The time distribution of biological phenomena – illustrated with the London marathon

Miguel Franco Corresp. [1]

[1] School of Biological and Marine Sciences, University of Plymouth

Corresponding Author: Miguel Franco
Email address: M.Franco@plymouth.ac.uk

Background. The time distribution of biological phenomena (phenology) is a subject of wide interest, but a general statistical distribution to describe and quantify its essential properties is lacking. Existing distributions are limiting, if not entirely inappropriate, because their parameters do not in general correlate with biologically relevant attributes of the organism and the conditions under which they find themselves. **Methods**. A distribution function that allows quantification of three essential properties of a biological dynamic process occurring over a continuous timescale was derived from first principles. The distribution turned out to have three parameters with clear meanings and units: (i) a scaled rate of completion (dimensionless), (ii) a measure of temporal concentration of the process (units: time^{-1}), and (iii) an overall measure of temporal delay (units: time). Its performance as an accurate description of the process was tested with completion data for the London Marathon employing non-linear regression. **Results**. The parameters of the distribution correlated with biological attributes of the runners (gender and age) and with the maximum temperature on the day of the race. These relationships mirrored known differences in morphology and physiology of participants and the deterioration of these biological attributes with age (senescence), as well as the known effects of hypo- and hyperthermia. **Discussion**. By relating the variation in parameter values to possible biological and environmental variables, the marathon example demonstrates the ability of the distribution to help identify possible triggers and drivers of the duration, shape and temporal shift of its temporal distribution. This more detailed account of the effect of biological and environmental variables would provide a deeper insight into the drivers of a wide variety of phenological phenomena of high current interest, such as the shifting patterns of leafing, flowering, growth, migration, etc. of many organisms worldwide.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27175v1 | CC BY 4.0 Open Access | rec: 6 Sep 2018, publ: 6 Sep 2018

The time distribution of biological phenomena – illustrated with the London marathon

(A biologically interpretable time distribution)

Miguel Franco

School of Biological and Marine Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA

M.Franco@plymouth.ac.uk
ORCID: 0000-0002-7249-4981
Tel +44 (0)1752 584626
Fax +44 (0)1752 584605

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27175v1 | CC BY 4.0 Open Access | rec: 6 Sep 2018, publ: 6 Sep 2018

Abstract

Background. The time distribution of biological phenomena (phenology) is a subject of wide interest, but a general statistical distribution to describe and quantify its essential properties is lacking. Existing distributions are limiting, if not entirely inappropriate, because their parameters do not in general correlate with biologically relevant attributes of the organism and the conditions under which they find themselves.

Methods. A distribution function that allows quantification of three essential properties of a biological dynamic process occurring over a continuous timescale was derived from first principles. The distribution turned out to have three parameters with clear meanings and units: (i) a scaled rate of completion (dimensionless), (ii) a measure of temporal concentration of the process (units: time^{-1}), and (iii) an overall measure of temporal delay (units: time). Its performance as an accurate description of the process was tested with completion data for the London Marathon employing non-linear regression.

Results. The parameters of the distribution correlated with biological attributes of the runners (gender and age) and with the maximum temperature on the day of the race. These relationships mirrored known differences in morphology and physiology of participants and the deterioration of these biological attributes with age (senescence), as well as the known effects of hypo- and hyperthermia.

Discussion. By relating the variation in parameter values to possible biological and environmental variables, the marathon example demonstrates the ability of the distribution to help identify possible triggers and drivers of the duration, shape and temporal shift of its temporal distribution. This more detailed account of the effect of biological and environmental variables would provide a deeper insight into the drivers of a wide variety of phenological phenomena of high current interest, such as the shifting patterns of leafing, flowering, growth, migration, etc. of many organisms worldwide.

Keywords: endurance sports, gender differences, phenology, senescence, time distribution

Introduction

The time course of biological phenomena, often measured as the time to completion of a particular event (such as hatching, leafing, flowering, germination or the completion of an athletics race) is of theoretical and practical interest (Berry et al. 1988; Edwards & Richardson 2004; El Helou et al. 2012; Johnson-Groh & Lee 2002). How biological phenomena unfold with time is determined by the interaction of intrinsic biological features of the organism (genetics, morphology, physiology) and environmental influences acting upon it (nutriments, conditions). These interacting influences determine the triggering of the phenomenon, its rate of occurrence, its duration, and generally the shape of the resulting statistical distribution, which is rarely normal (e.g., Fig. 2 in El Helou et al. 2012). Both symmetrical (e.g., Gaussian) and asymmetrical distributions (e.g., Richards function) are often employed to quantify these time courses. However, although their statistical moments are useful in themselves, their parameter values cannot be interpreted in a meaningful way, a fact recognized by Richards himself (Richards 1959). It would therefore be ideal to have a model whose parameters identify specific aspects of the distribution that account, for example, for biological differences between organisms and the environments under which they find themselves. If specific biological and/or environmental variables affect individual parameter values in predictable ways, the parameters would provide useful insight into the possible biological mechanisms involved. In addition to a good statistical fit, the most important aspect of the distribution must surely be its ability to account for these effects, especially if aimed, beyond description and quantification, at a mechanistic understanding of the process under study.

With these ideas in mind, the aims of this investigation were: (i) to develop a model of the time course of biological phenomena from first principles, (ii) to obtain its essential statistical properties, and (iii) to illustrate the insight that it provides on essential components of a biological time course. Given the completeness of records for the London Marathon, I chose to illustrate the usefulness of the model employing data from several instances of this athletics race. It would seem, however, that the model is potentially applicable to a vast number of temporal distributions (phenologies), perhaps including molecular and cellular processes too.

The model

The simplest time distribution is one that occurs at a constant rate. If hatching, invasion or completion of a race occurred at a constant rate, the completion of events would naturally follow the exponential distribution, $^{y}/_{y_{max}} = 1 - e^{x\ln(1-r)}$or $^{y}/_{y_{max}} = 1 - (1-r)^{x}$, where y/y_{max} is the fraction of the final number of completed events (y_{max}) after x units of time, and r is their rate of occurrence. Alternatively, the time course of completions from y=0 to y=y_{max} would be described by:

$$y = y_{max}(1-(1-r)^{x}) \qquad (1)$$

If a biological time course does not follow this exponential distribution, the simplest conclusion is that r is not constant. The question then becomes whether r changes in a systematic, predictable fashion. The completion of events can be thought of as a probabilistic manifestation of a phenomenon determined by a variety of attributes of the organism and the conditions under which it finds itself. This probabilistic feature is appropriately described by the logit, the logarithm of the odds, i.e., the ratio of “non-event” to “event” in a binomial process. The inverse logit converts the logarithm of the odds into a probability (the probability of hatching, invading or completing the race) making the inverse logit (the logistic function) a natural choice to describe an expected monotonic change in the probability of the event occurring with time under a given set of conditions. The general form of the logistic function (including a “position” or time-delay parameter t, which would seem necessary for any biological process) is: $\frac{1}{1+e^{-c(x-t)}}$, where c and t are constants.

Applying this function to the rate of completion of events, r, results in

$$r = \frac{r_{max}}{1+e^{-c(x-t)}} \qquad (2)$$

where r_{max} is the maximum biologically achievable proportional rate of completion under particular conditions. Substitution of equation 2 into equation 1 yields

$$y = y_{max}\left(1-\left(1-\frac{r_{max}}{1+e^{-c(x-t)}}\right)^{x}\right) \qquad (3)$$

This equation describes the essential features of the temporal dynamics of a biological process. Beyond the exponential distribution, it probably is the simplest interpretable form of the time distribution of a biological process. Normalization of equation 3 (by dividing both sides by y_{max})

yields the underlying cumulative distribution function (*cdf*) of the time course. $y_{\max}$ simply scales the *cdf* to the total count, i.e., the final number of organisms hatching, invading or completing a race. Parameter $r_{\max}$ represents the maximum proportional rate of completion (or of whatever process the model is applied to: hatching, leafing, invasion, etc.). Parameter c is the rate of realisation of $r_{\max}$ (r tends towards $r_{\max}$ faster as c increases) and t measures the delay in the realisation of $r_{\max}$. It should be clear, however, that while $r = r_{\max}$ is the constant explicitly quantified by the model, the instantaneous (cumulative) r increases with time towards $r_{\max}$ (eqn. 2).

The derivative of equation 3 yields the probability density function (*pdf*; Supplementary Material 1). This equation is lengthy, which may explain the reason it has remained undiscovered as a general model of the temporal distribution of a biological process (formulas for the distribution's statistical moments are also presented in SM1).

The *cdf* mimics the variety of monotonically increasing shapes of the completion of events, and the influence that each parameter has on the shape of both *cdf* and *pdf* is clear (Fig. 1). Parameter r determines the rate with which the *cdf* rises, producing "diverging" trajectories. Parameter c, on the other hand, shortens the timespan over which the majority of the process occurs: increasing values of c reduce the spread of the *pdf* (c is a measure of concentration of the time distribution). Note that *cdf*s differing only in their value of c intersect each other at $x = t$. Finally, parameter t delays the process of completion producing "parallel", delayed *cdf*s and delayed *pdf*s. The model can produce normal-looking as well as truncated, left- or right-skewed, platykurtic or leptokurtic distributions. It must be emphasised, however, that its main strength is that it provides us with the ability to interpret effects in terms of the magnitude with which each parameter contributes to the shape of the distribution. This is crucial to understand, for example, whether temperature increases the rate (*r*), concentration (*c*), temporal shift (*t*), or a combination of them in a phenological process. Since, as shown in Fig. 1, these are three aspects of the model which are distinct from its statistical moments, common statistical distributions defined by their mean and standard deviation are insufficient to account for these biologically interpretable effects.

Materials and Methods

Full datasets for the London Marathon in years 2010, 2011 and 2016 were obtained from the London Marathon website (London Marathon 2016). In addition, data were downloaded for the period 2001-2009 from marathonguide.com (Marathon Guide 2016). Because results from marathonguide.com could only be downloaded in subsets of 100 completions, data were obtained for the first 100 completions in each set of 1000 consecutive runners, i.e. runners finishing in positions 1-100, 1001-1100, 2001-2100, etc., plus the last minimum consecutive 100 runners beginning at a “hundred and one” position. For example, in 2008, 23574 men completed the race and the data downloaded and used in the analyses consisted of the first 24 subsets of 100 runners between positions 1 and 23100, plus the last 174 runners occupying positions 23401 to 23574. This meant that sample size was larger at the end of the distribution, but this was preferred to the alternative of leaving a wider gap between the last two subsets of data. This allowed downloading of the data faster and had a minimal influence on the results.

Two sets of analyses were conducted. In the first one, the model was fitted to different subsets of participants in the London Marathon 2016: (i) for all runners combined, (ii) separately for men and women, and (iii) separately within each gender for each of 5 age categories: 18-39, 40-49, 50-59, 60-69 and ≥70 years. Although runners in the three intermediate categories are classified in five-year intervals (40-44, 45-49, etc.), given the larger range of ages in the two end categories (youngest and oldest), the original six five year-long intermediate categories were combined into three ten year-long categories. In the second set of analyses, the model was fitted separately to men and women for each marathon event in the period 2001-2011, for which it was possible to investigate the possible influence of prevalent weather conditions on the date each race was run.

As explained before, the model is independent of y_{max}, which only scales the distribution to the number of completions, and thus the proportion of completions (y/y_{max}) was used for model fitting. The *cdf* was fitted to each set/subset by non-linear least squares regression employing the Levenberg-Marquardt algorithm in IBM SPSS 23 (IBM 2016). The initial parameter values employed (r=0.4, c=1.6, t=4.8) produced convergence in ≤7 iterations in all cases but one, women in age category 70+. In this case, r was estimated to be >1, which is a biological impossibility. This parameter was then constrained to different values in the range (0.3 to 0.6) and the performance of the fit judged by the standard error of parameter estimates, which tended to be high compared to those in other age categories. Given the observed trend in the decrease of r with age obtained from the other four age categories, a value of $r = 0.31$, which kept all three parameter errors as small as it seemed possible, was chosen as acceptable (see Results). Once all parameters had been estimated, the *pdf* and statistical moments (SM1) were calculated (SM2) in Maple 2016 (Maple 1996-2016).

For the first set of analyses, parameters values were correlated with age employing the mean age of the first four categories listed above (28.5, 44, 54, 64 years) and an approximate guess for the mean age in the last category (74 years; the maximum age was 75 and 74 years in (Lara et al. 2014) and (Zavorsky et al. 2017), respectively). For the second set of analyses, the weather conditions reported for Heathrow airport at each marathon date were obtained from the Met Office. In order to investigate if weather conditions influenced desertion, the organizers of the London marathon supplied the figures for the number of runners who started the race, but these figures were separated by gender only for the period 2005-2011. Because of this, the overall proportion of completions combining men and women was used. Proportions were logit transformed before analysis (Warton & Hui 2011). Only relationships between model parameters and maximum temperature are presented because minimum and maximum temperatures were correlated and yielded similar relationships with the model's parameter. Besides, the race takes place in the daytime, when the maximum temperature is reached. Rainfall and sunshine hours did not correlate with each other, with temperature, or with distribution parameters, and are therefore not referred to in the Results.

Results

Influence of gender and age

The London Marathon 2016 was completed by 23940 men and 15048 women, providing ample sample sizes for model fit. While a single model fitted all 38988 data points (figure not shown), separating runners by gender provided equally good fits (Fig. 2; SM3 Table S1). Men and women differed in the values of all three parameters (95% confidence intervals for all three parameters ≤ 0.003 from their estimated values in all three fits). Men ran faster (higher r), had a higher rate of realisation of r (c value), and took a shorter time to run the race (smaller t) than women (SM3 Table S1). A one-way analysis of variance of completion times yielded a significant difference between the sexes ($F_{1,38986} = 3899.6$, $P < 0.001$; mean for men = 4.81 hours, mean for women = 4.20 hours) but Levene's test of homogeneity of variances indicated heteroscedasticity ($F_{1,38986} = 48.94$, $P < 0.001$), which is confirmed by their statistical moments (SM3 Table S1). Indeed, all distributions (including those discussed below, where runners were classified by age, and those in the next section, where the effect of temperature is investigated) had large positive skew and excess kurtosis (SM3 Tables S1 & S2). In the majority of cases, mode < median < mean (in a few cases median < mode < mean), and the arithmetic mean consistently overestimated the mean calculated from the model. The right-skewed and consistently leptokurtic nature of the distribution makes the model more realistic and accurate in the estimation of statistical moments than the normal distribution. For comparison, standard deviation, skewness and kurtosis calculated using the standard formulas for sample moments yielded values <1 in all cases, clearly underestimating them.

Parameter r showed a declining relationship with age, the difference between the curves for the two genders becoming smaller with age (Fig. 3a; quadratic fit: men $R^2 = 0.975$, $P = 0.002$; women $R^2 = 0.930$, $P = 0.008$). Parameter c increased with age for the first four age categories in both sexes, with women's values lower than men's, but dropped and converged for both sexes in the last age category (Fig. 3b); it did not neatly fit a continuous function (e.g., quadratic). Parameter t showed a quadratic relationship with age for both men and women, (Fig. 3c; men $R^2 = 0.983$, $P = 0.001$; women $R^2 = 0.922$, $P = 0.01$). Mode, median and mean showed significant quadratic relationships with age similar to that of t, with which they were highly correlated ($R > 0.965$, $P < 0.01$ for all pair combinations of t, mode, median and mean within each sex). Only the relationship between median and age is shown (Fig. 3d; men $R^2 = 0.996$, $P < 0.001$; women $R^2 = 0.965$, $P = 0.003$). The quadratic relationships for t and median had optima (minimal values) for men at ages 43.1 and 40.9, respectively; while for women these optima occurred at ages 44.3 and 42.5 years, respectively (Figs. 3c and d).

The number of participants decreased with age (N in SM3 Table S1). This smaller sample size was accompanied by an increase in the error of parameter estimates, particularly for women in the oldest category, even after constraining parameter r (Figs. 3a-c).

Influence of temperature

Sampling just above 10% of runners for the years 2001-2009 produced similarly high R^2 values and only slightly larger standard errors to those obtained with the full 2010 and 2011datasets (SM3 Table S2). Maximum temperature (T_{max}) produced a significant linear relationships with c (Fig. 4b), t (Fig. 4c), mode, median, mean and entropy, but not with r (Fig. 4a) and standard deviation, for both sexes (SM3 Table S3). On the other hand, skewness and kurtosis in men, but not in women, showed significant linear relationships with T_{max} (SM3 Table S3). While there was no evidence of a different relationship between c and T_{max} for men and women (Fig 4b), t and measures of central tendency indicated parallel relationships, with women taking longer to complete the marathon (Fig. 4c). As with the data for 2016, women had lower values of parameter r than men (Fig. 4a; men's mean $r = 0.421$, SD = 0.011; women's mean $r = 0.390$, SD = 0.021). Finally, desertion from the race increased away from an optimal T_{max} for completion of 17.1 °C (Fig. 4).

Discussion

Model suitability

As a description of a time course, the model is conceptually simple and has three characteristics that make it preferable over other distributions. First, it is derived from basic principles that take into account the essential elements of a biological time course. This circumvents conceptual and practical problems derived from the use of ad-hoc sigmoid functions designed to describe radically different phenomena. Second, by differentiating between the different aspects or parameters the model allows their unique characterization and quantification. Third, the quantification of these properties allows investigation of the influence that biological (gender and age) and environmental factors (temperature) have on each of them. All three parameters were influenced by gender and age, which are under biological control (genetics/life history), but only *c* and *t* were affected by temperature, an environmental effect. These results indicate that permanent biological attributes (permanent in relation to the duration of the race) influenced the runners' intrinsic speed (*r*), the degree of concentration of the race (*c*), and the delay in its completion (*t*) (Fig. 3). On the other hand, the maximum ambient temperature of the day, an environmental influence, spread the finishing times (lowering *c*) without apparent distinction of gender (Fig. 4b), and increased the race's duration (*t* and other measures of central tendency) in a similar fashion in both sexes (parallel lines in Fig. 4c), but did not significantly affect the runner's intrinsic speed (*r*) (Fig. 4a).

Gender and age differences

The completion of a marathon has been a popular method by which physical and physiological differences between sexes and age groups have been investigated (Connick et al. 2015; Conoboy & Dyson 2006; Jokl et al. 2004; Zavorsky et al. 2017). To begin with, a variety of morphological, physiological and behavioural differences exists between the sexes (Ellis et al. 2008). Although the source of some of these differences in humans may be controversial, there is ample support for biological differences in athletics performance between the sexes (Connick et al. 2015; Lippi et al. 2008). It is generally accepted that the main factor contributing to endurance running is aerobic capacity, as measured by maximum oxygen uptake and its interaction with muscle mass distribution, liver and muscle glycogen content and exercise intensity as a fraction of aerobic capacity (Mahler & Loke 1985; Rapoport 2010; Sjodin & Svedenhag 1985). Thus, the difference in marathon completion times between men and women is primarily attributed to the larger aerobic capacity and muscular strength of men (Cheuvront et al. 2005), and this difference is reflected in all three parameters of the model (Fig. 3). In particular, parameter *r* follows the known accelerating decline of aerobic capacity with age (e.g., Fig. 6b and Fig. 8 in Tanaka & Seals 2003, though their Fig. 8 was fitted to a straight line). The difference between the sexes in parameter values decreased with age (Fig. 3), indicating that men's performance drops faster with age than it does in women. The quadratic models fitted to parameters *r* and *t* (Fig. 3) predict that the curves for men and women would cross at ages 95 years and 82 years, respectively. Taken as measures of the rate of senescence, these patterns predict that, by senescing more slowly, women should reach older ages than men.

The idea that men age faster than women is supported by demographic and physiological studies (Austad 2006; Blagosklonny 2010; Celermajer et al. 1994; Gallagher et al. 2000; Graves et al. 2006), and the oldest known living people are women (Wikipedia 2018). This idea is also consistent with the observation that larger organisms within the same species senesce faster than smaller ones (Austad 2010; Kraus et al. 2013), though the regularity of this pattern and, more importantly, the reasons for it are unclear. There are, however, reports on sport performance where a faster drop was found in women than in men. For example, a faster drop in women's than in men's performance with age was observed in a 10 km athletics race and in 1500 m swimming (Tanaka & Seals 2003), and a similar result was obtained for elite marathon runners in Germany (Leyk et al. 2007). It is recognised, however, that such result may be due to "selection bias": the lower participation of women in many sports, particularly at more advanced ages (Tanaka & Seals 2003); there is indeed evidence of small sample size affecting the statistical trends observed (fewer points at advanced ages that show convergence between the sexes) in their figures 1 and 2 . Even then, it must be emphasised that the assumption of normality in all these studies misrepresents the values of all statistical moments. Sample size was indeed a limitation in the analysis of the oldest age categories for the London Marathon data (witness error of parameter estimates in Figs. 3a-c). However, by fitting a model of the expected course of the marathon from which to derive its distribution parameters, the model presented here aims to quantify comparable parameter values and distribution moments for the whole participant population (and subpopulations) that may be more robust to variation in sample size. Other studies have focused on elite athletes or on a small sample of competitors, and all have assumed a normal distribution of finishing times (Cejka et al. 2015; Connick et al. 2015; Lara et al. 2014; Leyk et al. 2007; Zavorsky et al. 2017). Those results are therefore not fully comparable with the results presented here.

The consistent right skew and leptokurtic nature of the distributions is likely a consequence of the dwindling number of participants with age and the lower proportion of women. No matter how we define the subsamples, there is a tail of fewer older individuals, and of women who also participate in lower numbers (Fig. 2; witness also the higher values of standard deviation, skewness, kurtosis and entropy for women compared to men, and their increase and converge with age, some with a drop in the last age class, in SM3 Table S1). There must also be other uncontrolled differences. As an example, the frequency of self-imposed handicaps, such as the varied costumes that some runners sport, likely varies with gender and age. This effect is difficult to isolate because the records do not contain such information – except the fact that a runner may be linked to a charity. The heterogeneity found in older categories with fewer competitors also makes it difficult to estimate the parameters accurately, so sex differences in aging may remain difficult to ascertain without more detailed *in-situ* morphological and physiological measures of runners, e.g. before, during and/or after the race.

A morphological attribute thought to have explanatory power on endurance running is calcaneus length (the endurance running hypothesis, Carrier et al. 1984). Clear differences exist in calcaneus length between the sexes (Riepert et al. 1996), and differences between Neanderthals and humans (corrected for body size) have been linked to the capacity for endurance running (Raichlen et al. 2011). The current evidence suggests that shorter calcaneus length, and consequently shorter moment arm of the Achilles tendon (Scholz et al. 2008) confer advantage in endurance running, and it would be interesting to investigate the role of these and possibly other morphological attributes on the running economy of marathon participants employing the model presented here.

Environmental influences

With regard to environmental effects, evidence suggests that colder temperatures favour faster marathons (Adams et al. 1975; Ely et al. 2007; Montain et al. 2007), but these can also impair some runners (Jones et al. 1985). The only detectable influence at the London Marathon, which is run in April each year, was maximum temperature. Within the range of temperatures observed, the intrinsic ability of the competitors (*r*) was not significantly affected (Fig. 4a). However, temperatures were sufficient to produce evidence of exhaustion (longer time to completion/larger *t*) and thus spreading of the race (lower *c*) as T_{max} increased (Fig. 4b-c), but maximum race completion was predicted to occur at 17.1 °C (Fig. 4d). Around this temperature, runners are not discomforted by lower than preferred temperatures and are less prone to become overheated.

If the composition of the race changes every year, as it must surely do, parameter values may change even in the absence of environmental differences between events. On the other hand, temperature affects runners in relation to their running ability (Ely et al. 2008). Both effects would be expected to combine and produce the variability around the lines of best fit in figure 4. The large sample size, however, would be expected to override these compositional differences. Finally, there are psychological factors affecting the decision to abandon the race or encourage clumping, which must also influence the variation in the trends shown in figure 4d. Clumping has been analysed and discussed by other authors (Allen et al. 2016; Alvarez-Ramirez et al. 2007), but for reasons of space it is not explored here. A detailed analysis of the residuals from the model fit would provide insight on this matter. Extreme marathons would also provide useful comparisons, but they often involve longer distances (ultra-marathons) and/or special footwear/clothing (e.g., the North Pole Marathon) that modify other aspects/conditions of the race. These marathons also tend to have few competitors, making the estimation of parameters unreliable – a record of 56 participants in the North Pole Marathon for 2016 (North Pole Marathon 2016).

Conclusion

By defining and quantifying three essential aspects of a biological time course, the time distribution presented here provides a standard with which to evaluate specific hypothesis regarding the influence of biological and environmental variables. The richness of information that the model provides allowed exploration of several hypothesis posed in the literature regarding the role that morphological (e.g., body size and calcaneous length), physiological (e.g., endurance as measured by aerobic capacity), life history evolution (senescence) and environmental (maximum air temperature) variables may exert on marathon completion times. More generally, the example illustrates the potential applicability of the distribution to many other biological phenomena under the broad heading of phenology. Previously, we used the distribution to confirm the hypothesis of a fast-slow continuum of plant life histories measured on reproductive value (Mbeau-Ache & Franco 2013), a parameter of interest in evolutionary theories of senescence (Fisher 1930; Partridge & Barton 1996) and my research group is currently working on several long-term datasets of phenology in both plants and animals.

Acknowledgements

I thank David Franco for his help with the derivation of the distribution's statistical moments. I thank Jordan Golubov, Mick Hanley, María C. Mandujano, Carlos Montaña, Alma Orozco, Paul Ramsay and Nicola Steer for useful discussion during the development of this work.

References

Adams WC, Fox RH, Fry AJ, and MacDonald IC. 1975. Thermoregulation during marathon running in cool, moderate, and hot environments. *Journal of Applied Physiology* 38:1030-1037.

Allen EJ, Dechow PM, Pope DG, and Wu G. 2016. Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science* 63:1657-1672. 10.1287/mnsc.2015.2417

Alvarez-Ramirez J, Rodriguez E, and Dagdug L. 2007. Time-correlations in marathon arrival sequences. *Physica A: Statistical Mechanics and its Applications* 380:447-454. 10.1016/j.physa.2007.03.008

Austad SN. 2006. Why women live longer than men: Sex differences in longevity. *Gender Medicine* 3:79-92. 10.1016/S1550-8579(06)80198-1

Austad SN. 2010. Animal Size, Metabolic Rate, and Survival, Among and Within Species. In: Wolf NS, ed. *The Comparative Biology of Aging*. Dordrecht: Springer Netherlands, 27-41.

Berry GJ, Cawood RJ, and Flood RG. 1988. Curve fitting of germination data using the Richards function. *Plant, Cell and Environment* 11:183-188. 10.1111/j.1365-3040.1988.tb01135.x

Blagosklonny MV. 2010. Why men age faster but reproduce longer than women: mTOR and evolutionary perspectives. *Aging (Albany NY)* 2:265-273.

Carrier DR, Kapoor AK, Kimura T, Nickels MK, Satwanti, Scott EC, So JK, and Trinkaus E. 1984. The Energetic Paradox of Human Running and Hominid Evolution [and Comments and Reply]. *Current Anthropology* 25:483-495. 10.1086/203165

Cejka N, Knechtle B, Rüst CA, Rosemann T, and Lepers R. 2015. Performance and Age of the Fastest Female and Male 100-km Ultramarathoners Worldwide From 1960 to 2012. *The Journal of Strength & Conditioning Research* 29:1180-1190. 10.1519/jsc.0000000000000370

Celermajer DS, Sorensen KE, Spiegelhalter DJ, Georgakopoulos D, Robinson J, and Deanfield JE. 1994. Aging is associated with endothelial dysfunction in healthy men years before the age-related decline in women. *Journal of the American College of Cardiology* 24:471-476. 10.1016/0735-1097(94)90305-0

Cheuvront SN, Carter R, DeRuisseau KC, and Moffatt RJ. 2005. Running Performance Differences between Men and Women. *Sports Medicine* 35:1017-1024. 10.2165/00007256-200535120-00002

Connick MJ, Beckman EM, and Tweedy SM. 2015. Relative Age Affects Marathon Performance in Male and Female Athletes. *Journal of Sports Science & Medicine* 14:669-674.

Conoboy P, and Dyson R. 2006. Effect of aging on the stride pattern of veteran marathon runners. *British Journal of Sports Medicine* 40:601-604. 10.1136/bjsm.2006.026252

Edwards M, and Richardson AJ. 2004. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* 430:881-884. 10.1038/nature02808

El Helou N, Tafflet M, Berthelot G, Tolaini J, Marc A, Guillaume M, Hausswirth C, and Toussaint J-F. 2012. Impact of Environmental Parameters on Marathon Running Performance. *PLoS ONE* 7:e37407. 10.1371/journal.pone.0037407

Ellis L, Hershberger S, Field E, Wersinger S, Pellis S, Geary D, Palmer C, Hoyenga K, Hetsroni A, and Karadi K. 2008. *Sex differences: summarizing more than a century of scientific research*. New York: Psychology Press.

Ely MR, Cheuvront SN, Roberts WO, and Montain SJ. 2007. Impact of Weather on Marathon-Running Performance. *Medicine & Science in Sports & Exercise* 39:487-493. 10.1249/mss.0b013e31802d3aba

Ely MR, Martin DE, Cheuvront SN, and Montain SJ. 2008. Effect of ambient temperature on marathon pacing is dependent on runner ability. *Medicine & Science in Sports & Exercise* 40:1675-1680.

Fisher RA. 1930. *The genetical theory of natural selection* Oxford, UK: Clarendon Press.

Gallagher D, Ruts E, Visser M, Heshka S, Baumgartner RN, Wang J, Pierson RN, Pi-Sunyer FX, and Heymsfield SB. 2000. Weight stability masks sarcopenia in elderly men and women. *American Journal of Physiology - Endocrinology And Metabolism* 279:E366-E375.

Graves BM, Strand M, and Lindsay AR. 2006. A reassessment of sexual dimorphism in human senescence: Theory, evidence, and causation. *American Journal of Human Biology* 18:161-168. 10.1002/ajhb.20488

IBM. 2016. IBM SPSS Statistics for Windows, Version 23.0. Armonk, New York: IBM Corp.

Johnson-Groh CL, and Lee JM. 2002. Phenology and demography of two species of Botrychium (Ophioglossaceae). *American Journal of Botany* 89:1624-1633.

Jokl P, Sethi PM, and Cooper AJ. 2004. Master's performance in the New York City Marathon 1983–1999. *British Journal of Sports Medicine* 38:408-412. 10.1136/bjsm.2002.003566

Jones BH, Rock PB, Smith LS, Teves MA, Casey JK, Eddings K, Malkin LH, and Matthew WT. 1985. Medical Complaints After a Marathon Run in Cool Weather. *The Physician and Sportsmedicine* 13:103-110. 10.1080/00913847.1985.11708904

Kraus C, Pavard S, and Promislow DEL. 2013. The size–life span trade-off decomposed: why large dogs die young. *The American naturalist* 181:492-505. 10.1086/669665

Lara B, Salinero JJ, and Del Coso J. 2014. The relationship between age and running time in elite marathoners is U-shaped. *AGE* 36:1003-1008. 10.1007/s11357-013-9614-z

Leyk D, Erley O, Ridder D, Leurs M, Rüther T, Wunderlich M, Sievert A, Baum K, and Essfeld D. 2007. Age-related changes in marathon and half-marathon performances. *International Journal of Sports Medicine* 28:513-517. 10.1055/s-2006-924658

Lippi G, Banfi G, Favaloro EJ, Rittweger J, and Maffulli N. 2008. Updates on improvement of human athletic performance: focus on world records in athletics. *British Medical Bulletin* 87:7-15. 10.1093/bmb/ldn029

London Marathon. 2016. London Marathon. *Available at* https://www.virginmoneylondonmarathon.com/en-gb/ (accessed 24/04/2017).

Mahler DA, and Loke J. 1985. The physiology of marathon running. *The Physician and Sportsmedicine* 13:84-97. 10.1080/00913847.1985.11708729

Maple. 1996-2016. Maple User Manual. Waterloo, Canada: Maplesoft, Maple, Inc.

Marathon Guide. 2016. Marathon Guide. *Available at* http://www.marathonguide.com/ (accessed 24/04/2017).

Mbeau-Ache C, and Franco M. 2013. The time distribution of reproductive value measures the pace of life. *Journal of Ecology* 101:1273-1280. 10.1111/1365-2745.12131

Montain SJ, Ely MR, and Cheuvront SN. 2007. Marathon performance in thermally stressing conditions. *Sports Medicine* 37:320-323. 10.2165/00007256-200737040-00012

North Pole Marathon. 2016. North Pole Marathon. *Available at* http://www.npmarathon.com/ (accessed 15/01/2017).

Partridge L, and Barton NH. 1996. On measuring the rate of ageing. *Proceedings of the Royal Society of London Series B: Biological Sciences* 263:1365-1371. 10.1098/rspb.1996.0200

Raichlen D, Armstrong H, and Lieberman D. 2011. Calcaneus length determines running economy: implications for endurance running performance in modern humans and Neandertals. *Journal of human evolution* 60:299-308. 10.1016/j.jhevol.2010.11.002

Rapoport BI. 2010. Metabolic factors limiting performance in marathon runners. *PLOS Computational Biology* 6:e1000960. 10.1371/journal.pcbi.1000960

Richards FJ. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany* 10:290-301. 10.1093/jxb/10.2.290

Riepert T, Drechsler T, Schild H, Nafe B, and Mattern R. 1996. Estimation of sex on the basis of radiographs of the calcaneus. *Forensic Science International* 77:133-140. 10.1016/0379-0738(95)01832-8

Scholz MN, Bobbert MF, van Soest AJ, Clark JR, and van Heerden J. 2008. Running biomechanics: shorter heels, better economy. *The Journal of Experimental Biology* 211:3266-3271. 10.1242/jeb.018812

Sjodin B, and Svedenhag J. 1985. Applied Physiology of Marathon Running. *Sports Medicine* 2:83-99. 10.2165/00007256-198502020-00002

Tanaka H, and Seals DR. 2003. Invited Review: Dynamic exercise performance in Masters athletes: insight into the effects of primary human aging on physiological functional capacity. *Journal of Applied Physiology* 95:2152-2162. 10.1152/japplphysiol.00320.2003

Warton DI, and Hui FKC. 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92:3-10.

Wikipedia. 2018. Oldest people. *Available at https://en.wikipedia.org/wiki/Oldest_people* (accessed 10/06/2018).

Zavorsky GS, Tomko KA, and Smoliga JM. 2017. Declines in marathon performance: Sex differences in elite and recreational athletes. *PLoS ONE* 12:e0172121. 10.1371/journal.pone.0172121

Figure legends

Fig. 1. The influence of parameters *r* (panels a and b), *c* (c and d) and *t* (e and f) on the shape of the cumulative distribution function (left panels) and the probability density function (right panels). In the order blue, red and green lines, parameter values are: a & b: r=0.4, 0.6, 0.9, c=1.5, t=4.5; c & d: r=0.4, c=1.5, 2.5, 4.5, t=4.5; e & f: r=0.4, c=1.5, t=4.5, 5.5, 6.5.

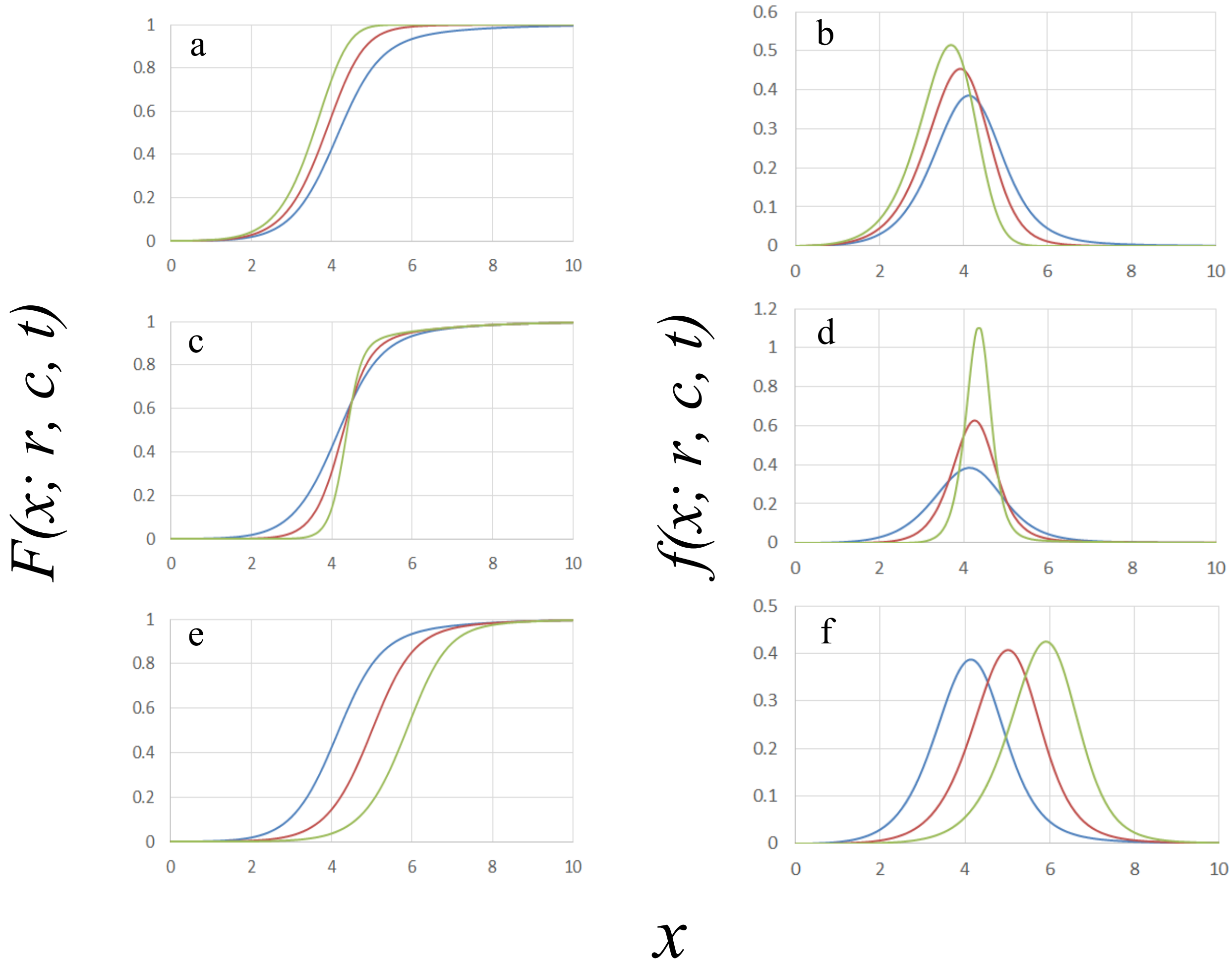
Fig. 2. The cumulative distribution functions and corresponding probability density functions for men and women completing the London Marathon 2016. 23940 male runners (blue circles and lines) and 15048 female runners (red circles and lines) completed the race. Note that the fitted *cdf*s (continuous lines in first panel) are only visible at both ends where the model departs from the observed completion times.

Fig. 3. The relationship between model parameters estimated for the London Marathon 2016 and the mean age of competitors in each category. Men: blue symbols and lines; women: red symbols and lines. Error bars represent standard error of parameter estimates (*r*, *c* and *t*). Medians estimated from the estimated parameter values are exact.

Fig. 4. The relationship between model parameters for the London Marathon 2001-2011 and maximum temperature on the day of each race. Standard errors of parameter estimates were smaller than the diameter of the points (see Table S3) and are therefore not visible (panels a-c). The proportion of completions is a scalar for each race and it is not possible to calculate an error. Men: blue symbols and lines; women: red symbols and lines.

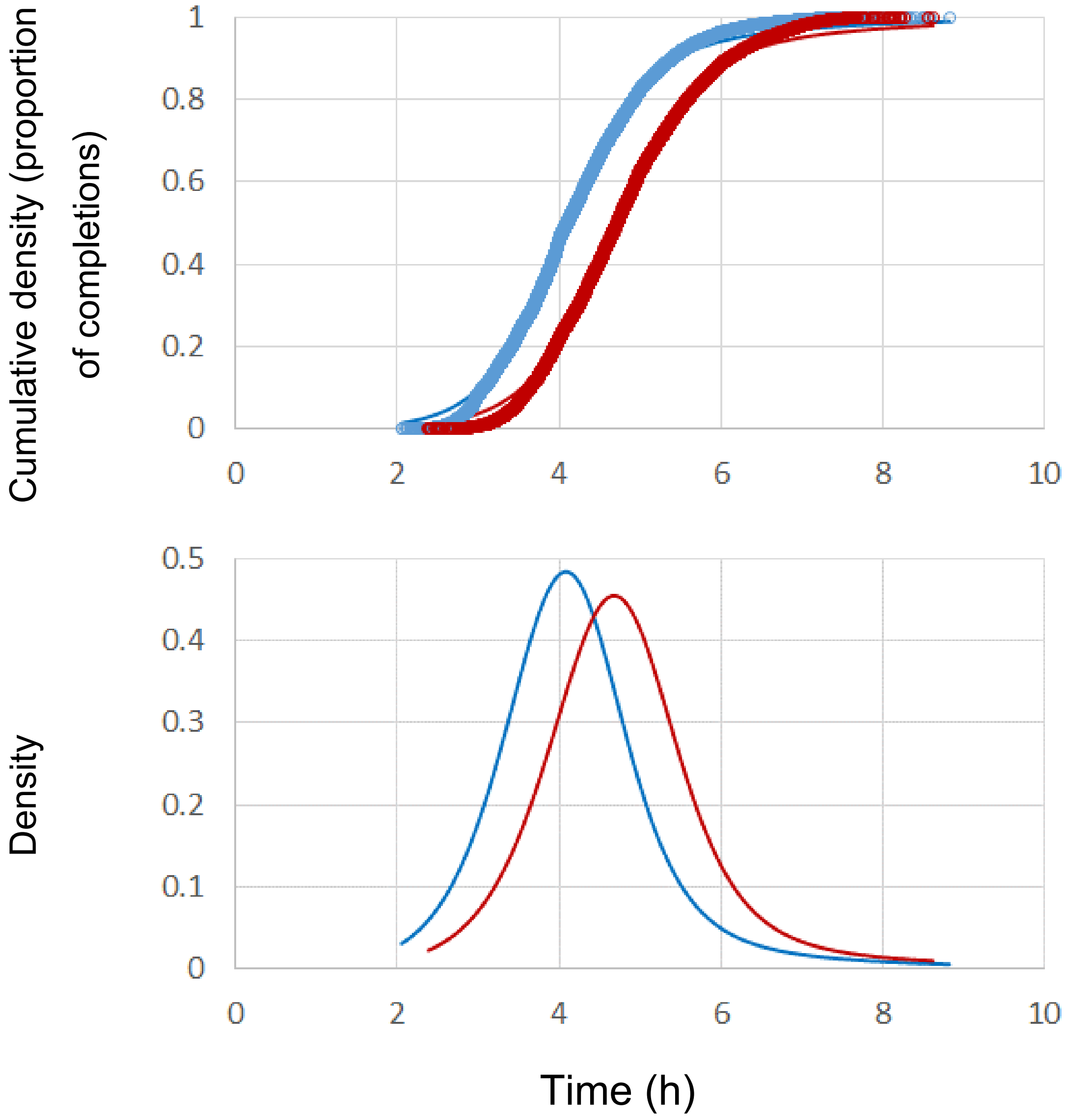
502 Figure 1

503



505 Figure 2

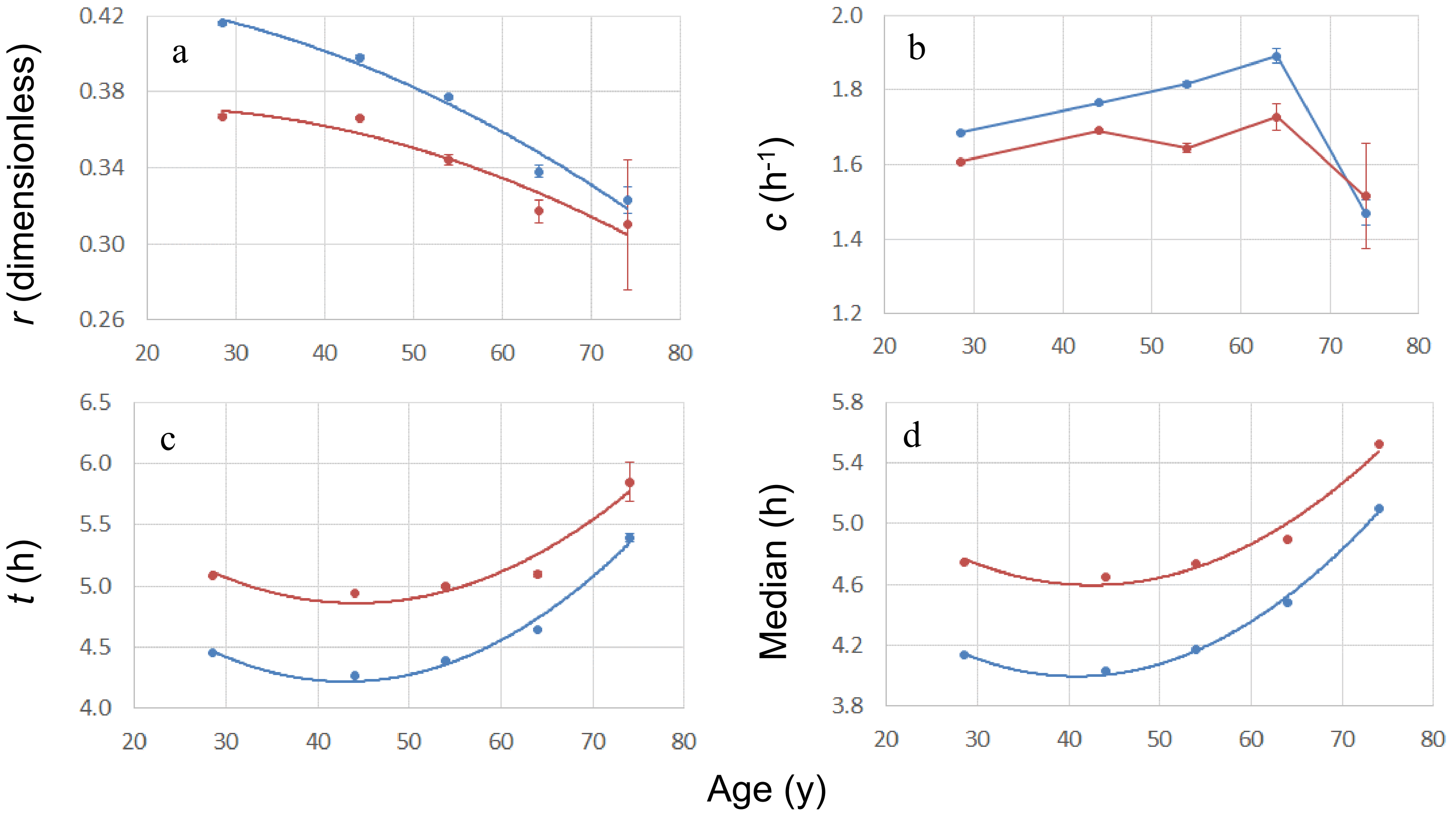
506



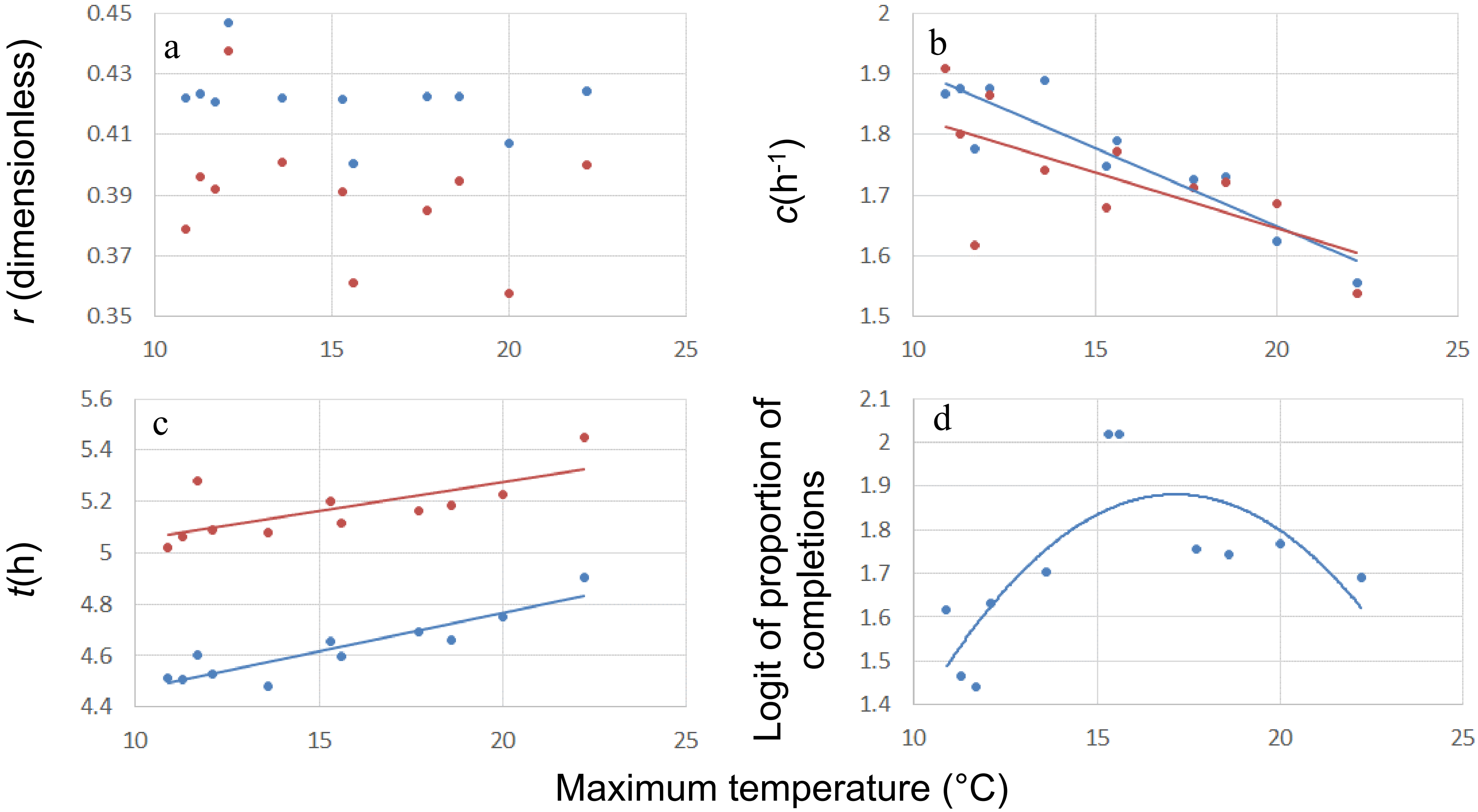
507

508 Figure 3

509



511 Figure 4
512



513