Dynamic demand estimation on large scale networks using Principal Component Analysis: the case of non-existent or irrelevant historical estimates

Moeid Qurashi[a], Qing-Long Lu[a], Guido Cantelmo[b] and Constantinos Antoniou[a]

[a]Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
[b]Technical University of Denmark, Bygningstorvet 116, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
DTA model calibration
OD estimation
Demand estimation
Large networks
Principal Component Analysis
SPSA
Dimension reduction

ABSTRACT

Calibrating DTA models is complex due to the involved indeterminacy, non-linearity, and dimensionality, restricting the application of conventional calibration approaches, especially for larger networks. For this, Principal Component Analysis (PCA) is slowly establishing itself as the new state of the art because it can greatly tackle two well known challenges - i.e. problem dimensionality and non–linearity. PCA application limits the optimization search space in a lower dimension space, defined by orthogonal Principal Components, evaluated upon a set of historical estimates. In this paper, we solve practical implementation problems for PCA–based calibration techniques. Specifically, we formulate a data–assimilation framework to propose multiple OD historical data–set generation methods which allows the use of PC–based algorithms in case the historical data is irrelevant or unavailable, often the case for large–scale DTA models. Furthermore, we propose a simplified problem formulation that leverages properties of the novel data–set generation framework and helps for faster and more efficient calibration. The methodology is implemented using the PC–SPSA algorithm, which combines PCA with the popular Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm, commonly used to calibrate smaller networks. The approach is tested on one of the largest case studies reported to date, the Munich metropolitan urban network, with encouraging calibration results. The proposed data–assimilation framework can account for spatial, temporal, and day–to–day variations in the demand. Different methods and combinations are tested and compared. The results suggest that all these correlations should be used in order to avoid over–fitting issues. Furthermore, the implementation properties of PCA and PC–SPSA are also explored using different sensitivity analyses to assess the toll and benefits of using PCA i.e., ease in SPSA hyper–parameter, role of historical data–set generation parameters and the algorithm's performance against different target demand fluctuations. The analysis shows encouraging results for PC–SPSA robustness and helps establishing simplified guidelines for implementing such PCA–methods practically on large–scale DTA models.

1. Introduction

Dynamic Traffic Assignment (DTA) models have been successfully applied as decision support tools for the evaluation of traffic planning and traffic management solutions for many years. They not only offer the opportunity to estimate and predict the transport network traffic state, but also to evaluate different transport measures, therefore quantifying their effectiveness (Ben-Akiva et al., 2001; Mahmassani, 2001; Tampere et al., 2010). Given the importance of these models, DTA calibration is a long–hauled research topic and the literature within the last decade is filled with many efforts trying to propose different calibration techniques with better application towards large DTA models (Balakrishna, 2006; Antoniou et al., 2009, 2015). Such a calibration problem is extremely complex as DTA models are highly non–linear and require a large set of parameters to be calibrated (Marzano et al., 2009).

As a wrong demand pattern will also generate a biased simulation output, mobility demand is one of the most important inputs for a DTA model. Typically, mobility demand is represented as an Origin–Destination (OD) demand matrix, where each cell of the matrix represents the number of trips travelling from a certain origin to a certain destination, during a specific time interval. The main problem is that state of the art measurement systems, such as loop

moeid.qurashi@tum.de (M. Qurashi); qinglong.lu@tum.de (Q. Lu); guica@dtu.dk (G. Cantelmo); c.antoniou@tum.de (C. Antoniou)
ORCID(s): 0000-0002-0135-6450 (M. Qurashi); 0000-0002-6087-8670 (Q. Lu); 0000-0001-5218-2609 (G. Cantelmo); 0000-0003-0203-9542 (C. Antoniou)

detectors, measure the effect of the demand on the network rather than the demand itself (Frederix et al., 2011). As a consequence, practitioners usually turn to demand generation models in order to estimate the OD demand matrix (McNally, 2007). Although demand generation models provide an initial guess about the demand, the estimated OD matrix is at most an approximation of the average demand. Unfortunately, daily demand patterns show substantial fluctuations with respect to the average demand, because of partially predictable phenomena such as weather conditions (Balakrishna, 2006). These deviations can be corrected by using traffic measurements, such as loop detectors, to update the existing (a-priori) OD matrix. This problem, which is known in the literature as the Dynamic Demand Estimation Problem (DODE), searches for time-dependent OD demand matrices able to best fit measured traffic data (Cascetta and Postorino, 2001).

Depending on the specific DTA application, several formulation frameworks have been proposed in the literature to solve the DODE problem. A first distinction is between *offline* (Balakrishna et al., 2007b; Cipriani et al., 2011; Antoniou et al., 2015; Osorio, 2019) and *online* models (Antoniou et al., 2007; Prakash et al., 2018; Cantelmo et al., 2020), where the former focus on medium-long term planning, while the latter are frequently adopted for real-time applications, such as route guidance. In addition, the DODE can be formulated as an optimization (Balakrishna et al., 2007b; Cipriani et al., 2011; Antoniou et al., 2015; Qurashi et al., 2019) or a state–space problem (Ashok and Ben-Akiva, 2000; Antoniou et al., 2007; Prakash et al., 2018; Cantelmo et al., 2020). The state–space formulation is especially used for capturing day–to–day dynamics (Zhou and Mahmassani, 2007) or for on-line demand estimation (Ashok, 1996; Ashok and Ben-Akiva, 2000). However, studies that implement a state-space model in the context of off-line also exist (Balakrishna et al., 2005). Finally, we can further divide existing models into assignment–matrix based and assignment–matrix free algorithms (Cantelmo et al., 2014b). Assignment–matrix based algorithms explicitly use an analytical representation of the relationship between demand and traffic flows to estimate the most likely demand matrix (Cascetta and Postorino, 2001; Toledo and Kolechkina, 2012). However, this relationship is usually assumed to be linear. As this is not the case in reality, other authors proposed assignment–matrix free algorithms, using the DTA model to indirectly capture this correlation (Balakrishna et al., 2007b; Vaze et al., 2009). These models can be further divided into gradient–based (Cipriani et al., 2011; Antoniou et al., 2015; Qurashi et al., 2019) and gradient–free (Zhang et al., 2017; Osorio, 2019). Similarly, attempts to use Machine–Learning (neural networks) to solve the DODE problem have also been proposed (Wu et al., 2018; Krishnakumari et al., 2019).

Regardless of the specific application, recent years have witnessed a shift towards assignment matrix–free methods. Matrix–free algorithms solve two of the main issues common to all DODE formulations. First, they allow to accurately model the relationship between supply and demand. Second, assignment–matrix free formulations allow to incorporate any data source and do not require defining an analytical relationship between data and observations (e.g. between Bluetooth data and mobility demand). To include additional data is in fact a crucial aspect, as the DODE is traditionally a highly underdetermined problem, due to the fact that the number of variables to be estimated far exceeds the available amount of information (Marzano et al., 2009). One such approach, named 'Simultaneous Perturbation Stochastic Approximation' (SPSA) (Spall, 1998), has been one of the most popular algorithms for DTA model calibration (Balakrishna et al., 2007a). SPSA, due to its ability to deal with non–linear and stochastic systems, a generalized problem formulation, and ease of implementation, has been used frequently by many researchers (Balakrishna et al., 2005; Cantelmo et al., 2014a; Barceló et al., 2010; Ros-Roca et al., 2021). However, conventional algorithms, including the SPSA, often fail in convergence with large–scale problems, because their performance deteriorates rapidly with the increase of the problem scale and complexity. For example, SPSA's gradient approximation gets highly sensitive against: 1) definition of hyper–parameters (objective function gets more expensive, making trail–based setup infeasible); 2) more varying OD magnitudes, which increase exponentially with DTA model size and are also sparsely correlated with traffic measurements.

Most of the literature, which aims to improve the application scalability of DTA model calibration, has followed two major domains i.e., reducing problem dimensions or reducing problem non–linearity (adding structural/correlation information in the objective function). Within the dimension reduction domain, approaches tend to reduce the number of estimation variables by e.g., using a statistical technique i.e. Principal Component Analysis (PCA) (Djukic et al., 2012; Prakash et al., 2018; Qurashi et al., 2019), using a correlation assumption i.e., quasi dynamic (Cascetta et al., 2013; Cantelmo et al., 2014b), clustering the model parameters (Tympakianaki et al., 2015), redefining the problem formulation i.e., utility–based formulations (Cantelmo et al., 2018, 2020). While, in the other domain of catering problem non–linearity, approaches tend to add additional structural/correlation information spatially or temporally among model parameters and traffic measurements e.g., adding a weight matrix for correlation between ODs and network (Cantelmo et al., 2014a; Lu et al., 2015; Antoniou et al., 2015), using response surface methods or (physical) metamodels which approximate the DTA simulation's input/output relationship using differentiable analytical functions (Zhang et al., 2017; Osorio, 2019).

Within all such efforts for improving the objective functions of different conventional approaches, the application of PCA stood out for being significantly more efficient in reducing problem dimensions (from the scale of 10^3 to 10^1) and non–linearity, as it transforms the OD vector into a lower dimensional space defined by orthogonal/uncorrelated PCs extracted from the variance of historical OD estimates. Hence, the application of Principal Component Analysis (PCA) has been widely adopted for both offline and online calibration problems to do dimension reduction. For DTA model calibration, it is first proposed by Djukic et al. (2012), followed by many other approaches e.g., PC–GLS (Prakash et al., 2017), PC-EKF (Prakash et al., 2018), and PC–SPSA Qurashi et al. (2019). In all these PCA-based OD estimation frameworks, given a series of historical estimates, PCA leverages strong patterns and correlations to represent the problem with a few orthogonal/uncorrelated Principal Components (PCs) in a low dimensional space.

PC–based methods, although being powerful and intuitive, strongly rely on the presence and quality of the historical estimates, by which they extrapolate estimation patterns. PCA provides a considerable advantage through dimension reduction, providing a lower dimensional search space based on PCs evaluated from historical data–set. Hence, application and performance of PCA–based methods is limited by the presence and quality/relevance of the historical data–set relative to the target solution. This in general is not possible for large–scale DTA models, for which such PCA–based methods are proposed, because conventional calibration techniques struggle to calibrate them and PCA application requires historical estimates. Considering this limitation, this paper aims to further explore, optimize, and establish better implementation methods for the application of demand estimation models based on Principal Component Analysis and PC–SPSA. First, we propose a new methodology to use PC–based methods, when a reliable historical data–set is not available, including both cases of non–existent or irrelevant historical estimates. Then, we also establish a simplified problem formulation to define the objective function which alongside improving the computational times significantly, increases both the least error convergence and solution quality. Later, to test the novel procedures, we use the PC–SPSA algorithm to calibrate one of the largest case studies reported in DTA calibration literature i.e., the Munich city network. Comparisons among all different possible historical OD generation methods and conventional versus simplified problem formulation are performed to understand and depict the associated benefits. Later, to explore and understand the practical implementation of PC–SPSA, we perform a series of sensitivity analyses focusing on two aspects. First, to test the ease of setting up SPSA hyper–parameters for PC–SPSA calibration. Second, to test the influence of different historical data–set characteristics (i.e., size, variance, and dimension reduction) and their quality/relevance with respect to target demand. Finally, using all different empirical results, we also propose a set of guidelines helpful to conveniently setup PCA based methods and PC–SPSA.

Below, the contributes of this paper are further discussed in detail.

1. **Properties of Principal Component Analysis**

   (a) **Generation of historical OD estimates:** This paper contributes to define a data-assimilation framework for both generating historical estimates data–set and controlling their quality. As mentioned above, PCA, using the historical data–set, extracts a set of PCs, which are then used to transform and estimate the model parameters in a lower–dimensional space, restricting the search space of the model. Hence, the application and performance of any PCA–based method is limited by the presence and quality of these historical estimates. The data–assimilation framework proposed in this paper explores all possible correlation in the existing demand matrix and generates a set of (artificial) historical estimates from a given historical OD matrix. In addition, this method also provides the possibility to derive these correlations from different available data sources which can help further reduce the residual errors.

   (b) **Simplified problem formulation:** This paper proposes a simplified problem formulation for DODE. Since OD demand cannot be observed and the DODE problem is under–determined, the objective function comprises two minimizing error terms: the error between simulated and traffic measurements and the one between calibrated and initial OD estimate. The second term constraints the calibrated OD demand from moving away off the starting estimate and overfitting the traffic data. It also limits the calibration performance due to added noise and complexity in the objective function. We show in this research that the application of PCA does not require to constraint the calibrated OD pattern, as the proposed data-assimilation framework allows to include information about the historical matrix directly into the principal components, even when historical estimates are not available. The search space is therefore already restricted by the variance of historical data–set, hence the DODE problem formulation can be simplified by using only the error term between traffic measurements. Comparisons between the simplified and the conventional problem formulation show that simplified formulation requires significantly lesser number of iterations and converges to the least error and best OD solution quality. That because, conventional formulation leads to a sub-optimal solution due to an over representation of the historical demand in the objective function.

2. **Implementation properties of PC–SPSA**

   (a) **Ease of hyper–parameters tuning:** This paper performs sensitivity analyses for robustness of PC–SPSA against SPSA hyper–parameters. Although (Spall, 1998) has proposed general guidelines for SPSA hyper–parameters definition, there is no set rule to define these parameters generically for SPSA and its variants. Besides, even other approaches that aim for calibrating large scale DTA models require regress effort to set up specific to a DTA model (e.g., defining appropriate metamodel functions (physical metamodel as per (Zhang et al., 2017; Osorio, 2019)), create network correlation weight matrices (Antoniou et al., 2015)). In previous studies, sensitivity analysis was often use to identify case–specific SPSA hyper–parameters (Cantelmo et al., 2014a). In this research, we show that PC-SPSA converges on high quality solutions even when a sub–optimal set of hyper–parameters is used. This advantage of skipping problem-specific manual input, especially with large–scale DTA models, is the reduction of additional computational effort for running simulations repeatedly during this trial/definition phase. In this paper, we show that PCA gives the ease to tune SPSA's hyper–parameters because of showing significant robustness to a range of different hyper–parameter values.

   (b) **Value of added structural information:** Many literature approaches use different techniques to add information within the objective function for improving their application scalability (Antoniou et al., 2015; Cantelmo et al., 2014a; Tympakianaki et al., 2018; Osorio, 2019). Similarly, PCA also incorporates OD structural patterns from the historical estimates to reduce non–linearity and computational requirements. But how to implement PCA in real-life applications is still an open question. It requires determining the optimum number of historical estimates, the effect of the amount of variance present in the system, and the optimum amount of PCs. Defining these set of hyper–parameters adds up as a requirement due to the PCA application. Hence, in this paper, we perform multiple sensitivity analyses to measure the impact of varying historical data–set characteristics (i.e., size, variance, and number of PCs) on PC–SPSA calibration performance. The analysis helps to understand the value of structural information added in the objective function. It also provides directions to control model over fitting.

   (c) **Computational efficiency:** Most calibration methods, let alone their capability to calibrate large–scale networks, require significant computational efforts due to the higher simulation run–times, large set of estimation variables, and iterative nature. Methods proposed in this paper address this practical challenge, and calibrate one of the largest calibration experiment for DODE to date i.e., the Munich network. First, the results show the direct benefits of PCA i.e., the increase in dimensionality and non–linearity/complexity for Munich network doesn't directly translate into an equivalent increase in optimization complexity and estimation variables. Moreover, exploiting PCA properties, the ease of SPSA hyper–parameters tuning eliminate the need of recursive simulations for trail–based setting. Similarly, we also eliminate the requirement of using multiple gradient replications in SPSA (otherwise used in all SPSA methods to remove gradient biased). Also, the simplified problem formulation provides significant improvements for the required number of iterations. Hence overall, the calibration of Munich network is shortened between 2-6 iterations with practically almost 1 simulation run–time required per iteration (with parallel replications and SPSA gradient evaluation), making PC–SPSA even feasible for online calibration.

The rest of the paper is structured as follows. Section 2 describes the overall methodology followed in this research. After introducing PCA in the OD estimation context, we discuss the proposed data-assimilation framework for historical data matrix generation, the simplified problem formulation, and our implementation of PC–SPSA. Then, section 3 describes the experimental setup, network case study, and the calibration results for PC–SPSA. It also includes the comparisons for different historical OD generation methods and conventional versus simplified problem formulations. Later, section 4 covers the sensitivity analyses performed on PCA and PC–SPSA implementation properties alongside the guidelines for their setup. Finally, section 5 concludes the paper describing the overall contributions and findings of the research alongside its future implications and possible research directions.

2. PCA based OD estimation

Principal Component Analysis (PCA) is already a standard for problem dimension reduction. It allows to dimensionally reduce a large set of decision variables $\boldsymbol{\theta}$ or $\boldsymbol{x}$ (i.e., the starting OD vector for DODE) into few number of PC–scores using a lower dimensional space. This space is defined by a set of Principal Components (PCs) estimated by the application of PCA on the time series historical data of the decision vector. The optimization problem is then formulated and estimated using the PC–scores, solved by a suitable optimization approach. For OD estimation, Djukic et al. (2012) is the first to apply PCA on the time series OD matrices, extracting the spatial–temporal correlation among different OD pairs. Although the idea of PCA's application is of dimension reduction, it also gives other favorable properties. For example, it gives an orthogonal/uncorrelated OD demand representation which otherwise is sparsely correlated and it keeps the search space limited in the variance captured from historical estimates resulting in good quality OD solutions.

2.1. Dimension reduction

Principal Components (PCs) are linear vectors combinations containing the variance information of a time series data. All PCs have their subsequent coefficients (named 'PC–directions') which define the amount of variance captured by them. The value of these PC–directions decrease in an ascending order i.e., the first PC captures the highest sample variance in the data followed by the second PC with the second–highest variance captured and soon. The estimation of PCs requires a time series OD demand information which can be supplemented using historical OD estimates (calibrated offline or online). Given the availability of historical estimates, they are set in a data matrix $\boldsymbol{D}$ with dimensions $[n_k \times n_x]$, where n_k is the number of historical data points and n_x is the size of OD vector estimate. Then, Singular Value Decomposition (SVD) is applied on this historical data matrix $\boldsymbol{D}$ to evaluate the PCs, given as:

$$\boldsymbol{D} = U \Sigma V^T \tag{1}$$

The unitary matrix V with dimension $[n_x \times n_x]$ contains vectors of the orthogonal PCs and their corresponding PC–directions are stored in the rectangular–diagonal matrix Σ with dimension $[n_k \times n_x]$. U is a $[n_k \times n_k]$ unitary matrix with orthogonal vectors. A time series historical estimates data–set of n_k data points result in n_k PCs (Djukic et al., 2012), hence the first n_k columns of unitary matrix V are PCs and the diagonal n_k values of matrix Σ are their PC-directions. The evaluated PCs can be further reduced to retain only the first few significant PCs n_d, which can explain most of the time series variance from the historical estimates (Djukic et al., 2012), hence V is further reduced to $\hat{V}$:

$$\hat{V} = [\, v_1 \; v_2 \; v_3 \; ... \; v_{n_v} \,] \tag{2}$$

The starting OD vector $\boldsymbol{x}$ (otherwise used directly for estimation) is transformed into a lower dimensional PCs space. The reduced $\hat{V}$ unitary matrix containing n_v significant PCs is used to transform $\boldsymbol{x}$ into to set of PC scores z of dimension $[n_v \times 1]$, as:

$$z = \hat{V}^T x \tag{3}$$

These PC scores are instead then used for estimation, while the OD vector can be re–approximated as:

$$x \approx \hat{V} z \tag{4}$$

2.2. Historical data matrix generation

Historical OD estimates used for estimating PCs are critical for application of PCA–based methods. These historical estimates should be relevant temporally (i.e., day–to–day historical estimates of the same time intervals ($\mathcal{H} = \{1, 2, ...h\}$), to ensure similar OD spatial/structural patterns as of the target solution. This implies that different historical data–sets should be constructed between e.g., morning and evening peak hours, peak and off–peak hours, weekdays and holidays. If relevant estimates are not available then PCA–Based models will give poor quality solutions. Setting the relevance property aside, the existence/availability of historical OD estimates is even more critical (especially for large scale DTA models). As stated before (under section 2.1), it is evident from the literature that conventional models, such as SPSA, are in fact not capable of being used to calibrate large–scale networks and therefore the presence of calibrated/estimated historical OD data–set is impractical, hence limiting the use of PCA–based techniques in practice.

In this section, we propose a data-assimilation framework for applicability of PCA–methods in scenarios of irrelevant or non–existing historical estimates. In such scenarios, there exists a possibility to synthetically generate historical OD estimates using the available OD estimate. As discussed previously, PCA limits the search space by projecting each OD pair into a few principal components capable of explaining their variance. Traditionally, principal components are obtained from time series of data - i.e. the historical estimates. The data assimilation framework allows to incorporate historical information from one single historical demand matrix into the principal components of the problem. This means that, while previous approaches rely on historical estimates, in this case the Principal Components represent the historical (seed) matrix, which can be easily obtained with any demand model, from the gravity model to Synthetic Population. Given a single demand matrix x, we perturb the demand and artificially generate variations within the data. Different types of demand fluctuations are considered, such as spatial, temporal, and day–to–day variations. This allows us to use PCA–based algorithms, without even the need to first obtain the historical estimates, which is often infeasible in practice. Additionally, by artificially perturbing the demand in three different dimensions, the proposed approach allows to have control over the search space definition (e.g., define a narrow search space if small variations are assumed and hence good OD quality is retained in reference to the initial OD estimate; or a broader search space with more variance is considered in case the model error does not converge to a good solution).

2.2.1. Correlations among time–dependent OD flows

Dynamic OD demand is mostly represented as time-dependent OD flows $(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_h})$, which are individual sets of OD matrices $\boldsymbol{x}$ each representing a single time interval h. Demand fluctuations among such time–dependent OD flows can correlate in three possible dimensions. Figure 1 presents the conceptual directions for each of these three correlation dimension in a OD demand time series plot, where each vertical vector represents a single time–dependent OD for a given time interval. Further, we describe these dimensions as:

- **Spatial correlation:** The spatial correlation presents the spatial structure of the OD demand over the network, i.e., how all the OD pairs $x_{n_{ij}}$ are spatially correlated among themselves. This correlation dimension should help in capturing the demand fluctuations triggered spatially e.g., the changes in trip distribution among different OD pairs. The source of these fluctuations can variate from long–term changes of land-use to short–term changes in trip attractions and distributions among OD pairs due to consistently varying network travel times or traffic congestion patterns.

- **Temporal correlation:** The temporal correlation presents the times series evolution of demand, i.e., the time–dependent fluctuations of each OD pair $x_{n_{ij}}$ between all time intervals t (or previously said $\mathcal{H} = 1, 2, ...h$). This correlation dimensions helps in capturing the demand fluctuations or distributions for departure time choice of the overall demand for each OD pair. Individual departure time choice decisions depend on factors such as trip purpose/activity, network state/congestion and person demographics.

- **Day–to–day correlation:** Mobility demand is correlated to the demand for activities. As such, it follows a structure and day–to–day variations are likely to occur. Hence, day–to–day correlations presents the correlation of each OD pair $x_{n_{ij}}$ among different days d. This correlation dimension should capture the day–to–day demand fluctuations for individual OD pairs due to change in their trip generation/attractions for different trip activities which are influenced by e.g., day–of–the-week, weather conditions, seasons, special events like sales, festivals, sport events etc.

2.2.2. Historical data–set generation methods

After developing our understanding on the above mentioned correlation dimensions for time–dependent ODs, we consider that the demand fluctuations within the historical OD estimates should naturally follow these correlations. Hence, synthetic historical data–sets can be generated by perturbing the starting OD vector $\boldsymbol{x}$ among them. Since, these three correlation dimensions cover the possible user behaviors, we propose six different historical OD generation methods exploiting them. Intuitively, more correlations should lead to a more realistic representation of the behaviour. However, this will also requires a larger time series, which also means more principal components and therefore more variables to be calibrated. To mathematically express the proposed methods, we first define the utilized notations in table 1, followed by the definitions of all methods.

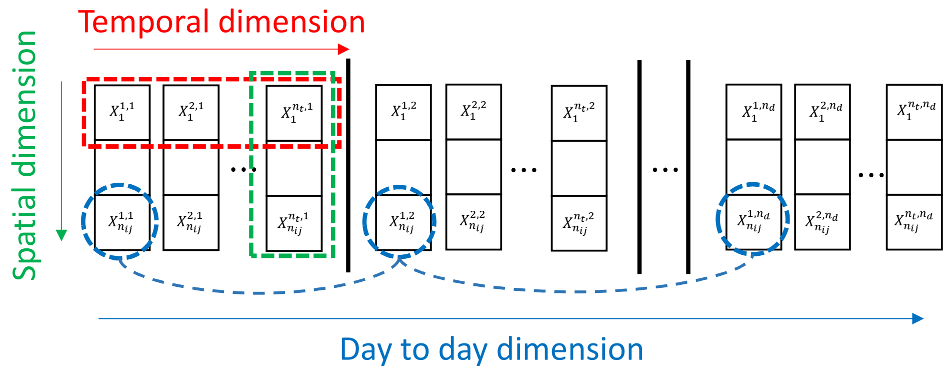


Figure 1: Different correlation dimensions among time–dependent OD flows

Table 1
List of Symbols

$\boldsymbol{D}$	Historical data matrix with dimensions $[n_{ij} \times (n_t n_d)]$
$\boldsymbol{\Delta}_T$	Perturbation matrix for correlation of type T
$\boldsymbol{x}$	Current/prior OD estimate matrix with dimensions $[n_{ij} \times n_t]$
$\boldsymbol{X}$	Augmented matrix of multiple x sets with dimensions $[n_{ij} \times (n_t n_d)]$
$\mathcal{N}_{od}$, $\mathcal{N}_t$, $\mathcal{N}_d$	Gaussian distributions of size n_{ij}, n_t and n_d, mean μ and standard deviation σ
R_{od}, R_t	Perturbation/weight coefficient for sizing the effect of spatial and temporal correlation variance
R_{min}	The smaller value within R_{od} and R_t
n_{ij}, n_t, n_d	Number of OD pairs, time intervals and historical days

- **Method 1: Spatial correlation**
  This method considers the spatial correlation to generate the historical OD data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta}_{od}$ is generated using $\mathcal{N}_{od}$ Gaussian distribution. The mathematical expression is given as:

$$\boldsymbol{D} = (\mathbf{1} + R_{od}\boldsymbol{\Delta}_{od}) \odot \boldsymbol{X} \tag{5}$$

where $\boldsymbol{X}$ is an augmented matrix given by:

$$\boldsymbol{X} = \underbrace{(x|x|\dots|x)}_{n_d} \tag{6}$$

while x is the initial OD estimate matrix of n_{ij} OD pairs and n_t time intervals. The $\odot$ operation achieves the Hadamard (element-wise) product to perturb the augmented matrix X. Note that, R_{od} is the perturbation factor for sizing the effect of perturbation matrix $\boldsymbol{\Delta}_{od}$.

- **Method 2: Temporal correlation**
  This method considers the temporal correlation to generate the historical data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta}_t$ is generated using $\mathcal{N}_t$ Gaussian distribution. The mathematical expression is given as:

$$\boldsymbol{D} = (\mathbf{1} + R_t\boldsymbol{\Delta}_t) \odot \boldsymbol{X} \tag{7}$$

where R_t is the perturbation factor for sizing the effect of perturbation matrix $\boldsymbol{\Delta}_t$.

- **Method 3: Spatial and temporal correlation**
  This method considers both spatial and temporal correlations to generate the historical data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta_{od,t}}$ is generated using the Gaussian distributions $\boldsymbol{\mathcal{N}_{od}}$ and $\boldsymbol{\mathcal{N}_t}$ in spatial and temporal directions (see fig. 1). The mathematical expression is given as:

$$\boldsymbol{D} = (\boldsymbol{1} + R_{min}\boldsymbol{\Delta}_{od,t}) \odot \boldsymbol{X} \tag{8}$$

  where R_{min} is the lowest of the perturbation factors R_{od} and R_t for sizing the effect of perturbation matrix $\boldsymbol{\Delta_{od,t}}$.

- **Method 4: Spatial and day–to–day correlation**
  This method considers both spatial and day–to–day correlations to generate the historical data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta_{od,d}}$ is generated using the Gaussian distributions $\boldsymbol{\mathcal{N}_{od}}$ and $\boldsymbol{\mathcal{N}_d}$ in spatial and day–to–day directions (see fig. 1). The mathematical expression is given as:

$$\boldsymbol{D} = (\boldsymbol{1} + R_{od}\boldsymbol{\Delta}_{od,d}) \odot \boldsymbol{X} \tag{9}$$

- **Method 5: Temporal and day–to–day correlation**
  This method considers both temporal and day–to–day correlations to generate the historical data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta_{t,d}}$ is generated using the Gaussian distributions $\boldsymbol{\mathcal{N}_t}$ and $\boldsymbol{\mathcal{N}_d}$ in temporal and day–to–day directions (see fig. 1). The mathematical expression is given as:

$$\boldsymbol{D} = (\boldsymbol{1} + R_t\boldsymbol{\Delta}_{t,d}) \odot \boldsymbol{X} \tag{10}$$

- **Method 6: Spatial, temporal and day-to-day correlation**
  This last method considers all possible correlation dimensions possible in time–dependent ODs. To estimate the historical data–set $\boldsymbol{D}$. The perturbation matrix $\boldsymbol{\Delta_{od,t,d}}$ is generated using the Gaussian distributions $\boldsymbol{\mathcal{N}_{od}}$, $\boldsymbol{\mathcal{N}_t}$ and $\boldsymbol{\mathcal{N}_d}$ in spatial, temporal and day–to–day directions (see fig. 1). The mathematical expression is given as:

$$\boldsymbol{D} = (\boldsymbol{1} + R_{min}\boldsymbol{\Delta}_{od,t,d}) \odot \boldsymbol{X} \tag{11}$$

The six proposed generation methods capture all possible combinations between spatial, within–day temporal and day–to–day temporal correlations. Note that, in current methodology we use Gaussian distributions with zero mean to define the perturbation matrices Δ_T but an additional value of these generation formulations is that these correlation distributions (currently $\boldsymbol{\mathcal{N}_{od}}$, $\boldsymbol{\mathcal{N}_t}$, and $\boldsymbol{\mathcal{N}_d}$) can be derived by other data sources, such as mobile phone network data and survey data. Finally, this leads to a framework that is more general - as it does not depend on an historical database - and is more flexible - as the structure of the PCs would reflect both OD flows as well as other spatial-temporal dynamics.

2.3. Simplification of DODE problem formulation

The DTA calibration problem is generally formulated as an optimization problem, minimizing the specified objective function by optimizing the model parameter values with the given constraints (to decide a feasible parameter space). A generic problem formulation for DTA model calibration is given as:

$$\underset{\beta,x}{Minimise}\ z(\boldsymbol{y}, \boldsymbol{y'}, \boldsymbol{x}, \boldsymbol{x^p}, \boldsymbol{\beta}, \boldsymbol{\beta^p}) \tag{12}$$

Where y/y' represent the observed/simulated traffic measurements, x and β indicate the current values for the origin–destination demand flows and for the behavioural parameters, respectively, while x^p and β^p are their historical (or prior) estimates. The traditional DODE problem focuses on only estimating time–dependent OD flows $\{x_1, x_2, ..., x_h\}$, while other model parameters β are kept constant. The objective function formulation for time-dependent DODE problem can be reformulated as:

$$\underset{x}{Minimise}\ \sum_{H}^{h=1}[w_1 z_1(\boldsymbol{y_h}, \boldsymbol{y'_h}) + w_2 z_2(\boldsymbol{x_h}, \boldsymbol{x^p_h})] \tag{13}$$

subject to:

$$\boldsymbol{y}'_h = f(\boldsymbol{x}_1, ...\boldsymbol{x}_h; \boldsymbol{\beta}; \boldsymbol{G}_1, ..., \boldsymbol{G}_h)$$
$$\boldsymbol{l}_x \leq \boldsymbol{x} \leq \boldsymbol{u}_x$$

where the calibration time period is defined in intervals $\mathcal{H} = \{1, 2, ..., H\}$ and:

$\boldsymbol{x}_h$: Time–dependent demand parameters i.e., OD flows

$\boldsymbol{y}_h$: Observed time–dependent traffic measurements

$\boldsymbol{y}'_h$: Simulated time–dependent traffic measurements

$\boldsymbol{\beta}$: Other fixed DTA model parameters

$\boldsymbol{x}^p_h$: Prior values for time–dependent demand parameters i.e., OD flows

$\boldsymbol{G}_h$: Road network and other supply parameters

The minimization of the DODE objective function (equation 13) heavily relies on z_1, which measures the goodness of fit between observed and simulated traffic measurements, while z_2 (i.e., the goodness of fit between estimated and prior OD demand) help to restrain the estimated solution closer to the prior/starting OD. The weight factors w_1 and w_2 are used to scale the reliance (or reflect uncertainty) on both observed traffic measurements y_h and prior OD flows x^p_h information. The simulated traffic data y'_h detected in time interval h are explicitly modelled through a (non-linear) function $f(\cdot)$ (i.e., DTA simulator) of all OD flows x, model parameters β and the road network/supply parameters till time interval h. Using this optimization–based problem formulation with any non-assignment based approach provides an advantage of including any available traffic data y_h for estimation (requiring $f(\cdot)$ to be a DTA simulator).

Since the nature of DODE problem is highly underdetermined (far more estimation variables against traffic measurements), reliance on using z_2 can be seen throughout the literature because its keeps the calibrated OD solution close to the prior/starting estimate, considering it the most reliable available estimate. For PCA–based models, we propose to simplify the DODE formulation releasing the z_2 error term. This is a generalization of PCA–based models where the use of PCs help us include historical OD information in the objective function, allowing us to release z_2 from equation 13 and simplify the DODE problem formulation (equation 14). It also allows to simplify the problem through dimension reduction (as we solve it in PC space). The new simplified problem formulation is given as:

$$\underset{x}{Minimise} \sum_{H}^{h=1} [z_1(\boldsymbol{y}_h, \boldsymbol{y}'_h)] \tag{14}$$

subject to:

$$\boldsymbol{y}'_h = f(\boldsymbol{x}_1, ...\boldsymbol{x}_h; \boldsymbol{\beta}; \boldsymbol{G}_1, ..., \boldsymbol{G}_h)$$
$$\boldsymbol{l}_x \leq \boldsymbol{x} \leq \boldsymbol{u}_x$$

This simplification is possible only by the use of PCA, where previously the standard approach (equation 13) requires the term z_2 to include prior information about the historical demand. This information, however, is already included within the PCA components, where the vector of Principal Components $\hat{V}$ is in fact directly obtained by the time series historical demand, which means that the PCs defined search space is already constrained within the variance present in the historical estimates. This keeps all the patterns of the calibrated OD estimate within those present in historical estimates, which is also the purpose of using the error term z_2. Hence, for all PCA–based methods, the purpose of using the error term z_2 is already fulfilled by PCA's dimension reduction.

2.4. Estimation setup

As discussed before in section 1, SPSA is arguably the most popular assignment matrix–free method due to its generalized problem formulation and ability to deal with non–linear and stochastic systems. Therefore, to demonstrate the significance of the proposed PCA methods, we choose it as the optimization problem solver to estimate the DODE problem formulated in PC space (Qurashi et al., 2019). Below, we describe the SPSA setup for PCA–based DODE and emphasize on the ease in requirement of defining SPSA hyper–parameters alongside proposing some modifications to exploit the properties of PCA application. Similarly, we also discuss the PCA application setup to understand the role of new hyper–parameters required to define the characteristics of historical data matrix and dimension reduction.

2.4.1. SPSA for PCA–based estimation

SPSA (Spall et al., 1992) is a Stochastic Approximation (SA) algorithm with a unique advantage of approximating a noisy gradient with only two objective function evaluations using simultaneous perturbation. Qurashi et al. (2019) proposed a modified SPSA to solve PCA–based DODE problem. Equation 15 shows the modified gradient estimation method to estimate PC–scores z, where Δ is a p-dimensional vector generated randomly from a ± 1 Bernoulli distribution (where P is the length of decision vector $\boldsymbol{z_k}$).

$$\boldsymbol{g'} = \frac{f(\boldsymbol{z_k} + \boldsymbol{z_k} \times c_k \boldsymbol{\Delta_k}) - f(\boldsymbol{z_k} - \boldsymbol{z_k} \times c_k \boldsymbol{\Delta_k})}{2c_k} \left[\Delta_1 \; \Delta_2 \; .. \; \Delta_p\right]^T \tag{15}$$

The estimated gradient is used to minimize the solution using a modified form of SA approach (equation 16).

$$\boldsymbol{z_{k+1}} = \boldsymbol{z_k} - \boldsymbol{z_k} \times a_k \boldsymbol{g'_k}(\boldsymbol{z_k}) \tag{16}$$

$$c_k = c/k^{\gamma} \qquad a_k = a/(k+A)^{\alpha} \tag{17}$$

Note that, the coefficients of perturbation c_k and minimization a_k evolve over the number of iterations $\mathcal{K} = \{1, 2, 3, ...k\}$ and are evaluated based on the set of pre–defined hyper–parameters c, a, γ, α, and A (equation 17). Apart from the general guidelines proposed by Spall (1998), their is no set rule to define these hyper–parameters for SPSA or any of its variants. Hence, it requires a trail–based method to find appropriate values which can result in good convergence. When combining PC-SPSA with the data–set generation method proposed in Section 2.2.2, the number of hyper–parameters further increases, as the model requires to define both the number of historical estimates n_d as well as the mean and the variance for the spatio/temporal distributions $\boldsymbol{\mathcal{N}_{od}}, \boldsymbol{\mathcal{N}_t}, \boldsymbol{\mathcal{N}_d}$, which regulate the link between historical demand and PCs. However, the application of PCA drastically reduces the required number of iterations (Qurashi et al., 2019) and the modified SPSA (equation 16) applies a percentage change instead of absolute increase/decrease in estimation variables $\boldsymbol{z_k}$, as in the traditional SPSA. Therefore, the sensitivity of the model to changes in the hyper–parameter decreases significantly. as shown in Section 4.1. Additionally, by combining the proposed data–set generation method with the simplified formulation discussed in Section 2.3, the number of iterations of PC-SPSA futher decreases making the calibration of the parameters γ, α and A unnecessary, as the model converges for a low value of k. Finally, SPSA requires multiple gradient replications for DODE (Balakrishna et al., 2007b) and almost all SPSA based literature works use it to reduce gradient bias (e.g., Cantelmo et al. (2014a); Tympakianaki et al. (2015)) due to correlations and non–linearity present in DODE variables. We show in this paper that this becomes unnecessary with PCA because all PCs are orthogonal and uncorrelated. Hence, we also propose to remove this requirement and all experiments ran in this paper use only a single gradient estimate per SPSA iteration. A step–wise PC–SPSA algorithm is given in appendix 1.

2.4.2. PCA application setup

Recalling from section 2.1, to transform the OD flows in lower dimensional space, PC–directions V^T are used. These PC–directions are evaluated from the historical data matrix D (see equation 1) and represent the variance present in it. Note that, the optimization search space for PCA–based methods is confined within this variance. In other words, it is the additional demand information added to the DODE objective function. Hence, it is important to better understand the impact of this added variance information and control it characteristics accordingly. The variance present in PC–directions can be controlled by certain parameters which define the characteristic of historical data matrix D. These parameters include the number of historical estimates i.e., size n_d of data matrix D (equation 6), number of PCs retained n_v (equation 2), and control of the variance present in historical estimates (defined by R and σ from equation 5-11, i.e., in case of using historical generation methods). Note that, both in case of availability or unavailability of historical OD estimates, the variance information can be controlled. But, it also increases the overall set of required hyper–parameters for manual setup.

3. Case study: Munich city

3.1. Experimental setup

3.1.1. Network and simulation setup

We implement the case study on the Munich regional network (about 900 km^2). As shown in Figure 2, the network is divided into 73 zones resulting in 5,329 OD pairs, including 10 external zones (green circles) at major radial motorways entering the city. The network consists of a total of 8,761 links (Figure 4), excluding residential roads to reduce the route choice burden for the simulation experiment. A total of 507 detector locations are used for the case study. As described previously, this leads to a highly *underdetermined system* (5,329 *unknowns* per interval with only 507 *traffic measurements*) and renders the application difficulties of conventional calibration methods.

An open–source traffic simulator, Simulation of Urban MObility (SUMO, Lopez et al. (2018)), is assembled with the proposed calibration algorithm for experiments. All simulations are implemented at the mesoscopic level via the trip–based (one–shot) stochastic user route choice assignment method. To focus on DODE problem, we fix the route choice and supply side parameters (e.g., jam threshold). Also, to cater for the stochasticity of the traffic simulations we used outputs averaged from 10 simulation replications. Overall, the run–time for a single simulation (for morning peak hours i.e., 6am - 10 am) is 12 minutes and the 10 simulation replications are run in parallelization. One iteration of SPSA needs minimum 2 simulation run–time of 24 minutes. Given the sizes of the network, SPSA cannot be used to calibrate the DTA model. Additionally, historical estimates are not available as the network has never been calibrated before. Therefore, we use the procedure explained in section 2.2.2 combined with the PC–SPSA algorithm to calibrate the network under the simplified problem formulation, described in section 2.3.

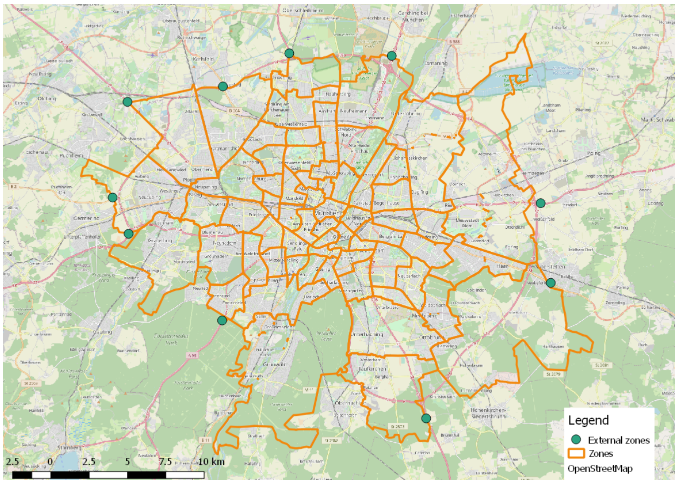


Figure 2: Traffic zones of Munich major region.

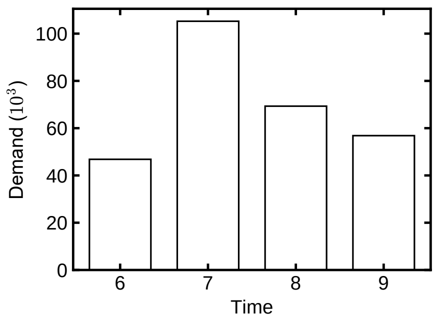


Figure 3: Network Demand (6 am to 10 am)

3.1.2. Demand scenarios

To explore the effectiveness and efficiency of PC–SPSA on the DTA model calibration problem, we apply PC–SPSA to calibrate the demand from 6 am to 10 am represented in 15–minute intervals, which contains characteristics of very low demand (6 am – 7am), normal off–peak (8 am – 10 am) and peak traffic (7 am – 8 am). To process the procedure, we specify the demand scenario following the benchmarking framework standardized by Antoniou et al. (2016) for testing calibration algorithms. The method has also been used in many recent works on developing calibration algorithms (Qurashi et al., 2019; Cantelmo et al., 2020). To create the scenario, the target/true demand is synthetically perturbed with the latest previous estimate x_{p_1} and its simulated outputs are taken as true outputs. Two coefficients of reduction (Red) and randomization ($Rand$) are used for perturbation. Different values of these two coefficients are used to create different types of true demands as in reality. The demand scenario generation is specifically expressed as:

$$x_c = (Red + Rand \times \delta) \times x_{p_1} \tag{18}$$

where δ is the random perturbation vector following Gaussian distribution. In this case study, we apply $Red = 0.7$ and $Rand = 0.15$ (i.e., $x_c = (0.7 + 0.15\delta) \times x_{p_1}$), and $\delta \sim N(0, 0.333)$ (99.7% of values located in [-1,1]), resulting in the demand distribution shown in Figure 3 (aggregated into one hour for easy illustration).

Figure 4: Used Munich Network overview

3.1.3. PC–SPSA algorithm settings

Recall Equation (17), we need to update the gains for perturbation (c_k) and minimization (a_k) to control the step size and convergence at each step. In all following experiments, A, α and γ are set to be 25, 0.3 and 0.15, respectively. For the experiments within this section, c and a are set to be 0.15 and 1, respectively. Note that, c and γ control the perturbation percentage of the PC–scores. For example, at the first step, the PC-scores are perturbed with $\pm(15\%)$. On the other hand, a, A and α control the actual moving step in the searching space. All historical data–set generation methods introduced in Section 2.2 are applied for comparison, for which R_{od}, R_t, and R_d are set as 0.3, 0.4 and 1, respectively, while the Gaussian distributions $\mathcal{N}_{od}$, $\mathcal{N}_t$, and $\mathcal{N}_d$ are generated using $\sim N(0, 0.333)$ setting. The demand of 100 historical days is thus generated. Furthermore, to reserve enough variance contained in the historical data–set for tracking the patterns and achieve the goal for dimension reduction at the same time, the number of PCs expressing 95% of the total variance are used.

3.1.4. Goodness of fit

Given that PC–SPSA is a non-assignment matrix based algorithm it requires the DTA model simulation to map the OD matrix into measurable traffic measurements, such as vehicle counts recorded by detectors. These generated traffic counts are then compared with the observed traffic counts to evaluate their difference which is used as an indicator for DODE minimization (i.e., z_1 in equation 14). In this study, we apply Root Mean Square Normalized error (RMSN) to measure the Gof of the simulated traffic counts and thus evaluate the estimated OD matrix. RSMN is specifically used extensively for DODE problem (Qurashi et al., 2019; Antoniou et al., 2015) because it finds the normalized root mean distance between all counts helpful to estimate closer patterns towards the target solution. The calculation of RMSN is given by:

$$RMSN = \frac{\sqrt{n \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}}{\sum_{i=1}^{n} y_i} \tag{19}$$

where y and $\hat{y}$ are the observed traffic counts and simulated traffic counts, respectively. n is the number of detectors.

3.2. Results

3.2.1. Convergence analysis and calibration quality

Figure 5 displays PC–SPSA's convergence results for calibrating 15-minute demand intervals of the peak hour from 7 am to 8 am as shown in fig. 3. The results include convergence plots for all six historical OD generation methods described in section 2.2.2. Despite the large study area, PC–SPSA is able to converge to a low RMSN error values within the first few iterations, confirming the improved application scalability of PC–SPSA on large scale DTA models. Figure 6 illustrates the quality of model calibration comparing observed and simulated traffic counts at all detector locations using a 45° plot. The results depicted are only for method 6 (figure 5). We refer to section 3.2.2 for the discussion on the differences between the six method. Since all points are aligned closer to the 45° line, it is confirmed that the low error convergence is achieved at all detector locations. Figure 5 also shows that, while all generation methods perform fairly well, some of the proposed methods obtain drastic improvements in only one or two iterations.

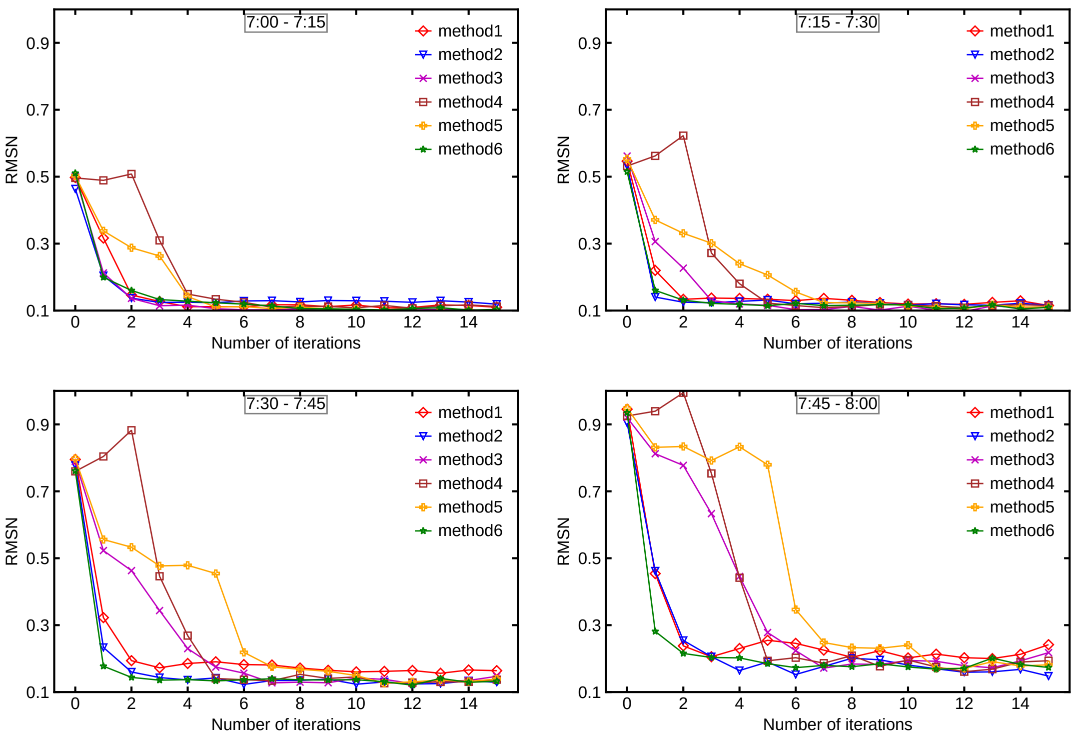
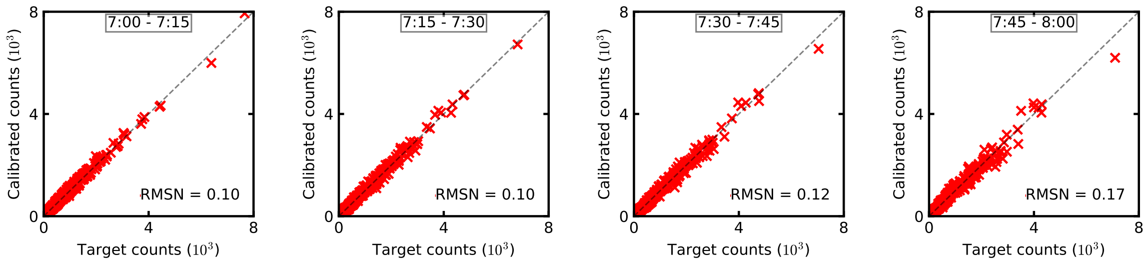


Figure 5: Comparison between generation methods for specific intervals.

Figure 7 and 8 are also plotted for method 6 and depict the quality of calibrated OD matrices by comparing it with the target and initial OD matrices on 45° plots. Overall, PC–SPSA is able to find a good quality solution and as per the property of PCA application (i.e., confining search space in historical OD variance) all OD pairs are close to the 45° line. Note that PC-SPSA is able to calibrate the reduction change of the target demand (i.e., plots in figure fig. 7 are around the 45 %) but it is not able to entirely converge the error due to random fluctuations (*Rand* in equation 18). This is an expected result when using PCA, as the PCs constraint the search space allowing for limited structural changes in the OD demand matrix. To understand better this behavior, we conduct a sensitivity analysis for different demand scenarios in section 4.2), and also compare the results from different historical OD generation methods which actually do behave differently for converging the random fluctuations (section 3.2.2).

3.2.2. Comparing different historical OD generation methods

Figure 5 deploys PC–SPSA's convergence results using all historical OD generation methods and despite that all methods show different converging speed, they can converge to almost the same level of error. This indicates restricted requirements and robustness of PC–SPSA on the historical OD estimates with respect to final error convergence.



(a) 7:00 - 7:15 (b) 7:15 - 7:30 (c) 7:30 - 7:45 (d) 7:45 - 8:00

Figure 6: Comparison of target and calibrated traffic counts.

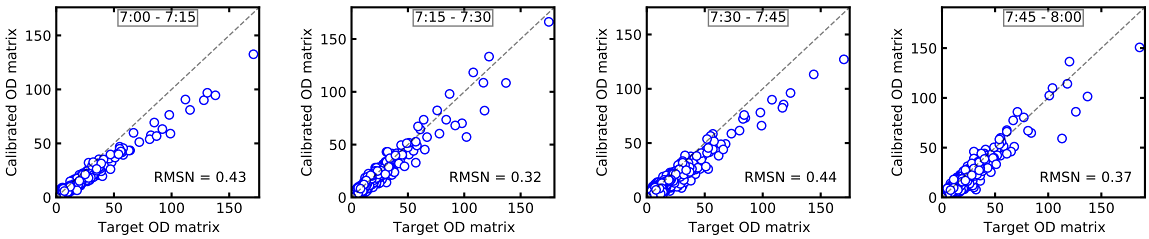


Figure 7: Comparison of target and calibrated OD matrices.

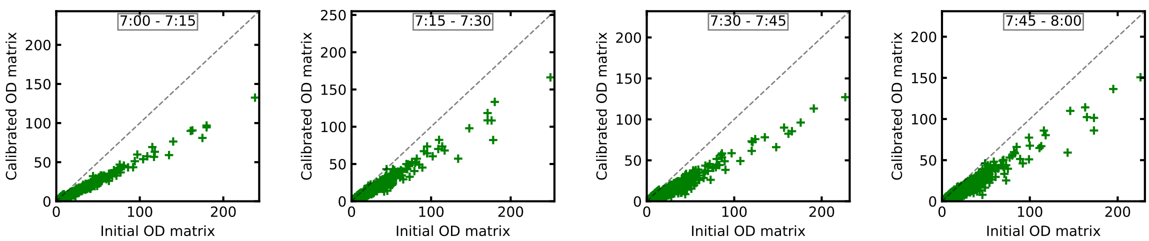
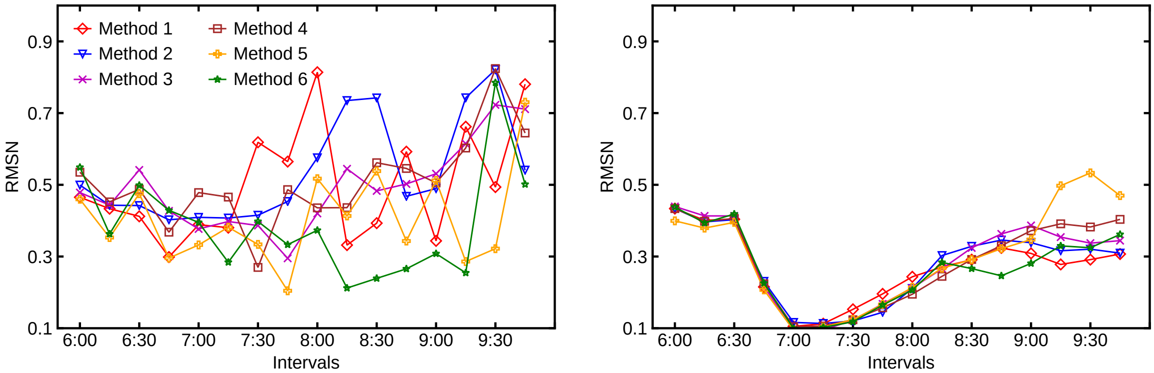


Figure 8: Comparison of initial and calibrated OD matrices.

However, in terms of the converging speed, the method capturing most correlations (method 6) and the methods considering only one–dimension correlation (method 1 and 2) outperform the others. For the latter, it is easy to understand as searching the pattern in a single correlated direction would be faster because the defined search space have more noise and randomness (local minimums). In contrary, when the correlations of two of three dimensions are fused (method 3, 4 and 5), they construct the search space with more accurate and sufficient information. Although the noise and randomness is reduced, now its presence probably hinders the SPSA algorithm to struggle finding the minimized solution. Surprisingly, method 6 which combines three dimension information, however, also leads to a fast convergence as method 1 and 2. This behavior may be due to the expectation that the space constructed by this method is more comprehensive and thus it directs the algorithm to find a faster direction compared with the ones with only two–dimensional information.

To better understand the above stated comparison, we further compare all the historical OD generation methods by their calibration quality. Figure 9 illustrates the quality of calibration for all generation methods with fig. 9(a) showing the quality of calibrated OD (RMSNs comparing with target OD) and fig. 9(b) showing the final convergence error achieved for the whole demand period. Moreover, as mentioned previously, literature efforts only considered temporal correlations for historical OD generation i.e., method 2, and hence we also show comparison of its calibrated OD with



(a) Best RMSN for OD matrix

(b) Best RMSN for counts

Figure 9: Comparison between all generation methods.

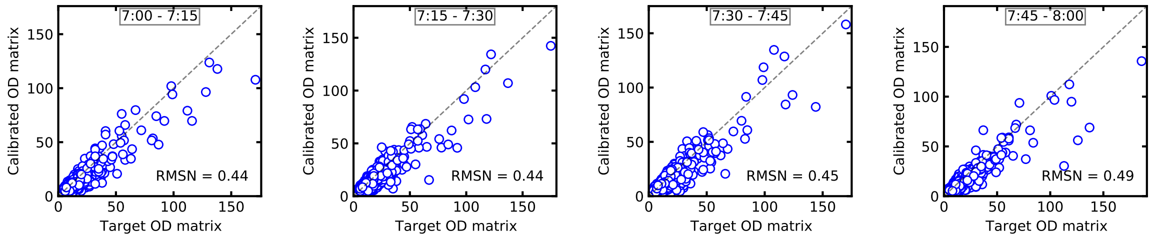


Figure 10: Comparison of target and calibrated OD matrices (method 2).

the target OD in figure 10. By analyzing fig. 9(a), we can validate the above mentioned arguments about the effects of using more correlated information in generation methods. In general, considering multiple correlations leads to a reduction in the demand error (fig. 9(a)) and a similar error in term of traffic counts (fig. 9(a)). Method 6, the one considering the highest number of spatio/temporal correlations, not only shows a faster convergence but it is also the most consistent in terms of OD calibration quality (i.e., least RMSN error from target OD). At the other end of the scale, the methods considering only one correlation dimension (method 1 and 2) are the most inconsistent with poor quality OD estimates (see time intervals from 7 to 9 am), meaning that the faster convergence is mostly due to the model over fitting the data. Especially, figure 10 shows that the calibrated OD from method 2 is more scattered as compared to figure 7 for method 6 (further comparison of the calibrated OD quality for method 2 and 6 is shown in section 4.2). The methods with two correlations (method 4 and 5) have a medium range of OD quality. Perceiving these results, it can be established that creating the OD estimates with more correlation information helps in better calibration quality and having lesser random perturbation or noise also pushes towards faster convergence. Lastly, analyzing fig. 9(b), it can be seen that all different historical OD generation methods are able to eventually converge on very similar RMSN errors, validating the robustness of PC–SPSA algorithm convergence performance with different methods.

3.2.3. Conventional versus simplified problem formulation

Simplified problem formulation removes the error term z_2 (between the calibrated and prior OD) from the conventional problem formulation equation 13. This is similar as setting the w_2 weight as 0 % in equation 13, which otherwise if set as w_2 >0 is following the conventional problem formulation. Figure 11 shows the convergence performance of PC–SPSA at different weight settings (i.e., 0%, 20%, 40% and 60% weight w_{od} for z_2). Similarly, figure 12(b) shows the least RMSN error achieved for traffic counts and figure 12(a) shows the OD solutions' quality for all different weight settings. It is clearly evident that the simplified problem formulation outperforms all other weight settings for much faster convergence towards the least RMSN error. Another surprising outcome is from figure 12(a) where the simplified problem formulation also results good OD solution quality consistently. Only 20% w_{od} gives better solution quality for some intervals but this comes at the cost of an increased error in the traffic counts.

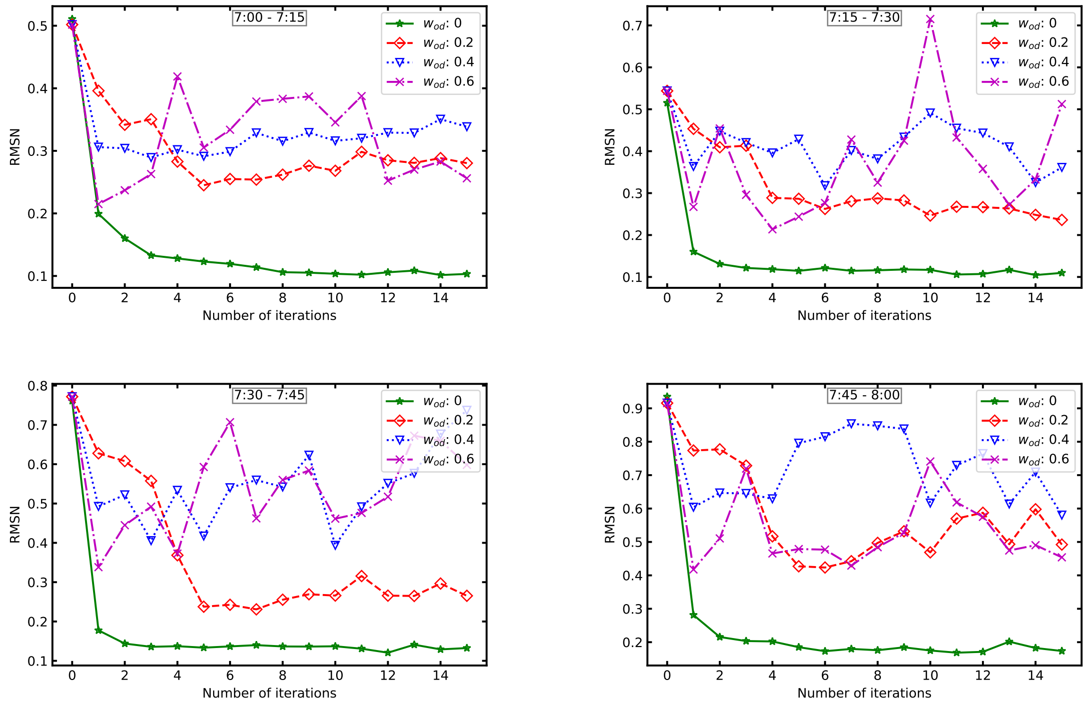
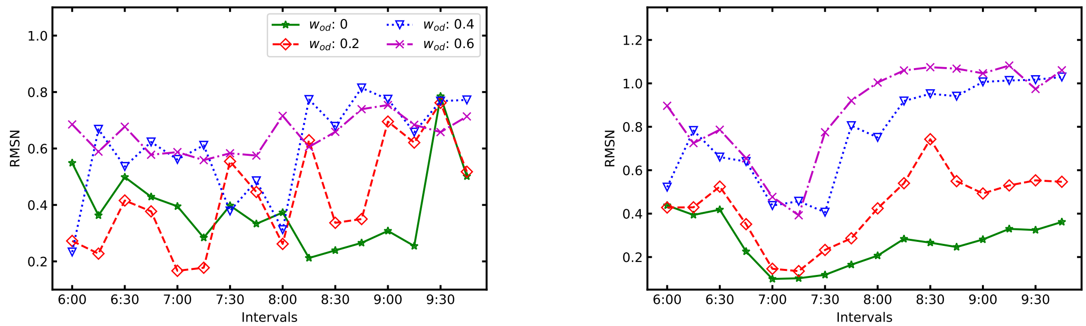


Figure 11: Comparison between objective weights for specific intervals.



(a) Best RMSN for OD matrix

(b) Best RMSN for counts

Figure 12: Comparison between different weights combination in the objective function.

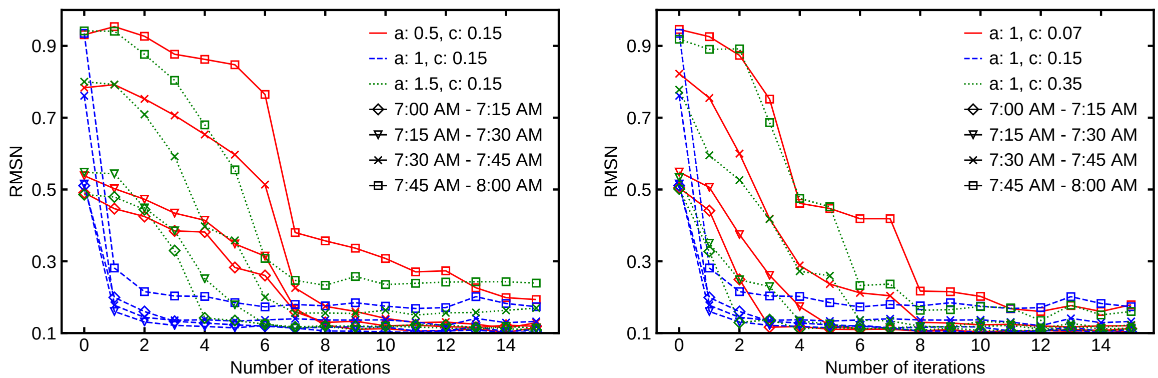
Note that, all the mentioned results confirm that we can utilize the benefits of using PCA application (i.e., limiting SPSA search space within the variance of historical OD estimates) for simplifying the DODE objective function. Since PCA application adds the required demand information in the PCs, further constraining the calibrated OD with prior/starting OD will have a double restraining effect adding unnecessary burden in the objective function. Moreover, even adding the weight w_{od} does not result in better OD solution quality indicating that PCA includes the OD information in a more structural way. Note that, as we increase the w_{od} weight for OD error term, the performance of the algorithm detoriates either it is in terms of convergence (figure 11), the least RMSN error (figure 12(b)) or the OD solution quality (figure 12(a)). Lastly, Figure 7 and 8 (plotted for method 6) also provide supplementary results for simplified problem formulation (showing the quality of calibrated OD matrices by comparisons with the target and initial OD matrices on 45° plots), where both plots show that the patterns of calibrated OD estimates are well estimated and are close to the target solution.

4. Sensitivity analysis

In this section, we perform sensitivity analysis on PC–SPSA with respect to SPSA parameters, demand conditions, and quality of historical estimates, respectively. The historical estimates are generated using method 6 (as per our analysis in section 3.2.2). Note that, the other parameters not specifically mentioned here remain the same as that in the previous section.

4.1. Robustness against SPSA parameters definition

In this section, we analyze the robustness of PC–SPSA against definition of SPSA hyper–parameters. SPSA is a random search stochastic algorithm and requires an appropriate definition of its hyper–parameters. These hyper–parameters can vary significantly for different problems and don't have any universally identified set of values (guidelines are given by Spall (1998)). Since SPSA parameters are only defined by trial–and–error method during implementation, we observe its sensitivity for the PC–SPSA algorithm. Figure 13 shows the convergence plots for calibrating the Munich network case study with different set of c and a hyper–parameters. c is used for defining the perturbation step size, while a is used for minimization step (equation 17). Analyzing the results from fig. 13, PC–SPSA appears to be significantly less sensitive to varying SPSA hyper–parameters. The values used for both c and a vary significantly since they act as a percentage change instead of an absolute change. Although, the convergence rate is different among these hyper–parameter settings, all experiments converge to the a similar RMSN error value within a few iterations.



(a) Robustness against a

(b) Robustness against c

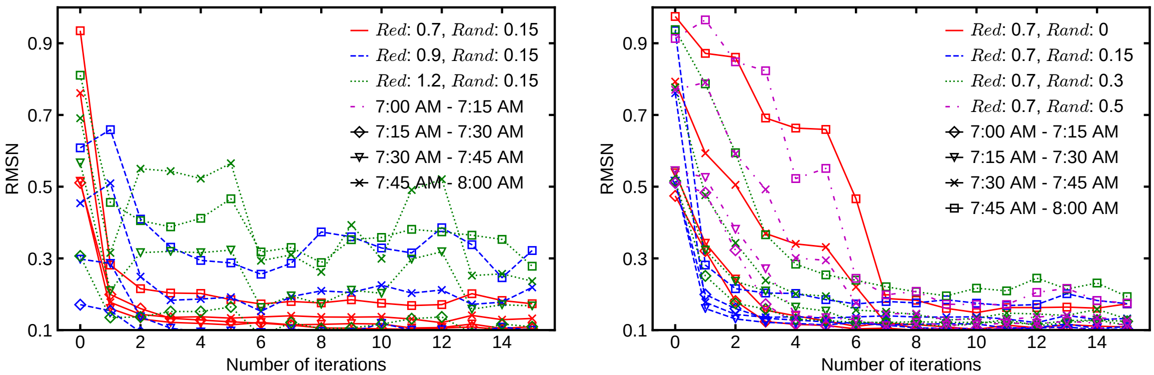
Figure 13: Comparison of using different SPSA parameter values (c and a).

We consider two reasons for PC–SPSA robust behavior, 1) the hyper–parameters act as the percentage change in perturbation and minimization (equation 15); and 2) faster convergence of PC–SPSA and properties of PC scores vector (i.e., very few estimation variables with even lesser being more significant). Also, since the rest of SPSA hyper–parameters i.e., γ, α and A are used for evolving the gain sequence parameters over the number of iterations, we do not add their sensitivity analysis as PC–SPSA converges in a handful number of iterations; making it insensitive to their definition (we use the default values given by Spall (1998)). Overall, we can establish that PC–SPSA being robust, requires significantly less manual input or trail–and–error method for setup.

4.2. Performance in different traffic conditions and demand fluctuations

In this section, we analyze the performance of PC–SPSA in different traffic conditions and demand fluctuations. More specifically, we define different demand scenarios using eq. (18) and analyze PC–SPSA convergence. Here its noteworthy to mention that, the historical demand matrix D is created using method 6 (section 2.2.2) with R_{min} as 0.3 and $\Delta_{od,t,d} \sim N(0, 0.333)$.

Figure 14(a) shows the PC–SPSA performance under different network conditions, where *Red* coefficient (from eq. (18)) are set to 0.7 (70%), 0.9 (90%) and 1.2 (120%) in reference to starting/current OD matrix while keeping the *Rand* coefficient constant as 0.15 (15%). These set of variables result in target demands with three different traffic conditions i.e., less–congested, normal/congested, highly congested. Analyzing fig. 14(a), PC–SPSA converges well for the first two scenarios converging to a low RMSN error, but struggles to calibrate the highly congested scenario. The zig–zag behavior of its convergence is due to the use of traffic counts in congested state, which adds more noise in the objective function. This is a known result for demand calibration and this is why, for practical implementation, it is suggested to always use a matrix that is less congested than the target one. This can be easily done by comparing the simulated and observed traffic data. Still overall, PC–SPSA is able to converge the RMSN errors for all different traffic conditions.



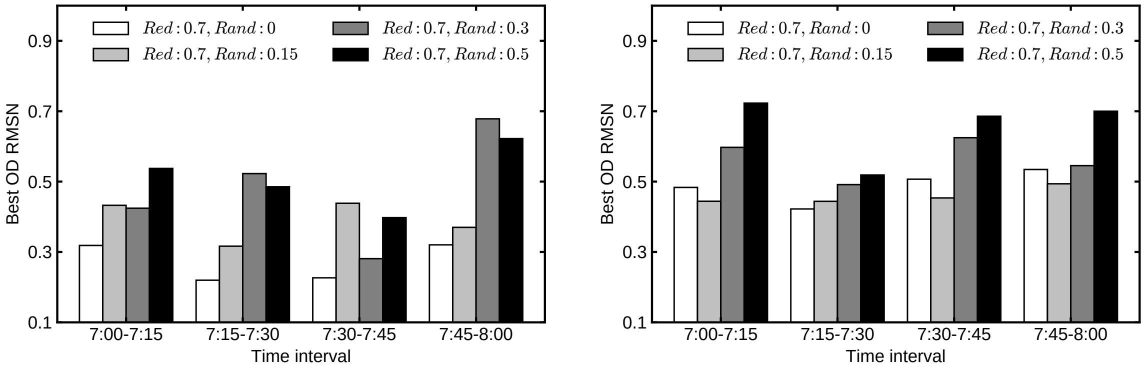
(a) Reduction sensitivity

(b) Randomness sensitivity

Figure 14: Demand scenarios sensitivity (method 6)

Similarly, fig. 14(b) shows PC–SPSA performance while calibrating against different magnitudes of random fluctuations in target demand generated using multiple *Rand* values in *eq*. (18), while fig. 15(a) illustrates the subsequent OD solution quality for all scenarios. As mentioned above the D historical data–set is generated with R_{min} as 0.3, hence the target demand generated equal or above $Rand = 0.3$ should contain more significant demand fluctuations than what are present in D data–set. Analyzing the results from fig. 14(b) and fig. 15(a), PC–SPSA using method 6 with 30% R_{min} is able to converge all demand fluctuations scenarios resulting in a low RMSN error but with varying solution quality (i.e., RMSN between calibrated and target OD). Comparing the scenarios results individually, $Rand = 0$ scenario has the target demand without any pattern changes and gets the best OD solution quality but PC–SPSA convergence is quite slower because the algorithm is still directly perturbing the OD patterns hence it also requires a few iterations to get back to closer solution (a reduced clone of initial OD). A similar convergence trend can been seen in $Rand = 0.5$ scenario, since the target demand patterns are highly fluctuated and is even more than the variance within historical demand D, hence it requires more time for converging to a low RMSN error and with poor OD solution quality (i.e., the possible solution within the variance of historical estimates satisfying the traffic measurements).

Note that, with the increase in *Rand* values both the OD solution quality and algorithm convergence performance deteriorates because the target solution has more demand fluctuations (i.e., higher *Rand* component) from initial OD. Hence, we can say that overall PCA–based methods have limited performance against estimating higher random demand fluctuations especially because the OD solution quality deteriorates significantly. Furthermore, figure 15 also compares the OD solutions' quality for method 2 and 6, where the latter is able to result better OD solutions consistently against all scenarios. This comparison validates the argument that using all three correlations dimensions (method 6) helps in establishing the search space more structurally around the initial OD. It is also noteworthy to mention that, the fact that PC–SPSA has limited performance against random demand fluctuations also signifies the importance of the proposed data–assimilation framework which allows derivation of the correlations from other data sources to form more realistic search space for PCA–based calibration.



(a) Method 6

(b) Method 2

Figure 15: Best OD RMSN of scenarios with different randomness.

4.3. Historical estimates setup

As already established, all PCA–based methods heavily rely on the quality of historical estimates. Previously, in section 2.2, we proposed a data assimilation framework to create estimates from an initial historical matrix for scenarios where they are unavailable or irrelevant. These established generation methods should also allow to control the quality of historical estimates and calibrated OD solution (in reference to starting/available OD estimates). In this section, we explore the effects of historical data–set D generation variables i.e., n_d the number of days historical data–set contains, R_{min} for resizing the variance within historical estimates and σ (standard deviation) for Δ (i.e. the correlated random matrix) defining the shape of variance.

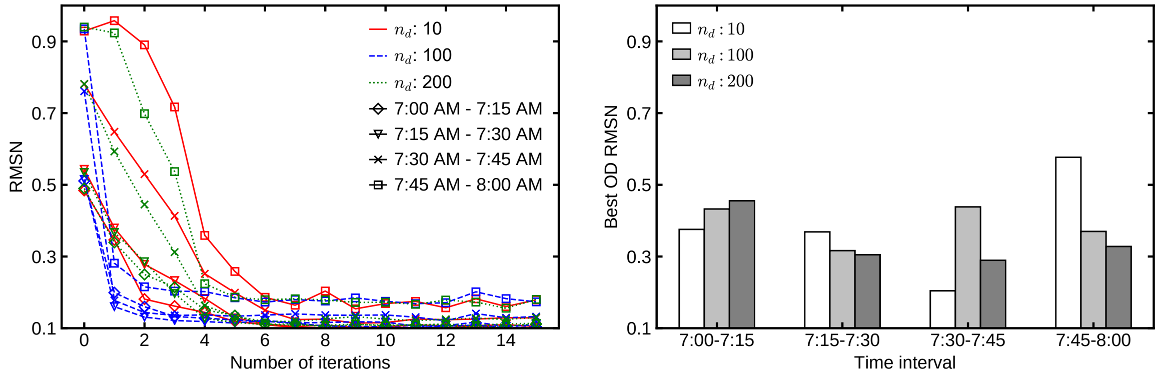
4.3.1. Size of historical data–set

The number of historical observations n_d is an additional parameter to be calibrated when using PCA in the context of the DODE. Figure 16 illustrates the PC–SPSA performance upon using three different sizes of D data–set. Analyzing these results, it is evident that the size of D data–set influences the convergence plots (fig. 16a) as if the n_d is too small or large, the convergence gets slower. Comparing the OD solution qualities for different D data–set sizes (fig. 16b), the increase in size seems to improve both the consistency and quality of estimated OD solution. The convergence results can be explained such that the size of D data–set defines the amount of variance which if is too small or large the algorithm needs more iterations for convergence, while given an appropriate set of n_d historical estimates, the algorithm performs faster. This is proven by the fact that for $n_d = 10$ both the convergence results and OD solution quality show larger fluctuations while on the other hand, a larger number of observations ($n_d = 200$) shows a much more consistent quality, which is explained by capability of the model to better incorporate the structure of the demand. Overall, it can be established that small size of D data–set contains less variance directing the algorithm to converge slower and with random OD estimate quality, while as the number of observations in D data–set increase the amount of variance generated also increases which till a certain optimum value improves convergence but later with further increase the convergence requires more time due to larger search space. But enlarging the variance or search space always helps to improve the consistency in OD solution quality.

4.3.2. Variance within historical data–set

Next, we perform the sensitivity analysis on defining the variance of historical data–set D. Different set of values are used for R_{min} and σ (i.e., the standard deviation for the Gaussian distributions defining Δ_T coorelation) to generate historical data–set using method 6. Note that, the effect of changing both R_{min} and σ is quite similar with a minor difference, where R_{min} widens/shrinks the shape of Gaussian distribution with increasing/decreasing the values of random distribution, σ directly effects the distribution of random numbers. We also perform the analysis for calibrating two different target demand fluctuations, setting $Rand$ in eq. (18) as 0.15 and 0.3, while the σ is set to 0.333 (i.e., $\delta \sim N(0, 0333)$. Figures 17 and 18 illustrates the convergence plots for both demand scenarios subsequently, while Figures 19 and 20 show the OD solution qualities for varying R_{min} and σ experiments.

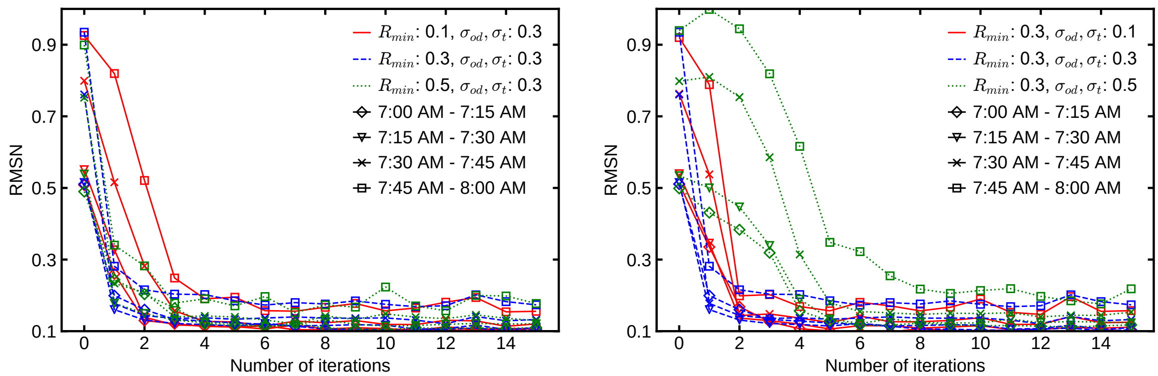
First analyzing the effect of varying R_{min} values, the calibration convergence plots are similar to the demand



(a) Traffic counts RMSN (b) OD RMSN

Figure 16: Historical data matrices size sensitivity.

fluctuation experiment from fig. 14(b) i.e., for scenarios where $R_{min} > Rand$ the convergence is much faster (see $R_{min} = 0.5$) and for $R_{min} \leq Rand$, the convergence is slower (see $R_{min} = 0.3$ for $Rand = 0.3$ scenario). While fig. 19 illustrates that lower R_{min} setting results in better OD solution quality and as we increase R_{min}, the error between target and calibrated OD also increase. The performance for varying R_{min} is consistent with the previous results from section 4.2, i.e., if we use larger values, the variance space increases and the algorithm converges faster but to a poor quality solution (see fig. 19). Hence, given the results it can be said that the use of lower values for R_{min} is more efficient unless either the solution is not converging and more variance space is required or a faster convergence is desired.

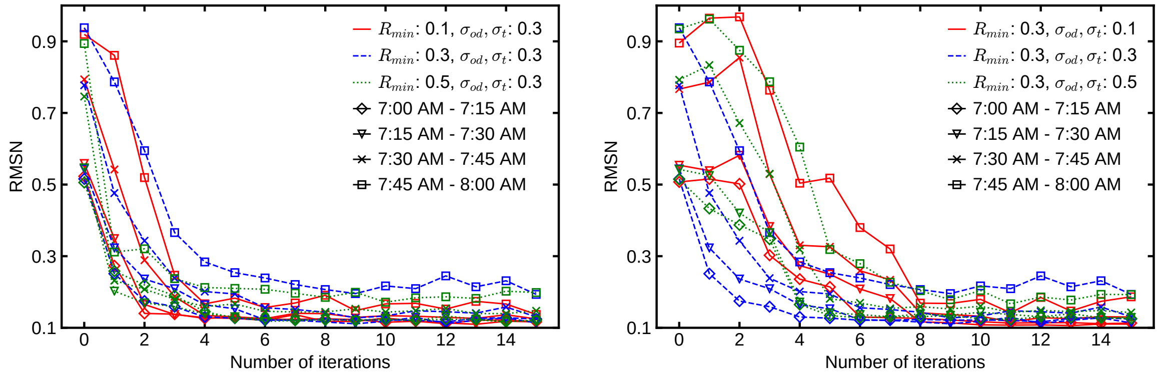


(a) Robustness against R_{min} (b) Robustness against σ_{od}, σ_t

Figure 17: Demand scenarios sensitivity (scenario: $Red = 0.7, Rand = 0.15$).

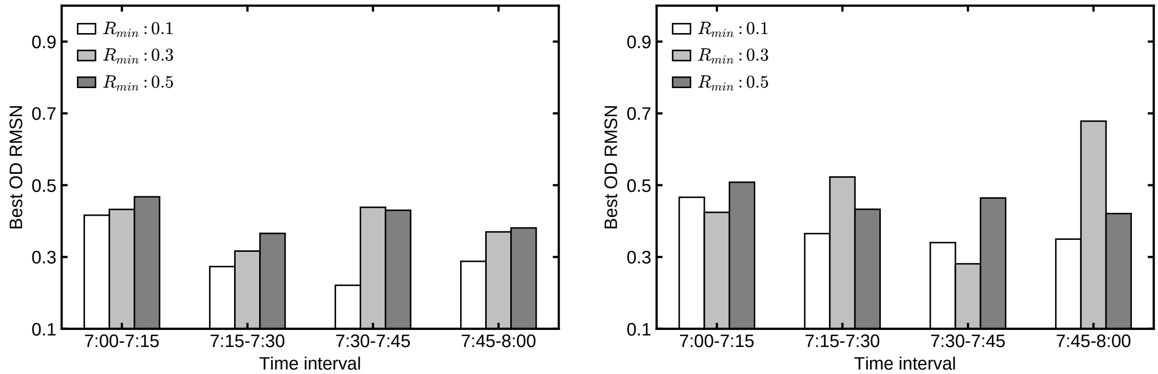
Next, analyzing the effect of varying σ values, the algorithm convergence is slower for both the smaller and larger σ values and is more optimum for middle value of $\sigma = 0.3$. Considering the OD solution qualities, note that similar to R_{min}, lowest value of σ result in the best calibrated ODs relative to the target solution. Hence, to achieve better calibration efficiency in solution quality, lower amount of variance is desirable. The convergence behavior of varying σ is similar as of varying sizes of D data–set (fig. 16) which also control the amount of variance and the middle optimum size gave faster convergence. But, it is noteworthy to understand that controlling the variance through R_{min} or σ is more systematic which create a more restrictive search space around initial OD estimate generating better OD solution qualities.

Comparing the results of varying R_{min} and σ experiments, first it is interesting to see that lower values of both parameters can converge much more fluctuating demand scenarios (i.e., with $Rand = 0.5$ and σ =0.333). Then, also



(a) Robustness against R_{min} (b) Robustness against σ_{od}, σ_t

Figure 18: Demand scenarios sensitivity (scenario: $Red = 0.7, Rand = 0.3$).



(a) scenario: $Red = 0.7, Rand = 0.15$ (b) scenario: $Red = 0.7, Rand = 0.3$

Figure 19: OD RMSN with different R_{min}.

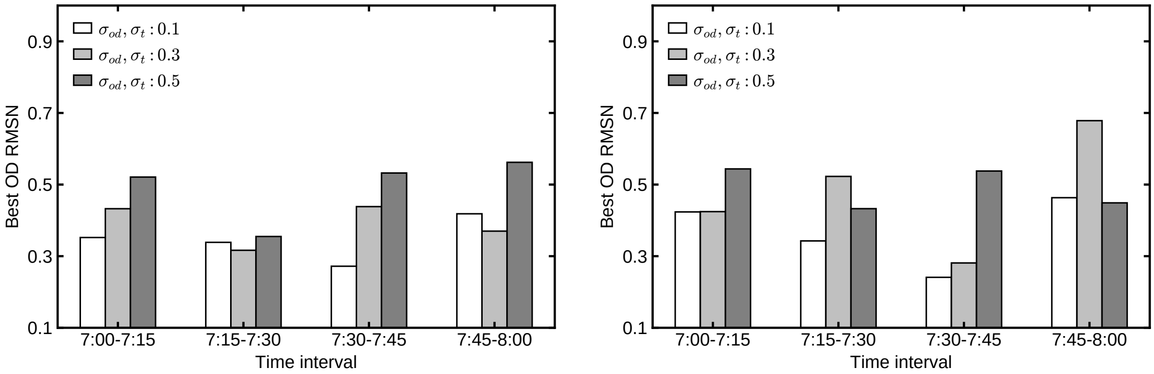
note that, in comparison to the lower σ value of 0.1 (with $R_{min} = 0.3$), the setting of $R_{min} = 0.1$ and $\sigma = 0.3$ gives much faster convergence. Hence, we can conclude that restricting the generated variance by directly reducing the random vector distribution is less efficient than keeping the random vector generation more distributed using higher σ and than tuning down the amount of variance by use of smaller R_{min} values.

4.4. Remarks

The combination of PCA's dimension and complexity reduction with simplified problem formulation gives significant boost to SPSA calibration performance. Also, the proposed framework for data-assimilation generation of historical estimates gives the flexibility to control the size and quality of generation historical variance i.e., the algorithm search space or directions for PCA–methods. Overall, the set of inputs required to use PC–SPSA in our proposed framework include: SPSA hyper–parameters (c, a, α, γ, A), historical data–set generation parameters (generation method, R_{min}, n_d, σ) and PCA application parameters (amount of dimension reduction i.e., V to $\hat{V}$, temporal limits for combined PCA application). In sections 3.2 and 4, we performed a set of experiments on different parameter inputs for PC–SPSA setup. Analyzing the empirical outputs of these experiments, we enlist the guidelines in below sections which can be followed for efficient calibration setup.

4.4.1. SPSA hyper–parameters

Although Spall (1998) gave guidelines for defining appropriate SPSA parameters, their definition remain problem specific with no universal values for different DTA models. For PC–SPSA, the perturbation c_k and minimization a_k



(a) scenario: $Red = 0.7, Rand = 0.15$

(b) scenario: $Red = 0.7, Rand = 0.3$

Figure 20: OD RMSN with different σ_{od} and σ_t.

coefficient behave as percentage change instead of absolute, hence they can be set similar for varying DTA models even having different magnitudes of the decision variable. The results from section 4.1 depict that PC-SPSA is even robust for significantly varying values of c and a but they still effect the convergence speeds. Hence, for efficient performance of the algorithm, c can be set in range of 0.1-0.2 resulting in c_k with 10-20% change at first iteration. Similarly, for setting a parameter, a range between 0.8-1.2 is optimum for the current network and using the RMSN as estimator but it depend on the resulting gradient values. Due to fast convergence, the other SPSA parameters α, γ and, A are insignificant because they only control the evolution of gain sequence parameters (c_k, a_k) over the increasing number of iterations.

4.4.2. Historical data–set generation

The proposed data-assimilation framework generate historical OD data–sets using all different correlations present in time-dependent ODs. The set of inputs given in these generation methods include number of correlation dimensions or generation method, size n_d of the historical estimates, R_{min} to control size of generated variance and σ to define the correlation distributions used to generate Δ_T perturbation matrices. In 4.3, sensitivity analysis on each of these stated parameters are performed to understand their effect on calibration convergence and OD solution quality. Below are the stated guidelines to be followed for each parameter:

- **Generation method:** Given the results in figure 9, method 6 which generates D data–set with all correlation dimensions outperform because of its consistency in convergence speeds and OD solution quality. Hence, it is recommended to use method 6 for implementing PCA–methods with the proposed data-assimilation framework.

- **Size of historical data–set:** In section 4.3.1, analysis upon different sizes of historical data–set is performed. For faster convergence of DODE, the optimum size of generated D data–set should be around 3-4 months (90-120 prior days). Further, to improve the quality and consistency of OD solution quality D data–set can be further extended to higher size but at an expense of reducing convergence speed.

- **Variance of historical data–set:** Section 4.3.2 gives the analysis on defining different variance characteristics within generated D data–set. Two parameters (i.e., R_{min} and σ) are set to control the variance. Individually, smaller values of both parameters (around 0.1) result in optimum OD solution qualities as they restrict the generated variance closer to the seed OD matrix. In terms of convergence, higher values of R_{min} always result in faster convergence but at an expense of more nosier/poor OD estimate, while very low or high σ values show slower convergence, hence optimum value of $\sigma = 0.3$ can result in faster convergence.
  For combined set of values for both R_{min} and σ, it is recommended to use larger σ value in range $0.3 - 0.5$ with smaller value of R_{min} in range $0.1 - 0.15$. This helps to generate a more distributed variance with higher σ but with a much smaller size contained by lower values of R_{min}. If convergence error results are not satisfactory, gradually increasing the R_{min} value is recommended due to probabilities of larger fluctuations in target demand. Note that higher value of R_{min} in such case with always reduce the OD solution quality.

4.4.3. PCA application

The application of PCA on DODE has been covered previously in literature. Djukic et al. (2012) showed the detailed concept of PCA application on OD estimation. Later many other approaches followed the use of PCA to develop variants of conventional approaches (Prakash et al., 2017, 2018; Qurashi et al., 2019). Once the historical OD estimates are available, two main inputs are required for PCA application: 1) the amount of dimension reduction or the number of PCs retained 2) Temporal settings of historical estimates to apply PCA.

The first input of PCs retained during dimension reduction (i.e, changes V to $\hat{V}$ in 2) is commonly given in terms of the level of variance explained by the retained PCs. Since mostly the first few PCs are the most significant, explaining the majority of variance, a cumulative variance of 95% is set for reducing the PCs matrix V. The second input about temporal settings of historical data–set D is defined inside matrix x of eq. (6) in our proposed framework. This input is the number of n_t time intervals set together for application of PCA. It is recommended to apply PCA for the time intervals which have a single activity pattern (e.g., morning or evening peak hours separately). It is also a work in progress for future research to do more systematic PCs extraction from discrete activity patterns and then use the combination of these PC-directions to do more efficient OD estimation.

5. Conclusion

In this paper, practical implementation methods for PCA–based calibration approaches are proposed and evaluated. The results suggest that these methods will facilitate the adoption of PCA-based methods for large–scale applications. PCA–based calibration has become a standard for improving the scalability of conventional algorithms towards large–scale DTA models. However, PCA implementation is based on the availability of historical estimates, which are usually not available in practise. This triggers a chicken and egg problem. To use PCA–based models there is a need for historical estimates, which can be obtained by calibrating the network. However, without PCA–based models it is not possible to calibrate large networks, therefore to have historical estimates. This is a major limitation of current PCA–based methodologies, which is addressed in this paper. In addition, while current approaches mostly focused on using PCA to reduce the number of variables in the problem, a significant gap still exists to exploit the properties of PCA based model calibration for simplifying the structure of the calibration process. Even when historical estimates are available, it is not clear to which extent the quality of these estimates influences prediction accuracy. This paper answers this question, bringing PCA-based algorithms one step closer to real–life applications.

The major contribution of this research is to propose a data–assimilation framework which allows to incorporate the structure of the historical (seed) demand into the Principal Components (PCs) without the need for historical estimates. Such a framework allows the use of all PC-based algorithms proposed in the literature when historical data is irrelevant or unavailable (a standard case for large–scale networks). Based on this data–set generation model, a simplified problem formulation for Dynamic Origin Destination matrix Estimation (DODE) is also presented, which allows removing the demand from the objective function. These extensions have been tested using PC-SPSA, an algorithm that combines PCA with the well known Simultaneous Perturbation Stochastic Approximation (SPSA) model. The paper shows that a better exploitation of the PCA properties leads to an enhanced algorithm that achieves faster convergence and provide more robust results even on large urban networks. Different historical OD generation models have been proposed and tested in this paper, each of which accounts for different types of correlations between the variables. These correlations model spatial, temporal, and day to day changes in the demand. The results suggest that the method that uses all three correlations outperforms others for convergence speed, robustness of the results, and calibrated OD solution quality. Approaches that use only one of these correlations also provide very good results in terms of reproducing the traffic measurements. However, the results show that in this case PCA–models are more likely to over-fit the data, as the PCs cannot model correlations between the ODs properly. Although the proposed framework currently uses Gaussian distributions for presenting the correlations, it also provides the flexibility to use data–driven spatial–temporal correlations extracted from other data sources, representing more realistic structure of PCs which can better reflect the historical OD flows' dynamics.

In this paper, we tested the model on the network of Munich, one of the largest DTA models ever used as a calibration case study. Even on such a scale (more than 8000 links and 20.000 variables to be calibrated), the results indicate that a very low number of iterations is required for convergence. Setting the model scale aside, the required number of simulations are still far less (around 10 simulation runs) when compared to conventional techniques like SPSA (almost 150-300 simulation runs) on much smaller networks. This is a crucial aspect, as a single simulation run for the Munich regional network can take several hours on high performance computing platforms and that, due to the iterative nature of the DTA calibration problem, opportunities for parallel computing are limited. Further, the PC–SPSA implementation used in this research shows robustness towards the definition of SPSA hyper–parameters depicted through the empirical results of a conducted sensitivity analysis. The proposed approach allows to introduce domain specific knowledge within the PCA algorithm by using probability distributions to describe spatial and temporal correlations. These distribution are characterized by a mean and a variance, which becomes additional hyper–parameters to be calibrated. While this allows for more control over the OD solution quality, it also increase the number of hyper–parameters to be tuned. Tuning SPSA parameters is a trial and error procedure that can require significant amount of time and additional simulations. These findings are summarized in Section section 4.4, which introduces implementation guidelines for PC-SPSA. These guidelines can also be used to combine enhanced SPSA algorithms, such as the W–SPSA, and PCA.

The research presented in this paper introduces the first building block to move PCA–based calibration models proposed in the literature from theory to practise. Existing works in fact rely on historical estimates of the demand, which are not necessarily always available. Based on the proposed concept of data-assimilation, many promising research directions are now opening up. In this paper, the data–assimilation framework is used to incorporate historical information within the PCs of the problem. In the future, we plan to use the same concept to incorporate synthetic populations, activity based models, and, in general, more information about the travel demand without increasing the complexity of the problem. A second interesting research direction is to incorporate different data sources, such as mobile phone network data, GPS trajectory data, and even social media data into the data–assimilation framework. Similarly to the historical demand, this procedure can allow to incorporate these data within the PCs, remove them from the objective function, and therefore significantly improve model performances. Another advantage of the proposed framework is that, beside reducing the number of variables, the proposed model drastically reduces the number of simulation runs required to calibrate the model. This is an important observation when the objective is to calibrate multimodal transport systems, where the number of variables to be calibrated as well as the simulation time are prohibitive already for small sized systems. Finally, traditional PCA–based are linear in their nature. However, there is not guarantee that data are linearly correlated, specifically when using different data sources or complex representations of travel behaviour, such as synthetic populations. Therefore, non linear PCA–based frameworks should also be investigated in the future.

Acknowledgements

This research was supported by the German Research Foundation - Deutsche Forschungsgemeinschaft (DFG) [Project number 392047120, Research grants in collaboration with China] and by the European Union Horizon 2020 research and innovation programme under the Marie-Curie grant agreement N. 754462 and MOMENTUM N.815069.

Appendix A PC–SPSA algorithm

Initialization at iteration 0

Estimate PCs: $\boldsymbol{D} = U\Sigma V^T$

Definition SPSA hyper–parameters: c,a A, γ, α

OD transformation to PC–scores: $z_0 = \hat{V}^T x_0$

Gain sequence update at iteration k

$c_k = c/k^\gamma$

$a_k = a/(k+A)^\alpha$

Perturbation

$$\boldsymbol{z}_{\boldsymbol{k}}^{\pm} = \boldsymbol{z}_{\boldsymbol{k}} \pm \boldsymbol{z}_{\boldsymbol{k}} \times c_k \Delta$$

OD approximation

$$\boldsymbol{x}_{\boldsymbol{k}}^{\pm} \approx \hat{V} \boldsymbol{z}_{\boldsymbol{k}}^{\pm}$$

Gradient evaluation

$$\boldsymbol{g}_{\boldsymbol{k}}'(\boldsymbol{x}_{\boldsymbol{k}}) = \frac{f(\boldsymbol{x}_{\boldsymbol{k}}^{+}) - f(\boldsymbol{x}_{\boldsymbol{k}}^{-})}{2c_k} \left[\Delta_1 \; \Delta_2 \; .. \; \Delta_p\right]^T$$

Minimization

$$\boldsymbol{z}_{\boldsymbol{k+1}} = \boldsymbol{z}_{\boldsymbol{k}} - a_k \boldsymbol{g'}_{\boldsymbol{k}}(\boldsymbol{x}_{\boldsymbol{k}})$$

OD approximation at convergence iteration $\mathcal{K}$

$$\boldsymbol{x}_{\boldsymbol{\mathcal{K}}} \approx \hat{V} \boldsymbol{z}_{\boldsymbol{\mathcal{K}}}$$

References

Antoniou, C., Azevedo, C.L., Lu, L., Pereira, F., Ben-Akiva, M., 2015. W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. Transportation Research Part C: Emerging Technologies 59, 129–146.

Antoniou, C., Balakrishna, R., Koutsopoulos, H.N., Ben-Akiva, M., 2009. Off–line and on–line calibration of dynamic traffic assignment systems. IFAC Proceedings Volumes 42, 104–111.

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., Ciuffo, B., Djukic, T., Hoogendoorn, S., Marzano, V., Montero, L., Nigro, M., Perarnau, J., Punzo, V., Toledo, T., van Lint, H., 2016. Towards a generic benchmarking platform for origin-destination flows estimation/updating algorithms: Design, demonstration and validation. Transportation Research Part C: Emerging Technologies 66, 79–98.

Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2007. Nonlinear kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Transactions on Intelligent Transportation Systems 8, 661–670.

Ashok, K., 1996. Estimation and prediction of time-dependent origin-destination flows. Ph.D. thesis. Massachusetts Institute of Technology.

Ashok, K., Ben-Akiva, M.E., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. Transportation science 34, 21–36.

Balakrishna, R., 2006. Off-line calibration of dynamic traffic assignment models. Ph.D. thesis. Massachusetts Institute of Technology.

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., Wen, Y., 2007a. Calibration of microscopic traffic simulation models: Methods and application. Transportation Research Record 1999, 198–207.

Balakrishna, R., Ben-Akiva, M., Koutsopoulos, H.N., 2007b. Offline calibration of dynamic traffic assignment: simultaneous demand-and-supply estimation. Transportation Research Record 2003, 50–58.

Balakrishna, R., Koutsopoulos, H.N., Ben-Akiva, M., 2005. Calibration and validation of dynamic traffic assignment systems, in: Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic TheoryUniversity of Maryland, College Park.

Barceló, J., Montero, L., Marqués, L., Carmona, C., 2010. Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. Transportation research record 2175, 19–27.

Ben-Akiva, M., Bierlaire, M., Burton, D., Koutsopoulos, H.N., Mishalani, R., 2001. Network state estimation and prediction for real-time traffic management. Networks and spatial economics 1, 293–318.

Cantelmo, G., Cipriani, E., Gemma, A., Nigro, M., 2014a. An adaptive bi-level gradient procedure for the estimation of dynamic traffic demand. IEEE Transactions on Intelligent Transportation Systems 15, 1348–1361.

Cantelmo, G., Qurashi, M., Prakash, A.A., Antoniou, C., Viti, F., 2020. Incorporating trip chaining within online demand estimation. Transportation Research Part B: Methodological 132, 171–187. doi:https://doi.org/10.1016/j.trb.2019.05.010. 23rd International Symposium on Transportation and Traffic Theory (ISTTT 23).

Cantelmo, G., Viti, F., Cipriani, E., Nigro, M., 2018. A utility-based dynamic demand estimation model that explicitly accounts for activity scheduling and duration. Transportation Research Part A: Policy and Practice 114, 303–320.

Cantelmo, G., Viti, F., Tampère, C.M., Cipriani, E., Nigro, M., 2014b. Two-step approach for correction of seed matrix in dynamic demand estimation. Transportation Research Record 2466, 125–133.

Cascetta, E., Papola, A., Marzano, V., Simonelli, F., Vitiello, I., 2013. Quasi-dynamic estimation of o–d flows from traffic counts: Formulation, statistical validation and performance analysis on real data. Transportation Research Part B: Methodological 55, 171–187.

Cascetta, E., Postorino, M.N., 2001. Fixed point approaches to the estimation of o/d matrices using traffic counts on congested networks. Transportation science 35, 134–147.

Cipriani, E., Florian, M., Mahut, M., Nigro, M., 2011. A gradient approximation approach for adjusting temporal origin–destination matrices. Transportation Research Part C: Emerging Technologies 19, 270–282.

Djukic, T., Van Lint, J., Hoogendoorn, S., 2012. Application of principal component analysis to predict dynamic origin–destination matrices. Transportation research record 2283, 81–89.

Frederix, R., Viti, F., Corthout, R., Tampère, C.M., 2011. New gradient approximation method for dynamic origin–destination matrix estimation on congested networks. Transportation Research Record 2263, 19–25.

Krishnakumari, P., van Lint, H., Djukic, T., Cats, O., 2019. A data driven method for od matrix estimation. Transportation Research Part C: Emerging Technologies .

Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E., 2018. Microscopic traffic simulation using sumo, in: The 21st IEEE International Conference on Intelligent Transportation Systems, IEEE. URL: https://elib.dlr.de/124092/.

Lu, L., Xu, Y., Antoniou, C., Ben-Akiva, M., 2015. An enhanced spsa algorithm for the calibration of dynamic traffic assignment models. Transportation Research Part C: Emerging Technologies 51, 149–166.

Mahmassani, H.S., 2001. Dynamic network traffic assignment and simulation methodology for advanced system management applications. Networks and spatial economics 1, 267–292.

Marzano, V., Papola, A., Simonelli, F., 2009. Limits and perspectives of effective o–d matrix correction using traffic counts. Transportation Research Part C: Emerging Technologies 17, 120–132.

McNally, M.G., 2007. The four-step model. Emerald Group Publishing Limited.

Osorio, C., 2019. High-dimensional offline origin-destination (od) demand calibration for stochastic traffic simulators of large-scale road networks. Transportation Research Part B: Methodological 124, 18–43.

Prakash, A.A., Seshadri, R., Antoniou, C., Pereira, F.C., Ben-Akiva, M., 2018. Improving scalability of generic online calibration for real-time dynamic traffic assignment systems. Transportation Research Record 2672, 79–92.

Prakash, A.A., Seshadri, R., Antoniou, C., Pereira, F.C., Ben-Akiva, M.E., 2017. Reducing the dimension of online calibration in dynamic traffic assignment systems. Transportation Research Record 2667, 96–107.

Qurashi, M., Ma, T., Chaniotakis, E., Antoniou, C., 2019. PC–SPSA: Employing dimensionality reduction to limit SPSA search noise in DTA model calibration. IEEE Transactions on Intelligent Transportation Systems 21, 1635–1645.

Ros-Roca, X., Montero, L., Barceló, J., 2021. Investigating the quality of spiess-like and spsa approaches for dynamic od matrix estimation. Transportmetrica A: Transport Science 17, 235–257.

Spall, J.C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. IEEE Transactions on aerospace and electronic systems 34, 817–823.

Spall, J.C., et al., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE transactions on automatic control 37, 332–341.

Tampere, C.M., Viti, F., Immers, L.H.B., 2010. New Developments in Transport Planning. Edward Elgar Publishing.

Toledo, T., Kolechkina, T., 2012. Estimation of dynamic origin–destination matrices using linear assignment matrix approximations. IEEE Transactions on Intelligent Transportation Systems 14, 618–626.

Tympakianaki, A., Koutsopoulos, H.N., Jenelius, E., 2015. c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin–destination matrix estimation. Transportation Research Part C: Emerging Technologies 55, 231–245.

Tympakianaki, A., Koutsopoulos, H.N., Jenelius, E., 2018. Robust spsa algorithms for dynamic od matrix estimation. Procedia computer science 130, 57–64.

Vaze, V., Antoniou, C., Wen, Y., Ben-Akiva, M., 2009. Calibration of dynamic traffic assignment models with point-to-point traffic surveillance. Transportation Research Record 2090, 1–9.

Wu, X., Guo, J., Xian, K., Zhou, X., 2018. Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. Transportation Research Part C: Emerging Technologies 96, 321–346.

Zhang, C., Osorio, C., Flötteröd, G., 2017. Efficient calibration techniques for large-scale traffic simulators. Transportation Research Part B: Methodological 97, 214–239.

Zhou, X., Mahmassani, H.S., 2007. A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. Transportation Research Part B: Methodological 41, 823–840.