

Expected Goals and Expected Pass Completion models have become standard metrics to benchmark player/team performance in both soccer and hockey. Current models are built using event datasets that capture only the locations of on-ball/puck actions and lack context surrounding actions such as the amount of defensive pressure or the locations of the defenders surrounding the target. By combining event and tracking data sources, we develop an augmented dataset that accounts for the positions of all nearby offensive and defensive players at the time of the event. This allows us to model of defensive pressure on both the shooter/passers and their targets on a continuous scale. We then build xG and xPass models using this augmented event dataset and show that these models outperform those built using standard event datasets.

Hockey Expected Goals (xG) Model

Dataset & Methods

- Dataset
 - 2018-19 NHL regular season
 - All shot attempts
 - 54 Event Features
 - 14 Tracking Features
- XGBoost classifier
- Model ran for:
 - Event + Tracking data
 - Event-only data
 - Tracking-only data
 - Basic data- shot distance/angle

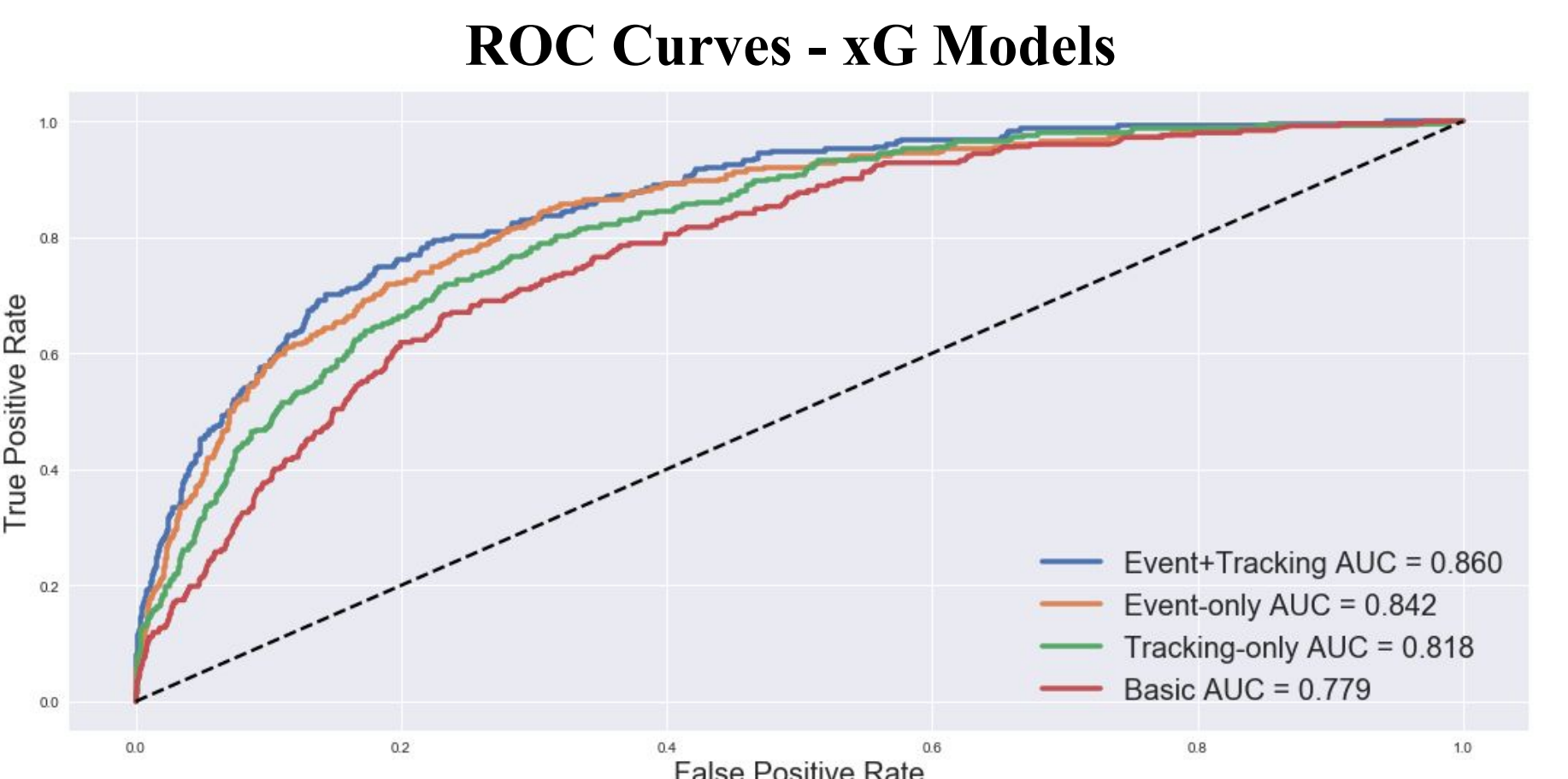
Event Data Features	
Shot Location	Distance, angle of shot
Pre-shot Movement	Distance, angle, time, velocity, angular velocity from reception to shot
Pre-reception Movement	Distance, angle, time, velocity, angular velocity from pre-reception event (e.g. pass) to reception
Game Context	Manpower situation, zone entry type, time since zone entry, empty net
Tracking Data Features	
Pressure	Distance, angle, speed of closest defender
Shot Clarity	Attackers/defenders in the shooting lane/slot
Team Shape	Area of convex hull for each team
Shooter	Speed, acceleration and body position of shooter

Results

- Sportlogiq's Event-only xG dataset has been feature engineered to provide robust performance
- The event-only model outperforms the tracking-only model across all measured metrics (log-loss, AUC, mean-squared error)
- Event+Tracking model outperforms both Event-only and Tracking-only models by a sizeable margin

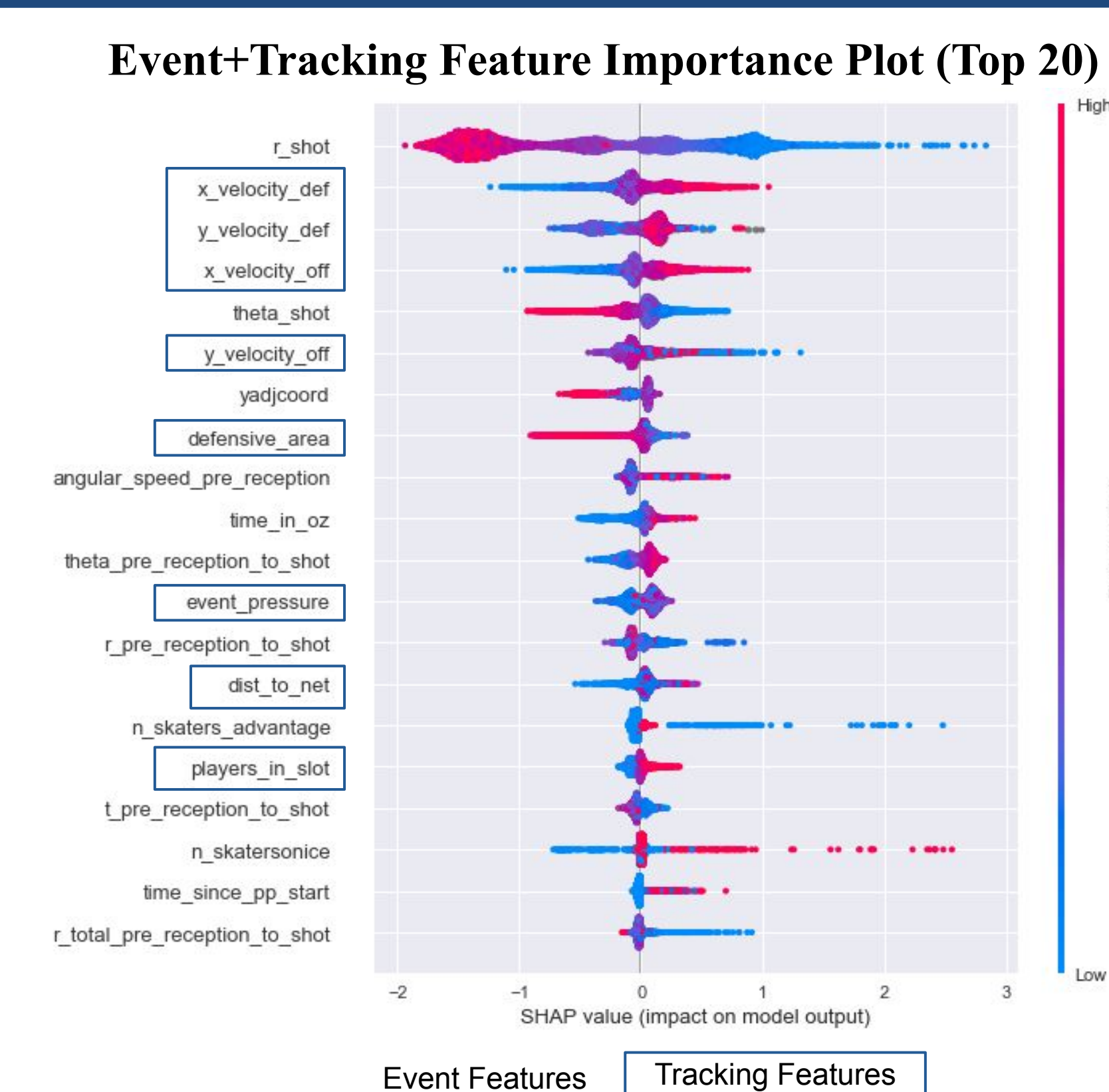
xG Model Fit Results

Model	Event+Tracking	Event	Tracking	Basic
Test log-loss	0.143	0.149	0.157	0.167
Test AUC	0.860	0.842	0.818	0.779
Test MSE	0.0375	0.0389	0.0405	0.0424



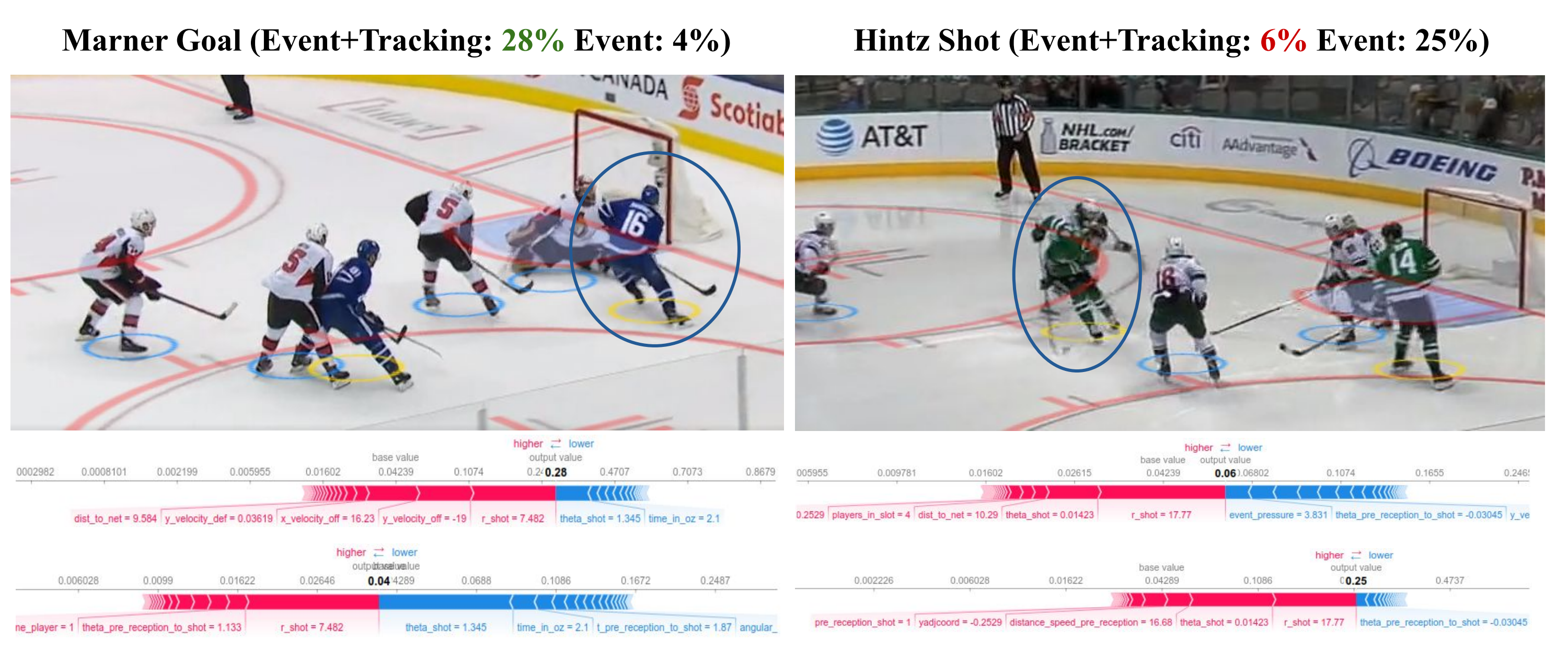
Feature Importance

- We use SHAP plots to help explain model outputs
- 8 of the top 20 most important features are derived from tracking data
- Most important tracking features are velocity of the closest defender/shooter (higher speeds towards the net leads to increased xG)
- The area of the defensive team and pressure distance also show high feature importance



Tracking improvements

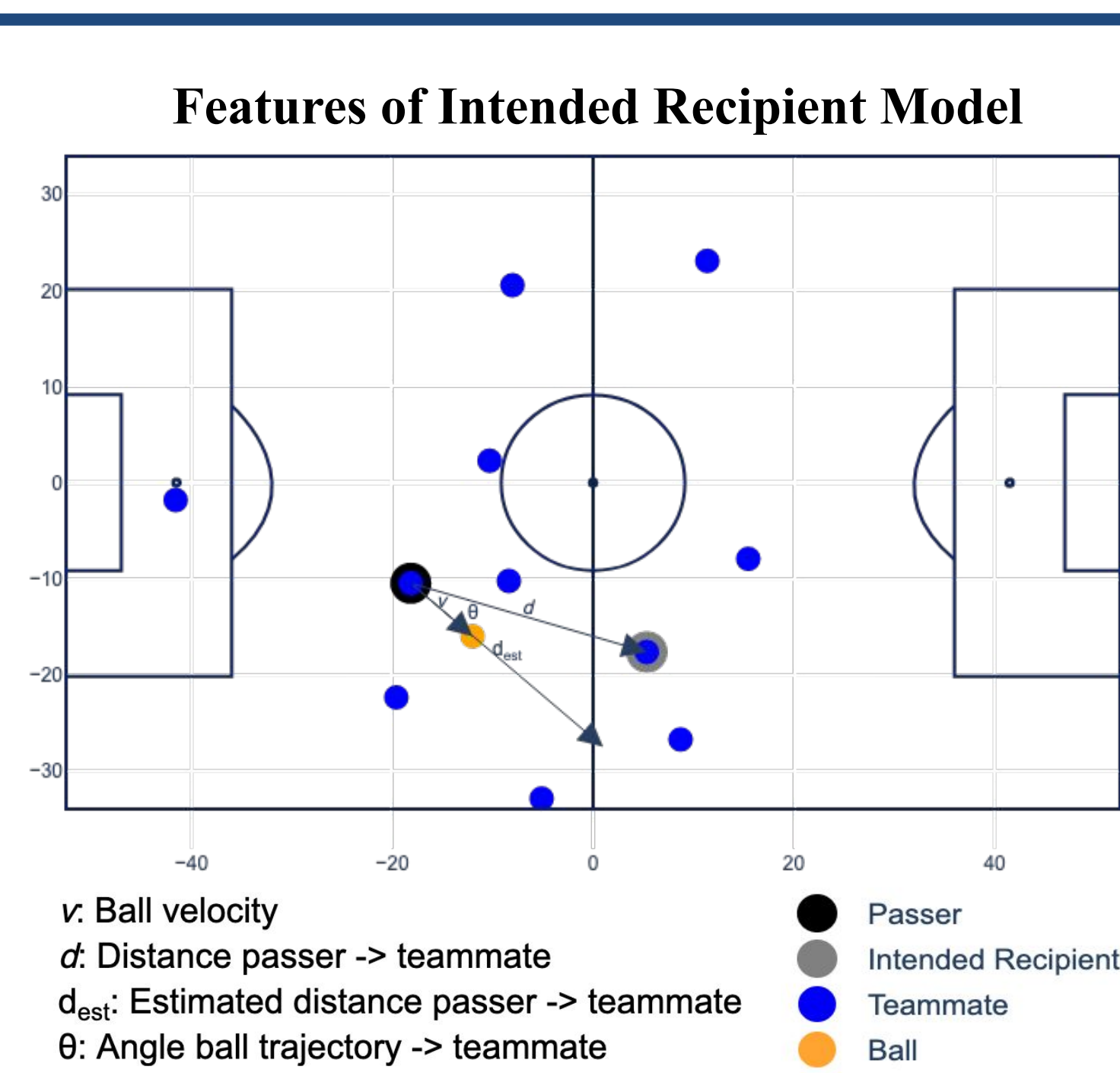
- The event+tracking and event-only models are well correlated (R-squared = 0.74)
- Examples where to identify value added by tracking data features



Soccer Expected Pass Completion (xPass) Model

Dataset & Methods

- Motivation
 - xPass models have inherent biases due to lack of contextual information, so that defenders and midfielders' passing abilities are over- and under-rated respectively.
- Dataset
 - Anonymised tracking and event data from a top European league (2018-19 season)
- Steps:
 - Estimate intended recipient on failed passes
 - Summarise opposition pressure on each pass
 - Build xPass model on the resulting features



Intended Recipient Model

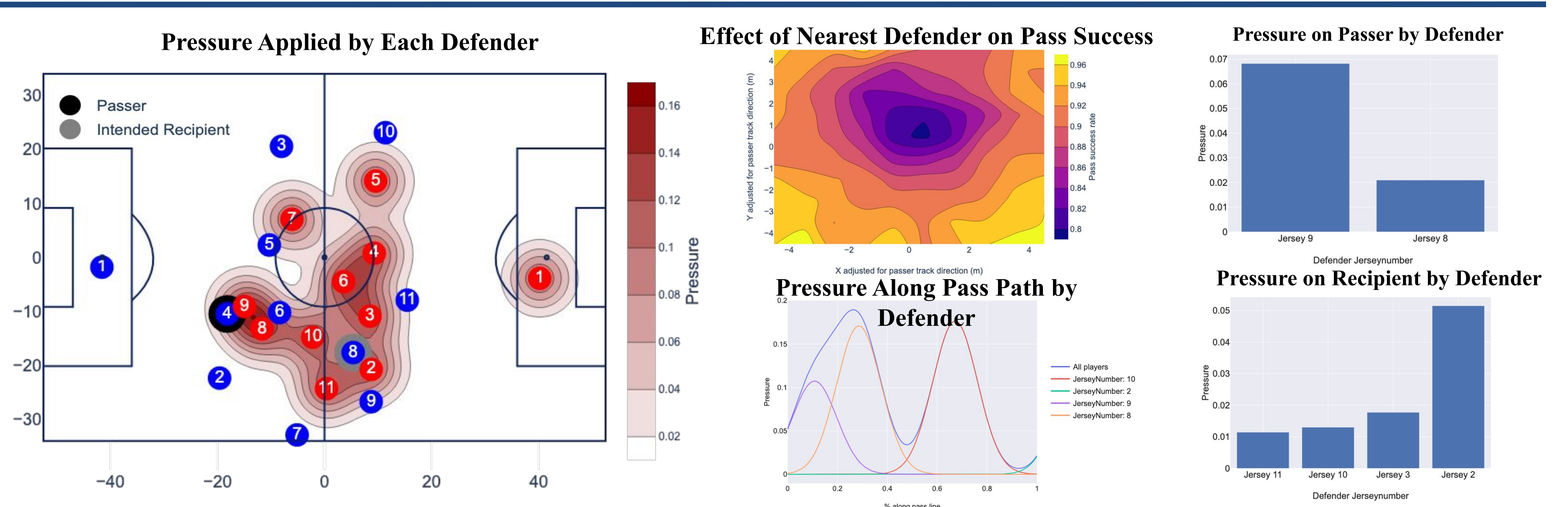
- XGBoost model trained on successful passes
- 10 observations per pass: one for each teammate
- Features
 - Expected pass distance
 - Distance between passer and teammate
 - Angle between ball trajectory and line between passer and teammate

Intended Recipient Model Results

Model	Test Accuracy
Naive model (random teammate)	10.0%
Model that selects nearest teammate	30.8%
Best model	87.9%

Feature Engineering

- Three measures summarise opposition positioning relative to the pass
 - Continuous pressure around the passing player
 - Continuous pressure around intended recipient
 - Proximity of opposition players to the path of the pass - "pressure along path of the pass"



Results

- The tracking-augmented model outperforms the standard xPass model
- Players in deep roles play passes high up the pitch when play is advanced and the threat of interceptions and blocks is low.
- This results in a bias among standard xPass models towards players who play deeper, so that midfielders are often rated much worse than defenders.
- The tracking-augmented xPass model corrects this bias.

xPass Model Fit Results

xPass Model	MSE	Log loss	Accuracy
Event-only	0.106	0.349	86.1%
Event+Tracking	0.095	0.306	86.7%

Per Pass Actual to Expected Pass % by Position

Position	Event-only	Event+Tracking
Defender	+3.1%	+0.8%
Midfielder	-0.9%	+0.3%
Forward	-7.2%	-3.7%

Feature Importance

- Tracking-derived features dominate the xPass model feature importances;
- X coordinate of the passer, the traditional rough proxy for context, has far less importance in the event+tracking model. This indicates that the pressure features are capturing this context effectively.

