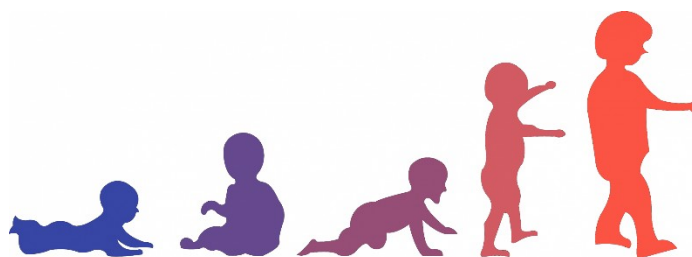


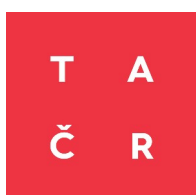
GROWTH



**COMPUTER APPLICATION FOR TRACKING CHILDREN'S GROWTH BASED ON
INDIVIDUAL GROWTH PATTERNS**

USER MANUAL

BRNO 2021



THE SOFTWARE WAS DEVELOPED WITH STATE BUDGET FUNDING PROVIDED BY THE TECHNOLOGY AGENCY OF THE CZECH REPUBLIC WITHIN THE ETA PROGRAMME.



IT4Innovations Centre of Excellence, Brno University of Technology, Brno

MUNI
SCI

Department of Anthropology, Faculty of Science, Masaryk University, Brno

Program GROWTH

A computer program designed to model and predict child growth, especially body height, in the form of an on-line application. A demonstration web application *GROWTH* for case analysis is available at: <<http://growth.sci.muni.cz>>.

This manual is intended for users of the web application..

The growthfd package, executable in R, containing the GROWTH computational routines and allowing batch processing of multiple cases simultaneously, is available at:

<<https://ondrej-klima.github.io/growthfd/index.html>>

Members of the team of the TAČR project number TL01000394 - *Počítačová podpora pro analýzu a predikci růstu a vývoje dítěte* (Computer support for analysis and prediction of child growth and development) participated in the development of the program:

Prof. Dr. Ing. Pavel Zemčík (VUT), Ing. Ondřej Klíma (VUT), Ing. Michal Španěl, Ph.D. (VUT), Assoc. Prof RNDr. Miroslav Králík, Ph.D. (MU), Mgr. Martin Čuta, Ph.D. (MU), Mgr. Lenka Polcerová (MU), Mgr. Anna Škultétyová (MU) a Mgr. Dominik Černý (MU).

The expert application guarantors of the TAČR project contributed a great deal of help in the development and testing of the program:

Assoc. Prof. MUDr. Lubomír Kukla, CSc. (General Practitioner for Children and Adolescents)

Prof. MUDr. Zdeněk Doležel, CSc. (Masaryk University, Faculty of Medicine, Department of Pediatrics)

The following contributed significantly to the development of the program:

Prof. Sławomir Koziel, PhD. (Polish Academy of Sciences, Hirszfeld Institute of Immunology and Experimental Therapy - provided test reference data from the Wrocław Growth Study and Wrocław Longitudinal Study of Twins)

Prof. Robert M. Malina PhD. et PhD. (Department of Kinesiology and Health Education, University of Texas at Austin - made significant contributions through expert consultation and collaboration in the production of publications.

Mgr. et Mgr. Daniela Kuruczová (Masaryk University, Faculty of Sports Studies, Incubator of Kinanthropology Research - programming the traditional growth models of Bock et al. (1973) and Preece and Baines (1978) into the R program script)

Assoc. Prof. RNDr. Petr Sedlak, Ph.D. (Charles University, Faculty of Science, Department of Anthropology and Human Genetics, provided data of selected pathological cases for testing the program)

Mgr. Tomáš Mořkovský (Masaryk University, Faculty of Science, Department of Anthropology, contributed significantly to the organizational and economic aspects of the TAČR project)

Organizing committee of the original Brno part of the ELSPEC project, with kind representation by its first national coordinator (Assoc. Prof. MUDr. Lubomír Kukla, CSc), contributed with provisional access to selected data from the database to test the functionality of the algorithms being developed.

How to cite

If you use this program, please kindly cite the following publication:

Králík, M., Klíma, O., Čuta, M., Malina, R. M., Koziel, 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). <https://doi.org/10.3390/children8100934>

Content

Initial motivation.....	5
Program resources.....	5
Methodological inspiration and applied principle.....	5
Data reference samples.....	6
Technical solution.....	6
Main features of the program - theoretical description.....	7
Fitting data with a growth curve.....	7
(1) The closest empirical trajectory.....	7
(2) FPCA model.....	7
Description of the web application.....	8
Input data and calculation settings.....	8
Personal Info.....	8
Examination Performator.....	8
Stature measurements – Measurement inputing.....	8
Closest trajectory number setting.....	9
Model and sample.....	9
Midparent.....	9
Actions.....	9
PROCESS.....	9
EXPORT PDF.....	9
Results and Outputs.....	10
DISTANCE PLOTS.....	10
REFERENCE PLOTS.....	12
PARENT REFERENCE PLOTS.....	14
EQUIPHASE PLOTS - Phase Synchronized Plots.....	15
TARGET HEIGHT.....	16
GROWTH MILESTONES.....	17
TEXT EVALUATION.....	17
Controlling other parts of the application.....	18
References.....	18

Initial motivation

The motivation for creating this program was that we could not find a suitable program that would allow us to fit a growth model to growth data with a small number of sparse measurements. A number of models for studying human growth, particularly stature, have been developed that work well when used to model/fit a large number of measurements recorded at short and regular time intervals. However, their potential for use with sparse data measured at long intervals in an irregular manner is limited, especially when we have only a few measurements with large intervals that also vary for each individual. This makes it difficult to model individual growth in practical fields that often work with such limited data, such as paediatrics, growth anthropology, and sports anthropology.

Theoretical considerations of the difficulties of previously used models in such a situation are briefly presented in the methodological article for the publication of our new growth modelling method method [20]. A problematic circumstance in modelling growth based on a small number of data points is that the best-fitting model (with minimal residuals) may not be the most accurate because the points are always measured with error (noise) and the progression of the curve exactly through the specified measurements in such a case may represent an unnatural curve along which no person's 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 a suitable model for real growth. While we can use this straight line to model, accurately and efficiently, mathematically the change from one particular age to another particular age, there is no doubt about that, however, it does not tell us very well what happens between these points - interpolation into the region between the points or extrapolation outside the empirical range of the specified points is likely to be wrong. So in such cases, the information from the input values alone cannot be sufficient to interpolate missing portions of the curve between the empirical points or to extrapolate outside the observed age range.

Program resources

Methodological inspiration and applied principle

Previous experience [11] has shown that, despite the generally consistent sequence of growth phases in all humans, there is a great deal of individual variation in various details of growth trajectories, for example, in the number and intensity of small growth spurts in childhood before puberty. The most commonly used models usually do not take this into account, which is usually not a bad thing. The reason is that it is not essential for estimating major growth milestones such as the age of maximum growth velocity at puberty (APV), nor, for example, for examining major/average differences in growth between two populations. However, when we have only a few measurements of a child's height over several years and our goal is to assess in detail the physiological nature of the individual child's growth and possibly detect his/ her growth imbalances in the short term, the models used to date may no longer be adequate.

The GROWTH program represents the first result of our efforts to develop a growth modelling method that will allow us to work also with this type of data and, in particular, to facilitate growth assessment for practitioners.

The main idea was to use computer methods to select, from a large number of real, empirically recorded growth trajectories (reference curves), the most appropriate one for the case to be analyzed and then to use it as a growth model for the individual to be evaluated. This is therefore a methodological application of empiricism – we are not mathematically calculating or creating a curve (we do not construct it according to some predefined equation), but simply selecting the most appropriate one from the curves that Mother Nature has already created. However, this approach has two pitfalls. First, both sides (reference trajectories and the evaluated case) are measured with error, which can seriously affect the comparison, and second, the reference measurements were made at different ages (and usually different from the evaluated case), so they are not directly comparable. In addition, interpolation of values between empirically measured points on the reference trajectories is required to calculate growth velocities and accelerations, and thus significant growth milestones on the growth curve. Therefore, it was necessary to first fit the empirical reference trajectories with a suitable growth curve and then work with these curves as representations of the reference set. In the role of the model fitted to the full reference data, we used six different methods (see below) whose fit to the reference trajectories (growth measurements) was appropriate for the further procedure.

During testing, however, it soon became apparent that even the large number of empirical trajectories in the reference set was insufficient, especially at the outlying regions of the trajectory distribution, i.e., farther away from the mean values. For many of the cases tested that were further from the population mean, no reference curve was found that ran directly through the measurements of a case being evaluated. It was therefore necessary to develop a procedure to obtain new, artefactual reference trajectories that were nevertheless based on the empirical curves of the reference population (BGS), covered the full range of growth variability, but were still natural human growth trajectories. The solution to this problem was then a method based on the description of

growth trajectories using Functional Data Analysis (FDA) followed by Functional Principal Component Analysis (FPCA).

This method has been described extensively in the literature [28,30]. It is based on splitting the growth into two independent models that describe its phase and amplitude separately. These two parts are then modelled separately using Functional Principal Component Analysis (FPCA), which provides the ability to generate new curves (PCA is generally a generative model). Unlike the SITAR model [9,10], which includes a procedure for applying the trained model to new data, Ramsay et al. [28–30] have not published a procedure for registering the FDA model to new measurements (i.e., how to use information about existing curves to fit a curve to data from a new, evaluated case). Hermanussen and Meigen [14] 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 [12]. In addition to registration, they also thought 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 using the Shifting algorithm, classical PCA methods) and therefore did not use the full potential of the functional data analysis methods described by Ramsay et al. [28–30]. Meigen and Hermanussen [22] developed a web-based application for growth modelling, but did not separate growth phase from amplitude and described growth curves using only one common model. We incorporated the FDA approach with two independent FPCA models (one for phase and one for amplitude) into our method [20] and it is included in the presented web application GROWTH. We used the computational capabilities of the *fda* package for R [27] and described the application of this method in a methodological paper [20], including a visualization of the substeps and a graphical representation of each principal component. The application of FDA and FPCA in GROWTH is described below for a description of program functionality. Simply put, this approach allows us to generate any number of empirically possible growth curves with a fully continuous range of properties based on empirical variability, fill the space missing in the actual empirical trajectories, and thus optimally fit each evaluated case.

Data reference samples

Currently, the program allows the comparison or calculation and construction of individual models based on the longitudinal growth data of four studies:

- (a) BGS – Brno Growth Study [7];
- (b) WLST.M – Wroclaw Longitudinal Study of Twins, monozygotic twins [2,3,18,19];
- (c) WLST.D – Wroclaw Longitudinal Study of Twins, dizygotic twins [2,3,18,19];
- (d) WGS – Wroclaw Growth Study [4,17,36];
- (e) ELSPAC – Czech part of the European Longitudinal Study of Pregnancy and Childhood [21].

Technical solution

The aim 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 datasets. Therefore, the front-end was designed as a web-based application, while the computational core consists of scripts for the R environment [26]. The advantage of the web-based interface is that it is ready to use and does not require software installation on the user's computer, making the application available without much effort.

The R environment [26] is widely used in the research community, so the core uses up-to-date packages according to the state-of-the-art and 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 in the Masaryk University cloud infrastructure (IaaS) (<https://cloud.muni.cz/>) in an OpenStack environment. The R environment is loaded inside the Apache2 server via the R-Apache module (<https://github.com/jeffreyhorner/rapache>), which enables online execution and interpretation of R scripts. The web interface is based on the CMS Wordpress, and a plugin we developed provides the connection to the core R scripts. The transfer of input data and results between the user interface and the computational part is done interactively via Ajax, and 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 started for every user request. In addition, the module supports preloading of data and execution of initialization scripts, so that environment setup occurs only once when Apache2 is started, when R packages and data with trained models are loaded.

Basic scripts for growth model computations and other statistical operations use the *fda* [27], *nlme* [24], *sitar* [9], *gamlss* [31], and *pracma* [6] for computations, *ggplot2* [37], *Cairo* [35], and *png* [34] for visualizing results, and *jsonlite* [23] for communicating with the front-end.

The use of CMS Wordpress allows easy extensibility of the user interface with new features and is suitable for support.

Main features of the program - theoretical description

Fitting data with a growth curve

The program allows the fitting of individual longitudinal data recorded during growth period with a growth model that is based on the empirical variability of the realized growth trajectories of individuals examined in growth studies (reference data). In this sense, the program offers two different approaches:

(1) The closest empirical trajectory

In the first approach, which we refer to as the search for the *closest empirical trajectory*. In doing so, the program fits the more or less complete individual longitudinal data of the reference population – children who have already grown up and have been measured over time – with the selected growth model. The choices are: *Bock* [5], *Preece and Baines model 1–3* [25,32], *SITAR* [1,8,10] a *FDA* [28,30], with the SITAR model currently available for all reference samples. We use these methods to obtain growth curves for all reference specimens. By interpolating on all these curves, we obtain the body height values of these models for the values of age that are measured for the analyzed case. These interpolated height values are then no longer affected to such an extent by measurement noise (error). The program then finds the reference individual from the entire reference sample whose estimates of stature are closest to the case being analyzed based on the *least squares method* of residuals. The growth model of this reference case – the closest empirical growth trajectory – is then used as the model of the case under analysis. The reference data thus enter the process as values on the smoothed curves of the respective models, while the data from the case being evaluated enter the process as raw measurements.

(2) FPCA model

As a second and more advanced form of fitting, the program uses the FPCA model, whose detailed description and some tests have been published in previous studies [16,20]. The approach is based on a statistical generative growth curve model that is fitted to the measured points so that the resulting curve is always biologically plausible and most likely relative to the reference population, while being robust to a reasonable degree of measurement imprecision.

We chose B-spline with sensitively adjusted parameters so that it responds to local growth waves within the 18-year trajectory and respects the data well, but is no longer affected by short-term fluctuations and noise. However, this was only possible due to the large number of regular measurements of the Brno Growth Study (BGS) reference sample, for which it was easy to distinguish the curvature of the whole curve from the rebound/fluctuation of a single point due to isolated measurement inaccuracies. Therefore, the FPCA method is so far only available for the Brno Growth Study reference data.

The data collected in the Brno Growth Study (BGS) [7] contained mostly 39 height measurements of 334 individuals from birth to 18 years of age. Based on these data, two generative growth models were constructed, one for boys and one for girls. Height data for each individual were interpolated using a B-spline curve following the procedure described by Ramsay et al. [29]. These curves were then superimposed using a registration that uses the age at peak growth velocity at puberty (APV) found on each curve based on inflection point searches. This was a key step that allowed the construction of a model describing growth phase and amplitude separately using independent parameters, thus generating time-warping functions describing the shift in growth phase of each B-spline curve relative to the population mean. Using the time-warping functions, the growth curves were aligned to the same phase. The final generative model for each sex was obtained by applying Functional Principal Component Analysis (FPCA) to sets of phase-aligned curves and inverse time-warping functions to generate amplitude and phase modes, respectively.

The generation of a model describing a specific curve corresponding to certain parameters is performed by first generating an amplitude curve, which is then time (phase) fitted/deformed. Fitting the model to the measurements of the newly analyzed case (i.e., fitting this model with the evaluated data) is performed by numerically optimizing the model parameters using a Levenberg-Marquardt solver [15:56–58], minimizing two sets of residuals that describe the differences between the measured and modeled values of body height and the likelihood of the modeled curve, respectively. To ensure the biological plausibility of the fitted model, both phase and amplitude parameter scores are constrained to ± 3 standard deviations during optimization. For further details on the estimation method, see Králík et al. [20].

Thus, by both of these methods – (1) the *closest empirical trajectory* and (2) the *FPCA model* – it is possible to obtain an individual growth curve of an evaluated case, regardless of how many measurements of the evaluated individual are available, in the extreme case from a single measurement. However, the reliability of the

estimates depends on the reliability of the input data, and we can only obtain an approximate result from a single measurement.

Based on the growth model (curve) obtained, it is then possible to:

- 1) Interpolate any values of body height between measurements of the evaluated case;
- 2) Extrapolate beyond the empirical data range, including the prediction of adult stature height;
- 3) Calculate a velocity and acceleration curve for a given individual;
- 4) Compare the growth of the individual being evaluated with reference population(s);
- 5) Compare his/her growth to predictions based on the height of the parents;
- 6) Estimate significant pubertal and adolescent growth milestones;
- 7) Detect growth abnormalities.

Some predictive properties have been tested and are presented in a publication dedicated to the design of the FPCA model [20]. This involves testing the reliability of the estimation of APV and other growth milestones at puberty, depending on how the empirical range of values measured and entered into the model covers (encompasses) the region of the estimated milestone. *Supplementary materials* to this paper include detailed plots of the residuals of the estimates of these comparisons. The results show that the better the empirical data region covers the age of the estimated milestone, the more reliable the estimate, with data from the pre-APV age being more important for its reliable estimation than data from the post-APV age.

All of the procedures implemented in GROWTH require extensive validation on other populations, and we will add the results of these tests over time.

Description of the web application

Input data and calculation settings

Personal Info

Personal code

Enter the name or personal number/code of the individual under investigation here.

Biological sex

Select the biological sex from a choice of male or female. Currently we do not have any reference sample for intersex individuals, also we do not recommend using the app for transgender individuals.

Date of birth

Here we enter the exact date of birth. If an age for the height measurement data is not entered in the next step, but a measurement date is, the program will automatically calculate the age from the birth date and measurement date entered here and use it to calculate the growth model.

Examination Performator

Name and title

Fill in here the identity of the person carrying out the evaluation (medical doctor, anthropologist). These details will be displayed in the output and there will be space for a signature in the final printable report.

Institution

Enter here the institution that will appear in the protocol.

Stature measurements – Measurement inputing

Number of measurements

Select a count for the number of values to be entered and the program will create the appropriate number of empty input data fields. A table is generated with three columns labeled Date, Age (yrs) and Stature (cm), in which you must enter the measurement date (day entered manually or a selection from a calendar table), age (years with decimal point and division into tenths and hundredths) and stature height in centimeters with any number of decimal places after the decimal point. If birth date is entered in the previous step and a measurement date is entered here, but not an age, the age is automatically calculated from the birth date and measurement date and used for the calculation. If both the measurement date and an age are entered, only the age entered manually is used for the calculation and the birth date is ignored.

Closest trajectory number setting

Here you specify the number of nearest empirical growth trajectories you want to display in the *distance growth plot* (A). The nearest empirical growth trajectory is found by interleaving all cases in the reference sample (4 available reference samples) with the selected growth model (its selection is in the next step) and using the least squares method of residuals to find the trajectory whose model is closest to the specified points of the evaluated case.

Model and sample

Longitudinal model and reference sample

Select the growth model that will be used to fit the reference data and select the reference data. You can always select one of several combinations of reference sample and growth model. The models have been listed and cited in the explanation of the main features of the program above and described in the previous paper [16]. The default is the set of Brno Growth Study children whose growth trajectories were fitted with the FDA model.

Compute FPCA model

Select whether to display the results of the main FPCA growth model designed and published by our team [16,20]. In this case, the reference set is always children from the Brno Growth Study.

Percentile reference

A percentile (usually cross-sectional) growth chart derived from the literature and other sources for traditional entry and comparison of growth data for the child being assessed.

Midparent

Enter the height of the mother's stature (in cm) and the height of the father's stature (in cm) in adulthood. These values are further used to calculate midparental height as a basis for the child's target height and to compare the child's height with his or her prediction based on the percentile position of his or her parents in adulthood.

Actions

Save inputs as xml file

Once all the data has been entered, it can be saved to an xml file. The program will save the data you have entered (including the verbal evaluation you may have entered in the *Text Evaluation* sheet) and settings on your computer. The name of the file consists of the case identifier entered as *Personal code* above (e.g. child's name), calendar date and time of saving. The structure of this file can be used to populate other dates.

Load input from xml file

Here it is possible to load data in xml format and structure that matches the structure of the xml file above that the program saves. Thus, it is possible to reload the data of an already entered and saved case, analyze it in an alternative way, add another part of the verbal evaluation, etc.

PROCESS

Button that starts the calculation of the growth model. Only after clicking on it, the program will calculate the growth model based on the entered data and set options, display the results in graphs and prepare tabular and textual outputs of the evaluation of the specified growth curve.

Because optimizing the growth curve is computationally intensive iterative procedure, the entire calculation can take from one to tens of seconds. During this time, the program displays a spinning clock and a notification that the calculation is in progress.

EXPORT PDF

By clicking on this button, the program generates a report of all the analyses performed, including the evaluation in the text field, and saves it on the computer in PDF format.

Results and Outputs

DISTANCE PLOTS

A. Distance growth plot

Distance growth plot of the body height of the evaluated case (**Figure 1**). The blue points represent the entered values. The green points represent the most similar empirical growth trajectory from the selected reference sample, fitted with the model chosen in the *MODEL & SAMPLE* selection (Longitudinal model and sample). The red curve represents the individual growth model calculated by the FPCA method based on the reference data of healthy individuals of the Brno Growth Study. Based on the FPCA model, the ATO (age at take-off or age at minimum growth velocity before puberty) estimate is calculated, shown as the blue dashed vertical, and APV (age at peak velocity or age at pubertal maximum growth velocity) is shown as the red dashed vertical. The horizontal dashed lines show the body height at the respective ATO and APV points. For comparison, the grey curve represents the mean curve of the Brno Growth Study sample for a given sex.

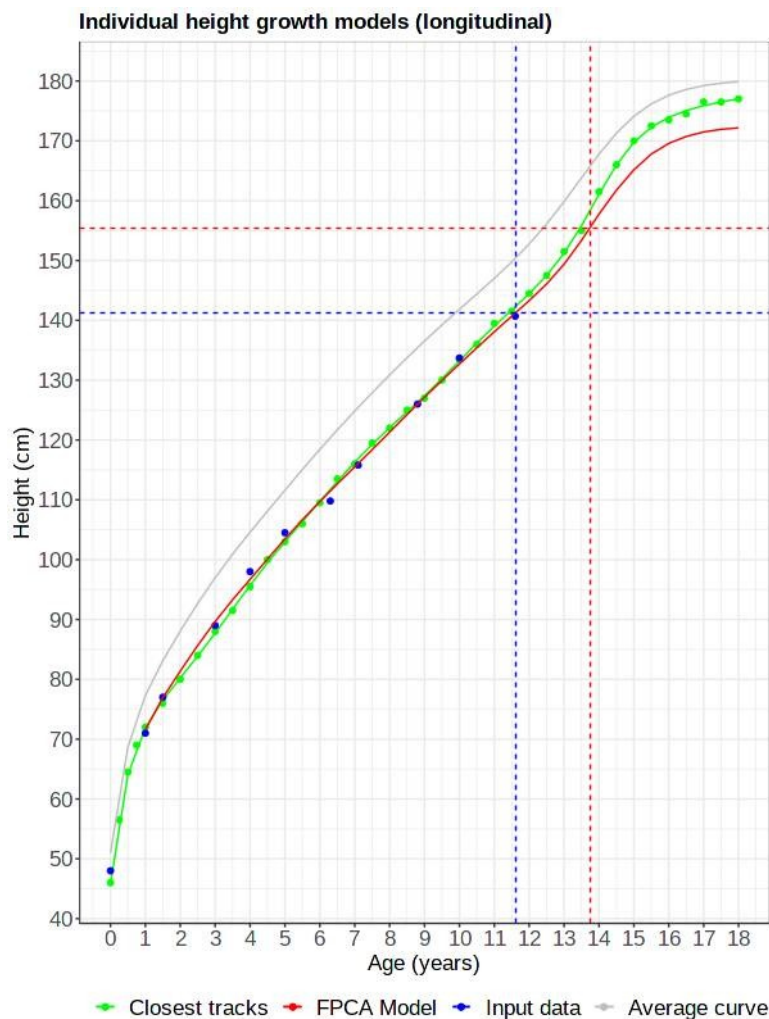


Figure 1 Distance growth plot (A); The assessed individual (blue points) is substantially below the average BGS curve (grey solid line) throughout the recorded growth period, but has a consistent position as the nearest empirical trajectory (green), and the FPCA model (red) respect well the recorded growth of the assessed individual and are very similar to each other over the empirical range of the assessed individual's age, but the prediction (extrapolation) to higher ages diverges, with the nearest empirical trajectory deviating significantly higher towards the mean curve than the generative FPCA model.

Table of values and residuals of growth models

This table expresses the values shown in the *Distance growth plot A*. The first column (No.) is the number of the measurement of the case being evaluated, given as input data. The second column (*Track heights*) represents the value of the stature height (in cm) of the nearest empirical growth trajectory for the relevant calendar age (i.e.

the age at which the measurement of the evaluated individual was taken, denoted as No.); this is an extrapolation from the growth model, not an empirical value measured on a reference individual nearest to the evaluated case. The third column (*Track residuals*) represents the difference (in cm) between the value of the stature height of the evaluated case and the value estimated from the nearest empirical growth trajectory given in the second column as *Track heights*. The following column (*FPCA heights*) represents the value of stature height (in cm) extrapolated from the FPCA growth model of the assessed case for the relevant calendar age. The next column (*FPCA residuals*) then shows the difference (in cm) between the value of the evaluated case height and the value estimated from the FPCA model.

B. Percentile distance plot

Plot of the height growth of the evaluated case (black dots connected by lines) projected onto the reference percentile plot of the population selected when entering the *MODEL & SAMPLE* selection (*Centile sample*). This type of chart (**Figure 2**) is standardly used when entering children's growth data into a paediatric reference frame. However, it is not a longitudinal, but a cross-sectional record, whose percentile boundaries (suggesting certain growth "lines", and thus in the growth standard a kind of subconscious growth "guide line") certainly do not represent any "norm" of the shape of the growth curve for individual growth trajectories.

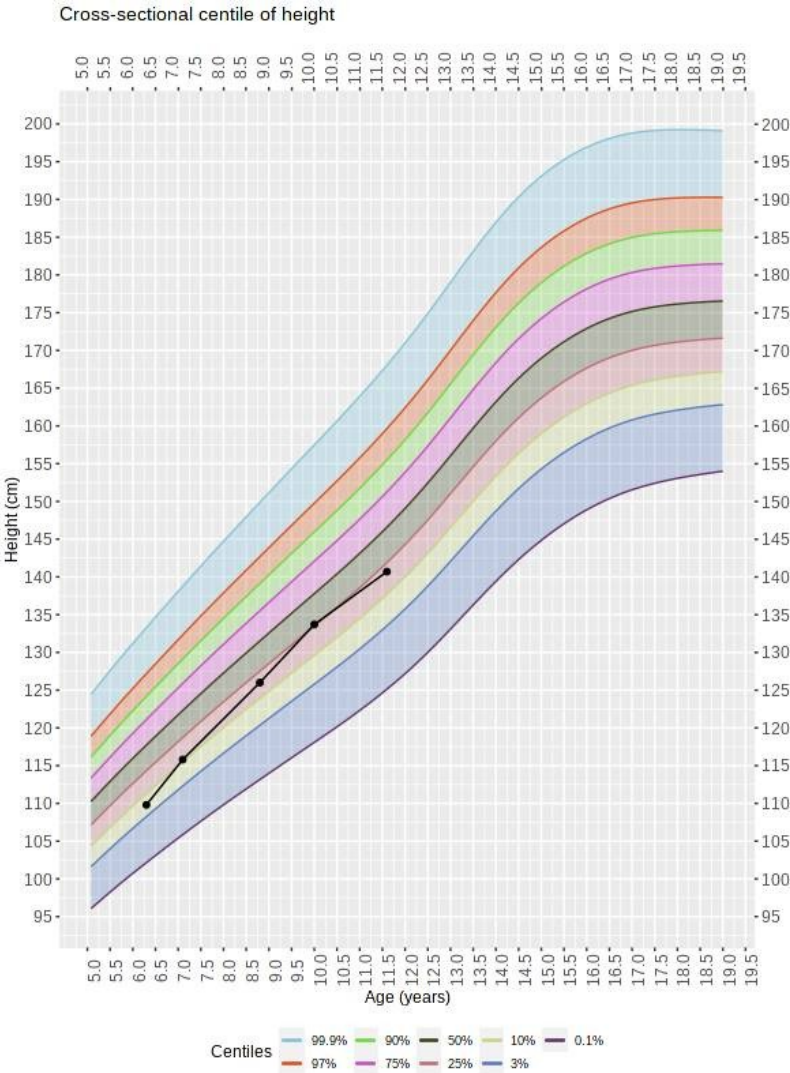


Figure 2 Percentile cross-sectional plot of stature growth in boys (WHO), showing the trajectory (black dots and solid lines) of the case under evaluation; the individual is below the mean of the comparison population throughout, but never below the 3rd percentile, usually between the 10th and 25th percentiles; the course of his curve does not parallel the percentile boundaries, but it is not clear from this display to what extent this exceeds the rate common in the healthy population.

REFERENCE PLOTS

C. Distance plot against reference models

All cases for a given sex of the Brno Growth Study were modeled using Functional Data Analysis (FDA), which is set as the primary (default), or another combination of growth model and reference data (selected when MODEL AND SAMPLE is entered). From the models, values of stature height for the values you specified for the age of the case being evaluated are extrapolated and these extrapolations were plotted using a boxplot. The evaluated individual is always modelled using the FPCA model. The color of each point of the evaluated case then corresponds to the percentile value of that point relative to the reference boxplot at that age. The points of this plot (**Figure 3**) connected by *dotted lines* represent the specified values of the evaluated case (raw measurements), and the points connected by *dashed lines* represent the values of the height interpolation *from the model* for a given calendar age.

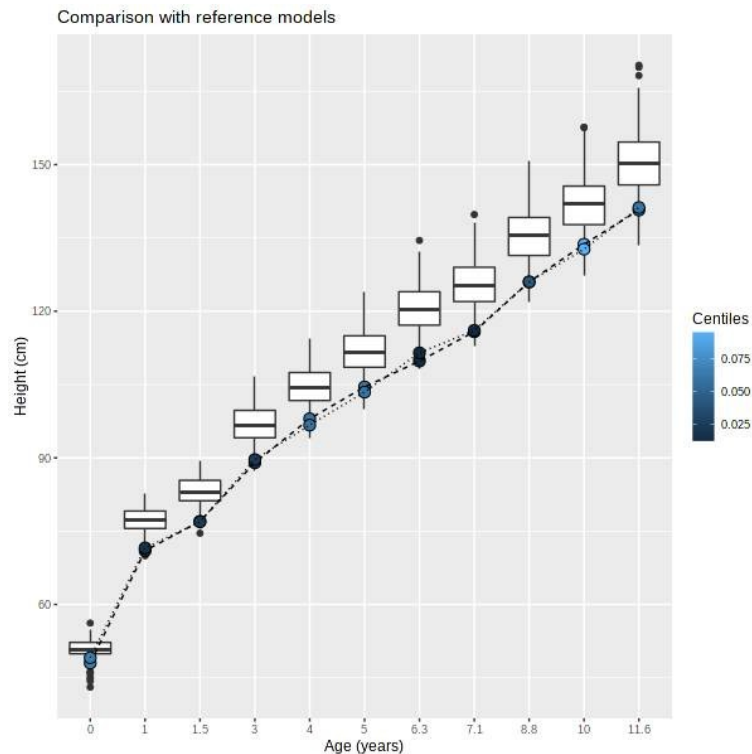


Figure 3 Comparison of the stature height of the evaluated case (colored points, raw values and interpolations from the FPCA model) with the ranges of values of the reference model (boxplots, selected in the *Model and Sample* menu) for the values of the age of the evaluated individual; both the raw measured values of the subject (points connected by a *dashed line*) and the corresponding points of the same measurements interpolated from the FPCA model (points connected by a *dotted line*) show the lower position within the variability of the reference distribution; the raw values and the interpolated values are very close together in this type of display and difficult to distinguish, as is the relative growth trend to the reference population in the full height range.

Table of estimates against the BGS longitudinal reference

This table numerically represents the values shown graphically in Plot A and C. The first column (No.) is the measurement number of the case being evaluated, and the second column (*Measured percent.*) is the raw measurement value expressed as a percentile relative to the BGS reference distribution for that age (points in Plot C connected by a *dotted line*). The next column (*Track percent.*) is the value of stature height, extrapolated from the nearest empirical growth trajectory (Plot A – *green curve*) and again expressed as a percentile of the BGS reference distribution. The last column (*FPCA percent.*) is the value of stature height extrapolated from the FPCA model (Plot A – *red curve*, Plot C – points connected by *dashed lines*) and again expressed as a percentile of the BGS reference distribution. The last row of bolded numbers indicates for all three columns the maximum percentile range over the entire measured trajectory of the individual being assessed. Values derived by interpolation from the models (closest curve, FPCA model) can be expected to have a lower percentile range because they represent smoothed trajectory values with limited influence of measurement noise.

D. Plot of growth velocity against the reference model - calendar age

This plot (**Figure 4**) represents a comparison of the growth velocity of the case being evaluated with the growth velocity of the Brno Growth Study reference sample for the calendar age values of the case under evaluation. It is calculated and displayed only if at least 2 measurements of the evaluated case are given. For each pair of measurements (range of two calendar age values) the program calculates the difference in height, relates it to the range of age and converts it to the growth velocity in centimeters per year. For a given age range, the program calculates an extrapolation of the body height of the reference BGS sample and the growth rate in the same way. The growth rate of the reference file is then displayed as a boxplot and the evaluated case as a point. The points of this plot connected by *dotted lines* represent the growth velocities calculated from the raw values of the evaluated case, the points connected by *dashed lines* represent the growth velocities calculated from the interpolations of the height from the FPCA model of the evaluated case.

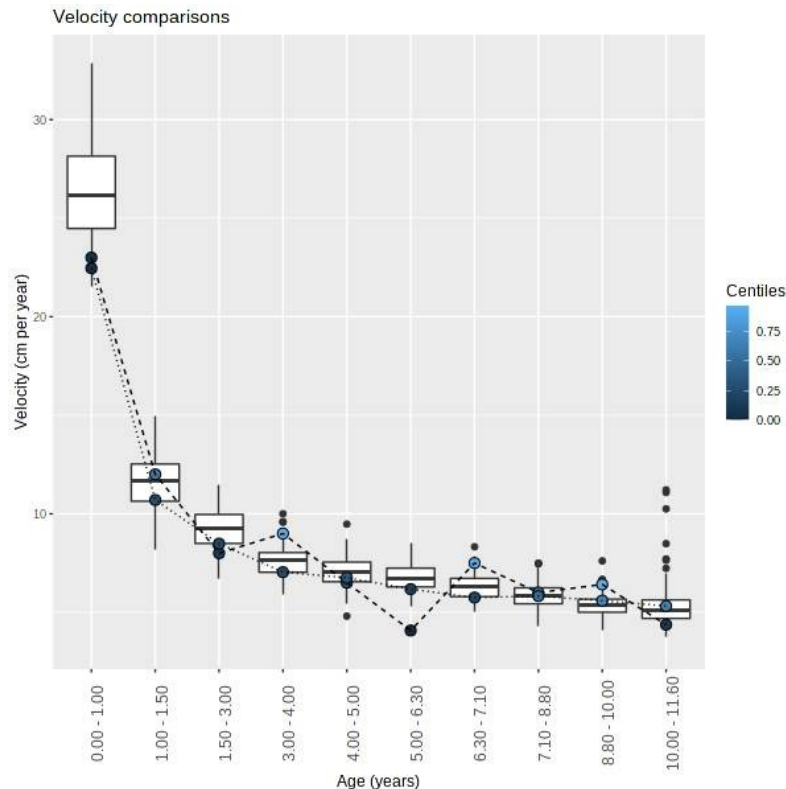


Figure 4 Comparison of the growth velocity of the evaluated case with the reference growth velocity (boxplots) for the values of the age intervals between the measurements of the evaluated case; the growth velocity of the individual is below the average of the reference until about 7 years and then gets above the average. The plot shows very well the difference between the unsteady rates from the raw measurements (points connected with dotted line) and the gradually changing position of the growth rate calculated from the FPCA model curve (dotted line points). Although the measurement error (noise) is not very obvious in the distance plots (all plots above this one), it becomes much more apparent after the growth rate is calculated. Therefore, we recommend taking growth rates calculated from raw data with a large caution (due to the high risk of false positive findings of extreme velocities) and rather evaluate the velocities using values derived from growth models.

Table of growth velocity estimates against the BGS longitudinal reference

The table of growth velocity estimates expresses the values shown graphically in Plot D. The first column (*No.*) is the range of measurement numbers of the evaluated case for which the velocity is calculated. The second column (*Age (years)*) is the value of the range of ages for which the velocity is calculated. The next column (*Measured percentile.*) represents the value of the growth velocity of the evaluated case calculated from the raw measurements (*dotted line* connected points in plot D) as a percentile of the BGS growth velocity reference values for the age range (boxplots in Plot D). The last column (*FPCA percentile*) represents the value of the growth velocity of the evaluated case calculated from the FPCA model (points of plot D connected by *dashed lines*) in the form of percentile of the BGS growth velocity reference values for the given age range (boxplots in plot D). For both columns, the last row of bolded numbers indicates the maximum percentile range over the entire measured trajectory of the assessed individual. Values derived by interpolation from the model (FPCA

model) can be expected to have a lower percentile range because they represent smoothed trajectory values with limited influence of measurement noise.

PARENT REFERENCE PLOTS

E. Plot comparing the child's growth with the assumption from the parents' height in adulthood

This diagnostic plot compares the child's growth to the assumption made from knowing the height of both biological parents in adulthood. From the height of the father and mother of each reference case, the program calculates an estimate of a *target height* based on some modification of *midparental height*; currently it uses the calculation of Gray (1948) [13], but we will gradually add other methods of estimating target height. The current estimate method is as follows:

For boys:

$$\text{target height} = (\text{height of father} + (\text{height of mother} * 1.083333)) / 2$$

For girls:

$$\text{target height} = ((\text{height of father} * 0.9230769) + \text{height of mother}) / 2$$

Based on the target height estimate from the parental data, the program calculates the Z-score to which this target height corresponds at age 18 of the reference population (BGS by default). Next, the interpolation of the height for a given age of measurement is calculated according to the model curve (i.e. the height of each child at each age of measurement adjusted for noise). The difference between each child's height at a given age and the percentile estimate of their height based on the target height percentile (height minus the estimate of height from the target height) is then calculated. Thus, the difference on the scale on both the positive and negative axes corresponds to whether the child is taller or shorter in stature at a given age than would correspond to the percentile of his or her target height estimate (**Figure 5 left**). The distribution of these differences is converted into a standardized distribution (Z-score) using the mean of 0 and the standard deviation of the distribution of these differences of the reference sample of children, and finally, the values of the evaluated case is computed in the same manner and plotted on this standardized scale (**Figure 5 right**). The evaluated case is then also plotted as a Z-score against this reference. If the evaluated case differs in height from the Z-score of the target height (i.e. his or her parents) by more than $\pm 2SD$ (more than about 95% of all cases in the reference population), the points are shown in red and indicate a growth problem. The points in this graph connected by a *dotted line* represent the Z-score calculated from the raw values of the case being evaluated, the points connected by a *dashed line* represent the Z-score calculated from interpolating the height from the selected growth model of the case being evaluated.

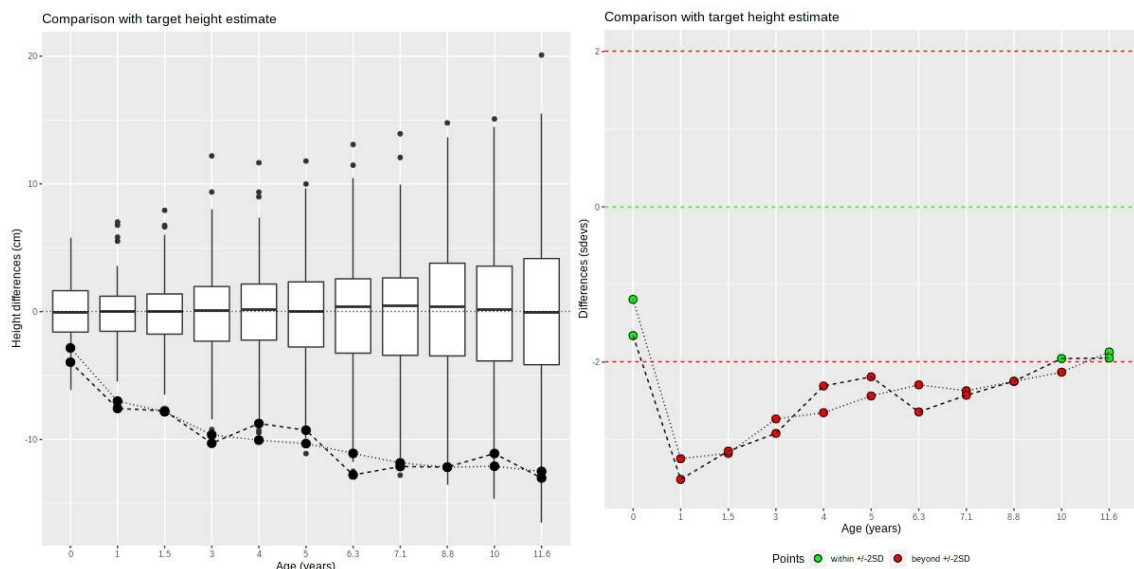


Figure 5 Comparison of the differences between the height of the evaluated case and the assumption from the target height according to the parents in the original values in centimeters (**left**) and the same converted into a standardized scale calibrated by the standard deviations of these differences in the reference population (**right**); the evaluated individual is deeply below average throughout growth, even below the $-2SD$ level of the reference population (red points), but is gradually approaching this threshold.

Table comparing standardized values of the differences between height and the assumption from the target height

This table shows the standardized values of the differences between the child's height and the assumption derived from the parents' target height (points from the left graph of **Figure 5**). In the first column (*No.*) is the measurement number of the evaluated case for which the difference is calculated. In the second column (*Age (years)*) is the value of the age for which the difference is calculated. In the next column (*Measured SD*) is the value of the standardized difference for the raw measurements and in the last column (*FPCA SD*) is the value of the standardized difference for the measurements derived by interpolation from the FPCA model.

EQUIPHASE PLOTS - Phase Synchronized Plots

F. Pubertal growth velocity curve – biological age

The FPCA method is based on two separate FPCAs, one for phase and the other for amplitude, which are summed during model generation. Thus, it allows to analyze both aspects of the growth curve separately. This can be used to synchronize the phase of all growth curves. All the reference curves are registered at point 0, which corresponds to the APV, their phase is neglected and their shape is deformed (using the *landmarkreg* function from the *fda* package [27]) so that the overall magnitude matches the average growth curve of the reference sample. The curves then differ only in amplitude. The curve of the individual being evaluated is identically processed. As a result, we can compare the growth of individuals of the same biological age, free from differences in their calendar age (or from differences in the calendar timing of the pubertal growth phase), in other words: below each other in the graph are values that correspond in the timing of their phases for all individuals. In this plots, we use the first derivative of such synchronized curves and compare the growth velocity of the evaluated case with the growth velocity of a reference set at the same growth age. In Plot F (**Figure 6**) we see a comparison of the phase-synchronized velocity curve of the evaluated case with the average velocity curve of the reference sample during the pubertal growth acceleration period. For this procedure, the results of the FPCA method are needed; if we do not select it when selecting methods (*Compute FPCA model*), the plot will not be displayed.

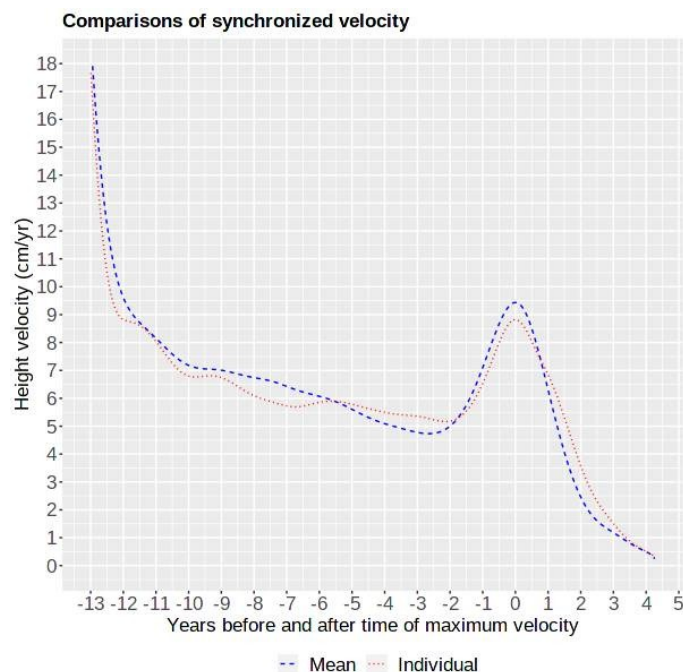


Figure 6 Comparison of the average body height velocity curve of the reference population (blue dashed line) and the phase-synchronized body height velocity growth curve of the evaluated individual (red dotted line). We can assess the growth rate relative to the mean for a given growth phase regardless of calendar age; for example, the evaluated individual grew slower than the mean during the spurt, but faster during the cessation of growth.

G. Plot of growth velocity against a reference model of the same growth age

This plot (**Figure 7**) represents a comparison of the phase-synchronized growth velocity curve of the case under evaluation (points connected by a solid line) with the reference, phase-synchronized growth velocities

values for all measured ages of the case under evaluation (box plots). We can then assess what the growth rate is relative to other individuals in the same growth phase. For this procedure, the results of the FPCA method are needed; if we do not select it when selecting methods (*Compute FPCA model*), the plot will not be displayed.

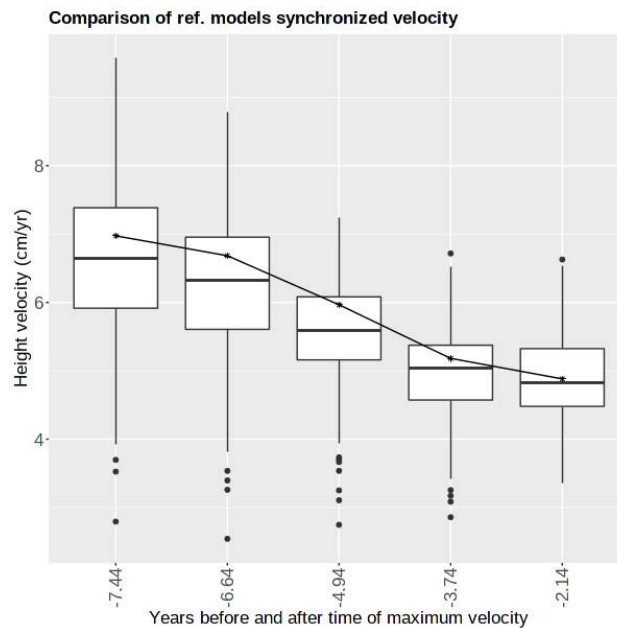


Figure 7 Comparison of the mean velocity curve of the reference population (boxplots) and the phase-synchronized velocity value of the evaluated individual (black points connected by solid line). We can compare the growth velocity of a given individual with the reference at the moments developmentally corresponding to the measurements of the evaluated individual.

Table of the height growth velocity after phase synchronization

The table shows the growth velocity values of body height for the respective growth age, expressed in relation to the APV of the average reference growth velocity curve (in years). The first column (*No.*) is the measurement number of the case being evaluated for which the velocity is calculated. The second column (*Age (years)*) is the value of the age for which the rate is calculated. The next column (*FPCA Percentiles*) is the value of the growth rate for that measurement after synchronization. The last row is the overall range of the standardized distribution over which the speed of the individual being evaluated varies over the entire set of comparisons (once is the standard deviation of the growth rate).

TARGET HEIGHT

Calculating the target height

Target height is an estimate of the height a child will reach as an adult (or at age 18) based on currently available information. Since genetic predispositions (“genetic growth potential”) resulting from the combined genetic contribution of both parents have always been considered as a strong determinant, the target height of a child is often derived from the body height of the biological mother and biological father. Different methods of calculating the target height then take into account in different ways the height of the mother and father, the sex of the child and possibly also secular trends (average intergenerational differences) in stature or differences between populations more generally. However, genetic theory suggests that each child (even within siblings of the same sex) of the same parents represents a slightly different combination of the father's and mother's traits, so it can be assumed that different children of the same sex of the same parental couple will in fact have different heights in adulthood, even though the target height of the child of the two parents will be only one.

The table of results currently includes these estimates:

Midparent 1 – the method of Tanner et al. [33], represents the average of the father's and mother's stature (midparental height), to which 6.5 cm is added for boys and the same value is subtracted for girls. This method does not take into account a possible secular trend in increasing height in the 20th century.

Midparent 2 – the method of Molinari et al. (1984), follows a similar procedure, but adds 6.4 cm to the midparental height for boys and subtracts 6.4 cm for girls, but accounts for the secular trend in height increase by adding 3.8 cm to all estimates.

Centile – this method obtains the Z-score of the mother's height relative to a reference sample of mothers of BGS children, and similarly the Z-score of the father's height relative to a reference sample of fathers of BGS children, and calculates the average of the two Z-scores. This value is then multiplied by the standard deviation of the interpolations of the BGS children's growth patterns (of the relevant sex) for age 18, and this multiple is finally added to the average value of the interpolation of the BGS children's stature height (of respective sex) for age 18, thus getting back to the stature height scale in centimeters and obtaining an estimate of the stature height of the individual being assessed at age 18. This method addresses the intergenerational difference (secular trend) in absolute stature height between the parents' and children's generation by transferring it through the relative Z-score, but *de facto* increases the stature height of children relative to their parents to a degree corresponding to the difference in their means. It is therefore appropriate in a situation where the difference between the parents' and children's generation will be similar to the BGS study.

Linear Regression – our original multiple linear regression method, based on BGS data, where child height is estimated directly from parental height by fitting it to a linear equation.

For boys, the equation takes the form:

Son's height at adulthood (cm) = 21.36057 + 0.52969*mother's height (cm) + 0.41558*father's height (cm)

For girls, the equation takes the form:

Daughter's height at adulthood (cm) = 30.62437 + 0.48017*mother's height (cm) + 0.33279*father's height (cm)

This regression model is derived from the BGS and accounts for the specific change between the generation of parents and the children born to them in the 1960s in Brno.

Longitudinal – this estimate of target height is the only one included so far in the Program that does not take into account the height of the parents, but the previously measured height of the child whose target height we are estimating. The measured stature heights to date are fitted with the FPCA model and the target height estimate is an extrapolation of the stature height for this model at age 18. The fewer the number of measurements of the child available and the further the age at the time of measurement from age 18, the less accurate the estimate will be. At the same time, given the behavior of the model [20], it can be expected that in this case the estimates will tend towards average values, i.e. the stature height of smaller individuals will tend to be overestimated and the stature height of taller individuals will tend to be underestimated.

GROWTH MILESTONES

Calculation of growth milestones at puberty

From the growth curve obtained for the individual under evaluation, the program calculates growth milestones using the *getPeak* and *getTakeoff* functions from the *sitar* package [9,10] available for the R environment [26]. For each individual modeled, the following growth milestones/parameters are determined from the model curves:

APV (Age at Peak Velocity) - age at peak velocity at pubertal growth spurt in body height (in years).

VPV (Velocity at Peak Velocity) - maximum velocity at pubertal height growth spurt (in cm/year).

HPV (Height at Peak Velocity) - body height at the time of peak velocity in puberty (in cm).

ATO (Age at Take Off) - age of onset of pubertal height spurt (in years), corresponds to the point of lowest growth velocity before puberty, the position of the foot at the beginning of the pubertal velocity wave.

VTO (Velocity at Take Off) - velocity at the ATO point, the lowest velocity of the pre-pubertal growth spurt of body height (in cm/year)

HTO (Height at Take Off) - body height at the ATO point, the height reached at the start of the pubertal growth spurt in height (in cm).

TEXT EVALUATION

In this text field, the person evaluating the body height (physician, anthropologist) can enter his/her textual evaluation of the program outputs, which will then also appear in the report downloaded in PDF format. It is recommended to follow the individual outputs and comment on the message of the corresponding table or plot. The main focus should be on situations where the values of the examined person deviate from the specified

reference frame. In some plots the degree of deviation should be assessed using scale units, in others there are defining limits and colored highlighting of deviating values.

Controlling other parts of the application

Similar to the height analysis part of the app, two other sections are dedicated to body weight and body mass index (BMI) analysis. Switching to these variables is possible in the top bar of the program.

However, the range of analyses incorporated is smaller and includes in this version of the application the basic graphs of models based on the SITAR method [10], as the trajectories are much more variable and may include negative changes (decreasing values) even during periods of growth.

References

- [1] Ken J. Beath. 2007. Infant growth modelling using a shape invariant model with random effects. *Stat. Med.* 26, 12 (2007), 2547–2564. DOI:<https://doi.org/10.1002/sim.2718>
- [2] P. Bergman (Ed.). 1988. *Bliźnięta wrocławskie Vol. I*. Państwowe Wydawnictwo Naukowe, Wrocław.
- [3] P. Bergman (Ed.). 1995. *Bliźnięta wrocławskie Vol. II*. Arboretum, Wrocław.
- [4] T Bielicki and H Waliszko. 1975. Wrocław Growth Study. Part I: females. *Stud Phys Anthr.* 2, (1975), 53–83.
- [5] R. Darrell Bock, Howard Wainer, Anne Petersen, David Thissen, James Murray, and Alex Roche. 1973. A Parameterization for Individual Human Growth Curves. *Hum. Biol.* 45, 1 (1973), 63–80.
- [6] Hans W. Borchers. 2021. *pracma: Practical Numerical Math Functions*. Retrieved from <https://CRAN.R-project.org/package=pracma>
- [7] Marie Bouchalová. 1987. *Vývoj během dětství a jeho ovlivnění. Brněnská růstová studie*. Avicenum, Zdravotnické nakladatelství, Praha.
- [8] T. J. Cole, H. Pan, and G. E. Butler. 2014. A mixed effects model to estimate timing and intensity of pubertal growth from height and secondary sexual characteristics. *Ann. Hum. Biol.* 41, 1 (January 2014), 76–83. DOI:<https://doi.org/10.3109/03014460.2013.856472>
- [9] Tim Cole. 2020. *sitar: Super Imposition by Translation and Rotation Growth Curve Analysis*. Retrieved from <https://CRAN.R-project.org/package=sitar>
- [10] Tim J Cole, Malcolm D C Donaldson, and Yoav Ben-Shlomo. 2010. SITAR—a useful instrument for growth curve analysis. *Int. J. Epidemiol.* 39, 6 (December 2010), 1558–1566. DOI:<https://doi.org/10.1093/ije/dyq115>
- [11] Martin Čuta. 2014. *Modelování lidského růstu. Dynamický fenotyp*. Akademické nakladatelství CERM, Brno.
- [12] B. S. Gottfried and J. Weisman. 1973. *Introduction to Optimization Theory*. Prentice Hall, Englewood Cliffs, NJ.
- [13] H Gray. 1948. Prediction of Adult Stature. *Child Dev* 19, 3 (1948), 167–175.
- [14] M Hermanussen and C. Meigen. 2007. Phase variation in child and adolescent growth. *Int. J. Biostat.* 3, 1 (2007).
- [15] C. T. Kelley. 1999. *Iterative Methods for Optimization*. Society for Industrial and Applied Mathematics. DOI:<https://doi.org/10.1137/1.9781611970920>
- [16] Ondřej Klíma, Martin Čuta, Lenka Polcerova, Pavel Zemčík, Anna Škultétyová, Dominik Černý, and Miroslav Králík. 2021. GROWTH: Functional Data Analysis of the human height postnatal growth in an on-line application. *Anthropol. Integra* 12, 2 (2021), 7–17. DOI:<https://doi.org/10.5817/AI2021-2-7>
- [17] Sławomir Koziel. 2001. Relationships among tempo of maturation, midparent height, and growth in height of adolescent boys and girls. *Am. J. Hum. Biol.* 13, 1 (2001), 15–22. DOI:[https://doi.org/10.1002/1520-6300\(200101/02\)13:1<15::AID-AJHB1002>3.0.CO;2-P](https://doi.org/10.1002/1520-6300(200101/02)13:1<15::AID-AJHB1002>3.0.CO;2-P)
- [18] Miroslav Králík, Stanisław Gronkiewicz, Pavlína Ingrová, and Sławomir Koziel. 2014. Changes in digit ratio during puberty: X-ray sample from the Wrocław Longitudinal Study of Twins. In *Mikulov Anthropology Meeting*, Sandra Sázelová, Adela Hupková and Tomáš Mořkovský (eds.). Academy of Sciences of the Czech Republic and Masaryk University, Brno, 187–189. Retrieved from <http://www.sci.muni.cz/anthrop/wp-content/uploads/2015/06/013-MAM-komplet.pdf>
- [19] Miroslav Králík, Pavlína Ingrová, Sławomir Koziel, Adela Hupková, and Ondřej Klíma. 2017. Overall trends vs. individual trajectories in the second-to-fourth digit (2D:4D) and metacarpal (2M:4M) ratios

- during puberty and adolescence. *Am. J. Phys. Anthropol.* 162, 4 (2017), 641–656. DOI:<https://doi.org/10.1002/ajpa.23153>
- [20] Miroslav Králík, Ondřej Klíma, Martin Čuta, Robert M. Malina, Sławomir M. Koziel, Lenka Polcerová, Anna Škultétyová, Michal Španěl, Lubomír Kukla, and Pavel Zemčík. 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 (2021). DOI:<https://doi.org/10.3390/children8100934>
- [21] Lubomír Kukla. 2008. Evropská dlouhodobá studie těhotenství a dětství - ELSPAC. *Vox Pediatr.* 8, 5 (2008), 24–26.
- [22] C Meigen and M Hermanussen. 2003. Automatic analysis of longitudinal growth data on the Website willi-will-wachsen.de. *Homo* 54, 2 (2003), 157–161. DOI:<https://doi.org/10.1078/0018-442x-00067>
- [23] Jeroen Ooms. 2014. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *ArXiv14032805 StatCO* (2014). Retrieved from <https://arxiv.org/abs/1403.2805>
- [24] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2020. *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-148*. <https://CRAN.R-project.org/package=nlme>. Retrieved from <https://CRAN.R-project.org/package=nlme>
- [25] M. A. Preece and M. J. Baines. 1978. A new family of mathematical models describing the human growth curve. *Ann. Hum. Biol.* 5, 1 (January 1978), 1–24. DOI:<https://doi.org/10.1080/03014467800002601>
- [26] R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- [27] J. O. Ramsay, Spencer Graves, and Giles Hooker. 2020. *fda: Functional Data Analysis*. Retrieved from <https://CRAN.R-project.org/package=fda>
- [28] J. O. Ramsay and B. W. Silverman. 2005. *Functional Data Analysis* (2th ed.). Springer Science+Business Media, Inc., New York.
- [29] James O. Ramsay, Giles Hooker, and Spencer Graves. 2009. *Functional Data Analysis with R and MATLAB*. Springer, Dordrecht - Heidelberg - London - New York. Retrieved from DOI 10.1007/978-0-387-98185-7
- [30] James O. Ramsay and Bernard W. Silverman. 2002. *Applied Functional Data Analysis: Methods and Case Studies* (1st Edition ed.). Springer-Verlag, New York - Berlin - Heidelberg.
- [31] R. A. Rigby and D. M. Stasinopoulos. 2005. Generalized additive models for location, scale and shape, (with discussion). *Appl. Stat.* 54, (2005), 507–554.
- [32] Adrian Sayers, Mike Baines, and Kate Tilling. 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. *Ann. Hum. Biol.* 40, 3 (2013), 298–299. DOI:<https://doi.org/10.3109/03014460.2013.772655>
- [33] J M Tanner, H Goldstein, and R H Whitehouse. 1970. Standards for children’s height at ages 2-9 years allowing for heights of parents. *Arch. Dis. Child.* 45, 244 (December 1970), 755–762. DOI:<https://doi.org/10.1136/adc.45.244.755>
- [34] Simon Urbanek. 2013. *png: Read and write PNG images*. Retrieved from <https://CRAN.R-project.org/package=png>
- [35] Simon Urbanek and Jeffrey Horner. 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*. Retrieved from <https://CRAN.R-project.org/package=Cairo>
- [36] A Waliszko and W Jedlińska. 1976. Wrocław Growth Study. Part II: Males. *Stud Phys Anthr.* 3, (1976), 27–48.
- [37] Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>