UNIVERSITY OF NAPLES “FEDERICO II”

Doctoral School of Hydraulic, Transportation and Urban Planning Engineering

Doctoral Thesis

Estimation of origin-destination matrices from traffic counts: theoretical and operational development

Iolanda Vitiello

Dean of the Doctoral School:
Full Professor Elvira Petroncelli

Supervisor:
Associate Professor Andrea Papola

Assistant Supervisors:
Assistant Professor Fulvio Simonelli
Assistant Professor Vittorio Marzano

Abstract

This thesis deals with the o-d estimation problem from indirect measures, addressing two main aspects of the problem: the identification of the set of indirect measures that provide the maximum information with a resulting reduction of the uncertainty on the estimate; once defined the set of measures, the choice of an estimator to identify univocally and as much reliable as possible the estimate.

As regards the former aspect, an innovative and theoretically founded methodology is illustrated, explicitly accounting for the reliability of the o-d matrix estimate. The proposed approach is based on a specific measure, named Synthetic Dispersion Measure (SDM), related to the trace of the dispersion matrix of the posterior demand estimate conditioned to a given set of sensors locations. Under the mild assumption of multivariate normal distribution for the prior demand estimate, the proposed SDM does not depend on the specific values of the counted flows – unknown in the planning stage – but just on the locations of such sensors. The proposed approach is applied to real contexts, leading to results outperforming the other methods currently available in the literature. In addition, the proposed methodology allows setting a formal budget allocation problem between surveys and counts in the planning stage, in order to maximize the overall quality of the demand estimation process.

As regard the latter aspect, a “quasi-dynamic” framework is proposed, under the assumption that o-d shares are constant across a reference period, whilst total flows leaving each origin vary for each sub-period within the reference period. The advantage of this approach over conventional within-day dynamic estimators is that of reducing drastically the number of unknowns given the same set of observed time-varying traffic counts. The quasi-dynamic assumption is checked by means of empirical and statistical tests and the performances of the quasi-dynamic estimator - whose formulation is also given – are compared with other dynamic estimators.

Contents

List of Figures

List of Tables

Introduction and motivation

Transport demand estimation is a crucial issue in both planning and management of transport systems because the knowledge of the demand is a necessary condition for a design of effective policy and operational measures and for a proper evaluation of their effects. In practice, transport demand can be estimated through direct methods (e.g., surveys or other inference-based statistical methods), indirect methods (e.g., mathematical models linking transport demand to a set of explanatory variables related to the socioeconomic characteristics of the study area and to the performance of the transport system under analysis), or both. In the case of indirect methods, such models would require the completion of specific surveys for their calibration and validation. Costs for performing surveys, either for direct demand estimation or for calibration and validation of demand models, are normally high and often incompatible with the budget. Therefore, models previously calibrated in different contexts are usually applied in practice, leading to quite unreliable demand estimates and the task of improving the quality of such estimates is left to the "o-d flow estimation from traffic counts" procedure. For this reason, the transport demand estimation problem based on the use of indirect measures has received considerable interest in the literature and its fundamental aspects, along with direct and indirect estimation methods, are summarized in Chapter 1.

Unfortunately, the demand flow estimation problem based on link counts is, in almost all cases, an underdetermined problem, since in real situations the number of measurements – that is the number of available equations – is far less than the number of variables to be estimated. Therefore, information provided by the measurements should be improved by means of information derived from other sources: generally, one can refer to a prior estimate of the demand flows and can correct this estimate by forcing it to reproduce the measured flows. In this context, two problems arise: 1) the Network Sensor Location Problem (NSLP), that is the identification of the optimal locations of traffic counts allowing to get an estimate as much reliable as possible; 2) the choice of an estimator that, starting from the observed measures, allows to identify univocally and, once again, as much reliable as possible the estimate.

As regards the first aspect, in accordance with the existing literature on this problem, the proper definition of an NSLP should be based on a rigorous measure of the quality – or equivalently of the variability – of the posterior demand vector, i.e. the result of the updating of the prior o–d matrix using the flows counted in the optimal sensor locations. Formally, this measure should be related to the statistical distribution of the posterior demand vector conditional upon the assignment equations related to the counted links. In a Bayesian perspective, this means accounting also for the statistical distribution of the prior o–d matrix, in turn related to the estimation methodology (e.g. direct surveys and/or mathematical models) and/or to the subjective judgement of the analyst (e.g. different data sources adopted for prior estimation may impact differently on each o–d pair). Furthermore, since the actual link flows in the count sections are unknown in the planning stage, a proper NSLP should treat them as random variables. Finally, from a practical standpoint, the mathematics underlying the NSLP and the corresponding solution algorithms should allow for feasible and effective applications to real networks. Starting from those premises, Chapter 2 proposes an innovative and theoretically founded NSLP formulation, explicitly accounting for the variability of the posterior o–d matrix estimate. The problem is formulated first in the general case and then particularized to the noteworthy case of multivariate normal distribution for the prior demand. Indeed, this assumption allows for a substantial simplification of the proposed approach, leading also to the implementation of a sequential NSLP solving algorithm applicable to real networks. Importantly, the proposed framework can be effectively applied in a context of budget allocation problem for demand estimation: that is, since the variability of the posterior demand estimate depends both on the variability of the prior demand estimate and on the information related to the optimal link count locations, a budget allocation problem can be formulated in terms of trade-off between surveys and traffic counts, in order to minimize the overall demand variability.

As regards the second aspect, that is the choice of the estimator, the problem of the balance between equations and unknowns is usually solved by introducing additional information related to historical (or prior) estimates. Unlike the stationary contexts, where this approach appears to be the only adoptable, in dynamic contexts it is possible to formulate several hypotheses about the evolutionary mechanisms of the demand between the temporal intervals of interest; in this way, a reduction of the number of unknowns in the problem and, in some cases, a balance between equations and unknowns can be achieved. Following this approach, a new hypothesis on the evolutionary mechanisms of the demand, called "quasi- dynamic" is proposed and tested on real data in Chapter 3.

1 TRANSPORT DEMAND ESTIMATION

1.1 Introduction

Information on the origin-destination (o-d) matrix [1] of a transport network is a fundamental requirement in transport systems analysis, planning and managing. Many researchers investigated methods for reconstructing, estimating or predicting the o-d traffic flows.

Hazelton (2001) defined the basic concepts of this three problems:

- Reconstruction. The aim is to estimate the actual number of trips between each o-d pair that occurred during the observational period.
- Estimation. The aim is to estimate the expected number of o-d trips.
- Prediction. The aim is to estimate future o-d traffic flows.

O-d estimation procedures can be based on surveys of mobility (direct estimation) or on mathematical models (indirect estimation) where o-d flows are correlated with a series of known explanatory variables related to the study area. Direct estimation provides results whose reliability is closely related to the sampling rate and therefore to the cost of the surveys. Indirect estimation is cheaper than the direct one, but its reliability is generally limited. In particular, the most difficult aspect to reproduce is the spatial structure of the trips that depends on variables difficult to observe and on aspects which are sometimes random and difficult to include in a model. Nevertheless, while direct estimation methods allow to estimate only the present demand, properly calibrated models can be used to estimate the present demand and to forecast the (hypothetical) future demand.

The result of these considerations is that researchers and practitioners generally use models calibrated and tested in different territorial realities to obtain a prior o-d matrix for the working area, and much cheaper indirect measures, i.e. flow counts on some

[1] Transport demand of a study area can be spatially represented by an origin-destination (o-d) matrix. The generic element *ij* of the matrix represents the flow of users travelling from origin *i* to destination *j*, during a reference period.

elements of the considered network, to improve/update this available prior o-d matrix, that is to get an o-d matrix as closer as possible to the true one.

Estimation/updating of o-d matrix based on traffic counts is a widely discussed problem in literature and its fundamental aspects, along with direct and model estimation methods, are summarized in this chapter.

1.2 Direct sample estimation of the transport demand

“Direct” methods are based on the use of surveys to a sample of system users and on statistical techniques which, starting from these surveys, allow getting an estimate extended to overall analysis system.

Sampling surveys for direct estimation of travel demand are usually known as “origin-destination surveys”. These surveys can be conducted with several techniques depending on characteristics of the information that one wants to get from them. “On board surveys” are usually conducted to estimate internal-external and external-internal demand flows with reference to a study area; in this case they are also called “cordon surveys”, because of the survey sections locations. Differently, “household surveys” are interviews with a sample of families or persons living within the study area, conducted at interviewees domicile or by mail, internet, telephone. Interviews at travellers homes are usually more expensive; however, more reliable and precise information is generally obtained because of the direct interaction between the interviewer, conveniently trained, and the interviewee.

As described in Cascetta (2009), the statistical design of a sampling survey for travel demand estimation consists of several standard phases:

- definition of the sampling unit (person, family, vehicle, etc.) and of the method for enumerating the universe (e.g., list of residents or list of telephone subscribers);
- definition of the sampling strategy, that is the method for extracting the sample of individuals to be interviewed;

- definition of the estimator to be adopted, that is the function used to estimate the unknown quantities from the information obtained by the survey;
- definition of the number of units in the sample (sample size).

The definition of the sampling unit is largely influenced by the type of survey and the availability of information about the universe. In applications, the most commonly used probabilistic sampling strategies are the *simple random sampling* (discussed in Section 1.2.1), the *stratified random sampling* (discussed in Section 1.2.2) and the *cluster sampling* (discussed in Section 1.2.3). The choice of the estimator to be adopted and the sample size depend on the sampling strategy considered.

1.2.1 Simple random sampling

In the case of *simple random sampling*, all the elements of the population have an equal probability of belonging to the sample.

Let n and N be the sample size and universe size, respectively. The sampling rate is denoted by $\alpha = n/N$. Let d_{od} denote the demand flow between the origin o and the destination d to be estimated and let n_{od}^{i} be the number of these trips undertaken by the i-th element of the sample. The total of trips obtained from the sample is given by $n_{od} = \sum_{i=1,\dots,n} n_{od}^{i}$ and the sample estimate $\hat{d}_{od}$ of the demand flow for the overall universe can be obtained as follows:

$$\hat{d}_{od} = \frac{n_{od}}{\alpha} = N\bar{n}_{od} \tag{1.1}$$

where $\bar{n}_{od} = n_{od}/n$ is the average number of trips with the desired characteristics undertaken by an element of the sample.

The variance of $\hat{d}_{od}$ can be estimated as:

$$Var[\hat{d}_{od}] = \frac{N^2}{n}\hat{s}^2(1-\alpha) \tag{1.2}$$

where $\hat{s}^2$ is the sample estimate of the variance of the random variable n_{od}^{i}:

$$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1,\dots,n}(n_{od}^i - \bar{n}_{od})^2 \qquad (1.3)$$

$(1-\alpha)$ is a correction coefficient which accounts for the fact that the population, over which the random variable is defined, has a finite number of units; therefore, if a census were conducted, that is if $\alpha = 1$, one could obtain an estimate corresponding to the true value with zero variance.

In some surveys, a sample element undertakes at most one trip with the required characteristics. In other surveys, the required information is whether the sample element has a given characteristic. In such cases, n_{od}^i is either zero or one, and $\bar{n}_{od}$ is the sampling estimate of $\hat{P}_{od}$, that is the percentage of travellers who have undertaken a trip with the given characteristics. In this case, the sampling estimate of the variance of n_{od}^i can be expressed as the variance of a Bernoulli random variable:

$$\hat{s}^2 \cong \hat{P}_{od}(1-\hat{P}_{od}) \qquad (1.4)$$

From the estimate of the variance, one can evaluate the confidence limits of the estimate $\hat{d}_{od}$, that is the values of the upper bound (LS_γ) and the lower bound (LI_γ) of the interval which, depending on a predetermined value γ, includes the true value of the estimated variable with a probability equal to (1- γ). If the sample is enough numerous to be able to apply the central limit theorem[2], one can assume that the estimator $\hat{d}_{od}$ has a normal distribution and the confidence limits can be obtained as follows:

$$LS_\gamma = \hat{d}_{od} + z_{1-\gamma/2} Var\lfloor \hat{d}_{od} \rfloor^{1/2} \qquad (1.5)$$

$$LI_\gamma = \hat{d}_{od} + z_{\gamma/2} Var\lfloor \hat{d}_{od} \rfloor^{1/2} \qquad (1.6)$$

[2] The *central limit theorem* states that the sum of *n* independent and identically distributed random variables with mean μ and variance σ^2, will tend to a normal distribution with mean $n\mu$ and variance $n\sigma^2$, as *n* increases.

$z_{1-\gamma/2}$ and $z_{\gamma/2}$ are respectively the $(1-\frac{\gamma}{2})$ and $\frac{\gamma}{2}$ percentiles of the standard normal variable.

The ratio IR(1- γ), between the confidence interval size and the value to be estimated, is defined relative confidence interval at (1- γ) percent of the estimate $\hat{d}_{od}$:

$$IR(1-\gamma) = \frac{LS_\gamma - LI_\gamma}{d_{od}} \tag{1.7}$$

From a prefixed confidence interval $IR(1-\gamma)$, if the coefficient of variation $cv = s/\bar{n}_{od}$ of the variable n_{od}^{i} is known, one can obtain beforehand the sample size *n* as follows:

$$n \approx 4\frac{cv^2 z_{1-\gamma}^2 (1-\alpha)}{IR(1-\gamma)^2} \tag{1.8}$$

This is rarely possible because it would be necessary to know the parameters values that usually are obtained during the survey and because the size *n* necessary to obtain estimates enough accurate would be too big. Therefore, in order to choose the sample size, one can usually refer to surveys with similar characteristics that produced good results.

1.2.2 Stratified random sampling

Stratified random sampling considers a division of the population into non-overlapping and exhaustive groups (strata). A simple random sampling is then conducted in each stratum, that is a sample of elements is drawn from each stratum and elements of a same stratum have an equal probability of belonging to the sample, while elements of different strata can have different probability. In cordon surveys, the users passing through the several survey sections represent the strata, while, in household surveys, the strata are represented by the families living within each zone.

Let N and N_k be the universe size and the size of the stratum k, respectively. The ratio $w_k=N_k/N$ is so the weight of the stratum k with respect to the universe. Let n_k denote the size of the sample drawn at random from the stratum k and let n_{od}^{ik} be the number of trips between the origin *o* and the destination *d* undertaken by the *i*-th element in the sample of stratum *k*. The sampling estimate $\hat{d}_{od}$ of the total demand flow between *o* and *d* can be obtained as follows:

$$\hat{d}_{od} = N \sum_k w_k \sum_{i=1,..,n_k} n_{od}^{ik} / n_k = N \sum_k w_k \bar{n}_{od}^{k} \qquad (1.9)$$

where $\bar{n}_{od}^{k}$ is the average number of trips observed in the *k*-th stratum.
The variance of the stratified sampling estimate can be estimated as follows:

$$Var[\hat{d}_{od}] \approx N^2 \sum_k w_k^2 \, \hat{s}_k^2 (1 - \alpha_k)/n_k \qquad (1.10)$$

where α_k is the sampling rate in the *k*-th stratum and $\hat{s}_k^2$ is the sample estimate of the variance of n_{od}^{ik}:

$$\hat{s}_k^2 = \frac{1}{n-1} \sum_{i=1,\dots,n_k} (n_{od}^{ik} - \bar{n}_{od}^{k})^2 \qquad (1.11)$$

As in the case of simple random sampling, it is possible to calculate the confidence limits of the estimate.

1.2.3 Cluster sampling

In the case of *cluster sampling*, sampling units are grouped in clusters, which are then extracted randomly with a pre-determined probability of belonging to the sample (*simple random cluster sampling*) or subdivided into strata and sampled with different probabilities in the different strata (*stratified random cluster sampling*). A further possibility is the *two-stage cluster sampling*, in which a sample of clusters is first selected, and then a sample of individuals within each cluster is extracted; in this case, the probability that an individual belongs to the sample is the product of the probability of selecting the cluster to which he or she belongs and the probability that the individual is then extracted within the cluster.

With both stratification (Section 1.2.2) and clusters the population is partitioned into subgroups (strata or clusters). Nevertheless, in the former case (stratification) the sample is drawn from all subgroups, while in the latter case (clusters) all the units are sampled from a subset of subgroups.

Let N_c and n_c be the number of clusters in the population and the number of clusters in the sample, respectively. Let N_k and n_k denote the size of the cluster k and the size of the sample drawn at random from the cluster k, respectively. Let n_{od}^{ik} be the number of trips between the origin o and the destination d undertaken by the i-th element in the sample of cluster k. In the case of *two-stage cluster sampling*, the sampling estimate $\hat{d}_{od}$ of the total demand flow between o and d can be obtained as follows:

$$\hat{d}_{od} = \frac{N_c}{n_c \sum_{k=1,..,N_c} N_k} \sum_{k=1,..,n_c} \frac{N_k}{n_k} \sum_{i=1,\ldots,n_k} n_{od}^{ik} \qquad (1.12)$$

Let $\bar{M} = \frac{\sum_{k=1,..,N_c} N_k}{N_c}$ be the average cluster size. The variance of $\hat{d}_{od}$ can be estimated in this case as:

$$Var\left[\hat{d}_{od}\right] = \frac{N_c - n_c}{N_c} \frac{1}{n_c N_c \bar{M}^2} s_b^2 + \frac{1}{n_c N_c \bar{M}^2} \sum_{k=1}^{n_c} {N_k}^2 \left(\frac{N_k - n_k}{N_k}\right) \frac{s_k^2}{n_k} \qquad (1.13)$$

where

$$s_b^2 = \frac{\sum_{k=1}^{n_c} \left(\frac{N_k}{n_k} \sum_{i=1,\ldots,n_k} n_{od}^{ik} - \bar{M}\hat{d}_{od}\right)^2}{n_c - 1}$$

and

$$s_k^2 = \frac{\sum_{k=1}^{n_k} \left(n_{od}^{ik} - \frac{1}{n_k} \sum_{i=1,\ldots,n_k} n_{od}^{ik}\right)}{n_k - 1}$$

1.3 Disaggregate estimation of demand models

Mathematical models, applied to estimate travel demand, have to be properly specified, calibrated and validated. Information, resulting from surveys to a sample of users, are necessary to proceed in such operations. These surveys can be related to behaviour actually observed or demonstrated by the users in a real context (*Revealed Preference* or *RP surveys*), or to behaviour declared by users in hypothetical contexts (*Stated Preference* or *SP surveys*).

Model specification consists in defining its mathematical structure, i.e. the functional form of the discrete choice model (Logit, Probit) used, and in identifying the explanatory variables to be used in it.

After the specification stage, the model is calibrated, that is its parameters are estimated. The most widely used method to calibrate a disaggregate demand model is the *Maximum Likelihood* (ML) *method*, which provides the values of the unknown parameters by maximizing the sample likelihood, that is the probability of observing the choices made by a sample of users, expressed as a function of the unknown parameters themselves.

In the case of *simple random sampling* of *n* users, the likelihood function is expressed as

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1,\dots,n} p^i[j(i)](\mathbf{X}^i, \boldsymbol{\beta}, \boldsymbol{\theta}) \tag{1.14}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are the vectors of the model parameters, $p^i[j(i)]$ is the probability that each user *i* chooses *j(i)*, that is the alternative actually chosen by him or her, and $\mathbf{X}^i$ is the vector of the explanatory variables for the user *i*.

The maximum likelihood estimate $[\boldsymbol{\beta}, \boldsymbol{\theta}]_{\mathrm{ML}}$ of the vectors of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is obtained by maximizing (1.14) or, more conveniently, its natural logarithm (the log-likelihood function):

$$[\boldsymbol{\beta}, \boldsymbol{\theta}]_{\mathrm{ML}} = \arg\max \ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \arg\max \sum_{i=1,\dots,n} \ln p^i[j(i)](\mathbf{X}^{\mathrm{i}}, \boldsymbol{\beta}, \boldsymbol{\theta}) \tag{1.15}$$

In the case of *stratified random sampling*, the probability of observing the sample choices and therefore the likelihood function, depends on the method used to identify the strata. If the population is stratified using, either directly or indirectly, the attributes **X** but not the choices to be modelled, the strategy is known as *exogenous stratified sampling* (e.g. geographical stratification and income stratification). For samples obtained through *exogenous stratified sampling*, considered that for each stratum h (h = 1, …, H) the sampling rate is $\alpha_h = n_h/N_h$, it can be demonstrated that the log-likelihood function is:

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{h=1,\dots,H} \sum_{i=1,\dots,n_h} \ln p^i[j(i)](\mathbf{X}^{\mathrm{i}}, \boldsymbol{\beta}, \boldsymbol{\theta}) + \text{cost} \tag{1.16}$$

which, apart from a constant term, coincides with the function (1.15) obtained for a simple random sample with size $n = \sum_{h=1,\dots,h} n_h$.

If the stratification is based on the choices made by the users, the sampling strategy is known as *choice-based stratified sampling*. The exact closed form log-likelihood function is rather complex for this sampling strategy. As an approximation, the maximum likelihood estimator with exogenous weights can be adopted and the log-likelihood function is expressed as:

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{h=1,\dots,H} \left(\frac{w_h}{\alpha_h}\right) \sum_{i=1,\dots,n_h} \ln p^i[j(i)](\mathbf{X}^{\mathrm{i}}, \boldsymbol{\beta}, \boldsymbol{\theta}) \tag{1.17}$$

which, apart from the weights w_h and α_h, coincides with (1.16) and therefore with (1.15).

To apply the maximum likelihood estimator with exogenous weights to a choice-based stratified sample, it is therefore necessary to have an estimate of the weight of each stratum, that is, of the fraction of the total population choosing each alternative. This information can be obtained from official statistics, or estimated from another simple random sample with smaller or less detailed questionnaires.

Under rather general assumptions, maximum likelihood estimators have many desirable asymptotic statistical properties such as consistency, efficiency, and normality, regardless of the model used to express the probabilities $p^i[j]$.

Furthermore, it is possible to obtain approximate estimates of the variances and covariances of the components of $[\boldsymbol{\beta}, \boldsymbol{\theta}]_{\mathrm{ML}}$; the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\theta}}$ is, in fact, asymptotically equal to the negative inverse of the log-likelihood function's Hessian, evaluated at the point $[\boldsymbol{\beta}, \boldsymbol{\theta}]_{\mathrm{ML}}$:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\theta}} = -\left[\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial(\boldsymbol{\beta},\boldsymbol{\theta})\partial(\boldsymbol{\beta},\boldsymbol{\theta})^T}\right]^{-1}_{[\boldsymbol{\beta},\boldsymbol{\theta}]_{\mathrm{ML}}} \tag{1.18}$$

If the sample is sufficiently large, expression (1.18) can be used to estimate variances and confidence limits for the coefficients. In addition, how the covariance matrix of the coefficients can influence the dispersion of the overall demand estimate could be an interesting aspect to be dealt with.

The specified and calibrated model is then validated. Validation stage consists in testing model's quality by using appropriate informal (that is qualitative) tests and formal (that is statistical) tests. The specification-calibration-validation cycle is usually repeated several times until a good demand model is obtained.

1.4 Estimation/Updating of o-d demand flows using traffic counts

Traffic counts are often collected to monitor traffic circulation. They measure the number of vehicles passing through a point (or section) during a specified time period. They are usually conducted to monitor and describe traffic characteristics such as peak hour volume, average daily traffic, average annual daily traffic, etc. In fact, these link traffic counts can be efficiently used to estimate an o-d matrix. The information on o-d flows contained in traffic counts is represented by the following system of equations:

$$\hat{\mathbf{y}} = \widehat{\mathbf{M}}\mathbf{d} + \boldsymbol{\varepsilon} \tag{1.18}$$

where $\hat{\mathbf{y}}$ is the vector of measured link flows; $\widehat{\mathbf{M}}$ is the assignment matrix, which maps the o-d demand flows to link flows, on the base of the route choice behaviour of the trip makers and of the flow propagation relationships; $\mathbf{d}$ is the "true" vector of demand o-d flows; $\boldsymbol{\varepsilon}$ is the vector which accounts for assignment and measurement errors. Even assuming that $\boldsymbol{\varepsilon}$ is null, the number of independent equations of this system is usually much less than the number of unknown o-d demand flows to be estimated. Indeed, to estimate the unknown o-d flows, information contained in the counts must be combined with other sources, that is, in general, a prior o-d matrix. Therefore, information contained in the flows observed on a certain number of links in the network can be used to update or improve a prior o-d matrix estimate already available for the study network. Demand flows estimation using traffic counts can be considered as the inverse assignment problem. In fact, by solving an assignment problem, one can calculate link flows starting from o-d flows, network topology and characteristics (supply system), and path choice models. Conversely, the considered o-d estimation problem is that of calculating o-d flows starting from measured link flows, supply system and path choice models. This approach is practically attractive, because traffic counts are readily available in many urban areas and relatively inexpensive to collect, thereby overcoming the more time-consuming and expensive traditional travel surveys.

Cascetta and Nguyen (1988) expressed the problem of estimating o-d flows by using traffic counts in general form as:

$$\mathbf{d}^* = \arg\min_{x\geq 0}\left[z_1\left(\mathbf{x},\hat{\mathbf{d}}\right) + z_2(\boldsymbol{\nu}(\mathbf{x}),\hat{\mathbf{y}})\right] \tag{1.20}$$

where the two functions $z_1\left(\mathbf{x},\hat{\mathbf{d}}\right)$ and $z_2(\boldsymbol{\nu}(\mathbf{x}),\hat{\mathbf{y}})$ can be considered as two "distance" measures: z_1 measures the "distance" of the unknown demand vector $\mathbf{x}$ from a prior estimate vector $\hat{\mathbf{d}}$, z_2 measures the "distance" of the flow vector $\boldsymbol{\nu}(\mathbf{x})$ obtained by assigning $\mathbf{x}$ to the network from the traffic counts vector $\hat{\mathbf{y}}$. An intuitive interpretation of the problem (1.20) is that it searches the vector $\mathbf{d}^*$ that is closest to the prior estimate, and, once it is assigned to the network, produces the flows $\boldsymbol{\nu}(\mathbf{d}^*)$ closest to the counts. In general, the functional form of the two terms $z_1(\cdot)$ and $z_2(\cdot)$, depends on the type of information available (experimental or non-experimental) and on the probability laws associated with such information.

The main approaches developed in literature are described in Section 1.4.1. Some contributions existing in literature, with reference to congested networks, are then recalled in Section 1.4.2.

1.4.1 Static o-d flow estimators based on traffic counts

In the static framework, three main static estimators for o-d matrices were proposed:

- Maximun Likelihood (Maher, 1983; Bell, 1983; Cascetta and Nguyen, 1988);
- Generalized Least Squares (Cascetta, 1984);
- Bayesian (Maher, 1983).

Maximum likelihood (ML) estimator is obtained by maximizing the probability of observing both the additional sampling survey results and the counted flows. Under the usually acceptable assumption that these two probabilities are independent, the maximum likelihood estimator $\mathbf{d}^{\mathrm{ML}}$ can be expressed as:

$$\mathbf{d}^{\mathrm{ML}} = \arg\max_{\mathbf{x}\in \mathrm{S}} \left[\ln \mathrm{L}\,(\mathbf{n}/\mathbf{x}) + \ln \mathrm{L}\,(\hat{\mathbf{y}}/\mathbf{x})\right] \tag{1.21}$$

where $\mathbf{x}$ is the “unknown” demand vector; $\mathbf{n}$ is the vector of o-d demand counts, that is the sampling survey results; $\hat{\mathbf{y}}$ is the vector of link counts; $\ln \mathrm{L}\,(\mathbf{n}/\mathbf{x})$ is the log-likelihood function of demand counts, that is, the logarithm of the probability of observing the sampling vector $\mathbf{n}$ if $\mathbf{x}$ is the true demand vector; $\ln \mathrm{L}\,(\hat{\mathbf{y}}/\mathbf{x})$ is the log-likelihood function of the traffic counts, that is, the logarithm of the probability of observing the vector of the counts $\hat{\mathbf{y}}$ if $\mathbf{x}$ is the true demand vector; S is the feasibility set of the true demand vector.

Therefore, the formulation of hypotheses on the probability laws of $\mathbf{n}/\mathbf{x}$ and of $\hat{\mathbf{y}}/\mathbf{x}$ is required to solve the problem (1.21). It is usually assumed that traffic counts are random variables with means given by the flows $\boldsymbol{\nu}(\mathbf{x})$ obtained by assigning the demand $\mathbf{x}$. Furthermore, the probability laws most widely used for count data are the Poisson and the multivariate normal.

The log-likelihood function of o-d demand counts depends on the type of sampling adopted. In the case of stratified random sampling by zone of origin, a multinomial distribution is considered. If the number of trips sampled at each origin is sufficiently large (a few dozen or more), the multinomial variable can be closely approximated by the product of independent Poisson variables.

Generalized Least Squares (GLS) or Aitken estimator provides the estimate of the unknown o-d demand flow vector $\mathbf{d}$, starting from a system of linear stochastic equations, which can be obtained by combining the following equations:

$$\hat{\mathbf{d}} = \mathbf{d} + \boldsymbol{\eta} \tag{1.22}$$

$$\hat{\mathbf{y}} = \widehat{\mathbf{M}}\mathbf{d} + \boldsymbol{\varepsilon} \tag{1.23}$$

Equations (1.22) and (1.23) represent the information contained in the prior estimate $\hat{\mathbf{d}}$ and in the observed link flows $\hat{\mathbf{y}}$, respectively. The vector $\boldsymbol{\eta}$ is to take in account of the presence of a sampling bias, if $\hat{\mathbf{d}}$ is a direct sample estimate, or of a bias due to the model misspecification, if $\hat{\mathbf{d}}$ is obtained by applying a demand model calibrated in a different situation.
Under the hypothesis that the random vectors $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ have zero means, i.e. $E(\boldsymbol{\eta}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and being $\mathbf{V}$ and $\mathbf{W}$ their covariance matrices, i.e. $Var[\boldsymbol{\eta}] = \mathbf{V}$ and $Var[\boldsymbol{\varepsilon}] = \mathbf{W}$, the GLS estimator of the demand vector can be expressed as:

$$\mathbf{d}^{\mathrm{GLS}} = \arg\min_{\mathbf{x}\epsilon\mathrm{S}} \left[\left(\hat{\mathbf{d}} - \mathbf{x}\right)^{\mathrm{T}} \mathbf{V}^{-1} \left(\hat{\mathbf{d}} - \mathbf{x}\right) + \left(\hat{\mathbf{y}} - \widehat{\mathbf{M}}\mathbf{x}\right)^{\mathrm{T}} \mathbf{W}^{-1} \left(\hat{\mathbf{y}} - \widehat{\mathbf{M}}\mathbf{x}\right) \right] \tag{1.24}$$

Note that, under the hypothesis that the random vectors $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ have zero means, the GLS estimator, is the best linear unbiased estimator (BLUE) of the demand vector $\mathbf{x}$, i.e. the estimator of minimum variance in the class of all unbiased estimators linear in the vectors $\hat{\mathbf{d}}$ and $\hat{\mathbf{y}}$. Furthermore, if a multivariate normal distributional is assumed for the random vectors $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$, the GLS estimator coincides with the ML estimator and therefore it is the minimum variance estimator among all unbiased ones.
Expression (1.24) is often applied assuming that the matrices $\mathbf{V}$ and $\mathbf{W}$ are diagonal, that is ignoring the covariances between the components of vectors $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$. This is done because these covariances are difficult to express and also to reduce memory

requirements and computing times. Therefore, the solution of the problem (1.24) is the demand vector of minimum weighted distance from the prior estimate $\hat{\mathbf{d}}$, and that, once assigned to the network, gives rise to value of link flows of minimum weighted distance from the observed link flows $\hat{\mathbf{y}}$. Furthermore, the distance of a component of $\hat{\mathbf{d}}$, or of $\widehat{\mathbf{M}}\mathbf{x}$, from the analogous component of $\mathbf{x}$, or of $\hat{\mathbf{y}}$, has a weight inversely proportional to the variance of its error. This means, for example, that a prior o-d component with a higher variance, being far from the analogous true o-d component, is weighted less.

Under the hypothesis of zero assignment and measurement errors, one can obtain the Constrained Generalized Least Squares (CGLS) estimator as:

$$\mathbf{d}^{\mathrm{GLS}} = \arg\min_{\mathbf{x}\in \mathrm{S}} \left[\left(\hat{\mathbf{d}} - \mathbf{x}\right)^{\mathrm{T}} \mathbf{V}^{-1} \left(\hat{\mathbf{d}} - \mathbf{x}\right)\right] \tag{1.25}$$

$$\widehat{\mathbf{M}}\mathbf{x} = \hat{\mathbf{y}} \tag{1.26}$$

$$\mathbf{x} \geq \mathbf{0} \tag{1.27}$$

In this case, the solution, that is $\mathbf{d}^{\mathrm{GLS}}$, according to the metrics introduced by $\mathbf{V}$**,** is a projection of the prior demand estimate $\hat{\mathbf{d}}$ on the space defined by the constraint (1.26) and (1.27).

Bayesian estimator for o-d demand flows can be found maximizing the posterior probability function of the unknown demand vector $\mathbf{x}$ conditional on a priori information $\hat{\mathbf{d}}$ and on experimental information $\hat{\mathbf{y}}$. According to Bayes' theorem, the posterior probability is proportional to the product of two factors: the prior probability function $\mathrm{g}(\mathbf{x}/\hat{\mathbf{d}})$, which expresses the distribution of subjective probability attributed to the unknown vector given the a priori estimate, and the probability, or likelihood, function $\mathrm{L}(\hat{\mathbf{y}}/\mathbf{x})$, which expresses the probability of observing the traffic counts conditional on the unknown demand vector. Bayesian estimator $\mathbf{d}^{\mathrm{B}}$ is therefore expressed as:

$$\mathbf{d}^{\mathrm{B}} = \arg\max_{\mathbf{x}\in \mathrm{S}} \left[\ln \mathrm{g}(\mathbf{x}/\hat{\mathbf{d}}) + \ln \mathrm{L}(\hat{\mathbf{y}}/\mathbf{x})\right] \tag{1.28}$$

The tractability of the mathematics here depends on the distributional assumptions made for $g(\mathbf{x}/\hat{\mathbf{d}})$ and for the observed flows.

1.4.2 O-d demand flows updating using traffic counts in congested networks

Several researchers dealt with the problem of updating an o-d matrix with reference to a congested network, where the path proportions are proportional to the path costs, which are flow dependent. For this aim, a bi-level programming approach was used, integrating the conventional generalized least squares estimation model and an equilibrium traffic assignment model into one process. In particular, the upper level problem seeks to minimize the sum of error measurements in traffic counts and o-d matrix, whereas the lower level problem represents a network equilibrium assignment that guarantees that the estimated o-d matrix and the corresponding link flows satisfies the user-equilibrium conditions.
Importantly, be-level programming problem proved to be difficult to solve because of their inherent non-convexity and non-differentiability. In addition, the second-level problem has to be solved at each step of the optimizing search process of o-d matrix. Nevertheless, Yang et al. (1992) showed that the bi-level programming approach could be used as an efficient technique to achieve the simultaneous estimation of the o-d matrix and the route choice under congested traffic conditions. Florian and Chen (1995) developed, then, a Gauss-Seidel type coordinate descendent method for solving the considered problem formulated as a bi-level optimization problem, proposing again the analysis of Spiess (1990) for the same problem. The difference between these two methods concerns only the direction employed for the solution of the upper level problem.
The network equilibrium model adopted by Yang et al. (1992) is concerned only with the situation where the link cost functions are independent of each other. Namely, the travel time on a given link depends only on the flow through that link and not on the flow through any other link. Therefore, Yang (1995) presented an extended model, where the network equilibrium problem is formulated as variational inequalities, and two computational techniques for solving the bi-level model with link flow interaction.

One is an existing heuristic algorithm that solves the o-d matrix estimation problem and the equilibrium traffic assignment problem iteratively. The other one is a heuristic descent algorithm, in which sensitivity analysis method is used to calculate the derivatives of link flows with respect to o-d demand. The two heuristic algorithms are similar in the sense that the variations of link flows with respect to o-d matrix adjustments are taken into account explicitly in the system optimizing search process but differ from each other in the manner in which link flow variations are evaluated, that is the changes of link flow pattern in response to o-d matrix adjustments are evaluated in different ways.

Furthermore, Cascetta and Postorino (2001) proposed different fixed-point algorithms, namely, Functional Iteration, Method of Successive Averages, and Method of Successive Averages with Decreasing Reinitialisation.

2 NETWORK SENSOR LOCATION PROBLEM

2.1 Introduction

A theoretical issue strictly linked to the o-d matrix estimation problem using traffic counts is the Network Sensor Location Problem (NSLP), that is the identification of the optimal locations of a given number of link count sections. Optimal locations are the ones that provide for maximum information of the underlying o-d matrix given a budget constraint (a number of traffic link counts). In other words, the NSLP can be stated as the problem of identifying the set of sensors, maximizing the quality – or symmetrically minimizing the variability – of the estimated o-d matrix given a budget constraint: therefore, it may be formulated straightforwardly as the mathematical problem of optimizing a proper measure of quality/variability related to the o-d matrix estimation process.

This chapter first presents a review on the main measures of quality and location methods introduced in literature; then, it describes an innovative and theoretically founded NSLP formulation, explicitly accounting for the variability of both prior and posterior o-d estimates, being the posterior estimate the result of the updating of the prior o-d matrix by using the flows measured in the optimal sensor locations. The problem is formulated first in the general case and then particularized to the noteworthy case of multivariate normal distribution for the prior demand. Indeed, this assumption allows for a substantial simplification of the proposed approach, leading to the implementation of a sequential NSLP solving algorithm applicable to real networks. Importantly, the proposed framework can be effectively applied in the context of a budget allocation problem for demand estimation: since the variability of the posterior demand estimate depends both on the variability of the prior demand estimate and on the information related to the optimal link count locations, a budget allocation problem can be formulated in terms of trade-off between surveys and traffic counts, in order to minimize the overall demand variability.

2.2 Literature review

2.2.1 Measures of quality

A fundamental question arising when one tries to solve the network sensor location problem is how to select a measure that can quantify information gain from sensor measurements at various locations. In literature, several statistical measures have been proposed to evaluate the quality of an o-d demand estimator, such as the root mean square error and mean absolute error. These performance indices can be expressed as deviations in terms of o-d demand or link flows. Nevertheless, in the sensor location problem, the link flow observations are not available before installing the sensors, and the true o-d demand matrix is also generally unknown, so sensor location models should tend to construct indirect quality measures that do not require knowledge of the exact values of flows.

Yang et al. (1991) introduced the Maximum Possible Relative Error (MPRE) as a measure of variability of an estimated o-d trip matrix. In particular, the MPRE is defined as the maximum possible relative deviation of the estimated o-d matrix from the (unknown) true one, when the traffic counts are error free and the route choice proportions are correctly specified, that is also the assignment matrix is error free. The MPRE is then formulated as a quadratic programming problem, as briefly explained hereafter.

Let $\hat{d}_w$ denote the estimated trips between the o-d pair w, and d_w^* the true ones. Let p_{lw} be the proportion of trips between the o-d pair w using link l, and $\bar{y}_l$ the traffic count on link l. Thus, the following equations, expressing measured flows as a function of o-d matrix entries, must be satisfied:

$$\sum_w p_{lw} d_w^* = y_l \qquad (2.1)$$

$$\sum_w p_{lw} \hat{d}_w = \bar{y}_l \qquad (2.2)$$

Subtracting eq. (2.2) from (2.1), one can obtain the following equation:

$$\sum_w p_{lw}\left(d_w^* - \hat{d}_w\right) = 0 \quad (2.3)$$

If $\lambda_w = \left(d_w^* - \hat{d}_w\right)/\hat{d}_w$ denotes the relative deviation of the estimated trips from the true ones for o-d pair w, eq. (2.3) becomes:

$$\sum_w p_{lw}\hat{d}_w\lambda_w = 0 \quad (2.4)$$

$d_w^* \geq 0$, $\hat{d}_w \geq 0$; thus:

$$\lambda_w \geq -1 \quad (2.5)$$

Therefore, the MPRE can be formulated as the following optimization problem:

$$\max \Phi(\lambda) = \sum_w (\lambda_w)^2 \quad (2.6)$$

subject to constraint equations (2.4) and (2.5).
It is noteworthy that the MPRE does not show how actually reliable an estimated o-d matrix is because the estimation also depends on the prior information, that is the variability of the prior o-d matrix to be updated.

Gan et al. (2005) introduced a modified MPRE formulation, termed Expected Relative Error (ERE), representing the expected – instead of the maximum – error between the (unknown) true o-d matrix and the estimated o-d matrix: its calculation requires numerical simulation, under the assumption of uniform demand distribution within the feasibility set.
Importantly, calculation of both MPRE and ERE requires knowledge of an o-d matrix estimate consistent with the observed link flows. Therefore, the MPRE cannot be adopted as a measure to be optimized within a NSLP, since observed flows are unknown in the planning stage, i.e. when the optimal locations of sensors have not yet been identified.

From a theoretical standpoint, proper measures of variability/quality can be defined based on the volume of the polytope representing the feasibility set of the problem, i.e.

the set containing all o-d matrices consistent with the counted link flows. In practice, since the calculation of this volume is unfeasible for large dimensions, Bierlaire (2002) introduced the Total Demand Scale (TDS), defined as the difference between the maximal and the minimal total demand volume (defined as the sum of the o-d matrix entries) consistent with the counted flows.

Let $\hat{\boldsymbol{d}}$ be the estimated o-d matrix and let $\hat{\boldsymbol{f}}$ be the link flows vector obtained when $\hat{\boldsymbol{d}}$ is assigned on the network. The vector $\hat{\boldsymbol{y}}$ contains only the entries of $\hat{\boldsymbol{f}}$ corresponding to links where flow observations are available. $\boldsymbol{M}$ is the assignment matrix and $\widehat{\boldsymbol{M}}$ is the matrix composed of the rows of $\boldsymbol{M}$ corresponding to $\hat{\boldsymbol{y}}$.

The TDS is calculable by means of two constrained linear programming problems under the hypothesis of error-free counts and error-free assignment matrix:

$$\boldsymbol{\varphi}_{min} = \min_{\boldsymbol{d}} \boldsymbol{d}^{\mathrm{T}} \boldsymbol{e} \tag{2.7}$$

$$\widehat{\boldsymbol{M}}\boldsymbol{d} = \hat{\boldsymbol{y}}$$

$$\boldsymbol{d} \geq 0$$

where $\boldsymbol{e}$ is a vector composed only of ones, and

$$\boldsymbol{\varphi}_{max} = \max_{\boldsymbol{d}} \boldsymbol{d}^{\mathrm{T}} \boldsymbol{e} \tag{2.8}$$

$$\widehat{\boldsymbol{M}}\boldsymbol{d} = \hat{\boldsymbol{y}}$$

$$\boldsymbol{d} \geq 0$$

The TDS $\Phi\left(\hat{\boldsymbol{d}}, \widehat{\boldsymbol{M}}\right)$ was then defines by Bierlaire (2002) as:

$$\Phi\left(\hat{\boldsymbol{d}}, \widehat{\boldsymbol{M}}\right) = \boldsymbol{\varphi}_{max} - \boldsymbol{\varphi}_{min} \tag{2.9}$$

The interval $[\boldsymbol{\varphi}_{min}, \boldsymbol{\varphi}_{max}]$ is the range, or scale, of the total level of demand in the network.

Chen et al. (2005) used the TDS as one of the quality measures when analysing the properties of a path flow estimator.

2.2.2 Network sensor location methods

In literature many optimization techniques or heuristics for finding a set of link counts have been proposed.

Yang and Zhou (1998) proposed four basic rules of locating traffic counting points, namely the *o-d covering*, the *maximal flow fraction*, the *maximal flow interception* and *link independence* rules, some prototypically introduced by Lam and Lo (1990)[3], and developed integer linear programming models and heuristic algorithms to determine the counting links satisfying these rules.

In order for the MPRE to be finite (the o-d estimation error has to be bounded), the traffic counting sections on the study network must satisfy the *o-d covering* rule, that is they must be located so that a certain portion of the trips between any o-d pair can be observed for at least one link of their path. The problem of determining the minimum number and locations of traffic counting sections satisfying the o-d covering rule was formulated by Yang et al. (1991) as the following integer linear programming problem:

$$\min Z = \sum_{l \in L} z_l \qquad (2.10)$$

subject to:

$$\sum_{l \in L} \delta_{wl} z_l \geq 1 \qquad (2.11)$$

where z_l is a binary integer variable, $z_l = 1$ if the traffic counting section is located on link l, and 0 if otherwise; δ_{wl} is a (0, 1) constant, $\delta_{wl} = 1$ if some trips on one or more paths between o-d pair w pass over link l, and 0 otherwise; L is the set of the links of the network.

[3] Lam and Lo (1990) proposed some heuristic procedures for identifying the order in which the links should be selected.

The *maximal flow fraction* rule, whose proposition is based again on the theory of MPRE, states that, for a particular o-d pair, the traffic counting points on a road network should be located at the links so that the flow fraction between this o-d pair out of flows on these links is as large as possible. Indeed, if $\alpha_{lw} = p_{lw}\, d_w / y_l$ denotes the fraction of the flow between o-d pair w in link flow y_l, the term $(\frac{1}{\alpha_{lw}} - 1)$ is thus the upper bound of the relative deviation λ_w, and, clearly, the larger the value of α_{aw}, the smaller the upper bound of λ_w.

The *maximal flow fraction* rule leads to conflicts among different o-d pairs and may not always be desirable in selection of counting links. For instance, counting a low volume link may be less warranted, although such a link may carry a high fraction of trips between a minor o-d pair. However, supposing one wishes to identify the order in which the links should be selected from the set of links L for an effective o-d matrix estimation, a compromise could be made by setting a threshold value of the o-d flow fraction out of flow on a link, beyond which an o-d pair is said to be effectively covered by this link. Then one can make a priority list for the links in the set L in such a simple way that the candidate links are arranged in a descending order according to the number of o-d pairs effectively covered by the relevant links.

Yang and Zhou (1998), by using a simple example, illustrated the other two location rules. The *maximal flow interception* rule states that, under a certain number of links to be observed, the chosen links should intercept as many flows as possible. The *link independence* rule states that the traffic counting points should be located on the network so that the resultant traffic counts on all chosen links are not linearly dependent.

Therefore, when the route or link choice proportions of o-d trips are known (which are assumed to be given in most existing estimation models), the minimum number and locations of counting points can be obtained by solving the above-given problem (equations 2.7 and 2.8). Let $\boldsymbol{q}_0$ and $\boldsymbol{z}^{\boldsymbol{q}_0}$ denote, respectively, the minimum number and the corresponding location vector of traffic counting points. Yang and Zhou (1998) proved that $\boldsymbol{z}^{\boldsymbol{q}_0}$ satisfies the link independence rule. In general, *o-d covering* rule can be considered to be a fundamental rule that should be satisfied. Link independence rule can exclude links whose flow does not include any new information (unnecessary links), and should be satisfied too. Therefore, these two rules can be treated as constraints in determining traffic counting locations, while the *maximal flow fraction* and the *maximal flow interception* rules can be incorporated in the objective function to be maximized.

Considering the *maximal flow intercepting* rule, a criterion to solve the network sensor location problem is to maximize the total net flows observed. With reference to the path flow pattern associated with the prior o-d matrix, based on Hodgson (1990), this problem was formulated by Yang and Zhou (1998) as:

$$\text{Maximize } F(\boldsymbol{z}) = \sum_{r \in R} f_r x_r \tag{2.12}$$

subject to:

$$\sum_{l \in L} z_l = q \tag{2.13}$$

$$\sum_{l \in r} z_l \geq x_r \tag{2.14}$$

$$\sum_{l \in L} \delta_{lw} z_l \geq 1 \tag{2.15}$$

where f_r is the path flow associated with the prior o-d matrix, x_r is a binary variable indicating whether path *r* is observed ($x_r = 1$) or not ($x_r = 0$), and q is the prescribed number of counting points. Constraints (2.15) imply *o-d covering* rule. *Link independence* rule can be ensured for the minimum number of counting sites.

Further refinements were provided, for example, by Chung (2001) and Ehlert et al. (2006). Chung (2001) added the cost of purchasing and installing detectors into the count location problem, generating two distinct problems; the first one minimises the budget subject to complete o-d coverage, and the second problem maximises the coverage of o-d pairs subject to budget restrictions. Ehlert et al. (2006) proposed two extensions to these formulations. Firstly, the locations of detectors existing beforehand is taken into account. Secondly, the information content of the prior o-d flows is (optionally) taken into account, by identifying o-d pairs of particular interest and then weight these appropriately. The concept of information was applied to the o-d matrix estimation problem by Van Zuylen and Willumsen (1980). The average information content of an o-d movement is proportional to

$$H_w = -\, p_w \ln p_w \tag{2.16}$$

where p_w is the probability of a trip selected at random from the set of all trips passing between o-d pair w, as opposed to any other o-d pair. This probability could be estimated from the prior o-d matrix as follows:

$$p_w = \frac{d_w}{\sum_w d_w} \tag{2.17}$$

being d_w the traffic between o-d pair w.

To solve the location problems (2.10-2.11 and 2.12-2.15) different solution methods can be used. Generally, solution methods can be divided into “exact” algorithms and ‘‘heuristic’’ approaches. Exact methods guarantee an optimal result to the problem by employing various techniques to search the solution space. Heuristic methods can produce near-optimal solutions, but without defining the distance from an unknown optimal result. Exact methods include, for example, Branch and Bound algorithm. This technique produces an optimal result by applying an implicit enumeration procedure which efficiently eliminates infeasible and non-optimal solutions. The major difficulty of the Branch and Bound algorithm is that optimality of the solution cannot be guaranteed until all possible solutions have been eliminated from consideration, i.e. when the enumeration tree has been searched completely.

As regards the heuristic methods, Yang and Zhou (1998) proposed heuristics to solve the location problem (2.12-2.15), whereby the number of count locations is increased until all flows in the network are observed at least once, or a prescribed number of counting sites has been located. Many researchers discussed other heuristics based on a geographical and/or topological disaggregation of link flows, e.g. screen lines. In most large urban areas traffic measurement is frequently carried out at cordons and screen lines for data collection. Through screen line traffic survey, all traffic movements with the origin on one side of a screen line and the destination on the other are intercepted. This is contrasted with the traffic counting location problem based on the rules proposed by Yang and Zhou (1998), in which only a certain portion of the trips between any o–d pair is required to be observed for at least one link of their path. Traditionally, the selection of cordon and screen lines has been primarily based on subjective choices, depending on political jurisdictions, census area boundaries, and natural boundaries or man-made barriers, such as rivers and railway tracks.

The screen line based traffic counting location problem can be states as:

a) how to select the optimal locations for a given number of traffic counting stations to separate as many o-d pairs as possible;

b) how to determine the minimum number of counting stations and their locations required for separating all o–d pairs.

Here an o-d pair is regarded as separated when each of its feasible paths passes through at least one of the counting links and hence all trips between any o-d pair are observed for at least one link of their path.

Gan and Yang (2001) made an interesting initial attempt to address the screen line-based traffic-counting location problems. In particular, they provided an integer nonlinear programming model which is hereafter shown.

Let τ_l be a virtual travel time on link $l \in L$ and suppose τ_l is a function of z_l that is simple defined as $\tau_l(z_l) = z_l$ for all $l \in L$. Because each link has a non-negative virtual travel time, an appropriate shortest path algorithm such as Dijkstra method can be used to find the virtual shortest path and its corresponding virtual travel time between each o-d pair.

Let μ_w be the virtual shortest travel time between o-d pair $w \in W$. Clearly, μ_w is a function of the integer decision variables $z = (z_l)_{l \in L}$, and one can easily understand that, if $\mu_w(z) > 0$, then the virtual shortest path between o-d pair $w \in W$ includes at least one counting link. Because of the definition of link travel time function, it is straightforward to see that, if $\mu_w(z) > 0$, then the origin and destination for o-d pair w are separated by at least one screen line. Otherwise, if $\mu_w(z) = 0$, there exists at least one virtual shorter path with zero virtual travel time from its origin to destination that does not go through any counting link or cross any screen line.

For a given number of traffic counting stations $q < L$, the objective is to select their locations or counting links that constitute one or more screen lines to separate as many o-d pairs as possible; then the problem of interest can be formulated as:

$$\max_{\boldsymbol{z}} \sum_{w \in W} \delta\big(\mu_w(z)\big) \tag{2.18}$$

subject to:

$$\sum_{l=1}^{L} z_l \leq q; \;\; z_l \in \{0,1\} \tag{2.19}$$

where the not linear delta function $\delta(\mu) = 1$ if $\mu > 0$ and 0 if $\mu = 0$. Evidently, the objective function is the total number of o-d pairs that are separated, and the double counting effect is not accounted for. The integer maximization problem embodies a shortest path calculation (calculation of μ_w) as an internal procedure. In addition, all o-d pairs are viewed as equally important in the objective function of eq. (2.18), but a weighted objective function according to the relative magnitude of o-d demand between each o-d pair (if available) can be used instead.

The problem of determining the minimum number of counting links to separate all the o-d pairs in the network can be formulated as:

$$\min_{\boldsymbol{z}} \varphi\left[q - \sum_{w \in W} \delta\big(\mu_w(z)\big)\right] + \sum_{l=1}^{L} z_l \tag{2.20}$$

where φ is a sufficiently large positive constant. As shown in Gan and Yang (2001), as long as φ is selected to be larger than the total number of links in the network, then the optimal solution of eq. (2.20) gives exactly the minimum number and locations of counting links to separate all o-d pairs in the network.

Gan and Yang (2001) solved the model (2.20) by applying a genetic algorithm. Yang et al. (2006) reformulated the model (2.20) as an equivalent integer linear programming model using link-path incidence information and solved it by combining a column generation procedure and a branch and bound technique.

Further contributions for determining optimal screen lines for the purpose of o-d matrix estimation were provided by Chootinan et al. (2005) and Chen et al. (2007). In

particular Chootinan et al. (2005) extended the two single-objective problems to a bi-objective binary integer program. Then they developed a distance-based genetic algorithm solution procedure to solve the multi-objective screen-line-based traffic counting location problem.
Chen et al. (2007), using the selected traffic counts, estimated the o-d matrix through a modified path flow estimator which is capable of handling traffic count inconsistency internally.

Bianco et al. (2001) proposed an approach which consists of a two-stage procedure. Stage one is intended to solve the network sensor location problem, that is to determine how many sensors should be located and where, so as to infer the complete set of traffic flows in a transport network at a minimum measurement cost. This is interpreted as the cost of installing a set of traffic sensors in a network where no sensor has been placed yet. It is assumed that traffic sensors are to be located on the network nodes. A sensor located at a given node measures all traffic in- and out-flows. Each node is labeled with a weight reflecting the intersection complexity. The weight of each node (hence its measurement cost) is simply associated with the number of links incident to that node.
Based on the results of stage one, stage two produces an estimate of the o-d matrix. The basic assumption of their approach is that the values of turning coefficients at each network node are known.

It is noteworthy that all the above mentioned approaches do not explicitly take into account various error sources in the o-d estimation process. In fact, the quality of historical o-d demand estimates could significantly vary, depending on the date and size of the original survey conducted to obtain the prior o-d estimate, that is the o-d matrix to be updated by means of measurements information. Interestingly, extending the traffic state learning framework proposed by Eisenman et al. (2006), Zhou and List (2010) adopted an information-theoretic approach to examine the inherent connection between the sensor location problem and the o-d estimation problem. In particular, they explicitly modelled the mean and variance of the available prior o-d estimate and assumed as a measure of variability the trace of the covariance matrix of the posterior o-d estimate. Zhou and List (2010), who adopted a Kalman filter as a posterior demand estimator, used so a linear measurement equation to relate the unknown o-d demand vector $\mathbf{d}$ to measurements $\mathbf{y}$:

$$\mathbf{y} = \mathbf{Md} + \boldsymbol{\varepsilon}$$

where $\mathbf{M}$ is the assignment matrix and $\boldsymbol{\varepsilon}$ is a multivariate normal random vector of measurement errors.

Essentially, the o-d demand estimation problem is to find a new estimate $\mathbf{d}^{+}$ that can combine and use information from a given prior estimate $\mathbf{d}^{-}$ and sensor measurements. Therefore, let $\mathbf{P}^{-}$ be the error covariance matrix of the prior demand estimate, a transition equation was defined so as to obtain an updated posterior estimate $\mathbf{d}^{+}$ (with error covariance matrix $\mathbf{P}^{+}$) expressed by:

$$\mathbf{d}^{+} = \mathbf{d}^{-} + \mathbf{K}(\mathbf{Md} + \boldsymbol{\varepsilon} - \mathbf{M}_{\boldsymbol{\Lambda}}\mathbf{d}^{-}) \tag{2.21}$$

Given $\mathbf{d}_{\mathbf{err}} = \mathbf{d} - \mathbf{d}^{+}$, the classic Kalman filter aims to minimize the mean square error:

$$\mathrm{E}\left\|{\mathbf{d}_{\mathbf{err}}}^{\mathbf{2}}\right\| = \mathrm{E}[(\mathbf{d} - \mathbf{d}^{+})(\mathbf{d} - \mathbf{d}^{+})^{\mathrm{T}}] = \mathrm{Cov}(\mathbf{d}_{\mathbf{err}}) = \mathbf{P}^{+} \tag{2.22}$$

Substitution of (2.21) into (2.22) yields:

$$\mathbf{P}^{+} = \mathrm{Cov}[\mathbf{d} - \mathbf{d}^{-} - \mathbf{K}(\mathbf{M}_{\boldsymbol{\Lambda}}\mathbf{d} + \boldsymbol{\varepsilon} - \mathbf{M}_{\boldsymbol{\Lambda}}\mathbf{d}^{-})]\mathbf{b} \tag{2.23}$$

Using standard algebra and mild statistical assumptions, Zhou and List (2010) reformulated Eq. (2.23) independently of $\mathbf{d}$ and $\mathbf{d}^{-}$. This yields both the optimal value of the gain matrix $\mathbf{K}_{\mathbf{opt}}$ which minimizes the trace of the posterior dispersion matrix $\mathbf{P}^{+}$, and also a closed-form expression for the posterior demand estimate.

The NSLP has also been tackled by using pure network-based approaches, i.e. not considering any prior o-d matrix. This was proposed for instance by Hu et al. (2009), who dealt with a procedure, requiring explicit path enumeration, for the identification of the whole set of link flows starting from a counted subset. Along these lines, Ng (2012) achieved the relaxation of the assumption of explicit path enumeration required by Hu et al. (2009).

Furthermore, the matter of finding the set of counting locations allowing for full o-d observability was addressed by Castillo et al. (2008a, 2008b, 2010). The problems of observability of traffic networks are in general stated as:

- determine if a subset of available (observed) traffic flows is sufficient to obtain the values of another subset of (observable) traffic flows;
- obtain a minimum set of observations (o-d and/or link flows) that allow full observability, i.e., observability of all network flows;
- identify observable flows, given a set of observed flows (partial observability).

Furthermore, three different types of observability problems are distinguished, depending on the flows considered: link observability problems, that is based on link flows; o-d observability problem, in which o-d flows are estimated in terms of other flows; route flow observability problem, aimed to observe all route flows. Due to the fact that knowledge of the route flows immediately leads to the knowledge of the o-d and link flows, the observability problem of route flows is the most adequate from a practical point of view and can be considered as the full observability problem. However, it is the most difficult of the observability problems because its solution requires the maximum amount of information.

For a given network, if the assignment matrix is known, the further knowledge of the o-d flows is sufficient for the link flows to become known. In this case, it can be said that the set of link flows becomes observable if the o-d flows are observed (become known). Anyway, if one is interested in determining the o-d flows in terms of a subset of observed links, the question consists of determining which subsets of link flows allow for this estimation. In this case, the o-d pair flows become the unobserved flows and the observed links become the data (observed).

In order to solve the above problems Castillo et al. (2008b) proposed a topological method, thus providing a refinement to a previously shown algebraic method (Castillo, 2008a), that, basically, is aimed to express the observable flows in terms of the actually observed flows. In this case, two algorithms are shown, aimed to provide two sets and one matrix of interest: a set containing the list of a minimum number of measurements required for the full observability - these measurements are denominated *essential measurements;* a set containing the list of redundant measurements for observability purposes, i.e., even if the measurements are lost, the network remains observable; a matrix (which can easily be obtained from the network topology) containing the coefficients of the linear combinations of the redundant measurements in terms of the required (essential) measurements.

In addition, Castillo et al. (2010) proposed two theorems, one lemma and one corollary providing the bases for optimizing the proposed numerical procedures to solve the observability problems.

2.2.3 Optimal location of plate scanning devices

The NSLP can also be formulated in the presence of different types of measurements, e.g. automatic vehicle identification and plate recognition techniques, as proposed for instance by Castillo et al. (2008c) and by Minguez et al. (2010).

Let (N, L) be a traffic network, where N is the set of nodes and L is the set of links. From N one can distinguish two subsets of nodes, O and D, corresponding to origins and destinations, respectively. Let $\Lambda \in L$ be the set of $n_{sc} \neq 0$ observed links, containing information about plate number I_K, link l_K, and time t_K of registration, i.e. the information provided consists of the set:

$$SP \equiv (I_K, l_K, t_K); \quad k = 1,\dots, m; \; l_K \in \Lambda$$

where k is the kth plate scanned, and m is the total number of plates scanned.

As shown in Castillo et al. (2008c), the plate scanning technique consists of registering plate numbers and the corresponding times of the vehicles at some subset of links to reconstruct vehicle routes by identifying identical plate numbers at different locations and times. Castillo et al. (2008c) also pointed out that the set of links to be scanned must be chosen adequately so that all different combinations of scanned links belong to a unique route, which means that the scanning process allows identifying uniquely the path of any scanned user. The scanned observation can be thus summarized as:

$\hat{f}_r: r \in OR$

where *OR* is the set of observed routes, *OR* = 1,…, $n_r \in R$; *R* is the set of all considered routes and n_r is the number of different C_r sets of scanned links which allows to uniquely identify every observed route r. This information is used for route flow estimation by means of the following model:

$$\text{minimize}_{f_{r;\forall r \epsilon R}} \sum_{\forall x \epsilon R} \sum_{\forall y \epsilon R} (f_x - f_x^0) \gamma_{xy} (f_y - f_y^0) \tag{2.24}$$

$$\text{subject to } \hat{f}_r = f_r \; \forall r \epsilon OR; \;\; \hat{v}_l = \sum_{\forall r \epsilon R} \delta_l^r f_r \;\; \forall l \epsilon \Lambda \tag{2.25}$$

where f_r^0 and $\hat{f}_r$ are the prior and observed flows through route r, respectively; $\hat{v}_l$ is the observed flow at link *l*, and γ_{xy} are the weights (normally the elements of the inverse of the covariance matrix). Constraint (2.25) allows to include in the estimation model the total link flows, which are also known from the scanning process; this constraint includes redundant information for links where all the passing routes are observable, but it improves the prediction of unobservable route flows. The aim of this approach consists of identifying uniquely as many routes as possible through scanner devices in links.

Importantly, Castillo et al. (2008c) proposed also a binary linear programming which selects the minimum number of links to distinguish the users of any pair of routes. The plate number observations over this set of links, supposing that the scanning process is error free, allow to have a full identifiability of all path flows. The problem was formulated as follows:

$$\text{minimize}_z \; n_{sc} = \sum_{l \epsilon L} z_l \tag{2.26}$$

$$\text{subject to} \sum_{l \epsilon L} z_l \, d(r, r_1, l) \geq 1 \;\; \forall (r, r_1) | r \neq r_1 \tag{2.27}$$

$$\sum_{l \epsilon L} z_l \, \delta_l^r \geq 1 \quad \forall r \tag{2.28}$$

Where z_l is a binary variable such that it takes value 1 if the link l is scanned, and 0, otherwise; r and r_1 are paths; δ_l^r are the elements of the incident matrix, $\delta_l^r = 1$ if path r contains link l, $\delta_l^r = 0$ otherwise; $d(r, r_1, l) = 1$ if $\delta_l^r \neq \delta_l^{r_1}$, $d(r, r_1, l) = 0$ otherwise.

Constraint (2.27) guarantees that the selected subset of scanned links is able to distinguish the users of any given pair of paths r and r_1 based on their scanned plate numbers, i.e. there exists at least one scanned link which is in path r and not in path r_1 or vice versa. In addition, constraint (2.28) ensures that any route or path contains at least one scanned link, and, therefore, information, not only of all o-d pairs but all the routes, becomes available.

The optimal solution n_{sc}^* is the minimum number of scanning device positions provided by model (2.26)–(2.28) that allows estimating the o-d matrix exactly (error free) if all possible routes between any o-d pair have been considered. Nevertheless, model (2.26)–(2.28) does not include considerations on the budget, so in case the number of possible links to be installed is limited or the scanning device costs are different between links, this method is not suitable to get the best possible scanner locations reproducing as exactly as possible the o-d matrix with minimum cost.

Minguez et al. (2010) enriched the model proposed by Castillo et al. (2008c), considering alternative mathematical programming formulations to take into account some practical issues: budget minimization subject to complete route identifiability; maximum route identifiability subject to budget constraints; consideration of existing plate scanners.

2.3 Formulation of the proposed NSLP approach

2.3.1 Formulation in the general case

The proposed NSLP approach refers to the static uncongested framework, that is a linear relationship is assumed between the demand vector **D** and the link flows vector **F** by means of the assignment matrix **M**:

$$\mathbf{F} = \mathbf{MD} \tag{2.29}$$

M is assumed known, i.e. its estimate $\widehat{\mathbf{M}}$ is assumed error-free.
Let Λ be a set of λ counting sections, **Y** the corresponding vector of link flows[4] and $\mathbf{M}_{\mathbf{\Lambda}}$ the sub-matrix of **M** made up by the rows corresponding to the counting sections Λ. Equation (2.29) yields:

$$\mathbf{Y} = \mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} \tag{2.30}$$

The prior available o-d estimate **D** is always characterized by an inherent degree of variability, quantifiable either formally based on the estimation method or subjectively based on the analyst’s expectations. Therefore, **D** should be treated as a multivariate random vector, with a known prior density function $\varphi_{\mathbf{D}}(\mathbf{d})$ defined in a feasibility domain $\mathbf{\Omega}_{\mathbf{D}}$. In turn, this implies that **Y** is a random vector, induced by means of the relationship (2.30), with density function $\varphi_{\mathbf{Y}}(\mathbf{y})$ defined in a feasibility domain $\mathbf{\Omega}_{\mathbf{Y}}$. On the other hand, a realization **y** of the random vector **Y** (i.e. a set of measurements from the sensors within Λ) provides additional information about the random vector **D**, which, under the assumption of error-free measurements, is expressed implicitly again by (2.30):

$$\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y} \tag{2.31}$$

[4] The link flows are unknown in the planning stage, in which the approach is applied.

Formally, equation (2.31) implies a transition from the unconditional density $\varphi_{\mathbf{D}}(\mathbf{d})$ of the demand vector $\mathbf{D}$ to the density $\varphi_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}(\mathbf{d})$ of the random vector $\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D} = \mathbf{y}$ conditional upon a specific realization $\mathbf{y}$ of the link flows vector $\mathbf{Y}$, defined in a feasibility domain $\mathbf{\Omega}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}} \subseteq \mathbf{\Omega}_{\mathbf{D}}$.

In general, the variability related to a random vector $\mathbf{X}$, fully characterized by its distribution $\varphi_{\mathbf{X}}(\mathbf{x})$, can be expressed by means of aggregated measures: for instance, following Zhou and List (2010), the trace $\mathrm{Tr}[\mathbf{\Sigma}_{\mathbf{X}}]$ of the covariance matrix $\mathbf{\Sigma}_{\mathbf{X}}$ of $\mathbf{X}$ is a suitable and practically tractable measure of variability. Therefore, within the proposed framework, $\mathrm{Tr}[\mathbf{\Sigma}_{\mathbf{D}}]$ may represent a measure of the initial variability of the random vector $\mathbf{D}$ and $\mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]$the residual demand variability after accounting for the information provided by the counted flows $\mathbf{y}$ through (2.31). In accordance with that, since the trace $\mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]$ generally depends on the vector $\mathbf{y}$ of counted flows, which is a realization of the random variable $\mathbf{Y}$ with density function $\varphi_{\mathbf{Y}}(\mathbf{y})$ defined in the domain $\mathbf{\Omega}_{\mathbf{Y}}$, a *synthetic dispersion measure* (SDM) of the posterior random vector $\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D} = \mathbf{y}$ can be defined straightforwardly as the average of $\mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]$ within $\mathbf{\Omega}_{\mathbf{Y}}$:

$$\mathrm{SDM} = E\left[\mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]\right] = \int_{\mathbf{\Omega}_{\mathbf{Y}}} \mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]\varphi_{\mathbf{Y}}(\mathbf{y})\,\partial \mathrm{y} \tag{2.32}$$

Notably, the approach underlying equation (2.32) assumes the mean of the posterior random vector $\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D} = \mathbf{y}$ as statistical estimator for the demand vector $\mathbf{D}$. In this respect, apart from a scale factor, the trace $\mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]$ coincides with the corresponding MSE expression for the estimator in question.

As a result, the proposed NSLP is formulated as the problem of finding the optimal set Λ of sensors which minimizes the SDM (2.32), that is the variability associated with the posterior random vector $\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D} = \mathbf{y}$ over $\mathbf{\Omega}_{\mathbf{Y}}$, subject to budget constraints:

$$\Lambda^* = \arg\min_{b_\lambda \le b_{max}} \mathrm{SDM}(\Lambda) = \arg\min_{b_\lambda \le b_{max}} \int_{\mathbf{\Omega}_{\mathbf{Y}}} \mathrm{Tr}\left[\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]\varphi_{\mathbf{Y}}(\mathbf{y})\,\partial \mathrm{y} \tag{2.33}$$

where b_λ is the budget for a single sensor, $\mathrm{b}_{\max}$ the overall available budget and λ the cardinality of the set Λ. Mathematically, the set of sensors Λ can be represented in the NSLP (2.33) as a Boolean vector with cardinality equal to the number of links in the network and entries equal to one for the links included in Λ.

The trace $\mathrm{Tr}\left[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\right]$ is a function of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}}$ of the prior estimate of the random vector $\mathbf{D}$. Therefore the minimum achieved by the NSLP (2.33) depends both on the initial value $\mathrm{Tr}[\boldsymbol{\Sigma}_{\mathbf{D}}]$ (and thus on the budget allocated for the surveys leading to the prior o-d direct estimate) and on the information coming from the set Λ of sensors (and thus on the budget allocated for the counts): this allows formulation of a budget allocation problem between surveys and counts in order to minimize the variability of the overall demand estimation process.

The multidimensional integral over the feasibility domain $\boldsymbol{\Omega}_{\mathbf{Y}}$ of the counted flows in the objective function leads to a generally cumbersome application of the NSLP (2.33) on real networks. However, a significant simplification can be achieved if the random demand vector $\mathbf{D}$ is multivariate normally distributed, an assumption which allows also establishing a formal relationship between the NSLP (2.33) and the constrained GLS-based o-d matrix correction.

With regard to the impacts of possible relaxations of the hypotheses underlying the proposed approach, it should be noted that the removal of the error-free link count assumption would lead to a modification of equation (2.31), which should include a vector $\boldsymbol{\varepsilon}$ of measurement errors. Notably, any specific realization $\boldsymbol{\varepsilon}_{\mathbf{i}}$ of the vector $\boldsymbol{\varepsilon}$ can be incorporated into the vector of the known terms $\mathbf{y}$ and, therefore, the corresponding residual variability of the posterior demand vector would be still given by $\mathrm{Tr}\left[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}+\boldsymbol{\varepsilon}_{\mathbf{i}}}\right]$. Formally, this implies that the SDM (2.32) and the NSLP (2.33) should be modified by adding an outer integral dimension over the joint distribution function of the vector $\boldsymbol{\varepsilon}$. In this respect, the distribution function of $\boldsymbol{\varepsilon}$ can be defined either in absolute terms (i.e. all measurements have an absolute intrinsic error not depending on the actual counted value) or in relative terms (i.e. the measurement error is a percentage of the counted value). The former case can be handled relatively easily, whilst the latter would imply the measurement error to be itself a percentage of a random variable. Notably, if measurement errors are defined in absolute terms and independently normally distributed, the simplification proposed under the assumption of multivariate normal distribution for the prior demand is still likely to hold. With an analogous conceptual vehicle, the removal of the assumption of error-free assignment matrix $\widehat{\mathbf{M}}$ would lead to a further error source in equation (2.31), which can be treated, in principle, with the same mathematical complications related to the presence of measurement errors.

Similarly, different measurement sources may be incorporated provided that they can be expressed through additional linear equations to be coupled with equations (2.31): this is the case of AVI data. Consistent with the above, the presence of possible different error distributions for each measurement source might be incorporated as well, but only if they are expressed in absolute terms.

With the same argument, the proposed procedure might be suitably adapted to a dynamic uncongested assignment – i.e. with still linearity between the demand flows and link flows – through a proper modification of equation (2.31). Notably, the dimension of the problem and its tractability would be in such a case very cumbersome, due to the larger number of dimensions (i.e. o-d pair, link, departure time slice, link running time slice) of the assignment matrix.
Finally, extension to congested assignment does not however appear feasible due to the non-linearity introduced in the problem, and would in any case not lead to any practical gain in the proposed conceptual framework.

2.3.2 Particularization in the case of multivariate normal distribution for the prior demand

The assumption of multivariate normal distribution for the prior demand allows for a substantial analytical simplification of the SDM (2.32) and of the corresponding NSLP (2.33). In this respect, it should be noted firstly that the conditional random vector $\mathbf{D}|\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$ can always be expressed in a canonical form, $\mathbf{D}'|\mathbf{D}_{\mathbf{\Lambda}}' = \mathbf{y}'$, by means of a specific rotation $\mathbf{D}' = \mathbf{RD}$ of the demand vector $\mathbf{D}$, based on a proper rotation matrix $\mathbf{R}$. The conditional demand vector $\mathbf{D}|\mathbf{M}_\mathbf{\Lambda}\mathbf{D} = \mathbf{y}$ is in fact not in a ‘‘canonical’’ form in the sense that the conditioning given by $\mathbf{y}$ is expressed in implicit form on the components of $\mathbf{D}$, through the system of linear equations $\mathbf{M}_\mathbf{\Lambda}\mathbf{D} = \mathbf{y}$. Conversely, a canonical form $\mathbf{D}'|\mathbf{D}_\mathbf{\Lambda}' = \mathbf{y}'$ is intended in the sense of expressing explicitly such conditioning through equations where a single demand component is encompassed in each equation.

Notably, if $\mathbf{D}$ is multivariate normally distributed, i.e. $\mathbf{D}$ ~MVN($\boldsymbol{\mu}_\mathbf{D}$, $\boldsymbol{\Sigma}_\mathbf{D}$), $\mathbf{D}' = \mathbf{RD}$ is a linear transformation of a multivariate normal random vector and is therefore still a multivariate normal random vector, whose covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}'}$ can be expressed easily as a function of the covariance matrix $\boldsymbol{\Sigma}_\mathbf{D}$ through the well known formula:

$$\boldsymbol{\Sigma}_{\mathbf{D}'} = \mathbf{R}\boldsymbol{\Sigma}_\mathbf{D}\mathbf{R}^\mathbf{T} \tag{2.34}$$

In addition, the assumption of normality of $\mathbf{D}$ (and therefore of $\mathbf{D}'$) allows the Schur complement expression to be used for the conditional covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}'|\mathbf{D}_\mathbf{\Lambda}'=\mathbf{y}'}$ of the conditional vector $\mathbf{D}'|\mathbf{D}_\mathbf{\Lambda}' = \mathbf{y}'$ (e.g. Eaton, 1983):

$$\boldsymbol{\Sigma}_{\mathbf{D}'|\mathbf{D}_\mathbf{\Lambda}'=\mathbf{y}'} = \boldsymbol{\Sigma}_{\mathbf{D}'\delta-\lambda,\delta-\lambda} - \boldsymbol{\Sigma}_{\mathbf{D}'\delta-\lambda,\lambda}\left(\boldsymbol{\Sigma}_{\mathbf{D}'\lambda,\lambda}\right)^{-\mathbf{1}}\boldsymbol{\Sigma}_{\mathbf{D}'\lambda,\delta-\lambda} \tag{2.35}$$

where δ and λ denote the number of o-d pairs and link counts respectively, and the elements after the equal sign are the following four blocks of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}'}$ of the unconditional random vector $\mathbf{D}'$:

$$\boldsymbol{\Sigma}_{\mathbf{D}'} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{D}'\delta-\lambda,\delta-\lambda} & \boldsymbol{\Sigma}_{\mathbf{D}'\delta-\lambda,\lambda} \\ \boldsymbol{\Sigma}_{\mathbf{D}'\lambda,\delta-\lambda} & \boldsymbol{\Sigma}_{\mathbf{D}'\lambda,\lambda} \end{bmatrix} \tag{2.36}$$

As a consequence, $\mathbf{\Sigma_{D'|D_{\Lambda}'=y'}}$ depends in the new reference system[5] only on the blocks, defined by (2.36) of the covariance matrix (2.34) of the unconditional distribution of the random vector $\mathbf{D'}$, i.e. it does not depend on the specific values of $\mathbf{y'}$. Therefore, its trace $\mathbf{\Sigma_{D'|D_{\Lambda}'=y'}}$ does not depend on the actual value of the counted links, but just on which components of the demand vector are conditional to the set λ of sensors corresponding to the random vector **Y**. Recalling that the trace operator is invariant with respect to a change of basis, a substantial simplification of the SDM (2.32) is therefore pursued:

$$\mathrm{SDM} = \int_{\mathbf{\Omega_Y}} \mathrm{Tr}\left[\mathbf{\Sigma_{D|M_{\Lambda}D=y}}\right] \varphi_{\mathbf{Y}}(\mathbf{y})\, \partial \mathrm{y} = \int_{\mathbf{\Omega_Y}} \mathrm{Tr}\left[\mathbf{\Sigma_{D'|D'_{\Lambda}=y'{}'}}\right] \varphi_{\mathbf{Y}}(\mathbf{y}')\, \partial \mathrm{y}' =$$

$$\mathrm{Tr}\left[\mathbf{\Sigma_{D'|D'_{\Lambda}=y'{}'}}\right] =$$

$$= \mathrm{Tr}\left[\mathbf{\Sigma_{D'}}_{\delta-\lambda,\delta-\lambda} - \mathbf{\Sigma_{D'}}_{\delta-\lambda,\lambda}\left(\mathbf{\Sigma_{D'}}_{\lambda,\lambda}\right)^{-\mathbf{1}} \mathbf{\Sigma_{D'}}_{\lambda,\delta-\lambda}\right] \tag{2.37}$$

which in turn leads to a simplified definition of the NSLP (2.33):

$$\Lambda^* = \arg\min_{b_{\lambda|\Lambda|} \le b_{max}} \mathrm{Tr}\left[\mathbf{\Sigma_{D'|D'_{\Lambda}=y'{}'}}\right] =$$

$$= \arg\min_{b_{\lambda|\Lambda|} \le b_{max}} \mathrm{Tr}\left[\mathbf{\Sigma_{D'}}_{\delta-\lambda,\delta-\lambda} - \mathbf{\Sigma_{D'}}_{\delta-\lambda,\lambda}\left(\mathbf{\Sigma_{D'}}_{\lambda,\lambda}\right)^{-\mathbf{1}} \mathbf{\Sigma_{D'}}_{\lambda,\delta-\lambda}\right] \tag{2.38}$$

Zhou and List (2010) also solved an NSLP based on the minimization of the variability of the posterior demand vector, measured through the trace of its covariance matrix. Importantly, a major difference is that they adopted the Kalman filter as a posterior demand estimator, whilst the NSLP (2.38) assumes the mean of the conditional vector $\mathbf{D|M_{\Lambda}D = y}$ as posterior demand estimator. In addition, Zhou and List (2010) dealt with only the assumption of prior normal distribution for the demand vector, whilst equation (2.38), based on an error-free measurement assumption, is derived as a

[5] $\mathbf{\Sigma_{D'|D_{\Lambda}'=y'}}$ has dimension $(\delta - \lambda)$ $(\delta - \lambda)$, whilst $\mathbf{\Sigma_{D|M_{\Lambda}D=y}}$ has dimension $(\delta \cdot \delta)$ and rank $(\delta - \lambda)$. It occurs that $\mathbf{\Sigma_{D|M_{\Lambda}D=y}} = \mathbf{\Sigma^0_{D'|D_{\Lambda}'=y'}} \cdot \mathbf{R^{-1}}$ where $\mathbf{\Sigma^0_{D'|D_{\Lambda}'=y'}} = \begin{bmatrix} \mathbf{\Sigma_{D'|D_{\Lambda}'=y'}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ is the matrix obtained from $\mathbf{\Sigma_{D'|D_{\Lambda}'=y'}}$ by adding λ rows and columns of zeros.

particular instance of the more general approach (2.33), which does not require any specific distributional assumption on the prior demand vector.

It is worth pointing out that the solution to the NSLP (2.38) is the same which would be obtained by imposing the minimization of the trace of the covariance matrix of the constrained GLS estimator (CGLS), i.e. a GLS correction of the prior demand estimate $\mathbf{D}$ wherein equation (2.31) represents a constraint in accordance with the assumption of error-free link count measurements. Indeed, as reported by Cascetta (1984), the CGLS correction $\mathbf{D}_{\mathrm{CGLS}}$ of the prior o-d estimate $\mathbf{D}$ based on the error-free link count measurements $\mathbf{y}$ is given by the following estimator:

$$\mathbf{D}_{\mathrm{CGLS}} = \mathbf{D} + \boldsymbol{\Sigma}_{\mathbf{D}}{}^{-1}\mathbf{M}_{\boldsymbol{\Lambda}}{}^{\mathbf{T}}\left(\mathbf{M}_{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}_{\mathbf{D}}{}^{-1}\mathbf{M}_{\boldsymbol{\Lambda}}{}^{\mathbf{T}}\right)^{-1}(\mathbf{y} - \mathbf{M}_{\boldsymbol{\Lambda}}\mathbf{D})$$

with posterior covariance matrix:

$$\boldsymbol{\Sigma}_{\mathbf{D}_{\mathrm{CGLS}}} = \boldsymbol{\Sigma}_{\mathbf{D}} - \boldsymbol{\Sigma}_{\mathbf{D}}\mathbf{M}_{\boldsymbol{\Lambda}}{}^{\mathbf{T}}\left(\mathbf{M}_{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}_{\mathbf{D}}\mathbf{M}_{\boldsymbol{\Lambda}}{}^{\mathbf{T}}\right)^{-1}\mathbf{M}_{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}_{\mathbf{D}} \quad (2.39)$$

Although covariance matrices (2.39) and (2.34) are different, it is easy to recognize that they have the same trace. Indeed, this is intuitive: since the GLS estimator is a best linear unbiased estimator (BLUE), it necessarily minimizes the errors between the posterior estimate and the true o-d matrix, i.e. the trace of the covariance matrix of the posterior random demand vector expressed by (2.39). Therefore, in this respect, the problem (2.38) can be interpreted as tantamount to the problem of finding the set Λ^* minimizing the trace of the covariance matrix (2.39) of the $\mathbf{D}_{\mathrm{CGLS}}$ posterior estimate performed with Λ^*.

From an algorithmic standpoint, the three approaches (i.e. that of Zhou and List (2010), the proposed approach (2.38) and the minimization of the trace of the covariance matrix (2.39) of the CGLS estimator) share the same asymptotic computational complexity, since they basically involve a matrix inversion operation of the same order. Importantly, however, it is worth noting that the mathematics underlying the proposed approach described above allows for the implementation of an effective algorithm, capable of outperforming the others when the number of sensors becomes significant.

2.3.2.1 Mathematical details on the rotation of the prior demand vector

The analytical simplification of the SDM (2.32) and of the NSLP (2.33), under the assumption of multivariate normal distribution for the prior demand, is based on the possibility of expressing the conditional vector $\mathbf{D}|\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$ in a canonical form $\mathbf{D}'|\mathbf{D}_{\mathbf{\Lambda}}{}' = \mathbf{y}'$, by means of a specific rotation $\mathbf{D}' = \mathbf{RD}$ of the demand vector $\mathbf{D}$. For this aim, relationship (2.31) may be regarded as an undetermined linear system with $\mathbf{M}_{\mathbf{\Lambda}}$ as coefficients matrix and $\mathbf{D}$ and $\mathbf{y}$ as vectors of the unknowns and of the constant terms respectively.

As known from the algebra (e.g. Shilov (1977), Sheldon (1997), Lay (2005)), the whole set of the solutions of such system can be expressed as the sum of the generic solution of the associated homogeneous system $\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{0}$ and of a specific solution of the complete system $\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$. In other words, since the space of the solutions of $\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{0}$ is a subspace $Ker(\mathbf{M}_{\mathbf{\Lambda}}) \subseteq R^{\delta}$ called kernel of $\mathbf{M}_{\mathbf{\Lambda}}$, with dimension $\delta - \lambda$ (δ being the number of o-d pairs and λ the number of counting sections), the space of the solutions of the system $\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$ is a translation of $Ker(\mathbf{M}_{\mathbf{\Lambda}})$. This means that all vectors $\mathbf{D}$ solutions of $\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$ share the same components orthogonal to $Ker(\mathbf{M}_{\mathbf{\Lambda}})$, and differ among themselves only for components parallel to $Ker(\mathbf{M}_{\mathbf{\Lambda}})$.

Therefore, in order to express $\mathbf{D}|\mathbf{M}_{\mathbf{\Lambda}}\mathbf{D} = \mathbf{y}$ in a canonical form, it is sufficient to rotate the reference system of the space of the demand vectors, so as to find a new reference systems with directions respectively parallel and orthogonal to $Ker(\mathbf{M}_{\mathbf{\Lambda}})$: the orthogonal components will be those conditional upon the set Λ of sensors. Since the components orthogonal to $Ker(\mathbf{M}_{\mathbf{\Lambda}})$ form its image $Im(\mathbf{M}_{\mathbf{\Lambda}}) = span(rows(\mathbf{M}_{\mathbf{\Lambda}}))$ with dimension λ, a proper rotation can clearly be performed by means of the following rotation matrix:

$$\mathbf{R}=\begin{bmatrix} b^{1,1}_{Ker(\mathbf{M}_\Lambda)} & \dots & b^{1,\delta-\lambda}_{Ker(\mathbf{M}_\Lambda)} & i^{1,\delta-\lambda+1}_{Im(\mathbf{M}_\Lambda)} & \dots & i^{1,\delta}_{Im(\mathbf{M}_\Lambda)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ b^{1,\delta}_{Ker(\mathbf{M}_\Lambda)} & \dots & b^{\delta,\delta-\lambda}_{Ker(\mathbf{M}_\Lambda)} & i^{\delta,\delta-\lambda+1}_{Im(\mathbf{M}_\Lambda)} & \dots & i^{\delta,\delta}_{Im(\mathbf{M}_\Lambda)} \end{bmatrix}=$$

$$=\left[\mathbf{b}^{1}_{Ker(\mathbf{M}_\Lambda)} \quad \dots \quad \mathbf{b}^{\delta-\lambda}_{Ker(\mathbf{M}_\Lambda)} \quad \mathbf{i}^{\delta-\lambda+1}_{Im(\mathbf{M}_\Lambda)} \dots \quad \mathbf{i}^{\delta}_{Im(\mathbf{M}_\Lambda)}\right]=\left[\mathbf{B} \quad \mathbf{I}\right]$$

(2.40)

where $\mathbf{B}=\left\{\mathbf{b}^{1}_{\mathrm{Ker}(\mathbf{M}_\Lambda)},\dots,\mathbf{b}^{\delta-\lambda}_{\mathrm{Ker}(\mathbf{M}_\Lambda)}\right\}$ and $\mathbf{I}=\left\{\mathbf{i}^{\delta-\lambda+1}_{Im(\mathbf{M}_\Lambda)} \dots \quad \mathbf{i}^{\delta}_{Im(\mathbf{M}_\Lambda)}\right\}$ are orthonormal bases of the subspaces $Ker(\mathbf{M_\Lambda})$ and $Im(\mathbf{M_\Lambda})$ respectively. As a result, the rotation matrix (2.40) leads to a rotated demand vector $\mathbf{D}' = \mathbf{RD}$ such that the conditional random vector $\mathbf{D}|\mathbf{M_\Lambda D} = \mathbf{y}$ becomes a canonical conditional random vector $\mathbf{D}'|\mathbf{D_\Lambda}' = \mathbf{y}'$.

Notably, the above translation of $Ker(\mathbf{M_\Lambda})$, representing the space of the solutions of the system $\mathbf{M_\Lambda D} = \mathbf{y}$, defines the feasibility set of the demand vectors consistent with the counted flows $\mathbf{y}$, which is the same as that considered by Bierlaire (2002) for the calculation of the TDS, by Yang et al. (1991) for the MPRE, and Gan et al. (2005) for the ERE. In the proposed approach a considerable difference is that a specific distribution of the demand is assumed within the feasibility set, based on which an aggregate measure of the overall demand variability is established (i.e. the trace of the covariance matrix of the posterior demand vector).

2.4 Analysis of the proposed NSLP approach on toy networks

In order to illustrate the proposed framework and its related mathematics, two toy networks are introduced in this section: a 3-link toy network, under the assumptions of uniform prior demand distribution and of normal prior distribution; a 5-link toy network, mimicking a real situation where the prior demand variability is related to the sampling rate of the underlying process of demand estimation.

2.4.1 3-link toy network

The first test network, depicted on the left-hand side of Figure 1, comprises three links and two o-d pairs *A-C* (demand value D_1) and *B-C* (demand value D_2).

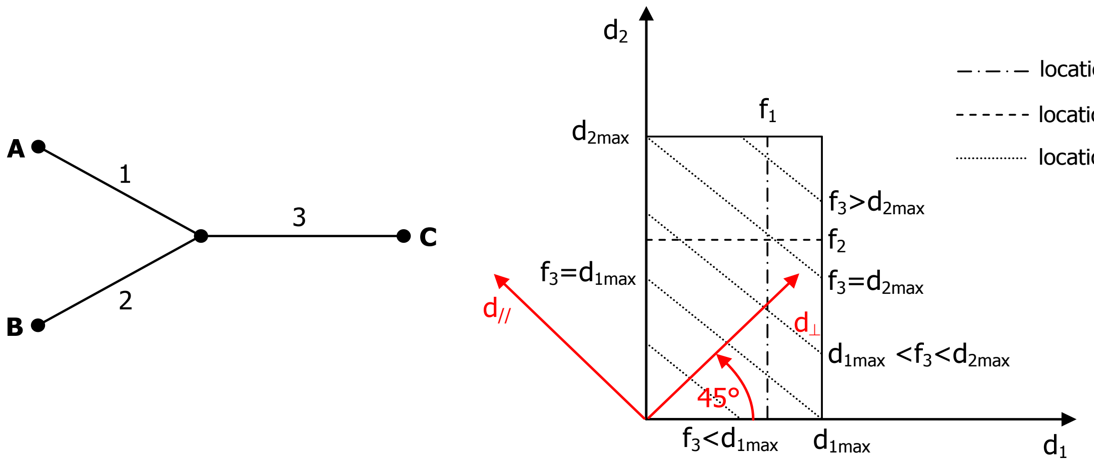


Figure 1 - 3-link toy network (left) and feasibility domains $\boldsymbol{\Omega}_{D|M_\Lambda D=y}$ depending on Λ for $\lambda = 1$ (right)

The objective is to find the best location for a single sensor, i.e. $\lambda = 1$. Both the o-d covering (whatever weights and thresholds) and the maximum flow fraction rules (Yang and Zhou, 1998) indicate link *3* as the best location, allowing also the feasibility set $\boldsymbol{\Omega}_{\mathbf{D}|\mathbf{M}_\Lambda \mathbf{D}=\mathbf{y}}$ to be bounded. Indeed, link *3* also represents the most reasonable subjective choice in the absence of any further prior information. However, in practice, prior information on $\boldsymbol{\Omega}_\mathbf{D}$ and $\varphi_\mathbf{D}(\mathbf{d})$ is normally available: for instance, upper bounds of the demand generated and attracted by a zone may be defined according to its population and workplaces respectively, and the variance of the generic demand value may be related to the adopted sampling strategy in the case of survey-based prior estimate. In such a situation, link *3* can be proved to be not the most effective choice through simple calculations, in line with Zhou and List (2010).

Firstly, let us consider availability of prior information on the domain $\boldsymbol{\Omega}_\mathbf{D}$ in terms of upper bounds on the demand generated, i.e. $D_1 \leq d_{1\,max}$ and $D_2 \leq d_{2\,max}$ with

$d_{2\,max} > d_{1\,max}$ (this assumption is obviously not restrictive), and absence of information on the prior distribution $\varphi_{\mathbf{D}}(\mathbf{d})$, which can thus be assumed uniformly distributed with independent marginals in the rectangle defined by the bounds $d_{1\,max}$ and $d_{2\,max}$. Its mean vector $\boldsymbol{\mu}_{\mathbf{D}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}}$ are therefore:

$$\boldsymbol{\mu}_{\mathbf{D}} = \begin{pmatrix} d_{1max}/2 \\ d_{2max}/2 \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_{\mathbf{D}} = \begin{bmatrix} d_{1max}^2/12 & 0 \\ 0 & d_{2max}^2/12 \end{bmatrix}$$

The right-hand side of Figure 1 depicts the feasibility set $\boldsymbol{\Omega}_{\mathbf{D}}$ and the corresponding sets $\boldsymbol{\Omega}_{\mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y}}$ given the possible locations of a single sensor (link *1*, *2* or *3*). Clearly, choosing link *1* implies fixing the demand value D_1 (i.e. $Var(D_1) = 0$ in the error-free measurements assumption) and similarly link *2* implies $D_2 = y_2$ (i.e. $Var(D_2) = 0$). In particular, for $\Lambda = \{1\}$ the mathematics introduced in Section 2.3 yields:

$$\mathbf{Y} \equiv Y_1; \quad \mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y} \rightarrow D_1=y_1; \quad \mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y} \rightarrow \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} | \, D_1=y_1 \rightarrow \begin{pmatrix} y_1 \\ D_2 \end{pmatrix}$$

$$\mathrm{Tr}[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y}}] = Var[D_2] = d_{2max}^2/12$$

Therefore, the trace $\mathrm{Tr}[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y}}]$ is independent of the unknown value of the counted flow y_1 in this particular situation, leading to the following SDM value from (2.31):

$$\mathrm{SDM}_{\Lambda=\{1\}} = \int_{\Omega_{\mathbf{Y}}} \mathrm{Tr}[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y}}]\, \varphi_{\mathbf{Y}}(\mathbf{y}) \partial\mathbf{y} = \mathrm{Tr}[\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{M}_{\Lambda}\mathbf{D}=\mathbf{y}}] = d_{2max}^2/12 \tag{2.41}$$

It is easy to verify that the same occurs for $\Lambda=\{2\}$, yielding:

$$\mathrm{SDM}_{\Lambda=\{2\}} = d_{1max}^2/12$$

(2.42)

Differently, the choice $\Lambda=\{3\}$ yields:

$\mathbf{Y}\equiv Y_3$; $\quad \mathbf{M}_\Lambda\mathbf{D}=\mathbf{y} \rightarrow D_1+D_2=y_3$; $\quad \mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y} \rightarrow \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} | D_1+D_2=y_3$

$$\mathrm{Tr}[\,\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\,] = \begin{cases} y_3^2/6 & \textit{if } 0 \le y_3 \le d_{1max} \\ d_{1max}^2/6 & \textit{if } d_{1max} \le y_3 \le d_{2max} \\ (d_{1max}+d_{2max}-y_3)^2/6 & \textit{if } d_{2max} \le y_3 \le d_{1max}+d_{2max} \\ 0 & \text{otherwise} \end{cases} \tag{2.43}$$

that is, the trace $\mathrm{Tr}[\,\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\,]$ depends now on the unknown value of the counted flow y_3, and an explicit calculation of the integral within (2.32) is required. In this respect, the components of the random vector **Y** are linear combinations of uniform random variables **D** by means of equation (2.30). Hence the density $\varphi_{\mathbf{Y}}(\mathbf{y}) = \varphi_{Y_3}(y_3)$ can be expressed as the convolution of the two uniform independent variables D_1 and D_2 defined respectively in $[0,d_{1max}]$ and $[0,d_{2max}]$, that is:

$$\varphi_{Y_3}(y_3) = \begin{cases} y_3/d_{1max}d_{2max} & \textit{if } 0 \le y_3 \le d_{1max} \\ 1/d_{2max} & \textit{if } d_{1max} \le y_3 \le d_{2max} \\ 1/d_{2max} - (y_3 - d_{2max})/d_{1max}d_{2max} & \textit{if } d_{2max} \le y_3 \le d_{1max}+d_{2max} \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

In turn, the SDM indicator becomes, by substituting (2.43) and (2.44) into (2.32):

$$\mathrm{SDM}_{\Lambda=\{3\}} = \int_{\Omega_{\mathbf{Y}}} \mathrm{Tr}[\,\mathbf{\Sigma}_{\mathbf{D}|\mathbf{M}_\Lambda\mathbf{D}=\mathbf{y}}\,]\,\varphi_{\mathbf{Y}}(\mathbf{y})\partial\mathbf{y} = \int_{y_3=0}^{d_{1max}+d_{2max}} \mathrm{Tr}[\,\mathbf{\Sigma}_{\mathbf{D}|D_1+D_2=y_3}\,]\,\varphi_{Y_3}(y_3)\partial y_3 = \frac{d_{1max}^2}{6} - \frac{d_{1max}^3}{12 d_{2max}} \tag{2.45}$$

Table 1 compares the SDM values (2.40), (2.42) and (2.44) for the three possible sensor locations for different values of d_{1max} and d_{2max} (and therefore for different values of the variances of D_1 and D_2). Unlike the best location indicated by both the o-d covering and the maximum flow fraction rules (i.e. link *3*), the best location indicated by the SDM indicator is always link *2,* apart from the case $d_{1max} = d_{2max}$ for which the choice is indifferent. This result is consistent with the theoretical principle that the link count location allowing for maximal reduction of the variability of the prior o-d estimate is the one gathering the most information about the demand component with maximal variance.

Table 1 - Best sensor location as a function of the SDM indicator for the Figure 1 network (hypothesis of uniform demand)

d_{1max}	d_{2max}	Var[D_1]	Var[D_2]	SDM			best location
				Λ={1}	Λ={2}	Λ={3}	
100	100	833	833	833	833	833	1 or 2 or 3
100	110	833	1008	1008	833	909	2
100	120	833	1200	1200	833	972	2
100	130	833	1408	1408	833	1026	2
100	140	833	1633	1633	833	1071	2
100	150	833	1875	1875	833	1111	2
100	160	833	2133	2133	833	1146	2
100	170	833	2408	2408	833	1176	2
100	180	833	2700	2700	833	1204	2
100	190	833	3008	3008	833	1228	2
100	200	833	3333	3333	833	1250	2

Let the random vector **D** now be multivariate normally distributed with covariance matrix:

$$\mathbf{\Sigma_D} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \tag{2.46}$$

As in the previous calculations, the cases Λ={*1*} and Λ={*2*} are trivial because they lead to conditional random vectors already expressed in "canonical" form and, therefore, equation (2.35) can be applied without any rotation of the reference system:

$$\mathbf{\Sigma}_{\mathbf{D}|D_1=y_1} = \sigma_2^2 - \sigma_{12} \cdot (\sigma_1^2)^{-1} \cdot \sigma_{12} = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \tag{2.47}$$

$$\mathbf{\Sigma}_{\mathbf{D}|D_2=y_2} = \sigma_1^2 - \sigma_{12} \cdot (\sigma_2^2)^{-1} \cdot \sigma_{12} = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \tag{2.48}$$

yielding the following SDM values:

$$\mathrm{SDM}_{\Lambda=\{1\}} = \mathrm{Tr}[\,\mathbf{\Sigma}_{\mathbf{D}|D_1=y_1}\,] = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \tag{2.49}$$

$$\mathrm{SDM}_{\Lambda=\{2\}} = \mathrm{Tr}[\,\boldsymbol{\Sigma}_{\mathbf{D}|D_2=y_2}\,] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \tag{2.50}$$

Conversely, in order to obtain a conditional random vector in canonical form for the case Λ={3}, the approach described in Section 2.3.2 should be followed, i.e. applying a 45° anticlockwise rotation of the reference system in order to obtain two axes $d_{//}$ and $d_{\perp}$ respectively parallel and orthogonal to the equation defined by the sensor Λ={3} (right side of Figure 1). The corresponding rotation matrix **R** can be obtained easily from (2.40), yielding, with also an intuitive geometrical interpretation:

$$\mathbf{R} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \tag{2.51}$$

The covariance matrix of the demand vector **D'** in the new reference system becomes, by substituting (2.46) and (2.51) into (2.34):

$$\boldsymbol{\Sigma}_{\mathbf{D'}} = \mathbf{R}\boldsymbol{\Sigma}_{\mathbf{D}}\mathbf{R}^{\mathrm{T}} = \begin{bmatrix} (\sigma_1^2 + \sigma_2^2)/2 + \sigma_{12} & (-\sigma_1^2 + \sigma_2^2)/2 \\ (-\sigma_1^2 + \sigma_2^2)/2 & (\sigma_1^2 + \sigma_2^2)/2 - \sigma_{12} \end{bmatrix}$$

and the covariance matrix of the conditional demand vector in the new reference system becomes, by applying equation (2.35):

$$\boldsymbol{\Sigma}_{\mathbf{D'}|D_{\perp}=y'_3} = \frac{\sigma_1^2 + \sigma_2^2}{2} + \sigma_{12} - \frac{\left(\sigma_2^2 - \sigma_1^2\right)^2}{4}\,\frac{1}{\frac{\sigma_1^2 + \sigma_2^2}{2} - \sigma_{12}} = \frac{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)}{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})} \tag{2.52}$$

which in accordance with (2.37) provides $\mathrm{SDM}_{\Lambda=\{3\}}$. The following Table 2 reports the calculation of the SDM (2.49), (2.50) and (2.52) depending on the position of the sensor, for $\sigma_1^2 = 1$ and for increasing values of σ_2^2 and σ_{12}.

Again, location 2 is the best, accordingly with the proposed framework, and increasing values of the covariance σ_{12} make count section 3 even worse with respect to section 2.

Table 2 - Best sensor location as a function of the SDM indicator for the network (hypothesis of normally distributed demand)

σ_1^2	σ_2^2	σ_{12}	correlation	SDM			best location
				Λ={1}	Λ={2}	Λ={3}	
1.00	1.00	0.00	0.0	1.00	1.00	1.00	1 or 2 or 3
1.00	1.00	0.10	0.1	0.99	0.99	1.10	1 or 2
1.00	1.10	0.00	0.0	1.10	1.00	1.05	2
1.00	1.30	0.00	0.0	1.30	1.00	1.13	2
1.00	1.70	0.00	0.0	1.70	1.00	1.26	2
1.00	2.00	0.00	0.0	2.00	1.00	1.33	2
1.00	2.00	0.14	0.1	1.98	0.99	1.46	2
1.00	2.00	0.28	0.2	1.92	0.96	1.58	2
1.00	2.00	0.42	0.3	1.82	0.91	1.69	2
1.00	3.00	0.00	0.0	3.00	1.00	1.50	2
1.00	3.00	0.17	0.1	2.97	0.99	1.63	2
1.00	3.00	0.35	0.2	2.88	0.96	1.74	2
1.00	3.00	0.52	0.3	2.73	0.91	1.84	2
1.00	3.00	0.69	0.4	2.52	0.84	1.93	2
1.00	3.00	0.87	0.5	2.25	0.75	1.98	2
1.00	3.00	1.04	0.6	1.92	0.64	2.00	2

2.4.2 5-link toy network

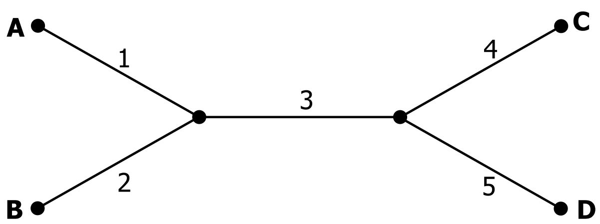
The 5-link toy network depicted in the left-hand side of Figure 2 has two origins (centroids *A* and *B*) and two destinations (centroids *C* and *D*). Each origin is associated with a synthetic population, wherein each individual has specific socio-economic and transport characteristics, corresponding to the aggregated values of inhabitants and generated trips reported in the right-hand side of Figure 2. In order to mimic a real application of the proposed methodology, a sampling estimate of the covariance matrix $\mathbf{S_D}$ of the prior demand vector is obtained by applying the theory of the statistical inference to a random sample selected within each origin, accordingly with a prefixed sampling rate. In particular, following Cascetta (1984), the sampling variances and covariances of the estimates are expressed as:

$$Var(\hat{d}_{ij}) = \frac{N_i^2}{n_i} P_{ij}(1 - P_{ij}) \quad (2.53)$$

$$Cov(\hat{d}_{ij}, \hat{d}_{ik}) = -\frac{N_i^2}{n_i} P_{ij} P_{ik} \quad (2.54)$$

$$Cov(\hat{d}_{ij}, \hat{d}_{lm}) = 0 \quad (2.55)$$

where N_i is the number of potential travellers, that is the synthetic population associated with origin i; n_i the number of travellers independently extracted; $P_{ij} = \frac{n_{ij}}{N_i}$, n_{ij} being the number of trips between origin i and destination j resulting from the sampling experiment.



	Centroid of origin	
	A	**B**
inhabitants	2000	5000
generated trips	1200	3300
to C	400	1650
to D	800	1650

Figure 2 - 5-link toy network (left) and setup of the experiment (right)

As in the previous section, the objective is to define the best location for a single sensor, under different combinations of sampling rates for each origin. Results are reported in the following Table 3 wherein, for different combinations of sampling rates, both the corresponding $\mathbf{S_D}$ estimate – together with its trace – and the SDM values for each single possible location are presented.

Notably, whilst the o-d covering and the maximal flow fraction rules indicate location *3* as the best choice, the sensor location minimizing the SDM is either link *4* or link *5* depending on the sampling rate.

Table 3 - Best sensor location for the network of Figure 2 under different sampling rates for direct estimation of prior demand (optimal locations highlighted in yellow)

sampling rate (%)		variances and covariances										Tr[S_D]	SDM				
zone A	zone B	σ^2_1	σ^2_2	σ^2_3	σ^2_4	σ_{12}	σ_{13}	σ_{14}	σ_{23}	σ_{24}	σ_{34}		Λ={1}	Λ={2}	Λ={3}	Λ={4}	Λ={5}
10	**10**	3200	4800	11055	11055	-1600	0	0	0	0	-5445	30110	28511	29786	29404	**20819**	20949
10	**20**	3200	4800	5528	5528	-1600	0	0	0	0	-2723	19055	17456	18893	18230	14381	**13866**
15	**10**	2133	3200	11055	11055	-1067	0	0	0	0	-5445	27443	26377	27119	26955	**17832**	18152
15	**15**	2133	3200	7370	7370	-1067	0	0	0	0	-3630	20073	19007	19858	19603	**13879**	13966
25	**20**	2133	3200	5528	5528	-1067	0	0	0	0	-2723	16388	15322	16227	15898	11806	**11715**
25	**30**	1280	1920	3685	3685	-640	0	0	0	0	-1815	10570	9931	10463	10282	7506	**7490**
35	**35**	914	1371	3159	3159	-457	0	0	0	0	-1556	8603	8146	8511	8402	**5949**	5986
40	**35**	800	1200	3159	3159	-400	0	0	0	0	-1556	8317	7918	8225	8142	**5636**	5698

2.5 Applications to real networks

In this section, two algorithms are introduced: *Algorithm 1*, which allows the objective function (2.37) of the NSPL (2.38) to be calculated without matrix inversion; *Algorithm 2*, which provides an effective sequential heuristic solution of the NSLP (2.38) based on *Algorithm 1*. Then the performances of the proposed method and of the most common NSLP approaches in the literature are compared by means of an application to a real highway context.

2.5.1 Algorithms to solve the proposed NSLP

Calculation of the objective function (2.37) of the proposed NSLP (2.38) first requires calculation of the rotation matrix **R** using expression (2.40), and then calculation of $\Sigma_{\mathbf{D'}}$ and $\Sigma_{\mathbf{D'|D_\Lambda'=y'}}$ using (2.34) and (2.35) respectively.

Asymptotically, the computational complexity of such operations is $O(\delta^3)$, the same exhibited by Zhou and List (2010) and by minimization of the trace of the CGLS covariance matrix (2.39).

In practice the most computationally demanding step in the above calculation of the objective function (2.37) is the inversion of the block $\mathbf{\Sigma}_{\mathbf{D}'\lambda,\lambda}$ in equation (2.35). Interestingly, for λ=1 (i.e. just one sensor) the matrix $\left(\mathbf{\Sigma}_{\mathbf{D}'\lambda,\lambda}\right)^{-1}$ in equation (2.35) becomes simply the reciprocal of the variance of the term $\sigma_{\mathbf{D}'\lambda,\lambda}$ of the rotated matrix $\mathbf{\Sigma}_{\mathbf{D}'}$, leading to a significant simplification of equation (2.35) itself:

$$\mathbf{\Sigma}_{\mathbf{D}'|\mathbf{D}_{\Lambda}'=\mathbf{y}'} = \mathbf{\Sigma}_{\mathbf{D}'\delta-\lambda,\delta-\lambda} - \mathbf{\Sigma}_{\mathbf{D}'\delta-\lambda,\lambda} \frac{1}{\sigma_{\mathbf{D}'\lambda,\lambda}} \mathbf{\Sigma}_{\mathbf{D}'\lambda,\delta-\lambda} \qquad (2.56)$$

Intuitively, given a set Λ of λ>1 sensors, this suggests a stepwise calculation of the objective function (2.37), in the form of the following *Algorithm 1*.

Algorithm 1 – *Stepwise calculation of the objective function (2.37) of the NSLP (2.38).*

- initialization: choose any arbitrary link processing order *L* within Λ and set $\mathbf{D_0}=\mathbf{D}$
- for each $i\leq\lambda$:
  - rotate the random demand vector $\mathbf{D_{i-1}}$ with dimension $\delta-(i-1)$ with respect to the single equation defined by link at position *i* in *L*, by means of the matrix $\mathbf{R_i}$ given by (2.40), yielding the rotated vector $\mathbf{D}_{\mathbf{i}-1}^{'} = \mathbf{R_i}\mathbf{D_{i-1}}$ with dimension still equal to $\delta-(i-1)$;
  - calculate the corresponding rotated covariance matrix $\mathbf{\Sigma}_{\mathbf{D}_{\mathbf{i-1}}^{'}} = \mathbf{R_i}\mathbf{\Sigma}_{\mathbf{D_{i-1}}}\mathbf{R_i}^{T}$ by means of (2.34);
  - calculate the covariance matrix $\mathbf{\Sigma}_{\mathbf{D}'_{\mathbf{i-1}}|(\mathbf{D}'_{\mathbf{i-1}})_{\Lambda=\{\mathbf{i}\}}=\mathbf{y_i}}$ (with dimension $\delta-i$) of the conditional random vector $\mathbf{D'_{i-1}}|(\mathbf{D'_{i-1}})_{\Lambda=\{i\}}=\mathbf{y_i}$ by means of (2.56);
  - set $\mathbf{D_i} = \mathbf{D'_{i-1}}|(\mathbf{D'_{i-1}})_{\Lambda=\{i\}}=\mathbf{y_i}$ and $\mathbf{\Sigma}_{\mathbf{D_i}} = \mathbf{\Sigma}_{\mathbf{D}'_{\mathbf{i-1}}|(\mathbf{D}'_{\mathbf{i-1}})_{\Lambda=\{\mathbf{i}\}}=\mathbf{y_i}}$.

At each iteration of *Algorithm 1* the starting matrix $\mathbf{D_{i-1}}$ has a stepwise reduced dimension with respect to the previous iteration. Coupled with the computational saving achieved by avoiding matrix inversion, this suggests a parsimonious heuristic for solving the NSLP (2.38), based on processing at the generic iteration *k* a set Λ_k of sensors obtained by adding just one more sensor to the set Λ_{k-1} processed in a previous iteration. In this way, the objective function at the generic iteration *k* can be calculated through just a single step of *Algorithm 1*, i.e. through a single one-dimension rotation.

This approach is synthesized in the following *Algorithm 2*.

Algorithm 2 – *Sequential heuristic solution of the NSLP* (2.38).

- initialization: define a maximal number of counting sections $\lambda_{\max} \leq n_l$ (being n_l the number of links of the network), based for instance on budget constraints. Set also $\Lambda_0=\{\varnothing\}$;

- for each $i \leq \lambda_{\max}$:

  - for each candidate link l_{cand} amongst the still available $n_l - i + 1$ links, i.e. those not belonging to the set Λ_{i-1} defined at the previous iteration;

  - augment the set Λ_{i-1} with link l_{cand}, i.e. $\Lambda_{i\,\mathrm{cand}} = \Lambda_{i-1} + \{l_{\mathrm{cand}}\}$;

  - calculate the objective function (2.38) for the set $\Lambda_{i\,\mathrm{cand}}$, by means of a rotation of the conditional random vector $\mathbf{D}_{\Lambda_{i-1}}$ given the set Λ_{i-1} (available from the previous iteration) with respect to the single equation defined by the link count location l_{cand} (i.e. a single step of *Algorithm 1*);

  - define Λ_i as the set amongst all $\Lambda_{i\,\mathrm{cand}}$ which minimizes the objective function (2.38).

2.5.2 Real case study: A3 Naples-Salerno motorway

The first application of the proposed approach concerns the A3 Naples-Salerno motorway, with 17 junctions over a 52 km length, 176 o-d pairs and 30 links (i.e. assignment matrix **M** with dimensions 30 x 176) corresponding to 29 independent measurement equations. A comprehensive survey at all toll stations was carried out in 2008, leading to the estimation of a prior o-d matrix together with its covariance matrix $\mathbf{S_D}$. The experiment consists of applying various NSLP methods and then performing a GLS o-d matrix correction based on each identified set of optimal count locations.

In detail, the proposed NSLP approach is compared with the most commonly quoted in the literature, that is:

- the maximal flow fraction and the o-d covering rule applied with different covering thresholds α (Yang and Zhou, 1998);

- the o-d weighted covering criterion by Ehlert et al. (2006).

Results of the different NSLP approaches are presented in Table 4 in terms of percentage reduction of the SDM (2.37) with respect to the trace of the prior covariance matrix, i.e. assuming the SDM as the correct measure to be optimized in an NSLP. For this reason, the benchmark is represented by the results of a branch and bound algorithm applied in order to find the global optimum of the NSLP (2.38). In detail, at each step a binary branching is performed by partitioning the sets of candidate sensors within the current nest into two nests, depending on the presence or not of the best link defined for that nest in accordance with a branching criterion. In practice, at the *i*-th step the branching criterion considers the *i*-th best link in the ordering identified by the proposed sequential heuristic. Table 4 also reports the percentage of all feasible combinations of sensor locations explored by the branch and bound algorithm for reaching the global optimum.

The o-d covering criterion requires 22 links for full coverage and for α>0 the number of links effectively contributing to the o-d covering decreases. Furthermore, the branch and bound algorithm was able to recognize the global optimum of the problem (2.38) for any number of link count locations. Most importantly, also the proposed *Algorithm 2* almost always achieves this global optimum (or finds a very close solution), outperforming the other methods.

The proposed case study also allows the calculation times of the NSLP approach by Zhou and List (2010) to be compared with those of the proposed NSLP approach solved by means of *Algorithm 2*. For this aim, the objective function by Zhou and List (2010) is embedded in the sequential heuristic of *Algorithm 2* in place of the objective function (2.37): this allows proper comparison of the computational effort required by the two approaches. Figure 3 below draws the calculation times (performed using an Intel Pentium (R)4 CPU 3.06 GHz with 4Gb RAM) for each possible number of sensors λ in the real case study presented above. Notably, for low λ the approach by Zhou and List (2010) is slightly more efficient but, as soon as the number of counts increases, the proposed NSLP approach leads to a faster response.

Table 4 - O-d matrix correction results for different NSLP approaches (% reduction of the SDM with respect to the trace of the prior covariance matrix)

Link counts	SDM (percentage reduction)								% of solutions explored by the branch and bound
	Proposed methodology (sequential)	Maximal flow fraction	o-d coverage			o-d weighted coverage		Branch and bound (global optimum)	
			α=0.00	α=0.02	α=0.05	variance	entropy		
1	3.93	2.97	0.45	2.74	0.51	3.76	3.76	3.93	-
2	9.07	8.58	4.33	3.11	2.36	8.58	8.46	9.07	39.54
3	13.60	9.81	7.35	3.69	5.45	9.81	13.28	13.60	20.17
4	18.43	14.50	9.36	4.21	5.82	14.50	14.50	18.43	8.40
5	22.49	17.40	9.72	4.53	9.71	17.40	19.04	22.49	3.60
6	26.71	19.25	14.02	8.99	11.03	20.64	23.35	26.71	1.33
7	30.78	22.50	14.54	9.07	11.49	25.79	26.25	30.78	0.46
8	33.82	22.74	15.71	11.70	15.41	29.50	28.12	33.82	0.21
9	36.50	25.44	17.69	13.67	19.11	33.94	31.38	36.50	0.10
10	38.91	30.59	17.76	15.89	20.50	35.82	31.61	38.91	0.05
11	40.89	34.32	22.62	20.78	20.63	40.03	34.32	41.11	0.03
12	42.85	35.27	25.59	21.01	-	41.03	35.27	43.52	0.01
13	44.78	37.00	25.81	21.53	-	41.25	39.71	45.49	0.01
14	45.87	41.44	26.28	-	-	43.44	41.44	47.42	0.00
15	47.68	41.80	31.23	-	-	46.16	41.80	49.07	0.00
16	49.34	42.17	35.65	-	-	46.57	42.17	50.40	0.00
17	50.64	42.25	36.95	-	-	48.29	46.46	51.81	0.00
18	51.93	46.55	38.94	-	-	50.21	46.55	53.15	0.00
19	53.38	48.76	40.98	-	-	50.59	48.76	54.21	0.00
20	54.44	50.68	41.17	-	-	50.68	50.68	54.68	0.00
21	54.91	51.20	41.48	-	-	51.20	51.20	55.12	0.01
22	55.35	53.03	42.03	-	-	53.03	53.03	55.60	0.01
23	55.81	54.46	-	-	-	54.46	53.45	56.01	0.03
24	56.22	54.52	-	-	-	54.88	54.88	56.36	0.10
25	56.59	54.94	-	-	-	54.94	54.94	56.63	0.39
26	56.87	56.33	-	-	-	56.33	55.48	56.87	1.67
27	57.01	56.37	-	-	-	56.88	56.88	57.01	10.81
28	57.15	56.91	-	-	-	56.91	57.15	57.15	-
29	57.19	57.19	-	-	-	57.19	57.19	57.19	-

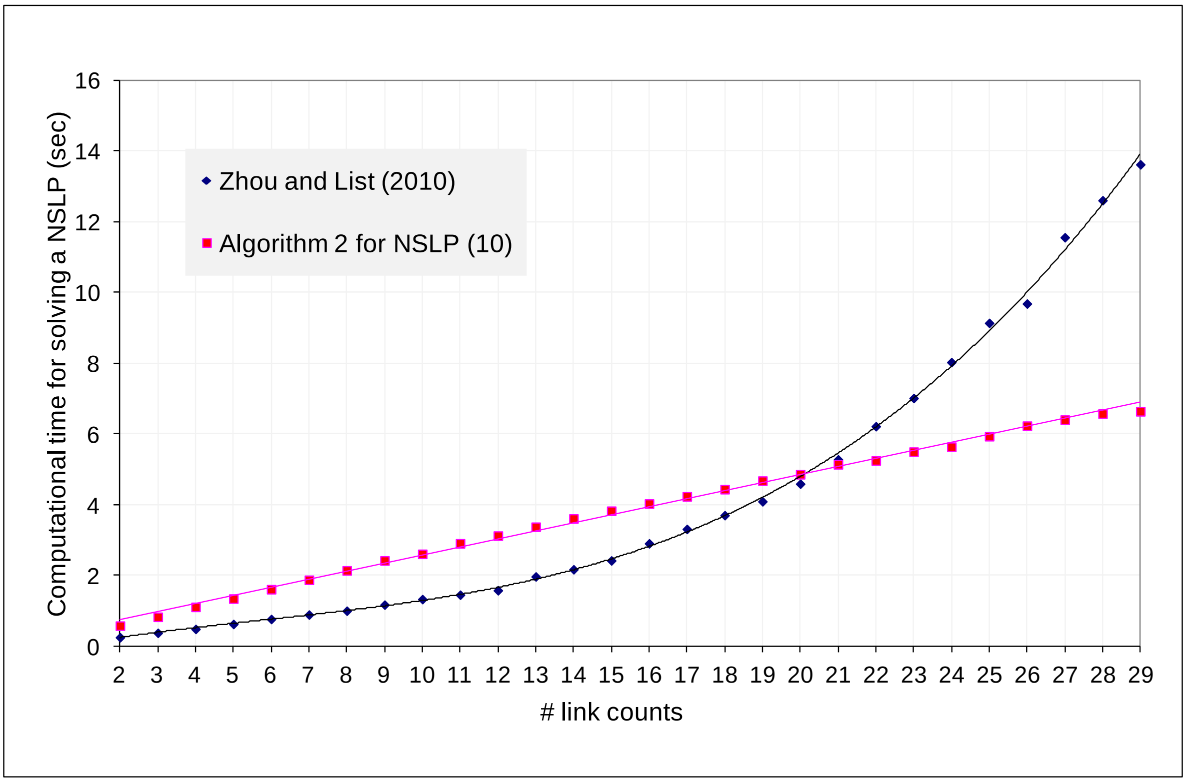


Figure 3 - Comparison of computational times between the proposed sequential algorithm and Zhou and List (2010) on the A3 motorway network

2.6 Budget allocation problem

This section deals with the practical instance of a budget allocation problem between surveys and traffic counts in the planning stage, mentioned in Section 2.3.1. The experiment deals with a common important planning decision in transport engineering, i.e. how to allocate a budget b_{max} between surveys and traffic counts in order to minimize the variability of the o-d matrix estimate. Indeed, in accordance with the proposed approach, a higher budget allocated to surveys will reduce the variability of the sampled covariance matrix $\mathbf{S_D}$ of the prior estimate, but fewer counts will be available for further correction. Vice versa – and this happens more frequently in the practice – allocating most of the budget to counts rather than to surveys leads to a more variable initial demand estimate but more link counts available for correction. Therefore, assuming a cost b_{sur} for a single survey and a cost b_{cs} for a count section, different combinations of the number n_{sur} of surveys and λ of link counts can be defined such that $n_{sur}b_{sur}+\lambda b_{cs}=b_{max}$, and for each combination a prior estimate of **D** can be inferred and a subsequent $\mathbf{D_{GLS}}$ correction can be performed. Therefore, for each pair (n_{sur}, λ) the mean coefficient of variation of the posterior demand estimate can be calculated, given by:

$$\frac{1}{\delta}\sum_{i}\frac{\sigma_{d_{GLS_i}}}{d_{GLS\,i}}$$

δ being the number of o-d pairs and $d_{GLS\,i}$ and $\sigma_{d_{GLS_i}}$ the mean and the standard deviation of the *i*-th component of the vector $\mathbf{D_{GLS}}$ respectively. Obviously, when performing only surveys, the mean variation coefficient is calculated on the basis of the mean and the standard deviation of the components of the vector D. Furthermore, the mean variation coefficient was used instead of the trace for a clearer interpretation of the results.

The real test site is the interurban network of the Province of Benevento (Italy), with about 300,000 inhabitants over 2000 km^2, made up by 1722 o-d pairs, 1800 links and 766 independent equations. Results are reported in Figure 4 below, where b_{sur}=10€ and two different values for b_{λ} are assumed, i.e. 50€ and 20€ respectively, so as to mimic the cases of both a full day of counts and just peak-hour counts.

The red line in Figure 4 depicts the reduction of the posterior demand variability as a function of a budget entirely allocated to surveys ("all-surveys" curve): the shape of this curve is convex, that is the marginal gain of a further survey is remarkably significant for low budgets (i.e. few surveys) and tends to decrease for higher sampling rates. On the contrary, the curves representing the decrease in posterior demand variability as a function of the number of available counts exhibit a concave shape, that is a significant gain is observed only for substantial network coverage (i.e. a high number of counts).

Interestingly, all count curves tend to fall under the all-surveys curve: therefore, the intersection between the all-surveys curve and the "all-counts" curve (i.e. the curve corresponding to a minimal budget allocation to surveys and all the remaining counts) identifies an indifference point where the budget can be allocated equivalently either on surveys or on counts, leading to the same posterior variance reduction.

As a consequence, a first clear outcome is that for low budgets – unable to achieve a good network coverage, i.e. before the aforementioned indifference point – a full allocation to surveys is preferable. Conversely, a budget overcoming the indifference point should be invested entirely on traffic counts up to a full network coverage, after which the residual budget, if any, can be allocated to surveys. Importantly, the budget threshold corresponding to the indifference point should be evaluated case by case, since it depends on specific factors (e.g. network topology, relative cost of surveys and counts). The example provided shows the optimal budget allocation between surveys and counts to be a non-trivial issue, to be assessed through proper quantitative tools and approaches.

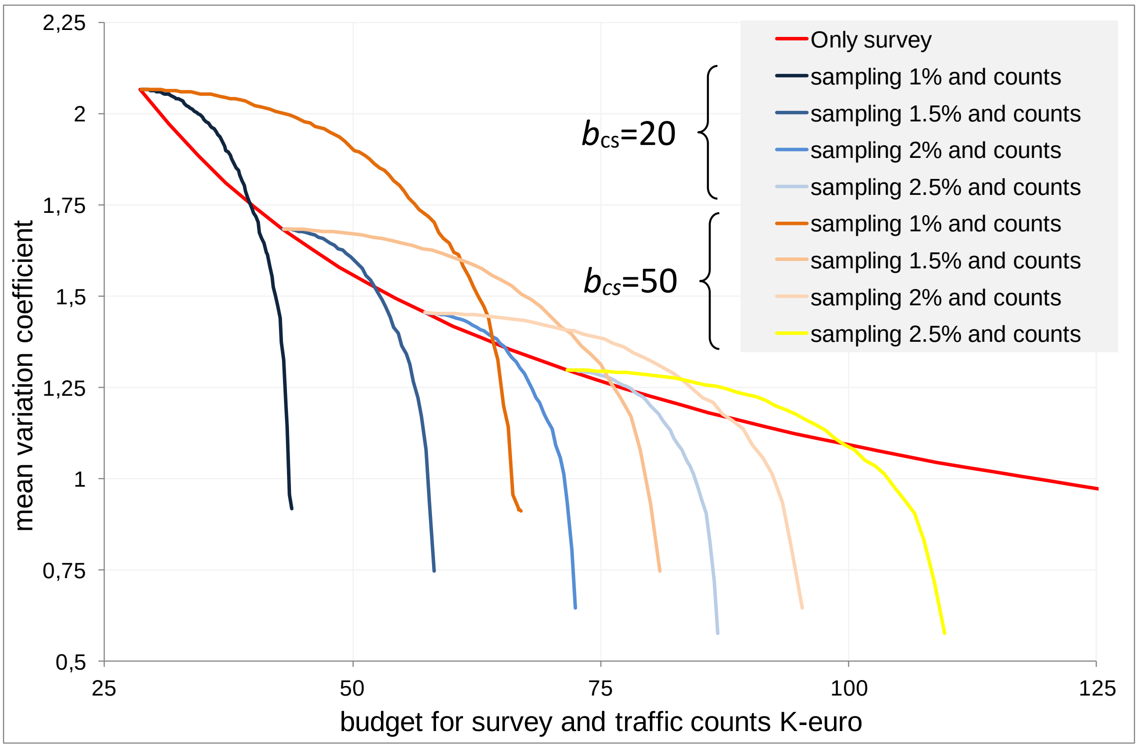


Figure 4 - Application of the proposed procedure to the network of the Province of Benevento, with different allocations of budget between surveys for prior estimates and traffic counts for posterior GLS correction.

3 DYNAMIC O-D ESTIMATION FROM TRAFFIC COUNTS

3.1 Introduction

As shown by the laboratory experiments on real-size synthetic networks carried out by Marzano et al. (2009), a satisfactory updating – regardless of the quality of the prior estimate – can be obtained only when the ratio between the number of equations (i.e. independent observed link flows) and the number of unknowns (i.e. o-d flows) is close to one. Unfortunately, such unknown/equation balance cannot be achieved in real networks under a static assumption, because the number of o-d pairs is always structurally larger than the number of counted links. On the other hand, achieving a proper unknown/equation balance is not straightforward also in dynamic contexts, where o-d estimation problem deals with the estimation of different o-d matrices, each corresponding to a time interval. Therefore, given a number of time intervals, the number of equations and the number of unknowns increase proportionally in the same way. However, under reasonable hypotheses on o-d flow variation across time slices, the number of unknowns in within-day dynamic systems can be bound, thus achieving unknowns/equations ratios close to one. In fact, in such dynamic contexts it is possible to formulate several hypotheses about the temporal evolution of the demand; the most commonly used approach consists in modelling the within-day evolution of o-d flows across time slices as an autoregressive process.

Along this research direction, this chapter proposes a "quasi-dynamic" framework for estimation of o-d flows, hinted by Marzano et al. (2009), in which o-d shares are assumed constant across a reference period, whilst total flows leaving each origin are assumed varying for each sub-period within the reference period. In particular, after a literature review on some dynamic o-d matrix estimators, this chapter checks whether real data confirm the quasi-dynamic assumption from an empirical perspective, and then compares the performances of the quasi-dynamic o-d matrix estimation with both classical off-line dynamic estimators and with other possible evolution rules characteristics of on-line dynamic estimation.

3.2 Literature review on dynamic estimation of o-d flows

In literature, two types of dynamic or time-dependent o-d estimation problems have been proposed: on-line (or real-time) and off-line. On-line estimation exploits the continuous flow of surveillance information to allow the dynamic updating of model inputs for each time interval of interest. Off-line estimation is aimed to determine a set of time-dependent o-d matrices given a time-series of link traffic counts (and other information such as travel times, historical o-d flows, etc). Off-line estimation is more relevant to planning or evaluation studies, while on-line estimation is required to provide – rapidly and recursively – o-d estimates for recent time slices together with predictions for subsequent time slices, to be used to generate traffic information and predictions for ITS applications.

3.2.1 On-line estimation of o-d flows

The first extension of the static o-d updating to the within-day dynamic framework – i.e. given a time horizon T of duration t_T divided into $n_\theta = t_T/t_\theta$ time slices θ of duration t_θ, estimating n_θ o-d matrices using n_θ vectors of link counts – was provided by Cascetta et al. (1993) through the proposition of two estimators: simultaneous and sequential.

While the simultaneous estimator - whose specification is shown in Section 3.2.2 - is usually used for off-line estimation, the sequential estimator is suitable for on-line estimation problems. Sequential estimator is based on the estimation at each step the o-d flows for a given time slice θ expressed as a function of the traffic counts within θ and (some) of the already estimated previous o-d matrices. This estimator provides two advantages: the first is the reduction of computational complexity by breaking down a large optimization problem into a number of smaller and more manageable ones; the second is the possibility of using the estimates obtained for a time slice as prior estimates for estimations in subsequent time slices.

The Kalman filter algorithm (Kalman 1960) has been widely used to accommodate the on-line requirements. The first approach was proposed by Okutani and Stephanades (1984) and subsequently generalized by Ashok and Ben-Akiva (1993 and 2000) and Ashok (1996), who formulated as a state-space model, modelling the within-day evolution of o-d flows across time slices as an autoregressive process and using a Kalman filter to predict o-d flows for the time slice $\theta+1$ based on link flow measurements at the time slice θ.

In order to develop a state-space model, a “state” should be defined. In particular, Ashok (1996) used the vector of deviations of o-d flows from best available historical estimates instead of the actual flows themselves as “state”, that is as unknown variables.

A state-space model typically consists of a set of transition equations and measurement equations: transition equations describe the evolution of the “state” over time and measurement equations, on the other hand, relate the unknown “state” variables to the other measured variables. Therefore, once a “state” is defined, one needs to specify transition and measurement equations. In this case, the transition equation expresses o-d flows related to a time slice θ as a result of the update of an historical estimate $\mathrm{d}_{\mathrm{Hod}}^{\theta}$ (typically the outcome of the filtering process at the previous day) by means of a within-day autoregressive process of order p based on the deviations from the historical estimates observed for the p time slices prior to θ:

$$\mathrm{x}_{\mathrm{od}}^{\theta+1} = \mathrm{d}_{\mathrm{Hod}}^{\theta+1} + \sum_{\theta'=\theta_{\mathrm{p}}}^{\theta} \varphi_{\mathrm{od}}^{\theta'} \left(\mathrm{x}_{\mathrm{od}}^{\theta'} - \mathrm{d}_{\mathrm{Hod}}^{\theta'} \right) + \alpha_{\mathrm{od}}^{\theta+1} \qquad \forall\ od,\ \theta \in T \tag{3.1}$$

wherein $\varphi^{\theta}{}_{od}$ are the coefficients of the autoregressive process, θ_p the farthest time slice included in the autoregressive process (i.e. the number of time slices between θ_p and θ included equals p) and $\alpha^{\theta}{}_{od}$ an error term.

The measurement equation is instead represented by the dynamic network loading equations:

$$\hat{y}_l^{\theta} = \hat{y}_{Hl}^{\theta} + \sum_{\theta'=\theta_l}^{\theta} \sum_{od=1}^{n_{od}} m_{odl}^{\theta'\theta}\left(x_{od}^{\theta'} - d_{Hod}^{\theta'}\right) + \mu_l^{\theta} \qquad \forall\ l \in \Omega_{lc},\ \theta \in T \tag{3.2}$$

wherein $\hat{y}_l^{\theta}$ represents the flow on link l at the time slice θ and $\hat{y}_{Hl}^{\theta}$ its historical value; $m_{odl}^{\theta'\theta}$ the percentage of users of the o-d pair starting from the origin o at the time slice θ' and crossing the link l at the time slice θ; $x_{od}^{\theta'}$ the flow starting from the origin o toward the destination d at time slice θ'; $\mu^{\theta}{}_l$ a measurement/assignment error; Ω_{lc} the set of counted links.

Equations (3.1) and (3.2) can be coupled and a direct closed form solution found, i.e. an optimal demand forecast for the time slice θ+1 based on traffic counts and prior estimates up to time slice θ. The filter applied at θ involves all the demand unknowns backward up to θ_l due to the dynamic assignment map in equation (3.2) and to θ_p due to the order of the autoregressive process in equation (3.1). Therefore, the filter encompasses jointly the demand unknowns for all the time slices between θ and the farthest between θ_p and θ_l or, equivalently, the o-d flow vector for each time slice is corrected a number of times equal to the number of time slices between θ and the farthest between θ_p and θ_l.

If a given network has n_{od} o-d pairs, let $\mathbf{X}_{\theta+1}$ be the $(n_{od} * 1)$ vector of all deviations in o-d flows at the time slice $\theta + 1$ and $\widehat{\mathbf{X}}_{\theta+1}$ its estimate. Denote by $\mathbf{Y}_{\theta+1}$ the vector of all deviations in the observed traffic counts and by $\mathbf{Q}_{\theta+1}$ and $\mathbf{R}_{\theta+1}$ the covariance matrices of the error terms of the equations (3.1) and (3.2) respectively.
In addition, consider the matrix $\mathbf{\Phi'}_{\theta}$, including all the coefficients $\varphi^{\theta}{}_{od}$ of the autoregressive process, and the matrix $\mathbf{M}_{\theta+1}$, including all the elements $m_{odl}^{\theta'\theta}$ of the equations (3.2).
The solution, comprising the Kalman Filter, was provided by Ashok (1996) through the following equations:

$$\mathbf{\Sigma}_{0|0} = \mathbf{P}_0 \tag{3.3}$$

$$\boldsymbol{\Sigma}_{\theta+1|\theta} = \boldsymbol{\Phi}_{\theta}\boldsymbol{\Sigma}_{\theta|\theta}\boldsymbol{\Phi}'_{\theta} + \mathbf{Q}_{\theta+1} \tag{3.4}$$

$$\mathbf{K}_{\theta+1} = \boldsymbol{\Sigma}_{\theta+1|\theta}\mathbf{M}'_{\theta+1}\left(\mathbf{M}_{\theta+1}\boldsymbol{\Sigma}_{\theta+1|\theta}\mathbf{M}'_{\theta+1} + \mathbf{R}_{\theta+1}\right)^{-1} \tag{3.5}$$

$$\boldsymbol{\Sigma}_{\theta+1|\theta+1} = \boldsymbol{\Sigma}_{\theta+1|\theta} - \mathbf{K}_{\theta+1}\mathbf{M}_{\theta+1}\boldsymbol{\Sigma}_{\theta+1|\theta} \tag{3.6}$$

$$\widehat{\mathbf{X}}_{0|0} = \overline{\mathbf{X}}_{0} \tag{3.7}$$

$$\widehat{\mathbf{X}}_{\theta+1|\theta} = \boldsymbol{\Phi}_{\theta}\widehat{\mathbf{X}}_{\theta|\theta} \tag{3.8}$$

$$\widehat{\mathbf{X}}_{\theta+1|\theta+1} = \widehat{\mathbf{X}}_{\theta+1|\theta} + \mathbf{K}_{\theta+1}\left(\mathbf{Y}_{\theta+1} - \mathbf{M}_{\theta+1}\widehat{\mathbf{X}}_{\theta+1|\theta}\right) \tag{3.9}$$

$\widehat{\mathbf{X}}_{\theta+1|\theta}$ represents a one-step prediction of the state $\widehat{\mathbf{X}}_{\theta+1}$. It is assumed that the initial state of the system $\mathbf{X}_0$ has known mean $\overline{\mathbf{X}}_0$ and variance $\mathbf{P}_0$. $\boldsymbol{\Sigma}_{\theta+1|\theta}$ and $\boldsymbol{\Sigma}_{\theta+1|\theta+1}$ represent the variances of $\widehat{\mathbf{X}}_{\theta+1|\theta}$ and $\widehat{\mathbf{X}}_{\theta+1|\theta+1}$. The matrix $\mathbf{K}_{\theta+1}$ is called *gain* matrix.

An important aspect concerning the filtering technique is the requirement that $\mathbf{P}_0$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{M}$, $\boldsymbol{\Phi}$ and $\overline{\mathbf{X}}_0$ be known. For this aim, repeated measurement over several days are needed. Since this may be unrealistic in most practical applications, no filter design is really optimal. This then raises the question of whether one could deduce non-optimal behaviour during operation and improve the quality of filter performance. Within certain limits, this is possible and is known in literature as *Adaptive Filtering*.

The Kalman filter can also be applied to non-linear systems. An algorithm used to dealt with the problem of non-linear estimation in dynamic systems is the *Extended Kalman Filter* algorithm, which involves a first-order Taylor linearization of the measurement equation about the best available estimate of the "state" vector. In other words, the non-linear system can be approximated by linearization during each time slice about the latest state estimate. Estimates obtained from the *Extended Kalman Filter* algorithm could be improved by performing successive iterations of linearization and re-estimation, leading to an *Iterated Extended Kalman Filter* algorithm.

Along this line, in a day-to-day framework, instead of an autoregressive process, Zhou and Mahmassani (2007) assumed a polynomial approximation for the structural deviation of the demand with respect to the historical estimate and for some of its derivatives. In particular, Zhou and Mahmassani (2007) considered the true demand partitioned into three components, namely, the regular pattern, structural deviations and random fluctuations. Theoretically, only the prior estimate of the regular demand, reflecting prior survey data and surveillance information up to the previous day, is available before performing real-time estimation on the current day. For this reason, the true demand was modelled by Zhou and Mahmassani (2007) as a linear combination of the prior estimate, structural deviation and random disturbance; the random disturbance term was assumed to follow a Normal distribution with zero mean and a polynomial trend model was introduced to describe the structural deviations.

Computational issues in within-day o-d estimation in large networks were addressed by Bierlaire and Crittin (2004), who proposed a least-square formulation of the real-time dynamic o-d estimation and prediction problem, based on a combination of the approaches by Cascetta et al. (1993) and Ashok and Ben-Akiva (1993).

The issue of dealing with an expanded measurement space, including not only traffic counts, was explored for instance by Barcelò et al. (2012), who proposed procedures, based on Kalman Filter, that explicitly exploit traffic data available from Bluetooth sensors. They assumed: flow counting detectors and ICT (Information and Communication Technologies) sensors located in a cordon and at each possible point for flow entry (centroids of the study area), and ICT sensors also located at intersections in urban networks, covering access and links to/from the intersection; flows and travel times available from ICT sensors for any selected time interval length higher than one second; trip travel times from origin entry points to sensor locations as measures provided by the detection layout. Expansion factors from equipped vehicles to total vehicles, in a given interval, could be estimated by using the inverse of the proportion of ICT counts to total counts at centroids, and were assumed to be shared by all o-d paths and pairs with common origin centroid and initial interval.

The linear formulation of the Kalman Filtering approach, proposed by Barcelo et al., used deviations of o-d paths flows as state variables, as suggested by Ashok and Ben-Akiva (1993 and 2000) and Ashok (1996), calculated in respect to DUE-based historical o-d path flows for equipped vehicles. Nevertheless, the approach of Barcelo at al. differs in that it doesn't require an assignment matrix, by using, instead, the subset of the most likely o-d paths flows identified from a DUE assignment (conducted with the historical o-d flows).

3.2.2 Off-line estimation of o-d flows

Simultaneous estimator - proposed by Cascetta et al. (1993) - jointly estimates all o-d matrices for all time slices using the whole set of traffic counts, assuming the dynamic assignment matrix known. Its specification, directly derived from estimator (1.18), is

$$\mathbf{d}^{\mathrm{GLS}} = \arg\min_{\mathbf{x}\geq\mathbf{0}} \sum_{\theta=1}^{n_\theta} \left(\hat{\mathbf{d}}_\theta \mathbf{x}_\theta\right) \mathbf{V}_{\boldsymbol{\theta}}{}^{-1} \left(\hat{\mathbf{d}}_\theta - \mathbf{x}_{\boldsymbol{\theta}}\right)^{\mathrm{T}} + \sum_{\theta=1}^{n_\theta} (\hat{\mathbf{y}}_\theta - \mathbf{M}_{\boldsymbol{\theta}} \mathbf{x}_{\boldsymbol{\theta}}) \mathbf{W}_{\boldsymbol{\theta}}{}^{-1} (\hat{\mathbf{y}}_\theta - \mathbf{x}_{\boldsymbol{\theta}})^{\mathrm{T}} \tag{3.10}$$

This estimator is usually used for off-line estimation problem, even if it can be shown to be inefficient for moderate size networks, as reported in Cascetta and Russo (1997) and Toledo et al. (2003).

The Kalman filter can be used also for off-line applications, as proposed by Balakrishna et al. (2005) and particularly by Gelb (1974), who suggested a double-step off-line estimation approach based firstly on a forward Kalman filter application (as in on-line estimation) and then on a backward Kalman smoothing, in order to account for the knowledge of link counts for all time slices in off-line contexts. Ashok (1996) claimed this double-step approach to provide more robust and reliable results with respect to a simple forward Kalman filter.

Cipriani et al. (2011) also proposed a method suitable for off-line applications. In particular, the problem of simultaneous estimation of time-dependent o-d matrices was formulated by Cipriani et al. (2011) without assignment matrices, by using measures such as traffic counts and speeds and introducing constraints with respect to the generated flows from the origins; a solution algorithm, that is a modification of the gradient based SPSA -Simultaneous Perturbation Stochastic Approximation, Spall (1998a, 1998b) - path search optimization method, was adopted. This algorithm works with a gradient approximation and can find a good solution when the starting point (that is the seed matrix) is assumed to be ‘‘near” the optimal one. The modification concerned both the step size computation and the gradient estimation and it was proposed in order to improve the efficiency of the algorithm.

Other relevant research directions within this field – not followed in this thesis - dealt with joint estimation of demand and supply parameters; this is the case of Liu and Fricker (1996) who, indeed, addressed networks with stochastic characteristics (that is travelers' perceptions of link costs vary) and extended the o-d estimation model to include the route choice parameters at the same time.

3.3 Quasi-dynamic estimation of o-d flows

3.3.1 Definition of quasi-dynamic o-d flows

A satisfactory estimation/updating of o-d flows – independent of the quality of the prior estimate – can be obtained only for an unknown/equation ratio close to one, as shown by Marzano et al. (2009). It is immediate to recognize that such condition cannot be met in static systems, where the number n_{od} of o-d pairs is much larger than the number n_l of links in the network and therefore even more larger than the number n_{lc} of counted links, i.e. $n_{od} >> n_l \geq n_{lc}$. Also in within-day dynamic contexts the balance is clearly always unsatisfactory: indeed, given a time horizon T of duration t_T divided into $n_\theta = t_T/t_\theta$ time slices of duration t_θ, the number of available link counts is $n_{lc} \cdot n_\theta$ and the number of o-d unknowns is $n_{od} \cdot n_\theta$. Therefore, from the above, $n_{od} >> n_{lc}$ implies $n_{od} \cdot n_\theta >> n_{lc} \cdot n_\theta$.

However, intuitive behavioural considerations may help in achieving a better balance between unknowns and equations in a within-day dynamic context. First of all, let the generic o-d flow d^{θ}_{od} for the time slice θ be expressed as the product between the demand g^{θ}_{o} generated by zone o during the time slice θ and the distribution probability $p^{\theta}_{d|o}$ of choosing the destination zone d moving from o within the time slice θ, i.e. $d^{\theta}_{od} = g^{\theta}_{o} \cdot p^{\theta}_{d|o} \, \forall \theta$. Taking into account the underlying phenomena leading to the commonly observed demand patterns, factors affecting g^{θ}_{o} are inherently within-day time varying (i.e. number of persons leaving from o in θ) while factors affecting $p^{\theta}_{d|o}$ are more within-day stationary (i.e. localizations of houses and workplaces, spatial impedances between pairs of zones). Consequently, given a sub-period $\tau \subseteq T$ of duration $t_{\tau} \leq t_{T}$ within the entire time horizon T encompassing a number $n_{\theta|\tau} = t_{\tau}/t_{\theta}$ of subsequent time slices θ, the distribution probability $p^{\theta}_{d|o}$ of the $n_{\theta|\tau}$ time slices θ within τ may be reasonably approximated by its average $p^{\tau(\theta)}_{d|o}$ over τ, yielding:

$$d^{\theta}_{od} = g^{\theta}_{o} p^{\theta}_{d|o} \cong g^{\theta}_{o} p^{\tau(\theta)}_{d|o} = d^{\theta,qd}_{od} \qquad \forall \theta \in \tau \subseteq T \tag{3.11}$$

wherein:

$$g^{\theta}_{o} = \sum_{d} d^{\theta}_{od} \qquad \forall \theta \in \tau \subseteq T \tag{3.12}$$

$$p^{\tau(\theta)}_{d|o} = \frac{d^{\tau}_{od}}{g^{\tau}_{o}} = \frac{\sum_{\theta \in \tau} d^{\theta}_{od}}{\sum_{\theta \in \tau} g^{\theta}_{o}} = \frac{\sum_{\theta \in \tau} d^{\theta}_{od}}{\sum_{\theta \in \tau} \sum_{d} d^{\theta}_{od}} \qquad \forall \theta \in \tau \subseteq T$$

(3.13)

and $\tau(\theta)$ maps time slices θ and corresponding sub-periods τ, i.e. $\tau(\theta)$ represents the specific sub-period τ including the time slice θ. In the following, $\mathrm{d}^{\theta,\mathrm{qd}}_{\mathrm{od}}$ will be referred to as a "quasi-dynamic" o-d flow, since it comes from the assumption of constant (i.e. static) distribution probabilities and variable (i.e. dynamic) generation profiles within τ.

Prior real evidence of this assumption is reported in the literature in the contributions by Cascetta et al. (1993) on Italian data, by Ashok (1996) on data related to the Boston area and by Van der Zijpp (1996) on Dutch data. Notably, both Van der Zijpp (1996), who assumed known generation profiles g^{θ}_{o}, and Ashok (1996), who modelled two different autoregressive processes for demand generation and distribution, seminally contains a quasi-dynamic formulation, but not inspired by the explicit purpose of reducing the number of unknowns and thus improving the quality of the o-d flows estimation process.

Interestingly, the quasi-dynamic assumption allows reducing the number of unknowns from $n_\theta \cdot n_{od}$ to $n_\theta \cdot n_o + n_\tau \cdot (n_{od} - n_o)$, where $n_\tau = t_T / t_\tau$ represents the number of sub-periods τ within T. Therefore, the unknown/equation ratio becomes:

$$\frac{n_\theta \cdot n_o + n_\tau \cdot (n_{od} - n_o)}{n_\theta \cdot n_{lc}} \tag{3.14}$$

which approaches the target value of one when the number of counted links becomes:

$$\frac{n_\theta \cdot n_o + n_\tau \cdot (n_{od} - n_o)}{n_\theta \cdot n_{lc}} \cong 1 \;\rightarrow\; n_{lc} \cong n_o + \frac{n_\tau \cdot (n_{od} - n_o)}{n_\theta} \;\rightarrow\; n_{lc} \cong n_o + \frac{1}{\bar{n}_{\theta|\tau}} \cdot (n_{od} - n_o) \tag{3.15}$$

where $\bar{n}_{\theta|\tau} = E\left[n_{\theta|\tau} \; \forall \tau \in T\right]$ is the average number of time slices θ per sub-period τ.

Equation (3.15) shows that, thanks to the quasi-dynamic assumption, the number n_{lc} of traffic counts needed to balance equations and unknowns should be at least equal to the number of origins plus a quantity opportunely limitable through $\bar{n}_{\theta|\tau}$. Importantly, this is generally possible in standard networks, wherein $\bar{n}_{\theta|\tau}$ may be augmented by increasing the duration t_τ of the sub-period τ and/or by reducing the duration t_θ of the time slice θ. In the first case, this implies assuming the quasi-dynamic hypothesis to hold for larger time intervals of duration t_τ, a circumstance which should be checked carefully on real data: the following Section 3.5 will be devoted specifically to this issue. In the second case, very limited durations t_θ may result in very high correlations between measurements across subsequent time slices, therefore limiting the possibility of obtaining further independent equations. In that respect, Marzano et al. (2009) already showed, through laboratory experiments, that an acceptable lower bound for t_θ is three minutes.

3.3.2 Formulation of the proposed QD-GLS estimator

The estimator (1.14) may be easily particularized for the quasi-dynamic assumption (3.11), yielding:

$$
\begin{aligned}
&\{\, \mathbf{g}^{*1},\ldots,\mathbf{g}^{*\theta},\ldots,\mathbf{g}^{*n_\theta}\,;\mathbf{p}^{*1},\ldots,\mathbf{p}^{*\tau},\ldots,\mathbf{p}^{*n_\tau}\,\} = \\
&= \underset{\substack{\mathbf{x}^1\ldots\mathbf{x}^{n_\theta}\in S_x \\ \boldsymbol{\pi}^1\ldots\boldsymbol{\pi}^{n_\tau}\in S_\pi}}{arg\,min}\left[z_1\left(\mathbf{x}^1\ldots\mathbf{x}^{n_\theta},\boldsymbol{\pi}^1\ldots\boldsymbol{\pi}^{n_\tau},\hat{\mathbf{d}}^1\ldots\hat{\mathbf{d}}^{n_\theta}\right)+z_2\left(f\left(\mathbf{x}^1\ldots\mathbf{x}^{n_\theta},\boldsymbol{\pi}^1\ldots\boldsymbol{\pi}^{n_\tau}\right),\hat{\mathbf{y}}^1\ldots\hat{\mathbf{y}}^{n_\theta}\right)\right]
\end{aligned}
\tag{3.16}
$$

wherein $\mathbf{g}^\theta$ is the $(n_o\cdot 1)$ vector of the generated demands g^θ_o for a given time slice θ, $\mathbf{p}^\tau$ is the $(n_o\cdot n_d)$ matrix of the distribution probabilities $p^\tau_{d|o}$ for a given sub-period τ, $\hat{\mathbf{d}}^\theta$ the $(n_o\cdot n_d)$ matrix of the prior demand estimates $\hat{d}^\theta_{od}$ for the time slice θ, and $\hat{\mathbf{y}}^\theta$ the $(n_{lc}\cdot 1)$ vector of the observed link counts $\hat{\mathbf{y}}^\theta_l$ for the time slice θ.

Therefore, the unknowns of the problem (3.16) are the demand generation profiles $\mathbf{x}^\theta$ for each time slice θ and the matrices of the distribution shares $\boldsymbol{\pi}^\tau$ for each sub-period τ, respectively variable in the feasibility sets S_x and S_π.

In equation (3.16), the functional forms of $z_1(\cdot)$ and $z_2(\cdot)$ depend on the general estimation approach (i.e. GLS, ML or Bayesian) and on the assumptions on the statistical multivariate distribution of the demand and of the counted flows, of which a realization is observed. Without loss of generality, a GLS estimation approach will be assumed in the following, corresponding to a normal multivariate distribution assumption, leading to the following expression of the quasi-dynamic o-d estimator (3.16) under the assumption of diagonal dispersion matrices:

$$\{\mathbf{g}^{*1},\ldots,\mathbf{g}^{*\theta},\ldots,\mathbf{g}^{*n_\theta};\mathbf{p}^{*1},\ldots,\mathbf{p}^{*\tau},\ldots,\mathbf{p}_o^{*n_\tau}\}=$$

$$=\underset{\substack{x_o^\theta\geq 0\ \forall o,\forall\theta\in T\\ 0\leq\pi_{d|o}^\tau\leq 1\ \forall\pi_{d|o}^\tau\in\boldsymbol{\pi}_{d|o}^\tau\ \forall\tau\in T\\ \sum_d\pi_{d|o}^\tau=1\ \forall o,\forall\tau\in T}}{arg\,min}\left\{\sum_{\theta=1}^{n_\theta}\sum_{od=1}^{n_{od}}\frac{\left(x_o^\theta\cdot\pi_{d|o}^{\tau(\theta)}-\hat{d}_{od}^\theta\right)^2}{\sigma_{od}^\theta}+\sum_{\theta=1}^{n_\theta}\sum_{l=1}^{n_{lc}}\frac{\left(\sum_{\theta'=\theta_l}^{\theta}\sum_{od=1}^{n_{od}}m_{odl}^{\theta'\theta}x_o^{\theta'}\cdot\pi_{d|o}^{\tau(\theta')}-\hat{y}_l^\theta\right)^2}{\sigma_l^\theta}\right\}$$

(3.17)

wherein $m_{odl}^{\theta'\theta}$ is the generic term of the dynamic assignment map linking time-dependent o-d flows with time-dependent link flows (i.e. it represents the fraction of o-d flow generated at the time slice θ' being on link l at the time slice θ), σ^θ_{od} and σ^θ_l are related to the dispersion matrix of the demand and of the counted flows distribution respectively, θ_l is the farthest time slice whose generated demand contributes to the link flows on θ, and the constraints follow directly from the quasi-dynamic assumption. Although formally present in the QD-GLS estimator (3.17), the last constraint on the sum of the distribution shares is in fact not necessary, leading to remarkable simplification of the solution algorithm. For this aim, it is sufficient to re-scale the matrix of the distribution shares $p_{d|o}^{*\tau,unc}$ estimated without the last constraint and the corresponding generation profiles $g_o^{*\theta,unc}$ in the following way:

$$p_{d|o}^{*\tau}=\frac{p_{d|o}^{*\tau,unc}}{\sum_d p_{d|o}^{*\tau,unc}}\quad\forall\tau\in T,\forall o$$

and

$$g_o^{*\theta}=g_o^{*\theta,unc}\cdot\sum_d p_{d|o}^{*\tau(\theta),unc}\;.$$

The proposed quasi-dynamic estimator (3.17), which will be termed in the following the QD-GLS estimator, may be seen also as a particularization of the simultaneous estimator by Cascetta et al. (1993); however, it should be noted that the estimator (3.17) is non linear, more specifically it is a bilinear form with respect to the unknowns $\mathbf{x}^{\theta}$ and $\boldsymbol{\pi}^{\tau}$.

Importantly, as stated by equation (3.11), the demand flow d^{θ}_{od} and the corresponding quasi-dynamic flow $d^{\theta,qd}_{od}$ do not coincide: their difference will be termed in the following the “intrinsic error” ie^{θ}_{od} of the quasi-dynamic assumption and also, by extension, of the QD-GLS estimator:

$$ie_{od}^{\theta} = d_{od}^{\theta} - d_{od}^{\theta,qd} \tag{3.18}$$

The presence of an intrinsic error has a key practical relevance, since it represents a lower bound for the effectiveness of the QD-GLS estimator: in other words, the QD-GLS estimator will be able at most to provide an o-d estimate differing from the true o-d flows of a quantity equal to the intrinsic error. Intuitively, the less reliable is the quasi-dynamic assumption (3.11) within τ, the larger is the intrinsic error: therefore, checking the acceptability of the assumption (3.11) on real data is fundamental also in the light of the quantification of the intrinsic error.

3.4 The test site for the empirical validation of the QD-GLS estimator

Both the quasi-dynamic o-d flows assumption and the consistent QD-GLS o-d flows estimator introduced in Section 3.3 require a thorough testing on real data: more specifically, the validity of the quasi-dynamic assumption needs to be assessed in order to quantify the magnitude of the intrinsic error, and the performances of the QD-GLS need to be compared with respect to other existing dynamic o-d estimators.

Observing time-varying “true” o-d flows for a significant urban context is still a very tricky issue, in spite of the availability of advanced collection methods (e.g. plate recognition techniques, Bluetooth, GPS). Therefore, following the literature in the field, an experimental field was extracted from the Y-shaped urban/suburban motorway system in the North-East of Italy made by the A23 Palmanova-Tarvisio and A4 Venezia Est-Trieste motorways (top of

Figure 5). The chosen motorway system is characterized by the presence of both commuting and long distance trips, representing therefore a challenging test site for the quasi-dynamic assumption. This closed system encompasses 17 junctions equipped with entrance/exit toll stations, leading to a full test site with 17 origins, 124 links and 272 o-d pairs, the farthest having a travel time of about 90 minutes. Without any loss of generalization, the A4 branch between Palmanova and Trieste was eliminated and replaced by a virtual junction close to the A4-A23 intersection, so as to decouple the network into two independent closed systems (one per carriageway) made by 13 origins, 91 o-d pairs and 49 links each (bottom of

Figure 5). Importantly, such test site does not allow for a proper balance between unknowns and equations in the static case, with a ratio of 91/49=1.86.

Available data consist of entry/exit junctions and times (in hours, minutes and seconds) for each vehicle, covering the whole day of four weeks in different seasonal periods (i.e. April, June, August and October 2008). Vehicles are disaggregated in five toll classes, which can be grouped into two clusters (passenger and freight). The entrance/exit toll system allows to obtain the true o-d flows and also a very reliable estimate of the dynamic assignment map – with respect to whatever aggregation in time slices θ and in sub-periods τ – based on the observed entry/exit times.

On average, the test site encompasses about 60.000 trips/day. For a proper interpretation of the experiments, it is worth analysing the distribution of the 10-minutes o-d flows, calculated for each o-d pair and for each time slice as the average over the three days 15^{th}-17^{th} April (i.e. weekdays from Tuesday to Thursday). For this aim, the overall 144·91 (i.e. $n_{od} \cdot n_{\theta}$) o-d flows can be sorted in decreasing order and partitioned into 10 different clusters, each containing a 10% of the overall demand volume (i.e. the sum of all 144·91 o-d flows). The result is reported in

Figure 6, which depicts the o-d flows in decreasing order (bleu curve) together with the average demand for each cluster (red curve) and the percentage of o-d pairs falling within each cluster (green curve). As it can be observed, more than 90% of o-d pairs are characterized by an average 10-minutes flow close to zero.

Figure 5 - Overall A4/A23 motorway system in North-Eastern Italy (top) and schematic representation of the one-way (eastbound direction) test site (bottom)

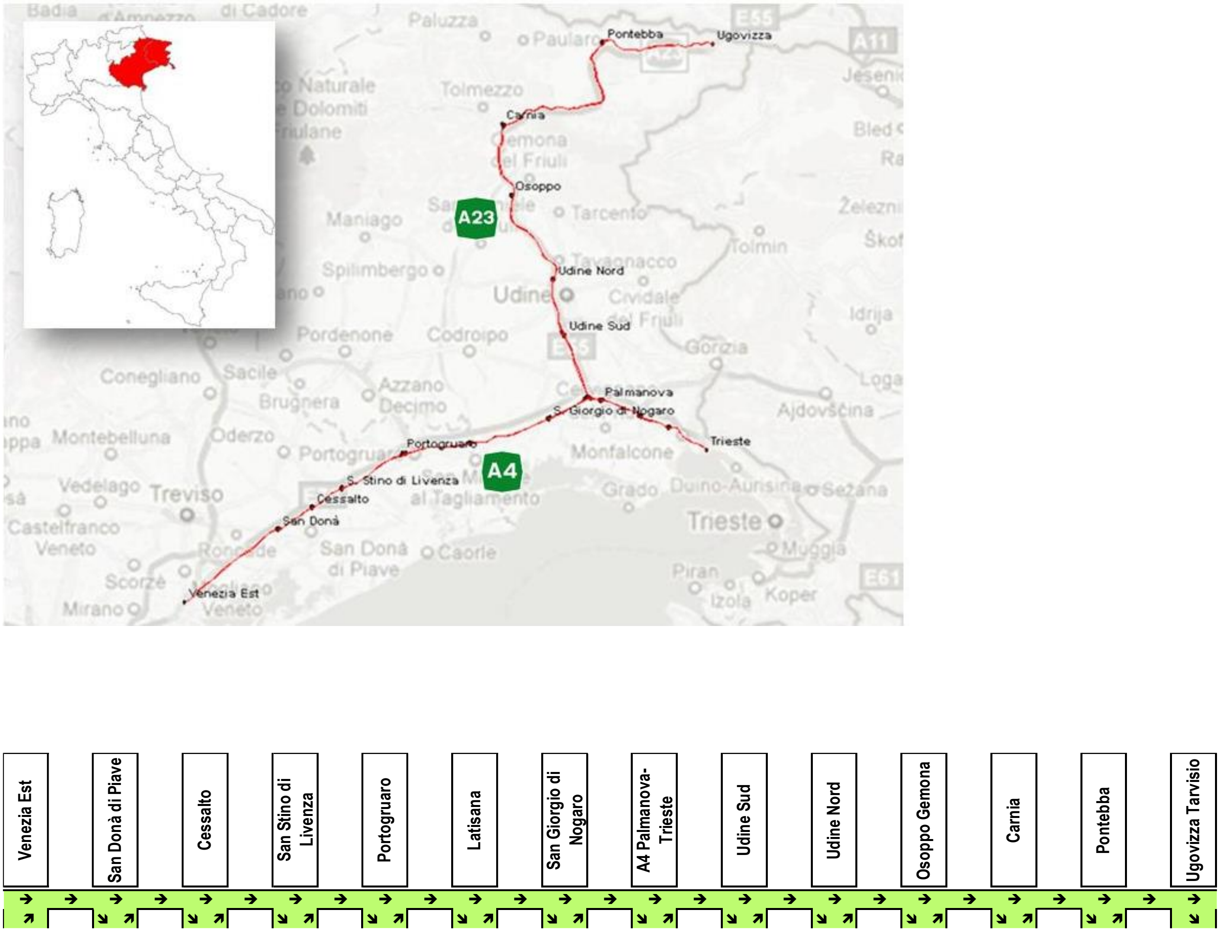
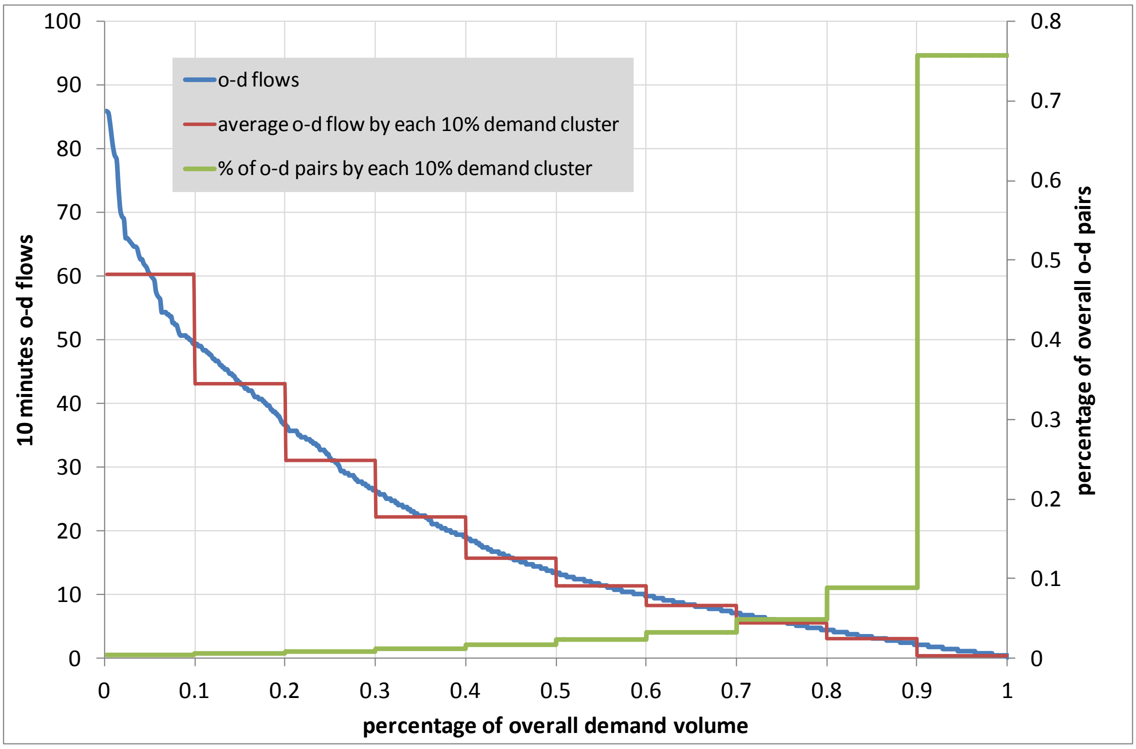


Figure 6 - Analysis of 10-minutes average o-d flows for the 15th-17th April days



Note: 10-minutes o-d flows are sorted for decreasing values.

3.5 Empirical and statistical analysis of the quasi-dynamicity assumption

A first key aspect of the QD-GLS estimator is the assessment of the magnitude of the intrinsic error, thanks to the observability of the o-d flows d_{od}^{θ} in the test site. For this aim, chosen a duration t_{τ} of the sub-period τ and identified the time slices θ included in τ, the quasi-dynamic flows $d_{od}^{\theta,qd}$ can be calculated from d_{od}^{θ} by combining the right-hand side of equation (3.11) with equations (3.12) and (3.13). This allows comparing for any duration t_{τ} the observed o-d flows d_{od}^{θ} and the quasi-dynamic o-d flows $d_{od}^{\theta,qd}$ by means of standard goodness-of-fit measures. In the remaining of the paper, the MSE (mean squared error) and the CV_{RMSE} (i.e. a coefficient of variation calculated using RMSE, the square root of MSE, in place of the standard deviation) will be adopted:

$$MSE = \frac{1}{n_\theta \cdot n_{od}} \sum_\theta \sum_{od} \left(d_{od}^{\theta,true} - d_{od}^{\theta,trueqd} \right)^2 = \frac{1}{n_\theta \cdot n_{od}} \sum_\theta \sum_{od} \left(ie_{od}^{\theta} \right)^2$$

$$CV_{RMSE} = \frac{RMSE}{\bar{d}^{true}} = \frac{\sqrt{MSE}}{\frac{\sum_\theta \sum_{od} d_{od}^{\theta,true}}{n_\theta \cdot n_{od}}}$$

The following Table 5 reports the values of both MSE and CV_{RMSE} for different durations t_τ from 30 minutes to 24 hours, disaggregated by vehicle class. Indicators are averaged across all days of the 14^{th}-18^{th} April week[6]: therefore, in order to quantify also the dispersion across observed days, Table 5 reports also the standard deviation of the MSE across weekdays.

Table 5 - Empirical analysis of the intrinsic error of the quasi-dynamic estimator for different durations of the sub-periods τ of constant distribution percentages: disaggregation by vehicle class

vehicle class	indicator	quasi-dynamic time interval (hours)						
		0.5	**1**	**2**	**3**	**6**	**12**	**24**
all	**MSE**	2.42	3.15	3.80	4.05	4.62	5.15	6.01
	MSE (st dvn)	*0.27*	*0.36*	*0.61*	*0.65*	*0.97*	*1.13*	*1.34*
	CV_{RMSE}	0.363	0.415	0.456	0.470	0.503	0.531	0.573
passenger	**MSE**	1.38	1.77	2.11	2.20	2.51	2.66	2.81
	MSE (st dvn)	*0.16*	*0.25*	*0.35*	*0.38*	*0.51*	*0.58*	*0.71*
	CV_{RMSE}	0.424	0.481	0.525	0.536	0.573	0.589	0.605
freight	**MSE**	0.83	1.05	1.23	1.29	1.42	1.67	2.10
	MSE (st dvn)	*0.10*	*0.11*	*0.17*	*0.16*	*0.20*	*0.26*	*0.41*
	CV_{RMSE}	0.603	0.679	0.734	0.753	0.789	0.858	0.961

As expected, the average intrinsic errors increase for larger durations t_τ, with a CV_{RMSE} for all vehicles moving from 0.36 (t_τ= 0.5 h) to 0.57 (t_τ= 24 h) and a MSE from 2.42 (t_τ= 0.5 h) to 6.01 (t_τ= 24 h), with a very limited variability across days. In absolute terms, the CV_{RMSE} values presented in Table 5 are affected by the presence of a remarkable number of 10-minutes o-d flows close to zero (see

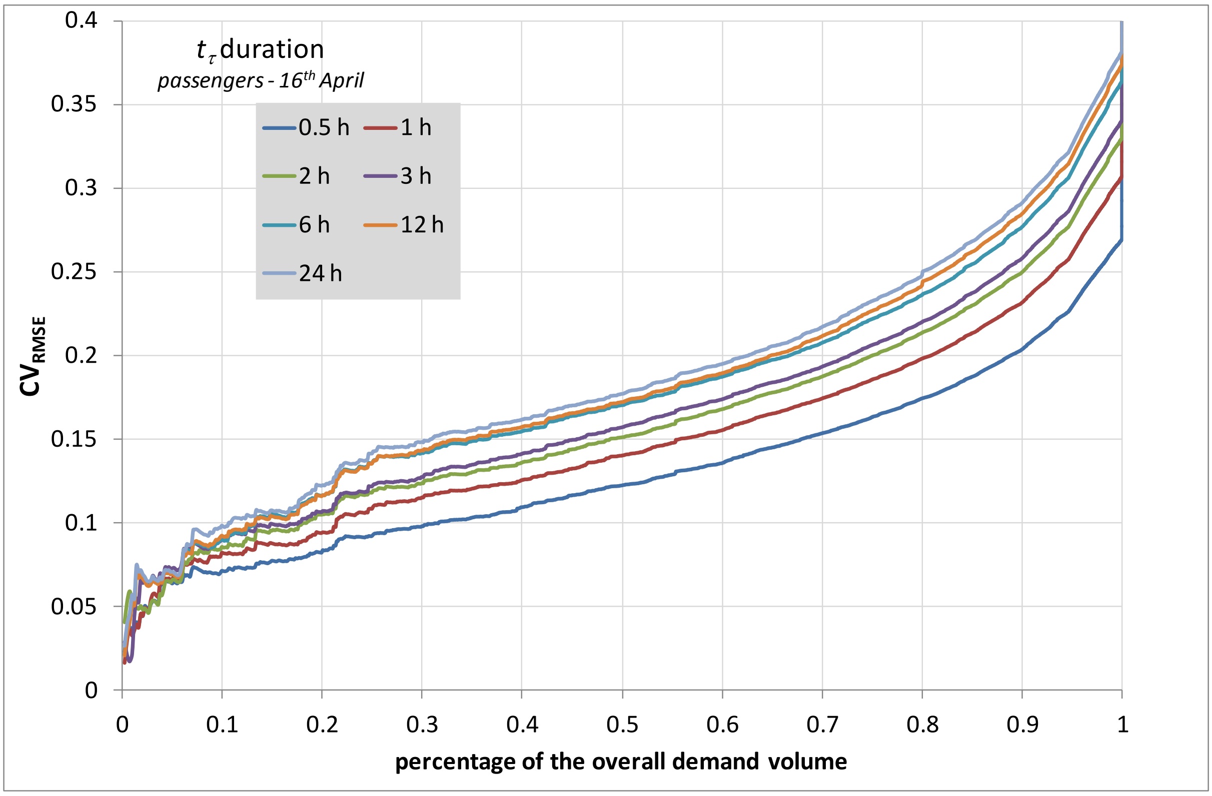
[6] Results for the other days/weeks were completely similar.

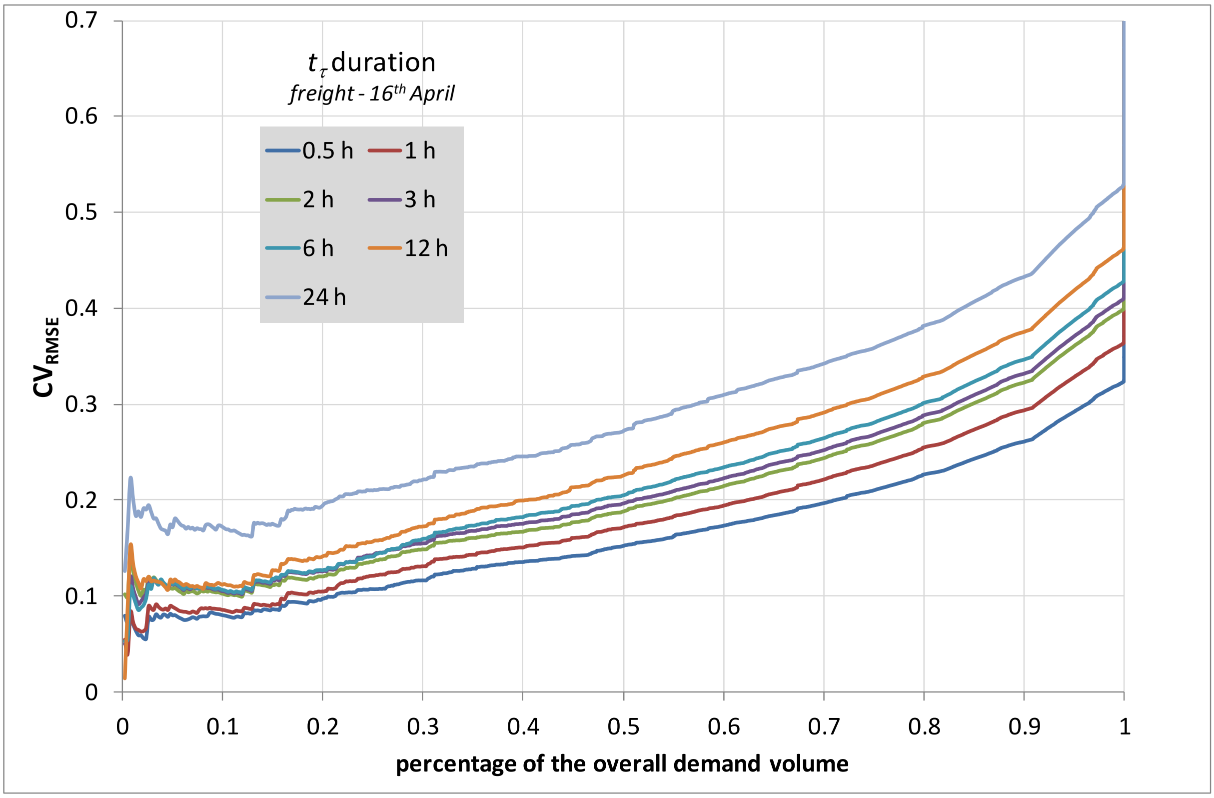
Figure 6). In order to overcome this issue, the $n_{od} \cdot n_{\theta}$ (i.e. 91·144 = 13104) o-d flows can be sorted for decreasing o-d flow values, and the CV_{RMSE} can be calculated just with reference to the first p-th percentage of the overall demand volume (i.e. with reference only to the o-d flows in decreasing order summing up to the *p*-th percentage of the overall demand volume). As a result, the whole distribution of the CV_{RMSE} as a function of the percentage *p* of the overall demand volume can be drawn, as reported in Figure 7 for both passenger and freight vehicles for different t_{τ} durations with reference to the 16th April data.

The CV_{RMSE} values are satisfactory up to a significant percentage of the overall o-d flows (e.g. the CV_{RMSE} for the 80% of the overall passenger demand equals 0.20 for t_{τ}=1h and 0.25 for t_{τ}=24h) and acceptable in practical terms. This provides a first evidence of the robustness of the quasi-dynamic assumption.

Interestingly, appreciable differences may be observed across vehicle classes both in Table 5 and in Figure 7 calculations, with a higher intrinsic error in terms of CV_{RMSE} for freight vehicles with respect to passenger vehicles. This result can be explained with a more stable demand pattern for passenger vehicles, typically representing local and commuting demand, with respect to freight vehicles, for which the share of non-recurrent trips cannot be discarded normally. From a practical standpoint, the presence of different demand distribution patterns for different vehicle classes may be accommodated through a multi-class version of the quasi-dynamic estimator (3.17), wherein each vehicle class is explicitly treated as a separate demand segment with its own quasi-dynamic assumptions. Such generalization does not imply any theoretical and/or operational difference with respect to the simple estimator (3.17), and will be therefore not considered in the remaining of the thesis.

Figure 7 - Distribution of the CV_{RMSE} as a function of the considered percentage of overall demand volume for passenger (top) and freight (bottom) vehicles for different t_τ durations: data for April 16th





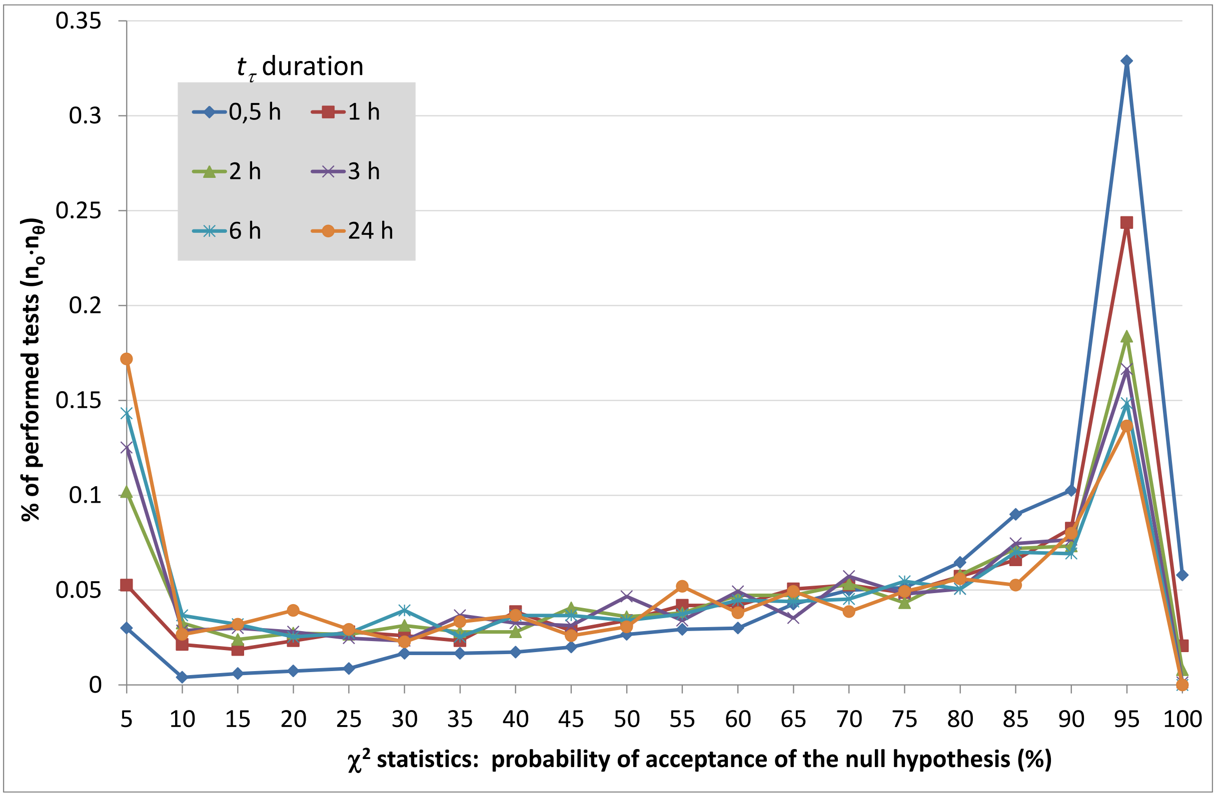
In addition to the empirical investigations presented above, formal statistical tests may be performed in order to check the hypothesis of existence of quasi-dynamic demand patterns within a given sub-period τ. In more detail, for each time slice θ and for each origin o, one is interested in testing whether the observed o-d flows $d^{\theta}_{od}=g^{\theta}_{o}\cdot p^{\theta}_{d|o}$ coming from the observed distribution shares $p^{\theta}_{d|o}$ and the corresponding expected quasi-dynamic o-d flows $d^{\theta,qd}_{od} = g^{\theta}_{o}\cdot\, p^{\tau(\theta)}_{d|o}$ coming from the quasi-dynamic distribution shares $p^{\tau(\theta)}_{d|o}$, are statistically different. For this aim, two different statistics can be adopted.

The first is a standard Pearson's chi-squared test (Kendall and Stuart, 1979), which tests formally the null hypothesis H_0 that the observed frequencies d^{θ}_{od} and the expected frequencies $d^{\theta,qd}_{od}$ come from the same distribution, based on the statistic:

$$\chi^2_{o\theta} = \sum_d \frac{\left(d^{\theta}_{od} - d^{\theta,qd}_{od}\right)^2}{d^{\theta,qd}_{od}} = \sum_d \frac{\left(g^{\theta}_{o}\, p^{\theta}_{d|o} - g^{\theta}_{o}\, p^{\tau(\theta)}_{d|o}\right)^2}{g^{\theta}_{o}\, p^{\tau(\theta)}_{d|o}} = \sum_d g^{\theta}_{o}\, \frac{\left(p^{\theta}_{d|o} - p^{\tau(\theta)}_{d|o}\right)^2}{p^{\tau(\theta)}_{d|o}} \tag{3.19}$$

with a number of degrees of freedom equal to $n_{d|o}-1$, being $n_{d|o}$ the number of destinations reachable from the origin o. Such test may be repeated $n_{\theta}\cdot n_{o}$ times for each day: for the sake of brevity, the following Figure 8 reports the distribution of the probability of acceptance of the null hypothesis of all tests performed for April 14th (being results for all other days remarkably similar) with reference to different t_{τ} durations (from 0.5 to 24 hours). Results are generally satisfactory, with a probability of acceptance larger than 80% exhibited by almost 60% of tests in the most favourable situation (t_{τ} = 0.5 h) and by about 40% of tests in the worst case (t_{τ} = 24 h). Interestingly, the test may be used also in order to identify the optimal subdivision of the whole day into a proper number n_{τ} of periods τ for which the quasi-dynamic assumption (3.11) holds.

Figure 8 - Distribution of the probability of acceptance of the null hypothesis of the χ2 test for April 14th, under different tτ durations

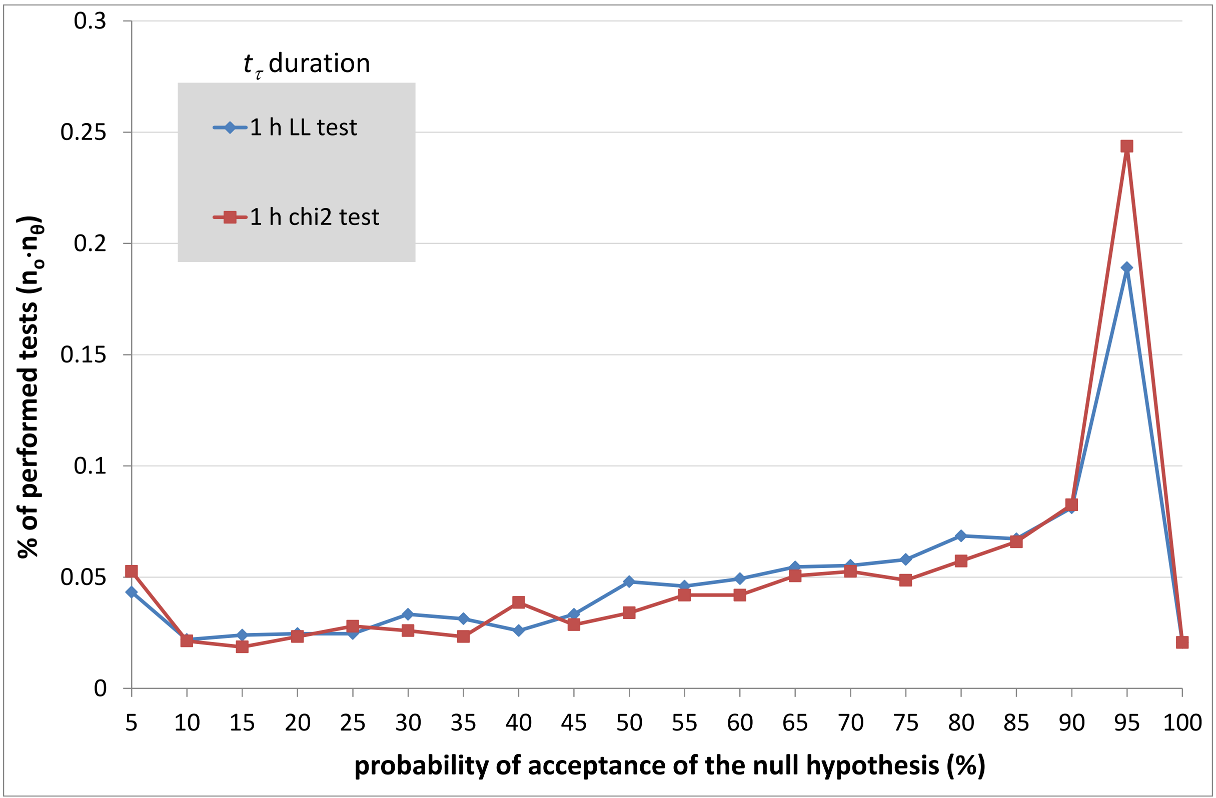


The second is a likelihood ratio (LR) test (Kendall and Stuart, 1979), testing the null hypothesis H_0 that the observed distribution shares $p^{\theta}_{d|o}$ and the corresponding quasi-dynamic distribution shares $p^{\tau(\theta)}_{d|o}$ come from the same distribution. In other words, the LR statistic may be constructed by considering that the quasi-dynamic distribution shares represent the null model and the observed distribution shares the alternative model, yielding:

$$\rho_o^{\theta} = -2\ln\frac{L_{qd}}{L_{true}} = -2\ln\frac{\prod_d p[d\,|\,o\tau(\theta)]^{d^{\theta}_{od}}}{\prod_d p[d\,|\,o\theta]^{d^{\theta}_{od}}} = -2\sum_d d^{\theta}_{od}\ln\frac{p[d\,|\,o\tau(\theta)]}{p[d\,|\,o\theta]} \tag{3.20}$$

with ρ_o^{θ} asymptotically distributed as a chi-squared variable with a number of degrees of freedom still equal to ($n_{d|o}$–1). The LR test (3.20) provides very similar results to the chi-squared test (3.19): by way of an example, the following Figure 9 compares the distribution of the probability of acceptance of the null hypothesis using the two different statistics, for t_τ = 1 hour.

Figure 9 - Comparison of the probability of acceptance of the null hypothesis of the χ2 test and of the LL test for April 14th (tτ =1 hour)



3.6 Empirical analysis of the QD-GLS estimator

The kernel of the validation of the QD-GLS estimator is the comparison of its performances with other o-d estimators available in the literature on real data. For this aim, Section 3.6.1 illustrates the theoretical characteristics of the chosen benchmark estimators, Section 3.6.2 describes the experimental setup, and Section 3.6.3 provides the results of all the experiments.

3.6.1 Benchmark estimators

A first natural benchmark estimator for the QD-GLS estimator is the simultaneous dynamic estimator proposed by Cascetta et al. (1993), having the following functional form:

$$\{ \mathbf{d}^{*1},\ldots,\mathbf{d}^{*\theta},\ldots,\mathbf{d}^{*n_\theta} \} = \underset{\mathbf{x}^\theta \geq 0 \ \forall \theta \in T}{arg\,min} \left\{ \sum_{\theta=1}^{n_\theta} \sum_{od=1}^{n_{od}} \frac{\left(x_{od}^{\theta} - \hat{d}_{od}^{\,\theta} \right)^2}{\sigma_{od}^{\theta}} + \sum_{\theta=1}^{n_\theta} \sum_{l=1}^{n_l} \frac{\left(\sum_{\theta'=\theta_l}^{\theta} \sum_{od=1}^{n_{od}} m_{odl}^{\theta'\theta} x_{od}^{\theta'} - \hat{y}_l^{\theta} \right)^2}{\sigma_l^{\theta}} \right\}$$

(3.21)

wherein $\mathbf{x}^\theta = \{ x_1^\theta,\ldots,x_{n_{od}}^\theta \} \ \ \forall \theta \in T$ represents the unknown demand vectors, $\mathbf{d}^{*\mathbf{0}} = \{ d_1^{*\theta},\ldots,d_{n_{od}}^{*\theta} \} \forall \theta \in T$ the corresponding optimal solutions and the other symbols were already introduced.

A further benchmark estimator is the Kalman filter (Ashok, 1996), typically applied in on-line estimation/prediction but suitably adaptable to off-line contexts, whose specification encompasses a measurement equation (3.2) and a transition equation (3.1).

Notably, the estimation of the coefficients φ^{θ}_{od} and of the dispersion matrix of the errors α^{θ}_{od} in the transition equation (3.1) requires knowledge of seed o-d flows for two consecutive days. In addition, in the off-line application of this estimator a backward smoothing of the estimates obtained recursively from the forward application of (3.1)-(3.2) over T may be applied, in order to improve the quality of the final estimate.

In the light of the quasi-dynamic assumption, the autoregressive process (3.1) may be interpreted as an alternative hypothesis on the within-day evolution of o-d flows able to bound the number of unknowns and, therefore, to help achieving an effective unknowns/equation balance. However, from the other side, this implies also the Kalman filter estimator to be characterized by an intrinsic bias (i.e. an ideal lower bound to its estimation/updating capabilities) due to the structural inability of the autoregressive process – or of whatever transition equation (3.1) – to reproduce exactly the true o-d flows. In addition, since the true o-d flows are generally not available for the estimation of the autoregressive process, the performances of the Kalman filter depend also on the quality of the historical estimate of o-d flows used for the estimation of the autoregressive process, practically leading to an intrinsic bias much higher than the aforementioned ideal lower bound. This is an important difference with respect to the QD-GLS estimator, whose evolution rule – differently of that assumed by the Kalman filter – does not need to be estimated, apart from the duration t_τ of the quasi-dynamic assumption. As a consequence, while the intrinsic error of the QD-GLS estimator depends only on the inherent approximation of the quasi-dynamic assumption, the intrinsic bias of the Kalman filter estimator depends on both the inherent approximation of the autoregressive process and on the quality of the historical o-d flows used for estimating the coefficients of the autoregressive process[7]. Therefore, since such historical o-d flows are generally provided by a simultaneous estimator, it is worth exploring whether using the QD-GLS estimates as historical o-d flows may help improving the performances of the Kalman filter (i.e. reaching the above mentioned ideal lower bound).

Finally, it is worth noting that the dynamic o-d matrices $\mathbf{d}^{*\theta}$ – provided by any of the aforementioned dynamic estimators (i.e. simultaneous, QD-GLS, Kalman filter) for each time slice $\theta \in T$ – can be aggregated (i.e. summed) over a time interval $T_S \in T$ in order to estimate o-d flows over T_S. Such “static” (i.e. resulting from aggregations of dynamic o-d matrices in the sense above defined) performances of the dynamic o-d estimators may be compared straightforwardly with the outcomes of the static GLS estimator proposed by Cascetta (1984) and applied to T_S, i.e.:

[7] In that respect, it is important to underline that performing the Kalman filter across a higher number of days might help improving the quality of the autoregressive process through repeated estimations.

$$\mathbf{d}^{*T_s} = \underset{\mathbf{x}^{T_s} \geq 0}{arg\,min} \left\{ \sum_{od=1}^{n_{od}} \frac{\left(x_{od}^{T_s} - \hat{d}_{od}^{T_s} \right)^2}{\sigma_{od}^{T_s}} + \sum_{l=1}^{n_l} \frac{\left(\sum_{od=1}^{n_{od}} m_{odl}^{T_s} x_{od}^{T_s} - \hat{y}_l^{T_s} \right)^2}{\sigma_l^{T_s}} \right\} \tag{3.22}$$

3.6.2 Description of the experiments

The chosen test site allows:

1. observation of o-d flows;
2. calculation of a realistic assignment map inferred from entry/exit times;
3. calculation of link flows, through the dynamic network loading of the observed o-d flows based on the assignment map, which will represent in the following the "observed" link flows of the test site.

The experiments consist firstly of a perturbation of the observed o-d flows (point 1 above), leading to perturbed o-d flows, which are used as seeds of the estimation/updating procedures in order to mimic the unreliability of real historical o-d estimates.

Estimation/updating is then performed starting from these seed o-d flows, on the basis of the assignment map (point 2 above) and on a subset of 15 link counts chosen – by means of the max flow method proposed by Yang and Zhou (1998) – amongst the observed link flows (point 3 above). In other terms, the estimation/updating procedures are performed in the most favourable conditions of error-free link counts and error-free assignment map[8], in line with Marzano et al. (2009), who showed that the tested estimators may be affected by significant issues and drawbacks also in these ideal conditions. Note that this updating implies the static GLS estimator to work with

[8] Nevertheless, measurement and assignment errors may be introduced directly in the proposed experimental framework, through proper random perturbations of the true link flows (so as to mimic measurement errors) and/or of the dynamic assignment map (so as to mimic assignment errors).

an unknowns/equations ratio equal to 91/15=6.07 for each hour. The same ratio applies also to the simultaneous estimator (3.21), characterized by 91·144 unknowns and 15·144 equations.

Once performed the estimation/updating, the quality of the tested estimator may be measured directly by comparing the updated o-d flows with the observed o-d flows through standard goodness-of-fit measures, such those introduced at the beginning of Section 3.5.

In addition, since only a subset of the observed link flows is used for the estimation/updating procedures, the remaining observed link flows may be used for a further hold-out validation, i.e. comparing them with link flows obtained by assigning the updated o-d flows.

The described experimental framework applies directly to the dynamic estimators (3.17), (3.21) and (3.1)-(3.2), but it may be adapted straightforwardly also to the static GLS estimator (3.23) and to the static performances of the dynamic estimators (3.17), (3.21) and (3.1)-(3.2) over a time interval T_s. For this aim, from one hand, observed o-d flows, perturbed seed o-d flows and true link flows over T_s should be computed (simply by summing observed o-d flows, perturbed seed o-d flows and true link flows for each time slice $\theta \in T_s$) in order to perform the static GLS correction (3.22) and to compute its performance indicators in terms of both o-d and link flows. From the other hand, the updated o-d flows and the corresponding assigned link flows obtained through the dynamic estimators (3.17), (3.21) and (3.1)-(3.2) should be summed so as to calculate the same performance indicators over T_s.

In terms of setup of the experiments, without loss of generalization, only o-d flows related to passenger vehicles are considered. Furthermore, a 10-minutes time slices disaggregation is taken into account, i.e. t_θ=10 minutes and n_θ=144 within the overall daily time horizon. The observed o-d flows d^θ_{od} are perturbed by perturbing the observed generation profiles g^θ_o $\forall \theta, o$ through independent draws from a normal distribution (truncated to non-negative values) with mean g^θ_o and standard deviation $0.3g^\theta_o$ (i.e. using a coefficient of variation of 0.3), and assuming uniform distribution shares $p^\theta_{d|o}$ across all destinations, i.e. $p^\theta_{d|o}=1/n_{d|o}\ \forall \theta, o$.

Thanks to this “controlled” perturbation, the dispersion matrix of the seed o-d flows – required by all estimators described in Section 3.6.1 – is also directly calculable: notably, the average coefficient of variation of the seed o-d flows identified by such perturbations is very high (about 2).

Finally, it is important to mention specific setups for each of the tested dynamic estimators:

- for the QD-GLS estimator, a duration t_{τ}=24 hours is assumed, i.e. the distribution shares are kept constant for the whole day; this implies the QD-GLS estimator to work with $n_{\theta}{\cdot}n_o+n_{od}–n_o$=144·13+91–13=1950 unknowns and $n_{\theta}{\cdot}n_l$=144·15=2160 equations, i.e. with an unknowns/equations ratio equal to 0.90;
- for the static GLS estimation, a duration T_s=1 hour is assumed, i.e. 24 daily static estimate/updates are performed;
- for the Kalman filter, three different experiments are carried out depending on the type of seed o-d flows. A first ideal experiment, named “true seeds”, consisted of using the true o-d flows both for the estimation of the autoregressive process and as seeds for the filter. Obviously, this experiment does not correspond to any possible real application, but it allows quantifying the intrinsic bias in the Kalman filter coming from the autoregressive process in the transition equation (3.1). In addition, two more standard experiments are carried out, based on considering as historical estimates respectively the simultaneous estimates (“simultaneous seeds” scenario) and the quasi-dynamic estimates (“quasi-dynamic” scenario). In all scenarios, the estimated autoregressive process leading to the best statistical result is a 5-order lagged process and, differently from what observed by Ashok (1996), the backward smoothing does not provide any appreciable improvement to the forward on-line estimates.

3.6.3 Experimental results

A first set of experiments refers to the application of the QD-GLS estimator (3.17), of the simultaneous estimator (3.21) and of the Kalman filter (3.1)-(3.2) to a standard within-day dynamic context. Results are reported in the following Table 6, which also includes the initial error embedded in the perturbed seed o-d flows, together with the intrinsic error of the QD-GLS estimator and the “true seed” scenario of the Kalman filter, representing a sort of optimal lower bound for the QD-GLS and the Kalman filter estimators respectively in the light of the above discussion.

In addition to the aggregated values reported in Table 6, a detailed assessment of the distribution of the cv_{RMSE} for increasing percentages of the overall demand volume (with o-d flows sorted in decreasing order as in Figure 7) may be performed, similarly to the analysis provided in the Section 3.5 with reference to the intrinsic error. In more detail, the following Figure 10 reports the cv_{RMSE} for each estimator as a function of the percentages of the overall o-d volume. Reported data refers to June 5th, which provided the worst results in terms of MSE for the QD-GLS estimation in middle working weekdays; however, the other observed days exhibit extremely similar figures.

Table 6 - Results of o-d updating with the QD-GLS, the simultaneous and the Kalman filter estimators: tests based on 15 count sections for the weeks April 14-18 (top) and June 3-6 (bottom)

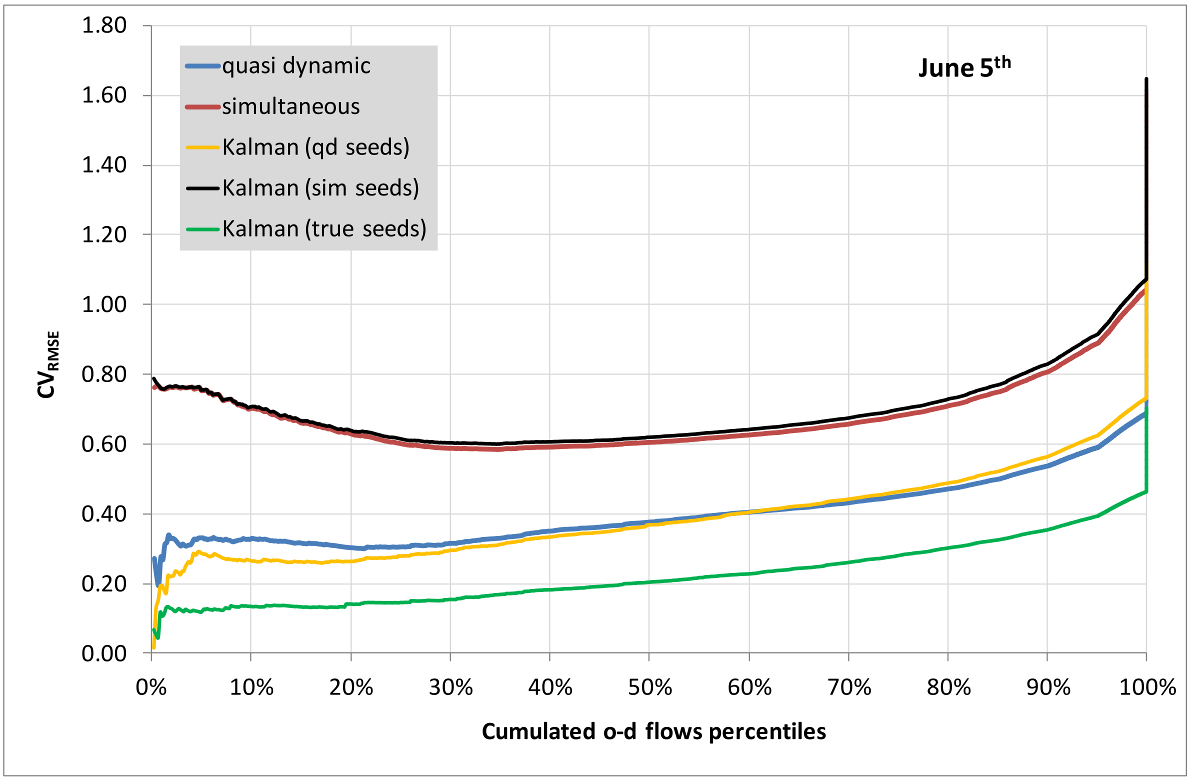
day		MSE											
		Seed matrix	Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
					intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
MO	April 14th	32.9	20.9	-37%	2.41	6.40	-81%	estimation of the autoregressive process					
TU	April 15th	35.1	21.0	-40%	2.38	7.50	-79%						
WE	April 16th	36.7	21.6	-41%	2.59	8.16	-78%	3.23	-91%	20.56	-44%	7.56	-79%
TH	April 17th	38.7	21.9	-44%	2.74	8.57	-78%	3.45	-91%	21.90	-43%	8.72	-77%
FR	April 18th	48.2	23.3	-52%	3.91	14.77	-69%	3.09	-94%	23.65	-51%	10.51	-78%

day		CV_{RMSE}											
		Seed matrix	Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
					intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
MO	April 14th	2.20	1.75	-20%	0.59	0.97	-56%	estimation of the autoregressive process					
TU	April 15th	2.27	1.76	-23%	0.59	1.05	-54%						
WE	April 16th	2.24	1.72	-23%	0.60	1.06	-53%	0.66	-70%	1.68	-25%	1.02	-55%
TH	April 17th	2.23	1.67	-25%	0.59	1.05	-53%	0.66	-70%	1.67	-25%	1.06	-53%
FR	April 18th	2.22	1.55	-30%	0.63	1.23	-45%	0.56	-75%	1.56	-30%	1.04	-53%

day		MSE											
		Seed matrix	Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
					intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
TU	June 3rd	33.93	17.78	-48%	2.69	8.03	-76%	estimation of the autoregressive process					
WE	June 4th	37.48	21.81	-42%	3.01	8.56	-77%						
TH	June 5th	38.27	22.32	-42%	3.50	9.60	-75%	4.30	-89%	23.55	-38%	10.95	-71%
FR	June 6th	50.00	23.64	-53%	4.37	16.80	-66%	4.55	-91%	23.42	-53%	11.89	-76%

day		CV_{RMSE}											
		Seed matrix	Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
					intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
TU	June 3rd	2.12	1.53	-28%	0.60	1.03	-51%	estimation of the autoregressive process					
WE	June 4th	2.15	1.64	-24%	0.61	1.03	-52%						
TH	June 5th	2.10	1.60	-24%	0.64	1.05	-50%	0.70	-66%	1.65	-22%	1.12	-47%
FR	June 6th	2.09	1.44	-31%	0.62	1.21	-42%	0.63	-70%	1.43	-32%	1.02	-51%

Figure 10 - Distribution of CV_{RMSE} for the QD-GLS, the simultaneous and the Kalman filter estimators as a function of the percentages of the overall o-d flow volume with o-d flows sorted for decreasing values (5th June data)



The main outcome of this experiment is the effectiveness of the quasi-dynamic o-d updating, which is always able to achieve a significant reduction of both MSE (-78% on average with respect to the initial perturbed seed o-d flows) and CV_{RMSE} (-53%), even in the most unfavourable assumption of constant distribution shares across the entire day.

The performances of the Kalman filter are entirely dependent on the quality of the seed o-d flows, as it can be recognized by comparing the "simultaneous" and "quasi-dynamic" indicators of the estimators (3.21) and (3.17) respectively with the "simultaneous seeds" and "quasi-dynamic seeds" indicators of the Kalman filter in Table 6. Therefore, the QD-GLS estimator is also very useful in supporting on-line applications, since using quasi-dynamic estimates as historical seeds allows the Kalman filter to provide good o-d flow estimates, slightly better with respect to those obtainable by applying directly the quasi-dynamic estimator.

On the contrary, the simultaneous estimator is substantially not effective, with estimation results remarkably far from the performances of the QD-GLS estimator and of the Kalman filter estimator feed with quasi-dynamic estimates as historical seeds. The good results of the QD-GLS are even more positive when looking at the distribution chart reported in Figure 10, wherein the CV_{RMSE} is about 0.40 up to the 60% of the overall o-d flow volume, whilst the simultaneous estimators exhibit much higher errors (with $CV_{RMSE} \geq 0.60$). Again, the performances of each Kalman filter scenarios are very similar (slightly better) with respect to those of the corresponding estimators used to obtain the seed o-d flows.

A subsequent set of experiments aims at checking the capability of the QD-GLS estimator (3.17), of the simultaneous estimator (3.21) and of the Kalman filter (3.1)-(3.2) to provide reliable o-d flows estimates over a larger period $T_s \in T$ by aggregating the corresponding o-d flow estimates for each time slice θ over T_s: in that respect, a direct comparison may be performed with the classical static estimates obtained through the uncongested GLS estimator (3.22). Results are reported in the following Table 7 which, as in the previous experiment, reports also the initial error embedded in the perturbed seed o-d flows, the intrinsic error of the QD-GLS estimator and the "true seeds" scenario within the Kalman filter experiments. Furthermore, in analogy with Figure 10, Figure 11 draws the distribution of the cv_{RMSE} indicator for the QD-GLS, the simultaneous and the Kalman filter estimators as a function of the percentages of the overall o-d flow volume, with o-d flows sorted in decreasing order, with reference to the average of the 16^{th}-17^{th} April period (other days/periods provided in general very similar results).

Table 7 - Results of hourly o-d estimates obtained with the static GLS estimator by Cascetta (1984) and by aggregating simultaneous, QD-GLS and Kalman filter estimates: tests based on 15 count sections for the weeks April 14-18 (top) and June 3-6 (bottom)

day		MSE													
		Seed matrix	Static updating		Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
							intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction	absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
MO	April 14th	1047	734	-30%	681	-35%	27	155	-85%	estimation of the autoregressive process					
TU	April 15th	1101	795	-28%	677	-39%	28	170	-85%						
WE	April 16th	1151	824	-28%	693	-40%	32	187	-84%	19	-98%	648	-44%	164	-86%
TH	April 17th	1221	839	-31%	702	-43%	33	192	-84%	17	-99%	697	-43%	183	-85%
FR	April 18th	1539	959	-38%	679	-56%	58	394	-74%	39	-97%	743	-52%	234	-85%

day		CV_{RMSE}													
		Seed matrix	Static updating		Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
							intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction	absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
MO	April 14th	2.06	1.73	-16%	1.66	-20%	0.33	0.79	-62%	estimation of the autoregressive process					
TU	April 15th	2.12	1.80	-15%	1.66	-22%	0.34	0.83	-61%						
WE	April 16th	2.09	1.77	-15%	1.62	-23%	0.35	0.84	-60%	0.27	-87%	1.57	-25%	0.79	-62%
TH	April 17th	2.08	1.73	-17%	1.58	-24%	0.34	0.83	-60%	0.25	-88%	1.57	-24%	0.81	-61%
FR	April 18th	2.09	1.65	-21%	1.39	-34%	0.41	1.06	-49%	0.33	-84%	1.46	-30%	0.82	-61%

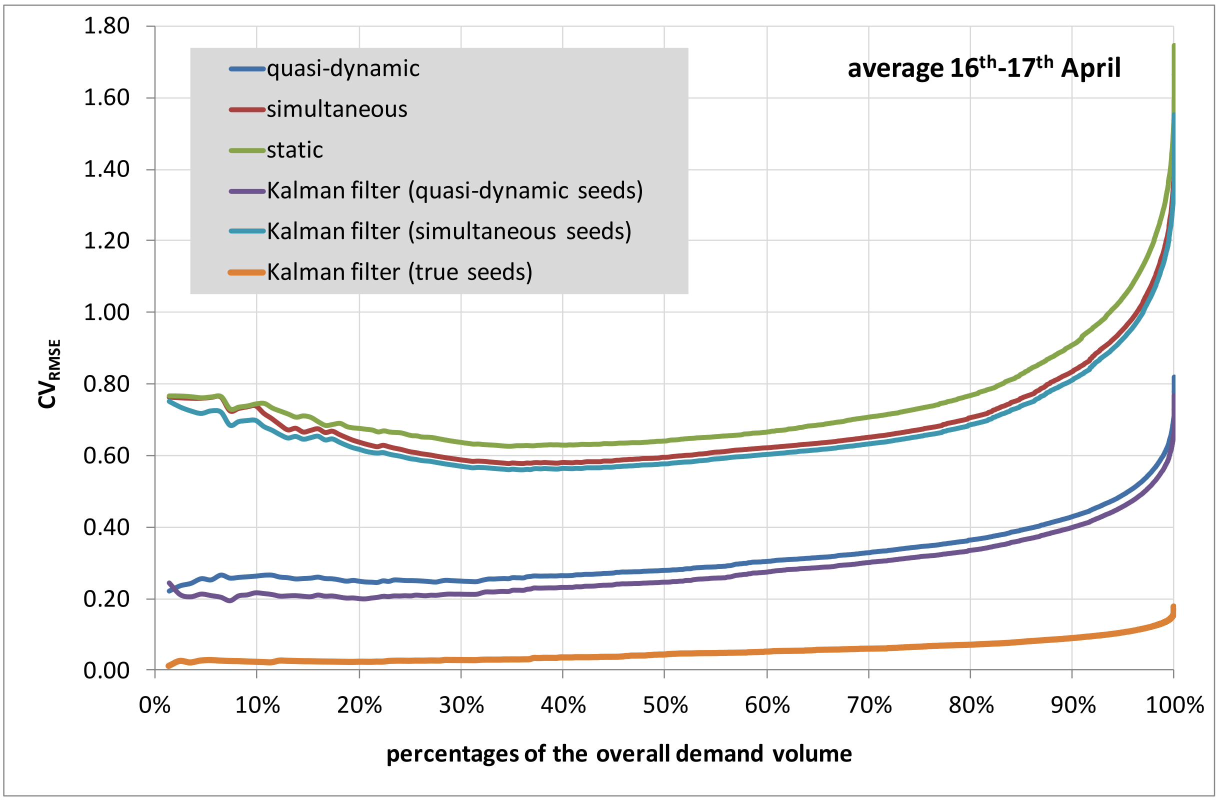
day		MSE													
		Seed matrix	Static updating		Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
							intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction	absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
TU	June 3rd	1032	719	-30%	537	-48%	29	161	-84%	estimation of the autoregressive process					
WE	June 4th	1177	857	-27%	696	-41%	39	195	-83%						
TH	June 5th	1196	867	-27%	708	-41%	55	225	-81%	49	-96%	745	-38%	233	-80%
FR	June 6th	1606	974	-39%	747	-53%	71	467	-71%	52	-97%	732	-54%	265	-83%

day		CV_{RMSE}													
		Seed matrix	Static updating		Simultaneous updating		Quasi-dynamic (τ=T=24h)			Kalman filter					
							intrinsic error	updating		true seeds		simultaneous seeds		quasi-dynamic seeds	
			absolute	% reduction	absolute	% reduction		absolute	% reduction	absolute	% reduction	absolute	% reduction	absolute	% reduction
TU	June 3rd	1.95	1.63	-17%	1.41	-28%	0.33	0.77	-61%	estimation of the autoregressive process					
WE	June 4th	2.01	1.72	-15%	1.55	-23%	0.37	0.82	-59%						
TH	June 5th	1.96	1.67	-15%	1.51	-23%	0.42	0.85	-57%	0.40	-80%	1.54	-21%	0.86	-56%
FR	June 6th	1.97	1.54	-22%	1.34	-32%	0.42	1.06	-46%	0.36	-82%	1.33	-32%	0.80	-59%

Note: results expressed in terms of hourly aggregation of the corresponding 10-minutes corrected matrices for the QD-GLS, the simultaneous and the Kalman filter estimators. The intrinsic error estimate reported in the table differs from that in the last column of Table 5 (t_τ=24h) because here is calculated on a 1-hour basis rather than on a 10-minute basis.

Results are obviously analogous to those obtained in the dynamic experiments, with good performances of the QD-GLS estimator, which outperforms both the static and the dynamic (i.e. simultaneous) GLS estimators. Again, the overall quality of the Kalman filter is shown to depend almost entirely on the quality of the seed o-d flows. This is even clearer in the chart of Figure 11, which evidences the QD-GLS estimates (and the related Kalman filter estimates) to provide low absolute values up to significant percentages of the overall o-d flow volume.

Figure 11 - Distribution of CV_{RMSE} for static, QD-GLS, simultaneous and Kalman filter estimators as a function of the percentages of the overall demand volume with o-d flows sorted in decreasing order (16h-17th April average)



A final set of experiments helps analyzing how the tested o-d estimators are capable to reproduce the observed link flows. In that respect, Table 8 reports the distance between the observed 10-minutes link flows and the link flows obtained assigning the updated o-d flows. In more detail, the MSE and the CV_{RMSE} are calculated on all the 49 links of the network, on the 15 counted links and on the remaining 34 unmonitored links (hold-out sample). The intrinsic bound for the quasi-dynamic estimator is also reported, intended as the distance between the observed link flows and the flows obtained assigning the quasi-dynamic o-d flows. The main outcome of Table 8 is that the QD-GLS estimator, even if intrinsically not able to reproduce exactly the counted links, is characterized by a very high robustness on the hold-out sample, outperforming for those links the simultaneous estimator and allowing the Kalman filter to obtain very effective results. Notably, referring to the 15 links selected for the estimation/updating, the intrinsic error embedded in the quasi-dynamic assumption is always lower than the updating performances of the QD-GLS estimator.

Table 8 - Distances between observed and assigned link flows for different updated o-d flows: tests based on 15 count sections for the weeks April 14-18 (top) and June 3-6 (bottom)

day		MSE																	
		all links						counted links						hold-out sample					
		simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman		
			intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds
MO	April 14th	369	24	80	-	-	-	10	49	106	-	-	-	527	13	113	-	-	-
TU	April 15th	423	23	130	-	-	-	9	41	96	-	-	-	606	15	145	-	-	-
WE	April 16th	422	24	137	16	349	92	12	44	118	3	12	12	603	15	145	22	498	127
TH	April 17th	407	27	137	15	406	122	9	49	122	4	14	100	582	18	143	20	579	132
FR	April 18th	70	46	264	31	398	129	84	85	469	20	32	136	603	28	174	35	559	126

day		CV_{RMSE}																	
		all links						counted links						hold-out sample					
		simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman		
			intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds
MO	April 14th	0.50	0.13	0.23	-	-	-	0.04	0.08	0.12	-	-	-	1.31	0.21	0.61	-	-	-
TU	April 15th	0.55	0.13	0.31	-	-	-	0.04	0.08	0.12	-	-	-	1.43	0.23	0.70	-	-	-
WE	April 16th	0.53	0.13	0.30	0.10	0.48	0.25	0.04	0.08	0.13	0.02	0.04	0.04	1.37	0.22	0.67	0.26	1.24	0.63
TH	April 17th	0.49	0.13	0.29	0.10	0.49	0.27	0.03	0.08	0.12	0.02	0.04	0.11	1.29	0.22	0.64	0.24	1.28	0.61
FR	April 18th	0.17	0.14	0.34	0.11	0.41	0.24	0.08	0.09	0.20	0.04	0.05	0.11	1.14	0.25	0.61	0.28	1.10	0.52

day		MSE																	
		all links						counted links						hold-out sample					
		simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman		
			intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds
TU	June 3rd	345	31	124	-	-	-	6	44	85	-	-	-	494	25	142	-	-	-
WE	June 4th	384	34	122	-	-	-	7	51	96	-	-	-	551	27	133	-	-	-
TH	June 5th	355	64	162	38	355	131	6	74	167	5	10	91	513	59	168	52	508	148
FR	June 6th	73	73	450	39	360	139	85	85	1022	56	62	160	490	68	197	32	492	130

day		CV_{RMSE}																	
		all links						counted links						hold-out sample					
		simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman			simultaneous	QD-GLS		Kalman		
			intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds		intrinsic	updating	true seeds	simultaneous seeds	quasi-dynamic seeds
TU	June 3rd	0.45	0.14	0.27	-	-	-	0.03	0.07	0.10	-	-	-	1.18	0.27	0.63	-	-	-
WE	June 4th	0.47	0.14	0.26	-	-	-	0.03	0.08	0.11	-	-	-	1.17	0.26	0.57	-	-	-
TH	June 5th	0.42	0.18	0.29	0.14	0.42	0.26	0.03	0.09	0.14	0.02	0.03	0.10	1.02	0.35	0.58	0.32	1.02	0.55
FR	June 6th	0.16	0.16	0.40	0.12	0.36	0.22	0.08	0.08	0.28	0.06	0.07	0.11	0.87	0.33	0.55	0.22	0.88	0.45

3.7 Research perspectives

From a theoretical point of view, the study of the properties of the QD-GLS estimator might be an interesting future research step. For this aim, in this section, the quadratic relations between variables and measures introduced by the QD-GLS estimator are described.

The dynamic relationship between d^{θ}_{od}, i.e. the flow starting from the origin *o* toward the destination *d* at the time slice θ, and the flow on the link *l* at the time slice θ_1, $y_l^{\theta_1}$, can be expressed through the matrix $\mathbf{A}_{r,\theta,l,\theta_1}$, whose generic element represents the percentage of users of the o-d pair *r*, starting from the origin at θ, and crossing the link *l* at θ_1 (zero in many cases, in particular for $\theta > \theta_1$):

$$y_l^{\theta_1} = \sum_r \sum_{\theta<\theta_1} a_{r,\theta,l,\theta_1} d_r^{\theta} \tag{3.23}$$

Considering the quasi-dynamic hypothesis, Eq. (3.23) becomes:

$$y_l^{\theta_1} = \sum_r \sum_{\theta<\theta_1} a_{r,\theta,l,\theta_1} g_o^{\theta} p_r^{\tau(\theta)}$$

Let *k* indicate the couple made up of *l* and θ_1, there are then $n_\theta \cdot n_{lc}$ equations, each of which can be expressed in vectorial form as:

$$\mathbf{x}\mathbf{A}^k\mathbf{x}^{\mathrm{t}} = \mathbf{y}$$

x is the vector of the variables x_i, which can divided in two groups: the $n_\theta \cdot n_o$ variables of generation, g_o^{θ}, and the $n_\tau \cdot n_{od}$ variables of distribution $p_r^{\tau(\theta)}$, with the following characteristics:

- $x_i \geq 0 \;\; \forall i \in [1, n_\theta \cdot n_o + n_\tau \cdot n_{od}]$
- only for the distribution variables: $\sum_{i=n_\theta \cdot n_o+1}^{n_\tau \cdot n_{od}} x_i = 1$.

For each equation *k*, corresponding to one measure (flow at time slice θ_1 on the link *l*), there is a different matrix $\mathbf{A}^{k}$ with the following characteristics:

- there are just mixed products, that is $a_{ii}^{k} = 0 \ \forall i, k$;
- $a^{k}_{ij} = 0$ if $i,j \in [1, n_{\theta} \cdot n_{o}]$ or $i,j \in [n_{\theta} \cdot n_{o}, n_{\theta} \cdot n_{o} + n_{\tau} \cdot n_{od}]$; furthermore $a^{k}_{ij} = 0$ if $i \in [1, n_{\theta} \cdot n_{o}]$ represents an origin o and $j \in [n_{\theta} \cdot n_{o}, n_{\theta} \cdot n_{o} + n_{\tau} \cdot n_{od}]$ represents an o-d pair r=o'd with o'≠ o or a subperiod $\tau \neq \tau(\theta)$;
- in the other cases, $a^{k}_{ij} = a_{r,\theta,l,\theta_1}$.

Conclusions

This thesis dealt with the o-d estimation problem from indirect measures, addressing two main aspects of the problem: 1) identifying the optimal locations of traffic counts providing the maximum information and therefore the maximum reduction of the uncertainty on the estimate; 2) choosing an estimator to identify univocally and as much reliable as possible the estimate.

The first issue was addressed with an innovative approach and, in accordance with the indications of the literature, in static contexts. Such a choice, however, does not affect in any way the application to dynamic contexts where the sensor location problem should be related to average conditions. The theoretically founded methodology, proposed for addressing the issue of the optimal location of link count sections, explicitly accounts for the variability of the o–d matrix estimate. For this aim, a specific measure based on the trace of the covariance matrix of the posterior demand estimate, conditional upon a given set of count locations, termed SDM (Synthetic Dispersion Measure), was introduced, and a corresponding network sensor location problem formulated accordingly. Under the mild assumption of multivariate normal distribution for the prior demand estimate, the proposed SDM does not depend on the specific values of the counted flows but just on the locations of such sections, and therefore could be effectively used as an objective function in a Network Sensor Location problem. Furthermore, a stepwise algorithm for the calculation of the proposed measure, given a set of link counts, was presented, allowing for the implementation of an effective sequential heuristic algorithm for the solution of the NSLP in real contexts. Some practical examples of the proposed methodology were presented as well, first on 3-link and then on 5-link toy networks, in order to show the rationale of the approach and to illustrate the underlying mathematics, and then on two real networks. The results allowed the proposed NSLP to be compared against the most common methods available in the literature. The proposed approach generally outperforms the other methods analysed, and the proposed sequential heuristic algorithm is almost always able to achieve or to get close to the global optimal solution of the proposed approach (which was found by means of a branch and bound algorithm). In addition, a prototypical application of a budget allocation problem between surveys and traffic counts was illustrated, showing how the proposed framework can help assessing properly a problem whose solution, involving mixed integer programming, is generally not trivial.

Future research developments might focus on the integration, from a modelling point of view, of different measures (normal counts, plate scanner, probe vehicles). In addition, the problem of the optimal allocation of the budget between surveys and counts might also be approached from an algorithmic point of view, looking for more efficient procedures to solve such a mixed integer optimization problem.

The second issue, that is the choice of the estimator, was addressed with reference to within-day dynamic contexts, where hypotheses about the demand evolution can make the estimation problem less underdetermined. A theoretical formulation of the quasi-dynamic o-d flow updating framework was proposed, i.e. assuming constant distribution shares across larger time horizons with respect to the within-day variation of the generation profiles, leading to an estimator which improves dramatically the unknowns/equations ratio. The proposed estimator was tested on a large real dataset with a twofold objective: firstly, to check whether real data supports the quasi-dynamic assumption; secondly, to compare the performances of the quasi-dynamic o-d estimator with both classical off-line dynamic estimators and other possible recursive estimators typically used for on-line dynamic estimation (e.g. the Kalman filter). Experiments were carried out on the real test site of the motorways A4-A23 in North-East Italy. The analysis of the observed o-d flows supported the assumption of quasi-dynamic o-d flows pattern, statistically tested using chi-squared and likelihood ratio tests, with acceptable goodness-of-fit measures even under the hypothesis of constant distribution shares for the whole day. Then, o-d flows updating experiments were performed using the proposed QD-GLS estimator, as well as the simultaneous estimator proposed by Cascetta et al. (1993) and the Kalman filter approach proposed by Ashok (1996). Three main findings may be summarized:

- the QD-GLS estimator outperforms the simultaneous estimator in reproducing dynamic o-d flows estimates;

- the quality of the Kalman filter estimates is quite close to the quality of its seed o-d flows: as a consequence, the QD-GLS estimator is also very useful in supporting on-line applications, since using quasi-dynamic estimates as historical seeds allows the Kalman filter to provide good o-d flow estimates;
- aggregating QD-GLS estimates for successive time slices represents also the most effective way to reproduce o-d flows estimates for larger time horizons (e.g. hourly estimates) for static applications, outperforming in such way estimations coming by both using the classical static estimator proposed by Cascetta et al. (1984) and by aggregating the simultaneous dynamic o-d estimates of the corresponding time slices.

Importantly, also the validation based on an hold-out sample of link flows (i.e. checking the capability of the various estimators to reproduce link flows not used in the o-d estimation/updating process) revealed the QD-GLS estimator to be overall more robust and effective with respect to the other tested estimators.

It is noteworthy that, as concerns the quasi-dynamic estimator, results are related to applications on a highway system; therefore, it would be interesting checking its robustness with respect to a more complex context which is the urban one. Moreover, from a theoretical point of view, the study of the properties of this estimator, which introduces quadratic relations between variables and measures, might be an interesting future research step.

References

Ashok, K., Ben-Akiva, M., 1993. *Dynamic origin-destination matrix estimation and prediction for real-time traffic management systems*. Proceedings of the 12th ISTTT, 465-484.

Ashok, K., 1996. *Estimation and prediction of time-dependent origin-destination flows*. PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Ashok, K., Ben-Akiva, M., 2000. *Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows*. Transportation Science 34(1), 21-36.

Ashok, K., Ben-Akiva, M., 2002. *Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Paths Flows and Link Flows*. Transportation Science 36(2), 184-198.

Balakrishna, R., Koutsopoulos, H.N., Ben-Akiva, M., 2005. Calibration and validation of dynamic traffic assignment systems. Proceedings of 16th ISTTT, 407-426.

Barcelo, J., Montero, L., Bullejos, M., Serch, O., Carmona, C., 2012. *A Kalman Filter Approach for the Estimation of Time Dependent OD Matrices Exploiting Bluetooth Traffic Data Collection*. Transportation Research Board 91st Annual Meeting.

Bell, M.G.H., 1983. *The estimation of origin-destination matrix from traffic counts*. Transportation Science 17(2), 198-217.

Bianco L., Confessore, G., Reverberi, P., 2001. *A Network Based Model for Traffic Sensor Location with Implications on O/D Matrix Estimates*. Transportation Science 35(1), 50-60.

Bierlaire, M., 2002. *The total demand scale: a new measure of quality for static and dynamic origin-destination trip tables*. Transportation Research Part B 36, 837-850.

Bierlaire, M., Crittin, F., 2004. *An Efficient Algorithm for Real-Time Estimation and Prediction of Dynamic OD Tables*. Operations Research 52(1), 116-127.

Cascetta E., 1984. *Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator*. Transportation Research Part B 18(4/5), 289-299.

Cascetta, E., Nguyen, S., 1988. *A unified framework for estimating or updating origin/destination matrices from traffic counts*. Transportation Research Part B 22(6), 437-455.

Cascetta, E., Inaudi, D., Marquis, G., 1993. *Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts*. Transportation Science 27(4), 363-373.

Cascetta, E., Russo, F., 1997. *Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance*. Transportation 24, 271-293.

Cascetta, E., Postorino, M.N., 2001. *Fixed point models for the estimation of O-D matrices using traffic counts on congested networks*. Transportation Science 35(2), 134-147.

Cascetta, E., 2009. *Transportation systems analysis: models and application*, 2nd Edition, Springer.

Castillo, E., Conejo, A. J., Menendez, J. M., Jimenez, P., 2008a. *The Observability Problem in Traffic Network Models*. Computer-Aided Civil and Infrastructure Engineering 23, 208-222.

Castillo, E., Jimenez, P., Menendez, J. M., Conejo, A. J., 2008b. *The Observability Problem in Traffic Models: Algebraic and Topological Methods*. IEEE Transactions on intelligent transportation system 9(2), 275-287.

Castillo, E., Menendez, J. M., Jimenez, P., 2008c. *Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations*. Transportation Research Part B 42, 455-481.

Castillo, E., Gallego, I., Sánchez-Cambronero, S., Rivas, A., 2010. *Matrix Tools for General Observability Analysis in Traffic Networks*. IEEE Transactions on intelligent transportation system 11(4), 799-813.

Chen, A., Chootinan, P., Recker, W., 2005. *Examining the Quality of Synthetic Origin-Destination Trip Table Estimated by Path Flow Estimator*. Journal of transportation engineering, 506-513.

Chen, A., Pravinvongvuth, S., Chootinan, P., Lee, M., Recker, W., 2007. *Strategies for selecting additional traffic counts for improving O-D trip table estimation*. Transportmetrica 3(3), 191-211.

Chootinan, P., Chen, A., and Yang, H., 2005. *A bi-objective traffic counting location problem for origin-destination trip table estimation*. Transportmetrica, 1, 65-80.

Chung, I.-H., 2001. *An optimum sampling framework for estimating trip matrices from day-to-day traffic counts*. Ph.D. Thesis, University of Leeds.

Cipriani, E., Florian M., Mahut, M., Nigro, M., 2011. *A gradient approximation approach for adjusting temporal origin–destination matrices*. Transportation Research Part C, 19(3), 270-282.

Cremer, M., Keller, H., 1981. *Dynamic identification of o-d flows from traffic counts at complex intersections*. Proceedings of 8th ISTTT.

Cremer, M., Keller, H., 1984. *A systems dynamics approach to the estimation of entry and exit o-d flows*. Proceedings of 9th ISTTT.

Eaton, M.L., 1983. *Multivariate Statistics: A Vector Space Approach*. John Wiley and Sons, pp. 116-117.

Ehlert, A., Bell, M.G.H., Grosso, S., 2006. *The optimisazion of traffic count locations in road networs*. Transportation Research Part B 40, 460-479.

Einsenman, S. M., Fei, X., Zhou, X., Mahmassani, H. S., 2006. *Number and Location of Sensors for Real-Time Network Traffic Estimation and Prediction*: A sensitivity

Analysis. Transportation Research Record: Journal of the Transportation Research Board 1964, 253-259.

Fei, X., Mahmassani, H. S., Einsenman, S. M., 2007. *Sensor Coverage and Location for Real-time Traffic Prediction in Large-Scale Networks*. Transportation Research Record: Journal of the Transportation Research Board 2039, 1-15.

Florian, M., Chen, Y., 1995. *A Coordinate Descendent Method for the Bi-level O-D Matrix Adjustment Problem*. Transportation Operations Research 2, 165-179.

Gan, L.,Yang, H., 2001. *Integer programming model for optimal selection of screen lines in road networks.*" Proc., 6th Conf., Society for Transportation Studies, Hong Kong, 122-130.

Gan, L., Yang, H., Wong, S. C., 2005. *Traffic Counting Location and Error Bound in Origin-Destination Matrix Estimation Problems*. Journal of transportation engineering, 524-534.

Gelb, A., 1974. *Applied Optimal Estimation*. M.I.T. Press.

Hazelton, M. L., 2001. *Inference for origin-destination matrices: estimation, prediction and reconstruction*. Transportation Research Part B 35, 667-676.

Hodgson, M. J., 1990. A flow-capturing location-allocation model. Geographical Analysis 22, 271-279.

Hu, S.R., Peeta, S., Chu, C.H., 2009. *Identification of vehicle sensor locations for link-based network traffic applications*. Transportation Research Part B 43, 873-894.

Kalman, R., 1960. *A new approach to linear filtering and prediction problems*. Transactions of the ASME-Journal of Basic Engineering 82 (Series D), 35-45.

Lam, W.H.K., Lo, H.P., 1990. *Accuracy of o–d estimates from traffic counts*. Traffic Engineering and Control 31 (6), 358-367.

Lay, David C., 2005. *Linear Algebra and Its Applications*, third ed. Addison Wesley.

Liu, S.S., Fricker, J.D., 1996. *Estimation of a trip table and the theta parameter in a stochastic network*. Transportation Research Part A 30(4) , 287-305.

Maher, M., 1983. *Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach*. Transportation Research Part B 17(6), 435-447.

Marzano, V., Papola, A., Simonelli, F., 2009. *Limits and perspectives of effective O-D matrix correction using traffic counts*. Transportation Research Part C 17(2), 120-132.

Mínguez, R., Sánchez-Cambronero, S., Castillo, E., Jiménez, P., 2010. *Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks*. Transportation Research Part B 44, 282-298.

Ng, M., 2012. *Synergistic sensor location for link flow inference without path enumeration: a node-based approach*. Transportation Research Part B 46 (6), 781-788.

Okutani, I., Stephanades, Y., 1984. *Dynamic Prediction of Traffic Volume through Kalman Filtering Theory*. Transportation Research Part B, 18(1), 1-11.

Sheldon, Axler, 1997. *Linear Algebra Done Right*, second ed. Springer.

Shilov, Georgi E., 1977. *Linear Algebra*. Dover Publications.

Spall, J.C., 1998a. *An overview of the simultaneous perturbation method for efficient optimization*. Johns Hopkins APL Technical Digest 19 (4), 482-492.

Spall, J.C., 1998b. *Implementation of the simultaneous perturbation algorithm for stochastic optimization*. IEEE Transactions on aerospace and electronic systems 34 (3).

Spiess, H., 1990. *A Gradient Approach for the O-D matrix Matrix Adjustment Problem*. Publication n. 693, CRT, Université de Montréal, Montréal.

Toledo, T., Koutsopoulos, H. N., Davol, A., Ben-Akiva, M., Burghout, W., Andreasson, I., Johansson, T. and C. Lundin (2003). *Calibration and validation of microscopic traffic simulation tools: Stockholm case study*. Transportation Research Record 1831, 65-75.

Van der Zijpp, N., 1996. *Dynamic origin-destination matrix estimation on motorway networks*. PhD thesis, Delft University of Technology, The Netherlands.

Van Zuylen, H.J., Willumsen, L.G., 1980. *The most likely trip matrix estimated from traffic counts*. Transportation Research Part B 14 (3), 281–293.

Yang H., Iida Y., Sasaki T., 1991. *An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts*. Transportation Research Part B 25(5), 351-363.

Yang, H., Sasaki, T., Iida, Y. and Asakura, Y., 1992. *Estimation of origin-destination matrices from link traffic counts on congested networks*. Transportation Research Part B 26(6), 417-434.

Yang, H., 1995. *Heuristics algorithms for the bilevel origin-destination matrix estimation problem*. Transportation Research Part B 29(4), 231-242.

Yang, H., Zhou, J., 1998. *Optimal traffic counting locations for origin-destination matrix estimation*. Transportation Research Part B 32(2), 109-126.

Yang, H., Yang, C., Gan, L., 2006. *Models and algorithms for the screen line-based traffic-counting location problems*. Computers & Operations Research 33, 836-858.

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 41, 823-840.

Zhou, X., List, G. F., 2010. *An Information-Theoretic Sensor Location Model for Traffic Origin-Destination Demand Estimation Applications*. Transportation Science 44(2), 254-273.