6
Zooming in on Human Growth

6.1 Introduction

The careful documentation of human growth is essential in order to define what we call normal growth, so that we can detect as early as possible when something is going wrong with the growth process. Auxologists, the scientists that specialize in the study of growth, also need high quality data to advance our understanding of how the body regulates its own growth. It may come as a surprise to learn that human growth at the macro level that we see in our children is not that well understood.

Growth data are exceedingly expensive to collect since children must be brought into the laboratory at preassigned ages over about a 20-year span. Meeting this observational regime requires great dedication and persistence by the parents, and the dropout rate is understandably high, even taking for granted the long-term commitment of maintaining a growth laboratory. The Fels Institute in Ohio, for example, has been collecting growth data since 1929, and is now measuring the third generation for some of its original cases.

The accurate measurement of height is also difficult, and requires considerable training. Height diminishes throughout the day as the spine compresses, but it also depends on other factors. Infants must be measured lying down, and when the transition is made to measuring their standing height, measurements shrink by around one centimeter. The most careful procedures still exhibit standard deviations over repeated measurements of about three millimeters.

Records of a child's height over 20 years display features, described below, that are difficult for a data analyst to model. The classic approach has been to use mathematical functions depending on a limited number of unknown constants, and auxologists have shown much ingenuity in developing these parametric models to capture these features. The best models have eight or more parameters, and are still viewed as possibly missing some aspects of actual growth.

Nonparametric modeling techniques developed over the last three decades, such as kernel and spline smoothing methods, have been applied to growth data. These methods have been successful at detecting new features missed by parametric models, but they are not guaranteed to produce smoothing curves that are monotonic, or strictly increasing. Even a small failure of monotonicity in a height curve can have serious consequences for the corresponding growth velocity, and even more so for acceleration curves, which are especially important in identifying processes regulating growth.

In this chapter we look at some new developments in growth data analysis. A recently developed method for monotonic smoothing is applied to some old and new data. This method is used for all the curves estimated below, and is described in Section 6.8.3. Another aspect of the analysis is the introduction of *curve registration* methods, which allow the separation of amplitude and phase variation.

6.2 Height measurements at three scales

Figure 6.1 shows, for each of 10 girls, the height function $H(t)$ as estimated from 31 observations taken between 1 and 18 years. These data were collected as part of the Berkeley Growth Study; for more details of these data, and the other data analyzed in detail in this chapter, see Section 6.8.2. Growth is the most rapid in the earliest years, but we note the increase in slope during the pubertal growth spurt (PGS) that occurs at ages ranging from about 9 to 15 years. One girl is tall for all ages, but some girls can be tall during childhood, but end up with a comparatively small adult stature. The intervals between height measurements are six months or more, and the picture from this long-term perspective is of a relatively smooth growth process.

Figure 6.2 zooms in on growth by using measurements of a boy's height at 83 days over one school year, with gaps corresponding to the school vacations. The measurement noise in the data, of standard deviation about 3 mm, is apparent. The trend is also noticeably more bumpy, with height increasing more rapidly over some weeks than others.

To zoom in further, more accurate measurements are essential. The length of the tibia of a baby measured to within about 0.1 mm is graphed

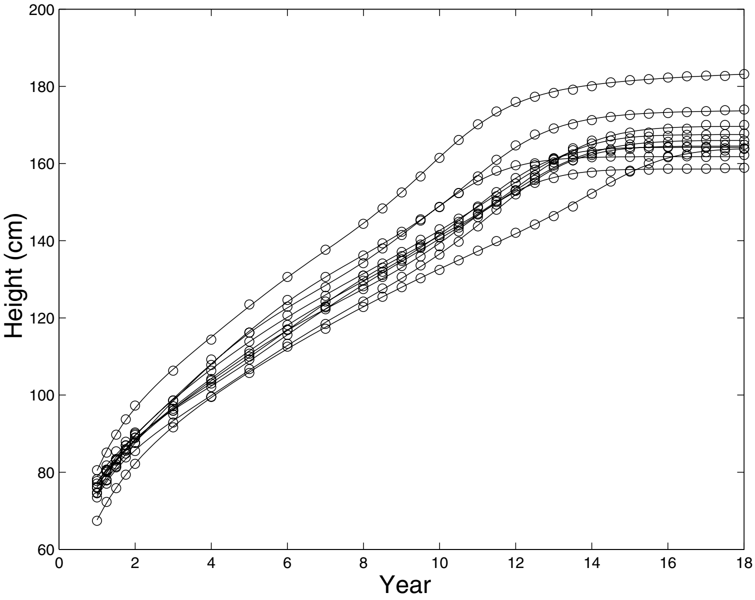

Figure 6.1. The heights of the first 10 females in the Berkeley Growth Study. Circles indicate the ages at which measurements were taken.

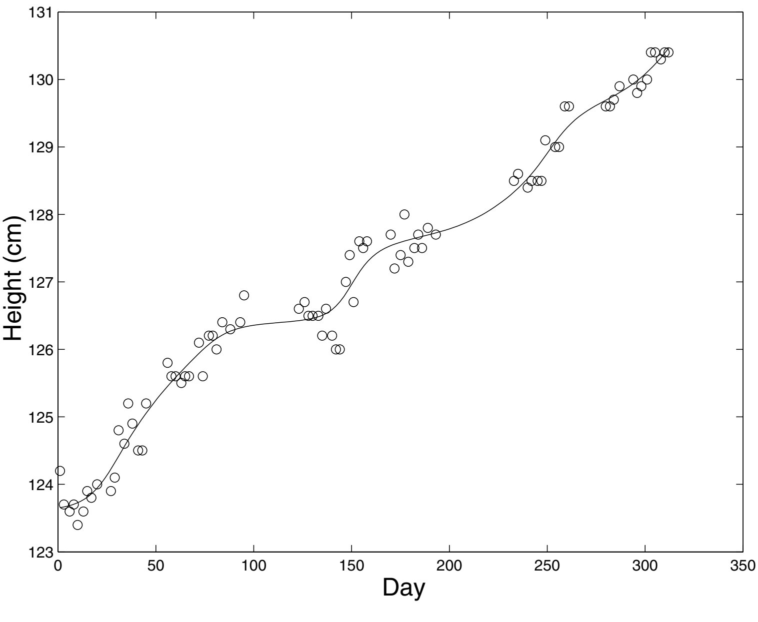

Figure 6.2. The circles are 83 measurements of height of a 10-year-old boy, and the solid curve is a smooth monotone fit to the data, as described in Section 6.8.3.

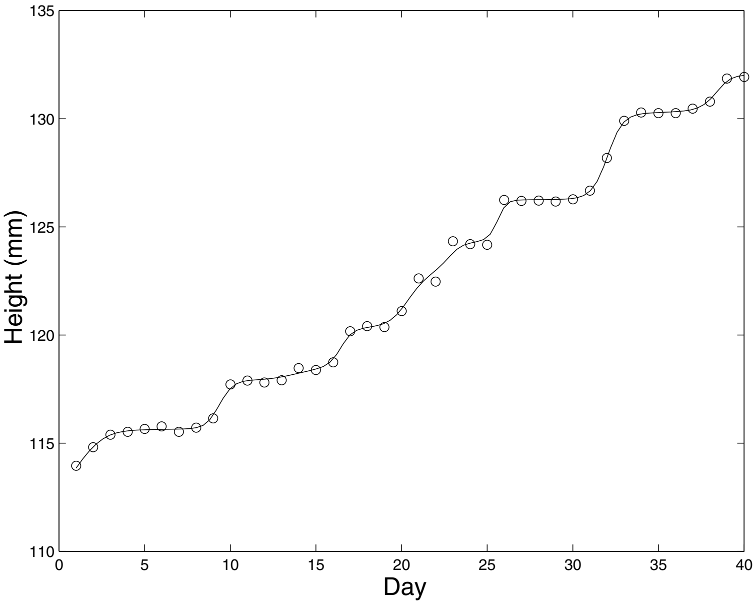

Figure 6.3. The dots indicate lengths of the tibia in the lower leg of a newborn infant, and the solid curve is a smooth monotone estimate of height.

in Figure 6.3. The jumps, or saltations, that we saw in the boy's growth are now much more visible. These data demand that we find a way to estimate just how much bone length changes over, say, a 24-hour period. Since bone length can only increase, it is essential that any smooth line, such as that in the figure, also be everywhere increasing, and this is one of the features of the smoothing method we use.

6.3 Velocity and acceleration

Although we commonly refer to data and curves such as shown in these figures as "growth curves," the term growth really means change. Hence, it is the velocity function $V(t)$, the instantaneous rate of change in height at time t, that is the real growth curve, and we should use the term "growth" to mean $V(t)$. Because height does not decrease (at least during the growing years), velocity or growth is necessarily positive. The height data only indirectly reflect growth, because they are measures of the *consequences* of growth.

If height observations are taken at time points t_i, we might consider estimating velocity by the difference ratio,

$$V(t_i) = [H(t_{i+1}) - H(t_i)]/(t_{i+1} - t_i),$$

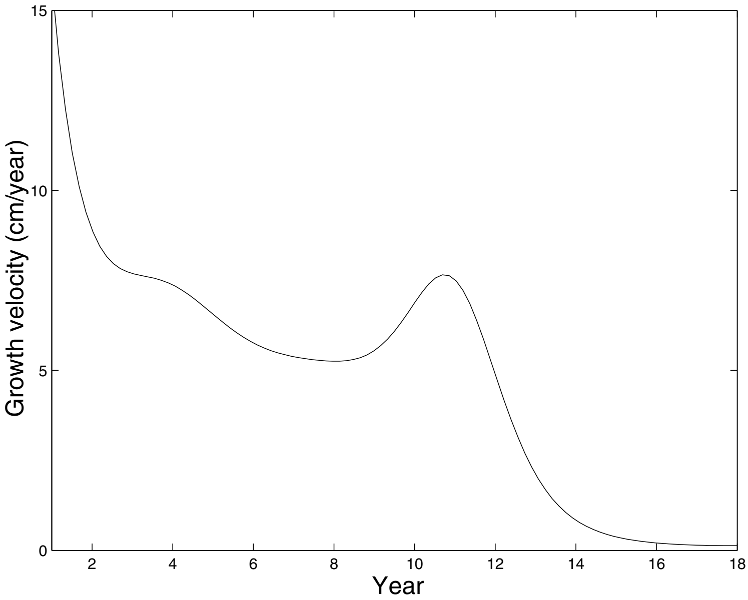

Figure 6.4. The estimated growth velocity, or rate of growth, of the first girl whose data are in Figure 6.1.

but this a bad idea from a statistical perspective, since even a small amount of noise in the height measurements will have a huge effect on the ratio, and this problem only gets worse as the time points get closer together. It is much better to fit the height data with an appropriate smooth curve, and then estimate velocity by finding the slope of this smooth curve.

Figures 6.4 through 6.6 display estimated velocity curves for the long-, medium-, and short-term growth examples considered above. Now we can see much more clearly what is happening. The pubertal spurt in Figure 6.4 is certainly more obvious, but even more impressive are the velocity surges for the 10-year-old boy. The peaks in velocity for the baby, exceeding two millimeters per day, are simply astonishing. We now know that we need to work hard to get good methods for estimating velocity, which at least during infancy is revealed to be a very intricate process.

We can get more understanding of the growth process by studying the rate of change in velocity; this is the *acceleration* in height, denoted by $A(t)$. Estimated acceleration curves for the 10 girls in the Berkeley data are given in Figure 6.7. Now we can see even more clearly what happens in the pubertal growth spurt. Naturally there is a big positive surge in velocity at the beginning of the PGS, followed by a return to zero when the velocity is no longer increasing, and finally a negative change in velocity in the final phase. It can be seen that the timing of the pubertal growth spurt varies a great deal from one girl to another, a feature we return to in

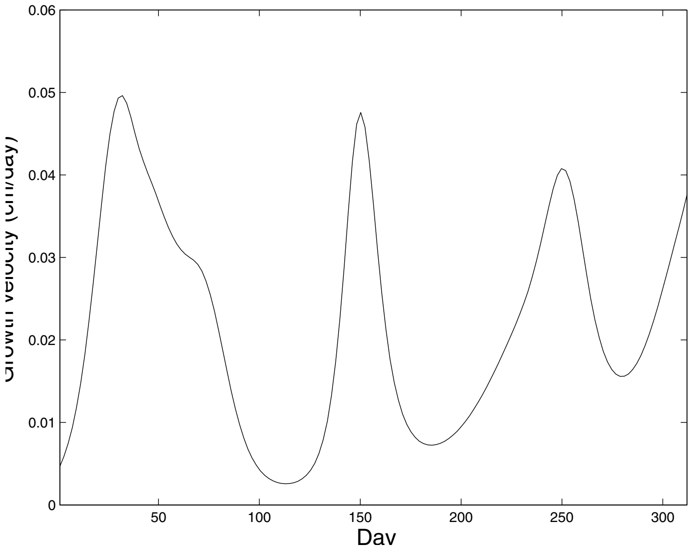

Figure 6.5. The estimated growth velocity of the boy whose data are in Figure 6.2.

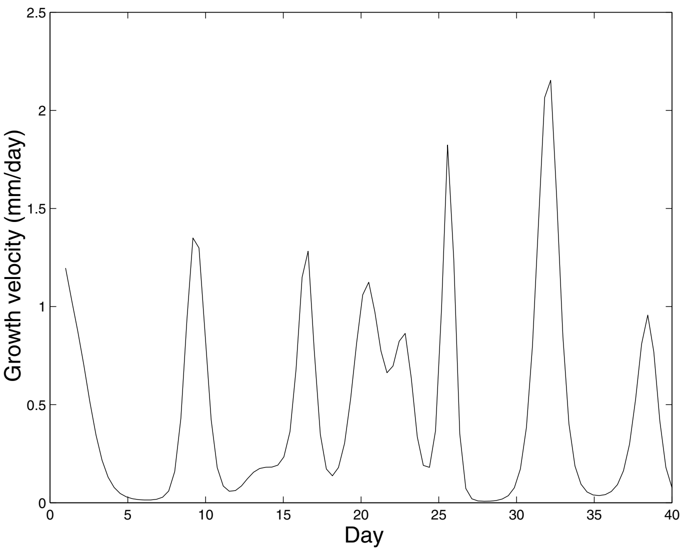

Figure 6.6. The estimated growth velocity of the baby whose data are in Figure 6.3.

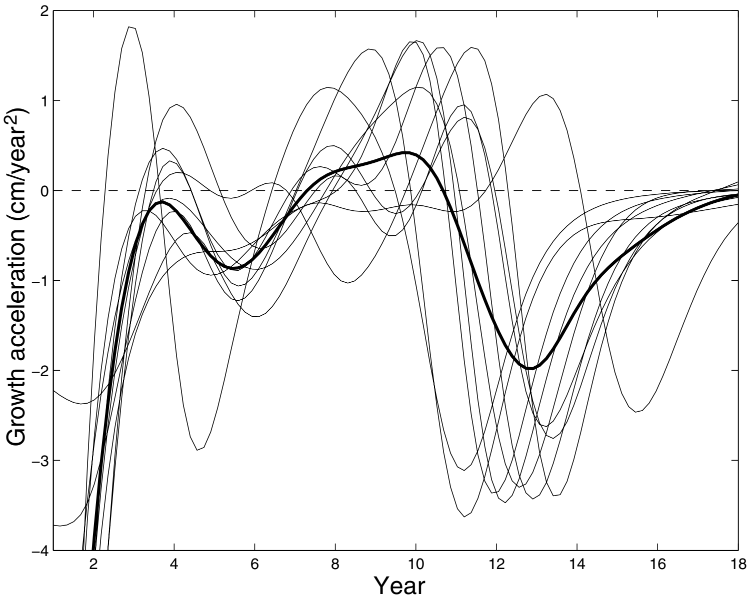

Figure 6.7. The estimated growth acceleration curves for the 10 girls whose data are shown in Figure 6.1. The heavy solid line is the average of these 10 curves.

Section 6.5. But what can also be seen, for several girls, are one or more smaller oscillations in acceleration before the pubertal growth spurt. The capacity to detect these so-called *midspurts* was one of the important early achievements of nonparametric curve estimation technology in this area.

6.4 An equation for growth

What causes the velocity $V(t_i)$ at age t_i to change to $V(t_{i+1})$ for the next observation time t_{i+1}? The question can be formulated by the following equation,

$$V(t_{i+1}) - V(t_i) = w_i V(t_i)(t_{i+1} - t_i). \tag{6.1}$$

This equation is not a model for growth, but merely a way of looking at it. It relates the velocity change over the interval $t_{i+1} - t_i$ to three factors.

- $t_{i+1} - t_i$ itself. The smaller this time interval, the less change there will be, and in the limit $\Delta t \to 0$, velocity will not change. This says that over very small time scales growth is essentially a smooth process, an assertion that seems beyond question since a jump in the rate of growth over an arbitrarily small time interval would seem inconceivable in terms of the body's physiology.

- $V(t_i)$, a term that measures growth changes on a percentage or relative basis. This is particularly useful in allowing for variations in height over the population, and, for instance, allows for comparison of growth patterns independently of people's ultimate adult height.

- w_i, a factor that determines the change in velocity. We make this factor depend on t_i because we imagine that this factor itself will change with time. This is the factor that really specifies how growth varies.

Asking a question in the right way is everything in science, and the formulation in (6.1) focuses our attention on the size of the factor w_i, which will be positive if velocity is increasing at age t_i, zero if there is no change, and negative if velocity is decreasing.

Here is a rearrangement of equation (6.1):

$$\frac{V(t_{i+1}) - V(t_i)}{t_{i+1} - t_i} = w_i V(t_i). \tag{6.2}$$

The left side of this equation is just an estimate of the instantaneous rate of change of $V(t)$, and becomes the acceleration $A(t)$ when $t_{i+1} - t_i \to 0$. Therefore, rather than defining w_i to satisfy (6.1) and (6.2) exactly, we replace it by a function $w(t)$ defined by

$$A(t) = w(t)V(t) \quad \text{or} \quad w(t) = \frac{A(t)}{V(t)}. \tag{6.3}$$

The continuously defined function $w(t)$ is now the ratio of acceleration to velocity, or what we can call *relative acceleration*, meaning acceleration of height measured as a fraction of velocity. We can rewrite (6.3) as the differential equation

$$\frac{d^2H}{dt^2} = w(t)\frac{dH}{dt}. \tag{6.4}$$

The general solution to this equation is

$$H(t) = C_0 + C_1 \int_0^t [\exp \int_0^u w(v)\, dv]\, du. \tag{6.5}$$

In this expression, C_0 and C_1 are arbitrary constants that will need to be estimated from data.

Equation (6.4) may be described as the fundamental equation of growth, in the sense that any intrinsically smooth growth process may be expressed in this way. The relative acceleration $w(t)$ is the *functional parameter of growth*. Our approach to thinking about growth is to model this function, rather than the height function itself. Once we have a way of estimating $w(t)$, we can check it against the data by using equation (6.5). To see how we estimate $w(t)$, go to Section 6.8.3.

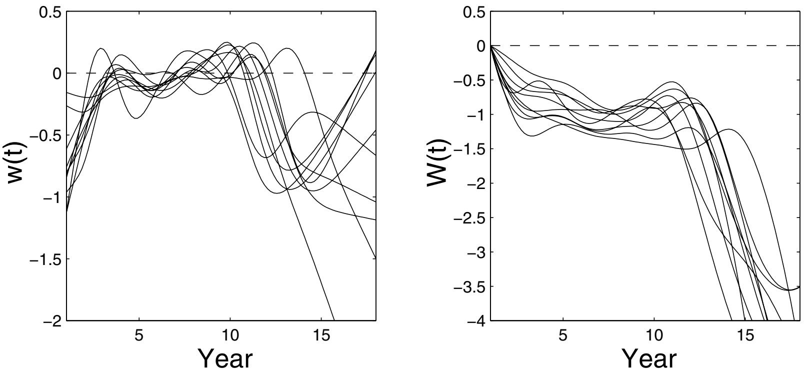

Figure 6.8. The left panel shows the relative acceleration function $w(t)$, and the right panel its integral $W(t)$.

For the 10 girls we have been studying, Figure 6.8 displays the functions $w(t)$, as well as their integrals

$$W(t) = \int_0^t w(u)du \ . \tag{6.6}$$

It can be shown from (6.5) that $W(t)$ is proportional to $\log H'(t)$, the logarithm of the instantaneous growth rate. We see that $w(t)$ looks rather like the acceleration curves in Figure 6.7 except at the end in late adolescence. This is a consequence of $w(t)$ being relative acceleration, as expressed in equation (6.3).

6.5 Timing or phase variation in growth

As in any data analysis, important aims for the long-term growth data are to estimate the average features of growth, and to get an impression of their variability across individuals. However, Figure 6.7 shows that these tasks, which are straightforward for univariate and multivariate data, present a new challenge. The heavy line, which is the mean of the 10 acceleration curves, does not have the characteristics of any of the observed curves. The PGS peak and valley for the average are much too small, but on the other hand the duration of the PGS for the average curve is longer than that of any single observed curve.

The problem is that the growth curves exhibit two types of variability. *Amplitude* variability pertains to the sizes of particular features such as the velocity peak in the pubertal growth spurt, ignoring their timings. *Phase* variability is variation in the timings of the features without considering their sizes. Before we can get a useful measure of a typical growth curve, we must separate these two types of variation, so that features such as the

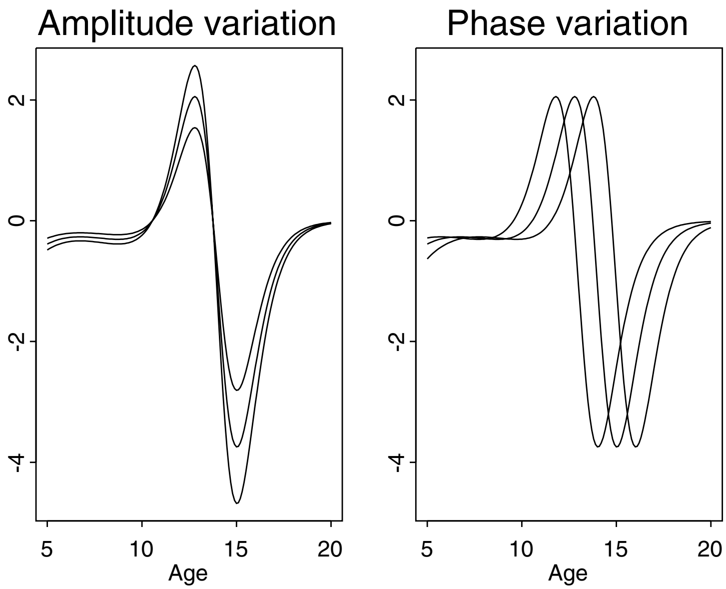

Figure 6.9. The left panel shows three height acceleration curves varying only in amplitude. The right panel shows three curves varying only in phase.

pubertal spurt occur at roughly the same "times" for all girls. The problem is expressed in schematic terms in Figure 6.9, where we see in the left panel two acceleration curves that differ only in amplitude, and in the right panel two curves with the same amplitude, but differing in phase.

By "time" here we now mean something like physiological time, which need not unfold at the same rate as physical or clock time. We mean that two girls in the middle of the pubertal spurt are, effectively, at the same physiological age, whatever their respective chronological ages. What we need is some way of mapping clock time t into its physiological counterpart. That is, we want a function $h_i(t)$ for girl i such that at physiological time t this girl has a chronological age of $h_i(t)$. For example, if $h(t) > t$, we have someone who is growing late, and if t is the physiological age at which the growth spurt takes place, then this person is having the PGS at a clock age that is later. The curve $h(t)$ is often called a *time warping* function. Figure 6.10 displays these functions $h(t)$ for our 10 girls. Remember that curves above the diagonal correspond to late growth, and curves below the diagonal to early growth.

But isn't time, too, a positive growth process? It always increases because days and years accumulate, and its velocity is defined by the time units that we use. Or at least, that is so for clock time, which increases linearly, corresponding to relative acceleration $w(t) = 0$. But physiological time, which is driven by factors such as hormonal secretions that are not constant

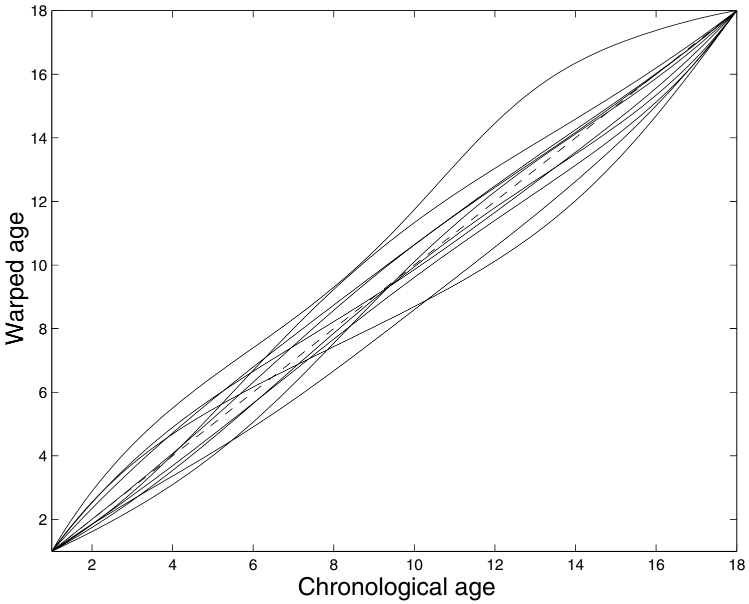

Figure 6.10. The time warping functions $h(t)$ for the 10 Berkeley girls. Curves above the diagonals indicate girls with a physiological age consistently earlier than chronological age, and therefore growing late.

across individuals, need not unfold in this elementary way. Even in Figure 6.1 we can see clearly that some girls are outstripping clock time, and maturing early, while other girls are lagging behind the clock, being late maturers.

Therefore, the warping function $h(t)$, which must be always increasing, reflects simply another type of growth curve, and may be characterized by the same mathematical representation that we have in equation (6.4), and therefore corresponds to its own relative acceleration $w_h(t)$. We defer further details on how we estimate $h(t)$ to Chapter 7, where registration is the main topic, and pass to what we see when the growth curves have been registered.

6.6 Amplitude and phase variation in growth

What do we do with warping functions $h_i(t)$ once we have estimated them? Recalling that for a late grower, $h(t) > t$, we see that we can think of $h(t)$ in such a case as "speeding up" clock time to make it match physiological time. This means that if we calculate the function

$$V^*(t) = V[h(t)] \tag{6.7}$$

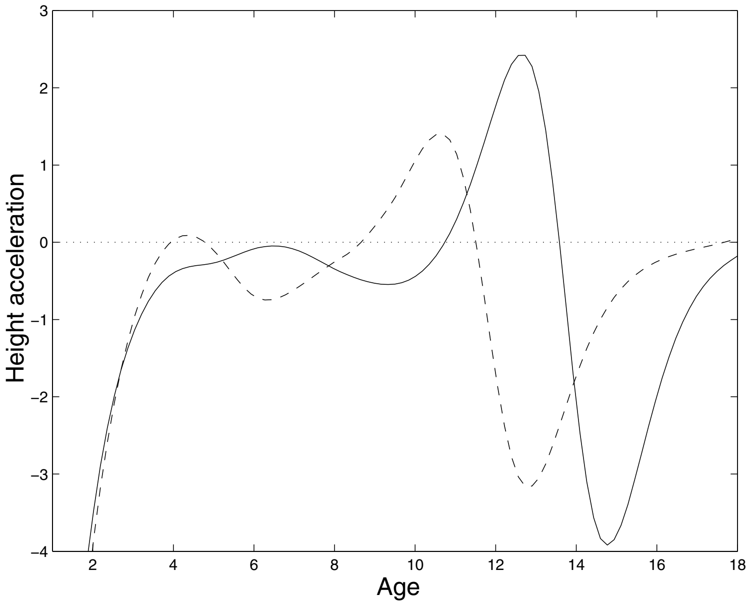

Figure 6.11. The solid curve is the average acceleration for the registered data from the boys, and the dashed curve is the registered acceleration average for the girls.

we now have a new velocity function $V^*(t)$ that shows the pubertal growth spurt, for example, as occurring at the "right time." Similarly, for $h(t) < t$, we can use the warping function to slow down clock time for an early grower. We also define the registered height and acceleration curves $H^*(t) = H[h(t)]$ and $A^*(t) = A[h(t)]$, respectively.

With these registered curves in hand, we can now carry out averaging and other analyses more meaningfully, since registered curves no longer have the phase variation that affected the average in Figure 6.7. Figure 6.11 superimposes the mean registered acceleration curves of girls and boys. Some new features now emerge. We see that the pubertal spurt is not the only spurt visible in long-term growth data, and we already know that there are even more spurts within medium- and short-term data.

We see in Figure 6.11 that girls and boys seem to go through the same pubertal growth cycles, but differ in two ways: the PGS is earlier in girls, but more intense in boys. The time shift prompts us to warp time for one gender in order to render its growth equivalent to the other. The left panel of Figure 6.12 displays the warping function $h(t)$ that registers the boys' data to the girls', and the right panel shows the registered average acceleration curves. We can see two major gender differences. The left panel demonstrates that male growth essentially lags behind female growth, with a gap that increases steadily until growth is finally finished. The right panel

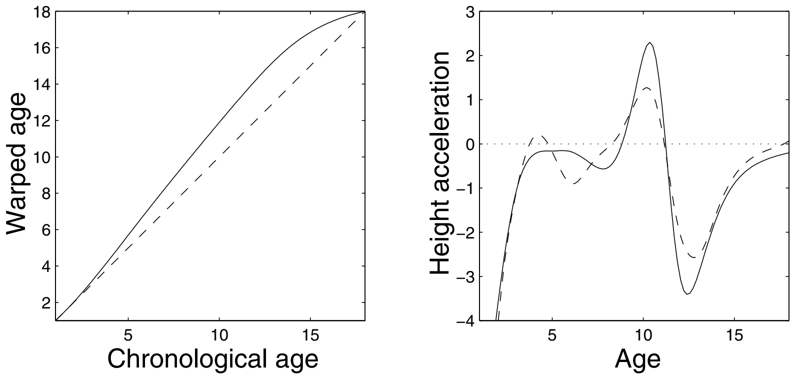

Figure 6.12. The left panel displays the warping function for registering the boys' average velocity to that of the girls. Because boys mature more slowly, the warping function is above the diagonal, shown as a dashed line. The right panel shows the registered average acceleration curves. The solid curve corresponds to the boys and the dashed curve to the girls.

shows that the intensity of the acceleration function during the pubertal spurt is greater for boys than for girls. These are the two main contributors to the gender difference in mean adult heights: boys grow over a longer period, and grow more intensely during the pubertal growth spurt.

The right panel of Figure 6.12 also shows some gender difference in earlier childhood. Closer examination of these data, and also of other larger data sets on growth such as the Fels Institute data, reveals that many children have more than one midspurt. Furthermore, both the number and the registered position of these midspurts is more variable in boys than in girls. This is partly because boys have a longer prepubertal period. It is the averaging out of this greater intergender variability that causes boys to have a flatter average registered acceleration curve.

What of the amplitude variation among the girls? A functional principal components analysis of the registered acceleration reveals that three principal components or harmonics account for 72% of their variation about the mean acceleration curve. After varimax rotation of these components, we get the three components displayed in Figure 6.13, and they account for nearly equal proportions of variance. Varimax harmonic 1 has to do only with variation during the pubertal spurt, and therefore captures the intensity of this event. The second and third harmonics, on the other hand, reflect variation only in the prepubertal years, but rather differently. The second harmonic shows an intensification of the two prepubertal spurts relative to the mean curve, but the third is more complex, capturing phase variation in these two earlier spurts that was not taken out by the registration process.

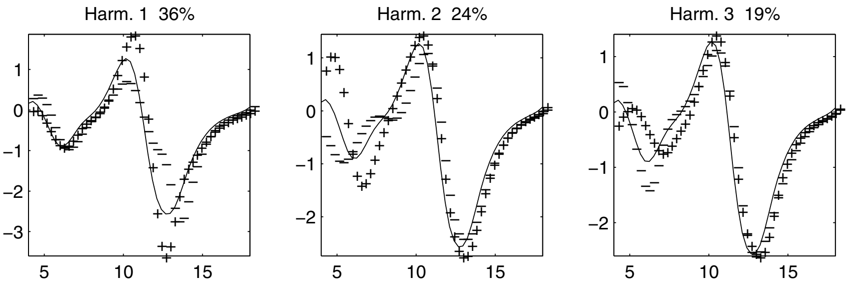

Figure 6.13. The varimax-rotated harmonics of registered acceleration for the girls. The amount of variation accounted for is indicated at the top of each harmonic. The solid curve is the mean acceleration, and the plus and minus symbols show the effects of adding and subtracting a multiple of the harmonic to the mean.

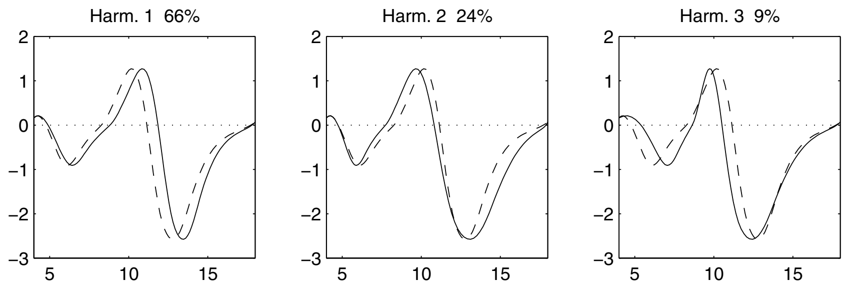

Figure 6.14. Results of a PCA of the warping functions, regarded as functional data in their own right. The dashed curves show the mean acceleration curve without time transformation, and the solid curves show what the mean acceleration curve would look like under the influence of each harmonic. The underlying data are the growth data for the 10 females in the Berkeley growth study.

We can also study phase variation by carrying out a PCA of the warping functions in Figure 6.10. The harmonics are displayed in Figure 6.14 by showing what the mean acceleration would look like if a multiple of the harmonic were added to clock time. In this case, the first three components explain 99% variation. The first harmonic corresponds to growth that is consistently late. The second shows early growth up to the deceleration phase of the PGS, and then slow recovery. The third indicates late prepubertal growth and early onset of puberty.

6.7 What we have seen?

Growth is not at all smooth over a short time scale. Our results hint that growth takes place by turning on and off the velocity function periodically.

In an infant the period is three or four days, but later the period seems to lengthen, until by 10 years it is of the order of a number of weeks. The discovery of these jumps or saltations is new, and we need much more data of the quality that we have for the baby before we can understand this process better. But perhaps what counts for growth is what turns it off; growth at the rate displayed in Figure 6.6 could actually be dangerous if sustained for much longer than a day or so.

On the methodological side, a formulation of the growth process in terms of the difference equation (6.1) or the differential equation (6.4) leads to a smoothing technology for growth data that respects the monotonicity of the height function $H(t)$ and the positivity of velocity $V(t)$, and also yields in the form of relative acceleration $w(t)$ a curve with a natural interpretation. An added bonus was the appreciation that the time warping function $h(t)$ that takes clock time into physiological time is also a growth process, and this story is taken up further in Chapter 7. The time warping functions for each individual can themselves be considered as functional data.

Finally, once we have teased apart, at least to some extent, amplitude and phase variation, we see that boys and girls do not differ strikingly in the shapes of their acceleration amplitudes, but that they do show a large amount of phase variation. Among the girls (and boys as well), amplitude variation seems to be primarily three-dimensional, and separable into components that reflect variation in the pubertal growth spurt, and others that show variation in prepubertal growth.

6.8 Notes and further issues

6.8.1 Bibliography

The work of this chapter is discussed in more detail in Ramsay and Bock (2002). They provide extensions and more details of the analyses presented here, apply the methods to the larger Fels Institute data set, and give further discussion and bibliographic references. The formulation of the growth process as a second-order linear differential equation, and the analysis of the growth data for the 10-year-old boy, are given in Ramsay (1998). The companion paper, Ramsay and Li (1998), applies this formulation to the registration problem, which is examined in more detail in Chapter 7.

There is already a large literature containing functional data analyses of growth data. Indeed, this field has provided one of the most important test beds for the development of curve estimation and analysis. The many contributions of T. Gasser and his collaborators, of which Gasser et al. (1990) is only one example, are especially important. A good deal of the research in this fascinating field appears in *Annals of Human Biology.*

Growth curve analysis has also inspired many contributions to the curve registration problem, and statistical issues in the use of features or landmarks to register growth curves has been studied by Kneip and Gasser (1992) and Gasser and Kneip (1995).

6.8.2 The growth data

The Berkeley Growth Study (Tuddenham and Snyder, 1954) recorded the heights of 54 girls and 39 boys between the ages of 1 and 18 years. Although larger studies of growth have since been completed, notably the Fels (Roche, 1992) and Zurich (Falkner, 1960) data, the Berkeley data have been published and are therefore freely available. Heights were measured at 31 ages for each child, and the standard error of these measurements was about 3 mm, tending to be larger in early childhood and lower in later years.

The data on the growth of the 10-year-old boy were collected as part of a study reported in Thalange et al. (1996), and generously made available to us by P. J. Foster at the University of Manchester. The short-term data on the growth of the tibia in a newborn infant are described in Hermanussen et al. (1998), and we thank Prof. Hermanussen for supplying them. This paper is one in a series of papers that provide details on the experimental procedure, and which report similar results in the growth of rats.

6.8.3 Estimating a smooth monotone curve to fit data

In this section, the monotone smoothing method is described briefly; for more details, see Ramsay (1998). Relevant software is available from the Web site corresponding to this chapter. We use the differential equation for growth $A(t) = w(t)V(t)$ to transform the problem of estimating the height function $H(t)$ that actually fits the height observations y_j observed at ages $t_j, j = 1, \ldots, n$ to one of estimating the relative acceleration function $w(t)$. Our task is made simpler by the fact that $w(t)$ is unconstrained in any way, unlike $V(t)$ which must be positive, or $H(t)$ which must always increase.

Our approach to estimating $w(t)$ is to express it as a linear combination of basis functions $\phi_k(t)$, as we already have done in previous chapters, so that

$$w(t) = \sum_{k=1}^{K} c_k \phi_k(t). \tag{6.8}$$

We can then fit the data by numerically minimizing the error sum of squares

$$\text{SSE} = \sum_{j=1}^{n} [y_j - H(t_j)]^2 \tag{6.9}$$

with respect to the coefficients c_k defining the basis function expansion (6.8).

Our choice of basis is the B-spline basis $\phi_k(t) = B_k(t)$ described briefly in Section 2.5 and in detail in Ramsay and Silverman (1997, Chapters 3 and 4). We tend to choose this basis for any function that is not periodic and that has no other restrictions on its shape. A B-spline basis is defined by a set of knots, and our strategy is to place a knot at each age t_j at which height is observed.

Putting knots at every data point allows considerable flexibility, but results in more basis functions than there are observations. We compensate for this overly rich basis by adding a roughness penalty to the error sum of squares criterion (6.9) and then minimizing the following penalized least squares criterion

$$\text{PENSSE} = \sum_{j=1}^{n} [y_j - H(t_j)]^2 + \lambda \int_0^T [w''(t)]^2 dt. \tag{6.10}$$

In this expression, T is the largest age at which we wish to estimate $H(t), V(t)$, and $A(t)$. Roughness in this expression is defined as the integral of the square of the second derivative $w''(t)$ of $w(t)$. Because of the nonlinear dependence of $H(t)$ on $w(t)$, the minimization of PENSSE will involve a numerical optimization over the vector of B-spline coefficients c_k.

The effect of varying the smoothing parameter λ in (6.10) is as follows. The closer λ is to zero, the less the roughness of $w(t)$ is penalized, and in the limit $H(t)$ will become a monotone curve that comes as close as any monotone curve can come to fitting the data, which, of course, may not be themselves strictly increasing. Such a curve is bound to have plateaus and points of very rapid increase, and would be unacceptable even for data as rough as those in Figure 6.3. At the other extreme, if λ were to increase without limit, $w(t)$ would approach a straight line, and $H(t)$ would become much too smooth to fit the data acceptably. In particular, $A(t)$ would become linear, and would not offer a plausible account of events such as the PGS. In the present context, we have found it satisfactory to choose subjectively the smallest value of λ that still provided a smooth and interpretable estimate of $A(t)$.