



Distributing xG: Giving Passes Due Credit



Kapil Khanal,¹ Nathan Moss,² and Jeremy Sanchez³

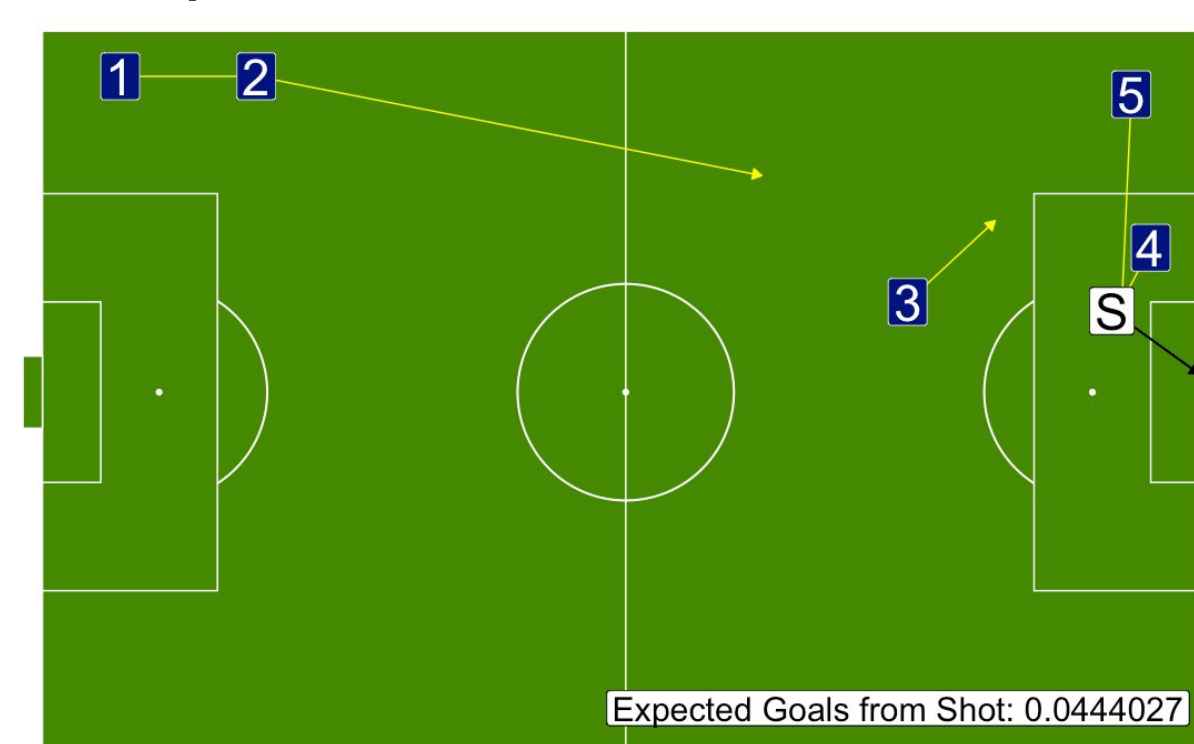
Project Supervisor: Kostas Pelechrinis⁴ Project Advisor: Xiaoyi Yang²

¹ Winona State University; ² Carnegie Mellon University; ³ University of Florida; ⁴ University of Pittsburgh

Background

- **Expected goals (xG)** - a measure of the likelihood that a shot results in a goal.
- **xG/pass** - distributes the expected goals at the end of a chain evenly to each pass.
- **What's the gap in the status quo?**
 - An even distribution seems too simplistic. Shouldn't better passes get more credit?
- **How are we filling the gap?**
 - We are building a model that gives better passes more credit for the shot.

Example Pass Chain and Distribution of xG

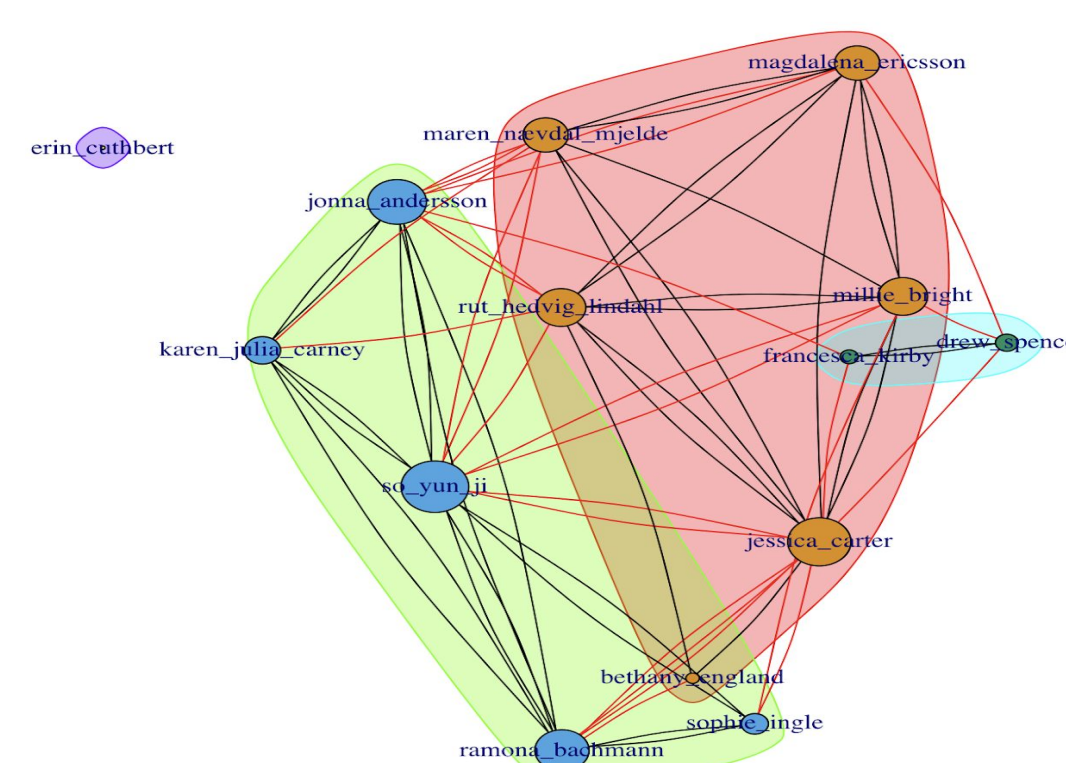


Expected Goals from Shot: 0.0444027

Data Processing and Variables

- The original event dataset, provided by **STATSBOMB**, detailed 258 matches from four different soccer leagues.
 - We subset data for **FA Women's Super League**:
 - Identified pass chains with and without shots at the end of the chain.
 - Gave passes the same xG as the shot if it was taken.
- At the end of this process, we had data with 100,235 events and ten variables.

VARIABLE	DESCRIPTION
ExpectedGoal (Response)	<ul style="list-style-type: none"> → xG based on StatsBomb's xG model. → Passes in a chain were given the xG of the shot. → Chains that didn't end in a shot were given 0 xG.
X, Y	→ X and Y coordinates of the origin of the pass/shot.
end.X, end.Y	→ X and Y coordinates of the place the pass was received (if successful) or intercepted (if not).
PageRank Origin/End	<ul style="list-style-type: none"> → Calculate passing network for each team by match. → PageRank is a measure of "Popularity and "influence" of each player. → Passes from and to an influential player gets a higher PageRank score.

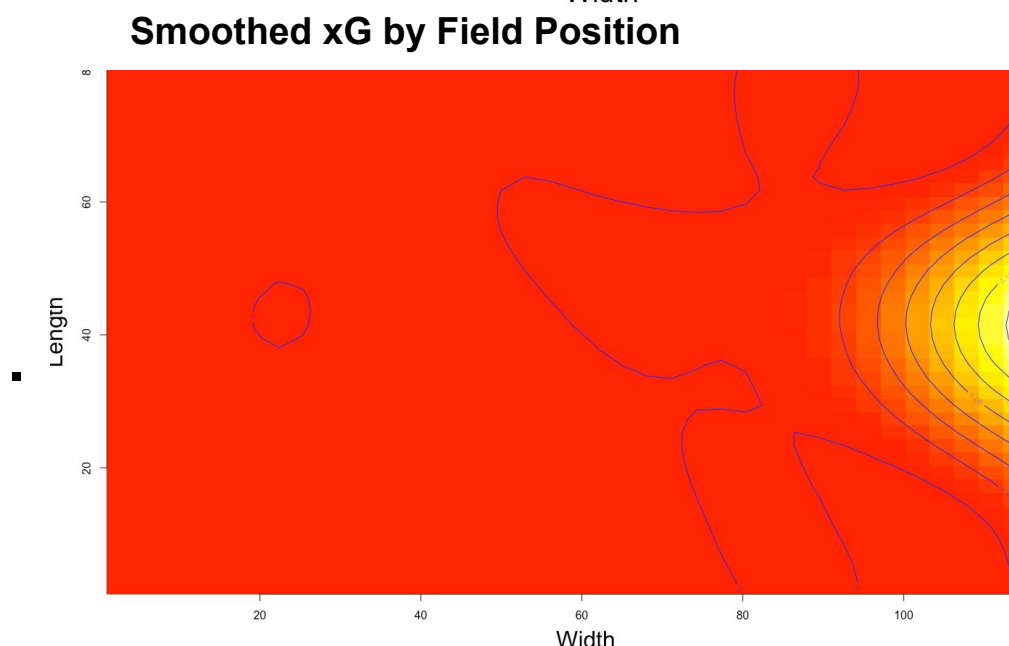
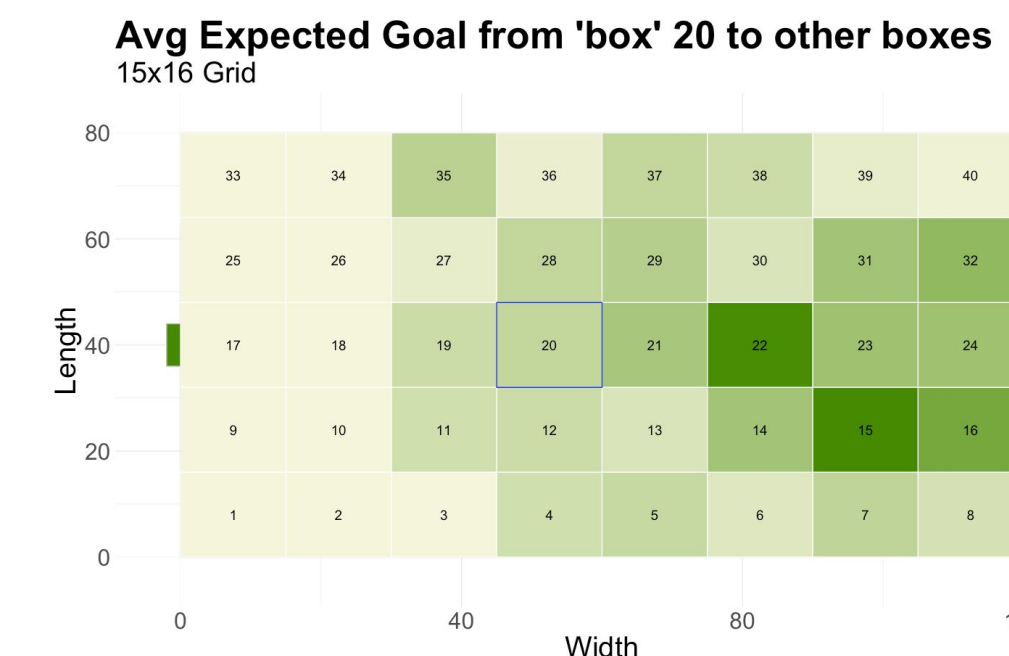


Passing Network with clusters as calculated by PageRank

Analysis and Results

Models:

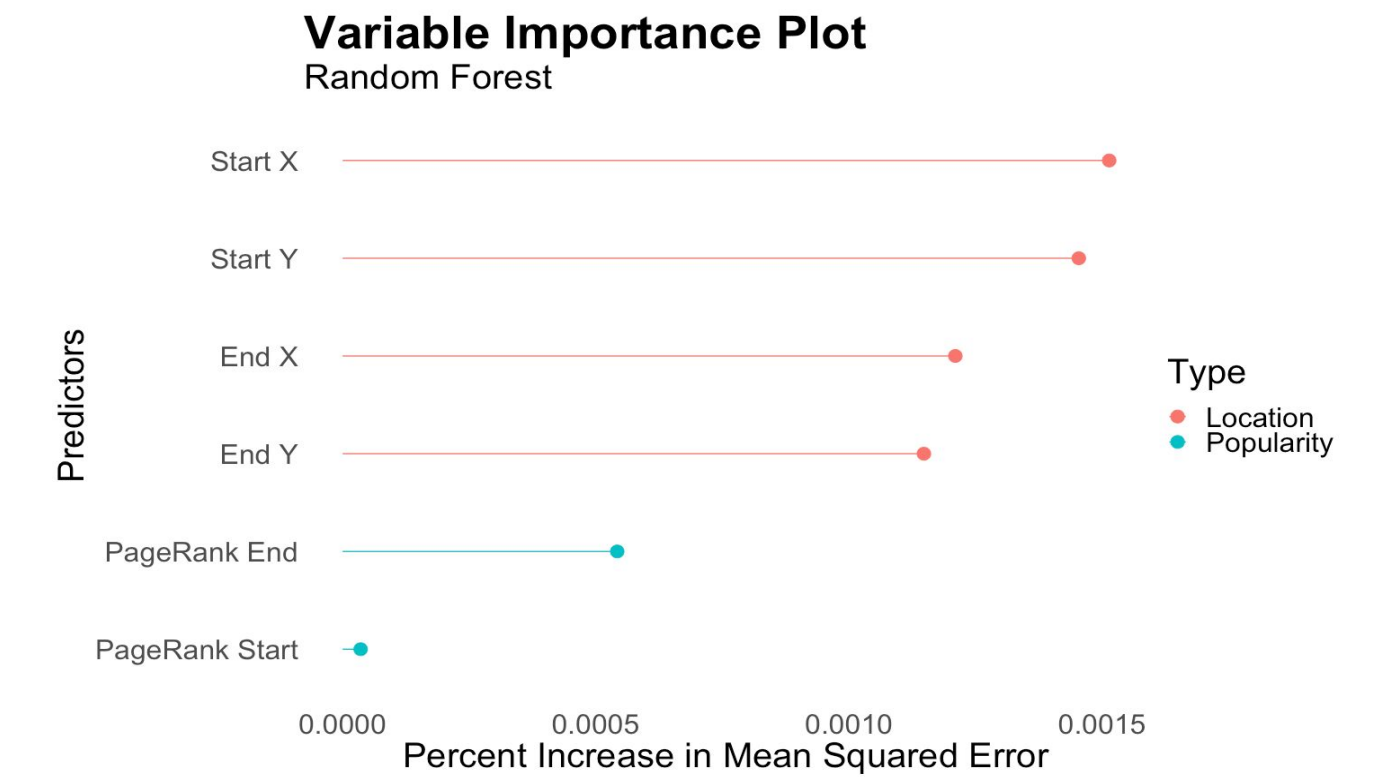
- **Grid-based model:**
 - Splits the field into boxes
 - Computes the average xG
 - Validate the box size in test set
 - 15 X 16 box size optimal
- **GAM:**
 - Local splines of location data
 - Smooths out the weight
 - REML Smoothness Selection.
- **Random Forest:**
 - Estimates xG with interactions
 - Number of trees: 500
 - Features at any split: 2



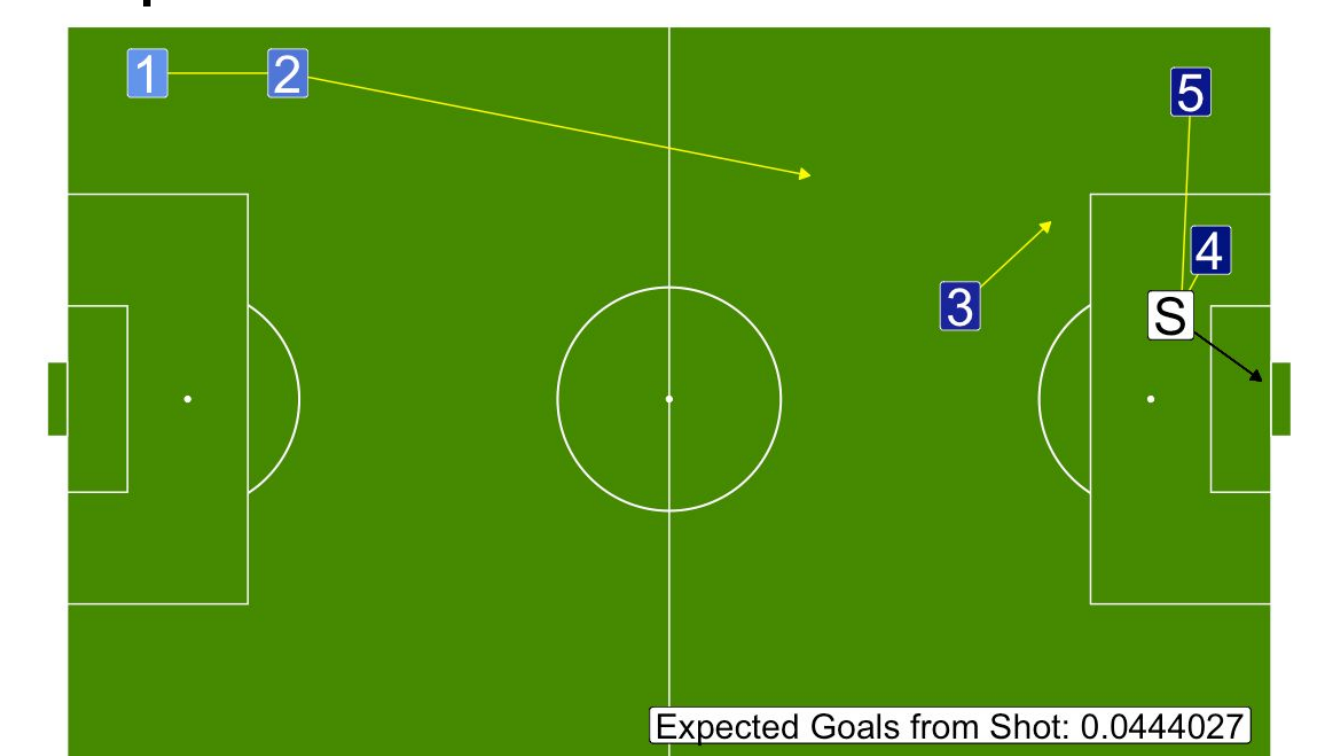
MODEL	MSE
Grid-based model	4.30 x 10 ⁻³
GAM: [X, Y]	4.26 x 10 ⁻³
GAM: [10 Variables]	3.95 x 10 ⁻³
RF: [10 variables]	3.44 x 10 ⁻³
RF: [Key variables]	2.47 x 10 ⁻³

Results and follow-up:

- Passing location matters the most, followed by the influence of the player.
 - We then created **xG/pass** by taking into account that each pass is part of a chain. We used the following formula, letting f be the value of a pass.
- $$xG/pass = \frac{f(pass)}{\sum_{i=1}^{chainLen} f(pass_i)} \cdot xGFromShot$$
- Now, passes colored dark blue are given more credit.



Improved Distribution of xG



References

- A PageRank Model for Player Performance Assessment in Basketball, Soccer and Hockey. Shael Brown. MIT Sloan Sport Analytics
- "Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data". Paul Power, Hector Ruiz, Xinyu Wei, Patrick Lucey. ACM

Contact Information

- Jeremy Sanchez: sanchezjeremy@ufl.edu
- Kapil Khanal: kkhanel16@winona.edu
- Nathan Moss: nmoss1@andrew.cmu.edu

Conclusions and Future Work

- **xG/pass is a better reflection of a team's work.**
 - Passes near the goal post are given more credit than those farther away.
 - Passes between influential players are given more credit.
- **Future work may include:**
 - Incorporating team tactics and players' position data in model
 - Evolving the passing network
 - Including shot as a node to reward player who takes the shot
 - Finding an effective way to quantify bad passes