Landelijk Netwerk
Mathematische Besliskunde

Stochastic Programming
LNMB PhD course 2021-2022

Project Report

Delft, August 4, 2022

Submitted by:

Nanda Kishor Panda	Neda Vahabzad
n.k.panda@tudelft.nl	n.vahabzad@tudelft.nl

Code used in this report can be found in GitHub

Q(a) Two-Stage recourse model with single discrete random parameter

To analyze the recourse model described in the case study, we will go through the following procedure:
Compute (i) the Expected Value Solution and (ii) the Expected Result (of the expected value solution). Based on the results, discuss the usefulness of solving the recourse model. Remark: To compute (i), the maximum demand in the first stage constraint can be adjusted.
Solve the recourse model and compare the outcome with (i) and (ii) above. In particular, discuss the so-called Value of the Stochastic Solution. Compute the Wait-and-See solution and compare the outcome with previous results. In particular, discuss the so-called Expected Value of Perfect Information

Deterministic model (MOD1)

In order to analyze the two-stage recourse model, we first need to formulate the deterministic mode as follows:

$$\min_{x\geq 0}\left\{\underbrace{\sum_{i=1}^{n} c_i x_i}_{\text{investment cost}} + \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{k} q_i T_j y_{ij}}_{\text{operating cost}}\right\} \tag{1}$$

subject to:

$$\sum_{i=1}^{n} x_i \geq \sum_{j=1}^{k} \xi_j^{max} \tag{2}$$

$$\sum_{i=1}^{n} c_i x_i \leq c_{max} \tag{3}$$

$$\sum_{j=1}^{k} y_{ij} \leq x_i \qquad i = \{1,\cdots,n\} \tag{4}$$

$$\sum_{i=1}^{n} y_{ij} \geq \xi_j \tag{5}$$

Expected Value Problem Formulation

For solving the problem with the expected value method, the expectation of each random parameter is used as input to the deterministic model. Therefore, to obtain the expected value solution, we solve MOD1 by changing the values of the random parameters in (5) by their expected value, as shown below:

$$\xi_j = \bar{\xi_j} \qquad \forall\, j = 1,2,3 \tag{6}$$

where;

$$\bar{\xi_j} = \mathbb{E}[\xi_j] = \sum_{l=1}^{3} \xi_j^{(l)} \mathcal{P}(\xi_j^{(l)}) \qquad \forall\, j = 1,2,3$$

Expected Value (EV):

$$EV := \min_{x \geq 0} \left(g(E[\xi], x) \right) = g(\bar{\xi}, \bar{x}) \tag{7}$$

$$\tag{8}$$

where, $\bar{x}$ is the optimal solution of MOD1 by using $\xi = \bar{\xi}$.

Results using expected value

The results needed for the analysis of the recourse model is obtained through computer simulations in Python using the Pyomo package. The stochastic solutions are obtained using the MPI-SPPY package that can use the models created using Pyomo.

(i) Expected Value Solution

Table 1 provides the expected solution ($\bar{x}_i$: Capacity of installed technology) along with the second stage variables, i.e. the capacity of each technology i effectively used in mode j. The deterministic maximum capacity (ξ^{max}) used in the constraint (5) is taken as [7, 4, 3] for each modes ξ_1, ξ_2, ξ_3 respectively. This is taken by selecting the maximum value from the discrete distribution of each individual random parameter (ξ_j). For this reason, an additional 4 units of technology-4 is installed to satisfy the constraint of maximum demand but are never used in the expected solution. The maximum value is chosen so there is no in-feasibility while solving for stochastic solution (TS) and expected result of expected value solution (EEV) due to certain scenarios where the total sum of loads in each mode is greater than their mean values.

Table 1: Expected value solution for Q(a)

Technologies	Installed capacity	Used capacity (j=1)	Used capacity (j=2)	Used capacity (j=3)
Technology-1	2.83	2.83	0	0
Technology-2	3	0	3	0
Technology-3	2.16	2.16	0	0
Technology-4	6	0	0	2

The result achieved for the expected value (EV), which is the total objective cost obtained by the expected value solution, and the results of the investment and operation costs are given as follows:

- Investment = 120
- Operation cost = 274.66
- Total objective cost (EV) = **394.67**

(ii) Expected Result of the Expected Value Solution

The expected value problem is a linear deterministic problem, which is much simpler to solve due to replacing the uncertain demands with their respective expected values. To judge how good is the result of this simple deterministic model, we evaluate the so-called expected result of the expected value solution (EEV). EEV measures the performance of the expected solution ($\bar{x}_i$) when used as a parameter to solve only the second-stage problem, as shown in (9).

$$EEV := \mathbb{E}_\xi \left[g(\bar{x}, \xi) \right] \tag{9}$$

EEV can be calculated for m discrete random variables at d values each by solving d^n linear programs. The EEV is calculated using the sample mean approximation as follows:

- For $i = 1, ..., L$ where $L = d^m$ and for this particular problem $d = m = 3$:
- Sample ξ_i from the discrete realizations
- Compute the optimal value of x_i for $\bar{\xi}$: $\bar{x}_i$
- $EEV_l = \sum_{i=1}^{n} c_i \bar{x}_i + Q(\bar{x}, \xi^{(l)})$ where, l is the l^{th} scenario. Here $l = 1, 2, ..., 27$
- Finally, $EEV = \frac{1}{L} \sum_{l=1}^{L} EEV_l$
- $Q(\bar{x}, \xi^{(l)})$ can be written as a linear program as follows:

$$\min_{y \geq 0} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{k} q_i T_j y_{ij} \right\} \tag{10}$$

$$\text{Subject to:} \tag{11}$$

$$\sum_{j=1}^{k} y_{ij} \leq \bar{x}_i \qquad i = \{1, \cdots, n\} \tag{12}$$

$$\sum_{i=1}^{n} y_{ij} \geq \xi_j^{(l)} \qquad j = \{1, \cdots, k\} \tag{13}$$

By solving the algorithm mentioned above, the value of EEV is found as **400.04**. The obtained amount for the EEV is used for calculating the Value of the stochastic solution (VSS) after solving the stochastic recourse model later.

The obtained solutions for the installed capacity of each technology using the expected values of the random parameters are feasible but not necessarily optimal. Because in real cases, demand can be higher or lower than the expected amount. Therefore, despite the simplicity, the result of the expected value problem may not be realistic and optimal. On the other hand, since the recourse model considers the uncertain nature of the demand in each mode, the result of this model can be optimal and hence more reliable. In real-life scenarios, decisions should be taken with due diligence about the worst case but not at the cost of overestimating the solution.

Stochastic recourse model

The 2-stage stochastic model can be formulated by changing the objective function of the deterministic model (MOD1) and adding the random variable ξ_j as follows:

$$\min_{x\geq 0} \underbrace{\left\{ \sum_{i=1}^{n} c_i x_i + \mathbb{E}_\xi \underbrace{\left[\min_{y\geq 0} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{k} q_i T_j y_{ij} \right\} \right]}_{\text{second stage}} \right\}}_{\text{first stage}} \tag{14}$$

subject to:

first stage constraints:

$$\sum_{i=1}^{n} x_i \geq \sum_{j=1}^{k} \xi_j^{max} \tag{15}$$

$$\sum_{i=1}^{n} c_i x_i \leq c_{max} \tag{16}$$

second stage constraints:

$$\sum_{j=1}^{k} y_{ij} \leq x_i \qquad i = \{1, \cdots, n\} \tag{17}$$

$$\sum_{i=1}^{n} y_{ij} \geq \xi_j \qquad j = \{1, \cdots, k\} \tag{18}$$

The given distributions of the independent random variables ξ_j are summarized in Table 2. A total of 27 scenarios are generated based on the provided data and their corresponding probabilities. These 27 realizations can be visualised in fig 1

Table 2: Distributions of the independent random variables ξ_j

Prob	ξ_1	ξ_2	ξ_3
0.3	3	2	1
0.4	5	3	2
0.3	7	4	3

To solve the above-mentioned two-stage stochastic programming problem with the obtained scenarios, different methods can be used, including Extensive Form, Progressive Hedging (PH), and Benders' Decomposition. We used Benders' Decomposition, also known as the L-shaped method, due to their ease in handling many scenarios.

The strategy behind Benders decomposition can be summarized as divide-and-conquer. In Benders decomposition, the variables of the original problem are divided into two subsets so that a first-stage master problem is solved over the first set of variables, and the values for

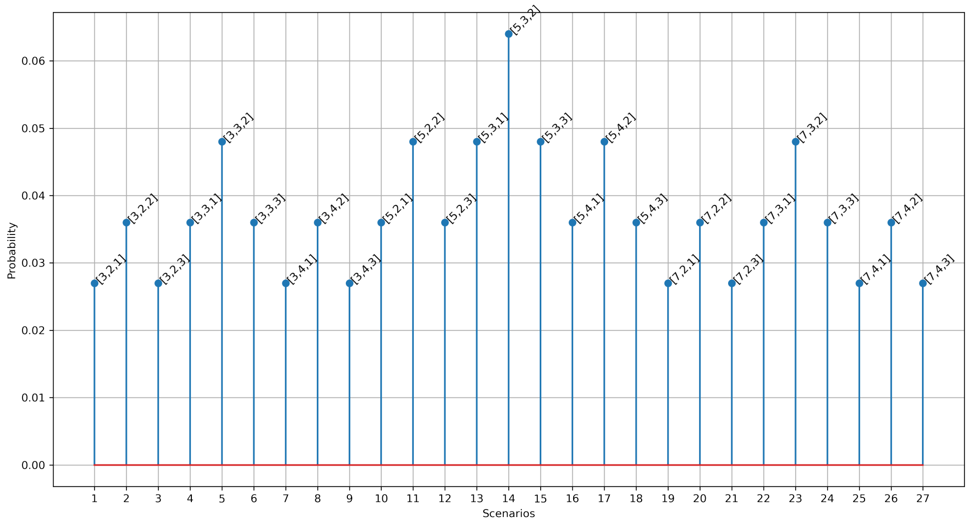


Figure 1: Probability distribution of ξ_i for each 27 scenarios

the second set of variables are determined in a second-stage sub-problem for a given first-stage solution. If the sub-problem determines that the fixed first-stage decisions are in fact infeasible, then so-called Benders cuts are generated and added to the master problem, which is then re-solved until no cuts can be generated. Since Benders' decomposition is a suitable technique for solving large linear stochastic programming problems in which the uncertainty is represented with several scenarios, we chose this interesting method to solve our problem. Fig 2 gives a better representation of Benders decomposition methodology.

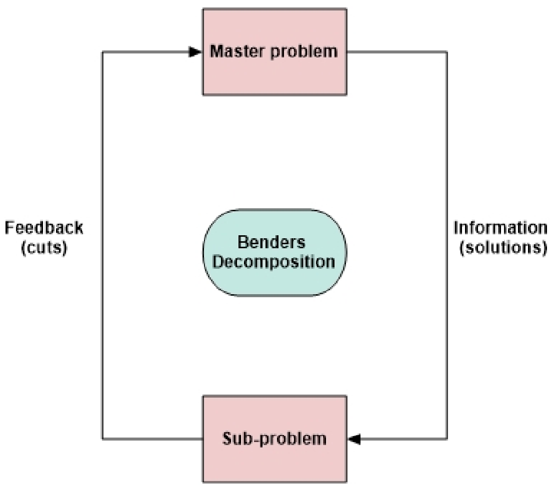


Figure 2: Benders decomposition

Results of stochastic solution

(i) Recourse model solution

The solution obtained by solving the stochastic problem with the generated scenarios as the input data for demand and with the help of Benders decomposition is depicted in Table 3.

Table 3: Recourse model solution

Technologies	Installed capacity
Technology-1	3.17
Technology-2	5
Technology-3	1.83
Technology-4	3.99

(ii) Recourse model result

The total objective cost obtained by the two-stage stochastic problem (TS) is equal to **397.75**.

By comparing $TS = 397.75$ and $EEV = 400.04$, we notice that $TS \leq EEV$. It is because the solution obtained by the simple deterministic model is feasible but not necessarily optimal for this stochastic problem.

(iii) Value of Stochastic Solution (VSS)

To evaluate the usefulness of the recourse model over the deterministic model, we use the so-called value of the stochastic solution (VSS), which is defined as the difference between EEV (the expected result of of the expected value solution) and TS (the optimal value of two-stage recourse model solution). It indicates the average price one pays for using the simplistic deterministic model rather than the stochastic model. The VSS for this case is as follows:

$$VSS = EEV - TS = 2.29$$

Wait-and-see solution

In the wait-and-see problem, we assume that it is possible to wait for the actual value of the uncertain demand before deciding upon the installed capacity of each technology (x_i) and the used capacity per technology per mode (y_{ij}). The wait-and-see problem is formulated by solving MOD1 for all possible realizations of ξ_j. As ξ_j is discrete and mutually independent, there are a total of 27 realizations for the value of the objective function. The mean value of the objective functions achieved by the wait-and-see solutions in each scenario, is denoted by WS and can be formulated as:

$$WS := \mathbb{E}_{\xi}\left[min_{x\geq 0}\, g(\xi, x)\right] = \mathbb{E}_{\xi}\left[g(\xi, \hat{x}(\xi))\right] \tag{19}$$

Results of wait-and-see

As the wait-and-see solution solves the deterministic model for every permutation and combination of the random parameters, it is convenient to look at the mean value of the solution for every scenario. The mean values of installed capacity for different technologies can be found in Table 4. Now, the wait-and-see result (WS) is the mean of the objective function for each scenario. The values of WS is provided in Table 5.

Table 4: Wait-and-see Solution

Technologies	Capacity
Technology-1	3.18
Technology-2	2.85
Technology-3	2
Technology-4	5.96

Table 5: Wait-and-see Result

Investment Cost	119.593
Operation cost	275.407
Total objective cost (WS)	395

Expected value of perfect information (EVPI)

If we compare $WS = 395$ and $TS = 397.75$, we notice that $WS \leq TS$. It is reasonable because the optimal solution to the original recourse model is feasible for each possible scenario, but not necessarily optimal. The difference between TS and WS is called the Expected Value of Perfect Information (EVPI), which is calculated as below:

$$EVPI = TS - WS = 2.75$$

Indeed, EVPI gives the average price one pays for knowing only the distribution of the random parameters rather than their actual values. It can be concluded that the wait-and-see model provides a lower bound for the optimal value of the two-stage stochastic recourse model, whereas the expected result gives an upper bound. By comparing all the results obtained for EV, EEV, TS and WS, we can relate between EV, WS, TS and EEV as shown in Fig 3.

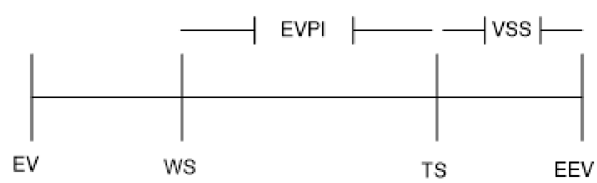


Figure 3: The relation between the performance indicators for analysis of recourse model.

Q(b) Addition of operational uncertainty: $\alpha_{1,2,3,4}$

Assume now, in addition, that the operational availability of each of the technologies is random: if a capacity of x_i is installed, the actual amount available is $\alpha_i x_i$, with α_i a random parameter, i = 1, .., n.

$$\begin{aligned} \alpha_1 &\sim \mathcal{U}(0.6{,}0.9) \\ \alpha_2 &\sim \mathcal{U}(0.7{,}0.8) \\ \alpha_3 &\sim \mathcal{U}(0.5{,}0.8) \\ \alpha_4 &\sim \mathcal{U}(0.9{,}1.0) \end{aligned}$$

Moreover, assume that it is possible to 'import' electricity to cover any observed power shortage. This virtual fifth technology has zero investment cost, but high production cost 10. Formulate the corresponding recourse model. (Hint: the possibility to import electricity has implications for both stages of the model). Analyze this model, and compare the results with (a). In particular, discuss the expected result (in model (b)) of the solution obtained under (a).

Formulation of the modified recourse model with operational uncertainty (MOD2)

We re-formulate the corresponding recourse model as follows (c.f. (20)-(23)), considering both the additional random parameter α_i (which we dealt with in the next sub-section), the effect of the newly added virtual technology and the original uncertain parameter of demand.

$$\min_{x\geq 0} \underbrace{\left\{ \sum_{i=1}^{n} c_i x_i + \underbrace{\mathbb{E}_\xi \left[\min_{y\geq 0} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{k} q_i T_j y_{ij} \right\} \right]}_{\text{second stage}} \right\}}_{\text{first stage}} \tag{20}$$

subject to:

first stage constraints:

$$\sum_{i=1}^{n} c_i x_i \leq c_{max} \tag{21}$$

second stage constraints:

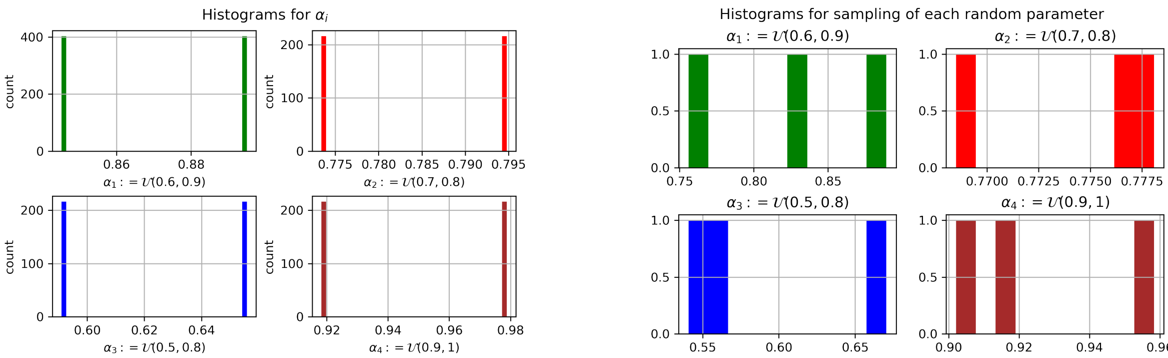
$$\sum_{j=1}^{k} y_{ij} \leq \alpha_i x_i \qquad i = \{1, \cdots, n\} \tag{22}$$

$$\sum_{i=1}^{n} y_{ij} \geq \xi_j \qquad j = \{1, \cdots, k\} \tag{23}$$

Here, the n goes from 1 till 5, to include the fifth virtual technology. By adding this fifth technology with 0 investment cost, the constraint $\sum_{i=1}^{n} x_i \geq \sum_{j=1}^{k} \xi_j^{max}$ becomes redundant as for any values of y_{ij}, this constraint can always be satisfied by any value of x_5 as it does not affect the total investment cost. Since the technology-5 is always available, we consider $\alpha_5 = 1$.

Sampling and scenario generations from continuous and discrete distributions

In this part another random parameter, the operational availability of each of the technologies, is also added to the problem, which has different continuous distribution for each technology. Since, discrete distributions are easy to deal with in algorithms than continuous distributions, we tried to discretize each continuous distribution (sampling). However, the discretization of continuous distributions always don't provide the optimal solution. This can be dealt with by increasing the number of samples which will increase the computational time exponentially. To evaluate this impact of increasing the number of samples on the recourse model, we solved the problem for 2 and 3 samples for each of the continuous distributions. These resulted in a total of 432 and 2187 scenarios respectively. For reproducibility of the samples a constant seed is selected when generating the samples from the continuous distribution. The samples selected for $m = 2, 3$; where m is the number of samples per continuous distribution is plotted in Fig 4.



(a) m=2 discrete samples

(b) m=3 discrete samples

Figure 4: Histograms for the discretely sampled uniform distribution for 2 and 3 discrete samples each.

After completing the sampling process, we generated scenarios based on different combinations of these independent discrete samples for each α_i and the previous 27 realizations of ξ_j. The process of scenario generation can be better visualized in Fig 5.

To have a better comparison between the results obtained in part (a), (b) and (c) from solving the expected value problem, recourse model and wait and see problem, as well as the relevant indicators, we provided the achieved values for EV, EEV, TS, VSS, WS and EVPI all together in a table at the end of the document after bringing the modeling and problem formulation of part (c).

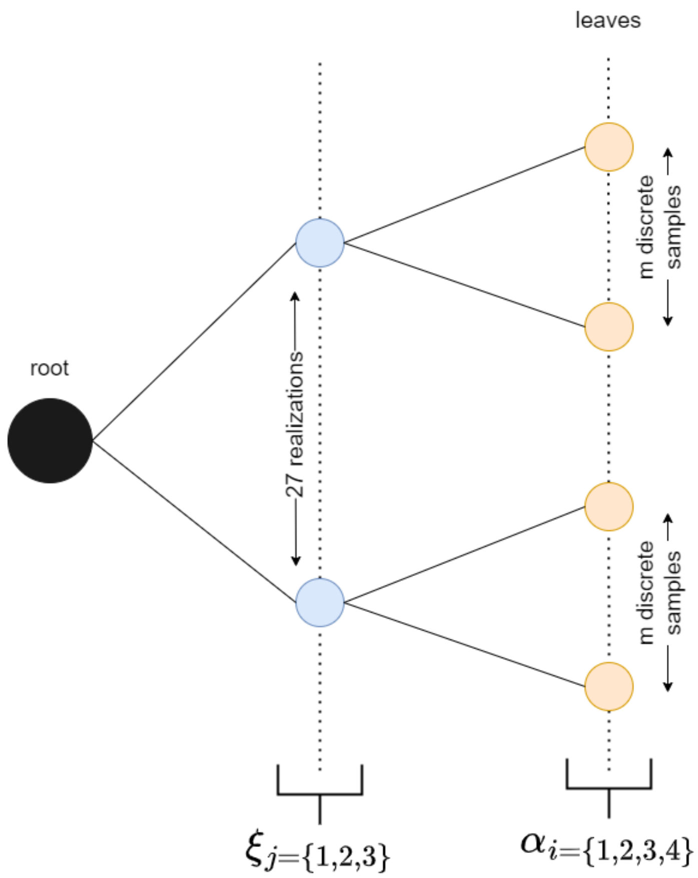


Figure 5: Scenario tree for seven discrete parameters $\xi_{i=\{1,2,3\}}$ and $\alpha_{i=\{1,2,3,4\}}$

.

Q(c) Addition of random parameter in mode duration for $T_{i=\{2,3\}}$

Assume, in addition, that $T_i = \tau_i$, i = 2, 3, where the distribution of the independent random variables τ_i is given as follows:

Probability	τ_2	τ_3
0.6	5	0.5
0.4	7.5	1.75

Formulation of the modified recourse model (MOD3)

The stochastic two-stage recourse model can be reformulated by adding the new random parameters to the previous recourse model (MOD2) as follows:

$$\min_{x\geq 0} \underbrace{\left\{ \sum_{i=1}^{n} c_i x_i + \mathbb{E}_\xi \underbrace{\left[\min_{y\geq 0} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{k} q_i \tau_j y_{ij} \right\} \right]}_{\text{second stage}} \right\}}_{\text{first stage}} \tag{24}$$

$$\text{first stage constraints:} \tag{25}$$

$$\sum_{i=1}^{n} c_i x_i \leq c_{max} \tag{26}$$

$$\text{second stage constraints:} \tag{27}$$

$$\sum_{j=1}^{k} y_{ij} \leq \alpha_i x_i \qquad i = \{1, \cdots, n\} \tag{28}$$

$$\sum_{i=1}^{n} y_{ij} \geq \xi_j \qquad j = \{1, \cdots, k\} \tag{29}$$

Sampling and scenario generations from continuous and discrete distributions

In this part, the duration of mode 2 and mode 3 are also added to the problem as random variables with a discrete distribution, having a total of 4 combined realizations. The process of scenario generation is similar to the previous models.

Results and conclusion

The detailed results after solving each of the above models is presented in Table 6. The important inference from these results can be summarised below:

- Expected value solutions can be a simple and computationally efficient approach to handle uncertainties, but at the cost which can be optimistic in many cases. In reality the actual costs after including uncertainty can be higher. We can see from Table 6 that the value of TS is higher than EV for each of the models, which tells us that the trade off for using a simple approach is an inaccurate solution (non-realistic solution).
- For the first model due to the unavailability of any virtual technology, the EV solution always invests on more units of the least costly technology to satisfy the constraint of maximum total deterministic demand. For example, it can be seen in Table :1 an extra of 4 units of Technology 3 is installed just to satisfy the max demand constraint. This is no longer needed in the rest of the models (MOD2, MOD3) due to the availability of a virtual technology with 0 investment cost.
- For MOD1, with only discrete random parameter, the TS was calculated quite efficiently due to small and fixed number of scenarios. In this case, the price which one pays to

have access to perfect information (EVPI) is low due to limited number of scenarios. For the same reason the average price for using a simple deterministic model (VSS) is low.

- In MOD2, we have an additional uncertain parameter α_i which makes the operational availability of each technology uncertain. Due to the continuous nature of this random parameter, achieving the optimal TS value is still challenging using discretization. For example, if we just consider the upper and lower bound of the continuous distribution the value of TS will be either overly optimistic or pessimistic. Hence there has to be a trade off between a number of discretized samples and computation time. As seen from Table 6, the increase in computation time is not linear with the increase in a number of discrete samples per continuous distribution.

- In general the TS value for MOD2 will be higher than the TS value of MOD1 as now the installed technologies are not always available 100%. For example, as seen in Fig :6, a total of 9.19 of Technology 1 is installed, but only 8 units are used in total among the different modes. This is to compensate for the effects of α_1.

```
+--------------+-------------------------+------------+------------+------------+
| Technologies | Installed Capacity (x)  | Usage (T1) | Usage (T2) | Usage (T3) |
+--------------+-------------------------+------------+------------+------------+
| Technology-1 |   9.194074809305528     |    5.0     |    3.0     |    0.0     |
| Technology-2 |          0.0            |    0.0     |    0.0     |    0.0     |
| Technology-3 |          0.0            |    0.0     |    0.0     |    0.0     |
| Technology-4 |          0.0            |    0.0     |    0.0     |    0.0     |
| Technology-5 |          2.0            |    0.0     |    0.0     |    2.0     |
+--------------+-------------------------+------------+------------+------------+
Total investment cost     = 91.94074809305528
Total operation cost = 292.0
Total objective cost (EV) = 383.9407480930553
```

Figure 6: Results of expected value solution for MOD2

- The value of EVPI for MOD2 and MOD3 are much higher than MOD1 due to the increase in a number of scenarios. Also, we can notice that with the increase in a number of scenarios for the same model, the EVPI increases due to the additional effort one has to pay to gain access to the perfect information of newer scenarios.

- It is interesting to note that for MOD2 and MOD3 the value of EEV are less than the TS, which results in a negative VSS. This is because EEV can not be solved for all the scenarios due to the lack of constraint on the maximum demand. There are many scenarios in which the expected solution cannot satisfy the demand with the inclusion of α_i. For example, to solve MOD3 for m=3; only 4104 scenarios were feasible out of 8748 total scenarios.

Table 6: Comparison of performance indices for all three models

Model	No. of Discrete Samples	No. of Scenarios	EV	WS	TS	EEV	EVPI	VSS	Computation time for TS [hours]
MOD1		*27*	394.67	395	397.75	400.04	2.75	2.29	0.0048
MOD2	m=*2*	*27*16*	383.94	384.076	398.13	388.40*	14.054	-	0.12
	m=*3*	*27*81*	385.73	385.95	404.42	388.36*	18.47	-	0.546
MOD3	m=*2*	*27*16*4*	383.94	388.062	398.13	328.44*	10.068	-	0.51
	m=*3*	*27*81*4*	388.715	390.105	400.42	325.36*	10.315	-	2.03

* The values of EEV are only for a part of the feasible scenarios and hence don't denote the true values