Hi Pawel,

Thanks for getting back to me and for inquiring about the big picture significance of the project. I’ve spent the past few days reconsidering my long-term goals and have come up with a procedure that I think could have legitimate strategic consequences. With your experience, I was hoping I could get your thoughts before I pitch a data request to AUDL.

The basic heatmap I attached in my previous email functionally is no more than a proof of concept. As you say, it is highly unlikely to reveal anything more than a higher score % when the disc is in the middle of the field and as it nears the endzone. As a result, I see it as a simple control on any dataset – if we take all the throws from a series of games, we should halt further analysis if this pattern is not seen.

From here I’ve developed a process for separating data into throws based on other variables within the dataset (personnel in action, force direction, offensive/defensive sets if that data is available). My initial plan was to create separate heatmaps for different game circumstances depending on combinations of these variables and compare their relative contours. I still think that this could result in some interesting visualizations, but I certainly would not expect anything groundbreaking to come out of it.

The fundamental issue with this approach is given a heatmap, any horizontal cross section, be it 50 yards from the endzone or 80, will have roughly the same distribution. In order to achieve the ‘spirit’ of a field heatmap, we must be able to differentiate the value of reaching a zone on the field after a cross field huck that rips the defensive set, as opposed to reaching that zone after a swing.

I think that the solution to this is the full realization of the potential of this data. I propose that we separate each zone into samples conditioned on the origin zone of the previous throw and calculate score % for each. Essentially, we should calculate how often a throw 20 yards from the endzone coming off a 60-yard huck, with the defense scrambling to catch up, eventually result in a goal, as well as how often throws from that position after a short under turn into goals. In this way we can produce an accurate estimate for the value of any throw on the field. (Ideally, we would condition on multiple previous throws but the data necessary for this may never be available)

An effective implementation of this method could have huge significance in terms of strategic development. Essentially, we can compare the relative value of attacking any two zones one the field from any position. Taking it a step further, if we add in completion probabilities for each throw (perhaps a ways from being analytically calculated, but estimates will suffice), we can begin to apply statistical decision theory to in game strategy and truly optimize the choices made. Down the road this can all be supplemented by applying the separation methods described above, for example selecting for score probabilities against a certain class of defensive set.

As I currently see it, the biggest challenge to achieve this final step is the quantity of data available. By my rough estimates there were around 75,000 throws logged in the past AUDL season. If we assume a uniform distribution of throw locations and require 100 throws in each zone, that allows for the field to be broken into just 27 zones. Performing this analysis is impossible using a single team’s dataset. I think that this still has value, but it may take a couple years (3 in order to use ~10 yd x ~10 yd zones) to reach full potential.

Apologies for rambling – if nothing else this has given me the opportunity to work through some ideas. If you get the chance to read through and have any thoughts, please let me know!

Thanks,

* Hiro