Mathematisches
Forschungsinstitut
Oberwolfach

Oberwolfach
Preprints

DUNCAN A. J. BLYTHE AND FRANZ J. KIRÁLY

Prediction and Quantification of Individual Athletic Performance

Oberwolfach Preprints (OWP)

Starting in 2007, the MFO publishes a preprint series which mainly contains research results related to a longer stay in Oberwolfach. In particular, this concerns the Research in Pairs-Programme (RiP) and the Oberwolfach-Leibniz-Fellows (OWLF), but this can also include an Oberwolfach Lecture, for example.

A preprint can have a size from 1 - 200 pages, and the MFO will publish it on its website as well as by hard copy. Every RiP group or Oberwolfach-Leibniz-Fellow may receive on request 30 free hard copies (DIN A4, black and white copy) by surface mail.

Of course, the full copy right is left to the authors. The MFO only needs the right to publish it on its website *www.mfo.de* as a documentation of the research work done at the MFO, which you are accepting by sending us your file.

In case of interest, please send a **pdf file** of your preprint by email to *rip@mfo.de* or *owlf@mfo.de*, respectively. The file should be sent to the MFO within 12 months after your stay as RiP or OWLF at the MFO.

There are no requirements for the format of the preprint, except that the introduction should contain a short appreciation and that the paper size (respectively format) should be DIN A4, "letter" or "article".

On the front page of the hard copies, which contains the logo of the MFO, title and authors, we shall add a running number (20XX - XX).

We cordially invite the researchers within the RiP or OWLF programme to make use of this offer and would like to thank you in advance for your cooperation.

Imprint:

Prediction and Quantification of Individual Athletic Performance

Duncan A.J. Blythe * [1,2] and Franz J. Király † [3]

[1] Bernstein Center for Computational Neuroscience Berlin,
Humboldt-Universität zu Berlin, Philippstr. 13, Haus 6, 10115 Berlin, Germany
[2]Department of Electrical Engineering and Computer Science,
Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
[3] Department of Statistical Science, University College London,
Gower Street, London WC1E 6BT, United Kingdom

Abstract

We present a novel, quantitative view on the human athletic performance of individuals. We obtain a predictor for athletic running performances, a parsimonious model, and a training state summary consisting of three numbers, by application of modern validation techniques and recent advances in machine learning to the thepowerof10 database of British athletes' performances (164,746 individuals, 1,417,432 performances).

Our predictor achieves a low average prediction error (out-of-sample), e.g., 3.6 min on elite Marathon performances, and a lower error than the state-of-the-art in performance prediction (30% improvement, RMSE). We are also the first to report on a systematic comparison of predictors for athletic running performance.

Our model has three parameters per athlete, and three components which are the same for all athletes. The first component of the model corresponds to a power law with exponent dependent on the athlete which achieves a better goodness-of-fit than known power laws in athletics. Many documented phenomena in quantitative sports science, such as the form of scoring tables, the success of existing prediction methods including Riegel's formula, the Purdy points scheme, the power law for world records performances and the broken power law for world record speeds may be explained on the basis of our findings in a unified way.

We provide strong evidence that the three parameters per athlete are related to physiological and/or behavioural parameters, such as training state, event specialization and age, which allows us to derive novel physiological hypotheses relating to athletic performance. We conjecture on this basis that our findings will be vital in exercise physiology, race planning, the study of aging and training regime design.

Note. *This manuscript is work in progress and has not yet been reviewed by an independent panel of experts. Once the manuscript has been reviewed and accepted by such a panel, this note will be removed. Until then, we would advise the reader to treat the presented results as preliminary and not to understand or present our findings as scientific fact but merely as a basis for scientific discussion.*

An overview on athletic performance prediction and our contributions

Performance prediction and modeling are cornerstones of sports medicine, essential in training and assessment of athletes with implications beyond sport, for example in the understanding of aging, muscle physiology,

---

*duncan.blythe@bccn-berlin.de

†f.kiraly@ucl.ac.uk

*,† both authors contributed equally to this work

and the study of the cardiovascular system. Existing research on athletic performance focuses either on (A) explaining world records [28, 19, 36, 22, 39, 17], (B) equivalent scoring [33, 34], or (C) modelling of individual physiology [20, 2, 43, 29, 1, 23, 30, 41]. Currently, however, there is no parsimonious model which simultaneously explains individual physiology (C) and collective performance (A,B).

We present such a model, a non-linear low-rank model derived from a database of UK athletes. It levers an individual power law which explains the power laws known to apply to world records, and which allows us to derive athlete-individual training parameters from prior performances data. Performance predictions obtained using our approach are the most accurate to date, with an average prediction error of under 4 minutes (2% rel.MAE and 3% rel.RMSE out-of-sample, see Tables 5, 6 and appendix S.I.b) for elite performances. We anticipate that our framework will allow us to leverage existing insights in the study of world record performances and sports medicine for an improved understanding of human physiology.

Our work builds on the three major research strands in prediction and modeling of running performance, which we briefly summarize:

(A) **Power law models of performance** posit a power law dependence $t = c \cdot s^{\alpha}$ between the duration of the distance run t and the distance s, for constants c and α. This is equivalent to assuming a linear dependence $\log t = \alpha \log s + \log c$ of log-time on log-distance. Power law models have been known to describe world record performances across sports for over a century [24], and have been applied extensively to running performance [28, 19, 36, 22, 39, 17]. These power laws have been applied by practitioners for prediction: the Riegel formula [35] predicts performance by fitting c to each athlete and fixing $\alpha = 1.06$ (derived from world-record performances). The power law approach has the benefit of *modelling* performances in a scientifically parsimonious way.

(B) **Scoring tables**, such as those of the international association of athletics federations (IAAF), render performances over disparate distances comparable by presenting them on a single scale. These tables have been published by sports associations for almost a century [32] and catalogue, rather than model, performances of equivalent standard. Performance predictions may be obtained from scoring tables by forecasting a time with the same score as an existing attempt, as implemented in the popular Purdy Points scheme [33, 34]. The scoring table approach has the benefit of *describing* performances in an empirically accurate way.

(C) **Explicit modeling of performance related physiology** is an active subfield of sports science. Several physiological parameters are known to be related to athletic performance; these include maximal oxygen uptake ($\dot{V}O_2$-max) and critical speed (speed at $\dot{V}O_2$-max) [20, 2], blood lactate concentration, and the anaerobic threshold [43, 6]. Physiological parameters may be used (C.i) to make direct predictions when clinical measurements are available [29, 1, 9], or (C.ii) to obtain theoretical models describing physiological processes [23, 30, 41, 13]. These approaches have the benefit of *explaining* performances physiologically.

All three approaches (A),(B),(C) have appealing properties, as explained above, but none provides a complete treatment of athletic performance prediction: (A) individual performances do not follow the parsimonious power law perfectly; (B) the empirically accurate scoring tables do not provide a simple interpretable relationship. *Neither* (A) nor (B) can deal with the fact that athletes may differ from one another in multiple ways. The clinical measurements in (C.i) are informative but usually available only for a few select athletes, typically at most a few dozen (as opposed to the 164,746 considered in our study). The interpretable models in (C.ii) are usually designed not with the aim of predicting performance but to explain physiology or to estimate physiological parameters from performances; thus these methods are not directly applicable without additional work.

The approach we present unifies the desirable properties of (A),(B) and (C), while avoiding the aforementioned shortcomings. We obtain (A) a parsimonious model for individual athletic performance that is (B) empirically derived from a large database of UK athletes. It yields the best performance predictions to date (2% average error for elite athletes on all events, average error 3-4 min for Marathon, see Table 6) and (C) unveils hidden descriptors for individuals which we find to be related to physiological characteristics.

Our approach bases predictions on *Local Matrix Completion* (LMC), a machine learning technique which posits the existence of a small number of explanatory variables which describe the performance of individual athletes. Application of LMC to a database of athletes allows us, in a second step, to derive a parsimonious physiological model describing athletic performance of *individual* athletes. We discover that a *three number-summary* for each individual explains performance over the full range of distances from 100m to the Marathon. The three-number-summary relates to: (1) the endurance of an athlete, (2) the relative balance
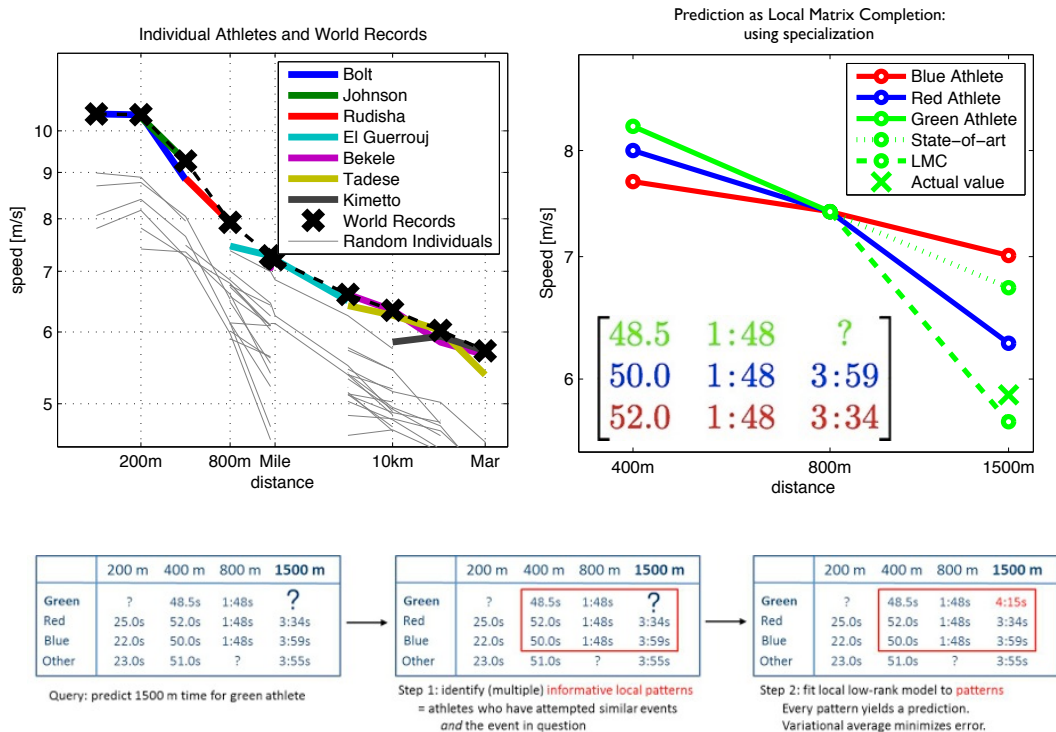
Figure 1: Non-linear deviation from the power law in individuals as central phenomenon. Top left: performances of world record holders and a selection of random athletes. Curves labelled by athletes are their known best performances (y-axis) at that event (x-axis). Black crosses are world record performances. Individual performances deviate non-linearly from the world record power law. Top right: a good model should take into account specialization, illustration by example. Hypothetical performance curves of three athletes, green, red and blue are shown, the task is to predict green on 1500m from all other performances. Dotted green lines are predictions. State-of-art methods such as Riegel or Purdy predict green performance on 1500m close to blue and red; a realistic predictor for 1500m performance of green - such as LMC - will predict that green is outperformed by red and blue on 1500m; since blue and red being worse on 400m indicates that out of the three athletes, green specializes most on shorter distances. Bottom: using local matrix completion as a mathematical prediction principle by filling in an entry in a (3×3) sub-pattern. Schematic illustration of the algorithm.

between speed and endurance, and (3) specialization over middle distances. The first number explains most of the individual differences over distances greater than 800m, and may be interpreted as the exponent of *an individual power law for each athlete*, which holds with remarkably high precision, on average. The other two numbers describe individual, non-linear corrections to this individual power law. Vitally, we show that the individual power law with its non-linear corrections reflects the data more accurately than the power law for world records. We anticipate that the individual power law and three-number summary will allow for exact quantitative assessment in the science of running and related sports.

Local Matrix Completion and the Low-Rank Model

It is well known that world records over distinct distances are held by distinct athletes—no one single athlete holds all running world records. Since world record data obey an approximate power law (see above), this implies that the individual performance of each athlete deviates from this power law. The left top panel of Figure 1 displays world records and the corresponding individual performances of world record holders in logarithmic coordinates—an exact power law would follow a straight line. The world records align closely to a straight line, while individuals deviate non-linearly. Notable is also the kink in the world records which makes them deviate from an exact straight line, yielding a "broken power law" for world records [39].

Any model for individual performances must model this individual, non-linear variation, and will, optimally, explain the broken power law observed for world records as an epiphenomenon of such variation over

individuals. In the following paragraphs we explain how the LMC scheme captures individual variation in a typical scenario.

Consider three athletes (taken from the database) as shown in the right top panel of Figure 1. The 1500m performance of the green athlete is not known and is to be predicted. All three athletes, green, blue and red, achieve similar performance on 800m. Any classical method for performance prediction which only takes this information into account will predict that green performs similarly on 1500m to the blue and the red, e.g., somewhere in-between. However, this is unrealistic, since it does not take into account event specialization: looking at the 400m performance, one can see that the red athlete is slowest over short distances, followed by the blue and then by the green athlete. Thus the red athlete is more an endurance type athlete than the blue, and blue is to a greater extent a speed type athlete than red; green specializes even more on speed than both red and blue. Using this additional information leads to the more realistic prediction that the green athlete will be out-performed by red and blue on 1500m. Supplementary analysis (S.IV) validates that the phenomenon presented in the example is prevalent throughout the data set.

LMC is a quantitative method for taking into account this event specialization. A schematic overview of the simplest variant is displayed in the bottom panel of Figure 1: to predict an event for an athlete (figure: 1500m for green) we find a 3-by-3-pattern of performances, denoted by A, with exactly one missing entry - this means the two other athletes (figure: red and blue) have attempted similar events and have data available. Explanation of the green athlete's curve by the red and the blue is mathematically modelled by demanding that the data of the green athlete is given as a weighted sum of the data of the red and the blue; i.e., more mathematically, the green row is a linear combination of the blue and the red row. By a classical result in matrix algebra, the green row is a linear combination of red and blue whenever the *determinant* of A, a polynomial function in the entries of A, vanishes; i.e., $\det(A) = 0$.

A prediction is made by solving the equation $\det(A) = 0$ for "?". To increase accuracy, candidate solutions from multiple 3-by-3-patterns (obtained from many triples of athletes) are averaged in a way that minimizes the expected error in approximation. We will consider variants of the algorithm which use n-by-n-patterns, n corresponding to the complexity of the model (we later show $n = 4$ to be optimal). See the methods appendix for an exact description of the algorithm used.

The LMC prediction scheme is an instance of the more general local low-rank matrix completion framework introduced in [26], here applied to performances in the form of a numerical table (or matrix) with columns corresponding to events and rows to athletes. The cited framework is the first matrix completion algorithm which allows prediction of single missing entries as opposed to all entries. While matrix completion has proved vital in predicting consumer behaviour and recommender systems, we find that existing approaches which predict all entries at once cannot cope with the non-uniform missingness and the noise associated with performance prediction in the same way as LMC can (see findings and supplement S.II.a). See the methods appendix for more details of the method and an exact description.

In a second step, we use the LMC scheme to fill in all missing performances (over all events considered—100m, 200m etc.) and obtain a parsimonious low-rank model - we remark that first filling in the entries with LMC and only then fitting the model is crucial due to data which is non-uniformly missing (see supplement S.II.a). The low-rank model explains individual running times t in terms of distance s and has the form:

$$\log t = \lambda_1 f_1(s) + \lambda_2 f_2(s) + \cdots + \lambda_r f_r(s), \tag{1}$$

with components $f_1, f_2, \ldots, f_r$ that are the same for every athlete, and coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$ which summarize the athlete under consideration. The number of components and coefficients r is known as the *rank* of the model and measures its complexity. The Riegel power law is a very special case, demanding that $\log t = 1.06 \log s + c$; that is, a rank 2 model with $\lambda_1 = 1.06$ for every athlete, $f_1(s) = \log s$, and an athlete-specific constant $\lambda_2 f_2(s) = c$. Our analyses will show that the best model has rank $r = 3$ (meaning above we consider patterns or matrices of size $n \times n = 4$ since above $n = r + 1$). This means that the model has $r = three$ universal components $f_1(s), f_2(s), f_3(s)$, and every athlete is described by their individual three-coefficient-summary $\lambda_1, \lambda_2, \lambda_3$. Remarkably, we find that $f_1(s) = \log s$ (for a suitable unit of distance/time, see S.II.b), yielding an individual power law; the corresponding coefficient λ_1 thus has the natural interpretation as an individual power law exponent.

Table 11 contains the exact form for the components f_1, f_2, f_3 in our model; they are also displayed in Figure 2 top left. More details on how to obtain components and coefficients can be found in the methods section, "obtaining the low-rank components and coefficients", and in supplementary experiment (S.II.b).

Data Set, Analyses and Model Validation

The basis for our analyses is the online database `www.thepowerof10.info`, which catalogues British individuals' performances achieved in officially ratified athletics competitions since 1954. The excerpt we consider contains performances between 1954 and August 3, 2013. It contains (after error removal) records of 164,746 individuals of both genders, ranging from the amateur to the elite, young to old, comprising a total of 1,417,432 individual performances over 10 different distances: 100m, 200m, 400m, 800m, 1500m, the Mile, 5km, 10km, Half-Marathon, Marathon (42,195m). All British records over the distances considered are contained in the dataset; the 95th percentile for the 100m, 1500m and Marathon are 15.9, 6:06.5 and 6:15:34, respectively. As performances for the two genders distribute differently, we present only results on the subset of 101,775 male athletes in the main corpus of the manuscript; female athletes and further subgroup analyses are considered in the supplementary results. The data set is available upon request, subject to approval by British Athletics. Full code of our analyses can be obtained from [download link will be provided here after acceptance of the manuscript].

Adhering to state-of-the-art statistical practice (see [14, 27, 15, 8]), all prediction methods are validated *out-of-sample*, i.e., by using only a subset of the data for estimation of parameters (training set) and computing the error on predictions made for a distinct subset (validation or test set). As error measures, we use the root mean squared error (RMSE) and the mean absolute error (MAE), estimated by leave-one-out validation for 1000 single performances omitted at random.

We would like to stress that out-of-sample prediction error is the correct way to evaluate the quality of *prediction*, as opposed to merely reporting goodness-of-fit in-sample; since outputting an estimate for an instance that the method has already seen does not qualify as prediction.

More details on the data set and our validation setup can be found in the supplementary material.

Findings on the UK athletes data set

(I) Prediction accuracy. We evaluate prediction accuracy of ten methods, including our proposed method, LMC. We include, as *naive baselines:* (1.a) imputing the event mean, (1.b) imputing the average of the k-nearest neighbours; as representative of the *state-of-the-art in quantitative sports science:* (2.a) the Riegel formula, (2.b) a power law predictor with exponent estimated from the data, which is the same for all athletes, (2.c) a power law predictor with exponent estimated from the data, with one exponent per athlete, (2.d) the Purdy points scheme [33]; as representatives for the *state-of-the-art in matrix completion:* (3.a) imputation by expectation maximization on a multivariate Gaussian [12] (3.b) nuclear norm minimization [10, 11].
We instantiate our *low-rank local matrix completion* (LMC) in two variants: (4.a) rank 1, and (4.b) rank 2.

Methods (1.a), (1.b), (2.a), (2.b), (2.d), (4.a) require at least one observed performance per athlete, methods (2.c), (4.b) require at least two observed performances in distinct events. Methods (3.a), (3.b) will return a result for any number of observed performances (including zero). Prediction accuracy is therefore measured by evaluating the RMSE and MAE out-of-sample on the athletes who have attempted at least three distances, so that the two necessary performances remain when one is removed for leave-one-out validation. Prediction is further restricted to the best 95-percentile of athletes (measured by performance in the best event) to reduce the effect of outliers. Whenever the method demands that the predicting events need to be specified, the events which are closest in log-distance to the event to be predicted are taken. The accuracy of predicting time (normalized w.r.t. the event mean), log-time, and speed are measured. We repeat this validation setup for the year of best performance and a random calendar year. Moreover, for completeness and comparison we treat 2 additional cases: the top 25% of athletes and athletes who have attempted at least 4 events, each in log time. More details on methods and validation are presented in the methods appendix.

The results are displayed in Table 2 (RMSE of log-time prediction) and supplementary Tables 3 (MAE of log-time prediction), 5 (rel.RMSE of time prediction) and 6 (rel. MAE of time prediction). Of all benchmarks, k-nearest neighbours (1.b), Purdy points (2.d) and Expectation Maximization (3.a) perform best. LMC rank 2 substantially outperforms k-nearest neighbours, Purdy points and Expectation Maximization (two-sided Wilcoxon signed-rank test significant on the validation samples of absolute prediction errors; $p \leq 2.0$e-8 on top 95% in log-time and $p \leq 1.4$e-11 for top 25% in log-time); rank 1 outperforms Purdy points on the year of best performance data ($p \leq 3.0$e-3) for the best athletes, and is on a par on athletes up to the 95th percentile.

Both rank 1 and 2 outperform the power law models ($p \leq$1.1e-42), the improvement in RMSE of LMC rank 2 over the power law models reaches over 50% for data from the fastest 25% of athletes.

(II) The rank (number of components) of the model. Paragraph (I) establishes that LMC is the best method for prediction. LMC assumes a fixed number of prototypical athletes, viz. the rank r, which is the complexity parameter of the model. We establish the optimal rank by comparing prediction accuracy of LMC with different ranks. The rank r algorithm needs r attempted events for prediction, thus $r+1$ observed events are needed for validation. Table 7 displays prediction accuracies for LMC ranks $r = 1$ to $r = 4$, on the athletes who have attempted $k > r$ events, for all $k \leq 5$. The data is restricted to the top 25% in the year of best performance in order to obtain a high signal to noise ratio. We observe that rank 3 outperforms all other ranks, when applicable; rank 2 always outperforms rank 1 (both $p \leq$1e-4).

We also find that the improvement of rank 2 over rank 1 depends on the event predicted: improvement is 26.3% for short distances (100m,200m), 29.3% for middle distances (400m,800m,1500m), 12.8% for the mile to half-marathon, and 3.1% for the Marathon (all significant at p=1e-3 level) (see Figure 5). These results indicate that inter-athlete variability is greater for short and middle distances than for Marathon.

(III) The three components of the model. The findings in (II) imply that the best low-rank model assumes 3 components. To estimate the components (f_i in Equation (1)) we impute all missing entries in the data matrix of the top 25% athletes who have attempted 4 events and compute its singular value decomposition (SVD) [18]. From the SVD, the exact form of components can be directly obtained as the right singular vectors (in a least-squares sense, and up to scaling, see methods appendix and S.II.b). We obtain three components in log-time coordinates, which are displayed in the left hand panel of Figure 2. The first component for log-time prediction is linear (i.e., $f_1(s) \propto \log s$ in Equation (1)) to a high degree of precision ($R^2 = 0.9997$) and corresponds to an *individual* power law, applying distinctly to each athlete. The second and third components are non-linear; the second component decreases over short sprints and increases over the remainder, and the third component resembles a parabola with extremum positioned around the middle distances.

In speed coordinates, the first, individual power law component does not display the "broken power law" behaviour of the world records. Deviations from an exact line can be explained by the second and third component (Figure 2 middle).

The three components together explain the world record data and its "broken power law" far more accurately than a simple linear power law trend—with the rank 3 model fitting the world records almost exactly (Figure 2 right; rank 1 component: $R^2 = 0.99$; world-record data: $R^2 = 0.93$).

(IV) The three athlete-specific coefficients. The three summary coefficients for each athlete ($\lambda_1, \lambda_2, \lambda_3$ in Equation (1)) are obtained from the entries of the left singular vectors (see methods appendix). Since all three coefficients summarize the athlete, we refer to them collectively as the *three-number-summary*. (IV.i) Figure 3 displays scatter plots and Spearman correlations between the coefficients and performance over the full range of distances. The individual exponent correlates with performance on distances greater than 800m. The second coefficient correlates positively with performance over short distances and displays a non-linear association with performance over middle distances. The third coefficient correlates with performance over middle distances. (All correlations significant at $p \leq$1.0e-4; t-distribution approximation to the distribution of Spearman's correlation coefficient.) The associations for all three coefficients are non-linear, with the notable exception of the individual exponent on distances exceeding 800m, hence the application of Spearman correlations. (IV.ii) Figure 4 top displays the three-number-summary for the top 95% athletes in the database. The athletes appear to separate into (at least) four classes, which associate with the athlete's preferred distance. A qualitative transition can be observed over middle distances. Three-number-summaries of world class athletes (not all in the UK athletics database), computed from their personal bests, are listed in Table 1; they and also shown as highlighted points in Figure 4 top right. The elite athletes trace a frontier around the population: all elite athletes are subject to a low individual exponent. A hypothetical athlete holding all the world records is also shown in Figure 4 top right, obtaining an individual exponent which comes close to the world record exponent estimated by Riegel [36] (1.08 for elite athletes, 1.06 for senior athletes). (IV.iii) Figure 4 bottom left shows that a low individual exponent correlates positively with performance in an athlete's preferred event. The individual exponents are higher on average (median=1.12; 5th, 95th percentiles=1.10,1.15) than the world record exponents estimated by Riegel. (IV.iv) Figure 4 bottom right shows that in cross-section, the individual exponent decreases with age until 20 years, and subsequently increases. (All correlations significant at $p \leq$1.0e-4; t-distribution approximation to the
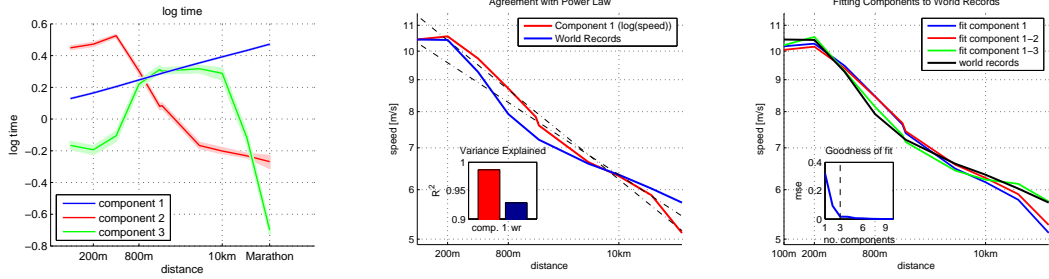
Figure 2: The three components of the low-rank model, and explanation of the world record data. Left: the components displayed (unit norm, log-time vs log-distance). Tubes around the components are one standard deviation, estimated by the bootstrap. The first component is an exact power law (straight line in log-log coordinates); the last two components are non-linear, describing transitions at around 800m and 10km. Middle: Comparison of first component and world record to the exact power law (log-speed vs log-distance). Right: Least-squares fit of rank 1-3 models to the world record data (log-speed vs log-distance).

### Component weights vs. individual performance
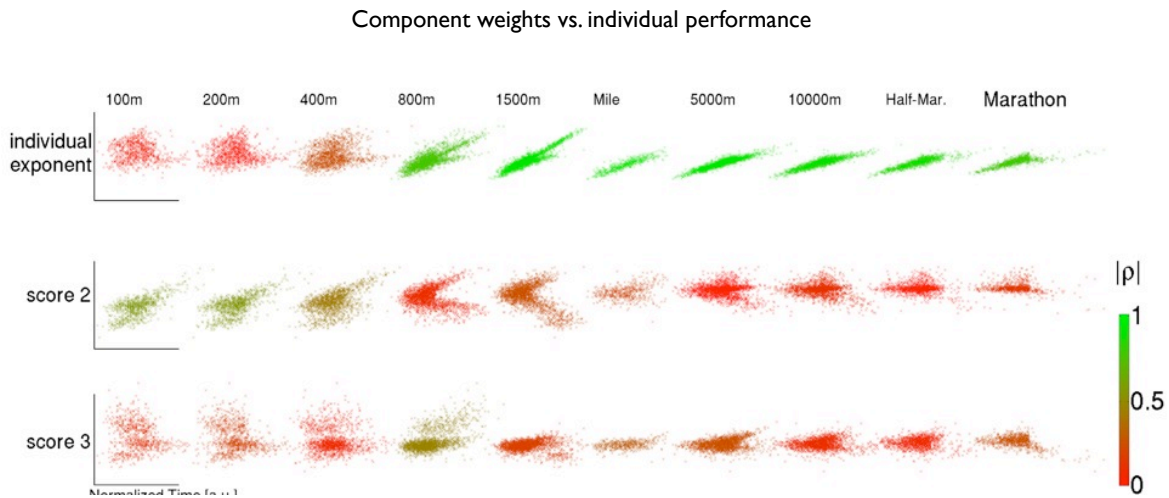
Figure 3: Matrix scatter plot of the three-number-summary vs performance. For each of the scores in the three-number-summary (rows) and each event distance (columns), the plot matrix shows: a scatter plot of performances (time) vs the coefficient score of the top 25% (on the best event) athletes who have attempted at least 4 events. Each scatter plot in the matrix is colored on a continuous color scale according to the absolute value of the scatter sample's Spearman rank correlation (red = 0, green = 1).

distribution of Spearman's correlation coefficient.)

(V) Phase transitions. We observe two transitions in behaviour between short and long distances. The data exhibit a phase transition around 800m: the second component exhibits a kink and the third component makes a zero transition (Figure 2); the association of the first two scores with performance shifts from the second to the first score (Figure 3). The data also exhibits a transition around 5000m. We find that for distances shorter than 5000m, holding the event performance constant and increasing the standard of shorter events leads to a decrease in the predicted standard of longer events and vice versa. On the other hand for distances greater than 5000m this behaviour reverses; holding the event performance constant, and increasing the standard of shorter events leads to an increase in the predicted standard of longer events. See supplementary section (S.IV) for details.

(VI) Universality over subgroups. Qualitatively and quantitatively similar results to the above can be deduced for female athletes, and subgroups stratified by age or training standard; LMC remains an accurate predictor, and the low-rank model has similar form. See supplement (S.II.c).
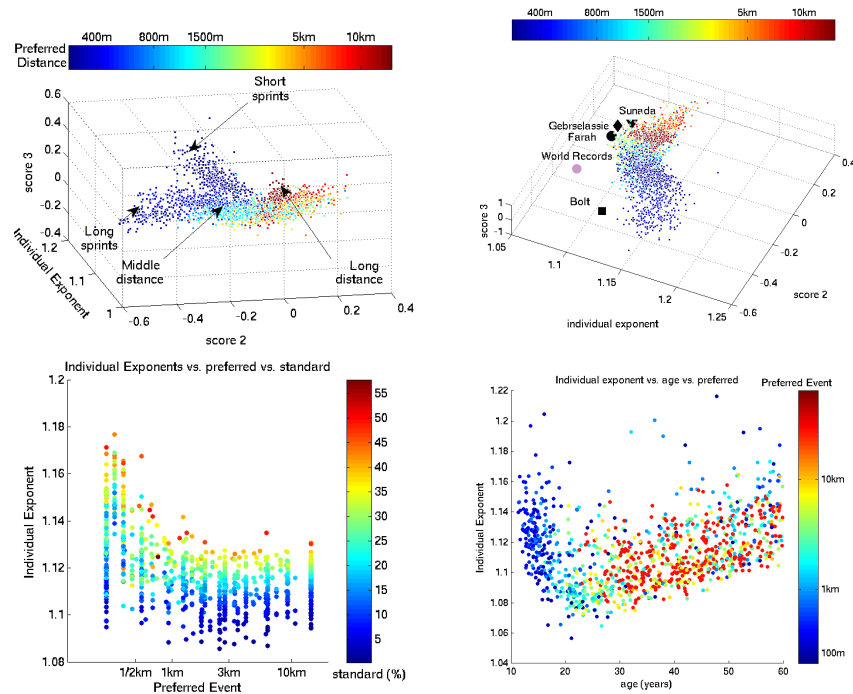
Figure 4: Scatter plots exploring the three number summary. Top left and right: 3D scatter plot of three-number-summaries of athletes in the data set, colored by preferred distance and shown from two angles. A negative value for the second score is a indicates that the athlete is a sprinter, a positive value an endurance runner. In the top right panel, the summaries of the elite athletes Usain Bolt (world record holder, 100m, 200m), Mo Farah (world beater over distances between 1500m and 10km), Haile Gabrselassie (former world record holder from 5km to Marathon) and Takahiro Sunada (100km world record holder) are shown; summaries are estimated from their personal bests. For comparison we also display the hypothetical data of an athlete who holds all world records. Bottom left: preferred distance vs individual exponents, color is percentile on preferred distance. Bottom right: age vs. exponent, colored by preferred distance.

Athlete	Specialization	Individual Exponent (λ_1)	Score 2 (λ_2)	Score 3 (λ_3)
Usain Bolt	Sprints	1.11	-0.367	0.081
Mo Farah	Middle-Long	1.08	0.033	-0.076
Haile Gabrselassie	Long	1.08	0.114	-0.056
Galen Rupp	Long	1.08	0.104	-0.040
Seb Coe	Middle	1.09	-0.085	-0.036
Takahiro Sunada	Ultra-Long	1.09	0.138	-0.010
Paula Radcliffe	Long (Female)	1.10	0.189	0.025

Table 1: Estimated three-number-summary (λ_i) for a selection of elite athletes. The scores λ_1, λ_2, λ_3 are as in Equation (1). Since component 1 is a power law (see the top-left of Figure 2), λ_1 may be interpreted as an individual exponent. See the bottom right panel of Figure 4 for a scatter plot of the athletes in the database.

Discussion and Outlook

We have presented the most accurate existing predictor for running performance to date—local low-rank matrix completion (finding I); its predictive power confirms the validity of a three-component model (finding II) that offers a parsimonious explanation for many known phenomena in the quantitative science of running, including answers to some of the major open questions of the field. More precisely, we establish:

The individual power law. In log-time coordinates, the first component of our physiological model is linear with high accuracy, yielding an individual power law (finding III). This is a novel and rather surprising finding, since, although world-record performances are known to obey a power law [28, 19, 36, 22, 39, 17], there is no reason to suppose *a-priori* that the performance of individuals is governed by a power law. This *parsimony a-posteriori* unifies (A) the parsimony of the power law with the (B) empirical correctness of scoring tables. To what extent this individual power law is exact is to be determined in future studies.

An explanation of the world record data. The broken power law on world records can be seen as a consequence of the individual power law and the non-linearity in the second and third component (finding III) of our low-rank model. The breakage point in the world records can be explained by the differing contributions in the non-linear components of the distinct individuals holding the world records.

Thus both *the power law and the broken power law on world record data can be understood as epiphenomena of the individual power law and its non-linear corrections.*

Universality of our model. The low-rank model remains unchanged when considering different subgroups of athletes, stratified by gender, age, or calendar year; what changes is only the individual three-number-summaries (finding VI). This shows the low-rank model to be universal for running.

The three-number-summary reflects an athlete's training state. Our predictive validation implies that the number of components of our model is three (finding II), which yields three numbers describing the training state of a given athlete (finding IV). The most important summary is *the individual exponent* for the individual power law which describes overall performance, especially on middle and longer distances (IV.iii). The second coefficient describes whether the athlete has greater endurance (positive) or speed (negative), the third describes specialization over middle distances (negative) vs short and long distances (positive). All three numbers together clearly separate the athletes into four clusters, which fall into two clusters of short-distance runners and one cluster of middle-and long-distance runners respectively (IV.i). Our analysis provides strong evidence that the three-number-summary captures physiological and/or social/behavioural characteristics of the athletes, e.g., training state, specialization, and which distance an athlete chooses to attempt. While the data set does not allow us to separate these potential influences or to make statements about cause and effect, we conjecture that combining the three-number-summary with specific experimental paradigms will lead to a clarification; further, we conjecture that a combination of the three-number-summary with additional data, e.g. training logs, high-frequency training measurements or clinical parameters, will lead to a better understanding of (C) existing physiological models.

Some novel physiological insights can be deduced from leveraging our model on the UK athletics database:

- We find that the individual exponent correlates with performances at middle distances above 800m and especially long distances above 5km (finding III). We also find that LMC is most effective for the longer-sprints and middle distances; the improvement of the higher rank over the rank 1 version is lowest over the marathon distance (supplement S.I.c). This indicates that the variability in performances on long distances may to a large extent be explained by a single factor, which may imply that there is **only one way to be a fast marathoner** (in the database). For example, some middle-distance runners use a high maximum velocity to coast whereas other runners use greater endurance to run closer to their maximum speed for the duration of the race; if the type of running (coasting vs endurance) is a physiological correlate to the specialization summary (as hypothesized above), it could imply that the "one way" corresponds to possessing a high level of endurance—as opposed to being able to coast relative to a high maximum speed. In any case, the low-rank model predicts that a marathoner who is not close to world class over 10km is unlikely to be a world class marathoner.

- The **phase transitions** which we observe (finding V) provide additional observational evidence for a transition in the complexity of the physiology underlying performance between long and short distances. This finding is bolstered by the difference we observe between the increase in performance of the rank 2

predictor over the rank 1 predictor for short/middle distances over long distances. Our results may have implications for existing hypotheses and findings in sports science on the differences in physiological determinants of long and short distance running respectively. These include differences in the muscle fibre types contributing to performance (type I vs. type II) [38, 21], whether the race length demands energy primarily from aerobic or anaerobic metabolism [6, 16], which energy systems are mobilized (glycolysis vs. lipolysis) [7, 42] and whether the race terminates before the onset of a $\dot{V}O_2$ slow component [5, 31]. We conjecture that the combination of our methodology with experiments will shed further light on these differences.

- An open question in the physiology of aging is whether sprinting power or endurance capabilities diminish faster with age. Our analysis provides cross-sectional evidence that **training standard decreases with age, and specialization shifts away from endurance**: a larger exponent is correlated with worse performance on endurance type events (finding IV.i), and exponent increases, in cross-section, with age (finding IV.iv). This confirms observations of Rittweger et al. [37] on masters world-record data. There are multiple possible explanations for this, for example longitudinal changes in specialization, or selection bias due to the distances older athletes prefer; our model renders these hypotheses amenable to quantitative validation.

- We find that there are a number of **high-standard athletes who attempt distances different from their inferred best distance**; most notably a cluster of young athletes (< 25 ys) who run short distances (mostly in accordance with legal limitations of participation), and a cluster of older athletes (>40 y) who run long distances, but who we predict would perform better on longer resp. shorter distances. Moreover, the third component of our model implies the existence of **athletes with very strong specialization in their best event**; there are indeed high profile examples of such athletes, such as Zersenay Tadese, who holds the half-marathon world best performance (58:23) but has as yet to produce a marathon performance even close to this in quality (best performance, 2:10:41).

We also anticipate that our framework will prove fruitful in **equipping the practioner with new methods for prediction and quantification**:

- Individual predictions are crucial in **race planning**—especially for predicting a target performance for events such as the Marathon for which months of preparation are needed; the ability to accurately select a realistic target speed will make the difference between an athlete achieving a personal best performance and dropping out of the race from exhaustion.
  N.B.: *We would like to stress that using a prediction as part of marathon preparation without professional support may lead to injury and is entirely at the risk of the user.*

- Predictions and the three-number-summary yield a concise description of the runner's specialization and training state and are thus of immediate use in **training assessment and planning**, for example in determining the potential effect of a training scheme or finding the optimal event(s) for which to train.
  N.B.: *We would like to stress that our study is not able to assign a conclusive meaning to the three-number summary, due to the limitations of the data set; therefore decisions should not be based on a hypothesized interpretation without consideration.*

- The presented framework allows, in principle, for the derivation of novel and more accurate **scoring schemes**, including scoring tables for any type of population.
  N.B.: *We would like to stress that the form of the derived scoring tables may depend on the selection of the data from which they are derived.*

- Predictions for elite athletes allow for a more precise **estimation of quotas and betting risk**. For example, we predict that a fair race between Mo Farah and Usain Bolt is over 492m (374-594m with 95% confidence), Chris Lemaitre and Adam Gemili have the calibre to run 43.5 (±1.3) and 43.2 (±1.3) resp. seconds over 400m. Kenenisa Bekele is capable, in a training state where he can achieve his personal bests over 5km, 10km and the half-marathon, of a 2:00:36 marathon (±3.6 mins).
  N.B.: *We would like to stress that such predictions need to be taken with much caution, as they are only insofar correct as our model extends, from the general population of British runners (who successfully participated in official events), to the very extremes of human performance.*

We further conjecture that the physiological laws we have validated for running will be immediately transferable to any sport where a power law has been observed on the collective level, such as swimming, cycling, and horse racing.

Acknowledgments

Author contributions

DAJB conceived the application LMC to athletic performance prediction, and acquired the data. DAJB and FJK jointly contributed to methodology and the working form of the LMC prediction algorithm. FJK conceived the LMC algorithm in higher ranks; DAJB designed its concrete implementation and adapted the LMC algorithm for the performance prediction problem. DAJB and FJK jointly designed the experiments and analyses. DAJB carried out the experiments and analyses. DAJB and FJK jointly wrote the paper and are jointly responsible for presentation, discussion and interpretation.

Persons under the age of 18. No British regulations concerning determinants are known to the authors.

Methods

The following provides a guideline for reproducing the results. Raw and pre-processed data in MATLAB and CSV formats is available upon request, subject to approval by British Athletics. Complete and documented source code of algorithms and analyses can be obtained from [download link will be provided here after acceptance of the manuscript].

Data Source

The basis for our analyses is the online database `www.thepowerof10.info`, which catalogues British individuals' performances achieved in officially ratified athletics competitions since 1954, including Olympic athletic events (field and non-field events), non-Olympic athletic events, cross country events and road races of all distances.

With permission of British Athletics, we obtained an excerpt of the database by automated querying of the freely accessible parts of `www.thepowerof10.info`, restricted to ten types of running events: 100m, 200m, 400m, 800m, 1500m, the Mile, 5000m (track and road races), 10000m (track and road races), Half-Marathon and Marathon. Other types of running events were available but excluded from the present analyses; the reasons for exclusion were a smaller total of attempts (e.g. 3000m), a different population of athletes (e.g. 3000m is mainly attempted by younger athletes), and varying conditions (steeplechase/ hurdles and cross-country races).

The data set consists of two tables: `athletes.csv`, containing records of individual athletes, with fields: athlete ID, gender, date of birth; and `events.csv`, containing records of individual attempts on running events until August 3, 2013, with fields: athlete ID, event type, date of the attempt, and performance in seconds.

The data set is available upon request, subject to approval by British Athletics.

Data Cleaning

Our excerpt of the database contains (after error and duplication removal) records of 164,746 individuals of both genders, ranging from the amateur to the elite, young to old, and a total of 1,410,789 individual performances for 10 different types of events (see previous section).

Gender is available for all athletes in the database (101,775 male, 62,971 female). The dates of birth of 114,168 athletes are missing (recorded as January 1, 1900 in `athletes.csv` due to particulars of the automated querying); the date of birth of six athletes is set to missing due to an recorded age at recorded attempts of eight years or less.

For the above athletes, a total of 1,417,476 attempts was recorded, out of which 1,410,789 remained in the data set after cleaning: 192,947 over 100m, 194,107 over 200m, 109,430 over 400m, 239,666 over 800m, 176,284 over 1500m, 6,590 at the Mile distance, 96,793 over 5000m (the track and road races), 161,504 over 10000m (on the track and road races), 140,446 for the Half-Marathon and 93,033 for the Marathon. 6,643 duplicate events were removed, and a number of 44 events whose reported performances are better than the official world records of their time, or extremely slow. Dates of the attempt were set to missing for 225 of the attempts that recorded January 1, 1901, and one of the attempts that recorded August 20, 2038.

Data Preprocessing

The events and athletes data sets are collated into $(10 \times 164,746)$-tables/matrices of performances, where the 10 columns correspond to events and the $164,746$ rows to individual athletes. Rows are indexed increasingly by athlete ID, columns by the type of event. Each entry of the table/matrix contains one performance (in seconds) of the athlete by which the row is indexed, at the event by which the column is indexed, or a missing value. If the entry contains a performance, the date of that performance is stored as meta-information.

We consider two different modes of collation, yielding one table/matrix of performances of size $(10 \times 164,746)$ each.

In the first mode, which in Tables 2 ff. is referenced as **"best"**, one proceeds as follows. First, for each individual athlete, one finds the best event of each individual, measured by population percentile. Then, for each type of event which was attempted by that athlete within a year before that best event, the best

performance for that type of event is entered into the table. If a certain event was not attempted in this period, it is recorded as missing.

For the second mode of collation, which in Tables 2 ff. is referenced as **"random"**, one proceeds as follows. First, for each individual athlete, a calendar year is uniformly randomly selected among the calendar years in which that athlete has attempted at least one event. Then, for each type of event which was attempted by that athlete within the selected calendar year, the best performance for that type of event is entered into the table. If a certain event was not attempted in the selected calendar year, it is recorded as missing.

The first collation mode ensures that the data is of high quality: athletes are close to optimal fitness, since their best performance was achieved in this time period. Moreover, since fitness was at a high level, it is plausible that the number of injuries incurred was low; indeed it can be observed that the number of attempts per event is higher in this period, effectively decreasing the influence of noise and the chance that outliers are present after collation.

The second collation mode is used to check whether and, if so how strongly, the results depend on the athletes being close to optimal fitness.

In both cases choosing a narrow time frame ensures that performances are relevant to one another for prediction.

Athlete-Specific Summary Statistics

For each given athlete, several summaries are computed based on the collated matrix.

Performance **percentiles** are computed for each event which an athlete attempts in relation to the other athletes' performances on the same event. These column-wise event-specific percentiles, yield a percentile matrix with the same filling pattern (pattern of missing entries) as the collated matrix.

The **preferred distance** for a given athlete is the geometric mean of the attempted events' distances. That is, if $s_1, \ldots, s_m$ are the distances for the events which the athlete has attempted, then $\tilde{s} = (s_1 \cdot s_2 \cdot \ldots \cdot s_m)^{1/m}$ is the preferred distance.

The **training standard** for a given athlete is the mean of all performance percentiles in the corresponding row.

The **no. events** for a given athlete is the number of events attempted by an athlete in the year of the data considered (**best** or **random**).

Note that the percentiles yield a mostly physiological description; the preferred distance is a behavioural summary since it describes the type of events the athlete attempts. The training standard combines both physiological and behavioural characteristics.

Percentiles, preferred distance, and training standard depend on the collated matrix. At any point when rows of the collated matrix are removed, future references to those statistics refer to and are computed for the matrix where those have been removed; this affects the percentiles and therefore the training standard which is always relative to the athletes in the collated matrix.

Outlier Removal

Outliers are removed from the data in both collated matrices. An outlier score for each athlete/row is obtained as the difference of maximum and minimum of all performance percentile of the athlete. The five percent rows/athletes with the highest outlier score are removed from the matrix.

Prediction: Evaluation and Validation

Prediction accuracy is evaluated on row-sub-samples of the collated matrices, defined by (a) a potential subgroup, e.g., given by age or gender, (b) degrees-of-freedom constraints in the prediction methods that require a certain amount of entries per row, and (c) a certain performance percentiles of athletes.

The row-sub-samples referred to in the main text and in Tables 2 ff. are obtained by (a) retaining all rows/athletes in the subgroup specified by gender, or age in the best event, (b) retaining all rows/athletes with at least **no. events** or more entries non-missing, and discarding all rows/athletes with strictly less than **no. events** entries non-missing, then (c) retaining all athletes in a certain percentile range. The percentiles referred to in (c) are computed as follows: first, for each column, in the data retained after step

(b), percentiles are computed. Then, for each row/athlete, the best of these percentiles is selected as the score over which the overall percentiles are taken.

The accuracy of prediction is measured empirically in terms of out-of-sample root mean squared error (RMSE) and mean absolute error (MAE), with RMSE, MAE, and standard deviations estimated from the empirical sample of residuals obtained in 1000 iterations of leave-one-out validation.

Given the row-sub-sample matrix obtained from (a), (b), (c), prediction and thus leave-one-out validation is done in two ways: (i) predicting the left-out entry from potentially all remaining entries. In this scenario, the prediction method may have access to the performance of the athlete in question which lie in the future of the event to be predicted, though only performances of other events are available; (ii) predicting the left-out entry from all remaining entries of other athletes, but only from those events of the athlete in question that lie in the past of the event to be predicted. In this task, temporal causality is preserved on the level of the single athlete for whom prediction is done; though information about other athletes' results that lie in the future of the event to be predicted may be used.

The third option (iii) where predictions are made only from past events has not been studied due to the size of the data set which makes collation of the data set for every single prediction per method and group a computationally extensive task, and due to the potential group-wise sampling bias which would be introduced, skewing the measures of prediction-quality—the population of athletes on the older attempts is different in many respects from the more recent attempts. We further argue that in the absence of such technical issues, evaluation as in (ii) would be equivalent to (iii); since the performances of two randomly picked athletes, no matter how they are related temporally, can in our opinion be modelled as statistically independent; positing the contrary would be equivalent to postulating that any given athlete's performance is very likely to be directly influenced by a large number of other athlete's performance history, which is an assumption that appears to us to be scientifically implausible. Given the above, due to equivalence of (ii) and (iii), and the issues occurring in (iii) exclusively, we can conclude that (ii) is preferrable over (iii) from a scientific and statistical viewpoint.

Prediction: Target Outcomes

The principal target outcome for the prediction is "performance", which we present to the prediction methods in three distinct parameterisations. This corresponds to passing not the raw performance matrices obtained in the section "Data Pre-processing" to the prediction methods, but re-parameterized variants where the non-missing entries undergo a univariate variable transform. The three parameterizations of performance considered in our experiments are the following:

(a) **normalized**: performance as the time in which the given athlete (row) completes the event in question (column), divided by the average time in which the event in question (column) is completed in the sub-sample;

(b) **log-time**: performance as the natural logarithm of time in seconds in which the given athlete (row) completes the event in question (column);

(c) **speed**: performance as the average speed in meters per second, with which the given athlete (row) completes the event in question (column).

The words in italics indicate which parameterisation is referred to in Table 2. The error measures, RMSE and MAE, are evaluated in the same parameterisation in which prediction is performed. We do not evaluate performance directly in un-normalized time units, as in this representation performances between 100m and the Marathon span 4 orders of magnitude (base-10), which would skew the measures of goodness heavily towards accuracy over the Marathon.

Unless stated otherwise, predictions are made in the same parameterisation on which the models are learnt.

Prediction: Models and Algorithms

In the experiments, a variety of prediction methods are used to perform prediction from the performance data, given as described in "Prediction: Target Outcomes", evaluated by the measures as described in the section "Prediction: Evaluation and Validation".

In the code available for download, each method is encapsulated as a routine which predicts a missing entry when given the (training entries in the) performance matrix. The methods can be roughly divided in four classes: (1) naive baselines, (2) representatives of the state-of-the-art in prediction of running performance, (3) representatives of the state-of-the-art in matrix completion, and (4) our proposed method and its variants.

The naive baselines are:

(1.a) **mean**: predicting the the mean over all performances for the same event. (1.b) k-**NN**: k-nearest neighbours prediction. The parameter k is obtained as the minimizer of out-of-sample RMSE on five groups of 50 randomly chosen validation data points from the training set, from among $k = 1, k = 5$, and $k = 20$.

The representatives of the state-of-the-art in predicting running performance are:

(2.a) **Riegel**: The Riegel power law formula with exponent 1.06. (2.b) **power law**: A power law predictor, as per the Riegel formula, but with the exponent estimated from the data. The exponent is the same for all athletes and estimated as the minimizer of the residual sum of squares. (2.c) **ind.power law**: A power law predictor, as per the Riegel formula, but with the exponent estimated from the data. The exponent may be different for each athlete and is estimated as the minimizer of the residual sum of squares. (2.d) **Purdy**: Prediction by calculation of equivalent performances using the Purdy points scheme [33]. Purdy points are calculated by using the measurements given by the Portugese scoring tables which estimate the maximum velocity for a given distance in a straight line, and adjust for the cost of traversing curves and the time required to reach race velocity. The performance with the same number of points as the predicting event is imputed.

The representatives of the state-of-the-art in matrix completion are:

(3.a) **EM**: Expectation maximization algorithm assuming a multivariate Gaussian model for the rows of the performance matrix in log-time parameterisation. Missing entries are initialized by the mean of each column. The updates are terminated when the percent increase in log-likelihood is less than 0.1%. For a review of the EM-algorithm see [3]. (3.b) **Nuclear Norm**: Matrix completion via nuclear norm minimization [10, 40].

The variants of our proposed method are as follows:

(4.a-d) **LMC rank** r: local matrix completion for the low-rank model, with rank $r = 1, 2, 3, 4$. (4.a) is LMC rank 1, (4.b) is LMC rank 2, and so on.

Our algorithm follows the local/entry-wise matrix completion paradigm in [26]. It extends the rank 1 local matrix completion method described in [25] to arbitrary ranks.

Our implementation uses: determinants of size $(r + 1 \times r + 1)$ as the only circuits; the weighted variance minimization principle in [25]; the linear approximation for the circuit variance outlined in the appendix of [4]; modelling circuits as independent for the co-variance approximation.

We further restrict to circuits supported on the event to be predicted and the r log-distance closest events.

For the convenience of the reader, we describe the exact way in which the local matrix completion principle is instantiated, in the section "Prediction: Local Matrix Completion" below

In the supplementary experiments we also investigate two aggregate predictors to study the potential benefit of using other lengths for prediction:

(5.a) **bagged power law**: bagging the power law predictor with estimated coefficient (2.b) by a weighted average of predictions obtained from different events. The weighting procedure is described below. (5.b) **bagged LMC rank 2**: estimation by LMC rank 2 where determinants can be supported at any three events, not only on the closest ones (as in line 1 of Algorithm 1 below). The final, bagged predictor is obtained as a weighted average of LMC rank 2 running on different triples of events. The weighting procedure is described below.

The averaging weights for (5.a) and (5.b) are both obtained from the Gaussian radial basis function kernel $\exp\left(\gamma\Delta\Delta^\top\right)$, where $\Delta = \log(\mathbf{s}_p) - \log(s_{p'})$ and $\mathbf{s}_p$ is the vector of predicting distances and $s_{p'}$ is the predicted distance. The kernel width γ is a parameter of the bagging. As γ approaches 0, aggregation approaches averaging and thus the "standard" bagging predictor. As γ approaches $-\infty$, the aggregate prediction approaches the non-bagged variants (2.b) and (4.b).

Prediction: Local Matrix Completion

The LMC algorithm we use is an instance of Algorithm 5 in [26], where, as detailed in the last section, the circuits are all determinants, and the averaging function is the weighted mean which minimizes variance, in

first order approximation, following the strategy outlined in [25] and [4].

The LMC rank r algorithm is described below in pseudo-code. For readability, we use bracket notation $M[i, j]$ (as in R or MATLAB) instead of the usual subscript notation M_{ij} for sub-setting matrices. The notation $M[:, (i_1, i_2, \ldots, i_r)]$ corresponds to the sub-matrix of m with columns $i_1, \ldots, i_r$. The notation $M[k, :]$ stands for the whole k-th row. Also note that the row and column removals in Algorithm 1 are only temporary for the purpose of computation, within the boundaries of the algorithm, and do not affect the original collated matrix.

Algorithm 1 – Local Matrix Completion in Rank r.
Input: An athlete a, an event s^*, the collated data matrix of performances M.
Output: An estimate/denoising for the entry $M[a, s^*]$

1: Determine distinct events $s_1, \ldots, s_r \neq s^*$ which are log-closest to s^*, i.e., minimize $\sum_{i=1}^{r} (\log s_i - \log s^*)^2$
2: Restrict M to those events, i.e., $M \leftarrow M[:, (s^*, s_1, \ldots, s_r)]$
3: Let v be the vector containing the indices of rows in M with no missing entry.
4: $M \leftarrow M[(v, a), :]$, i.e., remove all rows with missing entries from M, except a.
5: **for** $i = 1$ to 400 **do**
6:     Uniformly randomly sample distinct athletes $a_1, \ldots, a_r \neq a$ among the rows of M.
7:     Solve the circuit equation $\det M[(a, a_1, \ldots, a_r), (s^*, s_1, \ldots, s_r)] = 0$ for s^*, obtaining a number m_i.
8:     Let $A_0, A_1 \leftarrow M[(a, a_1, \ldots, a_r), (s^*, s_1, \ldots, s_r)]$.
9:     Assign $A_0[a, s^*] \leftarrow 0$, and $A_1[a, s^*] \leftarrow 1$.
10:     Compute $\sigma_i \leftarrow \frac{1}{|\det A_0 + \det A_1|} + \frac{|\det A_0|}{(\det A_0 - \det A_1)^2}$
11:     Assign the weight $w_i \leftarrow \sigma_i^{-2}$
12: **end for**
13: Compute $m^* \leftarrow \left(\sum_{i=1}^{400} w_i m_i \right) \cdot \left(\sum_{i=1}^{400} w_i \right)^{-1}$
14: Return m^* as the estimated performance.

The bagged variant of LMC rank r repeatedly runs LMC rank r with choices of events different from the log-closest, weighting the results obtained from different choices of $s_1, \ldots, s_r$. The weights are obtained from 5-fold cross-validation on the training sample.

Obtaining the Low-Rank Components and Coefficients

We obtain three low-rank components $f_1, \ldots, f_3$ and corresponding coefficients $\lambda_1, \ldots, \lambda_3$ for each athlete by considering the data in log-time coordinates. Each component f_i is a vector of length 10, with entries corresponding to events. Each coefficient is a scalar, potentially different per athlete.

To obtain the components and coefficients, we consider the data matrix for the specific target outcome, sub-sampled to contain the athletes who have attempted four or more events and the top 25% percentiles, as described in "Prediction: Evaluation and Validation". In this data matrix, all missing values are imputed using the rank 3 local matrix completion algorithm, as described in (4.c) of "Prediction: Models and Algorithms", to obtain a complete data matrix M. For this matrix, the singular value decomposition $M = USV^\top$ is computed, see [18].

We take the components f_2, f_3 to be the the 2-th and 3-rd right singular vectors, which are the 2-nd and 3-rd column of V. The component f_1 is a re-scaled version of the 1-st column v of V, such that $f_1(s) \approx \log s$, where the natural logarithm is taken. More precisely, $f_1 := \beta^{-1} v$, where the re-scaling factor β is obtained as the ordinary least-squares regression coefficient of the linear explanatory model $v(s) = \beta \log s + c$, where s ranges over the ten event distances, which is $\beta = 0.0572$. A more detailed study of v and the regression coefficient can be found in supplementary experiment (S.II.b).

The three-number-summary referenced in the main corpus of the manuscript is obtained as follows: for the k-th athlete we obtain from the left singular vector the entries U_{kj}. The second and third score of the three-number-summary are obtained as $\lambda_2 = U_{k2}$ and $\lambda_3 = U_{k3}$. The individual exponent is $\lambda_1 = \beta \cdot U_{j1}$.

The singular value decomposition has the property that the f_i and λ_j are guaranteed to be least-squares estimators for the components and the coefficients in a projection sense.

Computation of standard error and significance

Standard errors for the singular vectors (components of the model of Equation 1) are computed via independent bootstrap sub-sampling on the rows of the data set (athletes).

Standard errors for prediction accuracies are obtained by bootstrapping of the predicted performances (1000 per experiment). A method is considered to perform significantly better than another when error regions at the 95% confidence level (= mean over repetitions $\pm$ 1.96 standard errors) do not intersect.

Predictions and three-number-summary for elite athletes

Performance predictions and three-number-summaries for the selected elite athletes in Table 1 and Figure 4 are obtained from their personal best times. The relative standard error of the predicted performances is estimated to be the same as the relative RMSE of predicting time, as reported in Table 2.

Calculating a fair race

Here we describe the procedure for calculating a fair racing distance with error bars between two athletes: athlete 1 and athlete 2. We first calculate predictions for all events. Provided that athlete 1 is quicker on some events and athlete 2 is quicker on others, then calculating a fair race is feasible. If athlete 1 is quicker on shorter events then athlete 2 is typically quicker on all longer events beyond a certain distance. In that case, we can find the shortest race s_i whereby athlete 2 is predicted to be quicker; then a fair race lies between s_i and s_{i-1}. The performance curves in log-time vs. log-distance of both athletes will be locally approximately linear. We thus interpolate the performance curves between $\log(s_i)$ and $\log(s_{i-1})$—the crossing point gives the position of a fair race in log-coordinates. We obtain confidence intervals by repeating this procedure after sampling data points around the estimated performances with standard deviation equal to the RMSE (see Table 2) on the top 25% of athletes in log-time.

Supplementary Analyses

This appendix contains a series of additional experiments supplementing those in the main corpus. It contains the following findings:

(S.I) Validation of the LMC prediction framework.
(S.I.a) Evaluation in terms of MAE. The results in terms of MAE are qualitatively similar to those in RMSE; smaller MAEs indicate the presence of outliers.
(S.I.b) Evaluation in terms of time prediction. The results are qualitatively similar to measuring prediction accuracy in RMSE and MAE of log-time. LMC rank 2 has an average error of approximately 2% when predicting the top 25% of male athletes.
(S.I.c) Prediction for individual events. LMC outperforms the other predictors on each type of event. The benefit of higher rank is greatest for middle distances.
(S.I.d) Stability w.r.t. the unit measuring performance. LMC performs equally well in predicting (performance in time units) when performances are presented in log-time or time normalized by event average. Speed is worse when the rank 2 predictor is used.
(S.I.e) Stability w.r.t. the events used in prediction. LMC performs equally well when predicting from the closest-distance events and when using a bagged version which uses all observed events for prediction.
(S.I.f) Stability w.r.t. the event predicted. LMC performs well both when the predicted event is close to those observed and when the predicted event is further from those observed, in terms of event distance.
(S.I.g) Temporal independence of performances. There are no differences between predictions made only from past events and predictions made from all available events (in the training set).
(S.I.h) Run-time comparisons. LMC is by orders of magnitude the fastest among the matrix completion methods.

(S.II) Validation of the low-rank model.
(S.II.a) Synthetic validation. In a synthetic low-rank model of athletic performance that is a proxy to the real data, the singular components of the model can be correctly recovered by the exact same procedure as on the real data. The generative assumption of low rank is therefore appropriate.
(S.II.b) The individual power law component, and the distance/time unit. The first singular component can be explained by a linear model in log-distance (R-square 0.9997) with slope $\beta = 0.0572 \pm 0.0003$ and intercept $c = -0.136 \pm 0.003$.
(S.II.c) Universality in sub-groups. Quality of prediction, the low-rank model, its rank, and the singular components remain mostly unchanged when considering subgroups male/female, older/younger, elite/amateur.

(S.III) Exploration of the low-rank model.
(S.III.a) Further exploration of the three-number-summary. The three number summary also correlates with specialization and training standard.
(S.III.b) Preferred distance vs optimal distance. Most but not all athletes prefer to attend the event at which they are predicted to perform best. A notable number of younger athletes prefer distances shorter than optimal, and some older athletes prefer distances longer than optimal.

(S.IV) Pivoting and phase transitions. The pivoting phenomenon in Figure 1, right panel, is found in the data for any three close-by distances up to the Mile, with anti-correlation between the shorter and the longer distance. Above 5000m, a change in the shorter of the three distances positively correlates with a change in the longer distance.

(S.I.a) Evaluation in terms of MAE. Table 3 reports on the goodness of prediction methods in terms of MAE. Compared with the RMSE (Table 2, the MAE tend to be smaller than the RMSE, indicating the presence of outliers. The relative prediction-accuracy of methods when compared to each other is qualitatively the same.

(S.I.b) Evaluation in terms of time prediction. Tables 5 and 6 report on the prediction accuracy of the methods tested in terms the relative RMSE and MAE of predicting time. Relative measures are chosen to avoid bias towards the longer events. The results are qualitatively and quantitatively very similar to the log-time results in Tables 2 and 3; this can be explained that mathematically the RMSE and MAE of a logarithm approximate the relative RMSE and MAE well for small values.

(S.I.c) Individual Events. Prediction accuracy of LMC rank 1 and rank 2 on the ten different events is displayed in Figure 5. The reported prediction accuracy is out-of-sample RMSE of predicting log-time, on the top 25 percentiles of Male athletes who have attempted 3 or more events, of events in their best year of performance. The reported RMSE for a given event is the mean over 1000 random prediction samples, standard errors are estimated by the bootstrap.
The relative improvement of rank 2 over rank 1 tends to be greater for shorter distances below the Mile. This is in accordance with observation (IV.i) which indicates that the individual exponent is the best descriptor for longer events, above the Mile.

(S.I.d) Stability w.r.t. the measure of performance. In the main experiment, the LMC model is learnt on the same measure of performance (log-time, speed, normalized) which is predicted. We investigate whether the measure of performance on which the model is learnt influences the prediction by learning the LMC model on either measure and comparing all predictions using the log-time measure. Table 9 displays prediction accuracy when the model is learnt in any one of the measures of performance. Here we check the effect of calibration in one coordinates system and testing in another. The reported goodness is out-of-sample RMSE of predicting log-time, on the top 25 percentiles of Male athletes who have attempted 3 or more events, of events in their best year of performance. The reported RMSE for a given event is the mean over 1000 random prediction samples, standard errors are estimated by the bootstrap.
We find that there is no significant difference in prediction goodness when learning the model in log-time coordinates or normalized time coordinates. Learning the model in speed coordinates leads to a significantly better prediction than log-time or normalized time when LMC rank 1 is applied, but to a worse prediction with LMC rank 2. As overall prediction with LMC rank 2 is better, log-time or normalized time are the preferable units of performance.

(S.I.e) Stability w.r.t. the event predicted.
We consider here the effect of the ratio between the predicted event and the closest predictor. For data of the best 25% of Males in the year of best performance (**best**), we compute the log-ratio of the closest predicting distance and the predicted distance for Purdy Points, the power law formula and LMC rank 2. See Figure 6, where this log ratio is plotted by error. The results show that LMC is far more robust to error for predicting distances far from the predicted distance.

(S.I.f) Stability w.r.t. the events used in prediction. We compare whether we can improve prediction by using all events an athlete has attempted, by using one of the aggregate predictors (5.a) bagged power law or (5.b) bagged LMC rank 2. The kernel width γ for the aggregate predictors is chosen from $-0.001, -0.01, -0.1, -1, -10$ as the minimizer of out-of-sample RMSE on five groups of 50 randomly chosen validation data points from the training set. The validation setting is the same as in the main prediction experiment.
Results are displayed in Table 10. We find that prediction accuracy of (2.b) power law and (5.a) bagged power law is not significantly different, nor is (4.b) LMC rank 2 significantly different from (5.b) bagged LMC rank 2 (both $p > 0.05$; Wilcoxon signed-rank on the absolute residuals). Even though the kernel width selected is in the majority of cases $\sigma = -1$ and not $\sigma = -10$, the incorporation of all events does not lead to an improvement in prediction accuracy in our aggregation scheme. We find there is no significant difference ($p > 0.05$; Wilcoxon signed-rank on the absolute errors) between the bagged and vanilla LMC for the top 95% of runners. This demonstrates that the relevance of closer events for prediction may be learn from the data. The same holds for the bagged version of the power law formula.

(S.I.g) Temporal independence of performances. We check here whether the results are affected by using only temporally prior attempts in predicting an athlete's performance, see section "Prediction:
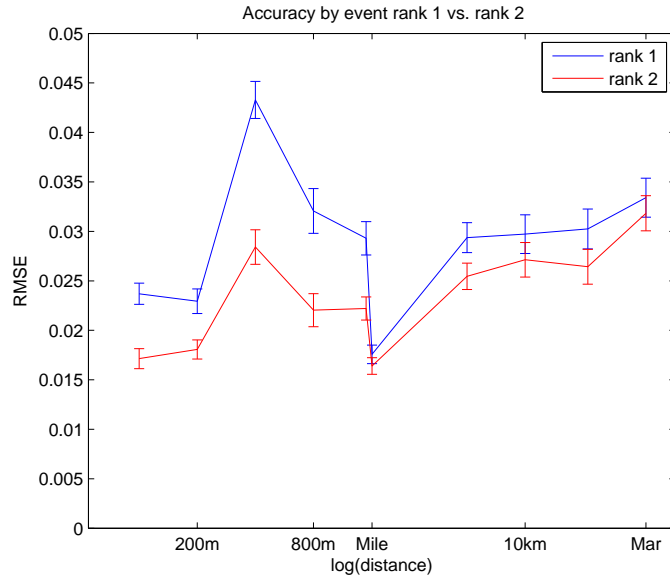
Figure 5: The figure displays the results of prediction by event for the top 25% of male athletes who attended ≥ 3 events in their year of best performance. For each event the prediction accuracy of LMC rank 1 (blue) is compared to prediction accuracy in rank 2 (red). RMSE is displayed on the y-axis against distance on the x-axis; the error bars extend two standard deviations of the bootstrapped RMSE either side of the RMSE.
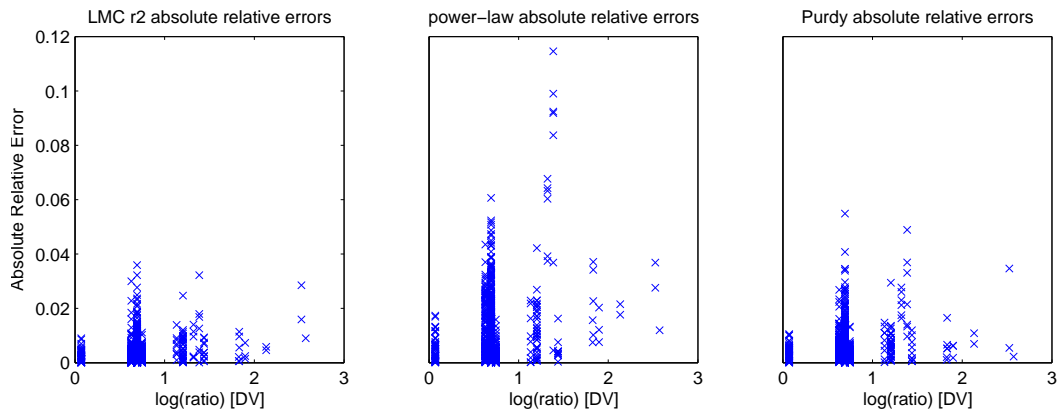
Figure 6: The figure displays the absolute log ratio in distance predicted and predicting distance vs. absolute relative error per athlete. In each case the log ratio in distance is displayed on the x-axis and the absolute errors of single data points of the y-axis. We see that LMC rank 2 is particularly robust for large ratios in comparison to the power law and Purdy Points. Data is taken from the top 25% of male athletes with **no. events**≥ 3 in the **best** year.
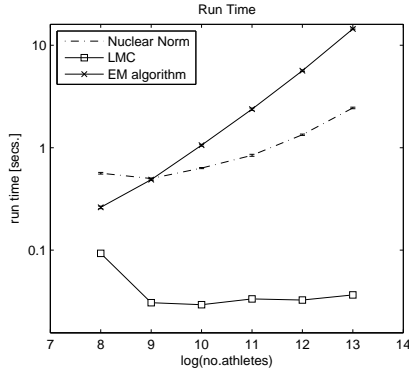
Figure 7: The figure displays mean run-times for the 3 matrix completion algorithms tested in the paper: Nuclear Norm, EM and LMC (rank 3). Run-times (y-axis) are recorded for completing a single entry in a matrix of size indicated by the x-axis. The averages are over 100 repetitions, standard errors are estimated by the bootstrap.

Evaluation and Validation" in "Methods". To this end, we compute out-of-sample RMSEs when predictions are made only from those events.

Table 4 reports out-of-sample RMSE of predicting log-time, on the top 25 percentiles of Male athletes who have attempted 3 or more events, of events in their best year of performance. The reported RMSE for a given event is the mean over 1000 random prediction samples, standard errors are estimated by the bootstrap.

The results are qualitatively similar to those of Table 2 where all events are used in prediction.

(S.I.h) Run-time comparisons. We compare the run-time cost of a single prediction for the three matrix completion methods LMC, nuclear norm minimiziation, and EM. The other (non-matrix completion) methods are fast or depend only negligibly on the matrix size. We measure run time of LMC rank 3 for completion of a single entry for matrices of $2^8, 2^9, \ldots, 2^{13}$ athletes, generated as described in (S.II.a). This is repeated 100 times. For a fair comparison, the nuclear norm minimization algorithm is run with a hyper-parameter already pre-selected by cross validation. The results are displayed in Figure 7; LMC is faster by orders of magnitude than nuclear norm and EM and is very robust to the size of the matrix. The reason computation speeds up over the smallest matrix sizes is that 4×4 minors, which are required for rank 3 estimation are not available, thus the algorithm must attempt all ranks lower than 3 to find sufficiently many minors.

(S.II.a) Synthetic validation. To validate the assumption of a low-rank generative model, we investigate prediction accuracy and recovery of singular vectors in a synthetic model of athletic performance.

Synthetic data for a given number of athletes is generated as follows:

For each athlete, a three-number summary $(\lambda_1, \lambda_2, \lambda_3)$ is generated independently from a Gaussian distribution with the same mean and variance as the three-number-summaries measured on the real data and with uncorrelated entries.

Matrices of performances are generated from the model

$$\log(t) = \lambda_1 f_1(s) + \lambda_2 f_2(s) + \lambda_3 f_3(s) + \eta(s) \tag{2}$$

where f_1, f_2, f_3 are the three components estimated from the real data and $\eta(s)$ is a stationary zero-mean Gaussian white noise process with adjustable variance. We take the components estimated in log-time coordinates from the top 25% of male athletes who have attempted at least 4 events as the three components of the model. The distances s are the same ten event distances as on the real data. In each experiment the standard deviation of $\eta(s)$

Accuracy of prediction: We synthetically generate a matrix of 1000 athletes according to the model of Equation (2), taking as distances the same distances measured on the real data. Missing entries are randomized according to two schemes: (a) 6 (out of 10) uniformly random missing entries per row/athlete. (b) per row/athlete, four in terms of distance-consecutive entries are non-missing, uniformly at random.
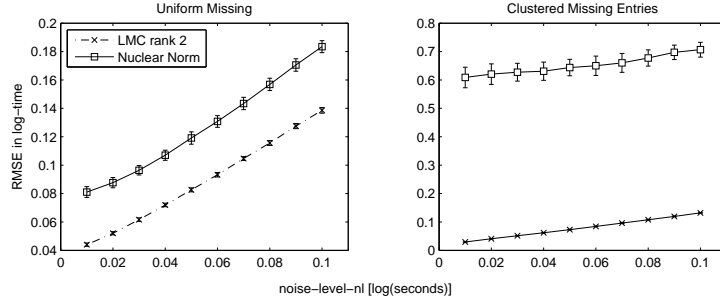
Figure 8: LMC and Nuclear Norm prediction accuracy on the synthetic low-rank data. x-axis denotes the noise level (standard deviation of additive noise in log-time coordinates); y-axis is out-of-sample RMSE predicting log-time. Left: prediction performance when (a) the missing entries in each ros are uniform. Right: prediction performance when (b) the observed entries are consecutive. Error bars are one standard deviation, estimated by the bootstrap.
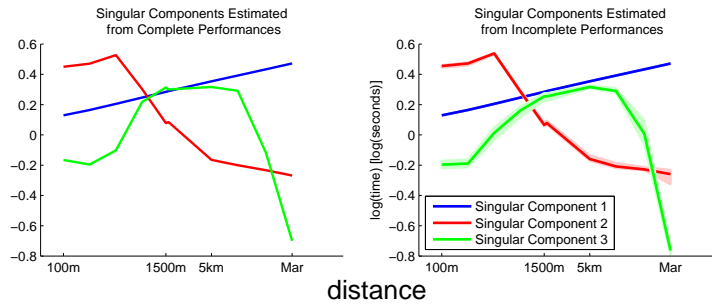
Figure 9: Accuracy of singular component estimation with missing data on synthetic model of performance. x-axis is distance, y-axis is components in log-time. Left: singular components of data generated according to Equation 2 with all data present. Right: singular components of data generated according to Equation 2 with missing entries estimated with LMC rank 3; the observation pattern and number of athletes is identical to the real data. The tubes denote one standard deviation estimated by the bootstrap.

We then apply LMC rank 2 and nuclear norm minimization for prediction. This setup is repeated 100 times for ten different standard deviations of η between 0.01 and 0.1. The results are displayed in Figure 8.

LMC performance outperforms nuclear norm; LMC performance is also robust to the pattern of missingness, while nuclear norm minimization is negatively affected by clustering in the rows. RMSE of LMC approaches zero with small noise variance, while RMSE of nuclear norm minimization does not.

Comparing the performances with Table 2, an assumption of a noise variance of $\mathrm{Std}(\eta) = 0.01$ seems plausible. The performance of nuclear norm on the real data is explained by a mix of the sampling schemes (a) and (b).

Recovery of model components. We synthetically generate a matrix which has a size and pattern of observed entries identical to the matrix of top 25% of male athletes who have attempted at least 4 events in their best year. We set $\mathrm{Std}(\eta) = 0.01$, which was shown to be plausible in the previous section.

We then complete all missing entries of the matrix using LMC rank 3. After this initial step we estimate singular components using SVD, exactly as on the real data. Confidence intervals are estimated by a bootstrap on the rows with 100 iterations.

The results are displayed in Figure 9.

One observes that the first two singular components are recovered almost exactly, while the third is a slightly deformed. This is due to the smaller singular value of the third component.

(S.II.b) The individual power law component, and the distance/time unit. We examine linearity of the first singular vector v, as listed in Table 11 and as described in methods section "Obtaining the Low-Rank Components and Coefficients". In an ordinary least squares regression model explaining v by $\log s$ and an intercept, we find that $v \approx \beta \log s + c$ with an R-squared of 0.9997 (Table 12), where the scaling factor is $\beta = 0.0572 \pm 0.0003$ and the intercept is $c = -0.136 \pm 0.003$. The intercept corresponds to a choice of unit, the scaling factor to a choice of basis for the logarithm. Thus re-scaling v with β^{-1}, that is, setting
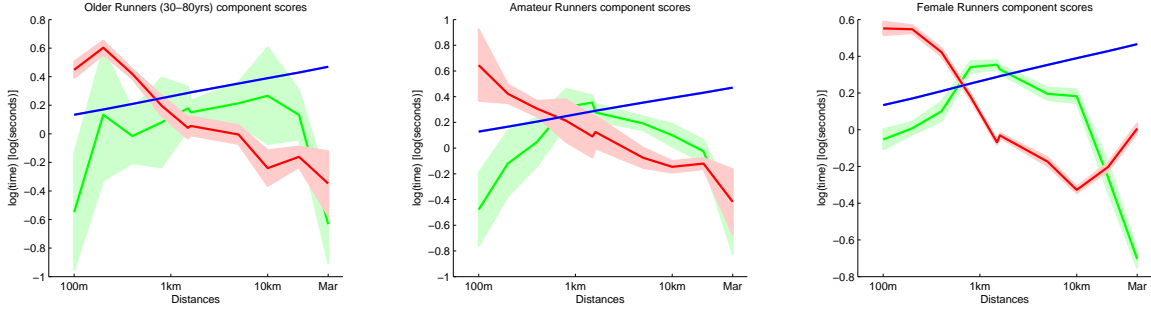
Figure 10: The three components of the low-rank model in subgroups. Left: for older runners. Middle: for amateur runners = best event below 25th percentile. Right: for female runners. Tubes around the components are one standard deviation, estimated by the bootstrap. The components are the analogous components for the subgroups described as computed in the left-hand panel of Figure 2.

$f_1 := \beta^{-1} v$ in the low-rank model, and re-scaling the first individual coefficient with β, corresponds to the choice of the natural basis.

The residuals of the the linear model appear to be plausibly explained by the second and third singular component (Table 12), though the small number of fitting nodes which is 10 does not allow a for an assessment that is more than qualitative.

(S.II.c) Universality in sub-groups. We repeat the methodology for component estimation described above and obtain the three components in the following sub-groups: female athletes, older athletes (> 30 years), and amateur athletes (25-95 percentile range of training standard). Male athletes were considered in the main corpus. For female and older athletes, we restrict to the top 95% percentiles of the respective groups for estimation.

Figure 10 displays the estimated components of the low-rank model. The individual power law is found to be unchanged in all groups considered. The second and third component vary between the groups but resemble the components for the male athletes. The empirical variance of the second and third component is higher, which may be explained by a slightly reduced consistency in performance, or a reduction in sample size. Whether there is a genuine difference in form or whether the variation is explained by different three-number-summaries in the subgroups cannot be answered from the dataset considered.

Table 8 displays the prediction results in the three subgroups. Prediction accuracy is similar but slightly worse when compared to the male athletes. Again this may be explained by reduced consistency in the subgroups' performances.

(S.III.a) Further exploration of the three-number-summary. Scatter plots of preferred distance and training standard against the athletes' three-number-summaries are displayed in Figure 11. The training standard correlates predominantly with the individual exponent (score 1); score 1 vs. standard—$r = -0.89$ ($p \leq 0.001$); score 2 vs. standard—$r = 0.22$ ($p \leq 0.001$); score 3 vs. standard—$r = 0.031$ ($p = 0.07$); all correlations are Spearman correlations with significance computed using a t-distribution approximation to the correlation coefficient under the null. On the other hand preferred distance is associated with all three numbers in the summary, especially the second; score 1 vs. log(specialization)—$r = 0.29$ ($p \leq 0.001$); score 2 vs. log(specialization)—$r = -0.58$ ($p \leq 0.001$); score 3 vs. log(specialization)—$r = -0.14$ ($p = \leq 0.001$); The association between the third score and specialization is non-linear with an optimal value around the middle distances. We stress that low correlation does not imply low predictive power; the whole summary should be considered as a whole, and the LMC predictor is non-linear. Also, we observe that correlations increase when considering only performances over certain distances, see Figure 2.

(S.III.b) Preferred event vs best event. For the top 95% male athletes who have attempted 3 or more events, we use LMC rank 2 to compute which percentile they would achieve in each event. We then determine the distance of the event at which they would achieve the best percentile, to which we will refer as the "optimal distance". Figure 12 shows for each athlete the difference between their preferred and optimal
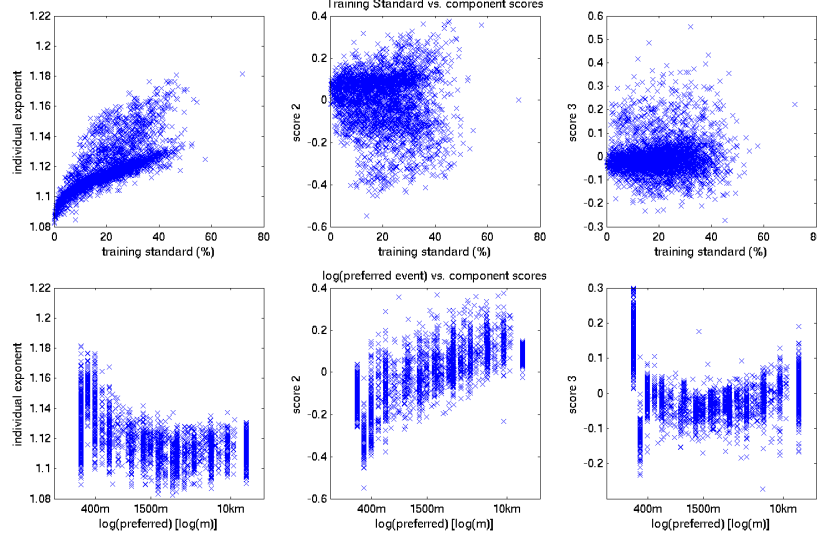
Figure 11: Scatter plots of training standard vs. three-number-summary (top) and preferred distance vs. three-number-summary. In each case the individual exponents, 2nd and 3rd scores (λ_2, λ_3) are displayed on the y-axis and the log-preferred distance and training standard on the x-axis.

distance.

It can be observed that the large majority of athletes prefer to attempt events in the vicinity of their optimal event. There is a group of young athletes who attempt events which are shorter than the predicted optimal distance, and a group of old athletes attempting events which are longer than optimal. One may hypothesize that both groups could be explained by social phenomena: young athletes usually start to train on shorter distances, regardless of their potential over long distances. Older athletes may be biased to attempting endurance type events.

(S.IV) Pivoting and phase transitions. We look more closely at the pivoting phenomenon illustrated in Figure 1 top right, and the phase transition discussed in observation (V). We consider the top 25% of male athletes who have attempted at least 3 events, in their best year.

We compute 10 performances of equivalent standard by using LMC rank 1 in log-time coordinates, by setting a benchmark performance over the marathon and sequentially predicting each lower distance (marathon predicts HM, HM predicts 10km etc.). This yields equivalent benchmark performances $t_1, \ldots, t_{10}$.

We then consider triples of consecutive distances s_{i-1}, s_i, s_{i+1} (excluding the Mile since close in distance to the 1500m) and study the pivoting behaviour on the data set, by performing the analogous prediction displayed Figure 1.

More specifically, for each triple, we predict the performance on the distance s_{i+1} using LMC rank 2, from the performances over the distances s_{i-1} and s_i. The prediction is performed in two ways, once with and once without perturbation of the benchmark performance at s_{i-1}, which we then compare. Intuitively, this corresponds to comparing the red to the green curve in Figure 1. In mathematical terms:

1. We obtain a prediction $\widehat{t}_{i+1}$ for the distance s_{i+1} from the benchmark performances t_i, t_{i-1} and consider this as the unperturbed prediction, and

2. We obtain a prediction $\widehat{t}_{i+1} + \delta(\epsilon)$ for the distance s_{i+1} from the benchmark performance t_i on s_i and the perturbed performance $(1 + \epsilon)t_{i-1}$ on the distance s_{i-1}, considering this as the perturbed prediction.

We record these estimates for $\epsilon = -0.1, 0.09, \ldots, 0, 0.01, \ldots, 0.1$ and calculate the relative change of the perturbed prediction with respect to the unperturbed, which is $\delta_i(\epsilon)/\widehat{t}_i$. The results are displayed in Figure 13.
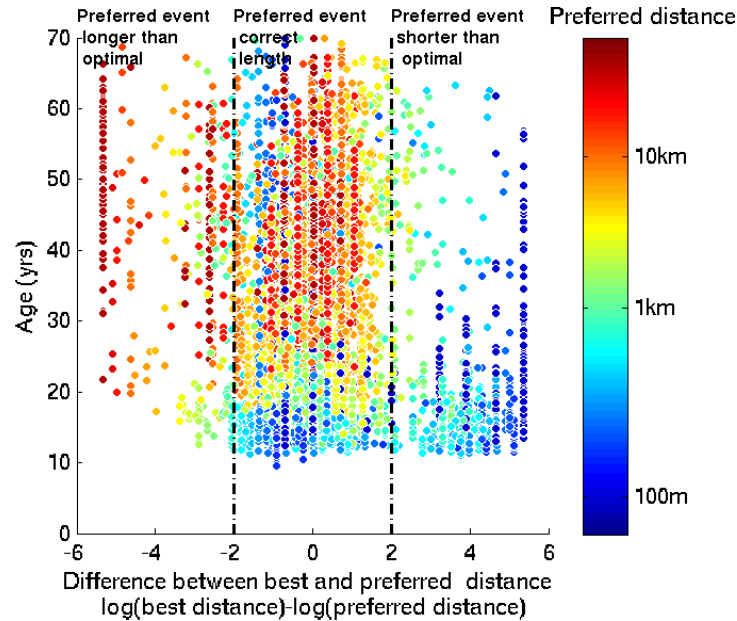
Figure 12: Difference of preferred distance and optimal distance, versus age of the athlete, colored by specialization distance. Most athletes prefer the distance they are predicted to be best at. There is a mismatch of best and preferred for a group of younger athletes who have greater potential over longer distances, and for a group of older athletes who's potential is maximized over shorter distances than attempted.

We find that for pivot distances s_i shorter than 5km, a slower performance on the shorter distance s_{i-2} leads to a faster performance over the longer distance s_i, insofar as this is predicted by the rank 2 predictor. On the other hand we find that for pivot distances greater than or equal to 5km, a faster performance over the shorter distance also implies a faster performance over the longer distance.

References

[1] L Véronique Billat. Use of blood lactate measurements for prediction of exercise performance and for control of training. *Sports medicine*, 22(3):157–175, 1996.

[2] L Véronique Billat and J Pierre Koralsztein. Significance of the velocity at VO2max and time to exhaustion at this velocity. *Sports Medicine*, 22(2):90–108, 1996.

[3] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[4] Duncan AJ Blythe, Louis Theran, and Franz Kiraly. Algebraic-combinatorial methods for low-rank matrix completion with application to athletic performance prediction. *arXiv preprint arXiv:1406.2864*, 2014.

[5] F Borrani, R Candau, GY Millet, S Perrey, J Fuchslocher, and JD Rouillon. Is the VO2 slow component dependent on progressive recruitment of fast-twitch fibers in trained runners? *Journal of Applied Physiology*, 90(6):2212–2220, 2001.

[6] Laurent Bosquet, Luc Léger, and Patrick Legros. Methods to determine aerobic endurance. *Sports Medicine*, 32(11):675–700, 2002.

[7] George A. Brooks and Jacques Mercier. Balance of carbohydrate and lipid utilization during exercise: the "crossover" concept. *Journal of Applied Physiology*, 76(6):2253–2261, 1994.

[8] Michael W. Browne. Cross-validation methods. *Journal of Mathematical Psychology*, 44(1):108–132, 2000.

[9] Matthew W. Bundle, Reed W. Hoyt, and Peter G. Weyand. High-speed running performance: a new approach to assessment and prediction. *Journal of Applied Physiology*, 95(5):1955–1962, 2003.

[10] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
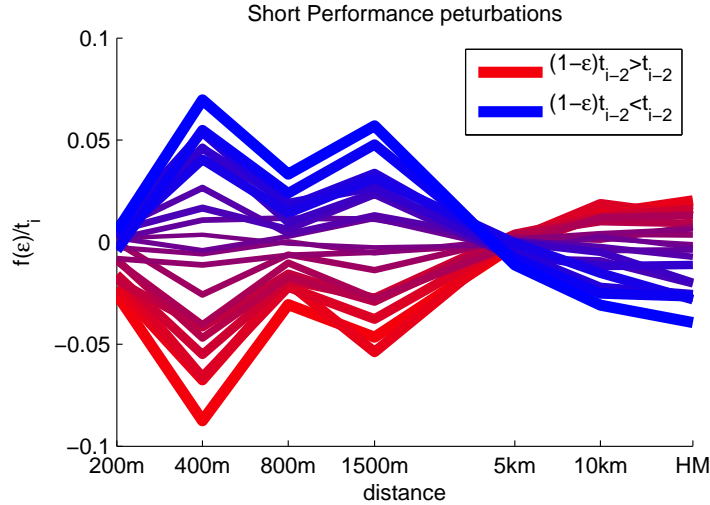
Figure 13: Pivot phenomenon in the low-rank model. The figure quantifies the strength and sign of pivoting as in Figure 1, top right, at different middle distances s_i (x-axis). The computations are based on equivalent log-time performances t_{i-1}, t_i, t_{i+1} at consecutive triples s_{i-1}, s_i, s_{i+1} of distances. The y-coordinate indicates the signed relative change of the LMC rank 2 prediction of t_{i+1} from t_{i-1} and t_i changes, when t_i is fixed and t_{i-1} undergoes a relative change of $1\%, 2\%, \ldots, 10\%$ (red curves, line thickness is proportional to change), or $-1\%, -2\%, \ldots, -10\%$ (blue curves, line thickness is proportional to change). For example, the largest peak corresponds to a middle distance of $s_i = 400m$. When predicting 800m from 400m and 200m, the predicted log-time t_{i+1} (= 800m performance) decreases by 8% when t_{i-1} (= 200m performance) is increased by 10% while t_i (= 400m performance) is kept constant.

[11] Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5):2053–2080, 2010.

[12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[13] Pietro Enrico Di Prampero. Factors limiting maximal performance in humans. *European journal of applied physiology*, 90(3-4):420–429, 2003.

[14] Bradley Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331, 1983.

[15] Bradley Efron and Robert Tibshirani. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

[16] Oliver Faude, Wilfried Kindermann, and Tim Meyer. Lactate threshold concepts. *Sports Medicine*, 39(6):469–490, 2009.

[17] Juan M. García-Manso, Juan M. Martín-González, Nancy Dávila, and Enrique Arriaza. Middle and long distance athletics races viewed from the perspective of complexity. *Journal of theoretical biology*, 233(2):191–198, 2005.

[18] Gene H. Golub and Christian Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5):403–420, 1970.

[19] Franklin M. Henry. Prediction of world records in running sixty yards to twenty-six miles. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 26(2):147–158, 1955.

[20] Archibald V Hill, CNH Long, and H Lupton. Muscular exercise, lactic acid, and the supply and utilisation of oxygen. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, pages 84–138, 1924.

[21] Hans Hoppeler, Hans Howald, Kevin Conley, Stan L. Lindstedt, Helgard Claassen, Peter Vock, and Ewald R. Weibel. Endurance training in humans: aerobic capacity and structure of skeletal muscle. *Journal of Applied Physiology*, 59(2):320–327, 1985.

[22] Leon Katz and J Sylvan Katz. Fractal (power law) analysis of athletic performance. *Research in Sports Medicine: An International Journal*, 5(2):95–105, 1994.

[23] Joseph B Keller. A theory of competitive running. *Physics today*, page 43, 1973.

[24] Arthur Edwin Kennelly. An approximate law of fatigue in the speeds of racing animals. In *Proceedings of the American Academy of Arts and Sciences*, pages 275–331. JSTOR, 1906.

[25] Franz J. Király and Louis Theran. Obtaining error-minimizing estimates and universal entry-wise error bounds for low-rank matrix completion. *NIPS 2013*, 2013. arXiv 1302.5337.

[26] Franz J. Király, Louis Theran, and Ryota Tomioka. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research*, 2015. accepted, to appear. arXiv 1211.4116.

[27] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, volume 14, pages 1137–1145, 1995.

[28] Milton H. Lietzke. An analytical study of world and olympic racing records. *Science*, 119(3089):333–336, 1954.

[29] Timothy D. Noakes, Kathryn H. Myburgh, and Robert Schall. Peak treadmill running velocity during the VO2max test predicts running performance. *Journal of sports sciences*, 8(1):35–45, 1990.

[30] François Péronnet and Guy Thibault. Mathematical analysis of running performance and world running records. *Journal of Applied Physiology*, 67(1):453–465, 1989.

[31] David C. Poole, Thomas J. Barstow, Glenn A. Gaesser, Wayne T. Willis, and Brian J Whipp. VO2 slow component: physiological and functional significance. *Medicine and science in sports and exercise*, 26(11):1354–1358, 1994.

[32] J Gerry Purdy. Computer generated track and field scoring tables: I. historical development. *Medicine and science in sports*, 6(4):287, 1974.

[33] J Gerry Purdy. Computer generated track and field scoring tables: II. theoretical foundation and development of a model. *Medicine and science in sports*, 7(2):111–115, 1974.

[34] J Gerry Purdy. Computer generated track and field scoring tables: III. model evaluation and analysis. *Medicine and science in sports*, 9(4):212–218, 1976.

[35] Peter S. Riegel. Time predicting. *Runner's World Magazine*, Aug, 1977.

[36] Peter S. Riegel. Athletic records and human endurance. *American Scientist*, 69(3):285–290, 1980.

[37] Jörn Rittweger, Pietro Enrico di Prampero, Nicola Maffulli, and Marco V. Narici. Sprint and endurance power and ageing: an analysis of master athletic world records. *Proceedings of the Royal Society B: Biological Sciences*, 276(1657):683–689, 2009.

[38] B. Saltin, J. Henricksson, E. Hygaard, and P. Andersen. Fibre types and metabolic potentials of skeletal muscles in sedentary man and endurance runners. *Annals of the New York Academy of Sciences*, pages 3–29, 1977.

[39] Sandra Savaglio and Vincenzo Carbone. Human performance: Scaling in athletic world records. *Nature*, 404(6775):244–244, 2000.

[40] Ryota Tomioka, Kohei Hayashi, and Hisashi Kashima. On the extension of trace norm to tensors. In *NIPS Workshop on Tensors, Kernels, and Machine Learning*, page 7, 2010.

[41] Gerrit Jan Ingen van Schenau, Jos J de Koning, and Gert de Groot. Optimisation of sprinting performance in running, cycling and speed skating. *Sports Medicine*, 17(4):259–275, 1994.

[42] Michelle C. Venables, Juul Achten, and Asker E. Jeukendrup. Determinants of fat oxidation during exercise in healthy men and women: a cross-sectional study. *Journal of applied physiology*, 98(1):160–167, 2005.

[43] Karlman Wasserman, Brian J. Whipp, SN Koyl, and WL Beaver. Anaerobic threshold and respiratory gas exchange during exercise. *Journal of applied physiology*, 35(2):236–243, 1973.

				Generic Baselines		State of art Performance Predictors				State of art Matrix Completion		Proposed Method: LMC	
evaluation	percentiles	no.events	data type	r.mean	k-NN	individual power law	riegel	power law	purdy	nuclear norm	EM	LMC rank 1	LMC rank 2
log time	0-95	3	best	0.1308 ±0.0032	0.0618 ±0.0027	0.1033 ±0.0042	0.0982 ±0.0046	0.0973 ±0.0046	0.0610 ±0.0031	0.3909 ±0.0457	0.0566 ±0.0028	0.0586 ±0.0026	0.0515 ±0.0027
normalized	0-95	3	best	0.1364 ±0.0044	0.0716 ±0.0046	0.1067 ±0.0048	0.1059 ±0.0066	0.1050 ±0.0065	0.0684 ±0.0043	0.1900 ±0.0120	0.0634 ±0.0045	0.0643 ±0.0038	0.0576 ±0.0039
speed	0-95	3	best	0.6655 ±0.0147	0.3057 ±0.0146	0.6096 ±0.0245	0.5467 ±0.0251	0.5415 ±0.0243	0.3077 ±0.0176	26.6210 ±11.4828	0.2922 ±0.0165	0.3123 ±0.0149	0.2530 ±0.0129
log time	0-95	3	random	0.1380 ±0.0032	0.0544 ±0.0025	0.0931 ±0.0035	0.0931 ±0.0038	0.0919 ±0.0038	0.0591 ±0.0028	0.4416 ±0.0428	0.0561 ±0.0031	0.0567 ±0.0027	0.0471 ±0.0024
normalized	0-95	3	random	0.1450 ±0.0043	0.0623 ±0.0037	0.0951 ±0.0039	0.1011 ±0.0048	0.0998 ±0.0046	0.0682 ±0.0038	0.2046 ±0.0117	0.0634 ±0.0039	0.0640 ±0.0037	0.0538 ±0.0035
speed	0-95	3	random	0.6935 ±0.0143	0.2585 ±0.0117	0.5917 ±0.0312	0.5052 ±0.0176	0.4979 ±0.0167	0.2835 ±0.0137	24.7206 ±10.9157	0.2801 ±0.0196	0.2863 ±0.0120	0.2261 ±0.0105
log time	0-95	4	best	0.1268 ±0.0032	0.0735 ±0.0030	0.0777 ±0.0024	0.0819 ±0.0032	0.0822 ±0.0032	0.0581 ±0.0023	0.1779 ±0.0199	0.0529 ±0.0024	0.0536 ±0.0021	0.0467 ±0.0022
log time	0-25	3	best	0.0557 ±0.0015	0.0416 ±0.0014	0.0806 ±0.0031	0.0683 ±0.0026	0.0720 ±0.0026	0.0411 ±0.0012	0.3008 ±0.0275	0.0383 ±0.0013	0.0411 ±0.0014	0.0306 ±0.0011

Table 2: Out-of-sample RMSE for prediction methods on different data setups. Predicted performance is of the 25 top percentiles of male athletes, in their best year. Standard errors are bootstrap estimates over 1000 repetitions. Compared method classes are (1) generic baselines, (2) state-of-the-art in performance prediction, (3) state-of-the-art in matrix completion, (4) local matrix completion (columns). Methods are (1.a) r.mean: predicting the mean of all athletes (1.b) k-NN: predicting the nearest neighbor. (2.a) riegel: Riegel's formula (2.b) power law: power law with free exponent and coefficient. Exponent is the same for all athletes. (2.c) ind.power law: power law with free exponent and coefficient. (2.d) purdy: Purdy points scheme (3.a) EM: expectation maximization (3.b) nuclear norm: nuclear norm minimization (4.a) LMC with rank 1 (4.b) LMC with rank 2. Data setup is specified by (i) evaluation: what is predicted. log-time = natural logarithm of time in seconds, normalized = time relative to mean performance, speed = average speed in meters per seconds, (ii) percentiles: selected percentile range of athletes, (iii) no.events tried = sub-set of athletes who have attempted at least that number of different events, (iv) data type: collation mode of performance matrix; best = 1 year around best performance, random = random 2 year period. LMC rank 2 significantly outperforms all competitors in either setting.

				Generic Baselines		State of art Performance Predictors				State of art Matrix Completion		Proposed Method: LMC	
evaluation	percentiles	no.events	data type	r.mean	k-NN	individual power law	riegel	power law	purdy	nuclear norm	EM	LMC rank 1	LMC rank 2
log time	0-95	3	best	0.1054 ±0.0025	0.0421 ±0.0014	0.0696 ±0.0024	0.0661 ±0.0023	0.0654 ±0.0023	0.0423 ±0.0014	0.1282 ±0.0115	0.0387 ±0.0013	0.0402 ±0.0014	0.0336 ±0.0012
normalized	0-95	3	best	0.1062 ±0.0027	0.0441 ±0.0018	0.0700 ±0.0026	0.0681 ±0.0026	0.0674 ±0.0025	0.0441 ±0.0017	0.0907 ±0.0051	0.0400 ±0.0016	0.0413 ±0.0016	0.0347 ±0.0014
speed	0-95	3	best	0.5463 ±0.0122	0.2118 ±0.0067	0.3989 ±0.0145	0.3640 ±0.0130	0.3600 ±0.0126	0.2197 ±0.0070	2.2846 ±0.8163	0.2023 ±0.0065	0.2130 ±0.0072	0.1716 ±0.0058
log time	0-95	3	random	0.1119 ±0.0026	0.0373 ±0.0012	0.0655 ±0.0021	0.0656 ±0.0021	0.0646 ±0.0021	0.0411 ±0.0014	0.1501 ±0.0131	0.0376 ±0.0014	0.0385 ±0.0013	0.0320 ±0.0011
normalized	0-95	3	random	0.1136 ±0.0029	0.0397 ±0.0016	0.0660 ±0.0021	0.0686 ±0.0025	0.0676 ±0.0023	0.0438 ±0.0016	0.1013 ±0.0055	0.0395 ±0.0016	0.0405 ±0.0015	0.0338 ±0.0013
speed	0-95	3	random	0.5727 ±0.0126	0.1825 ±0.0060	0.3795 ±0.0145	0.3552 ±0.0114	0.3497 ±0.0112	0.2066 ±0.0061	2.5636 ±0.7841	0.1913 ±0.0065	0.1995 ±0.0065	0.1607 ±0.0052
log time	0-95	4	best	0.1014 ±0.0024	0.0514 ±0.0016	0.0557 ±0.0017	0.0551 ±0.0020	0.0553 ±0.0020	0.0406 ±0.0013	0.0716 ±0.0051	0.0350 ±0.0013	0.0366 ±0.0013	0.0310 ±0.0010
log time	0-25	3	best	0.0424 ±0.0012	0.0294 ±0.0009	0.0559 ±0.0019	0.0479 ±0.0015	0.0507 ±0.0016	0.0310 ±0.0008	0.0970 ±0.0092	0.0282 ±0.0008	0.0300 ±0.0009	0.0221 ±0.0007

Table 3: Out-of-sample MAE for prediction methods on different data setups. Predicted performance is of the 25 top percentiles of male athletes, in their best year. Standard errors are bootstrap estimates over 1000 repetitions. Compared method classes are (1) generic baselines, (2) state-of-the-art in performance prediction, (3) state-of-the-art in matrix completion, (4) local matrix completion (columns). Methods are (1.a) r.mean: predicting the mean of all athletes (1.b) k-NN: predicting the nearest neighbor. (2.a) riegel: Riegel's formula (2.b) power law: power law with free exponent and coefficient. Exponent is the same for all athletes. (2.c) ind.power law: power law with free exponent and coefficient. (2.d) purdy: Purdy points scheme (3.a) EM: expectation maximization (3.b) nuclear norm: nuclear norm minimization (4.a) LMC with rank 1 (4.b) LMC with rank 2. Data setup is specified by (i) evaluation: what is predicted. log-time = natural logarithm of time in seconds, normalized = time relative to mean performance, speed = average speed in meters per seconds, (ii) percentiles: selected percentile range of athletes, (iii) no.events tried = sub-set of athletes who have attempted at least that number of different events, (iv) data type: collation mode of performance matrix; best = 1 year around best performance, random = random 2 year period. LMC rank 2 significantly outperforms all competitors in either setting.

				Generic Baselines		State of art Performance Predictors				State of art Matrix Completion		Proposed Method: LMC	
evaluation	percentiles	no.events	data type	r.mean	k-NN	individual power law	riegel	power law	purdy	nuclear norm	EM	LMC rank 1	LMC rank 2
log time	0-95	3	best	0.1398	0.0637	0.1065	0.0100	0.0991	0.0639	0.3624	0.0574	0.0622	0.0545
				±0.0066	±0.0057	±0.0054	±0.0063	±0.0063	±0.0066	±0.0849	±0.0060	±0.0054	±0.0050
normalized	0-95	3	best	0.1483	0.0792	0.1103	0.1051	0.1042	0.0724	0.1769	0.0658	0.0694	0.0620
				±0.0097	±0.0104	±0.0068	±0.0066	±0.0068	±0.0097	±0.0222	±0.0096	±0.0078	±0.0086
speed	0-95	3	best	0.7153	0.3308	0.6553	0.5827	0.5772	0.3383	19.2009	0.3067	0.3410	0.2918
				±0.0349	±0.0348	±0.0356	±0.0368	±0.0385	±0.0440	±9.9799	±0.0393	±0.0310	±0.0321
log time	0-95	3	random	0.1380	0.0544	0.0931	0.0931	0.0919	0.0591	0.4416	0.0561	0.0567	0.0471
				±0.0032	±0.0027	±0.0035	±0.0039	±0.0038	±0.0027	±0.0435	±0.0031	±0.0027	±0.0023
normalized	0-95	3	random	0.1450	0.0623	0.0951	0.1011	0.0998	0.0682	0.2046	0.0634	0.0640	0.0538
				±0.0044	±0.0037	±0.0038	±0.0049	±0.0049	±0.0039	±0.0124	±0.0041	±0.0038	±0.0033
speed	0-95	3	random	0.6935	0.2585	0.5917	0.5052	0.4979	0.2835	24.7206	0.2801	0.2863	0.2261
				±0.0147	±0.0121	±0.0329	±0.0171	±0.0167	±0.0134	±10.7164	±0.0199	±0.0121	±0.0112
log time	0-95	4	best	0.1368	0.0763	0.0823	0.0859	0.0862	0.0620	0.2371	0.0608	0.0599	0.0531
				±0.0075	±0.0060	±0.0042	±0.0060	±0.0059	±0.0038	±0.0423	±0.0064	±0.0041	±0.0040
log time	0-25	3	best	0.0539	0.0425	0.0810	0.0675	0.0710	0.0412	0.2479	0.0358	0.0417	0.0318
				±0.0027	±0.0030	±0.0056	±0.0050	±0.0051	±0.0026	±0.0600	±0.0022	±0.0030	±0.0022

Table 4: Prediction only from events which are earlier in time than the performance to be predicted. The table shows out-of-sample RMSE for performance prediction methods on different data setups. Predicted performance is of the 25 top percentiles of male athletes, in their best year. Standard errors are bootstrap estimates over 1000 repetitions. Legend is as in Table 2.

				Generic Baselines		State of art Performance Predictors				State of art Matrix Completion		Proposed Method: LMC	
evaluation	percentiles	no.events	data type	r.mean	k-NN	individual power law	riegel	power law	purdy	nuclear norm	EM	LMC rank 1	LMC rank 2
time	0-95	3	best	0.1295	0.0627	0.0959	0.0973	0.0964	0.0596	0.1785	0.0560	0.0569	0.0499
				±0.0027	±0.0027	±0.0035	±0.0064	±0.0065	±0.0025	±0.0105	±0.0028	±0.0023	±0.0024
time	0-95	3	random	0.1357	0.0535	0.0874	0.0907	0.0895	0.0585	0.1961	0.0544	0.0550	0.0461
				±0.0029	±0.0022	±0.0028	±0.0031	±0.0031	±0.0026	±0.0116	±0.0025	±0.0022	±0.0020
time	0-95	4	best	0.1232	0.0745	0.0750	0.0782	0.0785	0.0566	0.1167	0.0525	0.0522	0.0455
				±0.0025	±0.0031	±0.0021	±0.0027	±0.0027	±0.0021	±0.0084	±0.0029	±0.0019	±0.0019
time	0-25	3	best	0.0559	0.0422	0.0760	0.0668	0.0704	0.0406	0.1579	0.0377	0.0402	0.0302
				±0.0015	±0.0016	±0.0025	±0.0022	±0.0023	±0.0012	±0.0113	±0.0012	±0.0014	±0.0001

Table 5: Exactly the same table as Table 2 but relative root mean squared errors reported in terms of time. Models are learnt on the performances in log-time.

				Generic Baselines		State of art Performance Predictors				State of art Matrix Completion		Proposed Method: LMC	
evaluation	percentiles	no.events	data type	r.mean	k-NN	individual power law	riegel	power law	purdy	nuclear norm	EM	LMC rank 1	LMC rank 2
time	0-95	3	best	0.1057	0.0424	0.0669	0.0654	0.0647	0.0420	0.0876	0.0384	0.0397	0.0333
				±0.0023	±0.0015	±0.0022	±0.0023	±0.0024	±0.0014	±0.0048	±0.0013	±0.0013	±0.0012
time	0-95	3	random	0.1116	0.0372	0.0635	0.0651	0.0642	0.0410	0.0980	0.0373	0.0381	0.0318
				±0.0024	±0.0012	±0.0018	±0.0019	±0.0020	±0.0013	±0.0055	±0.0013	±0.0013	±0.0011
time	0-95	4	best	0.1006	0.0519	0.0547	0.0540	0.0543	0.0401	0.0605	0.0348	0.0362	0.0307
				±0.0023	±0.0016	±0.0016	±0.0018	±0.0018	±0.0013	±0.0032	±0.0013	±0.0012	±0.0011
time	0-25	3	best	0.0425	0.0296	0.0542	0.0476	0.0504	0.0308	0.0688	0.0280	0.0297	0.0220
				±0.0011	±0.0001	±0.0017	±0.0015	±0.0016	±0.0008	±0.0046	±0.0008	±0.0008	±0.0007

Table 6: Exactly the same table as Table 2 but relative mean absolute errors reported in terms of time. Models are learnt on the performances in log-time.

no events.	r1	r2	r3	r4
3	0.0411	0.0306	—	—
	±0.0014	±0.0011		
4	0.0446	0.0328	0.0309	—
	±0.0016	±0.0013	±0.0012	
5	0.0518	0.0408	0.0400	0.0408
	±0.0032	±0.0033	±0.0034	±0.0036

Table 7: Determination of the true rank of the model. Table displays out-of-sample RMSE for predicting performance with LMC rank 1-4 (columns) Predicted performance is of the 25 top percentiles of male athletes, in their best year, who have attempted at least the number of events indicated by the row. The model is learnt on performances in log-time coordinates. Standard errors are bootstrap estimates over 1000 repetitions. The entries where **no. events** $\geq$ rank are empty, as LMC rank r needs $r + 1$ attempted events for leave-one-out-validation. Prediction with LMC rank 3 is always better or equally good compared to using a different rank, in terms of out-of-sample prediction accuracy.

subgroup	RMSE
Amateur	0.0305
	±0.0002
Female	0.0305
	±0.0003
Old	0.0326
	±0.0003

Table 8: Prediction in three different subgroups: amateur athletes, female athletes, older athletes. Table displays out-of-sample RMSE for predicting performance with LMC rank 2.

rank	log time	speed	normalized
1	0.0410	0.0376	0.0399
	±0.0014	±0.0011	±0.0013
2	0.0304	0.0315	0.0305
	±0.0011	±0.0011	±0.0001

Table 9: Effect of performance measure in which the LMC model is learnt. The model is learnt on three different measures of performance: log-time, time normalized by event mean, speed (columns). The table shows out-of-sample RMSE for predicting log-time performance with LMC rank 1,2. Standard errors are bootstrap estimates over 1000 repetitions. Performance is of the 25 top percentiles of male athletes, in their best year of performance.

percentiles	no.event	bagged LMC r2	bagged power-law	LMC r2	power-law
0-25	3	0.0310	0.0654	0.0308	0.0666
		±0.0011	±0.0025	±0.0011	±0.0025
0-95	3	0.0529	0.0898	0.0512	0.0948
		±0.0031	±0.0040	±0.0028	±0.0039
0-95	4	0.0480	0.0762	0.0467	0.0825
		±0.0034	±0.0029	±0.0021	±0.0030

Table 10: Comparison of prediction using all distances, to prediction using only closest distances. Table displays out-of-sample RMSE of predicting log-time, for (5.a) the bagged power law and (5.b) the bagged LMC rank 2 predictor, compared with the un-bagged variants, (2.b) and (4.b). Predicted performance is of the 25 top percentiles of male athletes, in their best year. Standard errors are bootstrap estimates over 1000 repetitions. The results of the bagging predictors are very similar to the unbagged one.

s	100m	200m	400m	800m	1.5km	Mile	5km	10km	HM	M
f_1	2.254	2.875	3.574	4.305	4.964	5.049	6.179	6.844	7.555	8.243
f_2	0.4473	0.4721	0.5265	0.3045	0.0798	0.0806	-0.1597	-0.1983	-0.2279	-0.2785
f_3	-0.1750	-0.2004	-0.1145	0.2224	0.3263	0.3092	0.3157	0.2717	-0.1153	-0.6912
v	0.1291	0.1647	0.2047	0.2466	0.2843	0.2892	0.3539	0.3920	0.4327	0.4721

Table 11: The three components of the low-rank model 1. An entry in the rows $i = 1, 2, 3$ is $f_i(s)$, where s is the column header; HM is the half-Marathon, M is the Marathon. The components are obtained as described in methods, "obtaining the low-rank components and coefficients". v is the raw singular vector described there from which f_1 is obtained by rescaling. v, f_2, f_3 are displayed in Figure 2 top left with standard error tubes per entry. The entries for v have, on average, an estimated standard error of 0.005, the entries for f_2 have, on average, an estimated standard error of 0.02, and the entries for f_3 have, on average, an estimated standard error of 0.04.

	variables	β	β_2	β_3	c
model 1	$\log s$	0.0572 ± 0.0003			-0.136 ± 0.003
model 2	$\log s, f_2$	0.0547 ± 0.0007	-0.017 ± 0.004		-0.115 ± 0.006
model 3	$\log s, f_2, f_3$	0.0554 ± 0.0007	-0.013 ± 0.004	0.002 ± 0.001	-0.120 ± 0.006

	t_1	$p(X > |t_1|)$	t_2	$p(X > |t_2|)$	t_3	$p(X > |t_3|)$	t_c	$p(X > |t_c|)$
model 1	168	1.7e-15					-51	2.3e-11
model 2	81	1.1e-12	-3.9	5.9e-3			-21	1.5e-7
model 3	80	2.5e-10	-3.0	2.5e-2	1.8	0.13	-21	7.1e-7

	F	$P(X > F)$	RSE	R-squared
model 1	2.8e+4	1.7e-15	0.0020	0.9997
model 2	3.9e+4	6.6e-15	0.0012	0.9999
model 3	3.4e+4	4.4e-13	0.0011	0.9999

Table 12: Explaining the first singular component, v. The following explanatory linear models are fitted: v explained by $\beta \log s + c$ (model 1); v explained by $\beta \log s + \beta_2 f_2 + c$ (model 2); v explained by $\beta \log s + \beta_2 f_2 + \beta_3 f_3 + c$. The β, β_2, β_3 are the estimated coefficients, $\pm$ one standard error. t_1, t_2, t_3 are the t-statistics of β, β_2, β_3; t_c is the t-statistic of c. The F-statistic of the respective model is F, RSE is the residual standard error.