
SUOMEN PALLOLIITTO
VOITA VIERASPELIMATKA
1

Extended Model for Expected Threat in Soccer

Jirka Poropudas (SportIQ) and Ville-Pekka Inkilä (Football Association of Finland)
NESSIS 2021

What is the value of ball possession?

In other words, what is the probability of scoring when a team has the ball at a given location?

SEP. 14, 2021, AT 10:00 AM

Possession Is The Puzzle Of Soccer Analytics. These Models Are Trying To Solve It.

By John Muller

Filed under Soccer

More people have started trying to measure the vast, muddled majority of soccer that happens in between shots. SHAUN BOTTERILL / GETTY IMAGES

https://fivethirtyeight.com/features/possession-is-the-puzzle-of-soccer-analytics-these-models-are-trying-to-solve-it/

Expected Threat (xT)

Originally introduced by Karun Singh in a blog post in 2018.

- Simple Markov chain
- Ball event data
  - No locations for other players
- Offensive model
  - Goals for
- Short time scope
  - Next 5 events

Soccer possession models are gaining steam

Key soccer possession models by publication year, with type of model and possession information

NAME	CREATOR	DEBUT	METHOD	WINDOW	OFF-BALL INFORMATION
Markov Chains	S. Rudd	2011	Markov chain	One possession	Defensive states tagged in event data
Possession-Based Model	N. Mackay	2016	Logistic regression and GAM	One possession	None
Expected Threat (xT)	K. Singh	2019	Markov-like	Next 5 actions (goal for)	None
Valuing Actions by Estimating Probabilities (VAEP)	KU Leuven DTAI	2019	Gradient-boosted trees	Next 10 actions (goal for or against)	Possession history proxies
Expected Possession Value (EPV)	J. Fernández et al.	2019	Multiple models	Next goal (for or against) or end of half	Full tracking data
Possession Value (PV)	Stats Perform	2019	Gradient-boosted trees	Next 10 seconds (goal for)	Possession history proxies
Goals Added (g+)	American Soccer Analysis	2020	Gradient-boosted trees	Two possessions	Possession history proxies
On-Ball Value (OBV)	StatsBomb	2021	Gradient-boosted trees	Two possessions	Broadcast freeze frames (in development)

FiveThirtyEight

https://karun.in/blog/expected-threat.html

Expected Threat (xT)

- "Probability of scoring a goal within the next 5 events, when in possession of the ball at a given location"
- The field is divided into a grid of locations
- Model considers two types of events
  - Shot at goal from the current location
  - Movement to another location
- Event probabilities estimated from real-life event data
  - Scoring probability from any suitable expected goals (xG) model
  - Transition probabilities derived from real-life event data

Expected Threat (xT)

Definition of expected threat: $xT = S \cdot xG + M \cdot P_M \cdot xT$

Decision model:

- Shoot
- Move
  - Probabilities depend on location

Dynamics:

- xG
  - Scoring probability
- P_M
  - Move to new location
  - Transition probability matrix

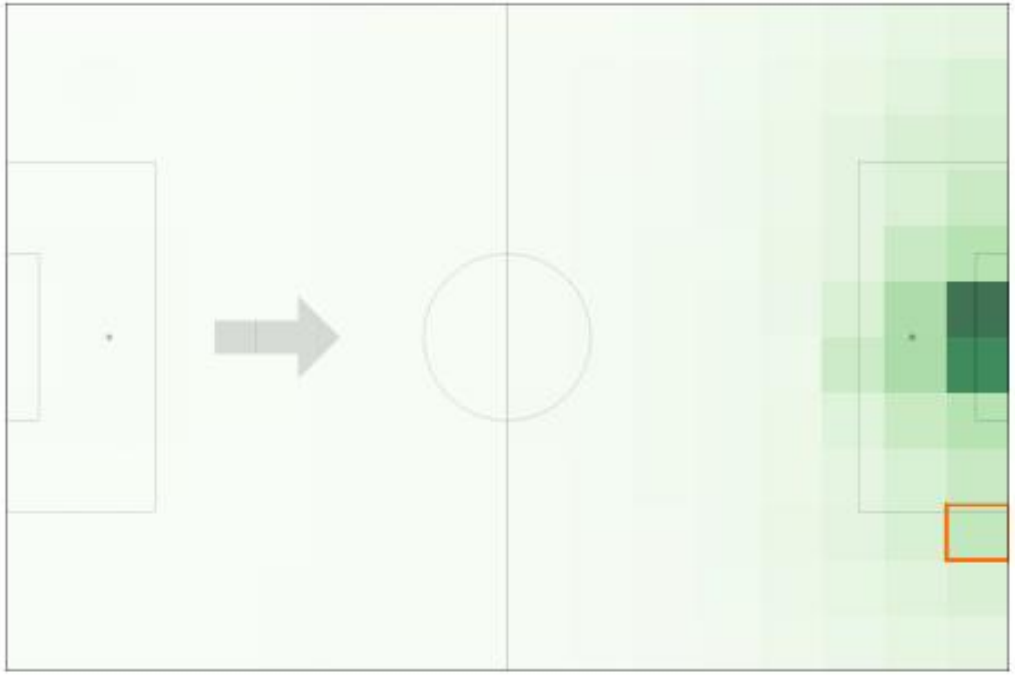
The numerical value of xT is solved by iteration.
That is, by repeating the equation five times.

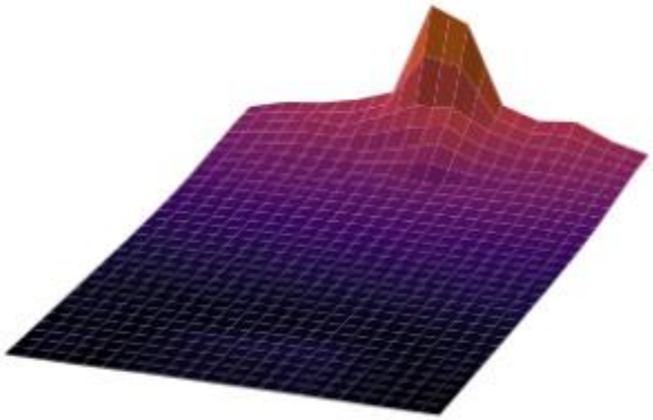
https://karun.in/blog/expected-threat.html

Expected Threat (xT)

Expected Threat (xT) = 0.136

i.e. when the team has the ball in the highlighted zone, they will score in the next **5** actions **13.6%** of the time.





https://karun.in/blog/expected-threat.html

Extended xT model

Two-way xT model

Definition: $xT = S \cdot xG + M \cdot P_M \cdot xT + T \cdot P_{TO} \cdot (-xT)$

"Decision model":

- Shoot
- Move
- Turnover
  - Probabilities depend on location

Dynamics:

- xG
  - Scoring probability
- P_M
  - Move to new location
- P_{TO}
  - Move to new location following a turnover

The minus sign in last term denotes loss of possession!

Extended xT model

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

"Decision model":

- Shoot
- Short or Long move
- Short or Long turnover
  - Probabilities depend on location

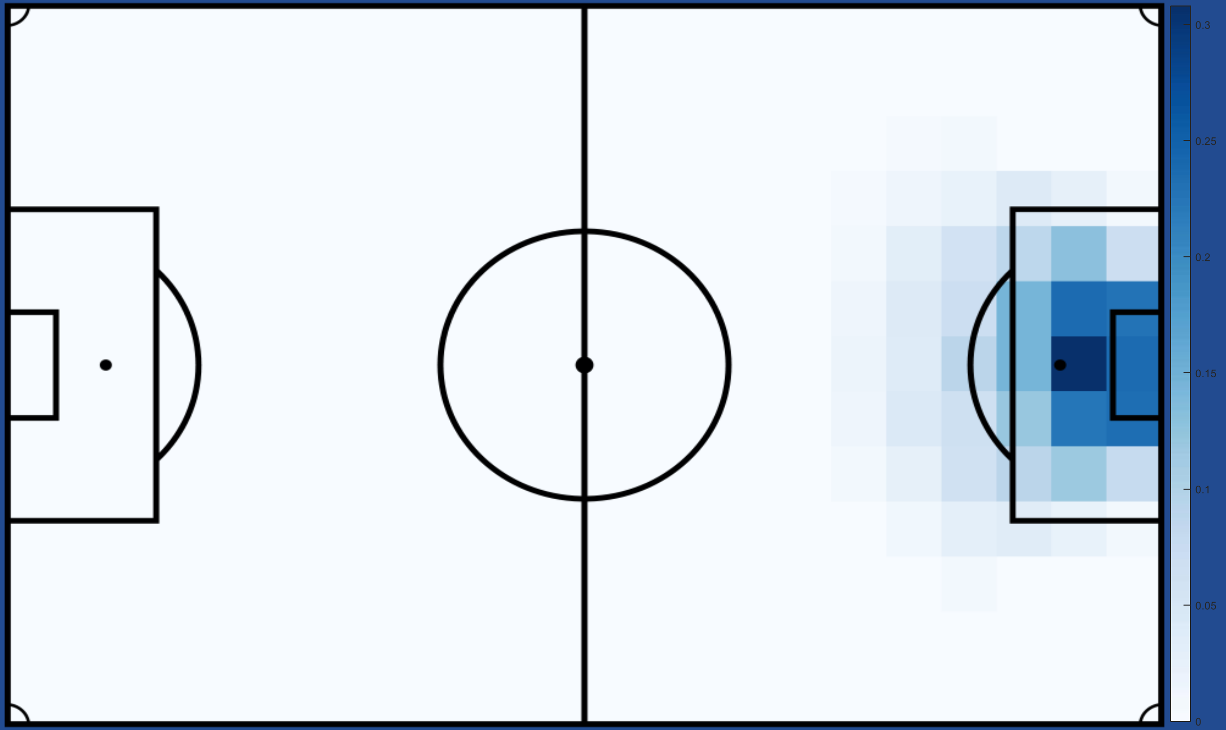
Dynamics:

- xG
  - Scoring probability
- P_{MS} or P_{ML}
  - Move to new location
- P_{TOS} or P_{TOL}
  - Move to new location following a turnover

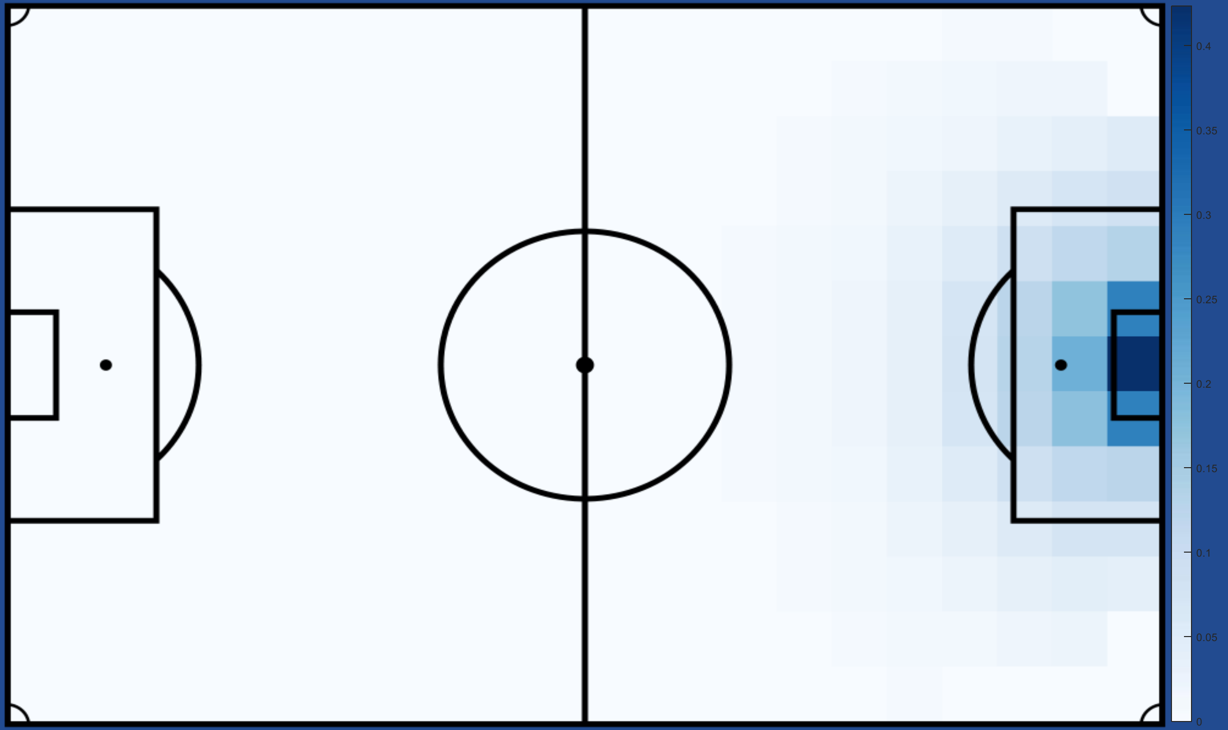
The minus sign denotes loss of possession!

Parameters of the extended xT model

Shot probability

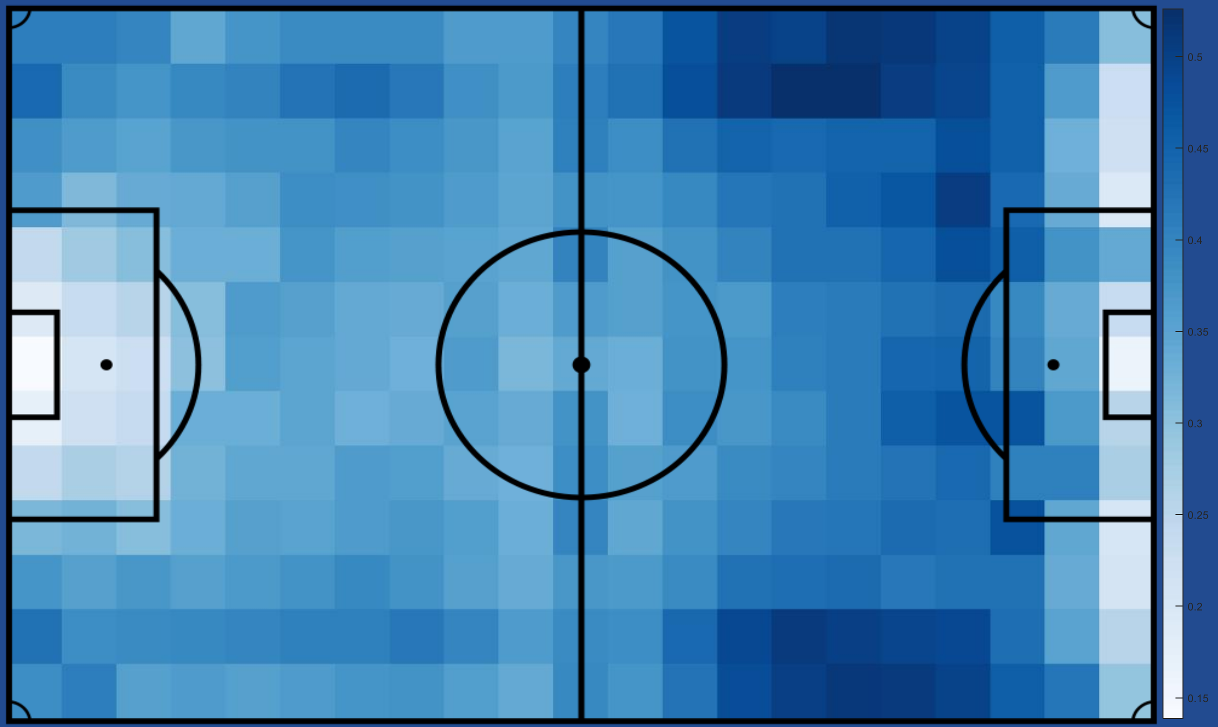


Expected goals

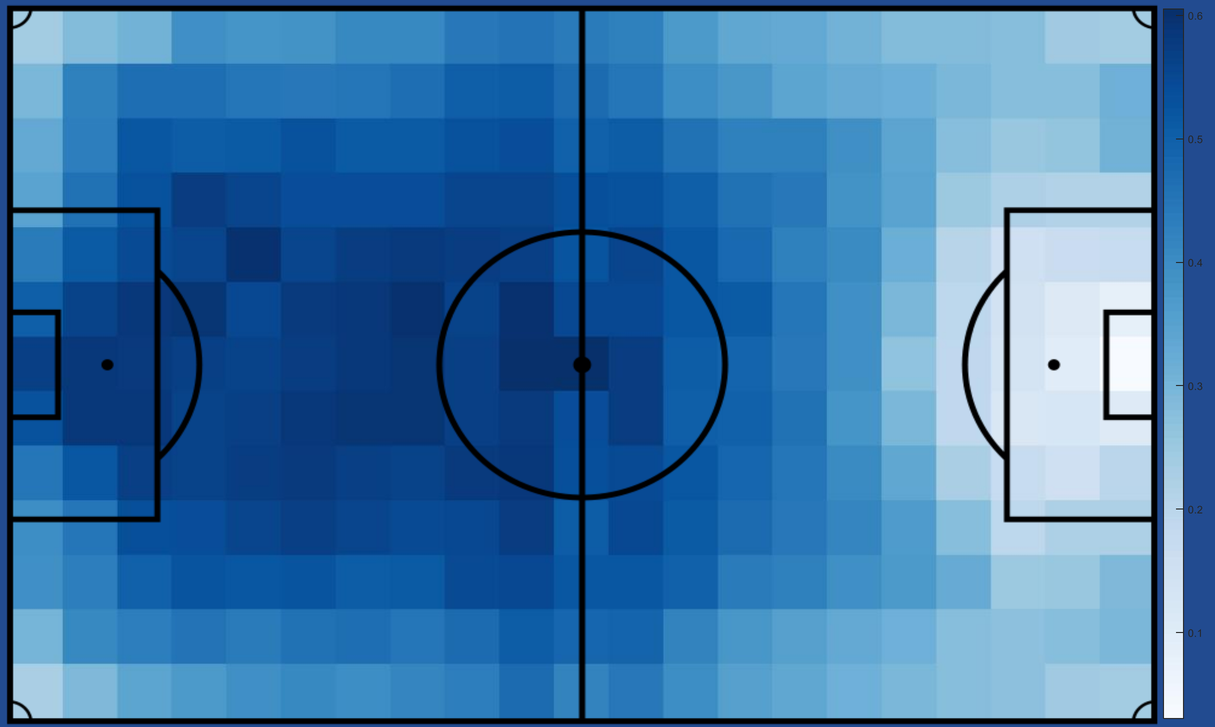


Move probabilities

Short move

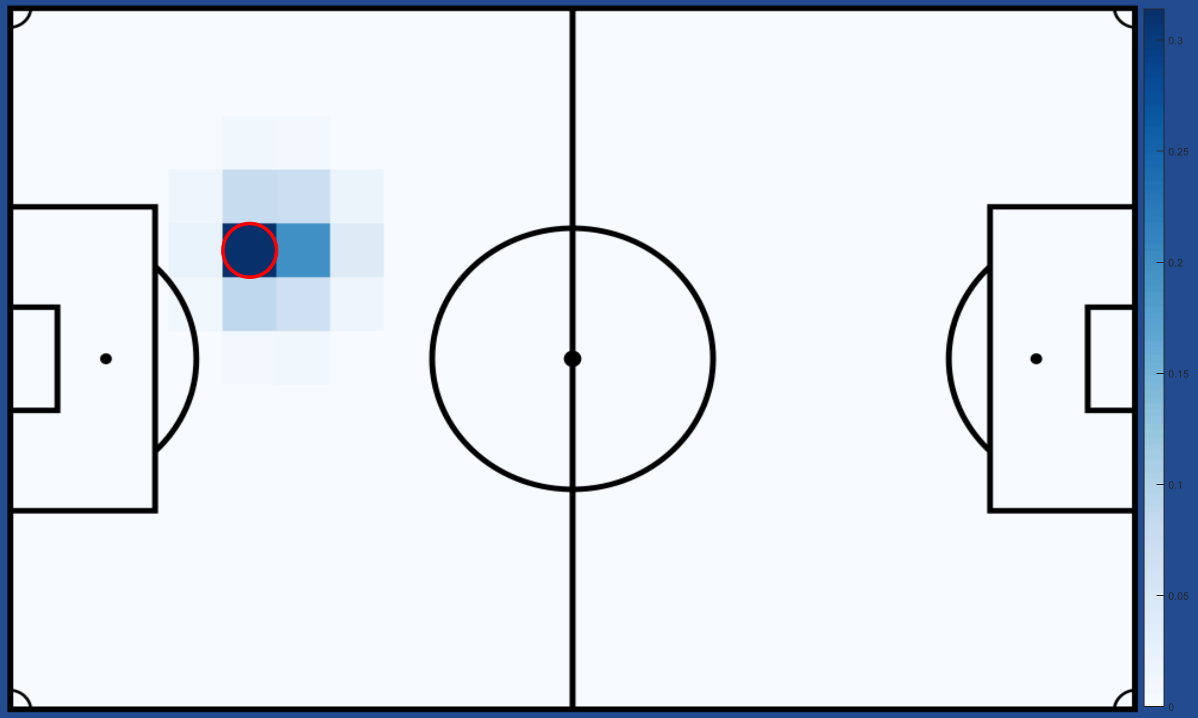


Long move (10+ m)

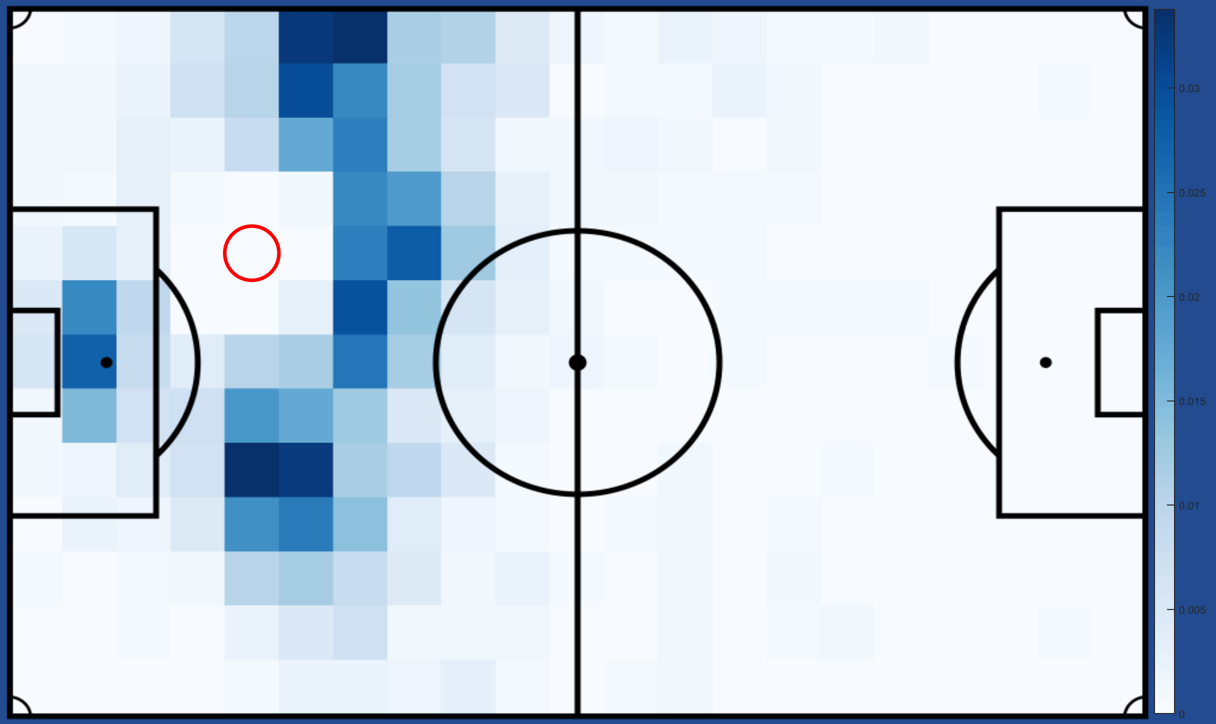


Example: Transition probabilities

Short move

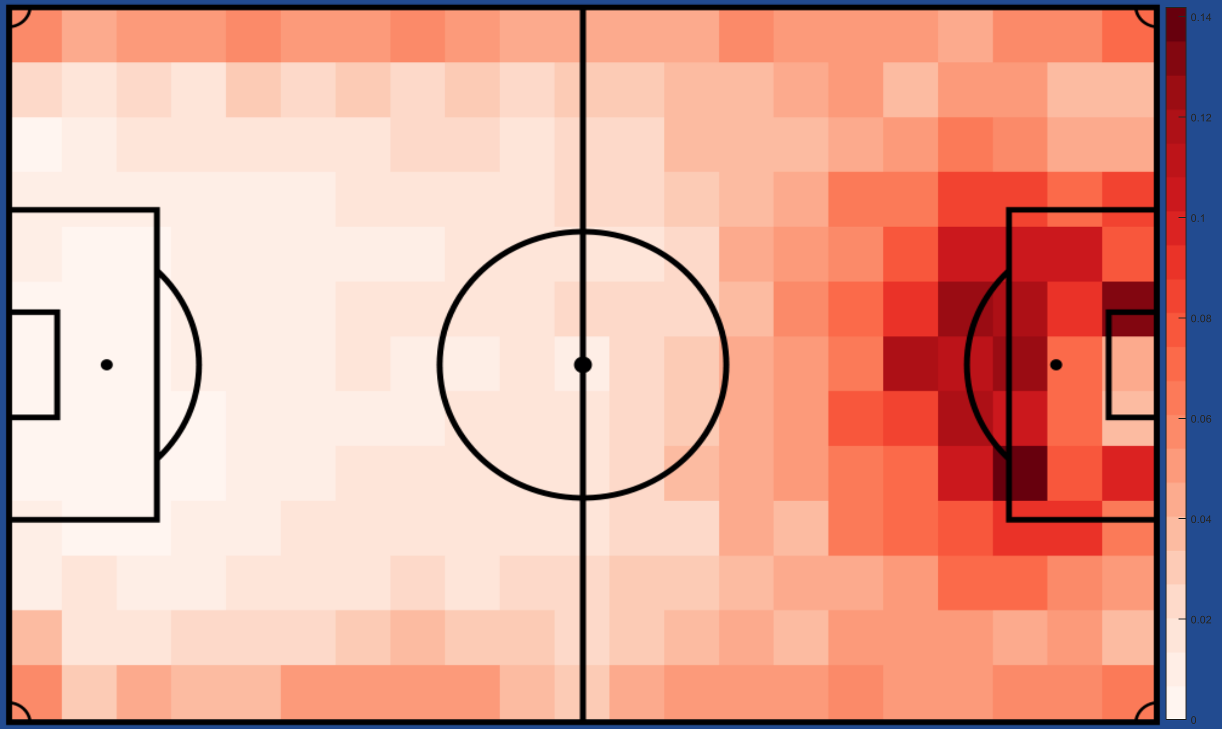


Long move (10+ m)

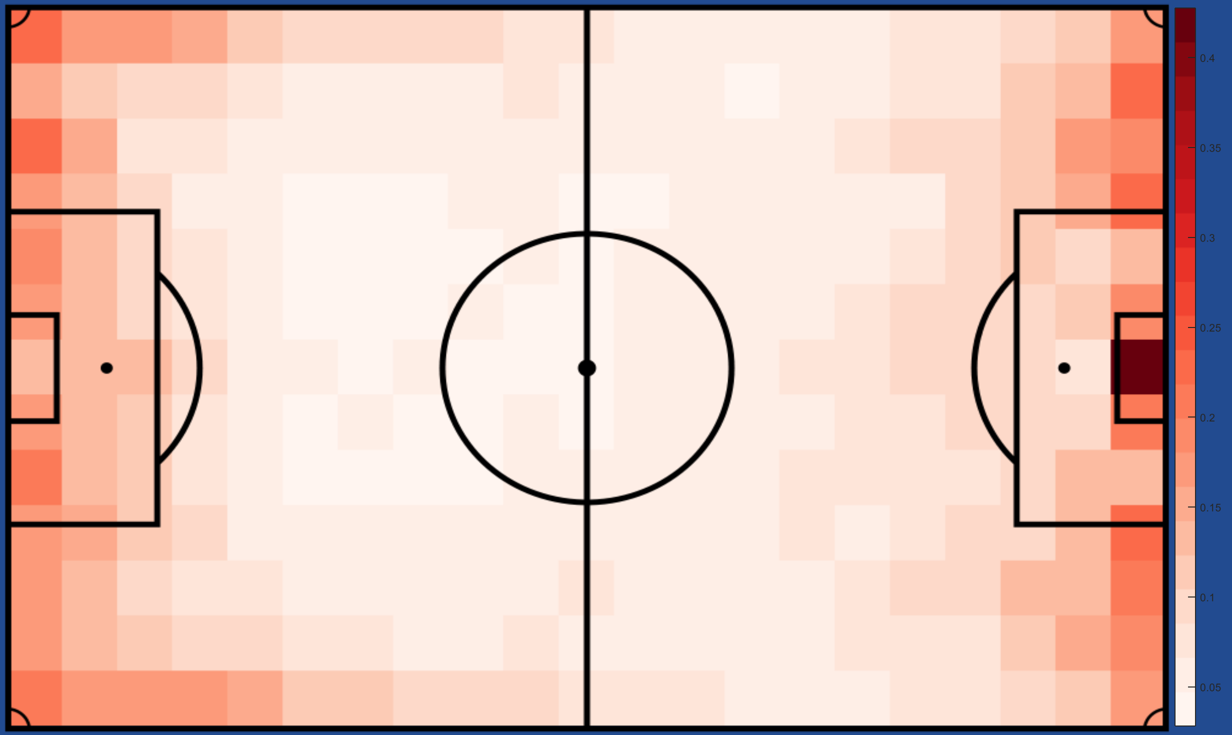


Turnover probabilities

Short turnover

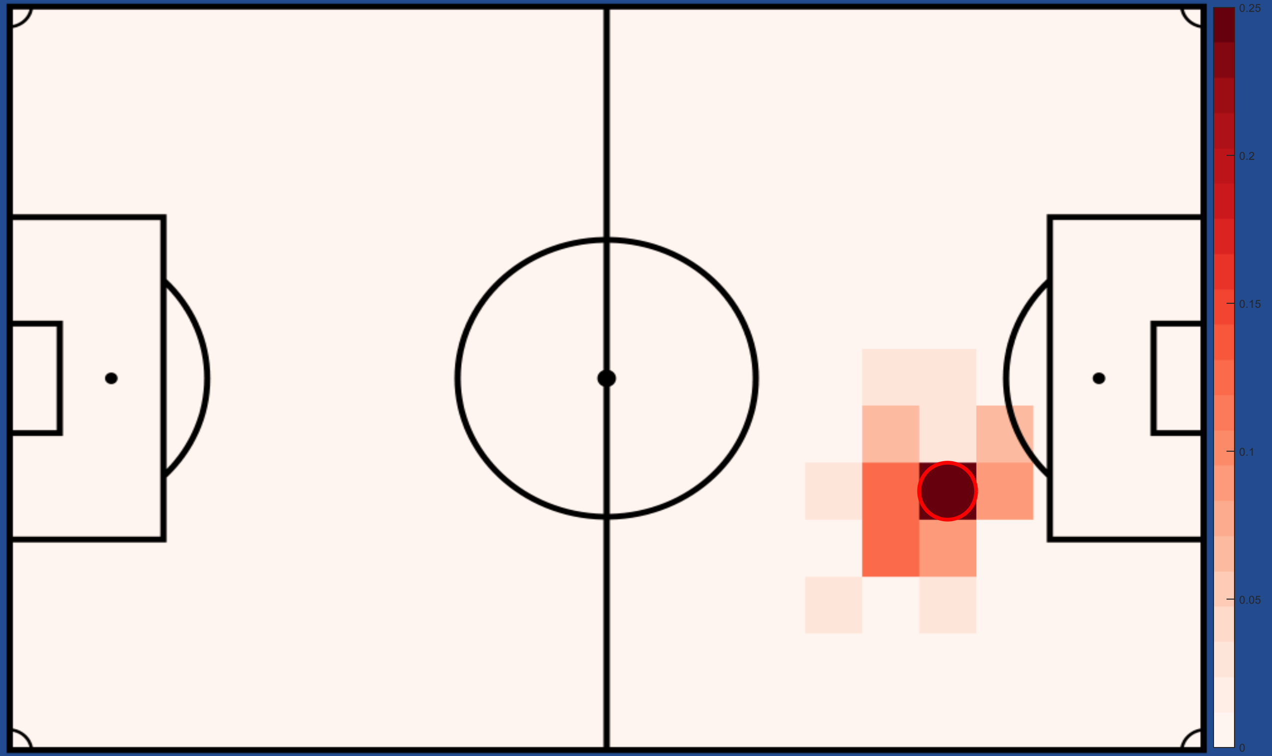


Long turnover (10+ m)

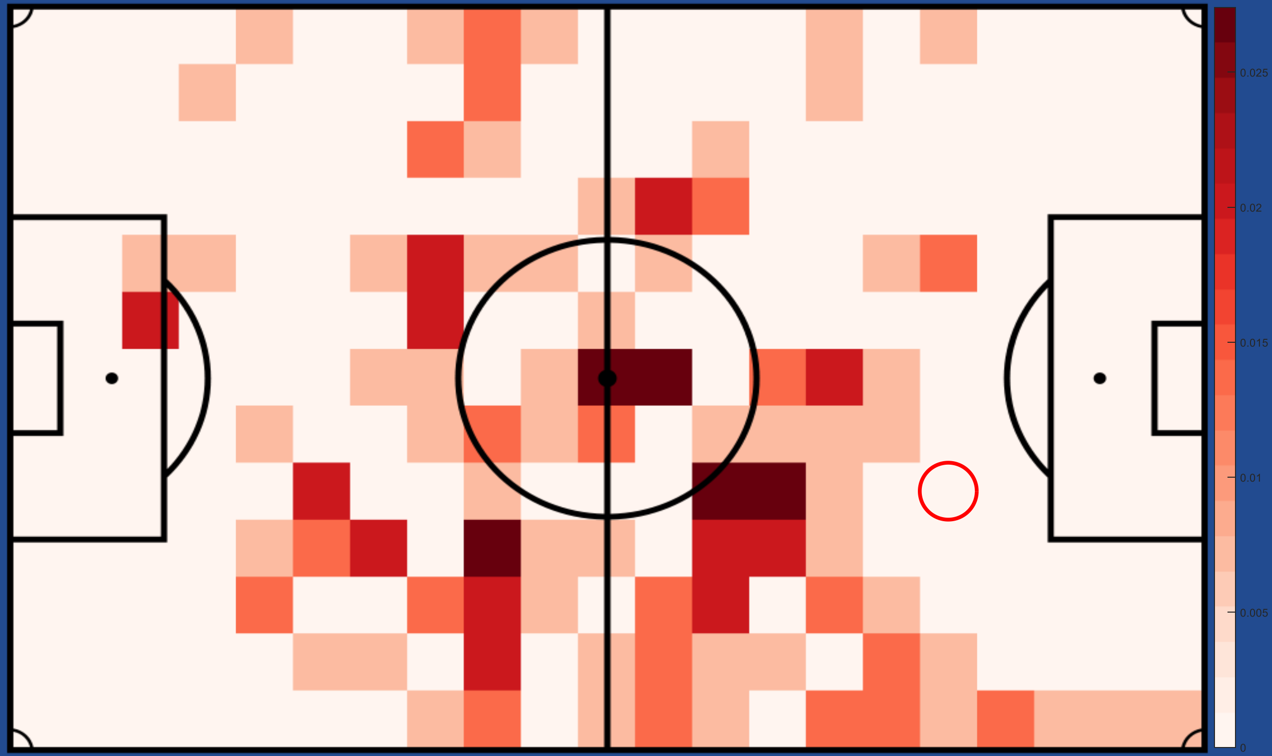


Example: Transition probabilities for (forced) turnovers

Short turnover



Long turnover (10+ m)



Solution of xT

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

$$= S \cdot xG + W \cdot xT$$

Suddenly, the equation looks almost the same as the Singh's original xT model!

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

$$= S \cdot xG + W \cdot xT$$

$$= S \cdot xG + W \cdot (S \cdot xG + W \cdot xT)$$

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

$$= S \cdot xG + W \cdot xT$$

$$= S \cdot xG + W \cdot S \cdot xG + W^2 \cdot xT$$

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

$$= S \cdot xG + W \cdot xT$$

$$= (I + W + W^2 + W^3 + W^4 + \cdots) \cdot S \cdot xG$$

Solution of xT

$$xT = S \cdot xG + M_S \cdot P_{MS} \cdot xT + M_L \cdot P_{ML} \cdot xT + T_S \cdot P_{TOS} \cdot (-xT) + T_L \cdot P_{TOL} \cdot (-xT)$$

$$= S \cdot xG + (M_S \cdot P_{MS} + M_L \cdot P_{ML} - T_S \cdot P_{TOS} - T_L \cdot P_{TOL}) \cdot xT$$

$$= S \cdot xG + W \cdot xT$$

$$= (I + W + W^2 + W^3 + W^4 + \cdots) \cdot S \cdot xG$$

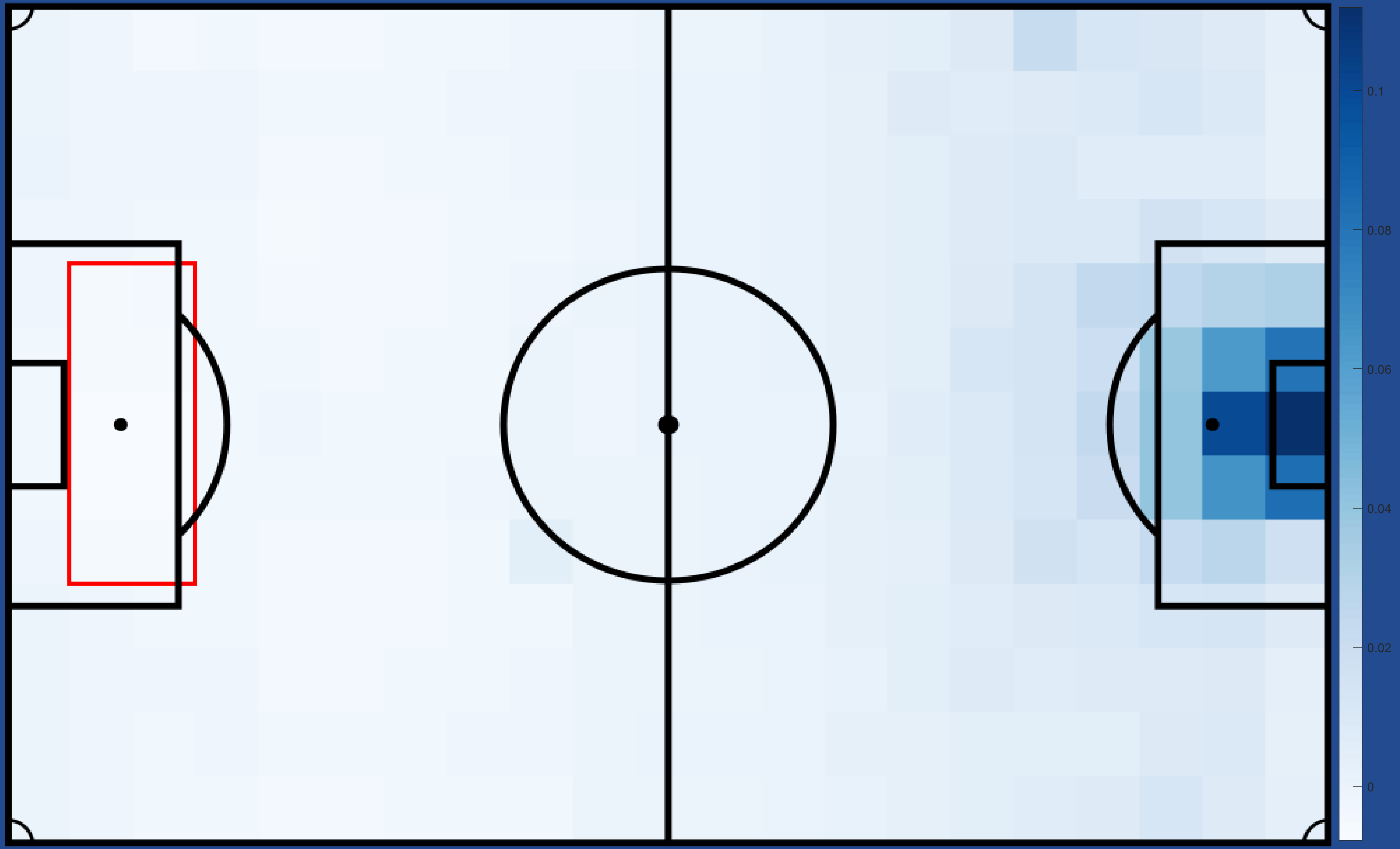
$$= (I - W)^{-1} \cdot S \cdot xG$$

The inverse matrix should exist, unless there are some extreme pathologies in the data.

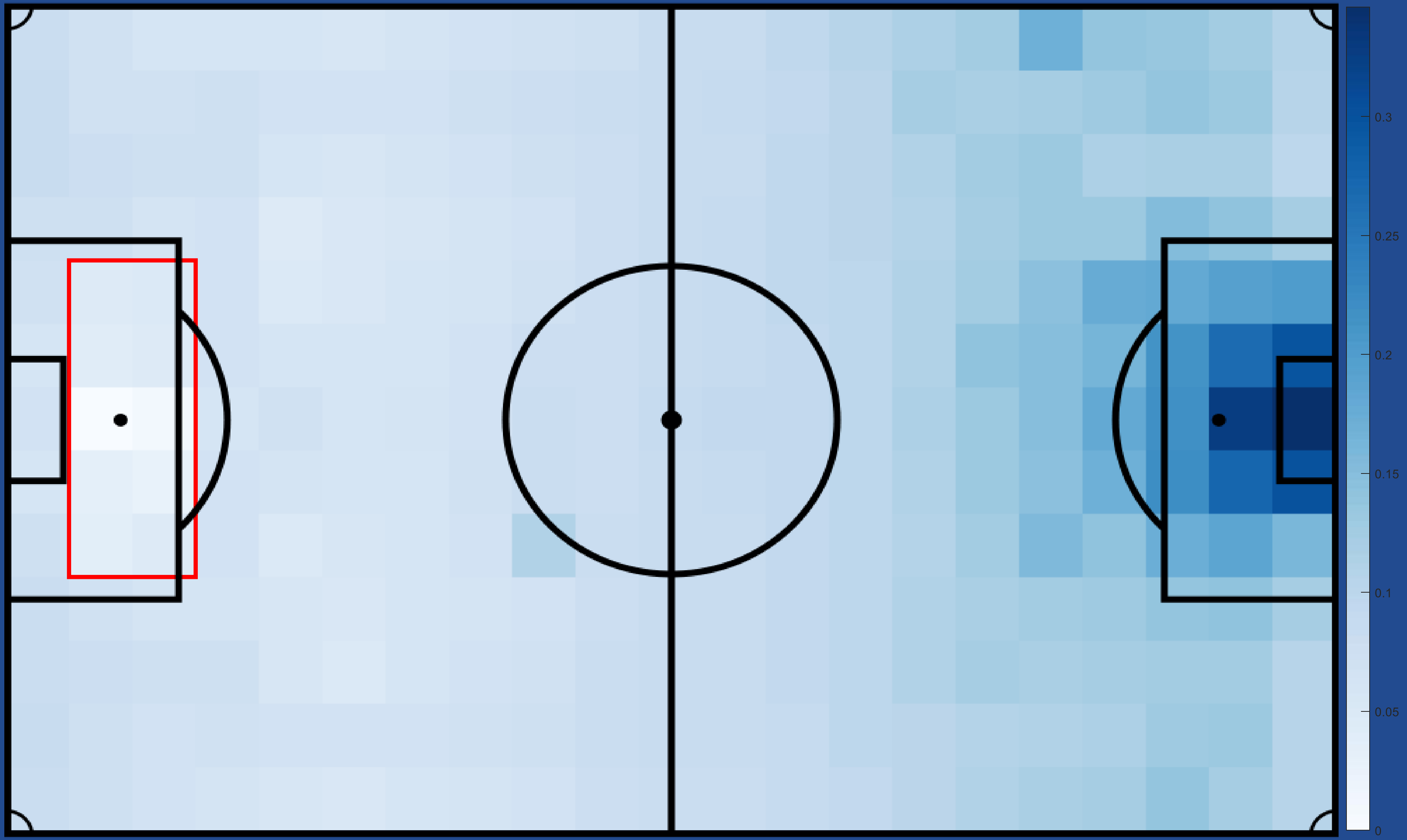
“The Soccer Equation”

$$xT = (I - W)^{-1} \cdot S \cdot xG$$

Extended xT

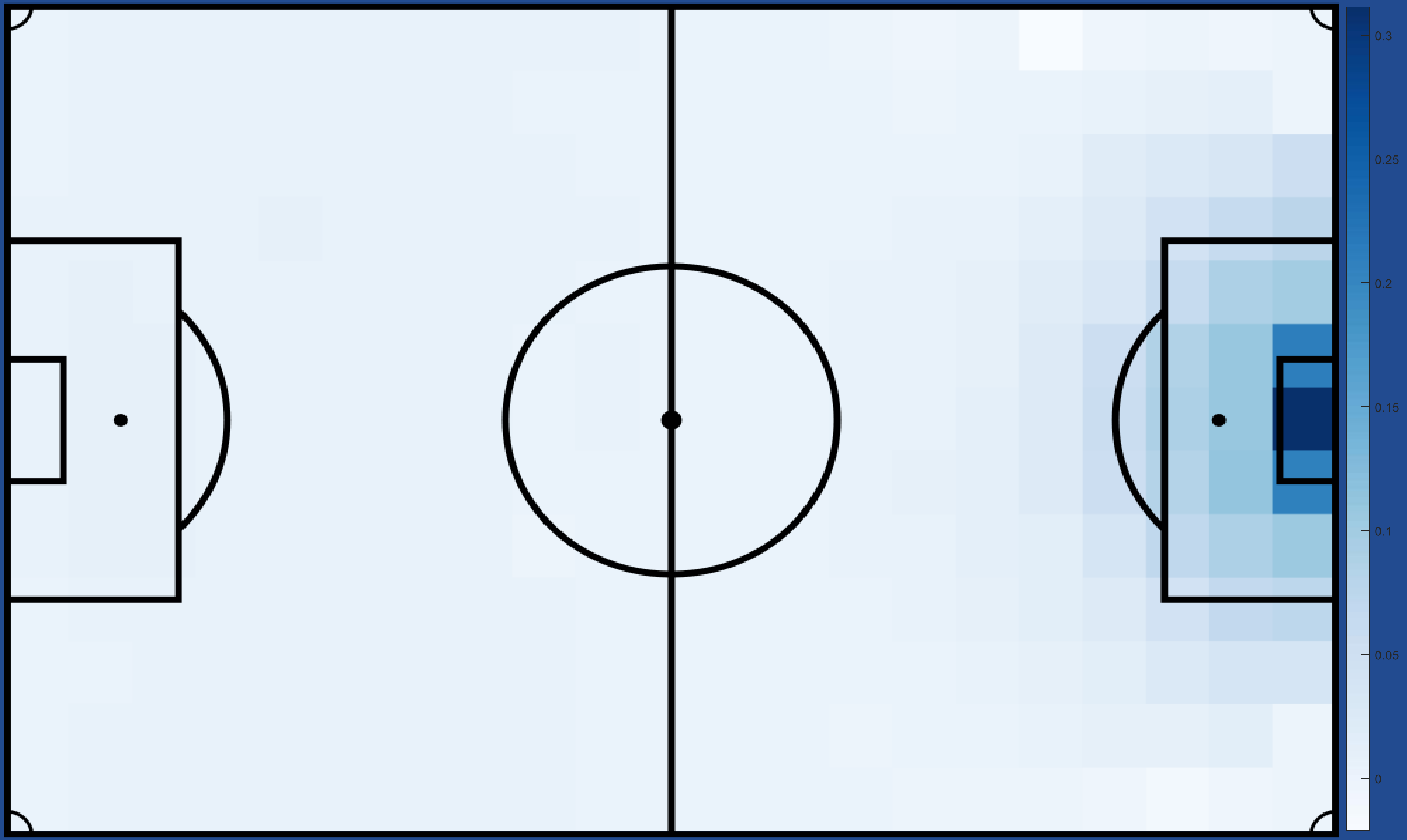


Extended xT (re-scaled)



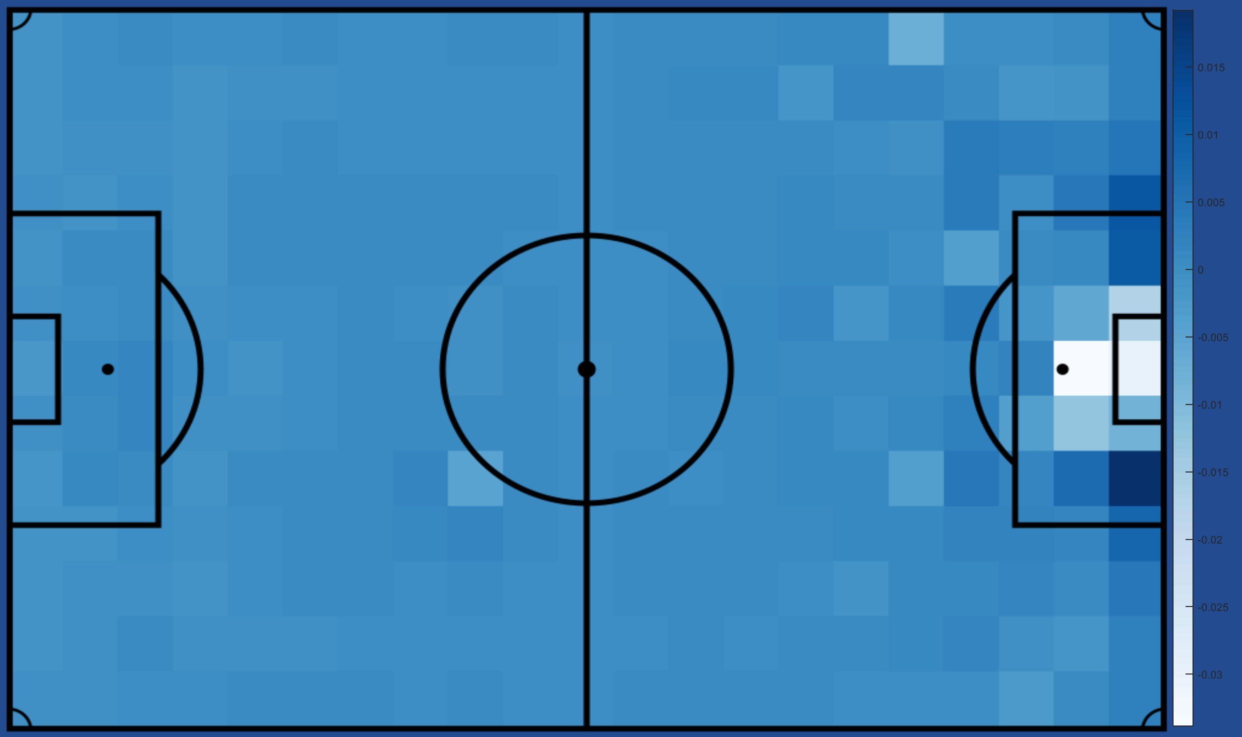
Expected value of events

Expected value of shot

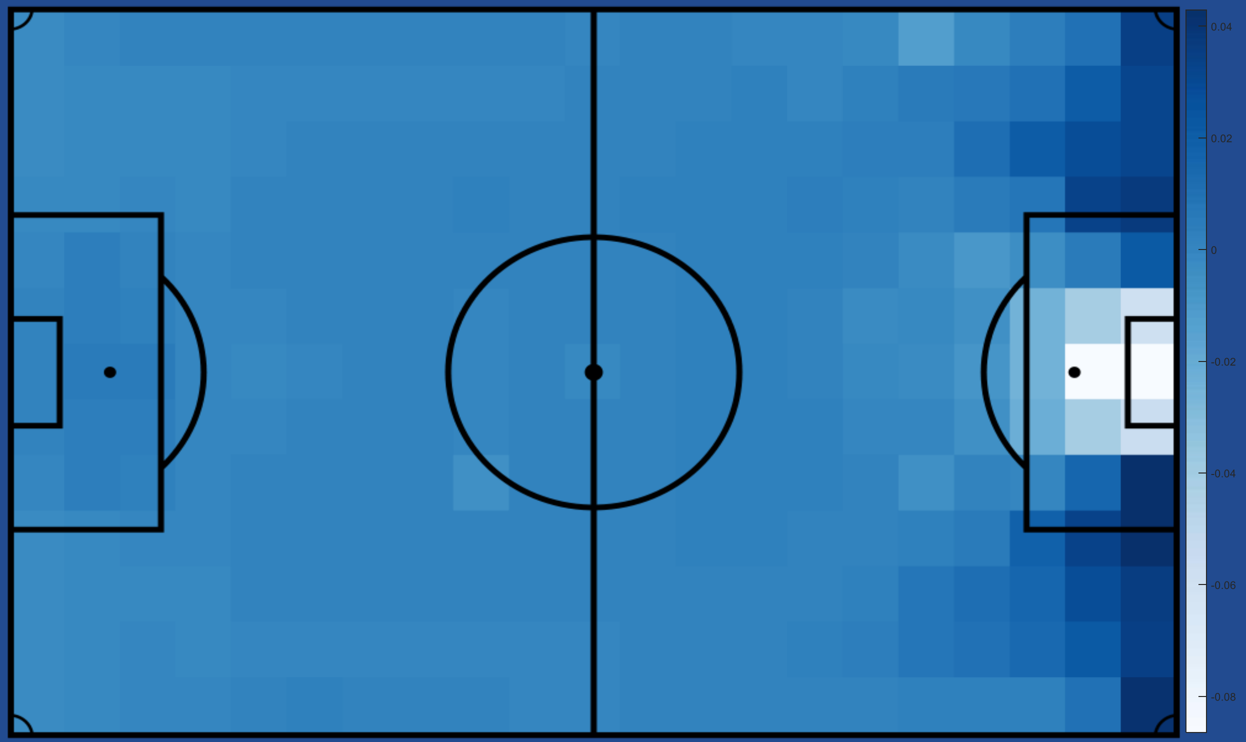


Expected value of move events

Short move

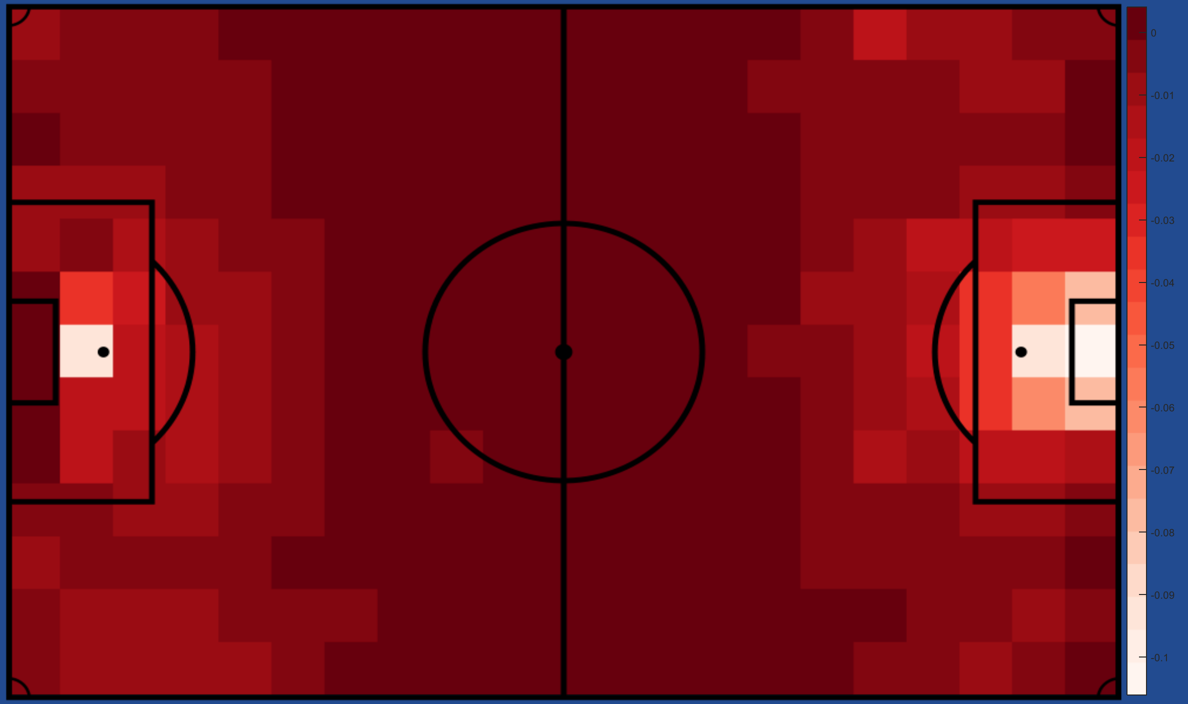


Long move (10+ m)

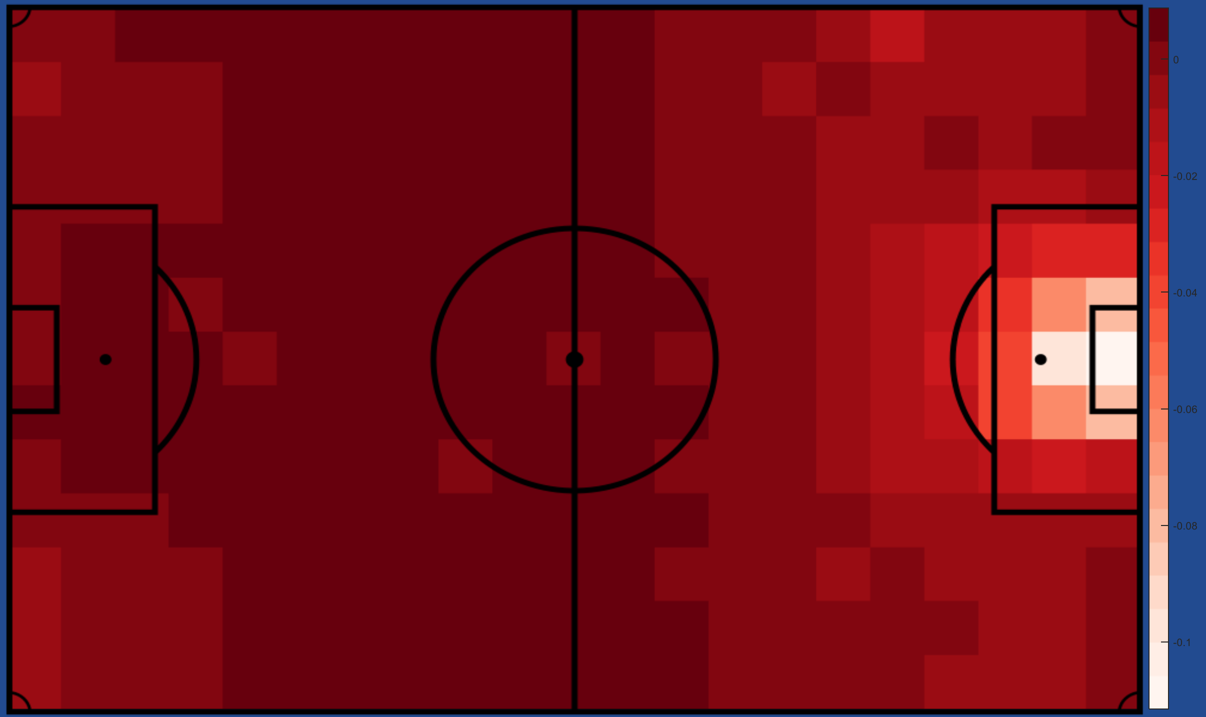


Expected cost of turnovers

Short turnover



Long turnover (10+ m)



Conclusions

- Extended xT models
  - Improved representation of soccer dynamics
- Two-way models include also opponent's scoring
  - Time horizon can be extended up to the next shot
- More detailed state transitions
  - Expected value/cost of events
- Simple and transparent model

  => Surprisingly believable results!

Future work

- Fine tuning for the extended xT model
  - Smoothing of probability values to reduce noise
- “xT added” models
  - Reward players properly for their actions
  - Include the expected value of actions into the calculation
- Include player locations into the xT model
  - Number of defenders between the ball and the goal
  - Defensive pressure (e.g., a defender within given radius)
  - For example, StatsBomb 360 data
- With larger state space, more advanced estimation techniques may be necessary

Any questions or comments?

Jirka Poropudas
@hamahakkimies
jirka.poropudas@gmail.com