

Minimising GEH in Static OD estimation

Aleix Ruiz de Villa¹, Jordi Casas¹, Martijn Breen¹, Josep Perarnau¹

¹TSS – Transport Simulation Systems

aleix.ruizdevilla@aimsun.com

Abstract

One of the essential quantities in macro traffic simulation is the ‘who goes where’ OD matrix. This is usually estimated through an optimization problem of minimizing the squared differences between observed and simulated traffic counts. The choice of the mean square error loss is due to its good statistical properties and the fact that the related quadratic optimization problem is easy to deal with. However, in many applications following the recommendations of validation criteria suggested in many guidelines published by different road administrations, such as FHWA in USA, ARRB in Australia or Highways Agency in UK, the quality of a static OD adjustment is assessed with other quantities such as the relative error or the GEH index.

A direct minimization of the GEH index involves technical difficulties. However, recent developments in optimization theory give the opportunity of solving such problem using gradient descent techniques. We apply this method to some static OD estimation problems in real traffic networks. Empirical evidence show that GEH minimization converges slower than least squares. On the other hand the former is more robust against measurement errors on data.

This paper show numerical results of real networks, comparing the OD estimation using least squares and GEH index, concluding with the benefit of including this optimization method for static OD estimation.

1. Introduction

The static Origin-Destination (OD) estimation or adjustment matrix is a classical problem applied and solved in Transport Planning. This class of problems can take on many different formats, depending upon what is considered to be known and what assumptions are made to derive the missing parameters. The description of generic problem is described below.

Given an OD matrix Q , denote by v_a the volume on a link a as a result of a particular traffic assignment with OD matrix Q . In practice, we are often interested in the reverse problem, that is, given a set of links $(a)_{a \in A}$ and observed volumes in such set $(v_a^o)_{a \in A}$, find an OD matrix Q such that the assigned volumes v_a are as close as possible to the measured ones. Usually an initial matrix Q_0 , regarding prior information about traffic behavior (obtained for instance by a survey), is required and the objective is to modify Q_0 to fit current data. This process is called matrix adjustment or estimation. When traffic flows are modelled time-independently the matrix adjustment process is called static. Otherwise it is called dynamic. See (Bera & Rao, 2011) and the references therein for a summary of the main techniques used to approach this problem. In this paper, we will focus on bi-level static based techniques where the optimization problem has two levels: the inner or lowest is, for a given OD matrix, the traffic assignment, and the outer or upper is the minimization with respect to OD matrices of the errors between observed and assigned flows. Spiess (Spiess, 1990) solved this problem using a gradient based approach with minor modifications to ensure preservation of unused paths. In (Yang et al, 1994) and (Yang, 1994) heuristic approaches were proposed to deal with the bi-level mathematical programming. A Gauss-Seidel type coordinate decent approach can be found in (Florian & Chen, 1995).

Traditionally square error loss has been chosen to measure disagreement between assigned and measured volumes. The reason is that it is a widely studied loss (with a central role in

statistics) and it is easy to use (differentiable and easy to calculate). Thus, the matrix adjustment problem can be stated as

$$\min_Q \sum_{a \in A} (v_a - v_a^o)^2. \quad (1)$$

Another relevant quantity in transportation is the so called GEH. The GEH of two traffic volumes, usually measured and predicted or simulated nonnegative data, is defined as

$$Geh(y, \hat{y}) = \sqrt{\frac{2(y - \hat{y})^2}{y + \hat{y}}} = \frac{|y - \hat{y}|}{\sqrt{(y + \hat{y})/2}}$$

The GEH formula gets its name from Geoffrey E. Havers (UK Highways Agency, 1992). Although its mathematical form is similar to a chi-squared test, is not a true statistical test. Rather, it is an empirical formula that has proven useful for a variety of traffic analysis purposes. The main trait of the GEH formula is that, unlike the relative error, differences in high volumes are relatively more important than in lower volumes. This index is commonly used as validation criteria in many applications following the recommendations suggested in many guidelines published by different road administrations, such as FHWA in USA, ARRB in Australia or Highways Agency in UK. For traffic modelling the different guidelines has the following consensus: a GEH index of less than 5.0 in a measurement point is considered a good match between the modelled and observed volumes, and 85% of the volumes in a traffic model should have a GEH less than 5.0 for all measurement points.

Analogously to problem (1), one can consider the following optimization problem

$$\min_Q \sum_{a \in A} Geh(v_a, v_a^o).$$

Such a problem can become difficult because both the numerator and denominator depend on Q . For this reason we consider the approximation

$$\min_Q \sum_{a \in A} \frac{|v_a - v_a^o|}{\sqrt{v_a^o}},$$

provided that $\frac{v_a + v_a^o}{2} \approx v_a$.

Another interesting optimization problem would be to find an OD matrix such that the 85% of the assigned volumes have a GEH less than 5.0. This could be solved by minimizing directly the number of detectors that have a geh greater than 5. However, this problem turns out to be a much more difficult problem, since the loss function is not even continuous. Thus, the introduction of techniques for minimizing the sum of gehs can be thought as a first step towards the solution of such difficult problem.

In addition to the square error loss and the geh, another typical quantity is the relative error, yielding to the optimization problem

$$\min_Q \sum_{a \in A} \frac{|v_a - v_a^o|}{v_a^o}.$$

The aim of this paper is to study differences in OD estimation arising from the loss choice. For completeness (and symmetry), we will include three loss functions more. The set of studied losses and their related optimization problems in this paper becomes then:

- Square error (which will be denoted by L2),

$$\min_Q \sum_{a \in A} (v_a - v_a^o)^2.$$

- Absolute deviation error,

Minimising GEH in Static OD estimation

$$\min_Q \sum_{a \in A} |v_a - v_a^o|.$$

- Weighted square error with exponent one (which will be denoted by L2₁),

$$\min_Q \sum_{a \in A} \left(\frac{v_a - v_a^o}{\sqrt{v_a^o}} \right)^2 = \min_Q \sum_{a \in A} \frac{1}{v_a^o} (v_a - v_a^o)^2.$$

- Geh (with some abuse of notation),

$$\min_Q \sum_{a \in A} \frac{|v_a - v_a^o|}{\sqrt{v_a^o}}.$$

- Weighted square error with exponent two (which will be denoted by L2₂),

$$\min_Q \sum_{a \in A} \left(\frac{v_a - v_a^o}{v_a^o} \right)^2 = \min_Q \sum_{a \in A} \frac{1}{(v_a^o)^2} (v_a - v_a^o)^2.$$

- Relative error,

$$\min_Q \sum_{a \in A} \frac{|v_a - v_a^o|}{v_a^o}.$$

The use of the absolute deviation loss

$$g(v_a, v_a^o) = |v_a - v_a^o|,$$

brings technical difficulties. Namely, it is not differentiable, and thus classical gradient descent techniques cannot be directly implemented. Fitting linear models using absolute deviation loss, called median regression, is a well researched area. See (Koenker, 2005) for a comprehensive treatment on the subject. It is well known that one of the strengths of using absolute value loss is that the resulting models are more robust against outliers. Usually median regression models are solved transforming the problem into a linear programming one and applying techniques such as the simplex or interior point methods. Instead, in this paper, we follow a more flexible approach based on Staines and Barber (2012). First we find differentiable functions $g_\sigma \rightarrow g$ as $\sigma \rightarrow 0$. Then, for a selected sequence σ_n , we find Q_n minimizing

$$\min_Q \sum_{a \in A} g_{\sigma_n}(v_a, v_a^o), \quad (2)$$

using classical gradient descent techniques. In this way, the resulting Q_n approximates the optimal Q^* .

We have to make a distinction here, between the objective loss, which we will call *evaluation loss*, and the loss used in our optimization method, which will be denoted *method loss*. Our main interest should rely on optimizing the evaluation loss, since it contains the way we choose to measure the performance of our estimations. A priori, no loss is better to any other. So the choice is rather subjective. One may think that the chosen evaluation loss represents our interests the best possible way. On the other hand, the method is just the technique used to achieve such goal. Following this principle, one should choose a loss first for its *meaning* and after that look for a way of optimizing it. This has not always been the case with the L2 loss: often it is used because of its good optimization properties, instead of its implications on further decisions in traffic modelling.

It is natural to use the same evaluation and method losses, expecting that other methods will be less efficient. However this is not always the case, as we will see in the results section

with the absolute error. If one does not have a clear preference for a particular loss, he or she should take a method that has a good behaviour among several evaluation losses.

2. Algorithm

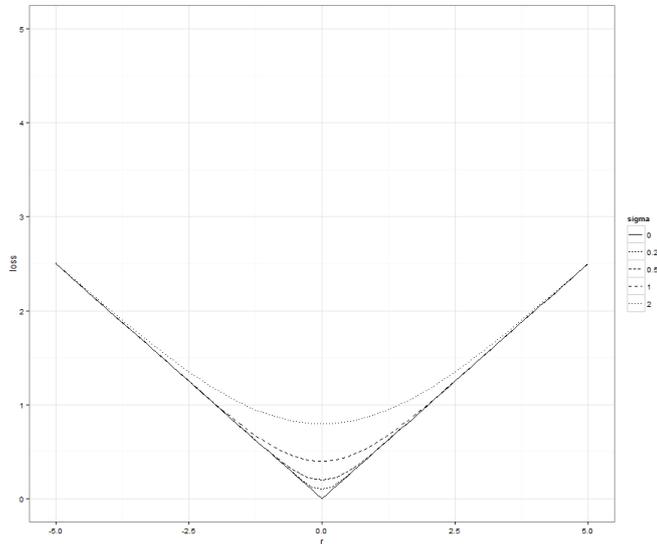
The algorithms used in this paper are based on Spiess (1990), which describes an optimization algorithm for estimating OD matrices using the L2 loss. Since the $L2_1$ and $L2_2$ cases are weighted versions, they can be easily implemented with minor modifications. Spiess's method is gradient descent type algorithm. It assumes that path flows are locally constant to find an approximate gradient at each step. In addition, he finds a closed formula to approximate the optimal step size.

Since the loss function $g(v_a, v_a^o) = |v_a - v_a^o|$ is nondifferentiable, we cannot directly calculate the gradient. Instead, in order to apply the approximation scheme described above, we find the functions g_σ following Staines and Barber (2012): we consider the convolution of the absolute error with the gaussian kernel $\eta_\sigma(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}}$,

$$g_\sigma(r) = g * \eta_\sigma(r) = \frac{1}{\sqrt{2\pi}\sigma} \int |s| e^{-\frac{(s-r)^2}{2\sigma^2}} ds = \frac{2\sigma}{\sqrt{2\pi}} e^{-\frac{r^2}{2\sigma^2}} + r \left(1 - 2\Phi\left(-\frac{r}{\sigma}\right) \right),$$

where Φ is the cumulative distribution function of the standard Gaussian kernel $\eta = \eta_1$. When σ is small, the function $g_\sigma(r)$ at point r can be regarded as a weighted average of the absolute value function with the points near r . It is a calculus exercise to see that $g_\sigma(r) \rightarrow g(r)$ uniformly as $\sigma \rightarrow 0$. Moreover, the functions g_σ are infinitely differentiable. In Figure 1 we show a sequence of approximations with $\sigma \in \{2, 1, 0.5, 0.25, 0\}$.

Figure 1 Functions g_σ



Now, it is easy to modify Spiess's algorithm so as to minimize problem (2). Analogously to the square error case, the relative error and (simplified) geh losses can be easily implemented via a weighted version of the algorithm. The interested reader can find the details in Ruiz de Villa et al (2013).

Figure 2 Test network with detectors

Minimising GEH in Static OD estimation



3. Results

We tested our method on a network of a medium sized city of Spain. The implementation has been done in a combination of Aimsun (2013) and R (2012).

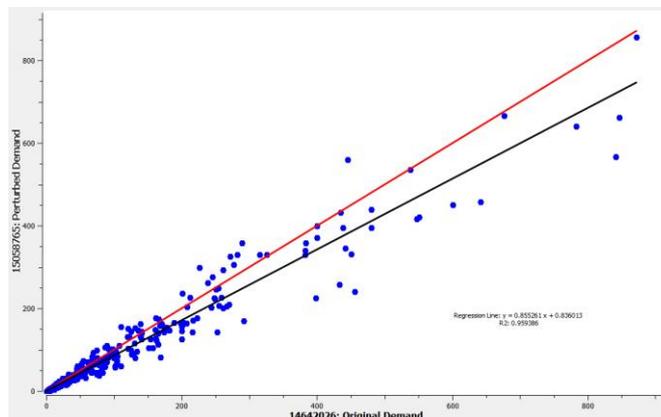
3.1 Network

The network that has been used to perform the experiments is a medium urban network with a highway stretch north of the city centre. The network comprises a total road length of 631 km and the different zones are modelled with 57 centroids. The network detection consists of 362 count detectors which are located both in the urban centre as on the highway stretch. In the Figure 2 the distribution of the detectors is shown in which the green dots represent the count detectors.

The assignment period is the evening peak hour in which a total of 35000 cars are assigned to the road network. With this demand a synthetic dataset has been created which will be used to validate the O/D estimation procedure.

The original demand matrix is perturbed in order to be used as a starting point for the O/D estimation procedure. Each cell is multiplied by a random normal distribution with mean 0.9 and standard deviation 0.2. In the Figure 3 the trip values of the original matrix are plotted against the perturbed trip values to show the degree of perturbation of the original matrix.

Figure 3 Perturbed cells



The assignment method that has been used for this experiment is the Frank-Wolfe algorithm, which is available in the traffic simulation package Aimsun (2013). The network achieves an

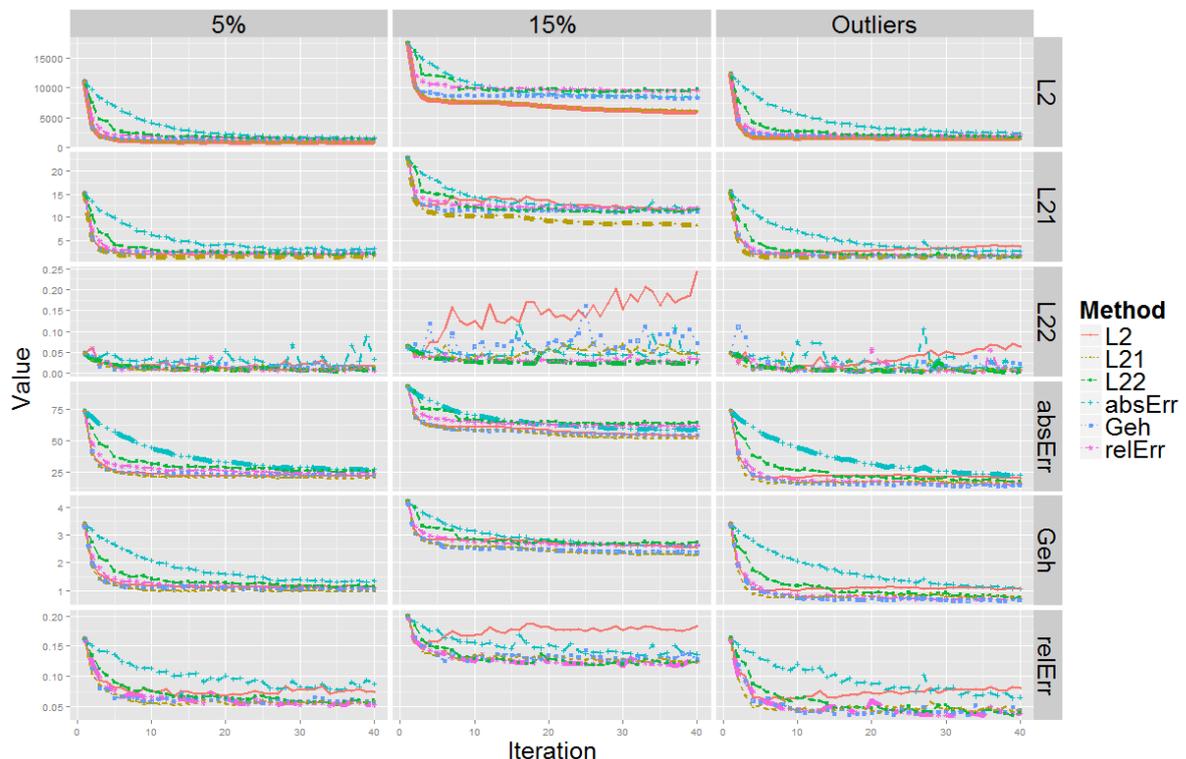
equilibrium status with a relative gap below 1% after 50. From the assignment result the path proportions are extracted to be used in the O/D estimation.

3.2 Experiments

Heuristically we found that often, with the absolute deviation loss and its weighted versions, the proposed direction by the algorithm was not a real descent direction, since it led to bigger losses. In this case we applied a naïve line search: we tempted smaller step sizes (up to 20 equally distributed values between 0 and the proposed step size) to find a reliable descent direction. In order to compare the performance of the proposed method with the least squares estimation of Spiess (1990), we also applied this strategy to the L2 and related losses.

In the following figures we distinguish between the optimization method, called *method*, and the evaluated loss or objective function, called *evaluation*. Two types of perturbations have been considered. On the one hand, in applications, since we do not know the OD matrix that generates the observed data, an estimated initial OD matrix is obtained from external sources, such as a survey. Thus, the original matrix in our experiments has been perturbed as explained above with some random noise. On the other hand, usually observed data is perturbed due to measurement errors and many other causes. For this reason we have also noising the assigned volumes from the original matrix in two ways. In the first one, all of the assigned volumes from the original matrix have been perturbed with random gaussian distributions with factors of 5% and 15%. The second set of experiments recreates the effect of outliers: the volume of four relevant detectors (with high flow) has been increased in a 30%.

Figure 4 Method/Evaluation error comparison



The results are shown in Figure 4. Different line types denote different minimization algorithms. The evaluation loss is displayed in the vertical axis of the panel. The horizontal axis of the panel is used for the type of perturbation. Matching methods and evaluations (for

instance when the L2 loss is used in both the minimization and the evaluation) are marked with thicker lines.

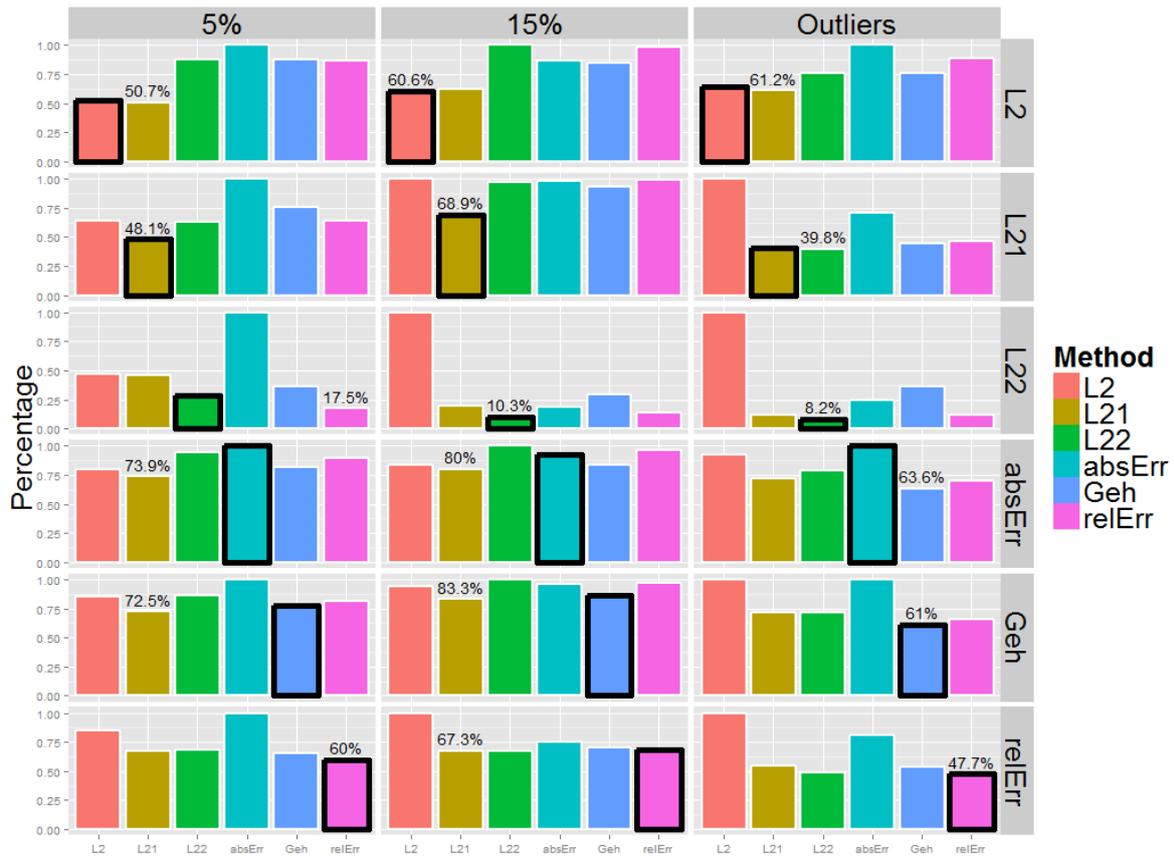
Generally speaking the minimization of a quantity forces the others also to decrease. We see that for each evaluation loss, its corresponding optimization method is among the ones that achieve better results, with the exception of the absolute loss. In this case we attribute the lack of efficiency of the method to its slowness. In fact, in all experiments the absolute error is the slowest one. We also highlight the fact that the L2 loss does not have a monotonically decreasing trajectory if evaluated with the relative error or its squared counterpart, the $L2_2$. It also has a slightly increasing trajectory with the geh and $L2_1$ evaluations.

To have a more quantitative comparison of the different methods, we introduce Figure 5. As seen in Figure 4, different losses lead to different ranges of values, thus the efficiency of the methods is not comparable among evaluation losses. In an attempt to elude this effect, we rescale the results up to the maximum loss achieved, thus showing a percentage with respect to the greatest value. In Figure 5, such percentages are displayed from the last iteration (number 40). The matching methods and evaluations are put into a black frame and the minimum value achieved is displayed above the related method.

The first thing to notice is that the traditional L2 method is not necessarily the best among the tested in this paper. It outperforms the rest of methods if we evaluate also with the L2 loss, but has poor performance within the rest of evaluations. So, unless one has a clear preference for the L2 evaluation, the L2 method should not be an automatic choice. On the other hand, as expected from our previous results, the absolute loss is among the worse methods, having poor performances even if we use the absolute error evaluation.

Among the remaining methods, we see that the L21 method has relevant results in the 5% and 15% sets and good performances in the outliers set. Comparing per blocks, differentiable (L21 and L22) vs no differentiable (geh and relative error), we see that, excluding the L2 evaluation, both have parallel results. It is also remarkable that the efficiency of the no differentiable block improves slightly in the presence of outliers: all the minimums with the absolute error, geh and relative errors belong to the no differentiable block and in the L21 case there is a significant decrease of errors with respect to the 5% and 15% sets. As presented at the introduction, in statistics it is well known that the absolute error loss is more robust against outliers than the L2 loss. We thus regard our results as a confirmation of this effect.

Figure 5 Last iteration error comparison



Now, we analyze the differences of the methods when evaluated with the denoted “#geh<5” function, that is, the number of detectors with geh less than 5. As explained in the introduction, it is commonly requested that 85% of detectors have a geh less than 5. In Figure 6, we display the evolution of the #geh<5. In Figure 7 we show the percentage of detectors having a geh less than five at the last iteration. The main differences arise in the 15% set, where the L21 outperforms the rest of methods. Instead, in the outliers set, differences among sets decrease. In Figure 8, we display the minimum number of iterations needed to achieve that the percentage of detectors with a geh less than 5 is greater than 85%. The methods with 40 iterations in the graph are the ones that did not achieve such result.

Minimising GEH in Static OD estimation

Figure 6 #geh<5 error comparison

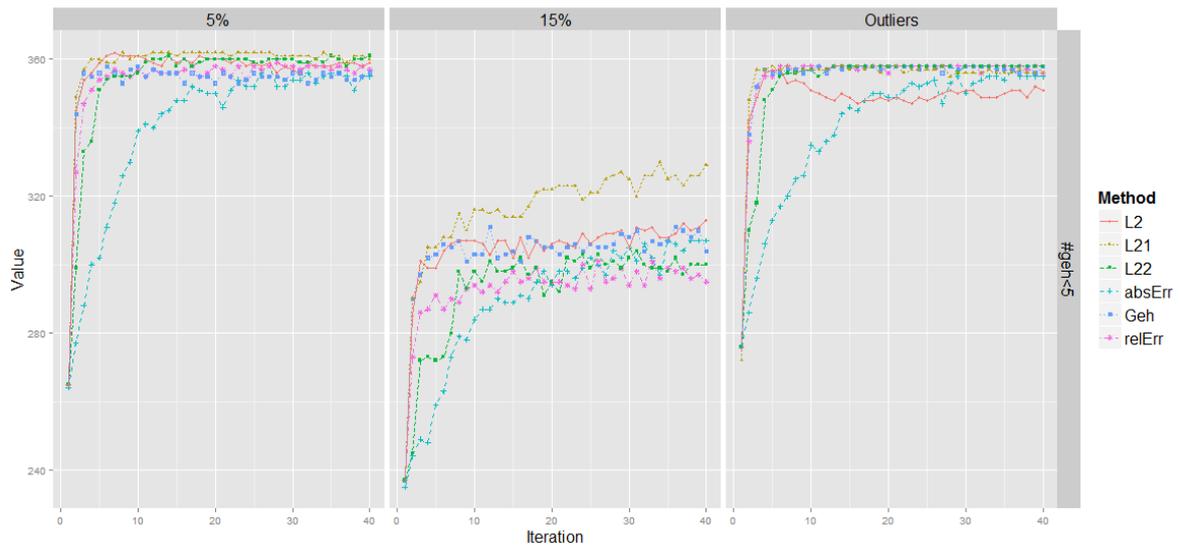


Figure 7 geh<5% last iteration

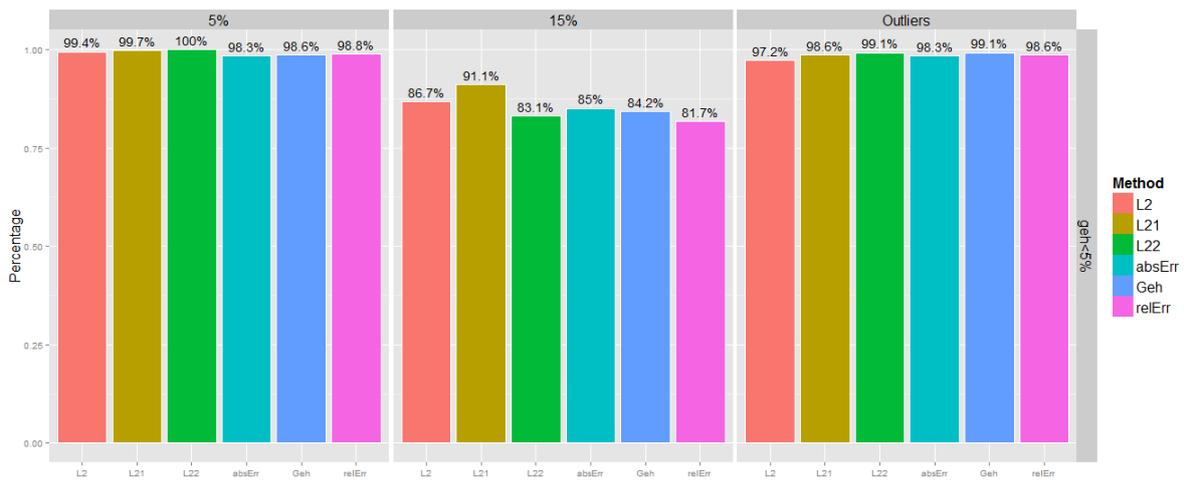
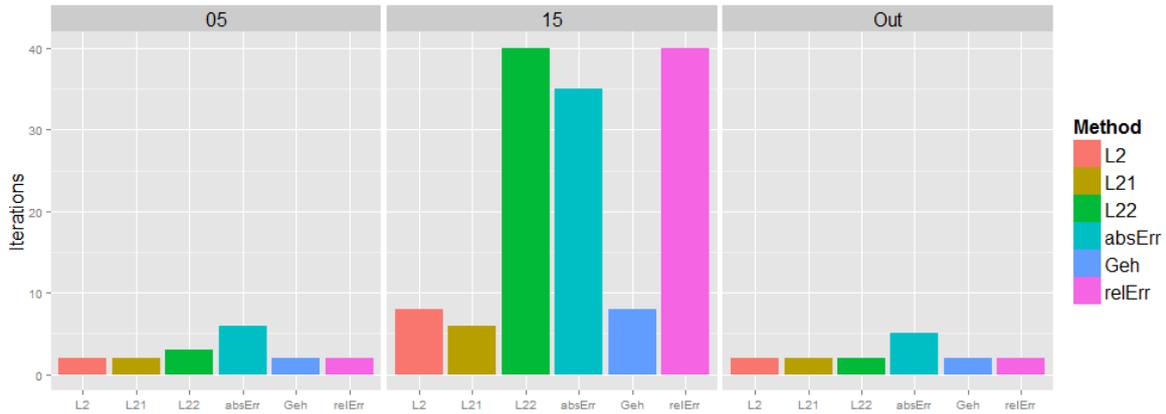
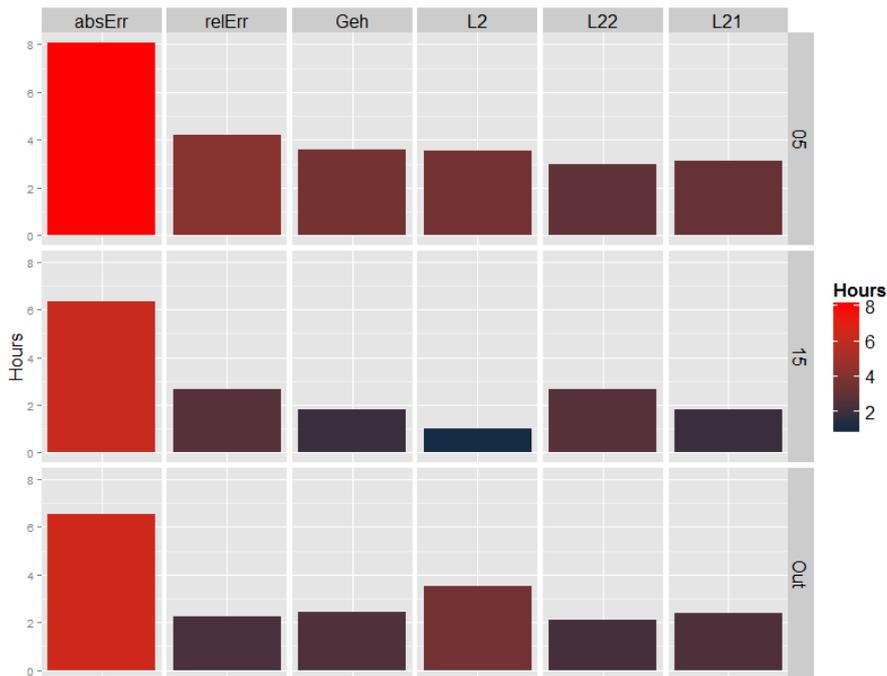


Figure 8 Minimum number of iterations



Finally we show the computational times for each method. Recall that we applied a line search strategy, thus the total number of iterations are not the ones showed in the graphs. This also leads to different computational times between methods. Clearly the absolute error needs more computational effort. Notice that, on average, the 15% set is the fastest set. We attribute this to the fact that since the two others was previously more optimized (with less initial error), it was harder to find reliable descent directions.

Figure 9 Computational Time



4. Conclusions

The main objective of this paper is to provide an alternative to the traditional mean square error in static OD estimation problems. This is done by introducing the absolute error loss and some variations of it. We give the technical tools to deal with such problems in a flexible way.

We have compared through empirical studies six different methods and the effect they have on the widely used “geh<5%” index. We have compared the methods using two types of perturbations, gaussian and with outliers, and we have to conclude that losses involving the absolute deviation loss have a relevant place in the presence of outliers. The authors suspect that more differences could arise if other types of perturbations were introduced, namely higher levels of outliers or nonsymmetric perturbations. These could be subject for further studies. Finally, we have seen that using the so called L21 loss has generally shown better results than the others, while the presented algorithm for the absolute error loss is by far the slowest.

5. References

Aimsun Version 8.0 (2013). *User's Manual*. TSS-Transport Simulation Systems.

Bera, S. & Rao, K. (2011) Estimation of origin-destination matrix from traffic counts: the state of the art. *European Transport \ Trasporti Europei*, 49, 3-2.

Florian, M. & Chen, Y. (1995). A coordinate descent method for the bi-level od matrix adjustment problem. *Int. Trans. Opt. Res.*, vol. 2, n. 2, 165-179.

Koenker, R. (2005) *Quantile regression*. Econometric Society Monographs, 38, Cambridge University Press.

R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Ruiz de Villa, A., Casas, J., Breen, M. and Perarnau, J. (2013) Static OD estimation minimizing the relative error and the GEH index. *Preprint*.

Spiess, H. (1990). A gradient approach for the o-d matrix adjustment problem. *Centre de Recherche sur les Transports de Montréal*, 69.

Staines, J. & Barber, D., (2012) Variational Optimization. *Technical report*. <http://arxiv.org/pdf/1212.4507v2.pdf>.

UK Highways Agency (1992). *The Design Manual for Roads and Bridges*, volume 12.

Yang, H. (1994). Heuristic algorithms for the bi-level origin-destination matrix estimation problem. *Transportation Research, Part B: Methodological* 2, 231-242.

Yang, H., Sasaki, T., Iida, Y. & Asakura, Y. (1992) Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research, Part B: Methodological* 2, 417-433.