Collective Dynamics

Towards a Reference Database for Pedestrian Destination Choice Model Development

Christopher King · Nikolai Bode

University of Bristol, Department of Engineering Mathematics, Bristol, UK,
E-mail: aa18187@bristol.ac.uk, nikolai.bode@bristol.ac.uk

Received: 15 July 2022 / Last revision received: 31 October 2022 / Accepted: 11 November 2022
DOI: 10.17815/CD.2022.140

Abstract The move towards publishing research data openly has led to the formation of reference databases in many fields. The benefits of such resources are numerous, particularly in the development of models. While these exist in research on other aspects of pedestrian behaviour, no reference database is available for modelling pedestrian destination choice, the process by which pedestrians choose where they wish to visit next. This work seeks to construct such a database from the literature. The resulting data obtained are described and potential ways in which they could be used to calibrate a simple pedestrian destination choice model are presented. It contains four datasets that include destination choices for hundreds of pedestrians in settings ranging from university campuses and music festivals to highly structured stated preference surveys. A case study using one of these datasets to calibrate a simple pedestrian destination choice model is provided. These efforts highlight some general issues from creating and using reference data openly. Discussing these issues will hopefully guide the development of reference data and accelerate the development of accurate pedestrian destination choice models that can be applied generally.

Keywords Pedestrian destination choice · reference database · model calibration · pedestrian dynamics · choice modelling · open data

1 Introduction

As the urban population grows, it becomes increasingly important for overseers of buildings, event sites, and cities, to understand and predict crowd behaviour. It is widely believed that people walk to places due to a desire to perform activities [1], so understanding

Licensed under CC BY

how and why people choose their next destination, *i.e.* pedestrian destination choice, is crucial in understanding and forecasting crowd movement.

Pedestrian destination choice models, and scientific models in general, are often developed similarly: first, the model is specified, where the researcher decides what assumptions are made regarding the process being studied and how observable and unobservable sources of error are taken into account. This is also where the mathematical form to describe the process, determining which predictors are relevant and how they interact with each other, using either prior knowledge or any collected data. Next, model calibration (also known as model estimation) is conducted, which determines the optimal weighting of said predictors to best fit a set of data. The next step is to validate the model, which involves performing checks that the model can reproduce observed behaviour. This can be qualitative or quantitative. The order of these steps can vary, and/or some steps can be skipped entirely. Some or all of these steps can be repeated, where predictors can be added or removed to improve the model's ability to explain the data, depending on the objectives of the researcher. But once completed, a model can be used to either explain observed behaviour or to predict the behaviour of people in hypothetical scenarios.

In the context of pedestrian destination choice models, calibration gives an indication of the relative importance of the predictors in choosing any particular destination. Calibration is therefore an important step in model development and relies on data being collected and available. When it comes to pedestrian behaviour research, the most popular aspects are: routing - determining the most likely route a person takes through an environment, and mobility - understanding the physical characteristics of humans moving around a space. A great deal of data has been collected in these areas and more recently, reference databases have appeared [2, 3]. These have the potential to allow considerable progress to be made in improving the accuracy, robustness, and validity of the models that represent these facets of human behaviour. For example, Zhou *et al.* [3] have released a database [4] of videos of crowds moving in 62 locations, along with the individual trajectories. This data was collected in the context of identifying collective motion in crowds and has been used by a variety of authors within this field [5–9]. Wu et al. [9] use the video data to identify collective motion in crowds using the curl and divergence of a crowd motion vector field. Liu et al. [7] instead use the data to assess the efficacy of different methods for tracking individuals from videos. Additionally, the data has been used to assess algorithms for identifying and segregating collective motion in crowds [5, 6, 8]. The same cannot be said for pedestrian destination choice, where comparatively fewer data have been collected, and no reference datasets are available.

There has been a recent increase in demand for openly available data, particularly for use in areas such as machine learning [10]. This has seen a number of funding bodies include stipulations that data collected must be published openly, where possible. Such data is often collated into reference datasets that are openly available, easily interpretable, and well organised. These have appeared in a number of fields, such as in wireless networks [11], computer vision (e.g. [12]), and medicine [13][1]. Having such data available not only allows new models to be calibrated and validated, but also makes model comparison

[1] Accessed: 14/07/2022

easier and more reliable. Consider pedestrian movement as an example, entries in the data archive hosted by the Institute of Advanced Simulation at the Forschungszentrum Jülich [2] has been utilised in a variety of ways. Data originally collected by Seyfried *et al.* [14] has been used to: calibrate the Optimal Steps and Social Force model [15], validate an evacuation simulator which includes both cooperative and competitive agents [16], and recalibrate a collision avoidance algorithm [17], among others [18–20]. Other entries in this database have also been used to compare and assess pedestrian behaviour models. For example, the data provided by Cao *et al.* [21] was used to compare results from a pedestrian movement simulator using a velocity-based model [22] the data collected by Zhang *et al.* [23] is used to assess a generalised centrifugal-force model for pedestrian movement in work by Rathinakumar and Quaini [24]. Zhao et al. [25] also use this dataset to train, test, and validate a neural network which predicts pedestrian velocities.

Therefore, in an attempt to accelerate pedestrian destination choice model development, this contribution aims to create a database of reference data which can be used by other researchers in the field. For example, researchers could use it to compare and/or assess their own models. With further work to formalise each dataset, the database could also act as a standard for calibration, validation, and comparison of pedestrian destination choice models. It should also make reproducing and replicating published results much easier [26]. To do this, a search for viable data sources is conducted in the pedestrian destination choice literature, which is described in Sec. 2. This contribution also hopes to highlight the lack of openly-available data and serve to encourage researchers in this field to consider publishing their data openly in the future.

Reference datasets are often subject to strict standards for publishing, use, and access. This is done to streamline the process of using any data published and provides a set of guidelines for publishers [27]. Typically, data collected that was not initially intended for open publication will have issues, such as missing entries, inaccessible formats, and structures that are hard for others to understand. This requires careful consideration when constructing and using a reference database. The other contribution of this work is to therefore describe and discuss the general issues that can arise when collating and using reference data for model development (Sec. 4). These issues are illustrated by means of a case study, where one dataset is used to calibrate a simple pedestrian destination choice model (Sec. 3).

2 Literature Search

This section describes the methods used to search for, select, and include previously published data (Sec. 2.1). Results of this process are also presented here, describing, among other things, the numbers of candidates identified, the outcomes, and the issues encountered, at each stage (Sec. 2.2).

2.1 Methodology

There are many ways of searching for and obtaining literature, with more reproducible and systematic techniques being developed [28–31]. These include searching online services such as Web of Science[2] and SCOPUS[3] using specific keywords and phrases. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [28] is an example of an orderly approach to writing literature review papers. It consists of a checklist of 27 items that help ensure a robust, reproducible, and clear literature review has been carried out. It has been used in several recent works (e.g. [32–34]). Several aspects of this methodology which are applicable to the search for data, rather than literature, are utilised here. Another, often complementary, method of finding literature is backward and forward snowballing [35] which involves identifying relevant references and citations from a set of initial papers, respectively. Each discovered article can themselves have snowballing applied to them, and potentially numerous relevant literature can be obtained in this way. Another more collaborative approach involves reaching out to key researchers in the area of exploration [31].

The search for data employs several methods described above. First, an ad-hoc search for potentially available datasets in the literature was conducted. This is because data can be found in a huge variety of research fields and topics and is often not mentioned in titles, abstracts, or keywords. Instead, initial candidate papers were selected: [36–45]. Candidate papers were selected if their data satisfied the following conditions:

1. **Individual** - Measurements of aggregate properties of crowds are often made [46–48], such as pedestrian flows, densities, and mean speeds. However, decisions of individuals are required for subsequent analysis by destination choice models. Therefore, data in which individual participants can be uniquely identified is a necessary stipulation.

2. **Walking only** - in many research fields, such as transport modelling, the concept of ‘activity-based demand’ [1] is widespread when attempting to explain and predict flows of people on transport networks, such as roads. This concept states that demand and flows are driven by individual desires to visit destinations, so an individual’s choice of destination is important. However, modelling of these decisions often involves an addition choice of transport mode, e.g. walking, car, train, etc. Therefore, only data of people travelling by walking only will be included in this database.

3. **Spatial scales** - The size of the data collection area must also be considered, as individual destination choice data can be collected on scales anywhere from countries [49–51] to within a couple of metres [41,52]. This depends partly on the scope through which data is collected, but also on how one defines a ‘destination’. Destinations in this work are particular areas where a person can perform activities, but

[2] `https://login.webofknowledge.com/`, accessed: 14/07/2022

[3] `https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030`, accessed: 14/07/2022

other definitions can be a country or region to visit for a holiday [49, 50, 53] or a hypothetical waypoint along a route [54–56]. Therefore, spatial scales in which an individual can visit several areas of potential interest are required. Depending on the context, this can be anything from the size of a floor of a building to the size of a city block.

4. **Temporal scales** - sometimes individual-level data is only recorded for a few minutes, e.g. for calculating and/or predicting pedestrian trajectories [2]. Other times, it can be recorded over many weeks, months, or even years, e.g. in household activity surveys [37, 57, 58]. In this work, viable data for inclusion must have been collected over a period anywhere in the order of tens of minutes to several days. No exact thresholds are used, as the timescale of data depends on the size of the area studied as well as the context for which it was collected.

Potential privacy concerns can arise from the first point, especially in light of recent legislation regarding data privacy (e.g. GDPR in Europe[4]). However, no personal information that could be used to uniquely identify individuals is required for the database, only that each datum can be related back to a unique individual. Depending on the context and kind of data collected, it can be straightforward to achieve this by removing such personal information while giving each individual a unique numerical identifier. However, in other contexts, such as daily mobility data, this may not be sufficient, as patterns within the data can still be used to identify sensitive personal information about individuals, so researchers must be careful.

While there is no explicit criteria on when data was collected, the likelihood of finding relevant data gathered from further back in time is assumed to decrease. This may be because of a lack of available technology and methodologies that could produce data that meets the criteria outlined above. Also, data that was collected long ago is potentially more likely to be lost, or otherwise unavailable for online publication.

Fig. 1 summarises the data search procedure and provides the numbers of candidate papers identified at each stage. This search was conducted over the course of several months from March to December 2021. Any candidates that had already published their data openly were immediately included in the database. Otherwise, the corresponding authors of the papers were contacted via email about potentially publishing their data. If the author was willing to publish, then the data were added to the database either once the data was publicly available or if the author shared the data directly (with the expectation that the online location of said data would become available). Any suggestions made by these authors as to who or where to find appropriate data were also investigated.

In order to expand this search, backward and/or forward snowballing [35] is used on the initial candidates. This is supplemented by approaching contacts in the author's research network and posting on the professional social media sites LinkedIn[5] and ResearchGate[6].

[4] https://gdpr-info.eu/, accessed: 14/07/2022

[5] https://gb.linkedin.com/, accessed: 25/02/2022. Job advertisement website where professionals from around the world can connect and spread information.

[6] https://www.researchgate.net/, accessed: 25/02/2022. Website where scientists can share papers, ask questions, and find collaborators.

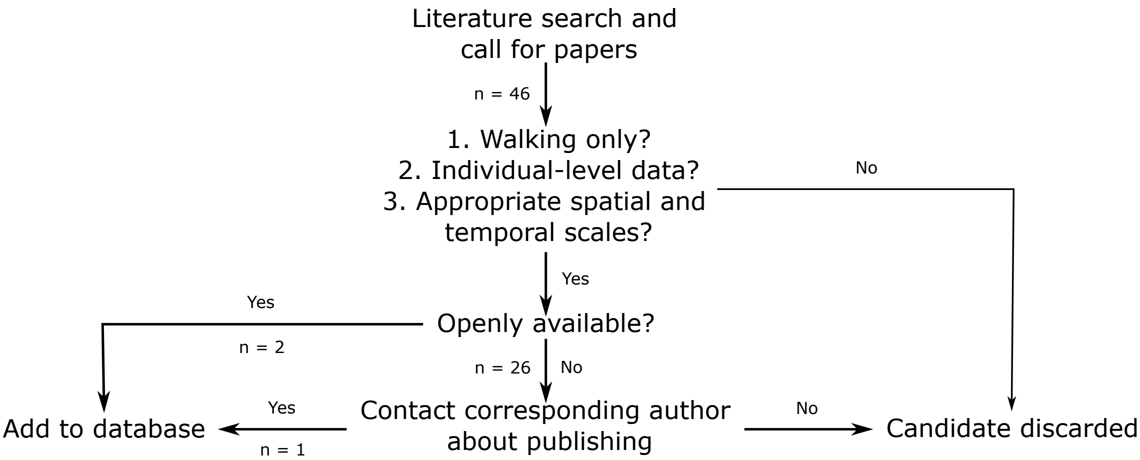
Literature search and call for papers
n = 46
1. Walking only?
2. Individual-level data?
3. Appropriate spatial and temporal scales?
No
Yes
Yes
n = 2
Openly available?
n = 26
No
Yes
n = 1
Add to database
Contact corresponding author about publishing
No
Candidate discarded

Figure 1 Summary of the data selection process from the literature.

Additionally, the author made use of recommendations suggested by Science Direct[7] and ResearchGate based on previous papers viewed. Finally, if an author publishes several potential candidate papers, then they are contacted via email asking if they have any data that could be published. Each candidate paper identified in these ways is subjected to the same procedures shown in Fig. 1.

2.2 Results

This section describes the results of the literature search described in Sec. 2.1 and gives detailed overviews of the datasets obtained.

Starting from the ten initial papers selected (listed in Sec. 2.1), an additional 46 papers were identified via backward and forward snowballing. No additional candidate papers were identified through the social media posts or through liaison with the author's research network. Of the additional papers identified, 28 of these contained data which satisfied the criteria outlined in Sec. 2.1, of which only two had published their data openly. After emailing the corresponding authors of the remaining 26 candidates, nine responded. Several authors could not release any data for a variety of reasons, including privacy concerns, author's not owning or no longer having access to the data, and the data being commercially sensitive. Five of the authors suggested others to ask and/or places to look, but no further viable data was identified. Only two of the authors contacted shared their data. However, for one of these, the mapping of physical locations to nodes of the network used to represent the space was missing, so the data could not be shared.

Tab. 1 shows the datasets included in the database. The first three were found using the literature search methods described in Sec. 2.1, while the fourth is data collected in previous work by the authors. The information on finding these datasets is given in Sec. 5. For the remainder of this section, each dataset acquired is briefly introduced, describing how it was collected, for what purpose, and what information is provided.

[7] `https://www.sciencedirect.com/`, accessed: 25/02/2022. Bibliographic database of scientific and medical publications of Dutch publisher Elsevier.

2.2.1 Danalet data

The data provided by Danalet *et al.* [59] (henceforth referred to as 'Danalet's data') was one of the two openly available datasets found during the search. A full description of how the data was generated can be found in [41]. It was collected between May and July 2012 on the campus of the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. They recorded the positions over time (trajectories) of Wi-Fi-enabled devices of both students and staff. This data was combined with additional information, such as campus layout, building capacities, and course timetables, to infer the most likely sequence of destinations visited by individuals. To determine a device's approximate location, triangulation was performed on the Received Signal Strength Indicator (RSSI) of devices to nearby Wi-Fi access points [59]. This data is used to calibrate and validate statistical models of activity and destination choice, which were then used to make predictions [41].

Danalet's data consists of three parts: the Wi-Fi data, the 'Semantically-enriched Routing Graph' (SERG) of the EPFL campus, and the 'Potential Attractivity Measures' (PAMs). The SERG is a graph representing the space in which data was collected, with nodes representing places and links representing the routes between them. Both nodes and links contain additional information to better represent the underlying space. For example, nodes can have attributes such as opening times, seating capacities, and facilities available. Links can be weighted according to distance travelled and whether there is a floor change and/or one-way systems. Nodes are given spatial coordinates of which destinations are a subset. The role of the SERG is to link the Wi-Fi data with the environment and allows for realistic routing between destinations. PAMs attempt to quantify the desirability of a destination over time and are context specific, e.g. for classrooms, it is the number of students enrolled on the course taking place there, for offices, it is the work rates of employees stationed there, and for restaurants, it is the total seating capacity. These attractivity measures are used to help infer the most likely destination visited by an individual over time.

There are two sets of Wi-Fi data present: one that records the movements of the lead author of the work leading to this data over the course of a working day, and one that records the movements of anonymous campus users, which is described in the first row of Tab. 1. The former data was originally used to estimate the accuracy of the overall activity sequence assignment process as the true sequence of activities performed is known.

The movement data for campus users is itself split into two versions: raw and cleaned. Both versions give information on the device location, its associated user, and the time at which it was detected. The cleaned data was generated from the raw data by removing data from individuals where a partial trajectory is collected (see [60] for more details). However, they are formatted differently, with the raw data being in the format that it was originally collected (without duplicated rows) and the cleaned data being designed for use in a Structured Query Language (SQL) database. Both formats can be read by any commonly-used text editor. The data and accompanying description can be found in Sec. 5.

Table 1 Summary information of the datasets available in the database.

Authors	Collection Date	Location	Collection Method	Participants	Sample Size
Antonin Danalet, Bilal Farooq, Michel Bierlaire [59]	May - July 2012	EPFL campus Switzerland	Wi-Fi positioning using triangulation of received signal strength indicators (RSSIs)	Staff and students	4257[1] 4180[2]
Antonin Danalet, Loïc Tinguely, Matthieu de Lapparent, Michel Bierlaire [61]	May - July 2012	EPFL campus Switzerland	Bayesian inference of destination choice using Wi-Fi positioning (same method as above), and other data	Staff and students	211[3]
Bram Bonné, Arno Barzan, Wim Lamotte, Peter Quax [62]	August 2012	Hasselt, Belgium	Wi-Fi localisation - binary indication of when device enters/leaves detector range.	Festival attendees	138641[4]
Christopher King Nikolai Bode	July 2021	N/A	Online surveys - questions asking for choice of destination in hypothetical scenarios.	People aged 18 and older	813

[1] raw, 3490 staff

[2] cleaned

[3] 145 staff

[4] Number of unique devices detected over all days

2.2.2 Tinguely data

The data in the second row of Tab. 1 (henceforth called ‘Tinguely’s data’) is generated from Danalet’s data using techniques that are fully elaborated in Danalet *et al.* 2016 [61], Danalet 2015 [41], and their associated references. This data provides information on the choice of catering location for students and staff on the EPFL campus between 16th May and 4th July 2012. This data is used to estimate and validate a variety of catering destination choice models. The model that best fits this data is then used to predict the effect of adding a new catering location on the preferences of campus users. The number of individuals in this dataset is all individuals who chose to visit a catering location during opening hours over the data collection campaign. This data does not show consecutive choices made by each individual, only their choice(s) of catering location. This is distinctly different from the Danalet data, which only gives the location of Wi-Fi-enabled devices over time, along with supplementary information of destinations and the environment. Tinguely’s data also includes additional information on the individual (e.g. staff or student), the times at which the individual arrived and left the location, and destination traits, such as the capacity and services provided.

The data can be found as part of a repository on the website Zenodo[8] and is linked in Sec. 5. This repository is initially split into two sections; one section contains the aforementioned data and the code to analyse the different choice models studied, the other contains the code to generate the accompanying technical description. Within the first section, there are three subdivisions: the first contains the data, split into three parts: the data used to calibrate the models, subsets of the calibration data used to validate the models, and the estimation data with additional information about a new destination, which is used for making predictions. Only the data used to calibrate the models is discussed in this work. The second subdivision contains the Python code used to run the aforementioned analyses for each destination choice model. This also includes code for running sensitivity analysis on the models. The final subdivision stores the output of said code. The technical description provides more information on what is available in this repository.

2.2.3 Bonné data

The third dataset collected by Bonné *et al.* [62] (named ‘Bonné’s data’ forthwith) was donated to the database as a result of contacting the authors. This data was collected during the Pukkelpop festival held in Hasselt, Belgium. The festival lasted for three days from 15th-18th August 2012, with additional data being collected on 14th August, where a subset of destinations was made available to a limited audience. This data was collected as one of two case studies that demonstrate a new method of tracking the movements of people using existing Wi-Fi infrastructure and so illustrates the capabilities of this technology for use in crowd monitoring during large-scale events. The link to this data repository is provided in Sec. 5.

The primary data provided is the mobility data of devices over the course of the festival. This data is separated by the four collection days, where each day is split by detector.

[8] `https://www.zenodo.org/`, accessed: 14/07/2022

These files contain: a unique device ID, a binary indicator giving the first time a device is detected or not detected, and the associated measurement time. This method of detecting devices means that a device is assumed to be present (*i.e.* within the detector's coverage area) for the entire time between when it is first detected and when it is first not detected. Additionally, a map of the festival space, which has the positions of each detector annotated, is provided, along with the schedule of performances at each destination over the three days of the festival. The latitude and longitude coordinates of each detector are also present within the detector filenames. For further information on the nature of these data and the repository structure, please refer to the accompanying description document.

2.2.4 King data

The fourth row of Tab. 1 describes the data collected in previous work by the authors [63]. This data was collected between 9th and 20th July 2021 using an online survey platform[9]. The surveys gave respondents the same task of choosing up to five destinations to visit within a hypothetical time in a given environment of six possible destinations. After each choice made, the respondents were told how much time remained. The amount of time elapsed between choices depended on both the distance and busyness of the chosen destination, participants were told this information before starting. If a respondent ran out of time before choosing five destinations, then the survey would end, but the sequence of destinations chosen would still be recorded. Five different variations (or experimental conditions) were considered:

1. Base case - Reference case, using an open environment. Information presented as an abstract screenshot.
2. Schedule chosen - Before choosing which destinations to visit, participants are required to create a schedule of the destinations they would like to visit and in what order. Otherwise, identical to the base case.
3. Schedule given - participants are provided with a suggested order of destinations to visit in the environment. Otherwise, identical to the base case.
4. Closed environment - A different environment which constrains direct travel between destinations. Otherwise, identical to the base case.
5. Photo - The information on each destination in the environment is provided by photos taken from a human observer's perspective. Otherwise, identical to the base case.

The repository is split into four sections:

- Processed data — the data used to generate the results in the accompanying paper.
- Raw data — the data directly produced from the online surveys with no processing.

[9] Online Surveys, URL: `https://www.onlinesurveys.ac.uk/`, accessed: 14/07/2022

- Schedule chosen variants — the schedule chosen experimental condition data had additional restrictions imposed. Versions of the data with and without these restrictions are provided here.

- Images — contains all images used for the surveys.

The first three sections are further split by the attempts made by participants. For the raw and processed data, these are divided by experimental condition. Within each experimental condition, there is a text file containing all the data necessary to calibrate the destination choice model specified in Eq. 1 and 2. There is also a file containing the socio-demographic data of all participants. Additionally, for the schedule chosen and given conditions, there is an additional file providing the schedules of every respondent and how they change over the course of the survey. The Schedule chosen variants contains choice, schedule, and socio-demographic data for all respondents for each possible variant of the schedule chosen data.

3 Case Study

To demonstrate how these kinds of data could be utilised by future researchers in the field of pedestrian destination choice, a case study on one of the datasets is performed. The main objective of this study is to attempt to calibrate a simple destination choice model using data collected by others. In doing so, the considerations needed to repurpose the data for calibration of a different destination choice model can be illustrated. Any subsequent issues that arise during this procedure will also be identified and discussed in detail.

Bonné's data was chosen for this case study, as it is the only dataset that has been subject to little analysis. This should inform how this particular dataset could be analysed as well as highlight potential limitations. The issues and considerations highlighted are examples of generic problems when using reference data for model calibration.

This section is organised as follows: first, a description of the data processing and model calibration is given (Sec. 3.1). Then in Sec. 3.2, the results of the model calibration are presented, along with their interpretations as well as a discussion of any further issues not mentioned in Sec. 3.1.

3.1 Methodology

A description of the Bonné data and the information provided has been given in Sec. 2.2.3. The link to this data, and all other datasets in the database, can be found in Sec. 5.

To the best of the author's knowledge, this is the first time this data has been analysed in detail. The paper by Bonné *et al.* [62] provides a description of the data and suggests possible applications. The aim is to investigate how this data can be used to calibrate a simple destination choice model (see below). Doing so shall highlight any issues and simplifications that need to be made during this process. This will hopefully inform future researchers about any problems with using reference data, both in general and specifically for this dataset.

Discrete choice models [64] can be used to broadly understand pedestrian destination choice. These models assign the probability of a decision-maker choosing a particular alternative out of a set of possibilities. This probability is often a function of potentially influential factors weighted by parameters. The primary assumption made is that human decision-making is a trade-off between different factors (known as predictors forthwith) that result in an optimal outcome for the decision-maker (e.g. [41], [65], [38]). Many modelling frameworks exist, each with advantages and disadvantages [64]. The pedestrian destination choice model used in this study is a multinomial logit, where the errors in any unobserved predictors are assumed to be independently and identically Gumbel distributed and is the same as the one used in previous work [66]. Though this is model has restrictive assumptions that make it unsuitable in many real-world situations [64], it is used in this case study because of its simplicity and ease of use. As mentioned above, the aim of this case study is not to give accurate results, but to explore how this data could be analysed. It is given by the equations below:

$$P_i = \frac{e^{U_i}}{\sum_{j \in C} e^{U_j}} \tag{1}$$

where P_i is the probability of choosing destination i out of the set of all possible destinations (the choice set) C, and

$$U_i = \beta_{occ}\hat{n}_i + \beta_{dist}\hat{d}_i + \beta_{des}\hat{q}_i \tag{2}$$

where $\hat{n}_i$, $\hat{d}_i$, and $\hat{q}_i$ are the occupancy of, distance to, and desirability of i, respectively. They are denoted with a hat symbol to show that these are normalised between zero and one.

These three factors are commonly used in pedestrian destination choice research (see below). This observation supports the intuition that these factors are influential in various pedestrian destination choice contexts, for instance, shopping centres, transport hubs, or mass events. How these factors influence a person depends on both the context and the qualities inherent to the person, such as familiarity with the environment and current mental state. A commuter at a train station will probably avoid busy places as they constrained by time. They are likely to be familiar with the environment and so would be comfortable visiting an alternative destination which is further away. A tourist, on the other hand, may not have such a time pressure but may be unfamiliar with the environment. Therefore, they may be more likely to choose a busy but familiar route to a known destination as a result, even if there is a less busy destination elsewhere. This case study is only used to illustrate how Bonné's data, and other datasets presented in this database, can be analysed, so including potentially complicating factors, such as situational context and internal mental factors of the decision-maker, are not considered here, even if doing so could create a model that would explain the data better (though this can be included in future work).

Occupancy is often an influential factor when choosing a destination [38,40,67,68]. In some situations, such as festivals or tourist attractions, it can be attractive as it indicates something worth visiting [69]. In other scenarios, for example when shopping or taking public transport, it can be a repulsive factor.

In most scenarios, distance from a destination can be considered a repulsive factor (e.g. [39,41,70–76]). However, if the decision-maker wishes to exercise, has no time constraints, or just enjoys the journey as much as reaching the destination, then it can also be attractive.

The intrinsic motivation of pedestrians to visit a destination is ubiquitous in pedestrian destination choice modelling. Pedestrians often visit destinations with one or more activities to perform, with some activities being more important to complete. Some pedestrians will therefore form an itinerary or schedule of activities and this is likely to influence the decisions made as to which destinations to visit, with destinations where more important activities can be performed being more likely to be chosen. Research suggests many ways of capturing this intangible influencing factor, such as from market research [77,78], using decay functions based on individual opinions [67,79], the number of transitions between destinations [80], or using properties such as seating capacity [41], and floor space [71]. The desirability is a predictor with a context-dependent interpretation and one that can be used to capture a group of possible influential choice attributes. In this case study, the desirability designed to quantify how much a person wants to visit a particular destination and its interpretation depends on the context. For example, when shopping, shops which sell a necessary product might be more desirable than those which don't, or when at a theme park, certain rides may be more preferable depending on the individual's taste and personality. This model is not necessarily the most suitable for this particular context, but it is appropriate for exploring the data.

Both β_{occ}, and β_{dist} can take positive and negative values, representing the potential for occupancy and distance to have an attractive or repulsive effect, respectively. The parameter β_{des} can only be positive, because it represents the effect of an individual's chosen or given destination schedule. Negative values would mean that an individual does the opposite to what they desire. It is expected in the context of a music festival, that distance will be negative, as attendees might be more unwilling to walk longer distances. It is unclear what sign and magnitude the occupancy parameter might take in this context, as some attendees may wish to avoid large crowds while some may be drawn to them. The magnitude of the desirability parameter may be smaller than those of the other parameters, as it is expected that the importance of performance schedule will vary over time and attendees.

This model will be calibrated using Maximum Likelihood Estimation (MLE), where the model parameters are optimised to maximise the likelihood of this model explaining the data. Four pieces of information are required to calibrate the model described in Eq. 1, and Eq. 2 using MLE:

1. Observed decisions made by individuals. This is used to calculate P_i from the data as part of determining the likelihood.

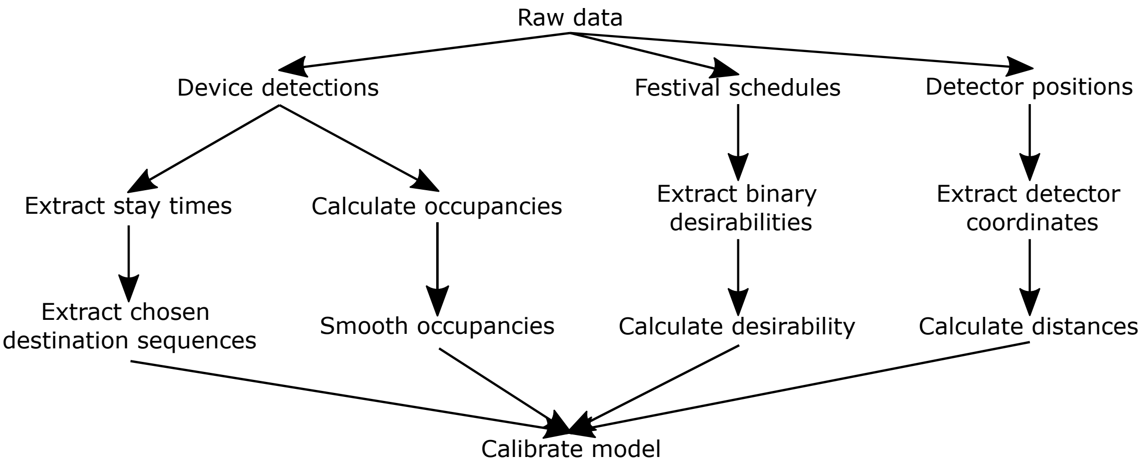

Figure 2 Overview of the Bonné data extraction and analysis process, separated according to the three main sources of information present. The 'device detections' branch is described in detail in Sec. 3.1.1, the 'detector positions' branch is described in detail in Sec. 3.1.2, and the 'festival schedules' branch is described in detail in Sec. 3.1.3.

2. Destination occupancies, n - how many people are present at all destinations when a decision was made.
3. Destination distances, d - how far away are destinations from the decision maker's current position.
4. Destination desirability, q - how 'desirable' are all destinations at the time the decision was made.

Fig. 2 gives an overview of the process used to extract the four pieces of information necessary for model calibration as described above. The values of the three predictors must be available for all destinations at all decision times, not just for the chosen destination, in order to calculate the normalisation constant in the denominator of Eq. 1. All the processes described in this section are performed using the R programming language [81].

Before extracting the necessary information, however, the set of destinations (*i.e.* the choice set) in this context must be defined. Fig. 3 provides a rough sketch of the layout of the festival, along with the approximate positions of all detectors. This sketch is based on both the map and the latitude and longitude coordinates of each destination as provided by the Bonné data. The names of each detector are extracted from the data and are also displayed. In this case study, a destination is defined as an area in which one or more activities can be performed. In the context of a music festival, these activities could include: watching a performance, getting something to eat, and going to the toilet. The performance schedule for the festival shows eight areas where performances occur and this is where eight of the 14 detectors have been placed, with the other six placed at thoroughfares. But, since the size of the area in which a device can be detected (henceforth known as 'coverage') for each detector is unknown and is likely time-dependent due to a variety of factors that can influence the detection range of devices [82], it is possible that a

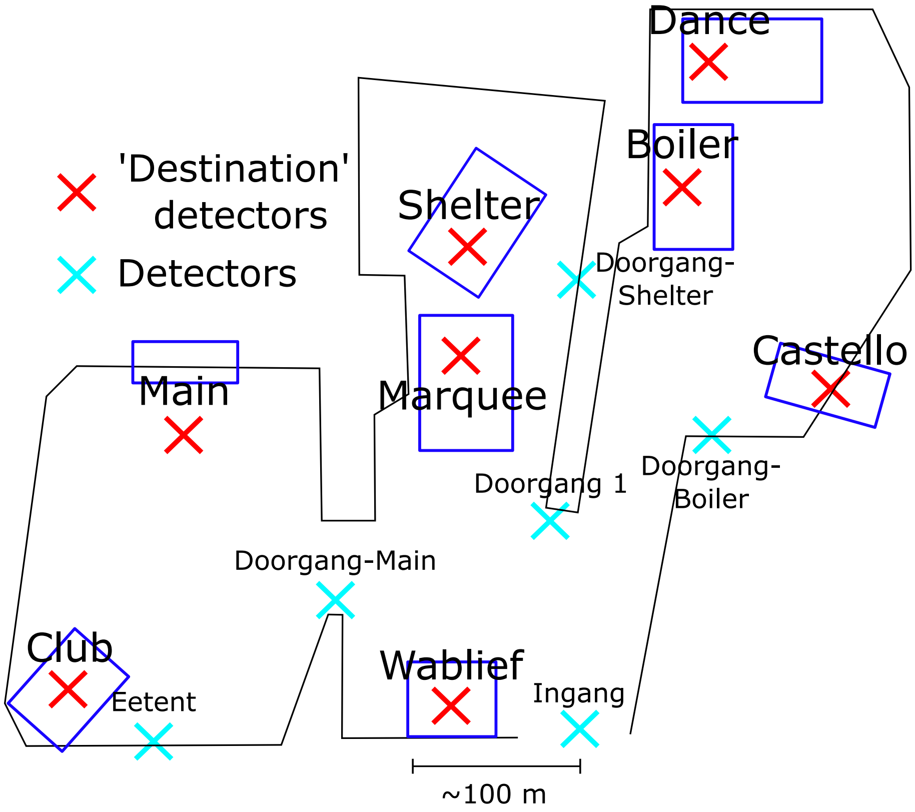

Figure 3 Sketch of the festival area used in the Bonné data. The approximate placement of all detectors are given, with detectors classed as destinations in red.

detector can have multiple activities associated with it. Due to this lack of information, it is assumed that detectors at performance areas capture only one kind of activity: attending a performance, and the detectors placed at thoroughfares have no associated activities. Therefore, the eight performance areas are considered as the set of possible destinations, and these are marked in red in Fig. 3. The data from the remaining six detectors (marked in cyan in Fig. 3) are not used. Each destination is identified by arranging the detector names in alphabetical order and numbering according to their position: Boiler = 1, Castello = 2, Club = 3, Dance = 4, Main = 11, Marquee = 12, Shelter = 13, Wablief = 14.

The following subsections describe each branch of Fig. 2 in turn, showing how each of the three main pieces of information present in the Bonné data: mobility data, festival map, and performance schedules, are used to infer the required data above.

3.1.1 Mobility data

As described in Sec. 2.2.3, the mobility data provides the times at which devices first enter and leave the detection area for each detector over all four days of data collection. The festival itself occurs over three days (Day 1 to 3). Data was also collected the day before the festival (Day -1), where the area was available to workers and a select few attendees. This is why the values for this day are smaller than on the other main festival

Table 2 Quantitative summary of the Bonné raw mobility data. Day -1 is the day before the festival started.

Day	Total number of measurements	Number of devices	Collection time / HH:MM:SS	Median stay time / s
-1	79731	18835	16:00:06-05:59:56	577
1	290244	52646	11:00:00-05:59:46	1010
2	251324	49203	11:00:00-05:59:59	1124
3	212604	46565	11:00:00-05:59:59	1414

days. There is also no performance schedule for this day, so the data from this day are not used in model calibration. This data is summarised in Tab. 2. The first column describes the total number of measurements recorded by all destination detectors over each day. The stay time of a device at a destination is the difference in the time at which a device is no longer detected (henceforth referred to as ‘absence time’) and the time at which a device is first detected (henceforth referred to as ‘detection time’). The median stay time over all devices and destinations is provided in the fourth column. The table shows that for the main days of the festival, Days 1 to 3, the total number of unique devices recorded diminishes, suggesting that fewer people attend as the festival progresses. The detectors record for almost 17 hours per day, assumedly covering the most busy periods. The median time spent at any given destination over the course of the festival increases, indicating that devices are spending more time at destinations as the festival progresses.

To calibrate the destination choice model, the sequences of chosen destinations for all individuals are needed, these are comprised of three pieces of information: the time at which choices were made, the current destination, and the destination chosen, for each individual. Here, it is assumed that there is a one-to-one relationship between a device and an individual, so the movements of one unique device correspond to one unique individual throughout. Therefore, the sequence of detectors where a given device is observed gives the sequence of current destinations visited by an individual. The destination chosen is the next destination visited, therefore, the chosen destination sequence is the current destination sequence shifted one destination later. The chosen destination sequence is one destination shorter than the current destination sequence as there is no information about where the decision to visit the first destination was made. There is no way of knowing exactly when the individual associated with a device made a decision, so for the purposes of calibration, the decision time was decided to be the absence time.

For example, consider a situation where a device visits the following sequence of destinations: Main, Boiler, Wablief. The device is detected arriving at these destinations at 09:30:00, 10:30:00, and 11:30:00 and leaving at 10:00:00, 11:00:00, and 12:00:00, respectively. Therefore, the decision times for this device are: 10:00:00, 11:00:00, 12:00:00, with a corresponding current destination sequence: 11, 1, 14, and chosen destination sequence: 1, 14. This assumes that the individual associated with the device does the following: arrive at Main to perform an activity for 30 minutes, then visit Boiler, stay there to perform an activity for 30 minutes, then visit Wablief and spend half an hour performing an activity.

However, some devices will belong to people outside the festival or will be stationary devices not associated with any one person. To help filter these out from the data and obtain valid chosen destination sequences, only devices detected at more than one destination are considered.

It is also important to try to distinguish between people moving through the coverage of the detector with people visiting the area to perform an activity. To do this, the distribution of stay times over all devices, destinations, and days is created. The 5% and 95% quantiles of this distribution are estimated and used as the criteria for a device being counted as 'visiting' a destination. If a device has a detection time but no absence time, then it is assumed to have stayed at the destination for 130 s, which is the average length of time that a device should be detected if in range. This value was determined empirically by Bonné *et al.* 2013 [62]. If a device has an absence time without an associated detection time, then no stay time is estimated.

Fig. 4 shows the stay time distribution after the artificial 130 s stay times have been applied. The 5% and 95% quantiles are estimated at 236 s (3 min 56 s) and 5354 s (1 hr, 29 min 14 s), respectively. Only the subset of devices that had stay times within these bounds were considered subsequently. There is no noticeable peak in the stay time distribution at 130 s, so the artificial durations described above have little influence on the stay time distribution.

If the application of the stay time constraints leave a device with only one visited destination, then it is discarded, as at least two visited destinations are required to calculate choice probabilities. For each remaining device, the detection and absence times, and the current and chosen destination sequences are extracted and separated by day. They are sorted chronologically by detection time.

Upon examining the extracted sequences, a new problem became apparent: presence conflicts. These arise when the absence time of the previous destination is later than the detection time at the subsequent destination, suggesting that a device is at two places simultaneously. It is possible for a device to be detected at two or more Wi-Fi detectors [83], but this is not sensible when inferring a person's physical location from a device. Between 26% (Day 1) and 32% (Day 3) of all choices are involved in a presence conflict. Even destinations which are far apart from each other can be involved in these conflicts, suggesting that this problem is not solely due to overlapping coverage of detectors. One possible source of these conflicts could come from the imperfect detection or absence of devices. Perhaps this is an issue with the new detector technology, e.g. it is possible that an unpaired detection at a destination could be mistakenly paired with the absence time of a separate measurement at that destination.

This presents a problem for calibration, as there is no way of discerning from the information provided which destination involved in the conflict is the true destination visited. Therefore, two versions of the destination sequences are generated, the original ('conflicted'), and one where all conflicted choices were removed ('clean'). If any devices have no choices as a result of this cleaning process, then they are discarded. Both versions will be used to calibrate the model separately to examine any effect on the results that may arise from neglecting a significant proportion of the available data. Details of the number of devices and choices remaining after each stage of this extraction process

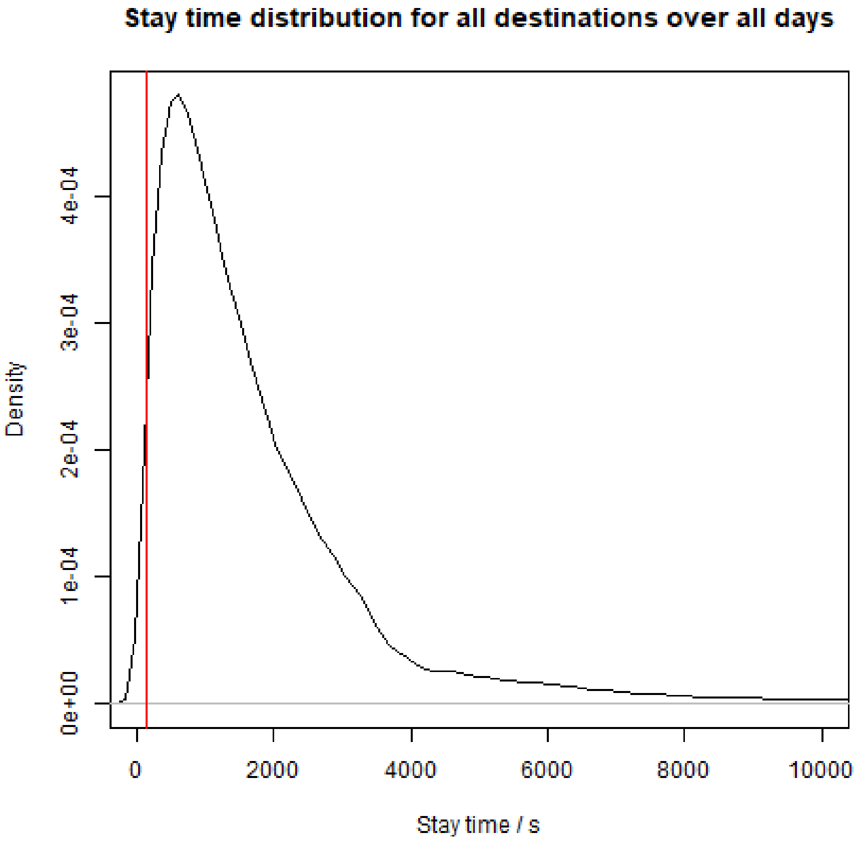

Figure 4 Distribution of stay times for all valid detections over all destinations and collection days. The red dashed line marks the position of the artificial duration used in cases where a device was detected arriving but not detected leaving a detector.

are given in Tab. 5, and Tab. 6 in App. 7, respectively.

The total number of detected devices present within the detection area of each destination (*i.e.* occupancy) can also be calculated from the mobility data. This generates a time series for each destination where the occupancy increases by one at each detection time and decreases by one at each absence time. One requirement for calibration is that values for each predictor for all destinations are available for each decision time. However, detectors do not take measurements at a constant frequency, only when a device arrives or leaves its range. This means that different detectors are likely to have slightly different measurement times. Therefore, the occupancy time series of each destination are supplemented with any measurement times which are present in at least one other destination. The occupancy for these new times is interpolated as the occupancy for the last measurement made by the destination.

Consider an example with two detectors: detector 1 has measurements at 09:30:00, 09:30:02, 09:30:03, with occupancies 5, 6, 5, respectively, while detector 2 has measurements at 09:30:00, 09:30:01, 09:30:02 with occupancies 2, 3, 2, respectively. First, the measurement times for each detector are supplemented: 09:30:00, 09:30:01, 09:30:02, 09:30:03, then, the occupancies are interpolated. The resulting occupancies for the detectors are: 5, 5, 6, 5 for detector 1, and 2, 3, 2, 2 for detector 2.

Fig. 5(a) shows the supplemented occupancies for all destinations during Day 2, acting as a representative example for the other collection days. The occupancy of most destinations gradually increases over the first few hours of recording and decrease in the final hours of recording, with a distinct saw-tooth structure displayed throughout. Wablief is unique in that it does not show any such structure and remains constantly low throughout the collection period. Perhaps it was not a popular destination or maybe the coverage of this detector is smaller than the others.

Fig. 5(a) shows periods of time when no measurements are taken by some detectors, e.g. Club between roughly 17:45:00 and 23:15:00, with destination Main being missing entirely. This is likely to be due to detector malfunctions, as alluded to in Bonné *et al.* [62]. Choices made during these periods must therefore be discounted from any data used in model calibration. These are prevalent for at least one destination over all data collection days, so only data from a period of time outside any of these gaps is chosen for calibration. The period between 11:00:00-15:36:11 on Day 2 is chosen (highlighted in Fig. 5) as it is the longest uninterrupted measurement period over the entirety of data collection. It also starts when the detectors start taking measurements, making it easier to segregate from the rest of the data. Missing destinations artificially reduce the choice set, altering the calculated choice probabilities. There is no way to avoid or correct this with the data provided, as Days 2 and 3 have at least one destination missing and the Shelter detector for Day 1 seems to malfunction for the majority of its data collection time.

This truncation of available time during Day 2 must also be reflected in the destination sequences. Therefore, the sequence for each device is truncated such that only decision times within the time window are included. As before with the other stages of sequence processing, any devices with insufficient choices after this are discarded. This applies to both the clean and conflicted versions of the sequences.

Fig. 5(a) also shows that almost every detector (except Wablief) displays ‘saw-tooth’ behaviour during their operation over all collection days. These peaks show that the average rate at which devices leave a destination can be up to around 80 devices per second, which seems unrealistic, despite the coverage of detectors being unknown. These peaks do not correspond to the performance schedule of destinations (see Fig. 6), so it is unlikely to be due to performances taking place. The correct functioning of the data analysis code was ascertained by completing the following tests. First, by verifying that the total number of devices observed over the course of the day does not exceed the predicted number of visitors. Second, by applying the code to four simple scenarios involving one detector over 20 seconds where the outcome over time is known beforehand. The first of these involves five devices arriving one after the other for the first 10 seconds and then leaving in the order they arrived over the last 10 seconds. The second scenario is identical to the first, however, in the latter half, one of the devices which leaves returns later before leaving again. The third scenario is identical to the first, but one of the devices never leaves the detector. Scenario four involves each device entering and leaving the detector with no other devices present. As mentioned previously, this could also be an issue with

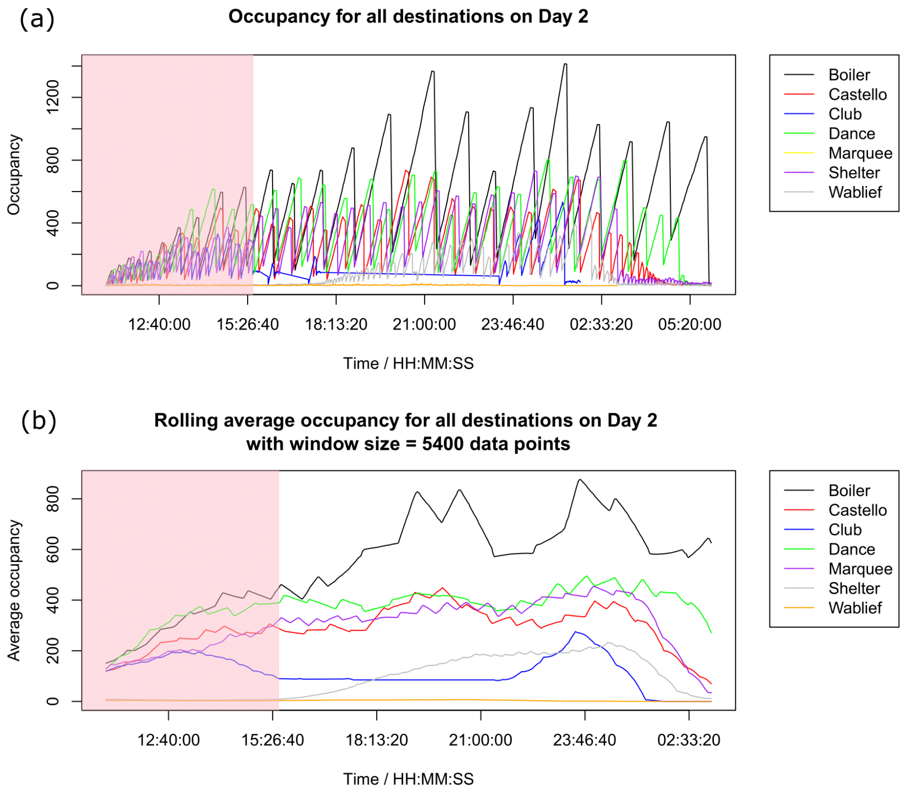

Figure 5 Raw (a) and smoothed (b) occupancy time series for Day 2 for all present destinations. The highlighted area shows the data used in model calibration.

the detectors themselves, but with the information available it is impossible to ascertain this.

These occupancy peaks suggest sharp and significant decreases in the number of devices at detectors, which is unrealistic. Therefore, the occupancy time series are smoothed using a rolling average, making the changes in occupancy less pronounced overall. The averaging window width was chosen such that the peak structure of the occupancies is smoothed out without losing the general trends. The window width was therefore chosen to be 5400 data points wide, with the resultant smoothed occupancies for each destination in Day 2 shown in Fig. 5b. The definition of the rolling average means that the smoothed time series are shorter than the originals, losing the last window width of data points. However, this is not a problem for calibration, which will use the occupancies from the first few hours of the day. This smoothing obfuscates the information contained within the raw occupancies, which will impact the occupancy parameter during calibration. These smoothed occupancies are normalised by dividing by the maximum observed occupancy for that destination on that day.

3.1.2 Distance

As mentioned in Sec. 2.2.3, the approximate latitude and longitude coordinates of each detector are provided. The Great-Circle distance [84] between each detector over the surface of the Earth can then be estimated. This is calculated using the Vincenty Ellipsoid method [85] using the values from WGS84 [86], providing Euclidean distances between each destination pair. These are constant over time as the detectors are assumed not to have moved. Before calibration, the distances are normalised by dividing by the largest calculated distance.

3.1.3 Desirability

As mentioned in Sec. 3, desirability represents a person's innate desire to visit a certain destination at a certain time. In the context of a music festival, the desirability of a venue could be whether a performance is taking place there. The desirability of other possible destinations, such as places to eat, or toilets, could be a quantity that reflects the individuals bodily needs. Since the destinations are only performance venues, desirability can be defined as whether a performance is occurring or predicted to occur by the time the venue is reached. No information is available about any attendee's preferences for different performances. Therefore, the desirability of each destination is a binary variable such that when a performance is scheduled to occur, the desirability is one, otherwise, it is zero. Using the performance schedules for each festival day, such as the one illustrated in Fig. 6, the binary desirability for each measurement time can be determined. The schedule does not provide exact start times for performances, so the start time is determined by eye to the nearest five minutes.

It is likely that venues will be of different sizes depending on the predicted popularity of their performances, such that larger venues will house more popular performances. Therefore, a larger venue could be, on average, more desirable than a smaller venue for any given individual. To account for this, the binary desirability of a destination is multiplied by the maximum occupancy for that destination observed over all days. The desirability is then normalised by dividing by the largest occupancy recorded over all destinations.

For example, consider a destination which has a performance starting at 12:00:00, the detector records device 1 arriving at 11:55:00, and device 2 at 12:00:05, the binary desirability for the destination for device 1 is zero, but one for device 2. Suppose the greatest measured occupancy over all destinations is four. The maximum observed occupancy at the destination is half that of the busiest recorded destination, so the desirability for device 1 is zero, while it is 0.5 for device 2.

3.1.4 Calibration

Once the extraction of the necessary information to calibrate the choice model specified in Eq. 1 and Eq. 2 was complete, the predictors for each destination must be matched and conflated with each choice made. The occupancy and desirability are both time-dependent, while the distance is constant. The distance for a given choice is the distance

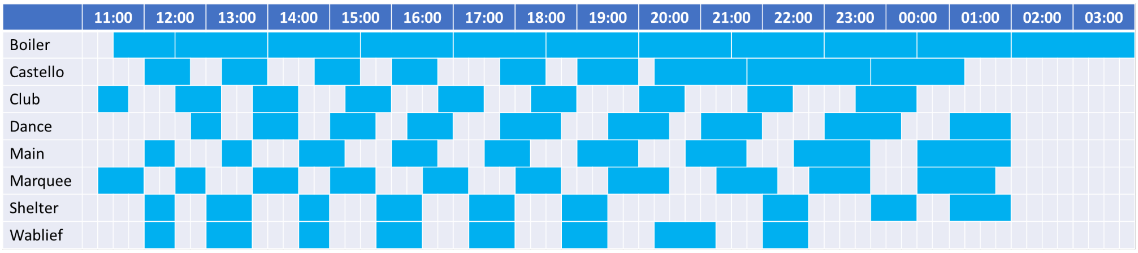

Figure 6 Illustration of the performance schedule for each destination on Day 2 of the festival. Blue areas indicate periods where a performance is taking place.

between the device's current destination and the chosen destination. The occupancy at the time at which the choice is assumed to be made (*i.e.* absence time) is used. The desirability at the detection time of the chosen destination is used. This is because it is assumed that a person who wishes to see a particular performance will leave their current destination with enough time to reach their desired destination.

Once the specified information has been extracted and collated from the raw data, model calibration can be performed. As described in Sec. 3.1.1, the choices made between 11:00:00 and 15:36:11 on Day 2, along with the predictors for all available destinations, is used as input. The clean and conflicted destination sequences are calibrated separately. The cleaned data contains 6567 choices spread over 2598 devices, while the conflicted data contains 9142 choices spread over 3050 devices. The three predictors; occupancy, distance, and desirability, are all normalised so that only the relative weighting of the model parameters affect the choice probabilities. Model calibration is conducted via Maximum Likelihood Estimation through the 'optim' function in R using the Nelder-Mead method. To obtain confidence intervals for the parameter estimates, bootstrap resampling of the calibration process is performed.

For each destination sequence version, the calibration process begins by selecting a random sample of 25% of the total available data. In the conflicted version, if a choice is selected that is part of a conflict, then the chosen destination is randomly allocated from the outcomes of the conflicted choices. Next, initial parameter values are chosen by sampling from a uniform distribution bounded between -5 and 5, which is fed into the numerical optimiser along with the selected data. To avoid local minima, this procedure is repeated 10 times and the final parameter estimates that generate the minimum negative log-likelihood are used as the result for the bootstrap replicate.

5000 bootstrap replicates are performed for each destination sequence version. The average value for each parameter is used as the final estimate for the calibration, with the 5% and 95% quantiles of the resultant distributions of each parameter over all bootstrap replicates taken as the confidence intervals.

3.2 Results

The parameter estimates, as well as their confidence intervals, for the clean and conflicted sequences are shown in Tab. 3 and Tab. 4, respectively.

Table 3 Model parameter estimates obtained from clean destination sequences, showing the upper and lower bootstrap confidence intervals and the bootstrap mean over 5000 replicates. Values are given to 3 d.p.

Clean	Lower bound	Mean	Upper bound
β_{occ}	1.082	1.394	1.724
β_{dist}	-2.172	-1.950	-1.737
β_{des}	0.078	0.238	0.399

Table 4 Model parameter estimates obtained from conflicted destination sequences, showing the upper and lower bootstrap confidence intervals and the bootstrap mean over 5000 replicates. Values are given to 3 d.p.

Conflicted	Lower bound	Mean	Upper bound
β_{occ}	1.470	1.686	1.912
β_{dist}	-1.527	-1.378	-1.229
β_{des}	0.394	0.513	0.629

In both versions, the occupancy and desire parameters are positive, while the distance is negative. This indicates that closer destinations which currently have more people and/or have a performance scheduled by the time the decision-maker arrives are more likely to be visited by the festival attendees who have a Wi-Fi-enabled device on their person. There is no overlap of the confidence intervals in both versions, suggesting that the weighting of each predictor is distinct. The confidence intervals also do not change sign, affirming the weighting the average individual gives each predictor. This makes sense intuitively, as people often try to minimise their walking distance/travel time between destinations and, in the context of a music festival, a destination is more attractive to an individual if there is a performance occurring. A destination with a large crowd could also be attractive to festival attendees, especially those with no plans, as a performance by more popular artists are likely to attract larger crowds. Also, perhaps festival attendees will be more naturally drawn to larger groups of other attendees, in order to share their experience and enjoyment with others.

The distance parameter has the largest magnitude, indicating that individuals consider this the most important predictor out of those presented. The relative magnitudes of the occupancy and desire parameters suggest that the average individual considers occupancy much more favourably than whether a performance is taking place. This could be due to how desirability was defined, being based on what was predicted to have happened, rather than what actually happened. It is possible that performances started/ended at different times or were even cancelled, both of which would have a significant impact on desirability. Also, the definition of desirability is strict and does not consider individuals hanging around destinations between performances. Plus, if an individual arrives earlier than when the performance is scheduled to begin, then the destination would still have a desirability of zero, even if the individual desired to visit. This definition of desirability also does not take into account individual preferences or personality. Some people are more impulsive than others, and it is fair to assume that most individuals will want to see certain performances more than others. Also, instead of scaling binary desirability of a destination by the maximum observed occupancy, which has its own problems (see below), it could be scaled by the popularity of the artist/group performing. Proxy measures for the popularity of artists/groups could be any arbitrary measure of success that can be easily obtained online, such as the number of iTunes and/or Spotify downloads, the number of albums/singles released, or number of followers on various social media platforms. This could be a better indication of the drawing power of each performance scheduled during the festival and thus be a better measure of desirability for the average festival attendee.

There is some difference in the parameter values between the clean and conflicted data, indicating that removing the conflicted choices does introduce a bias in the parameter estimates. When all choices are considered and conflicts are resolved at random, the occupancy and desirability parameters are generally higher and the distance parameter is generally lower. Now the average individual considers occupancy to be the most important, followed by distance, and then desirability. This shows that caution is needed when interpreting parameter magnitudes, as they change depending on the exact data used during calibration. However, the signs of the parameters are unchanged, so it seems that individuals see the distance as detrimental, while occupancy and desirability are both attractive, regardless of the data used for calibration.

The effect of rolling average window width used to smooth occupancies on the results of the calibration was investigated. The results are shown in App. 8. Though small differences in the magnitude of the occupancy parameter were detected, the sign and order of parameter magnitudes remained unchanged for both clean and conflicted data.

In addition to the issues mentioned above and in Sec. 3.1, there are several other assumptions and simplifications made during data processing that may have influenced the results. For instance, only the Euclidean distance between destinations can be calculated. But the possibility of interim stops and the environmental layout means that this is not necessarily equal to the actual distance travelled by any individual. The actual distance travelled cannot be gleaned from the data, but a potentially more accurate value of distance travelled by a device could be obtained by taking into account any interim detections of the device. For example, consider a device that visits Dance and then Club, but is detected at Ingang in between. The distance travelled can then be estimated as the distance between Dance and Ingang plus the distance between Ingang and Club. However, due to presence conflicts and the fact that the coverage of each detector is unknown, any subsequent model will still need to consider distance carefully.

The method for determining whether an individual has visited a destination does not take into account the prevailing conditions at the destination, such as crowd density, which can impact the time taken for a device to pass through a detector's coverage area. It is therefore possible to misidentify devices as visiting, artificially increasing the number of destinations chosen. If the coverage of each detector was known, then the minimum time could be weighted such that detectors with larger coverage would have a larger minimum time. A more accurate minimum time for visiting could be calculated if the coverage of each detector and the local pedestrian densities over time were also known, but this is impossible to do accurately with the data available. Alternative definitions of visiting a destination have been defined based on the speed of an individual device [87], but this is not available from the data and could still lead to misidentifications.

The total number of individuals detected at a destination depends on the coverage of the associated detector, where a larger coverage will likely detect more individuals than a smaller coverage. This could partly explain why Boiler consistently shows the largest occupancies, despite its size being comparable to other destinations. Arguably, the performances anticipated to be more popular would be held on the Main stage, where the large open area should allow the largest crowds to form, yet this does not appear to be reflected in the occupancies. Without knowing each detector's coverage, it is unclear if the number of individuals detected is a valid measure of the occupancy. Also, there is no information on how many people actually attended the music festival or attended each destination. The organisers predicted up to 100,000 attendees and Bonné et al. report that 29,296 unique devices were detected (after filtering out stationary devices and devices from people outside the festival) [62] leading to an average of 29.3% of people possessing a Wi-Fi-enabled device, assuming this proportion is constant throughout the festival, then the number of detected devices could serve as a suitable proxy for the actual occupancy. One other thing this assumption neglects is the possibility that individuals could have more than one Wi-Fi-enabled device or that such a device could be shared among a group. The former could, in theory, be detected if two or more unique devices are detected arriving and leaving the same sequence of destinations at roughly the same time.

4 Discussion

In this section, the general issues of using the datasets as reference data for model development are discussed. These include: the compromises required in repurposing data, the quality of the data itself and that of the accompanying technical description, the accessibility of the data, and the extent to which the data has been processed.

Attempting to repurpose data that was collected with different aims, scope, and context in mind can be challenging. Often the information that is needed for any new application is not directly available from the data, if at all. Therefore, it falls to the researcher to consider how to extract this information and what caveats and assumptions must be made during the process.

For example, the Bonné data was collected with several applications in mind: real-time crowd monitoring at mass events, supporting opportunistic communications network simulations, and use in the development of 'smart' buildings [62]. Yet in this work, it is used to infer the impact of certain generic predictors on individual destination choice, which required extensive processing before it could be used (see Sec. 3). These processes introduce additional sources of error that would not otherwise be present if new data was collected with said purpose in mind. For example, extracting destination sequences and occupancies from King's data is straightforward, but not for Bonné's data.

This is not so great an issue in Danalet's data, as it was collected with similar research aims in mind - analysis of activity choice models. Despite this, some predictors given in Eq. 2 are not present in the data. For example, real-time occupancies are not available, and must be inferred either from the presence of detected individuals, or using estimated/assumed capacities, both of which have issues. The individuals in the data may not homogeneously mix with other campus users across both time and space, making it unclear whether the number of detected individuals is correlated with the actual number of individuals present. Using the capacities of destinations instead of occupancy is also problematic, as they are unlikely to be representative over all time.

Tinguely's data requires little in the way of repurposing, as the main aim is exploration of catering destination choice models. In comparison to Danalet's and Bonné's data, there is expected to be a lot less processing needed in order to extract the necessary information to calibrate a different choice model, such as the one specified in Eq. 1 and Eq. 2. That being said, the issues with extracting occupancies are similar to those from Danalet's data, again highlighting the issues with inferring new information from existing data. These problems are not present when analysing the King data, where the whole scope of data collection revolves around the destination choice model specified in this work. However, any alternative model specification could run into these kinds of issues.

The quality of the reference data is important as it can determine the scope of its use in model development. Model calibration, and often validation, is based on data, and the quality of the resulting calibration directly depends on the quality of the data supplied and what kind of errors are present [66].

As described in previous sections, Bonné's data has many issues in data quality that arose during the case study. This is probably because the data was collected to illustrate the capabilities of a new detection system, where unexpected errors and malfunctions are likely to occur. Here, these issues become evident when attempting to derive chosen destination sequences and occupancies (Sec. 3.1.1), with some devices being detected in two places at once and large gaps in recorded measurements. These problems are so significant that great care is needed when interpreting the results of the case study.

Danalet's data also has some issues, particularly with the Wi-Fi data; a small proportion of measurements have missing data, however, these can be ignored without impacting subsequent analysis. Other potential issues with this data might arise if analysed in more detail.

In King's data, the basic information for a couple of the experimental conditions is missing. This was due to a technical error when implementing the surveys. Also, for one particular experimental condition, the survey was altered after being released, so the data after that time cannot be included in subsequent analyses. These issues are described in detail in King and Bode.

Tinguely's data does not appear to have any quality issues, possibly because it is the result of processing the raw Danalet data. Though potential issues with this data might arise if analysed in more detail.

A full and clear explanation of reference data is needed to assess its suitability for a set of research objectives and make it easier to understand, explore, and manipulate. This should include; how the data was collected and in what context and experimental conditions, what is available, and in what format. This can be difficult and time-consuming to do, especially if there is a lot of different data which are used in a variety of ways. Such appears to be the case with Danalet's data, where there is detailed and extensive descriptions of the Wi-Fi mobility data, but not so much for the Potential Attractivity Measures (PAMs). Also, there is also some discrepancies between the data as it has been reported in the published work [59] and what is present in the dataset. This makes it difficult to understand where and when this data was collected, and how it is used in the original work.

The other datasets are all adequately explained for researchers to begin using them, perhaps because they contain fewer kinds of data and information.

Reference data should be stored in accessible formats, so that anyone with different backgrounds and different skills can make use of it. Any reference data should therefore be in easily-read formats which see widespread use in programming languages, analytical software, and operating systems. Part of Danalet's data requires knowledge of PostgreSQL[10], and the requisite software, in order to be accessed. While this is a standard format used by researchers for data analysis and processing, it may be even more accessible to the wider public if it were stored in text or Comma-Separated Value (CSV) formats.

Finally, reference data should be presented in its most raw form, as this allows the greatest flexibility in its use. Any pre-processing that has been done can introduce inherent assumptions and uncertainties that have to be worked around by subsequent researchers. Tinguely's data is one such example. However, if the original purpose of such data aligns closely to the intentions of the researcher, this can be an advantage, as the data will typically be cleaner and require fewer additional processing steps.

To summarise, there are many potential issues with creating and using reference data, and each of the datasets included in this database exemplify one or more or these issues. The Bonné and Danalet data need considerable re-purposing to be used in the context of pedestrian destination choice calibration. Both suffer issues with data quality, where data is either missing or not inaccurate, but this can be readily taken care of after preprocessing, for example, Tinguely's data does not suffer these issues as it is the result of processing Danalet's data.. The explanation of Danalet's data is also incomplete and it is difficult to cross-reference the properties of the data provided versus what was reported in the literature. Part of Danalet's data could also be stored in even more accessible formats. In order to provide the most flexibility in potential applications, reference data should be in its rawest form, with as little preprocessing done as possible. These issues do not devalue the data, they are simply representative of data colection challenges with a given technology. Ultimately, it is up to the researcher to decide which, if any, datasets currently available in this database are the most suitable given their goals.

[10] IBM: Structured Query Language, URL: https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/db2/10.5?topic=reference-sql, accessed: 11/07/2022

5 Data Locations

This section details where each dataset described in the database can be found. Danalet's data can be found on Zenodo. It has the following Digital Object Identifier (DOI): 10.5281/zenodo.15798. Tinguely's data can be found on Zenodo. It has the following DOI: 10.5281/zenodo.1038622. Bonné's original and processed data, along with the R code used to process and anlyse the data as described in Section Sec. 3 can be found on the University of Bristol's Research Data Storage Facility here. It has the following DOI: 10.5523/bris.7ob8ukji8iwp2n0l94n4ykc08. King's data can be found on the University of Bristol's Research Data Storage Facility here. It has the following DOI: 10.5523/bris.249d43dprgg8x2td33cdmg8kax.

6 Conclusion

Openly-available reference databases have benefitted model development in several research fields, including pedestrian mobility and route choice behaviour, such a resource is not currently available in pedestrian destination choice behaviour. A literature search reveals that while data for development of destination choice models have been collected by researchers, only 2 or 3 examples were discovered. This paper attempts to build a reference database for use primarily in the development of pedestrian destination choice models. The data collated for this database are described and the potential ways in which they could be used is illustrated through a case study. This revealed several general issues around creating and using reference data in model development and these are discussed in detail in the context of the other datasets. Ultimately this work hopes to form the basis for an extensive reference database in the field of pedestrian destination choice and provide guidance on future publication of reference data. It is hoped that this database will grow in the future.

Acknowledgements The authors would like to thank Professor Eddie Wilson for several insightful discussions which helped shape this work. Special thanks goes to Professor Wim Lamotte and the other co-authors of the Bonné data for making their data available for analysis. The authors also appreciate the support from all researchers who responded during the data search, particularly Professor Aloy Borgers for being willing to share his data.

Christopher King acknowledges funding from the Engineering and Physical Sciences Research Council (EPSRC) in the form of a doctoral scholarship.

Ethics Statement This study repurposes openly published and fully anonymised data.

Author Contributions Christopher King: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review and editing, Visualisation / Nikolai Bode: Conceptualisation, Methodology, Validation, Data curation, Writing - review and editing, Supervision, Funding acquisition

References

[1] Miller J., H.: Activity-Based Analysis. In: Fischer, M.M., Nijkamp, P. (eds.) Hand book of Regional Science, vol. 2, pp. 705–724. Springer Berlin Heidelberg, Berlin (2014). DOI: 10.1007/978-3-662-60723-7_106

[2] Institute of Advanced Simulation: Pedestrian Dynamics Data Archive (2017). URL https://ped.fz-juelich.de/da, accessed: 15/07/2022

[3] Zhou, B., Tang, X., Wang, X.: Measuring Crowd Collectiveness. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '13, pp. 3049–3056. IEEE Computer Society, USA (2013). DOI: 10.1109/CVPR.2013.392

[4] Zhou, B.: Collective Motion Database (2013). URL http://mmlab.ie.cuhk.edu.hk/projects/collectiveness/dataset.htm, accessed: 15/07/2022

[5] Wang, L., Li, G., Lei, J., Wang, T., Zhang, Y.: Density-Based Manifold Collective Clustering for Coherent Motion Detection. In: Proceedings of the International Conference on Machine Vision and Applications, ICMVA 2018, pp.22–27. Association for Computing Machinery, New York, NY, USA (2018). DOI: 10.1145/3220511.3220521

[6] Gao, M.L., Wang, Y.T., Jiang, J., Shen, J., Zou, G.F., Liu, L.N.: Crowd motion segmentation via streak flow and collectiveness. In: 2017 Chinese Automation Congress (CAC), pp. 4067–4070 (2017). DOI: 10.1109/CAC.2017.8243492

[7] Liu, W., Chan, A.B., Lau, R.W.H., Manocha, D.: Leveraging Long-Term Predictions and Online Learning in Agent-Based Multiple Person Tracking. IEEE Transactions on Circuits and Systems for Video Technology **25**(3), 399–410 (2015). DOI: 10.1109/TCSVT.2014.2344511. Conference Name: IEEE Transactions on Circuits and Systems for Video Technology

[8] Lin, W., Mi, Y., Wang, W., Wu, J., Wang, J., Mei, T.: A Diffusion and Clustering-Based Approach for Finding Coherent Motions and Understanding Crowd Scenes. IEEE Transactions on Image Processing **25**(4), 1674–1687 (2016). DOI: 10.1109/TIP.2016.2531281. Conference Name: IEEE Transactions on Image Processing

[9] Wu, S., Yang, H., Zheng, S., Su, H., Fan, Y., Yang, M.H.: Crowd Behavior Analysis via Curl and Divergence of Motion Trajectories. International Journal of Computer Vision **123**(3), 499–519 (2017). DOI: 10.1007/s11263-017-1005-y

[10] Mueller, J.P., Massaron, L.: Machine Learning for Dummies. John Wiley & Sons, Incorporated, Hoboken, United States (2016). URL http://ebookcentral.proquest.com/lib/bristol/detail.action?docID=4526803

[11] Kotz, D., McDonald, C., Henderson, T., Gaughan, S., Proctor, P., Shubina, A., Yeo, J., Akestoridis, D.G., Cannon, H., Dimov, R., Jacobson, J., Johnson, S., Kelk, K., Malajian, D., Muckle-Jones, R., Neophytou, C., Pietikainen, P., Riina, A., Seletsky, O.: CRAWDAD. URL https://crawdad.org/about.html, accessed: 31/01/2022

[12] Sarafraz, A.: Datasets (2020). URL https://computervisiononline.com/datasets, accessed: 31/01/2022

[13] University College San Diego: Metabolomics Workbench. URL https://www.metabolomicsworkbench.org/, accessed: 14/07/2022

[14] Seyfried, A., Passon, O., Steffen, B., Boltes, M., Rupprecht, T., Klingsch, W.: New Insights into Pedestrian Flow Through Bottlenecks. Transportation Science **43**(3), 395–406 (2009). DOI: 10.1287/trsc.1090.0263. Publisher: INFORMS

[15] Godel, M., Bode, N.W.F., K ¨ oster, G., Bungatz, H.J.: Bayesian inference methods to calibrate crowd dynamics models for safety applications. Safety Science **147**, 105586 (2022). DOI: 10.1016/j.ssci.2021.105586. Publisher: Elsevier

[16] Cao, R.F., Lee, E.W.M., Yuen, A.C.Y., Chen, T.B.Y., De Cachinho Cordeiro, I.M., Shi, M., Wei, X., Yeoh, G.H.: Simulation of competitive and cooperative egress movements on the crowd emergency evacuation. Simulation Modelling Practice and Theory **109**, 102309 (2021). DOI: 10.1016/j.simpat.2021.102309

[17] Haghpanah, F., Ghobadi, K., Schafer, B.W.: Multi-hazard hospital evacuation planning during disease outbreaks using agent-based modeling. International Journal of Disaster Risk Reduction **66**, 102632 (2021). DOI: 10.1016/j.ijdrr.2021.102632

[18] Li, N., Guo, R.Y.: Simulation of bi-directional pedestrian flow through a bottleneck: Cell transmission model. Physica A: Statistical Mechanics and its Applications **555**, 124542 (2020). DOI: 10.1016/j.physa.2020.124542

[19] Zhuang, Y., Liu, Z., Schadschneider, A., Yang, L., Huang, J.: Exploring the behavior of self-organized queuing for pedestrian flow through a non-service bottleneck. Physica A: Statistical Mechanics and its Applications **562**, 125186 (2021). DOI: 10.1016/j.physa.2020.125186

[20] Li, H., Zhang, J., Yang, L., Song, W., Yuen, K.K.R.: A comparative study on the bottleneck flow between preschool children and adults under different movement motivations. Safety Science **121**, 30–41 (2020). DOI: 10.1016/j.ssci.2019.09.002

[21] Cao, S., Seyfried, A., Zhang, J., Holl, S., Song, W.: Fundamental diagrams for multidirectional pedestrian flows. Journal of Statistical Mechanics: Theory and Experiment **2017**(3), 033404 (2017). DOI: 10.1088/1742-5468/aa620d. Publisher: IOP Publishing

[22] Xu, Q., Chraibi, M., Seyfried, A.: Anticipation in a velocity-based model for pedestrian dynamics. Transportation Research Part C: Emerging Technologies **133**, 103464 (2021). DOI: 10.1016/j.trc.2021.103464

[23] Zhang, J., Klingsch, W., Schadschneider, A., Seyfried, A.: Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions. Journal of Statistical Mechanics: Theory and Experiment **2011**(06), P06004 (2011). DOI: 10.1088/1742-5468/2011/06/P06004. Publisher: IOP Publishing

[24] Rathinakumar, K., Quaini, A.: A microscopic approach to study the onset of a highly infectious disease spreading. Mathematical Biosciences **329**, 108475 (2020). DOI: 10.1016/j.mbs.2020.108475

[25] Zhao, X., Zhang, J., Song, W.: A Radar-Nearest-Neighbor based data-driven approach for crowd simulation. Transportation Research Part C: Emerging Technologies **129**, 103260 (2021). DOI: 10.1016/j.trc.2021.103260

[26] Munafo, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., Ioannidis, J.P.A.: A manifesto for reproducible science. Nature Human Behaviour **1**(1), 1–9 (2017). DOI: 10.1038/s41562-016-0021. Number: 1 Publisher: Nature Publishing Group

[27] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data **3**(1), 160018 (2016). DOI: 10.1038/sdata.2016.18. Number: 1 Publisher: Nature Publishing Group

[28] Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, T.P.: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine **6**(7), e1000097 (2009). DOI: 10.1371/journal.pmed.1000097. Publisher: Public Library of Science

[29] Wee, B.V., Banister, D.: How to Write a Literature Review Paper? Transport Reviews **36**(2), 278–288 (2016). DOI: 10.1080/01441647.2015.1065456. Publisher: Routledge

[30] Kitchenam, B., Charters, S.: Guidelines for performing Systematic Literature Reviews in software engineering. Technical, University of Durham, Durham,

UK (2007). URL https://www.researchgate.net/publication/258968007_Kitchenham_B_Guidelines_for_performing_Systematic_Literature_Reviews_in_software_engineering_EBSE_Technical_Report_EBSE-2007-01

[31] . Greenhalgh, T., Peacock, R.: Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ (Clinical research ed.) **331**(7524), 1064–1065 (2005). DOI: 10.1136/bmj.38636.593461.68

[32] Feng, Y., Duives, D.C., Daamen, W., Hoogendoorn, S.P.: Data collection methods for studying pedestrian behaviour: A systematic review. Building and Environment **187**, 107329 (2021). DOI: 10.1016/j.buildenv.2020.107329

[33] Seaborn, K., Miyake, N.P., Pennefather, P., Otake-Matsuura, M.: Voice in Human–Agent Interaction. ACM Computing Surveys **54**(4), 1–43 (2022). DOI: 10.1145/3386867

[34] Xu, M., Nie, X., Li, H., Cheng, J.C., Mei, Z.: Smart construction sites: A promising approach to improving on-site HSE management performance. Journal of Building Engineering **49**, 104007 (2022). DOI: 10.1016/j.jobe.2022.104007

[35] Jalali, S., Wohlin, C.: Systematic literature studies: Database searches vs. backward snowballing. In: Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, pp. 29–38. IEEE, Lund, Sweden (2012). DOI: 10.1145/2372251.2372257

[36] Ying, F., Wallis, A.O.G., Beguerisse-D´ıaz, M., Porter, M.A., Howison, S.D.: Customer mobility and congestion in supermarkets. Physical Review E **100** (2019). DOI: 10.1103/PhysRevE.100.062304. ArXiv: 1905.13098

[37] Arentze, T.A., Timmermans, H.J.P.: Deriving performance indicators from models of multipurpose shopping behavior. Journal of Retailing and Consumer Services **8**(6), 325–334 (2001). DOI: 10.1016/S0969-6989(00)00038-2

[38] Beaulieu, A., Farooq, B.: A dynamic mixed logit model with agent effect for pedestrian next location choice using ubiquitous Wi-Fi network data. International Journal of Transportation Science and Technology **8**(3), 280–289 (2019). DOI: 10.1016/j.ijtst.2019.02.003

[39] Ettema, D., Bastin, F., Polak, J., Ashiru, O.: Modelling the joint choice of activity timing and duration. Transportation Research Part A: Policy and Practice **41**(9), 827–841 (2007). DOI: 10.1016/j.tra.2007.03.001

[40] Hui, S.K., Bradlow, E.T., Fader, P.S.: Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior. Journal of Consumer Research **36**(3), 478–493 (2009). DOI: 10.1086/599046

[41] Danalet, A.: Activity choice modeling for pedestrian facilities. Ph.D. thesis, Swiss Federal Institute of Technology, Lausanne (2015). URL https://infoscience.epfl.ch/record/214544#

[42] Tinguely, L.: Exploiting pedestrian WiFi traces for destination choice modeling. Master's thesis, EPFL, Lausanne, Switzerland (2015). URL https://infoscience.epfl.ch/record/209732

[43] Vukadinovic, V., Dreier, F., Mangold, S.: A simple framework to simulate the mobility and activity of theme park visitors. In: Proceedings of the 2011 Winter Simulation Conference (WSC), pp. 3248–3260 (2011). DOI: 10.1109/WSC.2011.6148022

[44] Yao, R., Bekhor, S.: Data-driven choice set generation and estimation of route choice models. Transportation Research Part C: Emerging Technologies **121**, 102832 (2020). DOI: 10.1016/j.trc.2020.102832

[45] Zhu, W., Timmermans, H.J.P.: Modeling pedestrian shopping behavior using principles of bounded rationality: model comparison and validation. Journal of Geographical Systems **13**(2), 101–126 (2011). DOI: 10.1007/s10109-010-0122-8

[46] Duives, D.C., Daamen, W., Hoogendoorn, S.P.: State-of-the-art crowd motion simulation models. Transportation Research Part C: Emerging Technologies **37**, 193–209 (2013). DOI: 10.1016/j.trc.2013.02.005

[47] Feliciani, C., Nishinari, K.: Estimation of pedestrian crowds' properties using commercial tablets and smartphones. Transportmetrica B: Transport Dynamics **7**(1), 865–896 (2019). DOI: 10.1080/21680566.2018.1517061. Publisher: Taylor & Francis

[48] Voskamp, H.a.W.: Measuring the influence of congested bottleneck on route choice behavior of pedestrians at Utrecht Centraal. Master's thesis, Delft University of Technology, Delft (2012). URL http://resolver.tudelft.nl/uuid:6a272595-d98b-4769-99b4-1eec7ab2b620

[49] Bigano, A., Hamilton, J.M., Tol, R.S.J.: The Impact of Climate on Holiday Destination Choice. Climatic Change **76**(3), 389–406 (2006). DOI: 10.1007/s10584-005-9015-0

[50] Karl, M., Reintinger, C., Schmude, J.: Reject or select: Mapping destination choice. Annals of Tourism Research **54**, 48–64 (2015). DOI: 10.1016/j.annals.2015.06.003

[51] Fotheringham, A.S.: Modelling Hierarchical Destination Choice. Environment and Planning A: Economy and Space **18**(3), 401–418 (1986). DOI: 10.1068/a180401

[52] Andion, J., Navarro, J.M., L ´ opez, G., ´ Alvarez Campana, M., Due ´ 34 nas, J.C.: Smart Behavioral Analytics over a Low-Cost IoT Wi-Fi Tracking Real Deployment. Wireless Communications and Mobile Computing **2018**, e3136471 (2018). DOI: 10.1155/2018/3136471. Publisher: Hindawi

[53] Raun, J., Ahas, R., Tiru, M.: Measuring tourism destinations using mobile tracking data. Tourism Management **57**, 202–212 (2016). DOI: 10.1016/j.tourman.2016.06.006

[54] Andresen, E., Chraibi, M., Seyfried, A.: A representation of partial spatial knowledge: a cognitive map approach for evacuation simulations. Transportmetrica A: Transport Science **14**(5-6), 433–467 (2018). DOI: 10.1080/23249935.2018.1432717

[55] Tong, Y., Bode, N.W.F.: Higher investment levels into pre-planned routes increase the adherence of pedestrians to them. Transportation Research Part F: Traffic Psychology and Behaviour **82**, 297–315 (2021). DOI: 10.1016/j.trf.2021.07.019

[56] Shao, W., Terzopoulos, D.: Autonomous Pedestrians. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA ’05, pp. 19–28. ACM, New York, NY, USA (2005). DOI: 10.1145/1073368.1073371

[57] Axhausen, K.W., Zimmermann, A., Schonfelder, S., Rindsf ¨ user, G., Haupt, T.: Observing the rhythms of daily life: A six-week travel diary. Transportation **29**(2), 95–124 (2002). DOI: 10.1023/A:1014247822322

[58] Shiftan, Y.: Practical Approach to Model Trip Chaining. Transportation Research Record **1645**(1), 17–23 (1998). DOI: 10.3141/1645-03

[59] Danalet, A., Farooq, B., Bierlaire, M.: A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies **44**, 146–170 (2014). DOI: 10.1016/j.trc.2014.03.015

[60] Danalet, A., Bierlaire, M.: A path choice approach to activity modeling with a pedestrian case study. p. 45. Unpublished, Monte Verita, Ascona, Switzerland (2014). DOI: 10.13140/2.1.4099.8728

[61] Danalet, A., Tinguely, L., Lapparent, M.d., Bierlaire, M.: Location choice with longitudinal WiFi data. Journal of Choice Modelling **18**, 1–17 (2016). DOI: 10.1016/j.jocm.2016.04.003

[62] Bonne, B., Barzan, A., Quax, P., Lamotte, W.: WiFiPi: Involuntary tracking of visitors at mass events. In: 2013 IEEE 14th International Symposium on “A World of Wireless, Mobile and Multimedia Networks” (WoWMoM), pp. 1–6. IEEE, Madrid, Spain (2013). DOI: 10.1109/WoWMoM.2013.6583443

[63] King, C., Bode, N.W.F.: A virtual experiment on pedestrian destination choice: the role of schedules, the environment and behavioural categories. Royal Society Open Science **9**(7), 211982 (2022). DOI: 10.1098/rsos.211982

[64] Train, K.: Discrete choice methods with simulation, 2nd ed edn. Cambridge University Press, Cambridge; New York (2009). URL https://eml.berkeley.edu/books/train1201.pdf.OCLC:ocn349248337

[65] Hoogendoorn, S.P., Bovy, P.H.L.: Pedestrian route-choice and activity scheduling theory and models. Transportation Research Part B: Methodological **38**(2), 169–190 (2004). DOI: 10.1016/S0191-2615(03)00007-9

[66] King, C., Koltsova, O., Bode, N.W.F.: Simulating the effect of measurement errors on pedestrian destination choice model calibration. Transportmetrica A: Transport Science pp. 1–41 (2022). DOI: 10.1080/23249935.2021.2017510. Publisher: Taylor & Francis

[67] Kielar, P.M., Borrmann, A.: Modeling pedestrians' interest in locations: A concept to improve simulations of pedestrian destination choice. Simulation Modelling Practice and Theory **61**, 47–62 (2016). DOI: 10.1016/j.simpat.2015.11.003

[68] Saarloos, D., Fujiwara, A., Zhang, J.: The Interaction between Pedestrians and Facilities in Central Business Districts: An Explorative Case Study. Journal of the Eastern Asia Society for Transportation Studies 7(0), 1870–1885 (2007). DOI: 10.11175/easts.7.1870. Publisher: Eastern Asia Society for Transportation Studies

[69] Kwak, J., Jo, H.H., Luttinen, T., Kosonen, I.: Modeling Pedestrian Switching Behavior for Attractions. Transportation Research Procedia **2**, 612–617 (2014). DOI: 10.1016/j.trpro.2014.09.102

[70] Arentze, T.A., Ettema, D., Timmermans, H.J.P.: Location choice in the context of multi-day activity-travel patterns: model development and empirical results. Transportmetrica A: Transport Science **9**(2), 107–123 (2013). DOI: 10.1080/18128602.2010.538870

[71] Ettema, D., Borgers, A.W.J., Timmermans, H.J.P.: Simulation model of activity scheduling behavior. Transportation Research Record **1413**, 1–11 (1993). URL https://www.persistent-identifier.nl/urn:nbn:nl:ui:25-0287ad73-721b-4094-962c-fa7cc87db416

[72] Fesenmaier, D.R.: Integrating activity patterns into destination choice models. Journal of Leisure Research **20**(3), 175–191 (1988)

[73] Kurose, S., Borgers, A.W.J., Timmermans, H.J.P.: Classifying Pedestrian Shopping Behaviour According to Implied Heuristic Choice Rules. Environment and Planning B: Planning and Design **28**(3), 405–418 (2001). DOI: 10.1068/b2622

[74] Ton, D.: Navistation: A study into the route and activity location choice behaviour of departing pedestrians in train stations. Ph.D. thesis, Delft University of Technology (2014)

[75] van der Hagen, X., Borgers, A.W.J., Timmermans, H.J.P.: Spatiotemporal Sequencing Processes of Pedestrian in Urban Retail Environments. The Journal of the RSAI **70**(1), 37–52 (1991). DOI: 10.1007/BF01463442

[76] Zhu, W., Timmermans, H.J.P., De, W.: Temporal Variation in Consumer Spatial Behavior in Shopping Streets. Journal of Urban Planning and Development **132**(3), 166–171 (2006). DOI: 10.1061/(ASCE)0733-9488(2006)132:3(166)

[77] Fahmy, S.A., Alablani, B.A., Abdelmaguid, T.F.: Shopping center design using a facility layout assignment approach. In: 2014 9th International Conference on Informatics and Systems, pp. ORDS–1–ORDS–7. IEEE, Cairo, Egypt (2014). DOI: 10.1109/INFOS.2014.7036689

[78] Yang, Y., Fik, T., Zhang, J.: Modeling Sequential Tourist Flows:Where is the Next Destination? Annals of Tourism Research **43**, 297–320 (2013). DOI: 10.1016/j.annals.2013.07.005

[79] Dijkstra, J., Timmermans, H.J.P., de Vries, B.: Activation of Shopping Pedestrian Agents—Empirical Estimation Results. Applied Spatial Analysis and Policy **6**(4), 255–266 (2013). DOI: 10.1007/s12061-012-9082-3

[80] Greenwood, D., Sharma, S., Johansson, A.: Mobility Modelling in a Process Constrained Environment: Modelling the Movements of Nurses in a Neonatal Intensive Care Unit. In: Chraibi, M., Boltes, M., Schadschneider, A., Seyfried, A. (eds.) Traffic and Granular Flow '13, pp. 233–241. Springer International Publishing, Cham (2015). URL https://www.semanticscholar.org/paper/Mobility-Modelling-in-a-Process-Constrained-the-of-Greenwood-Sharma/f8a8c139eb0657f6f2863df4419ed6cd43b8a11

[81] R Core Team: R: A language and environment for statistical computing (2020). URL https://www.R-project.org/

[82] Cisco Meraki: Understanding Wireless Performance and Coverage. Technical (2020). URL https://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/Understanding_Wireless_Performance_and_Coverage

[83] Chilipirea, C., Dobre, C., Baratchi, M., Steen, M.v.: Identifying Movements in Noisy Crowd Analytics Data. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM), pp. 161–166 (2018). DOI: 10.1109/MDM.2018.00033. ISSN: 2375-0324

[84] Kern, W.: Solid Mensuration With Proofs (1938). URL http://archive.org/details/in.ernet.dli.2015.205959

[85] Vincenty, T.: Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations. Survey Review **23**(176), 88–93 (1975). DOI: 10.1179/sre.1975.23.176.88. Publisher: Taylor & Franci

[86] National Geospatial-intelligence Agency: NGA Geomatics - WGS 84. URL https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs84

[87] Zhai, Y., Baran, P.K., Wu, C.: Spatial distributions and use patterns of user groups in urban forest parks: An examination utilizing GPS tracker. Urban Forestry & Urban Greening **35**, 32–44 (2018). DOI: 10.1016/j.ufug.2018.07.014

7 Bonné’s Data after each Step of the Data Analysis Procedure

This Appendix gives quantitative information at the various stages of the destination sequence extraction process on the Bonné mobility data as described in Sec. 3.1. This supplementary information will hopefully guide any reader who wishes to repeat and reproduce the work described in Sec. 3.1 and Sec. 3.2. Only the Day 2 data is truncated, so only the that data is affected by the truncation process, as shown in Tab. 5 and Tab. 6.

Table 5 How the total number of devices changes with each stage of processing of the Bonné mobility data.

Day	**Detected at** ≥ 1 **destination**	**Detected at** > 1 **destination**	**Visiting** > 1 **destination**	**After** **cleaning**	**After** **truncation**	**After cleaning** **and truncation**
-1	12351	3379	2914	2344	2914	2344
1	32362	9822	9193	8101	9193	8101
2	31907	9257	8646	7569	3050	2598
3	30133	9092	8294	6930	8294	6930

Table 6 How the total number of choices changes with each stage of processing of the Bonné mobility data.

Day	**Detected at** ≥ 1 **destination**	**Detected at** > 1 **destination**	**Visiting** > 1 **destination**	**After** **cleaning**	**After** **truncation**	**After cleaning** **and truncation**
-1	42273	12192	9823	7263	9823	7263
1	127273	40810	35657	26371	35657	26371
2	126056	40567	34946	25133	9142	6567
3	110206	34951	30207	20416	30207	20416

8 How Model Parameter Estimates Change with Occupancy Smoothing Window Width

This Appendix displays how the values of the occupancy, distance, and desire parameters of the model specified in Sec. 3 vary with the window width of the occupancy smoothing average. Results are shown for calibration on both cleaned and conflicted destination sequences. The average values and 95% confidence intervals from 5000 bootstrap replicates are shown.

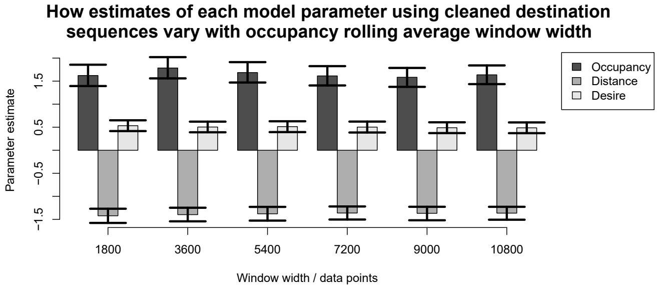

Figure 7 Model parameter estimates calibrated on cleaned destination sequences using different rolling average window widths for smoothing occupancies. Error bars represent the 95% confidence bootstrap confidence intervals. 5000 bootstrap replicates were taken.

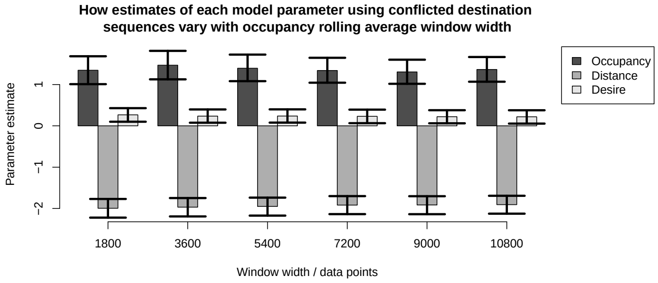

Figure 8 Model parameter estimates calibrated on conflicted destination sequences using different rolling average window widths for smoothing occupancies. Error bars represent the 95% confidence bootstrap confidence intervals. 5000 bootstrap replicates were taken.