



GROWTH: Functional Data Analysis of the human height postnatal growth in an on-line application

Ondřej Klíma¹ – Martin Čuta² – Lenka Polcerová² – Pavel Zemčík¹ – Anna Škultétyová² – Dominik Černý² – Miroslav Králík^{2*}

1 IT4Innovations Centre of Excellence, Brno University of Technology, Božetěchova 1/2, 612 66 Brno, CZ, Czech Republic

2 Laboratory of Morphology and Forensic Anthropology (LaMorFA), Department of Anthropology, Faculty of Science, Masaryk University, Kotlářská 267/2, Brno 61137, Czech Republic

* Corresponding author, Miroslav Králík: 18313@muni.cz

Received 14th November 2021; accepted 15th December 2021

GROWTH: ON-LINE APLIKACE PRO FUNKČNÍ ANALÝZU DAT POSTNATÁLNÍHO RŮSTU VÝŠKY POSTAVY ČLOVĚKA

ABSTRAKT Sledování individuálního růstu je v pediatrické praxi vyžadováno denně a často je doprovázeno potřebou podrobnějších analýz. Analýzu lidského růstu potřebují také sportovní antropologové a výzkumníci v oblasti biologie člověka. Přínosem by tedy byla pokročilá a zároveň snadno použitelná a bezplatná aplikace, která by pediatrům, auxologům a výzkumným pracovníkům v oblasti biologie člověka umožňovala provádět hloubkovou analýzu postnatálního růstu. Aplikace GROWTH byla vyvinuta na základě pochopení biologických procesů lidského růstu a matematických přístupů, které poskytují nejvhodnější model pro individuální (longitudinální) empirická data. Aplikace je navržena tak, aby ji bylo možné používat v každodenní pediatrické praxi. Poskytuje lékařům nástroje pro sledování růstu, předpovídání dosažené výšky a diagnostiku patologických růstových vzorců. Pokročilá analýza zahrnuje odhad časování hlavních růstových milníků. Současná verze je vyvrcholením několikastupňového vývoje aplikace a je založena na metodě FPCA (funkční analýza hlavních komponent) s numerickou optimalizací. Výstupní parametry jsou snadno použitelné a zobrazují se numericky i graficky.

KLÍČOVÁ SLOVA tělesná výška; postnatální růst; modelování růstu; růstová křivka; funkční analýza dat

ABSTRACT Monitoring of individual growth is required on daily basis in paediatric practice, often accompanied by need for more in-depth analyses. Sports anthropologists and researchers in the field of human biology also need to analyze human growth. To allow clinicians and researchers to monitor human growth and analyze it in advanced manner, easy-to-use free application would be beneficial. The application, GROWTH, was developed based on an understanding of the biological processes of human growth and mathematical approaches that provide the most appropriate model for individual (longitudinal) empirical data. The application is designed to be used in daily paediatric practice. It provides clinicians with tools to monitor growth, predict attained height and diagnose pathological growth patterns. Advanced analysis includes the timing of major growth milestones. The current version is the culmination of a multi-stage development of the application and is based on Functional Principal Components Analysis method with numerical optimization. The output parameters are easy to use and are displayed both numerically and graphically.

KEY WORDS body height; postnatal growth; growth modelling; growth curve; Functional Data Analysis

INTRODUCTION

From an evolutionary perspective, every living individual strives to adapt optimally to its environment. Ontogenesis—individual development and growth—can be understood as

a long-term mechanism of this adaptation (Gilbert 2012; Gilbert – Epel 2015), in which an organism grows and matures, realizing the form of characteristics (body, behavior) that its internal and external resources and constraints—the characteristics of its genes (from its parents) and the conditions of its

external environment (womb, family, school, society)—allow. Since each person inherits a different combination of hereditary traits from his/her parents and also develops in a more or less different environment from others, each person has his/her own optimal developmental path. Thus, there is not one correct developmental path for all that can be considered the norm, but there are as many paths as there are people. So, when we evaluate a characteristic of an individual (body size and shape, behavioral form, etc.), each person has an individual “size” of the observed characteristic, an individual timing of developmental events (i.e., when each stage of their development begins), and an individual rate of progression of each stage (how long it takes for the stage to occur). Differences between individuals in these three values of developmental trajectories reflect individual-specific “solutions” to life strategy. This fact leads to several significant and (by our) society insufficiently reflected and certainly not respected consequences. In the course of postnatal development, calendar (chronological, absolute) age and developmental age, i.e., the stage of development and growth of the body, psyche and social maturation, diverge in different ways in different people. One of the above consequences is the difficulty of drawing a generally valid line between a “healthy” or “physiological” (correct, normal) development and incorrect, pathological development, which requires special intervention (in medicine, psychology, social relations). The second consequence is the relatively difficult categorization of the majority/healthy/normal population by society (state, school, health care, etc.) into some ontogenetically internally homogeneous groups. In fact, children of identical calendar age differ significantly in the values/levels of some developmental indicators (body height, teeth eruption, cognitive abilities, social competence), in a natural (healthy, physiological) way. The variability of these differences increases steadily in the postnatal phase and can amount to several years in the pubertal period, which is practically equivalent to the difference between a prepubertal child and a sexually and physically almost adult person of the same calendar age (Geithner et al. 1998). In the case of school-age children this is especially evident in the second stage of elementary school.

Usually, however, norms of human growth and development are established on the basis of chronological (calendar) age (Cole 2012a; 2012b). The values of the observed trait are averaged for individuals of the same calendar age and the average value is considered the “norm”—i.e., that value a person of a particular calendar age should “ideally” have. Therefore, there is a slightly different view of individuals below and above the norm. Individuals who are above the average are seen as cases of good development, while individuals who are below the average are seen as late, deficient individuals, regardless of the fact that they are equally far from that “ideal” on each side of the value distribution. Both of these are misconceptions. It is indeed an average, but one calculated from a variety of developmentally incommensurable states. Therefore, such an average cannot be the norm for the whole group or for some individuals within the group. This approach is used not only

in interpreting data from cross-sectional studies but unfortunately also in processing longitudinal data.

This method of assessment is useful in finding the boundary/decision between physiological and pathological development (human medical auxology). If we have an appropriate frame of reference—for example, a developing population in the form of a percentile plot—cases that fall outside the distribution of most (albeit developmentally diverse) individuals are indeed very likely to be developmentally pathological (Barstow – Rerucha 2015; Nwosu – Lee 2008). However, the population approach may not be ideal when looking at development and growth within a more or less healthy majority, where a broad population grid may not be detailed/precise/fine enough to capture smaller developmental imbalances. Here it is necessary to turn to examining the growth of the individual itself comparing individual growth curves, preferably through an analysis of growth velocities/rates (Lampl – Thompson 2007). The need for an individualized approach to developmental processes has been evident since early modern European studies of child development and growth (Bogin 1999). Explicitly then, at least since the beginning of scientifically conceived growth anthropology (Human Auxology), this topic was already opened, for example, by Franz Boas (1892; 1930), who pointed out the need to compare individuals of the same biological and not calendar age.

Tracing the relationships between postnatal development at the individual level and environmental factors was the theme of the first large longitudinal studies in the United States that sought to examine the “whole child” (Bogin 2021, p. 48). This included examining the many factors that influence child growth. Because the greatest differences in growth and final body size were observed between developing countries and developed Western economies (most striking in the comparison between immigrant parents from their home countries with their children born and developing in Western countries), the main factors examined were nutrition (quantity and quality of food), health care (hygiene, pathogens), and physical stress on children as they grew. Poor living conditions in the country of origin, or their improvement after arrival in a more developed country, were considered the main reason for children’s improved growth and larger body size in adulthood. In the last decade, the focus of studies has shifted both to the social and psychological determinants of growth and to the level of epigenetic processes in the prenatal period. Indeed, growth has been shown to be significantly predisposed by processes in the prenatal period (Stout et al. 2015; Zambrano et al. 2014), but after birth it is also significantly influenced by social and psychological factors, possibly including some adjustment of one’s growth based on the growth of peers—theories have been referred to as “strategic growth adjustments” and “competitive growth” (Bogin et al. 2018; Charmandari et al. 2012). This modern view of child growth is gradually being applied in compendia and auxological monographs and textbooks (Hermanussen 2013; Preedy 2012).

Thus, the complexity of modelling human growth is due to the fact that, on the one hand, growth processes (cell prolif-

eration, tissue growth) in the organism occur on the basis of relatively simply defined physiological processes and can be modelled by simple curves, but, on the other hand, their combination with the evolutionarily rooted folding of human life history and the adaptive nature of ontogeny itself, realized in a constantly changing natural and social environment, makes each growth trajectory a completely individual path. It is therefore impossible, in principle, to find a single curve that could model the growth of every human being with the same quality.

GROWTH MODELLING

The key step in growth modelling is to fit an appropriate individual model to the individual data so that it accurately and correctly describes individual growth, i.e., on the one hand it ignores noise as much as possible and, on the other hand, it does not level/suppress the diagnostically relevant signal. At the same time, it should be defined/described by the same set of parameters (e.g., coefficients of an equation) for each case (subject) so that all individuals can be compared with each other and the data can be further statistically analyzed, utilized to make predictions, to reveal the influence of external growth factors, etc.

The fundamental difference of human growth—compared to the growth of other animals—that poses a serious challenge for growth modelling is its multi-stage nature (Bogin 1999). Because of the changes that have occurred during unique human evolution, our growth is characterized by a significant lengthening of some developmental stages and the addition of some new stages. Therefore, the postnatal growth trajectory of humans cannot be modelled by a simple (e.g., logistic) curve. This is only possible for certain separate/limited periods of postnatal growth, for example puberty. In this situation, the shape of the growth curves is very similar across all individuals (Cole et al. 2010), though not exactly the same, and most of the variance in values at a given calendar age is caused by the shift of a particular developmental event/phase to a different calendar age and a different slope of the curves (growth rate). Of course, it is a difficult task to accurately determine these periods within the entire postnatal trajectory to be comparable. However, attempting to model the whole postnatal growth in one curve remains a complex problem, as both the individual periods and their timing and relationships to each other vary from person to person. Moreover, each trait (stature, weight, etc.) grows differently and must be modelled with a different curve.

In terms of how the model is created, the methods proposed and used so far to model individual human height growth trajectories can be divided into several groups/types.

Advanced mathematical models

One approach is polynomials and other types of mathematically described curves. Probably the best known and also most

commonly used model is *Preece and Baines model 1* (Preece – Baines 1978; Sayers et al. 2013), a nonlinear model with 5 parameters. However, a number of other more or less similar curves have been proposed (Goldstein 1986; McKeague et al. 2011). For overview of the different models, see, for example, Roche and Sun (2005). These curves are quite complex, and the structure of their coefficients allows to take into account the specificities of human growth (however, the most adapted are those for height growth). The problems of this approach include both the different quality of the model fit to the data in different individuals (especially in the most variable phases in the curve) and the unclear biological interpretation of the obtained coefficients.

Partitioning into separate growth periods or growth-phase components

Another approach is to divide the entire growth from birth to adulthood into phases and model each phase separately using a simpler curve (most commonly logistic, Gompertz, and exponential curves) and then combine them into a final model. Here we can mention the popular *Infancy-Childhood-Puberty* (ICP) model of Karlberg (1989a; 1989b), which divides human growth into three separate “components”, each driven by a separate model (equation). The difficulty with this approach is dealing with the transition between growth phases, i.e., which point on the curve belongs to the previous phase and which already belongs to the next phase. As in the previous type of methods, the problem lies in the biological interpretation of the obtained coefficients. The advantage, on the other hand, lies in the possibility of increasing the lengths of the individual growth phases differently for each individual independently. The *Dynamic Phenotype Model* by Ludvík Novák also belongs to this type of growth models (Čuta 2014; Novák et al. 2007; 2008). The advantage of his approach within this category of models is that it is based on physiological growth characteristics and the parameters (coefficient values) obtained after fitting can be unambiguously interpreted biologically. However, the problem still remains in connecting the models of the different phases.

Functional data analysis

Functional Data Analysis (FDA), generally represents a statistical discipline that focuses on the functional description and analysis of the change of a signal along a continuous variable, usually space or time. In the field of human growth modelling, the analyzed continuum is a growth curve (change in a measured variable along time) which is transformed from raw data into the form of a function which is further analyzed. FDA can be further combined with Principal Component Analysis (here Functional Principal Component Analysis, FPCA) to quantify variations in the analyzed functional data. The FDA method was developed in the second half of the 20th century and it is mathematically described in the books by Ramsay and Silverman (2002; 2005).

Fitting a sample mean curve

Another approach was proposed by Beath (2007) and elaborated by Tim Cole for the purpose of human growth curve analysis (Cole et al. 2010; 2014). The procedure consists of *SuperImposition by Translation And Rotation* (SITAR for short). In this method, the original individual trajectories are superimposed as tightly as possible, then a single model (average) curve is first obtained as an optional and optimized spline by nesting these superimposed trajectories, and finally the average curve is used as a growth model and transformed back into the individual growth data by using affine shape transformations (translation and uniform shear/rotation). The deviation from the average curve (with minus or plus signs) is then described by three coefficients of this transformation: x -axis position—timing relative to age, y -axis position—size, rotation—rate of change. This approach is suitable not only for studying the development of traits on a continuous scale, such as body height, but also categorical traits, such as the reaching of a certain developmental stage (e.g., Tanner scales) or a sub-skill, such as cognitive (Cole et al. 2014).

The average curve is based on the evaluated sample and will be different if we use a different sample. This can be an advantage as the curve best matches, on average, the sample which was just analyzed. In addition, the resulting curve can be compelled even if none of the original individual trajectories are that long, since all of the original trajectories are combined in the resulting curve. However, the properties of the recorded data in terms of the superposition procedure and the formation of the average curve are affected by their age range and the coverage of growth milestones, and the model curve may then not match the actual human growth in the case of an inappropriate sample (we must not forget that the measured data are raw and contain noise). Another problematic aspect of this approach may be related to growth rate variability—if the curve is the same for all individuals, this may also mean that the model fits well to an average growth rate of the individual but may not be as accurate (in terms of fit; residuals) with respect to variable growth rate changes along the individual trajectories.

ISSUES ON MODELLING OF GROWTH ON A SMALL NUMBER OF SPARSE LONGITUDINAL GROWTH DATA

The above approaches work reasonably well, and all are applicable and used, though not uniform, when we have dense longitudinal data of the entire postnatal growth, e.g., one measurement every six months for eighteen years from birth to adulthood. However, fitting a suitable and analyzable curve to some empirical longitudinal data is only the first part of the problem. The situation is much more difficult when we have only a few measurements, further apart and for each individual at a slightly different age or even growth stage. Theoretical considerations of the difficulties of standard models in such

a situation are briefly presented in a methodology paper for the publication of our new growth model method (Králik et al. 2021). In short, when the number of points is small, the best-fitting model may not be the most accurate, because the points are measured with error (noise), and a perfect fit in such a case may represent an unnatural curve along which no human growth actually occurs. In the extreme case, we have only two points and their ideal model is a straight line, but this is certainly not an appropriate model for actual growth. Thus, the information from the modeled values alone cannot be sufficient to interpolate missing curve segments between empirical points or to extrapolate beyond the observed age range.

As part of the project (see Funding) we sought a solution to this problem, including its application that could be readily used for practical purposes in pediatric auxology and sports anthropology. Previous experience of some team members (Čuta 2014) has shown that despite the relatively good performance of the above models in fitting growth trajectories, there is a great deal of individual variability in various details of individual growth trajectories, for example, in the number and intensity of growth spurts in childhood to pre-puberty. The most commonly used models do not usually take these into account (they are not part of the proposed model of the growth trajectory “waveform”), which is usually not detrimental, as it is not essential for estimating major growth milestones such as the age at peak velocity (APV), nor, for example, for examining major differences in growth between two populations. However, when only a few years of growth records are available and our goal is not to estimate population major milestones, but rather a detailed assessment of the physiological nature of an individual child’s growth and the possible detection of growth imbalances at the short-term level, standard models may no longer be sufficient to fit the model with a limited amount of data or to describe these data in detail.

The first idea of our solution was to model incomplete, limited data simply by empirically documented complete growth tracking, i.e., without any mathematical model. That is, we used already realized growth trajectories of people who were already adults—children documented in the Brno Growth Study—and from a large number of such empirical trajectories we selected one or a few trajectories that were most similar to the modeled data and used them as model trajectories. This approach has two pitfalls. First, both sides (reference trajectories and modelled data) are measured with an error, which can severely affect the comparison, and second, the reference measurements were taken at different ages, so they are not directly comparable. Furthermore, interpolation of values between empirically measured points is necessary to calculate the growth rate and acceleration, and thus the significant point on the growth curve.

It was therefore necessary to fit the empirical reference trajectories with a suitable curve. We chose a B-spline with sensitively adjusted parameters so that it responded to local growth waves within the 18-year trajectory and respected

the data well but was no longer affected by short-term fluctuations and noise. However, this was only possible because of the large number of regularly taken and spaced measurements of the reference sample (BGS), where the curvature of the whole curve could be easily distinguished from the bounce/fluctuation of a single point (a measurement error). These curves are then used to calculate height values corresponding to the age(s) at which the assessed case/subject was measured. From all these interpolated values of body height/stature of the reference sample, the least squares method is then used to select one or more trajectories that are closest to the measured values of the assessed case. (The reference data thus enter the process as values on the smoothed splines, while the data from the evaluated case enter the process as raw measurements.)

However, in testing the functionality of this procedure, it soon became apparent that even the number of empirical curves available in a large-scale growth study such as the BGS was insufficient. For many of the cases tested, no reference curve was found that ran directly through the measurements of the newly evaluated case. It was therefore necessary to develop a procedure to obtain new, artifactual reference trajectories that were nevertheless based on the empirical curves of the reference set (BGS), covered the entire growth range, but were still natural human growth trajectories. The solution to this problem was then a method based on the description of the growth trajectories using the FDA followed by Functional Principal Components Analysis (FPCA). This method was described in detail in the methodological article (Králik et al. 2021), including visualization of sub-steps and graphical representation of each principal component. Here we take the liberty of a brief summary of the method.

THE NEW METHOD FOR GROWTH MODELLING

The approach is based on a statistical generative model of the growth curve that is non-rigidly fitted to the measured points in such a constrained manner that the resulting curve is always biologically plausible and most probable with respect to the reference population, while being robust to a reasonable degree of measurement inaccuracy. The statistical model describes the curve by two sets of independent parameters, separating effects of phase and amplitude of growth. To obtain a model instance that describes a particular individual, the parameters were adjusted using a numerical optimizer so that the model curve intersects the measured points with certain tolerance.

Two such generative growth models, one for boys and one for girls, were constructed based on reference samples collected as part of the Brno Growth Study (BGS) (Bouchalová 1987), which contained typically 39 height measurements from birth to age 18 years for 334 individuals. Height data for each individual were interpolated using a B-spline curve, following a procedure described by Ramsay et al. (2009). These B-spline curves were then brought into correspondence using a regis-

tration that exploited the age of peak velocity (APV) detected on each curve. This was a key step that allowed the construction of a model describing the growth phase and amplitude separately by independent parameters, generating the time-warping functions describing the shift in the growth phase of each B-spline curve with respect to the population mean. Using the time-warping functions, the growth curves were aligned to the same phase. The final generative model for each group was obtained by applying Functional Principal Component Analyses (FPCA) to the sets of phase-aligned curves and the inverse time-warping functions, generating the amplitude and phase modes, respectively.

An instance of a model describing a particular curve corresponding to certain parameters was obtained by first generating an amplitude curve and then performing the time warping. Fitting the models to the sparse measurements of the newly analyzed case is accomplished by numerical optimization of the model parameters using a Levenberg–Marquardt solver (Kelley 1999, p. 56–58), minimizing two sets of residuals, describing the differences between the measured and modeled heights and the modeled curve probability respectively. To ensure the biological plausibility of the fitted model, the scores of both the phase and amplitude parameters were constrained to ± 3 standard deviations during optimization. For further details of the estimation method see Králik et al. (2021).

From the resulting continuous curve of the individual, growth milestones are obtained using the *getPeak* and *getTakeoff* functions from the *sitar* package (Cole 2020) available for the R environment (R Core Team 2020). For each modeled individual, these milestones from model curves were determined:

APV – age at peak velocity in pubertal body height growth spurt (in years)

VPV – peak velocity in the pubertal body height growth spurt (in cm/years)

HPV – body height at the point of peak velocity in puberty (in cm)

ATO – age at take-off before pubertal body height spurt (in years)

VTO – velocity at point of take-off before body height spurt (in cm/years)

HTO – body height at the point of take-off before spurt in puberty (in cm)

PRESENT STATE OF THE GROWTH MODELLING APPLICATION

Technical solution

The goal of the architectural design was to develop software that was both easy to use and practical for paediatricians, and suitable for the research community to perform batch analysis of large and diverse data sets. Therefore, the front-end was

designed as a web application while the computational core consists of scripts for the *R* environment. The advantage of the web interface is that it is ready to use and does not require installation of the software on the user's computer, making the application available without much effort.

The *R* environment (R Core Team 2020) is widely used in the research community, so the core uses current state-of-the-art packages, and the newly developed methods are also returned to the community in the form of *R* packages.

The solution runs on an *Apache2* server hosted on an Ubuntu Linux virtual machine located on the cloud infrastructure (IaaS) of Masaryk University <<https://cloud.muni.cz/>>, in the OpenStack environment. The *R* environment is loaded inside the *Apache2* server via an *R-Apache* module <<https://github.com/jeffreyhorner/rapache>>, which enables online execution and interpretation of *R* scripts. The web interface is based on the *WordPress CMS*, and the connection with the *R* core scripts is ensured by a custom developed plugin. The transfer of input data and results between the user interface and the computational part is done interactively via *Ajax*, the data itself is transferred in *JSON* format, including binary data (graphs stored as *png* images) encoded in *Base64* format.

Using the *R* environment as an *Apache2* module is advantageous for performance reasons, as the environment runs as part of the server and is therefore not restarted for each user request. In addition, the module also supports preloading data and executing startup scripts, so the environment only needs to be restarted once at the *Apache2* startup when the *R* packages and data containing trained models are loaded.

The core scripts use the packages *fda* (Ramsay et al. 2020), *nlme* (Pinheiro et al. 2020), *sitar* (Cole 2020), *gamlss* (Rigby – Stasinopoulos 2005), and *pracma* (Borchers 2021) for the computations, *ggplot2* (Wickham 2016), *Cairo* (Urbanek – Horner 2020) and *png* (Urbanek 2013) for the visualization of the results and *jsonlite* (Ooms 2014) for the communication with the front-end.

The use of *WordPress CMS* allows easy extensibility of the user interface with new features and is convenient for support.

Description of the input

A demonstration application that computes our proposed growth model and interleaves these curves with newly entered data is available at: <<http://growth.sci.muni.cz>>.

Currently, the application allows computing of the FPCA model and its derived estimates for one specified longitudinal series of body height measurements of an individual – only one case (child) can now be entered in a single calculation run. Therefore, batch processing of multiple individuals simultaneously is not possible, which is envisaged for further program development.

When the user opens the application page (fig. 1), a drop-down toolbar appears on the left side for entering information about the case being evaluated. When the “Personal info” menu is clicked, the user will need to enter the biological sex of the case being assessed and the date of birth. Based on the

sex entered, the program will now work with data for that sex only. From the date of birth and date of measurement entered below, the program calculates the calendar age at the time of measurement.

In the next menu labeled “Stature measurements”, the user enters the number of unmeasured values to be counted (“Measurements Count”) and the program opens the corresponding number of empty fields for entering individual measurements. As for entering time/age data is concerned, the user can either enter the date of the measurement or directly the calendar age. If the date of birth of the person being assessed was entered previously and now the date of measurement is entered, the program automatically calculates the calendar age. If the calendar age is entered directly, the program ignores the date of birth and the date of measurement in the calculation.

In the “Plot setting” menu the user can select how many of the most similar growth curves from the reference population he or she wishes to display (1–3 are recommended for clarity).

In the next menu, “Model and Sample” the user selects the method and reference sample for the longitudinal model (Longitudinal model and sample)—we recommend keeping the described FPCA method based on the sample of children from the Brno Growth Study, which is set as the default option. Next, the user can select the data source for the traditional percentile plot.

Another menu labeled “Midparent” allows the user to enter the height of the father and mother, which can be used in calculating of the target height of the child under study and other operations.

There is also an “Actions” menu that allows the user to save the entered data to a backup text file or to reload such a data file at another time.

And finally, the user arrives to the “PLOT” button. When it is clicked, the program calculates the growth model based on the entered data and the set options, displays the results in graphs and prepares tabular and textual outputs of the evaluation of the entered growth curve. Since the optimization of the growth curve is computationally intensive, the entire calculation can take units up to tens of seconds.

Description of the graphical outputs

After the calculation, several graphical outputs are displayed. The first is the main distance curve graph (fig. 2), which plots body height against the calendar age. It shows the estimated growth curve, the empirical curve from the reference population that is the closest to it, and the optimized FPCA model used to calculate the ATO and APV values for age and body height. The second graph projects the assessed case onto a percentile (cross-sectional) graph of the selected reference population (fig. 3). The third graph provides a comparison of the measured values with the distribution of values of the reference sample of the longitudinal model at the age at which the assessed case was measured (fig. 4).

When there is more than one measurement for an evaluated case, the fourth graph shows a comparison of the growth rate

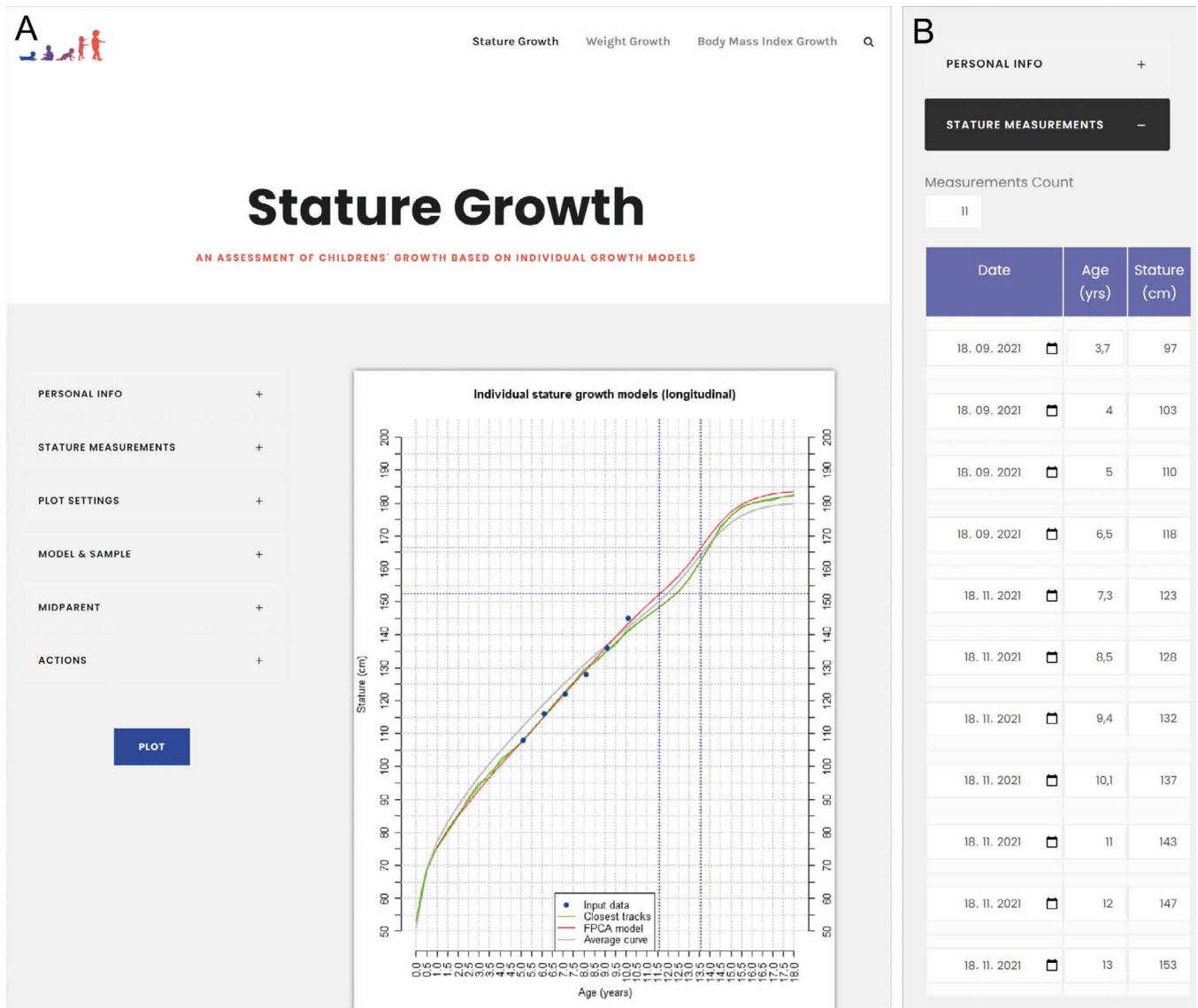


Figure 1 GROWTH main page with left menu for input data (A), drop-down menu for inputting measurement age and height (B).

of the evaluated case with the distribution of values (box-and-whisker plot) of the reference set at the same age interval (fig. 5). This plot allows a much more sensitive assessment of the growth trends of a particular case in each growth interval. Another sensitive indicator of the growth position and trend of the case being evaluated may be a comparison of body height at day age with the value retrospectively predicted for that calendar age to the value of the midparental stature (if the values of the adult height of the child's father and mother are available), which is the next graph provides.

Description of the text and numerical outputs

All values of the indicators presented in the graphs are also available as numerical values in tables and text data. These are the input data themselves, as well as the values of body height with the indicated ages of the nearest empirical growth curve, the values extrapolated from the FPCA model, the residuals

of the FPCA model, the values of growth rates at the specified intervals, and the statistical parameters of the reference population growth rate values, as well as the relative position of the case being evaluated in relation to the extrapolation of the midparental value for the indicated age. In addition, the results include the calculation of target height using several methods and the estimation of target height by the FPCA model. User data and results are not permanently stored anywhere, but the user can store them in numerical and/or graphical form on his/her computer.

Commentary to the proposed method and its application

The proposed application GROWTH is based on the idea of modelling human growth using previously documented growth trajectories of real people. In the first phase, we tested the use of raw growth trails. In the next phase of developing

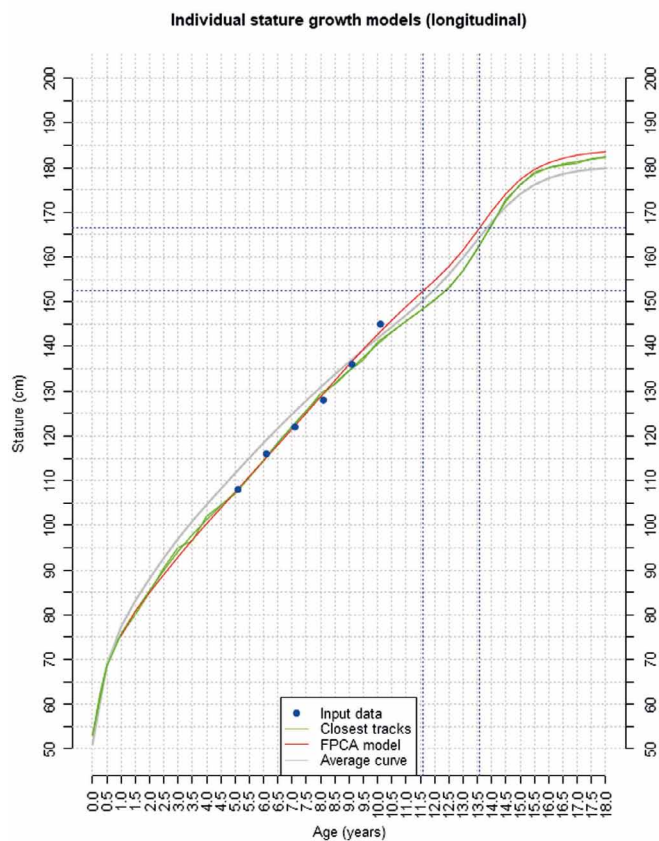


Figure 2 Plot of the distance growth curve for body height in GROWTH (males); blue points – input data (evaluated case of a hypothetical boy), green curve – most similar empirical trajectory from Brno Growth Study, red – model curve based on FPCA method, grey – population mean curve of Brno Growth Study, blue dashed horizontals and verticals – stature height and age at ATO (Age at Take-off) and APV (Age at Peak Velocity) from FPCA model curve.

this method, we fitted the curves using the SITAR method (Cole 2020; Cole et al. 2010; 2014), which allows us to work easily with a smaller number of unevenly distributed measurements. In any case, SITAR is a very robust method and has been proven in practical applications (Cole et al. 2014; Malina – Coelho-e-Silva, et al. 2021; Malina – Kozieł, et al. 2021). However, we ultimately moved to pure B-splines and the FDA method because the individual-specific B-spline provides a more accurate fit of the individual growth trajectory than the average spline used by the SITAR method. This does not imply a large difference in the estimates of major growth milestones (Králik et al. 2021), but rather the advantage of being able to account for the details of individual child growth on a shorter time scale, which was one of the goals in developing this auxological application.

The resulting method, used in the presented web application, is based on the FDA procedure described by Ramsay et al. (2009; 2002; 2005). It is based on splitting the growth into two independent models that describe its phase and amplitude separately. The two parts are then modelled separately using Functional Principal Component Analysis (FPCA), resulting

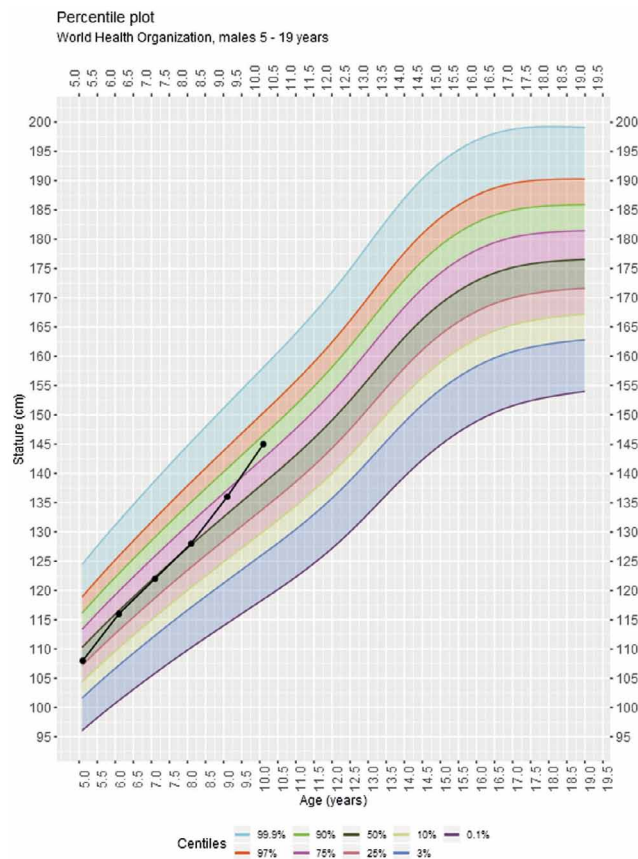


Figure 3 Projection of the evaluated case (black points and line) from Figure 2 onto a traditional percentile plot based on WHO data.

in the possibility of generating new curves (PCA is generally a generative model). Although the use of the SITAR model is now more widespread, mainly because of its ease of use, we have focused more on the FDA model introduced by Ramsay et al. However, while the SITAR model includes a way to apply it to new data, Ramsay et al. (2009) only published the model itself, without a procedure to register it to new measurements. Hermanussen and Meigen (2007) then developed a procedure that formulated the registration based on the Maximum Likelihood principle and solved it using numerical optimization, in particular the Hooke-Jeeves method (Gottfried – Weisman 1973). Besides registration, they also had the idea of detecting pathological cases by checking the parameters of the “registered” PCA model using the False Discovery Rate method to reduce the number of falsely detected pathological cases. However, they worked with discrete methods and data (growth curve registration with the Shifting algorithm, classical PCA methods), and therefore they did not exploit the full potential of the Functional Data Analysis methods described by Ramsay et al. (2009; 2002; 2005). Meigen and Hermanussen (2003) developed a web application for growth modelling, but did not separate growth phase and amplitude and described growth curves with only one common model. The above approach was also tested in our project but was deemed to be less suitable due to its shortcomings. We incorporated

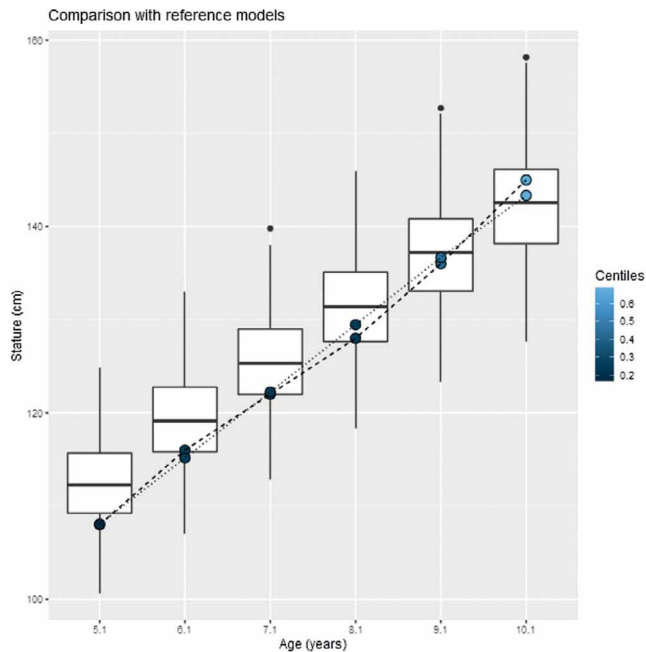


Figure 4 Measured values of the case visualized in Figures 2 and 3 in comparison with the distribution of values of the reference sample (BGS) at the ages at which the assessed case was measured; points connected with a dashed line – empirical values, points connected with a dotted line – values extrapolated from FPCA model of the case.

the approach with two independent FPCA models (one for phase and one for amplitude) into our method (Králik et al. 2021) and it is part of the presented GROWTH web application.

The proposed software currently handles several main tasks, which may vary depending on the user's specialization. When the task is to fit the growth data with a curve within the measured and specified age range, i.e., to complete the curve between empirical measurements (*to interpolate*), the FPCA model used in our program performs as well or better than traditional/older models. When the aim of the fitting and interpolation is also to obtain estimates of growth milestones (APV, VPV, etc., computed based on inflection points of the model curve), the method also performs reliably and comparably to the SITAR method as long as the computed milestone is within the range of empirically measured and entered values. This can be seen in the test section of our model publication (Králik et al. 2021).

When the purpose of the model is to predict long-term growth over a range of many years beyond the specified range of empirical values (*i.e., to extrapolate*), the principle of the model calculation implies that the estimate will certainly be more approximate and more influenced by the tendency to the average curve. Permutation test approach of our model on samples of five measurements (with a span of 1 year between successive measurements) systematically selected from empirical data (of the same testing population sample—BGS by means of leave-one-out method) at varying distances from

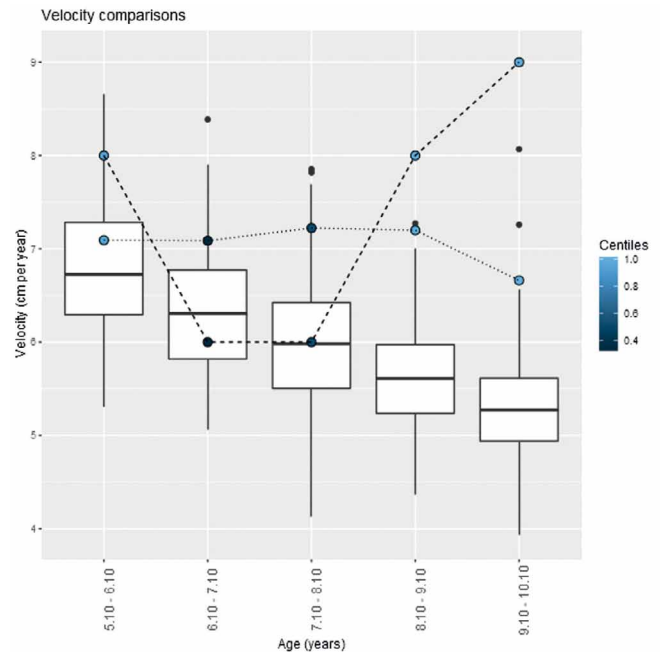


Figure 5 Comparison of the growth velocities of the evaluated case (from figures 2–4) with the reference velocities (Brno Growth Study) at the same age intervals; points connected with a dashed line – empirical values, points connected with a dotted line – values extrapolated from FPCA model of the case.

the APV showed that the FPCA model provides reasonable estimates even when extrapolated outside the empirically specified age range (*i.e., predicting future growth or estimating past growth*), but the resulting estimates have overall lower variance than the reference (true) APV values secured on the curves (Králik et al. 2021). Reducing the variance of the estimates during extrapolation beyond limited empirical data is a general feature of growth models based on any optimization and is certainly better than the opposite extreme (increasing the variance of the estimates and large variability of the curves). Regression of the model curve toward the mean curve and reduction in the variance of the estimates may not matter in some population comparisons of two or more population samples matching the nature of the input data, but they will matter a great deal when individual-level extrapolation is used, and interpretation of the individual estimates is attempted somehow. It is therefore important to note that the further the age is from the empirically entered data for which stature is estimated, the less precise the estimate becomes. At the same time, we have also found that the APV estimation is more reliable for empirical data that are outside the APV when we have empirical measurements *before* the APV than *after* the APV (Králik et al. 2021). Thus, for some reason, the growth onset information is more significant for detecting the APV than growth completion after the APV (*i.e., when predicting back in age*). Moreover, it is apparent that future growth depends on future external growth factors (which are not known or even do not exist yet) and an exact prediction

is not possible in principle. Therefore, the use of this application for interpolation (filling the curve within the range of measured values) and extrapolation (estimation/prediction outside the range of empirical values) should be considered as two completely different disciplines and their accuracy should be assessed completely differently.

One of the potential practical applications of our program is to test whether an evaluated child with incomplete growth (e.g., at age 4, 8, or 13) is growing “normally” or should be considered abnormal. It is possible to visually compare his/her past growth trajectory with the most similar empirical growth curve to assess the normality (typicality) of the FPCA growth model (and its individual components of the two FPCAs). However, it is probably most sensitive to compare the case with the extrapolation of the midparental height (whether it is smaller or larger than the midparent at a given age) and, in particular to assess the growth rates between each measurement relative to the growth rates of the reference sample.

Future plans for development

There are a number of aspects that are not yet (or not fully) addressed that we would like to add in the future. Due to population differences, it is desirable to complete the FPCA model for other longitudinal data as well. It would also be useful to include assessments of other body measures, e.g., sitting height, head circumference or some body proportions. To make the model curves more usable, we also plan to create a package called *growthfd* for the R-software that would allow all functions to be computed on one's own computer directly in the R-software (which would certainly speed up the entire computation), as well as batch processing of multiple individuals simultaneously. Currently, a pre-release version of the package is available on github <<https://github.com/ondrej-klima/growthfd>>. The most important new feature allows the users to batch process individual data. New features will be added in the near future.

CONCLUSIONS

The application GROWTH is an advanced yet easy-to-use web-based tool designed to provide a complex analysis of an individual child's growth, including the assessment of timing of major milestones. At the same time, the application can also be used for a simple visualization of a child's growth trajectory. The application is primarily intended for paediatricians, who can use it to monitor growth, predict attained height and diagnose pathological growth patterns. The development of the application involved several stages. The current version of the application, which provides the best fit of individually modeled curves uses an FPCA-based model with numerical optimization and uses machine learning to analyze complete and scarce data. These advanced techniques are computed in the backend of the application and provide the user with easy-to-use numerical and graphical output.

ACKNOWLEDGEMENTS

We thank all students who participated in sorting, cleaning, and preparation of computerized archival data. The authors would like to express their gratitude to Tomáš Mořkovský, MSc. for administrative support within the related grant project.

FUNDING

This study was supported by Technology Agency of the Czech Republic (Technologická agentura České republiky), grant number TL01000394.

REFERENCES

- Barstow, C. – Rerucha, C. (2015): Evaluation of Short and Tall Stature in Children. *American Family Physician*, 92(1), p. 43–50. <https://www.aafp.org/afp/2015/0701/p43.html>
- Beath, K. J. (2007): Infant growth modelling using a shape invariant model with random effects. *Statistics in Medicine*, 26(12), p. 2547–2564. <https://doi.org/10.1002/sim.2718>
- Boas, F. (1892): The Growth of Children. *Science*, 19(483), p. 256. <https://doi.org/10.1126/science.ns-19.483.256>
- Boas, F. (1930): Observations on the Growth of Children. *Science*, 72(1854), p. 44. <https://doi.org/10.1126/science.72.1854.44>
- Bogin, B. (1999): *Patterns of Human Growth*. 2nd Edition. Cambridge: Cambridge University Press, 472 p.
- Bogin, B. (2021): *Patterns of Human Growth*. 3rd Edition. Cambridge: Cambridge University Press, 590 p.
- Bogin, B. – Hermanussen, M. – Scheffler, C. (2018): As tall as my peers? Similarity in body height between migrants and hosts. *Anthropologischer Anzeiger*, 74(5), p. 365–376.
- Borchers, H. W. (2021): *pracma: Practical Numerical Math Functions*. Online at <https://CRAN.R-project.org/package=pracma>
- Bouchalová, M. (1987): *Vývoj během dětství a jeho ovlivnění. Brněnská růstová studie*. Praha: Avicenum, Zdravotnické nakladatelství, 384 p.
- Charmandari, E. – Achermann, J. C. – Carel, J.-C. – Soder, O. – Chrousos, G. P. (2012): Stress Response and Child Health. *Science Signaling*, 5(248), p. mr1. <https://doi.org/10.1126/scisignal.2003595>
- Cole, T. (2020): *sitar: Super Imposition by Translation and Rotation Growth Curve Analysis*. Online at <https://CRAN.R-project.org/package=sitar>
- Cole, T. J. (2012a): Growth references and standards. In: Cameron, N. – Bogin, B., eds., *Human growth and development*, Amsterdam: Elsevier, p. 537–566. Online at <http://dx.doi.org/10.1016/B978-0-12-383882-7.00021-0>
- Cole, T. J. (2012b): The development of growth references and growth charts. *Annals of Human Biology*, 39(5), p. 382–394. <https://doi.org/10.3109/03014460.2012.694475>
- Cole, T. J. – Donaldson, M. D. C. – Ben-Shlomo, Y. (2010): SITAR—a useful instrument for growth curve analysis. *International Journal of Epidemiology*, 39(6), p. 1558–1566. <https://doi.org/10.1093/ije/dyq115>
- Cole, T. J. – Pan, H. – Butler, G. E. (2014): A mixed effects model to estimate timing and intensity of pubertal growth from height and secondary sexual characteristics. *Annals of Human Biology*, 41(1), p. 76–83. <https://doi.org/10.3109/03014460.2013.856472>
- Čuta, M. (2014): *Modelování lidského růstu. Dynamický fenotyp*. Brno: Akademické nakladatelství CERM, 107 p.
- Geithner, C. A. – Woynarowska, B. – Malina, R. M. (1998): The adolescent spurt and sexual maturation in girls active and not active in sport. *Annals of Human Biology*, 25(5), p. 415–423. <https://doi.org/10.1080/03014469800006662>
- Gilbert, S. F. (2012): Ecological developmental biology: environmental signals for normal animal development. *Evolution & Development*, 14(1), p. 20–28. <https://doi.org/10.1111/j.1525-142X.2011.00519.x>

- Gilbert, S. F. – Epel, D. (2015): *Ecological Developmental Biology. The Environmental Regulation of Development, Health and Evolution*. New York: Oxford University Press Inc, 576 p.
- Goldstein, H. (1986): Efficient statistical modelling of longitudinal data. *Annals of Human Biology*, 13(2), p. 129–141. <https://doi.org/10.1080/03014468600008271>
- Gottfried, B. S. – Weisman, J. (1973): *Introduction to Optimization Theory*. Englewood Cliffs, NJ: Prentice Hall, p. 571.
- Hermanussen, M., ed. (2013): *Auxology. Studying Human Growth and Development*. Stuttgart, Germany: Schweizerbart Science Publishers, 324 p.
- Hermanussen, M. – Meigen, C. (2007): Phase variation in child and adolescent growth. *The International Journal of Biostatistics*, 3(1), Article 9, 14 p. <https://doi.org/10.2202/1557-4679.1045>
- Karlberg, J. (1989): A Biologically-Oriented Mathematical Model (ICP) for Human Growth. *Acta Paediatrica Suppl.*, 78(s350), p. 70–94. <https://doi.org/10.1111/j.1651-2227.1989.tb11199.x>
- Karlberg, J. (1989): On the Construction of the Infancy-Childhood-Puberty Growth Standard. *Acta Paediatrica* 78(s356), p. 26–37. <https://doi.org/10.1111/j.1651-2227.1989.tb11199.x>
- Kelley, C. T. (1999): *Iterative Methods for Optimization*. Frontiers in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 180 p.
- Králík, M. – Klíma, O. – Čuta, M. – Malina, R. M. – Kozieł, S. M. – Polcerová, L. – Škultétyová, A. – Španěl, M. – Kukla, L. – Zemčík, P. (2021): Estimating Growth in Height from Limited Longitudinal Growth Data Using Full-Curves Training Dataset: A Comparison of Two Procedures of Curve Optimization—Functional Principal Component Analysis and SITAR. *Children*, 8(10), p. 934. <https://doi.org/10.3390/children8100934>
- Lamp, M. – Thompson, A. L. (2007): Growth chart curves do not describe individual growth biology. *American Journal of Human Biology*, 19(5), p. 643–653. <https://doi.org/10.1002/ajhb.20707>
- Malina, R. M. – Coelho-e-Silva, M. J. – Martinho, D. V. – Sousa-e-Siva, P. – Figueiredo, A. J. – Cumming, S. P. – Králík, M. – Kozieł, S. M. (2021): Observed and predicted ages at peak height velocity in soccer players. *PLoS ONE*, 16(7), p. e0254659. <https://doi.org/10.1371/journal.pone.0254659>
- Malina, R. M. – Kozieł, S. M. – Králík, M. – Chrzanowska, M. – Suder, A. (2021): Prediction of maturity offset and age at peak height velocity in a longitudinal series of boys and girls. *American Journal of Human Biology*, 33(6), p. e23551. <https://doi.org/10.1002/ajhb.23551>
- McKeague, I. W. – López-Pintado, S. – Hallin, M. – Šiman, M. (2011): Analyzing growth trajectories. *Journal of Developmental Origins of Health and Disease*, 2(6), p. 322–329. <https://doi.org/10.1017/S2040174411000572>
- Meigen, C. – Hermanussen, M. (2003): Automatic analysis of longitudinal growth data on the Website willi-will-wachsen.de. *Homo*, 54(2), p. 157–161. <https://doi.org/10.1078/0018-442x-00067>
- Novák, L. – Kukla, L. – Čuta, M. (2008): Child and Adolescent Longitudinal Growth Data Evaluation Using Logistic Curve Fitting with Use of the Dynamic Phenotype Method. *Scripta Medica*, 81(1), p. 31–46.
- Novák, L. – Kukla, L. – Zeman, L. (2007): Characteristic Differences between the Growth of Man and the other Animals. *Prague Medical Report*, 108(2), p. 155–166.
- Nwosu, B. U. – Lee, M. M. (2008): Evaluation of short and tall stature in children. *American Family Physician*, 78(5), p. 597–604. <https://www.aafp.org/afp/2008/0901/p597.html>
- Ooms, J. (2014): The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO]. Online at <https://arxiv.org/abs/1403.2805>
- Pinheiro, J. – Bates, D. – DebRoy, S. – Sarkar, D. – R Core Team (2020): *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-148*. Online at <https://CRAN.R-project.org/package=nlme>
- Preece, M. A. – Baines, M. J. (1978): A new family of mathematical models describing the human growth curve. *Annals of Human Biology*, 5(1), p. 1–24. <https://doi.org/10.1080/03014467800002601>
- Preedy, V. R., ed. (2012): *Handbook of Growth and Growth Monitoring in Health and Disease*. New York: Springer-Verlag.
- R Core Team (2020): *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Online at <https://www.R-project.org/>
- Ramsay, J. O. – Graves, S. – Hooker, G. (2020): *fda: Functional Data Analysis*. Online at <https://CRAN.R-project.org/package=fda>
- Ramsay, J. O. – Hooker, G. – Graves, S. (2009): *Functional Data Analysis with R and MATLAB*. Dordrecht – Heidelberg – London – New York: Springer, 207 p.
- Ramsay, J. O. – Silverman, B. W. (2002): *Applied Functional Data Analysis: Methods and Case Studies*. 1st Edition. New York – Berlin – Heidelberg: Springer-Verlag, 190 p.
- Ramsay, J. O. – Silverman, B. W. (2005): *Functional Data Analysis*. 2nd Edition. New York: Springer Science+Business Media, Inc., 426 p.
- Rigby, R. A. – Stasinopoulos, D. M. (2005): Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, 54, p. 507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Roche, A. F. – Sun, S. S. (2005): *Human Growth: Assessment and Interpretation*. New York: Cambridge University Press, 324 p.
- Sayers, A. – Baines, M. – Tilling, K. (2013): A new family of mathematical models describing the human growth curve—Erratum: Direct calculation of peak height velocity, age at take-off and associated quantities. *Annals of Human Biology*, 40(3), p. 298–299. <https://doi.org/10.3109/03014460.2013.772655>
- Stout, S. A. – Espel, E. V. – Sandman, C. A. – Glynn, L. M. – Davis, E. P. (2015): Fetal programming of children's obesity risk. *Psychoneuroendocrinology*, 53, p. 29–39. <https://doi.org/10.1016/j.psyneuen.2014.12.009>
- Urbaneck, S. (2013): png: Read and write PNG images. Online at <https://CRAN.R-project.org/package=png>
- Urbaneck, S. – Horner, J. (2020): *Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output*. Online at <https://CRAN.R-project.org/package=Cairo>
- Zambrano, E. – Guzmán, C. – Rodríguez-González, G. L. – Durand-Carbajal, M. – Nathanielsz, P. W. (2014): Fetal programming of sexual development and reproductive function. *Molecular and Cellular Endocrinology*, 382(1), p. 538–549. <https://doi.org/10.1016/j.mce.2013.09.008>

AUTHORS

Klíma, Ondřej (1989, Brno), is a computer vision researcher. His research is focused on methods for morphometric analysis of medical and anthropological data using statistical shape models.

Čuta, Martin (1979, Brno), is a biological anthropologist. His research focuses on human growth and development and the study of stressors that influence them. His research also focuses on human body composition using advanced imaging methods and their connection with body surface features applied in ergonomics.

Polcerová, Lenka (1992, Brno), is a Ph.D. candidate in biological anthropology. Her research focuses on the diversity of quantitative dermatoglyphic traits with respect to interpopulation differences. At the same time, she is actively involved in the implementation of modern computational methods in the field of biological anthropology.

Zemčík, Pavel (1965, Brno), is a computer graphics researcher. His research is focused on algorithm design for a real-time video processing on embedded devices for traffic surveillance.

Škultétyová, Anna (1992, Hodonín), is a Ph.D. student and researcher in the field of anthropology. Her research and scientific interest focuses on biological anthropology especially growth, variability and sexual dimorphism of human hands and hand bones.

Černý, Dominik (1989, Čáslav), is a biological anthropologist. The main object of his studies is the use of static and dynamic features of the human body in forensic anthropology, especially in virtual environment.

Králík, Miroslav (1973, Brno), is a biological anthropologist. His research focuses on human sexual dimorphism in physical features and behavior, human ontogenetic adaptations and developmental stress factors.

