

Automatic detection and quantification of growth spurts

FRÉDÉRIC DANDURAND

CNRS and Aix-Marseille University, Marseille, France

AND

THOMAS R. SHULTZ

McGill University, Montreal, Quebec, Canada

Growth phenomena are often nonlinear and may contain spurts, characterized by a local increase in the rate of growth. Because measurement error and noise may produce apparent spurts, it is important to identify systematic and reliable spurts. We describe a system, *automatic maxima detection* (AMD), for statistically identifying significant spurts and computing (1) *point of maximal velocity*, when the spurt was most intense; (2) *start*, when the spurt started; (3) *amplitude*, the intensity of the spurt; and (4) *duration*, the length of the spurt. We also introduce a software implementation of AMD in MATLAB. In growth of height data, AMD showed a reliable pubertal growth spurt for most children and a reliable prepubertal spurt for some children. In simulated growth of vocabulary, AMD showed a large global spurt and several minispurts. In real vocabulary growth, AMD showed a few spurts. Advantages of AMD include improvements in objectivity, automaticity, quantification, and comprehensiveness.

Growth is fundamental to complex systems as varied as biological organisms, national economies, and social networks. Models and theories of development often attempt to accurately describe and predict growth and to explain its underlying processes. A common example is physical growth in humans (Eveleth & Tanner, 1990; Jolicœur, Pontier, & Abidi, 1992). Figure 1 illustrates an example of growth in physical height of a child in the Berkeley study (Tuddenham & Snyder, 1954). The figure presents height measurements from 1 to 18 years of age and a spline-based continuous function fitted to the data with the first three derivatives (velocity, acceleration, and jerk). This example illustrates how growth can be bumpy and nonlinear, involving complex dynamics that can be seen on the graphs of the derivatives.

Studying development often involves detecting departures from baseline growth. Such departures—that is, growth spurts—may indicate important changes in the underlying growth processes. If we want to test competing theories of development, we need a good assessment of what the data are really like. When the underlying growth process is unknown, one of the first questions one may ask is the following: Are data increasing linearly or not? This might seem like a simple question, but because developmental data are usually noisy and measurement techniques have a limited precision, it is often difficult to determine whether observed deviations from baseline growth are genuine or, instead, caused by random variations and measurement artifacts.

Visual inspection of the growth curve itself is often insufficient because it can be difficult to determine whether an apparent spurt is a significant departure from baseline growth or present at all. The height curve in Figure 1 provides an example of an apparent spurt at adolescence, but no other spurt is apparent by visual examination alone. The method we introduce in this article is able to detect two statistically valid growth spurts for this boy, one at just over 7 years of age and another one later at adolescence, around 13 years of age.

When a model of the underlying growth process exists (e.g., modeled using asymptotic, logistic, or exponential functions), parametric approaches are well suited to detecting departures from expected growth—that is, spurts. Here, however, we address the more difficult problem of detecting spurts in data for which little, if anything, is known about the process that generates them. We thus take a nonparametric approach that consists of detecting significant departures from linearity. We define spurts as local or temporary departures from some linear, baseline rate of growth in continuous data. Such local spurts begin with a significant increase in rate of growth and end with a significant decrease. To exclude spurious changes due to noise, we count as significant only increases and decreases that cannot be accounted for by the variance or uncertainty of the data (more on this in the Method section). Decreases following a spurt mark a slowing down of growth but do not require growth to become negative. Identifying growth spurts is part of a class of statistical

F. Dandurand, frederic.dandurand@gmail.com



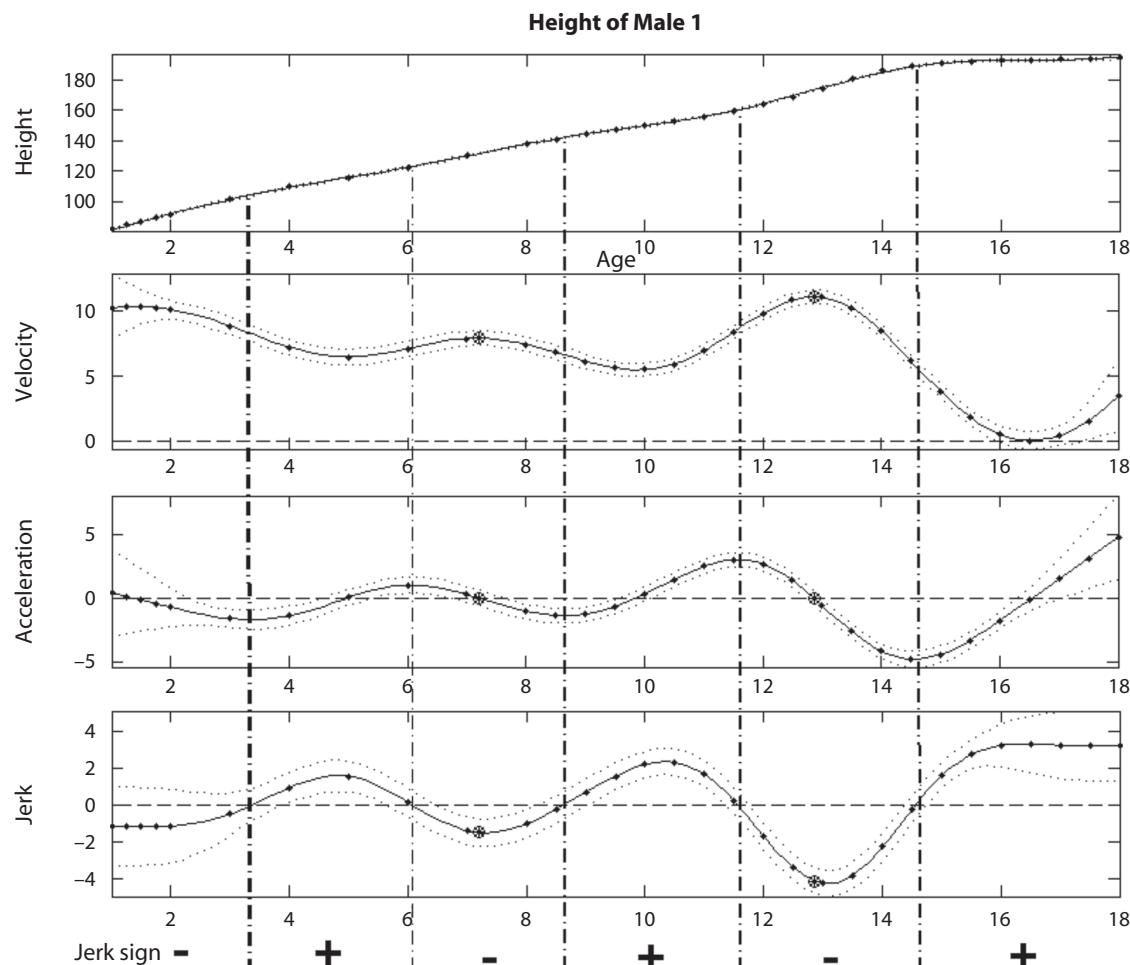


Figure 1. Height of a child in the Berkeley study (Tuddenham & Snyder, 1954) aligned with its first three derivatives (velocity, acceleration, and jerk), estimated using functional data analysis. Dotted lines around curves represent 95% confidence bands. Automatic maxima detection (AMD) uses signs of the jerk function (third derivative) as important markers. In all AMD-generated plots presented in this article, significant spurts are marked as large circles at the location in time at which they occur (i.e., point of maximal velocity).

problems called *change point analysis* (e.g., Carlstein, Müller, & Siegmund, 1994).

In this article, we introduce a technique called *automatic maxima detection* (AMD) that automates this identification process. Our work has two goals: first, automating the spurt detection process, and second, precisely quantifying spurts—when they started, when they were most intense, how large they were, and how long they lasted. Furthermore, our system computes the variance associated with those four measures. These spurt characteristics can be subjected to further statistical analysis. Our method, which takes as input a smoothed curve with confidence bands, is systematic, automated, and problem independent. AMD makes few assumptions about the data, because it uses local features in individual records. Thus, AMD is especially interesting for exploratory analyses when little is known about the data. AMD can also be used to support confirmatory analyses.

Some analysis techniques capitalize on global features such as trends, periodicity, and replication across individ-

uals. For example, Fourier series could be used to identify spurts that occur cyclically (Ramsay & Silverman, 2005). Also, when spurts synchronously occur across individuals, data from individual records can be pooled to reduce noise and increase detection power. However, pooling growth data often requires curve registration. In contrast to these techniques, AMD makes very few assumptions about the data.

AMD takes as input a smooth estimate of the function that generated the observed data sample, its first three derivatives (sometimes referred to as velocity, acceleration, and jerk), and confidence bands for the first derivative (velocity). Any technique that provides these features can be used, including general kernel methods (Shawe-Taylor & Cristianini, 2004), Loess (Cleveland, Grosse, & Shyu, 1992), penalized splines (Green & Silverman, 1994; O'Sullivan, 1986), restricted cubic splines (Harrell, 2001), random walks (Hughes, 1996), and so forth. Confidence bands can be numerically approximated for these techniques by using jackknife or bootstrap approaches (Efron, 1981).

Researchers may select any appropriate smoothing technique for their data. A large literature exists on smoothing methods (Eubank, 1999; Ramsay & Silverman, 2005; Simonoff, 1996). However, because current techniques have limitations, caution should be exercised when selecting an automated smoothing method (Ramsay, Hooker, & Graves, 2009; Ramsay, pers. comm., July 2007); automation may never completely replace subjective decisions (Ramsay, Bock, & Gasser, 1995, p. 420). Consequently, appropriate smoothing is often problem specific, and thus a general discussion is beyond the scope of this article. However, we present an illustration using height data of how we can systematically study the effect of the amount of smoothing on the AMD measures of spurts.

The MATLAB implementation presented here uses functional data analysis (FDA) as a spline regression technique (Ramsay & Silverman, 2005). AMD code can be downloaded from <http://Inslab.org/lib/AMD/>. The latest version of the FDA code is available at www.psych.mcgill.ca/misc/fda/. See Ramsay et al. (2009) for documentation on using FDA with MATLAB.

An important advantage of using FDA is that the computation of derivatives is particularly well integrated. FDA has proven useful for the analysis of growth curves and detection of developmental spurts. In previous research, however, identification of spurts and plateaus had to be performed manually by an expert (Ramsay et al., 1995; Shultz, 2003). AMD aims at automating this identification process.

As illustrations, we tested our system on physical data of children's height from the Berkeley growth study (Tuddenham & Snyder, 1954), on simulated data of children's vocabulary size (McMurray, 2007; Mitchell & McMurray, 2008), and on experimental data of children's vocabulary size (Corrigan, 1978). We chose the first two problems because both of them have uncontroversial growth spurts that we expected AMD to find, a well-documented spurt during adolescence for physical growth (Tanner, 1978) and a large central spurt for simulated vocabulary data. The two problems have, however, different characteristics. On the one hand, physical growth data are usually sampled sparsely and change rather smoothly. On the other hand, vocabulary data are often densely sampled and change more abruptly, resulting in a staircase-like curve. Finally, we include an analysis of real vocabulary growth to show how the quantitative rigor of AMD can help inform controversial issues of development—in this case, whether vocabulary spurts in children are genuine (Ganger & Brent, 2004). Using the physical growth data, we assess the effect of the smoothing parameter, which also serves to illustrate how spurt characteristics measured with AMD can be further analyzed using standard statistical techniques. In a discussion following the presentation of these results, we will argue that AMD can inform research on stages of development.

METHOD

Before presenting AMD per se, we first describe a preprocessing step that involves estimating smoothed functions, derivatives, and confidence bands. We describe this

process in FDA, but AMD can be used equally well with other smoothing techniques.

Preprocessing Using FDA

The analysis begins with a sample of data pairs (y_j, t_j) , where y_j is a measure of interest (e.g., height or vocabulary size) that varies with t_j (which is often the time of measurement). A first step in FDA is to estimate a smooth and continuous function $x(t)$ ¹ from these discrete observations as a weighted linear combination of B-spline basis functions, φ_k , $k = 1, \dots, K$:

$$x(t) = \sum_k^K c_k \varphi_k(t). \quad (1)$$

B-spline basis functions are piecewise polynomials of fixed degree joined end to end at locations called knots. Whereas, on the axis of the independent variable (time), knots are often positioned to coincide with observation points t_j , on the other axis, estimated values $x(t)$ are not required to—and in general will not—coincide with y_j .

FDA uses a roughness penalty approach to curve smoothing. A roughness penalty consists of limiting or penalizing the size of some higher order derivative of the smoothed function. Large values of higher derivatives imply rapid changes in the value of the function.

Coefficients c_k are selected to minimize a penalized sum of squared errors (PENSSE) between the estimated function and observed data vector \mathbf{y} :

$$\text{PENSSE}(\mathbf{y} | \mathbf{c}) = (\mathbf{y} - \Phi \mathbf{c})' \mathbf{W} (\mathbf{y} - \Phi \mathbf{c}) + \lambda \mathbf{c}' \mathbf{R} \mathbf{c}, \quad (2)$$

where \mathbf{c} is a vector of coefficients c_k , \mathbf{W} is a symmetric positive definite weight matrix, Φ is the $n \times K$ matrix of basis function values $\varphi_k(t_j)$, and λ is a smoothing parameter. The fitted curve $x(t)$ becomes increasingly smooth as λ increases. \mathbf{R} is a roughness penalty matrix:

$$\mathbf{R} = \int \left[\frac{d^2}{dt^2} \varphi(t) \right] \left[\frac{d^2}{dt^2} \varphi'(t) \right] dt. \quad (3)$$

For AMD, the roughness of the third derivative is controlled by penalizing the size of the fifth derivative. Splines of polynomial degree six are used to ensure that the third derivative of $x(t)$, needed for AMD, is smooth. See Ramsay and Silverman (2005) for further details, including a discussion of smoothing parameter selection in the context of modeling growth data.

Computation of confidence bands in FDA. In addition to a smoothed function and its first three derivatives,² AMD requires confidence bands of the first derivative of the function. In FDA, variance of the fitted function x is computed as follows:

$$\text{Var}[x] = \Phi' \text{Var}[c] \Phi, \quad (4)$$

where Φ is a matrix of basis function values at the observation points (see above); $\text{Var}[c]$ is the variance of coefficients c_k :

$$\text{Var}[c] = (\Phi' \mathbf{W} \Phi)^{-1} \Phi' \mathbf{W} \Sigma_e \mathbf{W} \Phi (\Phi' \mathbf{W} \Phi)^{-1}; \quad (5)$$

\mathbf{W} is a symmetric positive definite weight matrix (see above); and Σ_e is the variance-covariance matrix of the residual vector \mathbf{e} .

Here, using the standard model ($\Sigma_e = \sigma^2 I$), $\text{Var}[c]$ simplifies to

$$\sigma^2(\Phi'\Phi)^{-1}. \quad (6)$$

These FDA estimates of variance can be easily converted into a standard error and then into confidence bands for a given user-specified target alpha value. For instance, for alpha equal to .05 as used here (i.e., 95% confidence intervals), the width of the confidence band is 2×1.96 times the standard error. These bands are called *pointwise* because they indicate the precision with which a curve is estimated at a *fixed* point t , rather than the entire curve over the whole range of t . The details of the method used to estimate these bands in FDA, including some limitations and warnings, can be found elsewhere (Ramsay & Silverman, 2005).

Automatic Maxima Detection

In this section, we describe how the AMD system works to automatically detect candidate spurts and quantify their times of maximal velocities, starts, ends, durations, and amplitudes. Then we explain how AMD determines spurt validity—that is, how it distinguishes, among candidate spurts, those that are likely genuine from those due to random variations in the data. As a precondition to using AMD, users need a smooth function fitted to their raw data, three derivatives of this function, and confidence bands for the first derivative.

Ramsay et al. (1995) and Shultz (2003) have provided examples of growth data modeling using FDA. They noted that velocity peaks can identify spurts, whereas velocity valleys indicate plateaus in growth data. They also noted that rapid acceleration descents indicate velocity peaks and, thus, growth spurts, whereas acceleration increases represent relative growth plateaus (cf. zones of negative and positive jerk, respectively, in Figure 1). Finally, they noted that an upward bend in the growth curve, signaled by high curvature (peak acceleration) and followed by straightening (decreasing acceleration), marks high growth velocity and, thus, a spurt in growth. Our system operationalizes these ideas.

Locating and quantifying spurts: Start, point of maximal velocity, duration, and amplitude. AMD first identifies zero crossings in the first three derivatives as important landmarks in the function; a similar approach was taken in Ramsay et al. (1995). This section describes how to identify candidate spurts, which may turn out to be genuine (significant) or spurious (nonsignificant). The following section describes how to determine the significance of these candidate spurts.

Figure 2 presents a visual example of how spurts are measured.

We define the point of maximal velocity of a spurt as the point t_{spurt} where velocity is maximal and, thus, by the fundamental theorem of calculus, acceleration equals zero:³

$$\frac{d^2}{dt^2} x(t_{\text{spurt}}) = 0. \quad (7)$$

We define the beginning (start) of a spurt, t_{start} , as the point where acceleration is locally maximal and its sign is

positive. This point occurs before the point of maximal velocity is reached. Using the fundamental theorem of calculus, such a point of maximal acceleration corresponds to the jerk equaling zero:

$$\frac{d^3}{dt^3} x(t_{\text{start}}) = 0. \quad (8)$$

Similarly, we define the ending of a spurt t_{end} as the point following the spurt where the magnitude of acceleration is locally maximal but with a negative sign, which also corresponds to a zero crossing of the jerk curve:

$$\frac{d^3}{dt^3} x(t_{\text{end}}) = 0. \quad (9)$$

We define spurt duration as the length of the interval from start to end:

$$\text{duration} = t_{\text{end}} - t_{\text{start}}. \quad (10)$$

Note that over the spurt interval, because acceleration is positive at the beginning of a spurt and negative at the end, we can actually characterize a spurt as the region of negative jerk bounded by two zero crossings (sign changes), using the fundamental theorem of calculus:

$$\frac{d^3}{dt^3} x(t) < 0 \text{ for all } t \in [t_{\text{start}}, t_{\text{end}}] \quad (11)$$

Equation 11 represents the definition of a spurt in AMD. On the example given in Figure 1, spurts (i.e., regions of negative jerk) are indicated.

Finally, we define spurt amplitude as the energy in the velocity of the function—that is, the area under the jerk curve during the spurt. Because this area is negative, we need to invert the sign. Using the fundamental theorem of calculus, the energy in the spurt—that is, its amplitude—is the difference in acceleration between points t_{end} and t_{start} :

$$\begin{aligned} \text{spurt amplitude} &= -\int_{t_{\text{start}}}^{t_{\text{end}}} \frac{d^3}{dt^3} x(t) \\ &= \frac{d^2}{dt^2} x(t_{\text{start}}) - \frac{d^2}{dt^2} x(t_{\text{end}}). \end{aligned} \quad (12)$$

Since data sampling is discrete, the true location of a spurt feature relating to time (point of maximal velocity, start, and end) will virtually never coincide with a sampling point and, instead, will fall between sampling points. Because the spline-based approach of FDA generates a smooth fitted function, AMD can measure spurt features at an estimation of their true location, including between the sampling locations.⁴ In other words, once the points y_j are modeled as a continuous function $x(t)$ using FDA, calculations of derivatives rely directly on this continuous function and not on sampling locations.

To compute the variance of each of these four measures (start, point of maximal velocity, duration, and amplitude), we use a standard bootstrapping approach called *resampling residuals* (Efron & Tibshirani, 1993). We compute residuals between the estimated values in the FDA-fitted curve and the raw data. We generate new data by adding residuals that were resampled with replacement to the original fitted curve. With these new data, we estimate a

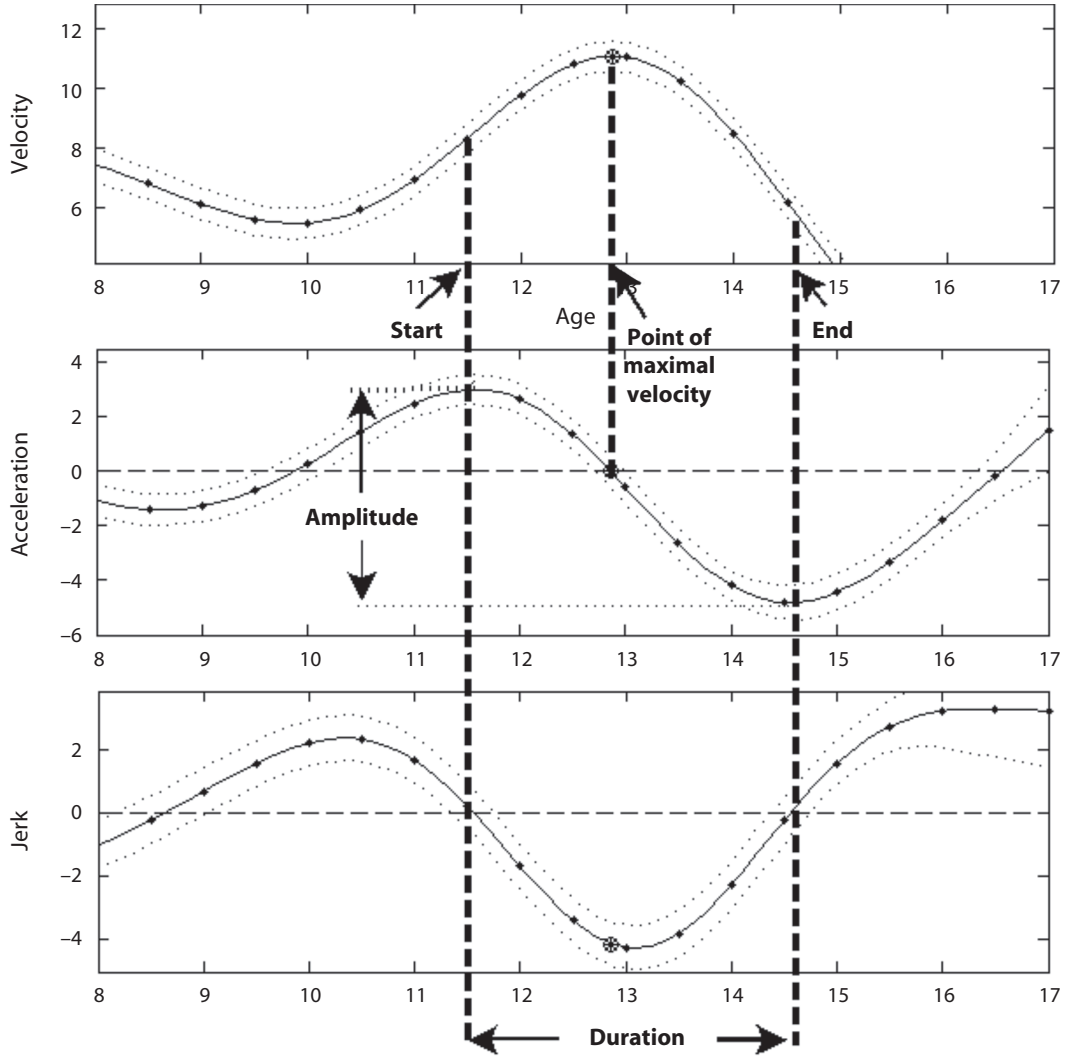


Figure 2. Visual representation of spurt start, end, point of maximal velocity, amplitude, and duration. This illustration corresponds to the adolescent spurt of the child in Figure 1.

new curve using FDA, on which we run our spurt identification system to find values of spurt measures (point of maximal velocity, start, duration, and amplitude). The process is repeated many times (e.g., 10,000) with different random sampling of the residuals distribution in order to generate a large sample of curves whose noise characteristics are similar to the originally fitted curve. Finally, we compute spurt measures on each curve in the sample. The variance of those measures in the sample is an estimate of the variance in the original curve.

Identifying genuine spurts. After having detected candidate spurts, AMD distinguishes genuine spurts from spurious spurts that can be explained away as noise. To determine whether a spurt is genuine, AMD verifies that the velocity at the point of maximal velocity of the spurt is significantly larger than the velocities of the very next local velocity minima preceding ($t_{\min 1}$) and following ($t_{\min 2}$) the spurt. As is illustrated in Figure 3, the interval defined by $[t_{\min 1}, t_{\min 2}]$ is wider than, but encompasses,

the interval of the spurt $[t_{\text{start}}, t_{\text{end}}]$. If we assume that those three points (a maximum and two adjacent minima) are independent (uncorrelated), this criterion can be implemented as having nonoverlapping confidence intervals:

$$\text{low}\left(\frac{d}{dt}x(t_{\text{spurt}})\right) > \text{high}\left(\frac{d}{dt}x(t_{\min 1})\right) \quad (13)$$

and

$$\text{low}\left(\frac{d}{dt}x(t_{\text{spurt}})\right) > \text{high}\left(\frac{d}{dt}x(t_{\min 2})\right), \quad (14)$$

where

$$\text{low}\left(\frac{d}{dt}x(t_{\text{spurt}})\right)$$

is the lower bound of the confidence band of the velocity where the spurt occurs and

$$\text{high}\left(\frac{d}{dt}x(t_{\min 1})\right)$$

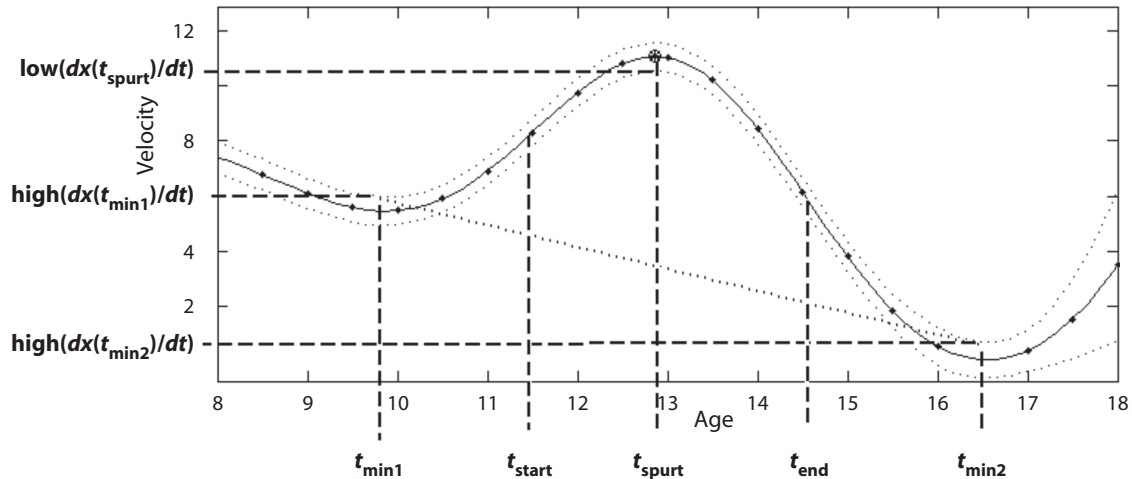


Figure 3. Example of a significant spurt.

and

$$\text{high}\left(\frac{d}{dt}x(t_{\min 2})\right).$$

are the upper bounds of the confidence bands of the velocity minima surrounding the spurt. We return to the independence assumption in the Discussion section.

Figure 3 presents an example of a genuine—that is, statistically significant—spurt. For this spurt, the lower bound of the confidence band at the point of maximal velocity of the spurt is larger than the higher bound of the confidence bands on the two surrounding minima. Furthermore, the dotted line joining the higher parts of confidence bands of the two minima is not fully included in the function's confidence band. Therefore, at the .05 level, we reject the hypothesis that a straight line can account for the data, and thus a significant increase in velocity (i.e., a spurt) is present.

Figure 4 presents an example of a spurious—that is, not statistically significant—spurt. Dotted lines surrounding

the velocity function represent 95% confidence bands. We see that the lower bound of the confidence band at the point of maximal velocity of the spurt,

$$\text{low}\left(\frac{d}{dt}x(t_{\text{spurt}})\right)$$

is smaller than the higher bound of the confidence bands on the two surrounding minima,

$$\text{high}\left(\frac{d}{dt}x(t_{\min 1})\right)$$

and

$$\text{high}\left(\frac{d}{dt}x(t_{\min 2})\right).$$

Furthermore, the dotted line joining the upper parts of confidence bands of the two minima is completely included in the function's confidence band, and therefore, we cannot reject the hypothesis that a straight line between the minima (i.e., without an increase in velocity) can account for the data.

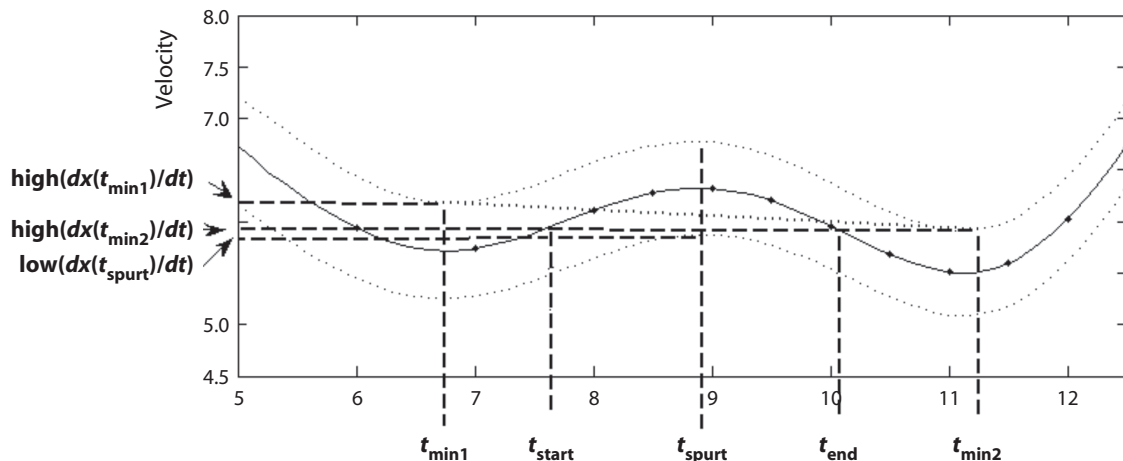


Figure 4. Example of a nonsignificant spurt.

Visualization tools. In addition to automatic detection and quantification of significant spurts, we included visualization tools in our AMD system. First, each individual case can be plotted (see Figures 1, 13, 14, and 15 for examples). On the top level, the plot shows a graph of the fitted function using FDA overlapped with the raw data (y_j, t_j). The lower graphs show the first three derivatives: velocity, acceleration, and jerk curves, respectively. For these graphs of derivatives, dots indicate where observations were taken [i.e., $d^n/dt^n x(t_j), t_j$ where x is the fitted function]. Dotted lines mark lower and upper bounds of the confidence intervals. Circled dots indicate where significant spurts are located in the derivative functions.

Second, group data are summarized on a single graph using an innovative graphing technique. Figures 5 and 6 are examples of height data as a function of age from the Berkeley study (Tuddenham & Snyder, 1954), for boys and girls, respectively. These data will be further discussed in the Results section. Each horizontal line represents a significant spurt. Line thickness is proportional to the amplitude, with the amplitude scale adjusted to fit the largest spurt in the data to be plotted. Line length indicates spurt duration. Line ends coincide with the start and the end of the spurt. The number superposed on the line indicates the order of the spurt (if there are more than one, the first spurt has a label “1,” the second “2,” and so on). The position of each such number label indicates the point of maximal velocity. The results shown in Figures 5 and 6 are sorted by points of maximal velocities, but AMD also provides other sorting options.

RESULTS

We applied our AMD technique to two problems: (1) physical growth data measuring children’s height at various ages and (2) simulated and real vocabulary growth in children. In all the analyses presented, a significance level (α) of .05 was used.

Physical Growth

To study physical growth, we obtained data from the Berkeley growth study (Tuddenham & Snyder, 1954) in which various physical measures were collected, including height. The study comprised 66 boys and 70 girls. The height of the children was measured four times a year from 0 to 2 years of age, then annually from 2 to 8 years, and finally twice a year from 8 to 18 years. As in Ramsay et al. (1995), we selected the data from 1 to 18 years of age. Because the data were sparse, we excluded children who had any data missing between 1 and 18 years old; these were 9 girls (ID = 309, 311, 316, 321, 331, 332, 334, 368, 383) and 12 boys (ID = 201, 203, 208, 209, 210, 215, 218, 222, 233, 260, 269, 274). Our analysis was based on the remaining 54 boys and 61 girls.

When applied to these data, AMD successfully detected and measured three important and well-known phenomena of physical growth of children: (1) an adolescent growth spurt in virtually all the children; (2) an earlier age of onset

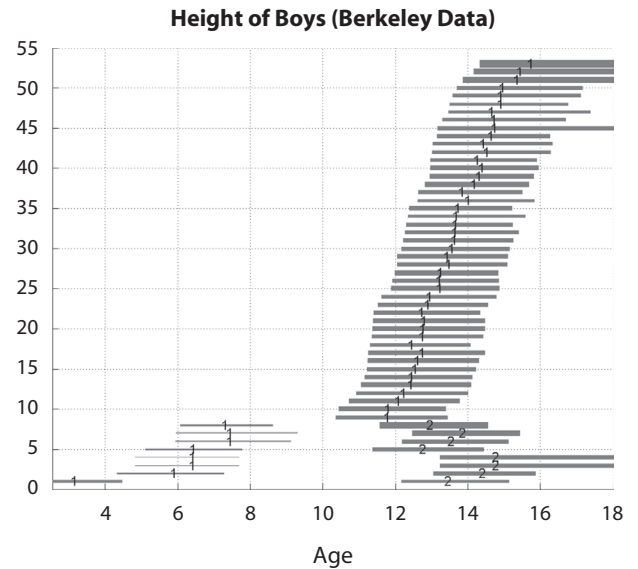


Figure 5. Summary plot of spurt characteristics for boys in the Berkeley growth study. Line thickness represents amplitude, line length represents duration, integer values represent the spurt number, and integer placement indicates the point of maximal velocity.

for girls’ adolescent growth spurts than for boys’ (e.g., Frisch & Revelle, 1971); and (3) a smaller preadolescent growth spurt in some children (such spurts tend to be small and difficult to detect; Molinari, Largo, & Prader, 1980).

Figure 5 graphically summarizes the spurt results for all the boys in the Berkeley study, using a smoothing level λ of 0.1. AMD showed a statistically significant adolescent

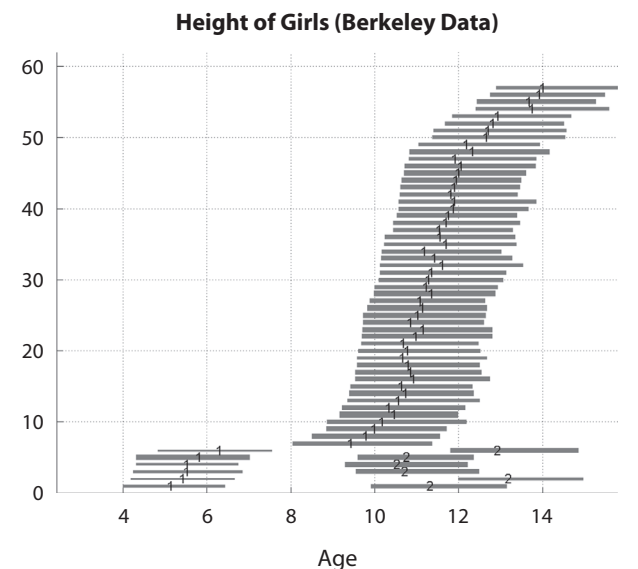


Figure 6. Summary plot of spurt characteristics for girls in the Berkeley growth study. Line thickness represents amplitude, line length represents duration, integer values represent the spurt number, and integer placement indicates the point of maximal velocity.

spurt in 53 of the 54 boys (98%) and a mid-growth (prepubertal) spurt in 8 boys (15%). No boy had more than two significant spurts. Figure 6 graphically summarizes the spurt results for all the girls in the Berkeley study, using the same smoothing level λ of 0.1. AMD found a statistically significant adolescent spurt in 58 of the 61 girls (95%) and a mid-growth (prepubertal) spurt in 6 girls (10%). No girl had more than two significant spurts.

Smoothing parameter analysis. To explore the effect of smoothing, we compared results obtained with previously used smoothing λ values of 0.01 (Shultz, 2003) and 0.34 (Ramsay et al., 1995), also adding 0.1 as an intermediate value. Shultz chose a smoothing value of 0.01 because it generated sensible results in accord with the literature. This serves as an illustration of how AMD measures of spurts can be subjected to further statistical analysis.

We did not tailor the level of smoothing for each individual because we hypothesized that noise in the data would likely be similar across individuals. Because children's heights were measured in similar experimental conditions, variance could be expected to be similar across individuals. Also, for the purpose of this analysis, we assumed that the variance of the biological processes involved in growth would be similar across children. Finally, it was not clear how to establish the best smoothing level for each case.

Contingency tables of the frequency of preadolescent and adolescent growth spurts for the three levels of smoothing are presented in Tables 1 and 2 for boys and girls, respectively. A chi-square test revealed that frequency of AMD-detected spurts was not statistically different for the three levels of smoothing for boys [$\chi^2(2) = 1.3, p > .05$] or for girls [$\chi^2(2) = 2.1, p > .05$].

Next, we subjected adolescent spurt measures to a two-way mixed ANOVA, using gender as an independent factor and the three levels of smoothing as a repeated factor. Separate ANOVAs were performed on start, point of maximal velocity, amplitude, and duration. To avoid missing

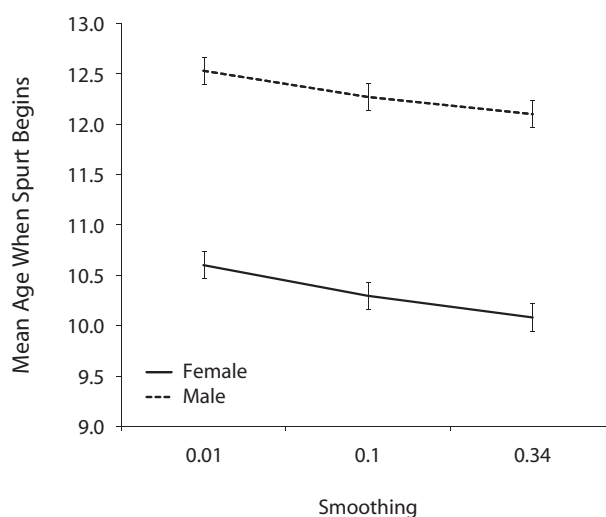


Figure 7. Age at which adolescent growth spurts begin, by gender and smoothing level. Error bars represent standard errors.

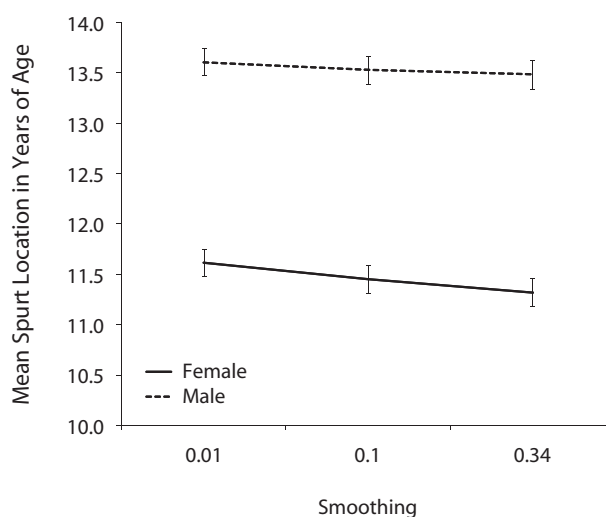


Figure 8. Age at which adolescent growth spurts occur, by gender and smoothing level. Error bars represent standard errors.

Table 1

Contingency Table of the Frequency of Preadolescent and Adolescent Growth Spurts for the Three Values of Smoothing for the 54 Boys in the Berkeley Growth Study

	Level of Smoothing (λ)			Total
	0.01	0.1	0.34	
Preadolescent	4	8	6	18
Adolescent	54	53	53	160
Total	58	61	59	178

Table 2

Contingency Table of the Frequency of Preadolescent and Adolescent Growth Spurts for the Three Values of Smoothing for the 61 Girls in the Berkeley Growth Study

	Level of Smoothing (λ)			Total
	0.01	0.1	0.34	
Preadolescent	6	6	2	14
Adolescent	57	58	57	172
Total	63	64	59	186

data, only children who had adolescent spurts that replicated across the three levels of smoothing were included in this analysis (53 males and 55 females).

The ANOVA of spurt start revealed a main effect of gender [$F(1,106) = 100, p < .001$], a main effect of smoothing [$F(2,212) = 455, p < .001$], and a smoothing \times gender interaction [$F(2,212) = 4.1, p < .05$]. As we can see in Figure 7, adolescent spurts started earlier in girls than in boys and earlier with more smoothing. The small interaction indicates a slightly larger effect of smoothing for girls than for boys.

The ANOVA of point of maximal velocity also revealed a main effect of gender [$F(1,106) = 113, p < .001$], a main effect of smoothing [$F(2,212) = 351, p < .001$], and a smoothing \times gender interaction [$F(2,212) = 58, p < .001$]. Figure 8 indicates similar effects for point of

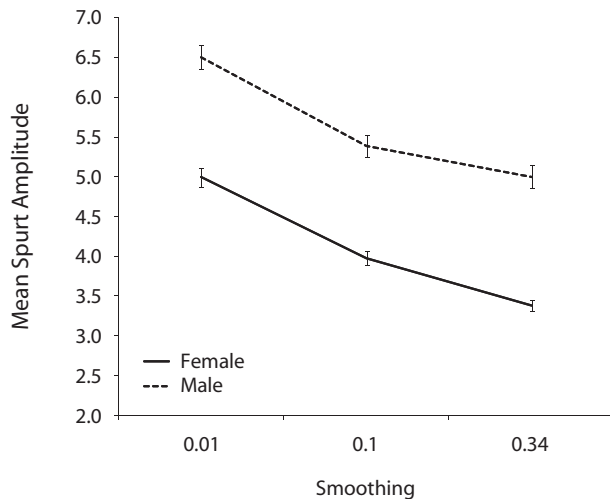


Figure 9. Amplitude of adolescent growth spurts, by gender and smoothing level. Error bars represent standard errors.

maximal velocity as for spurt start; that is, girls tend to have their adolescent spurt at a younger age than do boys, and AMD reports these spurts as occurring at a slightly earlier age with increased smoothing.

The ANOVA of spurt amplitude revealed a main effect of gender [$F(1,106) = 101, p < .001$], a main effect of smoothing [$F(2,212) = 261, p < .001$], but no interaction. Inspection of Figure 9 reveals that boys had more intense adolescent spurts than did girls and that increased smoothing resulted in smaller amplitudes.

The ANOVA of spurt duration revealed a main effect of gender [$F(1,106) = 16, p < .001$], a main effect of smoothing [$F(2,212) = 359, p < .001$], and a smoothing \times gender interaction [$F(2,212) = 12, p < .001$]. In Figure 10, we see that adolescent spurts lasted longer for boys than for girls and that duration increased with smoothing. The interaction reflects that duration was less sensitive to smoothing for girls than for boys.

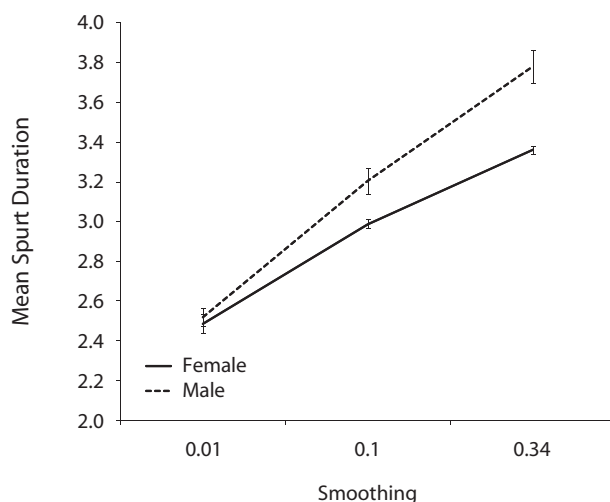


Figure 10. Duration of adolescent growth spurts, by gender and smoothing level. Error bars represent standard errors.

In short, we found a huge gender effect in spurt start, point of maximal velocity, and amplitude, at all levels of smoothing we tested. Girls tended to have their adolescent growth spurt at younger ages than did boys. Spurts tended to be less intense and shorter for girls than for boys. Increased smoothing made spurts appear longer and less intense, and changes of the smoothing parameter had a small, yet significant, impact on the timing of the adolescent spurt.

Simulated Vocabulary Growth

As a second illustration of the AMD technique, we analyzed simulated data generated by a model of vocabulary growth in children (McMurray, 2007; Mitchell & McMurray, 2008, 2009). This mathematical model simulated the cumulative count of words learned by children. More difficult words take more time to learn, and difficulties have some statistical distribution. Here, we used a normal distribution of difficulties. At each time step, each word accumulated one point, and when its points surpassed its difficulty, the word was considered as learned. The simulations showed that a vocabulary spurt naturally emerges as a mathematical consequence of words' varying in difficulty level and being learned in parallel, suggesting that various psychological explanations are not necessary to explain a vocabulary spurt (McMurray, 2007; see van Geert, 1991, for an alternative explanation).

In comparison with height data, where the dependent variable is continuous, vocabulary growth exhibits the challenging and interesting feature that the dependent variable is discrete. When measuring children's vocabulary size, only uttered words that are complete and recognizable are counted, and not failed attempts (which might be considered as fractions of words). Similarly, in the McMurray (2007; Mitchell & McMurray, 2008, 2009) model, cumulative word counts are discrete integer values representing complete words, computed as the integral of a single smooth distribution. Depending on the theoretical approach taken, the jitter caused by the discrete nature of measurements of the dependent variable (here, the word count) might be treated as noise or error.

Thus, as an additional benefit of using these artificial data, we can investigate whether the number of spurts detected by AMD depends on data sampling. In a number of data sets including actual and simulated conservation acquisition and physical growth, more densely sampled data tend to show more spurts (Shultz, 2003).

Results. As can be seen in Figure 11, there is a large global spurt with maximal velocity at around 2,000 time steps. In addition, numerous minispurts due to the staircase nature of the cumulative sum of integer values can be seen on shorter time scales, as illustrated in Figure 12.

We analyzed these simulated vocabulary growth data with AMD, manipulating sampling density: a high sampling rate using all 4,000 data samples and a low sampling rate using every 20th sample, for a total of 200 samples (4,000/20). Lowering the sampling rate acts as a low-pass filter that can reduce the jitter of the cumulative vocabulary curve. The staircase nature of the complete cumulative word count curve implies high-frequency

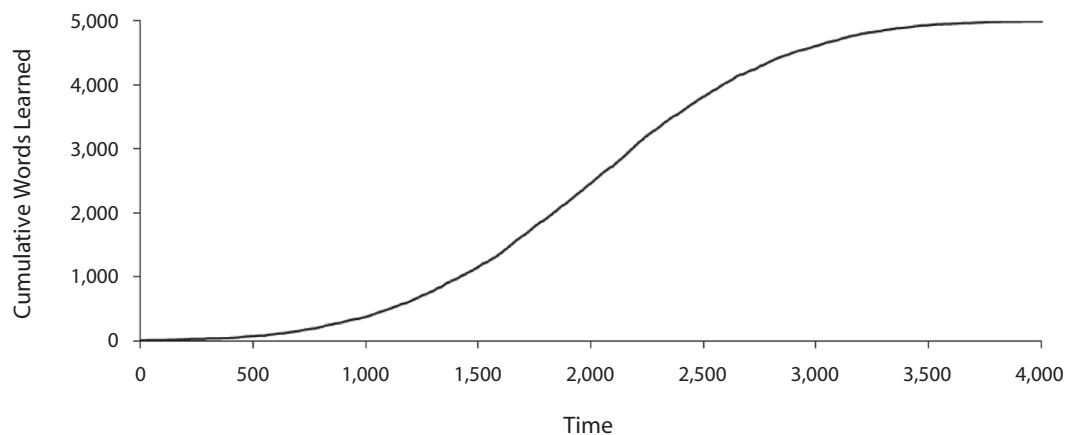


Figure 11. Cumulative words learned as a function of simulated time steps (complete curve).

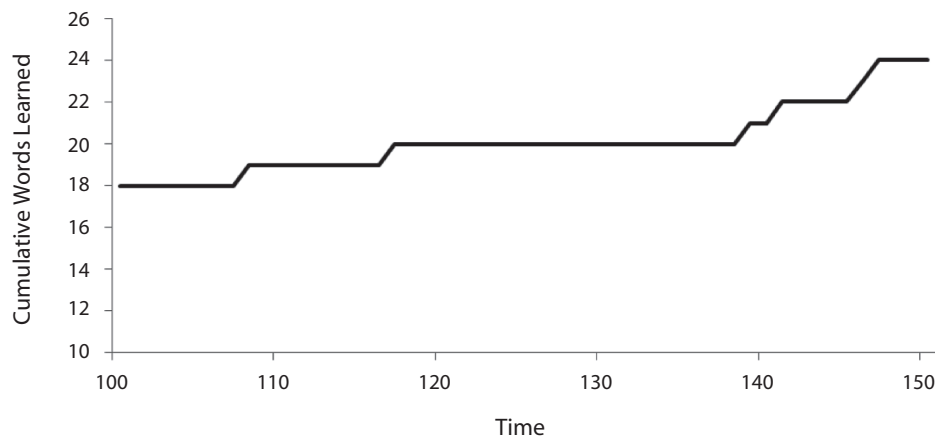


Figure 12. Cumulative words learned as a function of simulated time steps 100 to 150.

spectral components. We empirically determined that a smoothing value λ equal to 1.0×10^6 was appropriate for these data.

As was expected, AMD showed one large global spurt under low sampling. As is shown in Figure 13, we indeed observe one circle indicating a single significant spurt. Furthermore, we anticipated finding many local minispurts under high sampling. Indeed, AMD showed 68 smaller spurts under high sampling, represented as circles in Figure 14.

In short, this analysis shows how spurt detection is sensitive to smoothing and sampling. We found one large spurt in sparsely sampled data and many smaller spurts in more densely sampled data. Given the theoretical assumption that these minispurts are due to noise, one would prefer the lower sampling. An exhaustive discussion of the effect of sampling rate on spurt detection can be found in Adolph, Robinson, Young, and Gill-Alvarez (2008).

Real Vocabulary Growth

Finally, we investigated vocabulary growth in real data from Corrigan (1978). These data were collected from

three English-speaking children every third week between 9 and 27 months of age. Children’s vocabulary sizes were measured in 30-min videotaped sessions. We analyzed these data with AMD using a smoothing value of 0.01, as in a less systematic analysis of the same data (Shultz, 2003).

Consistent with what Shultz (2003) noted, the child named Ashley exhibited three velocity peaks (see Figure 15), and all of them were statistically significant. AMD showed two and one significant spurts, respectively, for the other two children, for an average of 2.0 spurts among all three children.

Our preliminary results applying AMD to simulated and real vocabulary growth suggest that vocabulary spurts do exist and that, depending on sampling density, there can be more than a single spurt. For the data presented here, density of the real vocabulary (18 data points) is much lower than the simulated vocabulary (200+ data points). For this reason, the simulated data have more high-frequency signals and, thus, require more smoothing (1×10^6 vs. 0.01 for the real vocabulary). This may suggest that these real data are undersampled for the purpose of identifying growth spurts.

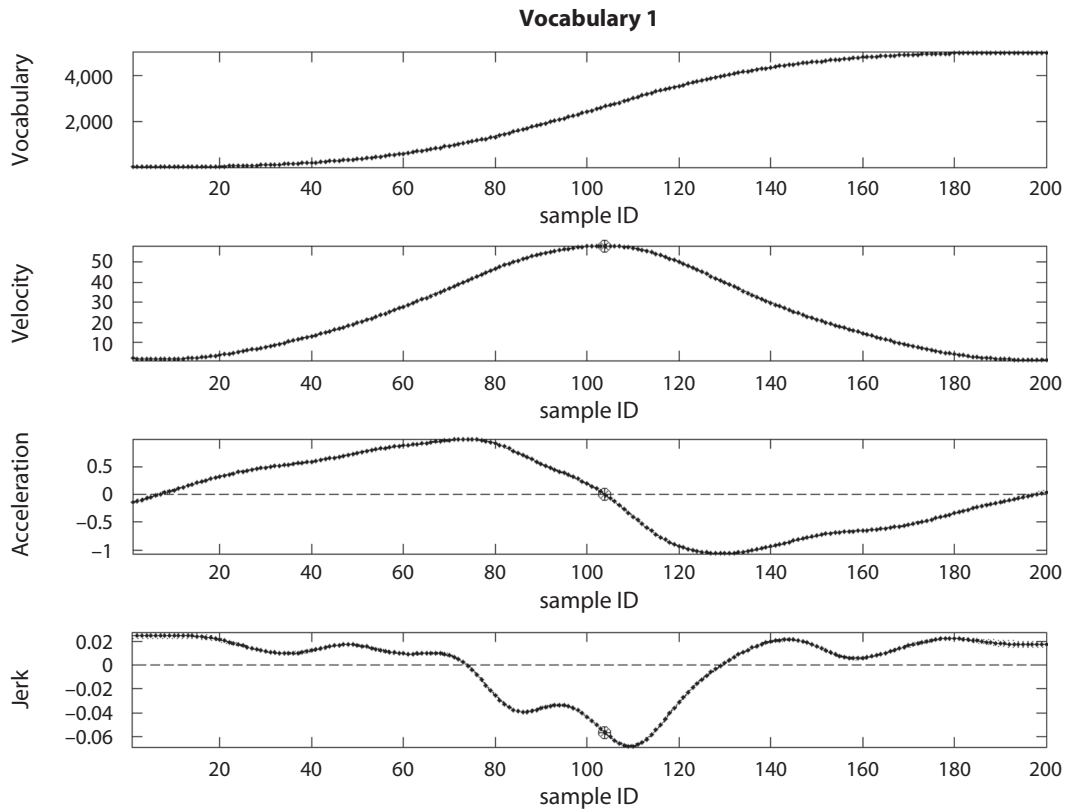


Figure 13. Results of the McMurray (Mitchell & McMurray, 2008, 2009) model for simulated vocabulary data under low sampling (1:20).

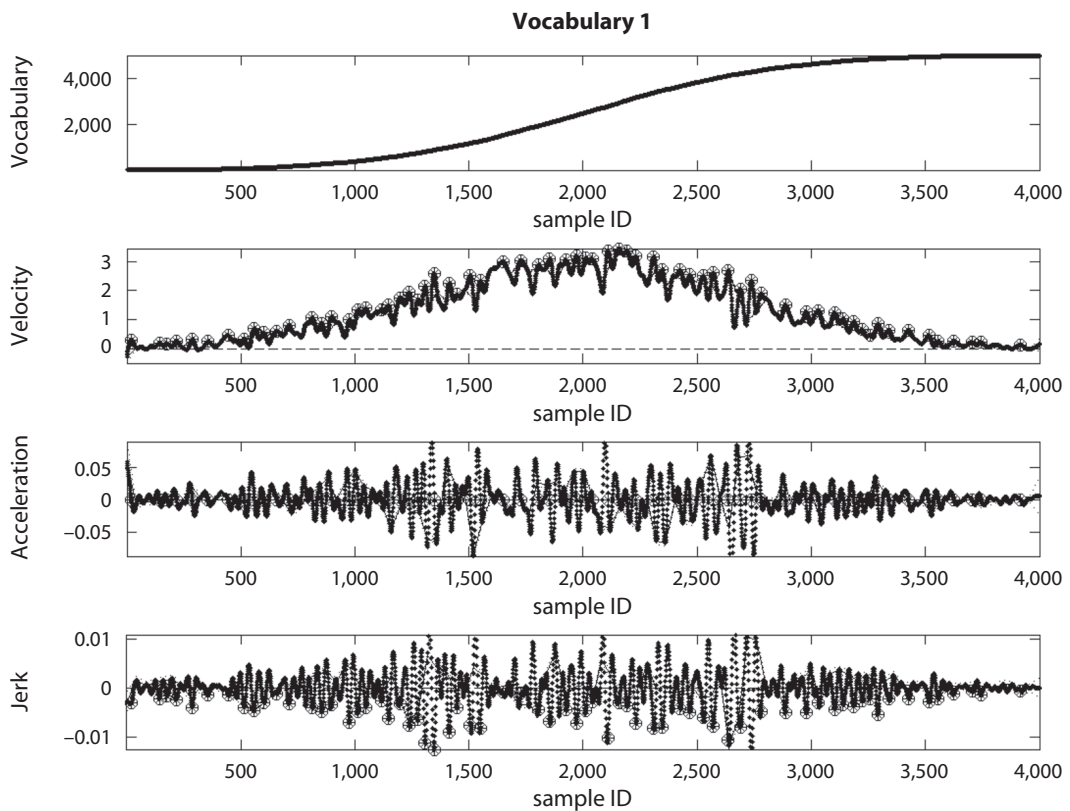


Figure 14. Results of the McMurray (Mitchell & McMurray, 2008, 2009) model for simulated vocabulary data under high sampling (1:1).

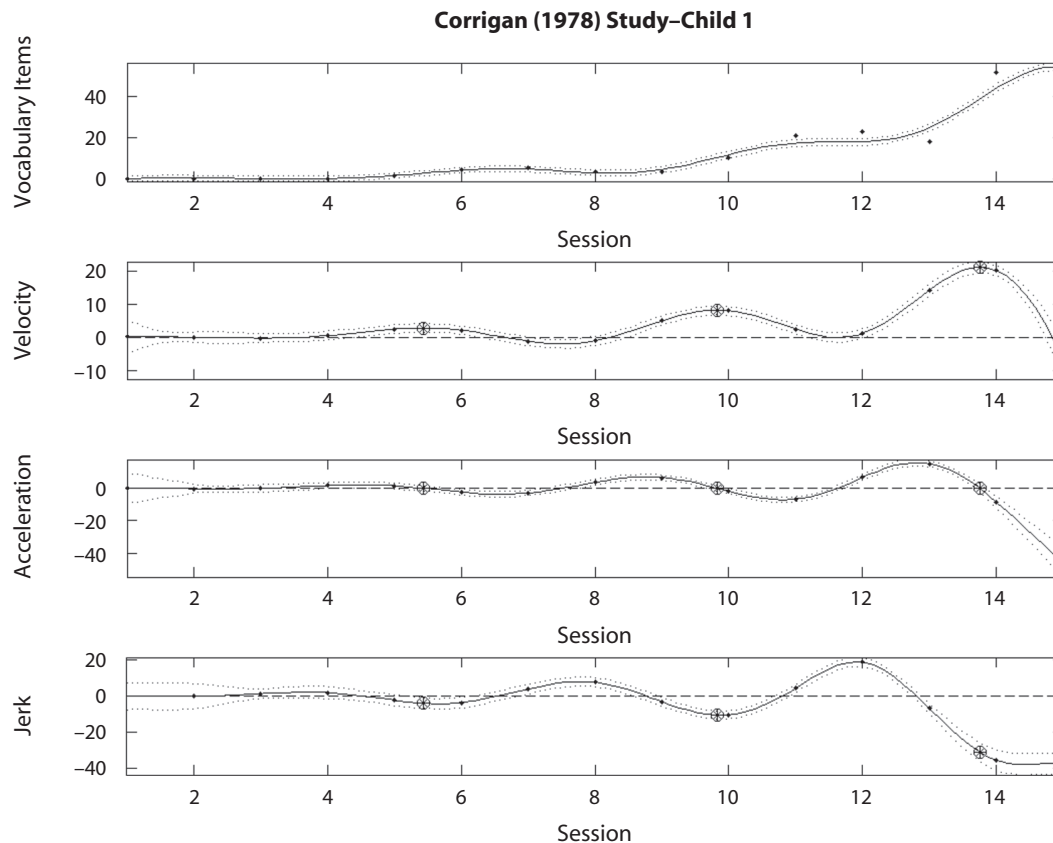


Figure 15. Results for the child named Ashley in the Corrigan (1978) study.

DISCUSSION

Our AMD method identifies systematic and reliable spurts and quantifies the point of maximal velocity, start, amplitude, and duration of the spurts. AMD automated detection of growth spurts, replicated well-known phenomena in physical and vocabulary growth, allowed us to explore the effect of smoothing and sampling on identified spurts, and informed debates about the existence of such spurts.

Other techniques have been proposed to deal with the important problem of reliable spurt detection. For example, Harezlak and Heckman (1999) analyzed the Berkeley growth data, defining spurts as local maxima in the velocity curve, similar to our definition. Their system tests the hypothesis that a curve has, at most, k spurts against the alternative hypothesis that the curve has more than k spurts, with p values computed using bootstrapping. Their technique successfully identified an adolescent growth spurt in at least 98% of the children.

In addition, Molinari et al. (1980) introduced a statistical technique based on the null hypothesis that velocity growth curves are convex between birth and puberty. They reported a small but significant mid-growth spurt in height in most children.

Finally, van Dijk and van Geert (2007) took a different theoretical approach to growth centered around the con-

cept of discontinuities rather than spurts. They designed a system that tests the null hypothesis that some continuous model with noise is capable of generating the observed data. Discontinuities are identified when this null hypothesis is rejected. Such discontinuities are analogous to spurts because they correspond to sufficiently large increases in rate of growth. Using their technique, van Dijk and van Geert successfully identified discontinuities in language development data.

Strengths of the AMD Method

First, AMD is a useful data-mining and exploratory tool when the sheer amount of data would make manual identification of spurts prohibitively costly—for example, imaging data such as event-related potentials and data from eyetracking devices.

Second, AMD is more comprehensive than other approaches to spurt detection. For instance, the techniques proposed by van Dijk and van Geert (2007) and Harezlak and Heckman (1999) can identify spurts, but not where they occur, how large they are, and how long they last. The technique used by Molinari et al. (1980) allows identification of the time when the spurt is most intense (similar to the point of maximal velocity in AMD) and the velocity at that point. But it does not locate when the spurt starts, how intense it is (amplitude), and how long it lasts.

Third, AMD is more automated and less subjective than predecessors that require expert judgments to identify spurts (Ramsay et al., 1995; Shultz, 2003).

Fourth, like other techniques, AMD can be applied both to individual records and to averaged sample-level analyses. Averaging across curves can help to remove noise present in individual curves. However, averaging across subjects can easily obscure spurts when they occur at different times for different children (James, 2008). With AMD, features measured on individual records can be used for population-level analyses. For instance, in this article, we compared spurt amplitudes, points of maximal velocities, and durations in two populations (boys and girls), using ANOVAs.

Local and Global Spurts

Although AMD is designed to detect local spurts, it can also be used to detect more global features, with appropriate data preprocessing. We saw in the vocabulary growth example that lowering data sampling and increasing smoothing act as a low-pass filter that allows more global features to emerge. Periodic growth such as a sine wave would require Fourier basis functions in FDA, and we would expect AMD to show periodic spurts in such data.

Spurts and Smoothing

As was mentioned, no universally agreed-upon method for smoothing exists. For confirmatory analyses, users may already have sufficient expertise and knowledge of the data to manually select an appropriate smoothing value. As we saw, the case in which the dependent variable (e.g., vocabulary size) is observed and measured as a discrete (stairlike) variable is especially challenging. A possible theoretical perspective that can be taken is to treat these apparent abrupt increases as jitter or noise, not as spurts, and adjust smoothing and sampling accordingly.

For exploratory analyses, AMD can be used to perform sensitivity analyses to smoothing, as we did for height data. Researchers may find it valuable to know how smoothing influences the number and characteristics of the reliable spurts detected. In addition, AMD could be used for the reverse problem. Given unknown noise but theoretical grounds for expecting a certain number of spurts or spurt characteristics, one can infer how much smoothing is necessary and, thus, may learn something about noise in the data.

Interpreting Spurts

As with any other statistical tool, AMD should be used with caution, and domain expertise may be required for correctly interpreting the spurts that were detected, since some cases may be ambiguous. For instance, the simulated vocabulary growth illustrated the challenge of interpreting discrete dependent values. Observed data, as shown in Figure 12, exhibited a stairlike profile and, consequently, appeared as spurts that AMD correctly detected. Here, we know that the pattern was generated by a continuous underlying growth process that could be observed only in a

discrete manner. However, a discontinuous process would arguably also produce a similar pattern. The nature of the measurement (uttered words) does not allow us to discriminate the two on the basis of the observed data alone, and thus other, domain knowledge is required.

Limitations of the AMD Method

First, no automated method, including ours, can extract anything meaningful from very noisy or overly sparse data. Statistical techniques require good data, and this is especially true of AMD, because it extensively uses higher derivatives, which are particularly sensitive to noise (Ramsay et al., 1995).

Second, the identification of significant spurts as non-overlapping confidence bands relies on the assumption that the maxima and the minima of the function are independent and uncorrelated. Although it is simple and intuitively appealing, this approach is unrealistic and may result in an under- or overestimation of the width of the confidence bands. A more complex approach based on the variance-covariance matrix estimated by FDA would provide more accurate estimates of confidence bands when maxima and minima are correlated (see Ramsay & Silverman, 2005, for details).

Finally, our relatively conservative definition of spurts excludes some events that other methods may deem to be spurts. Our system requires both a significant increase in velocity before the spurt and a significant decrease after the spurt. When only one of these conditions is met, AMD excludes a possible spurt. This may explain why we detect fewer spurts than do some other researchers, such as Molinari et al. (1980). Conservatism in spurt detection does have advantages. For instance, if a child has a long period of relatively constant growth, followed by a second, long period of constant growth but at a lower rate, it is unclear whether either of these periods should be considered a spurt. It seems more correct to consider both periods as plateaus, without any spurt.

Spurts and Stages of Development

Identifying spurts in growth curves may inform us about stages of development. As we argued in the introduction, noise is inevitable in experimental data, causing apparent spurts that are not genuine. Thus, an important first step in studying stages of development is to identify statistically reliable spurts. Rigorous spurt detection criteria like those of AMD appear vital to achieving this goal.

Many results and theories in psychology suggest such stages; some are well accepted (e.g., many of the phenomena studied by Piaget and his colleagues [Piaget & Inhelder, 2000] and Siegler's [1998] balance scale problem), but others may be more contentious (e.g., vocabulary growth; Ganger & Brent, 2004). AMD could help address such controversial issues, providing a rigorous basis for researching difficult issues in linearity, nonlinearity, and possible stages of development. When psychological growth can be measured as a continuous variable, detected spurts could identify potential transitions between stages, which in turn could be charac-

terized as relative plateaus between spurts. Suggested stage locations could then be tested further by examining knowledge representations at each suggested stage to see whether they are qualitatively different (Flavell, 1971) or fulfill other stage criteria.

Using FDA for Other Kinds of Data

Identifying spurts in time-based data such as physical growth is only a specific case of a more general problem of identifying maxima in some derivative of two-dimensional functional data. Recall that spurts are defined as velocity maxima. AMD can identify and quantify maxima in a fitted function and any of its derivatives—hence, the use of *maxima*, rather than *spurt*, in its name. In particular, locating maxima in the fitted function itself could be very useful—for example, in brain-imaging data such as EEG and ERP. Users would need to instruct AMD to identify maxima in the fitted function itself, rather than in its first derivative (velocity). Harezlak and Heckman (1999) similarly extended their system to look for spurts in derivatives of curves.

In this article, we assumed that the independent variable was time because growth is generally defined over time, but AMD can be used to analyze any two-dimensional functional data where the independent variable represents a property other than time. The meaning of maxima in the function or in its first derivative would depend on the nature of those variables.

Monte Carlo Approach to Testing AMD

We used McMurray's (2007; Mitchell & McMurray, 2008, 2009) vocabulary model to generate artificial data known to possess a spurt at the inflection point of a logistic function and then determined that AMD could identify that spurt. A possible avenue for future research would be to apply this Monte Carlo approach more generally and intensively. Namely, samples of artificial data of known characteristics (including some parameter variance) could be generated, and spurt detection mechanisms such as AMD could be tested and compared on the basis of their performance in detecting and quantifying these spurts. This approach would allow numerical estimation of important characteristics, such as rates of false alarms (Type I errors) and of false negatives (Type II errors) in spurt detection.

CONCLUSIONS

AMD uses an innovative and rigorous approach based on confidence bands to largely automate the spurt detection process. Although the generation of a smoothed FDA function still requires manual selection of the smoothing parameter λ , AMD surpasses other approaches in its degree of automation. Automation enforces a strict criterion for significant spurts, which reduces the need for subjective judgment. Also, AMD quantifies more aspects of spurts than other methods do. In short, AMD goes further than other methods in being objective, quantitative, and comprehensive.

AUTHOR NOTE

This work was supported by a McGill Major scholarship to F.D. and by a grant to T.R.S. from the Natural Sciences and Engineering Research Council of Canada. We thank James Ramsay, François Rivest, and Signe Bray for providing seminal ideas, and Kristine H. Onishi, Fermin Moscoso Del Prado, and Bob McMurray for insightful comments and suggestions. Correspondence concerning this article should be addressed to F. Dandurand, Laboratoire de Psychologie Cognitive, CNRS, Aix-Marseille University, 3, place Victor Hugo, 13331 Marseille, France (e-mail: frederic.dandurand@gmail.com).

REFERENCES

- ADOLPH, K. E., ROBINSON, S. R., YOUNG, J. W., & GILL-ALVAREZ, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*, 527-543. doi:10.1037/0033-295X.115.3.527
- CARLSTEIN, E. G., MÜLLER, H.-G., & SIEGMUND, D. (Eds.) (1994). *Change-point problems* (Vol. 23). Hayward, CA: Institute of Mathematical Statistics.
- CLEVELAND, W. S., GROSSE, E., & SHYU, W. M. (1992). Local regression models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 309-376). New York: Chapman & Hall.
- CORRIGAN, R. (1978). Language development as related to stage 6 object permanence development. *Journal of Child Language*, *5*, 173-189.
- EFRON, B. (1981). Nonparametric estimates of standard error: The jack-knife, the bootstrap and other methods. *Biometrika*, *68*, 589-599.
- EFRON, B., & TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- EUBANK, R. L. (1999). *Spline smoothing and nonparametric regression* (2nd ed.). New York: Dekker.
- EVELETH, P. B., & TANNER, J. M. (1990). *Worldwide variation in human growth*. Cambridge: Cambridge University Press.
- FLAVELL, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, *2*, 421-453.
- FRISCH, R. E., & REVELLE, R. (1971). The height and weight of girls and boys at the time of initiation of the adolescent growth spurt in height and weight and the relationship to menarche. *Human Biology*, *43*, 140-159.
- GANGER, J., & BRENT, M. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, *40*, 621-632.
- GREEN, P. J., & SILVERMAN, B. W. (1994). *Nonparametric regression and generalized linear models*. London: CRC Press.
- HAREZLAK, J., & HECKMAN, N. E. (1999). *CrisP—A tool for bump hunting*. Vancouver: University of British Columbia, Department of Statistics.
- HARRELL, F. E. (2001). *Regression modeling strategies with applications to linear models, logistic regression and survival analysis*. New York: Springer.
- HUGHES, B. D. (1996). *Random walks and random environments*. Oxford: Oxford University Press.
- JAMES, G. M. (2008). *Moments based functional synchronization*. Los Angeles: University of Southern California.
- JOLICÉUR, P., PONTIER, J., & ABIDI, H. (1992). Asymptotic models for the longitudinal growth of human stature. *American Journal of Human Biology*, *4*, 461-468.
- McMURRAY, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*, 631.
- MITCHELL, C. C., & McMURRAY, B. (2008). A stochastic model for the vocabulary explosion. In *Cognitive Science Conference Proceedings* (pp. 1919-1924). Washington, DC: Cognitive Science Society.
- MITCHELL, C. C., & McMURRAY, B. (2009). On leveraged learning in lexical acquisition and its relationship to acceleration. *Cognitive Science*, *33*, 1503-1523. doi:10.1111/j.1551-6709.2009.01071.x
- MOLINARI, L., LARGO, R. H., & PRADER, A. (1980). Analysis of the growth spurt at age seven (mid-growth spurt). *Helvetica Paediatrica Acta*, *35*, 325-334.
- O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, *1*, 502-518.
- PIAGET, J., & INHELDER, B. (2000). *The psychology of the child*. New York: Basic Books.
- RAMSAY, J. O., BOCK, R. D., & GASSER, T. (1995). Comparison of height

- acceleration curves in the Fels, Zurich, and Berkeley growth data. *Annals of Human Biology*, **22**, 413-426.
- RAMSAY, J. O., HOOKER, G., & GRAVES, S. (2009). *Functional data analysis with R and MATLAB*. New York: Springer.
- RAMSAY, J. O., & SILVERMAN, B. W. (2005). *Functional data analysis*. New York: Springer.
- SHAWE-TAYLOR, J., & CRISTIANINI, N. (2004). *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.
- SHULTZ, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.
- SIEGLER, R. S. (1998). *Children's thinking* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- SIMONOFF, J. S. (1996). *Smoothing methods in statistics*. New York: Springer.
- TANNER, J. M. (1978). *Growth at adolescence* (2nd ed.). Oxford: Blackwell Science.
- TUDDENHAM, R. D., & SNYDER, M. M. (1954). *Physical growth of California boys and girls from birth to eighteen years*. Berkeley: University of California Press.
- VAN DIJK, M., & VAN GEERT, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant & Child Development*, **16**, 7-33.
- VAN GEERT, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, **98**, 3-53.

NOTES

1. We use the following notation to refer to the estimated function and its derivatives:

$x(t)$	the estimated function itself;
$\frac{d}{dt}x(t)$	the first derivative, or velocity, of the estimated function;
$\frac{d^2}{dt^2}x(t)$	the second derivative, or acceleration;
$\frac{d^3}{dt^3}x(t)$	the third derivative, or jerk.

2. Although AMD requires three derivatives, we just saw that FDA requires five derivatives for its roughness penalty method. However, other smoothing techniques may require different numbers of derivatives.

3. Acceleration of concave and convex functions does not change sign; thus, they contain no spurt, according to this definition.

4. Note that FDA is able to numerically estimate derivatives at any point, not only knots.

(Manuscript received June 24, 2009;
revision accepted for publication February 20, 2010.)