Developing models for winter processes in grasslands – Bayesian calibration and sensitivity analysis

Utvikling av modeller for vinterprosesser i eng – Bayesiansk kalibrering og sensitivitetsanalyse

Philosophiae Doctor (PhD) Thesis

Stig Morten Thorsen

Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences

Ås 2010

Thesis number 2010: 9
ISSN 1503-1667
ISBN 978-82-575-0921-7

Acknowledgements

This dissertation constitutes the written part of the requirements for the doctoral degree Philosophiae Doctor (PhD). My work is carried out at the Norwegian Institute for Agricultural and Environmental Research, Grassland and Landscape Division (Bioforsk), and the Department of Mathematical Sciences and Technology, the Norwegian University of Life Sciences (UMB). This thesis was funded as part of the strategic institute programme *Climate Change effects on winter survival of perennial forage crops and winter wheat, and on plant diseases and weed growth and control at high latitudes* (WINSUR). The thesis contains an introductory section where the methods used for model calibration and sensitivity analysis are illustrated, summary of the four included papers and a section with concluding remarks.

Firstly, I would like to thank Prof. Arkadi Ponossov (main supervisor) for introducing me to the field of applied mathematics, and for accepting the responsibility of being my main supervisor. I would also like to thank Dr. Lars Egil Haugen (ass. supervisor) for introducing me to the field of soil science. It is vital to have a basic understanding of what happens below the soil surface when working with crop related modelling.

Secondly, I would like to thank Dr. Marcel Van Oijen for sharing his knowledge and experience in the field of modelling as well as model calibration. He has patiently answered all my questions and thoroughly reviewed paper manuscripts. I am grateful for all the opportunities we had to meet throughout this work.

Thirdly, I would like to thank Dr. Mats Höglind (ass. and local supervisor). He is my partner in modelling, and his door is always open. We are currently the only ones working with model development at our research station, and I am grateful for his companionship and valuable discussions. I also appreciate the outstanding cooperation with my fellow PhD-student Anne-Grete Roer on our two papers. Although she is located at Ås, time was always taken when a discussion was called for.

I am grateful for all my good colleagues at Bioforsk, both locally at Særheim and at the other research stations. Especially I would like to thank Dr. Anne Kari Bergjord at Bioforsk Midt Kvithamar for her cooperation with the frost tolerance model for winter wheat, and Dr. Trond Rafoss at Bioforsk Plantehelse for his enthusiasm and encouragement.

Finally, I would like to thank my girlfriend and wife Heidi for patiently listening to all my frustrations and for her unconditional support; and my daughter Oda for dragging dad out of his scientific bubble.

Sandnes, January 2010
Stig Morten Thorsen

Contents

1 List of papers

This following papers are included in this thesis. They will be referred to by their roman numerals.

I Anne-Grete Roer, Stig Morten Thorsen, Trond Rafoss, Marcel Van Oijen and Trygve Almøy. Fine-Tuning Bayesian calibration for complex systems with application to a snow depth model. *Submitted to Environmental Modelling & Software.*

II Stig Morten Thorsen, Anne-Grete Roer and Marcel Van Oijen. Modelling the dynamics of snow cover, soil frost and surface ice in Norwegian grasslands. *Polar Research.* In press.

III Stig Morten Thorsen and Mats Höglind. Modelling cold hardening and dehardening in timothy. Sensitivity analysis and Bayesian model comparison. *Submitted to Agricultural and Forest Meteorology.*

IV Stig Morten Thorsen and Mats Höglind. Assessing winter survival of forage grasses in Norway under future climate scenarios by simulating potential frost tolerance in combination with simpler agroclimatic indices. *Submitted to Agricultural and Forest Meteorology.*

2 Summary

For grass-based agriculture at high latitudes, poor overwintering of perennial forage grasses often has economic consequences due to yield loss and re-establishment of grass fields. In order to assess the performance of grass cultivars currently used in Norway under a future changing climate, a whole-year grassland model has been developed. The basis of this whole-year model was a grassland model developed for the growing season. In order to incorporate the winter season, this grassland model needed additional sub-models for simulating snow cover, soil frost, ice encasement and the development of frost tolerance in the plants. The main objective of this thesis has been to develop these additional sub-models, calibrate them using Bayesian methods and identifying key parameters using sensitivity analysis. The sub-models were also used to construct agroclimatic indices in order to assess the impact of climate change on the winter survival of two forage grasses.

There are several challenges emerging when applying Bayesian calibration to a dynamic model. The Bayesian approach regards parameters as random and allows integration of prior knowledge. Using the snow cover sub-model as case study, it is here demonstrated how prior information and new data affect the calibration process, parameters and model outputs, with focus on uncertainty. Point estimates and uncertainties are calculated and visualized for both parameters and model outputs. Generally, uncertainty decreased when new data were incorporated. Uniformly distributed priors gave the best fit for this model according to root mean square error, while the more informative beta distributed priors gave more physically meaningful parameter estimates. Markov chains of samples from the posterior distribution of the parameters were obtained by the random walk Metropolis-Hastings algorithm. Crucial points when using these methods are reaching and determining convergence of these chains. In order to reach convergence faster, informative priors, Sivia's likelihood, reflection and updating the proposal distribution with parts of the data gave successful results. To determine convergence objectively and cor-

rectly, the use of multiple chains and the Gelman Rubin method was found useful. Several decisions must be made when implementing Bayesian calibration, and we highlight and visualize the choices that were found to be most effective.

We developed a simple model SnowFrostIce which simulates depth of snow cover, the lower frost boundary of the soil and the freezing of surface puddles. We parameterised the model by means of Bayesian calibration, and identified important model parameters using the sensitivity analysis method of Morris. Verification of the model suggests that the results are reasonable. Due to the simple model structure, some overestimation occurs in snow and frost depth. Both the calibration and the sensitivity analysis suggested that the snow cover module could be simplified with respect to snow melt and liquid water content. The soil frost module should be kept unchanged, while the surface ice module should be changed when more detailed topographical data become available, such as better estimates of the fraction of the land area where puddles may form.

Timothy (*Phleum pratense* L.) is the most important forage grass in Scandinavia and it is therefore highly interesting to study how it will perform in a changing climate. In order to model winter survival, the dynamics of hardening and dehardening must be simulated with satisfactory precision. We investigated an early timothy frost tolerance model (LT50 model), and an LT50 model for winter wheat. Based on the assumption that timothy has no vernalization requirement, unlike winter wheat, but does have the ability to adapt to cold temperatures in a process linked to stage of development, two alternative versions of the winter wheat model were also constructed. These four models were calibrated by a Bayesian approach using observations on LT50 for the timothy cultivar Engmo. The models were validated using independent observations at different locations reflecting differences in climate. A sensitivity analysis using the Morris method to identify important model parameters suggested that there is a connection between frost tolerance and stage of plant development, even if there is no vernalization requirement. The simplified winter wheat

model was selected as the best candidate based on model selection criteria and its ability to capture the hardening and dehardening processes. The results from the Bayesian calibration suggests that there are no major regional differences in Norway calling for regional calibration. However, cultivar-specific calibration is probably required, since there are hardy and less hardy cultivars within the same species. A functional LT50 model would allow risk assessments to be made of future winter survival using specifically tailored and downscaled climate scenarios.

We assessed the impact of climate change on the winter survival of timothy (*Phleum pratense* L.), and perennial ryegrass (*Lolium perenne* L.) under Norwegian conditions using agroclimatic indices and a simulation model of frost tolerance. Available to this study was locally adjusted climate scenarios (two for the period 2071-2100; one for the period 2020-2049) for six important agricultural regions, represented by one location each. We proposed and validated a rough way to estimate the daily minimum air temperatures from scenario data. Compared to the control period 1961-1990, the hardening period would be shortened by up to 21 days. As a consequence the modelled maximum frost tolerance is expected to be reduced by up to 3.9℃ and 1.9℃ for timothy and perennial ryegrass, respectively, under the warmest scenario. In spite of this reduction in hardiness, the plants are expected to be hardy enough to withstand the predicted autumn frosts, and also we expect a general reduction in risk of winter frost injuries. The plant data available to this study suggests that the agroclimatic indices developed for Canadian conditions can be useful for assessing the hardening status in timothy and perennial ryegrass. They are, however, less suitable for assessing the risk of plant injury related to frost and ice encasement in Norway since the dynamics of cold adaptation is not accounted for by these indices. Although less snow is expected, this is in most cases not accompanied by an increase in the risk of ice encasement injuries. There is little risk of winter injuries related to frost and ice encasement in the hardier grass species timothy. The better overwintering conditions in general indicate that it will be possible to grow perennial ryegrass in areas where it is

not grown today, given that the risk of fungal diseases is not increased.

3 Sammendrag

For norsk grasbasert landbruk vil ofte dårlig overvintring av enga få økonomiske konsekvenser i form av tapt avling, eller omsåing. For å kunne studere hvordan ulike grassorter vil klare seg under forskjellige klimascenarioer er det utviklet en helårsmodell for simulering av grasvekst. Denne helårsmodellen er basert på en eksisterende grasmodell for vekstsesongen. For å kunne utvide denne sesongbaserte grasmodellen til også å omfatte vinteren, var det nødvendig å utvikle nye delmodeller til å simulere snødekke, teledyp og dannelse av isdekke samt en delmodell for plantenes evne til å utvikle frosttoleranse. Hovedmålet med denne avhandlingen har vært å utvikle disse delmodellene, kalibrere dem ved hjelp av Bayesianske metoder og identifisere de viktigste parameterne ved hjelp av sensitivitetsanalyse. Disse delmodellene har blitt brukt til å konstruere agroklimatiske indekser som i sin tur har blitt brukt til å vurdere effekten av et endret klima på vinteroverlevelsesevnen til to arter fôrgras.

Man støter på flere utfordringer når Bayesianske metoder skal brukes til å kalibrere en modell. I modellkalibrering ved bruk av den Bayesianske tilnærmingen betrakter man parametere som tilfeldige variable, og tillater integrering av tidligere (*a priori*) kunnskap om parameterne. Ved å bruke delmodellen for snødekke ble det demonstrert hvordan *a priori* kunnskap og nye observasjoner påvirket kalibreringsprosessen, og hvordan dette førte til at usikkerheten til både parameterne og modellresultatet ble påvirket. Generelt ble usikkerheten redusert når flere observasjoner ble benyttet. Uniforme *a priori* fordelinger til parameterne gav best modellresultat med hensyn på prediksjonsfeil (root mean squared error), mens mer informative beta *a priori* fordelinger gav mer fornuftige parameterverdier. Markov kjeder, som representerer et utvalg fra den estimerte *a posteriori* fordelingen til parameterne, ble fremskaffet ved å bruke Metropolis-Hastings algoritmen (random walk). Et nøkkelelement ved bruken av disse metodene er å avgjøre når Markov kjedene har konvergert.

For å oppnå forholdsvis hurtig konvergens brukte vi informative beta *a priori* fordelinger, Sivias sannsynlighetsfunksjon (likelihood) og refleksjon ved parameternes grenser. Det ble også gjort innledende forsøk med å bruke en oppdatert forslags-fordeling basert på et begrenset datasett; noe som viste seg å gi lovende resultater. For å avgjøre konvergens til Markov kjedene på en objektiv måte valgte vi å bruke flere kjeder samt Gelman Rubins metode. Når Bayesianske metoder skal brukes til modellkalibrering må flere beslutninger tas. Vi understreker og visualiserer de beslutningene vi fant som viste seg å være mest effektive.

Vi utviklet en enkel modell SnowFrostIce for å simulere snødekke, den nedre telegrensa i jorda og frysing av vann i overflatedammer. Denne modellen ble kalibrert ved hjelp av Bayesianske metoder. Vi brukte Morris' metode til å identifisere de viktigste parameterne. Verifisering av SnowFrostIce-modellen viste at modellresultatene var fornuftige. På grunn av modellens enkle form forekommer noe overestimering av både snø- og teledyp. Både kalibreringen og sensitivitetsanalysen antyder at snømodulen kan forenkles noe med hensyn på simuleringen av flytende vann i snølaget. Teledypmodulen kan i første omgang beholdes som den er. Modulen som simulerer isdekket bør endres når mer detaljerte opplysninger om topografi på jordoverflaten blir tilgjengelig. På denne måten kan bedre estimater av landområder som potensielt kan bli dekket av dammer oppnås.

Timotei (*Phleum pratense* L.) er det viktigste fôrgraset i Skandinavia, og derfor veldig interessant å studere med tanke på hvordan den vil klare seg under et endret klima. For å simulere frosttoleranse i timotei må både herding og avherding simuleres tilfredsstillende. Vi undersøkte en tidligere frosttoleransemodell (LT50 modell) for timotei og en LT50 modell for høsthvete. Basert på antakelsen om at timotei ikke har noe vernaliseringskrav, i motsetning til høsthvete, men allikevel har evnen til å tilpasse seg lave temperaturer som følge av utviklingsstadium, foreslo vi to alternative LT50 modeller for timotei basert på høsthvete modellen. Disse fire kandidatmodellene ble så kalibrert med den Bayesianske metoden der vi brukte observerte LT50 verdier til timoteisorten

Engmo. De fire kandidatmodellene ble validert mot observasjoner hentet på forskjellige steder med ulikt klima. Sensitivitetsanalysen av høsthvete modellen, som identifiserte de viktigste parameterne, indikerte en sammenheng mellom frosttoleranse og plantens utvikling selv om timotei ikke har vernalisering. En av de forenklede høsthvete modellene ble valgt som den beste kandidaten basert både på utvalgskriterier og dens evne til å fange herdings- og avherdingsprosessene. Resultatene fra den Bayesianske kalibreringen antydet at det ikke var noen regionale forskjeller i Norge som skulle tilsi at modellen må kalibreres lokalt. På den andre siden vil sortspesifikk kalibrering sannsynligvis være påkrevd siden det finnes både herdige og mindre herdige sorter innen samme art. Dersom lokalt nedskalerte klimascenarioer blir tilgjengelig kan LT50 modellen brukes til å vurdere fremtidig risiko for vinterrelaterte skader i eng.

Vi vurderte effekter av et endret klima på vinteroverlevelsesevnen til timotei (*Phleum pratense* L.) og flerårig raigras (*Lolium perenne* L.) for norske forhold ved beregning av agroklimatiske indekser og en simuleringsmodell for frosttoleranse. Tilgjengelige klimadata var lokalt tilpassede klimascenarioer (to for perioden 2071-2100; og ett for perioden 2020-2049) for seks viktige landbruksregioner. Vi foreslo og validerte også en enkel metode for å estimere døgnminimums temperatur ut fra scenario data. Sammenliknet med kontrollperioden 1961-1990, vil herdeperioden bli kortet inn med inntil 21 dager. En konsekvens av økt temperatur om høsten er at den simulerte maksimale frosttoleranse blir redusert med opp til 3.9℃ and 1.9℃ for henholdsvis timotei og flerårig raigras under det varmeste scenarioet. På tross av denne reduksjonen i frosttoleranse forventer vi at plantene oppnår tilstrekkelig med herdighet til å klare frost om høsten, samt at vi også forventer en generell reduksjon i risikoen for vinterrelaterte skader. De observasjonene på frosttoleranse som var tilgjengelige for denne studien antyder at agroklimatiske indekser utviklet for kanadiske forhold kan være nyttige for å anslå graden av herdighet til timotei og flerårig raigras. Disse kanadiske indeksene er imidlertid mindre egnet til å vurdere risiko under norske forhold for skader forårsaket av frost og innkapsling i is

siden dynamikken i herdingsforløpet ikke blir fanget opp. Selv om simuleringene antyder mindre snø under klimascenarioene er dette ikke etterfulgt av økt risiko for isdekkeskader. Det forventes lav risiko for skader relatert til frost og is-innkapsling for den herdige timotei planten. Generelt indikeres bedre overvintringsforhold, noe som kan gjøre det mulig å dyrke flerårig raigras i områder det ikke gjøres i dag. Det tas da forbehold om at risiko for biotiske skader (som sopp) ikke øker samtidig.

4 Introduction

4.1 Background

The global temperature is increasing, and accompanied by increasing amounts of CO2 in the atmosphere. Conclusions made by Olesen and Bindi (2002) suggests these changes will be positive for agriculture at high latitudes. Although the overall picture appears positive, there may still be some challenges for agriculture in these regions related to changed overwintering conditions (Bélanger *et al.* 2002). Also, plant diseases and many weeds are likely to become more serious constrains under warmer and a more humid climate; a consequence of which might be an increased use of fungicides and herbicides to avoid yield losses (Olesen and Bindi 2002). In Norway, the yearly average air temperature is expected to increase by 2,3 to 4,6℃ by the end of this century (Hansen-Bauer and Førland 2009). The largest temperature increase is expected during winter, and in the northernmost county, Finnmark, the temperature is expected to increase by 3 to 5,4℃ on a yearly basis. The smallest increase is expected along the western coast, where an increase is expected to lie between 1,9 and 4,2 ℃. The yearly average precipitation in Norway is expected to increase by 5 to 30% by the end of this century, specifically winter precipitation is expected to increase by up to 40% and more of the winter precipitation will fall as rain at the expense of snow. A consequence of the predicted increase in temperature and changed precipitation pattern during winter is that spring

floods occurs earlier, and the risk of floods during late fall and winter increases. For crops this means increased risk of water-logging (the soil is completely saturated with water, and water accumulates in puddles), and less snow cover to protect the plants from lethal subzero temperatures. These climate predictions are uncertain, and hence the effects of climate change on agriculture are more uncertain. Prediction models might serve as a useful tool to investigate the effects of climate change on agricultural systems.

4.2 Purpose and scope of this thesis

Some of the most important segments of Norwegian agriculture is grass-based milk and livestock production. The proportion of perennial forage grasses in Norwegian feed ratios are typically high. In 2007 these segments comprised 70% of the income in the Norwegian agriculture (Rognstad and Steinset 2008). Poor overwintering of perennial forage grasses often has economic consequences for these farm systems due to substantial yield loss and subsequent costs related to the re-establishment of grass fields. In this context, the prediction of overwintering of forage grasses under a changing climate is of great importance for Norwegian agriculture.

In order to study the ability of grass plants to withstand winter conditions like e.g. reduced photosynthesis due to snow cover, anaerobic conditions caused by ice encasement and periods of physiological drought we developed a model for their overwintering ability. This model was in turn linked to an existing model for the growing season resulting in a whole-year grassland model. This grassland model needed additional sub-models for simulating snow cover, soil frost, ice encasement and the development of frost hardiness in the plant.

The purpose of this thesis was to develop sub-models for snow cover, soil frost and soil surface ice cover in order to simulate these winter conditions. We also developed a model for frost tolerance in grasses in order to simulate the response of the plants to winter conditions. The philosophy behind the grassland model was to keep it as simple as possible since parameter rich models are

difficult to keep track of, and they easily can give right answers for the wrong reasons. The soil part of the grassland model consists of one layer, and this placed constrains on the choices we had to make, e.g. for the soil frost module. We therefore wanted to keep the sub-models developed during this work simple and robust, and with as few parameters as possible. The sub-models were to undergo both uncertainty and sensitivity analysis.

Models in the literature

Snow depth and soil frost models

In the literature there are numerous examples on approaches how to model the accumulation and ablation of snow. Models simulating snow cover are designed and implemented in a wide range of areas, which in turn dictates the allowed degree of complexity. In the simplest end of the complexity scale, we find degree-day index models (low complexity), in the mid range we find the snow models used in land surface schemes for general circulation models, and in the high end we find snow-physics models used for instance in avalanche forecasting. Rutter *et al.* (2009) presents a comprehensive evaluation of thirty-three snowpack models varying in complexity and purpose, with the intention to find a suitable candidate for modelling snow water equivalent (SWE) in forested areas. Melloh (1999) presents a review of seven selected one dimensional (considers only vertical direction) snow melt models used in operational model systems.

There are also quite a few methods developed for estimating soil freezing. From simple index-based methods for estimating maximum frost depth, to parameter rich state-of-the-art simulation models coupling heat and mass transport in soil. Kennedy and Sharratt (1998) compares the performance of four models soil frost models using two different approaches. Two models which are based on the finite difference method (SHAW by Flerchinger and Saxton (1989), and SOIL by Jansson (1991)) and two models which are based on balancing heat fluxes.

Models for frost tolerance in plants

In the literature there exist quite a few papers on the ability of plants to acclimate to freezing temperatures (often referred to as cold hardening) and the effect of freezing temperatures on plant survival. A review of current research on this topic is summarized by Kalberer *et al.* (2006). Only a few examples of models for frost tolerance in crops that can be grown in Norway today are available in the literature. Relevant models are the frost tolerance models for winter wheat presented by Bergjord *et al.* (2007); Fowler *et al.* (1999); Lecomte *et al.* (2003), a model for freezing injury in alfalfa presented by Kanneganti *et al.* (1998), and a model for the separate processes of cold hardening and dehardening in ryegrass is presented by Gay and Eagles (1991). I was unable to find any published work on models regarding frost tolerance in timothy; one of the most important grass species in Norwegian agriculture, and also the grass species for which the grassland model in the WINSUR project was developed together with perennial ryegrass.

4.3 Aspects of modelling

A model is called upon whenever a scientist wants to describe the interconnection of processes underlying a certain set of observations. The purpose could be for instance to *explain* underlying processes of the observations, or use the model for *predictions*. The purpose of the models presented in this thesis is mainly prediction (depths of snow cover and soil frost, occurrence of soil surface ice and frost tolerance in timothy) although especially the frost tolerance model can also be seen as the first step towards an explanatory model with sub-processes. One important issue to bear in mind, is that all models are simplifications of a limited part of reality (often referred to as a system), and can never represent the full spectrum of behavior of the real system (de Wit 1993). Thus, models fail if they are used beyond their scope. One of the

challenges in modelling is therefore to simplify wherever possible, and focus on the parts of reality that are essential for the problem at hand (Goudriaan and Van Laar 1994). Another important issue to consider when building a model is setting the boundaries between the real system, and the limited part we are currently interested in. These boundaries should sought to be selected in such a way that the environment might affect the system, but the system should affect the environment as little as possible. For example, in the SnowFrostIce model that was developed as part of this thesis (see Papers I and II), we use inputs from the environmental inputs (air temperature and precipitation) to model the change in depths of snow cover, soil frost and ice cover; but we are only considering one dimension (the vertical direction). These types of one dimensional models (other examples are COUP and SHAW) are therefore often referred to as point-scale models in hydrology literature. However we make the assumptions that the meteorological inputs (air temperature and precipitation) are homogeneous, and that there is no spatial variation at the model scale. The spacial scale of the model we consider is typically a part of a field where we assume homogeneous environmental conditions (i.e. soil conditions, snow cover, plants and climate). For the frost tolerance models, we consider a grass population at the same plot size as for the SnowFrostIce model.

When building models, it is considered good modelling practice to perform uncertainty and sensitivity analysis, followed by validation of the model using independent observations (Jakeman *et al.* 2006). These additional tasks in the modelling process often lead to suggestions on how to further develop and eventually simplify the original model. Together with the documentation of the original model and the steps taken to modify the original model, this comprises what can be regarded as the pedigree of the final model. This is the framework within which the models presented in this thesis was developed and documented.

4.4 Methods used in this thesis

In this thesis I have focused especially on the application of methods for performing calibration and sensitivity analysis of simulation models. To illustrate the methods of calibration and sensitivity analysis I have used, I have constructed an example snow model which is presented together with these methods below.

Description of Snow2par

This section presents the simple snow model constructed for visualizing the calibration and sensitivity methods employed in this thesis. The snow model consists of two parameters. The first parameter T_{ms} (℃) is the threshold temperature determining both precipitation form (rain or snow) and snow melt. It is assigned as prior distribution a normal distribution with zero mean and standard deviation of 3 ℃. The second parameter is the snow melt rate M (mm ℃-1 day-1). Since this is a so called scaling parameter, I followed the suggestion by Sivia (2006) and assigned as prior the Jeffrey's prior ($P \propto 1/M$). The prior range of the parameter M was set to (0.01, 10.0) mm ℃-1 day-1. The state variable in the model is the snow water equivalent SWE (mm) (the amount of liquid water obtained if the snow cover melted instantaneously). The observations used when calibrating the model Snow2par was snow depth (m) observations from Kise, Norway, the winter of 1988-1989. To relate SWE to snow cover depth, we used a common assumption regarding the water equivalent of falling snow of 10:1, i.e. 10 cm of fresh snow melts down to 1 cm of water (Judson and Doesken 2000). This corresponds to assuming a density of new snow of 100 kg m-3. A conceptual representation of Snow2par is presented in Figure 1.

Calibration of Snow2par

When faced with the challenge of calibrating a model which contains several parameters, and provides multiple outputs, the Bayesian framework is very

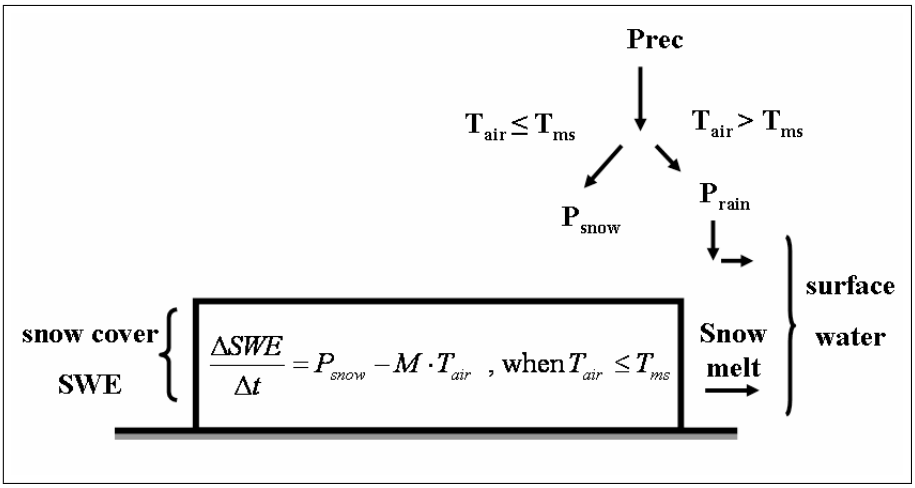


Figure 1: A conceptual representation of the snow model Snow2par

well suited. We consider the model parameters $\boldsymbol{\theta}$ not to be known precisely, and that this uncertainty can be represented as a probability distribution over the parameters $\pi(\boldsymbol{\theta})$. Given a dataset $\mathbf{D}$ on model output, we can use Bayesian calibration to update the distribution of the parameters $\boldsymbol{\theta}$ as new observations becomes available by applying Bayes' theorem: $\pi(\boldsymbol{\theta}|\mathbf{D}) = \pi(\boldsymbol{\theta})f(\mathbf{D}|\boldsymbol{\theta})/f(\mathbf{D})$, where $\pi(\boldsymbol{\theta}|\mathbf{D})$ is the posterior distribution of $\boldsymbol{\theta}$ given the data $\mathbf{D}$, $f(\mathbf{D}|\boldsymbol{\theta})$ is the likelihood of the data given the model outputs using parameters $\boldsymbol{\theta}$, and $f(\mathbf{D})$ is a normalisation constant (see Van Oijen *et al.* (2005) for a full description of this method). The parameters of the models presented in this thesis are all considered to be uncertain, as is the structure of the models. We use Bayesian calibration to reduce the uncertainty of the parameters, and thereby also the model output. To visualize the calibration method used, a simple snow model containing two parameters, called Snow2par was used (see Figure 1). Details regarding the implementation of this calibration method are presented in Papers I and II in this thesis.

Relevant information on model parameters (e.g. found from literature or previous studies) are considered as prior knowledge. Observations on snow

depth together with model output are evaluated through a likelihood function, and together with the prior information this forms the basis for a joint posterior probability distribution of the parameters. All this is achieved by using the Bayes theorem. The general idea used in this thesis when calibrating the models is to randomly walk through the multidimensional parameter space, eventually ending up at the highest peak indicating the region with the highest probability of the parameter values given our observations. This sampling technique is a Markov Chain Monte Carlo (MCMC) method known as the Metropolis-Hastings random walk. When the sample sequence eventually stabilizes, meaning that the resulting distribution of the sampled parameter values reach stationarity, the chain is said to have converged. This way of sampling the parameter space forms eventually a Markov chain (i.e. when the sequence is converged) since each new candidate point is selected as a multivariate random step away from the current point in parameter space. In all calculation steps in the implementation of the MCMC-algorithms we used logarithms to avoid numerical problems, as suggested by Van Oijen *et al.* (2005).

Figure 2 shows a 3D plot of the joint posterior distribution of the two parameters T_{ms} and M. In this figure, the joint posterior probability is calculated at each grid point with the intention to visualize the landscape map of the joint posterior distribution. Superimposed on the x-y-plane in this figure are the contours of the joint posterior probability indicating the region with the desired peak. Superimposed on the joint posterior surface are parameter values from each 50th step of the random walk from the calibration showing that the calibration eventually stabilized at the peak. Figure 3 shows histograms of the parameter values from the Markov chain obtained from the calibration. We can see that marginal posterior distributions in this constructed example was different from the prior distributions, suggesting the calibration procedure reduced the initial uncertainty for both parameters.

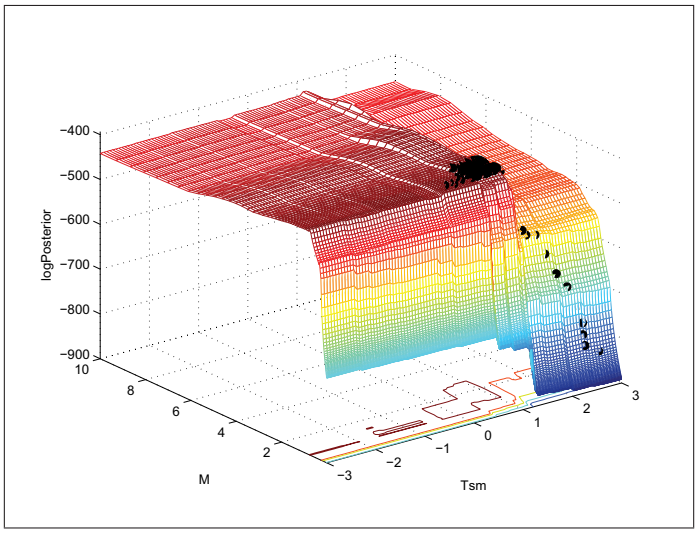


Figure 2: 3D-surface showing log-posterior calculated for the parameters T_{ms} for all combinations of parameter values. Superimposed as black dots are the log-posterior corresponding to each 50th step from the calibration of the model Snow2par. The limits for T_{ms} was set to $[-3, 3]$ for convenience.

Sensitivity analysis

When working with models, sensitivity analysis (hereafter referred to as SA) is recommended as part of the development process (Jakeman *et al.* 2006). A review of various SA methods is presented by Saltelli *et al.* (2006). For the SA to be meaningful, the modeller should decide beforehand on what type of question the SA is expected to answer (Saltelli *et al.* 2008). When developing models, they tend to rapidly increase in complexity, so as a candidate model evolves, the question of simplification often arises. In this thesis this was the question addressed to the models described in Papers II and III, i.e. we wanted to identify which parameters of the model were the important ones. A suitable SA method for this purpose is the screening method of Morris. The Morris method is designed to identify the important parameters within a complex

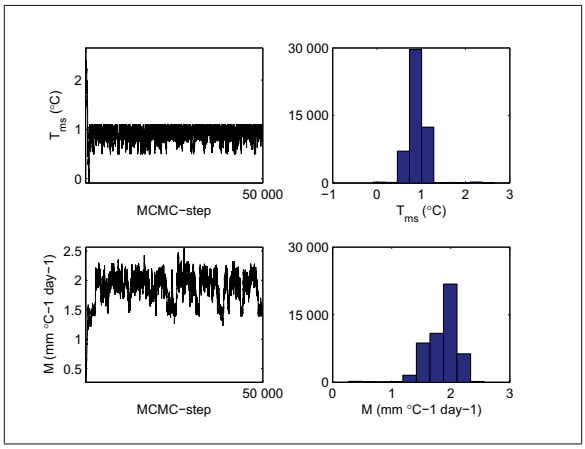


Figure 3: Trace plots from the calibration (left) and marginal posterior distributions for the temperature threshold parameter T_{ms} and the melt rate M from the calibration of the model Snow2par

model. The results from this SA then prepares the grounds for further model simplification by suggesting which parameters can be fixed anywhere within their prior bounds without affecting model outputs, i.e. which parameters are not important. Although this method is a One Parameter at a Time (OPT) approach, (Campolongo *et al.* 2007) empirically showed this method to be an acceptable proxy for a more sophisticated variance-based sensitivity measure. Despite it's simplicity, the Morris method is still seldomly used (ibid) by the modelling community.

The Morris method proposes two sensitivity measures, whose main purpose is to determine which of the parameters of the model that can be considered to be either (i) not important, (ii) linear and additive, or (iii) non-linear or involved in interactions with other parameters. For each of the parameters, two sensitivity measures are computed; μ, which evaluates the overall influence of the parameter on the model output (main effect, or elementary effect EE), and σ, which evaluates collectively all the higher order effects due to non-linearity and/or to interaction with other parameters.

The Morris method was originally used for parameters following uniform distributions in $[0, 1]$. However, if parameters follow other distributions, Campolongo *et al.* (1999) suggest that rather than sampling directly from these distributions, the sampling should be performed in the space of the quantiles of these distributions (e.g. each parameter distribution is discretized into p levels, were each quantile q_p varies in $[0, 1]$). The actual parameter values would subsequently be derived from their known distributions. These quantiles will then comprise the sub-set Ω from which parameter sets $\boldsymbol{\theta}$ are sampled, i.e. $\boldsymbol{\theta} \in \Omega$. For a model with k parameters, the elementary effect of parameter θ_i is defined as

$$EE_i(\boldsymbol{\theta}) = \frac{f(\theta_1, \ldots, \theta_i + \Delta, \ldots, \theta_k) - f(\boldsymbol{\theta})}{\Delta} \tag{1}$$

where $f(\boldsymbol{\theta})$ is a scalar representation of model output, Δ is the sampling step, and $\boldsymbol{\theta}$ is a point in Ω which is chosen such that $(\boldsymbol{\theta} + \boldsymbol{e}_i\Delta)$ remains within Ω, and $\boldsymbol{e}_i$ is a vector of zeros except for its ith element having the value one. What follows is when sampling random parameter sets $\boldsymbol{\theta}$ from Ω according to the Morris method, we obtain the distribution associated with the elementary effect of each parameter. This distribution is finite and can be denoted as $EE_i(\boldsymbol{\theta}) \sim D_i$. As suggested by Campolongo *et al.* (2007), a convenient choice for p is an even number, and a convenient choice for the sampling increment Δ is setting $\Delta = p/\left[2(p-1)\right]$. The number of elements in D_i is $p^{k-1}\left[p - \Delta(p-1)\right]$ (ibid.). The sensitivity indices for parameter θ_i from the Morris method is the mean μ and standard deviation σ of D_i. The design proposed by Morris (1991) is to sample r elementary effects from each D_i by constructing r trajectories of $k+1$ parameter sets from Ω, where each trajectory provides k elementary effects (one elementary effect for each parameter). A detailed description of the implementation of the Morris method is presented in Saltelli *et al.* (2008).

In this example, the parameter values were sampled from $p = 6$ equidistant quantiles of their prior distributions, and using the increment $\Delta = 3/5$ yields the following subset Ω of parameter values:
$T_{ms} \in \{-9.7, -2.7, -0.8, 0.7, 2.4, 10.7\}$ and

$M \in \{0.01, 0.045, 0.17, 0.61, 2.65, 10.0\}$. We sampled three trajectories in the example with the Snow2par model (see Figure 4). Outputs from Snow2par

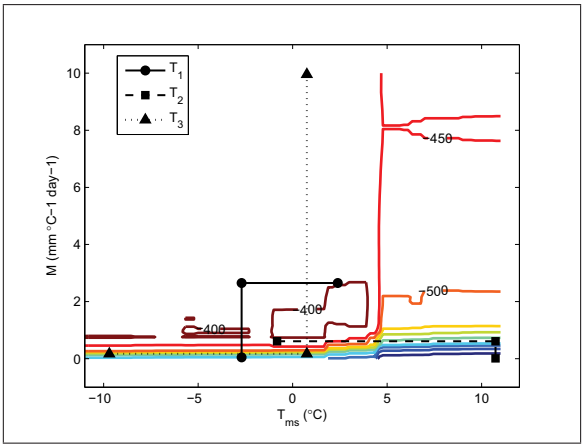


Figure 4: Trajectories obtained via the Morris method. The parameter values are sampled in six equidistant quantiles from their prior distributions.The temperature parameter is assumed to have a normal prior, i.e. $\pi(T_{ms}) \sim N(0, 3)$; the snow melt index parameter is assumed to follow Jeffreys' prior, i.e. $\pi(M) = \frac{1}{M}$

are time series, and the Morris method requires a scalar output value for the SA. Thus, for the simulation runs required, we used as scalar output the log-transformed likelihood from the calibration; the likelihood being the probability of the observed data given a certain model parameterisation $\boldsymbol{\theta}$. If the model is non-monotonic, some elements of D_i may have opposite signs. This may cause elementary effects to cancel each other out resulting in a μ-value close to zero. To remedy this Campolongo *et al.* (2007) suggests an alternative sensitivity index, called μ^*, which is the mean of the absolute values of D_i. The three trajectories used in this example are shown in Figure 4 where they are superimposed on a contour plot of the posterior distribution obtained from the calibration. The sensitivity indices μ^* and σ are interpreted as follows: a high σ_i value for parameter θ_i implies that the corresponding EE_i-value for θ_i

at one point in Ω is considerably different from another EE_j-value ($i \neq j$) for the same parameter θ_i located somewhere else in Ω, i.e. that this particular EE-value is influenced by the values of the other parameters or nonlinearities. A low value for σ_i suggests that the EE_i-value associated with θ_i is less influenced by the values of the other parameters and thus it is not involved in interactions or nonlinearities.

Since the example model Snow2par only has two parameters (which are equally important) the results from the sensitivity analysis presented in Papers II and III are more interesting due to more complex models.

5 Summary of included papers

The outline of the papers is as follows. Paper I describes the calibration methods we used, highlighting some implementing issues where the literature was scarce. Solutions to these issues are suggested. This calibration method is further used throughout this thesis. The case study in Paper I is a snow cover model which is calibrated for one location. In paper II we describe the development of the SnowFrostIce model. This model contains a further development of the snow model investigated in paper I and includes new modules to simulate soil frost and surface ice cover. In addition to calibrate the model, we perform a sensitivity analysis and describe the method used. In paper III we compare four candidate models simulating frost tolerance in the forage grass timothy. The SnowFrostIce model is an essential part of all these frost tolerance models. We used Bayesian methods for both calibration and model comparison. In paper IV we used the best frost tolerance model together with the SnowFrostIce model to develop agroclimatic indices used to assess the impact of climate change scenarios on Norwegian agriculture.

5.1 Paper I

The objective in paper I was to explore challenges emerging when applying Bayesian methods to calibrate complex models. Bayesian methods are becoming increasingly popular as means of both calibrating complex models and comparing the performance of different candidate models for specific data set. When implementing Bayesian calibration methods we encountered challenges not described properly in the literature. Paper I describes these challenges and proposes a set of solutions. The case study in this paper was a model which simulates accumulation and melting of a snow cover. This snow model is a combination of a previous snow depth model proposed by Riley and Bonesmo (2005) and a snow water equivalent (SWE) model proposed by Sorteberg *et al.* (2001). In the Bayesian calibration, model parameters are described by distributions rather than estimates of true values. Instead of searching for the best possible parameter values, we search for the probability distribution of these parameters which reduce model uncertainty.

The effect of adding more observations was a reduction in uncertainty of both parameter estimates and model output. On the other hand, reaching convergence became increasingly difficult when adding more data.

Using informative priors (beta-distributions) compared to uniform priors, resulted in more realistic parameter values and faster convergence of the chains. Using uniform priors gave on the other hand better model fit. We used beta-distributions for the parameters found in the literature.

We investigated the use of Gaussian distribution and a uniform distribution as proposal distribution in the Metropolis Hastings random walk. The Gaussian was most efficient, however, it's covariance matrix needed adjustments by trial and error to obtain the desired acceptance rate. We successfully used this sampling algorithm with reflection at parameter boundaries to avoid sampling outside the prior intervals. The sampling with reflection was used when calibrating all models.

We investigated both a Gaussian likelihood function, and Sivia's constrained

Gaussian likelihood function. The latter gave faster convergence, and this was therefore used in the following works.

5.2 Paper II

In paper II we present the simple winter model SnowFrostIce, constructed with the intention (and thereby also accompanying restrictions) to be incorporated in a larger grassland model designed to simulate the regrowth dynamics and winter survival of forage grasses in Norway. The SnowFrostIce model simulates the depth of snow cover, the lower frost boundary of the soil and the freezing of surface puddles. The snow depth module in SnowFrostIce is the same snow model as the case study in Paper I, whereas the modules for soil frost and soil surface ice are new additions.

The snow module simulates snow accumulation based on a threshold temperature (deciding whether precipitation should be simulated as snow or rain) and a calibrated parameter describing the density of falling new snow. Snow melt is simulated based on a degree-day temperature melting index K which is described by a sinusoidal curve. The reason for predetermining K in this way is to incorporate the seasonal variation. Incoming radiation increases and albedo decreases in spring, and since Norway is located at a high latitude, this sinusoidal is defined as having period of one year with trough and crest coinciding with the dates of the solstice.

The lower frost boundary is estimated using an energy balance approach. The soil frost module relies heavily on certain key assumptions. First, following Benoit and Mostaghimi (1985), we assume that the daily mean *soil surface* temperature on snow free days can be estimated by the daily mean *air* temperature at 2 m. Secondly, when simulating the soil surface temperature, we assumed a constant ratio of the thermal conductivities of frozen soil and the the snow cover, respectively (Jansson and Karlberg 2001). Thirdly we assume a linear variation in soil temperature with respect to soil depth in the frozen soil layer, and that all available soil water within this frozen layer freezes.

The SnowFrostIce model is one-dimensional, meaning spatial variation is not considered. We assume that within the grid area where the model operates, a fraction of this area might be flooded with water if the soil conditions prohibit infiltration. The model does not simulate hydraulic properties of the soil, so the only situations when infiltration will be prohibited is when an adequately thick frozen soil layer is simulated (the implementation of the soil water balance of SnowFrostIce is described in Höglind *et al.* (2001)). We assumed this limit to be 20 cm, i.e. liquid water from rainfall or snow melt is drained as long as the simulated frost depth is less than 20 cm. This assumption is supported by observations Iwata *et al.* (2008). When a surface puddle is formed, the water may freeze and form a basal ice layer. By regarding this puddle as an extremely dilute soil, and setting the water content to unity, we use the same approach to estimate the thickness of this ice layer as we did for estimating soil frost depth.

The available data set containing observations on snow depth and lower frost boundary was divided in two parts; one part used for calibration and one part used for validation. We had data at four locations available for calibration and validation: Kise (6 years of data), Kvithamar (four years of data), Vågønes (five years of data) and Holt (five years of data). We calibrated SnowFrostIce using observations on snow depth and lower frost boundary. The frost boundary was measured using a frost tube. Observations on surface ice was scarce, and data were only available for two locations. Based on literature review we defined upper and lower boundaries and mode value for the nine model parameters. For parameters where both range and mode value were suggested, we used a beta distribution as prior. Otherwise we used a uniform distribution as prior. We used the likelihood function suggested by Sivia (2006) which was slightly modified to account for model discrepancy. Also, as in Paper I, we used a Gaussian proposal distribution where the covariance matrix was adjusted by trial and error to obtain the desired acceptance rate. Convergence of the MCMC was determined by running two chains in parallel and calculating the scale reduction factor suggested by Gelman and Rubin (1992). Model

uncertainty was assessed by sampling 20 parameter sets from the posterior distribution and calculating and plotting the subsequent means and standard deviations of the model outputs. The resulting marginal posterior distributions for the parameters were either multi-modal, skewed or both. For most of the parameters, their initial uncertainty was reduced. This is shown by comparing the marginal posterior to the respective prior distributions. For the parameters related to liquid water within the snow cover, it was evident that snow depth observations alone was not adequate for reducing their initial parameter uncertainty. It was also difficult to obtain convergence of the Markov chains for these parameters.

We conducted a sensitivity analysis (SA) of the SnowFrostIce model to identify the least important parameters. For this task we used the simple yet efficient method originally proposed by Morris (1991) and improved by Campolongo *et al.* (2007). We did not find suggestions in literature on how to handle correlated parameters when using the Morris method, and therefor parameter sets were sampled from the assumed uncorrelated prior distributions. The SA was performed at the same locations used in the calibration. Results from the sensitivity analysis shows that the least important parameters are those related to liquid water within the snow cover. This implies that varying these parameters within their prior bounds will not affect model output markedly. These results are in accordance with the results from the calibration, in the sense that their prior uncertainty was not much reduced, implying that no new information was added through the observations.

5.3 Paper III

In paper III we incorporated the SnowFrostIce model in four empirical candidate models for the estimation of frost tolerance in forage grasses. These candidate models were subjected to calibration and sensitivity analysis using the methods outlined in this thesis. Results from the calibration was also used for model selection, as one of three model selection criteria.

Based on model selection criteria in combination with visual assessment of model performance, a candidate model was selected. This candidate model was developed from a model predicting frost tolerance in winter wheat. Results from the calibration suggests that there are no major regional differences in Norway calling for regional calibration. This is verified in the validation plots where the same parameterization $\boldsymbol{\theta}^{\mathrm{MAP}}$ (the posterior distribution) was used at all locations. Therefore the parameter distribution obtained from the multi-site calibration can be used for our specific grass species. The model most likely needs to be calibrated for each grass cultivar separately, since there are both hardy and less hardy cultivars within the same species, but this assumption has to be verified in further modelling studies.

5.4 Paper IV

Climate plays a very important role in agriculture because it directly influence production. In Norwegian agriculture, milk and livestock production are amongst the most important segments. Poor over wintering of forage grass therefore represents a potential for economic loss both in relation to yield and to reestablishment of fields with dead plants. The SnowFrostIce model from Paper II, in combination with the frost tolerance model from Paper III was used to develop agroclimatic indices designed to assess the impact of three future climate change scenarios on the development of frost tolerance in two important forage grasses. The most important results for Norway under these scenarios was a delay in the hardening period, and earlier start of growth in spring. This means a prolonged thermal growing season. Additional harvests and cultivation of more productive but less winter hardy crops are possible outcomes. However, some coastal locations may experience a slightly increased risk of frost injury after start of growth in spring which is related to an earlier onset of growth in spring. A slightly increased risk of ice encasement injury may also be expected at some locations where the length of the snow period is dramatically decreased. However, the general trend is for a reduced risk of

plant injury related to ice encasement. The results suggests agroclimatic indices can be useful supplements when assessing the hardening status of forage grasses.

5.5 Concluding remarks and suggestions for continued work

Calibration:
In this thesis predictive models have been developed. The level of complexity of these models was kept at an absolute minimum because they were designed to be incorporated in a larger grassland model, and because over-parameterized models easily can give the right answers for the wrong reasons. The models can therefore be characterized as being empirical. This implies that few parameters are directly measurable in experiments, and therefore literature values for the parameters are non-existent, or at least hard to come across. In situations like this the Bayesian method of calibrating complex models enables us to reduce the uncertainty of the model by using observations on state variables in combination with limited prior knowledge on model parameters.

One way of obtaining convergence faster is to use part of the observation data set in a preliminary investigation to obtain an effective proposal distribution. In this way it is possible to construct a full covariance matrix, and not only use a diagonal matrix as we did with the Metropolis Hastings random walk with reflection. If we use a full covariance matrix, then the steps taken during the MCMC sampling will not have the symmetry properties required when using the reflection at the boundaries, therefore we must use the standard Metropolis Hastings without reflection.

Sensitivity analysis:
In this work we implemented and used one simple yet effective method for sensitivity analysis in parallel with the calibration. In future applications we will perform the sensitivity analysis at an earlier stage, i.e. before calibrating

the models. This is also one of the reasons why perform a sensitivity analysis. It would be interesting to compare results obtained by first doing sensitivity analysis, secondly calibrate the model; to an alternative approach of using part of the available data to design a proposal distribution to be used in the calibration. Which of these two paths will converge to the desired posterior first, and which path leads to the largest reduction in model uncertainty are interesting questions.

It would also be interesting to investigate other sensitivity analysis methods, i.e. based on variance decomposition. These methods are more expensive to perform (in terms of model runs) but this is affordable if the CPU time of one model run is low ($<$ 1 min.). They are on the other hand not as straight forward to implement as the Morris method. One option is to investigate the applicability of existing available implementations, e.g. the SimLab framework (http://simlab.jrc.it/).

Model limitations:
It is tempting to continue developing the physical model SnowFrostIce with respect to all it's constituents. The grassland model sets the premises for any sub-model to be incorporated. Regarding the simulation time step (which is one day in the grassland model) the sub-models should not operate using a shorter time step, i.e. the processes with diurnal variation e.g. those related to snow cover, soil frost and surface ice still must be approximated by daily averages.

Further restrictions in the current grassland model is the soil layer. The grassland model has one soil layer (thickness of 60 cm). This relates to the depth of the rooted zone of a grassland field. This means that the soil water balance, which is based on a tipping bucket approach, considers the average soil water content within this single soil layer. With regards to modelling soil frost, which is directly related to the soil water content, we simulate the dynamics of the lower frost boundary based on an algebraic expression using this average soil water content. If the simulated frost boundary exceeds the rooted

zone, we assume the soil water content at this depth is at field capacity. Other models, like e.g. COUP (Jansson and Karlberg 2001) and SHAW (Flerchinger and Saxton 1989) uses a finite element approach where the soil profile is divided into layers according to the user, and soil frost is simulated with the vertical resolution of these layers.

Simulating surface ice cover requires knowledge on field geometry. This kind of information was not available for this project. The fractional area of a field where surface puddles may form is a kind of information that can be interpreted as field geometry. The literature might hold references to surface runoff models that could be useful here. When we can simulate ice cover in these puddle areas in more detail it is also possible to make better estimations of lost yield via the grassland model.

Snow cover:
When simulating snow cover depth, the density of the falling new snow is important. Also estimation of the density of the snow cover is needed. Both these snow related densities are difficult to cope with (since they are both influenced by several climatic factors), yet in simple, empirical models they are important. If the simulated output was snow water equivalence (SWE), on the other hand, the density of falling new snow might not be necessary. The standard meteorological variable of precipitation rate would suffice. However, SWE is not a common variable to measure at meteorological stations, and therefore the data available for calibration at relevant agricultural locations would be scarce. Results from the sensitivity analysis identified parameters related to liquid water within the snow cover as being less important. These same non-important parameters were the ones for which convergence was most difficult to obtain, and they had the least reduction in parameter uncertainty. SWE is frequently used as output variable in hydrological models, but we did not have access to measurement data gathered at locations which also recorded temperature and precipitation. An interesting approach to simulate snow cover depth would be to incorporate a functional description of the density of falling new

snow, based not only on daily average temperatures but also incorporate daily minimum and maximum temperatures.

Soil surface temperature:
The soil surface temperature turned out to be an important variable. It is used as driving temperature both in the soil frost dynamics and in the frost tolerance model to simulate the temperature around the plant when a snow cover is present When simulating soil surface temperature in SnowFrostIce we assume constant values for the average thermal conductivities of both the frozen soil layer and the snow cover; only the first of these thermal conductivities was calibrated. The reason for this was that we had only observations on the lower frost boundary and not observations that could be linked to the thermal properties of the snow cover. A step to improve this approach could be to include an average thermal conductivity of the snow cover in the calibration. For this to be effective, observations on the soil surface temperature is required. Observations on soil surface temperature are not performed today, and since it is an important climatic variable for crops it should be incorporated in the automated meteorology services of Bioforsk (http://lmt.bioforsk.no).

Soil frost:
The literature has numerous examples of methods for simulating soil frost dynamics. Thawing of soil frost in spring takes place both from the bottom of the frozen soil layer, and from the soil surface, however the thawing from above is limited as long as a snow cover is present. The upper layers of the soil has the highest root density. Also, freezing of soil water is experienced by plants as drought. This is a motivation to investigate other simple approaches to simulate soil frost dynamics. An alternative approach to simulate soil frost on a daily time scale could be to estimate daily soil temperature profiles and identifying the depth where the soil temperature is 0℃.

Frost tolerance:

The empirical frost tolerance model for timothy showed very promising results following multi-site calibration. For use with other cultivars of timothy, the model is likely to need re-calibration since there are both hardy and less hardy cultivars. We will investigate the effect of linking the calculation of hardening rate to the availability of carbohydrates within the plants, and also alternative methods of relating the hardening rate to plant development.

Bibliography

Bélanger, G., Rochette, P., Gastonguay, Y., Bootsma, A., Mongrain, D., and Ryan, D., 2002. Climate change and winter survival of perennial forage crops in eastern canada. *Agron. J.* 94, 1120–1130.

Benoit, G. and Mostaghimi, S., 1985. Modeling soil frost depth under three tillage systems. *Transactions of the ASAE* 28, 1499–1505.

Bergjord, A., Bonesmo, H., and Skjelvåg, A., 2007. Modelling the course of frost tolerance in winter wheat i. model development. *European Journal of Agronomy* 28, 321 – 330.

Campolongo, F., Cariboni, J., and Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software* 22, 1509–1518.

Campolongo, F., Tarantola, S., and Saltelli, A., 1999. Tackling quantitatively large dimensionality problems. *Computer Physics Communications* 117, 75–85.

de Wit, C., 1993. *On Systems Analysis and Simulation of Ecological Processes*, *Current Issues in Production Ecology*, volume 4, chapter 1, 318. Kluwer Academic Publishers, Dordrecht, 2 edition.

Flerchinger, G. and Saxton, K., 1989. Simultaneous heat and water model of a freezing snow-residue-soil system i. theory and development. *Transactions of the ASAE* 32, 565–571.

Fowler, D., Limin, A., and Ritchie, J., 1999. Low-temperature tolerance in cereals: model and genetic interpretation. *Crop Science* 39, 626 – 633.

Gay, A. and Eagles, C., 1991. Quantitative analysis of cold hardening and dehardening in *Lolium*. *Annals of Botany* 67, 339–345.

Gelman, A. and Rubin, D., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7, 457–511.

Goudriaan, J. and Van Laar, H., 1994. *Modelling Potential Crop Growth Processes*, *Current Issues in Production Ecology*, volume 2. Kluwer Academic Publishers, Dordrecht.

Hansen-Bauer, I. and Førland, E., 2009. Klima i norge 2100. *Norsk offentlig utredning* 1, 1–136.

Höglind, M., Schapendonk, A., and Van Oijen, M., 2001. Timothy growth in scandinavia: Combining quantitative information and simulation modelling. *New Phytologist* 151, 355–367.

Iwata, Y., Hayashi, M., and Hirota, T., 2008. Comparison of snowmelt infiltration under different soil-freezing conditions influenced by snow cover. *Vadose Zone Journal* 7, 79–86.

Jakeman, A., Letcher, R., and Norton, J., 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software* 21, 602–614.

Jansson, P., 1991. Soil water and heat model technical description. Report 165, Swedish University of Agricultural Science, Dept. of Soil Science, P.O. Box 7014, S-750 07 Uppsala, Sweden.

Jansson, P. and Karlberg, L., 2001. *Coupled heat and mass transfer model for soil-plant-atmosphere systems*. Royal Institute of Technology, Dept. of Civil and Environmental Engineering, Stockholm, Sweden. Internet: ftp://www.lwr.kth.se/CoupModel/CoupModel.pdf.

Judson, A. and Doesken, N., 2000. Density of freshly fallen snow in the central rocky mountains. *Bull. Amer. Meteor. Soc.* 81, 1577–1587.

Kalberer, S., Wisniewski, M., and Arora, R., 2006. Deacclimation and reacclimation of cold-hardy plants: Current understanding and emerging concepts. *Plant Science* 171, 3–16.

Kanneganti, V., Rotz, C., and Walgenbach, R., 1998. Modeling freezing injury in alfalfa to calculate forage yield: I. model development and sensitivity analysis. *Agronomy Journal* 90, 687–697.

Kennedy, I. and Sharratt, B., 1998. Model comparisons to simulate soil frost depth. *Soil Science* 163, 636–645.

Lecomte, C., Giraud, A., and Aubert, V., 2003. Testing a predicting model for frost resistance of winter wheat under natural conditions. *Agronomie* 23, 51–66.

Melloh, R., 1999. A synopsis and comparison of selected snowmelt algorithms. Report 99-8, U.S. Army Cold Regions Research and Engineering Laboratory.

Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33, 161–174.

Olesen, J. and Bindi, M., 2002. Consequences of climate change for european agricultural productivity, land use and policy. *European Journal of Agronomy* 16, 239–262.

Riley, H. and Bonesmo, H., 2005. Modelling of snow and freeze-thaw cycles in the eu-rotate n decision support system. Report Planteforsk Grønn Kunnskap 112, Bioforsk. Http://www.bioforsk.no/dok/senter/adm/gke/gke_9_112_bonesmo.pdf.

Rognstad, O. and Steinset, T., 2008. Landbruket i norge 2007 (in norwegian). *Statistiske analyser SSB* .

Rutter, N., Essery, R., Pomeroy, J., and other authors, 2009. Evaluation of forest snow process models (snowmip2). *Journal of Geophysical Research* 114.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., 2008. *GLOBAL SENSITIVITY ANALYSIS. The Primer*. Wiley & Sons Ltd, West Sussex, England.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F., 2006. Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety* 91, 1109–1125.

Sivia, D., 2006. *Data Analysis: A bayesian tutorial*. Oxford University Press, Oxford, England, 2 edition.

Sorteberg, H., Engeset, R., and Udnæs, H., 2001. A national network for snow monitoring in norway: Snow pillow verification using observations and models. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial And Planetary Science* 26, 723–729.

Van Oijen, M., Rougier, J., and Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology* 25, 915–927.

Paper I

Anne-Grete Roer, Stig Morten Thorsen, Trond Rafoss, Marcel Van Oijen and Trygve Almøy

Fine-Tuning Bayesian calibration for complex systems with application to a snow depth model.

Submitted to Environmental Modelling & Software.

Fine-tuning Bayesian calibration for complex systems with application to a snow depth model

Anne-Grete Roer*[ab], Stig Morten Thorsen[cd], Trond Rafoss[a], Marcel van Oijen[e], Trygve Almøy[b]

[a] Norwegian Institute for Agricultural and Environmental Research, Plant Health and Plant Protection Division. Høgskoleveien 7, N-1432 Ås, Norway

[b] Norwegian University of Life Science, Department of Chemistry, Biotechnology and Food Science, N-1432 Ås, Norway

[c] Norwegian Institute for Agricultural and Environmental Research, Grassland and Landscape Division. Postvegen 213, N-4353 Klepp st., Norway

[d] Norwegian University of Life Science, Department of Mathematical Science and Technology, N-1432 Ås, Norway

[e] CEH-Edinburgh, Bush Estate. Penicuick, EH26 0QB, UK

*Corresponding author. Tel.: +47 922 83 427; fax: +47 649 46 110.

Email address: anne-grete.roer@bioforsk.no, (Anne-Grete Roer)

Abstract

This paper explores the challenges emerging when applying Bayesian calibration to a complex deterministic dynamic model. The Bayesian approach regards parameters as random and allows integration of prior knowledge. It is here demonstrated how prior information and new data affect the calibration process, parameters and model outputs, with focus on uncertainty. Point estimates and uncertainties are calculated and visualized for both parameters and model outputs. Generally, uncertainty decreased when new data were incorporated. Uniformly distributed priors gave the best fit for this model according to root mean square error, while the more informative beta distributed priors gave more physically meaningful parameter estimates. Markov chains of samples from the posterior distribution of

the parameters were obtained by the random walk Metropolis algorithm. Crucial points when using these methods are reaching and determining convergence of the chains. In order to reach convergence faster, informative priors, Sivia's likelihood, reflection and updating the proposal distribution with parts of the data gave successful results. To determine convergence objectively and correctly, the use of multiple chains and the Gelman Rubin method was found useful. Several decisions must be made when implementing Bayesian calibration, and we highlight and visualize the choices that were found to be most effective.

Key words: convergence diagnostics, model uncertainty, parameter estimates, parameter uncertainty, random walk Metropolis

INTRODUCTION

The potential effects of climate change on Norwegian agriculture are studied in the Norwegian Research program WINSUR (winter survival). A primary goal is to predict the impact of climate change on winter wheat and grass by making climate scenario driven plant growth models. Van Oijen et al. (2005b) developed a plant model for timothy and perennial ryegrass to forecast winter climate impacts on forage crops. Motivated by the need for daily information about snow depth, which is an important climate factor for winter survival of perennial plants, a model is built for predicting the not-commonly-measured variable snow depth, based on two commonly-measured variables, air temperature and precipitation (Thorsen and Haugen 2007). The model is based on a model computing the snow water equivalent developed by Vehvilainene (1992) and the parametrization is based on previous modeling work of Riley and Bonesmo (2005) for a site located at Bioforsk Arable Crops Division, Kise, Norway.

Our snow model is an example of a large category of environmental models, which are deterministic and dynamic and aim to represent the processes underlying the behaviour of the system. The processes are defined by differential equations which the model solves by – computationally demanding - numerical simulation. No such environmental model simulates the real world system perfectly, but still predictions are often made conditional on the model being correct. Predictions related to climate change will not be tested until several decades or maybe a century ahead. It is therefore important to provide decision makers with predictions that are transparent with respect to uncertainty (Thyer et al. 2009). There are three major

sources of uncertainty related to any model (Goldstein and Rougier 2006): (1) the model contains parameters whose values are not certain, (2) the model is an imperfect analog of the system and (3) the collected data contain measurement error. In this paper, we shall focus on uncertainty derived from (1) and (3). Our approach is that of Bayesian calibration (Van Oijen et al. 2005a) which unifies the two goals of model parameterization and uncertainty quantification. Uncertainty with respect to model structure (2) can be addressed in this framework as well, provided multiple models of the same system are available, but we do not carry out Bayesian model comparison in this paper.

In practice, parameter values of environmental models are either inferred from the literature or found by trial and error when little information is available. Calibration is the process of finding the best parameter estimate for the model using data from the system. Maximum likelihood (Miller and Miller 1999) is a well used traditional calibration routine that maximizes the probability of the data given the parameters, $f(\boldsymbol{x}|\boldsymbol{\theta})$. Limitations of the maximum likelihood approach are that uncertainties can not readily be quantified and conclusions made by the modeler are conditional on the model being correct. An alternative Bayesian approach is more rarely used for complex models (Van Oijen et al. 2005a, Hue et al. 2008, Luo et al. 2009), partly because of practical problems addressed in the present paper. A key issue is the computationally demanding numerical solution of differential equations, which limits the number of model evaluations for calibration that is feasible. Despite these computational problems, the application of the Bayesian method to environmental models has been increasing in recent years (e.g. Reinds et al. 2008, Lehuger et al. 2009) because it improves on the traditional approach by automatically including uncertainty quantification (Campbell 2006). It also allows for prior information about the parameters and conclusions are made conditional on the data.

Much pioneering work on the Bayesian calibration of environmental models has been carried out in hydrology (e.g. Kavetski et al. 2006, Smith and Marshall 2008, Thyer et al. 2009), often for stochastic models of water flow in response to precipitation. However, experience with the approach for the slow environmental models is sill limited and, as Campbell (2006) states in a recent review of calibration of computer simulators, “much work is still to be done to place calibration on a sound and practical statistical footing”.

The main objective for this paper has been to apply Bayesian calibration to a complex model in order to explore practical problems with the calibration as well as work out solutions. Point

estimates are calculated and uncertainties visualized for both the parameters and the model outputs. In order to obtain convergence of the chains (and thus reasonable results) by simulation in a limited amount of time, the usefulness of informative priors, Sivia's constrained likelihood, the reflection method and different proposal distributions including optimizing the proposal distribution with parts of the data are tested. To detect the state of convergence, we have checked the usefulness of multiple chains and Gelman-Rubin.

SNOW DEPTH MODEL

The SnowFrost model, described in detail in Thorsen and Haugen (2007) is a one dimensional model which simulates the dynamics of depth of snow cover S_{depth} (m) and soil frost F_{depth} (m). SnowFrost is integrated in a grassland model which simulates the regrowth dynamics of timothy (*Phleum pratense* L.). This grassland model by Van Oijen et al. (2005a) is under further development. There are two main modules in SnowFrost; one module relates to the dynamics of the snow cover, and one module relates to the formation of soil frost. In SnowFrost, the formation of soil frost is affected by the presence of a snow cover, but the snow cover is not affected by the presence of soil frost. In this paper the focus is on the snow module (Figure 1) and the calibration of its parameters, and thus we leave out issues related to soil frost. Preliminary calibration of SnowFrost suggested some modifications, and this new snow depth model is presented below.

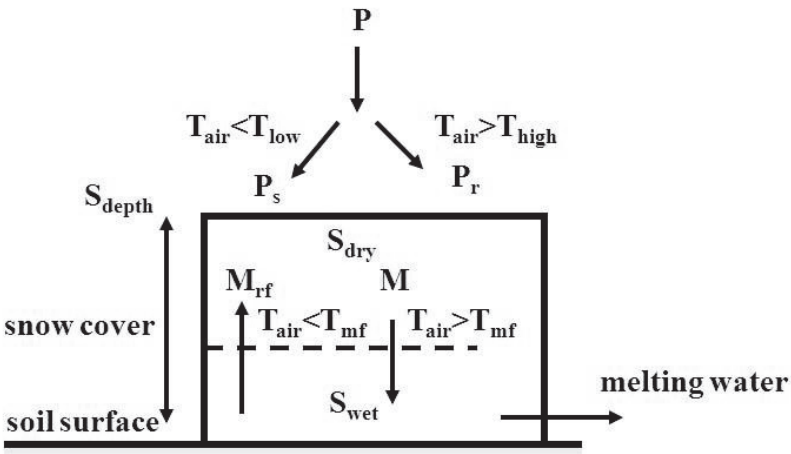


Figure 1: Description of the system simulated in the snow depth model.

Table 1: Symbol and description of the 9 parameters in the snow depth model.

$\{\boldsymbol{\theta_i}\}$	**Interpretation**	**Symbol**
1	Precipitation falls as rain if $T_{air} > T_{high}$ (°C)	T_{high}
2	Precipitation falls as snow if $T_{air} < T_{low}$ (°C)	T_{low}
3	Threshold temperature for snow melt M (°C) and refreezing S_{wet}	T_{mf}
4	Densification of snow cover (mm mm-1 day-1)	ξ
5	The difference between the maximum and minimum value for the melting rate of snow pack K (mm °C-1 day-1)	ΔK_{max}
6	Minimum value for the melting rate of snow pack K (mm °C-1 day-1)	K_{min}
7	Degree-day temperature refreezing index (mm °C-1 day-1)	SW_{rf}
8	The density of fresh snow (kg m-3)	ρ_{ns}
9	The retention capacity of snow cover (mm mm-1)	SW_{ret}

Based on precipitation rate P (mm day-1), mean air temperature T_{air} (ºC) and two threshold temperatures T_{high} (ºC) and T_{low} (ºC) (see Table1), the snow model determines the precipitation form (rain P_r (mm day-1) or snow P_s (mm day-1), where falling new snow has the density ρ_{ns} (kg m-3)) by calculating a fraction of liquid water f_w (mm mm-1) of the precipitation according to

$$f_w = \begin{cases} 1.0 & \text{if } T_{high} < T_{air} \\ \dfrac{T_{air} - T_{low}}{T_{high} - T_{low}} & \text{if } T_{low} \leq T_{air} \leq T_{high} \\ 0 & \text{if } T_{air} < T_{low} \end{cases}$$

where the corresponding amounts of P_r and P_s are

$$P_r = f_w P$$
$$P_s = (1 - f_w) P$$

The snow cover consists of water in solid state S_{dry} (mm) (snow and ice), and liquid water within the snow cover S_{wet} (mm). In SnowFrost snow melt occurs when T_{air} exceeds the base temperature T_{bm} (ºC), and refreezing of S_{wet} occurs when T_{air} drops below T_{bf} (ºC). Preliminary calibration of the SnowFrost model showed that the marginal posterior distributions for the two threshold temperatures T_{bm} and T_{bf} was practically equal. We therefore replaced T_{bm} and T_{bf} with one threshold temperature T_{mf} (ºC), that determines whether snow is melting M (mm day-1), when $T_{air} > T_{mf}$, or liquid water within the snow is refreezing M_{rf} (mm day-1), when $T_{air} < T_{mf}$. The snow cover, being a porous medium, can retain a limited amount of liquid water S_{wet} resulting from rain or melted snow. Similar to

Engseth et al. (2000), we estimate the potential retention capacity of the snow cover as $SW_{ret} \cdot S_{dry}$ where SW_{ret} (mm mm-1) is the retention capacity of the snow cover. Liquid water within the snow cover may refreeze at the rate SW_{rf} (mm ºC-1 day-1). Also, following the idea of Engseth et al. (2000), we calculate the rate of snow melt using a temperature dependent rate K (mm ºC-1 day-1) described by a sinusoidal curve; the period is one year with maximum snow melt rate K_{max} (mm ºC-1 day-1) occurring on 23. June, and minimum snow melt K_{min} (mm ºC-1 day-1) on 23. December. To avoid situations like $K_{max} < K_{min}$ during the calibration, we replaced K_{max} by $\Delta K_{max} = K_{max} - K_{min}$ and calibrate ΔK_{max} (mm ºC-1 day-1). If the entire snowpack melted instantaneously, the resulting depth of water is known as the snow water equivalent SWE (mm). SWE is defined as the sum of S_{dry} and S_{wet}, and the density of the snow cover ρ_s (kg m-3) is defined as SWE/S_{depth} (note: 1 mm of precipitation equals 1 kg m-2). Densification of the snow cover due to change in physical properties (e.g. change in shape of snowflakes and the increase in weight of overlying snow following accumulation) is incorporated through the empirical compaction parameter ξ (mm mm-1 day-1). We use the following equations for the snow cover dynamics:

$$
\begin{aligned}
\Delta \mathrm{S_{dry}}/\Delta \mathrm{t} &= P_s + M_{rf} - M \\
\Delta \mathrm{S_{wet}}/\Delta \mathrm{t} &= P_r + M - M_{rf} \\
\Delta \mathrm{S_{depth}}/\Delta \mathrm{t} &= P_s/\rho_{ns} - M/\rho_s - \xi S_{depth} \\
K &= \Delta K_{max}/2 \sin(2\pi t/365 + 3\pi/8) + (K_{min} + \Delta K_{max}/2) \\
M &= K(T_{air} - T_{mf}) \\
M_{rf} &= SW_{rf}(T_{mf} - T_{air})
\end{aligned}
$$

Snow depth model parameters to be calibrated are listed in Table 1.

STATISTICAL METHOD

The model, $Y(\boldsymbol{\theta}, \boldsymbol{x})$ simulates output variables $\hat{\boldsymbol{y}}$ using input variables $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$. In the Bayesian calibration approach, parameters are regarded as random variables and thus follow some probability distribution. Instead of searching for the best parameter estimates $\boldsymbol{\theta}^*$, we actually search for the probability distribution of these parameters. The calibration routine collects samples from these distributions and parameter uncertainties may be visualized together with point estimates.

Bayesian learning

Bayes theorem is the building block in Bayesian calibration. It was formulated by Thomas Bayes in 1763 (Berger 1985), and may be reformulated as

$$\pi(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{f(\boldsymbol{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(\boldsymbol{d})}$$

where the parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ ($\boldsymbol{\Theta}$ is the whole parameter space) and $\boldsymbol{d} = (d_1, d_2, \dots, d_M)$ is the collected data. The formula reverses conditional probabilities by looking at the unknown parameter set $\boldsymbol{\theta}$ as random variables. The posterior probability distribution $\pi(\boldsymbol{\theta}|\boldsymbol{d})$, is the probability distribution of the parameters given the collected data. According to Bayes formula it is found by combining the original parameter uncertainty, expressed by a prior probability distribution $\pi(\boldsymbol{\theta})$ and the conditional probability density function of the collected data given the parameters, $f(\boldsymbol{d}|\boldsymbol{\theta})$ (often called the likelihood function and denoted $L_d(\boldsymbol{\theta})$). The so called 'evidence' or 'integrated likelihood' term $f(\boldsymbol{d})$ is constant and found by the integral $f(\boldsymbol{d}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\theta$. This gives us the proportionality

$$\pi(\boldsymbol{\theta}|\boldsymbol{d}) \propto f(\boldsymbol{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \qquad (1)$$

which shows that the posterior information is a combination of prior knowledge and new information incorporated through the likelihood function of the collected data.

The prior distribution

The prior distribution quantifies the original uncertainty we have about the parameters. According to Ellison (1996) there are three different interpretations of the prior distribution: (1) a frequency distribution based on existing data. As long as the same data is not used twice, a part of the collected data can be used, or existing data from an earlier investigation, (2) an "objective" statement of what is rational to believe about the parameters and (3) a subjective measure of what the investigator actually believes about the parameter values.

Although limited, the prior information that reflects the initial population basis will assist in the probability distribution of the posterior prediction (Gelman et al. 1996b and Marshall et al. 2004). If the prior dominates the likelihood, the prior will have much greater effect on the posterior probability function than the subsequent experiment can supply. Most of the

criticism of Bayesian inference is that Bayesian analysis can produce results consistent with any point of view when specifying a subjective prior based on personal belief (Dennis 2004). It is therefore of great importance, not to use unrealistically informative prior. If non-informative prior distributions were used for all the individual parameters, then the model would fit the data very closely, but often also with scientifically unreasonable parameters. This may motivate the researcher to specify a prior distribution using external information (Gelman 2002). If no prior information of the parameters is available, non-informative priors (one approach introduced by Jeffrey (1961)) may be used, so that the inferences are unaffected by information external to the experiment (Gelman et al. 1996 b). As usual, we will assume prior independence between the parameters. When having more than one parameter, the joint prior density can be written

$$\pi(\boldsymbol{\theta}) = \pi(\theta_1, \theta_2, \ldots, \theta_L) = \prod_{l=1}^{L} \pi_l$$

where L is the total number of parameters in the model.

The Likelihood function

The likelihood function is the data distribution, conditional on the model used and, expressed as a function of the model parameter values. Measurements $\boldsymbol{d}$ made in the true observable quantity $\boldsymbol{y}$ are not perfect. At the same time, the model $Y(\boldsymbol{\theta},\boldsymbol{x})$ is a simplification of the real world system.

$$\boldsymbol{d} \equiv Y(\boldsymbol{\theta}, \boldsymbol{x}) + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon}$ is both measurement and representational error. After some simplifications (Rougier 2007), the likelihood function can be written as:

$$\begin{aligned} L_d(\boldsymbol{\theta}^*) &= f(\boldsymbol{d}|\boldsymbol{\theta}^*, \boldsymbol{x}^*) \\ &= \prod_{m=1}^{M} f(d_m|\boldsymbol{\theta}^*, \boldsymbol{x}^*) \\ &= \prod_{m=1}^{M} \varphi(d_m - Y_m(\boldsymbol{\theta}^*, \boldsymbol{x}^*); 0, \sigma_m) \end{aligned}$$

where φ is the univariate normal probability density function and σ_m is the m'th diagonal element of the diagonal variance matrix Σ of the errors. The likelihood function can then be written:

$$L_d(\boldsymbol{\theta}^*) = \prod_{m=1}^{M} \frac{1}{\sigma_m (2\pi)^{1/2}} exp\left(-\frac{1}{2}(d_m - Y_m(\boldsymbol{\theta}^*, \boldsymbol{x}^*))^2/\sigma_m^2\right) \quad (2)$$

Outliers in the collected data may produce bad results. Sivia (2006) solves this problem by formulating a constraint on the Gaussian likelihood function. He used a variant of Jeffreys' prior to specify a lower boundary (σ_0) for the standard deviation

$$\pi(\sigma|\sigma_0) = \sigma_0/\sigma^2$$

for $\sigma \geq \sigma_0$, and zero otherwise.

The formula of the constraint likelihood function with the unknown σ integrated out is written

$$L_{d_m}^{(1)}(\boldsymbol{\theta}^*) = \frac{1}{\sigma_{0_m}(2\pi)^{1/2}}(1 - \exp(-R_m^2/2))/R_m^2$$

Where σ_{0_m} is the lower bound of the standard deviation and $R_m = (d_m - Y_m(\boldsymbol{\theta}^*, \boldsymbol{x}^*))/\sigma_{0_m}$. The equation is not defined for $R_m = 0$, but the limit likelihood when $R_m$goes through zero is found as

$$L_{d_m}^{(0)}(\boldsymbol{\theta}^*) = \lim_{R_m \to 0} L_1(\boldsymbol{\theta}^*) = \frac{1}{2\sigma_{0_m}(2\pi)^{1/2}}$$

By series expansion (not shown here). The total constraint Gaussian likelihood function is finally defined as

$$L_d(\boldsymbol{\theta}^*) = \prod_{R_m \neq 0} L_{d_m}^{(1)}(\boldsymbol{\theta}^*) \cdot \prod_{R_m = 0} L_{d_m}^{(0)}(\boldsymbol{\theta}^*) \quad (3)$$

The variance-covariance matrix, Σ_0 of model and measurement error in the likelihood function is unknown. The measurement error may be found by investigating how reliable the measurement instrument used is. The representational error including both simplifications in the model and the fact that the model and the data are not talking about the exact same parameter (the model predicts mean snow depth while the measurements are point estimates),

would be much harder to find. The problem of estimating the covariance matrix of model errors has been simplified by using a fixed diagonal covariance matrix. Conform Van Oijen et al (2005a), we set the standard deviation of each measurement to 30 % of the mean observed value.

$$\hat{\sigma}_{0m} = 0.3 \cdot d_m$$

To avoid a standard deviation of zero, when no snow is observed, the standard deviation was redefined to be

$$\hat{\sigma}_{0m} = \max\left(0.1, 0.3 \cdot d_m\right)$$

This gives a standard deviation of 0.1 if the mean collected snow depth is less than 0.33m.

The difference between observed and simulated output ($d_m - Y_m(\boldsymbol{\theta}^*, \boldsymbol{x}^*)$), will be dominated by zeros, because no snow depth will be both observed and simulated most of the year. A student-t distribution, having a fatter tail (Miller and Miller 1999) is an alternative recommended when outliers occurs, but is not tested here. Probably a zero-inflated distribution (Agarwal et al. 2002) would be an even better choice

The likelihood function is what modifies the prior knowledge into a posterior distribution. According to Bayes theorem, the more experimental data added, the more will the likelihood dominate the prior, and have much greater effect on the posterior probability distribution.

Random Walk Metropolis

There are four different techniques that can be used to find the posterior distribution in Bayesian calibration; exact calculation, analytical approximation, numerical integration and Monte Carlo simulation. Integration problems makes the exact calculation impossible, especially when $\boldsymbol{\Theta}$ is high dimensional. We will use a MCMC algorithm, random walk Metropolis.

The Metropolis Algorithm is the cornerstone of all Markov chain-based Monte Carlo methods. It was proposed as early as in 1953 in a short paper by Metropolis et al. (1953). The idea is of great simplicity and power, and its variations are in use by many researchers in several different scientific fields.

We implemented the random walk Metropolis algorithm in Matlab. We start with some initial parameter values, $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_L^0)$, where L is the number of parameters in the model. For each iteration step $i \in 1{:}\, I$, we have these steps:

1. Draw $\boldsymbol{\theta}' \sim g_\sigma$, where g_σ is a spherically symmetric distribution, independent distributed for different i, centered at the current state.
2. Compute the ratio $r = \frac{\pi(\boldsymbol{\theta}'|\boldsymbol{d})}{\pi(\boldsymbol{\theta}^{i-1}|\boldsymbol{d})} = \frac{\pi(\boldsymbol{\theta}')\cdot f(\boldsymbol{d}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{i-1})\cdot f(\boldsymbol{d}|\boldsymbol{\theta}^{i-1})}$
3. Draw $u \sim \mathrm{U}[0,1]$, where U is the uniform probability density function, and set

$$\boldsymbol{\theta}^i = \begin{cases} \boldsymbol{\theta}' & \text{if } u \le r \\ \boldsymbol{\theta}^{i-1} & \text{otherwise} \end{cases}$$

The draws $\theta^1, \theta^2, \ldots, \theta^I$ will in the long run converge to the posterior distribution of the parameter set (Liu 2001).

To avoid the joint likelihood to be too large to be represented by a digital computer, that round off to infinity, the natural logarithm was used in all steps in the random walk Metropolis algorithm.

The step length δ is the distance between the current and the proposed parameter vector. Small δ ensures that the proposed parameter vector is close to the current position, so the probability of accepting it is high. With small average δ, the Markov chain will converge slowly since all its moves will be small. On the other hand, a large step length δ places the new proposed parameter further away from the current parameter vector, which leads to a low probability of accepting it. The Metropolis algorithm will then reject a too high proportion of its proposed moves. Most of the computation time goes to costly evaluation of the posterior density. The step length δ therefore controls the effectiveness of the Metropolis algorithm. According to Roberts et al. (1997), an acceptance rate of roughly 0.23 is desired. We considered an acceptance rate between 0.15 and 0.5 to be acceptable (Roberts 1996).

The choice of a proposal distribution may be a crucial factor for convergence of the algorithm. Adaptive MCMC algorithms (Andrieu and Thoms 2008, Smith and Marshall 2008) solve this problem by using the chain history in order to continually tune the proposal distribution.

Convergence Diagnostics

The random walk Metropolis algorithm produces a Markov chain whose stationary distribution is the target posterior distribution. If the iterations before stationarity are used to summarize the target distribution, they can give false answers. To detect the state of stationarity (``burn-in'' state) or lack of stationarity, different methods exist. Gelman and Rubin (1992) pointed out that in many problems, lack of convergence can be easily determined from multiple independent sequences but can not be diagnosed using simulation outputs from any single sequence. The sequence may remain in a region heavily influenced by the starting point, although it has not converged to the true posterior distribution (Gelman et al. 1996b). In contrast, Geyer (1992) states that one should concentrate all computational resources in a single chain, since it is more likely that for example the latter 90.000 iterations from a single run of 100.000 iterations come from the target distribution than the final samples from 10 parallel runs of 10.000.

When running parallel sequences, the most obvious approach to assess convergence is to plot the chains as time series and assess by visual inspection whether the sequences have converged to each other. A more quantitative approach based on the ratio of between- and within-chain variance was formulated by Gelman and Rubin (1992). Convergence is identified when the empirical distribution of simulations obtained separately from each sequence is approximately the same as the distribution obtained by mixing all the sequences together. Before the parallel sequences have converged, the simulations from each sequence will be much less variable than the sequence combined. We assume J parallel simulations, each of length I and with starting points drawn randomly from the prior distribution that is over-dispersed in the sense of being more variable than the target posterior distribution. The first $I/2$ iterations are discarded to diminish the effect of the starting distribution. The estimated potential scale reduction factor $\sqrt{\hat{R}}$ is calculated at each iteration step

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{I/2-1}{I/2} + \frac{J+1}{J\,I/2}\frac{\hat{B}}{\widehat{W}}\frac{df}{df-2}\right)}$$

where B is the variance between the sequence means and W is the average within-sequence variance. df refers to the degree of freedom in a t-distribution approximation to the empirical distribution of θ. For large number of samples, $df/(df-2)$ can be ignored.

$$\hat{B} = \frac{I/2}{J-1}\sum_{j=1}^{J}\left(\overline{\theta}_{J\cdot} - \overline{\theta}_{\cdot\cdot}\right)^2$$

$$\widehat{W} = \frac{1}{J(I/2-1)}\sum_{i=1}^{I/2}\sum_{j=1}^{J}\left(\theta_{ji} - \overline{\theta}_{J\cdot}\right)^2$$

When $\sqrt{\hat{R}}$ is close to 1 (less than 1.2 in practice (Gelman 1996)), the parallel Markov chains are essentially overlapping. We should also make sure that the mixture of sequence variance $\hat{V}$ and the within sequence variance $\widehat{W}$ stabilizes as a function of I (Brooks and Gelman 1998).

$$\hat{V} = \frac{I/2-1}{I/2}\widehat{W} + \frac{J+1}{J\,I/2}\hat{B}$$

The iterations before the ``burn-in'' state are discarded. Typically one will discard only a small fraction of the run. So, if after ``burn-in'' state you are left with less than half the run, you haven't run the iterations for long enough (Kass et al. 1998).

DATA

The snow depth model is calibrated using snow depth data from Kise, Norway, which is situated 60.77N, 10.8E, 127 meters above sea level. Kise has a continental climate, and the landscape is dominated by arable land and the largest lake in Norway, Mjøsa. The model is calibrated using data from 10 years, 1988 to 1998, and it is tested using data from the 5 following years. Temperature and precipitation observations are obtained from Bioforsk Agrometeorological service, the snow depth observations are obtained from both Bioforsk Agrometeorological service (1988-1997) and from The Norwegian Water Resources and Energy Directorates service (1997-2003). On average there was snow cover 120 days of the year, with an average depth of 0.16 m. Variation between years was from 160 days with snow cover and an average of 0.36 m the winter 1993/1994 to only 77 days with snow cover and an average of 0.06 m the winter 1989/1990,

RESULTS AND DISCUSSION

Tuning the MCMC

To run the Bayesian calibration algorithm, several decisions must be made by the researcher.

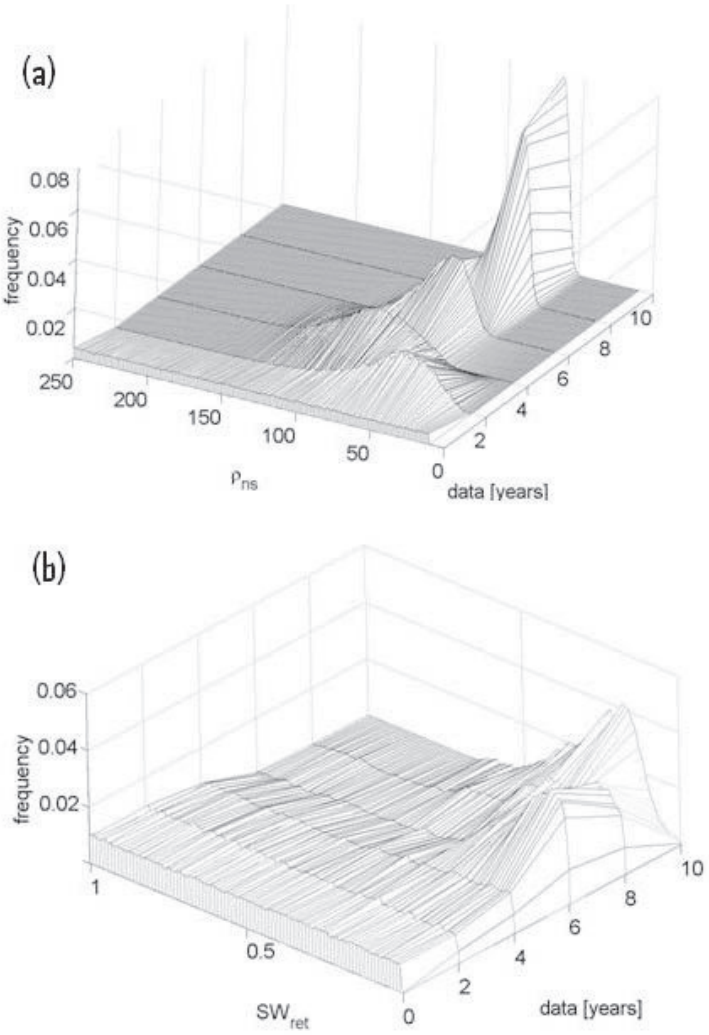
The number of observed data

Usually, we will use all available data. Here, 10 years of snow depth data are used to fit the model and 5 years to test the model. To see the effect of the number of data used to fit the model, the calibration algorithm was also run using 2, 4, 6 and 8 years of collected data. The results from ρ_{ns} (θ_8) and SW_{ret} (θ_9) are plotted as three dimensional figures to visualize the change in uncertainty about the respective parameter when adding data (Figure 2). In agreement with Bayesian learning (Equation 1), we can see that the uncertainty, i.e. the width of the histograms, decreases when adding more data. Also the position of the histograms changes, most dramatically up to 8 years, but also from 8 to 10 years. The weather situation varies from year to year, and the parameter estimates do depend on what kind of years used. A period of 8 years does not contain all variability in weather, and the estimates do therefore change further when adding two more years. Most probably, the estimates will still change when adding more years of data, until the whole specter of weather situations are included.

Reaching convergence for the posterior chains was easy when 2 or 4 years of collected data were used. With 6 or more years of data, convergence became much harder to reach.

The prior distribution

We have chosen to use relatively wide uniform prior distributions. Usually, we will not consider all values between the upper and lower limit in the prior distribution as equally believable. We therefore constructed a beta distribution between the boundaries and used results from Engseth et al. 2000 (Table 2) as modal values. Comparing the results of calibration starting from beta distributions rather than uniform ones, showed that the more informative beta priors gave much easier convergence and different point and interval estimates for the parameters. These new estimates permitted more meaningful physical interpretation, but showed worse fit according to RMSE for both the training and the test data.



Figur 2: Changes in parameter uncertainty when using respectively 0, 2, 4, 8 and 10 years with collected data in the calibration routine. Figure (a) shows parameter uncertainty for $\boldsymbol{\rho_{ns}}$ $(\boldsymbol{\theta_8})$and $\boldsymbol{SW_{ret}}$ $(\boldsymbol{\theta_9})$.

RMSE is now commonly reported in environmental modeling in comparsions of different calibration approaches (e.g. Reinds et al. 2008, Lehuger et al. 2009). The lower RMSE with the uniform prior was not necessarily expected because our likelihood function was not Gaussian nor did our data all have the same standard deviation.

The Likelihood function

Both the Gaussian likelihood function (Equation 2) and the Gaussian likelihood function with constraints (Equation 3) were tested and the constrained function gave much faster parameter convergence than the ordinary Gaussian. Less iteration were needed to reach convergence and since each iteration step is time requiring; only the constrained function was used.

The proposal distribution

The choice of an effective proposal distribution for the MCMC is essential in order to obtain convergence in a limited amount of time. Both a uniform and a Gaussian proposal distribution centered at the current state were tested. The Gaussian distribution, which predominantly samples close to zero turned out to be the most efficient one and was therefore used. The covariance matrix was defined as diagonal with the *l*'th diagonal element proportional to the width of the prior interval for the respective parameter $\left(a_l = c \cdot \left(\theta_l^{max} - \theta_l^{min}\right)\right)$. In order to achieve efficient convergence, the constant c was set by trial and error to produce an acceptance rate of roughly 0.23 (Roberts et al. 1997). To prevent the most sensitive parameters T_{high} (θ_1), T_{low} (θ_2), ξ (θ_4) and ρ_{ns} (θ_8) (according to sensitivity analysis calculated for the entire *SnowFrostIce* model (Thorsen et al. 2009)) to control the whole accept/reject procedure, the constant c was individually corrected up for all other parameters. The sensitive parameter ξ (θ_4) proved most challenging. If its proposal stepsize was not weighted down, it controlled most of the accept/reject procedure and made it impossible for all other parameters to converge. At the same time, if weighted down enough, convergence is not reached within a proper time when having a starting value for the parameter far away from the target posterior distribution. Our solution was to weigh the parameter down after a number of iterations. To keep the rule that the step length has to be identically distributed for different iterations (Liu 2001), this is done during the ``burn-in'' phase only.

Adaptive MCMC algorithms (Andrieu and Thoms, 2008, Smith and Marshall 2008) were not used, but we did use adaptation informally, as follows. The information obtained by the calibration using two years of observations was used to form a suitable proposal distribution for the calibration using all ten years of observations. The new covariance matrix was calculated from the variances of the proposal distribution and the correlation matrix calculated from the resulting parameter chains after convergence when using two years of observations.

Then only a scaling factor for the entire covariance matrix had to be found by trial and error to produce an efficient acceptance rate. Preliminary tests of this method showed it to be highly efficient for the calibration of our model, but it was not used to produce the results reported here.

Convergence

The calibration algorithm was run for 300.000 iterations, requiring about 40 hours computing time. The usefulness of running parallel sequences to detect convergence was found during model development from the first version of the snow model (Thorsen and Haugen 2007) to this version. The Markov chain for T_{bm} (Lower limit temperature for snow melt) is plotted in Figure 3a for 150.000 iterations, which objectively seems like a large enough number. Running only one chain, we would determine ``burn-in" after 10.000 iterations by eyes, and treat the remaining 140.000 iterations as draws from the true posterior distribution. When running two chains in parallel instead (Figure 3b), we found that the two chains had not converged to each other during this run. We therefore ran the algorithm for 150.000 more

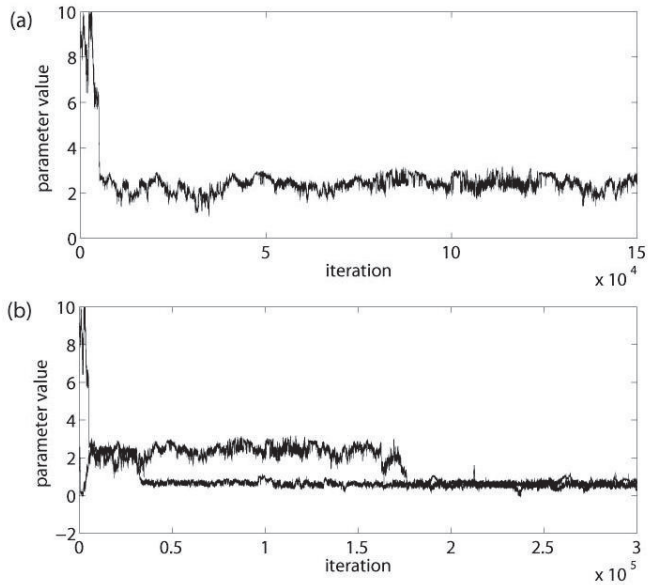


Figure 3: Markov chain of $\boldsymbol{T_{bm}}$ (parameter not included in the finial version of the snow depth model) (a) one chain for 150.000 iterations and (b) two chains in parallel for 300.000 iterations

iterations and clearly they converge to each other after 175.000 iterations. With more confidence, we can now treat the last 125.000 iterations as draws from the posterior distribution. In most cases, four sequences were run in parallel, but during the development of the model only two. The method of Gelman and Rubin was used to detect ``burn-in''.

Reflection

When using upper and lower limits in the prior distribution, new proposal parameters may be generated outside these boundaries and consequently be rejected in the random walk Metropolis algorithm. Here, T_{low} (θ_2), which stabilized near the lower boundary of the prior interval, caused a high rejection rate. To avoid proposal parameters generated outside the prior boundaries, reflection at the boundaries (Yang 2006) is used. If the proposal parameter θ' is outside the prior interval ($[\theta_{min}, \theta_{max}]$) the excess is reflected back into the interval; that is, if $\theta' > \theta_{max}$, θ' is reset to $\theta_{max} - (\theta' - \theta_{max})$, and if $\theta' < \theta_{min}$, θ' is reset to $\theta_{min} + (\theta_{min} - \theta')$. The proposal parameter distribution will still be symmetric($p(\theta' | \theta^{i-1}) = p(\theta^{i-1} | \theta')$) and thereby the acceptance of the Metropolis algorithm correct, since $p(\delta) = p(-\delta)$ and a step length δ from the current state θ^{i-1} with reflection will give the state θ' ($\theta' = \theta^{i-1} + \delta = 2\theta_{min} - (\theta^{i-1} + \delta)$), while the same step length δ from θ' will reflect the proposal parameter back to current state θ^{i-1} ($\theta' + \delta = 2\theta_{min} - (\theta' + \delta) = \theta^{i-1}$).

Uncertain vs. fixed values of the parameters

To reach convergence for all parameters turned out to be difficult. We therefore reduced the number of parameters by setting some of them to fixed values in the early investigation. This showed that convergence problems emerged when four or more parameters were included in the calibration. It was therefore tempting to treat some of the less sensitive parameters as fixed values. But this has the problem of producing inaccurate estimates and it underestimates uncertainty when the parameters to be fixed are not known accurately (Gelman et al. 1996a). Especially the posterior distribution of T_{low} (θ_2) changed dramatically when the last three parameters were fixed. In the results, all 9 parameters are treated as uncertain. By performing a sensitivity analysis, e.g. using the Morris method as described in Campolongo et al. 2007, it

is possible to identify the least important parameters which are candidates for fixed values. This was done for a similar snow depth model by Thorsen et al. 2009.

Statistical Inference

The model is calibrated using wide uniform prior distributions (Table 2) and four sequences are run in parallel, each for 300.000 iterations. All parameters converged after less than 100.000 iterations (according to Gelman and Rubins criteria) with an acceptance rate of approximately 0.25. The potential scale reduction factor, $\sqrt{\hat{R}}$ at the end of the calibration is listed in Table 3. The highest value is 1.02 for both ΔK_{max} (θ_5) and SW_{ret} (θ_9) suggesting that additional simulation might reduce the posterior interval for these parameters by only up to a factor of 1.02. The 95 % highest posterior density (HPD) interval is calculated for each parameter and listed in Table 3. Each interval is a measure of how certain we are about the respective parameter, and we can clearly see a decreased uncertainty compared to the prior information (Table 2) for all parameters except SW_{rf} (θ_7).

Three point estimates are calculated, the mean $(\overline{\boldsymbol{\theta}})$, the mode $\left(\widehat{\boldsymbol{\theta}}_{mode}\right)$ and the maximum a posteriori estimate ($\widehat{\boldsymbol{\theta}}_{MAP} = \arg\max\ \pi(\boldsymbol{\theta}|\boldsymbol{x})$) (Gilks et al. 1996). Both the mean and the mode estimates are calculated for each parameter one by one, while the MAP estimate is the largest mode for the joint posterior distribution. All three estimates are different (Table 3)

Table 2: Minimum and maximum values used to define limits in the uniform prior interval and parameter estimates from Engseth et al. (2000), $\widehat{\boldsymbol{\theta}}_E$. Lack of number indicates that the specific parameter does not occur in Engseth's model.

$\{\boldsymbol{\theta}_i\}$	Symbol	$\boldsymbol{\theta}_{min}$	$\boldsymbol{\theta}_{max}$	$\widehat{\boldsymbol{\theta}}_E$
1	T_{high}	-5	10	0.5
2	T_{low}	-10	5	0.5
3	T_{mf}	-10	10	0.5
4	ξ	0	1	
5	ΔK_{max}	0	10	1.25
6	K_{min}	0	10	2
7	SW_{rf}	0	10	0.01
8	ρ_{ns}	10	250	
9	SW_{ret}	0	1	0.1

which may be explained by skewness and several peaks in the parameter densities and by correlations (Table 3) between the different parameters. The MAP estimate, which is the only estimate considering the entire parameter set concurrently is the estimate giving the smallest root mean square error (RMSE) for both the training data and for the test data. It is important to note that the MAP estimate is the parameter set having the largest posterior density among our 800.000 iterations. Each of our four parallel chains gave a different MAP estimate, suggesting that the parameter space is not totally searched and additional simulation might give even better MAP estimates. The Bayesian calibration method does not search for the best parameter estimates, but for the posterior distribution of them.

Bayesian calibration simply combines prior parameter information with the likelihood of the data given the model. Since the model is not a perfect representation of the system, parameter estimates may deviate from physically meaningful values. Here, especially T_{low} (θ_2) seems unrealistic (Table 3). To reach a more realistic estimate, the model could have been improved, a more informative prior distribution could have been used or a separating parameter could replace T_{high} (θ_1) and T_{low} (θ_2). Since the purpose of our model is not to learn about the system, but prediction and we do not want a too complex model, both the model and the parameters are retained while a more informative prior was tested.

Table 3: Results from Bayesian Calibration using MCMC with chains of 300000 iterations. The potential scale reduction factor $\left(\sqrt{\widehat{R}}\right)$, the mean parameter value $(\overline{\boldsymbol{\theta}})$, the mode parameter value $(\widehat{\boldsymbol{\theta}}_{\boldsymbol{mode}})$, the maximum a' posteriori estimate $(\widehat{\boldsymbol{\theta}}_{\boldsymbol{MAP}})$, the coefficient of variation (CV), the 95 % HPD interval and the parameter with which parameter is correlated at greater absolute values than 0.3 (underlined if negative).

$\{\boldsymbol{\theta_i}\}$	$\sqrt{\widehat{\boldsymbol{R}}}$	$\overline{\boldsymbol{\theta}}$	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{mode}}$	$\widehat{\boldsymbol{\theta}}_{\boldsymbol{MAP}}$	CV	95 % HPD	Correlated $\{\boldsymbol{\theta_i}\}$
1	1.00	1.95	1.87	1.99	0.12	[1.60, 2.42]	
2	1.00	-9.32	-10.00	-9.92	-0.06	[-10.00, -8.11]	$\{\theta_8\}$
3	1.01	0.58	0.54	0.61	0.14	[0.43, 0.74]	$\{\theta_6\}$ $\{\underline{\theta_9}\}$
4	1.01	0.03	0.03	0.03	0.10	[0.02, 0.03]	$\{\underline{\theta_8}\}$
5	1.02	2.27	1.05	1.63	0.73	[0, 5.40]	
6	1.01	6.69	7.01	6.69	0.10	[5.36, 7.86]	$\{\theta_3\}$
7	1.00	5.07	1.05	1.56	0.56	[0.51, 10]	
8	1.00	65.12	63.25	62.10	0.08	[54.96, 75.75]	$\{\theta_2\}$ $\{\underline{\theta_4}\}$
9	1.02	0.25	0.18	0.31	0.46	[0.06, 0.50]	$\{\underline{\theta_3}\}$

The coefficient of variation (CV) is a normalized measure of dispersion of a probability distribution defined as the ratio of the standard deviation to the mean. Three parameters $\left(\Delta K_{max}\ (\theta_5), SW_{rf}\ (\theta_7) \text{and } SW_{ret}\ (\theta_9)\right)$ stand clearly out with greatest CV values (Table 3). These parameters are also the parameters having the largest relative distance between the three point estimates, having the longest ``burn-in'' phase and having the smallest relative decrease in uncertainty when comparing the prior interval with the 95 % HPD interval. This indicates that the information from new data had little effect on these three parameters, not only on their parameter values, but on general knowledge about the parameter characterized by their posterior distribution. A sensitivity analysis done for the whole SnowFrostIce model (Thorsen et al. 2009) gave the result that changes in these three parameters also gives the smallest rate of changes in the output of the model. In summary, the results from the Bayesian calibration can tell us how new information from data are allocated within the model and accumulated as increased knowledge for some parameters while leaving others unaffected. While sensitivity analysis tells us about the sensitivity of model outputs to changes in parameters, Bayesian calibration tells us about how new information affects our knowledge about the parameters and model outputs.

Predictive uncertainty in model outputs is visualized in Figure 4 together with snow depth observations for the two test years 1998/1999 and 1999/2000. Prior and posterior uncertainties are calculated by sampling randomly 100.000 samples from the prior distribution and from the posterior chains respectively. Model outputs are then calculated for each parameter set, and the uncertainty plotted as one standard deviation above and below the mean model output for each day. Wide prior intervals were used for the parameters and we can see that the predictive posterior uncertainty is much reduced compared to the prior uncertainty for the outputs. The calibrated model approximates the data fairly closely, except for some underestimation during periods of prolonged large snow depth. Standard goodness of fit assessment is also done by constructing a predictive qq plot (Dawid 1984 and Thyer et al. 2009), checking if the predictive distribution is consistent with the observed data (Figure 5). If the cumulative distribution function (cdf) of the predictive distribution of snow cover (assumed Gaussian distributed with mean and standard deviation calculated from the 100.000 random samples drawn from the posterior chains) is independent uniform U[0,1] variables, the observations are realizations of the predictive distribution. The shape of the qq plot (Figure 5) indicates underestimated predictive uncertainty (according to Thyer et al. (2009)). We expect that this underestimation in the current case is caused by representational error,

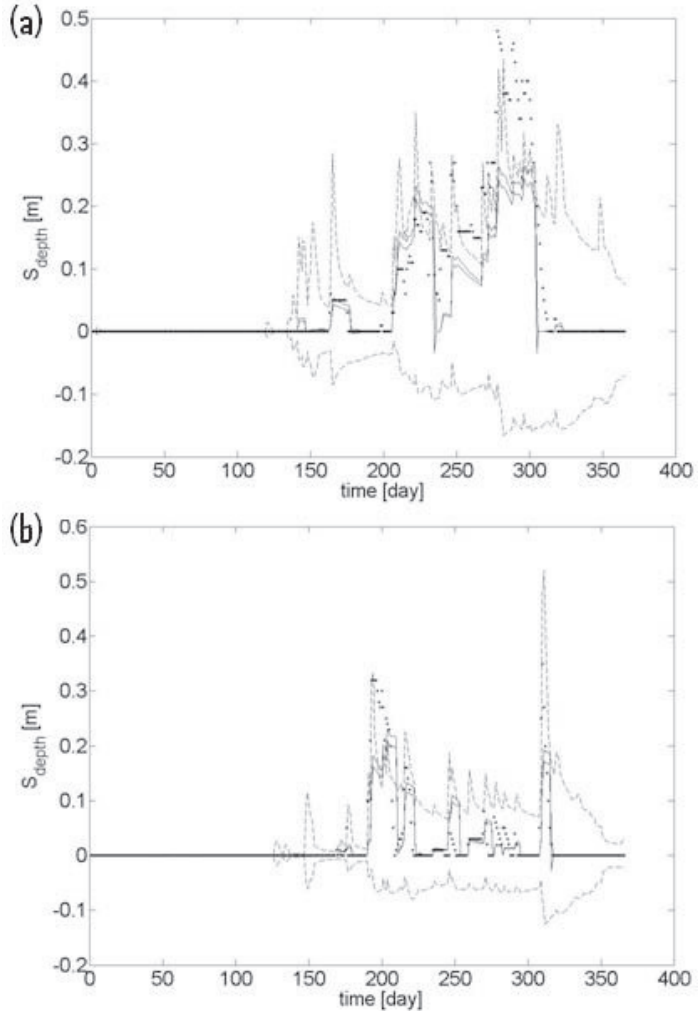


Figure 4: Prior and posterior uncertainties in the model output snow depth plotted as one standard deviation above and below the mean output from 100.000 model runs for (a) 1998/1999 (b) 1999/2000. Dotted line denotes prior uncertainty and solid line denotes posterior uncertainty. The stars denote observed values.

which we have not quantified (see introduction), but also measurement error not included in the likelihood of the data may add to the predictive uncertainty. Further more, the information in the calibration data may not span the variation range sufficiently, causing an additional parameter uncertainty (e.g. Figure 2). The posterior uncertainty is also visualized by frequency histograms of estimated snow depth values for each day during the winter 1998/1999 (Figure 6). We can see snowing and snow melt periods as changes in the snow

depth position of the frequency histograms over time. We can also see a larger uncertainty for larger snow depth values as wider frequency histograms for the larger snow depth values.

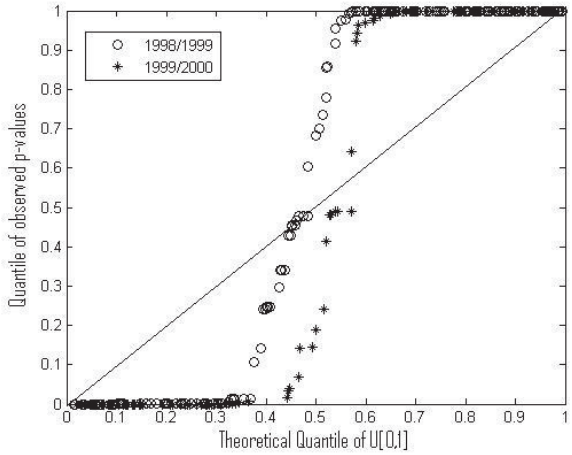


Figure 5: Predictive QQ plot.

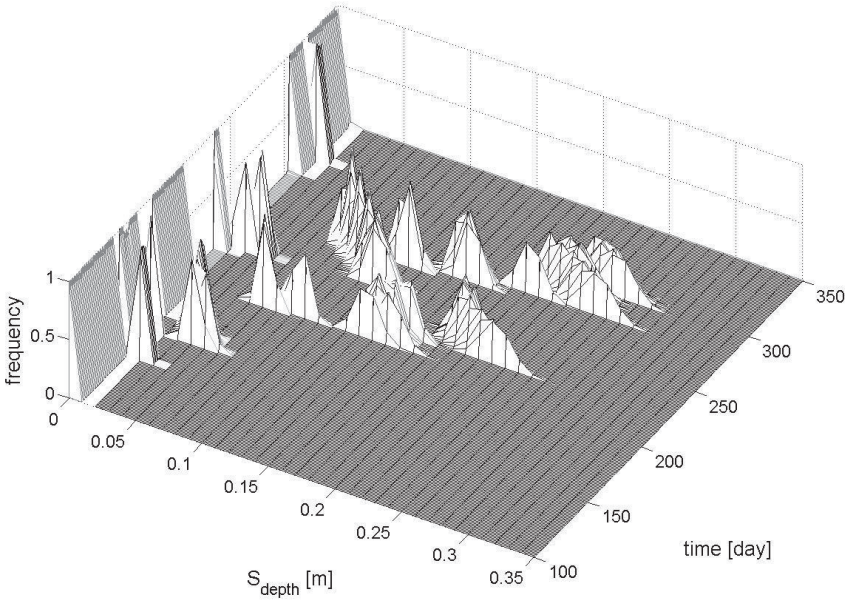


Figure 6: Posterior uncertainties on the output for the period 1. October 1998 to 1. May 1999, plotted as frequency histograms of the output each day.

CONCLUSIONS

We have used Bayesian calibration to calibrate a complex model of snow depth. The Bayesian approach regards parameters as random and prior information of the parameters is combined with observed data to form a joint posterior parameter distribution.

Here, point estimates were calculated and uncertainties visualized for both parameters and model outputs. Clearly, the uncertainty of both the parameters and the model outputs decreased when adding more data. Also, the amount of data affected the parameter estimates since the input data varied from year to year and the data used did not include the whole specter of varieties in the input space. The best fit of the model was found when using less informative priors, while more informative priors gave more meaningful physical values for the parameters. To detect ``burn-in'' both objectively and correctly, both multiple chains and Gelman Rubin were found to be useful. The choice of treating some uncertain parameters as fixed values simplified the calibration procedure, but changed the parameter estimates and led to underestimated parameter uncertainty.

We used the Markov chain Monte Carlo algorithm, random walk Metropolis. Both the idea and the implementation of the algorithm are relatively simple, but the use of the method for calibration of the complex model was far from straightforward in practice. The major problem was to obtain convergence of the chains in a limited amount of time. With regard to the prior parameter distributions, we found that informative beta distributions led to faster convergence of the posterior parameter chains than less informative uniform priors. Faster convergence was also achieved by the use of Sivia's constraint likelihood rather than the more common Gaussian likelihood function. The choice of an effective proposal distribution was difficult, but optimizing the proposal distribution with parts of the data was found to be useful. To avoid spending time on proposal parameters generated outside the prior interval, the reflection method was successfully used.

ACKNOWLEDGEMENTS

We want to thank the two anonymous referees on the first version of this paper who made extensive detailed suggestions on how to improve it. We thank them for their effort, which has been of great help in rewriting the paper.

REFRENCES

Agarwal, D. K., Gelfand, A. E., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics **9 341-355**.

Andrieu, C., Thoms J., 2008. A tutorial on adaptive MCMC. Stat. Comput., **18 343-373**.

Berger, O. J., 1985. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York.

Brooks, S. P., 1998. General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics **7 434-455**.

Campbell, K. 2006. Statistical calibration of computer simulations. Reliability Engineering and System Safety **91 1358-1363**.

Campolongo, F., Cariboni J., Saltelli A., 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling and Software 22 **1509-1518**.

Dawid, A. P., 1984. Statistical theory: The Prequential Approach. J. R. Stat. Soc., Ser. A, **147 278-292**

Dennis, B., 2004. Statistics and scientific method in ecology. The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations **327-378**.

Ellison, A. M., 1996. An Introduction to Bayesian Inference for Eqological Research and Environmental Decision-Making. Ecological Applocation **6 1036-1046**.

Engeseth, R. V., Sorteberg, H. K., Udnæs, H., 2000. NOSiT, utvikling av NVEs operasjonelle snøinformasjonstjeneste. Norges vassdrags- og energidirektorat.

Gelman, A., 1996. Inference and Monitoring Convergence. In: Gilks, W. R., Richardson, S., Spiegelhalter, D. J., (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall.

Gelman, A., 2002. Prior distribution. Encyclopedia of Environmetrics **3 1634-1637**

Gelman, A., Rubin, D. B., 1992. Inference form Iterative Simulation Using Multiple Sequences. Statistical Science **7 457-511**.

Gelman, A., Bois, F., Jiang, J., 1996a. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. Journal of the American Statistical Association **91 1400-1412**.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 1996b. Bayesian Data Analysis. Chapman & Hall.

Geyer C., 1992. Practical Monte Carlo Markov chain (with discussion). Statistical Science **7 473-511**.

Goldstein, M., Rougier, J., 2006. Bayes Linear Calibrated Prediction for Complex Systems. Journal of the Americal Statistical Association **101 1131-1143**.

Hue, C., Tremblay, M., Wallach, D. J., 2001. A Bayesian Approach to Crop Model Calibration Under Unknown Error Covariance. Journal of the Americal Statistical Association **101 355-365.**

Jeffreys, H., 1961. Theory of probability. Oxford University Press

Kass, R. E., Carlin, B. P., Gelman, A., Neal, R. M., 1998. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. The American Statistican **52 93-100**.

Kavetski, D., Kuczera, G. & Franks, S.W. 2006. Bayesian analysis of input uncertainty in hydrological models: 1. Theory. Water Resources Research 42: W03407.

Lehuger, S., Gabrielle, B., Van Oijen, M., Makowski, D., Germon, J.-C., Morvan, T. & Hénault, C. 2009. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agriculture, Ecosystems and Environment **133 208-222.**

Liu, J. S., 2001. Monte Carlo Strategies in Scientific Compution. Springer-Verlag, New York.

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X. & Zhang, L. 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. Ecological Applications **19 571-574.**

Marchall, L., Nott, D., Sharma, A., 2004. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resource Research. **40**

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. Journals of Chemical Physics **21 1087-1092**.

Miller, I., Miller, M., 1999. Freund's Mathematical Statistics. Prentice-Hall, Upper Saddle River.

Reinds, G.J., Van Oijen, M., Heuvelink, G.B.M. & Kros, H. 2008. Bayesian calibration of the VSD soil acidification model using European forest monitoring data. Geoderma **146: 475-488.**

Riley, H., Bonesmo, H., 2005. Modelling of snow and freezing-thaw cycles in the EU-rotate_N secision support system. Grønn kunnskap **9,1-8**.

Roberts, B. O., 1996. Markov Chain concepts related to sampling algorithms. In: Gilks, W. R., Richardson, S., Spiegelhalter, D. J., (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall.

Roberts, G. O., Gelman, A., Gilks, W. R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. The Annals of Applied Probability **7 110-120**.

Rougier, J., 2007. Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations. Climatic Change **81 247-264**.

Sivia, D. S., 2006. Data Analysis, A Bayesian Turtorial. Oxford University press.

Smith, T.J. & Marshall, L.A. 2008. Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. Water Resources Research 44: W00B05.

Thorsen, S. M., Haugen, L. E., 2007. Development of the SnowFrost model for the simulation of Snow Fall and Soil Frost. Bioforsk FOKUS **2 1-23**.

Thorsen, S. M., Roer, A-G., Van Oijen, M., 2009. Modelling the dynamics of snow cover, soil frost and surface ice in Norwegian grasslands. Submitted for publication.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W. & Srikanthan, S. 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological

modeling: A case study using Bayesian total error analysis. Water Resources Research 45: W00B14.

Van Oijen, M., Rougier, J., Smith, R., 2005a. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiology **25 915-927**.

Van Oijen, M., Höglind, M., Hanslin, H. M., Caldwell, N., 2005b. Process-Based Modelling of Thimothy Regrowth. Agronomy Journal **97 1295-1303**.

Vehviläinen., 1992. Snow cover models in operational watershed forecasting. PhD Thesis, Helsinki University of Technology.

Yang, Z., 2006. Computational Molecular Evolution. Oxford University press.

Paper II

Stig Morten Thorsen, Anne-Grete Roer and Marcel Van Oijen
Modelling the dynamics of snow cover, soil frost and surface ice in Norwegian grasslands.
Polar Research. In press.

Modelling the dynamics of snow cover, soil frost and surface ice in Norwegian grasslands

Stig Morten Thorsen[*†], Anne-Grete Roer[‡§], Marcel Van Oijen[**]

Abstract

Studying the winter survival of forage grasses under a changing climate requires models that can simulate the dynamics of soil conditions at low temperatures. We developed a simple model which simulates depth of snow cover, the lower frost boundary of the soil and the freezing of surface puddles. We parameterised the model by means of Bayesian calibration, and identified the least important model parameters using the sensitivity analysis method of Morris. Verification of the model suggests that the results are reasonable. Due to the simple model structure, some overestimation occurs in snow and frost depth. Both the calibration and the sensitivity analysis suggested that the snow cover module could be simplified with respect to snow melt and liquid water content. The soil frost module should be kept unchanged, while the surface ice module should be changed when more detailed topographical data become available, such as better estimates of the fraction of the land area where puddles may form.

Keywords

Modelling; snow cover; frost depth; ice cover; calibration; sensitivity analysis

Correspondence

Stig Morten Thorsen, Norwegian Institute for Agricultural and Environmental Research, Grassland and Landscape Division, Postvegen 213 N-4353 Klepp st., Norway. Email: stig.thorsen@bioforsk.no

Introduction

Grasslands are important components of Norwegian terrestrial ecosystems. In order to investigate the impacts of climate change, parts of the Norwegian strategic research programme WINSUR are dedicated to developing a grassland model to study the winter survival of different crops. The grassland model, currently simulating the regrowth dynamics of timothy (*Phleum pratense* L.), has been developed by Van Oijen *et al.* (2005a). The same model will be adapted to simulate the regrowth dynamics of perennial ryegrass (*Lolium perenne* L.). During the winter, a significant number of plants may die due to frost, ice encapsulation and other physical and biological stresses (Larsen 1994). Snow cover provides insulation from lethal freezing temperatures, while also reducing the amount of photosynthetically active radiation at plant level. However, more variable winter climate in Norway (Beldring *et al.* 2008) may lead to less snow cover and thus increase plant exposure to killing frosts (Belanger *et al.* 2002).

[*] Norwegian Inst. for Agricultural and Environmental Research, Grassland and Landscape Division
[†] Dept. of Mathematical Sciences and Technology, Norwegian University of Life Sciences
[‡] Norwegian Inst. for Agricultural and Environmental Research, Plant Health and Plant Protection Division
[§] Dept. of Chemistry, Biotechnology and Food Science, Norwegian University of Life Scinences
[**] CEH-Edinburgh, Bush Estate, Penicuik, EH26 0QB, UK

If the ground is frozen, water (rain or snow melt) can accumulate in small depressions, freeze and cause plants to be encapsulated in ice. Ice encasement can severely reduce gas exchange between the plant and the surrounding atmosphere, leading to a transition from aerobic to anaerobic respiration and accumulation of respiration products (esp. CO_2) to toxic levels (Gudleifsson and Larsen 1993).

In order to make predictions on the effects of climate change on plant performance over more than one growing season, the grassland model needs an additional set of functions to describe the winter survival of the sward. The grassland model must be able to simulate effects of winter climate on soil and soil surface processes. The main objective of this work is to develop a simple winter module that can easily be incorporated into the existing grassland model. Therefore the structure of the winter module needs to be kept as simple as possible.

Regarding the simulation of winter climate effects on soil and soil surface conditions (e.g. snow cover and soil frost), the literature provides examples of different approaches (Benoit and Mostaghimi 1985; DeGaetano *et al*. 2000; Engeset *et al*. 2000; Flerchinger and Saxton 1989; Jansson and Karlberg 2001; Jordan 1991; Kokkonen *et al*. 2006; Melloh 1999; Vehvilainen 1992). We implemented and tested different algorithms for snow cover and soil frost already being applied for Nordic conditions. Based on preliminary modelling work, including site-specific model calibration, we developed a new snow module using ideas from a snow model currently being used by the Norwegian Water Resources and Energy Directorate (NVE) (Engeset *et al*. 2000). The NVE model has 10 parameters, and is used throughout Norway for operational snow forecasts. This model simulates snow accumulation based on daily precipitation rates and daily mean air temperature. Snow melt is a function of a degree-day temperature index described by a sinusoidal curve and daily mean air temperature. The NVE model is mainly designed for hydrological purposes (hydroelectricity production and spring flood warnings), and thus simulates the liquid water equivalent of snow SWE (mm) and snow melt runoff, but not snow depth.

Different models for simulating snow accumulation and snow melt are described in the literature, ranging from hydrological purposes (Engeset *et al*. 2000; Jordan 1991; Kokkonen *et al*. 2006) to agricultural and hydrological applications (Flerchinger and Saxton 1989) and soil-plant-atmosphere systems (Jansson and Karlberg 2001). These models simulate point estimates of a single-layered homogeneous 1-D (z-direction) snow cover, whereas Jordan (1991) presents a multi-layered 1-D snow model. Melloh (1999) provides a review of several snow melt models. Comprehensive state-of-the-art snow models such as the COUP model (Jansson and Karlberg 2001) (graphical user interface) and the SNTHERM model (Jordan 1991) (FORTRAN-77 code) are very complex and rich in parameters (> 100). The COUP model was considered as a potential candidate early in the project, but the model version available at that time required a special graphical user interface and therefore could not be incorporated into the grassland model, which was developed using another programming environment (MATLAB® and Simulink®). The combination of a special user interface and extensive data requirements (as regards number of parameters and driving climate variables) makes it very difficult to incorporate state-of-the-art snow cover and soil frost models as sub-modules into other models. The ability to incorporate a snow and soil frost model into a larger grassland model was our main motivation for developing a new model. Our proposed model is simple; it only requires nine calibrated parameters and two

input variables to simulate daily values of the depths of snow cover, soil frost and surface ice and the temperature between the soil surface and the snow cover.

A study comparing four models simulating soil frost (Kennedy and Sharratt 1998) (the two finite difference models SHAW and SOIL, and the two energy balance models by Benoit and Gusev, respectively) concluded that the simpler energy balance models generally overestimate frost depth. However, one weakness of all four models compared by Kennedy and Sharratt (1998) is the estimation of snow depth (the Gusev model uses snow depth as input). Snow cover has a strong influence on the estimation of soil frost depth, e.g. through snow depth and snow density; both affecting the thermal conductivity of the snow cover. Therefore accurate simulation of snow cover is important for the simulation of soil frost depth.

As regards modelling the formation of ice on the soil surface, we did not find examples in the literature of models simulating this process or ice encapsulation of the ground vegetation.

Following the conclusions by Kennedy and Sharratt (1998), the present work describes a new model that simultaneously simulates the depths of snow and soil frost and surface ice and explains how it was calibrated for sites across Norway using Bayesian methods. We also conducted a sensitivity analysis of the model using the Morris method, which identifies the parameters to which the model is most sensitive.

Materials and methods

The snow model

Our snow module is based on ideas presented by Engeset *et al.* (2000) and Melloh (1999, and references therein). While snow models used for hydrological purposes usually simulate snow water equivalent (SWE), SnowFrostIce also simulates the actual depth of the snow cover S_{depth} (m). To run SnowFrostIce, the only required meteorological inputs are daily values of mean air temperature T_{air} ($^{\circ}C$) and precipitation rate P (mm d-1). These parameters, which need to be locally calibrated, are listed in Table 1.

In SnowFrostIce the precipitation form is determined by a threshold temperature T_{rs} ($^{\circ}C$). If $T_{air} > T_{rs}$, precipitation fall as rain P_r (mm d-1). Otherwise it falls as snow P_s (mm d-1), having the density ρ_{ns} (kg m-3). There is no intermediate form as sleet. The snow cover consists of water in solid state S_{dry} (mm) (snow and ice), and liquid state S_{wet} (mm). The threshold temperature T_{mf} ($^{\circ}C$) determines whether snow is melting M (mm d-1), when $T_{air} < T_{mf}$, or liquid water within the snow cover is refreezing M_{rf} (mm d-1), when $T_{air} < T_{mf}$. The numerical values of T_{rs} and T_{mf} are sampled from the posterior distribution obtained in the Bayesian calibration. Since the model is calibrated locally, the estimates of T_{rs} and T_{mf} are different for each location. Instead of using a constant melt rate (mm snow melt per degree celsius and day, also known as the degree-day temperature index method), we use a degree-day temperature index K (mm $^{\circ}C$-1 day-1), which is described by a sinusoidal curve (see Equation (4)). The reason for describing K by a sinusoidal curve is to incorporate the seasonal variation. Incoming radiation increases and albedo decreases in spring. In Norway, located between latitudes 58° N and 71° N in the northern hemisphere, the dates of the solstice are December 21 and June 21.

Table 1: Parameter description for the SnowFrostIce model. θ^{min} and θ^{max} represent parameter lower and upper boundaries; θ^{mode} and θ^{def} represent parameter mode and default values, respectively. When θ_i^{mode} values are presented, a beta prior distribution is used for parameter θ_i, otherwise a uniform prior distribution is assumed between θ^{min} and θ^{max}.

Symbol	Unit	θ_i^{min}	θ_i^{max}	θ_i^{mode}	θ_i^{def}	References
T_{rs}	°C	-5	5	0.5	0.5	Engeset et al. (2000)
T_{mf}	°C	-5	5	0.5	0.5	Engeset et al. (2000)
ξ	mm mm-1 day-1	0	1	-	0.02	Thorsen and Haugen (2007)
ΔK_{max}	mm °C-1 day-1	0	5	1.25	1.25	Engeset et al. (2000)
K_{min}	mm °C-1 day-1	0	5	2	2	Engeset et al. (2000)
SW_{rf}	mm °C-1 day-1	0	5	0.01	0.01	Engeset et al. (2000)
ρ_{ns}	kg m-3	10	250	-	100	Judson and Doesken (2000)
SW_{ret}	mm mm-1	0	1	0.1	0.1	Engeset et al. (2000)
λ_{fs}	J m-1 °C-1 day-1	$8.6 \cdot 10^4$	$21.6 \cdot 10^4$	-	$17.3 \cdot 10^4$	Jansson and Karlberg (2001)

The sinusoidal curve is therefore defined as having a period of one year; with a trough, termed K_{min} (mm °C-1 day-1), on December 21, and a crest, termed K_{max} (mm °C-1 day-1), on June 21. The simulated snow melt intensity M is proportional to the number of degrees above T_{mf} (see Equation (5)). To avoid situations such as $K_{max} < K_{min}$ during the calibration, we replaced K_{max} by $\Delta K_{max} = K_{max} - K_{min}$ and calibrated ΔK_{max} (mm °C-1 day-1) (see Table 1).

Liquid water within the snow cover may refreeze. The simulated refreezing intensity M_{rf} is proportional to the number of degrees below T_{mf} (see Equation (6)) where SW_{rf} (mm °C-1 d-1) is a degree-day temperature index for refreezing. We calculated the potential retention capacity of the snow cover as $SW_{ret} * S_{dry}$ where SW_{ret} (mm mm-1) is the retention capacity of the snow cover. The snow water equivalent SWE is defined as the sum of S_{dry} and S_{wet}, and the density of the snow cover ρ_s (kg m-3) is defined as SWE/S_{depth}. As snow is accumulated on the surface of a snow cover, there is a rapid metamorphosis as snow crystals break down, and at lower snow depths densification occurs at a slower rate which is largely determined by the overburden pressure (Gray and Morland 1995). In SnowFrostIce we make the assumption that the combined effects of the metamorphosis of snow crystals and the densification of the lower snow layers is captured by the empirical compaction parameter ξ (mm mm-1 d-1). We use the following Equations (1-6) to describe the snow cover dynamics:

$$\frac{\Delta S_{dry}}{\Delta t} = P_s + M_{rf} - M \quad (1)$$

$$\frac{\Delta S_{wet}}{\Delta t} = P_r + M - M_{rf} \quad (2)$$

$$\frac{\Delta S_{depth}}{\Delta t} = \frac{P_s}{\rho_{ns}} - \frac{M}{\rho_s} - \xi S_{depth} \quad (3)$$

$$K = \frac{\Delta K_{\max}}{2} \sin\left(\frac{2\pi t}{365} + \frac{3}{8}\pi\right) + \left(K_{\min} + \frac{\Delta K_{\max}}{2}\right) \quad (4)$$

$$M = K\left(T_{air} - T_{mf}\right) \quad (5)$$

$$M_{rf} = SW_{rf}\left(T_{mf} - T_{air}\right) \quad (6)$$

Snow model parameters to be calibrated are listed in Table 1.

The soil frost model

When modelling soil frost we use an energy balance approach. Our simple approach does not include an annual energy budget for the soil system. SnowFrostIce simulates only the lower frost boundary F_{depth} (m), resulting in one frozen soil layer ranging from the soil surface to F_{depth}. For the soil water balance we use the routines implemented in the grassland model by Höglind et al. (2001) to obtain daily values of available soil water content x_w (m3 m-3) (i.e. what is left from surplus liquid water after transpiration and evaporation is subtracted), which is used in the calculation of F_{depth}. The soil layer is parameterised as in that grassland model. SnowFrostIce requires site-specific soil type parameters for soil water retention, but the only soil parameter to be calibrated is the thermal conductivity of the frozen soil λ_{fs} (J m-1 °C -1 d-1).

Our way of estimating the lower frost boundary F_{depth} is based on certain assumptions. Regarding surface temperature, we follow along the lines of the assumption made by Benoit and Mostaghimi (1985) that in any given 24-h period, the mean surface temperature of the soil or snow cover can be approximated by the daily mean air temperature for that same period. However, instead of using the daily mean air temperature at the snow cover surface (as Benoit and Mostaghimi (1985)) when calculating F_{depth}, whenever a snow cover is simulated we use a simulated soil surface temperature T_{surf} (°C) from Equation (15) as an approximation to the soil surface temperature to incorporate the insulating effect of the snow cover (note to Equation (7): during snow-free periods we assume T_{surf} can be approximated by T_{air}). We assume a unidirectional stationary flow of heat between F_{depth} and the soil surface, ignoring additional heat from e.g. lower unfrozen soil layers, percolating water, radiation and no freeze-point depression. We further assume a linear variation in soil temperature $T(z)$ (°C) with respect to soil depth z (m) in the frozen soil layer, and that all available soil water x_w within this layer freezes. It is the temperature difference between the soil surface and F_{depth} that drives the process of soil frost formation in the model:

$$T(z) = T_{surf} + z\frac{T^* - T_{surf}}{F_{depth}} \quad (7)$$

where T_{surf} is the simulated temperature just above the soil surface, T^* (°C) is the temperature where soil water freezes (we assume $T^* = 0$ °C). Following the assumption regarding $T(z)$, Equation (7) is only valid when $F_{depth} > 0$. We denote the heat flux density released when the soil water freezes Q_E (J m-2 d-1). Following an existing idea (Thorsen and Haugen 2007), we express Q_E using the above assumptions as:

$$Q_E = -x_w \rho_w L_f \frac{\partial F_{depth}}{\partial t} \quad (8)$$

where x_w is available soil water content, ρ_w (1000 kg m-3) is density of water and L_f (335 kJ kg-1) is latent heat of fusion. When the soil cools down during autumn and winter, the heat released Q_E when the soil frost penetrates deeper into the soil is transported through the previously frozen soil. Using Fourier's Equation for heat transport in one-dimensional form, we express the heat transport through the frozen soil, termed Q_{fs} (J m-2 d-1), as:

$$Q_{fs} = -\lambda_{fs} \frac{\partial T(z)}{\partial z} \quad (9)$$

From the assumption of linear variation in soil temperature $T(z)$ with depth z in frozen soil, we obtain $\frac{\partial T(z)}{\partial z}$ from Equation (7), and insert this into Equation (9):

$$Q_{fs} = -\lambda_{fs} \frac{T^* - T_{surf}}{F_{depth}} \quad (10)$$

Equating Equations (8) and (10) and using the assumption $T^* = 0$ °C, we obtain an algebraic expression for the rate of change in F_{depth}:

$$\frac{\partial F_{depth}}{\partial t} = -\frac{\alpha}{F_{depth}} \quad (11)$$

where $\alpha = \frac{\lambda_{fs} T_{surf}}{x_w \rho_w L_f}$. If we neglect the diurnal variation in T_{surf} and x_w, and consider Equation (11) as $\frac{\mathrm{d}F_{depth}}{\mathrm{d}t} = -\frac{\alpha}{F_{depth}}$, by solving this Equation we can express the daily increase in frost depth as $F_{depth}^{(t+1)} = \sqrt{\left(F_{depth}^{t}\right)^2 - 2\alpha}$. Provided $\left(F_{depth}^{(t)}\right)^2 - 2\alpha > 0$, we can express the rate of change in F_{depth} as follows:

$$\frac{\Delta F_{depth}}{\Delta t} = \frac{1}{\Delta t}\left(\sqrt{\left(F_{depth}^{(t)}\right)^2 - 2\alpha} - F_{depth}^{(t)}\right) \quad (12)$$

The presence of a snow cover has an insulating effect on the soil. Following Jansson and Karlberg (2001) we assume a steady state heat flow through the frozen soil layer and the snow cover. The heat flux density through the frozen soil Q_{fs} from Equation (10) thereby equals the heat flux density through the snow cover Q_{snow} (J m-2 d-1):

$$Q_{fs} = Q_{snow} \quad (13)$$

$$-\lambda_{fs} \frac{T^* - T_{surf}}{F_{depth}} = -\lambda_s \frac{T_{surf} - T_{air}}{S_{depth}} \quad (14)$$

where λ_s (J m-1 °C -1 d-1) is the thermal conductivity of the snow cover. The parameter λ_s is treated as a constant and not calibrated. According to Jansson and Karlberg (2001), a reasonable estimate for the ratio λ_{fs}/λ_s in our situation is $\lambda_{fs}/\lambda_s \approx 10$. We rearrange the above Equation to derive the following approximation of T_{surf}:

$$T_{surf} \approx T_{air} / \left(1 + 10\left(S_{depth} / F_{depth}\right)\right) \quad (15)$$

(note: during calculations $F_{depth} > 0$ when soil frost is present). In the case of an existing snow cover but no soil frost ($F_{depth} = 0$), we assume T_{surf} to lie around 0°C. This assumption is in accordance with observations made by Iwata et al. (2008), and it is incorporated by an additional empirical expression preserving the insulating effect of the snow cover:

$$T_{surf} \approx T_{air} \mathrm{e}^{\left(-\gamma S_{depth}\right)} \quad (16)$$

where the empirical parameter γ (m-1) is set to 65. This parameter γ is not calibrated.

Puddle formation and infiltration of meltwater

Since we were unable to obtain topographical information for any location during this study, we assume the hypothetical field of interest to be an even, rectangular surface sloping at a low angle towards a water-blocking barrier at the lower end. The height of this barrier determines the maximum depth of the surface puddle. This maximum storage level is set to 50 mm. Baker and Spaans (1997) report that infiltration from puddles can occur despite the presence of a frozen soil layer of 20-40 cm. Based on this observation, surface water (snow melt and rain) in SnowFrostIce is allowed to infiltrate into the soil if F_{depth} < 20 cm. This assumption is also confirmed by Iwata et al. (2008). In reality, the surface water transfers heat to the soil, and since the frozen soil initially remains cold this may create a thin ice layer at the soil surface which impedes water infiltration and increases surface runoff (Stähli et al. 2004). Therefore, when F_{depth} penetrates below 20 cm, we assume that the soil becomes impermeable to any further infiltration and the surface water is re-directed to the puddle area. If the maximum depth of the barrier at the end of the field is exceeded, the additional surface water runs off. When the soil starts thawing we let the infiltration rate of the puddle water follow the thawing rate (according to observations by Hayashi et al. (2003)) until F_{depth} > 20 cm, when the remaining puddle water is drained as if the soil were unfrozen.

Formation of ice layer

When a surface puddle is formed, the water may freeze and form a basal ice layer. By regarding the puddle as an extremely dilute soil and setting the water content to unity, we use the same approach to calculate I_{depth} (mm) as we do for the soil frost. Provided $\left(I_{depth}^{(t)}\right)^2 - 2\beta > 0$, we get the following expression for the daily change in I_{depth}:

$$\frac{\Delta I_{depth}}{\Delta t} = \frac{1}{\Delta t}\left(\sqrt{\left(I_{depth}^{(t)}\right)^2 - 2\beta} - I_{depth}^{(t)}\right) \quad (17)$$

where $\beta = \frac{\lambda_i T_{surf}}{\rho_w L_f}$, the thermal conductivity of ice is λ_i ($19.4 \cdot 10^4$ J m-1 °C -1 d-1), the density of water is ρ_w and the latent heat of fusion is L_f.

Description of locations and data used in calibration

The SnowfrostIce model was calibrated using observed depths of snow cover and lower frost boundary. The snow cover depth was measured in cm according to the Norwegian Meteorological Institute. The depth of the lower frost boundary was measured in cm using a frost tube (as used by DeGaetano *et al.* (2001) and described in Iwata *et al.* (2008)). We were unable to obtain information on the accuracy of the observations. We were also unable to obtain information on normal depths of snow cover and soil frost. We therefore present values of mean air temperature and precipitation sums from autumn to spring and frost sums. Table 2 presents a geographical description of the locations, and Tables 3-6 provide a summary of the climate for each location for the current normal period in Norway (1961-1990) and for the calibration and validation periods. For each location we calculated the following from autumn to spring (i.e. from 1 September to 30 April): the mean 2 m air temperature, denoted mean(T_{air}); the temperature sum for days when $T_{air} < 0$, denoted $\sum T_{air}$; the sum of daily precipitation rates, denoted $\sum \mathrm{Prec}$.

Table 2: Locations used for calibrating the SnowFrostIce model

Location	Grid	Elevation (m.a.s.l.)	Climate	Measurement period Calibr.	Validation
Kise	60° 77' N 10° 8' E	127	Interior, lake	1993-1996	1996-1999
Kvithamar	63° 49' N 10° 88' E	40	Coastal	2001-2003	2003-2005
Vågønes	67° 28' N 14° 45' E	30	Coastal	1998-2001	2001-2003
Holt	69° 65' N 18° 91' E	20	Coastal	1996-1999	2005-2007
Karasjok	69° 28' N 25° 31' E	149	Interior	-	1998-1999

During the calibration period at Kise (Table 3), the first and third winter were both colder and had more frost compared with the normal period. The second winter was milder and had less frost. The first winter received more precipitation compared with the normal period, while the latter two were dryer. In the validation period all winters were slightly milder and had less frost than normal; the first winter was dryer than normal while the latter two were wetter.

Table 3: Climate summary for Kise. Values are calculated for the months September-April for the current normal period in Norway (1961-1990), and for the respective calibration and validation periods. Mean(Tair) (°C) is the average 2 m air temperature, ΣT_{air} (°C day) is the temperature sum on frost days and Σ *Prec* (mm) is the precipitation sum

Sept - Apr	61/90	93/94	94/95	95/96	96/97	97/98	98/99
Mean(T_{air})	-1	-2.5	1.1	2.2	0.8	1.2	0.2
ΣT_{air}	-761	-1068	-400	-1214	-629	-439	-611
Σ *Prec*	340	368	294	188	273	421	436

At Kvithamar (Table 4), both winters in the calibration period were milder than normal, but they had more frost. The first winter was wetter, and the second winter was dryer than normal. In the validation period, both winters was milder and wetter compared with the normal period.

Table 4: Climate summary for Kvithamar. See caption to Table 3 for explanations

Sept - Apr	61/90	01/02	02/03	03/04	04/05
Mean(T_{air})	1.5	3.2	1.7	3.2	3.3
ΣT_{air}	-269	-272	-385	-245	-225
Σ *Prec*	597	682	508	604	891

At Vågønes (Table 5), all winters in the calibration period were milder compared with the normal period, but the first and third winters had more frost days, while there were fewer frost days in the second winter. The first two winters were wetter, and the third winter was dryer than normal. In the validation period, both winters had more frost than normal, but only the first winter was milder than normal. The first winter was wetter than normal, and the second was dryer.

Table 5: Climate summary for Vågønes. See caption to Table 3 for explanations

Sept - Apr	61/90	98/99	99/00	00/01	01/02	02/03
Mean(T_{air})	1.3	1.9	2.6	2.3	2.6	1.2
ΣT_{air}	-284	-323	-264	-368	-330	-372
Σ *Prec*	811	902	1156	561	983	735

At Holt (Table 6), all winters in the calibration period were milder and had more frost than normal. The first winter was wetter while the latter two were dryer than normal. Both winters in the validation period were milder, had less frost and were wetter when compared with the normal period.

Table 6: Climate summary for Holt and Karasjok. Values within brackets represent the normal period for the Karsjok location. 98/99† represents the Karasjok location. See caption to Table 3 for explanations

Sept - Apr	61/90	96/97	97/98	98/99	05/06	06/07	98/99†
Mean(T_{air})	-0.8 (-8.3)	-0.1	0.2	0.3	1.6	1.1	-8.5
$\Sigma\ T_{air}$	-375 (-2199)	-468	-483	-432	-317	-322	-2295
Σ *Prec*	765 (172)	804	627	578	831	817	207

The winter in the validation period at Karasjok (Table 6) was approximately the same as the normal period, but slightly wetter.

In addition to simulating S_{depth} and F_{depth}, SnowFrostIce simulates the thickness of ice I_{depth} resulting from the freezing of soil surface puddles. However, data on surface ice were scarce and there was no description of field topography available, forcing us to make assumptions on field topography. Thus we present full simulation results for only two locations; Holt in Troms county and Karasjok in Finnmark county. Based on data availability, we chose four locations for site-specific calibration of the model spanning the south-north variation in regional climate. Table 2 gives a brief description of these locations. Karasjok was not included in the calibration.

Observations of surface ice cover were scarce and data were only available for two sites: Holt (97/98, 98/99) and Karasjok (98/99). Ice observations from Holt came at a later stage in the project, and thus we had to use observations on snow cover and frost depth from the calibration period.

Bayesian calibration of the SnowFrostIce model

The SnowFrostIce model represents a simplification of different physical processes. Parameters used in process-based models have a physical meaning, but these are seldom precisely known, or are at best difficult to measure. We represented this uncertainty as a probability distribution over the parameters. Thus, if we define a parameter vector $\boldsymbol{\theta}$ for the model, then $\pi(\boldsymbol{\theta})$ is said to be a joint probability density function (pdf) expressing our initial prior belief in the parameters. Given a data set **D** of model outputs, we update the joint pdf of the parameters by applying the Bayes Theorem: $\pi(\theta \mid \mathrm{D}) = \pi(\theta) f(\mathrm{D} \mid \theta) / f(\mathrm{D})$ where $\pi(\boldsymbol{\theta}|\mathbf{D})$ is the posterior distribution of $\boldsymbol{\theta}$ given the data **D**, $f(\mathbf{D}|\boldsymbol{\theta})$ is the likelihood of the data given the model outputs using parameters $\boldsymbol{\theta}$, and $f(\mathbf{D})$ is a normalisation constant. In Bayesian calibration of dynamic models, a large number of model runs are carried out, often in a Markov Chain Monte Carlo (MCMC) approach. We used the MCMC algorithm known as the Metropolis Random Walk. For further details on using Bayesian methods to calibrate complex models see Van Oijen *et al.* (2005b). The target posterior distribution was the stationary distribution of the Markov chain produced by the Metropolis Random Walk.

Metropolis Random Walk

The general idea is to randomly walk through the parameter space, running the model at each visited point, eventually forming a Markov chain. The starting point of this chain $\boldsymbol{\theta}_0$ is randomly chosen from the prior distributions for the parameters.

A new proposal parameter vector $\boldsymbol{\theta}'$ is then chosen based on the current parameter vector $\boldsymbol{\theta}^t$

$$\theta' = \theta^t + \delta \qquad (18)$$

where $\boldsymbol{\delta}$ is the step length vector. It is also important that $p(\boldsymbol{\delta}) = p(-\boldsymbol{\delta})$ i.e. there is equal probability of stepping in either direction from the current point. We then compute the so-called Metropolis ratio

$$r = \frac{\pi(\theta' \mid \mathrm{D})}{\pi(\theta^t \mid \mathrm{D})} = \frac{\pi(\theta')f(\mathrm{D} \mid \theta')}{\pi(\theta^t)f(\mathrm{D} \mid \theta^t)} \qquad (19)$$

The next step is to generate a uniform random number $u \sim U(0,1)$ and accept the proposal parameter vector $\boldsymbol{\theta}'$ as the new $\boldsymbol{\theta}^{t+1}$ if $u \le r$. Otherwise let $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. The chain consisting of all $\boldsymbol{\theta}^t$ forms our Markov chain, which is our sample from the posterior distribution.

The posterior distribution is thus a combination of prior knowledge and new information obtained from the data using the likelihood function. Measurement errors are used in the determination of how likely a model-data mismatch might be, i.e. if the data are informative and have a sharply peaked distribution (i.e. a small variance), the resulting posterior distribution will be narrower and more peaked than the prior distribution. This indicates that the parameter uncertainty is reduced.

Defining prior probability distributions of the parameters

Based on literature review, we defined the likely ranges $\left[\theta_i^{\min}, \theta_i^{\max}\right]$ and mode values for the nine parameters. For parameters where range and mode value were suggested, we used a beta distribution as prior. For parameters ξ, ρ_{ns} and λ_{fs} only a suitable range was found. For these three parameters we selected a flat uniform distribution within their range $\left[\theta_i^{\min}, \theta_i^{\max}\right]$. In the calibration process we assumed the parameters to be independent *a priori*, implying that their joint prior distribution is equal to the product of their individual marginal pdfs. The parameters, along with their prior distributions, are presented in Table 1.

Defining the data-likelihood function

We used measurements on snow depth and lower frost boundary for the calibration of SnowFrostIce. Specific information about the precision of the measurements was not available, and thus we used the same approach as Van Oijen *et al.* (2005b) and chose the standard deviation of each measurement to be 30% of the mean value. To avoid a standard deviation of zero (if the observed variable was zero), the standard deviation was redefined as $\sigma_{ij}^o = \max\left(0.1; 0.3 \cdot \mathrm{D_{ij}}\right)$ where $\mathbf{D}_{ij}$ are the measurements on output j at time i. Assuming the measurement errors to be independent and Gaussian, we used Sivia's formulation (Sivia 2006) which was slightly modified to account for model discrepancy:

$$f(\mathrm{D} \mid \theta) = \prod_{j=1}^{G}\left[\prod_{R_{ij}\neq 0}\prod_{i=1}^{M}\frac{1}{\sigma_{ij}^o\sqrt{2\pi}}\left[\frac{1-\mathrm{e}^{-R_{ij}^2/2}}{R_{ij}^2}\right]\prod_{R_{ij}=0}\frac{1}{2\sigma_{ij}^o\sqrt{2\pi}}\right] \qquad (20)$$

where σ_{ij}^{o} represent the lower bounds on the data noise and the residual is represented by $R_{ij} = \left(\mathrm{D}_{ij} - M_{ij}(\theta, X)\right) / \sigma_{ij}^{o}$, where $M_{ij}\left(\theta^{*}, X\right)$ are model outputs using input variables **X** and parameterisation $\boldsymbol{\theta^{*}}$.

Determining jumps in the Metropolis Random Walk algorithm

The step length vector $\boldsymbol{\delta}$ in the Metropolis Random Walk algorithm is very important in order to obtain convergence of the Markov chain produced, i.e. the targeted posterior distribution of the parameters. In our implementation the new candidate value θ_i' for parameter *i* was $\theta_i' = \theta_i^t + \delta_i$, where $\delta_i \sim N(0, a_i)$. If the elements in the step length vector $\boldsymbol{\delta}$ are too small, the random walk algorithm will not move far enough from the current point in parameter space $\boldsymbol{\theta}^t$ when proposing a new candidate parameter vector $\boldsymbol{\theta'}$, and consequently the acceptance rate will be too large; and *vice versa*. In our case, choosing a_i so that the acceptance rate was between 0.15 and 0.5 (according to Roberts (1996)) was done by trial and error. Each element a_i of the vector $\boldsymbol{\delta}$ was chosen according to $a_i = c_i\left(\theta_i^{\max} - \theta_i^{\min}\right)$, where c_i is a constant found by trial and error, and $\left(\theta_i^{\max} - \theta_i^{\min}\right)$ is the width of prior pdf of parameter $\boldsymbol{\theta}_i$.

Determining convergence of the Markov chains

A central issue when using an iterative simulation method such as the Metropolis Random Walk algorithm is to determine when the chain has converged to the desired posterior distribution. One option, suggested by Gelman and Rubin (1992), is to generate multiple chains followed by calculation of the scale reduction factor $\sqrt{\hat{R}}$, which is used to determine the length of the "burn-in" phase. The "burn-in" of the chain is the first part where the chain is influenced by the starting point until it reaches stationarity. We determined the "burn-in" phase to last until $\sqrt{\hat{R}} < 1.2$ according to Gelman (1996); when $\sqrt{\hat{R}}$ is near 1 it means that the Markov chains are essentially overlapping. We randomly sampled two starting points from the prior distribution, and used the $\sqrt{\hat{R}}$ to determine when the two chains had converged to the desired posterior distribution.

Sensitivity analysis of SnowFrostIce

When working with models, sensitivity analysis (hereafter referred to as SA) is recommended as part of the process (Kokkonen *et al*. 2006). For the SA to be meaningful, the practitioner should decide beforehand on how to define the importance of the parameters, i.e. the type of question the SA is expected to answer (Saltelli *et al*. 2008). In our case, we would like to know which of the parameters can be fixed anywhere within their prior bounds without affecting model outputs, i.e. which parameters are not important. This is helpful in relation to model simplification.

In order to identify non-important parameters in the model, we carried out a screening exercise using the improved sensitivity indices from the Morris method as described by Campolongo *et al*. (2007). This method is relatively simple to implement.

The Morris method proposes two sensitivity measures, the main purpose of which is to determine the model k parameters that can be considered to be either (i) not important, (ii) linear and additive, or (iii) non-linear or involved in interactions with other parameters. For each of the parameters, two sensitivity measures are computed; μ, which evaluates the overall influence of the parameter on the model output (main effect, or elementary effect EE), and σ, which evaluates collectively all the higher order effects due to non-linearity and/or to interaction with other parameters. The Morris method was originally used for parameters following uniform distributions in [0, 1]. If the k parameters follow other distributions, Campolongo *et al.* (1999) suggest that rather than sampling directly from these distributions, the sampling should be performed in the space of the quantiles of these k distributions (i.e. each parameter is discretised into p levels, and each quantile q_p varies in [0,1], producing as sampling space a k-dimensional unit hypercube). The actual parameter values would subsequently be derived from their known distributions. In this SA of SnowFrostIce, we investigated the k = 9 parameters from the calibration (Table 1). The input space we used was the sub-space Ω comprised of the k-dimensional unit hypercube of the p = 6 equidistant quantiles in [0,1] from the prior distribution of the parameters $\pi(\theta)$. Outputs from SnowFrostIce are time series, and for this SA we needed a scalar value. Thus, for the simulation runs required in the SA, we used as output the log-transformed likelihood from Equation (20), i.e. $\log(f(\mathbf{D}|\boldsymbol{\theta}))$; the likelihood being the probability of the observed data $\mathbf{D}$ given a certain model parameterisation $\boldsymbol{\theta}$.

By randomly sampling parameter vectors $\boldsymbol{\theta}$ from Ω and calculating EE (see Campolongo *et al.* (2007) for details) for each of the nine parameters, we obtained a sample from the distribution for each EE, termed $EE_i \sim F_i(\mu_i, \sigma_i)$. The sensitivity measures μ_i and σ_i proposed by Morris are the mean and standard deviation of F_i, respectively. To estimate μ_i and σ_i, the sampling strategy proposed by Morris is to create r trajectories in parameter space Ω. Each of these r trajectories contains (k+1) points and results in k elementary effects (i.e. estimates of one EE per parameter), leading to a total of $r(k+1)$ sample points corresponding to the number of model runs required for the complete SA. A very nice stepwise presentation of this method is presented in Saltelli *et al.* (2008).

A high σ_i value for parameter θ_i implies that the corresponding EE_i-value for θ_i at one point in Ω is considerably different from another EE_j-value ($i \neq j$) for the same parameter θ_i located somewhere else in Ω, i.e. that this particular EE-value is influenced by the values of the other parameters or nonlinearities. A low value for σ_i suggests that the EE_i-value associated with θ_i is independent from the values of the other parameters and thus it is not involved in interactions or nonlinearities.

To avoid type II errors of failing to identify important parameters, Campolongo *et al.* (2007) suggest replacing μ by μ^*; an estimate of the mean of the distribution of the absolute values of the elementary effects G_i, i.e. $|EE_i| \sim G_i(\mu^*_i, \sigma_i)$. To properly characterise non-influential parameters one must therefore simultaneously consider the vectors $\boldsymbol{\mu}^*$ and $\boldsymbol{\sigma}$ (see Figure 1 for SA results for the Kise site).

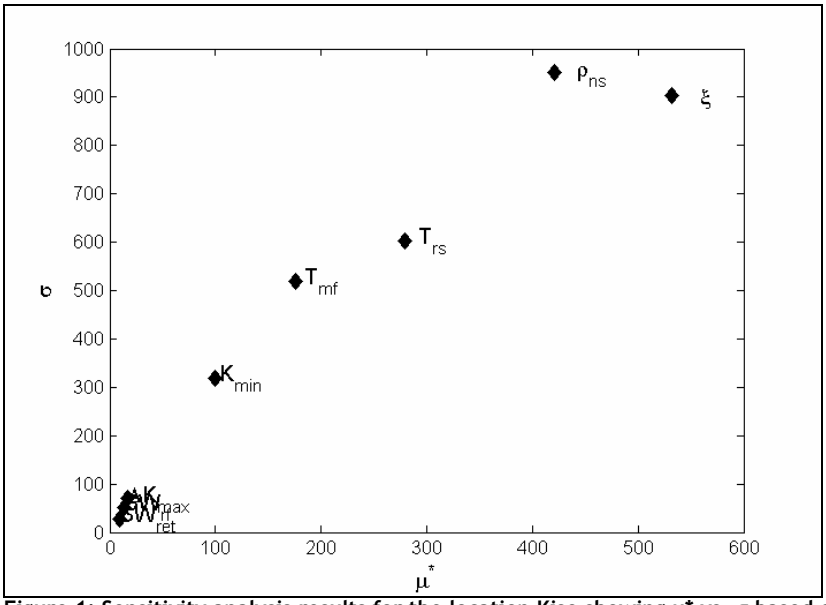


Figure 1: Sensitivity analysis results for the location Kise showing μ* vs. σ based on r = 100 trajectories. Low values for both μ* and σ identifies SWret, SWrf and ΔKmax as the least important parameters.

When conducting the SA, we tried the same approach for all locations. First we sampled trajectories from the prior distribution and calculated $\boldsymbol{\mu}^*$ and $\boldsymbol{\sigma}$. Then we sampled from the posterior distribution and calculated $\boldsymbol{\mu}^*$ and $\boldsymbol{\sigma}$. Sampling trajectories from the prior distribution gave very similar results for all sites (as those for Kise, Figure 1). When we sampled from the correlated posterior distribution the results in $\boldsymbol{\mu}^*$ and $\boldsymbol{\sigma}$ were different when comparing sites. For all but one site the same non-important parameters were identified, but highly correlated parameters influenced the results. For example, at Kise, the parameter T_{rs} was wrongly recognised as being non-important. This illustrates that the Morris method can produce different results depending on whether the parameters are correlated or not. We did not find examples in the literature of how to handle correlated parameters when using the Morris method. T_{rs} is an important parameter, as was clearly shown when sampling trajectories from the prior distribution. Based on this observation, we decided to use the assumed uncorrelated prior distribution when sampling trajectories for the screening exercise.

Model validation and predictive uncertainty

The data sets for each location were divided in two and one part was used for calibration and one for validation (Table 2). To evaluate the predictive uncertainty of the model after calibration, we sampled 20 parameter sets from the posterior distribution and calculated the subsequent mean and standard deviations of the model outputs.

Results

Results from the Bayesian calibration

The main result of the Bayesian calibration procedure is the estimated joint posterior distribution of the model parameters. This correlated multi-dimensional joint distribution is difficult to visualise, so we present the marginal posterior distribution for single parameters.

We determined the success of the calibration by evaluating the estimated marginal posterior distributions. If they are narrower than their corresponding prior distribution, this indicates that the parameter uncertainty has been reduced. The calibration at each location used two chains of length 300,000, and a unique step length vector for that location.

The part of the Markov chains succeeding the burn-in point, which we determined as the point from where $\sqrt{\hat{R}}$ remains below 1.2, comprises the marginal posterior distribution of the parameter (Gelman 1996). The right column of Figure 2 shows plots of $\sqrt{\hat{R}}$ for the parameters ξ, ρ_{ns} and T_{rs}, and the centre column shows the estimated marginal posterior distribution for the same parameters. Panels in the left column in Figure 2 show trace plots of the Markov chains for parameters ξ, ρ_{ns} and T_{rs} calibrated at the Kise site. These trace plots are used to verify that the two chains for each parameter stabilise around the same value, and that the posterior distribution is properly explored.

Table 7: Parameter values for SnowFrostIce which gave highest posterior density θ_{MAP}, and the median values $\tilde{\theta}$ for the sites Kise, Kvithamar, Vågønes and Holt

Parameter	Kise		Kvithamar		Vågønes		Holt	
θ_i	$\theta_{MAP(i)}$	$\tilde{\theta}_{(i)}$	$\theta_{MAP(i)}$	$\tilde{\theta}_{(i)}$	$\theta_{MAP(i)}$	$\tilde{\theta}_{(i)}$	$\theta_{MAP(i)}$	$\tilde{\theta}_{(i)}$
T_{rs}	-0.1	-0.1	-0.6	1	2.3	2	3.1	3
T_{mf}	-1.4	-1.5	3.1	2.1	0.7	1.3	-3	-2.3
ξ	0.02	0.02	0.025	0.12	0.15	0.13	0.01	0.02
ΔK_{max}	4.5	3.6	0.79	1.5	1.8	1.5	0.5	2
K_{min}	1.1	1	0.43	1.6	0.2	2.3	2.6	3.5
SW_{rf}	0.002	0.48	0.87	0.68	2.61	0.63	3.65	0.78
ρ_{ns}	128	124	216	89	84	95	250	231
SW_{ret}	0.32	0.07	0.21	0.22	0.18	0.22	0.35	0.2
λ_{fs} (10^4)	8.6	8.8	12.6	10.3	17.6	13.1	13.7	13.8

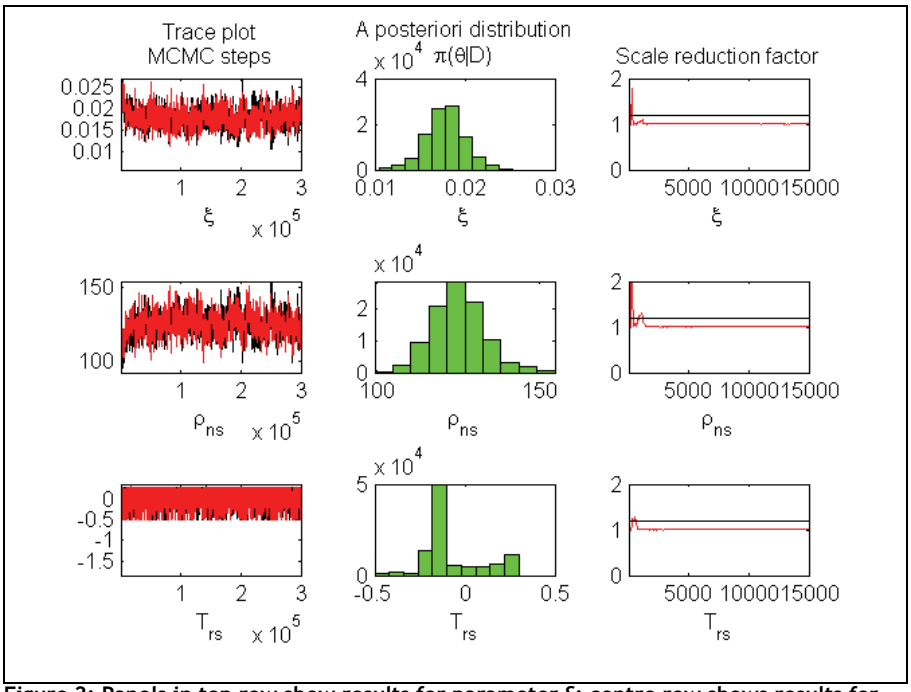


Figure 2: Panels in top row show results for parameter ξ; centre row shows results for parameter ρns; bottom row shows results for parameter T_{rs}. Panels in the left column show trace plots of the two parallel chains (red - chain 1, black - chain 2). Panels in the centre column show the marginal posterior distribution of the parameter $\pi(\theta_i|D)$. Panels in the right column show the scale reduction factor $\sqrt{\hat{R}}$, calculated at each 20th iteration.

In order to visualise the marginal posterior distributions for all locations simultaneously, we fitted continuous distributions to the samples from the posterior generated by the MCMC. They are shown, together with the prior distributions, in Figure 3. The marginal posterior distributions are either multi-modal, skewed or both. It was thus informative to present both the maximum posterior estimate and the median value of $\boldsymbol{\theta}$ ($\boldsymbol{\theta}_{MAP}$ and $\tilde{\theta}$, respectively) from the marginal posterior distributions as summary statistics (see Table 7), complemented by plots of the marginal posterior distributions in Figure 3 showing posterior parameter uncertainty. The parameter vector $\boldsymbol{\theta}_{MAP}$ represents the single best parameter vector at the different locations. For most of the parameters, when comparing the marginal posterior distributions in Figure 3 to their respective prior distribution (black lines), it is clear that the calibration process reduced the prior parameter uncertainty. However, for the parameters related to liquid water in the snow cover, SW_{rf} and SW_{ret}, we can see that measurements on snow depth alone did not provide enough information to depart from our prior estimates (for the Kise site, they are more peaked). For the precipitation threshold temperature T_{rs} the parameter uncertainty was least reduced at Kvithamar compared with the other locations. For T_{mf} the parameter uncertainty was reduced more at Kise and Holt than at Kvithamar and Vågønes. For Kvithamar the median value of T_{mf} (see Table 7) was larger than the median value of T_{rs} (this is shown in Figure 4 (b) where the green line is located slightly above the red line). The uncertainty in ΔK_{max} and K_{min} was reduced for Kise and Holt, but for Kvithamar and Vågønes there was not much

improvement. The parameter uncertainty was reduced for the remaining ξ, ρ_{ns} and λ_{fs}.

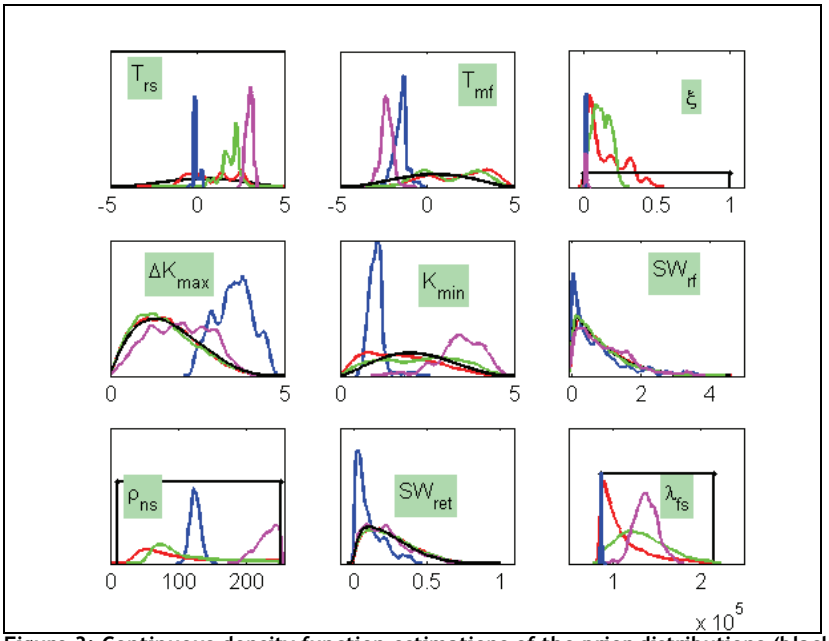


Figure 3: Continuous density function estimations of the prior distributions (black) and the marginal posterior distributions for all locations: Kise (blue), Kvithamar (red), Vågønes (green), Holt (magenta).

Results from the sensitivity analysis

At each of the locations used in the calibration, we randomly generated r = 100 different trajectories for the computation of EE; i.e. $r(k+1)$ = 1000 parameter vectors were sampled from Ω and thus 1000 model runs were used for the SA. The results were very similar for each location. Figure 1 shows the sensitivity indices $\boldsymbol{\mu}^*_i$ and $\boldsymbol{\sigma}_i$ for each parameter for the Kise site. We find the parameters SW_{ret}, SW_{rf} and ΔK_{max} in the lower left-hand corner, and the remaining parameters almost linearly spread. Inspection of histograms of the sampled parameter values suggests that the ranges of the prior intervals were adequately explored.

The parameter λ_{fs} was excluded from the SA since S_{depth} affects F_{depth} and not *vice versa*.

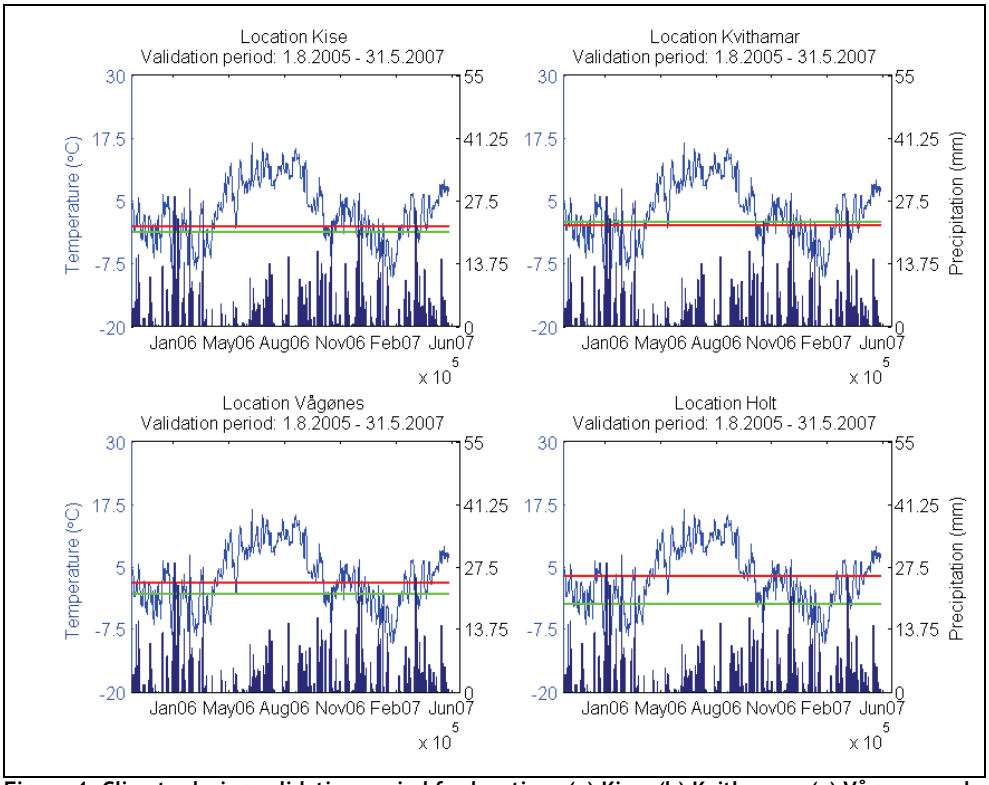


Figure 4: Climate during validation period for locations (a) Kise, (b) Kvithamar, (c) Vågønes and (d) Holt. Solid lines show T_{air} (blue) and bars show daily precipitation; median values for T_{rs} (red) and T_{mf} (green). See Table 7 for parameter values.

Validation of the model

The SnowFrostIce model was validated at all locations used in the calibration. For each of the locations we sampled 20 parameter vectors from the posterior distribution and calculated the mean and standard deviation of the model output. Variation in model output is shown as the mean ± one standard deviation. Figure 5 shows validation results for each of the four locations. The sub-figures show the variation between model output and observed values on depths of snow and soil frost. Figure 4 shows the meteorological variables during the validation period at each location; daily mean air temperature, daily precipitation rate and the median values from the posterior distribution of the threshold temperatures for precipitation T_{rs} and snow melt/refreezing T_{mf}. If the median value of T_{rs} is close to that of T_{mf} they appear as one line in the sub-figures. See Table 7 for these parameter values. The validation at the Kise site shows little variation in model output. At this site S_{depth} is overestimated during the winter of 97/98. This is as expected when considering that $T_{air} < T_{rs}$ for most of the precipitation events (see Figure 4 (a)). Frost depth at Kise during 97/98 is initiated earlier than observed, in addition to being slightly underestimated. F_{depth} during the 98/99 winter is overestimated; too high frost rates initially caused F_{depth} to be shifted downwards

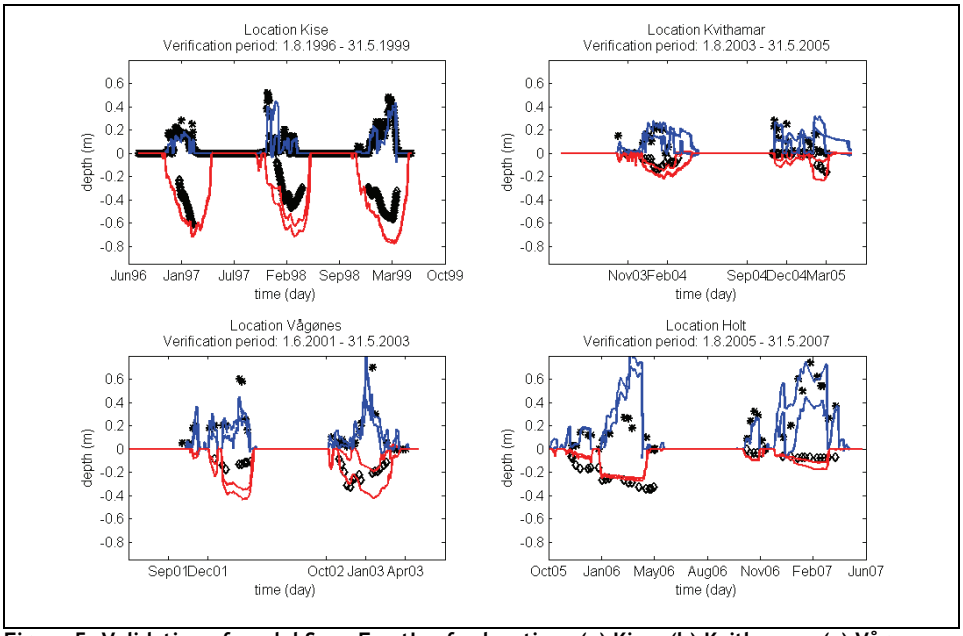


Figure 5: Validation of model SnowFrostIce for locations (a) Kise, (b) Kvithamar, (c) Vågønes and (d) Holt. Solid lines (mean ± one standard deviation) show Sdepth (blue) and Fdepth (red); observed snow cover depth (*); observed lower frost boundary (◇).

compared with observations. The validation for Kvithamar, Figure 5 (b), shows more variability in model output compared with Kise, especially towards the end of springtime for S_{depth}. The data points here are captured within this variation. At Vågønes, Figure 5 (c), model performance for S_{depth} is quite good, but F_{depth} is overestimated (more severely during 01/02 than 02/03). At Holt, Figure 5 (d), S_{depth} is overestimated during 05/06 (as with Kise 97/98) because $T_{air} < T_{rs}$ for most of the precipitation events in this winter. Note especially events between January and May 2006 with $P > 20$ mm, see Figure 4 (d), where precipitation is simulated as snow. F_{depth} looks reasonably accurate, but complete thaw is predicted too early for both validation years. Variation in model output is in general higher for Kvithamar and Holt than for Vågønes and Kise. Figure 6 shows all output (snow cover, soil frost and surface ice) for Holt (97/98) and Karasjok (98/99). We had no data to calibrate SnowFrostIce for Karasjok. For Karasjok we sampled parameter values from $\pi(\boldsymbol{\theta}|\mathbf{D})$ obtained for Kise; both locations having interior climate. Holt and Karasjok were the only locations where ice observations were available.

Discussion and conclusions

In this paper we present a new model for the simulation of snow depth, soil frost depth and depth of surface ice cover. We calibrated the model by means of well-documented Bayesian methods, and conducted a qualitative sensitivity analysis. As far as we know this practice is still relatively new for this kind of model. The results presented here, both regarding assumptions on prior pdfs and the resulting posterior pdfs, and the simple yet very effective method of sensitivity analysis, are useful for the modelling community.

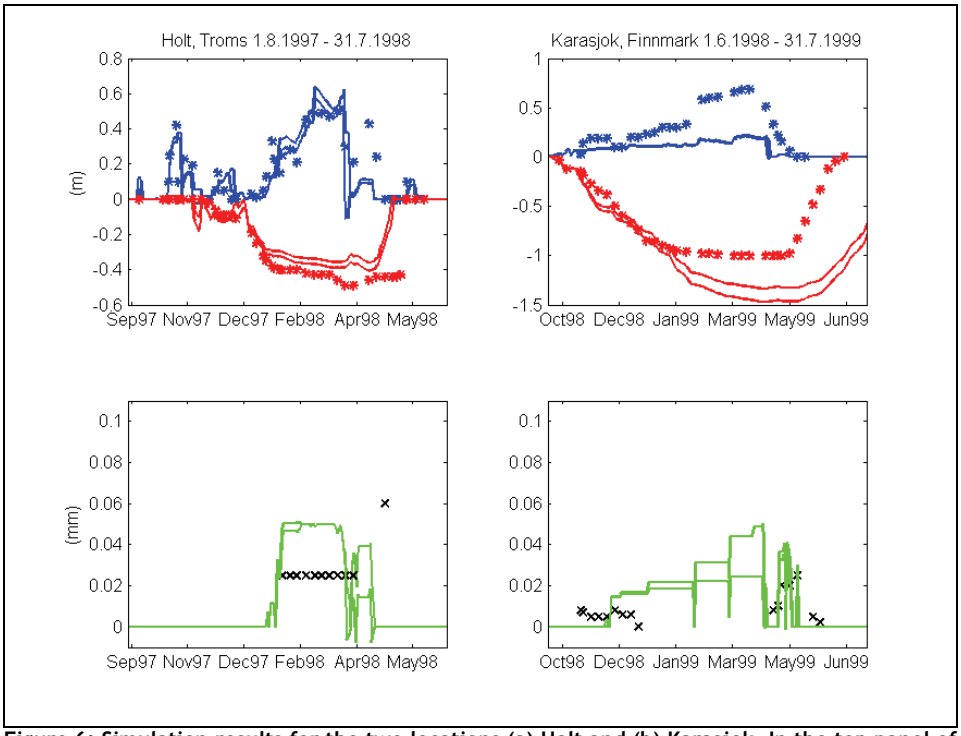


Figure 6: Simulation results for the two locations (a) Holt and (b) Karasjok. In the top panel of the sub-figures, solid lines show simulated values (mean ± one standard deviation) and (*) indicate observed values; blue represents snow cover, and red represents lower frost boundary. In the bottom panel, solid green lines represent simulated (mean ± one standard deviation) ice cover depth, and (x) indicate observed values for ice cover depth.

A study comparing four models simulating soil frost (Kennedy and Sharratt (1998)) (the two finite difference models SHAW and SOIL, and the two energy balance models by Benoit and Gusev, respectively) concluded that the simpler energy balance models generally overestimate frost depth. However, one weakness of the models (investigated by Kennedy and Sharratt (1998)) that also simulate snow cover is the estimation of snow depth. Snow cover has a strong influence on the estimation of soil frost depth, e.g. through snow depth and snow density; both affecting the thermal conductivity of the snow cover. Therefore accurate simulation of snow cover is important for the simulation of soil frost depth.

Our new model SnowFrostIce for simulating the effects of winter climate on soil surface is designed to be included in a grassland model. This restricts SnowFrostIce with regards to the number of parameters included. We calibrated SnowFrostIce against independent data from four locations in Norway, capturing climatic variation from south to north and coastal to inland. We also identified the key parameters by conducting a sensitivity analysis.

It is important to bear in mind that SnowFrostIce represents simplifications of real world processes which are described at various levels of complexity. Some of the parameters used have a physical interpretation, but they are seldom measured and quantitative data are scarce in the literature. This means the parameters and thereby also the model outputs are subjected to uncertainty. The Bayesian method we used aims to quantify and reducing these uncertainties rather

than maximizing model fit. When selecting an optimal parameter set for a simulation run for a specific location, we chose the parameter values which maximized the posterior distribution $\boldsymbol{\theta}_{\mathrm{MAP}}$ (Table 7). A consequence of this procedure was that these specific parameter values must be interpreted accordingly (i.e. reducing model uncertainty) rather than given a clear-cut physical interpretation.

When carrying out the Bayesian calibration, it was difficult to obtain convergence of the Markov chains for the parameters relating to liquid water in the snow cover (SW_{ret} and SW_{rf}). This may imply that the calibration data were not sufficient for improving the prior knowledge related to these parameters.

The estimated posterior distribution is different for each location. We expected some regional differences for the melting parameters K_{min} and ΔK_{max} due to e.g. regional differences in radiation, altitude (m.a.s.l.) and ocean vicinity, but not for the threshold temperatures for precipitation T_{rs} and snow melt/refreezing T_{mf}, or the density of new snow ρ_{ns}. This might indicate that the model needs geographical adjustments and a functional description of ρ_{ns}. The differences in the results for the thermal conductivity of frozen soil λ_{fs} were expected, since the soil types are different for each of the locations.

A reason for erroneous estimation of S_{depth} could be that the calibrated value of T_{rs} is wrong, leading to observed rain being simulated as snow or *vice versa*. In addition, by using daily mean air temperatures the model might associate incorrect air temperatures to precipitation events. For instance, observed air temperature could be below $0\,^{\circ}C$ for most of the day, followed by above $0\,^{\circ}C$ at the end of the day, resulting in mean daily temperature below T_{rs}. If precipitation had been observed as rain by the end of the day, it would still have been simulated as snow. The overestimation of S_{depth} might be due to important processes being omitted, e.g. the heat content of rain is not incorporated in the model and thus this kind of additional snow melt is not included. A third reason for erroneous estimation of S_{depth} might be redistribution of snow by wind, a factor not taken into account in the model.

The amount of available observations for the calibration is important. Using data from two and three years is not sufficient to capture the inter annual variation in snow cover and soil frost. The limited amount of observations on both snow cover and soil frost at the same location has an effect on the results from the calibration. In a preliminary study, the snow module of SnowFrostIce was calibrated for the Kise location using two, four, six, eight and finally 10 years of snow depths observations (unpubl. ms.). Including increasingly more data resulted in a more narrow posterior distribution, but convergence was also increasingly harder to obtain. Including more observations also resulted in a shift in the location of the posterior parameter distribution. This showed that the inter annual variation in winter weather will affect the results of the calibration. As long as more data are included, the results are likely to keep varying until the whole spectrum of weather situations is included. Ideally we should have had observations comprising a full climate period (thirty years) to capture the variation within a normal period. In a study comparing thirty-three snowpack models by Rutter *et al.* (2009) only two years of observations were available. In the present study, the data set was split in two in order to conduct the validation which otherwise would have to be postponed until more observations became available.

The parameters related to snow melt (T_{mf}, K_{min} and ΔK_{max}) are less uncertain for Kise than for the other locations (see Figure 3). This contributes to less

uncertainty in the snow depth simulation at Kise compared to the other locations. At Kvithamar, in addition to the uncertainty of T_{mf}, this parameter also has a high numerical value (see Table 7) compared to the other locations. This leads to more uncertainty in the melting period at Kvithamar and also a delayed onset of snow melt in the simulations compared to e.g. Kise. The results from the sensitivity analysis showed that ΔK_{max} was the least important parameter related to snow melt. It is therefore reasonable to attribute the uncertainty and delay in snow melt mainly to the uncertainty of the parameters T_{mf} and K_{min}.

In this study we used the likelihood of a sampled parameter set given the data (see Equation (20)) as a scalar output when calculating the sensitivity indices μ^* and σ. If we on the other hand were to use e.g. daily simulated snow depth values as scalar output in the SA, we would have to calculate one pair of μ^* and σ for each of these S_{depth} values. This would provide an answer to the question of which parameters that were most important on which day during the whole simulation period. However, by performing two SA where the first SA uses depth of snow cover on a specific day during midwinter, and the second SA uses depth of snow cover on a specific day towards the end of winter might give further indications on which parameters are most important regarding snow melt in cold and mild periods, respectively.

The purpose of our SA was to identify key parameters in the model. Here we used the Morris screening method to identify the non-important parameters. In Figure 1, the parameters SW_{rf}, SW_{ret} and ΔK_{max} are recognised as being less important (low values for μ^* and σ). Following how we defined parameter importance in our SA, the SA results suggest that varying these parameters within their prior bounds would not markedly affect model output. We can also find support for this conclusion in the calibration results. Figure 3 shows that the posterior distribution for the three non-important parameters has not changed much compared with the prior distribution, implying that no new information is added through the data.

The ability of SnowFrostIce to simulate the maximum depth of the snow cover (solid black circle ●), and the maximum depth of the soil frost (solid black square ■) is presented in Figure 7. Each pair of points (● and ■) represent one of the winters from the validation periods. Thus, three pairs correspond to Kise, two pairs each to Kvithamar, Vågønes and Holt, and one pair to Karasjok. The points lie close to the 1:1 line, indicating satisfactory model performance. The maximum depths of snow cover and soil frost are good indices to show the trends of the snow cover and soil freezing in each winter, and they are both appropriately estimated by the model.

The approach to calculating soil frost, by balancing energy, is similar to that proposed by Benoit and Mostaghimi (1985). Our soil frost model, although we made some critical assumptions (e.g. estimation of the soil surface temperature, a constant thermal conductivity of frozen soil, a constant thermal conductivity of snow), is even simpler than the Benoit model and, as shown when tested on independent data sets (see Figure 7), the ability of SnowFrostIce to estimate maximum lower frost boundary is also quite good.

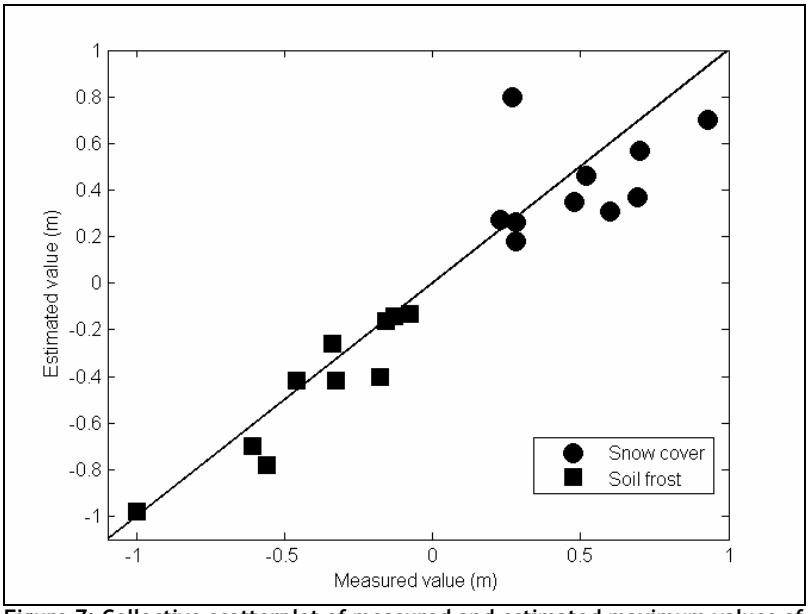


Figure 7: Collective scatterplot of measured and estimated maximum values of the depths of snow cover Sdepth (●) and soil frost Fdepth (▪); each point represents one independent set of data from the validation periods (including Karasjok).

The simulation of ice cover at Holt and Karasjok were based on the assumptions outlined in the sections *'Puddle formation and infiltration of meltwater'* and *'Formation of ice layer'*. The results shown in Figure 6 indicate that our simple approach is a sound starting point for further development of this module.

We conclude that our simple yet effective method for modelling depths of snow cover, lower frost boundary and soil surface ice provides reasonable results, making it suitable for incorporation into more complex models.

Continued work

In order to simulate crop damage due to e.g. ice encasement, there is a need for better description of local field topography, such as quantifying the area of the field that can potentially be covered by surface puddles. This will be of help when simulating the amount of plants dying due to ice-related stresses. These refinements should be followed by further model validation.

The results from the calibration and SA indicate scope for model improvement. A modification motivated by the calibration results is a functional description of ρ_{ns}. In addition, the results from the SA suggest lumping together (or disregarding) the processes related to liquid water within the snow cover and replacing the sinusoidal snow melt function by a constant melt rate.

Acknowledgements

Comments from Dr. Mats Höglind are much appreciated. This study was funded by the Norwegian Research Council as part of the strategic institute programme *Climate change effects on winter survival of perennial forage crops and winter wheat, and on plant diseases and weed growth and control at high latitudes* (WINSUR). The authors are grateful to the two anonymous reviewers for their insightful comments on an earlier version of this manuscript.

References

Baker, J. and Spaans, E.J.A., 1997. Mechanics of Meltwater Movement Above and Within Frozen Soil. *In* Iskandar, I., Wright, E., Radke, J., Sharratt, B., Groenevelt, P., and Hinzman, L., eds., *International Symposium on Physics, Chemistry, and Ecology of Seasonally Frozen Soils, Special Report,* volume 97, 31-36. U.S. Army Cold Regions Research and Engineering, CRREL, Hanover, NH.

Bélanger, G., Rochette, P., Gastonguay, Y., Bootsma, A., Mongrain, D., and Ryan, D., 2002. Climate change and winter survival of perennial forage crops in eastern canada. *Agron. J.* 94, 1120–1130.

Beldring, S., Engen-Skaugen, T., Forland, E., and Roald, L., 2008. Climate change impacts on hydrological processes in norway based on two methods for transferring regional climate model results to meteorological station sites. *Tellus* 60A, 439–450.

Benoit, G. and Mostaghimi, S., 1985. Modeling soil frost depth under three tillage systems. *Transactions of the ASAE* 28, 1499–1505.

Campolongo, F., Cariboni, J., and Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software* 22, 1509–1518.

Campolongo, F., Tarantola, S., and Saltelli, A., 1999. Tackling quantitatively large dimensionality problems. *Computer Physics Communications* 117, 75–85.

DeGaetano, A., Cameron, M., and Wilks, D., 2001. Physical simulation of maximum seasonal soil freezing depth in the united states using routine weather observations. *Journal of Applied Meteorology* 40, 546–555.

Engeset, R., Sorteberg, H., and Udnas, H., 2000. NOSIT - utvikling av NVEs operasjonelle snoinformasjonstjeneste. (NOSIT - development of NVE's operational snow information services). Dokument 1, NVE.

Flerchinger, G. and Saxton, K., 1989. Simultaneous heat and water model of a freezing snow-residue-soil system i. theory and development. *Transactions of the ASAE* 32, 565–571.

Gelman, A., 1996. *Markov Chain Monte Carlo in Practice*, chapter 8, 131–

144. Interdisiplinary Statistics Series. Chapman & Hall, Suffolk, GB, 1 edition.

Gelman, A. and Rubin, D., 1992. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7, 457–511.

Gray, J. and Morland, L., 1995. The compaction of polar snow packs. *Cold Regions Science and Technology* 23, 109–119.

Gudleifsson, B. and Larsen, A., 1993. *Advances in Plant Cold hardiness*, chapter Ice encasement as a component of winter kill in herbage plants. CRC Press, USA.

Hayashi, M., van Der Kamp, G., and Schmidt, R., 2003. Focused infiltration of snowmelt water in partially frozen soil under small depressions. *Journal of Hydrology* 270, 214-229.

Höglind, M., Schapendonk, A., and Van Oijen, M., 2001. Timothy growth in scandinavia: Combining quantitative information and simulation modelling. *New Phytologist* 151, 355–367.

Iwata, Y., Hayashi, M., and Hirota, T., 2008. Comparison of snowmelt infiltration under different soil-freezing conditions influenced by snow cover. *Vadose Zone Journal* 7, 79–86.

Jansson, P. and Karlberg, L., 2001. *Coupled heat and mass transfer model for soil-plant-atmosphere systems*. Royal Institute of Technology, Dept. of Civil and Environmental Engineering, Stockholm, Sweden. Internet: ftp://www.lwr.kth.se/CoupModel/CoupModel.pdf.

Jordan, R., 1991. A one-dimensional temperature model for a snow cover. technical documentation for sntherm.89. Special Report 91-16, U.S.Army Cold Regions Research and Engineering Laboratory.

Judson, A. and Doesken, N., 2000. Density of freshly fallen snow in the central rocky mountains. *Bull. Amer. Meteor. Soc.* 81, 1577–1587.

Kennedy, I. and Sharratt, B., 1998. Model comparisons to simulate soil frost depth. *Soil Science* 163, 636–645.

Kokkonen, T., Koivusalo, H., Jakeman, T., and Norton, J., 2006. Construction of a degree-day snow model in the light of the ten iterative steps in model development. In Voinov, A., Jakeman, A., and Rizzoli, A., eds., *Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modelling and Software Society*. Burlington, USA. Internet: http://www.iemss.org/iemss2006/sessions/all.html.

Larsen, A., 1994. Breeding winter hardy grasses. *Euphytica* 77, 231-237.

Melloh, R., 1999. A synopsis and comparison of selected snowmelt algorithms. Report 99-8, U.S. Army Cold Regions Research and Engineering Laboratory.

Roberts, G., 1996. *Markov Chain Monte Carlo in Practice*, chapter 3, 45–57. Interdisiplinary Statistics Series. Chapman & Hall, Suffolk, GB, 1 edition.

Rutter, N., Essery, R., Pomeroy, J., and et. al, 2009. Evaluation of forest snow process models (snowmip2). *Journal of Geophysical Research* 114, 1–18.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., 2008. *GLOBAL SENSITIVITY ANALYSIS. The Primer*. Wiley & Sons Ltd, West Sussex, England.

Sivia, D., 2006. *Data Analysis: A bayesian tutorial*. Oxford University Press, Oxford, England, 2 edition.

Stähli, M., Bayard, D., Wydler, H., and Fluhler, H., 2004. Snowmelt infiltration into alpine soils visualized by dye tracer technique. *Arctic, Antarctic and Alpine Research* 36, 128–135.

Thorsen, S. and Haugen, L., 2007. Development of the snowfrost model for the simulation of snow fall and soil frost. Bioforsk FOKUS 9. Internett: http://www.bioforsk.no/PublicationAttachment.aspx?publicationid =10048&attachmentid=1179.

Van Oijen, M., Hoglind, M., Hanslin, H., and Caldwell, N., 2005a. Process-based modelling of timothy regrowth. *Agronomy Journal* 97, 1295–1303.

Van Oijen, M., Rougier, J., and Smith, R., 2005b. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology* 25, 915–927.

Vehvilainen, B., 1992. *Snow cover models in operational watershed forecasting*. Ph.D. thesis, Helsinki University of Technology.

Nomenclature

F_{depth}	simulated depth of lower frost boundary (m)
I_{depth}	simulated thickness of surface ice cover (m)
K	degree-day temperature index for snow melt (mm °C-1 d-1)
K_{min}	minimum value of K (mm °C-1)
K_{max}	maximum value of K (mm °C-1)
L_f	latent heat of fusion (J kg-1)
M	snow melt rate (mm d-1)
M_{rf}	refreezing rate (mm d-1)
P	precipitation rate (mm d-1)
P_r	simulated daily precipitation rate as rain (mm d-1)
P_s	simulated daily precipitation rate as snow (mm d-1)
Q_E	heat flux density from freezing of soil water (J m-2 d-1)
Q_{fs}	heat flux density through frozen soil (J m-2 d-1)
Q_{snow}	heat flux density through snow cover (J m-2 d-1)
S_{dry}	water constituent of snow cover in solid state (snow and ice) (mm)
S_{wet}	liquid water constituent of snow cover (mm)
SWE	snow water equivalent (mm)
SW_{ret}	retention capacity of snow cover (mm mm-1)
SW_{rf}	degree-day temperature index for refreezing of liquid water within snow cover (mm °C-1 d-1)
S_{depth}	depth of simulated snow cover (m)
T_{air}	daily mean air temperature at 2 m height (°C)
T_{surf}	simulated temperature in void between soil surface and snow cover (°C)
T_{rs}	daily mean air temperature below which precipitation is simulated as snow (°C)
T_{mf}	daily mean air temperature below which water within snow cover refreezes (°C)
T^*	daily mean air temperature below which soil water freeze (°C)
x_w	volumetric content of available soil water (m3 m-3)
z	soil depth (m)
γ	empirical parameter (m-1)
λ_{fs}	thermal conductivity of frozen soil (J m-1 °C -1 d-1)
λ_i	thermal conductivity of surface ice cover (J m-1 °C -1 d-1)
λ_s	thermal conductivity of snow cover (J m-1 °C -1 d-1)
ρ_{ns}	density of falling new snow (kg m-3)
ρ_w	density of water at 0 \, °C (kg m-3)
ρ_s	density of snow cover (kg m-3)
ξ	snow cover compaction parameter (mm mm-1 d-1)

Paper III

Stig Morten Thorsen and Mats Höglind
Modelling cold hardening and dehardening in timothy. Sensitivity analysis and Bayesian model comparison.
Submitted to Agricultural and Forest Meteorology.

Modelling cold hardening and dehardening in timothy. Sensitivity analysis and Bayesian model comparison

Stig Morten Thorsen[*,a,b], Mats Höglind[a]

[a] *Norwegian Institute for Agricultural and Environmental Research, Grassland and Landscape Division. Postvegen 213, N-4353 Klepp st., Norway.*
[b] *Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.*

Abstract

Timothy (*Phleum pratense* L.) is the most important forage grass in Scandinavia and it is therefore highly interesting to study how it will perform in a changing climate. In order to model winter survival, the dynamics of hardening and dehardening must be simulated with satisfactory precision. We investigated an early timothy frost tolerance model (LT50 model), and an LT50 model for winter wheat. Based on the assumption that timothy has no vernalization requirement, unlike winter wheat, but does have the ability to adapt to cold temperatures in a process linked to stage of development, two alternative versions of the winter wheat model were also constructed. These four models were calibrated by a Bayesian approach using observations on LT50 for the timothy cultivar Engmo. The models were validated using independent observations at different locations reflecting differences in climate. A sensitivity analysis, using the Morris method, to identify important model parameters suggested that there is a connection between frost tolerance and stage of plant development, even if there is no vernalization requirement. The simplified winter wheat model was selected as the best candidate based on model selection criteria and its ability to capture the hardening and dehardening processes. The results from the Bayesian calibration suggest that there are no major regional differences in Norway calling for regional calibration. However, cultivar-specific calibration is probably required, since there are hardy and less hardy cultivars within the same species. A functional LT50 model would allow risk assessments to be made of future winter survival using specifically tailored and downscaled climate scenarios.

Key words: Modelling, frost tolerance, timothy, Bayesian calibration, sensitivity analysis

1. Introduction

Overwintering is a central issue in forage production in Scandinavia. Plants adapt to winter conditions by reallocating assimilates from growth to storage organs and by undergoing a number of physiological changes aimed at avoiding or mitigating cellular injuries caused by sub-zero temperatures. This acclimation to cold temperatures is often referred to as cold hardening. When

[*]Corresponding author. Tel.: +47 404 75 561; fax: +47 514 26 744.
Email address: stig.thorsen@bioforsk.no, (Stig Morten Thorsen)

springtime comes, or during mild spells in mid-winter, the plants deharden (de-acclimate), and if the mild period is sufficiently long or the temperature is sufficiently high, growth is resumed. Timothy (*Phleum pratense* L.) is the most important forage grass in Scandinavia, and it is therefore interesting to study how it will perform in a changing climate. In order to model winter survival, the dynamics of hardening and dehardening must be simulated with satisfactory precision. The hardening status of plants is often quantified by measuring the LT50, i.e. the temperature at which 50% of the plants are killed in controlled freezing experiments. In the present study we explored different ways of simulating LT50 in timothy. The starting point was the model for timothy presented by Van Oijen et al. (2008). Preliminary studies showed that this model did not simulate LT50 with satisfactory precision (ibid.), and therefore a model for winter wheat developed by Bergjord et al. (2007) was also included. As this latter model seems to capture the LT50 dynamics of winter wheat in Norway with relatively good precision, we wanted to test whether it could also be used for simulating LT50 in timothy. Timothy and winter wheat are both graminoids, and as relatives they have many common morphological and physiological traits. However, there are also important differences. Winter wheat needs to be vernalized in order to develop flowering stems, whereas timothy does not (Heide, 1994). Moreover, in winter wheat there is a link between vernalization and hardening, more specifically plants that are fully vernalized cannot harden at the same rate as pre-vernalization plants (Fowler et al., 1999; Fowler and Limin, 2004). The lack of vernalization requirement in timothy suggests that the wheat model can be simplified in order to simulate LT50 in timothy. However, although timothy has no vernalization requirement, the ability to harden may still be linked to plant development (Rapacz, 2002; Kalberer et al., 2006). Based on this assumption, we constructed two alternative versions of the winter wheat model, with the vernalization requirement replaced by simpler functions describing plant development, and with links between these functions and plant hardening ability.

The aim of the work was to present different versions of the model and to identify the best candidate for further work, based on comparison of simulations and observations of LT50 data from field experiments carried out at five different locations in Norway. Data from one year at three different locations were used for multi-site calibration, and independent data from four locations were used for site-specific validation. We used a Bayesian approach to obtain an estimate of the joint posterior distribution of the model parameters based on non-informative prior distributions for the parameters and an observational data set. We also conducted a sensitivity analysis identifying the least important parameters in order to aid in the model simplification process. Obtaining an LT50 model that performed satisfactorily would make it possible to perform risk analyses for future winter survival using specifically tailored and downscaled climate scenarios.

2. Methods

2.1. Frost tolerance (LT50) models

We compared four LT50 models, each simulating the course of frost tolerance in timothy. In the following we give a description of each model, using the original parameter names to facilitate comparisons with the original literature. All model parameters were assumed to be cultivar-dependent. A common feature of the LT50 models investigated in this study is that we used the SnowFrostIce model (Thorsen et al., 2009) to simulate winter climate effects (snow cover, soil frost and soil surface temperature) on the soil surface. This is where the main apex of the grass plant is located during winter. The simulated soil surface temperature T_{surf} (°C) was

used as driving temperature in all LT50 models. Driving variables for the SnowFrostIce model, in turn, are observed daily mean air temperature T_{air} (°C) and precipitation Prec (mm day-1) (see Table 1). The differential equations for LT50 were solved numerically using the Euler method with a fixed time step of one day. This also applied for the SnowFrostIce model.

2.1.1. Model 1: LINGRA timothy model

The first model candidate is based on the key assumptions: (1) a target LT50 value exists for the test cultivar that is a linear function of the ambient soil surface temperature T_{surf}; (2) one equation applies to both hardening and dehardening rates. Both the hardening and dehardening rates are proportional to the difference between target and actual LT50 value, using the same proportionality constant KLT50 (°C °C-1). The daily change in LT50 for Model 1 is calculated as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{LT50}) = \frac{1}{\mathrm{TCLT50}}\left([\mathrm{LT50MX} + \alpha(T)\,\mathrm{KLT50}] - \mathrm{LT50}\right) \tag{1}$$

$$\alpha(T) = \begin{cases} T_{surf} - T_{Hard} & \text{if } T_{surf} < T_{Hard} \\ 0 & \text{otherwise} \end{cases}$$

where TCLT50 is a scaling parameter (d), LT50MX (°C) is the maximum LT50 value (i.e. the LT50 value of unhardened plants), KLT50 (°C °C-1) is a proportionality constant, T_{Hard} (°C) is the threshold temperature for hardening and LT50 (°C) is the current LT50 value.

2.1.2. Model 2: FROSTOL wheat model

A detailed description of Model 2 is presented by Bergjord et al. (2007). This model simulates LT50 in winter wheat and is based on a Canadian model for LT50 in winter wheat proposed by (Fowler et al., 1999) and further developed by Bergjord *et al.* for use in an oceanic climate. As input data, Model 2 requires observed soil temperature at 2 cm depth and snow cover depth. Our implementation of Model 2 was linked to the SnowFrostIce model, which provides estimates for both the soil temperature at 2 cm depth and snow cover. Thus Model 2 only requires daily input on mean air temperature and precipitation rates. The frost tolerance in Model 2 is calculated as:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{LT50}) = \mathrm{RATED} + \mathrm{RATES} + \mathrm{RATER} - \mathrm{RATEH} \tag{2}$$

where frost tolerance is increased by hardening (RATEH), and decreased by dehardening (RATED) and two stress-related functions. The first of these stress functions (RATES) is related to the exposure of plants to very low temperatures, and the second function is related to anaerobic conditions (RATER) when the soil temperature is approximately 0 °C and the soil is simultaneously covered by snow. All rates have units (°C d-1). Hardening takes place when $T_{surf} < 10$ °C:

$$\mathrm{RATEH} = \mathrm{Hparam}(10 - T_{surf})(\mathrm{LT50} - \mathrm{LT50MN}) \tag{3}$$

where Hparam (d-1 °C-1) is a scaling parameter originally assumed to be independent of cultivar, LT50 is the current LT50 value, and LT50MN is the maximum frost tolerance of the cultivar. The loss of frost tolerance due to dehardening RATED is also calculated according to Fowler et al. (1999):

$$\mathrm{RATED} = \mathrm{Dparam}(LT_{50i} - \mathrm{LT50})(T_{surf} + \kappa)^{\xi} \tag{4}$$

where Dparam (d-1 °C-ξ) is a scaling parameter assumed to be independent of cultivar, κ (°C) is a temperature shift constant and LT_{50i} is the upper threshold for dehardening for the cultivar,

Table 1: Variables and parameters used in the LT50 models. Original parameter names have been kept to facilitate comparisons with source literature

Common variables and parameters:	LT50	simulated frost tolerance (°C)
	T_{air}	daily mean air temperature (°C)
	Prec	daily precipitation rate (mm d-1)
	T_{surf}	simulated soil surface temperature (°C)
	LT50MX	maximum LT50-value (°C)
	T_{Hard}	threshold temperature for hardening or dehardening (°C)
Model 1:	TCLT50	scaling parameter (d)
	KLT50	proportionality constant (°C °C-1)
Model 2: FROSTOL winter wheat model	RATER	respiration under snow stress (°C d-1)
	RATES	low temperature stress (°C d-1)
	VR	daily rate of vernalization (-)
	VD	vernalization days (d)
	T_{maxV}	maximum temperature for vernalization (°C)
	T_{minV}	minimum temperature for vernalization (°C)
	T_{optV}	optimum temperature for vernalization (°C)
	Sparam	low temperature stress coefficient (°C-1)
	Rparam	respiration stress coefficient (d-1)
	$S_{depthMN}$	min snow depth for respiration stress (m)
Common parameters for models 2, 2a and 2b:	RATED	daily rate of dehardening (°C d-1)
	RATEH	daily rate of hardening (°C d-1)
	LT50MN	minimum LT50-value (°C)
	Hparam	hardening coefficient (d-1 °C-1)
	Dparam	dehardening coefficient (d-1 °C-ξ)
	ξ	empirical exponent (-)
	κ	temperature shift constant (°C)
Model 2a	d_{ψ}	start day of reduction in rehardening ability (d)

calculated according to Fowler et al. (1999) as $LT_{50i} = -0.6 + 0.142$ LT50MN. The constant κ was originally equal to 4 in the FROSTOL model, and was therefore not calibrated for this model.

The calculation of both RATEH and RATED requires an estimate of the state of vernalization. At the time of vernalization completion, the threshold temperature for dehardening is lowered, and the plant's ability to harden and reharden is switched off. The daily rate of vernalization (VR) is calculated according to Wang and Engel (1998) as:

$$\mathrm{VR} = \frac{2(T_{surf} - T_{minV})^{\alpha}(T_{optV} - T_{minV})^{\alpha} - (T_{surf} - T_{minV})^{2\alpha}}{(T_{optV} - T_{minV})^{2\alpha}} \tag{5}$$

where $\alpha = \ln 2 / \ln\left[(T_{maxV} - T_{minV})/(T_{optV} - T_{minV})\right]$. Necessary inputs to Equation (5) are the critical temperatures T_{minV}, T_{optV} and T_{maxV} (all having units (°C)) and representing minimum, optimum and maximum temperatures for vernalization, respectively. No vernalization is assumed if $T_{surf}<T_{minV}$ or $T_{surf}>T_{maxV}$. For technical reasons, in the calibration and sensitivity analysis we introduced an alternative formulation of the temperature parameters T_{maxV} and T_{optV}:

$$\begin{aligned} T_{maxV} &= T_{minV} + \Delta T_{maxV} \\ T_{optV} &= T_{minV} + \Delta T_{optV} \end{aligned}$$

where we calibrated the parameters T_{minV}, ΔT_{maxV} and ΔT_{optV}. The accumulated VR gives a number of vernalization days VD. If the plants are exposed to a temperature equal to T_{optV} for one day, then one VD is acquired. Otherwise a fraction of VD is acquired. When T_{surf} remains within the range of the critical temperatures T_{minV} and T_{maxV}, the duration of the vegetative phase of winter wheat is shortened and, as the vernalization requirement is gradually fulfilled, the level of maximum frost tolerance is reached. Streck et al. (2003) presents a function relating the number of accumulated VD to the winter wheat plant's stage of development when the transition from vegetative to reproductive growth is initiated (i.e. the plant's state of primary induction):

$$f(\mathrm{VD}) = \frac{(\mathrm{VD})^5}{\left[(22.5)^5 + (\mathrm{VD})^5\right]} \tag{6}$$

where the value 22.5 represents the number of days until half saturation is reached. f(VD) has a sigmoid shape varying from 0 (unvernalized plants) to 1 (fully vernalized plants). Experiments have shown winter wheat plants to be fully vernalized after 49 VD (Bergjord et al., 2007, and references therein). According to Equation (6), f(VD)>0.98 when VD>50. Bergjord et al. (2007) assumed in Model 2 that when f(VD)≥ 0.99 the plants would be fully vernalized. When this vernalization requirement is met, Model 2 does not simulate additional hardening. In addition, the threshold temperature for dehardening is lowered, from 10 °C to -4 °C.

In accordance with Fowler et al. (1999), the low temperature stress function RATES (which contributes to decreased winter survival) calculates the effects of exposing the plants to near-lethal temperatures over a longer period of time as:

$$\mathrm{RATES} = \left(\frac{\mathrm{LT50} - T_{surf}}{\exp\left(-\mathrm{Sparam}(\mathrm{LT50} - T_{surf}) - 3.74\right)}\right)\frac{1}{\mathrm{d}t} \tag{7}$$

where Sparam (°C-1) is a scaling parameter assumed to be independent of cultivar. When winter wheat plants are exposed to near-lethal temperatures for some time, their winter survival is decreased.

The respirational stress RATER, assumed to be caused by anaerobic conditions under the snow cover, which also contributes to decreasing the frost tolerance, is calculated as:

$$\text{RATER} = \text{Rparam} \times \text{RE} \times f(\text{snow depth}) \tag{8}$$

where Rparam (d-1) is a scaling parameter assumed to be independent of cultivar, RE (°C) is a respiration variable which, according to Bergjord et al. (2007) and references therein, is calculated as $\text{RE} = \left[\exp(0.84 + 0.051\, T_{surf}) - 2\right]/1.85$. Finally, f(snow depth) is a linear function of snow depth which increases from 0 to 1 as the snow depth reaches a maximum level S_{depthMN} (m).

2.1.3. Model 2a: Modified wheat model

Using Model 2 as a reference model, we investigated two simpler versions. In the first of these, Model 2a, we kept the calculations of hardening RATEH and dehardening RATED rates and the parameter κ. We discarded the low temperature and respirational stress functions. From hardening and dehardening experiments performed at Særheim, it is clear that dehardening in timothy can occur from the onset of the hardening process until spring and therefore we did not include a switch-function for the hardening and dehardening processes as in Model 2. The ability of the timothy plants to reharden declines towards the springtime. We assumed that their ability to reharden was lost completely by 1 April. We thus scaled the hardening rate RATEH using a linear function which is switched on at the day of year called d_ψ (d) and declines to zero on a date which was always set to April 1 (day of year 91):

$$\frac{\mathrm{d}}{\mathrm{d}t}(\text{LT50}) = \text{RATED} - \text{g}(d) \times \text{RATEH} \tag{9}$$

where g(d) is the linear scaling function by which d_ψ is calibrated.

2.1.4. Model 2b: Modified wheat model

The other simplified model version, Model 2b, is a further simplification of Model 2a. In this model version the linear declination function g(d) was preset to start on 1 January, i.e. the parameter d_ψ was fixed at value 366 and, as for Model 2a, the function g(d) was set to reach zero on 1 April.

2.2. Datasets

The LT50 observations used to calibrate and validate the models were taken from frost tolerance experiments performed at different locations in Norway. Særheim (58.45°N, 5.39°E, 80 m.a.s.l.), Fureneset (61.18°N, 5.3°E, m.a.s.l.), and Kvithamar (63.29°N, 10.52°E, 20 m.a.s.l.) are all situated in or close to coastal regions, while Apelsvoll (60.42°N, 10.51°E, 250 m.a.s.l.) is inland and Holt (69.39°N, 18.54°E, 20 m.a.s.l.) is on the northern coast. Weather data were collected from automated stations located within 300 m of the experimental fields. We used daily observations on mean air temperature T_{air} (°C) and precipitation rate P (mm d-1) in our study. Table 2 shows seasonally-averaged (autumn: September- November; winter: December-February; spring: March-May) values for air and soil temperature and cumulative precipitation. We estimated the number of days with a snow cover depth of more than 10 cm for each location using the SnowFrostIce model (Thorsen et al., 2009). The estimated snow cover at Særheim during the calibration winter of 05-06 lasted 36 days, and for the validation winter of 06-07 zero days. At Fureneset, the estimated snow cover for the validation winter 05-06 lasted 12

days. At Apelsvoll, the estimated snow cover for the calibration winter 91-92 lasted 13 days. At Kvithamar, the estimated snow cover for the validation winter 06-07 lasted 121 days. At Holt, the estimated snow cover for the calibration winter of 05-06 lasted 94 days and for the validation winter of 06-07 103 days.

The LT50 data from Apelsvoll originate from Sunde (1996), while the data from the other locations originate from the WINSUR project presented by Höglind et al. (2008). The total LT50 dataset was divided into two parts, one for calibration and one for validation (see Table 3).

2.3. Parameterization and model uncertainty

The models presented here represent simplifications of complex, real world processes. They try to describe, at various levels of complexity, the temporal variation in frost tolerance in timothy. Our models simulate the combined dynamics of hardening and dehardening. Some of the parameters used have a physical interpretation, but they are seldom measured and quantitative data are scarce in the literature. Hence, the parameters and thereby also the model outputs are influenced by uncertainty. We used Bayesian methods to quantify the uncertainty in model parameters and model output, with the intention of reducing these uncertainties rather than maximizing model fit, which is the classical way to calibrate and develop models. Van Oijen et al. (2005) give a detailed description of how to use Bayesian methods to calibrate complex plant models.

Calibration in the Bayesian way calls on the Bayes theorem to decrease both parameter and model uncertainty based on prior knowledge of the parameters and measurements. The prior knowledge is reflected by the prior probability density function for the parameters (pdf). If we say that all model parameters to be calibrated are elements of the vector $\boldsymbol{\theta}$, then $\pi(\boldsymbol{\theta})$ is a joint pdf expressing our initial belief about plausible values for the parameter. Given a dataset $\mathbf{D}$ on variables that are also model outputs, we can update the joint pdf of the parameters $\pi(\boldsymbol{\theta})$ by applying the Bayes theorem: $\pi(\boldsymbol{\theta}|\mathbf{D}) = \pi(\boldsymbol{\theta}) f(\mathbf{D}|\boldsymbol{\theta}) / f(\mathbf{D})$, where $\pi(\boldsymbol{\theta}|\mathbf{D})$ is the distribution of the model parameters given the observations $\mathbf{D}$, often referred to as the posterior distribution for the parameters. The likelihood of the observations given a model output generated using the particular parameter set $\boldsymbol{\theta}$ is denoted by the likelihood function $f(\mathbf{D}|\boldsymbol{\theta})$, and $f(\mathbf{D})$ is a normalization constant. In order to calibrate a dynamic model, a large number of parameter sets must be sampled and the corresponding model runs must be carried out. To sample these parameter sets, we used the Markov Chain Monte Carlo (MCMC) algorithm known as the Metropolis-Hastings Random Walk. Our target posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{D})$ was the stationary distribution of the Markov chain produced by the Metropolis-Hastings Random Walk.

2.3.1. Metropolis-Hastings Random Walk

The general idea of the MCMC is to randomly walk through the parameter space, running the model at each point visited. The first step in the random walk is to generate a starting point (i.e. $t = 0$) of the Markov chain. This starting point, denoted $\boldsymbol{\theta}^0$, is randomly sampled from the prior distributions of the parameters. The starting point could also have been chosen as e.g. the mid-point of the prior distribution. We randomly sampled three different $\boldsymbol{\theta}^0$ because we used three different Markov chains in our analysis. As the MCMC algorithm proceeds, every new proposal parameter vector $\boldsymbol{\theta}'$ is chosen based on the current parameter vector $\boldsymbol{\theta}^t$:

$$\boldsymbol{\theta}' = \boldsymbol{\theta}^t + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is a stochastic step length vector. It is important to note that the probability of jumping from the point $\boldsymbol{\theta}^t$ to the point $\boldsymbol{\theta}^t + \boldsymbol{\delta}$ is equal to the probability of jumping from the point $\boldsymbol{\theta}^t + \boldsymbol{\delta}$

to the point $\boldsymbol{\theta}^t$. We will shortly return to the step length vector $\boldsymbol{\delta}$ when describing its importance in the Metropolis-Hastings random walk. After sampling a proposed parameter set $\boldsymbol{\theta}'$, the model was run using this parameter set to obtain $\pi(\boldsymbol{\theta}'|\mathbf{D})$. Subsequently the so-called Metropolis ratio was computed:

$$r = \frac{\pi(\boldsymbol{\theta}'|\mathbf{D})}{\pi(\boldsymbol{\theta}^t|\mathbf{D})} = \frac{\pi(\boldsymbol{\theta}')\, f(\mathbf{D}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^t)\, f(\mathbf{D}|\boldsymbol{\theta}^t)}$$

The next step when deciding whether or not to accept $\boldsymbol{\theta}'$ was to generate a uniform random number $u \sim U(0,1)$, accepting the proposed parameter vector $\boldsymbol{\theta}'$ as the new $\boldsymbol{\theta}^{t+1}$ if $u \leq r$, or otherwise keeping the current parameter vector, $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. When the chain consisting of all $\boldsymbol{\theta}^t$ converged (i.e. reached stationarity), it formed the Markov chain serving as our sample from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{D})$.

The posterior distribution is thus a combination of prior knowledge and new information obtained from the data using the likelihood function. Measurement errors are used in the determination of how likely a model-data mismatch might be. For example, if the data are informative and have a sharply peaked distribution, the resulting posterior distribution will be narrower and more peaked than the prior. This indicates that the parameter uncertainty is reduced.

2.3.2. Defining the prior probability distribution for the parameters

Since literature is scarce regarding LT50 models, we had to select non-informative distributions as prior distributions with generously set minimum and maximum values. Regarding scaling parameters, we defined the boundaries so wide that they spanned orders of magnitude. Therefore as priors for the scaling parameters, we used the Jeffreys prior according to Sivia (2006) ($\pi(\theta) \propto 1/\theta$). For the remaining parameters we used flat uniform distributions with generously set boundaries (see Tables 4-7).

In the calibration process we assumed that the parameters were independent a priori, implying that their joint prior distribution is equal to the product of their individual marginal pdfs.

2.3.3. Defining the data-likelihood function and multi-site calibration

We used measurements on LT50 in timothy cv. Engmo at three different locations in the calibration of the LT50 models, as described above. The likelihood associated with a particular parameter set $\boldsymbol{\theta}^*$ is derived by comparing model output (using $\boldsymbol{\theta}^*$) with data. Outliers in the observed data can produce poor results. According to Sivia (2006), this problem can be solved by formulating a constraint on the Gaussian likelihood likelihood function. We assumed the measurement errors to be independent, and we used the formulation for the constraint on the Gaussian likelihood function $f(\mathbf{D}|\boldsymbol{\theta})$ according to Sivia (2006):

$$f(\mathbf{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sigma_i^o \sqrt{2\pi}} \left[\frac{1 - \exp^{R_i^2/2}}{R_i^2} \right] \tag{10}$$

where σ_i^o represents the lower boundary of the expected mismatch between a datum D_i and model output M_i at time $t = i$ (i.e. the standard deviation), $R_i = (M_i - D_i)/\sigma_i^o$ is the residual for the datum D_i, and N is the number of data points. The standard deviations of the Apelsvoll data were not available. From the estimation of LT50 values in timothy, 95% of the observations fell within ± 1°C (Höglind et al., 2008). We thus decided to use $\sigma^o = 2$°C for all data when calibrating the models.

Table 2: Seasonal average values for air and soil temperature (°C) and cumulative precipitation (mm). Air temperature and precipitation values are from the 30-year normal period 1961-1990, while soil temperatures are based on all available observations from on-location weather stations. Soil temperature was measured at 10 cm depth in an adjacent short-cut grass field

	Fall (SON)			Winter (DJF)			Spring (MAM)		
Location	T_{air}	T_{soil}	Prec	T_{air}	T_{soil}	Prec	T_{air}	T_{soil}	Prec
Særheim	8.2	9.2	156.7	1.0	2.5	101.7	5.7	6.8	70.0
Fureneset	7.7	8.9	244.3	1.7	2.5	181.0	5.7	5.9	113.3
Apelsvoll	4.6	6.5	59.0	-6.9	-0.1	34.0	2.5	3.8	35.0
Kvithamar	5.5	7.0	96.3	-2.8	0.4	67.7	4.3	3.7	52.7
Holt	3.4	4.6	111.7	-3.2	0.2	93.3	1.2	1.5	58.3

Table 3: Locations used for calibration and validation of the LT50 models. Number of data points in brackets

	Year	
Location	Calibration	Validation
Særheim	05-06 (5)	06-07 (3)
Fureneset	-	05-06 (5)
Apelsvoll	91-92 (12)	-
Kvithamar	-	06-07 (3)
Holt	05-06 (5)	06-07 (2)

Table 4: Description of parameters calibrated for Model 1, the LINGRA timothy model: Prior intervals (the Jeffreys prior indicated by †, otherwise Uniform prior), maximum a posteriori estimates $\boldsymbol{\theta}^{\mathrm{MAP}}$ and mode values $\tilde{\boldsymbol{\theta}}$. Non-influential parameters (their value having little effect on model outputs) are indicated by ‡

Symbol	Unit	Prior interval	$\boldsymbol{\theta}^{\mathrm{MAP}}$	$\tilde{\boldsymbol{\theta}}$	References
$T_{\mathrm{Hard}}^{\ddagger}$	°C	$[5, 15]$	11.1	13.5	This study
LT50MX‡	°C	$[-7, -3]$	-6.9	-6.8	This study
KLT50	°C°C-1	$[0, 200]^{\dagger}$	1.49	1.34	This study
TCLT50‡	d	$[0, 15]^{\dagger}$	5.0	5.2	This study

Table 5: Description of parameters calibrated for Model 2, the FROSTOL wheat model: Prior intervals (Jeffreys prior indicated by †, otherwise Uniform prior), maximum a posteriori estimates $\boldsymbol{\theta}^{\mathrm{MAP}}$ and mode values $\tilde{\boldsymbol{\theta}}$. Non-influential parameters (their values having little effect on model outputs) are indicated by ‡

Symbol	Unit	Prior interval	$\boldsymbol{\theta}^{\mathrm{MAP}}$	$\tilde{\boldsymbol{\theta}}$	References
Dparam‡	d-1 °C-ξ	$[0, 0.0027]^{\dagger}$	1.05E-6	5.02E-5	Bergjord et al. (2007)
LT50MN	°C	$[-35, -20]$	-33.5	-26.9	This study
ξ	-	$[0, 10]$	5	3	Bergjord et al. (2007)
LT50MX‡	°C	$[-7, -3]$	-6.9	-6.9	This study
Rparam‡	d-1	$[0, 50]^{\dagger}$	2.98	0.081	Bergjord et al. (2007)
$\mathrm{S}^{\ddagger}_{depthMN}$	m	$[0, 0.3]$	0.06	0.20	Bergjord et al. (2007)
Sparam‡	°C-1	$[0, 200]^{\dagger}$	178.41	0.93	Bergjord et al. (2007)
Hparam	d-1 °C-1	$[0, 1]^{\dagger}$	0.0017	0.0097	Bergjord et al. (2007)
$\mathrm{T}^{\ddagger}_{minV}$	°C	$[-3.0, 1.0]$	0.97	0.88	This study
$\Delta T^{\ddagger}_{maxV}$	°C	$[13, 19]$	13.2	13.1	This study
$\Delta T^{\ddagger}_{optV}$	°C	$[3, 7]$	4.9	5.4	This study
$\mathrm{T}_{\mathrm{Hard}}$	°C	$[5.0, 15.0]$	14.4	14.8	This study

Table 6: Description of parameters calibrated for Model 2a, the modified wheat model: Prior intervals (Jeffreys prior indicated by †, otherwise Uniform prior), maximum a posteriori estimates $\boldsymbol{\theta}^{\mathrm{MAP}}$ and mode values $\tilde{\boldsymbol{\theta}}$. Non-influential parameters (their values having little effect on model outputs) are indicated by ‡

Symbol	Unit	Prior interval	$\boldsymbol{\theta}^{\mathrm{MAP}}$	$\tilde{\boldsymbol{\theta}}$	References
Dparam‡	d-1 °C-ξ	$[0, 0.0027]^{\dagger}$	2.5E-3	5.3E-6	Bergjord et al. (2007)
ξ	-	$[0, 10]$	1	1	Bergjord et al. (2007)
LT50MX‡	°C	$[-7, -3]$	-6.7	-6.8	This study
$\kappa^{\ddagger}$	d	$[0, 10]$	9.9	7.5	This study
$\mathrm{T}^{\ddagger}_{\mathrm{Hard}}$	°C	$[5, 15]$	11.2	9.3	This study
LT50MN	°C	$[-35, -20]$	-31.5	-23.3	This study
Hparam	d-1	$[0, 1]^{\dagger}$	8.0E-3	0.0172	Bergjord et al. (2007)
$d^{\ddagger}_{\psi}$	d	$[300, 451]$	352	381	This study

Table 7: Description of parameters calibrated for Model 2b, the modified wheat model: Prior intervals (Jeffreys prior indicated by †, otherwise Uniform prior), maximum a posteriori estimates $\boldsymbol{\theta}^{\mathrm{MAP}}$ and mode values $\tilde{\boldsymbol{\theta}}$. Non-influential parameters (their values having little effect on model outputs) are indicated by ‡

Symbol	Unit	Prior interval	$\boldsymbol{\theta}^{\mathrm{MAP}}$	$\tilde{\boldsymbol{\theta}}$	References
Dparam‡	d-1 °C-ξ	$[0, 0.0027]^{\dagger}$	6.40E-6	8.00E-7	Bergjord et al. (2007)
ξ	-	$[0, 10]$	4	4	Bergjord et al. (2007)
LT50MX‡	°C	$[-7, -3]$	-6.8	-6.9	This study
κ‡	d	$[0, 10]$	6.0	5.8	This study
$T^{\ddagger}_{\mathrm{Hard}}$	°C	$[5, 15]$	13.5	10.5	This study
LT50MN‡	°C	$[-35, -20]$	-25.2	-23.3	This study
Hparam	d-1	$[0, 1]^{\dagger}$	0.0187	0.0299	Bergjord et al. (2007)

2.3.4. Determining jumps in the Metropolis-Hastings Random Walk algorithm

The step length vector $\boldsymbol{\delta}$ in the Metropolis-Hastings Random Walk algorithm is very important in order to obtain effective convergence of the Markov chain produced, i.e. a sample from the targeted posterior distribution of the parameters. In our implementation a new proposal parameter set was generated by sampling from a multivariate normal distribution with variance proportional to the prior ranges of the parameters. We successfully used a variance equal to the square of 5% of the prior range, and zero covariances. If the elements in the step length vector $\boldsymbol{\delta}$ are too small, the random walk algorithm will on average not move far enough from the current point in parameter space $\boldsymbol{\theta}^{t}$ when proposing a new candidate parameter vector $\boldsymbol{\theta}'$, and consequently the acceptance rate will be too large, or *vice versa*. The key issue when determining the step length is selecting a value to ensure an adequate exploration of the parameter space, and ideally obtaining an acceptance rate between 0.15 and 0.5 (Roberts, 1996).

2.3.5. Determining convergence of the Markov chains

A central issue when using an iterative simulation method such as the Metropolis-Hastings Random Walk algorithm is to determine when the chain has converged to the desired posterior distribution. One option, suggested by Gelman and Rubin (1992), is to generate multiple chains followed by calculation of the scale reduction factor GR for each parameter, which is used to determine the length of the 'burn-in' phase. The burn-in of the chain is the first part, where the chain is still influenced by the starting point in the walk through parameter space, prior to reaching stationarity. Following Gelman (1996) we determined the burn-in phase to last until GR remained below 1.2 (when GR is near 1 it means that the Markov chains are essentially overlapping). This meant that our independent chains had located the same region in parameter space. We randomly sampled three starting points from the prior distribution, and used the GR to determine when the three chains had converged to the desired posterior distribution.

2.4. Sensitivity analysis

In order to identify non-influential parameters that can be omitted from the calibration, we carried out a screening exercise using the improved sensitivity indices from the Morris method as described by Campolongo et al. (2007). This method is relatively simple to implement, yet it

provides a good estimate for the main effects of the parameters (Campolongo et al., 2007). The results from this screening are also helpful in relation to model simplification.

The Morris method proposes two sensitivity measures, the main purpose of which is to determine those of the model's k parameters that can be considered to be (i) not important, (ii) linear and additive, or (iii) non-linear or involved in interactions with other parameters. For each of the parameters, two sensitivity measures are computed; μ, which evaluates the overall influence of the parameter on the model output (main effect, or elementary effect), and σ, which collectively evaluates all the higher order effects due to non-linearity and/or interactions with other parameters. The Morris method was originally used for parameters following uniform distributions in $[0, 1]$. Campolongo et al. (1999) suggest how to treat parameters with different distributions. Rather than sampling directly from the prior distribution for the parameters, the sampling should be performed in the space of equidistant quantiles of these k distributions (i.e. each prior distribution is discretized into a predetermined number of p levels, where each quantile q_p varies in $[0, 1]$, producing as sampling space a k-dimensional unit hypercube). We performed the sensitivity analysis (SA) for each of the candidate models, sampling from the quantiles of the prior distributions listed in Tables 4 - 7. We call this sampling space Ω.

Outputs from the LT50 models are time series, and for this SA we needed a scalar value as response variable. Thus instead of performing the SA at one single point in time, we used as scalar response the log-transformed likelihood from Equation (10), i.e. $\log(f(\mathbf{D}|\boldsymbol{\theta}))$; the likelihood being the probability of the observed data $\mathbf{D}$ given a certain model parameterization $\boldsymbol{\theta}$.

By sampling random parameter sets $\boldsymbol{\theta}$ from Ω according to the Morris method, we obtained the distribution associated with the elementary effect of each parameter. This distribution is finite and can be denoted $EE_i(\boldsymbol{\theta}) \sim D_i$. As suggested by Campolongo et al. (2007), a convenient choice for p is an even number, and a convenient choice for the sampling increment Δ is setting $\Delta = p/\left[2(p-1)\right]$. The number of elements in D_i is $p^{k-1}\left[p - \Delta(p-1)\right]$ (ibid.). The sensitivity indices for parameter θ_i from the Morris method are the mean μ and standard deviation σ of D_i. The design proposed by Morris (1991) is to sample r elementary effects from each D_i by constructing r trajectories (each trajectory having $k + 1$ points in Ω), where each trajectory provides k elementary effects (one elementary effect for each parameter). This leads to a total of $r(k + 1)$ sampling points, giving us the number of model runs required for the SA. A detailed description of the implementation of the Morris method is presented in Saltelli et al. (2008).

The sensitivity indices are interpreted as follows. A high σ_i value for parameter θ_i implies that the corresponding elementary effect EE_i for this parameter θ_i at one point in Ω is considerably different from the elementary effect for the same parameter calculated somewhere else in Ω, i.e. that this particular EE value is also influenced by the values of the other parameters or nonlinearities/interactions. A low value for σ_i suggests that the EE_i value associated with θ_i is less influenced by the values of the other parameters and thus that it is not involved in interactions or nonlinearities.

If the distribution D_i contains both positive and negative elements (i.e. the model is non-monotonic), we might get the situation where effects are cancelling each other out and thereby producing a low μ_i value, even for an important parameter. Therefore the value of μ_i should be considered together with σ_i. This kind of cancelling effect yielding a low μ_i would be accompanied by a high σ_i value. If the model is non-monotonic (as the LT50 models), Campolongo et al. (2007) suggest replacing μ by μ^*; an estimate of the mean distribution of absolute values of the elementary effects G_i, i.e. $|EE_i| \sim G_i(\mu_i^*, \sigma_i)$. This solves the problem of opposite signs on the elementary effects. However, to properly characterize non-influential parameters, one must therefore simultaneously consider μ^* and σ (the standard deviation of the distribution

D_i). Important parameters are those having large values of μ^* and σ compared with the results for the other parameters. This means that the non-influential parameters will be identified by having values of μ^* and σ lying in the lower left-hand corner of the scatter plot μ^* vs. σ. A non-influential parameter has the property that when it is varied freely within its boundaries (as defined for the SA), it will not markedly affect the model output. Therefore by identifying the non-influential parameters, these may be left out of the calibration and prepare the ground for model simplification.

2.5. Parameter estimation and model selection

We used the same MCMC settings with regard to number of iterations and sampling steps in the Metropolis-Hastings random walk for all four models. From the prior distributions we randomly sampled three sets of initial values, and ran three chains for 50,000 iterations each. We assessed the convergence of the chains for each model by investigating time-series plots of GR. Convergence was obtained when GR remained below 1.2. To make sure we obtained a sample from the stationary posterior distribution, we used the final 20,000 iterations from each of the three chains as our posterior. From this final posterior sample, consisting of $n = 60,000$ elements, we identified two parameter vectors: the parameter vector with the highest posterior probability $\boldsymbol{\theta}^{\mathrm{MAP}}$, and the parameter vector consisting of the mode values of the marginal posterior distributions $\widetilde{\boldsymbol{\theta}}$.

To quantify the uncertainty in LT50 simulations from each model, we sampled 9,000 parameter sets from the posterior $\pi(\boldsymbol{\theta}|\mathbf{D})$ and ran the model using weather data from the validation data set. The resulting LT50 time series are the members of what we call the ensemble LT50 for the corresponding model. The average of all ensemble members is referred to as the ensemble mean LT50, termed $\overline{M}$. As suggested by Iizumi et al. (2009) we quantified the model uncertainty by calculating root-mean-square deviation (RMSD_L) for each model at each location using $\overline{M}$. We calculated RMSD_L as follows:

$$\mathrm{RMSD}_t = \frac{1}{E}\sqrt{\sum_{e=1}^{E}(M_{e,t} - \overline{M}_{\bullet,t})^2} \tag{11}$$

$$\mathrm{RMSD}_L = \frac{1}{J}\sum_{t=1}^{J}\mathrm{RMSD}_t \tag{12}$$

where $\overline{M}_{\bullet,t}$ is the ensemble mean at simulation time t, $M_{e,t}$ is the model output of ensemble member e at simulation time t, and E and J are the total numbers of ensemble members and simulation time steps, respectively. The resulting RMSD reported in Table 8 is the weighted mean of the RMSD_L calculated for each location.

To evaluate the model accuracy we calculated RMSE using the independent observations in the validation data set. The RMSEs are presented in Table 8, and calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - D_i)^2} \tag{13}$$

where M_i is model output, D_i is observed LT50 value at time t_i, and N is the total number of observations used. For each model, we calculated two sets of RMSE based on two different sets of model output. The first set, called $\mathrm{RMSE}_{\mathrm{MAP}}$, was calculated based on simulation results

Table 8: Model selection criteria (AIC and BIC) calculated from the MCMC results and Goodness of Fit (root-mean-squared errors RMSE) and model uncertainty (root-mean-squared deviation RMSD) calculated using validation data

	Selection criteria		RMSE		RMSD
LT50 model	AIC	BIC	M^{MAP}	$\overline{M}$	
Model 1	141.25	145.62	5.47	5.19	2.54
Model 2	135.47	148.56	8.83	6.98	12.85
Model 2a	**131.26**	**139.98**	**4.00**	**4.19**	8.44
Model 2b	132.43	140.10	4.93	4.57	6.33

obtained using the parameter set that produced the maximum posterior probability, denoted $\boldsymbol{\theta}^{\mathrm{MAP}}$. M_t^{MAP} is model output at time t using $\boldsymbol{\theta}^{\mathrm{MAP}}$ parameterization. The second set of RMSE, denoted $\mathrm{RMSE}_{\overline{M}}$, was calculated using the ensemble mean LT50, where $\overline{M}_t$ is the ensemble mean LT50 at time t.

To aid in the model selection process, we calculated two commonly used model selection criteria: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC):

$$AIC = -2\log(ML) + 2p \tag{14}$$

$$BIC = -2\log(ML) + p\log(N) \tag{15}$$

where ML is the estimated maximum likelihood from the MCMC, p is the number of model parameters and N is number of data points used in model calibration (Kass and Raftery, 1995). As pointed out in Hastie et al. (2001), there is no clear choice between AIC and BIC for model selection purposes. We calculated both to compare the two different ways of weighting model complexity.

As suggested by Van Oijen et al. (2005), we also performed a Bayesian model comparison by calculating Bayes factors. The Bayes factor represents the evidence provided by the data in favour of one model as opposed to another (Kass and Raftery, 1995). The BIC criterion provides a rough approximation to the logarithmic Bayes factor (ibid.). The Bayes factor is the ratio of integrated likelihoods, here referred to as $f(\mathbf{D}|\mathrm{Model}_i)$, which represent the probability of the observations $\mathbf{D}$ given Model_i. According to Kass and Raftery (1995), the integrated likelihood $f(\mathbf{D}|\mathrm{Model}_i)$ can be estimated from the MCMC results by taking the harmonic mean of the likelihood:

$$f(\mathbf{D}|\mathrm{Model}_k) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{f(\mathbf{D}|\boldsymbol{\theta}^i)} \right\}^{-1} \tag{16}$$

where $n = 60,000$ is the size of the sample from the posterior (comprised of the last 20,000 iterations of each of the three Markov chains for each model).

3. Results

The Markov chains for all models converged relatively fast (GR<1.2) except for Model 2, which reached convergence after 30,000 iterations. Calibration results for the parameters of each model are shown in Table 4-7.

The calibration data set contained more information from the hardening period than the validation data. This is likely to have affected the model performance when comparing simulation results to calibration data and validation data.

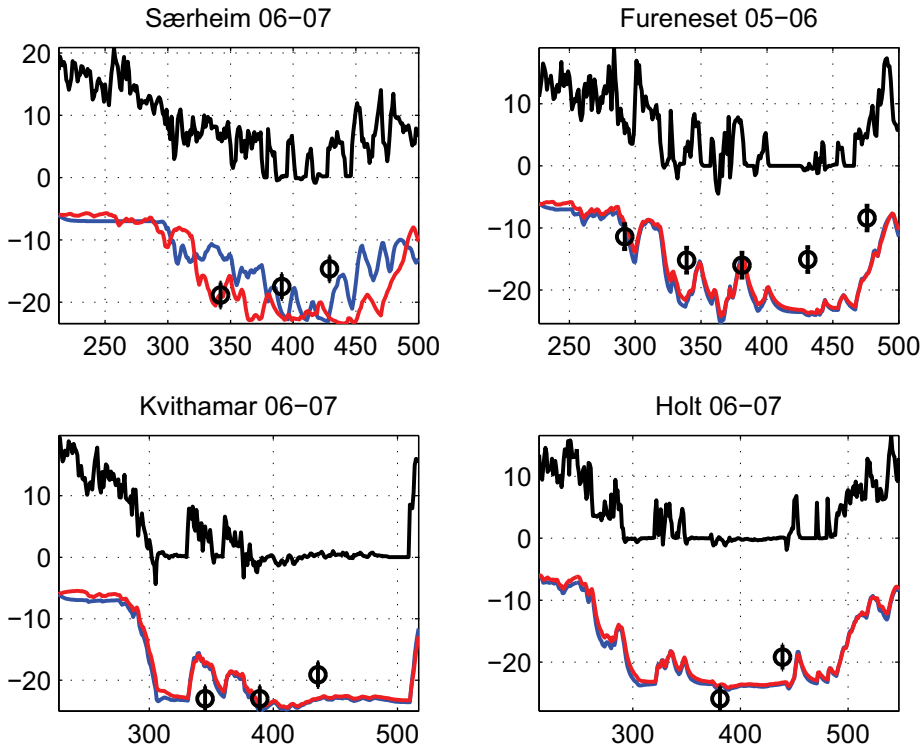


Figure 1: LT50 simulated with Model 1 using the best parameter set M^{MAP} (blue) and ensemble mean $\overline{M}$ (red). Observations with standard deviations indicated with circles. Simulated daily soil surface temperature T_{surf} (black).

3.1. Model evaluation

Figure 1 shows LT50 time series produced by Model 1, the LINGRA timothy model, using the best parameter set M^{MAP} (blue) and ensemble mean $\overline{M}$ (red). Model 1 captured the hardening process reasonably well, but stayed at lower LT50 for a longer time, leading to underestimation at nearly all subsequent data points. This means the plants remained hardened too long. This was also reflected by the high RMSE values for Model 1. Hardening is also simulated towards the end of the dehardening period. The RMSE for Model 1 was 5.47 and 5.19, when using M^{MAP} and $\overline{M}$, respectively. Model 1 had the highest RMSE of all candidate models but produced the lowest RMSD (2.54) at all locations (Table 8). The non-influential parameters in Model 1, i.e. the parameters that least affected the model output when varied within their boundaries, were TCLT50, LT50MX and T_{Hard}.

Figure 2 shows LT50 simulated with Model 2. The hardening process was captured relatively well for all locations except Holt. Based on M^{MAP}, the vernalization requirement from Equation (6) was met on December 24 ± one day for Særheim, Fureneset and Kvithamar, respectively. This led to a lowering of the threshold temperature for dehardening from 10°C to -4°C in the model,

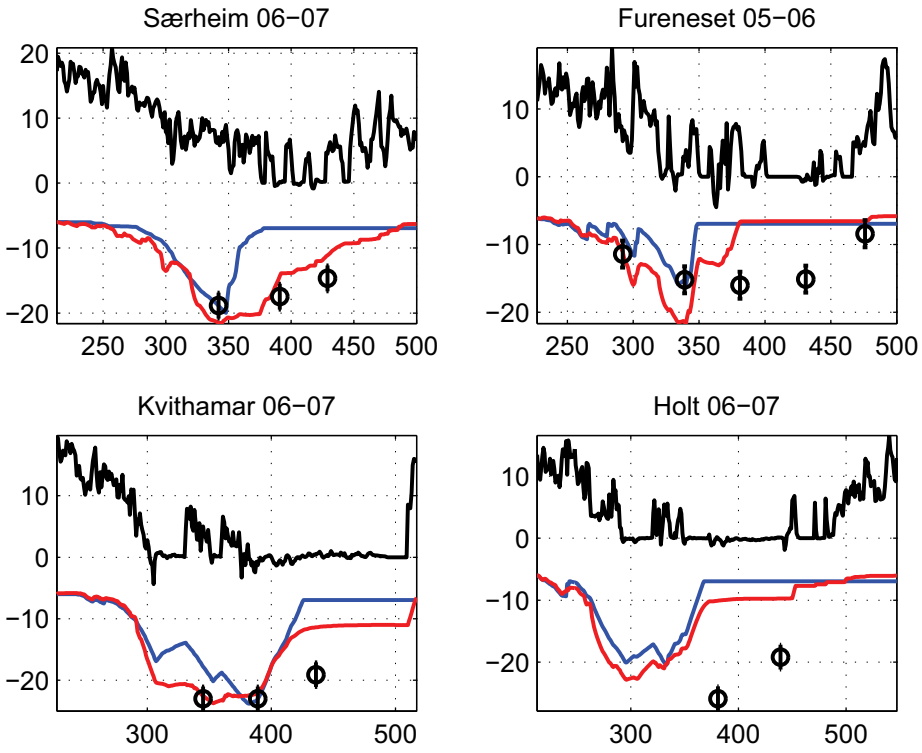


Figure 2: LT50 simulated with Model 2 using the best parameter set M^{MAP} (blue) and ensemble mean $\overline{M}$ (red). Observations with standard deviations indicated with circles. Simulated daily soil surface temperature $\mathrm{T_{surf}}$ (black).

with the consequence that dehardening was simulated too early, and LT50 was overestimated compared with the observations. At the Holt location, the vernalization requirement was met even earlier, on November 12. The hardening process was therefore terminated too soon, and due to the insulating effect of the snow cover ($\mathrm{T_{surf}}$ at around 0 °C) the dehardening was completed too early. The RMSE for Model 2 was 8.83 and 6.98 when using M^{MAP} and $\overline{M}$, respectively. The large difference in RMSE for Model 2 is due to the difference between the two LT50 time series, M^{MAP} and $\overline{M}$ (Figure 2). The RMSD for Model 2 was the highest amongst the candidate models, with value 12.85. The non-influential parameters in Model 2 were ΔT_{optV}, $\mathrm{S_{depthMN}}$, ΔT_{maxV}, T_{minV}, LT50MX, Rparam, Dparam, and Sparam (data not shown).

Figure 3 shows LT50 simulated with Model 2a. LT50 was underestimated at Særheim and Fureneset, but had a very good fit at Kvithamar and Holt. Gay and Eagles (1991) report that hardening rates are generally lower than dehardening rates. This trait appears to be captured best by Model 2a (see Figure 3). The RMSE for Model 2a was 4.00 and 4.19 when using M^{MAP} and $\overline{M}$, respectively. Model 2a had the overall lowest RMSE, and thus had the best fit at all locations. The RMSD for Model 2a was 8.44. The non-influential parameters in Model 2a were d_ψ, κ,

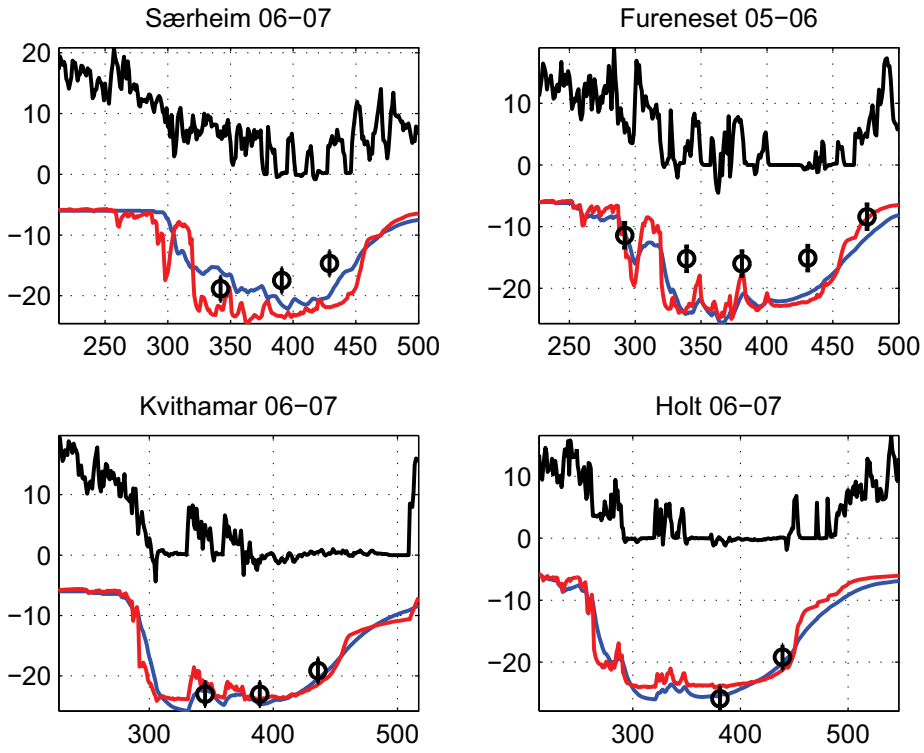


Figure 3: LT50 simulated with Model 2a using the best parameter set M^{MAP} (blue) and ensemble mean $\overline{M}$ (red). Observations with standard deviations indicated with circles. Simulated daily soil surface temperature T_{surf} (black).

LT50MX, T_{Hard} and Dparam (Figure 4).

Figure 5 shows LT50 simulated with Model 2b. The hardening process was captured well, but the loss of hardiness towards the beginning of spring was underestimated (i.e. the ability to reharden was overestimated), which resulted in the LT50 being underestimated for a longer period compared with Model 2a. The RMSE for Model 2b was 4.93 and 4.57 when using M^{MAP} and $\overline{M}$, respectively. The RMSD for Model 2b was 6.33. The least important parameters in Model 2b were Dparam, κ, T_{Hard} and LT50MX (not shown).

At the Kvithamar location, there were two distinctive mild episodes during the winter (T_{surf}>0°C), while at Holt there was one such episode. These mild episodes were captured well by all models except Model 2 in the sense that the simulated dehardening was followed by rehardening.

The selection criterion AIC ranked model 2a as the best model, slightly better than model 2b. Model 2 and model 1 were ranked third and fourth, respectively. According to the BIC criterion, models 2a and 2b were ranked first and second, respectively, while models 2 and 1 were ranked fourth and third, respectively. This indicates that the likelihood for models 2a and 2b was higher than that for models 1 and 2, and thereby they were ranked higher. One reason for the difference

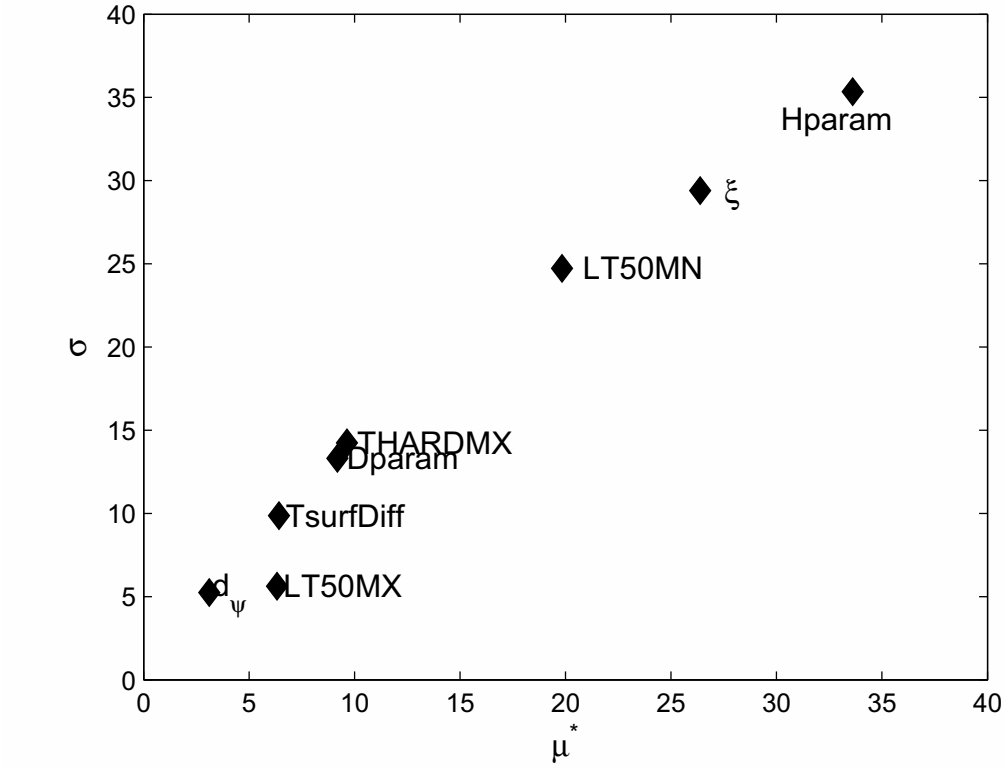


Figure 4: Sensitivity indices μ^* and σ for Model 2a. A high value for μ_i^* indicates a large elementary main effect for parameter θ_i, while a high value for σ_i indicates interaction or non-linear effect associated with parameter θ_i.

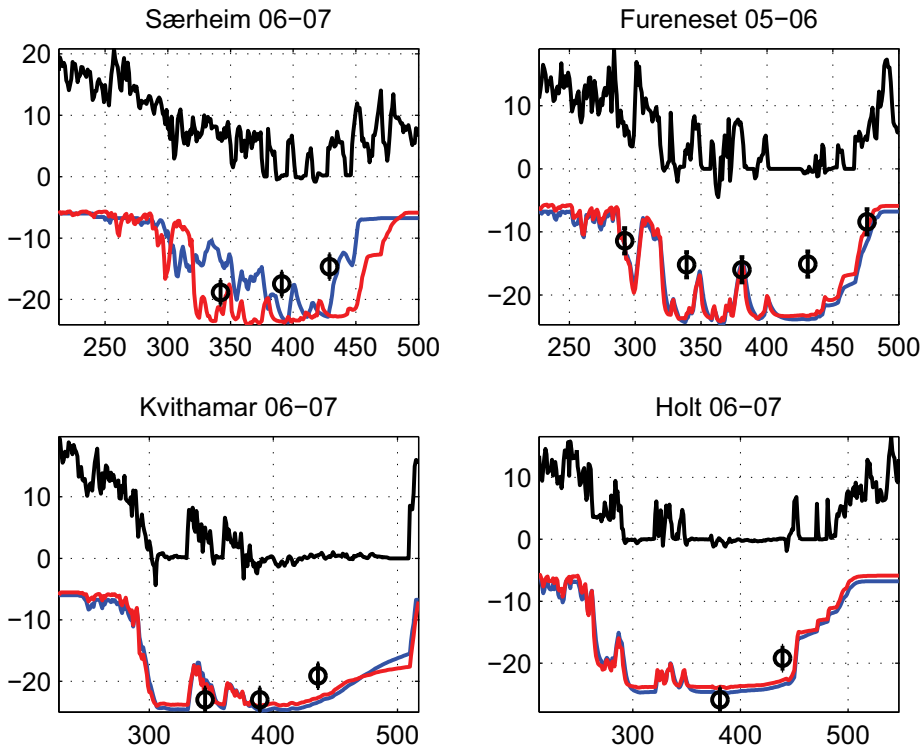


Figure 5: LT50 simulated with Model 2b using the best parameter set M^{MAP} (blue) and ensemble mean $\overline{M}$ (red). Observations with standard deviations indicated with circles. Simulated daily soil surface temperature T_{surf} (black).

Table 9: Bayesian model comparison using integrated likelihoods $f(\mathbf{D}|\text{Model}_k)$ from MCMC. For example, Model 2a is compared with Model 1 using the ratios of their integrated likelihoods, i.e. $f(\mathbf{D}|\text{Model}_{2a})/f(\mathbf{D}|\text{Model}_1)$

LT50 model	$f(\mathbf{D}\|\text{Model}_k)$	Model 1 vs.	Model 2 vs.	Model 2a vs.	Model 2b vs.
Model 1	4.29E-31		43.06	189.65	289.45
Model 2	1.85E-29	0.02		4.40	6.72
Model 2a	8.14E-29	0.01	0.23		1.53
Model 2b	1.24E-28	0.00	0.15	0.66	

between AIC and BIC for models 1 and 2 might be that BIC tends to favour simpler models more than AIC (Kass and Raftery, 1995) due to the heavier penalty on complexity (Hastie et al., 2001). The results for AIC and BIC are presented in Table 8.

We calculated the Bayes factors according to Kass and Raftery (1995) using equation (16). Model 1 and Model 2 were ranked last according to Bayes factor, suggesting that these two models were least supported by the data. For the two simplified wheat models, the simplest version (Model 2b) received most support from the data, thereby obtaining the highest value of the Bayes factor. The results for the Bayes factors are presented in Table 9.

4. Discussion

The earlier work of Gay and Eagles (1991) presents a method for quantitatively investigating the processes of hardening and dehardening as dependent on time and temperature. Their approach considers these two processes separately. In the present study we propose and describe four candidate models that simulate the combined effect of hardening and dehardening in timothy (*Phleum pratense* L.) cv. Engmo based on standard meteorological variables. Each candidate model was calibrated by means of Bayesian methods, and validated using independent observations.

Calibration of models by means of Bayesian methods begins with stating the initial degree of belief about parameter values. This is what is done when defining the prior distributions. If initial knowledge about the parameters is limited, the most common approach is to use so-called non-informative priors and to assume that the parameters are uncorrelated. In this study we used either uniform distributions with generously set maximum and minimum values, or the Jeffreys prior for scaling parameters (Sivia, 2006). The price we paid for choosing non-informative priors was that the resulting posterior distribution was less accurate, i.e. the resulting variance in model output was larger. Other aspects to consider when evaluating the results from the calibration are the data quality and the number of observations available. Only one location had data covering the whole hardening and dehardening period (location Apelsvoll, n=12). The other two locations used in the calibration had 5 data each for the whole period. Together with non-informative priors, this resulted in less accurate predictions, and more variability in model output.

In the sensitivity analysis we sampled parameter values from six equidistant quantiles of the posterior. This rough partitioning might cause important parameter values to be left unchecked, causing the Morris method to miss peaks of the likelihood. On the other hand, by following the sampling scheme of the Morris method, the main regions of the parameter space were explored, yielding a qualitative assessment of parameter sensitivities.

Judging by the model selection criteria AIC and BIC and the RMSEs, Model 2a (modified wheat model) was the best candidate (Table 8). This is also clear when comparing model output from Model 2a (Figure 3) with outputs from the other candidate models. The Bayes factors (Table 9) ranked Model 2b as having the highest probability in view of the data and prior information. On the other hand, the model selection criteria AIC and BIC both favoured Model 2a over the other three candidate models (Table 8).

The RMSE presented here describe the goodness of fit for the independent validation data, and thereby the capabilities of the individual models in capturing variation in the observations. For models 1, 2 and 2b, the RMSE obtained when using $\overline{M}$ was smaller than that obtained when using M^{MAP}. However, this result is of little use in prediction because it requires a large number of ensemble members from which to estimate an ensemble mean. One reason why $\mathrm{RMSE}_{\overline{M}} > \mathrm{RMSE}_{\mathrm{MAP}}$ for only one model might be insufficient reduction in parameter uncertainty in the calibration, which might have caused the posterior to have a broad, flat peak rather than being narrow and sharp, in turn causing many parameter sets to have similar likelihood values. Thus by sampling extensively from the posterior distribution, the ensemble mean simulation $\overline{M}$ might result in a lower RMSE than the use of only one parameter set, namely $\boldsymbol{\theta}^{\mathrm{MAP}}$. The RMSE for Model 2a calculated from M^{MAP} was slightly lower than the RMSE obtained using the ensemble mean. This indicates that the use of only one parameter set, the maximum posterior parameter set $\boldsymbol{\theta}^{\mathrm{MAP}}$, for this particular model gave good results compared with the results obtained using the ensemble mean $\overline{M}$. Considering the performance of the candidate models using the validation data and the selection criteria AIC and BIC, we proceeded with models 2a and 2b. These models are modest with respect to number of parameters and both the hardening and dehardening processes are well simulated. Results from the sensitivity analysis of Model 2 suggest that the parameters related to the particular stress functions of low temperatures and anaerobic conditions (RATES and RATER) as used by Bergjord et al. (2007), and also parameters related to the vernalization requirement, as used by Fowler et al. (1999), are not needed when modelling frost tolerance in timothy, at least under our climate conditions. The removal of these stress functions and the vernalization requirement constituted the main differences between Model 2 and the two simplified versions 2a and 2b. The findings from the sensitivity analysis thus support our assumption that the stress and vernalization functions are not necessary in our case. Our suggestion of gradually reducing plant ability to reharden by using a linear reduction of the hardening rate (in contrast to not reducing this rehardening ability at all as in Model 1) was supported, indicating that there is still a connection between frost tolerance and plant developmental stage (Rapacz, 2002; Kalberer et al., 2006), even if this is not linked to a vernalization requirement. A similar approach of modifying hardening rates linearly has been suggested by Kanneganti et al. (1998), who also distinguished between dehardening due to high temperatures in the hardening period and dehardening towards the initiation of the growth period. Models 2a and 2b underestimated the frost tolerance in late winter/early spring at Særheim and Fureneset. These are the two locations with the mildest winters (see Table 2). It may be that the lack of snow cover at these locations, leading to more fluctuating soil surface temperatures, puts the plants under more severe stress, resulting in additional dehardening not accounted for in the models. There may also be a connection between plant carbohydrate content and LT50 that is not accounted for in the models. In perennial ryegrass, both rate and maximum level of hardening have been shown to increase with the level of carbohydrates in the stubble up to a threshold value (Hanslin and Höglind, 2009).

5. Conclusions

From the model selection criteria in Tables 8 and 9 and the capability of the model to capture the hardening and dehardening processes, we concluded that Model 2a is the best candidate for further use when simulating the winter survival of timothy. Although Model 2b was ranked above Model 2a according to Bayes factors, Model 2a was better developed with regard to plant development. The results from the Bayesian calibration suggest that there are no major regional differences in Norway calling for regional calibration, since in the validation plots the same parameterization $\boldsymbol{\theta}^{\mathrm{MAP}}$ (the posterior distribution) was used at all locations. Therefore the parameter distribution obtained from the multi-site calibration can be used for a specific grass species. However, the model most likely needs to be calibrated for each grass cultivar separately, since there are both hardy and less hardy cultivars within the same species, but this assumption has to be verified in further modelling studies.

In coming works, we will examine whether the models can be improved further by making the rate of hardening dependent on the availability of carbohydrates within the plant, i.e. restricting the rate of hardening at carbohydrate contents below a threshold value. We will also test more refined ways of relating the rate of hardening to plant development. More observations capturing a larger part of both the hardening and dehardening periods would be helpful in increasing the precision and reducing the uncertainty in the candidate models.

6. Acknowledgements

The authors are grateful to Marcel Van Oijen and Anne Kari Bergjord for their detailed comments on the manuscript. This work was supported by the Norwegian Research council.

References

Bergjord, A., Bonesmo, H., Skjelvåg, A., 2007. Modelling the course of frost tolerance in winter wheat i. model development. European Journal of Agronomy 28, 321 – 330.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling and Software 22, 1509–1518.

Campolongo, F., Tarantola, S., Saltelli, A., 1999. Tackling quantitatively large dimensionality problems. Computer Physics Communications 117, 75–85.

Fowler, D., Limin, A., 2004. Interaction among factors regulating phenlogcal development and acclimation rate determine low temperature tilerance in wheat. Annals of Botany 94, 717–724.

Fowler, D., Limin, A., Ritchie, J., 1999. Low-temperature tolerance in cereals: model and genetic interpretation. Crop Science 39, 626 – 633.

Gay, A., Eagles, C., 1991. Quantitative analysis of cold hardening and dehardening in *Lolium*. Annals of Botany 67, 339–345.

Gelman, A., 1996. Markov Chain Monte Carlo in Practice, 1st Edition. Interdisiplinary Statistics Series. Chapman & Hall, Suffolk, GB, Ch. 8, pp. 131–144.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7 (4), 457–511.

Hanslin, H., Höglind, M., 2009. Differences in winter-hardening between phenotypes of lolium perenne with contrasting water-soluble carbohydrate concentrations. Grass and Forage Science 64, 187–195.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning Data, Mining, Inference, and Prediction. Springer, Ch. 7.

Heide, O., 1994. Control of flowering and reproduction in temperate grasses. New Phytologist 128, 347–362.

Höglind, M., Jørgensen, M., Østrem, L., Bakken, A., Thorsen, S., 2008. Overwintering of timothy and perennial ryegrass in norway from a climate change perspective. In: Hopkins, A., Gustafsson, T., Bertilsson, J., Dalin, G., Nilsdotter-Linde, N., Spörnly, E. (Eds.), Biodiversity and Animal feed. Future Challenges for Grassland Production. Proceedings of the 22nd General Meeting of the European Grassland Federation, Uppsala, Sweden 9-12 June 2008. Vol. 13 of Grassland Science in Europe. pp. 203–205.

Iizumi, T., Yokozawa, M., Nishimori, M., 2009. Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a bayesian approach. Agricultural and Forest Meteorology 149, 333–348.

Kalberer, S., Wisniewski, M., Arora, R., 2006. Deacclimation and reacclimation of cold-hardy plants: Current understanding and emerging concepts. Plant Science 171, 3–16.

Kanneganti, V., Rotz, C., Walgenbach, R., 1998. Modeling freezing injury in alfalfa to calculate forage yield: I. model development and sensitivity analysis. Agronomy Journal 90, 687–697.

Kass, R., Raftery, A., 1995. Bayes factors. J. A. Stat. Assoc. 90, 773–795.

Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174.

Rapacz, M., 2002. Cold-deacclimation of oilseed rape (brassica napus var. oleifera) in response to fluctuating temperatures and photoperiod. Annals of Botany 89, 543–549.

Roberts, G., 1996. Markov Chain Monte Carlo in Practice, 1st Edition. Interdisiplinary Statistics Series. Chapman & Hall, Suffolk, GB, Ch. 3, pp. 45–57.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. GLOBAL SENSITIVITY ANALYSIS. The Primer. Wiley & Sons Ltd, West Sussex, England.

Sivia, D., 2006. Data Analysis: A bayesian tutorial, 2nd Edition. Oxford University Press, Oxford, England.

Streck, N., Weiss, A., Baenziger, P., 2003. A generalized vernalization response function for winter wheat. Agronomy Journal 95, 155–159.

Sunde, M., 1996. Effects of winter climate on growth potential, carbohydrate content and cold hardiness of timothy (*Phleum pratence* l.) and red clover (*Trifolium pratence* l.). Ph.D. thesis, Agricultural University of Norway.

Thorsen, S. M., Roer, A.-G., Van Oijen, M., 2009. Modelling the dynamics of snow cover, soil frost and surface ice in norwegian grasslands. Submitted for publication.

Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiology 25, 915–927.

Van Oijen, M., Thorsen, S., Schapendonk, A., Höglind, M., 2008. Process-based modelling of timothy survival in winter. In: Organizing committee of IGC/IRC Congress (eds.) (Ed.), Multifunctional Grasslands in a changing world. Vol. 1. Guangdong People's Publishing House, Guangzhou, Peoples Republic of China, p. 893.

Wang, E., Engel, T., 1998. Simulation of phenological development of wheat crops. Agrc. Syst. 58, 1–24.

Paper IV

Stig Morten Thorsen and Mats Höglind

Assessing winter survival of forage grasses in Norway under future climate scenarios by simulating potential frost tolerance in combination with simpler agroclimatic indices.

Submitted to Agricultural and Forest Meteorology.

Assessing winter survival of forage grasses in Norway under future climate scenarios by simulating potential frost tolerance in combination with simple agroclimatic indices

Stig Morten Thorsen[*†‡], Mats Höglind[†]

Abstract

Norwegian agriculture is mainly dominated by grass-based milk and livestock production, so winter damage to overwintering grasses may have large economic consequences. We assessed the impact of climate change on the winter survival of timothy (*Phleum pratense* L.) and perennial ryegrass (*Lolium perenne* L.) under Norwegian conditions using agroclimatic indices and a simulation model of frost tolerance. This study was based on locally adjusted future climate scenarios (two for the period 2071-2100; one for the period 2020-2049) for six important agricultural regions, represented by one location each. We proposed and validated a rough way to estimate the daily minimum air temperatures from scenario data. Compared with the control period 1961-1990, the future hardening period will be shortened by up to 21 days. As a consequence, the modelled maximum frost tolerance is expected to be reduced by up to 3.9°C and 1.9°C for timothy and perennial ryegrass, respectively, under the warmest scenario. In spite of this reduction, the plants are expected to be hardy enough to withstand the predicted autumn frosts, and we also expect a general reduction in the risk of winter frost injuries. The plant data available to this study suggest that agroclimatic indices developed for Canadian conditions can be useful for assessing the hardening status in timothy and perennial ryegrass. However, such indices are less suitable for assessing the risk of plant injury related to frost and ice encasement in Norway, since they do not account for the dynamics of cold adaptation. Although less snow is expected, in most cases this will not be accompanied by an increase in the risk of ice encasement injuries. However, a slight increase in the number of ice encasement events was predicted for one location. An earlier start of growth was predicted for all locations, accompanied at one coastal location by a slightly increased predicted risk of spring frosts. There is little risk of winter injuries related to frost and ice encasement in the hardier grass species timothy. The better overwintering conditions in general indicate that it

[*] Corresponding author. Tel.: +47 404 75 561; fax: +47 514 26 744. *Email address:* stig.thorsen@bioforsk.no (Stig Morten Thorsen)
[†] *Norwegian Institute for Agricultural and Environmental Research, Grassland and Landscape Division. Postvegen 213, N-4353 Klepp st., Norway.*
[‡] Dept. of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.

will be possible to grow perennial ryegrass in areas where it is not grown today, provided the risk of fungal diseases does not increase.

Key words: agroclimatic indices, climate change, frost tolerance, timothy, perennial ryegrass

1. Introduction

Two of the most important sectors in Norwegian agriculture are grass-based milk and livestock production. In 2007 these sectors accounted for 70% of the total income in Norwegian agriculture (Rognstad and Steinset, 2008). Poor overwintering of forage grasses often has negative economic consequences due to substantial yield losses and subsequent costs related to the re-establishment of grass swards and the purchase of supplementary feedstuffs. The global climate is changing, and during the next 100 years the global temperature is expected to increase by 1.4-5.8°C. In Norway, this temperature increase is likely to be larger in northern areas of the country than in the south, and the winter weather is projected to include milder temperatures with more precipitation as rain (NOU, 2009). These changes may affect the winter survival of perennial crops (Bélanger et al., 2002), although no scientific analysis has yet been carried out on the consequences of climate change for overwintering of forage grasses in Norway or in the rest of Scandinavia.

A quick way to characterise plant-climate interactions is by calculating agroclimatic indices (often based on degree-day heat units, which are related to plant growth and development) as opposed to using process-based models for this characterisation (Eitzinger et al., 2008). Bélanger et al. (2002) assessed the potential impact of climate change on overwintering of perennial forage crops in Eastern Canada using agroclimatic indices reflecting the risks of winter injuries. The results suggested that the risk of winter injury in that area will likely increase because of less cold hardening during autumn and reduced protective snow cover during the cold period, which will increase the exposure of plants to killing frosts, soil heave and ice encasement. Similar calculations have not been performed for future Norwegian climate conditions. In addition, the climate in the area studied by Bélanger et al. (2002), eastern Canada, is different from the Norwegian climate in many ways. This study therefore investigated the indices proposed by Bélanger et al. (ibid.) for Norwegian climatic conditions.

In the study, we considered two of the most important forage grasses in Norway, timothy (*Phleum pratense* L.) and perennial ryegrass (*Lolium perenne* L.), to investigate how the

changing climate will affect their winter survival abilities. Grass plants prepare for winter by re-allocating sugars produced during photosynthesis to storage organs, along with other physiological adjustments aimed at increasing their tolerance to sub-zero temperatures and other abiotic and biotic winter stress factors (Andrews, 1987). The process of acquiring frost tolerance is often referred to as cold acclimation, or cold hardening. When springtime comes, or during mild spells in mid-winter, the plants de-acclimate (deharden), and if the mild period is long enough or the temperature high enough, growth is resumed and dehardening becomes irreversible (Kalberer et al., 2006). The more hardened plants are, the better they survive stressful winter conditions. The state of frost tolerance of plants, i.e. their hardening status, can be characterised by their LT50 value, i.e. the temperature at which 50% of the plants are killed. Thorsen and Höglind (2009) proposed a model which simulates LT50 in timothy, and this model was used here for assessing the risk of frost-related injuries under future climate scenarios.

In the present study, we assessed the impact of climate change on winter survival in timothy and perennial ryegrass in Norway i) by calculating three of the agroclimatic indices suggested by Bélanger et al. (2002) reflecting the conditions for hardening in autumn, the risk of dehardening in winter and the risk of injuries related to rain in winter; and ii) by calculating the risk of frost and ice encasement-related injuries using a model that simulates LT50 in timothy and perennial ryegrass (Thorsen and Höglind, 2009), together with a model for soil surface temperature and the development of snow and ice in the field (Thorsen et al., 2009). In addition, we calculated the length of the autumn hardening period, the winter period and the starting day of the thermal growth period. The calculations were carried out for six locations and three climate scenarios (one for the period 2025-2040; two for 2071-2100). The resulting risks of injuries were compared against those obtained from similar calculations performed using a control period (1961-1990). Neither models nor indices can provide 100% reliable answers, but they can serve as alternative approaches to the problem at hand – assessing the risk of winter damage to forage grasses.

2. Materials and Methods

2.1 Definition of autumn, winter and thermal growth periods

We assumed that the plant year consisted of three periods: the thermal growth period (GP), the hardening period (HP) and the winter period (WP). The hardening period was defined to

start at the end of the thermal growth period, and the winter period was defined to end at the start of the thermal growth period. The literature mentions several ways to define the thermal growth period (e.g. Walther and Linderholm, 2006), and thus implicitly the hardening and winter periods. According to Bonesmo (1999), the start of the thermal growth period for timothy can be estimated as being the fifth day of the first five-day spell with daily mean air temperature above 5°C on snow-free ground. We used this definition by Bonesmo (1999) to estimate the start of the growing period, GP_{start}. In the field, hardening is initiated when the daily mean air temperature drops below approximately 10°C, and increases as the air temperature continues to drop below 5°C (Bélanger et al., 2002). We assumed that the end of the thermal growth period could be approximated in the opposite way to the start and thus we defined the start of the hardening period, HP_{start}, as the fifth day of the first five-day spell when $T_{air} \leq 5$°C. For the plants to achieve maximum frost tolerance, a final phase of sub-zero temperatures is required (Kacperska-Palacz, 1978). According to frost tolerance experiments carried out in Norway (Larsen, 1994; Höglind et al., 2006), grass plants usually reach maximum frost tolerance towards the end of the year. To estimate the end of the hardening period, $HP_{stop,}$ we used simulations of LT50 produced using the model developed by Thorsen and Höglind (2009). We located the day of the year when the lowest LT50 occurred and defined this day as HP_{stop}. We defined the start of the winter period, WP_{start}, as the day following HP_{stop}. The end of the winter period, $WP_{stop,}$ was defined as the day prior to GP_{start}. The three periods HP, WP and GP, and the agroclimatic indices used in this study are presented in Table 1.

Table 1 Description of periods and agroclimatic indices used in this study. Units in brackets

Period	Description
HP	Autumn hardening period (d) §
WP	Winter period (d) §
GP	Thermal growth period (d) §
Index	
FH-COLD	Net accumulation of cold degree-days during HP (°C d) †
W-THAW	Mean daily accumulation of degree-days during WP (DD0 d-1) †
W-RAIN	Mean daily rainfall during WP (mm d-1) †
AF	Autumn frost, number of days with lethal temperatures during HP (d) ‡
WF	Winter frost, number of days with lethal temperatures during WP (d) ‡
ID	Ice damage, number of days with ice damage conditions during WP (d) ‡
SF	Spring frost, number of days after start of growth when frost occurs (d) ‡

§ defined in section 2.1
† defined in section 2.2.1
‡ defined in section 2.2.2

2.2 Agroclimatic indices

2.2.1. Previously proposed agroclimatic indices

Bélanger et al. (2002) proposed five agroclimatic indices for assessing the relative risks associated with climatic causes of damage to perennial forage crops during autumn and winter. Two indices express the influence of temperature and rainfall on the acquisition of frost hardiness during fall (FH-COLD, FH-RAIN), and three indices reflect the loss of hardiness due to high temperatures (W-THAW), damage due to frost exposure (W-COLD) and damage due to frost heave and ice encasement during winter (W-RAIN). For the present study, we chose the indices FH-COLD, W-THAW, W-RAIN, and ignored the other two.

The fall hardening index FH-COLD is calculated as the net accumulation of cold degree-days during the autumn (see McMaster and Wilhelm (1997) for interpretation of degree-days). The assumption underlying FH-COLD is that degree-days above 5°C (DD5) (the temperature sum using 5°C as base temperature) and cold degree-days, between 0°C and 5°C (CDD5), can be used to express the impact of temperature on growth and hardening, respectively, and that DD5 and CDD5 have equal but opposite effects. Thus, accumulated CDD5 could be negated by a following accumulation of an equal amount of DD5. In our implementation, the index FH-COLD was calculated based on daily mean air temperature as follows:

$$\text{FH-COLD} = \sum_{i=HP_{\text{startB}}}^{HP_{\text{stopB}}} CDD5_i - \sum_{i=HP_{\text{startB}}}^{HP_{\text{stopB}}} DD5_i \qquad (1)$$

where HP_{startB} and HP_{stopB} mark the start and stop of the hardening period, CDD5 is degree-days when $0 < T_{air} \leq 5$, and DD5 is degree-days using 5°C as base temperature. HP_{startB} and HP_{stopB} were calculated according to Bélanger et al. (2002). HP_{startB} and HP_{stopB} are not to be confused with HP_{start} and HP_{stop} as defined in section 2.1.

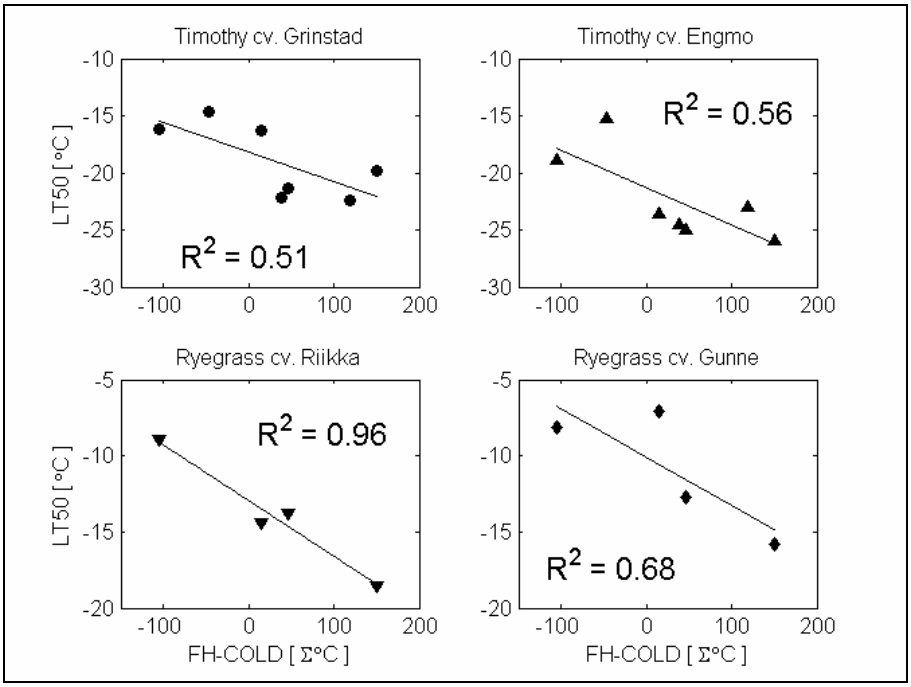


Figure 1: Scatter plot of fall hardening index FH-COLD versus observed LT50 values for timothy cultivars Grindstad and Engmo, and perennial ryegrass cultivars Riikka and Gunne.

Bélanger et al. (2002) proposed that the potential loss of hardiness during winter could be expressed by the index W-THAW, calculated as the mean daily accumulation of degree-days above 0°C (DD0) during the winter period divided by the length of the winter period. In the present study, the index W-THAW (units DD0 d-1), was calculated based on the simulated soil surface temperature T_{surf}. This estimated soil surface temperature (Thorsen et al., 2009) represents the daily mean temperature between the soil surface and the snow cover. Using this T_{surf} temperature instead of the daily mean air temperature incorporates the insulating effect of the snow cover. The refined index W-THAW was calculated as follows:

$$\text{W-THAW} = \sum_{j=HP_{\text{stop}}}^{GP_{\text{start}}} \frac{DD0_j}{[GP_{\text{start}} - HP_{\text{stop}}]} \qquad (2)$$

where DD0 is degree-days above 0°C based on T_{surf}, which is the simulated soil surface temperature, and the difference (GP_{start} - HP_{stop}) represents the duration of winter. FH-COLD and W-THAW were originally constructed for Canadian environmental conditions, which differ in many aspects from Norwegian conditions. Before applying them to Norwegian conditions, we therefore evaluated their usefulness by comparing calculated indices with

observed LT50 values from Norwegian field experiments on the winter survival of timothy and perennial ryegrass. Based on the results presented in Figures 1 and 2, we concluded that the indices FH-COLD and W-THAW were indeed useful for Norwegian conditions.

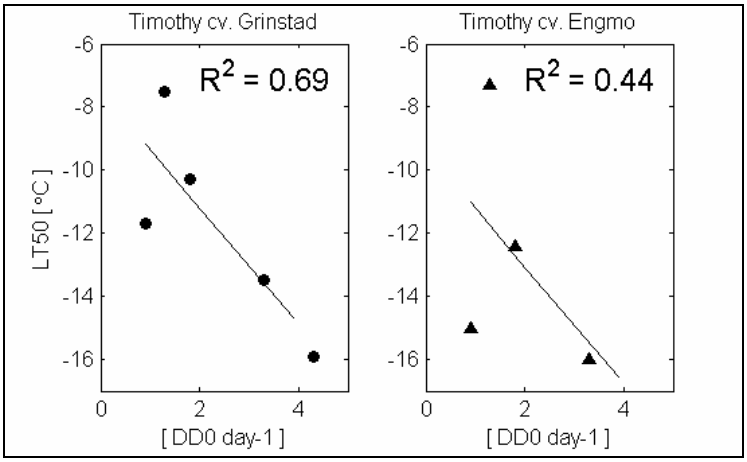


Figure 2: Scatter plot of winter dehardening index W-THAW versus observed LT50 values for timothy cultivars Grindstad and Engmo.

Bélanger et al. (2002) proposed that damage related to ice encasement and soil heave could be expressed by the index W-RAIN. This index was calculated as the mean daily rainfall during the cold period, a period lasting from the date of the first occurrence of $T_{min} \leq -15°C$ to the date of the last occurrence of $T_{min} \leq -15°C$. In our study, W-RAIN was calculated as the average daily rainfall during the winter period WP.

Norwegian experiments reviewed by Höglind et al. (2008) provided inconclusive results on the effect of autumn rainfall on frost hardening in forage grasses. We therefore omitted the index related to rainfall intensity during autumn hardening, FH-RAIN (Bélanger et al., 2002) from the present study. We also omitted the index expressing the number of days the grasses are potentially exposed to freezing temperatures, W-COLD, which is defined as the difference between days with a snow cover depth more than 10 cm and the number of days in the cold period. This gives a relatively rough estimate of the risk of frost injury, although it does not take into account the frost tolerance of the plants at the time of frost exposure. As an alternative, we constructed a new winter frost index WF as described below.

2.2.2. New agroclimatic indices: frost and ice-related indices

In order to assess the risk of exposure to lethal temperatures during the hardening and winter periods, we used simulated LT50 time series obtained using the frost tolerance model developed by Thorsen and Höglind (2009) to estimate the number of days when lethal temperatures occurred, i.e. the number of days when the estimated daily minimum air temperature T_{min} was lower than the simulated LT50-value, $T_{min} < LT50$ or, if snow was present, the number of days when $T_{surf} < LT50$. The estimated number of days was called Autumn Frost (AF) or Winter Frost (WF) depending on whether these days occurred during the autumn hardening or winter period. Index subscripts T and R indicate timothy and ryegrass, respectively. The frost tolerance model was originally developed for timothy grown in Norway, an application for which it has been validated (Thorsen and Höglind, 2009). Simulations of LT50 for perennial ryegrass were performed using the same parameterisation as for timothy except that two parameter values were changed: initial LT50 was set to -4°C, and minimum LT50-value was set to -16°C (see Thorsen and Höglind (2009) for details on LT50-model) based on data for the perennial ryegrass cultivars Riikka and Gunne, which are well adapted to Norwegian conditions (Höglind et al., 2006). This parameterisation yielded good agreement with observed LT50 for the perennial ryegrass cultivars for most of the autumn and winter (Thorsen and Höglind, unpublished data), as illustrated in Figure 3 with examples from two locations (Særheim and Holt), together with examples for timothy cv. Engmo.

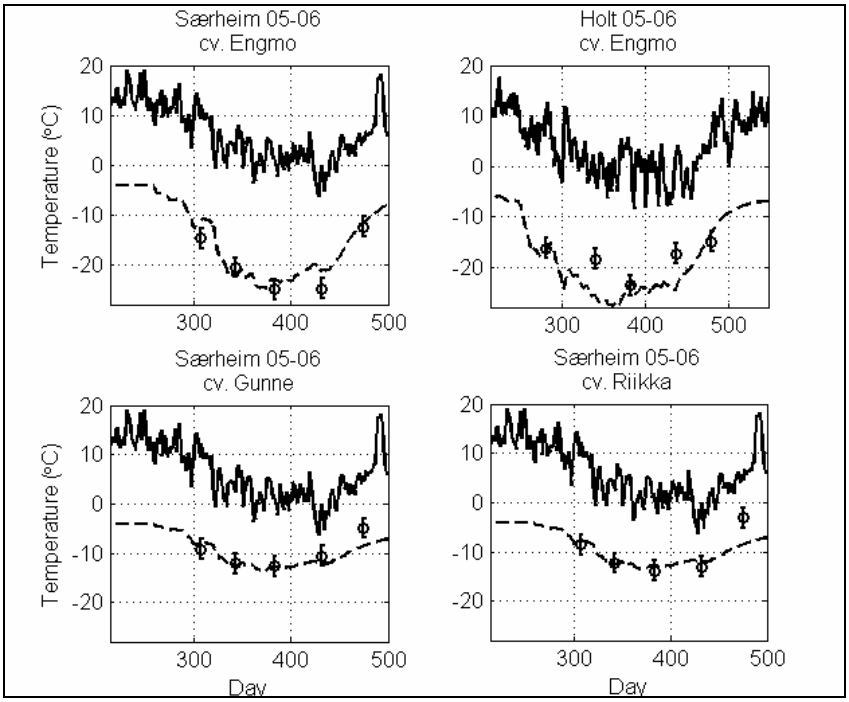


Figure 3: Simulated LT50 values (dashed) for timothy and perennial ryegrass, and simulated soil surface temperature (solid). Circles with vertical line segments denote observed LT50 values ± one standard deviation. The top two diagrams show results for timothy cv. Engmo at Særheim (located approximately 13 km south of Sola) and Holt, both for winter 05-06. The bottom two diagrams show results for perennial ryegrass cvs. Gunne and Riikka for Særheim during winter 05-06.

If water from snowmelt or rain were to accumulate in soil depressions due to the reduced infiltration capacity of frozen soil, a surface ice cover might be induced, leading to ice encasement and damage caused by anoxia and accumulation of CO_2, lactic acid, ethanol and ethylene (Gudleifsson, 1993). The number of days plants can withstand being encapsulated in ice is correlated to their degree of frost tolerance at the start of the ice period. Höglind et al. (2008) found that the ice encasement tolerance of plants, expressed as LD50, was linearly related to their estimated LT50 value at the start of the ice period. For timothy, using underlying data from Höglind et al. (2008) and Jørgensen et al. (2008), we estimated LD50 as:

$$LD50_T = 1.3403 - 2.1128\ LT50_T \qquad (3)$$

where the coefficient of determination was $r^2 = 0.65$ based on n = 12 observations.

For ryegrass, we estimated LD50 as:

$$LD50_R = -7.7041 - 1.5387\ LT50_R \quad (4)$$

where the coefficient of determination was $r^2 = 0.90$ based on n = 12 observations. $LT50_T$ and $LT50_R$ in equations (3) and (4) represent simulated LT50 values for timothy and ryegrass, respectively. The occurrence of ice encasement in the field was simulated using the SnowFrostIce model (Thorsen et al., 2009), which has been calibrated and partly validated for Norwegian conditions. Outputs from the SnowFrostIce model include depth of snow cover and depth of soil frost for different soil types, using standard meteorological measurements on daily average air temperature and precipitation as driving variables. If surface water is simulated in soil surface depressions during winter, the SnowFrostIce model simulates formation of an ice layer. As long as surface water is present and the simulated soil surface temperature is below 0°C, ice is simulated. A thick snow cover can protect plants from frost exposure. In the present study we therefore considered periods when the simulated snow cover was relatively thin (< 20 cm), and estimated the number of days when a simulated ice cover coincided with this reasonably thin simulated snow cover when assessing the risk of ice-related injuries. We then compared the duration of this snow and ice cover period with the estimated LD50 values for timothy and ryegrass at the beginning of the snow and ice period. If the duration of the ice and snow cover outlasted the ice encasement tolerance estimated for the start of the ice encasement period (snow and ice days > LD50) for timothy and ryegrass, this was characterised as one ice damage event. The indices ID_T and ID_R were then calculated as the total number of such ice damage events in timothy and ryegrass, respectively, during one winter.

After the start of thermal growth in spring, here called GP_{start}, plants are particularly vulnerable to the occurrence of low temperatures, as they have generally lost most of their frost tolerance. We proposed an index to reflect the risk of spring frost, defined as the number of days after GP_{start} when the daily minimum air temperature, T_{min}, dropped below -5°C, reflecting the fact that unhardened plants of timothy and perennial ryegrass typically have an LT50 value around -5°C (Gay & Eagles, 1991). If a snow cover was simulated during this period, we used the temperature T_{surf} if the snow cover depth was less than 10 cm. This index, reflecting the risk of injuries due to spring frost following start of growth in spring, was termed SF.

2.3. Climate scenarios and weather stations

The data sets used in this study were empirically adjusted outputs from two different atmosphere-ocean general circulation models (GCM). We selected one GCM model from the Hadley Centre for the control period 1961-1990 (HaCtrl), and the two emission scenarios A2 and B2 for the period 2071-2100 (HaA2 and HaB2 respectively); and output from one GCM from the Max Planck model using the IS92a emission scenario (MPI) for the period 2020-2049 (available from http://noserc.met.no/effect/dynamic/PM1/index.html). These model outputs were downscaled to produce locally adjusted scenario data sets (Engen-Skaugen, 2007). The data sets consist of daily mean air temperature and daily precipitation sum. We used the HaCtrl data set as the baseline climate for comparisons with the other scenarios.

Since the scenario data sets did not include daily minimum air temperatures, T_{min}, these needed to be estimated. We selected six different locations in Norway, representing important agricultural regions, where both empirically adjusted climate scenarios and automatically recorded weather data were available (Table 2).

Table 2 Description of the six representative locations included in this study

Agricultural region, climate	Represent. location	Elevation m.a.s.l.	Latitude North	Longitude East
South-west, coastal	Sola	7	58° 53'	5° 38'
South-east, lowland	Oslo	94	59° 56'	10° 43'
South-east, inland	Løken	525	61° 7'	9° 4'
Central Norway, fiord	Værnes	12	63° 27'	10° 56'
Northern, coastal	Bodø	11	67° 16'	14° 26'
Northern, coastal	Tromsø	100	69° 39'	18° 55'

From weather stations at each of the six representative locations in this study, we obtained data on daily average air temperature and daily minimum air temperature for each day of the calendar year. At each weather station, approximately 20 observations were available for each day. We estimated mean and standard deviation of the difference between daily average and daily minimum air temperatures. Based on these daily temperature differences, termed ΔT_d, we estimated the daily minimum temperature on day d, $T_{min,d}$ as:

$$T_{min,d} = \begin{cases} T_{air,d} - \Delta T_d & \text{if Prec} > 0 \\ T_{air,d} - (\Delta T_d + 0.68\,\sigma_{\Delta T}) & \text{if Prec} = 0 \end{cases}$$

where $T_{air,d}$ is the daily mean temperature from the climate scenario, ΔT_d is the estimated daily temperature difference, and $\sigma_{\Delta T}$ is the standard deviation of ΔT_d. The net global radiation is less on cloudy days than on clear days, and thus on cloudy days the temperature amplitude is likely to be smaller. The climate scenarios did not contain information on cloudiness, so we assumed days with precipitation to be cloudy, and days with no precipitation to be clear. To incorporate this information, we used the temperature difference ΔT_d on days with precipitation, while on days with no precipitation we used the temperature difference ΔT_d plus a $\sigma_{\Delta T}$ from the 75% quantile (assuming the difference ΔT_d follows a normal distribution, this means adding 0.68 $\sigma_{\Delta T}$ to the temperature difference ΔT_d).

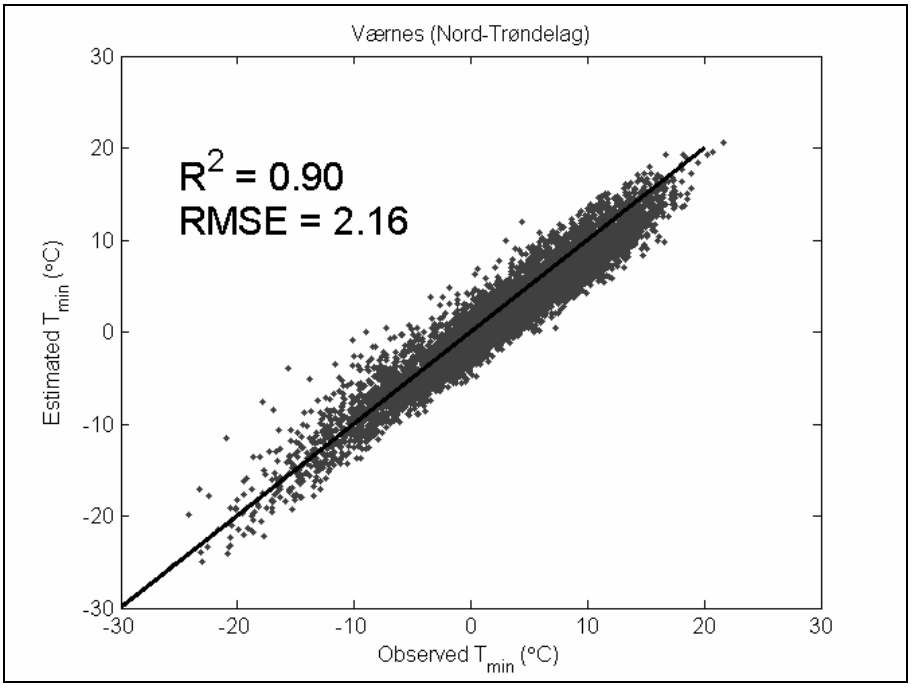


Figure 4: Scatter plot of observed versus estimated daily minimum air temperature at Værnes for the period using data from 1988-2008.

Figure 4 shows a scatter plot of observed versus estimated daily values of T_{min} for Værnes (Nord-Trøndelag) using observations from the period 1.9.1988 – 31.8.2008. The root mean

squared error was calculated as $RMSE = \sqrt{\left(\frac{1}{n}\sum_{d=1}^{n}(T_{minObs,d} - T_{min,d})^2\right)}$, where $T_{minObs,d}$ was observed daily minimum air temperature on day *d*. The median value of the residuals ($T_{minObs,d}$ - $T_{min,d}$) was 1.23, indicating that $T_{min,d}$ was underestimated. Figure 4 indicates that this underestimation mostly occurred on days when the minimum air temperature was above 0°C.

2.4. Statistical analysis

To assess whether the results for the different scenarios differed from those for the control period, we used the Wilcoxon rank sum test and did a pair-wise comparison of the results from the control period and the scenarios. This test performs a two-sided rank sum test of the null hypothesis that indices, calculated for the control period and the different scenarios, respectively, are independent samples from identical distributions with equal medians, against the alternative that the medians are not equal. We rejected the null hypothesis at the 5% significance level. Significant differences are marked in Figures 6-13 with asterisks (*). All calculations were performed using MATLAB.

3. Results

3.1. Length of autumn hardening period HP

Mean dates of the estimated start and end of the autumn hardening period (HP_{start} and HP_{stop}) for all locations and scenarios are presented in Table 3, together with the mean length of the autumn hardening period (HP_{stop} - HP_{start}). Both HP_{start} and HP_{stop} were delayed for all locations and future scenarios in comparison with the control period, and the hardening period was shortened as a result of HP_{start} being more delayed than HP_{stop}. This is also illustrated graphically for the warmest scenario HaA2 in Figure 5. The delay in HP_{start} varied between 8 and 34 days and was significant for all scenarios and locations. The delay in HP_{stop} varied between zero and 21 days and was significant for all scenarios at the Tromsø site; for two scenarios at Bodø, Værnes, Oslo and Sola, and for one scenario at Løken. As a result, the autumn hardening period was shortened for all scenarios and locations except MPI at Tromsø.

The largest reduction, 21 days, was observed for HaB2 at Sola. Significant reductions in the length of HP are indicated by * in Table 3.

Table 3 Mean dates of start and end of the hardening period (HP_{start} and HP_{stop}, respectively), mean length of the hardening period (HP_{stop} - HP_{start}) and mean dates for start of growth in spring (GP_{start}) for the control period and the three future climate scenarios. Significant reduction in length of hardening period indicated by *

Location Scenario	HP_{start}	HP_{stop}	Length of HP	GP_{start}
Sola				
HaCtrl	13 Nov	02 Jan	50	09 Mar
MPI	22 Nov	02 Jan	41 *	31 Jan
HaB2	15 Dec	13 Jan	29 *	08 Feb
HaA2	14 Dec	20 Jan	37	09 Feb
Oslo				
HaCtrl	24 Oct	20 Dec	57	15 Apr
MPI	02 Nov	27 Dec	55	01 Apr
HaB2	13 Nov	03 Jan	51	02 Apr
HaA2	21 Nov	31 Dec	40 *	28 Mar
Løken				
HaCtrl	30 Sep	3 Dec	64	16 May
MPI	10 Oct	12 Dec	63	12 May
HaB2	26 Oct	16 Dec	51 *	02 May
HaA2	25 Oct	11 Dec	47 *	27 Apr
Værnes				
HaCtrl	2 Oct	25 Dec	65	11 Apr
MPI	01 Nov	30 Dec	59	26 Mar
HaB2	15 Nov	11 Jan	57	27 Mar
HaA2	22 Nov	14 Jan	53	24 Mar
Bodø				
HaCtrl	19 Sep	23 Dec	65	06 May
MPI	27 Oct	26 Dec	60	01 Apr
HaB2	11 Nov	03 Jan	53	29 Mar
HaA2	13 Nov	07 Jan	55	02 Apr
Tromsø				
HaCtrl	26 Sep	8 Dec	73	02 Jun
MPI	7 Oct	20 Dec	74	09 May
HaB2	26 Oct	26 Dec	64	26 Apr
HaA2	30 Oct	29 Dec	60 *	16 Apr

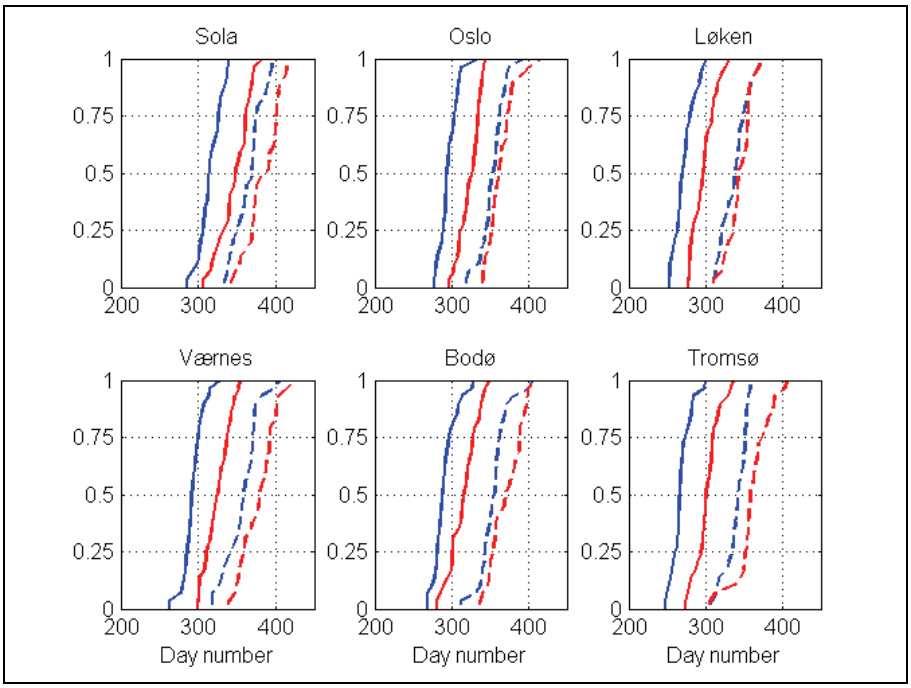


Figure 5: Cumulative frequency curves for the start (solid) and the end (dashed) of the hardening period. Blue lines represent the control period 1961-1990, and red curves represent climate scenario HaA2 for the period 2071-2100.

3.2. Fall hardening index FH-COLD

Values of the FH-COLD index, representing the hardening status of the plants according to Bélanger et al. (2002), are shown in Figure 6.

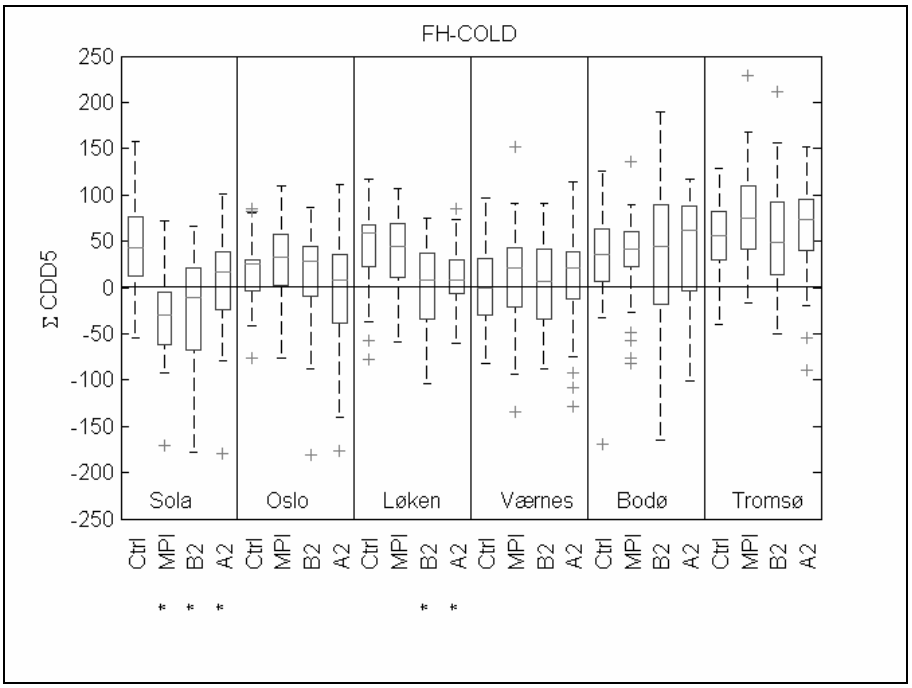


Figure 6: Box plots of the index FH-COLD calculated for the different locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentiles, line within boxes indicates median, whiskers extend to extreme values and possible outliers marked by +. Asterisks* indicate significant differences.

For some combinations of location and scenario, a decrease in FH-COLD compared with the control period was predicted, whereas for others no change or a non-significant increase in FH-COLD was predicted. The predicted decrease in FH-COLD was significant for all three scenarios at Sola. Here, the largest reduction was expected under MPI, where the median FH-COLD value was reduced from 43.2 CDD5 to -30.3 CDD5. The negative values of FH-COLD for Sola suggest more days with temperatures above 5°C during the hardening period. At Løken a significant reduction was found for HaB2 and HaA2. For the remaining locations, none of the differences between scenario and control period were significant.

3.3. Maximum frost tolerance LT50

Values of maximum frost tolerance LT50 at the end of autumn hardening are shown in Figure 7 for timothy (top panel) and perennial ryegrass (lower panel). For Sola, a significant decrease in frost tolerance was predicted for all scenarios compared with the control period. For both timothy and perennial ryegrass, the largest reduction was expected under the HaA2 scenario,

where the median maximum LT50-value was reduced by 3.9°C and 1.9°C for timothy and ryegrass, respectively. The predicted decrease at the other locations varied between 0 and 2°C. A significant decrease in maximum frost tolerance was also predicted for scenarios HaB2 and HaA2 for Oslo, Løken, Værnes and Bodø, and for HaA2 for Tromsø.

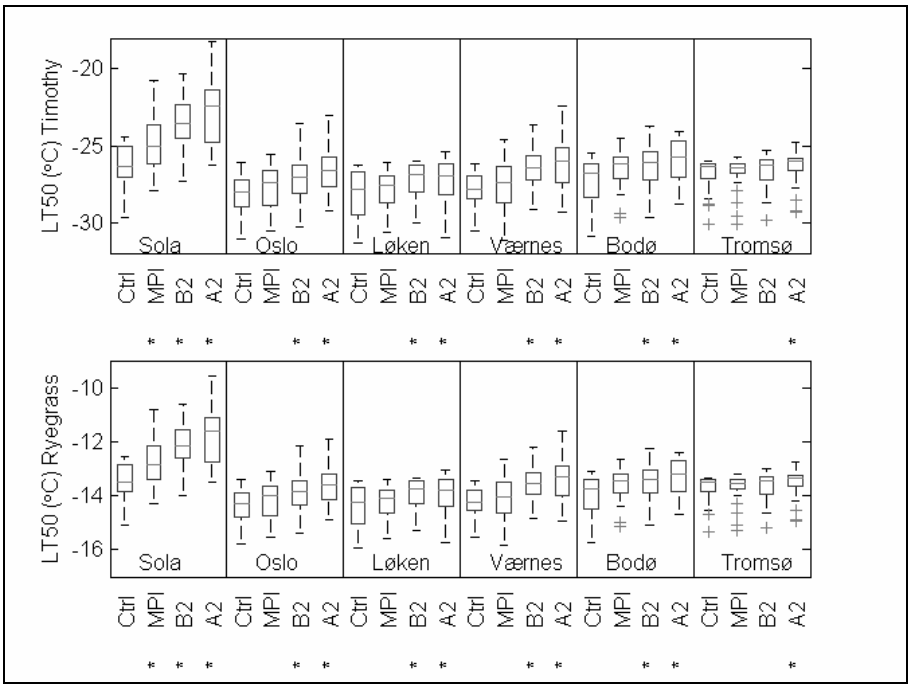


Figure 7: Box plots of simulated maximum frost tolerance (LT50) for timothy (top panel) and perennial ryegrass (bottom panel) calculated for the different locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentiles, line within boxes indicates median, whiskers extend to extreme values and possible outliers marked by +. Asterisks * indicate significant differences.

3.4. Risk of frost injury during autumn hardening AF_R and AF_T

Values of the autumn frost index for perennial ryegrass (AF_R) are shown in Figure 8. For Oslo, a significant decrease in AF_R was predicted for HaB2 and HaA2 (both to zero days) compared with the control period (one day), while for Sola there was a significant decrease to zero occurrences for the MPI scenario.

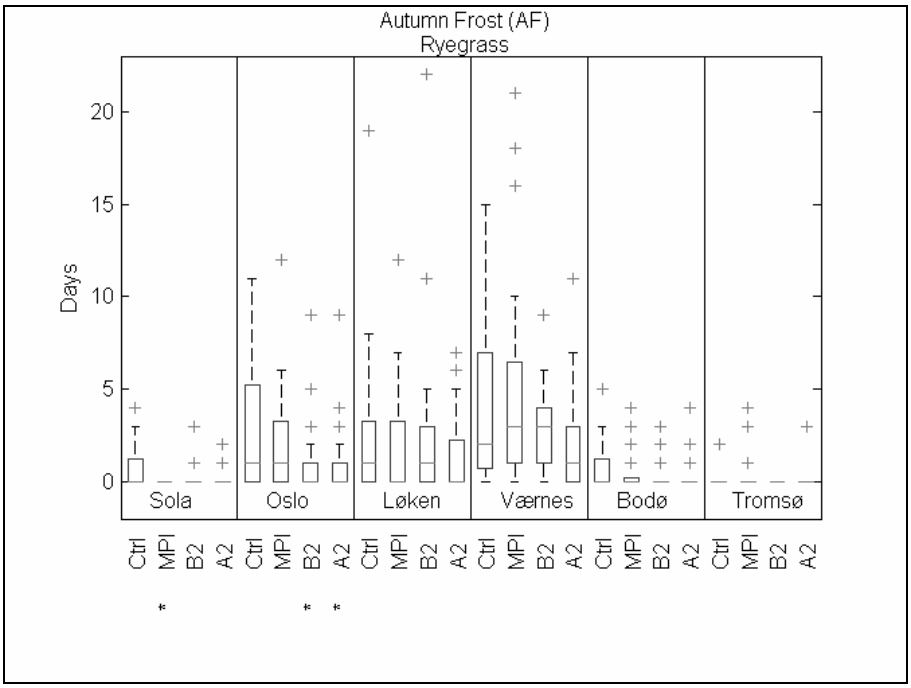


Figure 8: Box plot of the autumn frost index AF for perennial ryegrass calculated for the different locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentiles, line within boxes indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

At Værnes, the results indicated a relative large but non-significant increase in AF_R for MPI and HaB2. The remaining results for AF_R were small and non-significant. There was no significant effect of climate change on AF_T at any location.

3.5. Risk of frost injury during winter WF_R and WF_T

Values for WF_R are shown in Figure 9. Oslo and Værnes had the largest risk of frost injuries among locations during the control period. A significant decrease from 3 days in the control period to 1 day was predicted for all scenarios at Oslo and for HaA2 at Værnes. A smaller significant reduction was predicted for two scenarios at Sola and Bodø, and for one scenario at Løken. For the remaining locations, the main difference between the control period and the scenarios was a tendency for fewer WF_R days overall under the scenarios.

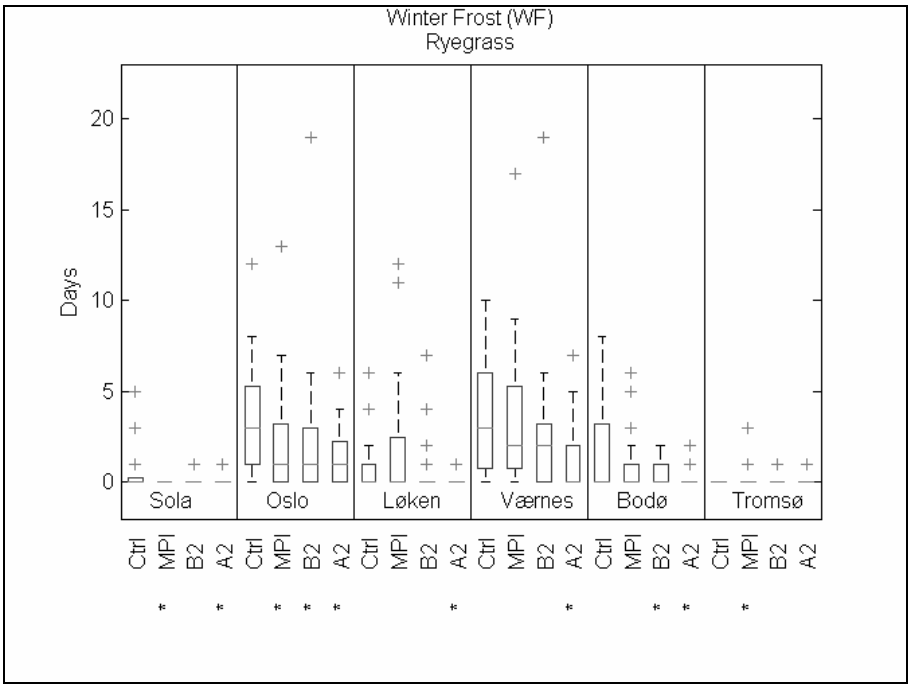


Figure 9: Box plot of the winter frost index WF for perennial ryegrass calculated for all locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentiles, line within boxes indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

There were practically no occurrences of winter frost events for timothy at any of the locations for all three scenarios and the control period (not shown).

3.6. Risk of ice damage ID and W-RAIN

Values for the index ID are presented in Figure 10. For Oslo, a significant decrease was predicted from 1 event under HaCtrl to 0 events under all scenarios. A significant decrease in ID compared with the control period was also predicted for two scenarios at Værnes and two at Bodø. For Tromsø, a small but significant increase was predicted for HaB2. There were no significant changes in ID for Sola and Løken.

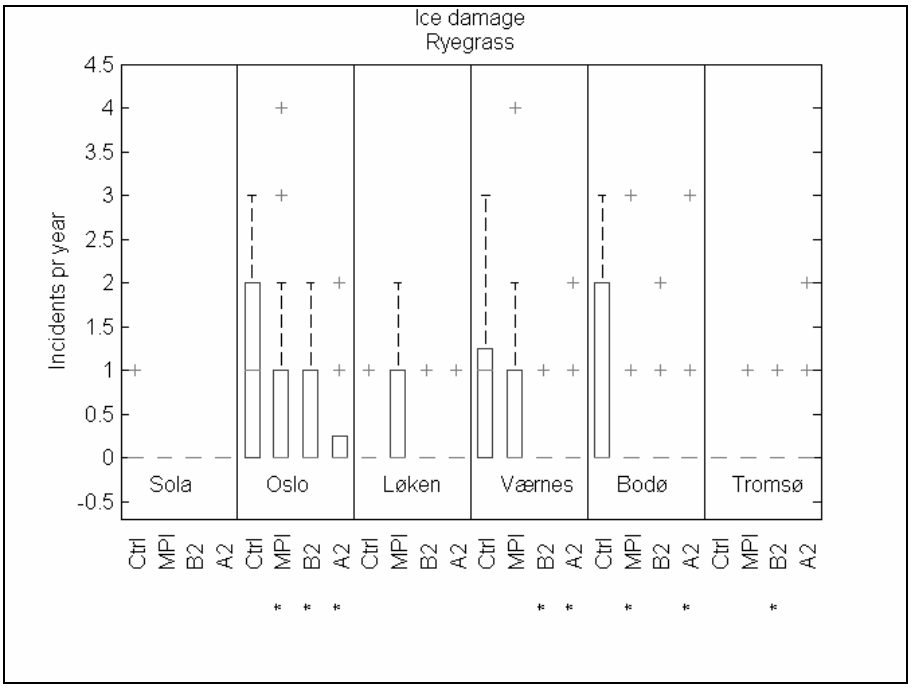


Figure 10: Box plot of results for the number of ice damage events ID per year for perennial ryegrass calculated for all locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentiles, line within boxes indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

Values for W-RAIN for the different scenarios and locations are shown in Figure 11. For most combinations of location and scenario, an increase in W-RAIN was predicted compared with the control period. The predicted increase was significant for all scenarios for Oslo and Tromsø, for two of the scenarios (MPI and HaB2) for Bodø and for one of the scenarios (MPI) for Sola and Værnes. The largest change was predicted for Sola, where W-RAIN was expected to increase from 2.6 mm per day under HaCtrl to 5.0 mm per day under MPI.

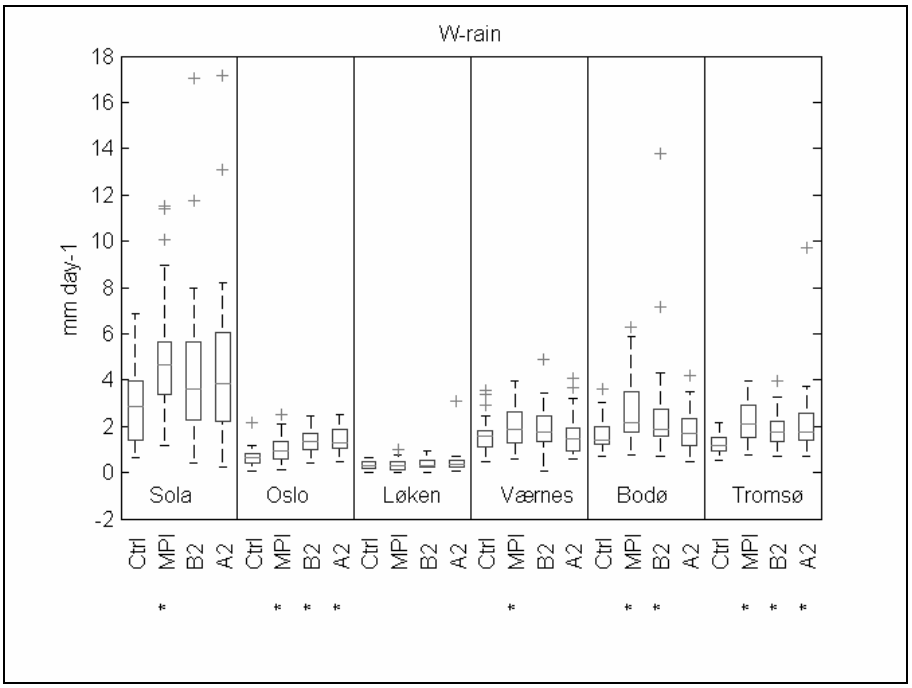


Figure 11: Box plots of results for the ice damage index W-RAIN calculated for all locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentile, line within box indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

3.7. Risk of dehardening during winter W-THAW

Mean values for W-THAW are presented in Figure 12. For all but one combination of location and scenario, W-THAW was predicted to increase, suggesting an elevated risk of dehardening. This increase was significant for all three scenarios compared with the control period at all locations except Løken, where W-THAW was significantly larger only for scenario HaA2. The largest increase was expected at Sola, where W-THAW increased from 3.0 DD0 per day under HaCtrl to 5.0 DD0 per day under MPI.

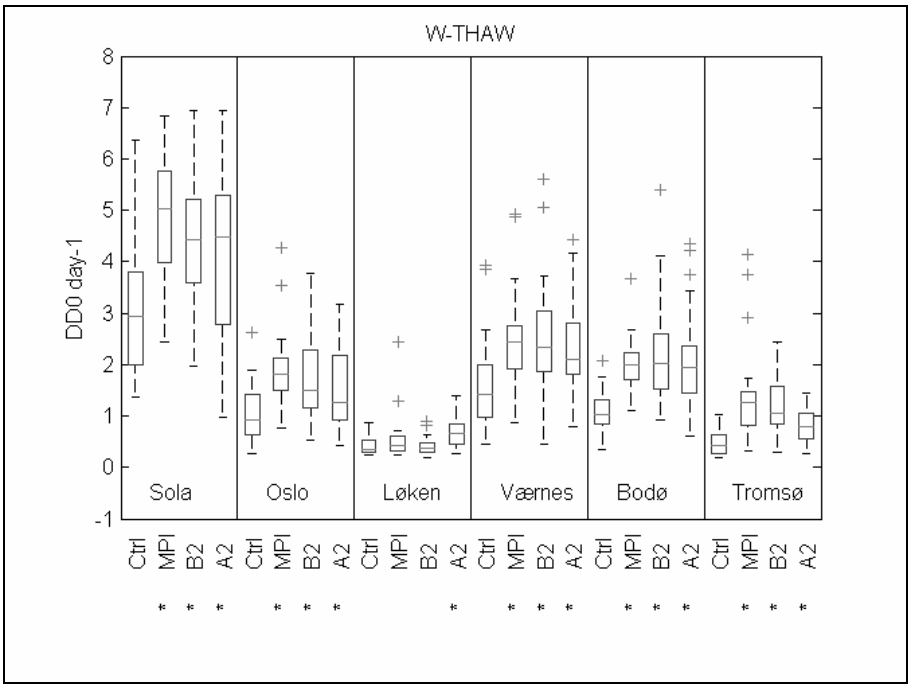


Figure 12: Box plot of results for the dehardening index W-THAW calculated for the different locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentile, line within boxes indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

3.8. Risk of frost injury during spring growth SF

Values of SF for the different locations and scenarios are shown in Figure 13. A significant increase was predicted for all scenarios at Bodø. No significant changes were predicted for the remaining locations and scenarios. However, the results shown in Figure 13 indicate that Sola was expected to experience more SF events under the scenarios, although the increases were not significant according to the Wilcoxon tests.

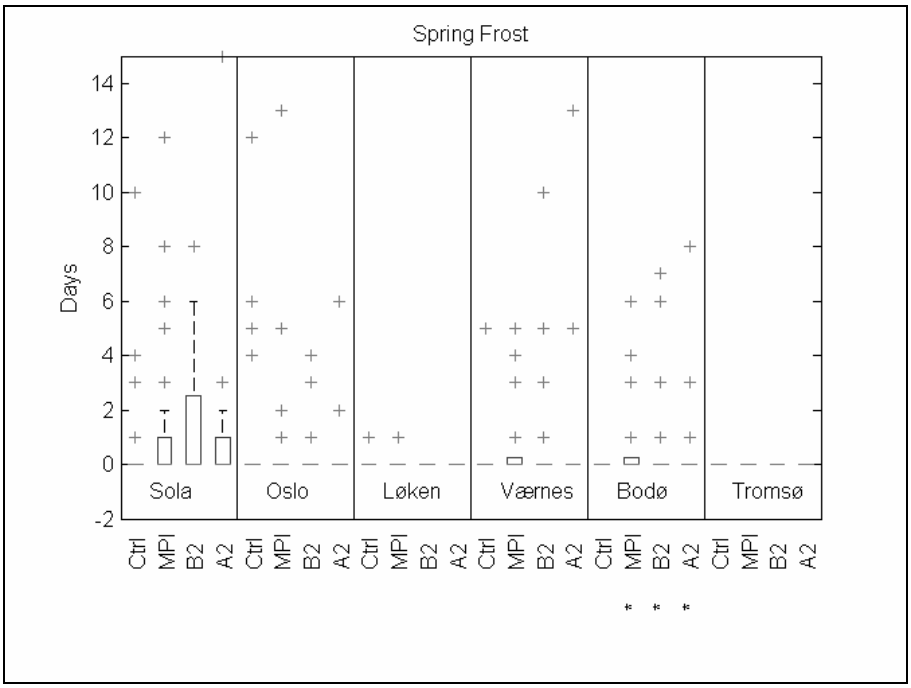


Figure 13: Box plots of results for the spring frost index SF calculated for all locations and scenarios; upper and lower edges of boxes indicate 75th and 25th percentile, line within boxes indicates median, whiskers extend to extreme values, possible outliers marked by +. Asterisks * indicate significant differences.

4. Discussion

The climate data used in this study were generated by two global circulation models (one from the Hadley Centre and one from the Max Planck Institute) and further locally adjusted using dynamic downscaling by a regional climate model followed by empirical local adjustments (Engen-Skaugen, 2007). As a baseline climate we used the control period 1961-1990 from the Hadley Centre (termed HaCtrl) when comparing effects under the projected scenarios IPCC SRES A2 and B2 (termed HaA2 and HaB2, respectively) and IS92a (termed MPI). The general pattern in Norway under these scenarios is for a marked west-east and south-north temperature gradient, with the highest increases in the east and north (under the scenario HaA2 the increase was < 3°C in the west to > 3.5°C in the east and > 4.0°C in the north-east) (Engen-Skaugen, 2007). Another trait in these scenarios is for the temperature increase to be larger during winter, i.e. the winters are projected to be milder, with more precipitation as rain (Beldring et al., 2008).

All calculations made using the simulation model were based on the assumption that current cultivars, well adapted to Norwegian conditions, were used. Any possible change in winter hardiness as a result of plant breeding were not considered. In general, plant breeding could be expected to result in cultivars with improved performance. However, it is important that cultivars adapted to the specific photoperiodic pattern at higher latitudes are available (Rognli, 1998).

4.1. The hardening period

Our calculations suggest that the increase in temperature will have largest negative effect on frost hardening in autumn in areas of Norway already experiencing a variable autumn climate, such as the south-west region, represented in this study by Sola. Regions like this coastal area, currently receiving little snow which may serve as protective insulation, are in this respect particularly vulnerable to freezing temperatures. However, at the same time minimum air temperatures are less extreme at these coastal locations. Our index AF (as well as WF) captured this by calculating the risk of frost actually hitting the plant, accounting for the insulating effect of snow cover if present. Hardening was predicted to start later at all locations studied, especially under the HaA2 scenario (between 3 to 5 weeks; Table 3). For inland locations predicted to receive much snow and experience stable winters even in the future (e.g. Løken), the end of hardening will not change much, but the hardening period will still be shortened due to delayed onset of hardening. Our finding of up to three weeks shorter hardening periods in the future confirmed earlier results produced by Engen-Skaugen and Tveito (2004) using a slightly different definition of thermal growth period than ours. In comparison, Bélanger et al. (2002) predicted an average decrease in the length of hardening period in the range of 1 to 6 days for Canadian conditions.

The results for the index reflecting the hardening status of the plants by the end of the hardening period, FH-COLD, gave similar results to the simulations of the maximum frost tolerance LT50 for the locations Sola, Løken and Tromsø. This was in contrast to Oslo, Værnes and Bodø, where LT50 and FH-COLD gave opposite results. It would be interesting to run our frost tolerance model for the Canadian climate scenarios used by Bélanger et al. (2002), to see whether we get similar disagreement of predictions by FH-COLD and simulated maximum LT50, and to clarify which method gives the most accurate results.

The results of the index FH-COLD do not indicate quantitatively how much differently the plants will harden in the future, only that there will be a change. In contrast, the model simulations give a quantitative estimate. Based on the simulations of LT50, we concluded that even if the hardening period becomes shorter in the future, it will still be long enough to provide substantial hardening, so in the end the maximum frost tolerance will only be reduced by 3°C at worst, and most often by less than 2°C. The predicted decrease in frost tolerance is in agreement with observations from an open top chamber experiment at Særheim, 15 km from Sola, where a 2°C increase in air temperature throughout autumn and winter compared with ambient temperatures caused a decrease in frost tolerance of about 1°C (Höglind et al., 2008).

According to AF_R and AF_T, the risk of frost injury during the autumn period will decrease or remain unchanged in the future compared with today. Thus, even if the frost tolerance is reduced during the hardening period, the plants will still be hardy enough to withstand the predicted autumn frosts as they are generally less intense than those in the control period.

4.2. The winter period

Our model simulations (which formed the basis for calculating WF_R, WF_T and ID) indicated that even if plants harden less in the future, frost and ice encasement-related injuries during winter will decrease at most locations. General warming of the climate is likely to be accompanied by less frost exposure for plants and soil, leaving the soil more able to maintain its drainage capabilities (little or no ice in the soil pores to impede soil water infiltration). Better drainage capability will lead in turn to a reduced risk of ice encasement episodes, or at least a reduction in the duration of such episodes, as reflected by the ice damage index ID. One exception is Tromsø, for which an increased risk of ice encasement injuries is predicted due to a rather dramatic decrease in snow cover (snow data not shown here) accompanied by increased precipitation as rain in combination with frost episodes where ice can form on the surface. Other locations in Norway with similar winter conditions to Tromsø may also experience an increased risk of ice encasement. However, the increase will not be dramatic, and the general trend for Norway will be a reduction in the risk of ice encasement events.

The results from the W-RAIN index mostly contradicted the results from the ice damage index ID. The cold inland regions of Norway have a longer protective snow cover than milder

coastal regions in Norway or in Eastern Canada. For these inland regions of Norway, rain during winter might just as well be absorbed by the snow cover (the snow cover being a porous medium with the ability to retain a limited amount of liquid water) rather than forming an ice cover. The Canadian winter conditions are expected to have less protective snow cover than in the Norwegian inlands (Bélanger et al., 2002), leaving the crops and soil more exposed to rain and frost events. In the milder regions of Norway there is currently not much soil frost, so rain during winter is more likely to cause waterlogging than ice sheeting. Thus the index W-RAIN, which was developed for eastern Canadian conditions, might be less suitable for Norwegian conditions. On the other hand, the SnowFrostIce model has only been partly validated for the ice cover simulation due to lack of data (Thorsen et al., 2008). When more data become available, the model should be revalidated, and if necessary the scenario simulations should be repeated with an improved ice encasement model.

The south-west location Sola had the highest W-THAW values, indicating that this region is most vulnerable to a temperature increase during winter in terms of the risk of warm spells causing the plants to deharden. The high W-THAW values for Sola can be related to the shorter winter period predicted for this location compared with the others. Locations having a relatively long and stable snow cover, e.g. Løken, are less prone to experience serious dehardening according to W-THAW (Figure 9). W-THAW gives a rough estimate of the risk of dehardening but does not quantify the number or intensity of dehardening events followed by temperature drops causing frost kills. The index Winter Frost (WF) gives a more detailed picture by calculating the number of events where the simulated LT50 value is lower than the simulated daily minimum temperature. For ryegrass, WF_R showed a significant decrease in the risk of frost events during winter for Oslo, whereas for the remaining locations no significant change was predicted, although the general trend was for a slight decrease in the number of events.

The index WF_T suggests there is little or no risk of frost injuries in timothy swards in winter, either under the future scenarios or in the control scenario at any location.

4.3. Spring frost

The risk of frost injury in spring depends on how much the growing season is extended into the winter period, the number of air frosts during the prolonged growing season, and the

absence or presence of a protective snow cover during frost exposure. Spring growth will start earlier at all locations according to our calculations. At Bodø, the earlier start of growth will be accompanied by an increased risk of frost injury after the start of growth, as reflected by the increased SF compared with the control period. For the other locations, the risk of frost injury in spring will not change significantly, although a strong tendency for increased SF was also evident for Sola. At coastal locations such as Bodø and Sola, the sea acts like a thermal buffer, keeping temperatures high enough so that growth can start early in spring. However, freezing temperatures may still occur during the initial growth phase, which is rather long compared with inland locations where the temperature curve after start of growth often has a steeper slope. The spring frost (SF) and start of growth (GP) indices are dependent upon accurate simulation of snow cover, and modelling studies have shown that snow simulation in coastal areas is more difficult than in more continental regions (Thorsen et al., 2008). Hence these indices (SF and GP) are also subject to the uncertainty related to the snow cover simulations. Even though an increase in SF was predicted for Bodø, it should be noted that this increase was relatively small and mainly due to a limited number of years when SF > 6 days.

4.4. Agronomic implications

The results of our research indicate a longer growing season in autumn and an earlier start of growth in spring throughout Norway. This implies that farmers might get one or two extra grass harvests in the future, and/or a prolonged grazing period, compared with the situation in the control period. On the other hand, if the increase in the growing season in autumn is accompanied by more precipitation as rain, the fields may be too wet to harvest the extra regrowth. In countries such as Norway with relatively harsh winter conditions, winter survival is a critical factor. The present study indicates better overwintering conditions in general for timothy and perennial ryegrass in the future. However, at some locations, some types of winter stress may increase, e.g. the risk of spring frosts may increase at some coastal locations such as Bodø, and there may also be an increased risk of ice encasement injury in areas where the length of snow period is drastically reduced, as in Tromsø. The better overwintering conditions in general indicate that it will be possible to grow perennial ryegrass in areas where it is not grown today, provided that the risk of fungal diseases does not increase.

It should be borne in mind that the present study deals with the risk of abiotic winter injuries. When models for biotic winter injuries become available, it is greatly important that the present study be supplemented with a study of winter survival in future climate scenarios in relation to the risk of winter injuries related to biotic factors (e.g. snow mould fungi).

4.5. Usefulness of combining model simulations with simpler agroclimatic indices

The plant data available for this study suggest that the agroclimatic indices developed for Canadian conditions (Bélanger et al., 2002) can be useful for assessing the hardening status in Norwegian forage grasses, but they are less suitable for assessing the risk of plant injury related to frost and ice encasement in Norway since the dynamics of hardening are not accounted for. The simulated frost tolerance in timothy and perennial ryegrass was useful when assessing a more complex interaction between winter survival and climate effects, such as snow and ice cover. Results from simulation models can be summarised and presented as agroclimatic indices for the use of decision-makers and agricultural advisors.

5. Conclusions

Climate scenarios for Norway indicate an increase in temperature for all seasons. Based on simulations, we concluded that even if the hardening period is shorter in the future, it will remain long and cold enough to provide substantial hardening. Thus, the maximum frost tolerance will in most cases be reduced by less than 2°C. For most locations, the risk of frost-related injuries during the hardening, winter and spring growth periods will be somewhat reduced according to existing scenarios. However, some coastal locations may experience a slightly increased risk of frost injury after start of growth in spring. A slightly increased risk of ice encasement injury may also be expected at some locations where the length of the snow period is dramatically decreased. However, the general trend is for a reduced risk of plant injury related to ice encasement. Our predictions are directly related to snow and ice cover simulations, and thereby subject to the uncertainty in the underlying snow model.

Acknowledgements

The authors would like to thank Hans Martin Hanslin for valuable comments on an earlier version of this manuscript. This work was supported by the Norwegian Research Council and Bioforsk as part of the WINSUR strategic research program, ‘Climate change effects on

winter survival of perennial forage crops and winter wheat and on plant diseases and weed growth and control at higher latitudes', 2004-2009 (Grant number 158 934 / I 30).

References

Andrews, C.J., 1987. Low-temperature stress in field and forage crop production. Canadian Journal of Plant Science 67, 1121-1133

Bélanger, G., Rochette, P., Gastonguay, Y., Bootsma, A., Mongrain, D., Ryan D., 2002. Climate Change and Winter Survival of Perennial Forage Crops in Eastern Canada. Agronomy Journal 94, 1120-1130

Beldring, S., Engen-Skaugen, T., Førland, E.J., Roald, L.A., 2008. Climate change impacts on hydrological processes in Norway based on two methods for transferring regional climate model results to meteorological station sites. Tellus 60A 3, 439-450

Bonesmo, T., 1999. Spring Growth and Regrowth Rates of Timothy and Meadow Fescue Swards. PhD Thesis, Agricultural University of Norway

Eitzinger, J., Thaler, S., Orlandini, S., Nejedlik, P., Kazandjiev, V., Vucetic, V., Sivertsen, T.H., Mihailovic, D.T., Lalic, B., Tsiros, E., Dalezios, N.R., Susnik, A., Kersebaum, K.C., Holden, N.M., Matthews,R. 2008. Agroclimatic indices and simulation models. i*n* Nejedlik,P. and Orlandini S. (editors) *Survey of agrometeorological practices and applications in Europe regarding climate change impacts.* Report, COST Action 734.

Engen-Skaugen, T., 2007. Refinement of dynamically downscaled precipitation and temperature scenarios. Climate Change 84, 365-382

Engen-Skaugen, T., Tveito, O.E., 2004. Growing-season and degree-day scenario in Norway for 2021-2050. Climate Research 26, 221-232

Engen-Skaugen, T., Haugen, J.E., Tveito, O.E., 2007. Temperature scenarios for Norway: from regional to local scale. Climate Dynamics 29, 441-453

Gay, A.P., Eagles, C.F., 1991. Quantitative analysis of cold hardening and dehardening in *Lolium*. Annals of Botany, 67, 339-345.

Gudleifsson, B.E., 1993. Metabolic and cellular impact of ice encasement on herbage plants. *In* M.B. Jackson and C.R. Black (ed.) Interacting stresses on plants in a changing climate. NATO ASI Ser., Vol. 16, Springer-Verlag, Berlin. p. 407-421.

Gudleifsson, B., Bjørnsson, H., 1989. Methods for estimating ice encasement tolerance of grasses in the laboratory. Icelandic Agricultural Sciences 2, 99-103

Hansen-Bauer, I., Førland, E.J., 2009. Climate in Norway 2100. NOU 1, 1-136 (in Norwegian)

Höglind, M., Jørgensen, M. Østrem, L., 2006. Growth and development of frost tolerance in eight contrasting cultivars of timothy and perennial ryegrass during winter in Norway. Proceedings of NJF Seminar 384 10-12 August 2006, Akureyri, Iceland: 50-53.

Höglind, M., Jørgensen, M., Østrem, L, Bakken AK, Thorsen S.M., 2008. Overwintering of timothy and perennial ryegrass in Norway from a climate change perspective. Grassland Science in Europe 13: 203-205.

Jørgensen, M.., Østrem, L., Bakken A.K., Höglind M., 2008. Effects of ice encasement during winter on different cultivars of Phleum pratense and Lolium perenne in Norway. In: Organizing Committee of IGC/IRC Congress (eds) Multifunctional Grasslands in a changing world, Volume 1, p. 915. Guangzhou Peoples Republic of China: Guangdong People's Publishing House.

Kacperska-Palacz, A., 1978. Mechanisms of cold acclimation in herbaceous plants. In: Li PH, Sakai A. (eds). Plant cold hardiness and freezing stress. Academic Press New York, p. 139-152.

Kalberer, S.R., Wisniewski, M., Rajeev, A., 2006. Deacclimation and reacclimation of cold-hardy plants: Current understanding and emerging concepts. Plant Science 171: 3-16

Larsen A., 1994. Breeding winter hardy grasses. Euphytica 77, 231-237.

McMaster, G., Wilhelm, W.W., 1997. Growing degree-days: one equation, two interpretations. Agricultural and Forest Meteorology 87: 291-300

NOU, 2009. Climate in Norway 2100 (in Norwegian). Available online at: http://www.regjeringen.no/upload/MD/Kampanje/klimatilpasning/Bilder/NOU/Klima-Norge-2100_lavoppl_2opplag_okt2009.pdf

Rognli, O.A. 1998. Species aspects of breeding herbage varieties for northern marginal regions. Journal of Agricultural Science in Finland 60, 181-189.

Rognstad, O., Steinset, T.A., 2008. Landbruket i Norge 2007. Jordbruk-Skogbruk-Jakt. Statisktiske analyser 101. Statistics Norway (www.ssb.no)

Thorsen, S.M., Roer, A.G., Van Oijen, M., 2009. Modelling the dynamics of snow cover, soil frost and surface ice in Norwegian grasslands. *In press.*

Thorsen, S.M., Höglind, M., 2009. Modelling frost tolerance in timothy; Bayesian calibration and sensitivity analysis. In prep.

Walther, A., Linderholm, H.W., 2006. A comparison of growing season indices for the Greater Baltic Area. International Journal of Biometeorology 51, 107-118