

The growing problem of comparing elite sport performances: The Olympic speed skating case

BERTUS TALSMA¹, GERARD SIERKSMA² , MARCEL TURKENSTEEN³

¹ORTEC Sports, The Netherlands

²University of Groningen, The Netherlands

³Aarhus University, Denmark

ABSTRACT

The increased performance densities in the top of elite sports, sometimes mutual contest results are within the error margins of the measuring systems, has caused major problems in comparing performances and deciding on winners. In case of Dutch speed skating, the pool of highly competitive athletes is large, and, since there is a limit on the number of Olympic participants per country, the Olympic selection procedure is obviously a precarious affair. Because more than 100 years of data is available, we are able to study in this respect the well-known Gould hypothesis: When the sport matures, the variation in performance, especially at the top, decreases, and extreme events, where one athlete outperforms all his competitors, become rarer. Since Gould's hypothesis only holds with 'unchanged rules of the game', several data corrections, for example on the introduction of the klap skate, are needed. A major goal of this paper is to analyze the possible role of Gould's hypothesis concerning the growing performance densities for top speed skating. **Key words:** SPORT STATISTICS, OLYMPIC GAMES, PERFORMANCE ANALYSIS, GOULD'S HYPOTHESIS, SPEED SKATING.

Cite this article as:

Talsma, B., Sierksma, G., & Turkensteen, M. (2017). The growing problem of comparing elite sport performances: The Olympic speed skating case. *Journal of Human Sport and Exercise*, 12(3proc), S892-S907. doi:<https://doi.org/10.14198/jhse.2017.12.Proc3.13>

 **Corresponding author.** University of Groningen, The Netherlands.

E-mail: g.sierksma@rug.nl

11th World Congress of Performance Analysis of Sport, 16-18 November 2016. International Society of Performance Analysis of Sport. Alicante, Spain.

JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

© Faculty of Education. University of Alicante

doi:10.14198/jhse.2017.12.Proc3.13

1. INTRODUCTION

Comparing performances of athletes is a tedious aspect of all sports. Especially when athletes are compared for selection purposes and the used results are obtained under different circumstances, making a fair comparison may be very difficult. Errors in measuring and judgment of results may have disproportionate consequences.

Even in a direct comparison it may be difficult to determine the winner in a contest. For example, was the Olympic gold medal of Michael Phelps in 2004 correct or was the difference with the silver medal winner within the error margins of the measurement systems? In elite soccer there is an increasing number of matches that end up with not more than one goal, so that referee mistakes very often determine the outcome of the match. The difficulties grow when one compares performances between athletes or teams that do not compete directly, for example, when speed skaters race in different pairs.

A popular topic is the comparison of athletes over time. Of course, since technology, circumstances, and competition level have changed over time, the question of how athletes from the past would thrive in the current environment is not simple to answer directly. The results of such a comparison can be used, among others, to rank baseball players from different eras [1], or to determine potential future world records in speed skating [10].

In this paper we develop a 'fair' way of measuring and comparing speed skaters over time. Obviously, it is unfair to base a comparison on absolute performance results. A possible solution is to use relative times, such as the difference with the best time during a race. However, as we will see below, relative performances change over time as well. One of the causes is the following phenomenon.

It has been observed in many sports that the level of competition has increased over the years but the variation in performance at the top has decreased. Moreover, extreme events, i.e., performances where athletes are much better than their rivals, become more and more rare. This phenomenon has been identified for the first time in baseball in [4], and is known as Gould's hypothesis; see [1], [3], [11], and [12].

Stephan Jay Gould (1941-2002) was a rewarded American evolutionary biologist, who studied, among others, the development of complex biological systems. He was also an enthusiast baseball fan. One of his scientific claims was that, when complex systems improve over time and when the best performers play by the same rules during this process, the performances of the participants equilibrate and the variation of the top performances decreases. For the case of sports, we can reformulate these statements as follows.

There is a constant improvement of the level of competition due to just practicing, called the maturation process. More and more the limits of what is humanly possible are reached. This leads to a certain leveling of performances at the top, and extreme events, where some athletes are much better than their rivals, become rare: the differences in performances between elite athletes, and between top teams become smaller and smaller over time.

Gould's paper [4] discusses the disappearance of 0.400 baseball hitters, i.e., of baseball players that are able to hit an average of over 40% of the balls during one season. Such hitting averages were commonplace in the twenties and thirties of the twentieth century but have not occurred since 1941. Commentators used the absence of such events to argue that the performance level of the baseball players has gone down. Gould came up with the opposite explanation, namely that the performance level has increased. The disappearance

of 0.400 hitters is in his view caused by an improvement in the performance of the pitchers and of the field players along with that of the batters. The hitting averages of all players have been more or less constant, but because the average performances of both hitters and pitchers has increased over time, it has become harder for the top hitters to achieve averages of 0.400 or higher. Similar results are found in basketball [3]. But how does this phenomenon play a part in speed skating?

The usual explanation is that, due to limitations of the human body, there is an absolute boundary to sports performances, which is approached by the performances of the very best athletes; see [2], [3], and [5]. In [4] the average performance level among top athletes at a certain time instance is called the general performance level (GPL) at that time.

An important aspect of Gould's hypothesis is the assumption that the rules of the game remain unchanged over time. We will discuss and analyze the changes in speed skating that are related to this question, among others changes in, what we call, the five T's (Section 3), such as the increased number of top speed skaters, the introduction of indoor skating rinks, and the use of equipment innovations such as the klap skate.

For each year (starting in around 1900) the GPL will be calculated by correcting the data in such a way that Gould's assumption on 'unchanged game rules' is satisfied. Box plots for the corrected data show a decreasing trend in faster race results and an increase of the performance density (decreasing variance). We will apply regression analysis on the medians of the box plots, using the well-known three biological growths models: Exponential, Weibull, and Gompertz. It turns out that in fact the performance variation decreases over time, but that, interestingly enough, the occurrence of world records, which can be considered as extreme events, did not decrease over time; see Section 4.3.

In Section 4, we perform statistical analyses on publicly available data bases with skating performances such as to correct the performances for the above mentioned factors and assertions.

The basic rules in speed skating are given in Section 2, and the major developments and technological innovations are discussed in Section 3. In Section 4, we analyze the GPL for speed skating and show the general trend towards better performances using the development of world records. We also perform some computational experiments, and measure the influence of the introduction of new technologies and new tournaments. Finally, Section 6 contains the conclusions and directions for further research. This study should be considered as a first attempt to explain the development of the GPL in sports; a more thorough study is needed to determine the actual significance of the observed phenomena; see also Talsma [12].

2. SPEED SKATING: BASICS AND RULES

The history of skating goes back a long way. In 1893, the International Skating Union (ISU) was established and races were organized on 400 meter oval rinks. Formal measurements of race results started in that year. The rink is divided into two separate lanes, an inside and an outside lane. The two competing skaters are in the different lanes during the race; they change lanes after each lap.

After skating the full distance, the times of both skaters are recorded at the finish line. The finish time determines the ranking of the skater during that tournament.

The ISU supports international races on the distances 500, 1000, 1500 meters (called short distances), and 3000, 5000, 10000 meters (called long distances). We denote the distances by 500m, 1000m, and so on.

These distances are skated in various combinations at allround tournaments, sprint tournaments, and single distance tournaments. Currently, there are seven international tournaments organized by the ISU, namely the World Championships Allround, the European Championships Allround, speed skating at the Olympic Winter Games, the World Championships Sprint, the World Cups, and the World Championships Single Distances. We use these tournaments for our data base, because all top skaters are able to participate there (if qualified). National and regional championships are therefore excluded from our analysis. In Table 2, we have summarized a number of statistics concerning the considered speed skating tournaments.

For championships where the distances are skated individually, the skater with the fastest race time on that distance is the winner, with the exception of the 500m. In Hjort (1994) [6], it is shown that male skaters starting in the inner lane at the 500m have an advantage of 0.05 seconds, on average. Based on this finding, the ISU decided to skate the 500m twice by each skater during Olympic Games, starting once in the inner and once in the outer lane. The same rules are used for the 500m of the World Championships Single Distances. However, since 2017 this decision has been reversed; see Kamst et al., [8] and [9].

Table 1: Speed skating tournaments

Tournament name	For men, since	For women, since	Type	Abbreviation
World Championship Allround	1892	1936	Allround	WChA
European Championship Allround	1892	1936	Allround	EChA
World Championship Sprint	1970	1970	Sprint	WChS
Olympic Winter Games	1924	1960	Distance	OG
World Cups	1979	1979	Distance	WC
World Championships Single Distance	1996	1996	Distance	WChSD

Alternatively, there are championships where the weighted average of each skater's times over multiple distances is taken. In 'allround' championships, a time-based point system, called the samalog system, is used. The times of the four distances are converted to '500m times' and then added. For instance, the 1500m time is divided by 3, the 5000m by 10, and the 10000m by 20. The sum of these four '500m times' yields a point total and the skater with the lowest point total is the winner. The Sprint Championships use the same method. In our analysis in Section 4, we use the samalog system to normalize the skated times at the mentioned championships to a fixed distance (the 500 m), which we call the r-time. It is denoted by and defined as, for each measured time s and each race over d meters:

$$r(s, d) = (s \times 500)/d. \quad (1)$$

(1) See http://nos.nl/artikel/353729-wereldrecords-terug-tot_aardse-properties.html (in Dutch)

3. DEVELOPMENTS IN SPEED SKATING

When the American speed skater Eric Heiden completed the 10000m in 14 minutes and 28 seconds at the Winter Olympics in 1980, the Dutch commentator shouted that this performance would never be improved. However, the current world record on this distance is just above 12 minutes and 36 seconds, almost two minutes faster. In this section we take a closer look at the developments that speed skating has undergone since 1893.

Figure 1 shows the progress of the world records on the classical distances, measured in r-times, for both male and female skaters. The steep declines in the graphs are mainly due to technological innovations and, although to a lesser extent, to rink improvements; see Kuper, Sterken [10].

Other factors, that have influenced the improvements in world record, are the introduction of tight skating suits, new training methods, and the more frequent occurrence and perhaps earlier recognition of exceptional athletic skills. Brownlie et al. [2] focus in particular on the influence of tight skin skating suits. In this paper we also take into account the increase in the number of tournaments, namely: the number of annual tournaments has increased from two in the initial years to 17 now.

The following five T's concern the performance improvement factors in our analysis:

- Talent pool: the increasing number of participants;
- Training: the improvement of training methods;
- Tracks: the modernization of the rink (for example, the introduction of indoor skating rinks);
- Tools: the innovation of the equipment of skaters (such as tight fit suits and klap skates);
- Tournaments: the changes in the tournament calendar (such as an increase of the number of tournaments, leading to specialization).

Clearly, the GPL does not improve as fast as the world records, because GPLs concern 'averages'. On the other hand: we may expect that the GPL-values converge to the world records. The increased size of the talent pool, the better training methods, as well as the increase in specializations, have resulted in improved performances. However, faster skating rinks, improved equipment, and changes in the tournament set-up may be considered as rule changes. So one has to be careful when Gould's hypothesis is tested. In the following section, we take a closer look at these factors.

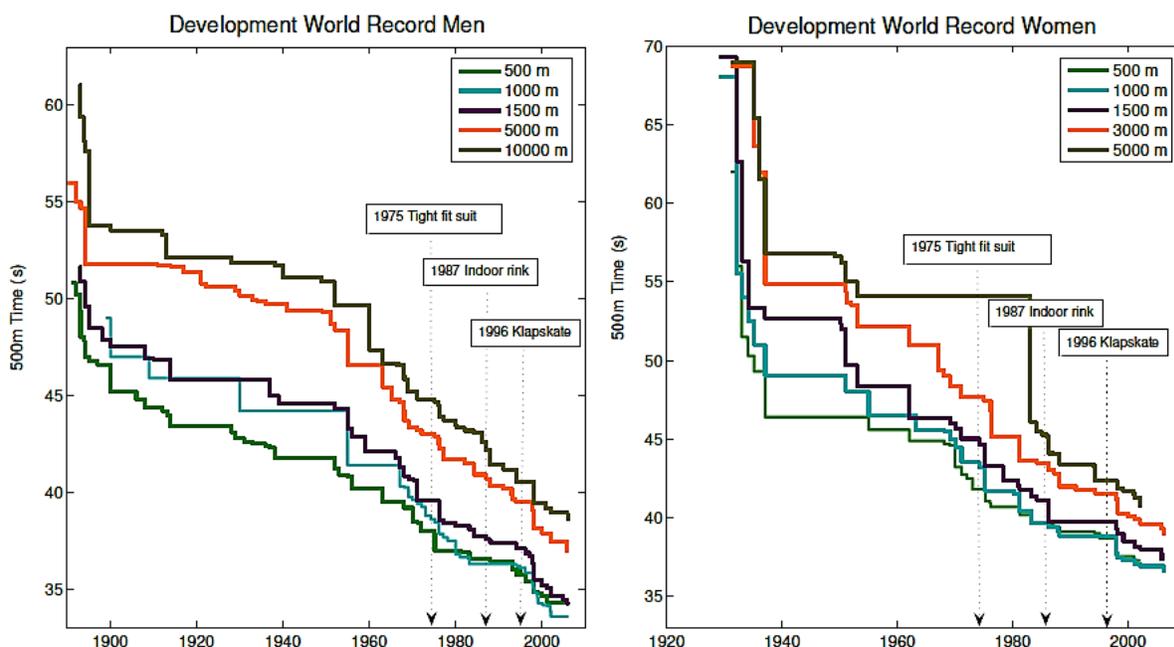


Figure 1: Progress of the world records

3.1. Skater specific factors

An important aspect of Gould's hypothesis is the phenomenon that an increase in active participants results in a larger pool of highly talented athletes. This may be a result of better training methods, and more specialization due to changes in tournament calendar. These factors are called skater specific improvements. Also the fact that talented skaters are now recognized much earlier and accompanied to the top much better than before (see [10]) has resulted in an increase in the size of the talent pool. Figure 2 shows the increase in the number of active skaters world- wide on international top tournaments. The more skaters become competitive, the larger the talent pool.

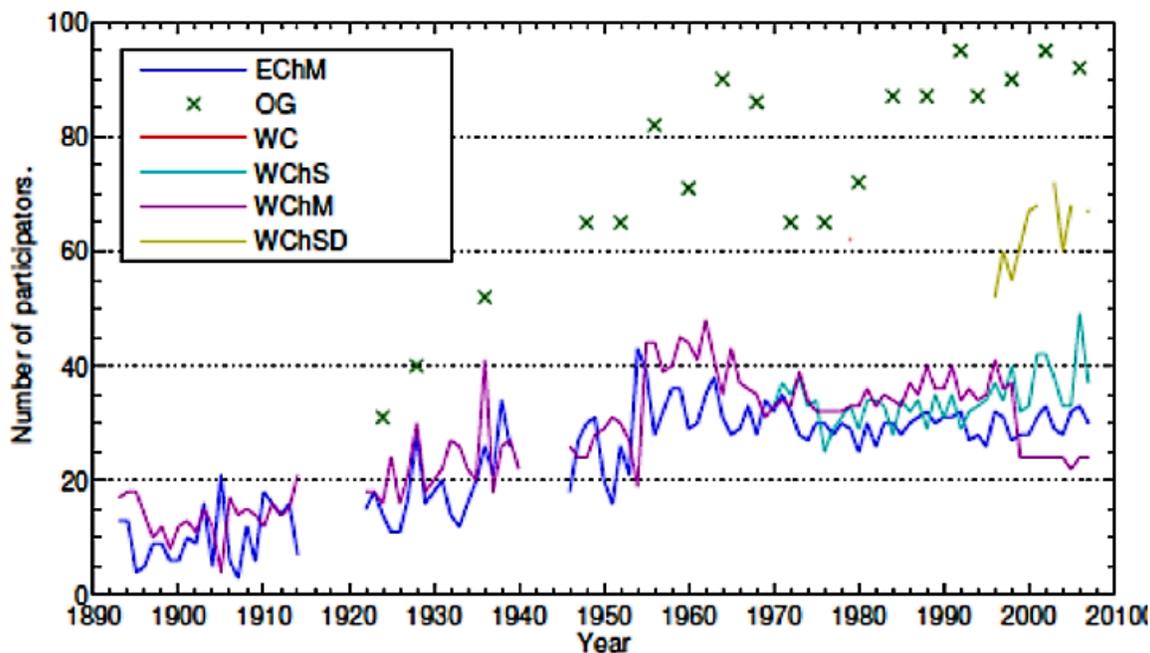


Figure 2: Total number of participants per tournament, men's races

The increase in the size of the talent pool in combination with changes in the tournament calendar has led to more specialization. In the early years, a few allround championships were organized, while the current tournament calendar contains many disciplines, mostly specialist (i.e., over individual distances). The importance of 'allround' results has decreased, and it has become more and more difficult to excel at all skating disciplines. Speed skaters nowadays tend to specialize on at most two distances. As skaters focus on their best distances, performances at individual distances improve.

Figure 3 shows the development over the years of the number of participants per distance. Notable exceptions are the 10000m for men and the 5000m for women (not depicted), due to the fact that these long distances are skated only once or twice during a World Cup season.

A difficulty in the comparison between results from the 'old' allround tournaments and the more recent specialized tournaments is the fact that sprinters had to participate in the 5000m on allround tournaments, because special single sprint contests did not exist. This may have had a negative influence on the GPL of that time, because sprinters did not perform very well on the 5000m, resulting in lower averages.

New training methods have clearly boosted the performance of the skaters. This also holds for better nutrition, and for the appearance of sponsored skating teams around 1995. We may argue that this better preparation of skaters boosts the average performance of top skaters, bringing it closer to the boundary.

All these factors are considered as skater specific and have increased the GPL, while bringing the GPL closer to Gould's boundary.

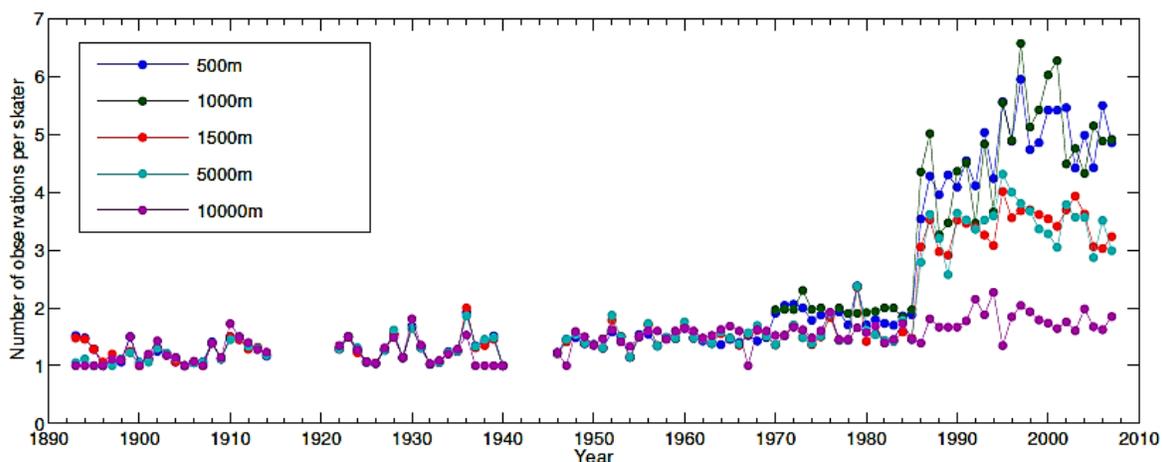


Figure 3: Observations per skater per distance

3.2. Skater specific factors

A number of developments have changed the 'circumstances' of the sport and should be considered as 'changing the rules of the game'. The factors that concern these developments are the T's (from the beginning of this section): 'Tracks' and 'Tools'.

Several innovations in skating equipment (tools) are introduced over time over the years. The innovations range from the introduction of iron blades in the early days and experiments on the thickness of them to the introduction of the so-called tight suits by the Swiss speed skater Franz Krienbuhl in 1976. The last major innovation (in 1996) was the klap skate with the blade fixed at one hinged point on the shoe, introduced by the Dutch scientist Gerrit Jan van Ingen Schenau. This new skate led to significant improvements in almost all personal and world records; see Houdijk et al. [7]. The effect of the klap skate on world records is discussed in Kuper and Sterken [10], where the world records in 2006 are predicted based on the gains in the first years after the introduction. All predictions turned out to be too pessimistic, meaning that the effect of the klap skate appears to have been underestimated. This was caused by the fact that the first users of the klap skate grew up with conventional skates and needed to unlearn the traditional way of skating to learn the new technique.

Also the improvement in skating rinks (tracks) has had a positive effect on skating performances, but not uniformly. The last five years almost all world records are skated in Calgary or in Salt Lake City on high altitude indoor rinks. These rink innovations make it impossible to use uncorrected race times. For example, an r-time of 38 seconds on an outdoor rink is much better than the same time on the fast rink of Calgary.

The introduction of new equipment has had an almost immediate effect on race results. Improvements on the rinks have a less visible effect, because they are introduced not at one moment of time, but rink by rink over the years.

In [4], Gould mentions the fact that the introduction of technological innovations produce 'sudden blips' in the performance graphs. The introduction of such innovations may then be the starting point of a new convergence process if the innovation leads to a shift in the 'boundary', although the average performance adapts only slowly.

4. COMPUTATION OF PERFORMANCE LEVELS

In this section we use a statistical approach for comparing performances over time by eliminating the 'rule changing' effect of the various developments discussed in Section 3.2.

In Gould [4], the set of all performances in one year is represented as a statistical distribution in which the best performances lie at the right tail of this distribution. The distribution is bounded on the right hand side. While the GPL increases, more and more athletes come closer to this boundary, and the difference between top and average athletes becomes smaller.

Figure 4 shows the development of r-times from the start of the measurements until recently and shows a steady improvement in these results, while also the variance decreases.

4.1. Statistical analysis of the General Performance Level

We have observed that the GPL depends on a number of factors called the 5 T's. We will now apply statistical corrections in order to make the GPL only dependent on the skater specific factors. As said, the GPL is a measure of the quality of the top skaters in a certain year.

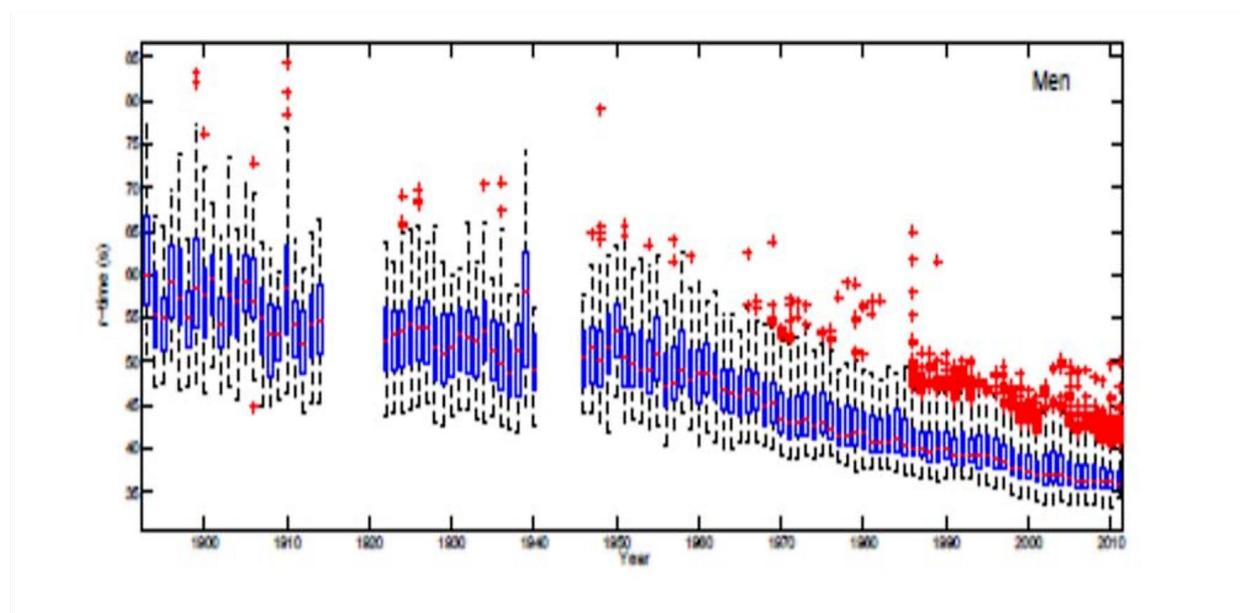


Figure 4: Box plot of r-times

We will investigate how far the GPL has come closer to the best possible performances, and to what extent this has led to a decrease in variation in performances. Measures of variance are given in Section 4.2. Finally, we determine whether or not we can observe a decrease in the occurrence of extremely good performances over time as well as the occurrence of extremely small differences between top performers, as 'predicted' by Gould's hypothesis. Although we focus on men's skating results, similar analyses can be performed for women's skating.

The data set contains all results on the mentioned championships from 1893 to 2006; see <http://www.isu.org/>. In order to make the data comparable across disciplines and eras, we perform the following standard operations on the data. First, we compare the various distances by computing the average race times using the r-times. Secondly, since the number of participants on tournaments has increased, we only use the results of the 24 best skaters from each contest, as explained below.

For each race results ranking we will use time differences with a fixed reference time of that race. For example, if the reference time is 38 seconds and a skater has completed the 500m in 40 seconds, we use the difference of 2 in our analysis. The difference with a reference time is called the AV („absolute difference“) performance of the skater. We use as the reference time the average of the five best r-times of a race, which we call the AV 5-performance. (We deviate from this in the early times of the sport). So, for each race, the average r-time of the first five times is subtracted from all r-times of that race. We did not take the fastest time as the reference time, because in that case, all winners are rated equal and possible superiority of a winner is not recognized anymore.

Table 2: Notation of tournaments

SD	Single Distance Championships
OV	Allround Championships
SP	Sprint Championships
OG	Olympic Winter Games

In order to justify the '5' in AV 5, we have applied sensitivity analysis on this number. We refer for this analysis to Talsma [12]. Since we do not use other values for this set of averages we will delete the '5' from the symbols. So we write AV instead of AV 5, among others.

Each tournament type uses a different set of distances. We use the symbols from Table 2. Define, $K = \{SD, OV, SP, OG\}$, $D = \{500m, 1000m, 1500m, 3000m, 5000m, 10000m\}$, the years in our data base $T = \{1893, \dots, 2006\}$ excluding the periods 1914 - 1918 and 1940 - 1945. In this paper ‘year’ means the year that the skater’s season ends; usually the season runs from September in year $t - 1$ through March in year t . We only use skaters in our data set that have finished their race; this set of skaters is denoted by S . We denote the data base containing all relevant skating times by DB . An observation in DB is denoted by the indices $i(\in S)$, $k(\in K)$, $t(\in T)$, and $d(\in D)$. For each observation $(i, t, k, d) \in DB$, let I_{tkd}^5 be the set of skaters who finished within the first 5 positions on distance d at tournament k in year t . The average of the r -times of the first 5 skaters, $\bar{r}T_{tkd}$, is denoted by and defined as:

$$\bar{r}T_{tkd} = \frac{1}{5} \sum_{i \in I_{tkd}^5} rT_{itkd}.$$

and

$$AV_{itkd} = rT_{itkd} - \bar{r}T_{tkd}.$$

Hence, AV_{itkd} is the difference in time (seconds) between the average time of the first 5 skaters and the r -time of skater i on distance d of tournament k in year t . In the following section, we will show that, based on a number of assumptions, these AV -values measure the relative individual performance of skaters and that they are not influenced by ‘rule-changing’ innovations.

In Figure 5, we have depicted the AV -values of all top 24 ranked male skaters over the years. We may observe a decreasing trend in these values. This may be an indication that these AV -values converge to a best performance. Observe that the variation among top skaters also decreases over time. The interested reader is asked to see Talsma [12] for sensitivity analysis on the number 24.

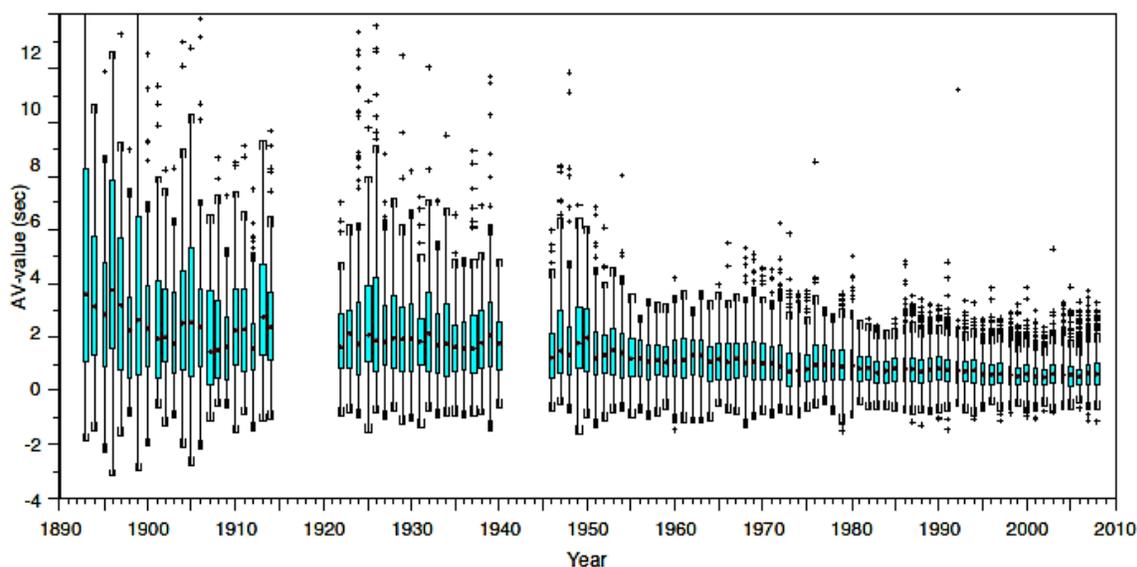


Figure 5: Box plot of all AV -values, 24 best times per tournament

4.2. Measuring the variation in performances

In order to analyze the decreasing trend of the GPL, we look for an appropriate statistical gauge. We have decided to use the average and the median of the AV - values of the first 24 skaters at each tournament; this set of skaters is indicated by:

I^24_{tkd} . Hence, for each $t \in T$ and $d \in D$, we have that:

$$MAV_{td} = \text{median} \bigcup_{k \in K} \{AV_{itkd} \mid i \in I^24_{tkd}, \}$$

For each discipline $L \in \{SD_1, \dots, SD_5, SP, OV\}$, and $t \in T$, define

$$MAV_{tL} = \begin{cases} MAV_{td} & \text{for } L \in SD \\ \frac{1}{4} \sum_{d \in L} MAV_{td} & \text{for } L = OV \\ \frac{1}{2} \sum_{d \in L} MAV_{td} & \text{for } L = SP. \end{cases}$$

Note that for $L = OV$ and $L = SP$, the average of the median values are taken over the number of distances in the tournament, four in OV and two in SP (namely, twice the 500m and twice the 1000m). In order to determine the development of the expected MAV -values of each discipline over time, we perform regression on the computed MAV -values with time t as the independent variable. Our model should take care of the property that the skating times cannot decrease indefinitely (e.g., to negative values), but should converge to some lower asymptote. A similar problem is faced in Kuper and Sterken [10], where the purpose is to find a number of suitable candidate functions for fitting the development of world records over time. The authors use statistical regression based on biological growth models, namely the Exponential, the Weibull, and the Gompertz curves; see Table 3. These functions satisfy our requirements on converging to a lower asymptote and are therefore our considered candidates.

Table 3: Candidate Functions

Function	Type
$Y = \beta_1 + \beta_2 \exp^{\beta_3 t}$	Exponential
$Y = \beta_1 + \beta_2 \exp^{\beta_3 t_4^\beta}$	Weibull
$Y = \beta_1 + \beta_2 \exp^{-\exp \beta_3 (t - \beta_4)}$	Gompertz

The three functions are given in Table 3, where β_1 is the horizontal asymptote, β_3 is the growth parameter, while the other parameters are extra fitting parameters. In our case, the variance may continue to decrease over time, but cannot surpass the boundary level $\beta_1 (> 0)$. The estimated $\tilde{\beta}$ -parameters can be found by minimizing the weighted residuals. The residuals are weighted by the number of residuals per year in order to obtain the best fit through medians of years with a large number of observations.

For each discipline L and $t \in T$, we model the MAV -values as

$$MAV_{tL} = Y(\beta_L) + \epsilon_{tL} \quad (2)$$

where β_L is the vector of chosen function parameters and depending on the type of regression function (so for discipline L , we have to estimate the parameters $\beta_{L1}, \beta_{L2}, \beta_{L3}$), while ϵ_{tL} represents the residuals, assumed to be normally distributed. The function $Y(\beta_L)$ describes, for each L , the dependency on the choice of β_L .

We illustrate the estimation of MAV -values for the discipline 1500m in Figure 6. One reason to use the 1500m is that this distance has always been a key distance on the allround tournaments and is still the most spectacular one. The values of $MAV_{t,1500m}$ are plotted as dots. The three regression curves (Exponential, Weibull and Gompertz) are shown as well. Note that these functions coincide more or less for the 1500m discipline. More extensive computations show that the Gompertz curve has the best fit for each discipline.

We use the regression result $Y(\beta_L)$ with the estimated parameter values β as our measure of the (adjusted) General Performance Level. For each $t \in T$ and $k \in K$, we denote the estimated GPL by $G\tilde{P}L_{tL}$ and compute it as

$$G\tilde{P}L_{tL}^{\alpha_t} = Y(\tilde{\beta}_L). \quad (3)$$

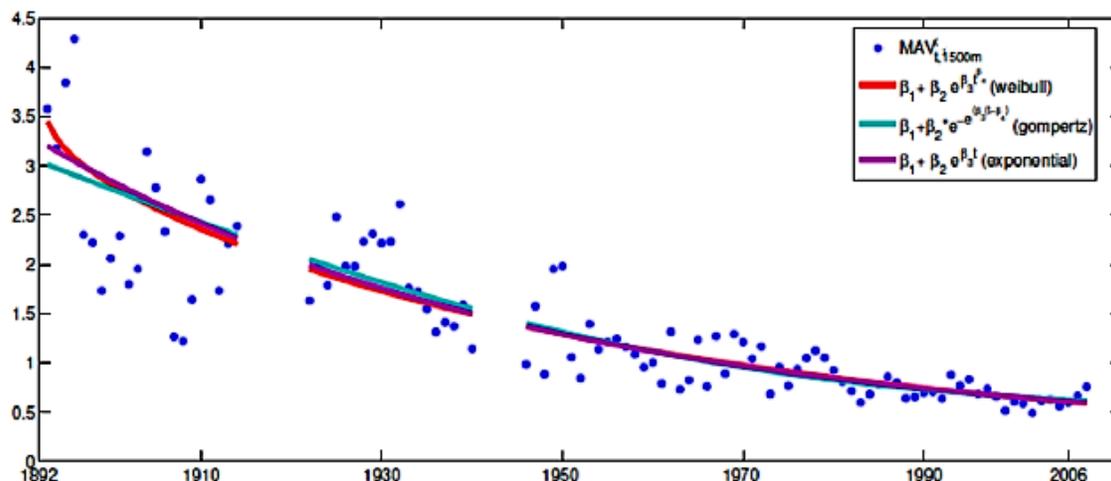


Figure 6: Estimation of the MAV-values of 1500m results

We may observe that the MAV-values are best modeled with a slowly decreasing function of the Gompertz type. We do not report the estimations for the other disciplines, including those for the women. The above analysis applies mutatis mutandis for the other disciplines.

4.3. Extreme events: world records

The reduction in variation and the improvement in average times, compared to the Gould boundary, should reduce the frequency of extreme events. As candidate extreme events, we may look at world records, as they are commonly seen as, at least, special events among skating enthusiasts. As said before, according to Gould's hypothesis, extreme events become rarer and rarer as the variation in top performances decreases. Figure 7 illustrates the number of world records per five year period. Actually, the number of world records increases, in spite of the fact that the variation in performance has decreased. A possible explanation is that innovations and technical improvements have led to better performance under equal physical performance. Moreover, this increase in world records may be explained from the increase in the number in contests and from the increasing competition: where there were previously a small number of skaters capable of skating world records, a larger number of skaters can do so now. A commonly overlooked aspect here concerns the increased accuracy of the measuring systems. Nowadays, world records with less than 0.001 seconds time differences are assigned. On the other hand, time specialists agree on the fact that such time differences are within the error margins of the measuring systems. It would be interesting to correct for these factors and then see what this means for the occurrence of new world records.

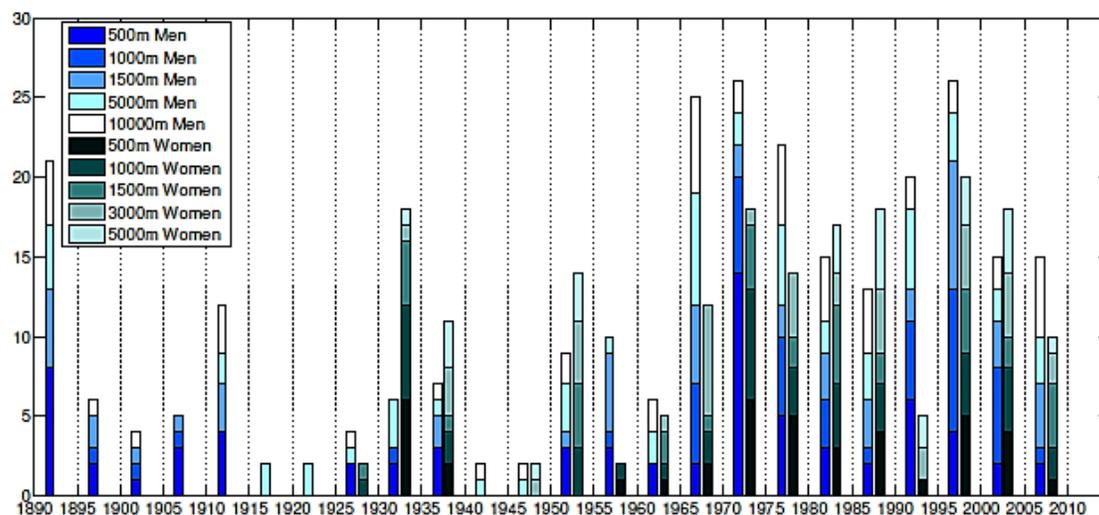


Figure 7: Number of world records per five year period.

5. SUMMARY

We have analyzed the development and the variation in performance of speed skating using a measure of the GPL that was insensitive to ‘changes of the rules of the game’, namely the technology (e.g., klap skates), tools (e.g., skating suits), and tracks (e.g., improvements in skating rinks). We used the average of the five best times at each tournament or discipline as a reference time, which was subtracted from the actual r-times to obtain a skater’s ‘normalized performance’, called the AV -value.

In our second step, we took into account the increase in the competition (the talent pool), and the fact that more skaters now have access to professional training methods. To that end, we measured for each discipline the average performance of a typical skater at a tournament as the median value of all (but at most the 24 best) skaters at that tournament, which was called the M AV performance. Based on Gould’s hypothesis, we would expect a decreasing trend in the M AV -values over time. Actually, it turned out that the so-called Gompertz model, that satisfies this trend, provides the best estimate of the M AV -values over time. The Gompertz curve is characterized by a decay where the predicted values slowly approach an asymptote (the value of which is estimated as well).

The results can also be used for comparing skaters across different eras (see Berry et al. [1]). The regression results are then used to construct a typical performance on a given discipline in a given year. If the AV -values of two skaters in, say 1921 and 1989, divided by the expected M AV -values, are the same, we may conclude that the performance of both skaters is similar. If we use sufficiently large sets of observations for each skater, we may rank skaters across different eras; see Talsma [12].

6. CONCLUSIONS AND FUTURE RESEARCH

This paper considers the Olympic sport speed skating and analyzes performances over time. In this respect, we also consider Gould’s hypothesis. This hypothesis states for the case of sports that, by just practicing, performances of the athletes improve, the so-called maturation process, while top performances tend to reach a certain ceiling. This leads in turn to increasing mutual performance densities caused by an

increasing number of athletes that have approached the ceiling. Besides this phenomenon, extreme performance events become more and more rare, mainly due to decreasing variances of top performances. A major aspect in Gould's hypothesis is the assumption of 'nonchanging rules of the game' during the maturation process.

In case of speed skating several 'rule changing' events have taken place through the years. These innovations certainly have increased the absolute speed of the skaters, but cannot be seen as athlete specific improvements. Therefore, in order to 'test' the Gould hypothesis, we needed to apply (statistical) corrections on the data used. Actually, we have changed the results of any race into relative times by subtracting from each race time the average of the first five times in that race. These values are called the AV -values. By applying statistical regression with the Gompertz model on the AV -values, we have calculated performance gauges with which Gould's hypothesis could be investigated.

This study should be seen as a start of more extensive research, among them the development of variance bounds of these gauges. Also other types of gauges should be analyzed in this respect (robustness analysis on the validity of the gauges). Finally, Gould's hypothesis should be tested in all its aspects for other sports as well, among them the track and field sports (see, Haake et al. [5]).

We conclude this paper with a number of practical consequences when the 'increasing density trend' continues.

- More and more ex aequo situations;
- Athletes are declared winners, where the advantage of the 'winner' with the 'runner up' is within the error margins of the measuring systems;
- Selection procedures for, e.g., Olympic Games, become precarious affairs. When the pool of potential medal winners is larger than the (IOC) participants quota, potential medal winners have to stay home;
- In many cases already, it takes too much time and spectators have to wait, before the contest jury has decided on the winner.

Although we cannot conclude that top elite sport is in a 'performance judgement' crises, a quick look at internet sites on 'close finishes' reveals that unfairnesses in judgements are already no exceptions.

REFERENCES

1. S.M. Berry, S.C. Reese, and P.D. Larkey. Bridging different eras in sports. *Journal of the American Statistical Society*, 94:661–687, 1999.
2. L.W. Brownlie, C.R. Kyle, E. Harber, R. MacDonald, and M.R. Shorten. Reducing the aerodynamic drag of sports apparel: development of the Nike Swift sprint running and SwiftSkin speed skating suits. *The Engineering of Sport*, 5(1):90–96, 2004.
3. S. Chatterjee and M.R. Yilmaz. The NBA as an evolving multivariate system. *The American Statistician*, 53:257–262, 1999.
4. S.J. Gould. *Full House: the Spread of Excellence from Plato to Darwin*, chapter 4, pages 77–132. Three Rivers Press, 1996.
5. S. Haake, D. James, and L. Foster. An improvement index to quantify the evolution of performance in field events. *Journal of Sports Sciences*, 33(3):255–267, 2015.

6. N.L. Hjort. Should the Olympic sprint skaters run the 500 meter twice? Technical report, MyScienceWork, 1994.
7. H. Houdijk. Scientific explanation for success of klapskate. www.eurekalert.org/pub_releases, 2001.
8. R. Kamst, G.H. Kuper, and G. Sierksma. The Olympic 500-m speed skating; the inner-outer lane difference. *Statistica Neerlandica*, 64(4):448–459, 2010.
9. R. Kamst, G.H. Kuper, G. Sierksma, and B.G. Talsma. Inner-outer lane advantage in Olympic 1000m speed skating. *Journal of Economics and Statistics*, 232(3):293–317, 2012.
10. G.H. Kuper and E. Sterken. Endurance in speed skating: the development of world records. *European Journal of Operational Research*, 148:293–301, 2003.
11. G. Sierksma. Introduction: Olympics, track & field. In J.J. Cochran, J. Bennett, and J. Albert, editors, *The Oxford Anthology of Statistics in Sport: Volume 1: 2000-2004*, pages 73–76. Oxford University Press, 2017.
12. B.G. Talsma. *Performance Analysis in Elite Sports*. PhD thesis, University of Groningen. SOM Research School, 2013.