3

Watching Children Grow Taught Us All We Know

James O. Ramsay
McGill University, Montréal, QC

Michael Hermanussen
Christian-Albrechts-Universität, Kiel, Germany

A statistician, like Alfred Hitchcock, aims to reveal the shocking in seemingly ordinary data. His or her technology is a large variety of mathematical models, powerful computational devices and a wide selection of data display methods. But in the end, great statistics require great data; meaning information that many people care deeply about, that has a structure that is genuinely informative, and in a quantity that provides answers with useful precision. Data on growth of children works on all three counts; and in this chapter a Canadian statistician and a German pediatrician/auxologist team up to surprise the reader and record our own discoveries along the way.

3.1 Introduction

It is often not the reader of statistical analyses who gets the first shocks. The statistician arrives at the construction site, opens his toolbox, and discovers that there is nothing there that precisely meets the requirements of the job. This collision between science and data is where much of the evolution in the discipline of statistics happens, and we highlight a number of challenges that we encountered in our work together. It was unexpected that the study of human growth would require so many new ideas and approaches. These are now bundled into the larger enterprise of functional data analysis, which is methodology for the analysis of data distributed over continua such as time and space.

What, then, are functional data? The left panel of Figure 3.1 illustrates much of the answer. The data, from Hermanussen et al. (1998), are measurements, accurate to a tenth of a millimeter, of the length of a baby's lower leg

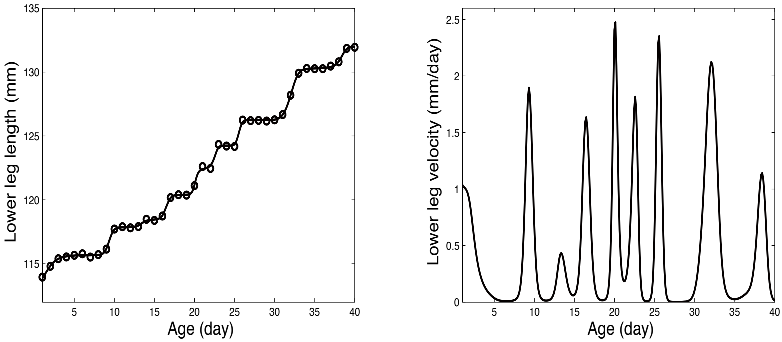

FIGURE 3.1: The left panel contains the measurements of lower leg length for a baby on its first forty days, along with a smooth monotonic curve estimated from these data. The right panel displays the first derivative or height velocity for the fitted curve.

over its first 40 days, the lower leg length being around eleven centimeters at birth. The plot also contains a smooth curve that approximates each point. Of course we have many babies available for measurement, and so we can think of these data as representative of a reasonably large sample of growth curves. We were amazed by the pulse-like nature of growth, with changes over a single day of two or more millimeters, or about three percent. These pulses are separated by three or four days of negligible growth, which seems essential since a sustained 3% per day growth would produce in a year a lower leg length of over three kilometers! If this were a heating system, we'd look for the thermostat that shuts the furnace down to prevent overheating.

The right panel of Figure 3.1 presents the first derivative of the curve in the left panel, which we will call velocity. There we see eight complete growth peaks, suggesting a growth spurt roughly every five days. There is some suggestion that later peaks tend to be smaller, broader, and spaced farther apart.

Functional data analysis is the exploration of variation in samples of curves, most often defined over time, but also over space, wavelength, and other continua. The data themselves are seldom continuous, but we assume that they are sufficiently accurate and sufficiently densely sampled that a curve fit to them can be considered to have an error that is small relative to the inter-curve variation that we wish to study. Functional data analysis has a short but vigorous history. The first paper to use the term was by Ramsay and Dalzell (1991). The field was defined more completely in the monographs by Ramsay and Silverman (1997), Ramsay and Silverman (1997, 2005) and Ramsay et al. (2009).

3.2 Modeling Growth

Since data, like our lower leg lengths, do not conform to any obvious mathematical model, we need a flexible strategy for expressing an arbitrary curve in mathematical notation to any desired accuracy. The oldest and still the most useful approach is to express a function $x(t)$ as a blend of known functions $\phi_k(t)$, called basis functions. That is, $x(t)$ is expressed as the linear combination

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t). \tag{3.1}$$

By manipulating and estimating the coefficients c_k, and making K large enough, we hope to fit data as well as we wish. A function family very much in vogue for this purpose are splines, constructed by joining polynomials together at fixed locations called knots, in such a way as to ensure a smooth transition over a knot from one polynomial to another. The more knots there are over the interval of approximation (the first 40 days in this example), the more flexible the fit to the data.

But here's the first nasty surprise. We expect growth to be positive, so we need a curve that is monotone, or everywhere at least non-decreasing. Series expansions like (3.1) don't do this well; if they fit the data closely, they will almost surely violate monotonicity. Or, to put the problem another way, monotonicity is a powerful but demanding constraint on a function's shape, and the science of human growth, auxology, requires that growth models have this feature.

Growth, as we have said, is all about change in a data-fitting curve x, and we keep our notation clean by using $v(t)$ (for velocity) rather than dx/dt. Now since v is required for our purposes to be everywhere positive, it can be expressed as the exponential of an unconstrained function W, i.e.,

$$v(t) = e^{W(t)} \quad \text{or} \quad W(t) = \log v(t). \tag{3.2}$$

This is a differential equation in x, and the solutions to the equation can be expressed as

$$x(t) = C_0 + \int_0^t e^{W(v)} \mathrm{d}v, \tag{3.3}$$

where C_0 is a constant indicating the lower limb length of the baby at birth. We have now expressed two constrained functions, x and v, in terms of an unconstrained function W, and this function in turn can be expressed in terms of a basis function series. Moreover, this is a universal expression for a strictly monotonic differentiable function. Ramsay (1998) first articulated this approach, and provided fast and reasonably accurate numerical methods for computing values of x.

3.3 Phase/Plane Plots of Derivative Interactions

Conventional plots like those in Figure 3.1 display what we see at the level of the data or a single derivative, but don't convey much information about how the function and one or more of its derivatives interact. In part, this limitation arises because one of the axes is used to display variation over something that we already understand, namely clock time. A good graph highlights the informative but downplays the obvious. Can we rise to this second challenge by replacing time by something more revealing?

The capacity to accurately estimate, inspect and model features of functions like their slope and curvature is the aspect of functional data analysis that most separates it from multivariate data analyses of the raw discrete data. The concept of energy, defined in physics as the capacity to do work, leaps to our attention in Figure 3.1 and seems natural in a context where growth represents work done in terms of adding to a baby's mass and size. A baby adds to its energy pool by ingesting nutrients in order to compensate for energy lost through growth.

We now want to consider acceleration in lower leg length, denoted here as $a(t) = d^2x/dt^2$, as offering further information on growth. The following equation offers a useful breakdown of what produces a change $v(t_2) - v(t_1)$ in velocity over two times $t_2 > t_1$:

$$v(t_2) - v(t_1) = w(t_1)v(t_1)(t_2 - t_1).$$

That is, a change in velocity is proportional to (i) the current velocity (usually the bigger the velocity, the bigger the change) and (ii) the time difference (the bigger this difference, the bigger the potential change). The instantaneous constant of proportionality is $w(t)$, which may be positive, negative or zero, but will most likely vary with t. Dividing both sides by the time difference and allowing this difference to go to zero leads to

$$a(t) = w(t)v(t). \tag{3.4}$$

This is a first order differential equation in velocity $v(t)$, the solutions to which are of the form (3.2), where $W(t)$ in (3.3) is given, for some constant C_1, by

$$W(t) = C_1 + \int_0^t w(z)\mathrm{d}z.$$

Figure 3.2 shows how velocity and acceleration are coupled together over these forty days by plotting acceleration as a function of velocity. Each peak appears as a balloon shape with the mouthpiece at point $(0, 0)$ where both velocity and acceleration vanish. On the up-side of each peak, both the acceleration and velocity are positive, whereas the down-side has negative acceleration. The symmetry of each peak shows up as symmetry of these balloons

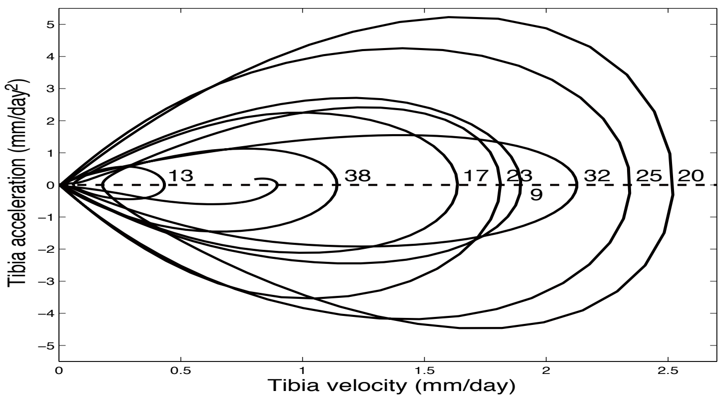

FIGURE 3.2: The second derivative or height acceleration for the baby plotted against height velocity. This is called a phase/plane plot. The numbers to the right of each loop are the day of the corresponding peak velocity in the right panel of Figure 3.1.

about a horizontal zero axis. The “power” of a peak, or work done per unit time, corresponds roughly to the area of the corresponding balloon. The two thinnest loops correspond to the last two peaks, which seem here as well as in Figure 3.1 to have less power. Time now plays the much humbler role of separating one balloon from another. We use the term phase/plane plot, borrowed from dynamic systems jargon, for a plot of one derivative against another.

How long does the bursty nature of growth that we see in Figures 3.1 and 3.2 persist? Thalange et al. (1996) report daily data on the growth of a ten-year old boy, and we estimate from their data the same phase/plane diagram shape as Figure 3.2, except that the velocity peaks reach only .5 mm/day, the acceleration swings through $\pm.05\,\mathrm{mm/day}^2$, and the peaks are about 100 days apart. It is remarkable that velocities between the peaks are no more than a tenth of the peak heights. Applications like these of functional statistical methods have led auxologists to some profound rethinking about the nature of human growth.

Let’s see what the phase/plane plot brings to the study of growth over the entire first 18 years of a child’s life. The left panel of Figure 3.3 shows the heights of ten girls in the Berkeley growth study (Tuddenham and Snyder, 1954) along with strictly monotone fitting curves, and the right panel displays the corresponding height velocities. Notice the pubertal growth spurt, which is a peak in velocity for each girl at times varying around about 12 years. Figure 3.4 presents the acceleration curves. Data like these in Figure 3.3 are typical of older growth studies, including the Fels (Roche, 1992) and Zürich (Falkner, 1960) studies, where the focus was on long-term skeletal growth, and

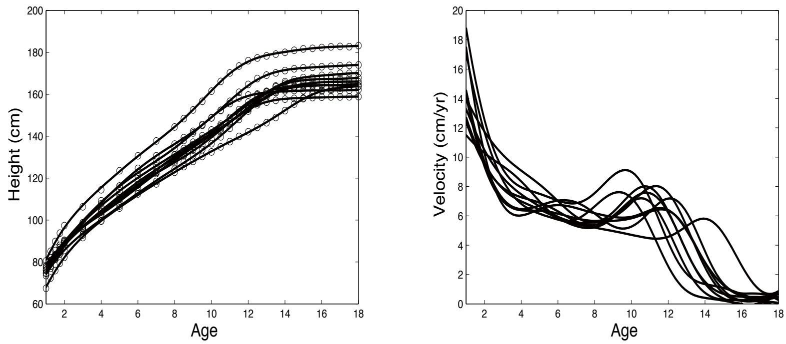

FIGURE 3.3: The left panel shows the height measurements and fitted curves for the first ten girls in the Berkeley growth data. The right panel displays the corresponding height velocity functions.

it was not considered necessary to measure heights any more frequently than twice a year. We now know better.

Figure 3.5 plots acceleration $a(t)$ against velocity $v(t)$ for each girl. Now we see that there are three distinct phases to long-term growth. The first, starting from the lower right of the figure, shows a nearly linear relation between acceleration and velocity such as would be exhibited by an exponential decay process, and that is precisely what we see in the left panel of Figure 3.3 where velocity decreases exponentially up to the age at which the pubertal growth spurt (PGS) begins. The second phase, which begins quite abruptly for some girls, consists of a loop spanning the pubertal growth spurt, quite similar in shape to the balloon in the baby's phase/plane plot. In this case, however, there is no $(0, 0)$ point in the loop, but rather the loops center on

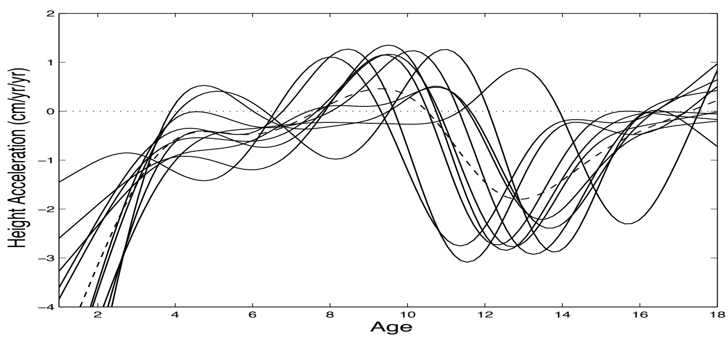

FIGURE 3.4: The height acceleration functions for the first ten girls in the Berkeley growth data. The dashed line indicates the mean of these curves.

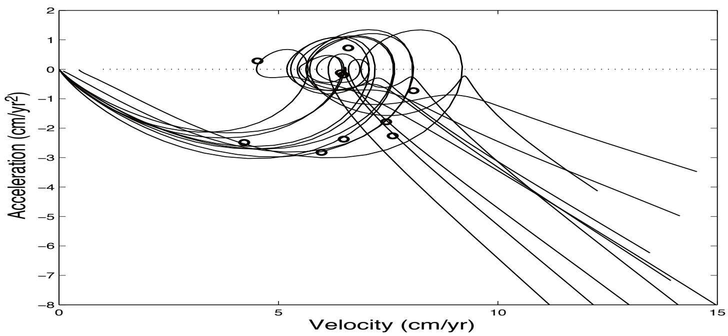

FIGURE 3.5: A phase/plane plot of the fitted curves for the first ten girls in the Berkeley growth data. The circle on each curve indicates the mean center of the pubertal growth spurt, 11.7 years.

a velocity of about seven cm/year and zero acceleration. Finally, there is the half-balloon phase in which both velocity and acceleration decay to zero as the girl approaches her adult height. Of course, we now know that all this is too simple, since right up to puberty there are distinct cycles between high and low intensity growth. The PGS is only the final growth blowout, after which energy is diverted to, among other things, reproductive behavior.

We will say more about this interesting plot in the next section, but already we can conclude that plots of one derivative against another can highlight a lot of information that might escape our attention in ordinary data plots. In effect, a phase/plane plot is a graphical analog of a differential equation like (3.4), where a higher order derivative is expressed as a function of lower order derivatives and the function itself.

3.4 Identification and Analysis of Phase or Tempo Variation

The PGS is an example of what we call a feature in a curve, typically seen as a peak, valley or crossing of a fixed threshold. In Figure 3.5 we have put a small circle on each curve at the average PGS age, 11.7 years; girls having an early PGS will, at 11.7 clock years, be already in the final wind-up phase of growth in the left portion of their curves, while girls with late PGS's will have the small circle in the right lead-up portion prior to the loop. We see two types of variation across curves: in intensity or size of features, called amplitude variation, and in timing of the feature, called phase variation in classical mathematics. Because "phase" carries such a heavy semantic burden in the mathematical sciences, we rather like borrowing the term "tempo" from music for the idea of time as an elastic medium.

The drastic consequences for statistical methodology of mixing tempo and amplitude variation were not obvious to Ramsay and Silverman (1997, 2005) when they began their work. These consequences are nicely illustrated by the dashed line in Figure 3.4, which displays the cross-sectional mean of the ten acceleration curves. The mean, which is supposed to capture typical behavior in data, fails miserably here. The amplitude of the mean PGS acceleration is less than that of any observed curve, and the duration of the PGS shown in the mean as peak-to-valley time substantially exceeds the four years that most girls actually take.

One can presume that, if the mean doesn't work, not much else in classical statistics will either. Such is the case; standard deviation functions, correlation surfaces and other summaries are distorted by tempo variation, and data exploration tools such as principal components analysis show that even a small amount of tempo variation in the data inflates the number of modes of important variation detected. An approach to quantifying the amount of tempo variation developed by Kneip and Ramsay (2008) estimates that over 50% of the variation in the Berkeley growth data is in tempo, and similar percentages have been reported for the Fels and Zürich data. Moreover, one sees tempo variation in almost all sets of functional data; for example, about 35% of seasonal variation in temperature is due to tempo, which is mostly attributable to the early or late arrival of winter.

Our third challenge, then, was to discover either how to simultaneously estimate tempo and amplitude variation, or to first estimate tempo variation and then remove it from the data so as to leave pure amplitude variation. We called the process curve registration. Conceptually, at least in the growth context, tempo variation is due to growth evolving at different rates for different girls. This is to say that there is a physiological or growth time associated with each girl's growth relative to which the PGS is always centered on 11.7 growth years; and growth time and clock time are nonlinearly related. We denote this relationship as "clock time $= h_i$(growth time)" for the ith girl, and we call function h_i that girl's time-warping function. Each function h is, of course, strictly monotonic, so that we can consider time, too, as a growth process, and we can, at least for convenience, add the additional restriction $h(0) = 0$ and $h(18) = 18$.

An elementary tempo estimation method called landmark registration calculates for each girl a smooth transformation of clock time such that transformed feature times all occur at corresponding fixed values. For the growth curves, the single landmark time is the time at which the acceleration curve crosses zero with negative slope, and this is visible in all curves, so that landmark registration works well and is easy to apply. The left panel of Figure 3.6 shows the warping functions h for each of the ten girls; values above the diagonal indicate late growth, and values below this diagonal indicate early growth. The right panel of Figure 3.6 shows the registered velocities, along with the

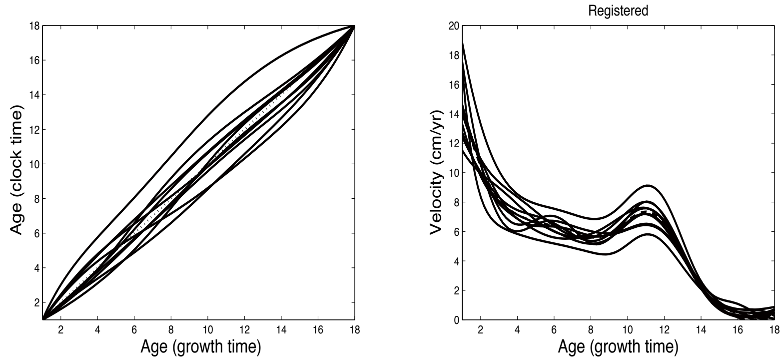

FIGURE 3.6: The left panel contains the warping functions $h(t)$ transforming clock time into growth time for the first ten girls in the Berkeley growth data. Warping functions above the dashed diagonal line correspond to late growth, and those below correspond to early growth. The right panel shows the registered velocity functions for the first ten girls in the Berkeley growth data. The dashed line indicates the mean of these velocity curves.

mean of these velocities. These curves represent pure amplitude variation, and now we see that the mean does a nice job of summarizing the data.

However, because landmarks are not always distinct, and registering all features to a fixed time can seem artificial, more powerful continuous registration methods have been developed by an ever-increasing number of statisticians beyond what is detailed in Ramsay and Li (1998) and Ramsay and Silverman (2005). Indeed, the mathematical analysis of tempo and amplitude variation is now taking us deeply into differential geometry, and it appears that there is much yet to learn about this fascinating problem.

Once amplitude and tempo variation are separated, it is natural ask if there are any patterns in their joint variation. In both Figure 3.4 and Figure 3.5 we notice that early PGS's tend to be intense, as reflected in the large swing between positive and negative acceleration and the larger loops in the phase/plane plot. Conversely, late PGS's are usually mild. This makes sense from the standpoint of growth energy: girls with early PGS's need large growth power during puberty in order to compensate for the extra years of growth that late puberty girls enjoy, and as a consequence adult height does not depend very much on the timing of the PGS.

3.5 Conclusion

The fundamental theme in this chapter is the power of differential equations for modeling functional data. The importance of differential equations as models for data has led to new approaches to the estimation of the parameters that define these equations Ramsay et al. (2007). Growth data have inspired much of the functional data analysis enterprise, but many other data contexts could have served the purpose of motivating dynamic systems models in statistics just as well. Ramsay and Silverman (2005), Ferraty and Romain (2011) and many more current publications promise a new era of partnerships between statisticians and the rapidly increasing number of researchers working with functional data.

About the Authors

James O. Ramsay is a professor emeritus of psychology and an associate member in the Department of Mathematics and Statistics at McGill University in Montréal. He received a PhD from Princeton University in 1966 in quantitative psychology. His research interests include pyschometrics, dynamic systems and functional data analysis, of which he was a founder. He has been president of the Psychometric Society and the Statistical Society of Canada, and is a fellow of the Canadian Psychological Association. He was awarded the SSC Gold Medal for research in 1998, and honorary membership by the Statistical Society of Canada in 2012.

Michael Hermanussen is a German pediatrician and professor at Christian-Albrechts-Universität zu Kiel. He studied medicine and worked as a pediatrician at the University of Kiel from 1982 until 1989. He investigated growth and child development (auxology) and first described mini growth spurts. Since 1990 he has cooperated in international projects, and he organizes national and international meetings on growth and nutrition. From 2003 to 2011 he was a member of the scientific board of the Deutsche Gesellschaft für Anthropologie and chief editor of *Anthropologischer Anzeiger*. He is the founder and head of the Deutsche Gesellschaft für Auxologie.

Bibliography

Falkner, F. (1960). *Child Development: An International Method of Study.* Karger, Basel.

Ferraty, F. and Romain, R. (2011). *The Oxford Handbook of Functional Data Analysis.* Oxford University Press, Oxford.

Hermanussen, M., Thiel, C., von Büren, E., Rol de Lama, M., Pérez Romero, A., Ariznaverreta Ruiz, C., Burmeister, J., and Tresguerres, J. A. F. (1998). Micro and macro perspectives in auxology: Findings and considerations upon the variability of short term and individual growth and the stability of population derived parameters. *Annals of Human Biology*, 25:359–385.

Kneip, A. and Ramsay, J. O. (2008). Combining registration and fitting for functional linear models. *Journal of the American Statistical Association*, 103:1155–1165.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B*, 60:365–375.

Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society, Series B*, 53:539–572.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Series B*, 69:741–796.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* Springer, New York.

Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B*, 60:351–363.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis,* Second Edition. Springer, New York.

Roche, A. (1992). *Growth, Maturation and Body Composition: The Fels Longitudinal Study 1929–1991.* Cambridge University Press, Cambridge.

Thalange, N. K., Foster, P. J., Gill, M. S., Price, D. A., and Clayton, P. E. (1996). Model of normal prepubertal growth. *Archives of Disease in Childhood*, 75:427–431.

Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development*, 1:183–364.