



King, C. J., Koltsova, O., & Bode, N. W. F. (2022). Simulating the effect of measurement errors on pedestrian destination choice model calibration. *Transportmetrica A: Transport Science*.  
<https://doi.org/10.1080/23249935.2021.2017510>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1080/23249935.2021.2017510](https://doi.org/10.1080/23249935.2021.2017510)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Taylor and Francis at <https://doi.org/10.1080/23249935.2021.2017510>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



## Simulating the effect of measurement errors on pedestrian destination choice model calibration

Christopher King, Oksana Koltsova & N. W. F. Bode

To cite this article: Christopher King, Oksana Koltsova & N. W. F. Bode (2022): Simulating the effect of measurement errors on pedestrian destination choice model calibration, Transportmetrica A: Transport Science, DOI: [10.1080/23249935.2021.2017510](https://doi.org/10.1080/23249935.2021.2017510)

To link to this article: <https://doi.org/10.1080/23249935.2021.2017510>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 08 Mar 2022.



[Submit your article to this journal](#)



Article views: 254




[View related articles](#)



[View Crossmark data](#)

# Simulating the effect of measurement errors on pedestrian destination choice model calibration

Christopher King <sup>a</sup>, Oksana Koltsova<sup>b</sup> and N. W. F. Bode <sup>a</sup>

<sup>a</sup>Department of Engineering Mathematics, University of Bristol, Bristol, UK; <sup>b</sup>Bentley Systems UK Limited, London, UK

## ABSTRACT

Accurately calibrated pedestrian destination choice models help explain and predict foot traffic in public places by describing how individuals choose locations to visit. Model calibration relies on empirical data, which is subject to measurement errors that can obfuscate calibration. This contribution adds errors to simulated data in a controlled and realistic way which can be applied to many model specifications, demonstrated on a pedestrian destination choice model. Results show that errors can cause calibrated models to generate dynamics that differ substantially from the true dynamics, along with causing bias in parameters and decreased prediction accuracy. By quantifying the size of errors and the impacts on calibration, this work aims to guide researchers in pedestrian destination choice modelling on what level of error is acceptable given the scope of their research.

## ARTICLE HISTORY

Received 23 April 2021  
Accepted 4 December 2021

## KEYWORDS



Crowd simulation; pedestrian dynamics; destination choice modelling; statistical model calibration; measurement error


## ABBREVIATIONS

MLE: maximum likelihood estimation; MNL: multinomial logit; DC: destination coverage; OIQR: occupancy interquartile range; NLL: negative log-likelihood; GPS: global positioning system; IR: infra-red; PP: predictive power; KS: Kolmogorov–Smirnov; RMKS: relative median Kolmogorov–Smirnov

## 1. Introduction

As the global population grows, particularly in urban areas, it is increasingly important to understand and predict the flows and density of crowds in buildings and during large-scale events to ensure safety and improve people's experience of the environment (Murakami et al. 2020). People walk to places because of a desire to perform activities (Miller 2014), so understanding how pedestrians decide where to walk to and in what order they visit locations is crucial for understanding and predicting pedestrian traffic. A commonly-used framework for defining the behaviour of pedestrians is that proposed by Hoogendoorn and Bovy (2004) which splits pedestrian behaviour into three levels: strategic, tactical, and operational. The strategic level is involved with deciding which activities to perform, where, and

**CONTACT** Christopher King  chris.king@bristol.ac.uk; aa18187@bristol.ac.uk; N. W. F. Bode  nikolai.bode@bristol.ac.uk

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/23249935.2021.2017510>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in what order. Tactical level relates to routing and navigation behaviour. The operational level manages the step-by-step movement and interactions.

Theoretical models, often based on the broader framework of discrete choice modelling (Train 2009), have been developed to explain and predict the processes of pedestrian destination choice (Danalet 2015). Common to most of these models is the assumption that pedestrians trade off different factors against each other when making decisions (e.g. (Danalet 2015; Hoogendoorn and Bovy 2004; Beaulieu and Farooq 2019)). Many factors could be relevant, such as sociodemographic attributes of individuals (Wang and Li 2011), a person's habits and general mental state (Kielar and Borrmann 2016), and any spatial and temporal constraints of the situation (Ettema, Borgers, and Timmermans 1993). Therefore, the key challenges in pedestrian destination choice modelling are then to determine the mechanism by which information about different factors is combined and to establish the relative weighting or importance of these factors in determining the decisions of individuals. This contribution focuses on the latter problem and investigates challenges in learning the relative weighting of these factors from data.

A common approach to determine the relative weighting of factors in pedestrian destination choice is to calibrate theoretical models (Beaulieu and Farooq 2019; Dai 1998; Danalet 2015; Ettema et al. 2007; Ying et al. 2019; Yang, Fik, and Zhang 2013; Tinguely 2015). Models encapsulate the assumed integration of information and their parameters determine the relative importance of factors considered. Model calibration is the process of deciding on values for model parameters that best fit available data. However, almost all data collected from the real world contains errors. For example, measuring instruments can only measure to a certain precision and will contain sources of random and/or systematic errors (Viswanathan 2005). These errors have the potential to disrupt model calibration by obfuscating the true relationships between factors and observed dynamics. Therefore, it is important to assess the effects such errors have on the calibration process, including estimates of model parameters, the accuracy of predictions made by the model, and any alteration in dynamics predicted by models due to erroneous calibration. This contribution is concerned with the effects of errors in data on the calibration of a pedestrian destination choice model.

One of the earliest works suggesting measurement errors have an effect on discrete choice model calibration was Kao and Schnell (1987), who found that not accounting for errors introduces bias in the parameter estimates; a systematic deviation from the original estimate, and demonstrated a simple means of accounting for said bias. This result was corroborated by Hausman, Abrevaya, and Scott-Morton (1998), who, by performing a simulation study on a binary choice model using misclassification probabilities to add errors, found that neglecting measurement error causes bias in Maximum Likelihood Estimates (MLE) of hazard-duration models, illustrating their result on empirical data. Meyer and Mittag (2017) also perform a simulation study on a binary choice model and confirm the results by Hausman *et al.* and further show that the misclassification probabilities depend on the predictor being considered.

Bhatta and Larsen (2011) explore the impact of random errors in variables for a multinomial logit choice model in the context of transport mode choice for households at the urban level. These authors use empirical data and apply measurement errors to access/egress times and distances by drawing from normal, log-normal, and triangular distributions with

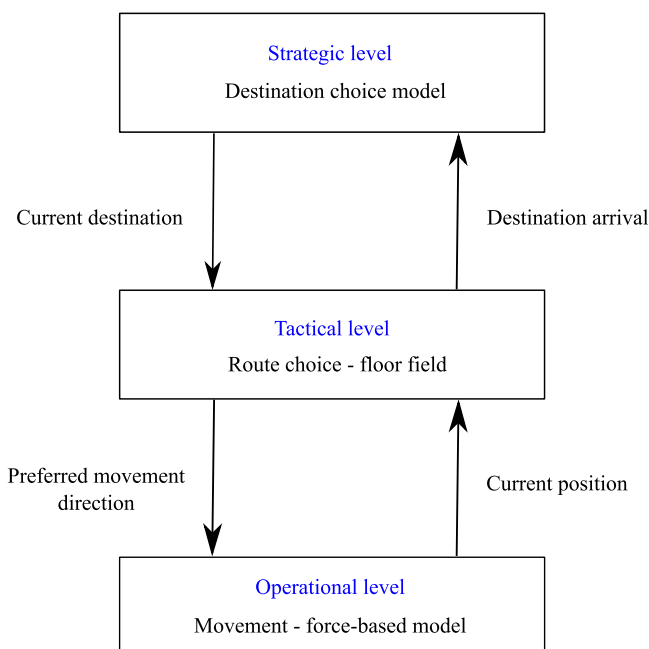
different variances. They find that estimates of the parameters associated with erroneous factors are downward biased (*i.e.* the parameter estimate is lower than expected from the data without errors added), while the bias in other parameters is less predictable.

Jang, Rasouli, and Timmermans (2017) extend the work of Bhatta and Larsen (2011), looking at the effect of measurement errors on both random utility and random regret choice model formulations. They also use empirical data but add additional noise to each variable as a proportion of the variable's intrinsic variance. Not only do they look at estimate bias, confirming the results seen by Bhatta and Larsen, they also look at the impact on prediction accuracy and find that it decreases with increasing measurement error. They find that the sensitivity of parameters to changes in their associated predictors also increases with increasing error.

Hausman, Abrevaya, and Scott-Morton (1998) and Meyer and Mittag (2017) conduct simulation studies for binary choice models, rather than models of choices with more than two alternatives, while Jang, Rasouli, and Timmermans (2017) and Bhatta and Larsen (2011) investigate multinomial choice models allowing more than two options, but use empirical data. Empirical data already contains inherent errors, and so studies using it measure the impact of additional error relative to this already erroneous data, rather than determining the absolute effect of errors. To the best of the authors' knowledge there is currently no simulation study on the effect of erroneous data on the calibration of a multinomial choice model. There is also currently no consistent, systematic, and realistic method for artificially adding measurement errors to data which does not depend on the specification of the model. Having this would be a useful model calibration tool to assess whether the amount of error present in candidate data will have a significant effect on model calibration before such a model is used to make predictions or derive explanations for phenomena. Additionally, while previous work shows that measurement error introduces bias in estimates of model parameters, it is unclear whether this bias is significant enough for the estimated model to produce significantly different dynamics from the unbiased model. For example, if a researcher is more interested in using their model to replicate the dynamics of a system, rather than obtaining accurate parameter estimates, then the presence of measurement errors may not be so important.

This paper attempts to address these gaps in the literature by assessing the effect of errors in data on model calibration using data from an agent-based simulator as the starting point. It is intended to be a guide for both empirical data collection and in choosing data to calibrate a specified model, establishing whether model calibration is feasible, either in replicating observed dynamics or in accurately explaining observations, for different amounts of measurement error in one or more variables.

The rest of this paper is organised as follows, the modelling framework is described first (Section 2.1), followed by details on characterising the dynamics of the model (Section 2.2). Before results are presented (Section 3), details of how errors are introduced to simulated data (Section 2.3.1), the metrics used for assessing the success of model calibration (Section 2.3.2), and the simulations performed (Section 2.3.3), are described. The interpretation and implication of these results as well as the limitations of this work (Section 4) are then discussed in detail before concluding remarks are made (Section 5).



**Figure 1.** The simulator according to Hoogendoorn's framework (2004).

## 2. Methodology

### 2.1. Simulation model

The computational model used in this study is written in Java by the authors and simulates the full microscopic movement dynamics of a fixed number of pedestrians in a confined space that contains a fixed number of destinations. Pedestrian behaviour described by the model can be viewed as occurring at three different levels, as shown in Figure 1, following Hoogendoorn and Bovy's (2004) framework. The main focus of this work is on the calibration of the model describing the strategic level – destination choice. The other behavioural levels are modelled explicitly, as they determine inputs for destination choice, such as the number of pedestrians at a destination which depends on how long it takes people to walk between destinations, for example.

Operational-level behaviour handles the step-by-step movement of agents in continuous space; describing the interactions between each other and the environment using a force-based model derived from (Helbing, Farkas, and Vicsek 2000). Further details on the implementation of this model can be found in (Bode and Codling 2013) and the same values of all parameters as in this previous work are used throughout. In general, any pedestrian movement model can be used, as the aim of this simulator is to look at destination choice behaviour of pedestrians only.

Tactical-level behaviour describes how agents choose a route from their current position to a chosen destination. Agents navigate using the shortest-distance route that is implemented via discrete floor fields, as described in (Bode and Codling 2013). To conveniently access the shortest routes towards all possible destinations, separate floor fields are implemented for each destination. The preferred movement direction used in the force-based

movement model of an agent is based on the local gradient in the floor field at its current position.

A probabilistic discrete choice model grounded in Random Utility Theory (McFadden 1974) is used to describe agent destination choice in the simulator. There are many discrete model specifications, each with their own benefits and limitations (Train 2009). For the purpose of this contribution, the multinomial logit (MNL) model is capable of describing a sufficiently broad range of behaviours:

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

where  $P_i$  is the probability of choosing alternative  $i$  from the set of all possible alternatives,  $C$ . This model makes use of a utility function for each alternative,  $U_i$ , which quantifies the value of each alternative to a given decision-maker. From this specification, the alternative with the highest utility is most likely to be chosen by a decision-maker.

### 2.1.1. Destination choice model specification

The destination choice model used here only considers three factors which are now briefly discussed in turn: the occupancy of destinations, the distance that needs to be covered to reach them, and how desirable they are according to the intrinsic preferences of each individual.

How busy a destination is, or its occupancy, often impacts the likelihood of people visiting it (Beaulieu and Farooq 2019; Kielar and Borrmann 2016; Saarloos, Fujiwara, and Zhang 2007; Hui, Bradlow, and Fader 2009). Sometimes occupancy can be attractive when it indicates something worth visiting, such as acts at festivals, tourist attractions, and promotional events (Kwak et al. 2014). At other times, pedestrians seek to avoid high occupancy destinations, for example when shopping, or buying tickets for public transport. This factor is often embedded within other factors, such as time spent at a location (Ton 2014), seating capacity (Danalet 2015), or floor space (Zhu, Timmermans, and De 2006; Borgers and Timmermans 1986).

For most people in the majority of contexts, destinations further away are less likely to be chosen than those nearby (e.g. Danalet 2015; Arentze, Ettema, and Timmermans 2013; Ettema, Borgers, and Timmermans 1993; Ettema et al. 2007; Fesenmaier 1988; Kurose, Borgers, and Timmermans 2001; Li et al. 2019; Ton 2014; van der Hagen, Borgers, and Timmermans 1991; Zhu, Timmermans, and De 2006). The investment of time and effort involved in walking further could only be an attractive feature of a destination if the decision-maker wishes to exercise, has no time constraints, or just enjoys the journey as much as reaching a destination.

The intrinsic motivation of pedestrians to visit a destination is harder to quantify than effects of occupancy or distance. Nevertheless, it is ubiquitous in pedestrian destination choice modelling. Many people visit a place with a certain set of activities to perform, often in a certain order of 'priority' (Li and Allbeck 2011). This itinerary or schedule (related to the Strategic Level in Hoogendoorn and Bovy's (Hoogendoorn and Bovy 2004) framework) of activities influences the destination choice behaviour of people in terms of the order of destinations to visit. There have been many ways of quantifying the desirability of a destination in the literature, such as from market research (Fahmy, Alablani, and Abdelmaguid 2014; Yang, Fik, and Zhang 2013), using decay functions based on the supposed opinions

of individuals towards destinations (Dijkstra, Timmermans, and de Vries 2013; Kielar and Borrmann 2016), the number of transitions between destinations (Greenwood, Sharma, and Johansson 2015), or using features of the location/building such as seating capacity (Danalet 2015), and floor space (Ettema, Borgers, and Timmermans 1993).

This discussion suggests that depending on the context, factors can make destinations appealing or undesirable for pedestrians. This can result in fundamentally different dynamics depending on the context. Consider, for example, a situation where high occupancies are appealing. This would lead to a positive feedback loop causing popular destinations to become even more popular. In contrast, if pedestrians avoid high occupancies, they spread more evenly across destinations. This also highlights the importance of accurate calibration. An extreme example would be a wrong sign for the estimated parameter capturing the response to destination occupancies, representing a failure to accurately distinguish between the two scenarios outlined above.

The three common predictors for destination choice identified above are considered in the utility,  $U_i$  for each alternative  $i$ :

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

where  $\beta_{occ}$ ,  $\beta_{dist}$ , and  $\beta_{des}$  are the parameters corresponding to the predictors,  $\hat{n}_i$  is the normalised occupancy of destination  $i$  (number of people at destination  $i$ ),  $\hat{d}_i$  is the normalised distance to destination  $i$  from the agent, and  $\hat{q}_i$  is a normalised measure of desire to visit a destination  $i$  by the decision-maker (defined in more detail below). The predictors are normalised for each decision to between zero and one by dividing each value by the maximum observed value of that predictor in the decision time-step. There are many reasons for this normalisation, firstly it avoids the numerical computation problem of computing very large or very small numbers. Second, it ensures that all predictors have the same range such that the relative effects of each predictor on the choice probability depend only on the relative values of the parameters. It also has the effect of setting a limit for the maximal contribution of each predictor and means that parameters capture the relative strength of the effects predictors have rather than effects per unit increase in predictors. The values of these predictors are captured at the time the decision was made.

$\beta_{occ}$  and  $\beta_{dist}$  can take positive and negative values, representing the potential for occupancy and distance to have an attractive or repulsive effect, respectively. The parameter  $\beta_{des}$  can only be positive, because it represents the effect of a person's activity schedule. It is assumed that parameters have the same values for all agents.

There have been many ways of quantifying the desirability of a destination in the literature (see above). In this work, the attractiveness of a destination for an agent is based on its destination schedule.

$$q_{i,k}(A_k, \Lambda) = e^{-\Lambda s_i} \quad (3)$$

where  $q_{i,k}$  is the desirability of destination  $i$  for agent  $k$ ,  $A_k$  is the schedule of destinations for  $k$ ,  $s_i$  denotes the position of  $i$  in the schedule of  $k$ , and  $\Lambda$  is a parameter which controls the strength of adherence of agents to their schedule. High values of  $\Lambda$  indicate a strong tendency of agents to rigidly follow their schedule. Destinations further along the agent's schedule are of lower priority than those near the start. If a destination is not in an agent's activity schedule, then it is given a desirability of  $q = 1 \times 10^{-121}$ . This value was chosen as even if the destination is not in an agent's schedule, it could still have some small



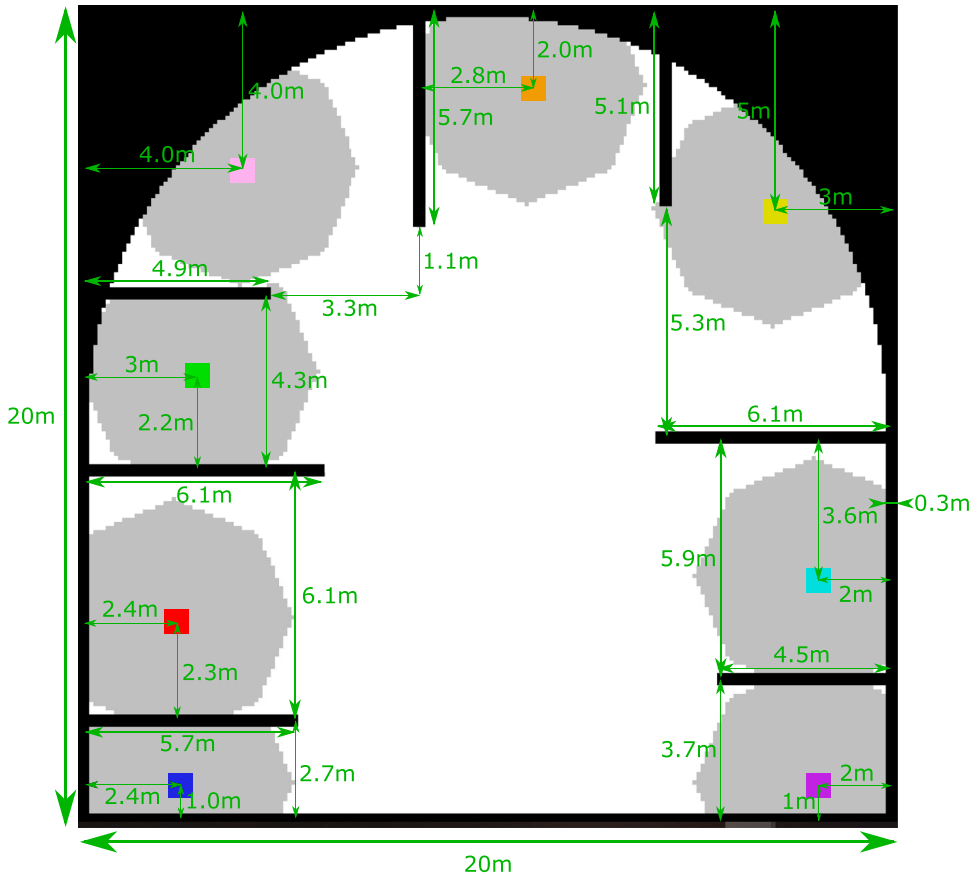
default desirability. To illustrate this, consider the situation where an agent has destination schedule  $A = (1, 2, 3, 4)$  and can choose between these four destinations. So  $\mathbf{q} = (q_1, q_2, q_3, q_4) = (e^{-\Lambda}, e^{-2\Lambda}, e^{-3\Lambda}, e^{-4\Lambda})$  are the desirabilities for each destination. If the agent in the example visits destination one, its schedule after the visit is now  $A = (2, 3, 4)$  and  $\mathbf{q} = (1 \times 10^{-200}, e^{-\Lambda}, e^{-2\Lambda}, e^{-3\Lambda})$ . Due to the requirement that agents do not choose to visit the same destination that they are currently located at, in practice destination 1 and therefore  $q_1$  would not be considered when choosing the next destination to visit.

### 2.1.2. Simulations

Though the processes described in this section are general to any model specification and simulated environment, the results of this paper are applied to one particular environment shown in Figure 2. The simulated environment is primarily characterised by the locations of walls and destinations inside a square of 20 m by 20 m, though other environment shapes and sizes are possible in general. Its shape resembles that of a horseshoe, and it could be thought of as representing a shopping centre with an open space in the middle and shops (destinations) lining the outer edge. This placement of walls in this environment is deliberately designed to reduce symmetry ensuring a spread of distances from each destination to every other. Highly symmetric environments can result in special dynamics that are illustrated in the appendix (see Appendix A). The grey areas represent the size of each destination. Any agent within these areas is counted as being present at the destination for calculating occupancy and for determining whether an agent has arrived at its chosen destination. These areas are circular with a radius of 3 m, any free space that lies within this distance from the destination coordinate (coloured squares) is included in this area. All simulations for the parameter scan and for the error study below are performed for 80 agents over 10,000 time-steps (= 500 s) in the horseshoe environment shown in Figure 2.

At the start of simulations, each agent is given a random initial position, along with a schedule of destinations, which describes the order in which an agent wishes to visit destinations. The length of this schedule is fixed, but each agent has a random arrangement of all possible destinations, where each destination only appears once in the schedule. The schedule length is fixed such that every destination is present only once in an agent's schedule. This was inspired by reality, as in many situations, people do not perform the same activity at the same place more than once during a given trip. Different destination schedules are possible and may even be likely in reality. Random schedules were chosen as they reduce correlations between desirability and the other predictors, as well as correlations between agents. An example in the appendix illustrates effects of destination schedules fixed over agents (see Appendix A, Figure A4). Schedules are updated every time an agent chooses a new destination by removing the first instance of a destination the agent has just visited if it appears in the schedule. Otherwise, the schedule remains unchanged. The initial destination agents visit is chosen at random from all destinations at the start of simulation. The quantities  $\hat{d}_i$ ,  $\hat{n}_i$ , and  $\hat{q}_i$  in Equation (2) are determined from the initial positions of the agent, of all other agents, and the schedule of the agent, respectively.

For each time-step in simulations (0.05 s), the position, speed, and direction of movement for each agent are updated according to their current destination and its associated floor field using the operational level model (for details, see (Bode and Codling 2013)). The occupancy of destinations is recorded at each time-step and defined as the number of agents inside each destination area, an unobstructed region of radius 2 m around the



**Figure 2.** The environment used in simulations. White space indicates free space outside of destinations areas, grey is free space within destination areas that are defined as unobstructed areas within a fixed radius of destination centres (shown as coloured squares). Walls are shown in black. The discrete nature of the floor field is visible.

centre point of the destination (see Figure 2). If an agent enters the area of its chosen destination, it is assigned a waiting time during which its preferred movement direction is set to the centre of the destination area. Waiting times for all agents and destinations in simulations are sampled from an exponential distribution with parameter  $\lambda = 0.003$  and with a constant  $t_0 = 20$  s added to reflect a non-zero minimal time agents spend at a destination. Once their waiting time has passed, agents choose a new destination that must be distinct from their current destination.

## 2.2. Destination choice model characterisation

To demonstrate the range of dynamics the destination choice model can produce, a parameter scan of the model is performed. The following set of integer combinations of parameters are considered:  $\beta_{occ}, \beta_{dist} \in \pm 5, \pm 4, \pm 3, \pm 2, \pm 1, 0$  and  $\beta_{des} \in 0, 1, 2, 3, 4, 5$ . These values were chosen to ensure a representative spread of observable dynamics from the destination choice model that gave accurate and robust calibration from numerical estimation (see Section 2.3). This means that  $11 \times 11 \times 6 = 726$  different parameter combinations

are considered in the three-dimensional parameter space of the model. The parameters are integers out of convenience, as long as the combinations considered represent the different dynamical regimes of the model, the actual values of the parameters do not matter.

To quantify the dynamics observed, two summary statistics are defined. Destination Coverage (DC) is defined as the proportion of unique destinations visited by each agent, averaged over all agents in the simulation, where a high DC indicates that most agents visit most destinations over the course of the simulation. Occupancy Interquartile Range (OIQR) is defined as the average interquartile range of occupancies across destinations over all timesteps. A high OIQR indicates that most destinations show large variation in occupancy over time.

The average values for these summary statistics is computed over 100 simulation replicates. Results from the parameter scan are used to identify a set of parameter combinations for the error study that capture a representative range of model behaviours. Combinations are denoted as vectors  $(\beta_{occ}, \beta_{dist}, \beta_{des})$ .  $(0, 0, 0)$ , corresponding to random destination choice, is included to act as a reference case.

### 2.3. Error study

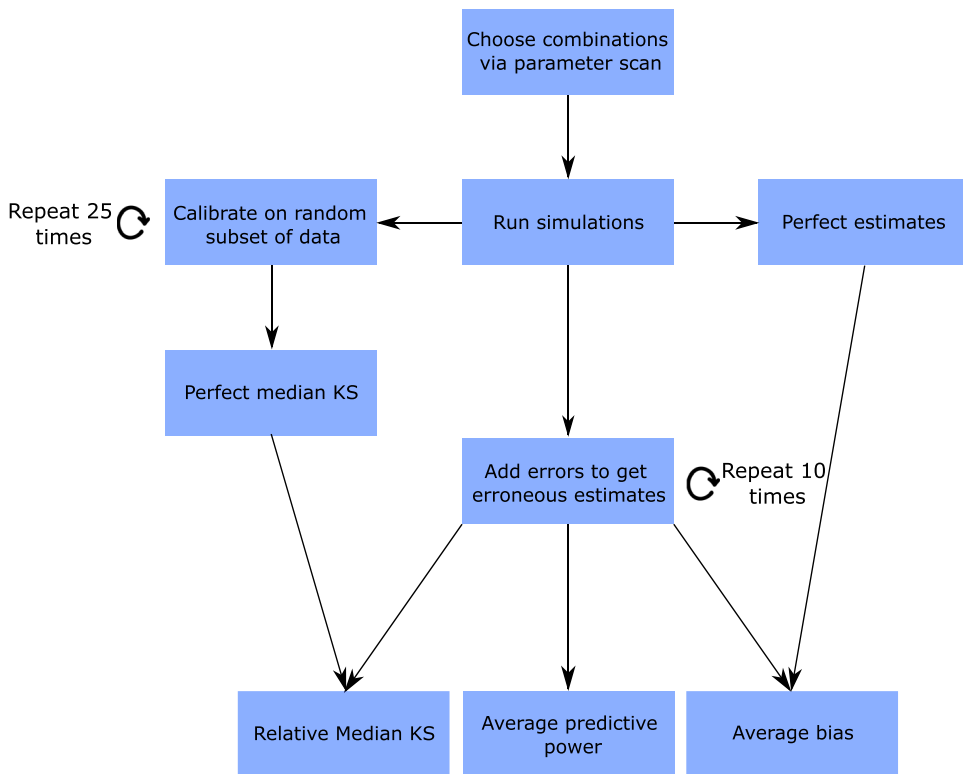
Figure 3 provides an overview of the simulation study performed to investigate the impact of measurement errors on model calibration. This section will explain the protocol of adding errors (Section 2.3.1), describe the three metrics used to assess calibration success (Section 2.3.2), and provide details for the simulations undertaken (Section 2.3.3).

#### 2.3.1. Adding errors

The approaches for adding errors to observed destination sequences, destination occupancies, and distances between destinations will now be discussed in turn.

Previous works suggest the sequence of destinations can be inferred in two main ways: first, using spatial positioning instruments, like Global Positioning System (GPS), Wi-Fi, or video cameras e.g. (Danalet 2015; Ton 2014; Ying et al. 2019; Yoshimura et al. 2014; Greenwood, Sharma, and Johansson 2015). These can be used to create trajectories for individuals. Combining this with information about the locations of places of interest, the destinations visited by people can be inferred. Second, surveys or questionnaires can be used as a way for individuals to self-report the sequence of destinations they visited (e.g. Ettema, Borgers, and Timmermans 1993; Zhu, Timmermans, and De 2006; Arentze and Timmermans 2004; Ruiz, Chebat, and Hansen 2004; Wang and Li 2011). Either of these approaches to collecting data on the sequence of destinations visited by an individual can result in errors when destinations are misidentified, missed or erroneously detected. Only misidentifications are considered here, as adding or deleting sequence elements could artificially alter the overall amount of data available for calibration which would hinder a principled comparison of calibration success (e.g. see predictive power in Section 2.3.2).

Destinations can be misidentified resulting in the substitution of one destination for another. This could occur due to recall error of individuals in surveys, or measurement errors in location data. For example, consider the measurement error in GPS signals which can be on the order of several metres. In data collection this may make it difficult to distinguish locations that are in close proximity.



**Figure 3.** Overview of the error study procedure.

Destination substitutions are implemented on the chosen destination sequences of agents. For each element in the sequence of chosen destinations of an agent, a substitution is performed with a fixed probability. This process is performed for every agent in the simulation. The size of the probability adjusts the sequence error, and represents the expected proportion of altered destinations across all agents in a simulation. As this procedure is run over many destination sequences, on average, the altered proportion of destinations will be very close to the sequence error value.

For any substitution, the set of available destinations are those which are different to both the original destination and the destinations immediately before and after it in the sequence. This is to ensure that destination sequences contain no identical consecutive elements, as these could be detected and filtered out in data cleaning (although relaxing this assumption is possible). If the destination is at the beginning or end of a sequence, then only the destination immediately after or before the chosen destination is considered, respectively. These constraints on substitutions can have the effect of introducing artificial structure to substituted sequences, particularly if there are few destinations to choose from (see Appendix A).

For illustration, consider a simulation in an environment with four destinations. In this example, each destination has a unique integer identifier from 0 to 3. Consider a sequence of chosen destinations: (0, 2, 3, 1, ...). Suppose a sequence error of 0.25 (or an average of 25% of all chosen destinations being altered) is applied to this data and that the second

sequence element in the sequence above is selected to be substituted. The only destination available for substitution for this sequence element is 1, as the preceding, current and following destinations form the set 0, 2, 3. Thus, the resulting sequence after the sequence error is applied is: (0, 1, 3, 1, ...). As stated above, this procedure is performed for all sequence elements and agents in a simulated data set.

Errors in occupancy observations can arise from measurement errors. The occupancy of a place/room is often measured by using some kind of counting instrument, such as infra-red (IR) scanners (Ton 2014), manual counting (Benezeth et al. 2011), or vision-based sensors (Scovanner and Tappen 2009). These methods can miscount the number of people, especially when occupancy is high. Occupancies must be non-negative integer values, as there cannot be a negative number of people in a space, nor can there be only part of a person in a space. With this in mind, a discrete distribution bounded at 0 is appropriate. Therefore, occupancy errors are implemented by drawing a new occupancy from a binomial distribution  $B(m, p)$ , where  $m$  and  $p$  are parameters. These are calculated by requiring that its mean,  $\mu = mp$ , is the true occupancy and by letting the size of the occupancy error be  $q = 1 - p$ . The variance of the binomial distribution is given by  $\sigma^2 = mp(1 - p)$ , so the error value controls the spread of values around the true occupancy. The upper bound for the occupancy error is given by the case when the binomial distribution has maximal variance,  $\sigma_{\max}^2$ , which occurs when  $p = q = 0.5$ . The remaining, smaller, error values are obtained by setting different values for this probability to achieve different proportions of  $\sigma_{\max}^2$ . The relation between variance and mean of the binomial distribution implies that the spread of possible erroneous occupancies increases as the true occupancy increases, as seen in reality.

Errors in estimating the distances between destinations can arise from measurement errors or difficulties in defining this distance in cases when the spatial extent of a destination is not clear. Distances are non-negative continuous values, so a continuous distribution bounded at zero is appropriate. A log-normal distribution,  $\text{Lognormal}(\nu, \gamma^2)$  satisfies these conditions, where  $\nu$  and  $\gamma$  are parameters. The true distance is set to be the mode of the distribution  $\theta = \exp(\nu - \gamma^2)$ .  $\gamma$  acts as a proxy for the error size, so the value of  $\nu$  can be determined. However, being bounded at 0 makes the distribution asymmetric about the mode, with a larger range of possible distances above the mode than below it, this is unavoidable if the distribution is non-negative. But this asymmetry decreases as the mode increases.

These error specifications are designed to replicate real sources of error in these measured quantities. Errors in the schedules of agents were not considered explicitly. In real data, it is often impossible to know a person's true desires and priorities during data collection. Uncertainties around this aspect of behaviour are only considered implicitly via sequence errors. The implications of this and the possibility for extending this contribution are discussed in Section 4.

### 2.3.2. Metrics

Several metrics are used to assess the effect of errors on model calibration. First, the impact of errors on the ability of the model to explain the data, and hence make accurate predictions is given by the predictive power (PP), which is the final optimised negative log Likelihood (NLL) (Bhatta and Larsen 2011). A large NLL indicates that the model fits the data poorly and any subsequent predictions made by the model will be less accurate, *i.e.* a small

predictive power. The model was calibrated on data with and without added errors from the same 40 replicate simulations for each parameter combination to prevent any effect on PP due to using different amounts of data in calibration. To facilitate comparison of the effect errors have on PP across parameter combinations, the following normalised quantity is reported:

$$\widehat{PP}_g = \frac{\max(\mathbf{PP}) - PP_g}{\max(\mathbf{PP}) - \min(\mathbf{PP})} \quad (4)$$

where  $\mathbf{PP}$  is a vector of PP values obtained for each value of an error type for a given parameter combination,  $PP_g$  is the PP for the  $g^{\text{th}}$  error value, and  $\max$  and  $\min$  are functions which find the maximum and minimum element in a vector, respectively. It is also assumed that there are at least two unique elements of  $\mathbf{PP}$ , so that  $\max(\mathbf{PP}) \neq \min(\mathbf{PP})$ .

To investigate the effect of errors on the values of the parameter estimates, the estimate bias is defined as in Jang, Rasouli, and Timmermans (2017):

$$b = \frac{\beta_e^*}{\beta^*} \quad (5)$$

where  $\beta_e^*$  is the estimate of a parameter using data with added errors, and  $\beta^*$  is the corresponding estimate from perfect data.  $b > 1$  indicates that errors cause positive bias in the estimate and vice-versa for  $b < 1$ . The bias is not calculated for the (0,0,0) reference case, as the parameter estimates are small and can readily change sign, which would cause large biases to be observed. It should be noted that this definition of bias allows the effect of errors on each unique parameter value to be quantified. Though comparison of bias magnitudes between different parameter values is impossible, due to different parameters having different  $\beta^*$ .

To assess quantitatively whether the model dynamics produced by two sets of parameter values are similar, the Kolmogorov–Smirnov (KS) statistic is used to compare the distributions of the destination coverage (DC) and occupancy interquartile range (OIQR) produced via simulations using the different sets of parameter values. It is possible for different parameter combinations to produce similar dynamics, even if the magnitude of  $b$  described above is large.

To facilitate comparison of the effect errors have across parameter combinations, a relative KS value is used. The model is calibrated 25 times on the data without added errors and each calibration uses a different subset of 40 replicates from the original 100 performed as part of the parameter scan. Following calibration, a further 100 simulations were performed for each of these 25 parameter estimates to obtain estimated summary statistic distributions. These were compared to the original distributions obtained from the parameter scan. This resulted in 25 KS values for each chosen parameter combination, the median of which was used as the reference KS value for that combination (see Figure 3).

To obtain relative KS values indicating the effect of errors, the KS value between the original summary statistic distributions from the parameter scan and those from simulations of models calibrated on data with added errors are generated for each parameter combination and error value. Repeating this over multiple replicates of adding errors to data (see Section 2.3.1) produces a set of KS values for each chosen combination and error value. The reference median KS for each combination is then subtracted from the median of this set to obtain the Relative Median KS (RMKS). The RMKS thus quantifies the change in dynamics

in model simulations when the model is calibrated on data with added errors compared to a calibration on data without added errors.

Together, these metrics aim to give a comprehensive quantitative assessment of the impact of errors on model calibration.

### 2.3.3. Error study simulations

The following error values were applied to the perfect data: occupancy error ( $1 - p$  in the binomial distribution) (0.0675, 0.125, 0.25, 0.375, 0.5), sequence error (average percentage of chosen destinations substituted over all agents) (1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100), and distance error ( $\gamma$  in the log-normal distribution) (0.05, 0.1, 0.15, 0.2, 0.25). The process of adding each of these errors individually to the data was repeated 10 times for all types and levels of error, as the process of adding errors is random, and so could have different impacts on estimates and metrics. The average effect of each error on each metric alongside two standard deviations is reported.

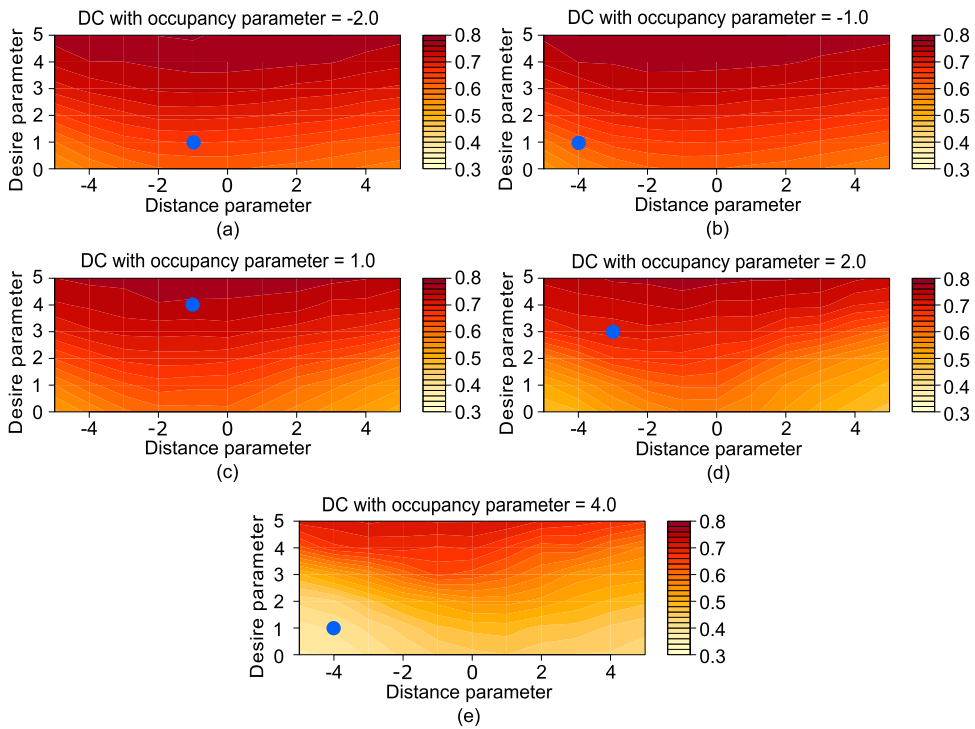
The choice model is calibrated on simulated data (or simulated data with added errors, as described above) using Maximum Likelihood Estimation (MLE), implemented with the optim function in the R programming environment (R Core Team 2020). By using the destination sequences as the observed data, with observed destination occupancies, destination distances, and destination desirabilities being substituted into the utility function in Equation (2), which is thence substituted into Equation (1), estimates of model parameters,  $\beta_{occ}^*$ ,  $\beta_{dist}^*$ ,  $\beta_{des}^*$  are obtained. For given input data, the calibration process is repeated 10 times with random initial parameter values in the optimisation routine to avoid only detecting local maxima in the Likelihood. Estimates from the calibration with the lowest final Negative Log Likelihood (NLL; corresponding to the maximal Likelihood) are retained. Importantly, if no errors are added to the simulated data this implies perfect knowledge of destination occupancy, the distances between destinations, and the schedule of agents, which is unlikely to be available in all real-world applications (further discussed in Section 4).

The error study is performed on a selection of parameter combinations, as explained in Section 2.2, and uses simulated data without errors as its starting point. Parameters are calibrated on the combined data of a random sample of 40 out of the 100 replicates for each parameter combination selected from the parameter scan, as this resulted in sufficiently stable estimates. Subsequently, errors are introduced to this sample of simulated data and the calibration success is measured by comparison to the calibration on the data without added errors.

Model calibration on the various kinds of data with added errors was conducted in the same manner as for the data without added errors. This generates parameter estimates and PP values for each combination for each value of each type of error. The estimates generated are used to calculate the estimate bias,  $b$ , and as input for the simulations to compute the Relative Median KS (RMKS).

## 3. Results

This section first provides the results of the parameter scan; a characterisation of the range of dynamical states that can occur in simulations of the model and the parameter combinations chosen. It then demonstrates how the calibration of the model is affected when different levels of errors are introduced in the data.



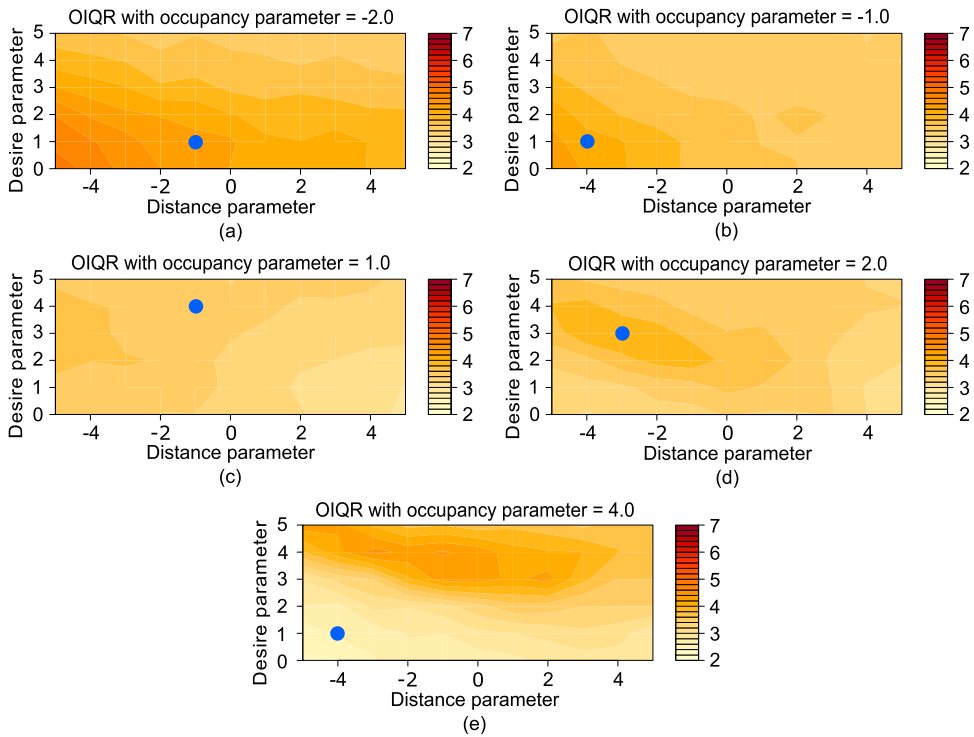
**Figure 4.** Destination coverage for different distance, desire, and occupancy parameter values. The combinations used in the error study are marked by blue dots. (DC = Destination Coverage).

Figures 4 and 5 show how changing the parameters of the destination choice model affects the dynamics in simulations, as measured via the summary statistics Destination Coverage (DC) and Occupancy Interquartile Range (OIQR).

Since the occupancy, distance, and desirability of destinations are normalised, the relative size of parameter values is crucial in determining the behaviour of agents. Figure D2 shows that agents visit more unique destinations when the distance parameter is small and the desire parameter is large. This can be explained by the fact that the desire parameter controls how influential the agent's schedule is on their choice of destination and since the schedule of agents is a random arrangement of every destination, agents are likely to visit many different destinations. Conversely, a high distance parameter, either positive or negative, encourages agents to visit only destinations furthest from or closest to their current destination, respectively, resulting in only a selection of destinations being visited frequently. A large positive occupancy parameter with a large negative distance parameter shows the lowest DC, because agents are compelled to visit the busiest destination, which is likely to be close to them due to influence of the distance parameter, resulting in pairs of close destinations being visited often, with random fluctuations allowing these pairs to change (see video 1 in supplementary material).

For negative occupancy parameters, the OIQR only varies within a narrow range as the other two parameters are varied and shows a maximum for large negative distance and

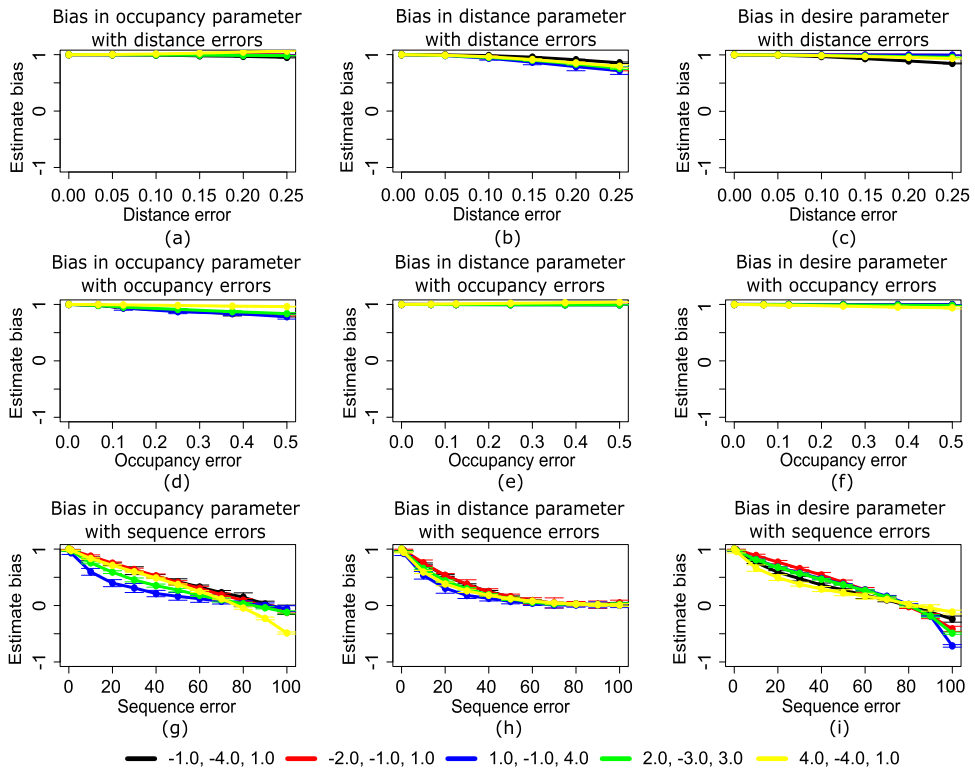




**Figure 5.** Occupancy Interquartile range for different distance, desire, and occupancy parameter values. The combinations used in the error study are marked by blue dots. (OIQR = Occupancy Interquartile Range).

small desire parameter values. Here the influence of agent's diverse schedules is negligible, and agents are compelled to visit the closest destination while avoiding busier destinations. The latter effect means that destinations will oscillate between busy and empty, leading to a wide range of occupancies across destinations at any one time (see video 2 in supplementary material). The behaviour for positive occupancy parameters instead shows a region of high OIQR within the parameter range considered that increases in size as the occupancy parameter increases. This reflects agent's increasing desire to visit the busier destinations. Combining this effect with the desire to visit closer destinations, this results in a pair of close destinations becoming consistently busy, leading to a small OIQR across destinations on average over time.

These results show that the model can describe different dynamics ranging from agents concentrating at a few destinations to them spreading out more evenly across destinations. Based on these results, the following set of parameter combinations was chosen for the error study to capture a representative range of the DC and/or OIQR values that are attained in model simulations:  $(-2, -1, 1)$ ,  $(-1, -4, 1)$ ,  $(1, -1, 4)$ ,  $(2, -3, 3)$ ,  $(4, -4, 1)$ . For example, combination  $(\beta_{occ}, \beta_{dist}, \beta_{des}) = (4, -4, 1)$  has both a low DC and OIQR, and  $(2, -3, 3)$  sits in an 'island' of high OIQR. The parameter values for the chosen combinations are represented by blue dots in Figures 4 and 5.

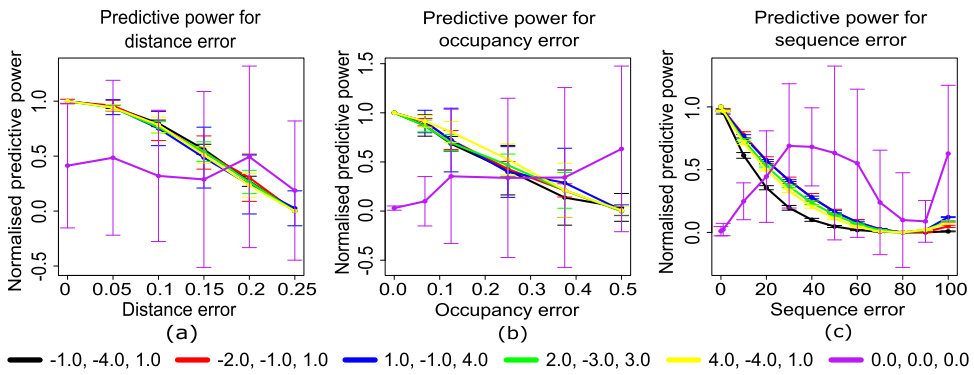


**Figure 6.** Average bias in each model parameter due to each type of error. Error bars show two times the standard deviation of bias over 10 repetitions of adding errors to perfect data. Each line shows a different combination of original parameter values. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ .

The remainder of this section describes the results of the error study. Each metric described in Section 2.3.2 is considered in turn for all chosen parameter combinations, error types and values.

Figure 6 shows the bias in each parameter with each type of error. Firstly, there is no obvious trend in bias with occupancy error on the distance parameter estimate, and vice-versa across parameter combinations (Figure 6a,e). However, distance and occupancy errors cause downward bias in their respective parameter estimates (Figure 6b,d). This makes sense considering that the distance from a destination is independent of its occupancy, so any error applied to one of these should not impact the estimate of the other's parameter. The attenuation indicates that the presence of errors reduces the effect of their accompanying predictors, such that with more errors introducing randomness to their respective predictors causes the calibrated model to become more like the random choice model.

Sequence errors appear to have the greatest effect on the bias of all three parameters, with each showing strong attenuation as errors increase (Figure 6g, h, i). This significant effect could arise from the fact that these sequences are at the heart of MLE for the model, and so any significant changes to these sequences can have drastic effects on any trends in destination choice due to any of the three predictors.

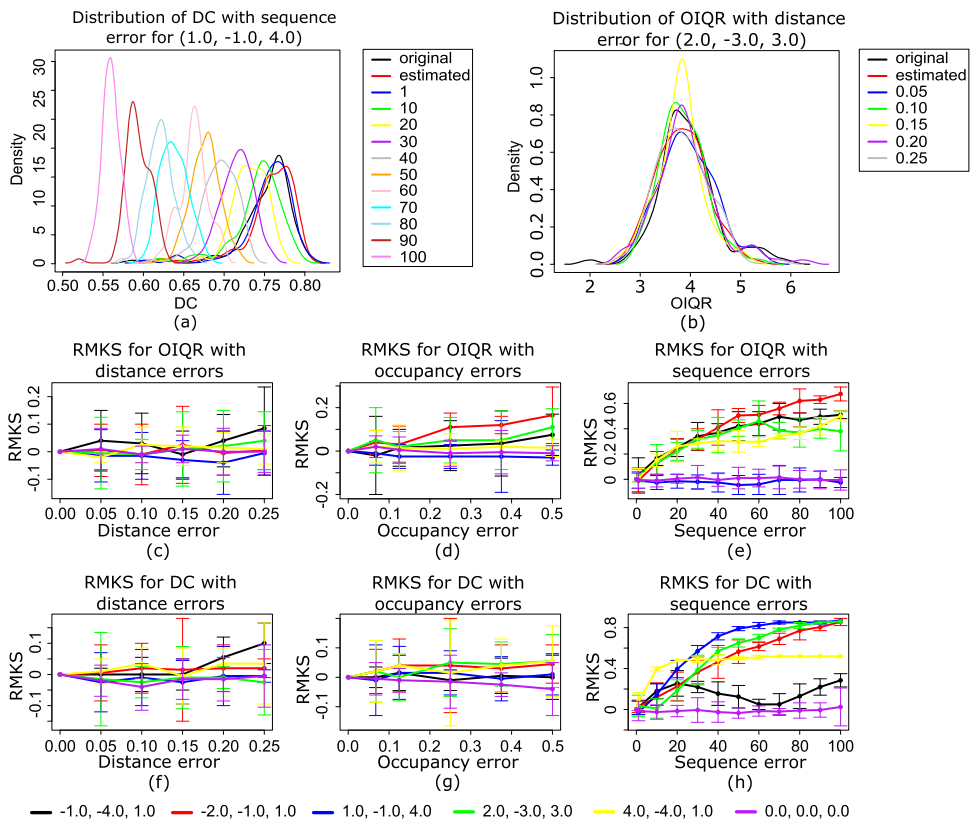


**Figure 7.** Average normalised predictive power of the choice model for all combinations with each type error. Error bars show two times the standard deviation of predictive power over 10 repetitions of adding errors to perfect data. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ .

There also appears to be downward bias in the desire parameter with distance errors (Figure 6c), albeit the strength of this effect seems to depend on the parameter combination. One possible source of correlation between these parameters could come from the way in which destination schedules of agents relate to the configuration of destinations in the environment even for randomised schedules (for a clear link in this regard see Appendix A, Figure A4).

The results for the normalised predictive power are shown in Figure 7 (see Equation (4) in Section 2.3.2). This confirms what might be expected intuitively: the predictive power generally decreases with the introduction of any error for all parameter combinations except  $(\beta_{occ}, \beta_{dist}, \beta_{des}) = (0, 0, 0)$ . For sequence errors, predictive power reaches a minimum when 80% of all chosen destinations are altered and subsequently increases again for larger sequence errors (Figure 7c). This increase could result from artificial structure being introduced into the data due to the constraint that agents have to visit new destinations when destinations are substituted. The model could then be fitted to this artificial structure with a different set of parameters. This notion is supported by the estimate bias in Figure 6h, i, where the signs of the occupancy and desire parameters begin to become negative above this sequence error. The strength of this effect depends on the environment, specifically how many destinations there are and how they are arranged, as shown in the appendix at the example of an empty room with four destinations arranged symmetrically in a square (see Figure A5). The parameter combination  $(0, 0, 0)$  shows completely different behaviour to the other combinations, with much larger variation over error study replicates. This makes sense, as this is the random destination choice model. Adding a random source of error to a set of random data does not make it any less random. It could be argued that for sequence errors, the predictive power of  $(0, 0, 0)$  peaks at 50% error, but the large overlap of error bars suggests a high degree of variability that makes any interpretations of such effects tentative.

The Relative Median KS-statistic (RMKS) values shown in Figure 8 indicate that the distance errors have no effect on the distributions of either summary statistic – for all parameter combinations there is no significant increase above zero. This suggests that distance



**Figure 8.** Average difference in median erroneous KS values and median perfect KS values for destination coverage and occupancy interquartile range for every parameter combination plotted against each error type. An example of when the summary statistic distributions change significantly (a) and minimally (b) are shown to illustrate the kinds of differences seen between distributions. Error bars show two times the standard deviation of RMKS over 10 repetitions of adding errors to perfect data. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ . (DC = Destination Coverage, OIQR = Occupancy Interquartile Range).

errors do not affect the calibration of the model substantially enough to result in different dynamics, as measured by DC and OIQR. This is supported by Figures B1 and B4 in the appendix, which show that both summary statistic distributions for the estimated model with distance errors are not significantly different from the original estimated distribution. Similarly, occupancy errors have no significant effect on the RMKS for DC, which is corroborated by Figure B5 in the appendix. Occupancy errors cause little significant change in RMKS for OIQR, except for parameter combinations  $(-2, -1, 1)$  and  $(2, -3, 3)$ , which show a small increase, but the overlapping error bars suggest that this effect is subject to substantial variability. Figure B2 in Appendix B shows that for  $(-2, -1, 1)$ , the original estimated distribution is strongly peaked at around 4.5, but this peak decreases and shifts down slightly in occupancy as errors increase. The opposite seems to occur with  $(2, -3, 3)$ , where the original estimated distribution has two smaller peaks, while the distributions for data with added errors show one larger peak. Again, this suggests that occupancy errors do not affect the calibration of the model substantially enough to result in different dynamics, as measured

by DC and OIQR. Figure 8b is a representative example of how similar the summary statistic distributions are in this case.

The RMKS with sequence errors shows slightly different behaviour for the two summary statistics. The parameter combination (0, 0, 0) shows little change in either summary statistic. This makes sense as adding additional random noise to a random model will not cause any significant shifts in dynamics. Even though (1, -1, 4) shows little change in the OIQR distributions, it shows some of the highest changes in DC distributions, as evidenced by Figure 8a, where DC decreases significantly with increasing sequence error. (-2, -1, 1) shows a continuous increase in RMKS for both summary statistics, indicating that the sequence errors disrupt the calibration enough to fundamentally change the dynamics. The distributions of this combination with sequence errors (Figures B3 and B6) show a gradual downwards shift in the distributions as sequence errors increase. For parameter combination (-1, -4, 1) the RMKS for OIQR increases continuously, but the RMKS for DC seemingly oscillates at low values, indicating that sequence errors between 20% and 50% actually make the DC distributions more similar. From Figure B6 in Appendix B, it can be seen that initially, the distributions shift upwards slightly for small sequence errors, before shifting back downwards for subsequent errors. Parameter combination (2, -3, 3) has a continually increasing RMKS for DC like (-2, -1, 1), but its RMKS for the OIQR appears to plateau after 30% sequence errors. Figure B3 in the appendix reflects this trend, showing that the original distribution is quite broad, and sequence errors narrow the distribution and shift the peak downwards, but the distributions are similar for larger sequence errors. Parameter combination (4, -4, 1) shows a sharp increase in the RMKS for DC before plateauing after 30% errors, showing that the DC distribution changes little beyond this point, as shown in Figure B6 in Appendix B, where the distributions shift upwards and become more peaked, converging at  $DC \approx 0.6$ . The RMKS behaves similarly for OIQR, but begins to increase again above 80% error. Again, this is observed in the distributions shown in Figure B3 in Appendix B, where the distributions for data with added errors shift to higher OIQR values and narrow, stabilising at peak OIQR of  $\approx 3.5$ , before starting to shift upwards again. This shows that each summary statistic captures a different aspect of the dynamics and that both are needed to fully quantify such dynamical changes. It also highlights that the change in dynamics in calibrated models due to measurement error depends on the original dynamics and the type of measurement error applied.

In reality, errors exist in all measured data simultaneously, and it is unclear how their effects on model dynamics might interact. To investigate this, the error study is repeated with all three errors implemented. The error study procedure is repeated five times and the average results are reported. To keep the number of error combinations at a manageable level, the number of sequence error values is reduced to (10, 30, 50, 80, 100) and only the (1, -1, 4), (-1, -4, 1), and (-2, -1, 1) are considered. These particular combinations were chosen because they are representative of the different RMKS behaviours over both summary statistics. Heatmaps of how the RMKS for both DC and OIQR change with both occupancy and sequence error at constant distance error values are shown in Appendix D. Comparing these with Figure 8, it is clear that combining errors does not lead to significant changes in how the dynamics vary with each error type. For example, the DC RMKS for (-1, -4, 1) combination shows the same oscillatory behaviour with changing sequence error in the combined error study as seen in the single error study. However, the OIQR RMKS for (1, -1, 4) does indicate that the impact of sequence errors does vary a little with occupancy error,

but this could be noise, as the change in RMKS is quite low here. With this exception, these results suggest that, in this particular case, these errors act independently in altering the dynamics of the calibrated model.

#### 4. Discussion and future work

The key contribution of this research is a systematic investigation of how measurement errors in data affect the calibration success of destination choice models in pedestrian dynamics. A novel and rigorous methodology for introducing errors to data is presented, inspired by sources of real measurement error. Findings confirm that measurement errors can cause significant bias in destination choice model parameters, decrease the ability of the model to make accurate predictions, and in some cases, can cause significant changes to the dynamics. A parameter scan and suitably chosen summary statistics show that a simple destination choice model, the multinomial logit model with three predictors, can generate a variety of pedestrian choice dynamics. Importantly, the effect of measurement errors on calibration success can depend on where in the parameter space of the model the true dynamics are situated.

The results concerning bias in calibrated model parameters are in agreement with previous work (Bhatta and Larsen 2011; Jang, Rasouli, and Timmermans 2017) and thus add to the consensus that measurement errors reduce the impact of their associated predictors on decision-making. However, the impact of errors directly related to one parameter on other model parameters is impossible to predict, as previously stated by Bhatta and Larsen (2011). The reduction in predictive power of the calibrated model with increasing errors also agrees with previous work (Bhatta and Larsen 2011; Jang, Rasouli, and Timmermans 2017). This result is not unexpected, as a model fitted to data which is generated using the same model will explain almost all variation and is likely to make very accurate predictions. Errors obfuscate the data, adding variability not explained by the model, and hence reduce the accuracy of any model predictions. The change in model dynamics in calibrated models compared to the true dynamics due to errors is less straightforward to explain. This effect depends on where in the parameter space of the model the true dynamics are located, and on the simulation environment considered.

Two main applications for this research are envisaged. First, as a guide to qualitatively inform researchers planning empirical data collection with a view to calibrating models where to direct their efforts. Second, as a way for researchers to gauge the likely calibration success for a model calibrated on given data. For both cases, the key observation of this research is that errors in some measurements have more of an impact on model calibration than others. Specifically, measurement errors in the predictors, in this case occupancy and distance, have a substantially weaker effect than errors in the sequence of visited destinations in the findings presented here. While the precise details of how errors affect model calibration are likely to also depend on the amount of data available, the model, and the specific context considered (see also discussion below), the findings presented here are a starting point for considerations on what level of error in which measurements is acceptable, if the goal is to obtain accurate and consistent parameter estimates and/or seek to replicate the observed dynamics. For example, if chosen or visited destinations are being inferred, either from positioning technologies like Wi-Fi or through surveys, great care must

be taken to minimise and quantify the error, as errors in these inferences can cause drastic changes to both calibration estimates and the observed dynamics.

This contribution is intended as a rigorous starting point for considering the effect measurement errors in data have on the calibration of destination choice models in pedestrian dynamics. There are several ways in which this research could be extended.

First, only one discrete choice modelling framework (multinomial logit) and one approach to calibration (MLE) was considered here, even though a range of different approaches have been used previously (Carroll et al. 2006; Dai 1998; Schennach 2016). The approach presented here for generating simulated data, applying errors to it, and assessing model calibration provides a recipe that is applicable across modelling and calibration methods.

Second, simulated data is used to test the effects of errors on model calibration. While this data allows us to know the true mechanisms which generated the data, it is unclear how well the results presented here translate to real data.

Third, the simulated environments considered here are comparatively small ( $20 \times 20$  m) and there is evidence to suggest that the layout of the environment can impact the results (see Appendix A). Thus, it may be necessary to perform separate error studies for environments that differ substantially from those shown here. This is part of a larger issue which is that the dynamics observed are dependent on the simulated environment, as the simulator also incorporates the microscopic movement of pedestrians. The extent to which the destination choice model itself depends on the environment would depend on the predictors involved and whether a normalisation of said predictors was undertaken. In this instance, occupancy and distance values depend on the environment and/or the microscopic behaviour. However, since they are normalised, these effects are minimised.

Fourth, in these scenarios, it is assumed that each destination has one unique activity that can be performed there. Whereas in reality, destinations can have several activities associated with them, depending on the activity and destination area specifications. For example, a coffee shop could have 'buying a coffee' and 'sitting at a table' as two possible activities or the shop could have two destinations within: the counter and the tables. There could also be multiple coffee shops available to a decision-maker to buy coffee. Lifting these constraints could therefore improve the realism of the scenarios.

Finally, the application of errors to data could be refined further. For example, it is more likely for destinations that are close to each other to be mistakenly identified using positioning technologies, while further apart destinations are highly unlikely to be mistaken for each other (Baba 2017; Rieser-Schüssler 2012). Thus, using a distance-based function for the probability of substituting one destination for another, rather than assuming equal substitution probabilities for all destinations could add another layer of realism to the research. Additionally, other distributions could have been used to sample distance errors, for example, a gamma or truncated normal distribution. However, it is unlikely that using a different distribution would cause any significant difference in the presented results, unless there was significant skew or kurtosis in such distributions. Moreover, knowledge of individuals' schedules has been assumed here with uncertainties about this only being considered indirectly via sequence errors. Thus, an important extension could be to either explicitly infer schedules from data or to define errors to desirability that are meaningful and reflect uncertainties in estimating the schedules of individuals (e.g. via surveys). Inferring schedules from data could be achieved by inferring the most likely sequences of desired locations

from data using Hidden Markov Models in a similar approach to methodology that has been used for inferring behavioural states in animals (Bode and Seitz 2017).

## 5. Conclusions

This paper investigates the effect of measurement errors on discrete choice model calibration in the context of pedestrian destination choice. A novel coupling of a destination choice model to an agent-based simulator is presented. Characterising the dynamics of this model through a parameter scan shows that a wide range of dynamical regimes are attainable. A novel protocol for adding errors to different measured data inspired by real error sources is applied to simulated data to observe what effect errors have on choice model calibration. The results show how different types of errors affect calibration success as measured by bias in parameter estimates, prediction accuracy, and changes in the dynamics in simulations of calibrated models compared to the true data. Errors in the sequence of destinations visited by individuals had a substantially stronger detrimental effect on calibration success than errors related to properties of destinations, such as the number of people visiting them or the distance between them. This work presents a principled starting point for informing data collection protocols for empirical pedestrian destination choice research and for indicating the likelihood of a successful calibration of models on available data.

## Acknowledgements

The authors would like to acknowledge Professor Eddie Wilson for many helpful discussions, ideas, and insights.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

Christopher King acknowledges funding via an Engineering and Physical Sciences Research Council Doctoral Training Partnership scholarship.

## ORCID

Christopher King  <http://orcid.org/0000-0002-3172-2941>

N. W. F. Bode  <http://orcid.org/0000-0003-0958-5191>

## References

- Arentze, Theo A., Dick Ettema, and Harry J.P. Timmermans. 2013. "Location Choice in the Context of Multi-Day Activity-Travel Patterns: Model Development and Empirical Results." *Transportmetrica A: Transport Science* 9 (2): 107–123.
- Arentze, Theo A., and Harry J. P. Timmermans. 2004. "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38 (7): 613–633.
- Baba, Asif Iqbal. 2017. "Removing Object Bouncing from Indoor Tracking Data." In *Proceedings of the SouthEast Conference, ACM SE '17*, New York, NY, USA, Apr., 134–141. Association for Computing Machinery.



- Beaulieu, Alexandra, and Bilal Farooq. 2019. "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.
- Benezeth, Y., H. Laurent, B. Emile, and C. Rosenberger. 2011. "Towards a Sensor for Detecting Human Presence and Characterizing Activity." *Energy and Buildings* 43 (2): 305–314.
- Bhatta, Bharat P., and Odd I. Larsen. 2011. "Errors in Variables in Multinomial Choice Modeling: A Simulation Study Applied to a Multinomial Logit Model of Travel Mode Choice." *Transport Policy* 18 (2): 326–335.
- Bode, Nikolai W. F., and Edward A. Codling. 2013. "Human Exit Route Choice in Virtual Crowd Evacuations." *Animal Behaviour* 86 (2): 347–358.
- Bode, Nikolai W. F., and Michael J. Seitz. 2017. "Using Hidden Markov Models to Characterise Intermittent Social Behaviour in Fish Shoals." *The Science of Nature* 105 (1): 7.
- Borgers, A., and H. J. P. Timmermans. 1986. "City Centre Entry Points, Store Location Patterns and Pedestrian Route Choice Behaviour: A Microlevel Simulation Model." *Socio-Economic Planning Sciences* 20 (1): 25–31.
- Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. CRC Press. Google-Books-ID: 9kBx5CPZCqkC.
- Dai, Jian. 1998. "Calibration and Test of a Discrete Choice Model with Endogenous Choice Sets." *Geographical Analysis* 30.
- Danalet, Antonin. 2015. "Activity Choice Modeling for Pedestrian Facilities." PhD diss., Swiss Federal Institute of Technology, Lausanne.
- Dijkstra, Jan, Harry J. P. Timmermans, and Bauke de Vries. 2013. "Activation of Shopping Pedestrian Agents—Empirical Estimation Results." *Applied Spatial Analysis and Policy* 6 (4): 255–266. Accessed 29 October 2018. <https://link-springercom.bris.idm.oclc.org/article/10.1007/s12061-012-9082-3>.
- Ettema, Dick, Fabian Bastin, John Polak, and Olu Ashiru. 2007. "Modelling the Joint Choice of Activity Timing and Duration." *Transportation Research Part A: Policy and Practice* 41 (9): 827–841. Accessed 24 July 2019. <https://www-sciencedirectcom.bris.idm.oclc.org/science/article/pii/S0965856407000328>.
- Ettema, D., Aloys Borgers, and Harry Timmermans. 1993. "Simulation Model of Activity Scheduling Behavior." *Transportation Research Record*, 1–11.
- Fahmy, S. A., B. A. Alablani, and T. F. Abdelmaguid. 2014. "Shopping Center Design Using a Facility Layout Assignment Approach." In 2014 9th International Conference on Informatics and Systems, Dec., ORDS–1–ORDS–7.
- Fesenmaier, D. R. 1988. "Integrating Activity Patterns into Destination Choice Models." *Journal of Leisure Research* 20 (3): 175–191.
- Greenwood, David, Shrikant Sharma, and Anders Johansson. 2015. "Mobility Modelling in a Process Constrained Environment: Modelling the Movements of Nurses in a Neonatal Intensive Care Unit." In *Traffic and Granular Flow '13*, edited by Mohcine Chraïbi, Maik Boltes, Andreas Schadschneider, and Armin Seyfried, 233–241. Cham: Springer International Publishing. Accessed 22 March 2021. <https://www.semanticscholar.org/paper/Mobility-Modelling-in-a-Process-Constrained-the-of-Greenwood-Sharma/f8a8c139eb0657f6f2863df4419ed6cd43b8a110>.
- Hausman, J. A., Jason Abrevaya, and F. M. Scott-Morton. 1998. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87 (2): 239–269.
- Helbing, Dirk, Illés Farkas, and Tamás Vicsek. 2000. "Simulating Dynamical Features of Escape Panic." *Nature* 407 (6803): 487–490. Accessed 16 October 2018. <https://www.nature.com/articles/35035023>.
- Hoogendoorn, S. P., and P. H. L. Bovy. 2004. "Pedestrian Route-Choice and Activity Scheduling Theory and Models." *Transportation Research Part B: Methodological* 38 (2): 169–190.
- Hui, Sam K., Eric T. Bradlow, and Peter S. Fader. 2009. "Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior." *Journal of Consumer Research* 36 (3): 478–493. Accessed 21 March 2019. <https://academic.oup.com/jcr/article/36/3/478/1847025>.

- Jang, Sunghoon, Soora Rasouli, and Harry Timmermans. 2017. "Bias in Random Regret Models Due to Measurement Error: Formal and Empirical Comparison with Random Utility Model." *Transportmetrica A: Transport Science* 13 (5): 405–434. Publisher: Taylor & Francis eprint: doi:10.1080/23249935.2017.1285366.
- Kao, Chihwa, and John F. Schnell. 1987. "Errors in Variables in the Multinomial Response Model." *Economics Letters* 25 (3): 249–254.
- Kielar, Peter M., and André Borrmann. 2016. "Modeling Pedestrians' Interest in Locations: A Concept to Improve Simulations of Pedestrian Destination Choice." *Simulation Modelling Practice and Theory* 61: 47–62.
- Kurose, Shigeyuki, Aloys W J Borgers, and Harry J P Timmermans. 2001. "Classifying Pedestrian Shopping Behaviour According to Implied Heuristic Choice Rules." *Environment and Planning B: Planning and Design* 28 (3): 405–418. Accessed 9 January 2019. <https://journals.sagepub.com/doi/10.1068/b2622>.
- Kwak, Jaeyoung, Hang-Hyun Jo, Tapio Luttinen, and Iisakki Kosonen. 2014. "Modeling Pedestrian Switching Behavior for Attractions." *Transportation Research Procedia* 2: 612–617.
- Li, Weizi, and Jan M. Allbeck. 2011. "Populations with Purpose." In *Motion in Games*, edited by Jan M. Allbeck and Petros Faloutsos, 132–143. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer.
- Li, Yuan, Linchuan Yang, Han Shen, and Zhonglong Wu. 2019. "Modeling Intradestination Travel Behavior of Tourists Through Spatio-Temporal Analysis." *Journal of Destination Marketing & Management* 11: 260–269. Accessed 11 August 2020. <https://www.sciencedirect.com/science/article/pii/S2212571X17303839>.
- McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3 (4): 303–328.
- Meyer, Bruce D., and Nikolas Mittag. 2017. "Misclassification in Binary Choice Models." *Journal of Econometrics* 200 (2): 295–311.
- Miller, Harvey J. 2014. "Activity-Based Analysis." In *Handbook of Regional Science*, edited by Manfred M. Fischer and Peter Nijkamp, Vol. 2, 705–724. Berlin: Springer Berlin Heidelberg.
- Murakami, Hisashi, Claudio Feliciani, Kenichiro Shimura, and Katsuhiro Nishinari. 2020. "A System for Efficient Egress Scheduling During Mass Events and Small-Scale Experimental Demonstration." *Royal Society Open Science* 7 (12): 201465. Publisher: Royal Society.
- n.d.a. "CRAFTSMAN 165-ft Outdoor Laser Distance Measurer with Backlit Display Lowes.com." Accessed 24 June 2021. <https://www.lowes.com/pd/CRAFTSMAN-165-ftIndoor-Outdoor-Laser-Distance-Measurer-with-Backlit-Display/1001091160>.
- n.d.b. "FLIR T560 Professional Thermal Camera | Teledyne FLIR." Accessed 24 June 2021. <https://www.flir.com/products/t560/>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rieser-Schüssler, Nadine. 2012. "Capitalising Modern Data Sources for Observing and Modelling Transport Behaviour." *Transportation Letters* 4 (2): 115–128.
- Ruiz, Jean-Paul, Jean-Charles Chebat, and Pierre Hansen. 2004. "Another Trip to the Mall: A Segmentation Study of Customers Based on Their Activities." *Journal of Retailing and Consumer Services* 11 (6): 333–350.
- Saarloos, Dick, Akimasa Fujiwara, and Junyi Zhang. 2007. "The Interaction Between Pedestrians and Facilities in Central Business Districts: An Explorative Case Study." *Journal of the Eastern Asia Society for Transportation Studies* 7: 1870–1885. Publisher: Eastern Asia Society for Transportation Studies.
- Schennach, Susanne M. 2016. "Recent Advances in the Measurement Error Literature." *Annual Review of Economics* 8 (1): 341–377. eprint: doi:10.1146/annurev-economics080315-015058.
- Scovanner, Paul, and Marshall F Tappen. 2009. "Learning Pedestrian Dynamics from the Real World." In 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Sep., 381–388. IEEE.
- Tinguely, LÖic. 2015. "Exploiting Pedestrian WiFi Traces for Destination Choice Modeling." Master's thesis, EPFL, Lausanne, Switzerland. Accessed 1 April 2021. <https://infoscience.epfl.ch/record/209732>.

- Ton, D. 2014. "Navistation: A Study into the Route and Activity Location Choice Behaviour of Departing Pedestrians in Train Stations." PhD diss., Delft University of Technology.
- Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge; New York: Cambridge University Press. OCLC: ocn349248337, Accessed 16 February 2021. <https://eml.berkeley.edu/books/train1201.pdf>.
- van der Hagen, Xavier, Aloys Borgers, and H. J. P. Timmermans. 1991. "Spatiotemporal Sequencing Processes of Pedestrian in Urban Retail Environments." *The Journal of the RSAI* 70 (1): 37–52. Accessed 21 January 2019. <https://link-springercom.bris.idm.oclc.org/article/10.1007/BF01463442>.
- Venkatnarayan, Raghav Hampapur, and Muhammad Shahzad. 2019. "Demo: Measuring Distance Traveled by an Object Using WiFi-CSI and IMU Fusion." In 2019 IEEE 27th International Conference on Network Protocols (ICNP), Oct., 1–2. ISSN: 2643-3303.
- Viswanathan, Madhu. 2005. *Measurement Error and Research Design*. Thousand Oaks: Sage.
- Wang, Donggen, and Jiukun Li. 2011. "A Two-Level Multiple Discrete-Continuous Model of Time Allocation to Virtual and Physical Activities." *Transportmetrica* 7 (6): 395–416.
- Yang, Yang, Timothy Fik, and Jie Zhang. 2013. "Modeling Sequential Tourist Flows: Where is the Next Destination?" *Annals of Tourism Research* 43: 297–320. Accessed 19 July 2019. <https://www.sciencedirect.com/science/article/abs/pii/S0160738313001023>.
- Ying, Fabian, Alisdair O. G. Wallis, Mariano Beguerisse-Díaz, Mason A. Porter, and Sam D. Howison. 2019. "Customer Mobility and Congestion in Supermarkets." *Physical Review E* 100, ArXiv: 1905.13098, Accessed 25 March 2021. <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.100.062304>.
- Yoshimura, Yuji, Stanislav Sobolevsky, Carlo Ratti, Fabien Girardin, Juan Pablo Carrascal, Josep Blat, and Roberta Sinatra. 2014. "An Analysis of Visitors' Behavior in the Louvre Museum: A Study Using Bluetooth Data." *Environment and Planning B: Planning and Design* 41 (6): 1113–1131. Accessed 4 March 2019. <https://journals.sagepub.com/doi/10.1068/b130047p>.
- Zhu, Wei, Harry Timmermans, and Wang De. 2006. "Temporal Variation in Consumer Spatial Behavior in Shopping Streets." *Journal of Urban Planning and Development* 132 (3): 166–171.

## Appendices

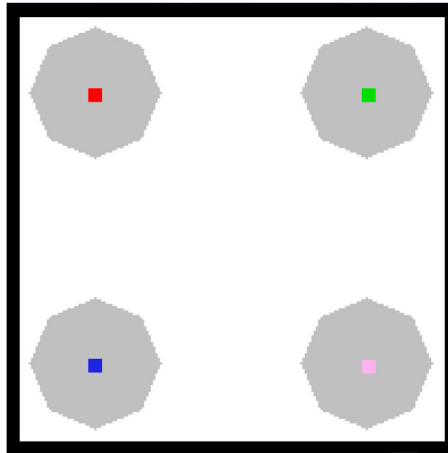
### Appendix A. The empty room

This appendix details results for an error study on another setting, an empty room, shown in Figure A1. These additional results for a highly abstracted and simple setting serve to indicate the relevance of the spatial configurations of destinations on the calibration process in the presence of errors. The space is  $20 \times 20$  m with the destinations arranged at the corners of a square centred at the middle of the space with sides of length 16m. There are no obstacles besides the bounding outer walls.

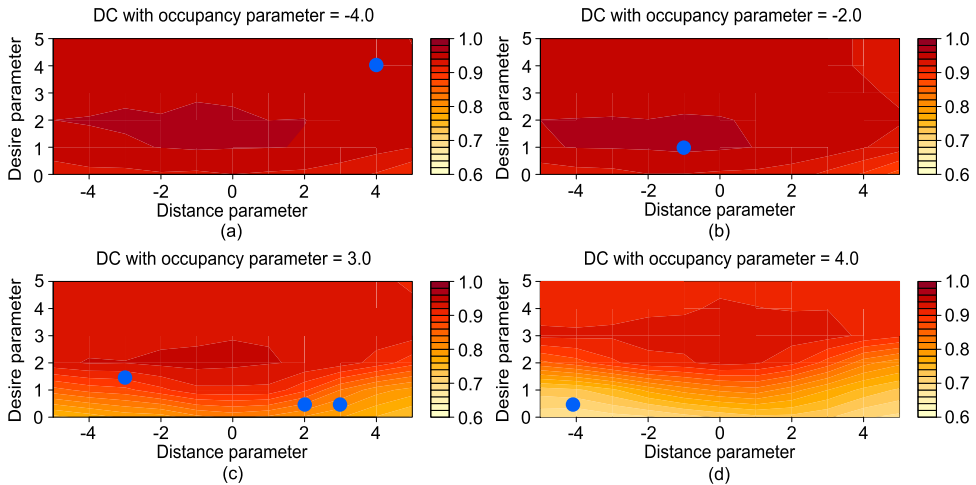
The entire error study process, summarised in Figure 3, was conducted on this environment with the simulator conditions detailed in Table A1.

**Table A1.** Simulator inputs for the error study on an empty room. All other simulator conditions are identical to those used in the main text.

Input	Value
Number of agents	100
Timestep / s	0.05
Number of Timesteps	10,000
Simulation Replicates	100
Number of destinations	4



**Figure A1.** Empty room environment. White space indicates free space outside of destinations, grey is free space within a destination area. Coloured squares are destination centres. Obstacles and walls are shown in black. The discrete nature of the floor field is visible.

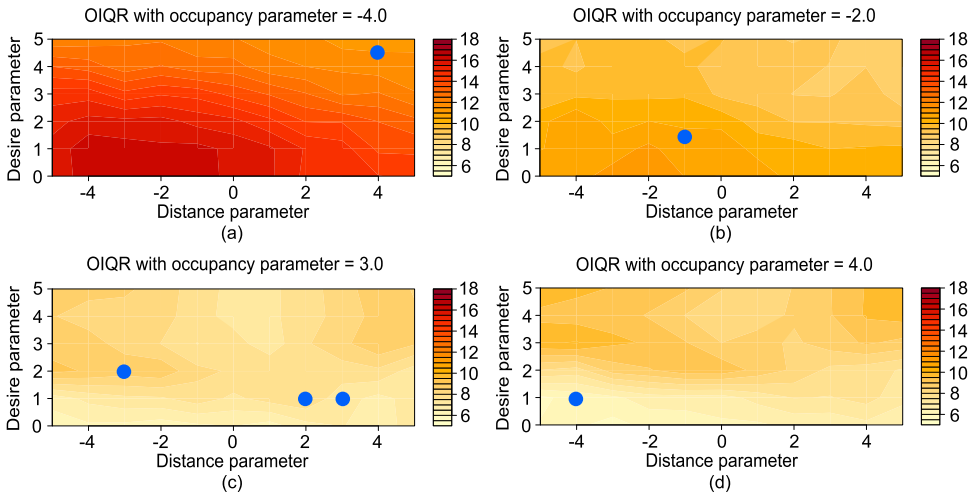


**Figure A2.** Destination coverage for different distance, desire, and occupancy parameter values. The combinations used in the error study are marked by blue dots. (DC = Destination Coverage).

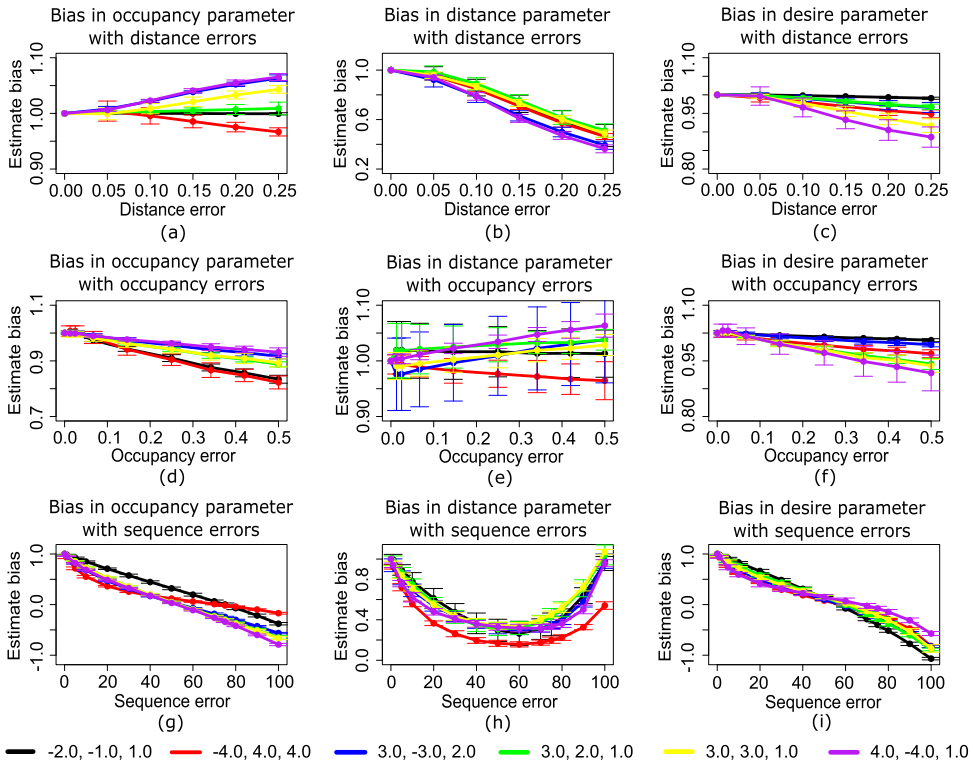
As for the environment considered in the main text, there was no variation in  $\beta_{occ}$ ,  $\beta_{dist}$ , and  $\beta_{des}$  across agents. The destination schedule for all agents was the sequence '1234' repeated 100 times, so that every destination would have a non-negligible desirability for every decision made by every agent during simulation. The parameter scan was performed with the same constraints on  $\beta_{occ}$ ,  $\beta_{dist}$ , and  $\beta_{des}$  as in the main text.

Results from a parameter scan are shown in Figures A2 and A3. As for the environment considered in the main text, the combination (0,0,0) acts as the reference case. The following additional combinations are chosen based on their high or low DC and/or OIQR:  $(-2, -1, 1)$ ,  $(3, 2, 1)$ ,  $(3, 3, 1)$ ,  $(3, -3, 2)$ ,  $(4, -4, 1)$ ,  $(-4, 4, 4)$ , these combinations are marked by blue dots in Figures A2 and A3.

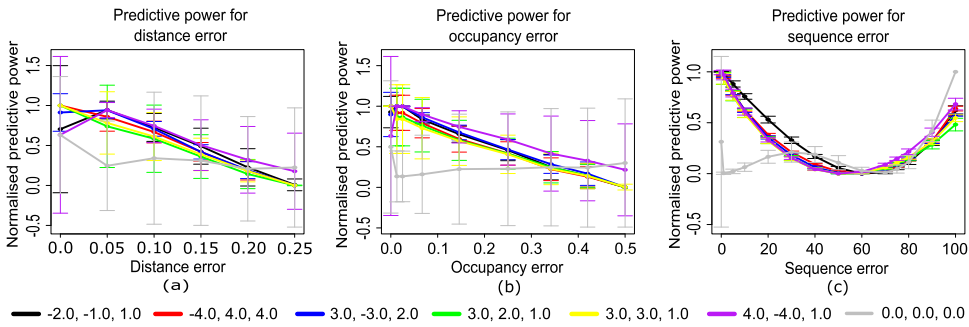
Figures A2 and A3 show that DC is close to one in the vast majority of combinations considered, with the smallest DC at around 0.7. This is due to the environment and simulator conditions. There are only four destinations in this environment, with no obstacles to impede agent's progress, so in



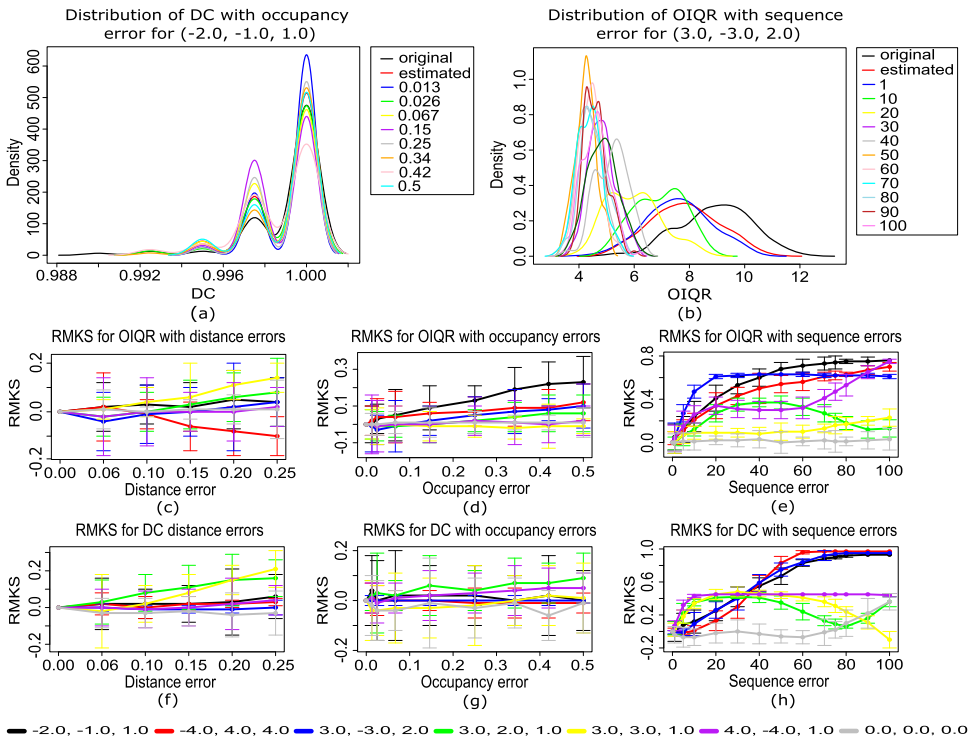
**Figure A3.** Occupancy Interquartile range for different distance, desire, and occupancy parameter values. The combinations used in the error study are marked by blue dots. (OIQR = Occupancy Interquartile Range).



**Figure A4.** Average bias in each model parameter due to each type of error. Error bars show two times the standard deviation of bias over 30 repetitions of adding errors to perfect data. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ .



**Figure A5.** Average normalised predictive power of the choice model for all combinations with each type error. Error bars show two times the standard deviation of predictive power over 30 repetitions of adding errors to perfect data. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ .



**Figure A6.** Average difference in median erroneous KS values and median perfect KS values for destination coverage and occupancy interquartile range for every parameter combination plotted against each error type. An example of when the summary statistic distributions change significantly (a) and minimally (b) are shown to illustrate the kinds of differences seen between distributions. Error bars show two times the standard deviation of RMKS over 30 repetitions of adding errors to perfect data. Parameter combinations are shown in the format  $\beta_{occ}, \beta_{dist}, \beta_{des}$ . (DC = Destination Coverage, OIQR = Occupancy Interquartile Range).

10,000 timesteps, the vast majority of agents are able to visit all four destinations, except in extreme cases. These extreme cases occur when the distance parameter is large, occupancy parameter is large and positive, and desire parameter is small. In these situations, agents will be driven to visit either the destinations at opposite vertices only (distance parameter positive) or destinations at either of the neighbouring vertices (distance parameter negative), since the square configuration means two destinations are comparable in distance, with the other being further away. The positive occupancy parameter means agents will be drawn to the busier destinations. These two effects together mean that it is likely that not all destinations will be visited by all agents. These results, combined with those in the main text, show that though the environment and initial simulation conditions used strongly influence the values of the two summary statistics, a wide range of behaviours can be captured.

Figure A4 shows the bias in each parameter estimate with each type of error for the empty room and can be compared to Figure 6 in the main text. First, notice that occupancy errors have a negligible effect on the distance parameter and vice-versa, as seen in the main text, though there are significant differences in the trend of occupancy bias with distance error across combinations. This reinforces the notion that these two predictors are independent of each other and this does not depend too much on environment. Another result shared by both environments is that distance and occupancy errors cause downward bias in their associated parameters, confirming the results of previous work (Bhatta and Larsen 2011; Jang, Rasouli, and Timmermans 2017). Both environments also show a small downward bias in desire parameter with both distance and occupancy errors, though this is more pronounced for the empty room. This shows that there is some correlation between desirability and the other predictors that is only partially explained by the environment and the initial simulator conditions, such as the initial schedules of agents.

Sequence errors have the greatest impact on estimate bias with occupancy and desire parameters obtaining an almost equal and opposite value from their original estimates. The effect on distance parameter is the most interesting – all combinations show a parabolic relationship in estimate bias with sequence error, with a minimum at around sequence error = 60%. This implies that the effect of distance is actually recovered when over half the chosen destinations are altered. Two effects contribute to this: the destination layout and the new next destination constraint in destination substitutions. There are only four destinations, and for any one destination, there are two of which are equidistant and one which is further away. The new next destination constraint means that there are at most only two possible destinations that can be substituted, as the destination must be different from the current one, and cannot be the same as the destinations before and after it, if all these are different, then there is only one destination that can be substituted, if the destinations before and after are the same, then there are two possible substitutions. In the case where distance is strongly negative, the previous and subsequent destinations are likely to be one of the closer destinations, if they are distinct (the agent is moving around the square), then the only possible destination that can be substituted is the one further away, otherwise, the further one has a 50% chance of being substituted. In the case where distance parameter is strongly positive, then it is likely that both the previous and subsequent destinations are the one furthest from the current destination, so it is likely that this destination will be substituted for one of the closer ones. These effects are reduced in asymmetric environments with more destinations, as can be seen in the main text.

Figure A5 shows the predictive power for the empty room in analogy with Figure 7 in the main text. As for the environment in the main text, the predictive power consistently decreases as distance and occupancy error increase for all non-zero parameter combinations, as expected. (0, 0, 0), however, seems relatively unchanged with distance or occupancy errors, with large variance obfuscating any clear trends. This makes sense, as adding random noise to an already random model will not have any impact on the accuracy of such predictions. With sequence errors, there is a clear parabolic effect, with a minimum at sequence error = 60%. This shows that a new data structure emerges when the majority of destinations are altered by substitution which satisfies the new next destination constraint. In tandem with the sequence error biases, this data is best fit by a model with occupancy and desire parameters with similar magnitudes but opposite signs to the original and the same distance parameter value. It could be argued that (0, 0, 0) follows the rest of the combinations for high sequence error, this too supports the idea that a new, non-random structure is emerging in the sequences which can be explained by a non-random model.

The RMKS for the empty room for each summary statistic and error type are shown in Figure A6, along with representative examples of how similar and different the estimated summary statistic distributions are compared to the originals. This figure is analogous to Figure 8 in the main text.

The RMKS for both summary statistics are not greatly affected by distance errors, though the exact trend in RMKS depends largely on the parameters considered. It is interesting that OIQR RMKS  $(-4, 4, 4)$  becomes slightly negative with distance errors, indicating that the resultant summary statistic distributions actually become more similar to the original summary statistic distribution than the distribution obtained using estimates from non-erroneous data.

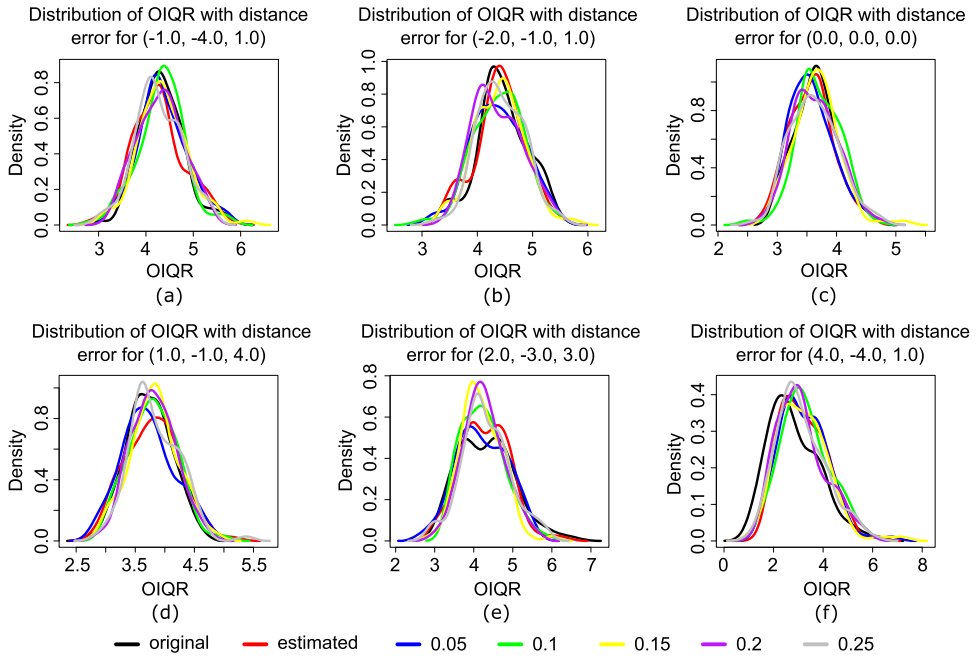
Occupancy errors also have little effect on the RMKS for both summary statistics for all combinations. However, it could be argued that  $(-2, -1, 1)$  OIQR RMKS increases, but the large error bars indicate large variation, showing that this may not be a significant effect. This is supported by Figure A6a, which shows that the overall shape of the distribution does not change much with occupancy error, but the height of the peaks varies and the amount by which each peak changes depends on the error value considered. This suggests that occupancy errors do not affect the calibration of the model substantially enough to result in different dynamics in most cases, as measured by DC and OIQR.

As with the horseshoe, sequence errors have the most dramatic effect on RMKS for both summary statistics, with each combination considered showing unique behaviour in one or both summary statistics.  $(0,0,0)$  demonstrates little change in OIQR RMKS and for DC RMKS up to 60% error. However, beyond this, DC RMKS begins to increase, indicating that the dynamics are changing in a systematic way that is not captured by the random model, lending further credence to the idea that a new data structure is created from the sequence errors.  $(3, 2, 1)$ ,  $(3, 3, 1)$ , and  $(4, -4, 1)$  all show oscillatory behaviour in at least one RMKS plot. This is despite the significant change in parameter estimates as sequence error increases (see Figure A4), showing that different model parameters can produce similar dynamics in this environment, as measured by DC and OIQR. Despite having almost identical parameter values,  $(3, 2, 1)$  and  $(3, 3, 1)$  show quite distinct trends in each RMKS plot. This would indicate that the summary statistics can be very sensitive to small changes in parameter values. The OIQR RMKS of combination  $(4, -4, 1)$  rises quickly before plateauing after sequence errors  $> 20\%$ . This can be explained by Figure A6b, which shows that OIQR distributions become very similar for  $> 20\%$  error, peaking at lower OIQR and showing a decreased spread of possible values. This could also be the case for the DC RMKS for  $(3, -3, 2)$ , which shows similar behaviour. These results show that the extent to which the dynamics of the calibrated model is altered compared to the original dynamics due to sequence errors is incredibly varied and depends in part on the original parameter values and the summary statistics used to specify the dynamical space. A similar diversity of RMKS trends is seen for the horseshoe environment, indicating that the environment may not be an important factor when assessing the change in dynamics due to errors.

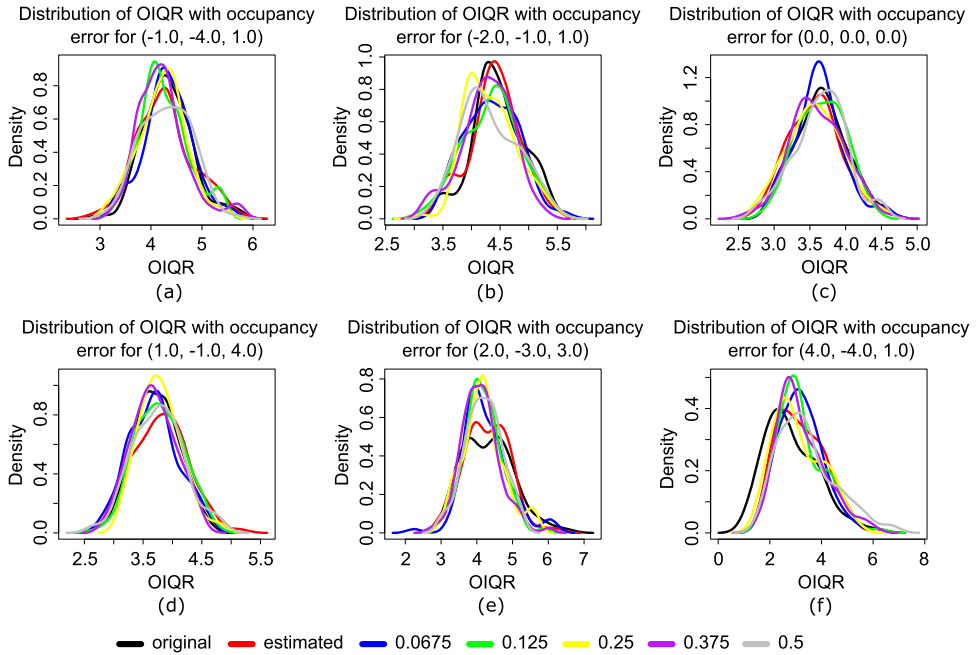
## Appendix B. Summary statistic distributions

This appendix demonstrates how the distributions of both summary statistics, destination coverage (DC) and occupancy interquartile range (OIQR), from simulations run using the choice model described in Equation (1) can vary when estimated on data containing differing amounts of error. These distributions come from one of the 10 error study replicates completed on the horseshoe environment (Figure 2). The differences between these distributions underpin the relative median Kolmogorov-Smirnov (RMKS) results shown in Figure 8.

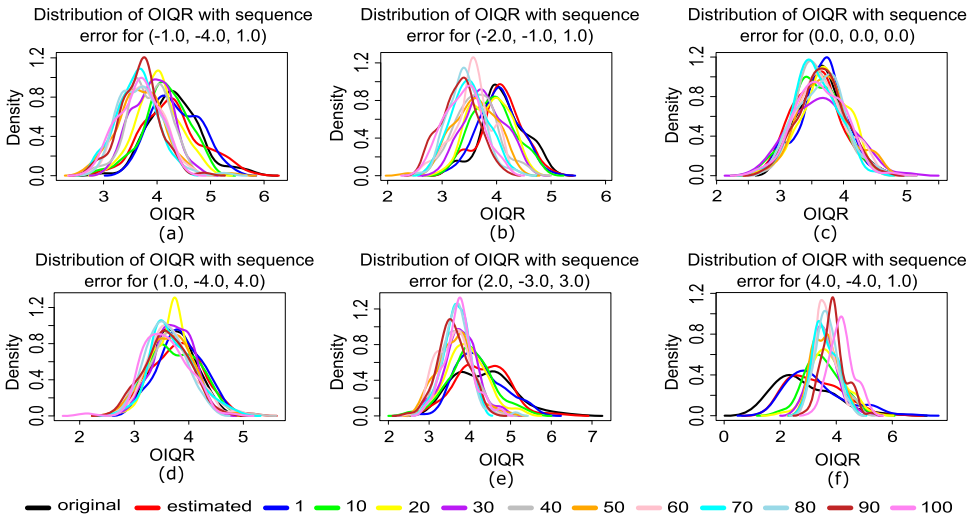




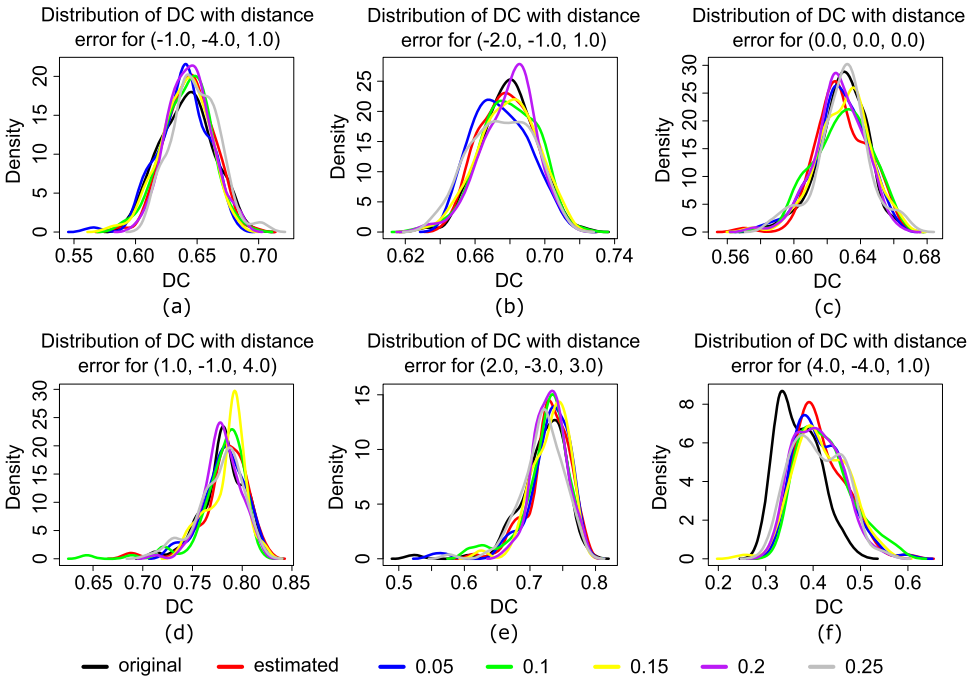
**Figure B1.** The occupancy interquartile range (OIQR) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of distance error added for one error study replicate.



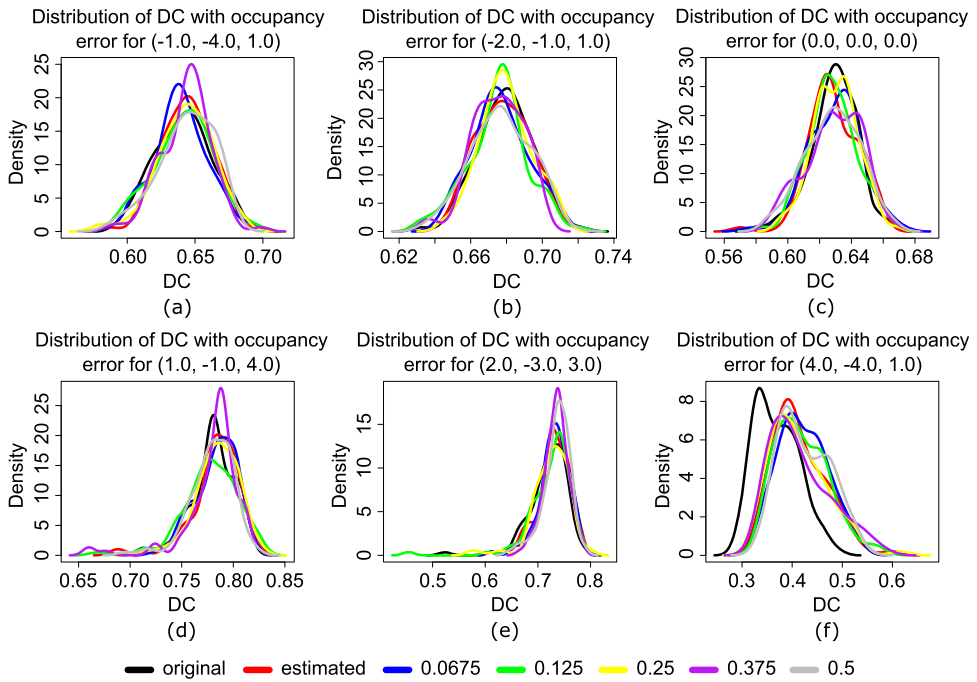
**Figure B2.** The occupancy interquartile range (OIQR) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of occupancy error added for one error study replicate.



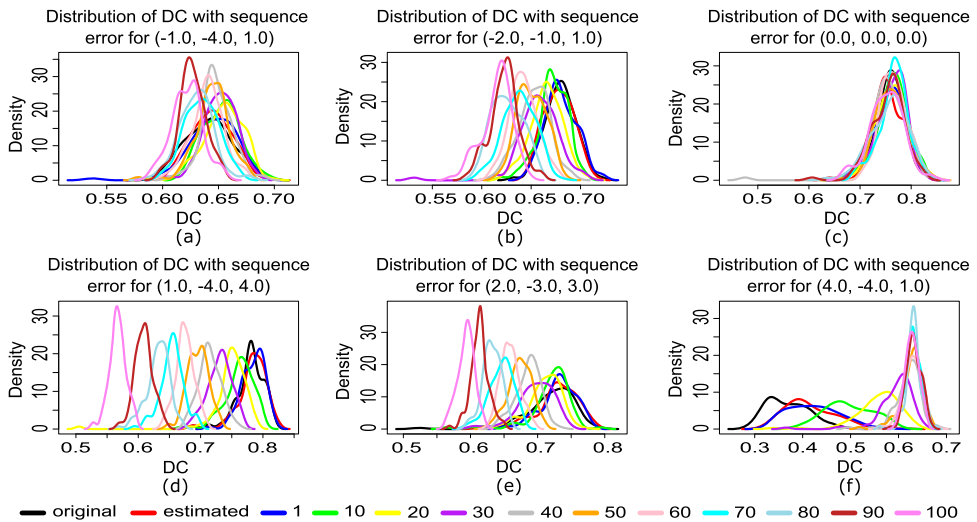
**Figure B3.** The occupancy interquartile range (OIQR) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of sequence error added for one error study replicate.



**Figure B4.** The destination coverage (DC) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of distance error added for one error study replicate.



**Figure B5.** The destination coverage (DC) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of occupancy error added for one error study replicate.



**Figure B6.** The destination coverage (DC) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of sequence error added for one error study replicate.

### Appendix C. Distance and occupancy error distributions

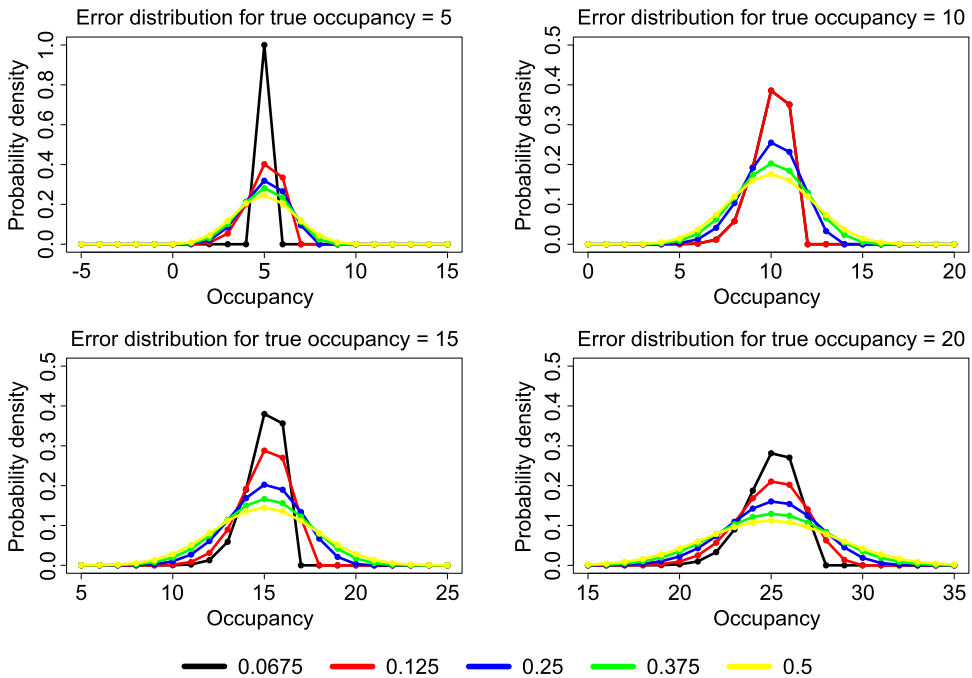
This appendix illustrates how the sampling distributions of both occupancy and distance errors change when their associated parameters change. As explained in Section 2.3.1, the occupancy and distance errors are sampled from a binomial distribution with parameters  $m$  and  $p$ , and a log-normal distribution with parameters  $\nu$  and  $\gamma$ , respectively.

Figure C1 shows how the shape of the sampling distribution for occupancy errors changes with both error value ( $1 - p$ ) and true occupancy. As mentioned in Section 2.3.1, the true occupancy is set to the mean and by using the error value,  $m$  can be calculated. Table C1 shows the values of each parameter for each true occupancy value.

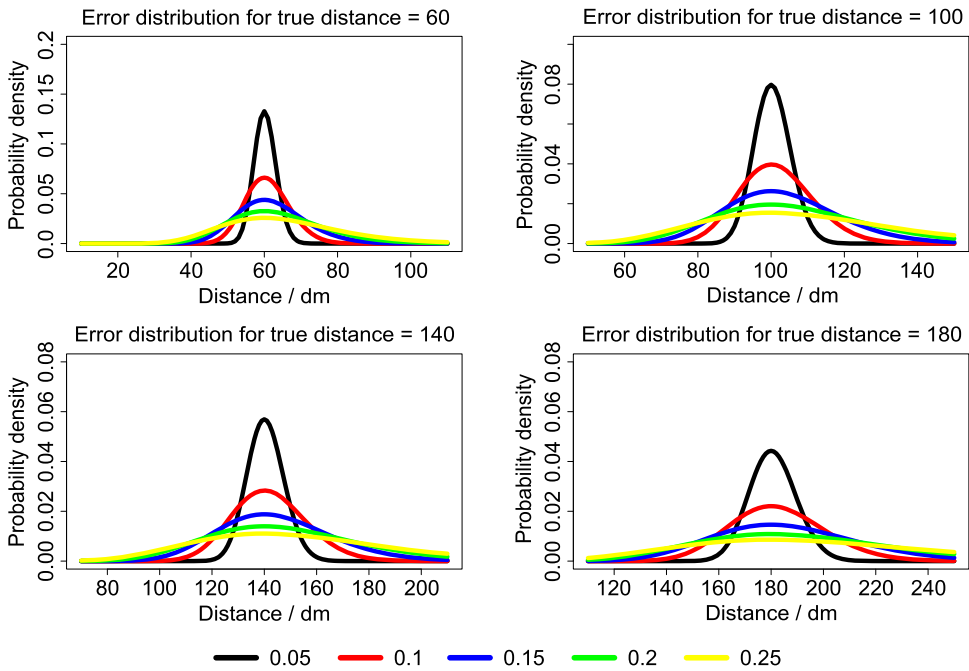
From Figure C1, it is clear that the spread of possible occupancy values increases with both the error value and the true occupancy. This reflects the fact that it is harder to get an accurate count when there are large numbers of people in an area.

Figure C2 shows how the shape of the sampling error distribution changes with both the true distance measured in the simulation and the error value,  $\gamma$ . As mentioned in the main text,  $\nu$  is calculated by setting the mode of the distribution as the true distance and solving for  $\nu$  using the different values of  $\gamma$ . Table C2 shows the values of each parameter for each true distance value.

Figure C2 shows that, in analogy with occupancy errors, the spread of possible distance values increases with both true distance and  $\gamma$ . This reflects that some distance-measuring instruments become less accurate as the measured distance increases, e.g. (n.d.a; n.d.b; Venkatnarayan and Shahzad 2019).



**Figure C1.** The destination coverage (DC) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of sequence error added for one error study replicate.



**Figure C2.** The destination coverage (DC) distribution from the original model parameter values vs the distributions for the estimates obtained from the unaltered data and from the data with different amounts of occupancy error added for one error study replicate.

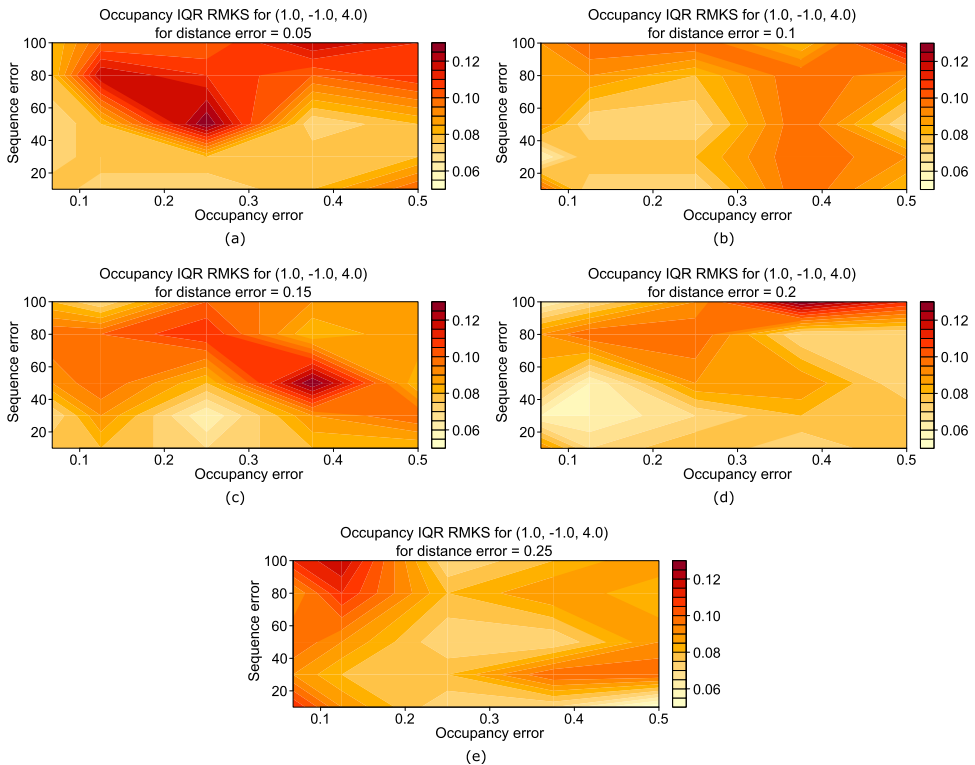
**Table C1.** Occupancy error distribution parameter values with true occupancy values and occupancy error values shown in Figure C1.

True Occupancy	$m$ 3 s.f.	Occupancy error 3 s.f.	$p$ 4 d.p.
5	5	0.0675	1.0000
10	11	0.0675	0.9091
15	16	0.0675	0.9375
20	21	0.0675	0.9524
5	6	0.125	0.8331
10	11	0.125	0.9091
15	17	0.125	0.8824
20	23	0.125	0.8170
5	7	0.250	0.7143
10	13	0.250	0.7692
15	20	0.250	0.7500
20	27	0.250	0.7407
5	8	0.375	0.6250
10	16	0.375	0.6250
15	24	0.375	0.6250
20	32	0.375	0.6250
5	10	0.500	0.5000
10	20	0.500	0.5000
15	30	0.500	0.5000
20	40	0.500	0.5000

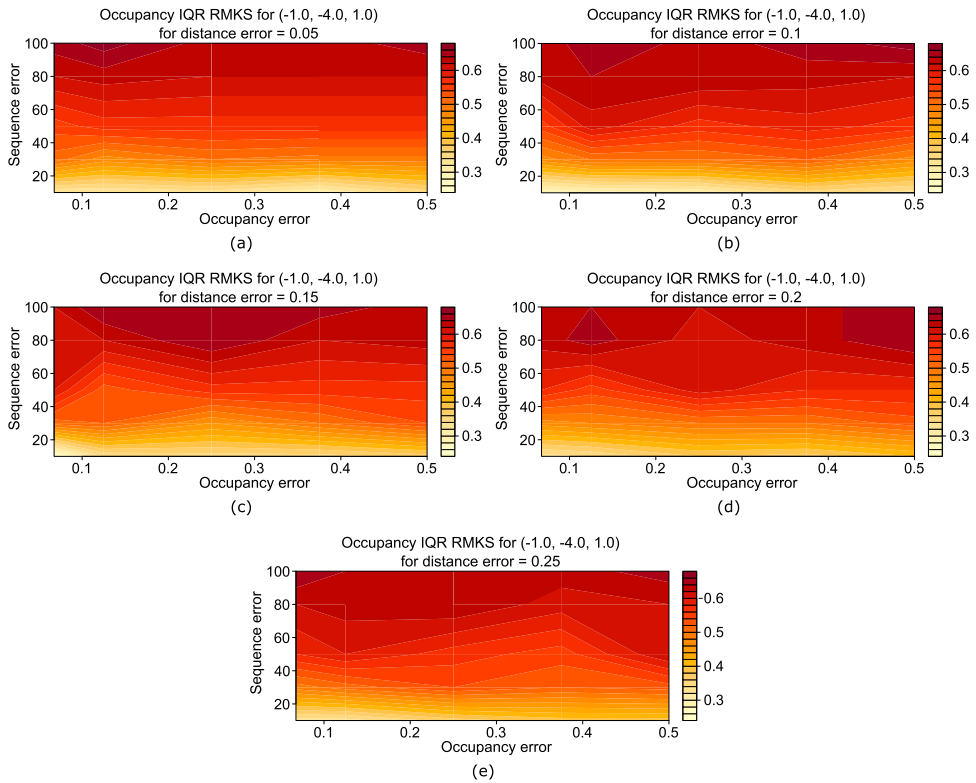
**Table C2.** Distance error distribution parameter values with true distance values and distance error values shown in Figure C2.

True Distance / dm	$\nu$ 3 s.f.	$\sigma$ 2 d.p.
60	4.10	0.05
100	4.61	0.05
140	4.94	0.05
180	5.20	0.05
60	4.10	0.10
100	4.62	0.10
140	4.95	0.10
180	5.20	0.10
60	4.11	0.15
100	4.63	0.15
140	4.96	0.15
180	5.22	0.15
60	4.13	0.20
100	4.65	0.20
140	4.98	0.20
180	5.23	0.20
60	4.16	0.25
100	4.67	0.25
140	5.00	0.25
180	5.26	0.25

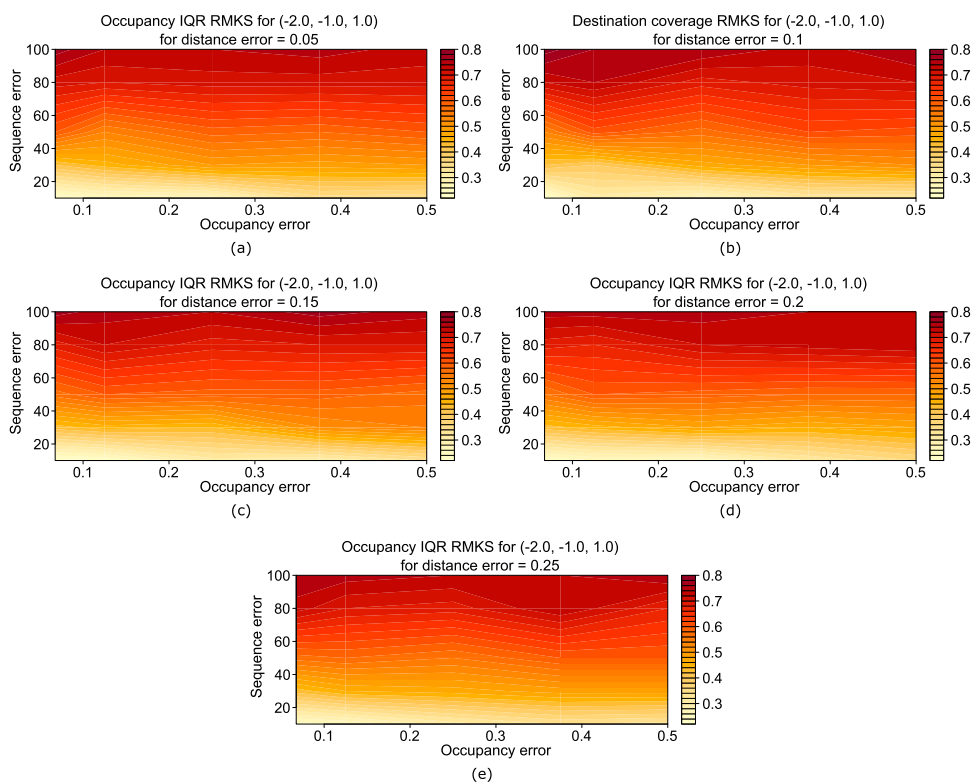
### Appendix D. Combined error study RMKS



**Figure D1.** Heatmaps of the average OIQR RMKS for (1, -1, 4) over five error study replicates against both occupancy and sequence errors at constant distance errors.

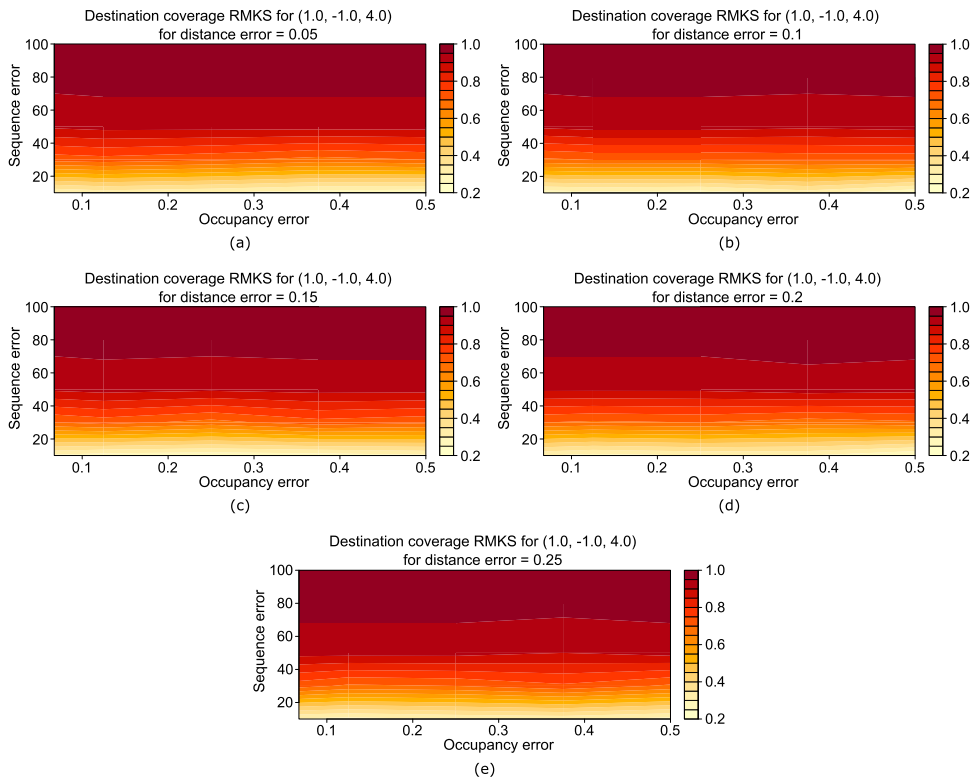


**Figure D2.** Heatmaps of the average OIQR RMKS for  $(-1, -4, 1)$  over five error study replicates against both occupancy and sequence errors at constant distance errors.

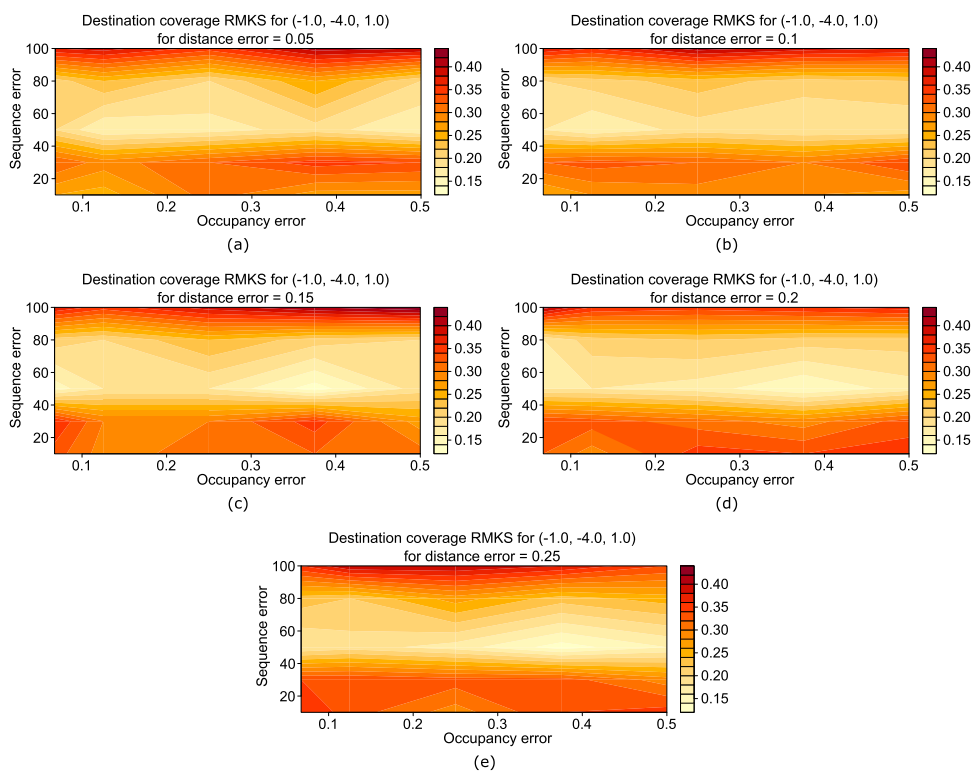


**Figure D3.** Heatmaps of the average OIQR RMKS for  $(-2, -1, 1)$  over five error study replicates against both occupancy and sequence errors at constant distance errors.

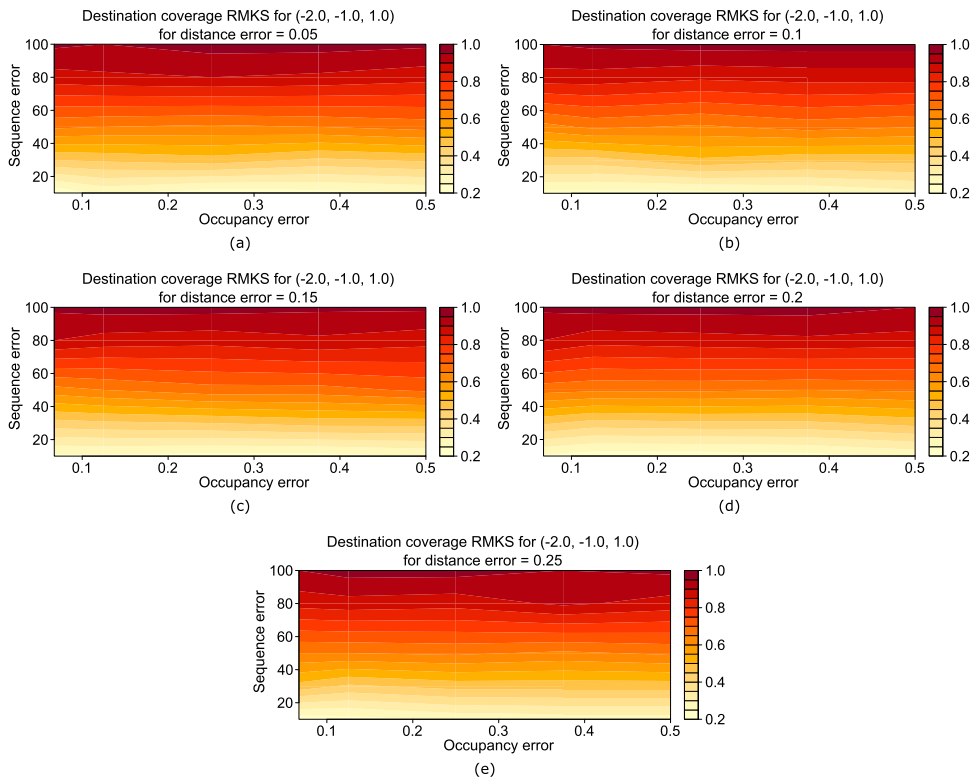




**Figure D4.** Heatmaps of the average DC RMKS for  $(1, -1, 4)$  over five error study replicates against both occupancy and sequence errors at constant distance errors.



**Figure D5.** Heatmaps of the average DC RMKS for  $(-1, -4, 1)$  over five error study replicates against both occupancy and sequence errors at constant distance errors.



**Figure D6.** Heatmaps of the average DC RMKS for  $(-2, -1, 1)$  over five error study replicates against both occupancy and sequence errors at constant distance errors.