Density-Functional Fluctuation Theory: A Common Statistical Framework for Predicting Football Outcomes from Player Positioning

Boris Barron, Max Planck Institute for Demographic Research
Irena Chen, Max Planck Institute for Demographic Research
Tahmeed Tureen, Independent Researcher

Introduction

Spatiotemporal tracking and freeze-frame data have created new opportunities to analyze football (soccer) at both the player and team levels. These data record the locations of the players and ball over the course of a match. As a result, they capture the spatial configurations (or positioning) of the players during any relevant football action such as a pass, shot, or corner kick etc. Statistical learning models can use these configurations to predict outcomes such as where a player will pass and whether that pass will be completed. Extracting useful information from a configuration, however, remains challenging because different outcomes may depend on different aspects of player positioning. Predetermined tactical features in a model—for example, the distance to the nearest defender or the number of players in a passing lane—necessarily encode assumptions about which players, regions, and interactions matter and may therefore omit informative patterns in the full spatial configuration. At the same time, the number of possible player arrangements far exceeds the number observed in any available dataset. Therefore, our primary research aim is to develop a common statistical framework that learns generalizable outcome-relevant positional patterns from *limited* data without requiring a predetermined set of tactical features.

To address this challenge, we adapt density-functional fluctuation theory (DFFT) to football. DFFT builds on classical density-functional theory (DFT), a physics-based framework for describing many-body systems through spatial densities. DFFT uses observed fluctuations in those densities to infer spatial preferences and interactions [1]. In the National Basketball Association (NBA), this approach has been used to predict player locations, relate player arrangements to scoring outcomes, evaluate individual defenders' positioning, and quantify player gravity (the tendency of particular attacking players to attract defenders) [2].

The central conceptual shift in DFFT is to represent configurations through local densities and treat the probability distributions of these densities as the core statistical objects. For player positioning, we begin by describing each frame through the local concentrations of attacking and defending players. Several attacking players gathered in one region produce a high attacking density there, whereas few nearby defenders produce a low defending density. Following the previous application to basketball [2], we represent each player by a Gaussian centered at their recorded position and sum these contributions

separately for the two teams. The resulting fields provide a smooth description of attacking and defending presence across the pitch. We then integrate these fields over spatial bins and discretize the resulting values, assigning each bin a pair of attacking and defending density levels for that frame. Figure 1 illustrates this transformation and the spatial detail retained after discretization. Alternative player arrangements that result in the same discretized densities (Figure 1g,h) remain close to the original, even when several players are repositioned together. Implementation details are provided in the Methods.

Each frame is therefore represented by a pair of attacking and defending density values in every spatial bin across the pitch. To see how this density representation can be used to predict an outcome, consider the direction of a pass from a given player arrangement. In the specific example here, consider that the ball is near $(x, y) = (6, 14)$ m and is passed either front left (FL) or right (R), as shown in Figure 2a. For a chosen bin b, we collect the attacking and defending density values separately for FL and R passes using frames available in our training data. The resulting joint probability distributions describe how frequently different density combinations occur in this bin for each outcome. For a new frame, we can then evaluate how probable the observed density combination is to occur using the FL distribution compared to the R distribution, providing local (bin-specific) evidence of which outcome is more likely. Combining evidence across spatial bins allows us to estimate which outcome is more likely for the player arrangement as a whole.

The central statistical challenge is therefore to estimate these distributions reliably in each spatial bin from limited data. Figure 2b,c illustrates this difficulty for FL passes: the empirical distributions from the training and held-out testing samples differ, and some density combinations observed in testing are absent from training. DFFT addresses this by combining distribution structure shared across spatial bins with bin-specific corrections [2]. Here, we develop more flexible corrections to capture more detailed local variation (see Methods). The resulting DFFT distributions, fitted to the training data, are shown in Figure 2d,g.

In this example, the DFFT fits describe the empirical training and held-out testing distributions comparably well, as measured by the Kullback–Leibler divergence (Figure 2d,g), suggesting generalization beyond the training set. The differences between the FL and R distributions also illustrate how density patterns provide an understanding of player arrangement and enable predictions. The considered bin b in Figure 2 lies in the FL direction from the ball: when both attacking and defending densities in this bin are near zero, the local evidence favors an R pass, consistent with little player presence in the FL direction. Substantial attacking density combined with little defending presence favors an FL pass, consistent with an available teammate to receive the pass without a defender nearby. Greater defending density alongside the attacker shifts the evidence back toward R (Figure 2j). Passing direction is therefore a particularly simple but useful illustration of the framework: the density distributions capture familiar tactical relationships without requiring those relationships to be specified in advance, and every bin across the pitch provides its own evidence regarding which outcome is more probable. Notably, the procedure is not specific to passing direction and can be applied to other outcomes for which the positional patterns are substantially more subtle and the relevant tactical features are not immediately apparent.

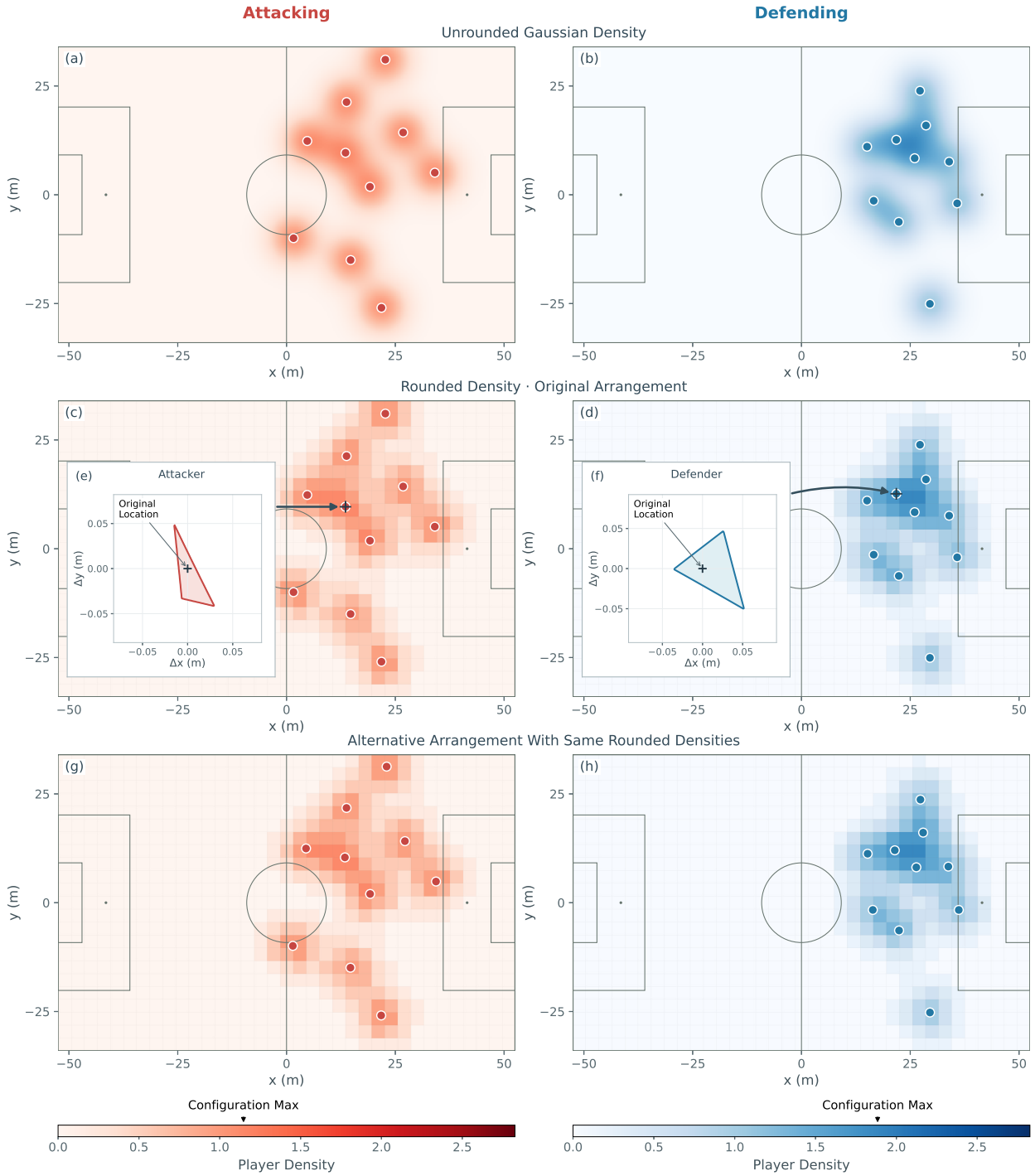


Figure 1: **From player configuration to density representations.** Attacking players are shown in red and defending players in blue. **(a, b)** Gaussian density fields before discretization, using a standard deviation of 4 m. **(c, d)** The corresponding densities integrated over spatial bins of 3×3 m and rounded to discrete levels. Values are scaled so that 1 corresponds to the contribution of a single player standing at the center of a bin. The upper density cutoff is selected from the training data (see Methods). **(e, f)** Insets show numerically estimated regions within which the highlighted player can move without changing any discretized bin density, with all other players fixed; crosses mark the original positions. **(g, h)** Alternative joint arrangements that reproduce exactly the discretized densities in (c, d) differ only slightly from the original configuration, demonstrating the spatial detail retained after discretization.

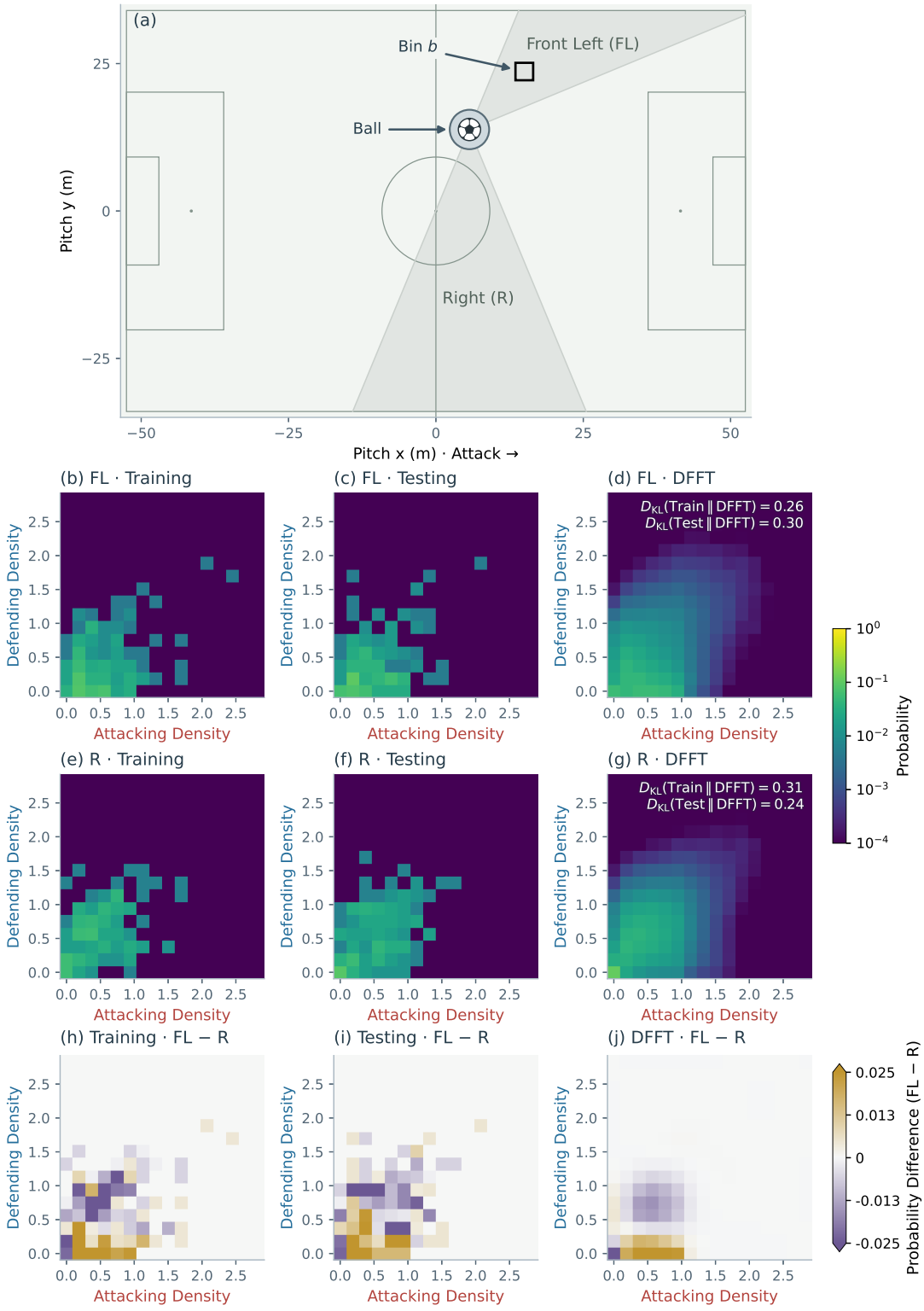


Figure 2: **Illustrating passing-direction prediction through local densities. (a)** Ball-origin region, front-left (FL) and right (R) passing directions, and bin b. **(b–g)** Joint attacking and defending density distributions in bin b for FL (top) and R (bottom). Columns show training data, held-out testing data, and DFFT fits obtained from the training data. The KL divergences (in nats) show that DFFT represents training and testing distributions with similar efficacy, demonstrating an ability to generalize. **(h–j)** FL-minus-R differences: gold indicates density combinations more probable under FL, purple under R. The color scale saturates at ± 0.025.

To examine this broader applicability, we develop and evaluate DFFT across a range of football prediction tasks, from passing direction and completion to corner-kick outcomes. We begin by distinguishing between two passing directions at a given ball location, then extend the approach to joint prediction across eight directions—the first multi-outcome version of DFFT. We next construct a *positional baseline* for pass completion, using ball and player positions *without specifying the direction, distance, or body part used for the pass*. Comparing this baseline with xPass [3] allows us to examine players that tend to attempt passes that are harder or easier than expected from the player arrangement alone, and then determine whether the observed player-specific success rate exceeds these expectations. Finally, we consider corner-kick outcomes, from delivery to team first-touch. Across these applications, we retain the same core density-based construction, examining how reliably it can extract predictive information when the observations available for different outcomes vary substantially.

Data

We use Hudl StatsBomb event and 360 freeze-frame data from the 2025/26 season across five European top-division football leagues: the English Premier League, Spanish La Liga, Italian Serie A, French Ligue 1, and German Bundesliga. The event data record individual match events, including the player involved, the location on the pitch, and the outcome. Related-event identifiers link associated actions, allowing us to examine events connected to a pass or corner kick.

The 360 freeze-frame data provide spatial context by recording the coordinates of players visible in the broadcast frame at the time of an event. These positions are linked to the ball location through the corresponding event data. Players outside the visible area are not recorded. The data also include the vertices of a polygon describing the visible area, allowing the spatial coverage of each frame to be identified. Figure E.15 illustrates the StatsBomb coordinate system.

We were additionally provided access to Hudl broadcast tracking data for 30 matches from the 2025/26 Italian Serie A season. These data provide estimated player positions throughout the match, including positions inferred for players outside the broadcast frame. Such off-camera positions are therefore model-based estimates rather than positions directly observed in the corresponding image.

For all analyses, we retain only frames in which all outfield players are visible; for corner kicks, we additionally require the defending goalkeeper to be visible. These complete observations provide a consistent basis for developing the framework across prediction tasks, with a particular focus on extracting predictive information from limited frames of player arrangements. Extending the framework to partial visibility would require accounting for how unobserved players affect the density distributions linked across spatial bins. We leave this extension for future work, focusing here on developing and evaluating the methodology for learning from limited samples of fully observed player arrangements.

2.1 Processing

Each pass or corner kick is linked to its corresponding freeze frame or broadcast tracking positions, representing the player arrangement when the ball is played. We refer to this arrangement as the *start*

frame and use it as the spatial input for predicting the outcome of interest. Recorded pass endpoints define passing directions and delivery regions, while event outcomes and related subsequent actions identify pass completion and corner-kick outcomes. The outcome definitions and selection criteria are specified for each analysis.

2.2 Coordinate System

Player and ball positions are expressed in meters relative to the center of the pitch. The positive x-direction points toward the goal attacked by the passing or corner-taking team, and the positive y-direction points to that team's left. This reference team remains fixed when analyzing subsequent events, including actions by the opposition. Where actual pitch dimensions are unavailable, positions are mapped to a nominal pitch of $105 \times 68\,\mathrm{m}$. Coordinate transformations and the alignment of related events are described in Supplementary Section E; any additional reflection used to combine symmetric situations is specified with the corresponding analysis.

2.3 Training and Testing

Training and testing data are separated by match, with matches assigned randomly and all observations from a given match kept together. The passing-direction analyses are restricted to one ball location and therefore contain relatively few eligible passes. A 50/50% split allocates comparable numbers of matches to fitting and evaluation. The pass-completion analyses include passes across the pitch, providing enough testing observations to use an approximately 80/20% split and retain more data for fitting. The splits are being done at the match-level, so the the number frames in the training/testing sets do not follow them exactly.

Direct comparisons between model variants use identical testing matches. Where different training-set sizes are compared, the smaller training set is contained within the larger one, and both models are evaluated on a common set of matches excluded from training.

Methods

The method first estimates distributions of attacking and defending densities for each outcome, then learns how strongly different spatial bins contribute to predicting the outcome. We extend the earlier DFFT formulation [2] by allowing more flexible local corrections of density distributions. The intention is to capture distribution differences from different bins and outcomes while limiting sensitivity to sampling noise. We first describe the density representation, then explain how the distributions are estimated and their evidence combined into outcome probabilities.

3.1 Representing Player Densities

Each player is represented by a Gaussian centered at their recorded location. The player's contribution to a spatial bin is given by the Gaussian mass within that bin, normalized so that a player standing at the bin's center contributes one. Summing these contributions separately for the two teams gives the local attacking and defending densities.

For a bin b centered at coordinates $\mathbf{x}_b = (x_b, y_b)$, these densities are

$$a_b = \frac{\sum_{p \in A} \int_b G_\sigma(\mathbf{x} - \mathbf{x}_p)\,\mathrm{d}\mathbf{x}}{\int_b G_\sigma(\mathbf{x} - \mathbf{x}_b)\,\mathrm{d}\mathbf{x}}, \qquad d_b = \frac{\sum_{p \in D} \int_b G_\sigma(\mathbf{x} - \mathbf{x}_p)\,\mathrm{d}\mathbf{x}}{\int_b G_\sigma(\mathbf{x} - \mathbf{x}_b)\,\mathrm{d}\mathbf{x}}. \tag{1}$$

Here, A and D are the sets of attacking and defending players relevant in the analysis, $\mathbf{x}_p$ is the position of player p, and G_σ is a two-dimensional Gaussian with standard deviation σ along each coordinate axis. Each integral is taken over the full area of the bin.

Analyses in this work use $\sigma = 4\,\mathrm{m}$ and a common grid of $3 \times 3\,\mathrm{m}$ square bins. The 35×23 grid covers the nominal $105 \times 68\,\mathrm{m}$ pitch, extending $0.5\,\mathrm{m}$ beyond each of its two long sidelines. Passing analyses use all 805 bins. Corner analyses use the 414 bins whose centers lie on the halfway line or in the attacking half.

For each analysis, a common upper density limit is selected from the training frames and used for all outcomes and model variants being compared. It is chosen so that, in every modeled bin, attacking and defending densities each exceed this limit in fewer than 1% of training frames. Each density is rounded to the nearest of 16 equally spaced levels between zero and this limit, with larger values capped at the highest level. This gives 256 possible attacking–defending density combinations per bin. The same cutoff and discretization are applied to validation and testing data. In the following sections, a and d denote these discretized density values. For each model fit, bins whose rounded attacking and defending densities are both zero in every frame used for fitting are excluded.

The passing representation includes all 20 outfield players and excludes both goalkeepers. For corners, the attacking density excludes the corner taker and attacking goalkeeper, while the defending density includes the defending goalkeeper.

3.2 Learning Density Distributions

Let $p_{kb}(a, d)$ denote the empirical training distribution: the relative frequency with which attacking density a and defending density d occur together in bin b, among training frames with outcome k. The aim is to represent the underlying distribution as accurately as the available data allow, taking into account sampling noise. A shared reference distribution $q_0(a, d)$ captures structure common across bins and outcomes, with corrections specific to each bin and outcome allowing the distributions to differ. These corrections are expressed using functions $\phi_j(a, d)$. Their training averages, $\sum_{a,d} \phi_j(a, d) p_{kb}(a, d)$, describe aspects of the distribution such as the mean attacking density.

The original NBA formulation used local corrections linear in the two densities [2]. Here, polynomials up to degree three in each density, including their products, allow more flexible corrections while retaining a small number of local coefficients. The constant term cancels upon normalization, leaving 15 nonconstant functions. These polynomials provide a convenient way for local distributions to differ from the reference, although other functions could also be considered.

To limit overfitting, penalties control the size of the local corrections and their differences across outcomes. The aim is to capture meaningful differences between the density distributions while limiting sensitivity to sampling noise.

Let $N_{\text{densities}} = 16^2 = 256$ denote the number of allowed joint density combinations. The functions used in this work are centered and orthonormalized:

$$\frac{1}{N_{\text{densities}}}\sum_{a,d}\phi_j(a,d) = 0, \qquad \frac{1}{N_{\text{densities}}}\sum_{a,d}\phi_i(a,d)\phi_j(a,d) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{2}$$

Orthonormalization puts the functions on a common scale and removes overlap between them over this grid, making penalties on their coefficients comparable.

Suppose that, for a given bin b and outcome k, the averages of the functions $\phi_j(a,d)$ are known. Information projection provides a way to update the reference distribution to reproduce these averages. Formally, information projection is

$$\begin{aligned} q^*_{kb} &= \operatorname*{argmin}_q D_{\mathrm{KL}}(q\|q_0), \\ \text{subject to} \qquad & q(a,d) \geq 0, \qquad \sum_{a,d} q(a,d) = 1, \\ & \sum_{a,d}\phi_j(a,d)q(a,d) = \sum_{a,d}\phi_j(a,d)p_{kb}(a,d) \qquad \text{for every } j. \end{aligned} \tag{3}$$

The first two conditions ensure that q is a probability distribution, whereas the final condition requires it to reproduce the selected averages. The projection therefore finds the distribution closest to the reference, as measured by the KL divergence, among those matching these averages. Provided the constraints can be satisfied, the optimal probability distribution is unique.

Applying the method of Lagrange multipliers shows that the optimal distribution takes the form

$$q_{kb}(a,d) = \frac{q_0(a,d)\exp\left[-\sum_j \lambda_{kbj}\phi_j(a,d)\right]}{\sum_{a',d'} q_0(a',d')\exp\left[-\sum_j \lambda_{kbj}\phi_j(a',d')\right]}. \tag{4}$$

The exact projection q^*_{kb} is obtained by choosing the coefficients λ_{kbj} to satisfy the specified average constraints.

The benefit of this construction follows from the Pythagorean identity satisfied by the exact projection:

$$D_{\mathrm{KL}}(p_{kb}\|q_0) = D_{\mathrm{KL}}(p_{kb}\|q^*_{kb}) + D_{\mathrm{KL}}(q^*_{kb}\|q_0). \tag{5}$$

The identity shows that matching the correct averages improves upon the reference distribution, and holds for any distribution sharing those averages. For a fixed reference distribution, the KL divergence from the empirical distribution to the fitted distribution is convex in the correction coefficients. Moving partway along the line from zero corrections toward the optimal coefficients therefore also improves the representation, unless the reference already satisfies the averages. Exact matching is not required to obtain an improvement.

In practice, the averages are estimated from finite samples. The challenge is therefore to retain the information they provide while limiting sensitivity to sampling noise. In the implemented method, the reference distribution and local coefficients are learned together using the objective below, without imposing exact agreement with the training averages. This balances agreement with the observed

density distributions against the size of the local corrections, allowing information across bins and outcomes to support estimates where observations are limited. An additional prediction term encourages density differences that help distinguish outcomes. Alternative functions, including functions learned from data, could be considered within the same construction in future work.

3.3 Controlling Flexibility in Density Fitting

The additional coefficients allow richer distributions than those of Barron et al. [2], but require regularization when observations are limited. We separate each coefficient into its average across outcomes and an outcome-specific departure:

$$\lambda_{kbj} = \mu_{bj} + \delta_{kbj}, \qquad \mu_{bj} = \frac{1}{K}\sum_{k=1}^{K} \lambda_{kbj}, \tag{6}$$

where K is the number of outcomes. Penalizing μ limits the average local departure from the shared distribution. Penalizing δ encourages outcomes to share their local corrections unless the data support a difference.

We also allow predictive performance to influence density fitting. Let π_k denote the observed fraction of outcome k in the training data. These fractions can serve as fixed baseline probabilities, so that equal density evidence across outcomes reproduces their training frequencies. More generally, the baseline probabilities can be fixed at other estimates of the overall outcome frequencies, which is useful when reliable estimates are available. In the extension used here, these baseline probabilities are instead learned through outcome intercepts α_k fitted alongside the density distributions. For each frame, a provisional prediction combines these intercepts with the local density probabilities:

$$\widetilde{P}_i(k) = \frac{\exp(\alpha_k)\prod_b q_{kb}(a_{ib}, d_{ib})}{\sum_\ell \exp(\alpha_\ell)\prod_b q_{\ell b}(a_{ib}, d_{ib})}, \qquad \mathcal{C} = -\frac{1}{N}\sum_{i=1}^{N} \log \widetilde{P}_i(y_i), \tag{7}$$

where a_{ib} and d_{ib} are the observed rounded densities in bin b, y_i is the actual outcome, and N is the number of training frames. The loss $\mathcal{C}$ decreases as more probability is assigned to the outcomes that occurred.

Using p_{kb} for the empirical training distribution defined above, we fit the shared distribution, local coefficients, and outcome intercepts by minimizing

$$\mathcal{J} = \frac{1}{B}\sum_{b,k} \pi_k D_{\mathrm{KL}}(p_{kb} \| q_{kb}) + \eta\mathcal{C} + \frac{\rho_\mu}{B}\sum_{b,j} \mu_{bj}^2 + \frac{\rho_\delta}{KB}\sum_{k,b,j} \delta_{kbj}^2. \tag{8}$$

Here, B is the number of included bins: those with nonzero attacking or defending density in at least one training frame. The first term measures how well the fitted distributions represent the observed density distributions, averaging across bins and weighting outcomes by their training frequencies. The second term encourages distributions that help predict outcomes, with η controlling its contribution. The final two terms limit the common and outcome-specific corrections, with larger ρ_μ and ρ_δ imposing stronger penalties. The intention is to balance density fit, predictive usefulness, and flexibility.

3.4 Frame Weighting

Nearby ball locations can provide additional training frames for estimating the local density distributions. To incorporate these observations while emphasizing the ball location of interest, the Gaussian-weighted passing analyses assign each training frame a weight

$$\omega_i = \begin{cases} \exp\left(-\dfrac{r_i^2}{2h^2}\right), & r_i \le 10\,\mathrm{m}, \\ 0, & r_i > 10\,\mathrm{m}, \end{cases} \qquad h = 3\,\mathrm{m}, \tag{9}$$

where r_i is the distance between the frame's ball location and the target location. Frames closer to the target therefore contribute more strongly to density fitting.

The empirical density distributions, outcome fractions, and provisional prediction loss are then calculated using these weights:

$$p_{kb}^{(\omega)}(a,d) = \frac{\sum_{\substack{i:\, y_i=k \\ (a_{ib},d_{ib})=(a,d)}} \omega_i}{\sum_{i:\, y_i=k} \omega_i}, \qquad \pi_k^{(\omega)} = \frac{\sum_{i:\, y_i=k} \omega_i}{\sum_i \omega_i}, \qquad \mathcal{C}^{(\omega)} = -\frac{\sum_i \omega_i \log \widetilde{P}_i(y_i)}{\sum_i \omega_i}. \tag{10}$$

These weighted quantities replace p_{kb}, π_k, and $\mathcal{C}$ in Equation (8); the penalty terms remain unchanged. The frame weights and weighted outcome fractions remain fixed during fitting, while the outcome intercepts are learned alongside the density distributions.

The subsequent spatial prediction weights and intercepts are fitted using equally weighted training frames within 3 m of the target ball location. Validation also uses equally weighted frames within this region. For single-location passing-direction prediction, testing uses the same 3 m region; the unweighted comparisons also restrict density fitting to this region. For pass-completion prediction across the pitch, each testing pass is assigned to one model according to its ball-location grid cell, as described with that analysis. Corner analyses use equal frame weights throughout.

3.5 Combining Evidence Across the Pitch

Different bins need not be equally informative, and neighboring densities contain overlapping information. We therefore learn nonnegative spatial weights and an intercept separately for each pair of outcomes.

For outcomes k and ℓ, the comparison score is

$$S_i^{k\ell} = \beta_{k\ell} + \sum_b w_b^{k\ell} \log \frac{q_{kb}(s_{ib})}{q_{\ell b}(s_{ib})}, \qquad s_{ib} = (a_{ib}, d_{ib}). \tag{11}$$

A bin supports outcome k when its observed density combination is more probable under k than under ℓ, while the weight determines how strongly this bin's evidence contributes to the final prediction.

The resulting probability of k, when comparing between two possible outcomes, can be written

$$r_i^{k\ell} = \frac{1}{1+\exp(-S_i^{k\ell})} = \frac{\exp(\beta_{k\ell}) \prod_b q_{kb}(s_{ib})^{w_b^{k\ell}}}{\exp(\beta_{k\ell}) \prod_b q_{kb}(s_{ib})^{w_b^{k\ell}} + \prod_b q_{\ell b}(s_{ib})^{w_b^{k\ell}}}. \tag{12}$$

The intercept $\beta_{k\ell}$ can either be fixed or learned. To fix the prediction baseline at observed or otherwise estimated outcome frequencies π_k, set

$$\beta_{k\ell} = \log \frac{\pi_k}{\pi_\ell}.$$

When the positional evidence is equal, Equation (12) then returns $\pi_k/(\pi_k + \pi_\ell)$, recovering the corresponding frequency ratio. In this work, $\beta_{k\ell}$ is instead fitted alongside the spatial weights, separately from the density-stage intercepts α_k.

The logistic rule balances evidence favoring one outcome with entropy, which favors retaining uncertainty. For a fixed score $S = S_i^{k\ell}$, it selects the probability r that maximizes $rS + H(r)$, where $H(r) = -r \log r - (1-r)\log(1-r)$ is the entropy of the two-outcome prediction. Since $H'(r) = \log[(1-r)/r]$, setting the derivative to zero gives $r = 1/(1+\exp[-S])$. Thus, the logistic form arises naturally: stronger evidence produces a more decisive prediction, while finite scores retain some probability for both outcomes.

3.5.1 Learning Spatial Weights

For each outcome pair, the spatial weights and intercept are fitted by minimizing prediction loss with a penalty on differences between neighboring weights. Let $\mathcal{I}_{k\ell}$ contain the eligible fitting frames with outcome k or ℓ, let $N_{k\ell} = |\mathcal{I}_{k\ell}|$, and set $z_i = 1$ for outcome k and $z_i = 0$ for outcome ℓ. The objective is

$$\begin{aligned}\mathcal{L}_{k\ell} = & -\frac{1}{N_{k\ell}} \sum_{i \in \mathcal{I}_{k\ell}} \left[z_i \log r_i^{k\ell} + (1 - z_i) \log(1 - r_i^{k\ell}) \right] \\ & + \frac{\kappa}{2 \max(E, 1)} \sum_{(b,c) \in \mathcal{E}} \left(w_b^{k\ell} - w_c^{k\ell} \right)^2,\end{aligned} \tag{13}$$

subject to $w_b^{k\ell} \geq 0$, with the intercept $\beta_{k\ell}$ unrestricted and unpenalized. Here, $\mathcal{E}$ contains horizontal and vertical neighboring bin pairs in the analysis region, each counted once, and $E = |\mathcal{E}|$. The penalty encourages neighboring bins to contribute similarly, with $\kappa = 100$ throughout. Bins whose attacking and defending densities are both zero in every pair-specific fitting frame have their weights fixed to zero. Neighbor pairs are defined on the full bin layout, including those connected to these fixed-zero weights.

3.5.2 Combining Multiple Outcomes

With two outcomes, Equation 12 directly gives the prediction. With more outcomes, each pair has its own spatial weights and intercept, so the pairwise probabilities need not agree perfectly with one another.

For each frame, we combine the pairwise predictions into a single set of outcome probabilities $P_k > 0$, with $\sum_k P_k = 1$. When considering only outcomes k and ℓ, these probabilities imply a probability $P_k/(P_k + P_\ell)$ for outcome k. We choose the overall probabilities to make these implied comparisons agree as closely as possible with the learned pairwise predictions, by minimizing

$$-\sum_{k<\ell} \left[r_{k\ell} \log \frac{P_k}{P_k + P_\ell} + (1 - r_{k\ell}) \log \frac{P_\ell}{P_k + P_\ell} \right]. \tag{14}$$

Each pair contributes equally. If the pairwise predictions are mutually consistent, the resulting probabilities reproduce them exactly; otherwise, the optimization finds the best compromise under this loss.

If the pairwise intercepts are fixed at $\beta_{k\ell} = \log(\pi_k/\pi_\ell)$, equal positional evidence across all outcomes gives $P_k = \pi_k$. This property is not imposed when the intercepts are fitted. Related outcomes can subsequently be combined by summing their probabilities. For example, summing attacker-first-touch and defender-first-touch probabilities for one destination gives the probability of delivery to that destination.

3.6 Selecting Hyperparameters and Evaluating Predictions

Testing matches are held out before fitting. Within the training data, we create ten fixed validation blocks formed from whole matches. For each block, both the density distributions and the spatial prediction weights and intercepts are fitted using only the remaining matches. The resulting model predicts the outcomes of eligible frames in the held-out block.

The search settings for the passing and four-region corner analyses are summarized below. The default starting values are $(\eta, \rho_\mu, \rho_\delta) = (0.01, 0.01, 1)$. For pass completion, an available solution from a neighboring ball location is used instead when available. The maximum change in any logarithmic coordinate in one step is $\log 2$. The candidate with the lowest validation loss encountered during the search is retained.

Setting	Passing Direction and Four-Region Corners	Pass Completion
Bounds for η	$[10^{-7}, 1]$	$[10^{-7}, 10]$
Bounds for ρ_μ, ρ_δ	$[10^{-7}, 10^3]$	$[10^{-7}, 10^3]$
Maximum objective evaluations	150	100
Maximum optimization steps	100	80
Projected-gradient tolerance	10^{-5}	10^{-6}

The projected-gradient tolerance is applied in logarithmic coordinates. The spatial smoothness strength is fixed rather than included in this search.

We select the hyperparameters η, ρ_μ, and ρ_δ by minimizing the mean validation log loss:

$$\mathcal{L}_{\text{val}} = -\frac{1}{N_{\text{val}}} \sum_{i=1}^{N_{\text{val}}} \log P_{i,\text{val}}(y_i), \tag{15}$$

where N_{val} is the total number of validation frames across the blocks. Each validation frame contributes equally to this loss.

The search uses bounded BFGS optimization in logarithmic coordinates. Implicit differentiation provides gradients through the density fits, spatial weights and intercepts, and combination of pairwise predictions, allowing us to avoid separate fits for small perturbations of each hyperparameter. For fixed hyperparameters, the fitting problems are convex in their implemented parameterizations; the outer validation objective is not generally convex.

After selection, both the density distributions and the spatial prediction weights and intercepts are refitted using all training matches, with the frame selection and weighting specified for each analysis. We then evaluate the resulting model on the held-out testing matches.

Applications of DFFT

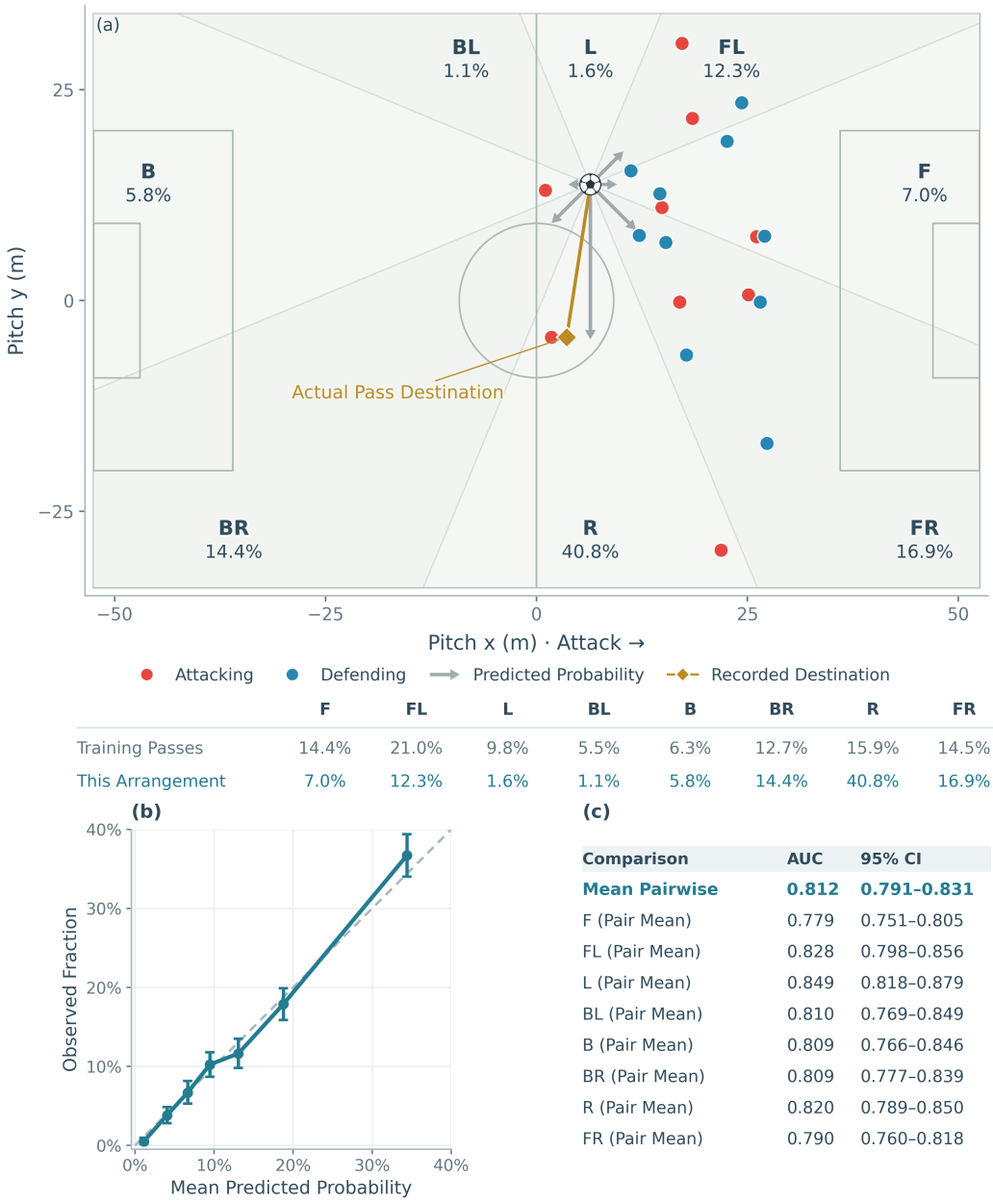
4.1 Predicting Passing Direction

We first examine whether the arrangement of players can be used to predict passing direction. Building on the two-direction illustration in the Introduction, we apply the multi-outcome formulation to predict eight passing directions jointly. Training uses approximately half of the matches, including 1,119 passes in the local ball-origin region and additional nearby passes incorporated through Gaussian weighting of the ball location. The testing sample contains 1,008 passes from 350 held-out matches.

Figure 3a shows a held-out player arrangement, with grey arrow lengths indicating the predicted probabilities. Compared with the direction frequencies in the local training sample, this arrangement favors R, which is also the recorded direction; the gold diamond marks the actual destination. Additional examples are provided in Appendix A.2.

Across the testing sample, the mean pairwise AUC is 0.812 (95% confidence interval: 0.791–0.831), averaged over the 28 direction pairs. Each direction's mean AUC across its seven pairwise comparisons ranges from 0.779 for F to 0.849 for L (Figure 3c). Thus, the predictive information is not confined to the FL–R comparison used in the introductory example. Across all eight probabilities from each testing frame, predicted probabilities broadly agree with observed frequencies (Figure 3b). Individual-direction and pairwise calibration plots, together with all pairwise AUCs, are provided in Appendix A.3.

The eight-direction model also retains similar FL–R discrimination to a dedicated two-direction model on the same testing passes (AUCs of 0.850 and 0.854, respectively; Appendix A.4).



	F	FL	L	BL	B	BR	R	FR
Training Passes	14.4%	21.0%	9.8%	5.5%	6.3%	12.7%	15.9%	14.5%
This Arrangement	7.0%	12.3%	1.6%	1.1%	5.8%	14.4%	40.8%	16.9%

Comparison	AUC	95% CI
Mean Pairwise	**0.812**	**0.791–0.831**
F (Pair Mean)	0.779	0.751–0.805
FL (Pair Mean)	0.828	0.798–0.856
L (Pair Mean)	0.849	0.818–0.879
BL (Pair Mean)	0.810	0.769–0.849
B (Pair Mean)	0.809	0.766–0.846
BR (Pair Mean)	0.809	0.777–0.839
R (Pair Mean)	0.820	0.789–0.850
FR (Pair Mean)	0.790	0.760–0.818

Figure 3: **Predicting passing direction from player positioning. (a)** A held-out arrangement with attackers in red and defenders in blue; the ball symbol obscures the player in possession. Grey arrow lengths indicate predicted probabilities, and the gold diamond marks the recorded pass destination. Percentages compare local training frequencies with predictions for this arrangement. F/B and L/R denote front/back and left/right. **(b)** Calibration of all eight probabilities from each of 1,008 testing passes, grouped into seven probability groups. The dashed line denotes perfect calibration. **(c)** Mean AUC across all 28 direction pairs and each direction's seven comparisons. Error bars and intervals show 95% confidence intervals (see Methods). Additional AUC and calibration results are provided in Appendix A.3.

4.2 Predicting Pass Completion

Pass-completion DFFT estimates whether a pass will be completed from ball and player positions, without specifying the direction, distance, or body part used for the pass. It therefore provides a *positional baseline* for completion before the attempted pass is specified. Figure 4 summarizes predictions for 22,481 passes from 364 held-out matches, comparing completed passes with blocked, intercepted, or recovered passes. We train a seperate model for each ball location on the pitch, so the effective sample size of training frames is in the low thousands, and often in the hundreds for failed passes. The calibration curves compare predicted and observed completion rates over the 60–100% range across these ball-location specific models, which contains 95.6% of DFFT predictions. High-confidence xPass predictions tend to underestimate completion, this discrepancy is smaller when all non-completion events are treated as failure outcomes, but appears to not disappear (Appendix B.3).

To examine how well the model distinguishes successful and failed passes, Figure 4b compares AUC within each of the eight directions we considered for passing in the previous subsection. We emphasize that direction is used only to group the testing passes, and the same pass-completion DFFT is evaluated throughout that was not trained on direction-specific data. Despite this, DFFT and xPass have similar AUC estimates for BR (0.878 versus 0.885), R (0.880 versus 0.894), and L (0.844 versus 0.850). For BL, the estimates are 0.788 and 0.754, respectively, with wider uncertainty, while xPass performs better for F and FR, likely because of greater variation in passing distance. These comparisons are promising given the limited local training data—the median effective sample size for failed passes in the local density fits is 728, ranging from 88 to 1,357 across locations—and DFFT's omission of pass-specific information used by xPass, including angle, distance, and body part.

Comparing the positional baseline with xPass also helps distinguish passing choices from completion relative to those choices (Figure 4c). Players to the right have lower mean xPass than mean DFFT, suggesting they attempt harder passes relative to the positional baseline. The vertical axis shows completion relative to this baseline, while the dashed line indicates completion matching xPass. For example, Joris Chotard lies below the DFFT baseline but close to the dashed line: his observed completion is close to the xPass expectation for the passes attempted. The profiles combine validation and testing predictions. Regional comparisons, player selection, and ranked summaries are provided in Appendix B.

4.3 Corner Kicks

We next examine corner kicks, considering both the delivery destination and which team makes first touch. Summing over teams gives delivery probabilities, while summing over regions gives first-touch probabilities. Both predictions therefore use the player arrangement at the kick, without supplying the delivery destination. The model is fitted to 982 corners from 465 matches and evaluated on 985 corners from 465 different matches, and the four delivery regions here are chosen so that in the training set there is roughly an equal number of delivies into each.

Figure 5a,b illustrates how the player arrangement changes the predicted delivery probabilities relative to the training frequencies. For the shown corner, the probability of delivery to region 3 increases from 23.0% to 35.0%; the recorded delivery also ends in this region. Across the testing sample, delivery probabilities broadly agree with observed frequencies, with a mean pairwise AUC of 0.673 (95% CI:

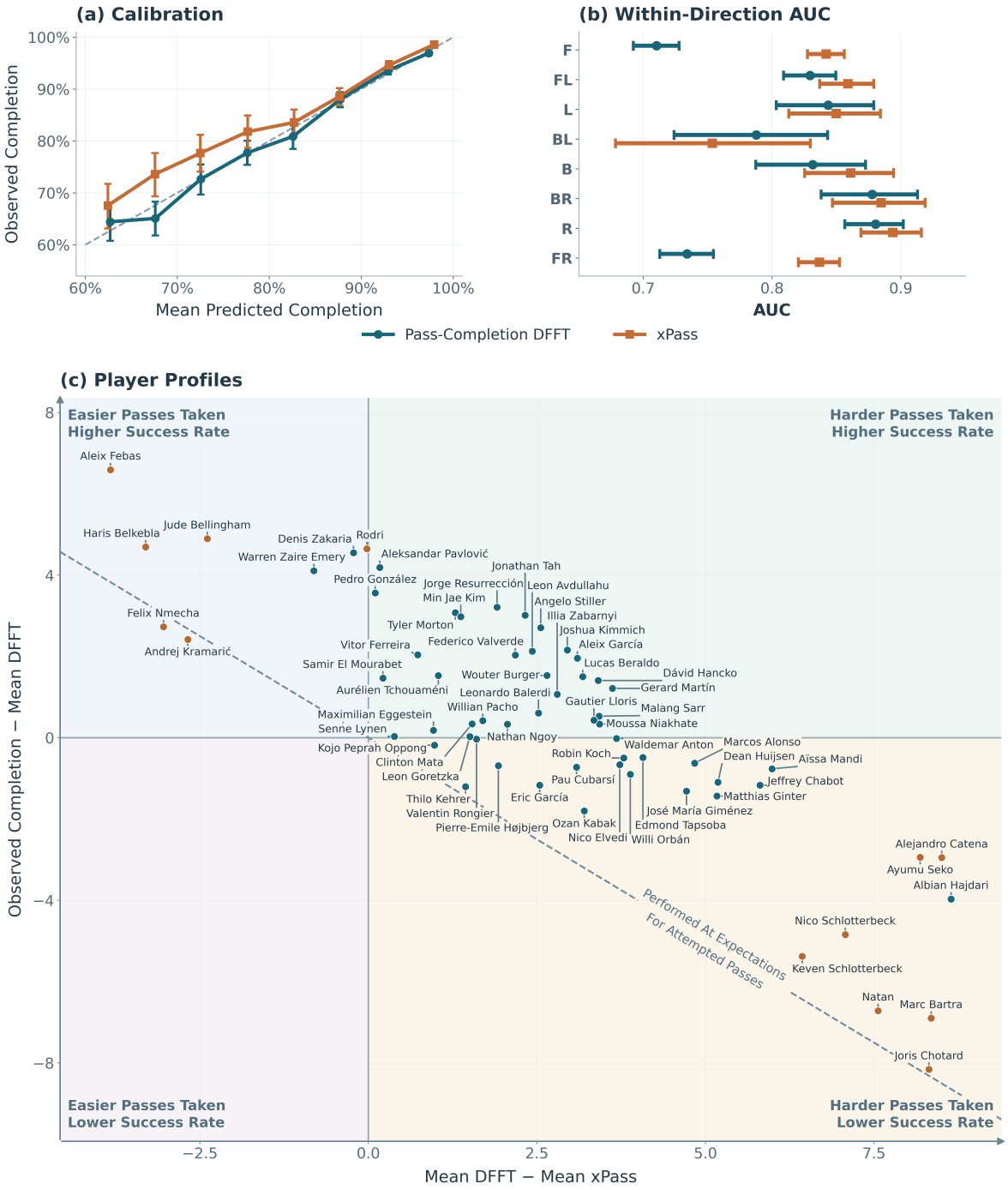


Figure 4: **Pass completion and player profiles. (a)** Calibration of DFFT and xPass on held-out matches, using five-percentage-point probability groups. **(b)** AUC for successful versus failed passes within each direction, using all testing predictions. Direction is not supplied to DFFT. Intervals in (a,b) are 95% confidence intervals from resampling whole testing matches. **(c)** Player profiles combining validation and testing predictions. The axes compare mean DFFT with mean xPass, and observed completion with mean DFFT, in percentage points. The dashed line indicates observed completion equal to mean xPass. Teal points show the 50 most-observed players; orange points show 13 additional profiles selected as described in Appendix B.2.

0.646–0.699; Figure 5c). Individual region comparisons range from 0.625 to 0.756 (Appendix C).

For first touch, the AUC is 0.557 (95% CI: 0.519–0.594), indicating a modest predictive signal from the initial arrangement (Figure 5d). The additional information supplied by positioning varies across delivery regions, with the clearest discrimination in region 4. Appendix C provides region-specific analysis for first-touch in the regions themselves.

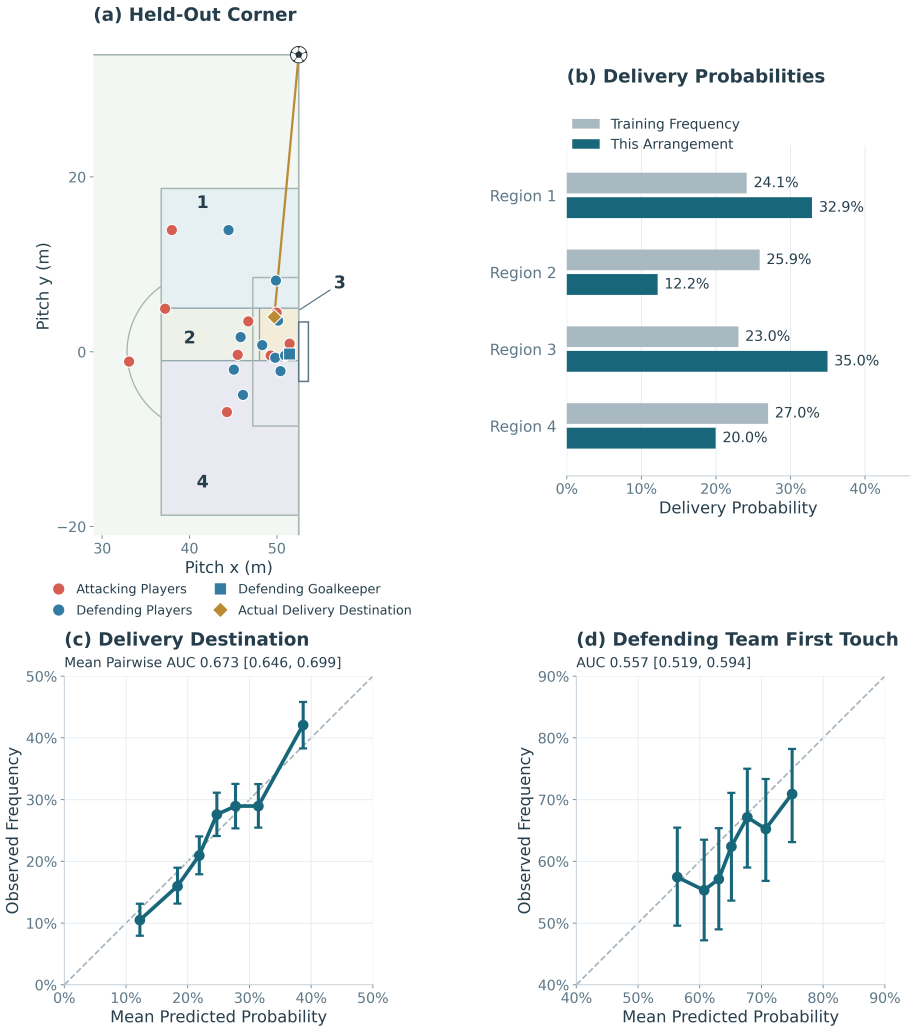


Figure 5: **Predicting corner delivery and first touch from player positioning. (a)** A held-out corner and the four delivery regions. Attacking players are red, defending players blue, and the defending goalkeeper is a square. The ball marks the corner-taking position; the gold line and diamond indicate the recorded delivery endpoint. **(b)** Training frequencies and predicted delivery probabilities. **(c)** Calibration using all four delivery probabilities from each testing corner. **(d)** Calibration of the probability of defending-team first touch. Both predictions come from the same joint model using only kick-time positioning, with seven calibration groups. AUCs are shown above the panels; all intervals are 95% whole-match bootstrap intervals from 10,000 resamples.

Exploratory Extension: Sequential Analysis of Corner Kicks

5.1 Motivation

The preceding corner-kick analysis shows that DFFT can predict delivery regions and extract a modest first-touch signal directly from the player configuration when the kick is taken. That analysis establishes that an explicit model of the intervening play is not required to form a pre-kick prediction. It does not, however, describe how the starting configuration becomes the configuration that ultimately contests the delivery.

A corner begins from a controlled dead-ball situation, but its outcome emerges through a sequence of related events. The initial configuration may indicate where the ball will be delivered; while the ball is in flight, attacking runs, defensive marking, and player interactions alter that configuration; and the configuration that forms around the delivery location influences which team makes first contact. These changes motivate a test of whether explicitly representing the delivery and spatial transition adds useful information beyond the initial configuration alone.

We investigate this question using a hybrid sequential framework with three linked stages:

$$\rho_0 \longrightarrow P(Z \mid \rho_0) \longrightarrow P(\rho_1 \mid \rho_0, Z) \longrightarrow P(Y \mid \rho_1, Z),$$

where ρ_0 denotes the configuration when the corner is taken, Z the delivery region, ρ_1 the configuration at first contact, and Y whether the corner-taking team makes that contact. This factorization separates uncertainty about the delivery, the subsequent spatial change, and the first-contact outcome. In the present implementation, DFFT models delivery choice and first-contact control, whereas a PCA–ridge transition estimates the conditional mean $\widehat{\mathbb{E}}[\rho_1 \mid \rho_0, Z]$ rather than the full distribution $P(\rho_1 \mid \rho_0, Z)$. We use this composition to test whether predictable spatial redistribution improves first-contact prediction, not to assume that an explicit transition is necessary.

This exploratory extension has three aims. First, we illustrate how the density representation describes overload and congestion in realized first-contact contests. Second, we assess how well the delivery destination and later density configuration can be predicted from the corner start. Third, we test whether composing the delivery, transition, and contact models improves first-contact prediction beyond a corresponding model composition that does not use the transition. These analyses examine a possible direction for sequential DFFT research; the paper's central evidence for DFFT does not depend on the sequential model providing an improvement.

5.2 Question and Relation to the Direct Corner Model

The four-region model in Section 4.3 establishes that the kick-time configuration contains information about both corner delivery and first touch. The initial configuration does not, however, describe the later configuration that contests the delivered ball, and the paired frames show substantial spatial redistribution before first contact (Appendix D.3). We therefore investigate whether explicitly modeling the delivery destination and this configuration transition improves first-contact prediction beyond a model composition using the starting configuration alone.

The two analyses serve different purposes. The core analysis jointly predicts four approximately

balanced delivery regions and first touch from kick-time positioning. The exploratory analysis instead uses seven groups—including tactically familiar near- and far-post deliveries—to connect destination prediction to group-specific transition and contact models. Different eligibility rules, frozen match partitions, and AUC summaries mean that their results are not directly comparable; Appendix D.1 provides a compact comparison.

5.3 Study Cohort

The source data contain 16,720 corners from the five leagues in this study. The delivery-model cohort contains 1,878 of these corners with complete start configurations (minimum ten attacking players, eleven defending players visible). We define the delivery as the recorded endpoint of the corner and assign it to one of the seven groups in Figure 6. We use *first contact*—equivalently termed *first touch* in Section 4.3—to mean the first recorded player contact following the delivery. Its outcome indicates whether that player represents the corner-taking team or the opposition.

The model-specific cohorts are summarized in Table 1. The paired cohort contains both frames for the same corner, with the same ten attacking outfield players, ten defending outfield players, and defending goalkeeper represented at each time.

Cohort	Training	Held-out test	Total
Delivery prediction	1,510	368	1,878
Inside first-contact model	569	124	693
Paired start–contact transition	478	106	584

Table 1: Cohorts for the three-stage sequential corner modeling pipeline.

Training and test partitions are match-disjoint: all corners from one match remain in the same partition. The delivery split contains 720 training and 188 held-out matches; the paired split contains 324 training and 77 held-out matches. In total, 1,035 complete first-contact frames are available, of which 693 belong to an inside delivery group.

5.4 Spatial Scales for Tactical Summaries

We summarize attacking and defending densities at four scales: the 3-m endpoint cell, the original detailed delivery zone, the seven-group delivery region, and the complete stored half-pitch. The first three describe increasingly broad versions of the contest around the delivery, while the complete surface is used for map-level diagnostics. Figure 6 shows how the twelve detailed zones are aggregated for the seven-group delivery model.

5.5 Spatial Configurations as Tactical Representations

The density representation in Figure 1 can also be summarized using familiar tactical concepts. For spatial bin b, let A_b and D_b denote the attacking and defending densities. We define local congestion and overload as

$$C_b = A_b + D_b \qquad \text{and} \qquad O_b = A_b - D_b.$$

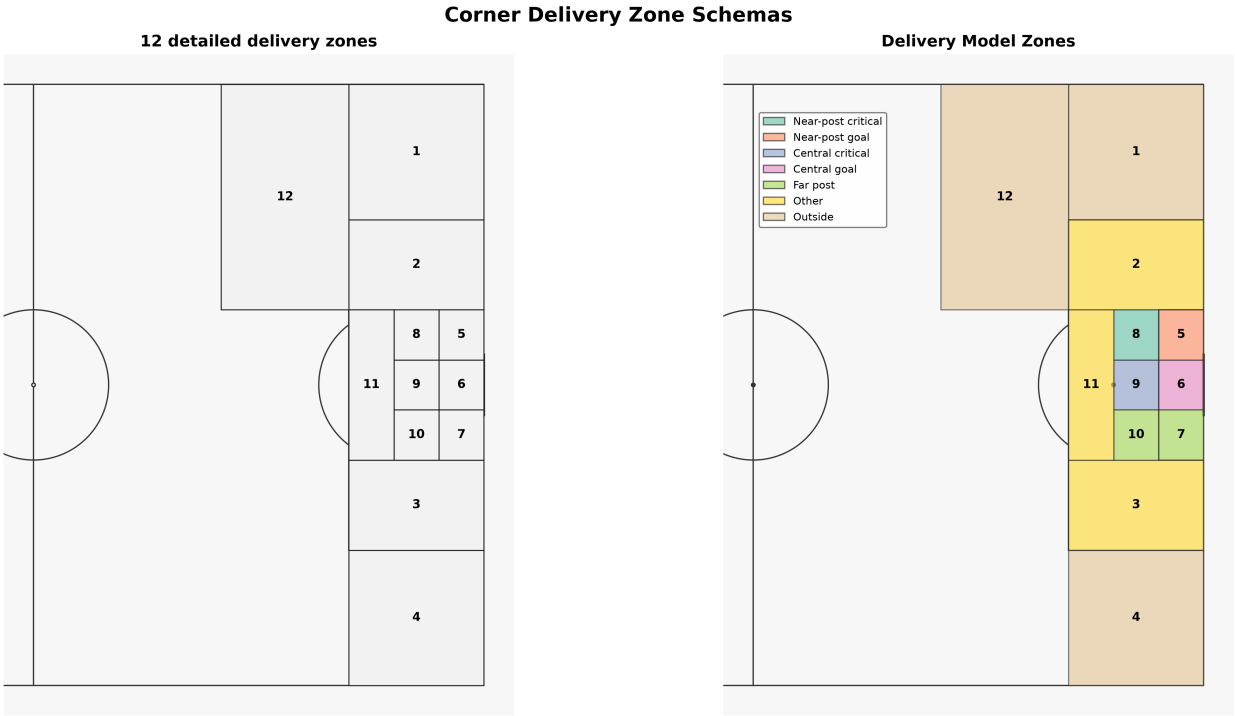


Figure 6: Twelve detailed corner-delivery zones and their aggregation into the seven delivery-model groups. Zones are defined relative to the corner side; a left-sided corner is shown, with right-sided corners using mirrored geometry.

Congestion describes how densely occupied a location is, while the sign of overload identifies which team has greater local presence. A positive value favors the attacking density and a negative value favors the defending density; neither quantity alone represents possession or causal control of the area.

Figure 7 shows two deliberately contrasting corners at observed first contact. VfL Wolfsburg wins the first contact against RB Leipzig after a far-post delivery with endpoint congestion 8.74 and overload 0.83. FC Köln wins the other contact against Hamburger SV with similarly high congestion, 7.73, but a defense-favoring overload of -4.70. The examples illustrate why congestion and team balance describe different properties of the realized contest.

These scalar summaries are interpretable projections of the density representation, not replacements for it. DFFT retains the joint attacking and defending state in every bin, allowing similar overloads to have different implications at different congestion levels and pitch locations. Because the examples use contemporaneous first-contact configurations and were selected for contrast, they illustrate interpretable spatial features rather than validate a predictive or causal mechanism.

5.6 Stage 1: Delivery Prediction

Stage 1 predicts one of the seven delivery groups from the corner-start configuration using the common DFFT construction developed above. For delivery groups k and h, it forms weighted spatial evidence and couples all pairwise comparisons into one probability vector:

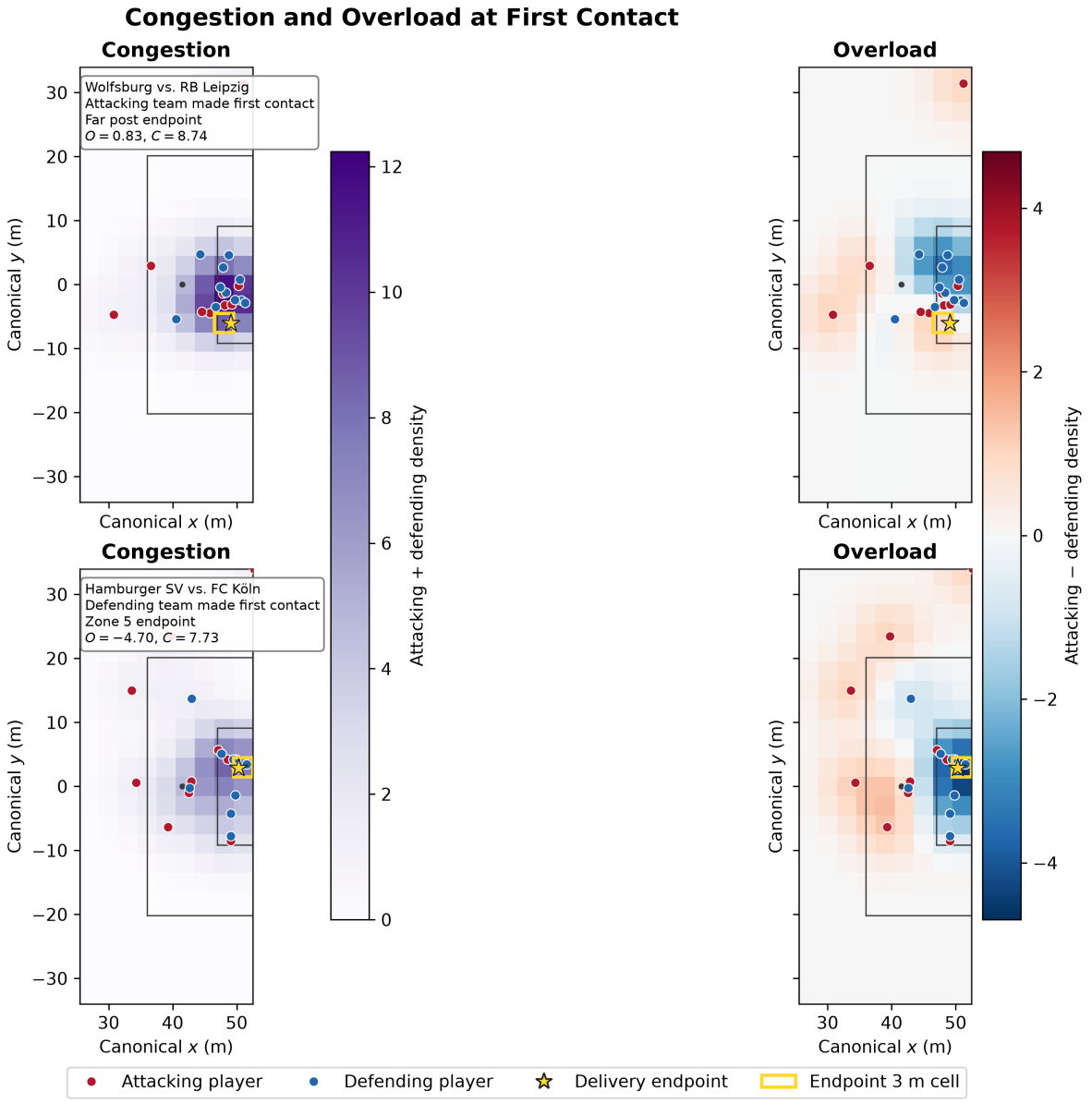


Figure 7: **Congestion and overload at observed first contact.** Rows show two selected corner kicks. Player markers show the observed configuration; density panels show total attacking-plus-defending density and attacking-minus-defending density. The gold star marks the delivery endpoint and the highlighted square its spatial bin. These examples were selected for interpretation and are not pre-kick forecasts or evidence that changing overload would cause a different outcome.

$$r_{ikh}(\rho_{0i}) = c_{kh} + \sum_{b \in \mathcal{B}_{kh}} w_{kh,b} \log \frac{f_{k,b}(s_{ib})}{f_{h,b}(s_{ib})}, \qquad w_{kh,b} \geq 0,$$
$$\widehat{\mathbf{p}}_i = \mathcal{C}\Big(\big\{\text{logit}^{-1}(r_{ikh})\big\}_{k<h}\Big), \qquad \widehat{p}_{ik} = \widehat{P}(Z_i = k \mid \rho_{0i}). \tag{16}$$

Here, s_{ib} is the joint attacking–defending density state in bin b, $f_{k,b}$ is its learned distribution for group k, and $\mathcal{C}$ is the pairwise coupling procedure described in Section 3. In practical terms, bins whose observed states are more typical of one destination contribute evidence toward that destination.

On 368 corners from held-out matches, DFFT gives log loss 1.762, Brier score 0.793, macro AUC 0.656, and support-weighted AUC 0.649. The strongest evaluated conventional specification, start-density PCA, gives 1.799, 0.811, 0.644, and 0.638, respectively. These are modest numerical advantages among the evaluated specifications, not an equal-tuning comparison of model families. Temperature scaling estimated from fold-excluded training predictions changes little; its probability vector is used by Stage 2 so that delivery uncertainty is propagated rather than replaced by one selected destination. Comparator definitions and complete results are reported in Appendix D.2.

5.7 Observed-Contact Reference Model

Before estimating the transition, we ask whether the realized first-contact configuration adds outcome information beyond knowing where the ball was delivered. For each of the six supported inside groups, DFFT predicts whether the corner-taking team makes first contact from the observed attacking and defending densities at contact. Because this model uses both the realized destination and contact configuration, it is an information-rich diagnostic rather than a pre-kick forecast.

For realized delivery group z, the model is

$$\text{logit}\, P_{\text{DFFT}}(Y_i = 1 \mid \rho_{1i}, Z_i = z) = \alpha_z + \sum_{b \in \mathcal{B}_z} w_{z,b} \log \frac{f^{(1)}_{z,b}(s_{ib})}{f^{(0)}_{z,b}(s_{ib})}, \qquad w_{z,b} \geq 0, \tag{17}$$

where s_{ib} is the observed attacking–defending density state in bin b, and the density ratio measures whether that state is more typical of the corner-taking team or opposition making first contact.

On 124 match-disjoint held-out contacts, a comparator using only the realized delivery group's training frequency already achieves AUC 0.811, Brier score 0.188, and log loss 0.560. Adding the observed contact configuration with DFFT gives point estimates of 0.831, 0.175, and 0.522, respectively. The incremental differences—0.020 in AUC, -0.013 in Brier score, and -0.038 in log loss—suggest additional spatial information within delivery groups, while showing that much of the standalone discrimination is attributable to the realized destination. Full specifications and diagnostics are provided in Appendix D.4.

These are observed held-out advantages, not proof that the contact configuration was predictable before the kick. The contacting player is also present in the observed frame, so the model describes the realized spatial contest rather than a causal mechanism or formal performance ceiling. Stage 2 therefore asks how much of this later configuration can be reconstructed using only information available at the

corner start.

5.8 Stage 2: Start-to-First-Contact Transition

The observed-contact model indicates that the realized first-contact configuration contains information about which team makes first contact, but that configuration is unavailable before the kick. Stage 2 therefore tests how well the later configuration can be reconstructed from the start. This stage is not DFFT: with only 478 paired training corners, we use PCA to compress the start and contact density maps and ridge regression to predict the later representation for each possible delivery group. Complete specifications are provided in Appendix D.5.

For every corner, the model produces a candidate first-contact map $\widehat{\rho}_{1i}(z)$ for each delivery group and combines those maps using the calibrated Stage 1 delivery probabilities $\widehat{p}_{iz}$:

$$\widehat{\rho}_{1i} = \sum_{z} \widehat{p}_{iz}\widehat{\rho}_{1i}(z). \tag{18}$$

This weighted map represents the expected first-contact configuration while retaining uncertainty about the destination and is used to evaluate Stage 2 reconstruction. For Stage 3, the destination-specific maps are retained separately so that each is evaluated by its corresponding contact model before probabilities are averaged. On 106 match-disjoint held-out corners, treating the start map as the contact map gives a complete-map error of 77.55. The deployable additive and interaction transitions reduce this to 61.09 and 61.59, respectively—an improvement of about 21%. The interaction model is retained for Stage 3 because it performed slightly better in the training-fold comparison, although the additive model was marginally better on the held-out set.

Stage 2 therefore shows that some spatial redistribution is predictable from the corner start. It does not show that the reconstructed maps improve the final outcome prediction: they are smooth conditional means rather than realized runs, blocks, or marking decisions. Stage 3 tests whether the reduction in map error produces a corresponding gain in first-contact prediction.

A natural extension would treat the subsequent spatial configuration as the DFFT outcome and learn a conditional transition distribution $P_{\mathrm{DFFT}}(\rho_1 \mid \rho_0, Z)$. Rather than producing one mean map, a dynamic DFFT could generate multiple plausible spatial consequences of attacking runs and defensive responses from the starting configuration. Propagating these configurations through the contact model would preserve uncertainty about how play may evolve and could extend beyond corners to other sequential football actions.

5.9 Stage 3: End-to-End First-Contact Prediction

Stage 3 composes the three fitted components. For each supported inside delivery group $z \in \mathcal{I}$, its DFFT contact model evaluates the corresponding Stage 2 map and returns $\widehat{q}_{iz}$. The pre-kick, conditional-inside prediction is

$$\widehat{P}(Y_i = 1 \mid \rho_{0i}, Z_i \in \mathcal{I}) = \frac{\sum_{z\in\mathcal{I}} \widehat{p}_{iz}\widehat{q}_{iz}}{\sum_{z\in\mathcal{I}} \widehat{p}_{iz}}. \tag{19}$$

The numerator combines each destination's predicted contact probability with its Stage 1 delivery probability, while the denominator conditions the result on delivery to a supported inside group. Although this probability can be calculated before the kick, the evaluation includes only corners whose realized delivery was inside; it is therefore not an unconditional forecast over all corners. Full definitions are provided in Appendix D.6.

Stage 3 evaluates each delivery-group-specific map separately, but each map is still one conditional-mean estimate of many possible contact configurations. If g_z denotes the nonlinear DFFT contact model for group z, then in general $g_z(\mathbb{E}[\rho_1 \mid \rho_0, z]) \neq \mathbb{E}[g_z(\rho_1) \mid \rho_0, z]$. Better reconstruction of the mean map therefore need not produce better outcome probabilities. This is a candidate explanation for the end-to-end result, not a demonstrated cause; destination uncertainty, transition error, and the small evaluation sample are not separately identified.

The information available to each model and its performance on the 83 inside deliveries from 63 held-out matches are summarized in Table 2. The interaction-minus-no-transition log-loss difference is -0.001 (95% match-clustered interval: -0.018 to 0.016), while its difference from prevalence is -0.014 (-0.044 to 0.018).

Model	Information available	Log loss	AUC
Prevalence	Training frequency; pre-kick baseline	0.687	0.500
No-transition composition	Start map and predicted delivery; pre-kick	0.674	0.604
Sequential interaction chain	Predicted delivery and contact maps; pre-kick	0.673	0.598
Observed-group frequency	Realized delivery; diagnostic	0.579	0.788
Observed-contact DFFT	Realized delivery and contact map; diagnostic	0.535	0.804

Table 2: First-contact performance on the common 83-corner conditional-inside cohort, ordered by information availability.

The transition models improved reconstruction of first-contact density maps relative to persistence, but this improvement did not translate into an established gain in first-contact prediction. On the common inside-delivery cohort, the sequential chain and no-transition composition have similar point estimates, and the interval spans no difference. This does not establish that the models are equivalent or rule out useful transition signal in a larger sample. It shows that the current hybrid composition predicts part of the spatial redistribution but does not yet demonstrate an end-to-end advantage.

Conclusion

We have shown that a common DFFT construction can extract predictive information from player positioning across passing and corner-kick tasks that would ordinarily use different tactical features. The strongest evidence comes from passing direction and corner delivery. For pass completion, DFFT provides a complementary positional baseline: despite excluding pass direction, distance, and body part, it achieves comparable AUC to xPass in several directional groups and shows different calibration at high completion probabilities. First-touch prediction from the kick-time configuration is weaker, indicating that not every outcome is equally recoverable from the initial arrangement.

The exploratory sequential corner analysis tests a further possibility: whether explicitly modeling the delivery and subsequent spatial transition improves first-contact prediction. The transition models reduce contact-map reconstruction error by approximately 21% relative to persistence, showing that some spatial redistribution is predictable. This improvement does not translate into an established advantage for the end-to-end forecast over the no-transition composition. The extension therefore identifies both a promising direction and a current limitation; the paper's central evidence for DFFT does not depend on the sequential model outperforming the no-transition composition.

This represents the first application of DFFT to football, and several limitations motivate further work. Requiring complete player positions substantially reduces the available data; adapting the method to partial visibility could increase sample sizes and reduce reliance on Gaussian weighting across nearby ball locations and mirroring symmetric situations. The results also demonstrate useful prediction from limited local outcome samples, but they do not establish superior sample efficiency without a matched comparison. Future work could develop team- or season-specific models and a dynamic DFFT that replaces the conditional-mean corner transition with a distribution over plausible future configurations. Across the evaluated tasks, DFFT provides a promising common framework for incorporating player positioning into football analysis while retaining explicit uncertainty about what the observed configurations can predict.

References

[1] J Felipe Méndez-Valderrama, Yunus A Kinkhabwala, Jeffrey Silver, Itai Cohen, and TA Arias. Density-functional fluctuation theory of crowds. *Nature communications*, 9(1):3538, 2018.

[2] Boris Barron, Nathan Sitaraman, and Tomás Arias. Analyzing nba player positions and interactions with density-functional fluctuation theory. *Scientific Reports*, 15(1):19830, 2025.

[3] Hudl StatsBomb. Hudl StatsBomb release new models: Expected pass (xPass) and pass clustering. Hudl blog, October 2023. URL https://www.hudl.com/blog/statsbomb-models-expected-pass-xpass-and-pass-clustering. Published 5 October 2023; accessed 8 November 2026.

Appendix

Passing-direction prediction

A.1 Analysis and Samples

The analysis considers eight directions for passes originating within 3 m of $(x, y) = (6, 14)$ m: front (F), front-left (FL), left (L), back-left (BL), back (B), back-right (BR), right (R), and front-right (FR). Frames with a negative ball y-coordinate are reflected across the x-axis, including the players and pass endpoint, to combine corresponding situations on the two sides of the pitch. Unless stated otherwise, results use the Gaussian-weighted eight-direction model presented in the main text.

The eight directions are equal 45-degree sectors centered on F, FL, L, BL, B, BR, R, and FR. Angles are measured counterclockwise from the positive x-axis after reflection, with F containing angles from $-22.5°$ up to but not including $22.5°$. This analysis uses 360 data from the English Premier League, German Bundesliga, and Italian Serie A, together with the broadcast tracking data.

The match-level split contains 529 training matches and 528 testing matches. Within the local ball-origin region, these contribute 1,119 training passes from 372 matches and 1,008 testing passes from 350 matches. The training sample contains 1,039 passes represented by 360 freeze frames and 80 by broadcast tracking; the corresponding testing counts are 863 and 145. When both sources describe the same event, only the tracking version is retained. All frames contain coordinates for the 20 outfield players, although broadcast tracking can include estimated off-camera positions.

Gaussian-weighted density fitting includes training passes within 10 m of the reference ball location, using a Gaussian standard deviation of 3 m. This gives 10,615 passes and an effective sample size of approximately 4,333, calculated as $(\sum_i w_i)^2 / \sum_i w_i^2$, where w_i is the frame weight. Spatial prediction weights are fitted using the 1,119 local training passes, and testing uses local held-out passes with equal weight. The Gaussian weighting of ball locations is distinct from the 4 m Gaussian used to represent each player.

All comparisons use the same match split, player-density grid, training-derived density cutoff, and 15 local correction functions. Model selection uses ten validation blocks formed from whole training matches. Confidence intervals use 10,000 whole-match bootstrap resamples of the testing set, keeping the fitted models and probability groups fixed, as described in Section **??**.

A.2 Additional Held-out Examples

Figure A.1 illustrates three different prediction patterns: a preference for FL, a preference for BR, and similar probabilities for F and FR. The frames come from distinct testing matches and were selected to show varied predictions and clear spatial arrangements. They illustrate individual predictions; performance is assessed using the full testing sample.

A.3 Discrimination and Calibration Across Eight Directions

Figure A.2 shows calibration separately for each direction, using all 1,008 testing passes. Figures A.3 and A.4 show calibration and AUC for all 28 direction pairs. Each pairwise comparison includes only

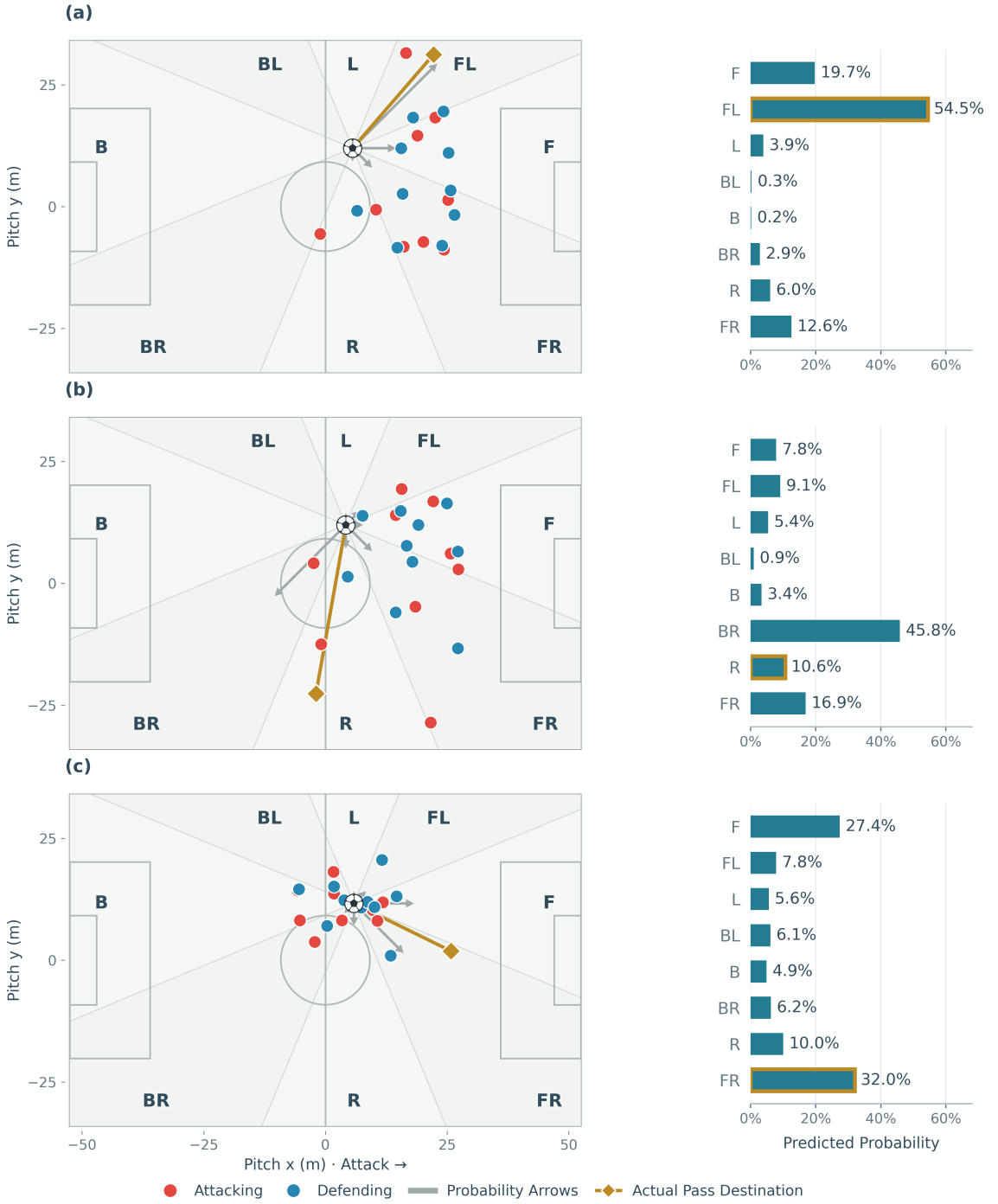


Figure A.1: Passing-direction predictions for additional held-out arrangements. Attacking players are red and defending players blue; the ball marks the pass origin. Grey arrow lengths are proportional to predicted probabilities. The gold diamond marks the recorded pass destination, and the corresponding probability bar is outlined in gold. **(a)** FL receives 54.5% probability and is the recorded direction. **(b)** BR receives 45.8%, although the recorded pass is R, assigned 10.6%. **(c)** F and FR receive 27.4% and 32.0%, respectively; the recorded pass is FR.

passes recorded in one of its two directions, k or ℓ, and uses

$$P(k \mid k \text{ or } \ell) = \frac{P(k)}{P(k) + P(\ell)}. \tag{20}$$

These probabilities are obtained from the eight-direction predictions without additional fitting. Averaging equally over the 28 pairs gives a mean AUC of 0.812 (95% confidence interval: 0.791–0.831), with individual comparisons ranging from 0.645 for BL versus B to 0.888 for FL versus BL.

Calibration points compare mean predicted probabilities with observed frequencies. Six initial probability groups are used for individual directions and direction pairs, preserving tied predictions and merging adjacent groups where needed to retain at least 30 passes from five matches per point. Axis limits follow the range of the observations and their confidence intervals, with the same scale on both axes within each panel.

A.4 Comparison With a Dedicated Two-direction Model

Both models are evaluated on the same 389 testing passes from 206 matches: 217 FL and 172 R. The dedicated model uses 413 local training passes, with nearby FL and R passes contributing to density fitting through Gaussian weighting. Conditioning the eight-direction predictions on FL or R gives an AUC of 0.850, compared with 0.854 for the dedicated model (Figure A.4, final two panels). The eight-direction formulation therefore retains similar discrimination for this comparison while providing probabilities for all eight directions.

A.5 Including Nearby Passes Through Gaussian Weighting

Gaussian-weighted fits are compared with fits using equally weighted passes inside the local 3 m region. Including nearby passes increases the density-fitting sample from 413 to 4,084 for FL–R and from 1,119 to 10,615 for eight directions, with effective sample sizes of approximately 1,691 and 4,333. Spatial prediction weights use the same local training passes, and each comparison uses the same testing passes.

Gaussian-weighted fitting gives higher testing AUCs for both tasks: 0.854 versus 0.814 for FL–R, and a mean pairwise AUC of 0.812 versus 0.796 for eight directions (Figure A.5). This comparison assesses the combined effect of including nearby passes and weighting them by distance.

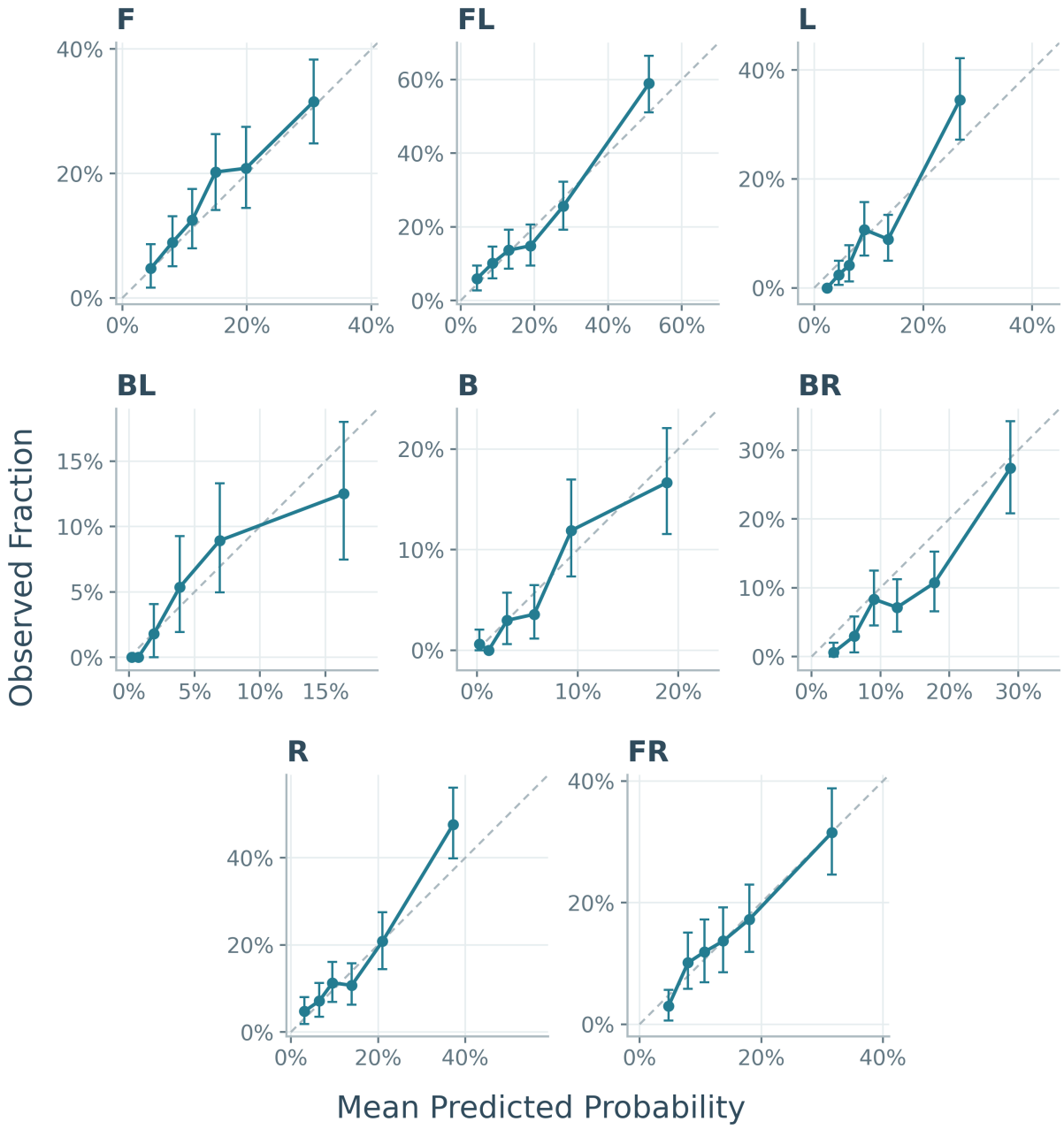


Figure A.2: Calibration of individual passing directions. Each panel compares the predicted probability of its named direction with its observed frequency among all 1,008 testing passes. Dashed lines denote perfect calibration; error bars are pointwise 95% confidence intervals from resampling whole testing matches. Axis ranges vary between panels.

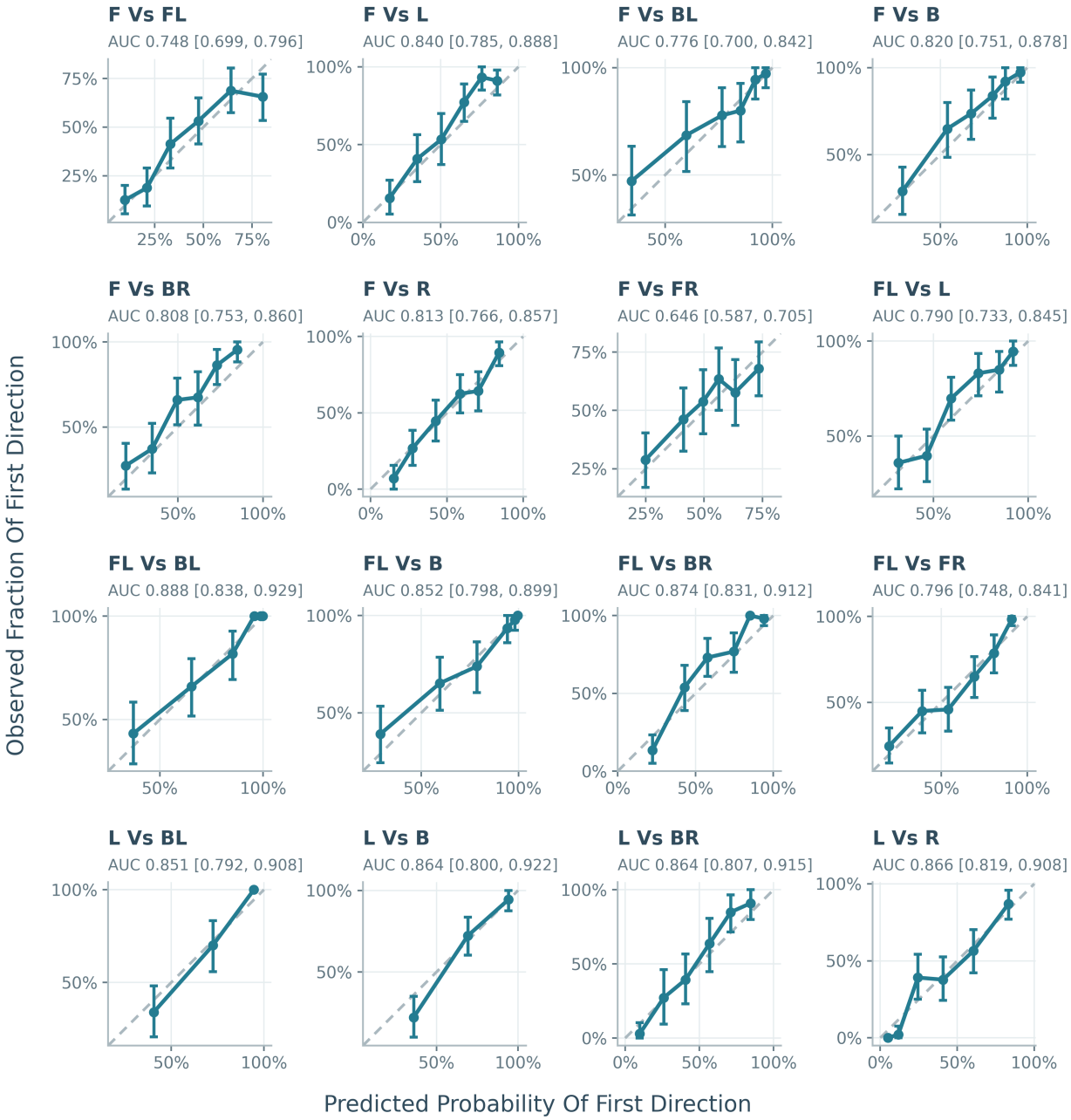


Figure A.3: Pairwise passing-direction calibration and discrimination, part 1 of 2. Each panel uses testing passes recorded in either named direction and compares the conditional probability of the first direction with its observed frequency. Predictions come from the eight-direction model. Panel annotations give AUCs and 95% confidence intervals; error bars show pointwise 95% intervals for observed frequencies. Intervals are obtained by resampling whole testing matches. Dashed lines denote perfect calibration.

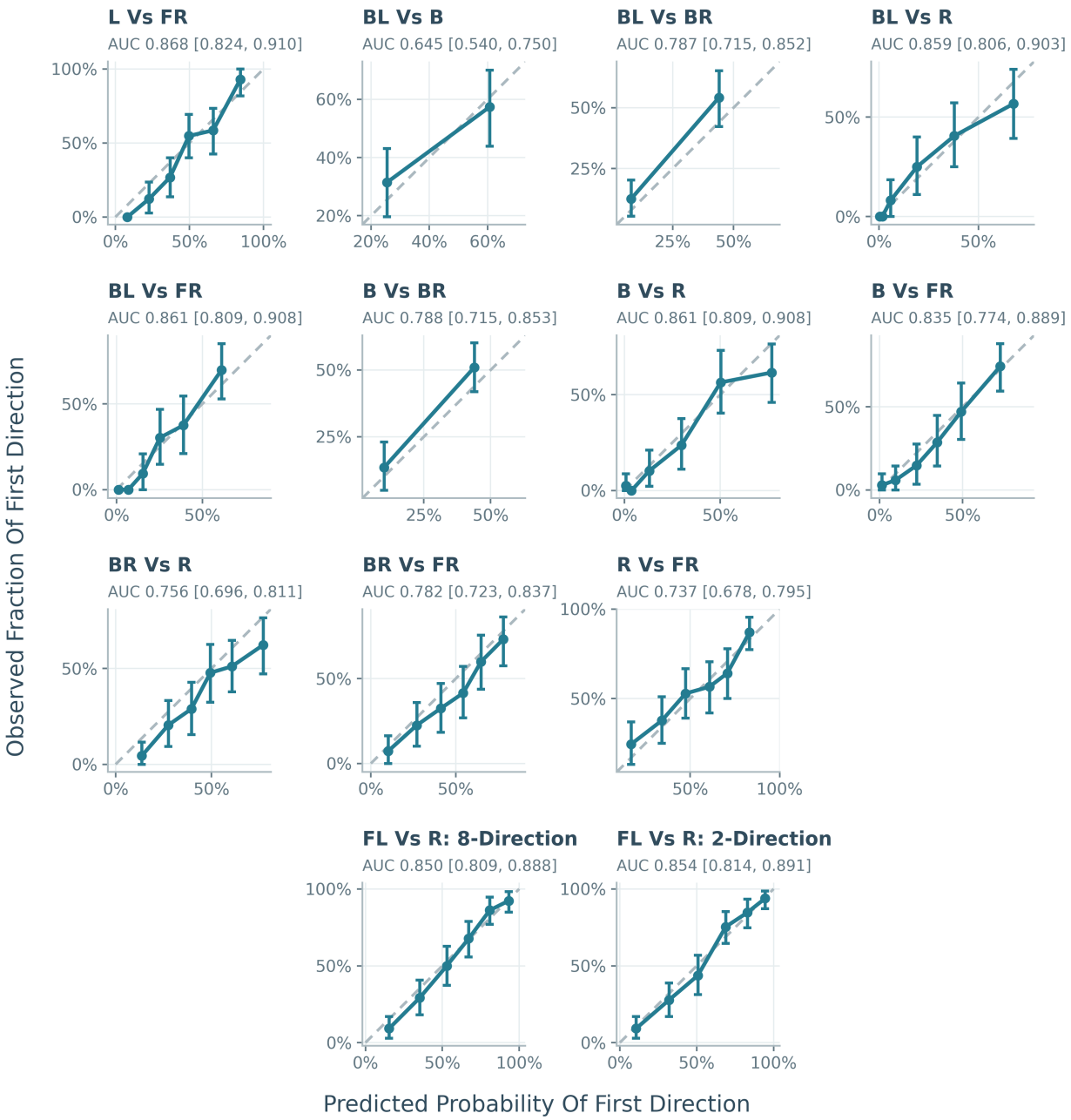


Figure A.4: Pairwise passing-direction calibration and discrimination, part 2 of 2. The remaining twelve direction pairs, with an additional panel for the dedicated two-direction FL–R model. The final two panels compare conditional eight-direction predictions with dedicated FL–R predictions on the same testing passes and axis scales. Other definitions follow Figure A.3.

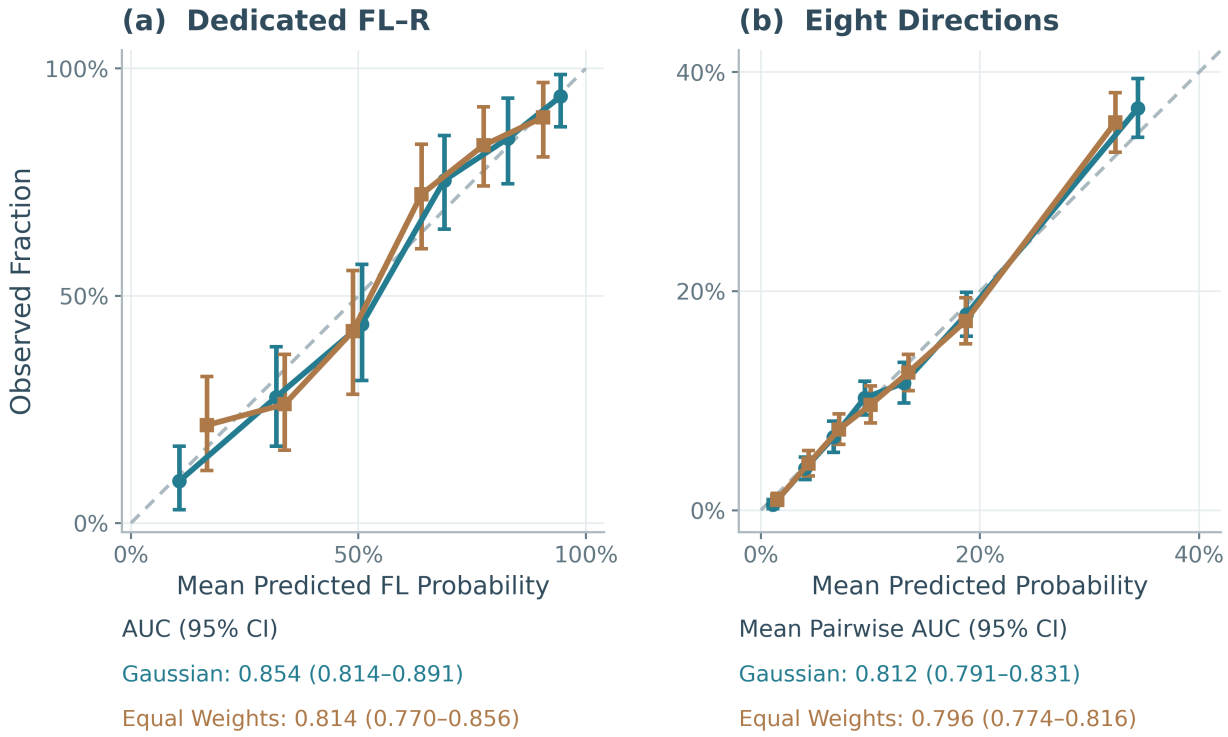


Figure A.5: Comparison of Gaussian-weighted and local equal-weight training. (a) Dedicated FL–R models evaluated on the same 389 testing passes. **(b)** Eight-direction models evaluated using all eight probabilities from each of the same 1,008 testing passes. Teal circles show Gaussian-weighted fits; brown squares show local equal-weight fits. Probability groups are formed separately for each model, using six initial groups in (a) and seven in (b). Dashed lines denote perfect calibration; error bars are pointwise 95% confidence intervals from resampling whole testing matches. AUC summaries and their 95% intervals are shown below each panel.

Pass-Completion Prediction

B.1 Ball Locations and Regional Comparisons

Figure B.6 shows the ball-origin locations used for pass-completion prediction. Locations on opposite sides of the pitch share a model after reflecting the player and ball positions. These model locations are distinct from the spatial bins used to represent player densities. For regional comparisons, the evaluated area is divided into three equal-width longitudinal bands, with boundaries at approximately $x = -9.58\,\text{m}$ and $x = 9.58\,\text{m}$.

The model locations lie on a triangular grid with 2.5 m nearest-neighbor spacing, retaining centers within $|x| \leq 27.5\,\text{m}$ and $0 \leq y \leq 24\,\text{m}$ after reflection. Each location owns its full hexagonal cell. A testing pass uses the model for the cell containing its ball location and contributes once; passes outside these cells are excluded from evaluation. The 3 m neighborhood described in the Methods is used for spatial fitting and validation.

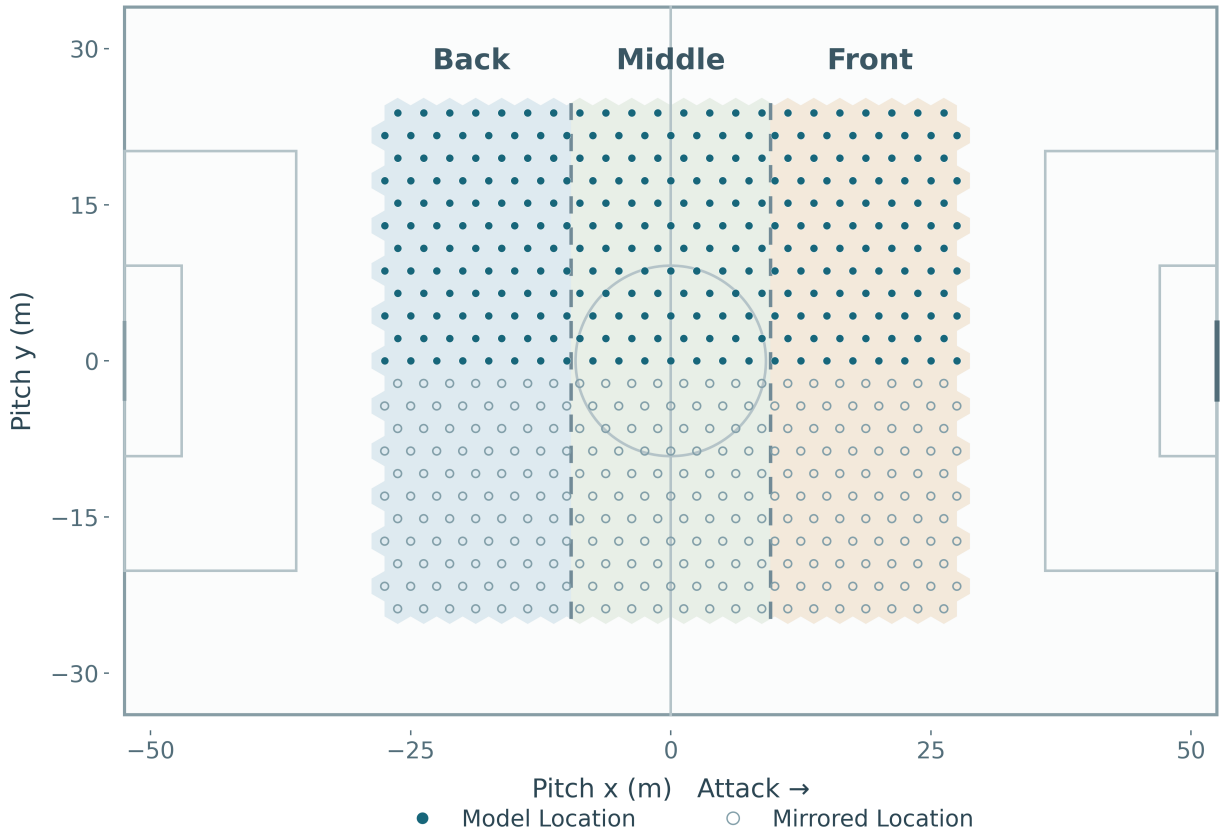


Figure B.6: Ball-origin locations and regional divisions. The grid contains 270 local model locations with nearest-neighbour spacing of 2.5 m. Filled points show model locations and hollow points their reflections across the longitudinal centre line ($y = 0$); reflected locations use the same model. Shading marks the evaluated area, divided into back, middle, and front regions. The positive x-direction points toward the attacking goal.

All performance comparisons use the same 22,481 testing passes from 364 held-out matches. The completion target distinguishes completed passes from passes recorded as blocked, intercepted, or recovered. Eligible passes are made by outfield players and have finite, in-bounds recorded positions for

all 20 outfield players. Open-play passes are identified by a missing pass-type label or a label of Interception or Recovery. Passes are excluded when a recorded goalkeeper lies in the same 45-degree directional sector as the pass, unless that goalkeeper is more than 25 m from the ball and another attacker in that sector is closer to the ball and at least 10 m from the goalkeeper. Missing goalkeeper positions do not by themselves exclude a pass. Each pass contributes once, at its assigned grid location. Figures B.7–B.9 repeat the main-text analysis for the back, middle, and front regions, containing 6,450, 11,380, and 4,651 testing passes, respectively. The models are unchanged; only the observations used for each regional summary differ. Passing direction is used to group the AUC comparisons and is not supplied to DFFT.

Calibration uses five-percentage-point groups between 60% and 100%, displaying groups containing at least 30 passes. This range contains 95.6% of DFFT predictions and 88.2% of xPass predictions in the overall testing sample. AUC uses all testing predictions within each direction. Performance intervals are pointwise 95% intervals from 10,000 whole-match bootstrap samples, using the same resampled matches for both models and keeping the fitted predictions fixed.

B.2 Player Selection and Completion Relative to Expectations

Player profiles combine out-of-fold validation predictions with held-out testing predictions. The full prediction sample contains 87,246 validation passes and 22,481 testing passes. For each player, let D and X denote mean DFFT and xPass completion probabilities, and let O denote the observed completion rate. The profile coordinates are $100(D - X)$ and $100(O - D)$. Their sum is $100(O - X)$, so the diagonal line indicates observed completion equal to the xPass expectation for the attempted passes.

The displayed players comprise the 50 most-observed players and the union of the five highest and five lowest values of each profile coordinate among players with at least 150 observations. Removing duplicates gives 63 players. The same players are retained in all regional figures, although their regional sample sizes differ.

Figure B.10 orders these players by observed completion minus mean DFFT and also shows completion relative to xPass. Intervals are obtained from 10,000 whole-match bootstrap samples of the combined validation and testing observations, with predictions held fixed. They describe uncertainty for each displayed profile and do not account for the selection of extreme profiles.

B.3 High-Confidence xPass Predictions and Failure Definitions

The main comparison restricts failed passes to blocked, intercepted, and recovered outcomes. To examine whether this restriction explains the calibration difference for xPass, Table 3 compares xPass predictions of at least 90% under this definition and when all known failure outcomes are included. Mean predicted completion is approximately 96.95% in both cases, while observed completion is 97.81% and 97.50%, respectively. Including the additional failures reduces the discrepancy, but some underprediction remains.

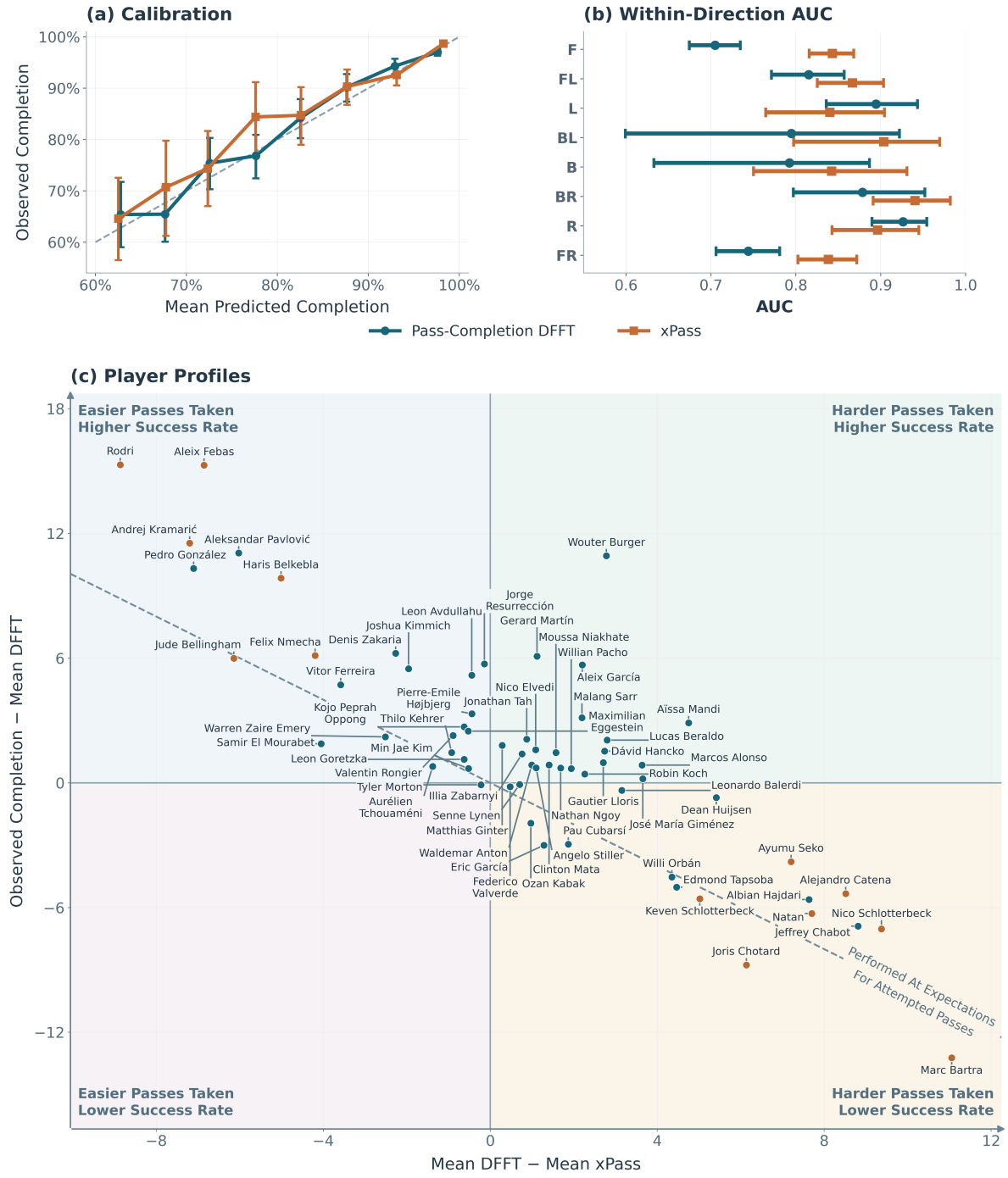


Figure B.7: Pass completion in the back region. (a,b) Calibration and within-direction AUC for 6,450 testing passes originating in the back region of Figure B.6. **(c)** Validation and testing profiles for the same 63 players shown in Figure 4, restricted to this region. Profile differences are in percentage points. Colours, axes, reference lines, and uncertainty estimates follow the main-text figure.

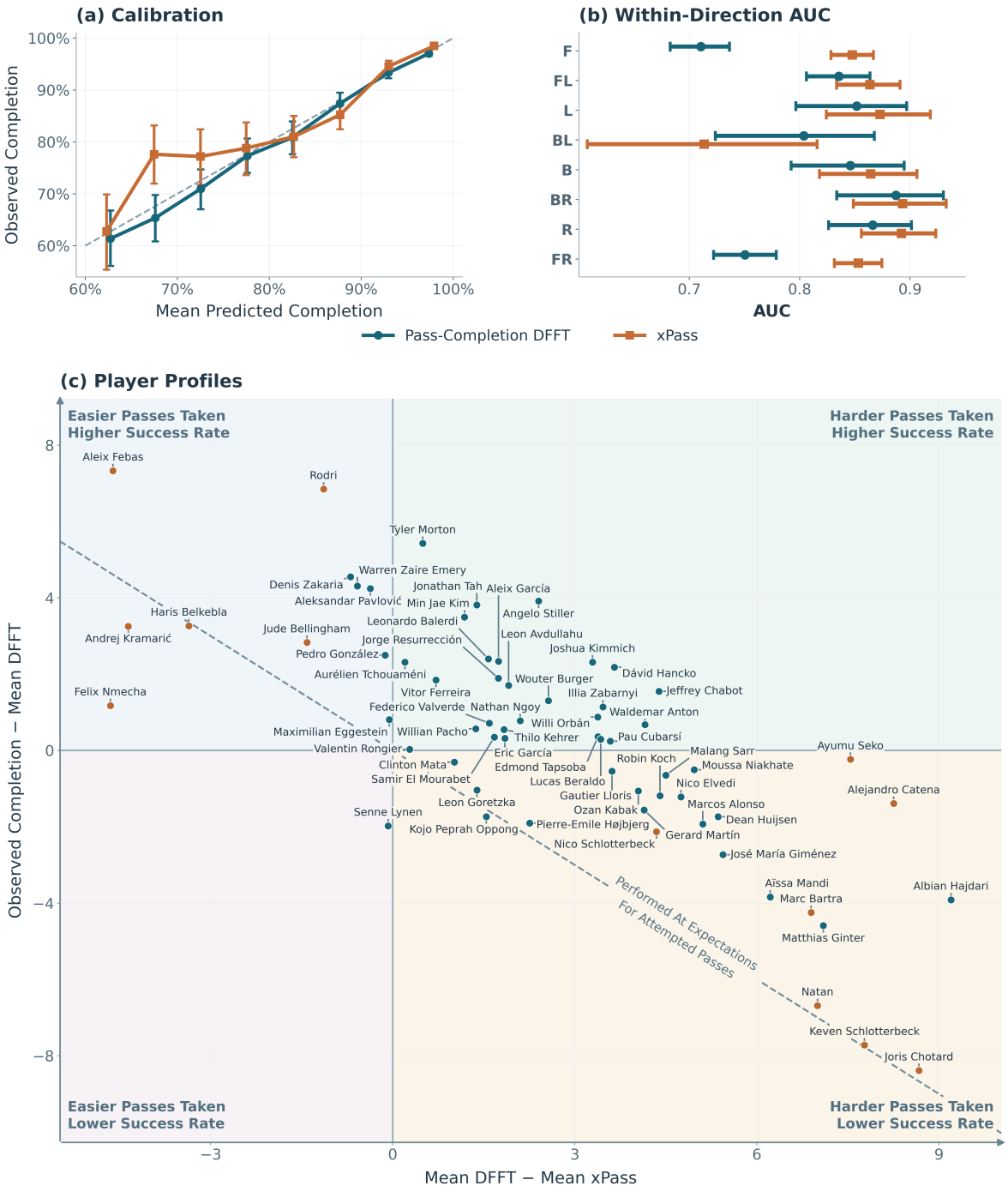


Figure B.8: Pass completion in the middle region. The analysis in Figure B.7 is repeated for the middle region of Figure B.6. Panels (a,b) use 11,380 testing passes; panel (c) combines validation and testing observations for the same 63 players. Profile differences are in percentage points.



Figure B.9: Pass completion in the front region. The analysis in Figure B.7 is repeated for the front region of Figure B.6. Panels (a,b) use 4,651 testing passes; panel (c) combines validation and testing observations for the same 63 players. Profile differences are in percentage points. The player-selection threshold applies to overall observations, so some regional profiles contain substantially fewer passes.

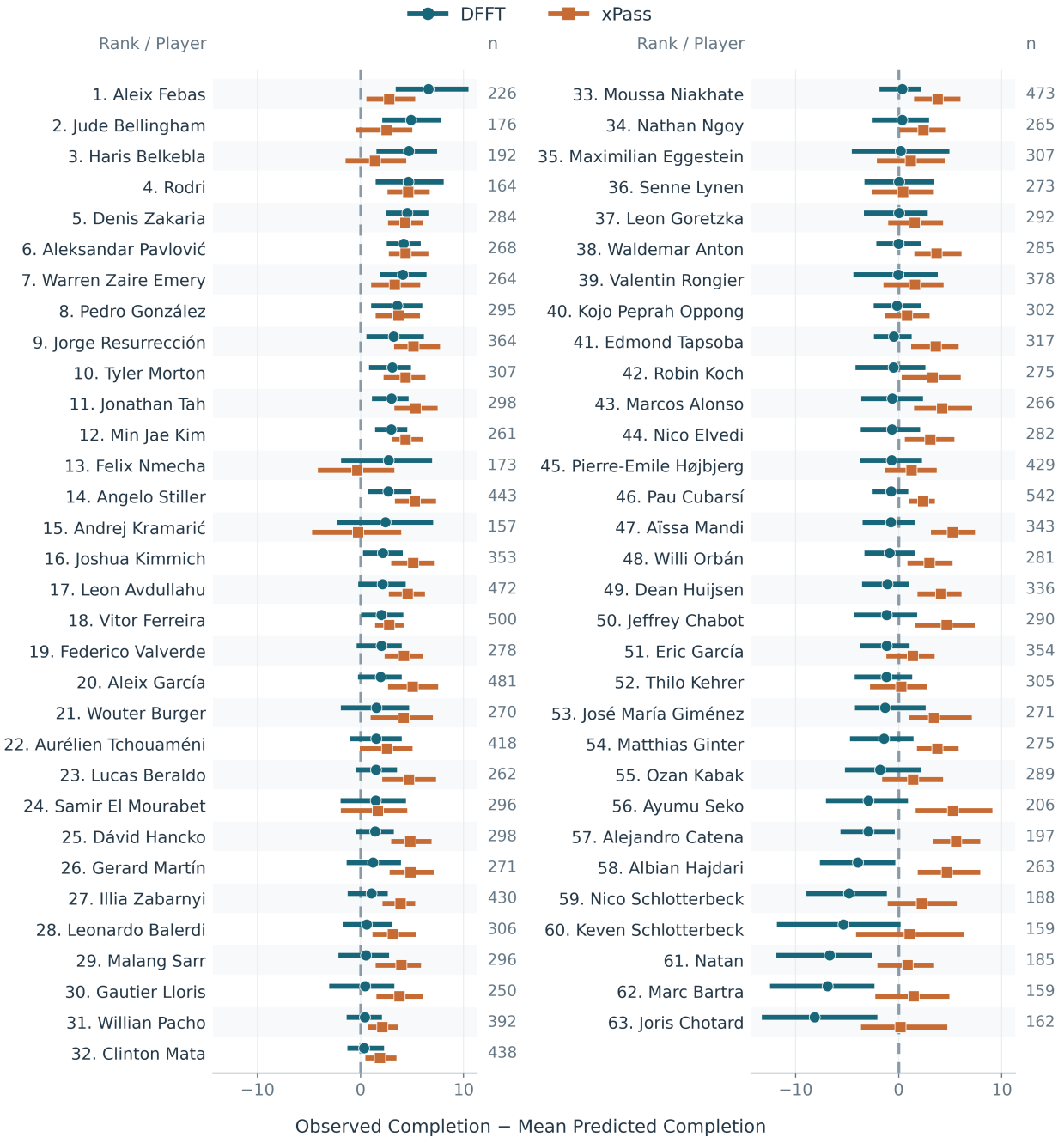


Figure B.10: Observed completion relative to DFFT and xPass. The 63 selected players are ordered by observed completion minus mean DFFT, from top to bottom in the left column and then the right. Points show observed minus mean predicted completion in percentage points; bars are pointwise 95% whole-match bootstrap intervals. The column n gives the number of validation and testing passes contributing to each player's profile.

Failure definition	Passes	Mean xPass	Observed	Difference [95% CI]
Three retained types	15,781	96.95%	97.81%	$0.85\ [0.62, 1.08]$
All known failures	15,830	96.95%	97.50%	$0.56\ [0.31, 0.80]$

Table 3: **High-confidence xPass calibration.** Testing passes with xPass at least 90%, from 361 matches. Differences are observed minus mean predicted completion, in percentage points; intervals are based on 10,000 whole-match bootstrap samples.

Corner-kick prediction

C.1 Delivery regions and evaluation

Corners are aligned so that the attacking goal is on the right and the corner taker is above it, reflecting all player and ball coordinates across the pitch's longitudinal axis where necessary. The analysis retains deliveries into the four regions in Table 4, with known first-touch outcomes and the complete player observations described in the Data section. Internal boundaries were obtained by rounding training-derived cuts to whole meters, giving roughly balanced delivery counts. Each corner belongs to one region.

Table 4: **Corner delivery regions and sample sizes.** Coordinates are in meters after corner-side alignment. Shared boundaries belong to the region above or to the right. Each split contains 465 distinct matches.

Region	x range	y range	Training	Testing
1	36.75–52.5	5–18.7	237	266
2	36.75–48	−1–5	254	216
3	48–52.5	−1–5	226	216
4	36.75–52.5	−18.7–−1	265	287
Total			982	985

Figure C.11a–f examines all six region pairs. Each comparison uses only testing corners delivered to either region, with their predicted probabilities normalized to sum to one. Intervals for this four-region analysis use 10,000 bootstrap resamples of the 465 testing matches, drawn with replacement. Predictions and calibration-group boundaries remain fixed; intervals are the 2.5th and 97.5th percentiles.

C.2 Which team makes first touch?

The joint model also gives the probability that either team makes first touch. Let R denote delivery region, T the first-touch team, and ρ_0 the player densities at the kick. Before the delivery region is known, the defending-team probability is obtained by summing over regions. If region r is known, the prediction can instead be conditioned on it:

$$\begin{aligned} P(T = \mathrm{D} \mid \rho_0) &= \sum_r P(R = r, T = \mathrm{D} \mid \rho_0), \\ P(T = \mathrm{D} \mid R = r, \rho_0) &= \frac{P(R = r, T = \mathrm{D} \mid \rho_0)}{\sum_{t \in \{\mathrm{A},\mathrm{D}\}} P(R = r, T = t \mid \rho_0)}. \end{aligned} \tag{21}$$

Here, A and D denote the attacking and defending teams. Both predictions use the player configuration at the kick; the conditional evaluation additionally supplies the observed delivery region. The attacking-team probability is one minus the defending-team probability.

Conditioning on the observed region gives a first-touch AUC of 0.674 (95% CI: 0.639–0.708). However, regional first-touch frequencies estimated from training data alone already give 0.662 (0.627–0.695). The paired AUC difference is 0.012 (-0.007–0.032), so the aggregate comparison does not establish a statistically significant improvement over this baseline. Much of the discrimination in this comparison is captured by differences in first-touch frequencies between delivery regions, and delivery appears to be substantially easier to predict.

Figure C.11g–j examines predictions within each delivery region. The clearest discrimination is in region 4, with AUC 0.608 (0.542–0.672); region 1 gives 0.565 (0.491–0.639). Regions 2 and 3 show little discrimination from initial positioning, although their first-touch frequencies differ substantially: defenders make first touch in 58.3% and 84.3% of their testing corners, respectively. Across the four regions, the mean AUC is 0.539 (0.497–0.580), suggesting a presence of weak signal.

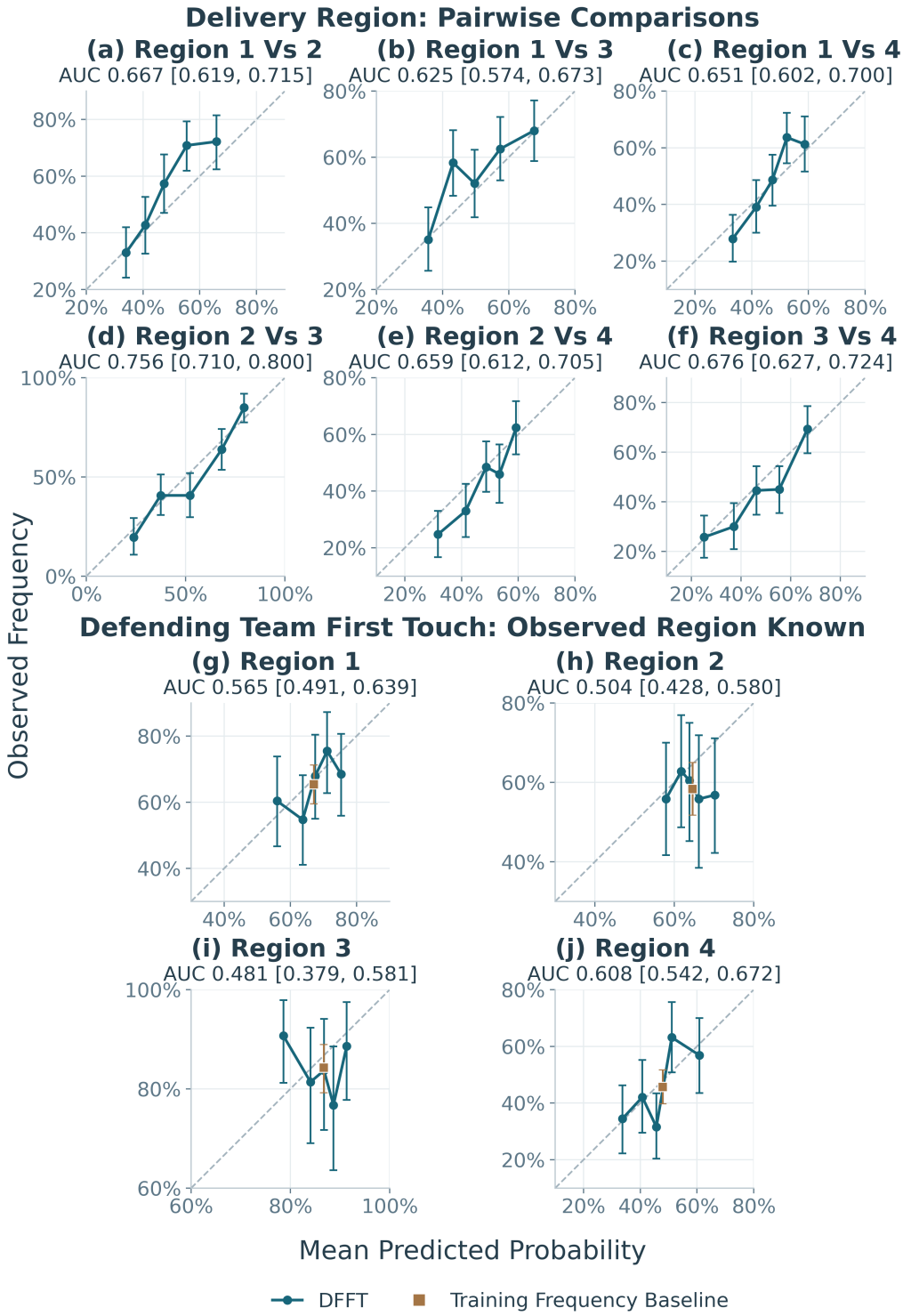


Figure C.11: Corner delivery and first-touch calibration. (a–f) Pairwise delivery predictions, with probabilities referring to the first named region. **(g–j)** Defending-team first-touch predictions conditional on the observed delivery region, using player positions at the kick. Circles summarize five probability groups. Squares compare regional training frequencies (horizontal) with testing frequencies (vertical). Dashed lines indicate perfect calibration; error bars and AUC intervals are 95% whole-match bootstrap intervals. Predicting delivery location based on the starting configuration appears substantially easier than first touch for a given region.

Exploratory Sequential Corner Extension

D.1 Relationship Between the Core and Sequential Corner Analyses

Both analyses use corners from the same five-league source data, but they answer different questions and are treated as distinct evaluation cohorts. The core four-region model jointly predicts delivery and first touch from the kick-time configuration. The sequential extension uses a separate seven-group delivery target as the first component of a hybrid delivery–transition–contact pipeline. In the extension, *first contact* is synonymous with the core analysis's *first touch*: the first recorded player contact following the delivery. The seven-group archive carries its own frozen match-disjoint partition; it is not a matched re-evaluation of the four-region test cohort. Consequently, differences between their reported metrics should not be interpreted as performance gains or losses.

Analysis	Purpose	Training	Test	Primary delivery summary
Four-region joint model	Core DFFT application using kick-time positioning to predict delivery and first touch	982	985	Mean AUC across the six pairwise region comparisons
Seven-group delivery model	Stage 1 input to the exploratory sequential analysis	1,510	368	Macro and support-weighted one-versus-rest AUC

Table 5: Relationship between the two corner-delivery analyses. The four-region split contains 465 training and 465 testing matches; the seven-group split contains 720 training and 188 testing matches. Different targets, eligibility rules, partitions, and AUC summaries preclude direct numerical comparison.

D.2 Stage 1: Seven-Group Delivery Model and Comparators

Let $Z_i \in \{1, \ldots, 7\}$ denote the delivery group, L_i the league, S_i the corner side, A_i the attacking team, and D_i the defending team. Let $\widetilde{\mathbf{c}}_i$ denote standardized principal-component scores from the concatenated attacking and defending start-density maps. Each learned comparator uses the multinomial form

$$P(Z_i = k) = \frac{\exp(\eta_{ik})}{\sum_{h=1}^{7} \exp(\eta_{ih})}.$$

Delivery-group frequency. The simplest comparator assigns every corner the delivery-group proportions in the fitting partition:

$$\widehat{P}(Z_i = k) = \frac{n_k}{\sum_{h=1}^{7} n_h}.$$

Pre-corner context. This model uses league, corner side, attacking team, and defending team:

$$\eta_{ik} = \alpha_k + \gamma_{L_i,k} + \delta_{S_i,k} + a_{A_i,k} + d_{D_i,k}.$$

League and side enter as fixed effects; attacking- and defending-team indicators receive ordinary ridge penalties.

Start density only. This model uses PCA scores from the concatenated attacking and defending 414-bin start maps:

$$\eta_{ik} = \alpha_k + \boldsymbol{\beta}_k^\top \widetilde{\mathbf{c}}_i.$$

Pre-corner context plus start density. This model combines both feature sets:

$$\eta_{ik} = \alpha_k + \gamma_{L_i,k} + \delta_{S_i,k} + a_{A_i,k} + d_{D_i,k} + \boldsymbol{\beta}_k^\top \widetilde{\mathbf{c}}_i.$$

The three preceding learned comparators minimize multinomial negative log likelihood plus $\lambda\|\boldsymbol{\theta}\|_2^2/2$, with $\lambda = 2$ and an unpenalized intercept.

Hierarchical context plus start density. This model uses the same linear predictor as the context-plus-density model, but treats the team effects as exchangeable:

$$a_{t,k} \sim \mathcal{N}(0, \sigma_A^2), \qquad d_{t,k} \sim \mathcal{N}(0, \sigma_D^2), \qquad \sigma_A = \sigma_D = 0.5.$$

League, side, and PCA coefficients retain the ridge penalty $\lambda = 2$. The Gaussian team priors partially pool team effects toward zero, and an unseen team receives the population-mean effect. The model is fitted by maximum a posteriori estimation; it does not integrate coefficient or prior-scale uncertainty through posterior sampling.

For every density-based comparator, PCA and score standardization are fitted within the relevant training partition. PCA targets 90% of the training variation subject to a maximum of 30 components. Categorical levels and all model coefficients are likewise estimated without using the corresponding validation or test observations.

D.2.1 Stage 1 Evaluation and Results

All models predict the same seven delivery groups. The primary evaluation uses 368 corners from 188 matches excluded from model fitting. Ten match-grouped folds within the training partition provide fold-excluded predictions for stability assessment, but DFFT hyperparameters were selected using these folds; the held-out test is therefore the primary comparison. A scalar temperature of 1.032 is estimated from fold-excluded training predictions and applied unchanged to the test probabilities.

The conventional comparators use the fixed regularization settings specified above, whereas DFFT hyperparameters are selected using the match-grouped training folds. Their results therefore compare the fitted specifications actually evaluated; they should not be interpreted as an algorithm-level comparison under identical hyperparameter-search budgets.

Multiclass log loss evaluates the probability assigned to the observed group, and the Brier score is the mean summed squared error across the seven predicted probabilities. Classwise expected calibration error (ECE) measures the average one-versus-rest difference between predicted and observed

frequencies. We also calculate one-versus-rest AUC, averaging either equally across groups (macro AUC) or in proportion to group support (weighted AUC). Lower log loss, Brier score, and ECE are preferred; higher AUC is preferred.

Model	Log loss	Brier	ECE	Macro AUC	Weighted AUC
Delivery-group frequency	1.898	0.843	0.013	0.500	0.500
Pre-corner context	1.995	0.873	0.056	0.522	0.527
Start-density PCA	1.799	0.811	0.031	0.644	0.638
Context plus density	1.914	0.850	0.059	0.615	0.611
Hierarchical plus density	1.859	0.832	0.046	0.626	0.623
DFFT	1.762	0.793	0.025	0.656	0.649
Calibrated DFFT	1.761	0.793	0.027	0.656	0.648

Table 6: Stage 1 performance on the 368-corner match-disjoint test set for the evaluated model specifications.

Uncalibrated DFFT has the best log loss, Brier score, macro AUC, and weighted AUC among the evaluated specifications. Temperature scaling slightly improves log loss but otherwise changes performance little. Start-density PCA is the strongest conventional comparator, and the numerical DFFT advantage over it is modest. The hierarchical model improves on the ordinary context-plus-density model but remains weaker than density alone, while pre-corner context underperforms the delivery-frequency comparator on both proper scoring rules.

D.3 Shared Diagnostic: Change from Start to First Contact

Figure D.12 illustrates how the attacking and defending densities change between the corner start and first contact. The right-hand panels show the bin-wise difference, with green indicating increased density and orange indicating decreased density. The example is the held-out corner nearest the median individual persistence error, rather than a best- or worst-case transition. It shows that both teams redistribute their spatial presence while the ball is in flight, particularly around the eventual delivery location.

We quantify this change using persistence, which assumes that the first-contact density map is identical to the start map. For corner i, persistence error is the summed squared difference over both team maps and all spatial bins:

$$E_i^{\text{persist}} = \sum_{t \in \{A,D\}} \sum_{b} \left(\rho_{1itb} - \rho_{0itb}\right)^2 .$$

Both frames contain the same ten attacking outfield players, ten defending outfield players, and defending goalkeeper, so the comparison isolates spatial change rather than a change in represented actors. Mean persistence error across the 106 match-disjoint held-out corners is 77.55. This value quantifies the error incurred by treating the initial configuration as the contact configuration. Because it is not scale-free, its meaning comes from comparison with transition-model errors: a lower error indicates that the model recovers part of the spatial evolution that persistence ignores. The observed change

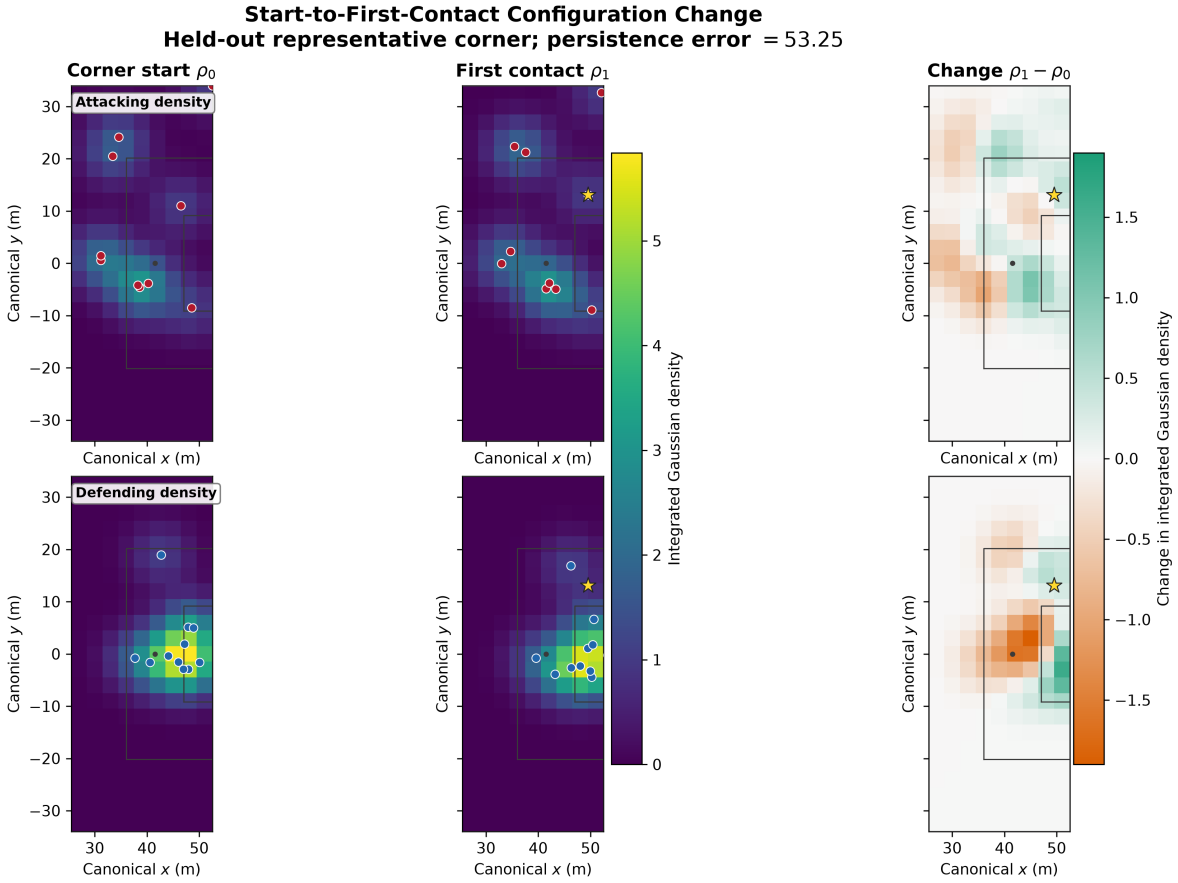


Figure D.12: Attacking and defending densities at the corner start, at first contact, and their bin-wise differences for a representative held-out corner. Player locations are shown on the observed start and first-contact maps, and the star marks the delivery endpoint.

motivates testing a transition model, but it does not by itself establish that explicit transition modeling will improve first-contact prediction.

D.4 Observed-Contact Reference Model (Oracle Diagnostic)

D.4.1 Cohort and Prediction Protocol

The reference-model cohort contains 693 observed first-contact frames assigned to one of the six inside delivery groups: 569 training contacts and 124 held-out contacts. No native first-contact classifier is available for the outside group, so outside deliveries are excluded. Each training row receives a prediction from the saved DFFT model that excluded its fold, while held-out rows receive predictions from the final model fitted to all 569 training frames. Because DFFT hyperparameters were selected using the training folds, the fold-excluded predictions are exploratory rather than nested model-selection estimates; the match-disjoint held-out partition is the primary benchmark.

The comparator assigns each corner the attacking-contact rate of its realized delivery group estimated from training data. It therefore controls for differences in first-contact success across destinations but does not use the spatial configuration within a group. Like the DFFT reference model, it uses the future realized destination and is diagnostic rather than deployable.

D.4.2 DFFT Contact Model

At first contact, the density representation contains ten attacking outfield players, ten defending outfield players, and the defending goalkeeper. The attacking and defending density in each of the 414 bins is rounded to one of 16 levels and combined into a joint state $s_{ib} \in \{0, \ldots, 255\}$. For each inside group z, DFFT learns separate bin-level state distributions conditional on attacking or defending first contact. Their weighted log-density ratio contributes evidence to the binary log odds shown in the main text. A cell with zero spatial weight is inactive; the sign of its contribution depends on the observed state-specific density ratio rather than on the weight alone.

Equation (17) gives the group-specific classifier, where $f_{z,b}^{(1)}$ and $f_{z,b}^{(0)}$ are the learned state distributions when the corner-taking team and opposition make first contact, respectively, and $\mathcal{B}_z$ is the set of active bins for group z.

D.4.3 Evaluation and Results

AUC evaluates ranking across attacking- and defending-contact outcomes. Brier score and log loss evaluate probability quality, while accuracy reports classification after applying a fixed 0.5 threshold. Because accuracy depends on that threshold, AUC and the proper scoring rules are the primary measures.

Model	Accuracy	AUC	Brier	Log loss
Observed delivery-group frequency	0.710	0.811	0.188	0.560
Observed-contact DFFT	0.677	0.831	0.175	0.522

Table 7: First-contact performance on 124 match-disjoint held-out contacts from the six supported inside delivery groups. Both models use the realized delivery group; DFFT additionally uses the observed contact configuration.

The observed-contact DFFT has numerically better AUC and proper scores than delivery-group

frequency, but the comparator is already strong because the realized destination contains substantial outcome information. We therefore interpret the differences as observed held-out advantages rather than attributing the full DFFT performance to the contact configuration. The lower 0.5-threshold accuracy does not contradict the probability-score differences: it reflects one fixed decision threshold rather than the quality of the full probability ranking.

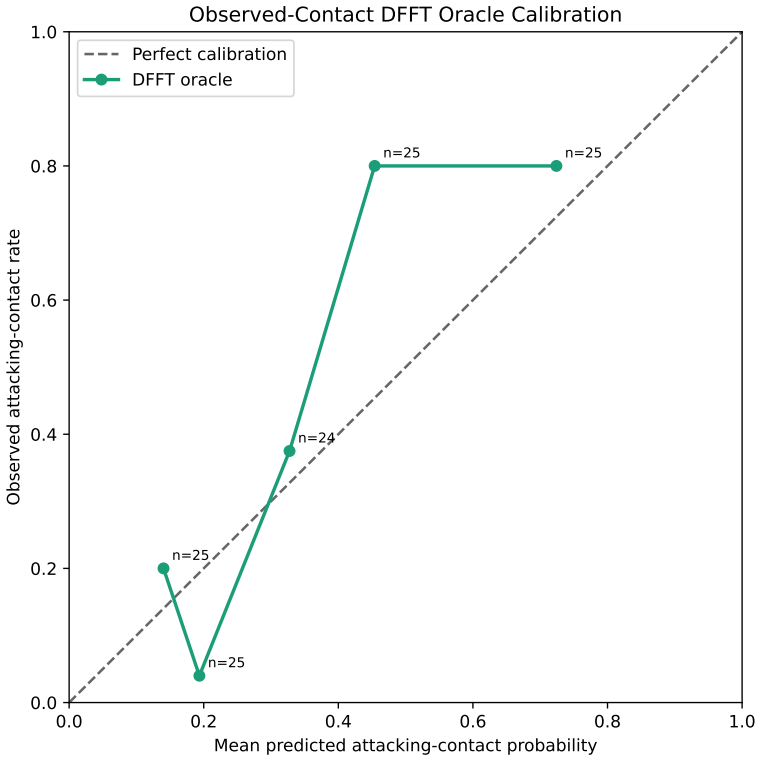


Figure D.13: Held-out calibration of the observed-contact DFFT reference model using five equal-frequency bins. Labels give the number of contacts in each bin. The curve is descriptive and imprecise because each bin contains only 24 or 25 contacts.

The mean DFFT probability on the held-out set is 0.368, compared with an observed attacking-contact rate of 0.444. The binned curve is uneven and should not be interpreted as evidence of perfect calibration. It is retained as a diagnostic alongside Brier score and log loss rather than used as the primary justification for the model.

D.4.4 Interpretation Guardrails

The reference model observes ρ_1 contemporaneously with the outcome, and the contacting player is represented in that frame. Its performance therefore shows that DFFT recognizes information in the realized spatial contest; it does not show that the contest was predictable at the corner start or that any density pattern caused first-contact control. The model also conditions on the realized inside delivery group and cannot score outside deliveries. It is therefore an information-rich diagnostic, not a deployable corner forecast or a guaranteed upper performance bound. Moreover, its gap from the sequential chain combines differences in destination knowledge and contact-state knowledge and does not isolate transition-map error.

D.5 Stage 2: Start-to-First-Contact Transition Models

D.5.1 Paired Configurations and Dimensionality Reduction

The transition cohort contains 584 corners with both a start frame and a first-contact frame: 478 training corners and 106 match-disjoint held-out corners. The primary matched-actor representation contains the same ten attacking outfield players, ten defending outfield players, and defending goalkeeper at both times. This prevents apparent temporal change from being created by adding or removing an actor between frames.

We concatenate the attacking and defending maps into an 828-dimensional vector. PCA is fitted separately to the start and first-contact vectors within each fitting partition. Each PCA targets 90% of the training variation subject to a maximum of 30 components, after which the start scores are standardized using training-partition quantities. Let $\widetilde{\mathbf{c}}_{0i}$ denote the standardized start score and $\mathbf{c}_{1i}$ the first-contact score.

D.5.2 Transition Specifications

Persistence copies the start map into the contact frame, $\widehat{\rho}_{1i} = \rho_{0i}$, and provides the principal no-transition benchmark. The delivery-group mean instead assigns the mean first-contact map observed for the realized group. Because it uses the future destination, the group mean is diagnostic rather than deployable.

The observed-group additive model is

$$\mathbf{c}_{1i} = \boldsymbol{\alpha} + B\widetilde{\mathbf{c}}_{0i} + G\mathbf{e}_{Z_i} + \boldsymbol{\epsilon}_i,$$

where $\mathbf{e}_{Z_i}$ is the indicator vector for the realized delivery group. This model estimates a common relationship between the start and contact configurations while allowing the expected contact state to shift by destination. The observed-group interaction model additionally permits the start-to-contact relationship to vary across delivery groups:

$$\mathbf{c}_{1i} = \boldsymbol{\alpha} + B\widetilde{\mathbf{c}}_{0i} + G\mathbf{e}_{Z_i} + \sum_z H_z\widetilde{\mathbf{c}}_{0i}\,\mathbf{1}(Z_i = z) + \boldsymbol{\epsilon}_i.$$

Both models are fitted by multi-output ridge regression with penalty $\lambda = 10$. The observed-group versions measure transition performance when the future destination is known and are therefore diagnostic. For deployable prediction, the additive and interaction models instead generate one candidate contact map for every delivery group and combine them using the calibrated Stage 1 probabilities:

$$\widehat{\rho}_{1i}^{\,\text{deployable}} = \sum_z \widehat{P}_{\text{DFFT}}(Z_i = z \mid \rho_{0i})\widehat{\rho}_{1i}(z).$$

The candidate maps retain all seven delivery groups, including outside. Training rows receive fold-excluded Stage 1 probabilities and transition predictions; held-out rows receive Stage 1 probabilities

from the delivery model fitted on all training corners and transition predictions from the model fitted on all paired training corners.

Stage 2 covers all seven delivery groups, but the observed-contact DFFT model covers only the six inside groups. The native Stage 3 handoff therefore retains a projected candidate map for each inside group and combines their contact probabilities conditional on the delivery being inside. Outside remains available for statistical transition analyses but has no corresponding native DFFT contact classifier.

D.5.3 Nonnegative Projection

Inverse PCA reconstruction is linear and can produce negative density values. Each reconstructed attacking and defending map is therefore projected separately to the nonnegative space while preserving its nonnegative total mass. For reconstructed values $\widetilde{\rho}_b$, define

$$m = \max\left(\sum_b \widetilde{\rho}_b, 0\right), \qquad \Pi_+(\widetilde{\rho})_b = m\frac{\max(\widetilde{\rho}_b, 0)}{\sum_h \max(\widetilde{\rho}_h, 0)}.$$

Each group-conditional candidate is projected before probability weighting. Before projection, 28.97% of cells in the deployable interaction reconstructions on the held-out set are negative. Projection removes all negative values, preserves team-map mass to numerical precision, and changes complete-map error only slightly, from 61.71 to 61.59.

D.5.4 Evaluation and Results

For corner i, complete-map error is the squared difference summed over attacking and defending maps and all 414 bins. We report its mean across corners:

$$\mathrm{MSSE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\{A,D\}}\sum_b (\rho_{1itb} - \widehat{\rho}_{1itb})^2 .$$

Secondary measures evaluate whether the predicted map preserves overload within the delivered-to group: mean absolute error, correlation with observed overload, and accuracy of the overload sign. The complete-map error is the primary transition criterion because it evaluates the full Stage 2 output rather than one regional summary.

The observed-group additive model reduces held-out error from 77.55 under persistence to 57.84, showing that the individual start configuration contains substantial information about the later state once the destination is supplied. The deployable models must also account for destination uncertainty, but both still reduce error by approximately 21% relative to persistence. Their held-out errors differ by only 0.50. The interaction model has the lower training-fold error, 62.74 versus 62.93, and was therefore designated as the primary Stage 3 handoff before comparing their held-out performance. The additive model's slightly lower held-out error is reported as a test-set result rather than used to revise that selection.

The deployable interaction model also gives a slightly higher delivered-group overload correlation than the additive model, 0.842 versus 0.840. These high correlations indicate that both models preserve

Transition model	Training-fold error	Test error	Availability
Persistence	78.41	77.55	Pre-kick
Delivery-group mean	72.09	76.19	Observed group
Additive, observed group	59.69	57.84	Observed group
Interaction, observed group	63.92	59.55	Observed group
Additive, predicted group	62.93	61.09	Pre-kick
Interaction, predicted group	62.74	61.59	Pre-kick

Table 8: Mean summed squared bin error for matched-actor start-to-first-contact transitions. Observed-group models use the realized delivery destination and are diagnostic; predicted-group models use only information available when the corner is taken.

much of the relative local balance in the eventual delivery region, even though their conditional mean maps cannot reproduce the exact player movements in a realized corner. These are reconstruction results, not evidence of improved outcome prediction. Because the downstream DFFT classifier is nonlinear and was fitted on observed contact configurations, lower map error does not guarantee better first-contact probabilities; the end-to-end Stage 3 evaluation tests that question directly.

A dynamic DFFT transition would instead estimate $P_{\mathrm{DFFT}}(\rho_1 \mid \rho_0, Z)$, treating the future density configuration as a distribution rather than a single regression target. Samples from this distribution could be passed through the group-specific contact classifier and averaged, approximating $\mathbb{E}[g_Z(\rho_1) \mid \rho_0, Z]$ while retaining within-destination uncertainty. Developing such a model would require learning temporally conditional density distributions and their spatial dependence; it is a proposed extension rather than a component evaluated here.

D.6 Stage 3: End-to-End Conditional-Inside Prediction

D.6.1 Native DFFT Composition and Cohort

Stage 2 supplies a projected candidate map $\widehat{\rho}_{1i}(z)$ for each of the six inside groups. The corresponding frozen DFFT classifier returns $\widehat{q}_{iz}$, and Stage 1 supplies the calibrated delivery probability $\widehat{p}_{iz}$. Their normalized weighted sum is the conditional-inside probability reported in the main text. Training rows use fold-excluded Stage 1, Stage 2, and contact-model predictions; held-out rows use models fitted on the complete training partitions.

The native evaluation contains 390 inside training corners and 83 inside held-out corners. The remaining 111 paired corners include outside deliveries and cannot be scored by the native contact model. Without an outside classifier, the unconditional first-contact probability is only bounded:

$$\sum_{z \in \mathcal{I}} \widehat{p}_{iz}\widehat{q}_{iz} \le P(Y_i = 1 \mid \rho_{0i}) \le \sum_{z \in \mathcal{I}} \widehat{p}_{iz}\widehat{q}_{iz} + \widehat{p}_{i,\mathrm{outside}}.$$

The conditional-inside evaluation avoids making an assumption about first-contact success outside the modeled regions, but its probabilities should not be interpreted as unconditional forecasts over all

corners.

All held-out predictions are generated by models fitted without the corresponding test matches. Where predictions are passed from one fitted stage to another during training, they are match-fold-excluded. The eligible samples differ across stages because start frames, contact frames, and supported delivery groups are not available for every corner. The held-out partition was inspected during Stages 1 and 2; the interaction transition was retained using the training-fold comparison, and the Stage 3 variants were specified without selecting among them on Stage 3 held-out performance.

D.6.2 Native Comparators

Prevalence assigns every held-out inside corner the attacking-contact rate from the inside training cohort. Observed-group frequency assigns the training contact rate for the realized delivery group; because it uses the future destination, it is diagnostic rather than deployable. The no-transition composition applies every inside-group contact classifier to the matched-actor start map and averages the resulting probabilities using Stage 1 weights. It is a deployable ablation of Stage 2, not a classifier trained directly on start maps. The additive and interaction chains instead apply the classifiers to their respective Stage 2 candidate maps. The observed-contact reference model uses the realized ρ_1 and realized delivery group and is not available pre-kick.

D.6.3 Conditional-Mean Approximation

The sequential chain does not first average the candidate maps across delivery groups. It evaluates each group-specific map $\widehat{\rho}_{1i}(z)$ with the corresponding frozen contact classifier g_z and then averages the resulting probabilities using the Stage 1 delivery weights. Destination uncertainty is therefore propagated at the probability level.

Within each delivery group, PCA–ridge supplies one conditional-mean map. The prediction currently uses

$$g_z\Big(\widehat{\mathbb{E}}[\rho_{1i} \mid \rho_{0i}, Z_i = z]\Big)\,,$$

whereas a model representing the full transition distribution would target

$$\mathbb{E}[g_z(\rho_{1i}) \mid \rho_{0i}, Z_i = z]\,.$$

These quantities are generally unequal because g_z is nonlinear. Consequently, reducing squared error for the conditional-mean map does not guarantee better first-contact probabilities. The observed results are consistent with this approximation and with a mismatch between the realized maps used to train g_z and the smoother predicted maps supplied at deployment. They do not isolate these mechanisms from destination uncertainty, transition error, limited sample size, or other model misspecification.

D.6.4 Held-Out Performance

Log loss and Brier score evaluate probability quality, AUC evaluates ranking, and expected calibration error (ECE) compares predicted and observed frequencies. The native models are evaluated on the

same 83 held-out inside deliveries.

Model	Information availability	Log loss	Brier	AUC	ECE
Prevalence	Pre-kick baseline	0.687	0.246	0.500	0.082
Observed-group frequency	Realized delivery; diagnostic	0.579	0.197	0.788	0.079
No-transition composition	Pre-kick; conditional inside	0.674	0.240	0.604	0.103
Additive transition + DFFT	Pre-kick; conditional inside	0.676	0.242	0.592	0.075
Interaction transition + DFFT	Pre-kick; conditional inside	0.673	0.240	0.598	0.092
Observed-contact DFFT	Realized delivery and contact; diagnostic	0.535	0.181	0.804	0.147

Table 9: Stage 3 performance on 83 match-disjoint held-out inside deliveries. Observed-group frequency and the observed-contact reference model use future information and are diagnostic.

The additive and interaction chains perform similarly. Most importantly, the sequential interaction chain does not show an established advantage over the no-transition composition: their log losses are 0.673 and 0.674, and their AUCs are 0.598 and 0.604. Observed-group frequency and the observed-contact model are information-rich diagnostics because they use the realized destination or contact state.

D.6.5 Match-Clustered Uncertainty

Corners from the same match are not independent. We therefore resample the 63 held-out matches with replacement 5,000 times, retaining every eligible corner from each selected match. For each bootstrap draw, we calculate the candidate-minus-reference difference in log loss, Brier score, and AUC. Negative log-loss and Brier differences favor the interaction chain; positive AUC differences favor it. The intervals describe sampling variation across the observed held-out matches and do not account for model-selection uncertainty or distribution shift.

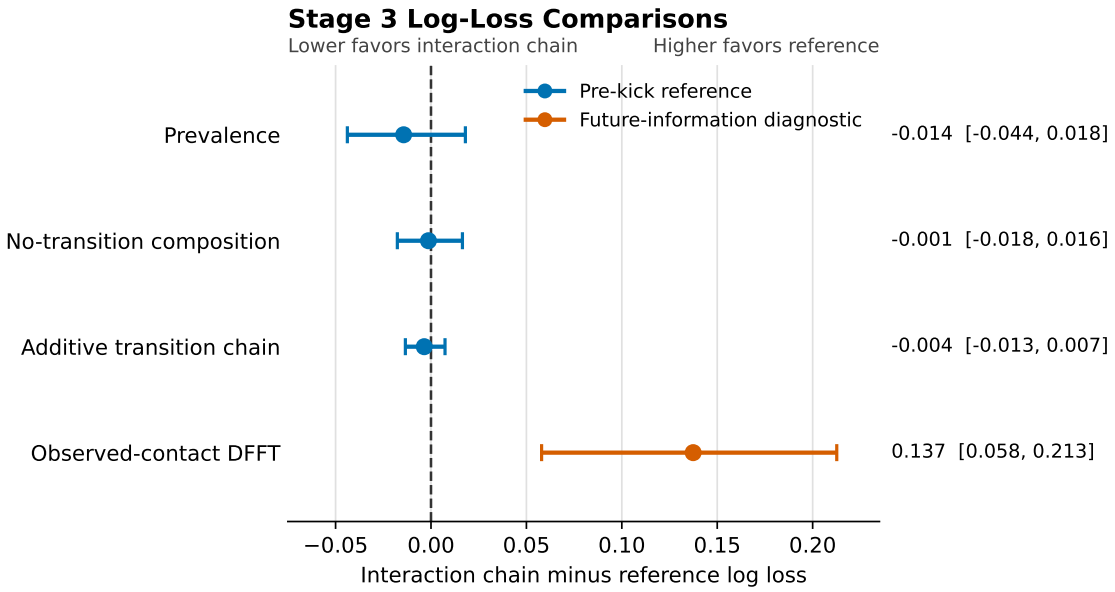
Reference	Δ log loss	Δ Brier	Δ AUC
Prevalence	-0.014 $[-0.044, 0.018]$	-0.006 $[-0.019, 0.008]$	0.098 $[-0.040, 0.220]$
No-transition composition	-0.001 $[-0.018, 0.016]$	-0.000 $[-0.008, 0.007]$	-0.006 $[-0.090, 0.067]$
Additive transition	-0.004 $[-0.013, 0.007]$	-0.002 $[-0.006, 0.003]$	0.005 $[-0.037, 0.044]$
Observed-contact DFFT	0.137 $[0.058, 0.213]$	0.059 $[0.025, 0.092]$	-0.206 $[-0.346, -0.076]$

Table 10: Interaction-chain metric differences and 95% match-clustered bootstrap intervals. Each entry is interaction transition plus DFFT minus the listed reference.

Every interval comparing the interaction chain with prevalence, the no-transition composition, or the additive chain contains zero. These results do not establish an improvement, but an interval spanning zero is not evidence that the models are equivalent. In contrast, all three intervals comparing the interaction chain with the observed-contact model exclude zero and favor the observed-contact model. The data therefore support a stable observed-to-predicted gap but not a clear ordering among the deployable native models.

D.6.6 Conventional Predicted-Map Diagnostics

The frozen DFFT contact classifiers were trained on observed contact configurations but receive smooth predicted maps in the deployable chain. To test whether Stage 2 maps contain outcome signal when the downstream model is trained on the same type of input it receives at test time, we also fit conventional



Points are paired estimates; bars are 95% intervals from 5,000 whole-match bootstrap samples (83 corners, 63 matches).

Figure D.14: Paired Stage 3 log-loss differences. Points show interaction-chain minus reference log loss, and bars show 95% intervals from 5,000 whole-match bootstrap samples of the 83-corner, 63-match conditional-inside cohort. Negative values favor the interaction chain. Blue comparisons use pre-kick references; the orange observed-contact DFFT reference uses future information and is diagnostic. Intervals crossing zero do not establish a difference.

ridge-logistic models to the fold-excluded Stage 2 maps. These are statistical diagnostics rather than DFFT models and are evaluated on all 106 held-out paired corners, including outside deliveries.

Model	Log loss	Brier	AUC
Prevalence	0.707	0.257	0.500
Full additive, local	0.617	0.217	0.688
Full interaction, local	0.616	0.216	0.691
Observed group only	0.504	0.162	0.859
Observed-contact model, local	0.452	0.146	0.879

Table 11: Selected conventional Stage 3 diagnostics on all 106 held-out paired corners. Models using the observed group or observed contact map contain future information and are not deployable.

The deployable conventional interaction model improves substantially over prevalence, indicating that Stage 2 maps retain outcome information. Its stronger performance relative to the native DFFT chain is consistent with the conditional-mean and input-distribution limitations described in Appendix D.6.3, but the experiment does not isolate those explanations from other sources of error. These analyses also use different cohorts and model classes, so their metric values should not be treated as a direct head-to-head comparison.

D.6.7 Interpretation Guardrails

The end-to-end pipeline uses DFFT at Stages 1 and 3 but PCA–ridge at Stage 2; it is therefore a DFFT hybrid rather than a fully dynamic DFFT. Native metrics condition on the realized delivery being inside, while the conventional diagnostics include all paired delivery groups. The observed-group and observed-contact references use future information. The improved Stage 2 reconstruction error did not establish improved first-contact prediction, and the current sample does not establish equivalence between the sequential and no-transition compositions. Finally, the held-out partition was inspected during Stages 1 and 2, so Stage 3 variants were pre-specified or selected using training-fold results rather than chosen for their Stage 3 test performance.

Event processing and coordinate transformations

E.1 Linking player arrangements to outcomes

Passes and corner kicks are identified within each match and linked to their corresponding 360 freeze frames using event identifiers. The player positions in each freeze frame describe the arrangement at the time of the action. Players are assigned to the attacking or defending team relative to the passing or corner-taking team, which remains the reference team throughout the analyzed sequence.

The start and end locations of a pass determine its recorded displacement and passing direction. For corner kicks, the recorded endpoint identifies the delivery location. Event outcomes and linked subsequent actions provide additional outcome information, including pass completion and first contact after a corner. These subsequent actions are used to define outcomes and are not included in the start-frame representation. The specific labeling rules are given with each prediction task.

E.2 Coordinate conversion and team alignment

StatsBomb locations are recorded on a standardized pitch with coordinates ranging from 0 to 120 longitudinally and from 0 to 80 laterally. Event coordinates are oriented toward the attacking goal of the team performing the action. For an event already expressed in the reference team's orientation, a location $(x_{\mathrm{SB}}, y_{\mathrm{SB}})$ is converted as

$$x = \left(\frac{x_{\mathrm{SB}}}{120} - \frac{1}{2}\right) L, \qquad y = -\left(\frac{y_{\mathrm{SB}}}{80} - \frac{1}{2}\right) W, \tag{22}$$

where L and W are the pitch length and width in meters. The origin is therefore at the pitch center, with positive x toward the reference team's attacking goal and positive y toward that team's left. Where actual pitch dimensions are unavailable, we use $L = 105\,\mathrm{m}$ and $W = 68\,\mathrm{m}$.

A subsequent event performed by the opposition may be recorded in the opposing team's orientation. Its converted coordinates are then rotated by $180°$,

$$(x, y) \longmapsto (-x, -y), \tag{23}$$

to place the event in the original reference team's frame. This transformation applies to the entire event, including any associated player positions, endpoints, and visible-area vertices. Individual defending players within the original start frame are not rotated separately, because all positions in that frame already share the same orientation.

Lateral reflection used to pool situations from opposite sides of the pitch is a separate operation. Where such pooling is applied, the transformation $(x, y) \mapsto (x, -y)$ is applied consistently to the player positions, ball position, event endpoint, and visible-area polygon. The original side is retained in the metadata, and any side-dependent direction labels are transformed consistently. The use of this pooling is specified for each analysis.

E.3 Pitch markings and interpretation of distances

The conversion in Equation (22) provides a common metric representation, but does not reconstruct exact physical distances on pitches of unknown dimensions. In particular, uniformly rescaling StatsBomb coordinates does not preserve the physical dimensions of every pitch marking. For example, the 44-unit width of the penalty area becomes 37.4 m when mapped to a pitch width of 68 m.

Regions defined by pitch markings therefore use the corresponding StatsBomb boundaries, transformed consistently into the coordinates used for analysis. On the nominal pitch, the attacking penalty area is bounded by $36.75 \leq x \leq 52.5\,\mathrm{m}$ and $-18.7 \leq y \leq 18.7\,\mathrm{m}$. This preserves classification relative to the recorded penalty-area boundaries. Distances reported in these coordinates should be interpreted as distances on the standardized pitch.

E.4 Pitch coordinates

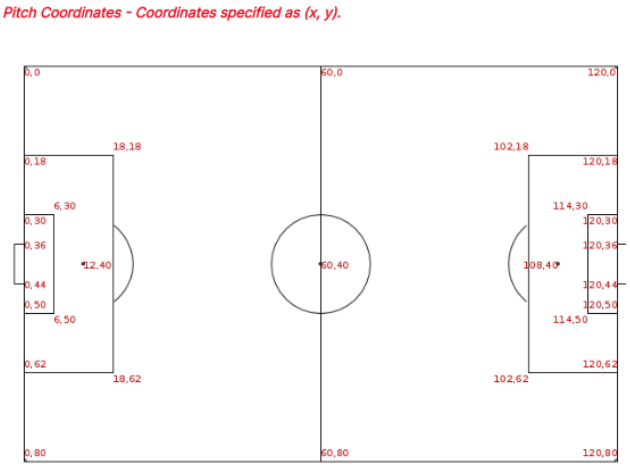


Figure E.15: StatsBomb pitch coordinates, specified as (x, y). Source: StatsBomb API Events Structure and Data Specification v6.0.0.