Track and Field Performance Data and Prediction Models: Promises and Fallacies

Yuanlong Liu

Western Michigan University
yuanlong.liu@wmich.edu

1 Introduction

Prediction is always a fascinating obsession no matter what we predict. We predict what tomorrow's weather will be, which team will win the game, who will be the next president of the United States, how fast one will run and how high one will jump, and even how much money we will make next year. Mathematical and statistical models have been used to predict future events in different disciplines. Some prediction models can be used to predict specific magnitudes of an event in the future. For example, because of the high technology development in the last two decades, the models used for weather forecasting became more and more accurate. In some other fields, the prediction models are actually not for a specific magnitude of an event in the future but for predicting various developmental trends in the future. For example, statistical models are commonly used for predicting the developmental trends in stocks but not for predicting specific values of the stocks.

In track and field, the curiosity with respect to the limits of human athletic performances has always been of interest to athletes, coaches and scientists (Liu 2002). Since track and field performances are the oldest competitive sports, the interest in a human being's ability and accomplishment in running, jumping and throwing has a long and diversified history. Mathematicians, statisticians, physiologists and operational researchers have developed numerous models to predict future performances, world records and 'ultimate' performances in track and field, and to compare male and female performances (Liu and Schutz 1998).

There have been many publications regarding the prediction of track and field performances. Some of them are opinion based and lack scientific support. Others used track and field data and mathematical and statistical models. There has been some confusion as to what the track and field performance data provide and what the prediction models show. A review of the prediction models in literature leads us to believe that there is no prediction model available to accurately predict the future magnitude of any performance in track and field (Liu 2002, Liu and Schutz 1998). So, what are the promises and fallacies of the prediction models in track and field? The primary focus of this chapter is a discussion of the following issues in predicting track and field performances: Track and field performances data and its validity in predicting future performances and the promises and fallacies of the prediction models.

2 The Track and Field Data and Prediction Models

2.1 Track and Field Performance Data

The sources of track and field performance data

Collecting track and field performance data is different from collecting data in laboratories. The major advantage of collecting track and field performance data is that it can be obtained easily from different published sources. The International Association of Athletic Federation (IAAF, http://www.iaaf.org) is a governing authority for athletic programs, standardized technical equipment and world records. It provides the verified world records in track and field since 1912. The federations in individual countries affiliated to IAAF have the system to record and provide the track and field performance data in each countries of the world. The International Olympic Committee (http://www.olympic.org) provides all the records of the modern Olympic games. Track and field performances data can also be obtained from newspapers and magazines (e.g., Track and Field News), yearbooks and organization web-pages.

Track and field data in predicting future performances

The track and field performances data are usually accurately recorded with great precision (e.g., to the hundredth of a second in track competitions). There are many different ways to collect the track and field performance data. Men's and women's records at different competition levels (world, Olympic, national, college, and high school, etc.) provide historical documentation for human beings breaking their limits in running, jumping and throwing. However, one should notice that the discontinuity of the record data (e.g., the world records and Olympic records) may seriously affect the validity when it is used in the purpose of prediction, because there may be no records for several years or several records in one single year. The best performance per year data provide better continuous longitudinal trend without missing magnitudes. Coaches use individual athlete's longitudinal performance data to evaluate and predict his/her future performances.

Scientists use track and field performance data to develop prediction models. Historically, there are three types of data used in developing the prediction models and in comparing male and female track and field performances in literature. Some researchers used the performance percentages of improvement over certain time period in evaluating and predicting the track and field performances (Dyer 1982, Ryder et al. 1976, Stefani 1977). World records and Olympic records data are also commonly used (Whipp and Ward 1992, Gembris et al. 2002). Liu and Schutz (1998) pointed out that the annual best performances should be used in the development of the track and field prediction models and these data would overcome the drawback of discontinuity in improvement percentage and world record data. Annual best performance data have also been used in academic research publications (Liu 2002, Liu and Schutz 1998, Schutz and Liu 1998).

Validity of the track and field performance data

As shown above, all the three types of data have been employed to develop the prediction models and to compare male and female track and field performances. However, the discontinuous characteristic of the improvement percentage and world records data may seriously affect the validity of the prediction models. For example, between 1960 and 1967, there has been only one new world record established in men's 1500 meters run. The current women's 200 meters world record has stood for 21 years.

Using only world records or the improvement rates, and ignoring the fact that there has been no improvement over years would not provide a valid model for reality. Liu (1993) used the best annual performance data and world record data to fit nonlinear regression and linear models and to examine the impact of the discontinuity in the world record data on the validity of prediction models. The results showed that the parameter estimates of the prediction models were much more unstable with using the world record data than with the annual best performance data. Schutz and Liu (1993) indicated that using word records would lead to an inflation of the estimate of linearity and/or improvement rates that cannot be empirically justified. Because of the continuous property and the stable estimation of the parameters in the prediction models, the annual best performance data has been shown to be the best data in developing deterministic prediction models in track and field.

2.2 Prediction Models in Track and Field

A brief overview of existing prediction models

Evaluating and predicting track and field performances has a long history and there have been numerous models developed to evaluate and predict future track and field performances and to compare male and female performances. Kennelly (1906) is probably the earliest researcher to examine the track and field data using a statistical modeling technique. He showed that a linear relation between velocity and running distance was stable over all the examined running events. Physiologists Hill (1925) and Peronnet and Thilbault (1989) developed metabolism based models to provide a physiological explanation of the time-distance relation and to predict record performance. Keller (1974) and Senator (1982) used biomechanics to develop mechanics-based models to predict the ultimate performances. However, models based on either physiology or biomechanics do not work well when evaluating and predicting the performance trend in track and field.

The most common approach in developing the prediction models for track and field performances are the ones used by mathematicians, statisticians, and psychometicians (e.g., Chateerjee and Chatterjee 1982, Deakin 1967, Liu 2002). These models usually examine longitudinal improvement of running, jumping and throwing through types of asymptotic mathematical function.

Mathematical models

The following section provides a brief description of the types of mathematical models that have been utilized to predict track and field records (also see Schutz and Liu 1998 for partial review).

Historical Date versus Time

In these models a timeyear, velocityyear, or a distanceyear relationship is developed separately for each event. An exponential model relates time, velocity, or distance and historical date for a specific event:

$$T(n) = b_0 + b_1 e^{-b_2 n}, \tag{1}$$

where $T(n)$ is the predicted performance in year n, b_0 is the estimated asymptotic value of $T(n)$ as n approaches infinity, and b_1 and b_2 are calculated parameters which govern the shape of the curves (Chatterjee and Chatterjee 1982, Lucy 1958, Schutz and McBryde 1983).

A linear model relates time, velocity, or distance and historical date for a specific event:

$$T(n) = b_0 + b_1 n, \tag{2}$$

where $T(n)$ is the predicted performance in year n, b_0, the intercept, is the world record or best performance in year zero (usually 1900), and b_1, the slope, is the estimated yearly amount of improvement (e.g., Ballerini and Resnick 1985, Ballerini 1987, Whipp and Ward 1992).

A polynomial model relates velocity and historical time for a specific event:

$$V(n) = b_0 + b_1 n + b_2 n^2 + ... + b_m n^m, \tag{3}$$

where $V(n)$ is velocity, n is the historical year, and b_0, b_1, b_2, $\ldots$, $b_m$are calculated parameters which determine the shape of the curve (Mognoni et al. 1982).

Time versus Distance

In these models the time or velocity versus distance relationship, for all running events, is developed separately for specific years. By comparing the different yearly relationships, future projections are made via a second level modeling of the derived parameters.

An exponential model relates mean running velocity and distance for all events in a specific year:

$$V(d) = b_0 + b_1 e^{-b_2 d}, \tag{4}$$

where $V(d)$ is the mean velocity for a specific year, d is the running time or distance, and b_0, b_1, and b_2 are calculated parameters reflecting the shape of the curve (Furusawa et al. 1927, Ward-Smith 1985).

A power function model relates running time and distance for all events in a specific year:

$$T(d) = b_1 d^{b_2}, \tag{5}$$

where $T(d)$ is running time, d is distance, and b_1 and b_2 are calculated parameters (Lloyd 1966, Riegel 1981). This model is equivalent to a linear relationship between time and distance on a logarithmic scale (e.g., Kennelly 1906).

A linearlog function model relates velocity and distance for all events in a specific year:

$$V(d) = b_1 f(\log(d)), \tag{6}$$

where $V(d)$ is velocity, d is distance, and $f()$ is a function which is varies with different researchers (Craig 1963, Francis 1943, Hill 1925).

3 The Promises and Fallacies of the Prediction Models

3.1 The Expectation and Reality on the Predictions in Track and Field

When we talk about predictions or forecasting, we usually expect to accurately know what would happen. For example, when we listen to weather forecasting, we would like to know whether it will be rainy or sunny tomorrow. Because of the development of technology and science, short term weather forecasting is becoming trustworthy. The mathematical and statistical models are also used to predict the stock market. However, the prediction for stocks for tomorrow may or may not be very accurate. The inaccuracy of the stock prediction for tomorrow is due to many external factors that are out of the researchers' control and the unsystematic effect cannot be exactly modeled. Yet the models in stock predictions usually work very well to predict the trend of the stocks over a period of time.

What do we expect from the mathematical models in track and field performance predictions? Similar to predicting tomorrow's weather, it is not uncommon for people to expect that a mathematical model would provide an accurate magnitude of the next world record and the year in which that the record may be established. Therefore, there has been so much effort to predict future world records and waiting time for establishing new world records. The journal *Nature*, in spite of its reputation around the world, had published several predictions that were far from the reality. Whipp and Ward (1992) with limited literature review on prediction modeling published an essay in *Nature*. They used discontinuous word record data with a linear regression model and proposed that the track and field performances would improve linearly. They asserted that female performances in track and field would catch up with male performances as early as 1998 in some events. Based on their assertion, one might run 100 meters at no time in some day! *Nature* published another manuscript that again ignored the academic literature and used the discontinuous record data (Gembris et al. 2002). The Figure 1 in the article already shows that the proposed model was not valid and the models could not represent the trend of the improvement of their track and field data.

Schutz and Liu (1998) summarized some published predictions of the world records literature (Peronnet and Thibault 1989, Ryder et al. 1976, Schutz and McBryde 1983, Stefani 1977). The reality differs from the model predictions. Some people say that the track and field performances would improve linearly they do not. Others say that in some events the performances have reached an asymptotic level, however new world records are still established. The dilemma is that either all these models do not demonstrate what they promise to show or what we expected is not what the models are supposed to reveal. In the next section, the promises and fallacies of the prediction models in track and field are discussed.

3.2 The Promises and Fallacies of the Prediction Models in Track and Field

Research has shown that the nonlinear regression modeling technique with the annual best performance data is the most appropriate method to describe track and field performances and predict the deterministic trend of improvement in track and field. When the general trend of the performances has met the asymptotic level of the deterministic models, regression models cannot assess the future performances. Other approaches such as a random sampling model suggested by Schutz and Liu may be used to predict the general trend of the world records (Schutz and Liu 1998).

Regression type models only promise to describe the past performance patterns over time and predict the trend of future performances in track and field. Although one may wish to use the prediction models to predict the specific magnitude of future world records, however it is not what the regression models promise to provide.

In the regression models shown in the previous sections, the parameters in the regression equations are derived based on the Least Squares (LS) and Maximum Likelihood (ML) estimating theory, and the regression line is fitted through the performance data (Ratkowsky 1983). When track and field performance data and historic year were used in the regression modeling, the residual variance or error variance is always nonzero magnitude. Therefore, precise prediction of the track and field performances is not what the prediction models can offer. Due to this reason, it may not be appropriate to evaluate the prediction models by comparing the predicted world records in order to see which model makes more accurate predictions. However, it is a valid comparison in which a model has a better fit to the data measured by some goodness of fit indexes (see Liu and Schutz 1998 for a detailed model comparisons).

In summary, it is a fallacy to use the prediction models to predict accurate track and field performance in the future. The prediction models worked well to show the historic trend of the track and field performances and to predict the general trend of the performances in the future.

The factors that affect goodness of fit about the prediction models

Speaking about the goodness of fit, there are many external factors that affect prediction models. Liu (2002) and Schutz and Liu (1998) compared the mathematical models in the literature and pointed out that the variability among the model predictions is large because different models have utilized different data sets and assumptions.

In fact, many factors affect accuracy and validity of prediction models. Internal factors include the validity of the data and the methods used for model development. External factors that are out of researchers' control include potentials and limitations of the human body, psychological effect in competitions, science and technology influence, performance-enhancing drugs, and political and economic situations in a specific time period. Because the asymptotic mathematical models have treated the external effects as random variables for evaluating and predicting the longitudinal trend of the track and field, it is more valid and preferable than other models in the literature. Liu (2002) summarized the external factors and discussed how these factors would affect future track and field performances and prediction models.

4 Conclusion

This chapter focused on the discussion of prediction models in track and field performances and their promises and fallacies. Although one might wish to use the prediction models to predict the specific magnitude of future world records, it is not what regression models promise to provide. The promise of all the regression type of models in track and field is to describe the past performance patterns over time and predict the trend for future performances. After the asymptotic level is reached, the regression models cannot be used for the evaluation and prediction of the track and field performances. Some new approaches such as the random sampling model (Liu 1993, Schutz and Liu 1998) should be employed for this purpose. The future holds surprising changes, in equipment, in technology, in performance enhancing drugs, in social and economic status, etc. It is still true that predicting future track and field performances is good fun, and probably about as exact as some of the uses that mathematics is put to elsewhere (Deakin 1961).

References

1. Ballerini, R. and S. I. Resnick. Records from improving populations. *Journal of Applied Probability,* 22: 487-502, 1985.
2. Ballerini, R. Records, in the presence of a linear trend. *Advances in Applied Probability,* 19: 801-828, 1987.
3. Chatterjee, S., and S. Chatterjee. New Lamps for old: An exploratory analysis of running times in Olympic Games. *Applied Statistics,* 31, 14-22, 1982.
4. Craig, A. B. Evaluation and predictions of world running and swimming records. *Journal of Sport Medicine,* 3: 14-21, 1963.
5. Deakin, B. Mathematics of athletic performance. *Matrix,* Melbourne University Mathematical Society Magazine, issue 4, 1961.
6. Deakin, B. Estimating bounds on athletic performance. *The Mathematical Gazette,* 51: 100-103, 1967.
7. Dyer, K.F. *Catching up the men: Women in sport.* Junction Books, 1982.
8. Francis, A. W. Running records. *Science,* 98: 315-316, 1943.
9. Furusawa, K., A. V. Hill, J. L. Parkinson The dynamics of sprint running. *Proceedings Royal Society,* B(102): 29-42, 1927.
10. Gembris, D., J.G. Taylor, and D. Suter. Trends and random fluctuations in athletics. *Nature,* 417: May 506, 2002.
11. Hill, A. V. The physiological basis of athletic records. *British Association for the Advancement of Science, Report of the 93rd Meeting:* 156-173, 1925.
12. Keller, J. B. Optimal velocity in a race. *The American Mathematical Monthly,* 81: 474-480, 1974.
13. Kennelly, A. An approximate law of fatigue in speeds of racing animals. *Proceedings of the American Academy of Arts of Scienc,* 42: 275-331, 1906.
14. Liu, Y. *Evaluation and prediction of world records and ultimate performance in track and field.* Unpublished Master Thesis, University of British Columbia, 1993.
15. Liu, Y. and R. W. Schutz. Prediction models for track and field performances. *Measurement in Physical Education and Exercise Science*, 2: 205-223, 1998.
16. Liu, Y. Track and field performances and competitions in the new millennium: Will it be a thematic change? *International Sports Journal*, 6: 29-35, 2002.

17. Lloyd, B. B. The energetics of running: An analysis of world records. *British Association for the Advancement of Science, Report of the 133rd Meeting:* 515-530, 1965.
18. Lucy, L. B. Future progress in the mile. *IOTA,* 1: 8-11, 1958.
19. Mognoni, P., C. Lafortuna, G. Russo, and A. Minetti. An analysis of world records in three types of locomotion. *European Journal of Applied Physiology,* 49: 287-299, 1982.
20. Peronnet, F. and G. Thibault. Mathematical analysis of running performance and world running records. *Journal of Applied Physiology,* 67: 453-465, 1989.
21. Ratkowsky, D. A. *Nonlinear Regression Modeling.* Marcel Dekker, Inc, 1983.
22. Riegel, P. S. Athletic records and human endurance. *American Scientist,* 69: 285290, 1981.
23. Ryder, H.W., H.J. Carr, P. Herget. Future performance in footracing. Scientific American, 234: 109-119, 1976.
24. Schutz, R. and Y. Liu. Statistical Modeling in Track and Field. In Bennett, J. (editor). *Statistics in Sport,* pp. 173-197. London: Arnold. 1998.
25. Schutz, R. and Y. Liu. Track and field world records: optimal models, datasets and projections. *American Statistical Association 1993 Proceedings of the Section on Statistics in Sport:* 52-57, 1993.
26. Schutz, R. W. and J. P.McBryde. The prediction of ultimate track and field performance: Past, present, future. *Proceedings of the FISU Conference Universiade' 83 in Association with the Xth HISPA Congress*: 498-515, 1983.
27. Senator, M. Extending the theory of dash running. *Journal of Biomechanical Engineering,* 104: 209-213, 1982.
28. Stefani, R.T. Trends in Olympic winning performances. *Athletic Journal,* Dec.: 44-46, 1977.
29. Ward-Smith, A. J. 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.
30. Whipp, B.J. and S. A. Ward. Will women soon outrun men? *Nature*, 335(2): January 25, 1992.