Author(s): Eyvindson, Kyle; Kangas, Annika

Title: Comment on "A simple way to incorporate uncertainty and risk into forest harvest scheduling"

Year: 2017

Version:

Comment on "A simple way to incorporate uncertainty and risk into forest harvest scheduling"

Kyle Eyvindson*, Annika Kangas

[1] University of Helsinki, Department of Forest Sciences, Latokartanonkaari 7 (P.O. Box 27) 00014 University of Helsinki, Finland, kyle.eyvindson@helsinki.fi

[2] Economics and Society, Natural Resources Institute Finland (Luke), P.O. Box 68, 80101 Joensuu, Finland

Highlights:

Managing risk requires an understanding of the preferences of the decision maker

Minimizing downside risk measures can be more preferentially appropriate

It is important understand how minimizing the risk impacts the forest management

Risk management is accomplished through some form of trade-off

Abstract:

In a recent research article, Robinson et al. (2016. For. Ecol Manage **359**:11-18) described a method of estimating uncertainty of harvesting outcomes by analyzing the historical yield to the associated prediction for a large number of harvest operations. We agree with this analysis, and consider it a useful tool to integrate estimates of uncertainty into the optimization process. The authors attempt to manage the risk using two different methods, based on deterministic integer linear programming. The first method focused on maximizing the 10th quantile of the distribution of predicted volume subject to area constraint, while the second method focused on minimizing the variation of total quantity of volume harvested subject to a harvest constraint. The authors suggest that minimizing the total variation of the harvest could be a useful tool to manage risk. Managing risks requires trade-offs, however, typically less risk involves higher costs. The authors only superficially stated the costs and did not consider if these costs are reasonable for the management of risk. In this comment, we specifically develop the models used in their article, and demonstrate a method of managing the downside risk by utilizing the Conditional Value at Risk.

# 1.	Introduction:

Being able to incorporate uncertainty, and the management of risks associated with harvesting decisions are important questions for both small scale and large scale forest owners (Pasalodos-Tato et al. 2013). This topic has been recently explored in a research article published by this journal (Robinson et al. 2016) (hereafter RMM), which demonstrated a method of incorporating uncertainty through utilizing historical information about harvesting quantities, linking to the predicted values. The authors then demonstrated a method that minimized the variance while achieving a specified harvest target. While the paper demonstrated a useful method of integrating uncertainty through empirical data, the method which they strove to minimize the risk (measured as variance) did not assess the required trade-off to manage the risk.

In order to manage risk, it is important to really understand the meaning behind what risk is. In an application to manage the risk of an investment portfolio, Markowitz (1952) defined risk as the variation of the distribution and that a risk-averse individual would want to minimize the variation with a specific decision. More recently, the ISO has defined risk in a slightly more general fashion as the "effect of uncertainty on objectives" (ISO Guide 73:2009). For both of these definitions, both unexpected losses and gains are treated as undesired.

Other definitions of risk can be more context specific, such as downside risk measures. Some examples are the downside mean semideviation (Krzemienowski and Ogryczak 2005), the Value at Risk (VaR: a given quantile of the distribution, Duffie and Pan 1997) and the Conditional Value at Risk (CVaR: expected shortfall, Rockafellar and Uryasev 2000) (Figure 1). The key feature of downside risk measures is that only losses below the target set are of importance. Even though RMM have identified a context specific preference towards risk (specifically "the loss

associated with under-prediction is held to be less than the loss associated with over-prediction"), the authors have decided to ignore this preferential information, and utilize the more general definition of risk, rather than utilize a downside risk measure. An example of managing downside losses for a multi-period forest harvest scheduling problem is found in Eyvindson and Kangas (2015b).

Figure 1.

By focusing solely on the minimization of the variation of the total harvest quantity, RMM did not analyze the increasing costs required to minimize the variation. By focusing on minimizing the variation, the authors seemed to be willing to accept a cost of doubling the area harvested (from 997 ha to 1983 ha) for a modest improvement in the variation. The implicit meaning behind such a large increase of harvested area is that lower volume and possible less productive stands will be harvested. This is a very large cost. To aid in the decision making process, those individuals tasked with making the decision may wish to see the trade-off between increased harvesting area and decreased harvesting variation.

As the authors have identified, managing risk requires a balance of trade-offs. For the investment portfolio problem of Markowitz (1952), this trade-off occurs between the expected mean return of the investment, and the variation of the expected return of the investment. A choice must be made by the investor, a risk-averse investor would like to minimize the variation, while a risk-seeking investor would solely be interested in maximizing the mean return. For the case described by RMM, the trade-off occurs between the variance of the total volume harvested at a

specific target and the total area harvested. To represent the trade-off more accurately, an analysis could be done as the trade-off between the total harvest variation and the total cost of conducting the harvest.

In this comment, we clarify the trade-offs that are occurring as the decision maker shifts from being risk-averse to risk-seeking. Additionally we suggest an alternative approach to managing the risk associated with determining the appropriate harvest schedule. Rather than minimizing variation, the objective should be to minimize the total area harvested which provides a specific target volume, while ensuring that a specific downside risk measure is met. This is done by including the CVaR, a coherent risk measure (Artzner et al. 1999; Krokhmal et al. 2011), as a constraint to the optimization model. In this way, the key constraint is placed in the objective function, and the associated downside risk is set as a constraint. We believe that this kind of optimization process better reflects the intent of area constraint, and focuses harvest on only a limited forest area.

2. Materials and Methods

2.1 Methods

To ensure clarity for the models, we will mathematically describe the models used by RMM (both based on the written description and the supplementary R file). From there, we will emphasize the (probably unintended and most likely unwanted) effects of the optimization models on the forest structure and harvesting costs. We then suggest an alternative way to achieving a specific target with a specified level of risk. While we are not utilizing the same data used by RMM, we are using the simulated data which they suggest is realistic, and we expect is a reflection of the real dataset that they used. As the result that we obtained from the simulated

data has the same properties as described in the article, we believe that an appropriate comparison can be made.

The first model used by RMM was a simple integer linear programming model, which maximized the expected harvest, subject to an area constraint (and in the R code a constraint limiting the number of stands harvested). We identify this model as Model 1:

[1] $$\max \sum_{j \in N} \mathbb{E}(c_j) x_j$$

      subject to:

[2] $$\sum_{j \in N} a_j x_j \leq t_a$$

[3] $x_j \in \{0,1\}, \forall j \in N$

where $\mathbb{E}(c_j)$ is the expected timber harvested for stand j, N is the set of stands under consideration, x_j is the decision variable defining if stand j is to be harvested or not, a_j is the area of stand j, and t_a is the limit to the area harvested . This is a simple integer linear programming model, where the objective is to maximize the quantity of timber harvested, subject to: an area constraint ([2]). A note: in the R model provided as a supplement to RMM, there is a constraint limiting the total number of stands harvested ([4] where t_s is the limit to the stands harvested), while in the paper there is no mention of this constraint. For this comment, we will not include an additional constraint limiting the total number of stands harvested.

[4] $$\sum_{j \in N} x_j \leq t_s$$

The second model presented by RMM is one which maximizes the 10th quantile of the timber harvest predictions, rather than the expected timber harvest. The only change is in the objective function. We identify this model as Model 2:

[5] $$\mathbf{max} \sum_{j \in N} c_j^q x_j$$

subject to [2] and [3], where c_j^q is the qth (in this case $q = 10$) quantile of the distribution of predicted timber harvested for stand j.

The third model is one which minimizes the variation of the total timber harvested, subject to a targeted harvest constraint. We identify this model as Model 3:

[6] $$\mathbf{min} \sum_{j \in N} v_j x_j$$

subject to:

[3]

[7] $$\sum_{j \in N} \mathbb{E}(c_j) x_j \geq b$$

where v_j is the variation of the timber harvested from stand j, and b is the target for the total harvested volume. From the description provided by Robinson et al. (2016) they indicate that the

area constraint was replaced by the volume constraint. However, in the R script from the supplementary material, this model also has an area constraint ([2]). In the R code the maximum area was set to 2,000 ha, which can explain the reason why this model has nearly 2x the area harvested than the models one and two of RMM.

While this model does manage the risk (positive and negative gains), it does so indiscriminately. The authors of this paper have identified that decision makers are probably more averse to losses associated with under prediction, so this should be the focus of the model. As a demonstration of how the negative losses can be managed, we developed a simple stochastic programming model which can ensure that the CVaR does not exceed a specified limit. In comparison to model 3, this model aims at minimizing the total area harvested while ensuring that the CVaR relating to a specific harvest target is met. The stochastic model, identified as Model 4, is as follows:

[8] $$\mathbf{min} \sum_{j \in N} a_j x_j$$

subject to:

[9] $$CVaR_\alpha \leq g$$

[10] $$CVaR_\alpha = \left(Z + \frac{\mathbf{1}}{(\mathbf{1} - \alpha)|I|} \sum_{i \in I} [L_i - Z]^+ \right)$$

[11] $$L_i = \left[\sum_{j=1}^{J} c_{ij} x_j - b \right]^+, \forall i \in I$$

and [3], where $CVaR_\alpha$ is the Conditional Value at Risk at the α protection level, g is the target level for the $CVaR_\alpha$, c_{ij} is the timber harvested according to scenario i and schedule j, Z is the Value at Risk, I is the set of scenarios (a sample of values from the distribution of predicted harvested volumes), L_i is the negative loss associated with scenario i. The square brackets indicate that only positive values will be returned, if the value inside the brackets is negative, a zero is returned.

2.2   Materials:

As we did not have access to the actual data, we have used the simulated data produced by the method found in the supplementary materials in RMM. This simulated data represents a forest holding which is 34,858 ha in size, and has a high proportion of stands which have a small amount of timber (Table 1).

To allow for a comparison between the simulated data set, and the results from the real data, we need to run the simulated data through all of the different models. For ease of comparison, none of the models included a constraint limiting the number of stands harvested. The models were run sequentially, as a volume target is required for models 3 and 4. To understand the influence of adding an area constraint, three versions of model 3 were run: model 3a, b, and c have an area constraint of 1,000 ha, 2,000 ha, and 34,858 ha respectively. Additionally, for model 4, preferential information was required for managing the CVaR. For this case, we arbitrarily selected a value of 5,000 m3 for the CVaR with the confidence level set at 5% ($\alpha = 0.05$). To allow for readers to explore the models, we have added a supplementary R code which builds upon the work done by RMM. All of the models have been solved by R, with the exception of model 4. The size of the problem is too large for R to solve quickly, so a commercial alternative

was used (CPLEX). The R code provides the data, a set of 1,000 scenarios, and the associated model can be found in the supplementary materials. For this case, a scenario represents a fixed set of harvest volumes for each stand. For one scenario, each stand is assigned a specific volume based on the mean and variance for the specific stand. As the scenario based stochastic program is only an approximation of the true problem, it is wise to evaluate an estimate of the gap. To do this, we used the stochastic average approximation method (Kleywegt et al. 2002), for an application in forestry, readers are referred to Eyvindson and Kangas (2015a). With a set of 1,000 scenarios the estimated gap in the CVaR constraint is slightly greater than 200 m3, which we believe indicates that the set of scenarios is sufficient to approximate the stochastic problem.

To conduct a trade-off analysis between cost and risk, we have made a few assumptions regarding the costs of conducting a harvest. Harvesting cost can be separated into two parts, a fixed cost and a variable cost. The fixed cost relates to the transportation of equipment to a specific site to conduct the harvesting, and the variable cost relates to the quantity of timber harvested at the particular site. For this trade-off analysis, we have decided to set the fixed cost to 100 units (i.e. €or $) and the variable cost is related to the density of the stand to be harvested. We will use costs as indicated in table 1 for different densities of volume. Additionally, for both models we solved a series of problems, by making adjustments to the constraints. For model 3, we adjust the area constraint ([2]) and for model 4, we adjust the CVaR constraint ([9]).

Table 1.

3. Results:

To compare what actually happens in the forest with these different models, we analyze the proportion of stands harvested in different volume classes. The harvesting areas for the different volume classes are found in table 2.

In a similar fashion as RMM, we also ran model 3 with a range of volume constraints, however the range of volume constraints included values near the maximum harvested volume for model 1. To understand the influence of the area constraint, we used this approach with all three versions of model 3. The results for model 3b and 3c are reflective of the results obtained by RMM, however the results for model 3a highlight the importance of area constraints (Table 3). The values described in Table 3 are not the results from the optimization process, as the standard deviation and coefficient of variation cannot be evaluated through a deterministic optimization model. These values were generated by evaluating the decision over a 10,000 representations of the possible harvesting possibilities of each stand.

The results from the trade-off analysis can be seen in Figure 2. While the models are not entirely comparable, as the focus of both models are different, the trade-off analysis highlights the different approaches in managing risk. A series of three variables are compared against the total area harvested, while the last variable reflects the m^3 cost of using the different approaches.

3.   Discussion:

There are two ways to cope with uncertainty when a decision needs to be made: a decision maker can make a choice with the current information available, or the decision maker can reduce the uncertainty by collecting additional relevant information. Minimizing the variance of harvested volume can provide a less variable amount of timber, however the reduced variation does require some form of trade-off. In a forest management context, the trade-off can be associated with the

area harvested, or with the costs of harvesting. Without considering the trade-offs, risk management can be very costly.

There is always a tradeoff between mitigating the risk involved in decision making and the objective. In the RMM, the cost of reducing uncertainty in the harvested volume was quite steep, an increase of harvested area from 997 to 1983 hectares, which is most likely not acceptable to decision makers. Such solution would be poor both from silvicultural and wood procurement perspectives. It would make more sense to make the trade-off between uncertainty and expected harvest explicit, and let the decision maker to decide how much he is willing to pay for reducing the uncertainty. Moreover, in RMM the risk is defined as the variation of harvested volume, so that also deviations to positive side are also avoided. For most decision makers the approach ensuring a specific downside risk would be more appropriate.

If we look closely at what is happening in the different models, it is possible to understand how preference information can be implicitly incorporated into the different models. As RMM have noted, there is not much of a change between model 1 and model 2. Depending on the area constraints, there can be a significant shift in harvesting for model 3. If the area constraint is equal to model 1 and 2, then there is no change. This can be expected, as obtaining the maximum volume harvested from a limited area has a very limited decision space. When the area constraint is relaxed, a much larger proportion of low volume stands are harvested. So the reduction in total harvested volume variation requires a substantial harvesting in very low volume stands. In order to meet the CVaR constraint, some additional harvests are conducted in model 4. While the total variance is not minimized with the CVaR approach, the downside risk is restricted according to the preferences of the decision maker.

In Table 2, it is possible to compare the the CVaR, the total area harvest and the area harvested according to the density of the stands. For models 1, 2, 3a and 4, the total area harvested and the distribution of the density of the stands harvested are all rather similar. In model 4, the primary difference is that there is some additional area harvested, focused on the 100-200 m^3/ha cohort. In this way, it is rather clear that managing for downside risk reflects the original optimization problem (i.e. model 1).

RMM were also rather interested in the influence of using model 3 on the coefficent of variation, and were of the impression that the CV decreases as volume constraint increases. In table 3 we examine this property, and our results suggest that the CV is dependent on how the optimization model is formulated. For model 3a, the area constraint is rather tight and the CV increases as the volume constraint increases. For model 3b, the area constraint is loosened considerably, and the CV is rather constant as the volume constraints increase. For model 3c, we find similar results as RMM, that with essentially no area constraints, the CV decreases as volume constraints increase. These results are to be expected, as there is a noticeable trade-off between variation and increased total harvested volume. This trade-off is not as noticeable with model 3b or 3c, as there are possibilities to harvest larger quantities of timber with only a minor increase in variation.

Risk management requires some form of trade-off, this occurs in both model 3 developed by RMM and model 4 developed here. This can be analyzed visually in Figure 2. As expected, for model 3 the variance decreases with an increased area harvested, while for model 4 there is a slight increase in the variance. The CVaR for model 3 decreases slightly as the area harvested increases, but remains above 13,000 m^3. For model 4, there is a steady increase in CVaR, as the total harvested area decreases. So, as the decision maker becomes more risk seeking, the harvested area decreases. The third variable describes how risk is accounted for with according

to model 4. The volume harvested is essentially constant for model 3, however steadily increases for model 4. To ensure a CVaR above a specific threshold, the total volume harvested increases. The last variable highlights the cost associated with the approaches. For both models, the cost increases as the appetite for risk decreases. The per-unit cost is substantially lower for model 4 and has a much smaller range (7.5 – 8) than model 3. In this analysis, we defined the costs of variables in a way that reflects current operations (pers. comm. V.P. Kivinen 2016). The key assumption we make is that it is more costly (per m^3) to harvest sites with lower volume density.

The key to manage risk is to appropriately define the risk preferences of the decision maker. If the sole consideration of the decision maker is to minimize the variance, with a specific increase in the area harvested, then model 3 could be appropriate. However, if the decision maker wishes to ensure that a specific threshold of timber volume is harvested according to a specific risk level, the model 4 could be more appropriate. To manage this risk, model 4 increases the expected harvest level, and this increases the probability of meeting the target volume. By minimizing the area harvested, costs are simultaneously minimized.

5.   Conclusion

Being able to manage risk requires an understanding of the decision maker's preferences towards the management of the forest. The model proposed by RMM suggested that the key aspect of was to minimize the variation of the total harvested timber. While this seemingly makes intuitive sense, we need to also really consider the preferential interpretations of the model. In linear programming models, the constraints are key preferential information and are of more importance than those variables in the objective function. Thus, when integrating risk into the problem formulation, all adjustments need to be carefully thought through, and the impacts on

14

the management of the forest needs to be considered. Large changes in the management practices (i.e. a doubling in the area harvested) should be questioned, and should be analyzed if the adjustments are justifiable and reasonable to a decision maker.

References:

Artzner, P., Delbaen, F., Eber, J.M., and Heath, D. 1999. Coherent measures of risk, Mathematical Finance 9(3) 203–228.

Duffie, D., and Pan, J. 1997. An overview of value at risk. The Journal of Derivatives 4(3), 7-49.

ISO Guide 73. 2009. Risk Management Terminology. Geneva, Switzerland: International Organization for Standardization (ISO).

Eyvindson, K and Kangas, A. 2015a. Evaluating the required scenario set size for stochastic programming in forest management planning: incorporating inventory and growth model uncertainty. Canadian Journal of Forest Research. In Press.

Eyvindson, K and Kangas, A. 2015b. Integrating risk preferences in forest harvest scheduling. Annals of Forest Science. In Press.

Kleywegt, A.J., Shapiro, A., and Homem-de-Mello, T. 2002. The sample average approximation for stochastic discrete optimization. Siam. J. Optim. (12:2) 479-502.

Krokhmal, P., Zabarankin, M., and Uryasev, S. 2011. Modeling and optimization of risk. Surveys in Operations Research and Management Science, 16(2), 49-66.

Krzemienowski, A., and Ogryczak, W. 2005. On extending the LP computable risk measures to account downside risk. Comput. Optim. Appl. 32:133-160.

Markowitz, H. 1952. Portfolio Selection. The Journal of Finance, 7(1): 77-91.

Pasalodos-Tato, M., Mäkinen, A., Garcia-Gonzalo, J., Borges, J.G., Lämås, T., and Eriksson, L.O.2013. Review. Assessing uncertainty and risk in forest planning and decision support

systems: review of classical methods and introduction of new approaches. Forest Systems, 22(2), 282-303.

Robinson, A.P., McLarin, M. and Moss, I. 2016. A simple way to incorporate uncertainty and risk into forest harvest scheduling. Forest Ecology and Management. 359:11-18.

Rockafellar, R. T., and Uryasev, S. 2000. Optimization of conditional value-at-risk. J. Risk, 2: 21-42.

Table 1. The variable costs of harvesting.

Quantity of timber on stand (m^3/ha)	Unit cost per m^3
0-100	10
100-200	8
200-350	7
350+	6

Table 2. Harvesting of different volume cohorts, according to the different models.

	Unit	Model 1	Model 2	Model 3a	Model 3b	Model 3c	Model 4	AREA of cohort
Volume Target	(m^3)	-	-	213,360	213,360	213,360	213,360	-
Area Constraint	(ha)	1,000	1,000	1,000	2,000	34,770	-	-
0-50 m^3/ha	(ha)	0.0	0.0	0.0	384.1	10,215.9	0.0	30,156.6
50-100 m^3/ha	(ha)	0.0	0.0	0.0	1,021.2	551.2	22.1	3,533.5
100-200 m^3/ha	(ha)	708.8	710.9	708.8	462.4	90.4	852.3	879.1
200-300 m^3/ha	(ha)	160.3	158.1	160.3	60.5	25.4	160.3	160.3
300-400 m^3/ha	(ha)	102.7	102.7	102.7	47.0	0.0	102.7	102.7
400-500 m^3/ha	(ha)	12.7	12.7	12.7	12.7	0.0	12.7	12.7
>500 m^3/ha	(ha)	13.6	13.6	13.6	10.0	5.5	13.6	13.6
CVaR*	(m^3)	22,332	23,377	22,106	15,396	11,342	4,323	-
TOTAL	(ha)	998.2	998.0	998.2	1,997.9	10,888.4	1,163.8	34,858.6

* Note that the CVaR is re-evaluated in R using a larger set of scenarios, so for model 4, the CVaR will differ slightly than the expectation of nearly 5,000 m^3

Table 3. Results of variance minimization utilizing three different constraints for total area harvested. These values are an evaluated of the set of decisions by a simulation of 10,000 possible scenarios.

| Volume target (thousand m³)	Model 3a				Model 3b				Model 3c			
	Mean (thousand m³)	Standard Deviation of volume (thousand m³)	Coefficient of Variation	Area harvested (ha)	Mean (thousand m³)	Standard Deviation of volume (thousand m³)	Coefficient of Variation	Area harvested (ha)	Mean (thousand m³)	Standard Deviation of volume (thousand m³)	Coefficient of Variation	Area harvested (ha)
110	110.03	6.04	0.055	998.36	110.06	4.94	0.045	1999.81	109.98	4.23	0.038	7297.04
115	114.99	6.16	0.054	998.24	114.91	5.21	0.045	1997.84	115.02	4.40	0.038	7428.94
120	120.11	6.53	0.054	999.82	119.98	5.46	0.045	1999.16	120.13	4.51	0.038	7796.49
125	125.03	6.89	0.055	998.79	125.02	5.65	0.045	1999.47	124.96	4.66	0.037	8194.74
130	130.13	7.16	0.055	996.41	130.05	5.86	0.045	1999.47	129.91	4.82	0.037	8166.59
135	135.08	7.42	0.055	999.80	135.09	6.04	0.045	1998.91	134.93	4.94	0.037	8297.70
140	140.01	7.60	0.054	999.17	140.00	6.13	0.044	1999.67	139.95	5.07	0.036	8698.66
145	145.24	7.94	0.055	999.08	145.06	6.47	0.045	1999.39	145.00	5.23	0.036	9196.58
150	149.91	8.33	0.056	999.01	150.14	6.52	0.043	1998.89	150.10	5.30	0.035	9081.56
155	155.13	8.65	0.056	997.77	155.08	6.86	0.044	1999.57	154.99	5.49	0.035	9607.76
160	160.07	8.86	0.055	998.62	160.10	6.98	0.044	1997.67	160.03	5.60	0.035	9910.31
165	165.21	9.24	0.056	999.87	165.05	7.22	0.044	1998.31	164.89	5.74	0.035	9740.66
170	170.17	9.58	0.056	998.76	170.09	7.43	0.044	1999.19	170.06	5.78	0.034	9976.37
175	175.16	9.79	0.056	997.36	175.09	7.66	0.044	1999.95	174.94	6.08	0.035	10215.58
180	180.11	10.26	0.057	998.65	180.09	7.91	0.044	1997.25	180.09	6.07	0.034	10424.93
185	184.98	10.47	0.057	999.98	185.00	8.17	0.044	1997.62	185.05	6.19	0.033	10192.84
190	190.20	11.03	0.058	998.96	189.95	8.32	0.044	1998.72	190.05	6.38	0.034	10299.68
195	195.00	11.46	0.059	999.82	194.94	8.65	0.044	1998.31	195.10	6.48	0.033	10718.40
200	200.06	12.32	0.062	997.91	200.12	8.83	0.044	1997.55	200.01	6.54	0.033	10723.33
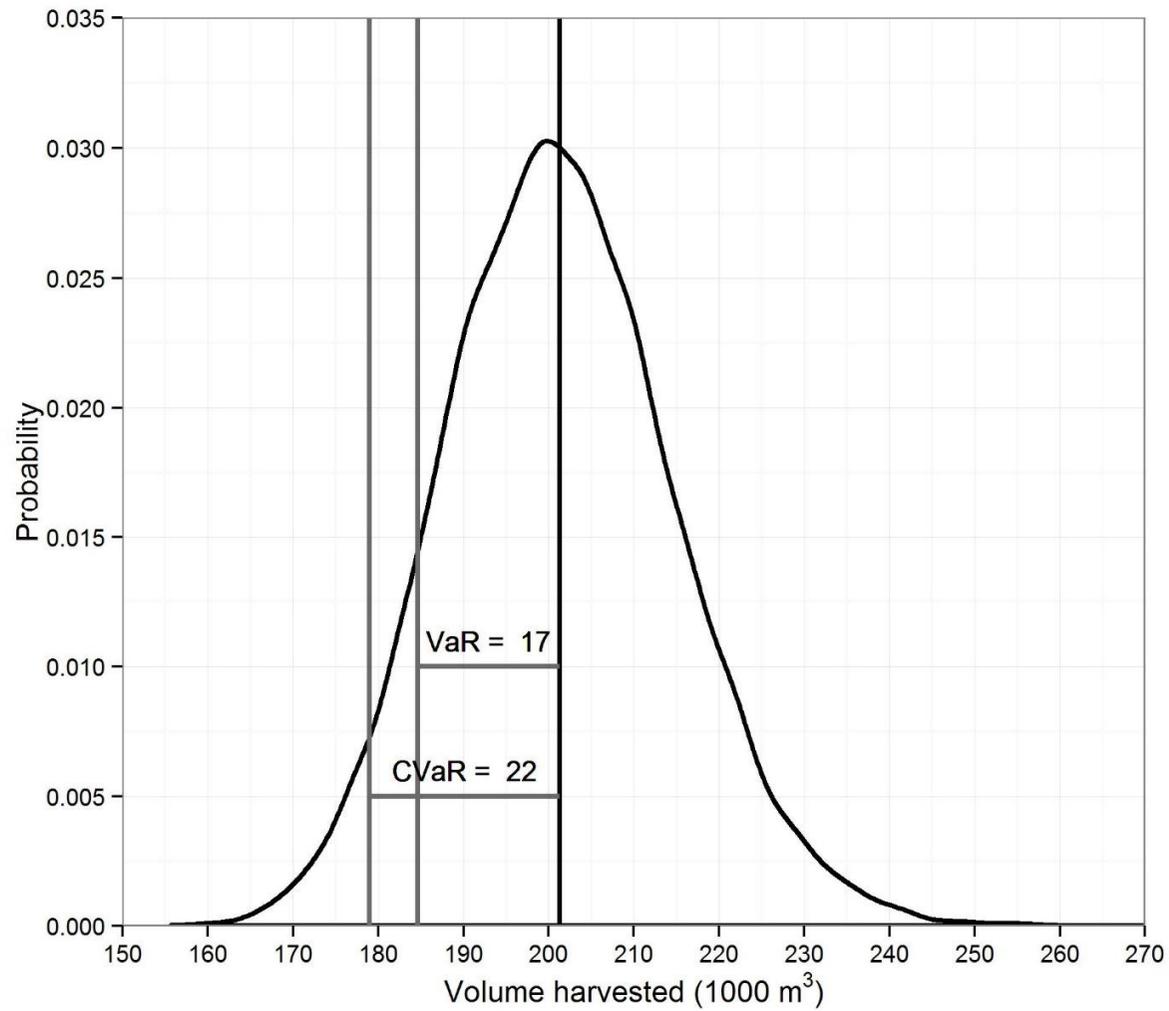
Figure 1. A representation of the possible harvest distribution for the simulated forest. The black vertical line is the expected harvest (which is used as the target for evaluating the downside risk measures), while the VaR and CVaR when α=0.1 are identified for losses exceeding 5,000 m^3 from the mean result are in grey.

Model 3   Model 4

1
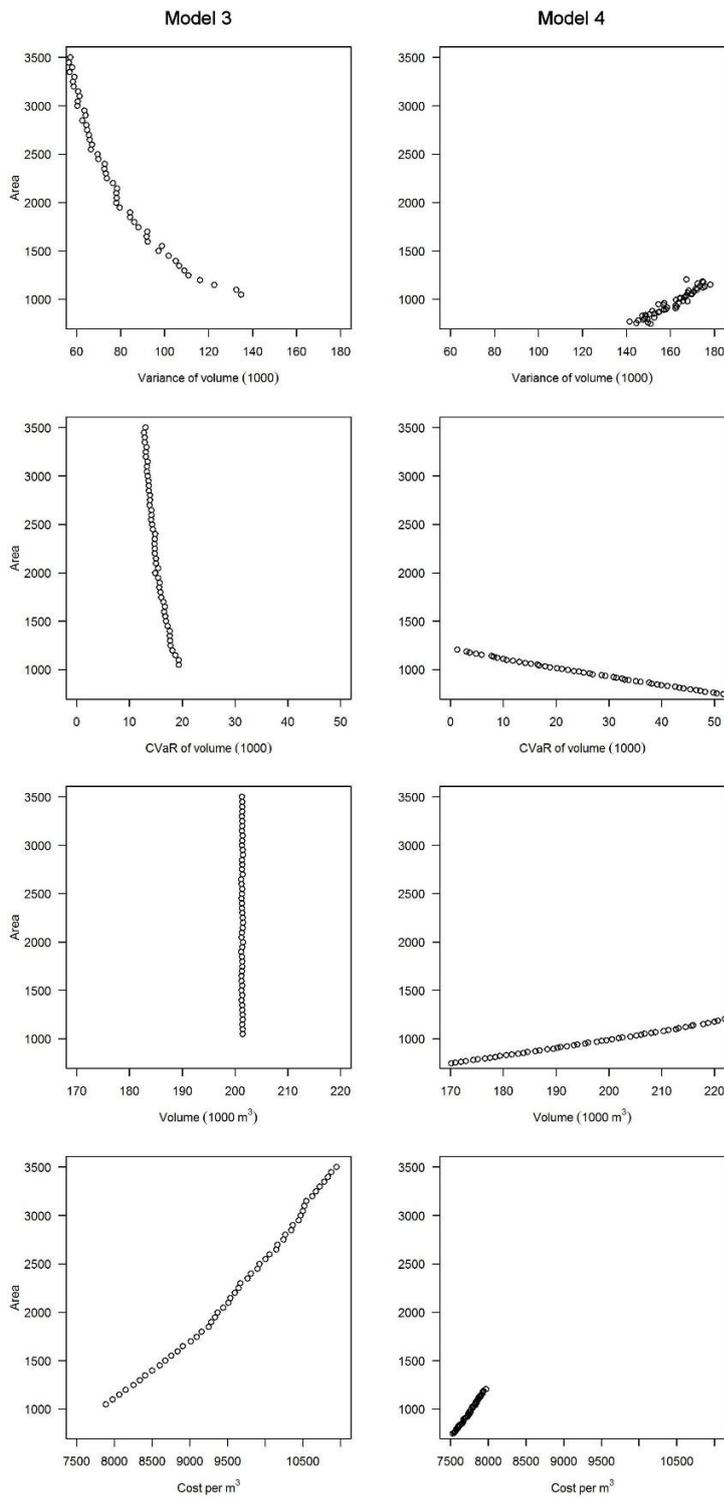
Figure 2. The trade-off analysis of the two models. Model 3 is on the left and Model 4 is on the right. The
areas harvested according to the different models have different ranges, Model 3 has a range (1000 –
3500 ha) while Model 4 has a range (750- 1200 ha)