International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Investigating the performance of SPSA in Simulation-Optimization approaches to transportation problems

Xavier Ros-Roca[a,b], Lidia Montero[b], Arne Schneck[a], Jaume Barceló[a,*]

[a]*PTV Group, Karlsruhe 76131, Germany*
[b]*Universitat Politècnica de Catalunya, Barcelona 08034, Spain*

Abstract

While optimization models play a key role in transportation analysis, the objective function to be optimized, however, cannot be defined analytically. It is therefore necessary to resort to non-differentiable optimization methods that usually pivot on evaluating the objective function. Special cases of particular interest are those in Dynamic OD Estimation, which cannot evaluate the objective function analytically and thus the formulation falls in the computational framework of Simulation-Optimization. SPSA is not limited to the inputs from conventional traffic counts and can be easily extended to account for the measurements of traffic variables supplied by emerging sensors exploiting Information and Communication Technologies (ICT). Numerical experiments have been conducted, and the results have been analyzed from two different perspectives: performance and solution quality. This allows understanding the behavior of the SPSA algorithm and new variants, which altogether contribute to the aim of adding ICT measures in the future. Their sensitivity to the initial values, the effect of bounding the variables and scaling techniques are analyzed. This paper will report on the results of the numerical experiments, their analyses, conclusions and further research.

Keywords: dynamic OD matrix estimation; SPSA; Dynamic User Equilibrium; sensitivity analysis

1. Introduction

The bi-level optimization approach is usually considered to be the most convenient formulation adjustment and means of correcting OD matrices by combining available sources of data, because it explicitly takes into account the congestion effects that influence the use of paths between OD pairs. The most common formulation considers that the available sources of data are a historical OD matrix X^H − which is usually provided by either a household survey or a former demand model − and indirect traffic volumes measured − link flow counts, $\{\hat{Y}_l\}$, which are measured by detection stations, at a subset of links $l \in \hat{L} \subseteq L$ in the network, where L is the set of links −. The problem is then formulated as the optimization problem:

$$\min Z(X, Y) = w_1 F_1 \left(Y, \hat{Y}\right) + w_2 F_2 \left(X, X^H\right)$$
$$Y = Assignmt(X) \tag{1}$$
$$X \geq 0$$

where F_1 and F_2 are distance functions between estimated and observed values and w_1 and w_2 are weighting factors reflecting the uncertainty of the information contained in $\hat{Y}$ and X^H, respectively. The underlying hypothesis is that $Y(X) = \{y_l\}$ are the link flows predicted by assigning the demand matrix X to the network, which can be expressed by a proportion of the OD demand flows x_{nr} passing through the count location at link l at time period t. Therefore, in terms of an assignment matrix $A(X)$, function of the demand X, it can be stated as:

$$Y = A(X)X \tag{2}$$

where $\{a_{l,nr}\}$ is the assignment matrix whose entries $a_{l,nr}$ are the fractions of the OD demand x_{nr} of $n - r$ OD-pair that passes through link l, and the problem can then be formulated as:

$$\textbf{(A)} \quad \min Z(X) = w_1 F_1 \left(A(X)X, \hat{Y}\right) + w_2 F_2 \left(X, X^H\right)$$
$$X \geq 0 \tag{3}$$

This formulation allows resorting to a variety of analytical solutions, such as, Spiess, H. (1990), Lundgren, J. T. et al. (2008) and Toledo, T. et al. (2013), just to mention a few examples. However, this formulation allows for an analytical solution only when link flows $\{\hat{Y}_l\}$, $l \in \hat{L} \subseteq L$ are the unique observable reference for adjusting the historical OD matrix X^H in order to find a new OD matrix X that will likely explain the observed link flows. This is because Equation (2) then sets a direct analytical relationship between the observed $\{\hat{Y}\}$ and the estimated $\{Y\}$ link flows. However, it is unclear how to include them into an analytical formulation when other indirect measurements are available, such as those supplied by ICT sensors, (e.g., Bluetooth measurements of travel times, GPS tracking measurements , cellular phone tracking or electronic toll collection). Difficulties arise because travel times between arbitrary pairs of points in the network are not direct outputs of an analytical model, which results in the need for an alternative approach. If the *Assignment* in Problem (1) is replaced by a Dynamic User Equilibrium (DUE), for instance, Mahut, M. et al. (2010) and ADB30 (2010), then time-dependent link travel times can be available as output and they can subsequently be employed to calculate the used path travel times between any pair of points in the network. This can be included in other algorithmic approaches, such as those based on a Simulation-Based Optimization approach Osorio, C. et al. (2013, 2015). In this case, the upper level optimization function in (1) is no longer an analytical function, so it cannot be evaluated analytically. The most common computational approach for solving the DUE in order to evaluate the objective function in Problem (1) is a Simulation Based Assignment, Barceló, J. (2010). Therefore the Stochastic Perturbation Stochastic Approximation (SPSA), Spall, J. C. (1998) appears to be a natural candidate for solving the upper level optimization problem. Successful examples of SPSA applications for solving the bi-level approach to OD adjustment (when additional traffic measurements are available) can be found in Cipriani, E. (2011) or Bullejos, M. et al. (2014). However, there is evidence not only on the practical difficulties of employing SPSA, but also on the difficulty of getting consistently good results when moving from one instance of a type of problem to another or even when changing the type of problem.

The purpose of the research reported in these notes is to investigate the robustness and stability of SPSA when applied to the dynamic estimation of origin to destination matrices as well as when analyzing the influence of the factors that play a key role in the numerical process and when setting rules and patterns of behavior for the parameters that condition the convergence of the process. In order to learn from a conventional case before applying SPSA to it with additional ICT measurements, we have therefore analyzed:

- The role of the initial matrix, and its perturbations.
- The effects of bounding the changes in the OD values by taking into account not only the numerical role in the optimization process but also its physical meaning in terms of the underlying transportation system.
- The role of normalizing the terms of the objective function to avoid scale effects when the terms correspond to the measurements of different variables and magnitudes, as analyzed in some computational experiments.
- The scaling of the OD values to be estimated in order to avoid significant changes related to their magnitudes.
- The role of the number of independent gradient estimations averaged in the SPSA algorithm.

2. SPSA Procedures

The optimization problem is non-convex and non-linear, and the function cannot be represented analytically as a function of the OD values. Therefore, it is non-differentiable with respect to the OD values. Furthermore, the function cannot be computed analytically; consequently, it must usually be numerically evaluated by simulation, which makes the problem a natural candidate for simulation-based optimization methods Osorio, C. et al. (2015). This prompts resorting to non-differentiable optimization methods such as SNOBFIT, Genetic Algorithms, SPSA and others. This work uses the optimization procedure SPSA, which was originally designed in Spall, J. C. (1998) to solve the non-differentiable simulation-based optimization problems that are typical in simulation applications. It is therefore a natural candidate for solving Problem (1). SPSA is even more appealing than other candidates, since all non-differentiable methods strongly rely on evaluations of the objective function, which are computationally expensive in simulations. Furthermore, SPSA requires fewer evaluations than other approaches. In demand estimation for traffic simulation models, SPSA is one of the most commonly used algorithms. SPSA represents a critical advance for those minimizations where the analytical gradient cannot be computed and numerical gradient approximations are extremely expensive in regard to computational resources, meaning in other words that large computation time is needed for each dynamic traffic assignment calculation. In the simplest form of SPSA, only two assignments per iteration are needed to compute a gradient estimation. Using the same notation as in Problem (3), the algorithm starts from an initial OD matrix, typically the historical OD matrix $X_0 = X^H$, and the iterative procedure is a first order descent method:

$$X_{k+1} = X_k - a_k \cdot \hat{g}_k(X_k) \tag{4}$$

where $\hat{g}(X_k)$ is the approximation to the gradient direction using $\Delta_k \sim Be(1/2, \pm 1)$, where each Δ_k follows a Bernoulli distribution vector with average probability of $1/2$ for each ± 1 outcome. The perturbation chosen in this work for computing the gradient is the asymmetric design shown below.

$$\hat{g}_Z(X_k) = \frac{Z(X_k + c_k\Delta_k) - Z(X_k)}{c_k} \cdot \Delta_k^{-1} = \frac{Z(X_k + c_k\Delta_k) - Z(X_k)}{c_k} \cdot \begin{bmatrix} \Delta_1^{-1} \\ \vdots \\ \Delta_d^{-1} \end{bmatrix}_k \tag{5}$$

where c_k, a_k are inherent algorithm coefficients specified in Spall, J. C. (1998), and they are computed from the starting coefficients a, A, c, α, γ, which must be selected carefully, as $a_k = a/(A + k + 1)^\alpha$ and $c_k = c/(k + 1)^\gamma$.

As already mentioned, the intention of this work is to test the robustness of SPSA by experimenting with the different forms of the method for the original minimization Problem (3). Thus, the original problem has been modified to two different realistic approaches:

2.1. Constrained SPSA

In traffic simulation models, practitioners have access to some historical data in the form of an OD matrix, X^H, which, with a certain degree of uncertainty, provides some prior information about the mobility pattern of the study area. In these cases, Problem (3) can be defined by adding some constraints to the OD values, where a range can be added using a factor, β, to state the confidence interval of the historical data. The problem can be rewritten as follows:

$$\textbf{(B)} \quad \min Z(X) = w_1 F_1\left(A(X)X, \hat{Y}\right) + w_2 F_2\left(X, X^H\right)$$
$$X \in G = \left\{(1 - \beta)x_{nr}^H \le x_{nr} \le (1 + \beta)x_{nr}^H, \ \forall (n, r) \in OD\right\} \tag{6}$$
$$X \ge 0$$

This variation implies some modifications to the originally presented SPSA algorithm. Sadegh, P. (1997) proposed adding a projection to the set G during the iterative procedure shown in Equation (4). This projection affects only the iterative procedure $X_{k+1} = \pi_G(X_k - a_k \cdot \hat{g}_k(X_k))$, while $X_k + c_k\Delta_k$ can fall outside of the feasible set G; and $Z(X_k + c_k\Delta_k)$ can be computed subject to non-negative OD values.

2.2. Penalized SPSA

In a similar way, Wang, I.-J. et al. (1999) proposed a modified SPSA based on penalty functions, which is an equivalent formulation of Problem (6) but with implicit constraints on the objective function. With this new formulation, let Problem (3) take the following form:

$$\textbf{(C)} \quad \min \tilde{Z}(X) = w_1 F_1\left(A(X)X, \hat{Y}\right) + w_2 F_2\left(X, X^H\right) + r_k P\left(X, X^H\right)$$
$$X \geq 0 \tag{7}$$

where r_k is defined as other sequences : $r_k = r/(k+1)^{0.1}$. The drawback to this formulation is that this adds another algorithm parameter to be tuned: r. $P\left(X, X^H\right)$ is a set of penalization functions for the set of constraints which delimits the constraints set G of Problem (6). Actually, the constraints set can be rewritten as:

$$G \triangleq \left\{q_{nr}\left(X, X^H\right) \leq 0 \,,\, (n, r) \in OD\right\} = \left\{x_{nr} - (1 + \beta)x_{nr}^H \leq 0 \,,\, (1 - \beta)x_{nr}^H - x_{nr} \leq 0 \,,\, \forall(n, r) \in OD\right\} \tag{8}$$

and the penalty function $P\left(X, X^H\right)$, which has to be a differentiable, non-negative and increasing function, can be:

$$P\left(X, X^H\right) = \sum_{(nr) \in OD} \omega_{nr} \cdot p\left(q_{nr}\left(X, X^H\right)\right) \tag{9}$$

where ω_{nr} are weighting factors and $p(\cdot)$ is also a penalty function; and $p(x) = 0$ if and only if $x \geq 0$, such as in the quadratic form $p\left(q_{nr}\left(X, X^H\right)\right) = \max\left\{0, q_{nr}\left(X, X^H\right)\right\}^2$, which will be used in the experimentation. Problem (7), based on penalizing functions, modifies the iterative procedure and converts it to:

$$X_{k+1} = X_k - a_k \cdot \hat{g}_{\tilde{Z}}(X_k) = X_k - a_k \cdot \hat{g}_k\left(X_k\right) - a_k \cdot r_k \cdot \nabla P\left(X_k, X^H\right) \tag{10}$$

3. Experimental Design

A model of a subarea of the city of Hillsboro, Oregon, USA, has been selected for these experiments, Figure 1a. It consists of 618 links and 58 zones, and the simulation runs over a time horizon from 08AM to 09AM in 3 periods of 20 minutes. The assignment engine used is the Simulation Based Assignment (SBA) in PTV VISUM 17 (2017), exploiting the use of multiple parallel assignments via Visum Scenario Manager, with stopping criteria of maximum 100 iterations or 3% relative gap, ADB30 (2010). Each single assignment takes less than 1 minute using a Intel(R) Core(TM) i7-7820HQ CPU @ 2.90GHz. The computational experiments are synthetic in nature, using a Ground Truth OD Matrix $\left(X^{GT}\right)$ to generate the input data. Thus, the counts for the 80 most used links are taken as the real counts in the network. In Figure 1b, the detection layout is colored in purple and in orange (see Section 4 for color details) in an assigned schema of the network, where link bars represent assigned volumes.
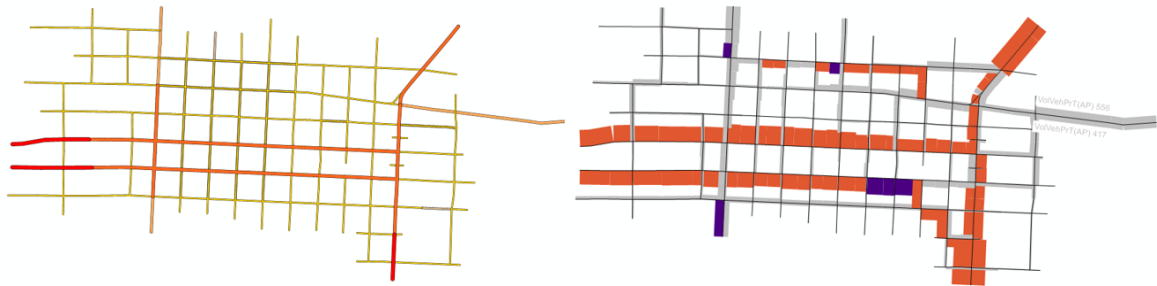

Fig. 1. (a) Network of a subarea of the city of Hillsboro (colors represent hierarchical links type); (b) Detection Layout schema.

For the historical OD matrices, 4 different OD matrices have been built in order to address the behavior of SPSA in different realistic situations:

- Inc$^+$: Incrementing all the OD values by a fixed percentage, i.e., Inc$^+$ = $X^{GT} \cdot (1 + \delta)$, and $\delta = 0.2$.
- Inc$^-$: Decrementing all the OD values by a fixed percentage, i.e., Inc$^-$ = $X^{GT} \cdot (1 - \delta)$, and $\delta = 0.2$.
- Mix$^+$: Incrementing all the OD values by a random variable percentage, i.e., Mix$^+$ = $X^{GT}_{nr} \cdot (1 + \delta_{nr})$, and $\delta_{nr} \sim U([0.18, 0.22])$.
- Mix$^-$: Decrementing all the OD values by a random variable percentage, i.e., Mix$^+$ = $X^{GT}_{nr} \cdot (1 - \delta_{nr})$, and $\delta_{nr} \sim U([0.18, 0.22])$.

The experiments explore, firstly, whether or not the congestion of the network has influence on the performance and, secondly, whether or not there is any influence from breaking the natural structure of the Ground Truth OD matrix, by perturbing each value independently. These perturbations increment or decrement only values different from zero. All 4 Historical matrices are compounded by $58 \times 58 \times 3 = 10092$ OD values, but most of them are lower than 2 and will not be included in the calibration procedure. Therefore, the number of OD values subject to optimization is around 800 and 1000.

The objective function of a Dynamic OD matrix estimation (DODE) problem is typically a set of distance measures to available real data. There is a wide range of measure functions that can be considered in this minimization problem, but some previous works, Toledo, T. et al. (2013), use absolute, quadratic and relative differences of the available data. In this case, the objective function measures relative differences:

$$Z(X) = w_1 F_1\left(A(X)X, \hat{Y}\right) + w_2 F_2\left(X, X^H\right) = w_1 \sum_{l \in \hat{L}} \frac{|y_l - \hat{y}_l|}{\hat{y}_l} + w_2 \sum_{(n,r) \in OD} \frac{\left|\tilde{x}_{nr} - \tilde{x}^H_{nr}\right|}{\tilde{x}^H_{nr}} \tag{11}$$

where tildes in $\tilde{x}_{nr}, \tilde{x}^H_{nr}$ indicate scaling of the OD matrix values to uniformize their magnitudes with the intention of applying the same effect to them during the SPSA process. For instance, without the scaling, the simultaneous perturbation applied will provoke uncontrolled oscillations in the low values at the early iterations or in later convergences because of the slight changes to the highest OD values, depending on the tuning of SPSA coefficients a, A, α. As already mentioned, SPSA coefficients play a critical role in the SPSA performance; and, since it was not in the initial scope for this work, guidelines in Spall, J. C. (1998) have been used to select appropriate values. The values for these experiments are the following: $a = 0.08, A = 30, \alpha = 0.602, c = 0.15, \gamma = 0.101, r = 50, w_1 = 1$. For the OD values part of the objective function, w_2, their weight has been set to 0 or 150 in different experiments in order to evaluate its effect on the convergence. Spall, J. C. (1992) shows that averaging many independent estimations of the gradient (see Equation (5)) contributes to a more stable and quicker convergence of the SPSA method. To see this effect in this work, the number of independent calculations has been set to $n_g = 3$ and 5. Therefore, the gradient estimation is finally

$$\hat{g}_Z(X_k) = \frac{1}{n_g} \sum_{j=1}^{n_g} \hat{g}^j_Z(X_k) \tag{12}$$

where $\hat{g}^j_Z(X_k)$ is calculated as in Equation (5). The asymmetric design for the gradient saves a huge number of assignments, since all $\hat{g}^j_k(X_k)$, $\forall k$ share the mid-point X_k evaluation.

To summarize, there are three different SPSA methods (see Section 2), four different initializations of X^H, and combinations of parameters $\{n_g = 3, w_2 = 0\}, \{n_g = 3, w_2 = 150\}, \{n_g = 5, w_2 = 0\}$. Therefore, the total number of experiments conducted in this work are 36. The stopping criteria for all these experiments is set at a maximum number of 1200 assignments. Taking into account that the asymmetric design for the gradient requires $n_g + 1$ assignments per iteration, this criteria stops the SPSA procedure at 300 or 200 iterations when $n_g = 3$ or 5, respectively.

4. Results

In order to investigate the performance of SPSA in Simulation-Optimization approaches to transportation problems, two different perspectives are addressed: Firstly, the speed of convergence, stability and performance of the algorithm must be considered; followed by, secondly, the quality of the partial solution, its adjustment to real data and proximity to the Ground Truth OD Matrix. For the quality measurements, entropy has been chosen as a similarity measure

between OD matrices, Ortúzar, J. et al. (1990), which can be calculated as shown below in Equation (13), since the real counts of the experiment have been built artificially using the Ground Truth OD matrix. We will show the duality between these two perspectives.

$$E\left(X, X^{GT}\right) = -\log\left(W\left(X, X^{GT}\right)\right) = \sum_{(r,s)\in OD}\left(x_{rs}\log\left(\frac{x_{rs}}{x_{rs}^{GT}}\right) - x_{rs} + x_{rs}^{GT}\right) \tag{13}$$

Figure 2a shows the evolution of the objective function during the SPSA process for 6 experiments. These experiments have the same initial historical matrix and only the method and number of averaging gradients change. Note that for the three methods (A), (B) and (C) that are defined in Equations (3), (6) and (7), respectively, the experiments with $n_g = 5$ present a better initial descent and fewer oscillations. On the other hand, the basic SPSA problem (A) is the one that obtains a lower value, while Problems (B) and (C), which are restricted and penalized problems, converge on solutions with higher objective function values. Figure 2b shows the same experiments, but by plotting the entropy with respect to the Ground Truth OD matrix. The order of the methods is completely opposite: the best performance, in terms of similarity is the restricted SPSA; while the worst is the unrestricted SPSA, (A), whose entropy value clearly increases. Balancing the two plots of Figure 2, Penalized SPSA (C) achieves a higher quality solution while also delivering a good performance of the objective function minimization procedure.
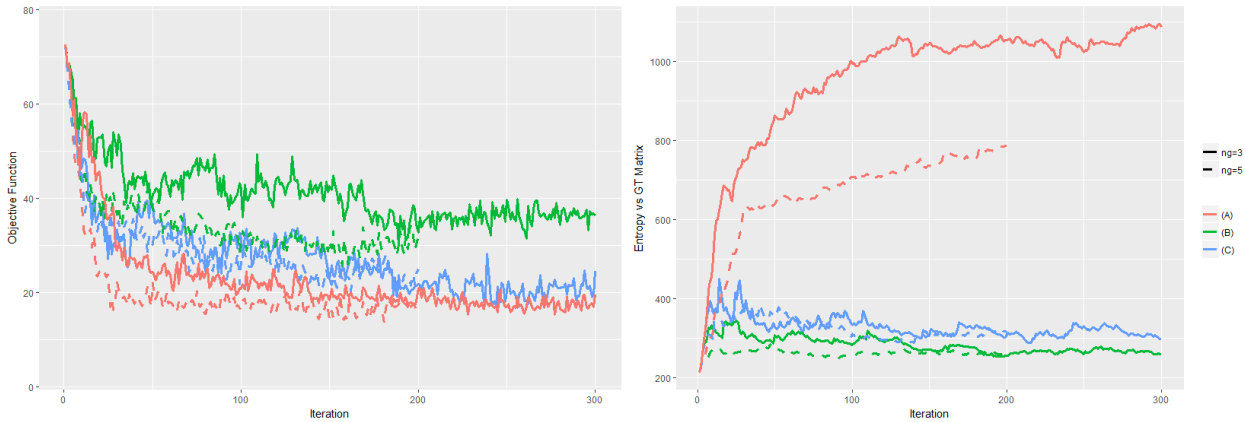

Fig. 2. (a) Objective Function evolution; (b) Evolution of entropy with respect to GT.

In Figure 3, different experiments using the penalized SPSA (Problem (C)) are plotted. The 8 experiments shown are those corresponding to the 4 different historical matrices and $n_g = 3$ or 5. On the left is the entropy with respect to the Ground Truth OD matrix. In terms of similarity, there is a notable difference between those incremented and decremented initializations as defined above. The incremented OD matrices present a worse solution in terms of quality, due to the additional congestion. On the right, the number of total trips at each iteration is shown. The Ground Truth OD Matrix has 9877.53 trips in total and, surprisingly, the SPSA process is not able to reduce or increase these values in order to reach the number of trips for the Ground Truth.

Table 1 contains a summary of indicators for the developed experiments. The rows show the SPSA algorithm factor with levels (A) to (C), where (A) is the basic SPSA Eq. (3), (B) is the constrained SPSA Eq. (6) and (C) is the penalized SPSA Eq. (7). In the main columns are the initialization factors with levels $\texttt{Inc}^-$, $\texttt{Inc}^+$, $\texttt{Mix}^-$ and $\texttt{Mix}^+$, as detailed in Section 2. The level result of each main column is split into as many columns as there are levels in the SPSA parameter factor (see the last paragraph of Section 2 for details). For each of the 36 experiments, the included goodness of fit (GoF) measures are: R_o^2, the coefficient of determination in the regression line between observed and estimated link counts at convergence; R_f^2, the coefficient of determination in the former regression line when outliers are removed (those not included in the 95% confidence band for prediction in regression); "# out" is the number of outliers on the first linear model; "% tot" is the percentage of OD values – subject to the optimization procedure – that are fully captured (all used paths) by the detection layout; and "% part" is the same percentage, also including those OD values that are partially captured (at least one used path).
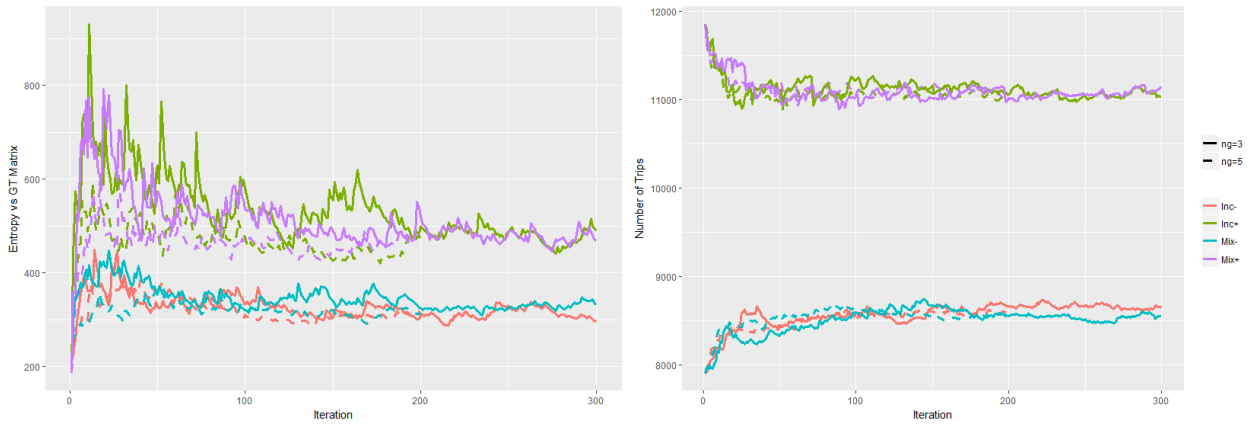
Fig. 3. (a) Entropy evolution in Problem (C); (b) Number of Trips evolution in Problem (C).

Table 1. Goodness of fit measures for the 36 experiments

		Inc^+			Inc^-			Mix^+			Mix^-		
		$n_g = 5$ $w_2 = 0$	$n_g = 3$ $w_2 = 0$	$n_g = 3$ $w_2 \neq 0$	$n_g = 5$ $w_2 = 0$	$n_g = 3$ $w_2 = 0$	$n_g = 3$ $w_2 \neq 0$	$n_g = 5$ $w_2 = 0$	$n_g = 3$ $w_2 = 0$	$n_g = 3$ $w_2 \neq 0$	$n_g = 5$ $w_2 = 0$	$n_g = 3$ $w_2 = 0$	$n_g = 3$ $w_2 \neq 0$
	R_o^2	0.8913	0.8838	0.8812	0.965	0.9522	0.9291	0.8367	0.8725	0.8634	0.8934	0.8866	0.921
	# out	8	10	13	12	16	8	16	16	12	9	17	9
(A)	R_f^2	0.9099	0.9128	0.9145	0.9764	0.967	0.9386	0.8852	0.9238	0.9013	0.9211	0.9148	0.9376
	% tot	68.99	66.46	64.24	65.82	61.39	61.39	64.56	64.56	67.09	64.87	64.87	65.51
	% part	81.96	84.18	83.86	82.59	84.81	82.28	84.49	83.86	84.18	85.13	85.13	82.59
	R_o^2	0.9583	0.959	0.9414	0.9597	0.9142	0.9165	0.966	0.9644	0.9474	0.9277	0.9162	0.9389
	# out	10	14	12	13	13	11	10	11	17	8	14	14
(B)	R_f^2	0.9718	0.9696	0.9587	0.9688	0.9365	0.9385	0.9763	0.9743	0.9657	0.9395	0.9522	0.9582
	% tot	64.87	63.61	68.67	62.66	64.56	63.61	65.19	68.99	66.46	64.56	62.97	65.51
	% part	85.76	86.08	86.71	86.08	81.96	84.49	81.33	86.39	86.39	85.44	83.23	86.71
	R_o^2	0.9569	0.9565	0.9533	0.9491	0.964	0.961	0.9554	0.9577	0.9473	0.9684	0.9523	0.9586
	# out	9	12	13	13	9	14	14	11	11	9	12	16
(C)	R_f^2	0.965	0.968	0.9667	0.9605	0.9727	0.9727	0.9746	0.9694	0.9647	0.9749	0.9714	0.9706
	% tot	62.66	66.77	67.41	64.87	62.34	61.71	63.92	66.77	66.46	65.82	62.97	59.81
	% part	82.91	83.54	85.44	84.81	83.54	80.38	82.59	85.13	83.86	81.65	85.44	85.13

For the basic SPSA algorithm (A), its R^2 fit to observed counts is in general poorer than the fit shown by (B) and (C), whether it be before or after removing the outliers whose numbers are similar for all runs. Detectors partially capture 81-87% of OD pairs, and they fully capture 59-69% of the optimized OD values; so there are many cells in the matrices that cannot benefit from traffic count adjustment, which thus confirms the high level of indetermination in the process. SPSA (B) is slightly better than (C) according to R_f^2. Once outliers are removed the SPSA parameter factor seems to be irrelevant. Looking at the third column in each block of Table 1, it appears that the second component of the objective function does not show a significant impact on the experiments, perhaps because the historical matrices X^H were built artificially.

Regarding the number of outliers, all 36 experiments have between 8 and 17 outliers, which is lower than the 8% of all available count measures ($80 \times 3 = 240$). However, there are some sensors (marked in purple in Figure 1b) that appear more often as outliers (around 30-40% of the experiments) and mostly in the third period of the simulation. This fact leads one to question the correct positioning of these sensors. Together with the detection layout's low coverage of the OD demand which is reminiscent of Bierlaire, M. (2002), who proposed a new measure of the indetermination of the OD estimation problem, the Total Demand Scale (TDS). From the conducted experiments, the 36 assigned with A_{T1}, A_{T2}, A_{T3} (the assignment matrices for each simulation period) have been used to estimate the associated minimization problem of TDS, which have result unbounded in all cases. All these results, taken together, confirm the hypothesis that the available data is not enough to cover the indetermination of the calibration problem.

5. Conclusions

The basic SPSA proposal, Problem (A), presents a nice convergence speed and good fit of estimated flows to real counts at convergence. Alternatively, when adding the experimental experience of SPSA variants (B) and (C), the convergence speeds in both cases do not seem so good. However, according to the graphics in Figures 2 and 3, an entropy comparison with respect to the Ground Truth OD matrix shows a discrepancy in the matrix pattern for the basic SPSA (A) case, which can be avoided using alternatives (B) and (C). This is a severe drawback for practical applications that enforces the need for restricting large OD trip changes during the optimization procedure. In any case, the number of trips in the Ground Truth OD matrix cannot be achieved (Figure 3b). The SPSA (A) variant increments or decrements the estimated OD matrices more sharply than alternative variants. After analyzing the OD pattern structure of the SPSA (A) variant by using the entropy measure, large discrepancies appear; while the consistency of obtained quality results for (B) and (C) variants is better. The SPSA (A) variant behaves as a meta-regression model, but the quality of the estimated matrices is worse. And last but not least, the TDS indicator is unbounded for all the experiments, pointing at deficiencies in sensor location since many of the OD paths are not captured by the current detection layout. Further experimentation is needed to establish the impact of the detection layout on the quality of the estimated OD matrices by SPSA variants. In summary, SPSA variants behave differently when the performance and quality of the solution are considered. In addition, depending on the actual historical and available measurements, the penalized and constrained SPSAs are highly recommended. Furthermore, when using relative distance functions in the objective function, SPSA is a natural choice for including different magnitude measures in the calibration procedure, and this work presents a preliminary good behavior for it.

Acknowledgements

References

ADB30, 2010. A Primer for Dynamic Traffic Assignment, Transportation Network Modeling Committee Transportation Research Board.

Barceló J. (editor), 2010. Fundamentals of Traffic Simulation, Springer, ISBN: 978-1-4419-6142-6.

Bierlaire M., 2002, The total demand scale: a new measure of quality for static and dynamic origin-destination trip tables, Trans. Res. B 36, pp 837–850.

Bullejos M., Barceló, J., Montero, L., 2014, A DUE based bilevel optimization approach for the estimation of time sliced OD matrices, International Symposium of Transport Simulation, Ajaccio.

Cipriani E., Florian M., Mahut M. and Nigro M., 2011, A gradient approximation approach for adjusting temporal origin destination matrices, Transportation Research Part C 19, pp 270-282.

Lundgren, J. T. and Peterson, A., 2008. A heuristic for the bi-level origin-destination matrix estimation problem. Transp. Res. B 42, pp 339–354.

Mahut, M. and Florian, M., 2010. Traffic Simulation with Dynameq, Chapter 9 in Fundamentals of Traffic Simulation, J. Barceló, Editor, Springer, ISBN: 978-1-4419-6142-6.

PTV AG. PTV VISUM 2017. Karlsruhe Germany. http://vision-traffic.ptvgroup.com/en-us/products/ptv-visum/.

Ortúzar, J. and Willumsen, L. G., 1990. Modelling Transport. John Wiley and Sons, Ltd, United Kingdom.

Osorio, C. and Chong, L., 2015. A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transportation Problems. Transportation Science 49(3): 623–636.

Osorio, C. and Bierlaire, M., 2013. A Simulation-Based Optimization Framework for Urban Transportation Problems. O. R. 61(6) p1333-1345.

Sadegh, P., 1997. Constrained Optimization via Stochastic Approximation with a Simultaneous Perturbation Gradient Approximation. Automatica 33(5) pp 332-341.

Spall, J. C., 1992. Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation. IEEE Trans. on Autom. Con. 37(3) pp 482-492.

Spall, J. C., 1998. Implementation of the Simultaneous Perturbation Algorithm for Stochastic Optimization. IEEE Trans. Aerosp. Electron. Syst. 34(3), pp 817-823.

Spiess, H., 1990. A Gradient Approach for the OD Matrix Adjustment Problem. Publication No. 693, CRT, Université de Montréal.

Toledo, T. and Kolechkina T., 2013. Estimation of dynamic origin-destination matrices using linear assignment matrix approximations. IEEE Transactions on ITS 14(2), pp 618–626.

Wang, I.-J. and Spall, J.C., 1999. A Constrained Simultaneous Perturbation Stochastic Approximation Algorithm Based on Penalty Functions. Proceedings of the American Control Conference, San Diego, California.