Single-commodity network design with random edge capacities

Biju K. Thapalia and Teodor G. Crainic and Michal Kaut and Stein W. Wallace

in: European Journal of Operational Research. See also BibTeX entry below.

BibTeX:

```
@article{ThaWalKau10,
  author = {Biju K. Thapalia and Teodor G. Crainic and Michal Kaut and Stein W. Wallace},
  title = {Single-commodity network design with random edge capacities},
  journal = {European Journal of Operational Research},
  year = {2012},
  volume = {220},
  pages = {394--403},
  number = {2},
  doi = {10.1016/j.ejor.2012.01.026}
}
```

Single-commodity network design with random edge capacities

Biju K. Thapalia* Teodor Gabriel Crainic† Michal Kaut‡
Stein W. Wallace§

January 2012

Abstract

This paper examines the single-commodity network design problem with stochastic edge capacities. We characterize the structures of the optimal designs and compare with the deterministic counterparts. We do this partly to understand what constitutes robust network designs, but also to construct a heuristic for the stochastic problem, leading to optimality gaps of about 10%. In our view, that is a rather good result for problems that otherwise cannot be solved at all.

Keywords: Network flows, Single-commodity network design, Survivable networks, Edge failure, Correlations, Robustness

1 Introduction

Many important physical networks, such as distribution networks for water, oil and gas pipelines, road systems, or distribution channels are integral parts of our lives. These networks are made to last for a long time and are often subjected to daily routine operational decisions. If any parts of these networks are down, major portions of society will be affected. Focus on cost savings tends to make these networks sparser, and hence more vulnerable to disruptions, failures, maintenance-induced capacity reduction, congestion phenomena, etc. that may occur (see discussions in Ball et al. (1995) and Balakrishnan et al. (1998)). The owners of these networks must therefore design and maintain them, often under strict budgetary regimes, so that they work well even in the case of reduced capacity or broken links in the network. Hence, in our view, there is an increased need to understand what constitutes a good network design in light of random arc capacities modeling such uncertainties. In particular, we wish to see if it is important to use models explicitly expressing the randomness in capacities when designing the networks. And if the answer is yes, we would like to understand in what ways the designs from deterministic models fall short of designs from models explicitly considering the uncertainty. So, in our view, deterministic solutions are not just good or bad. Even if they are bad in their own rights, they may carry useful information.

Even deterministic network design problems of industrial size are very hard to solve. Stochastic problems of the same dimensions (in terms of the number of edges and nodes)

*Purbanchal University, Biratnagar, Nepal
†Université du Québec à Montréal (UQAM), Canada
‡SINTEF Technology and Society, Trondheim, Norway
§Lancaster University Management School, England

simply cannot be solved today, not even with decent heuristics. It is reasonable to assume that we shall always be able to solve larger deterministic than stochastic instances. Hence, we want to investigate the following very simple heuristic (which does *not* generally provide good solutions to stochastic integer programs): What if we decide which edges to open based on the deterministic design problem (hence fixing all discrete variables) and then use a stochastic continuous model to set capacities? Will that result in a good solution to the stochastic design problem? If it does, our ability to find good solutions to stochastic design problems will grow with our ability to solve the deterministic versions. That is certainly not the case today.

Good robust designs trade off deterministic initial costs versus expected future costs (or gains) in a good way. Most often, by increasing initial costs, the network is provided with more operational flexibility, and expected future costs decrease. (Note though, that at times, uncertainty induces lower initial investments. This may, for example, take the form of decision postponements.) What is more important to this paper, though, is that many networks, having more or less the same initial investment costs, may react very differently to disturbances. So even though it is clear that increasing the number of edges will increase the quality of a network design, we would like to understand more specifically what characterizes a good way to increase the number of edges. And, again, we would like to understand if deterministic models will guide us well in designing the networks, or if they will lead us astray.

There are many application areas where network design with random edge capacities are important. Possibly the most famous one is the design of survivable multi-commodity networks in telecommunications. Much of this work originates with Suurballe and Tarjan (1984). Water, oil, and gas distribution systems are other central cases. If we drop the pressure constraints, these pipeline design problems simplify and can be expressed as single commodity design problems with underlying continuous single commodity flow problems (Brimberg et al., 2003).

A water pipeline network services different consumers, of which some are very sensitive to disruption in supply, such as hospitals and certain industries. Disruption in the supply may arise due to pipeline ruptures, leakages or blockage, which affect the overall flow in the network. Many studies were performed on designing optimal and reliable water distribution systems (Chung et al., 2009; di Pierro et al., 2009; Montalvo et al., 2008). A seminal paper in gas distribution is by Rothfarb et al. (1970). The papers by De Wolf and Smeers (1996), Martin et al. (2006), and Brimberg et al. (2003) address design issues for oil and gas distribution networks. The work by Midthun et al. (2009) shows how the network structure and the physical properties impacts the operation and development of natural gas transportation networks.

In their work on fleet management (see, e.g., Cheung and Powell (1996)), Powell and his co-authors use random edge capacities in a single commodity network to represent random demand. Bounds on the recourse problem in this situation are discussed in Wallace (1987).

It is evident that a network with random edge capacities must function reasonably well in many situations with partial or full breakdown of capacities. A common way to investigate this situation is to perform single- or multi-parameter sensitivity analysis in order to understand how the optimal solution changes as a function of these breakdowns. This approach might seem appropriate, but in fact it is not. This is outlined in detail in Wallace (2000) and Higle and Wallace (2003). Logically, when performing sensitivity analysis, one is assuming that the design can be postponed until after breakdowns have become known. This is hardly ever pointed out, though. So, whether sensitivity analysis is performed or not, one ends up with a solution not created for robustness, and hence, may have to face difficult operational decisions when breakdowns occur.

Much of the literature in the field of survivable networks discusses various heuristics but does not address the resulting network structure. We know that a deterministic solution might perform very badly when used in a stochastic environment, that is, when subjected to the uncertainties that were suppressed when the deterministic model was solved, see for example Thapalia et al. (2012) for the case of random demand. The reason is simply that it is not made to handle uncertainties in a good way. This paper studies the structural differences between stochastic and deterministic designs, in order to understand what flexibility means in the optimal network structure for a single commodity flow problem with single or multiple sources and sinks. We also develop a simple heuristic for the stochastic case.

The remainder of the paper is organized as follows. Section 2 defines the problem and introduces the mathematical formulation. Section 3 describes the experimentation set-up and scenario generation. Section 4 presents the computational results with discussions and, finally, Section 5 concludes the paper.

2 Problem description and modeling issues

Given a set of nodes (made up of source, demand, and transshipment nodes) and a set of potential edges connecting these nodes, the *single-commodity network design problem with random edge capacities* aims to determine a subset of the edges to open and their capacities, so as to satisfy the demand, at demand nodes, at minimal cost, given the supply at the source nodes and the potential failures of the edges to provide the full installed capacity.

The design is based on minimizing the sum of the fixed costs of selecting edges connecting the nodes, the total costs to add capacity on the edges and flow the commodity on them, and the penalty costs for not satisfying demand. The last three terms are continuously linear. Not satisfying demand can have many interpretations, such as sending the flow at a later point in time, with another mode, or a straightforward rejection. We find it crucial to include the possibility of not satisfying all the demand, as it is unlikely to have a network which satisfies all demand in all situations. The same formulation is used in both the stochastic and deterministic models, to make the results comparable. We view supply as a capacity, and hence, do not consider unused supply as a problem.

When we wish to compare a stochastic network design model with its deterministic counterpart, we need to be careful about how we define the deterministic model. For random demand, this is not so difficult. If historical data is available, for example, demand will usually be the average observed demand (or possibly some other forecast value based on the history). Hence, it is not unreasonable to compare the stochastic model with a deterministic model where all demands are replaced with their mean values.

It is not quite as easy for the case of random edge capacities. If the starting point is the stochastic model, and we ask "What is the natural deterministic counterpart?", the answer is most likely a model where edge capacities are replaced by their means. But if the starting point is that of setting up a deterministic network design model (possibly realizing that edges might fail, but not wanting to model it), it is rather likely that edges will be treated with capacities equal to fully-operational values, that is, their maximal capacities from a stochastic perspective. One will argue: this edge costs a\$ and has a capacity of b. One will not use expected capacity taking possible failures into account. Consequently, in what follows, for each stochastic case, we shall consider two deterministic cases: average capacity and maximal (design) capacity.

The expected behavior of a deterministic solution can be made arbitrarily bad by setting the penalties for unsatisfied demand sufficiently high. However, showing that the deterministic solution is bad is not our goal, rather we wish to understand how the stochastic and deterministic solutions relate to each other. Hence, we have set the penalties at moderate levels, thereby facilitating the comparisons.

To define the mathematical formulation, let $G = (\mathcal{N}, \mathcal{E})$ be a network defined by a set $\mathcal{N}$ of n nodes, made up of sets $\mathcal{C}$ of source nodes, $\mathcal{D}$ of demand nodes, and $\mathcal{T}$ of transshipment nodes, and a set $\mathcal{E}$ of m edges (undirected arcs), where

$$\mathcal{E} \subset \{k = (i, j) : i \in \mathcal{N}, j \in \mathcal{N} \text{ and } i < j\}.$$

Each edge is indexed either by i, j or by k.

The randomness in the capacity of the edges is described by a set of scenarios $\mathcal{S}$, where each individual scenario $s \in \mathcal{S}$ has one capacity realization for each edge. We shall discuss in Section 3.2 the scenario generation procedure. Define the following parameter notation:

M_k upper bound on capacity installable on edge $k \in \mathcal{E}$;
R unit cost of unsatisfied demand;
P^s probability of scenario $s \in \mathcal{S}$;
C_k unit flow cost on edge $k \in \mathcal{E}$;
G_k fixed setup cost for edge $k \in \mathcal{E}$;
H_k unit capacity installation cost for edge $k \in \mathcal{E}$;
V_k initial capacity on edge $k \in \mathcal{E}$, if any;
D_i for $i \in \mathcal{D}$, demand at demand node i; $D_i < 0$;
for $i \in \mathcal{C}$, supply at source node i; $D_i > 0$;
Δ_k^s part of capacity on edge $k \in \mathcal{E}$ that works in scen. $s \in \mathcal{S}$, $0 \le \Delta_k^s \le 1$.

Define the decision variables

$x_k^s = x_{ij}^s$ flow on edge $k = (i, j) \in \mathcal{E}$, going in direction $i \to j$, in scen. $s \in \mathcal{S}$;
$z_k^s = z_{ij}^s$ flow on edge $k = (i, j) \in \mathcal{E}$, going in direction $j \to i$, in scen. $s \in \mathcal{S}$;
u_k new capacity installed on edge $k \in \mathcal{E}$;
e_i^s for $i \in \mathcal{D}$, unsatisfied demand at node i in scenario $s \in \mathcal{S}$;
for $i \in \mathcal{C}$, unused capacity of source i in scenario $s \in \mathcal{S}$;
y_k 1 if edge $k \in \mathcal{E}$ is selected, 0 otherwise.

We assume that total supply, coming from equally-sized source nodes equals the maximal demand in the network, so that

$$D_i = -\sum_{j \in \mathcal{D}} D_j \,/\, |\mathcal{C}| \tag{1}$$

for all $i \in \mathcal{C}$, where $|\mathcal{C}|$ is the number of source nodes. By letting supply come from a set of equally-sized nodes, we avoid, as much as possible, that variation in supply affects the interpretation of results, which, after all, should concern edge capacities. The model hence is:

$$\min \sum_k G_k y_k + \sum_k H_k u_k + \sum_s P^s \left\{ \sum_k C_k \left(x_k^s + z_k^s\right) + R \sum_{i \in \mathcal{D}} e_i^s \right\} \tag{2}$$

Subject to:

$$\sum_{j:\,(ij)\in\mathcal{E}} \left(x^s_{ij} - z^s_{ij}\right) - \sum_{j:\,(ji)\in\mathcal{E}} \left(x^s_{ji} - z^s_{ji}\right) = \begin{cases} 0 & \forall i \in \mathcal{T},\ \forall s \in \mathcal{S} \\ D_i - e^s_i & \forall i \in \mathcal{C},\ \forall s \in \mathcal{S} \\ D_i + e^s_i & \forall i \in \mathcal{D},\ \forall s \in \mathcal{S} \end{cases} \tag{3}$$

$$x^s_k + z^s_k \leq \Delta^s_k \left(u_k + V_k\right) \qquad \forall k \in \mathcal{E},\ \forall s \in \mathcal{S} \tag{4}$$

$$u_k \leq M_k\, y_k \qquad \forall k \in \mathcal{E} \tag{5}$$

$$0 \leq e^s_i \leq |D_i| \qquad \forall i \in \mathcal{C} \cup \mathcal{D},\ \forall s \in \mathcal{S} \tag{6}$$

$$x^s_k, z^s_k, u_k \geq 0 \text{ and } y_k \in \{0,1\} \qquad \forall i \in \mathcal{C} \cup \mathcal{D},\ \forall k \in \mathcal{E},\ \forall s \in \mathcal{S} \tag{7}$$

The objective function (2) minimizes the total cost of the network. The first part is the fixed setup cost for all new edges, the second part the costs of adding the new capacities, the third part is the expected total cost of the flows through the edges, and the penalty for unsatisfied demand. Constraints (3) model conservation of flow at nodes for all scenarios. The left-hand side is the net outflow from node i, which must be zero for transshipment nodes $\mathcal{T}$, equal to the used capacity for source nodes $\mathcal{C}$, and equal to the satisfied demand for demand nodes $\mathcal{D}$ (D_i is negative in this case).

Constraints (4) represent the flow limit in each edge. The left hand side of the equation is the net flow on edge k, which should be less than or equal to the total capacity of the edge. Since we do not start with any initial existing capacity, we always have $V_k = 0$. Note that, in an optimal solution, an edge will never have flow in both directions. Constraints (5) show that new capacity u_k can be added up to M_k, but only if edge k is built. Constraints (6) bound the unused supplies and rejected demands and, finally, (7) ensure that all variables are non-negative and the edge selection decisions binary.

For the deterministic counterpart (as mentioned in Section 2) we replace the stochastic edge capacities by their expectations and their maximal values, yielding two deterministic cases.

3 Experimentation and scenario generation

We first discuss the test instances and their sources before turning to scenario generation and the question of stability relative to the chosen scenarios. Our experiments are designed to achieve two goals: Firstly, we wish to understand how deterministic designs perform in stochastic environments and to what extent information from deterministic designs are useful for the stochastic problem, partly to see if they can be used to generate heuristics for the stochastic case. Secondly, we aim to characterize the stochastic designs, so that we can qualitatively describe good designs and use this knowledge to evaluate a given design without making any serious calculations.

3.1 Test instance generation

We took five networks also used in Thapalia et al. (2011, 2012). The instances named Germany, Nobel-EU, France, and Pdh are telecommunication networks from the SNDlib library

Table 1: The test instances.

Problem name	# nodes	# edges	# dem. n.	# src. n.	# tests
Germany_SS	29	48	10	1	4
Germany_MS	29	48	10	3	4
France_SS	16	30	10	1	4
France_MS	16	30	10	3	3
Montreal_r06.1_SS	10	38	5	1	4
Montreal_r06.1_MS	10	38	7	3	3
Pdh_SS	11	30	7	1	4
Pdh_MS	11	30	7	4	4
Nobel-EU_SS	28	38	8	1	4
Nobel-EU_MS	28	38	8	4	4

(Orlowski et al., 2010), while Montreal_r06 is obtained from CIRRELT (Interuniversity Research Centre on Enterprise Networks, Logistics and Transport), Montreal. The names of the test networks do not mean anything in our computational setup.

In total, 76 test instances were constructed using these five networks, 40 single-source and 36 are multi-source. The instances were created in the following way: for each of the five networks, we created single-source and multi-source test cases by selecting one or multiple source nodes. This yielded 38 instances of which 20 are single-source and 18 are multi-source. Since we know that correlations may play important roles in the design of a network, we created positively correlated and uncorrelated cases for each instance. In the case of positive correlations, adjacent edges were given correlations of 0.5, while edges separated by one edge were given correlations of 0.2. Edges which are separated by two edges have a correlation of 0.1. This is an intuitive setting for representing natural calamities. Whenever one edge is hit hard, there is a chance that nearby edges are hit as well, see, e.g., Chen et al. (2002).

It is worth noting that, originally, these instances were multi-commodity network design problems. Therefore, to adapt them to the single-commodity context, we only kept the node coordinates (where available) and the fixed setup edge costs. The values for unit flow and capacity costs were fixed proportional to the Euclidean distance between the edge node. The cost of unfulfilled demand was derived for each instance using a multiple of the highest value among all edge costs in the network. We made sure, however, that the penalty costs are not driving the solutions. The results in the first part of Section 4.1 are based on instances with these cost structures. The Montreal instance does not have node coordinates. We therefore used Graphviz (Gansner and North, 2000) to draw the graph using the fixed setup costs as distance measure. The graphs of the instances coming from Nobel-EU are planar, whereas the other graphs are non-planar. The test instances are described in Table 1.

We selected four potential source nodes for instances with single-source networks. When one of them is a source, the others are transshipment nodes. We selected four (three for Montreal_r06 and France) sets of source nodes for the multi-source networks to create four instances from each network. The number of source nodes for each case is listed in the fifth column of Table 1. In the pictures that follow, the distance between two nodes reflects not only the actual distance, but also the levels of the unit capacity and flow costs. If an edge is twice as long as another, it is also twice as costly with respect to these two costs.

Given the difficulty of solving the stochastic network design problem to optimality, we kept n (the number of nodes) below 30 and m (the number of edges) below 50. Using AMPL and solving to optimality using CPLEX 9.0, the solution time varies from a few seconds to 5 hours depending on the instance, on a PC with 3 GHz Intel® CPU and 8 GB of RAM. Hence, these instances are the largest we can handle if we want optimal solutions to the stochastic formulations.

3.2 Scenario generation and stability tests

Stochastic programs need discrete probability distributions. A scenario is a vector of length m containing a possible capacity for each edge. We have created scenarios with equal probabilities of occurring, using a variant of the moment-matching method from Høyland et al. (2003).

Lacking specific knowledge, we have chosen a triangular distribution on the $[0, 1]$ interval, with mode at one, which gives an expected value of 0.67. Note that a mode below one would imply that the edge (in the continuous case) has an extremely low probability of being close to fully operational. We feel the chosen distribution is a reasonable description of edge failures.

The decision on the number of scenarios used to represent the stochastics is critical as we want to be sure we study the effects of randomness on our model, and not some random side-effect of the scenario generating procedure. For a given scenario generation procedure, there is normally a trade-off between the number of scenarios and the time needed to solve the stochastic program to optimality. The task is thus to find the smallest number of scenarios that still gives solutions that are both in- and out-of-sample stable, in the sense described in Kaut and Wallace (2007).

We ran our in-sample stability test with different numbers of scenarios and ended up with 200, considering the solution time and stability. The deviation (measured by standard deviation of the objective values of all runs divided by the mean of the objective values) in all cases is less than 1% except for the case of Montreal_r06, where it is 1.5% for single-source cases and 2% for multi-source cases. Out-of-sample stability tests, using a reference tree with 2000 scenarios, are all within 1%. With these values, we are satisfied that we have stability.

3.3 Comparison tests

As outlined in the Introduction, the deterministic solution, by construction, has a worse expected behavior than its stochastic counterpart. However, we would like to understand more about why this is the case, and in what sense it is worse. This is partly motivated by what we found in Thapalia et al. (2011, 2012), where we discussed random demand: the edges (if not their capacities) from the deterministic solution provided a good starting point for the stochastic case. This is unusual for stochastic integer programs.

In order to check the quality of the deterministic designs, as well as compare them to the stochastic ones, we have set up two tests, named *comparisons*. Whenever a comparison is performed, we take the deterministic and stochastic designs (or parts thereof) and evaluate them using reference trees – in our case trees with 2000 scenarios, to make sure we have good approximations of the true distributions. The costs from the design and evaluation phases are added up, making the reported costs comparable across all tests.

A The classical test where the whole first-stage solution is evaluated out-of-sample. This amounts to solving a 2000-scenario stochastic program with all first-stage variables (designs and capacities) fixed, so in fact this equals the solution of 2000 independent second-stage problems. Since the second stage does not involve any integer variables, this is very fast.

B Only information on which edges should be opened is imported from the first stage. Then, in a 2000-scenario stochastic program, all discrete variables y describing opened and closed edges—we call the resulting network a *skeleton*—are fixed and the stochastic program is run. Hence, the model is allowed to install any capacity on the opened edges (also lower than in the deterministic case), but not to open new ones.

Applied on the deterministic solution, Comparison A tests the quality of the deterministic solution. The purpose of Comparison B is to check if the deterministic solution possibly has a good structure, but badly chosen capacities (typically too low). If this is the case, to find a reasonably good solution to the stochastic design problem, we first find the skeleton, using a deterministic model, and then set capacities using a stochastic *continuous* model. This way, we can find good solutions to the stochastic design problem whenever we can solve the deterministic one. This is not true with today's technology.

In what follows, we discuss the major findings, details are given in the Appendix. Our first need is to understand the relationship between the stochastic and deterministic solutions.

For each of the 38 instances, we solve two deterministic problems, one with full capacity available and the other with the mean (in our case 67% of full) capacity on each edge. These choices were motivated in Section 2. In addition, we solve two stochastic versions of each instance, one with uncorrelated and one with positively correlated edge failures. For each stochastic version, we take the solution of the stochastic model and the two deterministic solutions and evaluate them out-of-sample on a reference tree, i.e., a tree with 2000 scenarios and the same correlations as those used to solve the stochastic programs.

Our measure of the quality of a solution (or a partial solution like a skeleton) is the ratio between the expected costs using the deterministic solution and the expected costs using the stochastic solution. As the expected costs are never close to zero in our problems, there is no danger of running into issues amounting to a division by zero. Note that, since both the stochastic and deterministic solutions are evaluated out-of-sample, the ratio might become slightly smaller than 1.

We also want to explore the relationships between the capacity building cost and the performance of the deterministic skeleton in the stochastic environment. The hypothesis is that as the capacity costs increase, the stochastic skeleton will look increasingly like the deterministic one (which is a tree) due to the cost of opening more capacity than what is absolutely necessary. Similarly, as the capacity costs costs increase, the deterministic skeleton tends toward the stochastic one, in terms of the number of open edges, as it is (relatively speaking) governed more and more by installed capacity and less and less by the number of opened edges. In other words, we postulate that as costs of installing capacity decrease, the expected behavior of the deterministic skeleton in a stochastic setting becomes increasingly bad, and this is true for both multi- and single-source cases.

To achieve this, we started with the instances identified above, which we define as base cases. From each base case, we built five additional instance by using 33%, 66%, 133%, 166%, and 200% of the unit capacity installation cost of the base case.

Certainly, the mean-capacity deterministic problem is equivalent to the full-capacity deterministic problem with $1/0.67 = 1.5$ times higher capacity costs. Consequently, this parametric analysis of the capacity costs contains the analysis of the relationship between the two de-

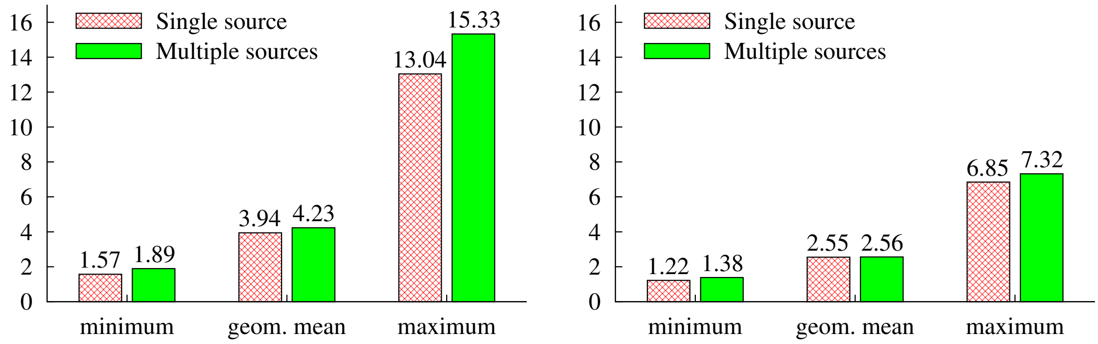

Figure 1: Results of the Comparison A. Quality of the solutions to the deterministic problems with full (left) and mean (right) edge capacities, measured as a ratio of their out-of-sample costs relative to the costs of the corresponding stochastic solutions.

terministic cases. However, to keep the interpretations apart, we have chosen this approach instead of reading one set of deterministic results from within the results of the other.

4 Computational results

We present and discuss the computational results of the experiments indicated above. The full results can be found in A, where Tables 3 and 4 include results presented in Figures 1 and 2, while Table 5 corresponds to Figure 3.

4.1 Inheritance from the deterministic solutions

Figure 1 presents results of Comparison A for the two deterministic solutions, split between single- and multi-source cases. The figure shows that both deterministic designs are bad, but differences may be observed, When mean capacities are used, results are better (around 2.5 times higher costs on average for both the single- and multi-source cases) than when the full capacities are used (around 4 to 4.25 times higher costs on average). The reason is simply that when mean capacities are used, the edges seem to have less capacity, and hence more is installed. Moreover, as we have observed in earlier papers – and that is confirmed here –, deterministic designs do not only suffer in terms of structure (skeletons) but also in terms of too low capacities. Therefore, doing what many practitioners do, namely running deterministic models with a pessimistic view on edge capacities, is indeed a good idea.

Figure 2 shows that when we use the deterministic skeleton and apply a stochastic program to set capacities, the results are rather good, implying that the skeletons perform quite well. We can also see that the multi-source cases do better than the single-source ones. This observation holds for most of the tested levels of capacity costs, as shown in Figure 3. As the capacity installation costs increase, the loss of using a deterministic skeleton decreases, and more so for the multi-source cases. This is true for both maximal and mean value edge capacities. This observation confirms that the quality of a design is a function of both connectivity and capacity. When the installation cost is low, the deterministic designs are guided by the shortest routes in terms of fixed costs of opening the edges, and the resulting skeletons perform relatively badly in a stochastic environment. But when the capacity costs increase, the deterministic designs change. They are no longer primarily guided by the shortest

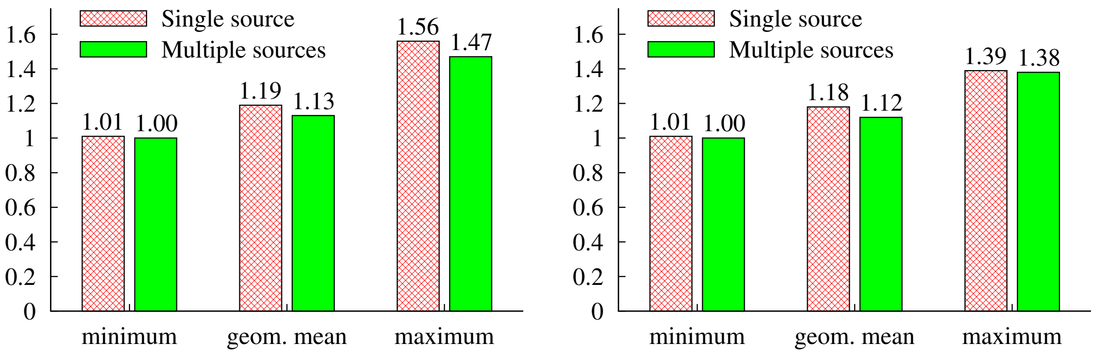

Figure 2: Results of the Comparison B. Quality of the solutions to the deterministic problems with full (left) and mean (right) edge capacities, measured as a ratio of their out-of-sample costs relative to the costs of the corresponding stochastic solutions.

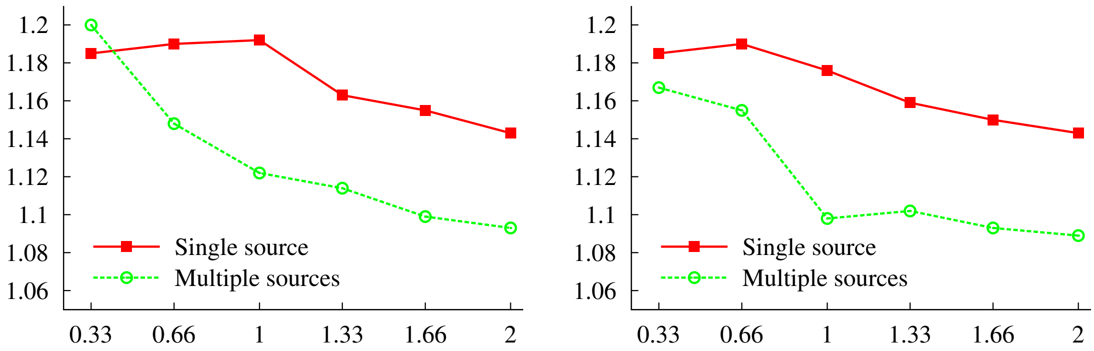

Figure 3: Graph of Comparison B test values for different variable setup cost. The X-axis shows the percentage of variable setup cost as compared to the base case and the Y-axis the quality of the solutions to the deterministic problems with full (left) and mean (right) edge capacities, measured as a ratio of their out-of-sample costs relative to the costs of the corresponding stochastic solutions.

routes in terms of fixed edge selection costs (mostly implying as few edges as possible), but also edge capacity-installation costs. This results in more edges being opened as total installed capacity, rather than the number of edges with capacity is the primary driver of costs. These skeletons, when used in a stochastic environment, perform better as they contain more connections. In the multi-source case, this happens more quickly because there are more edges than in the single-source case. Since the demand does not vary within each small tree in the forest, contrary to the case with random demand, the fact that the trees effectively cut the design into smaller parts is not a problem. Therefore, what is needed here is high density (many paths) and enough installed capacity within each tree to make sure the demand within the tree is satisfied with a high probability so as to achieve low penalty costs.

What we see from this analysis is that, if the observations carry over to larger cases, something that, as usual, cannot be tested, we now have a heuristic for the stochastic case: First solve the deterministic problem to optimality, then use a stochastic continuous program to set capacities. That will result in an optimality gap of about 10% in the better cases. Although 10% is far from ideal, it is better than any other method that we are aware of, for problems of this size and similar computation efforts. Such a heuristic would work as long as

Table 2: Average number of edges and average edge capacity for all tests in each single-source (SS) and multi-source (MS) case, for various correlation values (ρ).

	Number of edges			Capacity per edge		
	det.	$\rho \geq \mathbf{0}$	$\rho = \mathbf{0}$	det.	$\rho \geq \mathbf{0}$	$\rho = \mathbf{0}$
France_SS	11.75	16.25	18	624	1096	843
Germany_SS	16.75	28.75	32.75	7	5	4
Montreal_r06_SS	5.5	9.5	10	104	162	111
Nobel-EU_SS	13.62	25.25	29	28	24	18
Pdh_SS	7.37	9.5	10.5	131	186	154
France_MS	13	17	17.67	475	745	646
Germany_MS	19	32.25	36	5	4	3
Montreal_r06_MS	7	9	9.33	53	188	130
Nobel-EU_MS	15.75	23	25.5	13	19	15
Pdh_MS	7.89	8.75	9	87	170	159

we can solve the deterministic problem to optimality and take only marginally more time for the continuous stochastic program.

The heuristic could be applied to even bigger problems if we address the deterministic problem by some heuristic, but we know little about the quality of the result in that case. On the other hand, there is no feasible alternative. Note that this approach, though in principle available for any stochastic mixed integer program, normally does not deliver at all.

4.2 Structural characteristics

This section examines the structures of the deterministic and stochastic network designs from the tests mentioned in Section 3.1, and focuses on a few important observations shedding light on the characteristics of the stochastic designs under random edge capacities. It will be observed that these characteristics, once presented, are rather obvious. We see that as necessary for the characterization to be useful. A property that does not appear obvious or natural ex-post, is hardly useful since that would imply it did not teach us anything. So, it is our view that although our observations are obvious in this sense, they still provide understanding, which is one of the goals of this paper.

Network density

The second, third, and fourth columns of Table 2 show that stochastic designs have more edges than their deterministic counterparts. Similarly, columns five, six, and seven show that installed edge capacity is higher in the stochastic designs except for the cases of Germany (both for multi- and single-source case) and Nobel-EU_SS. The total installed capacity in the stochastic designs is higher in all cases, however. The reasons for this are that, firstly, with more edges in the network, there are alternative ways to reach demand nodes in the event of reduced edge capacities, and, secondly, higher installed capacities ensure that there are reasonably high capacity paths reaching the demand nodes even when there are edge failures.

We can also see from Table 2 (third and fourth columns) that, designs for the uncorrelated cases generally have more edges than the corresponding positively correlated cases. This can

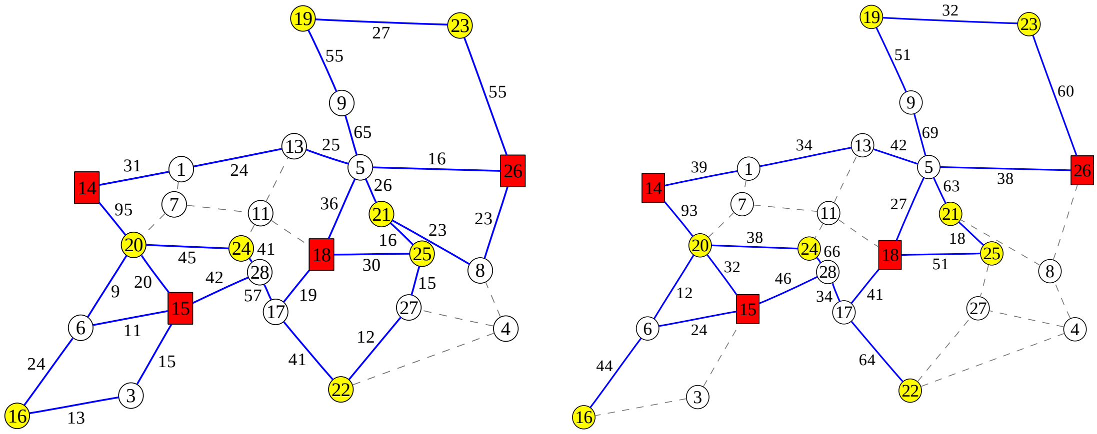

Figure 4: Network Density. Uncorrelated (left) and positively correlated (right) stochastic structure of Nobel-EU showing higher number of edges for the uncorrelated case and generally higher installed capacities on edges for the positively correlated case.

be explained by the fact that, in the uncorrelated case, it is very useful to have alternate paths by setting up extra edges. Then, when one path leading to a demand node has reduced capacity, another might work well — capacities are uncorrelated. In the positive correlation case, all edges incident to a node are positively correlated, so even though there may be many paths, they will tend to have difficulties at the same time. This reduces the value of alternative paths. The positive correlation cases compensate by installing more capacity on the selected edges (sixth and seventh columns of Table 2). The latter is of course a function of how we defined edge failures – as a percentage of installed capacity. These effects can be seen in Figure 4, where the uncorrelated case has more edges than the positively correlated one and the positively correlated case generally has higher installed edge capacity. In the figure, solid (blue) edges are installed with the given capacities, dotted edges are not installed. The dark (red) nodes are the source nodes, the white ones transshipment nodes. The shaded (yellow) nodes are demand nodes. The same color scheme is followed throughout the paper.

Alternative paths

We observe the creation of alternative paths in the stochastic network structures, even for the case of positively correlated edge failures. This is due to the fact that, with alternative paths, the network increases the chance that demand is at least partly satisfied even when one of the paths fails or works at low capacity. Figure 5 illustrates this observation. Demand node 16 in the stochastic structure is served by three paths, one approaching via transshipment node 19, one via transshipment node 6, and finally one via a collection of other demand nodes (through node 17). These paths may help fulfill demand at node 16 when one of them is down or has low capacity. Even though this is less useful when failures are positively correlated, the alternative paths still provide some extra chance of reaching demand node 16.

We observed, when studying optimal designs under random demand (see Thapalia et al., 2011, 2012), that consolidation was the major tool for hedging against uncertainty. By having several demand nodes share paths, one node could use the path when the others do not need it. With random edge capacities, this need does not emerge since demand is known. Hedging

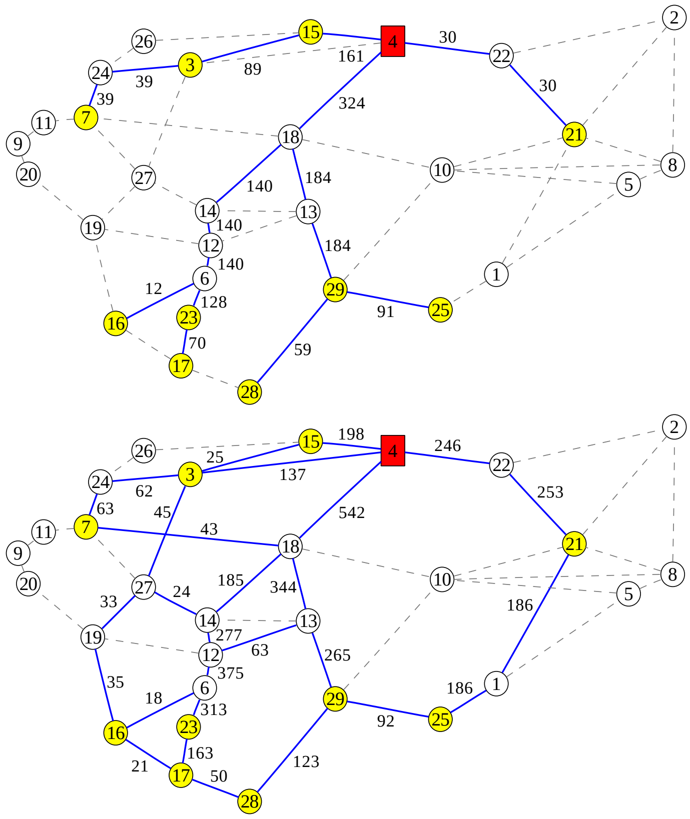

Figure 5: Deterministic (top) and stochastic (bottom) network structures for Germany_SS_04, showing alternative paths in the stochastic structure.

comes instead from having alternative paths, and of course, generally higher installed capacity since edges may fail. Consolidation-like phenomena are hence observed in some cases, but they come from the same phenomena as in deterministic cases: paths to demand nodes share edges (even if the paths become slightly longer) so that fewer fixed selection costs need to be paid, and, of course, two shortest paths (including edge selection and capacity installation costs) may simply happen to share edges.

Cycles

The formation of cycles is quite visible in networks designed for stochastic edge capacities. Cycles are formed among the demand nodes, including or excluding the source nodes, or by joining the leaves of the trees. When we compare this with the networks for stochastic demand, it is far more prominent here. The main reason for cycle formation is the need to provide alternate paths to fulfill demand when some edges are (partly) down. Cycles have the advantage that they can be used both ways. So somewhat high capacities (which characterizes the stochastic designs) combined with cycles provide alternative paths to demand nodes. We can see this in Figure 5 where cycles are seen in the stochastic network design, but not in the deterministic structures.

Removing edges

It is observed from the test results that in almost all cases, the stochastic skeleton contains the deterministic one. The additional edges are providing flexibility to the network structure. However, as we increase the fixed edge selection cost or the capacity installation costs compared to the base case, the additional edges, which were seen in the stochastic skeletons, disappear and finally very few are left. We can observe this in Figure 6. As we increase the fixed costs, keeping all other costs the same (left drawings), more of the edges that were specific to the stochastic solution disappear, making the stochastic design closer to the deterministic one (top structure of Figure 5). The first edges and partial paths to disappear are those with small capacities, such as 16–17, 18–10–21, or 3–27–19–16. The same effect is seen when increasing the capacity installation costs (right-column drawings).

5 Conclusions

We have seen that optimal stochastic designs for both the single- and multi-source cases differ from the deterministic ones. The flexibility, which gives the stochastic designs better expected performance, comes from a higher number of edges and higher installed capacities. With a higher number of edges there exist more paths to demand nodes and hence it also becomes easier to find alternative routes to the demand nodes in the case of edge failures. So while the sharing of paths is the main vehicle for hedging in the case of random demand, here it comes from providing alternative paths from sources to demand nodes.

The deterministic solutions as such are not good in the stochastic setting as their expected performance is bad. But borrowing the skeleton from the deterministic solution is rather good. With increased costs for adding edges or for adding capacities in the edges, the stochastic skeletons start to look more and more like the deterministic ones. Thus, for cases of this type, using a deterministic method to set the skeleton, and solving a stochastic continuous program to set capacities, is a promising heuristic resulting in optimality gaps of about 10%.

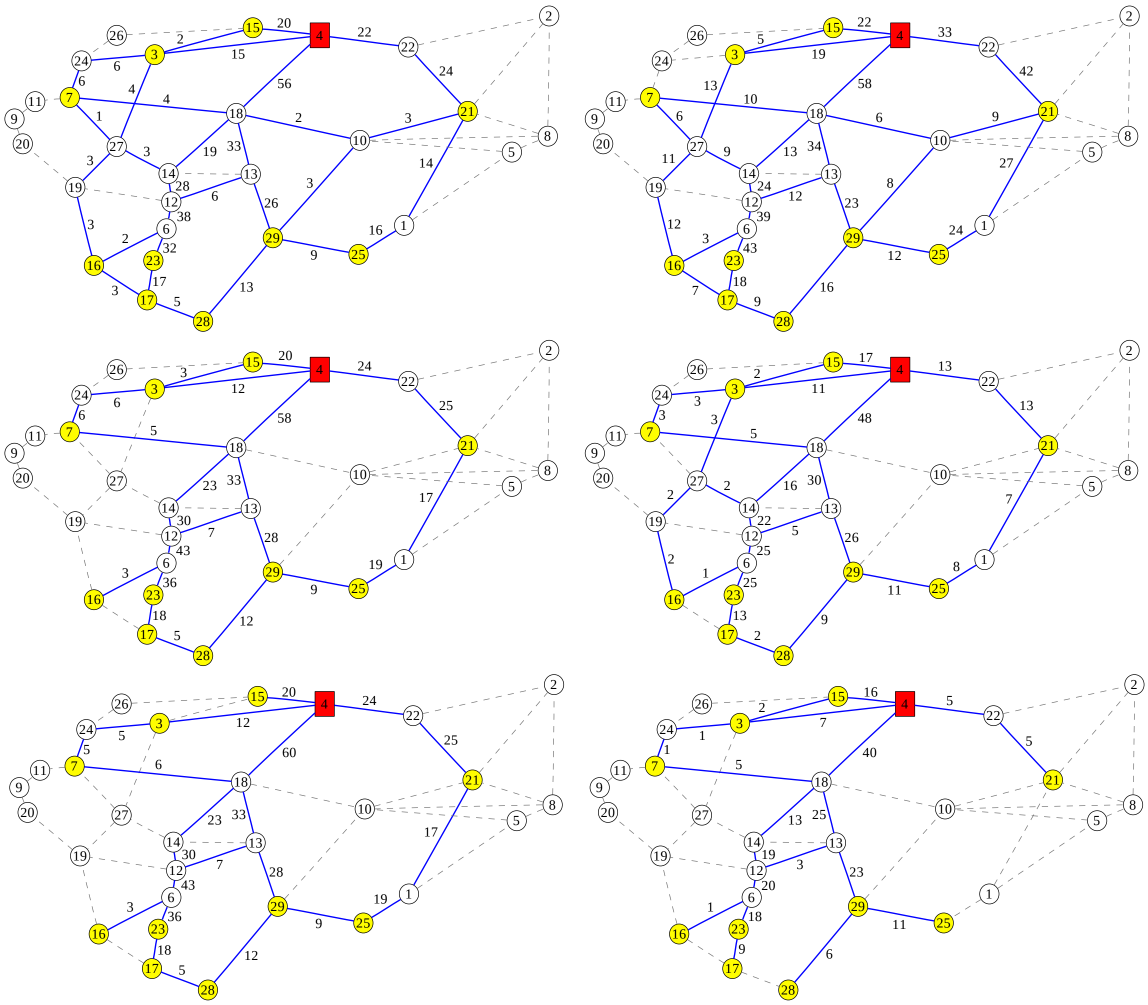

Figure 6: Disappearing Edges. Stochastic structures of Germany_SS with increasing fixed edge selection costs (left) and with increasing capacity installation costs (right) showing that, as the costs increase, the stochastic designs look increasingly like the deterministic ones. (Note that all figures are divided by 10.)

Naturally, we cannot test this for larger cases, since we cannot solve the stochastic versions to optimality.

Using the deterministic skeleton is slightly better if based on average edge capacities rather than maximal ones. The reason is somewhat subtle: Using average rather than maximal edge capacities is equivalent to increasing capacity installation costs. That reduces the importance of the fixed edge selection costs, generally leading to more edges being opened, and hence a better starting point for the stochastic linear program.

Correlations have important effects on the structure of the design. With uncorrelated edge failures, the stochastic designs have more edges than when edge failures are positively correlated. With positively correlated edge failures, the networks have higher installed capacities. The reason is simply that with positively correlated failures, all paths to a node tend to have difficulties at the same time, providing less hedging from multiple paths.

Cycles are present in the stochastic networks due to a combination of two phenomena.

The first is the one we observe for consolidation in the deterministic problem: avoid paying too many fixed costs. The second is the characteristics of a ring network. It provides two connections between any pair of nodes in the ring, and the ring can be used in both directions. For these reasons, cycles are much more prominent here than with random demand.

So, network designs for stochastic edge capacities are fundamentally different from network designs for stochastic demand. With stochastic edge capacities there are more edges, more cycles, and more installed capacities as compared to the design for stochastic demand. A major reason is that there is less consolidation. For stochastic edge capacities we only see consolidation of the type we see in deterministic designs, mostly caused by savings in the fixed edge selection costs. Instead, alternative connections become more important as a hedge against edges having reduced capacities. Skeletons generally do better here than with random demand as trees in the forests – typical for deterministic designs – no longer have the need to contact each other when randomness strikes, as each tree has enough supply.

Acknowledgments

This project was supported in part by grant 171007/V30 from The Research Council of Norway. While working on this project, T.G. Crainic was the NSERC Industrial Research Chair on Logistics Management, ESG UQAM, and Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal, and the Department of Economics and Business Administration, Molde University College, Norway. Partial funding for this project has been provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Industrial Research Chair and Discovery Grants programs.

References

A. Balakrishnan, T. Magnanti, and P. Mirchandani. Designing hierarchical survivable networks. *Operations Research*, 46(1):116–136, 1998.

M. O. Ball, C. J. Colbourn, and J. S. Provan. Network reliability. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Models*, volume 7 of *Handbooks in Operation Research & Management Science*, chapter 11. North-Holland, Amsterdam, 1995.

J. Brimberg, P. Hansen, K.-W. Lih, N. Maldenović, and M. Breton. An oil pipeline design problem. *Oper. Res.*, 51(2):228–239, March-April 2003.

A. Chen, H. Yang, H. K. Lo, and W. H. Tang. Capacity relaibility of a road network: an assessment methodology and numerical results. *Transportation Research Part B: Methodological*, 36(3):225–252, 2002. doi:10.1016/S0191-2615(00)00048-5.

R.-M. Cheung and W. Powell. An algorithm for multistage dynamic networks with random arc capacities, with an application to dynamic fleet management. *Operations Research*, 44 (6):951–963, 1996.

G. Chung, K. Lansey, and G. Bayraksan. Reliable water supply system design under uncertainty. *Environmental Modelling & Software*, 24:449–462, 2009.

D. De Wolf and Y. Smeers. Optimal dimensioning of pipe networks with application to gas transmission networks. *Operations Research*, 44(4):596–608, 1996. doi:10.1287/opre.44.4.596.

F. di Pierro, S.-T. Khu, D. Savić, and L. Berardi. Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms. *Environmental Modelling & Software*, 24:202–213, 2009.

E. R. Gansner and S. C. North. An open graph visualization system and its applications to software engineering. *Software: Practice and Experience*, 30(11):1203–1233, 2000. doi:10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.0.CO;2-N.

J. L. Higle and S. W. Wallace. Sensitivity analysis and uncertainty in linear programming. *Interfaces*, 33:53–60, 2003.

K. Høyland, M. Kaut, and S. W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2–3):169–185, 2003. doi:10.1023/A:1021853807313.

ILOG. *CPLEX 9.0 User's Manual*. ILOG, S. A., 2003.

M. Kaut and S. W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271, 2007.

A. Martin, M. Möller, and S. Moritz. Mixed integer models for the stationary case of gas network optimization. *Mathematical Programming*, 105(2–3):563–582, Feb 2006. doi:10.1007/s10107-005-0665-5.

K. T. Midthun, M. Bjørndal, and A. Tomasgard. Modelling optimal economic dispatch and system effects in natural gas networks. *The Energy Journal*, 30(4):155–180, 2009. doi:10.5547/ISSN0195-6574-EJ-Vol30-No4-6.

I. Montalvo, J. Izquíerdo, R. Pérez, and M. M. Tung. Particle swarm optimization applied to the design of water supply systems. *Computers and Mathematics with Applications*, 56: 769–776, 2008.

S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly. SNDlib 1.0 – Survivable Network Design library. *Networks*, 55(3):276–286, 2010. doi:10.1002/net.20371.

B. Rothfarb, H. Frank, D. M. Rosenbaum, K. Steiglitz, and D. J. Kleitman. Optimal design of offshore natural-gas pipeline system. *Operations Research*, 18:992–1020, 1970.

J. Suurballe and R. E. Tarjan. A quick method for finding shortest pairs of paths. *Networks*, 14:325–336, 1984.

B. K. Thapalia, T. G. Crainic, M. Kaut, and S. W. Wallace. Single-commodity stochastic network design with multiple sources and sinks. *INFOR*, 49(3):193–211, 2011. doi:10.3138/infor.49.3.003.

B. K. Thapalia, T. G. Crainic, M. Kaut, and S. W. Wallace. Single source single-commodity stochastic network design. *Computational Management Science*, 9(1):139–160, 2012. doi:10.1007/s10287-010-0129-0. Special issue on 'Optimal decision making under uncertainty'.

S. W. Wallace. A piecewise linear upper bound on the network recourse function. *Mathematical Programming*, 38:133–146, 1987.

S. W. Wallace. Decision making under uncertainty: Is sensitivity analysis of any use? *Operations Research*, 48(1):20–25, 2000.

A Results of the numerical tests

This appendix provides detailed results from the tests in Section 3.1.

Table 3: The ratios for the single source cases corresponding to Figures 1 and 2, split by correlation structure.

	Full capacity				Mean value capacity			
	comp. A		comp. B		comp. A		comp. B	
Test Name	$\rho = 0$	$\rho \geq 0$	$\rho = 0$	$\rho \geq 0$	$\rho = 0$	$\rho \geq 0$	$\rho = 0$	$\rho \geq 0$
Germany_SS_04	1.76	1.62	1.20	1.11	1.39	1.26	1.18	1.09
Germany_SS_10	1.71	1.57	1.17	1.08	1.34	1.22	1.17	1.08
Germany_SS_13	2.12	1.92	1.20	1.10	1.57	1.39	1.20	1.10
Germany_SS_27	1.99	1.81	1.21	1.10	1.55	1.37	1.21	1.10
France_SS_06	6.31	5.46	1.27	1.13	3.15	3.23	1.27	1.13
France_SS_10	4.82	4.41	1.56	1.43	2.74	2.66	1.39	1.28
France_SS_13	5.04	4.42	1.31	1.18	3.44	2.74	1.31	1.18
France_SS_16	5.91	5.30	1.22	1.11	3.98	2.91	1.22	1.11
Montreal_r06_SS_01	12.29	11.45	1.24	1.17	6.17	5.71	1.24	1.17
Montreal_r06_SS_03	11.16	9.35	1.30	1.10	5.92	4.89	1.30	1.10
Montreal_r06_SS_04	12.62	11.17	1.33	1.15	6.61	5.69	1.33	1.15
Montreal_r06_SS_08	13.04	10.84	1.32	1.14	6.85	5.65	1.32	1.14
Pdh_01_SS	1.97	1.81	1.10	1.03	1.49	1.33	1.10	1.03
Pdh_02_SS	2.48	2.36	1.07	1.03	1.67	1.56	1.07	1.03
Pdh_04_SS	2.36	2.25	1.04	1.01	1.62	1.51	1.04	1.01
Pdh_08_SS	2.15	2.00	1.13	1.07	1.53	1.40	1.13	1.07
Nobel_EU_SS_04	3.88	3.56	1.42	1.32	2.81	2.54	1.39	1.29
Nobel_EU_SS_05	4.86	4.24	1.34	1.19	3.28	2.79	1.34	1.19
Nobel_EU_SS_15	4.02	3.53	1.33	1.18	2.87	2.45	1.37	1.21
Nobel_EU_SS_18	4.83	4.27	1.31	1.18	3.14	2.74	1.31	1.18

Table 4: The ratios for the multi-source cases corresponding to Figures 1 and 2, split by correlation structure.

	Full capacity				**Mean value capacity**			
	comp. A		**comp. B**		**comp. A**		**comp. B**	
Test Name	$\rho = \mathbf{0}$	$\rho \geq \mathbf{0}$	$\rho = \mathbf{0}$	$\rho \geq \mathbf{0}$	$\rho = \mathbf{0}$	$\rho \geq \mathbf{0}$	$\rho = \mathbf{0}$	$\rho \geq \mathbf{0}$
Germany_1_MS	2.05	1.89	1.14	1.07	1.54	1.40	1.14	1.07
Germany_2_MS	2.23	2.00	1.22	1.12	1.63	1.43	1.22	1.12
Germany_3_MS	2.10	1.95	1.17	1.09	1.52	1.38	1.17	1.09
Germany_4_MS	2.15	1.96	1.12	1.04	1.57	1.41	1.12	1.04
France_MS_1	6.57	5.98	1.17	1.08	3.56	3.20	1.17	1.08
France_MS_2	5.83	5.35	1.38	1.26	3.59	3.30	1.38	1.26
France_MS_3	6.33	5.72	1.47	1.35	3.48	3.10	1.22	1.13
Montreal_r06_MS_1	15.33	14.78	1.09	1.04	6.90	6.76	1.09	1.04
Montreal_r06_MS_2	14.29	13.23	1.20	1.10	6.88	6.36	1.20	1.10
Montreal_r06_MS_3	15.31	14.04	1.14	1.02	7.32	6.69	1.14	1.02
Pdh_1_MS	2.50	2.42	1.02	1.00	1.60	1.52	1.02	1.00
Pdh_2_MS	2.43	2.30	1.03	1.00	1.58	1.48	1.03	1.00
Pdh_3_MS	2.65	2.58	1.02	1.00	1.65	1.59	1.02	1.00
Pdh_4_MS	2.65	2.49	1.08	1.03	1.74	1.60	1.08	1.03
Nobel_EU_MS_plus_01	4.50	4.21	1.28	1.18	2.88	2.70	1.17	1.10
Nobel_EU_MS_plus_02	4.61	4.32	1.20	1.13	2.72	2.54	1.20	1.13
Nobel_EU_MS_plus_03	5.05	4.67	1.26	1.17	2.94	2.70	1.26	1.17
Nobel_EU_MS_plus_04	4.89	4.52	1.23	1.13	3.08	2.77	1.23	1.13

Table 5: The ratios for Comparison B corresponding to Figure 3, for full capacities and mean capacities at different capacity installation costs.

	Single source		**Multi-source**	
Var. setup cost	**Full Cap.**	**Mean Cap.**	**Full Cap.**	**Mean Cap.**
0.33	1.18	1.18	1.2	1.17
0.66	1.19	1.19	1.15	1.16
1	1.19	1.18	1.12	1.1
1.33	1.16	1.16	1.11	1.1
1.66	1.16	1.15	1.1	1.09
2	1.14	1.14	1.09	1.09