

A nonlinear regression model for distance-velocity curve of 100m sprint

Katsuyuki HAGIWARA and Masaaki SUGITA

Faculty of Education, Mie University

Abstract Position data during a 100m sprint are usually recorded by using the LABEG(LASer VELOCITY Guard) system. The recorded position data is transformed into a velocity course by a certain method and a distance-velocity property is often focused in analyzing a 100m sprint. The transformed velocity data is noisy in general if we apply a simple numerical differentiation. Therefore, we need a device for extracting a smooth distance-velocity property from the raw velocity data. In this paper, we proposed a nonlinear regression model for this purpose. The nonlinear function in our model is composed of a sum of two functions which represent a velocity increase in the early stage of a 100m sprint and a velocity decrease in the later stage respectively. The former function is an exponential function to represent a rapid speed increase at a sprint start and the latter function is a polynomial function to represent a gradual decrease due to fatigue in the later state of a sprint. We apply this model to analyze collected LABEG data of students in elementary and junior high school under an appropriated pre-processing for position data. The model parameters were estimated by the gradient descent method for the least squares estimation. As a result, we verified that this model can well represent distance-velocity properties of the collected data. By the analysis based on the estimated curves, we clarified that the maximum velocity is the most important factor for a 100m sprint time. This is consistent with the well known speculation. We also found that the start acceleration and velocity decrease at the later stage are also related to the 100m sprint time. Furthermore, we introduced a derivative of the estimated distance-velocity curve and found that it's values at around sprint start are well related to a sprint time. Fortunately, we could confirm that the similar insights were obtained by estimated parameter values. This is because we have assumed a structured model in which the parameters directly reflect characteristics of each phase of a 100m sprint.

1 Introduction

For the purpose to investigate the characteristics of a 100m sprint, position data during the 100m sprint are recorded by using the LABEG(LASer VELOCITY Guard) system[5]. The characteristics of a 100m sprint is often analyzed based on velocity change during the sprint[6, 5];i.e. time-velocity course or distance-velocity course in which the distance is calculated from the start line. Analysis of the velocity change is important as in below.

- As well known, a 100m sprint is mainly divided into the three phases that are acceleration phase, maximal speed phase and speed maintenance phase[1];e.g. see [2] for more detailed categorization. These phases correspond to the early, middle and later stage of a 100m dash and are clearly found in the velocity change during the sprint.
- In the analysis of a 100m sprint, for example, it is well known that a 100m sprint time is very correlated to the maximum velocity[1, 3]. The maximum velocity may be directly estimated if we have velocity data.
- For capturing the characteristics of a 100m sprint, [7] has applied the factor analysis for velocity data calculated at several portions of distances from a start line.

A naive method to calculate velocity courses from recorded time-position data is to apply a numerical differentiation. However, it is very noisy since the position data include measurement noise and it is amplified by the numerical differentiation. [5] has considered moving average and 3 point digital filtering of time-position data before a simple numerical differentiation. In these methods, however, it is difficult to determine an appropriate number of points for averaging and/or cut-off frequency of a filter. Actually, the main theme in [5] is to explore those parameters heuristically. On the other hand, [6] has proposed the amount of deviation of subject's running velocity(ADV) as an index for velocity decrement in a 100m sprint. It is defined by the accumulation of the difference between a 5 order polynomial fitting for recorded time-velocity data and Furusawa's theoretical velocity. The latter is given by an exponential function whose parameters are roughly estimated based on data. In the analysis of [6], it has been clarified that a sprint time is related to a combination of the maximum velocity and ADV. In this method for identifying velocity change, the validity of polynomial fitting is questionable.

In this paper, we consider a curve fitting of noisy distance-velocity data. This can be viewed as a nonlinear regression problem. Especially, we consider to model a distance-velocity property by taking account of the above mentioned three phases. More specifically, our model is a parametric nonlinear regression model that reflects the property of a 100m sprint. As a pre-processing for the curve fitting, we apply a moving average for the velocity data which is obtained by a simple numerical differentiation. However, it is just moderately smoothed so as to keep the mean shape of a distance-velocity course. Therefore, the number of points for a moving average is not needed to be determined precisely but is roughly set to a small value. This is because we fit the pre-processed noisy velocity data by a pre-determined nonlinear function with parameters; i.e. the smoothed distance-velocity curve is represented by the known nonlinear function in the context of a nonlinear regression problem. The function has two components which bear the early and later parts of a sprint respectively. The rapid increase at a sprint start is represented by an exponential function as in Furusawa's theory and the gradual decrease at the later stage is represented by a polynomial function. Therefore, our model can be viewed as an extension of analysis in [6]. The important difference is that our regression model is structured to represent the known phases of an entire 100m dash, by which we capture many characteristics of a 100m sprint other than velocity decrease as in [6]. The model parameters are estimated by the gradient descent method for the least squares estimation. We also give some analysis of 100m dash based on the estimated distance-velocity curve and/or the estimated parameter. This paper is organized as follows. In the section 2, we describe data, pre-processing of data and a considered model. In the section 3, we give details of the parameter estimation for our model defined in the section 2. In the section 4, we show the fitting results and some analysis of the 100m sprint based on the estimated distance-velocity curve and/or the parameter. The section 5 is devoted to conclusions and future works.

2 Nonlinear regression for sprint data

2.1 Data description

For a 100m sprint, positions during the dash are recorded per 0.01[s] by using the laser system (LAVEG)[5]. The subjects participated in this record are male and female students in elementary and junior high school; i.e. the range of ages of subjects is 9–15. The number of total subjects is 109. The number of subjects for each sex and grade in school is shown in Table.1, in which, for example, E4 denotes the 4th grade of elementary school and J1 denotes the 1st grade of junior high school. Here, we only have an information on grade in school and do not have an exact information on age.

The position data is calibrated to be 0[m] at a start line; i.e. it is therefore distance data from a start line. It includes extra measurement data before start and after goal; i.e. it includes minus values and values that are greater than 100. Therefore, as a pre-processing, we need to pick up a portion of a 100m sprint from this raw data. The extracted data are denoted by x_1, \dots, x_n , in which x_1 is greater than and nearest to 0[m] and x_n is greater than and nearest to 100[m]. Note that x_1 and x_n are possibly not equal to 0[m] and 100[m] respectively. Note also that the number of data is different for each student. The minimum value, 1st quantile, median, mean, 3rd quantile and maximum value of the number of data are 1215, 1414, 1533, 1547, 1656 and 2143 respectively. By using the number of data, we can calculate

a sprint time by $0.01n[s]$ since the measurement interval is $0.01[s]$. The minimum value, 1st quantile, median, mean, 3rd quantile and maximum value of sprint times are 12.15, 14.14, 15.33, 15.47, 16.56 and 21.43 respectively.

Let $z_i, i = 1, \dots, n$ be an instantaneous velocity that is calculated by $z_i = (x_i - x_{i-1})/0.01$, where we set $x_0 = 0$ approximately. This is a simple numerical differentiation. We then have $\{(x_i, z_i), i = 1, \dots, n\}$ as data for a distance-velocity curve. x_i may include measurement errors and such errors may be expanded in calculating instantaneous speeds by using a difference approximation. This leads to a large variance of z_1, \dots, z_n . Therefore, we consider to smooth those by applying a local averaging. Let M be a positive integer that determine an interval for averaging; i.e. width of window for averaging. We define y_i by the mean of $z_{i-M}, \dots, z_i, \dots, z_{i+M}$, in which we exclude z_k for $k < 1$ and $k > n$. We then have $\{(x_i, y_i), i = 1, \dots, n\}$ as data for a distance-velocity curve that characterizes a 100m sprint. If M is small then the calculated velocity is very noisy. Conversely, if M is large then data is squashed and is close to a constant function. The purpose of smoothing here is to suppress a large variance of velocity data and a smooth distance-velocity curve is identified by using curve fitting. Therefore, M is not needed to be precisely determined if it is moderate for suppressing noisy data. In this reason, we choose $M = 5$ throughout this paper. The example of data for $M = 5$ is shown by gray dots in Figure.1.

2.2 Model description

In this paper, we consider to estimate a distance-velocity curve under a nonlinear regression model. This is a curve-fitting problem using a nonlinear function. For the purpose, we first discuss the choice of the nonlinear function here.

As argued in [1], a 100m sprint is mainly divided into the three phases that are acceleration phase, maximal speed phase and speed maintenance phase; e.g. see [2] for more precise categorization. These phases correspond to early, middle and later stage of a 100m dash. Existence of these phases are actually found in our data shown in Figure.1, in which velocity may steeply increase in the early to middle part of sprint and almost flat or gradually decrease in the middle to later part. This trend is common for all data while, of course, the details are different depending on the subject sprinter. We then formulate a distance-velocity curve by the sum of two functions that represent the early to middle and the middle to later part of a 100m sprint respectively. More precisely, for a position x , our nonlinear regression function is defined by

$$f_{\mathbf{a}}(x) = g_{\mathbf{a}_1}(x) + g_{\mathbf{a}_2}(x), \quad (1)$$

$$g_{\mathbf{a}_1}(x) = a_{13} \left[\tanh \left(\frac{a_{11}x}{100} \right) \right]^{a_{12}}, \quad (2)$$

$$g_{\mathbf{a}_2}(x) = a_{22} \left(\frac{x}{100} \right)^{a_{21}} \quad (3)$$

where $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$ is a parameter vector in which $\mathbf{a}_1 = (a_{11}, a_{12}, a_{13})$ and $\mathbf{a}_2 = (a_{21}, a_{22})$. Under an appropriate choice of \mathbf{a}_1 , $g_{\mathbf{a}_1}$ is a monotone increasing function with steep change around $x \simeq 0$ and saturates for a large x . On the other hand, under an appropriate choice of \mathbf{a}_2 , $g_{\mathbf{a}_2}$ is a monotone decreasing function with a gradual change. $g_{\mathbf{a}_1}$ and $g_{\mathbf{a}_2}$ is therefore expected to represent velocity changes at the early to middle and the middle to later part of a sprint respectively. Note that, in this representation, interaction between both functions may be small; i.e. $g_{\mathbf{a}_2}$ shows small change in the early stage and $g_{\mathbf{a}_1}$ is almost constant in the later stage. We will see later that this is true in actual fitting of data. According

Table. 1: The number of subjects for each sex and age (grade in school) group.

Grade (Age)	E4 (9–10)	E5 (10–11)	E6 (11–12)	J1 (12–13)	J2 (13–14)	J3 (14–15)
Female	5	4	7	16	15	7
Male	4	2	7	12	14	16

to the known categorization of a sprint phase, g_{a_1} mainly represents a velocity change from acceleration phase to maximal speed phase and g_{a_2} mainly represents a velocity change from maximal speed phase to maintenance phase. The important point is that the entire velocity change is represented by sum of these two essential components that play own parts. In other words, distance-velocity curve is decomposed into two meaningful components. It is worthy to note that steep increase in the early part is represented by an exponential type function as in (2) and gradual decrease in the later part is represented by a polynomial type function as in (3). Also, we employ a linear and nonlinear parameters in both of g_{a_1} and g_{a_2} . These are needed for a fine tuning of a distance-velocity curve at the early to middle and the middle to later parts respectively. In other words, those are used for avoiding biased regression curve in estimation. In g_{a_1} and g_{a_2} , the position, x , is divided by 100 for a normalization. An example of estimated g_{a_1} , g_{a_2} and f_a are shown in Figure.1, in which g_{a_1} and g_{a_2} are found to be mainly used for fitting data in the early and later part respectively as desired. The details of parameter estimation used for obtaining the curves in Figure.1 will be presented in the next section.

We address some remarks on parameters of (1).

- a_1 is related to a velocity increase in the early to middle parts of sprint. a_{13} is the hight of g_{a_1} and represents his/her ideal maximum speed. It decays during a 100m dash due to physical/mental limitations that is represented by g_{a_2} in our model. The maximum velocity given by $\max_x f_a(x)$ is determined by the balance between these two functions. Since the decay due to g_{a_2} may be small in the early part of sprint, a_{13} may well approximate the maximum velocity that is known to be related to an over all sprint time[1, 3]. $a_{11}(> 0)$ and $a_{12}(> 0)$ determine a slope shape of g_{a_1} and is expected to characterize a velocity increase in the early part of sprint. If a_{11} is large and/or a_{12} is small then the slope is large, thus, g_{a_1} steeply approaches to 1. Note that this does not necessarily imply a high acceleration after a sprint start. It is also related to a_{13} in addition to a_{11} and a_{12} .
- $a_{22}(< 0)$ determines a degree of velocity decrease since $g_{a_2}(x) = a_{22}$ for $x = 100$. On the other hand, $a_{21}(> 0)$ controls a slope shape of g_{a_2} . Both of a_{21} and a_{22} are related to a velocity decrease in the middle to later parts of sprint, in which there is an interaction between them. For example, if a_{21} and a_{22} are moderate values then g_{a_2} gradually decreases. If a_{21} is large under a fixed a_{22} then the slope of decrease of g_{a_2} is small. On the other hand, if a_{22} is large under a fixed $a_{21} \neq 0$ then the slope of decrease of g_{a_2} is large.

3 Parameter estimation

We here consider to apply the least squares method for estimating a distance-velocity curve by (1). The mean of squared errors is defined by

$$S_w(\mathbf{a}) = \frac{1}{n} \sum_{i=1}^n e_i^2(\mathbf{a}) \quad (4)$$

$$e_i(\mathbf{a}) = y_i - f_a(x_i). \quad (5)$$

For minimizing (4), we usually apply an iterative optimization algorithm started with appropriate initial values for parameters. The most basic and simple non-constrained optimization method for minimizing (4) is the gradient descent method which is an incremental update procedure of a parameter vector[4]. We define a parameter vector at the k th step in the procedure by $\mathbf{a}(k) = (a_{11}(k), a_{12}(k), a_{13}(k), a_{21}(k), a_{22}(k))$. The gradient descent method starts from a randomly chosen initial vector $\mathbf{a}(0)$ and repeats

$$\mathbf{a}(k) = \mathbf{a}(k-1) - \eta \Delta \mathbf{a}(k-1) \quad (6)$$

$$\Delta \mathbf{a}(k-1) = \left. \frac{\partial S_w(\mathbf{a})}{\partial \mathbf{a}} \right|_{\mathbf{a}=\mathbf{a}(k-1)} \quad (7)$$

for $k = 1, \dots, K$, where K is the number of repetitions, η is known as a step size and

$$\frac{\partial S_w(\mathbf{a})}{\partial \mathbf{a}} = \left(\frac{\partial S_w(\mathbf{a})}{\partial a_{11}}, \frac{\partial S_w(\mathbf{a})}{\partial a_{12}}, \frac{\partial S_w(\mathbf{a})}{\partial a_{13}}, \frac{\partial S_w(\mathbf{a})}{\partial a_{21}}, \frac{\partial S_w(\mathbf{a})}{\partial a_{22}} \right). \quad (8)$$

By (1), (2) and (3), $\partial S_w(\mathbf{a})/\partial \mathbf{a}$ can be calculated as

$$\frac{\partial S_w(\mathbf{a})}{\partial a_{11}} = -\frac{2a_{13}a_{12}}{n} \sum_{i=1}^n e_i u_i \tanh^{a_{12}-1}(a_{11}u_i) \frac{1}{\cosh^2(a_{11}u_i)} \quad (9)$$

$$\frac{\partial S_w(\mathbf{a})}{\partial a_{12}} = -\frac{2a_{13}}{n} \sum_{i=1}^n e_i g_{a_1}(x_i) \log(\tanh(a_{11}u_i)) \quad (10)$$

$$\frac{\partial S_w(\mathbf{a})}{\partial a_{13}} = -\frac{2}{n} \sum_{i=1}^n e_i g_{a_1}(x_i) \quad (11)$$

$$\frac{\partial S_w(\mathbf{a})}{\partial a_{21}} = -\frac{2}{n} \sum_{i=1}^n e_i g_{a_2}(x_i) \log(u_i) \quad (12)$$

$$\frac{\partial S_w(\mathbf{a})}{\partial a_{22}} = -\frac{2}{n} \sum_{i=1}^n e_i g_{a_2}(x_i) \quad (13)$$

where $u_i = x_i/100$. The resulting estimate of \mathbf{a} , denoted by $\hat{\mathbf{a}}$, is obtained by $\hat{\mathbf{a}} = \mathbf{a}(K)$, where $\hat{\mathbf{a}} = (\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2)$, $\hat{\mathbf{a}}_1 = (\hat{a}_{11}, \hat{a}_{12}, \hat{a}_{13})$ and $\hat{\mathbf{a}}_2 = (\hat{a}_{21}, \hat{a}_{22})$.

The resulting function $f_{\hat{\mathbf{a}}}$ may well capture a distance-velocity curve buried in data by the gradient descent method while it may not be guaranteed that the resulting $g_{\hat{\mathbf{a}}_1}$ and $g_{\hat{\mathbf{a}}_2}$ are respectively estimated to represent curves of the early to middle part and the middle to later part of sprint; i.e. to play their proper roles. Fortunately, we can expect and, indeed, have experimentally confirmed that a parameter vector converges to a desired solution when we set initial parameter values roughly around a desired solution. It is not difficult since shapes of underlying distance-velocity curve of all subjects are roughly similar but are different in details. Since the degree of freedom of our nonlinear model is small compared with the number of data and the gradient descent method is not so powerful method, a large variation of model output may not occur in iteration if model output have already captures a rough shape of data. We here do not need a global minimum solution that strictly minimizes $S_w(\mathbf{a})$ but do need a local minimum solution at which the resulting model is interpretable; i.e. $g_{\hat{\mathbf{a}}_1}$ and $g_{\hat{\mathbf{a}}_2}$ are respectively responsible for representing the early to middle and the middle to later part of sprint. Fortunately, in the next section, we can see that this desired solution is actually obtained by using the gradient descent method given in (6) and (7).

4 Results

4.1 Results of parameter estimation

In the parameter estimation, we set $\mathbf{a}(0) = (5, 1.5, 7, 1, -1)$ as an initial vector, $K = 5000$ as the number of repetitions and $\eta = 0.01$ as a step size. Note that $\mathbf{a}(0)$ is set to be a rough approximation of a distance-velocity curve buried in data. The transition of the residual mean square errors (RMSE) for all subjects during the repetitions of the gradient descent method is depicted in Figure.2, in which the vertical axis is shown in log scale. In all cases, the errors are smoothly converged until the maximum iteration. Note that the solution obtained here may not be a global minimum and may be a local minimum. However, the important point is that the solution gives a relatively good fitting to data and is interpretable in analyzing a 100m sprint. It should be note that, in the parameter estimation, $a_{11} > 0$, $a_{12} > 0$, $a_{13} > 0$, $a_{21} > 0$ and $a_{22} < 0$ were satisfied for all subjects. In Figure.1, we show an example of estimated curve (thick line) with two estimated components (thin line); i.e. g_{a_1} for velocity increase in the early part and g_{a_2} for velocity decrease in the later part. In Figure.1, data are also shown by the gray dots. We can find that data are well fitted by the resulting curve, in which the two components are properly estimated to play their own roles. As important examples, we show data and estimated curves for the fastest and slowest sprinters in Figure.3.

For checking a degree of fitting, we calculate the multiple correlation coefficients (MCCs) between fitted values and velocity data for all subjects. The minimum value, 1st quantile, median, mean, 3rd quantile and maximum value of MCCs are 0.806, 0.920, 0.944, 0.933, 0.955 and 0.971 respectively. Since

the smallest MCCs is 0.806 that is sufficiently large, this confirms that the estimated curve fits data well for almost all subjects. In Figure.4, we show estimated distance-velocity curves that give the smallest and largest MCCs. The curves apparently fit data well even for the smallest MCC case. For the smallest MCC case, we can see an accidental change at the middle part of sprint in data. This is a reason for an avoidable fitting error while the MCC is still relatively large; i.e. it is 0.806. Note that we may not need to take account of over-fitting here since the number of data is sufficiently larger than the number of adjustable model parameters; i.e. the number of parameters is five and the number of data is larger than a thousand. It may be important to note that the correlation coefficient between MCC and sprint time is -0.8 . Therefore, data of a subject with a better sprint time can be well fitted by our model. In other words, our model is more suitable for relatively fast sprinters. Slow sprinters may tend to show somewhat irregular pattern as seen in Figure.3 (b) and also Figure.4 (b). This pattern cannot be captured by our model. As seen in later, subjects in elementary school naturally tend to set a slow sprint time and it may be difficult to expect a regular (well-modeled) sprint for some subjects.

4.2 On estimated distance-velocity curve

In Figure.5, we show scatter plots between a sprint time and some sprint characteristics that can be calculated from estimated curves.

- In Figure.5 (a), we show the scatter plot between sprint time and maximum velocity that is approximately calculated by a maximum value in $\{f_{\hat{a}}(u) : u \in U\}$, where \hat{a} is the obtained estimator and $U = \{0, 0.1, 0.2, \dots, 100\}$; i.e. a sequence from 0 to 100 by $0.1[m]$. This may be an acceptable approximation since $f_{\hat{a}}$ is a smooth function and U is sufficiently dense in $[0, 100]$. In Figure.5 (a), we can see that the maximum velocity is highly correlated with a sprint time. This result says that a sprinter with high maximum velocity tends to set a better sprint time. This fact is well known in general; see e.g. [1, 3].
- A position at which the maximum velocity is achieved is called by velocity maximizing position here. It is approximately obtained as u^* at which $f_{\hat{a}}(u)$ is maximized on U ; i.e. $\max_{u \in U} f_{\hat{a}}(u) = f_{\hat{a}}(u^*)$. In Figure.5 (b), we show the scatter plot between sprint time and velocity maximizing position. We can see that correlation between them is high. Indeed, the correlation coefficient between the velocity maximizing position and the maximum velocity is 0.73. This implies that, as a natural consequence, it takes a long distance to achieve the maximum velocity when it is high. This can be apparently understood by comparing the fastest and slowest sprinter's curves in Figure.3. Note that if the velocity maximizing position is large then the period at which acceleration is positive is long. Therefore, the period of positive acceleration may be related to a sprint time.
- We define the velocity decrease by the mean of $g_{\hat{a}_2}(u)$ for $u \in U$. This is closely related to ADV in [6]. The velocity decrease represents a degree of speed down mainly in the later part of sprint. It is easily understood that the subject with small absolute value of the velocity decrease possibly produces a better sprint time. In Figure.5 (c), we show the scatter plot between sprint time and velocity decrease. We can find that the velocity decreases for all subjects take negative values. It implies that a_2 is successfully estimated to play its role in characterizing a dash; i.e. to represent the velocity decrease in the later part of sprint. Basically, there is a large negative correlation between averaged velocity decrease and sprint time. This is notable for sprinters with high velocity decrease and poor sprint time while this does not apply for fast sprinters. This point will be discussed in a later section.
- Let i^* be a data index at which estimated velocity is maximized; i.e. $f_{\hat{a}}(x_{i^*}) = \max_{1 \leq i \leq n} f_{\hat{a}}(x_i)$. Of course, it is close to the maximum velocity calculated above. We consider to approximately calculate the acceleration at x_i by $(f_{\hat{a}}(x_i) - f_{\hat{a}}(x_{i-1}))/0.01$ for $i = 2, \dots, n$ since measurement interval is $0.01[s]$. Then the average of approximate acceleration up to achieving the maximum velocity on data points is given by $(f_{\hat{a}}(x_{i^*}) - f_{\hat{a}}(x_1))/(0.01(i^* - 1))$. This represents an averaged acceleration at early part of sprint and may especially determine a start performance. We call this value by start acceleration. In Figure.5 (d), we show the scatter plot between sprint time and

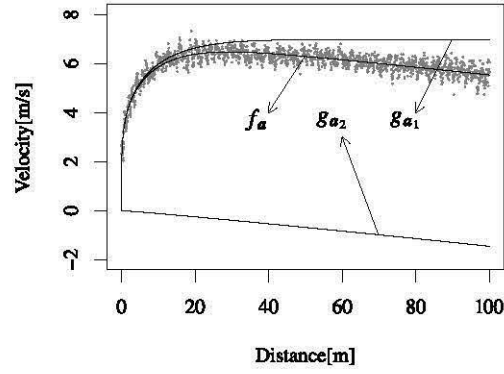


Figure. 1: Example of data and estimated distance-velocity curve, where gray dots, solid thick line and two solid thin lines show data, f_a and two functions that are g_{a_1} and g_{a_2} respectively.

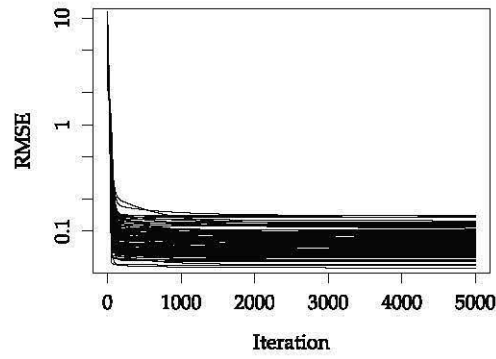


Figure. 2: Transition of the residual mean square error for all subjects during the repetition of the gradient descent estimation.

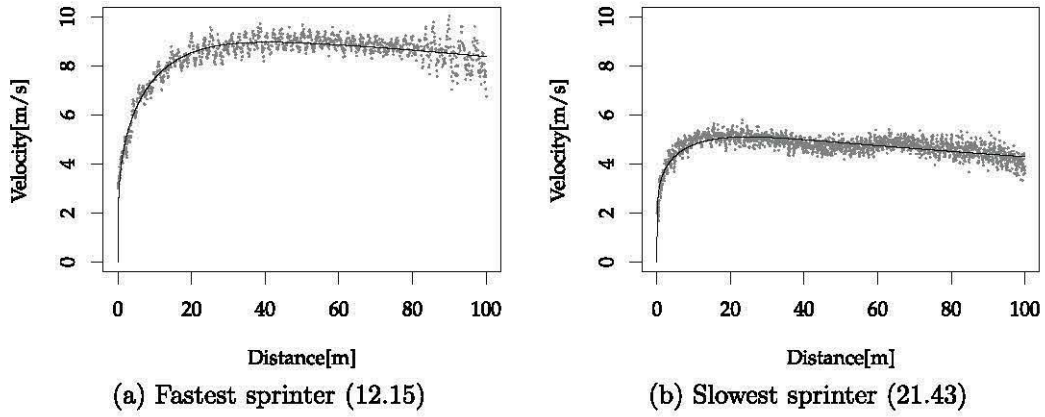


Figure. 3: Estimated distance-velocity curve for the fastest and slowest sprinter. The value in the bracket is sprint time in seconds.

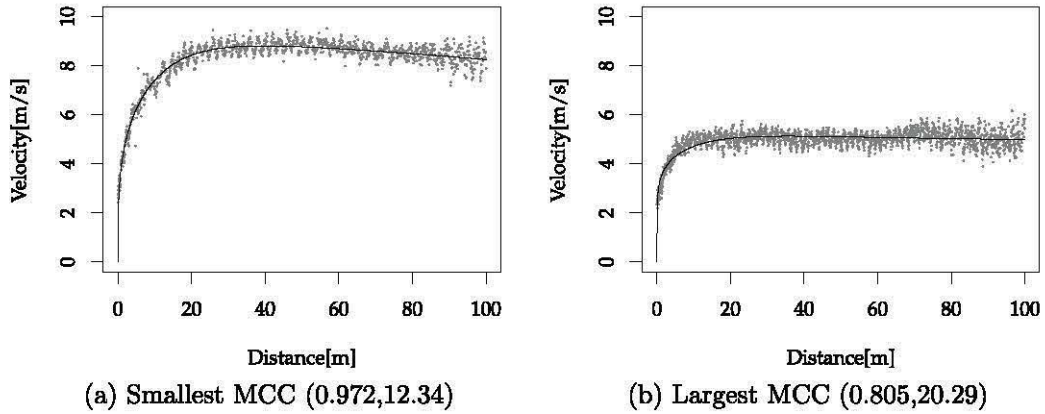


Figure. 4: Estimated distance-velocity curves that give the smallest and largest MCC. The paired values in the bracket are MCC and sprint time in seconds respectively.

start acceleration. As seen in Figure.5 (d), the correlation between the sprint time and the start acceleration is high. Therefore, we can say that an acceleration at early part of sprint affects overall sprint time, or, a sprinter with a better sprint time shows a good performance at the sprint start.

- On the other hand, by (1), we can calculate a derivative of a distance-velocity curve by

$$f'_a(x) = \frac{a_{11}a_{12}a_{13}}{100} \frac{\tanh^{-1+a_{12}}(a_{11}x/100)}{\cosh^2(a_{11}x/100)} + \frac{a_{21}a_{22}}{100} (x/100)^{-1+a_{21}}. \quad (14)$$

This is a velocity change rate in terms of a position change. Since we have

$$\frac{df_a(x(t))}{dt} = \frac{df_a(x(t))}{dx(t)} \frac{dx(t)}{dt}, \quad (15)$$

(14) is a time depending link coefficient between velocity and acceleration. This can be a characteristic of a sprint. We call $f'_a(x)$ velocity change rate at x . In Figure.6 (a), we show correlation coefficients between sprint time and $f'_a(u)$ at each $u \in U$; i.e. at each fixed $u \in U$, correlation coefficient between sprint times and values of $f'_a(u)$ for all subjects is calculated. We can see that the correlation coefficients near the sprint start is very close to -1 . Since the velocity monotonically increases for a while after the sprint start, we define start velocity increase rate by $f'_a(u)$ at $u = 0.1[m]$ which is a position just after a sprint start. The start velocity increase rate naturally represents a quality of sprint start. In Figure.6 (b), we show a scatter plot between sprint time and start velocity increase rate. We can see that the start velocity increase rate actually highly correlated with a sprint time. A large start velocity increase rate implies a good performance of start dash. Therefore, this result says that high quality of start dash produces a better sprint time as expected.

The important point here is that our nonlinear regression model capture a smooth distance-velocity curve well in which its properties are highly related to a sprint performance and are interpretable.

4.3 On estimates of parameters

In Table.2, we show the mean and standard deviation of the estimates of parameters. In the table, we also show the correlation coefficients between estimated parameters and a sprint time.

- As seen in Table.2, a_{13} is found to be the most correlated parameter to sprint time, in which correlation coefficient is very close to -1 . a_{13} is a high of g_{a_1} and, thus, approximates the maximum velocity since the influence of g_{a_2} upon g_{a_1} is small in the early part of a dash. Therefore, we can say that a_{13} represents a characteristics of a 100m sprint as the maximum velocity does.
- a_{12} nonlinearly controls a slope of velocity increase just after a sprint start. If a_{12} is small then the velocity increase tends to be steep since $0 < \tanh(u) < 1$ for $u > 0$. As seen in Table.2, correlation between a_{12} and sprint time is relatively high. This implies that a slope of velocity increase near the sprint start is not steep for a subject with a better sprint time. Note that this does not imply a small start acceleration for such subject since start acceleration is determined by all of a_{11} , a_{12} and a_{13} . Actually, large acceleration at early stage of sprint produces a better sprint time as seen in Figure.5 (d). If a_{12} is small and a_{13} is large then the velocity increase is steep and the maximum velocity is large. This may be an ideal sprint. However, Table.2 says that this is difficult in general; i.e. a velocity increase after sprint start tends to be mild for a fast sprinter. This may be due to physical constraints.

On the other hand, correlation between a_{11} and sprint time is low. Therefore, a_{11} may not be essential for a sprint time while it contributes as a constant for determining a basic slope shape common for all sprinters.

- Both of a_{21} and a_{22} control velocity decrease from the middle to later part of sprint. As in Table.2, these parameters are well correlated with sprint time. If a_{21} is large then a velocity decrease is

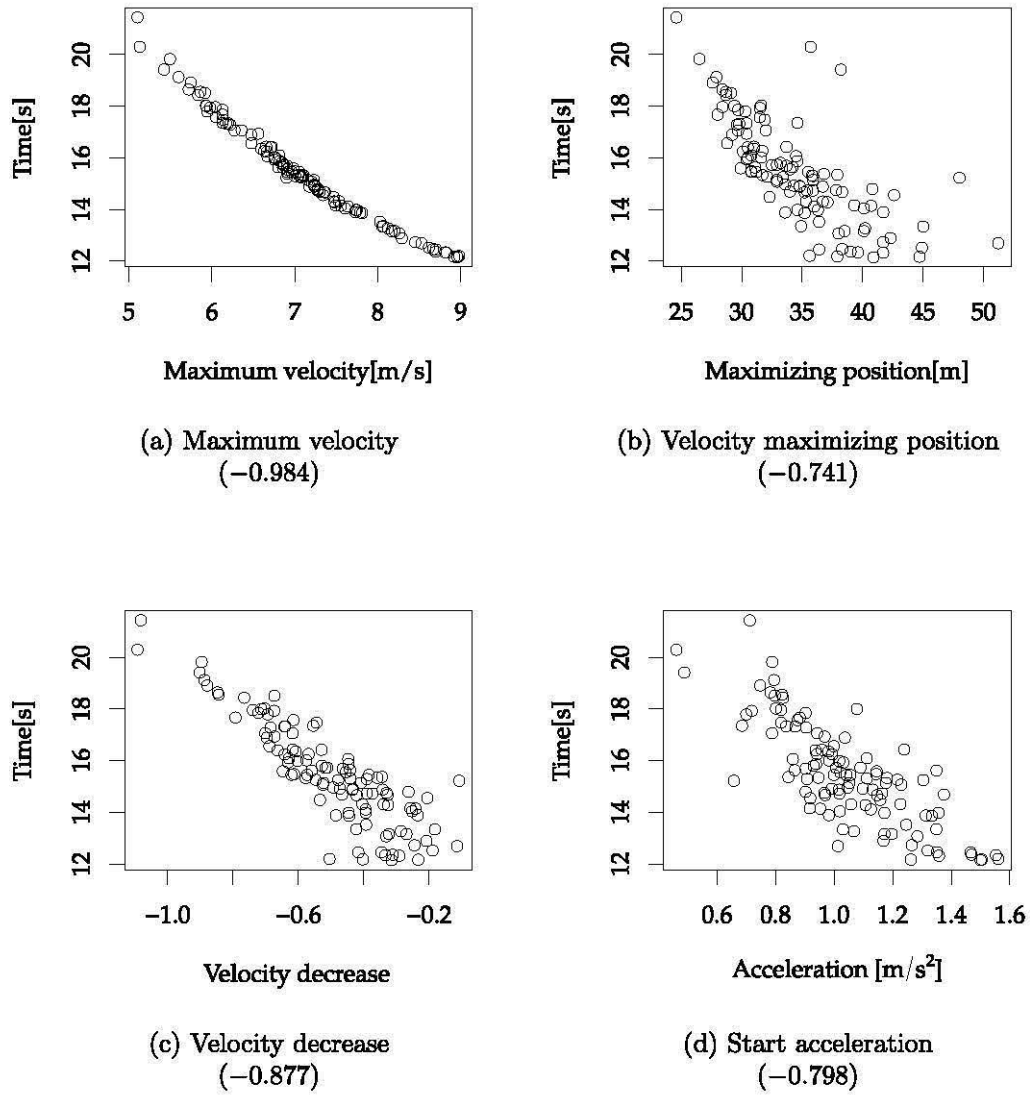


Figure. 5: Scatter plot between 100m sprint time and the characteristics of distance-velocity curve.

Table. 2: The statistics related to estimates of parameters.

	a_{11}	a_{12}	a_{13}	a_{21}	a_{22}
Mean	5.330	0.232	7.508	1.090	-1.010
S.D.	0.131	0.024	0.765	0.302	0.306
Correlation Coef.	-0.08	-0.81	-0.961	-0.839	-0.666

small until 100m. This may cause a better sprint time and correlation coefficient is actually high in the minus direction. On the other hand, if the absolute value of a_{22} is small then, again, a velocity decrease is small. This may also cause a better sprint time and correlation coefficient is actually high in the minus direction again. These results for a_{12} and a_{22} correspond to the result for the velocity decrease shown in Figure.5 (c).

4.4 PCA for estimates of parameters

We apply PCA(principal component analysis) for estimates of the parameter vectors \mathbf{a} except a_{11} which is not essentially related to a sprint time. The cumulative variances of the 1st, 2nd, 3rd and 4th principal components are 0.737, 0.929, 0.968 and 1.000 respectively, in which the first two principal components may be found to be sufficient for explaining the given data. The factor loadings are shown in Table.3, in which, for example, PC1 indicates the first principal component. In the table, we can see that the first principal component(PC1) relatively relates to a_{12} and a_{13} which determine g_{a_1} , thus, the properties of the early to middle part of sprint. By the factor loadings, there are positive correlations between PC1 and the both estimates. This implies that a_{12} and a_{13} are large if PC1 score is large. Therefore, for a subject with a large PC1, the maximum velocity is large and velocity increase after the sprint start is steep. Thus, a subject with a large PC1 score tends to show a better performance at the early to middle phase of sprint. That may result in a better sprint time. On the other hand, by Table.3, the second principal component(PC2) mainly relates to a_{21} and a_{22} which determine g_{a_2} , thus, the properties of velocity decrease in the middle to later part of sprint. We know that if a_{21} is large and/or $|a_{22}|$ is small then velocity decrease is small. Therefore, velocity decrease is small for a sprinter with a large PC2 score.

In Figure.7 (a), we show the score plot of PCA, in which a number showing the order of sprint time is plotted at the corresponding coordinate. In the figure, we can see that a sprint time strongly depends on the first principal component; i.e. score of PC1 is large for a subject with a better sprint time. This is consistent with the above discussion based on the factor loadings of PC1. On the other hand, in Figure.7 (b), we show estimated curves for the 11th fastest subject(dotted line) and 4th fastest subject(solid line). As easily found, the 4th fastest subject exhibits a better performance at the early stage of sprint but a poor performance at the later stage compared to the 11th fastest subject. Note that, of course, the 4th fastest subject shows a better total performance. A negative value for the score of PC2 implies a large velocity decrease as discussed above based on factor loadings of PC2. If we restrict our attention to the top 10 sprinter, velocity decrease is relatively large. Therefore, they tend to show excellent performances at the early to middle stage of sprint but do poor performance at the middle to later stage as an example of the 4th fastest subject in Figure.7 (b). However, for sprinters with low PC1 scores, negative PC2

Table. 3: Factor loadings of PCA for estimates of parameters.

	a_{12}	a_{13}	a_{21}	a_{22}
PC1	0.9235	0.9529	0.9074	0.6024
PC2	-0.2178	0.0004	-0.3055	0.7934
PC3	0.3155	-0.1305	-0.1924	0.0126
PC4	0.0144	-0.2739	0.2152	0.0870

scores tend to imply a worse total sprint time; i.e. the order according to sprint time tends to be larger for subject with a negative score of PC2. Therefore, we can say that, for slow sprinters, velocity decrease may be an essential factor for a total sprint time. This also found in Figure.5 (c), in which sprint time tends to be correlated to velocity decrease especially for slow sprinters. Note that most of slow sprinters belong to a low age group. For those subjects, velocity decrease may be notably caused by consumption or fatigue.

5 Conclusions and future works

In this paper, we proposed a nonlinear regression model for estimating a distance-velocity curve of a 100m sprint. The nonlinear function in our model is composed of a sum of two functions which represent a velocity increase in the early stage of 100m sprint and a velocity decrease in the later stage respectively. The former function is an exponential function to represent a rapid speed increase at a sprint start and the latter function is a polynomial function to represent a gradual decrease due to fatigue in a later stage of sprint. We apply this model to analyze collected LABEG data of students in elementary and junior high school under an appropriate pre-processing such as a moving average. The model parameters was estimated by the gradient descent method for the least squares estimation. As a result, we confirmed that this model can well represent the distance-velocity curve of the collected data. By the analysis based on the estimated curves, we clarified that the maximum velocity is the most important factor for determining a 100m sprint time. This is consistent with the well known speculation; e.g. [1, 3]. We also found that acceleration in the early stage and velocity decrease in the later stage is also related to the 100m sprint time. Furthermore, we introduced a derivative of the estimated velocity curve and found that it's value at just after a sprint start characterizes a sprint time well. Fortunately, we confirmed that the similar insights are obtained also by estimated parameter values. This is because the parameters in our modeling reflect the property of each phase of a 100m sprint directly. In the next challenge, we will apply our nonlinear regression method for a top athlete. We will also consider the development of a method for improving an athlete's performance by applying the analysis based on our nonlinear regression model.

References

- [1] M.Misjuk, M.Viru, "Running velocity dynamics in 100m sprint: comparative analysis of the world top and estonian top male sprinters", *Acta Kinesiologiae Universitatis Taruensis*, 17, 131-138, 2011.
- [2] R.Saunders, "Five components of the 100m sprint. *Modern Athlete and Coach*, 42, 1-3, 2004.
- [3] L.Cunha, "The relation between different phases of sprint run and specific strength parameters of lower limbs", *Proceedings of XXIII International Symposium on Biomechanics in Sports*, Beijing, China, 183-186, 2005.
- [4] T.Hastie, R.Tibshirani, J.Friedman, "The elements of statistical learning", Springer, 2001.
- [5] H.Kintaka, "The examination of processing of raw kinematic data form measuring of velocity courses in the sprint events with the use of the laser velocity measuring device", *The Bulletin of National Institute of Fitness and Sports in Kanoya*, 20, 9-108, 1999 (in Japanese).
- [6] K.Takahashi, T.Kurokawa, S.Usui, "The new index of velocity change in the 100m sprint:in terms of the amount of velocity deviation", *Japan Journal of Physical Education, Health and Sport Sciences*, 50, 189-200, 2005 (in Japanese).
- [7] K.Sugiyama, Y.Murata, H.Watanabe, K.Takayuki, M.Iwase, "Factor Analysis of Sprint Phases of the Speed Curve of the 100m Dash", 18th International Symposium on Biomechanics in Sports Conference, Hong Kong, China, June 25 30, 2000.

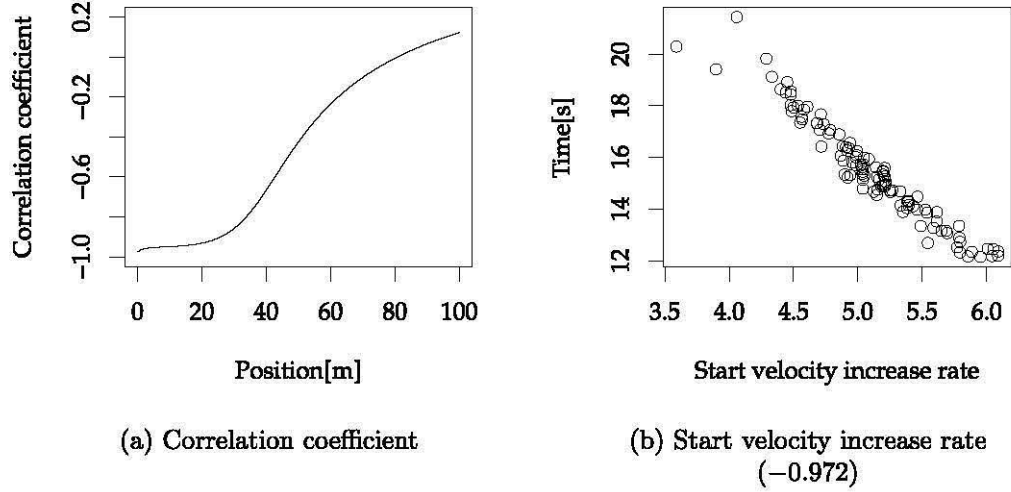


Figure. 6: Velocity change rate. (a) Transition of correlation coefficient between sprint time and velocity change rate. (b) Start velocity increase rate. The value in the bracket is correlation coefficient with sprint time.

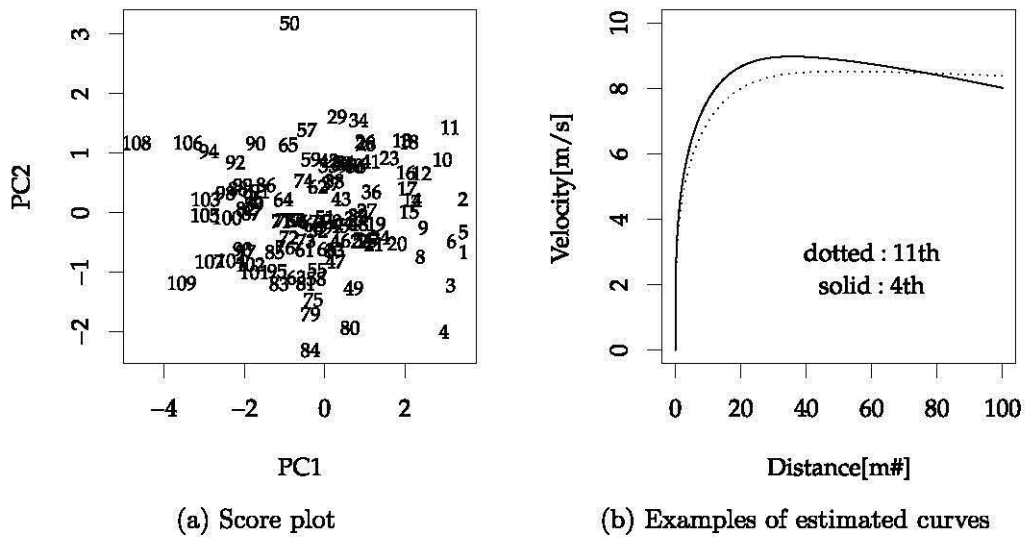


Figure. 7: Analysis based on PCA for estimates of parameters. (a) Score plot in which a number in descending order of sprint time is plotted at the coordinates with respect to scores of PC1 and PC2. (b) Examples of estimated curves for the 4th and 11th fastest sprinter.