Reliability assessment of dynamic OD estimation methods based on structural similarity index

Conference Paper · January 2013

DOI: 10.13140/RG.2.1.4174.1929

1 author:

Tamara Djukic
TSS - Transport Simulation Systems, Barcelona, Spain

TRB 13-XXXX

Reliability assessment of dynamic OD estimation methods based on structural similarity index

Tamara Djukic, Ms.C. (corresponding author)
T.Djukic@tudelft.nl
Transportation and Planning Department
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, P.O. Box 5048,
2600 GA Delft, the Netherlands

Professor S.P. Hoogendoorn, Ph.D., Ms.C.
S.P.Hoogendoorn@tudelft.nl
Transportation and Planning Department
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, P.O. Box 5048,
2600 GA Delft, the Netherlands

J.W.C. van Lint, Ph.D., Ms.C.
J.W.C.vanLint@tudelft.nl
Transportation and Planning Department
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, P.O. Box 5048,
2600 GA Delft, the Netherlands

Word count:
Words (including abstract and references): 6132
Tables and figures: 8 * 250 = 2000
Total: = 8132

Submitted for publication and presentation for the 92nd meeting of the Transportation Research Board, 13-17 January 2013, Washington D.C.

Abstract

The statistical measures used for quality assessment of estimated OD matrices typically quantify the difference between the estimated OD matrix and available true/reference OD matrix. Although the underlying rational makes sense intuitively, the actual statistical measures in literature, such as MSE, do not capture the most important aspect: the structural similarity of the estimated and reference OD matrix. In this paper we propose a new quality measure that does incorporate such a term, so-called Structural SIMilarity (SSIM) index.

In this paper we explore the application perspectives of SSIM index for this purpose. First, we investigate the properties of SSIM index compared to some statistical measures. Then, we show how SSIM index can be used as additional performance measure for benchmarking the dynamic OD estimation methods. More over, we provide insight into how SSIM index can be used further as a new performance function to estimate dynamic OD matrices.

INTRODUCTION

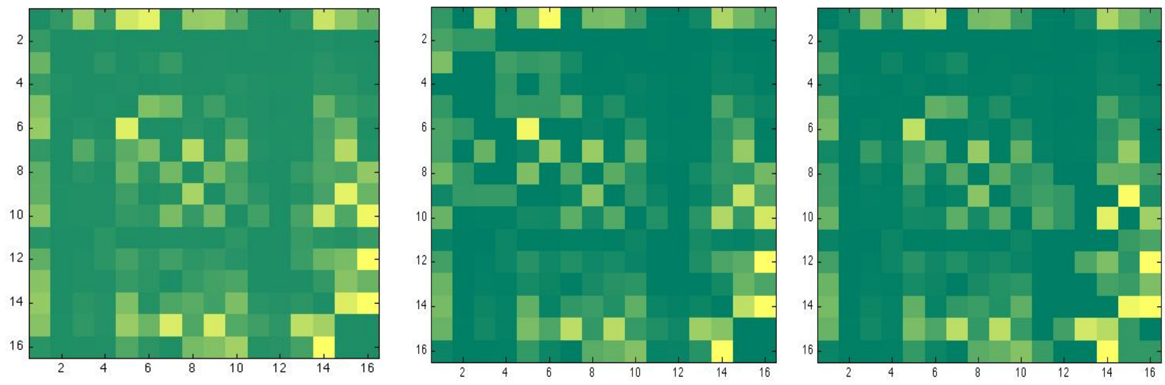
The ex-post and ex-ante evaluation of traffic and demand management measures, and transport policy measures requires a very high quality of the traffic variables. In particular, important input to the models used for assessing such measures are OD matrices. The important role of OD demand has resulted in a variety of mathematical approaches to estimate and predict dynamic OD matrices, such as the Generalized least square models *(1-3)*, the Maximum Entropy models *(4, 5)*, the Maximum likelihood *(6)*, Bayesian inference model *(7, 8)* and the Kalman filter models *(9, 10)*. Thus, there is increased need for quantitative objective measures to assess the quality of existing OD estimation methods to pinpoint their strengths and weakness and applicability and validity under different circumstances.

In this paper we will focus on an important and often-overlooked aspect in the benchmark of different OD estimation methods, namely the selection of an appropriate measure or a set of criteria that can be used to compute the quality of estimated OD matrices. Few studies have focused on evaluation of the reliability and accuracy of the estimated OD matrices in absence of ground truth OD matrix *(1, 11)*, and on available ground truth OD matrix such as *(12, 13)*. A number of statistical measures have been proposed and used in literature to evaluate the quality of an OD estimator, such as root mean square error and mean percentage error. However, the basic principal of these performance indices is that they are expressed as deviations in terms of OD demand or traffic counts in respect to ground truth data. These statistical measures are widely used because they are simple to calculate and have clear meanings. It is worth nothing that these statistical measures are not very well matched to capture the structural patterns between estimated and ground truth OD pairs. In the ideal case, the estimated OD matrix reflects the OD matrix which is very close to the actual OD matrix, in the sense that it has a similar structure, for example in terms of distribution of trips over destinations, and the trip length distribution.

The OD matrices may be determined from different sources of information (e.g. land-use models, travel surveys) using different methods, which represent a common spatial and temporal behavior of travelers (e.g. choice of destination, departure time, mode choice). For example, the gravity model illustrates the macroscopic relationships between places (say homes and workplaces). It has long been postulated that the interaction between two locations declines with increasing (distance, time, and cost) between them, but at the same time, the interaction is positively associated with the amount of activity at each location. These rules yield the structural correlation between OD pairs. Such OD matrices are thus highly structured and stem from the combination of various kinds of information, such as OD matrix structure and the correlation between OD pairs.

Statistical measures that ignore this spatial correlation between OD pairs in OD matrices may fail to provide effective and accurate quality measures. To show this, we will use the mean square error (MSE) as an example. In FIGURE 1 (b) and (c), we compare the two estimated OD matrices using two different OD estimation methods with the available “ground truth” OD matrix (a). For better visual examination of the structure in the OD matrices, where the origin zones are given in rows and destinations in columns, they are represented as images where the number of trips per OD pair is used as index into the colormap that determine the color for each OD pair. Then, the number of trips in $X_{1,1}$ (FIGURE 1 (a)) represents the color of the OD pair at position as an index into the colormap. For example, if $X_{1,1} = k$; then the color of OD pair $X_{1,1}$ is the color represented by row *k* of the color map. In Figure 1, the OD pair with demand of 20 trips has a light yellow color, and OD pair with demand of 400 trips has a dark green color. The MSE between ground truth OD matrix and both of the estimated OD matrices are exactly the same. However, the visual examination of the two estimated OD matrices clearly indicate that they capture different structural patterns. Therefore, such measures are only designed to find the distances between a pair of attributes in a data set or overall distance amongst all data. They are unable to find and distinguish different correlation structures in data, and cannot provide an in depth explanation of the patterns in OD demand.

One possible solution approach to tackle such a problem is to propose a new quality assessment measure that will incorporate the structural correlation between OD pairs. In this paper we propose the application of the so-called Structural Similarity (SSIM) Index *(14)*, which is an measure to quantify the similarity of two data sets, taking into account their structure that is expressed through correlation between OD pairs. Origination from image processing and analysis, the SSIM approach was motivated by the observation that image signals are highly structured, meaning that samples of natural image signals have strong dependencies. These dependencies carry important information about the structures of the objects in the visual scene.



a) b) c)

FIGURE 1 Comparison of patterns in reference and estimated OD matrices: a) "ground truth" OD matrix; b and c): estimated OD matrices that have the same MSE with respect to the reference OD matrix, but different structural patterns

In this paper, we propose using the SSIM index as a new quality metric for sensitivity assessment of dynamic OD estimation methods, and study its applicability. The new metric would enable researchers and practitioners better insight into how to assess the quality of the estimated OD matrix, and to make a strict conclusion about the quality and efficiency of OD estimation methods. More specifically, we argue that traffic engineers have rethink whether the statistical measures such as MSE or RMSE are the most useful criteria of choice in their comparative studies and applications. Also, we explore application potentials of SSIM index as a performance function to estimate the OD demand.

The paper is organized as follows. In the first part of the paper we will outline the theoretical background of SSIM and explain its main properties. Next we will present the insensitivity of the statistical measure MSE to identify and evaluate the structural pattern in OD matrices. In the second part of the paper we will first define the experimental data set and benchmark framework to assess the reliability of the several OD estimation methods. Next, we assess the quality of estimated OD matrices with different statistical measures and SSIM index. The paper closes with a discussion on further application perspectives of SSIM index in estimation and prediction of OD demand and further research.

QUALITY ASSESMENT BASED ON STRUCTURAL SIMILARITY INDEX

The idea of structural similarity

The statistical measures used for quality assessment of estimated OD matrices typically measure the difference between the estimated OD matrix and available true/reference OD matrix. Although the underlying rational makes sense intuitively, the actual statistical measures in literature, such as MSE, do not encapsulate the most important aspect: the structural similarity of the estimated and reference OD matrix. By structural similarity, we mean that the *spatial and temporal behavior of travelers reflected in OD trip patterns have a strong spatial and temporal correlation reflected by the OD pairs*. In the ideal case, the estimated OD matrix reflects the OD matrix which is very close to the actual OD matrix, in the sense that it has a similar structure, for example in terms of distribution of trips over destinations, and the trip length distribution. The statistical measures that ignore correlation in OD matrix data may fail to provide effective and accurate quality measures. Therefore, the key idea in our approach is to define a quality metric, in such way that the structural pattern in estimated OD matrix is quantified adequately.

We propose the application of the Structural SIMilarity (SSIM) index *(14)* that is typically used as a method for measuring the similarity between two images, based on the degradation of the structural information in one image compared to the reference image. In general, images are highly structured: their pixels exhibit strong dependencies, especially when they are close to each other, and these dependencies carry important information about the structure of the objects in the visual scene. The key idea is now, that if we represent the OD demand in the matrix form, the OD pairs can be seen as pixels in image that exhibit strong dependencies as well. Hence, the SSIM seems to be a good measure to compare OD matrices as well.

In next subsection we define and explain in detail the Structural Similarity (SSIM) index and its application to OD matrices. The measure ensures that the amount of structural information in reference OD matrix is preserved in estimated OD matrix. We will demonstrate and compare the suitability of the MSE and SSIM index under different scenarios, to illustrate the key benefits of the SSIM.

The Structural Similarity Index

The Structural SIMilarity (SSIM) index that is typically used as a method for measuring the similarity between two images, based on the degradation of the structural information in one image compared to the reference one. For instance, if we assume that any prior OD matrix or available true OD matrix contains the best pattern information to our knowledge, then the SSIM index can be viewed as an indication of the quality of the estimated OD matrix compared to the prior OD matrix or true OD matrix, respectively. Note that Wang et al. *(14)* introduced the SSIM index in the context of similarity measure to explore and compare the structural information between images, a problem that is in a many respects similar to exploring the structures in OD matrices.

To explain the metric, we follow a similar rationale as in *(14)*. Assume that the OD demand for a particular time interval t is defined in form of the matrix where the rows of matrix represents the origins i, $i=1,2,\ldots I$, and columns represent the destinations j, with $j=1,2,\ldots J$, of trips. To evaluate the structural similarity between two OD matrices, let $d=\{d_n \mid n=1,2,\ldots,N\}$ and $\hat{d}=\{\hat{d}_n \mid n=1,2,\ldots,N\}$ be two vectors that have been extracted from the same spatial location from reference OD matrix $D=\{d_{ij}\}$ and estimated OD matrix $D=\{d_{ij}\}$, as is shown in FIGURE 2. The SSIM index is computed within a local $N\times N$ square box, which moves cell-by-cell over entire OD matrix.

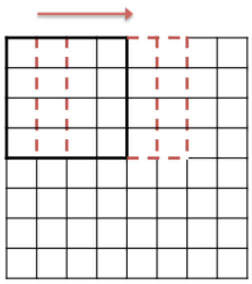


FIGURE 2 Computation of local SSIM index per sliding $N\times N$ square box

The most general form of the metric that is used to measure the structural similarity between two vectors d and $\hat{d}$ consist of *three main components* and is given as

$$SSIM(d,\hat{d})=[l(d,\hat{d})^{\alpha}][c(d,\hat{d})^{\beta}][s(d,\hat{d})^{\gamma}] \tag{1}$$

In this equation l is used as a distance metric to compare the mean values of the two matrices, c compares the standard deviation of the matrices, and finally s compares the matrix structure. Now let us look at each of the components in detail.

As said, the term $l(d,\hat{d})$ compares the mean values of the vectors d and $\hat{d}$, $\mu_d=\frac{1}{N}\sum_{n=1}^{N}d_n$, and is defined by the following expression

$$l(d,\hat{d})=\frac{2\mu_d\mu_{\hat{d}}+C_1}{\mu_d^2+\mu_{\hat{d}}^2+C_1} \tag{2}$$

The term $c(d,\hat{d})$ compares the standard deviation (the square root of variances) of the vectors, $\sigma_d=\sqrt{\frac{1}{N}\sum_{n=1}^{N}(d_n-\mu_d)^2}$, and takes the similar form given by

$$c(d,\hat{d})=\frac{2\sigma_d\sigma_{\hat{d}}+C_2}{\sigma_d^2+\sigma_{\hat{d}}^2+C_2} \tag{3}$$

Finally, the structure term $s(d,\hat{d})$ is defined as the correlation (inner product) between the normalized OD demand vectors d and $\hat{d}$, $d-\mu_d/\sigma_d$ and $\hat{d}-\mu_{\hat{d}}/\sigma_{\hat{d}}$, and is effective measure to quantify the structural similarity. This is equivalent to the correlation coefficient which measures the degree of linear correlation between vectors d and $\hat{d}$. Geometrically, $s(d,\hat{d})$ correspond to the cosine of the angle between two vectors $d-\mu_d$ and $\hat{d}-\mu_{\hat{d}}$, independent of the lengths of these vectors.

Thus, the structure term $s(d,\hat{d})$ is define as follows:

$$s(d,\hat{d})=\frac{\sigma_{d\hat{d}}+C_3}{\sigma_d\sigma_{\hat{d}}+C_3} \tag{4}$$

where $\sigma_{d\hat{d}}=\frac{1}{N-1}\sum_{n=1}^{N}(d_n-\mu_d)(\hat{d}_n-\mu_{\hat{d}})$.

The structure term $s(d,\hat{d})$ reflects the similarity between two OD demand vectors – it equals one if and only if the structures of the two demand vectors being compared are exactly the same.

The constants C_1,C_2,C_3 in Eqn. (2), (3) and (4) are used to stabilize the metric for the case where the means and variances become close to zero. The parameters in Eqn. (1), $\alpha>0,\beta>0$ and $\gamma>0$, are used to adjust the relative importance of the three components. In order to simplify the expression, as is recommended in *(14)* we set $\alpha=\beta=\gamma=1$, and $C_3=C_2/2$. This results in a final form of the SSIM index between two OD matrices

$$SSIM(d,\hat{d})=\frac{(2\mu_d\mu_{\hat{d}}+C_1)(2\sigma_{d\hat{d}}+C_2)}{(\mu_d^2+\mu_{\hat{d}}^2+C_1)(\sigma_d^2+\sigma_{\hat{d}}^2+C_2)} \tag{5}$$

Finally, at each step we calculate the local statistics (μ_d,σ_d and $\sigma_{d\hat{d}}$) and SSIM index within the square box. The overall quality measure of the entire estimated OD matrix is given as a mean of the local SSIM indexes as

$$MSSIM(D,\hat{D})=\frac{1}{M}\sum_{m=1}^{M}SSIM(d_m,\hat{d}_m) \tag{6}$$

where D and $\hat{D}$ are the reference and the estimated OD matrices, respectively, d_m and $\hat{d}_m$ are the OD matrix contents at the m^{th} local square box; and M is the number of local square boxes of the entire OD matrix.

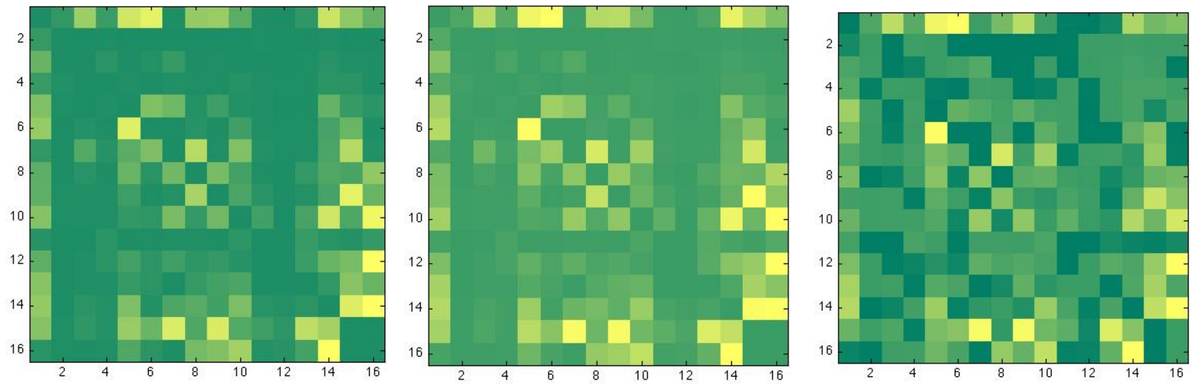
The SSIM index is symmetric: $SSIM(x,y)=SSIM(y,x)$, so that two OD matrices being compared give the same index value regardless of their ordering. It is also bounded: $-1\le SSIM(d,\hat{d})\le 1$, achieving the maximum value $SSIM(d,\hat{d})=1$ if and only if $d=\hat{d}$ and value $SSIM(d,\hat{d})=0$ represent that estimated OD matrix does not capture the spatial correlation between OD pairs as is given in reference OD matrix. Also, the order of origins in rows and destinations in columns in reference and estimated OD matrix must be the same. Otherwise, the SSIM index will give a bias result.

Although the computation of SSIM index seems a more complex then other statistical measures, it is efficient method to assess the quality of estimated OD matrices.

The reliability of the MSE error and SSIM index

To illustrate the properties and the advantages of the SSIM index over statistical measures that are often used in sensitivity analysis of OD estimation methods or in the optimization process, we examine the following several scenarios that reflect the importance of assumptions that an engineer is making when she/he decides to use the MSE. For better visual examination of the structure in the OD matrices, they are represented as images where the values of the OD flows are used as indices into the colormap that determine the color for each OD pair.

In this example, we will show how use of the MSE error is not sufficient for researchers and practitioners to pinpoint the strengths and weakness of different OD estimation methods. For example, we want to assess the quality of estimated OD matrices from two different OD estimation methods (FIGURE 3 (b) and (c)) in respect to the reference OD matrix (e.g. ground truth OD), FIGURE 3(a). In the FIGURE 3, the first perturbed OD matrix (b) was obtained by adding a constant value to all cells in the “ground truth” OD matrix (a) and reflects the estimated OD matrix from one model. The second perturbed OD matrix (c) was generated by the same method, except that the signs of the constant were randomly chosen to be positive or negative, to reflect the estimated OD matrix from model two.

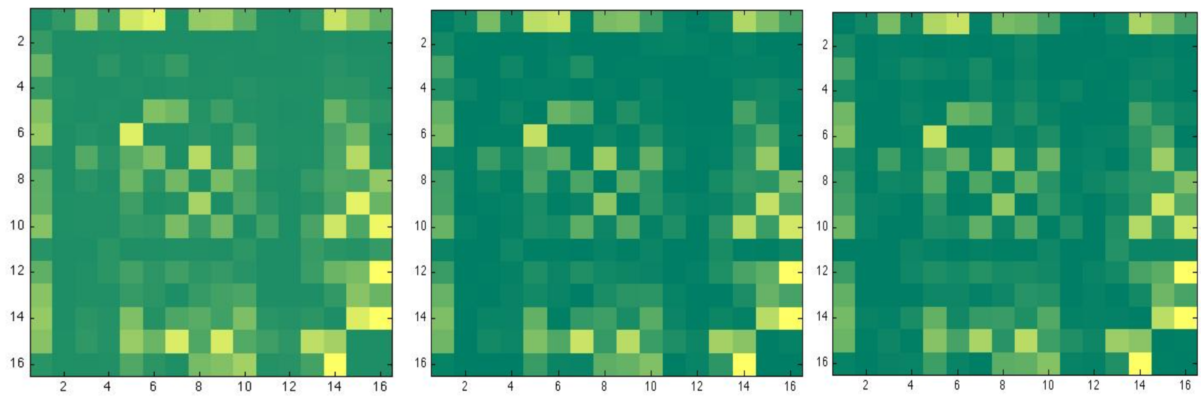


a) b) MSE = 200, SSIM = 0.9405 c) MSE = 200, SSIM = 0.7932

FIGURE 3 Comparison of patterns in reference and estimated OD matrices: a) "ground truth" OD matrix; b and c): estimated OD matrices that have the same MSE with respect to the reference OD matrix, but different structural patterns

The visual difference of the two estimated OD matrices is clearly different. Via visual inspection is it quite clear that the first perturbed matrix resembles the "ground truth" matrix much better than the second one. Yet, the MSE ignores the effect of signs and reports the same value for both perturbed OD matrices while the SSIM index captures the structural difference in the matrices. This result indicates that if we want to compare the performance of the two OD estimation methods and if we look only at the MSE error, we could conclude that both methods perform similarly. Contrary, if we look at the SSIM index values, we can conclude that the method one performs better then other.

Let us consider another example, where the estimated OD matrices have different MSE values but very similar patterns. In the FIGURE 4, the both perturbed OD matrices (b) and (c) were obtained by adding an independent Gaussian noise to the "ground truth" OD matrix (a).



a) b) MSE = 157, SSIM = 0.904 c) MSE = 326, SSIM = 0.904

FIGURE 4 Comparison of patterns in reference and estimated OD matrices: a) "ground truth" OD matrix; b and c): estimated OD matrices that have the same SSIM with respect to the reference OD matrix, but different MSE values.

Apparently, OD matrices that undergo small geometrical modifications have a very large MSE values relative to the "ground truth" OD matrix, yet show a negligible loss of perceived quality. In this case, results indicate that the method with estimated OD matrix (b) performs better then method estimated OD matrix (c), which is consistent with MSE value. Therefore, the SSIM index can be used as additional information (goodness of fit measure) in performance assessment of OD estimation methods.

The examples presented in FIGURE 3 and FIGURE 4, indicate that simplified OD estimation models with the associated assumptions about violated or ignored structural correlation may fail to provide efficient and accurate estimates of OD demand. Cascetta et al. *(15)* have shown that most OD demand estimators can be obtained by solving a constrained optimization problem, where distance functions are defined by considering the distance measures between unknown OD demand and the prior OD demand. These distance measures depend on particular estimation framework, such as maximum likelihood, generalized least square, Bayesian inference, et cetera. For example, the generalized least square function *(2, 15, 16)* is equal to the Mahanalobis distance measure which is often turned into Euclidean distance measure under commonly used simplifying assumptions, such as the variance covariance matrix is equal to the identity matrix or it is diagonal. The models that aim to minimize the Euclidean distance between the prior OD and estimated OD matrix can provide unrealistic estimation results due to fact that too little information is taken from prior OD matrix.

In the next section we will show the application perspectives of the SSIM index as an additional performance measure to statistical measures in benchmark study. In the rest of the paper, we provide insight into how SSIM index can be used further as a new performance function to estimate dynamic OD matrices.

BENCHMARKING FRAMEWORK: PERFORMANCE MEASURES AND CONSIDERED SCENARIOS

In this section, we assess the performance of two OD estimation methods with a least square modeling approach for solving the OD estimation problem. In this study, we consider offline estimation without prediction of the future OD flows. The use of least square approaches to solve dynamic OD estimation problem has been proposed by Cascetta *(15)*. In this study we consider two solution algorithms for the least square problem, Kalman filter algorithm proposed by Ashok and Ben Akiva *(9)* and LSQR algorithm proposed by Bierlaire *(3)* based on proposal of using deviations between historical and actual OD flows as state variables.

Our benchmark framework is as a simulation-based approach where the OD estimator is considered as black box, providing a certain outcome (i.e. the OD matrix estimate) given certain input. We applied the stratified sampling method (Latin hypercube (LHC) method from *(17)* that provides a computationally much more efficient alternative to random (Monte Carlo) sampling for estimating the conditional distribution. A requirement for applying this method is that prior distributions for the inputs need to be known. This feature allows us to perform a comprehensive sensitivity analysis with different assumptions on input data that exhibit certain properties of the OD estimation method. We refer to reader to the original paper on LHC method *(17)* for more details on the approach.

Given such an efficient sampling method exists, we can define a set of input scenarios varying in terms of network topology, traffic conditions, and data availability and quality. First, we will define the set of performance measures to assess the sensitivity of OD estimation methods. Then, in the next subsection we will provide some more detail on the method used to sample the scenarios. For more detail explanation of benchmark framework we refer to the paper *(13)*.

Performance measures

The choice of performance criteria plays an important role in benchmarking OD estimation methods under input uncertainty. From a macroscopic viewpoint, there are two kinds of performance measures that relate to the goodness of fit of the OD estimation algorithm, and to the computational efficiency. In this paper, we are interested in defining a set of statistical measures that allow us to compare the respective performance of the tested OD estimators with our new proposed performance measure, SSIM index. The output will provide us insights into the relative merit of the various performance measures.

The comparison of estimated and true OD demand can be assessed by traditional statistical measures (e.g. error measures such as RMSE, MSE, etc.). To examine the accuracy and robustness of OD estimation methods, we use the following set of well-know performance measures:

- The root mean square error (RMSE) is chosen as an estimate of the variance present in the estimated results (as the average magnitude that the estimate will deviate from the true value). Let $\hat{d}_n$ denote the n^{th} simulation, for $n = 1, 2, \ldots N$, estimated OD matrix for time interval t

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{d}_n - d_n)^2} \qquad (7)$$

Since the scale of OD flows considerably vary, where OD flows with larger volume might dominate comparison, we applied in addition the set of following relative error measures:

- Normalized root mean square error (NRMSE):

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{D}_n - D_n)^2}}{\sum_{n=1}^{N} D_n} \tag{8}$$

- Mean percentage error (MPE):

$$MPE = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\hat{D}_n - D_n}{D_n}\right] \tag{9}$$

The relative measures, which are unit free, eliminate the influence of the input data scale by calculating the error score relative to either the true measurement, or alternatively, the score of another estimation algorithm. In addition, the mean percentage error (MPE) indicates the existence of systematic under or over estimation in the estimated data.

We have seen in the previous section that we can use SSIM index to quantify how well the estimated OD matrices capture the pattern structure of the true OD matrix. Therefore, we include the SSIM index as a quality metric to assess the reliability of OD estimation methods.

Experiment: Synthetic data

In this section numerical experiments are presented to evaluate the performance of the OD estimation methods and solution algorithms in the terms of uncertainty in input data on academic network example and simulated data. First, the OD estimation methods used in the assessment will be described briefly. Then, we will describe the input data and assessment scenarios used in this paper. Finally, we will analyze the performance of OD estimation methods based on defined performance criteria.

Network topology

Prior to methods performance evaluation, we define a simplified synthetic network that consists of 5 nodes, 25 OD pairs with a single route between them and 16 corresponding links (FIGURE 5). This network was chosen because we could model and assume the availability of the “true” OD demand and assignment matrix to the analysts. The “true” assignment matrix is arbitrary derived assuming network is congested.

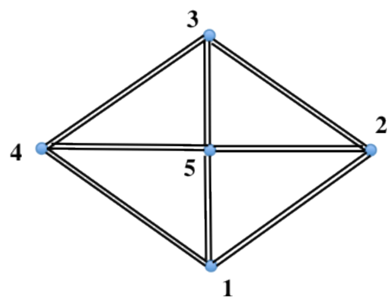


FIGURE 5 The synthetic network

Considered scenarios and results

A major problem with assessing the performance of OD estimators is to obtain meaningful evaluations of the algorithms results and performance, because the ground truth OD data are generally not available for comparison when working with real data. One solution is to use simulated OD demand data, where underlying sources and phenomena are known. To generate a simulated a priory OD matrix dataset for this purpose requires us to define an arbitrary model for OD demand generation, which represents a common spatial and temporal behavior of travelers.

Scenario 1:

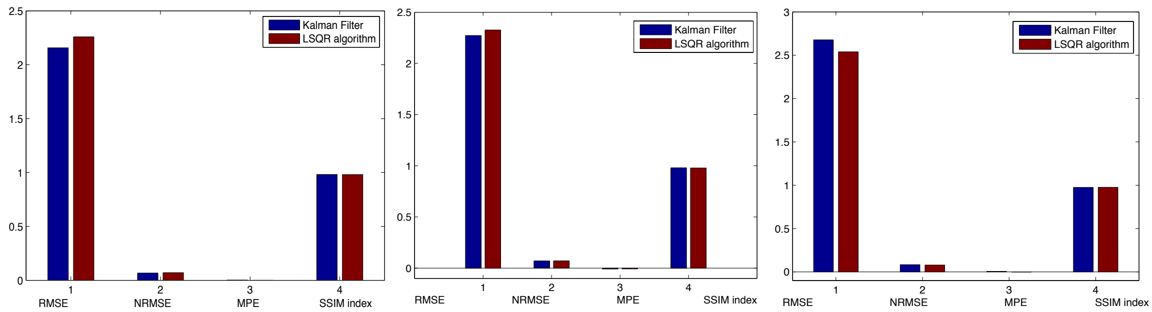
By implementing the LHC method described in previous section, we can simulate the fact that the dynamic prior OD matrix denoted as $\tilde{d}$ with elements $\tilde{d}_{i,t}$ may contain errors. This scenario is based on the assumption that the prior OD matrix is the best estimate of the mean of the dynamic OD matrices. In this case $\tilde{d}$ is varied by adding uniformly random components to the ground truth OD matrix, with standard deviation of 20% representing the difference between the smoothed historical estimate and the particular daily realization:

$$\tilde{d}_{i,t} = d_{i,t}[0.8 + 0.4u_{i,t}] \tag{9}$$

where $u_{i,t} \sim U\left[0,1\right]$, and the $d_{i,t}$ the assumed ground truth OD demand. Next to the prior OD, the available traffic data presents an important source of information. We assume that the true traffic counts $c_{l,t}$ resulting from the assignment of the true OD demand are available on all detectors and has been randomly perturbed to obtain the traffic count data. In this case $c_{l,t}$ is varied by adding uniformly random components to the true traffic counts, with standard deviation of 5%, 10% and 20%:

$$\begin{aligned} \tilde{c}_{l,t} &= c_{l,t}[(1-\delta) + 2u_{l,t}\delta] \\ u_{l,t} &\sim U[0,1]; \delta : [0.05, 0.10, 0.20] \end{aligned} \tag{10}$$

The results for Scenario 1 are presented in FIGURE 6. We can observe that the prior OD matrices with randomly distributed perturbations are improved by a fairly consistent percentage. This implies that the estimation accuracy that can be obtained is directly proportional to the random error in the prior OD matrix. Also, we can see from MPE value that both methods slightly systematically underestimate the OD demand. However, the high SSIM index values ($\text{SSIM}_{\text{KF}} = 0.9816$ and $\text{SSIM}_{\text{LSQR}} = 0.9803$) indicate that structural patterns in estimated OD matrices are preserved and very close to true OD matrix.



a) b) c)

FIGURE 6 Estimation results for random prior OD matrix with: a) traffic counts error $\sigma 2\%$; b) traffic counts error $\sigma 5\%$; and c) traffic counts error $\sigma 10\%$

Scenario 2:
This scenario addresses situations where the prior OD demand might contain other, *structural* errors besides the random daily fluctuations. The demand per each origin over destinations is generated from positively and negatively skewed mean values of distribution from random demand scenario defined in two prior OD data sets:

$$\tilde{d}_{i,t} = d_{i,t}[0.9 + 0.4u_{i,t}] \tag{11}$$

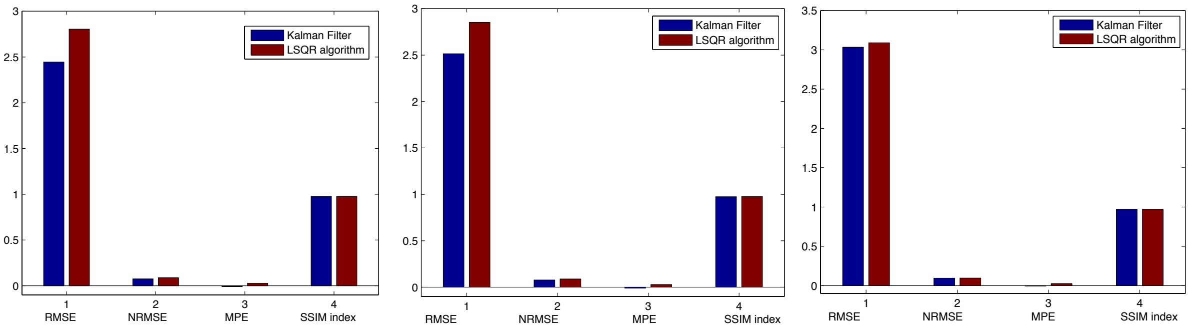
and

$$\tilde{d}_{i,t} = d_{i,t}[0.7 + 0.4u_{i,t}] \tag{12}$$

where $u_{i,t} \sim U\left[0,1\right]$, and the $d_{i,t}$ the assumed ground truth OD demand. The Eqn. (12) reflects the overestimated prior OD matrix in total demand compare to the true OD matrix, while Eqn. (13) represents the underestimated prior OD matrix. In this scenario we used the same traffic counts scenarios as defined in Scenario 1.

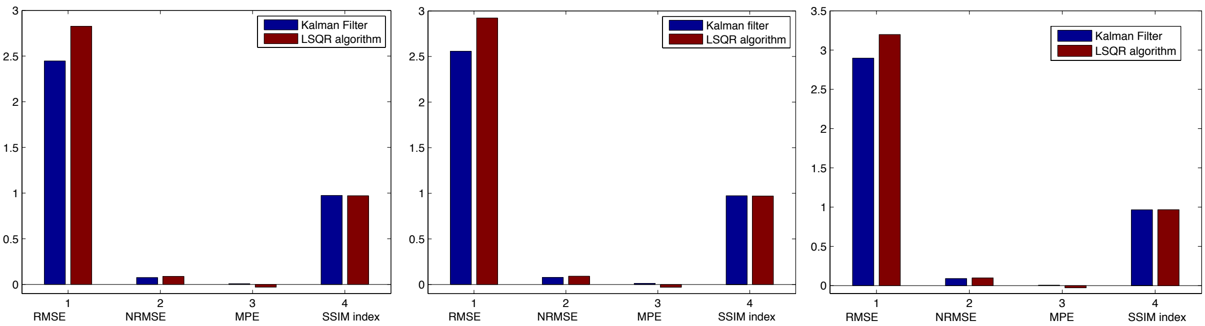
The results for Scenario 2 are presented in FIGURE 7 and FIGURE 8. The estimated OD matrices for the overestimated and underestimated prior OD matrix result in a slightly higher RMSE values, where the Kalman Filer solution algorithm for least square model formulation shows a better performance then LSQR algorithm. The improvement of the Kalman Filter algorithm can be explained by the structure of the used minimization algorithm. We can also observe that both methods show more sensitivity with the increase of deviations in traffic counts. Also, we can see from MPE value that both methods slightly systematically underestimate the OD demand. The SSIM index values ($\text{SSIM}_{\text{KF}} = 0.9671$ and $\text{SSIM}_{\text{LSQR}} = 0.9605$) are slightly lower then in Scenario

1, yet they still indicate that structural patterns in estimated OD matrices are preserved and very close to true OD matrix. Note that next to the RMSE-type measures, this provides us additional insight into this importation characteristic of the estimated OD matrix.



a) b) c)

FIGURE 7 Estimation results for overestimated prior OD matrix with: a) traffic counts error $\sigma 2\%$; b) traffic counts error $\sigma 5\%$; and c) traffic counts error $\sigma 10\%$



a) b) c)

FIGURE 8 Estimation results for underestimated prior OD matrix with: a) traffic counts error $\sigma 2\%$; b) traffic counts error $\sigma 5\%$; and c) traffic counts error $\sigma 10\%$

Although, we benchmark the performance of two OD estimation methods in a simple network example, we can observe that SSIM index provide more insight into the quality of estimation results. For more involved cases, with more complex networks and error structures, we expect the differences between the different measures to become more considerable.

TOWARDS A NEW OD ESTIMATION FRAMEWORK

In FIGURE 3 and FIGURE 4, we have seen that Euclidean distance measure as a performance function is only designed to find the distances between a pair of attributes in a data set or overall distance among all data. They are unable to find and distinguish different correlation structures in data, and cannot provide an in depth explanation of the patterns in OD demand.

One possible solution approach to tackle this problem is to use a performance function that will incorporate the structural correlation between OD pairs and that allows the modeler to control the trade off between simplicity of the model and the level of realism. In this section we propose such a framework that incorporate the structural similarity index in performance function as a penalty factor. Since the penalty factor has to be applicable to a whole range of OD estimation methods, the workings of which are not explicitly considered, the new performance function can be interpreted as consisting of two elements:

$$PerformanceFunction = Best\_fit \times SSIM\_index \tag{13}$$

A higher best fit likelihood leads to optimal OD matrix solution that can explain and fit the available data well, i.e. that have a low estimation error $\sum (d_{est} - d_{prior})^2$. However, if only this measure would be investigated the overfitting problem would occur as when the estimation error is used for prediction of OD matrices. Therefore, the models performance is penalized by the SSIM index, which always takes a value between -1 and 1. The estimated OD matrix that has a significantly different structure then the prior OD matrix has a lower SSIM index and therefore receives lower performance value.

The formulation of the new framework for the estimation (and assessment of the prediction) of OD demand given in Eqn. (13) has several advantages over existing OD estimation and prediction methods: (1) the most important feature is that it introduces additional information in the estimation process on the basis of structural patterns in the OD matrix. This allows in estimation process to select most probably best estimated OD matrix, taking into account both the estimation error as well as the structural similarity between OD matrices; (2) the structural OD pattern information is included when estimating the OD demand to rule out unrealistic estimation results due to the fact that too little information is taken from prior OD demand; (3) the approach can be used to combine with a weighted performance function, where the weighting value is determined by the corresponding covariance matrices representing the dispersion of collected data or by the analyst's relative confidence in either input information.

Note that the purpose of this section is to demonstrate the potential application of SSIM index and main features of the novel performance function in estimation and prediction of OD matrices. The presented framework is still academic in nature and must be interpreted as concept idea. In the future work we will rigorously derive the new performance function mathematically to ascertain that the method performs well in practice.

CONCLUSIONS AND FUTURE WORK

This paper discusses the potential of using the structural similarity (SSIM) index as a quality measure that quantifies the similarity between two OD matrices (e.g. between an OD matrix estimate and a reference OD, such as the prior OD matrix, or ground truth OD matrix). The most important feature of the new metric is that it includes additional information in evaluation process on the basis of the structural patterns of OD matrices. We showed that this quality metric has several advantages over existing statistical measures, such as the Mean Squared Error MSE. As an example, it turns out to be more sensitive to capture the structural correlation between OD pairs; it ignores the effect of the signs of the error in estimated OD matrix. Further, we recommend the application of SSIM index as a performance measure in addition to existing statistical measures in benchmarking studies.

Since our final objective was to show potential application of SSIM index as a new performance function, we presented a new framework for the estimation of OD demand. This allows in estimation process to select the optimal OD matrix, taking into account both the estimation error as well as the structural similarity between OD matrix to be estimated and prior OD matrix. The presented framework is still theoretical in nature and must be interpreted as such. In particular the mathematical characteristics of the measure needs to be investigated further (e.g. non-convexity) and its implications for estimation need to be formulated and if need be adapted. More results in more realistic settings will be obtained in future research to ascertain that the method performs well in practice.

ACKNOWLEDGEMENTS

This research is partly funded by the ITS Edulab, a collaboration between TUDelft and Rijkswaterstaat.

REFERENCES

1. Bierlaire, M., *The total demand scale: a new measure of quality for static and dynamic origin-destination trip tables.* Transportation Research Part B: Methodological, 2002. **36**(9): p. 837-850.
2. Bell, M.G.H., *The estimation of origin-destination matrices by constrained generalised least squares.* Transportation Research Part B: Methodological, 1991. **25**(1): p. 13-22.
3. Bierlaire, M. and F. Crittin, *An Efficient Algorithm for Real-Time Estimation and Prediction of Dynamic OD Tables.* OPERATIONS RESEARCH, 2004. **52**(1): p. 116-127.
4. Van Zuylen, H.J. and L.G. Willumsen, *The most likely trip matrix estimated from traffic counts.* Transportation Research Part B: Methodological, 1980. **14**(3): p. 281-293.
5. Wu, J., *A real-time origin-destination matrix updating algorithm for on-line applications.* Transportation Research Part B: Methodological, 1997. **31**(5): p. 381-396.
6. Watling, D.P., *Maximum likelihood estimation of an origin-destination matrix from a partial registration plate survey.* Transportation Research Part B: Methodological, 1994. **28**(4): p. 289-314.
7. Hazelton, M.L., *Inference for origin-destination matrices: estimation, prediction and reconstruction.* Transportation Research Part B: Methodological, 2001. **35**(7): p. 667-676.

8. Van Der Zijpp, N., *Dynamic Origin-Destination Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data.* Transportation Research Record: Journal of the Transportation Research Board, 1997. **1607**(-1): p. 87-94.
9. Ashok, K., M.E. Ben-Akiva, and T. Massachusetts Institute of, *Dynamic origin-destination matrix estimation and prediction for real-time traffic management systems.* Transportation and traffic theory., 1993.
10. Ashok, K. and M.E. Ben-Akiva, *Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows.* TRANSPORTATION SCIENCE, 2000. **34**(1): p. 21-36.
11. Yang, H., Y. Iida, and T. Sasaki, *An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts.* Transportation Research Part B: Methodological, 1991. **25**(5): p. 351-363.
12. Marzano, V., A. Papola, and F. Simonelli, *Limits and perspectives of effective O-D matrix correction using traffic counts.* Transportation Research Part C: Emerging Technologies, 2009. **17**(2): p. 120-132.
13. Djukic, T., J. van Lint, and S. Hoogendoorn, *Efficient Methodology for Benchmarking Dynamic Origin-Destination Demand Estimation Methods.* Transportation Research Record: Journal of the Transportation Research Board, 2011. **2263**(-1): p. 35-44.
14. Zhou, W., A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, *Image quality assessment: from error visibility to structural similarity.* Image Processing, IEEE Transactions on, 2004. **13**(4): p. 600-612.
15. Cascetta, E., D. Inaudi, and G. Marquis, *Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts.* TRANSPORTATION SCIENCE, 1993. **27**(4): p. 363-373.
16. Antoniou, C., M. Ben-Akiva, and H.N. Koutsopoulos, *Dynamic traffic demand prediction using conventional and emerging data sources.* Intelligent Transport Systems, IEE Proceedings, 2006. **153**(1): p. 97-104.
17. McKay, M.D., R.J. Beckman, and W.J. Conover, *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code.* Technometrics, 1979. **21**(2): p. 239-245.