

Supplementary Material for
*Stable Stylistic Axes in Football: Interpretable
Principal-Component Decomposition of Player
Data from the Top-Five European Leagues*

July 20, 2025

Contents

1	Glossary of Raw Data Variables	2
2	Extended Methods	7
2.1	Yeo-Johnson transformation	7
2.2	Feature Engineering	7
2.3	Individual PC loadings	8
3	Extended Results	11
3.1	Further PCs interpretation	11
3.2	Remaining latent space bi-plots	12
3.3	Latent space bi-plots over multiple seasons	15
3.4	Average values of components across leagues	19
3.5	Distribution of components values across selected leagues	21
4	Extended Robustness and Validation Checks	24
5	Data and Code Availability	25

1 Glossary of Raw Data Variables

- **player** – Player's full name.
- **nation** – Country the player represents internationally (FIFA code).
- **pos** – Primary position(s) the player occupied during the season.
- **age** – Age of the player on January 1st of the season.
- **90s** – Number of 90-minute equivalents played (minutes played divided by 90).
- **standard_gls** – Goals scored, excluding own goals.
- **standard_sh** – Total number of shots attempted.
- **standard_sot** – Shots on target (i.e., shots that would score without a goal-keeper).
- **standard_sot%** – Percentage of shots that were on target.
- **standard_sh/90** – Shots attempted per 90 minutes played.
- **standard_sot/90** – Shots on target per 90 minutes played.
- **standard_g/sh** – Goals scored per shot taken.
- **standard_g/sot** – Goals scored per shot on target.
- **standard_dist** – Average distance (in yards) from which shots were taken.
- **standard_fk** – Goals scored directly from free kicks.
- **standard_pk** – Penalty kicks scored.
- **standard_pkatt** – Penalty kicks attempted.
- **expected_xg** – Total expected goals (xG), based on shot quality and location.
- **expected_npxg** – Expected goals excluding penalty kicks.
- **expected_npxg/sh** – Non-penalty expected goals per shot.
- **expected_g-xg** – Difference between actual goals and xG (positive means over-performance).
- **expected_np:g-xg** – Difference between actual non-penalty goals and non-penalty xG.
- **total_cmp** – Total completed passes.
- **total_att** – Total attempted passes.
- **total_cmp%** – Pass completion percentage (completed ÷ attempted).
- **total_totdist** – Total passing distance in yards.
- **total_prgdist** – Progressive passing distance (yards toward opponent's goal).

- **short_cmp** – Completed short passes (5–15 yards).
- **short_att** – Attempted short passes.
- **short_cmp%** – Short pass completion percentage.
- **medium_cmp** – Completed medium passes (15–30 yards).
- **medium_att** – Attempted medium passes.
- **medium_cmp%** – Medium pass completion percentage.
- **long_cmp** – Completed long passes (30+ yards).
- **long_att** – Attempted long passes.
- **long_cmp%** – Long pass completion percentage.
- **ast** – Assists credited for a goal.
- **xag** – Expected assists (xA), based on shot quality from passes.
- **xa** – Total expected assists.
- **a-xag** – Difference between actual assists and expected assists (xA).
- **kp** – Key passes (a pass that leads directly to a shot).
- **1/3** – Completed passes into the final third.
- **ppa** – Completed passes into the penalty area.
- **crspa** – Crosses into the penalty area.
- **prgp** – Progressive passes (passes that advance the ball significantly forward).
- **att** – Total number of pass attempts.
- **pass_types_live** – Passes made while the ball was in play.
- **pass_types_dead** – Passes made from dead-ball situations (e.g., free kicks, corners).
- **pass_types_fk** – Passes from free kicks.
- **pass_types_tb** – Through balls.
- **pass_types_sw** – Switches of play (wide lateral passes).
- **pass_types_crs** – Crosses.
- **pass_types_ti** – Throw-ins.
- **pass_types_ck** – Corner kicks.
- **corner_kicks_in** – Inswinging corner kicks.
- **corner_kicks_out** – Outswinging corner kicks.

- **corner_kicks_str** – Straight corner kicks.
- **outcomes_cmp** – Passes completed.
- **outcomes_off** – Passes that were offside.
- **outcomes_blocks** – Passes that were blocked.
- **sca_sca** – Shot-creating actions (2 offensive actions before a shot).
- **sca_sca90** – Shot-creating actions per 90 minutes.
- **sca_types_passlive** – Shot-creating actions from live-ball passes.
- **sca_types_passdead** – Shot-creating actions from dead-ball passes.
- **sca_types_to** – Shot-creating actions from take-ons (dribbles past opponent).
- **sca_types_sh** – Shot-creating actions from shots.
- **sca_types_fld** – Shot-creating actions from fouls drawn.
- **sca_types_def** – Shot-creating actions from defensive actions.
- **gca_gca** – Goal-creating actions (2 offensive actions before a goal).
- **gca_gca90** – Goal-creating actions per 90 minutes.
- **gca_types_passlive** – Goal-creating actions from live-ball passes.
- **gca_types_passdead** – Goal-creating actions from dead-ball passes.
- **gca_types_to** – Goal-creating actions from take-ons.
- **gca_types_sh** – Goal-creating actions from shots.
- **gca_types_fld** – Goal-creating actions from fouls drawn.
- **gca_types_def** – Goal-creating actions from defensive actions.
- **tackles_tkl** – Total number of tackles made.
- **tackles_tklw** – Tackles won (where possession was retained).
- **tackles_def_3rd** – Tackles in the defensive third.
- **tackles_mid_3rd** – Tackles in the middle third.
- **tackles_att_3rd** – Tackles in the attacking third.
- **challenges_tkl** – Dribblers tackled.
- **challenges_att** – Dribblers challenged.
- **challenges_tkl%** – Percentage of challenges where the dribbler was tackled.
- **challenges_lost** – Number of times the dribbler got past the player.
- **blocks_blocks** – Total blocks (shots, passes, crosses).

- **blocks_sh** – Shots blocked.
- **blocks_pass** – Passes blocked.
- **int** – Interceptions.
- **tkl+int** – Combined tackles and interceptions.
- **clr** – Clearances (removing the ball from danger).
- **err** – Defensive errors leading to a shot.
- **touches_touches** – Total number of touches.
- **touches_def_pen** – Touches in the defensive penalty area.
- **touches_def_3rd** – Touches in the defensive third.
- **touches_mid_3rd** – Touches in the middle third.
- **touches_att_3rd** – Touches in the attacking third.
- **touches_att_pen** – Touches in the attacking penalty area.
- **touches_live** – Touches while the ball was in live play.
- **take-ons_att** – Take-on attempts (dribble attempts past an opponent).
- **take-ons_succ** – Successful take-ons.
- **take-ons_succ%** – Take-on success rate.
- **take-ons_tkld** – Times tackled while attempting a take-on.
- **take-ons_tkld%** – Percentage of take-ons that resulted in being tackled.
- **carries_carries** – Total number of ball carries.
- **carries_totdist** – Total distance carried with the ball (in yards).
- **carries_prgdist** – Total progressive carry distance (yards moved toward goal).
- **carries_prgc** – Number of progressive carries (carries that move the ball significantly forward).
- **carries_1/3** – Carries into the final third of the pitch.
- **carries_cpa** – Carries into the penalty area.
- **carries_mis** – Miscontrolled touches resulting in a turnover.
- **carries_dis** – Times dispossessed while carrying the ball.
- **receiving_rec** – Passes received by the player.
- **receiving_prgr** – Progressive passes received (passes moving the ball significantly forward).
- **mp** – Number of matches played.

- **playing_time_min** – Total minutes played.
- **playing_time_mn/mp** – Average minutes played per match.
- **playing_time_min%** – Percentage of total possible minutes played.
- **playing_time_90s** – Number of 90-minute equivalents played.
- **starts_starts** – Number of matches started.
- **starts_mn/start** – Average minutes played per start.
- **starts_compl** – Number of full 90-minute matches completed.
- **subs_subs** – Number of times the player came on as a substitute.
- **subs_mn/sub** – Average minutes played per substitute appearance.
- **subs_unsub** – Number of times the player was in the matchday squad but did not play.
- **team_success_ppm** – Points per match earned while the player was on the pitch.
- **team_success_on** – Goals scored by the team while the player was on the pitch.
- **team_success_onga** – Goals conceded while the player was on the pitch.
- **team_success_+/-** – Goal difference while the player was on the pitch.
- **team_success_+/-90** – Goal difference per 90 minutes played.
- **team_success_on-off** – Difference in goal difference with vs. without the player.
- **team_success_(xg)_onxg** – Expected goals scored while the player was on the pitch.
- **team_success_(xg)_onxga** – Expected goals conceded while the player was on the pitch.
- **team_success_(xg)_xg+/-** – Expected goal difference while the player was on the pitch.
- **team_success_(xg)_xg+/-90** – Expected goal difference per 90 minutes.
- **team_success_(xg)_on-off** – xG goal difference with vs. without the player.
- **performance_crdy** – Yellow cards received.
- **performance_crdr** – Red cards received.
- **performance_2crdy** – Second yellow card leading to a red.
- **performance_fls** – Fouls committed.
- **performance_fld** – Fouls drawn (fouled by an opponent).

- **performance_off** – Offsides.
- **performance_crs** – Crosses attempted.
- **performance_int** – Interceptions made.
- **performance_tklw** – Tackles won (retained possession).
- **performance_pkwon** – Penalties won.
- **performance_pkcon** – Penalties conceded.
- **performance_og** – Own goals scored.
- **performance_recov** – Loose ball recoveries.
- **aerial_duels_won** – Aerial duels won.
- **aerial_duels_lost** – Aerial duels lost.
- **aerial_duels_won%** – Percentage of aerial duels won.

2 Extended Methods

2.1 Yeo-Johnson transformation

The Yeo-Johnson transformation is defined by the following formulas:

$$Y_i(\lambda) = \begin{cases} \frac{(x_i+1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \text{ and } x_i \geq 0 \\ \ln(x_i + 1) & \text{if } \lambda = 0 \text{ and } x_i \geq 0 \\ \frac{-(-x_i+1)^{(2-\lambda)} - 1}{2-\lambda} & \text{if } \lambda \neq 2 \text{ and } x_i < 0 \\ -\ln(-x_i + 1) & \text{if } \lambda = 2 \text{ and } x_i < 0 \end{cases} \quad (\text{S2.1})$$

2.2 Feature Engineering

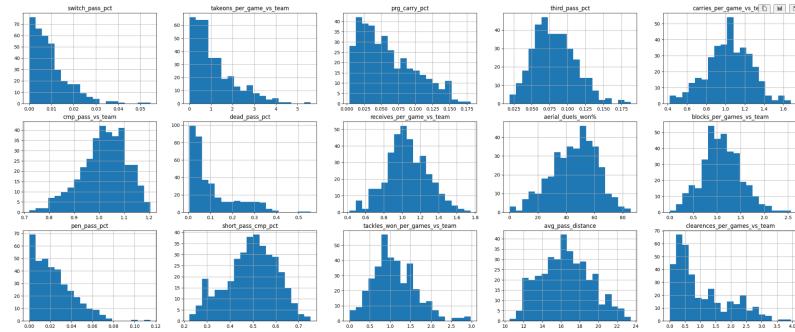


Figure S2.1: Distributions of selected raw pre-transformation features for the Premier League 2017/18 sample. Heavy skew and variable variance is prevalent.

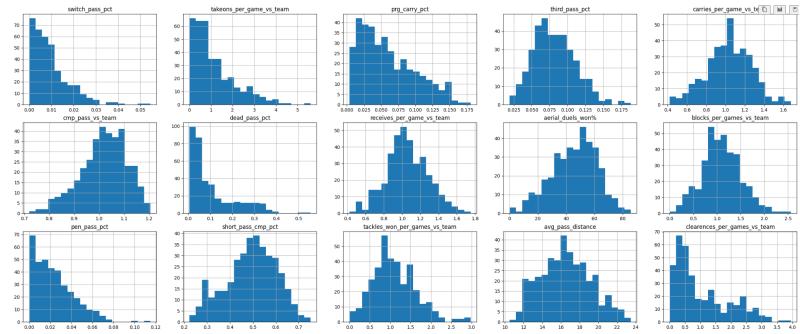


Figure S2.2: Distributions of selected post-transformation features for the Premier League 2017/18 sample. In line with Yeo-Johnson assumptions, values are centered around 0 and appear more uniform.

2.3 Individual PC loadings

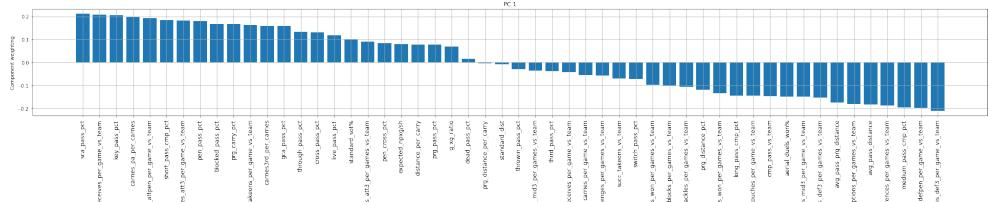


Figure S2.3: PC_1 loadings

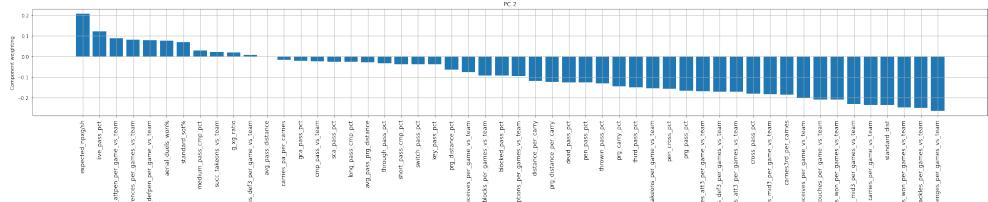


Figure S2.4: PC_2 loadings

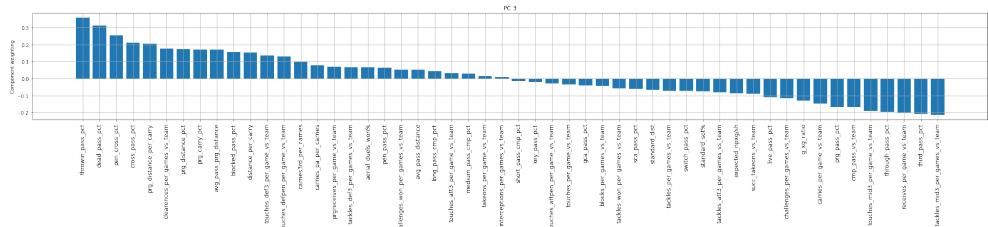


Figure S2.5: PC_3 loadings

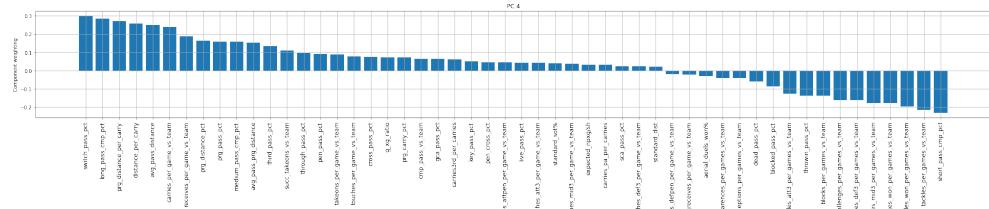


Figure S2.6: PC_4 loadings

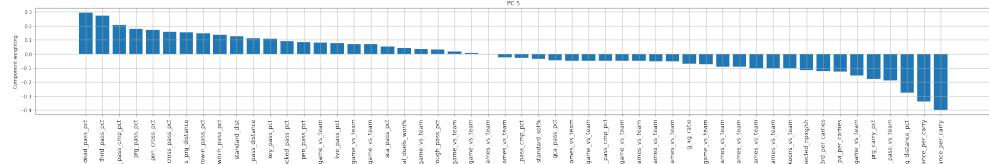


Figure S2.7: PC_5 loadings

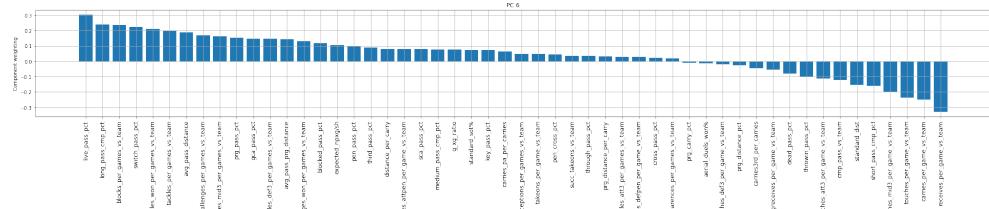


Figure S2.8: PC_6 loadings

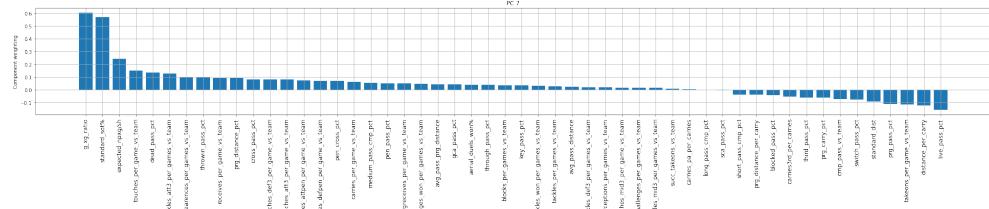


Figure S2.9: PC_7 loadings

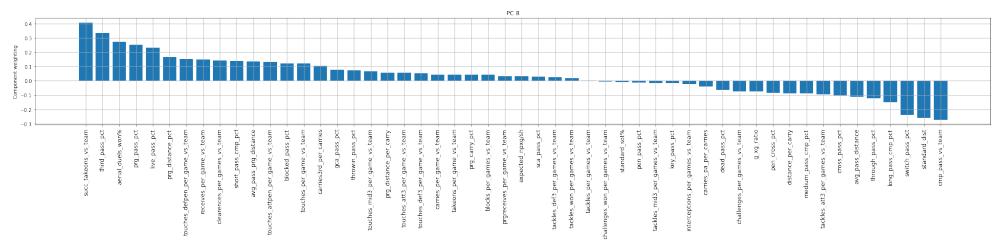


Figure S2.10: PC_8 loadings

3 Extended Results

3.1 Further PCs interpretation

Table S3.1: Highest positive and negative coefficients for the PCs 6-8, Premier League 2017/18.

PC label	Top + loadings	Top - loadings	Rationale
PC_6 Ball-winning vs Metronome	live_pass_pct (0.30) long_pass_cmp_pct (0.24) blocks_per_game_vs_team (0.24) switch_pass_pct (0.22) tackles_won_per_game_vs_team (0.21)	receives_per_game_vs_team (-0.33) carries_per_game_vs_team (-0.25) touches_per_game_vs_team (-0.24) touches_mid3_per_game_vs_team (-0.20) short_pass_cmp_pct (-0.16)	Axis from “high-intensity ball-winners” who regain possession and play longer forward passes to “high-touch metronomes” who receive constantly, favor short passes, and recycle play to retain control. This provides a finer role distinction during possession regains than PC_4 .
PC_7 Dead-ball & Finishing vs Live-play Progression	g_xg_ratio (0.60) standard_sot% (0.57) expected_npxg/sh (0.24) touches_per_game_vs_team (0.15) dead_pass_pct (0.14)	live_pass_pct (-0.16) distance_per_carry (-0.12) prg_pass_pct (-0.11) takeons_per_game_vs_team (-0.11) standard_dist (-0.09)	Differentiates restart-savvy finishers who thrive on dead-ball phases and clinical shot conversion from live-play progressors who advance possession through carries, switches and forward passing.
PC_8 Ball distribution range	succ_takeons_vs_team (0.41) third_pass_pct (0.34) aerial_duels_won% (0.27) prg_pass_pct (0.25) live_pass_pct (0.23)	cmp_pass_vs_team (-0.27) standard_dist (-0.26) switch_pass_pct (-0.24) long_pass_cmp_pct (-0.15) through_pass_pct (-0.12)	Separates compact short-pass recyclers who clear danger and keep possession from long-range distributors launching switches, diagonals and through-balls.

3.2 Remaining latent space bi-plots

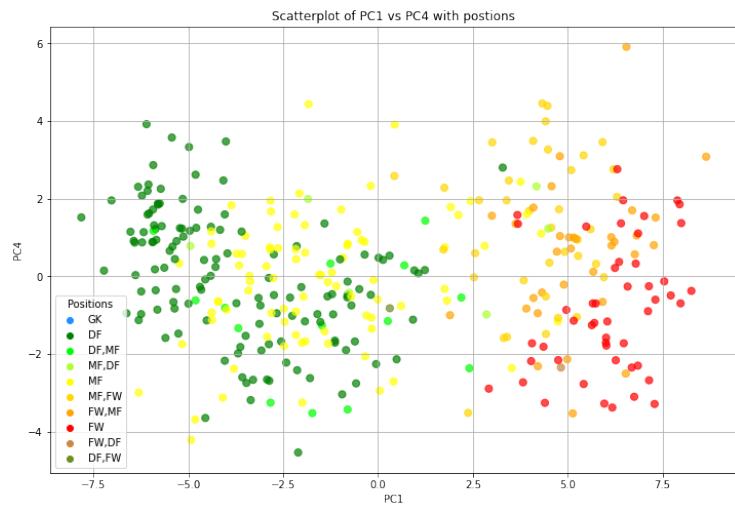


Figure S3.1: $PC_1 \times PC_4$

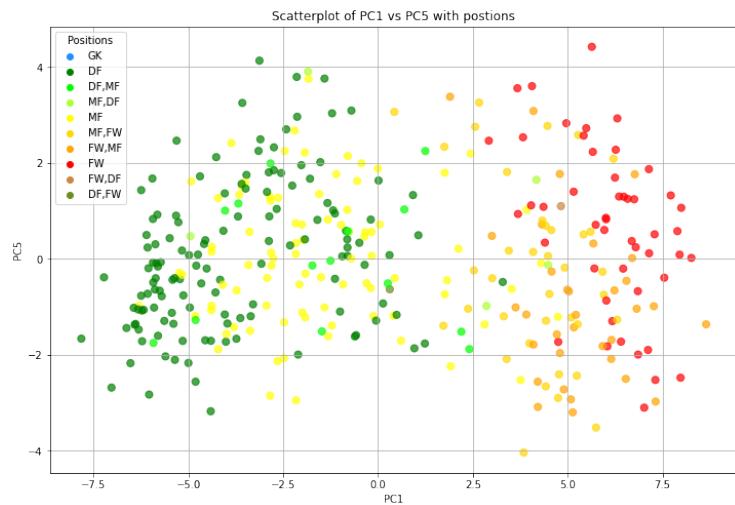


Figure S3.2: $PC_1 \times PC_5$

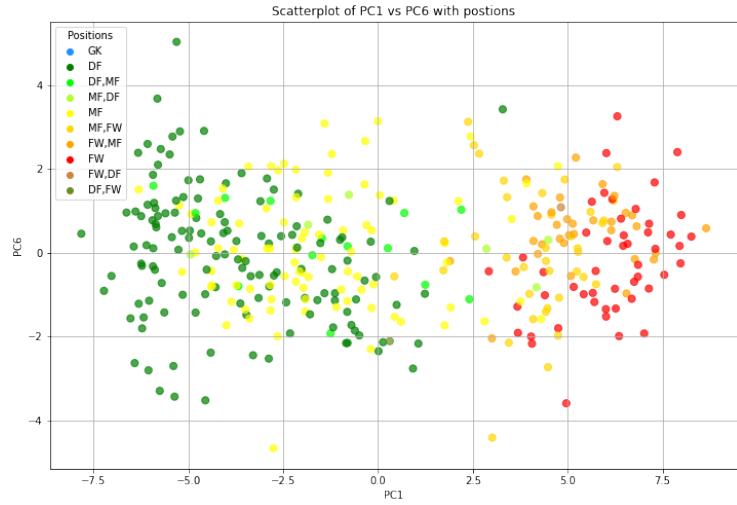


Figure S3.3: $PC_1 \times PC_6$

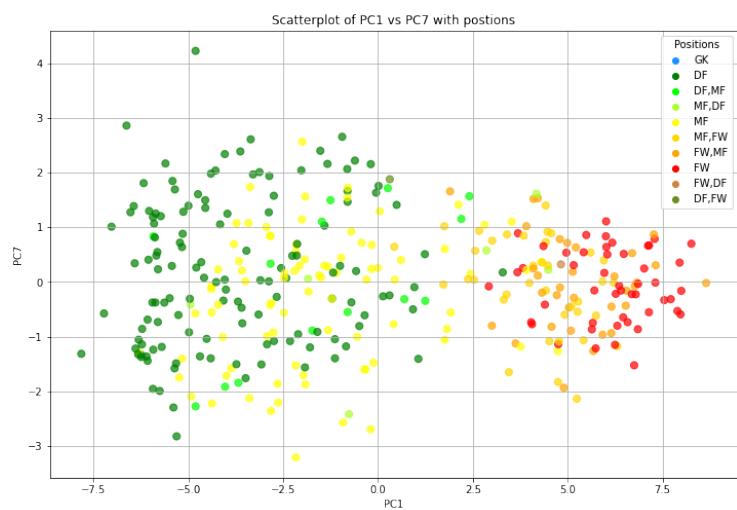


Figure S3.4: $PC_1 \times PC_7$

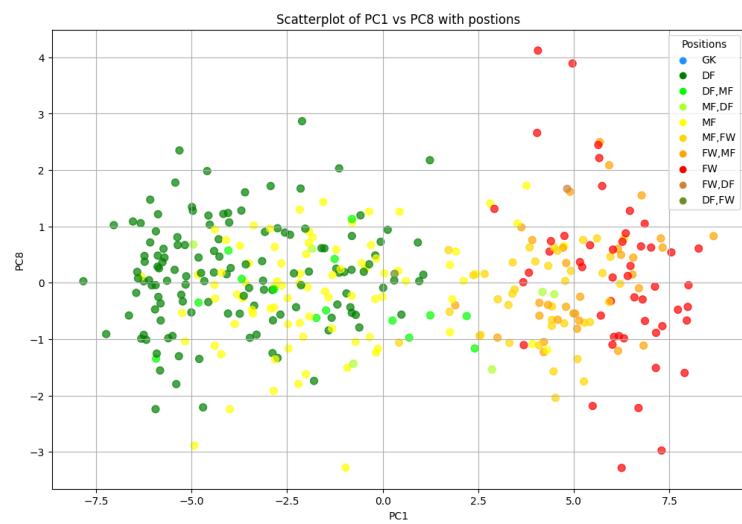


Figure S3.5: $PC_1 \times PC_8$

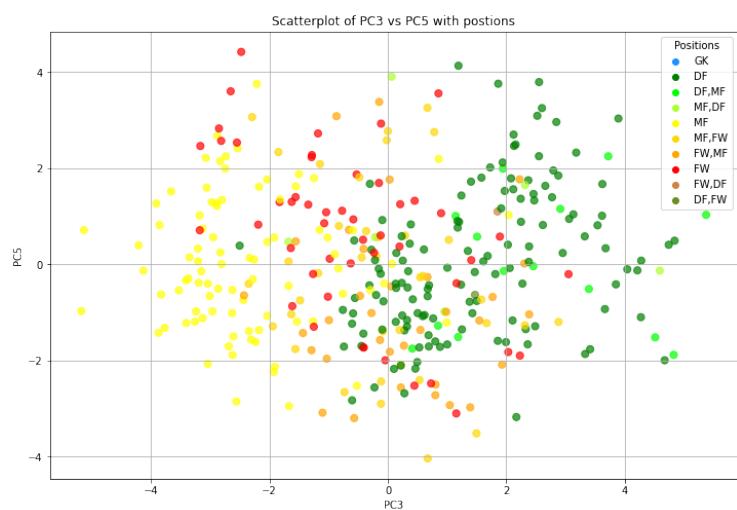


Figure S3.6: $PC_3 \times PC_5$

3.3 Latent space bi-plots over multiple seasons

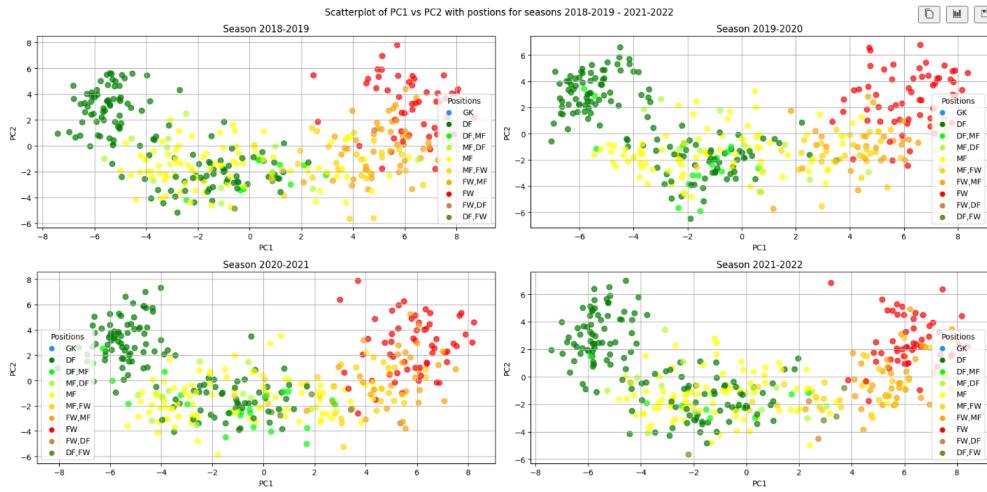


Figure S3.7: $PC_1 \times PC_2$ subplanes between 2018/19 - 2021/22 EPL seasons

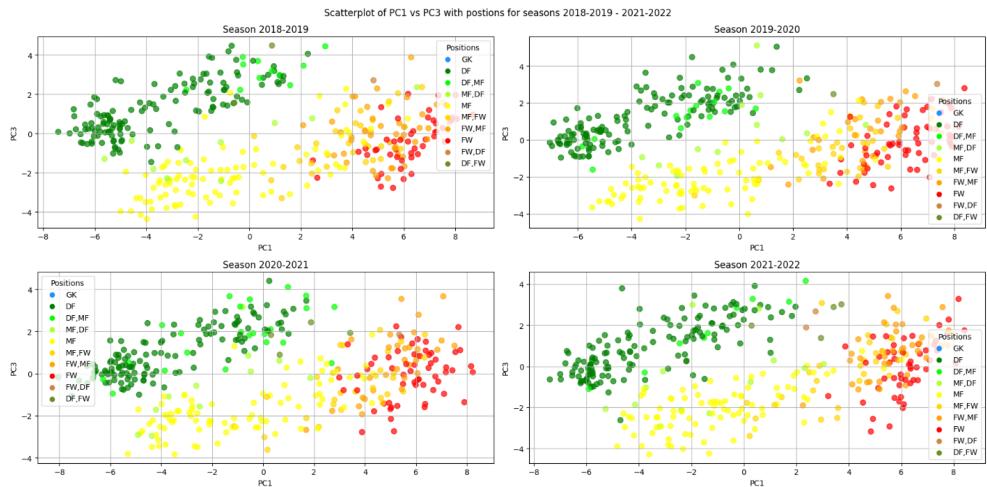


Figure S3.8: $PC_1 \times PC_3$ subplanes between 2018/19 - 2021/22 EPL seasons

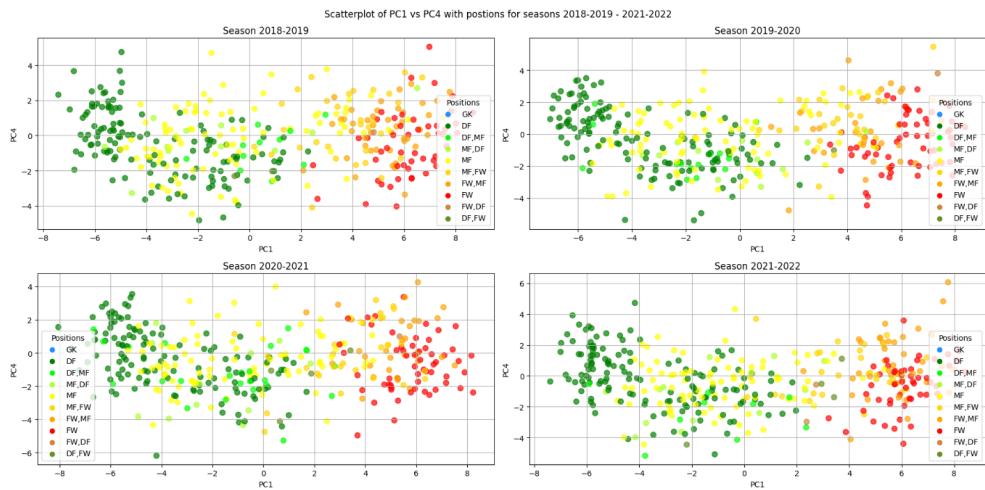


Figure S3.9: $PC_1 \times PC_4$ subplanes between 2018/19 - 2021/22 EPL seasons

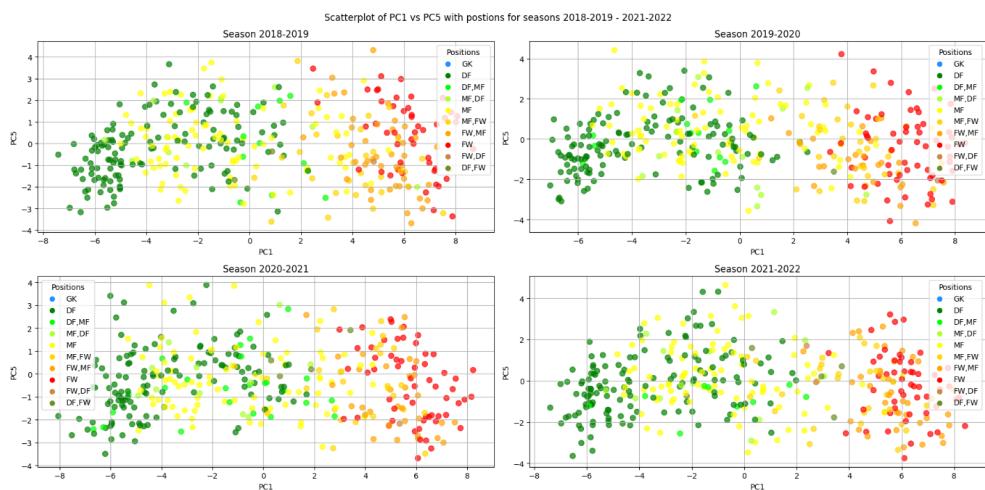


Figure S3.10: $PC_1 \times PC_5$ subplanes between 2018/19 - 2021/22 EPL seasons

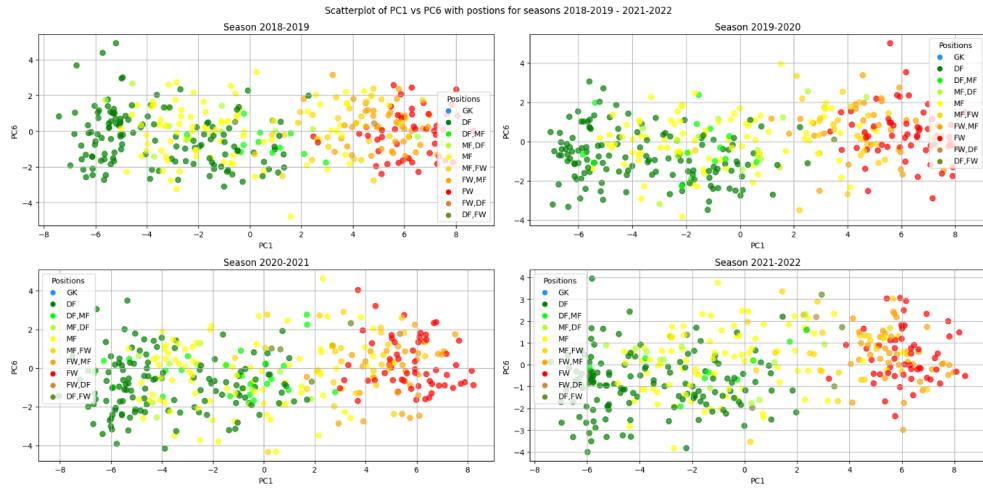


Figure S3.11: $PC_1 \times PC_6$ subplanes between 2018/19 - 2021/22 EPL seasons

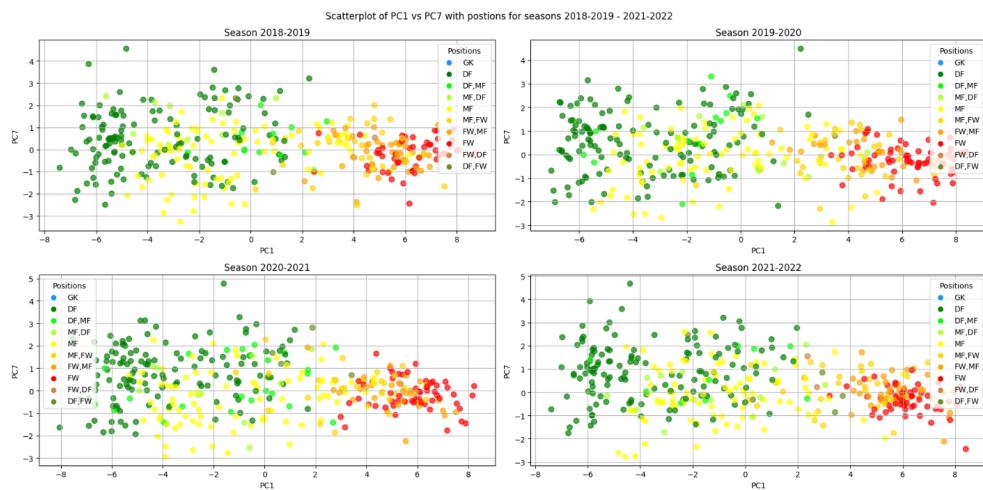


Figure S3.12: $PC_1 \times PC_7$ subplanes between 2018/19 - 2021/22 EPL seasons

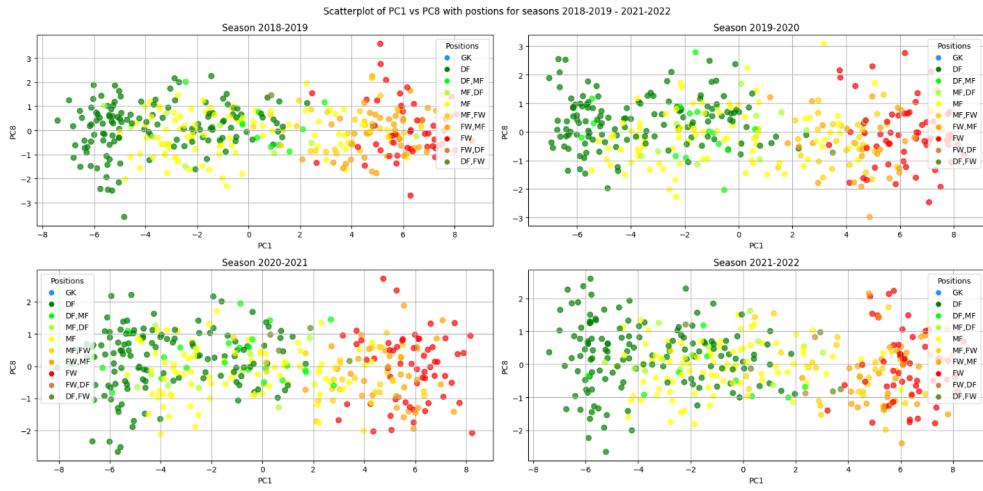


Figure S3.13: $PC_1 \times PC_8$ subplanes between 2018/19 - 2021/22 EPL seasons

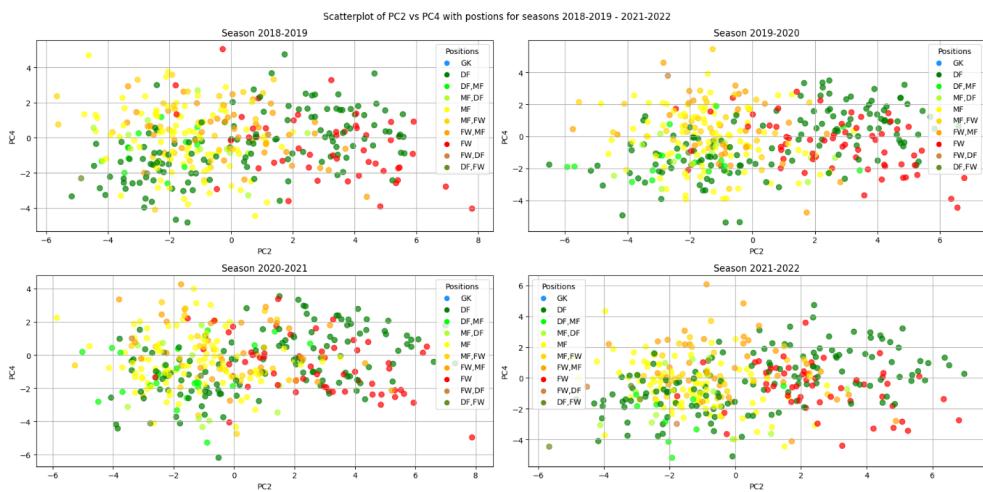


Figure S3.14: $PC_2 \times PC_4$ subplanes between 2018/19 - 2021/22 EPL seasons

3.4 Average values of components across leagues

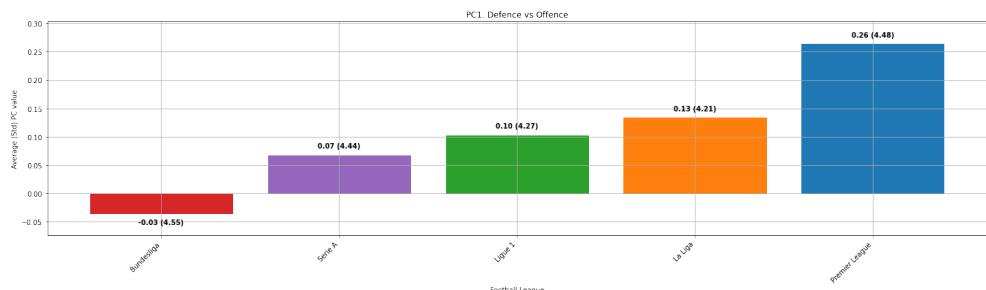


Figure S3.15: Average (standard deviation) values of PC_1

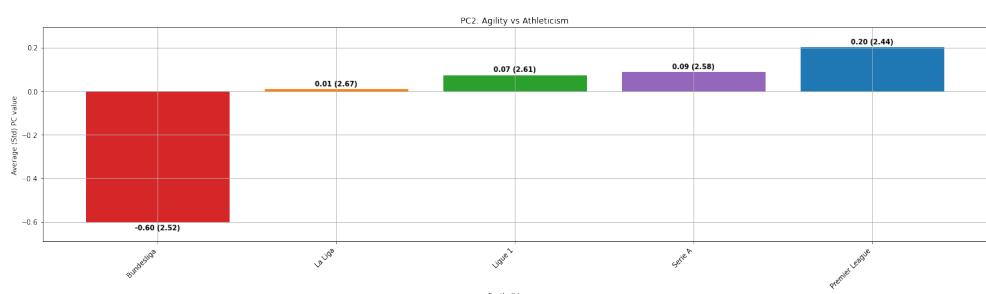


Figure S3.16: Average (standard deviation) values of PC_2

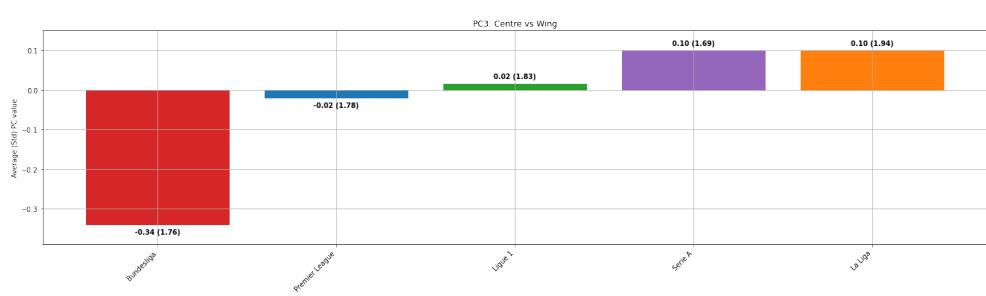


Figure S3.17: Average (standard deviation) values of PC_3

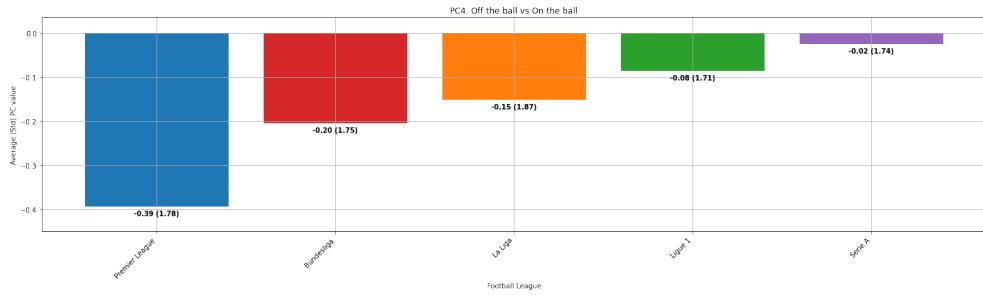


Figure S3.18: Average (standard deviation) values of PC_4

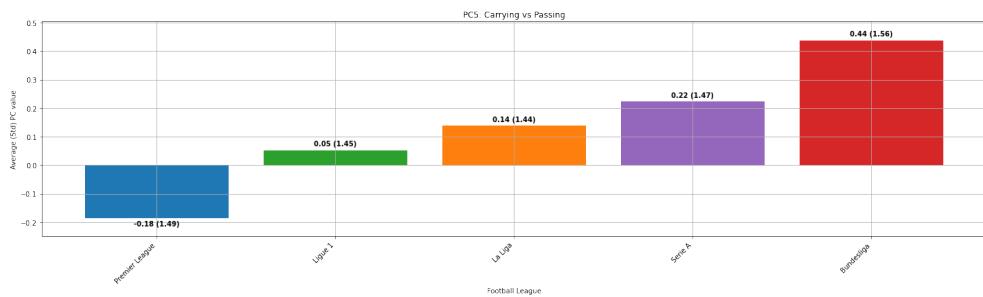


Figure S3.19: Average (standard deviation) values of PC_5

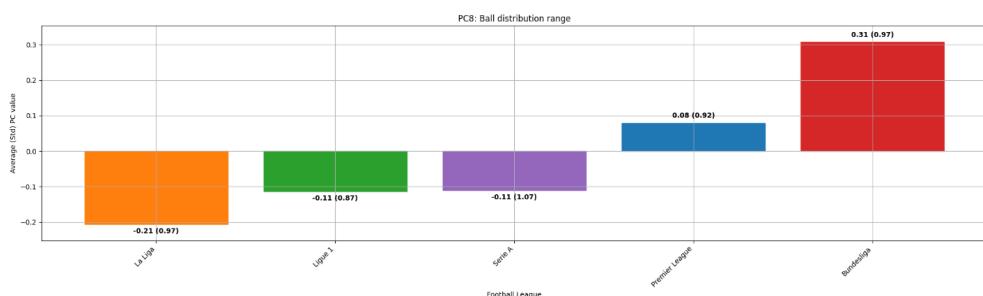


Figure S3.22: Average (standard deviation) values of PC_8

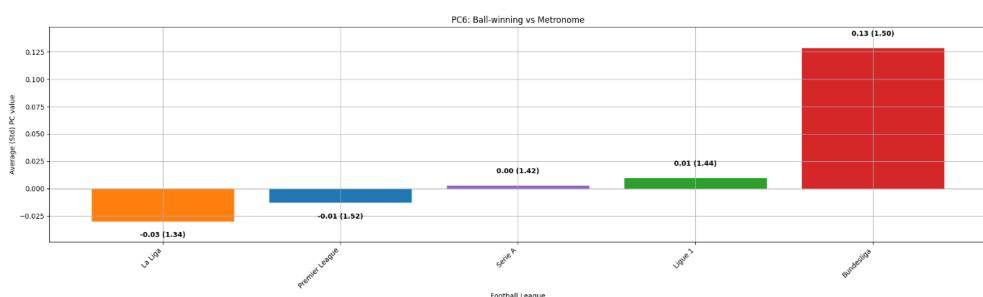


Figure S3.20: Average (standard deviation) values of PC_6



Figure S3.21: Average (standard deviation) values of PC_7

3.5 Distribution of components values across selected leagues

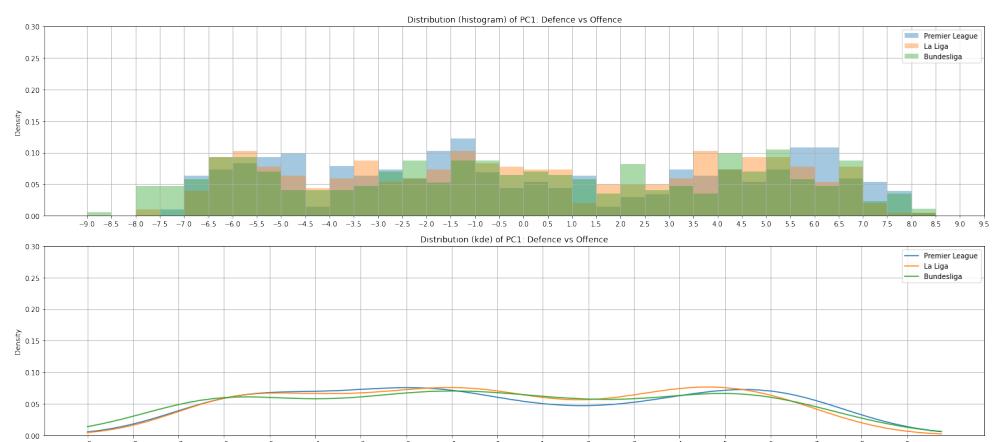


Figure S3.23: Distribution of PC_1 values

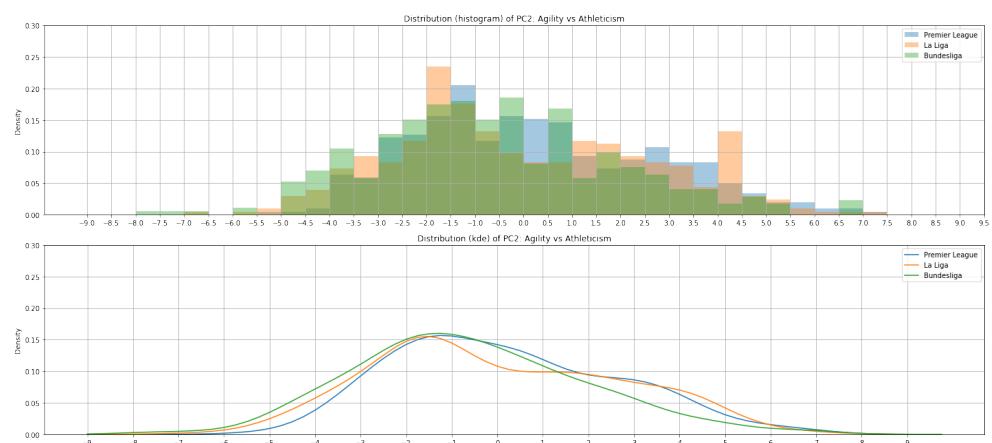


Figure S3.24: Distribution of PC_2 values

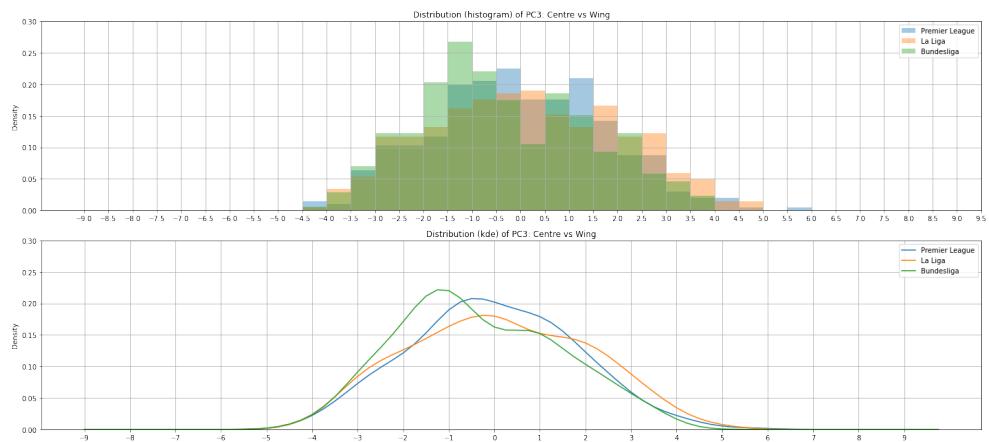


Figure S3.25: Distribution of PC_3 values

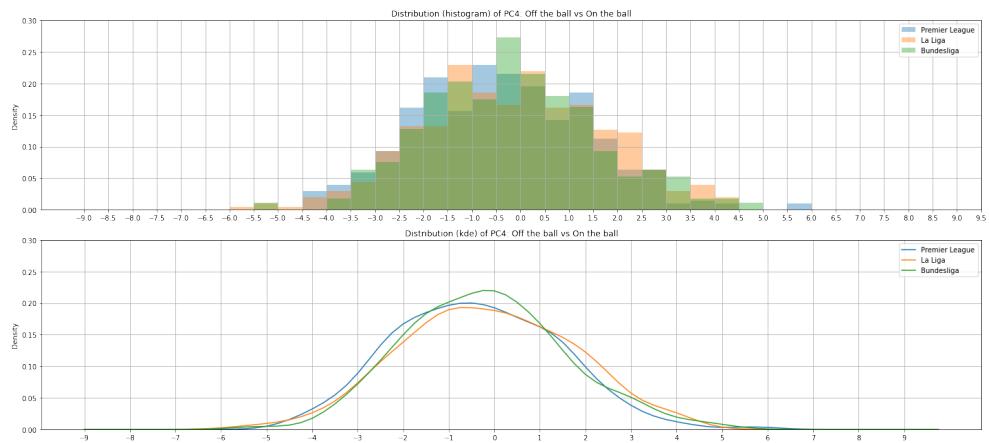


Figure S3.26: Distribution of PC_4 values

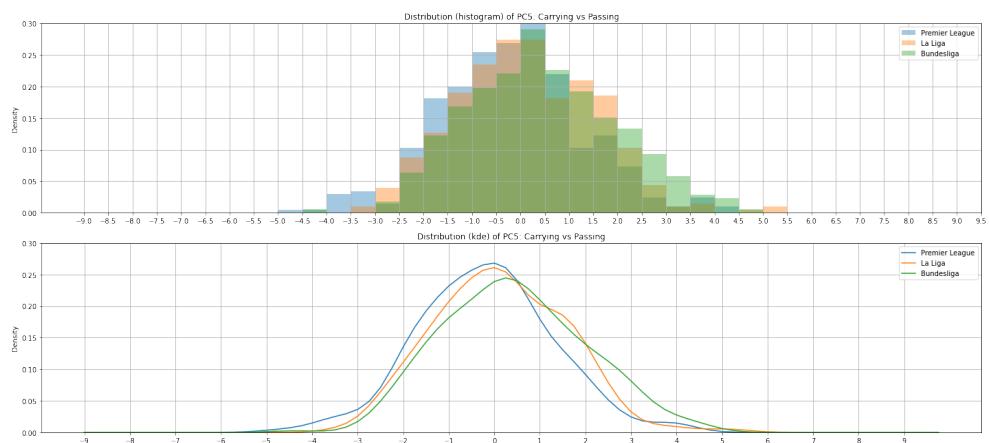


Figure S3.27: Distribution of PC_5 values

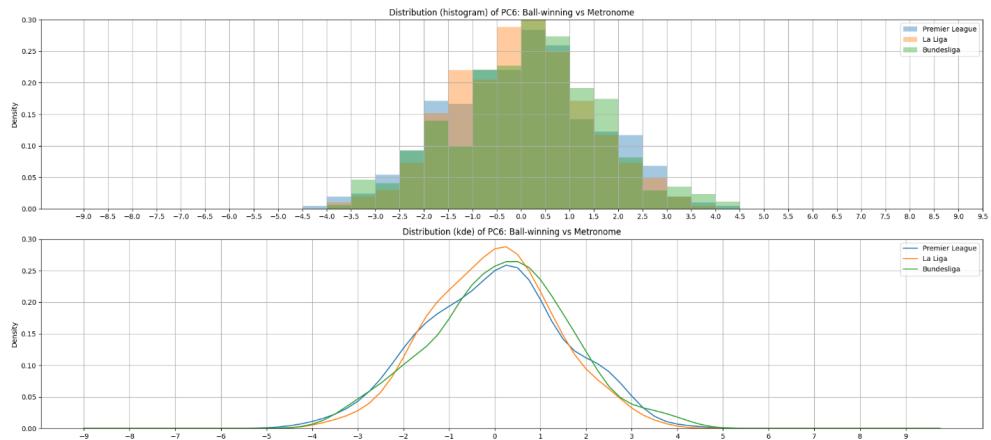


Figure S3.28: Distribution of PC_6 values

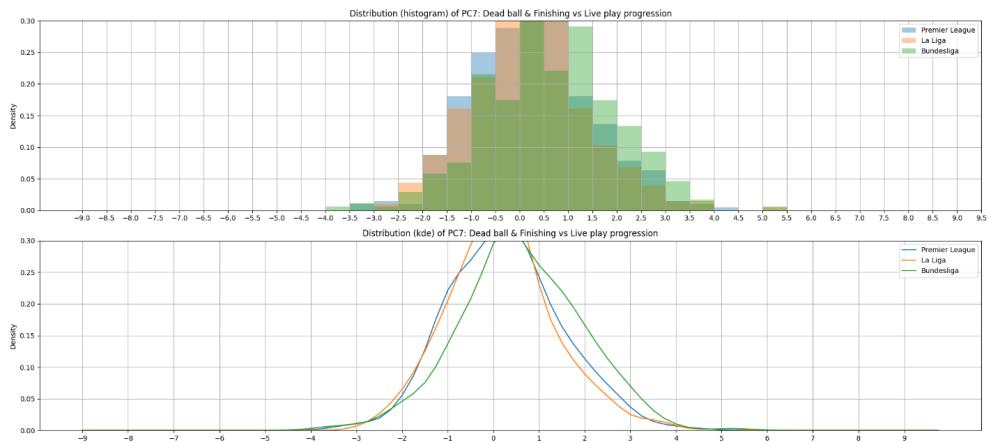


Figure S3.29: Distribution of PC_7 values

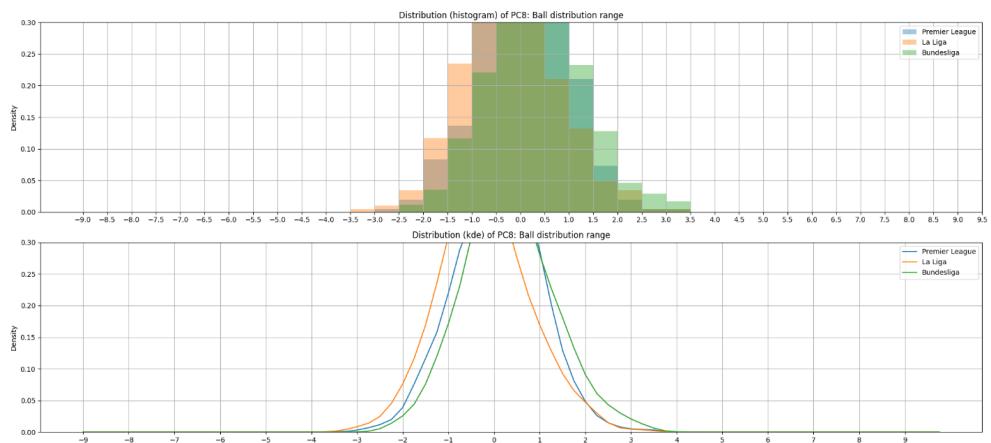


Figure S3.30: Distribution of PC_8 values

4 Extended Robustness and Validation Checks

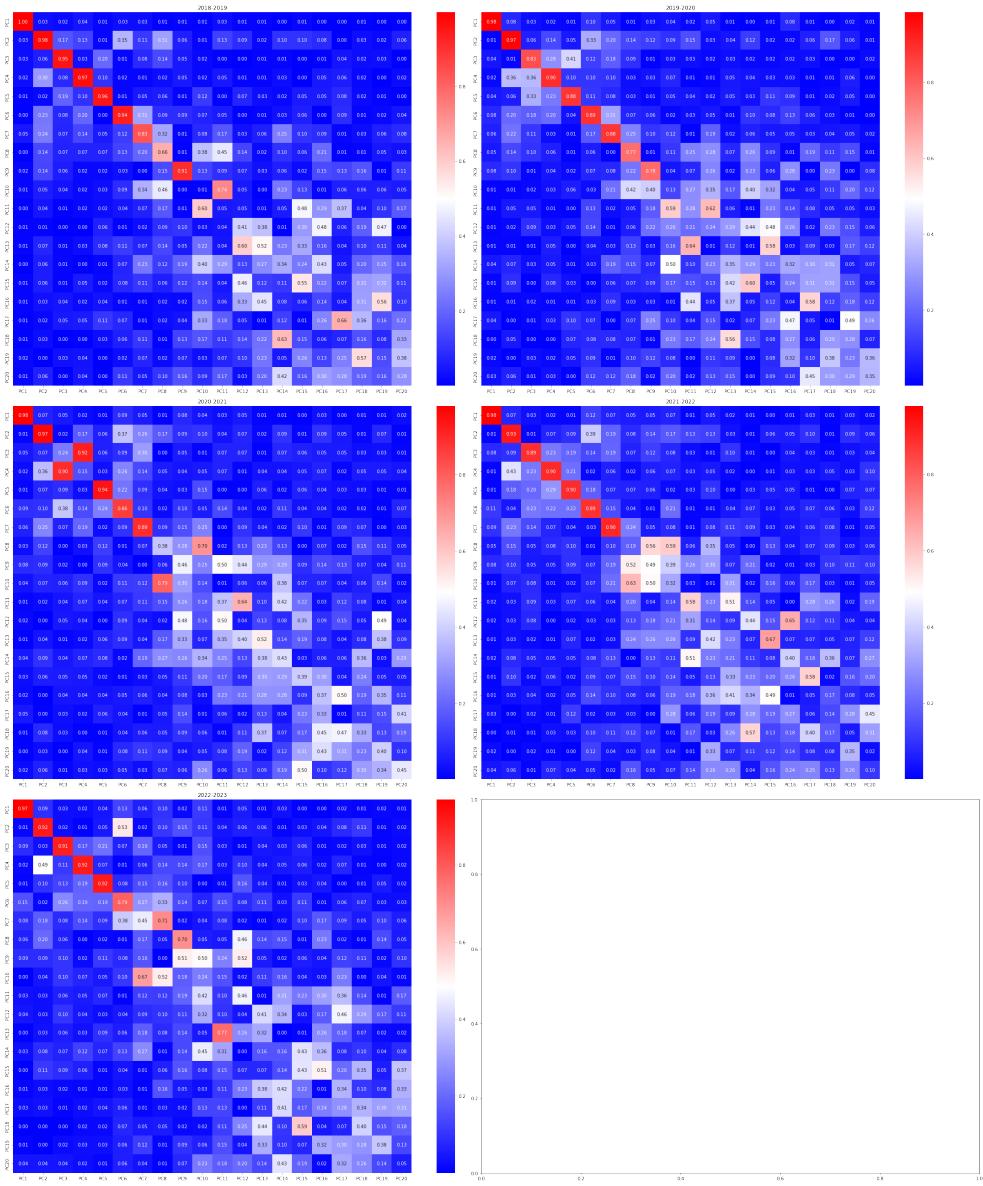


Figure S4.1: Correlations of EPL PC loadings with season 2017/18

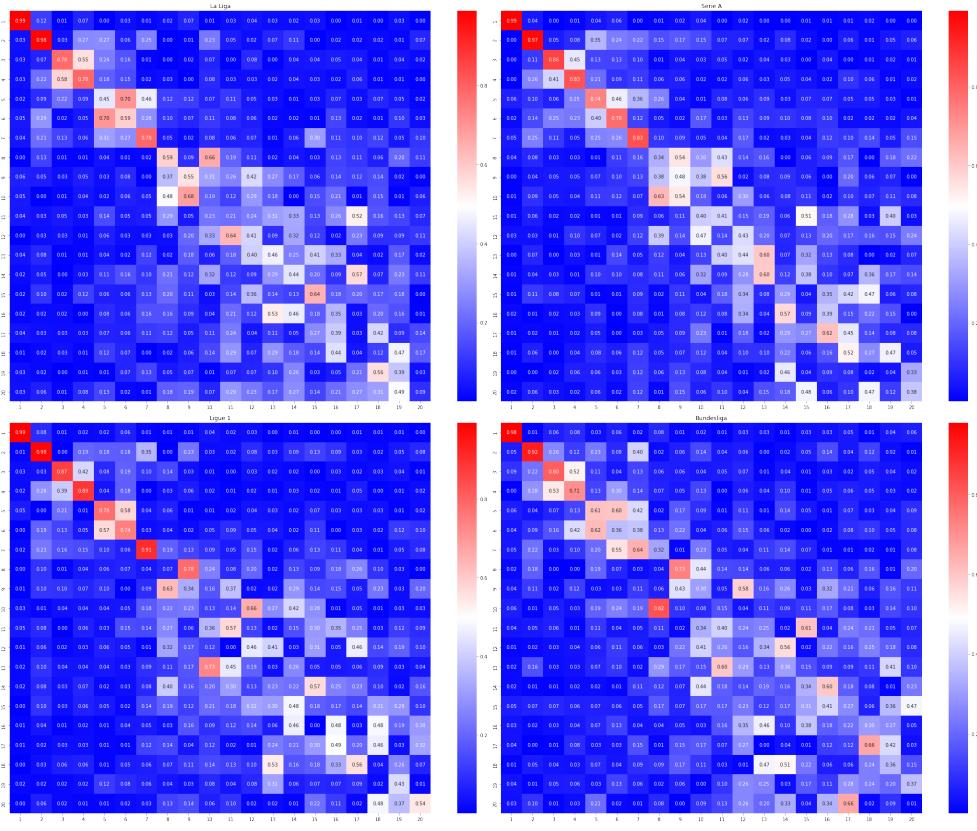


Figure S4.2: Correlations of league PC loadings with Premier League in 2017/18

5 Data and Code Availability

All scripts, processed feature matrices, and analysis notebooks are publicly available at https://github.com/CezaryKlimczuk/pca_football_profiling. The repository contains a CITATION.cff file for automatic citation export via GitHub’s “Cite this repository” button.