

The Multi-path Traveling Salesman Problem with Stochastic Travel Costs: Building Realistic Instances for City Logistics Applications

Original

The Multi-path Traveling Salesman Problem with Stochastic Travel Costs: Building Realistic Instances for City Logistics Applications / Maggioni, Francesca; Perboli, Guido; Tadei, Roberto. - In: TRANSPORTATION RESEARCH PROCEDIA. - ISSN 2352-1465. - ELETTRONICO. - 3:(2014), pp. 528-536. [10.1016/j.trpro.2014.10.001]

Availability:

This version is available at: 11583/2574145 since:

Publisher:

Elsevier

Published

DOI:10.1016/j.trpro.2014.10.001

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

17th Meeting of the EURO Working Group on Transportation, EWGT2014, 2-4 July 2014,
Sevilla, Spain

The Multi-Path Traveling Salesman Problem with Stochastic Travel Costs: Building Realistic Instances for City Logistics Applications

Francesca Maggioni ^a, Guido Perboli ^{b,c,*}, Roberto Tadei ^b

^a*Department of Management, Economics and Quantitative Methods, University of Bergamo, Via dei Caniana 2, 24127, Bergamo, Italy*

^b*DAUIN, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Turin, Italy*

^c*CIRRELT, Pavillon André-Aisenstadt, CP 6128, Succursale Centre-ville, QC H3C 3J7, Montréal, Canada*

Abstract

One of the main issues related to routing problems applied in an urban context with uncertainty related to the transportation costs is how to define realistic instances. In this paper, we overcome this issue, providing a standard methodology to extend routing instances from the literature incorporating real data provided by sensors networks. In order to test the methodology, we consider a routing problem specifically designed for City Logistics and Smart City applications, the multi-path Traveling Salesman Problem with stochastic travel costs, where several paths connect each pair of nodes and each path shows a stochastic travel cost with unknown distribution.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the Scientific Committee of EWGT2014

Keywords: City Logistics; Smart City; Stochastic Travel Costs; TSP; Routing.

1. Introduction

In the past decade, City Logistics pushed researchers towards the definition of a new paradigm of transportation and supply chain integration in urban areas. In recent years, this paradigm has been extended with the introduction of

* Corresponding author. Tel.: +39 011 090 7097; fax: +39 011 090 7099.

E-mail address: guido.perboli@polito.it

the concept of Smart City (Chourabi et al., 2012), where "smart" implies to incorporate a plethora of methods and disciplines in a holistic vision in order to mitigate problems generated by the population growth and its rapid urbanization. In this context, new transportation issues emerge, bringing researchers to define new transportation problems and, in particular, to incorporate information about uncertainty due to the transportation costs and multiple attributes (Perboli et al., 2012; 2011; Tadei et al., 2012). One of the main issues related to routing problems applied in an urban context with uncertainty related to the transportation costs is how to define realistic instances (Tadei et al., 2014). In fact, the instances present in the literature are often too much artificial to really reflect the peculiarities of urban transportation, and the freight transportation in particular. In this paper, we overcome this issue, providing a standard methodology to extend routing instances from the literature in order to incorporate real data provided by sensors networks. This type of data start to be freely available thanks to several European and National grants (see, for example, the data provided by the AperTO project of the municipality of Turin and the Open Data Services of the Piedmont Regional Council).

In more detail, the introduction of this problem is also motivated by a real-life Smart City application, the PIE_VERDE project, a project funded by the ERDF - European Regional Development Fund for the development of new planning tools for freight delivery in urban areas by means of electrical vehicles. In this project, one of the goals is to plan and manage a two-echelon delivery service, where trucks are not allowed to enter into the city. Freight is then consolidated in small peripheral depots and from them brought to customers by means of environmental-friendly vehicles (Crainic et al., 2012; Perboli et al., 2011). In this application context, a crucial part is played by the planning of periodic tours between recurrent nodes and the usage of environmental friendly vehicles as hybrid and electric LDV. In this case, the aim is to plan a tour for each vehicle valid for a given time horizon. Unfortunately, at the time of planning, the decision maker has only a partial idea of the different paths interconnecting any pair of nodes of the transportation network. Moreover, due to congestion the travel time profile of these paths rapidly changes during the day. Similar problems can be found also in other applications, like garbage collection or periodic replenishment of medium-sized grocery stores, where the truck tours are designed in advance and cannot be changed for a fixed number of weeks. Recently, the European project CITYLOG, a joint project between IVECO and TNT, presented the BentoBox, a modular system of containers for envelopes delivery. The containers are usually placed in malls and shopping centres and the company needs to design fixed tours in order to store the envelopes into them.

As test routing problem, we consider the multi-path Traveling Salesman Problem with stochastic travel costs (mpTSPs), a recently introduced stochastic variant of the Traveling Salesman Problem (Tadei et al., 2014). Given a set of nodes, where each pair of nodes is connected by several paths and each path shows a stochastic travel cost with unknown distribution, the mpTSPs aims at finding an expected minimum Hamiltonian tour connecting all nodes. Despite the PIE_VERDE project, the mpTSPs arises in City Logistics when one has to design tours to provide services such as garbage collection, periodic delivery of goods in urban grocery distribution, and periodic checks of shared resources as in bike sharing services. In these situations, the decision maker must provide tours that will be used for a time horizon, which spans from one to several weeks. In such a case, the different paths connecting pairs of nodes in the city are affected by uncertainty due to different time-dependent travel time distributions of the different paths. Moreover, in many cases even an approximated knowledge of the travel time distribution is made difficult by the large size of the data involved and the high variability of the travel times.

This paper contributes to extend the Transportation literature along different directions. First, we introduce a standard methodology to build realistic instances with time-dependent travel times starting from a network of speed sensors (Deflorio et al, 2012). The methodology we propose is easy to replicate, not requiring a deep statistical analysis on the historical data, and can be adapted to other cases where only a partial information about the road network is available. Second, we deeply analyse the results obtained by combining the academic instances with real data on the traffic flows taken from the city of Turin in Italy, showing the benefits and the limits of the methods presently available in the literature to solve the mpTSP.

The paper is organized as follows. In Section 2 the mpTSPs is introduced and the relevant literature is presented. Section 3 is devoted to sketch the experimentation plan, with a special focus on the design of the instances and the traffic scenario generation. The computational results of the newly defined instance sets are discussed in Section 4. Finally, Section 5 summarizes the lessons learned by this computation study and gives some highlights on future works.

2. The mpTSPs problem

Given any pair of nodes, we consider the set of paths between the two nodes. Each path is characterized by a travel cost, which is composed by a deterministic travel cost plus a random term, representing the travel cost oscillation due to the path congestion. Such travel cost oscillations randomly depend on different time scenarios and are actually very difficult to be measured. This implies that the probability distribution of these random terms must be assumed as unknown.

Let it be

- N : set of nodes
- S : set of time scenarios
- K_{ij} : set of paths between nodes i and j
- c_{ij}^k : unit deterministic travel cost of path $k \in K_{ij}$
- $\tilde{\theta}_{ij}^{ks}$: random travel cost oscillation of path $k \in K_{ij}$ under scenario $s \in S$. $\tilde{\theta}_{ij}^{ks}$ are considered independent random variables
- $\tilde{c}_{ij}^{ks}(\tilde{\theta}_{ij}^{ks}) = c_{ij}^k + \tilde{\theta}_{ij}^{ks}$: unit random travel cost of path $k \in K_{ij}$ under time scenario $s \in S$
- x_{ij}^k : boolean variable equal to 1 if path $k \in K_{ij}$ is selected, 0 otherwise
- y_{ij} : boolean variable equal to 1 if node j is visited just after node i , 0 otherwise.

The mpTSP_s is formulated as follows

$$\min_{\{x,y\}} \mathbb{E}_{\{\tilde{\theta}_{ij}^{ks}\}} \left[\sum_{i \in N} \sum_{j \in N} y_{ij} \sum_{k \in K_{ij}} \sum_{s \in S} \tilde{c}_{ij}^{ks}(\tilde{\theta}_{ij}^{ks}) x_{ij}^k \right] \quad (1)$$

subject to

$$\sum_{j \in N: j \neq i} y_{ij} = 1 \quad i \in N \quad (2)$$

$$\sum_{i \in N: i \neq j} y_{ij} = 1 \quad j \in N \quad (3)$$

$$\sum_{i \in U} \sum_{j \notin U} y_{ij} \geq 1 \quad \forall U \subset N \quad (4)$$

$$\sum_{k \in K_{ij}} x_{ij}^k = y_{ij} \quad i \in N, \quad j \in N \quad (5)$$

$$x_{ij}^k \in \{0,1\} \quad k \in K_{ij}, \quad i \in N, \quad j \in N \quad (6)$$

$$y_{ij} \in \{0,1\} \quad i \in N, \quad j \in N \quad (7)$$

The objective function (1) expresses the minimization of the expected total travel cost over all scenarios $s \in S$; (2) and (3) are the standard assignment constraints; (4) are the subtour elimination constraints; constraints (5) link the variables x_{ij}^k to y_{ij} . Finally, (6) – (7) are the integrality constraints. Model (1)-(7) is the so-called Wait-and-See approach (WS), where under each scenario, the realization of the uncertainty is completely known (Maggioni et al., 2014).

In the literature, several stochastic TSP and routing problems can be found. In these problems, a known distribution affecting some problem parameters is given and the theoretical results are strongly connected with the hypotheses on such distribution. The main sources of uncertainty are related to the arc costs, the arc location, and the subset of cities to be visited (Goemans et al., 1991; Maggioni et al., 2009; Maggioni et al., 2014; Maggioni and Wallace, 2012; Mu et al. 2011; Bertazzi, & Maggioni, 2014; Maggioni et al., 2009). Only a little set of papers deal with the choice among multiple paths and they do that by considering the shortest path problem (Eiger et al., 1985; Fu, & Rilett; 1998; Psaraftis, & Tsitsiklis, 1993).

In Tadei et al. (2014) the authors introduced the mpTSP, giving a stochastic formulation of the problem and, under a mild assumption on the unknown probability distribution, a deterministic approximation of the stochastic problem. In more details, by considering $\tilde{\theta}_{ij}^{ks}$ as defined by a unknown probability distribution and independent and identically distributed only among the scenarios s , the authors show that model (1) – (7) can be approximated by the following deterministic problem

$$\min_{\{y\}} \sum_{i \in N} \sum_{j \in N} -\frac{1}{\beta} y_{ij} \ln A_{ij} \quad (8)$$

subject to (2) – (7)

where

- $\beta > 0$ is the parameter of the Gumbel probability distribution (Gumbel, 1958)
- $A_{ij} = \sum_{k \in K_{ij}} e^{-\beta c_{ij}^k}$ $i \in N$, $j \in N$ is the accessibility, in the sense of Hansen (1959), of the pair of nodes i, j to the set of paths between i and j .

In the following, we will refer to model (8), (2) – (7) as the deterministic approximation of the mpTSP_s problem.

3. Experimentation plan

The only instance sets available for the mpTSP_s are the ones presented by Tadei et al. (2014). Unfortunately, those instances, although the costs were generated according to realistic rules, do not fully reflect real cases of City Logistics applications. In particular, the arc costs are only related to the arc length, while, in urban areas, they are more linked to travel times, and then to the vehicle speed profile distributions. Moreover, no real data taken from an existing city were used to generate them. In order to fulfil this gap, in the following we discuss how we generated new instances heavily based on real traffic sensor networks. Thus, we apply two different speed profile distributions: an empirical one, whose values are obtained by data from a real sensor network in the city of Turin, and a theoretical one where the speed values are distributed accordingly to a given distribution. Being our deterministic approximation based on the extreme values theory, we choose the Gumbel distribution for the second speed profile. In this way, the theoretical distribution allows us to measure the error due to the bias introduced by our approximation itself. Hence, the comparison between the empirical and theoretical speed distribution results shows the error due to the bias introduced by our approximation and the error due to the real data distribution.

According to Kenyon and Morton (2003), we generated our inputs as follows:

- **Instances.** We considered all instances in the TSP Library set with a number of nodes up to 200. In particular, we split those instances into two sets: 11 instances with up to 100 nodes (set N100) and 15 instances with number of nodes between 101 and 200 (set N200).
- **Nodes.** Given the portion of plane containing the nodes of the original TSP instances and their position, they are mapped over a square of 14 km edge, which is equivalent to a medium sized city like Turin (see Figure 1 for a depiction of the position of the speed sensors network). The set of nodes is partitioned into two subsets:
 - **Central nodes:** nodes belonging to city center, which are the nodes in the circle with the center coincident with the geometric center of the 14 km square and a radius equal to 7 km;
 - **Suburban nodes:** nodes that are not central.
- **Arc types.** Arcs can be homogenous or heterogeneous.
 - **Homogenous.** They are arcs where the starting node i and the destination node j are both central. In this case, all multiple paths between the nodes present the empirical speed profile of a central speed sensor.
 - **Heterogeneous.** They are arcs where at least i or j belongs to the suburban set. In this case multiple paths between the nodes present the empirical speed profile of a central speed sensor for 1/3 of the paths and a suburban one for the 2/3 of them if the paths are more than 1. If there is only one path between i and j , it has a suburban speed profile.
- **Multiple paths.** The number of paths per each pair of nodes was set to 3 and 5.
- **Speed profile distributions.** The speed profiles can be empirical, i.e. derived by an empirical distribution obtained from the network of sensors, or theoretical, i.e., computed from a given known distribution.
 - **Empirical** speed profile distributions. We generated central and suburban speed profile distributions from real data on the traffic of Turin available at the website <http://www.5t.torino.it/5t/>. The data of the mean vehicle speed, expressed in kilometres per hour (km/h), are accessible with an accuracy of 5 minutes. We aggregated them into blocks of 30 minutes, for a total of 48 observations per day. The instances refer to 9 central speed sensors

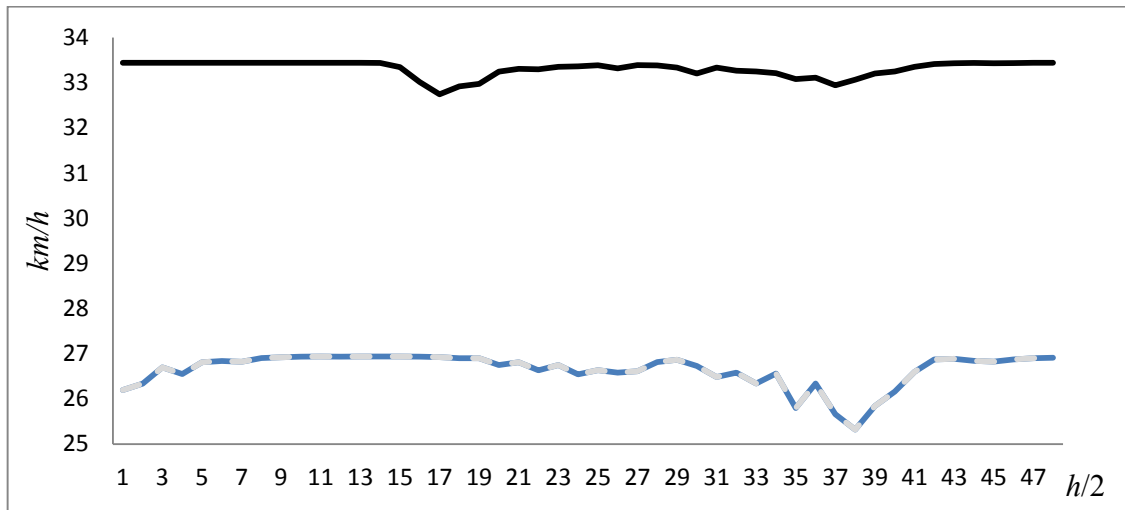


Fig. 2. Two observed speed profiles (in km/h) over 24 hours (48 observations) during a week (gray) and weekend days (black).

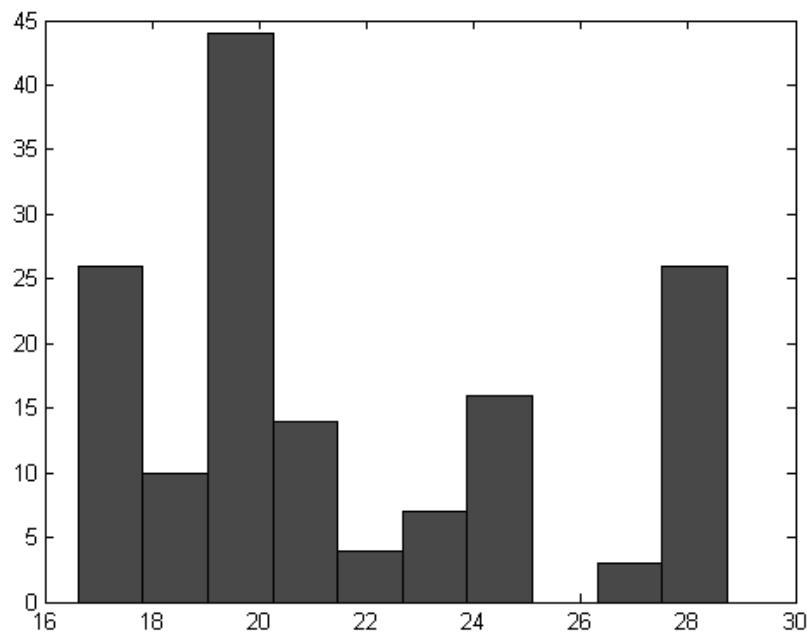


Fig. 3. Frequencies histogram of a simulated central speed scenario according to the procedure described in Section 3.

4. Computational results

In this section, we use the instances defined by the methodology described in Section 3 and we apply them to the stochastic model (1)-(7) and its deterministic approximation (8), (2)-(7). The goals are both to qualify the instances and to evaluate the effectiveness of the deterministic approximation of the mpTSPs we derived. In our computational experiments, travel costs are associated to travel times. We do this by comparing our deterministic approximation with the Perfect Information case (Wait-and-See solution), computed by means of a Monte Carlo simulation over all

possible realization of random speed scenarios. In more details, the model (1)-(7) is solved by means of a Monte Carlo simulation, while for the deterministic approximation (8), (2)-(7) we used the Concorde TSP solver (Applegate et al., 2007).

Aim of the Monte Carlo simulation is to compute the value of the objective function (1) over a set of scenarios S . We decided for a Monte Carlo approach in order to obtain the distribution of the objective function of the stochastic model, as well as the distribution of its variables. Thus, the Monte Carlo simulation gives us the Wait-and-See solution of each problem instance. The Monte Carlo iterates I times the following overall process:

- Create S scenarios with the random speeds v_{ij}^{ks} generated as described in Section 3 and derive by (9), (10), and (11) c_{ij}^k and \tilde{c}_{ij}^{ks} .
- Solve each scenario as follows. Build a TSP with the node set equal to the node set of the stochastic problem. Set the cost c_{ij} between nodes i and j as $c_{ij} = \min_k \{c_{ij}^k + \theta_{ij}^k\}$. Indeed, when a cost scenario becomes known, its optimal solution is obtained by using, as path between the two nodes, the path with the minimum random travel cost. The scenarios are solved to optimality by means of the Concorde TSP solver (Applegate et al., 2007; Cook, 2012).
- Given the scenario optima, compute the expected value of the total cost over all scenarios considered.
- Compute the distribution of the expected value of the total cost for the scenario-based simulations.

In order to obtain the most reliable results of the Monte Carlo simulation, we performed a set of tuning tests by using a subset of instances (5 from N100 and 5 from N200). The values for the parameters R (number of repetitions) and S (number of scenarios) have been set such that the standard deviation of the distribution of the expected value was less than 1% of its mean. These values were $I = 10$ and $S = 100$.

For the deterministic approximation, we used a similar approach. In fact (8) states that in the approximation we can collapse the uncertainty due to the different paths in an arc cost $c_{ij}^* = 1/\beta \ln A_{ij}$. Thus, the resulting problem is a standard TSP with c_{ij}^* as arc costs, which is solved by means the Concorde TSP solver.

Each instance is solved by using the Empirical speed profile and the Theoretical one. This is done to give a comparison between the ideal situation for the deterministic approximation (the Theoretical speed distribution) and the Empirical one, in order to catch the bias introduced by the correlation of the data in a real situation, when compared with the Theoretical one. Table 1 summarizes these data. Column 1 and 2 give the set and the number of paths. Columns 3 and 4 report the percentage gap between the deterministic approximation and the Monte Carlo simulation with the Empirical and the Theoretical speed distributions. Finally, Columns 5 and 6 show the computational effort of the Monte Carlo and the approximation expressed in seconds, respectively. According to these results, we can see how the usage of the Empirical distribution, where the hypothesis of having the complete independency of the variables is not completely fulfilled, means an increase of the gap of about 2%. In particular, we can see how, differently from the Theoretical distribution, in the Empirical one there is a large increase of the gap when moving from 1 to 3 paths, while this increment becomes less evident when the paths becomes 5, showing a sort of asymptotic behavior. From the point of view of the computational time, the deterministic approximation presents a substantial reduction of it, with a gap of about two orders of magnitude. Thus, the data confirm that the deterministic approximation can be used also in a City Logistics environment as a predictive tool for the cost function, while reducing by a large amount the computational time. This makes the deterministic approximation usable in a larger DSS system, letting the user to perform what-if analysis on the overall system, as well as to enlarge the size of the considered instances. Moreover, the small increase of the gap due to the usage of Empirical speed distributions highlights how the deterministic approximation is robust even when the underlying correlation between the different sensors is present. These results become more remarkable when one considers that the deterministic approximation requires no hypothesis on the probability distributions of the stochastic path costs (and then of the stochastic path speeds).

5. Conclusions

In this paper we considered the issue of building realistic instances for urban routing problems in Smart City and City Logistics applications starting from real data taken from a sensor network. In particular, we deal with the situation, typical in real applications, where the network of sensors cope only a subset of the streets.

We used the proposed instance generation approach on the multi-path Traveling Salesman Problem with stochastic travel costs, a stochastic variant of the Traveling Salesman Problem where each pair of nodes of the graph are connected by several arcs with different costs arising in Smart City and City Logistics applications.

In particular, we applied the new set of instances to a Monte Carlo based simulation and a deterministic approximation of the stochastic problem, qualifying the instances and showing how the deterministic approximation is able to find accurate forecasts of the optimal tours when empirical data are provided.

Table 1. Results of the deterministic approximation with the Empirical and Theoretical distributions.

Set	Path	Empirical	Theoretical	T_M	T_{approx}
N100	1	2.72%	0.87%	489.2	4.3
	3	4.05%	2.85%	587.4	5.1
	5	4.93%	2.73%	569.8	7.8
	Avg	3.90%	2.15%	548.8	5.7
N200	1	2.57%	1.14%	1462.2	15.7
	3	6.93%	3.84%	1743.1	18.2
	5	7.52%	5.16%	1634.5	18.6
	Avg	5.68%	3.38%	1613.3	17.5
Global avg		4.79%	2.76%	1081.0	11.6

Acknowledgements

Partial funding was provided by the Regional Council of Piedmont under the "PIE VERDE" project, P.O.R. - FESR 2007-2013 - "AUTOMOTIVE" and by the Italian University and Research Ministry under the UrbeLOG project -Smart Cities and Communities.

References

- Applegate D.L., Bixby R.E., Chvatal V., & Cook W.J., 2007. The Traveling Salesman Problem: A Computational Study, Princeton NJ: Princeton University Press.
- Bertazzi, L., & Maggioni, F., 2014. The Stochastic Capacited Traveling Salesmen Location Problem: a Computational Comparison for a United States Instance. *Procedia Social and Behavioral Sciences*, 108, 47-56.
- Chourabi H., Nam T., Walker S., Gil-Garcia J.R., Mellouli S., Nahon K., Pardo T.A. , & Scholl H.J., 2012. Understanding smart cities: An integrative framework, *Proceedings of the 2012 45th Hawaii International Conference on System Sciences, HICSS '12*, 2289-2297, Washington: IEEE Computer Society.
- Cook W.J., 2012. In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation, Princeton NJ: Princeton University Press.
- Crainic,T.G., Mancini S., Perboli G., & Tadei R., 2012. Impact of generalized travel costs on satellite location in the Two-Echelon Vehicle Routing Problem. *Procedia - Social and Behavioral Sciences* 39, 195-204.
- Deflorio, F., Gonzalez-Feliu, J., Perboli, G., & Tadei, R., 2012. The influence of time windows on the costs of urban freight distribution services in city logistics applications. *European Journal of Transport and Infrastructure Research* 12, 256-274.
- Eiger A., Mirchandani P.B., & Soroush H., 1985. Path preferences and optimal paths in probabilistic networks. *Transportation Science* 19, 75-84.
- Fu, L., & Rilett L.R., 1998. Expected Shortest Paths in Dynamic and Stochastic Traffic Networks. *Transportation Research Part B: Methodological* 32, 499-516.
- Goemans M.X., & Bertsimas D., 1991. Probabilistic analysis of the Held and Karp lower bound for the Euclidean traveling salesman problem. *Mathematics of Operations Research* 16, 72-89.

- Gumbel E.J., 1958. *Statistics of Extremes*, New York: Columbia University Press.
- Hansen W., 1959. How accessibility shapes land use. *Journal of the American Institute of Planners* 25, 73-76.
- Kenyon A.S., & Morton D.P., 2003. Stochastic vehicle routing with random travel times. *Transportation Science* 37, 69-82.
- Maggioni, F., Allevi, E., & Bertocchi, M., 2014. Bounds in Multistage Linear Stochastic Programming, *Journal of Optimization, Theory and Applications* 163(1), doi: 10.1007/s10957-013-0450-1.
- Maggioni, F., & Wallace, S.W., 2012. Analyzing the quality of the expected value solution in stochastic programming. *Annals of Operations Research* 200, 37-54.
- Maggioni, F., Kaut, M., & Bertazzi, L., 2009. Stochastic Optimization Models for a Single-Sink Transportation Problem. *Computational Management Science* 6(2), 251-267.
- Maggioni, F., Potra, F., Bertocchi, M., & Allevi, E., 2009. Stochastic second-order cone programming in mobile ad hoc networks. *Journal of Optimization, Theory and Applications* 143, 309-328.
- Mu Q., Fu Z., Lygaard J., & Eglese R., 2011. Disruption management of the vehicle routing problem with vehicle breakdown. *Journal of the Operational Research Society* 62, 742-749.
- Perboli G., Tadei R., & Baldi M.M., 2012. The stochastic generalized bin packing problem. *Discrete Applied Mathematics* 160, 1291-1297.
- Perboli G., Tadei R., & Vigo D., 2011. The two-echelon capacitated vehicle routing problem: Models and math-based heuristics. *Transportation Science* 45, 364-380.
- Pillac V., Gendreau M., Guret C., & Medaglia A.L., 2013. A review of dynamic vehicle routing problems. *European Journal of Operational Research* 225, 1-11.
- Psaraftis, H. N., & Tsitsiklis J.N., 1993. Dynamic shortest paths in acyclic networks with Markovian arc costs. *Operations Research* 41, 91-101.
- Tadei R., Perboli G., & Perfetti F., 2014. The multi-path Traveling Salesman Problem with stochastic travel costs. *EURO Journal on Transportation and Logistics*, doi:10.1007/s13676-014-0056-2 .
- Tadei R., Perboli G., Ricciardi N., & Baldi M.M., 2012. The capacitated transshipment location problem with stochastic handling utilities at the facilities. *International Transactions in Operational Research* 19, 789-807.