Reliability in public transport is understood as the certainty travellers have regarding the level of service they will experience in their trips (van Oort, 2011). The travel time, arrival time, or the comfort level they will experience inside the vehicle are some of the most important reliability attributes. Besides, reliability is usually neglected from travel behavioural models (Petersen & Vovsha, 2006; Raveau et al., 2014), especially in Latin America where there is a lack of studies addressing its impact in travellers' choices.

Furthermore, the availability of large volumes of information regarding public transport systems' operation generated in an automated way has been increasing year after year. This represents a source of data with a level of detail that, properly processed, allows a detailed understanding of the system operation and how variable the level of service provided is (Birr et al., 2014; Bucknell et al., 2017; Cham, 2006; Fadaei & Cats, 2016; Furth & Muller, 2006; Gschwender et al., 2016). This type of information comes mostly from sensors strategically placed inside vehicles (such as GPS systems) and from records left by users with their means of payment when boarding or alighting the vehicles or stations.

This study proposes an in-depth analysis and characterisation of travel time reliability in a public transport system, taking Transantiago (the public transport system of Santiago, Chile) as a study case. The effects of travel time reliability on travellers' mode choice decisions are also analysed, as well as the impact of dedicated infrastructure (such as segregated bus corridors) on reliability indicators.

To address the proposed objectives, a statistical analysis of the available sources of information was carried out. Both the travel demand (origin, destination and specific services) and the travel times in each vehicle for the bus stages were obtained from the "trip-legs table" obtained with the methodology proposed by Munizaga & Palma (2012). In this methodology, there is no detailed information for metro travel times, since it is not known which train each user boarded and what route they took to reach their destination. To solve this issue, two extra sources of information were used: firstly, the arrival and departure times of all trains at each station in the network for the week of analysis and, secondly, proportions of use for the routes that connect the same pair of stations obtained from user surveys. With this information, it is possible to graphically represent the variability of travel time for trips of similar length and different mode.

So far, the available automated data has not been used in Santiago to understand how travel time variability has an impact (if any) on user's decisions. Based on the available information, this study develops an aggregate mode choice model, in which the explanatory variables are both average level-of-service indicators and indicators of their variability. This analysis further emphasises the importance of travel time reliability.

Overall, this study provides evidence of significant differences among travel time dispersion for trips of similar length and different mode. The standard deviation of in-vehicle travel time (and other measures of dispersion) can be quite high for bus trips, while in the case of metro is kept smaller than 5 minutes independent of trip length. Regarding the aggregate public transport mode choice model, a previous model using data for 2015 showed that the coefficient of variation for in-vehicle travel time shows a non-linear and significant impact. It was found that the users of the most unreliable services would accept travelling 6 minutes longer to completely avoid travel time unreliability, which is equivalent to a 22.2% fare increment (based on the social value of time). This model was updated with new data for 2017, showing similar results. More importantly, the new version of this aggregate choice model will also account for headway variability and its effects on waiting time. This was possible as more data regarding bus arrival times at stops was available for this year in comparison to 2015. These new data sources have already been processed and right now preliminary models are being estimated. This study also shows that segregated corridors not only reduce average travel times, but also reduce travel time variability which prove it to be an effective way to improve bus reliability.

It is important to emphasize that all the analysis in this study was conducted by only using passive-data, without the need of any kind of survey or external information. The data used comprises smartcard validation, buses’ GPS position and trains’ time schedules. Although the demand model is quite general (as no individual information, such as gender or income, is recorded in the smartcards), to the best of the authors’ knowledge, revealed preferences have not been used to analyse the impact of reliability on the preferences of public transport travellers. In a world where passive-data collection technologies rapidly gain importance over former techniques, studies similar to the one presented here represent a novel and quite promising approach for choice modelling and will help to understand passive-data capabilities and limitations better.