Value of Information Analysis of Snow Measurements for the Scheduling of Hydropower Production

Heidi Liljeblad Ødegård[1], Jo Eidsvik[1] and Stein-Erik Fleten[2]
1) Department of Mathematical Sciences, NTNU, Norway (jo.eidsvik@ntnu.no)
2) Department of Industrial Economics and Technology Management, NTNU, Norway

Abstract

The scheduling of a hydropower plant is challenging because of inflow uncertainty. During spring there is increased uncertainty when the snow melts. By gathering snow measurements, one learns more about the future inflow, and this might lead to lower spillage risk or higher efficiency. In this paper the value of information of snow measurements is studied. The value of information is representative of how much a test is worth. If the price of acquiring and processing snow measurements is less than the value of information, the test is worth doing. The notion of value of information is also useful for comparing various kinds of snow measurements in different situations.

For scheduling a least squares Monte Carlo method is used in this paper. The uncertain inflow is represented by discrete scenarios, while the time-varying spot price is assumed known. Data from a Norwegian power plant are used to fit the inflow and snow distributions as well as prices, water reservoir limits and production release alternatives. The numerical tests show that snow measurements have little value when the reservoir is large compared to the total inflow. When the reservoir is smaller, the probability of overflow is bigger, and the snow measurements can be valuable for the scheduling when the data have high accuracy. The increase in value by using the snow measurements is between 0 and 10 % in the different parameter settings considered here.

1 Introduction

The scheduling of hydropower plants is important for reliable energy production. The goal is to time releases of water from the reservoirs so as to maximize expected value, but uncertainties in the inflow of water complicates the scheduling procedures. There has been a lot of research on optimal scheduling of hydropower plants, and the models have become more complex and adapted to each power plant or systems of power plants [21]. Most approaches use some kind of dynamic programming [15]. In this paper we

use an approximation method called least squares Monte Carlo (LSMC) to find reasonable scheduling plans [4] for our case. This approach handles uncertainty by representing inflow as discrete scenarios.

The scheduling relies on all available information. For hydroelectric scheduling [10], [20] and [5] show that obtaining more complete hydrology information has merit when operators emphasize water shortage or flooding risk. In addition to the common hydrological information one can in some situations gather snow measurements ([12]; [14]; [13]). Such snow measurements are then assimilated to update the model for inflow, and this can improve the scheduling. The snow measurements of course come with a price, and one might ask when they have significant value?

The focus of this paper is on value of information (VOI) analysis to guide the acquisition and processing of snow measurements. The VOI is an information criterion that can represent monetary units. If the VOI exceeds the price of snow measurement acquisition and processing, the data are worthwhile gathering. The VOI is also useful for comparing different kinds of tests [6]. Computational challenges associated with VOI analysis are not straightforward for complex decision situations such as hydropower scheduling, and we approximate the VOI building on the LSMC approach.

In Sect. 2 we describe the situation with hydropower scheduling and discuss how one can use snow measurements in this context. In Sect. 3 we present the LSMC approach for scheduling under uncertain inflows. In Sect. 4 we define the VOI and outline a method for VOI analysis of snow measurements for hydropower scheduling. In Sect. 5 we present results on a case from a Norwegian reservoir, where the inflow consists of both rainwater and melted snow from the catchment area. Sect. 6 provides closing remarks.

2 Background and notation

Hydropower companies strive to deliver the maximum value of energy over a long-term horizon, with various constraints on the water level and the production equipment. There are of course multiple challenges related to this goal. We focus on optimal scheduling of a single hydropower plant, when the main uncertainty is water inflow, and when one considers gathering snow measurements to learn about future inflow (Fig. 1). The background required for this limited scope is described next. The notation we use throughout the paper is summarized in Table 1.

Time is discretized and represented by week $t \in [0, T]$, where T is the finite time horizon. The inflow q_t at time t is a random variable (exogenous

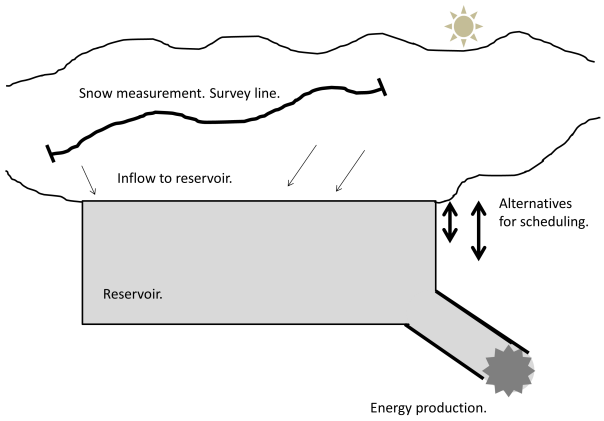

Figure 1: Schematic illustration of the reservoir water levels, inflow, and snow measurement.

Table 1: Definition of input and output variables.

Description	Variable
Time horizon	T
Number of scenarios	K
Number of production controls	J
Inflow scenarios	$q_t^{(k)}$, $k = 1, \ldots, K$, $t = 0, \ldots, T$
Reservoir levels	L_t, $t = 0, \ldots, T$
Alternatives or controls	$a_{t,j}$, $t = 1, \ldots, T$, $j = 1, \ldots, J$
Spot price	c_t, $t = 1, \ldots, T$
Classes for snow data	Y
Snow data	$y \in \{1, \ldots, Y\}$
Immediate pay-off	$\pi(q_t, a_t, L_t, c_t)$
Future value	$V_{t+1}(q_{t+1}, L_{t+1})$
Initial (prior) value	$V_0(q_0, L_0)$
Fitted regression surface value	$\tilde{V}_{t,j}$, $t = 1, \ldots, T$, $j = 1, \ldots, J$
Optimal scheduling	α
Prior value	PV
Posterior value	PoV
Value of information	VOI

variable) representing the weekly inflow from rain and melting snow. In the scheduling procedure, the uncertain inflow is represented by K scenarios (realizations) of time series. The inflow is revealed each week. (In practice this is observed from the difference in the water level in the reservoir and the releases of water for energy production.)

The alternative or control variable (endogenous variable) a_t represents the amount of water (or rate) released for electricity production at time t. It is here discretized according to different production levels. The release controls are ordered such that $a_{t,j} > a_{t,j-1}$. In our presentation we use two possible controls; $a_t \in \{a_{t,1}, a_{t,2}\}$, but this can easily be generalized like we do in the case study in Sect. 5. We further assume that these possible controls are constant over time. For mathematical clarity and the understanding of our algorithm, we note that a 'wait-and-see' approach is used, where the inflow q_t is revealed before the release control a_t is selected. An alternative is to use a 'here-and-now' strategy.

The water level in the reservoir at time t is denoted L_t. In the simplest case $L_{t+1,j} = L_t + q_t - a_{t,j}$, when applying release control j at time t. However, the reservoir has upper and lower limits $L_{\min}$ and $L_{\max}$, and if the water level crosses these boundaries, the payoff is penalized. The initial water level is specified by L_0.

The spot price c_t is deterministic in our example. This is set from market data, see e.g. [8] and [16]. The price has yearly variations caused by seasonal variability in supply and demand. At time t, the energy production gives profits $c_t a_t$ if the limits of the reservoir are not violated by the chosen control. When the price is high, it appears optimal to produce much energy. But this of course reduces the water level, and there is less potential later, especially if the expected inflow is low. Hydropower scheduling is used to optimize the overall (long-term) production of energy.

Scheduling is difficult when there is much uncertainty in the inflow of water from precipitation and snow melting. The challenge is to find the right reservoir level prior to the spring flood. If average snow melting and rainfall is assumed, the risk of spill of water is underestimated, see e.g. [19]. Snow measurements can be gathered before the melting season. This additional information is denoted y, and it will be indicative of the future water inflow to the reservoir. Recall that the inflow is revealed over operation time. The snow measurements would be available ahead of scheduling, to further guide the decisions about controls. But note that there can be poor precision in the snow measurements. Moreover, these measurements come in addition to all currently available data provided by weather forecasting and precipitation gauges. Thus, the snow measurements do not necessarily change the prior

flood forecast very much, which would be a requirement to influence the scheduling. We will conduct VOI analysis to gauge the effect of the snow measurements on the hydropower scheduling.

3 Hydropower scheduling and least squares Monte Carlo

We now present the equations underlying hydropower scheduling. We further outline a suggested approximate method for scheduling which has useful properties in our setting.

3.1 Scheduling

We ignore short-term operations [7] and concentrate on tactical hydro scheduling. The level of the reservoir decreases with the energy production and increases with the inflow. We enforce penalty terms for reaching the reservoir limits, so the immediate payoff is specified by

$$\pi(q_t, a_t, L_t, c_t) = \begin{cases} c_t \max\{0, L_t + q_t - a_t - L_{\min}\}, & L_t < L_{\min}, \\ c_t a_t, & L_{\min} < L_t < L_{\max}, \\ c_t \max\{0, L_{\max} - L_t - q_t + a_t\}, & L_t > L_{\max}. \end{cases} \tag{1}$$

This means that if the reservoir level moves above the upper limit $L_{\max}$, water flows over the dam, and it gives no income. Similarly, if the reservoir level is below $L_{\min}$, the power plant must avoid large releases to get inside the bounds, and at this level there is no income for this.

The goal of the scheduling is to maximize the expected income from future energy production. The scheduling is then defined by the following optimization problem:

$$V_0(q_0, L_0) = \max_{\boldsymbol{a}} \mathrm{E}\left[\sum_{t=0}^{T-1} \pi(q_t, a_t, L_t, c_t) + V_T \;\middle|\; L_0, q_0\right], \tag{2}$$

where we assume that current inflow q_0 and reservoir level L_0 is known, and V_T is the value of the reservoir at the final time. The expected values are with respect to the inflows with joint probability model $p(\boldsymbol{q})$. When the current inflow is not yet known, the value is $V_0(L_0) = \mathrm{E}[V_0(q_0, L_0)]$. In the long run the values will not be very influenced of q_0.

Given the current information at time step t, the dynamic behavior of the values in the reservoir can be phrased as

$$V_t(q_t, L_t) = \max_{a_t} \{\pi(q_t, a_t, L_t, c_t) + \mathrm{E}[V_{t+1}(q_{t+1}, L_{t+1}) \mid L_t, q_0, \ldots, q_t]\}, \quad (3)$$

where the first term on the right is the immediate pay-off, while the latter term is the expected value of the future pay-off, starting at the next reservoir level, and given the current information. No discounting is used here. The expectation is with respect to the conditional distribution $p(q_{t+1}|q_t, \ldots, q_0)$, and hence the conditioning on both the current reservoir level and the past inflow. For an autoregressive process we have $p(q_t|q_{t-1}, \ldots, q_0) = p(q_t|q_{t-1})$. In practice the uncertain inflows will be represented by time-series scenarios in our case. Recall that the current release is determined after the current inflow is revealed. The optimization becomes a sequential problem with expected values and maximization over the release controls. The solution to this dynamic program must be solved by approximations, see e.g. [15] and [17].

3.2 Least squares Monte Carlo

We build on an approximate solution to equation (2) using LSMC [4]. The expectations are then approximated by regression analysis using results from Monte Carlo sampling of inflows. The optimal path is constructed sequentially by comparing regression surfaces at each stage. A working assumption of this algorithm is the Markovian structure of the problem, where the current level, control and inflow bring the reservoir to the next level, but it is irrelevant how the past brought us to the current reservoir level. This means that there is a transition relation with an inverse which (in the simplest setting) is $L_t = L_{t+1,j} + a_{t,j} - q_t$, for controls $j \in \{1, 2\}$. The relation allows the sequential construction of regression surfaces and a forward selection scheme. Even though the particular LSMC algorithm is not crucial for the more VOI-focused contribution of the current paper, it is important to understand the simulation-regression idea, so we present the main parts of the algorithm in some level of detail.

Main inputs to the algorithm are: i) Discrete time-series scenarios of uncertain inflows from $p(\boldsymbol{q})$, represented by $q_t^{(k)}$, $t = 0, \ldots, T$, $k = 1, \ldots, K$. ii) Levels at the end of the time horizon $L_T^{(k)}$, $k = 1, \ldots, K$, from the uniform distribution between $L_{\min}$ and $L_{\max}$. Scenario values at time T are then $V_T^{(k)} = L_T^{(k)} c_T$, $k = 1, \ldots, K$. The algorithm uses the Markovian structure for the reservoir level to wind up scenario paths backwards in time, with an optimal forward selection at each stage.

Consider now time step t and $t+1$ only. (The algorithm will start at time $t = T-1$ and roll back to $t = 0$.) For scenario k, the water level at time $t+1$ is $L_{t+1}^{(k)}$. At time t one either used control $a_{t,1}$ or $a_{t,2}$. When we know the inflow scenario $q_t^{(k)}$, these controls entail time t water levels $L_{t,1}^{(k)}$ or $L_{t,2}^{(k)}$ (Fig. 2, left).

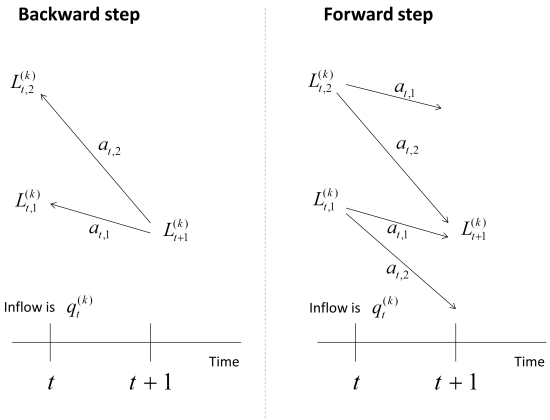

Figure 2: Illustration of the backward and forward relations connecting the reservoir levels at time t and $t+1$. Here L represents reservoir level, a release control and q is the inflow.

The algorithm finds the optimal forward selection (Fig. 2, right) from the expected value corresponding to each forward control. For comparing the different forward choices, LSMC fits regression surfaces for controls $a_{t,j}$, $j \in \{1,2\}$ using all scenarios, and starting from the levels derived from the backward step [4]. These regression surfaces are calculated separately for $j = 1, 2$ by regressing

$$\pi(q_t^{(k)}, a_{t,j}, L_{t,j}^{(k)}, c_t) + V_{t+1}(q_{t+1}^{(k)}, L_{t+1}^{(k)}), \quad \text{on } (L_{t,j}^{(k)}, q_t^{(k)}). \tag{4}$$

Note that the regression surfaces are calculated by using both controls to all samples.

We denote the fitted regression surfaces by $\tilde{V}_{t,j}$, $j \in \{1,2\}$. The parametric regression model that we use in Sect. 5 is based on second order polynomials (six parameters) for each of the alternatives j. We then have $\tilde{V}_{t,j}(q_t, L_t) = \beta_{0,j} + \beta_{1,j} q_t + \beta_{2,j} L_t + \beta_{3,j} q_t^2 + \beta_{4,j} L_t^2 + \beta_{5,j} q_t L_t$. Fig. 3 shows results of the regression fitting at one time t, with $K = 500$ scenarios. The

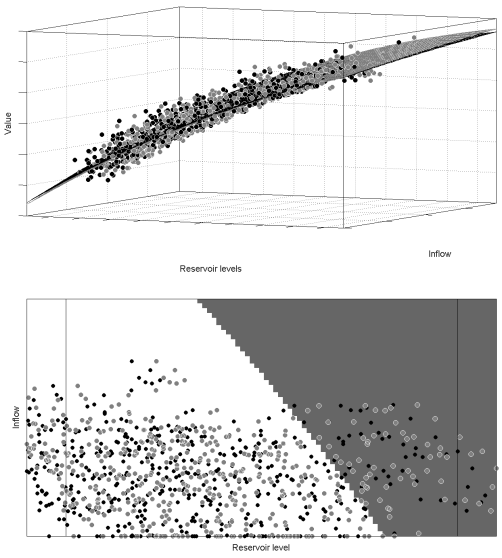

Figure 3: Example of fitted regression surfaces (top) and the selection of the largest surface (bottom) when there are two controls and $K = 500$ samples.

top display shows the scenario outputs as gray or black circles, and the corresponding fitted regression surfaces for small and larger production controls. These two surfaces are very similar in this illustration, but he gray surface is above the black one for high reservoir levels. The bottom display is made by evaluating both of the fitted regression surfaces on a regular grid of reservoir level (first axis) and inflow (second axis), and grayscale color-coding according to the largest surface. In the white region $\tilde{V}_{t,1} > \tilde{V}_{t,2}$. This means that for inflow and reservoir levels in the white region it is optimal to choose alternative $a_{t,1}$, while $a_{t,2}$ is optimal in the gray region. We denote the optimal control by

$$\alpha(q_t, L_t) = \text{argmax}_j \left\{ \tilde{V}_{t,j}(q_t, L_t) \right\}.$$

Note that the two optimal decisions $\alpha(q_t, L_{t,1})$ and $\alpha(q_t, L_{t,2})$ will be different only if the points lie on each side of the border between the white and gray area in Fig. 3, and $L_t^{(k)}$ is set as the reservoir level with optimal decision leading to $L_{t+1}^{(k)}$. Hence, these optimal decisions are used to wind the reservoir levels backward in time for each scenario. Assuming they are the same, and in Fig. 2, $\alpha(q_t^k, L_{t,1}^{(k)}) = \alpha(q_t, L_{t,2}^{(k)}) = a_{t,2}$, then the optimal

control is $a_{t,2}$, and the reservoir level at time t becomes $L_t^{(k)} = L_{t,2}^{(k)}$ since this is the volume that ends up in $L_{t+1}^{(k)}$ by release control $a_{t,2}$. If the optimal controls are not the same, the strategy is decided randomly. This happened only a few times in our application, and special decisions have to be made in these situations (see below). This rolling-back procedure continues until time 0, where we fit the initial value V_0 for reservoir level L_0 by averaging over all K inflow scenarios:

$$\hat{V}_0(L_0) = \frac{1}{K} \sum_{k=1}^{K} \tilde{V}_{0,\alpha(q_0^{(k)},L_0)}(q_0^{(k)}, L_0). \tag{5}$$

Regression fitting is often most accurate when the design span the covariate domain widely. In our setting we would at the same time avoid having too many extreme scenarios, because there is a chance that they break outside the reservoir limits $L_{\min}$ and $L_{\max}$ [4]. (See points outside vertical line limits in Fig. 3.) In our implementation we assign limits $L_{\max}^{+}$ and $L_{\min}^{-}$ which are more extreme maximal and minimum levels. If a scenario goes beyond any of these extreme limits, we randomly reassign a level uniformly in the feasible area. In our examples this happened only about 5 % of the time. To avoid end problems we use a rather long time period, T, set several years ahead. But the relevant time horizon we are looking at in the example is that of spring and summer.

Algorithm 1 summarizes the scheduling algorithm. We illustrate hydropower scheduling with parameters inspired from our case, where the scheduling starts in winter/spring before the snow flooding. We use Algorithm 1 with two possible production controls each week. Fig. 4 illustrates how the average production decisions (top) and reservoir levels (middle) vary together with the price (bottom). (The controls are set rather extreme

Algorithm 1 Value of scheduling using simulation of inflow scenarios and regression fitting.

Require: Inflow realizations $(q_0^{(k)}, \ldots, q_T^{(k)})$ and uniform random reservoir levels between $L_{\min}$ and $L_{\max}$ at time T; $L_T^{(k)}$, for $k = 1, \ldots, K$.

1: **for** $t = T - 1$ to 0 **do**

2: Applying the Markov transition on $L_{t+1}^{(k)}$, for $j = 1, \ldots, J$, build regression surfaces $\tilde{V}_{t,j}(q_t, L_t)$, $j = 1, \ldots, J$ using the scenarios (equation (4)).

3: Find the optimal forward control $\alpha(q_t^{(k)}, L_t^{(k)})$ and associated water levels $L_t^{(k)}$.

4: **end for**

5: **return** For initial water level L_0, compute estimated value from regression surface $\hat{V}_0(L_0)$ (equation (5)).

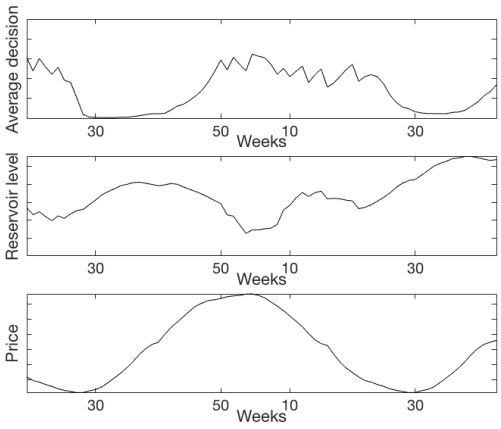

Figure 4: Illustration of scheduled production with two controls, reservoir trajectory and price path.

in this illustration.) In summer, when the price is low and there is little demand for energy, the decision is to produce little, and the reservoir level rises. There appears to be a lag in the system, where the highest (lowest) reservoir level is reached a little after the lowest (highest) price. The small zig-zag shape occurs when one is almost indifferent about which control to use.

4 Value of information of snow measurements

The VOI is defined as the expected additional value obtained by data. See e.g. [11] for foundational background on decision analysis and VOI. In this context hydropower scheduling as described in Sect. 3 gives the prior value PV $= V_0$ (equation (2)). But before scheduling, in winter or spring, the decision maker can choose to purchase snow measurements (at a price). This would lead to better scheduled production, and a higher expected value than what was attained without this snow information. The expected value with snow measurements is called the posterior value.

Snow measurements, denoted $\boldsymbol{y}$, are informative of the uncertain inflow. The value of scheduling conditional on this data is using the posterior dis-

tribution $p(\boldsymbol{q}|\boldsymbol{y})$, and it becomes

$$V_{0|\boldsymbol{y}}(q_0, L_0) = \max_{\boldsymbol{a}} \mathrm{E}\left[\sum_{t=0}^{T-1} \pi(q_t, a_t, L_t, c_t) + V_T \;\middle|\; L_0, q_0, \boldsymbol{y}\right]. \tag{6}$$

Note that the scheduling is still based on the sequential inflow information over time. The snow measurements are additional to this, and they are available up front, before the melting period. If the first inflow variables are very informative about the latter inflows, knowing the snow data would not help so much in the scheduling. But when there is larger variability in the inflows, knowing the snow data (and hence about the aggregated future inflow) could provide important information for the scheduling.

Similar to what was done in Sect. 3.2, the optimal strategy $\tilde{V}_{0,\alpha(q_0^{(k)},L_0|\boldsymbol{y})}$ is computed, but now using the subset of scenarios consistent with the data. The average is then taken over the relevant inflow scenarios, gives the approximate value

$$\hat{V}_{0|\boldsymbol{y}}(L_0) = \frac{1}{|\mathcal{K}_y|} \sum_{k \in \mathcal{K}_y} \tilde{V}_{0,\alpha(q_0^{(k)},L_0|\boldsymbol{y})}(q_0^{(k)}, L_0), \tag{7}$$

where $\mathcal{K}_y$ is the set representing the fraction of scenarios consistent with measurement y.

The data are typically modeled by a likelihood model $p(\boldsymbol{y}|\boldsymbol{q})$ connecting the snow measurements and the inflow variables. The marginal distribution for the data is obtained by marginalizing over the inflows; $p(\boldsymbol{y}) = \sum_{\boldsymbol{q}} p(\boldsymbol{y}|\boldsymbol{q})p(\boldsymbol{q})$, assuming a discrete sample space for the inflows. The posterior value is defined by taking the expectation over all data;

$$\mathrm{PoV} = \sum_{\boldsymbol{y}} \hat{V}_{0|\boldsymbol{y}}(L_0) p\left(\boldsymbol{y}\right), \tag{8}$$

assuming a discrete sample space for the data. Comparing the posterior value defined by equations (6), (7) and (8) with the prior value in equation (2), we see that the production controls can now vary depending on the outcome of snow data. The informed decisions about controls mean that, on expectation, the releases are timed more effectively, and the posterior value is larger or equal to the prior value. Note that even with statistical approximations involved, this increase in value can be guaranteed by using double expectation $E(V_0) = E(E(V_0|y))$ for the prior value.

For a risk-neutral decision maker the VOI is the difference between posterior and prior value:

$$\mathrm{VOI} = \mathrm{PoV} - \mathrm{PV}. \tag{9}$$

In Sect. 3 the prior value was approximated by LSMC in Algorithm 1. Our suggested approach for approximating the posterior value and the VOI builds on similar scheduling approximations (Sect. 4.2).

4.1 Snow measurements

The posterior value calculation requires a quantitative assessment of the snow measurements, and a link between the inflow and the snow measurements. The first is related to the processing of measured snow data. The second is tied to the likelihood model for the snow measurements, given the inflow variable.

One typically acquires measurements in the catchment and combines them to predict inflow, see e.g. [14] or [13]. Measurements are collected by a survey team going along line segments in a snow scooter (Fig. 1). Point measurements are acquired with radar instruments giving the snow water equivalent(SWE) at the locations where they are made. The SWE combines the density and thickness of the snow to provide the volume of water this snow will bring to the reservoir. The SWE measurements can be calibrated by calculating both the inflow and the precipitation in the catchment area, and taking the difference between the two. When doing this kind of calibration one assumes that all the precipitation will either be stored as snow in the catchment area or inflow to the reservoir. [1] suggest extensions of current methods for assimilating climate models and snow density measurements to improve the prediction of SWE. Others have suggested methods using weather forecasts and satellite data to estimate the SWE in a catchment [12]. [2] suggest using public web images.

The additional information provided by such snow measurements, $\boldsymbol{y}$, can be included in various ways. Looking at historical data, it is common practice to aggregate the snow measurements to one scalar value y. We will adopt this view here, and further assume that the snow measurement is categorized in one of Y classes. These classes are representative of the total amount of snow in the catchment, and the classification is based on historical data (usually acquired along the same geographical line segments). The K inflow scenarios are similarly divided into Y classes for the aggregated inflow in the realizations. This means that the likelihood maps the largest accumulated snow realizations to the largest measurement class, and similarly for the other classes. If one measures a lot of snow, the high-inflow realizations dominate the posterior. The classification is such that groups are equally likely, i.e. $p(y) = 1/Y$. (The sensitivity to this likelihood model is studied in the case study below.)

Decision analysts often use the concept of clairvoyance or perfect information to study the effect of knowing an uncertainty before a decision is made. In our setting, perfect information entails knowing the future inflows at all times, which is unrealistic. Aggregated perfect information means knowing the accumulated future inflow, which in our case is obtained as the limit when the number of snow classes Y increases.

Note that this classification entails an enormous reduction of the actual data size in the snow measurements. It does not account for the spatial variables of the situation, which can be important to get the full benefits of such geographical data. However, it turns out that the spatial elements associated with snow measurements are difficult to integrate in a unified spatial statistical model [3] because there is so much heterogeneity. In our case, the higher-resolution snow measurements were extremely sensitive to local variations in the terrain. Possibly, better monitoring systems obtained by new sensors and automated measurements will be available in the future, and this would allow decision makers richer data gathering opportunities, such as high-fidelity testing at subsets of the spatial domain [6]. However, because of the large heterogeneity in the spatial snow measurements, it is currently more robust to stick with the aggregated view where measurements are indicative of the snow level category in the catchment.

4.2 Algorithm and VOI analysis

The VOI computation builds on the LSMC approach for scheduling presented in Sect. 3, but this is now computed conditional on different data outcomes, and then averaged.

When the snow measurement is in class y, Algorithm 1 is used for the associated subset of inflow scenarios. The resulting value is denoted $\hat{V}_{0|y}(L_0)$. The regression surfaces are here estimated using this subset of realizations alone. Clearly, the regression surfaces vary, and the fitted strategies can also differ between the data $y \in \{1, \ldots, Y\}$. The approximated posterior value is

$$\hat{\text{PoV}} = \sum_{y=1}^{Y} \hat{V}_{0|y}(L_0) p(y). \tag{10}$$

Algorithm 2 summarizes our approach for approximating the VOI.

The snow flooding during spring is of interest, and in our implementation the value calculations go on for 20 weeks (from week 16 to 36). Then the value of the reservoir is added as the water volume times the mean spot price from time step 20 until the time horizon T. This is done to avoid

Algorithm 2 Value of information of snow measurements for hydropower scheduling.

Require: Inflow scenarios $(q_0^{(k)}, \ldots, q_T^{(k)})$, $k = 1, \ldots, K$, initial inflow q_0 and water level L_0, and measurement categories $y \in \{1, \ldots, Y\}$.

1: **for** $y = 1, \ldots, Y$ **do**
2: Pick the subset of scenarios corresponding to data class y.
3: Find the value $\hat{V}_{0|y}(L_0)$ based on the scenarios picked from data class y. See Algorithm 1.
4: **end for**
5: **return** $\hat{\text{PoV}} = \sum_{y=1}^{Y} \hat{V}_{0|y}(L_0)p(y)$. Subtract the prior value from Algorithm 1 to get the VOI.

impact from different scheduling in the future. We used $K = 50,000$ Monte Carlo samples in the case study. Most likely this could be reduced through variance reduction techniques, but for our case the run time of Algorithm 2 was only one minute on a laptop computer.

In Fig. 5 we illustrate the average decision and reservoir levels in the simplest situation (from Sect. 3.2) with only two rather extreme production controls and two snow classes (high or low). The display shows that the

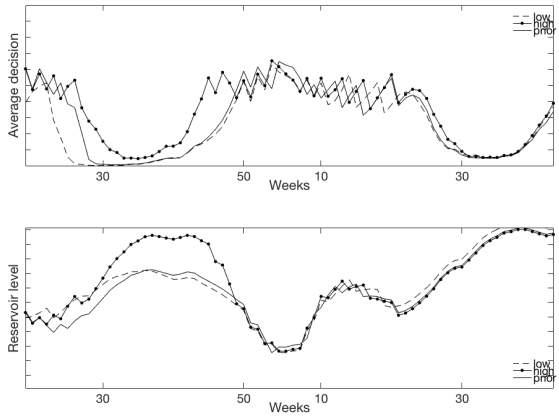

Figure 5: Average decision and reservoir levels based on prior inflows and posterior inflows (high or low snow measurements).

scheduling gives higher production when there is more snow. There is a longer period of intensive energy production in the spring, and the production also steps up earlier in the fall. During summer the reservoir level is higher for the case with much snow. The opposite happens when data tell us there is little snow. Since the information leads to different scheduling

compared with that obtained without snow measurements (prior), the VOI is positive here.

The VOI should be compared with the price of data acquisition and processing. If the VOI exceeds the price, it is worthwhile purchasing the data. However, rather than being just a single number or result, the VOI is commonly regarded as a basis for discussion. In particular, it is considered good practice to look at the sensitivity of the VOI to various input parameters in the model. Moreover, the VOI is often used to compare multiple data gathering opportunities, and different accuracies of the data, compared with the price. We show examples of this for our case study.

5 Case study from Norwegian power plant

We consider scheduling and VOI analysis of snow measurements for a Norwegian hydropower plant. We first provide some background for this case and then conduct VOI analysis.

5.1 Reference case

We study a high-drop medium-sized power plant located in Southern Norway. The hydropower plant reservoir levels (top) and energy production (bottom) for recent years are shown in Fig. 6. Because of price and in-

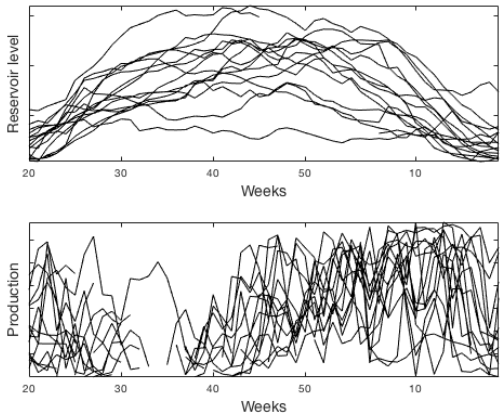

Figure 6: Time series of reservoir level (top) and production (bottom) from the Norwegian power plant.

flow expectations, and the scheduling used by the hydropower company, the reservoir has a tendency to be low before the melting period. It reaches its highest level around October. The production is highest when the reservoir level decreases (winter), or when there is much inflow from snow (spring). The production is small during summer since there is little demand. The data are comparable to the scheduled reservoir level and production in Fig. 4, but note that there were only two ($J = 2$) weekly production rates in Fig. 4.

Inflow data from recent years are shown in Fig. 7. We note that there

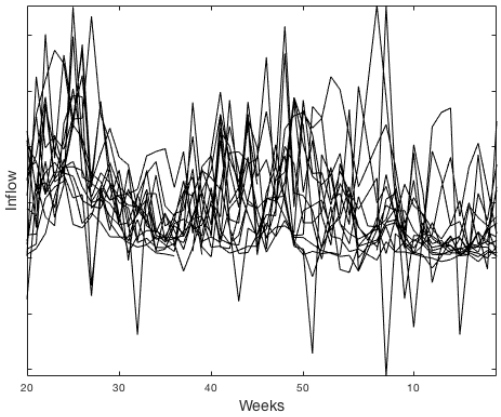

Figure 7: Time series of inflow from the Norwegian power plant.

is more inflow in week 20-30 (spring) due to melting snow. This will be relevant for our assessment of the value of snow information. The display further shows large variability in the inflows, especially in late fall/early winter because there is more precipitation and some snow in short cold periods, which melts again relatively fast when the temperature increases.

The scheduling approach requires that we generate realizations of inflow, and we use data from the last ten years to fit a time series model. The inflow is a weekly variable; q_t, $t = 0, \ldots, 52$. We model a trend and a heterogeneous variance component: For each week we set mean $E(q_t) = \hat{\mu}_t = \frac{1}{10}\sum_{l=1}^{10} q_{t,l}$ and variance $\mathrm{Var}(q_t) = \hat{\sigma}_t^2 = \frac{1}{9}\sum_{l=1}^{10}(q_{t,l} - \hat{\mu}_t)^2$, where l is an index for the subsequent years. Using a Gaussian model with these means and variances, we estimate the autoregressive correlation coefficient, $\rho = \mathrm{Corr}(q_t, q_{t+1})$. By likelihood maximization for our dataset this gives an autocorrelation of $\hat{\rho} = 0.2$. This is relatively low, meaning that there is limited dependence in the time series scenarios. Realizations of inflow are then generated by drawing

K independent length $T = 52$ time series; $(q_0^{(k)}, \ldots, q_T^{(k)})$, $k = 1, \ldots, K$, from the fitted multivariate Gaussian distribution.

5.2 VOI results

We will be interested in studying the value of snow data for different reservoir volumes, control levels and data accuracy. In Fig. 8 historical snow data are plotted. By week 30 the snow has usually melted. It starts to build up

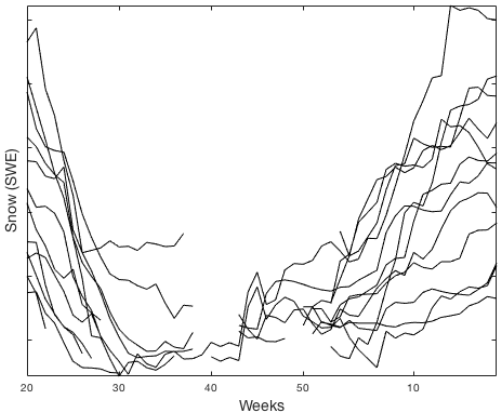

Figure 8: Time series of snow water equivalent (SWE) from the Norwegian power plant.

again in the fall. The snow flooding from the catchment area will contribute to increased inflow. Together with the rainwater it makes the total inflow during spring bigger than for the rest of the year (Fig. 7, week 20-30). However, the amount of snow and the form of decline during the melting period vary much from year to year, dependent on the temperature and the amount of precipitation during winter. As discussed in Sect. 4, the likelihood model (and associated posterior model) is constructed by dividing snow and inflow in groups. The grouping is based on snow amounts in week 16 and aggregated inflows until week 30. The times were chosen based on the correlation in inflow and snow amount.

Numerical testing showed that the average value of the total production one year increased between up to 10 % by including the snow measurement. This might seem like a small VOI, but the annual production in Norway is 130 TWh in total, worth more than $ 500 million. Not all of this is related

to snow flooding, but even so, a small value improvement could result in large profits, and this can justify the snow measurement acquisition in some cases. The VOI should be compared with the actual price of data acquisition and processing. This work consists of a small team that plans and acquires the data, and then the subsequent processing of data. The largest cost for the hydro-power company is man-hours, typically a couple of weeks job for a small team of persons.

We next different study how key input parameters influence the VOI. Consider different classes $Y = 3$, 6, 9 and 12 for the snow measurement category, and assume there are $J = 2$ or $J = 4$ production controls each week. The controls are now tuned to realistic levels for the reservoir case study. The results are shown in Table 2. For $Y = 3$ the VOI is zero for both two

Table 2: VOI results with different snow measurement classes and production controls. The displayed numbers are relative values in % of largest value with four controls and 12 snow classes.

Classes	3	6	9	12
2 production rate controls	0	0	9	21
4 production rate controls	0	23	47	100

and four production controls. This is caused by very small differences in the optimal strategies (especially when there are few production alternatives), with or without data. Neither of the (rather large) subclasses of posterior inflows give strategies that are very close to the upper or lower reservoir limits. This means that even though the inflow can be large or small, the reservoir capacity can store this inflow. This is also seen in the historical data which has very little overflow. By letting the number of snow classes increase, the data influence increases as we get less imperfect information, and the posterior inflow scenarios become more separated. The VOI is then larger. In our application, 12 snow classes means a very accurate inflow prediction from the snow data. Further, Table 2 shows that effect of additional flexibility in the production controls. With more controls (4 instead of 2), one can run the reservoir levels closer to the upper (lower) limits. The accurate snow data will indicate which of these more extreme strategies are rewarding to the company. This added flexibility of having more production controls would clearly depend on the time resolution. With very small time intervals and many controls the method could show insignificant differences in the regression surfaces, at least when using relatively few scenarios.

We study the sensitivity to the reservoir size in more detail. The snow

will always contribute to a flood of water in spring, but the fraction of the total inflow that is originating from the snow will of course depend on the situation. For our case, the maximum measurement of the snow reservoir corresponds to about 15 percent of the total inflow for one year. This means that the snow catchment area is quite small compared to the total inflow. When we reduced the capacity (upper limit) to 60 percent of the original, the VOI is positive even for two production alternatives and three snow classes. Now that the risk of overflow is much higher, the need for an optimal strategy gets more important, and the snow measurements carry more information. In this case, snow measurements are likely worthwhile collecting even when there are few control opportunities. Then again, for very small reservoirs with large inflow, it is likely best to produce as much as possible at all times and data would not have much influence on this decision.

A low VOI may be the case for a number of reservoir plants. For example, [18] find that the value of reducing SWE error is small for a hydropower system in California.

In the above results and discussion, the snow measurements were assumed to clearly separate the aggregated inflow scenarios, and with many snow classes the information goes toward perfect information about the accumulated future inflow. We now relax this assumption by introducing a misclassification probability in the likelihood model. This means that the snow measurement is less accurate: even though we condition on a large snow amount, some of the low-inflow scenarios can enter the posterior model. The VOI decreases by about 25 percent when there is a 40 percent probability of inflows from another (random) group. This reduction is rather constant across different number of alternatives and measurement classes. Here, 0 percent misclassification probability represents the VOI in Table 2, while 100 percent probability would simply reproduce the prior value for the posterior value via randomization of samples. Note that this is just one of several other likelihood models : There could be larger misclassification rates for much snow, or one could use multiple snow measurements to perhaps indicate the time profile of the snow flooding and the more probable inflow scenarios.

The VOI is often used to compare various information gathering schemes for a range of prices. Data acquisition comes with the cost of snow scooters and radar equipment as well as man-hours [9]. The processing requires testing and interpretation to get useful SWE data at the catchment scale. An example is made here assuming the snow data are acquired (with price P_1) and then processed in two different ways. Processing method 1 has price P_2 while method 2 has price $P_2/5$. We assume that the higher price corre-

sponds to an accurate snow measurement method, which is here represented by the direct link between the snow measurement and the inflow. The less expensive method corresponds to inaccurate processing, which is here represented by misclassification of data with 40 percent chance of inflow from another (random) group. Decision regions are constructed by selecting the data processing scheme that has the largest VOI compared with its price. The best data gathering option (No data, Method 1, Method 2) is in this situation defined via

$$\text{argmax}\{0, \text{VOI}_1 - P_1 - P_2, \text{VOI}_2 - P_1 - P_2/5\}. \tag{11}$$

The resulting decision regions are plotted in Fig. 9. The price P_1 of acquiring

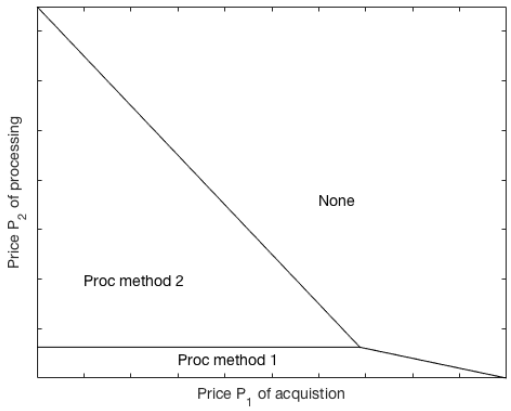

Figure 9: Decision regions for snow measurement acquisition using the VOI results.

data is on the first axis and the price P_2 of processing the accurate test is on the second axis. When the price of processing increases, method 2 is preferred. When both acquisition and processing prices are high, none of the tests are worth conducting. Such diagrams can be constructed for various reservoir limits and model input parameters. It lays a constructive foundation for discussion among decision makers that must choose among various data acquisition or processing opportunities.

6 Closing remarks

We present an approach to conduct VOI analysis of snow measurements for hydropower scheduling. Results show that we need high-accuracy snow measurements to get significant VOI of the snow data in our case study. One reason for this is that the catchment is rather small compared to the reservoir size. For a smaller reservoir we get narrower margins, and snow measurements can then improve the scheduling results, giving a larger VOI. In practice this means that snow measurements are unlikely to add much value for reservoirs that are large compared with the inflow. For reservoirs that are small compared with the inflow, snow information can add value through the use of intelligent survey designs for the snow measurements. This will give accurate measurements of the snow level which are indicative of the future inflows. However, surveys with large coverage and much detail might be costly - a less expensive practice is to use similar measurement conditions over different years so that a robust likelihood classification of the snow data, and associated inflows, can be made from historical data. We recommend conducting a VOI analysis to compare different survey opportunities for snow measurements, and to evaluate whether the additional value of this information can justify the cost of getting the snow measurements.

The presentation is focused on LSMC for hydropower scheduling, using scenarios of water inflows as input. The methodology for VOI analysis builds on LSMC scheduling policies over subsets of the inflows to approximate the posterior value. The run time of Algorithm 1 is linear in time and in the number of scenarios. VOI computations in Algorithm 2 scale accordingly, just by another linear factor for the number of measurement classes. In the application we use two or four possible release controls at each time step. There is only a moderate increase in the computational cost of more controls using our approach - it only entails the approximation of one regression surfaces for each control. The LSMC scheduling algorithm that we employed here is however not directly applicable to situations with more than one reservoir, which would give more coupling than what is modeled in the Markovian transitions for the reservoir water level. The suggested VOI approximation would however also be applicable with other simulation-regression approaches for scheduling.

Acknowledgments

We thank the hydropower company for letting us analyze and present their dataset, and for providing insight on relevant questions. We further thank Oddbjørn Bruland, Knut Sand and Yisak Sultan for discussions on snow measurements and hydropower production, and Tord Olsen for scientific discussions and careful reading of the paper. Fleten acknowledges support from the Research Council of Norway through project 268093, and recognizes the Norwegian Research Centre for Hydropower Technology - HydroCen (project 257588).

References

[1] Bruland, O., Færevåg, A., Steinsland, I., Liston, G., Sand, K.: Weather SDM: Estimating snow density with high precision using snow depth and local climate. Hydrology Research **46**(4), 494–506 (2015)

[2] Castelletti, A., Fedorov, R., Fraternali, P., Giuliani, M.: Multimedia on the mountaintop: Using public snow images to improve water systems operation. In: Proceedings of the 2016 ACM on Multimedia Conference, MM '16, pp. 948–957. ACM, New York, NY, USA (2016)

[3] Cressie, N.A.C., Wikle, C.K.: Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics. Wiley (2011)

[4] Denault, M., Simonato, J.G., Stentoft, L.: A simulation-and-regression approach for stochastic dynamic programs with endogenous state variables. Computers & Operations Research **40**(11), 2760–2769 (2013)

[5] Desreumaux, Q., Côté, P., Leconte, R.: Role of hydrologic information in stochastic dynamic programming: a case study of the Kemano hydropower system in British Columbia. Canadian Journal of Civil Engineering **41**(9), 839–844 (2014)

[6] Eidsvik, J., Mukerji, T., Bhattacharjya, D.: Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis. Cambridge University Press, Cambridge (2015)

[7] Fleten, S.E., Kristoffersen, T.K.: Short-term hydropower production planning by stochastic programming. Computers & Operations Research **35**(8), 2656–2671 (2008)

[8] Fosso, O.B., Gjelsvik, A., Haugstad, A., Mo, B., Wangensteen, I.: Generation scheduling in a deregulated system. the Norwegian case. IEEE Transactions on power systems **14**(1), 75–81 (1999)

[9] Goninon, T., Pretto, P., Smith, G., Atkins, A.: Estimating the economic costs of hydrologic data collection. Water Resources Management **11**(4), 283–303 (1997)

[10] Guariso, G., Rinaldi, S., Zielinski, P.: The value of information in reservoir management. Applied Mathematics and Computation **15**(2), 165–184 (1984)

[11] Howard, R.A., Abbas, A.: Foundations of decision analysis. Prentice Hall (2015)

[12] Kolberg, S., Gottschalk, L.: Updating of snow depletion curve with remote sensing data. Hydrological Processes **20**(11), 2363–2380 (2006)

[13] Lundberg, A., Granlund, N., Gustafsson, D.: Towards automated 'ground truth' snow measurements-a review of operational and new measurement methods for Sweden, Norway, and Finland. Hydrological Processes **24**(14), 1955–1970 (2010)

[14] Marshall, H.P., Koh, G.: FMCW radars for snow research. Cold Regions Science and Technology **52**(2), 118–131 (2008)

[15] Nandalal, K.D.W., Bogardi, J.J.: Dynamic programming based operation of reservoirs: applicability and limits. Cambridge University Press (2007)

[16] Paraschiv, F., Fleten, S.E., Schürle, M.: A spot-forward model for electricity prices with regime shifts. Energy Economics **47**, 142–153 (2015)

[17] Powell, W.B.: Approximate Dynamic Programming. John Wiley & Sons, Inc. (2011)

[18] Rheinheimer, D.E., Bales, R.C., Oroza, C.A., Lund, J.R., Viers, J.H.: Valuing year-to-go hydrologic forecast improvements for a peaking hydropower system in the Sierra Nevada. Water Resources Research **52**(5), 3815–3828 (2016)

[19] Séguin, S., Fleten, S.E., Côté, P., Pichler, A., Audet, C.: Stochastic short-term hydropower planning with inflow scenario trees. European Journal of Operational Research **259**(3), 1156–1168 (2017)

[20] Tejada-Guibert, J.A., Johnson, S.A., Stedinger, J.R.: The value of hydrologic information in stochastic dynamic programming models of a multireservoir system. Water Resources Research **31**(10), 2571–2579 (1995)

[21] Wallace, S., Fleten, S.E.: Stochastic programming models in energy. In: A. Ruszczynski, A. Shapiro (eds.) Stochastic Programming, pp. 637–677. Vol. 10 of *Handbooks in Operations Research and Management Science*. Elsevier Science (2003). URL `http://dx.doi.org/10.1016/S0927-0507(03)10010-2`