Master Thesis

Forcing and evaluation of numerical snowpack models with snow stratigraphy measurements

Léo Viallon-Galinier

Météo-France/CNRS, CNRM-GAME UMR3589
CEN (Centre d'Étude de la Neige)
Grenoble, France

March – July 2017

Author :
Léo Viallon-Galinier
École polytechnique
Promotion X2014
3rd-year student

Supervisors :
Pascal Hagenmuller
Matthieu Lafaysse

Abstract

[English] Snowpack detailed models, such as the Crocus model used in this study, do not assimilate snowpack observations to adjust simulations, as commonly done for atmosphere models. Furthermore, the models are usually evaluated on bulk or surface variables whereas they are mainly used to represent the detailed stratigraphy relevant for avalanche hazard forecasting.

However, snowpack stratigraphic measurements are performed regularly by observer networks. This work proposes methods to evaluate model results and to reinitialize the model during the winter in order to reduce the accumulation of errors through the snow season. For this purpose, the observed data are first converted into model state variables. Then, a matching algorithm is performed between simulation and observation profiles so that it allows to reinitialize some model state variables from observation data during the winter despite the discrepancies in the layering between observations and simulations. Finally, a way of evaluating error of simulation regarding to manual observations is developed to be able to evaluate model and reinitialization skill.

When model is sufficiently far from the observation, the reinitialization is shown to successfully reduce the simulation errors. The method is fully adjustable : even partial observations could be used with partial reinitializations. The results obtained encourage to keep going with more advanced assimilation data algorithms. This work is also useful to evaluate the models in a more comprehensive way, checking the whole stratigraphy : it highlighted model biases and inconsistencies between some simulated variables and observed ones.

[Français] Les modèles détaillés de manteau neigeux, comme le modèle Crocus utilisé dans cette étude, n'utilisent pas d'observations nivologiques pour corriger les simulations, comme cela est courant dans les modèles atmosphériques. De plus, ces modèles sont généralement évalués à partir de variables intégrées sur la hauteur ou de variables de surface alors qu'ils sont utilisés parce qu'ils représentent le détail de la stratigraphie, notamment en ce qui concerne l'évaluation du risque d'avalanche.

Pourtant, des observations du manteau neigeux sont réalisées régulièement par un réseau d'observateurs. Ce travail propose une méthode pour évaluer les résultats de simulation et réinitialiser cette dernière durant l'hiver afin de réduire l'accumulation des erreurs tout au long de la saison. Pour cela, les données d'observation sont d'abord traitées pour les traduire en variables d'état du modèle. Ensuite, un algorithme d'appariement entre simulation et observation est mis en place afin d'utiliser simultanément certaines variables du modèle et des données d'observation de manière cohérente malgré les possibles incohérences de découpage des couches entre observation et simulation. Enfin, une méthode d'évaluation est développée pour évaluer le modèle et l'effet des réinitialisations.

Lorsque l'erreur du modèle vis à vis de l'observation est suffisante, les réinitialisations permettent d'améliorer significativement les résultats de la simulation. La méthode est totalement flexible : même des observations partielles peuvent être utilisées pour réinitialiser partiellement la simulation. Les résultats obtenus encouragent à poursuivre avec des algorithmes d'assimilation plus complexes. Ce travail est également utile pour l'évaluation des résultats du modèle en prenant en compte l'ensemble de la stratigraphie : il peut permettre de repérer des biais du modèle ou des incohérences entre variables mesurées et variables simulées.

Acknowledgements

Je souhaite remercier en tout premier lieu mes tuteurs, Pascal Hagenmuller et Matthieu Lafaysse pour leur accueil chaleureux et leur encadrement tout au long du stage : ils ont réussi à me donner accès, en quelques mois, à un domaine sur lequel je n'avais a priori pas beaucoup de connaissances. En particulier merci à Pascal pour les explications sur le comportement du matériau neige, sa structure et son observation et à Matthieu pour la découverte et la prise en main du modèle numérique de neige Crocus. Je remercie également Jean-Louis Dumas et Delphine Charlieu pour m'avoir fait découvrir le centre et pour tous les détails administratifs, toujours dans la bonne humeur. Merci à Marie Dumont, pour l'intégration dans son équipe et à Samuel Morin, chef de centre, pour m'avoir accueilli et m'avoir permis de bien comprendre le fonctionnement et l'organisation du Centre d'Étude de la Neige et ses relations avec les autres composantes de Météo-France, en espérant qu'il acceptera encore de m'accueillir à nouveau.

Du côté de l'École polytechnique, je remercie Hervé Le Treut d'avoir accepté ce stage sous sa responsabilité, me permettant à la fois de découvrir un matériau passionnant à étudier : la neige et un possible débouché de mon cursus d'Ingénieur des Ponts, Eaux et Forêts avec le Centre National de Recherches Météorologiques.

Merci à Philippe Lapalus, François Tuzet, Jacques Roulle et Yves Lejeune pour m'avoir permis d'aller sur le terrain au Col de Porte et au Col du Lautaret et pour m'avoir expliqué les mesures nivologiques sur lesquelles j'ai été amené à travailler. Merci également à Daniel Goetz d'avoir pris du temps pour me présenter le réseau nivo-météo d'observation dont sont issues les observations utilisées dans cette étude et avoir guidé le choix des sites.

Quoi qu'ils en disent ou en pensent, Eveyne Pougatch et Yannick Danielou de l'équipe informatique permettent (entre autres) d'avoir accès à un environnement de travail efficace et aux bases de données d'observation, sans eux la science avancerait sans doute moins vite.

Merci également à Cécile Coléou, de la cellule montagne et nivologie, pour les discussions sur la prévision du risque d'avalanche et les missions de Météo-France sur le climat et l'information sur les risques en montagne.

Je n'oublierai certainement pas Louis Queno et Coline Bouchayer, mes collègues de bureau pour l'ambiance à la fois studieuse et agréable, entre bonne humeur et collaborations fructueuses.

Enfin, un grand merci à tout le personnel des Centre d'Étude de la Neige, de la Cellule Montagne et Nivologie et du Centre Météorologique des Alpes du Nord pour leur accueil, leur bonne humeur et les cafés, thés ou repas partagés.

Contents

1 Introduction and context

Internship context

This report is the result of my 3rd-year internship in the Météo-France snow research center from 27 March to 28 July 2017. I was immersed myself in the research group in charge of the study of snowpack structure and stability from micrometric to mountain ranges scale, the development of numerical models mainly for avalanche hazard forecasting but also for other purposes (hydrology, climate-snow interactions, etc.). My project focused on the forcing of numerical snowpack model with snow stratigraphy measurements. I first discovered snow issues, snowpack modeling with Crocus model and snow pit observations with some outings in the field. Forcing simulation with observations also bring me to an evaluation of these snowpack models to be able to evaluate the method, but this evaluation could have a larger impact.

This report is based on an article I wrote (with helpful advice and reviews of my supervisors), to be submitted in *Cold Regions Science and Technology* in coming months.

Scientific context

The understanding of snow cover evolution is a key element for a lot of issues, at different scales. At a large scale, it is a critical point for understanding climate change: snow, with the whole cryosphere being a huge water and energy storage at the Earth scale. But at a smaller scale, it is also interesting to have a detailed knowledge of snowpack state, including its detailed layering. It is a main issue both for science, economy and civil safety. The knowledge of snowpack state is important to estimate water storage, critical for estimating flood hazard or for regulating lakes level and monitoring hydroelectricity production. From a scientific point of view, the snowpack is an important factor in the mountain meteorology. But it has also an economic importance : ski resorts are interested in a better understanding and predictability of snow evolution. It has a special interest for civil safety, with the estimation of avalanche hazard which requires an accurate representation of snowpack stratigraphy, in other words snowpack layering and characteristics of each layer, allowing to compute a local snowpack stability : small details such as thin refrozen layers could have a great importance in the whole snowpack stability.

Knowing accurately the snowpack stratigraphy is a real challenge because of great spatial variability, even at local scale, or its fast evolution, governed by interrelated complex processes. Snowpack evolve with precipitation of snow but also with temperature gradient metamorphism. Presence of liquid water have a significant impact on snow evolution and stability and could appear in case of melting or rain fall. Mechanical effects such as wind effect or settlement also have a great importance in snow stability and evolution.

To help monitoring the snowpack evolutions, it is possible to rely on observations or on numerical modelling. National institutions, like Météo-France or WSL/SLF (Swiss equivalent) have networks of observers, recording weekly (more or less) snowpack state by realising snow pits. But observations provides only local and discontinuous informations, so that the representation of detailed snowpack with models is very useful for avalanche hazard forecasting. Several research teams have also developed numerical models. For instance, Swiss researchers developed a model called Snowpack [Bartelt and Lehning, 2002] while was developed the model Sntherm in the United States [Jordan, 1991]. In France, Meteo-France also developed a model called Crocus [Vionnet et al., 2012]. These three models aim to represent the detailed stratigraphy of the snowpack, by the description of a set of layers with different physical properties. These models were first developed to estimate the mechanical stability of the snowpack and further the avalanche hazard, which requires the representation of a detailed layering, discriminating weak layers from stable layers.

Use of observations for snowpack modelling

Models presented below were developed to represent the detailed stratigraphy of the snowpack. However, model evaluations were mainly on total snow depth, snow water equivalent or surface albedo, that is to say surface or bulk variables, and not on the detailed stratigraphy which made its interest. these evaluations are interesting, and sufficient for many applications (e.g. Boone and Etchevers, 2001; Essery et al., 1999, 2013), but the avalanche hazard forecasting application requires more advanced evaluations. For Crocus, the evaluation has until now mainly focused on total snow height [Brun et al., 1989], sometimes with surface temperature [Brun et al., 2012], snow water equivalent, bulk density or albedo [Lafaysse et al., 2017] but very little on its detailed stratigraphy. Some studies, like [Morin et al., 2013; Domine et al., 2013] or [Carmagnola et al., 2014] added recently some subjective evaluations on other parameters such as SSA (Specific Surface Area) and density profiles but there is not a proper quantification of the model skills about the details of stratigraphy, and it remains largely incomplete. For the Swiss Snowpack model, Lehning et al., 2001 proposes a deeper evaluation, but this method has been limited to SNOWPACK. Furthermore, this method presents the interest of a mapping between observation and simulation layers to be able to discriminate error from layering and error from intensive variables. But this mapping allows layers inversions, which is not fully satisfactory. No other studies were performed on this topic.

An other issue is that nowadays, snowpack simulations are started during summer, on a bare ground. Then a meteorological forcing (from measurements or from a model, which also could assimilate meteorological measurements) is provided to simulate snowpack evolution until next summer. As a result errors from snowpack model [Essery et al., 2013; Lafaysse et al., 2017] and from weather forcing [Raleigh et al., 2015] propagate throughout the winter, which could lead to significant error especially at the end of the season. Crocus uses only meteorological observations whereas SNOWPACK uses observed snow depth data to drive the model instead of precipitation amount from meteorological forcing. But neither of the two use detailed snowpack observations to reinitialize the model during the season.

So, there is two way of monitoring snow evolution: manual observations in a punctual network or snowpack models. These two approaches are complementary but models do not use extensively data from observations. However, it is a huge dataset to evaluate and improve simulations, and this is a common practice for atmospheric models [Seity et al., 2011]. Data assimilation of snowpack observations into the snowpack models has recently been investigated. Charrois et al., 2016 shows the interest of assimilating surface optical reflectance (spectral albedo) into Crocus model, but this study was carried out on synthetic data, because spectral albedo measurements from satellites, spectral albedo at ground level and spectral albedo from simulations are not already well linked together. Magnusson et al., 2014 also assimilates snow water equivalent in a less detailed model. , but a lot of information from manual snow pit measurements have never been used, whereas they contain a lot of informations on the snowpack state at the observation point.

Goals of the study

The starting point of the study was to investigate the interest of using such manual snow pit observations into the snowpack model. On a first way on using observation in snowpack modelling, we propose a new method to totally or partially reinitialize and evaluate a snowpack model by these observations. In order to reinitialize with informations from both observation and simulation, a matching method is needed to map simulation and observation snow layers. The reconstruction of model state variables from observations is also presented.

To evaluate our actions we also propose a method to evaluate a distance between simulation and observations. Simulation adequation to observation is quantitatively computed for a set of variables or more globally to produce concise indicators of model error. This method is generic, and could be used more largely to evaluate snowpack models regarding to observations.

In this study, these methods were applied to the Crocus model, on a 15-season (2000 to 2015) and 3-points (in French Alps) set. We demonstrate how the use of observations could improve simulation, reducing the accumulation of errors throughout the winter.

In Section 2 the observation dataset available is presented, as well as the model used (SURFEX/ISBA/Crocus). Then, we explain in detail our method for reinitializing simulation and computing a distance between observation and simulation. The obtained results are presented in Section 3 and discussed with its prospects in Section 4.

2 Material and Methods

2.1 Measurements of snow profiles

Manual snowpack measurements and observations by trained observers is a key information to produce the avalanche bulletin. Therefore, avalanche warning services such as the one supported by Meteo-France or the WSL-SLF maintain extensive networks of manual snowpack punctual observations ([Pahaut and Giraud, 1995]). Among these observations, snow profiles, that is to say the snow characteristics determined layer by layer, is reported, in a standardized manner, on a weekly to monthly basis. These stratigraphic data are used as input of the presented re-initialization method.

2.1.1 Study plots and period

In the French mountains, the observer network supported by Meteo-France and its partners in ski resorts is composed of about 120 study plots with altitudes in the range between 1100 m and 3000 m (Figure 1, for the French Alps). Around 1300 measured snow profiles are reported each year. For this study, three sites were selected based on their high rate of reporting and low exposure to wind: Col de Porte (Chartreuse, 1325 m, slope 5°, aspect N), Tignes (Haute-Tarentaise, 2400 m, 5°, E) and La Plagne (Vanoise, 2160 m, 5° , NE). They are identified by yellow dots on Figure 1. The study period covers the winter seasons 2000-2001 to 2014-2015. In this time period and for the selected sites, 709 complete standard snow profiles are reported, that is to say about one profile each week during winter. Note that the Col de Porte site is also a long-term research study plot where additional measurements are available [Morin et al., 2012].

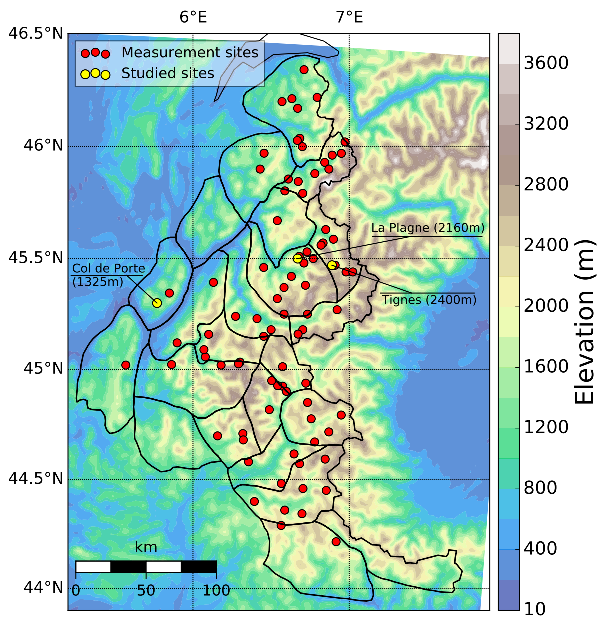

Figure 1: Map of the observer network of Meteo-France in the French alps. The three points used for this study are highlighted in yellow.

2.1.2 Measured snow layer variables

First, a measurement of ram resistance is performed with a manual penetrometer (ramsonde, data not used in this study). Then, after excavating the snowpack down to the ground, the observer partitions the snowpack into individual homogeneous layers. Each layer is characterized in terms of vertical position, grain type and size, density, hand hardness and humidity according to a standard procedure [Fierz et al., 2009]. an example of standard observation is shown on Figure 3a. Grain type and size are determined by visual inspection with a 10x lens. Density is measured by weighing a defined snow volume extracted using a cylinder-type cutter (6 cm diameter). Density measurements are usually limited to snow layers thicker than the cutter diameter. Hand hardness is evaluated by pushing the fist (F), four fingers (4F), one finger (1F), a pen (P) or a knife (K) into the snowpack. The hand hardness then corresponds to the biggest element the can be inserted while not exceeding a force of about 10 N. Humidity is classified into five classes: dry, moist, wet, very wet and soaked according to the snow behaviour if pressed in the glove. In addition to this layer-by-layer observations, a temperature profile is also measured on an independent vertical grid adjusted by the observer according to temperature gradient.

Except for density whose measurement uncertainties are limited [Proksch et al., 2016] and temperature, these measurements inevitably contain some operator subjectivity. This uncertainty is increased because observers vary in time and space. Moreover, at the study plot scale (10 m), little spatial variability is unavoidable [Harper and Bradford, 2003], even if the sites were selected based on their low exposure to wind.

For the specific site of Col de Porte, other data are available. In this study, the time series of albedo is used, obtained from measurements of incoming and upcoming short-wave radiations [Morin et al., 2012]. Besides, Carmagnola, 2013, sect. 3.2.2 conducted additional measurements at Col de Porte during winter 2011-2012 : vertical profiles of specific surface area were measured with the Alpine Snowpack Specific Surface Area Profiler (ASSSAP, light version of POSSUM [Arnaud et al., 2011]) and penetration resistance was measured weekly with the SnowMicroPenetrometer (SMP, [Schneebeli et al., 1999]).

2.2 Snowpack simulation : model SAFRAN-SURFEX/ISBA/Crocus

The snowpack evolution was simulated with using the modeling chain SAFRAN-SURFEX/ISBA/Crocus, presented in Durand et al., 1999; Lafaysse et al., 2013, on the selected sites.

2.2.1 Atmospheric forcing

The SAFRAN analysis system (Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie, [Durand et al., 2009]) provides the atmospheric forcing driving the snowpack evolution. Provided variables are wind speed, air temperature, humidity, atmospheric pressure, incoming short-wave and long-wave radiations and precipitation (amount and phase), on a hourly basis. This system adjusts a guess from numerical weather prediction model at 40 km grid spacing with weather data available from observation sources (but no snow observations are included). On the French Alps, the analysis is performed on 23 areas (the so-called massif) which present a certain climatological homogeneity, and on 300 m elevation band or directly at station elevation in the case of punctual simulations such as in this study. To be more accurate on incoming radiations, masks adapted to local topography are taken into account.

2.2.2 Snowpack model

The model SURFEX/ISBA/Crocus (called Crocus hereafter) is a one-dimensional multilayer physical snowpack model [Brun et al., 1989; Vionnet et al., 2012]. Crocus represents the snowpack as a set of up to 50 snow layers. Each layer is characterized by its density, age, enthalpy, snow water equivalent and two variables representing the snow grain morphology: sphericity and specific surface area (SSA). An additional state variable, called historic variable is included to indicate whether there once was liquid water or faceted crystals in the layer (see Table 9). SSA represents the total surface area per unit of mass. Sphericity varies between 0 and 1 and describes the ratio of rounded versus angular shapes. Crocus reproduces the time-evolution of these state variables by accounting for new snow deposition, metamorphism, settlement, heat exchanges, melting and refreezing for each layer, at a time step of 15 minutes, as summarized with Figure 2. Crocus is a unidimensional model, along the vertical axis, so that the representation at a mountain scale is a set of reinitialization points, with different altitudes, slopes, orientations, and sometimes location but no interactions between points.

Crocus is coupled to the soil scheme ISBA [Decharme et al., 2011] in the SURFEX platform [Moigne, 2012; Masson et al., 2013]. Numerous physical options are now available. In this study, we used the default configuration as defined in [Lafaysse et al., 2017]. This version includes the C13 formalisation of snow metamorphism developed by Carmagnola, 2013, representing grain by SSA and sphericity, instead of grain size, sphericity and dendricity. An example of simulation result presenting all state variable is presented in Figure 3b.

2.3 Forcing simulation initial conditions with measured profiles and evaluation

In this section, the method to re-initialize the simulation of the time-evolution of the snowpack stratigraphy with measured snow profiles is presented. As shown previously (Figure 3), the partition of the snowpack into layers and their characterization generally differ between observations and simulations. Therefore a method to convert the observation into model state variables is required (see Section 2.3.1). Furthermore, to combine simulated and

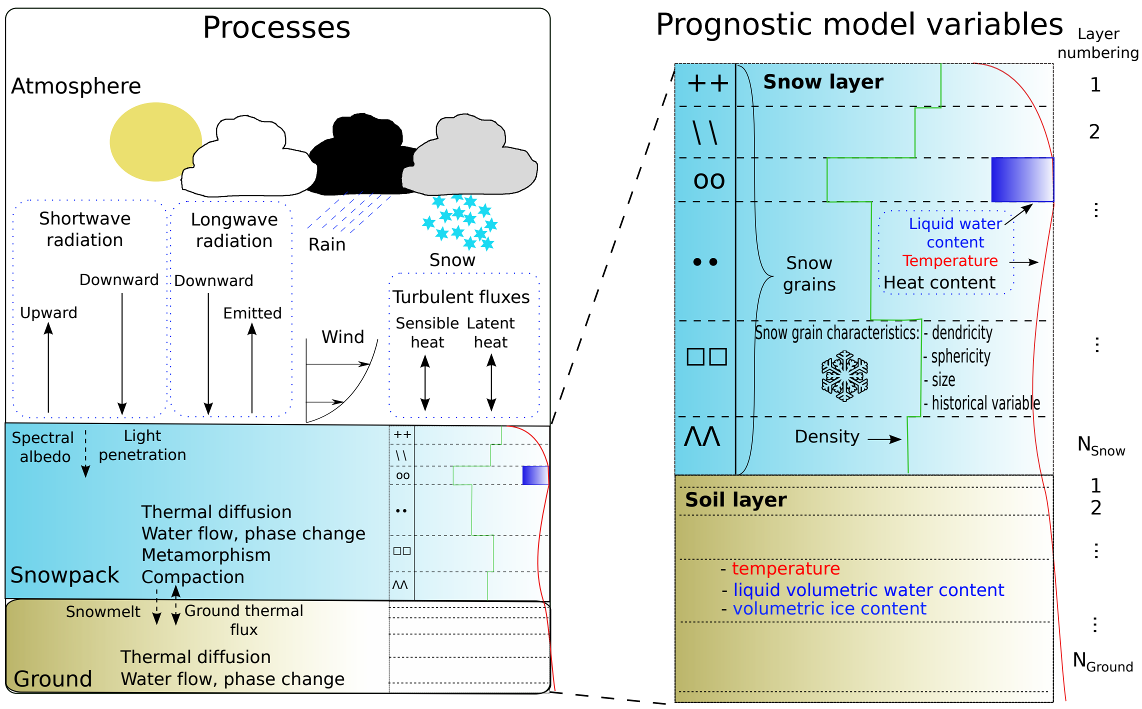

Figure 2: Crocus detalied snowpack numeric model : main variables and physical processes taken into acount. Source : [Vionnet et al., 2012]

observed variables, the simulated layers need to be associated to observed layers, which may bet at different vertical position. This association requires both a metric (or distance) between layers (Section 2.3.2) and a matching algorithm (Section 2.3.3).

2.3.1 Conversion of observed variables to simulation state variables

The simulated layers are described in terms of snow water equivalent, density, SSA, historic variable, sphericity, enthalpy and age (Figure 4). The measured layers are described in terms of depth, grain type and size, density, humidity and hardness. An independent temperature profile is also recorded (Figure 4). Only complete observations on grain type and humidity class are used, but other variables, could be incomplete. Crocus can derive some of the observed variables from the simulation state variables [Vionnet et al., 2012] but the implemented relations are not all injective nor surjective. Therefore, to recover the simulation state variables from observations, additional relations need to be introduced (Figure 4):

Density We use the manually recorded density as recorded by the observer, considering density is uniform on the layer thickness. Too thin layers (thickness under cutter diameter) could not always be measured and the density of some thicker layers may be missing. For these layers, typical density values are derived from the grain type (see Appendix A).

Enthalpy Enthalpy is a function of temperature T and density ρ in case of dry layer ($H = (Tc - L_{fus})\rho$) and liquid water content LWC in case of wet layer ($H = L_{fus}(LWC - \rho)$). L_{fus} is mass enthalpy of fusion for water under atmospheric pressure (333700 $\mathrm{J\,kg^{-1}}$) and c is the mass heat capacity of water (2106 $\mathrm{J\,kg^{-1}\,K^{-1}}$). The liquid water content (LWC in $\mathrm{kg\,m^{-3}}$) is assessed from measured humidity class as following: as 0, 2.5 and 5 % of pore volume for dry, moist and wet/very wet/soaked classes, respectively. Note that 5 % of pore volume corresponds to the liquid water content saturation in Crocus.

Grain morphology variables Sphericity, historic variable and SSA represent all together grain morphology. They could be treated independently but all are reinitialized using table from grain type presented in Appendix A.

Age Age could also be inferred approximately from grain type, using a very low value (1 day) for precipitation particles, a mid-value for decomposing and fragmented snow, depending on altitude (between 2 days at low

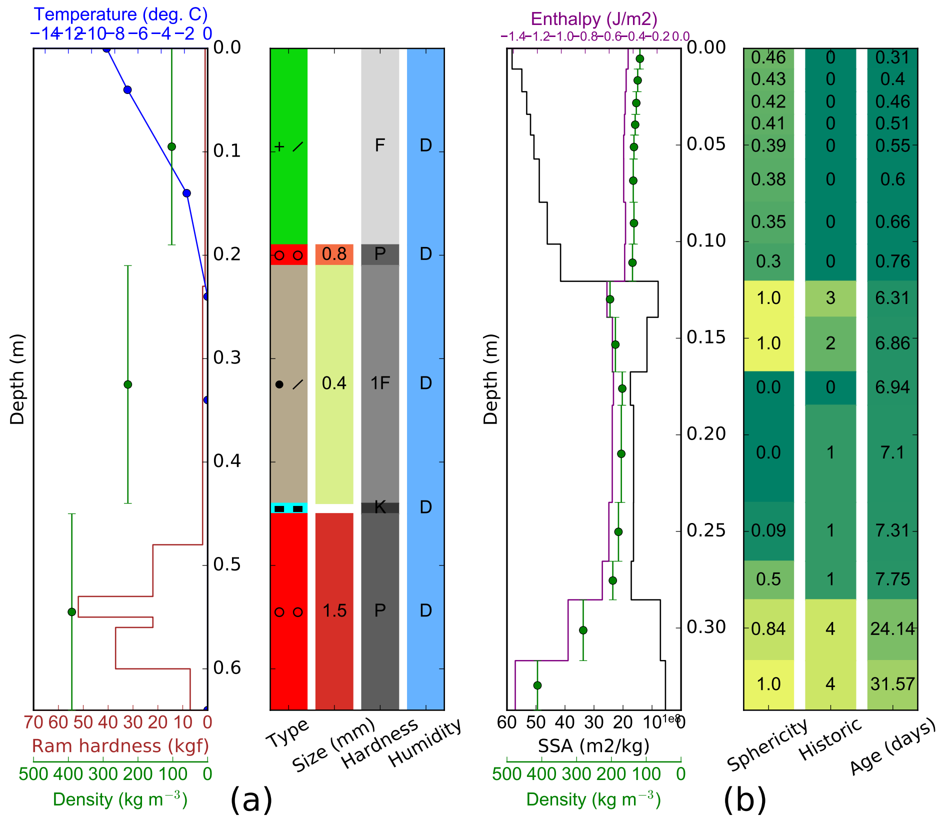

Figure 3: Typical (a) observation record and (b) simulation output. Example of December 29 at Col de Porte. Representation of grain type, hand harness and humidity class correspond to the international classification [Fierz et al., 2009].

altitude, to 6 days at high altitude), a mean value for mixed precipitation particles and decomposing snow and a high value (20 days) for all other grain types.

2.3.2 Layer metric

In order to evaluate agreement between simulation and observation, a distance between two layers is required (layers from observation or simulation). Several metrics are considered, for each layer or for each point:

- The bias (difference between measured value and Crocus value)
- The absolute error which is the absolute bias
- Bias could be renormalized by the third quartile of the values encountered on seasons 2000 to 2015
- Absolute error could also be renormalized in the same way so that all renormalized error could be compared together.

The use of these metrics will be detailed further. These metrics could be computed for total snow height, density, temperature and SSA (computed as described below for observation). But this is not applicable to grain type or liquid water content because it implies discrete values, an other way to compute errors is defined for these variables:

Liquid water content: Three liquid water classes (LWC) are defined (dry, moist, saturated), associated with arbitrary values of 0, 0.5 and 1 so that bias and absolute error could be calculated respectively as the difference and the absolute difference. For observation, classes are defined according to [Fierz et al., 2009], that is to say dry and moist are two international class and saturated correspond to higher liquid water content. For the model, snow is considered as dry if liquid water content is less than 1 % of the model maximum and saturated when liquid water content is more than 75 % of the maximum.

Grain type: To be able to determine a distance between two grain type we use the Table 1 of Lehning et al., 2001 which is a table of scores between grain types. Contrarily to him, we need a distance (0 value for a perfect

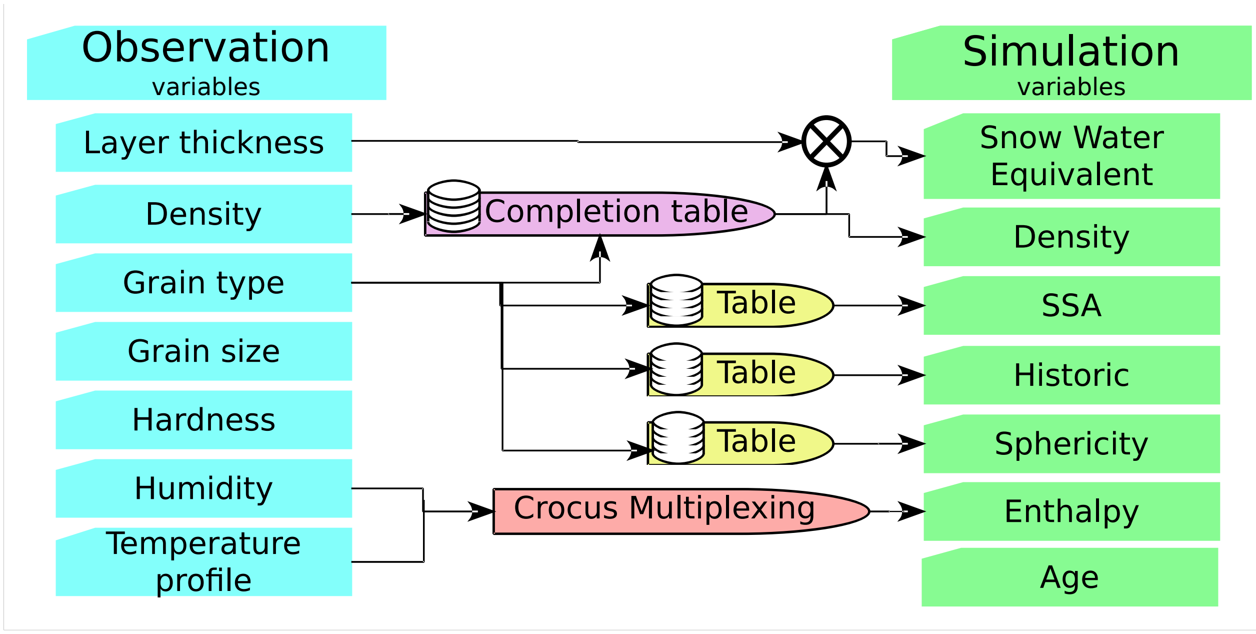

Figure 4: Diagram of the method to reinitialize model state variables with manual observation data

agreement), so we modify values to this end and add ice forms. The table used is presented in Appendix A. Then we calculate agreement between principal grain d_1, secondary grains d_2 and cross-distances d_{12} and d_{21}. As the Crocus model do not have the concept of main grain and secondary one, we do not associate a different weight to cross-distances and direct distances and chose the absolute error $0.5 \cdot \min(d_1 + d_2, d_{12} + d_{21})$. This allows not to give a higher weight to the ordering of grain type, in comparison to Lehning et al., 2001, because Crocus model have not the concept of main and secondary grain type. To understand well this distance, it is necessary to be aware that our model do not represent all grain types (can be described neither ice layers nor surface hoar or graupel). So, a minimal absolute error value higher than zero is intrinsically expected for this variable.

2.3.3 Vertical matching between measured and simulated profiles

Layers at the same depth in the observation and simulation are not necessarily at the same position in the stratigraphy. In order to combine measured and simulated profiles to re-set the snowpack state, observed and simulated profiles need to be matched.

The matching approach used in this study for snowpack profiles was proposed by [Hagenmuller and Pilloix, 2016; Schaller et al., 2016] to identify common features in sets of highly resolved measured snow profiles. It was adapted here to match a simulated profile to a measured profile. The procedure goes as following:

Variables standardization: Comparison requires standardisation between observation and simulation variables and complete profiles, so that a completion of density is done as presented in Section 2.3.1. SSA is computed from observation to be compared to model SSA, humidity classes of observation and liquid water content of simulation are converted into 3 classes (LWC) as presented in Section 2.3.2. Temperature and snow depth are already present on both sides.

Layer support : In case of error on total snow height, the thickness of the simulated layers is uniformly adjusted so that the simulated total depth corresponds to measured total depth. The measured and adjusted simulated profiles are then re-sampled on the same vertical layer grid (generally 1 mm thickness).

Dynamic Time Wrapping: Global stretching is not sufficient : local discrepancies could remain [Hagenmuller and Pilloix, 2016], therefore a matching is compulsory. Dynamic Time Warping (DTW) is a well-known technique to find an optimal alignment between two given discrete sequences and under certain conditions [Sakoe and Chiba, 1978]. Here the sequence corresponds to a profile, that is to say a two dimensional array with the first dimension being the index of the grid layer and the second dimension being the considered layer variables (density, SSA, grain type, humidity...). The general idea of DTW is to consecutively assign the layer of one sequence to one of the other sequence, whereby each layer can be matched with at maximum two layers of the other one. To find the best fit, a matrix D is calculated where D_{ij} is the lower error that leads to the i-th element of the first data set being connected to the j-th element of the second one. With these constraints, layer thickness are stretched up to +100 % or thinned up to -50 %. The best fit minimizes the total cost to link the top and bottom of the two considered profiles.

The cost matrix D is computed as the sum of the layer metric described in details in Section 2.3.2. It includes renormalized absolute error on density, LWC, grain type, temperature and SSA. Lighter weight are applied to SSA and temperature, because determination of SSA lacks of accuracy and temperature is more linked to depth rather than snow layer. Additional cost is added in order to avoid too important displacements : the difference of original depth multiplied by $10^{-1} c \cdot s$,s being the grid step and c the sum of coefficients.

An example of the use of this algorithm (correction of total snow height and matching) is presented between Figures 5b and c.

Modified simulation layers support: Thickness of simulation layers are adjusted according to the matching. It is now possible to average values from observation on numeric layers to reinitialize simulation as presented on Figure 5d.

An overview of the reinitialization method is presented on one example on Figure 5.

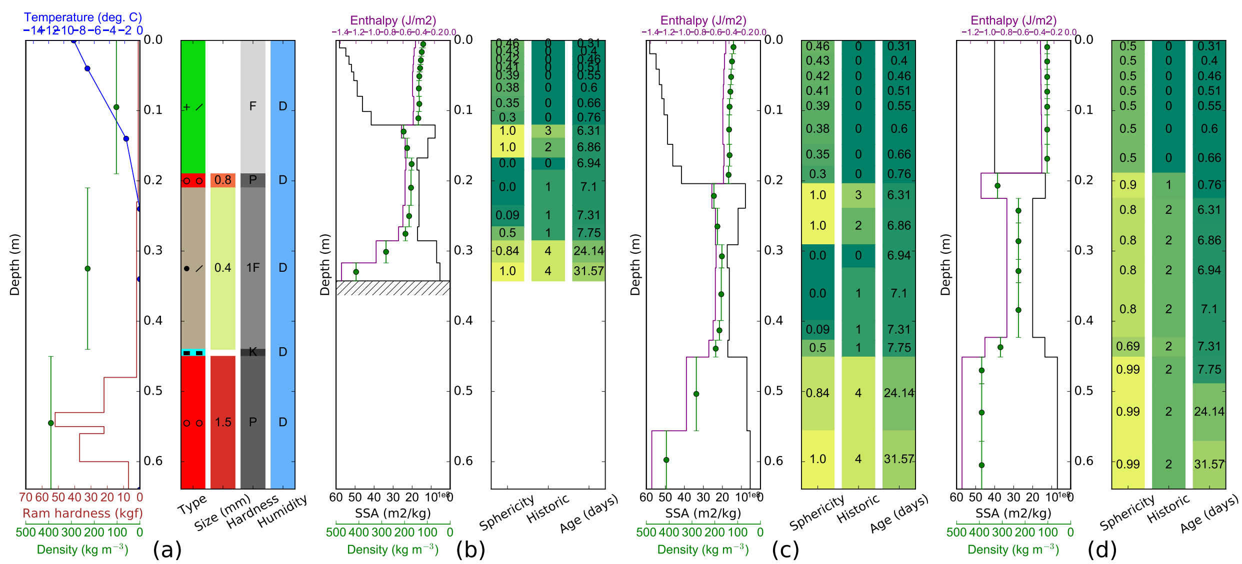

Figure 5: Overview of the reinitialization method on the example of 12th Dec. 2003 at Col de Porte. (a) : observation, (b) : reference simulation, (c) : matched reference simulation and (d) : simulation reinitialized with historic and sphericity from observed grain type, SSA from grain type, enthalpy from temperature and humidity and density from density.

2.3.4 Special case when model have no snow on the ground

If a sufficient amount of snow is present on the ground, the model already has a correct layering for numerical resolution so this basis is used as presented above. On the other hand, it could be necessary to redefine a layering when there is no convenient layering present (too few snow) in the model simulation.

Numeric snow layers To this end, we first use the stratigraphic layer bounds, then we cut the first layers so that they match the reference profile needed by Crocus [Vionnet et al., 2012, sect. 3.2]. Finally, we cut each central layer in equal parts so that they don't exceed the model maximum thickness.

Numeric ground layers If the model has no snow on the ground on the day of observation, it is also necessary to modify ground temperature. As temperature evolution is very slow, a temperature of 0 °C is set on the upper ground layer and we do not change ground temperature under 20 cm. Intermediate ground temperatures are linearly interpolated.

2.3.5 Selection of variables

The previously described methodology was applied with different set of variables in order to chose variables to be injected in the model for an optimal reinitialization, according to our norms.

Preliminary experiments reveals that, using the ASSSAP SSA, the grain type is incorrectly identified and the score of grain adequacy between reinitialized simulation, just after reinitialization collapse. Then, the same effect was compared between SMP density and manually recorded density. The mean scores observed are the same using SMP density or manual density. Being available only at Col de Porte, not every year and not giving a real improvement, it was decided not to use the SMP density and ASSSAP SSA.

It remains 3 groups of variables : density, grain type (sphericity, SSA, historic variable) and enthalpy (computed from temperature and liquid water class). Different combinations of these variables were tested as summarized in Table 1.

	Basic	Without		
	set	SSA	Density	Enthalpy
Snow water equivalent	X	X	X	X
Density	X	X		X
SSA	X		X	X
Historic	X	X	X	X
Sphericity	X	X	X	X
Enthalpy	X	X	X	
Age				

Table 1: Set of variables tested in reinitialized simulations to select the optimal set

2.3.6 Evaluation of agreement between observation and simulation

To be able to measure improvements, the method is quite similar : depth differences are corrected by the same matching, then, layers could be compared point-to point with layer metric described in details in Section 2.3.2. To produce more concise indicators, these indicators are averaged on the snowpack height.

These error indicators are computed immediately after reinitialization (3h), one week after (the next observation) and one month after (four observations later).

3 Results

Simulations were performed on our 3 test sites : Col de Porte, Tignes and La Plagne, on winters 2000-2001 to 2014-2015, reinitializing the model at each observation with methods presented previously.

3.1 A first example : Col de Porte on winter 2003-2004

Figure 6 show the season 2003-2004 at Col de Porte, simulated from bare soil in August 2003 and with reinitialization each week by observations showing achievement of redefining chosen state variable of the snowpack (here layers thickness, density, enthalpy, SSA, sphericity and historic variable).

On this season, the reinitialization leads to a large snow depth correction. This could be highlighted by computation of indicators such as mean absolute error on all observations for snow depth in reference (without reinitialization) which is about 28 cm and which is only 9 cm comparing observation with simulation reinitialized at all previous observation dates. This snowdepth error is due to snowfall in December as the first reinitialization have to double the amount of snow on the ground but also on late season melting or precipitations where reinitializations have to reduce the snowdepth. The same effect is visible on density with a mean reference absolute error of 68 $\text{kg}\,\text{m}^{-3}$ divided by two when previous observations are taken into account. It is also possible to notice a more accurate detection of deep hoar (correction of detection of deep hoar with Jan 8-th observation and around the Jan 28-th : the calculated absolute error decreases for grain type from 0.27 to 0.20) and a global diminution of liquid water content compared to the standalone model simulation. This reinitialization method implies to a discontinuous snowpack representation, at each observation date.

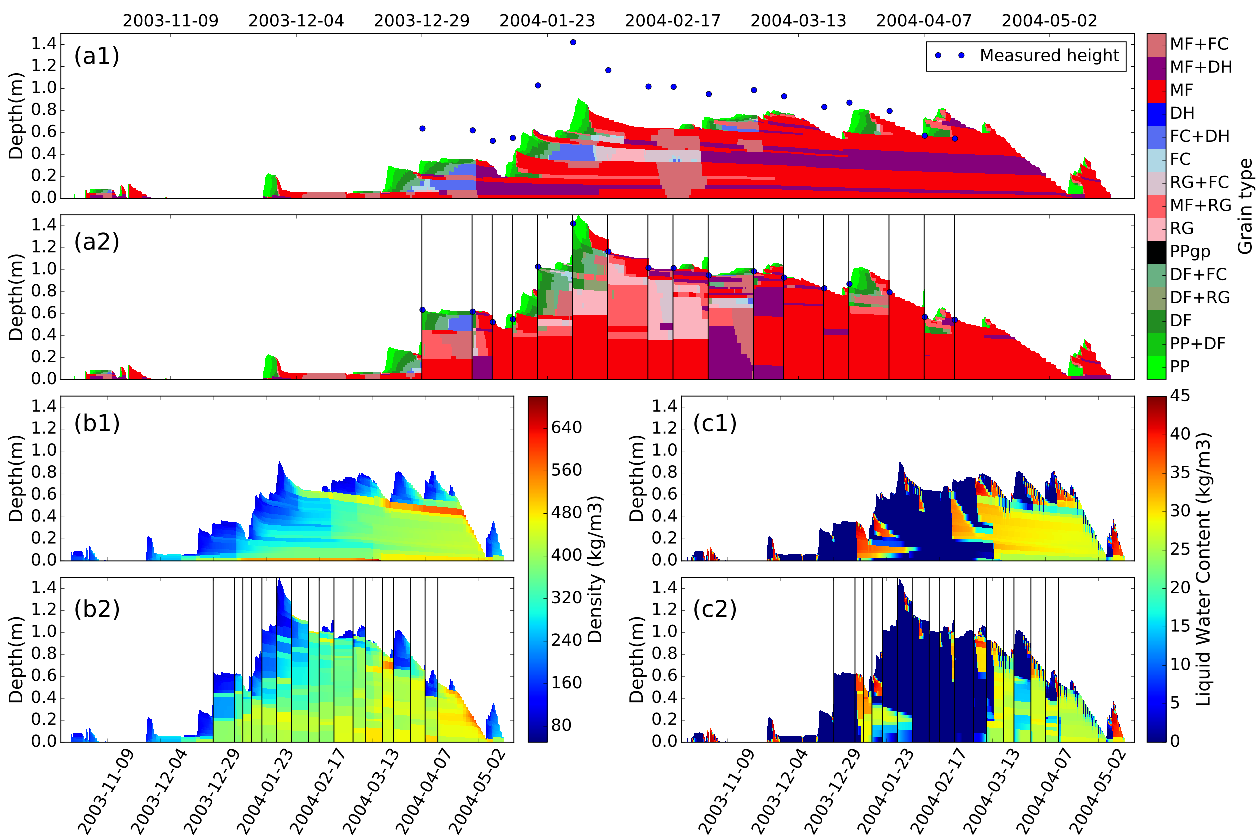

Figure 6: (1) Reference simulations (never reinitialized, start on August 1st on bare ground) and (2) simulation reinitialized at each observation (bottom). Black lines highlights observation dates. Example of snowpack evolution at col de Porte station on season 2003-2004 for variables (a) grain type, (b) density and (c) liquid water content. Blue dots on (a) reports measured snow height.

3.2 Evaluation on a long period (2000-2015)

3.2.1 Example of one station : Tignes

Once it is made possible to simulate a snowpack integrating partially observations, it is necessary to evaluate the result simulation, computing absolute error for each simulation at three time steps (just after the reinitialization, one week after and one month after). Figure 7 shows improvements obtained with reinitialization at the three time steps for one station on the whole studied period. Improvement is the difference between absolute error with reference and absolute error with reinitialized simulation (so that a positive value points out a positive effect of reinitialization whereas a negative value indicates a negative effect). At a first sight, it is possible to notice a large variability of absolute errors and improvements depending on the observation date. But there is also strong trends : the absolute error is always reduced immediately after, which means that simulations are correctly reinitialized. On the mid-term (one week), an improvement is globally observed but on some experiment the reference was better. On the long term (one month), when absolute errors are rather small on the reference, there is no mean improvement, but when absolute errors are important, especially on total height (seasons 2007-2008, 2013-2014 or 2014-2015), the improvement continues over one month.

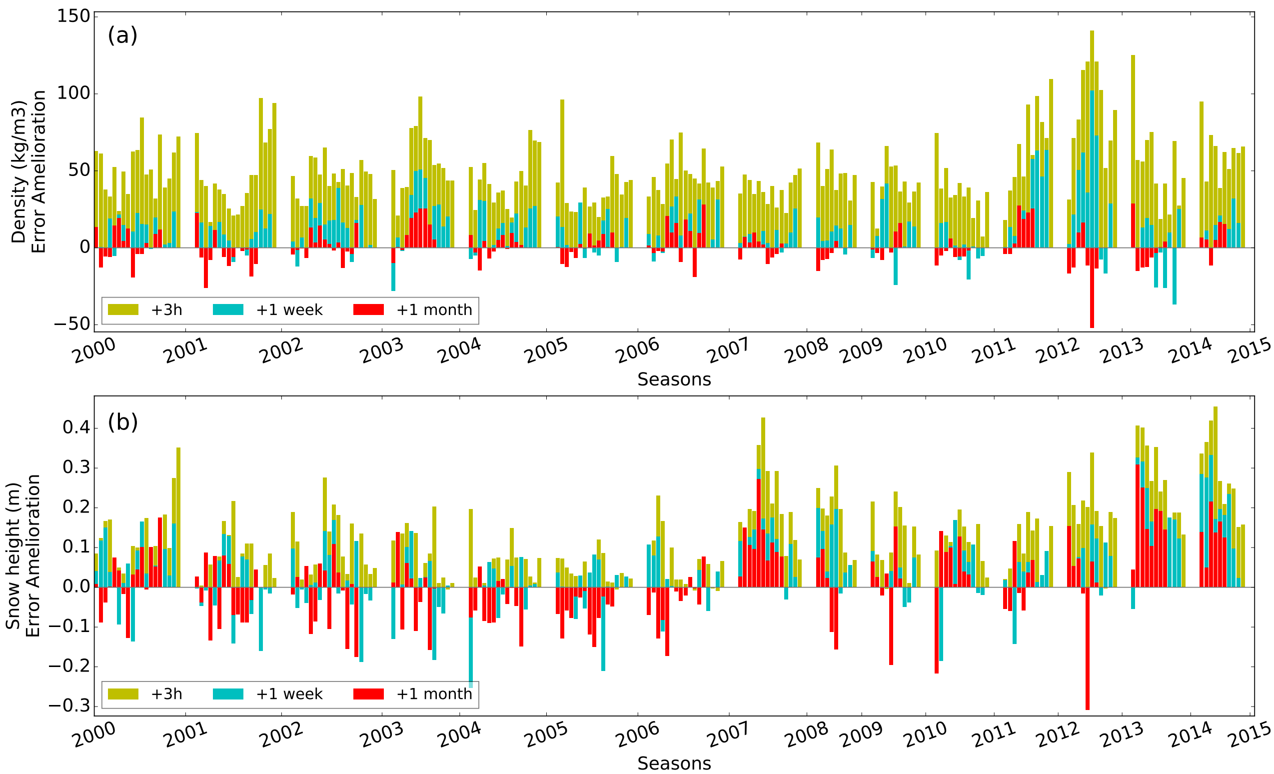

Figure 7: Improvement obtained after reinitialization : immediately after (3h, yellow), one week after (blue) and one month after (red). Only periods with available observations. Data from La Plagne station, winters from 2000-2001 to 2014-2015. Example of (a) density and (b) total snow height. Variables injected are layers thickness, density, enthalpy, SSA, sphericity and historic variable. Improvement is the difference between absolute error with reference and absolute error with reinitialized simulation (so that a positive value points out a positive effect of reinitialization whereas a negative value indicates a negative effect)

3.2.2 Generalisation to the complete dataset

The previous graphical representation gives a first visual evaluation but to compare more synthetically the results averages are computed : Figure 8 shows the mean absolute error and bias on our three stations on seasons 2000-2015, as well as the absolute error immediately after (3h), one week after and one month after. The snow depth correction 3h after is almost perfect whereas error on temperature is still 0.2 °C : it is easily explained because near-surface temperature is highly sensitive to air temperature, and temperature of surface layers could evolve significantly in

3 hours. The improvement on density is important but not perfect, because almost immediately the model bring this variables back to a mechanical equilibrium, not checked for reinitialization. SSA also evolve quickly, especially for high SSA, or under mechanical pressure. Considering these reservations, depth, density, temperature and SSA are shown to be correctly reinitialized. Liquid water content is also a phenomenon with fast evolution, this explains the importance of the 3 hours error of about 0.07. On the contrary, error on albedo do not evolve with reinitialization : it was expected because albedo mainly depends on snow age which is not reinitialized. Between these two extremes, grain type have an immediate error around 0.25. To understand this value, it have to be considered that 0.2 is the typical error between the two closest different grain types, so a value of 0.25 does not show a high error.

On the long-term trend, the error is always important regarding to error immediately after reinitialization, and error after one month is of the same magnitude as the reference error. This phenomenon is studied in detail in Section 3.4.

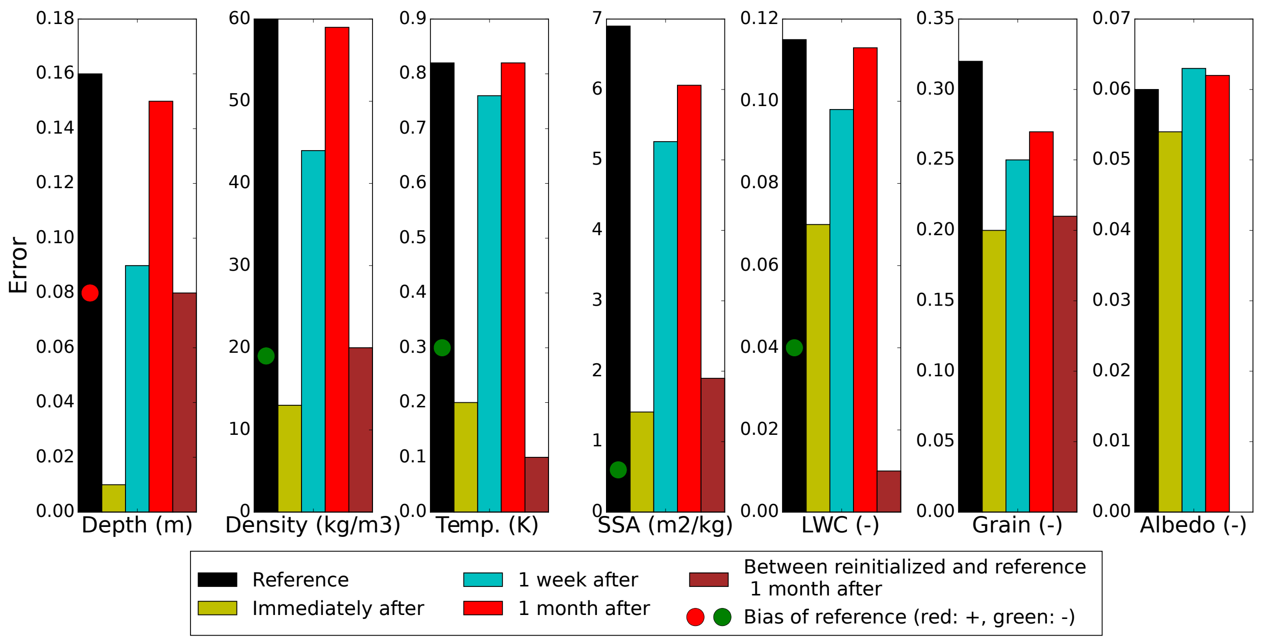

Figure 8: Absolute error between observations and reference (black) , and absolute error for each variable just after reinitializations (yellow), one week (blue) and one month (red) after, bias (dots, sign according to color : positive n green and negative in red) between observation and reference and absolute error between reference and simulation reinitialized one month before (brown), averaged on 3 stations and on 15 years (2000-2015) (except for Albedo, measured only at Col de Porte). Variables injected are layers thickness, density, enthalpy, SSA, sphericity and historic variable.

3.3 Sensitivity to variables selection

Figure 9 show mean absolute errors immediately after and one week after for a set of variable and for differents set of injected variables, as presented in Table 1. As including all of these 3 groups do not deteriorate the results (as shown by Figure 9) and taking off one of these variables only damage the scores. Therefore, the work was performed on a basis of reinitialization of density, sphericity, historic, SSA and enthalpy.

3.4 Time persistence and long-term evolution

If observations are available at a weekly time step over the three test stations, there are often more scarce in various stations of the network : it raises the problem of durability of the correction given by an observation. To visualise this evolution, improvements were plotted depending on time since the reinitialization on Figure 10. There is a large variability of improvement depending on the simulation. To be able to see the trend, a mean curve is plotted for simulations in each quartile of the absolute error before reinitialization. Figure 10 highlights the durability issue showing the decrease of improvement with time, and allows to quantify it : Table 2 gives the critical time to

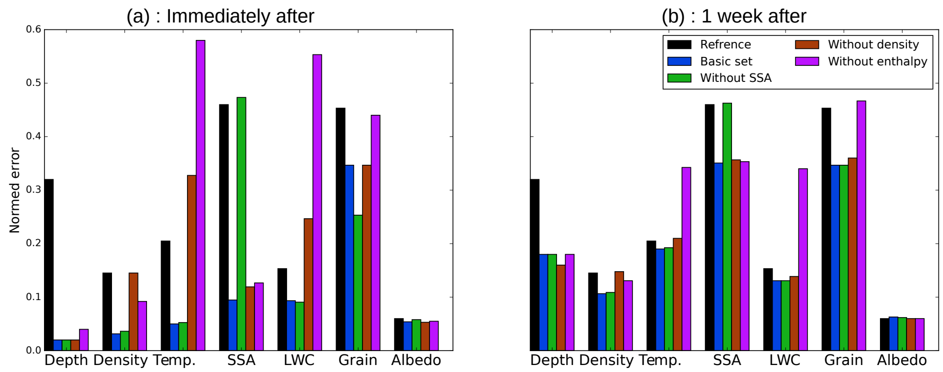

Figure 9: absolute error for a set of variable, (a) immediately after reinitialization and (b) one week after for different set of injected variables : in black, noting injected (reference), in blue our basic set chosen (density, sphericity, historic, SSA and enthalpy), in green our basic set without SSA, in red, our basic set without density and in purple our basic set without enthalpy. Data generated from averaging absolute errors on seasons 2000 to 2015 on our 3 experimental sites (except Albedo available only at Col de Porte).

lose three quarters of the achieved improvement just after reinitialization. This critical time depends on variable considered : temperature is largely influenced by air temperature and incident radiations so this information is lost quickly but snow height and SSA are still improved after 1 month. Figure 10 also allow to see that worse simulations are more improved than the better : the 4-th quartile mean improvement is still above the others and improvement lasts longer.

Variable	Time of persistence (25 % improvement)
Snow height	$>$ 1 month
Grain type	$>$ 1 month
Density	2 week
Temperature	1 day
SSA	2 week
LWC	1 week

Table 2: Time of persistence of each variable after a reinitialization, given on a weekly scale, and defined as persistence of 25 % of the improvement achieved just after reinitialization, computed from average improvement on our 3 stations on seasons 2000-2015 and on experiments which initial improvement is above the median.

When the model is already very efficient (always according to our norms), the precision of observations do not allow to improve the results. In Figure 11, shows the improvement as a function of the reference absolute error, for four example variables, at 3h-term, one week and one month time interval. Loss of improvement is still visible but the interesting point is that more important is the model absolute error before reinitialization, more the both short-term and long-term improvement are important whereas when reference is already not so bad (low reference absolute error) improvements are weak on the short-term and zero if not negative on the long-term. Linear trend lines are added to be able to have an idea of when it is interesting to reinitialize the model : when error to be expected at the next observation available is not worse than the reference one.

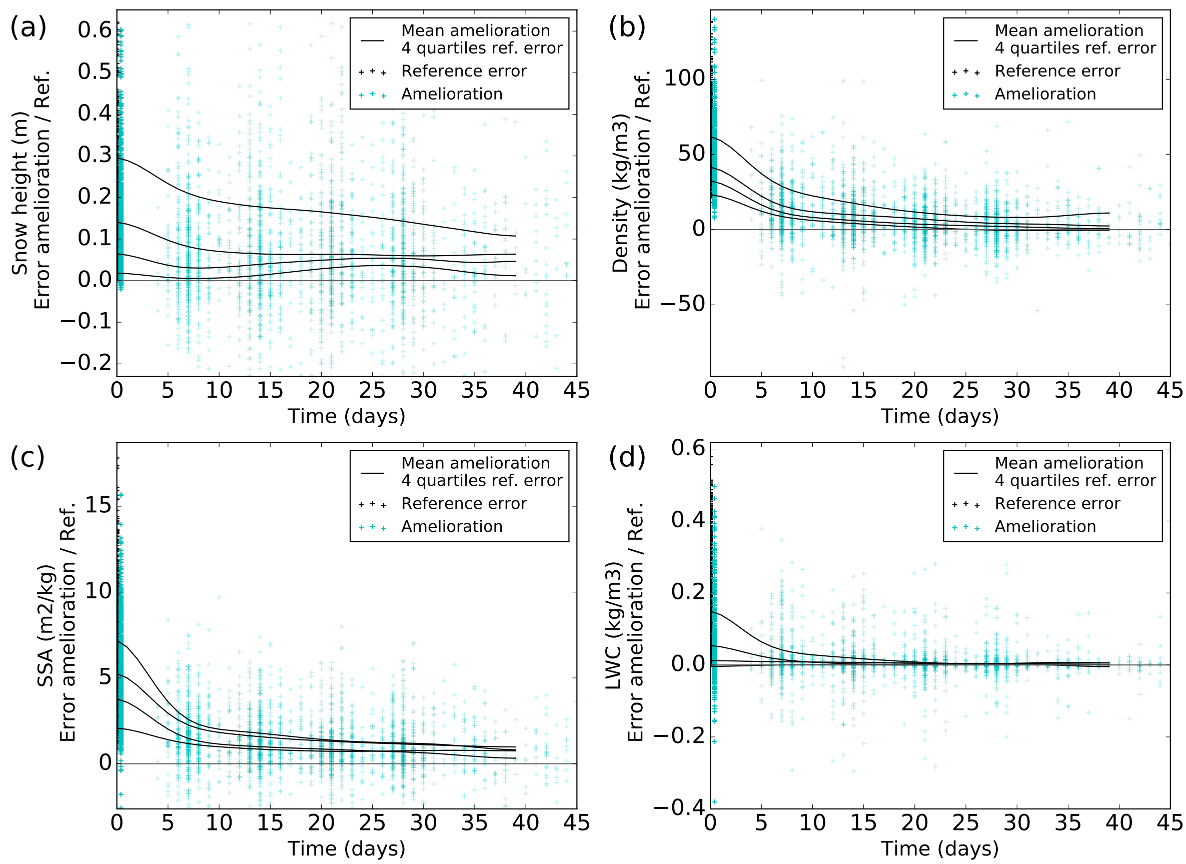

Figure 10: Reference absolute error and improvement of absolute error on the four next observations depending on time between reinitialization and evaluation observation. Example of some variables : (a) snow depth, (b) density, (c) SSA and (d) LWC. Solid black line are smoothed (Gaussian filter, $\sigma = 3$ days) mean improvement for the four quartiles of reference absolute error.

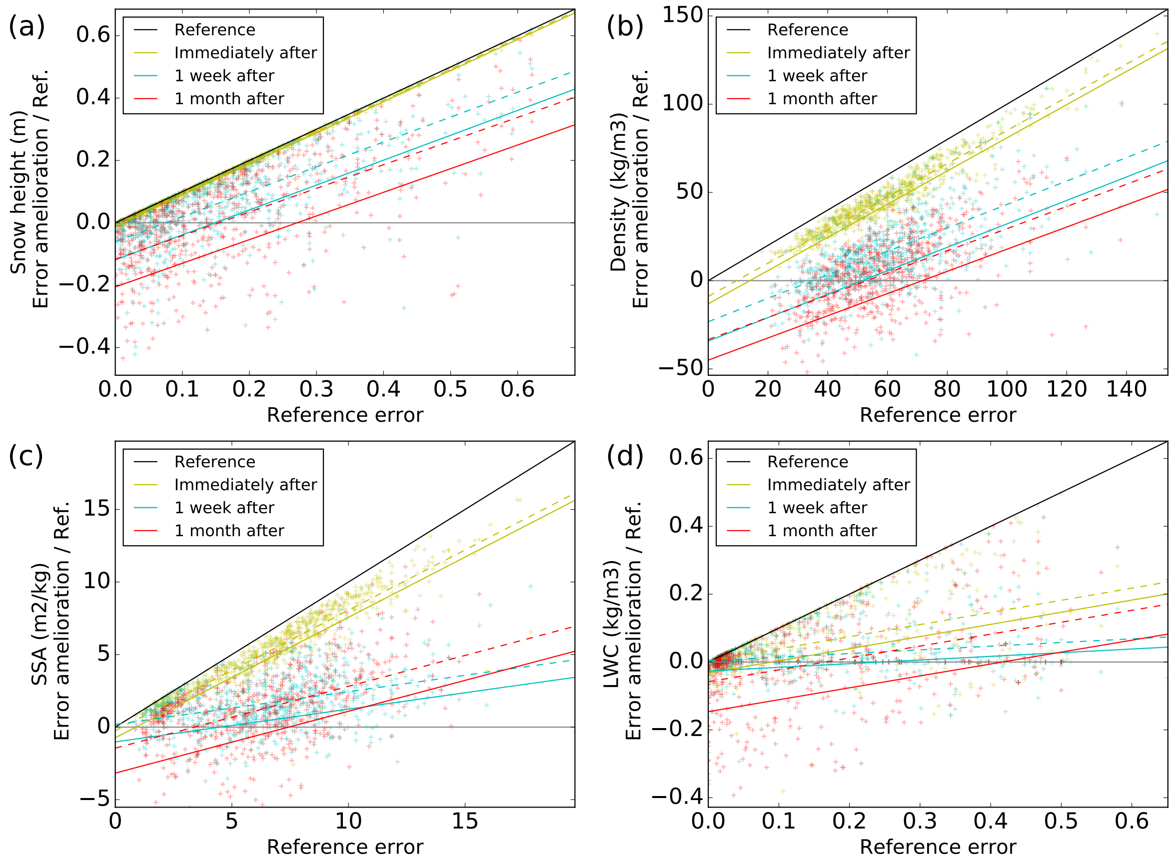

Figure 11: Absolute error improvement immediately after reinitialization (3h, green), one week after (blue) and one month (red) for all observations on season 2000 to 2015 on our 3 stations. Example of some variables : (a) snow depth, (b) density, (c) SSA and (d) LWC. Dashed line are linear regressions. 80 % of the experiences are above (better improvement) the solid line associated. Black solid line is 1:1.

4 Discussion

4.1 Encouraging results for a customisable method

Results presented in Section 3 highlights a significant improvement in using observations in the snowpack model Crocus, as shown in Figure 8. Error on density, temperature or snow depth is almost totally corrected after reinitialization, even though representation of grain type is not perfect and liquid water drifts quickly. Furthermore, all state variables except age were reintialized in our study because of the availability of accurate observations, but it is also possible to reinitialize only partially these variables. The same method could be applied to incomplete observations, choosing whether state variable to reinitialize or not depending on available information.

The model evaluation is also fully adjustable to available data : it could be done on any chosen variable depending on the goal of the evaluation and on available observation data. Reinitialization is also totally adjustable, variables reinitialized being chosen depending on data available, precision of measurements, etc.

Integration of observations in the operational model for avalanche hazard estimation could highly improve its efficiency, especially at the end of the season when the model has accumulated a lot of errors since its initialisation in previous August. Today, forecasters used to no longer look at the simulation results when important drifting was diagnosed previously in the season : they lack confidence in late-season simulations. Injecting observations could correct these main errors as shown in Section 3.4, improving results of avalanche hazard estimation, and give back confidence in the model to forecasters.

4.2 Sources of errors between simulation and observations

In Section 3.4 it was observed that some variables evolve quickly, and improvement decreases significantly on one month. Model could not be improved largely with this method on the long-term. Errors from snowpack measurements could be important and are not taken into account in this method because the observations are directly used as a reference for comparison and reinitialization. But there is also other sources of error, easier to study and correct.

4.2.1 Role of erroneous precipitation forcing

A first source of error is the meteorological forcing [Raleigh et al., 2015]. The impact of precipitation forcing was expected to be the main cause of error. In this case, it is expected that error was concentrated in the upper layers created since the reinitialization. To investigate this assumption, that the estimation of obtained absolute error only for the layers which age indicates there were present before the last reinitialization was computed. The obtained mean absolute error was quite similar to error on the whole snowpack indicating that error from new precipitation is not prevailing regarding to the others. To study more largely impact of the meteorological forcing (other variables and not only solid precipitation), the use of an ensemble of possible meteorological forcing could be interesting either by statistical methods [Raleigh et al., 2015] or by outputs of ensemble numerical weather prediction systems [Vernay et al., 2015].

4.2.2 Partial discrepancy between model state variables and measures

To be efficient, the reinitialization must inject physical quantities as they are implemented into the model. Crocus evolution rules are self-consistent: there were evaluated so that they give a correct agreement on snow height until the end of the season, but they don't necessarily represent exactly the actual phenomenon because of high equifinality between different physical processes [Lafaysse et al., 2017].

For instance, reinitialization of the whole snowpack with density while model represents snow compaction and the model could represent a too fast or too slow compaction in this period. Reinitializing snow density could introduce errors on the long term because real compaction and simulated compaction were not compatible. That is why it will be necessary to evaluate changes on the model on the detailed snow profiles in the future to detect these inconsistencies.

It was also tried to use a measure of SSA with an automatic profiler, the ASSSAP instrument (Alpine Snowpack Specific Surface Area Profiler, the light version of POSSUM [Arnaud et al., 2011]), available on season 2011-2012 at the Col de Porte station. It measures SSA with optical reflectance methods with a high resolution (1 mm step). But error estimation was greater with this instrument than without reinitializing SSA and much greater than using a very simple table from Crocus code for estimating SSA from grain type. This could be explained by the fact that SSA measured by ASSSAP is not exactly the same variable as SSA used by Crocus. Figure 12 presents the ASSSAP SSA depends on SSA simulated by Crocus. This shows that Crocus does not represent the SSA as

ASSSAP measures it. Moreover, ASSSAP SSA could not be used to redefine Crocus SSA : for instance the mean value measured for deep hoar or melt forms is 11 $m^2\,kg^{-1}$ (data presented in Carmagnola, 2013 from Domine et al., 2007) whereas Crocus will identify deep hoar only in the range $0-6.2$ and melt forms between 0 and 11.9 $m^2\,kg^{-1}$. No obvious relationship could be deduced from the scatter plot, without more research to determine which of the two SSA is closer of reference SSA (methane measurement or micro-tomography). As a result, ASSSAP is useless to reinitialize Crocus simulations.

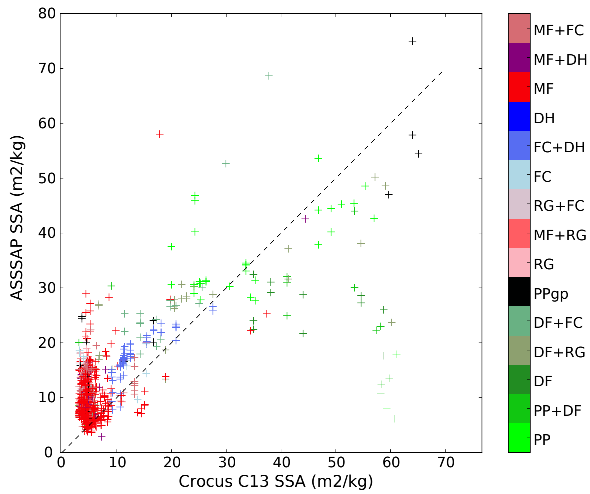

Figure 12: SSA measured by ASSSAP depends on SSA used by Crocus determined by a Crocus simulation at Col de Porte on season 2011-2012. Color represents the grain type identified by observer. Dashed black line is 1:1.

These examples show the necessity of having in the model observable variables to be able to use observations to correct it, or at least reliable relations between observation variables and model ones. Nowadays, the model being tested mainly on bulk and surface variables, the model variables for detailed layering do not exactly correspond to observable ones. This is an important limitation on future improvement based on this work. One next step could be to reevaluate evolution laws with large observation dataset to evolve representation of physical phenomena so that they match better with reality, but this may represent a large extent of work. Furthermore, all physical processes would have to be dealt simultaneously for an overall improvement of the simulations due to the equifinality between them.

Furthermore, in the best reinitialization, all variables need to be consistent all together. If Crocus uses enthalpy preventing inconsistent liquid water content and temperature (which is a classical issue in observations because of thermometer bias), grains are identified by sphericity, size (computed from SSA) and historic variable, so these three variables have to be consistent. In our set of variables used, these three variables are computed from grain type, but this explains the poorer identification of grain type when SSA is not reinitialized with sphericity and historic variable (see Figure 9). Moreover, the three set of variables (density, grain type and enthalpy) could have an influence on each other : for instance, a layer with a too weak density will evolve quickly with mechanical compaction, changing all the properties of the layer. This is illustrated by Figure 9, and most visible on the reinitialization without density, which harms a lot grain type, SSA and enthalpy variables.

4.2.3 Model bias identification

It was observed that on the long term, improvement tends to ease, if not to reduce to zero (absolute error tends to reference absolute error). This could be because the reinitialized simulation tends to reference but also because the simulations tends to a different state which error is of the same magnitude as the reference. To better understand that, absolute difference between reference and reinitialized simulation was computed, one month after reinitialization. Results presented in Figure 8 show that on the long term, this distance increases between observation and reinitialized simulation whereas it decreases between reinitialized simulation and reference, even though it do not completely come back to reference, especially on density and grain type. Effect is more visible on variables influenced more quickly by the weather such as LWC or temperature.

This could be explained because variables used are not the same variable as the model ones, so model evolution rules are not consistent with our data and tends to evolve towards its own variables. But this also could illustrate some model biases. For instance, the bias observed for LWC could be explained by error in representation of liquid water content but it could also represent the bad water retention and percolation in the snowpack. On Figure 6, this could be seen that on Feb. 24th and Mar. 10th, observer record on large depth ranges a dry snowpack whereas Crocus simulate a liquid water content of more than 10 $\mathrm{kg\,m^{-3}}$. This was observed on many other events. Even though humidity classes are quite questionable, the dry class is rather precise, so it could indicate a model bias. This gives a first prospect of our method : the identification of model bias and physical processes requiring improvement. On that topic, D'Amboise et al., 2017 are working on implementing a more physical representation of liquid water percolation using the Richards equations. To be affirmative on model errors, it is nevertheless necessary to take into account errors on observations.

4.3 DTW matching

The usefulness of the DTW matching could be estimating looking at the result without this matching and computing the mean displacement. The mean displacement on our three sites on 15 seasons (2000-2015) is 17 mm, this is the value after global stretching which mean displacement is about 16 cm and this is more than the precision of manual measurement (around 1 cm) so the matching is useful and fully operating. This could also be seen on examples as in Figure 5, with the change of layers proportions between (b) and (c) to match melt forms layers and fresh snow layers.

The second question is on the necessity to have observations each week so that the snowpack description is all the time close to the observations, for the matching to be efficient. Reinitializations each week was tried as well as reinitializing at each observation on a Crocus simulation started in August on bare soil. The maximum score variations are 10 % and generally 5-6 % but the global score is unchanged. This tends to indicate that matching does not depend significantly on the time location of the previous observation, at least when the model is not worse than Crocus in its long-term snowpack simulation.

However, this type of matching is very efficient when used with data with good dynamics and perfectly comparable, because distance is only computed on absolute differences, and must vary quickly for a better identification. Observations and simulation used grid on typical thickness 1 cm and absolutes values are not always perfectly comparable as shown in Section 4.2.2. With our example sites, it was quite efficient because observers reports a lot of details giving a sufficient dynamics, but this is not always the case in the observation network. This matching was conserved in our study because of its simplicity and time efficiency. It was not necessary to have a very accurate matching because most of variables were reinitialized, and only age was conserved, so that small error on matching have a little effect on final error. But matching is something that could be improved. For instance, introducing peak identification (to match crusts for instance) could help to see layers limit. It is also possible to add some knowledge from snowpack comparison such as matching first new snow to eliminate the amount of precipitations due for instance at the time lapse between simulation reinitialization and observation or at a time shift in meteorological forcing (made at a massif scale so time of precipitation could vary from one end to the other). This could improve matching for reinitialization as well as error estimation.

4.4 Interest of reinitialisation depending on frequency of observations and variables used

4.4.1 Variables to be used

According to Figure 9, reinitializing all variables that seem to be sufficiently compatible with model ones have no averaged negative effect and improve significantly total scores. Reinitializing all variables on which relevant information are available also allow to start the simulation with a more consistent snowpack. That is why once observer has correctly reported grain type, density (even partially), humidity and temperature profile, it is possible to reinitialize all the variable set that is to say enthalpy (from temperature and humidity class), density and SSA, sphericity and historic variable (from grain type).

But it is interesting to notice that if only incomplete observations are available, they could be used too. For instance, on automatic stations where only snow height is available and not the detailed stratigraphy. scores for reinitialisations with only snow height integrated were computed. This corrects on the short-term the snow height and do not deteriorate scores on other variables regarding to the reference.

Currently, the most questionable variable seems to be the SSA, as a low improvement on grain type identification is observed when SSA is not reinitialized. Indeed, it is computed with reduced informations because of the lack of measurements compatible with Crocus values. While no direct, quick and cheap measurement can be performed over an extensive network, advances in the link between grain type observations and Crocus values of SSA is required to improve significantly grain reinitialization which appears as a difficult variable to reproduce in this study.

It was also tried to use more accurate observations, with the use of the automatic penetrometer Snow Micro Pen (SMP, [Schneebeli et al., 1999]). The SMP signal is not usable directly because our model computes a penetration resistance for a ramsonde (40 mm diameter whereas SMP has a 5 mm diameter) which is not similar to SMP signal. So SMP was used to estimate a high-resolution density profile with the method presented by Löwe and van Herwijnen, 2012 and improved by Proksch et al., 2015. Depth grid was adjusted by matching well visible variation of stratigraphic density with SMP density profile. Quite accurate manual densities were used to unbias computed density. In this case, mean improvement is not better than if manual recorded density was used, filled with Table of Appendix A. Indeed, the model have not the high resolution of SMP, so that resolution of manual measurements are sufficient : all resolution carried by SMP is lost when averaging on numeric layers.

4.4.2 Frequency of observations and when to use it for reinitialization

The next question is the frequency necessary for reinitializations (thus observations). Except for temperature, which could vary quickly due to the weather (mainly), Section 3.4 show that improvements have a typical persistence time around 1 week. This is today the frequency of observations on our best experimental stations, and this seems to be consistent with the typical time of error evolution in the model.

It was not possible to test more frequent reinitializations because of the lack of observations, but Figure 11 shows that if the model is too close from the observation, observation precision is a limitation that leads to a degradation of the simulation scores in case of reinitialization.

To quantify this effect, thresholds could be determined, under which the simulation should not be reinitialized. To this end, plots like Figure 11 are used. An immediate effect could be expected if model error is sufficient so that correction improve results in more than 80 % of cases immediately after and a mid-term effect could be expected if one week after 80 % of simulations reveals an improvement compared to reference. This gives Table 3, with threshold under which reinitializing have a more disputable interest on the long-term and even could be worse than letting simulation continue if the reference score is below the immediate (3h) improvement threshold. This table could be a basis to choose whether or nor reinitializing the simulation depending on the score computed with the reference just before observation.

Variable	Error threshold for immediate improvement	Error threshold for 1-week improvement
Temperature (K)	0.2	1.2
Total height (m)	0.01	0.13
Density ($\mathrm{kg\,m^{-3}}$)	10	50
LWC (-)	0.03	0.25
SSA ($\mathrm{m^2\,kg^{-1}}$)	1	4
Layer thickness (cm)	0	2.5

Table 3: Absolute error threshold on reference, under which more than 20 % of reinitialized simulation gives results worse than the non-reinitialized simulation 3 h after reinitialization (short-term) and 1 week after (long-term).
.

4.5 Distance choice

We have chosen a distance measurement method for each variable, consistent with data and precision, and adapted to the goal of the simulation. If density distance (and all continuous variables) is quite natural, LWC is based on thresholds, between dry and slightly wet and between very wet and saturated. This could give a relevant indicator but could also over-estimate error easily if model or observation is not far from one of the thresholds.

In all cases, this is a choice, and many other distances are possible. Other choices could change the improvement obtained and give a different threshold for reinitialization.

The model being designed and used (among others goals) for avalanche hazard forecasting, it is also possible to include distances specific of this topic. for instance, it is possible to evaluate ram resistance agreement. The problem with this variable is that ram resistance diagnosed by Crocus is designed to highlight the presence of weak faceted layers : variations are largely over-estimated by the model and ram resistance computed is very low compared to measurements so that it gives a lower bound of snowpack stability to forecasters. So, before using this variable, it is compulsory to compute an observable diagnostic ram resistance in Crocus. An other idea could be using the avalanche risk level, from a numerical tool (Meteo-France developed on, called MEPRA, [Giraud, 1991], to compute it from Crocus output) or a risk level estimated by forecasters. These indicators currently lacks of reliability, so this study focused on more basic but more reliable indicators from quite well defined variables.

4.6 Future prospect : ensemble assimilation

4.6.1 Forcing ensemble

A method to quantify errors between model and observations was established and good results of reinitializing were obtained, which means that observed variables could be simulated correctly by the model.

However, our method only injects observations in the model, forgetting the simulation results. This does not take into account observation uncertainty, no more than the model accuracy. Besides, improvement could only be obtained on points where observations are performed.

All of this lead us to prefer in the future using data assimilation algorithm. This requires to build an appropriate ensemble of meteorological forcing (from ensemble-forecasting models or build with perturbations around a deterministic model), or physics ensemble like in Lafaysse et al., 2017. In a model with variable dimension length, a lagrangian discretization, and highly non linear processes, simple ensemble algorithms such as a particle filter can be a good option [Magnusson et al., 2014; Charrois et al., 2016].

Improvement results on a single point with rough reinitialization gives good prospects for data assimilation of vertical stratigraphies whereas all previous studies on that topic have been focused on assimilating a bulk or surface observations on the snowpack. The methodology described below to compute a distance between an observed and a simulated profile could be directly applied for that purpose. Nevertheless, note that the use of this method in spatialized observations is questionable due to the lack of observations.

4.6.2 Multiphysical ensemble

When changes are made on the model, this method gives the possibility to compare more precisely the impact than with the only total snow depth, total snow equivalent or surface albedo commonly used for these impact studies. Given that it have been shown observations could give relevant information on model errors, intercomparison and validation of detailed snowpack models as Crocus should also include an evaluation on the detailed stratigraphy. Studies as Lafaysse et al., 2017 could be improved reducing the equifinality by comparing also the detailed stratigraphy obtained.

5 Conclusion

5.1 Scientific conclusion

This study proposes a method to evaluate snowpack simulation on its detailed stratigraphy usi<ng manual snow pit observations and a method to reinitialize simulations with these observations during the winter. This was applied to the SAFRAN-SURFEX/ISBA/Crocus model chain, with observations from the Meteo-France observer network, on seasons from 2000-2001 to 2014-2015, on 3 stations in the french Alps.

This method allows to compare model results with snow pit observations. This has never been done before for Crocus model, which was evaluated mainly on bulk and surface variables whereas it represents the detailed stratigraphy of the snowpack. A set of distances to compare agreement between snowpack manual observation and simulation results was provided on the main variables of the model. Future evaluation of detailed snowpack models should integrate such detailed comparison as soon as a sufficient amount of observations is available. This could also be a mean of detecting model intrinsic error and physical processes to be improved.

It was also found that snow pit measurement could usefully be used to correct model error during the winter, especially when model errors are high, by reinitializing some model state variable with measurements. This method is quite basic but allows to see potential and limits of these data.

The next step would be to apply more advanced assimilation algorithms to take into account all together observation, model and forcing uncertainties. The set of distances proposed here for comparison between observation and simulation could directly be applied to some of these algorithms. This approach, integrating snowpack observations in the simulation during the winter, could improve significantly simulations, correcting model or forcing errors. These improvements are expected to increase the accuracy of punctual simulations and consequently the confidence of the forecasters in these simulations.

5.2 Internship conclusion

This internship brings me a better view of the scientific research environment and reinforced my idea of continuing with a PhD thesis after my Masters in Multiscale Approaches for Materials and Structures in École des Ponts et Chaussées. It allows me to both discover an application field of mechanical research with the very interesting material which is snow and a possible prospect of my professional path with the Météo-France research center (CNRM). It also allows me to better understand the organisation of large public organisations with the example of Météo-France, such organisations which could employ me as ingénieur des Ponts, Eaux et Forêts.

Appendix A Tables used

To be able to reinitialize variables that are not measurable, such as sphericity, we need to infer from other measured variables. Grain type is the more evident for variables used then to redefine a grain type (SSA, sphericity, historic variable), and have a sufficient range of possible variables to be quite precise.

Table for reinitializing sphericity (Table 4) and density (Table 5) comes from existing codes used at Météo-France to compute model variables such as penetration resistance from stratigraphic variables. Historic variable is used only to identify grain type, preventing all possible reverse evolution. the meaning of this variable is presented in Table 9 and reinitialization table is presented in Table 6. It comes from the inversion of the grain identification table of Crocus. This source could also give a table of SSA depending on the grain type, which is completed, to be more precise on variations between grain types, by measurements from Carmagnola, 2013 and Domine et al., 2007, because Crocus gives only 3 ranges of values. This table is more simple because of the lack of data, measurable SSA being different of Crocus SSA as shown in Section 4.2.2.

Secondary grain	PP	DF	RG	FC	DH	MF	IF	SH	PPgp
Main grain									
PP	0.50	0.50	0.75	0.25	0.00	0.99	0.50	0.50	0.45
DF	0.50	0.50	0.70	0.30	0.00	0.99	0.50	0.50	0.45
RG	0.90	0.80	0.99	0.60	0.50	0.99	0.99	0.90	0.75
FC	0.10	0.20	0.40	0.00	0.00	0.40	0.30	0.00	0.10
DH	0.00	0.00	0.60	0.00	0.00	0.35	0.25	0.00	0.50
MF	0.99	0.99	0.99	0.60	0.55	0.99	0.99	0.90	0.90
IF	0.50	0.50	0.99	0.30	0.25	0.99	0.50	0.50	0.50
SH	0.50	0.50	0.90	0.50	0.50	0.90	0.50	0.50	0.50
PPgp	0.45	0.45	0.65	0.10	0.50	0.75	0.50	0.50	0.50

Table 4: Sphericity associated with grain type, taking into account the main grain type and the secondary one (here represented by their international coding, see [Fierz et al., 2009])
.

Secondary grain	PP	DF	RG	FC	DH	MF	IF	SH	PPgp
Main grain									
PP	100	150	150	100	100	180	180	100	120
DF	150	180	230	200	200	250	250	180	180
RG	200	230	300	250	250	350	450	200	200
FC	180	200	250	250	280	350	450	180	200
DH	180	200	250	280	300	350	450	180	200
MF	180	250	350	350	350	400	450	400	350
IF	180	250	450	450	450	450	450	450	450
SH	100	180	200	180	180	400	450	100	250
PPgp	120	180	200	200	200	350	450	250	250

Table 5: Density ($kg.m^{-3}$) associated with grain type, taking into account the main grain type and the secondary one.

Secondary grain	PP	DF	RG	FC	DH	MF	IF	SH	PPgp
Main grain									
PP	0 (2)	0 (2)	0 (2)	0 (3)	1 (3)	2 (2)	2 (2)	1 (3)	0 (2)
DF	0 (2)	0 (2)	0 (2)	1 (3)	1 (3)	2 (2)	2 (2)	1 (3)	0 (2)
RG	0 (2)	0 (2)	0 (0)	0 (0)	1 (3)	2 (2)	2 (2)	1 (3)	0 (2)
FC	1 (3)	1 (3)	0 (0)	0 (0)	1 (3)	3 (3)	3 (3)	1 (3)	1 (3)
DH	1 (3)	1 (3)	1 (3)	1 (3)	1 (1)	3 (3)	3 (3)	1 (3)	1 (3)
MF	2 (2)	2 (2)	2 (2)	3 (3)	3 (3)	2 (2)	2 (2)	3 (3)	2 (2)
IF	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)
SH	1 (3)	1 (3)	1 (3)	1 (3)	1 (3)	3 (3)	3 (3)	1 (3)	1 (3)
PPgp	0 (2)	0 (2)	0 (2)	1 (3)	1 (3)	2 (2)	1 (3)	1 (3)	0 (2)

Table 6: Historic variable associated with grain type, taking into account the main grain type and the secondary one and the liquid water class. The values in brackets are used when humidity is detected in the layer.

Grain type	PP	DF	RG	FC	DH	MF	IF	SH	PPgp
SSA	40	30	20	25	4	7	2	4	20

Table 7: SSA ($m^2.kg^{-1}$) associated with grain type. SSA is computed from main grain type and secondary one as the barycenter of values for main grain type with weight 2/3 and secondary one with weight 1/3.

Grain type	PP	DF	RG	FC	DH	MF	IF	SH	PPgp
PP	0.0	0.2	0.5	0.8	1.0	1.0	1.0	1.0	0.8
DF	0.2	0.0	0.2	0.6	1.0	1.0	1.0	1.0	0.6
RG	0.5	0.2	0.0	0.6	0.9	1.0	0.0	1.0	0.5
FC	0.8	0.6	0.6	0.0	0.2	1.0	0.0	1.0	0.2
DH	1.0	1.0	0.9	0.2	0.0	1.0	0.0	1.0	0.3
MF	1.0	1.0	1.0	1.0	1.0	0.0	0.2	1.0	1.0
IF	1.0	1.0	0.0	0.0	0.0	0.2	0.0	1.0	1.0
SH	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
PPgp	0.8	0.6	0.5	0.2	0.3	1.0	1.0	1.0	0.0

Table 8: Distance between two grain types, from [Lehning et al., 2001], completed.

Valeur	Signification
1	Has been angular but was never in contact with liquid water
2	Has been in contact with liquid water but was never angular
3	Has been in contact with liquid water and has been angular
4	Same as 2 and has undergone several melt-freeze cycles
5	Same as 3 and has undergone several melt-freeze cycles
0	All other cases

Table 9: Values of the historical variable and the meaning in Crocus model

List of Figures

References

Arnaud, L., Picard, G., Champollion, N., Domine, F., Gallet, J., Lefebvre, E., Fily, M., and Barnola, J. (2011). Measurement of vertical profiles of snow specific surface area with a 1 cm resolution using infrared reflectance: instrument description and validation. *Journal of Glaciology*, 57:17–29.

Bartelt, P. and Lehning, M. (2002). A physical snowpack model for the swiss avalanche warning. *Cold Regions Science and Technology*, 35(3):123 – 145. doi:10.1016/S0165-232X(02)00074-5.

Boone, A. and Etchevers, P. (2001). An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an alpine site. *Journal of Hydrometeorology*, 2(4):374–394. doi:10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2.

Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C. (1989). An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of glaciology*, 35(121):333–342.

Brun, E., Vionnet, V., Morin, S., Boone, A., Martin, E., Faroux, S., LeMoigne, P., and Willemet, J.-M. (2012). Le modèle de manteau neigeux crocus et ses applications. *La Météorologie*, 76:44–54.

Carmagnola, C. (2013). *Mesure, analyse et modélisation des processus physiques du manteau neigeux sec*. PhD thesis, Université de Grenoble.

Carmagnola, C., Morin, S., Lafaysse, M., Domine, F., Lesaffre, B., Lejeune, Y., Picard, G., and Arnaud, L. (2014). Implementation and evaluation of prognostic representations of the optical diameter of snow in the surfex/isba-crocus detailed snowpack model. *The Cryosphere*, 8(2):417–437.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. *The Cryosphere*, 10:1021–1038. doi:10.5194/tc-10-1021-2016.

D'Amboise, C. J. L., Müller, K., Oxarango, L., Morin, S., and Schuler, T. V. (2017). Implementation of a physically based water percolation routine in the crocus (v7) snowpack model. *Geoscientific Model Development Discussions*, 2017:1–32. doi:10.5194/gmd-2017-56.

Decharme, B., Boone, A., Delire, C., and Noilhan, J. (2011). Local evaluation of the interaction between soil biosphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *Journal of Geophysical Research: Atmospheres*, 116(D20). doi:10.1029/2011JD016002.

Domine, F., Morin, S., Brun, E., Lafaysse, M., and Carmagnola, C. (2013). Seasonal evolution of snow permeability under equi-temperature and temperature-gradient conditions. *The Cryosphere*, 7(6):1915–1929. doi:10.5194/tc-7-1915-2013.

Domine, F., Taillandier, A.-S., and Simpson, W. R. (2007). A parameterization of the specific surface area of seasonal snow for field use and for models of snowpack evolution. *Journal of geophysical research*, 112:F0231. doi:10.1029/2006JF000512.

Durand, Y., Giraud, G., Brun, E., Mérindol, L., and Martin, E. (1999). A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. *Journal of Glaciology*, 45(151):469–484. doi:10.1017/S0022143000001337.

Durand, Y., Laternser, M., G.Giraud, P.Etchevers, B.Lesaffre, and Mérindol, L. (2009). Reanalysis of 44 yr of climate in the french alps (1958–2002): Methodology, model validation, climatology, and trends for air temperature and precipitation. *Journal of applied meteorology and climatology*, 48:429–449. doi:10.1175/2008JAMC1808.1.

Essery, R., Martin, E., Douville, H., Fernandez, A., and Brun, E. (1999). A comparison of four snow models using observations from an alpine site. *Climate Dynamics*, 15:583–593.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. (2013). A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55:131–148. doi:10.1016/j.advwatres.2012.07.013.

Fierz, C., Armstrong, R., Durand, Y., Etchevers, P., Greene, E., McClung, D., Nishimura, K., Satyawali, P., and Sokratov, S. (2009). The international classification for seasonal snow on the ground. *IHP-VII Technical Documents in Hydrology*, 83.

Giraud, G. (1991). Mepra: Modèle expert d'aide à la prévision du risque d'avalanches. *ANENA, Comptes rendues Syposium de Chamonix CISA-IKAR*, pages 248–254.

Hagenmuller, P. and Pilloix, T. (2016). A new method for comparing and matching snow profiles, application for profiles measured by penetrometers. *Frontiers in Earth Science*, 4:1–13. doi:10.3389/feart.2016.00052.

Harper, J. and Bradford, J. (2003). Snow stratigraphy over a uniform depositional surface: spatial variability and measurement tools. *Cold Regions Science and Technology*, 37(3):289–298. doi:10.1016/S0165-232X(03)00071-5.

Jordan, R. (1991). A one-dimensional temperature model for a snow cover: Technical documentation for sntherm 89. Technical report, Cold regions research and engineering lab, Hanover NH.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S. (2017). A multiphysical ensemble system of numerical snow modelling. doi:10.5194/tc-11-1173-2017.

Lafaysse, M., Morin, S., Coléou, C., Vernay, M., Serca, D., Besson, F., Willemet, J.-M., Giraud, G., and Durand, Y. (2013). Towards a new chain of models for avalanche hazard forecasting in french mountain ranges, including low altitude mountains. *Proceedings of International snow science workshop Grenoble*.

Lehning, M., Fierz, C., and Lundy, C. (2001). An objective snow profile comparison method and its application to snowpack. *Cold Regions Science and Technology*, 33:253–261. doi:10.1016/s0165-232x(01)00044-1.

Löwe, H. and van Herwijnen, A. (2012). A poisson shot noise model for micro-penetration of snow. *Cold Regions Science and Technology*, 70:62–70. doi:10.1016/j.coldregions.2011.09.001.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T. (2014). Assimilation of point swe data into a distributed snow cover model comparing two contrasting methods. *Water Resources Research*, 50(10):7816–7835. doi:10.1002/2014WR015302.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A. (2013). The surfexv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, 6(4):929–960. doi:10.5194/gmd-6-929-2013.

Moigne, P. L. (2012). Surfex scientific documentation. www.umr-cnrm.fr/surfex/IMG/pdf/surfex_scidoc_v2-2.pdf.

Morin, S., Domine, F., Dufour, A., Lejeune, Y., Lesaffre, B., Willemet, J.-M., Carmagnola, C., and Jacobi, H.-W. (2013). Measurements and modeling of the vertical profile of specific surface area of an alpine snowpack. *Advances in Water Resources*, 55:111–120. doi:10.1016/j.advwatres.2012.01.010.

Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P., and Sudul, M. (2012). An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (col de porte, france, 1325 m alt.) for driving and evaluating snowpack models. *Earth System Science Data*, 4:13–21. doi:10.5194/essdd-5-29-2012.

Pahaut, E. and Giraud, G. (1995). Avalanche risk forecasting in france: results and prospects. *La Météorologie*.

Proksch, M., Löwe, H., and Schneebeli, M. (2015). Density , specific surface area , and correlation length of snow measured by high-resolution penetrometry. *Journal of Geophysical Research: Earth Surface*, 120:346–362. doi:10.1002/2014JF003266.

Proksch, M., Rutter, N., Fierz, C., and Schneebeli, M. (2016). Intercomparison of snow density measurements: bias, precision, and vertical resolution. *The Cryosphere*, 10:371–384. doi:10.5194/tc-10-371-2016.

Raleigh, M., Lundquist, J., and Clark, M. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Science*, 19:3153–3179. doi:10.5194/hess-19-3153-2015.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49. doi:10.1109/TASSP.1978.1163055.

Schaller, C. F., Freitag, J., Kipfstuhl, S., Laepple, T., Steen-Larsen, H. C., and Eisen, O. (2016). A representative density profile of the north greenland snowpack. *The Cryosphere*, 10(5):1991–2002. doi:10.5194/tc-10-1991-2016.

Schneebeli, M., Pielmeier, C., and Johnson, J. B. (1999). Measuring snow microstructure and hardness using a high resolution penetrometer. *Cold Region Science and Technology*, 30:101–114.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V. (2011). The arome-france convective-scale operational model. *Monthly Weather Review*, 139(3):976–991. doi:10.1175/2010MWR3425.1.

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., and Morin, S. (2015). Ensemble forecasting of snowpack conditions and avalanche hazard. *Cold Regions Science and Technology*, 120:251–262. doi:10.1016/j.coldregions.2015.04.010.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Moigne, P. L., Martin, E., and Willemet, J.-M. (2012). The detailed snowpack scheme crocus and its implementation in surfex v7.2. *Geoscientific model development*, 5:773–791. doi:10.5194/gmd-5-773-2012.