Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Evaluating the required scenario set size for stochastic programming in forest management planning: incorporating inventory and growth model uncertainty

Journal:	<i>Canadian Journal of Forest Research</i>
Manuscript ID	cjfr-2014-0513.R2
Manuscript Type:	Article
Date Submitted by the Author:	17-Aug-2015
Complete List of Authors:	Eyvindson, Kyle; University of Helsinki, Kangas, Annika; University of Helsinki
Keyword:	Stochastic programming, Sample average approximation, Even-flow forestry, Risk, Uncertainty

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

**Evaluating the required scenario set size for stochastic programming in
forest management planning: incorporating inventory and growth model
uncertainty**

Kyle Eyvindson*, Annika Kangas

*University of Helsinki, Department of Forest Sciences, Latokartanonkaari 7 (P.O. Box 27) 00014
University of Helsinki, Finland

Keywords: Stochastic programming, Sample average approximation, Even-flow forestry, Risk,
Uncertainty

Abstract

Developing a plan of action for the future use of forest resources requires a way to predict the development of the forest through time. These predictions require the use of inventory data and growth models which contain a large number of uncertainties. These uncertainties impact the quality of the predictions and if the uncertainties are not accounted for, can lead to the selection of a suboptimal management plan. To account for and manage the uncertainties and the associated risk we have explored the use of stochastic programming. Stochastic programming can integrate uncertainty into the optimization process by solving the problem for a large number of potential scenarios of the forests future development. The selection of an appropriately sized set of scenarios involves a trade-off between tractability issues and problem representation issues. In this paper an analysis of the trade-offs is conducted. Two cases are studied, one where only the uncertainty of the inventory data is included, and a second where both growth model and inventory data uncertainty are included. The impact on increasing the number of scenarios to the problem representation is examined through a simple even-flow problem.

Introduction

In forest management, the development of high quality plans requires detailed knowledge regarding the current situation of the forest holding, the ability to forecast the development of the holding, and the preferences of the stakeholders that may be included in the decision making process. The management plan is a set of future actions to be taken in the forest over a set period of time. In forest management planning these decisions could be the timing of a harvest or thinning, the planting, seeding or natural regeneration of a stand, or silvicultural treatments to promote the establishment and growth of a stand.

1 These management actions are applied to a potential future forest holding, predicted using growth
2 models (for an example of the Finnish context for stand-level models see: Vuokila and Väliäho, 1980;
3 Mielikäinen, 1985; Huuskonen and Miina, 2007, for tree-level models Hynynen et al. 2002). Even with
4 accurate models, forecasting the future is never error-free (Diebold, 2001). The growth models predict
5 the future state of the forest using the most recent inventory of the forest holding. As the current state
6 of the forest is uncertain, this will have an additional impact on the predictions of future forest
7 resources. Even using current state-of-the-art forest inventory techniques, there is still a minor but
8 significant level of uncertainty (Næsset 2004; Packalén and Maltamo 2007). Thus, the forecast of forest
9 resources will be imperfect even if perfect growth models of forest resources were available. Moreover,
10 the growth process includes natural variability (e.g. weather) which cannot be reduced with improved
11 models (Ferson and Ginzburg 1996).

12 Evaluating the uncertainty of forest inventory methods and growth models is an ongoing research area.
13 Both inventory method and growth models should be validated by obtaining estimates of the precision
14 and accuracy of the model or method (Gregoire and Valentie 2008; Pretzch 2009). Sample based
15 inventory methods utilize analytic formulas to assess accuracy, while the assessment of the stand level
16 inventory is often compared to field data of very high quality. The assessment of the accuracy of
17 growth models are evaluated through time series data. Haara and Leskinen (2009) have evaluated the
18 effects of both uncertainties, and used the information to update stand-level inventory data.

19 When creating optimized forest management plans, the standard method of handling the uncertainties
20 contained within the models and data has been to conduct the optimizations under the assumption that
21 there was no uncertainty. These assumptions allowed for the use of relatively simple deterministic
22 optimization processes, through the use of linear programming or heuristic optimization approaches
23 (Kangas et al. 2008). However, some studies have been conducted in managing uncertainty. For

instance, some researchers have focused on ensuring the feasibility of meeting the demands of the constraints under uncertainty. Hof and Pickens (1991) used chance-constrained and chance maximizing approaches to minimize the probability of not meeting a specified demand target. Palma and Nelson (2009) used robust optimization to ensure feasibility of yield and demand of timber constraints subject to a specified protection level, while maximizing net present value. The use of stochastic goal programming (Eyvindson and Kangas, 2014) has been used to develop forest management plans which minimize the expected deviations from the targets set for those criteria of interest to the decision maker. Kangas et al. (2014a) have utilized a two-stage stochastic programming method to define the optimal measurement strategy to maximize the net present value of the forest holding. Additionally, Kangas et al. (2015) have utilized a multi-stage stochastic programming framework to balance between the cost of obtaining higher quality (or perfect) information and the improvement in the decisions made.

Stochastic programming has additional properties which makes it more attractive than the deterministic counterpart. As stochastic programming deals with the uncertainty of the problem, this allows for the management of different kinds of risk (Birge and Louveaux 2011). Different approaches to risk management could be utilized, such as the minimization of the downside mean semideviation (Krzemienowski and Ogryczak 2005) or the Conditional Value at Risk (Rockafeller and Uryasev 2000). The selection of the specific risk measure used should depend upon both the general problem formulation preferences and risk preferences of the decision maker.

Earlier studies have evaluated the value of information (VOI), or the losses of making improper decisions for problems which maximize net present value (NPV). Holmström et al. (2003) and Eid et al. (2004) used a cost-plus-loss analysis to evaluate different sampling strategies and inventory methods. Alternatively, Bayesian decision analysis (Lawrence 1999) could also be used to evaluate the

1 VOI. In a forestry context, Kangas (2010) used Bayesian decision analysis to evaluate the value of
2 improving the accuracy of information. These methods evaluate the VOI on a stand level, and cannot
3 include holding level constraints, such as maintenance of a specified amount of old growth forest or
4 even-flow requirements. When these kinds of constraints are included, the VOI can be calculated
5 through stochastic programming (chapter 4 of Birge and Louveaux, 2011).

6 The use of stochastic programming in forest science has primarily focused on evaluating the impacts of
7 uncertain events such as forest fires (Gassmann 1989; Boychuk and Martell 1996). The research has
8 focused on how to prepare for fire events and manage the risk (Hof and Pickens 1991; Bevers 2007). In
9 a forest management context, Eriksson (2006) highlighted how stochastic programming could be used
10 to plan for the uncertainty regarding climate change.

11 Depending on how the stochastic program is structured, adaptive management alternatives can be
12 integrated into the optimization process (Lohmander 2007). Adaptive management is a structured
13 process of learning about the problem and adapting the management decisions based on that learning
14 (Williams 2011). In a stochastic process, learning could be interpreted as the resolution of some
15 uncertainty. Two-stage or multi-staged stochastic programs with general recourse (rather than simple
16 recourse) could integrate possible outcomes of learning and suggest appropriate management decisions
17 based on the different possible outcomes of what had been learned. In the case of inventory / growth
18 model uncertainty, learning and adaptive management would require including probability of a new
19 inventory into the problem formulation (Kangas et al. 2014, 2015).

20 Even when new inventory is possible, the choice of utilizing general recourse over simple recourse
21 should depend upon the potential improvement to the decision process and the tractability of the
22 problem. Allowing for the resolution of uncertainty through additional stages adds complexity to the

1 model. When adding complexity to any model, the issue of parsimony must be addressed. In a decision
2 framework, the question must be asked: Does this improve the quality of the decision enough to justify
3 the added complexities? Sometimes it is simply not of much benefit to create a general recourse
4 problem, because you can simply re-run the simple recourse problem once the uncertainty becomes
5 resolved. Thus, even though options for adaptive management may not be explicitly detailed in the
6 optimization process, this does not preclude the option of, when uncertainty becomes resolved, that the
7 model can be re-run utilizing the updated information. If there is no possibility of resolving uncertainty,
8 simple recourse is the only option. Stochastic programming incorporates the known or assumed
9 uncertainty directly into the problem formulation (Birge and Louveaux 2011). When the estimate of
10 uncertainty is a continuous variable, the uncertainty can be approximated deterministically through
11 several different methods, such as Monte Carlo methods (Shapiro 2003) or Quasi-Monte Carlo methods
12 (Lemieux 2009). Through the creation of a scenario tree, the stochastic program becomes a
13 deterministic approximation, and can be solved using standard linear programming methods. When
14 generating the scenario tree, there are trade-offs which must be considered. If the number of scenarios
15 required to approximate the problem is too large, then the problem may be intractable (King and
16 Wallace 2012). On the other hand, if too few scenarios are used, then the problem may not have an
17 appropriate representation. One approach to manage this problem is to use an approach which
18 optimizes the development of the scenario tree, to satisfy particular statistical properties (Gülpinar et al.
19 2004).

20 The ability to quantify how well a set of scenarios approximates the true stochastic problem is an area
21 which has had significant attention in the stochastic programming community. In general, determining
22 how well the approximation represents the ‘true’ problem is based on finding a statistical bounds for
23 the optimal solution; other tools for evaluating the scenario generation methods are detailed by Kaut

1 and Wallace (2007). For Monte Carlo scenario generation one method is to use the Sample Average
2 Approach (Kleywegt et al. 2002) which uses an algorithm to determine the appropriate number of
3 scenarios required to produce a solution with an acceptable optimality gap and variability. A variety of
4 other methods can also be used to establish optimality bounds for the stochastic problem. Each method
5 has different advantages and disadvantages, and the methods may be better suited to specific problem
6 types and sampling methods used. For instance, Higle and Sen (1996) use stochastic decomposition for
7 stochastic linear programs, which incorporate approximations which are statistically motivated. Mak et
8 al. (1999) use Monte Carlo sampling and focus on estimating the optimality gap for discrete
9 approximations of the stochastic programs.

10 The objective of this paper is to show how the problem formulation and uncertainty involved affects
11 the required number of scenarios to ensure a specific optimality gap. The impact of how large the
12 scenario size must be for a corresponding optimality gap is analyzed with respect to changing risk
13 preferences (from risk neutral to risk averse) and to the level and nature of the uncertainty (only
14 inventory errors, and both inventory and growth model errors). For stochastic programming problems,
15 the number of scenarios required to approximate the true stochastic programming problem need not be
16 excessive. In this paper, the SAA method is used to demonstrate the optimality gap between the
17 estimated ‘true’ stochastic programming problem and the approximated stochastic program. The
18 method relies upon statistical techniques which should be rather familiar to forest inventory specialists
19 and forest biometricians. This paper details the SAA method, provides a case study, and discusses the
20 importance of including uncertainty into forest management planning.

21 **Methods**

22 The Sample Average Approach Method

The quality of the solutions created using different sized sets of scenarios is evaluated using the SAA method (Kleywegt et al. 2002). The SAA methodology allows us to approximate a stochastic program in such a way that allows us to estimate the optimality of the solutions it yields.. The algorithm stops according to a specified stopping criterion, either when the optimality gap is small enough, or when the predefined number of iterations has been exceeded.

To calculate an estimate of the optimized objective function, we must first explain the general formulation. The general stochastic programming formulation follows:

$$[1] \quad \min_{x \in \mathcal{S}} \{f(x) := \mathbb{E}_P F(x, W)\}$$

where W is a random vector with the probability distribution P , $\mathcal{S}$ is set of decisions, $F(x, W)$ is a function of two vector variables x and W , and $\mathbb{E}_P F(x, W)$ is the expected value function.

The optimal values for the solution to the problem will be denoted as v^* :

$$[2] \quad v^* := \min_{x \in \mathcal{S}} f(x)$$

An approximation of the general stochastic programming formulation is the sample average approximation:

$$[3] \quad \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^N F(x, W^j)$$

1 where each $W^j, j=1, \dots, N$, is a realization of the random vector W , and N is the number of realizations.
 2 F indicates the problem formulation with only one scenario, while f represents the stochastic
 3 formulation.

4 The corresponding approximate optimal values for the solution to the problem will be denoted as $\hat{v}_N$:

$$[4] \quad \hat{v}_N := \min_{x \in \mathcal{S}} \hat{f}_N(x),$$

or if a maximization problem: $\hat{v}_N := \max_{x \in \mathcal{S}} \hat{f}_N(x)$,

5 An estimator of v^* , is the average of a large number of $\hat{v}_N$:

$$[5] \quad \bar{v}_N^M := \frac{1}{M} \sum_{m=1}^M \hat{v}_N^m$$

6 where M is the total number of repetitions of the algorithm. This estimator provides an expected lower
 7 bound of the true optimal value, and $\bar{v}_N^M$ is a conservative estimate (Mak et al. 1999). This bias
 8 monotonically decreases as N increases.

9 For each $\hat{v}_N^m$ and for each m in M , there exists an optimal solution $\hat{x}_N^m$ to the corresponding SAA
 10 problem. To calculate the estimator for the performance bounds, an estimator which uses a much larger
 11 number of scenarios can be used:

$$[6] \quad \bar{f}_{N'}(\hat{x}_N^M) := \frac{1}{M} \sum_{m=1}^M \hat{f}_{N'}(\hat{x}_N^m)$$

12

13 where $N' \gg N$. This is an unbiased estimator of $f(\hat{x})$.

14 From these values, it is relatively easy to calculate the optimality gap. The true optimality gap is:

$$[7] \quad f(x) - v^*$$

1 The estimated optimality gap is calculated as:

$$[8] \quad \bar{f}_{N'}(\hat{x}_N^M) - \bar{v}_N^M$$

2 The variance of the estimates can be calculated, and the variance of $\bar{v}_N^M$ is estimated as:

$$[9] \quad \frac{S_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M (\hat{v}_N^m - \bar{v}_N^M)^2$$

3 and the variance of the optimality gap, $\bar{f}_{N'}(\hat{x}_N^M) - \bar{v}_N^M$ is estimated as:

$$[10] \quad \frac{\bar{S}_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M [(\hat{f}_{N'}(\hat{x}_N^m) - \hat{v}_N^m) - (\bar{f}_{N'}(\hat{x}_N^M) - \bar{v}_N^M)]^2$$

4

5 For a specific N , it is possible to estimate the probability that the solution produced will be within a
6 specific optimality gap.

$$[11] \quad \hat{P}[\bar{v}_N^M - \hat{f}(\hat{x}_N^m) \leq g \bar{v}_N^M] = M_g / M$$

7 where M_g is the total number times from $m=1, \dots, M$ that $\bar{v}_N^M - \hat{f}(\hat{x}_N^m) \leq g \bar{v}_N^M$, and g is the specific
8 optimality gap under consideration. Through the calculation of the estimate of the variance, confidence
9 limit can be produced. The estimated variance is calculated as:

[12]

$$\widehat{Var}[\hat{P}] = \frac{M_d \left(1 - M_d/M\right)}{M(M-1)}$$

1

2 Then the 95 % confidence limit is calculated as:

[13]

$$2(\widehat{Var}[\hat{P}])^{1/2}$$

3 The algorithm (Kleywegt et al. 2002) is:

4 1- Select the initial sample sizes for N , N' , and M . M should be set to allow for an accurate
5 estimation of the variances involved.

6 2- For $m=1, \dots, M$ do the following steps:

7 2.1. Generate a sample size of N and solve the problem to calculate $\hat{v}_N^m$ (equation [4])
8 and the solution $\hat{x}_N^m$.

9 2.2. Generate a sample size of N' and calculate $\hat{f}_{N'}(\hat{x}_N^m)$ (equation [3]).

10 3- Estimate the optimality gap for all m in M using equation [8] and the variance of the gap
11 estimator (equation [10]).

12 4- If the estimates of the gap or variance of the gap estimator is larger than desired, increase
13 the number of sample sizes of N and/or N' , and return to step 2.

14 5- Select the best solution ($\hat{x}$) amongst the candidate solutions ($\hat{x}_N^m$) using a screening
15 selection method and then stop.

16 To summarize, the goal of the SAA method is to find the number of scenarios required which provide
17 solutions which are within the acceptable gap of the estimated optimal solution possible.

The forest planning problem

An example of the SAA method will be applied to a forest management planning case. The objective of the problem is to maximize first period timber while ensuring an even-flow of timber provided by forest operations and ensuring continued even-flow possibilities past the planning horizon through an end inventory constraint (A more detailed analysis of a similar problem is presented in Eyvindson and Kangas 2015). The harvest scheduling problem used in this paper is a well-known traditional problem, which allows readers to focus on the application of the stochastic programming models. This approach can be used for any properly developed stochastic programming formulation.

The general problem can be formulated as:

$$[14] \quad \max z = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{K_j} p_i c_{jki1} x_{jk} - \lambda \sum_{i=1}^I p_i \left(\sum_{t=1}^T (w_t^- d_{it}^-) + w_E^- d_{iE}^- \right)$$

Subject to:

$$[15] \quad \sum_{j=1}^J \sum_{k=1}^{K_j} c_{jkit} x_{jk} - \sum_{j=1}^J \sum_{k=1}^{K_j} c_{jki1} x_{jk} - d_{it}^+ + d_{it}^- = 0, \forall i = 1, \dots, I, t = 2, \dots, T$$

$$[16] \quad \sum_{j=1}^J \sum_{k=1}^{K_j} PV_{jkiT} x_{jk} - d_{iE}^+ + d_{iE}^- = \sum_{j=1}^J PV_{j0i}, \forall i = 1, \dots, I,$$

$$[17] \quad d_{it}^+, d_{it}^- \geq 0, \forall i = 1, \dots, I, t = 2, \dots, T$$

$$[18] \quad d_{iE}^+, d_{iE}^- \geq 0, \forall i = 1, \dots, I$$

$$[19] \quad \sum_{k=1}^{K_j} x_{jk} = 1, \forall j = 1, \dots, J$$

$$[20] \quad x_{jk} \in [0,1]$$

1 where the problem inputs are comprised of: c_{jkit} is the quantity of timber harvested from stand j ,
2 schedule k , scenario i , and period t , PV_{jkiT} is the productive value of the stand j in schedule k for
3 scenario i at the end of the planning horizon and PV_{j0i} is the productive value in the beginning of the
4 planning horizon for scenario i . The productive value is the value of the forest stock predicted using the
5 models from Pukkala (2005), using a set of variables as predictors (basal area, mean diameter at breast
6 height, discount rate, site variables and timber price). w_t^- (w_t^+) and w_E^- (w_E^+) are the weights associated
7 with negative (positive) deviations, λ is a risk coefficient which is dependent on the DM's risk
8 preferences. The λ balances between the deviations of the even-flow problem and the maximization of
9 the first period timber. The number of periods is T (6), J is the number of stands and K_j the number of
10 treatment schedules for stand j , I is the total number of scenarios under consideration, and p_i refers to
11 the probability of scenario i occurring. The decision variables are: x_{jk} selects the proportion of stand j
12 being managed using schedule k , where a schedule is a set of harvesting (e.g. thinning or final felling)
13 or management actions (e.g. planting, fertilizing or pre-commercial thinnings) taken during the
14 planning periods, d_{it}^- (d_{it}^+) is the negative (positive) deviation for scenario i , in period t , and d_{iE}^- (d_{iE}^+)
15 is the negative (positive) deviation for the end-inventory value of the present value of the forests for
16 scenario i . This is a simple recourse formulation which manages the downside risk of not meeting

either the even-flow constraints or end volume constraint. In this formulation, the weights reflect the importance of minimizing the expected negative deviations.

In this problem formulation, there are two key aspects which should be described in detail. The first objective is to maximize the first period harvest, and the second objective is to minimize the weighted negative deviations for the subsequent period harvests while also minimizing the weighted negative deviations of the end productive value of the forest. A trade-off exists between maximizing even-flow of harvests (within the planning horizon) and ensuring that the end productive value is maintained (ensures continued sustainability beyond the planning horizon). By ensuring that the productive value is maintained, the sustainability of obtaining future timber resources is secured.

Materials

A small forest holding will be used to demonstrate how the estimated optimality gap decreases as the number of scenarios increases. The forest holding is located in North Karelia, Finland, and is composed of 41 stands, with an area of 47.3 ha. The holding is composed mainly of Scots pine (*Pinus sylvestris* L.) with a minor component of both Norway spruce (*Picea abies* (L.) Karst.) and birch (*Betula pendula* and *Betula pubescens*). Key forest inventory information of the holding can be found in figure 1. The planning horizon is 30 years, with six 5-year periods. To calculate the productive value, a discount rate of 2% was used.

The set of scenarios were generated using a Monte Carlo process. To allow for a comparison between the amount of uncertainty contained in the set of scenarios two different processes were used. The first Monte Carlo process incorporated only the uncertainty caused by inventory measurements, while the

second process incorporated the uncertainty caused by both inventory measurements and growth modeling errors. The measurement errors were introduced into both the dominant height and basal area. The errors were assumed to be normally distributed, had no bias and a relative standard error of 20%, which reflects state-of-the art inventory methods (Næsset, 2004). Recent studies have shown only a small variable correlation, thus the measurement errors were considered uncorrelated (Haara and Korhonen, 2004; Mäkinen et al. 2010). In this case, we used a stand-level growth model for the sake of simplicity. The growth model errors were assumed to follow a first-order autoregressive process (AR(1)). For details on the exact process involved in simulating growth model errors, readers are referred to Phase II of the Materials and Methods section in Pietilä et al. (2010).

To highlight how the optimality gap improves as the number of scenarios used increases, the algorithm was run using a stopping criteria which relied on the number of iterations. For both cases N' was set at 1,000 and M was set at 500. For the case where only inventory errors were introduced into the model, the algorithm was run a total of 16 times, using an N starting at one, increasing each iteration by one until ten, then increasing by an interval of five until reaching an N of 40. For the case where both inventory errors and growth model errors were introduced into the model, the algorithm was run a total of 15 times, using an N starting at ten, increasing each iteration by ten until 100, then increasing by an interval of 20 until reaching an N of 200.

As a method to analyse the influence of risk preferences has on the optimality gap, the algorithm was run using two different weights for λ . While the weights (w_t^- , w_t^+ , w_E^- and w_E^+) should be set by the decision maker, for this example a surrogate was used to approximate the risk-neutral case. The weights were set to the shadow prices for the special case when the inventory errors and growth model errors are assumed to be zero, and when the weights are set to an initially high arbitrary value. This model reflects the standard deterministic model currently used in forestry. The use of shadow prices as

weights is justified, as they describe an underlying implicit utility model (Lappi 1992). The algorithm was run for the cases where $\lambda = 1.05$ (the nearly risk neutral case) and $\lambda = 4$ (the moderately risk averse case).

Results

The impact of increasing the number of scenarios on the optimality gap can be clearly seen in figures 2 and 3. The cases where only inventory errors were included in the analysis are shown in figure 2, while figure 3 shows the cases where both inventory errors and growth model errors were included. The figures show the probability that the solution created with N will be within a specified tolerance limit from the optimal value (eq. 11), with the 95% confidence interval as error bars (eq. 13).

Figures 2 and 3.

Both figures clearly show an increased proportion of small deviations in the optimality gap as the number of scenarios increase. This increase starts rather quickly for the case where only inventory error and a risk neutral perspective are taken. The number of scenarios required for the case where both inventory error and growth model error, and a risk-averse perspective is much greater if you want to have a very low optimality gap. This effect is seen in Table 1 through both the expected optimality gap ($\bar{f}_{N'}(\hat{x}) - \bar{v}_N^M$) and in the variance of the optimality gap (equation [10]). For both cases, as the number of scenarios increase, both the optimality gap and variance decrease quickly.

Table 1.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

Discussion

The results of the SAA method clearly highlight the trade-off between solution tractability and the optimality gap in the stochastic solution. For all cases, as the sample size increases the optimality gap steadily decreases. The sample size required to maintain a specified optimality gap between cases increases dramatically as both the quantity of uncertainty increases and as the importance of managing the downside risk increases. For both of these problems, a sample size of 200 scenarios can ensure an optimality gap of less than 2%.

An interesting aspect of the analysis highlights the need for an increased number of scenarios as importance of managing risk increases. This is clearly shown in figures 2 and 3 with a movement from $\lambda = 1.05$ (relatively risk neutral) to $\lambda = 4$ (relatively risk averse). To maintain a similar optimality gap, the number of scenarios required to approximate the problem needs to roughly increase ten-fold (Table 1). This requirement to increase the amount of scenarios to appropriately manage the risk features of the problem is very reasonable. For the risk neutral case, the negative deviations are not very important, and the emphasis is on maximizing the first period timber. For the risk-averse case, the negative deviations are rather important, and the set of scenarios to approximate the problem must include a more diverse set of scenarios. As risk management is a key differentiation between deterministic and stochastic programming, it is important to acknowledge the importance accounting for risk has on the required size scenario set.

The case study analysed in this paper is rather small, involving only 41 stands. While small, this problem can be considered representative of the typical forest holding by private individuals in Finland. The results of this study highlight the idea that the use of stochastic programming for small-scale

owners can be technically implemented into current forest management tools. While the use of stochastic programming is technically possible, the value of using stochastic programming needs to be presented to both the forest owners and the forest planners. This is perhaps a greater challenge than the technical challenges of integrating stochasticity to the management tools. It will require a shift from optimizing a forest holding in a deterministic setting, to optimizing a forest holding in a stochastic setting and accounting for the forest owners risk preferences towards risk.

For larger scale forest management problems, such as regional plans or a forest company level plan, computational power required to solve this problem can be immense. Thus, approaches which enable the use of large computational power may ease the problem. Alternatives such as parallel computing and using computational grids (Linderoth and Wright 2003) could be used to focus computational power onto the problem. Alternatively, stochastic decomposition methods (Higle and Sen 1996) could be used as an alternative to the SAA method, which focuses on limiting the size of the stochastic programming problem.

The development of large forest management plans also tend to focus on strategic level concepts, thus the problem formulation may be significantly different. Some of the large management plans may not be possible to solve using deterministic approaches, and they may require a hierarchical approach to planning to solve the problem (Kangas et al. 2014b). Thus, in a similar fashion, it could be possible to separate the stochastic programming problem into a hierarchical structure, which solves a set of relatively small problems.

The stochastic programming example used in this study utilized simple recourse. Simple recourse models evaluate the penalty associated with not achieving the specified goals. In this case, the penalty was the expected negative deviations from achieving an even-flow of harvested timber. More advanced

models could be developed which allow for the resolution of some of the uncertainty. For instance, a two-stage stochastic model could identify the optimal time to conduct the next inventory. This would allow for a resolution of the growth model uncertainties, but would depend upon the cost of conducting the next inventory, the interest rate and the risk preferences of the decision maker.

Once solved, the stochastic program provides a single implementable set of decisions for the forest holding. To allow for an adaptive management of the forest, a new plan could be generated. The frequency for the needs of new plans would depend on the demands of the decision maker. Even though the plan is for a longer time horizon, this does not imply that the plan needs to be followed for the entire horizon. In fact, the decision maker would be wise to update his/her forest plan when preferences change, or when some uncertainty has been resolved (i.e. after conducting management decisions, or after an updated inventory).

The results highlight the quick decrease in the optimality gap and variance as the number of scenarios increase. The decrease is more rapid for the case where there are fewer errors included, and when there is less of an importance in risk management in the problem formulation. Through the SAA method it is possible to highlight the drastic improvement in the solution compared to the deterministic solution.

The improvement from the deterministic case to the stochastic case can be viewed for the special cases where only one scenario was used to represent the distribution (when $\lambda = 1.05$). For these cases, the optimality gap and variance is rather large, and the decrease in these values highlights the improvement in the objective just by including the uncertainties into the analysis. So even for those cases where the optimality gap is large, the solution to the stochastic programming formulation is a drastic improvement from the deterministic equivalent.

For future research, there are numerous factors and considerations which should be analyzed. In this case, we examined the specific case where the error distributions were unbiased and were normally distributed. It would be interesting to examine if the improvement in the solution is as dramatic if the growth model errors were biased, or if the inventory errors were biased (e.g. overestimating BA in a sparse stand and underestimating BA in a dense stand). It is also possible to examine the influence of using different distributions to describe the errors. For instance, an analysis could be done to estimate the cost associated with using a normal distribution when the “true” distribution is something else (e.g. log-logistic distribution). Additionally, the impact of the use of different risk measures could also be more demanding (e.g. Conditional Value at Risk) and they may require larger scenario sets to produce solutions with an acceptable optimality gap.

Conclusions

Determining how many scenarios are required to deterministically approximate the stochastic forest management problem depends upon several key factors. These are the amount of uncertainty integrated into the problem, the risk preferences of the decision maker and the acceptability threshold of the decision maker for an optimality gap. For this study, only the inventory error and growth model errors were included in the analysis. Other uncertainties could also be included, such as the impacts of climate change, the potential of natural damage (e.g. storms, insects or fire) and timber price variations. Acquiring the risk preferences of the decision maker should not be restricted to lessen the need for additional scenarios. If technical constraints limit the quality solution, the impact should be on the optimality gap rather than on how the decision maker formulates the problem. Thus, while structuring and preparing the stochastic problem is much more intensive than the deterministic counterpart, it is important to remember that the decision will be more robust as it is based on additional relevant information.

1 **References**

- 2 Bevers, M. 2007. A chance constrained estimation approach to optimizing resource management under
3 uncertainty. *Can. J. For. Res.* 37(11): 2270–2280.
- 4 Birge, J.R., and Louveaux, F. 2011. Introduction to stochastic programming. Second edition. Springer,
5 New York. 499 p.
- 6 Boychuk, D., and Martell, D.L. 1996. A multistage stochastic programming model for sustainable
7 forest-level timber supply under risk of fire. *For. Sci.* 42(1): 10–26.
- 8 Diebold, F.X. 2001. Elements of Forecasting. Second Edition. South-Western, Thomson Learning.
9 Cincinnati, 392 p.
- 10 Eid, T., Gobakken, T., and Næsset, E. 2004. Comparing stand inventories for large areas based on
11 photo-interpretation and laser scanning by means of cost-plus-loss analyses. *Scand. J. For. Res.*
12 19(6):512-523.
- 13 Eriksson, L.O. 2006. Planning under uncertainty at the forest level: a systems approach. *Scand. J. For.*
14 *Res.* 21(S7): 111–117.
- 15 Eyvindson, K., and Kangas, A. 2014. Stochastic goal programming in forest planning. *Can. J. For. Res.*
16 44(10) 1274-1280.
- 17 Eyvindson, K., and Kangas, A. 2015. Integrating risk preferences in forest management planning. *Ann.*
18 *For. Sci.* Accepted manuscript.
- 19 Ferson, S., and Ginzburg, L.R. 1996. Different methods are needed to propagate ignorance and
20 variability. *Reliab. Eng. Syst. Safe.* 54, 133-144.

- 1 Gassmann, H.I. 1989. Optimal harvest of a forest in the presence of uncertainty. Can. J. For. Res.
2 19(10): 1267–1274.
- 3 Gregoire, T. G., and Valentine, H.T. 2008. *Sampling strategies for natural resources and the*
4 *environment*. CRC Press.
- 5 Gülpınar, N., Rustem, B., and Settergren, R. 2004. Simulation and optimization approaches to scenario
6 tree generation. J. Econ. Dyn. Control, **28**(7): 1291–1315.
- 7 Haara, A., and Korhonen, K. 2004. Kuvioittaisen arvioinnin luotettavuus. Metsätieteen aikakauskirja
8 4/2004:489-508 [in Finnish].
- 9 Haara, A., and Leskinen, P. 2009. The assessment of the uncertainty of updated stand-level inventory
10 data. Silva Fennica, 43(1): 87-112
- 11 Higle, J.L., and Sen S. 1996. Duality and statistical tests of optimality for two stage stochastic
12 programs. Math. Program. **67**(1-3) 143-168.
- 13 Hof, J., and Pickens, J. 1991. Chance-constrained and chance-maximizing mathematical Programs in
14 renewable resource management. For. Sci. **38**(1): 308-325.
- 15 Holmström, H., Kallur, H., Ståhl, G. 2003. Cost-plus-loss analyses of forest inventory strategies based
16 on kNN-assigned reference sample plot data. Silva Fennica, 37 (3):381-398.
- 17 Huuskonen. S., and Miina, J. 2007. Stand-level growth models for young Scots pine stands in Finland.
18 For. Ecol. Manage. **241**: 49–61.

- 1 Hynynen, J., Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H., and Haapala, P. 2002. Models for
2 predicting stand development in MELA system. Finnish Forest Research Institute Paper, 835
3 (116 pp.).
- 4 Kangas, A.S. 2010. Value of forest information. *Eur. J For. Res.* 129(5):863-874.
- 5 Kangas, A., Hartikainen, M., and Miettinen, K. 2014(a). Simultaneous optimization of harvest schedule
6 and measurement strategy. *Scand. J. For. Res.* 29:224-233.
- 7 Kangas, A., Hartikainen, M., and Miettinen, K. 2015. Simultaneous optimization of harvest schedule
8 and data quality. *Can. J. For. Res.* 45(8):1034-1044.
- 9 Kangas, A., Nurmi, M., Rasinmäki, J. 2014(b). From a strategic to a tactical forest management plan
10 using a hierarchic optimization approach. *Scand. J. For. Res.* 29:154-165.
- 11 Kangas, A., J. Kangas, J., and Kurttila, M. 2008. Decision support for forest management (Vol. 16).
12 Springer.
- 13 Kaut, M. and Wallace, S.W. 2007. Evaluation of scenario-generation methods for stochastic
14 programming. *Pacific Journal of Optimization* 3(2):257-271.
- 15 Kleywegt, A.J., Shapiro, A., and Homem-de-Mello, T. 2002. The sample average approximation for
16 stochastic discrete optimization. *Siam. J. Optim.* (12:2) 479-502.
- 17 King, A.J., and Wallace, S.W. 2012 *Modeling with Stochastic Programming*, Springer, New York
- 18 Krzemienowski, A., and Ogryczak W. 2005. On extending the LP computable risk measures to account
19 downside risk. *Comput. Optim. Appl.* **32**:133-160.

- 1 Lappi, J. 1992. JLP: A linear programming package for management planning. Research Papers 414,
2 The Finnish Forest Research Institute, 134 pp.
- 3 Lawrence, D.B. 1999. The economic value of information. Springer, New York, NY
- 4 Lemieux, C. 2009. Monte Carlo and quasi-Monte Carlo sampling. Vol. 20. Springer, New York.
- 5 Linderoth, J., and Wright, S. 2003. Decomposition algorithms for stochastic programming on a
6 computational grid. Computational Optimization and Applications, 24 (2-3), 207-250.
- 7 Lohmander, P. 2007. Adaptive optimization of forest management in a stochastic world. In Handbook
8 of operations research in natural resources. Edited by A.Weintraub, C. Romero, T. Bjorndal and
9 R. Epstein. Springer US. pp. 525-543.
- 10 Mak, W., Morton, D.P., and Wood, R.K. 1999. Monte Carlo bounding techniques for determining
11 solution quality in stochastic programs. Numerical Techniques for Stochastic Optimization 33-
12 64.
- 13 Mielikäinen, K. 1985. Koivusekoituksen vaikutus kuusikon rakenteeseen ja kehitykseen.
14 Communicationes Instituti Forestalis Fenniae 133. [In Finnish].
- 15 Mäkinen, A., Kangas, A., and Mehtätalo, L. 2010. Correlations, distributions and trends of inventory
16 errors and their effects on forest planning. Can. J. For. Res. 40(7): 1386–1396.
- 17 Næsset, E. 2004. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic
18 full-scale operational project. Scand. J. For. Res. 19(6): 554–557.
19 doi:10.1080/02827580410019544.

- 1 Packalén, P., and Maltamo, M. 2007. The k-MSN method in the prediction of species specific stand
2 attributes using airborne laser scanning and aerial photographs. *Remote Sens. Environ.* **109**(3):
3 328–341. doi:10.1016/j.rse.2007.01.005.
- 4 Palma, C. D., and J.D. Nelson. 2009. A robust optimization approach protected harvest scheduling
5 decisions against uncertainty. *Can. J. For. Res.* **39**(2), 342-355.
- 6 Pietilä, I., Kangas, A., Mäkinen, A., and Mehtätalo, L. 2010. Influence of growth prediction errors on
7 the expected losses from forest decisions. *Silva Fenn.* **44**(5): 829-843.
- 8 Pretzsch, H. 2010. *Forest dynamics, growth and yield: From measurement to model.* Springer-Verlag
9 Berlin Heidelberg.
- 10 Pukkala T (2005) Metsikön tuottoarvon ennustemallit kivennäismaan männiköille, kuusikoille ja
11 rauduskoivikoille (Prediction models for productive value of pine, spruce and birch stands in mineral
12 soils). *Metsätieteen aikakauskirja* 3/2005: 311-322 [in Finnish]
- 13 Rockafellar, R. T., and Uryasev, S. 2000. Optimization of conditional value-at-risk. *J. Risk*, **2**: 21-42.
- 14 Shapiro, A. 2003. Monte Carlo sampling methods. In *Stochastic programming. Handbooks in*
15 *Operations Research and Management Science*, Vol. 10. Edited by A. Ruszczyński and A.
16 Shapiro. Elsevier, Amsterdam. pp. 353–425.
- 17 Vuokila, Y., and Väliäho, H. 1980. Viljeltyjen havumetsiköiden kasvatusmallit.
18 *Metsäntutkimuslaitoksen julkaisu* 99(2). [In Finnish].
- 19 Williams, B. K. 2011. Adaptive management of natural resources—framework and issues. *J. Environ.*
20 *Managem.* **92**(5), 1346-1353.

Table 1. The estimates of the optimality gap and variance for the two different risk preferences analysed and for the two different combinations of error.

$\lambda = 1.05$					$\lambda = 4$				
Scenarios	Inventory Errors only		Inventory and Growth Model Errors		Scenarios	Inventory Errors only		Inventory and Growth Model Errors	
	Gap	Variance	Gap	Variance		Gap	Variance	Gap	Variance
1	4.9%	0.7%	4.8%	0.6%	10	12.3%	0.1%	16.5%	0.1%
2	3.8%	0.3%	3.7%	0.3%	20	7.6%	0.0%	10.2%	0.0%
3	3.2%	0.2%	3.1%	0.2%	30	5.0%	0.0%	7.3%	0.0%
4	2.6%	0.1%	2.5%	0.1%	40	3.7%	0.0%	5.7%	0.0%
5	2.1%	0.1%	2.1%	0.1%	50	2.9%	0.0%	4.6%	0.0%
6	2.0%	0.1%	1.9%	0.1%	60	2.4%	0.0%	3.9%	0.0%
7	1.9%	0.1%	1.8%	0.1%	70	2.0%	0.0%	3.3%	0.0%
8	1.6%	0.1%	1.6%	0.1%	80	1.8%	0.0%	3.0%	0.0%
9	1.4%	0.1%	1.3%	0.1%	90	1.6%	0.0%	2.7%	0.0%
10	1.2%	0.1%	1.2%	0.1%	100	1.4%	0.0%	2.4%	0.0%
15	1.0%	0.0%	0.9%	0.0%	120	1.1%	0.0%	2.1%	0.0%
20	0.8%	0.0%	0.8%	0.0%	140	1.0%	0.0%	1.8%	0.0%
25	0.6%	0.0%	0.6%	0.0%	160	0.9%	0.0%	1.6%	0.0%
30	0.5%	0.0%	0.5%	0.0%	180	0.8%	0.0%	1.4%	0.0%
35	0.5%	0.0%	0.5%	0.0%	200	0.7%	0.0%	1.3%	0.0%
40	0.5%	0.0%	0.4%	0.0%

Figure 1. (a) age-class distribution of the holding at the beginning of the planning period; and (b) diameter distribution at 1.3 m height at the beginning of the planning period.

Figure 2. The probability of the solution $\hat{x}_N$ being within a tolerance limit of the estimated optimal solution, for the case where only inventory errors were included in the analysis. The risk-neutral ($\lambda=1.05$) case is presented, with only inventory errors considered on the left, and both growth model and inventory errors considered on the right.

1

2 **Figure 3.** The probability of the solution $\hat{x}_N$ being within a tolerance limit of the estimated optimal solution, for the case where both
3 the inventory errors and growth model errors were included in the analysis. The risk-averse ($\lambda=4$) case is presented, with only
4 inventory errors considered on the left, and both growth model and inventory errors considered on the right.