

# Overview of Big Data Bowl 2023 Open Metric and Undergrad Finalists

Ron Yurko

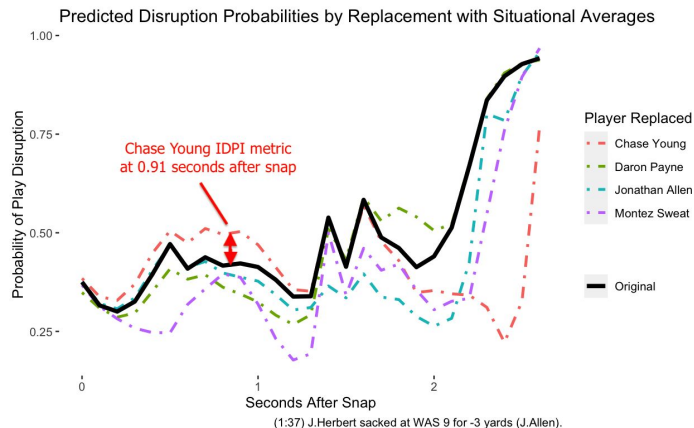
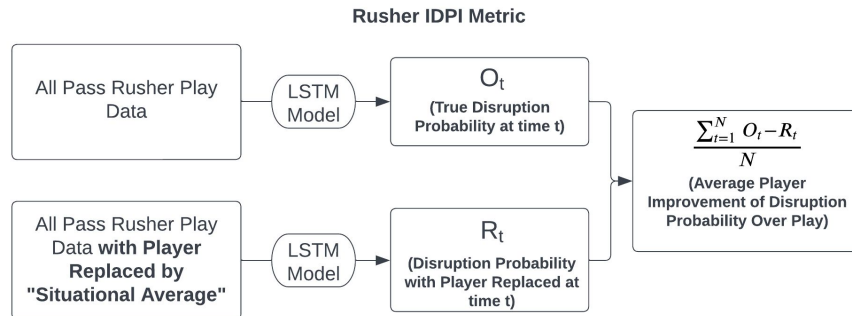
Honorable Mention I found interesting:

<https://www.kaggle.com/code/benjenkins96/causal-impact-of-offensive-linemen-on-pass-plays>

# IDPI: A Situational Metric for Pass Rushers

Nick Bachelder

- Instantaneous Disruption Probability Increase (IDPI) metric
  - Model team's probability of pass disruption (i.e., sack, hit, hurry) with LSTM
- LSTM model is trained on equally weighted sum of SoftMax outputs in the hidden layers of the model - trains the probabilities of a play disruption at all points in time during a play
- Replace individual level features with average values across plays in same start-of-play groups
- Compare observed prediction with player's observed features to prediction with averages

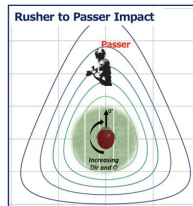
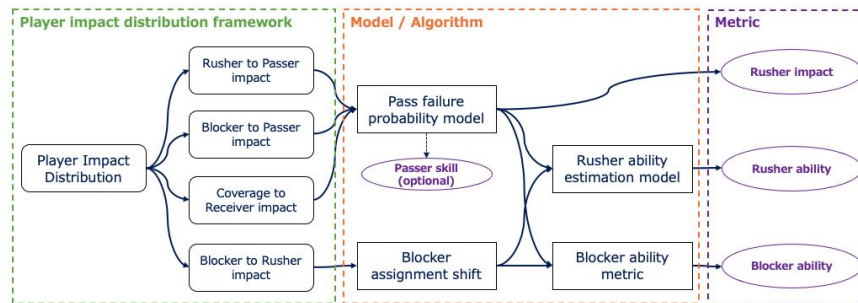


# Evaluate linemen using Player Impact Distribution

Sho Sekine, Nao Sekine

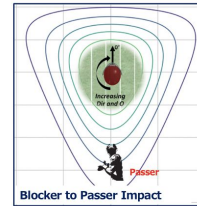
- Use Gaussian-Gamma mixture distribution to model players' impact on each other
  - Greater in the direction they are moving or facing, and NOT on their backs or in the direction they cannot see
  - Combines 1-way directionality provided by the Gamma with the spherically decaying characteristics provided by the Gaussian
- Then used Bayesian models from [item-response theory](#) to get player ratings?
- Extremely light on details and hard to follow...

Analysis overview



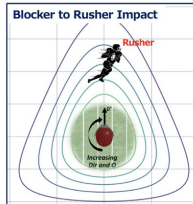
The contour lines indicate the degree of impact of the Rusher on the Passer.

This degree of impact is the **pressure** or the threat that the Rusher exerts on the Passer and can be viewed as a risk to the passer.



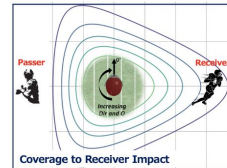
The contour lines indicate the degree of impact of the Blocker on the Passer.

A Blocker can bring **safety** to the rear by impeding the progress of the enemy in front of it. This means that the Blocker can generate a "**pocket**" in its back direction.



The contour lines indicate the Blocker's **attention** on the Rusher.

The Blocker can neutralize the Rusher by capturing the Rusher in front and impeding its progress.



The contour lines indicate the degree of impact of the Coverage on the Receiver.

Coverage is an interceptor and is a **threat** if it is in the path of the pass. Coverage has a sphere of influence toward the direction vector from the Passer to the Receiver.

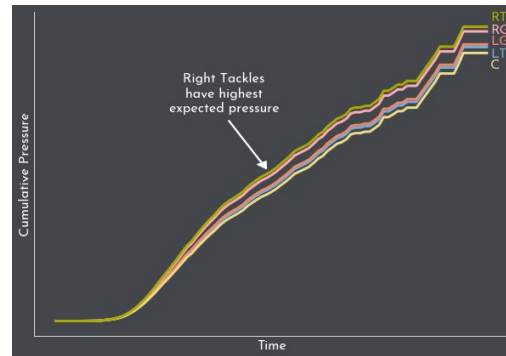
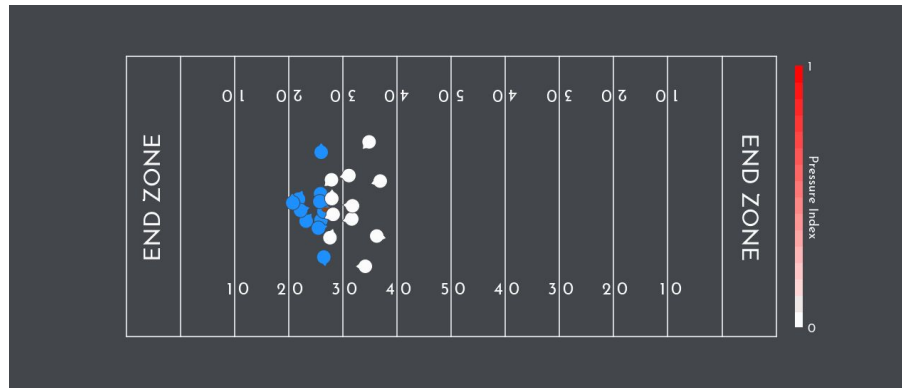
# Completions Added Through Suppression of Pressure

Vincent Karpick

Evaluates offensive linemen based on how many pass completions they created through preventing their opponents from pressuring the quarterback

1. Determining when opposition created pressure via a convolutional neural network (CNN)
2. Find expected pressure rates at each time for different positions with a penalized Cox model
3. Model expected completion rate on plays with and without pressure at various times
4. Combining Cox and completion model to find each player's completions over expected

Hungarian algorithm to assign blockers to pass rushers



# Between the Lines: How Do We Measure Pressure?

Hassaan Inayatli, Aaron White, Daniel Hocevar, University of Toronto

- Continuous Pocket Pressure (CPP) - model continuous time QB pressure probability
  - Using [Fernandez and Bornn](#) approach
- Survival analysis of CPP exceeding threshold

## 1 Frame Pocket Integrity as a Survival Analysis Problem

For each play, construct a table containing the status of the pocket at each frame. We consider a pocket to be "Dead" if the CPP > 0.915, and "Alive" otherwise.

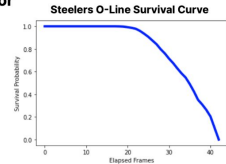
Frame	1	2	3	...	40
Status	Alive	Alive	Alive	...	Dead

## 2 Construct a Kaplan-Meier Estimator

Considering all plays in which a team is involved, estimate the survival probability at frame  $f$  using a Kaplan-Meier Estimator:

$$P(f) = \frac{\text{Alive}_f}{\text{Alive}_f + \text{Dead}_f}$$

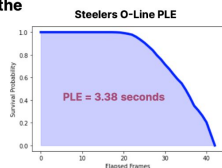
We can visualize the estimated survival curve by plotting the values produced by the Kaplan-Meier Estimator.



## 3 Apply Simpson Integration to Find the Pocket Life Expectancy

The pocket life expectancy (PLE) is the area under the Kaplan-Meier survival curve. We can compute this area by applying Simpson integration.

$$PLE = \int_0^{40} P(f) df$$



## 4 Create Separate Metrics to Evaluate O-Lines and D-Lines

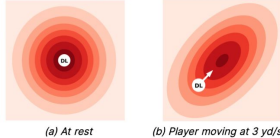
**OPLE:** Pocket life expectancy for offensive lines. A higher value indicates a better offensive line.

**DPLE:** Pocket life expectancy for defensive lines. A lower value indicates a better defensive line.

## 1 Defensive Player Influence Model

$$I_{dp} \sim N(\mu, \sigma)$$

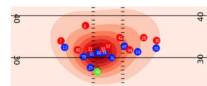
Defender's region of influence is affected by their velocity.



## 3 Defensive Team Influence Model

The defensive team's influence is the sum of each individual defender's influence.

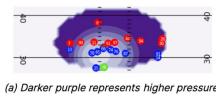
$$DTI \sim \sum_{dp \in D} I_{dp}$$



## 5 Continuous Pressure

Intuitively, our continuous pressure metric represents the percentage of the total player influence that belongs to the defensive team at a particular location.

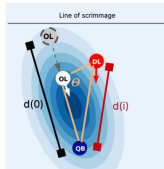
$$CP \sim \frac{DTI}{DTI + OPI}$$



## 2 Offensive Player Influence Model

$$I_{op} \sim \frac{d(i) \cdot \Theta \cdot N(\mu, \sigma)}{180 \cdot d(0)}$$

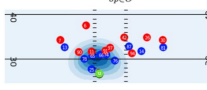
The normalized distance between defensive lineman and the quarterback as well as the angle between the QB, OL and DL is used to scale the offensive player's influence.



## 4 Offensive Team Influence Model

The offensive team's influence is the sum of each offensive player's influence.

$$OTI \sim \sum_{op \in O} I_{op}$$



## 6 Continuous Pocket Pressure

$$CPP \sim \frac{\iint N(\mu_{qb}, \sigma_{qb}) \cdot CP(x, y) dx dy}{\iint N(\mu_{qb}, \sigma_{qb}) dx dy}$$

Multiply the CP by a normal distribution centered at the QB. Then integrate over the domain of the field, and normalize the resulting CPP values so they are between [0, 1.0].



## 3. Measure Surplus Pressure:

- Subtract DPLE when the player is off the field from the DPLE when on the field

# Open SPACE: Spatial Survival Probabilities

Jay Sagrolikar, Paul Ibrahim, University of Chicago

SPACE: Survival Probability Above Contextual Expectation

- Response variable: indicator for whether a pressure occurs on the following frame of a play
- Model with random forests, NOT survival models...
- Construct features with Voronoi tessellations: partition the field into sections based on proximity to a player
- Model pressure at the start(?) then within play...

To compare the two, we calculate the final survival probability for an  $n$ -frame play via

$$P_s = \prod_{i=1}^n P_{s_i},$$

the cumulative product of survival probabilities for each frame. Since we train the model on *frames*, not plays (where probabilities would certainly affect each other), this is a well-defined metric. To define SPACE, we use the expression

$$\text{SPACE} = \frac{1}{n} \sum_{i=1}^n (P_{s_i} - C_p),$$

which gives us the survival probability for a team/player over expected.

