Progress in Applied Mathematics
Vol. 1, No. 1, 2011, pp. 98-105
www.cscanada.org

ISSN 1925-251X [Print]
ISSN 1925-2528 [Online]
www.cscanada.net

Kinetic Energy Analysis for Soccer Players and Soccer Matches

Zengyuan YUE[1,2,*]
Holger Broich[3]
Florian Seifriz[1,2]
Joachim Mester[1,2]

Abstract: A simple and practical semi-empirical formula for the calculation of the *Mean Specific Kinetic Energy* (*MSKE*) of a soccer player over a match is developed based on the statistical analysis of the detailed records of the time courses of locations of all players for a single match. This formula is further supported by additional tests and some theoretical analysis. The formula is then applied to five matches. This formula can be used not only to compare different players, but also to compare different teams, to see how active and how energetic they are during a match. This formula can also be used to test whether there is a significant effect of a training program on the kinetic energy of players of a team.

Key Words: Soccer; Energy; Statistics; Training

1. INTRODUCTION

The *Specific Kinetic Energy (SKE)*, i.e. the kinetic energy per unit body mass given in the unit Joule/kg, is an important parameter for a soccer player for two reasons: First, it shows how active and how energetic one player is; second, it has to do with physiological parameters, e.g. oxygen uptake and heart rate etc. Thus, it would be desirable to find a simple way to calculate the *Mean Specific Kinetic Energy (MSKE)* over a period of time, particularly over the first and the second halves of a match or over the entire match for a given player. We would then be able to compare not only different players, but also different teams, to see how active and how energetic they are over the match. Some physiological parameters, e.g. oxygen uptake and heart rate, are interesting but can not be easily measured during the match. *MSKE* is expected to have positive correlation with energy expenditure and therefore positive correlations with those physiological parameters. Thus, at least for a given player, higher *MSKE* in a match would mean higher energy expenditure and higher oxygen uptake and heart rate.

Of course, if we know the entire time course of locations for a given player, we would be able to calculate the instantaneous *SKE*, which would simply be one half of the square of the instantaneous speed.

[1]Institute of Training Science and Sport Informatics, German Sport University Cologne. E-mail address: z.yue@dshs-koeln.de (Zengyuan YUE).
[2]The German Research Centre of Elite Sport.
[3]Bayer 04 Leverkusen Fussball GmbH.
[*]Corresponding author.
[†] Received 11 December 2010; accepted 5 January 2011.

However, it is usually difficult to get the detailed data for the entire time courses of locations for different players. Such detailed records can be made and kept by special companies which built the facility and the software for such detailed records. Even if a soccer team buys the *Summary Report* from such a company for a given match, the *Summary Report* would include only some overall parameters, e.g. the distance coverage of each player over the first half and over the second half of the match, instead of the detailed time courses of locations of all players during the match. Thus, it would be desirable to develop a simple and practical formula to calculate *MSKE* based only on some overall parameters, e.g. the distance coverage over the first half and over the second half of the match. The purpose of this paper is to develop such a desired formula and to apply this formula to some soccer matches.

The previous dynamic studies of soccer games are rather limited in the literature. They have been restricted to the statistics of a single aspect of the game, e.g. the goal distribution[1-3], or the distance coverage[4-6], or the possession and motion of the ball[7, 8], or team strength parameter[9], or the temporal behavioural pattern[10], or the portions of time for different movements of different groups of players[11], over ensemble of matches (cups and championships).

In our last papers on mathematical analysis of a soccer match (Yue et al.[12, 13]), we used the raw data of the detailed time courses of the 2D locations of all players and of the ball for a soccer match between two top-level German teams held in May 2007. Such detailed data were made available by the software Amisco Pro and the facilities Amisco System both produced by SUP (Sport Universal Process, Nice, France). Detailed statistical analysis for various aspects of the match was made in those papers. Various parameters, e.g. the *geometrical centres*, the *radii*, and the *expansion speeds* etc., were obtained as functions of time. The major ranges of different groups of players (forwards, defenders, midfielders) at different phases (attacking phase, defending phase, the entire first half) as well as the distance coverage of each player and its average over different groups of players at different phases were calculated. In the present paper, we will use the same set of raw data to develop a simple formula for *MSKE*, in which only some overall parameters would be involved, so that the formula could be used for the matches where only the overall parameters are known through the *Summary Report*.

2. METHOD AND RESULTS

The raw data for the above mentioned soccer match give the detailed time courses of the 2D locations of all players and of the ball during the first and the second halves of the match with a sampling frequency 10 Hz. Thus, we can calculate the velocity components and the specific kinetic energy for any moment, and get the *MSKE* for each player for the first half. On the other hand, we can also calculate the distance coverage of each player for the first half, denoted by D. We also know the time duration of the first half, denoted by T. Figure 1 is the scatter plot of the points (D^2/T^2, *MSKE*) for the 22 players who participated in the first half of the match. As we see, the 22 points show very good and very significant correlation between D^2/T^2 and *MSKE* ($r = 0.9948$, $p < 0.000001$), where r is the Pearson's correlation coefficient, and p is the significance parameter. Since *MSKE* would be 0 for $D = 0$, the best fitting straight line should pass through the origin (0, 0). This best fitting straight line can be obtained by some calculations, and the result is, accurate to the second digit after the decimal

$$MSKE = 0.75D^2 / T^2. \tag{1}$$

This straight line gives very good fit to the 22 points as shown in Figure 1. Note, generally speaking, the Pearson's correlation coefficient r has the range $-1 \leqslant r \leqslant 1$. Usually, if $|r|$ is larger than 0.9, the correlation is considered to be good enough for prediction[14]. Namely, one parameter could be predicted from the other by the best fit. The extremely high value $r = 0.9948$ in our case means that Equation (1) can be used as a good approximation for the estimate of *MSKE* from the distance coverage D and the time duration T for each player. Thus, we do not need the raw data for the detailed time courses, which are usually difficult to get, in order to calculate *MSKE*. This is the major advantage of Equation (1).

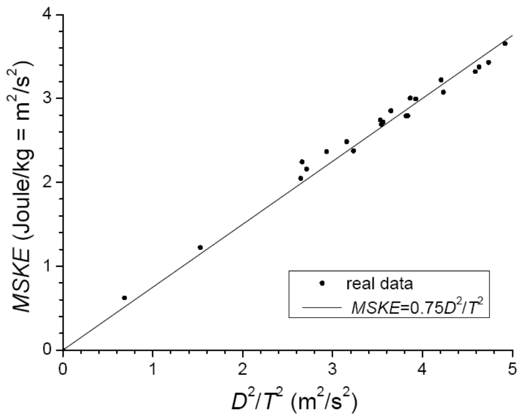


Figure 1: Scatter plot of the points (D^2/T^2, *MSKE*) for the 22 players in the first half of the match between two top-level German teams held in May 2007. The 22 points show very good and very significant correlation between D^2/T^2 and *MSKE* ($r = 0.9948$, $p < 0.000001$). The straight line is the linear fit given by Equation (1)

One question could be raised. Equation (1) is obtained only from the 22 players in the first half of a single match. Can it be applied to other players and other matches? This question has been partly answered by the extremely low value of p ($p < 0.000001$). Usually, $p < 0.05$ would be accepted to be significant because $p < 0.05$ would mean that we would have more than 95% confidence to claim that the correlation is true instead of just by chance. Now we have $p < 0.000001$, an extremely high significance. This means that we have extremely high confidence to claim that the very high correlation between *MSKE* and D^2/T^2 is a true correlation, instead of by the accidentally chosen 22 players and the accidentally chosen single match. In other words, if we had chosen a different match and different players, a similar very good correlation between *MSKE* and D^2/T^2 would still be expected.

As a further test of the general applicability of Equation (1), we consider the second half of the match, where 7 new players participated in. Figure 2 and Figure 3 show how the straight line given by Equation (1) fits the points (D^2/T^2, *MSKE*) over the second half of the match for the players who participated also in the first half and for the new players respectively. As we see, the fits in Figures 2 and 3 are both very good. The correlations between D^2/T^2 and *MSKE* in both cases are very high and extremely significant: For Figure 2, we have $r = 0.9929$, $p < 0.000001$, and for Figure 3, we have $r = 0.9972$, $p < 0.000001$.

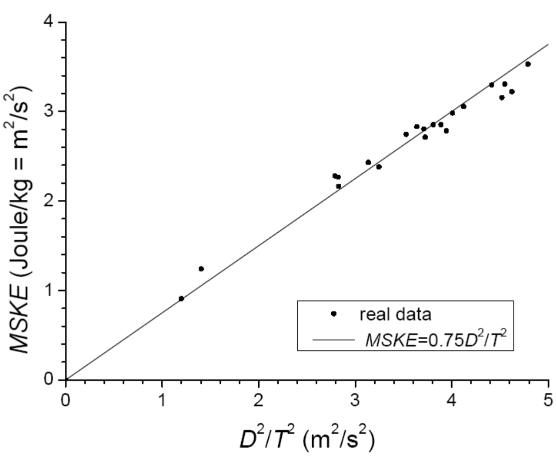


Figure 2: Scatter plot of the points (D^2/T^2, *MSKE*) for the players in the second half who have also participated in the first half of the match. The plot also shows very good and very significant correlation between D^2/T^2 and *MSKE* (r = 0.9929, $p < 0.000001$). The straight line is the linear fit given by Equation (1)

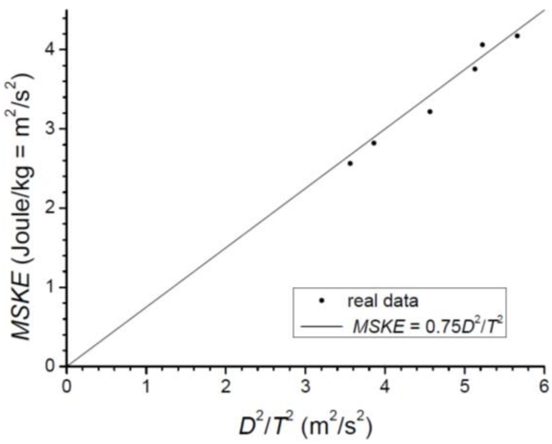


Figure 3: Scatter plot of the points (D^2/T^2, *MSKE*) for the new players in the second half who have not participated in the first half of the match. The plot also shows very good and very significant correlation between D^2/T^2 and *MSKE* (r = 0.9972, $p < 0.000001$). The straight line is the linear fit given by Equation (1)

As a further support to the general applicability of Equation (1), let us consider the speed of any player, denoted by *W*, which is the square root of the sum of the squares of the longitudinal and the lateral velocity components. For any given player, *W* is a time series because it changes from time to time. It is easy to show that, for any given time duration *T*, e.g. the first half or the second half of the match, we have

$$MSKE = \frac{1}{2}(\overline{W}^2 + \sigma^2). \tag{2}$$

In this equation, $\overline{W} = D/T$ is the mean speed over the time duration *T*, where *D* is the distance coverage of the player over the time duration *T*, and σ is the standard deviation of *W*. Thus, Equation (1) is equivalent to

$$\sigma^2 = \frac{1}{2}\overline{W}^2 = \frac{1}{2}D^2/T^2. \tag{3}$$

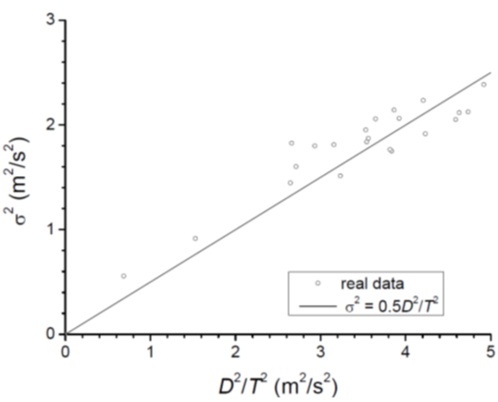


Figure 4: The correlation between σ^2 and $\overline{W}^2 = (D/T)^2$ for the 22 players in the first half of the match. Again, the correlation is very strong and extremely significant (r = 0.9459, p < 0.000001). The linear fit is given by Equation (3)

Actually, every player's speed varies from time to time, ranging from 0 to his individual maximal speed. Thus, qualitatively, it is easy to understand that the larger the mean speed is, the larger the standard deviation of the speed would be, as shown in Equation (3). Now, as a quantitative test, Figure 4 shows the correlation between σ^2 and D^2/T^2 for the 22 players in the first half of the match and the linear fit given by Equation (3). Again, the correlation is very strong and extremely significant (r = 0.9459, p < 0.000001).

3. APPLICATION OF EQUATION (1) TO OTHER MATCHES

Now we apply Equation (1) to the following five matches between Bayer Leverkusen team and other teams. These are

Match A: Bayer Leverkusen vs. Bayern München (Sep. 29, 2007);

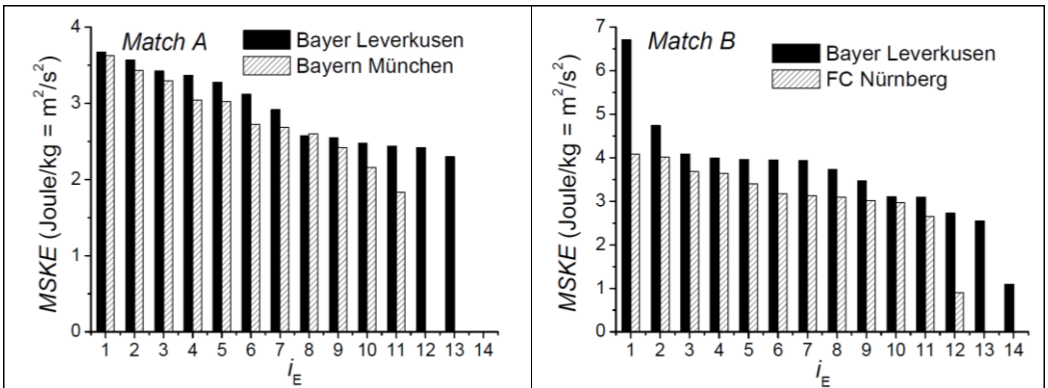
Match B: Bayer Leverkusen vs. FC Nürnberg (Mar. 16, 2008);

Match C: Bayer Leverkusen vs. Eintracht Frankfurt (Mar. 29, 2008);

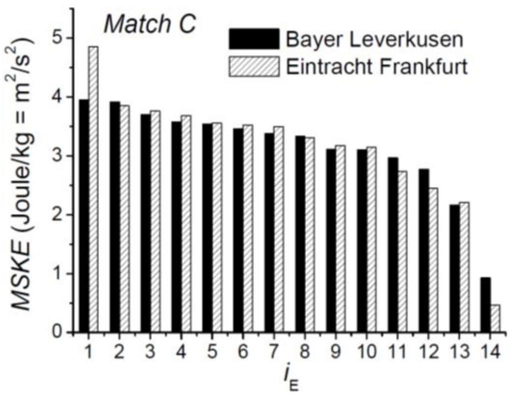
Match D: Bayer Leverkusen vs. VfB Stuttgart (Apr. 13, 2008);

Match E: Bayer Leverkusen vs. VfL Wolfsburg (Apr. 27, 2008).

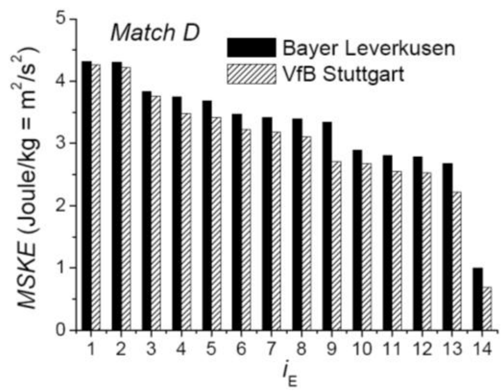
For these matches, Bayer Leverkusen team has bought the *Summary Reports* from the company which made the detailed records of the matches. However, as usually, the *Summary Reports* include only some overall parameters, e.g. distance coverage of each player in the first and the second halves of the corresponding match. The raw data of the detailed time courses of locations of all the players and of the ball during the entire matches were not sold. Nevertheless, we can now calculate *MSKE* for each player by Equation (1) only based on the distance coverage and the time duration of each player in the first and the second halves of each match without the need of the detailed raw data. In this way, we can see how active and how energetic each player is, so that we can compare different players from the energy point of view. We can also calculate the mean of *MSKE* over each team in each match, denoted by *MMSKE*, to see how active and how energetic each team as a whole is compared to its opponent team in the match. Part of the results is shown in Figure 5 and Figure 6.



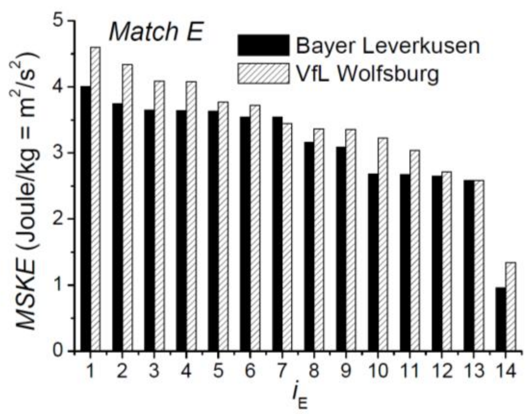
(a) (b)



(c)



(d)



(e)

Figure 5: The *Mean Specific Kinetic Energy* (*MSKE*) of each player, calculated by Equation (1), in (a) *Match A*, (b) *Match B*, (c) *Match C*, (d) *Match D*, and (e) *Match E* respectively, where $i_E = 1$ means the player who has the highest *MSKE* in the team, $i_E = 2$ means the player who has the second highest *MSKE* in the team, and so on. In the calculation of *MSKE*, the distance coverage and the time duration refer to the entire distance coverage and the time duration which the player participated in for the whole match, including both the first and the second halves

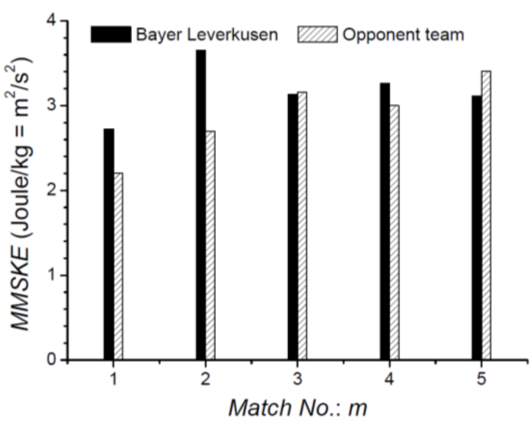


Figure 6: *MMSKE* comparison in the five matches, where m = 1 to 5 mean *Matches A* to *E* respectively, and *MMSKE* is the mean of *MSKE* over the team in the match

Figure 5 (a) to (e) give the individual *MSKE* of each player of each team in the *Matches A* to *E* respectively. In these figures, we did not give the true number of each player because the purpose of this paper is to present the simple and practical method for calculating *MSKE*, rather than making comments on the performances of different players. Thus, in these figures, $i_E = 1$ only means the player who has the highest *MSKE*, $i_E = 2$ means the player who has the second highest *MSKE*, and so on. Figure 6 shows the *MMSKE* comparison in the five matches, where m = 1 to 5 mean *Matches A* to *E* respectively.

4. DISCUSSIONS

Different groups of players (forwards, midfields, defenders, goalkeepers) have different tasks. Therefore, for example, it would not be fair to compare the performances of a forward and the goalkeeper by their *MSKE*. However, *MSKE* would indeed be an important parameter to compare the players within the same group because they have similar tasks.

Formula (1) can also be used to test whether there is a significant effect of a training program on *MSKE* of players.

5. SUMMARY

A simple and practical formula for calculating the *MSKE* of a player is developed, where only the overall parameters, i.e. the distance coverage and the time duration, are involved. Thus, we can easily calculate *MSKE* for each player and the mean of *MSKE* (*MMSKE*) over each team based only on the *Summary Report* of the match, which is relatively easy to get, rather than the detailed raw data for the time courses of the locations of the players and of the ball during the entire match, which are usually very difficult to get. By using this formula, we can easily compare different players and different teams to see how active and how energetic they are in their performances during the match, and we can also test whether a training program has significant effect on the kinetic energy of players or not.

REFERENCES

[1] Malacarne, L. C., & Mendes, R. S. (2000). Regularities in football goal distributions. *Physica A, 286*(2), 391-395.

[2] Onody, R. N., & de Castro, P. A. (2004). Complex network study of Brazilian soccer players. *Physical Review E, 70*(3), 037103: 1-4.

[3] Bittner, E., Nussbaumer, A., Janke, W., & Weigel, M. (2009). Football fever: Goal distributions and non-Gaussian statistics. *European Physical Journal B, 67*(3), 459-471.

[4] Bangsbo, J., Norregaard, L., & Thorso, F. (1991). Activity profile of competition soccer. *Canadian Journal of Sport Sciences – Revue Canadienne des Sciences du Sport, 16*(2), 110-116.

[5] Rienzi, E., Drust. B., Reilly. T, Carter, J. E. L., & Martin, A. (2000). Investigation of anthropometric and work-rate profiles of elite South American international soccer players. *Journal of Sports Medicine and Physical Fitness, 42*(2), 162-169.

[6] Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F. J., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine, 28*(3), 222-227.

[7] Lago, C., & Martin, R. (2007). Determinants of possession of the ball in soccer. *Journal of Sports Sciences*, *25*(9), 969-974.

[8] Mendes, R. S., Malacarne, L. C., & Anteneodo, C. (2007). Statistics of football dynamics. *European Physical Journal B, 57*(3), 357-363.

[9] Glickman, M. E., & Stern, H. S. (1998). A state-space model for National Football League scores. *Journal of the American Statistical Association*, *93*(441), 25-35.

[10] Borrie, A., Jonsson, G. K., & Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sports: an explanation and exemplar data. *Journal of Sports Sciences, 20*(10), 845-852.

[11] Ali, A., & Farrally, M. (1991). A computer-video aided time motion analysis technique for match analysis. *Journal of Sports Medicine and Physical Fitness*, *31*(1), 82-88.

[12] Yue, Z., Broich, H., Seifriz, F., & Mester, J. (2008). Mathematical analysis of a soccer game. Part I: Individual and collective behaviours. *Studies in Applied Mathematics,121*(3), 223-243.

[13] Yue, Z., Broich, H., Seifriz, F., & Mester, J. (2008). Mathematical analysis of a soccer game. Part II: Energy, spectral and correlation analyses. *Studies in Applied Mathematics*, *121*(3), 245-261.

[14] Vincent, W. J. (2005). *Statistics in Kinesiology* (3rd edition). New Zealand: Human Kinetics.