RESEARCH ARTICLE

The Analysis & Forecasting of Male Cycling Time Trial Records Established within England, Wales and Northern Ireland

Bryce Dyer[1]*, Hossein Hassani[2] and Mehran Shadi[2]

[1]*Faculty of Science & Technology, Bournemouth University, Poole; UK*

[2]*The Statistical Research Centre, Bournemouth University, Poole; UK*

Abstract

The format of cycling time trials in England, Wales and Northern Ireland involves riders competing individually over several fixed race distances of 10-100 miles in length and using time constrained formats of 12 and 24 hours in duration. Drawing on data provided by the national governing body that covers the regions of England, Wales and Northern Ireland, an analysis of six male competition record progressions was undertaken to illustrate its progression. Future forecasts are then projected through use of the singular spectrum analysis technique. This method has not been applied to sport-based time series data before. All six records have seen a progressive improvement and are non-linear in nature. Five records saw their highest level of record change during the 1950-1969 time period. Whilst new record frequency generally has reduced since this period, the magnitude of performance improvement has generally increased. The SSA technique successfully provided forecasted projections in the short to medium term with a high level of fit to the time series data.

Keywords: *cycling, forecasting, performance, singular spectrum analysis*

*Corresponding author: brdyer@bournemouth.ac.uk

Introduction

Cycling is regarded as an exercise for physiologically generated mechanical power to overcome any external resistant forces (Jeukendrup et al., 2000). Within competitive cycling, a unique cycling discipline is the *individual time trial.* This discipline requires an individual rider to cover a fixed distance or duration at the highest possible speed. To achieve this requires the highest possible physiologically generated power (Jeukendrup et al., 2000), the reduction and optimisation of aerodynamic drag of the rider and bicycle (Lukes et al., 2005) and the maximization of mechanical efficiency of the cyclist's equipment (Zamparo et al., 2002).

There has been previous analysis of cycling time trial records such as the 'Hour Record' (Padilla et al., 2000), masters 200m and 500m sprint records (Ransdell et al., 2009), the 4000m team pursuit (Schumacher & Mueller, 2002) and as part of a larger study into elite road cycling (Schumacher et al. 2006). However, no attention to date has been placed exclusively on road-based individual time trial cycling. This is possibly since time trials are internationally fewer in number as standalone events and instead, typically act as a partial influence within stage races such as the Tour de France. Within the UK though, this philosophy is fundamentally different and cycling time trials are a distinct and separate strand of cycle sport competition (Whitfield, 2013). These have taken place using several distance or time based formats since 1890 (Whitfield, 2013) with more established fixed records in place since 1935 (www.cyclingtimetrials.org.uk). Whilst the time trial records are often proposed as 'national records', these are in fact records regulated by a single governing body that spans England, wales and Northern Ireland. However, these cannot be classified as British records since Scotland operates time-trialling using slightly different equipment rules.

The consistent format of national cycling time trials involves riders competing individually over several fixed race distances of 10-100 miles in length or using time constrained formats of 12 and 24 hours in duration (www.cyclingtimetrials.org.uk). Each separate distance or duration based time trial has its own national record. A new record can be established and ratified at any point of the year when achieved in an event of the same race length. Whilst not historically always the case, both amateur and professional cyclists are currently eligible to participate and obtain these records. Whilst the distance or duration remains the same, the event race topography a time trial cyclist will race over is an open environment which means any performance is influenced by external factors such as weather, road gradients and the influence of passing motor vehicle traffic. However, whilst these conditions are not standardised or definable per se', their philosophical influence has remained consistent in principle since the sports inception.

Monitoring a sports performance over a period of time through the application of statistical analysis is an indirect method of investigating mathematically how a sports based society has evolved (Lippi et al., 2008; Carbone & Savaglio, 2001). As a result, by illustrating its progression to date and generating forecasts for its future may offer some insight into time-trialling's longer term development. As a result, both a statistical analysis and a forecasting strategy of male time trial records Ratified by the governing body of England, Wales and Northern Ireland) were undertaken.

Methods

The progression of national men's time trial records established in England, Wales and Northern Ireland across several distances were plotted and reviewed. The data was obtained from the sport's governing body website (www.cyclingtimetrials.org.uk) and annual handbook (CTT, 2014). The fixed distances that were investigated were the 10, 25, 50 and 100 mile distances. In addition, the 12 hour and 24 hour fixed duration events were also investigated. A record may be broken more than once in the same event but only the fastest of these is ratified as the new record holder when this occurs. However, a record may be broken and ratified more than once in the same year. The historical analysis investigated the record progression (RP), the number of records achieved per decade (RPD), the number of riders achieving a record per decade (RD) and the record to subsequent record percentage improvement (RRPI).

To act as a forecasting method, the *Singular Spectrum Analysis* (SSA) technique was applied to the historical record progression data. The SSA technique is a novel and powerful technique of time series analysis. Whilst mathematical forecasting with sports records is not new (Desgorces et al., 2008), this particular technique has not been applied to sports-based time series data before.

The possible applications areas of SSA are diverse: from mathematics and physics to economic and financial mathematics, from social science and market research to meteorology and oceanology (see for example: Hassani et al., 2013; Silva and Rajapaksa, 2014; Sanei et al. 2010; Hassani, 2007 and references therein). The aim of SSA is to make a decomposition of the original series into the sum of a small number of independent and interpretable components - such as a slowly varying trend, oscillatory components and a structureless noise (Hassani, 2007).

Basically SSA first decomposes the whole time series into all of the elements which have created the time series (including noise and signal). By then using the special techniques which form part of this method, it tries to distinguish signal from noise and then finally reconstructs the time series creating a time series which is noise free.

SSA is a very useful tool in order to find trends of different resolution, smoothing, extraction of seasonality components, simultaneous extraction of cycles with small and large periods, extraction of periodicities with varying amplitudes, simultaneous extraction of complex trends and periodicities, finding structure in short time series and change-point detection (Hassani, 2007).

The four different steps in SSA are: embedding, singular value decomposition, grouping and diagonal averaging.

Embedding can be regarded as a mapping that transfers a one dimensional time series $Y_T = (y_1,...,y_T)$ into multidimensional series $X_1, ..., X_K$ with vectors $X_i = (y_1,...,y_{i+L-1}) \in R^L$ where $K = T - L + 1$ (Hassani, 2007). The only parameter in this part is window length (L) which should be sufficiently large (Hassani, 2009). The result of this step is trajectory matrix $\mathbf{X} = [X_1, . . . , X_K] = (xij)_{i,j=1}^{L,K}$(Hassani, 2007).

The *SVD* step, makes the Singular value decomposition of the trajectory matrix and represents it as a sum of rank-one -bi-orthogonal elementary matrixes (Hassani, 2007). This step calculates the eigenvalues of XX^T and denote by $\lambda_1, ..., \lambda_L$and then for each eigenvalue (λ_i) calculate the right and left singular vectors (U_i, V_i) of X (Hassani, 2007). The collection of $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i^{th}

Formatted: Font: Italic

eigentriple of X. (Awichi, 2013). Then the SVD of trajectory matrix can be written as:

$$\mathbf{X} = \mathbf{X_1} + \cdot \cdot \cdot + \mathbf{X_D}$$

Where $\mathbf{X}_i = \sqrt{\lambda}\ U_i\ V_i^T$ (Hassani, 2007).

The *grouping* step corresponds to splitting the elementary matrices X_I into several smaller groups and summing the matrixes within each group (Hassani, 2007). If we consider $I = \{i_1, . . . , i_p\}$ as a group of indices $i_1, \ldots, i_p$, then the matrix X_I corresponding to the group I , can be defined as $\mathbf{X}_I = \mathbf{X}_{I1} + \cdot \cdot \cdot + \mathbf{X}_{IP}$ (Hassani, 2007). The split of the set of indices $J = 1, . . . , d$ into the disjoint subsets $I_1, \ldots \ldots, I_m$ corresponds to the representation:

$$\mathbf{X} = \mathbf{X}_{I1} + \cdot \cdot \cdot + \mathbf{X}_{IM}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping (Hassani, 2007). In this process, selected eigentriples are regarded as the signal while the rest of them are regarded as the noise (Briceno, 2013).

Diagonal averaging transfers each matrix *I* into a time series which is an additive component of the initial series Y_T (Hassani, 2007). If we consider z_{ij} as an element of matrix Z, then the *k-th* term of the resulting series is obtained by averaging z_{ij} overall *I, j* such that *i+j=k+1* (Hassani, 2007). This procedure is called diagonal averaging of Hankelization of matrix Z (Hassani, 2007). The result of Hankelization of matrix Z is the Hankel matrix *H*Z, which is the trajectory matrix corresponding to the series obtained as a result of the diagonal averaging. Hankel matrix *HZ* defines the series by relating the values in diagonals to the values in series. By applying Hankelization procedure after grouping section we obtain another expansion:

$$\mathbf{X} = \widetilde{\boldsymbol{X}}_{I1} + \ldots + \widetilde{\boldsymbol{X}}_{Im},$$

where $\widetilde{\boldsymbol{X}}_{I1}$= *HZ*. This is equivalent to the decomposition of the initial series Y_T= $(y_{T1}, \ldots, y_T)$ into a sum of m series:

$$y_t = \sum_{k=1}^{m} \tilde{y}_t^{(k)}$$

where $\tilde{Y}_T^{(K)} = (\tilde{y}_1^{(k)}, \ldots\ldots, \tilde{y}_T^{(k)})$ correspond to the matrix X_{I_K} (Hassani, 2007).

The R software (Bell Laboratories, United States) was used to produce the SSA results. The projected SSA forecasts using SSA were made in the short to medium term covering the projected years in five year gaps of 2015, 2020, 2025 and 2030.

Results

Historical Analysis

The RP is illustrated in figure 1 a-f.

Figure 1 a-f here

It can be seen that all six records have all improved progressively since their inception and ~~are~~ demonstrate a non-linear ~~in nature~~curve. In some cases there are examples of record plateau or stagnation. This has occurred several times in the 25 mile record but not typically in the other RP examples. There are also some occasions whereby a record has been broken by a visually significant amount and is indicated with a steep gradient of the graph trace. This can be seen with the 10 mile distance in 1981 and 2014, the 25 mile distance circa 1990, the 50 mile distance circa 1970 and the 100 mile distance circa 1980.

In the case of the 25 mile, 50 mile and 24hr (and to a lesser extent in the 100 mile and 12hr), there ~~is a time~~ are periods of time during the records progression ~~period~~ whereby the frequency of successful records has been high. However, this frequency then ~~but then this period has~~ lengthened over time. This effect is specifically illustrated and compared in figure 2.

Figure 2 here

It can be seen that the greatest volume of records in five of the six distances were typically obtained between 1950 and 1969. The exception to this is the 10 mile distance which was only established relatively recently. Despite this different inception date, its RPD during its existence is not dissimilar to the five other distances.

It could be argued that RPD could be skewed based upon the number and diversity of riders setting them. To consider this, the RD is shown in figure 3.

Figure 3 here

By considering both figures 2 and figure 3, not only was the greatest frequency of new records typically achieved with the 25 mile, 50 mile, 100 mile, 12 hour, and

24hr records between 1950-1969 but also that rider diversity was also high during this period. This means that both strength and the depth of British time trial riding was highest during that period and has also reduced over time.

Finally, the RRPI is illustrated in figures 4a-f.

Figure 4a-f here

As per figure 2, in figure 4, all six distances experience a period of high frequency record success followed by a gradual decline in this effect. Interestingly, the 25, 50 and 100 mile (as well as the 24 hour) have all seen larger percentage increases in the records when they have occurred after 1970. This effect also occurs to a lesser extent with the 12 hour records. The 10 mile record does not reflect this trend but the time frame is much narrower than all the other records, i.e. being only 30-40 years in existence (compared to the 70-80 of the others). However, its most recent change in 2014 was the largest it has seen since 1981.

SSA Forecast Analysis

The w-correlation between signal and noise was used to evaluate the accuracy of SSA reconstruction stage (see, for example, Hassani et al. 2009). The results indicate a good separability (w-correlation < 0.01).

After applying SSA technique to the time series, the trend line (first Eigenvalue) interpreted almost 100% of all information of the time series in all cases (outputs of the SSA Caterpillar software). By analysing the other components (Eigenvalues) of the time series in the grouping stage, they have been considered as the noise. As we believe that our model has identified all of the effective elements in the time series, we are not expecting to see any harmonic components or trends in the residual diagrams. Figures 5-10 illustrate the residual diagrams for the six record types.

Figure 5-10 here

As it can be seen in figures 5-10, there is not a distinctive trend or any kind of oscillation. The values appear randomised which do not follow any pattern. That means that the model has identified all of the effective elements in the time series.

The projected forecasts over a 15 year period using the SSA technique are shown in table 1.

Table 1 here

In the case of the duration fixed records, the SSA forecasts maintain its historical ‘stepped’ characteristics. As a result, the 10 mile and both the 12 and 24 hour based records do not always see progression every five years.

In addition, the amount of time the distance records will improve by from their current level to those projected in 2030 increase disproportionately with increasing race duration. The 10 mile record will improve by 39 seconds, the 25 3:58, the 50 7:47 and the 100 18:48. The reason that the forecast does not increase at the proportional increase in race distance could be sociologically/-environmentally based, scientifically grounded or a combination thereof. For example, the participation and popularity of one distance over another may be different - therefore the statistical likelihood of a new record being broken may change at different race lengths. Secondly, it is evident that the average speed of a rider will be less as the duration of an event lengthens. It is obvious that fatigue would be a factor as event duration lengthens. In addition, it is known that the power required to overcome air resistance is proportional to the bicycle speed cubed (Kyle & Bassett, 2003). As a result, the level of air resistance will reduce in a non-linear fashion meaning that the magnitude of time increase from one record to another will not be linear or directly proportional and would ultimately be greater as the race increases in duration.

Discussion

Historical Analysis

All six records have seen a progressive improvement and displayed a ~~are~~ non-linear decaying trend~~in nature~~. This non-linearity has been demonstrated before in cycling, pole vault and 100m sprinting (Haake, 2009), rowing, cross country skiing and speed skating (Desgorces et al., 2008) and swimming and weightlifting (Berthelot et al., 2008). In some cases, a steep gradient change or spike in the graphs trace has been suggested as being attributed to a sociological, environmental or technological change (Haake, 2009). The most notable change in British time trial based technology history has been proposed as the introduction of aerobars in 1991 (Donovan, 1991). The 25 mile record does illustrate a period of frequent record progression around this period. In addition, the rider diversity also increased during the same period. This could be attributed to the innovation of the aerobar but conversely, this effect was not witnessed in the 10 mile, 50 mile, 100 mile or 24hr records. This suggests the unique nature of predominantly British based time trialling (such as being conducted on an open road with the impact of passing traffic or being held throughout the year with different weather conditions) could shroud such technological impacts. These which have been shown to be clearly distinctive in other forms of time trial based cycling (Haake, 2009). Whilst this doesn't mean that technological change isn't of value per se', it does mean that other sociological or environmental improvements (such as course topography design and the time of year an event is held) is as critical as the impact of new equipment developments. This will also mean that time trial data between professional events and those within British based time-trialling are likely not directly comparable. A further limitation of this study is that the dynamic influence of sociological and environmental factors that have influenced performance in the past cannot be known with certainty whether they will do so in the future or do so to the same magnitude. This could include factors such as population changes or traffic volume increases. As a result, some degree of uncertainty will exist when extrapolating historical trends to predict future records.

Five records have had their highest level of record frequency and RD during the 1950-1969 time period. This could therefore be described as a 'golden age' of the sport. At a smaller level, there was a similar effect in early 1990. Conversely, whilst record frequency generally has reduced since the 1950-1969 period, the rate of performance improvement after this has often been far higher. It is not clear why this would occur since with a decaying trend since established over a longer period of time, the progression of a sport can be less frequent (Nevill et al., 2007) and in some cases reaching its asymptotic limits (Nevill & Whyte, 2005). It is possible that records have become far harder to obtain,. However, but this coupled with a reduction in RD may suggest that elite level performers (, possibly those above a typical amateur level physiologically), have the abilities to overcome the typical trend. For example, Chris Boardman has held the 25 mile record and later went on to hold the World Hour Record three times (Kyle & Bassett, 2003). Likewise, Graeme Obree has held the 10 and 50 mile records and later went on to hold the World Hour Record twice (Kyle & Bassett, 2003). More recently, Michael Hutchinson has held the 10, 25, 50 and 100 mile records and was twice 4th at the Commonwealth Games. It should be noted though that 'professional' riders were not always allowed to contribute to such records. As such, a limitation of this study is that in some cases, a full time rider may have been deterred from competing due to preventative legislation. There is no evidence present to determine the impact of this. Either wayAs such, these types ofprofessional riders may still be able to break the decaying trend in the future as they are not representative of a typical British record holder demographic. A cross generational comparison of cyclists has been undertaken (Bassett et al., 1999) but only when aerodynamic factors are assumed and that the rider has performed within the standardised environment of a closed cycling velodrome. As such, the characteristics of the record holders in comparison to physiologically talented amateurs could never be known.

SSA Forecast Analysis

The SSA technique successfully provided forecasted projections for (each record specifically). The success of the SSA technique in modeling and providing aa definitive sound forecast in this instance is mainly attributable to its ability of differentiating between the signal and noise in the time series. As noted earlier, the *w*-correlation statistic showed this was indeed the case with reported values of less than 0.01 indicating a good separation between noise and signal. The feasibility of the SSA forecasts could typically be verified using basic cycling power equations such as those defined by Wilson (2004) or using Martin et al. (2008) seminal cycling power mathematical model. However, in this case, the unique racing conditions - such as favourable course topography (high speed descents not retraced or favourable wind and shielding) and notably the impact of aerodynamic drag of passing traffic are not factors not accounted for in these models. As such, these models would overestimate the rider power actually required to achieve the predicted speeds. Instead, the use of variable or dynamic power-based pacing strategies such as those proposed by Boswell (2012) may be a preferred alternative in such scenarios.

Conclusion

All six male British time trial record progressions have seen an ongoing margin of improvement over time. The progression of all six records remains non-linear in nature. Five of these records saw their highest level of record change during the 1950-1969 time period. Whilst record frequency generally has reduced since this period, the magnitude of performance improvement after this timeframe when a record has been successfully broken has generally increased across several of the record types. Contrary to many other sports, the impact of the introduction of significant technological change was not as distinctive large in the records progression as was expected. This suggests that the unique nature of British time trialling being held in open conditions may be more crucial than the impact of technological change. Whilst this doesn't mean that technological change isn't of value, it does mean that other sociological and environmental effects (such as course design) and or when the event is held will be critical if future record progression is to continue. The novel application of the singular spectrum analysis to provide forecasts for such future progression provided a high level of fit to the raw data in the short to medium term.

References

Awichi, R., & Muller, W.~~,~~ (2013). Improving SSA prediction ~~By~~by inverse distance weighting. *Statistical Journal.*~~, *11*~~, *11*, 105-119.

Formatted: Font: Italic

Bassett D., Kyle, C., Passfield, L., Broker, J., & Burke, E. (1999). Comparing cycling world hour records, 1967-1996: modeling with empirical data. *Medicine and Science in Sports and Exercise*, *31*, 1665-1676.

Berthelot, G., Thibault, V., Tafflet, M., Escolano, S., El Helou, N., Jouven, X., & Toussaint, J. F. (2008). The citius end: world records progression announces the completion of a brief ultra-physiological quest. *PLoS One*, *3*, e1552.

Boswell, G. (2012). Power variation strategies for cycling time trials: A differential equation model. *Journal of Sports Sciences*, *30*, 651-659.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

CTT, 2014. *CTT 2014 Handbook*.

Briceno, H., Rocco, C., & Zio, E.~~,~~ (2013). Singular spectrum analysis for forecasting of electric load demand. *Chemical Engineering Transactions*, *13*, 919-924.

Desgorces, F., Berthelot, G., El Helou, N., Thibault, V., Guillaume, M., Tafflet, M., & Toussaint, J. (2008). From Oxford to Hawaii ecophysiological barriers limit human progression in ten sport monuments. *PLoS One*, *3*, e3653.

Donovan, D. (1991). Tri-bars get onboard. *Cycling Weekly*, Dec 8th, 5165, 3-4.

Haake, S. (2009). The impact of technology on sporting performance in Olympic sports. *Journal of Sports Sciences*, *27*, 1421-1431.

Hassani, H., (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science, 5*, 239-257.

Hassani, H., & Zhigljavsky, A. (2008). Singular spectrum analysis: methodology and application to economics data. *Journal of Systems Science & Complexity, 22,* 372-394.

Hassani, H., Soofi, S., & Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society Series A, 176*, 743-760.

Hassani, H., Mahmoudvand, H., & Zokaei, M. (2011). Seperability and window length in singular spectrum analysis. *Comptes Rendus Mathematique*, 349, 987-990.

Jeukendrup, A., Craig, N., & Hawley, J. (2000). The bioenergetics of world class cycling. *Journal of Science and Medicine in Sport*, *3*, 414-433.

Kyle, C., & Bassett, D. (2003). In: Burke, E. *High Tech Cycling*, US: Human Kinetics, 175-196.

Lippi, G., Banfi, G., Favaloro, E. J., Rittweger, J., & Maffulli, N. (2008). Updates on improvement of human athletic performance: focus on world records in athletics. *British Medical Bulletin*, *87*, 7-15.

Lukes, R., Chin, S., & Haake, S. (2005). The understanding and development of cycling aerodynamics. *Sports Engineering*, *8*, 59-74.

Martin, J., Gardner, A., Barras, M., & Martin, D. (2006). Modeling sprint cycling using field-derived parameters and forward integration. *Medicine and Science in Sports and Exercise*, *38,* 592.

Nevill, A., Whyte, G., Holder, R., & Peyrebrune, M. (2007). Are there limits to swimming world records?. *International Journal of Sports Medicine*,*28*, 1012-1017.

Nevill, A., & Whyte, G. (2005). Are there limits to running world records?. *Medicine and Science in Sports and Exercise*, *37*, 1785-1788.

Padilla, S., Mujika, I., Angulo, F., & Goiriena, J. (2000). Scientific approach to the 1-h cycling world record: a case study. *Journal of Applied Physiology*, *89*, 1522-1527.

Ransdell, L., Vener, J., & Huberty, J. (2009). Masters athletes: an analysis of running, swimming and cycling performance by age and gender. *Journal of Exercise Science & Fitness*, *7*, S61-S73.

Sanei, S., Ghodsi, M., & Hassani, H. (2010). An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *Medical Engineering and Physics, 33*, 399-411.

Schumacher, Y., Mroz, R., Mueller, P., Schmid, A., & Ruecker, G. (2006). Success in elite cycling: A prospective and retrospective analysis of race results. *Journal of Sports Sciences*, *24*, 1149-1156.

Schumacher, Y. O., & Mueller, P. (2002). The 4000-m team pursuit cycling world record: theoretical and practical aspects. *Medicine and Science in Sports and Exercise*, *34*, 1029.

Silva, E., & Rajapaksa, C. (2014). Evaluating the effectiveness of parametric and nonparametric energy consumption forecasts for a developing country. *International Journal of Energy and Statistics, 2,* 89-101.

Vincenzo, C., & Sandra, S. (2001). Scaling laws and forecasting in athletic world records. *Journal of Sports Sciences, 19,* 477-484.

Whitfield, P. 2013. *Time, speed and truth: The history of time-trialling 1890-2010*.

Wilson, D., 2004. *Bicycling Science 3rd Edition*, The MIT Press.

Zamparo, P., Minetti, A., & Di Prampero, P. (2002). Mechanical efficiency of cycling with a new developed pedal–crank. *Journal of Biomechanics*, *35*, 1387-1398.