LUND UNIVERSITY

Statistical model fitting and model selection in pedestrian dynamics research

Bode, Nikolai; Ronchi, Enrico

Statistical Model Fitting and Model Selection in Pedestrian Dynamics Research

Nikolai W.F. Bode[1] · **Enrico Ronchi**[2]

[1] Department of Engineering Mathematics, University of Bristol, Bristol, UK,
E-mail: nikolai.bode@bristol.ac.uk
[2] Department of Fire Safety Engineering, Lund University, Lund, Sweden,
E-mail: enrico.ronchi@brand.lth.se

Abstract Pedestrian dynamics is concerned with understanding the movement patterns that arise in places where more than one person walks. Relating theoretical models to data is a crucial goal of research in this field. Statistical model fitting and model selection are a suitable approach to this problem and here we review the concepts and literature related to this methodology in the context of pedestrian dynamics. The central tenet of statistical modelling is to describe the relationship between different variables by using probability distributions. Rather than providing a critique of existing methodology or a "how to" guide for such an established research technique, our review aims to highlight broad concepts, different uses, best practices, challenges and opportunities with a focussed view on theoretical models for pedestrian behaviour. This contribution is aimed at researchers in pedestrian dynamics who want to carefully analyse data, relate a theoretical model to data, or compare the relative quality of several theoretical models. The survey of the literature we present provides many methodological starting points and we suggest that the particular challenges to statistical modelling in pedestrian dynamics make this an inherently interesting field of research.

Keywords Pedestrian dynamics · collective behaviour · statistical modelling · statistics

1 Introduction

One of the transformative developments in the study of pedestrian dynamics has been the development and application of mathematical and computational models. Research into pedestrian dynamics is inherently multidisciplinary and concerned with understanding the movement patterns that arise in places where more than one person walks. Understanding
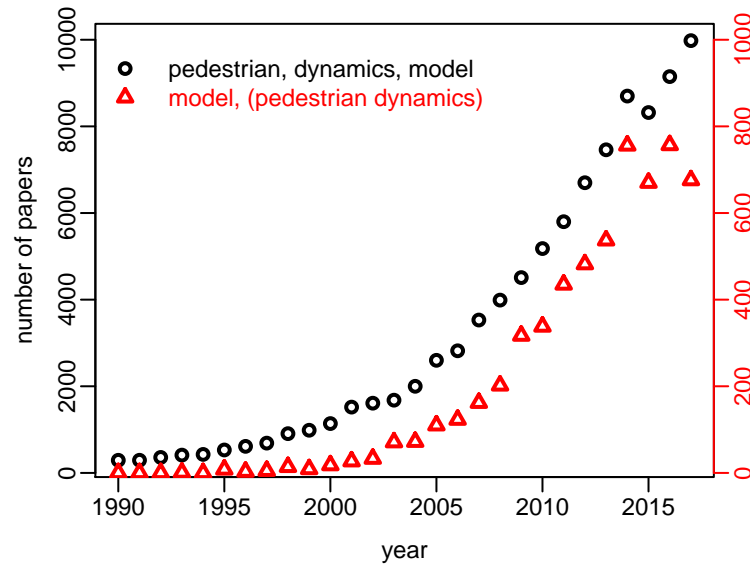
Figure 1 Number of annual publications on pedestrian dynamics models, estimated by performing an online search with the search engine Google Scholar using the search terms indicated in the figure legend (https://scholar.google.co.uk/; accessed 26 October 2018). Words in brackets indicate that the entire expression has to appear in the publication.

how individual walking behaviour leads to larger-scale dynamics is not only an interesting fundamental research question, but also has direct applications in building design, event planning and traffic management, for example. It is therefore not surprising that when theoretical models were developed to explain and possibly even predict pedestrian behaviour at an individual or aggregated level, they quickly became very successful and popular tools in research and industry [1–3]. Estimates of the number of publications on pedestrian dynamics models can serve as an indication for the extent of this success (see Fig. 1; an alternative perspective on the literature shows similar qualitative trends [4]).

An additional and equally transformative development in the field of pedestrian dynamics has been the increase in observational and experimental data that is being recorded in research [5] and industry (unfortunately most data in industry is not published at present). In analogy to other quantitative research domains, the increased use of theoretical models and data in pedestrian dynamics research raises two fundamental questions that we focus on here:

Question 1: *How and to what extent is it possible to infer the behavioural mechanisms underlying pedestrian dynamics from data?*

Question 2: *There are many theoretical models for pedestrian dynamics. How can we rigorously compare different models to decide which ones are most appropriate in different scenarios?*

Statistical modelling, the subject of this review, is one available methodological framework for approaching these questions. Statistics is a well-established discipline and many textbooks, literature reviews and journals are available to inform the use of statistical modelling across research fields. So why is there a need for a review on statistical model

fitting and model selection specifically for pedestrian dynamics? We decided to write this review for two main reasons. First, we suggest there are many opportunities for an increased use of carefully applied statistics in pedestrian dynamics research and with this review we hope to highlight some of these opportunities. By working towards addressing the two questions above, statistical modelling can help to consolidate the experimental and theoretical advances in the field. Second, in our opinion pedestrian dynamics is an intrinsically interesting application domain for statistical modelling. The fact that the observed behaviour of pedestrians often arises from interactions between individuals, such as avoiding collisions or attempting to stay close to friends, presents interesting challenges for statistical analysis and modelling that we will discuss below.

The purpose of this review is neither to present an exhaustive list and critique of statistical approaches, nor is it a detailed "how to" guide for statistical modelling in pedestrian dynamics. We attempt to cover the relevant literature but we do not claim to have found all relevant work. Simple descriptive methods and measures, some of them rooted in statistics, are frequently used in pedestrian dynamics research. We deliberately exclude these from our discussion to focus on reviewing key statistical modelling techniques adopted in the field and possible pitfalls in their use. We present our view on this topic, founded in the literature, and hope this is a useful starting point for an increased use of statistical modelling in pedestrian dynamics research.

The remainder of this review is structured into seven parts. In the four following sections, sections 2-5, we set the scene by introducing relevant methodological background in general terms. This includes a discussion of statistical model fitting (section 2) and model selection (section 3) in the context of pedestrian dynamics and a categorisation of the possible uses of this methodological framework (section 4). In the last section of the background we outline typical steps in statistical modelling in an attempt to start a discussion on best practices (section 5). Section 6 contains a survey of previous work in statistical model fitting and model selection in pedestrian dynamics. We hope that this section is a useful reference point for researchers wishing to develop their own statistical models. Before summarising and discussing the current state of the art in section 8, we highlight several common pitfalls in statistical modelling in section 7.

2 Statistical model fitting and pedestrian dynamics

An informal description of statistical models is that they describe the relationship between different variables, typically measured from data, by using probability distributions. This use of probability distributions is what distinguishes statistical models from other mathematical models. It means that we can use statistical models to make probabilistic statements about the processes or data they describe, expressing our certainty for outcomes. When formulating statistical models, we make assumptions related to the probability distributions inherent in them. Similarly, the process of fitting statistical models to data, or deciding on values for model parameters given data, is informed by the probabilistic nature of the models and requires assumptions on the statistical properties of the data used. This conceptual framework is very flexible and there are many different approaches for

fitting statistical models. Before discussing the particular challenges for statistical model fitting in pedestrian dynamics, we will illustrate some of the key concepts in examples. Throughout this review, we will only consider statistical models that go beyond standard statistical tests that assess hypotheses about populations means or distributions, such as the T-test, Chi-squared, Kolmogorov-Smirnov test, Wilcoxon signed-rank test, Fisher's exact test, Bootstrapping to name but a few.

In the first example, we consider the scenario depicted in Fig. 2: pedestrians are queuing in front of a narrow bottleneck and pass through it, one after the other. A key quantity in this context is the time gap, Δt, between consecutive pedestrians passing through the bottleneck. It relates to the flow of pedestrians, but variations in Δt could also tell us about temporary interruptions of the flow that could be caused by competitive behaviour in front of the bottleneck, for example. To study the dynamics of this system, we can develop a statistical model for the quantity Δt [6,7]. Based on the distribution of Δt observed in data, a Gamma distribution is a plausible distribution to use in a statistical model. A statistical model could thus assume that observed time gaps Δt_i, where the index i indicates the time of observation, follow a Gamma distribution with mean μ_i and variance σ and thus take the general form:

$$\Delta t_i \sim \Gamma(\mu_i, \sigma) \tag{1}$$

Importantly, in this formulation, the mean μ_i does not have to be a constant model parameter, but can depend on the relative positions and movement of pedestrians in front of the exit. For example, we could assume that the time gap depends on the distance of the closest person B to the exit, d_B, and on the difference in distance between the two closest pedestrians to the exit, $d_C - d_B$ (see Fig. 2). The expression for μ_i in our model could then be:

$$\mu_i = \beta_0 + \beta_1 d_B + \beta_2 (d_C - d_B), \tag{2}$$

where β_0, β_1 and β_2 are model parameters and d_B and d_C will vary over time. Other model formulations are possible and very useful for studying the scenario in more detail, as discussed below.

One approach of fitting such models makes use of the Likelihood function, $L(\theta)$. $L(\theta)$ is a function of model parameters θ and it describes the plausibility of parameter values for the model to describe observed data, for given data. Parameter values associated with higher values of L are therefore preferred. In Maximum Likelihood Estimation (MLE), statistical models are fitted to data by selecting parameter values that maximise the Likelihood function L. When possible, the parameters that maximise L can be found analytically, but very often this optimisation is performed numerically (examples in pedestrian dynamics are [6, 8–16]).

An alternative approach to Maximum Likelihood Estimation derives from Bayesian statistics and it combines the Likelihood function with prior knowledge about the parameters (prior distribution) to obtain information about likely parameter values (posterior distributions) and model fit (e.g. Marginal Likelihood; [13]).

For the example on time gaps introduced above, if we define $f_\Gamma(\Delta t_i; \mu_i, \sigma)$ to be the probability density function of a Gamma distribution with mean μ_i and variance σ evaluated at Δt_i. Then under some assumptions detailed below the likelihood, L, of the statis-
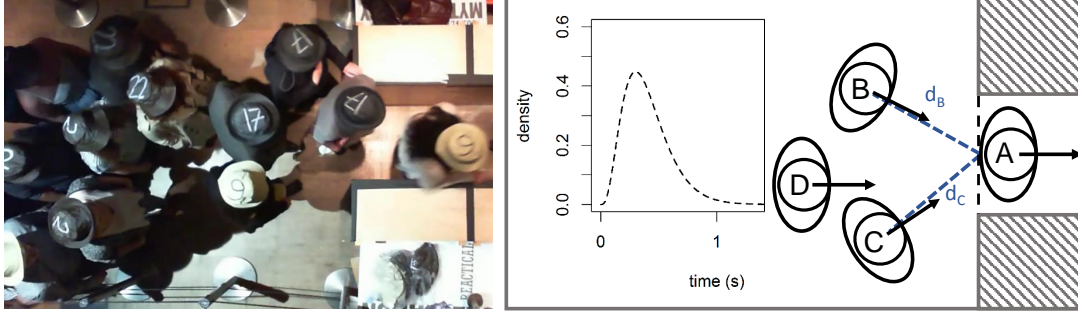
Figure 2 Example for a statistical model used to investigate pedestrian dynamics at bottlenecks [6, 7]. The left-hand panel shows a still image from an experiment with volunteers. The statistical model describes the time gap between consecutive pedestrians that pass through the bottleneck. In the right-hand panel, individual A has just entered the bottleneck. The model assumes that the time gap until the next pedestrian (B, C, or D) enters the bottleneck comes from a Gamma distribution, similar to the one shown, with mean dependent on the dynamics and positioning of pedestrians in front of the bottleneck. For example, a simple model assumes that the mean of the time-gap distribution is the distance of the closest pedestrian to the exit (d_B here).

tical model described above is the product of the probability densities over all observed time gaps:

$$L = \prod_i f_\Gamma(\Delta t_i; \mu_i, \sigma) \tag{3}$$

Since μ_i depends on model parameters, L is a function of the model parameters in Eq. 3.

Importantly, when formulating the Likelihood as in Eq. 3, we make the assumption that the different observed data points are conditionally independent. For our example, this means that we assume our model captures any correlations or dependencies between consecutive or indeed between any Δt_i and Δt_j. This assumption means we can compute the Likelihood as a product over the probability densities associates with each Δt_i. In the context of pedestrian dynamics, this assumption creates particularly interesting challenges that we will discuss further below. Authors concerned that such assumptions underlying statistical models do not hold, sometimes use a pseudo-likelihood approach to fit their models, effectively acknowledging that some of the model assumptions may be violated (e.g. [17]). The Likelihood concept does not rely on conditional independence. In principle, Likelihoods can be formulated differently, such as $L = f(\{\Delta t_i\}_i; \{\mu_i\}_i, \sigma)$, where f is a probability density function that simultaneously accounts for the set of all data points $\{\Delta t_i\}_i$, taking any spatio-temporal dependencies into account. However, in practice it is difficult to formulate Likelihood functions in this way and none of the examples from the pedestrian dynamics literature we discuss below takes this alternative approach to making independence assumptions.

Statistical models can be used to describe the full range of scenarios typical in pedestrian dynamics, ranging from models that describe the probability for individuals to select one from a set of discrete options, such as exit doors (Fig. 3(A)), to models that describe the movement path of individuals using multivariate probability distributions to express probabilities for positions in two or more dimensions (Fig. 3(B)). While the former sce-
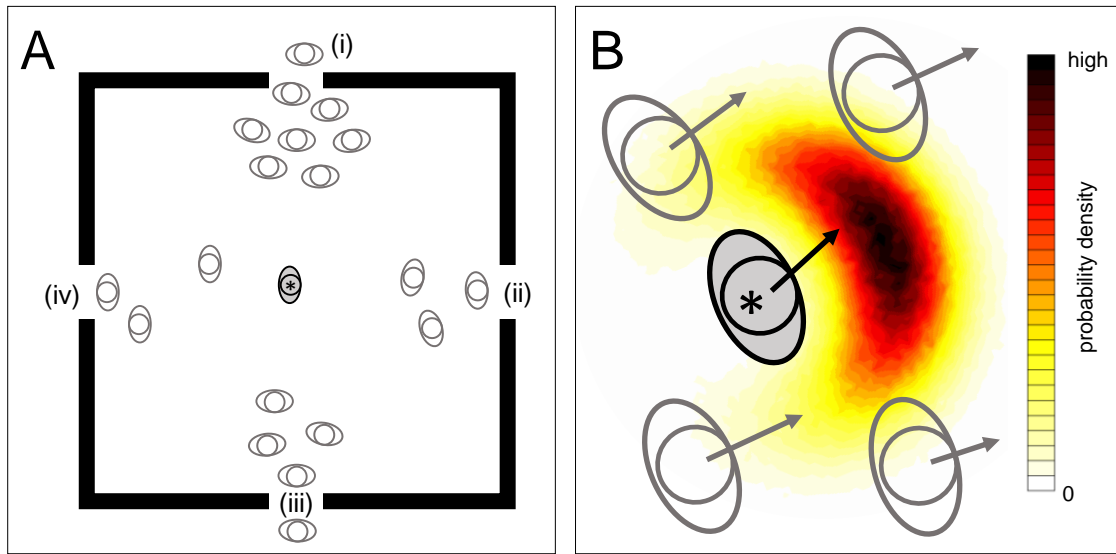
Figure 3 Examples for statistical modelling scenarios typical for pedestrian dynamics. In panel A, a statistical model is used to describe the exit choice of individuals. Consider the pedestrian highlighted in grey and with an asterisk, *. For example, a statistical model could express the probability of this individual to choose one of the exits, labelled (i)-(iv), based on how far away or busy they are (e.g. via Multinomial logistic regression). In panel B, a statistical model describes the location a pedestrian moves to within a fixed time interval. For example, the model may predict a probability distribution, shown in colour, for the location of a focal individual (highlighted in grey and with an asterisk), based on the location of the individual and other pedestrians nearby. Arrows indicate the actual pedestrian movement that may be observed.

nario can be investigated using statistical models with clearly defined likelihood functions (e.g. [10–12, 14, 15, 18]), models for the latter scenario often have to make simplifying assumptions to formulate a likelihood function (e.g. [9, 16]) or, alternatively, use techniques that replace an explicit Likelihood function with many simulations of models, also called Likelihood-free methods (e.g. [19]). A list of commonly used statistical model fitting approaches alongside references for their use in pedestrian dynamics can be found in Tab. 1.

At a glance, statistical model fitting, MLE in particular, appears to be similar to other approaches that define an objective function which measures the similarity between data

Statistical model fitting technique	References (examples)
Maximum Likelihood estimation (analytical)	[20, 21]
Maximum Likelihood estimation (numerical)	[6, 8–16]
Pseudo-likelihood (numerical)	[17]
Bayesian analysis (with Likelihood function)	[13]
Bayesian analysis (Likelihood-free)	[19, 22]

Table 1 List of model fitting techniques alongside references for their use in pedestrian dynamics.

and model and calibrate models by determining parameter values that optimise the values of this function (e.g. [23, 24]). An example for such an objective function is the sum of the squared difference between the model predictions and the data for all observations (incidentally, minimising this function is identical to MLE for a class of statistical models assuming Normal error distributions). However, there are two important differences. First, statistical model formulation and fitting requires explicit assumptions about the use of probability distributions to describe the relationship between observed variables. Assessing the validity of these assumptions for fitted models is often possible and thus facilitates checking the appropriateness of models directly (see Sec. 4.3; Likelihood-free methods are an exception). Second, as already mentioned above, the formulation of statistical models facilitates probabilistic statements about data, but also about parameters estimates (e.g. hypothesis tests on parameter values) and comparisons between models (see Sec. 3). In this sense, statistical model fitting is a form of model calibration that offers an established framework for additional analysis and model checking.

As already indicated above, the pedestrian dynamics context presents specific challenges to performing statistical model fitting. The underlying reason for this is that pedestrian dynamics is concerned with the movement dynamics arising over time from the interactions between multiple pedestrians, as well as interactions of pedestrians with the environment. This means that there are both temporal dependencies (e.g. the next step of a pedestrian may depend on previous steps) and spatial dependencies (e.g. the movement of one pedestrian depends on movement of other pedestrians nearby) in most pedestrian dynamics contexts (e.g. [17]). The challenge for statistical modelling in pedestrian dynamics is to formulate models that describe meaningful aspects of the dynamics, but can ideally still be expressed in terms of sensible probability distributions. In addition, the variables modelled and the model description should preferably ensure that it is possible to formulate a Likelihood function which typically requires that conditional independence of modelled data points holds. This can be challenging. For example, consider statistical models describing the movement of multiple pedestrians in two dimensions. This would require formulating joint, multi-variate distributions describing the movement of all individuals over time. Approaches specifying probability distributions separately for individuals would have to ensure all spatial and temporal dependencies are accounted for for each individual.

A further issue arises from the observation that pedestrian behaviour can change over time. This can be seen clearly in simple experiments, such as the one on pedestrians walking through a bottleneck discussed above. At the start of such experiments, pedestrians might rush to get through the bottleneck first, before settling into unhurried queuing behaviour that can turn into deliberately waiting towards the end of experiments to avoid standing in a crowd. Such intrinsic changes in behaviour or changes in behaviour caused by extraneous effects that may not be known or considered in an analysis lead to non-stationary time series of pedestrian behaviour observations. If unaccounted in statistical models, the non-stationarity of time series mean any independence assumptions in Likelihood functions are not valid (see Sec. 4.3 for an example).

One approach to these problem is to use Likelihood-free methods that avoid the problem of having to specify the relationship between variables in terms of probability distri-

butions (e.g. [19,22]). However, such approaches are typically computationally expensive and importantly they do not lend themselves to model checking and hypothesis testing in the same way as other statistical model fitting approaches do [19, 22]. Another approach is to select variables of interest for modelling statistically, such as the decision between discrete options (e.g. exit routes; see [10–12, 14, 15, 18]), or time gaps between pedestrians passing through a bottleneck, as discussed above [6, 7]. However, such approaches do not guarantee that issues of temporal or spatial dependencies are avoided and they inevitably describe a selection of the dynamics. To avoid the issue of non-stationary time series, data from time intervals for which the behaviour or system dynamics are stationary could be selected using stationary state detection methods (e.g. [25]). However, this risks discounting highly informative data. Alternative methods to account for auto-correlations in data exist, such as re-sampling techniques or generalised least squares, but their use limits the applicability of standard statistical techniques for model fitting or inference. Despite different approaches, there is no clear solution to this problem to date, which means pedestrian dynamics remains an interesting domain for formulating and fitting statistical models.

In an approach somewhat different to statistical modelling, the variability in data or model simulations is studied using statistical techniques. Rather than relating models that incorporate distributional assumptions directly to data via statistical model fitting, this approach analyses the variability in empirical data or model simulations separately. For example, this can be used to assess if a deterministic modelling approach that only captures average dynamics without any variability is appropriate given observed variability in data [26]. Other examples for this work assess how many replicate simulations of models that include variability have to be performed to ensure the convergence of average dynamics and variability in dynamics [27–29]. However, as this work is not directly concerned with statistical modelling, we will not discuss it further here.

3 Statistical model selection and pedestrian dynamics

Statistical model selection is the process of selecting a statistical model from a set of candidate models [30, 31]. For example, considering the statistical model for time gaps between pedestrians passing through a bottleneck discussed above, we may wish to formally compare the model specified in Eq. 2 with a much simpler model that assumes time gaps have a constant mean: $\mu_i = \beta_0$. Note that such a model would suggest that time gaps do not depend on the dynamics in front of the bottleneck and it would assume that the entire system is stationary meaning that pedestrians arrive and leave at a constant rate over time – assumptions which are unlikely to hold. Provided theoretical models in pedestrian dynamics can be considered within a statistical modelling framework, statistical model selection is a process that directly addresses the second main question posed in the introduction (how to rigorously compare models). More broadly, it relates to a key aspect of the knowledge generation process using theoretical models. Different theoretical models can be viewed as formalising different hypotheses for the mechanisms underlying a system. Statistical model selection is thus one approach for deciding which one of a number

of competing hypotheses is the most plausible, given observed data.

Statistical model selection is not a single generally applicable and valid methodology. Instead, this process aims to balance two main guiding principles and there is a range of techniques available to formally compare models. The guiding principles for statistical model selection are first, how well a model captures data and second, how simple a model is, typically measured in terms of the number of free model parameters. The first principle requires little explanation, as it is evidently desirable that models explain data well or that the fit of models to data is good. The second principle is based on a concept widely accepted as part of the scientific method known as Occam's razor or the principle of parsimony which states that simpler models or hypotheses are preferable [32, 33]. This includes avoiding overfitting, which occurs when overly complex models are used. An extreme example for overfitting is the case of a model that has as many parameters as there are data points. Such a model captures the data very well, but is meaningless in terms of describing or measuring general trends in the data.

It should be noted that although simpler models are preferred, the other extreme, overly simplistic or reductionist models should also be avoided (see e.g. [34] for a discussion).

Different approaches and corresponding techniques can be used to perform model selection and a list of commonly used methods is included in Tab. 2. We will briefly demonstrate these different approaches with reference to examples for their use in pedestrian dynamics research.

One approach that can be used to exclude models from a set of candidate models is to demonstrate that the assumptions underlying these models, mentioned in Sec. 2, can be demonstrated not to be valid. We will re-visit and demonstrate this concept below using an example from the pedestrian dynamics literature (Sec. 4.3; [6]).

Another approach for model selection is based on examining if different components of models substantially add to explaining the data. In practice, this is achieved by performing hypothesis tests on individual model parameters or groups of parameters. The tests typically compute probabilities for null hypotheses which state that a parameter is equal to zero or that the maximum value of the Likelihood does not improve substantially if parameters are added to a model (see hypothesis tests for single or multiple parameters in Tab. 2). In this way nested models, i.e. a set of models where simpler models are contained within more complex models, can be compared. An example for this is the use of nested statistical models to establish that the presence of simulated social groups is unlikely to have an effect on time gaps between consecutive pedestrians passing through a narrow bottleneck in an experiment [7]. A different study uses a similar conceptual approach to compare nested models for microscopic pedestrian movement and suggests that finite reaction times help to explain walking behaviour [9]. More generally, the approach and technique of testing hypotheses on model parameters is used very widely to establish trends in data, as further discussed in Sec. 4.1.

To compare models that are not nested, more general measures for the relative quality of models can be used. These are often based on the estimated maximal value of the Likelihood and more sophisticated measures explicitly or implicitly penalise models with more parameters (e.g. AIC, BIC, Bayes Factor in Tab. 2). A range of studies employ this approach. For example, a comparison of non-spatial models using Bayes factors

Use	Examples for methods
Hypothesis tests for single parameters	T-test, Likelihood-ratio test
Hypothesis tests for several parameters	F-test, Likelihood-ratio test
Measures for relative quality of models	R^2, Likelihood, AIC, BIC, Bayes Factor

Table 2   List of model selection techniques. The T-test, F-test and R^2 are based on the squared difference between model fit and data, whereas the remaining methods are based on the Likelihood of models. Details can be found in any comprehensive statistics textbook (e.g. [30, 31]). Note that none of these techniques are informative about the correctness of a model (see model checking, Sec. 4.3). Note also that some methods make specific assumptions and are not generally applicable for all statistical models.

suggests when and how crossing pedestrian streams at four-way intersections interact [22] and a model comparison using the Akaike Information Criterion (AIC) suggests that a macroscopic model for multi-directional, time-varying and congested pedestrian flows outperforms simpler models that do not consider anisotropic pedestrian walking speeds [17]. More examples for applications of these model selection approaches in pedestrian dynamics can be found in Sec. 6.

It is worthwhile to briefly contrast statistical model selection to a different model selection approach frequently used in machine learning. In this approach a given data set is divided into a training and a test set that are used to first calibrate models and subsequently assess its goodness of fit (examples in pedestrian dynamics include [35–37]). Repeating this procedure multiple times using cross-validation approaches helps to prevent overfitting. In contrast to statistical model selection techniques, measures of model quality based on this approach do not immediately lend themselves to making probabilistic statements relating to model parameters or the relative quality of models.

This discussion indicates that statistical model fitting is already widely and successfully used in pedestrian dynamics research (see also Sec. 6). A particular feature of pedestrian dynamics research are the close links between industry and research which means that theoretical innovations may be of direct interest for use in real-world applications, such as crowd simulators that are used in fire safety planning [38, 39]. Efforts to systematically contrast theoretical models that could be or are already used in industry are being made (e.g. [38, 39] and beyond academia[1,2]), and we suggest that statistical model selection could usefully contribute to such undertakings.

---

[1]e.g. the registered association RiMEA, see https://rimea.de/, accessed 16/01/2019.

[2]in addition, the ISO (International Standards Organization) has built a Task Group within the Fire Safety Engineering sub-committee (TC92/SC4 technical comitee 92, Sub-Committe 4 https://www.iso.org/committee/50552.html, accessed 08/02/2019) with the scope of developing a standard document for the verification and validation of pedestrian evacuation models used for fire safety engineering applications in buildings

4 Uses for statistical model fitting and selection

Statistical model fitting and selection can be used to inform different aspects of research that relates theoretical models to data. There are three main categories for uses that we will discuss in turn in this section: inference, prediction and model checking.

4.1 Inference

By far the most common use in pedestrian dynamics research of statistical model fitting and selection is inference, the process of inferring information from data and testing assumptions about data. This process directly links with the first main question posed in the introduction. To many authors, the process of inference is synonymous to statistical analysis. However, as we do not include standard statistical tests in this review, the process of inference we refer to here seeks to find models that explain phenomena at a conceptual level by assessing dependencies between observed variables. These models are thus often called explanatory models [40].

Typical examples in pedestrian dynamics research include using statistical models to test which out of a selection of observed factors influence decision-making in pedestrians (e.g. [10–12, 14, 15, 18]), or which out of a set of hypotheses (models) for pedestrian movement is the most plausible given data (e.g. [9, 16, 17, 22]).

Statistical models that are used for inference in pedestrian dynamics are typically derived systematically from anecdotally observed or hypothesised behavioural mechanisms. For example, observations may indicate that pedestrians follow others or that they avoid queues. These observations can be formalised and subsequently tested in statistical models for pedestrian route choice (e.g. [10–12, 15]). Inference using these models aims to ideally uncover causal relationships between variables (e.g. longer queues at an exit lead to a reduced probability of pedestrians choosing it). However, uncovering such causal relationships from data is only possible, if an inferential analysis is combined with a careful experimental design that exposes pedestrians to a representative and relevant range of situations whilst keeping other factors external or internal to pedestrians constant (see also Sec. 5). If this is not given, inferential analysis is still useful, as it can be used to detect and formally test correlations between variables. For example, pedestrians may take queue lengths into account when choosing exits, but their decisions could also be explained by other factors, such as signage, lighting or the movement of people they are related to that may not have been measured and are thus not included in statistical models.

Statistical models that are designed for inference may successfully uncover dependencies between variables in data, but this does not directly imply that they will also be useful for predicting future outcomes [40]. To give an example, a statistical model may help to test a hypothesised relationship between pedestrian route choice and the crowdedness of routes. However, this does not necessarily mean that the model will also be able to successfully predict route choices of individuals. It may have been designed with a strong emphasis on contrasting responses to crowdedness and failed to incorporate other factors influential for individuals' route selection, including internal factors, such as stress levels or personal preferences. This indicates the differences in purpose between statistical

models designed for inference and models designed for another use, namely prediction.

4.2 Prediction

Statistical model fitting and selection can also be used to produce predictions for future behaviour that can be measured. Models developed for this purpose are often called predictive models [40]. In contrast to inference, prediction does not necessarily require or aim at an understanding of the system modelled. Statistical models used for prediction can be derived entirely from data, rather than being informed by theory or hypothesised relationships between variables. Fitting such models and selecting between candidate models emphasises predictive success. In practice, this can mean authors de-emphasise the importance of ensuring that model assumptions hold (e.g. relating to the probability distributions used or conditional independence). In short, when it comes to this use of theoretical models, the distinction between statistical modelling and the wider field of machine learning can become blurred.

We are not aware of examples in pedestrian dynamics research that employ statistical model fitting and selection specifically for the purpose of prediction. Some studies fit and select statistical models and subsequently explore their predictive potential (e.g. [7, 11, 12, 22]). Other studies employ machine learning techniques for prediction and do not view these through a statistical modelling perspective [36, 37, 41, 42]. For example, one study uses multi-agent reinforcement learning to train the behaviour for individual agents and subsequently shows that this behaviour can be scaled up to predict emergent collective behaviour in larger crowds without the need for further training [42]. A different example illustrates the potential of supervised machine learning to learn a simpler model from simulations of a microscopic model that can subsequently be used in an iterative search for optimal solutions for simple pedestrian transport problems, thus demonstrating that the behaviour of the microscopic model is predicted sufficiently accurately [41].

Other models in pedestrian dynamics, such as simulators for the movement of pedestrian crowds, are regularly used to make predictions (e.g. in industry, see [3, 38, 39]). Although some of these models make use of data and model fitting to inform their core functionalities, they are typically not formulated as statistical models in their entirety and the complete models are certainly not calibrated and assessed using statistical model fitting and selection.

The processes of inference and prediction are not mutually exclusive and it is therefore possible and even desirable that a statistical model developed for one purpose can also be useful for the other purpose. In the discussion in Sec. 4.1 and this section, we aim to indicate that it is possible to employ statistical models in pedestrian dynamics for different purposes and that this can lead to differences in the emphasis on aspects in statistical model fitting and model selection that are considered. The difference between statistical analysis for predictive or for inferential (explanatory) purposes has also been discussed more thoroughly elsewhere [40].

4.3 Model checking

An entirely different use of statistical model fitting in pedestrian dynamics is to validate the appropriateness of models. This type of model checking makes use of the fact that statistical model fitting often requires explicit assumptions relating to the probability distribution of data and the conditional independence of data points (see Sec. 2). These assumptions can be tested explicitly and therefore present a convenient approach for model validation.

A flexible way to assess the assumptions underlying statistical models is to consider residuals, measures for the difference between observed values and values estimated by the fitted model. Depending on the distributional assumptions of a model, the distribution of residuals is expected to show particular properties. There are many different ways to compute and to investigate residuals that are firmly established [30, 31]. For example, model assumptions often imply that the residuals follow a particular distribution, such as the Normal distribution, and in this case a Kolmogorov-Smirnov test can be used to assess if this assumption holds. To give a more detailed example, we demonstrate some of the general principles using an example from pedestrian dynamics [6]. Recall the statistical model for time gaps between consecutive pedestrians passing through a narrow bottleneck introduced in Sec. 2. Fig. 4 shows plots of the deviance residuals obtained when fitting such a model to experimental data and to data simulated using a computational model for pedestrian movement (deviance residuals depend on the likelihood of the fitted model [31]; the computational model assumes forces acting between individuals and is based on [43], known as the "social force model"). One popular and often appropriate way of assessing model assumptions is to plot residuals against the values predicted by the model (Fig. 4(A, B)). We would expect that the mean of the residuals is approximately constant and does not depend on the size of the predicted values (as appears to be the case in Fig. 4(A, B)). Trends in residual mean would indicate that the model does not capture important variation in the data. Another residual plot that is particularly relevant for time series data, is a plot of residuals over time. Changes in the mean of residuals over time, could be indicative of auto-correlations in data that are not captured by the model, meaning that the assumption of conditional independence does not hold. Such trends in residuals indicate that models fail to account for non-stationarity in the time series data. In Fig. 4(D) the mean of residuals increases towards the end of simulations. This suggests that the statistical model used is not appropriate for this data. Moreover, as this issue does not occur when fitting the same statistical model to experimental data (Fig. 4(C)), it could even be suggested that the computational model used for the simulations produces different movement dynamics to those observed in the experiment (see [6] for details). This example indicates how residuals can be used to assess model assumptions and therefore learn about the appropriateness of statistical models.

An alternative indicator that can be used in model checking relates to model parameters. Statistical model fitting provides information on model parameter estimates, either relating to errors associated with the estimates, or in Bayesian analysis by producing a posterior probability distribution indicating how likely different parameter values are given data and prior knowledge. This information can be used to assess if models can be
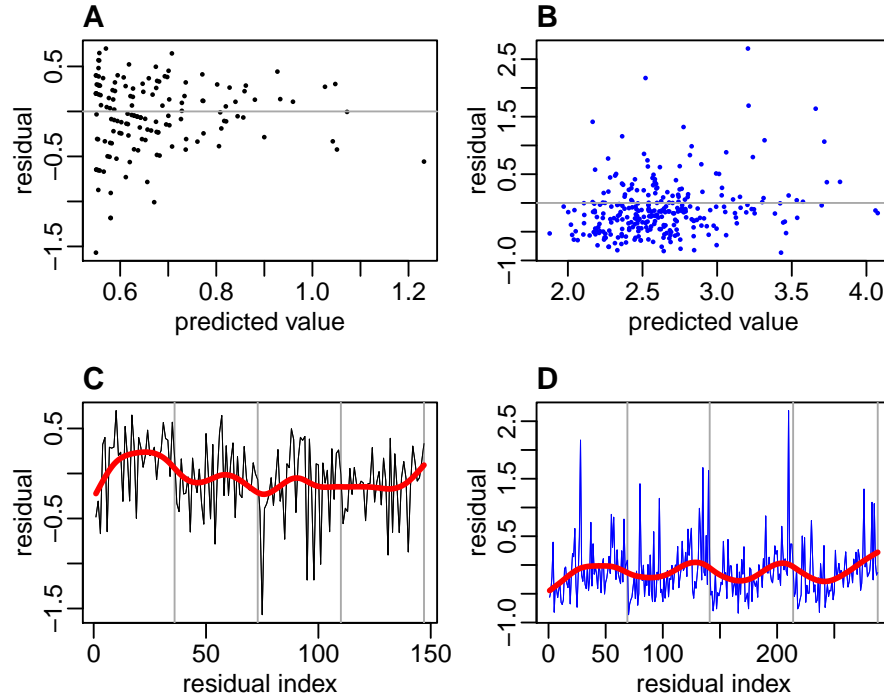
Figure 4   Example for statistical model checking using residuals (redrawn from [6]). A and B show deviance residuals versus predicted values for an experimental data set (A) and a simulated data set (B). C and D show deviance residuals plotted as a time series for experimental and simulated data, respectively (the residual index is incremented over time). Grey vertical lines indicate the start of a new experimental run under the same conditions (C) or the start of a new crowd simulation (D). The red lines in A and B are spline fits to the residual time series that emphasise overall trends visible.

calibrated successfully. For example, posterior parameter distributions can indicate that a large degree of uncertainty about parameters remains after model fitting and thus cast doubt over whether a model is appropriate for a given data set (e.g. a large variance or multi-modality in posterior distributions [19, 22]).

Either of these approaches to model checking are useful for single models, but can also be used in model selection to deselect models that are demonstrably not appropriate.

In summary, in this section we discuss three uses for statistical model fitting: inference, prediction and model checking, providing examples of each in recent literature. Before presenting a more comprehensive survey of relevant work in pedestrian dynamics, we will briefly discuss typical steps in statistical modelling with the intention to start a discussion on best practice for statistical modelling in pedestrian dynamics research.

5 Typical steps in statistical modelling

This section contains a brief summary of typical steps in analysis involving statistical modelling. A list of typical steps can be found in Tab. 3. Depending on the intended use

of statistical models, it may not be necessary to perform all of these steps and the order in which they are performed can also vary.

(i) Data collection. Ideally a statistical modelling analysis already informs data collection. For an inferential use of statistical models, it can be useful to carefully design controlled experiments to ensure the aspects of interest are covered comprehensively and in an unbiased way by the data collected and that enough data is collected to make the desired inferences [30, 31, 40]. In particular, "power analysis" methods allow experimenters to compute required sample sizes relative to expected effect sizes and desired confidence levels [30, 31]. Examples for such experimental design in pedestrian dynamics studies include [13–15, 44, 45]. For predictive modelling, as discussed in Sec. 4.2, it may be more important to focus on collecting a large enough quantity of data, to ensure out-of-sample testing can be performed adequately (e.g. splitting data into training and test sets and performing cross-validation).

(ii) Exploratory analysis of data. An exploratory analysis of raw data is good practice whenever data is analysed. In statistical modelling, this can help to develop hypotheses for dependencies between variables for inferential modelling, it can highlight outliers in the data that may affect subsequent analysis and it can help to identify biases in the data collection (e.g. automated observational data collection only functioned properly during the day, but not at night). These are just a few examples for why an exploratory analysis is useful and more examples can be found in any comprehensive textbook on statistical modelling (e.g. [30, 31]).

(iii) Decide on candidate models. This step is concerned with the formulation of the theoretical models that are to be analysed. In inferential uses of statistical modelling, this may involve deciding on the variables that should be included in a model (e.g. [18]) or deciding on a number of candidate models based on theoretical considerations (e.g. [22]). It could also include fundamental decisions on modelling approaches relating to distributional assumptions or a choice between classical statistical models and computational models with less concrete distributional assumptions. For prediction, different models may be selected because of their known distinct properties or to demonstrate the superiority of a novel model over existing ones (e.g. [35], although this study is closer in spirit to machine learning rather than statistical modelling).

(iv) Model fitting. This process is a crucial part of any statistical modelling analysis and has already been discussed in Sec. 2.

(v) Model selection. The relevance and importance of this step has already been covered above in Sec. 3. Most inferential statistical modelling studies either explicitly or implicitly perform model selection. For example, even if only one model is fitted to data, results of parameter-specific hypothesis tests could be interpreted as indications for how the model could be adapted.

(vi) Check model assumptions hold. This step has been discussed in Sec. 4.3. As mentioned before, it can also form part of the model selection process and if models are to be used purely for prediction, authors may put a lower emphasis on this step. Nevertheless, it is good practice to always perform this step and to report on it.

(vii) Use models for inference/prediction. This step is self-explanatory, but it is worthwhile to highlight the different possible uses of statistical models (inference and

(i)	Data collection (experimental design)
(ii)	Exploratory analysis of data
(iii)	Decide on candidate models
(iv)	Model fitting
(v)	Model selection
(vi)	Check model assumptions hold
(vii)	Use models for inference/prediction
(viii)	Interpret findings
(ix)	Reporting of model and results

Table 3 Summary of typical steps in statistical modelling analysis. Depending on the intended use of a statistical model, it may not be necessary to perform all of these steps and the order in which they are performed can also vary.

predictions) again.

(viii) Interpret findings. The importance of correctly interpreting findings from statistical modelling and model selection has to be stressed. At a basic level, in contrast to deterministic models, the findings of statistical modelling take the variability in data into account and this should be considered. For example, estimates for model parameters should not be viewed as point estimates only, but available information on the uncertainty about the estimates, such as standard errors, should also be considered. The use of statistical models were designed for should also be considered. For example, if a model performs well in model selection for inference, this does not immediately imply that it is also useful for prediction, and vice-versa [40]. Considering inferential uses of statistical modelling, the correct interpretation of the outcomes of hypothesis tests (p-values) is crucial and their misuse is subject to considerable controversy that goes beyond the scope of this review (see e.g. [46–48]).

(ix) Reporting of results. The comprehensive reporting of the outcomes of statistical model fitting and model selection is important and could, in our view, be improved in pedestrian dynamics research. Often researchers, ourselves included, only report on the aspects of statistical models that are of immediate interest to their work. This could be the outcomes of an inferential analysis stating which factors have an effect on the movement decisions of pedestrians, for example. However, it is also important to report on elements of a statistical analysis that may at first glance not be as informative. For example, reporting residual plots, or simply publishing residuals alongside papers could be very useful, as residual plots can highlight shortcomings of models that could be improved on in future work. It is widely understood that at a minimum studies should report in sufficient detail on their analysis so that others could reproduce it. This includes publishing data and stating software packages used.

6 A survey of statistical model fitting and selection in pedestrian dynamics

This section contains a survey of previous work in pedestrian dynamics that uses statistical model fitting. As already mentioned above, the main inclusion criterion for studies is that they make use of statistical modelling that goes beyond standard statistical tests. The resulting literature clusters into three broad categories regarding the application domain of statistical modelling. Studies in the first category have in common that they investigate pedestrian dynamics via meaningful summary statistics that focus on specific aspects of pedestrian behaviour. The two remaining categories are predominantly defined by the aim to relate theoretical models for pedestrian movement to data by calibrating models and by comparing the relative quality of alternative models. We have chosen to distinguish between microscopic and macroscopic models in this general approach, where the former model type describes the system dynamics at the level of individuals and the latter type describes system dynamics aggregated over pedestrians. In the following, we introduce each of these three categories in turn.

6.1 Regression analysis on summary statistics

The most common application of statistical modelling in pedestrian dynamics, perhaps because it is the most amenable to this type of analysis, relates to investigating the relationships between variables that summarise specific aspects of pedestrian behaviour. This type of analysis can be described as regression and Tab. 4 summarises studies that fall into this category.

A context where such variables arise naturally is route choice, where pedestrians have to choose between a number of discrete and mutually exclusive options, such as exit doors. Consequently, many studies develop models for the probability of pedestrians to choose one of two or more exit routes, testing the effect environmental aspects, such as the width or crowdedness of exits, or the influence of psychological aspects, such as proxies for stress or social connections, on the decisions of individuals (e.g. [10–15, 49]). Different contexts also naturally give rise to discrete variables capturing pedestrian behaviour. Examples include the question of whether pedestrians help others or not, depending on how much time they have to invest (e.g. [50]), or the decisions pedestrians make prior to evacuating (e.g. [18]). Alternatively, studies use questionnaires in which participants report their behaviour or perception in pedestrian dynamics contexts via discrete categories [51]. These contexts of predicting the probability of discrete outcomes can be modelled using existing and well-established statistical models, such as logistic regression or logit models and their extensions [30, 31].

Other studies identify continuous variables that can take any value or any value within a given range for investigation. These can arise naturally from the specific context under investigation, such as the time gaps between consecutive pedestrians passing through a narrow bottleneck (cf Sec. 2, [6, 7]). Alternatively, researchers select aspects of crowd dynamics they wish to investigate via surveys, observations or experiments, and devise

measures for these aspects, as well as for factors that may help to explain their value. Examples include modelling changes of the self-reported level of perceived safety [20], risk-taking behaviour [45, 52] or average walking speeds of pedestrians [53–56] in response to factors, such as pedestrian density, age or gender.

Tab. 4 shows that only very few studies explicitly test if the underlying assumptions of the statistical model used are valid. None of the studies found use statistical modelling primarily for the purpose of prediction, as described in Sec. 4.2. Some studies calibrate models and subsequently use them to make predictions, but typically this does not involve a systematic investigation including test and training data (e.g. [7, 12, 49]).

6.2 Calibrating and comparing microscopic models

Microscopic models describe pedestrian dynamics at the level of individuals' movement. Their flexibility and ability to capture emergent effects in pedestrian dynamics has made them a popular tool in research and industry [1–3]. Two key challenges in research concerned with such models are to calibrate them to data and to compare several models to decide which model is most suitable in general or for a given context. Statistical modelling has been used as one possible approach to address these challenges and Tab. 5 summarises relevant studies.

Statistical model fitting for microscopic models has been performed in three ways. First, existing deterministic models have been adapted by adding noise, a random variable from an appropriate or convenient probability distribution, to the deterministic model dynamics (e.g. [9, 57–59]; see Fig. 3(B) for an example). This process converts deterministic predictions into random variables, as required for statistical model fitting. Second, microscopic models are formulated in terms of probabilistic outcomes from the start. For example, cellular automata may describe the probability of pedestrians to move from one cell to neighbouring cells (e.g. [16]). Third, researchers adopt Likelihood-free techniques that only require simulations of microscopic models (e.g. [19]). This is the most flexible approach, but it does have drawbacks, such as being computationally expensive and often requiring the selection of summary statistics extracted from pedestrian trajectories for model fitting.

Typical assumptions in statistical model fitting include the conditional independence of data points, meaning that all spatial and temporal dependencies need to be accounted for (cf Sec. 2). However, microscopic models describe the movement of all individuals over time under the assumption that individuals interact somehow (e.g. by attempting to maintain their personal space). Therefore, the properties of microscopic models mean that the assumptions of statistical model fitting should be subject to careful scrutiny, but in practice this is rarely reported (see Tab. 5).

It should be noted that in addition to statistical model fitting approaches, a wide range of alternative techniques for calibrating and comparing microscopic models for pedestrian dynamics have been developed (e.g. [23, 24, 60–63]). One of the distinct advantages of statistical model fitting and model selection over alternative approaches is that model assumptions have to be made explicit and there is a range of techniques available for testing their validity (cf Sec. 4.3).

6.3 Calibrating and comparing macroscopic models

Macroscopic models describe the movement dynamics of pedestrians at an aggregated level and do not consider individuals separately. To date, few studies on macroscopic models in pedestrian dynamics have made use of statistical model fitting (Tab. 6). One possible explanation for this could be that many macroscopic models are inherently deterministic and this, in combination with typically small numbers of parameters, means that calibration is often performed via simple curve fitting [1]. At first glance, it could be tempting to assume that investigating aggregated dynamics may reduce the issues encountered by microscopic models in statistical model fitting (cf Sec. 6.2). However, this is not the case and researchers resort to adopting pseudo-Likelihood approaches [17] or to using Likelihood-free techniques when closed-form descriptions of Likelihood functions are difficult to formulate or unavailable [22]. While some attempts to investigate model assumptions are being made (e.g. testing distributional assumptions for time intervals between events [22]), in general the underlying assumptions of statistical models remain untested (Tab. 6).

7 Common pitfalls

In this section, we provide a brief summary of some of the common pitfalls in statistical modelling. This collection is neither complete, nor exclusive to pedestrian dynamics research, but we hope it will be useful in promoting best practices. More detail can be found in statistics textbooks (e.g. [30, 31]) and a dedicated literature on the subject (e.g. [48]). We would encourage anyone interested in applying statistical modelling in pedestrian dynamics to carefully consult this literature.

Multicollinearity arises when variables used in statistical models are correlated. This can be data-based (i.e. inherent in the data collected) or structural when new variables for prediction are created by combining existing ones. Multicollinearity can make parameter estimates and inference on parameters unreliable.

Model assumptions are violated. In Sec. 2 and Sec. 4.3 We have already discussed the assumptions that typically underlie statistical model fitting and how to assess if the assumptions are valid. If model assumptions do not hold, inference results using the model, including any hypothesis tests, cannot be relied upon. As also discussed before, if a statistical model is only used for the purpose of prediction, some authors reduce the emphasis on valid model assumptions (cf Sec. 4.2). We would advise to take great care when adopting such a perspective.

Extrapolation beyond the scope of the model. Trends identified in data by a statistical model do not necessarily hold beyond the range of values or scenarios covered by the data used in model fitting. There is a danger that results from statistical modelling are overgeneralised and assumed to hold for populations beyond the ones being studied. This can be highly problematic as such extrapolation may simply be inappropriate. Thus, the boundaries of the applicability for a given model should be clearly stated and conditions for which the findings might be generalise should be discussed carefully.

Brief description	SMF/SMS use	SMF aproach	SMS?	SMF test?	References
Regression analysis testing the effect of several factors on a continuous measure of interest related to crowd dynamics	inference	MLE†	no	no	[20, 21, 44, 45, 52, 53, 64, 65]
Regression analysis testing the effect of several factors on a continuous measure of interest related to crowd dynamics	model checking, inference	MLE	yes	yes	[6, 7]
Regression analysis testing the effect of several factors on a continuous measure of interest related to crowd dynamics	model checking, inference	Bayesian inference	no	no	[54–56]
Analysis of factors influencing decisions in binary choice scenario	inference	MLE	no	no	[14, 15, 45, 50, 66]
Analysis of factors influencing decisions in binary choice scenario	inference	MLE	yes	no	[67, 68]
Analysis of factors influencing decisions in binary choice scenario	inference	Bayesian inference using Annealed Importance Sampling	yes	no	[13]
Analysis of factors influencing decisions in discrete choice scenario (more than two options)	inference	MLE	no	no	[10–12, 18, 49, 51]
Analysis of factors influencing decisions in discrete choice scenario (more than two options)	inference	MLE	yes	no	[69]

Table 4 Survey of previous work on using statistical model fitting (SMF) and statistical model selection (SMS) in pedestrian dynamics research (Regression analysis on summary statistics). †MLE is short for *Maximum Likelihood Estimation* The column labelled "SMF test?" indicates if studies test the validity of some of the statistical model assumptions explicitly.

Brief description	SMF/SMS use	SMF approach	SMS?	SMF test?	References
Calibrating a microscopic model on trajectory data or measures derived from trajectories	inference	MLE[†]	no	no	[58, 59]
Calibrating microscopic models on trajectory data or measures derived from trajectories	inference	MLE	yes	no	[9, 16]
Calibrating microscopic models on trajectory data or measures derived from trajectories	inference	Bayesian inference using simulated annealing	yes	no	[57]
Calibrating microscopic models on trajectory data using summary statistics	inference	Likelihood-free method (Approximate Bayesian Computation)	yes	no	[19]

Table 5   Survey of previous work on using statistical model fitting (SMF) and statistical model selection (SMS) in pedestrian dynamics research (Calibrating and comparing microscopic models). [†]MLE is short for *Maximum Likelihood Estimation* The column labelled "SMF test?" indicates if studies test the validity of some of the statistical model assumptions explicitly.

Brief description	SMF/SMS use	SMF aproach	SMS?	SMF test?	References
Comparison of non-spatial, macroscopic models to study temporal dynamics of crossing pedestrian streams at four-way intersections	inference	Likelihood-free method (Approximate Bayesian Computation)	yes	no	[22]
Calibrating and comparing macroscopic models for pedestrian flows	inference	pseudo MLE†	yes	no	[17]

Table 6 Survey of previous work on using statistical model fitting (SMF) and statistical model selection (SMS) in pedestrian dynamics research (Calibrating and comparing macroscopic models). †MLE is short for *Maximum Likelihood Estimation* The column labelled "SMF test?" indicates if studies test the validity of some of the statistical model assumptions explicitly.

Excluding important predictors. This can lead to models that contain misleading associations between variables. For example, in a hypothetical route choice scenario, individuals may appear to ignore signage if light levels are not accounted for in a model describing individual's decisions. These issues can only be avoided by careful experimental design, data exploration and by considering all available background information on data (cf Sec. 5).

Parameter interpretation. A common misconception is that parameter estimates always measure the effect of one variable onto the variable under investigation independent from other variables. Instead parameter estimates and effects of variables always have to be considered in the context of other variables and effects measured in statistical models.

Another misinterpretation is that significant p-values in hypothesis tests for a parameter indicate a cause-and-effect relationship. Unless all other possible effects are controlled for, e.g. via experimental design, they do not.

A further issue related to parameter interpretation is that parameter estimates are presented without considering the variability underlying the estimates. A typical example for this is the reporting of data and model fits via trend lines that give no indication of the extent to which they capture the variability in the data (e.g. a model fit line is presented without error bars and without a scatter plot of the data the model was fitted to).

Overfitting. We have already discussed this problem in Sec. 3. An extreme case of overfitting is to use as many model parameters as there are data points. In less extreme cases, including too many variables into models makes interpretation difficult.

Sample size. Small data sets can lead to poorly fitted models with large uncertainty associated with parameter estimates. In general, the more data, the more reliable the results that can be inferred using statistical modelling. General quantitative guidelines, such as a number of data points per variable investigated, are not possible, as they depend on the context (e.g. effect size, variability in the data).

While it is very important to avoid such conceptual mistakes in analysis involving statistical modelling, overly strict adherence to model assumptions should not suppress research in pedestrian dynamics, in our opinion. For example, if certain model assumptions cannot be guaranteed to hold, it is still possible to produce insightful research using a pseudo-likelihood model fitting approach (e.g. [17]). However, whenever such an approach is taken, it is crucial that findings from model fitting and selection are not over-interpreted and that the limitations of the analysis are clearly stated and evidenced quantitatively (e.g. via residual plots or by tests on distributional assumptions of models, as in the supplement of [22]). Explicit information on the shortcomings of theoretical models when related to data is likely to be very useful for future work.

8 Discussion

At the start of this review, we state two fundamental questions and claim that statistical modelling can help to address them. The first question relates to the possibility of inferring behavioural mechanisms underlying pedestrian dynamics from data and the second question relates to rigorously comparing the quality of different theoretical models for

pedestrian dynamics. We have discussed the principles of statistical model fitting and model selection (Sec. 2 and 3), demonstrating how statistical model selection is an approach that directly addresses the second question we pose. In outlining three main uses for statistical modelling (Sec. 4), we illustrate how using statistical models for inference presents a rigorous and well-established approach for inferring information from data, as required in the first question we pose (Sec. 4.1). Our very brief outline of the main analysis steps (Sec. 5) and of some of the common mistakes (Sec. 7) will hopefully help to start a discussion on best practices in the use of statistical modelling for addressing our two fundamental questions in pedestrian dynamics. The survey of existing approaches in pedestrian dynamics provides many starting points for authors to identify techniques that most suit their research (Sec. 6). While this provides an overview of the literature, some questions relating to statistical model fitting and model selection in pedestrian dynamics remain.

Why should statistical modelling and model selection be used in pedestrian dynamics? There are alternative approaches to calibrating and comparing theoretical models that have been used in pedestrian dynamics (e.g. [23, 24, 60–63]) and these may produce reliable results. However, there are some features of statistical modelling that distinguishes it from other approaches and that make it an attractive option. First, it is a well-established methodology for which many techniques, textbooks, tutorials and software tools already exist. Second, statistical modelling is very flexible and can be used to address most questions on relating theoretical models to data in pedestrian dynamics. Third, statistical modelling explicitly considers uncertainty in data and models and facilitates probabilistic statements about aspects, such as parameter estimates or relative model quality. Importantly, this also means that the outcome of an analysis using statistical modelling can be inconclusive, if there is not enough data, for example (e.g. probability of one model being better than another is no different from random). Finally, statistical model selection accounts for model complexity, or the number of model parameters, and thus help to prevent overfitting.

Are new methods specifically developed for pedestrian dynamics needed? The survey of studies in Sec. 6 shows that a wide range of problems in pedestrian dynamics can be addressed using existing methods for statistical model fitting and model selection. In our opinion the challenge does not lie in developing methodology for fitting and comparing models, but in developing statistical models for pedestrian dynamics. There is no shortage of theoretical models in pedestrian dynamics (Fig. 1), but most of them are not formulated as statistical models. Some models can be re-formulated as statistical models (e.g. microscopic models, Sec. 6.2), but the difficulty is to develop models that account for the most important spatial and temporal dependencies in data, as discussed in Sec. 2. Alternatively, researchers can focus on identifying measures that comprehensively capture certain aspects of pedestrian dynamics and that facilitate statistical modelling, such as selecting from discrete options in route choice (Sec. 6.1).

Is there a need for one unified framework? The range of problems investigated in pedestrian dynamics is broad, ranging from the route choice of individuals and predicting the next step of pedestrians to predicting large-scale density fluctuations in crowds. Therefore, we suggest that a search for one methodological framework is misguided. Instead,

the approach most appropriate for a given problem should be selected.

Should statistical model fitting and model selection be used more in pedestrian dynamics? This review shows that statistical modelling is already used to great effect in pedestrian dynamics. Nevertheless, we argue that pedestrian dynamics would benefit from an even wider use of statistical modelling. The first reason for doing so is that it requires the explicit mathematical statement of model assumptions, including the dependencies between variables, but importantly also relating to the extent and type of variability in the model. This makes it possible to directly scrutinise model assumptions. The second main reason is that statistical model selection could be used to approach the task of comparing the many different theoretical models for pedestrian dynamics. This may not lead to a generally valid hierarchy of models, but instead may help to identify more or less suited modelling approaches for different contexts.

Perhaps more important than an increased use is the correct use of statistical modelling in pedestrian dynamics. To promote best practices, we have provided a list of typical steps in statistical modelling (Sec. 5) and summarised some of the common pitfalls (Sec. 7). There is also a wide literature on additional issues in statistical analyses of data (e.g. [48]). An important first step is for all researchers to openly engage with and report on the limitations of their statistical modelling approaches.

Why are model checking results under-reported in pedestrian dynamics? The survey in Sec. 6 indicates that only few studies, this authors' work included, report model checking results, such as residual plots or other tests for validating model assumptions. This under-reporting is likely not unique to research in pedestrian dynamics and it does not necessarily imply that researchers do not perform model checking. One possible explanation could be that some model checking analysis does not provide indisputable results. For example, ambiguity could arise from questions on the level of correlation that is permissible when considering auto-correlations in residuals of time series analysis (cf Sec. 4.3) or on the number and severity of outliers in data. Even though these interpretation issues are a natural consequence of applying a mathematical analysis to real-world data, researchers may understandably shy away from trying to explain why their statistical model assumptions are appropriate. An alternative explanation for the under-reporting of model checking results could be that researchers perform model checking, are happy with the results, and therefore do not see the need to report their findings. We would like to make the case for an increased reporting on model checking in pedestrian dynamics. This will inform future work by highlighting specific shortcomings of existing models or particular features in data that could serve as starting points for novel models. While still relevant, this is less important in standard regression analysis, such as the ones discussed in Sec. 6.1, but it is extremely useful when fitting complex microscopic models to data (Sec. 6.2). One example for the use of model checking results is discussed in Sec. 4.3 and [6]. Encouraging the reporting of model checking results will also depend on other researchers being accepting of the inconclusiveness of this analysis.

Why are statistical models predominantly used for inference in pedestrian dynamics? In Sec. 4 we outlined three main uses for statistical modelling in pedestrian dynamics: inference, prediction and model checking. The literature survey in Sec. 6 shows that few, if any, studies use statistical modelling for the purpose of prediction (see also Sec. 4.2).

The reason for this could be that training statistical models for prediction is thought to require a large amount of diverse data [40], which is only starting to become available in pedestrian dynamics research. Therefore, we expect that in the near future there will be an increased emphasis on developing predictive models in pedestrian dynamics. Many of these models may be developed in a machine learning context, but it is important to keep in mind that it is also possible to develop statistical models for the same purpose, or to formulate machine learning techniques as statistical models (e.g. logistic regression for binary classification). As already discussed above, in our opinion a statistical modelling perspective could provide added benefits relating to model checking and model selection.

We hope that the list of pitfalls we present will be useful as a check-list when discussing results of research findings based on statistical modelling in pedestrian dynamics and that it will help researchers to avoid methodological or interpretational issues or at least make them aware of them. We believe this is important to help pedestrian dynamics avoid issues found by meta-analyses in other fields of research [48].

Based on what we have discussed here, we suggest that statistical model fitting and model selection has already contributed substantially to the field of pedestrian dynamics. Considering the need for comparing existing theoretical models and for making sense of the increasingly available data, we propose that an enhanced consideration of the ideas outlined here would be useful. Given the particular challenges to statistical modelling in pedestrian dynamics, we expect that this endeavour will be inherently interesting scientifically and methodologically.

References

[1] Schadschneider, A., Klingsch, W., Klüpfel, H., Kretz, T., Rogsch, C., Seyfried, A.: Evacuation dynamics: Empirical results, modeling and applications. In: Encyclopedia of complexity and systems science, pp. 3142–3176. Springer (2009)

[2] Duives, D.C., Daamen, W., Hoogendoorn, S.P.: State-of-the-art crowd motion simulation models. Transportation research part C: emerging technologies **37**, 193–209 (2013)

[3] Chraibi, M., Tordeux, A., Schadschneider, A., Seyfried, A.: Modelling of pedestrian and evacuation dynamics. Encyclopedia of Complexity and Systems Science pp. 1–22 (2018)

[4] Vermuyten, H., Beliën, J., De Boeck, L., Reniers, G., Wauters, T.: A review of optimisation models for pedestrian evacuation and design problems. Safety science **87**, 167–178 (2016)

[5] Haghani, M., Sarvi, M.: Crowd behaviour and motion: Empirical methods. Transportation research part B: methodological (2017)

[6] Bode, N.W., Codling, E.A.: Statistical models for pedestrian behaviour in front of bottlenecks. In: Traffic and Granular Flow'15, pp. 81–88. Springer (2016)

[7] Bode, N.W.: The effect of social groups and gender on pedestrian behaviour immediately in front of bottlenecks. In: Proceedings of Pedestrian and Evacuation Dynamics 2016, Collective Dynamics, pp. 92–99. Springer (2016). doi:10.17815/CD.2016.11

[8] Bode, N., Holl, S., Mehner, W., Seyfried, A.: Disentangling the impact of social groups on response times and movement dynamics in evacuations. PLoS ONE **10**(3), e0121227 (2015)

[9] Hoogendoorn, S.P., Daamen, W.: Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data. In: Traffic and Granular Flow05, pp. 329–340. Springer (2007)

[10] Haghani, M., Sarvi, M., Shahhoseini, Z.: Accommodating taste heterogeneity and desired substitution pattern in exit choices of pedestrian crowd evacuees using a mixed nested logit model. Journal of choice modelling **16**, 58–68 (2015)

[11] Haghani, M., Sarvi, M.: Stated and revealed exit choices of pedestrian crowd evacuees. Transportation Research Part B: Methodological **95**, 238–259 (2017)

[12] Haghani, M., Sarvi, M.: Following the crowd or avoiding it? empirical investigation of imitative behaviour in emergency escape of human crowds. Animal behaviour **124**, 47–56 (2017)

[13] Bode, N.W., Wagoum, A.U.K., Codling, E.A.: Information use by humans during dynamic route choice in virtual crowd evacuations. Royal Society open science **2**(1), 140410 (2015)

[14] Bode, N.W., Wagoum, A.U.K., Codling, E.A.: Human responses to multiple sources of directional information in virtual crowd evacuations. Journal of The Royal Society Interface **11**(91), 20130904 (2014)

[15] Bode, N.W., Codling, E.A.: Human exit route choice in virtual crowd evacuations. Animal Behaviour **86**(2), 347–358 (2013)

[16] Lovreglio, R., Ronchi, E., Nilsson, D.: Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization. Physica A: Statistical Mechanics and its Applications **438**, 308–320 (2015)

[17] Hänseler, F.S., Lam, W.H., Bierlaire, M., Lederrey, G., Nikolić, M.: A dynamic network loading model for anisotropic and congested pedestrian flows. Transportation Research Part B: Methodological **95**, 149–168 (2017)

[18] Lovreglio, R., Ronchi, E., Nilsson, D.: A model of the decision-making process during pre-evacuation. Fire Safety Journal **78**, 168–179 (2015)

[19] Bode, N.W.: Parameter calibration in crowd simulation models using approximate bayesian computations. In: Proceedings of Pedestrian and Evacuation Dynamics 2018. Springer (in press)

[20] Alnabulsi, H., Drury, J.: Social identification moderates the effect of crowd density on safety at the hajj. Proceedings of the National Academy of Sciences **111**(25), 9091–9096 (2014)

[21] Daamen, W., Hoogendoorn, S.: Emergency door capacity: influence of door width, population composition and stress level. Fire technology **48**(1), 55–71 (2012)

[22] Bode, N.W., Chraibi, M., Holl, S.: The emergence of macroscopic interactions between intersecting pedestrian streams. Transportation Research Part B: Methodological **119**, 197–210 (2019)

[23] Seer, S., Rudloff, C., Matyus, T., Brändle, N.: Validating social force based models with comprehensive real world motion data. Transportation Research Procedia **2**, 724–732 (2014)

[24] Liao, W., Zhang, J., Zheng, X., Zhao, Y.: A generalized validation procedure for pedestrian models. Simulation Modelling Practice and Theory **77**, 20–31 (2017)

[25] Liao, W., Tordeux, A., Seyfried, A., Chraibi, M., Drzycimski, K., Zheng, X., Zhao, Y.: Measuring the steady state of pedestrian flow in bottleneck experiments. Physica A: Statistical Mechanics and its Applications **461**, 248–261 (2016)

[26] Alvear, D., Abreu, O., Cuesta, A., Alonso, V.: A new method for assessing the application of deterministic or stochastic modelling approach in evacuation scenarios. Fire safety journal **65**, 11–18 (2014)

[27] Ronchi, E., Reneke, P.A., Peacock, R.D.: A method for the analysis of behavioural uncertainty in evacuation modelling. Fire Technology **50**(6), 1545–1571 (2014)

[28] Lovreglio, R., Ronchi, E., Borri, D.: The validation of evacuation simulation models through the analysis of behavioural uncertainty. Reliability Engineering & System Safety **131**, 166–174 (2014)

[29] Tavares, R.M., Ronchi, E.: Uncertainties in evacuation modelling: Current flaws and future improvements. In: 6th international symposium: Human Behaviour in Fire, 2015, pp. 185–196. Interscience Communications (2015)

[30] Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning, vol. 1. Springer series in statistics New York, NY, USA: (2001)

[31] Wasserman, L.: All of statistics: a concise course in statistical inference. Springer Science & Business Media (2013)

[32] Gauch, H.G., Gauch Jr, H.G.: Scientific method in practice. Cambridge University Press (2003)

[33] Burnham, K.P., Anderson, D.R.: Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media (2003)

[34] Seitz, M.J., Templeton, A., Drury, J., Köster, G., Philippides, A.: Parsimony versus reductionism: how can crowd psychology be introduced into computer simulation? Review of General Psychology **21**(1), 95 (2017)

[35] Wang, H., Ondřej, J., OSullivan, C.: Trending paths: a new semantic-level metric for comparing simulated and real crowd data. IEEE transactions on visualization and computer graphics **23**(5), 1454–1464 (2017)

[36] Gödel, M., Köster, G., Lehmberg, D., Gruber, M., Kneidl, A., Sesser, F.: Can we learn where people go? In: Proceedings of Pedestrian and Evacuation Dynamics 2018. Springer (in press)

[37] Tordeux, A., Chraibi, M., Seyfried, A., Schadschneider, A.: Prediction of pedestrian speed with artificial neural networks. arXiv preprint arXiv:1801.09782 (2018)

[38] Ronchi, E., Colonna, P., Capote, J., Alvear, D., Berloco, N., Cuesta, A.: The evaluation of different evacuation models for assessing road tunnel safety analysis. Tunnelling and Underground Space Technology **30**, 74–84 (2012)

[39] Ronchi, E.: Testing the predictive capabilities of evacuation models for tunnel fire safety analysis. Safety science **59**, 141–153 (2013)

[40] Shmueli, G., et al.: To explain or to predict? Statistical science **25**(3), 289–310 (2010)

[41] Crociani, L., Lämmel, G., Vizzari, G., Bandini, S.: Learning observables of a multiscale simulation system of urban traffic. In: Proceedings of the Tenth International Workshop on Agents in Traffic and Transportation (ATT 2018). CEUR Workshop Proceedings (CEUR-WS.org) (2018)

[42] Martinez-Gil, F., Lozano, M., Fernandez, F.: Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models. Simulation Modelling Practice and Theory **74**, 117–133 (2017)

[43] Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. Nature **407**(6803), 487–490 (2000)

[44] Drury, J., Cocking, C., Reicher, S., Burton, A., Schofield, D., Hardwick, A., Graham, D., Langston, P.: Cooperation versus competition in a mass emergency evacuation: a new laboratory simulation and a new theoretical model. Behavior research methods **41**(3), 957–970 (2009)

[45] Bode, N.W., Codling, E.A.: Exploring determinants of pre-movement delays in a virtual crowd evacuation experiment. Fire Technology pp. 1–21 (2018)

[46] Sterne, J.A., Smith, G.D.: Sifting the evidencewhat's wrong with significance tests? Physical Therapy **81**(8), 1464–1469 (2001)

[47] Wasserstein, R.L., Lazar, N.A., et al.: The asas statement on p-values: context, process, and purpose. The American Statistician **70**(2), 129–133 (2016)

[48] Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., Ioannidis, J.P.: A manifesto for reproducible science. Nature Human Behaviour **1**(1), 0021 (2017)

[49] Haghani, M., Sarvi, M.: How perception of peer behaviour influences escape decision making: The role of individual differences. Journal of Environmental Psychology **51**, 141–157 (2017)

[50] Bode, N.W., Miller, J., OGorman, R., Codling, E.A.: Increased costs reduce reciprocal helping behaviour of humans in a virtual evacuation experiment. Scientific reports **5**, 15896 (2015)

[51] Ronchi, E., Nilsson, D., Kojić, S., Eriksson, J., Lovreglio, R., Modig, H., Walter, A.L.: A virtual reality experiment on flashing lights at emergency exit portals for road tunnel evacuation. Fire technology **52**(3), 623–647 (2016)

[52] Granié, M.A.: Effects of gender, sex-stereotype conformity, age and internalization on risk-taking among adolescent pedestrians. Safety science **47**(9), 1277–1283 (2009)

[53] Nikolić, M., Bierlaire, M., Farooq, B., de Lapparent, M.: Probabilistic speed–density relationship for pedestrian traffic. Transportation Research Part B: Methodological **89**, 58–81 (2016)

[54] Xie, S., Wong, S., Lam, W.H., Chen, A.: Development of a bidirectional pedestrian stream model with an oblique intersecting angle. Journal of transportation Engineering **139**(7), 678–685 (2013)

[55] Xie, S., Wong, S.: A bayesian inference approach to the development of a multi-directional pedestrian stream model. Transportmetrica A: transport science **11**(1), 61–73 (2015)

[56] Xie, S., Wong, S., Lam, W.H.: A bayesian modeling approach to bi-directional pedestrian flows in carnival events. Frontiers of Engineering Management **4**(4), 483–489 (2017)

[57] Corbetta, A., Muntean, A., Vafayi, K.: Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. Mathematical biosciences and engineering **12**(2), 337–356 (2015)

[58] Ko, M., Kim, T., Sohn, K.: Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. Transportation **40**(1), 91–107 (2013)

[59] Zeng, W., Chen, P., Yu, G., Wang, Y.: Specification and calibration of a microscopic model for pedestrian dynamic simulation at signalized intersections: A hybrid approach. Transportation Research Part C: Emerging Technologies **80**, 37–70 (2017)

[60] Johansson, A., Helbing, D., Shukla, P.K.: Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. Advances in complex systems **10**(supp02), 271–288 (2007)

[61] Wolinski, D., J. Guy, S., Olivier, A.H., Lin, M., Manocha, D., Pettré, J.: Parameter estimation and comparative evaluation of crowd simulations. In: Computer Graphics Forum, vol. 33, pp. 303–312. Wiley Online Library (2014)

[62] Anvari, B., Bell, M.G., Sivakumar, A., Ochieng, W.Y.: Modelling shared space users via rule-based social force model. Transportation Research Part C: Emerging Technologies **51**, 83–103 (2015)

[63] Chraibi, M., Ensslen, T., Gottschalk, H., Saadi, M., Seyfried, A.: Assessment of models for pedestrian dynamics with functional principal component analysis. Physica A: Statistical Mechanics and its Applications **451**, 475–489 (2016)

[64] Kwee-Meier, S.T., Mertens, A., Schlick, C.M.: Age-related differences in decision-making for digital escape route signage under strenuous emergency conditions of tilted passenger ships. Applied ergonomics **59**, 264–273 (2017)

[65] Tang, C.H., Wu, W.T., Lin, C.Y.: Using virtual reality to determine how emergency signs facilitate way-finding. Applied ergonomics **40**(4), 722–730 (2009)

[66] Kinateder, M., Comunale, B., Warren, W.H.: Exit choice in an emergency evacuation scenario is influenced by exit familiarity and neighbor behavior. Safety science **106**, 170–175 (2018)

[67] Wagoum, A.K., Tordeux, A., Liao, W.: Understanding human queuing behaviour at exits: an empirical study. Royal Society open science **4**(1), 160896 (2017)

[68] Zhuang, Y., Schadschneider, A., Cheng, H., Yang, L.: Estimating escalator vs stairs choice behavior in the presence of entry railing: A field study. KSCE Journal of Civil Engineering **22**(12), 5203–5214 (2018)

[69] Daamen, W., Bovy, P.H., Hoogendoorn, S.P., Van de Reijt, A.: Passenger route choice concerning level changes in railway stations. In: Transportation Research Board Annual Meeting, vol. 1930, pp. 12–20 (2005)