

# On the Power of Modern Statistical Methodology in Soccer Analysis

## Archetypal soccer players

Manuel J. A. Eugster  
with Jasmin Abedieh, Christian Schnell, and Thomas Augustin

Institut für Statistik  
Ludwig-Maximilians-Universität München

Statistische Woche, Universität Leipzig, 2011

DIGITALE SPIELANALYSE

# Ein Spiel ist drei Gigabyte groß

Analysten und ihr technologisches Wettrüsten in der Fußballbundesliga

von David Binnig | 04. August 2011 - 08:00 Uhr

© Alexander Hassenstein/Bongarts/Getty Images



DIGITALE SPIELANALYSE

## Ein Spiel ist drei Gigabyte groß

Analysten und ihr technologisches Wettrüsten in der  
Fußball-Bildschirmwelt

VON D

[...] sagt Martin Lames, "was nützt es mir, zu wissen,  
wer wie viele Prozent der Zweikämpfe gewonnen hat?  
Das sind Werte für Journalisten."

ges



### Forum

#### Inland

Keine Einigung im Fluglärmstreit  
Schaffhausen rutscht tief in die roten Zahlen  
Krankenkassenprämien steigen eher mässig  
Gaskombikraftwerke sollen im Ausland kompensieren

#### Wirtschaft

Radio- und Fernsehgebühren für alle  
Leuthard sieht eine Zukunft für Printmedien  
«Eine Gesellschaft der neuen Energien bauen»  
Dubiose Datensammler jagen Schweizer Unternehmen

#### Ausland

Walliser Raclette künftig auch in der EU geschützt  
Griechenland entlässt Staatsangestellte  
Und er stand auf und ging dahin  
Mehrere Tote bei Razzien in Syrien

#### Sport

Luzerner Justiz verurteilt 13 Krawallmacher  
Hoffnungsschimmer für den FC Sion: Teilsieg erreicht  
Besser aus den Startlöchern kommen  
Biel drei Monate ohne Gloor

#### Kultur

# Hitzfeld wie Ramsey: Never change a winning team

**Ottmar Hitzfeld lässt Spekulationen zur Startformation zu, doch: Am Ende bleibt wohl alles gleich.**

fest / Quelle: [Si](#) / Montag, 30. März 2009 / 22:30 h

Am Montag liess Ottmar Hitzfeld zweimal trainieren, wobei am Morgen nur die Ersatzspieler von Chisinau im Einsatz waren. Der Eindruck ist, dass der Nationalcoach am Mittwoch gegen Moldawien den gleichen elf Spielern vertraut wie beim 2:0 vom Samstag. Der Coach offiziisierte seine Wahl diesmal jedoch nicht frühzeitig. Sir Alf Ramsey, Englands Weltmeister-Coach von 1966, hatte eine der berühmtesten Weisheiten des Fussballs geprägt: never change a winning team. So abgedroschen die Phrase klingen mag, so sehr verhalten sich nach wie vor viele Trainer nach diesem Grundsatz. Ob Ottmar Hitzfeld auch zu ihnen gehört, liess er zwei Tage vor dem Rückspiel gegen Moldawien offen. Anders als letzte Woche, in den Tagen vor der Partie in Chisinau, hat sich der Coach 48 Stunden vor dem Spiel (offiziell) noch nicht auf eine Formation festgelegt.



## Oft gelesene Meldungen

Ticker: Biel besiegt Chiasso  
Hoffnungsschimmer für den FC Sion: Teilsieg erreicht  
Djokovic gewinnt auch das US Open  
FC Barcelona gegen AC Milan - Giganten im Direktduell

## Nati

«Es liegen spannende Wochen vor mir» Mit dem 3:1 gegen Bulgarien hat die Schweizer Nationalmannschaft eine weitere Hürde Richtung EM-Endrunde genommen. Coach Ottmar Hitzfeld gab seinen Spielern eine klare Botschaft für die nächsten vier Wochen auf den Heimweg.  
**Fortsetzung**

«Wir sind auf dem besten Weg» Nach dem 3:1-Sieg über Bulgarien herrschte beim Schweizer Team grosse Zufriedenheit. Selbst Bulgaren-Coach Lothar Matthäus musste eingestehen, dass der Erfolg verdient war. **Fortsetzung**

## International

FC Barcelona gegen AC Milan - Giganten im Direktduell Gleich zum Auftakt der Gruppenspiele der Champions League treffen heute Abend Titelverteidiger FC Barcelona und die AC Milan aufeinander. **Fortsetzung**

Reto Ziegler erstmals für Fenerbahce Der Schweizer Internationale Reto Ziegler debütierte am Montagabend für

Keine Einigung im Fluglärmstreit  
Schaffhausen rutscht tief in die roten Zahlen  
Krankenkassenprämien steigen eher mässig  
Gaskombikraftwerke sollen im Ausland kompensieren

Radio- un alle  
Leuthard  
Printmedi «Eine Ges Energien b Dubiose I Schweizer

Walliser R der EU ges Griechenland entlässt Staatsangestellte Und er stand auf und ging dahin Mehrere Tote bei Razzien in Syrien

Luzerner Justiz verurteilt 13 Krawallmacher Hoffnungsschimmer für den FC Sion: Teilsieg erreicht Besser aus den Startlöchern kommen Biel drei Monate ohne Gloor

# Hitzfeld wie Ramsey: Never change a winning team

**Ottmar Hitzfeld lässt Spekulationen zur Startformation zu, doch: Am Ende bleibt wohl alles gleich.**

fest / Quelle: [Si](#) / Montag, 30. März 2009 / 22:30 h

Am Montag liess Ottmar Hitzfeld zweimal trainieren, wobei am Morgen nur die Ersatzspieler von Chisinau im Einsatz waren. Der Eindruck ist, dass der Nationalcoach

## Logistic modeling and structural changes

(\*) See Schnell (2011)

dem Artikel geht es um die Partie gegen Rumänien letzte Woche, in den Tagen vor der Partie in Chisinau, hat sich der Coach 48 Stunden vor dem Spiel (offiziell) noch nicht auf eine Formation festgelegt.



## Oft gelesene Meldungen

Ticker: Biel besiegt Chiasso Hoffnungsschimmer für den FC Sion: Teilsieg erreicht Djokovic gewinnt auch das US Open FC Barcelona gegen AC Milan - Giganten im Direktduell

## Nati

«Es liegen spannende Wochen vor mir» Mit dem 3:1 gegen Bulgarien hat die Schweizer Mannschaft eine weitere A-Endrunde unter Ottmar Hitzfeld in Aussicht. Eine klare Siegesserie steht auf den nächsten vier Heimwegen.

sten Weg» er Bulgarien rizer Team Selbst Matthäus musste eingestehen, dass der Erfolg verdient war. Fortsetzung

## International

FC Barcelona gegen AC Milan - Giganten im Direktduell Gleich zum Auftakt der Gruppenspiele der Champions League treffen heute Abend Titelverteidiger FC Barcelona und die AC Milan aufeinander. Fortsetzung

Reto Ziegler erstmals für Fenerbahce Der Schweizer Internationale Reto Ziegler debütierte am Montagabend für



Monday, October 11, 2010  
Updated: October 21, 9:41 AM ET

## The best league in the world?

By Leander Schaerlaeckens  
ESPN.com

The Germans, to borrow a cliché, aren't so much known for being unpredictable. The country's soccer has never been confused for being high on octane, scoring and entertainment. But all bets are off in Europe's healthiest soccer league.

In its first eight rounds of topsy-turvy play this season, the Bundesliga has seen Michael Bradley's Borussia Monchengladbach beat Bayer Leverkusen 6-3 before losing its next games by 4-0 to Eintracht Frankfurt (which had yet to garner any other points) and 7-0 to Stuttgart (a team that has mustered just one draw from its seven other games). Two matches later, Stuttgart lost 4-1 to the very Leverkusen that had gotten trounced by Gladbach. Meanwhile, powerhouses Schalke 04 and defending champions Bayern Munchen languish in second-to-last and 10th place, respectively. Mainz, which came in ninth last year and wasn't even in the Bundesliga the season before that, sat at the top of the table with an unblemished record until this past weekend.



The Bundesliga is the best-attended league in Europe, thanks in part to Franck Ribery's Bayern Munich.

While the English Premier League is seen as the gold standard, a scenario is unfolding in which the Bundesliga could surpass it in attendance numbers. If the trend continues, with the Bundesliga increasing its lead over the EPL, the



Monday, October 11, 2010  
Updated: October 21, 9:41 AM ET

## The best league in the world?

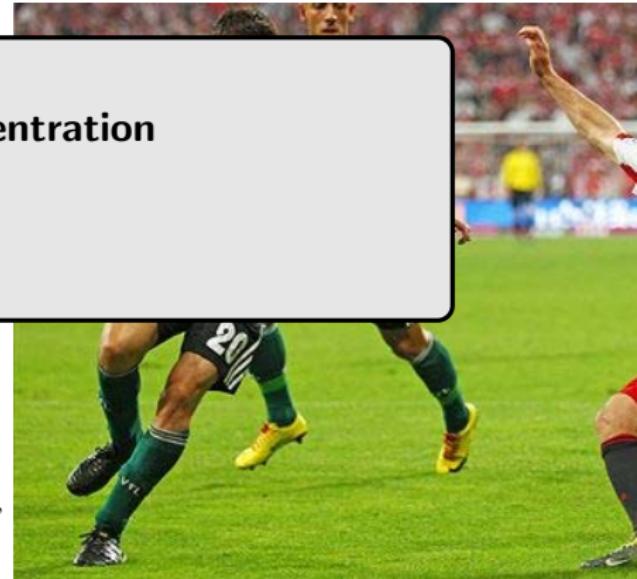
By Leander Schaerlaeckens  
ESPN.com

The Germans,  
being unpredict-  
able and  
confused for the  
first time in  
years, are

In its first eight matches of the season, the Bundesliga has been dominated by Borussia Monchengladbach, which has won its next game 4-0 against Hoffenheim (not garnering any other points) and 7-0 to Stuttgart (a team that has mustered just one draw from its seven other games). Two matches later, Stuttgart lost 4-1 to the very Leverkusen that had gotten trounced by Gladbach. Meanwhile, powerhouses Schalke 04 and defending champions Bayern Munchen languish in second-to-last and 10th place, respectively. Mainz, which came in ninth last year and wasn't even in the Bundesliga the season before that, sat at the top of the table with an unblemished record until this past weekend.

### Market concentration

(\*) See Abedieh (2011)



The Bundesliga is the best-attended league in Europe, thanks in part to Franck Ribery's Bayern Munich.

While the English Premier League is seen as the gold standard, a scenario is unfolding in which the Bundesliga could surpass it in attendance numbers. If that's the case, it would provide another reason to consider the Bundesliga as the

## Is Cristiano Ronaldo the world's best player?

---



By Vicki Hodges

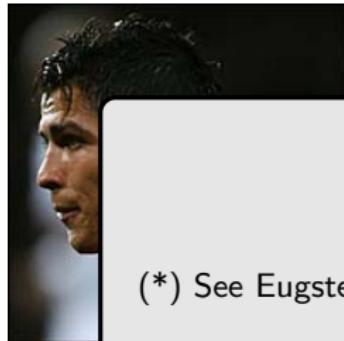
2:11PM GMT 20 Mar 2008

Managers, players and football pundits are falling over themselves to lavish praise on man-of-the-moment Cristiano Ronaldo.

The Manchester United winger took his goal tally to 33 for the season with a two-goal salvo against Bolton and the tributes labelling him as a footballing "genius" have spread far and wide.

Ronaldo's team-mate Nani has hailed the 23-year-old as the world's best player on current form. But will the

## Is Cristiano Ronaldo the world's best player?



### Archetypal analysis

(\*) See Eugster (2011)

By Vicki Hodges

2:11PM GMT 20 Mar 2008

Managers, players and football pundits are falling over themselves to lavish praise on man-of-the-moment Cristiano Ronaldo.

The Manchester United winger took his goal tally to 33 for the season with a two-goal salvo against Bolton and the tributes labelling him as a footballing "genius" have spread far and wide.

Ronaldo's team-mate Nani has hailed the 23-year-old as the world's best player on current form. But will the

## **Skill ratings from the PES Stats Database:**

Community-based data base to create accurate statistics for soccer players (originally for the Pro Evolution Soccer Series).

The skills are rated from 0 to 100. Note that we assume that the differences are interpretable, i.e., the ratings are on a ratio scale.

The extracted data set consists of 25 skill ratings of 1658 players from the *Bundesliga*, *La Liga*, *Premier League*, and *Serie A* (collected on September 16, 2011).

# Archetypal analysis

**Archetypal analysis** (Cutler and Breiman, 1994) has the aim to find a few, not necessarily observed, extremal observations (the archetypes) in a multivariate data set such that

1. all the observations are convex combinations of the archetypes, and
2. all the archetypes are convex combinations of the observations.

## Optimization problem:

Consider an  $n \times m$  matrix  $X$ . For given  $k$ , the archetypal problem is to find the matrix  $Z$  of  $k$   $m$ -dimensional archetypes. More precisely, to find the two  $n \times k$  coefficient matrices  $\alpha$  and  $\beta$  which minimize the residual sum of squares

$$\text{RSS} = \|X - \alpha Z^\top\|_2 \text{ with } Z = X^\top \beta$$

subject to the constraints

$$\sum_{j=1}^k \alpha_{ij} = 1 \text{ with } \alpha_{ij} \geq 0 \text{ and } i = 1, \dots, n,$$

$$\sum_{i=1}^n \beta_{ji} = 1 \text{ with } \beta_{ji} \geq 0 \text{ and } j = 1, \dots, k.$$

### **Approximation of the convex hull:**

The archetypes lie on the boundary of the convex hull of the data; let  $N$  be the number of data points which define the boundary of the convex hull, then if ...

$1 < k < N$ , there are  $k$  archetypes on the boundary which minimize RSS;

$k = N$ , exactly the data points which define the convex hull are the archetypes with  $\text{RSS} = 0$ ;

$k = 1$ , the sample mean minimizes the RSS.

**Algorithm:**

Cutler and Breiman (1994) present an alternating constrained least squares algorithm: it alternates between finding the best  $\alpha$  for given archetypes  $Z$  and finding the best archetypes  $Z$  for given  $\alpha$ :

### **Algorithm:**

Cutler and Breiman (1994) present an alternating constrained least squares algorithm: it alternates between finding the best  $\alpha$  for given archetypes  $Z$  and finding the best archetypes  $Z$  for given  $\alpha$ :

- At each step  $n + k$  convex least squares problems are solved.

### **Algorithm:**

Cutler and Breiman (1994) present an alternating constrained least squares algorithm: it alternates between finding the best  $\alpha$  for given archetypes  $Z$  and finding the best archetypes  $Z$  for given  $\alpha$ :

- At each step  $n + k$  convex least squares problems are solved.
- The algorithm converges, but not necessarily against the global minimum; use different initializations.

## **Algorithm:**

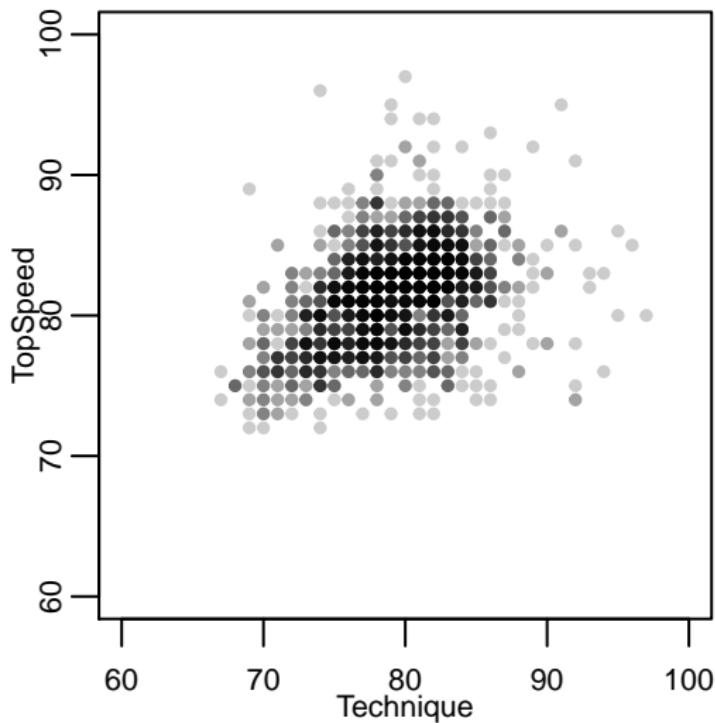
Cutler and Breiman (1994) present an alternating constrained least squares algorithm: it alternates between finding the best  $\alpha$  for given archetypes  $Z$  and finding the best archetypes  $Z$  for given  $\alpha$ :

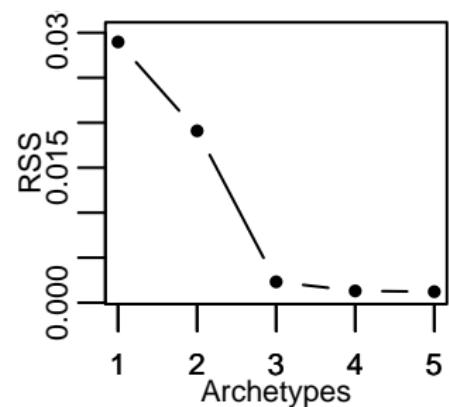
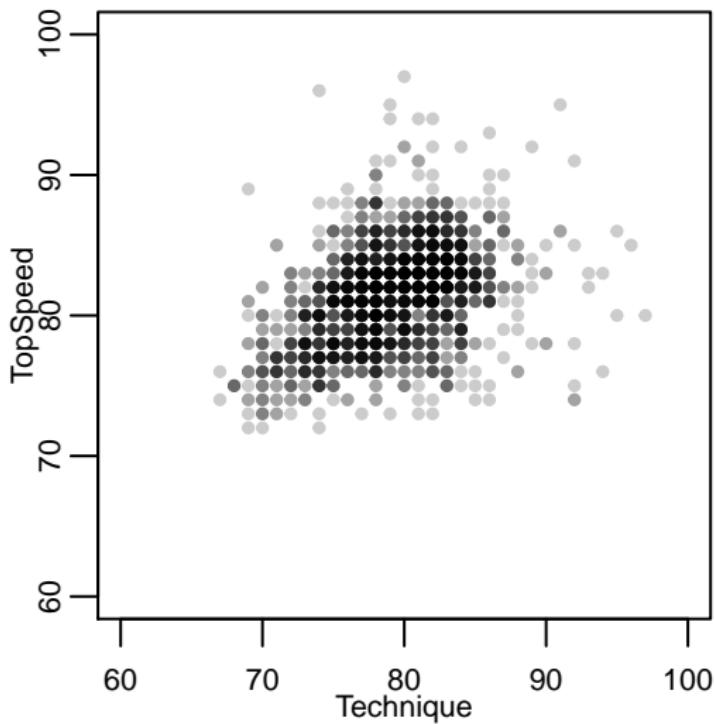
- At each step  $n + k$  convex least squares problems are solved.
- The algorithm converges, but not necessarily against the global minimum; use different initializations.
- There is no rule for the correct number of archetypes  $k$ ; use the “elbow criterion” to find an appropriate value  $k$ .

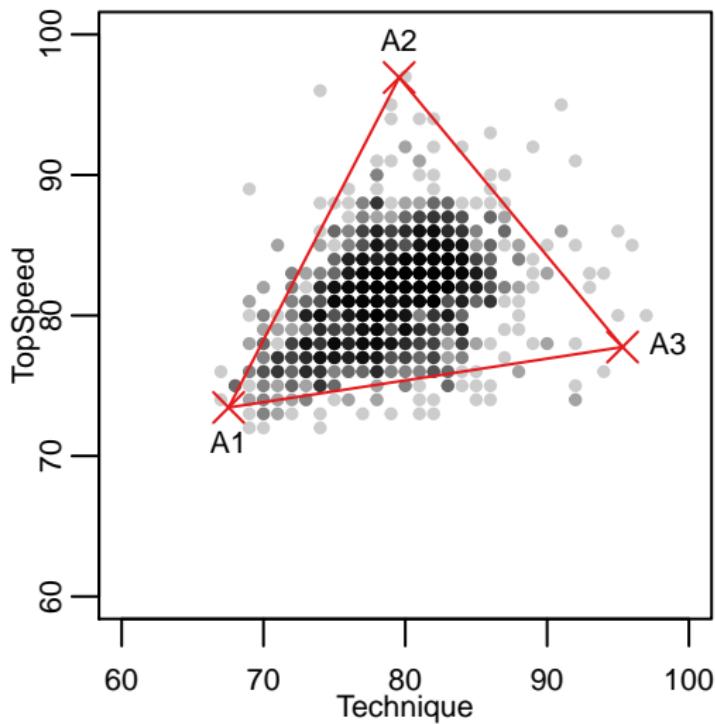
## **Algorithm:**

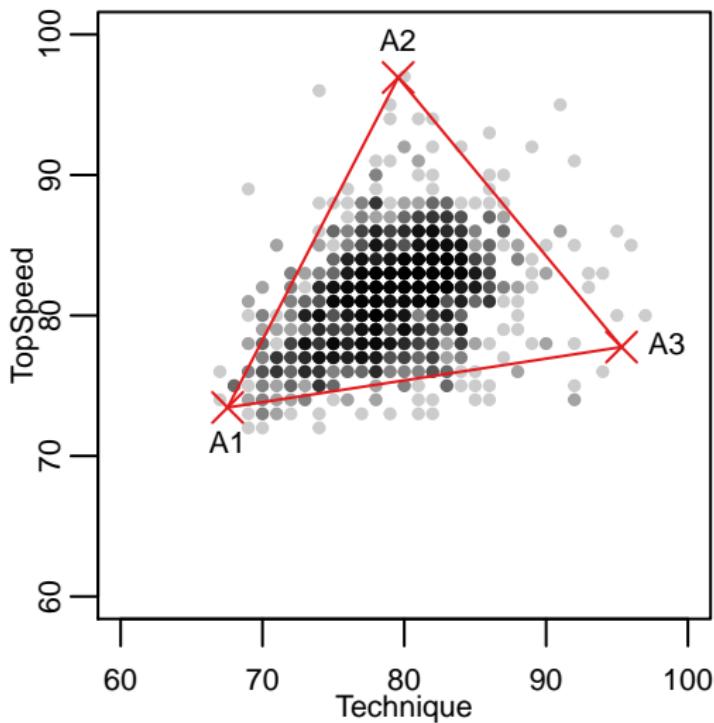
Cutler and Breiman (1994) present an alternating constrained least squares algorithm: it alternates between finding the best  $\alpha$  for given archetypes  $Z$  and finding the best archetypes  $Z$  for given  $\alpha$ :

- At each step  $n + k$  convex least squares problems are solved.
- The algorithm converges, but not necessarily against the global minimum; use different initializations.
- There is no rule for the correct number of archetypes  $k$ ; use the “elbow criterion” to find an appropriate value  $k$ .
- The method is (per definition) highly sensitive to “outliers”; weighted and robust extension by Eugster and Leisch (2011).

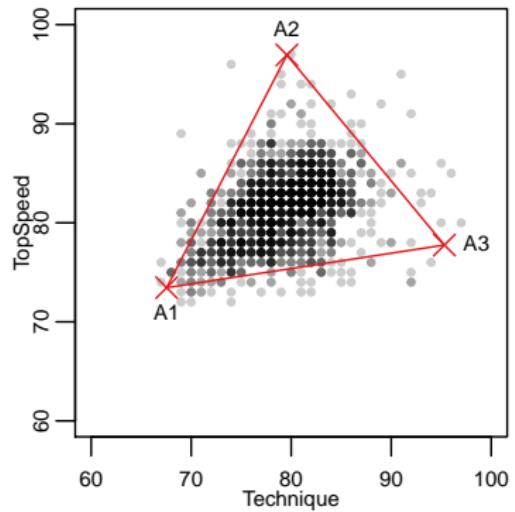
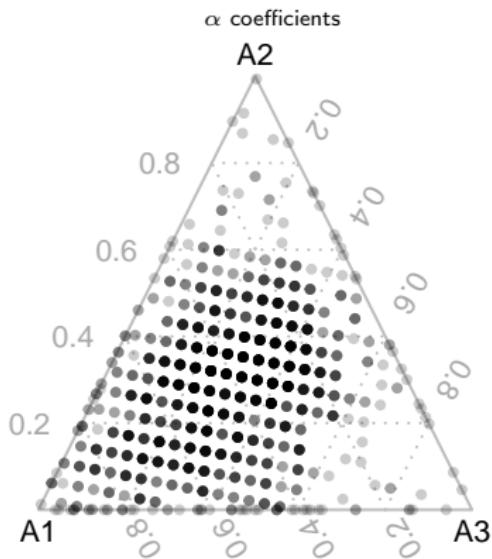


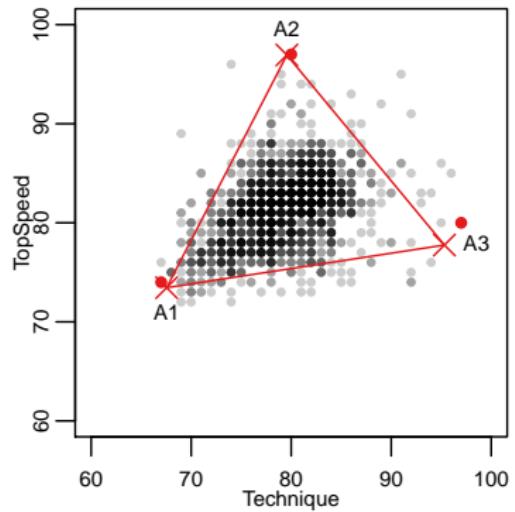
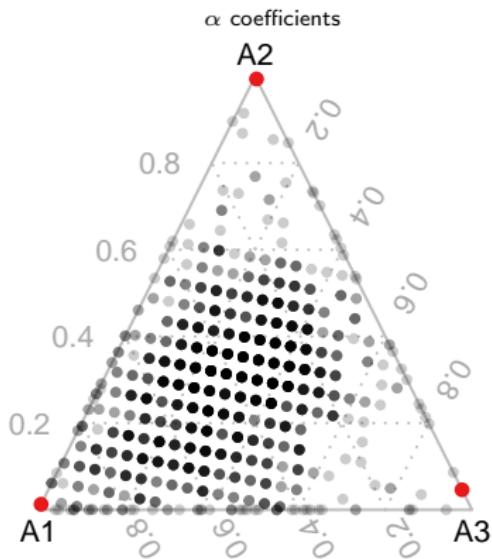




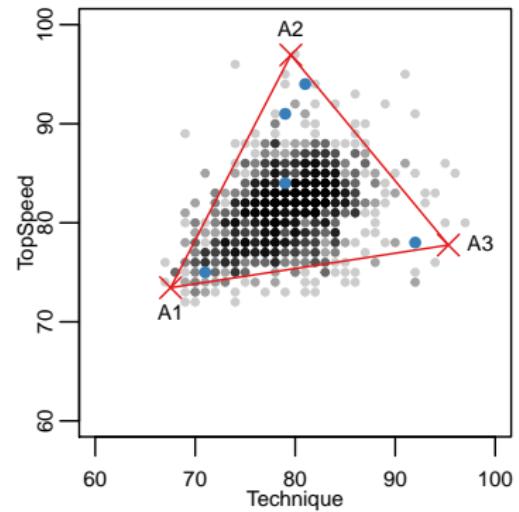
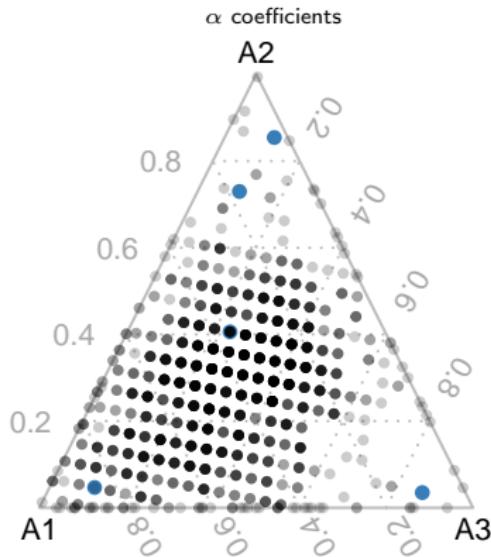


- A1** is the “slow and low ball control” archetype.
- A2** is the “fast and medium ball control” archetype.
- A3** is the “slow and high ball control” archetype.

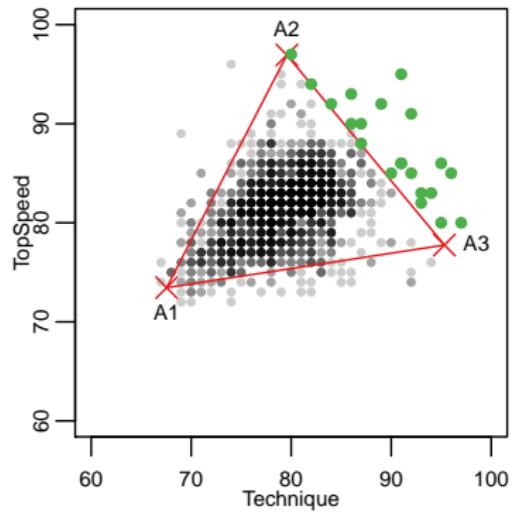
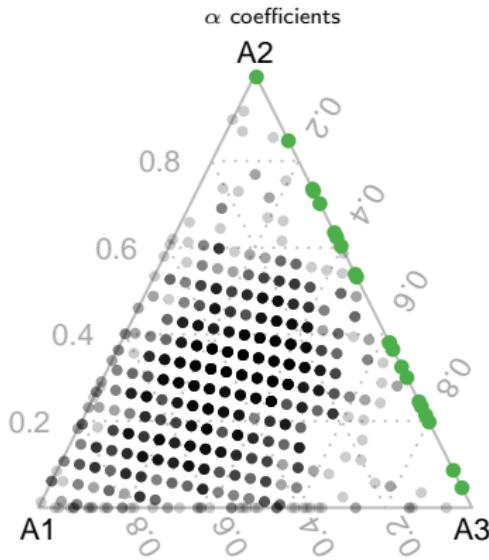




	Name	Club	Position	Technique	TopSpeed	$\alpha_{.1}$	$\alpha_{.2}$	$\alpha_{.3}$
A1	Claudio Terzi	AC Siena	Defender	67.00	74.00	0.99	0.01	0.00
A2	Theo Walcott	Arsenal FC	Forward	80.00	97.00	0.00	0.99	0.01
A3	Antonio Cassano	AC Milan	Forward	97.00	80.00	0.00	0.05	0.95



Name	Club	Position	Technique	TopSpeed	$\alpha_{.1}$	$\alpha_{.2}$	$\alpha_{.3}$
Michael Thurk	FC Augsburg	Forward	79.00	84.00	0.36	0.41	0.24
Slobodan Rajkovic	Hamburger SV	Defender	71.00	75.00	0.85	0.05	0.10
Gabriel Agbonlahor	Aston Villa FC	Forward	81.00	94.00	0.03	0.85	0.11
Alessandro Del Piero	Juventus FC	Forward	92.00	78.00	0.10	0.04	0.87
Pablo Armero	Udinese Calcio	Defender	79.00	91.00	0.17	0.73	0.10

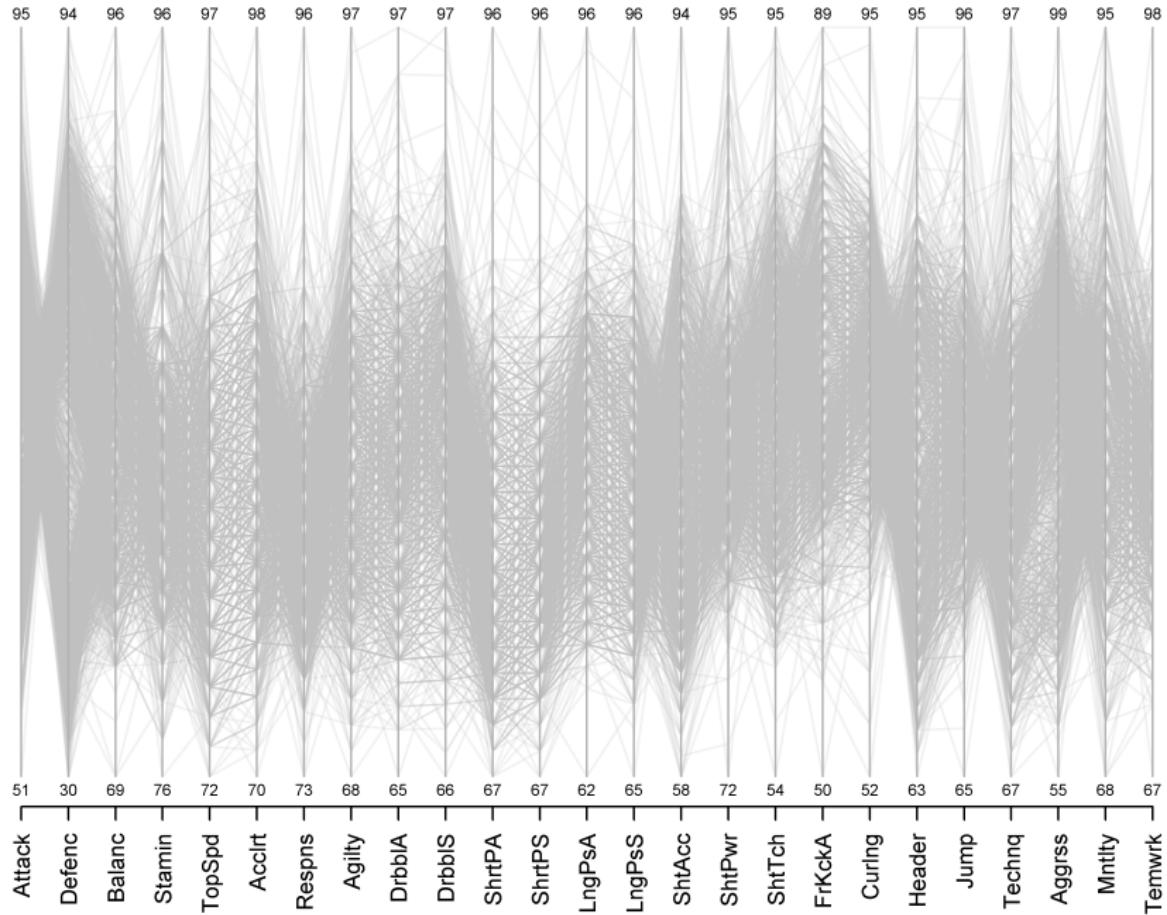


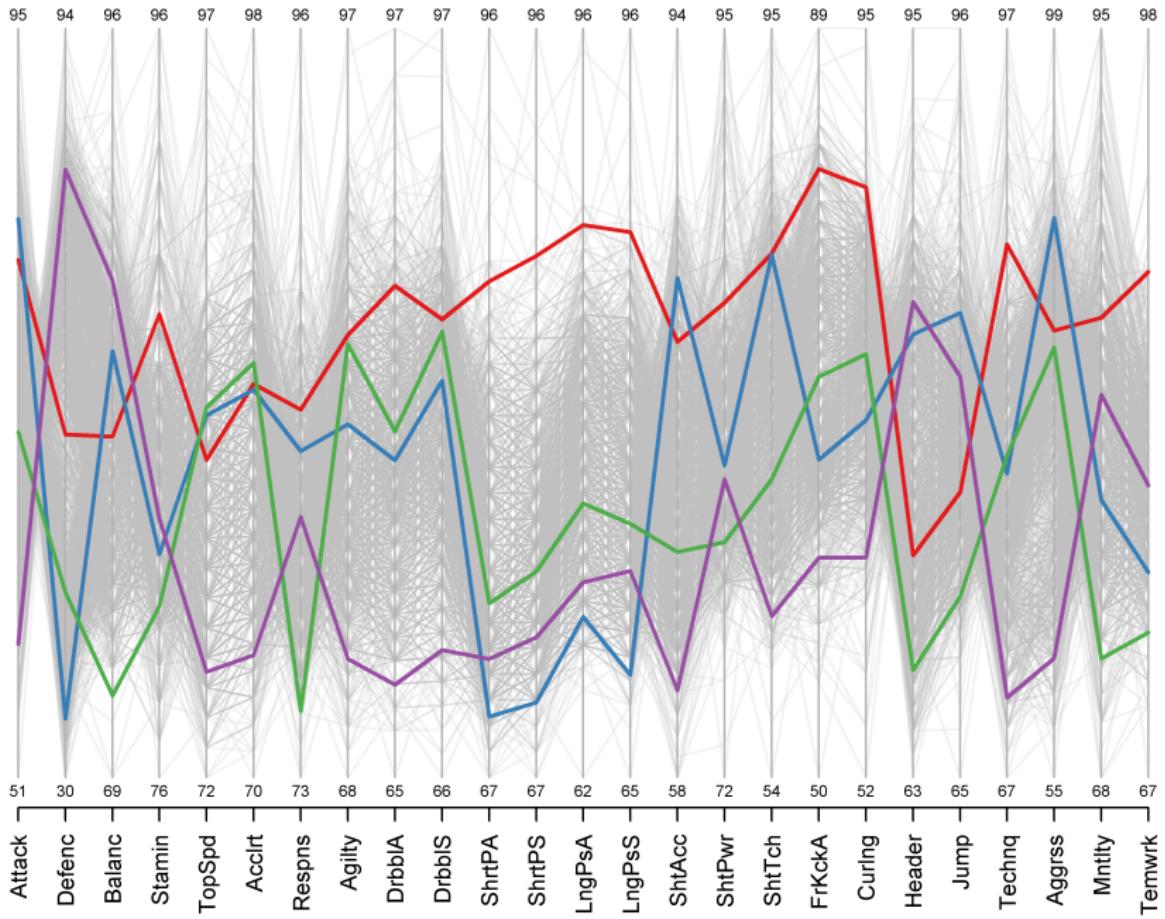
Name	Club	Position	Technique	TopSpeed	$\alpha_1$	$\alpha_2$	$\alpha_3$
Robin Van Persie	Arsenal FC	Forward	92.00	85.00	0.00	0.32	0.67
Jesus Navas	Sevilla FC	Midfielder	87.00	88.00	0.00	0.53	0.47
Cristiano Ronaldo	Real Madrid CF	Forward	91.00	95.00	0.00	0.70	0.30
Milos Krasic	Juventus FC	Midfielder	84.00	92.00	0.00	0.73	0.26
Alexandre Pato	AC Milan	Forward	87.00	90.00	0.00	0.60	0.40
Mesut Ozil	Real Madrid CF	Midfielder	90.00	85.00	0.00	0.36	0.63
Franck Ribery	Bayern Muenchen	Midfielder	92.00	91.00	0.00	0.54	0.46
Theo Walcott	Arsenal FC	Forward	80.00	97.00	0.00	0.99	0.01
Leo Messi	FC Barcelona	Forward	95.00	86.00	0.00	0.30	0.70
Gareth Bale	Tottenham Hotspur FC	Midfielder	82.00	94.00	0.00	0.85	0.15

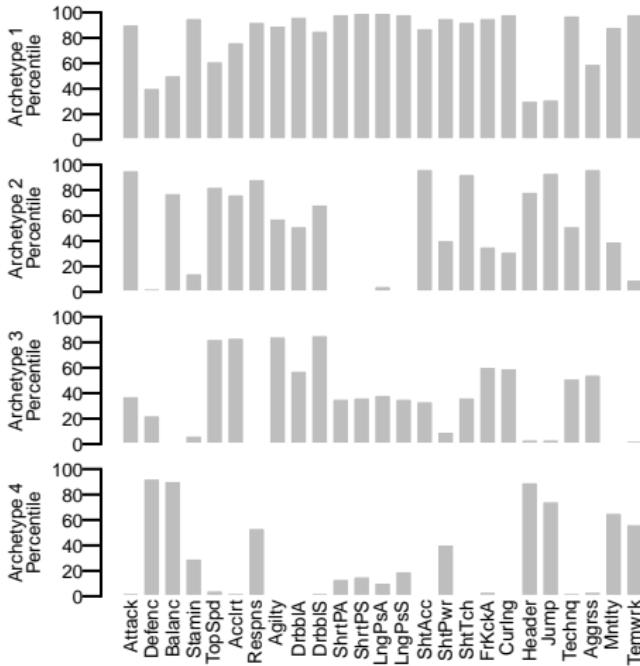
# Archetypal soccer players

or

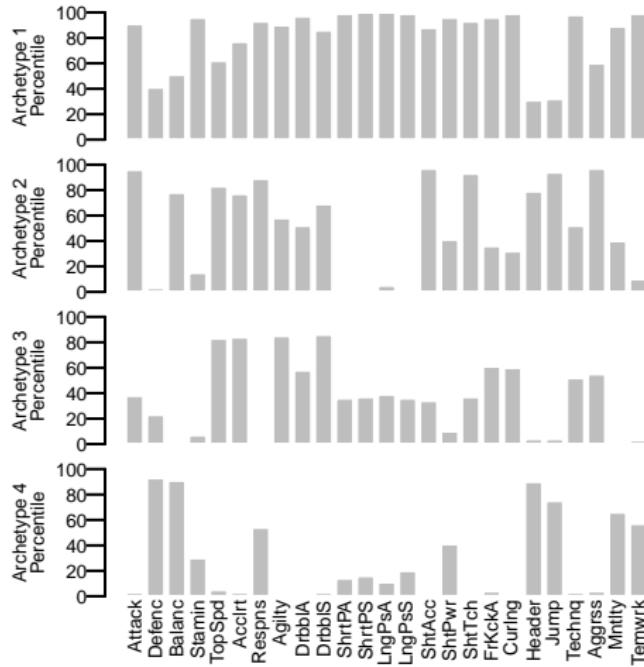
Is Cristiano Ronaldo the world's best player?







- A1** is the archetypal midfielder: all high expect defense, balance, header and jump.
- A2** is the archetypal center forward: high attack, shot acceleration, header and jump; low passing skills.
- A3** is the archetypal weak player.
- A4** is the archetypal defender: high defense, balance, header and jump.



**A1** is the archetypal midfielder: all high expect defense, balance, header and jump.

**A2** is the archetypal center forward: high attack, shot acceleration, header and jump; low passing skills.

**A3** is the archetypal weak player.

**A4** is the archetypal defender: high defense, balance, header and jump.

	Name	Club	$\alpha_{.1}$	$\alpha_{.2}$	$\alpha_{.3}$	$\alpha_{.4}$
A1	Xavi Hernandez	FC Barcelona	1.00	0.00	0.00	0.00
A2	Filippo Inzaghi	AC Milan	0.00	1.00	0.00	0.00
A3	Florin Matei	AC Cesena	0.00	0.00	1.00	0.00
A4	Massimo Paci	Novara Calcio	0.00	0.00	0.00	1.00

**Subjective definition:** The world's best player is a combination of Archetype 1 and Archetype 2 with Archetype 1 contributing more than Archetype 2.

Name	Club	$\alpha_{.1}$	$\alpha_{.2}$	$\alpha_{.3}$	$\alpha_{.4}$
Mladen Petric	Hamburger SV	0.53	0.47	0.00	0.00
Didier Drogba	Chelsea FC	0.52	0.48	0.00	0.00
Luis Suarez	Liverpool FC	0.53	0.47	0.00	0.00
Carlos Tivez	Manchester City FC	0.66	0.34	0.00	0.00
Dimitar Berbatov	Manchester United FC	0.60	0.40	0.00	0.00
Wayne Rooney	Manchester United FC	0.82	0.18	0.00	0.00
Xavi Hernandez	FC Barcelona	0.52	0.48	0.00	0.00
Leo Messi	FC Barcelona	0.79	0.21	0.00	0.00
Cristiano Ronaldo	Real Madrid CF	0.68	0.32	0.00	0.00
Giuseppe Rossi	Villarreal CF	0.51	0.49	0.00	0.00
Adrian Mutu	AC Cesena	0.60	0.40	0.00	0.00
Zlatan Ibrahimovic	AC Milan	0.54	0.46	0.00	0.00
Diego Forlan	FC Internazionale Milano	0.64	0.36	0.00	0.00
Antonio Di Natale	Udinese Calcio	0.67	0.33	0.00	0.00

# Summary

## Archetypal analysis:

- Estimate data-driven (multidimensional) extreme values, i.e., archetypal players.
- Identify archetypes as different types of “good” and “bad” players.
- Set observations in relation to the archetypes (using the  $\alpha$  coefficients)

**SportsAnalytics:** <https://r-forge.r-project.org/projects/soccer/>

**archetypes:** <http://cran.r-project.org/package=archetypes>

<http://www.statistik.lmu.de/~eugster/>

# References

- Jasmin Abedieh. Anwendungen von Konzentrationsmaße im Sport – Analyse von Ungleichheiten und  
mannschaftspezifischen Fähigkeiten im Fußball. Bachelorarbeit, Institut für Statistik,  
Ludwig-Maximilians-Universität München, Februar 2011. Supervised by Manuel J. A. Eugster and Thomas  
Augustin.
- David Binnig. Digitale Spielanalyse: Ein Spiel ist drei Gigabyte groß – Analysten und ihr technologisches  
Wettrüsten in der Fußballbundesliga. DIE ZEIT, Nr. 32, August 4, 2011. URL  
<http://www.zeit.de/2011/32/T-Videoanalyse-Fussball>.
- Adele Cutler and Leo Breiman. Archetypal analysis. *Technometrics*, 36(4):338–347, 1994.
- Manuel J. A. Eugster. Archetypal athletes. Working paper, September 2011.
- Manuel J. A. Eugster and Friedrich Leisch. Weighted and robust archetypal analysis. *Computational Statistics and Data Analysis*, 55(3):1215–1225, 2011. doi: 10.1016/j.csda.2010.10.017.
- Vicki Hodges. Is Cristiano Ronaldo he world's best player? The Telegraph, March 20, 2008. URL <http://www.telegraph.co.uk/sport/football/2295007/Is-Cristiano-Ronaldo-the-worlds-best-player.html>.
- PES Stats Database. PSD – PES Stats Database. Website, 2011. <http://pesstatsdatabase.com/>; visited on September 22, 2011.
- Leander Scherlaeckens. The best league in the world? ESPN.com, October 11, 2010. URL  
[http://espn.go.com/sports/soccer/news/\\_/id/5673975/success-germany-bundesliga](http://espn.go.com/sports/soccer/news/_/id/5673975/success-germany-bundesliga).
- Christian Schnell. Never Change a Winning Team – Eine Fussballweisheit auf dem statistischen Prüfstand.  
Bachelorarbeit, Institut für Statistik, Ludwig-Maximilians-Universität München, Februar 2011. Supervised by  
Manuel J. A. Eugster and Thomas Augustin.
- Felix Steinbild. Hitzfeld wie Ramsey: Never change a winning team. nachrichten.ch, March 20, 2008. URL  
<http://www.nachrichten.ch/detail/379937.htm>.