


Article

Multistage Sample Average Approximation for Harvest Scheduling under Climate Uncertainty

Martin B. Bagaram^{1,2*} , Sándor F. Tóth¹¹ School of Environmental and Forest Sciences, University of Washington, Box 352100, Seattle, WA 98195, USA² Industrial and Systems Engineering, University of Washington, Seattle, WA 98195-2650, USA

* Correspondence: martb@uw.edu

Abstract: Forest planners have traditionally used expected growth and yield coefficients to predict future merchantable timber volumes. However, because climate change affects forest growth, the typical forest planning methods using expected value of forest growth can lead to sub-optimal harvest decisions. We proposed in this paper to formulate the harvest planning with growth uncertainty due to climate change problem as a multistage stochastic optimization problem and use sample average approximation (SAA) as a tool for finding the best set of forest units that should be harvested in the first period even though we have a limited knowledge of what future climate will be. The objective of the harvest planning model is to maximize the expected value of the net present value (NPV) considering the uncertainty in forest growth and thus in revenues from timber harvest. The proposed model was tested on a small forest with 89 stands and the numerical results showed that the approach allows to have superior solutions in terms of net present value and robustness in face of different climate scenarios compared to the approach using the expected growth and yield. The SAA methods requires to generate samples from the distribution of the random parameter. Our results suggested that a sampling scheme that focuses on generating high number of samples in distant future stages is favorable compared to having large sample sizes for the near future stages. Finally, we demonstrated that, depending on the level of forest growth change, ignoring this uncertainty can negatively affect forest resources sustainability.

Keywords: data-driven optimization; climate change; harvest planning; optimality gap; forest

1. Introduction

Climate change is arguably one of the most challenging issues that contemporary forest planners need to address. In general, forest practitioners need to provide a long term forest harvest scheduling plan that takes into account the preference of the stakeholders while complying with environmental, regional and ecological restrictions. Usually, forest planning aims at scheduling which forest units should receive a specific treatment such as harvesting, thinning, etc., in each period in order to achieve the management objectives. This planning also known as harvest scheduling, expands several decades, and is subject to many sources of uncertainties due to factors that include natural hazards (fire, windthrow, insects, etc), and technological limitation such as forest inventory errors, growth prediction errors, poor foresight of the price of forest products on the market [1,2]. It is important to incorporate all or some of these uncertainties in forest planning.

The need to incorporate uncertainty in forest harvest scheduling has been acknowledged several decades ago [3]. Subsequently, [4] urged for consideration of forest growth uncertainty in the planning. The authors asserted that the cost of ignoring such uncertainty at a forest scale can be detrimental. When the uncertainty is ignored, we make decisions assuming the growth in the future will be the expected value of growth (average growth). However, it is rare that the actual growth will be the average one and therefore, many harvest decisions made in early periods of the planning horizon were not optimal. This leads to a failure to meet the planning objectives and satisfying many of the aforementioned restrictions. Owing to the computational challenges that the introduction of

uncertainty in harvest modeling involves, [5] recommended that each forest harvest scheduling plan should be accompanied by an estimation of its uncertainty or its reliability. Notwithstanding all the exhortation to incorporate uncertainty in forest planning, harvest scheduling models, for a long time, have ignored uncertainty or were limited at reporting the sensitivity of the harvesting plans in case of change of one or many input parameters of the modeling. It is only recently that there has been a prolific number of papers addressing the integration of uncertainty in harvest planning. The most common uncertainty that has been addressed is the wood price and demand uncertainties [6–10]. Recently, there is an interest to incorporate climate uncertainty in forest planning. The first study that we are aware of that explicitly addresses the issue was conducted by [11]. The authors assessed how climate change affected the management decisions of a Eucalyptus forest in Portugal with a planning horizon of 15 years. A similar study was conducted by [12,13] using the same dataset.

In this paper, we model harvest planning as a stochastic optimization problem and use sample average approximation (SAA) as a tool for identifying the best set of actions one can take to meet the management objectives. In other words, we use SAA for solving harvest planning models. Stochastic optimization is a modeling framework for optimization problems dealing with uncertainty. In multistage stochastic programming, decisions are made sequentially at stages and the uncertainty unfolds during periods. Thus, the decision maker needs to implement a decision at the beginning of the planning horizon (first stage decisions) without knowing what the value of the uncertain parameter will be. After a period, in which the uncertainty is revealed, the decision maker can take recourse actions at subsequent stages. Therefore, it is important to make optimal decisions in the early stages of the planning without knowing the magnitude of the uncertainty that will unfold. Specifically, in strategic harvest scheduling, the decision maker needs to prioritize the set of forest units (stands) that should be harvested here and now or during the ongoing decade or period. After that period, the decision maker will have the opportunity to revisit the management in the following periods. In this case, the goal of the harvest scheduling is to decide the set of actions that managers should apply immediately “here and now” (first stage decisions).

SAA is a Monte Carlo simulation based approach for solving stochastic optimization problems. It consists of drawing repetitively samples from the distribution of the random parameter and solving the resultant average sample deterministic optimization problem. Each sample might lead to a different solution. However, if the sample size is large enough, the sample objective function value will approximate the true optimal value. [14] showed that for minimization problems, the expectation of samples’ objective value corresponds to the lower bound on the true optimal value of the stochastic optimization problem and that the bound monotonically increases as the sample size increases. SAA has been successfully applied in several fields such as portfolio selection [15], supply chain network design [16,17], facility location problem [18] and personnel assignment [19]. Notwithstanding all these applications, and the fact that SAA is well suited for problems where the objective function is difficult to evaluate such as harvest planning under climate uncertainties [15], to the extent of our knowledge, SAA has never been applied in forestry domain. In addition, unlike stochastic programming which requires to know the probability distribution of the samples, SAA known as a data-driven optimization method, considers the samples to be equiprobable. The idea is that if the sample size is large enough, then the sample statistics will represent those of the actual population the sample is taken from. Hence, there is no need to know the actual distribution of the uncertainty. This is particularly important for harvest scheduling under climate uncertainty because climate change is forecast as possible futures without any probabilities. Moreover, although the method performs well for two-stage stochastic problems [20], it is unclear how the method will perform on a multistage harvest scheduling problem.

The objective of this paper is to fill the gap in the literature on SAA and multistage stochastic harvest scheduling models. For multistage stochastic optimization problems, each sample of the uncertainty is known as a scenario. It is known that the number of scenarios to consider grows exponentially with the number of stages in multistage stochastic programs. Forest harvesting

scheduling models are typically characterized with long planning horizons with many stages. The contributions in this paper are as follows:

1. Introduce a method to handle climate change uncertainty in forest harvest scheduling that allows to make sound decisions in early stages of planning. Poor decisions in early stages are significantly more detrimental to many businesses compared to decisions in later stages. Hence, later stage decisions can be considered as secondary.
2. Propose a method to generate the set of scenarios to reduce the sample size and keep the optimization model tractable. If we generate possible scenarios of forest growth in i.i.d (independent identical distributed) fashion, then we might have a very large sample size before SAA solution converges to optimality. This large sample size can make the problem computationally intractable. We propose a sampling scheme that focuses at having higher number of replications for distant stages.

We will show that stochastic harvest scheduling has an advantage over the expected value approach (deterministic model presented in Appendix A) because it allows to implement policies that are optimal for all foreseen forest growth change by providing management decisions that would be implemented if we consider the uncertainty and if we do not. The rest of this document is structured as follows. In Section 2, we present the materials and methods used in this research. We describe in that section the SAA method and the scenario generation scheme. We dedicate Section 3 to presenting our findings. Finally, in Section 4, we discuss the results and outline future research.

2. Material and methods

2.1. Problem description and formulation

We present in this section the multistage stochastic harvest scheduling problem with forest growth uncertainty due to climate change. Let $t \in \mathcal{T} = \{1, \dots, T\}$ be the set of stages in which forest units are eligible for harvest in the future. We reserve $t = 0$ to the time "now" (first stage decision), where there is no uncertainty on the forest growth. We define $\zeta := \{\zeta_1, \dots, \zeta_T\}$ to be the random vector characterizing forest growth change due climate change. Hence, ζ_t is the random parameter of ζ at time t . Each ζ_t has a support Ξ_t which is the range of values that the random parameter ζ_t can take. The meaning of parameters, variables and sets is given in Table 1.

The objective of the decision maker is to maximize the expected net present value (NPV) from timber harvest. This can be formulated in a generic form as (1) and (2):

$$\max \sum_{s \in \mathcal{S}} r_s x_s + \mathbb{E} \left[R(x, \zeta) \right] \quad (1)$$

$$\text{subject to } x \in \{0, 1\}^{|\mathcal{S}|}, \quad (2)$$

where we take the expectation with respect to ζ and, $R(x, \zeta^i)$ is the optimal value of the harvest scheduling given the harvest in the first stage (x) and the occurrence of one climate scenario ζ^i . In this formulation, the decision variables are only the first stage variables which means that the decision maker mainly cares about how to select the stands that can be harvested in the first period so that the expected NPV is maximized. The value of $R(x, \zeta^i)$ is obtained by solving the following optimization problem:

$$\max \sum_{s \in \mathcal{S}} \left[\sum_{t \in \mathcal{T}} r_{st}^{\zeta^i} y_{st}^{\zeta^i} + r_{s0}^{\zeta^i} w_s^{\zeta^i} \right] \quad (3)$$

subject to

$$x_s + w_s^{\zeta^i} + \sum_{t=1}^T y_{st}^{\zeta^i} = 1 \quad s = 1, \dots, |\mathcal{S}| \quad (4)$$

$$\sum_s v_s x_s = H_0 \quad (5)$$

$$\sum_s v_{st}^{\zeta^i} y_{st}^{\zeta^i} = H_t^i \quad t = 1, \dots, T \quad (6)$$

$$H_t^i \leq \alpha H_{t-1}^i \quad t = 1, \dots, T \quad (7)$$

$$H_t^i \geq \beta H_{t-1}^i \quad t = 1, \dots, T \quad (8)$$

$$H_t^i \leq \gamma H_{t-2}^i \quad t = 2, \dots, T \quad (9)$$

$$H_t^i \geq \lambda H_{t-2}^i \quad t = 2, \dots, T \quad (10)$$

$$\sum_{s \in \mathcal{S}} a_s \left[\sum_{t=1}^T \text{age}_{st} (x_s + y_{st}^{\zeta^i}) + \text{age}_{s0} w_s^{\zeta^i} \right] \geq \sum_s a_s \text{age}_s. \quad (11)$$

$$w_s^{\zeta^i} \in \{0, 1\}^{|\mathcal{S}|}, y_{st}^{\zeta^i} \in \{0, 1\}^{|\mathcal{S}| \times T}, H_t^i \geq 0 \quad \forall t \in \{0, 1, \dots, T\}. \quad (12)$$

Table 1. Variable, set and parameters used in problem formulation

Sets

| | |
|---------------|--|
| \mathcal{S} | set of stands or forest management units ($s \in \mathcal{S}$) |
| \mathcal{T} | set of time period in which the decision are implemented or stage at which the decisions are taken ($t \in \mathcal{T}$) |
| Ω | set of scenarios ($\zeta^i \in \Omega$) |

Variables

| | |
|--------------------|--|
| x_s | binary variable: 1 if management unit s is scheduled to be harvested in the first period (now); and 0 otherwise |
| $w_s^{\zeta^i}$ | binary variable: 1 if management unit s should not be harvested during the whole planning horizon given the scenario ζ^i ; and 0 otherwise |
| $y_{st}^{\zeta^i}$ | binary variable: 1 if the stand s should be harvested in t given the scenario ζ^i ; 0 otherwise |
| H_t^i | volume harvested in year t under scenario ζ^i (m^3). We omit the superscript i for $t = 0$ |

Parameters

| | |
|--------------------|---|
| r_s | discounted net harvest revenue from stand s if the stand is harvested now (\$) |
| $r_{st}^{\zeta^i}$ | net harvest revenue from stand s if the stand is harvested in year t according to scenario ζ^i (\$) |
| $r_{s0}^{\zeta^i}$ | discounted value of the stand at the end of the planning horizon if it is not harvested during the whole planning horizon under scenario ζ^i (\$) |
| v_s | merchantable yield of stand s in the first period (m^3/ha) |
| $v_{st}^{\zeta^i}$ | projected merchantable yield of stand s if harvested in year t according to scenario ζ^i (m^3/ha) |
| a_s | area of stand s (ha) |
| age_{st} | age of stand s at the end of the planning horizon if harvested in year t (yr) |
| age_s | current age of the stand s (yr) |
| age_{s0} | age of stand s at the end of the planning if not harvested during the planning horizon (yr) |
| α | acceptable lower bound on the fluctuation of volume of wood from one period to the next |
| β | acceptable upper bound on the fluctuation of volume of wood from one period to the next |
| γ | acceptable lower bound fluctuation of volume of wood between two non consecutive periods |
| λ | acceptable upper bound fluctuation of volume of wood between two non consecutive periods |

The objective function (3) maximizes the net present value from subsequent stages after the current harvest (x). This objective function includes the cost of harvest and re-plantation for each forest unit. In addition, this objective function accounts for the value of the stands that are not harvested

during the planning horizon because those stand have a monetary value. Constraint set (4) imposes that if a stand is harvested now, then it cannot be harvested in subsequent years. In other words a forest unit can only be harvested once during the whole planning horizon. The use of w_s variables is to account the stands that are not scheduled for harvest in the whole planning horizon. Constraint set (5) and (6) compute the volume of wood harvested now and in the future periods during the planning horizon, respectively. The volume computed in each period depends on the scenario ζ^t . Constraint sets (7) and (8) impose that the volume fluctuation between two consecutive stages should be within a given lower and upper bounds, respectively. These set of constraints are also known as even flow constraint since they ensure that the volume of timber harvested is almost evenly distributed in time. Even with constraint sets (7) and (8), there is a possibility that the volume harvested declines with time. Hence volume at $t = 4$ might be much lower than volume at $t = 1$, for instance. To attenuate this effect, we impose supplementary wood flow constraints which are constraint sets (9) and (10). These two sets of constraints impose flow restriction between two non-consecutive stages. Constraint set (11) states that the age of the forest at the end of the planning horizon should be greater or equal to the current age of the forest. This constraint is a proxy for sustainability; it ensures that forest resources are not depleted during the planning horizon. Finally, the definition of the variables is given in (12).

2.2. Climate change data

In this paper the uncertainty in forest growth stems from climate change. We present in this section how climate change was translated into forest growth to be suitable to the harvest planning model. Climate change in this project refers to the change in forest growth. Hence, the change can be positive or negative. The change is small in near future, stage 1, compared to the distant future (stage 4). There is a ten year difference between two consecutive stages. In this project, we consider the case of Pacific Northwest where most models forecast that climate change will lead to the increase of forest growth. We report in Table 2 the growth change used for the analysis. This growth change data is based on [21, Table 3]. Forest growth is age dependent and this is reflected in growth change. In addition, this growth change modeling reflect a Geometric Brownian Motion process where the absolute increment of growth are not independent from one period to another although the percentages of change are. The Lower and Upper in the table correspond to the lowest and highest possible growth change, respectively, at each stage. Formally speaking, [Lower, Upper] at stage t , represents Ξ_t of the random parameter ζ_t .

To assess the performance of the proposed modeling framework, we change the lower bound on the growth change by multiplying it by a factor ϵ . This allows to assess the performance of the proposed method for climate change scenarios that forecast decrease in forest growth. We assessed values of ϵ equal 1, 20 and 40. Only $\epsilon = 40$ corresponds to forest decline whereas, ϵ of 20 corresponds to a decrease in forest growth.

Table 2. Forest growth change (%) due to climate change

| Stage | Lower | Upper |
|-------|-------|-------|
| 1 | -1.2 | 11.1 |
| 2 | -2.4 | 22.2 |
| 3 | -3.6 | 33.3 |
| 4 | -4.8 | 44.4 |

2.3. Multistage sampling

We have presented both the optimization model we intend to solve in this paper and the uncertain parameter. In this section, we describe how to sample from the distribution of the uncertain parameter. To achieve that objective, let $\mathcal{N} = \{N_1, \dots, N_T\}$ be a sequence of positive integers. At the first stage (stage 1), we generate N_1 replications drawn from Ξ_1 , the support of the random parameter ζ_1 . To

minimize the number of replications necessary, we generate the N_1 by dividing Ξ_1 into N_1 intervals. From each interval, we sample uniformly one realization of ξ_1 . We repeat this procedure for stage 2 and so forth for the following stages. At the end, we connect the realization at $t = 0$ to all realizations at $t = 1$. We connect all realizations at $t = 1$ to all others at $t = 2$ and we continue until the last stage. The result of this procedure is a scenario tree with a total number scenarios or sample size $n = \prod_{t=1}^T N_t$. Each path from this scenario tree can be viewed as a scenario with probability $1/n$. Varying the values of N_t allows to have different sample sizes which can be solved using SAA as described in the following section. It is not clear, however, whether having $N_1 \geq N_2 \geq \dots \geq N_T$ is preferable to a sampling scheme with $N_1 \leq N_2 \leq \dots \leq N_T$. To answer this question, we tested the solution by considering a sampling scheme with $N_1 = N_2 = \dots = N_T$ in Section 3.1 and evaluate in Section 3.2, the solution behavior when we use the two other sampling schemes.

2.4. Solution method

We used SAA as the method for solving the stochastic harvest scheduling model since the main challenge of the decision maker is to find the set of stands that are suitable for harvest in the first period. SAA method is a Monte Carlo simulation based approach for solving stochastic optimization problems. A random sample is generated with uniform probability distribution, and then the expected value function of the stochastic problem is approximated by the corresponding sample average function. The sample average approximation problem which is specified by the generated sample is then solved by a deterministic optimization technique which is mixed integer programming in this study. This procedure is repeated by increasing the size of the samples to obtain solutions that are close enough to the true optimal solution.

To formally introduce the SAA method we used to solve the harvest scheduling problem, we reformulate the harvest scheduling problem in (1) and (2) in a compact form, just for practicability. We can rewrite the problem as

$$\max_{x \in X} f(x, \zeta) := \max_{x \in X} \mathbb{E} \left[\sum_{s \in \mathcal{S}} [r_s x_s] + R(x, \zeta) \right], \quad (13)$$

where X is the feasible set (constraints (2), (4) to (12)). We can write (13) because the term $r_s x_s$ has no uncertainty.

Let ζ^1, \dots, ζ^n be a sample of size n drawn from the distribution of ζ . The sample is generated using the method described in Section 2.3. We can write the sample average objective function as:

$$f_n(x) := \frac{1}{n} \sum_{i=1}^n f(x, \zeta^i). \quad (14)$$

where

$$f(x, \zeta^i) = \sum_{s \in \mathcal{S}} r_s x_s + \sum_{s \in \mathcal{S}} \left(\sum_{t \in \mathcal{T}} r_{st}^{\zeta^i} y_{st}^{\zeta^i} + r_{s0}^{\zeta^i} w_s^{\zeta^i} \right). \quad (15)$$

The idea of SAA is to solve (16) instead of (13)

$$z_n = \max_{x \in X} f_n(x) \quad (16)$$

We now describe the steps for solving a stochastic optimization problem using SAA as outlined in Figure 1.

Step a & b: Generate a large sample of size (N) and solve it using (16) to obtain a first stage candidate solution \hat{x} . The choice of N depends on the model that can be solved in a reasonable time. The fact that this problem is only solved once partially motivates on choosing the largest possible sample size.

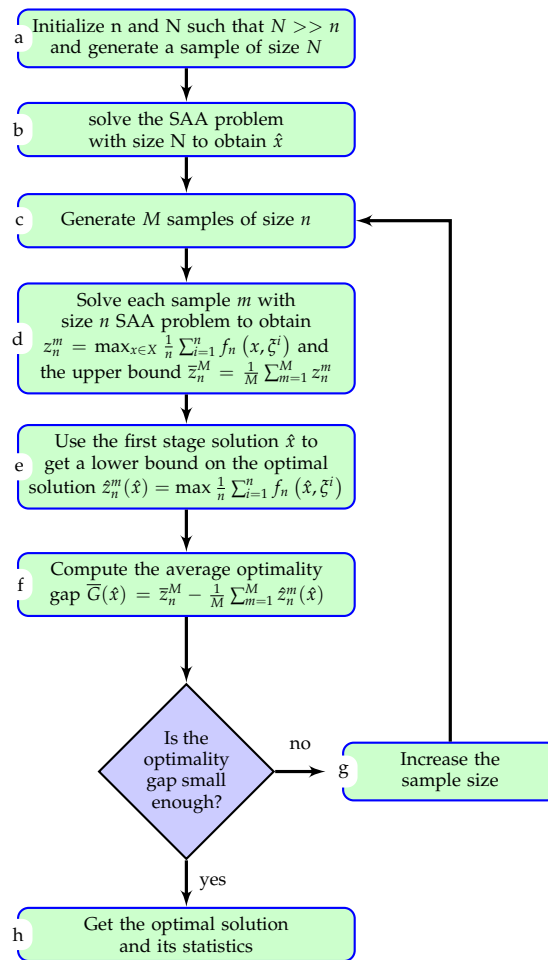


Figure 1. Flow chart of SAA

Step c: Generate M samples of size n . The idea is to start with lower values of n and increase the sample size progressively. [22] describes a procedure for choosing the value of M . In a nutshell, M should be chosen in a way that allows to compute different statistics such as the mean, variance and confidence interval.

Step d: Solve SAA problem of each sample m to get $z_n^m = \max_{x \in X} \frac{1}{n} \sum_{i=1}^n f_n(x, \zeta^i)$ and the upper bound $\bar{z}_n^M = \frac{1}{M} \sum_{m=1}^M z_n^m$. \bar{z}_n^M is an upper bound because it is the average of upper bounds z_n^m obtained by using only a set of scenarios, and therefore the solution is more optimistic.

Step e: For each sample m , fix the first stage variables to \hat{x} obtained from step a & b to get a lower bound on the optimal solution $z_n^m(\hat{x}) = \max \frac{1}{n} \sum_{i=1}^n f_n(\hat{x}, \zeta^i)$

Step f: Compute the average optimality gap as $\bar{G}(\hat{x}) = \bar{z}_n^M - \frac{1}{M} \sum_{m=1}^M z_n^m(\hat{x})$. We can also compute for each sample m the optimality gap as

$$G_n^m(\hat{x}) = \max_{x \in X} \frac{1}{n} \sum_{i=1}^n f_n(x, \zeta^i) - \frac{1}{n} \sum_{i=1}^n f_n(\hat{x}, \zeta^i) \quad (17)$$

where $f_n(\hat{x}, \zeta^i)$ is obtained by fixing the first stage variable in f_n to \hat{x} from Step a. Notice that $G_n^m(\hat{x}) \geq 0$. We can then compute the average optimality gap $\bar{G}(\hat{x}) = \frac{1}{M} \sum_{m=1}^M G_n^m(\hat{x})$ and its variance $s_{\bar{G}}^2 = \frac{1}{M(M-1)} \sum_{m=1}^M (G_n^m(\hat{x}) - \bar{G}(\hat{x}))^2$. [14] proved that the optimality gap depends on the sample size n . In fact, they proved that as $n \rightarrow \infty$, the optimality gap converges to zero and the SAA objective function value converges to the true optimal objective value with probability one. In other words, the sample size is large enough if we have a small optimality gap.

Step g: If the optimality gap computed in step f is not sufficiently small, then increase the sample size n and go back to step c. The decision of whether the optimality gap is sufficiently small and its variance is sufficiently low is domain specific.

Step h: If the optimality gap is small enough, then we have obtained the optimal solution and we can compute the confidence interval on the optimality gap as: $\mu_{\hat{x}} \in [0, \bar{G}(\hat{x}) + t_{M-1, \alpha} S_{\bar{G}}]$ with a confidence of $100(1 - \alpha)$ with α being Type I error.

2.5. Importance of SAA

To assess the advantage of solving the stochastic optimization problem with SAA instead of the deterministic model that only consider the average scenario (the model that ignores climate change uncertainty), we compute the so-called value of stochastic solution (VSS) [23]. Let \bar{x} be the first stage solution we would get if we implemented the average growth solution, respectively. Similarly, let x^* be the optimal solution of the stochastic model obtained from SAA. We denote by z^i and \bar{z}^i the net NPVs of the scenario ζ^i when the first stage variables are fixed to x^* and \bar{x} , respectively. Finally, let $z(\bar{x}) = \frac{1}{|\Omega|} \sum_{\zeta^i \in \Omega} \bar{z}^i$ whereas, $z(x^*) = \frac{1}{|\Omega|} \sum_{\zeta^i \in \Omega} z^i$. We compute VSS as follows: $VSS = z(x^*) - z(\bar{x})$. In term of relative value, we present VSS in basis points (bps)¹ as:

$$VSS(bps) = \frac{VSS}{z(\bar{x})} * 10,000 \quad (18)$$

In case where some scenarios are not feasible after fixing the first stage to \bar{x} , we report the number of infeasible scenarios.

2.6. Experiment

We present in this section, the values of several parameters introduced in the methods in order to conduct a numerical experiment. The numerical experiment was run on a Windows desktop with an AMD 8 Core processor of 4 GHz and 8 GB of RAM. We have used CPLEX 12.8 to solve the stochastic model using python 3 as the modeling language. We solved each model to 0.5% mixed integer programming (MIP) optimality gap. For each model, we generated 30 independent replications ($M = 30$). We used bootstrapping to compute the 95% confidence level for different metrics. N was set to 625 scenarios and we varied the sample size n from 16 to 625. We tested the modeling on Phyllis Leeper forest with 89 stands (<http://ifmlab.for.unb.ca/fmos/datasets/PhyllisLeeper/>). Model parameters α and γ were set to 0.85, whereas β and λ were set to 1.15. The planning horizon was 50 years divided into 5 planning periods of 10 years each.

3. Results

3.1. Effect of sample size on the optimality gap

The results of the experiments are summarized in Figure 2. The sample size is the number of scenarios. These results were obtained by setting $N_1 = N_2 = N_3 = N_4$. The solutions show that as the sample size increases, the NPV decreases converging toward the optimal objective function value. In addition, we have tighter confidence intervals as the sample size increases. A similar behavior is observed for the optimality gap. However, larger sample sizes lead to a significant increase in solution time. We can conclude that solving problems with up to 256 scenarios is sufficient to capture the variability and obtain solutions that are close enough to the true optimal solution since increasing the sample size to 625 did not significantly reduce the optimality gap and its variance.

¹ 1 Basis points (BPS) is equivalent to 0.01%. It is commonly used in economics and finance

Furthermore, although we need larger sample size to reduce the optimality gap and its variance, we need to compromise on solution time. The solution time increases exponentially with the sample size.

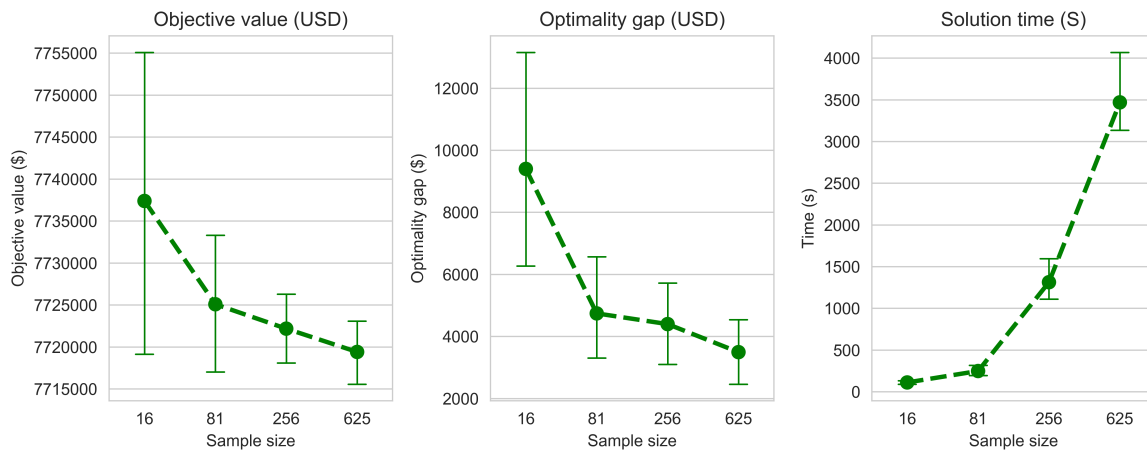


Figure 2. Performance measurement of SAA for harvest scheduling. Error bars indicate 95% confidence interval

3.2. Effect of sampling scheme on the optimality gap

In this section, we test the effect of scenario generation scheme on the optimality gap. The scheme used to generate scenarios for the four stages is presented in Figure 3. In the figure, the sampling scheme represents four digits corresponding to N_1 , N_2 , N_3 and N_4 . For instance, 1155 means that $N_1 = 1$, $N_2 = 1$, $N_3 = 5$ and $N_4 = 5$ leading to $1 \cdot 1 \cdot 5 \cdot 5 = 25$ scenarios. This means that the schemes 1555 and 5551 lead to the same number of scenarios (125 scenarios). As we can see from Figure 3, the optimality gap and its variability when $N_1 = 1$ is larger than when $N_1 > 1$. The lowest optimality gap is obtained with $N_1 = 3$. In general having higher values of N_t for higher t seems to be favorable.

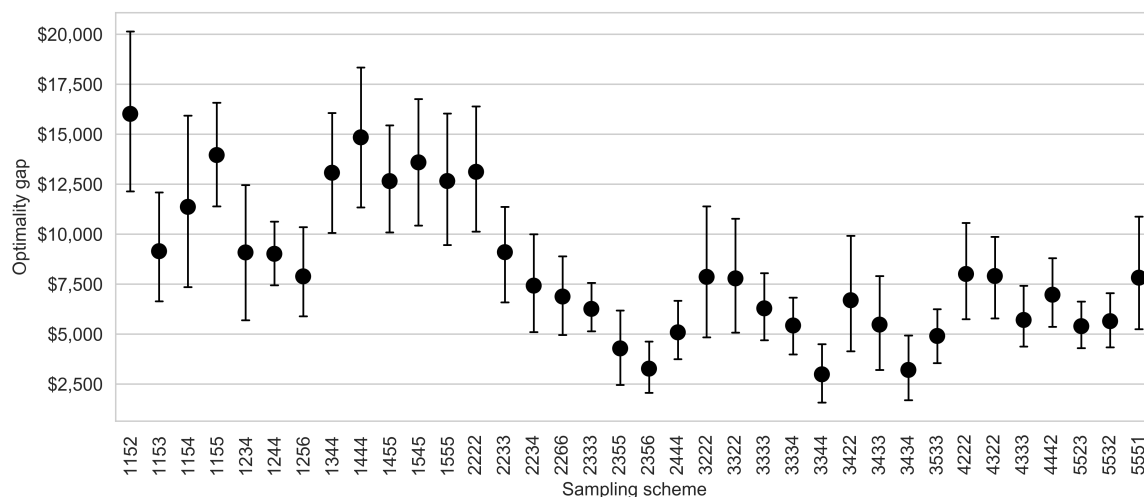


Figure 3. Change of optimality gap values due to the sample size and the configuration of the sampling scheme using severe climate change data

3.3. Advantage of SAA in stochastic harvest scheduling over the deterministic approach

The stochastic optimization solution allows to have superior solution in terms of NPV. In addition, it allows to have solutions that are robust (feasible) for a set of climate change scenarios that are not

feasible if we implement here and now the deterministic solution (Table 3). As the uncertainty on forest growth change magnitude increases, the value of stochastic solution increases as well. When we consider $\epsilon = 20$, out of 200 scenarios, 3 scenarios were infeasible.

Table 3. Value of the stochastic solution

| $ \Omega $ | ϵ | $z(\bar{x})$ | $z(x^*)$ | infeasible scen. | VSS | VSS(bp) |
|------------|------------|--------------|-----------|------------------|-----------|----------|
| 256 | 1 | 7,775,108 | 7,782,679 | 0 | 7,571 | 9.96 |
| 200 | 20 | 7,160,950 | 7,256,069 | 3 | 95,119 | 132.83 |
| 200 | 40 | 4,940,204 | 6,737,368 | 60 | 1,797,163 | 3,637.83 |

4. Discussion and conclusions

Climate change is a serious issue in forest management planning. In a study conducted in Norway, the majority of forest managers showed the importance of addressing climate change in forest planning [24]. Several researchers conducted similar studies in different ecosystems and reached analogous results [25,26]. To ensure the sustainability of forest resources, forest managers need to incorporate forest growth uncertainty in harvest scheduling models. In this work, we formulated a stochastic harvest scheduling model with forest growth uncertainty due to climate change and solved the model using sample average approximation (SAA). We tested the modeling and solution using climate change data transformed to forest growth change. We compared the robustness of the stochastic solution to the deterministic one by randomly generating a set of scenario and comparing the expected NPV if we implement the stochastic solution and the NPV if we use the deterministic solution. The numerical results showed that SAA allows to have stochastic solutions that are close enough to the true optimal solution when the sample size is large enough. However, large sample size lead to an exponential increase of solution time. This pattern of the computational complexity growing exponentially with the sample size was previously suggested by [22].

One of the main limitation of the proposed method is the computation required in step d of the algorithm presented in Figure 1. As we discussed in the previous paragraph, the increase of the sample size leads to an exponential increase in solution time. It is therefore important to have a strategy that allows to reduce the sample size while producing solutions that allow the convergence of SAA solutions to the true optimal solutions. Our tests suggest that with the appropriate sampling scheme, it is possible to reach convergence with smaller sample sizes. For instance, the sampling schemes 2356 and 3344 yielding sample sizes of 180 and 144, respectively, have smaller optimality gaps and variances compared to a sample size of 625 which stemmed from a sampling scheme of 5555. In conclusion, when we adopt the adequate sampling strategy that allows to sufficiently explore the first stage and generate large samples for future stages, we can limit the number of scenarios necessary for the SAA solution to converge to the true optimal solution.

The proposed model not only allows the managers to make intelligent decision now, but also allows the preservation of forest resources that take time to replenish once depleted. Indeed, the stochastic solution is robust to different climate scenario whereas, the deterministic one is infeasible to many of the tested climate scenarios. These results are in line with the ones obtained by [12,13]. The infeasibility of the deterministic solution stems mainly from the fact that the wood flow constraints are violated due to an intensive harvest in early periods.

In addition, the solution method proposed in this paper is easy to develop and implement by many forest managers. Unlike the methods used in [11,13] that require to know the probability distribution of the random parameter, SAA does not require such information. The method relies on the fact that if the sample size is large enough, then the sample statistics will approximate those of the actual population. The method is also suitable for many applications where the objective function cannot be computed in a closed form such as the so-called black-box optimization problems [27] and the stochastic knapsack problem [22]. In forestry, this method is well suited for harvest scheduling problem with wood price

and demand uncertainties because we can extract samples from historical demand and price without the need to model the price like done in [10,28]. Compared to stochastic programming which require the so-called non-anticipativity constraints [13,29], SAA model is relatively smaller in terms of number of constraints (and possibly in terms of number of variables, depending on the formulation) since it does not require such constraints.

Regarding climate change data, we used the information of climate change forecast with statistical models developed in [30]. Unlike the data from [11–13], which originated from a process-based modeling, statistical models of forest growth under climate change are much more common (for e.g [31]). Hence, the method developed in this study can easily be extended to other forest systems. Moreover, it is straight forward to translate forecast of precipitation, air moisture and temperature into forest growth compared to processed-based models which require the expertise in plant biology.

Finally, this research can be extended by considering other sources of uncertainty such as the price of wood, or the demand of forest products. In this research we assumed that climate change does not lead to species transition. In some cases, one might need to consider the species shift because of climate change. It would be interesting to include this information in the decision-making process and provide to forest practitioners a range of options when implementing harvest scheduling plans and the choice of species for regeneration.

Funding: This research received funding of Precision Forestry Cooperative of the School of Environmental and Forest Sciences, University of Washington

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Deterministic harvest scheduling model

The objective function (A1) is to maximize the net present value from forest harvest. This objective function includes the cost of harvest and re-plantation for each forest unit. In addition, this objective function accounts for the value of the stands that are not harvested during the planning horizon because those stand have a monetary value. Constraint set (A2) imposes that if a stand is harvested now, then it cannot be harvested in subsequent years. In other words a forest unit can only be harvested once during the whole planning horizon. The use of n_s variables is to account the stands that are not scheduled for harvest in the whole planning horizon. Constraint set (A3) compute the volume of wood harvested now and in the future periods during the planning horizon, respectively. Constraint sets (A4) and (A5) impose that the volume fluctuation between two consecutive stages should be within a fixed lower and upper bounds, respectively. These set of constraints are also known as even flow constraint since they ensures that the exploitation of forest resources is evenly distributed in time. Even with constraint sets (A4) and (A5), there is a possibility that the volume harvested declines with time. Hence volume at $t = 4$ is much lower than volume at $t = 0$, for instance. To attenuate this effect, we impose supplementary wood flow constraints which are constraint sets (A6) and (A7). These two set of constraints impose flow restriction between two non consecutive stages. Constraint set (A8) states that the age of the forest at the end of the planning horizon should be greater or equal to the current age of the forest. This constraint is a proxy for sustainability; it ensures that forest resources are not depleted during the planning horizon. Finally, the definition of the variables are given in (A9).

$$\max \sum_{s=1}^{|\mathcal{S}|} \left[\sum_{t=0}^T r_{st} x_{st} + r_{s0} w_s \right] \quad (\text{A1})$$

subject to

$$n_s + \sum_{t=0}^T x_{st} = 1 \quad s = 1, \dots, |\mathcal{S}| \quad (\text{A2})$$

$$\sum_s v_{st} x_{st} = H_t \quad t = 0, \dots, T \quad (\text{A3})$$

$$H_t^i \leq \alpha H_{t-1} \quad t = 1, \dots, T \quad (\text{A4})$$

$$H_t^i \geq \beta H_{t-1} \quad t = 1, \dots, T \quad (\text{A5})$$

$$H_t \leq \gamma H_{t-2} \quad t = 2, \dots, T \quad (\text{A6})$$

$$H_t^i \geq \lambda H_{t-2} \quad t = 2, \dots, T \quad (\text{A7})$$

$$\sum_{s \in \mathcal{S}} a_s \left[\sum_{t=0}^T \text{age}_{st} x_{st} + \text{age}_{s0} w_s \right] \geq \sum_s a_s \text{age}_s. \quad (\text{A8})$$

$$w_s \in \{0, 1\}, \forall s, x \in \{0, 1\}^{|\mathcal{S}| \times T+1}, H_t \geq 0 \quad \forall t \quad (\text{A9})$$

References

- Mäkinen, A.; Borges, J.G. Review . Assessing uncertainty and risk in forest planning and decision support systems : review of classical methods. *Forest Systems* **2013**, *22*, 282–303.
- Ross, K.L.; Tóth, S.F. A model for managing edge effects in harvest scheduling using spatial optimization. *Scandinavian Journal of Forest Research* **2016**, *31*, 646–654. doi:10.1080/02827581.2016.1213877.
- Dixon, B.L.; Howitt, R.E. Resource Production Under Uncertainty: A Stochastic Control Approach to Timber Harvest Scheduling. *American Journal of Agricultural Economics* **1980**, *62*, 499–507. doi:10.2307/1240204.
- Kooten, G.C.V.; Kooten, R.E.; Brown, L.G. Modeling the effect of uncertainty on timber harvest: A suggested approach and empirical example. *Journal of Agricultural and Resource Economics* **1992**, *17*.
- Pukkala, T. Multiple risks in multi-objective forest planning : integration and importance. *Forest Ecology and Management* **1998**, *111*, 265–284. doi:10.1016/S0378-1127(98)00339-9.
- Alonso-Ayuso, A.; Escudero, L.F.; Guignard, M.; Quinteros, M.; Weintraub, A. Forestry management under uncertainty. *Annals of Operations Research* **2011**, *190*, 17–39. doi:10.1007/s10479-009-0561-0.
- Veliz, F.B.; Watson, J.P.; Weintraub, A.; Wets, R.J.B.; Woodruff, D.L. Stochastic optimization models in forest planning: a progressive hedging solution approach. *Annals of Operations Research* **2014**, pp. 259–274. doi:10.1007/s10479-014-1608-4.
- Piazza, A.; Pagnoncelli, B.K. The optimal harvesting problem under price uncertainty. *Annals of Operations Research* **2014**, *217*, 425–445. doi:10.1007/s10479-014-1559-9.
- Pagnoncelli, B.K.; Piazza, A. The optimal harvesting problem under price uncertainty: the risk averse case. *Annals of Operations Research* **2017**, *258*, 479–502. doi:10.1007/s10479-015-1963-9.
- Alonso-Ayuso, A.; Escudero, L.F.; Guignard, M.; Weintraub, A. Risk management for forestry planning under uncertainty in demand and prices. *European Journal of Operational Research* **2018**, *267*, 1051–1074. doi:10.1016/j.ejor.2017.12.022.
- Garcia-Gonzalo, J.; Pais, C.; Bachmatiuk, J.; Weintraub, A. Accounting for climate change in a forest planning stochastic optimization model. *Canadian Journal of Forest Research* **2016**, *46*, 1111–1121. doi:10.1139/cjfr-2015-0468.
- Álvarez-Miranda, E.; Garcia-Gonzalo, J.; Ulloa-Fierro, F.; Weintraub, A.; Barreiro, S. A multicriteria optimization model for sustainable forest management under climate change uncertainty: An application in Portugal. *European Journal of Operational Research* **2018**, *269*, 79–98. doi:10.1016/j.ejor.2017.04.052.
- Garcia-Gonzalo, J.; Pais, C.; Bachmatiuk, J.; Barreiro, S.; Weintraub, A. A Progressive Hedging Approach to Solve Harvest Scheduling Problem under Climate Change. *Forests* **2020**, *11*, 224. doi:10.3390/f11020224.
- Mak, W.K.; Morton, D.P.; Wood, R.K. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters* **1999**, *24*, 47–56. doi:10.1016/S0167-6377(98)00054-6.
- Wang, W.; Ahmed, S. Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters* **2008**, *36*, 515–519. doi:10.1016/j.orl.2008.05.003.
- Schütz, P.; Tomasgard, A.; Ahmed, S. Supply chain design under uncertainty using sample average approximation and dual decomposition. *European Journal of Operational Research* **2009**, *199*, 409–419. doi:10.1016/j.ejor.2008.11.040.
- Chunlin, D.; Liu, Y. Sample Average Approximation Method for Chance Constrained Stochastic Programming in Transportation Model of Emergency Management. *Systems Engineering Procedia* **2012**, *5*, 137–143. doi:10.1016/j.sepro.2012.04.022.

18. Emelogu, A.; Chowdhury, S.; Marufuzzaman, M.; Bian, L.; Eksioglu, B. An enhanced sample average approximation method for stochastic optimization. *International Journal of Production Economics* **2016**, *182*, 230–252. doi:10.1016/j.ijpe.2016.08.032.
19. Pour, A.G.; Naji-Azimi, Z.; Salari, M. Sample average approximation method for a new stochastic personnel assignment problem. *Computers and Industrial Engineering* **2017**, *113*, 135–143. doi:10.1016/j.cie.2017.09.006.
20. Mohammadi Bidhandi, H.; Patrick, J. Accelerated sample average approximation method for two-stage stochastic programming with binary first-stage variables. *Applied Mathematical Modelling* **2017**, *41*, 582–595. doi:10.1016/j.apm.2016.09.019.
21. Latta, G.; Temesgen, H.; Adams, D.; Barrett, T. Analysis of potential impacts of climate change on forests of the United States Pacific Northwest. *Forest Ecology and Management* **2010**, *259*, 720–729. doi:10.1016/j.foreco.2009.09.003.
22. Kleywegt, A.J.; Shapiro, A.; Homem-De-Mello, T. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* **2001**, *12*, 479–502.
23. Birge, J.R. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming* **1982**, *24*, 314–325. doi:10.1007/BF01585113.
24. Heltorp, K.M.A.; Kangas, A.; Hoen, H.F. Do forest decision-makers in Southeastern Norway adapt forest management to climate change? *Scandinavian Journal of Forest Research* **2018**, *33*, 278–290. doi:10.1080/02827581.2017.1362463.
25. Scheller, R.M.; Parajuli, R. Forest management for climate change in New England and the Klamath Ecoregions: Motivations, practices, and barriers. *Forests* **2018**, *9*. doi:10.3390/f9100626.
26. Liu, K.; He, H.; Xu, W.; Du, H.; Zong, S.; Huang, C.; Wu, M.; Tan, X.; Cong, Y. Responses of Korean pine to proactive managements under climate change. *Forests* **2020**, *11*, 1–19. doi:10.3390/f11030263.
27. Kim, S.; Ryu, J.h. The sample average approximation method for multi-objective stochastic optimization. Proceedings of the 2011 Winter Simulation Conference (WSC). IEEE, 2011, pp. 4021–4032. doi:10.1109/WSC.2011.6148092.
28. Rios, I.; Weintraub, A.; Wets, R.J. Building a stochastic programming model from scratch: a harvesting management example. *Quantitative Finance* **2016**, *16*, 189–199. doi:10.1080/14697688.2015.1114365.
29. Bagaram, M.B.; Jaross, W.; Weintraub, A. A Parallelized Variable Fixing Process for Solving multistage Stochastic Programs with Progressive Hedging An Application to Harvest Scheduling in the Face of Climate Change A Parallelized Variable Fixing Process for Solving multistage Stochastic Programs. *preprint* **2020**. doi:10.13140/RG.2.2.21150.15687/1.
30. Latta, G.; Temesgen, H.; Barrett, T.M. Mapping and imputing potential productivity of Pacific Northwest forests using climate variables. *Canadian Journal of Forest Research* **2009**, *39*, 1197–1207. doi:10.1139/X09-046.
31. Elli, E.F.; Sentelhas, P.C.; Bender, F.D. Impacts and uncertainties of climate change projections on Eucalyptus plantations productivity across Brazil. *Forest Ecology and Management* **2020**, *474*, 118365. doi:10.1016/j.foreco.2020.118365.