



**HAL**  
open science

# **PSOLxG : Une nouvelle métrique pour décoder la performance des gardiens de but et des attaquants au football**

Rémi Verronneau, Bastien Cadiou, Vianney Perchet

## ► **To cite this version:**

Rémi Verronneau, Bastien Cadiou, Vianney Perchet. PSOLxG : Une nouvelle métrique pour décoder la performance des gardiens de but et des attaquants au football. Journée commune EGC/AFIA Gestion et Analyse de données Sportives (GAS'25), May 2025, Caen, France. <hal-05090045>

**HAL Id: hal-05090045**

**<https://hal.science/hal-05090045v1>**

Submitted on 29 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# PSOLxG : Une nouvelle métrique pour décoder la performance des gardiens de but et des attaquants au football



Rémi VERRONNEAU, Bastien CADIOU, Vianney PERCHET

CREST – ENSAE – CRITEO AI LAB – Footovision

## Motivations

- Le but est un événement rare et décisif en football. Les modèles classiques d'**Expected Goals** (xG) évaluent la probabilité de marquer avant le tir, mais ignorent des facteurs clés tels que la vitesse et la trajectoire finale du ballon. Ils se concentrent essentiellement sur la position et l'angle de tir, en négligeant l'influence du gardien, pourtant déterminante.
- Nous introduisons les **Post-Shot-On-the-Line Expected Goals** (PSOLxG), une métrique conçue pour évaluer plus précisément les performances des gardiens et des tireurs.
- Nous proposons une méthode de projection pour estimer l'évolution des performances individuelles en cas de transfert.

## Challenges

- Non-linéarités dans les trajectoires de tir** : Les tirs peuvent suivre des trajectoires complexes — effets, courbes, changements de direction — et varier rapidement en vitesse. Ces dynamiques non linéaires et multi-dimensionnelles sont difficiles à modéliser avec des approches traditionnelles. Nous utilisons les *Explainable Boosting Machines* (EBMs), qui étendent les modèles GAMs en intégrant des interactions entre variables, tout en restant interprétables.
- Importance de la calibration des modèles** : les probabilités de but doivent refléter les fréquences réellement observées. Sans cela, les performances des gardiens ou des attaquants peuvent être mal interprétées, ce qui compromet l'utilité des métriques produites.
- Concevoir des métriques pour gardiens et attaquants** : Développer des métriques robustes et interprétables, capables d'évaluer séparément les performances des gardiens et des attaquants. Cela implique de mieux modéliser l'impact réel des gardiens sur la ligne de but, souvent négligé dans les modèles xG classiques, et de quantifier la capacité des attaquants à générer des tirs dangereux via la puissance, le placement ou des trajectoires imprévisibles.
- Disponibilité des données**

## Méthode

**Objectif** : prédire la probabilité qu'un tir devienne un but, au moment où il atteint la ligne. Cela permet d'évaluer le gardien (réaction sur la ligne) et le tireur (qualité du tir).

- Représentation géométrique des tirs** : Chaque tir est modélisé par :
  - La position du tireur et du gardien au moment du tir.
  - Le point d'arrivée du ballon sur la ligne de but (projection)
  - La vitesse du tir et le nombre de défenseur dans le cône

Cette projection nous permet de comparer équitablement tous les tirs — qu'ils soient marqués ou arrêtés — sans que le modèle ne connaisse l'issue réelle de l'action.

- Calibration des probabilité** : Il ne suffit pas que les probabilités soient précises, il faut aussi qu'elles soient bien calibrées. On évalue la calibration grâce à :
  - des courbes de calibration
  - la métrique ECE (*Expected Calibration Error*)

- Modélisation — Explainable Boosting Machines (EBM)** : modèle interprétable basé sur des arbres de décision simples et transparents. Il apprend :
  - des effets non linéaires
  - des interactions entre variables

Cette approche permet de capter comme une variable seule affecte la vraisemblance ou comment une paire de variables interagit, tout en conservant une interprétabilité forte pour les analystes.

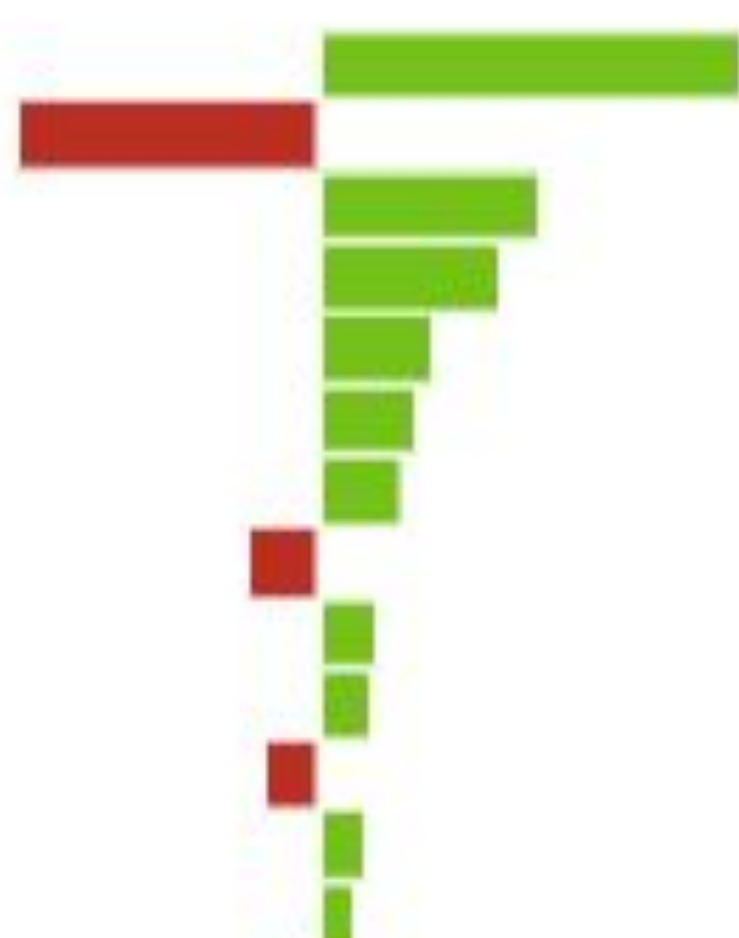
- Ajustement des paramètres (hyperparamètres)** : trouver le meilleur compromis entre précision (log-loss) et calibration (ECE). Pour cela on utilise l'optimisation bayésienne. On définit une fonction objectif combinée, pondérée et normalisée.
  - On modélise cette fonction par un processus gaussien, mis à jour à chaque essai
  - Une fonction d'acquisition choisit ensuite la meilleure combinaison d'hyperparamètres à tester.

## Résultats

### Explications locales d'un tir

Le modèle EBM nous permet de décomposer les log-odds d'un tir donné en contributions individuelles de chaque variable (et de leurs interactions par paire). Par exemple pour un tir qui à 74,8% de chance de finir en but (Image), on remarque que la position finale du ballon en ordonnée (près du poteau) et la vitesse augmentent la probabilité de marquer tandis qu'une position de départ éloignée du but (*X-coord start*) la diminue. Ces explications locales sont précieuses pour des entraîneurs.

Y-coord (end)  
X-coord (start)  
Speed (km/h)  
Y-coord (end and goalkeeper)  
Y-coord (start and end)  
Speed (km/h) and Z-coord (end)  
X-coord (goalkeeper)  
Y-coord (start)  
Z-coord (end) and header  
Z-coord (end)  
X-coord and Y-coord (start)  
X-coord (start) and Z-coord (end)  
Header



### Evaluation des gardiens et des attaquants

En sommant les valeurs PSOLxG de tous les tirs subis on obtient une métrique "Post-Shot Expected Goals conceded". On peut en déduire un indicateur sur la performance du gardien (GKR) et celle de l'attaquant (STR) :

$$GKR = \frac{\sum_i PSOLxG_i - TotalConcededGoals}{TotalConcededShots}$$

$$STR = \frac{\sum_i \max(PSOLxG_i, Goal_i) - \sum_i xG_i}{TotalShots}$$

Un GKR > 0 signifie que le gardien a concédé moins de buts qu'attendus et un STR > 0 signifie que les frappes de l'attaquant ont été plus dangereuses que prévues.

GoalKeeper	Sum PSOLxG	Sum Goals	Sum Shots	GKR
Sanchez (Che)	18.57	14	64	+0.07
Hermansen (Lei)	29.97	25	79	+0.06
Selz (Not)	17.03	14	60	+0.05
...	...	...	...	...
Martinez (AsV)	21.82	24	60	-0.035
Areola (WhU)	5.60	7	23	-0.06
Jose Sa (Wol)	10.56	14	39	-0.08

*Classement des gardiens à la mi-saison 2024-2025*

### Comparaison entre la ligue portugaise et la Première ligue

Pour chaque tir au Portugal, on calcule la moyenne des PSOLxG des trois tirs les plus similaires en Angleterre, puis on la compare à la valeur réellement obtenue au Portugal. Cela permet d'évaluer si un joueur portugais aurait fait mieux que des attaquants de Premier League dans des situations comparables.

On construit la métrique TRAM (TRANSfert Metric) :

$$TRAM = \frac{\sum_i PSOLxG_i - PSOLxG_{KNN\_England\_i}}{TotalShots}$$

où PSOLxG\_KNN\_England\_i représente la moyenne des PSOLxG des trois tirs les plus similaires au tir *i* dans le championnat anglais.

Striker	PSOL Port	PSOL Eng	TRAM	Mean xG Port	Mean xG Eng
Bruma (Br)	0.29	0.14	0.01	0.16	0.15
Dias (Fa)	0.27	0.08	0.009	0.06	0.05
Trincão (Sp)	0.28	0.13	0.007	0.11	0.10

Trois joueurs portugais affichent en moyenne un PSOLxG supérieur de 0,01 à celui d'attaquants anglais sur des tirs similaires.

## Références

- Y. Lou, R. Caruana, J. Gehrke, et G. Hooker. *Accurate intelligible models with pairwise interactions*. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*, pages 623–631, 2013.
- Mahdi Pakdaman Naeini, Gregory F. Cooper, et Milos Hauskrecht. *Obtaining well-calibrated probabilities using Bayesian binning*. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.
- Brian Macdonald. *An expected goals model for evaluating soccer shots*. *Journal of Sports Analytics*, 1(2), 169–187, 2012.