

# Formulation of Sprint Time Predictive Model for Olympic Athletic Games

**John E. Efiog**

Department of Computer Science, Wesley University, Ondo, Nigeria  
E-mail: john.efiog@wesleyuni.edu.ng

**Emmanuel A. Olajubu**

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria  
E-mail: emmolajubu@oauife.edu.ng

**Felix O. Aranuwa**

Department of Computer Science, Adekunle Ajasin University, Akungba – Akoko, Ondo State, Nigeria  
E-mail: felix.aranuwa@aaua.edu.ng

Received: 18 December 2018; Accepted: 20 January 2019; Published: 08 April 2019

**Abstract**—Olympic Games are international field and track events hosted within four years periods. Like other events, sprinting is a track event that requires rigorous and focused training. When training is done with little or no understanding of the possibilities of the games, the competition would leave more to be desired. This paper formulates, evaluates and validates a model for predicting the fastest sprinting time of Olympic athletes of 100m race for a-5 season appearances. Dataset was obtained from the Olympic official records of world best performances, typically Gold medalists in sprint for the male category from the inception in 1896 to the 2016 edition. The model was simulated on MATLAB. Cross-validation was done using residuals for whiteness and independence tests and model outputs. The results were evaluated based on Sum of Square Error (SSE), R-Square, adjusted R-Square, and Root Mean Square Error (RMSE) and benchmarked with existing models. The model outperformed the existing models with higher accuracy and goodness of fit. This prediction is a reasonable guide for predictive training, forecasting and future study on predictive algorithms.

**Index Terms**—Olympics, Predictive model, Sprint, Track, Gold Medalists.

## I. INTRODUCTION

Olympic Games are made up of 44 events in tracks and field which are divided into separate areas. The track events include running and walking while field events are throwing and vaulting events. A variety of other events, including decathlon and heptathlon also exist which make Olympic Games the most appealing events in the world history of sports. Sprinting is a track event that takes place within a short length of track, in this paper, 100 meters. A track is an oval-shaped path with semi-circles

at both ends and two straight areas linking the semi-circles together [1]. A standard track is built with a rubber surface, technically chosen to reduce slipping in wet weather conditions. Track events are of different ranges which should be covered within the earliest time possible for one to be declared a winner.

Many factors are suggestive of why some athletes consistently outperform others in competitions. References [2,3] opine that such exceptional qualities are hereditary and flow from genes, phenotypes and biological traits. Others like [4] hang their opinion on superior coaching and thorough training. Reference [5] asserts that speed capacities can be worked and improved upon by adopting science-based training methods. Such proven methods would help athletes overcome apparent challenges like high-speed running, speed endurance, strength and power, flexibility, neuro-muscular programming and mental preparation.

Subjecting oneself to rigorous training is what every sporting competition requires and every athlete knows this for a fact. Some coaches believe that good sprinters are born and not made, especially those who have the capacity to overcome the speed barrier – a situation where the athlete finds it extremely difficult to increase his or her running velocity irrespective of the number, intensity and quality of training sessions or level of coaching [5].

However, a vast majority of the trainers accept that training for the enhancement of inborn traits for gifted sprinters and the not-so-gifted sprinters cannot be overruled. They argue that the training must be deliberate and consistent in order for the athlete to garner strength and develop skills necessary for winning in the actual competitions. Technically, for sprint events, competitors will be expected to train having the maximal attainable velocity in mind and aiming at such – typically 20m-80m and 150m-300m distances for short and long races respectively. According to [6], to acquire mastery or earn

reasonable improvement towards the highest point, this training requires repetitive exercises and sessions. The most prominent of the exercises is high-velocity sprinting. There is, however, a psychological concern that repetitive actions of same exercises can result in dynamic stereotype problem, fatigue or retrogressive advances. This is called speed plateau [7]. Speed plateau can be managed through structured models or performing predictive modeling.

A model is used to forecast an outcome of a phenomenon at some future state or time based on changes to the inputs given to the model [8]. Predictive modeling is a technique that uses mathematical and computational methods to predict an event or outcome. A mathematical model uses an equation-based approach to describe the current state of a phenomenon under consideration and forecast its future state while the computational model uses simulation-based approach to achieve same purpose where the phenomenon cannot be described easily with equations. Predictive modeling is often performed using curve and surface fitting, time series regression, or machine learning approaches.

Regardless of the approach used, according to [8], the process of creating a predictive model is the same across methods which include:

1. cleaning the data by removing outliers and treating missing data;
2. identifying a parametric or nonparametric predictive modeling approach to use;
3. preprocessing the data into a form suitable for the chosen modeling algorithm;
4. specifying a subset of the data to be used for training the model;
5. training, or estimating model parameters from the training data set;
6. conducting model performance or goodness-of-fit tests to check model adequacy;
7. validating predictive modeling accuracy on data not used for calibrating the model and
8. using the model for prediction if satisfied with its performance

In sprint events prediction, several computational and mathematical models exist for performance enhancement through deliberate training, optimal performance prediction, performance measurement and scoring. This paper explores these models and proposes a novel approach for optimal performance prediction.

The paper is organized as follows; Section I contains the introduction of the paper. This introduction gives a sufficient background to the problem domain requiring exploration. Section II discusses and reviews related works. This is done by categorizing previous studies into works that dealt with performance enhancement techniques, optimal performance prediction (Mathematical) techniques and performances measurement techniques. Section III presents the methodology adopted and how it was implemented. This includes data sampling and experimental design. Section

IV is the detailed presentation of the results obtained and cross validation of the model. Discussions are made on the results in section IV and Section V contains the conclusion.

## II. RELATED WORKS

This section reviews related models that support sprint events modeling and prediction. These models are both computational and mathematical and represent different aspects of the event modeled.

### A. Performance Enhancement Techniques

The Deliberate Practice Model (DPM) is a performance enhancement technique that promotes conscious and structured training directed towards yielding good results. According to the pioneering work of [9], the DPM states that one's expertise in a skill depends largely on how one engages in training rather than how often the practice is done. The emphasis here is on the techniques rather than the number of times of the training. Albeit, [10] had shown that experience determines amazing success in sports and other areas of life. Experience, however, is obtained through continuous participation and training in line with best practices. This seeks to address the speed plateau problem identified by [6] and [7].

Reference [11] modeled forces generated by a sub-10 second 100m sprinter using Newton's Equations of Motion and obtained data from world-class sprint performances. The Newtonian laws show the relationship between force and motion as represented in the equation of impulse governing the velocity:

$$F_t = m(u - v) \quad (1)$$

In the equation (1),  $u-v$  represents the change in velocity ( $u$  is initial velocity and  $v$  is the final velocity),  $t$  is the change in time between each step and  $m$  is the mass of the sprinter. According to [11], the impulse that results in the forward motion of the sprinter is only applied when the foot of the sprinter is in contact periodically with the ground. That typically yields the equation for impulse as:

$$F_{total} \times t_{ground\ contact\ time} = m(u - v) \quad (2)$$

The study shows that after 30m, force in the horizontal direction becomes smaller and lighter than body weight which causes the influence of maximum strength to diminish, making the rate of force development the predominant factor [12]. This would imply that improvements to speed will require more force production within the same ground contact time which could be better achieved through explosive and deliberate strength training. This is largely so as explained by [13] that the values produced by Newtonian model represent forces in the horizontal direction only and are averaged over the period of contact.

According to [13], the Newtonian model of an elite sprinter is developed from a set of formulas that allow estimation of the fundamental quantities which can be related to the biomechanical position of an athlete during the actual step taken. That would allow basic elements of race event such as the force, time of force production and power which results in realized motion for a world record sprint race to be modeled effectively. With that allowance, the Newtonian model can be used as a guideline for the force, velocity, and force production times required of a world class sprinter. This gives the coach and the athlete a guideline as to what resistance to train against, at what speed to train at and in what position to use at training in order to provide specific velocity and movement specific training.

Another important model worthy of consideration under performance enhancement technique is the personal prediction model (PPM) [8]. The PPM addresses two different areas; firstly the general relationship between running velocity and running distance and lastly the corrective formulas for delays that occur during start-up [14]. The model considers the total time (t) and total distance (s) of a race without reflecting the dynamics during the race.

*B. Optimal Performance Prediction (Mathematical) Techniques*

The optimal performance prediction methodologies are basically mathematical models that attempt to predict and explain the chances of reaching the highest estimable velocity in sprint events using scientifically-based approaches and approximations. Reference [15] describes these as phenomenological models as they use mathematical expressions to describe observed phenomena. By demonstration of the phenomenological approach, [16] showed that athletics world records can be covered by power relationships between distance and exertion time while [17] demonstrated power law dependence between velocity and distance for the comparison of decays in energy over time in road and track events.

As cited by [18], reference [19] was the first to introduce Mathematical models that could be useful for race calculation. This was adapted using the principle of speed and energy loss during sprints by [10] to formulate a model that would predict possible race times for world champions. Their combined ideas yielded an equation of motion:

$$\dot{k}(t) = f(t) - \beta^{-1}k(t) \tag{3}$$

From the (3),  $f(t)$  = force per unit mass (N/Kg) which is imposed by the Athlete;  $\beta$  = decay constant, representing the physiological resistances experienced by the athlete, and  $k(0) = 0$ . The distance covered, d by the athlete on track is given as:

$$d = \int_0^T dtk(t) \tag{4}$$

The T is the time spent in the race, using the velocity,  $k(t)$ . Reference [6] instituted a constraint that the internal energy supply must equal the power  $f(t).k(t)$  as:

$$\frac{dEn}{dt} = \alpha - f(t)k(t) \tag{5}$$

The  $\alpha$  is a physiological element which helps to maintain balance in energy of the body, having an initial state of  $E(0) = E_0$  and non-negativity,  $E(t), (E(t) \geq 0)$ . Reference [10] adopted the following parameters in determining the sprint time:  $\beta = 0.892s$ ,  $f = 12.2m/s^2$ ,  $\alpha = 9.83cal / (Kg)$  and  $En_0 = 575cal / Kg$ .

The Keller model arrived at 10.07s as the optimal time for a 100 m race. Reference [20] modeled lower bounds of WR running times. The study presented mathematical models for estimating a lower bound on WR running times for different distances covering from the 100m sprint to the marathon, using regression and order statistics to model the lower bounds. Numerical analysis approaches on *Mathematica* were adopted to estimate parameters for equations that required a solution to a closed system and for addressing linear systems for polynomial equations. The model arrived at the optimal sprint time of 9.69 seconds.

Reference [21] launched a mathematical model for predicting swimming and athletic performances at the OG. This model was based on previous athletic and swimming records taken for both males and females before 1996, 1998, 2000 and 2004 OG performances. The model analyzed that predictions of the performances of males in 50m and females in both 50m and 100m observed times in swimming were close to estimated times. This comparison was not made of track events, though the study showed improvement on the performances of the athletes for both men and women in 100m and 400m races using the (6) and (7):

$$Y = b_0 + (b_1/t) \tag{6}$$

$$Y = e^{(b_0 + (b_1/t))} \tag{7}$$

with coefficients of determination (R2), 0.659 and 0.907 respectively, where  $b_0$  = a constant;  $b_1$  = regression coefficient;  $t$  = year, for each event.

Similarly, for gold-medal performances at OG, [22] presented a model that could measure the universality, limits and predictability of the athletes. The model showed that the relative increasing performances of winners of Olympic medals were normally distributed. This implies that the values of the evolution of performance would be viewed and approximated as an exponentially *a priori* system to an unknown limiting value. This was in view of removing the sophistication or unrealistic nature of earlier formulated models [23,24] that lacked a summarized theory for predicting sport performances [22]. The model, which adopted

self-consistent approach relying on the hypothesis of normality testing, was able to measure limiting performance values, with applicability both in athletics, jumping, throwing and swimming. However, applying a single model to varying degrees of sporting events that require a different level of training, techniques and winning strategies does not really appear realistic scientifically. Although, basic modules of a model could apply to others, the entire model will perform poorly in some. This was expected in the model by [22].

To address this, [25] predicted London Games using dataset obtained from OG from 1990 to 2000 for the Men's 100m category of sprint races. His model used simple log-linear regression technique to forecast the winning time which he arrived at 9.68 seconds. The study predicted that the model would shrink indefinitely, given the nature of the data.

### C. Performances Measurement Techniques

These techniques measure, compare and score performances using a set of scientific tools. To enhance experience gathered through training, [1] further developed a mathematical model for predicting linear races. The methodology employed used a simple least square fit method to measure the performances. The model only considered World record (WR) of race times, with no application on Olympic Games (OG). Reference [26] extended the work of [1] and modeled the energy loss of the athlete as a simple function of the force of centrifuge which acts on the athlete as he runs around the curve. He predicted the sprint times for both linear races and races around curves, including 200m and 400m for both indoor and outdoor athletic games using WR, just like the model by [1]. Reference [26] adopted the running

times of Donovan Bailey who held the 100m WR 1996 Olympic Games in Atlanta to benchmark his result.

In comparing athletic performances, [27] presented models which analyzed and modeled the change in the times with distance of the men and women in WR for the 2006 edition for between 100m and marathon. The study normalized athletic performances with respect to the plotted values (times against distances). The aim was to analyze the performances of each athlete at different distances in comparison with others. The study concluded that the decrease noticed in average speed with distance in athletic events is natural and useful for measuring performance which it attempted describing in parametric forms. The forms adopted seemed general for performance prediction through fitting and the model arrived at 9.87s as the optimal time for the 100m race of a Gold medalist.

References [28,29] performed a biomechanical analysis of the Sprint and Hurdles Events at the 2009 IAAF World Championships in Athletics. Using Usain Bolt's finishing time, [28] offers explanation on his performance advantage. By using a technique based on infrared laser measurement of the distance to the athlete and with comparison with previous top-level sprinters like Carl Lewis and Donovan Bailey, the study shows that Bolt outperforms his competitors in the acceleration phase of the race with performance differing considerably from other athletes with respect to the times for measured intervals in the race as seen in Figure 1. And in the phases of maximal velocity and sprint-specific endurance, Bolt defines new standards for athletes. By adopting a different method, this paper attempts to substantiate the claim by [25] and compares the optimal times of [1, 26] for gold medalists for the men category of OG.

	RT	20m	40m	60m	80m	100m	20-40m	40-60m	60-80m	80-100m
Bolt	0.146	2.88	4.64	6.31	7.92	<b>9.58</b>	1.76	1.67	1.61	1.66
Gay	0.144	2.92	4.70	6.39	8.02	<b>9.71</b>	1.78	1.69	1.63	1.69
Powell	0.134	2.91	4.71	6.42	8.10	<b>9.84</b>	1.80	1.71	1.68	1.74
Bailey	0.129	2.92	4.73	6.48	8.18	<b>9.93</b>	1.81	1.75	1.70	1.75
Thompson	0.119	2.90	4.71	6.45	8.17	<b>9.93</b>	1.81	1.74	1.72	1.76
Burns	0.165	2.94	4.76	6.52	8.24	<b>10.00</b>	1.82	1.76	1.72	1.76
Chambers	0.123	2.93	4.75	6.50	8.22	<b>10.00</b>	1.82	1.75	1.72	1.78
Patton	0.149	2.96	4.85	6.65	8.42	<b>10.34</b>	1.89	1.80	1.77	1.92

Fig.1. Split and interval times (sec) of the men's 100m final at the 2009 IAAF World Championships in Athletics [29]

### III. METHODOLOGY

The technique adopted for predicting the running events in this paper was the Curve fitting toolbox of MATLAB while the system identification toolbox was employed for cross-validating the model. The results were evaluated with parameters such as the Sum of Square Error (SSE), R-Square, adjusted R-Square, and Root Mean Square Error (RMSE).

#### A. Data

The dataset used for this study were obtained from the Olympic Champions' Records. These data were best

sprint performances (for Gold medalists) of male athletes of OG from 1896 to 2016 made available in the Olympic website. These data were extracted from the published results of athletics from Olympic official website by adopting type of sport-game-event search mechanism. The record sorted medalists according to significance with gold running top, followed by other medalists and other runner-ups in that order. This made extraction easy for the purpose of analysis.

The dataset were subjected to model predictor for the prediction of the optimal sprint time as represented in the model block diagram in Figure 2. The data were computed and plotted as seen in the time plot shown in

Figure 3. The upper plot is the Optimal Sprint Time, in seconds taken as the outputs of the model while the lower plot captures the Olympic Years as inputs into the model.

*B. Experimental Design*

The Figure 4 shows the experimental design of the proposed model. This contains the model components, including the data entry/importer, data viewer, data pre-processor, model predictor, model estimator, model validator and model viewer. A model predictor is a tool that uses known results to create, process and validate a model that can be used to forecast future outcomes. Predictors are variables that use data mining and probability techniques to possibly influence future results.

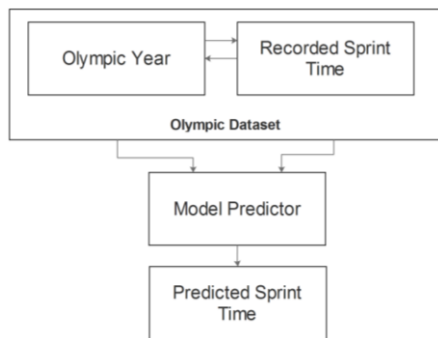


Fig.2. Model Block Diagram for Dataset Extraction

The datasets were imported from the workspace

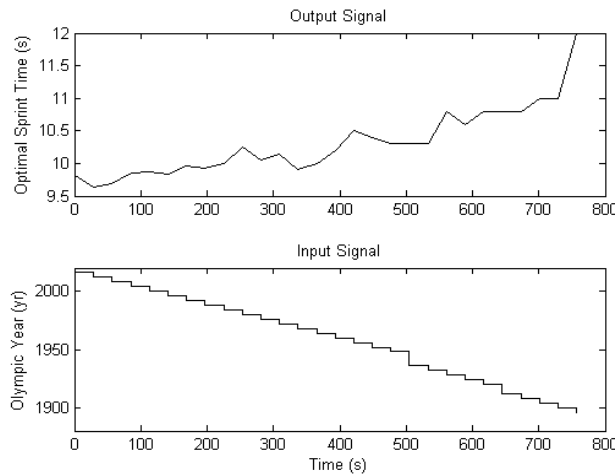


Fig.3. Time Plot of Data set

In finding the polynomial degree that would model our dataset, the differences of the sprint times were computed and the degree 7 is found suitable for the prediction. The choice of the septic polynomial with regression feature arises from the fact that in many real-life situations; one cannot find a simple model that fits data points exactly. A linear model polynomial of degree 7 is adopted for this system as shown in (8).

$$f(x) = p1 * x^7 + p2 * x^6 + p3 * x^5 + p4 * x^4 + p5 * x^3 + p6 * x^2 + p7 * x + p8 \tag{1}$$

through the data importer unto the data viewer, and were used as working data. These working data were used for prediction by the model predictor in their current state and pre-processed by data pre-processor to remove any unwanted constant factors before estimation and used for cross-validation as discussed in preceding sections.

A polynomial function is used to model the system. A polynomial function is a function with non-negative integer powers of x, where x is the term of the polynomial. Such functions may be quadratic, cubic, quartic, quintic, sextic, septic, octic, nonic, decic, etc. [30] depending on its degree. The degree of a polynomial function points to the highest power of a term in its expression in canonical form. This may linearly combine monomials. A constant (non-zero) polynomial has degree 0, linear has 1, quadratic 2, cubic 3 and quartics 4, quintic 5, sextic 6, septic 7, octic 8, nonic 9, and decic 10 [31]. In finding a polynomial function to model some dataset, it is important always to first find the degree of the function, using finite differences approach. The method allows finding the differences between the outputs of the data points. If the differences are constant, the process stops and the function takes the degree 1, otherwise, the difference of the differences just found are computed. And if they are not constant, the process is repeated until a constant value is obtained. That way, the number of times it takes for the differences to become constant becomes the degree of the polynomial.

To implement this model, a curve fitting mechanism was set in motion. Curve fitting is a systemic mechanism of building a curve or function mathematically which has the best fits to a series of datasets [32,33]. It requires capturing the trend in a dataset by assigning a single function across the entire range in order to identify the coefficients, 'p' such that the polynomial function f(x) fits the data well [34].

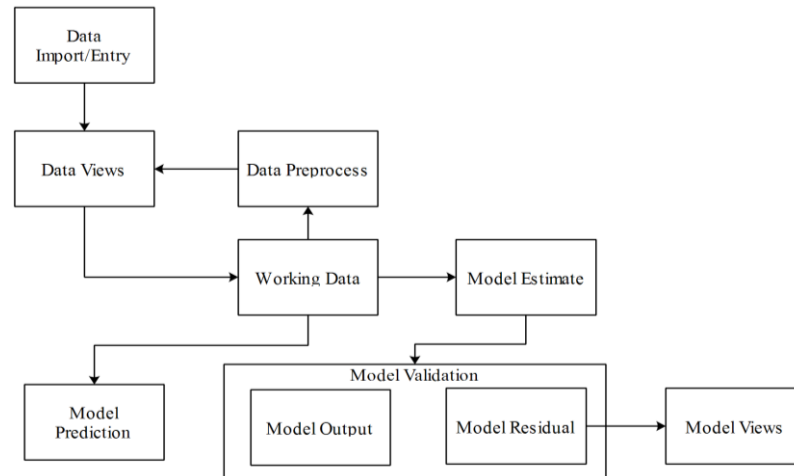


Fig.4. Proposed Model Design

In to determine the coefficients of the 7th degree polynomial that best fits our dataset in a least-squares method, the polyfit function was used with the syntax:

$$p = \text{polyfit}(x, y, n) \quad (9)$$

where:

- $x$  and  $y$  are vectors containing the  $x$  and  $y$  data to be fitted
- $n$  is the degree of the polynomial to return

At 95% confidence bounds, the coefficients were:

$$\begin{aligned}
 p1 &= -1.042e^{(-012)}(-3.13e^{(-012)}, 1.046e^{(-012)}) \\
 p2 &= 1.433e^{(-008)}(-1.425e^{(-008)}, 4.292e^{(-008)}) \\
 p3 &= -8.447e^{(-005)}(-0.0002522, 8.325e^{(-005)}) \\
 p4 &= 0.2766(-0.2701, 0.8233) \\
 p5 &= -543.3(-1613, 525.8) \\
 p6 &= 6.403e^{(005)}(6.141e^{(005)}, 1.895e^{(006)}) \\
 p7 &= -4.192e^{(008)}(-1.237e^{(009)}, 3.984e^{(008)}) \\
 p8 &= 1.176e^{(011)}(-1.107e^{(011)}, 346e^{(011)})
 \end{aligned}$$

With the understanding that in order to predict the next outputs of a model from former data, a system requires past inputs and outputs, this study considered a data for over a time horizon of 28 samples to predict for 5 prediction horizons. The time horizon is given as  $kTs$  time units, where  $Ts$  is the sampling interval and  $k$  is the prediction horizon. During the prediction, the algorithm harnesses the two data values - measured and calculated output in the difference equation to compute the future output. The predicted value  $y(t)$  is calculated from all available inputs  $u(s)$ , where  $s \leq t$ , and all available outputs  $y(s)$ , where  $s \leq (t - k)$ . The argument  $s$  represents the data sample number.

#### IV. RESULTS AND DISCUSSION

The implementation of the predictive algorithm described in the previous section was measured using as parameters - Sum squared error (SSE), R-square, Adjusted R-square and the root mean squared error (RMSE). The yielded results are shown in Table 1 and Table 2.

Table 1. Model Performance and Goodness of fit

Metrics	Value
SSE	0.5129
R-square	0.9304
Adjusted R-square	0.9060
RMSE	0.1001

The Sum squared error (SSE) measures the performance of the model according to the sum of squared errors [35] and returns the value of 0.5129. The R-square is a statistic which measures how successful the fit as could be used to explain the variation of the data, computed using equation (9):

$$R - \text{Square} = 1 - \frac{RSS}{TSS} \quad (10)$$

where  $RSS$  = residual sum of squared errors for the fitted model and  $TSS$  = total sum of squares. The model returned the R-square value of 0.9304 and an adjusted R-square value of 0.9060. The root mean squared error (RMSE) is the standard errors of both the fit and the regression, which the model gives as 0.1001.

Table 2. Predicted Olympic Optimal Running Time for the Next 5-Games Year

Olympic Year	Optimal Running Time
2020	9.71
2024	9.61
2028	9.37
2032	8.89
2036	8.01

The curve fitting analyses displayed in the Figures 5 and 6 enabled pin-pointing on the plots with help of a Data Cursor. The Data Cursor reads up the values of X-axis - projected year and the Y-axis - predicted Running time. At 95% level of confidence, the predicted optimal running times for an Athlete (Gold medalists) in the year

2020, 2024, 2028, 2032 and 2036 are 9.71s, 9.61s, 9.37s, 8.89s and 8.01s respectively. This implies that within the next 5 Olympic Athletic competitions, the 100m track is predicted to be expected to be covered below the 9s range.

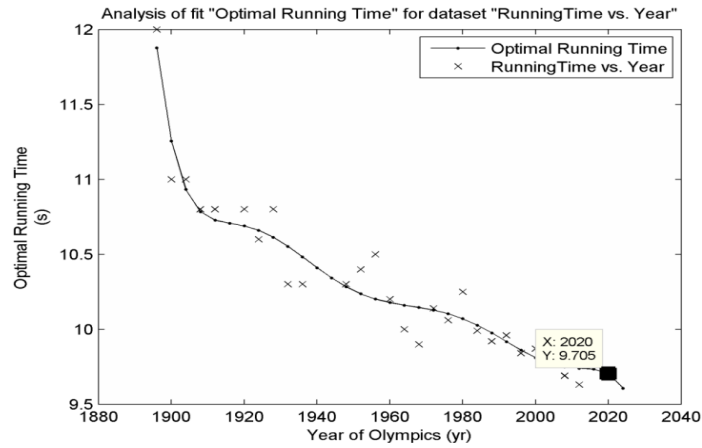


Fig.5. Optimal Time for Year 2020 Olympic Sprint for Male Category

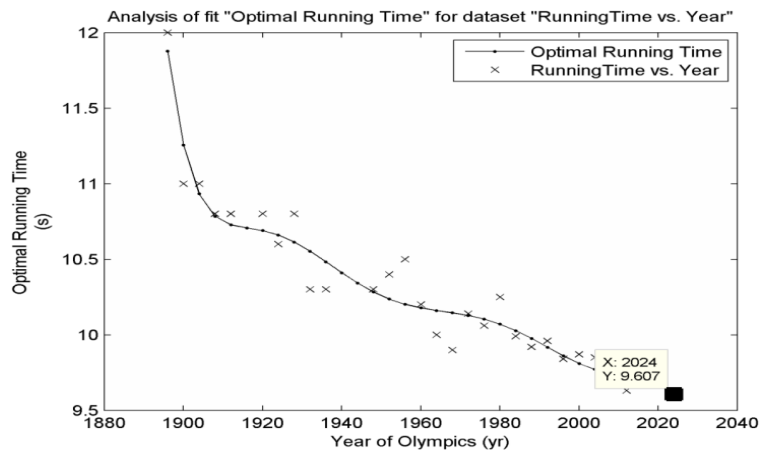


Fig.6. Optimal Time for Year 2024 Olympic Sprint for Male Category

A. Model Validation and Evaluation

Cross-validation is the process that allows a system model validation using an independent dataset. In this study, the Olympic Year was considered an independent dataset and used as input for the cross-validation. Two model validation approaches were adopted – Model

Residual method and Model Output method. To allow for cross-validation of the model, the datasets were pre-processed to remove the constant levels in the data sequences as shown in Figure 7. The pre-process includes mean removal, trend removal and range selection.

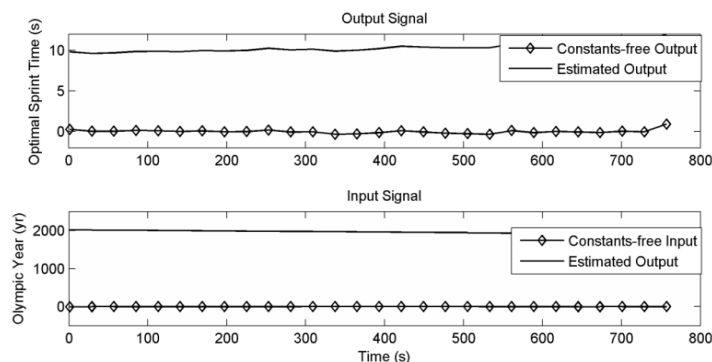


Fig.7. Constant levels-free Plot

### B. Residual Model Validation Method

The residual approach adopted for our time-domain validation data performs two tests on the model which are shown in two axes in the plot in Figure 8. These tests were whiteness and independence tests. The autocorrelation function of the residuals for each output were computed by the whiteness test – Optimal Sprint Time, while the independence test measures the cross-correlation between the input – Olympic Year and the residuals for each pair of input-output combination.

#### i. Whiteness Test

The whiteness test is represented in the top axes, where the horizontal scale is the number of lags, which is the time difference (in samples) between the signals at which

the correlation is estimated. This was taken at 20 samples. The horizontal dashed lines on the plot represent the confidence interval (CI) of the corresponding estimates. Notably, the lines within the CI are considered to be insignificant, implying success of the model in the whiteness test.

#### ii. Independence Test

The independence test is shown on the bottom axes of the same Figure 6. A good model is expected to have residuals uncorrelated with past inputs and exist within the defined CI. This implies that the model does not show how the inputs form corresponding outputs. In the study, the model peaked within the CI for lag 20.

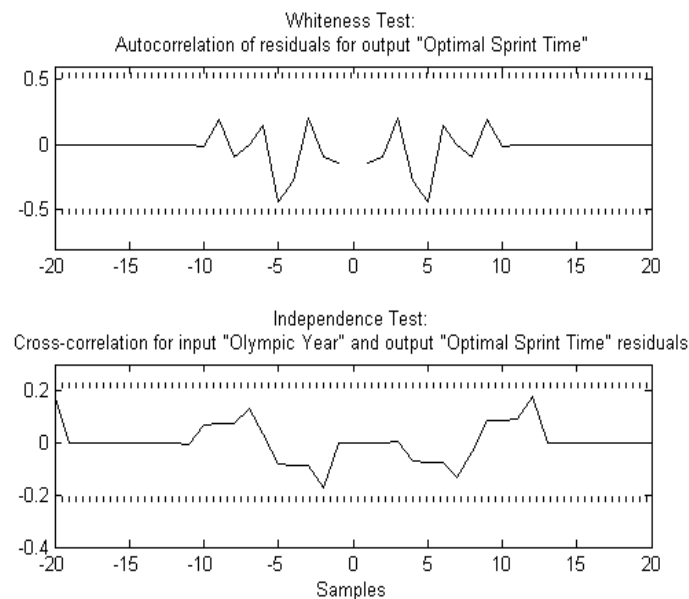


Fig.8. Residual Model Validation Plot show Whiteness and Independence Tests

### C. Model Output Validation Method

The time-domain validation data of the model output, shows simulated plot or predicted model output. In the study, the simulated and 5-step predicted outputs were captured and compared as seen in Figures 9 and 10 respectively. The simulated model output returned 99.98% working data best fit with the regressors used in the nonlinear block of model of Optimal Sprint Time( $t-1$ ), Optimal Sprint Time( $t-2$ ), Olympic Year( $t-1$ ) and Olympic Year( $t-2$ ). The forward prediction error was  $2.292e-008$ , while the loss function was  $2.926e-009$ . Under the same conditions, the 5-step prediction yielded 99.99% as the best fit for the data at the standard regressors in the order of  $n_a = 2$ ,  $n_b = 2$ ,  $n_k = 1$ .

#### i. Model Prediction

In prediction, ideal model behavior is observed when the sum squared error is close to 0 like in this model, where the random error element is smaller, and that the fit will be more useful for prediction. This shows that the model developed is good for the prediction in that regard. Similarly, the R-squared value gives the square of the

correlation between two values, which are the response values and the predicted response values. This indicates that a higher portion of the variance, about 93%, is covered by the model in explaining the total variation in the dataset about the average. The adjusted R-square is particularly useful when measuring the fit quality of a model. With the value closer to 1 in the model developed, indicates a better fit. The root mean square error on the other hand, is an estimate of the standard deviation of the random component in the dataset, and with a value (0.1001) closer to 0, further shows that the fit is more appropriate for the prediction. This implies that, the small value of the RMSE in the model is a good measure of accuracy [36], as it compares the forecasting errors in the model.

#### ii. Model Validation

Model validation is a set of processes and activities intended to show or prove that models are performing as expected or they are close to the real system. In this model, at the validation phase, the CI agrees with the range of residual values with a specific probability. Such



a probability must be insignificant statistically for the system modeled. Since the system identification toolbox of MATLAB in the model uses the estimated uncertainty [37] in the model parameters to determine CIs and considers that the estimates have a Gaussian distribution, the CI which was taken at 95% shows the region about zero, representing the set of residual values that have the probability of being that statistically insignificant. For

both the whiteness test and the independence test in Figure 7, the model succeeded in the autocorrelation of residuals for outputs and cross-validation of inputs and outputs. Furthermore, the model passed the tests of measured and simulated model outputs and 5-step prediction as seen in Figures 9 and 10 respectively. These show the validity of the model for prediction.

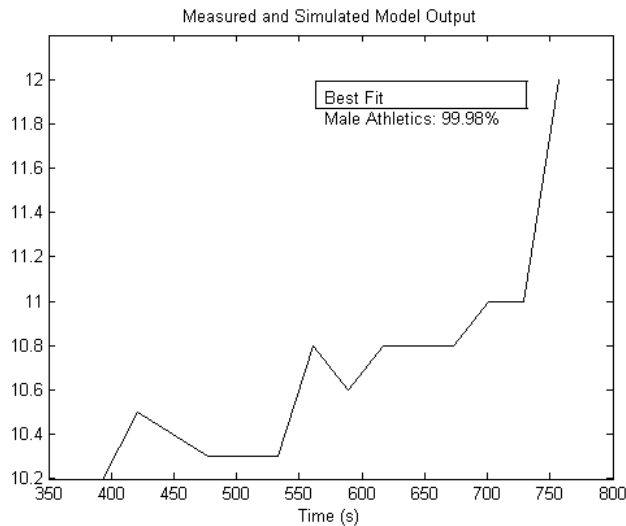


Fig.9. Measured and Simulated Model Output

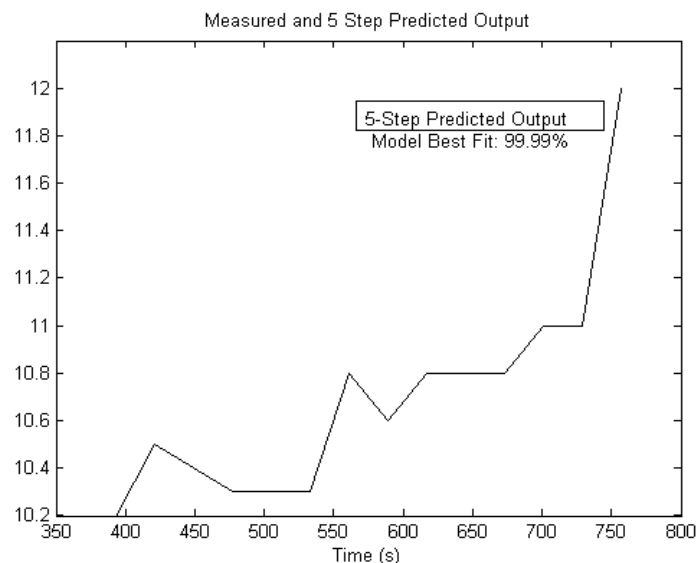


Fig.10. Measured and 5-step Predicted Model Output

## V. CONCLUSION

This study has developed and evaluated a predictive model for athletic sprint events of Olympic Games. The model predicts that the male 100m sprint events in 2020, 2024, 2028, 2032 and 2036 Olympic Games as 9.71s, 9.61s, 9.37s, 8.89s and 8.01s respectively. This is substantially reasonable as shown in the performance functions tested. The cross-validation indices further strengthen the validity of the model as a good fit for the

dataset. In comparison with the models formulated by [1, 10, 26], the proposed model here performs better. However, in agreement with [14], the model tends to shrink.

These results of the proposed model have shown better performance over existing models which is useful in forecasting trends and possibilities in view of helping athletes train consciously and assist coaches. It would help Olympic organizers/hosts gain an insight into what should be anticipated. Such would have an enhanced effect on the study of predictive algorithms.

## REFERENCES

- [1] J. B. Keller, "Optimal Velocity in a Race", American Mathematics Monthly, Vol.81, Issue. 474, pp. 331-333, 1974.
- [2] S. Pinker, "The Blank Slate: The Modern Denial of Human Nature", Viking, Press, New York, pp. 47-56, 2002.
- [3] M. Ridley, "Nature Via Nurture: Genes, Experience, and What Makes Us Human", HarperCollins Publishers, New York, pp. 98-102, 2003.
- [4] K. A. Ericsson, K. Nandagopal and R. W. Roring, "Toward a Science of Exceptional Performance: Attaining Superior Performance Through Deliberate Practice". Annals of the New York Academy of Sciences, Vol. 11, Issue. 72, pp. 199-217, 2009.
- [5] Jürgen Schiffer, "Training to Overcome the Speed Plateau", New Studies in Athletics, Vol. 2, Issue 1, 2011.
- [6] B. Tabachnik, "The Speed Chute". NSCA Journal, Vol. 14, Issue 4, pp. 75-80, 1992.
- [7] T. Kurz, "Science of Sports Training: How to Plan and Control Training for Peak Performance". Island Pond, VT: Stadion Publishing Company, 2001.
- [8] W. Westera, "Where Bolt and Bekele meet: the analytical basis of running performance estimates". International Journal of Sports Science and Engineering, 4, 3: 139-152, 2011.
- [9] K. A. Ericsson, R. Th. Krampe and C. Tesch-Romer, "The Role of Deliberate Practice in the Acquisition of Expert Performance". Psychological Review, Vol. 100, Issue. 3, pp. 363-406, 1993.
- [10] J. B. Keller, "A Theory of Competitive Running", Physics Today, No. 43, pp. 98-120, 1973.
- [11] Jeremy Richmond, "Modelling a Sub-10 Second 100m Sprinter Using Newton's Equations of Motion", New Studies in Athletics, Vol. 11, Issue 2, pp. 69-75, 2011
- [12] D. Schmidtbleicher, "Training for power events. In: Strength and Power in Sport". P.V. Komi, ed. pp. 381-395. Boston: Blackwell Scientific, 1992.
- [13] Jeremy Richmond, "Newtonian Model of an Elite Sprinter: How Much Force do Athletes Need to Produce Each Step to be World Class?", Fitness First Randwick Australia, University of Sydney, pp. 1-10, 2010
- [14] A. J. Ward-Smith, "A mathematical theory of running, based on the first law of thermodynamics, and its application to the performance of world class athletes". Journal of Biomechanics 18: 337-349, 1985.
- [15] Wim Westera, "Phenomenology of Sprinting and Endurance: Toward a Uniform Performance Assessment Model", New Studies in Athletics, Vol. 11, Issue 2, pp. 87-92, 2011
- [16] L. Katz, and J. S. Katz, Power laws and athletic performance. Journal of Sports Sciences; 17: 467-476, 1999.
- [17] J. Stanckiewicz, "On Track for a 2-hour Marathon", South African Journal of Science 101: 397-398, 2005.
- [18] K. F. Dyer, "The trend of the male-female differential in various speed sports 1936-84", Journal of Biosocial Science, Vol. 4, Issue. 1, 23-45, 1986.
- [19] A. V. Hill, "The Physiological Basis of Athletic Records", Report of the 93rd Meeting, British Association for the Advancement of Science, No. 156, pp. 56-78, 1925.
- [20] K. Heisler, "Modelling Lower Bounds of World Record Running Times", Senior Research Paper Presented to the Department of Mathematics and Computer Science of Stetson University in Partial Fulfilment of the Requirements for the Degree of Bachelor of Science, Stetson University, 1-35, 2009.
- [21] T. Heazlewood, "Prediction versus Reality: The Use of Mathematical Models to Predict Elite Performance in Swimming and Athletics at the Olympic Games". Journal of Sports Science and Medicine, The 8th Australasian Conference on Mathematics and Computers in Sport, 3-5 July 2006, Queensland, Australia, 5, pp. 541-547.
- [22] F. Radicchi, "Universality, Limits and Predictability of Gold-Medal Performances at the Olympic Games", Physics, Social and Physical Sciences. Departament d'Enginyeria Quimica, Universitat Rovira i Virgili, Spain, pp. 1-8, 2012.
- [23] W. M. Kolb, "Curve Fitting for Programmable Calculators". Syntec, Incorporated, 5-9, 1984.
- [24] K. A. Ericsson, M. J. Prietula & E. T. Cokely, "The Making of An Expert", Harvard Business Review, July-August, pp. 1-8, 2007.
- [25] M. Gesmann, "A Prediction for the Olympic Men's 100m Sprint. Revolutions". 2012. Retrieved from <http://blog.revolutionanalytics.com/2012/07/a-prediction-for-the-olympic-mens-100m-sprint.html>. Accessed 30/03/2017.
- [26] J. R. Mureika, "A Simple Model for Predicting Sprint Race Times Accounting for Energy Loss on the Curve", Department of Computer Science, University of Southern California, Los Angeles, California 90089 USA, September 27, Vol. 4, 28-39, 1997.
- [27] H. J. Grubb, "Models for Comparing Athletic Performances", Department of Applied Statistics, The University of Reading, pp. 1-19, 2007.
- [28] R. Graubner and E. Nixdorf, "Biomechanical Analysis of the Sprint and Hurdles Events at the 2009 IAAF World Championships in Athletics", New Studies in Athletics Vol. 26, Issue 1/2; pp. 19-53, 2011.
- [29] H. Muller and H. Hommel, "Biomechanical Research Project at the VIth World Championchips in Athletics, Athens 1997". New Studies in Athletics, Vol. 11, Issue 2, p 3, 2011.
- [30] S. Axler, "Linear Algebra Done Right", (2nd ed.), Springer Science & Business Media, USA, 1-198, 1997.
- [31] R. B. King, "Beyond the Quartic Equation", Mathematics, Springer Science & Business Media, USA, 1-165, 2009.
- [32] S. L. Arlinghaus, "PHB Practical Handbook of Curve Fitting", CRC Press, China, 1-87, 1994.
- [33] H. Yan, Y. Shen and J. Li. "Approach to calculating spatial similarity degrees of the same river basin networks on multi-scale maps", Geocarto International, Vol. 31, Issue. 7, pp. 22-36, 2016.
- [34] K. R. Gurley, "Numerical Methods Lecture 5: Curve Fitting Techniques", CGN-3421 Computer Methods, pp. 89-102, 2003.
- [35] C. Huang, "Gray Modeling and Studying of Olympics Track and Field Achievements", Advances in Intelligent and Soft Computing, Vol. 2, Issuee. 3, pp. 54-68, 2012.
- [36] B. Ozkaya, "Mathematical simulation and long term monitoring of leachate components from two different landfill cells", Journal of Hazardous Materials, Vol. 7, Issue. 31, 9-23, 2006.
- [37] Mathworks. "Predictive Modeling", 2013. Retrieved from <https://www.mathworks.com/discovery/predictive-modeling.html>. Accessed 10/07/2018.

### Authors' Profiles



**John E. Efiog** is a faculty in the Department of Computer Science, Wesley University Ondo, Nigeria. His interest covers Mobile/Pervasive Computing, Machine learning, Human-Computer Interaction, Software Engineering and Computing Technology Applications.



**Emmanuel A. Olajubu** is a Senior Lecturer in the Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria. His research area includes Mobile Agents/Networks, Internet of Things and Digital Communication Systems.



**Felix O. Aranuwa** is a senior faculty member in the Department of Computer Science, Adekunle Ajasin University, Akungba-Akoko, Nigeria. His research interest covers Data Mining, Machine Learning, Theory of Computation and Human-Computer Interaction.

**How to cite this paper:** John E. Efiog, Emmanuel A. Olajubu, Felix O. Aranuwa, "Formulation of Sprint Time Predictive Model for Olympic Athletic Games", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.11, No.4, pp.33-43, 2019. DOI: 10.5815/ijitcs.2019.04.04