Analyzing Movement Dynamics in Elite Female Soccer Athletes

Kendall Thomas[1] and Jan Hannig[1]

[1]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Full Paper

Introduction

Motivation: Understanding complex movement dynamics is key to optimizing performance and managing load in elite women's soccer. Traditional analyses, however, often overlook the multidimensional nature of athlete movement data.

Goal: Develop interpretable statistical models that quantify movement patterns in elite women's soccer and assess their links to athlete performance and match outcomes.

Approach:
1. Construct a 3D *quantile cube* (velocity, acceleration, and angle).
2. Compare movement distributions across match halves.
3. Apply dimensionality reduction to movement patterns in match contexts.
4. Model movement distributions as a function of athlete and match characteristics.

Methods

Data:
- GPS data from 9 elite female soccer athletes across both halves of 23 matches
  - 1Hz: timestamp, longitude, latitude
- Included athletes >25 min/half in ≥5 matches
  → 396 unique athlete-match-halves
- Covariates: Match- and athlete-level features for each athlete-match-half, stored in $\boldsymbol{X}_{n\times r=13}$

Processing & Metrics:
- Transformed coordinates to into (x, y) coordinates in meters and fit a 3rd degree spline
- Calculated velocity, acceleration, and angle (180° to 180°) from the spline
- Set velocities < 0.01 m/s and accelerations < 0.001 m/s² to 0, then log-transformed values

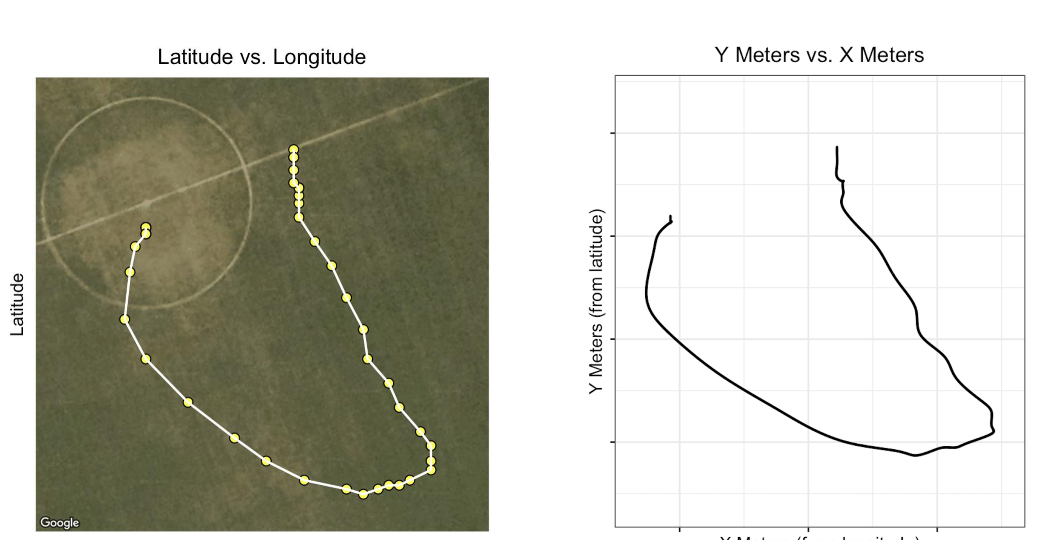

Fig 1. Left: Raw GPS data over 50 sec. Right: Interpolating spline for the same 50 sec.

Methods

Data Object: **Quantile Cube**
- A 3-D summary of athlete movement distributions
- Dimensions: velocity, acceleration, angle
- Discretized into quantiles: 5 velocity × 5 acceleration × 4 angle
  - Angle offset: starts from -30°

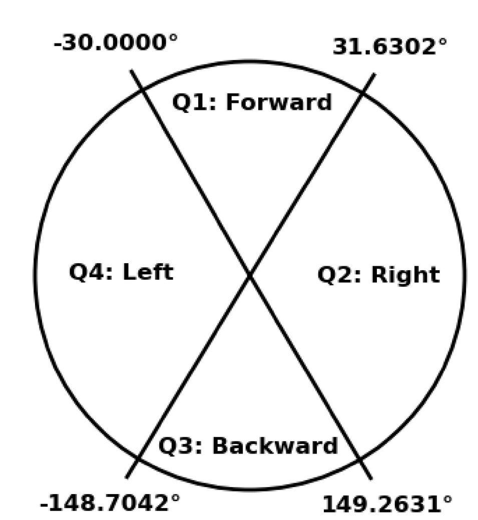

Fig 2. Angle Quantile Representation

Construction:
- Quantile cube created for each athlete-match-half from full-dataset quantiles
  → $5 \times 5 \times 4 = 100$-dimensional vector
- Represents either deciseconds or proportion of total time spent in each movement category
- Resulting dataset: $\boldsymbol{Y}_{n=396\times d=100}$

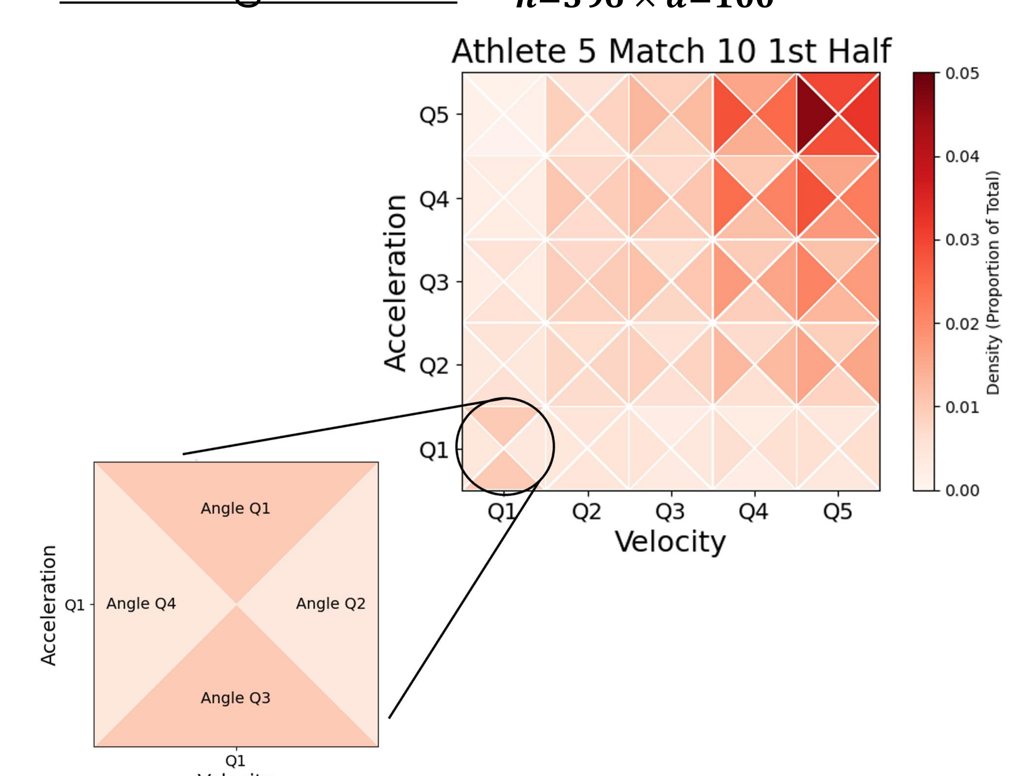

Fig 3. Quantile Cube for Athlete 5 Match 10 1st Half

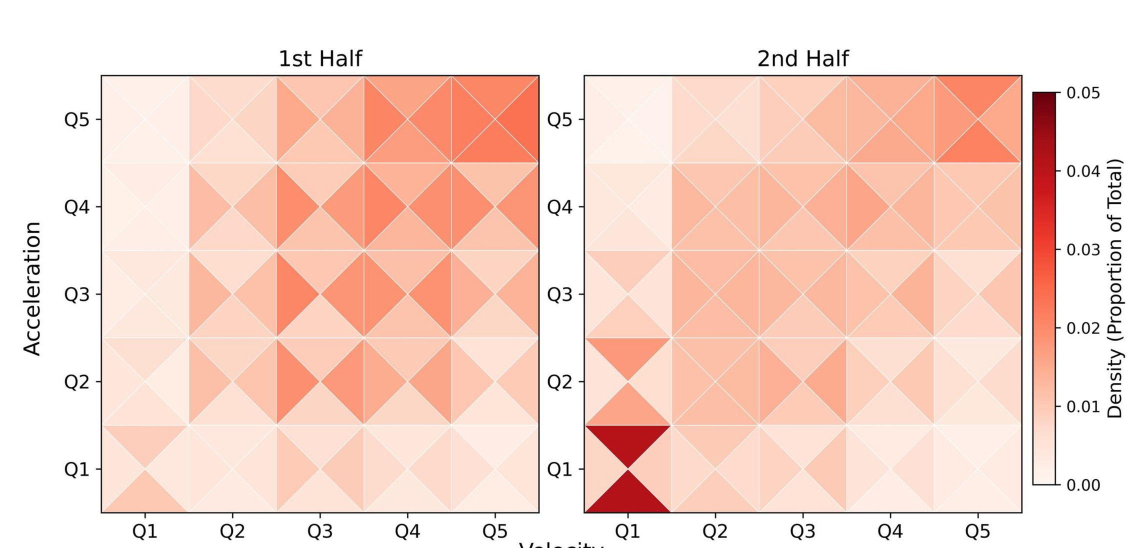

Fig 4. Quantile Cubes for Athlete 3 in the 1st and 2nd Halves of Match 12

Results

1. Quantify Differences Between Match Halves
- Using Hellinger distance, movement distributions differ significantly between the 1st and 2nd half for each player in every match. (see Fig. 4)

2. Reduce Dimensionality of Movement Patterns
- Principal Component Analysis (PCA) reduces the 100-dimensional movement data to 7 components explaining 90% of variance.
- PC1: Captures a shift toward polarized movement in Match 1's 2nd half, with less time in moderate-intensity movements

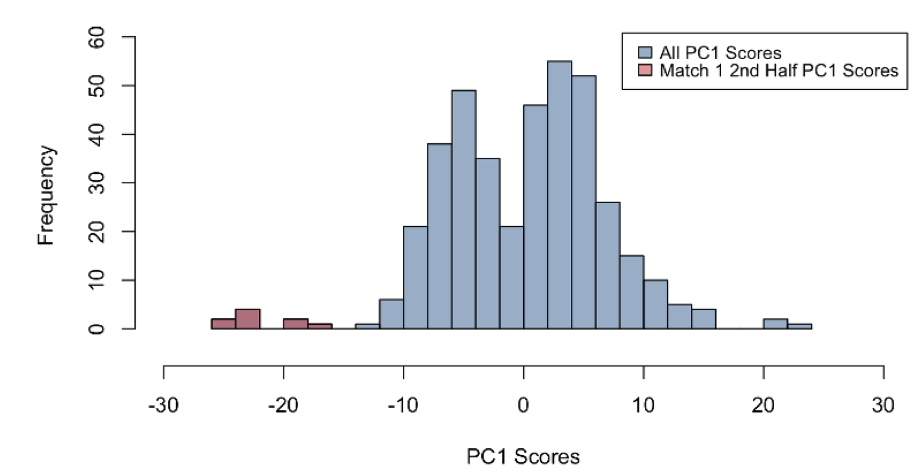

Fig 5. Histogram of PC1 Scores

- PC4: Highlights Match 23's unique pattern, with increased low-velocity, high-acceleration movements

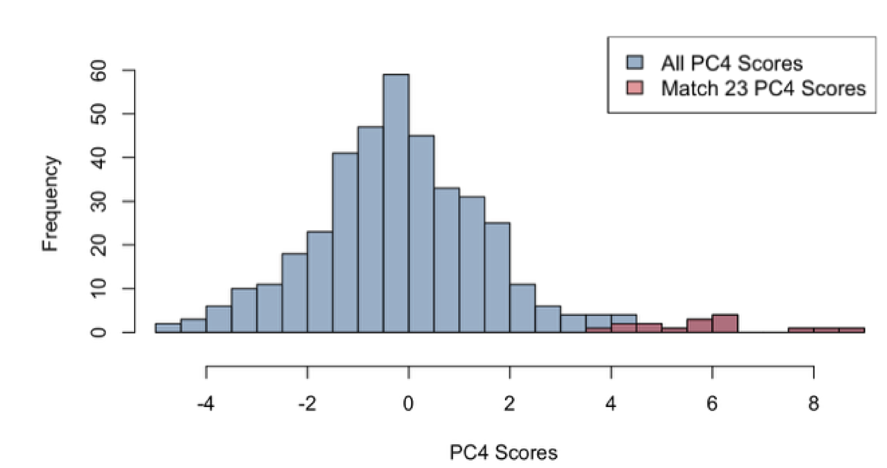

Fig 6. Histogram of PC4 Scores

3. Model Movement Distributions
- Relate athlete movement distributions to contextual factors using Dirichlet-multinomial regression (DMR)
- For each athlete-match-half i, the movement distribution $\boldsymbol{y}_i$ is modeled as a draw from a Dirichlet-multinomial (DM) distribution parameterized by $\boldsymbol{\eta}_i$.
- Each category j has a concentration parameter

$$\eta_{ij} = \exp\left(\beta_{j0} + \sum_{k=1}^{r} \beta_{jk} x_{ik}\right)$$

linking covariates to the proportion of time spent in that category.

Results

DMR Model: half, log(playing time), and position group

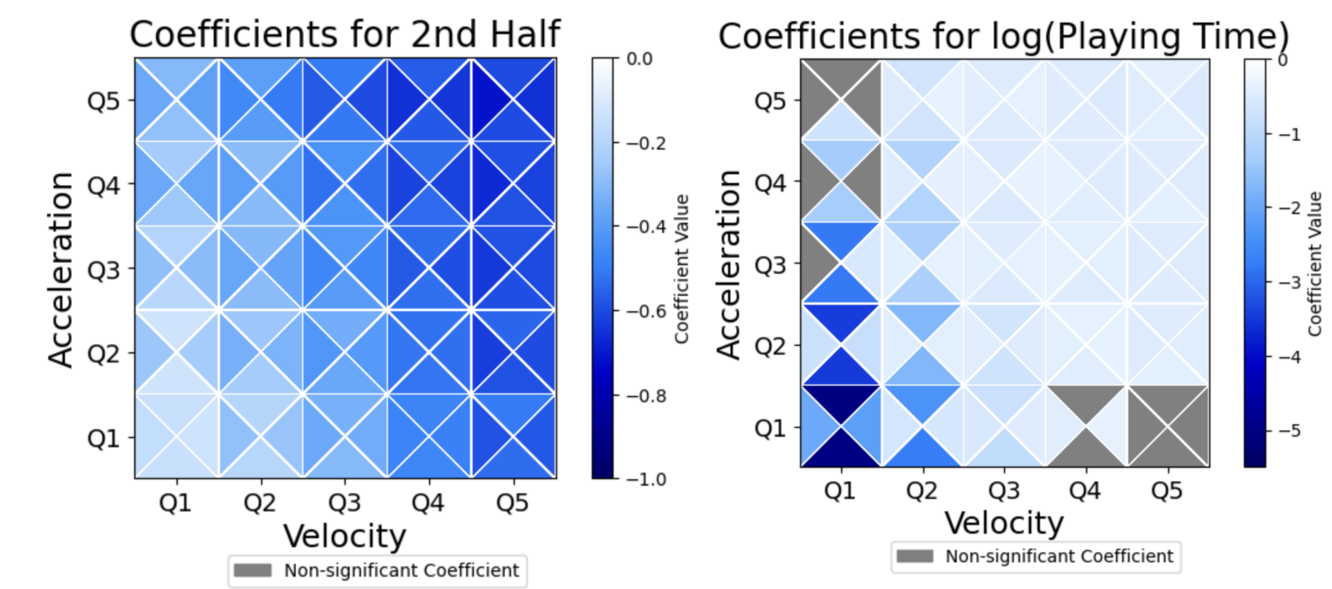

Fig 7. DMR Coefficients for 2nd Half (Baseline Category: 1st Half) and log(Playing Time)

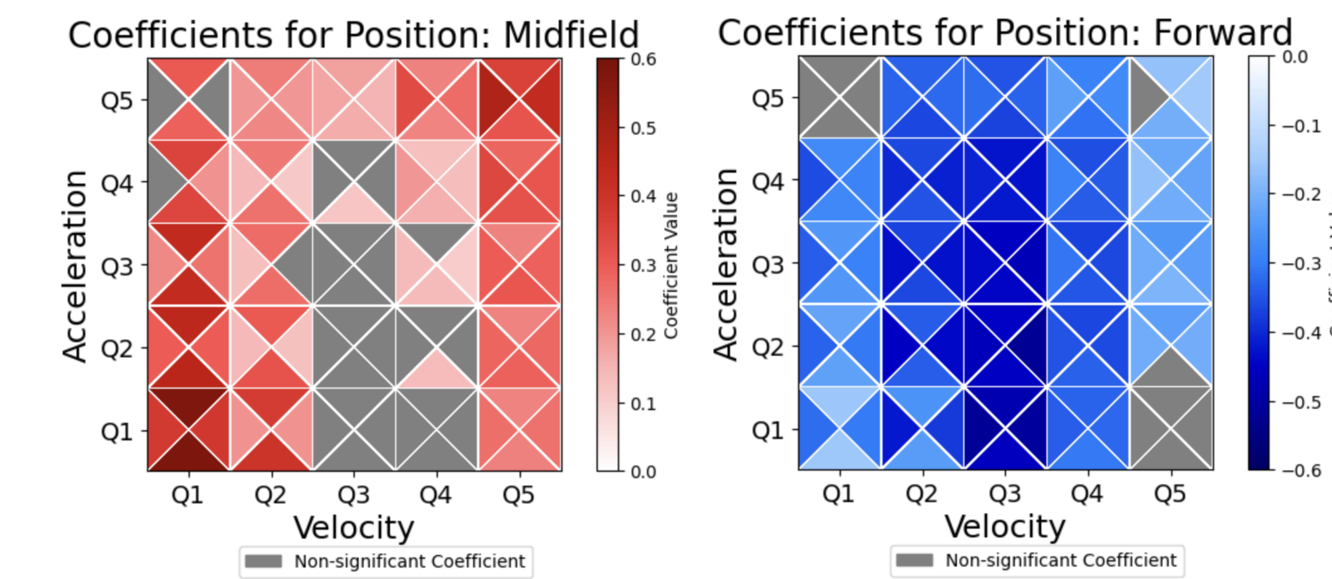

Fig 8. DMR Coefficients for Position Groups (Baseline Category: Defender)

Conclusions & Future Directions

The quantile cube effectively captures multidimensional movement dynamics, providing interpretable workload analytics.
- Movement distributions shifted significantly between halves, indicating potential acute fatigue or tactical shifts.
- PCA and DMR reveal role- and context-specific movement signatures, supporting probability-based monitoring for tailored training and recovery.

Future:
- Integrate longitudinal, multimodal data (IMU, RPE, wellness data)
- Validate the framework in real-world settings

Acknowledgements

We extend our gratitude to Elena I. Cantú, Sam R. Moore, and Dr. Abbie E. Smith-Ryan from the Applied Physiology Lab in the University of North Carolina at Chapel Hill's Department of Exercise and Sport Science for their generous contributions in gathering and providing access to the data.