Robust capacity expansion solutions for telecommunication networks with uncertain demands

F. Babonneau[†], O. Klopfenstein[*], A. Ouorou[*], J.-P. Vial[†]

August 2010

Abstract

We consider the capacity planning of telecommunication networks with linear investment costs and uncertain future traffic demands. Transmission capacities must be large enough to meet, with a high quality of service, the range of possible demands, after adequate routings of messages on the created network. We use the robust optimization methodology to balance the need for a given quality of service with the cost of investment. Our model assumes that the traffic for each individual demand fluctuates in an interval around a nominal value. We use a refined version of affine decision rules based on a concept of demand proximity to model the routings as affine functions of the demand realizations. We then give a probabilistic analysis assuming the random variables follow a triangular distribution. Finally, we perform numerical experiments on network instances from SNDlib and measure the quality of the solutions by simulation.

Keywords. Capacity expansion problem, Telecommunication networks, Robust optimization.

1 Introduction

Two conflicting criteria guide operators when they design or reinforce telecommunications networks: the costs of investment and the quality of service (QoS) (expressed here as a level of rejected requests). On an existing network, failure to achieve full service is caused by an excessive traffic load in conjunction with insufficient installed capacity. When facing uncertain traffic demands, the network designers may be tempted to over-estimate the traffic demand in order to achieve the required QoS. This is likely to result in over-sized networks entailing unnecessarily large investment costs. Our goal is to show that an approach based on robust optimization is able to achieve the required QoS at a lesser cost.

The need for new capacity occurs when the telecommunications operator offers new services or has to cope with a growing traffic. For strategical and operational reasons those investments must be planned well in advance and decisions must rely on traffic forecasts, which, by essence, are uncertain. Analysis of real-life data have revealed large errors between marketing forecasts and the actual future traffic in the network. Even at a one year horizon term, the difference can easily be larger than 10%. It is thus often the

[*]Orange Labs, R&D, France.
[†]Ordecsys, place de l'Etrier 4, CH-1224 Chêne-Bougeries, Switzerland.

case that when the planned network becomes operational, it turns out to be inadequately sized. Capacity adjustments must be made, at possibly much higher costs.

The capacity planning problem for telecommunications networks is essentially a multicommodity flow problem. Its objective involves the investment cost and possibly some cost associated with the traffic routing (e.g. travel times). Additional constraints to ensure traffic diversification may also be included. When demands are uncertain, the adequate formulation is a two-stage stochastic problem. In the first stage, capacities are allocated (deterministic variables) and routings to meet the uncertain demands are set in a second stage. Routings are thus stochastic by construction. This problem is a special case of more general formulations of network design under uncertain demands and possibly uncertain travel times with a single commodity and multiple sources and sinks or with multiple commodities with single source and a single sink for each commodity.

In the present paper we concentrate on capacity expansion for the multicommodity flow problem with capacity costs only and uncertain demands. Therefore, we do not review the literature on other related problems, unless the contributions are relevant to our problem. A first issue concerns the modeling of uncertain demands. Some contributions use a scenario based description with the condition that for each demand scenario there is a feasible flow that meets the demand and the capacity constraint on the installed capacities [17, 19, 20, 10, 21]. A solution that satisfies this requirement is claimed to be robust. In view of the linearity of all constraints, a solution that is robust with respect to all scenarios is also robust with respect to the convex hull of the scenarios. Therefore the demand uncertainty can be described via a polyhedral uncertainty set. This idea has been specifically exploited in [2, 9], where the set of traffic events is defined through inequalities rather than from a set of discrete scenarios. More general convex-compact uncertainty sets are considered in [4] (see also [5]). Those sets, just like the scenario-based uncertainty sets, do not cover the whole of possible demand realizations. A solution is claimed robust if for each demand in the uncertainty set there is a routing that meets the demands and is compatible with the installed capacity. A third alternative is to describe uncertainty via probability distribution and use a chance-constrained programming formulation [12]. It is known that the feasible set with probabilistic constraints is in general not convex, and possibly disconnected, a fact that also leads to considerable numerical difficulties.

The second issue concerns the handling of the routing decisions that are recourse decisions adjustable to meet the observed demands. As pointed out in [7, 3] fully adjustable solutions are liable to make the computation of robust solutions intractable, a fact that is recognized in [20, 14]. The get around intractability reference [7] propose a revival of the concept of linear decision rules (or Affine Decision Rules, ADR). In that framework, recourse decisions are formulated as affine functions of the observed demand uncertainties (see [18]). In our problem of interest when the number of demands is large this option can lead to a significant increase of the number of variables. In the present paper, we use a refined version of ADR based on a concept of demand proximity to overcome this difficulty.

The final issue deals with the handling of constraint violations. They can be represented either by individual 0-1 indicators (depending on whether or not the violation occurs) or as a the amount of unserved demands. Individual 0-1 indicators are useful in chance-constrained programming or in standard robust optimization. The latter case enables the measure of the the total amount of unserved demands; it can be used as objective function [17, 20], but this formulation unfortunately leads to intractable problems.

The current paper builds on an earlier work [18] of two of the present authors. The deterministic version of the model is a path-flow formulation. A set of admissible paths is proposed as part of the data. The number of paths used by the solution may be restricted to a fixed cardinality subset of the admissible paths. Individual demands are modeled as independent[1] uncertain variables. Each one varies around a nominal value on a symmetric interval. First stage decisions deal with the investment for capacity, while the recourse decisions are the actual routings on the network to meet the observed demands. Recourse decisions are modeled as Affine Decision Rules[2] (ADR). The uncertainty sets are polyhedral. We chose not to use ellipsoidal sets in order to remain in the realm of linear programming and make it easier the handling of first stage decision variables with integrality restrictions. We impose that the ADR's be such that the demand requirement is satisfied for all demand scenario within the uncertainty set. However ADR's spread the uncertainty into the capacity constraints and the flow nonnegativity constraints. Those two sets of constraints are robustified. To quantify the immunization factor in the equivalent robust counterpart, we perform a probabilistic analysis under the assumption that each demand is distributed according to a triangular symmetric distribution. A weaker assumption could be made, that would very similar, though weaker, results. For a detailed study, we refer to [4, 5, 3]. By construction, a robust solution tends to be conservative for three main reasons: i) ADR are suboptimal; ii) the immunization factor that guarantees a certain probability of satisfaction is over-estimated by the theoretical analysis based on probabilistic inequalities; iii) the indicators of constraint violation are not independent random variables from constraint to constraint. Consequently, the practical solution must be evaluated empirically, using simulations. This complementary study usually confirms that the robust solution is conservative, but also behaves consistently well.

The paper is organized as follows. In section 2 we present the basic capacity expansion problem. In section 3, we propose a model for the uncertainty on the demands and give the formulation of affine decision rules for the recourse. Section 4 formulates the robust counterpart of an uncertain constraint and collects few known results in robust optimization. Section 4 concludes with a probabilistic analysis with the triangular distribution. In section 5, we model the capacity planning problem with robust constraints. Section 6 is devoted to numerical results on test problems. The QoS achieved by the robust solutions is observed by simulation on a Monte-Carlo sample of demand scenarios, assuming a triangular distribution. Section 7 is a conclusion with some hints for further research.

2 Capacity expansion problem

Let $\mathcal{G}(\mathcal{N}, \mathcal{A})$ be a directed graph where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs. We denote by $\mathcal{K}$ the set of demands, characterized by origin-destination (OD) pairs. Let d_k be the traffic amount for demand $k \in \mathcal{K}$, and let $\mathcal{I}_k$ denote the set of available paths that can be used to route the demand k. This set $\mathcal{I}_k$ may contain all paths between the origin and the destination of demand k. However, it is often desirable, or even required, to restrict the number of potential paths for a given demand: indeed, not all the paths are acceptable by network managers; furthermore, from a computational viewpoint, this restriction will be helpful.

The path-flow formulation of the capacity expansion problem with linear expansion

[1]This choice is dictated by the situation that is prevalent in practice. More sophisticated models, involving, for instance, factors and correlations, can equally be handled by robust optimization. However, it turns out that, in our problem of interest of new telecommunications services, the influence factor matrices and the correlation matrices are generally not available or not even computable.

[2]or affinely adjustable variables in the parlance of [7],

cost is:

$$\min_{f\geq 0, c\geq 0} \quad \sum_{a\in\mathcal{A}} l_a c_a \tag{1a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k} \pi^a_{ik} f_{ik} \leq c_a, \quad a \in \mathcal{A} \tag{1b}$$

$$\sum_{i\in\mathcal{I}_k} f_{ik} = d_k, \quad k \in \mathcal{K}. \tag{1c}$$

In that formulation l_a is the cost of installing a unit capacity on arc a. The vector π_{ik} has component $\pi^a_{ik} = 1$ if path $i \in \mathcal{I}_k$ includes arc a, and 0 otherwise. The flow for the traffic demand k on the path i is represented by the variable f_{ik}. Finally, variable c_a is the capacity to be installed on the arc a.

In this problem, the number of admissible paths is restricted and is given in explicit form. Practitioners usually add the further restriction that only a small number of those admissible paths can be effectively used in the proposed solution. Let m_k be this number for the demand k. To deal with this restriction, we add binary variables and reformulate the original problem as the mixed integer linear problem (MIP)

$$\min_{f,c,y} \quad \sum_{a\in\mathcal{A}} l_a c_a \tag{2a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k} \pi^a_{ik} f_{ik} \leq c_a, \quad a \in \mathcal{A} \tag{2b}$$

$$\sum_{i\in\mathcal{I}_k} f_{ik} = d_k, \quad k \in \mathcal{K} \tag{2c}$$

$$\sum_{i\in\mathcal{I}_k} y_{ik} \leq m_k, \quad k \in \mathcal{K} \tag{2d}$$

$$f_{ik} \leq M_k y_{ik}, \quad k \in \mathcal{K},\ i \in \mathcal{I}_k \tag{2e}$$

$$f \geq 0, c \geq 0, \tag{2f}$$

$$y_{ik} \in \{0,1\}, \quad k \in \mathcal{K},\ i \in \mathcal{I}_k. \tag{2g}$$

The binary variables y indicate whether a path is opened or not. The parameter M_k is an upper bound on the flow f_{ik}. One can take $M_k = d_k$ or a tighter one if available. Constraints (2d) limit the number of active paths.

One of the major problems of this model lies in the values of traffic demand d to be used. In practice, network operators use statistical models, together with market surveys, to forecast the evolution of traffic. However, experience shows that forecasts are always wrong, and often far from the observed reality. Traffic measurements on a backbone network of France Telecom, compared with the amounts forecasted one year before, have revealed gaps up to 25% on the global amount of traffic in the network (i.e., $\sum_{k\in\mathcal{K}} d_k$), depending on the year. Note that the error may be positive, as well as negative, around the expected value. If no probablistic assumption is performed, this error of 25% should be reported on all demands $\{d_k\}_{k\in\mathcal{K}}$.

Under some circumstances, the error on individual demands may be far larger than that. Indeed, given the successive errors performed year after year, we estimate the average error and its standard deviation. In our data, we have observed a standard deviation of 17% for the error on $\sum_{k\in\mathcal{K}} d_k$. For the sake of simplicity, let us assume that the error process is Gaussian, and that demands $\{d_k\}_{k\in\mathcal{K}}$ are independent Gaussian random variables, with standard deviations $\sigma_k = c \cdot \mathbb{E}[d_k]$, for some $c > 0$. Then we have: $\left(17\% \cdot \mathbb{E}\left[\sum_{k\in\mathcal{K}} d_k\right]\right)^2 = \sum_{k\in\mathcal{K}} (c \cdot \mathbb{E}[d_k])^2$, which implies:

$$c = 17\% \cdot \frac{\sum_{k\in\mathcal{K}} \mathbb{E}[d_k]}{\sqrt{\sum_{k\in\mathcal{K}} \mathbb{E}[d_k]^2}}.$$

The relative error c of each individual demand d_k may thus be far larger than the relative error on the global traffic amount (c could reach $17\% \cdot \sqrt{|\mathcal{K}|}$, if all $\mathbb{E}[d_k]$ are equal).

Finally, the above quantitative analysis is based on forecasts performed one year in advance. Of course, the uncertainty is much larger if forecasts are performed earlier: even only 2 years in advance, we observed errors more than 40% on the global traffic amount. Notwithstanding, network operators usually perform network plans several years in advance.

The models developed in this paper aim at tackling this uncertainty issue.

3 Uncertainty and affine decision rules (ADR)

The capacity investment problem is essentially a two-stage problem with recourse. In the first stage, the capacity investment is selected. In the second stage, the decision concerns the routing of the messages to meet the observed demands. The sequence is thus: invest in capacities, observe the demands and route the messages. The important feature is the flexibility brought in by the recourse nature of the routing process. We propose to capture this feature via *decision rules*. To do so we must formalize first the demand process and next the decision rule.

In real life, defining a probability distribution for the demand may be difficult. It is far more natural and easier for practitioners to provide a range around a nominal value for each individual demand. This information can be captured by the equation:

$$d_k = \bar{d}_k + \hat{d}_k \xi_k, \tag{3}$$

where ξ_k represents a random factor taking values in [-1,1], $\bar{d}_k$ is the average demand and $\hat{d}_k$ is the demand dispersion. The demand range is thus $[\bar{d}_k - \hat{d}_k, \bar{d}_k + \hat{d}_k]$.

Remark 1 *More complex models could be used, in particular to capture correlation. In that case, one would adopt a model like*

$$d = \bar{d} + \hat{D}\xi$$

where $\bar{d} \in \mathbb{R}^{|\mathcal{K}|}$ is the reference demand (e.g., the estimated average, or the estimated mode), $\xi \in \mathbb{R}^p$ is the random factor of perturbation and $\hat{D}$ is a $|\mathcal{K}| \times p$ factor matrix. In general $p \leq |\mathcal{K}|$. If the first two moments of the probability distribution of the demand are known, $\bar{d}$ would be the mean demand and $\hat{D}$ the square root of the covariance matrix. Unfortunately, those data are not available in practice, and very little information can be obtained on the possible realizations of the demands. For these reasons, we stick with the simpler model (3).

Given the demand model, we can now define a decision rule as a function from the space of demand realizations to the space of recourse decision variables. The concept of decision rule captures the fact that the recourse decisions need not be fixed before the demand is realized and can be adjusted to fit the observed demand. An ADR is just a special case of this general setting. Linearity makes them computationally tractable, but possibly sub-optimal. In our case, the decision rule will apply to the routing decisions, that is to the flows on the admissible paths. An ADR will make each individual flow an affine function of all demand uncertainties:

$$f_{ik} = \alpha_{0ik} + \sum_{k' \in \mathcal{K}} \alpha_{k'ik} \xi_{k'}.$$

In this representation, the new decision variables are the coefficients α_{0ik} and $\alpha_{k'ik}$.

In view of the large number of demands, i.e., the large cardinality of $\mathcal{K}$, the number of variables is large and the robust model to be developed in the later section may become excessively large. Though well founded in theory, this approach may not lead to significantly better results than the simpler one in which the decision rule depends on a lesser number of factors. In [18], the authors restricted the decision rule for the OD pair k to be a function of only two factors: the traffic amount d_k and the sum of all other demands. In the present paper we intend to use a more refined version in which for each OD pair k, the set of other demands are partitioned on the basis of a proximity concept. The OD pairs $k' \in \mathcal{K} \setminus \{k\}$ that are deemed "close" to k are gathered in the set V_k. The other demands are collected into $R_k = \{k' \in \mathcal{K} | k' \neq k, k' \notin V_k\}$. With these notations we define the decision rule as the affine function:

$$f_{ik} = \alpha_{0ik} + \alpha_{1ik} \xi_k + \alpha_{2ik} \sum_{k' \in V_k} \xi_{k'} + \alpha_{3ik} \sum_{k' \in R_k} \xi_{k'}. \tag{4}$$

With this rule, the routing for each demand is defined by exactly four coefficients α. The key issue is of course the type of proximity to be used. We tried different types, but we eventually retained the following one.

Definition 1 *The traffic demand k' is close to the demand k if their respective shortest path has at least one arc in common.*

This proves to be sufficient in practice.

Denoting $\xi_k^v = \sum_{k' \in V_k} \xi_{k'}$ and $\xi_k^r = \sum_{k' \in R_k} \xi_{k'}$ we may write in a more compact form:

$$f_{ik} = \alpha_{0ik} + \alpha_{1ik} \xi_k + \alpha_{2ik} \xi_k^v + \alpha_{3ik} \xi_k^r. \tag{5}$$

Upon replacing the flow variables by their ADR in problem (2) we get, for any single traffic event ξ:

$$\min_{\alpha, c \geq 0, y} \quad \sum_{a \in \mathcal{A}} l_a c_a \tag{6a}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \pi_{ik}^a \left(\alpha_{0ik} + \alpha_{1ik} \xi_k + \alpha_{2ik} \xi_k^v + \alpha_{3ik} \xi_k^r \right) \leq c_a, \quad a \in \mathcal{A} \tag{6b}$$

$$\sum_{i \in \mathcal{I}_k} (\alpha_{0ik} + \alpha_{1ik} \xi_k + \alpha_{2ik} \xi_k^v + \alpha_{3ik} \xi_k^r) = \bar{d}_k + \xi_k \hat{d}_k, \quad k \in \mathcal{K} \tag{6c}$$

$$y_{ik} M_k \geq \alpha_{0ik} + \alpha_{1ik} \xi_k + \alpha_{2ik} \xi_k^v + \alpha_{3ik} \xi_k^r \geq 0, \quad k \in \mathcal{K}, i \in \mathcal{I}_k \tag{6d}$$

$$\sum_{i \in \mathcal{I}_k} y_{ik} \leq m_k, \quad k \in \mathcal{K} \tag{6e}$$

$$y_{ik} \in \{0, 1\}, \quad k \in \mathcal{K}, i \in \mathcal{I}_k. \tag{6f}$$

In this model, the uncertainty creeps into most constraints, either by definition like in (6c), or via the decision rule like in (6b) and (6d). We shall use the robust optimization paradigm to handle uncertainty in those constraints.

4 Robust counterpart of an uncertain constraint

In this section, we consider the single linear constraint with uncertain coefficient

$$\tilde{a}(\xi)^T x \leq b$$

where x is the decision variable and ξ is a random factor. Note that all the constraints of formulation (6) can be written under this form.

The first basic assumption on the uncertain parameters is that they depend on some random factor ξ in a linear way.

Assumption 1 *The uncertain vector $\tilde{a}$ is written as:*

$$\tilde{a}(\xi) = \bar{a} + P\xi,$$

where $\xi \in [-1,1]^m$ and P is an $n \times m$ matrix.

The certain vector $\bar{a}$ is usually named the normal factor. We can thus focus on the uncertain component of the constraint:

$$\underbrace{\bar{a}^T x}_{\text{certain}} + \underbrace{(P^T x)^T \xi}_{\text{uncertain}} \leq b. \tag{7}$$

4.1 Robust counterpart

The idea of robust optimization is to focus on a subset of all possible events that it is made of all realizations of the underlying uncertain factor ξ that the modeler deems necessary to protect against. This is the so-called *uncertainty set*. The robust version of the initial uncertain constraint $\tilde{a}(\xi)^T x \leq b$ consists in enforcing the uncertain constraint (7), not for *all* possible realizations, but only for those in the uncertainty set; that is, the less restrictive constraint

$$\bar{a}^T x + (P^T x)^T \xi \leq b, \text{ for all } \xi \in \Xi, \tag{8}$$

where $\Xi \subset \mathbb{R}^m$ is the uncertainty set. A solution to this constraint is called *robust* with respect to Ξ and (8) is the *robust counterpart* of the uncertain constraint (7). In the present formulation, the robust counterpart is not numerically tractable. However it can be given an equivalent form for a large class of uncertainty sets.

In this paper we shall use the general uncertainty set

$$\Xi = \{\xi \mid \|\xi\|_\infty \leq 1, \ \|\xi\|_p \leq \kappa_p\} \tag{9}$$

with $p = 1$ or $p = 2$. Using standard duality theory for convex programming one can easily derive an equivalent form of the robust counterpart of the uncertain constraint.

Theorem 1 *The robust equivalent of the robust constraint*

$$\bar{a}^T x + (P^T x)^T \xi \leq b, \text{ for all } \xi \in \Xi = \{\xi \mid \|\xi\|_\infty \leq 1, \ \|\xi\|_p \leq \kappa_p\},$$

is the constraint in x **and** *w*

$$\bar{a}^T x + \kappa_p \|P^T x - w\|_q + \|w\|_1 \leq b \tag{10}$$

with $q = \infty$ if $p = 1$ and $q = 2$ if $p = 2$.

For a proof of this standard result, we refer to [6]. The equivalent robust counterpart is a numerically tractable conic constraint for both values $p = 2$ and $p = 1$. In the former case, the constraint is conic quadratic. In the second case, the constraint is amenable to a set of linear constraints.

Theorem 2 *The conic constraint*

$$\bar{a}^T x + \kappa_1 \|P^T x - w\|_\infty + \|w\|_1 \leq b \tag{11}$$

has the same set of solutions as the system of linear inequalities

$$\begin{aligned} \bar{a}^T x + \kappa_1 t + e^T w &\leq b && (12a)\\ -(w + te) \leq P^T x &\leq w + te && (12b)\\ w \geq 0,\ t \geq 0. & && (12c) \end{aligned}$$

The l_1 and l_∞ norms are easily represented by linear inequalities. As mentioned in [6, Proposition 1 (p96)] the inequalities $-te \leq P^T x - w \leq te$ are in fact equivalent to (12b) with $w \geq 0$.

We now review two special cases of interest for our problem.

4.2 Constraints with upper and lower limits

The robust optimization paradigm implies that each constraint is immunized separately. However, the two sides of a two-sided inequality can be treated simultaneously. Indeed, let

$$\underline{b} \leq \tilde{a}(\xi)^T x \leq \bar{b}$$

be a two-sided inequality. It is easy to show that the same dual variable can be used on the two sides. Thus the set of vectors x satisfying the above two-sided inequality for all $\xi \in \Xi$ is equivalently written as follows:

$$\begin{aligned} \bar{a}^T x + \kappa_1 t + e^T w &\leq \bar{b} && (13a)\\ \bar{a}^T x - \kappa_1 t - e^T w &\geq \underline{b} && (13b)\\ te + w &\geq P^T x && (13c)\\ te + w &\geq -P^T x && (13d)\\ t \geq 0, w \geq 0. & && (13e) \end{aligned}$$

The practical advantage of this formulation is that the above robust equivalent has only one more scalar inequality than (12).

4.3 Uncertainty factors with identical coefficients

In the capacity expansion problem, for some constraints, several random factors in $\{\xi_k\}_{k \in K}$ have the same coefficient. We propose to exploit this property.

To ease the presentation, we shall denote $z = P^T x$. Suppose that there exists a set of indices $J \subset \{1, \ldots, m\}$ such that the $\{z_i\}_{i \in J}$ are all equal. Let $l \in J$, we have: $\forall i \in J, z_i = z_l$. Denoting $I = \{1, \ldots, m\} \setminus J$, we have:

$$z^T \xi = \sum_{i \in I} z_i \xi_i + z_l \sum_{i \in J} \xi_i.$$

Furthermore, from Theorem 1, the robust equivalent is: $\bar{a}^T x + \sigma^*(x) \leq b$, with:

$$\sigma^*(x) = \min_w \left\{ \kappa_1 \cdot \max \left\{ \max_{i \in I} |z_i - w_i|, \max_{i \in J} |z_i - w_i| \right\} + \sum_{i \in I} |w_i| + \sum_{i \in J} |w_i| \right\}.$$

It is easy to see that the condition $z_i = z_l$, for all $i \in J$, implies that the optimal solution w^* is such that the $\{w_i^*\}_{i \in J}$ take the same value. As a result: $\forall i \in J, w_i^* = w_l^*$. Hence:

$$\sigma^*(x) = \min_w \left\{ \kappa_1 \cdot \max_{i \in I \cup \{l\}} |z_i - w_i| + \sum_{i \in I} |w_i| + \text{card}(J) \cdot |w_l| \right\}.$$

Proposition 1 *Let $z = P^T x$. If there exists a set of indices $J \subset \{1, \ldots, m\}$ and $l \in J$ such that: $\forall i \in J, z_i = z_l$, then the robust equivalent for the robust constraint*

$$(\bar{a} + P\xi)^T x \leq b, \ \forall \xi \in \Xi = \{\xi \mid \|\xi\|_\infty \leq 1, \ \|\xi\|_1 \leq \kappa_1\}$$

is the single constraint in variables x **and** *w*

$$\bar{a}^T x + \kappa_1 \max_{i \in I \cup \{l\}} |z_i - w_i| + \sum_{i \in I} |w_i| + \text{card}(J) \cdot |w_l| \leq b,$$

with $I = \{1, \ldots, m\} \setminus J$.

4.4 Probabilistic analysis

Theorem 1 does not rely on probabilistic arguments. Nevertheless, it can be given a probabilistic interpretation provided some assumptions are made on the behavior of the uncertainty factor ξ. The standard approach, as described in [3, chapter 2], consists in selecting a class of distributions with some known characteristics, e.g., a finite support and a known mean. The typical result consists in showing that for a well-defined value of κ_p, any solution of the equivalent robust counterpart (10) will satisfy the uncertain constraint with probability at least $1 - \epsilon$, whatever is the true distribution within the considered class.

In this paper, we shall not consider a class of distribution, but a specific one, namely the symmetric triangular distribution on $[-1, 1]$.

Assumption 2 *The random factors ξ_j are independent, with asymmetric triangular distribution on the interval $[-1, 1]$.*

This assumption may look restrictive. We chose it in order to be consistent with our simulation study, but weaker assumptions would still leads to very similar results. The main difference is that the coefficient κ_p is larger when the assumption is weaker.

Our probabilistic result is as follows.

Theorem 3 *Assume that the random variables ξ_i, $i = 1, \ldots, m$ are i.i.d. with a symmetric triangular distribution on the range $[-1, 1]$. If the equivalent robust counterpart (10) is satisfied with*

$$\kappa_p = \begin{cases} \sqrt{\frac{1}{3} \ln \frac{1}{\epsilon}} & \text{if} \quad p = 2 \\ \sqrt{\frac{1}{3} \ln \frac{1}{\epsilon}} \sqrt{m} & \text{if} \quad p = 1, \end{cases}$$

then

$$Prob((\bar{a} + Px)^T \xi \leq b) \geq 1 - \epsilon.$$

Proof: We prove the theorem with $p = 2$ first. For the sake of simpler notations, we first rewrite the uncertain constraint (8) as

$$z_0 + z^T \xi \leq 0, \tag{14}$$

with $z_0 = \bar{a}^T - b$ and $z = P^T x$. We denote $Z = z_0 + z^T \xi$.

From the Markov inequality we bound the probability of violation of (14) by $\text{Prob}(Z \geq 0) = \text{Prob}(e^Z \geq 1) \leq E(e^Z)$. Noting that $\text{Prob}(Z \geq 0) = \text{Prob}(tZ \geq 0)$ holds for all $t > 0$, we have the stronger bound

$$\text{Prob}(Z \geq 0) \leq \inf_{t>0} E(e^{tZ}).$$

Let us explicit the right-hand side. Remembering that the ξ_j are independent random variables, we may write

$$E(e^{tZ}) = E(e^{t(z_0 + \sum_j z_j \xi_j)}) = e^{z_0} E\left(\prod_j e^{t z_j \xi_j}\right) = e^{z_0} \prod_j E(e^{t z_j \xi_j}).$$

We claim that the above quantity is bounded by $e^{\frac{t^2}{12}\sum_j z_j^2}$. To prove the claim, we focus on the typical component $E(e^{t z_j \xi_j})$. For the sake of simpler notation, we temporarily drop the index j and we denote $\tau = tz$. From the triangular distribution assumption on ξ, we have

$$\begin{aligned}
E(e^{\tau\xi}) &= \int_{-1}^{0} e^{\tau\xi}(1+\xi)d\xi + \int_{0}^{1} e^{\tau\xi}(1-\xi)d\xi && \text{(15a)}\\
&= \frac{e^{\tau} + e^{-\tau} - 2}{\tau^2} && \text{(15b)}\\
&= \frac{1}{\tau^2}\left[\sum_{k=0}^{\infty}\frac{\tau^k}{k!} + \sum_{k=0}^{\infty}\frac{(-\tau)^k}{k!} - 2\right] && \text{(15c)}\\
&= 2\sum_{k=0}^{\infty}\frac{(\tau^2)^k}{(2k+2)!} = \sum_{k=0}^{\infty} a^k (\tau^2)^k. && \text{(15d)}
\end{aligned}$$

The next step consists in bounding $E(e^{\tau\xi})$ by

$$e^{\alpha^2\tau^2} = \sum_{k=0}^{\infty}\frac{(\alpha^2\tau^2)^k}{k!} = \sum_{k=0}^{\infty} b^k (\tau^2)^k. \tag{16}$$

To prove the claim, we compare the two series term-wise and show that $a_k \leq b_k$ for all k Clearly, $a_0/b_0 = 1$. For $k = 1$, $\frac{a_1}{b_1} = \frac{2}{4\times3\times2}\frac{1}{\alpha^2}$. The smallest value for α^2 to ensure $a_1 \leq b_1$ is $1/12$. Assume $a_k/b_k \leq 1$ holds for some k and $\alpha^2 = 1/12$. Let us show that it holds for $k+1$. We have

$$\frac{a_{k+1}}{b_{k+1}} = \frac{a_k}{b_k}\frac{k+1}{(2k+3)(2k+4)}\frac{1}{\alpha^2} \leq \frac{k+1}{(2k+3)(2k+4)}\frac{1}{\alpha^2} \leq \frac{1}{2(2k+4)\alpha^2}.$$

One easily checks that for $k \geq 2$, the inequality $2(2k+4)\alpha^2 = (k+2)/3 > 1$ holds. We conclude that $a_k \leq b_k$ for all k. Hence (16) is component-wise larger than (15d). The claim is proved.

We can now bound the probability of interest:

$$\begin{aligned}
\text{Prob}(Z \geq 0) &\leq \inf_{t>0} e^{tz_0} \prod_j E(e^{tz_j\xi_j}) && \text{(17a)}\\
&\leq \inf_{t>0} e^{tz_0} \prod_j e^{\frac{t^2}{12}z_j^2} && \text{(17b)}\\
&\leq e^{\inf_{t>0}(tz_0 + \frac{t^2}{12}\sum_j z_j^2)}. && \text{(17c)}
\end{aligned}$$

Note that for $z_0 \geq 0$, the infimum in the right-hand side of (17c) is 1. In order to derive a useful bound, we shall assume $z_0 < 0$. Under this condition, the infimum is achieved at $t_{opt} = \frac{-6z_0}{||z||_2^2} > 0$ and

$$\text{Prob}(Z \geq 0) \leq e^{-\frac{3z_0^2}{||z||_2^2}}.$$

A sufficient condition to have the probability of constraint violation bounded above by ϵ is that the deterministic constraint, or robust equivalent,

$$z_0 + \sqrt{\frac{1}{3}\ln\frac{1}{\epsilon}}||z||_2 \leq 0 \tag{18}$$

be satisfied.

Finally, condition (18) can be made stronger without endangering the probabilistic statement. Indeed, let w be an arbitrary vector. Remembering that $-1 \leq \xi_j \leq 1$,

$$z_0 + z^T\xi = z_0 + (z-w)^T\xi + w^T\xi \leq z_0 + (z-w)^T\xi + ||w||_1. \tag{19}$$

Letting $\hat{z}_0 = z_0 + ||w||_1$ and $\hat{z} = z - w$, we have by (19)

$$\text{Prob}(z_0 + z^T\xi > 0) \leq \text{Prob}(\hat{z}_0 + \hat{z}^T\xi > 0).$$

We now apply (18) to the uncertain constraint $\hat{z}_0 + \hat{z}^T\xi \leq 0$ and obtain

$$\text{Prob}(\hat{z}_0 + \hat{z}^T\xi > 0) \leq \epsilon$$

whenever

$$z_0 + ||w||_1 + \sqrt{\frac{1}{3}\ln\frac{1}{\epsilon}}||z-w||_2 \leq 0$$

holds. This completes the first part of the proof.

To prove the last statement in the theorem, we just note that for $p = 1$, the coefficient $\sqrt{m}$ in the definition of κ_p comes into play because for any $a \in \mathbb{R}^m$ the ℓ_2-norm is bounded by $||a||_2 \leq \sqrt{m}\max_j |a_j| = \sqrt{m}||a||_\infty$.

■

Remark 2 *Similar results are obtained if Assumption 2 is modified to cover more general classes of distributions. We refer to [3, chap.2] for an extensive review of various classes. The main point is that the robust counterpart has almost the same structure as in Theorem 1. The difference is usually in the immunization factor κ_p. For instance, if one considers the class of distributions with support $[-1, 1]$ and mean $E(\xi) = 0$, Theorem 3 holds in its exact form with $\kappa_2 = \sqrt{2\ln\frac{1}{\epsilon}}$ instead of $\kappa_2 = \sqrt{\frac{1}{3}\ln\frac{1}{\epsilon}}$, that is an increase by a factor almost 2.5. This increase is the price to pay to compensate for a partial information on the true distribution. In our experiments, we choose to work with Assumption 2 and perform simulations for the validation process under the same conditions.*

5 Robust capacity expansion formulation

We now apply the concepts of robust optimization to the capacity expansion problem.

5.1 Demand constraints

The demand constraint requires that the total flow into a demand node meets the demand in a perfect equality. Since this demand constraint involves uncertain components, perfect equality cannot hold unless the equality is an identity with respect to each individual demand. In the case of the demand constraint associated with $k \in \mathcal{K}$, it means that the equation with uncertain coefficients

$$\sum_{i \in \mathcal{I}_k} (\alpha_{0ik} + \alpha_{1ik}\xi_k + \alpha_{2ik}\xi_k^v + \alpha_{3ik}\xi_k^r) = \bar{d}_k + \xi_k \hat{d}_k$$

holds for any possibly demand outcome only if the following equality constraints hold

$$\sum_{i \in \mathcal{I}_k} \alpha_{0ik} = \bar{d}_k, \ \sum_{i \in \mathcal{I}_k} \alpha_{1ik} = \hat{d}_k, \ \sum_{i \in \mathcal{I}_k} \alpha_{2ik} = 0, \ \sum_{i \in \mathcal{I}_k} \alpha_{3ik} = 0. \tag{20}$$

In formulation (20), each demand constraint is replaced by 4 equalities and thus we end up with $4|\mathcal{K}|$ equality constraints.

5.2 Capacity constraints

To obtain the robust counterpart of the capacity constraints, we consider the polyhedral uncertainty set (9) with $p = 1$

$$\Xi^{cap} = \{\xi \mid \|\xi\|_\infty \leq 1, \ \|\xi\|_1 \leq \kappa_1^{cap}\}$$

Hence applying theorems 1 and 2, the equivalent robust counterpart of the single robust capacity constraint

$$\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \pi_{ik}^a (\alpha_{0ik} + \alpha_{1ik}\xi_k + \alpha_{2ik} \sum_{k' \in V_k} \xi_{k'} + \alpha_{3ik} \sum_{k' \in R_k} \xi_{k'}) \leq c_a, \quad \forall \xi \in \Xi^{cap}. \tag{21}$$

with $a \in \mathcal{A}$, is the system of linear inequalities

$$\begin{aligned}
&\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \pi_{ik}^a \alpha_{0ik} + \kappa_1^{cap} v_a^{cap} + \sum_{k' \in \mathcal{K}} u_{ak'}^{cap} \leq c_a, \\
&u_{ak}^{cap} + v_a^{cap} \geq \pm\Big(\sum_{i \in \mathcal{I}_k} \pi_{ik}^a \alpha_{1ik} + \sum_{k' \in V_k, i \in \mathcal{I}'_k} \pi_{ik'}^a \alpha_{2ik} + \sum_{k' \in R_k, i \in \mathcal{I}'_k} \pi_{ik'}^a \alpha_{3ik}\Big), k \in \mathcal{K} \\
&u^{cap} \geq 0, \ v^{cap} \geq 0.
\end{aligned}$$

Finally the robust version of the capacity constraints has $2|\mathcal{A}| \times |\mathcal{K}|$ additional constraints and $|\mathcal{A}|(1 + |\mathcal{K}|)$ additional variables.

In Section 6, we experiment different values for κ_1^{cap} leading to different probabilities of satisfaction with Theorem 3.

5.3 Constraints of flow bounds

The ADR defines the flow in the recourse stage of the problem. This flow has to satisfy, for any demand $k \in \mathcal{K}$ and $i \in \mathcal{I}_k$:

$$y_{ik} M_k \geq \alpha_{0ik} + \alpha_{1ik}\xi_k + \alpha_{2ik}\xi_k^v + \alpha_{3ik}\xi_k^r \geq 0, \quad \forall \xi \in \Xi^{pos} \tag{22}$$

with

$$\Xi^{pos} = \{\xi \mid \|\xi\|_\infty \leq 1, \ \|\xi\|_1 \leq \kappa_1^{pos}\}.$$

According to Proposition 1, the robust counterpart is given by the following set of constraints:

$$\alpha_{0ik} - \kappa_1^{pos} v_{ik}^{pos} - u_{ik}^{pos} - n_k^v u_{kv}^{pos} - n_k^r u_{kr}^{pos} \geq 0 \tag{23a}$$
$$\alpha_{0ik} + \kappa_1^{pos} v_{ik}^{pos} + u_{ik}^{pos} + n_k^v u_{kv}^{pos} + n_k^r u_{kr}^{pos} \leq y_{ik} M_k \tag{23b}$$
$$u_{ik}^{pos} + v_{ik}^{pos} \geq \pm\alpha_{1ik} \tag{23c}$$
$$u_{ikv}^{pos} + v_{ik}^{pos} \geq \pm\alpha_{2ik} \tag{23d}$$
$$u_{ikr}^{pos} + v_{ik}^{pos} \geq \pm\alpha_{3ik} \tag{23e}$$
$$u^{pos} \geq 0, \ v^{pos} \geq 0. \tag{23f}$$

The above formulation of the flow bounds yields $6|\mathcal{I}|$ additional constraints and $4|\mathcal{I}|$ additional variables where $|\mathcal{I}|$ is the total number of paths.

In the experiments below, we set $\kappa_1^{pos} = 1$ ensuring from Theorem 3 a level of satisfaction for each constraint of 95%.

6 Numerical experiments

6.1 Test problems

The set of test problems consists of seven oriented graphs of various sizes. All of them correspond to physical networks, except `planar30` that was generated with a tool designed to produce network instances that simulate telecommunications problems. The corresponding data (topology, costs and demands) are publicly available either from `http://sndlib.zib.de/home.action` or from `http://www.di.unipi.it/di/groups/optimize/Data/MMCF.html` (in the case of `planar30`). The networks in the SNDlib being undirected, we chose to duplicate each arc in the graphs `polska`, `nobel-us`, `atlanta` and `france` in two arcs with opposite directions. This choice guarantees that there always exists a feasible path for each OD pair.

For each problem instance, we have constructed a set of distinct paths for each Origin-Destination (OD) pair according to the shortest distance (cost) criterion. To this end, we used the k-shortest paths algorithm [11]. In some problem instances it may happen that for a few OD-pairs the number of paths with distinct intermediary nodes is less than k. Table 1 provides, in the first three columns, for each instance the number of nodes, the number of arcs and the number of origin-destination pairs of demands. The next column gives the total number of generated shortest paths with $k = 4$. The next two columns give an evaluation of the corresponding number of constraints and number of variables for the robust model (6) without integer constraints. The last two columns give the same information for the determistic model (1).

Finally, the traffic demands are all uncertain with a range of variation ±50% of the nominal demand $\bar{d}_k$. This range of variation fits the observed yearly forecast errors.

	#nodes	#arcs	#OD	#paths*	robust version (6) #const*	#var*	determin. version (1) #const*	#var*
pdh	11	34	24	54	2140	1316	146	88
di-yuan	11	42	22	61	2405	1496	167	103
polska	12	36	66	264	6900	4560	402	300
nobel-us	14	42	91	364	10598	6818	539	406
planar30	30	150	92	368	30694	17044	760	518
atlanta	15	44	210	840	25244	16048	1138	884
france	25	90	300	1188	63606	36684	1668	1278

(*) Indicative values using the k-shortest paths algorithm with $k = 4$.
The nonnegativity constraints are included in the figures.

Table 1: Test problems.

6.2 Experiments

For each problem instance we considered four different models. They differ by the number of available paths per OD and the restriction on the number of usable paths per OD. The four models are as follows:

Model 1 The set of available paths for each OD pair reduces to a singleton ($k = 1$). The ADR is de facto restricted to a single path. (Results on Table 2.)

Model 2 The set of available paths for each OD pair consists of four paths ($k = 4$) and there is no restriction on the number of paths that the ADR can use. (Results on Table 3.)

Model 3 The set of available paths for each OD pair consists of four paths ($k = 4$), but the ADR is restricted to a single path. (Results on Table 4.)

Model 4 The set of available paths for each OD pair consists of four paths ($k = 4$), but the ADR is restricted to 2 paths. (Results on Table 5.)

For each model we performed experiments with different level of risk $1 - \epsilon$. According to Theorem 3 each level of risk defines a immunization coefficient $\kappa^{cap} = \sqrt{\frac{1}{3} \ln \frac{1}{\epsilon}}$. Recall that we do not apply the same risk target to the capacity constraints and the nonnegativity constraint on the flows. For the latter we uniformly use $\kappa^{pos} = 1$ in all experiments; it implies a level $1 - \epsilon = 95\%$,

Note that Model 1 and 3 both use a single path per OD, but in Model 1 the path is imposed from the outset, while in Model 3 its selection is to be made among four possible ones by the optimization process. All instances were solve with Cplex 11.0 on a computer running an Intel(R) Xeon(TM) processor 2.80GHz with 3Gb of RAM. A time limit of 5 hours was imposed on each instance.

6.3 Validation process

The goal of the validation process is to provide an empirical evaluation of the risk associated to a solution for the capacity expansion problem. This is done by generating a sample of demand scenarios and analyzing the performance of the selected solution on each scenario. The simulations are performed by a Monte-Carlo scheme in accordance with Assumption 2, that is, the demands d_k are assumed to be independent with a symmetrical triangular distribution on $[\bar{d}_k - \hat{d}_k, \bar{d}_k + \hat{d}_k]$, with $\hat{d}_k = 0.5\bar{d}_k$. At each simulation, a random matrix is sampled from the distributions. Recall that an optimal solution is a pair of installed capacities c^* and affine decision rules for the flows $f^*(d)$. The rules are such that the demand constraints $\sum_{i\in\mathcal{I}_k^*} f_{ik}^*(d) = d_k$ are satisfied by construction, but some capacity constraint $\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}_k^*} f_{ik}^*(d)\pi_{ik}^a \leq c_a^*$ may be violated. To check whether the violation is really due to insufficient capacity, we drop the ADR flow solution and solve the auxiliary multicommodity flow problem

$$\min_{f\geq 0, s\geq 0} \quad \sum_{k\in\mathcal{K}} s_k \tag{24a}$$

$$\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}_k^*} f_{ik}\pi_{ik}^a \leq c_a^*, \quad a\in\mathcal{A} \tag{24b}$$

$$\sum_{i\in\mathcal{I}_k^*} f_{ik} = d_k - s_k, \quad k\in\mathcal{K} \tag{24c}$$

where c^* is the optimal capacity expansion vector and I_k^* is the set of paths selected in the optimization process. If the objective value is positive, the capacities and the selected set of paths do not allow demand satisfaction in full. We use the ratio $\rho = \sum_{k\in\mathcal{K}} s_k / \sum_{k\in\mathcal{K}} d_k$ to measure the fraction of unserved demand in the simulation instance.

For each problem instance, we build a sample of 1000 simulations. We use the following statistics as indicators of the performance of the solution: i) the proportion of simulations for which the capacities turn out to be insufficient as an estimator of the probability of failure to serve the demands; ii) the relative traffic overflow ρ in case of capacity violation, that is an estimate of the conditional expectation of the proportion of traffic overflow, conditionally to $\rho > 0$. Note that the unconditional expectation is the product of the conditional expectation by the probability of failure.

6.4 Results

The results for Model 1 to 4 are reported in Tables 2 to 5. Tables 4 and 5 pertaining to Model 3 and 4 do not involve the two largest instances `atlanta` and `france`, because the solution time exceeded the time limit (5 hours). The difficulty stems from the complexity induced by the integral restriction on the number of paths (1 or 2 among 4) to be used for each OD.

The results for a particular problem are displayed on four rows in each table. The first row displays the solution cost ("Solutions"); the second row displays the proportion of simulations for which the traffic to be routed exceeds the installed capacity ("% of violations"); the third row shows the average conditional excess traffic in terms of fraction of lost traffic ("Rel. cond. viol."); and finally, the unconditional average amount of lost traffic ("Expected traffic loss"). We could have dispensed with this last row, as its figures are just the product of the two preceding rows. We chose to include them to facilitate the evaluation.

We comment here on some specific points worth noticing.

Controlling feasibility The percentage of violations is always much smaller than the targeted risk probability ϵ obtained from the theoretical analysis of Theorem 3. We can invoke several reasons for this. First, the robust counterpart is built on the assumption that ADR are used, while in the validation process we use fully adjustable recourse. The latter may be feasible, while the former turn out to be not. (See Table 8 and 9 for a refined analysis.) Second, we must recall that Theorem 3 only provides a lower bound on the probability of constraint satisfaction. The theorem makes it possible to interpret the equivalent robust counterpart as a "safe tractable approximation" of the chance constraint (see [3]) at the expense of some conservativeness.

	Total Protection	Robust solutions with $1-\epsilon$ for capacity constraints 85%	50%	10%	5%	Deterministic solution
			pdh			
Solutions	7.52E+008	7.52E+008	7.52E+008	7.32E+008	6.62E+008	5.02E+008
% of violations	0.00%	0.00%	0.00%	7.50%	79.30%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.06%	0.47%	8.24%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.37%	8.24%
			di-yuan			
Solutions	5.95E+006	5.94E+006	5.74E+006	5.20E+006	4.82E+006	3.97E+006
% of violations	0.00%	0.00%	0.10%	32.60%	81.80%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.15%	0.53%	0.95%	7.42%
Expected traffic loss	0.00%	0.00%	0.00%	0.17%	0.78%	7.42%
			polska			
Solutions	7.04E+006	6.90E+006	6.55E+006	5.66E+006	5.41E+006	4.69E+006
% of violations	0.00%	0.00%	0.00%	41.20%	81.70%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.39%	0.63%	4.32%
Expected traffic loss	0.00%	0.00%	0.00%	0.16%	0.51%	4.32%
			nobel-us			
Solutions	1.29E+008	1.28E+008	1.23E+008	1.08E+008	1.03E+008	8.60E+007
% of violations	0.00%	0.00%	0.00%	15.00%	58.50%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.19%	0.29%	4.51%
Expected traffic loss	0.00%	0.00%	0.00%	0.03%	0.17%	4.51%
			planar30			
Solutions	6.61E+007	6.44E+007	6.13E+007	5.34E+007	5.07E+007	4.31E+007
% of violations	0.00%	0.00%	0.00%	39.50%	84.00%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.42%	0.66%	5.20%
Expected traffic loss	0.00%	0.00%	0.00%	0.17%	0.55%	5.20%
			atlanta			
Solutions	4.59E+011	4.53E+011	4.41E+011	4.04E+011	3.87E+011	3.07E+011
% of violations	0.00%	0.00%	0.00%	0.40%	5.60%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.11%	0.12%	3.90%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.01%	3.90%
			france			
Solutions	6.65E+008	6.42E+008	6.10E+008	5.43E+008	5.22E+008	4.44E+008
% of violations	0.00%	0.00%	0.00%	1.90%	16.10%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.10%	0.12%	3.07%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.02%	3.07%

Table 2: Robust solutions and performances for Model 1 ($k = 1$: only one shortest path).

	Total Protection	Robust solutions with $1-\epsilon$ for capacity constraints 85%	50%	10%	5%	Deterministic solution
pdh						
Solutions	7.52E+008	7.52E+008	7.52E+008	7.00E+008	6.49E+008	5.02E+008
% of violations	0.00%	0.00%	0.00%	30.10%	72.40%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.53%	0.67%	8.42%
Expected traffic loss	0.00%	0.00%	0.00%	0.16%	0.49%	8.42%
di-yuan						
Solutions	5.95E+006	5.81E+006	5.59E+006	5.07E+006	4.74E+006	3.97E+006
% of violations	0.00%	0.00%	0.50%	24.20%	75.10%	100.00%
Rel. cond. viol.	0.00%	0.00%	1.54%	0.64%	1.04%	7.30%
Expected traffic loss	0.00%	0.00%	0.01%	0.15%	0.78%	7.30%
polska						
Solutions	7.00E+006	6.56E+006	6.10E+006	5.44E+006	5.27E+006	4.69E+006
% of violations	0.00%	0.00%	0.10%	16.40%	34.90%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.48%	0.51%	0.58%	3.97%
Expected traffic loss	0.00%	0.00%	0.00%	0.08%	0.20%	3.97%
nobel-us						
Solutions	1.28E+008	1.22E+008	1.16E+008	1.03E+008	9.92E+007	8.60E+007
% of violations	0.00%	0.00%	0.00%	1.60%	12.30%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.25%	0.27%	4.39%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.03%	4.39%
planar30						
Solutions	6.54E+007	6.25E+007	5.91E+007	5.22E+007	4.99E+007	4.31E+007
% of violations	0.00%	0.00%	0.00%	17.10%	57.70%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.31%	0.45%	5.13%
Expected traffic loss	0.00%	0.00%	0.00%	0.05%	0.26%	5.13%
atlanta						
Solutions	4.53E+011	4.41E+011	4.27E+011	3.93E+011	3.77E+011	3.07E+011
% of violations	0.00%	0.00%	0.00%	0.10%	3.20%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.34%	0.11%	3.88%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.00%	3.88%
france						
Solutions	6.18E+008	5.86E+008	5.56E+008	5.08E+008	4.92E+008	4.44E+008
% of violations	0.00%	0.00%	0.00%	0.90%	9.80%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.16%	0.14%	3.10%
Expected traffic loss	0.00%	0.00%	0.00%	0.00%	0.01%	3.10%

Table 3: Robust solutions and performances for Model 2 ($k = 4$ and all 4 shortest paths can support the solution).

	Total Protection	Robust solutions with $1-\epsilon$ for capacity constraints 85%	50%	10%	5%	Deterministic solution
			pdh			
Solutions	7.52E+008	7.52E+008	7.52E+008	7.29E+008	6.62E+008	5.02E+008
% of violations	0.00%	0.00%	0.00%	24.00%	79.90%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.28%	0.49%	8.27%
Expected traffic loss	0.00%	0.00%	0.00%	0.07%	0.39%	8.27%
			di-yuan			
Solutions	5.96E+006	5.88E+006	5.67E+006	5.13E+006	4.78E+006	3.97E+006
% of violations	0.00%	0.00%	0.10%	31.30%	81.00%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.43%	0.63%	0.93%	7.43%
Expected traffic loss	0.00%	0.00%	0.00%	0.20%	0.75%	7.43%
			polska			
Solutions	7.04E+006	6.85E+006	6.42E+006	5.61E+006	5.37E+006	4.69E+006
% of violations	0.00%	0.00%	0.10%	27.10%	62.60%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.04%	0.36%	0.50%	4.32%
Expected traffic loss	0.00%	0.00%	0.00%	0.10%	0.31%	4.32%
			nobel-us			
Solutions	1,29E+008	1,27E+008	1,21E+008	1,06E+008	1,01E+008	8,60E+007
% of violations	0,00%	0,00%	0,00%	5,80%	34,30%	100,00%
Rel. cond. viol.	0,00%	0,00%	0,00%	0,21%	0,24%	4,53%
Expected traffic loss	0,00%	0,00%	0,00%	0,01%	0,08%	4,53%
			planar30			
Solutions	6.60E+007	6.38E+007	6.05E+007	5.30E+007	5.04E+007	4.31E+007
% of violations	0.00%	0.00%	0.10%	44.80%	81.00%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.39%	0.43%	0.64%	5.18%
Expected traffic loss	0.00%	0.00%	0.00%	0.19%	0.52%	5.18%

Table 4: Robust solutions and performances for Model 3 (only one path among the four shortest paths can be used by the solution).

	Total Protection	Robust solutions with $1-\epsilon$ for capacity constraints 85%	50%	10%	5%	Deterministic solution
			pdh			
Solutions	7.52E+008	7.52E+008	7.52E+008	7.02E+008	6.50E+008	5.02E+008
% of violations	0.00%	0.00%	0.00%	30.40%	78.60%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.41%	0.72%	8.26%
Expected traffic loss	0.00%	0.00%	0.00%	0.12%	0.57%	8.26%
			di-yuan			
Solutions	5.95E+006	5.81E+006	5.59E+006	5.07E+006	4.74E+006	3.97E+006
% of violations	0.00%	0.00%	0.10%	30.70%	74.90%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.84%	0.69%	1.06%	7.32%
Expected traffic loss	0.00%	0.00%	0.00%	0.21%	0.79%	7.32%
			polska			
Solutions	7.01E+006	6.58E+006	6.13E+006	5.45E+000	5.27E+006	4.69E+006
% of violations	0.00%	0.00%	0.00%	14.80%	30.20%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.50%	0.60%	4.39%
Expected traffic loss	0.00%	0.00%	0.00%	0.07%	0.18%	4.39%
			nobel-us			
Solutions	1.28E+008	>5h	>5h	1.03E+008	9.94E+007	8.60E+007
% of violations	0.00%	-	-	2.40%	7.90%	100.00%
Rel. cond. viol.	0.00%	-	-	0.22%	0.29%	4.55%
Expected traffic loss	0.00%	-	-	0.01%	0.02%	4.55%
			planar30			
Solutions	6.54E+007	6.26E+007	5.92E+007	5.23E+007	5.00E+007	4.31E+007
% of violations	0.00%	0.00%	0.00%	20.10%	55.70%	100.00%
Rel. cond. viol.	0.00%	0.00%	0.00%	0.35%	0.43%	5.12%
Expected traffic loss	0.00%	0.00%	0.00%	0.07%	0.24%	5.12%

Table 5: Robust solutions and performances for Model 4 (at most two paths among the four shortest paths can be used by the solution).

A third reason for the discrepancy stems from the fact that the analysis treats the constraints as independent elements. This is certainly not the case. To handle the issue of global robustness in a model, one would have to replace the separate capacity constraints as a single one (for instance by using the maximum of auxiliary surplus variables). This approach would provide a stronger result, but it leads to an intractable robust counterpart. Clearly, to work with numerically tractable models, we have to live with the fact that the robust solution is by construction conservative. This is why it is suitable to use simulations to assess the experimental risk associated with the solutions.

The gap between the theoretical guarantee and the simulated feasibility tends to be larger for big networks, see instances `atlanta` and `france` in Table 2 and 3. Observe that for all instances, targeting theoretically a 50% feasibility ($1-\epsilon = 50\%$) still leads to solutions with almost no violated simulation.

Cost analysis Table 6 displays the saving in investment cost as a function of the risk level for Model 1. The saving is measured in terms of a (negative) percentage of the total protection cost. Table 6 (Model 1 with only one available path per OD pair) exhibits an average gain of 5% for a theoretical level $1-\epsilon = 50\%$ with almost zero violation across the simulations. We conclude that the 5% saving are achieved at no risk. If we use the conditional and unconditional average traffic loss as an alternative measure of risk, we observe through Tables 2-5 that those values never exceed 0.69% and 0.21% (and are much lower in most cases) for the solution with a theoretical level $1-\epsilon = 10\%$. If these traffic loss figures are deemed acceptable, then the solution with $1-\epsilon = 10\%$ allows for a gain of 14% in the average (Table 6).

Table 7 displays the gain in the overall investment cost when using up to four routed per demand (Model 2) instead of one (Model 1). Hence, this table gives the impact of path diversification. The gain can be large (up to 8.8% for network `france`, when $1-\epsilon = 50\%$). For the cases of practical interest ($1-\epsilon = 50\%$ or $1-\epsilon = 10\%$), the average gain related to path diversification is roughly 4%.

To better illustrate the above comments, Figure 1 displays the evolution of the cost with the expected traffic loss for four representative instances (`diyuan`, `polska`, `nobel-us` and `france`).

Impact of affine decision rules Affine decision rules (ADR) are used to control the numerical complexity of our models, but they are restrictive. It turns out that in some simulations the ADR violates the capacity constraints, while a feasible routing can be shown to exist. To capture the negative impact of the ADR, we use Model 2 (with all four paths available) and report two indicators. Table 8 gives the proportion of simulations for which the ADR meets the capacity constraints. With a theoretical risk level $1-\epsilon \geq 50\%$, the ADR solution meets the requirements in nearly 98% cases. Table 9 complements these results with an information on the average number of simulations for which the ADR fails (i.e., a positive entry in Table 8), but a feasible routing has been found. This number increases with $1-\epsilon$, but usually remains low when $1-\epsilon \geq 50\%$. This is a good indication that ADR perform very well, at least when the taken risk is moderate.

7 Conclusion

In the present study we implemented a robust optimization approach to cope with demand uncertainty in the capacity planning of telecommunications networks. The pro-

	Total	1-ϵ				Deterministic
	protection	85%	50%	10%	5%	solution
pdh	0%	0%	0%	-3%	-12%	-33%
diyuan	0%	0%	-3%	-13%	-19%	-33%
polska	0%	-2%	-7%	-20%	-23%	-33%
nobel-us	0%	-1%	-4%	-16%	-20%	-33%
planar30	0%	-3%	-7%	-19%	-23%	-33%
atlanta	0%	-1%	-4%	-12%	-16%	-33%
france	0%	-3%	-8%	-18%	-22%	-33%
average	0%	-2%	-5%	-14%	-19%	-33%

Table 6: Decrease of the investment cost with respect to the cost for total protection, with Model 1 (only one shortest path).

	Total	1-ϵ				Deterministic
	protection	85%	50%	10%	5%	solution
pdh	0.00%	0.00%	0.00%	-4.37%	-1.92%	0.00%
diyuan	0.00%	-2.14%	-2.66%	-2.42%	-1.60%	0.00%
polska	-0.54%	-4.93%	-6.80%	-3.78%	-2.64%	0.00%
nobel-us	-0.62%	-4.23%	-6.09%	-4.45%	-3.40%	0.00%
planar30	-1.01%	-2.87%	-3.46%	-2.28%	-1.52%	0.00%
atlanta	-1.42%	-2.61%	-3.18%	-2.80%	-2.41%	0.00%
france	-7.08%	-8.81%	-8.80%	-6.57%	-5.60%	0.00%
average	-1.52%	-3.66%	-4.43%	-3.81%	-2.73%	0.00%

Table 7: Decrease of the investment cost with Model 2 (routings allowed on the 4 shortest paths), with respect to Model 1 (routing on the unique shortest path).

	Total	1-ϵ				Deterministic
	protection	85%	50%	10%	5%	solution
pdh	100.00%	100.00%	100.00%	52.50%	17.10%	100.00%
diyuan	100.00%	100.00%	98.20%	40.70%	15.90%	100.00%
polska	100.00%	99.70%	90.80%	22.80%	7.60%	100.00%
nobel-us	100.00%	99.90%	97.10%	37.10%	10.40%	100.00%
planar30	100.00%	100.00%	97.60%	28.70%	5.70%	100.00%
atlanta	100.00%	100.00%	100.00%	75.20%	42.70%	100.00%
france	100.00%	100.00%	99.30%	50.30%	10.00%	100.00%
average	100.00%	99.94%	97.57%	43.90%	15.63%	100.00%

Table 8: Percentage of simulated data for which ADR are sufficient to route flows. These results are obtained for Model 2 (all 4 shortest paths can support the solution).

	Total	1-ϵ				Deterministic
	protection	85%	50%	10%	5%	solution
`pdh`	0.00%	0.00%	0.00%	17.40%	10.50%	0.00%
`diyuan`	0.00%	0.00%	1.30%	35.10%	9.00%	0.00%
`polska`	0.00%	0.30%	9.10%	60.80%	57.50%	0.00%
`nobel-us`	0.00%	0.10%	2.90%	61.30%	77.30%	0.00%
`planar30`	0.00%	0.00%	2.40%	54.20%	36.60%	0.00%
`atlanta`	0.00%	0.00%	0.00%	24.70%	54.10%	0.00%
`france`	0.00%	0.00%	0.70%	48.80%	80.20%	0.00%
average	0.00%	0.06%	2.34%	43.19%	46.46%	0.00%

Table 9: Percentage of simulated data for which ADR do not enable to route flows, while there is a solution. These results are obtained for Model 2 (routings are allowed on the $k = 4$ shortest paths).

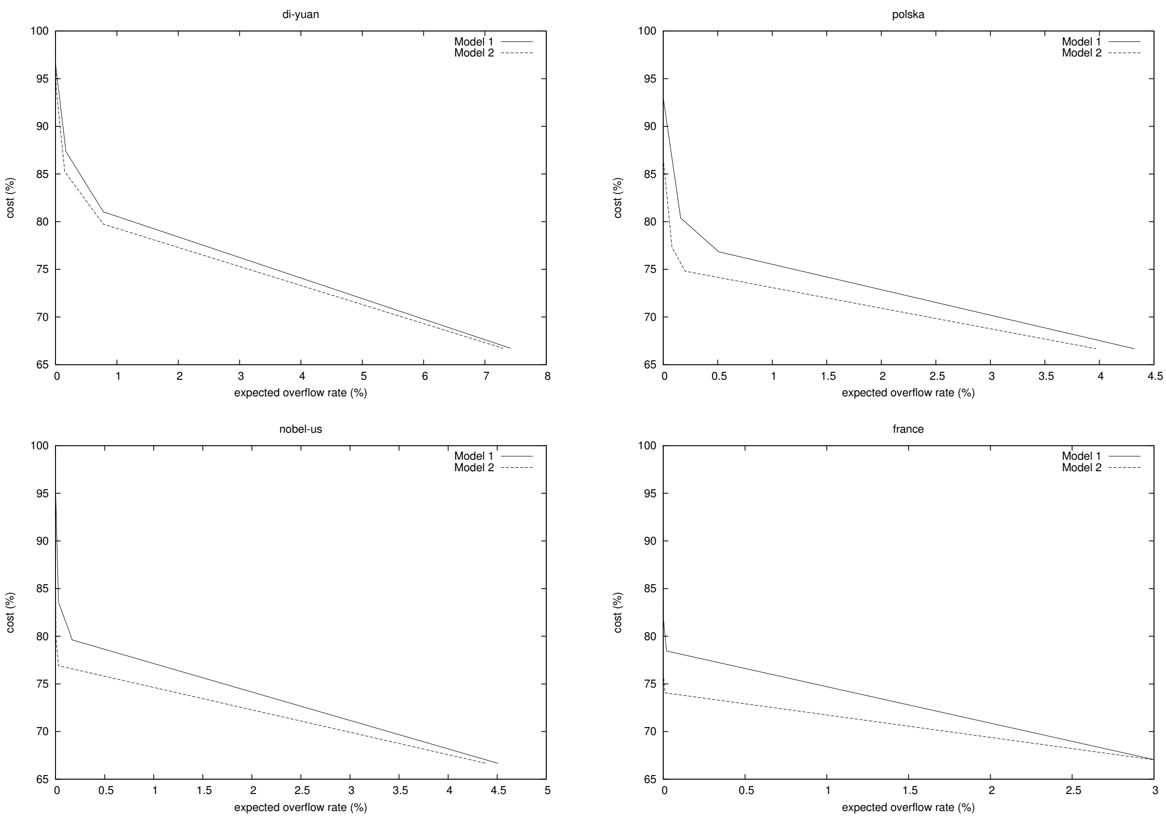

Figure 1: Evolution of the investment cost with the expected traffic loss for instances `diyuan`, `polska`, `nobel-us` and `france`. The 100%-cost corresponds to the case of total protection when routing for Model 1 (worst case).

posed methodology allows the decision maker to balance between the need of Quality of Service (QoS) and the investment costs. The main feature of the method is the use of a refined version of Affine Decision Rules (ADR) to model the recourse decisions, i.e., the flow routing, as linear functions of the revealed uncertainty. To reduce the dimension of the robust model induced by the ADRs, we used the concept of demand proximity. To calibrate the immunization factor κ in the equivalent robust counterpart of the uncertain constraints, we performed a probabilistic analysis under the assumption that each demand is distributed according to a triangular symmetric distribution. Finally we have tested our approach with numerical experiments on network instances from SNDlib and performed a validation analysis by simulation to study the quality of the solutions. The results obtained on the test problems suggest several interesting conclusions. First and contrarily to most of the classical approaches, the proposed robust formulations remain numerically tractable and even allow the introduction of integer constraints to bound the number of used paths. Second, although the introduction of refined ADR makes the recourse decisions very restrictive and conservative in theory, our numerical tests illustrate their relevance in practice. It appears that in most of the cases of practical significance, ADR are in fact not restrictive. The performances of the computed solutions are convincing; the investment cost savings are significant and the QoS very high.

Let us review now some directions for future research. First, we believe that the proposed methodology can be improved using the concept of globalized robustness [3] to control the magnitude of possible violations. Actually, our approach concentrates on solutions that remain feasible for all realizations within the uncertainty sets, but is silent about realizations that lie outside. Globalized robust optimization proposes an extension that admits possible constraint violations, but control their magnitude. Second it will be interesting to see whether our method performs well with others definitions of demand proximity for the ADR. In the present study, a demand is considered close to another if their respective shortest paths have at least one arc in common. We have tested without success some alternative definitions (using for example all admissible paths) but a more extensive study should be envisioned.

References

[1] A. Atamtürk, M. Zhang, "Two-stage robust network flow and design under demand uncertainty," *Operations Research*, vol. 55(4), pp. 662–673, 2007.

[2] W. Ben-Ameur, H. Kerivin, "Routing of uncertain demands", *Optimization and Engineering*, vol. 3, pp. 283-313, 2005.

[3] A. Ben-Tal, L. El Ghaoui, A. Nemirovski. *Robust Optimization.* Princeton University Press, 2009.

[4] A. Ben-Tal, A. Nemirovski, "Robust solutions of Linear Programming problems contaminated with uncertain data", *Mathematical Programming*, 88, pp. 411-424, 2000.

[5] D. Bertsimas, M. Sim, "The Price of Robustness", *Operations Research*, Vol. 52(1), pp. 35-53, 2004.

[6] F. Babonneau, J.-P. Vial, R. Apparigliato, *Handbook on "Uncertainty and Environmental Decision Making"*, chapter Robust Optimization for environmental and energy planning. International Series in Operations Research and Management Science. Springer Verlag, 2009.

[7] A. Ben-Tal, A. Goryashko, E. Guslitzer, A. Nemirovski, "Adjustable robust solutions of uncertain linear programs". *Mathematical Programming*, 99(2), pp. 351–376, 2004.

[8] GNU Linear Programming Kit (GLPK), version 4.11. http://www.gnu.org/software/glpk/ (2006).

[9] C. Lemaréchal, A. Ouorou, G. Petrou, "Robust Network Design in Telecommunications under Polytope Demand Uncertainty". *European Journal of Operational Research*, 206, pp. 634-641, 2010.

[10] A. Lisser, A. Ouorou, J.-Ph. Vial, J. Gondzio, "Capacity planning under uncertain demand in telecommunication networks", Technical Report 99.13, Logilab, Department of Management Studies, University of Geneva, Switzerland, 1999.

[11] E. Martins, M. Pascoal, "A new implementation of Yen's ranking loopless paths algorithm," *4OR*, vol. 1, pp. 121–133, 2003.

[12] E.A. Medova, J.E. Scott, "Evolving system architectures for multimedia network design", *Annals of Operations Research*, 104, pp. 163-180, 2002.

[13] K. Meesublak, "Network design under demand uncertainty", *Proc. of the Asia-Pacific Advanced Network Meeting*, 2008.

[14] M. Minoux, "Robust network optimization under polyhedral demand uncertainty is NP-hard," *Discrete Applied Mathematics*, vol. 158, pp. 597–603, 2010.

[15] S. Mudchanatongsuk, F. Ordónẽz, J. Liu, "Robust solutions for network design under transportation cost and demand uncertainty," *Journal of the Operational Research Society*, vol. 59, pp. 652–662, 2008.

[16] F. Ordóñez, J. Zhao, "Robust capacity expansion of network flows," *Networks*, vol. 50(2), pp. 136–145, 2007.

[17] A. Ouorou, "Robust capacity assignment in telecommunications", *Computational Management Science* vol.3, pp. 285–305, 2006.

[18] A. Ouorou, J.-P. Vial, "A model for robust capacity planning for telecommunications networks under demand uncertainty," in *Proceedings of the 6th International Workshop on Design and Reliable Communication Networks, DRCN 2007*, 2007.

[19] D. Paraskevopoulos, E. Karakitsos, B. Rustem, "Robust capacity planning under uncertainty," *Management Science*, vol. 37(7), pp. 787–800, 1991.

[20] G. Petrou, C. Lemaréchal, A. Ouorou, "An approach to robust network design in telecommunications", *RAIRO Operations Research*, vol. 41, pp. 411-426, 2007.

[21] S. Sen, R.D. Doverspike, S. Cosares, "Network planning with random demands", *Telecommunication systems*, vol. 3, pp. 11-30, 1994.

[22] SNDlib. http://sndlib.zib.de/.