Bayesian Assimilation of Snow Information

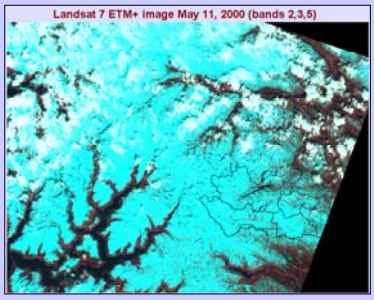
Sjur Kolberg, SINTEF Energy Research Sjur.Kolberg@sintef.no Lars Gottschalk, University of Oslo Lars.Gottschalk@geo.uio.no

Introduction

In each 1*1 km grid cell of a distributed model, a snow depletion curve (SDC) is used to represent the sub-grid heterogeneity of snow water equivalent (SWE). Since the snow covered area (SCA) reduces gradually through the melt season, it is desirable to update the model state by satellite observations of SCA.

However, observed SCA is a single number, whereas the model state consists of several variables. Which one should be adjusted? The presented routine provides updated estimates for the entire set, based on both the observation and the information available prior to update.

The method is a strict application of Bayes' theorem. Prior knowledge and uncertainty is formulated as a joint probability distribution in the model's state variables. Multiplying with a likelihood based on observations gives a posterior distribution, and comparing prior and posterior variance evaluates the effect of the added information.



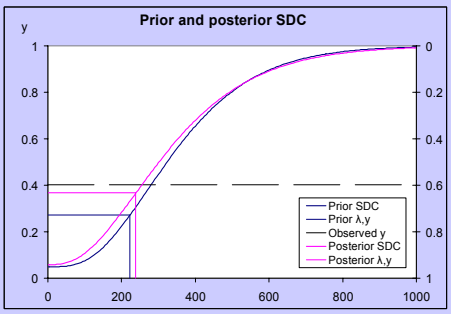
The Landsat ETM+ image is classified at 30 m resolution using a decision-tree algorithm which also estimates the uncertainty.

The model

The snow depletion curve is a cumulative probability distribution of point snow storage at the end of the accumulation season. Assuming homogeneous melt, it is also an expression of the bare-ground fraction as a function of accumulated melt depth. The SDC is thus a link between the snow cover mass balance, the accumulated melt depth, and the observable snow covered area fraction.

The current SDC model consists of three parameters; an initial bare ground fraction, and a 2-parameter Gamma distribution for the snow-covered part of the cell. These are static variables describing the situation at melt onset. Snowmelt is simulated dynamically by a temperature index equation, and the accumulated melt depth completes the snow pack state description.

The mass balance is given by the melt depth splitting the area above the curve in accumulated snowmelt volume and remaining SWE, respectively. The approach relies on identifying a transition from snow accumulation to snowmelt season, and treats mid-winter melt events and spring snowfall as special cases.



The SDC through updating: Both the curve and the dynamic state is updated. The posterior y=1-SCA does not equal the observed, due to observation variance.

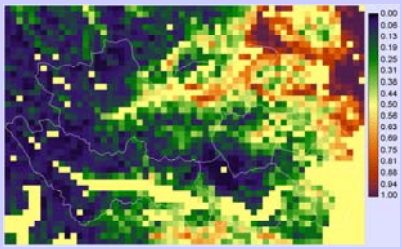
The Bayesian method

Bayes' theorem is used independently for each grid cell's SDC. It states how the posterior parameter distribution is found from the prior distribution and the conditional likelihood of the observations:

$$p(\theta|D) \propto p(\theta) \cdot p(D|\theta)$$

Here, θ contains the three SDC parameters and the melt depth, giving a 4-dimensional distribution for a single image. The four variables are assumed a priori independent, with 2 parameters in each marginal prior. The data D is the observed bare-ground fraction and its uncertainty. The likelihood is given a Beta model with expectation given by the prior parameters and variance estimated from the satellite image analysis.

With n images, the posterior has 6+4n parameters, and 3+n dimensions which are no longer independent. The posterior moments cannot be found analytically, and the distribution is sampled by MCMC. Since the posterior is not on a closed form, multiple-image updates are performed simultaneously, not in sequence. Careful adjustment of the MCMC sampling numerics is not feasible due to the large number of cells.



The observed y map: Bare-ground fraction y=1-SCA degraded to 1 km resolution. Yellow areas are lakes or outside edge of image.

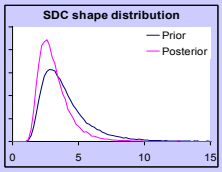
The variables:

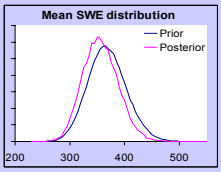
SDC shape α is a parameter in the Gamma formulation of the SDC, and describes the degree of small-scale spatial redistribution.

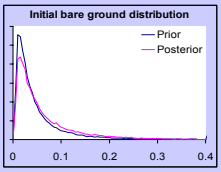
Mean SWE m is used rather than scale as parameter in the Gamma SDC, and is the expectation of SWE at the onset of the melt season.

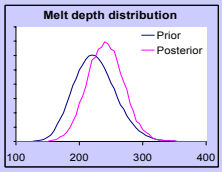
Initial bare ground y_0 is an offset to the Gamma SDC, describing the fractional bare-ground area at the onset of the melt season

Accumulated melt depth λ is the argument to the SDC. It is the only dynamic variable in the model; other dynamic states are calculated from λ.









Prior information relies on recent or historic snow surveys, and is given a Inverse Gamma distribution model to capture a low left tail.

Updating reduces the variance in 80% of the grid cells, with spatial variability and average/ median reduction as indicated in the figure below.

Prior information is interpolated from a gauge network, with a lapse rate added. A Gamma model is chosen, partly due to its additive properties.

Updating reduces the variance in 78% of the grid cells. As for melt depth, variance increases mainly occur at mid and high altitudes.

Prior information is sparse, but may be obtained from pre-melt images. A Lognormal model provides the highly skewed distribution.

Updating reduces the variance in 70% of the grid cells, while the increase in others is severe, in particular where low SCA is observed at high altitudes.

Prior information is taken from the melt model, variance proportional to mean. A Gamma model approaches symmetry as the estimate increases.

Updating reduces the variance in 63% of the grid cells. Increases in variance mainly appear at high altitutde, where simulated melt uncertainty is low.

Results

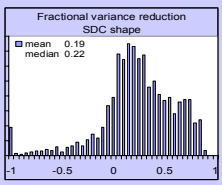
The proposed method can be used to update a operational model, by inserting the joint expectation of the posterior distribution into the model's state variables. It can also be used to assess the value of the added information, by comparing posterior variances to prior.

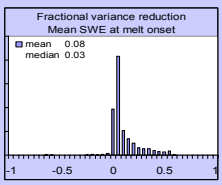
The reduction of uncertainty varies strongly with location, influenced by prior information, SCA sensitivity to parameters, the observed SCA value, and its difference from the a priori expectation. Using more than one image improves the results, and two images used separately gives correlated parameter updates.

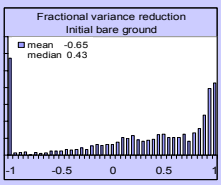
Uncertainty increases do occur, mainly where the observation is in strong conflict with the prior information. In some cases, the posterior is bimodal. Apart from being a sampling problem, such cases reveal a violated assumption or an underestimated prior variance. Quantitative updates or uncertainty estimates are doubtful is these cases.

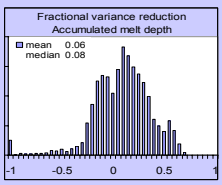
The median variance reduction is best for SCA and and the pre-melt bare ground fraction, moderate for the SDC shape and the mass balance, and small for pre-melt SWE m and melt depth λ. The parameters λ and m appear in the likelihood as a ratio only, and consequently the marginal uncertainty for these are not substantially reduced with a single image.

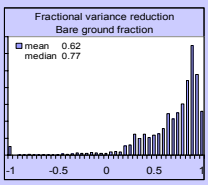
By its direct connection to the observation, the y_0 parameter often dominate in difficult cases, in particular when observed bare ground is larger than expected. It remains to find a good prior for y_0 to reduce this effect. Also, the degree-day melt equation has its biases, and the selected variance-proportional error model does not capture the melt depth uncertainty.

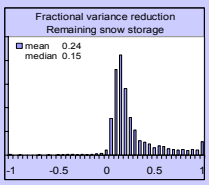




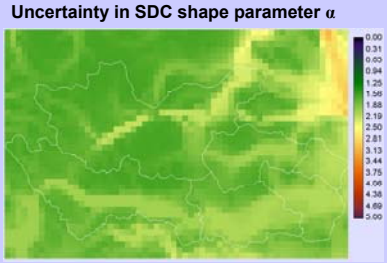




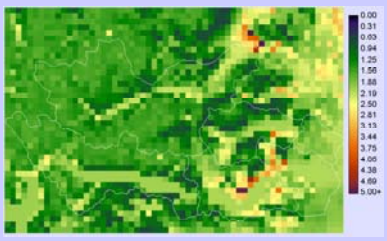




Uncertainty in SDC shape parameter α

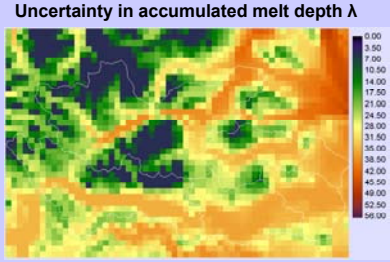


Prior SD[α]: Linked to the expectation, decreasing with elevation

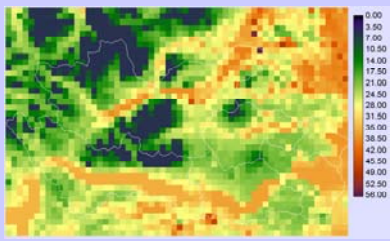


Posterior SD[α]: Generally decreased, but higher where the observation is surprisingly close to 0 or 1.

Uncertainty in accumulated melt depth λ

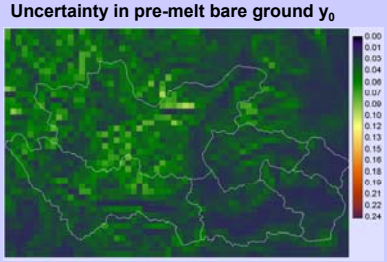


Prior SD[λ]: Linked to the expectation, decreasing with elevation

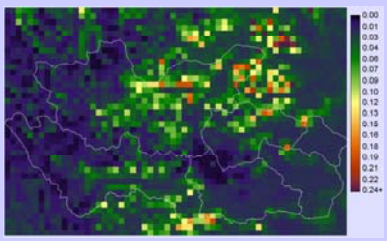


Posterior SD[λ]: Decreased at intermediate elevations in the eastern part of the region, where observed bare-ground is high.

Uncertainty in pre-melt bare ground y_0



Prior SD[y_0]: Linked to expectation, increasing with terrain roughness



Posterior SD[y_0]: Reduced where observed bare-ground is low, strongly increased where bare-ground is surprisingly high.

SINTEF

EBL KOMPETANSE