



# Considerations about the quality assessment of travel time and travel distance distributions in transport modelling: a proposal for a standardized methodology

Eric Pestel<sup>1</sup>

Published online: 10 March 2020  
© The Author(s) 2020

## Abstract

In travel demand modelling, trip distance distributions or trip time distributions are used to evaluate how well a model fits with observed sample data. Therefore, the comparison of distributions is an essential part in the model validation process. Despite its importance, the common modelling guidelines from the UK, the USA or Austria provide little information about the correct structure and handling of such distributions. Likewise, common statistical methods are not practicable for the validation of transport models. This lack of rules leads to individual solutions, which complicate a model validation and the comparison of models. For example, when comparing two distributions the quality indicator strongly depends on the number of classes. Therefore, guidelines for model validation need to suggest an appropriate way to determine the number of classes. The paper suggests a method for evaluating trip distance distributions and trip time distributions within the model validation process of a travel demand model. It proposes (a) indicators for a classification which consider mode-specific trip distances and trip times (b) a generic classification method based on an equiquantile class width, quality indicators for comparing two distributions and (c) to use relative frequencies instead of absolute frequencies for the calculation of the quality indicators.

**Keywords** Quality assessment · Travel demand modelling · Trip distance distributions · Trip time distributions · Frequency distributions · Key performance indicators

## Motivation

Within the process of building a travel demand model calibration and validation are two iterative and repetitive steps. Calibration describes the process of modifying the model, refinement of specifications, correction of (previously undiscovered) erroneous input data

---

Originally presented at hEART Conference 2018 in Athens (Greece).

---

✉ Eric Pestel  
eric.pestel@isv.uni-stuttgart.de

<sup>1</sup> Chair for Transport Planning and Traffic Engineering, Institute for Road and Transport Science, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany

or adjustment of (non-empirically verified) parameters. The validation on the other hand describes the process of testing if the model has reached a certain goodness-of-fit. This goodness-of-fit does not refer to a particular key performance indicator. Ideally, before the model building process starts, the tendering institution determines how accurately the model should reproduce e.g. number of trips, traffic volumes, mean trip distances and mean trip times for each mode and trip purpose, keeping in mind the availability of reference data and its quality. If the validation process concludes that the model quality is not sufficient, adjustments are made to the model in the step of calibration.

Frequency distributions, which are another type of key performance indicators, describe travel behaviour using specific indicators in discrete classes. There are two common cases: frequency distributions of (1) the time travelled, which are also known as trip time distributions or trip time frequency distributions, and (2) the distance travelled commonly known as trip distance distributions or trip length frequency distributions.

Among others, those distributions are an important tool to evaluate how well a model fits with observed sample data. Therefore, the comparison of distributions is an essential part in the model validation process. Despite its importance the common modelling guidelines from the UK (WEBTAG (Department for Transport 2014)), the USA (TRAVEL MODEL VALIDATION AND REASONABLENESS CHECKING MANUAL (Cambridge Systematics Inc. 2014)) or Austria (draft of QUALIVERMO (Sammer et al. 2012)) provide little information about the correct structure and handling of such distributions. Likewise, common statistical methods that check whether two samples derive from a common population are not practicable for application during the validation of transport models, as will be explained later. This lack of rules leads to individual solutions, which complicate the model validation process, the comparison of models and the definition of thresholds for quality indicators in guidelines.

The following Fig. 1 is intended to illustrate this issue. The same data set (number of trips with their corresponding trip length from a household travel survey and from a travel demand model) is classified into discrete classes in three different ways: the left figure is classified into 20 equidistant classes (1.5 km each), the middle figure into 10 equidistant classes (3 km each) and the right figure into 5 equidistant classes (6 km each). The overlap of the observed and modelled distribution, which is measured by the Coincidence Ratio (CR, which will be explained in more detail later), shows that the lower the number of classes, the less overlap there is. In other words, in an extreme case with only one or two classes, there would be a very high overlap. In this example, it would mean that the classes would only have to be large enough to reach the threshold of  $CR \geq 7$  proposed in the TRAVEL MODEL VALIDATION AND REASONABLENESS CHECKING MANUAL

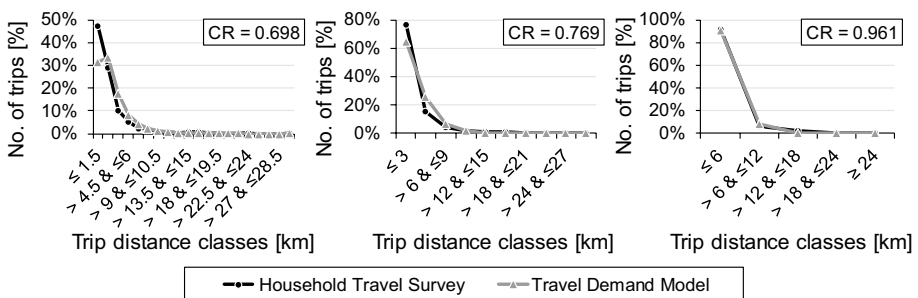


Fig. 1 Classifying the same data set with different classification rules leads to different valuation results

(Cambridge Systematics Inc. 2014). Having said this, the modelled distribution in the left figure would probably be rejected, while the others would be acceptable, although the underlying data set is the same in all three figures.

In general, this means that the resulting distribution evaluation is meaningless if there are no rules for creating the distribution, because even a poor model can be twisted so that the results look fine. Thus the overall goodness-of-fit is at risk. Therefore, guidelines for model validation need to suggest an appropriate way of how to build the distributions.

This is one of the reasons for many misunderstandings between the model users (tendering institutions) and the model developers (contractors) in the context of model development. When validating a model with distributions of time and distance the following questions inevitably appear and they should be answered beforehand by the tendering institutions:

- What indicator should be used for the classification?
- Which area is covered by the distribution?
- What is the reference for a comparison of distributions?
- How many classes should be distinguished? What is the appropriate size of each class? Should the class size increase with the distance travelled? How to proceed with empty classes?
- What quality indicator should be used when comparing two distributions? Should absolute or relative frequencies or both be evaluated?

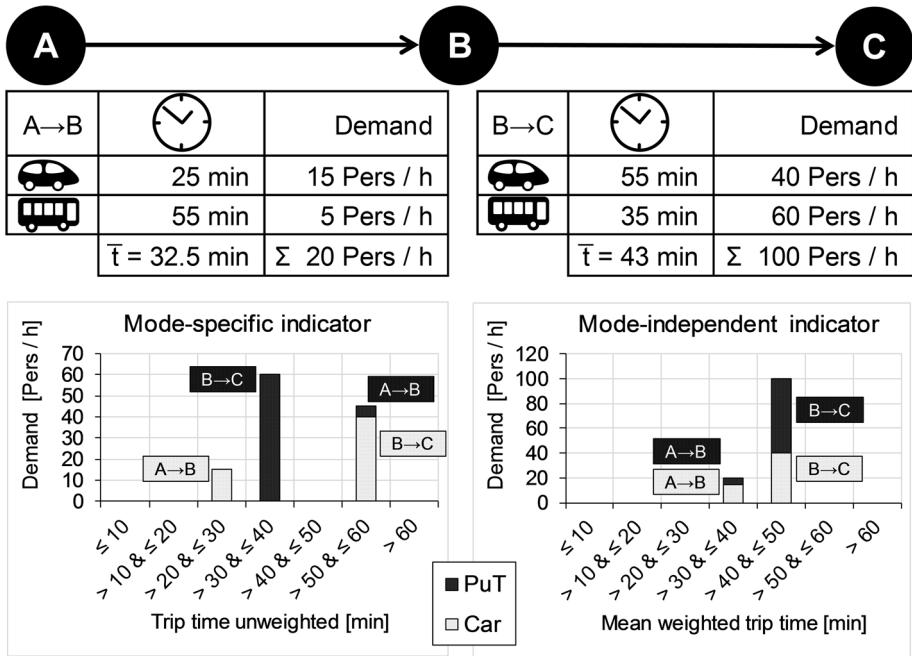
The questions above are also related to the issue that the classification often depends on the dimensions of the study area. This is another reason for individual and not comparable solutions. In order to provide general guidance on how to handle frequency distributions within the model validation process this paper proposes a standardized classification method that can be applied independent from the indicator used and from the dimensions of the study area. For the data set in the example in Fig. 1, this would lead to a single generic classification. It is therefore a necessary complement to the existing guidelines in order to make model validation results more evaluable and comparable.

## Selection of indicators for a classification

In a trip distance or a trip time distribution, the trips are assigned to discrete classes. This requires the number of trips as well as distance or time values on the level of origin–destination-pairs (OD-pairs).

Figure 2 shows a simple example with two OD-pairs (from A to B and from B to C) and two modes (car and public transport - PuT). The modes require different amounts of time to cover the distance between the OD-pairs. This can lead to cases where the trips of one OD-pair fall in different classes for different modes. Such cases occur frequently for trip time, but they can also occur for trip distance, as public transport trips are often longer than car trips. In the example of Fig. 2 the lower left chart shows that the travel demand for each OD-pair is assigned to different trip time classes, because a mode-specific indicator was used for classification.

This effect is difficult to understand in more complex multi-modal trip distributions. In order to avoid this effect entirely and for comparisons on OD-pair level (e.g. a



**Fig. 2** The classification by a mode-specific indicator causes the demand of one OD-pair to be assigned to different classes (left chart), whereas with a mode-independent indicator the demand of one OD-pair is assigned to the same class (right chart)

distance-dependent modal share), the travel demand of all modes on one OD-pair should fall into the same class.

To achieve this, a mode-independent indicator should be used for the classification. In case of trip distance distributions, the direct distance between an OD-pair can serve as a “natural” mode-independent indicator. For trip time distributions, such a “natural” mode-independent indicator does not exist. An indicator weighted with the number of trips provides a solution for this.

In the example of Fig. 2, the lower right chart displays a trip time distribution, which uses a mode-independent indicator for classification. As a result, both car and PuT travel demand fall in the same trip time class for each OD-pair.

Table 1 summarizes the options for selecting a frequency distribution indicator.

A standardized classification requires rules for handling intrazonal trips. In the case of a macroscopic travel demand matrix the main diagonal represents the intrazonal trips. Since this part of the demand is not assigned to the network model, indicators

**Table 1** Classification indicators for trip distance and trip time distributions

	Trip distance	Trip time
Mode-specific indicator	Mode-specific trip distance	Mode-specific trip-time
Mode-independent indicator	Direct distance or mean weighted trip distance	Mean weighted trip time

like trip time or trip distance cannot be calculated, but must be estimated in such a way that they represent an average movement within a zone. Those mean indicators will never fit to an observed intrazonal trip. Therefore, when using distribution comparisons for model validation it is advisable to exclude the intrazonal trips.

Nevertheless, as Bhatta and Larsen (2011) point out omitting intrazonal trips leads to different results in model estimation. Therefore, the model developers are obligated to check the intrazonal trips in a separate analysis. Such an analysis should investigate the number and the modal split of intrazonal trips. A separate examination for different zone sizes may also be reasonable.

## Selection of the study area and a reference distribution

A travel demand model computes demand for a defined study area. The model is calibrated with observed data from a household travel survey. This survey ideally covers a sample of the population from the entire study area. If the study area of the survey and the study area of the travel demand model do not match completely, the comparison of observed and modelled values should only include trips starting and ending in the common study area. This also means that observed trips leaving the common study area have to be excluded from the survey when comparing the observed and the modelled distributions. A comparison of two distributions is only meaningful if observed and modelled trips relate to the same area. A separate analysis is recommended for modelled trips where no comparison data is available.

An additional point to consider is that only home-based trips can be used for validation. For example, modelled trips from home to work can be compared directly with survey data, while modelled trips from shopping to leisure are not necessarily made by residents of the study area and therefore not comparable to a survey of the study area.

Calibrating the destination choice of a travel demand model requires observed distance or time distributions from a household travel survey as a reference distribution. Since interviewed persons tend to round estimates of distance and time for their reported trips, reported distance and time values should not be used directly from the survey. Instead, the distance and time values should be computed with the model using the geocoded origins and destinations from the survey (FGSV 2012).

As Sammer et al. (2018) point out, it should be noted that a household travel survey always contains a systematic error, which results from underreporting of travel behaviour. Hence, even a complete fit of observed and modelled distribution does not mean, that a realistic behaviour is modelled. Sammer et al. (2018) give advice how to increase the quality of a household travel survey.

The purpose of a model is to analyse impacts of scenarios compared to the impacts of a base case. For this, planners look at key performance indicators. Key performance indicators can be single values (e.g. volumes at certain locations, total distance travelled and total time spend) and distributions of travel time and distance. For comparing distributions of a base case and scenarios, distributions derived from the base case of the travel demand model replace observed distributions from the household travel survey as reference distribution.

## Selection of a classification method

### Class width versus class size

Usually distribution classes are determined in such a way that the class width and a number of classes are predefined. Based on the classification indicator the demand is then assigned to the resulting classes. If the class width is the same across all distribution classes (and the number of classes is infinite in an extreme case), the term “equidistant” distribution is used. However, in many cases insufficiently populated classes are aggregated, resulting in classes of varying width.

A different approach defines the share of demand in each class and uses it to determine the class width. The resulting classes vary in width but they are equally populated. This method of classifying distributions is called “equiquantile” (based on e.g. Paluš (1995)).

In the case of an equidistant classification, three questions need to be addressed:

- What is the size of the classes?
- How many classes should there be in total?
- Should classes be grouped together and, if so, how?

On the other hand, the method of equiquantile distribution only raises the question about the number of classes. For this reason, the equiquantile classification method is presented first. Based on this, a procedure is explained to answer the question about the class width of equidistant distributions.

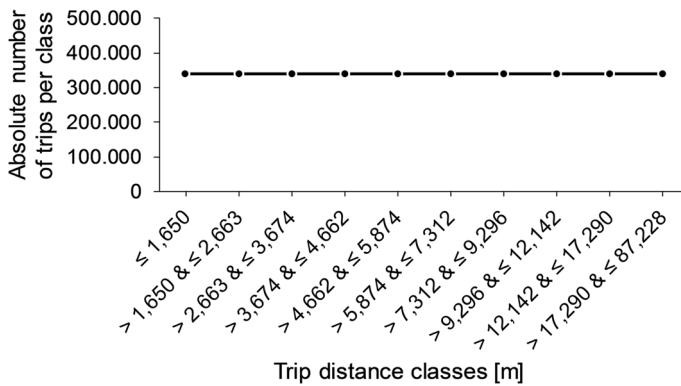
### Analysis with equiquantile classes

The class boundaries of this classification method are calculated using weighted quantiles, with the elements of the indicator matrix representing the classification variable and the demand matrix elements representing the weight. The calculation used for the classification and an example calculation are shown in the appendix.

Although all classes should have the same quantity, the actual demand per class may differ slightly from the desired quantity. This may be due to the following reasons:

- A discrete demand cannot always be distributed completely and evenly across all classes. This is the case, when  $(demand) \bmod (number\ of\ classes) \neq 0$ . As a result, at least one class has a greater demand than other classes. The smaller the population, the greater the deviation from theoretically equal classes. However, this effect is reduced as the sample size increases.
- A single OD-pair with a large demand can also lead to a distortion, for example when  $demand_{od} > demand_{total}/number\ of\ classes$ .
- A rounding of class boundaries prior to classification can lead to shifts in demand between neighbouring classes.

As mentioned above, the number of classes must be defined for the equiquantile classification. Dividing the demand into 10% steps, i.e. into ten classes, is one possible pragmatic assumption. Figure 3 shows a sample of a trip distance distribution. It is shown that each



**Fig. 3** Example of an equiquantile trip distance distribution

class has approximately the same number of trips and the resulting graph approximates a straight line.

Equiquantile classes can be applied for a particular demand segment, for the total demand and for the person distance travelled. Those applications are described in detail in the appendix.

### Analysis with equidistant classes

As previously mentioned, the number of classes and the class width must be predefined for the equidistant classification method. A common procedure to determine those parameters is Sturges' rule for constructing histograms, which is discussed e.g. by Hyndman (1995).

Another way of determining the class width is to define the smallest class of the *equi-quantile distributed total demand* as the equidistant class width. This class width represents the smallest class with 10% of the total demand. In a strictly equidistant distribution, the number of classes is not important, i.e. classes of equal width are created until the largest classification indicator is reached. However, in this case empty classes may occur.

In a variation of the equidistant classification, the statistically unreliable, i.e. low-occupied, classes are aggregated. However, in order to use this particular classification method the input parameters number of classes, size of the grouped classes and minimum occupancy of a class need to be defined in advance. QUALIVERMO (Sammer et al. 2012) suggests class widths for these cases:

- for metropolitan traffic in short distance areas: 2 km or 5 min,
- for regional and long-distance traffic: 5 km or 10 min.

### Application of distributions for validation and presentation purposes

The equidistant classification method requires three parameters as input (number, size, aggregation rules of classes). This leads to several reasonable combinations of parameters. In contrast, the equiquantile method requires only one parameter (number of classes). As the equiquantile method it the more generic representation of a demand distribution it should be used for model validation purposes.

Results from an equidistant classification display typical patterns of travel demand, e.g. decreasing demand with increasing distance. This may be helpful for interpreting results. Thus, an equidistant classification should be used for presentation purposes, but not for validation purposes.

## Selection of a quality indicator for distribution comparisons

To check the congruence of two distributions they can be plotted in a common diagram. However, a mere visual examination of congruence is often not adequate. In the following, a selection of different indicators to quantify the similarity or conformity of two distributions is presented.

### Comparison of distribution parameters

In general, distributions, i.e. their position, appearance and properties, can be described by distribution parameters. By a comparison of those distribution parameters, it is possible to determine the similarity between two distributions. Therefore, it is essential to record them for each distribution. These parameters are in particular:

- sample size  $N$  (this is equivalent to the total number of trips),
- weighted mean  $\bar{m}$  (this represents the mean indicator, e.g. mean trip distance),
- standard deviation  $s_m$ ,
- coefficient of variation  $V_m$ ,
- skew of the distribution  $\gamma_m$  and
- percentiles of the distribution, e.g.  $Q_{0.05}$ ,  $Q_{0.15}$ ,  $Q_{0.25}$ ,  $Q_{0.5}$ ,  $Q_{0.75}$ ,  $Q_{0.85}$ ,  $Q_{0.95}$ .

The calculation specifications of the parameters are shown in the appendix.

### Statistical tests

Various statistical tests can be used to check for differences between distributions and their parameters, respectively. For example, the  $t$ -test checks for significant deviations of the mean and the  $F$ -test checks for significant variance deviations (Backhaus et al. 2011).

In addition to these parametric tests, which check the distribution parameters, there are non-parametric or distribution-free tests, which check whether two samples derive from a common population. An example of this is the Kolmogorov-Smirnoff-test, suitable for interval scaled data. (Herz et al. 1992)

The Kolmogorov-Smirnoff-test initially determines the largest deviation in terms of absolute values between two discrete relative frequency distributions. A comparative measure is then obtained depending on the sample sizes and the accepted level of error. It should be noted that in this test only the position of relative frequencies is assessed.

In contrast to a household travel survey, a travel demand model represents a full census, which means that it covers a large sample size. Consequently, the test criterion is very strict, which often results in a negative test result (i.e. the distributions do not originate from a common population).



## Quality indicators

Quality indicators consider the similarity of two distributions  $x$  and  $y$ . This section will give an executive summary about available quality indicators. The corresponding calculation specifications for each indicator are shown in the appendix.

### Correlation Coefficient and Coefficient of Determination

The Correlation Coefficient  $R$  and the Coefficient of Determination  $R^2$  represent quality indicators that check the dependency on two (typically unclassified) datasets.  $R$  is a non-dimensional variable ranging between  $-1.0$  and  $1.0$  and reflecting the extent of the linear dependency between two data sets.  $R^2$  expresses the amount of variation explained by the independent variables of the model. The higher the Coefficient of Determination, the higher the proportion of data explained by the regression function.

It is to be considered critically that the Coefficient of Determination on the one hand does not indicate any causality of the observed correlation and on the other hand inevitably increases with an increasing number of examined factors (Cambridge Systematics Inc. 2014; Backhaus et al. 2011). Therefore,  $R^2$  is suitable as a quality indicator only to a limited extent. The thresholds for acceptable  $R^2$  values mentioned in the literature vary between 0.88, 0.95 (Cambridge Systematics Inc. 2014) and 0.98 (Department for Transport 2014).

### Mean Absolute Error, Euclidean Distance and Root Mean Squared Error

A common quality indicator is the Mean Absolute Error ( $MAE$ ). It describes, as the name implies, the average of all deviations between the modelled and observed values. One way to scale the  $MAE$  to a unit-less quality indicator is to divide the  $MAE$  by the sum of the observed values. The resulting quality indicator is called Relative Mean Absolute Error ( $\%MAE$ ). This  $\%MAE$  should not be confused with the Mean Absolute Percentage Error ( $MAPE$ ) (Vandeput 2019).

The Euclidean Distance  $d$  considers the sum of the squared deviations across all distribution classes. Since, as will be explained later, the use of relative frequencies is advisable, a standardization to a common unit, as recommended e.g. by Backhaus et al. 2011, is not necessary.

Another common quality indicator is the Root Mean Squared Error ( $RMSE$ ) and its relative and unit-less form ( $\%RMSE$ ). Compared to the  $MAE$ , the  $RMSE$  weights errors more strongly and thus has a stronger influence on the evaluation result (Cambridge Systematics Inc. 2014; Vandeput 2019).

### Theil's Forecast Accuracy Coefficient

Theil's Forecast Accuracy Coefficient examines the conformity of two distributions. Unfortunately, two indicators with the same name  $U$  have been developed. In order to distinguish them they are called  $U_1$  and  $U_2$ .

- $U_1$  ranges between 0 and 1, where 0 indicates a perfect match and 1 indicates the worst match. The worst match occurs either with a negative proportionality between

the considered distributions or when both distributions are continuously equal to zero. Since all distributions are evaluated better than a naïve forecast, it is not possible to interpret them unambiguously and therefore  $U_1$  should not be used (Bliemel 1973; Andres and Spiwoks 2000).

- $U_2$  ranges from 0 to  $\infty$ , where 0 indicates a perfect match. At  $U_2=1$  the comparative distribution has the same quality as a naïve forecast. Consequently,  $U_2 > 1$  indicates that the comparative distribution is worse to evaluate since even a naïve forecast is evaluated better.  $U_2$  is always preferable to  $U_1$  (Bliemel 1973, Andres and Spiwoks 2000).

Furthermore, the deviations between two distributions can be analysed more precisely with three components of error  $U^M$ ,  $U^S$  and  $U^C$  (FGSV 2006; Bliemel 1973; Andres and Spiwoks 2000).

- $U^M$  indicates the proportion of the mean squared error resulting from an inequality of the mean values, which in turn results from a systematic over- or underestimation.
- $U^S$  indicates systematic differences in the variances.
- $U^C$  indicates the absence of a linear correlation between the two distributions. It is an indicator for unsystematic, random errors.

According to Andres and Spiwoks (2000) and FGSV (2006) a good conformity is assumed if  $U^M$  and  $U^S$  are close to 0 ( $U^M, U^S \rightarrow 0$  or at least  $U^M, U^S < 0.2$ ) and when the total error is mostly due to unsystematic errors ( $U^C \rightarrow 1$ ).

### Vortisch's Indicator of Similarity

A distance indicator presented by Vortisch (2006) takes into account the similarity in form and position of two distributions (see Fig. 4).

The Correlation Coefficient  $R$  describes the similarity of the shapes. Furthermore, the positional similarity  $\theta$  and the overlapping of the domains  $\sigma$  are determined. The individual components are then combined to a general distance indicator  $\Delta$ . If the distributions match perfectly,  $\Delta=0$  will result and with increasing difference it will strive against  $\Delta=1$ .

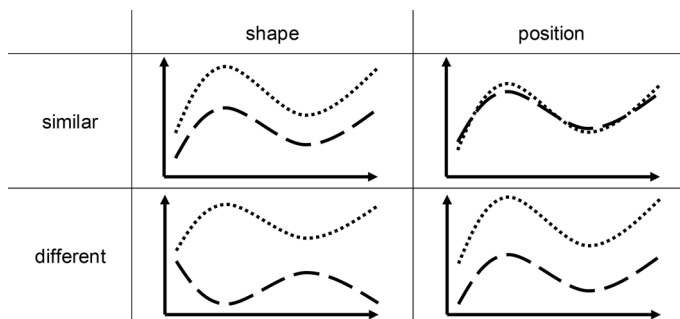


Fig. 4 Similarity in form and position of two distributions (modified figure according to (Vortisch 2006))

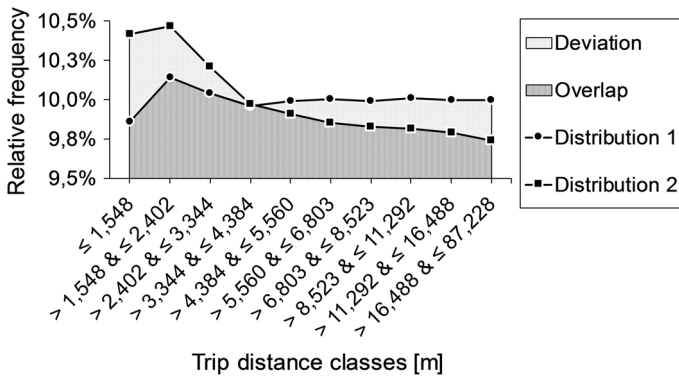


Fig. 5 Overlapping of distributions

### Coincidence Ratio

The Coincidence Ratio (*CR*) determines the degree to which two distributions overlap (see Fig. 5). Input value for the calculation are relative frequencies of classified distributions. The *CR* ranges from 0 to 1, where 1 indicates a perfect match and 0 indicates no match. According to TRAVEL MODEL VALIDATION AND REASONABLENESS CHECKING MANUAL (Cambridge Systematics Inc. 2014), a high degree of congruence applies if  $CR > 0.7$ .

### Absolute or relative frequencies

As shown in Table 2, a comparison of two distributions can either use relative or absolute frequencies. The choice of relative or absolute frequencies influences the value of the quality indicators. Hence, if two distributions are to be evaluated in terms of their similarity, it must be clarified beforehand whether the quality indicator used should operate with absolute or relative frequencies. As distance and time distributions derived from surveys are usually provided only as relative values, relative frequencies should be used. Furthermore, in a travel demand model, the absolute number of generated trips should be checked directly after trip generation. If this trip generation validation is omitted, systematic differences, such as permanently too few trips per class, are not recognized.

### Summary on quality indicators

Table 2 compares the presented quality indicators by the following criteria (☒ meaning “yes” and ☐ meaning “no”):

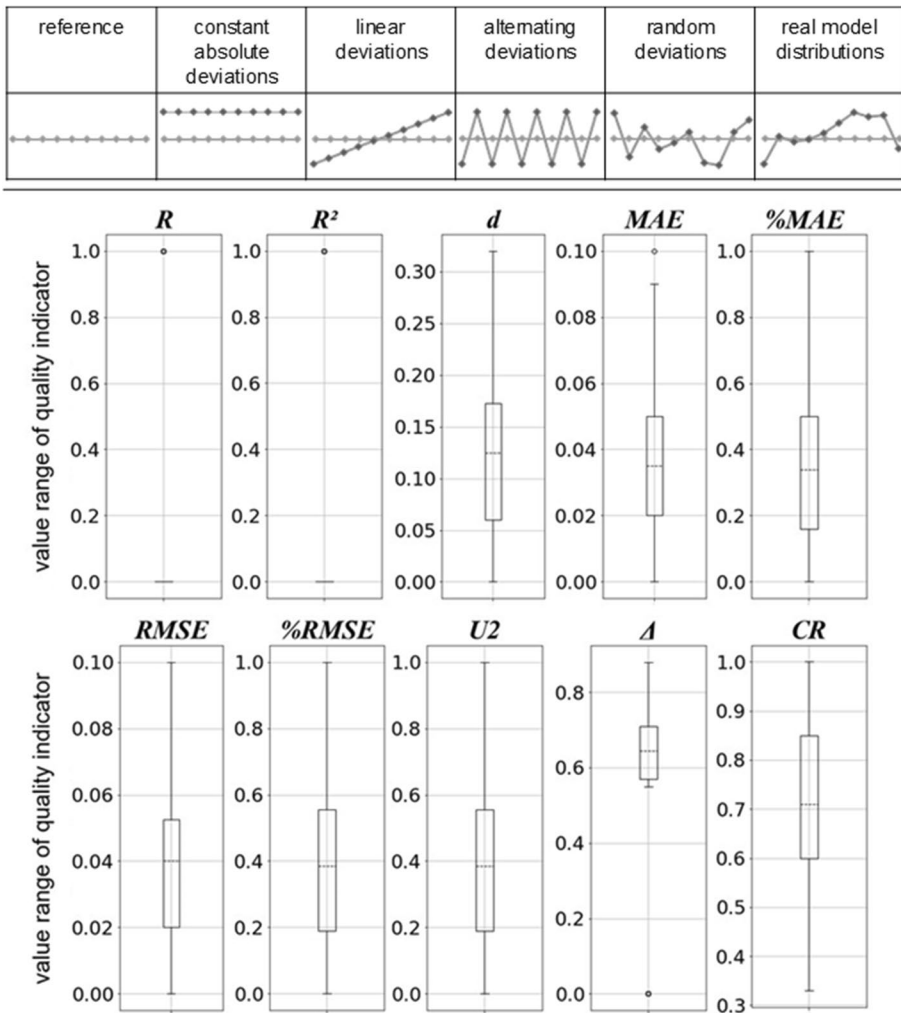
- What is the possible range of values of the indicators?
- Does the calculation specification require absolute or relative frequencies?
- Is there a difference in the resulting values of the indicator when using relative and absolute frequencies?
- Is there a possibility for in-depth analysis using the same indicator?
- Are there applicable thresholds for the indicator?

**Table 2** Comparison of quality indicators

Quality indicator	Range of values	frequencies recom- mended		Differences in resulting values when using relative and absolute frequencies	Possibility for in- depth analysis	Existing and applicable thresholds
		Relative	Absolute			
Correlation Coefficient $R$	$[-1, 1]$	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Coefficient of Determination $R^2$	$[0, 1]$	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Mean Absolute Error $MAE$	$[0, \infty]$	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Relative Mean Absolute Error $\%MAE$	$[0, \infty]$	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Euclidean Distance $d$	$[0, \infty]$	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/> (with standardized values)	<input type="checkbox"/>	<input type="checkbox"/>
Root Mean Squared Error $RMSR$	$[0, \infty]$	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Relative Root Mean Squared Error $\%RMSE$	$[0, \infty]$	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Theil's Forecast Accuracy Coefficient $U_2$	$[0, \infty]$	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Vortisch's Indicator of Similarity $\Delta$	$[0, 1]$	<input type="checkbox"/>	<input checked="" type="checkbox"/>	$R: \square; \theta, \Delta: \square$	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Coincidence Ratio $CR$	$[0, 1]$	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

In a benchmark comparison, over 30 equiquantile distributions with 10 classes were compared. These were derived from real travel demand models as well as systematically generated in order to examine certain characteristics of the indicators (see top Fig. 6). The box plots in bottom Fig. 6 show the result of this comparison. The following conclusions can be derived:

- Correlation Coefficient  $R$  and Coefficient of Determination  $R^2$  fail for ideal equiquantile distributions, because the mean of the reference distribution matches the demand per class, resulting in an unsolvable expression.
- The results for Euclidean Distance  $d$ , Mean Absolute Error  $MAE$  and Root Mean Squared Error  $RMSE$  show a very small bandwidth in the present case of relative fre-



**Fig. 6** Top figure: distribution types examined for a benchmark comparison, Bottom figure: result from the benchmark comparison of the different quality indicators

quencies. Therefore, their relative forms Relative Mean Absolute Error  $\%MAE$  or relative Root Mean Squared Error  $\%RMSE$  are to be preferred. A general statement as to which of these quality indicators is more suitable is not possible, since this depends, among other things, on the respective accuracy requirements of the model.

- The expressions for Relative Root Mean Squared Error  $\%RMSE$  and Theil's Forecast Accuracy Coefficient  $U_2$  are mathematically identical in this case of relative deviations and classified data.
- Due to the properties described above, Theil's Forecast Accuracy Coefficient  $U_2$  and Vortisch's Indicator of Similarity  $\Delta$  are suitable because of their sophisticated analysis options. However, the latter is less sensitive, which is reflected in the fact that it does not make full use of its value range even in the case of large deviations.
- The Coincidence Ratio  $CR$  is sensitive to changes, it can take values in its entire value range and its value range from 0 to 1 speaks for a straightforward interpretability. In addition, specific thresholds from the TRAVEL MODEL VALIDATION AND REASONABLENESS CHECKING MANUAL (Cambridge Systematics Inc. 2014) provide good orientation for the evaluation of quality.

In summary, the Coincidence Ratio  $CR$ , the Relative Root Mean Squared Error  $\%RMSE$  (or Theil's Forecast Accuracy Coefficient  $U_2$ ) and the Relative Mean Absolute Error  $\%MAE$  seem suitable for the evaluation of equiquantile classified relative frequency distributions. In addition, the three components of Theil's Forecast Accuracy Coefficient  $U^M$ ,  $U^S$  and  $U^C$  are useful for in-depth analyses.

## Conclusion and recommendations

Appropriate transport planning requires appropriate travel demand models. Currently, many discussions are taking place about what "appropriate" actually means. This paper is a contribution to those discussions. Frequency distributions are an important quality feature of travel demand models. In quality assessment it is particularly important that the classification method is known, because the current techniques that are used in everyday practice lead to individual classifications and thus to different evaluation results. This makes it impossible to compare validation results of different models and to define a quality indicator threshold in modelling guidelines.

In the previous chapters, various classification and quality determination methods have been presented. In order to ensure a uniform classification method in quality assurance, the following procedure is proposed for the standardized creation and assessment of distributions. This proposal is intended to serve as a basic concept. For specific travel demand model applications, the specifications have to be adjusted if necessary. Modelling guidelines should define a generally obligatory procedure in the future.

### 1. Selection of mode-independent classification indicators for each distribution:

- for trip distance distributions: direct distance,
- for trip time distributions: mean weighted trip time in the reference situation.
- The intrazonal trips are not taken into account in this evaluation. They have to be evaluated separately.

## 2. Specification of the study area:

- only areas covered by the reference *and* the model are considered,
- only trips with origin and destination inside the study area are considered,
- other trips have to be evaluated separately.

## 3. Specification of the reference distribution for each application:

- for calibrating and validating a travel demand model: a distribution from a household travel survey (keeping in mind that such surveys contain a systematic error),
- for comparisons of modelled scenarios: a distribution from the base case of a travel demand model.

## 4. Calculation of *ten equiquantile classes* for all relevant demand segments *and* for the total demand and

- visualization of the distributions with relative frequencies,
- presentation of the distribution parameters,
- using the Coincidence Ratio as quality indicator (calculation with relative frequencies). According to TRAVEL MODEL VALIDATION AND REASONABLENESS CHECKING MANUAL (Cambridge Systematics Inc. 2014), the Coincidence Ratio should be  $CR \geq 0.7$  for a high level of congruence.

## 5. Optional: Visualization of *equidistant* distributions. To determine the class width, it is possible to use the smallest class of the equiquantile classified of the total demand. Equidistant distributions should *not* be used for quality tests. They are created for display purposes only.

**Acknowledgements** Open Access funding provided by Projekt DEAL. This paper was developed as part of the research project „Influencing factors on the quality of macroscopic travel demand models“ (Einflussgrößen auf die Qualität von makroskopischen Nachfragemodellen im Personenverkehr) (Friedrich et al. 2015–2019) commissioned by the German Research Foundation (DFG). The classification methodology described was tested on three different macroscopic travel demand models in the research project „Quality assurance for transport model calculations“ (Rieser et al. 2015) commissioned by the Swiss Association of Transportation Engineers and Experts (SVI). Special thanks to PD Dr. Christian Schiller and Robert Simon from the Technical University of Dresden, Dr. Nadine Rieser and Bence Tasnády from EBP Schweiz AG, Dr. Juliane Pillat from PTV AG as well as Prof. Dr. Markus Friedrich from the University of Stuttgart. The in-depth discussions within the scope of the above-mentioned research projects have greatly enriched this paper. Last but not least, I would like to thank the reviewers for their comments and efforts towards improving the paper.

**Author contribution** The author Eric Pestel devised the conceptual idea as well as the formal definitions of the presented methodology and he wrote the manuscript including the presented examples.

## Compliance with ethical standards

**Conflict of interest** The corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons

licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

### Calculation specification and example of the equiquantile classification method

Initially, the indicator values  $v_n$  are sorted in ascending order. The demand values  $w_n$  are sorted according to the indicator sorting. Formula (1) is used to calculate the unweighted quantiles  $q_n$ .

$$q_n = \frac{(n - 0,5)}{N}; \quad q_n \in (0;1] \quad (1)$$

Formula (1) is then extended by the demand values (weight)  $w_n$  to formula (2):

$$q_{w,n} = \frac{(w_{cum,n} - 0,5 \cdot w_n)}{\sum_{n=1}^N w_n}; \quad q_{w,n} \in (0;1] \quad (2)$$

The desired (upper) class boundaries of the indicators  $v_q$  are the result of a linear interpolation between the relevant weighted discrete quantiles  $q_{w,n}$ . The demand is allocated to the resulting classes. This leads to the classified equiquantile demand  $w_q$ .

*Annotations:*

- $n$ :  $n$ th element of the distribution
- $N$ : population size (number of individual values)
- $q_n$ :  $n$ th quantile
- $q_{w,n}$ :  $n$ th weighted quantile
- $v_n$ :  $n$ th indicator
- $v_q$ : upper indicator class boundaries of the  $q$ th-quantile
- $w_n$ :  $n$ th demand element (weight)
- $w_{cum,n}$ : cumulated weight of the  $n$ th element
- $w_q$ : equiquantile demand within the  $q$ th-quantile

The following example with 20 OD-pairs illustrates the calculation procedure. Please note that with such a small sample size it is not possible to distribute the demand completely and evenly across all classes, but that this effect is reduced as the sample size increases.



• Input variables:

Desired quantile $q$	OD-indicators $v$	OD-demand $w$
0.1	92.0	627.0
0.2	17.0	841.8
0.3	53.0	34.1
0.4	86.0	592.2
0.5	83.0	846.6
0.6	15.0	846.6
0.7	3.0	196.6
0.8	54.0	550.4
0.9	20.0	223.6
1.0	21.0	403.7
	35.0	268.5
	90.0	255.2
	7.0	90.3
	62.0	301.6
	94.0	213.7
	37.0	506.5
	30.0	220.5
	1.0	562.5
	34.0	43.5
	43.0	814.0

• Calculation:

column index →	1	2	3	4	5	6
row index $n$	sorted indicators $v_n$	sorted Demand (weight) $w_n$	0.5 * col. 2	col. 2 cumulated	col. 4–col. 3	col. 5 / sum(col. 2) = weighted quantile $q_{w,n}$
1	1.0	562.5	281.3	562.5	281.3	0.033
2	3.0	196.6	98.3	759.1	660.8	0.078
3	7.0	90.3	45.2	849.4	804.3	0.095
4	15.0	846.6	423.3	1696.0	1272.7	0.151
5	17.0	841.8	420.9	2537.8	2116.9	0.251
6	20.0	223.6	111.8	2761.4	2649.6	0.314
7	21.0	403.7	201.9	3165.1	2963.3	0.351
8	30.0	220.5	110.3	3385.6	3275.4	0.388
9	34.0	43.5	21.8	3429.1	3407.4	0.404
10	35.0	268.5	134.3	3697.6	3563.4	0.422
11	37.0	506.5	253.3	4204.1	3950.9	0.468
12	43.0	814.0	407.0	5018.1	4611.1	0.546
13	53.0	34.1	17.1	5052.2	5035.2	0.597
14	54.0	550.4	275.2	5602.6	5327.4	0.631

column index →	1	2	3	4	5	6
row index $n$	sorted indicators $v_n$	sorted Demand (weight) $w_n$	$0.5 * \text{col. 2}$	col. 2 cumulated	col. 4–col. 3	col. 5 / sum(col. 2) = weighted quantile $q_{w,n}$
15	62.0	301.6	150.8	5904.2	5753.4	0.682
16	83.0	846.6	423.3	6750.8	6327.5	0.750
17	86.0	592.2	296.1	7343.0	7046.9	0.835
18	90.0	255.2	127.6	7598.2	7470.6	0.885
19	92.0	627.0	313.5	8225.2	7911.7	0.938
20	94.0	213.7	106.9	8438.9	8332.1	0.987
$\Sigma$		<b>8438.9</b>				

- Result:

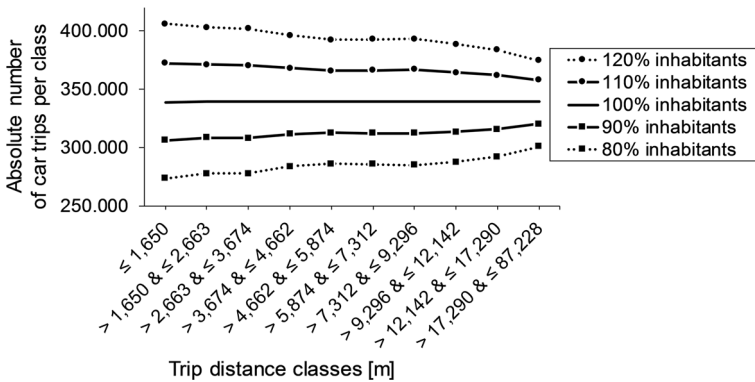
desired quantile $q$	upper indicator class boundaries $v_q$	absolute demand per class $w_q$ [–]	relative demand per class $w_q$ [%]
0.1	7.7	849.4	10.1
0.2	16.0	846.6	10.0
0.3	19.3	841.8	10.0
0.4	33.0	847.8	10.0
0.5	39.4	818.5	9.7
0.6	53.1	848.1	10.0
0.7	67.6	852.0	10.1
0.8	84.8	846.6	10.0
0.9	90.6	847.4	10.0
1.0	94.0	840.7	10.0

## Applications of the equiquantile classification method

### Equiquantile classes for a particular demand segment

Demand segments are a subset of the total demand. Usually demand is segmented by mode, by trip purpose or by the combination of mode and purpose. As each demand segment has a specific distribution, this leads to specific class widths for each segment. A demand segment “pedestrians” will have more short distance classes than a demand segment “car”. For comparing distributions of a particular demand segment in different scenarios, it is therefore recommendable to determine the class widths with the base case demand of that particular demand segment. The classification should be determined once for the reference distribution of the base case and then preserved for all scenarios.

Figure 7 shows an example of an equiquantile trip distance distribution of the mode car for a realism test, in which the number of inhabitants is changed by  $\pm 10\%$  and  $\pm 20\%$ . The base case serves as reference distribution. In the example, it can be seen



**Fig. 7** Example of an equiquantile trip distance distribution of mode car with different numbers of inhabitants

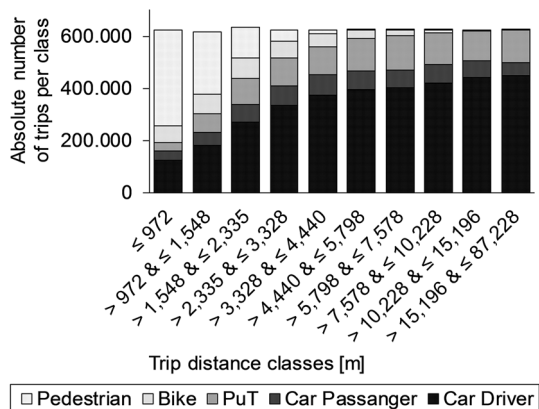
that in the case of less inhabitants (80% or 90%) the number of trips decreases across all classes. However, the decrease is not equal in all classes. Short distance classes show a lower share of the total demand, longer distance classes a higher share. This can be explained by the relationship between demand and supply, where lower demand leads to travel time reductions, which then increase travel distance. The opposite applies to the case with more inhabitants (110% or 120%).

**Equiquantile classes for the total demand**

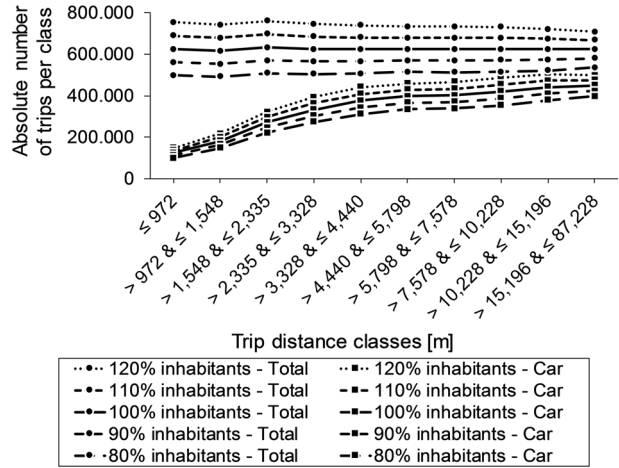
Similar to the evaluation for particular demand segments, it is possible to perform an evaluation of the total demand across all modes to analyse the changes of the total demand in the different distribution classes.

In addition, individual demand segments can also be presented with the classification of the total demand. For example, a distance-dependent modal split (see Fig. 8) or the development of individual modes in comparison to the total demand for different scenarios (see Fig. 9) can be displayed.

**Fig. 8** Example of an equiquantile trip distance distribution of all modes (distance-dependent modal split)



**Fig. 9** Example of an equiquantile trip distance distribution of mode car and for the total demand with different numbers of inhabitants

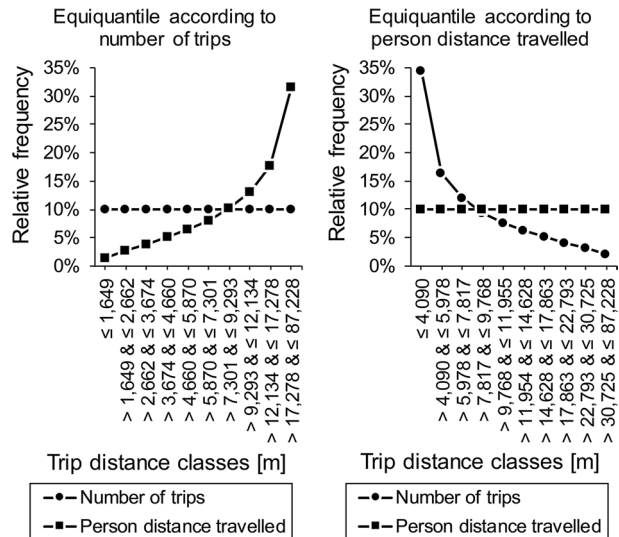


**Equiquantile classes from person distance travelled**

As a modification of the equiquantile classification, it is possible to use the combination of demand and trip distance (person distance travelled) as classification indicator. For vehicle distance travelled or person time travelled, an analogous approach has to be used.

This is important in the calibration process, where it is more important to match the person distance travelled than the actual number of trips. An equal number of trips in all distance classes leads to an increasing person distance travelled with increasing distance class, as is shown in Fig. 10 (left). In contrast, Fig. 10 (right) shows a distribution where each class contains the same amount of person distance travelled, while the number of trips is decreasing monotonously with increasing distance classes. In comparison to a “normal” equiquantile distribution, shown in Fig. 10 (left), it becomes obvious that the class width in the short-range distances increases while the long distance classes are displayed in greater detail.

**Fig. 10** Comparison of an equiquantile distribution according to the number of trips (left) and an equiquantile distribution according to the person distance travelled (right).



### Calculation specification and explanation of the distribution parameters

The calculation specifications presented refer to classified data—therefore it is irrelevant whether relative or absolute frequencies are used.

$$\bar{m} = \frac{\sum_{k=1}^K (x_k \cdot m_k)}{N} \tag{3}$$

$$s_m = \sqrt{\frac{\sum_{k=1}^K (x_k \cdot (m_k - \bar{m})^2)}{N - 1}} \text{ (for a sample)} \tag{4}$$

$$s_m = \sqrt{\frac{\sum_{k=1}^K (x_k \cdot (m_k - \bar{m})^2)}{N}} \text{ (for a finite population)} \tag{5}$$

$$V_m = \frac{s_m}{\bar{m}} \tag{6}$$

$$\gamma_m = \frac{\frac{\sum_{k=1}^K (x_k \cdot (m_k - \bar{m})^3)}{(N-1)}}{\sqrt{\left(\frac{\sum_{k=1}^K (x_k \cdot (m_k - \bar{m})^2)}{(N-1)}\right)^3}} \tag{7}$$

*Annotations:*

- $\bar{m}$ : weighted mean across all classes (mean indicator)
- $s_m$ : weighted standard deviation (standard deviation of the mean indicator)
- $V_m$ : coefficient of variation
- $\gamma_m$ : skewness of the distribution
- $x_k$ : frequency in class  $k$
- $m_k$ : mean of the class  $k$
- $K$ : number of classes
- $N$ : population size (number of individual values).

### Calculation specification and explanation of the quality indicators

**Correlation Coefficient and Coefficient of Determination (adapted for classified datasets)**

$$R = \frac{\sum_{k=1}^K (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^K (x_k - \bar{x})^2 \cdot \sum_{k=1}^K (y_k - \bar{y})^2}} \tag{8}$$

*Annotations:*

$R$ : Correlation Coefficient

$x_k$  or  $y_k$ : frequency of distribution  $x$  or  $y$  in class  $k$

$K$ : number of classes

$\bar{x}$  or  $\bar{y}$ : mean of dataset  $x$  or  $y$ .

### Mean Absolute Error and Relative Mean Absolute Error (adapted for classified datasets)

$$MAE = \frac{\sum_{k=1}^K |x_k - y_k|}{K} \quad (9)$$

$$\%MAE = \frac{MAE}{\frac{1}{K} \cdot \sum_{k=1}^K x_k} = \frac{\sum_{k=1}^K |x_k - y_k|}{\sum_{k=1}^K x_k} \quad (10)$$

*Annotations:*

$MAE$ : Mean Absolute Error

$\%MAE$ : Relative Mean Absolute Error

$x_k$  or  $y_k$ : frequency of distribution  $x$  or  $y$  in class  $k$

$K$ : number of classes

### Euclidean Distance (adapted for classified datasets)

$$d = \sqrt{\sum_{k=1}^K (x_k - y_k)^2} \quad (11)$$

*Annotations:*

$d$ : Euclidean Distance

$K$ : number of classes

$x_k$  or  $y_k$ : frequency of distribution  $x$  or  $y$  in class  $k$

### Root Mean Squared Error (adapted for classified datasets)

$$RMSE = \sqrt{\frac{\sum_{k=1}^K (x_k - y_k)^2}{K}} \quad (12)$$

$$\%RMSE = \frac{RMSE}{\frac{1}{K} \cdot \sum_{k=1}^K x_k} = \frac{\sqrt{K \cdot \sum_{k=1}^K (x_k - y_k)^2}}{\sum_{k=1}^K x_k} \quad (13)$$

*Annotations:*

$RMSE$ : Root Mean Squared Error

$\%RMSE$ : Relative Root Mean Squared Error

$x_k$  or  $y_k$ : frequency of distribution  $x$  or  $y$  in class  $k$

$K$ : number of classes

**Theil's Forecast Accuracy Coefficient (adapted for classified datasets)**

$$U_1 = \frac{\sqrt{\frac{1}{K} \sum_{k=1}^K (x_k - y_k)^2}}{\sqrt{\frac{1}{K} \sum_{k=1}^K (x_k)^2} + \sqrt{\frac{1}{K} \sum_{k=1}^K (y_k)^2}} \tag{14}$$

$$U_2 = \frac{\sqrt{\sum_{k=1}^K (x_k - y_k)^2}}{\sqrt{\sum_{k=1}^K (x_k)^2}} \tag{15}$$

$$U^M = \frac{(\bar{x} - \bar{y})^2}{\frac{1}{K} \sum_{k=1}^K (x_k - y_k)^2} \tag{16}$$

$$U^S = \frac{(s_x - s_y)^2}{\frac{1}{K} \sum_{k=1}^K (x_k - y_k)^2} \tag{17}$$

$$U^C = \frac{2 \cdot (1 - R) \cdot s_x \cdot s_y}{\frac{1}{K} \sum_{k=1}^K (x_k - y_k)^2} \tag{18}$$

*Annotations:*

$U_1$ : Theil's Forecast Accuracy Coefficient (old form; not recommended according to Bliemel (1973) and Andres and Spiwoks (2000).

$U_2$ : Theil's Forecast Accuracy Coefficient (new form)

$U^M$ : proportion of error from systematic differences in the means

$U^S$ : proportion of error from systematic differences in the variances

$U^C$ : proportion of error from unsystematic errors

$K$ : number of classes

$x_k$  or  $y_k$ : frequency of distribution  $x$  or  $y$  in class  $k$ , for  $U_2$  explicitly relative frequencies (Bliemel 1973)

$\bar{x}$  or  $\bar{y}$ : mean of distribution  $x$  or  $y$  across all classes

$s_x$  or  $s_y$ : standard deviation of distribution  $x$  or  $y$

$R$ : Correlation Coefficient

**Vortisch's Indicator of Similarity**

$$\theta = \frac{1}{|D(x, y)|} \sum_{k=1}^K \frac{\min\{x_k, y_k\}}{\max\{x_k, y_k\}} \tag{19}$$

$$\sigma = \frac{|D(x, y)|}{|D(x) \cup D(y)|} \tag{20}$$

$$\Delta = 1 - (\alpha \cdot R + (1 - \alpha) \cdot \theta) \cdot (\gamma \cdot \sigma + (1 - \gamma) \cdot 1) \quad (21)$$

The parameters  $\alpha$  and  $\gamma$  can be used to control the influence of shape, position and domain. Vortisch (2006) recommends  $\alpha = \gamma = 0.5$  by default.

Please note the following special cases:

- one of the two distributions is constant:  $R = 0$
- both distributions are constant:  $R = 1$
- if  $x_k = y_k = 0$ , then  $\theta = 1$  applies
- if  $|D(x, y)| = 0$ , then  $\theta = 0$  applies

*Annotations:*

$\Delta$ : Vortisch's Indicator of Similarity

$\theta$ : similarity of the shape

$\sigma$ : similarity of the position

$R$ : Correlation Coefficient

$K$ : number of classes

$x_k$  or  $y_k$ : absolute frequency of distribution  $x$  or  $y$  in class  $k$

$D(x, y)$ : common domain of distribution  $x$  and  $y$

$D(x)$  or  $D(y)$ : domain of distribution  $x$  or  $y$

$\alpha$ : influence of shape and position  $\in [0;1]$

$\gamma$ : influence of domain  $\in [0;1]$

## Coincidence Ratio

$$CR = \frac{\sum_{k=1}^K \min\{p_k, q_k\}}{\sum_{k=1}^K \max\{p_k, q_k\}} \quad (22)$$

where

$$p_k = \frac{x_k}{\sum_{k=1}^K x_k} \quad (23)$$

$$q_k = \frac{y_k}{\sum_{k=1}^K y_k} \quad (24)$$

*Annotations:*

$CR$ : Coincidence Ratio

$K$ : number of classes

$p_k$  or  $q_k$ : relative frequency of distribution  $x$  or  $y$  in class  $k$

$x_k$  or  $y_k$ : absolute frequency of distribution  $x$  or  $y$  in class  $k$ .



## References

- Andres, P., Spiwoks, M.: Prognosegütemaße. State of the Art der statistischen Ex-post-Beurteilung von Prognosen. Sofia-Studien zur Institutionenanalyse, Nr. 00-1. Sofia, Darmstadt (2000)
- Backhaus, K., Erichson, B., Plinke, W., Weiber, R.: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. In: 13th revised edition. Springer, Berlin, Dordrecht, London, New York (2011)
- Bhatta, B.P., Larsen, O.I.: Are intrazonal trips ignorable? *Transp. Policy* **18**(1), 13–22 (2011). <https://doi.org/10.1016/j.tranpol.2010.04.004>
- Bliemel, F.: Theil's forecast accuracy coefficient. A clarification. *J. Mark. Res.* **X**, 444–446 (1973)
- Cambridge Systematics Inc.: Travel Model Validation and Reasonableness Checking Manual, 2nd edn. Cambridge Univ Press, Cambridge (2014)
- Department for Transport (ed): TAG UNIT M3.1. Highway Assignment Modelling. Transport analysis guidance: WebTAG, M3.1 (2014)
- FGSV—Forschungsgesellschaft für Straßen- und Verkehrswesen (German Road and Transportation Research Association) (ed): Hinweise zur mikroskopischen Verkehrsflusssimulation. Grundlagen und Anwendungen, vol. 388. FGSV Verlag, Köln (2006)
- FGSV—Forschungsgesellschaft für Straßen- und Verkehrswesen (German Road and Transportation Research Association) (ed): Empfehlungen für Verkehrserhebungen. EVE, vol. 125. FGSV Verlag, Köln (2012)
- Friedrich, M., Schiller, C., Pestel, E., Simon, R., Schimpf, M.: Influencing Factors on the Quality of Macroscopic Travel Demand Models. Einflussgrößen auf die Qualität von makroskopischen Nachfragemodellen im Personenverkehr. DFG research project (FR 2666/3-1) (2015–2019)
- Herz, R., Schlichter, H.G., Siegner, W.: Angewandte Statistik für Verkehrs- und Regionalplaner. In: 2nd revised and extended edition, Werner-Ingenieur-Texte, vol 42. Werner, Düsseldorf (1992)
- Hyndman, R.J.: The Problem with Sturges' Rule for Constructing Histograms (1995). <https://robjhyndman.com/papers/sturges.pdf>
- Paluš, M.: Testing for nonlinearity using redundancies: quantitative and qualitative aspects. *Physica D* **1995**(80), 186–205 (1995)
- Rieser, N., Tasnády, B., Friedrich, M., Pestel, E., Vries, N. de, Rothenfluh, M., Fischer, R.: Quality assurance for transport models and their applications. Qualitätssicherung von Verkehrsmodellberechnungen; SVI research project (2015/001). (2018)
- Sammer, G., Gruber, C., Roeschel, G., Tomschy, R., Herry, M.: The dilemma of systematic underreporting of travel behavior when conducting travel diary surveys—a meta-analysis and methodological considerations to solve the problem. *Transp. Res. Procedia* **32**, 649–658 (2018). <https://doi.org/10.1016/j.trpro.2018.10.006>
- Sammer, G., Röschel, G., Gruber, C.: Qualitätssicherung für die Anwendung von Verkehrsnachfragemodellen und Verkehrsprognosen. Quality Management and Model Validation for Application of Transport Demand Modelling and Forecast. Straßenforschung, Heft 604. Österreichische Forschungsgesellschaft Straße, Schiene, Verkehr, Wien (2012)
- Vandeput, N: Forecast KPI: RMSE, MAE, MAPE and Bias (2019). <https://medium.com/analytics-vidhya/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d>. Accessed 06 Nov 2019
- Vortisch, P.: Modellunterstützte Messwertpropagierung zur Verkehrslageschätzung in Stadtstraßennetzen. Dissertation. Scientific series, University of Karlsruhe (TH), vol. 64, Karlsruhe (2006)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Eric Pestel** is a Research assistant at the Chair for Transport Planning and Traffic Engineering (University of Stuttgart) since 2015.