A deep dive into the replication, repeatability, and future utilization of the critical running speed model for predicting distance race performances

Brent Scott

Introduction

The beauty of the sport of running is in its simplicity - the winner is the one who completes the distance in the least time. Interestingly, the objectively fastest runner does not always win the race despite previously having the fastest time for the given race distance. Moreover, even having measures of athletes' maximal oxygen consumption (VO^2_{max}), which is often used as a measure of cardio-respiratory fitness, is not sufficient information to solely predict distance runner performance. Additional exercise physiology laboratory tests can contribute to performance predictions by allowing for more control over the physiological variables that govern performance (lactate threshold, aneorobic power, etc.). Notably, in 1991 Michael Joyner used some of the physiological determinants of running performance (VO2max, lactate threshold, and running economy) to model and predict that a human could run a marathon in under 2 hours, almost 20 years before Eliud Kipchoge ran 1:59:40. Ultimately, the physiology underlying distance track races (~800-10000m) performance is complicated as both aerobic and anaerobic energy pathways are utilized to meet the energy demands of the activity so the prediction of athletes' performances in these races becomes increasingly more complicated than compared to marathon distances where runners are exercising at intensities at or below their VO2max and lactate thresholds. Additional issues arise when addressing the concern of difficulty in providing training tools for coaches of high-school and collegiate athletes who do not have the time, money, or resources to readily measure all their athletes' physiological parameters in a laboratory setting. The Critical Speed Model (CS-Model) has been proposed to provide reliable methods at predicting distance running performance (@Jones2017). The CS-Model uses linear regression to model an athlete's time (x-axis) by the respective race distance (y-axis). Since the CS-Model is linear, the equation is of familiar form, y = mx + b, where the slope (m) is the athlete's critical speed in meters/second, and the y-intercept (b) is referred to as "D' Prime" in meters. D' Prime is a measure of anaerobic performance and theoretically describes how far a runner can sustain a given pace that is above their critical speed. Once D' Prime has been utilized a runner's maximum speed that can be sustained is their critical speed, at least until D' Prime can become "reconstituted" or restored. More recently, a so-called "D' Balance Model" has been shown to be highly correlated with performance in track distance races (even more so than critical speed or D' Prime) (@kirby2021). With these findings it was suggested that the "D' Balance Model" can become a useful tool for athletes and coaches to modify training and race preparations. However, the "D' Balance Model" has yet to be verified, replicated, or repeated outside the context of the original source. Additionally, for any potential tool to really be useful in practice it needs to be readily and easily accessible outside the scope of the original research. As such, here we sought to examine the replication of the "D' Balance Model" using the same 2017 World Championship Men's 5000m and 10,000m data, the repeatability of the model to predict the outcomes of other races (2020 Olympic Women's 5000m Final), and provide a potential framework for the future usability of the model for athletes and coaches.

Methods

Analysis

Analysis of all data was performed using R (version 4.1.2) in RStudio (version 1.4.1717). Source code for this entire document is available https://github.com/brentscott93/deep-dive-697. This repository contains all data, code, and figures to replicate this document which was originally generated via RMarkdown. Core analysis functionality required to replicate the analysis, including both the Critical Speed Model and D' Balance calculations, have convenient wrappers available in the {kin697u} R package which can be installed from Github. The {kin697u} package functions are used within this document as an example. Refer to the source code to see the internal working of the functions and package. The analysis of D' Balance is described further in the "Results" section.

Replication of Kirby 2021 data

In order to assess the replication of the "D' Balance Model" as utilized by the original authors, the publication figures from Kirby 2021 were digitized using *Engauge Digitizer* (Winchen et al.) in order to best estimate the data values from Figure 1 in Kirby 2021. This provided us with more reliable and repeatable estimates than choosing "by-eye" for the values of lap speed for the 5k and 10k (Kirby Figure 1a and 1c), and for the D' Balance curves (Kirby Figure 1b and 1d). To digitize an image three axis points are selected as anchors (indicated as red "+") and the x-y coordinates are calibrated accordingly. Desired data values are selected (blue "+") and their values were estimated by the program and saved as .csv files in the project's directory. This was performed for each of the four panels in Figure 1 of Kirby 2021. Critical speed and D'Prime values were copied verbatim from Tables 1 and 2 in Kirby 2021.
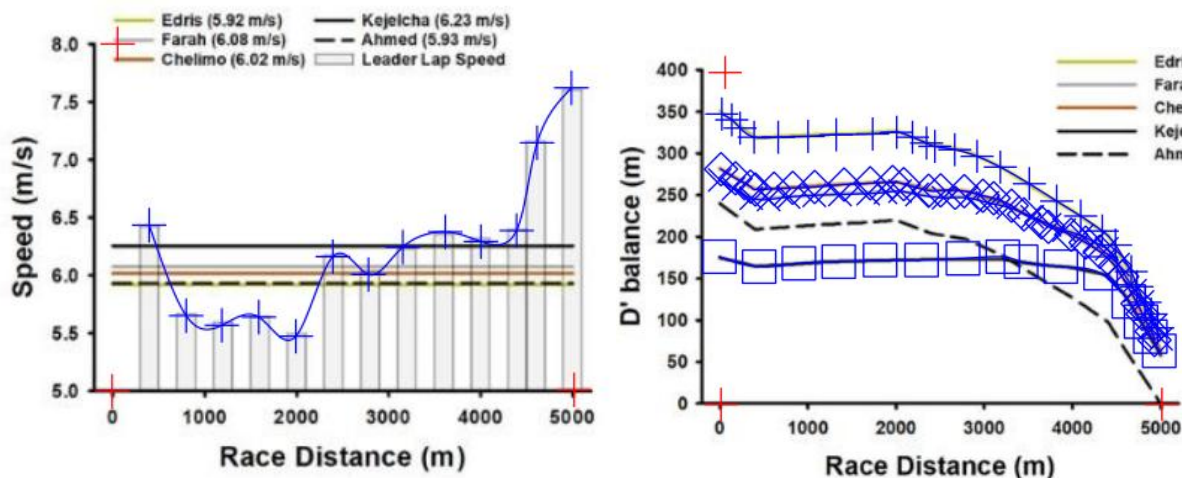

Figure 1: Digitized images from Kirby 2021. Red 'plus' signs indicate calibration points and blue markers are used to estimate data values.

Obtaining data from the 2020 Olympic Women's 5000m

To provide more frequent lap splits than provided on the official race results (1000m), the 2021 Olympic Women's 5000m was streamed on YouTube and lap splits were manually recorded. Unfortunately, the live broadcast did not always allow for consistent lap splits as other events or commercials interrupted the race coverage. Nonetheless, we were able to obtain reasonable estimation of lap distances, times, and race progression. The lap speed was calculated from the lap leader split by dividing the relative distance covered in meters by the relative time of the split in seconds.

Table 1: 2020 Olympic Women's 5k

Distance (m)	Lap Speed (m/s)
200	5.14
1000	5.64
1400	5.56
1800	5.51
2000	5.71
2200	5.50
2600	5.73
3000	5.45
3400	5.62
3800	5.64
4000	5.68
4200	5.81
4600	5.89
4800	6.76
5000	7.20

Critical speed and D' Prime were estimated for the top 4 finishers in the Women's 5000m by fitting the linear Critical Speed Model to the athlete's personal bests found on the internet (Wikipedia).

Table 2: 2020 Olympic Women's Top Finishers CS/D'

Athlete	CS (m/s)	D' (m)
Hassan	5.63	169
Obiri	5.47	212
Tsegay	5.53	197
Tirop	5.37	256

Results

Replication of Kirby 2021 data

Ultimately, we were able to replicate the results from the D' Balance model presented in Kirby 2021 with the same 2017 World Championship data; however, the equations that were provided coupled with the lack of description for application of the equation made this a challenging task as we were initially unable to calculate similar results solely using the original text as guidance. Here, we several steps and propose the solution we came up with which permitted replication of the results. The D' Balance model as presented in Kirby 2021 is based on work from Skiba et al. 2015 and describes D'balance as:

$$D'Balance = D'_0 - D'_{exp} \cdot e^{\frac{-D_{CS} \cdot t}{D'_0}}$$ (1)

where D'_0 is the initial D', D'_{exp} the expended D', D_{CS} is current speed minus the CS, and t is time in seconds. Additionally D'_{exp} is described as being:

$$D'_{exp} = D_{CS} \times t + D'_{prior}$$ (2)

where D'_{prior} is the preceding D'. This is the extent of the original instruction with no mention of how this was iteratively implemented for every athlete on every split. Since D' Balance is the relationship between

using (and draining) D' versus "reconstituting" (and gaining) D' we propose a model based on those of Kirby 2021 that is simply:

$$D'Balance = \begin{cases} D'_{drain}, & \text{if } Speed > CS \\ D'_{gain}, & \text{else} \end{cases} \tag{3}$$

This model describes that when the current running speed is greater than an athlete's critical speed they will be draining (i.e losing) D' and if the current running speed is slower than critical speed then the athlete can be gaining (i.e. "reconstiuting") their D'. Furthmore, the clarifiying use of the "if-else" cases provides useful intructions into how these equation can be programatically applied which is invaluable in the replication procedure. In this model D'$_{drain}$ is simply a renamed form of the D'$_{exp}$ term from Kirby 2021:

$$D'_{drain} = D_{CS} \times t + D'_{sol} \tag{4}$$

where D$_{CS}$ is the current lap speed minus the critical speed, t is the *relative lap time* (in seconds) and D'$_{sol}$ is the currently available D' at the **S**tart **O**f **L**ap. Simply, the calculation of D'$_{drain}$ describes the process of losing D' that is dependent on how great the speed is over the CS, how much D' there was to start with, and the duration of time spent at the speed.

D'$_{gain}$ is of similar form to the originally proposed D' Balance equation and describes the process of recovering back D' when the speed is less than critical speed:

$$D'_{gain} = D'_0 - D'_{expended} \cdot e^{\frac{-D_{CS} \cdot t}{D'_0}} \tag{5}$$

The important difference here is the introduction of D'$_{expended}$ which is calculated by subtracting the D' currently available at the start of lap (D'$_{sol}$) from the runner's initial D' value:

$$D'_{expended} = D'_0 - D'_{sol} \tag{6}$$

With the above equations defined including the scenarios in which which specific case an equation should be used we were able to replicate the Kirby 2021 data from Figure 1.

Repeatability with the 2020 Olympic Women's 5000m

Once we described the equations and procedure above we implemented the functionlity into an R package {kin697u} which provides convenient wrapper functions in order to increase to reproducibility of our proposed model. Using our R package we performed the same analysis using data from the 2020 Olympic Women's 5000m.

The results from the 2020 Olympic Women's 5000m suggest that the Bronze medalist (Tsegay) should have had potential for a silver medal. There could be several explanations for this result, one being the simulations assume all runners are at the same place (meter mark) at the same time and do not take in account distance from leader at start of each lap.

Future usability of the model

So far we have shown that with slight modifications the D' Balance Model can be replicated and repeated with relatively good performance at predicting the finishing position as was shown by Kirby 2021. Despite the advances in reproducibility made thus far with the implementation of the R package, this analysis still remains un-accessible to the average runner or trainer (unless they are also an R programmer). To decrease the entry barrier for performing this analysis we created a web-browser based application, WOUTER, using
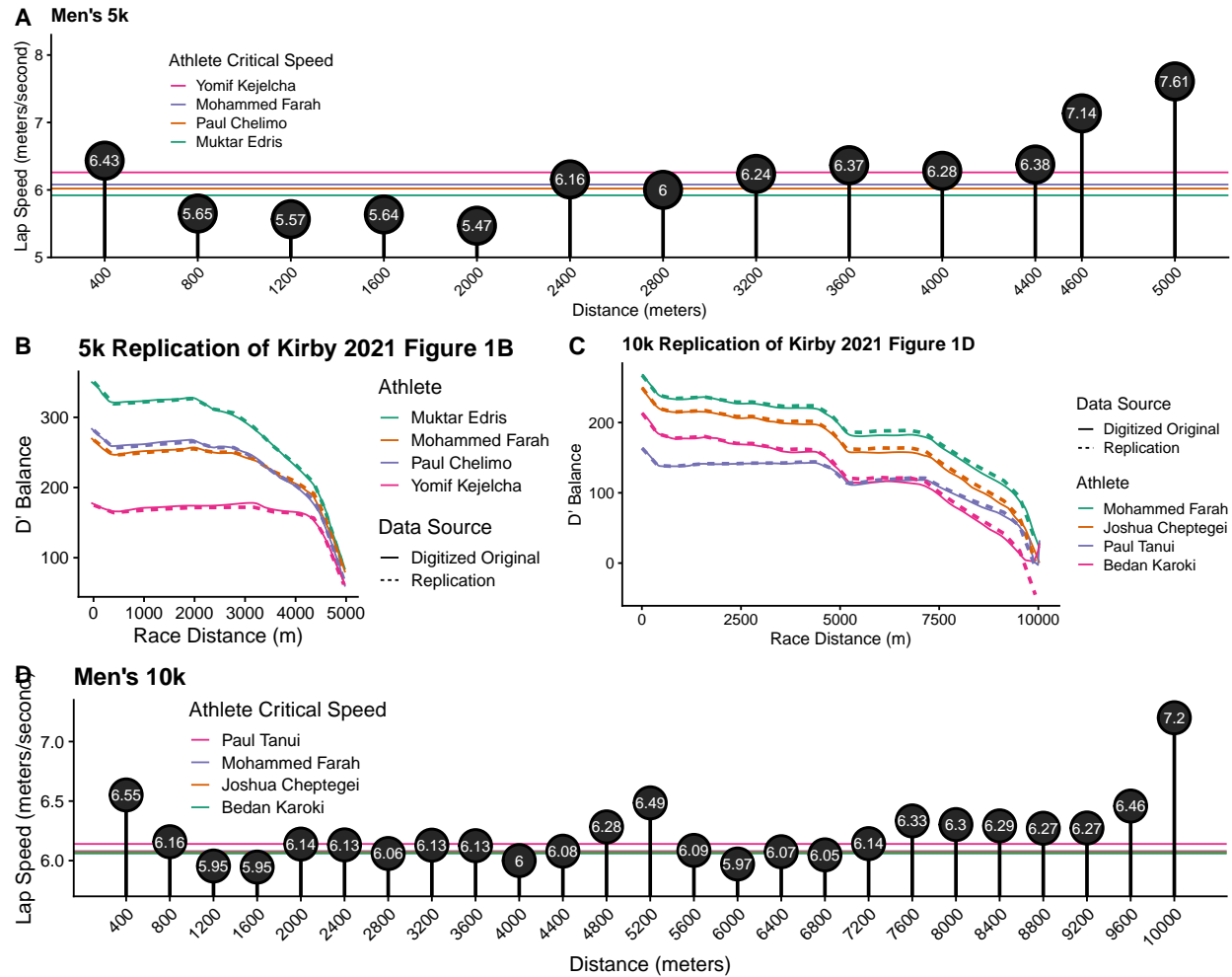
Figure 2: Replication of Kirby 2021 Figure 1. A&D) Lap speed throughout the races overlayed with athletes CS. B&C) Replication of the D' Balance model. Solid lines are the original data from the digitized images adn the dotted lines are the replicated results with the use of the new model presented here.

A **2021 Olympic Women's 5000m**

B

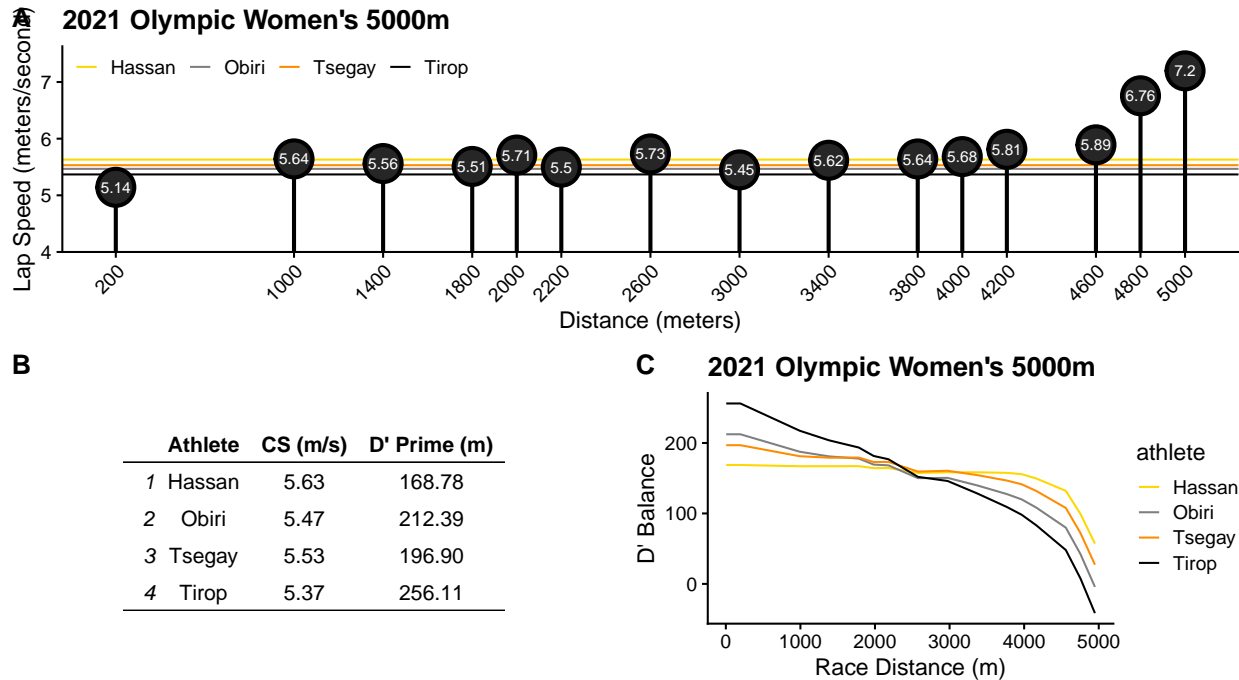|	Athlete	CS (m/s)	D' Prime (m)
1	Hassan	5.63	168.78
2	Obiri	5.47	212.39
3	Tsegay	5.53	196.90
4	Tirop	5.37	256.11

C **2021 Olympic Women's 5000m**

Figure 3: Results from 2021 Olympic Women's 5000m. A) Lap speed throughout the duration of the race. B) Critical speed an D' table. C) D' balance model. Lines colored to represent ultimate finish position.

the R-Shiny framework (Chang et al. 2021) to provide an easy-to-use graphical user interface (GUI) which provides a "point-and-click" experience for novice users that leverages the abilities of our {kin697u} R package. WOUTER is a recursive acronym which stands for Wouter Online Utility for Training Endurance Running. WOUTER can be accessed at [https://brentscott93.shinyapps.io/WOUTER]

Conclusions & Final Remarks

I probably should have spend less time coding and building apps, and **more** time writing/doing the actual assignment. . .