International Journal of Computer Science in Sport

Volume 12/2013/Edition 2

TABLE OF CONTENTS

Editorial

Arnold Baca

Department of Biomechanics, Kinesiology and Applied Computer Science,

ZSU, University of Vienna

Dear readers:

Welcome to the winter issue 2013 of the **International Journal of Computer Science in Sport (IJCSS).**

Four research papers and one project report have been included within this issue.

Allan Z. Maymin, Philip Z. Maymin, and **Eugene Shen** introduce a new framework to measure and to analyse the on-court synergy of basketball teams, which allows to optimize the composition of rosters.

Ralf Schneider, Oleksandr Kalentev, Tatyana Ivanovska, and **Stefan Kemnitz** studied the influence of an increased ball size as well as an increased net height on table tennis player and the game by computer simulations.

Nicole Ruch, Johanna Hänggi, Stephanie Zurbuchen, and **Urs Mäder** investigated the validity of the ActiSmile, a physical activity feedback device.

Moritz Vetterli and **Nicole Ruch** present a comparison of the energy expenditure estimates of activity-specific regressions, random forest, and regression trees using acceleration data from children.

The project report by **Robert H. Schmicker** demonstrates the application of SaTScan, a software for the the spatial, temporal, and space-time scan statistics, to evaluate the spatial distribution of corner kick goals.

If you have any questions, comments, suggestions, or points of criticism, please send them to me.

Best wishes for 2014!

Arnold Baca, Editor in Chief
University of Vienna, arnold.baca@univie.ac.at

NBA Chemistry: Positive and Negative Synergies in Basketball

Allan Z. Maymin[1], Philip Z. Maymin[2] & Eugene Shen[1]

[1]*AllianceBernstein*

[2]*NYU-Polytechnic Institute*

Abstract

We introduce a novel Skills Plus Minus ("SPM") framework to measure on-court chemistry in professional basketball. First, we evaluate each player's offense and defense in the SPM framework for three basic skill categories: scoring, rebounding, and ball-handling. Next, we simulate games using the skill ratings of the ten players on the court. Finally, we calculate the synergies of each NBA team by comparing their 5-player lineup's effectiveness to the "sum-of-the-parts." We find that these synergies can be large and meaningful. Because skills have different synergies with other skills, our framework predicts that a player's value depends on the other nine players on the court. Therefore, the desirability of a free agent depends on the current roster. Indeed, our framework generates mutually beneficial trades between teams. Other ratings systems cannot generate ex-ante mutually beneficial trades since one player is always rated above another. We find more than two hundred mutually beneficial trades between NBA teams, situations where the skills of the traded players fit better on their trading partner's team. We also find that differences in synergies between teams explain as much as six wins and that teams are no more likely to exhibit positive chemistry than negative chemistry.

KEYWORDS: NBA, SYNERGY, CHEMISTRY, SKILLS PLUS-MINUS

Introduction

> "My model for business is The Beatles. They were four guys who kept each other's negative tendencies in check. And the total was greater than the sum of the parts. Great things in business are not done by one person; they are done by a team of people."
>
> – Steve Jobs

Basketball, one of the world's most popular and widely viewed sports, is a timed game played by two teams of five players on a rectangular court1. While the exact playing regulations vary across different governing bodies, we focus on the National Basketball Association ("NBA"), which is widely considered the premier men's professional basketball league in the world. Teams alternate possession of the basketball and attempt to score points by shooting a ball through a hoop 18 inches in diameter and 10 feet high mounted to a backboard at each end of the floor. The team with the most points at the end of the game wins the game. In the NBA,

[1] Background information on the game of basketball draws from http://en.wikipedia.org/wiki/Basketball.

teams have 24 seconds to attempt a field goal. A successful field goal attempt is worth two points for the shooting team, or three points if the shooting player is behind the three-point line. A free throw is awarded to an offensive player if he is fouled while shooting the ball. A successful free throw attempt is worth one point. Each possession ends with either a field goal attempt, free throw attempt, or a turnover (if a player loses possession to the opposing team). Turnovers can occur when the ball is stolen (a "steal") or if the player steps out of bounds or commits a violation ("non-steal turnover"). A missed field goal attempt or free throw attempt results in a rebounding opportunity, where the teams fight to gain possession of the ball. Each possession ends with a finite number of possible outcomes, making the simulation of a game feasible.

The rules of basketball do not specify any positions whatsoever, and there are no special positions such as goalie. Over time, positions have evolved, where shorter and quicker players play "guard", a position that requires more ballhandling, passing and outside shooting. Meanwhile, taller and stronger players typically play "forward" or "center", operate closer to the basket, and grab more rebounds. Traditionally, teams play with two guards, two forwards, and one center, but it is possible to play with five guards or five centers, if a team so desires.

A box score summarizes the statistics of a game, detailing player contributions such as minutes played, field goal attempts, successful field goals, free throw attempts, successful free throws, rebounds, assists, steals, blocks and turnovers. Assists are awarded when a player passes the ball to a teammate who then scores a field goal. A block occurs when a defensive player legally deflects a field goal atatempt by an offensive player. In general, guards accumulate more assists, while centers block more shots. There have been many attempts to rate individual basketball players using box score statistics. Examples include Wins Produced or Win Shares (see Oliver 2004). These ratings systems generally agree with expert opinions on the best players in the league. For example, during the 2012-2013 season, both Wins Produced and Win Shares suggested that LeBron James and Kevin Durant were the two best players in the NBA. These two players also finished first and second in Most Valuable Player ("MVP") voting for that season.

While these box score ratings can measure an individual's contributions, they do not necessarily explain how players interact on the court. For example, it is possible that the five best players in the NBA are all centers. In this case, a team with five centers may not be the optimal lineup, since there would be no one to bring the ball up the court or guard the quicker opposing guards. Therefore to determine the optimal lineup, we would want to measure the "synergies" among players, and predict which players play well with each other. Our paper attempts to address this issue by introducing a Skills Plus Minus ("SPM") framework that decomposes a player's contributions into three skills: scoring, rebounding and ball-handling.

In sports, synergies are not often applied to individual athletes. Bollinger and Hotchkiss (2003) in evaluating baseball define team synergy as firm-specific productivity such as the signals and strategies unique to the team. MacDonald and Reynolds (1994) explicitly avoid attention to "synergy" or "chemistry" and focus only on the value of each baseball player on his own. Indeed, they hypothesize "a reasonably efficient market in player talent and a consequently quasi-efficient assignment of players among teams and within team line-ups." Idson and Kahane (2000) begin the path of testing this hypothesis by separating out the effects of individual and team productivity on salary determination in the National Hockey League, and indeed find that team attributes not only directly affect individual pay, but can also diminish certain individual productivity effects. Their results in fact hint at synergies: they find complementarity across some productive attributes but not others and they hypothesize that "larger, more significantly positive interactions might follow if certain positions are paired."

They leave this open as a fruitful subject of future research.

Here we are able to actually test this hypothesis by using a large dataset of repeated interactions combined with our Skills Plus Minus model and framework to decompose the players into their constituent skill groups and evaluate the synergies resulting from various combinations of those skill groups. We find that the allocation of players within teams is not efficient, and that there are hundreds of trades that would have benefitted both trading teams because of the effects on team chemistry.

An example helps frame our argument. With the third pick in the 2005 NBA draft, the Utah Jazz selected Deron Williams, a 6'3" point guard who played collegiately at Illinois. Using the very next pick, the New Orleans Hornets drafted Chris Paul, a 6'0" point guard from Wake Forest. Since the moment they entered the league, the careers of Williams and Paul have often been compared. Countless debates and discussions sparked about who is the better point guard. There are arguments for both sides.

The box score statistics seem to favor Paul. His career statistics (18.7 points per game, 4.6 rebounds, 9.9 assists, 2.4 steals, 0.571 true shooting percentage ("TS%")) are better than Williams across the board (17.2 points, 3.2 rebounds, 9.2 assists, 1.1 steals, 0.560 TS%). Paul has played in more All-Star games (4 vs. 2) and appeared on more All-NBA teams (3 vs. 2).

Meanwhile, supporters of Williams point to his better regular season record (0.590 winning percentage vs. 0.555 for Paul), relative playoff success (20 playoff wins vs. 10), head-to-head record against Paul, size, strength, and durability. They argue that Williams is a stronger one-on-one defender who does not gamble for steals.

At the end of the 2009-2010 season, if Utah had traded Deron Williams for Chris Paul, would they have been better off? If New Orleans had traded Chris Paul for Deron Williams, would they have been better off? Using the framework introduced in this paper, we can answer these questions: surprisingly, the answer is YES to both. A Williams-for-Paul swap would have made both teams better off and is an example of a mutually beneficial trade. Such a trade should not have been possible if team composition were efficient; at the very least, such a trade should have been consummated, but it never was.

This paper introduces a novel Skills Plus Minus framework to measure on-court chemistry in basketball. This SPM framework builds upon the Advanced Plus Minus ("APM") framework first introduced by Rosenbaum (2004). While APM evaluates each player based on the points scored while they are in the game, SPM evaluates each player based on the offensive and defensive components of three basic categories of skills: scoring, rebounding and ball-handling. For example, a player's "steal" ratings (part of the ball-handling category) are determined by how many steals occur while he is in the game. Like APM, SPM considers the other nine players on the court. A benefit of the APM and SPM framework is the ability to capture skills that are not found in traditional box score measures, such as off-the-ball defense, boxing out, and setting picks. Also, in contrast to other ratings such as Wins Produced, APM and SPM do not make position and team adjustments to the player ratings.

We use the SPM framework to simulate games using the skill ratings of the ten players on the court. These simulations incorporate how each play starts: out-of-bounds, steal, defensive rebound or offensive rebound. We find these starting conditions materially affect the outcome of the possession. The simulations are then used to measure the effectiveness of individual players and 5-player lineups.

We investigate which basketball skills have synergies with each other. Traditionally, team chemistry has been difficult to measure. Berri and Jewell (2004) use roster stability as a proxy for chemistry. While they acknowledge the "potential impact of disruptive players," (which we would call negative synergies in our framework) they note that "identifying and quantifying the impact of such players appears problematic." Our framework solves this problem.

Another method to measure chemistry compares the "lineup APM" versus the sum of the constituent single player APM's. The problem with that approach is that there are too many possible five-player lineup combinations. The APM's of the five-player lineups have small sample problems since the minutes played of any given five-player lineup can be small. Our innovation is that we are able to predict synergies while avoiding this problem.

We calculate the synergies of each NBA team by comparing their 5-player lineup's effectiveness to the "sum-of-the-parts." These synergies can be large and meaningful. Because skills have different synergies with other skills, a player's value depends on the other nine players on the court. Therefore the desirability of a free agent depends on the players currently on the roster.

Finally, our framework is able to generate mutually beneficial trades. Other ratings systems cannot generate mutually beneficial trades, since one player is always rated above another, c.f. Kubatko, Oliver, Pelton, and Rosenbaum (2007) for a review of most of them, or Berri (1999) or Berri (2008) for more detail on Wins Produced. Berri and Brook (1999) investigate whether trades are ex-post mutually beneficial and argue that trades can be ex-ante mutually beneficial if the ex-post distribution of minutes is known and different. In contrast, our framework generates ex-ante mutually beneficial trades without a change in the distribution of minutes played. Using our framework, we find many mutually beneficial trades, when the skills of the traded players fit better on their trading partner's team. One such mutually beneficial trade is Chris Paul for Deron Williams.

Methods

Description of the Data

While our primary innovation is a theoretical framework to model on-court chemistry, we use data to illustrate. Berri and Schmidt (2010) criticize APM because the player ratings are not stable from year-to-year. They favor ratings that use box score statistics (e.g. Wins Produced), because the ratings are more predictable from year-to-year. We acknowledge Berri and Schmidt's criticism and therefore use data from four NBA seasons (2006-2007 through 2009-2010) to achieve better estimates for player skills. While Fearnhead and Taylor (2011) allow their APM ratings to be time-varying, we estimate one rating for all four years.

The data we use is from basketballgeek.com, maintained by Ryan J. Parker, and represents a processed version of the play-by-play information from the NBA and ESPN. The data includes the names of all players on the court at each time, the location of the shots taken, result of possession, and more. The data set includes 4,718 games and 987,343 plays.

Tables 1-4 display summary statistics from our data set.

Table 1. Possession Start Variables

Possession Start	Count	Percent
Defensive Rebound	256,589	26.0%
Offensive Rebound	104,903	10.6%
Steal	59,329	6.0%
Out of Bounds	566,522	57.4%
Total	987,343	100.0%

Table 2. Possession Outcomes

Possession Outcomes	Count	Percent
Steal	68,460	6.9%
Non-steal turnover	66,912	6.8%
Missed FT – 2 pts	5,953	0.6%
Missed FT – 1 pt	15,068	1.5%
Missed FT – 0 pts	7,161	0.7%
Made FT – 3 pts	16,650	1.7%
Made FT – 2 pts	59,746	6.1%
Made FT – 1 pt	19,908	2.0%
Missed 3 FG	108,651	11.0%
Made 3 FG	60,652	6.2%
Missed 2FG	298,416	30.3%
Made 2 FG	257,524	26.1%
Total	985,101	100.0%

Table 3. Offensive Rebounds

Type	OReb	Missed Shots	OReb%
Field Goal	127,489	407,154	31.3%
Free Throw	3,749	28,218	13.3%

Table 4. Players Involved in the Most Plays in Our Data Set.

	Name	Plays	Name	Plays	Name	Plays	Name	Plays	Name	Plays	Name	Plays
1	Andre Iguodala	53,798	Samuel Dalembert	39,505	Ronnie Brewer	29,750	Brook Lopez	22,937	Mike James	17,691	Kris Humphries	12,515
2	Kobe Bryant	50,783	LaMarcus Aldridge	39,388	Antonio McDyess	29,712	Jason Maxiell	22,841	Michael Beasley	17,603	Josh Powell	12,290
3	Dwight Howard	49,297	Zach Randolph	39,098	Zydrunas Ilgauskas	29,589	Aaron Brooks	22,826	Eddy Curry	17,458	Leon Powe	12,131
4	LeBron James	49,254	Carlos Boozer	38,806	Luke Ridnour	29,559	Carlos Delfino	22,801	Jamaal Tinsley	17,428	Renaldo Balkman	12,010
5	Antawn Jamison	48,399	Allen Iverson	38,764	T.J. Ford	29,412	Jason Williams	22,734	C.J. Miles	17,399	Tony Battie	11,894
6	Jason Kidd	47,746	Mike Miller	38,742	Luis Scola	29,352	Jordan Farmar	22,712	Marko Jaric	17,225	Tyreke Evans	11,793
7	Andre Miller	47,515	Mike Bibby	38,565	Peja Stojakovic	29,175	Linas Kleiza	22,674	Josh Childress	17,156	Jamaal Magloire	11,681
8	Rudy Gay	47,238	Kevin Durant	38,436	DeShawn Stevenson	29,124	Daniel Gibson	22,402	Wally Szczerbiak	17,126	Ersan Ilyasova	11,617
9	Joe Johnson	47,209	Kirk Hinrich	38,322	Andres Nocioni	28,806	Dahntay Jones	22,334	Fabricio Oberto	17,051	Brent Barry	11,546
10	Dirk Nowitzki	47,053	Derek Fisher	38,318	Ricky Davis	28,802	Antoine Wright	22,124	Bobby Jackson	17,040	Joel Anthony	11,468
11	Vince Carter	46,936	Marvin Williams	38,152	Al Thornton	28,749	Darko Milicic	22,106	Sasha Vujacic	16,880	Ronnie Price	11,402
12	Deron Williams	46,845	Troy Murphy	38,081	Charlie Villanueva	28,184	Darius Songaila	22,103	Carlos Arroyo	16,803	Malik Allen	11,259
13	Stephen Jackson	46,780	Rafer Alston	37,792	Kyle Korver	28,084	Zaza Pachulia	22,071	Mark Blount	16,769	Chris Quinn	11,230
14	Raymond Felton	46,622	Kevin Martin	36,967	Brendan Haywood	27,983	Spencer Hawes	21,991	Kevin Love	16,669	Dan Gadzuric	11,164
15	Steve Nash	46,241	Andrea Bargnani	36,872	Kenyon Martin	27,980	Kelenna Azubuike	21,909	Joey Graham	16,504	Ruben Patterson	11,139
16	Danny Granger	44,800	Earl Watson	36,624	Trevor Ariza	27,905	Ime Udoka	21,876	Lou Williams	16,432	Mardy Collins	11,109
17	Rashard Lewis	44,796	Steve Blake	36,609	Michael Finley	27,314	Ronny Turiaf	21,834	Tony Allen	16,380	Hilton Armstrong	11,098
18	Carmelo Anthony	44,607	Corey Maggette	36,458	Maurice Evans	27,273	Desmond Mason	21,825	J.J. Redick	16,281	Shaun Livingston	11,069
19	Richard Jefferson	44,195	Udonis Haslem	36,362	Mickael Pietrus	27,156	Jamario Moon	21,783	Matt Harpring	16,124	Brandon Jennings	11,028
20	Amare Stoudemire	44,009	Devin Harris	36,074	Erick Dampier	27,145	Devin Brown	21,695	Jannero Pargo	16,092	Greg Buckner	10,956
21	John Salmons	43,992	Richard Hamilton	35,991	Mike Conley	27,132	Marc Gasol	21,243	Johan Petro	16,085	Louis Williams	10,837
22	Caron Butler	43,968	Kevin Garnett	35,817	Tracy McGrady	27,095	Luke Walton	21,210	Daequan Cook	16,052	Tyronn Lue	10,799
23	Baron Davis	43,896	Brad Miller	35,648	Andray Blatche	27,074	Marquis Daniels	21,140	Anthony Morrow	16,044	Sam Cassell	10,783
24	Josh Smith	43,851	Tony Parker	35,389	Elton Brand	26,906	Kurt Thomas	20,910	Brandon Bass	16,016	Shannon Brown	10,719
25	David West	43,690	Chris Duhon	35,372	O.J. Mayo	26,888	Gilbert Arenas	20,868	Jerry Stackhouse	15,986	Antoine Walker	10,706
26	Shawn Marion	43,629	Jeff Green	34,935	Thaddeus Young	26,800	Eddie House	20,834	DeSagana Diop	15,962	Jonny Flynn	10,695
27	Hedo Turkoglu	43,429	Rasheed Wallace	34,399	Nate Robinson	26,773	Trenton Hassell	20,735	Stephon Marbury	15,942	Travis Diener	10,637
28	Gerald Wallace	43,407	Rasual Butler	34,381	Travis Outlaw	26,601	Eric Gordon	20,699	Dorell Wright	15,676	Damon Jones	10,551
29	Ray Allen	43,312	Jose Calderon	34,105	Sebastian Telfair	26,548	Anthony Carter	20,679	Nazr Mohammed	15,627	Yakhouba Diawara	10,507
30	Chris Bosh	43,234	Raja Bell	34,066	Damien Wilkins	26,462	Joe Smith	20,623	Earl Boykins	15,455	Louis Amundson	10,496
31	Jamal Crawford	43,109	Nick Collison	33,773	Thabo Sefolosha	26,415	Antonio Daniels	20,580	Sergio Rodriguez	15,429	Ryan Hollins	10,496
32	Chauncey Billups	43,101	Andrew Bogut	33,578	Michael Redd	26,288	Vladimir Radmanovic	20,564	Brevin Knight	15,427	Gerald Green	10,446
33	Boris Diaw	42,703	Ben Wallace	33,548	Andrew Bynum	26,127	Joel Przybilla	20,361	Jose Juan Barea	15,270	Donte Greene	10,406
34	David Lee	42,058	Beno Udrih	33,220	Shaquille O'Neal	26,077	Quinton Ross	20,277	Bobby Simmons	15,147	Brian Scalabrine	10,220
35	Jason Terry	42,027	Charlie Bell	32,791	Francisco Garcia	25,833	Jason Thompson	20,148	Luc Richard Mbah a Mo	14,977	Damon Stoudamire	10,033
36	Jason Richardson	41,972	Chris Kaman	32,745	Mikki Moore	25,590	Corey Brewer	19,934	Rashad McCants	14,961	J.J. Hickson	10,000
37	Lamar Odom	41,838	Mike Dunleavy	32,577	Keith Bogans	25,362	Rasho Nesterovic	19,823	James Jones	14,839	Will Bynum	9,989
38	Monta Ellis	41,190	Andrei Kirilenko	32,491	Roger Mason	25,261	Luther Head	19,741	George Hill	14,710	Marco Belinelli	9,987
39	Anthony Parker	41,146	Matt Barnes	32,362	Derrick Rose	25,171	Morris Peterson	19,736	Chucky Atkins	14,471	Chris Douglas-Roberts	9,683
40	Al Harrington	41,112	Leandro Barbosa	32,337	Channing Frye	25,100	Jared Dudley	19,490	Chris Andersen	14,460	Marreese Speights	9,650
41	Emeka Okafor	41,065	Paul Millsap	32,061	Jason Kapono	25,015	Nenad Krstic	19,399	Juan Dixon	14,353	Kevin Ollie	9,635
42	Tayshaun Prince	40,991	Kendrick Perkins	31,802	Ronald Murray	24,884	Yi Jianlian	19,211	Jason Collins	14,319	Nicolas Batum	9,630
43	Ben Gordon	40,867	Nene Hilario	31,676	Bruce Bowen	24,745	Carl Landry	19,159	Devean George	14,240	Julian Wright	9,560
44	Paul Pierce	40,827	J.R. Smith	31,579	Chris Wilcox	24,716	Brandon Rush	18,997	Glen Davis	14,098	Taj Gibson	9,535
45	Pau Gasol	40,827	Jermaine O'Neal	31,564	Wilson Chandler	24,707	Nick Young	18,976	Danilo Gallinari	13,895	Eddie Jones	9,527
46	Shane Battier	40,740	Jameer Nelson	31,478	Tyrus Thomas	24,604	Matt Carroll	18,958	Rudy Fernandez	13,885	Ryan Anderson	9,435
47	Jarrett Jack	40,737	Tyson Chandler	31,467	Jeff Foster	24,261	Sasha Pavlovic	18,891	Stephen Curry	13,801	Goran Dragic	9,414
48	Mo Williams	40,617	Al Horford	31,233	Rodney Stuckey	24,191	Anthony Johnson	18,781	Shelden Williams	13,731	Jonas Jerebko	9,322
49	Tim Duncan	40,522	Josh Howard	31,201	Russell Westbrook	24,190	Ramon Sessions	18,706	Brian Skinner	13,715	Darren Collison	9,252
50	Chris Paul	40,417	Manu Ginobili	31,152	Jarvis Hayes	23,956	Reggie Evans	18,599	D.J. Augustin	13,689	James Singleton	9,169
51	Luol Deng	40,259	Anderson Varejao	31,125	Cuttino Mobley	23,915	Eduardo Najera	18,416	Roy Hibbert	13,626	JaVale McGee	9,120
52	Ryan Gomes	40,214	Andris Biedrins	30,943	Chuck Hayes	23,911	Josh Boone	18,415	Amir Johnson	13,586	Eric Snow	9,088
53	Al Jefferson	40,184	Quentin Richardson	30,442	Yao Ming	23,740	Arron Afflalo	18,406	Kwame Brown	13,554	Solomon Jones	9,052
54	Marcus Camby	40,171	Hakim Warrick	30,398	Jared Jeffries	23,737	Rodney Carney	18,385	Royal Ivey	13,401	Omri Casspi	8,748
55	Dwyane Wade	40,132	Willie Green	30,335	Craig Smith	23,697	Matt Bonner	18,327	Bostjan Nachbar	13,294	Kenny Thomas	8,659
56	Brandon Roy	40,028	James Posey	30,109	Kyle Lowry	23,638	Courtney Lee	17,996	Dominic McGuire	13,117	Stromile Swift	8,634
57	Ron Artest	39,947	Randy Foye	29,993	Keyon Dooling	23,530	Mario Chalmers	17,965	Marcus Williams	13,086	Juan Carlos Navarro	8,621
58	Mehmet Okur	39,831	Drew Gooden	29,973	Martell Webster	23,281	C.J. Watson	17,950	Adam Morrison	13,011	Jacque Vaughn	8,568
59	Rajon Rondo	39,647	Larry Hughes	29,899	Joakim Noah	23,155	Juwan Howard	17,883	Francisco Elson	12,653	Bonzi Wells	8,544
60	Grant Hill	39,532	Delonte West	29,880	Tim Thomas	23,000	Fred Jones	17,789	Smush Parker	12,626	Anthony Randolph	8,543

Description of the Model

In our Skills Plus Minus ("SPM") framework, we run a series of nested probit regressions to estimate the likelihood of various events for a given play. We order a series of events $\{EVTi, i = 1,\ldots n\}$ sequentially. We then define Pr_{EVTi}, the conditional probability of each $EVTi$ occurring, as:

$$\mathcal{N}(\mu_{EVTi} + B_{EVTiHC}HC + \sum_{n=1}^{3} B_{EVTiSTn} \cdot STn + \sum_{n=1}^{360} B_{EVTiOFFn} \cdot OFFn + \sum_{n=1}^{360} B_{EVTiDEFn} \cdot DEFn)$$

Pr_{EVTi} is the probability of the event i, conditional on all prior events in the sequence not occurring (since only one event can occur per play). $\mathcal{N}(\cdot)$ is the cdf of the standard normal distribution, μ_{EVTi} is a constant associated with the event, HC is the home court dummy variable, STn is the possession start variable, and $OFFn$ and $DEFn$ are player dummy variables. HC is 1 if the home team has possession, and 0 if the away team has possession. STn are dummy variables for either "Defensive Rebound", "Offensive Rebound", or "Steal".

"Out of Bounds" has been normalized to 0. $OFFn$ are the dummy variables that indicate the offensive players on the court during the play, while $DEFn$ are the dummy variables that indicate the defensive players.

We have dummy variables for the 360 players[2] who have participated in the most plays in our data sample, and define all others to be "replacement level" players. B_{EVTiHC}, $B_{EVTiSTn}$, $B_{EVTiOFFn}$, and $B_{EVTiDEFn}$ are coefficients associated with the variables, for event i. Each player has two ratings in any given event: offense and defense. Table 5 displays the regression results for steals.

Table 5. Probit Estimation of Steals

$$\mathcal{N}(\mu_{STL} + B_{STLHC}HC + \sum_{n=1}^{3} B_{STLSTn} \cdot STn + \sum_{n=1}^{360} B_{STLOFFn} \cdot OFFn + \sum_{n=1}^{360} B_{STLDEFn} \cdot DEFn)$$

	Estimate	Std. Err.	z value	Pr(>|z|)
(Intercept)	-1.4230	0.0187	-76.02	0.0%
Home Court	-0.0128	0.0039	-3.29	0.1%
Dreb	0.0598	0.0045	13.39	0.0%
Oreb	-0.1336	0.0069	-19.29	0.0%
Steal	0.0130	0.0082	1.57	11.5%

Offense	Estimate	Std. Err.	z value	Pr(>|z|)	Defense	Estimate	Std. Err.	z value	Pr(>|z|)
Chris Paul	-0.1483	0.0269	-5.51	0.0%	Thabo Sefolosha	0.1169	0.0207	5.66	0.0%
Vince Carter	-0.0927	0.0191	-4.85	0.0%	Trevor Ariza	0.1045	0.0201	5.19	0.0%
Leandro Barbosa	-0.0951	0.0200	-4.76	0.0%	Renaldo Balkman	0.1348	0.0265	5.09	0.0%
Kobe Bryant	-0.1149	0.0247	-4.65	0.0%	Gerald Wallace	0.1063	0.0214	4.96	0.0%
Joe Johnson	-0.1141	0.0249	-4.59	0.0%	C.J. Watson	0.1156	0.0239	4.84	0.0%
Tyreke Evans	-0.1521	0.0343	-4.43	0.0%	Chuck Hayes	0.1051	0.0226	4.66	0.0%
Stephon Marbury	-0.1132	0.0293	-3.86	0.0%	Ronnie Brewer	0.0970	0.0209	4.63	0.0%
LeBron James	-0.0856	0.0223	-3.84	0.0%	Monta Ellis	0.0856	0.0190	4.50	0.0%
Rajon Rondo	-0.0959	0.0254	-3.78	0.0%	Devin Harris	0.0929	0.0211	4.41	0.0%
Jannero Pargo	-0.0982	0.0261	-3.77	0.0%	Thaddeus Young	0.0939	0.0215	4.36	0.0%
… best 10 above, worst 10 below …					… best 10 above, worst 10 below …				
Dwight Howard	0.0739	0.0232	3.18	0.1%	J.J. Hickson	-0.1060	0.0340	-3.11	0.2%
Bonzi Wells	0.0981	0.0304	3.22	0.1%	Fabricio Oberto	-0.0835	0.0260	-3.21	0.1%
Brook Lopez	0.0955	0.0293	3.26	0.1%	Andres Nocioni	-0.0672	0.0198	-3.39	0.1%
Shaquille O'Neal	0.0652	0.0199	3.28	0.1%	Jermaine O'Neal	-0.0739	0.0203	-3.63	0.0%
Louis Amundson	0.1078	0.0327	3.30	0.1%	Wally Szczerbiak	-0.0909	0.0245	-3.71	0.0%
Andris Biedrins	0.0712	0.0206	3.45	0.1%	Joel Anthony	-0.1145	0.0301	-3.80	0.0%
Ryan Hollins	0.0993	0.0279	3.56	0.0%	Andrea Bargnani	-0.0809	0.0207	-3.91	0.0%
Andrew Bogut	0.0866	0.0242	3.57	0.0%	Amare Stoudemire	-0.1041	0.0241	-4.31	0.0%
Chris Kaman	0.0921	0.0204	4.51	0.0%	Erick Dampier	-0.1109	0.0248	-4.48	0.0%
Eddy Curry	0.1471	0.0286	5.15	0.0%	Mike Miller	-0.0918	0.0202	-4.54	0.0%

For example, if Rajon Rondo plays on the road on a team with four other replacement level players, against a team with five replacement level players, the probability of a steal for a possession that started out-of-bounds would be:

$$Pr_{STL} = \mathcal{N}(-1.423 - 0.0128 * 0 - 0.0959) = 6.4\% \text{ if Rondo's team has the ball}$$

$$Pr_{STL} = \mathcal{N}(-1.423 - 0.0128 * 1 + 0.0937) = 9.0\% \text{ if Rondo's opponent has the ball}$$

[2] We use 360 players since there are 30 NBA teams and twelve players are allowed to play in a given game. Thus, replacement players are those who would likely be the worst player on any team. If we change the number of players, then the PORP numbers will change, since the cutoff for a replacement player will be different. The other results, including synergies calculated, however, will not be materially different.

We bucket each event into the following "skill" categories:

Ball-handling Category: Steal, Non-steal turnover

Rebounding Category: Rebound of a missed field goal, Rebound of a missed free throw

Scoring Category: Made field goal (2 or 3 points), Missed field goal, Made free throw (1, 2, 3, or 4 points), Missed free throw (0, 1, 2, or 3 points).

Features of the Model

Uses simulations to estimate both mean and variance of outcomes

The SPM framework estimates how the start-of-play state variable (defensive rebound, offensive rebound, steal or out of bounds) affects the probability of an outcome. If we start a game with an out of bounds play, we are able to simulate an entire basketball game, since we can use the estimated coefficients to estimate the probability of every possible outcome and the resultant end-of-play state variable. We can then convert these simulations into winning percentages and point differentials. To rate each player, we simulate games with the player and four "replacement-level" players on one team, and five "replacement level" players on the other team.

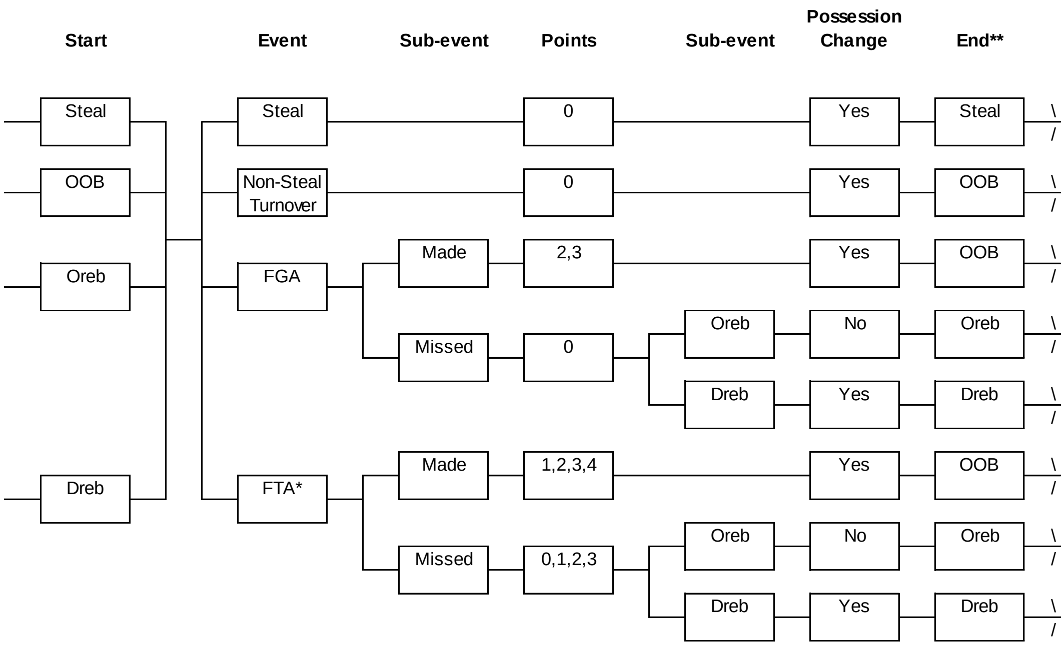


Figure 1. Flow chart of events.

Figure 1 shows the "flow chart" of the simulations. The probabilities associated with each node in the chart are calculated using the point estimates of the nested probit model we estimated. For the analysis done in this paper, we do not simulate games since each simulation is computationally time-consuming. Instead, we calculate a "steady-state" level of outcomes which would occur if a game has infinite length. We rank each player by the estimated point differential of an average length game that starts and ends in this "steady state." The results are not materially different from a simulation that starts with an out-of-bounds play. Using this

* Free throw events include "and-1" situations.

** Steals, Oreb, and Dreb sometimes end with an OOB situation if a timeout is taken or a non-shooting foul is committed, for example.

"steady state" approach, we do not calculate a range of outcomes. Instead, we calculate expected point differentials using the point estimates of the player skill parameters.

Models at the "play" level instead of the "possession level"

Imagine a situation where a team misses five consecutive field goals, and grabs five consecutive offensive rebounds, before finally making a field goal. Traditional APM will consider that sequence of events one possession which results in two points. Our SPM framework will instead count six plays, five of which end in missed field goals and offensive rebounds, and the sixth resulting in a made field goal. SPM will determine that the team with the ball has poor scoring skills but excellent offensive rebounding skills. Our framework distinguishes this sequence of events from a situation where the team immediately scores a field goal, since the outcomes were achieved in dramatically different ways. In the former scenario, the defensive team may want to counter with a defensive rebounder, while in the latter scenario, the defensive team could counter with a stronger on-the-ball defender.

Considers how a play starts

Unlike traditional APM, our framework identifies how each play starts: out-of-bounds, steal, defensive rebound or offensive rebound. We find that the start variable materially affects the outcome of the play. For example, we find that if a play starts with a steal, the average points scored increases from 0.83 to 1.04.

Reveals the strengths and weaknesses of each player

SPM provides granularity to a player's offensive and defensive ratings. If a player is a strong defender, is it because they create steals, prevent scoring, or grab defensive rebounds?

Results and Discussion

Individual Player Ratings

In this section we provide the results of the skill ratings of the 360 players who participated in the most plays in our data sample. See the Appendix for the various tables of player ratings. To estimate the contribution of each skill (e.g. steals), we isolate a player's "steals" ratings, and set his other skills to replacement levels. For example, we create a fictional player who has Ronnie Brewer's "steals" ratings, but is replacement level in all other skills. We then simulate games where one team consists of the fictional player and four replacement players, and their opponent utilizes five replacement players. The estimated point differential of this game is the player's ratings for that particular skill. For example, we estimate that Ronnie Brewer's defensive ball-handling skills are worth 3.2 points per game.

We rank the players by Points Over Replacement Player ("PORP"), the average expected point differential if the player plays an entire game with replacement players. For instance, a team with LeBron James and four replacement players would outscore a team with five replacement players by 15.1 points per game on average. The weighted average PORP across our data set is 2.82 points. The high rating of LeBron James provides some validation of our model, since many experts considered him the best player in the NBA during the four seasons in our data set[5]. Also, not surprisingly, a point guard (Chris Paul) is rated the best ball-handler, while the

[5] LeBron James received the most total votes for Most Valuable Player from 2006-2007 to 2009-2010. Source: www.basketball-reference.com.

best rebounders are generally power forwards and centers (e.g Jason Collins).

SPM Can Predict Which Skills Go Well with Each Other

To investigate synergies, we took the best players in the six skills and isolated their skills by setting their other skills to zero, or replacement level. We then tested $6 * 7/2 = 21$ combinations to see which skills have synergies. The six players are shown in Table 6.

Table 6. The best players in each of the six skills.

	Offensive	**Defensive**
Ballhandling	Chris Paul	Ronnie Brewer
Rebounding	Reggie Evans	Jason Collins
Scoring	Steve Nash	Kevin Garnett

We measured synergies by how many additional points a combination of two skills create. For example, Chris Paul's offensive ballhandling is worth 4.8 points, while Reggie Evans' offensive rebounding is worth 3.1 points. We calculate that a team with Chris Paul's offensive ballhandling and Reggie Evans' defensive rebounding will have a 8.1 point advantage. Therefore we calculate synergies as worth 0.2 points (8.1-4.8-3.1). Synergies are the difference between the point differential of the combined team and the sum of the two individual players; they tell us which types of players work well with one another. Table 7 has the results. We highlight a few of the bigger numbers.

Table 7. Synergies between skills.

	Oballhandling	Dballhandling	Oreb	Dreb	Oscore	Dscore
Oballhandling	-0.825					
Dballhandling	0.000	0.307				
Oreb	0.224	-0.052	0.293			
Dreb	0.071	-0.134	-0.002	-0.394		
Oscore	0.550	0.042	-0.191	0.254	-0.826	
Dscore	-0.064	-0.172	-0.132	0.128	-0.031	-0.284

Offensive ballhandling (preventing turnovers) has negative synergies with itself (-0.825) because a lineup with one great ballhandler does not need another. Defensive ballhandling (creating turnovers) has positive synergies with itself (0.307) because defenders who create turnovers feed off each other, creating more turnovers than they would individually. Offensive scoring has negative synergies with itself (-0.826) because players must share one ball. Defensive scoring has negative synergies with itself (-0.284) because most defensive stands end with a stop anyway.

Offensive rebounding has positive self-synergies (0.293), while defensive rebounding has negative self-synergies (-0.394). This differential sign illustrates a larger aspect of SPM. Because synergy is the excess to the total beyond the sum of the individual parts, any skill that adds to an event that is already likely to happen (such as securing a defensive rebound) will not give as much benefit as a skill that adds to an event that is unlikely to happen (such as securing an offensive rebound).

The cross-terms are more complex. Offensive ballhandling has positive synergies with offensive rebounding (0.550) because offensive ballhandling helps a team convert possessions into shot attempts, and offensive rebounding increases the number of possessions over which the ballhandler can protect the ball. Similarly, offensive ballhandling has positive synergies with offensive scoring (0.550) because the team receives more scoring opportunities, and those opportunities are good ones.

Offensive scoring has positive synergies with defensive rebounding (0.254) and negative synergies with offensive rebounding (-0.191) because defensive rebounding increases the number of potential scoring opportunities while offensive rebounding is more valuable when offensive scoring is low, since poor offensive players generate more offensive rebounding opportunities.

Empirical Evidence Suggests that Synergies Exist

Our framework predicts that skills that affect rare events (e.g. steals, offensive rebounds) will have positive synergies, while skills that contribute to common events (e.g. defensive rebounds) will have negative synergies. This feature is a result of our nested probit specification. Is this specification realistic? Do two players with strong defensive ballhandling skills create more turnovers than one? In this section, we investigate empirical evidence to validate our model.

We sorted the 987,343 observations into one hundred buckets, ordered by predicted steals. Within each bucket (each with 9873 or 9874 observations), we calculated the total predicted steals and the total actual steals. In the following scatterplot, we graph the one hundred data points, each representing a bucket of actual steals and predicted steals. If positive synergies in steals do not exist, then we would see that actual steals are less than predicted steals, for both low and high levels of predicted steals. For medium levels of predicted steals, we would see actual steals are higher than predicted steals. Instead, we see that actual steals are well within the 95% confidence intervals of predicted steals across all levels: only three points out of one hundred fall outside, two below and one above. This evidence suggests that our choice of probit to model the synergies in steals is a reasonable one.

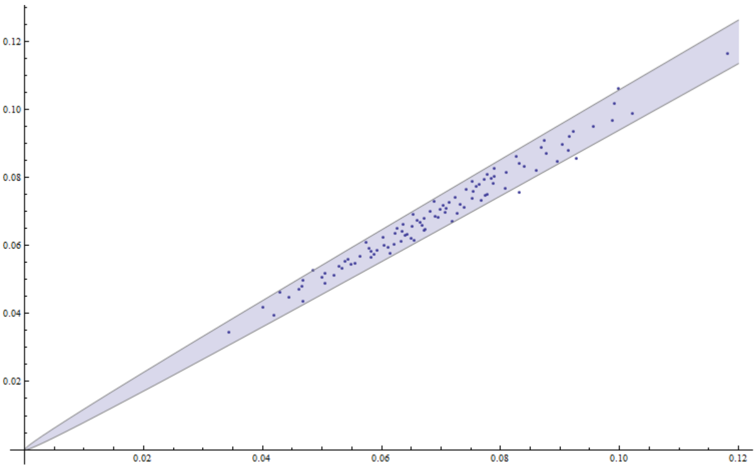


Figure 3: Actual Steals (y-axis) versus Predicted Steals (x-axis), with 95% probability confidence bands

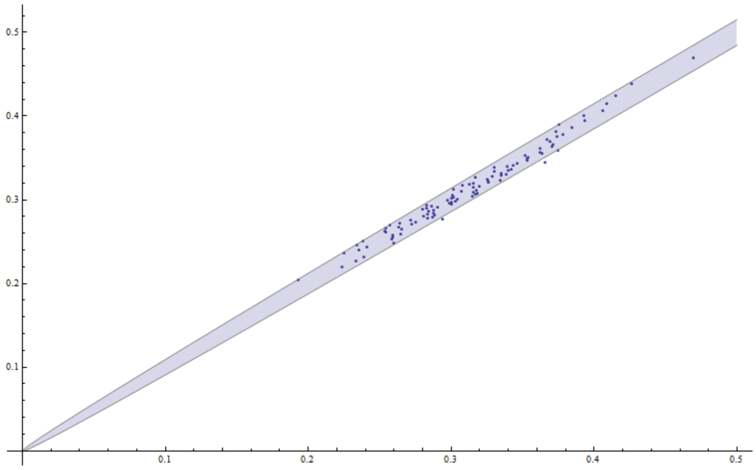


Figure 4: Actual Offensive Rebounds (y-axis) versus Predicted Offensive Rebounds (x-axis), with 95% probability confidence bands

Our framework also predicts that offensive rebounding has positive synergies with itself. Using the same methodology, we plot actual offensive rebounds versus predicted offensive rebounds. We have 407,154 missed field goals in our data set, so that each bucket contains 4,071 or 4072 observations. The above scatterplot shows that only four points out of one hundred fall outside the 95% confidence bands. These two scatterplots suggest that positive synergies do exist for both steals and offensive rebounds, as our framework predicts.

SPM Can Be Used to Calculate Synergies for Each NBA Team

For each NBA team, we formed lineups using the top five players in terms of plays played in our data sample. We calculated their ratings individually and as the 5-player lineup. For a given lineup of players x1, x2, x3, x4 and x5, define PORP(x1,x2,x3,x4,x5) to be the estimated point differential between a game played by this team of players against a lineup of replacement players ("RP").

We then define synergies as the difference of the sum-of-the-parts from the team total:

$$
\begin{aligned}
PORP(x1, x2, x3, x4, x5) &- PORP(x1, RP, RP, RP, RP) - PORP(x2, RP, RP, RP, RP) \\
&- PORP(x3, RP, RP, RP, RP) - PORP(x4, RP, RP, RP, RP) \\
&- PORP(x5, RP, RP, RP, RP)
\end{aligned}
$$

The results are in Table 8. Orlando's lineup has the highest amount of synergies, over one point per game, while Minnesota's negative synergies cost their lineup just under one point per game. Using the Pythagorean expectation formula with coefficients between 14 and 16.5 (c.f. Morey 1993), 1-2 points per game can translate into 3-6 wins per season (for a team that would otherwise score and allow 100 points per game). Thus a team that consistently fields a highly positively synergistic lineup will win up to six games more than if it consistently fields a highly negatively synergistic lineup. Such a differential could be the difference between making or missing the playoffs.

To investigate why Orlando's lineup has positive synergies, we replace players from their lineup one-by-one with replacement players and see how the synergies change. We find that Jameer Nelson and Hedo Turkoglu play well together. Our framework suggests that Nelson's superior ballhandling skills complement Turkoglu's offensive skills, since Nelson gives

Turkoglu more chances to score.

Using the same method, we find that Minnesota's Ryan Gomes and Randy Foye are not good fits since they are both good offensive players who protect the ball well. As noted earlier, our framework predicts negative synergies for both offense (since the players must share the ball) and offensive ball-handling (since one good ball-handler is enough for one lineup).

Table 8. Synergies within teams.

	Player1	Player2	Player3	Player4	Player5	Separate	Combined	Synergies
ORL	D. Howard	R. Lewis	H. Turkoglu	J. Nelson	K. Bogans	24.3	25.6	1.2
CLE	L. James	A. Varejao	Z. Ilgauskas	D. Gibson	M. Williams	30.7	31.8	1.1
IND	D. Granger	T. Murphy	M. Dunleavy	J. Foster	B. Rush	18.2	19.3	1.1
DEN	C. Anthony	N. Hilario	J. Smith	K. Martin	A. Iverson	14.9	16.0	1.1
SAC	K. Martin	B. Udrih	J. Salmons	F. Garcia	B. Miller	12.9	14.0	1.0
NOK	D. West	C. Paul	P. Stojakovic	T. Chandler	R. Butler	23.0	23.8	0.8
DAL	D. Nowitzki	J. Terry	J. Howard	J. Kidd	E. Dampier	25.4	26.0	0.6
LAL	K. Bryant	L. Odom	D. Fisher	P. Gasol	A. Bynum	28.1	28.6	0.4
NJN	V. Carter	D. Harris	B. Lopez	R. Jefferson	J. Kidd	23.6	24.0	0.4
SEA	K. Durant	J. Green	N. Collison	E. Watson	R. Westbrook	18.6	18.8	0.2
DET	T. Prince	R. Hamilton	R. Wallace	R. Stuckey	J. Maxiell	15.4	15.5	0.1
BOS	P. Pierce	R. Rondo	R. Allen	K. Perkins	K. Garnett	29.4	29.5	0.0
UTA	D. Williams	M. Okur	C. Boozer	A. Kirilenko	P. Millsap	25.0	25.0	0.0
HOU	S. Battier	L. Scola	R. Alston	T. McGrady	C. Hayes	22.3	22.3	0.0
GSW	M. Ellis	A. Biedrins	S. Jackson	B. Davis	K. Azubuike	18.0	18.0	0.0
PHI	A. Iguodala	S. Dalembert	A. Miller	W. Green	T. Young	18.6	18.5	-0.1
CHA	R. Felton	G. Wallace	E. Okafor	B. Diaw	M. Carroll	13.2	13.1	-0.2
LAC	C. Kaman	A. Thornton	C. Mobley	E. Gordon	B. Davis	10.2	10.0	-0.2
TOR	C. Bosh	A. Bargnani	J. Calderon	A. Parker	R. Nesterovic	19.1	18.9	-0.2
CHI	L. Deng	K. Hinrich	B. Gordon	D. Rose	J. Noah	19.8	19.5	-0.3
MIA	D. Wade	U. Haslem	M. Chalmers	M. Beasley	D. Cook	18.0	17.7	-0.4
NYK	D. Lee	N. Robinson	W. Chandler	J. Crawford	J. Jeffries	14.3	13.9	-0.4
ATL	J. Johnson	J. Smith	M. Williams	A. Horford	M. Bibby	20.0	19.6	-0.4
PHX	S. Nash	A. Stoudemire	L. Barbosa	G. Hill	R. Bell	26.2	25.6	-0.6
POR	B. Roy	L. Aldridge	T. Outlaw	S. Blake	M. Webster	19.6	19.0	-0.6
MEM	R. Gay	M. Conley	O. Mayo	H. Warrick	M. Gasol	10.0	9.4	-0.6
WAS	A. Jamison	C. Butler	D. Stevenson	A. Blatche	B. Haywood	18.2	17.6	-0.6
MIL	A. Bogut	C. Bell	M. Redd	C. Villanueva	M. Williams	14.6	13.9	-0.7
SAS	T. Duncan	T. Parker	M. Ginobili	M. Finley	B. Bowen	25.8	25.1	-0.7
MIN	R. Gomes	A. Jefferson	R. Foye	C. Brewer	C. Smith	8.2	7.3	-0.8

SPM Gives Context Dependent Player Ratings

An implication of the SPM framework is that player values depend upon the other players on the court. To illustrate this concept, we took the top four players in terms of plays played for each team. We then put everyone else into a "free agent" pool. For each team, we calculated which free agent would be the best fit for the remaining four players. In this analysis, Kevin Garnett is a "free agent" because he switched teams from Minnesota to Boston in our data sample, and played only the fifth highest number of minutes for Boston. Not surprisingly, he would be the most coveted free agent by every single team. Russell Westbrook, a "free agent" because he played only two seasons in our data sample, is likewise highly coveted. There are, however, significant differences among the more marginal players. For example, Eddie Jones, although retired, would fit well in a team like Minnesota (who rank him the fourth most desirable free agent), but would not fit in on the Spurs (who rank him seventeenth). Likewise, Marcus Camby would be coveted by the Knicks or Nets (ranked sixth), but not by the Pacers (ranked nineteenth).

Table 9 shows the “free agent” fits for each team.

Table 9: “Free agents” and synergies.

	Top Choice	2nd Choice	3rd Choice	4th Choice	5th Choice	6th Choice
CHI	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	C. Billups
PHX	K. Garnett	R. Hibbert	A. Johnson	R. Westbrook	N. Batum	B. Jennings
ATL	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	E. Jones
HOU	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	T. Young
IND	K. Garnett	R. Westbrook	A. Johnson	N. Batum	C. Billups	B. Jennings
LAC	K. Garnett	A. Johnson	R. Westbrook	N. Batum	R. Hibbert	C. Billups
MIL	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	T. Young
NOK	K. Garnett	A. Johnson	R. Westbrook	N. Batum	R. Hibbert	C. Billups
NYK	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	N. Batum	M. Camby
POR	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	C. Billups
TOR	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	N. Batum	C. Billups
WAS	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	N. Batum	C. Billups
DEN	K. Garnett	R. Westbrook	C. Billups	N. Batum	A. Johnson	B. Jennings
SAS	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	Y. Ming
CHA	K. Garnett	A. Johnson	R. Westbrook	N. Batum	R. Hibbert	T. Young
CLE	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	C. Billups
DET	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	T. Young
MIN	K. Garnett	A. Johnson	R. Westbrook	E. Jones	N. Batum	R. Hibbert
NJN	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	M. Camby
PHI	K. Garnett	A. Johnson	R. Westbrook	R. Hibbert	N. Batum	T. Young
SAC	K. Garnett	R. Westbrook	N. Batum	C. Billups	A. Johnson	R. Hibbert
SEA	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	C. Billups	N. Batum
UTA	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	N. Batum	T. Young
BOS	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	T. Young
DAL	K. Garnett	R. Westbrook	N. Batum	A. Johnson	R. Hibbert	C. Billups
MEM	K. Garnett	R. Westbrook	A. Johnson	C. Billups	N. Batum	E. Jones
LAL	K. Garnett	R. Westbrook	A. Johnson	R. Hibbert	N. Batum	C. Billups
MIA	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	E. Jones
ORL	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	B. Jennings
GSW	K. Garnett	R. Westbrook	A. Johnson	N. Batum	R. Hibbert	T. Young

Using SPM to Find Mutually Beneficial Trades

Other player rating systems like WP or Win Shares (see Oliver 2004) cannot generate ex-ante mutually beneficial trades because one player is always ranked higher than another (unless the distribution of minutes is changed). In contrast, the SPM framework can generate mutually beneficial trades because each potential lineup has different synergies. We examined every possible two player trade from one team’s starting five to another team’s starting five. There are a total of $30 \cdot 29/2 = 435$ possible team trading partners. Each pair of teams has $5 \cdot 5 = 25$ possible trades, so there are $435 * 25 = 10{,}875$ possible trades. We found 222 mutually beneficial trades, or 2% of all possible trades. These trades do not consider the distribution of minutes or the composition of the team’s bench. Table 10 lists a few trades.

Figure 2 shows the network of the 222 mutually beneficial trades among the various teams. Not surprisingly, the teams with the lowest synergies (Minnesota and San Antonio) have the most possible trading partners and are near the interior of this “trade network”. Meanwhile the teams with the highest synergies (Orlando and Cleveland) have the fewest trading partners and are on the perimeter.

Why is Chris Paul for Deron Williams a mutually beneficial trade? Overall, our SPM ratings rate Chris Paul and Deron Williams nearly the same, but with differences in skills. Paul is a better ballhandler, Williams a slightly better rebounder, and Williams is better at offense and defense. See Table 11.

Table 10. Some mutually beneficial trades.

Team 1	Team 2	Player 1	Player 2
PHX	MIN	Amare Stoudemire	Ryan Gomes
PHX	DAL	Amare Stoudemire	Erick Dampier
NOK	UTA	Chris Paul	Deron Williams
MIN	BOS	Al Jefferson	Rajon Rondo
MIN	MIA	Ryan Gomes	Udonis Haslem
MIN	MIA	Corey Brewer	Daequan Cook
DET	LAL	Rodney Stuckey	Derek Fisher
HOU	MIN	Luis Scola	Ryan Gomes
MIL	MIN	Mo Williams	Al Jefferson
ATL	MIA	Marvin Williams	Daequan Cook

Table 11. Comparison of Chris Paul and Deron Williams

	Off Ballhand.	Def Ballhand.	Off Rebound.	Def Rebound.	Off Scoring	Def Scoring
Chris Paul	4.8	1.2	-0.4	-1.4	4.7	-0.9
Deron Williams	1.9	-0.3	-1.7	0.1	6.5	1.4

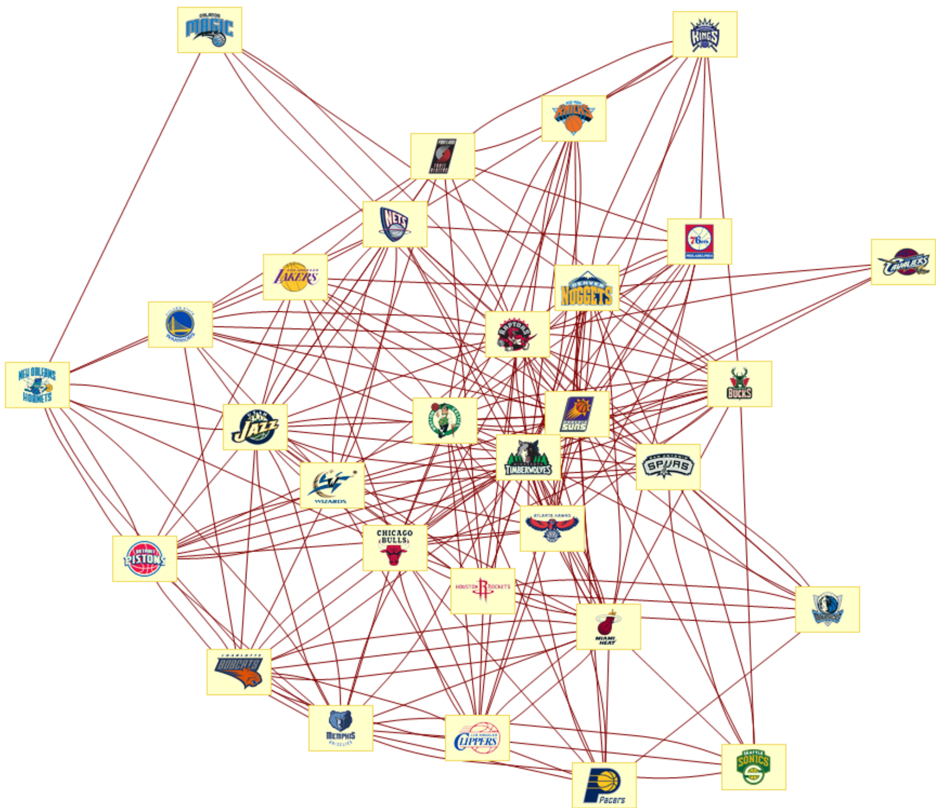


Figure 2. Trade network of mutually beneficial trades.

The SPM framework predicts that Chris Paul is a better fit for Utah because he creates a lot of steals (3.1 steals per 48 minutes (“SP48M”)), while no one else in the New Orleans lineup does (West 1.0 SP48M, Stojakovic 1.1, Chandler 0.7, Butler 0.9). Utah, on the other hand, has many players who create steals (Kirilenko 2.0, Boozer 1.5, Millsap 1.7, Okur 0.9, Williams 1.4). Because defensive steals has positive synergies in our system, Chris Paul's ballhawking skills fit better in Utah, where he can team up with others and wreak havoc to opponents' ballhandlers.

Conversely, why would New Orleans trade for Deron Williams? Our framework predicts that Williams is a better offensive fit with New Orleans. There are negative synergies between two good offensive players since they must share only one ball, and the New Orleans starters take fewer shots than Utah's. At New Orleans, Deron Williams would not need to share the ball with so many players.

The Utah lineup of Williams (PG), Okur (F-C), Boozer (F-C), Kirilenko (F) and Millsap (F) may seem big. The next player on Utah's roster in terms of plays in our sample is Ronnie Brewer (G-F). If we substitute Millsap for Brewer, the case for a Deron Williams for Chris Paul trade becomes stronger, since Brewer is good at steals (2.7 SP48M).

Conclusion

We provide a novel Skills Plus Minus ("SPM") framework that can be used to measure synergies within basketball lineups, provide roster-dependent rankings of free agents, and generate mutually beneficial trades. To our knowledge, the SPM framework is the first system that can generate ex-ante mutually beneficial trades without a change in the minutes played. Other ranking systems cannot generate mutually beneficial trades because one player is always ranked ahead of another.

Future research could use the SPM framework to calculate the optimal substitution patterns that maximize overall synergies given a fixed distribution of minutes played to each player, highlight the risks and exposures each team with respect to the specific skills, and evaluate the possibility of a separate synergy factor of players that may improve the skills of their teammates by even more than would be suggested by the synergies of the skills.

Acknowledgments

The authors thank the two anonymous reviewers, Kevin Arnovitz, David Berri, Jeff Chuang, Harry Gakidis, Matt Goldman, Shane Kupperman, Irwin Lee, Wayne Winston, and members of the APBRmetrics forum for their helpful feedback and comments.

References

Berri, D., Jewell, R. (2004). Wage inequality and firm performance: Professional basketball's natural experiment. *Atlantic economic journal*, *32*(2), 130-139.

Berri, D. (1999). Who is most valuable? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics*, *20*, 411-427.

Berri, D. (2008). A simple measure of worker productivity in the National Basketball Association. In Brad Humphreys and Dennis Howard (Eds.), *The Business of Sport*, 3 volumes (pp. 1-40). Westport, Conn.: Praeger.

Berri, D., Brook, S. (1999). Trading players in the National Basketball Association: For better or worse. In John Fizel, Elizabeth Gustafson, and Larry Hadley (Eds.), *Sports Economics: Current Research* (pp. 135-151).

Berri, D., Schmidt, M. (2010). *Stumbling on Wins: Two Economists Explore the Pitfalls on the Road to Victory in Professional Sports*. Princeton, NJ: Financial Times Press.

Bollinger, C., Hotchkiss, J. (2003). The upside potential of hiring risky workers: Evidence from the baseball industry. *Journal of labor economics*, *21*(4), 923-944.

ElGee (2011). Interpreting advanced statistics in basketball. Retrieved from http://www.backpicks.com/2011/01/24/interpreting-advanced-statistics-in-basketball.

Fearnhead, P., Taylor, B. (2011). On estimating the ability of NBA players. *Journal of quantitative analysis in sports*, *7*(3), Article 11.

Idson, T., Kahane, L. (2000), Team effects on compensation: An application to salary determination in the National Hockey League. *Economic inquiry 38*(2), 345-357.

Kubatko, J., Oliver, D., Pelton, K., Rosenbaum, D. (2007). A starting point for analyzing basketball statistics. *Journal of quantitative analysis in sports 3*(3), Article 1.

MacDonald, D., Reynolds, M. (1994). Are baseball players paid their marginal products? *Managerial and decision economics*, *15*(5), 443-457.

Morey, D. (1993). In John Dewan, Don Zminda, STATS, Inc. Staff, *STATS Basketball Scoreboard* (p. 17). STATS, Inc..

Oliver, D. *Basketball on paper*. Potomac Books Inc..

Pelton, K. (2010). WARP2 electric boogaloo. Retrieved from http://basketballprospectus.com/article.php?articleid=1209.

Rosenbaum, D. (2004). Measuring how NBA players help their teams win. Retrieved from http://www.82games.com/comm30.htm.

Appendix: Player Ratings

Best and Worst Overall

Best	PORP	Worst	PORP
LeBron James	15.1	Johan Petro	-3.3
Steve Nash	14.3	Gerald Green	-3.3
Dwyane Wade	13.5	Joel Anthony	-3.8
Kevin Garnett	13.3	Brian Skinner	-4.5
Kobe Bryant	10.2	Dominic McGuire	-4.5
Dirk Nowitzki	9.7	Hakim Warrick	-4.9
Tim Duncan	9.6	Earl Boykins	-5.4
Chris Bosh	9.5	Eddy Curry	-6.7
Manu Ginobili	9.4	Josh Powell	-7.8
Russell Westbrook	9.4	J.J. Hickson	-8.8

Best and Worst Offensive Ballhandling (preventing steals and turnovers)

Best	PORP	Worst	PORP
Chris Paul	4.8	Mikki Moore	-2.4
Brandon Jennings	4.6	Andrew Bogut	-2.4
Kobe Bryant	4.3	Louis Amundson	-2.5
Sasha Vujacic	3.8	Hilton Armstrong	-2.7
Sam Cassell	3.6	Kwame Brown	-2.8
LeBron James	3.3	Yao Ming	-2.8
Chauncey Billups	3.2	Ryan Hollins	-3.3
Mike Conley	3.1	Kendrick Perkins	-3.4
Daequan Cook	3.1	Joel Przybilla	-3.5
Jason Terry	3.0	Eddy Curry	-6.3

Best and Worst Defensive Ballhandling (creating steals and turnovers)

Best	PORP	Worst	PORP
Ronnie Brewer	3.2	Tim Duncan	-2.0
Gerald Wallace	2.9	Michael Finley	-2.3
Thabo Sefolosha	2.9	Brook Lopez	-2.4
Devin Harris	2.9	Aaron Brooks	-2.5
Monta Ellis	2.8	Andrew Bynum	-2.5
Renaldo Balkman	2.8	Taj Gibson	-2.6
Rajon Rondo	2.7	Joel Anthony	-2.8
Luc Richard Mbah a Moute	2.7	Amare Stoudemire	-3.3
C.J. Watson	2.7	Erick Dampier	-3.6
Eddie Jones	2.7	J.J. Hickson	-4.2

Best and Worst Offensive Rebounding

Best	PORP	Worst	PORP
Reggie Evans	3.1	Chris Quinn	-1.9
Matt Harpring	3.0	Jannero Pargo	-2.0
Kevin Love	2.9	Donte Greene	-2.0
Jeff Foster	2.7	Brandon Rush	-2.1
Jason Maxiell	2.6	Rashard Lewis	-2.3
Louis Amundson	2.5	Damon Stoudamire	-2.3
Leon Powe	2.2	Danilo Gallinari	-2.4
Amir Johnson	2.1	Travis Diener	-2.5
Joakim Noah	2.0	Stephen Curry	-2.8
Jared Jeffries	2.0	Jonny Flynn	-2.8

Best and Worst Defensive Rebounding

Best	PORP	Worst	PORP
Jason Collins	3.0	Francisco Garcia	-1.5
Tim Duncan	2.6	Sasha Vujacic	-1.5
Joel Przybilla	2.5	Eddie House	-1.6
Jeff Foster	2.5	Josh Childress	-1.6
Andrew Bogut	2.3	Dominic McGuire	-1.6
Zydrunas Ilgauskas	2.3	Darren Collison	-1.6
Nene Hilario	2.2	Charlie Bell	-1.7
Roy Hibbert	2.2	Jamaal Tinsley	-1.8
Rasho Nesterovic	2.2	Travis Diener	-2.1
Samuel Dalembert	2.0	Earl Boykins	-2.1

Best and Worst Offense (assuming no turnovers)

Best	PORP	Worst	PORP
Steve Nash	12.7	James Singleton	-2.3
Dwyane Wade	9.4	Josh Powell	-2.3
LeBron James	7.8	Hilton Armstrong	-2.4
Deron Williams	6.5	Louis Amundson	-2.4
Kevin Martin	6.4	Brian Skinner	-2.4
Kobe Bryant	6.3	Ben Wallace	-2.5
Goran Dragic	6.2	Jason Collins	-2.7
Dirk Nowitzki	5.9	Eric Snow	-3.0
Manu Ginobili	5.9	Renaldo Balkman	-3.4
Danny Granger	5.9	Nene Hilario	-3.7

Best and Worst Defense (assuming no turnovers)

Best	PORP	Worst	PORP
Kevin Garnett	6.2	Damien Wilkins	-3.0
Brendan Haywood	5.7	Josh Powell	-3.0
Tim Duncan	5.4	Kevin Martin	-3.0
Joel Przybilla	5.2	Gerald Green	-3.0
Amir Johnson	5.0	Marreese Speights	-3.2
Andrew Bogut	4.8	Juan Carlos Navarro	-3.2
Chris Andersen	4.5	Royal Ivey	-3.4
Jacque Vaughn	3.9	Jose Calderon	-3.4
Yao Ming	3.9	Sasha Vujacic	-3.7
Kendrick Perkins	3.9	Will Bynum	-4.2

Computer simulations of table tennis ball trajectories for studies of the influence of ball size and net height

Ralf Schneider[1], Oleksandr Kalentev[1], Tatyana Ivanovska[2] & Stefan Kemnitz[3]

[1]Institute of Physics, Ernst-Moritz-Arndt University Greifswald

[2]Institute of Community Medicine, Ernst-Moritz-Arndt University Greifswald

[3]Faculty of Informatics and Electrical Engineering, University Rostock

Abstract

One possible measure to increase the medial appeal of table tennis is to slow down the game by using bigger balls or higher nets. Usually, an empirical approach is followed to study the effect of such changes on the players and the game. In this work, a different approach is taken, namely solving numerically the equation of motion for table tennis balls for systematical, statistical studies of the impact of ball size and weight as well as of net height on the distribution functions of successful strokes.

The analysis confirms the empirical observation that the change of the ball in the year 2000 from a 38-mm to a 40-mm-ball can be compensated with other parameters such that their resulting trajectory distribution functions are nearly identical. This was also observed in reality, where adaptation of the player's technique compensated the larger ball size. A larger ball of 44 mm with small weight is one option for suppressing high velocities, coupled also to a reduction of the influence of spinning. As an alternative an increase of the net height is possible. A small increase of the net height could be one future option, where the basic character of the game is not strongly modified, but especially the influence of the service could be reduced.

KEYWORDS: SPORTS EQUIPMENT, PHYSICS COMPUTING, MONTE CARLO METHODS

Introduction

The medial appeal of table tennis seems to go down in terms of TV hours, at least outside Asia. One of the reasons is the fact that the speed of the game is nowadays so high that it is very hard for spectators to follow the balls (Nelson 1997, Djokic 2007). Possible counteractions to slow down the game are to use bigger balls or higher nets. Usually, empirical studies are done to study the effect of such changes on the players and the game. An alternative approach, followed in this work, is the use of computer simulations. The equation of motion for table tennis balls is solved numerically to allow systematical, statistical studies of the impact of ball size and weight as well as of net height on the distribution functions of successful strokes.

One key problem for the medial appeal of table tennis is that the spin of the ball, the rotation, is not visible for spectators, because they see only its effect. This makes it difficult to understand why a simple looking ball of the opponent leads to a mistake for the other player. Therefore, one intention of possible rule changes is to reduce the impact of spin on the game. Another goal is to reduce the speed of the balls to allow a better visual tracking during the rallies (Djokic 2007). Some rule changes, like a larger ball, different counting system, stricter limits for rubbers or new service rules, were already implemented and new modifications are under discussion (Djokic 2007). For the players all rule or technical changes have strong impacts on their techniques and strategies, requiring usually adaptations of their individual training programs. Therefore, players are rather hesitant to new rules.

The 40-mm-ball played today is 2 mm larger and 0.2 grams heavier than the 38-mm-ball used before. It has a larger air drag due to its larger cross sectional area reducing the maximum velocities (Bai 2005). The mass distribution of the larger ball is shifted further away from the center compared with the 38 mm ball. This creates a larger inertial moment and reduces the spin. The larger 40-mm-ball results in a velocity and spin reduction of about 5 to 10 percent (Li 2005, Iimoto 2002). However, the larger ball had practically no impact on the characteristics of table tennis, because larger exertions of forces by the players compensated the effects of the size increase (Liu 2005, Li 2005). As a consequence of the modified technique, the fitness of the individual player got more important. In modern table tennis the forces for a stroke are created not only by the arms but the whole body is used to support this. A stronger athletics allows more pronounced use of the legs producing larger forces on the ball, which are needed to compensate the size increase. In addition, the wrist has to be used more effective to produce spin. For the larger ball only the use of the forearm is no longer sufficient for spin, as it was the case for the 38-mm-ball. The needs for larger exertion of forces amplify possible technical mistakes, because the individual movement execution gets extended (Kondric 2007).

One obvious strategy to reduce the maximum velocity in table tennis rallies is to increase the net height. However, such a change will have a severe impact on the characteristics of table tennis, because this will limit very directly fast spins, shots and service. Therefore, up to now this change of rule was avoided and ball size was the preferred correction action. Nevertheless, a scientific data base is still missing for a decision.

In this work the impact of larger balls or higher nets on table tennis trajectories is studied using computer simulations. A data base is created to quantify the influence of such changes. Modifications in technique, tactics, strength and fitness are not considered in this analysis. For a huge number of initial conditions the effect on successful strokes is studied. This delivers the maximum amount of possible strokes for different conditions in terms of statistical distributions which can be compared and analyzed. This represents already the best possible adaptation to the changes, independent of what this would mean for the players in terms of changes in their training. In particular the impact of the changes on the ball velocity distributions will be discussed as motivated before.

After a short discussion of the effects of larger balls and higher nets as measures to slow down table tennis, the forces acting on a moving ball are introduced. The computer code solving the equation of motion is described and statistical analysis of trajectory distribution functions for different balls and net heights is done. Using for this a GPU (Graphics Processing Unit) by CUDA (Compute Unified Device Architecture, CUDA 2013) coding gives a very large speed-up compared to CPUs. Results for different cases are compared and analyzed. Finally, the results are summarized and discussed.

Methods

For a quantitative analysis of ball size and net height effects a computational approach is followed. The basic element of the simulation is the solution of the equation of motion for table tennis balls. The equation of motion needs a mathematical description of the acting forces. The flight trajectory of a table tennis ball is determined by the gravitational force of the earth and aero dynamical forces.

The gravitational force

$$\vec{F}_G = -m \cdot \vec{g}$$

alone results in a parabolic trajectory. This force acts towards the centre of the earth and depends on the mass m of the ball and the gravitational constant g (9.81 m/s²).

The aero dynamical forces modify the simple parabola by air drag and lift. Air drag acts as friction force against the direction of the movement of the ball. A simple example for this force is the back pushing of a hand held out of a driving car. A larger velocity gives stronger force acting against the direction of the car. This force also gets larger if one puts out not only a part of the hand, but the full hand. It scales with the cross sectional area. The mathematical expression is

$$\vec{F}_D = -\tfrac{1}{2} \cdot C_D \cdot \rho \cdot A \cdot v \cdot \vec{v},$$

with the density of air ρ, the cross sectional area A for a ball with radius r ($A = r^2 \cdot \pi$), the ball velocity v and an air drag coefficient C_D. This coefficient can be measured, e.g. in wind tunnel experiments.

The second important aero dynamic force is the air lift. The so-called "Magnus effect", named after his discoverer Heinrich Gustav Magnus (1802-1870), is the reason that a rotating ball experiences a deviation from its flight path. A famous example for this is a free kick goal from the Brazilian soccer player Roberto Carlos in a friendly game with France at the 3rd of June 1997. Carlos gave a lot of spin to the ball during the free kick hitting the ball right from the center of gravity with his left foot. The flight path of the ball got extreme passing around the defenders who formed a wall into the goal.

The Magnus effect is a surface effect, because around the spinning ball a co-rotating air layer is formed at the surface of the ball. The flying and spinning ball induces a pressure imbalance, because on one side the ball is rotating with the air flow created by the movement of the ball in the air, the other side opposite to it. On the side where counter-rotation exists, the total velocity of the air flow is reduced, because both velocities compensate partly. On the co-rotation side a larger flow velocity is created, because both velocities add up. Higher velocity in a flow means lower pressure and the pressure differences on the two sides lead to the deviating Magnus force, mathematically expressed with an air lift coefficient C_L as

$$\vec{F}_L = \tfrac{1}{2} \cdot C_L \cdot \rho \cdot A \cdot v \cdot \overrightarrow{e_\omega} \times \vec{v}$$

The air lift force acts perpendicular to the axis of rotation $\overrightarrow{e_\omega}$ and to the velocity $\vec{v}$.

Air drag and lift coefficients of a rotating ball (see Figure 1) as a function of the ratio of spinning velocity to translational velocity are implemented into the computer code as a fit of experimental data (Achenbach 1972, Bearman 1976, Davies 1949, Maccoll 1928, Mehta 1985) as a rational function y(x)

$$y(x) = \frac{a + b \cdot x + c \cdot x^2 + d \cdot x^3}{1 + e \cdot x + f \cdot x^2 + g \cdot x^3}$$

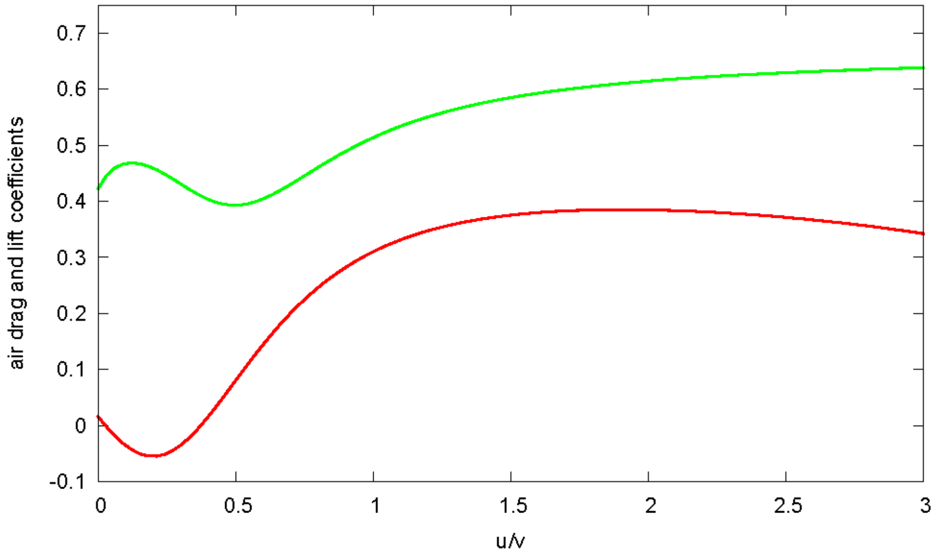


Figure 1: Air drag coefficient CD (upper green curve) and air lift coefficient CL (lower red curve) as a function of the ratio of spinning velocity u to the translational velocity v.

During a topspin shot with forward rotation the lift force acts downwards, during a backspin with backward rotation it acts upwards.

Swirling balls, often quoted in soccer and volleyball, can be created when the ball is hit with a critical velocity leading to the access of the inverse Magnus effect. It shows up in Figure 1 for low spinning velocities as a negative value of the air lift coefficient. This can lead also in table tennis to swirling balls, because during the flight path the regime of positive and negative air lift coefficients can change resulting in a swirling. However, for table tennis balls negative air lift coefficients exist only where the coefficient itself is already quite small. Therefore, the effect exists, but gives only deviations of some millimeters. The frequently quoted swirling balls with long pimples are therefore more a psychological effect than physics: the pre-programmed movement of the player anticipates a flight path of a strongly rotating ball from a normal rubber sponge. The balls from the long pimples with reduced rotation have a different flight path with less lift and fall down earlier such that the player is missing the ball and he complains, that the ball was swirling.

The computer code solves the equation of motion of table tennis balls for given initial positions, velocities and spins. An Euler solver was used, because its algorithmic simplicity allowed an easy transfer onto the GPU with CUDA. A commonly used Runge-Kutta algorithm was not chosen, because it has larger computational costs. A fourth order Runge Kutta approach needs to calculate four times the forces, which slows down the code performance in our case compared to the simple Euler method. This was not compensated by the larger time step possible with the Runge-Kutta method compared to the Euler method. The dependence of the aero dynamic forces on the velocity also does not allow the use of a Verlet algorithm. Therefore, we decided to stay with the Euler method.

One example of a table tennis ball trajectory is shown as a red line in Figure 2. The table tennis table region is marked in green, the net is blue. The orange sphere is the initial point of the trajectory, where the ball is hit. The spinning of the ball is taken constant during the flight. x and y are the spatial coordinates within the plane of the table tennis table. z is the height coordinate above the table. A time step of 0.0001 seconds was used.

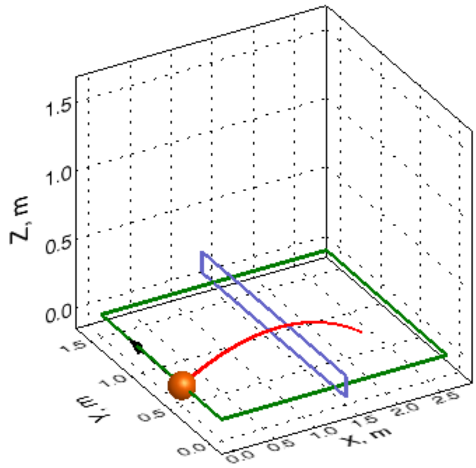


Figure 2: 3D trajectory of a table tennis ball

The ball in Figure 2 is hit at the baseline (x = 0 m) in the forehand part of the table (y = 0.6 m) on the height of the table (z = 0 m). The black arrow shows the rotation axis of the ball, which is here purely pointing into positive y-direction: the ball was a pure topspin without any sidespin.

Results

For a statistical analysis of the effects of ball sizes and net heights on trajectories of table tennis balls a Monte Carlo procedure was used. Many different initial conditions were solved: x was varied between 0.3 m to -3 m, representing hitting locations from 30 cm above the table to 3m behind the table. y was kept constant at 0.381 m, which is ¼ of the width of the table tennis table. This was chosen as a representative position, the exact location of the hitting point in y (forehand or backhand position) is not important for this numerical test. Initial height z was sampled from 0.4 m to -0.4 m. The direction of the initial velocity was determined in the following way: the horizontal angle was sampled between the limiting angles of the starting point to the net posts, the elevation angle was chosen randomly. The spin axis was also sampled randomly, that means topspin, backspin and sidespin were included.

The analysis was particularly aiming at fast shots. Therefore, only balls passing the net within 30 cm height distance were accepted. The absolute values of the translational velocities were limited from 20 to 200 km/h, the spinning velocities from 0 to 150 turns/s (which is equal to 9000 turns/min). These values were determined empirically before as limits for 38 mm balls (Wu 1993). These limits are probably different for other balls sizes and net heights, but in all case studies successful hits were not restricted by the accessible parameter space chosen here. A ball is counted as a successful ball if it passes the net within the height limit and hits the other side of the table tennis table.

Monte Carlo studies using random numbers were done for the 38-mm-ball with a weight of 2.5 g, used in tournaments until end of 2000, the actual 40-mm-ball with 2.7 g and a 44-mm-ball with a weight of 2.3 g, which was tested already in Japan. For the 40-mm-ball an increase of the net height for 1 and 3 cm was analyzed, too.

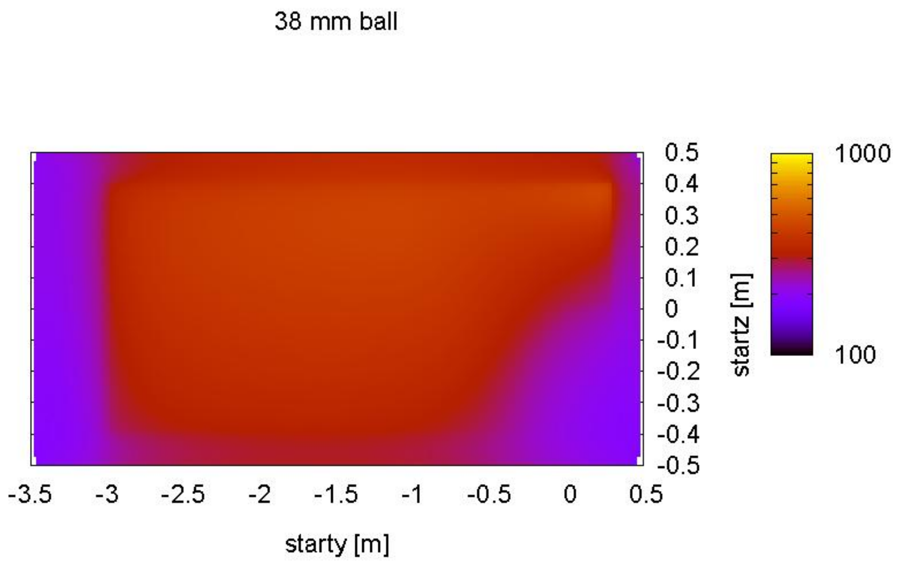
The sampling of such a large number of initial conditions guarantees to cover all possible combinations of initial parameters (positions, translational and spinning velocities) for the different cases creating a successful stroke. Clearly, for different balls and net heights the parameter space of initial conditions leading to successful strokes will be different. The database created in this study allows also an analysis of this effect.

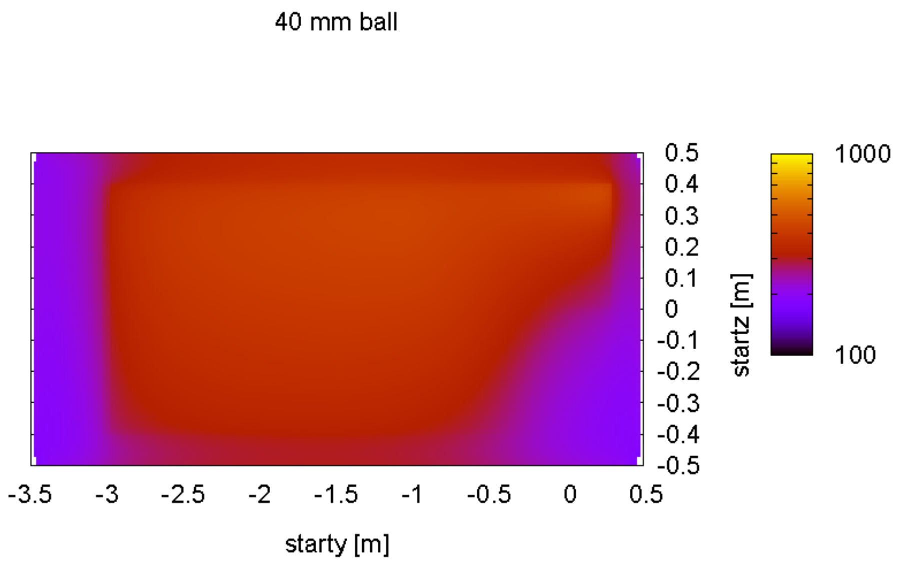
For each case $5*10^8$ initial conditions were sampled and trajectories calculated. Initially this was done on a Linux Cluster with 32 cores. The run-time for each core was 20 hours resulting in a total run time of 640 hours. Alternatively, GPU computing with CUDA was used on a Dell Precision T7500 Desktop with NVIDIA Quadro FX3800. Here, only 3 hours for the same calculation are needed. CUDA (CUDA 2013) is a programming interface to use the parallel architecture of NVIDIA GPUs for general purpose computing. CUDA library functions are provided as extensions of the C language, which allows for convenient and rather natural mapping of algorithms from C to CUDA. A compiler generates executable code for the CUDA device. The CPU identifies a CUDA device as a multi-core coprocessor. For the programmer, CUDA consists of a collection of threads running in parallel. A collection of threads, which is called a block, runs on a multiprocessor at a given time. The blocks form a so-called grid. They divide the common resources, like registers and shared memory, equally among them. All threads of the grid execute a single program called the kernel. All memory available on the device can be accessed using CUDA with no restrictions on its representation. However, the access times vary for different types of memory. Shared and register’s memory are the fastest, as they locate on the multiprocessor (on chip).The shared memory has the lifetime of the block and it is accessible by any thread on the block from which it has been created. This enhancement in the memory model allows programmers to better exploit the parallel power of the GPU for general purpose computing. Additionally, the texture memory which is off-chip allows for faster reading compared to the global memory due to caching.

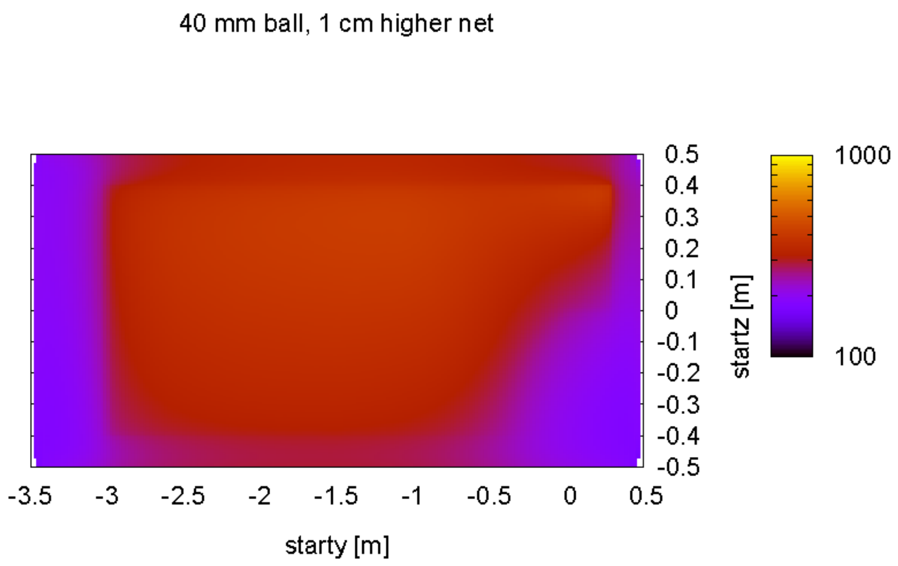
Our implementation consists of two main procedures. First, a predefined number of trajectories are initialized on the CPU side. Thereafter, the ball movements are implemented on the GPU. One step of the equation of motion for the ball’s trajectory, which includes the speed and the position of the ball, is computed in a kernel. The input parameter of the kernel function is the previous trajectory point. The calculations run for a maximal number of iterations. In each iteration step, the updates of the ball’s position and velocity are computed, if the trajectory has not stopped earlier, e. g., when the ball flew beyond the table.

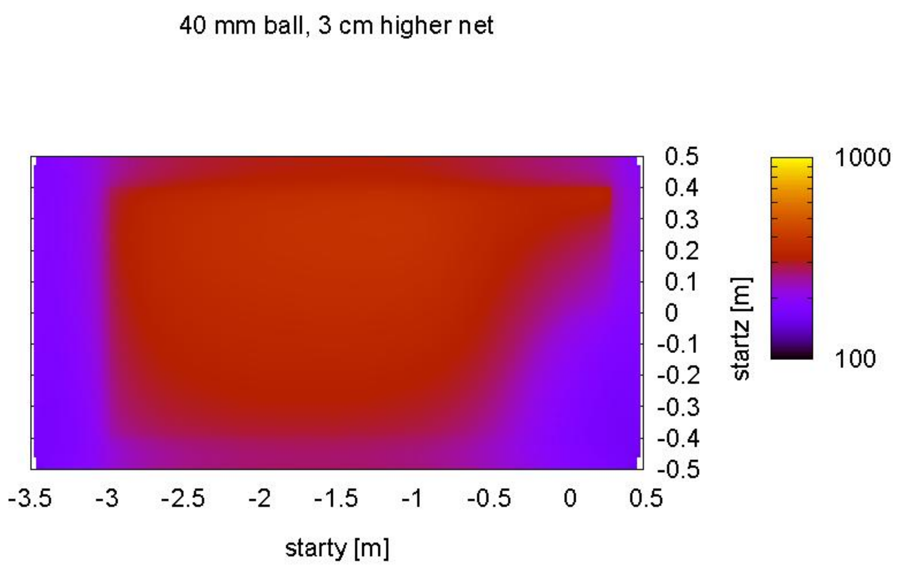
Figure 3 shows as a function of initial position in y and z the number of successful trajectories. The number of successful trajectories from half distance is nearly constant for all balls and net heights. Only for distances below one meter the number of successful strokes decreases continuously, because balls in this region have smaller probabilities hitting the table due to the smaller angle. Balls hit above the table can again reach easier the other side. There is practically no difference for the 38 and 40-mm-ball. Changes of the balls are compensated by other parameter changes. The 44-mm-ball allows more successful strokes even for negative height, because of its lighter weight and its higher air drag. A higher net affects strongly the balls hits above the table limiting there the number of successful trajectories.

In general, the differences between the different cases get more pronounced the higher the hitting point of the balls. A ball hit below the table must have a large spinning to reach the other table side within the height limit. Larger velocities are not possible, because then the balls are not able to reach the other table side and will pass beyond the baseline. Balls hit above the table, even above the height of the net, can be hit with much higher velocities for a successful strike.


38 mm ball
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
startz [m]
1000
100
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
starty [m]


40 mm ball
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
startz [m]
1000
100
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
starty [m]


40 mm ball, 1 cm higher net
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
startz [m]
1000
100
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
starty [m]


40 mm ball, 3 cm higher net
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
startz [m]
1000
100
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
starty [m]

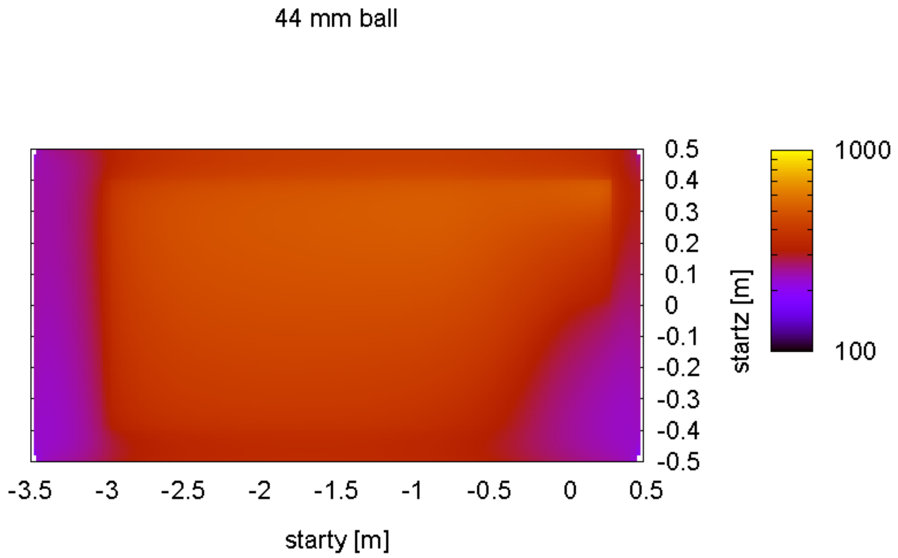


Figure 3: Number of successful trajectories as a function of initial y- and z-conditions

In Figures 4 and 5 the influence of the ball velocity on the distribution functions of the number of successful strokes is shown. Figure 4 shows the dependence on the initial velocity, Figure 5 the dependence on the final velocity. The velocity range used for sampling the initial velocity of 20-200 km/h is identical to 5.6-55.6 m/s.

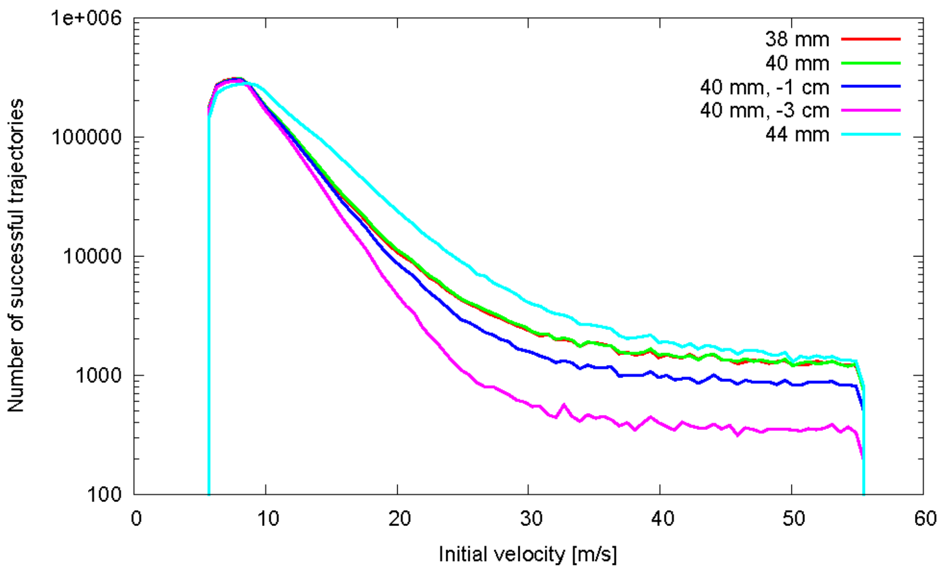


Figure 4: Number of successful trajectories as a function of initial velocity of the balls

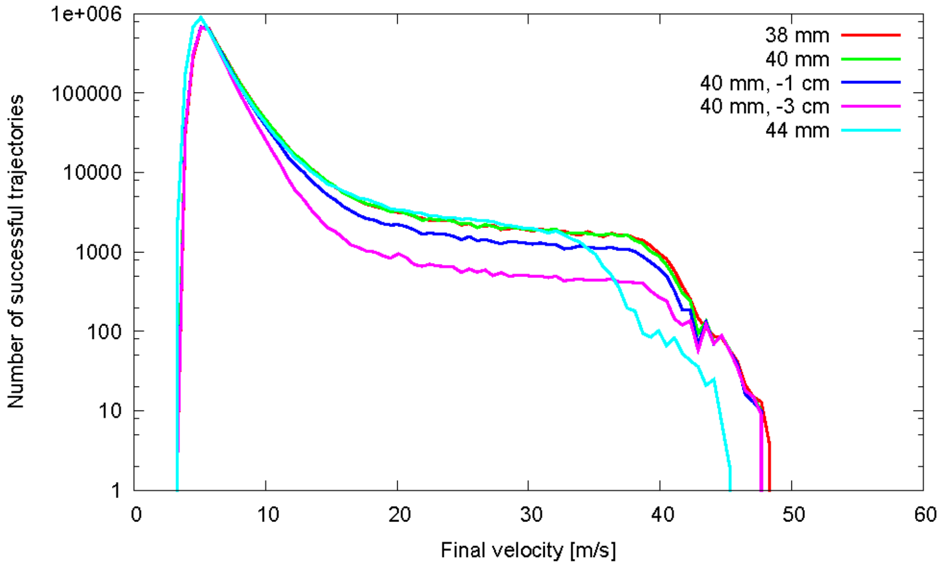


Figure 5: Number of successful trajectories as a function of final velocity of the balls

Again, the results for the 38 and 40 mm ball differ only marginally. For the 44-mm-ball one gets more successful trajectories compared to the 38 and 40-mm-ball for higher initial velocity, the distributions for the final velocities are nevertheless very close again. However, very high velocities above 35 m/s are suppressed earlier for the 44-mm-ball. A stronger influence is visible for the 40-mm-ball increasing the net height. Already for smaller initial and end velocities of about 10 m/s a reduction of successful trajectories shows up being equivalent to a slowing-down of the game. For very low velocities the impact of the air drag is not yet important resulting in larger number of successful trajectories.

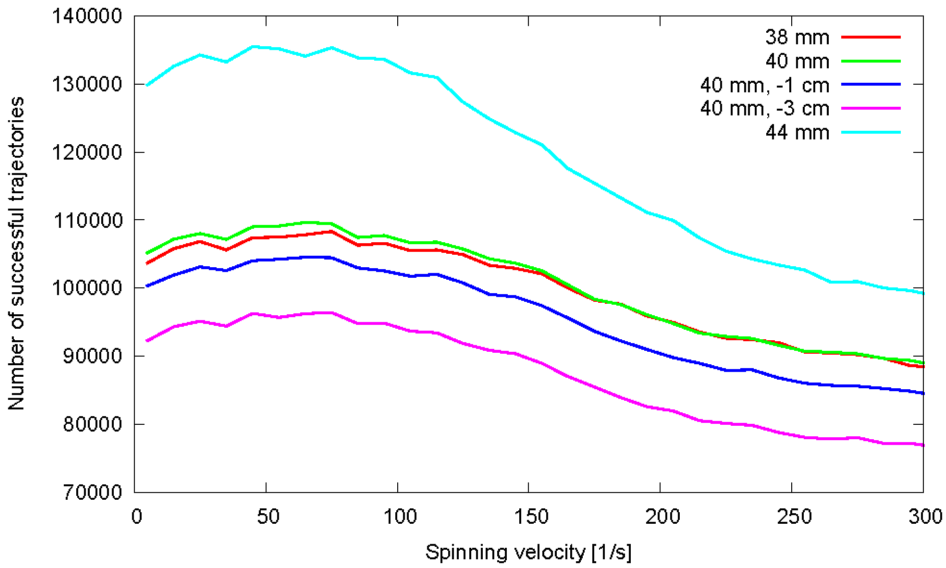


Figure 6: Number of successful trajectories as a function of spinning velocity

Figure 6 demonstrates that the influence of spin is rather weak, because all differences are within 20 percent. The number of successful trajectories is biggest for the 44-mm-ball, followed by nearly identical numbers for the 38 and 40 mm ball and the case with a 1 cm increase of the net height. As expected the highest net gives the smallest number of successful trajectories. The ratio of successful trajectories with strong spinning to those with little spinning is nearly the same in all cases with the exception of the 44-mm-ball. Here, the influence of spinning on the distribution is strongly reduced.

Conclusions

Statistical analysis of the influence of ball size and net height on the number of successful table tennis trajectories using computer modeling is used to quantify the effects on trajectory distribution functions. The analysis confirm the empirical observation that the change of the ball in the year 2000 from a 38-mm to a 40-mm-ball can be compensated such that their resulting trajectory distribution functions are nearly identical. This was achieved in reality by adaptations of the technique and the material. A larger ball of 44 mm with small weight is one option for suppressing high velocities, resulting also in a reduction of the influence of spinning. As an alternative option an increase of the net height is possible. For this, the character of the game will change more strongly, because the possibilities for successful trajectories are reduced limiting technical and tactical alternatives. A small increase of the net height could be one option, where the basic character of the game is not too strongly modified, but reducing especially the influence of the service.

Modifications of basic rules of table tennis like ball size and net height can reduce the maximum velocities, but such modifications will be linked with severe changes in the characteristics of table tennis: dynamics, technique and strategy will change strongly, too. The question is if a possible gain in attractivity of table tennis for TV by such changes is worth the loss of key elements of existing table tennis.

References

Achenbach, E. (1972). Experiments of the flow past spheres at very high Reynolds numbers; *American Journal of Physics*, 54, 565-575.

Bai, K.X. and Hong, X. and Hu, P. and Yin, H. (2005). Technical contrastive analysis after ping-pong diameter altering; *Proceedings of the 9th ITTF Sports Science Congress Shanghai 2005*.

Bearman, P.W. and Harvey, J.K. (1976). Golf Ball Aerodynamics; *Aeronautical Quarterly*, 27, 112-122.

CUDA 2013: Retrieved August, 7, 2013 from URL http://www.nvidia.com/

Djokic, Z. (2007). ITTF scored a goal (changes of rules in table tennis during 2000-2003); *Proceedings of the 10th International Table Tennis Sports Science Congress Zagreb 2007*.

Davies, J.M. (1949). The Aerodynamics of Golf Balls; *J. Applied Physics*, 20, 821-828.

Iimoto, Y. And Yoshida, K. And Yuza, N. (2002). Rebound characteristics of the new table tennis Ball; Differences between the 40 mm (2.7g) and 38 mm (2.5g) balls; *International Journal of Table Tennis Sciences No. 5, Proceedings of the 7th ITTF Sports Science Congress Osaka 2001*.

Kondric, M. and Medved, V. and Baca, A. and Kasovic, M. and Furjan-Mandic, G. and Slatinsek, U. (2007). Kinematic analysis of top spin stroke with balls of two different sizes; *Proceedings of the 10th International Table Tennis Sports Science Congress Zagreb 2007*.

Li, J.L. and Zhao, X. and Zhang, C.H. (2005). Changes and development: influence of new rules on table tennis techniques; *Proceedings of the 9th ITTF Sports Science Congress Shanghai 2005*.

Liu, Y.X. (2005). Comparative analysis and research of the impacts by 40 mm ball on the first-3-stroke skills of shake-hand looping style of world-class male table tennis players; *Proceedings of the 9th ITTF Sports Science Congress Shanghai 2005*.

Maccoll, J. (1928). Aerodynamics of a spinning sphere; *J.R. Aeronaut. Soc*., 32, 777.

Mehta, R.D. (1985). Aerodynamics of Sports Balls; *Annual Reviews of Fluid Mechanics*, 17, 151-189.

Nelson, R. (1997). Es geht um die Zukunft unseres Sports. *Deutscher Tischtennis-Sport*, *10,* 27. Münster: Philippka-Verlag.

Wu, H. (1993). Analysis of the training for the Chinese table tennis superiority from 1959 to 1989. *Sport Science*, *3,* 48-58. Beijing: The People’s Sport Publishing House.

Validation of the ActiSmile Physical Activity Feedback Device in the Activity Assessment

Nicole Ruch[1], Johanna Hänggi[2], Stephanie Zurbuchen[3] & Urs Mäder[1]

[1]Swiss Federal Institute of Sports, Magglingen, Switzerland

[2]University of Applied Sciences and Arts Northwestern Switzerland, School for Teacher Education, Brugg, Switzerland

[3]ETH Zürich, Switzerland

Abstract

The aim of this study was to investigate the validity of the ActiSmile (AS), a physical activity feedback device. Twenty-five participants (age: 36.5 ± 14.3 y) performed nine different activities including standing, running, cycling, playing badminton, sweeping the floor with a broom, and several types of walking for 3 min 25 s each. Each participant received two AS devices, one attached at the hip (AShip) and one carried in the trouser pocket (ASpocket). The AS classifies each min as inactivity, walking, or running. Recognized activity time was compared to observed activity. Indirect calorimetry was used to validate the energy expenditure (EE) estimated by the AS. The AShip correctly recognized 100% of standing as inactivity, between 97 and 100% of different walking modes and 97% of running. The ASpocket correctly recognized 100% of standing as inactivity, between 88% and 96% of different walking modes, and 96% of running. Good estimates of EE were reached for moderate walking (bias: < 1.1 kcal/min). The AS is therefore valid for recognizing both walking at individual intensities and running. It provides good group estimates of the EE during moderate walking. Therefore, the AS has the accuracy to be an effective tool for intervention studies.

KEYWORDS: VALIDITY, ACCELEROMETER, PHYSICAL ACTIVITY

Introduction

Physical activity (PA) is associated with a sustained energy balance and with improved cardiorespiratory, metabolic, musculoskeletal, and mental health (U.K. Department of Health, Physical Activity, Health Improvement and Protection, 2011; Haskell et al., 2007; Nelson et al., 2007; O'Donovan et al., 2010; U.S. Department of Health and Human Services, 2008). The accurate assessment of free-living PA has always been a challenge in studies that aim to identify the dose-response relationship between PA and health outcomes, describe the changes of PA levels in populations, investigate the effects of PA interventions and monitor adherence to PA guidelines. Numerous methods have been used to measure PA for these purposes. Generally, methods to assess PA fall into four different classes: subjective reports and observations, indirect calorimetry, double-labeled water, and portable monitors. The field of applications of these methods varies greatly and includes epidemiological research, intervention studies, clinical practice, and personal assessment (Chen et al., 2012). Doubly labelled water is an expensive method that is usually used for long- term investigations in small samples (Plasqui & Westerterp, 2007) and indirect calorimetry is not applicable for long-term measurements due to its burdensome setup. Self-reports and observations are influenced by subjective opinion and by social desirability (Adams et al., 2005). Therefore, the application of small, portable PA monitors has increased in recent years (Freedson, Bowles, Troiano & Haskell, 2012), as they allow a convenient, inexpensive, but objective assessment of PA. According to Chen et al. (2012), current PA monitors can be assigned to three different categories, movement sensors (e.g. Pedometers, inclinometers, gyroscopes, goniometers and accelerometers), physiological sensors (e.g. heart rate monitors, temperature and heat flux sensors) and contextual sensors (e.g. global and local contextual sensors, pressure sensors, passive infrared sensors). Integrated multisensory systems combine sensors of different categories described above (van Reemortel et al., 2012). Pedometers are monitors that estimate the number of steps taken based on mechanical or digital measurements of only the vertical plane. Ankle and hip-worn pedometers have been validated in measuring the number of steps (Foster et al., 2005; Le Masurier & Tudor-Locke, 2003; Oliver, Schofield, Kolt & Schluter, 2007; Tudor-Locke et al., 2004; Tudor-Locke, 2002), but they were reported to not accurately assessing distance or energy expenditure (EE) (Crouter, Schneider, Karabulut & Bassett, 2003; Foster et al., 2005). Moreover, pedometers cannot distinguish between steps performed during different types of activity such as running or walking. Uni-, bi- or triaxial accelerometers have been widely used to assess PA in children (Andersen et al., 2006; Bringolf-Isler et al., 2009; Ekelund et al., 2004) and adults (Troiano et al., 2008; Colley et al., 2011). They are an accepted method to measure the intensity, duration and frequency of activities in large populations (Bassett et al., 2012; Welk et al., 2012). Today, devices from different manufacturers are available and have been validated against a criterion measurement such as indirect calorimetry or doubly labelled water (Welk et al., 2012; van Remoortel et al., 2012). Integrated multisensory systems that measure the combination of heart rate and accelerometers have been increasingly used and been reported to precisely estimate the EE (Assah, Ekelund, Brage, Wright, Mbanya & Wareham, 2011; Brage et al., 2004; Brage, Franks, Ekelund & Wareham, 2005; Ojiambo et al., 2012). However, the measurement of heart rate might be burdensome for the user in long-term measurements, as the device usually has to be worn on the skin. Devices that include accelerometers and various other sensors are mainly used to identify PA types (Aminian & Najafi, 2004; Nawab, Roy & De Luca, 2004); Pärkkä et al., 2006). However, these setups have to be attached to different sites of the body (Aminian et al., 2004; Nawab et al., 2004; Pärkkä et al., 2006), which is burdensome for the user and therefore not applicable for long-term PA measurements. Accelerometers (Hendriksen, Lund, Moe-Nilssen, Biddal & Danneskiod-Samsoe, 2004) or their combination with other sensors (Bamberg, Benbasat, Scarborough, Krebs & Paradiso, 2008; Mayagoita, Nene &Veltink, 2002) have also been used for gait analysis. Although all of the mentioned devices have been reported to measure different aspects of PA accurately, they were developed for research purposes and were not designed to give simple feedback to the everyday user. On the contrary, these devices have been designed to be a black box for the user in order not to influence his usual PA or gait behavior. Therefore, accelerometers alone or combinations of accelerometers with other sensors that are designed for the use in research are not applicable for interventions as they provide no feedback to the user and store raw data that has to be further processed by the researcher, before the information on PA is accessible.

In contrast to accelerometer based devices, heart rate monitors have been used as a user feedback for the control of the personal training or the EE in the fitness sector for many years (Achten & Jeukendrup, 2003; Laukkanen & Virtanen, 1998). Recently, portable feedback monitors have been developed that monitor the daily PA of the inactive to moderately active user. Compared to PA monitors that are currently used in research, these devices are not sealed but provide real-time monitoring, feedback on activities performed and access to established measurement outputs (time spent in activitiy, steps, distance, EE). These monitors are developed for personal use and the everyday application. They could be effective PA intervention tools to increase the user's overall activity level since they meet several criteria identified to be effective in changing PA behavior. These criteria include providing a feedback on PA performance, prompting a specific PA goal, and setting graded tasks (Abraham & Michie, 2008; Michie, Johnston, Francis, Hardeman, & Eclles, 2008; Michie, Ashford, Sniehotta, Dombrowski, Bishop, & French, 2011). Some of these factors were previously mentioned in pedometer studies, which also emphasize the importance of keeping an activity diary (Bravata et al., 2007; Clemes & Parker, 2009; De Cocker, De Bourdeaudhuij, Brown, & Cardon, 2008; McKay et al., 2009; Rooney, Smalley, Larson, & Havens, 2003). However, only a few of accelerometer-based feedback monitors for the home use were validated so far.

Products that were released recently include the Nike fuel band, the Fitbit, the Personal Activity Monitor (PAM), the Polar Activity Watch 200 and the Polar FA20 activity watch, and the ActiSmile (Brugniaux et al., 2008; Kinnunen, Tanskanen, Kyröläinen, & Westerterp, 2012; Slootmaker, Chin A Paw, Schuit, van Mechelen & Koppes, 2009; Krauss, Solà, Renevey, Maeder, & Buchholz, 2008). The EE estimation of the PAM has been validated for different speeds of treadmill and stair walking (Slootmaker et al., 2009). However, treadmill is a specific setting not frequently performed in daily life. As accelerations on the treadmill may not be similar to free-living walking, the application of the PAM in free-living conditions has to be further investigated. In contrast, the EE estimated by the Polar Activity Watch 200 was analysed during long continuous hiking periods, which let to high correlations with measured values (r = 0.92-0.96) (Brugniaux et al., 2008). The EE estimation of the Polar FA20 has been reported to not significantly differ from the measured EE by doubly labelled water over the week (Kinnunen et al., 2012). The validation of the total EE over a week is very useful, however, might not reveal differences of the estimated versus measured EE during single activities, a fact that might be recognized by the everyday user. A device that especially fulfils the theory-linked techniques to increase PA in the user is the ActiSmile (AS) (ActiSmile, Kiesen, Switzerland), a portable feedback device that was developed to determine three types of activity and estimate EE (Krauss et al., 2008). The AS assesses inactivity, walking, and running. Furthermore, it gives feedback for a user-specific activity goal for graded, continuous activity bouts of five, ten, fifteen or twenty min depending on the fitness level of the user. The feedback is provided in the form of an icon resembling a smiling face with a growing smile for each bout of activity achieved. Compared to other personal monitoring devices, the AS is relatively low in cost and provides information on intensity and activity type, which has been mentioned as important in recent PA guidelines (Oja et al., 2010, Strong et al., 2005). Its feedback in the form of a smiley face, which is independent of age, language, or culture, makes it an especially interesting device for PA interventions. With the AS software, the accumulated activity can be recorded in an activity diary. By displaying the performed PA in a simple way, by enabling the user to set his activity goal graded to his fitness level, and by offering the opportunity to keep an activity diary, the AS seems to meet the conditions to be an effective intervention tool. As precise measurements are crucial in ensuring the credibility of the AS, the aim of the present study was to validate the AS for its classification of daily activities as inactive, walking, or running; for its EE estimates; and for its assessment of continuous bouts of activity.

Methods

Participants

Participants were recruited through personal contacts from the local area. The target population was adults age 20 to 65 with sedentary lifestyles, which is in accordance with the target population of the AS. Each participant was contacted by phone before the start of the study to assess if these requirements were met. Twelve women (age: 34.7 ± 13.5 y, height: 167.9 ± 5.8 cm, weight: 63.3 ± 7.3 kg, BMI: 22.5 ± 2.5) and thirteen men (age: 37.5 ± 15.8 y, height: 178.6 ± 7.0 cm, weight: 79.2 ± 9.0 kg; BMI: 24.8 ± 2.3) participated in the study. The local ethical committee of the canton Berne approved the study. All participants signed a written informed consent.

Measurement procedures

After the participants arrived at the laboratory, their weight and height were measured. The participants were then provided with two AS devices and a portable indirect calorimeter. The devices were worn during the entire measurement procedure. Nine different activities were performed in random order. Standing, moderate and fast walking, moderate walking uphill and downhill, and running were included since the AS was originally developed to recognize such activities (Krauss et al., 2008). All these activities were performed outdoors on a running track. Walking uphill and downhill was performed on a road with an approximate 8% incline. Other activities, such as cycling on a bike on a country lane, playing noncompetitive badminton without a net, and sweeping the floor (approx. 10x10m) with a broom, were included in the study to show how the device classifies activities other than locomotor activities. Participants were instructed to perform all activities at their own moderate pace giving them standardized instructions such as "walk in as speed as if you were walking to your work" for moderate walking and "walk as if you were late on your way to work" for fast walking to allow a natural variety in the speed as it would occur in field conditions. Instructions for all other activities were the similar to that for moderate walking. All activities were performed for 3 min and 25 s. The AS records blocks of 60 s and only the total number of minutes classified in each of the activity classes can be downloaded from the device. We therefore decided to give the device the tolerance time of 25 s to ensure that there was no influence of the start or end of the activity on the activity recognition or the EE estimation. The chosen activity length also allowed the measured EE to reach steady state conditions (Pearce & Milhorn, 1977; Whipp, Ward, Lamarra, Davis, & Wasserman, 1982). An additional continuous moderate walking bout was performed to validate the confirmative feedback that the AS shows the user after 5 min of continuous activity. Between all activities, the participants were given a break until their EE decreased to resting EE. The start and end of the activities were recorded in a protocol.

Instruments

Weight was measured to the nearest 0.1 kg using a standardized digital scale (Model 861, seca GmbH & Co., Hamburg, Germany). A stadiometer (Modell 213, seca GmbH & Co., Hamburg, Germany) was used to measure height to the nearest 0.5 cm. Participants were asked to remove their shoes and jackets and to empty their trouser pockets for both measures.

Two AS devices (4.0 cm x 4.0 cm x 2.0 cm, 20 g) (ActiSmile, Kiesen, Switzerland) were used, one mounted on the hip (AShip) and one carried in the trouser pocket (ASpocket) since the manufacturer suggests these measurement sites. During cycling, an additional AS was placed on the ankle of the participant (ASankle) by inserting it into the sock, as suggested by the manufacturer. The AS contains a three-axial accelerometer that measures at a range of ±2g and at a sampling frequency of 25 Hz with 12-bit precision. It assigns each time frame of 5 s to inactivity, walking, or running activities. Whenever a registered minute includes at least 11 time frames of 5 s walking or running, the current minute is assigned to the corresponding activity. Otherwise, a minute is classified as inactivity. The detailed specifications and the classification algorithms of the device have been described elsewhere (Krauss, et al., 2008). The AS was personalized with the associated software (ActiSmile, Version 3.5.8, ActiSmile AG, Kiesen, Switzerland) to create a personal profile for each user (in terms of body weight, height, sex, and activity level). The activity level (beginner, standard, advanced, or sports level) determines how much active time the user has to accumulate to obtain feedback on the monitor of the AS. Feedback is given in the form of a smiley face with a smile divided in three stages (Figure 1). For a beginner, stage 1 is completed after 5 min of continuous moderate walking or running, while stages 2 and 3 are completed after two further 5-min activity periods. In the standard, advanced, and sports levels, a smiling face signaling completion of stage 3 appears after 3 continuous 10-, 15-, or 20-min blocks of walking or running. In the present study, activity time during the continuous walking bout was measured until the smiley face appeared at the end of stage 1 for a beginner. After each activity the AS was connected to a laptop so that the number of minutes recognized as inactivity, walking, or running as well as estimates of the activity's EE could be calculated based on the user profile and displayed by the AS software.

Figure 1: The four stages of ActiSmile feedback.

EE was measured with the help of a light-weight (570 g) mobile indirect calorimetry (IC) device (MetaMax 3B, Cortex, Leipzig, Germany). A two-point calibration containing ambient and mixed gas was performed according to the manufacturer's guidelines before each test. The data storage on the mobile device ensures a wide range of possible actions and a high degree of flexibility. The validity and reliability of this device have been reported to be adequate (Vogler, Rice, & Gore, 2010). Breath-by-breath VO2 and VCO2 of the last minute of the EE measurement of each activity was analyzed using the formula of Elia and Livesey, (1992).

Data Analysis

All values were given as mean and standard deviations unless otherwise stated. The time intervals recorded for inactivity, walking, or running were given as proportions of the observed time during which an activity was performed. A Fisher test was used to compare the proportions of recognized activities by the AShip and the ASpocket, respectively, to the observed activity time during standing, moderate and fast walking, and walking uphill and downhill. The same procedure was used to compare activity time recognized by the different AS devices for biking, badminton, and sweeping the floor. The time measured until the stage 1 smiley icon appeared for a beginner was compared to the expected 5-min level by way of a Wilcoxon-signed-rank-test as a quantile comparison plot revealed that the data were not normally distributed. The agreement between the EE assessed by IC and the AShip and the ASpocket was determined using the Bland and Altman limits of agreement analysis (Bland & Altman, 1999) and %-bias from the mean measured values. A previous validation study reported 30-67% biases in the EE during treadmill and stair walking in the PAM (Slootmaker et al., 2009). In the present study, the bias will be considered to be accurate when <10%.

Coefficient of determinations for the estimated versus measured EE were r2 = 0.74-0.93 (which is equal to r = 0.86-0.96) (Slootmaker et al., 2008). These values were comparable to a study that investigated the EE during long, continuous hiking bouts estimated with the Polar Activity watch 200 (r = 0.92-0.96) (Brugniaux et al., 2008). When the estimated EE by the Polar FA20 and measured EE by doubly labelled water during a week were compared, correlation coefficients of r = 0.80-0.86 were found (Kinnunen et al., 2012), which the authors considered as a strong prediction. To assess the linear relationship between the measured EE by IC and the EE estimated by the AS, Spearman correlations were calculated and in line with previous literature considered as strong when $r \geq 0.80$ (Brugniaux et al., 2008; Kinnunen et al., 2012). They were considered as adequate when $r \geq 0.50$ which is in line with Cohen (1988) that considered $r \geq$ as large. The main precision in the estimated compared to the measured EE was determined by a Wilcoxon- signed rank test with Bonferroni corrections for multiple comparisons over the different activities, All statistical analyses were performed on R (R Project for Statistical Computing, Version 2.14.0, Bell Laboratories, Murray Hill, NJ, USA).

Results

The AShip and the ASpocket were accurate in measuring inactivity as they classified 100% of the observed time spent in standing as inactivity (Table 1). Recognition of moderate (100%) and fast walking (99%) did not differ significantly from the observed activity in the AShip. With the ASpocket, recognition of moderate walking (95%) did not differ from the observed activity. However, the activity recognized as walking during fast walking was significantly different from observed walking time ($p < 0.05$). There was no significant difference between uphill and downhill walking (100%) recognized by the AShip and the observed activity time. With the ASpocket, walking uphill (100%) did not differ from the observed activity time, but walking downhill did differ significantly (88%) ($p < 0.05$). Running was recognized correctly by the AShip (97%) and the ASpocket (96%), respectively. Recognition of cycling was significantly different when the AShip was compared to the ASpocket ($p < 0.05$) and the ASankle ($p < 0.05$), respectively, with the AShip classifying more time as inactivity. No significant difference between the devices was found for playing badminton. Sweeping the floor was significantly different between the AShip and the ASpocket ($p < 0.05$). Stage 1 (Figure 1) of the AS's smiley face icon was displayed, on average, after 5 min 3.5 s ± 17.7 s and was not significantly different from the 5-min target.

Table 1: Time the ActiSmile classified as no activity, walking, or running during different activities of adults (n = 25).

		Activity Classified by the ActiSmile		
Performed Activity	**Device**	**Inactivity (%)**	**Walking (%)**	**Running (%)**
Standing	AS_{hip}	100	0	0
	AS_{pocket}	100	0	0
Moderate walking	AS_{hip}	0	100	0
	AS_{pocket}	0	95	5
Fast walking	AS_{hip}	0	99	1
	AS_{pocket}[a]	1	92	7
Walking downhill	AS_{hip}	0	100	0
	AS_{pocket}[a]	4	88	8
Walking uphill	AS_{hip}	0	100	0
	AS_{pocket}	3	96	1
Running	AS_{hip}	1	2	97
	AS_{pocket}	0	4	96
Cycling	AS_{hip}	73	21	6
	AS_{pocket}[b]	20	78	2
	AS_{ankle}[b]	3	93	4
Playing badminton	AS_{hip}	0	100	0
	AS_{pocket}[b]	3	91	6
Sweeping the floor	AS_{hip}	84	16	0
	AS_{pocket}[b]	32	68	0

AS_{hip} = ActiSmile worn on the hip; AS_{pocket} = ActiSmile carried in trouser pocket; AS_{ankle} = ActiSmile worn on the ankle.
[a]Significantly different from observed activity time ($p < 0.05$).
[b]Significantly different from AS_{hip} ($p < 0.05$).

Between the measured and estimated EE from both AS devices, Bland and Altman limits of agreement analysis revealed a mean bias close to zero, narrow limits of agreement, and equally distributed variances in the AShip and ASpocket for moderate walking (Table 2). The EE estimation of the two devices during this activity was not significantly different from measured values, whereas the EE estimation during all other activities was significantly different ($p < 0.05$). It was the only activity, where the bias in the AShip (%-bias: 0.02%) and ASpocket (%-bias: 0.0%) was smaller than 10% and therefore the estimation was considered as accurate. For standing (AShip and ASpocket: -35%), fast walking (AShip: -31%, ASpocket: -29%), walking downhill (AShip: 27% and ASpocket: 22%), and running (-33% AShip: -33%, ASpocket: -32%), the bias was around one third of the measured values. For walking uphill (AShip and ASpocket: -45%), playing badminton (AShip: -42%, ASpocket: -43%), and sweeping (AShip: -55%, ASpocket: -45%), the bias was large. For biking, the bias was smallest when the device was attached to the ankle (%-bias: -30%). When the device was attached to the hip, it was largest (%-bias: -73%). Correlations were strong during standing in both devices. Correlations were adequate during fast walking in both devices, and during walking uphill, moderate walking and badminton in the AShip (Table 2). For cycling, the EEs with the AShip, ASpocket and ASankle (EE: 4.9 ± 2.0 kcal/min; difference to IC: -2.1 kcal/min; 95% CI: (-2.9, -0.8); $r = 0.209$; $p = 0.473$) were weakly correlated with the measurements of the IC. Likewise, correlation coefficients of the remaining activities were low. Correlation coefficient over all activities were $r = 0.58$ for the AShip and $r = 0.59$ for the ASpocket.

Discussion

This study found high validity in the AS, a physical activity feedback device used to detect standing and walking on different inclines and at varying intensities and to discriminate them from running. Furthermore, the AS gives accurate feedback on continuous walking or running bouts as well as a good estimate of the EE in moderate walking. Therefore, the AS has the potential to be an effective tool for future intervention studies.

The AS detected correctly inactive behavior by classifying standing as inactivity. The correct detection of inactivity is important because it is closely associated with health problems (Blair, 2009). However, when the device was attached to the hip or carried in the trouser pocket, other activities such as cycling were classified as inactivity although their intensity was high. Therefore, the AS might assign too much time to inactivity over the day depending on the activity types performed. The AS was able to recognize walking at individual intensities and on different inclines. However, for fast walking and downhill walking, we recommend wearing the AS on the hip since carrying the device in the trouser pocket during these activities leads to misclassifications. Nonetheless, the AS recognized most walking correctly, which is considered the most important activity to assess in typically sedentary populations (Masse et al., 1998) because it can provide essential health benefits (Morris & Hardman, 1997; Murphy, Nevill, Murtagh, & Holder, 2007). The AS successfully distinguished walking activities from running, a new feature in comparison to common pedometers, which cannot distinguish between different types of activities (Tudor-Locke & Myers, 2001). Correctly distinguishing between walking and running may serve as an important motivating factor for the user since the reliability of the device is increased. Given that the new PA recommendations focus more on the contribution of intensive activities to health benefits (Oja et al., 2010), the distinction between walking and running might be important for PA interventions that give participants precise PA guidelines concerning moderate and vigorous activities. In addition, recognizing the type of activity is important for PA recommendations in regard to certain health factors. For example, for bone health, PA guidelines recommend that children engage in high-impact activities such as running three times a week because of their bone strengthening effects (World Health Organization, 2010).

This study determined not only the AS's capability to recognize walking and running but also investigated the AS's classification of other activities regularly performed in daily life. This constitutes important information for the user in regard to the credibility of the device. As an example, cycling was recognized as walking by the ASankle and ASpocket, whereas the AShip classified most of the cycling time as inactivity. Everyday cycling was moderate in this study (5.2 MET, Table 2), which is comparable to the findings in the related literature (Ainsworth et al., 2000). Therefore, it is suggested that the activity be identified with regard to its intensity when the AS is carried in the pocket or in the sock. These results might be due to the higher activity generated in the lower extremities during cycling compared to the more stable hip position evident during this activity, or they may result from the device's loose position in the pocket, where it accumulates more acceleration than in a stable position on the hip. Similar

Table 2: EE measured by IC and estimated by the A_{hip} and the AS_{pocket} (n = 25).

Activity	IC (kcal/min)	IC (MET)	AS_{hip} (kcal/min)	AS_{pocket} (kcal/min)	Bias AS_{hip} (95% CI) (kcal/min)	Bias AS_{pocket} (95% CI) (kcal/min)	$r_{IC, AShip}$ (p-value)	$r_{IC,ASpocket}$ (p-value)
Standing	1.7 ± 0.4	1.3 ± 0.2	1.1 ± 0.2*	1.1 ± 0.2*	-0.6 (-0.9, 0.17)	-0.6 (-0.7, -0.4)	0.849 (< 0.001)	0.849 (< 0.001)
Moderate walking	4.7 ± 1.0	3.7 ± 0.6	4.6 ± 0.8	4.6 ± 1.0	-0.1 (0.3, -0.5)	-0.0 (0.4, -0.5)	0.586 (0.002)	0.464 (0.020)
Fast walking	6.8 ± 1.4	5.4 ± 0.8	4.7 ± 0.9*	4.9 ± 0.9*	-2.1 (-1.7, -2.6)	-2.0 (-1.6, -2.5)	0.621 (0.001)	0.576 (0.004)
Walking downhill	3.7 ± 1.1	2.5 ± 0.4	4.7 ± 1.3*	4.5 ± 0.8*	1.0 (1.6, 0.5)	0.8 (1.3, 0.3)	0.444 (0.026)	0.275 (0.183)
Walking uphill	8.3 ± 2.1	5.9 ± 0.8	4.6 ± 0.7*	4.6 ± 1.6*	-3.7 (-3.0, -4.4)	-3.7 (-2.8, -4.6)	0.642 (0.001)	0.332 (0.105)
Running	11.3 ± 3.1	9.4 ± 2.2	7.6 ± 1.4*	7.7 ± 1.6*	-3.7 (-2.5, -4.8)	-3.6 (-2.5, -4.7)	0.466 (0.022)	0.481 (0.017)
Cycling	7.0 ± 1.6	5.2 ± 1.3	1.9 ± 1.1*	3.7 ± 1.8*	-5.1 (-7.5, -2.7)	-3.3 (-2.3, -5.5)	0.207 (0.542)	0.295 (0.153)
Badminton	8.3 ± 2.4	6.3 ± 1.5	4.8 ± 1.0*	4.9 ± 1.7*	-3.6 (-2.8, -4.4)	-3.5 (-2.5, -4.4)	0.557 (0.004)	0.428 (0.033)
Sweeping	4.7 ± 1.5	3.6 ± 0.9	2.2 ± 1.9*	2.7 ± 2.5*	-2.6 (-1.7, -3.4)	-2.1 (-0.9, -3.3)	0.299 (0.146)	0.021 (0.920)

EE = energy expenditure; AS_{hip} = ActiSmile worn on the hip; AS_{pocket} = ActiSmile carried in trouser pocket. IC = indirect calorimetry. Bias AS_{hip} = AS_{hip}-IC. Bias AS_{pocket} = AS_{pocket} -IC. $r_{IC, AShip}$ = Spearman correlation coefficient between EE measured by IC and by AS_{hip}. $r_{IC, ASpocket}$ = Spearman correlation coefficient between EE measured by IC and by AS_{pocket}. Sweeping = sweeping the floor with a broom. * = significantly different from measured EE (p < 0.05).

results were found for sweeping the floor, an activity that was performed at moderate intensity (4.7 MET), comparable to values in the literature (Ainsworth, et al., 2000), and was mainly recognized as walking by the ASpocket. By contrast, the intensity of playing badminton was higher in our study (6.3 ± 1.5 MET) than in the literature (4.5 MET) (Ainsworth, et al., 2000) and was mainly classified into the walking activity, which is of moderate intensity (3 MET) according to the literature (Ainsworth, et al., 2000). Therefore playing badminton was not classified correctly according to its intensity. Furthermore, the measured intensity of sweeping was within moderate intensity (3.6 MET) and therefore comparable to the values for mopping in literature of (3.5 MET) (Ainsworth, et al., 2000). Sweeping was mainly classified into inactivity by the AShip whereas in the ASpocket, sweeping was mainly assigned to walking, which is the activity that is in compliance with the moderate intensity. Therefore AS might provide more plausible results for sweeping and cycling when it is carried in the pocket or in the sock (during cycling).

The AS gave feedback after the target time and is therefore a valid device for providing confirmative feedback to the user about his achievement of continuous PA bouts. The recognition of health-related activity bouts is crucial as they have been suggested to be important in previous PA recommendations (U.K. Department of Health, Physical Activity, Health Improvement and Protection, 2011; Haskell et al., 2007; Nelson et al., 2007; O'Donovan et al., 2010; U.S. Department of Health and Human Services, 2008). In newer PA recommendations the emphasis is on total weekly activity (Oja et al., 2010). However, the authors point out that the activity should be accumulated in at least in three weekly sessions to avoid overly large doses. Giving feedback about the length of time spent walking or runnning might help the user to accumulate activity time in reasonable amounts. Since the activity bout length can be graded according to the fitness level of the user, the AS has the potential to be a powerful tool for intervention studies (Abraham et al., 2008; Michie et al., 2008, 2011).

The EE estimated by the AS was not significantly different from the measured values for a moderate level of walking, independent of where the device was worn (pocket or hip). The small bias found in the EE during walking (0-0.2%) indicated a high level of accuracy and was smaller than the lowest bias found in the different models applied to the data of the PAM (30%) (Slootmaker et al., 2009). The correlation coefficients between EE and IC were adequate for moderate walking in the AShip but not in the ASpocket. Both coefficients were lower compared to the correlation coefficients between EE estimated by pedometers and EE measured by reference methods ($r = 0.68$) (Tudor-Locke & Myers, 2001) and they were lower than the correlations found in the estimations of the Polar FA20 for the daily EE with measured values ($r = 0.80$-0.86) (Kinnunen et al., 2012). However, the analysis of daily EE estimation might mask the differences in single activities. Given the low bias but the low correlation in comparison to other studies, the EE estimation of moderate walking can therefore be considered as accurate on a group level but not for individual estimations. The strong correlation between measured and estimated EE during standing might be due to the AS's estimation of resting EE by individual factors, which is a well-established method (Mifflin et al., 1990), however, the large bias and the significant difference in EE estimation for standing reveal that the AS does provide an estimation of the intensity of this activity performed by individuals but with a systematic bias. Likewise, the EE of fast walking was significantly different from the measured values and the bias was over 10%, however; the correlation coefficient was adequate in both devices. Therefore, the EE estimated by the AS during fast walking gives an estimate of the intensity on an individual level but with a systematic bias. Walking downhill was the only activity where EE was significantly overestimated by the AS. This might be due to its development for moderate walking and the lower EE during walking downhill. During walking downhill and running, biases were above 10%, and correlations were low, indicating that the feedback of the AS on EE during walking downhill and running should be interpreted carefully. The high relative bias during walking uphill is explainable as accelerations were lower during that activity concomitantly accompanied by a high EE. Nevertheless, correlation was adequate when the AS was worn on the hip, indicating that the device replicates the intensity of this activity performed by the individual but with a systematic bias. Cycling showed the best agreement between the EE assessed by IC and the AS when the AS was worn at the ankle. However, for all other cycling sites and during playing badminton, the devices underestimated EE with a large systematic bias. The correlation coefficients of badminton were adequate in the AShip. Therefore, the AShip gives an indication of the intensity of playing badminton in individuals but estimates the EE with a systematic bias. Biases were high and correlation coefficients were low for sweeping, indicating that estimates of EE by the AS are not valid for such an unstructured household activity. Therefore, the AS is valid in the prediction of EE for moderate walking when the device is worn on the hip and it provides accurate estimates on a group level when the AS is worn in the trouser pocket. The AS' EE feedback for standing, fast walking (AShip and ASpocket), walking uphill (AShip) and playing badminton (AShip) is also closely related to measured values, which provides users with information on the general intensity level of the performed activity but the estimation is always adhered to an estimation error.

The correlation coefficient between measured and estimated EE by the AS over all activities ($r = 0.58$ for the AShip and $r = 0.59$ for the ASpocket) was adequate but generally lower as in studies relating tri-axial accelerometer output to EE which were based on laboratory assessments of daily living activities. Correlation coefficients in these studies ranged from $r = 0.70$ to 0.95 (Van Remoortel et al., 2012), except for one study by Campbell et al. (2002) that showed a correlation coefficient of $r = 0.48$. In contrast to these PA monitors, which are mainly used by researchers, the AS was developed for personal use and does not require raw acceleration processing but comes along with an inbuilt algorithm that provides direct estimates of EE. The lower storage capacity and simpler algorithms due to the trade-off with long battery life and directly accessible feedback might explain the lower correlation coefficients between measured and estimated EE in the AS compared to PA monitors applied in research. The Polar Activity Watch 200 and the PAM, which were also developed for personal use, revealed stronger correlation coefficients for estimated versus measured EE (Polar Activity Watch: $r = 0.92$-0.96; PAM: $r2 = 0.74$-0.93, which is $r = 0.86$-0.96). However these studies did only include locomotor activities such as hiking, walking and stair walking and did not include other daily activities as in our study. This might explain the stronger correlation found in their studies (Brugniaux et al., 2010; Slootmaker et al., 2009). When the estimated EE by the Polar FA20 and measured EE by doubly labelled water during a week were compared, correlation coefficients of $r = 0.80$-0.86 were found (Kinnunen et al., 2012). However, analysing the EE of the user over a week might mask differences in the EE of single activities to measured values. Therefore, these values cannot be compared directly to our results. Concluding, the overall correlation found in our study was adequate. The lower values than in other studies result from limitations in the comparison with these studies and implicate that the EE of our included variety of activities is estimated adequately.

This study is limited since the activities were performed under laboratory conditions. However, this study introduces a new device that recognizes walking on different inclines at individually chosen intensities correctly and distinguishes them from running. This study not only evaluated the AS's capacity to recognize walking and running, which were the activities the AS was developed for, but also investigated the AS's classification of other activities performed in daily life, which might be crucial for the credibility of the device for the user. Although the AS was not developed as a measurement device, it accurately recognized walking at different intensities and inclines and precisely measured continuous bouts of walking or running. In addition, it provided a good estimate of EE for moderate walking. These results, in combination with its relatively low cost and its use of simple age-, language-, culture- independent feedback, indicate that the AS might be an effective PA intervention tool. Prompting a user to set a PA goal, specifying graded tasks, providing feedback on performed PA, and keeping a diary were suggested as the key motivational factors for increasing activity with pedometers and recent new technologies (Abraham et al., 2008; Bravata et al., 2007; Clemes & Parker, 2009; De Cocker, et al., 2008; McKay et al., 2009; Michie et al., 2008, 2011; Rooney et al., 2003). The AS fulfills these requirements by setting a graded goal according the user’s fitness level and provides a diary within the software; the AS appears to be a promising feedback device for future PA interventions.

Conclusion

The AS proved to be an easy-to-use and valid tool for the recognition of walking on different inclines at individually chosen intensities. It correctly distinguishes between walking and running, which makes it superior to ordinary pedometers and allows the user to follow recent guidelines that emphasize the importance of intensive activities. Furthermore, the AS gave reasonable estimates of time spent on inactivity, walking, and running during both, a household activity and a sport activity, which might be crucial information for the user in terms of the credibility of the device. It provides correct feedback about the continuous activity bouts accomplished by the user in a simple way that might be especially understandable by groups that are at risk with regard to PA. In addition the AS provided good estimates of EE during moderate walking on a group level. The AS’s smiley icon sets a user-specific activity goal and a feedback on fulfilled PA bouts might also support its potential as an intervention tool. Therefore, future research should investigate the effectiveness of the AS in increasing PA behavior in intervention studies.

Acknowledgements

The authors are grateful to the participants for their willingness to participate in the study. We hereby disclaim any competing interests. This study has been supported by the ActiSmile Company (Kiesen, Switzerland), but the authors performed all experiments and analyzed the data independently.

References

Achten, J. & Jeukendrup, A. E. (2003). Heart rate monitoring. *Sports Medicine 33*(7): 517–523.

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J. , & Hebert, J. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology, 161*(4), 389–398. doi: 10.1093/aje/kwi054.

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., . . . Leon, A. S. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine & Science in Sports and Exercise, 32*(9 Suppl), S498–504.

Aminian, K. & Najafi, B. (2004). Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications. *Computer Animation Virtual Worlds, 15(2)*, 79-94. doi: 10.1002/cav.2.

Andersen, L. B., Harro, M., Sardinha, L.B., Froberg, K., Ekelund, U., Brage, S. & Anderssen, S. A. (2006). Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet, 368*(9532), 299–304.

Araiza, P., Hewes, H., Gashetewa, C., Vella, C. A., & Burge, M. R. (2006). Efficacy of a pedometer-based physical activity program on parameters of diabetes control in type 2 diabetes mellitus. *Metabolism, 55*(10), 1382–1387. doi: S0026-0495(06)00214-9.

Assah, F. K., Ekelund, U., Brage, S., Wright, A., Mbanya, J. C., Wareham, N. J. (2011) Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. *International Journal of Epidemiology, 40*(1), 112 – 120.

Bamberg, S. J. M, Benbasat, A. Y., Scarborough, D. M., Krebs, D. E., Paradiso, J. A. (2008). Gait-analysis using a shoe-integrated wireless sensor system. *Institute of Electrical and Electronics Engineers, 12*(4), 413–423.

Bassett, D. R. Jr., Rowlands, A.V., Trost, S. G. (2012) Calibration and validation of wearable monitors. *Medicine and Science in Sports and Exercise, 44*(1 Suppl), S32 – S38.

Blair, S. N. (2009). Physical inactivity: the biggest public health problem of the 21st century. *British Journal of Sports Medicine, 43*(1), 1–2.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160.

Brage, S., Brage N., Franks, P. W., Ekelund, U., Wong, M.Y., Andersen, L. B. . . . Wareham, N. J. (2004). Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology, 96*(1), 343–351.

Brage, S., Brage, N., Franks, P.W., Ekelund, U., Wareham, N. (2005). Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical* Nutrition*, 59*, 561–570. doi:10.1038/sj.ejcn.1602118.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R. J., . . . Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health. *The Journal of the American Medical Association, 298*(19), 2296–2304.

Bringolf, B., Grize L., Mäder, U., Ruch, N. Sennhauser, F. H. & Braun-Fahrländer, C.(2009) Assessment of intensity, prevalence and duration of everyday activities in Swiss school children: a cross-sectional analysis of accelerometer and diary data. *International Journal of Behavioral Nutrition and Physical Activity, 6*(50). doi:10.1186/1479-5868-6-50.

Brugniaux J. V., Niva A., Pulkkinen I., Laukkanen R. M. T., Richalet, J. Pichon A., P. (2010). Polar Activity Watch 200: a new device to accurately assess energy expenditure. *British Journal of Sports Medicine, 44*(4), 245-249.

Campbell, K. L., Crocker, P. R., McKenzie, D. C. (2002). Field evaluation of energy expenditure in women using Tritrac accelerometers. *Medicine and Science in Sports and Exercise, 34*(10), 1667-1674.

Chen, K. J., Janz, K. F., Zhu, W. & Brychta R. J. (2012). Redefining the roles of sensors in objective physical activity monitoring. *Medicine and Science in Sports and Exercise, 44*(1 Suppl), S13–S23.

Clemes, S. A., & Parker, R. A. (2009). Increasing our understanding of reactivity to pedometers in adults. *Medicine and Science in Sports and Exercise, 41*(3), 674–680.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Colley, R. C., Garriguet, D., Janssen, I., Craig, C. L., Clarke, J.& Tremblay, M.S. (2011). Physical activity of Canadian adults: Accelerometer results from the 2007 to 2009

Canadian Health Measures Survey. *Statistics, Health Reports, 22*(1), Canada, Catalogue no. 82-003-XPE.

Crouter, S. E., Schneider, P. L., Karabulut, M. & Bassett, D. R., Jr. (2003). Validity of 10 electronic pedometers for measuring steps distance, and energy cost. *Medicine and Science in Sports and Exercise, 35*(8), 1455– 1460.

De Cocker, K. A., De Bourdeaudhuij, I. M., Brown, W. J., & Cardon, G. M. (2008). The effect of a pedometer-based physical activity intervention on sitting time. *Preventive Medicine, 47*(2), 179–181. doi: S0091-7435(08)00272-7.

Ekelund, U., Sardinha, L. B., Anderssen, S. A., Harro, M., Franks, P. W., Brage, S. . . . Froberg, K. (2004). Associations between objectively assessed physical activity and indicators of body fatness in 9- to 10-y-old European children: a population-based study from 4 distinct regions in Europe (the European Youth Heart Study). *American Journal of Clinical Nutrition, 80*(3), 584–590.

Elia, M., & Livesey, G. (1992). Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. *World Review of Nutrition and Dietetics, 70*, 68–131.

Freedson, P., Bowles, H. R., Troiano, R., Haskell, W. (2012). Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and Science in Sports and Exercise, 44*(1 Suppl), S1–S4.

Foster R. C., Lanningham-Foster, L. M., Manohar, C., McCrady, S. K., Nysse, L. J. Kaufman, K. R. Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Preventive Medicine, 41*(3-4), 778–783.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., . . . Bauman, A. (2007). Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation, 116*(9), 1081–1093. doi: CIRCULATIONAHA.107.185649.

Hendriksen, M., Lund, H., Moe-Nilssen, R., Biddal, H., Danneskiod-Samsoe, B. (2004). Test-retest reliability of trunk accelerometric gait analysis. *Gait & Posture, 19*(3), 288–297.

Kinnunen, H., Tanskanen, M., Kyröläinen, H., & Westerterp, K., R. (2012). Wrist-worn accelerometers in assessment of energy expenditure during intensive training. *Physiological Measurement, 33,* 1841–1854.

Krauss, J., Solà, J., Renevey, P., Maeder, U., & Buchholz, H. (2008). ActiSmile, a portable biofeedback device on physical activity. *Biomed '08 Proceedings of the Sixth IASTED International Conference on Biomedical Engineering*, 359–362.

Laukkanen, R. M. T. & Virtanen, P. K. (1998). Heart rate monitors: state of the art. *Journal of Sports Sciences, 16*(1), 3–7.

Le Masurier, G. C. & Tudor-Locke, C. E. (2003). Comparison of pedometer and accelerometer accuracy under controlled conditions. *Medicine and Science in Sports and Exercise, 35*(5), 867–871.

Masse, L. C., Ainsworth, B. E., Tortolero, S., Levin, S., Fulton, J. E., Henderson, K. A., & Mayo, K. (1998). Measuring physical activity in midlife, older, and minority women: issues from an expert panel. *Journal of Women's Health, 7*(1), 57–67.

Mayagoita, R. E., Nene, A. V., Veltink, P. H. (2002) Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. *Journal of Biomechanics, 35*(4), 537–542.

McKay, J., Wright, A., Lowry, R., Steele, K., Ryde, G., & Mutrie, N. (2009). Walking on prescription: the utility of a pedometer pack for increasing physical activity in primary care. *Patient Education Counseling, 76*(1), 71–76. doi: S0738-3991(08)00585-5.

Michie, S., Johnston, M., Francis, J., Hardeman, W., Eclles, M. (2008). From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology – An International Review, 57*(4), 660–680.

Michie, S., Ashford, S., Sniehotta, F. F., Dombrowski, S. U., Bishop, A. & French, D. P. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALORE taxonomy. *Psychology & Health*, 26(11), 1479–1498.

Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., & Koh, Y. O. (1990). A new predictive equation for resting energy expenditure in healthy individuals. *American Journal of Clinical Nutrition, 51*(2), 241–247.

Morris, J. N., & Hardman, A. E. (1997). Walking to health. *Sports Medicine, 23*(5), 306–332.

Murphy, M. H., Nevill, A. M., Murtagh, E. M. & Holder, R. L. (2007). The effect of walking on fitness, fatness and resting blood pressure: A meta-analysis of randomised, controlled trials. *Preventive Medicine, 44*(5), 377–385.

Nawab, S. H., Roy, S. H. & De Luca, C.J. (2004). Functional activity monitoring from wearable sensor data. *Conference Proceedings of the International Institute of Electrical and Electronics Engineers, Annual International Conference in Engineering in Medicine and Biology Society, 2*, 979-982.

Nelson, M. E., Rejeski, W. J., Blair, S. N., Duncan, P. W., Judge, J. O., King, A. C., . . . Castaneda-Sceppa, C. (2007). Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Circulation, 116*(9), 1094–1105. doi: CIRCULATIONAHA .107.185650

O'Donovan, G., Blazevich, A. J., Boreham, C., Cooper, A. R., Crank, H., Ekelund, U., . . . Stamatakis, E. (2010). The ABC of physical activity for health: a consensus statement from the British Association of Sport and Exercise Sciences. *Journal of Sports Sciences, 28*(6), 573–591. doi: 921352711

Oja, P., Bull, F. C., Fogelholm, M., & Martin, B. W. (2010). Physical activity recommendations for health: what should Europe do? *BioMed Central Public Health, 10*, 10. doi: 10.1186/1471-2458-10-10

Ojiambo, R., Konstable, K., Veidebaum, T., Reilly, J., Verbestel, V., Huybrechts, I., . . . Pitsiladis, Y. P. (2012). Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: the IDEFICS validation study. *Journal of Applied Physiology, 113*(10), 1530–1536.

Oliver, M., Schofield, G. M., Kolt, G. S. & Schluter, P. J. (2007) Pedometer accuracy in physical activity assessment of preschool children. *Journal of Science and Medicine in Sport, 10*(5), 303-310.

Pärkkä, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J. & Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *Institute of Electrical and Electronics Engineers Transaction on Information Technology in Biomedicine*, *10*(1), 119–128.

Pearce, D. H., & Milhorn, H. T., Jr. (1977). Dynamic and steady-state respiratory responses to bicycle exercise. *Journal of Applied Physiology, 42*(6), 959–967.

Plasqui, G. & Westerterp, K. (2007). Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity, 15(10), 2371–2379. doi: 10.1038/oby.2007.281.

Rooney, B., Smalley, K., Larson, J., & Havens, S. (2003). Is knowing enough? Increasing physical activity by wearing a pedometer. *Wisconsin Medical Journal, 102*(4), 31–36.

Slootmaker, S. M. Chin A Paw, M. J. M., Schuit, A. J., van Mechelen, W. & Koppes, L. L. J. (2009). Concurrent validity of the PAM accelerometer relative to the MIT Actigraph using oxygen consumption as a reference. *Scandinavian Journal of Medicine & Science in Sports, 19*, 36–43.

Strong, W. B., Malina, R. M., Blimkie, C. J. R., Daniels, S. R., Dishman, R. K., Gutn, B., . . . Trudeau, F. (2005). Evidence based physical activity for school-age youth. *Journal of Pediatrics*, *146*, 732–737.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T. & McDowell M. (2008). Physical activity in the United States: measured by accelerometer. *Medicine and Science in Sports and Exercise, 40*(1), 181–188.

Tudor-Locke, C. E., & Myers, A. M. (2001). Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Medicine, 31*(2), 91–100.

Tudor-Locke, C. E., Williams, J. E., Reis, J. P., Pluto, D. (2004) Utility of pedometers for assessing physical activity. *Sports Medicine, 32*(12), 795–808.

Tudor-Locke, C. E., (2004). *Taking Steps toward Increased Physical Activity: Using Pedometers To Measure and Motivate.* President's Council on Physical Fitness and Sports. U.S. department of education. Office of educational research and improvement. *Educational resources information center (ERIC).* Retrieved from http://www.eric.ed.gov/PDFS/ED470689.pdf.

U.K. Department of Health, Physical Activity, Health Improvement and Protection. (2011). *Start active, stay active: a report on physical activity for health from the four home countries' Chief Medical Officers.* London: Department of Health, Physical Activity, Health Improvement and Protection. Retrieved from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_128209

U.S. Department of Health and Human Services, Physical Activity Guidelines Advisory Committee. (2008). *Physical activity guidelines for Americans: be active, healthy, and happy*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1753-4887.2008.00136.x/abstract;jsessionid=EA6CD15913D564E59AE06CEF006E8D51.d04t01.

Van Remoortel, H. V., Giavedoni, S., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E. . . Troosters, T. (2012). Validity of activity monitors in health and chronic disease: a systematic review. International *Journal of Behavioral Nutrition and Physical Activity, 9*(84). Retrieved from http://www.ijbnpa.org/content/9/1/84.

Vogler, A. J., Rice, A. J., & Gore, C. J. (2010). Validity and reliability of the Cortex MetaMax3B portable metabolic system. *Journal of Sports Science, 28*(7), 733–742. doi: 921549046.

Welk, G. J., McClain, J, Ainsworth, B. E. (2012) Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise, 44*(1 Suppl), S39–S49.

Whipp, B. J., Ward, S. A., Lamarra, N., Davis, J. A., & Wasserman, K. (1982). Parameters of ventilatory and gas exchange dynamics during exercise. *Journal of Applied Physiology*, *52*(6), 1506–1513.

World Health Organization. (2010). *Global recommendations on physical activity for health.* Geneva, Switzerland: World Health Organization.

Energy Expenditure Estimation in Children by Activity-Specific Regressions, Random Forest and Regression Trees from Raw Accelerometer Data

Moritz Vetterli[1] & Nicole Ruch[2]

[1]*Institute of Exercise and Health Sciences, University of Basel, Switzerland*

[2]*Swiss Federal Institute of Sport SFIS, Magglingen, Switzerland*

Abstract

The aim of this study was to compare the energy expenditure (EE) estimates of activity-specific regressions (ASR), random forest (RFEE) and regression trees (treeEE) from raw accelerometer data in children. Forty-one children (age: 9.9 ± 2.2y) performed the activities of sitting, standing, walking, running, jumping, crawling, cycling and riding a scooter for 3.5 min., while 30 Hz raw accelerations were collected with one tri-axial hip-accelerometer and EE was measured using a portable device of spirometry. Twenty out of 42 accelerometer features calculated over 1-s windows were included into the prediction model of the RFEE according to their Gini-index. To provide the activity-specific information and the relevant features for the ASR, an a priori decision tree was used. The ASR accurately predicted the EE of sitting, walking, running, jumping, crawling and riding a scooter with biases of 0.04, 0.08, −0.33, −0.61, 0.08 and −0.41 MET, respectively. RFEE precisely estimated the EE of cycling, riding a scooter, jumping and running (bias: −0.18, −0.21, −0.57 and −0.29 MET) and the treeEE accurately predicted the EE of running and cycling (bias: −0.17 and −0.38 MET). The ASR predicted EE more accurately than RFEE or the treeEE. Using activity-specific information seems therefore to enhance the accuracy in assessing EE in children with raw accelerometer data.

KEYWORDS: ENERGY EXPENDITURE, RAW ACCELEROMETER DATA, CLASSIFICATION, CHILDREN, DATA MINING

Introduction

Physical activity (PA) is known to have beneficial effects on the mental (Biddle & Asare, 2011), bone (Boreham & McKay, 2011; Fuchs, Bauer, & Snow, 2001) and cardiovascular health of children (Andersen et al., 2006). Furthermore, it reduces the risk of overweight, obesity (Hills, Andersen, & Byrne, 2011) and type II diabetes (Boulé, Haddad, Kenny, Wells, & Sigal, 2001). Additionally, an active childhood is a good indicator of an active lifestyle in adulthood (Craigie, Lake, Kelly, Adamson, & Mathers, 2011; Telama et al., 2005). Activity guidelines for children based on self-reports recommend a minimum of one hour PA of moderate to vigorous intensity per day (Biddle, Sallis, & Cavill, 1998; Pate, 1995, Strong et al. 2005). However, measurements based on accelerometry recommend 90 min of moderate to vigorous PA per day (Andersen et al., 2006). Therefore, more objective methods may lead to an increase in the recommended health-enhancing dose of PA. Furthermore, objective measurements that define the exact amount of PA needed for each single health benefit are missing in the scientific literature (Biddle et al., 1998) and there is still no consensus on how to measure intermittent and spontaneous PA in children (Andersen et al., 2006; Biddle et al., 1998). Therefore, methods which objectively assess PA should be improved to achieve a deeper understanding of activity behaviour in childhood (Andersen et al., 2006; Biddle et al., 1998).

In the past few years, accelerometers became widely accepted as a valid tool to measure PA as they provide information about the intensity, duration and frequency of activities (John & Freedson, 2012; Nilsson et al., 2009; Riddoch et al., 2004). Accelerometers are especially advantageous because they are comfortable to wear (Janz, 1994), have a large storage capacity to collect data over several days (Freedson, Pober, & Janz, 2005; John & Freedson, 2012) and do not affect the activities. Cut-off-methods determine the intensity of activities by providing thresholds in accelerometer counts for the corresponding values in the metabolic equivalent of task (MET), in order to measure the time spent in sedentary, light, moderate or vigorous activities (Freedson, Melanson, & Sirard, 1998; Hänggi, Phillips, & Rowlands, 2013; Puyau, Adolph, Vohra, & Butte, 2002; Treuth et al., 2004). However, the intensity classification by this approach strongly depends on the activities performed during the development of the thresholds (Puyau et al., 2002; Welk, 2005). To estimate EE, single linear regression models have often been developed using accelerometer data as an independent variable (Eston, Rowlands, & Ingledew, 1998; Freedson et al., 2005; Puyau, Adolph, Vohra, Zakeri, & Butte, 2004; Trost et al., 1998) - though, it has been shown that this approach under- or overestimates the EE of certain activities (Freedson et al., 2005; Staudenmayer, Zhu, & Catellier, 2012) and the development of the regressions strongly depends on the included activities (Bassett, Rowlands, & Trost, 2012). The use of non-linear regression for the EE prediction led to comparable results as linear regressions (Tanaka, Tanaka, Kawahara, & Midorikawa, 2007).

Two or more activity-specific regressions (ASR) have been shown to be more precise than single prediction equations (Brandes, Van Hees, Hannöver, & Brage, 2012; Crouter, Clowers, & Bassett, 2006; Crouter, Horton, & Bassett, 2012). In parallel, an artificial neural network (ANN) has recently been suggested as a promising approach to predict EE from accelerometer data (Trost, Wong, Pfeiffer, & Zheng, 2012). Advantages and disadvantages of ASR models and ANN to predict EE from accelerometer data were recently conversely debated (Bonomi & Plasqui, 2012; Freedson, Lyden, Kozey-Keadle, & Staudenmayer, 2012), but it is not clear so far, which of these methods leads to more accurate results.

Most of the previous studies, which estimated EE by accelerometry, used counts per second. Recently features were extracted from accelerometer counts output over 10s windows to predict EE in children (Crouter et al., 2012; Trost et al., 2012). It has been suggested that the extraction of features from raw accelerometer data might be advantageous for assessing PA or EE (Liu, Gao, & Freedson, 2012). However, only a few studies measured raw accelerometer data to assess EE (Brandes et al., 2012; Van Hees et al., 2011). For EE predictions, they summed up the raw data to 1s window vector magnitude data. Until now, the extraction of multiple features and the selection of those that were most important for the EE determination has not been approached.

The aim of the present study was to compare EE estimation methods such as ASR, which consisted of a prior activity recognition and a subsequent linear regression per activity, to straight-forward approaches such as random forest (RF) and a regression tree ($tree_{EE}$) for several children-specific activities. The second aim of this study was to show the extraction of several different features directly from raw accelerometer data and the use of methods that help to choose the most important features for the EE estimation.

Methods

Participants

All children of a primary school and additionally, the participants of a voluntary sports week, where children could participate in a sports program every morning, were asked by an information letter to participate in the study. All children, who took part, and their parents signed an informed consent and the study was approved by the regional Ethics Committee. 46 children between the ages of 5 and 13 years agreed to participate in the study. Data of five subjects were excluded because of missing or deficient data due to issues with the instruments. Data of 41 children (21 girls and 20 boys, age: 9.9 ± 2.2 years) were used for further analysis.

Measurement devices

Body weight was measured using a digital scale (Modell 861, Seca GmbH & Co, Hamburg, Germany) to the nearest 0.1 kg and height was measured using a stadiometer (Modell 213, Seca GmbH & Co, Hamburg, Germany) to the nearest 0.5 cm. A light (570 g) indirect device of spirometry (MetaMax 3B, Cortex Biophysik GmbH, Leipzig, Germany) with an attached breathing mask (Hans Rudolph, Inc., Kansas City, KS, USA) was used to measure breath by breath oxygen uptake (VO_2) and carbon dioxide production (VCO_2). The device of spirometry was calibrated before each measurement according to the manufacturer's guidelines. Validity, reliability, and stability of the device were previously reported (Macfarlane & Wong, 2012). The collected data were downloaded to a laptop with the corresponding software (Metasoft, Version 3.9.8., CORTEX, Germany) and the equation of Elia and Livesey (1991) was used to determine EE from VO_2 and VCO_2. The values in the unit of MET were calculated relatively to the resting EE (REE). The REE was determined with the equations depending on weight and age from Schofield (1984), as it was recommended by Rodríguez, Moreno, Sarría, Fleta and Bueno (2002).

The ActiGraph (Version, GT3X, The ActiGraph, Pensacola, Florida USA) was used to collect 30 Hz raw acceleration data in three axes (longitudinal, anterior-posterior and lateral). The ActiGraph is a small (3.8 cm x 3.7cm x 1.8cm) and light (27g) accelerometer with a capacitive microelectromechanical sensor, which collects data in a range of ± 3 g (John & Freedson, 2012). A digital clock was synchronized with the timeline of the ActiGraph and the device of spirometry in order to synchronize their data with the start and end time of the activities.

Measurement procedure

When the children arrived in the lab, their body weight, height and age was determined. The accelerometer was placed on the right hip with an elastic belt and the children were equipped with the device of spirometry. The following activities were performed: sitting, standing, walking moderately, walking fast, running, jumping, crawling, biking and riding a scooter. These activities were the most frequently performed by children in a video observation study in the free-living conditions of Ruch, Rumo and Mäder (2011). First, the children were asked to perform the activities sitting and standing in order to not influence their EE by other previously performed activities. Afterwards, the remaining activities were performed in random order. Between the activities, the children rested for about 3 min until EE was at baseline level. Except for sitting, standing, crawling and jumping, the activities were carried out on a flat space outside in similar weather conditions (no rain or strong wind). The children were instructed to perform all the activities in their self-selected, moderate pace. If possible, the children rode their own bike, otherwise there were two sizes of children's bikes available. With the city scooter, the children rode around cones placed in a big eight. Jumping was exercised over a rope on the floor to a beat of 1 Hz. To make crawling interesting, the children were asked to carry a table-tennis ball with a teaspoon in one hand and to place the other hand on the floor. All activities were performed for 3.5 min. The last seconds of the data of every activity, were cut according to visual inspection. The data of the last minute before the cut were used for the calculation, so as to obtain steady state conditions (Pearce & Milhorn, 1977; Whipp, Ward, Lamarra, Davis, & Wasserman, 1982). Always the same minutes of the data of the accelerometer and device of spirometry were used for the analysis.

Statistical analyses

Based on the 30 Hz acceleration raw data in three axes, 42 features with a window size of 1s were calculated according to the suggestions of Liu et al. (2012). The features are considered to filter specific information from the raw data by summing up every second of the measured raw data (30 acceleration values) in different ways in order to reduce the noise in the data. They are considered to have a better correlation to EE or the activity classes than the raw data (Figure 1). The following features were generated for every axis: mean, standard deviation (sd), coefficients of variation (CV), minimum (min), maximum (max), percentiles (10th, 25th, 50th, 75th, 90th), sum, zero crossing (zc) and signal power (sp). Additionally, the vector magnitude (VM) and the distance between the minimum and the maximum of every second (d) were calculated (Table 1).

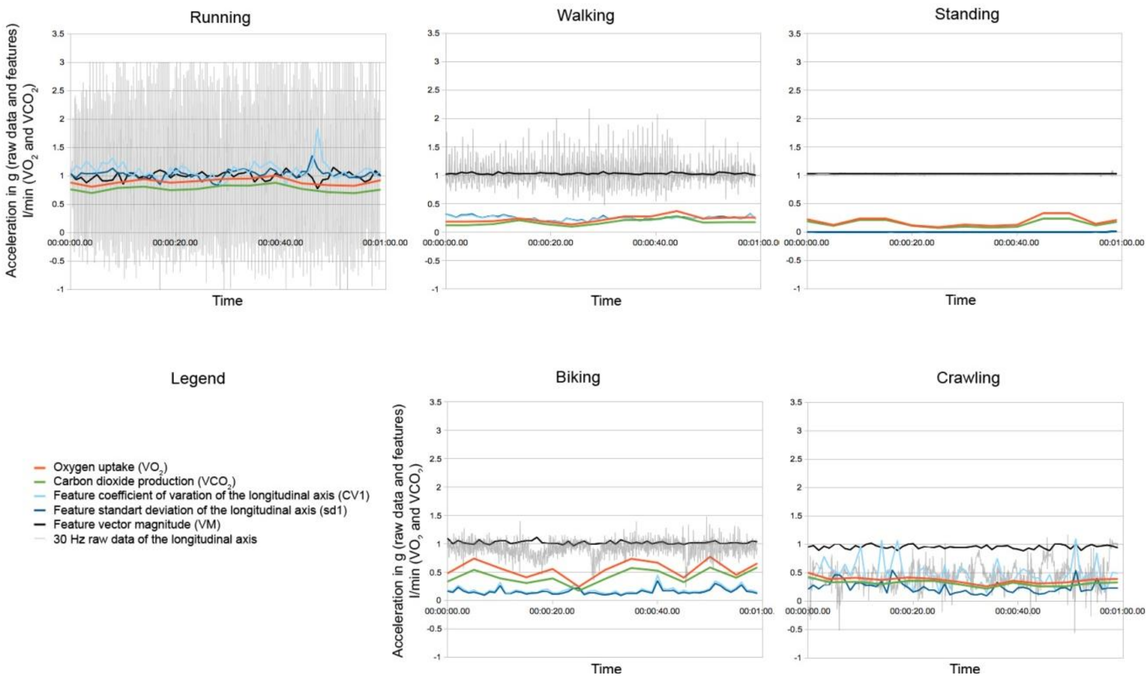


Figure 1. Measured raw data, the features coefficient of variation (CV1), standard deviation (sd1) of the longitudinal axis and the vector magnitude (VM), as well as the measured oxygen uptake (VO_2) and carbon dioxide production (VCO_2).

Table 1. All features, calculated over a non-overlapping window size of 1s from the 30Hz raw accelerometer data.

Feature	**Abbreviation**	**Formula**
Mean	mean	$\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$ s_i = raw acceleration
Standard deviation	sd	$\sigma_s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \frac{1}{N}\sum_{j=1}^{N} s_j)}$
Minimum	min	min (s_i)
Maximum	max	max (s_i)
Distance between min and max	d	$d_1 = \lvert\max(s)\rvert + \lvert\min(s)\rvert$, $d = \sqrt{d_1^2 + d_2^2 + d_3^2}$
Vector magnitude	VM	$VM = \sqrt{a_1^2 + a_2^2 + a_3^2}$ axes: 1 = longitudinal, 2 = anterior-posterior, 3= lateral
Corrected vector magnitude (the acceleration, without the earth acceleration)	VM2	$VM = \sqrt{(a_1 - 1)^2 + a_2^2 + a_3^2}$ (In units of g)
Sum	sum	$S = \sum_{i=1}^{N} s_i$
Coefficient of variation	CV	$CV = \frac{\sigma_i}{\lvert\bar{s}\rvert}$
Percentile (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th})	p_0.1, p_0.25, p_0.5, p_0.75, p_0.9	p_x = Value, so that x of the measured $\lvert s_i\rvert$ are below p_x
Signal Power	sp	$sp = \sum_{i=1}^{N} s_i^2$
Zero crossing	zc	Number of times the signal crosses 0

Calculating a multitude of features requires a selection for the EE prediction models and most of the other methods do not provide an automatic feature selection. EE of the different activities was estimated with $tree_{EE}$, RF_{EE} and ASR. These approaches were chosen, because each of them offers a possibility for a feature selection. In addition to the selected features, age, body weight and the sex of the children were introduced into the EE prediction models, because personal factors have been reported to affect EE (Ekelund, Yngve, Brage, Westerterp, & Sjöström, 2004; Schofield, 1984). The activities of moderate and fast walking were merged into the activity walking.

The ASR prediction required two work steps. First, the activity type of every second was recognized and second, the EE was estimated with a separate linear regression for each activity type, including the most important features and personal factors as variables. The prior activity classification of the ASR was achieved by using a decision tree (tree$_{Class}$), based on the calculated 1s window features. In each node of the tree the input data (all features of one second measured data) is disposed to one of the two sub braches, which end in another node with a new decision criterion, until the final decision is made and the activity class of the input data is determined. A cost complexity function was used as a measure for the proportion of the complexity of the tree and the error rate. It was chosen at 0.01 for the tree$_{Class}$ in the present study and led to a tree with ten splitting nodes and eleven branches (Figure 2).

Figure 2. Classification tree (tree$_{Class}$). To read from the top to the bottom, if the node criterion is full field, following the left branch, otherwise the right one (seen from the viewer's perspective). Scooter = riding a city scooter. Features: sd = standard deviation, CV = coefficient of variation, zc = zero crossing, p_x = xth percentile, max = maximum. Axes: 1 = longitudinal, 2= anterior-posterior, 3= lateral.

The tree$_{Class}$ chose its node criterions from all the available features according to the decrease of the Gini impurity, which measures the goodness of a variable as splitting criterion in a node of the tree. Gini impurity is:

$$i(\tau) = 1 - p_a^2 - p_b^2, \text{ where } \tau = \text{node}, \; p_k = \frac{n_k}{n}, \quad (1)$$

n = number of data points in the sample,

n_k = correctly classified data points to the class k (Menze et al., 2009).

The decrease of the Gini impurity is:

$$\Delta i(\tau) = i(\tau) - p_l \, i(\tau_l) - p_r \, i(\tau_r), \; \tau_{l/r} = \text{left/right subnode} \quad (2)$$

$$p_{l/r} = \frac{n_{l/r}}{n} \text{ (Menze et al., 2009)}$$

The decrease of the Gini impurity was calculated for every available variable (all features) and the one with the highest decrease was chosen as node criterion (Menze et al., 2009). According to the Gini impurity, standard deviation (sd1) and coefficients of variation of the longitudinal axis (CV1) were relevant features in the tree$_{Class}$ (Figure 2). The regressions for each activity were developed using the same two features; additionally, the ASR included the vector magnitude (VM) since this feature was considered to reflect the mechanical acceleration of the activity. In addition, the personal factors age, sex and weight were included in the regressions. From these selected features (sd1, CV1 and VM) and personal factors (age, sex and weight), the relevant variables for each regression were determined using multiple regressions models with backwards exclusions of the variables (Table 2). Hence, the ASR determined the activity type by the $tree_{Class}$ for each second of the measured data and applied the regression of the respective activity type to estimate EE.

Table 2. Activity specific regressions (ASR) per activity. VM = vector magnitude, sd1 = standard deviation of the longitudinal axe, CV1 = coefficient of variation of the longitudinal axe.

Activity	**Equation for the EE-prediction [kJmin^{-1}]**	**R^2**	**SEE** [kJmin^{-1}]	**p-value**
Sitting	1.40434 + 0.21455 age + 0.05283 weight	0.56	0.87	<0.001
Standing	0.81335 + 0.18471 age + 0.07587 weight	0.70	0.78	<0.001
Walking	70.69992 – 68.81717 VM – 86.25837 sd1 + 0.52874 age + 0.17639 weight	0.64	2.81	<0.001
Running	0.353 + 1.32144 age + 0.3436 weight	0.74	3.99	<0.001
Jumping	-1.01301 – 25.71135 sd1 + 25.83887 CV1 + 0.82935 age + 0.4458 weight	0.74	4.48	<0.001
Crawling	-2.05614 + 16.87322 sd1 + 1.06574 CV1 + 0.53126 age + 0.22102 weight	0.83	1.86	<0.001
Scooter	87.30692 – 88.29084 VM + 32.52044 sd1 + 0.36355 weight	0.66	3.93	<0.001
Cycling	0.85551 + 78.23134 sd1 + 0.20154 age	0.48	4.04	<0.001

The $tree_{EE}$ worked analogue the $tree_{Class}$ and disposed the input data for each node into one of two smaller regions, based on the condition of one feature. In contrast to the $tree_{Class}$ that recognized the activity class, it finally determined directly the EE values. The used $tree_{EE}$ had a cost complexity of 0.0005 (77 nodes) and chose the splitting variables from all 1-s window features and the personal factors age, sex and weight, according to the decrease of the Gini impurity, as like the $tree_{Class}$.

The RF_{EE} grows a defined amount of different regression trees based on a random vector of the input data (all features and personal factors) sampled independently for all trees of the forest. The result is the most supported class by the different trees (Breiman, 2001). RF can enhance the prediction because it is more robust in respect to noise and instabilities of small changes in the dataset than a single tree. But in the forest the influence of a variable becomes unclear and hard to interpret. Therefore, RF evaluates the feature according to their importance. One evaluation method is the decreases of the Gini index. The decrease of the Gini impurity ($\Delta i(\tau)$), used from the single trees to determine the splitting variable of every node, was accumulated over all nodes and all trees of the forest for each variable. Thus the Gini index of RF determines how often a variable was used for a splitting and how valuable the split was (Menze et al., 2009). RF_{EE} was trained with 1000 trees per decision with the 20 most useful features according to the decrease of the Gini index (Figure 3). These included all features of the longitudinal axis, the standard deviation, coefficients of variation, zero crossing, sum, the 25th percentile of the anterior-posterior axis, the standard deviation of the lateral axis and the distance between the minimum and the maximum of every second. In order to compare the results of the $tree_{Class}$, a classification by RF (RF_{Class}) was performed. RF_{Class} used a pre-defined amount of decision trees (1000 in this study) that were trained by a random set of the input data, like as RF_{EE} but determined in the end an activity class instead of an EE value. The model was developed the same way as the RF_{EE}. If only the 10 most useful features of the Gini index or all 42 features were included, RF_{Class} reached lower classification results than with 20 selected features (overall recognition of RF_{Class}: with all features: 72.7%; 10 selected features: 71.7%, 20 selected features: 73.5%).

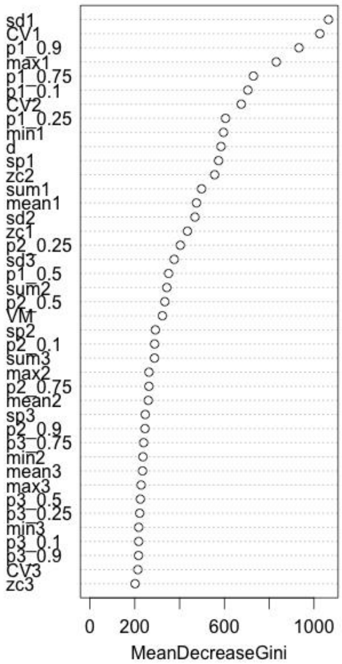


Figure 3. The mean decrease of the Gini index of the features, calculated by RF to evaluate the features after their importance. The higher the decrease the more important is the feature. The Gini index is the accumulation of the decreases of the Gini impurites, which were calculated by the trees to determine the splitting variable of the single nodes, over all variables and all nodes of every tree in the forest. Abbreviations of the features according to table 1.

A leave-one-out cross-validation was used to validate the activity recognitions and the EE estimations, as it was recommended by Staudenmayer et al. (2012). The method trains the algorithms with the data of n-1 children and tests it on the one left out. This procedure was repeated 41 times, so that the data of every child was tested once. Mean bias and root mean squared error (RMSE) of the estimated to the measured EE were determined for each activity. In order to compare the measured and estimated EE, a non-parametric Wilcoxon-Rank-Sum test with Bonferroni adjustments for multiple comparisons was used since the quantile-comparison plots of the data revealed a deviance from the normal distribution. The same procedure was applied to compare the bias of the ASR to that of the RF_{EE} and the $tree_{EE}$. A Chi-square test was used to compare the correctly classified and misclassified proportions of each activity and of the overall result by the $tree_{Class}$ and RF_{Class}. The feature calculation, all EE estimations by ASR, RF_{EE} and $tree_{EE}$, and activity classification by RF_{Class} and the $tree_{Class}$, as well as all statistical analysis were done using R project for statistical computing (R Project, Version 2.15.1, Statistics Department, University of Auckland, New Zeeland). In R, the package rpart was used for the $tree_{Class}$ and the $tree_{EE}$ and the package randomForest was used to perform RF_{EE} and RF_{CLass}.

Results

Energy Expenditure

The ASR accurately predicted EE of six activities, but they significantly underestimated EE during cycling and overestimated EE during standing ($p < 0.05$) (Table 3). RF_{EE} significantly overestimated EE of sitting, standing and walking ($p < 0.05$) and precisely predicted the remaining activities. The $tree_{EE}$ determined EE of the activities running and cycling accurately; however, it underestimated riding a scooter and overestimated the remaining activities ($p < 0.05$) (Figure 4, Table 3).

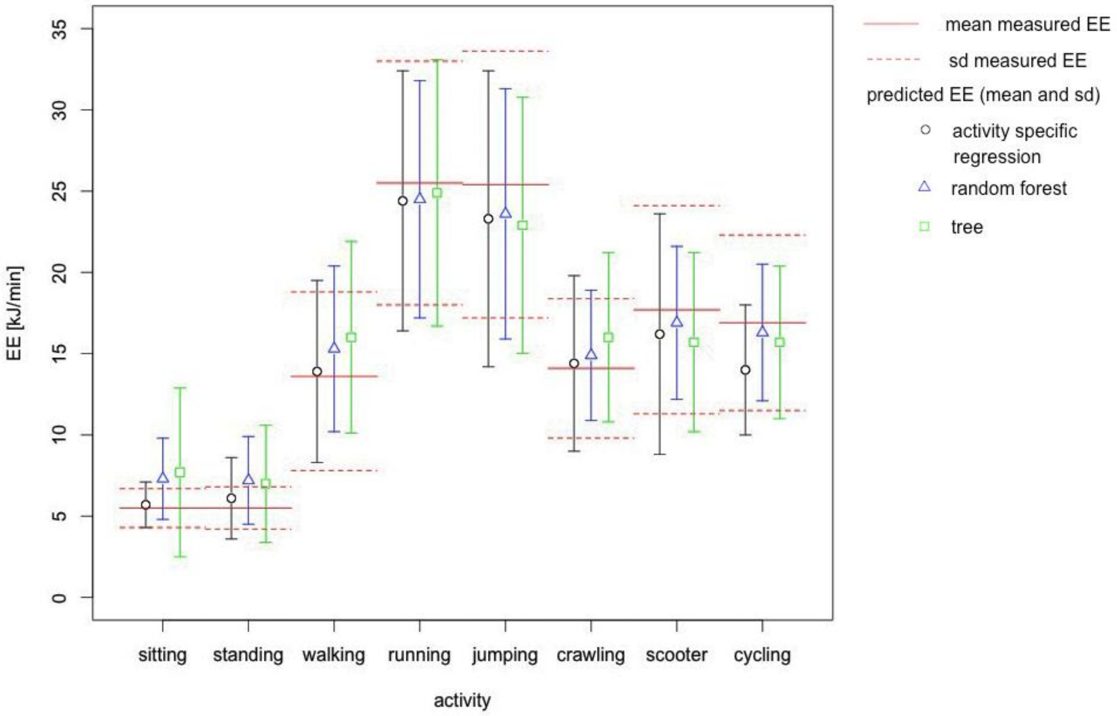


Figure 4. Measured and predicted energy expenditure in kJ/min for all activities (mean and standard deviation).

Table 3. Energy Expenditure, measured and predicted in MET (mean and standard deviation). * differs significantly from the measured value (p<0.05), scooter = riding a city scooter.

Activity	**Measured EE**	**Estimated EE**		
		ASR	RF	$Tree_{Reg}$
Sitting	1.5 ±0.3	1.6 ±0.4	2.0 ±0.7*	2.0 ±1.0*
Standing	1.5 ±0.2	1.7 ±0.6*	2.0 ±0.7*	1.9 ±0.9*
Walking	3.9 ±1.3	3.9 ±1.4	4.3 ±1.2*	4.5 ±1.3*
Running	7.1 ±1.3	6.8 ±1.6	6.9 ±1.3	7.0 ±1.6
Jumping	7.1 ±1.6	6.5 ±2.2	6.6 ±1.5	6.4 ±1.6*
Crawling	3.9 ±0.8	4.0 ±1.2	4.2 ±0.8	4.5 ±1.3*
Scooter	5.0 ±1.5	4.6 ±1.9	4.8 ±1.0	4.4 ±1.3*
Cycling	4.8 ±1.4	4.0 ±1.0*	4.6 ±0.8	4.4 ±0.9

The mean bias of the EE of sitting, standing, walking and crawling estimated by the ASR were within 0.17 MET. Except for crawling, this was significantly lower than the biases of the $tree_{EE}$

or the RF_{EE} ($p < 0.05$) (Table 4). Comparing the three methods of this study, RF_{EE} predicted the EE of cycling, riding a scooter and jumping most accurately (bias: -0.18, -0.21 and -0.53 MET), whereas the $tree_{EE}$ reached the most precise result for running with a bias of -0.17 MET.

Mean RMSE ranged from 0.87 to 5.93 kJ/min for the ASR,from 2.18 to 5.06 kJ/min for the $tree_{EE,}$ and from 2.24 to 5.51 kJ/min and RF_{EE} (Figure 5).

Table 4. Mean bias and RMSE of all methods for all activities in MET and kJ/min. * differs significantly from the aspR bias (p<0.05), scooter = riding a city scooter

Activity	Mean Bias						RMSE		
	ASR		RF		$Tree_{Reg}$		ASR	RF	$Tree_{Reg}$
	kJ min^{-1}	MET	kJ min^{-1}	MET	kJ min^{-1}	MET	MET	MET	MET
Sitting	0.11 ± 0.88	0.04 ± 0.41	1.72 ± 1.86*	0.49 ± 0.67*	2.16 ± 4.05*	0.52 ± 1.08*	0.41	0.83	1.20
Standing	0.65 ± 1.50	0.17 ± 0.55	1.70 ± 1.48*	0.48 ± 0.66*	1.56 ± 1.56*	0.42 ± 0.79*	0.57	0.82	0.89
Walking	0.32 ± 2.85	0.08 ± 1.57	1.67 ± 3.16*	0.47 ± 1.10*	2.41 ± 2.77*	0.65 ± 1.21*	1.57	1.19	1.38
Running	-1.15 ± 5.20	-0.33 ± 1.77	-1.04 ± 4.67	-0.29 ± 1.35	-0.60 ± 4.50	-0.17 ± 1.33	1.80	1.38	1.33
Jumping	-2.09 ± 5.62	-0.61 ± 2.32	-1.84 ± 5.14	-0.53 ± 1.59	-2.50 ± 4.45	-0.72 ± 1.50	2.40	1.67	1.66
Crawling	0.33 ± 1.90	0.08 ± 1.13	0.85 ± 2.94	0.27 ± 0.87	1.90 ± 3.47	0.58 ± 1.20	1.13	0.91	1.33
Scooter	-1.59 ± 3.80	-0.41 ± 2.05	-0.82 ± 5.12	-0.21 ± 1.53	-2.08 ± 3.82	-0.57 ± 1.38	2.09	1.54	1.49
Cycling	-2.83 ± 4.99	-0.82 ± 1.67	-0.54 ± 5.55*	-0.18 ± 1.57*	-1.14 ± 4.83*	-0.38 ± 1.35*	1.86	1.58	1.40

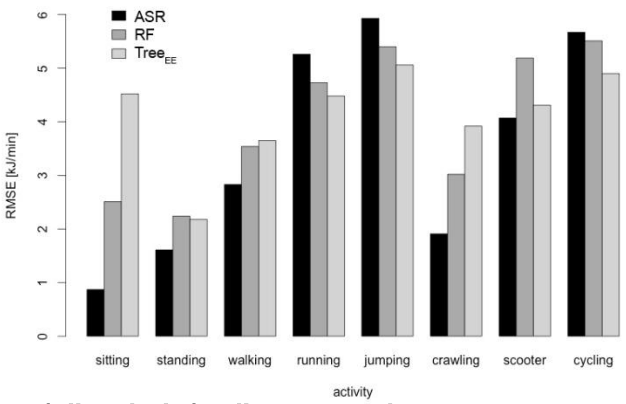


Figure 5. Mean RMSE of all methods for all activities in kJ/min.

Classification

The $tree_{Class}$ achieved an overall recognition of 70.2%. RF_{Class} achieved 73.5% and was significantly more accurate than the $tree_{Class}$ ($p < 0.05$). Sitting and standing were significantly better recognized by the $tree_{Class}$ ($p < 0.05$), whereas the RF_{Class} better classified all other activities - jumping, crawling and cycling significantly better ($p < 0.05$) (table 5). The highest recognition was reached for the activity walking by both methods. Most of the misclassifications were found between similar activities such as cycling and riding a scooter, sitting and standing or running, walking and jumping.

Table 5. Confusions matrix of the classification results. * The methods differ significantly from each other ($p < 0.05$).

Observed Activity	Method	Predicted Activity							
		Sitting (%)	Standing (%)	Walking (%)	Running (%)	Jumping (%)	Crawling (%)	Scooter (%)	Cycling (%)
Sitting *	$Tree_{Class}$	**76**	22	0	0	0	0	1	1
	RF_{Class}	**57**	29	5	3	1	3	1	1
Standing *	$Tree_{Class}$	15	**76**	0	0	0	4	3	2
	RF_{Class}	24	**69**	0	0	2	1	2	1
Walking	$Tree_{Class}$	1	0	**85**	0	2	0	9	1
	RF_{Class}	1	0	**87**	1	1	0	6	3
Running	$Tree_{Class}$	3	1	5	**76**	13	1	2	0
	RF_{Class}	2	1	1	**79**	10	1	2	0
Jumping *	$Tree_{Class}$	2	3	7	22	**52**	0	11	3
	RF_{Class}	0	1	16	7	**73**	1	7	4
Crawling *	$Tree_{Class}$	3	1	1	0	0	**78**	9	7
	RF_{Class}	1	3	4	0	1	**84**	4	7
Scooter	$Tree_{Class}$	0	1	29	0	3	4	**54**	9
	RF_{Class}	3	0	4	0	4	5	**58**	17
Cycling *	$Tree_{Class}$	3	0	12	0	0	7	26	**51**
	RF_{Class}	3	0	6	0	1	9	20	**60**

Discussion

Energy Expenditure

This study compared ASR, RF_{EE} and $tree_{EE}$ as methods to determine the EE of children during several activities using raw accelerometer data. The ASR reached the most accurate results and therefore, using activity-specific information before applying regression equations to activity type, seems to be a promising approach for the future. Regarding EE estimations or activity type recognitions, the calculation of features from raw accelerometer data and feature selection methods showed great potential.

The $tree_{EE}$ is a well-interpretable model but it predicted the EE of only two activities accurately. In contrast, the RF_{EE} estimated the EE of more activities accurately than the $tree_{EE}$ probably due to the use of a multitude of trees. This procedure might have increased the accuracy in comparison to using only one tree, because it can better balance variations and noise in the data. However, as the mode of the results of the single trees is taken as the final result, RF is still based on regression trees and depends, in the end, on the strength of these trees and the correlation between them (Breiman, 2001).

For the less intensive activities, the ASR showed significantly better results than the RF_{EE} or the $tree_{EE}$. The activity-specific regressions of the ASR might adapt more precisely to the intensity of single activities than $tree_{EE}$ or RF_{EE} that are trained with the whole data set. For sitting and standing, the regressions of the ASR were only based on age and weight. The prediction of EE for these sedentary activities was more accurate than the estimations of RF_{EE} and the $tree_{EE}$ that also included accelerometry in this estimation. Similarly to the ASR for sitting and standing, Schofield's (1984) REE equations for children only used body weight and age and reached an R of 0.83 (boys) and 0.81 (girls) for children between 3 and 10 years of age and 0.93 (boys) and 0.80 (girls) for children from 10 to 18 years of age. The regressions of the ASR for sitting achieved slightly lower values (R = 0.75) and for standing, the regression achieved a similar result (R = 0.83). Therefore, the ASR for stationary activities were comparable to existing equations and they were more accurate in estimating the EE of single activities than other methods such as RF_{EE} and $tree_{EE}$.

Brandes et al. (2012) calculated counts from raw accelerometer data and estimated the EE with ASR for walking, stair walking and cycling. With a lowest RSME of 2.15 kJ/min for walking or 6.44 kJ/min for cycling, the RMSE of Brandes et al. (2012) was a bit lower for walking and higher for cycling than that of the ASR, RF_{EE} or the $tree_{EE}$ in the present study. Brandes et al. (2012) worked with only three activities, in comparison to the eight activities used in the present study and they didn’t use a prior classifier to determine the activity type from the measured data. In contrast, the model of Crouter et al. (2012) provides a method that first distinguishes between inactivity, continuous walking or running and intermittent lifestyle activities and estimates the EE using two regressions: one for continuous walking or running and one for intermittent lifestyle activities. They set the EE of the category inactive to 1 MET. Depending on the used thresholds to separate the categories, the biases of Crouter et al. (2012) were within 0.6 - 0.8 of measured MET values for nearly all activities and therefore comparable to the present study. However, the biases of some activities were larger, for example running with a bias of 1.1 MET. This indicates that an even more activity-specific discrimination as used in the present study before applying the ASR might lead to more accurate results in the single activities.

Trost et al. (2012) used an ANN with features from accelerometer 1s counts over a window size of 10s to 60s and five activity categories to predicted EE. Trost et al. (2012) reached lower biases for running and walking than the ASR, $tree_{EE}$ and RF_{EE}. For sedentary activities, the ASR showed better results than the ANN of Trost et al. (2012). The other activity categories of Trost et al. (2012) are difficult to compare, because they merged different activities into a few categories. The biases of the EE of these categories were in a range of 0.01 to 0.62 MET and, therefore, were comparable to or slightly lower than the results of this study.

Overall, the results in the three studies are comparable; however, using ASR is an easy interpretable approach, whereas an ANN is limited in regard to its interpretation due to the complex algorithms. Furthermore, there is no foolproof method that determines the parameters that define ANN. Similarly, the RF_{EE} calculated 1000 trees for every decision and the mode determines the final class. Although every single tree can be extracted from the algorithm, there is no final tree that can be interpreted. Therefore, the RF remains a black box similar to ANN. In contrast to the straightforward approaches of ANN and RF, the ASR needs two steps (activity type classification and subsequent EE estimation), which might be considered as time-consuming. However, recognizing the activity type is an additional factor in describing children’s PA and therefore, the ASR might be the preferred approach for estimating EE in future studies with underlying activity-type classification, as it was recommended by Bonomi and Plasqui (2012).

In contrast to most of the previous studies that used accelerometer counts for the EE estimation, a multitude of different features were calculated in the present study and methods to evaluate them according to their importance were presented. According to the decrease of the Gini index, the features extracted from the longitudinal axis were the most important. Only a few features extracted from the other axis were selected for the model according to the Gini index. Therefore, measuring the vertical axis seems to be most important for a correct EE estimation.

Trost et al. (2012) showed that a longer window size (increase from 10s to 60s) caused a decrease of the average RMSE and average bias of the EE predictions with ANN. Therefore, using a longer time-window than the 1s window that was used, to extract features as in the present study might lead to more accurate EE estimates. As it is of interest to estimate the EE over a longer period of time and not per second, it might be recommended to estimate EE with longer window sizes. However, as children's PA is highly intermitted (Bailey et al., 1995; Baquet, Stratton, Van Praagh, & Berthoin, 2007), shorter intervals might be necessary when measuring EE in a child’s natural environment.

Concluding, comparing ASR with RF_{EE} and $tree_{EE}$, the ASR were most accurate in predicting EE. The results of the present study were comparable to previous studies (Brandes et al., 2012; Crouter et al., 2012; Trost et al., 2012); however, those studies did either not use a combined approach of a classification system and ASR (Brandes et al. 2012), they used only a two-regression model which might not estimate the EE of single activities correctly (Crouter et al., 2012), or they used a neural network which is difficult to interpret as it is laborious to extract its algorithms (Trost et al., 2012). With the more activity-specifically investigation of the EE than in previous studies and the easy interpretation of the algorithm, the ASR using extracted features from raw accelerometer data might be a promising approach for future studies.

Classification

This study showed that RF and decision trees could be used for an activity classification based on raw accelerometer data. A precise activity type recognition is essential for the subsequent application of the ASR. Errors in the classification will be transferred to the EE prediction. RF_{Class} achieved higher recognition rates than the $tree_{Class}$. As the latter was used to identify the activity type prior to the ASR, using RF_{Class} instead as the prior activity recognition method might improve EE predictions of the ASR. However, the $tree_{Class}$ is easy to interpret in contrast to RF_{Class}, where the decisions cannot be interpreted since no final tree is provided. The RF_{Class} reached better results when using the 20 most useful—not all 42—features. This shows that good results are depending on informative features and not on the quantity of them, so a feature selection is needed prior to the pattern recognition. The integrated algorithm that determines the decrease of the Gini index and indicates the importance of the variables and flexibility of RF_{Class} to adapt to complicated data by using bootstrap samples might be an advantage of a future use of this method. The used $tree_{Class}$ (Figure 2) is a simple approach and was based on only 10 features, which were selected within the development of the tree by reducing the Gini impurity for each split of the tree. The advantage of the $tree_{Class}$ is that the decision can easily be reconstructed and interpreted, but the price for it is a lower recognition rate and even if more branches are added, only a small improvement occurs. With a cost complexity of 0.001 (38 branches), the $tree_{Class}$ achieved an overall recognition of 70.9%, compared to its 70.2% with 11 branches (cost complexity of 0.01). Concluding, the RF_{Class} reached the significantly better overall classification result than the $tree_{Class}$, which had, for the most part, its strength in the easily interpretability of the approach.

Similar classification studies, working with one hip-accelerometer, reached similar or slightly better results for different classification methods but for fewer activities. For example, a quadratic discriminate analysis recognized 70.9% and a hidden Markov model 80.8% of four activities correctly (Pober, Staudenmayer, Raphael, & Freedson, 2006). Recognition rates of 72.4% to 80% were reached with an ANN in children (De Vries, Engels, & Garre, 2011; Trost et al., 2012). De Vires et al. (2011) reached slightly better results than the $tree_{Class}$ or RF_{Class} for all activities except for standing, where the $tree_{Class}$ achieved the better recognition rate. In comparison to de Vries et al. (2011), the present study didn’t include playing soccer but additionally included crawling and riding a scooter. Specifically, the later was difficult to recognize correctly (tree$_{Class}$ 54%, RF$_{Class}$: 58%). Trost et al. (2012) merged a multitude of activities into five classes. With longer feature window sizes (10s - 60s), they reached higher recognition rates for sedentary activities and walking, and similar or lower results for running than this study. A decision tree for a classification was also used by Ruch et al. (2011). They reported a recognition rate of 48% with one hip-accelerometer for nine categories. Collecting the data in free-living conditions probably explains the differences compared to the results in this study, which was performed in laboratory conditions.

Concluding, the activity type recognition in this study reached comparable or slightly lower classification results than previous studies, which might be explained by the fact that more activities were included. De Vries et al., (2011) and Trost et al. (2012) worked with features of a window size of 10s, in comparison to window sizes of 1s in the present study. As children's activities are highly intermitted and, on average, the intensive activities last only a few seconds, methods that account for short periods are crucial to accurately describe children's PA (Bailey et al., 1995; Baquet et al., 2007). Hence, the 1s features might be the choice for a more accurate PA description in children.

A limitation of the present study was that the sample of participants was not representative since some of the children were recruited from a voluntary sports week. However, the aim of the present study was not epidemiological but was to compare several methods on the same dataset. Therefore, the representativeness of the children was secondary. The number of participants in the study is lower than in other studies (Brandes et al., 2012; Crouter et al., 2012; Trost et al., 2012). By using a leave-one-out cross-validation this limitation has been countered. The acceleration of intensive activities such as running or jumping outbid the measuring range of 3g of the used accelerometer. Raw acceleration data of these activities often reached 3g and might have been a lot higher in reality. Using accelerometers, being able to measure higher accelerations, might likely lead to more precise data. However, even with this low measurement range, the activities running and jumping reached comparable results in the estimated EE to measured values in the present study. Even if the sampling rate with 30 Hz was comparable to other mentioned studies (Crouter et al., 2012; De Vries et al., 2011; Trost et al., 2012), a higher sampling rate of the accelerometer might lead to more precise measurements as well. In this study, linear accelerations in three axes were measured. As a lot of bodily movements contain rotations, results might be improved if additional sensors such as gyroscopes were added that are able to assess rotational acceleration. In this study, the ASR with its robust linear regressions achieved the best results but it has to be investigated if other, more sensitive models might be more precise when having additional data as basis. Despite all data processing still low correlation between the most important features for the ASR (CV1, sd1, VM) and oxygen uptake (VO_2) or carbon dioxide production (VCO_2) were measured for some activities, for example biking (Figure 1). Even with improvements there would rest a general difficulty in the correlation of mechanical to metabolic energy. In all methods, presented in this study, other measured data than acceleration could be integrated as well. It might be worth considering an addition to the acceleration data to achieve an improvement in EE estimation with these methods. This study was performed in laboratory conditions. The activity patterns of children in free-living conditions might differ from strict controlled lab conditions and the accuracy of EE estimation might decrease under field conditions (Gyllensten & Bonomi, 2011). Testing such algorithms in free-living conditions will be essential for their application in future monitoring or intervention studies. A limited amount of activities was included in this study. The performance of these methods has to be investigated in a broader variety of activities. This study was the first to investigate several approaches that provided a method to select the most important features extracted from accelerometer raw data in order to estimate EE of different child-specific activities. Regarding the EE estimation, ASR with its more activity-specific approach than previous studies, its easily interpretable algorithm and, in general, the use of extracted features from raw accelerometer data might be promising for future studies.

Conclusion

This study compared the accuracy of the ASR, RF_{EE} and $tree_{EE}$ for an EE prediction in children based on raw accelerometer data. The strength of the study was the use of raw accelerometer data and the extraction of a multitude of 1s features, which were evaluated according to their importance. The ASR was the most accurate method for predicting the EE of several activities. They were especially more precise for the less intensive activities. Therefore, using activity-specific information before applying regressions to accelerometer data might be a promising approach for future studies. In comparison to pattern recognition approaches, the ASR might be preferred to assess EE in children, since it is an easily interpretable method and additionally provides information on the performed PA type in a first step. Testing and improving the proposed methods under free-living conditions will be crucial for their application in future monitoring or intervention studies.

References

Andersen, L. B., Harro, M., Sardinha, L. B., Froberg, K., Ekelund, U., Brage, S., & Anderssen, S. A. (2006). Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet*, *368*(9532), 299–304. doi:10.1016/S0140-6736(06)69075-2

Bailey, R. C., Olson, J., Pepper, S. L., Porszasz, J., Barstow, T. J., & Cooper, D. M. (1995). The level and tempo of children’s physical activities: an observational study. *Medicine and science in sports and exercise*, *27*(7), 1033–1041.

Baquet, G., Stratton, G., Van Praagh, E., & Berthoin, S. (2007). Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: A methodological issue. *Preventive Medicine*, *44*(2), 143–147. doi:10.1016/j.ypmed.2006.10.004

Bassett, D. R., Jr, Rowlands, A., & Trost, S. G. (2012). Calibration and validation of wearable monitors. *Medicine and science in sports and exercise*, *44*(1 Suppl 1), S32–38. doi:10.1249/MSS.0b013e3182399cf7

Biddle, S. J. H., & Asare, M. (2011). Physical activity and mental health in children and adolescents: a review of reviews. *British Journal of Sports Medicine*, *45*(11), 886–895. doi:10.1136/bjsports-2011-090185

Biddle, S., Sallis, J. F., & Cavill, N. (1998). *Young and active?: young people and health-enhancing physical activity : evidence and implications*. Health Education Authority.

Bonomi, A. G., & Plasqui, G. (2012). “Divide and conquer”: assessing energy expenditure following physical activity type classification. *Journal of Applied Physiology*, *112*(5), 932–932. doi:10.1152/japplphysiol.01403.2011

Boreham, C. A. G., & McKay, H. A. (2011). Physical activity in childhood and bone health. *British Journal of Sports Medicine*, *45*(11), 877–879. doi:10.1136/bjsports-2011-090188

Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A., & Sigal, R. J. (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *JAMA: the journal of the American Medical Association*, *286*(10), 1218–1227.

Brandes, M., VAN Hees, V. T., Hannöver, V., & Brage, S. (2012). Estimating energy expenditure from raw accelerometry in three types of locomotion. *Medicine and science in sports and exercise*, *44* (11), 2235–2242. doi:10.1249/MSS.0b013e318260402b

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Craigie, A. M., Lake, A. A., Kelly, S. A., Adamson, A. J., & Mathers, J. C. (2011). Tracking of obesity-related behaviours from childhood to adulthood: A systematic review. *Maturitas*, *70* (3), 266–284. doi:10.1016/j.maturitas.2011.08.005

Crouter, S. E., Clowers, K. G., & Bassett, D. R. (2006). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, *100*(4), 1324–1331. doi:10.1152/japplphysiol.00818.2005

Crouter, S. E., Horton, M., & Bassett, D. R. (2012). Use of a Two-Regression Model for Estimating Energy Expenditure in Children. *Medicine and science in sports and exercise*, *44*(6), 1177–1185.

De Vries, S. I., Engels, M., & Garre, F. G. (2011). Identification of Children’s Activity Type with Accelerometer-Based Neural Networks. *Medicine & Science in Sports & Exercise*, *43*(10), 1994–1999. doi:10.1249/MSS.0b013e318219d939

Ekelund, U., Yngve, A., Brage, S., Westerterp, K., & Sjöström, M. (2004). Body movement and physical activity energy expenditure in children and adolescents: how to adjust for differences in body size and age. *The American journal of clinical nutrition*, *79*(5), 851–856.

Elia, M., & Livesey, G. (1991). Energy expenditure and fuel selection in biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. *World review of nutrition and dietetics*, *70*, 68–131.

Eston, R. G., Rowlands, A. V., & Ingledew, D. K. (1998). Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children’s activities. *Journal of Applied Physiology*, *84*(1), 362–371.

Freedson, P., Lyden, K., Kozey-Keadle, S., & Staudenmayer, J. (2012). Reply to Bonomi and Plasqui. *Journal of Applied Physiology*, *112* (5), 933–933. doi:10.1152/japplphysiol.01546.2011

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of Accelerometer Output for Children. *Medicine & Science in Sports & Exercise*, *37* (Supplement), S523–S530. doi:10.1249/01.mss.0000185658.28284.ba

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise*, *30*(5), 777–781. doi:10.1097/00005768-199805000-00021

Fuchs, R. K., Bauer, J. J., & Snow, C. M. (2001). Jumping Improves Hip and Lumbar Spine Bone Mass in Prepubescent Children: A Randomized Controlled Trial. *Journal of Bone and Mineral Research*, *16*(1), 148–156. doi:10.1359/jbmr.2001.16.1.148

Gyllensten, I. C., & Bonomi, A. G. (2011). Identifying Types of Physical Activity With a Single Accelerometer: Evaluating Laboratory-trained Algorithms in Daily Life. *IEEE Transactions on Biomedical Engineering*, *58*(9), 2656–2663. doi:10.1109/TBME.2011.2160723

Hänggi, J. M., Phillips, L. R. S., & Rowlands, A. V. (2013). Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. *Journal of Science and Medicine in Sport*, *16*(1), 40–44. doi:10.1016/j.jsams.2012.05.012

Hills, A. P., Andersen, L. B., & Byrne, N. M. (2011). Physical activity and obesity in children. *British Journal of Sports Medicine*, *45* (11), 866–870. doi:10.1136/bjsports-2011-090199

Janz, K. F. (1994). Validation of the CSA accelerometer for assessing children's physical activity. *Medicine and science in sports and exercise*, *26*(3), 369–375.

John, D., & Freedson, P. (2012). ActiGraph and Actical Physical Activity Monitors. *Medicine & Science in Sports & Exercise*, *44*, p.86–S89. doi:10.1249/MSS.0b013e3182399f5e

Liu, S., Gao, R. X., & Freedson, P. S. (2012). Computational methods for estimating energy expenditure in human physical activities. *Medicine and science in sports and exercise*, *44*(11), 2138–2146. doi:10.1249/MSS.0b013e31825e825a

Macfarlane, D. J., & Wong, P. (2012). Validity, reliability and stability of the portable Cortex Metamax 3B gas analysis system. *European Journal of Applied Physiology*, *112*(7), 2539–2547. doi:10.1007/s00421-011-2230-7

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, *10* (1), 213. doi:10.1186/1471-2105-10-213

Nilsson, A., Anderssen, S. A., Andersen, L. B., Froberg, K., Riddoch, C., Sardinha, L. B., & Ekelund, U. (2009). Between- and within-day variability in physical activity and inactivity in 9- and 15-year-old European children. *Scandinavian Journal of Medicine & Science in Sports*, *19* (1), 10–18. doi:10.1111/j.1600-0838.2007.00762.x

Pate RR, P. M. (1995). Physical activity and public health: A recommendation from the centers for disease control and prevention and the american college of sports medicine. *JAMA*, *273*(5), 402–407. doi:10.1001/jama.1995.03520290054029

Pearce, D. H., & Milhorn, H. T. (1977). Dynamic and steady-state respiratory responses to bicycle exercise. *Journal of Applied Physiology*, *42*(6), 959–967.

Pober, D. M., Staudenmayer, J., Raphael, C., & Freedson, P. S. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and science in sports and exercise*, *38*(9), 1626–1634. doi:10.1249/01.mss.0000227542.43669.45

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity research*, *10* (3), 150–157. doi:10.1038/oby.2002.24

Puyau, M. R., Adolph, A. L., Vohra, F. A., Zakeri, I., & Butte, N. F. (2004). Prediction of activity energy expenditure using accelerometers in children. *Medicine and science in sports and exercise*, *36* (9), 1625–1631.

Riddoch, C. J., Bo Andersen, L., Wedderkopp, N., Harro, M., Klasson-Heggebø, L., Sardinha, L. B., … Ekelund, U. (2004). Physical activity levels and patterns of 9- and 15-yr-old European children. *Medicine and science in sports and exercise*, *36* (1), 86–92. doi:10.1249/01.MSS.0000106174.43932.92

Rodríguez, G., Moreno, L. A., Sarría, A., Fleta, J., & Bueno, M. (2002). Resting energy expenditure in children and adolescents: agreement between calorimetry and prediction equations. *Clinical nutrition (Edinburgh, Scotland)*, *21*(3), 255–260.

Ruch, N., Rumo, M., & Mäder, U. (2011). Recognition of activities in children by two uniaxial accelerometers in free-living conditions. *European journal of applied physiology*, *111* (8), 1917–1927. doi:10.1007/s00421-011-1828-0

Schofield, W. (1984). Predicting basal metabolic rate, new standards and review of previous work. *Human nutrition. Clinical nutrition*, *39 Suppl 1*, 5–41.

Staudenmayer, J., Zhu, W., & Catellier, D. J. (2012). Statistical considerations in the analysis of accelerometry-based activity monitor data. *Medicine and science in sports and exercise*, *44*, S61–67. doi:10.1249/MSS.0b013e3182399e0f

Tanaka, C., Tanaka, S., Kawahara, J., & Midorikawa, T. (2007). Triaxial Accelerometry for Assessment of Physical Activity in Young Children. *Obesity*, *15*(5), 1233–1241. doi:10.1038/oby.2007.145

Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood: a 21-year tracking study. *American journal of preventive medicine*, *28* (3), 267–273. doi:10.1016/j.amepre.2004.12.003

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J., … Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine and science in sports and exercise*, *36*(7), 1259–1266.

Trost, S. G., Ward, D. S., Moorehead, S. M., Watson, P. D., Riner, W., & Burke, J. R. (1998). Validity of the computer science and applications (CSA) activity monitor in children. *Medicine and science in sports and exercise*, *30*(4), 629–633.

Trost, S. G., Wong, W.-K., Pfeiffer, K. A., & Zheng, Y. (2012). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and science in sports and exercise*, *44* (9), 1801–1809. doi:10.1249/MSS.0b013e318258ac11

Van Hees, V. T., Renstrom, F., Wright, A., Gradmark, A., Catt, M., Chen, K. Y., … Franks, P. W. (2011). Estimation of Daily Energy Expenditure in Pregnant and Non-Pregnant Women Using a Wrist-Worn Tri-Axial Accelerometer. *PLoS ONE*, *6*(7). doi:10.1371/journal.pone.0022922

Welk, G. J. (2005). Principles of Design and Analyses for the Calibration of Accelerometry-Based Activity Monitors. *Medicine & Science in Sports & Exercise*, *37* (Supplement), 501–S511. doi:10.1249/01.mss.0000185660.38335.de

Whipp, B. J., Ward, S. A., Lamarra, N., Davis, J. A., & Wasserman, K. (1982). Parameters of ventilatory and gas exchange dynamics during exercise. *Journal of Applied Physiology*, *52*(6), 1506–1513

An Application of SaTScan to Evaluate the Spatial Distribution of Corner Kick Goals in Major League Soccer

Robert H. Schmicker

Department of Biostatistics, University of Washington, Seattle, WA, USA 98105

Abstract

Identifying strategy that helps increase the number of goals in soccer is paramount to success. Opportunities exist to increase success in set pieces such as free kicks and corner kicks. We examined all corner kicks attempted during the 2010 Major League Soccer season to assess whether goals scored off of corner kicks were randomly spatially distributed. We separated the 18-yard box into 66, three by four yard boxes. A spatial scan statistic implemented in SaTScan, varied at-risk percentages to assess areas with higher than expected rates of goals scored. We examined data from 1859 corner kicks with an overall goal rate of 2.2%. A single box directly in the center of the box, 6-9 yards from goal was the only box with significantly higher rates of goals scored (5.0%) than expected. This result did not vary after adjusting for potential confounders' home field, kick trajectory or time of game. Our results, which consider goal rates, are consistent with previous research that suggest the most aggregate goals are scored in an area 6-12 yards from goal. Future research is needed to account for player movement on corner kick attempts.

KEY WORDS: CORNER KICKS, SPATIAL SCAN STATISTIC, GOAL RATES

Introduction

Goals are the most precious commodity in soccer. During the 2010 World Cup national teams averaged 1.1 goals/game; 73.4% of the 64 games either finished in a tie or were decided by one goal (FIFA, 2010). A single goal can be the difference between a loss and a tie or a tie and a win. In a tournament such as the World Cup, where the top two point earners from eight 4-team groups advance to a single elimination knockout phase, a single goal can be the difference between advancing in the tournament and being eliminated.

Soccer clubs are always in search of ways to score goals. Goals can be scored in numerous ways, but normally result from situations where the offensive team has an advantage over the defensive team. Such situations are scarce in soccer; finding a way to increase the number of advantages per game may be critical to success. Examples of such situations are set pieces, such as direct free kicks and corner kicks, which can result in higher quality shots on goal than during the run of play.

During the three most recent high profile international tournaments – the 2006 and 2010 World Cups and 2008 European Championships – there were between 9.8 and 10.2 corner kicks a game (FIFA, 2010; FIFA, 2006; Dunn, 2009; Baranda, Lopez-Riquelme & Ortega, 2011) with a goal scoring rate from corner kicks between 1.3% and 2.0% (FIFA, 2010; FIFA, 2006; Dunn, 2009; Baranda et al., 2011). The average number of corner kick attempts per game during these tournaments has been consistent in domestic leagues such as the English Premier League, the German Bundesliga, Italy's Series A and Spain's La Liga (Anderson, 2011a; Rudd, 2011; Lago-Penas, Lago-Ballesteros, Dellal & Gomez, 2010). Anderson and Rudd (2011) also estimate that the goal rate from corners is consistent with the English Premier League (EPL). Anderson (2011b) concludes that one goal per every 10 games makes corner kicks "useless in terms of scoring goals," a concept supported by the abandonment of long corner kicks in favor of short kicks by several teams, most notably FC Barcelona. Previous research has shown that 25%-40% of goals are scored off set pieces like corners (Taylor, James & Mellalieu, 2005; Dunn, 2009;Baranda & Lopez-Riquelme, 2012), suggesting that even though such opportunities may be inefficient, the magnitude of a single goal scored off a corner kick attempt can be of huge important.

The majority of shots on goal directly from corner kicks are headers (Baranda et al., 2011). Because players cannot achieve the same amount of power and accuracy from a header as from a kick, corner kicks closer to the goal are generally preferred. Previous literature has shown the greatest number of corner kicks are attempted in an area that is 6-12 yards directly in front of the goal box (Taylor et al., 2005), due mostly to the fact that corners kicks too close to the goal will get intercepted by the goalie. Because the location of goal kicks attempts vary spatially, it is possible that the highest goal rates may not be consistent with the highest number of goals.

We are interested if there are particular characteristics of corner kicks that are more likely to result in a goal. These can include corner kick location, game situation, player characteristics or spatial movement. For this study, we aim to describe corner kick attempts and then to assess whether corner kicks goals are randomly distributed spatially. We hypothesize that goals from corner kick attempts are randomly distributed across space. We use data from a single soccer league and do not account for player characteristics such as height, jumping ability or speed. We consider the spatial location of corner kick attempts as a snapshot in time.

Methods

Data

Major League Soccer (MLS) is the premier professional soccer league in the United States. In 2010, all 16 teams played each other twice - once at home and once on the road for a total of 240 regular season games. We used match reports from MLSsoccer.com to obtain the timing of all corner kick attempts for all 240 games. Between June 2010 and August 2011, we used the MLS Video Match Live (recently renamed to MLS Live, http://live.mlssoccer.com/mlsmdl/) service to visually review every corner kick attempted in a game. Video recordings existed for 239 of the 240 games.

The primary outcome was whether a goal was scored prior to change of possession. A change of possession was recorded at the time when the team defending against the corner kick first touched the ball. Our primary predictor was the location of the ball when initially touched by either team or the ground. Without proper graphing software, determination of the precise location was inaccurate. Using a technique pioneered by Hughes et al (1997), we divided the 18-yard box into several unique, non-overlapping boxes. Our analysis is more granular than those of Taylor et al. (2005), Dunn (2009) and Ergesoy et al. (2007) in that we create 66 distinct 3yard by 4yard boxes (Figure 1) compared to 5 unequal segments (Taylor et al, 2005). Each box was distinguishable through use of natural field markings like the 18 yard box, penalty spot and the 6 yard box.

Figure 1: The separation of the 18-yard box into 66 boxes

In addition to spatial location, we recorded additional data on every corner kick: the team attempting the corner kick, the difference in score at the time of the corner, the minute epoch of the corner and the trajectory of the kick (those with an arc that curved towards the goalie – inswing; those with an arc that curved away from the goal - outswing).

We conducted an intrarater reliability analysis on every goal scored and calculated the Cohen's kappa coefficient. The two reviews differed on 6 of the 40 goals, k=0.82.

Statistical Methods

We examined descriptive data on corner kicks and ran Chi-Square Goodness of Fit Tests to examine differences in corner kick goal rate across team, minute epoch and kick trajectories. The use of a Chi-Square Test to test whether goal rates differ by location would have two main drawbacks: 1) we would not be able to identify which box rejected the null hypothesis and 2) we would not be able to consider groups of boxes with higher than expected rates. To identify geospatial clustering of goals within the 18-yard box we use a spatial scan statistic as implemented in SaTScan™ (National Cancer Institute, Division of Cancer Control and Population Sciences, Statistical Research and Applications Branch Methods), whose methods have been discussed at length elsewhere. (Kuldorff, 1997; Kuldorff, 2006)

In brief, we define smaller, non-overlapping areas of varying size within the larger population. For each smaller area, we calculate a centroid for which all the areas attributes are attached. Population and area specific injury rates are then calculated.

At each centroid, SaTScan generates circular "windows" of increasing size until a pre-specified percentage of the total corner kick population (i.e., percent-at-risk) had been met. Clusters are created when centroids representing other areas fall within the boundaries of a window. Cluster specific injury rates are calculated and a primary cluster for the entire population is determined based on the cluster that is most likely not due to chance.

For the application to soccer corner kicks, we defined 66 distinct boxes within the 18 yard box. We generated centroids as the intersection of two diagonal lines within each box. We calculated an overall goal rate within the 18-yard box as well as individual goal rates for each box. Using a Poisson distribution, the null hypothesis is that the expected number of goals in each box is proportional to the population size. We chose to use the Poisson distribution for three reasons. First, we do not feel that the distribution of goals is inherently different than the distribution of the population. The cause of goals is not random - some goals may result from a deflection, but most long range shots from areas of low population will be blocked by one of the many bodies in the box. Second, Poisson distributions are preferable when adjusting for confounders as we will do in our secondary analyses. Finally, goal rates will be low, meaning that the Poisson distribution will be consistent with the Bernoulli. (Kuldorff, 1997)

For the Poisson model, the likelihood ratio is tested for significance using the Monte Carlo method. The likelihood ratio for a window constitutes the maximum likelihood ratio test statistic. Its distribution under the null hypothesis is obtained by repeating the same analytic exercise on a large number of random replications of the data set generated under the null hypothesis, with a *p* value obtained through Monte Carlo hypothesis testing. For this analysis, 999 Monte Carlo replications were used.

The likelihood function is maximized over all window locations and sizes, and the one with the maximum likelihood constitutes the most likely cluster. Because SaTScan is sensitive to the percent-at-risk parameter setting, (Chen, Roth, Naito, Lengerich & MacEachern, 2008) we varied the population percent-at-risk by the following levels: 50, 40, 30 and 20 percent. We did not consider 10 percent as two boxes would have been excluded due to their population exceeding 10% of the entire area. “Core clusters” were then identified as clusters of boxes that existed in all iterations.

Results

During the 2010 MLS Season, teams averaged 1.2 goals per game; 65.8% of games ended in a tie or were decided by one goal (MLS, 2010). Across the sample of 239 games, 2154 total corner kicks were assessed. We excluded 295 corner kicks (233 – short corners not kicked in the air into the box, 41 - technical problems such as no video recording or poor video recording or poor camera angle, 17 – kicked in the air but outside the box, 4 - fouls called while the ball was in the air) resulting in a population of 1859 corner kicks.

Forty (40) goals were scored off of corner kicks prior to possession change, resulting in an overall goal success rate of 2.2%. Home teams attempted more corner kicks (1035) than away teams (824). Corners that swung towards the goal accounted for 57% of all corner kicks attempted. Half of all corner kicks were attempted when the score of the game was tied. (Table 1)

Table 1: Corner Kick Characteristics

Games	239
Eligible Corner Kicks	1859
Attempted by home team, n (%)	1035 (55.7%)
In-Swing, n (%)	1060 (57.0%)
Score prior to corner	
-2 or higher, n (%)	149 (8.0%)
-1, n (%)	437 (23.5%)
0, n (%)	936 (50.3%)
1, n (%)	256 (13.8%)
+2 or higher, n (%)	81 (4.4%)
# Goals Scored	40

The total number of corner kicks attempted varied by team with a median of 116.5 and a range of 81 – 150. The median number of goals scored by team was 2.5 with a range from 0 to 6. Overall there was no significant difference in goal success rate across teams (p=0.14). Home teams scored a significantly lower percentage of attempts than away teams (1.5% vs. 2.9%, p <0.001). (Table 2)

Table 2: Goal Success Rates

	MLS
Home Corner Kicks	1035
Home Goals	16
Home Success Rate	1.5%
Away Corner Kicks	824
Away Goals	24
Away Success Rate	2.9%
Corner Kicks	1859
Goals	40
Success Rate	2.2%

The number of corner kicks attempted in every 10 minute period was consistent from minute 11 to minute 90. Goal success rates varied slightly across the 10 minute periods with a high after the 90 minute mark. (Table 3)

Table 3: Goal Success Rates, by Minute

Minute	Corner Kicks	Goals	Rate
1-10	175	3	1.7%
11-20	197	6	3.0%
21-30	200	4	2.0%
31-40	206	3	1.5%
41-50	207	1	0.5%
51-60	198	4	2.0%
61-70	200	4	2.0%
71-80	215	6	2.8%
81-90	197	6	3.0%
90+	64	3	4.7%

The majority (57%) of corner kicks were kicked with an in-swinging arc. The goal rate for in-swingers (2.0%) is slightly less than for out-swingers (2.4%). Ten of the 16 teams had a single player attempt more than half of its team's corner kicks. The remaining six teams averaged 8.5 corner kickers over the year. Goal rates for the two groups were 2.5% and 1.4% respectively.

Losing teams attempted more corner kicks with an in-swing trajectory (57%) and scored a significantly higher rate of goals (2.4% vs. 1.6%)

Figure 2 displays the population and goal rate by Box. Boxes 27, 39, 40 and 62 had the highest goal rates. There were no corner kick goals scored in 52 boxes; eight boxes had no corner kicks attempts. Box 39 is the only core box with higher than expected rates of goals after varying the percentage at risk (0.5, 0.4, 0.3, 0.2). A similar result was consistent within several subgroups: in-swing, out-swing, home, away (data not shown).

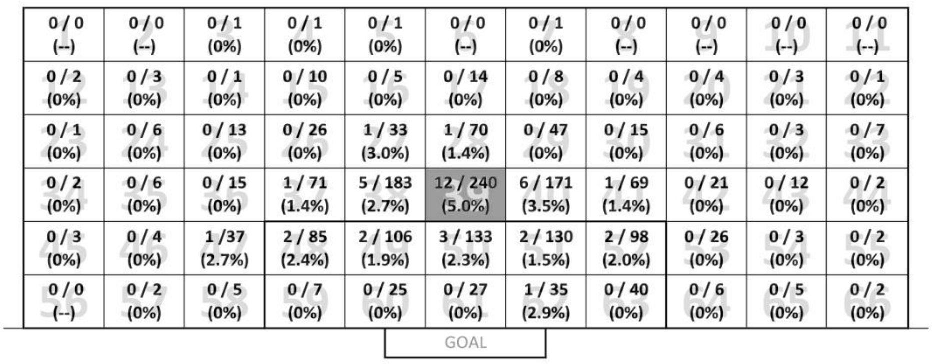


Figure 2: Goal rates for all 66 boxes with core clusters shaded in gray

Discussion

In our analysis, we examined the rate and location of goals scored off of corner kicks in a single domestic soccer league, Major League Soccer. We conclude that location of goals scored off of corner kicks are not random within the 18-yard box. In fact, one box(#39) located 6 to 9 yards from the goal line and 20 to 24 yards from the side of the 18-yard box had a higher than expected rate of goals in each repetition of our SaTScan statistic. This result is consistent with a previous analysis that showed that the highest number of goals is scored between the top of the 6-yard box and the penalty kick spot (12 yards from the goal). However, this analysis differs from Taylor et al. (2005) in that it considers neighboring boxes and not just the rate of goals for each box without respect to space. When individual rates are compared to the overall rate of the population, we were able to determine areas that have significantly higher rates of success than others. This eliminates any confounding due to varying denominators in each box.

Conventional wisdom suggests that the easiest way to score a goal is to be as close as possible to the goal. This may not be true since on corner kicks as goalies have a sizable advantage, being able to catch any balls kicked in their immediate vicinity. The fact that only 87 corner kicks (out of 1859) make first contact within 3 yards from the front of the goal is not surprising. Kicking a ball in an area where the goalie can catch it is simply an inefficient use of a corner kick. It is therefore not surprising that the highest number of goals and corner kicks are located in the center of the 18-yard box about 6 to 12 yards from the goal. Our analysis showed that a 3x4 area has a higher rate than what is expected by chance alone, suggesting that teams may want to focus greater effort in both practice and games on corner kick attempts.

Our analysis resulted in a few additional insights about corner kicks. First, the data show a moderate difference in success rates of corner kicks between teams playing at home and those on the road. Our data show that 56% of all corner kicks attempted were by the home team. However, this does not correspond to higher success rates, as home teams had a significantly lower ($p < 0.001$) success rate of 1.5% compared to 2.9% by the away team. We hypothesize that home teams are often more aggressive on offense than away teams because of the historical soccer strategy of playing to win games at home while playing for a tie on the road. We also hypothesize that away teams put more importance on the execution of their corner kick attempts since they often employ a more defensive strategy that results in limited offensive opportunities.

Second, the data show higher rates of goals off of corner kicks later in the game. The success rate for corner kicks attempted after the 70th minute was 3.2% compared to 1.8% in minutes 0-69. This rate was brought up by the 4.6% success rate in corner kicks taken after the 90th minute (i.e. – stoppage time). A large majority (81%) of goals scored off corner kick attempts after minute 70 were done so by teams either tied or trailing. As such, we hypothesize that teams are more focused on execution of corner kicks late in the game. However, it is also possible that their opponent is simply more tired at the end of the game and thus prone to poor defending, an idea presented by Dunn (2009), or that the random sequence of corner kick locations made it difficult to predict and thus prepare for.

Third, the data show that the majority of corners kicks are attempted in a way that the arc of the ball has an in-swinging trajectory (57%). The goal rate for in-swingers (2.0%) is slightly less than for out-swingers (2.4%). Unlike Baranda et al. (2011) and Taylor et al. (2005), who found a significantly higher goal rate for out-swingers than in-swingers, our data do not suggest a higher goal rate. Across MLS, 10 of 16 teams had 1 player attempt more than half of its team's corner kicks, with a goal rate of 2.5%. The other 6 teams, which averaged 8.5 corner kickers, had a goal rate of 1.4%. This latter group of teams primarily switched kickers to maximize in-swing attempts, suggesting that in fact this strategy may not be optimal. It is probable that the consistency of kicking method and ball trajectory achieved by a single player may be ideal. These data are in opposition to previous results on different leagues which suggest that in-swingers are more effective (Dunn, 2009)

When accounting for game status, losing teams attempted more corner kicks with an in-swing trajectory (57%) and scored a significantly higher rate of goals (2.4% vs. 1.6%) than for out-swing attempts. This is consistent with previous research that suggest teams alter the trajectory based on game status (Baranda et al., 2011). We hypothesize these changes in trajectory are a result of coaches' belief that the primary objective of in-swingers and out-swingers are different. Similarly, when winning, the rates of short corners are higher (Baranda & Lopez-Riquelme, 2012).

We feel that a strength of this analysis is the statistical method that examines the rate of goals rather than the aggregate goal count, which could simply be a function of the exposure to risk. In addition, categorizing the location of corner kick into 66 boxes provides a more robust analysis compared to previous attempts.

An obvious limitation is the lack of precision in our location data. We did not employ any graphing software but instead used natural line markings such as the 6 yard box, the penalty spot, the 18 yard box and the goal posts to estimate the location. Television broadcasts of MLS games often placed their cameras at midfield to capture all of the live action with the occasional replay. Our reliability analysis resulted in an intrarater reliability coefficient of 0.82, suggesting that there was high agreement between the two reviews. The data for the re-review were not different enough to cause a change in core clusters. While we only re-reviewed data for the 40 goals, we assume that any errors made on non-goals will also be random as there was no systematic bias.

The analyses preformed in SaTScan were dependent on the system used to divide the space into smaller regions. We divided the 18-yard box into unique boxes of the same size, each represented in the SaTScan analysis by its centroid. All centroids were equidistant from their neighbor centroid. As circles increased in size around each centroid, a minimum of 4 boxes were added every time. It was therefore possible to miss clusters of smaller size. To test this, we ran a sensitivity analysis where we added random noise of the form N(0,1) to each coordinate. In doing so, the results did not change – box 39 remains the core cluster. Even though it is a moot point, future analysis may want to avoid having equidistant centroids.

A final limitation is the absence of individual data such as height and vertical leap of the player making first contact, spatial movement within the box and flight characteristics of the ball. Our assumption that all players are equal is likely incorrect in that some players are more skilled at heading the ball than others. In ignoring spatial movement within the box, we are assuming that every corner attempted in the same box is exactly the same. The reality is that players start at one location within the box and run to another location in an attempt to arrive at corner kick before their defender does. Some players are not only more skilled at jumping and heading, but they are better at creating space between themselves and their defender. Finally, in ignoring the flight path of the ball prior to being touched by the first player, we assume that all kicks are equal. Two attempts that end in the same box can differ in height and trajectory such that the ball reacts differently in different situations.

If the above data were available, we could account for any confounding due to differences in player attributes and movement. We do not believe that taller players inherently move to the same location on all corner kick attempts. However, it is probable that the majority of the goals scored are not because of the location of the corner kick attempt but rather because of the mismatch between offensive and defensive player attributes. Data on potential confounders would aid in determining whether goals are due to location within the box or if they are simply due to player attribute.

Even though this analysis was performed on data from a single soccer league within the United States, we believe its results are applicable for other leagues in the world. Skill level will vary across leagues, but we feel that within leagues the variability in skill level is comparable. Some teams will be efficient with their use of corner kicks while others will not. However there is nothing to suggest that an analysis in any other league will yield different results. Goals off of corner kicks are not easy to come by. If teams want to maximize their use of corner kicks, then they should kick the ball within a 6 yard zone directly outside the 6 yard box.

Our results suggest a purely tactical observation which when combined with physical components may have a major effect on performance. As such, future work should address players' spatial movement within the box. Previous research in both ice hockey (Thomas, 2006) and soccer (Hirotsu, 2002; Rudd, 2011) have used a Markov process to examine the effect of starting state within a game on the number of goals. Taylor et al. (2008) define 13 on the ball behaviors in modeling technical behaviors with game status. If these methods can be adapted to this area of research this may aid in developing a method to assure that the player with the best skill at scoring goals off of corner kicks is in the best location is the next obvious step. If coaches are able to develop tactics that result in a succinct advantage, it has the potential to lead to more goals and thus more wins. At a current 2.0% success rate, there is great potential for improvement.

References

Anderson, C. (2011). The Distribution of Corners in the Big Leagues of Soccer: The 2010-11 Season to Date. In *Soccer By The Numbers.* Retrieved February 24, 2012, from http://www.soccerbythenumbers.com/2011/01/distribution-of-corners-in-big-leagues.html

Anderson, C. (2011). Why the Goal Value of Corners is (Almost) Nil: Evidence from the EPL. In *Soccer By The Numbers.* Retrieved June 14, 2012, from http://www.soccerbythenumbers.com/2011/05/why-goal-value-of-corners-is-almost.html

Baranda P.S., Lopez-Riquelme, D., Ortega, E., (2011). Criterios de eficacia ofensiva del saque de esquina en el mundial de alemania 2006: aplicaciones al entrenamiento. [Offensive performance criteria in the corner kick World Cup Germany 2006: Applications to training] In: Consejo General de Colegios Oficiales de Licenciados en Educación Física y en Ciencias de la Actividad Física y del Deporte. (Eds.) *Revista Espanola de Educacion Fisica y Deportes.* Número 395, Año LXIII, 4º trimestre, 2011 (nº 21, V época). 47-59.

Baranda, P.S., & Lopez-Riquelme, D. (2012). Analysis of corner kicks in relation to match status in the 2006 World Cup. *European Journal of Sport Science, 12*(2), 121-129.

Chen, J., Roth, R.E., Naito, A.T., Lengerich, E.J., MacEachern, A.M. (2008) Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. *International Journal of Health Geographics*, 7:18.

Dunn L. (2009). A Quantitative Analysis of Corner Kicks During UEFA Euro 2008, Austria & Switzerland. In the Video Analyst. Retrieved June 14, 2012, from http://www.thevideoanalyst.com/wp-content/uploads/2010/03/lee_dunn_article.jpg.

Egesoy, H. (2007). P-141 Comparison of corner kicks of host and visitor teams in

2005-2006 Super League and 3rd League matches in Turkey. Abstract presented at: VIth World Congress on Science and Football, Antayla, Turkey. *Journal of Sports Science and Medicine, Supplement. 10*, 207.

FIFA 2010 World Cup Scores (2010). Retrieved from http://www.fifa.com/worldcup/archive/southafrica2010/matches/groupstage.html

FIFA 2006 World Cup Scores (2006). Retrieved from http://www.fifa.com/worldcup/archive/germany2006/results/index.html

Hirotsu, N., & Wright, M. (2002). Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society, 53*(1174).

Hughes, M.D. and Franks, I.M. (1997). *Notational Analysis of Sport.* London: E. & F.N. Spon.

Kuldorff, M. (1997). A spatial scan statistic. *Communications in Statistics – Theory and Methods, 26,* 1481-1496.

Kuldorff, M. (2006) Information Management Services 1: SaTScan v7.0: Software for spatial and space-time scan statistics.

Lago-Penas, C., Lago-Ballesteros, J., Dellal A., Gomez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. *Journal of Sports Science and Medicine, 9*, 288-293.

MLS Soccer 2010 Schedule. Retrieved from http://www.mlssoccer.com/schedule?month=all&year=2010&club=all

Rudd, S. (2011). A framework for tactical analysis and individual offensive production assessment in soccer using Markov chains [PowerPoint slides]. Paper presented at the *2011 New England Symposium on Statistic in Sports*, Cambridge, MA. Retrieved from http://www.amstat.org/chapters/boston/nessis11/presentation_material/rudd.pdf

Taylor, J.B., James, N, Mellalien, S.D. (2005). Notational analysis of corner kicks in English Premier League soccer. In T Reilly, D Aranjo & J Cabri (Eds.), *Science and Football V: the Proceedings of the Fifth World Congress on Football.*, pp 229-234. Abingdon, Oxon: Routledge.

Taylor, J.B., Mellalieu, S.D., James, N., & Shearer, D.A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *Journal of Sports Sciences, 26,* 885-895.

Thomas, A. (2006). The Impact of Puck Possession and Location on Ice Hockey Strategy. *Journal of Quantitative Analysis in Sports, 2,* Article 6.