A survey on stochastic location and routing problems

Astrid S. Kenyon* 1, David P. Morton** 2

* Department of Mathematics B, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria (e-mail: kenyon@opt.math.tu-graz.ac.at)
** Graduate Program in Operations Research, University of Texas at Austin, Austin, Texas 78712, USA (e-mail: morton@mail.utexas.edu)

Abstract

We present a survey of the location and routing literature when the demands of the customers are random variables. We concentrate on mathematical programming models and survey the literature for the discrete location, p-median, capacity expansion, vehicle routing, inventory routing and location-routing problems.

1 Introduction

Discrete location and vehicle routing problems are well-studied in the Operations Research literature. In real-life applications, these problems often include random elements which cannot be neglected. Efforts have been made to incorporate this randomness in the modeling and solution processes. This paper reviews location and routing problems where the customer demands are random parameters and the arrival time of the demand is known. We

1 This author's research was supported through Spezialforschungsbereich F 003 "Optimierung und Kontrolle," Projektbereich Diskrete Optimierung.
2 This author's research was supported by the National Science Foundation through grant DMI-9702217.
concentrate on mathematical programming models. For a summary of other approaches, we refer the reader to the surveys of Louveaux [177] and Laporte and Osman [165].

This paper is organized as follows. We first review methods used to handle uncertainty in the parameters of a problem. We then summarize important properties of one of these approaches, stochastic linear programming (SLP), and more precisely stochastic integer programs (SIP). We present the methods used in the literature to address problems formulated as SIPs. This section was designed to be a toolbox for SIP, and consequently is not limited to location and routing problems. When appropriate, we specifically refer to these problems to illustrate the methods. Next, we take a problem-oriented approach to survey the literature for several applications, the stochastic discrete location, $p$-median, capacity expansion, vehicle routing, transportation location, inventory routing and location-routing problems.

2 Handling uncertainty

When constructing a realistic mathematical model, one is often faced with uncertainty in the parameters of the problem. It is sometimes appropriate to not include this uncertainty in the model itself and to study its impact on the solution to the deterministic problem by estimating how variance in the data influences an optimal decision and the optimal objective value. This approach is called sensitivity analysis and can be used when the uncertainty does not play a fundamental role in the application at hand. For examples of such studies for locations problems, see [76, 127, 154]. For a more general discussion of this issue see Wallace [249]. Labbé et al. [154] look at the problem of locating a new facility so as to minimize a weighted sum of distances from given demand points. The authors define methods to determine a trade-off curve of variability in the parameters versus the degree of optimality of a solution. They detail ranking procedures to obtain such a curve for two special cases, a location problem on a tree network, where the procedure is simple, and a location problem under a block norm, where a sequence of linear programs has to be solved. Hodgson [127] studies the stability of solutions to the $p$-median problem. Drezner [76] applies sensitivity analysis to the Weber problem. The author derives a set of differential equations that formulate the change in the location of the new facility when sites and weights of demands are changed and studies the set of all possible locations for the optimal solution when the demands are restricted to small areas and weights to given ranges.

For many applications, uncertainty plays a more fundamental role, and needs to be implicitly taken into consideration in the model. We distinguish stochastic models, where the uncertainty is represented by random distribution
functions, and set-valued problems, where one sets intervals for the values of the random parameters. In the latter case, the effect of the worst-case scenarios are studied via regret functions (e.g., [6]).

Important features differentiating stochastic models include the timing of the decisions and the timing of the release of information. The planning can be static, when all the decisions have to be taken before the operations start, or adaptive, when some decisions have to be taken during the operations, at one or several decision stages. Decisions taken during the operations are functions of the available information obtained at that point in time.

When the decisions are taken during the operations, the timing of the decision stages can be fixed (e.g., first decide on the location of the plants, then obtain information about the customer demands and allocate customers to the plants) or itself random (e.g., when an event occurs). The latter type of problem can be modeled using stochastic processes. Examples can be found in stochastic covering, capacity expansion, and some routing problems. In the stochastic covering problem [25], mobile units such as ambulances or emergency units are stationed at some locations in a transportation network. Customers, located at the nodes, request service to a dispatcher. If a service unit is available, it is dispatched to the customer location. If not, a congestion occurs, and two strategies can be used: either the customer is lost, or the customer enters a FIFO queue. Markovian models are defined when the service times can be considered exponentially distributed (e.g., [21, 190]). When the service times are dominated by the travel times, they depend on the locations of the servers and have non-Markovian distributions (e.g., [23, 61]). For more details on these models, see [25] for a review until 1990 and more recent publications such as [10, 111, 187, 188, 212]. The capacity expansion problem consists of choosing the location, size and timing of facilities needed to expand the operations of a company. Demands have been modeled as Poisson processes [70], Cox processes [70], diffusion processes [185, 239], birth-and-death processes [104, 16], nonlinear Brownian motion [16], or alternating renewal processes [48]. The resulting models are analyzed as stochastic processes (e.g., [70]) or transformed into deterministic equivalent models (e.g., [16, 105]). Routing problems where the demands vary over time, such as courier or pizza delivery services, are other examples of this approach. For these problems, the objective is to minimize the waiting time of the customers, not the traveling cost. Policies have to be implemented in real time. Bertsimas and Van Ryzin [33, 36] analyze this problem when the demands arrive according to a Poisson process and their locations are independent and uniformly distributed throughout the Euclidean service region. They find policies which are optimal in light traffic and policies that have results whose ratio to the optimum solution is within a constant factor in heavy traffic. Bertsimas and Van Ryzin extend their analysis to the case
where the demands are distributed with an arbitrary continuous density over the service region and arrivals form a renewal process [37]. Another type of approach is presented by Powell [205]. The author addresses the problem of real-time dispatching of truckload motor carriers. He presents a stochastic model of the dynamic booking process and shows how it can be integrated into a stochastic network model for planning vehicle movements. He also develops a methodology for evaluating and testing dynamic fleet management models in a continuous time setting using rolling horizon simulations.

In this paper, we concentrate our efforts on the stochastic programming approach. The uncertainty is represented by a discrete set of scenarios or by a more general probability distribution. Scenarios represent the possible values of the random parameters in the future. This modeling approach assumes that the distributions of the random parameters are known or can be evaluated, and that these distributions are independent from the decision vector. The decisions are taken at one or several decision stages with fixed timing.

3 Stochastic linear programming

In this paper, we focus on stochastic programming approach to stochastic location and routing problems. Before presenting the literature for these problems, we briefly review the terminology of SLP.

A two-stage linear program may be expressed

\[
\begin{align*}
\min & \quad cx + fy \\
\text{s.t.} & \quad Ax \geq b, \\
& \quad Bx + Dy \geq d, \\
& \quad x \geq 0, \quad y \geq 0,
\end{align*}
\]  

(1)

where \( x \) and \( y \) are vectors of decision variables, of dimension \( n_1 \) and \( n_2 \) respectively, \( A, B \) and \( D \) coefficient matrices of dimension \( m_1 \times n_1, m_2 \times n_1, \) and \( m_2 \times n_2, b \) and \( d \) right-hand side vectors of dimension \( m_1 \times 1 \) and \( m_2 \times 1, \) and \( c \) and \( f \) cost vectors of dimension \( 1 \times n_1 \) and \( 1 \times n_2. \) Many time-staged linear programs fit this, or a more general, staircase form (e.g., capacity expansion problems [45]).

Let us introduce random elements \( B, D, \) \( d \) and \( f \), and call the resulting vectors and matrices \( \bar{B}, \bar{D}, \bar{d} \) and \( \bar{f} \) to emphasize their stochastic nature. Introducing these elements directly in model (1) yields an incorrect formulation since a random objective cannot be minimized. Instead, a function of the random objective function, such as its expectation or its median, must be considered. We now present ways to formulate the stochastic version of
(1). For simplicity, we call the collection of all random quantities $\tilde{\xi}$ (i.e., if $\tilde{B}, \tilde{D}, \tilde{d} \text{ and } \tilde{f}$ are random, $\tilde{\xi} = \text{vec}(\tilde{B}, \tilde{D}, \tilde{d}, \tilde{f})$ where \text{vec} is an operator that reads its arguments columnwise).

In formulating a stochastic program, it is key to clearly specify the timing of when the decisions $x$ and $y$ are made relative to knowing the realization of the random vector $\xi$. If the decision vectors $x$ and $y$ can be taken after observing $\xi$, the optimal solution value is a random variable, and we have a \textit{wait-and-see} model [182].

If the decisions $x$ and $y$ must be made prior to observing $\xi$ and if infeasibility is accepted, but only if it occurs with a low probability $\alpha$, $0 \leq \alpha \leq 1$, the model can be extended by using constraints of the form

$$P(\tilde{B}x + \tilde{D}y \geq \tilde{d}) \geq 1 - \alpha.$$  \hspace{1cm} (2)

Instead of enforcing probabilistic constraints in this joint manner, we can enforce them individually. In this case, tolerance levels $\alpha_i$ are introduced for each constraint with random coefficients. The constraints then have the form

$$P(\tilde{B}_i x + \tilde{D}_i y \geq \tilde{d}_i) \geq 1 - \alpha_i, \quad i = 1, \ldots, m_2,$$  \hspace{1cm} (3)

where $\tilde{B}_i$, $\tilde{D}_i$ and $\tilde{d}_i$ denote the $i^{\text{th}}$ rows of $\tilde{B}$, $\tilde{D}$ and $\tilde{d}$. Constraints (2) or (3) define respectively joint \textit{chance-constrained models} and \textit{individual chance-constrained models}. Chance-constrained programming was first proposed by Charnes and Cooper [58]. In general, chance-constrained problems are not convex mathematical programs. For information on these models, refer to Prékopa [207], Ermoliev and Wets [90], Kall and Wallace [144] and Birge and Louveaux [45].

In another type of stochastic model, potential infeasibilities are avoided at a certain cost. The first-stage decision vector $x$ must be chosen here-and-now, before a realization $\tilde{\xi}(\omega)$ of the random vector becomes known. Once $\tilde{\xi}(\omega)$ is known, recourse actions $y$ are taken to avoid potential infeasibilities, and incur a recourse cost $\tilde{f}y$. These recourse actions are taken for a given decision vector $x$ and realization $\tilde{\xi}(\omega)$ of the random vector. It yields:

$$\begin{align*}
\min & \quad cx + E[h(x, \tilde{\xi})] \\
\text{s.t.} & \quad Ax \geq b, \\
& \quad x \geq 0, \\
\text{with} & \quad h(x, \tilde{\xi}) = \min \tilde{f}y \\
& \quad \text{s.t.} \quad \tilde{D}y \geq \tilde{d} - \tilde{B}x, \\
& \quad y \geq 0.
\end{align*}$$  \hspace{1cm} (4)

This model is known as a \textit{two-stage stochastic linear program with recourse} [15, 64]. The recourse costs are taken into the objective function, weighted
with their probability of occurrence. Thus the objective function of the two-stage recourse model is the sum of the first-stage cost and the expected value of the second-stage cost. The matrix $\bar{B}$ is called the technology matrix and $\bar{D}$ is the recourse matrix. The concept of recourse models has been generalized to multistage settings [45], corresponding to the idea that realizations of subsets of random parameters become available over time. A special case, called simple recourse [253, 259], rises when $\bar{D}$ is the identity matrix. In such case, the penalty to pay for infeasibility in the second-stage is proportional to the magnitude of the violation(s). Since no recourse actions are taken, this model is sometimes referred to as static. Simple recourse models are the "simplest of all 'nontrivial' stochastic programming problems" [253].

A further distinction in the class of recourse models deals with the question of feasibility. If a feasible solution to the second-stage problem can be found with probability one (wp1) no matter what feasible first-stage decision is made (i.e., for all $x \in X = \{x : Ax \geq b, x \geq 0\}$), the model is said to have relatively complete recourse. The recourse is called complete if the second-stage problem is feasible, wp1, for all $x \in \mathbb{R}^{n_1}$, i.e., even for $x \not\in X$. For more information on models with complete and relatively complete recourse, see [252]. In most models of real-world systems, relatively complete recourse is a reasonable assumption when "infeasibilities" can be modeled using a penalty-based formulation.

The function $E[h(x, \tilde{\xi})]$ defined in (4) is, in general, nonlinear, nondifferentiable and convex. Function and (sub)gradient evaluations of $E[h(x, \tilde{\xi})]$ involve an $n_\xi$-dimensional expectation (where $n_\xi$ is the dimension of the random vector $\tilde{\xi}$). These evaluations govern the computational difficulty of stochastic linear programs (SLP). Depending on the characteristics of the distributions, exact evaluations of $E[h(x, \tilde{\xi})]$ may be impossible even for small values of $n_\xi$.

Solution techniques for SLPs can be classified in three categories: exact solution methods, approximation and bounding, and sampling. For a survey of these methods, see [43]. When the vector of random parameters has a small number of realizations, it is possible to reformulate (4) as a single large-scale linear program [252]. Exact solution methods may be applied to solve this problem and include simplex-based algorithms that exploit the structure of the problem (e.g., [235]), decomposition schemes such as the L-shaped algorithm [229] and its variants [41, 218], interior point methods [55], Lagrangian schemes [194] and the Progressive Hedging algorithm [214]. For problems with many scenarios or continuous random parameters, developing an approximation is usually necessary. These approximations are based on deterministically valid bounds or on Monte Carlo sampling. Deterministically valid bounds are derived from approximations of the recourse function.
[40, 42, 84, 204] or of the underlying probability distribution [86, 101, 181]. These bounds are utilized in a conditional fashion within sequential approximation algorithms [85, 102, 130] that iteratively improve the bounds by refining partitions of the random vector's support. However, good deterministic bounds may sometimes be difficult to compute, e.g., when $\tilde{\xi}$ is high dimensional. In these cases, sampling techniques offer an attractive alternative. Dantzig and Glynn [65], Dantzig and Infanger [67] and Higle and Sen [122] employ Monte Carlo sampling within decomposition schemes. Approaches in which the sampling is done external to the solution procedure are considered in [213, 215] as well as [183].

4 Stochastic integer programming

In stochastic (mixed-) integer programming with recourse, integrality constraints are imposed on (some of) the first-stage and/or second-stage decision variables of model (4). Both stochastic location and routing problems have that property, and we now study in more details stochastic integer programs. Stochastic integer programs have been developed for many problems, such as planning and scheduling problems [46, 73, 103], location problems [68, 163, 175, 176, 211], resource acquisition [39], stochastic knapsack problem [57, 193] capacity expansion of hydrothermal systems [172], unit commitment in power generation problem [53, 238], asset-liability management [51, 75], menu planning [2], the sequencing of offshore oil and gas fields [141] and routing problems [135, 147, 148, 161]. For a recent review of stochastic integer programming, see [121].

In this section, we review structural properties of the recourse function, $E[h(x, \tilde{\xi})]$ and solution methods applied to SIPs. Two types of solution methods are found in the literature: the ones aiming at (near-)optimality and heuristics.

4.1 Structural properties of stochastic integer programs

When only the first-stage decision variables are required to be integer, the properties of the recourse function, such as convexity, are maintained (the recourse function being defined on the convex hull of the first-stage feasible region). Such problems can be approached using "fairly conventional adaptations of SLP methods," [233] as proposed in [258]. For example, Laporte et al. use branch-and-cut techniques to solve stochastic location [163] and routing [161] problems.

When integrality is also required for the second-stage decision variables, two difficulties arise. First, each evaluation of the second-stage problem requires the solution of an integer program, which is in general NP-hard. Next, the
recourse function does not conserve the desirable properties of continuous SLPs. For example, the objective function may be discontinuous when the random parameters have discrete distribution [233]. If the random parameters have continuous distributions, the objective function may no longer be discontinuous (see below), but is not necessarily convex [233]. Hence, methods which exploit convexity of SLPs cannot be applied directly to such SIPs.

Research on the properties of the recourse function has focused on two special cases of SIPs: complete mixed-integer recourse program (CMIR) and simple integer recourse (SIR) models. In these models, the second-stage cost coefficients \( f \) and the recourse matrix \( D \) are assumed to be deterministic. The second-stage right-hand side vector is random, whereas the technology matrix \( \tilde{B} \) is fixed in some studies. Note that stochastic location and routing problems with random demands fit this structure.

Schultz [219] derives continuity of \( E[h(x, \xi)] \) for CMIR under three conditions: (i) \( \tilde{d} \) is the only random element (\( \xi = \tilde{d} \)), (ii) every component of \( \tilde{d} \) is random, and (iii) \( \tilde{d} \) has an absolutely continuous distribution. He also studies the stability of these models under perturbations of the integrating probability measure. Sufficient conditions for Lipschitz continuity of the recourse function of the CMIR model with fixed technology matrix are given in [220]. These results are extended to the CMIR model with random technology matrix in [222].

Louveaux and Van der Vlerk [178] describe analytical and computational properties of the recourse function of SIR models with fixed technology matrix. These results are extended to SIR models with random technology matrix in [116].

### 4.2 Optimization methods

Several types of optimization methods are discussed in the literature. They attempt to exploit different properties of SIPs. Because SIPs are integer problems, methods based on cuts and branch and bound have been developed. Boolean methods have been applied to binary problems. Because the geometry of the second-stage problems does not vary much when only the right-hand side vector is random, Gröbner bases, a tool from commutative algebra [236], are used to speed up their evaluations. For a survey of these methods, see [221]. Finally, dynamic programming has been used for some SIPs with special structures.

#### 4.2.1 Integer L-shaped method.

The L-shaped method for two-stage stochastic linear programs with continuous decision variables was introduced by Van Slyke and Wets [229]. It is a
cutting-plane method that iteratively generates a sequence of first-stage decisions and associated supporting hyperplanes (i.e., cuts) of the convex recourse function, to construct a piecewise linear outer approximation of $E[h(x, \xi)]$. Wollmer [258] extends the method to two-stage models with binary first-stage variables and continuous second-stage variables. He generates cuts in an identical fashion to that of Van Slyke and Wets. Leopoldino et al. [172] use an extension of Benders' decomposition for a power system expansion planning problem with binary first-stage and continuous second-stage decision variables, and random arc reliabilities and demands. In their model the expected recourse cost does not appear in the objective function but is required to be below a certain maximum level of unsatisfied demands.

When the second-stage has integer-constrained decision variables, the recourse function is non-convex and this makes it more difficult to apply cutting-plane procedures. Averbakh [3] uses a Benders' decomposition scheme to solve problems with a finite set of discrete first-stage decision variables and continuous second-stage decision variables. The method assumes that the law of large numbers holds to avoid the exclusion of optimal solutions when applying the cuts. In [4], Averbakh extends this method to problems with constraints involving an expectation, discrete or continuous second-stage decision variables and continuous distribution of the random vector. The method assumes that the second-stage problems can be solved analytically.

Laporte and Louveaux [162] adapt the L-shaped method to problems with binary first-stage, continuous or binary second-stage decision variables and a discrete or continuous random vector. This method requires that exact evaluations of the recourse function are possible for a fixed first-stage decision vector. Laporte and Louveaux combine the original L-shaped method with a branch-and-bound procedure to deal with the binary variables. The adapted method is called the integer L-shaped method and has been applied to stochastic location problems [163] and vehicle routing problems with random demands [107, 164, 125].

Carøe and Tind [54] use integer programming duality theory to generalize the L-shaped method to cover two-stage integer recourse models with continuous first-stage decision variables, integer second-stage variables and a discretely distributed random vector.

4.2.2 Branch-and-Bound methods.

Ermoliev et al. [91] and Norkin et al. [198] present a stochastic branch-and-bound method. As for deterministic integer programs, the main idea is to partition the solution set into smaller subsets and to use associated bounds on the optimal objective value to guide the branching process and remove subsets that cannot (or, in this case, are unlikely to) improve upon the current
candidate solution. For a stochastic integer program, the authors estimate bounds by using Monte Carlo sampling and apply the method to financial and facility location problems.

Carøe and Schultz [52] combine a dual-decomposition method with a branch-and-bound procedure for multistage SIPs with mixed integer variables in each time stage and a finite number of scenarios. The bounding procedure is based on a Lagrangian relaxation of the non-anticipativity constraints, a set of constraints that ensure, e.g., that the first-stage decisions do not depend on the scenario which prevails in the second-stage. Carøe and Schultz present numerical results for test problems with up to 16 scenarios and subproblems with 10 binary and 65 continuous decision variables, 31 constraints and randomness in the right-hand side vector only. A successful application of this method to unit commitment in a German hydro-thermal power system is described in [53].

4.2.3 Pseudo-Boolean programming methods.

Ettinger and Hammer [92] examine expectations of functions with independent discrete coefficients and binary variables. They show how to transform such problems into equivalent deterministic pseudo-Boolean programs when the characteristic function of the pseudo-Boolean functions has a specific form. The resulting problem is solved by existing algorithms of Boolean programming.

4.2.4 Gröbner bases methods.

When randomness appears only in the right-hand side vector, solving an SIP requires the solution of many integer programs which differ only in their right-hand side. Gröbner bases methods exploit this by building a basis which captures the algebra of the second-stage. The number of elements in a Gröbner basis can be large and they can be expensive to compute. However, once this basis is found, solving the second-stage problem for different right-hand side vectors only requires a division of multivariate polynomials.

This method was used by Schultz et al. [221] to solve SIPs with complete recourse and discretely distributed right-hand side vectors. Tayur et al. [240] present an algorithm to solve chance-constrained integer programs. They apply this algorithm to the scheduling of job types on parallel machines when the demands are correlated random variables.

Schultz [223] suggests a preprocessing technique to reduce the search in the enumeration process of the algorithm presented in [221]. He uses Fourier-Motzkin elimination [248], a method eliminating variables using projections.
Nevertheless, because this elimination process is of exponential complexity, this method can only be used for problems of moderate size.

4.2.5 Dynamic programming.

Some deterministic combinatorial optimization problems are more efficiently solved using dynamic programming rather than integer programming methods. Similarly, some stochastic combinatorial optimization problems with special structures are more efficiently solved using stochastic dynamic programming. Lageweg et al. [156, 157] show how to formulate two-stage scheduling, bin packing, and multiknapsack problems as dynamic programs. As in the deterministic case, formulating a dynamic program with a low-dimensional state vector is key to computational tractability. Carraway et al. [56] apply dynamic programming to the stochastic traveling salesman problem, in which the objective is to find a tour with maximum probability of completion by a specified time.

4.3 Heuristics and other approximation methods

A common approach in solving hard combinatorial optimization problems is to not insist on optimality but to design fast and high quality approximation methods. This section presents several approximation approaches.

4.3.1 Hierarchical planning.

Hierarchical planning consists of splitting a problem into levels of decision. When two levels are defined, the aggregate level may be regarded as the first-stage problem in stochastic programming, and the more detailed one as the second-stage. A heuristic can be used to solve the detailed level (i.e., to approximate $h(x, \xi)$). This is most valuable when this second level is NP-hard. A heuristic may also be needed in the aggregate level to obtain an approximation of the expected recourse function, even when $h(x, \xi)$ can be computed easily for specific realizations of $\xi$ [170]. The success of this approach depends on the quality of the heuristics. This can sometimes be evaluated using results from probabilistic analyses of combinatorial optimization problems and heuristics. For such analyses for routing problems see, for example, [17, 146, 230].

The hierarchical approach for stochastic integer programs was "initiated by Dempster et al. [71, 73] and worked out by Lenstra et al. [170] and Stougie [233]" [234]. A survey paper [72] presents a general framework. Hierarchical planning methods have been designed for two-stage stochastic scheduling [44, 71, 103, 233], two-stage stochastic location, routing location [233, 226] and routing problems [186].
In hierarchical vehicle routing problems, a fleet has to be acquired in the first stage, without knowing exactly the customer locations. The first-stage decisions are made using a lower bounding approximation of the recourse function. Once the customer locations are known, vehicle routes are defined in the second-stage. The region on which the customers are spread is partitioned, and tours are built on each subregion and then combined in a suitable manner. For a survey of the hierarchical vehicle routing problems and the heuristics used, see [233].

Simchi-Levi [226] looks at location routing problems with random and identically distributed customer demands. He first locates the facilities and then designs the vehicle routes. The author shows that when customers are independently and uniformly distributed in a given area, the minimal expected distance traveled behaved asymptotically as the cost of the $p$-median, $p$ being the number of depots. The optimal number of depots is shown to minimize a given expression. Since this number may not be integer, it is rounded up or down. Depots should be located in a regular hexagonal pattern through the area. Asymptotically optimal routing strategies are then found using the iterated tour partitioning heuristic.

4.3.2 Convex approximations of simple integer recourse programs.

A second approach consists of finding convex approximations of the recourse function of the original problem and solving the approximation problem instead of the original one. For a review of such methods, see [245].

A first convex approximation of the recourse function is its convex hull, i.e., the largest lower semicontinuous function majorized by the recourse function. Haneyvel et al. [117] prove that the convex hull of the recourse function of a simple integer recourse program is equal to the recourse function of the related continuous simple recourse program plus a constant, after a suitable transformation of the random vector. The resulting continuous problem can be solved using existing algorithms for continuous simple recourse problems. The computational burden in this method is due to the presence of $E[h(x, \xi)]$ in the objective function and lies primarily in the construction of the convex hull. Haneyvel et al. [118] present an algorithm for computing the convex hull of the recourse function of SIR problems with discretely distributed right-hand side vectors with a finite number of mass points and fixed technology matrix.

Another approximation can be found by replacing the distribution of the random vector by another distribution such that the resulting recourse function is convex. The approximation is called convexity by perturbation of the underlying distribution and has been developed for SIPs with a continuous random right-hand side vector [244]. A description of a special class of distri-
butions that result in convex recourse functions is given in [119]. A uniform bound on the error of such approximations is developed in [120] for the "most elementary" of these distributions.

### 4.3.3 Approximations using the structure of the problem.

Another method of modifying the original problem is to exploit some of its properties. Bastian and Rinnooy Kan [14] look at single vehicle penalty models with random demands. They transform this model into a time-dependent traveling salesman problem (TDTSP, a TSP problem where the cost vector changes over time) and use dynamic programming to address the resulting model. They also propose a method to transform the chance-constrained version of this model into a TDTSP. Dror [82] looks at stochastic routing problems with random demand values. He modifies the network using properties of the problem based on the triangle inequality and assumptions on the maximum value of the demands and on the behavior of the vehicles when route failures occur (a route failure occurs when the demand at a customer location exceeds the load carried by a vehicle). The optimal solution on the new network is an Hamiltonian tour.

### 4.3.4 Other approximation methods.

Armstrong and Balintfy [2] adapt a block pivoting technique for multiple choice programming [9] to chance-constrained multiple choice programs with normally distributed technology matrix coefficients. The method consists of rewriting the problem constraints with the inverse distribution function of random elements. The resulting deterministic constraints are then used for the generation of a feasible solution for the problem and for the improvement of this feasible solution.

Løkketangen and Woodruff [174] embed tabu search in a progressive hedging algorithm to solve multistage mixed-integer zero-one programs.

Gendreau et al. [108] use tabu search to solve stochastic vehicle routing problems with random demands and where the customers are present with some probability. They compare the results obtained using this heuristic with known optimal solutions on problems with 6 to 46 customers: the procedure yields "an optimal solution for 89.45% of the cases, with an average deviation of 0.38% from optimality."

Higle and Sen [124] present a decomposition method for two-stage SIPs with integer variables in the second-stage. The method uses relaxations of both the master and subproblems based on estimates of the second-stage objective function.
5 Stochastic location problems with random demands

In this section, we are interested in optimizing the location of facilities delivering a commodity to a set of customers when the customer demands are random variables. We first present the general problem, the discrete facility location problem. Because this problem is a transportation problem when the locations are fixed, we also survey stochastic transportation problems. We then look at a special case of the discrete location problem, the $p$-median, where the number of locations to be opened is fixed. For reviews of the deterministic versions of these problems, see [69, 77, 97, 132, 155, 179, 192]. Finally, we review a related problem, the capacity expansion problem, where not only the size and location of the facilities but also the timing of the openings are decisions to be optimized. For a review of deterministic capacity expansion problems, see [180].

5.1 Discrete facility location problems

In the discrete facility location problem, a fixed number of customers must be supplied from facilities where a commodity is made available. The customer demands can only be satisfied from opened facilities. The problem consists of selecting the locations of the facilities from a finite set of candidate locations, and in deciding the fraction of the customer demand satisfied from a given open location. In the capacitated location problem, additional constraints on the total supply available from the facilities and the size of the facility to open can be added.

Louveau [177] reviews models for the stochastic location problem. Three models are studied in the literature. They differ by the timing of the decisions. In a first model ($M_1$), all the decisions have to be taken before the random vector is known, and costs are incurred for unsatisfied demands and excess supplies (e.g., [8]). In a second model ($M_2$), only the decisions about the locations of the facilities and their sizes are decided in the first stage, and the distribution decisions are made in the second stage (e.g., [175, 176]). Note that for a given location, the second stage is a transportation problem. A third modeling approach ($M_3$) is to decide on the facility locations and sizes in the first stage, as well as the allocation of customers to the plants. In the second stage, the amount of products sent from the opened facilities to their assigned customers is made (e.g., [163]). Jucker and Carlson [143] present models representing different risk-aversion strategies. They also study a special case of the discrete stochastic location problem, where the contribution to profit and fixed costs for a region for a certain plant is larger than for any other plant. They show that under this dominance condition the problem can be decomposed into easily solvable subproblems, generalizing the deter-
ministic result of Efroymson and Ray [87].

We now survey the methods used in the literature to address the different versions of the stochastic location problems. Three optimization approaches have been considered for the stochastic location problem with random demands: branch-and-bound methods, Benders' decomposition methods and dual-based procedures. For the main contributions for each of these methods, see [198], [163] and [176], respectively.

### 5.1.1 Methods to address model M1:

In this model, the location and distribution decisions are taken before the demands are known. After realization of the random vector, costs associated with shortage or excess can be incurred. Note that once the first-stage decisions are taken, these costs are fixed and there are no real second-stage decisions.

White and Francis [254] consider the problem of finding optimum warehouse sizes under random demands, but do not consider the location or distribution aspects of the problem. Leblanc [168] considers a discrete number of possible locations and random demands. He minimizes the expected holding, shortage, shipping and fixed construction costs. He proposes a heuristic that uses dual variables to guide the determination of whether a plant should be opened or not. Hodder [126] discusses the financial aspect of facility location problems under uncertainty. He compares models with variance- and covariance-based risk measures. The uncertainty in the demand is expressed as a random selling price rather than a target demand. The author suggests using branch-and-bound techniques or dual-based procedure to solve the uncapacitated versions of both models and the capacitated version of the model with the covariance-based objective function. Balachandran and Jain [8] look at the problem of finding optimal facility locations, their sizes, and the allocation of products to customers, when the production cost at each facility is a piecewise linear function. Once the random elements are known, the firms can experience a cost for unsatisfied demands or a penalty cost for excessive supplies. The authors propose an iterative branch-and-bound method based on the KKT conditions. For each subproblem, the production cost is assumed to be a linear function and the demand is probabilistic. To solve the subproblems, the authors use results from Charnes et al. [59] to show that the second-stage problem is equivalent to a constrained generalized median problem with absolute value terms in the objective function, and results from Garstha [106] to transform this objective function into an equivalent convex function.

Ermoliev et al. [91] propose a stochastic branch-and-bound method to address location problems with random demands. They partition the set of de-
cisions into smaller subsets and use bounds on the objective function within the subsets to guide this process, similarly to deterministic optimization. Exact bounds being hard to compute, the authors use recursive allocation of observations to the subsets to improve stochastic bounds. Two methods are proposed to estimate the lower bounds: interchange of minimization and mathematical expectation operators (see also [183]) and dual estimates based on Lagrangian multipliers (see also [123]). The authors write the latter bound for the location problem and show that it can be computed using standard linear programming techniques. For further developments, see [198].

5.1.2 Methods to address model M2:

In this model, the decisions related to the facilities are taken before the demand values are known, and the decisions related to the distribution of the commodity after the realization of the random vector. For given locations, the second-stage is a stochastic transportation problem [255].

Franca and Luna [96] use a generalized Benders' decomposition method to solve a mixed-integer nonlinear program modeling this problem. The method has convex stochastic transportation subproblems, whose optimal multipliers generate cuts for a master integer location problem.

Louveaux [175] models the location problem with random demands, production and transportation costs, and selling prices. Louveaux and Peeters [176] present a dual-based procedure for this problem. The algorithm can be used when the random vector has a limited number of possible realizations. This method has three main steps: (i) a heuristic dual-descent procedure, applied to the dual of the original problem where the integrality condition on the location binary variables are replaced by non-negativity constraints; this procedure produces a set of optimal or near-optimal feasible solutions; (ii) a procedure for constructing a corresponding primal solution that satisfies most complementary slackness conditions; (iii) a primal-dual adjustment procedure that alters the primal and dual solutions in order to reduce the sum of the complementary slackness violations. Results are presented for problems with 50 nodes, 16 or 25 potential locations, and 1, 3 or 5 possible realizations of the random vector. With 16 potential facility locations, CPU times of the order of 100s on an IBM 370/158 lead to results with a 0.5% duality gap. With 25 facilities, the duality gap is around 3% and the CPU times are not reported.

Lee [169] uses KKT conditions to guide decisions about which facilities to open, and cross-decomposition to solve stochastic transportation subproblems. This cross-decomposition exploits the primal and dual structure of the problem simultaneously and reduces the computational difficulty by incorporating Benders' decomposition and Lagrangian relaxation into a single
framework.

5.1.3 Methods to address model $M3$:

In this model, the location of the factories and the allocation of the customers to the factories are taken before the demands are known. The quantities delivered to the customers are second-stage decisions.

Logendram and Terrell [173] consider this problem in the context of price sensitive demands. The influence of price sensitivity is reflected in the mean of the distribution of the demand. Normal and uniform distributions are considered for the demands. The authors propose two optimization methods, a branch-and-bound procedure and an extreme point ranking approach, and a heuristic method that iteratively adds customers and drops plants according to priority rules. Wallace [247] looks at this formulation of the stochastic location problem for the location of fisheries. His solution method is based on a polyhedral decomposition of the second-stage program. Laporte et al. [163] solve this problems to optimality using a branch-and-cut algorithm, the integer L-shaped method of [162]. This algorithm requires that the distribution of the sum of the demands for any subset of customers can be computed. Computational results are reported on problems with 20 to 40 customers, 5 to 10 potential facility locations, and uniformly distributed random demands. CPU times vary between 0.3s and 553.9s on a VAX station model 60.

5.2 The $p$-median problem

The $p$-median problem consists of locating $p$ facilities on a graph so that the travel distance between demand nodes and facilities, weighted by the demand at each node, is minimized.

Frank [98] studies the $p$-median problem when demands at each node are independent random variables. The author seeks optimal locations that maximize the probability that the total travel distance is below a given threshold. For this, the author uses the central limit theorem to approximate the distribution of the sum of demands. He also considers the minimization of the variance of the objective function for isthmus in the graph (i.e., branches of the connected graph which, if deleted, divide the graph into two components) and gives a sufficient condition for the optimal solution to be on a specific branch when the graph is a tree. Frank [99] extends this work to cases when the demands cannot be considered to be independent random variables but the joint vertex demands can be approximated by multidimensional normal random variables. For this, he defines a procedure to enumerate the set of local optimum points, a local optimum being a point either minimizing the variance of the objective function, or maximizing the probability that the objective function value is under a given threshold. Carbone [50] uses
a chance-constrained model to maximize this threshold while ensuring that
the probability that the total travel distance is less than the threshold is
smaller than a factor \( \alpha \). As in [98, 99], Carbone use the central limit the-
orem to approximate the distribution of the total distance with a normal
distribution function and provides conditions under which the problem can
be transformed into an equivalent deterministic convex problem.

Louveaux [175] models the \( p \)-median problem with random demands as a
stochastic integer program. The location and the size of the facilities are the
first-stage decision variables, and the allocation of available service to the
customers are the second-stage variables. In his model, the expected value
factor contains terms representing penalties for unmet demands and trans-
portation costs. Investments and service costs are only considered in a budget
constraint. Louveaux presents several budget constraints formulations.

Mirchandani [191] proposes a model where both the customer demands and
the travel times on any arc are random. Weaver and Church [250] formulate
this problem as an equivalent deterministic linear program when the stochas-
tic parameters are discrete random variables. They propose an approximation
method based on vertex substitution or exchange. The heuristic starts with
an initial solution and exchanges nodes not in the current solution with nodes
in the solution if the substitution reduces the objective function value, i.e.,
the weighted distance. Exchanges are made until no improvements can be
made. The authors propose a second method, a bounding procedure based
on the subgradient optimization of the Lagrangian dual. The Lagrangian
relaxation is performed with respect to the assignment constraints.

Stochastic programs which incorporate an explicit objective function term to
capture risk have been labeled robust optimization models [195, 152]. Approp-
riate risk measures include, for example, minimizing up-side cost deviations
from a specified target and minimizing the cost of the worst-case scenario.

Kouvelis et al. [151] consider the 1-median problem on a tree graph with
random weights and transportation costs. The authors consider two repre-
sentations of the randomness: discrete scenarios and interval specification of
the parameters. They seek a location with a reasonable cost performance as
defined via several different robustness criteria. Polynomial algorithms for
the cases where the data are known via scenarios or a mix of scenarios and
intervals are given. The properties of the problem where all the data are spec-
ified by an interval are also studied. Note that more details on this approach
are presented in [152]. Averbakh and Berman [7] consider the minimax regret
median location on a network with uncertain weights of nodes. The goal is
to minimize the worst-case loss that may occur when the realization of the
weights becomes known. A lower and upper bound on each weight is known,
but the probability distribution of the weights is not. The authors show that this problem is polynomially solvable.

Burkard and Dollani [47] consider the 1-median problem on a network with uncertain or dynamically changing edge lengths and vertex weights. The robust 1-median problem with interval data for the random parameters, where the vertex weights can be negative, can be solved in linear time. Note that the case of negative weights corresponds to obnoxious facilities location, which is also treated by Parouch and Tapiero [201]. Parouch and Tapiero locate an obnoxious facility to minimize its expected impact on a population distributed on a line. They characterize probabilistically the pollution effects of the plant using a stochastic death process. They establish necessary movements on the optimal location by perturbing the population distribution.

Andreatta and Mason [1] look at a related problem, the $k$-eccentricity problem on probabilistic tree with random node weights and random arc lengths. There is a finite number of possible scenarios and the probability that the tree is in a given state is known. The $k$-eccentricity evaluated at a point of a graph is the sum of the weighted distances from that point to the $k$ vertices farthest from it. The absolute $k$-centrum is the set of points (not necessarily vertices) for which the $k$-eccentricity is a minimum. The authors show that the $k$-eccentricity is a convex function and that the absolute $k$-centrum is a connected set and is contained in an elementary path.

5.3 Stochastic transportation problems with random demands

Let us now consider a related problem, the transportation problem with random demand values. The location and transportation problems are related in the sense that once the locations are fixed, the allocation of products to customers is a transportation problem. Ferguson and Dantzig [95] look at the problem of allocating aircrafts to routes when the demands are discrete random variables. They propose a special version of the simplex method to solve the problem. Elmaghraby [88] shows that this method cannot be extended to continuous distributions, and presents an iterative method where a system of nonlinear equations must be solved at each iteration. The decision vector varies in such a way as to satisfy the KKT conditions. Williams [255] develops another iterative procedure, solving a succession of single destination problems. Szwarc [237] replaces the cost function by a linear approximation for problems where the density functions are step functions. Witten and Zimmermann [257] transform the stochastic problem into a deterministic model with partly convex, partly linear cost functions and apply a general algorithm. Wilson [256] presents an approximation technique to find starting solutions for exact algorithms. This technique reduces the problem to a linear deterministic program by estimating deterministic point estimates of the
demands. Shiode et al. [227] propose a two-stage stochastic programming model.

Cooper and Leblanc [63] use the Frank-Wolfe algorithm to solve the stochastic transportation problem. Qi [208, 209] presents a finitely convergent method to solve the stochastic transportation problem based on the network structure of the problem. The graph corresponding to the optimal solution of the stochastic transportation problem is a forest. At each step, a small number of one-dimensional monotone equations must be solved. Qi [210] extends this iterative method to generalized stochastic transportation problems. The graph corresponding to an optimal solution for such problems is an A-forest (each of its connected subgraphs is either a tree or a one-loop tree).

Holmberg and Jörnsten [128] use the Frank-Wolfe algorithm to solve the stochastic transportation problem with random continuous demands. Their method consists of two steps: the solutions of n continuous knapsack problems to derive the search direction, and a one-dimensional search to find the step length. This method can be easily implemented, but becomes slow as the accuracy requirements increase. The authors observe that the problem has a nonlinear separable objective function and linear constraints, and thus also propose a separable programming method. This method is fast, but has the disadvantages that one cannot determine the accuracy exactly before solving the problem and that a lot of computer memory is needed. They thus develop a third method, based on cross-decomposition [217], which combines the advantages of the first two methods, with speed and presetting of the desired accuracy. Nevertheless, this method cannot be used if non-accurate solutions are desired. There are two ways to separate the problem: a resource decomposition, where the amount shipped to the destinations are the complicating variables and can be computed by solving a Benders' master problem, and a price decomposition, for which the subproblems decompose into simple continuous knapsack problems and a bounded classical inventory problem. Prices for the latter decomposition can be obtained by solving a Dantzig-Wolfe master problem. The cross-decomposition method iterates between the two decomposition schemes until optimality is achieved or a convergence test fails. If a convergence test fails, one of the master problem is solved. This method is strongly linked to the structure of the problem.

Prékopa and Boros [206] search for conditions for the feasibility of the stochastic transportation problems on a network with random demands and arc capacities. They propose a method to find lower and upper bounds on the probability that a feasible flow exists. The method consists of first removing all redundant inequalities and computing the probabilities that sets of inequalities are satisfied. The upper and lower bounds are obtained by solving two linear programming problems where these probabilities are right-hand
side values.

Holmberg [129] compares several methods based on decomposition and linearization techniques: a separable programming approach, the Frank-Wolfe decomposition approach, the mean-value cross-decomposition, the Lagrangian relaxation with subgradient optimization, and combinations of these methods. Cheung and Powell [60] approximate the expected recourse function of a distribution problem by a convex, piecewise and separable function using the tree structure of the problem.

5.4 Capacity expansion under uncertain demand

The capacity expansion problem consists of choosing the size, location and timing of the facilities needed to expand the operations of a system. The objective is to maximize, say, the system's profit. For such problems, the uncertainty influences the investment, production and pricing decisions of the firms, the technological progress in the domain, and the demand. Capacity expansion problems occur, for example, in the context of electric power generation facilities, manufacturing and processing facilities, water resource projects and communication network planning. For a review of deterministic capacity expansion problems, see [180].

Two approaches can be found in the literature to address capacity expansion with uncertain demands. The first approach models the problem as a stochastic process. These models are then analyzed as stochastic processes (e.g., [70]) or transformed into deterministic equivalent models (e.g., [105]). For more details on this approach, see [16, 48, 104, 113, 185, 239]. The second approach, which we consider in this review, is to formulate capacity expansion problems as stochastic programming models. The effect of discretizing the planning period in time stages is studied in [199], where a sensitivity analysis on the impact of the discretization pattern on the optimal cost is presented.

Leondes and Nandi [171] solve capacity expansion problems in networks with uncertain discrete demands. The authors use a two-stage stochastic nonlinear program. They assume that the cost of expanding the capacity of each arc is a convex function and that there is a concave salvage value associated with excess capacity. Arc capacities are expanded in the first stage. When the demands become known, corrective actions are taken in the second stage to further increase or reduce capacity on the arcs. The authors propose an iterative deterministic method to solve that problem which uses network algorithms for the first-stage program. They suggest using sampling to extend the algorithm to continuous distributions.

Sherali et al. [225] consider marginal cost pricing in a capacity expansion problem for the electric utility industry under a finite and discrete stochastic
demand forecast. They develop a two-stage stochastic linear program with recourse to determine a marginal cost pricing strategy for sharing capital costs given an optimal capacity plan. They want to select a mix of equipment, which minimizes the total annual capital and operating cost. The first-stage decisions are the capacities for the different equipment types. The second-stage recourse consists of allocating the equipment to the demand. The objective function for the second stage is the fuel cost, and thus the problem decomposes into transportation problems for fixed first-stage decisions. To solve this program, the authors construct an expected load curve and use a result from [197] stating that the optimal solution resulting from this expected load curve provides a selection of equipment which is also optimal for the stochastic program. They then solve the transportation problems to allocate the equipment using the Northwest corner rule.

Eppen et al. [89] consider capacity problems where the demand can take three values, corresponding to pessimistic, neutral and optimistic realizations of the demands. The locations and the configurations of the plants must be chosen before the demands are known. The model maximizes the expected discounted cash flows subject to a linear constraint on risk. The authors use a standard MIP software to solve this problem.

Leopoldino et al. [172] model capacity expansion in the Brazilian power system as a stochastic integer program. The objective is to minimize investments in generating units and interconnection links, subject to constraints on supply reliability. The loads of the different regions and the availability of the generation units and of the transmission links are assumed to be random and to take a finite number of values. The authors solve this problem using Benders' decomposition. The master problem is an integer program solved by complete enumeration. The subproblems correspond to stochastic network flow problems, and are solved by a maximum flow algorithm and Monte Carlo simulation.

Berman et al. [26] study this problem when the demands are random and take a finite set of values. A scenario specifies the values of the demands at each location for each time period. Because of the small number of scenarios, the problem can be solved using traditional linear programming methods on the deterministic equivalent of the stochastic problem. The authors develop a more efficient approach using a Dantzig-Wolfe decomposition scheme. They show that the problem is a set of separable linear programs linked by a budget constraint. Each linear program has a special nested knapsack structure, and is thus efficiently solvable. They present computational results on problems with five time periods and five locations, and 5 to 100 demand scenarios. Run times on a 386-based PC are of the order of 0.1 minute for 5 scenarios, and 5 minutes for 100 scenarios.
Wagner and Berman [246] build on the model of [26]. They modify it so that the sales can be expressed in units different from those used for capacity. They study how different assumptions, about the timing of the decisions versus the time at which the demands become known and about the inclusion of the randomness in the models, lead to different expansion strategies. They conclude that the more flexibility the decision maker has to react to changes in the demands, the higher the expected profits. The authors present numerical results on two instances of the problem: a problem with a single location, two time periods and three scenarios, which can be solved using stochastic dynamic programming, and a problem with up to ten stores, five time periods, and a total of 243 possible demand scenarios, solved using the method of [26].

Marin and Salmeron [189] present a stochastic model for electric capacity expansion planning under uncertain continuous demands. Benders' Decomposition and Lagrangian relaxation methods are used to solve a nonlinear continuous model of this problem. The authors provide computational experience and compare their method to general purpose optimization package. Dentcheva and Røisland [74] present stochastic programming models whose constraints are loosely coupled across operating power units. The authors use this property to decompose the problem into stochastic single unit sub-problems using Lagrangian decomposition methods.

Dantzig et al. [66] consider a capacity expansion planning model in the electric utility industry. They argue that if capacity expansion plans over time must be committed to sufficiently far in advance then the multi-stage optimization problem collapses to a two-stage problem. Their model has stochastic equipment availability and random demand, and they propose solving it by using Monte Carlo sampling, with a variance reduction technique called importance sampling, within a Benders' decomposition algorithm.

Sen et al. [224] consider a capacity expansion planning problem on a communications network. First-stage decisions install capacity in the network. Random demands for point-to-point pair connections are then revealed and second-stage variables optimally route calls in this multicommodity network. The model is solved via the stochastic decomposition algorithm which incorporates sampling in a Benders' decomposition scheme [122].

Paraskevopoulos et al. [200] adopt a robust optimization approach that penalizes the sensitivity of the objective function to various types of uncertainty in the demand. For this, the objective function is augmented by a penalty term measuring the relative importance of uncertainty among the various error terms. This term is formulated using a risk-aversion parameter and the variance-covariance matrices of the error, forecast and parameter esti-
mation errors. The aim is to cast the problem in a deterministic framework to avoid the complexity of nonlinear stochastic programs. Numerical results show that as caution against the demand uncertainty increases, the variance of the objective function decreases. The cost of this robustness is a deterioration in the deterministic risk performance. The method is applied to a west European PVC industry for a ten-year period. The model has 143 decision variables, 33 linear constraints and 23 nonlinear constraints. Results show that increasing caution towards risk leads to excess capacity.

Malcolm and Zenios [184] develop a robust optimization model that generates capacity expansion plans which are both solution and model robust (i.e., the optimal solution is almost optimal and has reduced excess capacity for any realization of the demand scenarios). They generalize the model of [197]. The objective function of the model is composed of three terms: the expected cost of the system over all possible scenarios, the variance of the cost, weighted by a parameter \( \lambda \), and a term penalizing deviations from feasibility, weighted by a parameter \( \mu \). A measure of the solution and model robustness is obtained by varying \( \lambda \) and \( \mu \) and observing the changes in expected value and expected infeasibility. This model is equivalent to minimizing expected costs when \( \lambda \) and \( \mu \) are both set to zero. The authors compare the robust model solutions to the cost-minimizing solution and show that, for the problem instance they are considering, the latter solution has a lower expected cost, but results in idle capacity for some scenarios. Numerical results are given on instances with four possible supply options and four demands scenarios.

6 Stochastic vehicle routing problems

We now review stochastic routing problems with random demands. Two types of randomness are considered in the literature, and each of them has its own set of recourse strategies and methods. When the presence of the customer is random, vehicles follow a priori routes as planned and skip any absent customer. When the vehicles are capacitated and customer demand is random, a vehicle might be incapable of satisfying the sequence of demand realizations along its a priori route. In this case, a route failure is said to occur, and the vehicle then returns to the depot to load or unload, and resumes its route at the last visited customer. These problems are now presented in detail. They can be combined when both types of randomness appear in a problem [107, 108]. For a review of routing problems and stochastic routing problems, see [165, 166] and [109], respectively. For a review of stochastic dynamic routing problems, see also [139]. We also present in this section a related problem, inventory routing. In this problem, customers keep a local inventory of the considered commodity. The goal is to minimize long-term costs by scheduling the timing of the deliveries as well as the quantities delivered and the routes of the vehicles.
6.1 Vehicle routing problems with stochastic customers

One type of demand uncertainty concerns the presence of the customers, who generate binary demands with a given probability. Bartholdi et al. [13] look at the problem of designing meals-on-wheels routes, and use space-filling curves to generate solutions. The problem is more formally introduced by Jaillet [133] and is known as the probabilistic traveling salesman problem (PTSP). An a priori solution visiting all the customers is first defined. Once the set of customers requiring service is known, this solution is adapted to this realization of the random vector by skipping customers not present. The a priori solution is chosen to minimize the weighted average, over all possible realizations, of the objective function values obtained by applying the recourse strategy to each realization. There are two motivations for studying this problem. First, it is often impossible to reoptimize routes every day since the optimization process can be computationally intensive and information may be obtained only at the last minute. Second, even if reoptimization can be performed every day, it may not be desirable because it is too expensive to do so or because one wishes to eliminate unmodeled inefficiencies that could arise from rerouting vehicles on a daily basis. The driver's knowledge of the tour may also be a key for good service level (e.g., postal routes).

Jaillet [133] shows that the a priori solution obtained by solving a deterministic TSP can be arbitrarily bad, describes mathematical models, and gives a closed-form formulation of the objective of the PTSP which can be computed with \(O(n^2)\) work, where there are \(n\) cities. He also studies theoretical properties of the models. For example, an optimal solution to the PTSP defined in a plane may cross itself, contrary to what happens for the TSP.


Heuristics have been proposed to solve the PTSP. These include nearest-neighbor criterion or savings criterion such as that used in the Clarke-Wright heuristic [62, 140, 216], space filling curves [12, 13, 29], and 2-opt edge interchange [29]. Laporte et al. [164] apply an integer L-shaped method and solve to optimality instances involving up to 50 vertices.

A direct extension of the PTSP is the stochastic vehicle routing problem, PVRP, where customers are present with a certain probability and have deterministic, but not necessarily binary, demands. Bertsimas [34] proposes two adaptation strategies for the PVRP. The first strategy consists of visiting all the vertices according to the a priori tour, but to serve only customers that require service. The second strategy is similar to the first one, but the clients
with no demand are skipped. The author finds closed-form expressions of the objective value of the PVRP under both strategies, and proposes algorithms to compute these values under general probabilistic assumptions, as well as upper and lower bounds on their values. He also proposes using the cyclic heuristic [114] to address the PVRP, and analyzes the worst case performance and average behavior of this heuristic. The cyclic heuristic consists of considering special sequences of the nodes of the form \((0, i, \ldots, n, 1, \ldots, i - 1, 0), i = 2, \ldots, n\), and choosing the sequence which leads to the smallest expected a priori tour length. Bertsimas et al. [38] report computational results for several graph-based heuristics for the PTSP and PVRP presented in [34, 35] (space-filling curves and radial sorting, combined with 2-interchange local improvement or 1-shift local improvement for the PTSP, cyclic heuristic, Clarke-Wright and clustering heuristic for the PVRP) on ten (PTSP) and twenty (PVRP) instances of fifty nodes with different demand probabilities. For the PTSP, the nodes are randomly generated over the unit square. For the PVRP, the demands follow discrete or normal distributions. The performance are compared to a posteriori methods, i.e., optimizations performed after the random variables are known. Results indicate that the radial sorting with 2-opt or 1-shift procedures give “adequate” results for the PTSP, and that the cyclic heuristic performs well for the PVRP.

Berman and Simchi-Levi [24] treat the PTSP when the probabilities generated at the various customers are allowed to be different from each other. They derive a lower bound on the value of the objective function which can be obtained by solving a transportation problem and suggest using this bound in a branch-and-bound solution method.

The a priori optimization scheme has been extended to other optimization problems, more generally known as probabilistic combinatorial optimization problems, such as the probabilistic maximum independent set [196] and the scheduling and bin packing [19] problems. For a survey of probabilistic combinatorial optimization, see Bellalouna et al. [20].

### 6.2 Vehicle routing problems with stochastic demand values

For the second type of routing problems with stochastic demands, all the customers are present, and the value of the demand at each customer location is the only source of uncertainty (SVRP). Penalty-based and chance-constrained models without recourse decisions are presented in [14, 78, 82, 83, 112, 160, 231, 232]. Dynamic variants of the problem, which include planning horizons, are presented in [100, 203].

Golden and Stewart [232] transform the probabilistic constraint of a chance-constrained model into nonlinear deterministic constraints by explicitly writ-
ing the probability when the random variables are independent and the sum of the demands on a route has the same distribution as each individual demand. If the variance of the random variables is a constant multiple of their means, the deterministic constraints are linear. The problem is then equivalent to a capacitated VRP and classical VRP methods can be used. A modified Clarke-Wright heuristic and the generalized Lagrange multipliers method (where the capacity constraint is moved to the objective function and a 3-opt branch exchange procedure is used to find local minimum for the resulting problem) are used to solve both chance-constrained and penalty models on problems with 50 to 75 customers. Dror and Trudeau [78] illustrate the effects of route failure for the same problems and study the impact of the direction of a designed route. Bastian and Rinnooy Kan [14] assume that the demand at a node becomes known only when the vehicle arrives at the location, and require the vehicle to follow a fixed sequence in serving the customers. They describe three models, a chance-constrained model, a penalty model, and a full-service model, where the vehicle returns to the depot to (un)load whenever it is empty (full). They show that, if the demands are independent and identically distributed random variables, the three models reduce to a single-vehicle version of the time-dependent TSP, which is then solved using dynamic programming.

Dror [82] presents a multistage stochastic model for the SVRP and describes a graph representation of the model in which the solution of the SVRP corresponds to a Hamiltonian cycle. To form this graph, one must duplicate \( n - 1 \) times each customer location and \( n \) times the depot node. The \( n - 1 \) nodes representing a customer are interconnected by arcs of zero distance and connected to other nodes by arcs with the same distance as in the original graph. The \( n \) depot nodes are interconnected with arcs of zero distance and connected to the other nodes by arcs from the original distance matrix. Dror also describes a Markov decision process model representing the SVRP.

Dror et al. [83] consider SVRPs in which potential failures can occur only at one of the last \( k \) customers on a route. The chance-constrained model can then be solved as a deterministic VRP in which routes with more than \( k \) potential failures are eliminated. Under the assumption that only one route failure can occur, they develop three recourse policies: no reoptimization, reoptimization of the customer sequence after failure, and planning of a preventive break or risking a failure. They show that an optimal solution for these three models can be obtained by solving in a branch-and-bound manner a series of TSPs.

Laporte and Louveaux [160] formulate recourse models that depend on the timing of when the demands (or more exactly the supplies in their application) become known. If the realizations of the supplies become known before
the vehicles start their routes, breaks in the routes, allowing a vehicle to return to the depot to unload, can be scheduled at convenient times to avoid route failures. If the realization of a supply value becomes known only when a vehicle arrives at the customer location, the vehicles follow their a priori routes. When a vehicle's capacity is exceeded, it returns to the depot to unload and resumes its route from the last visited customer. Because these models are considered to be computationally intractable, the authors develop upper and lower bounds for the problem based in the solution of related simpler vehicle routing problems. Teodorović and Pavković [241] solve the latter model using a technique based on simulated annealing. Popović [202] extends their work by using a Bayesian approach to deal with the uncertainty when the distributions of the random demands are not known.

6.3 Vehicle routing problems with stochastic customers and demand values

This section combines the two problems described previously: both the customers' presence and the values of the demand at the customer locations are random. Gendreau et al. [107] propose an exact algorithm for this problem, based on the integer L-shaped method, and solve to optimality instances with 10 to 70 customers. Gendreau et al. [108] develop a tabu search approach to address larger instances. The heuristic is tried on instances with 6 to 46 customers and compared to optimal solutions obtained using the integer L-shaped method. For these instances, the probability of presence of each customer is uniformly distributed over [0, 1]. The random demand values are defined as follows: the nodes are assigned to one of three ranges of values, in equal proportions, and a value of the demand is generated in the appropriate range according to a discrete uniform distribution. Results indicate that the heuristic generates an optimal solution in 89.45% of the cases, with an average deviation of 0.38% from optimality. In 97% of the cases, the deviation from optimality is smaller than 5%.

6.4 Stochastic Inventory Routing Problem

The inventory routing problem is a distribution problem in which each customer maintains a local inventory of a product and consumes a certain amount of that product each day. The objective is to minimize the long-term delivery costs by timing the inventory replenishment and the delivery routes. Early deliveries to a customer increases the delivery costs in later time periods because the customer will have to be replenished sooner. Late deliveries increase the risk of paying a stockout penalty. For an industrial application of the inventory routing problem to the heating oil industry, see [18]. Note that a variant of this problem, called the strategic inventory routing problem [167, 251], focuses on estimating the minimum cost vehicle fleet
required to supply the inventories.

Although the demand at each customer location is a random variable, most of the literature studies the problem in a deterministic fashion, fixing the demands at their expected values \([18, 80, 167, 243, 251]\). A first attempt to consider random demands is presented in \([94]\). Federgruen and Zipkin \([94]\) consider a myopic version of the stochastic inventory routing problem which does not consider long-term but single-period planning. The inventory at each location at the beginning of the period is known. This information is used to determine the allocation of product among the customers for the day. After the deliveries are made, the demands are realized, and holding and shortage costs may be incurred at each location. The authors propose a simple recourse model and show that, for fixed allocation of customers to routes, the problem decomposes into inventory allocation and traveling salesman problems. They present numerical results with normally distributed demands and 50 to 75 locations. The allocation problem is solved using the algorithm of \([93]\) and the routing problems are solved using a modified interchange heuristic.

Long-term planning with random demands is addressed in \([81, 242]\). In these papers, the demands become known when the delivery vehicle visits the customer, and one must decide at the beginning of the operations time horizon whether or not to replenish a customer. If a customer stocks out before his planned replenishment time, he is serviced by an emergency service, for which the delivery company incurs extra costs.

Dror and Ball \([81]\) present a procedure to reduce the long-term version of the problem to a single-period problem, which can then be addressed using standard algorithms. The reduction procedure involves the definition of single-period costs that reflect the long-term costs, the definition of safety stock level, and a specification of the customer subset to be considered during a single period. The assignment of customers to days of the time period is done using an LP-based generalized assignment algorithm. Vehicle routing problems are then solved for each day with a modified version of the Clarke-Wright heuristic, and the resulting solutions are improved with interchange and inter-customer exchanges. For more details on these three steps, see \([80]\). Note that this solution method can have an extra initial step, selecting customers in function of a calculation of the "best" replenishment day of each customer \([79]\). Dror and Ball consider identically distributed random demands and look at the expected cost of the tradeoff between late and early deliveries. They show that if the distribution of the customer's consumption on a given day is normal, the expected cost function is convex. Kreimer and Dror \([153]\) extend this result for a number of other distributions.
Trudeau and Dror [242] consider continuous distributions of the customer demands. They improve upon the solution method of [79] by proposing a more robust method to select customers and developing a new algorithmic procedure for the routing problem solutions and the improvement of the resulting solutions with interchange algorithms. The authors present computational experiments on a twelve-week period with 2077 customers and 2 vehicles. The demands are normally distributed random variables, and generated by simulating their values using their estimated distribution parameters.

Bard et al. [11] look at the stochastic inventory problem with satellites facilities. Satellite facilities are locations other than the central depot where vehicles can be refilled (see [131]). The authors present a methodology to decompose the problem over the planning horizon and then solve daily rather than multi-day VRPs.

Campbell et al. [49] investigate two approaches to the inventory routing problem: a dynamic programming approach, using a discrete Markov decision process, and an two-phase integer programming approach. In the first phase, the authors determine when and how much to deliver to each customer on each day of the planning period. In the second phase, they determine sets of delivery routes for each day. This paper includes general observations on the problem and an interesting literature review on the subject. Kleywegt et al. [149, 150] also use a Markov decision process to model the inventory routing problem and propose approximation methods to find good solutions with reasonable computational efforts.

7 Stochastic location-routing problems

Location-routing problems consist of simultaneously locating a depot among a set of potential sites, determining the fleet size, and designing vehicle routes to visit customers. For a survey of such problems, see [158].

Laporte et al. [159] consider two one-vehicle stochastic location-routing problems. The first model minimizes the total cost, i.e., the sum of the depot operational cost, the vehicles’ fixed costs and the routing cost, while ensuring that the probability of route failure does not exceed a given threshold. The second problem minimizes the total cost in such a way that the expected penalty cost does not exceed a given fraction of the planned route length. The decisions must be taken before the demands are known. The authors propose two corrective strategies to modify this initial solution, which depend on the timing of the realizations of the demands. If the demands become known only when the vehicle arrives at the node, the vehicle follows its priori route until a route failure occurs. When a failure occurs, it proceeds to the depot to unload, returns to the same customer location and resumes its route. If
the demands are known after the decisions have been taken but before the vehicle starts its route, the vehicle does not proceed to a customer \( \gamma \) if it is known that a route failure will occur. Instead, it returns to the depot from the previous customer location and resumes its routes at \( \gamma \). The authors define the penalty cost of each strategy. For this, they assume that the supplies at the node are independently and identically distributed random variables and that the probability distribution of the sum of supplies on any subset of nodes can be computed or approximated. The problems are modeled as stochastic integer programs and solved to optimality using a branch-and-bound scheme. Computational results on problems with 10 to 30 nodes, 2 to 3 potential depot locations, and normally distributed demands are presented.

Averbakh et al. [5] take an a priori optimization approach, and consider separately the routing and location problems (i.e., they consider the routing problem when the location has already been chosen and the location problem when the routes have already been defined). They propose to solve the routing-location problem by solving a routing problem for each possible location and choosing the best location and the corresponding a priori tour. They consider five different objective functions, functions the total length of the tours and the waiting times of the customers. The contributions of this paper are the explicit expressions of these objective functions and efficient ways of calculating them.

The probabilistic traveling salesman location problem consists of finding a location and an a priori tour to minimize the expected distance traveled to service customers from this location. The customers randomly generate a request, with a known probability. If the triangle inequality holds, any node which always generates a demand (wp1) is an optimal location for the facility [28]. Berman and Simchi-Levi [24] prove that locating the facility at one of the nodes is optimal. They propose an \( O(n^3) \) algorithm to find the optimal location if the a priori tour is given [24, 29]. Bertsimas [29, 31] characterizes the worst-case performance of a nearest-neighbor location heuristic and of space filling curves. Results for this problem are reviewed in [27].

8 Concluding remarks

This paper presents the literature on stochastic location and routing problems with random demands. The main issues in addressing these problems are the following:

1. Are the distributions of the random parameters continuous or discrete? If the set of possible realizations of the random vector is small, the problem can be transformed into a deterministic model, and solved using (adapted versions of) traditional methods. If not, approximations or Monte Carlo sampling procedures should be used. In the case of
Monte Carlo sampling, this means increasing the number of constraints and variables, and the number of times the resulting problem should be solved.

2. *Where are the random parameters?* When the only source of uncertainty is the demand, the randomness is located in the right-hand side of the model. This structure may be useful, since the same problem may have to be solved several times with only small variations on the RHS coefficients. Efforts have been made to exploit this property with Gröbner bases methods when the demands are discrete random variables, but the method is too computationally demanding for large problems to be solved.

3. *What is the timing of the decisions with respect to the release of information?* Different models represent different settings.

4. *Is the timing of the decision stages fixed or random?* The examples considered in this paper assume fixed time decision stages.

5. *Where are the integer variables, if any?* The problem is easier to solve if there are no integer variables in the second stage (or subsequent stages) because the properties of the recourse function, such as convexity, are then maintained.

6. *Which recourse procedure can be implemented?* This question is both practical (i.e., what can be implemented for that problem) and computational (what can be solved).

7. *How many times does the problem have to be solved?*. The problem may need to be optimized once (i.e., locate a plant), or everyday, on a different set of data (i.e., the routes of a postal delivery truck). In the latter case, a solution is defined a priori, and a strategy is given to update the a priori solution to the instance at hand whenever the solution has to be applied.

*Acknowledgements:* The authors thanks an anonymous referee for insightful comments.

**References**


[88] Elmahraby S (1960) Allocation under uncertainty when the demand has a continuous distribution function. Management Science 7, 270-294


[131] Huang L (1997) The Inventory Routing with Satellite Facilities Problem, Ph.D. Dissertation, College of Business Administration, University of Texas, Austin


[177] Louveaux FV (1993) Stochastic Location Analysis. Location Science 1, 2:127-154


[236] Sturmfels B (1996) Gröbner bases and convex polytopes, University Lecture Series, 8, American Mathematical Society


