Supplementary Material

Detailed outlier detection process and decisions

1. Small-sided game backs (SSG-B)

The frequency that a player was identified as potential outlier by all the mathematical methods previously described within the framework is presented in Figure 1. At a player level, six players were identified as potential outliers by at least one detection method (Figure 1). Specifically, player ids 14 and 31 were identified as potential outliers by six and four strategies, respectively (Figure 1). A closer look to the observations related to player id 14 demonstrated that high-speed (>61%) running distance values were appropriate based on the median of the data frame (i.e., zero meters). In addition, the majority of Stagno's TRIMP values were considerate as legitimate values, apart from one observation (i.e., observation id 231) which was characterised by a Stagno's TRIMP value of zero. Similarly, player id 31 showed reasonable values for high-speed (>61%) running distance and the majority of Stagno's TRIMP values. However, player id 31 was characterised by one observation with Stagno's TRIMP value of zero, and seven values lower than 200 AU. In addition, player id 29 exhibited four observations characterised by zero for Stagno's TRIMP, whereas player ids 37 and 38 showed one observation with zero values for Stagno's TRIMP, which represent some suspicious observations.

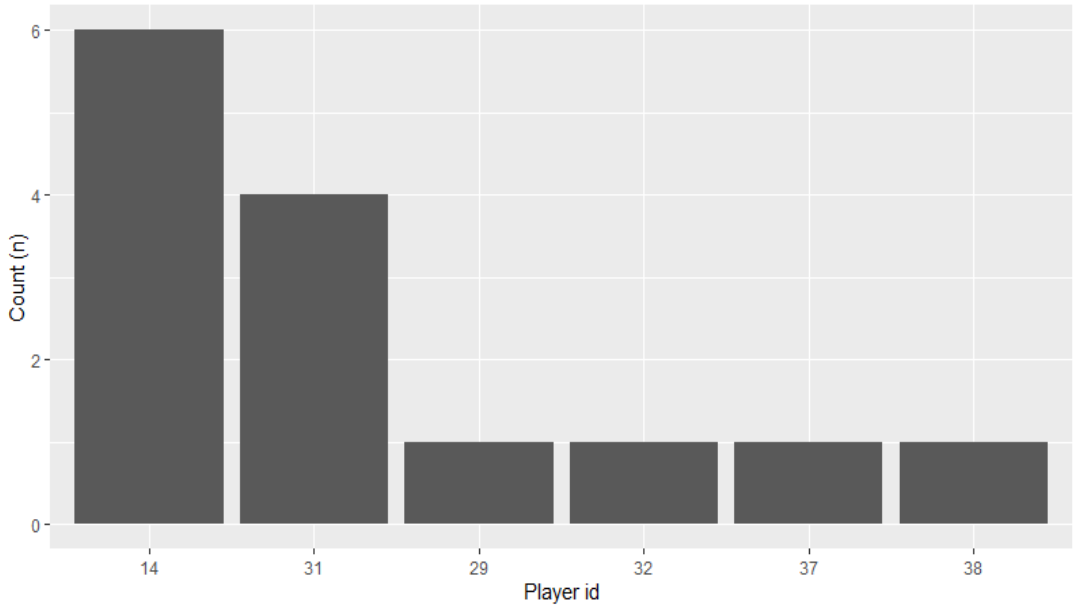

Figure 1: Player ids detected as potential outliers by multivariate strategies in SSG-B.

In terms of single observations, a total of 58 observations were detected as potential outliers, with the first 31 observations detected by more than three methods (Figure 2). All the observations demonstrated legitimate values for high-speed (>61%) running distance. The first four observations identified by the majority of the methods (i.e., ids 98, 231, 249, 41) were characterised by zero values for Stagno's TRIMP, thus further supporting the hypothesis that zero values may be a measurement error resulting from a detached heart rate sensor. In addition, considering the initial assumptions, inspection of all the 58 observations detected as potential outliers demonstrated that a total of 15 observations were characterised by Stagno's TRIMP values lower than 187.50 AU (i.e., eight zeros, 8.77, 141.95, 61.92, 179.01, 129.00, 171.01, 126.90 AU), corresponding to observation ids 98, 231, 249, 41, 265, 305, 431, 29, 179, 17, 5, 445, 158, 322, 356, respectively. In addition, to consider whether other observations with suspicious values for Stagno's TRIMP were not identified by the outlier detection strategies, the whole data frame was investigated, thus leading to the identification of other three observations characterised by Stagno's TRIMP values of 172.72, 182.48, 179.01 AU (observation ids: 175, 373, 445, respectively). Due to the suspicious data generating process of these observations, they were considered as error outliers. Consequently, a total of 18 observations were detected as potential outliers and labelled as error outliers. The strategy used to handle these outliers was their removal from the data frame.

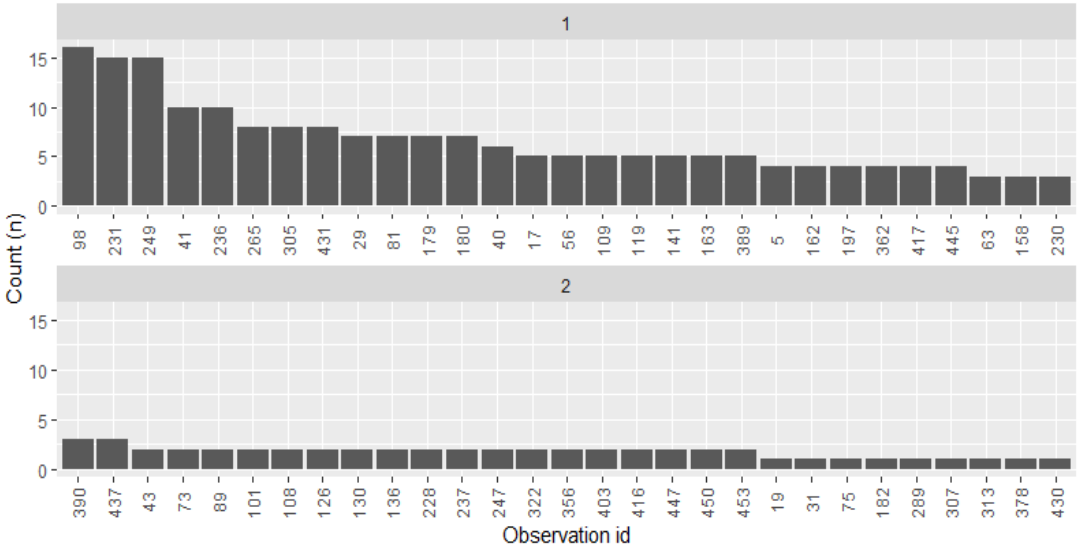

Figure 2: Observation ids detected as potential outliers by univariate and multivariate strategies in SSG-B.

2. Small-sided game forwards (SSG-F)

In SSG-F at a player level, player ids eight, 13, 15, and 19 were identified as potential outliers by more than one strategy (Figure 3). Among these player ids, high-speed (>61%) running distance showed reasonable values without including the maximum for the data frame (i.e., 16.22 meters). Conversely, player id eight showed one value for Stagno's TRIMP equal to zero, player id 13 showed one suspiciously low value (i.e., 120.63 AU), whilst player id 19 exhibited four suspiciously low values (i.e., 55.88, 93.98, 133.92, 162.56 AU). Player id 15 was characterised by legitimate values for Stagno's TRIMP.

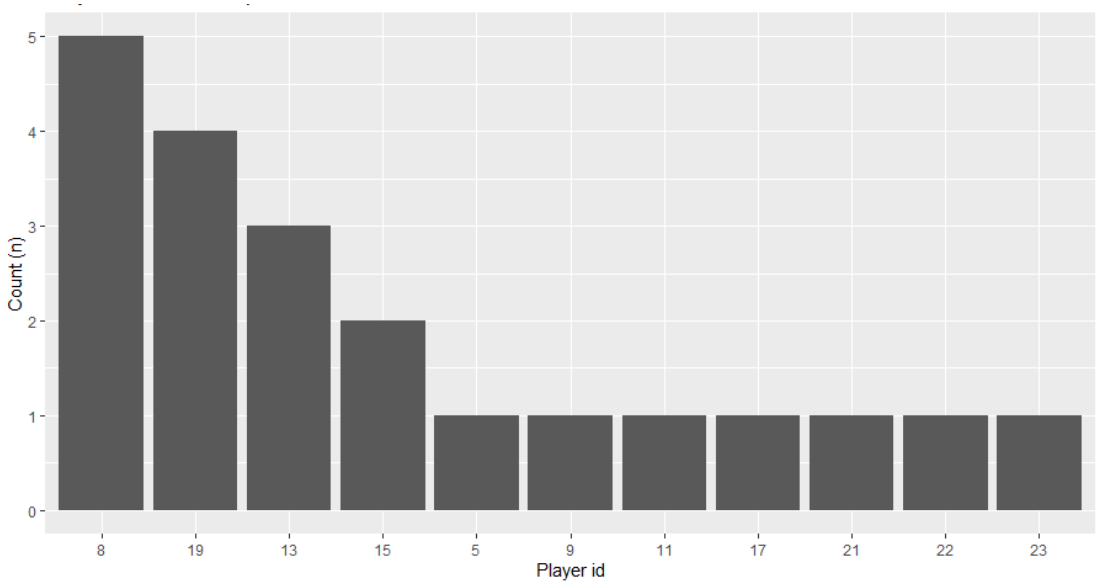

Figure 3: Player ids detected as potential outliers by multivariate strategies for SSG-F.

In terms of individual observations, 67 different observations were identified as potential outliers by all the outlier detection methods utilised (Figure 4). Among these observations, 21 were detected by more than five methods (Figure 4). These were characterised by legitimate values for high-speed (>61%) running distance, but some suspicious values were identified for Stagno's TRIMP. One observation was characterised by a Stagno's TRIMP value of zero (i.e., observation id 180), while five observations (i.e., observation ids 340, 503, 239, 300, 320) were characterised by suspiciously low values. A deeper investigation demonstrated that all the observations characterised by a Stagno's TRIMP value lower than 187.50 AU were identified as potential outliers by at least four methods. Consequently, seven observations (i.e., observation ids 180, 239, 266, 300, 320, 340, 503) were considered error outliers and removed from the data frame.

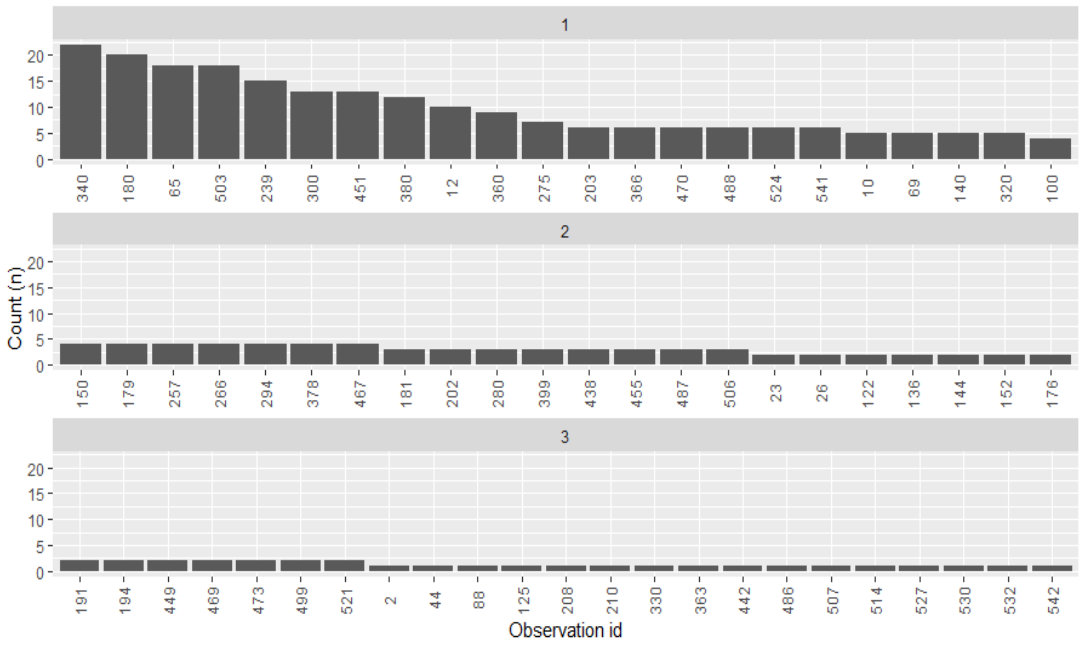

Figure 4: Observation ids detected as potential outliers by univariate and multivariate strategies for SSG-F.

3. Small-sided game backs forwards (SSG-BF)

Considering the results of all the methods together at a player level, player id 31 was identified seven times, followed by player id 27 identified five times, and player id 34 identified four times as potential outliers (Figure 5). A deeper investigation of the observations related to these player ids showed no Stagno's TRIMP observations that may have been the result of a measurement error, whilst a high value for high-speed (>61%) running distance (i.e., 44.02 meters) was observed for player id 31. However, this observation could be a legitimate value elicited by the constraints of the SSG-BF.

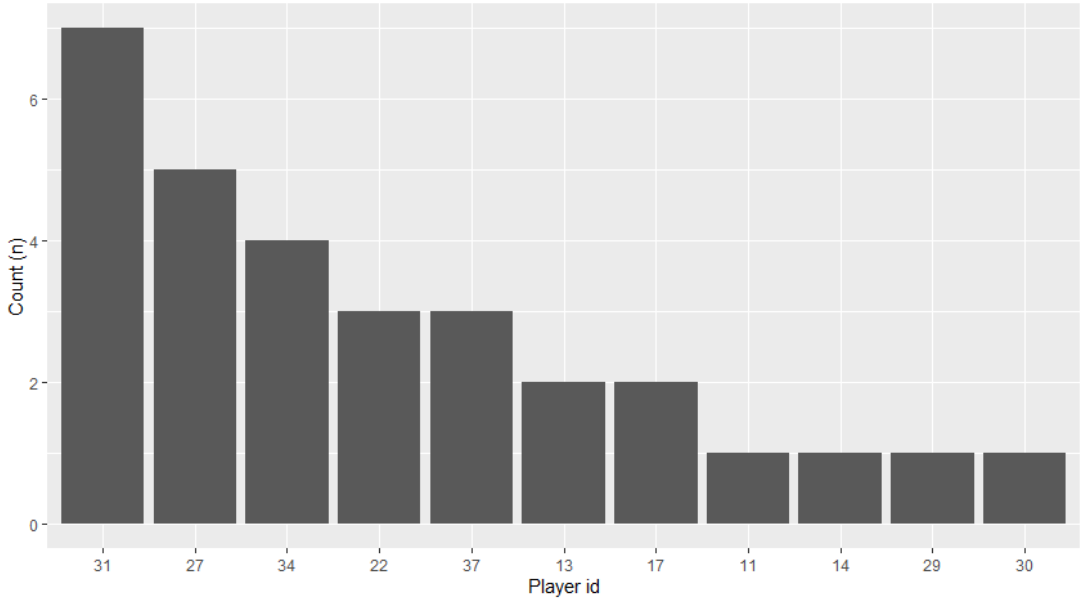

Figure 5: Players detected as potential outliers from the multivariate strategies for SSG-BF.

In terms of individual observations, the proposed framework identified 108 observations as potential error and/or influential outliers (Figure 6). An investigation of these observations demonstrated that 11 observations where characterised by zeros values for Stagno's TRIMP, thus supporting the initial concerns about possible measurement errors for Stagno's TRIMP due to misplacement of the heart rate sensor. In addition, another observation detected as potential outlier was characterised by the highest value recorded for high-speed (>61%) running distance (i.e., 54.95 meters). However, this value may have been generated by the constraints of the SSG, thus being a legitimate observation, or could have been generated by an inappropriate maximal speed used for that specific player. Nonetheless, the specific player was not identified as a potential outlier by the player id investigation and no further observations for that specific player were identified as potential outliers. Consequently, the observation characterised by the highest value for high-speed (>61%) running distance was not considered an error outlier, and it was maintained within the data frame.

Looking at the 108 observations identified as potential outliers, some suspicious low values for Stagno's TRIMP were identified. Based on the initial assumptions, Stagno's TRIMP values lower than 262.5 AU may indicate a possible displacement of the heart rate sensor. Specifically, 11 observations were associated with zero values for Stagno's TRIMP and six were characterised by Stagno's TRIMP values lower than 262.5 AU (i.e., 92.88, 132.40, 159.96, 167.88, 185.66, 205.99 AU). In addition, the data frame was investigated to identify further observations that may be characterised by a Stagno's TRIMP value lower than 262.5, but that were not detected by the outlier detection strategies. This led to identify two other observations characterised by low values for Stagno's TRIMP (i.e., 185.66, 250.59 AU), and

hence were labelled as error outliers. Observations identified as error outliers were removed from the data frame as they were not generated by the data generating process under investigation. This led to the removal of 19 observations from the data frame (i.e., observation ids: 5, 7, 37, 106, 143, 223, 224, 255, 259, 412, 421, 448, 464, 500, 554, 588, 622, 637, 656).

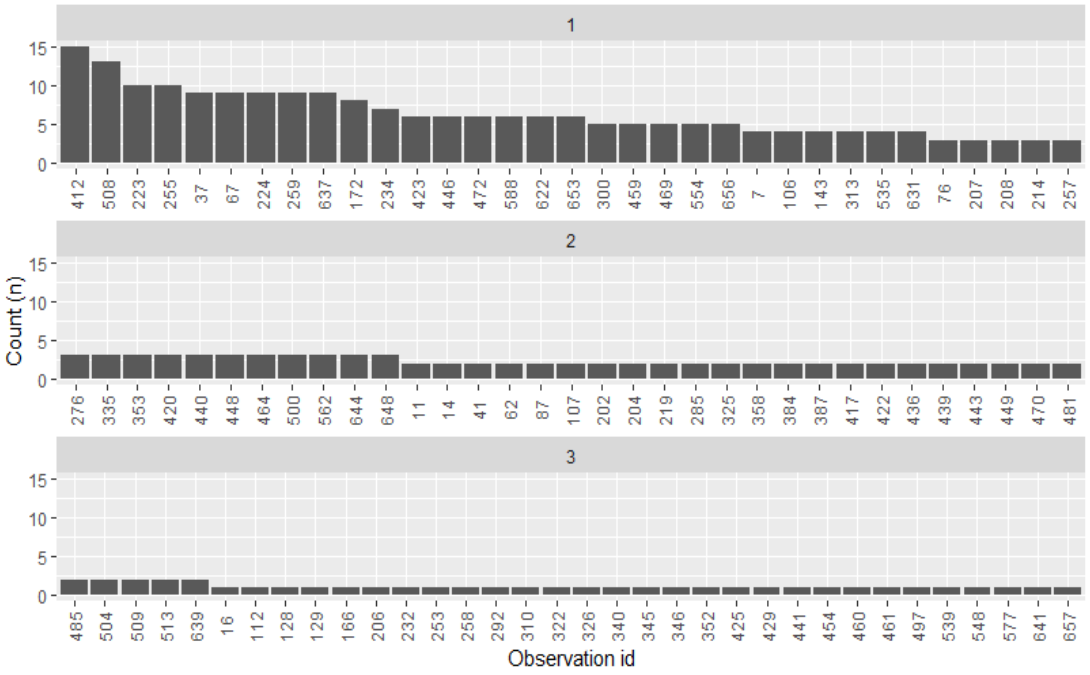

Figure 6: Observations detected as potential outliers from the univariate and multivariate strategies for SSG-BF.