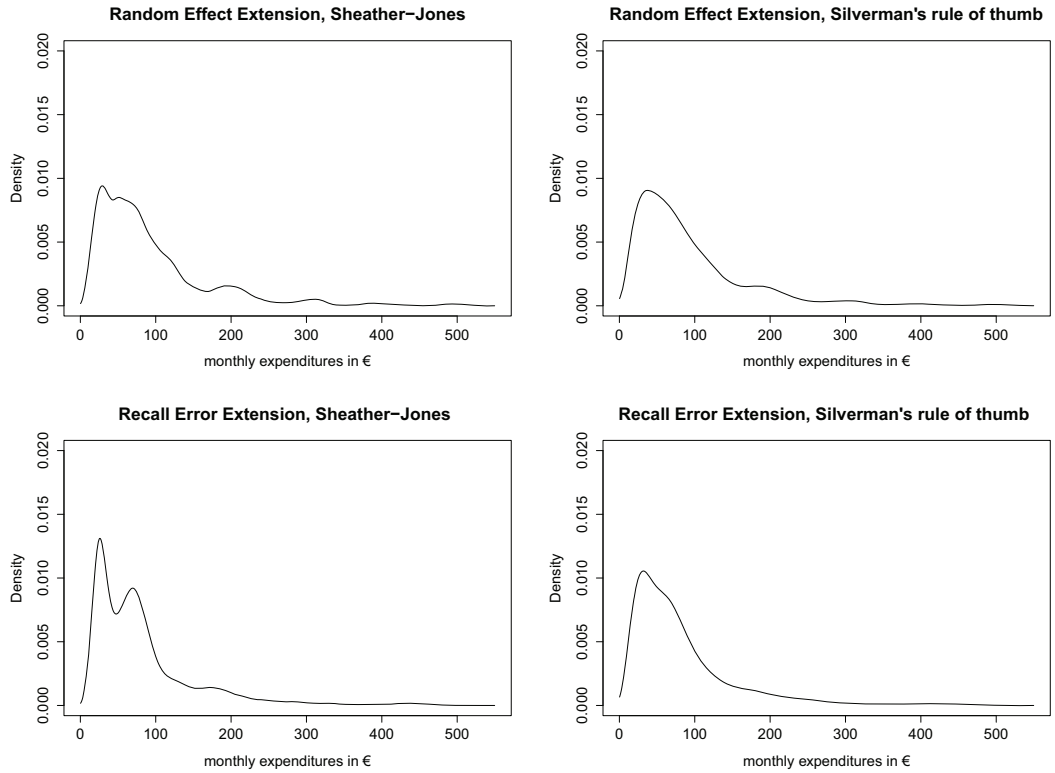

Figure 2.10: Kernel density estimation of food and drink expenditures outside home in € for the recall error (lower panel) and random effects extensions (upper panel) for the following bandwidth selection methods: Sheather-Jones (left), Silverman's rule of thumb (right).

Teil II

Messfehlermodelle für die Wirtschaftsarchäologie

Kapitel 3

Back to Good Shape: Biological Standard of Living in the Copper and Bronze Ages and the Possible Role of Food

3.1 Introduction: Anthropological Indicators of Nutrition

While archaeobotanical and archaeozoological remains (e.g. Röpke et al. in this volume; van Amerongen in this volume) as well as food-related artefacts inform us on the gross input of nutrients, anthropological data partially related to consumption practices can be used to assess alimentary practices and thus test our assumptions about food supply in a community in general or for various subgroups like women or men, age groups or social strata. Important parameters include demographic data, such as infant mortality and life expectancy, which were often linked to malnutrition (e.g. Martorell and Ho 1984) in pre-modern times. Other crucial proxies for nutritional practises can be present, such as individual physical signs of systemic stress, i.e. nutrient deficiency-related pathological markers, including tooth enamel hypoplasia or Harris Lines in the long bones, as well as the body height of the respective individuals (e.g. Haidle 2014). Additionally, dental caries and calculus as well as tooth wear are directly connected to the consumption of different food types (Lieverse Angela, 1999; Zero et al., 2009). Inferences regarding diet and subsistence can also be drawn from isotope analysis: isotopic data on δ15N retrieved from skeletal collagen can provide information on the trophic level of an individual on the vegan to carnivore scale as well as on the provenance of the protein from either marine or terrestrial sources. Additionally, δ13C values can provide estimates for the proportion of C4 and C3 plants (for details see Scheibner in this volume). Evidence ranging from isotopes (Salazar-García et al., 2014; Scott and Poulson, 2012) to actual food tissue remains (Henry and Piperno, 2008; Hardy et al., 2009) can even be extracted from dental calculus, a feature forming an intermediate class of source information between anthropological, archaeobotanical,

and archaeozoological material, much the same way as food crusts and lipid remains relate to the vessels in which they are found. While retrieving demographic data from archaeological/anthropological data is notoriously difficult, as selective burial customs as well as preservation conditions often turn cemeteries into unrepresentative samples, individual data, especially body height data, is easier to handle for early time periods and is the proxy used here. The first part of this work summarizes methodological problems arising when skeletal data is operationalized for body height studies and outlines the current state of research on the development of body height in Southwestern Asia and Europe in the Copper Age and the Bronze Age, i.e. what is commonly called Later Prehistory, starting with the global Old World appearance of cast copper technology at ca. 5000 cal BC. Since diet and especially protein intake in nutrition is the most important determinant for body height (see below), the second part examines the various protein sources available to prehistoric people in terms of protein content and quality, before changes in the food spectrum and composition are compared between the Neolithic and the Copper and the Bronze Ages and discussed as influential factors in net nutrition and thus body height. We thus recommend reading this work on some of our preliminary results after lunch or dinner, as we set the Bronze Age table not so much with speculations on the age of some traditional dishes, but mainly with amino acids and bioactive proteins. As a conclusion, we try to bring together body height and food development between the Neolithic and the Bronze Age in Southwestern Asia and Europe.

3.2 Body Height as a Proxy of Welfare

3.2.1 Genetic and environmental determination of human body height

The adult body height of an individual is the result of both the individual's genetic endowment and certain environmental factors prevalent during its growth period (Silventoinen, 2003). There is a consensus in auxology that the maximum height an individual can reach is determined genetically and that several genes are responsible in a polygenic process. While some responsible genes are located on the X and Y chromosomes and explain why women are in the mean shorter than men, other possible genes have been located on other chromosomes (Silventoinen, 2003; Weedon and Frayling, 2008), and it is still a matter of debate if there are genetic differences between African, European and Asian populations. Gene-driven timing of the adolescent growth spurt reacting to hormone-regulated growth factors, like IFG 1, is discussed as one possible reason (Silventoinen, 2003), which is corroborated by findings in the African 'pygmies' who miss the adolescent growth spurt altogether (Bozzola et al., 2009). Notoriously hard to quantify, studies into the hereditability of body height resulted in proportions varying between 40% and 90%, with most of them around 80%, thus leaving ca. 20% to environmental factors (Silventoinen, 2003). Moreover, there are indications that the proportion of the variation due to environmental factors is more important when environmental stress is strong (Silventoinen, 2003). Here, nutrition and disease appear to

be the most influential factors. In addition to macro-nutrients, such as carbohydrates, lipids and proteins, as well as various micro-nutrients, such as minerals and vitamins which are needed to maintain good nutritional health Biesalski (2010), protein appears to play the most important role for growth according to various studies (Silventoinen, 2003). While carbohydrates and fat account for most of our basal energy rate, only ca. 15% of that is met by proteins. Rather, nutritional protein is mainly used in the human body to synthesize functional proteins, e.g. hormones and structure proteins. Ca. 1/5 of our body mass, including the collagen in bones, is formed by means of amino acids, the components of proteins of which some are essential and cannot be synthesized by the body itself. In times of starvation, when the full energy intake is not sufficient, our body resorts to protein catabolism. A sufficient net supply in carbohydrates and fats beyond the necessary caloric requirements to keep the body temperature stable, move the body and fight diseases is thus a prerequisite for the adequate use of the ingested protein for tissue production (Vaupel and Biesalski, 2010). Evidently, growing individuals, like children and juveniles, need better protein quality, i.e. a few more amino acids than the essential amino acids needed by older people, which are consequently called semi-essential (Vaupel and Biesalski, 2010). Improving net nutrition in protein through an increased amount of highly bioavailable protein in the diet, better general calorie supply, a decreased workload and better medical support is thus to be regarded as the main reason for the increase in body height during the late nineteenth and the twentieth century in western industrialized nations (Cole, 2003), although a direct stunting influence of physical workload on growth is also debated (Ambadekar et al., 1999). Body height can thus be viewed as a proxy for the 'biological standard of living' and has been used in econometry since John Komlos' initial work (1989) on the economics of the Habsburg monarchy as an alternative economic key figure for times when and regions where other indicators, such as the gross national product or per capita income, are unreliable or not available, or as a controlling feature for those figures. Studies across populations have to check for population differences as a possible disruptive factor. Due to the uncertainties of the genetic component, at least a four-figure sample of body heights is needed to cancel out the influence of individuals when comparing different populations.

3.2.2 Operationalizing skeletal data for prehistoric body height studies

The transfer of the approach to skeletal material poses one advantage and a number of difficulties, which have to be overcome using statistical methods. While individuals who are already mature must be excluded from studies on living heights because of the shrinking of the inter-vertebral discs, all grown-up skeletons with fused long bone epiphyses (Szilvássy, 1988) can be included: the spine is not used in the formulae, and as long as the joints do not show excessive loss of tissue due to arthritic diseases, the long bones of even a senile individual whose back is bent by old age can inform us about the original adult body height Trotter and Gleser (1952). The stature (S),

as defined by Knußmann (1988), however, cannot be assessed directly as can be done on a living human or a dead body, but has to be inferred from bones. Due to the usually incomplete preservation of a skeleton, the method developed by Fully (1956) cannot be applied and body height has to be calculated from long bone lengths, e.g. from the maximum lengths of the femur (F1). A variety of methods have been developed in physical anthropology during the past one hundred years (for an overview cf. Krogman 1962; Wurm 1986; Rösing 1988; Formicola 1993; Siegmund 2010). As they were all generated using different reference populations, ranging from Mesolithic and Neolithic European skeletons that could be reconstructed using the Fully (1956) method (Formicola and Franceschi, 1996) to late nineteenth century anatomy corpses in Lyon (Manouvrier, 1892; Pearson, 1899; Mollison, 1910) and living women of the 1960s in the German Democratic Republic (Bach, 1965), reconstructed body height has to be regarded as an apparent value, which the individual would have had under the assumption that she or he had belonged to the reference population (Röhrer-Ertl, 1978; Rösing, 1988) – which is, of course, never the case. It is thus neither feasible to directly compare body heights that are reconstructed using different methods, nor to compare reconstructed body heights with body heights measured on living or dead bodies: comparability can only be achieved if all body heights are calculated using the same formula and the same reference population, the approach taken in this work. For

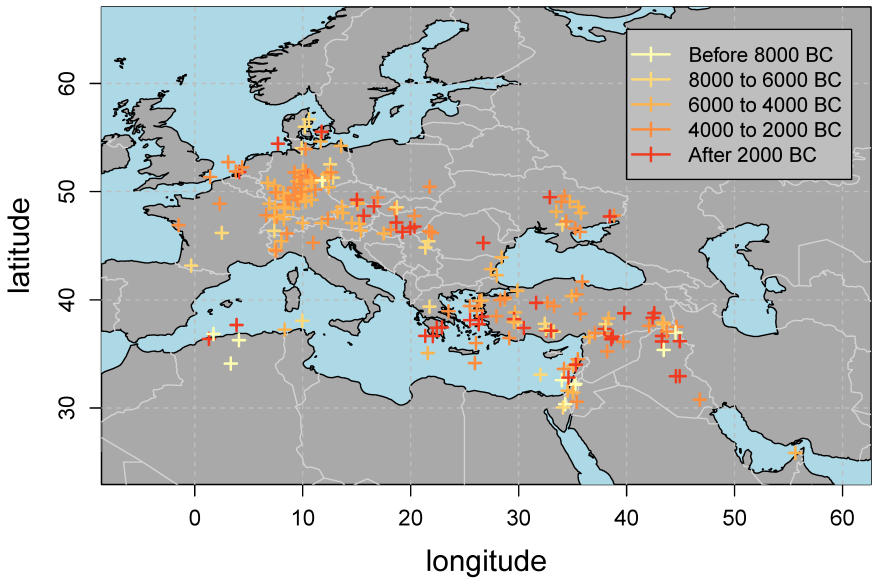

Figure 3.1: Distribution of archaeological sites over time categories (with small random error added to the latitude and longitude). For a list of sites and references see Table 3.1)

the present study, a preliminary dataset of 3052 skeletons from 219 sites (Fig. 3.1; for a list of sites and the respective references see Table 3.1), partially taken and updated with regard to chronology from the 'Mainzer Lochkartenarchiv für postcraniales

Skelettmaterial' (Perscheid, 1974) and partially collected from other printed sources, was used to calculate body height. Due to standardized long bone measures defined by Martin (1928), the inter-rater reliability should be sufficiently high, leading to negligible differences between the measures taken from different observers compared to the population variance. We are using the formulae of Pearson (1899), because they yield good approximations for both males and females in general (Formicola, 1993). The data was weighed according to a reliability category, reflecting the varying quality of literature references and the state of preservation of skeletons' long bones (number of skeletons with reliability category given in parentheses):

- Category 1: femur or humerus length available (1183)
- Category 2: other long bone measures available (460)
- Category 3: body height available with known formula applied to unknown long bone (1330)
- Category 4: body height available with unknown formula applied to unknown long bone (79)

Sex is an important characteristic not only because of the genetic height dimorphism (see above) but also because the reconstruction formulae are sex-specific. While the determination of an individual's sex is, in most cases, straightforward in living persons, it can only be assessed with some amount of uncertainty for skeletal remains unless DNA analyses are performed, depending on whether the pelvis is preserved or not (Sjøvold, 1988). If the skull is used for sexing (Acsadi, 1970; Graw et al., 1997, 2005), it has to be considered that some criteria are partly dependent upon the strength of the muscles attached, which, however, react to workload and thus environmental factors and render them less reliable. Such cases and cases in which sex is determined by the degree of muscle traces or the length of the long bones (as in e.g.Angel 1971; Çiner 1964) must be carefully controlled statistically. In our sample, the sex variable has five categories according to (Acsadi, 1970):

- Category 1: female (1074)
- Category 2: likely female (79)
- Category 3: uncertain (147)
- Category 4: likely male (145)
- Category 5: male (1607)

Another uncertainty concerns the usually broad archaeological dating spans: while in living individuals or historical height data, the birth date of a person is usually known or at least ratable, skeletons can often only be dated with a precision of centuries or more. While solutions tackling the problem at the stage of data input have been proposed (Nakoinz, 2012), here a Bayesian Errors-in-Variables model (Carroll et al.,

2006) was utilized to correctly account for these uncertainties in the estimation process (Groß (2016b) for details). Last but not least, as stable isotope studies have only yielded evidence for migration in a few cases Bentley (2004), the only proxy we have for the place or region where an individual has lived is the location of the burial of its skeleton.

Table 3.1: List of sites used in the data analyses.

Name	Country	References
Acemhöyük	Turkey	Güleç 1989
Afalou-Bou-Rhummel	Algeria	Vallois 1952
Ahlatlibel	Turkey	Güleç 1989
Ain Mallaha, Eynan	Israel/Palestine	Asmus 1973; Belfer-Cohen et al. 1991
Akkermen	Ukraine	Konduktorova 1956
Alaca Höyük	Turkey	Güleç 1989; Tunakan 1965; Şenyürek 1951
Alepotrypa Cave	Greece	Papathanasiou 2001
Altıntepe	Turkey	Çiner 1965
Arad	Israel/Palestine	Haas 1970
Arene Candide	Italy	Parenti and Messeri 1962
Argos	Greece	Charles 1958
Argusgrunden	Denmark	Bennicke 1987
Arma dell'Aquila	Italy	Parenti and Messeri 1962
Ascott-under-Wychwood	Great Britain	Galer 2007
Aşıklı Höyük	Turkey	Özbek 1991
Bab edh-Dhra	Jordan	Ortner and Fröhlich 2008
Babaköy	Turkey	Angel 1941; Angel 1953
Ballenstedt	Germany	Gerhardt 1953
Balmazújváros-Árkusmajor (Hortobágy)-Kettőzshalom	Hungary	Marcsik 1979
Battonya	Hungary	Szalai 1999
Bebertal	Germany	Bach 1978
Benzingerode	Germany	Berthold et al. 2008
Bernburg - Neu-Borna	Germany	Gerhardt 1953
Biebrich	Germany	Gerhardt 1953
Birsmatten	Switzerland	Asmus 1973
Bischleben	Germany	Bach 1978
Brachwitz	Germany	Bach 1978
Brackenheim	Germany	Orschiedt 1998
Bruchstedt	Germany	Bach 1978
Budakalász	Hungary	Köhler 2009
Butzbach	Germany	Preuschoft 1962
Büyük Güllücek	Turkey	Güleç 1989; Şenyürek 1950
Byblos	Lebanon	Vallois 1937
Cabeço Da Amoreira	Portugal	Cunha and Cardoso 2001
Çatalhöyük	Turkey	Ferembach 1982
Çayönü	Turkey	Özbek 1988
Columnata	Algeria	Chamla et al. 1970
Combe-Capelle	France	Asmus 1973
Darmstadt	Germany	Gerhardt 1953
Dévaványa-Barcéhalom	Hungary	Marcsik 1979
Dillingen/Steinheim	Germany	Haidle 2014
Dirmil	Turkey	Tunakan 1964
Durankulak	Bulgaria	Yordanov and Dimitrova 2002
Dürrenberg	Germany	Asmus 1973
Efringen	Germany	Gerhardt 1953
Eisleben	Germany	Bach 1978
El-Argar	Spain	Kunter 1990
Engen-Welschingen	Germany	Hald and Wahl 2009
Erfurt	Germany	Bach 1978
Erfurt-Löberkaserne	Germany	Gerhardt 1953
Erstein	France	Knussmann and Knussmann 1978
Evdi Tepesi	Turkey	Çiner 1964
Fulda "Schulzenberg"	Germany	Sangmeister and Gerhardt 1965; Gerhardt 1953
Gemeinlebarn	Austria	Neugebauer 1991
Gerlingen	Germany	Orschiedt 1998
Girikihaciyan	Turkey	Watson and LeBlanc 1990
Großbrembach	Germany	Ullrich 1961
Großkorbetha	Germany	Bach 1978
Großörner	Germany	Bach 1978
Großschwabhausen	Germany	Bach 1978
Grushkeva	Ukraine	Konduktorova 1974
Hagias Kosmas	Greece	Angel 1945
Haid	Austria	Jungwirth and Kloiber 1973 Kloiber and Kneidinger 1970
Hakkâri	Turkey	Gözlük et al. 2002
Halawa	Syria	Orthmann 1981
Halle-Trotha	Germany	Bach 1978; Gerhardt 1953
Hanaytepe, Hanai Tepe	Turkey	Angel 1951
Hankenfeld	Austria	Jungwirth and Kloiber 1973

Hauslabjoch	Italy	Hopfel et al. 1992
Hayonim Cave	Israel/Palestine	Belfer-Cohen et al. 1991
Heddesheim	Germany	Sangmeister and Gerhardt 1965
Heidesheim	Germany	Gerhardt 1953
Hisarlık/Troia	Turkey	Wittwer-Backofen and Kiesewetter 1997; Angel 1951
Hochstätt	Germany	Sangmeister and Gerhardt 1965
Hoëdic und Téviec	France	Asmus 1973
Hohlstedt	Germany	Gerhardt 1953
Hoiersdorf	Germany	Gerhardt 1953
İkiztepe	Turkey	Wittwer-Backofen 1985; Becker 1988
Ilıca	Turkey	Güleç 1989; Helmuth 1967
Ilıpınar	Turkey	Alpaslan-Roodenberg 2002
Ilvesheim	Germany	Gerhardt 1953
Jebel Buhais	Oman	Kiesewetter 2003
Jericho	Israel/Palestine	Röhrer-Ertl 1978
Kafer edj-Djarra, Kfar Jarra	Lebanon	Vallois and Ferembach 1962
Karataş	Turkey	Güleç 1989
Kebara	Israel/Palestine	Belfer-Cohen et al. 1991
Kétégyháza	Hungary	Marcsik 1979
Khirokitia	Cyprus	Angel 1953
Klein Hadersdorf	Austria	Tiefenböck 2010;Jungwirth and Kloiber 1973
Klingenberg	Germany	Orschiedt 1998
Koelbjerg	Denmark	Asmus 1973
Königsaue	Germany	Bach 1978
Korbetha	Germany	Bach 1978
Korsør	Denmark	Asmus 1973
Köthen	Germany	Gerhardt 1953
Kreuznach-Martinsberg	Germany	Knussmann and Knussmann 1978
Kültepe	Turkey	Angel 1951; Şenyürek 1949
Kumtepe	Turkey	Angel 1951; Şenyürek 1949
Kusura	Turkey	Güleç 1989; Kansu and Atasayan 1939
Kut	Ukraine	Konduktorova 1974
La Pollera Finale Lige	Italy	Parenti and Messeri 1962
La Trache Chaâteaubernard	France	Riquet 1962
Langendorf	Germany	Gerhardt 1953
Lauda-Königshofen	Germany	Menninger 2008
Lebendorf	Germany	Bach 1978
Lepenski Vir	Serbia	Schwidetzky 1973
Lerna	Greece	Angel 1971
Leuna-Daspig	Germany	Bach 1978
Lichtensteinhöhle	Germany	Schilz 2006
Lidar, Kamus Tepe	Turkey	Güleç 1989; Wittwer-Backofen 1987
Lingolsheim	France	Knussmann and Knussmann 1978
Lomovatoe	Ukraine	Konduktorova 1974
Ludwigshafen	Germany	Gerhardt 1953
Mangolding	Germany	Gerhardt 1968
Mannheim-Vogelstang "Am Schultheißenberg"	Germany	Orschiedt 1998
Mansfeld	Germany	Gerhardt 1953
Meckelstedt	Germany	Asmus 1939
Menteşe Höyük	Turkey	Alpaslan-Roodenberg 2001 Alpaslan-Roodenberg and Maat 1999
Mierzanowice	Poland	Galasizska-Pomykol and Szewko-Szwaykowska 1967
Minsleben	Germany	Bach 1978
Molenaarsgraaf	Netherlands	Knip 1974
Montigny-Esbly Lesches	France	Anthony and Manouvrier 1907
Mugem-Gruppe	Portugal	Asmus 1973
Mugharet El-Wad	Israel/Palestine	Belfer-Cohen et al. 1991
Mulhouse-Est (Rixheim)	France	Gerhardt and Gerhardt-Pfannenstiel 1985
Müskebi, Müsgebi	Turkey	Güleç 1989
Mykene	Greece	Angel 1973b
Mýtna Nová Ves	Slovakia	Jakab 1999
Nahal Mishmar Caves	Israel/Palestine	Haas and Nathan 1973
Naumburg	Germany	Bach 1978; Schafberg 1999
Nea Nikomedeia	Greece	Angel 1973a
Neckarsulm	Germany	Knöpke and Wahl 2010
Neudorf	Germany	Gerhardt 1953
Nohra	Germany	Gerhardt 1953
Novo Filippovka	Ukraine	Konduktorova 1956
Novo-Grigoryevka	Ukraine	Konduktorova 1974
Nürnberg	Germany	Knussmann and Knussmann 1978
Öküzini	Turkey	Özbek 2000; Şenyürek 1958
Ostorf	Germany	Schuldt 1961
Padina	Serbia	Živanović 1975
Pantano de los Bermejales	Spain	Palau and Palma 1997/Ferrer Palma 1997
Räpitz	Germany	Grimm 1960
R'as al Hamra, bei Maskat	Oman	Kunter 1990
Ras Shamra, Ugarit	Syria	Charles 1962
Roßleben	Germany	Bach 1978
Rothenschirmbach	Germany	Bach 1978
Roussolakkos (Crete)	Greece	Charles 1965
Rutzing	Austria	Jungwirth and Kloiber 1973 Kloiber and Kneidinger 1970

Safadi	Israel/Palestine	Ferembach 1958
Sarata Monteoru	Romania	Maximilian and Romîne 1962
Schafstädt	Germany	Grimm 1958
Schipluiden	Netherlands	Smits and Kooijmans 2006
Schmöckwitz	Germany	Asmus 1973; Bach 1978
Seehausen	Germany	Bach 1978
Şeyh Höyük	Turkey	Güleç 1989; Senyurek and Tunakan 1951
Sijbekarspel	Netherlands	Pasveer and Uytterschaut 1992
Singen	Germany	Haidle 2014
Skateholm	Sweden	Persson and Persson 1988
Skrystrup	Denmark	Bröste et al. 1956
Sondershausen	Germany	Bach 1978
Spergau	Germany	Bach 1978
Spiennes	Belgium	Knussmann and Knussmann 1978
Stångenäs	Sweden	Asmus 1973
Staré Město	Czech Republic	Asmus 1973
Steudnitz	Germany	Gerhardt 1953
Stillfried	Austria	Szilvássy and Kritscher 1991
Stuttgart-Mühlhausen "Viesenhäuser Hof"	Germany	Burger-Heinrich Burger-Heinrich
Taforalt	Morocco	Ferembach 1965
Tápé	Hungary	Farkas and Lipták 1975
Tauberbischofsheim	Germany	Sangmeister and Gerhardt 1965; Haidle 2014
Tell Chuera	Syria	Wahl 2010
Tilkitepe	Turkey	Güleç 1989
Trebur	Germany	Jacobshagen and Kunter 1999
Unseburg	Germany	Bach and Bruchhaus 1988
Ur	Iraq	Keith 1934
Varna	Bulgaria	Marinov 1978
Vasilevka	Ukraine	Gochman 1966; Asmus 1973
Veckenstedt	Germany	Gerhardt 1953
Velim	Czech Republic	Knüsel 2007
Villánykövesd	Hungary	Zoffmann 1968
Vovnigi	Ukraine	Konduktorova 1960
Wandersleben	Germany	Gerhardt 1953
Wassenaar	Netherlands	Smits and Maat 1996
Wechmar	Germany	Sangmeister and Gerhardt 1965
Wehrstedt	Germany	Gerhardt 1953
Weingarten	Germany	Knussmann and Knussmann 1978
Wiederstedt	Germany	Meyer et al. 2004
Worms	Germany	Gerhardt 1953
Würben	Poland	Gerhardt 1953
Yortan	Turkey	Angel 1951
Zawi Chemi Shanidar	Iraq	Ferembach 1970

3.2.3 The Copper and Bronze Age body height resurgence and possible reasons

A non-parametric, spatio-temporal analysis of the data was performed, which does not rely on the usual rather arbitrary factorization of time and space by using latitude, longitude and years cal BC. More specifically, a Tensor Product P-Spline (Eilers and Marx, 1996; Wood, 2006) was used to model the smooth spatio-temporal function. Additionally, as the data is structured hierarchically (level 1: archaeological site/occupation period of a site, level 2: skeletons/individuals), a population specific effect for capturing dependencies, such as similar genetic make-ups due to kinship or local cultural and environmental factors, should be introduced in the statistical model as well. Thus, an appropriate model class would be a semiparametric mixed model (Ruppert et al., 2003). A Berkson error model (cf. Buonaccorsi 2010) was used for sex and birth date and a Bayesian semiparametric mixed model with errors-in-variables was estimated using Markov Chain Monte Carlo Methods (MCMC) (Diaconis 2009; see Groß 2016b for further details). It becomes evident that after an initial drop in overall stature at the end of the Paleolithic, there is no significant height difference between Northern Europe and Southeastern Europe during the Mesolithic and Neolithic (Fig. 3.2).

Rather, a rise in body height is first observable in NW Central Europe after the 5th

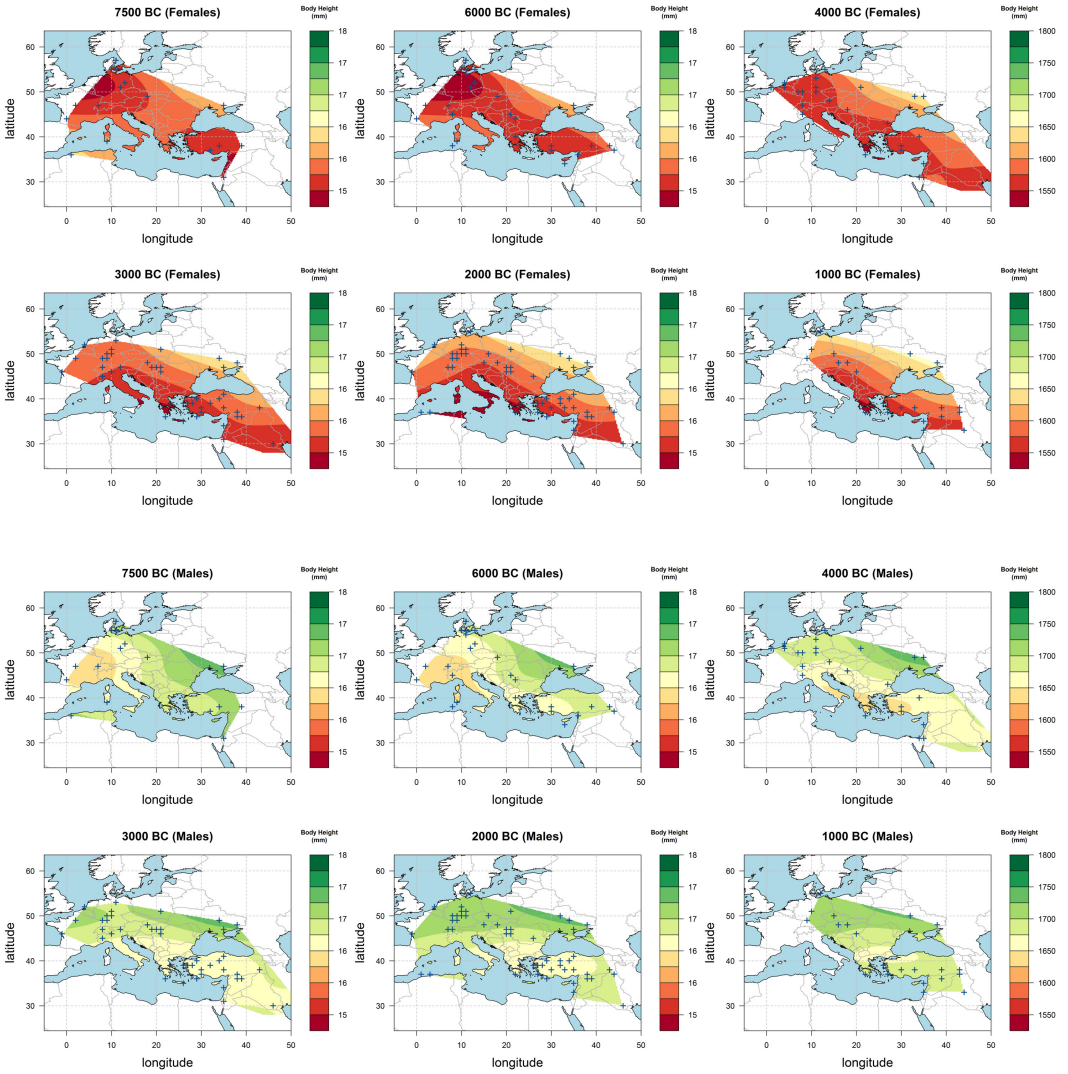

Figure 3.2: Estimated spatial mean body height in millimeters for different points in time. Upper: females, Lower: males. For a list of sites and references see Table 3.1

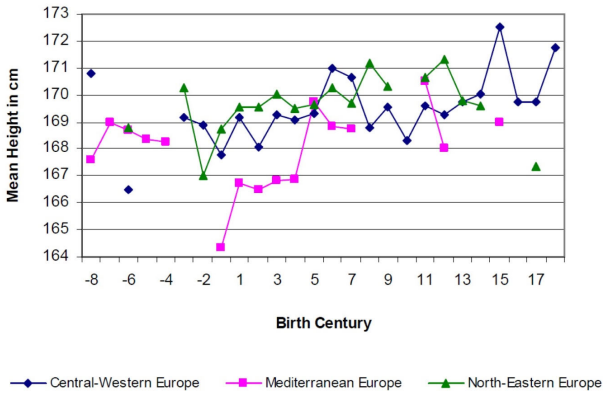

Figure 3.3: Body height in Europe over the last two to three millennia (Koepke and Baten, 2008).

millennium cal BC, while in the Near East, a slighter increase in body height establishes a NW-SE gradient of taller northerners and shorter Mediterranean and Near Eastern people from at least the 2nd millennium cal BC onwards, though it has to be kept in mind that skeletal material from Mesopotamia and its surroundings in later prehistory is either insufficient or the measurements relevant to our method were not published (d'Anna et al., 2014). Several regional studies, partially drawing on the same data (as e.g. Bennicke (1985) for modern Denmark, Jaeger et al. (1998) and (Siegmund, 2010) for Central Europe as well as Angel 1984 for Greece) confirm – despite their very rough chronological scales – the tendency towards increasing body height in later prehistory. The study by Koepke and Baten (2005) suggests that the NW-SE gradient continues throughout the 1st millennium into the historic period (Fig. 3.3; Table 3.1) and up to present times (Fig. 3.4).

In a pilot study (Rosenstock 2014), multivariate statistical analysis has shown that within the abandonment of the hunter-gatherer mode of subsistence for the Neolithic lifestyle, the adoption of plant cultivation had a significant negative effect on body height that was only slightly alleviated by the adoption of animal husbandry. These findings support, in accordance with the idea of the 'original affluent society' of the hunter-gatherers (Sahlins, 1968, 1972), that besides a heavy workload in cultivation, a shift to a diet dominated by plant products lowered the biological standard of living considerably during the Neolithic. Based on modern ethnographical data, it is assumed that prehistoric hunter-gatherer-fishers consumed between 19–35 % of proteins, whereas the diet of early farmers was mainly carbohydrate-plant-based (Cordain et al., 2000; Eaton and Konner, 1985). After this major change in diet, the Copper and Bronze Ages based their economy largely on the Neolithic package of domesticates, but there were few, yet important innovations in both production techniques and domesticates, as well as shifts in the relative importance of different foodstuffs that could have the potential to explain the improved biological standard of living in later prehistory. The effects of technical innovations are difficult to quantify, as the appearance of cattle traction at the end of the 4th millennium BC illustrates. This innovation was well-suited to

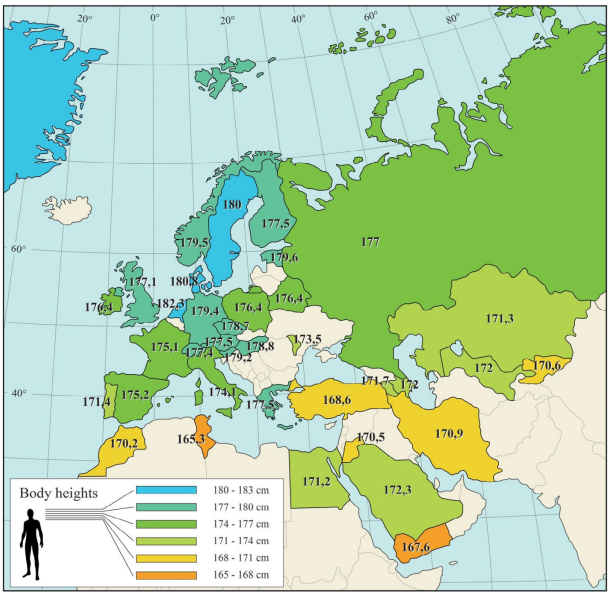

Figure 3.4: Male body height in Europe and the near East, 1970s birth cohort. For jemen und belarus the 1980s cohorts, for Greece and Poland the arithmetic mean of the 1960s and 1980s cohorts were used instead (data: Baten and Blum 2014

reduce the workload of prehistoric people, while the introduction of the ard probably further increased agricultural productivity (Kerig, 2007). Thus, cattle traction, on the one hand, appeared to be a benefit for prehistoric net nutrition. On the other hand, training traction animals and working with them requires that humans and animals live together in close proximity, a need that was probably facilitated by stabling and penning animals close to the settlements. Evidence of this increases at least in Central Europe from the 4th millennium onwards (Masson and Rosenstock, 2011). This development was accompanied by a higher risk of contracting diseases that can affect both animals and humans and which are transmitted by smear infections or via airborne pathogens, such as tuberculosis (Horwitz and Smith, 2000) – a circumstance which might have been apt to counteract positive effects. It is thus feasible for the present study to focus on protein-containing foodstuffs which can also be evaluated taking a qualitative and descriptive approach only: after all, they figure as the main factor in net nutrition (see above). This section thus complements models of subsistence, which are usually carbohydrate (e.g. Korfmann 1983) and caloriebased, ranging from 2000 kcal/person/day for Mesolithic hunter-gatherers in Denmark to 2500 kcal/person/day for Late Bronze Age Central European farmers and up to 7000 – 10,000 kcal/person/day for late-glacial/early Holocene hunter-gatherers in Poland or North America (Dennell, 1979; Jacomet and Karg, 1996).

3.3 Food as the Main Determinant for Body Height

3.3.1 Protein content of food

Two aspects are relevant for the assessment of the value of different foodstuffs in terms of protein supply: the overall proportion of protein in relation to the energy content (Table 3.2) and the quality of the protein. For the latter, several estimation methods and figures are currently used in nutritional medicine and ecotrophology: for the popular scale of the Biological Value (BV) of protein sources, eggs serve as the reference value. The Protein Digestibility Corrected Amino Acid Score (PDCAAS), however, is regarded as more reliable than the BV, but is less popular. As we are talking of rough ranges and tendencies here, both values are – as far as they are available for the foods used in prehistory – also provided in Table 3.2. Plant products oft en lack suffi cient quantities of one or more essential amino acids, which are termed limiting amino acids and are also provided in Table 3.2. As the total protein content of plant food highly depends on the growing conditions – especially on the availability of nitrogen – these figures can only give rough ranges for estimation: modern values in textbooks on nutrition derive from well-manured crops. It should be kept in mind, however, that not all essential amino acids in plant crops are increased equally by manuring, so that the bioavailability and thus the quality of the protein for human nutrition might be reduced while the overall content is increased: the relative content of lysine, the limiting amino acid in wheat protein, for instance, is lowered by nitrogen fertilizers (Crista et al., 2013). As some of the pulses attested in prehistory were marginalized by Phaseolus beans after 1492 and have thus not been a matter of scientific attention (Bermejo and León, 1994), values for some crops, such as Vicia faba and V. ervilia, are hard to find in the literature. The figures illustrate that among the cereals used in the Old World, oat ranks highest in both protein and energy content. Wheat has moderate values, while barley, millet and rye are poorer in both energy and protein. Yet, the lower protein content of barley and wheat is compensated by their relatively higher protein quality among the cereals. Most Old World pulses exhibit very similar energy levels, protein contents and biological values. As an exception to this, however, chickpeas (Cicer arietinum) are rather low in protein, although it is possible that their protein is of better quality than that of the other pulses: some researchers have argued that chickpeas have no limiting essential amino acid (Suárez López et al., 2006), while others claim that methionine and cystine are limiting factors (Iqbal et al., 2006). Before turning to animal products, a note on mushrooms seems appropriate, as, taxonomically, they are not considered as plants anymore, but as a kingdom of their own. Mushrooms contain considerable amounts of high-quality protein (Table 3.2), which, however, is trapped in chitin, and studies of the effects of different processing techniques, including grinding, chewing and cooking, are lacking so far (Kalač, 2009). Consequently, despite the fact that mushrooms in pieces commonly serve today as a meat substitute in Asian dishes due to their fleshy taste and texture, it has to be argued that mycoprotein can better be utilized by the human body if the mushrooms are heavily processed by grinding, pureeing or mashing. This is

yet another possible use of grinding stones that still awaits archaeological investigation (for an ethnographic sample see Jones 2011) before the protein benefit of edible fungi in prehistory can be assessed. The energy and protein content values of animal products given in Table 3.2 are derived from various sources, including modern high-performance breeds, so they should only serve as rough guidelines. Most animal products contain all essential amino acids in sufficient quantity for human needs, but there are some exceptions when it comes to invertebrates. The Burgundy snail (Helix pomatia), for instance, regarded as a delicacy in some regions of the Old World, lacks methionine, in this respect functioning more like pulses than meat in human nutrition (Miletic et al., 1991). All tissues that a bird fledgling or mammal cub consists of are synthesised from egg white and yolk or milk, so it comes as no surprise that, in terms of protein quality, eggs and milk rank even higher than meat, even for humans. Special attention has to be drawn to sheep's milk, as its protein content is the highest of all milked domesticates. Since ovicaprids only yield a quarter of the amount of milk of a cow (Table 3.3), the overall resulting protein output per animal and year is approximately 20 kg and thus about half of that of cattle. One cow, however, equals five goats or sheep in terms of fodder demand. Hence, the protein output calculated per livestock unit is the relevant figure: one livestock unit of sheep produces about twice as much milk protein as one livestock unit of cows. The fact that these amounts of protein can be gained for a number of years during the lifetime of an animal, in addition to the meat of the animals which can still be eaten afterwards, illustrates the importance of dairy farming for the net protein nutrition in human societies. Given that lysine is the limiting essential amino acid in all cereals, an entirely cereal-based diet would eventually lead to severe malnutrition. Cereals, however, contain enough methionine, cystine and tryptophane, the limiting amino acids in pulses, so combining cereals and pulses in a meal or at least over the course of a day (Young and Pellett, 1994) efficiently enhances the nutritious quality of a diet. Maize (*Zea mais*) tortillas eaten with a stew made from common beans (*Phaseolus sp.*) in roughly a 50:50 proportion is the classical New World example (Duranti and Gius, 1997) and reach a BV of 99 instead of only ca. 72 and 73 respectively for the separate ingredients (Vaupel and Biesalski, 2010). Similar traditional vegan Old World meals like wheat bread combined with hummus paste made from chickpeas (*Cicer arietinum*) or ful made from broad beans (*Vicia faba*), for which protein quality figures are not published, can thus be understood as an effective means of cultural adaptation to the amino acid problem in human diet if meat is scarce. Moreover, tahini is currently often mixed into hummus: it is made from the seeds of sesame (*Sesamum indicum*), an oil seed. Besides adding fat to the dish, it complements hummus with a high content of methionine in its protein fraction (Faris and Takruri, 2003). This might be true of other oil seeds as well (Sosulski and Sarwar, 1973): while the limiting amino acid in the protein fraction of flax seed (*Linum usitatissimum*) is lysine, and can thus be complemented by pulses, it is lysine and methionine in poppy seeds (*Papaver somniferum*). Hence, it is not surprising that both sesame and poppy seeds are the most common toppings on small breads like the German *Mohnbrötchen*

and Turkish *pide*, as the seeds or oils pressed from them complement cereal and pulse dishes in terms of amino acids. Even adding very small amounts of animal products to a plant-based diet enhances the diet's amino acid profile considerably (Young and Pellett, 1994) and is a common strategy if we look at classic dishes like the New World's chili con carne made from common beans (*Phaseolus sp.*) and meat, as well as most pulsebased stews in the Old World, such as lentils, which are for this reason traditionally eaten with sausages in Central Europe. Moreover, Spätzle-type pasta complements this dish in Southwestern Germany, illustrating that eggs and wheat mixtures add up to a BV of 118 (Vaupel and Biesalski, 2010). Pasta, but also *pancakes* and *pastries* made from egg and wheat are thus much more nutritious than egg with a BV of 100 and wheat with a BV of ca. 70 (see Table 3.2) eaten separately. The next best combination that can be found in the literature pertaining to the nutritional value of Old World foodstuffs is the combination of milk and egg with ca. 119, but the very best is milk and wheat: porridge made from ca. 3/4 milk and 1/4 wheat has a BV of 125 (Vaupel and Biesalski, 2010) instead of a BV of 85 and ca. 70 in the separate ingredients. Milk-wheat dishes are only topped by recipes like the Spanish *tortilla* and the German *Bauernfrühstück*: these post-1492 innovations, combining eggs with potatoes, if prepared in a proportion of ca. 1/3 egg and 2/3 potatoes, reach a BV of 136. They are therefore the recommended diet for persons with kidney problems who should eat a minimum amount of protein but still need to get all essential amino acids (Vaupel and Biesalski, 2010). It still remains to be investigated whether the introduction of potatoes into the Early Modern European societies led to different impacts in terms of net nutrition and protein supply, according to the amount and frequency that eggs were eaten in the respective regions and social classes. But we have to leave this question to economic history dealing with later periods. Nonetheless, the above example illustrates that, although a plant does not necessarily have to contain a lot of protein (the potato has only 2 % of protein), it can still contribute considerably to the overall protein supply if it contains essential amino acids apt to supplement the existing foodstuffs.

Table 3.2: Energy, protein content, protein bioavailability – expressed as biological value (BV) and Protein Digestibility Corrected Amino Acid Score (PDCAAS) – and limiting amino acids of selected cereals (wholemeal), pulses (podded, dry), oil seeds, mushrooms (dried), shellfish, snails, crustaceans and meats (carcass) and animal products (Sources: United States Department of Agriculture 2012 (1); BMVEL/MRI BMVEL/MRI (2); Schuster et al. 2000 (3); Abdullah et al. 2010 (4); Schaafsma 2000 (5); Suárez López et al. 2006 (6); Food and of the United Nations (FAO) 1995 (7); Iqbal et al. 2006 (8); Jaworska and Bernaś 2013 (9); Food and of the United Nations (FAO) 2012 (10); Lentner and Diem 1973 (11); Kallweit et al. 1988 (12); Horwitz and Rosen 2005 (13), Babiker et al. 1990 (14); Miletic et al. 1991 (15); Usydus et al. 2009 (16); Williams 2007 (17); Vaupel and Biesalski 2010 (18); Eklund and Ågren 1975 (19); Sosulski and Sarwar 1973 (20); Razavi et al. 2008 (21)).

	Energy (kcal/100g)	Protein (g/100g)	BV	PDCAAS	Limiting AA (if age-specific, for children)
Cereals					
Oat (Avena sativa)	350 – 390 (1,2)	12 – 17 (1,2)	60 (2)	56 (6)	Lysine (6)
Barley (Hordeum vulgare)	320 – 345 (1,2)	10 – 11 (1,2)	74 (2)	66 (6)	Lysine (6)
Millet (Panicum miliaceum)	350 – 370 (1,2)	10 – 11 (1,2)	47 (2)	n.a.	Lysine (7)
Rye (Secale cereale)	340 (1)	9 – 10 (1,2)	67 (2)	65 (6)	Lysine (6)
Wheat (Triticum sp.)	310 – 342 (1,2)	11 – 14 (1)	59 – 77 (2,18)	42 – 48 (5,6)	Lysine (6)
Pulses					
Chickpea (Cicer arietinum)	325 – 360 (1,2)	18 – 24 (1,2,8)	45 (2)	78 (6)	None (6), Methionine and Cystine (8)
Lentil (Lens culinaria)	310 – 340 (1,2)	24 – 26 (1,2,8)	33 – 45 (2,3)	63 (6)	Methionine and Cystine (6), Tryptophane (8)
Pea (Pisum sativum)	290 – 340 (1,2)	24 – 25 (1,2,8)	50 – 60 (2,3)	43 (6)	Methionine and Cystine (6), Tryptophane (8)
Bean (Vicia faba)	230 – 340 (1)	26 (1,2)	32 (2)	51 (6)	Methionine and Cystine (6), Tryptophane (8)
Bitter vetch (Vicia ervilia)	360 (4)	22 – 29 (4)	n.a.	n.a.	n.a.
Oil seeds and fruits					
Sesame (Sesamum indicum)	590 (1)	18 (1)	n.a.	n.a.	n.a.
Flax (Linum usitatissimum)	534 (1)	18 (1)	n.a.	n.a.	Lysine (20)
Olive (Olea europaea)	81 – 145 (1)	1 (1)	n.a.	n.a.	
Poppy (Papaver somniferum)	525 (1)	18 (1)	58 (19)	n.a.	Lysine, Methionine (19)
lallemantia (lallemantia sp.)	n.a.	26 (21)	n.a.	n.a.	n.a.
Mushrooms					
Agaric (Agaricus bisporus)	n.a.	30 (9)	n.a.	n.a.	None (9)
Yellow boletus (Boletus edulis)	n.a.	23 (9)	n.a.	n.a.	Lysine, Leucine (9)
Shellfish, crustaceans, poultry, meat					
Mediterranean mussel (Mytilus galloprovincialis)	86 (1)	10 – 12 (1,15)	81 (15)	94 (6)	None (6), Isoleucine (15)
Burgundy snail (Helix pomatia)	90 (1)	13 – 16 (1,15)	68 (15)	n.a.	Methionine (15)
Crayfish (Astacidae, Cambaridae)	77 (1)	16 (1)	n.a.	94 (6)	None (6)
Lobster (Homarus sp.)	77 (1)	17 (1)	n.a.	94 (6)	None (6)
Brown trout (Salmo trutta)	148 (1)	21 (1)	n.a.	94 (6)	None (6)
Atlantic herring (Clupea harengus)	158 (1)	18 (1)	n.a.	94 (6)	None (6)
Chicken (Gallus gallus), flesh	105 – 116 (1,10)	20 – 23 (1,10)	n.a.	94 (6)	n.a.
Chicken (Gallus gallus), egg	38 (10)	12 (10)	n.a.	97 – 118 (5,6)	None (6)
Goat (Capra hircus), flesh	109 (1)	21 (1,14)	n.a.	n.a.	n.a.
Goat (Capra hircus), milk	63 – 71 (11,12,13)	2.6 – 3.6 (11,12,13)	n.a.	n.a.	n.a.
Sheep (Ovis ammon), flesh	282 – 514 (1)	17 – 22 (1,7,14,17)	n.a.	94 (6)	None (6)
Sheep (Ovis ammon), intestines	n.a.	17 (17)	n.a.	94 (6)	None (6)
Sheep (Ovis ammon), milk	88 – 110 (1,11,12,13)	3.9 – 5.9 (1,11,12,13)	n.a.	n.a.	n.a.
Cattle (Bos taurus), flesh	323 – 498 (10,17)	17 – 23 (10,17)	87 (18)	92 – 94 (5,6)	None (6)
Cattle (Bos taurus), intestines	n.a.	17 (17)	n.a.	94 (6)	None (6)
Cattle (Bos taurus), milk	64 – 85 (11,12,13)	2.6 – 3.8 (11,12,13)	85 (18)	95 – 121 (5,6)	None (6)
Pig (Sus scrofa), flesh	472 (10)	11 (10)	n.a.	n.a.	None (6)
Horse (Equus caballus), milk	44 – 53 (11,12,13)	1.3 – 2.5 (11,12,13)	n.a.	n.a.	n.a.

3.3.2 Other food-related factors

Before the Neolithic amino acid spectra can be outlined and compared to the Copper Age and the Bronze Age spectra, however, a few other constituents of food as well as certain metabolic issues in humans should be discussed. The ability of the human organism to split starch into sugars, for instance, is not uniform in all current societies. Societies with a long tradition of a high starch intake in their diet, such as cereal-based agriculturalists like the American Europeans, the Japanese, or the African Hadza, who collect high proportions of starch rich tubers, have more copies of the AMY1 gene coding salivary alpha-amylase than societies with diets poor in starch, such as African rain-forest hunters of the Biaka and Mbuti, pastoralists like the African Datog, or pastoralist-fishermen of the Siberian Yakut. The former have a higher ability to split starch in the mouth, which is to be interpreted as an evolutionary adaptation to diet (Perry et al., 2007). Favism or G6PD deficiency, i.e. the lack of the ability to synthesize the enzyme glucose-6-phosphate dehydrogenase, is another example. Via changes in the sugar metabolism, this condition can lead to haemolytic processes if the organism is under oxidative stress. As the condition is most frequent among populations in regions where malaria is endemic, it is probably an evolutionary adaptation to this disease, similar to sickle cell anaemia (Mehta et al., 2000). The most important trigger of oxidative stress is the intake of pulses, especially broad or fava beans (Vicia faba) with their high content of the oxidative alkaloid glycosides vicine and covicine. Even though the term for the condition is derived from the fava bean, it has to be noted that most pulses contain toxins and alkaloids, so that their consumption requires – in comparison to cereals – extensive pre-treatment, such as soaking in water and cooking for long hours. People with favism should avoid the consumption of pulses and especially fava beans altogether, although interestingly they are a common type of food in those regions where G6PD deficiency is common. Although favism is the most frequent enzyme variation in the modern world population, lactose maldigestion has received much more attention so far. While in most people in the world the enzyme lactase ceases to be synthetized in later childhood, i.e. after weaning when the organism is no longer dependent on breast milk and thus lactose, certain populations in eastern and western Africa and especially in Central and Northern Europe are characterized by a high incidence of life-long persistence of lactase production, a trait thought to be genetically determined as an evolutionary adaptation to the consumption of animal milk as a foodstuff (Ingram et al., 2009). Lactose tolerant people can digest milk in unprocessed, i.e. unfermented form, whereas lactose intolerant people can to some extent only consume milk products like yoğurt, sour milk and their derivatives, such as cheese or kefir and koumiss, in which most of the original lactose has been split into lactate or alcohol due to the action of lactic acid bacteria or yeasts or both (Hertzler and Clancy, 2003). Despite the fact that the amino acid profile of the milk remains unaltered by these fermentation processes, and while the loss of a few grams of carbohydrates per kg milk appears negligible when dealing with protein supply, consuming fermented versus unfermented milk might still make a difference when it comes to hormone-like substances contained in milk such as

the insulin-like growth factor I (IGF-I). IGF-I is the most powerful growth factor and is regulated by the growth hormone (GH), but can also stimulate growth independently (Laron, 2001; Bernstein, 2010). IGF-I controls cell division and cell proliferation, plays – differing from IFG-II which mainly regulates fetal growth – an important role in postnatal and especially pubertal growth Bernstein (2010) and has a positive effect on growth rate and skeleton length according to most (e.g. Liu and LeRoith 1999; White et al. 1999; Laron 1993; Laron et al. 1980; Laron 2001; Ong et al. 2002 but not all studies (e.g. Beckett et al. 1998). IGF-I is found in the milk of all mammals, albeit in different concentrations (Donovan and Odle, 1994), and is particularly highly concentrated in colostrum (Klagsburn, 1978; Bernstein, 2010; Donovan and Odle, 1994). Elevated levels of IGF-I have been reported after the consumption of cow's milk, although it is not clear whether IGF-I is transmitted from the milk via the gastrointestinal tract into the human bloodstream, or whether milk consumption stimulates the production of human IGF-I (Wiley, 2009). The fact that the IGF-I concentration in humans is lowest in postnatal infants (Hammond, 2007), although the neonate period experiences the most intensive growth Bogin and Smith (1996), affirms the significance of growth hormone intake at least via the colostrum. Acid production in the gastrointestinal tract of neonates is still low (Speer and Gahr, 2009) and might thus prevent the denaturization of the IGF-I so it can unfold its full potential in the body of the growing consumer, whereas IGF-I is, like all other proteins, believed to be denatured due to the stomach acid and digestive enzymes in older individuals (Hammond, 2007): after all, in medical treatment of growth disorders, protein hormones are administered by injections (Laron and Kopchick, 2011; Rosenbloom, 2008). According to Gardner (1988), however, intact proteins can cross the gastrointestinal tract in adults, too, probably in interplay with binding proteins (Bernstein, 2010; Ottesen et al., 2001; Laron, 2001; Loui et al., 2004; Rechler and Clemmons, 1998; Blum and Baumrucker, 2008). 80–90 % of these proteins are bound to a ternary complex with an IGF-bounding protein (IGFBP 3) and an acid labile subunit (ALS) (Ottesen et al., 2001; Laron, 2001; Klagsburn, 1978). Nevertheless, it can be argued that compared to the IGF-I pool in the human body the IGF-I amount in cow's milk is low. Thus, it is assumed that even if all IGFs from consumed cow's milk would reach the blood stream undigested, they would not exert any noteworthy influence (Hammond, 2007). Experiments on rats, however, have shown that casein, which is the most common protein in milk, if administered orally together with IGF-I, can significantly increase the bioavailability of IGF-I (Kimura et al., 1997). Furthermore, mammalian milk also contains 'truncated' IGFs in small quantities, which increase their potency up to tenfold (RoPHA, 1999). Despite these open questions, we can conclude that the current state of research does not preclude the possibility that IGF-I ingested with animal milk can at least in some quantities pass the intestinal tract and enter the human organism or alternatively trigger its own IGF-I production without passing the intestinal tract and eventually affect the postnatal longitudinal growth of the milk consumer: a positive effect of milk consumption on body height that is not related to the protein content, but probably mediated via IGF-I

with or without calcium, casein or another milk constituent, has been demonstrated in prospective studies (Wiley, 2009). While IGF-I in bovine milk was not destroyed when pasteurized at 79° C for 45 seconds (Collier et al., 1991), fermentation processes seem to lower the amount of IGF in milk products significantly (Kang et al., 2006). This is in accordance with findings that fermented dairy product consumption was not positively associated with height (Wiley, 2009). It is therefore arguable that the body height of individuals with lactase persistence who consume animal milk during childhood and youth do not only benefit from the milk proteins as they would with fermented milk products but also from a higher level of IGF-I additionally triggering growth if all other factors relevant for growth like other hormones, minerals and nutrients are available. Such a mechanism might lie behind the correlation found between lactase persistence and body mass index (Kettunen et al., 2010), and if this holds true, the concept of body height as a proxy for net nutrition would be biased whenever lactose-intolerant individuals are compared with lactose-tolerant individuals who consumed unfermented milk during their growth period.

Table 3.3: Mean milk and protein yield per year for several domesticates, relating to one animal and one livestock unit, respectively (after Horwitz and Rosen 2005; Ruhr-Stickstoff 1988 data from modern high-output production were omitted).

Species	Milk yield kg/a (animal)	Protein yield in kg/a (animal)	Protein yield in kg/a (animal)
Goat	200 – 900	17	80
Sheep	200 – 900	26	130
Cattle	1000 – 2500	56	56
Horse	390 – 750	7	10
Camel	800 – 3600	60	60

3.3.3 The Neolithic versus the Copper and the Bronze Age food spectrum

The low trophic levels inferred from stable isotope studies of human skeletal material dating to the Near Eastern PPN B and C (ca. 9000–7000 cal BC) (Pearson et al., 2013) suggest that, despite the initial domestication of animals during that time period, their contribution to the diet was probably low. Cereals, already used in the Epipaleolithic, were cultivated from the PPN A onwards (Weiss et al., 2006), and it would be an interesting task to trace the AMY1 gene's (see above) history and its interplay between the Near Eastern Paleolithic, Epipaleolithic and Neolithic populations and their wild and then domestic cereal consumption as well as the gene's role in the Secondary Neolithization of Europe. It is not unlikely that early cereals, like wild einkorn (Triticum urartu) and emmer (T. dicoccum) as well as domestic einkorn (T. monococcum) and emmer (T. turgidum ssp. Dicoccum), did not contain enough gluten, a protein of low nutritional value but good binding qualities to allow for the baking of fluffy leavened bread, probably leaving early Neolithic eaters with gruel, flatbread and maybe beer (Braidwood et al., 1953). Closed ovens apt for baking bread are rarely found in the

archaeological record of the 9th and 8th millennia, with only one specimen found, for example, in Abu Hureyra (Moore et al., 2000). In contrast, they are found as a constant feature of Neolithic houses from the 7th millennium onwards as documented for Çatalhöyük East (Hodder, 2005), Karanovo (Hiller and Nikolov, 1997) and Djeitun (Müller-Karpe, 1984). Though one could argue that until the 7th millennium ovens were installations located outside the settlements and thus not detected by archaeologists, the coincidence seems striking that the first spelt and bread wheats containing the DD allele responsible for good baking qualities appear in the 7th millennium (Jacomet and Karg, 1996; Nesbitt, 2001). The DD- allele was inbred into wheats from the goatgrass Aegilops tauschii (also named A. squarrosa) and codes types of gluten that lead to coeliac disease, another genetic trait with an evolutionary history worth examining in the context of archaeology and archaeobotany (Molberg et al., 2005). Moreover, gluten contains exorphins, i.e. opiod-like substances, targeting the activity of grazing animals for the benefit of the surviving wheat plants, which can exert similar calming effects on humans, too: the borderline between drug and food has always been a matter of quantity (Zioudrou et al., 1979; Fukudome and Yoshikawa, 1992). Pulses, which were also already used in the Near East and the Mediterranean region in the Epipaleolithic, were first cultivated during the PPN B or even earlier (Weiss and Zohary, 2011), so wheat (flat-)bread with hummus as a dish could indeed be as old as the PPN B or older. The protein from pulses might have complemented the protein from the cereals, thus securing amino acid supply despite animal protein scarcity in the PPN B (Lösch et al., 2006). We could thus think of the few hunted and domesticated animals in the Early Primary Neolithic of the Near East as the providers of that little bit of meat that pushed the BV and PDCAAS of the cereal-pulse combination to ca. 100. Among the pulses in the PPN B and through the Neolithic and Chalcolithic of the Near East, Vicia faba appears to be present but usually in very small quantities (Jacomet and Karg, 1996). This may be due to the fact that broad beans are more fragile and more prone to disintegration in the archaeological record than small pulses such as lentils, bitter vetch and peas, but chickpeas are also quite large and more frequent (Tanno and Willcox, 2006). It could be that the high tryptophane content of chickpeas, an essential amino acid that positively affects the serotonine metabolisms and thus has a filling and calming effect on humans (Kerem et al., 2007), is the reason for the preference of chickpeas. But it is also not completely absurd to speculate on whether the infrequency of Vicia faba in the archaeological record of the Neolithic and Chalcolithic is a reaction to a long-term genetic history of favism in the Near East (Jacomet and Karg, 1996). Near Eastern lowlands are today infested by malaria and this situation could date back to at least the beginning of the Holocene, ca. 10,000 to 9000 cal BC, given the current state of research into the paleogenetics of the malaria agents Plasmodium falciparum and P. vivax (Schlagenhauf, 2004; Morgan-Forster, 2010). This is also supported by the occasionally high rate of porotic diseases in the Near East and the Mediterranean in the Holocene (Iezzi, 2009; Rathbun, 1984), which is often thought to be related to anemia – caused directly by favism and sickle cell disease – and/or malaria (Angel, 1970;

Wapler et al., 2004). Given the scarcity of Vicia faba cultivation, paleopathological evidence for anemia and the presence of malaria, it is not impossible that Near Eastern Neolithic people were often G6PD deficient and had to avoid the consumption of large quantities of pulses and especially broad beans. Although we have no clue as to what extent malaria posed a problem in temperate Europe during the Holocene climatic optimum, it does not seem unlikely that favism belonged to the genetic endowment of the founder populations bringing the Secondary Neolithization to Europe, thus causing them to consume limited amounts of pulses, too. Consequently, it might be worthwhile for future research to also take a closer look at the paleogenetics of G6PD deficiency. However, the probable limited consumption of pulses in comparison to cereals might not only be due to favism, but might also be the result of the low crop yield: wild pulse plants grow very patchy and the seeds do not ripen at the same time (Abbo et al., 2009). Furthermore, initial largescale consumption of pulses was also probably hindered due to even more alkaloids contained in the plants, especially at the beginning of the domestication process, or due to the necessary intensive pre-treatment (Valamoti et al., 2011): enhanced digestibility through the reduction of anti-nutritive substances was one effect of the domestication process probably only reached in later prehistory (Hopf and Zohary, 2001). The low PPN B trophic levels might also indicate that animal milk had not yet entered the diet to a significant amount, thus making it likely that the first milk fat remains detected in pottery (e.g. Evershed et al. 2008) from the second half of the 7th millennium cal BC onwards do indeed mark the time of initial animal milk use and not the invention of ceramic containers that also make their appearance in the 7th millennium cal BC. A general decline of the importance of pulses in the diet of the Near East during the later phases of the Neolithic (Miller, Miller) – maybe in part driven by the pathogen Didymella rabiei infesting chickpeas, a problem which was probably solved by the shift from summer to winter cropping by the Bronze Age (Abbo et al., 2003) – might therefore have been compensated by the possibility to eat cereal-cum-milk porridges. Such mixtures do not only have an even better BV than the cereal-pulse combination (Lösch et al., 2006) but they are additionally very rich in opioid-like substances, as they do not only contain gluten exorphins (see above), but also caso-morphin occurring in milk and milk products meant to sedate the suckling young animal for the benefit of the mother animal, also functioning in humans (Zioudrou et al., 1979; Sienkiewicz-Szłapka et al., 2009). This calming effect is still used today when milk-cereal porridge is recommended for children as an evening meal leading to better sleep. Until animal milk was available, the milk-cereal combination rich in amino acids was at best only accessible to still breastfed infants if cereal meals were used as complementary dishes during the weaning years. While evidence for a pastoral and milkor blood-based rather than a soil-tilling lifestyle exists in Europe from the Copper Age onwards, as, e.g., in the Beaker populations (e.g. Menninger 2008; Kolář et al. 2012), we note a higher visibility of dairying in the archaeological record in the Near East, e.g., the specialized churning vessels from the 4th millennium cal BC in Anatolia (Sauter et al. 2003; for the date Schoop 2009;

Figure 3.5: Milking frieze from el-Obed, 3rd millennium cal BC (after Orthmann 1975

Schoop et al. 2009) or the Early Dynastic milking frieze from el-Obeid showing dairy procedures (Gouin, 1993). It has to be noted that in the Neolithic of the Near East and Europe and with some probability also in the later prehistory of the Near East, milk was most likely consumed only as fermented milk products after childhood. As hunter-gatherers have no benefit from the ability to digest lactose beyond childhood, and as the ability is relatively rare in the Near East today while it is common in Northern Europe (Itan et al., 2009, 2010), the responsible genetic change most likely took place in Europe sometime during the Post-Palaeolithic. Alleles responsible for the ability to digest lactose could not be detected in several skeletons sampled from the Linear Pottery Culture (LBK) of the second half of the 6th millennium cal BC (Burger et al. 2007) despite a highly positive selection once the mutation has occurred, so the simulation by Itan et al. (2009) locating the allele in the preceding Starčevo culture of northern Southeast Europe seems unconvincing, leaving us with the LBK as a terminus post quem and the 1st millennium AD, in which the allele has been detected in Europe (Krüttli et al., 2014), as a terminus ante quem. In the 4th millennium cal BC, arboriculture of olives as an oil fruit started in the Levant and spread throughout the Mediterranean basin during the Bronze Age (Kaniewski et al., 2012). Although their protein has a balanced amino acid spectrum (Manoukas et al., 1973), the total protein content of only 1 g per 100 g renders their domestication rather unimportant in terms of amino acid supply, in contrast to poppy (Papaver somniferum), a component in the diet of the LBK in Central Europe and currently thought to be one of the rare examples of domesticates with Western European origin (Coward et al., 2008; Zohary et al., 2012). Wild poppy seeds, however, have been recently reported from a Levantine PPN C context, and after all, modern poppy production is centred in the Near East and Central Asia. Thus, taphonomy and archaeological sampling strategies might have biased its detection in earlier contexts in the Southeast of the Old World so far (Hnila, 2002). Thus, we cannot decide at this stage whether poppy indeed belongs to the spread of new foodstuffs around the Old World that starts in the Late Copper and the Early Bronze Age around ca. 3000 cal BC or had already been part of the Neolithic package. Poppy is found in Southeastern Europe, the Mediterranean, Egypt and Anatolia from the Early Bronze Age onwards, but interestingly not in Mesopotamia and other parts of the Near East (Jones and Valamoti, 2005; Hnila, 2002). Here, however, sesame (Sesamum indicum), a native from India, was first cultivated from ca. 3000 cal BC onwards (Bedigian and Harlan, 1986). Hummus with tahini can thus be viewed as

a typical Bronze Age innovation in the protein supply of the Near East. Lallemantia (Lallemantia iberica, L. canescens, L. peltata and L. roylenana) is another introduction into the food economy in the Early Bronze Age Mediterranean, probably stemming from the Near East (Jones and Valamoti, 2005), but nutritional data on the oil plant is hardly available because it is today at best used for the production of animal fodder or technical oils. New cereals such as millet (Panicum miliaceum) spreading from Eastern Asia into the Near East and Europe in the Bronze Age (Lightfoot et al., 2013; Spengler et al., 2014; Tafuri et al., 2009), and oat (Avena sativa) (Hubbard, 1980) were not major innovations in terms of protein content and quality, and probably mainly led to greater diversity and thus greater economic stability during bad years. The same appears to be true for rye (Secale cereale) (Behre, 1992; Abbo et al., 2013), already cultivated in the Neolithic of the Near East and either adopted or newly domesticated in the Late Bronze Age of Europe (Abbo et al., 2013; Weiss et al., 2006). Pulses, however, as a stable yet usually small component of the diet during the Neolithic and the Copper Age experience a general decline in the Bronze Age compared to the Neolithic in many settlements in Europe (Hopf and Zohary, 2001), though in some regions they gain considerable importance in the Late Bronze Age (LBA), i.e. in the circumalpine regions in the middle of the 2nd millennium cal BC. Especially broad beans (Vicia faba) are found there in large quantities in the LBA and most likely constituted a major component of the diet (Jacomet and Karg, 1996). It is worth noting in this context that in all three houses of the LBA settlement of Zug-Sumpf the storage finds of cereals and beans (Vicia faba) were counted in a proportion of approximately 2:1 (Jacomet and Karg, 1996), which cannot be explained by cultivation, harvesting or processing techniques, but equals the optimal proportion between cereals and pulses in terms of amino acid complementation (see above). If this is not just chance, but rather if such a proportion could be found regularly in storage finds, it would strengthen our assumption that not only contemporary traditional cuisine but also prehistoric people knew how a nutritious dish is prepared. Moreover, the regular consumption in the order of a hundredweight of broad beans per person and year calculated for LBA Zug-Sumpf (Jacomet and Karg, 1996) requires that LBA Central European people were not severely deficient in G6PD. Such an assumption could well be in line with evidence for increased genetic diversity in Central Europe from the 4th millennium onwards, when a decrease of the genetic influence of the genetic makeup of the Early Neolithic founder population goes with the reappearance of the hunter-gatherer component and the emergence of what has been termed the Late Neolithic-Early Bronze Age component (Brandt et al., 2013). Additionally, the latter group would be an interesting starting point to look for the origins of lactase persistence in Europe, which so far could not be dated (see above). If paleogenetic research could reveal that lactase persistence appeared during or shortly after the 4th millennium BC, mean body height data from the respective regions would have to be regarded as the result of both net nutrition and additional IGF-I intake.

3.4 Discussion: Bringing Food and Body Height Changes Together

The efficiency of the PPN B diet based on cereals and pulses and complemented with a bit of meat is corroborated, e.g., by the increased body heights in the PPN B of Jericho after a PPN A low (Röhrer-Ertl, 1978). Despite the additional introduction of animal milk products in the 7th millennium cal BC at the latest, mean body height, however, was not brought back to Paleolithic levels, although the amino acid package completed in the wake of the 'Second Neolithic Revolution' (Düring, 2010) enabled cultivators and livestock keepers to efficiently migrate into and settle temperate Europe. As a possible explanation, we could speculate that animal milk – probably in combination with cereals – did not complement, but at least partially replaced breast milk, an idea well in line with the population growth to be inferred for the Neolithic, as earlier weaning might have led to shorter intervals between births by means of a shortened lactational amenorrhea and thus more births (Armelagos et al., 1991; Sellen and Smay, 2001). Another possible reason we could think of is lowered per capita protein supply during Neolithic population growth (e.g. for the early to middle LBK growth cf. Zimmermann et al. 2009): if more mouths have to be fed, there is less for the individual, a similar mechanism to what has been termed the Malthusian check in economic history. It can be argued that the recovery of body heights starting in the Copper Age of both the Near East and Europe is merely the result of the in-migration of genetically taller people, if one wants to connect the genetic Late Neolithic- Early Bronze Age component with evidence about migrations during the Beaker cultures (Price et al., 2004). As the only region where body heights have been above the mean Neolithic values during the Early Holocene is Eastern Europe, a likely homeland for some of the Central European cultures of the 4th millennium cal BC from an archaeological point of view (Woidich, 2002), this explanation cannot completely be ruled out with regard to body height. However, we can equally think of the emergence of more pastoral ways of life based on milk and milk products during later prehistory as a possible reason. If that holds true, the general body height increase from the Copper Age onwards would be the result of the consumption of milk and milk products as an additional protein source, while the even more pronounced body height gain in Europe would additionally be due to a metabolic effect with the IGF-I contained in milk. By that time, the food spectrum of both Europe and the Near East had widened considerably in the course of the 3rd and 2nd millennia. Not only do the Near East and the Mediterranean now cultivate exotic new foodstuffs, such as sesame and lallemantia, which are apt to complement the amino acid supply in this region, but Europe and the Mediterranean region also base their diets on a wider range of cereals by the cultivation of millet and oat. Moreover, in some regions of Europe considerable quantities of pulse protein are added to the diet by the Late Bronze Age. It can, therefore, be argued that the sheer widening of the food spectrum in the Bronze Age may have led to an improved net nutrition and therefore a higher mean body height. It has to be kept in mind,

though, that the emergence of a deeper social differentiation in Later Prehistory might bias the skeletal evidence towards higher-ranked individuals who might have received inhumations that were more likely to enter the archaeological record. Higher rank as approximated in the grave goods has been associated with both taller body height (e.g. Teschler-Nicola 1986; Wason 2004) and higher trophic levels (Triantaphyllou et al., 2008). As an outlook, both zooarchaeological and archaeobotanical data have to be operationalized as variables that are as comparable through time and space as are body heights. Subsequently, multivariate statistical analysis (as in Rosenstock 2014) will be necessary to test whether wider food spectra, on the one hand, and the co-occurrence of certain amino acids on the other, are indeed correlated with taller mean height. Moreover, statistics has the potential to model the possible effects of metabolic factors, like AMY1 gene frequency and lactase persistence, in order to simulate the growth outcome of nutrition intake. Body heights can only be conclusively used as the 'rating agency' assessing the outcome of Neolithic vs. Copper and Bronze Age subsistence strategies if such analyses have been applied in advance.

Acknowledgements

The project 'LiVES' is funded by an Emmy-Noether-Grant of the German Research Foundation (DFG). We thank Stefanie Jacomet for sharing her knowledge on pulses as well as taphonomy and sampling strategies, Helmut Kroll for bringing forgotten oilseeds to our attention and Sandra Pichler for making us aware of the human genetic component, as well as Volker Heyd for important literature hints and Gerhard Hotz for critical comments on earlier versions of this work. We owe special thanks to Jörg Baten for giving us access to his dataset on modern body heights in Europe and the Near East before publication.

Kapitel 4

Modeling Body Height in Prehistory Using a Spatio-Temporal Bayesian Errors-in-Variables Model

4.1 Introduction

The LiVES project ('Lebensbedingungen und biologischer Lebensstandard in der Vorgeschichte Europas und Südwestasiens'; Rosenstock 2014) examines the biological standard of living and living conditions during prehistory, i.e. the period from the appearance of first modern humans in the Upper Paleolithic ca. 50000 to 40000 BC until the end of the Bronze Age ca. 1000 BC, in Southwest Asia and Europe. The measure used for approximating the biological standard of living is the body height, a key figure used in economics since the late 1980s (Komlos, 1989) as an alternative measure for welfare for times and regions where other economic indicators such as the gross domestic product per capita are unreliable or not available. Body height reflects the nutritional situation during childhood and youth, i.e. the human growth period. The genetically determined body height (Weedon and Frayling, 2008) can only be reached by optimal nutrition of the individual during childhood (Silventoinen, 2003). Subsistence was subject to several changes in prehistory, among which the transition from the mobile hunter-gathering way of life in the Paleolithic to a sedentary life based on plant cultivation and animal husbandry in the Neolithic starting in the Near East about 10000 BC and reaching Northern Europe by about 3000 BC was probably the most radical. But also Copper and Bronze Age innovations such as the cattle-drawn around 4000 BC might have had an impact. With the aim to connect environmental and cultural data as potential explaining variables with the body height variation, one of the central research questions of the project is: How did the body height, or respectively living standard, develop spatio-temporally in prehistory? The aim of this work is to develop a statistical model to answer this question.

To the author's best knowledge this is the first attempt to model long term trends in prehistoric living standards with respect to geographic differences. There are extensive reviews for the ancient and medieval period (Koepke and Baten, 2005, 2008) as well as modern history (e.g. Floud et al. 2011; Komlos 1989, 1995; Steckel 1995), while for prehistory a picture beyond regional level is missing. The Global History of Health Projects (Steckel, 2003) is aiming at a global picture, but has not yet come up with analysis. Larsen (1995) gives a review and states that the shift from foraging to farming led to a reduction in health status, increased physiological stress and led to a decline in nutrition and living standard. This was done using various skeletal and dental pathological condition measures whereas stature estimates showed no general trend. This finding is not surprising, as Larsen's paper does neither differentiate between different Neolithizations in the world, such as that of the Old World with a broad range of well suited plants and animals vs. the New world with only a limited spectrum of species, nor between the different modes of Neolithization in one Neolithization process, i.e. Primary Neolithization with local new domestication and Secondary or even Tertiary Neolithization by import of domesticates. The Near Eastern to European Neolithization trajectory (Schier, 2009), after all, displays a marked decrease in body height compared to the Paleolithic (Rosenstock, 2014). A recent publication of Mummert et al. (2011) provides additional evidence for a stature decline with the adaptation of agriculture. Nevertheless, several regional studies as e.g. Bennicke (1985) for modern Denmark, Jaeger et al. (1998) and Siegmund (2010) for Central Europe as well as Angel (1984) for Greece indicate a recovery of body height in the Copper and Bronze Ages. It is the intention of this article to combine anthropological measures from different prehistoric time periods and geographic locations and provide therefore a valuable retrospective picture of prehistoric living standards.

Body height, our variable of interest, is not directly observable for prehistoric individuals but has to be reconstructed from long bone measurements taken from skeletons preserved in excavations. There are several formulas typically based on linear regression found among the literature to compute body height by single or combinations of longbone measures. The most commonly used and most accurate formulas are the ones by Pearson (1899), Breitinger (1937)/Bach (1965) and Trotter and Gleser (1952). Moreover, the prehistoric archeological sites from which the long bone data is obtained are usually irregularly distributed over space and time. A division into categories of time or space as mostly done in archeological publications for the sake of convenience is very hard to justify. For that reason a model with a nonparametric spatio-temporal trend pattern is proposed here.

Specifically, we suggest an additive regression model, which is a common approach for spatio-temporal smoothing (see e.g., Currie et al. 2004; Lee and Durbán 2011). In the present case, the following model, with $\boldsymbol{Y}$ = body height, $\boldsymbol{s}$ = space and $\boldsymbol{t}$ = time, is proposed:

$$\boldsymbol{Y} = g_{male}(\boldsymbol{s},\boldsymbol{t}) + g_{female}(\boldsymbol{s},\boldsymbol{t}) + \boldsymbol{\epsilon},\ ,\ \boldsymbol{\epsilon} \sim N(0,\boldsymbol{\Sigma})$$

A 3-dimensional smoother $g(\boldsymbol{s},\boldsymbol{t})$ is considered here to account for a space-time in-

teraction effect and is modeled separately for males and females. Additionally, as the data is structured hierarchically – with the archaeological site/occupation period of a site on the higher level and skeletons/individuals at lower level – a population specific effect for capturing dependencies such as similar genetic make-ups due to kinship or local cultural and environmental factors should be introduced in the statistical model as well. Thus, a promising model candidate would be a semiparametric mixed model (Ruppert et al., 2003). In fact, a similar model has been proposed for estimation of trends in physical stature of Americans in the 19th century (Lang and Sunder, 2003). In our case, the upper formula is complemented by a random site-specific effect $\boldsymbol{\gamma}$ to account for this.

Unfortunately, one cannot observe the true values of time of birth and sex, but only estimate with some uncertainty. In general, ignoring such measurement error may lead to severely biased estimates and a loss of power and therefore measurement error should be accounted for in the statistical model (Carroll et al., 2006). More specifically, the individual date of birth is given only by a certain time interval, which is the same for all skeletons of an occupation period of a site. When radiocarbon data is available, the interval width can sometimes as short as a few dozen years, whereby in other cases it can be as large as two thousand years due to a rough resolution in archaeological dating. Additionally, sex is given on an ordinal 5-category scale ('female', 'likely female', 'uncertain', 'likely male', 'male'), so one has also to deal with misclassification. However, if we could estimate the proportion of males in the middle three categories from the data, potential underrepresentation of females caused by sex determination of anthropologists could be uncovered, e.g. if the data suggests that the proportion of males in the category 'likely female' exceeds 50%. This could provide an important hint for solving the mystery of low percentages of female individuals frequently observed in prehistoric findspots (Kemkes-Grottenthaler, 1997).

In this article a statistical model that is able to cope with measurement error in spatio-temporal data is introduced. Furthermore, this model is applied to a dataset with prehistoric long bone measurements. After some more specific information on the available dataset in Section 4.2, a short overview on (Bayesian) smoothing methods and errors-in-variables models is given in Section 4.3. Thereafter, the statistical model is presented in detail. Computational details and results are presented in Section 4.4, followed by a short discussion.

4.2 Data

One goal of the LiVES project is to collect a broad sample of prehistoric long bone measurements on the basis of a large skeleton database, the 'Mainzer Lochkartenarchiv für postkraniales Skelettmaterial' (Perscheid, 1974). Originally, this data was stored on punch cards, but has been digitized and converted to a MySQL database. All entries from the database have to be carefully examined and corrected since the chronological classification is mostly outdated. Additionally, we have searched for additional publications, where anthropological measurements of the excavated skeletons from prehistoric

sites are available. The corrected entries from the Mainz database and the ones collected from the literature were then merged into a new database. The extracted dataset contains long bone measurements of 3052 skeletons from 219 sites.

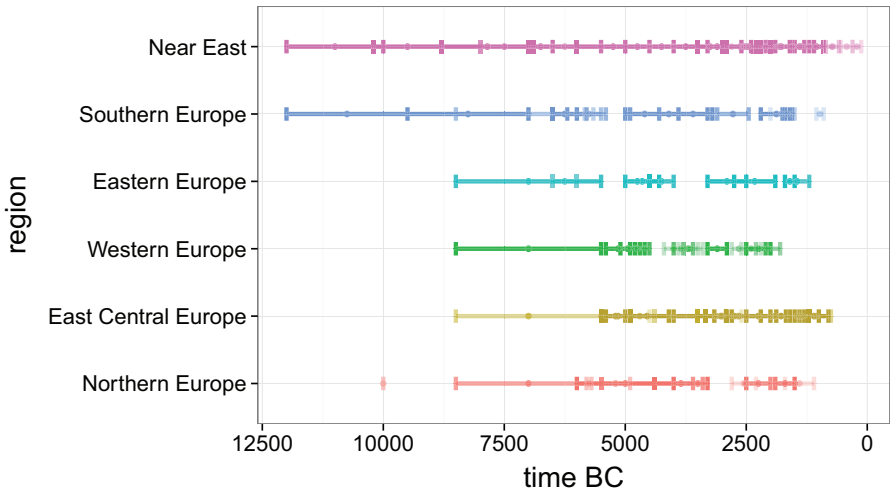

Figure 4.1: Assumed time intervals for the durations of individual sites for six regions. To demonstrate densities by overlapping intervals, an alpha (transparency) factor is introduced in plotting

Specifically, the dataset consists of the following variables: the reconstructed body height in mm $\boldsymbol{Y}$, latitude and longitude in degrees for the site $\boldsymbol{s} = (\boldsymbol{lat}, \boldsymbol{long})$, the 'observed' time of birth interval $\boldsymbol{t}_{obs} = (\boldsymbol{t}^-, \boldsymbol{t}^+)$, where $\boldsymbol{t}^-, \boldsymbol{t}^+$ denote the lower and upper interval borders determined by an expert archeologist or radiocarbon dating, the identification number of the site $\boldsymbol{id}$, the reliability category of the height measures $\boldsymbol{rel}$ (see below) and the observed sex category $\boldsymbol{sex}_{obs}$.

The sites are distributed across Europe, Western Asia and Northern Africa, from approximately 12000 BC to 200 BC, with a focus on the Neolithic as well as Copper and Bronze ages covering the time period from about 8000 BC to 1000 BC. Figures 4.1 and 4.2 show the distribution of archeological sites over time.

In the vast majority of cases, the prehistoric skeletons are only partially preserved. Thus, we have to rely on reconstruction formulas computed with an evaluation sample of recent individuals. The body height is calculated by the linear regression formulas of Pearson (1899), because in general they yield good approximations for both males and females (Formicola, 1993). The reliability category $\boldsymbol{rel}$ reflects the varying quality of literary references and the state of preservation of the skeletons' long bones:

- Category 1: femur or humerus length available ($n = 1183$)
- Category 2: other long bone measures available ($n = 460$)
- Category 3: body height available with known formula applied to unknown long bone ($n = 1330$)

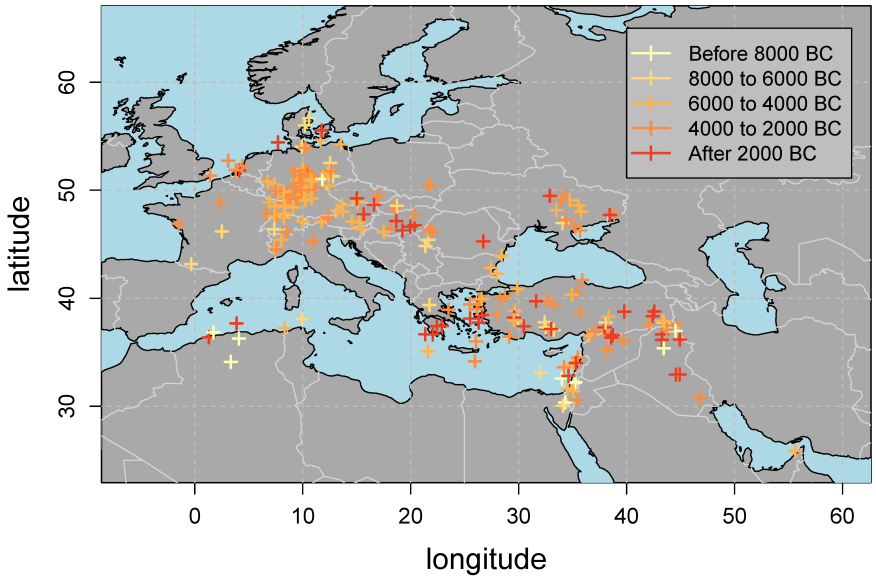

Figure 4.2: Distribution of archaeological sites over time categories (with small random error added to the latitude and longitude)

- Category 4: body height available with unknown formula applied to unknown long bone. ($n = 79$)

The femur and humerus are considered to be the most reliable long bones for body height reconstruction. Therefore, one would expect that reconstructed height values of individuals from category 1 should have a lower error variance than those from categories 2-4. Category 4 should be the least reliable, whereas it is hard to say whether category 2 or 3 is more reliable. In category 3 the long bone is unknown and therefore decreasing reliability, while it can contain body height reconstructed by femur or humerus measurements, which should lead to increased reliability compared to category 2.

The observed sex variable has 5 categories ('female', 'likely female', 'uncertain', 'likely male', 'male'), whereby it is assumed that the first and the last contain no misclassification, although this can occur sometimes but is very hard to quantify. The categories are determined by an expert using criteria independent of the long bone lengths and the body height. Table 4.1 illustrates frequency distribution of the observed sex category and their mean body height. We have only about 40% females in our sample which is common for prehistoric skeleton data as pointed out in Section 4.1.

Table 4.1: Body height means, standard deviations and frequencies of observed sex categories

Sex category	Mean	SD	n	%
1 (female)	1.568 m	6.787	1074	35.2
2 (likely female)	1.598 m	7.387	79	2.6
3 (uncertain)	1.658 m	6.817	147	4.8
4 (likely male)	1.682 m	6.209	145	4.8
5 (male)	1.683 m	5.837	1607	52.6

4.3 Methods

4.3.1 Spatio-temporal modeling

In general, spatio-temporal modeling considers a random process $\{\boldsymbol{Y}(\boldsymbol{s},\boldsymbol{t}) : \boldsymbol{s} \in D_s, \boldsymbol{t} \in D_t\}$ for a variable $\boldsymbol{Y}$ on the spatio-temporal index set $D_s \times D_t$ (Cressie and Wikle, 2011). Typically, $\boldsymbol{Y}(\boldsymbol{s},\boldsymbol{t})$ is decomposed into a smooth spatio-temporal surface $g(\boldsymbol{s},\boldsymbol{t})$ and some white noise $\boldsymbol{\epsilon}$. There exist many approaches for spatio-temporal smoothing as a special case of multivariate smoothing, which can be roughly divided into methods based on autocovariance functions (Gelfand et al., 2010), local smoothing methods and basis function methods (Fahrmeir et al., 2013). One popular approach of the latter class, which is pursued in this work, is the penalized B-Spline smoothing ('P-Spline', Eilers and Marx 1996), because of its low computational costs as a low-rank smoother and the ease of incorporation into the generalized linear mixed model framework. Instead of placing knots or basis functions at each covariate location, a sufficiently high number of knots evenly spread throughout the covariate space is chosen. The interesting function, the spatio-temporal surface, $g(\boldsymbol{s},\boldsymbol{t})$ is written as a weighted sum of B-spline basis functions $g(\boldsymbol{s},\boldsymbol{t}) = \boldsymbol{Z}\boldsymbol{\beta}$, where $\boldsymbol{Z}$ is a design matrix of B-Splines with dimension $n \times p$. To balance between roughness and smoothness ('bias-variance tradeoff') a penalty is introduced. In the case of a single covariate, the penalty consists of the squared regression parameters $\boldsymbol{\beta}$ weighted by the $p \times p$ penalty matrix $\boldsymbol{K}$ combined with a smoothing parameter λ, which controls the degree of smoothness. Typically, the second differences of the B-Spline functions are penalized to construct the penalty matrix $\boldsymbol{K}$ (see Eilers and Marx 1996 for specifics). For a linear model, $\boldsymbol{\beta}$ is estimated by minimizing the penalized residual sum of squares:

$$\text{argmin}_{\boldsymbol{\beta}}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}$$

For a multivariate version (Wood, 2006) a Kronecker Product of the marginal basis or design matrices $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, ..$ is applied to construct the multivariate basis-functions. The multivariate penalty matrix $\boldsymbol{K}^*$ can be constructed in a similar way by the marginal penalty matrices $\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3, ..$ and their smoothing parameters $\lambda_1, \lambda_2, \lambda_3, .. .$ This strategy is not restricted to P-Splines only. Any other appropriate smoothing basis is suitable as well. Additionally, the so-called (approximate) thin plate regression splines are a popular alternative in multivariate smoothing (Wood, 2003). They penalize the

integral of the m-th derivatives of the smooth function and have some desirable properties such as being rotation-invariant. However, the covariates are assumed to be on the same scale, i.e. they are not invariant to rescaling, since they are handled isotropically. This is not the case for spatio-temporal data as space and times are measured on completely different scales. A covariate rescaling is sometimes performed, but this is a rather ad-hoc workaround. Therfore, the author favors a tensor product P-Spline as introduced above. By having a unique penalty for each covariate, scale-invariance – but not rotation-invariance – is guaranteed. There is no real justification for this approach, such as being optimal in some sense, but it has been proven very useful in practical applications. The components of $\boldsymbol{\lambda}$ are generally estimated by searching on a grid of possible values and selecting the combination which minimizes the AIC or some cross-validation criterion, which can be computationally prohibitive for multiple smoothing parameters. Another possibility is to estimate $\boldsymbol{\lambda}$ within a mixed model by assuming a Gaussian distribution for $\boldsymbol{\beta}$ (Ruppert et al., 2003). In a Bayesian approach (Lang and Brezger, 2004), which is used here, the penalty parameter $\boldsymbol{\lambda}$ is considered as a random variable determining the precision of the (random) regression parameters $\boldsymbol{\beta}$.

4.3.2 Errors-in-variables models

Measurement error concepts

In practice, variables are often contaminated with measurement error. Mismeasured independent variables in regression models, for example, lead to biased estimates, a loss of power and tend to mask or smooth out features in the data (Carroll et al., 2006). The two most common types of measurement error are the so called 'classical' measurement error and the 'Berkson' (Berkson, 1950) measurement error.

Let W_i be the observed, mismeasured variable value and X_i be the unobserved, true value. The classical error model has the following form:

$$W_i = X_i + U_i \ , \ i = 1, .., n$$

with U_i and X_i independent, $E(U_i|X_i) = 0$. The errors U_i are often assumed to be i.i.d. normal, but other distributions, non-additive forms, heteroscedastic errors or even correlated errors are also conceivable.

Contrary to the classical measurement model, the errors U_i in the Berkson type error structure are assumed independent of the observed value W_i: $E(U_i|W_i) = 0$.

$$X_i = W_i + U_i \ , \ i = 1, .., n$$

Both formulas may seem quite similar at first glance so it is important to understand the differences and the implications of both error types. The classical measurement error model is suitable for an instrument, which can measure the true value X_i only with a limited precision or if we have to replace a population mean with a single measurement

W_i. The Berkson error model is in contrast appropriate when we have population means instead of individual exposures. Buonaccorsi (2010) introduces many practical examples of both error types.

In our case we have the information that an individual was born in a certain interval that is derived from knowledge about the related culture. Thus, we have information about the population respectively of the archaeological site or settlement but not about the skeleton itself, which can be translated into a Berkson measurement error as explained above.

In the special case of linear regression with a single mismeasured covariate with classical i.i.d errors, a naive estimation of the regression coefficient leads to attenuation, i.e. the slope parameter is biased towards zero. If the measurement error variance or the ratio of measurement error and residual error is known, one can correct for this attenuation (Frost and Thompson, 2000). In the case of Berkson type measurement error the estimator is still unbiased. However, in more complex models with more covariates and/or in nonlinear models, the effects of both classical and Berkson errors are not that obvious (Carroll et al., 2006). In general, the error distribution parameters have to be known, e.g. mean and variance for normally distributed error, or multiple measurements have to be available. Otherwise, the problem can be non-identifiable.

For discrete covariates, measurement error is equivalent to misclassification. Again, in our application we have Berkson type misclassification, because we are only interested in the probability of being male for the skeletons of a certain sex category and not the other way around. By knowing the misclassification probabilities one is able to construct the misclassification matrix Π with entries π_{xw}:

$$\pi_{xw} = P(X_i = x|W_i = w)$$

There exist a variety of approaches to correct for measurement errors such as regression calibration (Spiegelman et al., 1997), SIMEX (Simulation-Extrapolation; Cook and Stefanski 1994), method of moments estimators (Fuller, 2009), maximum-likelihood methods (Higdon and Schafer, 2001) or Bayesian methods (see e.g. Richardson and Gilks 1993; Gustafson 2003). Carroll et al. (2006) gives a thorough overview on these methods.

Bayesian errors-in-variables-models

The Bayesian approach to errors-in-variables models is to select an *observation model* as if all variables were observed without measurement error, then form an *error model* for the observed variables with measurement error and – only for classical errors – model the distribution of the true unobservable variables ('structural model'; see, for example, Carroll et al. 2006). The observations of the mismeasured variable are treated as additional parameters in the procedure. Furthermore, appropriate prior distributions have to be chosen. The Bayesian approach to EIV models has several advantages. Its modularity allows us to implement and compute relatively complex models in a straightforward manner. In this context we can specify the measurement error model and the

observation model seperately. They are easily combined to a complete likelihood because of their hierarchical structure afterwards. Furthermore, in combination with Markov Chain Monte Carlo (MCMC) we are able to perform inference on the model parameters (even on the unobserved, true parameters!) without resampling methods required for regression calibration or SIMEX. It is also notable that the Bayesian approach uses all information when imputing the true unobserved covariates contrary to regression calibration as Carroll et al. (2006) points out. Another advantage is that available prior information on parameters can be incorporated as well. Gustafson (2005) states that putting prior information on usually non-identified parameters such as measurement error variances is advantageous compared to using best-guess values. A disadvantage shared with maximum likelihood estimation is the high computational cost.

For a simple linear regression, with formula $\boldsymbol{Y} = \beta \boldsymbol{X} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N\left(0, \sigma_\epsilon^2\right)$, normally distributed Berkson errors and their known error variance σ_u^2, the posterior equals the following expression, whereby $[\cdot]/[\cdot|\cdot]$ denote the unconditional/conditional distribution $f(\cdot)/f(\cdot|\cdot)$:

$$\begin{aligned}[\beta, \boldsymbol{X}, \sigma_\epsilon^2 | \boldsymbol{Y}, \boldsymbol{W}] &\propto [\boldsymbol{Y}|\beta, \boldsymbol{X}, \sigma_\epsilon^2][\boldsymbol{X}|\boldsymbol{W}][\beta|\sigma_\epsilon^2][\sigma_\epsilon^2] \\ &\propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}(Y_i - X_i\beta)^2 - \frac{1}{2\sigma_u^2}\sum_{i=1}^{n}(X_i - W_i)^2\right) \times [\beta|\sigma_\epsilon^2][\sigma_\epsilon^2]\end{aligned}$$

This can be extended straightforwardly to heteroscedastic measurement errors or non-Gaussian distributions. For a linear model with a mismeasured categorical covariate (Berkson error), the posterior equals

$$\begin{aligned}[\boldsymbol{\beta}, \boldsymbol{X}, \sigma_\epsilon^2 | \boldsymbol{Y}, \boldsymbol{W}] &\propto [\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}, \sigma_\epsilon^2][\boldsymbol{X}|\boldsymbol{W}][\boldsymbol{\beta}|\sigma_\epsilon^2][\sigma_\epsilon^2] \\ &\propto \prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma_\epsilon^2}(Y_i - z_i'\boldsymbol{\beta})^2\right) \times \pi_{xw} \times [\boldsymbol{\beta}|\sigma_\epsilon^2][\sigma_\epsilon^2]\end{aligned}$$

with $\boldsymbol{Z}$ as design matrix constructed from $\boldsymbol{X}$ and known transition probabilities $\pi_{xw} = P(X_i = x|W_i = w)$. If the entries π_{xw} of the misclassification matrix $\boldsymbol{\Pi}$ are not exactly known one can place a suitable conjugate prior distribution on π_{xw}, the beta distribution $Be(A_{xw}, B_{xw})$:

$$\begin{aligned}[\boldsymbol{\beta}, \boldsymbol{X}, \sigma_\epsilon^2, \boldsymbol{\Pi} | \boldsymbol{Y}, \boldsymbol{W}] &\propto [\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}, \sigma_\epsilon^2][\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\Pi}][\boldsymbol{\Pi}][\boldsymbol{\beta}|\sigma_\epsilon^2][\sigma_\epsilon^2] \\ &\propto \prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma_\epsilon^2}(Y_i - z_i'\beta)^2\right) \times \pi_{xw} \times \pi_{xw}^{A_{xw}-1}(1-\pi_{xw})^{B_{xw}-1} \\ &\quad \times [\boldsymbol{\beta}|\sigma_\epsilon^2][\sigma_\epsilon^2]\end{aligned}$$

It is easy to see, that if only some values of a discrete (or continuous) covariate inherit some measurement error, as for sex in the application data, the transition probabilities π_{xw} (or measurement error variance) for the mismeasured values are identifiable.

In general, the posterior marginal distributions cannot be computed analytically. However, a MCMC sampler can be utilized to sample from these distributions. MCMC gives accurate results but is sometimes hard to implement and rather time consuming. Thus, alternative approximate estimation approaches such as the variational Bayes method (Bishop, 2006) or the integrated nested Laplace approximation algorithm (INLA; Rue and Held 2005) became popular recently. Variational Bayes was implemented by Pham et al. (2013) for linear models with classical measurement error, while Muff et al. (2015) applied INLA to parametric regression models with classical as well as Berkson measurement error. However, in this application there arose no serious problems when implementing the MCMC-method (see also Section 4.4).

4.3.3 Nonparametric regression in presence of measurement error

Several approaches have been introduced lately to combine nonparametric regression and measurement error models. Schafer (2001) developed an EM algorithm for semiparametric likelihood analysis in linear and nonlinear regression models. A kernel estimator and its convergence rates are derived by Fan and Truong (1993). This work is generalized from local constant estimators to local polynomial estimators by Delaigle et al. (2009). An approach using a modified least squares criterion is introduced in Taupin (2001). A fully Bayesian model using P-splines in the nonparametric part is proposed by Berry et al. (2002), while an estimator for heteroscedastic measurement errors using kernel methods is developed by Delaigle and Meister (2007). Carroll et al. (1999) proposed two estimators based on a SIMEX approach using kernel regression or penalized splines methods and an approach assuming a mixture of normals distribution for the unobserved covariates combined with regression splines. The case of Berkson errors is considered in Delaigle et al. (2006), while Carroll et al. (2007) is concerned with data contaminated with both Berkson and classical errors. However, little work was done for the error-in-variables problem in multivariate smoothing, which is relevant to answer the research question formulated in the introduction. A Bayesian approach is pursued in this work, since such complex models required for the research problem are relatively easy to handle within the Bayesian framework because of its modularity and other advantages stated in the previous subsection.

4.3.4 Statistical model for prehistoric anthropological data

To model the spatio-temporal trends in body height we introduce a Bayesian additive mixed model with errors-in-variables. Following Carroll et al. (2006) we begin with formulating our additive mixed model as if time ($\boldsymbol{t}$) and $\boldsymbol{sex}$ ($sex_i = 1$ for males and 0 for females) were observed:

$$Y_i = g_{male}(s_i, t_i) \times sex_i + g_{female}(s_i, t_i) \times (1 - sex_i) + \boldsymbol{u}_i'\boldsymbol{\gamma} + \epsilon_i \text{ ; for } i = 1, .., n$$

$g_{female}(\boldsymbol{s}, \boldsymbol{t}), g_{male}(\boldsymbol{s}, \boldsymbol{t})$ are the global spatio-temporal trend functions for males and females, which are modeled seperately to account for varying sexual dimorphism. A tensor product B-Spline, which was chosen here, or any other type of multidimensional

smoother constructing on penalized basis functions can implement this smooth function. $\gamma_m \sim N(0, \tau)$, $m = 1, .., M$, reflects the random population-specific effect of settlement or archaeological site containing local cultural, economic and environmental factors or genetic dependencies, with random effects design matrix $\boldsymbol{U}$ with rows $\boldsymbol{u}_i$ constructed from the site identification number $\boldsymbol{id}$. The residual error ϵ_i is assumed Gaussian and can be internally decomposed into the population specific variability and body height reconstruction uncertainty, wheareas the latter depends on the reliability category $rel_i \in \{1, 2, 3, 4\}$. Therefore, the error is assumed to be independent and heteroscedastically normally distributed, $\epsilon_i \sim N(0, \sigma^2_{rel_i})$, $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$. As a consequence, the height $\boldsymbol{Y}$ is multivariate normally distributed with mean $\boldsymbol{Z\beta} + \boldsymbol{U\gamma}$ and covariance $\boldsymbol{\Sigma}$. $\boldsymbol{Z} = (\boldsymbol{Z}_{male}|\boldsymbol{Z}_{female})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_{male}, \boldsymbol{\beta}_{female})$ denote the corresponding design matrix and parameter vector of the B-Spline basis.

Again, according to Carroll et al. (2006), the measurement error models have to be specified in a second step. We have Berkson errors for time ($\boldsymbol{t}$) and $\boldsymbol{sex}$. The distribution of the true time of birth given the observed time interval $t_{obs,i} = (t_i^-, t_i^+)$ is translated into an uniform distribution: $[t_i|t_{obs,i}] \sim U(t_i^-, t_i^+)$. The misclassification model for sex given the observed category and the associated transition probability is a simple Bernoulli distribution: $[sex_i|sex_{obs,i} = j, \pi_j] \sim Ber(\pi_j)$. The likelihood functions of the additive mixed model and the measurement error model can be easily combined into a complete likelihood because of their hierarchical structure. Together with prior distributions on the model parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{t}, \boldsymbol{sex}, \boldsymbol{\lambda}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \tau, \boldsymbol{\pi}\}$, the posterior can be formally written as:

$$\begin{aligned}[\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{s}, \boldsymbol{t}_{obs}, \boldsymbol{sex}_{obs}, \boldsymbol{id}, \boldsymbol{rel}] \propto\ & [\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{sex}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, \boldsymbol{id}] \\ & \times [\boldsymbol{t}|\boldsymbol{t}_{obs}][\boldsymbol{sex}|\boldsymbol{sex}_{obs}, \boldsymbol{\pi}] \\ & \times [\boldsymbol{\beta}|\boldsymbol{\lambda}][\boldsymbol{\lambda}][\boldsymbol{\Sigma}][\boldsymbol{\gamma}|\tau][\tau][\boldsymbol{\pi}]\end{aligned}$$

On the left the first line represents the posterior, and observed on the right, the likelihood of the model as if the covariates were measured without error. In the second line, a Berkson measurement error model for time ($\boldsymbol{t}$) and $\boldsymbol{sex}$ is introduced, while the third line contains the priors of the parameters. In general, weakly informative conjugate prior distributions were chosen except for the male proportion parameters. Specifically, for $[\boldsymbol{\beta}|\boldsymbol{\lambda}]$, with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, a Gaussian prior was used: $\boldsymbol{\beta} \sim N(0, \boldsymbol{K}^{-1})$. The same B-Spline basis was employed for both males and females leading to equal penalty matrices such that $\boldsymbol{K}$ has a block diagonal structure, $\boldsymbol{K} = blockdiag(\boldsymbol{K}^*, \boldsymbol{K}^*)$. The multivariate penalty matrix $\boldsymbol{K}^*$ is constructed from the marginal penalty matrices $\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3$ as described in Wood (2006): $\boldsymbol{K}^* = \lambda_1\boldsymbol{K}_1^* + \lambda_2\boldsymbol{K}_2^* + \lambda_3\boldsymbol{K}_3^*$, $\boldsymbol{K}_1^* = \boldsymbol{K}_1 \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{I}_{p_3}$, $\boldsymbol{K}_2^* = \boldsymbol{I}_{p_1} \otimes \boldsymbol{K}_2 \otimes \boldsymbol{I}_{p_3}$, $\boldsymbol{K}_3^* = \boldsymbol{I}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{K}_3$. Appropriate independent conjugate prior distributions were chosen for the smoothing parameters: $\lambda_{1,2,3} \sim GA(a_\lambda, b_\lambda)$, scale parameters: $\sigma^2_{1,..,4} \sim IG(a_\sigma, b_\sigma)$, $\tau \sim IG(a_\tau, b_\tau)$, random effects $\gamma \sim N(0, diag(\tau))$ and the proportion of males in the observed sex categories: $\pi_{2,3,4} \sim Be(A, B)$. As mentioned in the previous subsection, the probabilities are fixed to 0 and 1 for the first and last observed sex category: $\boldsymbol{\pi} = (0, \pi_2, \pi_3, \pi_4, 1)$. Vague but proper prior parameters were chosen for the inverse gamma ($a_\sigma, a_\tau = 0.0001, b_\sigma, b_\tau = 0.0001$) distributions. For the gamma distributions of the smoothing parameters $\lambda_{1,2,3}$ the values $a_\lambda = 1$ and

$b_\lambda = 0.0001$ were selected (see also Lang and Brezger 2004). For the beta distributions of π_j, Jeffreys' prior (Kass and Wasserman, 1996) with $A = 0.5$ and $B = 0.5$ or alternatively an informative prior with $(A_2, B_2) = (5, 20)$, $(A_3, B_3) = (12.5, 12.5)$ and $(A_4, B_4) = (20, 5)$ are candidates. An informative prior was chosen here, since statements as 'likely male', 'uncertain' and 'likely female' actually contain some valuable information about the sex variable. These priors correspond to distributions with mean proportions of 25%, 50% and 75% of males in the observed sex category, which seems reasonable. Figure 4.3 shows the informative prior distributions. After having specified the prior distributions, the complete posterior is proportional to the following expression presented in the same order as above, with $k_i = rel_i$ and $j_i = sex_{obs,i}$ for simpler annotation:

$$\begin{aligned}
&\propto \prod_{i=1}^{n} \left(\sigma_{k_i}^{-1/2} \exp\left(-\frac{1}{2\sigma_{k_i}^2}(Y_i - z_i'\boldsymbol{\beta} - \boldsymbol{u}_i'\boldsymbol{\gamma})^2 \right) \times \frac{I(t_i^+ < t_i < t_i^-)}{t_i^- - t_i^+} \times \underbrace{\pi_{j_i}^{sex_i}(1-\pi_{j_i})^{1-sex_i}}_{\text{only for } sex_{obs,i} \in \{2,3,4\}} \right) \\
&\times \prod_{l=1}^{3} \left(\exp\left(-\frac{1}{2}(\boldsymbol{\beta}'\lambda_l \boldsymbol{K}_l^* \boldsymbol{\beta}) \right) \times \frac{\lambda_l^{a_\lambda - 1}}{b_\lambda^{a_\lambda}\Gamma(a_\lambda)} \exp\left(-\frac{\lambda_l}{b_\lambda} \right) \right) \times \prod_{k=1}^{4} \frac{b_\sigma^{a_\sigma}}{\Gamma(a_\sigma)} \sigma_k^{-2(a_\sigma+1)} \exp\left(-\frac{b_\sigma}{\sigma_k^2} \right) \\
&\times \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'\tau^{-1}\boldsymbol{\gamma} \right) \times \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \tau^{-2(a_\tau+1)} \times \prod_{j=2}^{4} \left(\pi_j^{A_j - 1}(1-\pi_j)^{B_j - 1} \right)
\end{aligned}$$

Following some algebra the complete conditionals, i.e. the posterior distributions given all other parameters, for the model parameters are:

$$\boldsymbol{\beta}|. \sim N((\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{K})^{-1}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{\gamma}), (\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{K})^{-1})$$

$$t_i|. \propto \exp\left(\frac{(Y_i - z_i'\boldsymbol{\beta} - \boldsymbol{u}_i'\boldsymbol{\gamma})^2}{-2\sigma_k^2} \right) \times U(t_i^-, t_i^+)$$

$$sex_i|. \propto \exp\left(\frac{(Y_i - z_i'\boldsymbol{\beta} - \boldsymbol{u}_i'\boldsymbol{\gamma})^2}{-2\sigma_k^2} \right) \times Ber(\pi_j)$$

$$\pi_j|. \sim Be\left(A_j + \sum_{i:sex_{obs}=j} sex_i, B_j + \sum_{i:sex_{obs}=j} (sex_i - 1)_+ \right)$$

$$\boldsymbol{\gamma}|. \sim N\left((\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U} + diag(1/\tau))^{-1}\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta}), (\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U} + diag(1/\tau))^{-1} \right)$$

$$\sigma_k^2|. \sim IG\left(a_\sigma + n_k/2, b_\sigma + 0.5 \sum_{i:rel_i=k} \left(Y_i - z_i'\boldsymbol{\beta} - \boldsymbol{u}_i'\boldsymbol{\gamma} \right)^2 \right)$$

$$\tau|. \sim IG\left(a_\tau + M/2, b_\tau + 0.5 \sum_{m=1}^{M} \gamma_m^2 \right)$$

$$\lambda_l|. \sim GA\left(a_\lambda + rank(\boldsymbol{K}_l^*)/2, (b_\lambda + 0.5\boldsymbol{\beta}'\boldsymbol{K}_l^*\boldsymbol{\beta})^{-1} \right)$$

The computation is realized via MCMC. Gibbs sampling for drawing observations from the posteriors is performed using the complete conditionals in the order presented

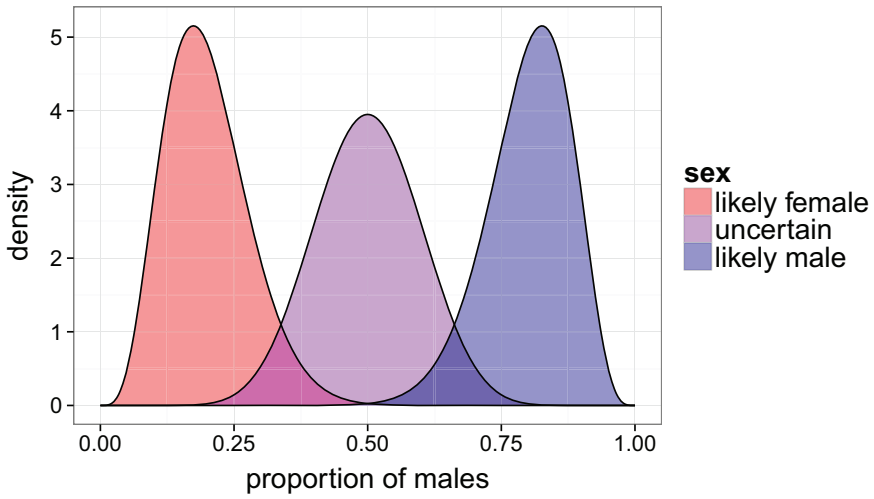

Figure 4.3: Beta priors for male proportions in observed sex categories

above. A Metropolis-Hastings step is necessary for t_i and sex_i, because the conditional distributions are not available for sampling directly as the design matrix $\boldsymbol{Z}$ is a highly complex function of $\boldsymbol{t}$ and $\boldsymbol{sex}$. Candidate observations for t_i are generated by a $U(t_i^-, t_i^+)$ distribution, while sex_i is drawn from a $Ber(0.5)$ distribution. The algorithm is quite slow, because the values of the design matrix $\boldsymbol{Z}$ are changed by iteration, which require computationally intensive matrix inversion steps for each one.

4.4 Computation and Application

All computations are done via R (R Development Core Team, 2008). To construct the spline basis functions and penalty matrices, the *mgcv* package was used (Wood, 2011). The R-code of the entire program is available on request from the author. For the MCMC algorithm, four different chains were started from widely varied starting values with a total size of 20000 samples each. Specifically, starting values in chains 1 to 4 were set to: (0.1,1,10,100) for scale paramaters; (0.2,0.4,0.6,0.8) for proportion parameters and (-1,0,1,10) for others. The computation took about 40 hours on an Intel Core i7 3.4 GHz processor with 16 GB RAM.

4.4.1 Convergence diagnostics

Convergence checks have been made visually by trace and running mean plots and by comparing the results of the four chains using the Gelman-Rubin statistic (Gelman and Rubin, 1992). The mixing was quite fast and all chains converged to the same parameter values. A burn-in period of 7500 burn-in samples was considered to be sufficient. With a thinning of m = 5 this results in a sample size of 2500 for each chain. Figure 4.4 is exemplary at displaying the running means for τ, π_2, σ_1^2 and λ_1 for the four MCMC chains. The Gelman-Rubin statistics exhibited potential scale reduction factors well below 1.1 for all parameters (e.g. 1.0003, 1.00002, 1.0002, 1.002 for τ, π_2, σ_1^2 and λ_1).

Results were virtually the same for moderate changes in the prior parameters of the

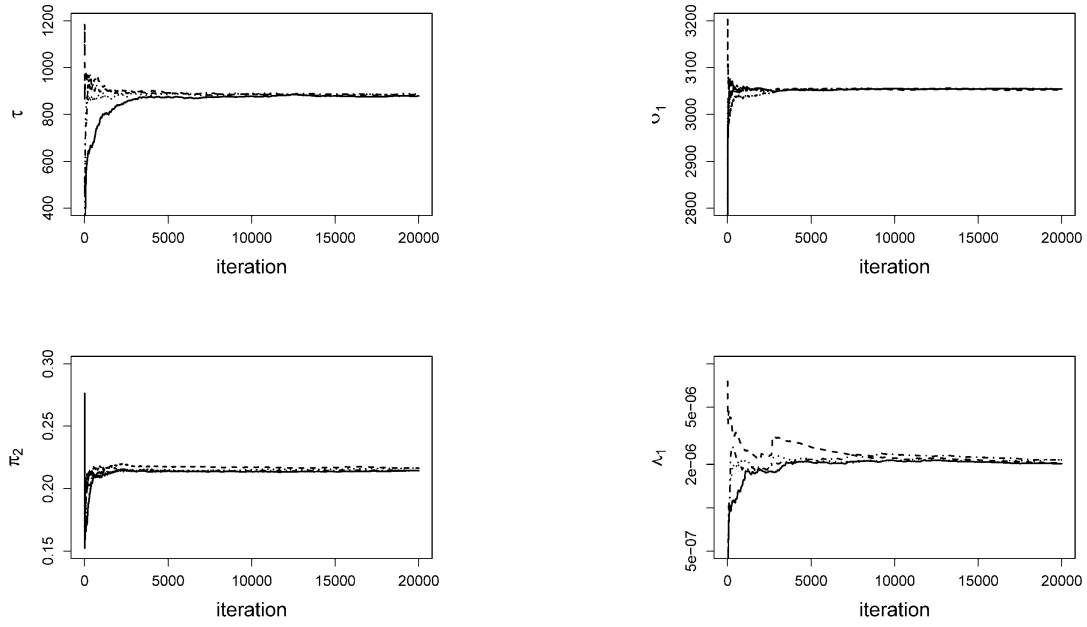

Figure 4.4: Running means for τ, π_2, σ_1^2 and λ_1

gamma and inverse gamma distributions (with $a_\sigma, a_\tau = 1$; $b_\sigma, b_\tau = 1$).

4.4.2 Model fit and comparison

To evaluate the benefit of our errors-in-variables model, three different models were estimated. Model 1 ignores the uncertainty in both sex and time of birth. This was done by setting the sex categories 'likely female' and 'likely male' to 'female' and 'male' respectively, assigning sex randomly with probability 0.5 for the 'uncertain' sex category and utilizing the center of the observed time interval as time of birth. Model 2 accounts for sex misclassification, while Model 3 considers the time of birth uncertainty, as described in the previous section. The fourth model combines both types of measurement errors. The performance of different models is evaluated by a modified version of the Deviance Information Criterion (DIC, Spiegelhalter et al. 2002), the DIC4, introduced in Celeux et al. (2006), because the parameter focus lies on the spatio-temporal time trend and not on the true, mismeasured covariate values. The DIC4 integrates over these latent variables and does not count them as model parameters. The corresponding number of effective parameters p_{D_4} are utilized as a measure for model complexity. The original motivation for the use of DIC4 was missing data models, but it is also applicable in this case because mismeasured values can be viewed as a milder case of missing ones (Blackwell et al., 2014). Table 4.2 compares the fit of the models. Accounting for sex misclassification in Model 2 results in a decent improvement in DIC4 and adjusted R^2, while this is rather marginal for uncertainty in time of birth in Model 3. However, the

combination of both in Model 4 leads to the best model fit. 512 (8^3) knots were chosen for tensor product P-Spline basis functions. Testing showed that a larger number of knots (10^3 or 12^3) did improve the model fit only marginally for an exceedingly high computational cost.

Table 4.2: Model comparison by adjusted R^2, DIC4 and *effective* number of parameters p_{D_4}

Model	*Adj.* R^2	*DIC4*	p_{D_4}
1 (no uncertainty controlled)	0.462	33868	146.38
2 (sex uncertainty controlled)	0.481	33691	120.42
3 (time uncertainty controlled)	0.467	33793	129.93
4 (sex + time uncertainty controlled)	0.488	33657	114.96

4.4.3 Results

In terms of R^2 sex, time and location are able to explain almost 50% of the variance in the estimated individual body height. Posterior means and credible intervals for variance components and other parameters of Model 4 are shown in Table 4.3. As

Table 4.3: Posterior means, standard errors and credible intervals of model parameters (Model 4)

Parameter	Posterior mean	Posterior sd	95% HPD interval
σ_1^2	3055	141.35	[2792.59, 3347.44]
σ_2^2	4432	333.29	[3825.29, 5126.51]
σ_3^2	3845	308.93	[3273.09, 4506.51]
σ_4^2	3576	678.35	[2505.47, 5068.91]
π_2	0.22	0.07	[0.10, 0.37]
π_3	0.54	0.08	[0.40, 0.70]
π_4	0.84	0.06	[0.71, 0.94]
τ	879.58	199.75	[539.85, 1319.42]
λ_1	1.99E-6	1.32E-6	[6.94E-7, 5.05E-6]
λ_2	1.98E-6	1.64E-6	[5.30E-7, 7.74E-6]
λ_3	5.04E-6	3.31E-6	[1.53E-6, 1.37E-5]

expected in the previous section, the reliability category 1 has the lowest residual variance. The estimated proportions of males in the sex categories are 22%, 54% and 84%. These proportions are shifted even more towards male when using Jeffreys' prior with mean posterior proportions of $\pi_{2,3,4} = (0.28, 0.70, 0.94)$. Hence, the model estimates do not give any evidence that the sex determination by the archaeologists is generally biased towards a higher proportion of males. The smooth functions $g_{female}()$ and $g_{male}()$ are visualized in Figure 4.5 (females) and 4.6 (males) for Model 4.

Most noticeable is that from about 5000 to 1000 BC a heavy increase of mean body height of about 10 cm can be observed in North Western Europe for females and to a lesser extent for males as well, whereas the body height elsewhere rather seems to

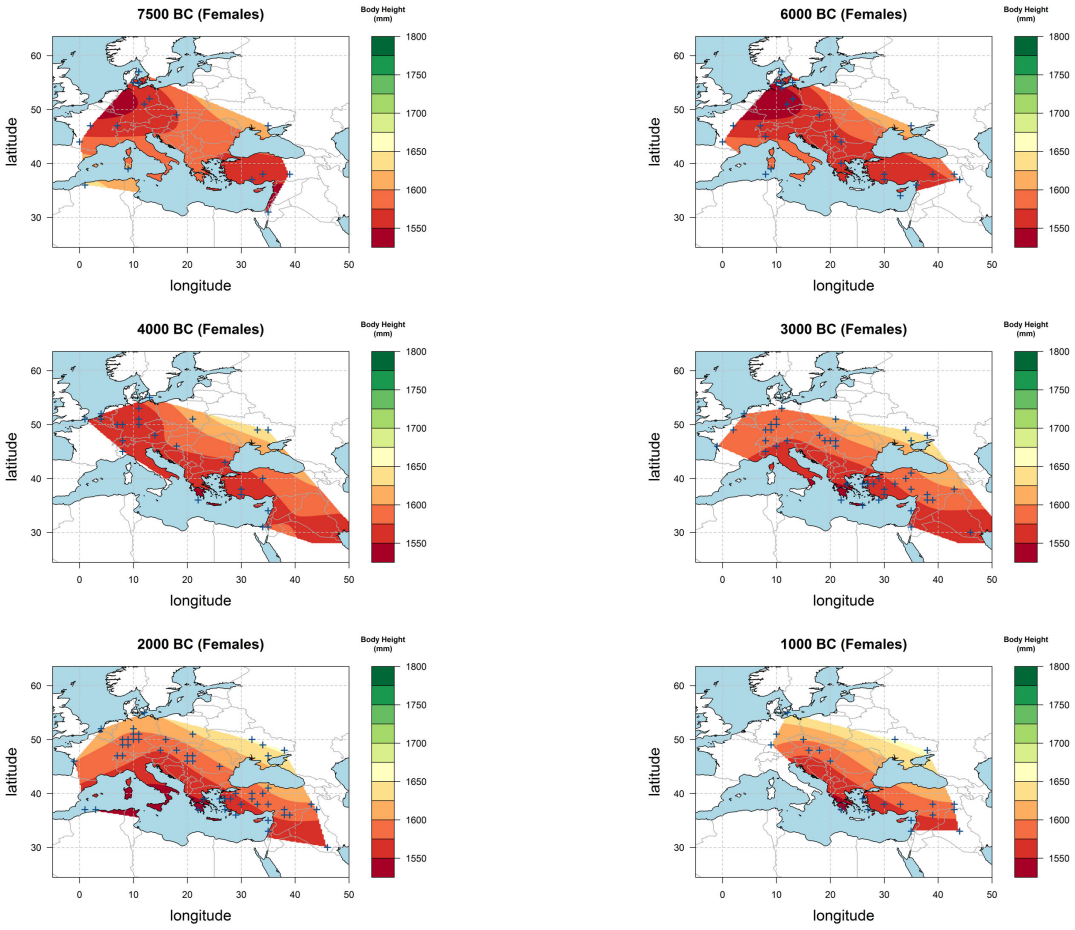

Figure 4.5: Estimated spatial mean body height in millimeters for females for different points in time (Model 4)

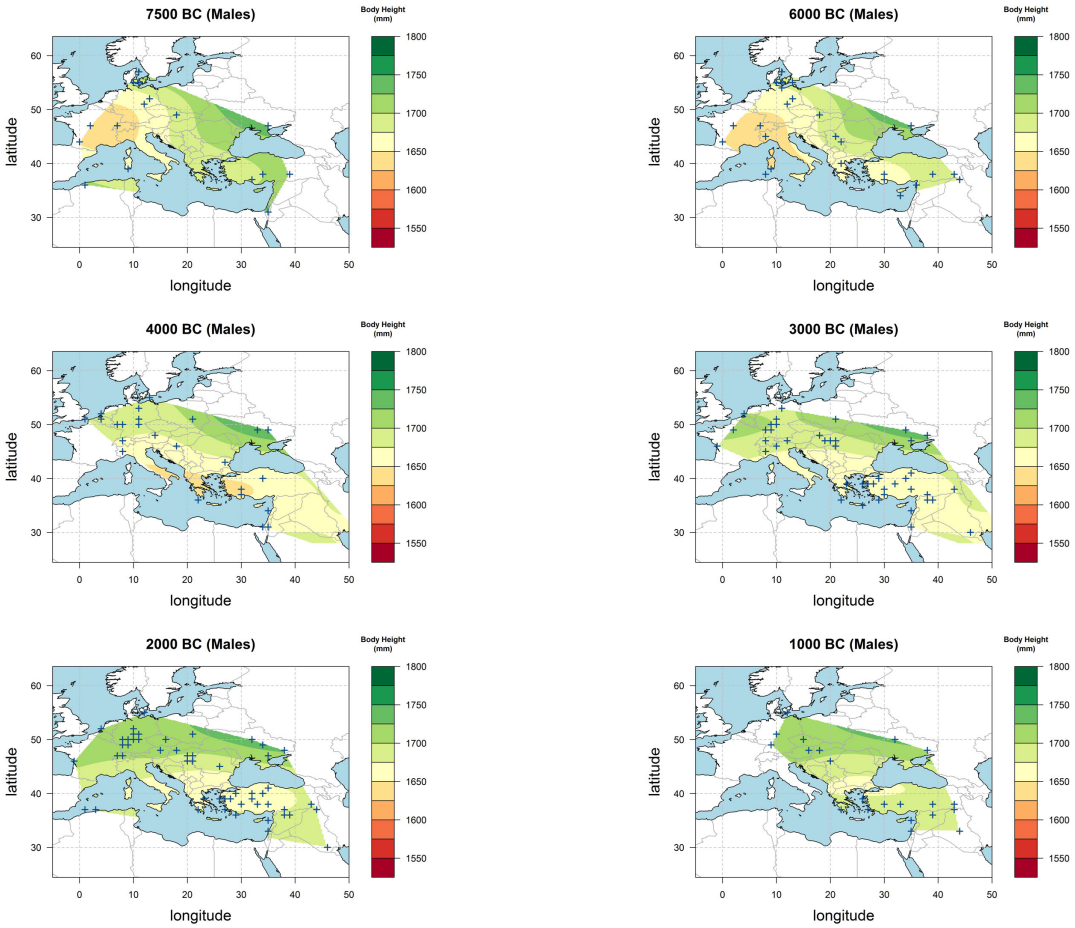

Figure 4.6: Estimated spatial mean body height in millimeters for males for different points in time (Model 4)

stay stagnant. Thus, an alternative illustration is given in Figure 4.8 for eight different locations shown in Figure 4.7. Most curves exhibit a U-shaped development pattern

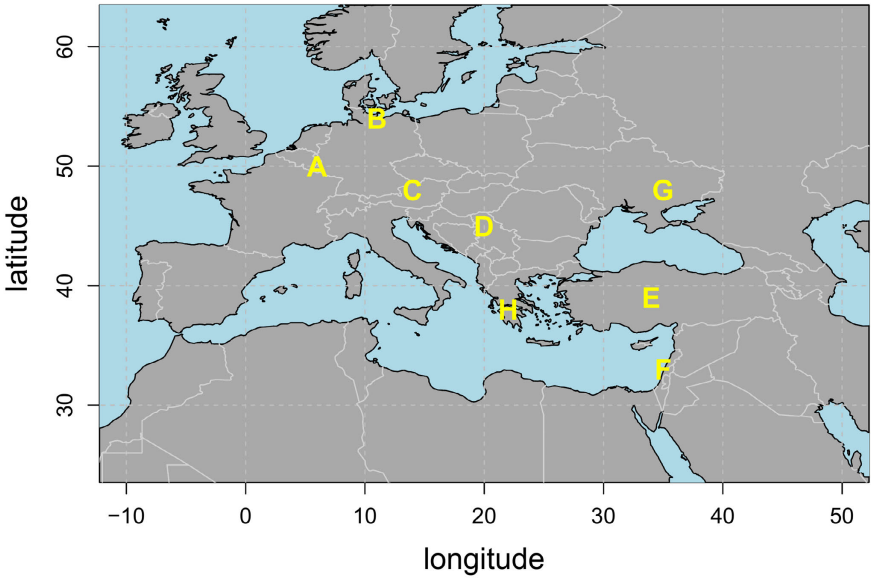

Figure 4.7: Positions of chosen locations for posterior inference

of body height in the considered time period with minima lying generally between 7000 to 4000 BC. The body height pattern at the end of the considered time span is quite similar to the one observed today (see e.g. Rosenstock et al. 2015 for a map of mean body heights by country). However, one should be careful in interpreting the results regarding the rather broad credible intervals. Additionally, the reader may ask what is gained by using advanced statistical methods and if the maps and other results from the complex Model 4 substantially differ from the results from simple Model 1. As it turned out the estimated spatio-temporal trends can differ quite considerably. Generally, the estimated trend functions for males and females were slightly more wiggly, which also explains why the effective number of parameters is higher for Model 1 than for Model 4. Moreover, they sometimes did even cross each other, which is somewhat unrealistic. Figure 4.9 shows this behaviour exemplary for two locations.

4.5 Discussion

Motivated by the anthropological long bone dataset a fully Bayesian additive mixed errors-in-variables model, allowing for both Berkson-type mismeasurement and misclassification, is proposed in this work. Accounting for measurement error leads to different and somewhat more reasonable results. Though the number of skeletons and settlements are not very high for a non-additive 3-dimensional smoother, this approach revealed some interesting spatio-temporal trends in body height or standard of living

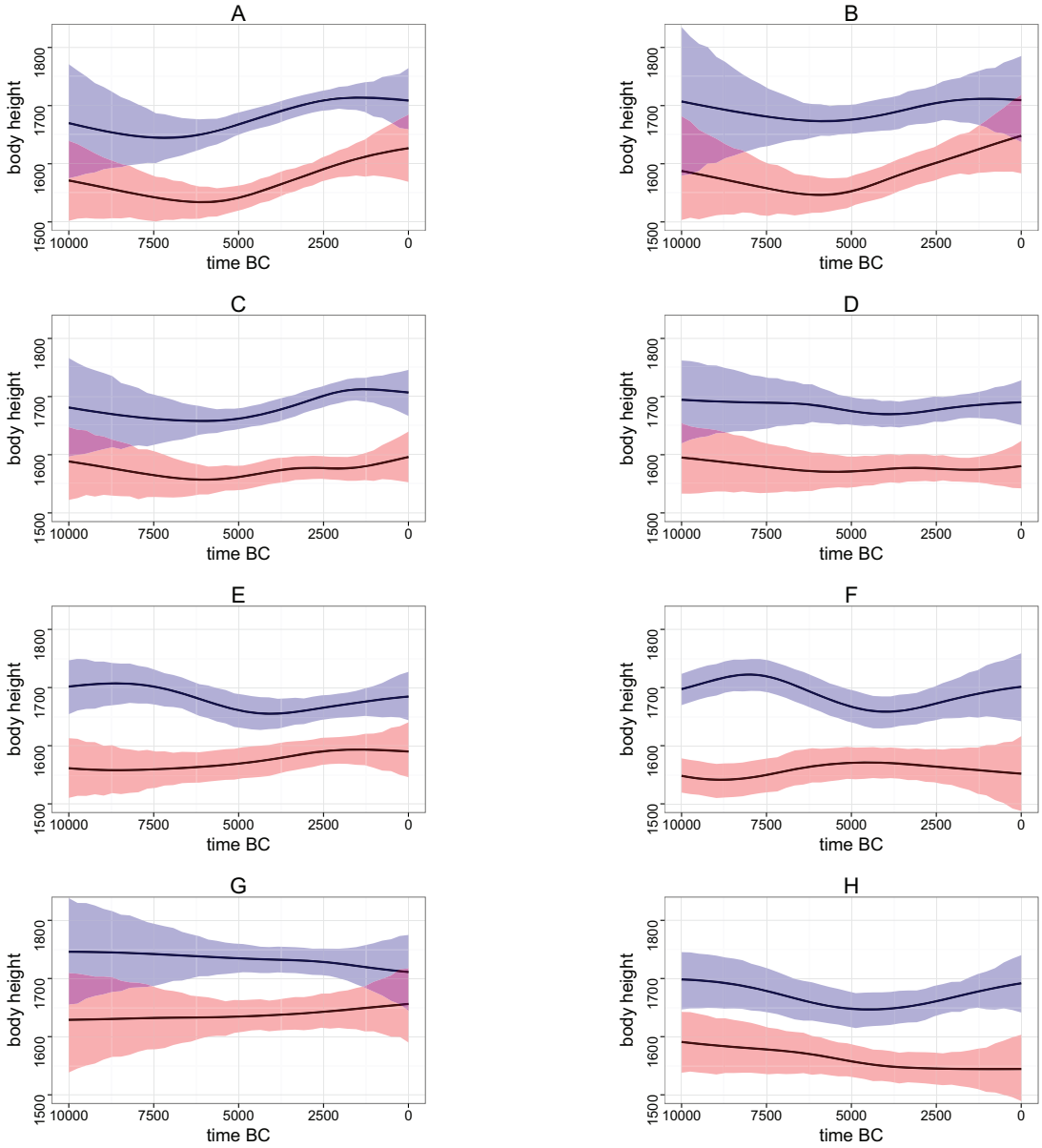

Figure 4.8: Estimated mean body height in millimeters through time for eight different locations (A-H) with 90% pointwise credible intervals for males (upper blue curve) and females (lower red curve) (Model 4)

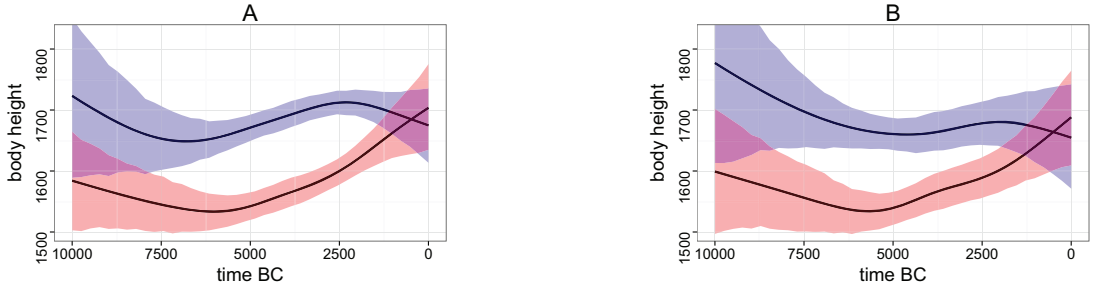

Figure 4.9: Estimated mean body height in millimeters through time for locations A and B with 90% pointwise credible intervals for males (upper blue curve) and females (lower red curve) (Model 1)

in the long-term perspective. The model results suggest that no negative effects on body height are notable during the primary neolithization in the Near East ca. 10000 BC, whereas the contemporary onset of the European Mesolithic and the spread of the Neolithic into Anatolia and temperate Europe by way of migration or acculturation in the 8th to 6th millennia (see Schier 2009 for a comprehensive map) witnessed a decrease in living standard. As discussed in more detail elsewhere (Rosenstock, 2014; Scheibner, 2015) this could be possibly explained by the diet shift leading to a lower protein consumption of the agricultural societies in the Neolithic Age. The subsequent body height recovery from ca. 5000 or 4000 BC onwards coincides with the neolithization of Northern Europe, the stabilization of the Neolithic way of life elsewhere and certain innovations such as cattle traction apt to reduce work load during the Later Neolithic, Chalcolithic and Bronze Ages. Moreover, the emergence of lactase-persistent human genotypes (Burger et al., 2007; Krüttli et al., 2014) and therefore the intake of unfermented animal milk with possible implications with growth factors has to be considered for Northwestern Europe during this time (Rosenstock, 2014). However, one must be very careful not to overemphasize the findings with respect to the rather low sample size and the resulting estimation uncertainty. Furthermore, it should be noted that a significant height increase or decrease can also be attributable to selection bias in the sample and confounding variables like climate changes or the burial practices/customs, because some cultures may have only buried individuals with a high social status. As a side result the data do not indicate that potentially biased sex determination of the experts is responsible for the mystery of low percentage of female individuals among prehistoric human skeletal remains.

Data correction and collection is an ongoing process of the project and we estimate that at least 10000 more skeletons can be added by literature research in the future. The LiVES project plans to make the data publicly available in 2018. As an extension of the model it is planned to incorporate additional variables such as domestication of animals or dairy farming to provide further insights about the association between cultural developments and living standards. Also non-uniform measurement errors for time of birth are conceivable and straightforward to include into the model. This is relevant because it is often not possible to give a single time period rather than multiple suggestions with different probabilities. This happens when the associated finds in the graves, which help to date the skeletons, may be assigned to different periods in time. The data are not available in such detail yet, but it is planned to incorporate this information later into the database.

Acknowledgements

This work originated from the LiVES project, which is funded by the Emmy-Noether-Program of the German Research Foundation (DFG). The author gratefully acknowledges Eva Rosenstock, the principal investigator of the LiVES project, for access to the data collection and giving her advice on archaeology. Moreover, thanks are due to Alisa Hujić for her counselation in anthropological issues. The author is also grateful to the editor and referees for helpful comments which led to an improved manuscript.

Kapitel 5

Reconstruction of Body Height from Long Bones for Prehistoric Individuals: New Methodological Concepts

5.1 Introduction

Among the variety of formulas available for the reconstruction of stature from long bone lengths (for an overview of all available formulas see e.g. Rösing 1988; Wurm 1985; Formicola 1993; Reichelt et al. 2003), those by Pearson (1899), Breitinger (1937)/Bach (1965) or Trotter and Gleser 1952 are most often used in Paleoanthropology, Prehistoric and Historic Anthropology. They all include the lengths of the humerus and radius as well as the femur and the tibia, and some additionally use the lengths of ulna and fibula (e.g. Telkkä 1950; Trotter and Gleser 1952, 1958; Bhavna and Nath 2009; Prasad et al. 2012), because it has been demonstrated that ulna and fibula have some minor importance for height development (Kurth et al., 1954; Siegmund, 2012). Correlations between individual long bone lengths and the corresponding stature are always given, and some authors additionally provide formulas for combinations of long bones, such as femur and humerus (e.g. Trotter and Gleser 1952, 1958; Olivier et al. 1978).

Consequently, when aiming to estimate the stature of skeletal finds, anthropologists have to carefully select not only which formula set they wish to use, but also which measurements or even combinations of measurements to include in the reconstruction. Different formula sets, i.e. the population-specific publications beginning with Pearson (1899), as well as different individual formulas, i.e. the formulas within a set which are specific to a certain long bone or their combinations, can yield vastly different results (see the example given in the results section). In individual cases of single skeletons, these decisions are usually determined by some degree of convention when it comes to the selection of the formula set, such as the Breitinger/Bach formulas being the preferred formula set for skeletons dating to the late Antiquity or early Mid-

dle Ages of Central Europe (Siegmund, 2010; Haberstroh and Harbeck, 2013). Also, the selection of long bone lengths to be included depends on the state of preservation of the skeleton: often a set is chosen that includes as many measurements as possible. Dealing with a population in which various combinations of preserved bones are possible, however, can be tricky, since there is no agreement on whether body heights calculated e.g. from upper limbs can really be compared with body heights calculated from lower limbs or even upper and lower limbs. In this methodological grey area, researchers have either resorted to choose randomly and accept that the final results do not represent real conditions in the past, regarding calculated body heights only as body heights a prehistoric individual would have reached if it had belonged to the actual reference population (Röhrer-Ertl, 1978; Rösing, 1988). Other authors, however, who are not content with such an abstract concept which represents – at best– only a mean of establishing comparability among populations, give recommendations based on methodological reasoning (e.g. Rösing 1988; Wurm 1986), population fitting (e.g. Rösing 1988; Formicola 1993) or mere compliance (Siegmund, 2012).

As especially the latter point does not hold up to empiric principles, we alternatively introduce two statistical concepts – already well-known in other fields – to the problem of stature estimation. First we propose a concept called "model averaging" applied to the set of formulas for different long bones. Computing all feasible reconstruction formulas according to the preservation of the individual skeleton has been proposed before by e.g. Pearson (1899), Telkkä (1950) and Sjøvold (1990), but their approach of simply averaging the resulting values is not ideal. Instead, we provide a statistically optimal way to weight these formulas appropriately by the so-called AIC-weights. This method is suitable for researchers who wish to utilize a formula set computed on a specific reference population. Second, to harmonize the published formulas from different research studies we introduce meta-analysis to stature estimation to provide a combined effect. As a result, new universal formulas for stature estimation are derived. A potential application for these methods is the LiVES ('Lebensbedingungen und biologischer Lebensstandard in der Vorgeschichte Europas und Südwestasiens') project (Rosenstock, 2014), which aims to model the body height in prehistory by collecting data from thousands of prehistoric skeletons from the available literature. Both approaches proposed here are illustrated with a practical example: a small subsample of a prehistoric population.

5.2 Statistical Methods for Stature Estimation from Long Bones

The goal in stature estimation from long bones is to get the best possible prediction for the body height of skeletal individuals with unknown body height. Since Rollet Rollet (1888) - following earlier attempts by Orfila and Lesueur (1831) and Topinard et al. (1885) - started to tackle the problem with a simple tabulation of long bone lengths and corresponding body heights derived from anatomy corpses from Lyon,

several statistical solutions have been proposed (Lorke et al., 1953).Manouvrier (1892) made a first attempt to formalize the relationship between the two figures and expressed stature (S) as a product of a long bone length, in this case the physiological length of the femur, later to be codified as F2 by Martin (1928), and a factor:

$$S_{F1} \;=\; \beta_1 F2$$

However, Manouvrier realized that one factor is not enough to describe the full range of relationships, so he defined a number of long bone length ranges and corresponding equations. Pearson (1899) was the first to understand that by introducing an intercept β_0, a linear equation

$$S_{F1} \;=\; \beta_0 + \beta_1 F1$$

is better suited for describing the relationship between long bone length – he now used the maximum femur length later to be defined as F1 Martin (1928) – and stature. After Pearson's notion that his linear regression formula set is valid not only for modern French material - but for all prehistoric, historic and modern populations - was questioned by the very different formula set for Chinese material published by Stevenson (1929), a number of structurally identical formulas have been published in the following century. They were based on different reference populations: Breitinger (1937) and Bach (1965), for instance, measured a sample of German sportsmen and female students and patients, Trotter and Gleser (1952) used data from corpses of white and black American soldiers and women from an anatomical collection, and (Olivier et al., 1978) measured a sample of deported young males, mostly French, from World War II and a sample of French women. With (reduced) major axis (Sjøvold, 1990) and piecewise linear regression (Duyar and Pelin, 2003), new methods were only applied in recent decades.

From a statistical point of view linear regression (also called ordinary least squares ? OLS) is preferable over the major axis method when performing prediction as Warton et al. (2006) clearly point out. The reduced major method (RMA) line describes the best bivariate fit between long bone and body height: the long bone lengths, after all, are a given fact in prehistoric and historic anthropology. Several authors like Formicola and Franceschi (1996) or Ruff et al. (2012) argue that very tall or very small individuals are better predicted with the reduced major axis (RMA) method (or 'line of organic correlation'). This is certainly true, but the overall prediction performance is overcompensated by a far worse prediction for individuals of average body height and very large or small long bone dimensions. This effect is very easy to verify by fitting the regression lines in a scatter plot or in a simulation. We performed such a simulation to illustrate this behavior. The true regression line was chosen to $y_i = 700 + 3x_i + \epsilon_i$with $i = 1, .., n$; where 100 samples of x_i and ϵ_i were drawn from Gaussian distributions, whose parameters were chosen in compliance with stature estimation: $x_i \sim N(300, 20)$ and $\epsilon_i \sim N(0, 50)$. Figure 5.1 shows one simulation run and illustrates in which cases OLS or RMA exhibit a superior prediction. Overall, OLS beats RMA by about 7% in

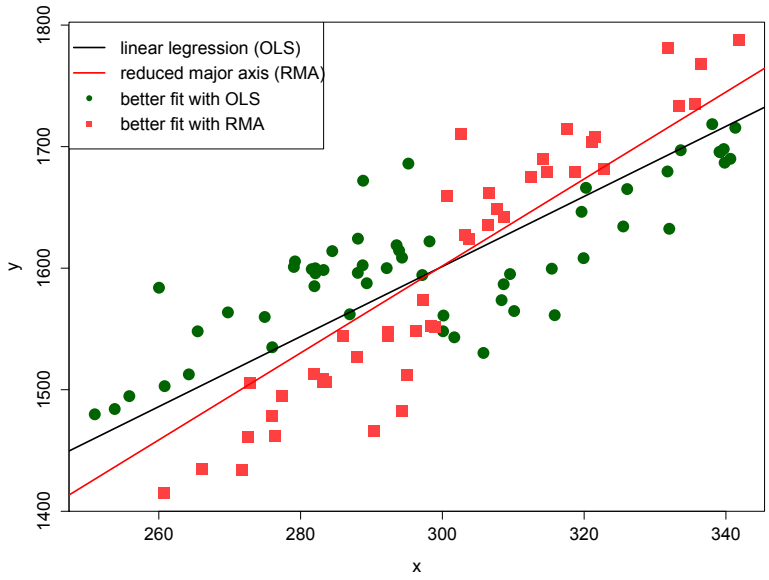

Figure 5.1: Comparison of linear regression (OLS) and reduced major axis (RMA) methods in a simulated example

terms of mean square error when repeating the simulation a large number of times. An extensive study conducted by Konigsberg et al. (1998) confirms this finding: as long as the reference sample is roughly representative for the individual case, OLS is the preferable method for stature estimation. The piecewise linear regression method is also not considered favorable by the authors, as it produces discontinuities in the estimates. Moreover, Breitinger (1937) and Trotter and Gleser (1952, 1958) empirically explored the shape of the association between long bone length and body height and could not find noteworthy deviations from linearity. Thus, the authors consider linear regression as the preferable method for stature estimation.

5.3 Weighting Long Bone Formulas via AIC

The *AIC* (Akaike, 1974) is a criterion for the selection of the best fitting model out of a variety of models for a given set of data. It is an estimate of the Kullback-Leibler divergence, a measure for the difference between two probability distributions. The AIC is the criterion of choice when the goal is to optimize prediction performance (Burnham and Anderson, 2004). It is defined as:

$$AIC = -2logL(\hat{\boldsymbol{\theta}}) + 2p$$

where p denotes the number of parameters in the model, i.e. in our case the number of long bones used in the formula + 1 (for the intercept β_0), and $\hat{\theta}$ the model parameter estimates. The second term penalizes the model complexity; hence following the parsimony principle or Ockham's razor, simpler models are preferred when their explanation

or prediction performance is equal. The specific AIC value has no direct interpretation but models with lower values indicate a better fit. For linear regression models the AIC has the following explicit formula:

$$AIC = n \ log(\hat{\sigma^*}^2_\epsilon) + 2p$$

$\hat{\sigma^*}^2_\epsilon = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2$ is the residual variance of the regression model. [1] The upper expression is advantageous for our purpose: with a few exceptions – Rollet 1888 underlying Pearson 1899 – neither the original data with which the reconstruction formulas were created nor the AIC is available. The residual variance or standard deviation, however, has always been published and enables us to compute the AIC for existing sets of reconstruction formulas in literature. As we usually have incomplete skeletons, however, the long bone formulas of a particular publication were estimated on a different sample size for each long bone. Therefore, we propose to set n in the upper AIC formula to the minimum sample size of all fitted models as a simple workaround. There arise two natural model selection strategies (Burnham and Anderson, 2004):

There arise two natural model selection strategies:

- Strategy 1: Select the model (long bone formula) with the lowest AIC.
- Strategy 2: Model-averaging: If several models have a comparable good prediction power or AIC it might be better to compute an ensemble of all models. One can do this by computing 'probabilities' from the AIC's that a certain model has the best prediction power and weight by them. Let $\boldsymbol{M}$ be the set of models and $\Delta AIC(M_i) = AIC(M_i) - \min_{m \in M}(AIC(m))$

$$w_i(M_i) = \frac{\exp(-1/2\Delta AIC(M_i))}{\sum_{k=1}^{K}\exp(-1/2\Delta AIC(M_i))}$$

While the first strategy allows us to identify the best long bone formula (for a single long bone or combinations of long bones) fitted on a sample, we are able to make use of all available data of a skeleton to predict its stature using the second strategy by optimally weighting the single predictions.

5.4 Meta-Analysis for Stature Estimation Formula Sets

Wurm (1986) states that searching for a universal formula applicable for every geographic region, time period, social strata et cetera may not be appropriate for stature estimation due to different body proportions. He recommends developing formulas for every thinkable population but warns of an outsized variety of formulas at the same time. The authors think that meta-analysis can provide an elegant compromise between these extremes. Meta-analysis is a well-established statistical concept usually employed in psychology Schmidt (1992) or clinical trials (DerSimonian and Laird, 1986) to assess

[1] This is the estimator without bias correction. To compute it from the usual estimator $\hat{\sigma}^2_\epsilon$, one has to premultiply $\hat{\sigma}^2_\epsilon$ by $\frac{n-p}{n}$. However, this difference is negligible for small values of $\frac{p}{n}$ as occurring in stature estimation.

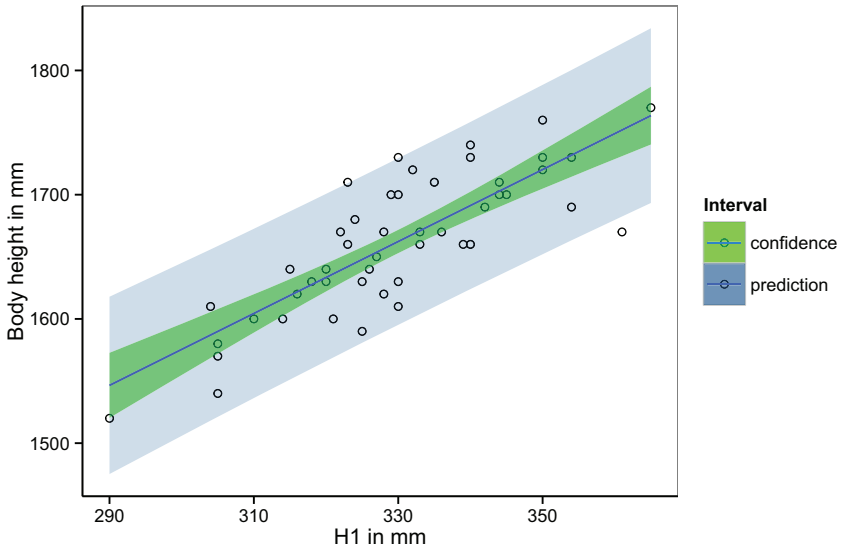

Figure 5.2: 95% confidence and prediction bands for humerus formula of Pearson (1899), males

the effect size of a certain treatment. It is used to aggregate evidence from multiple related studies (Cooper et al., 2009) by synopsis to compute a combined effect. Relying on a single study or – in our case – formula set in stature estimation as currently practiced by e.g. selecting the formula set with the best population fit, is not the best choice. One reason is the large estimation uncertainty induced by a small sample size: many stature estimation formulas are estimated on a sample smaller than 50 individuals, which leads to rather broad confidence intervals for the parameters of the linear regression formula. Figure 5.2 shows the confidence and prediction bands for the humerus (H1) formula of Pearson (1899) for males. The confidence bands include the uncertainty on the regression parameters, whereas the prediction bands additionally include the residual error ϵ. For humerus lengths far away from the sample mean, the 95% confidence bands (of the regression line) can be as large as 50 mm. That would mean that by the law of total variance the actual standard error of prediction would be about 35 mm compared to the residual standard error of 33 mm. Moreover, our estimates are certainly biased as the population sample, from which the regression line is estimated, is clearly different from the population we like to predict for. Also, the methods used to determine body height are subject to discussion and not perfectly reliable (see below). By combining formula sets from different studies we might be able to improve the prediction and get a realistic measure of the true estimation uncertainty which is underestimated when using a single formula set.

In general, one can differentiate between methodological differences and sample differences of the conducted studies. Some of the more frequent problems include the use of different long bone measures, as various different lengths measurements, such as in the case of the femur the maximum length (F1) or the natural or oblique length (F2), or in the case of the tibia the lateral condyle-malleolar length (T1), the spino-

malleolar length (T1a) and the medial condyle-malleolar length (T1b) etc. exist and – despite being codified by Martin (1928) – are often confused (in e.g. Breitinger 1937; for an overview see Siegmund 2010). , using long bones of the left or right side of the body or an average of both sides, reconstruction of bone length from bone fragments etc. Stature data can be gained as living stature (varying with posture, time of the day and age of the subject), corpse length (hanging or lying), or skeletal length (here, often various cartilage and soft tissues correction factors are applied). Differences in the sample can concern the population or region of settlement, economical structural change with altered dietary or physical exercise habits or certain demographic groups or social strata, respectively.

Currently, the approach to find a formula can be denoted as 'matching', i.e. one looks for a stature estimation formula which is estimated on a sample population which is as close as possible to the individuals we like to estimate the stature for and suffices certain quality criteria. Rösing (1988), for example, tried a categorization for some cases of applications and listed some quality criteria (e.g. sample size $n \geq 50$, separation by sex, 'normal' social structure or no secular acceleration). Assessing 'closeness' to prehistoric populations, however, is highly speculative. As first palaeogenetic insights suggest, the genetic variability between populations might have been much higher than today due to lower population density and lower mobility and therefore the dispersion in body proportions could be even larger than in recent populations, especially in certain time periods such as the glacial peaks and the early Holocene (Haak et al., 2015; Mathieson et al., 2015).

Thus, as long as both a population specific formula as well as a universal formula are prone to bias, a universal formula set derived from several formula sets is probably preferable as it is the only way to correctly assess the uncertainty of the estimation and is likely to have a smaller bias on an average prehistoric population. There are no particular helpful validation studies either. Formicola and Franceschi (1996) and Schmidt et al. (2007) evaluated some stature reconstruction formulas using prehistoric individuals reconstructed by the Fully (1956) method, but the sample sizes used for estimation were extremely low ($n < 30$). Additionally, even the Fully technique inherits a considerable error when estimating the living stature (Raxter et al., 2006). Thus, the authors propose a different strategy, i.e. to collect suitable publications (sufficing some quality standards) on regression formulas and perform a statistical meta-analysis to obtain a combined formula when trying to estimate the stature of certain (prehistoric) individuals.

Two variants of meta-analysis can be distinguished: the fixed effects and the random effects meta-analysis (Cooper et al., 2009). The fixed effect variant assumes that the underlying true effects and regression parameters respectively are identical in each study. The combined effect is then just the weighted average of the single effect where the weights are proportional to sample size. This is a very unrealistic assumption for our problem due to the differences in methodology and in the examined sample populations. The random effects meta-analysis assumes that there is a common effect (i.e. common

regression intercept and slope in stature estimation) which differs randomly from study to study. Formally, the combined regression slope can be expressed as:

$$\beta_{1i} = \beta_1 + \nu_i + \epsilon_i \; ; \; i = 1, \ldots, m$$

, where β_{1i} is the study-specific regression coefficient; β_1 is the common regression coefficient, $\nu_i \sim N(0, \sigma_\nu^2)$ denotes the deviation between the studies and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ is the error within the study. We are therefore decomposing the observed variance into its two components, the within-studies variation (expressed by σ_ϵ^2) and the between-studies variation (expressed by σ_ν^2). A meta-analysis is performed in the results section to compute common regression coefficients β_0 and β_1 using four popular formulas from literature on stature estimation.

5.5 Results

5.5.1 AIC-weighting for the formula set of Pearson (1899)

We computed the AIC for all formulas of Pearson (1899) using the data of Rollet (1888) for males. These can be found in Table 5.1. Terms like F1 + T1b denote simple linear regression with the sum of F1 and T1b as regressor while terms like F1, T1b denote a multiple linear regression with two regressors. The abbreviations correspond to the definitions of Martin (1928); see also Bräuer (1996).

Table 5.1: Summary AIC with data from Pearson (1899); males, $n = 41 - 50$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1	33.0	104.5	4.6	3.2
T1b	35.7	110.7	11.0	0.13
F1	32.7	103.7	3.8	4.5
R1	41.0	122.3	22.4	4.3E-6
Fib1	32.6	103.5	3.6	5.3
F1+T1b	31.2	99.9	0	31.8
H1+R1	33.9	106.7	6.8	1.1
F1,T1b	31.5	102.6	2.7	7.8
H1,R1	31.0	101.3	1.4	15.0
H1,F1	30.6	100.3	0.4	26.0
H1,F1,T1b,R1	30.3	103.5	3.6	5.3

Among the formulas with only a single long bone, the one using the fibula (Fib1) is estimated to have the best predictive power followed by the formula using the femur (F1) and humerus (H1). It is equally well as the formula using all four long bones considered by Pearson (1899) for long bone estimation due to the penalization term in the AIC. If the second strategy shall be applied, the weighted average of stature estimates by all formulas according to the AIC-weights (last column in Table 5.1) has to be computed. Prehistoric skeletons are very often incomplete and the model averaging strategy is still applicable for incomplete skeletons. We give an application

example at the end of the results sections for illustration. Admittedly, this approach requires more effort than just arbitrarily choosing the long bones formulas to use, but a simple Excel or R script will do the work and is available from the corresponding author. Additional AIC tables for females as well as for the formulas of Trotter and Gleser (1952) and Olivier et al. (1978) can be found in the appendix. One has to note that for increasing sample size a single formula is usually dominant and applying strategy 2 does not give a big advantage in this case. Therefore, the proposed strategy is rather advantageous for improving stature estimates using formulas computed on small to mid-sized reference populations.

5.5.2 New universal formulas generated by combining existing formulas using statistical meta-analysis

First, we perform an exemplary meta-analysis for stature reconstruction by the maximum humerus length H1 in females. We used four formulas commonly used for stature reconstruction of prehistoric individuals from Europe and the Near East, i.e. the formulas of Pearson (1899), Trotter and Gleser (1952)[2], Breitinger (1937) and Bach (1965) and Olivier et al. (1978). The metafor package (Viechtbauer, 2010) for the statistical Software R was utilized for the computation. Summary statistics for humerus (female) can be found in Table 5.2.

Table 5.2: Linear regression formulas for stature estimation by H1, females

Author, year	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}_\epsilon$	n
Pearson, 1899	703	2.75	35	50
Trotter/Gleser, 1952	580	3.36	45	63
Bach, 1965	984	2.12	39	500
Olivier et al., 1978	623	3.09	36	140

Bach (1965) might be considered as an outlier, but the random effects meta-analysis is relatively robust to it. Additionally, even though it is estimated on a rather specific population (sport students), we have no hint that populations in prehistory are less extreme. Figure 5.3 shows the resulting regression line together with prediction bands and the four original formulas.[3]

The resulting common regression coefficient estimates are $\hat{\beta}_0 = 732$ and $\hat{\beta}_1 = 2.80$. Cochran's Q-Test was significant ($p < 0.001$) for both coefficients. Thus, we have to reject the null hypothesis of equal regression coefficients, confirming that a random effects meta-analysis instead of a fixed effects one is the appropriate choice. The standard error of the meta-analysis regression line is now 22 mm (for H1 = 300 mm) which accounts for the methodological as well as sample population differences. It has to be added to the combined residual standard error of 38 mm resulting in a total prediction

[2]Based on the Terry collection.

[3]The standard errors of the regression coefficients are necessary for performing meta-analysis but are commonly not available. We approximated them by rescaling the standard errors of the Pearson (1899) data by the sample size which should work reasonably well.

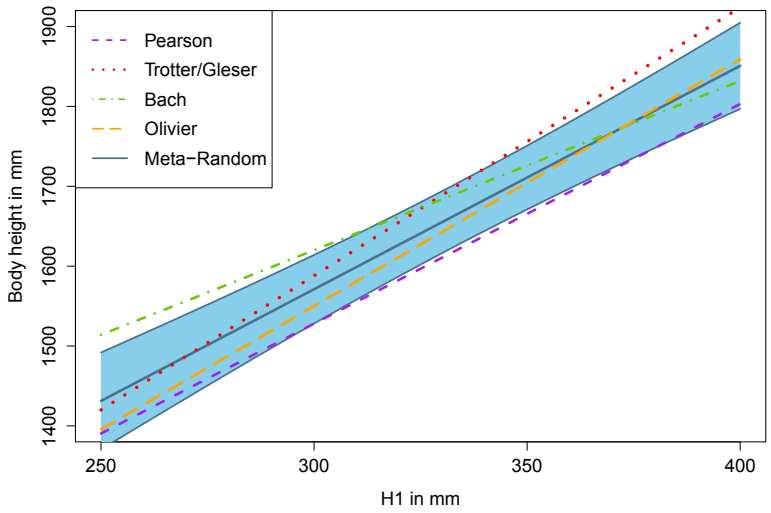

Figure 5.3: Stature estimation regression lines for H1, females: Pearson (1899), Trotter/Gleser (1952), Bach (1965), Olivier et al. (1978) and random effects meta-analysis line with 95 % prediction bands.

standard error of 44 mm. This is a considerably higher but more realistic estimate of the prediction uncertainty compared to the usage of a single formula from literature. Tables 3 (females) and 4 (males) show the results of the meta-analysis for six different long bones for males and females using the same formulas.[4]

Table 5.3: Results for meta-analysis (intercept $\hat{\beta}_0$, slope $\hat{\beta}_1$, residual standard error $\hat{\sigma}_\epsilon$ and prediction standard error for a typical long bone length in parenthesis) for diffferent long bones (females). *: Less than 4 formulas available

Long bone	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}_\epsilon$	Typical pred. error
H1	732.3	2.80	38.5	44.3 (300 mm)
F1	779.4	1.96	36.6	55.6 (430 mm)
T1	780.7	2.32	37.3	41.7 (340 mm)
R1	807.9	3.56	42.4	73.2 (220 mm)
Fib1*	672.6	2.66	35.8	40.8 (340 mm)
U1*	592.9	4.18	47.5	51.7 (240 mm)

Surprisingly, the formulas for tibia (T1) had the lowest typical prediction error for both males and females (leaving aside Fib1, for which only two instead of four formulas were available). This phenomenon can be assigned to the fact that the different formulas using tibia are relatively similar whereas they vary more widely for the radius, for example. The authors recommend estimating the stature by using the formula with the typical lowest prediction error of the available long bones.

[4]T2, F2 and R1b were converted to T1, F1 and R1 by the mean differences given in Rösing (1988). Formulas for the left and right body side were averaged.

Table 5.4: Results for meta-analysis (intercept $\hat{\beta}_0$, slope $\hat{\beta}_1$, residual standard error $\hat{\sigma}_\epsilon$ and prediction standard error for a typical long bone length in parenthesis) for diffferent long bones (males). *: Less than 4 formulas available

Long bone	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}_\epsilon$	Typical pred. error
H1	706.4	2.97	42.6	55.6 (330 mm)
F1	715.5	2.13	38.3	54.7 (460 mm)
T1	779.6	2.44	39.6	45.5 (360 mm)
R1	796.1	3.62	46.3	72.8 (240 mm)
Fib1*	708.2	2.67	36.8	43.8 (370 mm)
U1*	687.1	3.86	47.4	53.4 (260 mm)

5.5.3 Practical example

We illustrate the proposed methods for stature estimation on a small sample of prehistoric individuals from Stuttgart-Mühlhausen. The cemetery of Stuttgart-Mühlhausen ('Viesenhäuser Hof') consists of two areas dating to the oldest and older Linear Pottery Culture (LBK) ca. 5700-5400 cal BC (Area II) and middle to younger LBK ca. 5400-4900 cal BC (Area I) (Price et al., 2003; Kurz, 1993; Hujić, 2016), hence belonging to the first Central Europeans practicing plant cultivation and animal husbandry. Altogether 24 adult individuals, 12 (five females and seven males) from Area II and 12 (six females and six males) from Area I were chosen for this example. Table 5.5 shows the long bone lengths F1, T1b, H1 and R1. Stature estimates of these individuals using the formula sets of Pearson (1899) as well as estimates using the proposed methods - denoted as the AIC-methods 'Strategy 1' and 'Strategy 2' and 'Meta-Analysis' - can be found in Table 5.6. One remarkable skeleton is found in grave II-27. It has rather short upper limbs compared to its lower ones resulting in stature estimates ranging from 1588 mm to 1701 mm depending on which formula is applied. The simple strategy 1 selects the formula using the sum of F1 and T1b as regressor because it has the lowest AIC. Strategy 2, favored here, however, gives a more balanced estimate of 1660 mm as it is the weighted mean of all individual long bone formulas (with weights corresponding to the AIC-weights computed in Section 5.3; cf. last column of Table 5.1). The universal formula resulting from the meta-analysis (using T1 as it has the lowest prediction error of the available long bones) yields 1715 mm. This larger estimate is not surprising because the formulas of Pearson (1899) give rather low estimates compared to other stature estimation formulas (Fig. 5.3). An Excel file for application and reproduction is available from the corresponding author.

Table 5.5: Long bone measurements from Stuttgart-Mühlhausen (in mm)

Area I					
Grave	**Sex**	**F1**	**Ti1b**	**H1**	**R1**
I-20	m	399	334	-	221
I-25	m	487	394	338	263
I-36	m	410	342	291	233
I-43	m	-	-	-	231
I-47	m	439	367	311	247
I-48	m	423	339	308	228
I-21	f	376	-	272	245
I-33	f	415	334	292	226
I-34	f	377	300	269	-
I-37	f	-	-	288	214
I-55	f	405	-	284	-
I-61	f	424	340	304	227
Area II					
II-21	m	-	-	-	219
II-27	m	471	381	305	233
II-44	m	-	357	-	-
II-78	m	457	374	320	238
II-83	m	-	378	335	253
II-107	m	465	403	329	264
II-111	m	450	380	-	246
II-42	f	-	-	275	197
II-43	f	-	-	-	209
II-45	f	415	347	-	-
II-58	f	-	-	251	-
II-133	f	387	-	272	-

Table 5.6: Stature estimates in mm for individuals from Stuttgart-Mühlhausen (formulas of Pearson (1899). The last three columns denote the stature estimates of the proposed methods: The *AIC*-methods 'Strategy 1' and 'Strategy 2' as introduced in Section 5.3 and 'Meta-Analysis' as introduced in Section 5.4

Area I														
Grave	**Sex**	**H1**	**T1b**	**F1**	**R1**	**F1+T1b**	**F1,T1b**	**H1+R1**	**H1,R1**	**H1,F1**	**H1,F1,T1b,R1**	**Strategy 1**	**Strategy 2**	**Meta-Analysis**
I-20	m	-	1580	1563	1582	1562	1562	-	-	-	-	1562	1562	1600
I-25	m	1685	1723	1729	1720	1734	1734	1708	1685	1712	1716	1734	1717	1747
I-36	m	1549	1599	1584	1621	1584	1584	1575	1549	1559	1563	1584	1569	1620
I-43	m	-	-	-	1615	-	-	-	-	-	-	1615	1615	1632
I-47	m	1606	1659	1638	1667	1647	1646	1634	1607	1620	1626	1647	1630	1681
I-48	m	1598	1592	1608	1605	1596	1597	1596	1595	1599	1595	1596	1597	1613
I-21	f	1464	-	1460	1631	-	-	1541	1477	1457	-	1457	1461	1494
I-33	f	1519	1533	1536	1568	1535	1535	1542	1523	1530	1518	1518	1528	1561
I-34	f	1456	1453	1462	-	1454	1454	-	-	1455	-	1454	1454	1483
I-37	f	1508	-	-	1528	-	-	1516	1509	-	-	1509	1509	1539
I-55	f	1497	-	1516	-	-	-	-	-	1508	-	1508	1511	1528
I-61	f	1552	1547	1553	1571	1552	1552	1564	1554	1554	1543	1543	1549	1575
Area II														
II-21	m	-	-	-	1576	-	-	-	-	-	-	1576	1576	1589
II-27	m	1589	1692	1699	1621	1700	1701	1599	1588	1644	1659	1700	1660	1715
II-44	m	-	1635	-	-	-	-	-	-	-	-	1635	1635	1657
II-78	m	1632	1675	1672	1638	1676	1676	1634	1630	1653	1660	1676	1659	1698
II-83	m	1676	1685	-	1687	-	-	1686	1675	-	-	1675	1676	1708
II-107	m	1659	1744	1687	1723	1719	1717	1694	1660	1675	1690	1719	1692	1769
II-111	m	-	1690	1659	1664	1675	1674	-	-	-	-	1675	1673	1713
II-42	f	1472	-	-	1471	-	-	1468	1471	-	-	1471	1471	1502
II-43	f	-	-	-	1511	-	-	-	-	-	-	1511	1511	1552
II-45	f	-	1564	1536	-	1550	1550	-	-	-	-	1550	1549	1592
II-58	f	1406	-	-	-	-	-	-	-	-	-	1406	1406	1435
II-133	f	1464	-	1481	-	-	-	-	-	1472	-	1472	1476	1494

5.6 Discussion

This work introduced the statistical concepts of Akaike model weighting and random effects meta-analysis to stature estimation. These methods – which are already well known in other fields – help to improve prediction and – possibly more importantly – assess the prediction uncertainty in more realistic way by accounting for methodological and sample population differences. The concept of Akaike model weighting can be used to obtain the best prediction performance using a set of formulas estimated on a single reference population. When it is unclear whether the individuals one likes to estimate the stature of and the reference population are reasonably similar, one might be better off to use the universal formulas derived by meta-analysis in Section 5.4. The reader may ask how to combine both presented methodological concepts. Unfortunately, combining both the AIC weighting and the meta-analysis is a complicated problem in statistics where only approximate solutions exist (it is related to the denominator degrees of freedom in mixed models). This will be a subject of future research.

5.7 Appendix

Table 5.7: Summary AIC with data from Pearson (1899); females, $n = 43 - 50$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1	34.8	113.7	10.3	0.2
T1b	34.4	112.8	9.3	0.3
F1	32.8	108.7	5.2	2.7
R1	40.9	127.6	24.2	2.0E-4
Fib1	33.5	110.5	7.0	1.1
F1+T1b	30.9	103.5	0.1	35.2
H1+R1	36.1	116.9	13.4	4.3E-2
F1,T1b	31.0	104.7	1.3	19.1
H1,R1	33.6	112.7	8.2	0.6
H1,F1	32.3	108.3	4.9	3.2
H1,F1,T1b,R1	30.2	103.5	0	37.3

Table 5.8: Summary AIC with data from Trotter and Gleser (1952); males (living stature), $n = 545 - 710$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1	40.5	1533.2	330.8	9.5E-73
T1b	33.0	1310.0	107.6	2.8E-24
F1	32.7	1300.0	97.6	4.1E-22
R1	43.2	1603.5	401.1	5.1E-88
U1	43.2	1603.5	401.1	5.1E-88
Fib1	32.9	1306.7	104.3	1.4E-23
F1+T1b	29.9	1205.4	0	64.4
H1,R1	38.8	1488.5	286.1	4.8E-63
H1,T1b	32.6	1298.7	196.3	1.5E-43
F1,T1b	29.9	1204.4	2.0	23.7
H1,F1,T1b	29.9	1206.4	4.0	8.7
H1,R1,F1	31.5	1263.3	60.9	3.8E-14
H1,R1,F1,T1b	29.9	1208.4	6.0	3.2

Table 5.9: Summary AIC with data from Trotter and Gleser (1952); females (living stature), $n = 63$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1	44.5	194.6	28.4	1.6E-5
T1b	36.6	170.0	3.8	3.7
F1	37.8	174.1	7.9	0.5
R1	42.4	188.5	22.4	3.5E-4
U1	43.0	190.4	24.2	1.5E-4
Fib1	35.7	166.9	0.7	17.9
F1+T1b	35.5	166.2	0	25.5
H1,R1	40.4	183.4	17.2	4.6E-3
H1,T1b	36.7	171.3	5.1	1.9
F1,T1b	35.5	167.1	0.9	15.8
H1,F1,T1b	35.1	166.6	0.4	20.2
H1,R1,F1	36.6	171.9	5.7	1.4
H1,R1,F1,T1b	35.1	167.5	1.4	12.7

Table 5.10: Summary AIC with data from Olivier et al. (1978); males (living stature), $n = 140$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1 (r)	39.8	393.3	71.6	2.1E-16
T1b (r)	35.8	363.7	41.9	5.9E-10
F2 (r)	35.6	362.1	40.4	1.3E-9
R1b (r)	40.4	397.5	75.8	2.6E-17
U1 (r)	44.8	426.5	104.7	1.4E-23
Fi1 (r)	34.5	353.3	31.6	1.0E-7
H1,R1b	36.8	371.4	50.6	7.6E-12
H1,U1	37.9	380.6	58.9	1.2E-13
H1,F2	33.4	341.9	20.1	3.2E-5
H1,T1b	32.7	339.3	17.6	0.12
H1,Fi1	32.1	334.1	12.4	1.5E-3
R1b,F2	32.9	341.0	19.3	4.9E-5
R1b,T1b	34.1	351.0	29.3	3.3E-7
R1b,Fi1	33.4	345.2	23.5	6.0E-6
U1,F2	33.4	345.2	23.5	6.0E-6
U1,T1b	34.4	353.5	31.8	9.6E-8
U1,Fi1	33.7	347.7	26.0	1.7E-6
F2,T1b	31.7	330.6	8.9	0.089
F2,Fi1	31.0	324.4	2.6	20.4
H1,F2,T1b	31.2	328.1	6.4	3.1
H1,F2,Fi1	30.5	321.7	0	75.4

Table 5.11: Summary AIC with data from Olivier et al. (1978); females (living stature), $n = 140$

Long bone	$\hat{\sigma}_\epsilon$	AIC	ΔAIC	w in %
H1 (l)	36.2	366.7	83.2	8.3E-17
R1b (l)	35.3	359.7	76.2	2.8E-15
U1 (l)	35.3	359.7	76.2	2.8E-15
F2 (l)	35.6	362.1	78.5	8.4E-16
T1b (l)	38.5	384.0	100.5	1.5E-20
H1,R1b	32.8	340.1	56.6	4.9E-11
H1,U1	32.3	335.8	52.3	4.2E-10
H1,F2	31.3	327.0	43.5	3.4E-8
R1b,F2	29.9	314.2	30.7	2.1E-5
R1b,T1b	33.9	349.4	65.9	4.8E-13
U1,F2	29.9	314.2	30.7	2.1E-5
U1,T1b	33.5	346.1	62.5	2.5E-14
F2,T1b	27.6	291.8	8.3	1.5
H1,F2,T1b	26.7	283.5	0	98.4

Literaturverzeichnis

Abbo, S., S. Lev-Yadun, M. Heun, and A. Gopher (2013). On the 'lost' crops of the neolithic near east. Journal of Experimental Botany 64(4), 815–822.

Abbo, S., Y. Saranga, Z. Peleg, Z. Kerem, S. Lev-Yadun, and A. Gopher (2009). Reconsidering domestication of legumes versus cereals in the ancient near east. The Quarterly Review of Biology 84(1), 29–50.

Abbo, S., D. Shtienberg, J. Lichtenzveig, S. Lev-Yadun, and A. Gopher (2003). The chickpea, summer cropping, and a new model for pulse domestication in the ancient near east. The Quarterly Review of Biology 78(4), 435–448.

Abdullah, A. Y., M. M. Muwalla, R. I. Qudsieh, and H. H. Titi (2010). Effect of bitter vetch (vicia ervilia) seeds as a replacement protein source of soybean meal on performance and carcass characteristics of finishing awassi lambs. Tropical Animal Health and Production 42(2), 293–300.

Acevedo-Garcia, D., K. A. Lochner, T. L. Osypuk, and S. V. Subramanian (2003). Future directions in residential segregation and health research: a multilevel approach. American Journal of Public Health 93(2), 215–21.

Acsadi, Georgeand Nemeskéri, J. (1970). History of human life span and mortality. Budapest: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, IEEE Transactions on 19(6), 716–723.

Alpaslan-Roodenberg, S. (2001). Newly Found Human Remains from Menteşe in the Yenişehir Plain. Anatolica 27, 1–14.

Alpaslan-Roodenberg, S. (2002). Preliminary report on the human remains from the Early Bronze Age Cemetery at Ilıpınar-Hacılartepe. Anatolica 28, 91–107.

Alpaslan-Roodenberg, S. and G. Maat (1999). Human Skeletons from Menteşe Höyük near Yenişehir. Anatolica 25, 37–52.

Ambadekar, N., S. Wahab, S. Zodpey, and D. Khandait (1999). Effect of child labour on growth of children. Public Health 113(6), 303–306.

Angel, J. (1941). The Babaköy Skeleton. Archiv für Orientforschung 13, 28–31.

Angel, J. (1984). Health as a crucial factor in the changes from hunting to developed farming in the Eastern Mediterranean. In M. Cohen and G. Armelagos (Eds.), Palaeopathology at the Origins of Agriculture. New York: Academic Press.

Angel, J. L. (1945). Skeletal material from Attica. Hesperia 14(4), 279–363.

Angel, J. L. (1951). Troy, the human remains. American Journal of Physical Anthropology 9(4), 478–479.

Angel, J. L. (1953). The human remains from Khirokitia. pp. 416–430. London/New York/Toronto.

Angel, J. L. (1970). Appendix-human skeletal remains at Karatas. American Journal of Archaeology 74(3), 253–259.

Angel, J. L. (1971). The People of Lerna: Analysis of a Prehistoric Aegean Population. Washington: Smithsonian Books.

Angel, J. L. (1973a). Early Neolithic people of Nea Nikomedeia. Die Anfänge des Neolithikums vom Orient bis Nordeuropa 3, 103–112.

Angel, J. L. (1973b). Human skeletons from grave circles at Mycenae. In G. Mylonas (Ed.), Die Schlachtgräber, pp. 211–232. Athen: Archaiologike Etaireia.

Anthony, R. and L. Manouvrier (1907). Étude des ossements humains de la sépulture néolithique de montigny-esbly. Bulletins et Mémoires de la Société d'Anthropologie de Paris 8(1), 537–563.

Armelagos, G. J., A. H. Goodman, and K. H. Jacobs (1991). The origins of agriculture: Population growth during a period of declining health. Population and Environment 13(1), 9–22.

Armstrong, M. P., G. Rushton, and D. L. Zimmerman (1999). Geographically masking health data to preserve confidentiality. Statistics in Medicine 18(5), 497–525.

Asmus, G. (1939). Die vorgeschichtlichen rassischen Verhältnisse in Schleswig-Holstein und Mecklenburg, ein Beitrag zur Rassengeschichte des urgermanischen Raumes, Volume 4. Neumünster: K. Wachholtz.

Asmus, G. (1973). Mesolithische Menschenfunde aus Mittel-, Nord-und Osteuropa. In H. Schwabendissen (Ed.), Die Anfänge des Neolithikums vom Orient bis Nordeuropa. Teil Villa, Anthropologie, pp. 28–86. Wien: Böhlau Verlag.

Babiker, S., I. El Khider, and S. Shafie (1990). Chemical composition and quality attributes of goat meat and lamb. Meat Science 28(4), 273–277.

Bach, A. (1978). Neolithische Populationen im Mittelelbe-Saale-Gebiet. Zur Anthropologie des Neolithikums unter besonderer Berücksichtigung der Bandkeramiker, Volume 1. Weimar: Weimarer Monographien zur Ur-und Frühgeschichte.

Bach, A. and H. Bruchhaus (1988). Das mesolithische Skelett von Unseburg, Kr. Stassfurt. Jahresschrift für mitteldeutsche Vorgeschichte 71, 21–36.

Bach, H. (1965). Zur Berechnung der Körperhöhe aus den langen Gliedmaßenknochen weiblicher Skelette. Anthropologischer Anzeiger 29, 12–21.

Bar, H. Y. and D. R. Lillard (2012). Accounting for heaping in retrospectively reported event data–a mixture-model approach. Statistics in Medicine 31(27), 3347–3365.

Baten, J. and M. Blum (2014). Why are you tall while others are short? Agricultural production and other proximate determinants of global heights. European Review of Economic History 18(2), 144–165.

Battistin, E., R. Miniaci, and G. Weber (2003). What do we learn from recall consumption data? Journal of Human Resources 38(2), 354–385.

Becker, M. (1988). An analysis of the human skeletal remains from [the] necropolis at İkiztepe. In U. Alkım, A. H., and . Bilgi (Eds.), İkiztepe I. Turk Tarih Kurumu Basimevi, pp. 261–275. Ankara.

Beckett, P. R., W. W. Wong, and K. C. Copeland (1998). Developmental changes in the relationship between igf-i and body composition during puberty. Growth Hormone & IGF Research 8(4), 283–288.

Bedigian, D. and J. R. Harlan (1986). Evidence for cultivation of sesame in the ancient world. Economic botany 40(2), 137–154.

Behre, K.-E. (1992). The history of rye cultivation in Europe. Vegetation History and Archaeobotany 1(3), 141–156.

Belfer-Cohen, A., L. Schepartz, and B. Arensburg (1991). New biological data for the Natufian populations in Israel. pp. 1–424. Ann Arbor, Mich.: International Monographs in Prehistory, Archeological Series 1.

Bennicke, P. (1985). Palaeopathology of Danish Skeletons. A comparative study of demography, disease and injury. Ph. D. thesis, University Copenhagen.

Bennicke, P. (1987). Menneskeknogler frastenalderbo-pladsen på argusgrunden. beboernes udseende og helbredstilstand. fortidsminder og kulturhistorie. Antikvariske Studier 8, 94–104.

Bentley, A. (2004). Human mobility and the prehistoric spread of farming: isotope evidence from human skeletons. Archaeology International 8, 29–32.

Berkson, J. (1950). Are there two regressions? Journal of the American Statistical Association 45(250), 164–180.

Bermejo, J. E. H. and J. León (1994). Neglected crops: 1492 from a different perspective. Number 26. Rome: Food & Agriculture Org.

Bernstein, R. M. (2010). The big and small of it: How body size evolves. American Journal of Physical Anthropology 143, 46–62.

Berry, S., R. Carroll, and D. Ruppert (2002). Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association 97(457), 160–169.

Berthold, B., K. Alt, and B. Bramanti (2008). Die Totenhütte von Benzingerode: Archäologie und Anthropologie, Volume 7. Halle: Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt.

Bhavna, N. S. and S. Nath (2009). Use of lower limb measurements in reconstructing stature among Shia Muslims. Internet Journal of Biological Anthropology 2(2), 86–97.

Biesalski, H. (2010). Ernährung und Evolution. In H. Biesalski, S. Bischoff, and C. Puchstein (Eds.), Ernährungsmedizin, pp. 4–19. Stuttgart, New York.

Bishop, C. (2006). Pattern Recognition and Machine Learning. Information Science and Statistics. New York: Springer.

Blackwell, M., J. Honaker, and G. King (2014). A unified approach to measurement error and missing data: details and extensions. Technical report, Harvard University OpenScholar.

Blower, G. and J. E. Kelsall (2002). Nonlinear kernel density estimation for binned data: convergence in entropy. Bernoulli 8(4), 423–449.

Blum, J. W. and C. R. Baumrucker (2008). Insulin-like growth factors (IGFs), IGF binding proteins, and other endocrine factors in milk: role in the newborn. In Bioactive Components of Milk, pp. 397–422. New York: Springer.

BMVEL/MRI. Bundeslebensmittelschlüssel. Bundesministerium für Landwirtschaft; Ernährung und Verbraucherschutz/Max- Rubner-Institut- Bundesforschungsinstitut für Ernährung und Lebensmittel.

Bogin, B. and B. H. Smith (1996). Evolution of the human life cycle. American Journal of Human Biology 8(6), 703–716.

Bozzola, M., P. Travaglino, N. Marziliano, C. Meazza, S. Pagani, M. Grasso, M. Tauber, M. Diegoli, A. Pilotto, E. Disabella, et al. (2009). The shortness of pygmies is associated with severe under-expression of the growth hormone receptor. Molecular Genetics and Metabolism 98(3), 310–313.

Braidwood, R. J., J. D. Sauer, H. Helbaek, P. C. Mangelsdorf, H. C. Cutler, C. S. Coon, R. Linton, J. Steward, and A. L. Oppenheim (1953). Symposium: did man once live by beer alone? American Anthropologist 54(4), 515–526.

Brandt, G., W. Haak, C. J. Adler, C. Roth, A. Szécsényi-Nagy, S. Karimnia, S. Möller-Rieker, H. Meller, R. Ganslmeier, S. Friederich, et al. (2013). Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science 342(6155), 257–261.

Breitinger (1937). Zur Berechnung der Körperhöhe aus den langen Gliedmaßenknochen. Anthropologischer Anzeiger 14, 249–274.

Bröste, K., C. J. Becker, J. Brøndsted, and J. B. Jørgensen (1956). Prehistoric Man in Denmark: a Study in Physical Anthropology.

Bräuer, G. (1996). Osteometrie. In R. Knußmann (Ed.), Vergleichende Biologie des Menschen. Lehrbuch der Anthropologie und Humangenetik, pp. 106–231. Stuttgart: Fischer.

Buonaccorsi, J. (2010). Measurement Error: Models, Methods, and Applications. Chapman & Hall/CRC Interdisciplinary Statistics. Boca Raton, FL: Taylor & Francis.

Burger, J., M. Kirchner, B. Bramanti, W. Haak, and M. G. Thomas (2007). Absence of the lactase-persistence-associated allele in early neolithic europeans. Proceedings of the National Academy of Sciences 104(10), 3736–3741.

Burger-Heinrich, E. Die menschlichen Skelettreste aus dem Gräberfeld von Stuttgart-Mühlhausen "Viesenhäuser Hof". Anthropologische Befunde der Grabungen aus den Jahren 1977 und 1982. unpublished.

Burnham, K. P. and D. R. Anderson (2004). Multimodel inference understanding AIC and BIC in model selection. Sociological Methods & Research 33(2), 261–304.

Camarda, C. G., P. H. Eilers, and J. Gampe (2008). Modelling general patterns of digit preference. Statistical Modelling 8(4), 385–401.

Card, D. and J. Rothstein (2007, December). Racial segregation and the black-white test score gap. Journal of Public Economics 91(11-12), 2158–2184.

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. London: Taylor & Francis.

Carroll, R. J., A. Delaigle, and P. Hall (2007). Non-parametric regression estimation from data contaminated by a mixture of Berkson and classical errors. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(5), 859–878.

Carroll, R. J., J. D. Maca, and D. Ruppert (1999). Nonparametric regression in the presence of measurement error. Biometrika 86(3), pp. 541–554.

Celeux, G., D. Chauveau, and J. Diebolt (1996). Stochastic versions of the EM algorithm: an experimental study in the mixture case. Journal of Statistical Computation and Simulation 55(4), 287–314.

Celeux, G., F. Forbes, C. P. Robert, D. M. Titterington, et al. (2006). Deviance information criteria for missing data models. Bayesian Analysis 1(4), 651–673.

Chamla, M. C., J. N. Biraben, and J. Dastugue (1970). Les Hommes épipaléolithiques de Columnata, Algérie occidental: étude anthropologique, Volume 15. Paris: Arts et métiers graphiques.

Charles, R. (1958). Étude anthropologique des nécropoles d'Argos. Contribution à l'étude des populations de la Grèce antique. Bulletin de Correspondance Hellénique 82(1), 268–313.

Charles, R. (1962). Contribution a l'étude anthropologique de site du Ras Shamra. In C. Schaeffer (Ed.), Ugaritica, Volume 4, pp. 521–556. Paris.

Charles, R. P. (1965). Anthropologie archéologique de la Crète, Volume 14. Ecole française d'Athènes.

Çiner, R. (1964). Evdi tepesi ve civarında çıkarılan iskelet kalıntılarının tetkiki. Antropoloji 1, 78–98.

Çiner, R. (1965). Altıntepe (urartu) iskeletlerine ait kalıntıların tetkiki. Belleten 29(114), 225–244.

Cole, T. J. (2003). The secular trend in human physical growth: a biological view. Economics & Human Biology 1(2), 161–168.

Collier, R. J., M. Miller, J. Hildebrandt, A. Torkelson, T. White, K. Madsen, J. Vicini, P. Eppard, and G. Lanza (1991). Factors affecting insulin-like growth factor-I concentration in bovine milk. Journal of Dairy Science 74(9), 2905–2911.

Cook, J. R. and L. A. Stefanski (1994). Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association 89(428), 1314–1328.

Cooper, H., L. V. Hedges, and J. C. Valentine (2009). The Handbook of Research Synthesis and Meta-Analysis. New York: Russell Sage Foundation.

Cordain, L., J. B. Miller, S. B. Eaton, N. Mann, S. H. Holt, and J. D. Speth (2000). Plant-animal subsistence ratios and macronutrient energy estimations in worldwide hunter-gatherer diets. The American Journal of Clinical Nutrition 71(3), 682–692.

Coward, F., S. Shennan, S. Colledge, J. Conolly, and M. Collard (2008). The spread of neolithic plant economies from the Near East to Northwest Europe: a phylogenetic analysis. Journal of Archaeological Science 35(1), 42–56.

Crawford, F. W., R. E. Weiss, and M. A. Suchard (2015). Sex, lies, and self-reported counts: Bayesian mixture models for longitudinal heaped count data via birth-death processes. The Annals of Applied Statistics 9(2), 572–596.

Cressie, N. and C. Wikle (2011). Statistics for Spatio-Temporal Data. Hoboken, N.J: Wiley.

Crista, F., I. Radulov, L. Crista, A. Lato, F. Sala, A. Berbecea, L. Nita, K. Lato, et al. (2013). Changing quality indicators of wheat crops following fertilizers application. Research Journal of Agricultural Science 45(1), 3–12.

Cunha, E. and F. Cardoso (2001). The osteological series from Cabeço Da Amoreira (Muge, Portuga). Bulletins et Mémoires de la Société d'Anthropologie de Paris 13(3–4).

Currie, I. D., M. Durban, and P. H. Eilers (2004). Smoothing and forecasting mortality rates. Statistical Modelling 4(4), 279–298.

Czajka, J. L. and G. Denmead (2008). Income data for policy analysis: A comparative assessment of eight surveys. Technical Report PR08-62, Washington, DC: Mathematica Policy Research.

d'Anna, M., J. Jauß, and C. Johnson (2014). Food and urbanization. Material and textual perspectives on alimentary practice in early Mesopotamia. origini. rivista di preistoria e protostoria delle civiltà antiche. Prehistory and Protohistory of Ancient Civilizations, submitted.

Davies, T. M., M. L. Hazelton, and J. C. Marshall (2011). sparr: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in R. Journal of Statistical Software 39(1), 1–14.

De Lusignan, S. d., J. Belsey, N. Hague, and B. Dzregah (2004). End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care. Journal of Human Hypertension 18(4), 261–265.

Delaigle, A. (2007). Nonparametric density estimation from data with a mixture of Berkson and classical errors. Canadian Journal of Statistics 35(1), 89–104.

Delaigle, A. (2014). Nonparametric kernel methods with errors-in-variables: Constructing estimators, computing them, and avoiding common mistakes. Australian & New Zealand Journal of Statistics 56(2), 105–124.

Delaigle, A., J. Fan, and R. J. Carroll (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. Journal of the American Statistical Association 104(485).

Delaigle, A. and I. Gijbels (2004). Practical bandwidth selection in deconvolution kernel density estimation. Computational Statistics & Data Analysis 45(2), 249–267.

Delaigle, A., P. Hall, and P. Qiu (2006). Nonparametric methods for solving the Berkson Errors-in-Variables problem. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 68(2), pp. 201–220.

Delaigle, A. and A. Meister (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. Journal of the American Statistical Association 102(480), 1416–1426.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 39(1), 1–38.

Dennell, R. W. (1979). Prehistoric diet and nutrition: some food for thought. World Archaeology 11(2), 121–135.

DerSimonian, R. and N. Laird (1986). Meta-analysis in clinical trials. Controlled Clinical Trials 7(3), 177–188.

Desideri, J. (2015). ADAM : une base de données au service de l'Anthropologie Physique. Université de Genève. 2015-12-21.

Destatis (2009). Germany's Population by 2060 - Results of the 12th coordinated population projection. Wiesbaden: Federal Statistical Office Germany.

Diaconis, P. (2009). The markov chain monte carlo revolution. Bulletin of the American Mathematical Society 46(2), 179–205.

Diebolt, J. and E. Ip (1996). Stochastic EM: method and application. In W. R. Gilks, S. Richardson, and D. Spiegelhalter (Eds.), Markov Chain Monte Carlo in practice. London: Chapman & Hall.

Donovan, S. and J. Odle (1994). Growth factors in milk as mediators of infant development. Annual Review of Nutrition 14(1), 147–167.

Duong, T. (2014). ks: Kernel smoothing. Paris. R package version 1.9.0.

Duranti, M. and C. Gius (1997). Legume seeds: protein content and nutritional value. Field Crops Research 53(1), 31–45.

Düring, B. S. (2010). The prehistory of Asia Minor: from complex hunter-gatherers to early urban societies. Cambridge University Press.

Duyar, I. and C. Pelin (2003). Body height estimation based on tibia length in different stature groups. American Journal of Physical Anthropology 122(1), 23–27.

Eaton, S. B. and M. Konner (1985). Paleolithic nutrition. a consideration of its nature and current implications. The New England Journal of Medicine 312(5), 283–9.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. Statistical Science 11(2), 89–102.

Eklund, A. and G. Ågren (1975). Nutritive value of poppy seed protein. Journal of the American Oil Chemists' Society 52(6), 188–190.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association 90(430), 577–588.

Evershed, R. P., S. Payne, A. G. Sherratt, M. S. Copley, J. Coolidge, D. Urem-Kotsu, K. Kotsakis, M. Özdoğan, A. E. Özdoğan, O. Nieuwenhuyse, et al. (2008). Earliest date for milk use in the near east and southeastern europe linked to cattle herding. Nature 455(7212), 528–531.

Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). Regression: Models, Methods and Applications. Berlin, Heidelberg: Springer.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. The Annals of Statistics 19(3), 1257–1272.

Fan, J. and Y. K. Truong (1993). Nonparametric regression with errors in variables. The Annals of Statistics 21(4), 1900–1925.

Faris, M. A.-I. E. and H. R. Takruri (2003). Study of the effect of using different levels of tahinah (sesame butter) on the protein digestibility-corrected amino acid score (pdcaas) of chickpea dip. Journal of the Science of Food and Agriculture 83(1), 7–12.

Farkas, G. and P. Lipták (1975). Anthropologische Auswertung des bronzezeitlichen Gräberfelds bei Tápé. In O. Trogmayer (Ed.), Das bronzezeitliche Gräberfeld bei Tapé. Budapest: Akadémiai Kiadó.

Federation of Finnish Learned Societies' Open Journal Systems (2014). Planning for Aging Neighborhoods, Number 6 in Annual Architectural Symposium in Finland. Federation of Finnish Learned Societies' Open Journal Systems.

Ferembach, D. (1958). Note preliminaire sur les ossements humains découverts a Safadi (pres de Beersheva), Israel, par la mission archéologique francaise. unpublished.

Ferembach, D. (1965). Diagrammes crâniens sagittaux et mensurations individuelles des squelettes ibéromaurusiens de Taforalt (Maroc oriental). Travaux du Centre de Recherches Anthropologiques. Paris: Préhistoriques et Ethnographiques 117, Arts et Métiers graphiques.

Ferembach, D. (1970). Étude anthropologique des ossements humains proto-néolithiques de Zawi Chemi Shanidar Irak. Sumer 26(1-2), 21–65.

Ferembach, D. (1982). Mesures et indices des squelettes humains néolithiques de Catal-Hüyük (Turquie). Paris: Ecole pratique des hautes études, Laboratoire d'anthropologie biologique.

Feuerverger, A., P. T. Kim, and J. Sun (2008). On optimal uniform deconvolution. Journal of Statistical Theory and Practice 2(3), 433–451.

Floud, R., R. W. Fogel, B. Harris, and S. C. Hong (2011). The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700. Cambridge University Press.

Food and A. O. of the United Nations (FAO) (1995). Food and nutrition series: Sorghum and millets in human nutrition. `http://www.fao.org/docrep/t0818e/t0818e00.htm` (15.08.2014).

Food and A. O. of the United Nations (FAO) (2012). Composition of meat. `http://www.fao.org/ag/againfo/t_hemes/en/meat/backgr_composition.html` (06.08.2014).

Formicola, V. (1993). Stature reconstruction from long bones in ancient population samples: an approach to the problem of its reliability. American Journal of Physical Anthropology 90(3), 351–358.

Formicola, V. and M. Franceschi (1996). Regression equations for estimating stature from long bones of early holocene european samples. American Journal of Physical Anthropology 100(1), 83–88.

Frost, C. and S. G. Thompson (2000). Correcting for regression dilution bias: comparison of methods for a single predictor variable. Journal of the Royal Statistical Society: Series A (Statistics in Society) 163(2), 173–189.

Fukudome, S.-i. and M. Yoshikawa (1992). Opioid peptides derived from wheat gluten: their isolation and characterization. FEBS letters 296(1), 107–111.

Fuller, W. (2009). Measurement Error Models. Wiley Series in Probability and Statistics. New York: Wiley.

Fully, G. (1956). Une nouvelle méthode de déterminationde la taille. Annales de Médecine Légale 36, 266–273.

Galasizska-Pomykol, I. and I. Szewko-Szwaykowska (1967). Anthropological analysis of the skeletal remains from Mierzanowice. Materialy i Prace Antropologiczne 74, 217–244.

Galer, D. (2007). The human remains. In A. W. R. Whittle and D. Benson (Eds.), Building memories: the Neolithic Cotswold long barrow at Ascott-under-Wychwood, Oxfordshire. Oxford: Oxbow books.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute of Great Britain and Ireland, 246–263.

Gardner, M. L. (1988). Gastrointestinal absorption of intact proteins. Annual Review of Nutrition 8(1), 329–350.

Gelfand, A., P. Diggle, P. Guttorp, and M. Fuentes (2010). Handbook of Spatial Statistics. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis.

Gelfand, A. E., A. Kottas, and S. N. MacEachern (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. Journal of the American Statistical Association 100(471), 1021–1035.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. Statistical Science 7(4), 457–472.

Gerhardt, K. (1953). Die Glockenbecherleute in Mittel- und Westdeutschland. Ein Beitrag zur Paläanthropologie Eurafrikas. Stuttgart: Schweizerbart.

Gerhardt, K. (1968). Menschlichte Überreste aus bandkeramischen Gräbern von Mangolding, Ldkr. Regensburg-Süd. Vor allem ein Beitrag zur Paläopathologie. Quartär 19, 337–346.

Gerhardt, K. and D. Gerhardt-Pfannenstiel (1984/1985). Schädel und Skelette der Linearbandkeramik von Mulhouse-Est (Rixheim) im Elsaß. Acta Praehistorica et Archeologica 16/17, 55–69.

Gochman, I. (1966). Naselenie Ukrainy v Epochu Mesolita i Neolita. Moscow: Antropogicheskiy Ocherk.

Golyandina, N., A. Pepelyshev, and A. Steland (2012). New approaches to nonparametric density estimation and selection of smoothing parameters. Computational Statistics & Data Analysis 56(7), 2206–2218.

Gorr, W., M. Johnson, and S. Roehrig (2001). Spatial decision support system for home-delivered services. Journal of Geographical Systems 3, 181–197.

Gouin, P. (1993). Bovins et laitages en mésopotamie méridionale au 3ème millénaire. Iraq 55, 135–145.

Graw, M., H.-T. Haffner, and A. Czarnetzki (1997). Methode zur Untersuchung des Margo supraorbitalis als Kriterium zur Geschlechtsdiagnose-Reliabilität und Validität. Rechtsmedizin 7(4), 121–126.

Graw, M., J. Wahl, and M. Ahlbrecht (2005). Course of the meatus acusticus internus as criterion for sex differentiation. Forensic Science International 147(2), 113–117.

Grimm, H. (1958). Die Schnurkeramiker von Schafstädt. Jahresschrift mitteldeutsche Vorgeschichte 41(42), 299–314.

Grimm, H. (1960). Anthropologischer Befund bei dem Schnurkeramiker-Skelett von Räpitz. Ausgrabungen und Funde 5(2), 74–77.

Groß, M. (2016a). Kernelheaping: Kernel Density Estimation for Heaped Data. Berlin: Freie Universität. R package version 1.6.

Groß, M. (2016b). Modeling body height in prehistory using a spatio-temporal bayesian errors-in variables model. AStA Advances in Statistical Analysis, forthcoming.

Groß, M., U. Rendtel, T. Schmid, S. Schmon, and N. Tzavidis (2016). Estimating the density of ethnic minorities and aged people in Berlin: Multivariate kernel density estimation applied to sensitive geo-referenced administrative data protected via measurement error. Journal of the Royal Statistical Society: Series A (Statistics in Society), forthcoming.

Güleç, E. (1989). Paleoantropolojik verilere göre eski Anadolu bireylerinin boy açisindan incelenmesi. Araştirma Sonuçlari Toplantisi 5, 147–160.

Gustafson, P. (2003). Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. New York: CRC Press.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. Statistical Science 20(2), 111–137.

Gözlük, P., A. Sevim, A. Yığıt, and H. Yilmaz (2002). Hakkâri erken demir çağı iskeletlerininpaleoantropolojik açıdan incelenmesi. Arkeometri Sonucları Toplantısı 18, 31–40.

Haak, W., I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211.

Haas, N. (1970). Anthropological observations on the skeletal remains from Giv'at ha-Mivtar. Israel Exploration Journal, 38–59.

Haas, N. and H. Nathan (1973). An attempt at a social interpretation of the Chalcolithic burials in the Nahal Mishmar caves. In Y. Aharoni (Ed.), Excavations and Studies: Essays in Honour of Professor Shemuel Yeivin. Tel-Aviv, pp. 143–153. University, Institute of Archaeology, Tel-Aviv.

Haberstroh, J. and M. Harbeck (2013). Nekropolen des 5. Jahrhunderts n. Chr. in Bayern: eine Projektskizze aus archäologischer und anthropologischer Perspektive. 54.

Haidle, M. N. (2014). Mangel-Krisen-Hungersnöte? Ernährungszustände in Süddeutschland und der Nordschweiz vom Neolithikum bis ins 19. Jahrhundert. Archäologische Informationen 20(1), 185–188.

Hald, J. and J. Wahl (2009). Eine Gräbergruppe des Jung- bis Endneolithikums von Engen-Welschingen, Ldkr. Konstanz. In J. Biel, J. Heiligmann, and D. Krausse (Eds.), Landesarchäologie, Festschrift für Dieter Planck, pp. 87–106. Stuttgart.

Hammond, B. (2007). The food safety assessment of bovine somatotropin (bST). Food Safety of Proteins in Agricultural Biotechnology, 167–208.

Hanisch, J. U. (2007). Rounding of Income Data : An Empirical Analysis of the Quality of Income Data with Respect to Rounded Values and Income Brackets with Data from the European Community Household Panel. Ph. D. thesis, Goethe-Universität Frankfurt am Main.

Härdle, W. and D. W. Scott (1992). Smoothing by weighted averaging of rounded points. Computational Statistics 7, 97–128.

Hardy, K., T. Blakeney, L. Copeland, J. Kirkham, R. Wrangham, and M. Collins (2009). Starch granules, dental calculus and new perspectives on ancient diet. Journal of Archaeological Science 36(2), 248–255.

Heitjan, D. F. and D. B. Rubin (1990). Inference from coarse data via multiple imputation with application to age heaping. Journal of the American Statistical Association 85(410), 304–314.

Heitjan, D. F. and D. B. Rubin (1991). Ignorability and coarse data. Annals of Statistics 19(4), 2244–2253.

Helmuth, H. (1967). Anthropologische Untersuchung von Leichenbränden. In W. Orthmann and H. Helmuth (Eds.), Das Gräberfeld bei Ilica. Wiesbaden: F. Steiner.

Henry, A. G. and D. R. Piperno (2008). Using plant microfossils from dental calculus to recover human diet: a case study from Tell al-Raqā'i, Syria. Journal of Archaeological Science 35(7), 1943–1950.

Hertzler, S. R. and S. M. Clancy (2003). Kefir improves lactose digestion and tolerance in adults with lactose maldigestion. Journal of the American Dietetic Association 103(5), 582–587.

Higdon, R. and D. W. Schafer (2001). Maximum likelihood computations for regression with measurement error. Computational Statistics & Data Analysis 35(3), 283 – 299.

Hiller, S. and V. Nikolov (1997). Österreichisch-Bulgarische Ausgrabungen und Forschungen in Karanovo. Die Ausgrabungen im Südsektor 1984-1992. Salzburg: Berger & Söhne.

Hnila, P. (2002). Some remarks on the opium poppy in ancient Anatolia. In R. Aslan, S. Blum, G. Kastl, F. Schweizer, and D. Thumm (Eds.), Mauerschau. Festschrift für Manfred Korfmann, pp. 315–328. Grunbach.

Hodder, I. (2005). Çatalhöyük perspectives: reports from the 1995-99 seasons, Volume 6. McDonald Institute for Archaeological Research.

Hopf, M. and D. Zohary (2001). Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford University Press.

Hopfel, F., W. Platzer, and K. Spindler (1992). Der Mann im Eis, Band 1. In Bericht über das Internationale Symposium 1992 in Innsbruck. Veröffentlichungen der Universität Innsbruck. Universität Innsbruck.

Horwitz, L. K. and B. Rosen (2005). A review of camel milking in the southern Levant. In J. Mulville and A. Outram (Eds.), The zooarchaeology of Milk and Fats, pp. 121–131. Oxford.

Horwitz, L. K. and P. Smith (2000). The contribution of animal domestication to the spread of zoonoses: A case study from the southern Levant. Anthropozoologica 31, 77–84.

Hubbard, R. (1980). Development of agriculture in Europe and the Near East: evidence from quantitative studies. Economic Botany 34(1), 51–67.

Hujić, A. (2016). Das Kind in uns unter der Lupe der Isotopie, Allometrie und Pathologie. Zusammenhang zwischen δ15N und δ13C als Eiweißproxy und dem Knochenwachstum bei linienbandkeramischen Individuen von Stuttgart-Mühlhausen und Schwetzingen unter Berücksichtigung verschiedener Indikatoren für Nährstoffversorgung. Ph. D. thesis, Freie Universität Berlin.

Iezzi, C. (2009). Regional differences in the health status of the Mycenaean women of East Lokris. In L. Schepartz, S. Fox, and C. Bourbou (Eds.), New directions in the skeletal biology of Greece, Hesperia Supplement 43, pp. 175–192. Princeton.

Ingram, C. J., C. A. Mulcare, Y. Itan, M. G. Thomas, and D. M. Swallow (2009). Lactose digestion and the evolutionary genetics of lactase persistence. Human Genetics 124(6), 579–591.

Iqbal, A., I. A. Khalil, N. Ateeq, and M. S. Khan (2006). Nutritional quality of important food legumes. Food Chemistry 97(2), 331–335.

Itan, Y., B. L. Jones, C. J. Ingram, D. M. Swallow, and M. G. Thomas (2010). A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evolutionary Biology 10(1), 36.

Itan, Y., A. Powell, M. A. Beaumont, J. Burger, M. G. Thomas, et al. (2009). The origins of lactase persistence in Europe. PLoS Computational Biology 5(8), 1–13.

Izenman, A. J. (1991). Recent developments in nonparametric density estimation. Journal of the American Statistical Association 86(413), 205–224.

Jacobshagen, B. and M. Kunter (1999). Die mittelneolithische Skelettpopulation aus Trebur, Kreis Groß Gerau. Ergebnisse der Anthropologischen Bearbeitung. In H. Spatz (Ed.), Das mittelneolithische Gräberfeld von Trebur, Kreis Groß-Gerau. Materialien zur Vor-und Frühgeschichte von Hessen, Volume 19, pp. 281–332. Wiesbaden.

Jacomet, S. and S. Karg (1996). Ackerbau und Umwelt der Seeufersiedlungen von Zug-Sumpf im Rahmen der mitteleuropäischen Spätbronzezeitlicher Ergebnisse archäobotanischer Untersuchungen. In Die spätbronzezeitlichen Ufersiedlungen von Zug-Sumpf, Volume 1, pp. 198–303. Zug: Museum für Urgeschichte.

Jaeger, U., H. Bruchhaus, L. Finke, K. Kromeyer-Hauschild, and K. Zellner (1998). Säkularer Trend bei der Körperhöhe seit dem Neolithikum. Anthropologischer Anzeiger, 117–130.

Jakab, J. (1999). Anthropologische Analyse der Gräber mit Totenhäusern des frühbronzezeitlichen Gräberfeldes in Mỳtna Nová Ves. Praehistorische Zeitschrift 74(1), 58–67.

Jann, B. (2007). Univariate kernel density estimation. Statistical Software Component (S456410).

Jaworska, G. and E. Bernaś (2013). Amino acid content of frozen agaricus bisporus and boletus edulis mushrooms: Effects of pretreatments. International Journal of Food Properties 16(1), 139–153.

Jones, G. and S. M. Valamoti (2005). Lallemantia, an imported or introduced oil plant in Bronze Age northern Greece. Vegetation History and Archaeobotany 14(4), 571–577.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. Statistics and Computing 3(3), 135–146.

Jones, M. C., J. S. Marron, and S. J. Sheather (1996). A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91(433), 401–407.

Jones, U. (2011). Metates and Hallucinogens in Costa Rica. Papers from the Institute of Archaeology 2, 29–34.

Jungwirth, J. and Ä. Kloiber (1973). Die neolithischen Skelette aus Österreich. In H. Schwabedissen (Ed.), Die Anfänge des Neolithikums vom Orient bis Nordeuropa. Köln, Wien: Böhlau Verlag.

Kalač, P. (2009). Chemical composition and nutritional value of European species of wild growing mushrooms: A review. Food Chemistry 113(1), 9–16.

Kallweit, E., G. Fries, R. Kielwein, and S. Scholtyssek (1988). Qualität tierischer Nahrungsmittel. Fleisch - Milch - Eier. Stuttgart: Verlag Eugen Ulmer.

Kang, S., J. Kim, J. Imm, S. Oh, and S. Kim (2006). The effects of dairy processes and storage on insulin-like growth factor-I (IGF-I) content in milk and in model IGF-I–fortified dairy products. Journal of Dairy Science 89(2), 402–409.

Kaniewski, D., E. Van Campo, T. Boiy, J.-F. Terral, B. Khadari, and G. Besnard (2012). Primary domestication and early uses of the emblematic olive tree: palaeobotanical, historical and molecular evidence from the Middle East. Biological Reviews 87(4), 885–899.

Kansu, Ş. and M. Atasayan (1939). Afyonkarahisar Civarında Kusura Hafriyatında Meydana Çıkarılan Bakırçağı Ve Eti Devrine Ait İskeletler Üzerinde Tetkikler. Türk Antropoloji Mecmuası 15, 272–281.

Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. Journal of the American Statistical Association 91(435), 1343–1370.

Keith, A. (1934). Report on human remains. In L. Wooley (Ed.), Ur Excavations. The Royal Cemetery, Volume 2, pp. 214–240. New York.

Kemkes-Grottenthaler, A. (1997). Das Frauendefizit archäologischer Serien - ein paläodemographisches Paradoxon? Anthropologischer Anzeiger 55(3/4), 265–280.

Kerem, Z., S. Lev-Yadun, A. Gopher, P. Weinberg, and S. Abbo (2007). Chickpea domestication in the Neolithic Levant through the nutritional perspective. Journal of Archaeological Science 34(8), 1289–1293.

Kerig, T. (2007). "Als Adam grub...". Vergleichende Anmerkungen zu landwirtschaftlichen Betriebsgrößen In prähistorischer Zeit. Ethnographisch-Archäologische Zeitschrift 48(3), 375–402.

Kettunen, J., K. Silander, O. Saarela, N. Amin, M. Müller, N. Timpson, I. Surakka, S. Ripatti, J. Laitinen, A.-L. Hartikainen, et al. (2010). European lactase persistence genotype shows evidence of association with increase in body mass index. Human Molecular Genetics 19(6), 1129–1136.

Kiesewetter, H. (2003). The neolithic population at jebel buhais 18: remarks on funerary practices, palaeodemography and palaeopathology. In D. Potts, H. Al Naboodah, P. Hellyer, and I. Ashraf (Eds.), Archaeology of the United Arab Emirates: Proceedings of the First International Conference on the Archaeology of the UAE. London: Trident, pp. 35–44.

Kimura, T., Y. Murakawa, M. Ohno, S. Ohtani, and K. Higaki (1997). Gastrointestinal absorption of recombinant human insulin-like growth factor-i in rats. Journal of Pharmacology and Experimental Therapeutics 283(2), 611–618.

Klagsburn, M. (1978). Human milk stimulates dna synthesis and cellular proliferation in cultured fibroblasts. Proceedings of the National Academy of Sciences 75(10), 5057–5061.

Kloiber, Ä. and J. Kneidinger (1970). Die neolithische Siedlung und die neolithischen Gräberfundplätze von Rutzing und Haid, Gemeinde Hörsching, politischer Bezirk Linz-Land, Oberösterreich. Jahrbuch des Oberösterreichischen Museal-Vereines IIB 115(1), 21–38.

Knip, A. (1974). Late neolithic skeleton finds from Molenaarsgraaf. Analecta Praehistoria Leidensia 7, 379–395.

Knöpke, S. and J. Wahl (2010). Der urnenfelderzeitliche Männerfriedhof von Neckarsulm. Stuttgart: Theiss.

Knüsel, C. (2007). The human remains. Prague: Archeologickỳ ústav AV ČR.

Knußmann, R. (Ed.) (1988). Anthropologie: Handbuch der vergleichenden Biologie des Menschen; Lehrbuchs der Anthropologie. Wesen und Methoden der Anthropologie: Wissenschaftstheorie, Geschichte, morphologische Methoden. Stuttgart, New York: Fischer.

Knussmann, R. and R. Knussmann (1978). Die Skelettreste der Rössener und Michelsberger Kulturepoche. In H. Schwabedissen (Ed.), Die Anfänge des Neolithikums vom Orient bis Nordeuropa, pp. 164–217. Köln, Wien: Böhlau Verlag.

Koepke, N. and J. Baten (2005). The biological standard of living in Europe during the last two millennia. European Review of Economic History 9(1), 61–95.

Koepke, N. and J. Baten (2008). Agricultural specialization and height in ancient and medieval Europe. Explorations in Economic History 45(2), 127–146.

Köhler, K. (2009). The anthropological remains from the Budakalász cemetery. In B. M. and R. P. (Eds.), The Copper Age cemetery of Budakalász. Pytheas Kiadó, pp. 303–364. Budapest.

Kolář, J., I. Jarošová, G. Dreslerová, E. Drozdová, and M. Dobisíková (2012). Food strategies in Central Moravia (Czech Republic) during Final Neolithic: a case study of Corded Ware culture communities. Archeologické Rozhledy 64, 237–264.

Komlos, J. (1989). Nutrition and Economic Development in the Eighteenth-Century Habsburg Monarchy: an Anthropometric History. Princeton: University Press Princeton.

Komlos, J. (1995). The Biological Standard of Living in Europe and America, 1700-1900: Studies in Anthropometric History. Collected studies. Brookfield, Vt.

Konduktorova, T. (1956). Materialy po paleoantropologii ukrainy. Antropologiceskij Sbornik 1, 166–203.

Konduktorova, T. (1960). Paleoantropologicheskie materiali vovniz'kich pizn' oneoliticnich mogil'nikiv. Materiali z Antropologij Ukrajni 1, 66–97.

Konduktorova, T. (1974). The ancient population of the Ukraine (from the Mesolithic Age to the first centuries of our era). Anthropologie 12(1/2), 5–149.

Konigsberg, L. W., S. M. Hens, L. M. Jantz, and W. L. Jungers (1998). Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical

anthropology. American Journal of Physical Anthropology 107(Supplement 27), 65–92.

Korfmann, M. (1983). Demircihüyük 1: Architektur, Stratigraphie und Befunde. C. I&II, Mainz: Verlag Phillip von Zabern.

Krogman, W. (1962). The human skeleton in forensic medicine. Springfield, Illinois: Charles C. Thomas.

Krüttli, A., A. Bouwman, G. Akgül, P. Della Casa, F. Rühli, and C. Warinner (2014). Ancient DNA analysis reveals high frequency of European lactase persistence allele (t-13910) in medieval central Europe. PloS one 9(1), e86251.

Kunter, M. (1990). Menschliche Skelettreste aus Siedlungen der El Argar-Kultur: Ein Beitrag der Prähistorischen Anthropologie zur Kenntnis bronzezeitlicher Bevölkerungen Südostspaniens. Mainz: Verlag Phillip von Zabern.

Kurth, G., C. Dupertuis, and J. Hadden (1954). Ein Beitrag zur Vergleichbarkeit errechneter Körperhöhen (Nach Manouvrier, Pearson, Telkkä und Breitinger). Zeitschrift für Morphologie und Anthropologie 46, 317–370.

Kurz, G. (1993). Vorgeschichtliche Siedlungen und Gräber beim Viesenhäuser Hof, Stuttgart-Mühlhausen, pp. 61–64. Stuttgart: Konrad Theiss Verlag.

Kwan, M.-P., I. Casas, and B. C. Schmitz (2004). Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks? Cartographica: The International Journal for Geographic Information and Geovisualization 39(2), 15–28.

Lang, S. and A. Brezger (2004). Bayesian P-Splines. Journal of Computational and Graphical Statistics 13(1), 183–212.

Lang, S. and M. Sunder (2003). Non-parametric regression with BayesX: a flexible estimation of trends in human physical stature in 19th century America. Economics & Human Biology 1(1), 77–89.

Laron, Z. (1993). Disorders of growth hormone resistance in childhood. Current Opinion in Pediatrics 5(4), 474–480.

Laron, Z. (2001). Insulin-like growth factor 1 (IGF-1): a growth hormone. Molecular Pathology 54(5), 311.

Laron, Z. and J. Kopchick (2011). Laron syndrome-from man to mouse. Berlin, Heidelberg: Springer.

Laron, Z., A. Kowadlo-Silbergeld, R. Eshet, and A. Pertzelan (1980). Growth hormone resistance. Annals of Clinical Research 12(5), 269–277.

Larsen, C. S. (1995). Biological changes in human populations with agriculture. Annual Review of Anthropology 24, 185–213.

Lee, D.-J. and M. Durbán (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. Statistical Modelling 11(1), 49–69.

Lentner, K. and C. Diem (1973). Documenta Geigy Wissenschaftliche Tabellen. Baden: Wehr, Geigy Pharmazeutika.

Lieverse Angela, R. (1999). Diet and the aetiology of dental calculus. International Journal of Osteoarchaeology 9(4), 219–232.

Lightfoot, E., X. Liu, and M. K. Jones (2013). Why move starchy cereals? A review of the isotopic evidence for prehistoric millet consumption across Eurasia. World Archaeology 45(4), 574–623.

Liu, J.-L. and D. LeRoith (1999). Insulin-like growth factor i is essential for postnatal growth in response to growth hormone. Endocrinology 140(11), 5178–5184.

Long, J. P., N. E. Karoui, and J. A. Rice (2014). Kernel density estimation with Berkson error. arXiv preprint arXiv:1401.3362.

Lorke, D., H. Münzner, and E. Walter (1953). Zur Rekonstruktion der Körpergröße eines Menschen aus den langen Gliedmaßenknochen. Deutsche Zeitschrift für die gesamte gerichtliche Medizin 42(2), 189–202.

Lösch, S., G. Grupe, and J. Peters (2006). Stable isotopes and dietary adaptations in humans and animals at pre-pottery Neolithic Nevallı Çori, Southeast Anatolia. American Journal of Physical Anthropology 131(2), 181–193.

Loui, A., M. Elmlinger, F. Hochhaus, R. Grund, M. Obladen, and M. Ranke (2004). Insulin-like-growth-Faktoren (IGF) und Bindungsproteine (IGFBP) in Muttermilch von Früh-und Reifgeborenen. Zeitschrift für Geburtshilfe und Neonatologie 208(1), 208–235.

Manoukas, A. G., B. Mazomenos, and M. A. Patrinou (1973). Amino acid compositions of three varieties of olive fruit. Journal of Agricultural and Food Chemistry 21(2), 215–217.

Manouvrier, L. (1892). La détermination de la taille d'après les grands os des membres. Revuemensuelle de l'école d'anthropologie 4, 227–233.

Marcsik, A. (1979). The anthropological material of the Pit-Grave Kurgans in Hungary. In I. Ecsedy (Ed.), The people of the Pit-Grave Kurgans in Easterrn Hungary. Fontes Arch Hung, Budapest, pp. 87–98.

Marcus, J., R. Siegers, and M. M. Grabka (2013). Preparation of data from the new soep consumption module: Editing, imputation, and smoothing. Technical Report 70, Data Documentation, DIW.

Marinov, G. (1978). Preliminary data from studies of bone material from the Varna Chalcolithic necropolis during the 1972-1975 period. In G. Georgiev, L. M.,

and A. Fol (Eds.), Die Nekropole in Varna und die Probleme des Chalkolithikums. Internationales Symposium Varna, 19. - 23. April 1976, Sofia, pp. 60–67.

Marron, J. S. (1987). A comparison of cross-validation techniques in density estimation. The Annals of Statistics 15(1), 152–162.

Martin, R. (1928). Lehrbuch der Anthropologie in systematischer Darstellung. Jena: Fischer.

Martorell, R. and T. J. Ho (1984). Malnutrition, morbidity, and mortality. Population and Development Review, 49–68.

Masson, A. and E. Rosenstock (2011). Das Rind in Vorgeschichte und traditioneller Landwirtschaft: archäologische und technologisch-ergologische Aspekte. Mitteilungen der Berliner Gesellschaft für Anthropologie. Ethnologie und Urgeschichte 32, 81–105.

Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, et al. (2015). Genome-wide patterns of selection in 230 ancient eurasians. Nature 528(7583), 499–503.

Maximilian, C. and A. R. P. Romîne (1962). Sarata-Monteoru: Studiu Antropologic. Bukarest: Editura Academiei Republicii Populare Romîne.

McLachlan, G. and T. Krishnan (2007). The EM Algorithm and Extensions. Wiley Series in Probability and Statistics. New York: Wiley.

Mehta, A., P. J. Mason, and T. J. Vulliamy (2000). Glucose-6-phosphate dehydrogenase deficiency. Best Practice & Research Clinical Haematology 13(1), 21–38.

Menninger, M. (2008). Die schnurkeramischen Bestattungen von Lauda-Königshofen. Steinzeitliche Hirtennomaden im Taubertal? Ph. D. thesis, Universität Tübingen.

Merrill, R. M. and J. S. Richardson (2009). Peer reviewed: Validity of self-reported height, weight, and body mass index: Findings from the national health and nutrition examination survey, 2001-2006. Preventing Chronic Disease 6(4).

Meyer, C., O. Kürbis, and K. W. Alt (2004). Das Massengrab von Wiederstedt, Ldkr. Mansfelder Land: Auswertung und Gedanken zur Deutung im Kontext der Linienbandkeramik. Jahresschrift für mitteldeutsche Vorgeschichte 88, 31–66.

Miletic, I., M. Miric, Z. Lalic, and S. Sobajic (1991). Composition of lipids and proteins of several species of molluscs, marine and terrestrial, from the Adriatic Sea and Serbia. Food Chemistry 41(3), 303–308.

Miller, N. F. Tracing the development of the agropastoral economy in southeastern Anatolia and northern Syria. In R. Cappers and S. Bottema (Eds.), Studies in early Near Eastern production, subsistence, and environment: The dawn of farming in the Near East. Berlin.

Minnotte, M. C. (1998). Achieving higher-order convergence rates for density estimation with binned data. Journal of the American Statistical Association 93(442), 663–672.

Molberg, Ø., A. K. Uhlen, T. Jensen, N. S. Flæte, B. Fleckenstein, H. Arentz-Hansen, M. Raki, K. E. Lundin, and L. M. Sollid (2005). Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology 128(2), 393–401.

Mollison, T. (1910). Die Körperproportionen der Primaten. Morphologisches Jahrbuch 42, 79–304.

Moore, A. M. T., G. C. Hillman, and A. J. Legge (2000). Village on the Euphrates: from foraging to farming at Abu Hureyra. Oxford: Oxford University Press.

Morgan-Forster, A. H. (2010). Climate, Environment and Malaria during the Prehistory of Mainland Greece. Ph. D. thesis, University of Birmingham.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series C (Applied Statistics) 64(2), 231–252.

Müller-Karpe, H. (1984). Neolithisch-kupferzeitliche Siedlungen in der Geoksjur-Oase, Süd-Turkenistan, Volume 30. München: CH Beck.

Mummert, A., E. Esche, J. Robinson, and G. J. Armelagos (2011). Stature and robusticity during the agricultural transition: Evidence from the bioarchaeological record. Economics & Human Biology 9(3), 284–301.

Nakoinz, O. (2012). Datierungskodierung und chronologische Inferenz–Techniken zum Umgang mit unscharfen chronologischen Informationen. Praehistorische Zeitschrift 87(1), 189–207.

Nesbitt, M. (2001). Wheat evolution: integrating archaeological and biological evidence. The Linnean 3, 37–59.

Neugebauer, J. (1991). Die Nekropole F von Gemeinlebarn, Niederösterreich. Untersuchungen zu den Bestattungssitten und zum Grabraub in der ausgehenden Frühbronzezeit in Niederösterreich südlich der Donau zwischen Enns und Wienerwald. Römisch-Germanische Forschungen 49.

Olivier, G., C. Aaron, G. Fully, and G. Tissier (1978). New estimations of stature and cranial capacity in modern man. Journal of Human Evolution 7(6), 513–518.

Ong, K., J. Kratzsch, W. Kiess, and D. Dunger (2002). Circulating IGF-I levels in childhood are related to both current body composition and early postnatal growth rate. The Journal of Clinical Endocrinology & Metabolism 87(3), 1041–1044.

Orfila, M. J. B. and O. Lesueur (1831). Traité des exhumations juridiques: et considérations sur les changemens physiques que les cadavres éprouvent en se

pourrissant dans la terre, dans l'eau, dans les fosses d'aisance et dans le fumier, Volume 1. Paris: Béchet.

Orschiedt, J. (1998). Bandkeramische Siedlungsbestattungen in Südwestdeutschland. Rahden/Westf.: Verlag Marie Leidorf.

Orthmann, W. (1975). Der Alte Orient, Volume Kunstgeschichte 14. Berlin: Propyläen-Verlag.

Orthmann, W. (1981). Halawa 1977 bis 1979: vorläufiger Bericht über die 1. bis 3. Grabungskampagne. Saarbrücker Beiträge zur Altertumskunde 1.

Ortner, D. J. and B. Fröhlich (2008). The Early Bronze Age I Tombs and Burials of Bab edh-Dhra', Jordan, Volume 3. Washington D.C.: Rowman Altamira.

Otterbach, S. and A. Sousa-Poza (2010). How accurate are German work-time data? A comparison of time-diary reports and stylized estimates. Social Indicators Research 97(3), 325–339.

Ottesen, L. H., F. Bendtsen, and A. Flyvbjerg (2001). The insulin-like growth factor binding protein 3 ternary complex is reduced in cirrhosis. Liver 21(5), 350–356.

Özbek, M. (1988). Çayönü insanları ve sağlık sorunları. Toplantısı 4, 121–152.

Özbek, M. (1991). Aşıklı höyük neolitik insanları. Arkeometri Sonuçları Toplantısı 7, 145–60.

Özbek, M. (2000). Öküzini insanlarının antropolojik analizi. Arkeometri Sonuçları Toplantısı 15, 127–144.

Ozonoff, A., C. Jeffery, J. Manjourides, L. F. White, and M. Pagano (2007). Effect of spatial resolution on cluster detection: a simulation study. International Journal of Health Geographics 6(1), 1–7.

Palau, A. A. and J. E. F. Palma (1997). La necrópolis megalítica del pantano de Los Bermejales, Volume 39. Granada: Monogr. Hum. Arte y Arqu.

Papathanasiou, A. (2001). A Bioarchaeological Analysis of Neolithic Alepotrypa Cave, Greece, Volume 961. Oxford: British Archaeological Reports Ltd.

Parenti, R. and P. Messeri (1962). I resti scheletrici umani del Neolitico Ligure, Volume 50. Pisa: Paleontographia Italica.

Pasveer, J. and H. Uytterschaut (1992). Two late neolithic human skeletons, a recent discovery in the Netherlands. International Journal of Osteoarchaeology 2(1), 1–14.

Pearson, J., M. Grove, M. Özbek, and H. Hongo (2013). Food and social complexity at Çayönü Tepesi, southeastern Anatolia: Stable isotope evidence of differentiation in diet according to burial practice and sex in the early Neolithic. Journal of Anthropological Archaeology 32(2), 180–189.

Pearson, K. (1899). Mathematical contributions to the theory of evolution. V. On the reconstruction of the stature of prehistoric races. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 192, 169–244.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, et al. (2007). Diet and the evolution of human amylase gene copy number variation. Nature Genetics 39(10), 1256–1260.

Perscheid, M. (1974). Das Mainzer Lochkartenarchiv fur postkraniales Skelettmaterial prähistorischer Populationen. Homo 25(2), 121–124.

Persson, O. and E. Persson (1988). Anthropological report concerning the interred mesolithic populations from Skateholm, Southern Sweden. In L. Larsson (Ed.), The Skateholm Project. I. Man and Environment, pp. 89–105. Stockholm: Almqvist & Wiksell International.

Peterson, R. D., L. J. Krivo, and C. R. Browning (2008). Segregation and race/ethnic inequality in crime: New directions. In F. T. Cullen, J. Vright, and K. Blevins (Eds.), Taking Stock : The Status of Criminological Theory. New Brunswick, NJ: Transaction.

Pham, T. H., J. T. Ormerod, and M. Wand (2013). Mean field variational Bayesian inference for nonparametric regression with measurement error. Computational Statistics & Data Analysis 68, 375–387.

Prasad, A., V. Bhagwat, S. Porwal, and D. Joshi (2012). Estimation of Human Stature from length of ulna in Marathwada Region of Maharashtra. International Journal of Biological & Medical Research 3(4).

Preuschoft, H. (1962). Zur Anthropologie der Bandkeramiker aus Butzbach in Hessen. Fundberichte aus Hessen 2, 85–97.

Price, T. D., C. Knipper, G. Grupe, and V. Smrcka (2004). Strontium isotopes and prehistoric human migration: the Bell Beaker period in Central Europe. European Journal of Archaeology 7(1), 9–40.

Price, T. D., J. Wahl, C. Knipper, E. Burger-Heinrich, G. Kurz, and R. A. Bentley (2003). Das bandkeramische Gräberfeld von Stuttgart-Mühlhausen: Neue Untersuchungsergebnisse zum Migrationsverhalten im frühen Neolithikum. Fundberichte aus Baden-Württemberg 27, 23–58.

Pudney, S. (2008). Heaping and leaping: Survey response behaviour and the dynamics of self-reported consumption expenditure. Technical report, ISER Working Paper Series.

R Core Team (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

R Development Core Team (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rathbun, T. (1984). Skeletal pathology from the Paleolithic through the Metal Ages in Iran and Iraq. In C. M.N. and G. Armelagos (Eds.), Paleopathology at the Origins of Agriculture, pp. 137–168. Orlando: Academic Press, Inc.

Raxter, M. H., B. M. Auerbach, and C. B. Ruff (2006). Revision of the Fully technique for estimating statures. American Journal of Physical Anthropology 130(3), 374.

Razavi, S. M. A., T. Mohammadi Moghaddam, and A. Mohammad Amini (2008). Physical-mechanical properties and chemical composition of Balangu (Lallemantia royleana (Benth. in Walla.)) seed. International Journal of Food Engineering 4(5), 1–12.

Rechler, M. M. and D. R. Clemmons (1998). Regulatory actions of insulin-like growth factor-binding proteins. Trends in Endocrinology & Metabolism 9(5), 176–183.

Reichelt, E., M. Häckel, and H. Bruchhaus (2003). Die Schätzung der Körperhöhe am Beispiel eines mittelalterlichen Gräberfeldes – eine kritische Betrachtung. In N. Benecke (Ed.), Beiträge zur Archäozoologie und Prähistorischen Anthropologie, Volume 4, pp. 178–181. Konstanz.

Richardson, S. and W. R. Gilks (1993). A Bayesian approach to measurement error problems in epidemiology using conditional independence models. American Journal of Epidemiology 138(6), 430–442.

Riquet, R. (1962). Les ossements humains de la Grotte 2 de la Trache à Châteaubernard (Charente). Bulletin de la Société préhistorique de France 59(7-8), 456–463.

Robert Koch Institut (2014). Beiträge zur Gesundheitsberichterstattung des Bundes - Daten und Fakten: Ergebnisse der Studie "Gesundheit in Deutschland aktuell 2009". `http://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsB/GEDA09.pdf?__blob=publicationFile`.

Röhrer-Ertl, O. (1978). Die neolithische Revolution im Vorderen Orient: ein Beitrag zu Fragen der Bevölkerungsbiologie und Bevölkerungsgeschichte. München, Wien: Oldenbourg.

Rollet, E. (1888). De la mensuration des os longs des membres. Thèse pour le doctorat en médicine, 1st series. Lyon, Paris: Storck.

RoPHA (1999). Report on Public Health Aspects of the Use of Bovine Somatotrophin, 15-16 March. Scientific Committee on Veterinary Measures relating to Public Health, European Commission.

Rosenbloom, A. L. (2008). Insulin-like growth factor-I (rhIGF-I) therapy of short stature. Journal of Pediatric Endocrinology and Metabolism 21(4), 301–316.

Rosenstock, E. (2014). Eiweißversorgung und Körperhöhe: zur Übertragbarkeit anthropometrischer Ansätze auf die Archäologie. In W. Schier and M. Meyer (Eds.), Vom Nil bis an die Elbe. Forschungen aus fünf Jahrzehnten am Institut für Prähistorische Archäologie der Freien Universität Berlin.Internationale Archäologie, Studia Honoraria 36. Rahden/Westfalen: Verlag Maria Leidorf.

Rosenstock, E., M. Groß, A. Hujić, and A. Scheibner (2015). Back to good shape: biological standard of living in the copper and bronze ages and the possible role of food. In J. Kneisel, W. Kierleis, N. Taylor, M. dal Corso, and V. Tiedtke (Eds.), Setting the Bronze Age Table: Production, Subsistence, Diet and Their Implications for European Landscapes. Proceedings of the International Workshop 'Socio-environmental dynamics over the last 12.000 years: the creation of landscapes III (5th - 18th April 2013)', Kiel., Bonn. Habelt.

Rösing, F. (1988). Körperhöhenrekonstruktion aus Skelettmaßen. In R. Knußmann (Ed.), Anthropologie. Handbuch der vergleichenden Biologie des Menschen, Volume 1. Stuttgart, New York: Gustav Fischer Verlag.

Rowland, M. L. (1990). Self-reported weight and height. The American Journal of Clinical Nutrition 52(6), 1125–1133.

Rue, H. and L. Held (2005). Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. London: CRC Press.

Ruff, C. B., B. M. Holt, M. Niskanen, V. Sladék, M. Berner, E. Garofalo, H. M. Garvin, M. Hora, H. Maijanen, S. Niinimäki, et al. (2012). Stature and body mass estimation from skeletal remains in the European Holocene. American Journal of Physical Anthropology 148(4), 601–617.

Ruhr-Stickstoff, A. (1988). Faustzahlen für Landwirtschaft und Gartenbau. H. überarbeitete, ergänzte und erweiterte Auflage. Verlagsunion Agrar, Bochum.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). Semiparametric Regression. London: Cambridge University Press.

Rushton, G., M. Armstrong, J. Gittler, B. Greene, C. Pavlik, M. West, and D. Zimmerman (2007). Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. Taylor & Francis.

Sahlins, M. (1968). Notes on the original affluent society. In R. Lee and I. DeVore (Eds.), Man the Hunter, pp. 85–89. New York: Aldine.

Sahlins, M. D. (1972). Stone Age Economics. London: Transaction Publishers.

Salazar-García, D. C., M. P. Richards, O. Nehlich, and A. G. Henry (2014). Dental calculus is not equivalent to bone collagen for isotope analysis: a comparison between carbon and nitrogen stable isotope analysis of bulk dental calculus, bone and dentine collagen from same individuals from the Medieval site of El Raval (Alicante, Spain). Journal of Archaeological Science 47, 70–77.

Sangmeister, E. and K. Gerhardt (1965). Schnurkeramik und Schnurkeramiker in Südwestdeutschland. Badische Fundberichte Sonderheft, Staatliches Amt für Ur-und Frühgeschichte 8.

Saß, A., S. Wurm, and T. Ziese (2009). Somatische und psychische Gesundheit. In K. Böhm, C. Tesch-Römer, and T. Ziese (Eds.), Beiträge zur Gesundheitsberichterstattung des Bundes - Gesundheit und Krankheit im Alter, pp. 31–61. Berlin: Robert Koch Institute with Federal Statistical Office Germany and German Centre of Gerontology.

Sauter, F., L. Puchinger, and U.-D. Schoop (2003). Studies in organic archaeometry. VI1. Fat analysis sheds light on everyday life in prehistoric Anatolia: traces of lipids identified in Chalcolithic potsherds excavated near Boğazkale, Central Turkey. Arkivoc 15, 15–21.

Scarrott, C. J. and Y. Hu (2014). evmix: Extreme value mixture modelling, threshold estimation and boundary corrected kernel density estimation. Available on CRAN.

Schaafsma, G. (2000). The protein digestibility–corrected amino acid score. The Journal of Nutrition 130(7), 1865–1867.

Schafberg, R. (1999). Die linienbandkeramischen Bestattungen vom Otmarsweg in Naumburg (Saale), Ldkr. Burgenlandkreis. Jahresschrift für mitteldeutsche Vorgeschichte 81, 61–79.

Schafer, D. W. (2001). Semiparametric maximum likelihood for measurement error model regression. Biometrics 57(1), 53–61.

Scheibner, A. (2015). Prähistorische Ernährung in Vorderasien und Europa: eine Synopse der Quellen. Ph. D. thesis, Freie Universität Berlin.

Schier, W. (2009). Extensiver Brandfeldbau und die Ausbreitung der neolithischen Wirtschaftsweise in Mitteleuropa und Südskandinavien am Ende des 5. Jahrtausends v. Chr. Praehistorische Zeitschrift 84(1), 15–43.

Schilz, F. (2006). Molekulargenetische Verwandtschaftsanalysen am prähistorischen Skelettkollektiv der Lichtensteinhöhle. Ph. D. thesis, Universität Göttingen.

Schlagenhauf, P. (2004). Malaria: from prehistory to present. Infectious Disease Clinics of North America 18(2), 189–205.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. American Psychologist 47(10), 1173.

Schmidt, K., R. Bindl, and H. Bruchhaus (2007). Körperhöhenschätzung an ausgewählten neolithischen und bronzezeitlichen Skeletten. Archäologische Informationen 30(1), 51–69.

Schneeweiß, H. and J. Komlos (2009). Probabilistic rounding and Sheppard's correction. Statistical Methodology 6(6), 577–593.

Schoop, U.-D. (2009). Ausgrabungen in Çamlıbel Tarlası 2009. In A. Schachner (Ed.), Die Ausgrabungen in Bogazköy-Hattuša, pp. 56–69. Archäologischer Anzeiger.

Schoop, U.-D., P. Grave, L. Kealhofer, and G. Jacobsen (2009). Radiocarbon dates from Chalcolithic Çamlıbel Tarlası. In A. Schachner (Ed.), Die Ausgrabungen in Bogazköy-Hattuša, pp. 56–69. Archäologischer Anzeiger.

Schuldt, E. (1961). Abschliessende Ausgrabungen auf dem jungsteinzeitlichen Flachgräberfeld von Ostorf 1961. Jahrbuch für Bodendenkmalpflege in Mecklenburg 1961, 131–178.

Schuster, W. H., J. Alkämper, R. Marquard, A. Stählin, and L. Stählin (2000). Leguminosen zur Kornnutzung. Giessener Beiträge zur Enwicklungsforschung,Reihe 2 Monographien 11.

Schwidetzky, I. (1973). Die Anfänge des Neolithikums vom Orient bis Nordeuropa, Anthropologie Teil 1 ,VIIIa. Köln, Wien: Böhlau Verlag.

Scott, D. W. (2009). Multivariate Density Estimation: Theory, Practice, and Visualization. New York: Wiley.

Scott, D. W. and S. J. Sheather (1985). Kernel density estimation with binned data. Communications in Statistics - Theory and Methods 14(6), 1353–1359.

Scott, G. R. and S. R. Poulson (2012). Stable carbon and nitrogen isotopes of human dental calculus: a potentially new non-destructive proxy for paleodietary analysis. Journal of Archaeological Science 39(5), 1388–1393.

Sellen, D. W. and D. B. Smay (2001). Relationship between subsistence and age at weaning in "preindustrial" societies. Human Nature 12(1), 47–87.

Şenyürek, M. (1949). Truva Civarında Kumtepe'de Bulunmuş Olan İskeletlere Dair Bir Not. Coğrafya Fakültesi Dergisi 2, 295–304.

Şenyürek, M. (1950). Study of a skeleton of a Chalcolithic warrior from Büyük Güllücek. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi 8(3), 290–310.

Şenyürek, M. (1951). A Note on the Human Skeletons in the Alaca Höyük Museum. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi 9, 43–61.

Şenyürek, M. (1958). Study of a human skeleton found in Öküzini in the Province of Antalya. Belleten 22, 491–516.

Senyurek, M. and S. Tunakan (1951). The skeletons from Şeyh Höyük. Belleten 15(60), 431–45.

Sheather, S. and C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 53(3), 683–690.

Sheather, S. J. (2004, 11). Density estimation. Statistical Science 19(4), 588–597.

Shields, M., S. C. Gorber, and M. S. Tremblay (2008). Estimates of obesity based on self-report versus direct measures. Health Rep 19(2), 61–76.

Siegmund, F. (2010). Die Körpergröße der Menschen in der Ur- und Frühgeschichte Mitteleuropas und ein Vergleich ihrer anthropologischen Schätzmethoden. Books on Demand, Norderstedt 2010.

Siegmund, F. (2012). Wegleitung körperhöhenschätzung [guidelines for stature estimation]. Bulletin der Schweizerischen Gesellschaft für Anthropologie 18(2), 25–35.

Sienkiewicz-Szłapka, E., B. Jarmołowska, S. Krawczuk, E. Kostyra, H. Kostyra, and M. Iwan (2009). Contents of agonistic and antagonistic opioid peptides in different cheese varieties. International Dairy Journal 19(4), 258–263.

Silventoinen, K. (2003, 4). Determinants of variation in adult body height. Journal of Biosocial Science 35, 263–285.

Silverman, B. (1986). Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. London, New York: Taylor & Francis.

Sjøvold, T. (1990). Estimation of stature from long bones utilizing the line of organic correlation. Human Evolution 5(5), 431–447.

Sjøvold, T. (1988). Geschlechtsdiagnose am Skelett. In R. Knußmann (Ed.), Anthropologie: Handbuch der vergleichenden Biologie des Menschen. Stuttgart: Fischer.

Smits, E. and G. Maat (1996). An early/middle Bronze Age common grave of Wassenaar, the Netherlands. The physical anthropological results. Analecta Praehistorica Leidensia 26(2), 21–28.

Smits, L. and L. Kooijmans (2006). Graves and human remains. Schipluiden: a Neolithic settlement on the Dutch North Sea coast, 3500 cal. BC. Annalecta Praehistorica Leidensia 37/38, 91–112.

Sosulski, F. and G. Sarwar (1973). Amino acid composition of oilseed meals and protein isolates. Canadian Institute of Food Science and Technology Journal 6(1), 1–5.

Speer, C. P. and M. Gahr (2009). Pädiatrie. Heidelberg: Springer.

Spengler, R., M. Frachetti, P. Doumani, L. Rouse, B. Cerasetti, E. Bullion, and A. Mar'yashev (2014). Early agriculture and crop transmission among Bronze Age mobile pastoralists of central Eurasia. Proceedings of the Royal Society of London B: Biological Sciences 281(1783), 201–382.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(4), 583–639.

Spiegelman, D., A. McDermott, and B. Rosner (1997). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. The American Journal of Clinical Nutrition 65(4), 1179–1186.

Steckel, R. H. (1995). Stature and the standard of living. Journal of Economic Literature 33(4), 1903–1940.

Steckel, R. H. (2003). Research project: A history of health in Europe from the late paleolithic era to the present. Economics & Human Biology 1(1), 139–142.

Stefanski, L. A. and R. J. Carroll (1990). Deconvolving kernel density estimators. Statistics: A Journal of Theoretical and Applied Statistics 21(2), 169–184.

Stevenson, P. H. (1929). On racial differences in stature long bone regression formulae, with special reference to stature reconstruction formulae for the Chinese. Biometrika 21, 303–321.

Suárez López, M., A. Kizlansky, and L. B. López (2006). Evaluación de la calidad de las proteínas en los alimentos calculando el escore de aminoácidos corregido por digestibilidad. Nutrición hospitalaria 21(1), 47–51.

Szalai, J. (1999). Anthropologische Untersuchung der Skelette und der Reste von Leichenbränden aus den früh- und mittelbronzezeitlichen Gräberfeldern von Battonya. In Früh-und mittelbronzezeitliche Gräberfelder von Battonya, pp. 125–163. Budapest: Magyar Nemzeti Múzeum.

Szilvássy, J. (1988). Altersdiagnose am Skelett. In R. Knußmann (Ed.), Anthropologie: Handbuch der vergleichenden Biologie des Menschen. Stuttgart: Fischer.

Szilvássy, J. and H. Kritscher (1991). Preparation, reconstruction and interpretation of seven human skeletons from the late Bronze Age (urn-field-culture) found at a storage pit in Stillfried/March, Lower Austria. Anthropologischer Anzeiger 49(4), 303–324.

Tafuri, M. A., O. E. Craig, and A. Canci (2009). Stable isotope evidence for the consumption of millet and other plants in Bronze Age Italy. American Journal of Physical Anthropology 139(2), 146–153.

Tanno, K.-i. and G. Willcox (2006). The origins of cultivation of Cicer Arietinum L. and Vicia Faba L.: early finds from Tell el-Kerkh, north-west Syria, late 10th millennium BP. Vegetation History and Archaeobotany 15(3), 197–204.

Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. The Annals of Statistics 29(1), 66–93.

Taylor, A. W., E. D. Grande, T. K. Gill, C. R. Chittleborough, D. H. Wilson, R. J. Adams, J. F. Grant, P. Phillips, S. Appleton, and R. E. Ruffin (2006). How valid are self-reported height and weight? A comparison between CATI self-report and clinic measurements using a large cohort study. Australian and New Zealand Journal of Public Health 30(3), 238–246.

Telkkä, A. (1950). On the prediction of human stature from the long bones. Cells Tissues Organs 9(1-2), 103–117.

Teschler-Nicola, M. (1986). Soziale und biologische Differenzierung in der frühen Bronzezeit am Beispiel des Gräberfeldes F von Gemeinlebarn, Niederösterreich. Annalen des Naturhistorischen Museums in Wien. Serie A für Mineralogie und Petrographie, Geologie und Paläontologie, Anthropologie und Prähistorie 90, 135–145.

Tiefenböck, B. E. (2010). Die krankhaften Veränderungen an den linearbandkeramischen Skelettresten von Kleinhadersdorf, NÖ. Ph. D. thesis, Universität Wien.

Topinard, P. et al. (1885). Procédé de mensuration des os longs, dans le but de reconstituer la taille. Bulletins de la Société d'anthropologie de Paris 8(1), 73–83.

Triantaphyllou, S., M. P. Richards, C. Zerner, and S. Voutsaki (2008). Isotopic dietary reconstruction of humans from Middle Bronze age Lerna, Argolid, Greece. Journal of Archaeological Science 35(11), 3028–3034.

Trotter, M. and G. Gleser (1952). Estimation of stature from long bones of American Whites and Negroes. American Journal of Physical Anthropology 10(4), 463.

Trotter, M. and G. C. Gleser (1958). A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. American Journal of Physical Anthropology 16(1), 79–123.

Tunakan, S. (1964). Bodrum-Dirmil Kazısı İskeletleri. Belleten 28(111), 361–371.

Tunakan, S. (1965). Türk Tarih Kurumu Adına, 1964 Yazında, Alaca-Höyük'te Yapılan Kazıda Çıkarılan İki Eski Bronz Çağı İskeletinin İncelenmesi. Belleten 29(116), 571–584.

Ullrich, H. (1961). Anthropologische Untersuchungen am Skelettmaterial eines Aunjetitzer Gräberfeldes von Großbrembach, besonders im Hinblick auf die Frage nach der Herkunft der mitteldeutschen Auntjetitzer. Ph. D. thesis, Freie Universität Berlin.

United States Department of Agriculture, U. (2012). National nutrient database. `http://ndb.nal.usda.gov/` (21.07.2014).

Usydus, Z., J. Szlinder-Richert, and M. Adamczyk (2009). Protein quality and amino acid profiles of fish products available in poland. Food Chemistry 112(1), 139–145.

Utts, J. and R. Heckard (2014). Mind on Statistics. Boston, United States: Cengage Learning.

Valamoti, S. M., A. Moniaki, and A. Karathanou (2011). An investigation of processing and consumption of pulses among prehistoric societies: archaeobotanical, experimental and ethnographic evidence from Greece. Vegetation History and Archaeobotany 20(5), 381–396.

Vallois, H. and D. Ferembach (1962). Les restes humains de Ras Shamra et de Minet el-Beida: étude anthropologique. In Ugaritica: Découvertes des XVIIIe et XIXe campagnes, 1954-1955 fondements préhistoriques d'Ugarit et nouveaux sondages études anthropologiques porteries grecques et monnaies islamiques de Ras Shamra et environs, Volume 4, pp. 565–630. Paris.

Vallois, H. V. (1937). Note sur les ossements humains de la nécropole énéolithique du Byblos., Volume 1. Imprimerie de l'Institut français d'archéologie orientale, Bulletin du Musée de Beyrouth.

Vallois, H. V. (1952). Diagrammes sagittaux et mensurations individuelles des hommes fossiles d'Afalou-Bou-Rhummel: Travaux du Laboratoire d'anthropologie et d'archéologie préhistoriques du Musée du Bardo, Volume 5. Algier: Gouvernement général de l'Algérie, Dir. de l'intérieur et des beaux-arts, Service des antiquités.

VanWey, L. K., R. R. Rindfuss, M. P. Gutmann, B. Entwisle, and D. L. Balk (2005). Confidentiality and spatially explicit data: Concerns and challenges. Proceedings of the National Academy of Sciences of the United States of America 102(43), 15337–15342.

Vaupel, P. and H. K. Biesalski (2010). Proteine. In H. K. Biesalski, S. C. Bischoff, and C. Puchstein (Eds.), Ernährungsmedizin: Nach dem Curriculum Ernährungsmedizin der Bundesärztekammer und der DGE, pp. 109–131. Stuttgart, New York: Georg Thieme Verlag.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 36(3), 1–48.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The german socio-economic panel study (soep)-evolution, scope and enhancements. Schmollers Jahrbuch 127, 139–167.

Wahl, J. (2010). Tell Chuera - Die menschlichen Skelettreste aus dem 3. und 2. Jahrtausend v. Chr. In J.-W. Meyer, R. Hempelmann, and C. Falb (Eds.), Tell Chuera: Vorberichte zu den Grabungskampagnen 1998–2005: Vorderasiatische Forschungen der Max Freiherr von Oppenheim-Stiftung, Volume 2. Wiesbaden: Harrassowitz.

Wand, M. and M. Jones (1994). Multivariate plug-in bandwidth selection. Computational Statistics 9(2), 97–116.

Wang, B. and W. Wertelecki (2013). Density estimation for data with rounding errors. Computational Statistics & Data Analysis 65, 4–12.

Wang, H. and D. F. Heitjan (2008). Modeling heaping in self-reported cigarette counts. Statistics in Medicine 27(19), 3789–3804.

Wang, H., S. Shiffman, S. D. Griffith, and D. F. Heitjan (2012). Truth and memory: Linking instantaneous and retrospective self-reported cigarette consumption. The Annals of Applied Statistics 6(4), 1689–1706.

Wapler, U., E. Crubezy, and M. Schultz (2004). Is cribra orbitalia synonymous with anemia? Analysis and interpretation of cranial pathology in Sudan. American Journal of Physical Anthropology 123(4), 333–339.

Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby (2006). Bivariate line-fitting methods for allometry. Biological Reviews 81(02), 259–291.

Wasmer, M., E. Scholz, M. Blohm, et al. (2007). Konzeption und Durchführung der"Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften"(ALLBUS) 2008. GESIS-ZUMA.

Wason, P. K. (2004). The Archaeology of Rank. Cambridge University Press.

Watson, P. J. and S. LeBlanc (1990). Girikihaciyan: a Halafian site in southeastern Turkey, Volume 33. Los Angeles: Cotsen Institute of Archaeology.

Weedon, M. N. and T. M. Frayling (2008). Reaching new heights: insights into the genetics of human stature. Trends in Genetics 24(12), 595–603.

Weiss, E., M. Kislev, and A. Hartmann (2006). Autonomous cultivation before domestication. Science 312(5780), 1608–1610.

Weiss, E. and D. Zohary (2011). The Neolithic southwest Asian founder crops. Current Anthropology 52(S4), 237–254.

White, M., M. Hathaway, W. Dayton, T. Henderson, and A. Lepine (1999). Comparison of insulin-like growth factor-I concentration in mammary secretions and serum of small-and giant-breed dogs. American Journal of Veterinary Research 60(9), 1088–1091.

Wiley, A. S. (2009). Consumption of milk, but not other dairy products, is associated with height among us preschool children in NHANES 1999–2002. Annals of Human Biology 36(2), 125–138.

Williams, P. (2007). Nutritional composition of red meat, nutrition and dietetics. Nutrition & Dietetics 64, 13–119. Suppl. 4.

Wittwer-Backofen, U. (1985). Anthropologische Untersuchungen der Nekropole İkiztepe/Samsun. Arastirma Sonuclari Toplantisi 3, 421–428.

Wittwer-Backofen, U. (1987). Anthropological study of the skeletal material from Lidar. Arastirma Sonuclari Toplantisi 5(2), 191–202.

Wittwer-Backofen, U. and H. Kiesewetter (1997). Menschliche Überreste der Neuen Ausgrabungen in Troia – Funde der Kampagne 1989-1995. Studia Troica 7, 509–537.

Woidich, M. (2002). Westliche Kugelamphorenkultur. Die auf multivariate Statistik und GIS gestützte Analyse einer archäologischen Kultur. Ph. D. thesis, Freie Universität Berlin.

Wood, S. N. (2003). Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65(1), 95–114.

Wood, S. N. (2006). Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. Biometrics 62(4), 1025–1036.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1), 3–36.

World Health Organization (2005). Preventing chronic diseases: a vital investment. `http://whqlibdoc.who.int/publications/2005/9241563001_eng.pdf`.

Wurm, H. (1985). History of the determination body height from skeletal findings (body height determination for men). the proposed methods of body height determination from skeletal findings since the middle of the 20th century. Gegenbaurs morphologisches Jahrbuch 131(3), 383–432.

Wurm, H. (1986). Zur Geschichte der Körperhöhenschätzmethoden nach Skelettfunden. Vorschläge zur Körperhöhenschätzung nach Skelettfunden bis zur Mitte des 20. Jahrhunderts. Anthropologischer Anzeiger 44, 149–167.

Xu, S. (2014). Asymmetric kernel density estimation based on grouped data with applications to loss model. Communications in Statistics-Simulation and Computation 43(3), 657–672.

Yordanov, Y. and B. Dimitrova (2002). Results of an anthropological study of human skeletal remains of the prehistoric necropolis in the vicinity of Durankulak. In H. Todorova (Ed.), Durankulak: Die Prähistorischen Gräberfelder, Volume 2. Sofia: Deutsches Archäologisches Institut in Berlin.

Young, V. R. and P. L. Pellett (1994). Plant proteins in relation to human protein and amino acid nutrition. The American Journal of Clinical Nutrition 59(5), 1203–1212.

Zero, D. T., M. Fontana, E. A. Martínez-Mier, A. Ferreira-Zandoná, M. Ando, C. González-Cabezas, and S. Bayne (2009). The biology, prevention, diagnosis and

treatment of dental caries: scientific advances in the United States. The Journal of the American Dental Association 140, 25–34.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. The Annals of Statistics 18(2), 806–831.

Zhang, X., M. L. King, and R. J. Hyndman (2006). A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Computational Statistics & Data Analysis 50(11), 3009–3031.

Zimmermann, A., J. Hilpert, and K. P. Wendt (2009). Estimations of population density for selected periods between the Neolithic and AD 1800. Human Biology 81(3), 357–380.

Zinn, S. and A. Wuerbach (2015). A statistical approach to address the problem of heaping in self-reported income data. Journal of Applied Statistics, forthcoming.

Zioudrou, C., R. A. Streaty, and W. A. Klee (1979). Opioid peptides derived from food proteins. The exorphins. Journal of Biological Chemistry 254(7), 2446–2449.

Živanović, S. (1975). A note on the anthropological characteristics of the Padina population. Zeitschrift für Morphologie und Anthropologie 66, 161–175.

Zoffmann, Z. (1968). An Anthropological Study of the Neolithic Cemetery at Villanykövesd (Lengyel Culture), Hungary. Különnyomat a Janus Pannonius Múzeum 13, 25–38.

Zohary, D., M. Hopf, and E. Weiss (2012). Domestication of Plants in the Old World: The origin and spread of domesticated plants in southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press.

Zougab, N., S. Adjabi, and C. Kokonendji (2014). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. Computational Statistics & Data Analysis 75(11), 28–38.

Anhang

Kurzfassungen in englischer Sprache

Abstract: Estimating the Density of Ethnic Minorities and Aged People in Berlin: Multivariate Kernel Density Estimation Applied to Sensitive Geo-Referenced Administrative Data Protected via Measurement Error

Modern systems of official statistics require the timely estimation of area-specific densities of sub-populations. Ideally estimates should be based on precise geo-coded information, which is not available due to confidentiality constraints. One approach for ensuring confidentiality is by rounding the geo-coordinates. We propose multivariate non-parametric kernel density estimation that reverses the rounding process by using a measurement error model. The methodology is applied to the Berlin register of residents for deriving density estimates of ethnic minorities and aged people. Estimates are used for identifying areas with a need for new advisory centres for migrants and infrastructure for older people.

Abstract: Kernel Density Estimation for Heaped Data

In self-reported data usually a phenomenon called 'heaping' occurs, i.e. survey participants round the values of variables such as income, weight or height to some degree. Additionally, respondents may be more prone to round off or up due to social desirability. By ignoring the heaping process spurious spikes and bumps are introduced when applying kernel density methods naively to the rounded data. A generalized Stochastic Expectation-Maximization (SEM) approach accounting for heaping with potentially asymmetric rounding behaviour in univariate kernel density estimation is presented in this work. The introduced methods are applied to survey data of the German Socio-Economic Panel and exhibit very good performance in simulations.

Abstract: Back to Good Shape: Biological Standard of Living in the Copper and Bronze Ages and the Possible Role of Food

Body height has been put forward in the field of economics as a substitute measure of welfare in times and regions when usual proxies, such as GDP, are not reliable or available. Using skeletal remains, this concept can also be applied to archaeology. Taking earlier approaches (e.g. Jaeger et al. 1998; Koepke and Baten 2008; Siegmund 2010),

a database of relevant long bone measurements is currently being compiled within the framework of a research project entitled "Living conditions and biological standard of living in the prehistory of Europe and Southwest Asia (LiVES)", involving prehistoric archaeology, prehistoric anthropology and statistics. In addition to general methodological problems arising from the estimation and comparison of archaeological body height data as a proxy for the biological standard of living, research in later prehistory is additionally hampered by the increasing occurrence of collective burials where, in many cases, taphonomy and excavation techniques blur the skeletal contexts. Furthermore, cremation has led to the destruction of relevant parts of the skeletons. When properly circumvented using statistics, however, a preliminary sample shows specific diachronic as well as regional trends in the Neolithic and the Bronze Age when comparing the Near East and Europe. They suggest that the Old World body height pattern of tall northerners and short southerners, known since the Roman Period, has not been a constant phenomenon, but probably evolved during the Copper Age and the Bronze Age after a general Neolithic body height decline. Looking for causalities, analyses have to consider nutrition, work and disease load, as well as social status in multivariate models. They also have to overcome problems in the operationalization and interconnection of independent and dependent data. Moreover, possible genetically encoded metabolic changes in, for example, starch or lactose digestion in some prehistoric populations have to be taken into consideration.

Abstract: Modeling Body Height in Prehistory Using a Spatio-Temporal Bayesian Errors-in-Variables Model

Body height is commonly employed as a proxy variable for living standards among human populations. In the following, the human standard of living in prehistory will be examined using body height as reconstructed through long bone lengths. The aim of this work is to model the spatial dispersion of body height over the course of time for a large archeological long bone dataset. A major difficulty in the analysis is the fact that some variables in the data are measured with uncertainty, like the date, the sex and the individual age of the available skeletons. As the measurement error processes are known in this study, it is possible to correct this using so-called errorsin-variables models. Motivated by this dataset, a Bayesian additive mixed model with errors-in-variables is proposed, which fits a global spatio-temporal trend using a tensor product spline approach, a local random effect for the archeological sites and corrects for mismeasurement and misclassification of covariates. In application to the data, the model reveals long-term spatial trends in prehistoric living standards.

Abstract: Reconstruction of Body Height from Long Bones for Prehistoric Individuals: New Methodological Concepts

The variety of estimation formulas available in the physical anthropological literature leads to vastly different results if stature is reconstructed from long bone lengths. Despite the existence of several overviews giving general guidelines on this topic, the

selection of a particular formula as well as the selection of the long bone measurement or set of measurements to be used within this formula, is currently left to the individual researcher's judgement. A more formal perspective on this problem is therefore indicated. The new approach presented here introduces two statistical concepts: Model averaging by Akaike weights establishes a verifiable criterion for the selection of the best combination of long bone measurements within a chosen formula set such as Pearson (1899), whereas statistical meta-analysis can be used to create a new set of formulas for stature estimation of prehistoric individuals from the multitude of formula sets that have been published until now. An optimal weighting scheme for the formula set of Pearson (1899) is computed using Akaike model averaging. Furthermore, new universal formulas for stature estimation are derived by combining four frequently used stature estimation formulas. As an example, the two strategies described here are applied to a small subsample of skeletal finds from a Central European prehistoric population (early Neolithic Linear Pottery Culture (LBK), ca. 5700-5000 cal BC). It is shown that by such aggregation of information an improved and more reliable estimation of stature can be obtained. Both concepts lead to more balanced stature estimates and can provide a way out of the dilemma of the largely subjective applicant's choice of formula.

Kurzfassungen in deutscher Sprache

Zusammenfassung: Estimating the Density of Ethnic Minorities and Aged People in Berlin: Multivariate Kernel Density Estimation Applied to Sensitive Geo-Referenced Administrative Data Protected via Measurement Error

Moderne Systeme der amtlichen Statistik erfordern die zeitnahe Einschätzung flächenspezifischer Dichten von Teilpopulationen. Im Idealfall sollten sich Schätzungen auf exakte, geocodierte Informationen stützen, die jedoch aufgrund Vertraulichkeit bzw. datenschutzrechtlicher Bedenken nicht zur Verfügung stehen. Ein möglicher Ansatz für die Wahrung der Vertraulichkeit ist das Runden der Geokoordinaten. Wir schlagen eine multivariate nichtparametrische Kerndichteschätzung vor, die den Rundungsprozess mit Hilfe eines Messfehlermodells umkehrt. Die Methodik wird auf das Berliner Melderegister angewandt um Dichteschätzungen von ethnischen Minderheiten und älteren Menschen in Berlin zu erhalten. Diese Schätzungen werden für die Ermittlung von Regionen mit einem Bedarf an neuen Beratungsstellen für Bewohner mit Migrationshintergrund sowie geeigneter Infrastruktur für ältere Menschen genutzt.

Zusammenfassung: Kernel Density Estimation for Heaped Data

In Umfragen mit Selbstangaben tritt der Regel ein Phänomen namens "Heaping" auf, d.h. Umfrageteilnehmer runden die Werte ihres Einkommens, Gewichts oder ihrer Körperhöhe zu einem gewissen Grad. Darüber hinaus können die Befragten aufgrund

von sozialer Erwünschtheit anfälliger für Auf- oder Abrunden sein. Wenn das Verfahren der Kerndichteschätzung naiv, d.h. unter Ignorierung des Rundungsprozesses, auf die Daten angewandt wird, treten vermehrt Ausbuchtungen bzw. Modi an den Häufungswerten auf. Ein verallgemeinerter stochastischer Expectation-Maximization-Ansatz (SEM) zur Berücksichtigung von gehäuften Daten mit potenziell asymmetrischem Rundungsverhalten in univariater Kerndichteschätzung wird in dieser Arbeit vorgestellt. Diese Methoden wird auf Daten des Sozio-Oekonomischen Panels (SOEP) angewandt und zeigt sehr gute Ergebnisse in Simulationen.

Zusammenfassung: Back to Good Shape: Biological Standard of Living in the Copper and Bronze Ages and the Possible Role of Food

Körperhöhe wird in den Wirtschaftswissenschaften in Zeiten und Regionen, in welchen übliche Proxies wie das BIP, nicht zuverlässig oder verfügbar sind, zunehmend als Ersatzmaß für den Lebensstandard genutzt. Unter Verwendung von Skelettüberresten kann dieses Konzept auch auf die Archäologie übertragen werden. Unter Nutzung von früheren Ansätzen (z.B. Jaeger et al. 1998; Koepke and Baten 2008; Siegmund 2010) wird derzeit eine Datenbank mit relevanten Langknochenmessungen im Rahmen eines Forschungsprojekts "Lebensbedingungen und biologischer Lebensstandard in der Vorgeschichte Europas und Südwestasien (LiVES)", welches prähistorische Archäologie, prähistorische Anthropologie und Statistik miteinander verknüpft, zusammengestellt. Zusätzlich zu den allgemeinen methodischen Problemen bei der Schätzung und dem Vergleich archäologischer Körperhöhendaten als Proxy für den biologische Lebensstandard wird die Forschung in der späteren Vorgeschichte durch das zunehmende Auftreten von kollektiven Gräbern, wodurch in vielen Fällen die individuellen Skelettkontexte durch Ausgrabungstechniken verwischt werden, zusätzlich behindert. Darüber hinaus hat die Bestattungspraxis der Einäscherung oft relevante Teile der Skelette zerstört. Wenn dies mit modernen statistischen Methoden an einem vorläufigen Datenssatz berücksichtigt wird, zeigen sich jedoch diachrone sowie regionale Trends in der Jungsteinzeit und der Bronzezeit beim Vergleich des Nahen Osten und Europa. Sie lassen vermuten, dass das Verteilungsmuster der Körperhöhe, welches seit der römischen Zeit bekannt war – also große Menschen im Norden und kleinere im Süden Europas – keine ständige Erscheinung war, sondern sich wahrscheinlich während der Kupferzeit und der Bronzezeit nach einer allgemeinen Rückgang der Körperhöhe in der Jungsteinzeit entwickelte. Auf der Suche nach Kausalitäten sollten Analysen die Ernährung, Arbeits- und Krankheitsbelastung sowie den sozialen Status in statistischen Modellen berücksichtigen. Diese Modelle haben auch die Problemen bei der Operationalisierung und Assoziation von unabhängigen und abhängigen Daten überwunden. Außerdem müssen genetisch kodierte metabolischen Veränderungen z.B. bei der Verdauung von Lactose oder Stärke in einigen prähistorischen Populationen berücksichtigt werden.

Zusammenfassung: Modeling Body Height in Prehistory Using a Spatio-Temporal Bayesian Errors-in-Variables Model

Körperhöhe wird häufig als Proxy-Variable für den Lebensstandard menschlicher Populationen verwendet. Im Folgenden wird der menschliche Lebensstandard in der Vorgeschichte unter Verwendung von Körperhöhe, welche durch Lanknochen rekonstruiert wurde, untersucht. Das Ziel dieser Arbeit ist es, die räumliche Verteilung der Körpergröße im Laufe der Zeit mithilfe eines großen archäologischen Langknochen-Datensatzes zu modellieren. Eine Hauptschwierigkeit bei der Analyse ist die Tatsache, dass einige Variablen wie die chronologische Einordnung, das Geschlecht oder das Alter der Skelette nur mit einer gewissen Unsicherheit gemessen wurden. Da die Messfehlerprozesse in dieser Untersuchung bekannt sind, ist es möglich dies mithilfe von sogenannten Fehler-in-den-Variablen-Modellen zu korrigieren. Es wird ein Bayesianisches additiv gemischtes Modell mit Fehler-in-den-Variablen vorgeschlagen, welches eine globale räumlich-zeitliche Entwicklung unter Verwendung eines Tensor-Produkt-Spline Ansatzes, eines lokalen Zufallseffektes für die archäologischen Fundstätten sowie auf Messfehler und Missklassifikation von Kovariaten korrigiert. In der Anwendung auf die Daten macht das Modell langfristige räumliche Trends bezüglich des prähistorischen Lebensstandards sichtbar.

Zusammenfassung: Reconstruction of Body Height from Long Bones for Prehistoric Individuals: New Methodological Concepts

Die große Vielfalt der Schätzformeln zur Körperhöhenbestimmung anhand von Langknochen in der physisch-anthropologischen Literatur führt teilweise zu sehr unterschiedlichen Ergebnissen. Für den Anwender stellt sich daher die Frage, welche Langknochen und welche Formel für die Körperhöhenschätzung auszuwählen sind. Trotz der Existenz von mehreren systematischen Übersichtsarbeiten, welche allgemeine Empfehlungen zu diesem Thema geben, soll diese Arbeit das Problem aus statistischer Sicht und weniger aus inhaltlicher betrachten. Zwei neue methodische Konzepte werden dabei eingeführt: “model-averaging” durch AIC-Gewichte und statistische Meta-Analyse. Erstere etabliert ein verifizierbares Gewichtungsschema für eine gegebene Menge von Formeln, wie z.B. von Pearson (1899) publiziert, während die Meta-Analyse neue Universalformeln aus gegebenen Formeln verschiedener Quellen zur Körperhöhenschätzung erzeugt. Es wird gezeigt, dass durch eine solche Aggregation von Informationen eine verbesserte und zuverlässigere Abschätzung der Körperhöhe erreicht werden kann. Neue Formeln für die Körperhöhenschätzung von prähistorischen Menschen werden in diesem Zusammenhang vorgestellt und auf prähistorische Skelette einer zentraleuropäischen Population (Linearbandkeramik, ca. 5700-5000 cal BC) angewandt.