

Appendix 5.1 – Brief description of processing steps

The following describes the processing steps applied to these metrics, resulting in the 28 “raw-score” metrics and 26 “difference-score” metrics included in the Random Forests analyses (see Table 1).

1. Matches that resulted in a “Draw” (most likely resulting from automatically saved and uploaded replay files from matches that did not take place due to one player not connecting to the server, as draws cannot otherwise occur in Rocket League) were removed
2. Matches with a duration < 150 seconds to avoid overestimation of time normalised data
3. For the remaining matches (30,639 total), match length was calculated five different ways per match (ML1-ML5), using the following five different combinations of metrics that, when combined, covered the entirety of the match:

$ML1 = \text{time spent at slow speed} + \text{time spent at boost speed} + \text{time spent at supersonic speed}$

$ML2 = \text{time spent on the ground} + \text{time spent low in the air} + \text{time spent high in the air}$

$ML3 = \text{time spent in front of the ball} + \text{time spent goalside of the ball}$

$ML4 = \text{time spent in the defensive half} + \text{time spent in the offensive half}$

$ML5 = \text{time spent in the defensive third} + \text{time spent in the neutral third} + \text{time spent in the offensive third}$

The average (MLi) of ML1-ML5 was used as rounding discrepancies in the downloaded data could result in variance of up to 0.05sec depending on the combination of metrics used. Further, this average value varied by 3.45 ± 6.96 seconds for different players within the same match, hypothesised to resemble the discrepancy in initial match loading time (commonly seen as players can play one another on different platforms), however this is uncertain. Given this small discrepancy, we calculated the mean match length for a given match from the match lengths calculated for each player, and then

assigned that value to both players within the match (ML; see equations below).

$$MLi = \frac{ML1+ML2+ML3+ML4+ML5}{5}$$

$$Match\ Length\ (ML) = \frac{MLi(player)+MLi(opponent)}{2}$$

The calculation of ML described above was performed for all matches and resulted in an overall average all match duration of 360.45 seconds. We thus normalised all metrics not already expressed as a percentage to an average match length of 360 seconds, using the formula below:

$$Normalised\ Metric = \frac{Metric\ value}{ML} \times 360$$

4. After these calculations and time normalisation, we excluded the shooting percentage metric as its calculation (percentage of goals scored from all the shots taken) directly related to our outcome variable (GD).
5. *Average speed* was recalculated using the *total distance travelled* (TDT) metric and ML using the following equation.

$$Average\ speed = \frac{TDT}{ML} \tag{9}$$

6. Following the calculation of *average speed*, we excluded the *total distance travelled* metric from further analyses.
7. We calculated an additional metric called *true boost wastage*, using *total boost used while supersonic* and *total boost used* metrics. *True boost wastage* represents the proportion of “boost” used when a player is already travelling at max or near max speed. It is generally considered a measure of poor “boost” use, or wasted “boost” (Rocket Sledge, 2019; SquishyMuffinz, 2020).

Supplementary file 2 contains descriptions for “boost”, true boost wastage and all other metrics are described in greater detail. True boost wastage was calculated using the following equation

$$True\ boost\ wastage = \frac{total\ boost\ used\ while\ supersonic}{total\ boost\ used} \times 100$$

8. Data from five players (two Bronze, two Gold, one GC) were removed due to using no “boost” throughout a match, resulting in no *true boost wastage* metric value.
9. Upon identifying and calculating all the metrics to be included for subsequent analysis, we also calculated “difference-scores” for each metric (the difference between a given player and their opponent’s metric values). Difference scores were as per the following:
 - a. *Difference score for metric 'x' = players 'x' – opponents 'x'*
10. To eliminate the representation of matches twice in the same dataset, the data from one player in each match were removed (50%).
11. In many cases, certain players contributed many more matches than others (i.e., in the GC rank, one player contributed 812 matches). In order to maximise the independence of all the matches and data within a rank, only the ten most recent matches from the same in-game account name were retained (though the same account name could appear more than ten times as an opponent), with a maximum of one match retained for the same player-opponent combination. This strategy was chosen to both maximise the independence of the data while also maintaining a substantial amount (>45%) of matches for each rank (2527 Bronze, 7226 Gold, 7193 Diamond & 4643 GC*****).
12. Following this, in-game account name was removed to de-identify the data. Thus, from this point onwards, it was impossible to identify individuals from the data.
13. When one or more metrics were able to entirely explain the variance of another included metric, this was considered problematic from a collinearity standpoint, and so certain metrics were removed to mitigate this. Where this occurred, the “middle” metric was excluded (i.e., time spent in the neutral third was removed where time spent in the attacking and defensive thirds were retained; or time at boost speed was excluded where time spent at slow and supersonic speeds were retained). Furthermore, where positioning metrics were provided as “percentage spent in thirds” and “percentage spent in halves”, those expressed in “halves” were removed. Also, total boost collected/stolen from big and small pads were removed in favour of retaining ‘total boost collected/stolen’, as values for boost collected/stolen for big pads and small pads was also provided as a ‘count’ (which does not necessarily correspond to ‘total’ within Rocket League) as well. Overall, 24 metrics were removed in this way. These metrics are listed below.

- a. Raw and difference score metrics for: Total boost collected from big pads, total boost collected from small pads, total boost stolen from big pads, total boost stolen from small pads, total boost used while supersonic, time at boost speed, total time of powerslides, time low in the air, time in front of the ball, time in defensive half, time in offensive half, time in neutral third.
14. Finally, *shots conceded difference* and *demos taken difference* were removed, as these metrics mirrored *shots taken difference* and *demos inflicted difference* metrics respectively (see supplementary file 2 for elaboration).

Rocket Sledge. (2019). *How to Stop Wasting Boost in Rocket League* [Video File].

<https://www.youtube.com/watch?app=desktop&v=ykC-ozFSROg>

SquishyMuffinz. (2020). *HOW TO MANAGE YOUR BOOST LIKE A PRO PLAYER (TUTORIAL) / THE BEST TIPS FOR IMPROVING IN ROCKET LEAGUE*

<https://www.youtube.com/watch?app=desktop&v=eK3DLp-Yjwc>