

## Introduction

How can we measure offensive production beyond assists, goals, and expected goals in hockey?

Imagine a five-play possession that ended in a goal; the credited players would be the scoring player and up to two assisting players (last two touches). However, how are we crediting the other players who helped develop the possession? Which offensive action was the most impactful and significant to put the team in a position to score? These are the questions that our analysis answers.

We borrowed a soccer analytics concept called Expected Threat (xT), developed by Karun Singh to grade offensive production based on the field location at the start vs. the end of the play. In addition to adapting xT to hockey, we expanded it by adding time-remaining-in-period and turnover probability. We name the new metric: Net Expected Threat (nxT).

In soccer, the original xT model recognizes that at any point, a player possessing the ball has two options: to shoot (attempt to score) or to move the ball (dribble or pass). If a player chooses to shoot, the score-probability will depend on field location. If a player decides to move, there are multiple probable destinations, each with a different probability of success, depending on field location. xT generated (xTg) considers all possibilities and assesses by how much a player helped the team place the ball in a position to score, regardless of the actual outcome of the possession. To understand xT from a hockey perspective, interchange "ball" with "puck" and "dribble" with "zone-entry." One of the many benefits of this metric is the ability to divide credit. Take our five-play scoring-possession example. In that scenario, we can divide the xTg of each move action over the possession's total xTg. The result would be credit percentages to calculate which action or player had the highest impact.

In this submission, we explain how nxT expands xT to analyze a broader range of outcomes. Then, we describe the methodology behind nxT. Finally, we utilize nxT generated (nxTg) to grade offensive players for scouting purposes.



## Problem statement

Hockey analysts utilize standard metrics such as assists and goals to grade offensive hockey players. The analytics revolution brought metrics like expected goals xG, which have become a staple in the hockey analytics community. However, there is still a need to quantify each move action's impact to create a scoring threat. Expected Threat (xT) is a partial solution, but there are some limitations. xT only considers field position when assessing probabilities of shooting, scoring, attempting to move, and successfully moving the puck. According to our xG model, time-remaining-in-period is

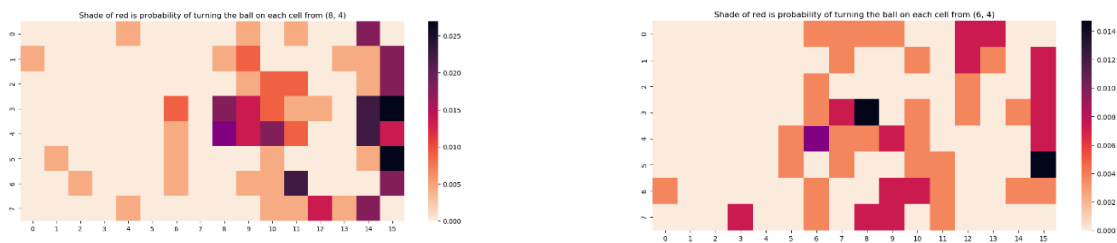
crucial when estimating score probabilities and significantly alters score probability regardless of field position. Also, xT does not account for the probability of losing the puck; therefore, we cannot measure the impact of unsuccessful move actions. As a result, xT can only quantify the effect of completed passes and successful zone-entries.

There is a glaring opportunity to generate a metric that quantifies successful and unsuccessful move actions and incorporates time-remaining-in-period. By successfully addressing these limitations, we would quantify the impact of every offensive move action. The new metric would facilitate the scouting process by quickly analyzing large amounts of data and identifying threat-generating players regardless of position.

## Solution

We decided to add time-remaining-in-period to an xG model, along with field location, to address one of xT's limitations. Then, we used the xG model to generate score-probability matrices and implemented them in our xT model. At this stage, our xT model is very similar to Singh's, except for adding an extra variable to generate the score-probability matrix.

Accounting for the probability of a turnover required a dynamic programming approach, similar to Singh's process when creating the move transition matrix  $T_{x,y}$  for his xT model. We split the field into 128 cells: 8 width cells and 16 length cells. Then, we estimated the probability of losing the puck at each cell, each probable turnover cell location, and the hypothetical opposing xT for each turnover cell location. By multiplying each cell's probability of turnover by their corresponding hypothetical xT and summing the results, we obtain Opposing Expected Threat (oxT).



We can account for the probability of turnover by obtaining the difference between the possession team's Expected Threat (xT) and the opposing team's Expected Threat (oxT). The resulting metric is called Net Expected Threat (nxT), which is our proposed solution.

$$nxT = oxT - xT$$

nxT tells us the level of threat that each team represents at every point based on field location (cell) and time-remaining-in-half. To quantify each move action's impact, we calculate the difference in nxT before and after each play.

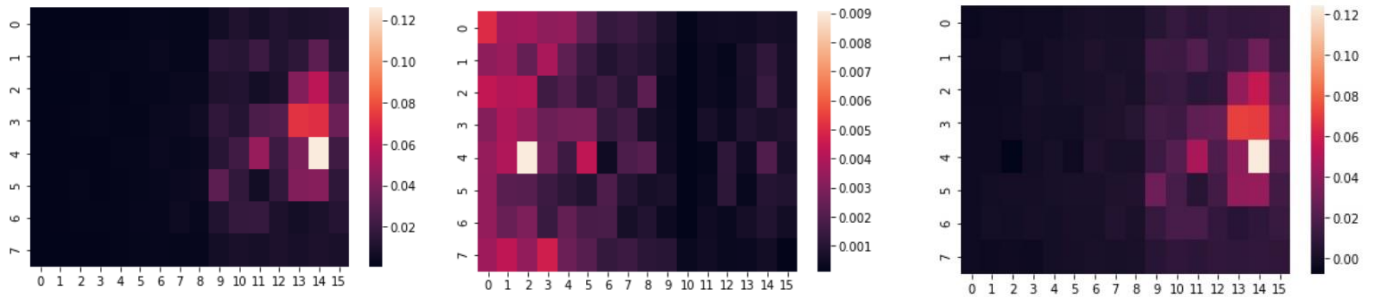
$$nxTg = nxT \text{ after the play} - nxT \text{ before the play}$$

By using oxT, we can measure the impact of a turnover by doing:

$$nxTg \text{ in turnover} = (oxT \text{ at the end of the play}) - nxT$$

nxT is the Expected Threat that team  $i$  had before the turnover (as the possession team). oxT is the Expected Threat that team  $i$  has after the turnover (now as the defensive team). This formula is different from equation b), which is the formula for nxTg when the play is successful.

As equation c) shows, oxT is needed to estimate the effect of turnovers. The previous methodology of xT was limited to equation b) and ignored unsuccessful actions: when the possession team loses the puck.



## Methodology

Before we describe our methodology, we would like to thank the socceraction team for their open-source module. We leveraged socceraction's xT model as a blueprint and codebase to build our own hockey xT module. We also thank Karun Singh for developing the methodology behind xT and nxT. This submission wouldn't have been [possible without his ground-breaking xT model and public-source code and articles.

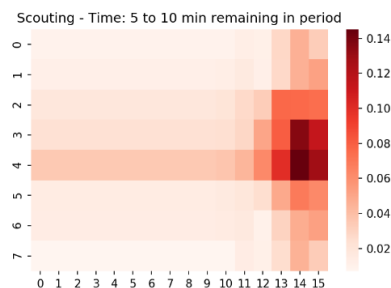
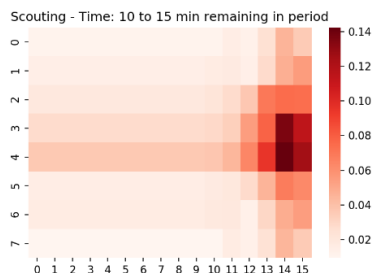
## Preparing Data

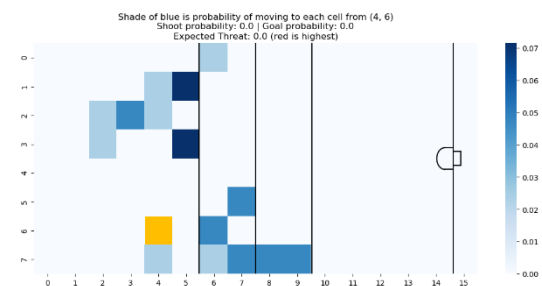
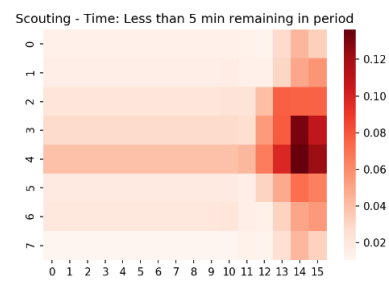
The first step is to generate the xT. The methodology in this section will be very similar to Singh's. To prepare the data, we need to estimate each play's start and end location. Also, we need to determine which actions will be considered move actions. In our model, the only move actions are pass and zone-entry. Finally, we need a success binary variable to determine whether the move action or the goal attempt was successful.

## Splitting field

The next step in data preparation is to split the field into zones or cells; this facilitates the analysis and avoids overfitting. We found a grid of 16x8 to be good enough. Every cell has a different:

- Move probability  $m_{x,y,t}$ : the probability of opting to move. Estimated by dividing the number of move actions over the number of total actions (move actions + goal attempts)
- Shoot probability  $s_{x,y,t}$ : probability for attempting to shoot. Following a similar approach to Move Probability
- Move transition matrix  $T_{x,y}$ : the probability of a player moving from the current zone to any other location. Each zone or cell has a matrix of length 16 x 8, totaling 128 different probabilities. For sample size reasons, this computation does not take time-remaining-in-period into consideration
- Goal probability  $g_{x,y,t}$ : Success probability when a player shoots from a given zone (x,y) during time-remaining-in-period t. Here is where our initial xT model differs from Singh's. Instead of calculating the score-rate per area, we developed an xG model including field x location, y location, Euclidean distance, and time-remaining-in-period. We created four different time-remaining bins: more than 15, between 10 and 15, between 5 and 10, and less than 5. Then, we created score probability matrices for each bin using our xG to predict score probability in different scenarios.





If the player moves, there are 128 possible destinations. Suppose the player moves to a destination  $(z,w)$ , which is the end location of a play that started in  $(x,y)$ . In that case, the expected payoff is the value of zone  $(z,w)$   $V_{z,w}$ . Since we have 128 possible destinations, we need to estimate the probability of moving to each area  $(z,w)$  and the expected payoff  $V_{z,w}$  of each potential destination. Finally, we proportionally weigh the payoffs. We can use the transition matrix  $T_{x,y}$  to estimate the expected payoff of moving the ball with the following equation:

$$\sum_{z=1}^{16} \sum_{w=1}^8 T_{(x,y) \rightarrow (z,w)} \times V_{z,w}$$

$$V_{x,y} = (s_{x,y} \times g_{x,y}) + (m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y) \rightarrow (z,w)} V_{z,w})$$

4

$$\mathbf{xT}_{x,y} = (s_{x,y} \times g_{x,y}) + (m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y) \rightarrow (z,w)} \mathbf{xT}_{z,w})$$

### This is where dynamic programming comes in handy

For the equation to work, we need to know  $\mathbf{xT}$  of all possible locations ( $\mathbf{xT}_{z,w}$ ), and right now, that is not the case. To solve that, Singh proposes assigning  $\mathbf{xT} = 0$  to all cells ( $x,y$ ) and evaluate the formula **iteratively until convergence**. During each iteration, assess the new  $\mathbf{xT}$  for each zone using  $\mathbf{xT}$  values from the **previous iteration**.

Each new iteration will utilize  $\mathbf{xT}$  values from the previous iteration as  $\mathbf{xT}_{z,w}$ . Therefore, the more iterations, the higher possible number of actions before the shoot-attempt we are considering. In other words, with 10 iterations, we are looking at up to 10 moves ahead of each scenario!

### Accounting for time remaining in half

During our model, the average number of iterations was 10. An important distinction: since we are adding an extra feature to the model – time-remaining-in-period  $t$ – we need to run this process 4 times, to utilize the corresponding score-probability matrix to the equation. We utilize categorical variable to split the training set in four different bins for time-remaining-in-half and trained to model correspondingly.

The wonder of Net Expected Threat

## Findings/Results

Here we present our findings: best players, create tables, charts, how nxTg relates to goals scored or win probability.

*\* I am sure Evan can add something cool here \**

## Action points

Using this metric, we were able to grade thousands of players in the scouting league in offensive production regardless of their position. Every move action generates xT irrespective of how close the play happened to the opposing net and whether the possession ended in a goal or not. nxT can quantify the effect of both successful and failed move attempts; therefore, players making mistakes are penalized accordingly.

We invite teams and scouting departments to implement a version of this metric to boost their scouting efforts and provide an extra layer of information to their evaluation and decision-making process. We create an open-source python module available **here** to facilitate our metric's implementation.

We are presenting examples of visualizations to analyze players and teams using nxTg.

*\* I am sure Evan can add something cool here \**