The Secret of Playing Football: 
Brazil vs. The Netherlands
Floris Goes, MSc, Dr. Michel Brink, Dr. Marije Elferink-Gemser, Prof. Dr. Koen Lemmink

Sports Science meets Data Science in Football Research

The Team (University of Groningen)

- Prof. Dr. Koen Lemmink – Head of the Group
- Dr. Marije Elferink-Gemser – Expertise: Talent
- Dr. Michel Brink – Expertise: Physical Performance
- Dr. Matthias Kempe – Expertise: Sport Data Science
- Floris Goes, MSc – PhD Student

Annually: 2-4 Bachelor Thesis projects and 2 Master Thesis projects on tactical analysis with tracking data

Notable work in Groningen in the past decade

- Groningen Talent Studies (Marije Elferink-Gemser)
- Load monitoring, stress and recovery in football (Michel Brink)
- First tactical analysis research with LPM system (Wouter Francken)
- Small-sides games in relation to match performance (Sigrid Olthof)
- Ongoing: Three PhD projects on football (Tactical analysis, scouting, hamstring injuries)

General aim of my PhD

- Compare unique characteristics of successful offensive tactics in different countries, age-groups and teams.
  - Focus Topics → Individual & collective decision making
  - Focus Areas → Sport & Data Science

Themes of my PhD project

1. Unlocking the potential of Big Data to support tactical performance analysis in professional football: A systematic review of data analytics and collaboration across scientific domains.
2. Introducing a new measure to rate individual pass effectiveness based on disruption of the defensive organization.
3. Effectiveness of passing sequences & determinants of decision-making.
4. Modelling individual physical & technical performance in tactical analysis models.
Part 1 – Systematic Review: Unlocking the potential of Big Data to support tactical performance analysis in professional football


Gandomi & Haider 2015
Sivarajah et al., 2017
Sarmento et al., 2014
Folgado et al., 2014
Castellano et al., 2013

Data Management
- Data Streams
- Tracking Data
- Meta (Notational) Data
- Sensor Data

Data Analytics
- Pre-processing
- Filtering & Cleaning
- Anonymization
- Annotation

Data Storage
- Access & Architecture
- Security
- Extraction

Data Analytics
- Feature Construction
- Spatial Aggregation
- Meta-data
- Web scraping

Modelling & Analysis
- Window Selection
- Hypothesis Testing
- Data Mining
- Interpretation

Valorisation
- Knowledge extraction

Developments

Aims/Research Question
- Review and compare contributions from both domains
- Review the contributions to data aggregation and analysis
- Review value and translation to practice of current methods
Methods

- Systematic literature search in 7 multidisciplinary databases
- Data Extraction:
  - Definition of Tactical Performance
  - Problem definition & research aims
  - Aggregation features & analysis techniques
  - Data collection, population, and link with match performance

Tactical Performance

- 14 studies (20%) defined tactical performance/tactical behaviour
  - Only 1 of the 14 studies was classified as computer/data science

  - Common elements:
    - Dynamic positioning and organisation
    - Interaction with opponent/ball and adaptation to the conditions
    - $1 + 1 = 2$

Translation → Link with performance

- 21 studies (30%) investigated the relation between tactical behaviour and successful performance
  - Low-level aggregate variables (i.e. centroid) did not show a relation
  - High-level complex methods (subgroups, composite variables, etc.) did show a relationship
Part 2: Measuring the effectiveness of a single pass

A tracking based measure of defensive disruptiveness

Recommendations for Future Research:
- Position data players and ball (automatic analysis)
- Take the interaction/opponent into account
- Investigate the link with successful performance
- Serial collaboration, not parallel
- Translate findings into actionable implications

Generating new questions based on outcomes of sport science research

Data-driven research: data-mining & knowledge discovery

Theory-driven research: integrating tools and testing hypotheses from CS

Part 2: Measuring the effectiveness of a single pass

What is a Good or Bad Pass?

Step 1: Introducing a new measure
- Tracking based pass-rating is often linked to goal scoring (Link et al., 2016; Power et al., 2017) and/or favours forward passes (Rein et al., 2017)
- Other methods rely on notational data
- Not every pass can be an assist, and most passes do not aim to directly create a scoring opportunity

What is a Good or Bad Pass?

Step 1: Introducing a new measure

• Tracking based pass-rating is often linked to goal scoring (Link et al., 2016; Power et al., 2017) and/or favours forward passes (Rein et al., 2017)

• Other methods rely on notational data

• Not every pass can be an assist, and most passes do not aim to directly create a scoring opportunity

Link, Lang & Semborsch, 2016
Rein, Raabe & Memmert, 2017
- Position data of all players and ball
- Automatic detection of ±10.000 passes in 18 Games
- Computation of 10 spatial aggregates

**D-Def & I-Mov definitions**

D-Def → Change in spatial organization of the defensive team between the moment of a pass ($T_0$) and 3 seconds later ($T_1$).

I-Mov → Total movement of all individual players of the defensive team between the moment of a pass ($T_0$) and 3 seconds later ($T_1$).

**Constructing the D-Def measure**

- Dimension reduction (PCA) of 10 spatial features → D-Def measure
- Centroid (X & Y), line-centroids (X & Y), Convex Hull, Frob. Norm.
- Validation with simple I-Mov measure

**Step 2: Validate the approach**

Few studies on tactical analysis investigate the link between tactical performance and actual performance.

1. Study the link between effectiveness and accuracy
2. Study the link between effectiveness and offensive performance

Model Construction

- 75,895 passes in 82 professional Dutch games
- Machine Learning (Ridge Regression with GL Cross Validation) to predict pass accuracy rate, assists per game, and key passes based on I-Mov & D-Def

Trade-off Accuracy & Effectiveness

Step 3: Individual ‘Fingerprint’

Inter-player differences?
Determinants of effective passers
Team composition → high risk & low risk players
Player profiling

Part 3: Passing Sequences & Passing Networks

From counting the frequencies of interactions to measuring the effectiveness of interactions & decision-making

Part 4: Integrating Physical Performance

Modeling individual action-capabilities in relation to tactical performance
Part 5: The Secret of Playing football

Determinants of successful attacks in different populations

Multidisciplinary collaboration

Pre-processing
Filtering
Automated event-detection
Anonymization

Statistical testing
Machine Learning
Graph Theory
Data Mining
Aggregation
Feature dev. & extraction

Communication with SQL
Visualization
Coach feedback

Mehrotra, de Leeuw, Goes & Knobbe (2018)