Dear Editor and Referees!

The authors want to thank you for your valuable comments and suggestions. We are sure they led to an improvement of the manuscript.

Referee comments RC1 to RC3 as well as short comment SC1 were answered in detail during the discussion process already (see AC1-4). Now we provide the revised paper as a fluently readable document and as a version where all changes are marked and commented (see below). All specific referee comments were considered according to AC1-4. In addition, all general criticism is addressed and the paper was improved accordingly:

- The paper was restructured and sharpened. Its aim – introducing and evaluating a new method to model *SWE* – was emphasized.
- The placement of the model within the snow modeling spectrum was specified, not least by providing Table 1.
- The entangled descriptions of model code, modules and snow physics were clarified, not least by including Table 2.
- Language and writing were improved; repetitions eliminated and parentheses avoided.
- Requested citations were analyzed and included, whenever possible.
- More details on the data used for calibration and validation were given. A respective table was added in the appendix (Table A1).
- The application example was moved to the appendix.
- Discussion-like parts from the initially submitted manuscript's Results section were moved to the Discussion section.
- Main text was shortened by ca. 3000 words (>20%).

We updated the results of our error analysis, initiated by RC2 but without providing relative numbers (see AC2). We now provide RMSEs for all *SWE* and SWE_{pk}, as well as gradings depending on absolute *SWE* (cf. Fig. 2, Table 4 etc.), instead of statistics of RMSEs from different stations. This changed some key values concerning model accuracy, but none of the overall statements. Revisiting data sources and providing Table A1 revealed some minor inconsistencies concerning the number of observations and years, which could be solved easily.

We hope the revised manuscript fulfills HESS journal requirements for a final revised paper, and we are looking forward to your answer.

With best regards,
Michael Winkler, Harald Schellander, Stephanie Gruber

Snow Water Equivalents exclusively from Snow Depths and their temporal Changes: The ΔSNOW.MODEL

Michael Winkler[1,*], Harald Schellander[1,2,*], and Stefanie Gruber[1]

[1]ZAMG – Zentralanstalt für Meteorologie und Geodynamik, Innsbruck, Austria
[2]Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria
[*]These authors contributed equally to this work.

Correspondence: Michael Winkler (michael.winkler@zamg.ac.at) and Harald Schellander (harald.schellander@zamg.ac.at)

Abstract. [1]Snow depths~~Snow heights~~[2] have been manually observed for many years, sometimes decades, at various places around the globe. These records are often of good quality. In addition, more and more data from automatic stations and remote sensing are available. On the other hand, records of snow water equivalent (SWE)~~SWE~~ – synonymous for snow load or mass – are sparse, although it might be the most important snowpack feature in ~~fields like~~ hydrology, climatology, agriculture, natural hazards research, etc. SWE very often has to be modeled, and respective~~those~~ models either depend on meteorological forcing or are not intended to simulate individual SWE values, like the substantial seasonal "peak SWE".

The ΔSNOW.MODEL is presented as a new method to simulate local-scale SWE. It solely needs a regular time series of snow depths as input.~~snow depths as input, though a gapless record thereof. Temporal resolution of the data series is no restriction per se.~~ The ΔSNOW.MODEL is a semi-empirical multi-layer model and freely available as R-package. Snow compaction is modeled following the rules of Newtonian viscosity. The model considers measurement errors, treats overburden loads due to new~~fresh~~[3] snow as additional unsteady compaction, and melted mass is stepwise distributed top-down in the snowpack.

Seven model parameters are subject to calibration. Snow observations of 67 winters from 14 stations, well-distributed over different altitudes and climatic regions of the Alps, are used to find an optimal parameter setting. Data from another 71 independent winters from 15 stations is used for validation.~~, which was performed using 71 winters from 14 stations, well-distributed over different altitudes and climatic regions of the Alps. Another 73 rather independent winters act as validation data.~~ Results are very promising: Median bias and root mean square~~squared~~[4] error for SWE are only $-3.0\,\mathrm{kg\,m^{-2}}$ and $30.8\,\mathrm{kg\,m^{-2}}$, and $+0.3\,\mathrm{kg\,m^{-2}}$ and $36.3\,\mathrm{kg\,m^{-2}}$~~$-4.0\,\mathrm{kg\,m^{-2}}$ and $23.9\,\mathrm{kg\,m^{-2}}$, and $+2.3\,\mathrm{kg\,m^{-2}}$ and $23.1\,\mathrm{kg\,m^{-2}}$~~ for peak SWE, respectively. This is a major advance compared to snow models relying on empirical regressions and even sophisticated thermodynamic snow models do~~, but also much more sophisticated thermodynamic snow models~~ not necessarily perform better.

~~Not least, this study outlines the need for comprehensive comparison studies on SWE measurement and modeling at the point and local scale.~~

[1]Changes based on specific and general reviewer comments are marked in blue and orange, respectively.
[2]Snow "height" was changed to snow "depth" throughout the paper and also in the title. The respective symbol H was changed to HS, and h was changed to hs.
[3]"Fresh" snow was changed to "new" snow throughout the paper.
[4]Root mean "squared" error was changed to root mean "square" error throughout the paper.

1 Introduction

Depth (HS)~~Total H~~ and bulk density (ρ_b) are fundamental characteristics of a seasonal snowpack (e.g., Goodison et al., 1981; Fierz et al., 2009). Equation (1) links them to the areal density $[\mathrm{kg\,m^{-2}}]$ of the snowpack, which – in hydrological applications – is usually referred to as snow water equivalent (SWE), as it resembles "the depth of water that would result if the mass of snow melted completely" (Fierz et al., 2009).

$$SWE = HS \cdot \rho_\mathrm{b} \qquad (1)$$
$$[1\,\mathrm{kg\,m^{-2}} \equiv 1\,\mathrm{mm\,water\,equivalent\,(w.e.)}]$$

1.1 Measurements of HS and SWE

Measuring HS is relatively easy (e.g., Sturm and Holmgren, 1998): Manual measurements at a certain point only require a rod or ruler (e.g., Kinar and Pomeroy, 2015), and decades-long series of daily HS measurements exist in many regions – in lowlands as well as in alpine areas (e.g., Haberkorn, 2019). In modern times more and more HS data from automated measurements (mostly sonic or laser distance ranging) become available, typically in sub-hourly resolution (McCreight and Small, 2014). In addition, remote sensing techniques currently increase the number of HS data significantly, having the advantage of an areal picture instead of point information but at the cost of accuracy and in most cases also~~– in most cases –~~ temporal resolution and regularity (Cf., e.g., Dietz et al. (2012) for a general review of methods, and Deems et al. (2013) for a review on lidar measurements. Painter et al. (2016) provide a thorough overview. Garvelmann et al. (2013) and Parajka et al. (2012), e.g., illustrate the potential of timelapse photography.)

In contrast, measurements of SWE (or ρ_b) are more difficult (e.g., Sturm et al., 2010): Manual measurements require some basic equipment like snow tubes or snow sampling cylinders, a bit of dexterity, and are time consuming. In case snow depth exceeds the sampling tool's size a pit has to be dug to consider the layered structure of the snowpack~~are time consuming (especially if a snow pit is dug), require some basic equipment like snow tubes or snow sampling cylinders and – not least – some dexterity~~ (e.g., Kinar and Pomeroy, 2015). As a consequence, SWE measurements are carried out at much fewer locations than HS measurements (e.g., Mizukami and Perica, 2008; Sturm et al., 2010), their accuracy is lower, and series are shorter. Only in very rare cases consecutive, decades-long measurement series are available (e.g., in Switzerland; cf. Jonas et al., 2009). Often they are only carried out at irregular time intervals ("snow courses") and even if regularly measured, temporal resolution is hardly ever higher than two weeks. A~~Besides these restrictions concerning manual measurements, a~~lso automatic measurements of SWE are not at all comparable in quality and quantity with automated HS measurements. They are quite expensive, often inaccurate, still at a developmental stage, and/or suffer from significant problems if not intensively maintained throughout the snowy season. Methods involve weighing techniques (snow scales; e.g., Smith et al., 2017; Johnson et al., 2015), pressure measurements (snow pillows; e.g., Goodison et al., 1981), upward-looking ground penetrating radar (~~GPR;~~ e.g., Heilig et al., 2009), passive gamma radiation (e.g., Smith et al., 2017), cosmic ray neutron sensing (~~CRNS;~~ e.g., Schattan et al., 2019), L-band Global Navigation Satellite Signals (~~GNSS;~~ e.g., Koch et al., 2019), etc. Presumably, the biggest and best serviced network of automated SWE measurements is SNOTEL with about 800 sites in Western North America (Avanzi et al., 2015).

SWE data from remote sensing are not operationally available for the local and point scale~~Furthermore, there is no way to directly monitor SWE by remote sensing techniques (Schaffhauser et al.,2008)~~, and deriving this snow property from satellite products at sub-kilometer resolution~~the local scale (< 1 km)~~ is still not possible (Smyth et al., 2019). On top of that, there is the issue of longterm availability: automated measurements and ~~(~~at least rough~~)~~ remote sensing of SWE have not been available for more than some twenty years at their best (e.g., SNOTEL, operated since the late 1990s), a fairly short timespan compared to decades-long daily HS data (e.g., Kinar and Pomeroy, 2015).

Regardless of these problematic circumstances accompanying SWE measurements, many hydrological, agricultural, and other~~hydrological and agricultural~~ applications depend on good estimates of SWE (e.g., Goodison et al., 1981; Sturm and Holmgren, 1998). Ultimately,~~Very often, in the end~~ the mass of water stored in the snowpacks matters very often and, therefore,~~, and that's why~~ the majority of those fields is especially interested in seasonal SWE maxima, i.e. "peak SWE" (SWE_{pk}). SWE_{pk} are also the main focus of different kinds of extreme value and climatic analyses, both of which additionally very much rely on longterm or even "historical" data. Not least, snow load standards (e.g., International Organization for Standardization, 2013) rely on extreme value analyses of longterm ~~snow load records (or~~ SWE records, as snow load is defined as the product of SWE and the gravitational acceleration.~~, respectively; snow load $SL = SWE \cdot g$, with gravitational acceleration $g = 9.8\,\mathrm{m\,s^{-2}}$)~~. These points reveal the great discrepancy between the good data situation in terms of HS on the one hand, and the insufficient availability of SWE data on the other.

1.2 Modeling SWE

Modern snow models like Crocus (e.g., Vionnet et al., 2012), SNOWPACK (e.g., Lehning et al., 2002), SNTHERM (Jordan, 1991), or the dual-layer model SNOBAL (Marks et al., 1998) resolve mass and energy exchanges within the ground-snow-atmosphere regime in a detailed way by depicting the layered structure of seasonal snowpacks. ~~(~~Echoing Langlois et al. (2009), these models will be termed "thermodynamic snow models" in the following.~~)~~ All of them need atmospheric variables as input, primarily precipitation, temperature, humidity, wind speed, and radiative fluxes. Also relatively simple thermodynamic models at least require temperature and/or precipitation (e.g., De Michele et al., 2013) or climatological means thereof (Hill et al., 2019). Avanzi et al. (2015) provide a good review. ~~Estimations of absolute model errors for SWE are rather scarce. However, it might be at the order of 10 to $15\,\mathrm{kg\,m^{-2}}$ (Langlois et al., 2009), but probably sometimes significantly more (Vionnet et al., 2012). See Sect. 3.1 and Table 4 for more details.~~[5] ~~All thermodynamic snow models need meteorological input data like temperature or radiation.~~ Unfortunately, many valuable longterm HS series ~~– which are so valuable for a variety of applications (see above) –~~ do not involve~~come along with~~ these data, and parameterizing or downscaling them from other sources in turn is susceptible to errors~~isn't straightforward either~~. ~~Even markedly simpler models at least require temperature and precipitation measurements as inputs (e.g., De Michele et al. (2013) or – in a very recent work by Hill et al. (2019) – climatological means thereof.~~ Thermodynamic snow models are not~~Consequently, no thermodynamic snow model is~~ applicable to derive SWE

[5] Moved to Discussion and Outlook section (Sect. 4.) NICHT VERGESSEN

exclusively from HS. (Avanzi et al. (2015) provide a good review and also introduce models including "statistical descriptions" which, in turn, need SWE measurements as input and, therefore, are not further addressed here. In this study these are counted as "thermodynamic snow models" as well.)

On the other side of the SWE modeling spectrum there are those models which – aside HS – only depend on date d (Pistocchi, 2016), d and altitude z (Gruber, 2014, see statistical approach therein), d and regional parameters (e.g., Mizukami and Perica, 2008; Guyennon et al., 2019) or d, z and regional parameters (e.g., Jonas et al., 2009)(e.g., Jonas et al. (2009); and applications thereof e.g., by Achleitner and Schöber (2017), who fitted the parameters to Austrian data). Again, Avanzi et al. (2015) provide a thorough listing of those models, which will be termed "empirical regression models" (ERMs)abbreviated as ERMs (for "empirical regression models") in the following. ERMs very much rely on the strong, near-linear dependence between HS and SWE (cf., e.g., Jonas et al., 2009). According to Gruber (2014) and Valt et al. (2018) HS describes 81% and 85% of SWE variance, respectivelymore than 80% of SWE variance (81% and 85%, respectively). This behavior bases on the narrow range within which the majority of bulk snow densities is found, and it leads to the well-known characteristic of HS–SWE–ρ_b datasets: log-normally distributed HS and SWE as well as normally distributed ρ_b (e.g., Sturm et al., 2010). Unfortunately, ERMs cannot adequately model (unchanged) SWE during periods with snow densification only due to metamorphism and deformationdeformation strain[6] (Jordan et al., 2010) but without (significant) mass loss. Sturm and Holmgren (1998) already state that HS and "load" (or SWE, respectively) play a more limited role in determining the compaction behavior in seasonal snow than "grain and bond characteristics" and temperature.

Interestingly, in most ERMs absolute, single-day HS observations are the only snow characteristics used. Depending on calibration focus they either adequately model mean-SWE or SWE_pk, mid winter or spring, etc. This is an inherent fact due to their model architecture. Those calibrated for good estimates of mean-SWE(necessarily) fail to model SWE_pk sufficiently well, those designed for SWE_pk often give bad SWE results during phases with shallow snowpacks. Typically, they simulate unrealistic mass losses during phases with compaction only by metamorphism and deformation, and the timing of SWE_pk as well as the duration of high snow loads cannot be modeled well. As it is honestly stated by Jonas et al. (2009) – the authors of one of the most influential ERMs – those models cannot be used to "convert time series of HS into SWE at daily resolution or higher" because they may "feature an incorrect fine structure in the temporal course of SWE". Therefore, ERMs are not suitable to calculate SWE for individual days.(Still, well made for means.)

McCreight and Small (2014) go an interesting step further and not only use single-day HS values for their regression model, but also the "evolution" of daily HS. They make use of the negative correlation of HS and ρ_b at short timescales (10 days) and their positive/negative correlation at longer timescales (3 months) during accumulation/ablation phases. This promising step of development is limited by the fact that the model parameters can only be estimated through regressions relying on at least three training datasets of HS and ρ_b (or SWE) from nearby stations. (McCreight and Small (2014) used ultrasonic HS measurements in conjunction with SWE pillow measurements.) Unfortunately, this disqualifies the model of McCreight and Small (2014) for assigning SWE to longterm and historical HS series as consecutive SWE measurements are not available for those.

[6] "Deformation strain", literally termed as such by Jordan et al. (2010), is called "deformation" throughout the revised manuscript.

An alternative approach that links HS and SWE throughout a snowy season without the need of further meteorological input is provided by Martinec (1977) and revisited by Martinec and Rango (1991). In some respect their semi-empirical model ~~Searching for alternative approaches that link HS and SWE throughout a snowy season without the need of meteorological input like temperature one might find Martinec and Rango(1991)'s work which provides a semi-empirical model that – in some respect –~~ bridges the gap between thermodynamic models ~~(needing lots of meteorological input)~~ and ERMs ~~(being "overregulated" by snow depth)~~. They use a method already developed by Martinec (1956) "to compute the water equivalent from daily total depths of the seasonal snow cover". Snow compaction is expressed as a time-dependent power function~~, and they end up with the following equation for each~~. Each layer's snow density ρ_n after n days is given by $\rho_n = \rho_0 \cdot (n+1)^{0.3}$,[7] where ρ_0 is the initial density of the snow layer. A fixed exponent of 0.3 is used, without going into detail. Martinec and Rango (1991) set ρ_0 to $100\,\mathrm{kg\,m^{-3}}$, Martinec (1977) varied it from 80 to $120\,\mathrm{kg\,m^{-3}}$. This computation is meant to give good results for the seasonal maximum snow water equivalent (SWE_{pk}). It is shown, the older the snow the less important is the correct choice of the crucial parameter ρ_0 (Martinec, 1977; Martinec and Rango, 1991).~~Martinec and Rango (1991) use a constant ρ_0 of $100\,\mathrm{kg\,m^{-3}}$ and a fixed exponent of 0.3, without going into detail how these values were found. They show that the error made by a bad choice of ρ_0 is rapidly decreasing with n and therefore the power function gives robust results at least for old(er) snow layers. They claim (without further explanation) that snow "luckily" does not settle according to an exponential curve, and show that in that case the error of ρ_0 would be independent from n and would not diminish while the snow layers are aging.~~ Their model interprets "each increase of total snow depth [. . .] as snow fall" and if "~~"[if]~~ the total snow depth remains higher than the settling by [the power function]~~[Eq. (2) of the article in hand]~~, this is also interpreted as new snow. If the snow depth drops lower than the value of the superimposed settling curve of the respective snow layers, it is interpreted as snowmelt, and a corresponding water equivalent is subtracted. In this way the water equivalent of the snow cover can be continuously simulated [. . .]" (Martinec and Rango, 1991). Rohrer and Braun (1994) improved this model particularly for the ablation season by further increasing density whenever melt conditions are modelled and by introducing a maximum possible snow density of $450\,\mathrm{kg\,m^{-3}}$.~~Martinec and Rango (1991) get promising results using this simple but robust snow compaction law. They only need daily snow depths as input and end up with a modeled SWE record at daily resolution.~~

Table 1 summarizes the classification of SWE models with respect to their essential input.[8]

1.3 Motivation for a New Approach

The question evolves, whether those~~such a~~ semi-empirical, layer-resolving snow model~~s like the rather old one of Martinec and Rango (1991)~~ can be improved and modernized, in order to provide an up-to-date snow model standard~~ somewhere~~ between sophisticated, thermodynamic models and modest~~, purely-statistical HS–SWE–ρ_b models, like~~ ERMs. Looking at the ease of Martinec (1977)'s and Rohrer and Braun (1994)'s approaches requiring only regular HS as input (see Table 1)~~Martinec and Rango (1991)'s approach~~, thinking about modern computational possibilities, and given the introductorily described strong need for an implementable method~~"something handy"~~, it seems interesting that there are no recent publications on this topic.

[7] Equation 2 of the submitted manuscript is not numbered in the revised manuscript.

[8] Table 1 was added during the revision. It was not part of the primarily submitted manuscript.

This leads to a possible null hypothesis of the paper-in-hand: “It is not possible to better model snow water equivalents than empirical regression models do, by exclusively using snow depths and their temporal changes as input.” This statement can be rejected after the following presentation of a new way to assign SWE values to the huge number of longterm, historical, and high-quality HS records of daily resolution. It follows Martinec and Rango (1991)’s key feature of considering daily *change* of snow depth as a proxy for the various processes altering bulk snow density ρ_b and snow water equivalent SWE, but further

In the following, an advancement of semi-empirical SWE models is presented, which maintains their key feature of considering daily *change* of snow depth as a proxy for the various processes altering bulk snow density and snow water equivalent, but further

- bases its (dry) snow densification function on Newtonian viscosity,
- provides a way to deal with small discrepancies between model and observation (in the order of HS measurement errors),
- takes into account unsteady compaction of underlying, older snow layers due to overburden snow loads, and
- densifies snow layers from top to bottom during melting phases without automatically modeling mass loss due to runoff. ;[9]

The ideas for the latter three features were already developed by Gruber (2014), but not suitably realized. TheThis new modeling approach is named ΔSNOW.MODEL and an easy-to-use R-package is available through https://cran.r-project.org/package=nixmass. The package is called *nixmass*, and it not only involves the ΔSNOW.MODEL, but also other models that use snow depth (*nix*. . . Latin for “snow”) to simulate SWE (i.e., snow *mass*).

The way how physical processes are coded in the ΔSNOW.MODEL is thoroughly described in theThe ΔSNOW.MODEL neither gives any new crucial insights in snow physics nor involves substantially new approaches. Still, the ΔSNOW.MODEL “rearranges” existing components in a physically consistent way and – as a whole – represents a new *method*. That is why it is described in the following Method section of this publication (Sect. 2). The calibration is outlined in Sect. 2 as well. Results, like best parameter choices and validation of the model output when compared to measurements, are given in Sect. 3. Section 4 provides an application of the ΔSNOW.MODEL for spatially modeling extreme snow loads in Austria. Sections 5 and 6 discuss possible future developments and provide concluding remarks.[10] In Sect. 4 model sensitivity, open questions, and possible future developments are discussed and Sect. 5 provides concluding remarks.

2 Method

As a successor of Martinec and Rango(1991)’s model the ΔSNOW.MODEL also builds on a semi-empirical approach and, therefore, can be regarded as standing between thermodynamic and empirical regression models. Its basic version was already presented by (Gruber, 2014, chapter 4) and a revision was presented by Gruber et al. (2018), but the ΔSNOW.MODEL described

[9] The last point, addressing the possible implementation of rain-on-snow, was omitted during the revision.

[10] Section 4 was moved to the appendix. In the revised paper the Discussion and Outlook section is numbered 4, Sect. 5 is the Conclusions section.

~~in the present article experienced a significant updating and recoding. It is designed for seasonal snowpacks and not intended for (multiannual) firn. The ΔSNOW.MODEL does not need any further input but a gapless snow depth record. It uses changes in observed snow depth to simulate a record of SWE. A scheme of the ΔSNOW.MODEL's principle is shown in Fig. 1.~~

Snow compacts over time due to various processes. Jordan et al. (2010) categorize them in snow drift, dry and wet metamorphism, and deformation. The ΔSNOW.MODEL cannot deal with snow drift, however, it differentiates between the latter processes:~~two processes.~~ Dry metamorphism is mainly processed in the *Dry Compaction module*[11] of the ΔSNOW.MODEL, in case of a significant increase in snow depth also in the *Overburden submodule* of the *New Snow module*. Wet metamorphism is treated in the *Drenching module*.~~and the Metamorphism and deformation are processed in the three modules *Dry Metamorphism* (2.1), *New Snow* (2.2.1), and *Drenching* (2.2.3).~~ The fourth module of the ΔSNOW.MODEL, the *Scaling module*~~, *Scaling* (2.2.2)~~, accounts for small discrepancies between model and observations. Table 2[12] correlates Jordan et al. (2010)'s compaction processes with the ΔSNOW.MODEL modules and outlines the processes that are ignored. The specific modules are described in Sects. 2.1 and 2.2, a scheme of the model principle is shown in Fig. 1.

2.0.1 Preliminary: The First Snow Layer

For non-zero snow depth observations ($HS_{\text{obs}} > 0$) after a snow-free period the ΔSNOW.MODEL assigns the following features to the model snowpack: There is one snow layer (layer counter $ly = 1$) and the age of this layer is set to $age = 1$. Thickness ~~Snow height~~[13] of this model layer (hs) and total model snow depth (HS) are equal, and set to observed snow depth: $hs = HS := HS_{\text{obs}}$. Analogously, the layer's snow water equivalent equals total snow water equivalent: $swe = SWE := \rho_0 \cdot HS_{\text{obs}}$, with new snow density ρ_0 being an important parameter of the ΔSNOW.MODEL (cf. Sect. 3). The treatment of the first snow event is illustrated at $t = 2$ in Fig. 1~~; it is processed within the New Snow Module (Sect. 2.2.1)~~.

2.1 Dry Compaction module~~Metamorphism~~

As it was mentioned in the Introduction, Martinec and Rango (1991) used a power function ~~(Eq. (2))~~ to describe densification of aging snow, because this way errors in initial density ρ_0 get less relevant over time. For the ΔSNOW.MODEL this kind of high error tolerance of ρ_0 is a rather feeble argument to use a power law, since it only holds for old snow and deep snowpacks, but with the ΔSNOW.MODEL also SWE of ephemeral snowpacks (e.g., at low elevation sites) should be modeled as good as possible. Furthermore, as the ΔSNOW.MODEL considers overburden load in a particular way (Sect. 2.2.1), it is not expedient to have a direct dependence between density and age of a layer. Aside from that drawbacks of a~~Martinec and Rango (1991)'s~~ power law compaction ~~(~~and in contrast to Martinec and Rango (1991)'~~s~~~~their~~ unproven claim "snow would [not] settle [. . .] according to an exponential curve"~~)~~, most modern snow models~~ very well~~ simulate snow compaction by way of Newtonian viscosity with associated exponential densification over time (e.g., Jordan et al., 2010). In the ΔSNOW.MODEL's *Dry*

[11] The *Dry Metamorphism module* was renamed to *Dry Compaction module* during the revision.

[12] Table 2 was added during the revision. It was not part of the primarily submitted manuscript.

[13] Changed throughout the paper according to Fierz et al. (2009) as a consequence of changing snow "height" to snow "depth" (see above). The respective symbol h was changed to hs.

Compaction module the densifying effects of dry metamorphism and deformation are combinedThe ΔSNOW.MODEL's Dry Metamorphism Module combines the effects of dry metamorphism and deformation, by applying the following adaption of Sturm and Holmgren (1998)'s relation, with the help of De Michele et al. (2013). (For wet metamorphism – as defined by Jordan et al. (2010) – see *Drenching module* in Sect. 2.2.3.)

$$\frac{hs(i,t-1)}{hs(i,t)} = 1 + \Delta t \cdot \frac{\widehat{\sigma}(i,t)}{\eta(i,t)}$$
$$\text{with } \widehat{\sigma}(i,t) = g \cdot \sum_{\hat{i}=i}^{ly(t)} swe(\hat{i},t) \tag{2}$$
$$\text{and } \eta(i,t) = \eta_0 \cdot e^{k\cdot\rho(i,t)}.$$

Model timestep Δt in general is arbitrary, but usually it is one day. If so, t can be explained as "today" and $t-1$ as "yesterday" here. Accordingly, $hs(i,t)$ is today's modeled thickness of the i-th snow layer. Snow layers are counted from bottom to top; layer $i = 1$ is the lowest and oldest layer. Today's depth of the total snowpack is $HS(t) = \sum_i hs(i,t)$.

The individual snow water equivalents of the layers are given by $swe(i,t)$, and their sum represents total mass of the snowpack $SWE(t) = \sum_i swe(i,t)$. The vertical stress at the bottom of layer i is given by $\widehat{\sigma}(i,t)$ (De Michele et al., 2013). It is constituted by the sum of loads overlying layer i (including layer i's own load), with $ly(t)$ being today's total number of snow layers or – in other words – $ly(t)$ is the index i of today's uppermost (i.e., "surface") layer.

The Newtonian viscosity of snow η is made density-dependent in the framework of the ΔSNOW.MODEL (following Kojima, 1967), but dependencies on temperature, grain characteristics etc. are very consciously ignored – due to the lack of information on it when dealing with pure snow depth data. Today's density of layer i is $\rho(i,t)$; it equals $\frac{swe(i,t)}{hs(i,t)}$. k and η_0 are tuning parameters of the *Dry Compaction module*Dry Metamorphism Module (see Sect. 3).

To avoid excessive compaction a crucial parameter is introduced in the ΔSNOW.MODEL, as it was already done by Rohrer and Braun (1994): ρ_{max}. It defines the maximal possible density of a snow layer and, (consequently,) also the maximum bulk snow density. Rohrer and Braun (1994) set ρ_{max} to $450\,\text{kg}\,\text{m}^{-3}$; finding its optimal value for the ΔSNOW.MODEL Finding its optimal value is subject to calibration (Sect. 2.3). ρ_{max} figures the density a snow layer (or the whole snowpack) can reach at most, unless it looses mass by melting. ρ_{max}, of course, is a model parameter and cannot be observed in real snowpacks. In case the *Dry Compaction module*Dry Metamorphism Module increases the density of one or more layers beyond ρ_{max}, $\rho(i,t)$ of the respective layer(s) is set equal to ρ_{max}.

According to Eq. (2) the rate of densification of a certain snow layer is linearly depending on the overlying snow load $\widehat{\sigma}(i,t)$ and exponentially depending on the layer's density $\rho(i,t)$. Sturm and Holmgren (1998) conclude that this difference is one reason why "snow load plays a more limited role in determining the compaction behavior than grain and bond characteristics and temperature". Nonetheless, higher overloads result in stronger compaction. Denser and older layers, respectively, compact less than newer layers with lower densities. This links the densification rate to the layer age, but indirectly by the use of density, and not directly. Therefore, ΔSNOW.MODEL's compaction is not directly depending on layer age as it was the case withif using

Martinec and Rango (1991)'s power law approach.~~The usage of an exponential function for compaction is one major difference between the ΔSNOW.MODEL and Gruber (2014), who uses a power law approach similar to Martinec and Rango (1991).~~

The *Dry Compaction module*~~Dry Metamorphism Module~~ of the ΔSNOW.MODEL is illustrated by the light blue arrows in Fig. 1. This module is applied at every point in time (except if there is no snow; see $t = 1$ in Fig. 1). The *Dry Compaction module*~~Dry Metamorphism Module~~ is the core~~"highest-ranking"~~ module because based on its result the ΔSNOW.MODEL decides between three different processes, realized by the other three modules:

2.2 Process Decisions

At every point in time, after the *Dry Compaction module* was run, ~~The ΔSNOW.MODEL simulates the layers' thicknesses for the next point in time $hs(i,t)$ using Eq. (2) (Dry Metamorphism Module). As one time step typically equals one day, figuratively spoken the Dry Metamorphism Module acts "over night", from "yesterday" to "today". "Today"~~ observed $HS_{\mathrm{obs}}(t)$ and modeled $HS(t)$ are compared. The ΔSNOW.MODEL's process decision algorithm now takes the result of the difference $\Delta HS(t) = HS_{\mathrm{obs}}(t) - HS(t)$ and confronts it with τ [m]. τ is another tuning parameter of the ΔSNOW.MODEL (see Sect. 2.3). ~~Its value is in the order of a few centimeters (see Sect. 3) since τ could also be regarded as a measure for observational error.~~ Technically, τ is a threshold deviation and defines a limit of $\Delta HS(t)$ whose overshooting, adherence, and undershooting heads for one out of the modules described in the following Sects. 2.2.1 to 2.2.3. Table 2 links them to snow physics.~~three branches, which – together with the Dry Metamorphism Module – build the four modules of the ΔSNOW.MODEL:~~

- ~~New Snow (2.2.1): In case observed snow depth is significantly higher than modeled snow depth ($\Delta HS(t) > +\tau$), a snowfall event is assumed. This means mass gain as well as enhanced compaction of underlying layers due to the overburden load.~~
- ~~Scaling (2.2.2): In case there is no significant difference between observed and modeled snow depth ($-\tau \leq \Delta HS(t) \leq +\tau$), neither mass gain nor loss is modeled. However, modeled snowpack is "scaled", i.e. compressed or stretched, to fulfill the condition $HS(t) = HS_{\mathrm{obs}}(t)$.~~
- ~~Drenching (2.2.3): In case observed snow depth is significantly lower than modeled snow depth ($\Delta HS(t) < -\tau$), it is interpreted as wet snow metamorphism. In the snowpack this "drenching" happens from top to bottom, resulting in the associated (strong) decline of snow depth. The drenching can either be caused by melt (mass loss) or rain (mass gain), whereas treating the latter is optional, not finalized in the current version of the ΔSNOW.MODEL, and not further detailed here.~~

2.2.1 New Snow module

In case $\Delta HS(t) > +\tau$, meaning observed snow depth is significantly higher than modeled snow depth, a new snow event is supposed and a new top snow layer is modeled by the ΔSNOW.MODEL (see at $t = 2$ and $t = 7$ in Fig. 1~~for a schematic illustration~~). This is a consequential step and nothing innovative at all. Other models have implemented this mechanism as well (e.g., Martinec and Rango, 1991; Lehning et al., 1999; Sturm et al., 2010)~~(e.g., Martinec and Rango, 1991; Sturm et al., 2010~~. However, the ΔSNOW.MODEL goes beyond and introduces another feature: It explicitly models the peculiar effect of overburden load on underlying layers, defined as their enhanced densification due to ~~(sudden)~~stress, which is put on by the weight of new snow~~ (or rain-on-snow)~~. Grain bonds get broken, grains slide, partially melt, and warp (Jordan et al., 2010), and the layers densify comparatively rapidly and strongly. The ΔSNOW.MODEL interprets overburden load as an "unsteady and discontinuous" stress on the snowpack, under which snow presumably does not react as a viscous Newtonian fluid. ~~(~~As long as ~~Δt –~~ the time between two consecutive observations Δt ~~–~~ is in the order of at least some hours, discontinuity is an intrinsic feature of the process. ~~Mostly daily snow depths are available, a new snow event is always an unsteady case then.) The part of Jordan et al. (2010)'s "deformation strain", that is created by the individual layer loads, is interpreted as a "continuous" effect and is processed by the Dry Metamorphism Module; see $\widehat{\sigma}$ in Eq. (2).~~

The *New Snow module* realizes the effect of overburden load through the *Overburden submodule* by reducing each layer's thickness $hs(i,t)$ with the help of the dimensionless "overburden strain" $\epsilon(i,t)$, defined as

$$\epsilon(i,t) = c_{\text{ov}} \cdot \sigma_0 \cdot e^{-k_{\text{ov}} \frac{\rho(i,t)}{\rho_{\max} - \rho(i,t)}} \quad \text{with } \sigma_0 = \Delta HS(t) \cdot \rho_0 \cdot g. \tag{3}$$

c_{ov} $[\text{Pa}^{-1}]$ is another tuning parameter of the model (see Sect. 2.3) and~~. It~~ controls the importance of the unsteady ~~sudden, enhanced~~ compaction due to overburden load. According to Sturm and Holmgren (1998) and in consistency with Eq. (2) snow load has a linear effect on the bulk density. Therefore, $\epsilon(i,t)$ is made linearly depending on the load, which the overlying new snow is putting on the underlying layers. This load~~ (or stress or pressure)~~ is well approximated by σ_0 [Pa]; the bigger the overburden load, the stronger the compaction. (The overburden load does not fully equal σ_0, since $\Delta HS(t)$ is not the depth of the new snow, but the difference between modeled depth ~~–~~"before" knowing about the new snow event ~~–~~and observed depth "after" the new snow event. An iterative calculation would be more precise, however, Eq. (3) proved to be an adequate compromise between simplicity and accuracy.) In order to avoid $\epsilon(i,t) > 1$, c_{ov} is restricted ~~(~~at least~~)~~ to the range of values between 0 and the minimum value of the data record for $\frac{1}{\sigma_0}$. As σ_0 hardly ever exceeds $1000\,\text{Pa}$, $\frac{1}{\sigma_0}$ normally is larger than $1 \times 10^{-3}\,\text{Pa}^{-1}$. This value, thus, marks a good upper bound for c_{ov} (Sect. 2.3). Dimensionless k_{ov} controls the role of a certain snow layer's density~~ (i.e., age, respectively)~~ on $\epsilon(i,t)$, and has to be specified by calibration (see Sect. 2.3). The density-dependence of $\epsilon(i,t)$ was chosen to be exponential, and using $\rho_{\max}$ in the denominator of Eq. (3)'s exponent secures that overburden loads cannot make snow layers denser than $\rho_{\max}$. The closer a snow layer's density is to the maximum density $\rho_{\max}$, the less it will be compacted by additional load. Relatively new and, therefore,~~ – therefore –~~ not very dense layers~~,~~ are exposed to greater densification, which is exactly what is observed in reality. As it will be shown in sections 2.2.2 and 2.2.3 $\rho_{\max}$ also governs mass loss and melt in the model. Not least, $\rho_{\max}$ illustrates the possible maximum density of a wet seasonal snowpack in the ΔSNOW.MODEL~~-~~ world and~~ – as it can be seen in Sect. 3 –~~ it is possible to assign a reasonable value to it (cf. Sect. 3). ~~(Sturm et al. (2010), who revisited Sturm and Holmgred (1998), already introduced a maximum density for seasonal snow.~~

They used it very prominently in their formula for modeling bulk density and defined five (snow) climate classes with different values of $\rho_{\max}$ ranging from 217 to 598 $\mathrm{kg\,m^{-3}}$.)

The "overburden strain" $\epsilon(i,t)$ theoretically lies between 0 and 1 and compresses all (old) snow layers of the model in case of a new snow event. Practically, $\epsilon(i,t)$ is often close to zero (in this study 90% of all computed ϵ are smaller than 0.09) and extremely rarely higher than 0.3 (in this study only 9 out of 10000).

The following intermediate (asterisked) variables are defined due to the overburden load. The compressed layer's masses, $swe(i,t)$, remain unaffected during this process.

$$
\begin{aligned}
&\epsilon(i,t) = \frac{hs(i,t) - hs^*(i,t)}{hs(i,t)} \qquad \text{leading to} \qquad hs^*(i,t) = (1-\epsilon(i,t)) \cdot hs(i,t) \\
&HS^*(t) = \sum_i hs^*(i,t) \\
&\rho^*(i,t) = \frac{swe(i,t)}{hs^*(i,t)}
\end{aligned}
\tag{4}
$$

A new snow event, identified by the condition $\Delta HS(t) > +\tau$, of course not only impacts the older snow and compacts it more strongly, but it also adds a new snow layer and mass to the snowpack (pink arrow at $t=2$ and $t=7$ in Fig. 1). The number of layers is increased by one and the following attributes are given to the new layer:

$$
\begin{aligned}
&age(ly,t) = 1 \\
&hs(ly,t) = HS_{obs}(t) - HS^*(t) \\
&swe(ly,t) = hs(ly,t) \cdot \rho_0
\end{aligned}
\tag{5}
$$

The total snow water equivalent is risen: $SWE(i,t) = SWE(i,t-1) + swe(ly,t)$, and the intermediate variables of Eq. (4) overwrite their originals: $hs(i,t) = hs^*(i,t)$, $HS(t) = HS^*(t) + hs(ly,t)$, and $\rho(i,t) = \rho^*(i,t)$. The model-snowpack with this new properties now again compacts according to Eq. (2), time t is risen by one increment, and at the next point in time"tomorrow" the process again starts with the decision described in Sect. 2.2. The *Overburden submodule* treatment of overburden snow loads as triggers for enhanced compaction of the underlying snow is illustrated with a purple arrow at $t=7$ in Fig. 1. Although the overburden strain ϵ in most cases hardly deviates from zero (see above), the value of this feature for the performance of the ΔSNOW.MODEL is supposed to be rather high, at least it should be worth the effort (cf. Sect. 4).[14]

2.2.2 Scaling module

EOf course, equations (2) and (3) are highly simplified representations of the complex viscoelastic behavior of snow and make no claims of being particularly precise. Still, also snow depth observations typically only show an accuracy of a few centimeters. The ΔSNOW.MODEL accepts these inherent inaccuracies and apparent discrepancies between model and measurements and

[14] The discussion of this issue is provided in the Discussion and Outlook section of the revised manuscript. NICHT VERGESSEN!

copes with them by not applying too strict criteria in the process decisions described in Sect. 2.2. The threshold deviation uncertainty measure τ (introduced above) acts as a buffer to avoid too frequent gain or loss of mass in the model world: In case $|\Delta HS| \leq |\tau|$ neither the snowpack looses mass nor gains mass, but mass is kept constant. In order to benefit from having a new measurement at every point in time, $HS(t)$ is intentionally set to $HS_{\text{obs}}(t)$ by the *Scaling module*.

The *Scaling module* forces a partial revaluationreversal of the previous compaction, which was modeled by the *Dry Compaction module*Dry Metamorphism Module between $t-1$ and t. The best-fitted parameter setting for η_0 is temporarily rejected and substituted by η_0^*. It would be straight forward to use one adjusted $\eta_0^*(t)$ for all layers. However, this leads to a rational function with multiple solutions for $\eta_0^*(t)$, . Consequently, this approach shows a clear non-physical behavior making it necessary to calculate different $\eta_0^*(i,t)$ for each layer i. See Appendix B for details on that.

$\eta_0^*(i,t)$ is then used instead of η_0 in Eq. (2) to recalculate the compaction of individual layers. $HS(t)$ now equals $HS_{\text{obs}}(t)$. In most cases all layers get "slightly more" or "slightly less" compacted by the *Scaling module* than by the *Dry Compaction module*Dry Metamorphism Module. Only at rare occasions the scaling does not compact, but a small "stretching" of the snowpack is necessary. This only happens if there was a small increase in observed snow depth *and* very little modeled dry metamorphic compaction; the condition $HS(t) + \tau > HS_{\text{obs}}(t) > HS_{\text{obs}}(t-1)$ has to be fulfilled. Of course, such "stretching" does not occur in reality, but also in the ΔSNOW.MODEL it is an infrequent case that only acts at a small scale. In: in any case the "stretching" is smaller than τ. The issue is accepted as a model artifact, not least, because the "stretching" enables the very valuable adjustment to HS_{obs} at every point in time – without forcing mass gains for any insignificant HS raises that are within the measurement accuracy.

In case the density of an individual layer exceeds ρ_{max} by the scaling process, the excess mass is distributed layerwise from top to bottom. SWE remains constant during scaling, unless it would be necessary to compact all layers beyond ρ_{max}. In this case the appropriate excess mass is taken from the model- snowpack and interpreted as runoff, SWE is reduced and all layer thicknesses are cut accordingly (see *Runoff submodule* in Sect. 2.2.3 for details). As τ turns out to be – reasonably and preferably – chosen in the order of a few centimeters (Sect. 3), the resulting reduction of SWE within the *Scaling module* is always quite small: (e.g., with $\tau = 2\,\text{cm}$ and maximum density chosenis $450\,\text{kg}\,\text{m}^{-3}$ (like Rohrer and Braun, 1994) the mass loss due to runoff, i.e. runoff, is only $9\,\text{kg}\,\text{m}^{-2}$).

The *Scaling module* is illustrated as black arrows in Fig. 1. Note, again that the scaling is nothing "physical", but also nothing "substantial" in terms of SWE, yet it is a smart way to utilize the advantage of having a measured snow depth at every point in time.

2.2.3 Drenching module

The *Drenching module*, finally, defines compaction due to liquid water percolating from top to bottom through the snowpack, loosening grain bonds and leading to densification (wet snow metamorphism). In case observed snow depth at a certain point in time is significantly lower than modeled snow depth ($\Delta HS(t) < -\tau$), the *Drenching module* is activated. Drenching compaction is the ΔSNOW.MODEL's synonym of wet snow metamorphism.

~~The drenching can either be caused by melt or rain and the ΔSNOW.MODEL can principally deal with both processes, which are (often) contradictory in terms of mass change (melt: mass loss or invariant; rain: mass gain or – only if combined with runoff – invariant or mass loss). However, distinguishing between them is indeed extremely difficult (if not impossible) if only snow depths are available. For the time being~~T ~~t~~he ΔSNOW.MODEL ignores rain on snow since it concentrates on modeling SWE for pure snow depth records without having any further information on e.g.~~e.g~~ precipitation, temperature, snowfall level, etc.~~...~~ Possibilities how rain could be addressed in~~at~~ future developments are outlined in Sect. 4 MACHEN, VIELLEICHT MITHILFE DER GESTRICHENEN SÄTZE. ~~This drawback seems disappointing, however, given the relative success of the ΔSNOW.MODEL "without rain" (see Sect. 3) one should not expect too much improvement when incorporating rain in one or another – potentially elaborate – way.~~

To cope with the model-observation-~~model/observation~~ discrepancy $\Delta HS(t) < -\tau$ the *Drenching module* densifies the model layers until ρ_{max} is reached, ~~–~~ starting from the uppermost one. Figuratively spoken, ~~it wettens or drenches~~ a certain layer gets drenched until saturation and meltwater is further distributed ~~until it is "saturated" and further distributes the melt water~~ to the underlying layer. This process is repeated until (transient, therefore asterisked) HS^* equals $HS_{\text{obs}}(t)$. One or more ~~(upper)~~ layers might reach ρ_{max}. In case $\Delta HS(t)$ is so negative that all model snow layers ~~(from top to bottom)~~ are compacted and densified to ρ_{max}, but still $HS^* > HS_{\text{obs}}(t)$ the *Runoff submodule* is activated and runoff $R(t)$ is defined as~~, the product of the remaining depth difference and the maximum density constitutes mass loss, i.e. runoff~~ ~~$R(t)$~~:

$$R(t) = (HS^* - HS_{\text{obs}}(t)) \cdot \rho_{\text{max}}. \tag{6}$$

All layer thicknesses are "cut" by a respective portion: $(HS^* - HS_{\text{obs}}) \cdot \frac{hs_i^*}{HS^*}$. This mechanism does not reduce total number of layers, but layers potentially get very thin. During the melt season, where most of the runoff is produced, the *Runoff submodule*~~*Drenching module*~~ is more or less continuously active until ~~it is snow-free~~ $HS_{\text{obs}}(t) = 0$ and all the snow has been converted to runoff. For a~~one~~ distinct snowpack ~~–~~ from the first snow fall (t_1) until getting snow-free again (t_2) ~~–~~ one has $\sum_{t_1}^{t_2} R(t) = SWE_{\text{pk}}$.

In Fig. 1 the *Drenching module* is shown by the brown ~~arrow (as long as there is no mass loss)~~ and its *Runoff submodule* by the green arrows~~cyan arrow (in case runoff is modeled)~~.

2.3 Calibration

The ΔSNOW.MODEL has seven parameters that can be used for calibration: ρ_0, ρ_{max}, η_0, k, τ, c_{ov}, and k_{ov} (cf. Tab. 3). For the first four parameters one finds suggestions and ranges in the literature:

Sturm and Holmgren (1998) do not address the criticality for the choice of new snow density, however, they use constant $\rho_0 = 75\,\text{kg}\,\text{m}^{-3}$. It is a well known characteristic of new~~freshly fallen~~ snow to show large variations in densities. Helfricht et al. (2018) reviewed many studies and give a general range of $10 - 350\,\text{kg}\,\text{m}^{-3}$, narrowing it down to "mean values" between $70 - 110\,\text{kg}\,\text{m}^{-3}$. Note, that this is daily densities. Sub-daily means of new snow densities are lower. Helfricht et al.

(2018), for example, come up with an average of $68\,\mathrm{kg\,m^{-3}}$ for hourly time intervals. During the calibration process for the ΔSNOW.MODEL ρ_0 was varied from 50 to $200\,\mathrm{kg\,m^{-3}}$.

The second density-related calibration parameter is ρ_{max}, the maximum possible density within the model framework. As mentioned, Rohrer and Braun (1994)Sturm et al. (2010) already setdefined such a maximum at $450\,\mathrm{kg\,m^{-3}}$. Also Sturm et al. (2010) defined it for five different climate classes, ranging. They range from 217 to $598\,\mathrm{kg\,m^{-3}}$. Glaciologists set the "critical density" before snow turns into firn (which is wetted snow that has survived one summer) to 400 to $800\,\mathrm{kg\,m^{-3}}$ (e.g., Paterson, 1998). Still, manual density measurements of seasonal snow used in previous studies hardly ever exceeded $\rho_{\mathrm{b}} =$$500\,\mathrm{kg\,m^{-3}}$ (e.g., Jonas et al., 2009; Guyennon et al., 2019). Armstrong and Brun (2010) limit it to approximately 400 to $500\,\mathrm{kg\,m^{-3}}$ too. In order to find the fittest value for ρ_{max} used in the ΔSNOW.MODEL, it was varied from 300 to $600\,\mathrm{kg\,m^{-3}}$.

Equation (2) needs η_0, the "viscosity at [which] ρ equals zero" (Sturm and Holmgren, 1998). It is found to be in the order of $8.5\times10^6\,\mathrm{Pa\,s}$ (Sturm and Holmgren, 1998), $6\times10^6\,\mathrm{Pa\,s}$ (Jordan et al., 2010), and $7.62237\times10^6\,\mathrm{Pa\,s}$ (Vionnet et al., 2012). During the calibration process for the ΔSNOW.MODEL η_0 was varied from 1 to $20\times10^6\,\mathrm{Pa\,s}$. Parameter k, the second necessary parameter in Eq. (2), was varied from 0.011 to $0.08\,\mathrm{m^3\,kg^{-1}}$ by Sturm and Holmgren (1998) depending on climate region and respective different types of snow. However, they cite Keeler (1969) in their Table 2 with values for k for "Alpine-new" snow of up to $0.185\,\mathrm{m^3\,kg^{-1}}$. In more complex snow models k is set to $0.023\,\mathrm{m^3\,kg^{-1}}$ (see Crocus: b_η in Vionnet et al. (2012)'s Equation (7); and also in Equation (2.11) of Jordan et al., 2010) or $0.021\mathrm{m^3\,kg^{-1}}$ (see SNTHERM: Equation (29) in Jordan, 1991). Its range for the ΔSNOW.MODEL calibration was set from 0.01 to $0.2\,\mathrm{m^3\,kg^{-1}}$, which is quite generous.

There are no references for the latter three parameters. Threshold deviation τ, as mentioned, might be interpreted as a measure of observation error, is regarded to be in the order of a few centimeters, and was modified from $1\,\mathrm{cm}$ to $20\,\mathrm{cm}$ for calibration. The last two parameters, c_{ov} and k_{ov}, determine the role of overburden strain and are newly introducted in kind of a specialty of the ΔSNOW.MODEL: c_{ov} and k_{ov}. At least the limits of c_{ov} could be defined (Sect. 2.2.1) as $c_{\mathrm{ov}} \in \left[0, \min(\frac{1}{\sigma_0})\right]$. k_{ov} is only known to be a dimensionless,(dimensionless) real, positive number. For calibrating the ΔSNOW.MODEL c_{ov} and k_{ov} were restrained by $[0, 10^{-3}\,\mathrm{Pa^{-1}}]$ and $[0.01, 10]$, respectively.

As mentioned, timestep Δt principally can be chosen arbitrarily. Mostly it might be one day, because many (longterm) snow depth measurements are on a daily basis. The calibration performed in this study is based on $\Delta t = 1\,\mathrm{day}$. Still, longer Δt (e.g., three days) as well as shorter Δt (e.g., one hour) are conceivable and could be handled by the ΔSNOW.MODEL. Note, however, at least some(at least some) calibration parameters will change significantly when changing Δt. This gets obvious when thinking about new snow density ρ_0, which of course is different if defined for one hour or for a three day timestep. The usage of this publication's calibration parameters can, therefore, only be suggested for daily snow depth records.

2.3.1 Calibration Data and Method

The calibration process needs SWE data, but SWE measurements are quite rare (see Sect. 1)data – either from observations or from a much more sophisticated snow model, whose simulated SWEs are sufficiently reliable. As it was outlined in the Introduction, SWE measurements are quite rare. FurthermoreHowever, for calibration not only SWE observations are needed, but also regular snow depthgapless snow depths records from the same places, at least at daily resolution. Gruber (2014) collected 14~~15~~ years of weekly SWE data from six stations in the Eastern Alps, measured by the observers of the Hydrographic Service of Tyrol (Austria) between winters 1998/99 and 2011/12~~2012/13~~. The measurements of snow depth and water equivalent were made manually in snow pits with rulers and snow sampling cylinders ($500\,\mathrm{cm}^3$), respectively. The sites range from 590 m to 1650 m altitude and are situated in relatively dry, inneralpine regions as well as in the Northern and Southern Alps, which are more humid due to orographic enhancement of precipitation (see Gruber, 2014, for details). The sites in the Southern Alps even show a moderate maritime influence due to their vicinity to the Mediterranean Sea, the most important source of moisture for this region (e.g., Seibert et al., 2007). These $6 \times 14 = 84$~~$6 \times 15 = 90$~~ winter seasons cover 1166 measured HS–SWE pairs. Besides these SWE measurements manual HS measurements are available for every day at the respective stations. Figure A1 and Table A1 provide a map and a list, respectively.[15]

The second source for SWE measurements used for calibration is Marty (2017). The Swiss SLF freely provides biweekly SWE and daily HS data from 11 stations in Switzerland~~(mostly in the Northern Alps, some inneralpine) spanning an altitude range from 1200 m to 2540 m~~. The ~~biweekly~~ HS measurements, accompanying~~(corresponding to~~ the biweekly SWE measurements,~~)~~ were compared with the contemporary value of the daily HS records. Only those sites and years were used for calibration where the respective values of the daily HS record match the values of the biweekly measurements. If this condition is fulfilled, it is supposed that SWE and HS measurements fit together sufficiently well, although they unfortunately cannot always be taken exactly at the same place, which introduces uncertainty (e.g., López-Moreno et al., 2020). Consequently, 9 stations were used, most of them in the Northern Alps, some inneralpine, spanning an altitude range from 1200 m to 1780 m ~~This was the case for 9 stations~~, with all in all 56 winters and 388~~363~~ pairs of HS and SWE measurements. Details are given in Fig. A1 and Table A1. ~~(~~Other stations and years suffer from discrepancies caused by too far distances between the measurements etc.~~)~~

In order to ensure an unperturbed validation, the observation data sets from Austria and Switzerland (1554~~1529~~ $SWE - HS$ pairs) were split in two almost equally big parts, one for model calibration (SWE_{cal}) and one for validation (SWE_{val}).~~into even years for model calibration (SWE_{cal}) and odd years for validation (SWE_{val}).~~[16] The two data sources (Gruber, 2014; Marty, 2017) do not address the accuracy of the manual SWE observations. Mostly, SWE measurements made with snow sampling cylinders are used as references in comparison studies, without addressing *their* accuracy (e.g., Sturm et al., 2010; Dixon and Boon, 2012; Kinar and Pomeroy, 2015; Leppänen et al., 2018). López-Moreno et al. (2020) provide a reported range of 3-13%, and condense the results of their own, very thorough and valuable experiments to an error range of 10-15% for bulk snow density. The majority of SWE_{cal} and SWE_{val} comes from the Hydrographic Service of Tyrol, Austria, where snow sampling cylinders ($500\,\mathrm{cm}^3$) are used (Sect. 2.3.1). The repeatability of this kind of measurement is estimated at $\pm 4\%$ for glacier mass balance studies (R. Prinz, Univ. of Innsbruck, Austria; pers. comm.). Roughly interpreting these density measurement "variabilities" as relative observation errors for SWE, the results for absolute accuracy would typically spread across the wide range of about 2 to $50\,\mathrm{kg\,m^{-2}}$.

[15] Figure A1 and Table A1 were added during the revision. They were not part of the primarily submitted manuscript.

[16] The following remarks on SWE observation accuracy were placed at the beginning of the Validation and Illustration section of the initially submitted manuscript and moved to this place during revision.

Model calibration was performed with the statistical software *R* (R Core Team, 2019) and the *R* package *optimx* (Nash, 2014). Results were obtained with optimization methods *L-BFGS-B* (Byrd et al., 1995) followed by *bobyqa* (Powell, 2009), which both are able to handle lower and upper bounds constraints. The function to be minimized was the root mean square error (RMSE) of SWEs from the ΔSNOW.MODEL and observed SWEs, using the calibration data set SWE_{cal}.

3 Results

The following evaluates the ability of the ΔSNOW.MODEL to calculate snow water equivalents exclusively from snow depths, and its practicability.[17] ~~See sections 3.2 and 4 for more. Before that, however, the model parameters have to be optimized during the calibration process (which was described in Sect. 2.3). The results thereof – the best-fitted parameters – are important model results per se, since they show insights how well model world and real world fit together.~~

3.1 ~~Optimized Parameters and Sensitivities~~

Table 3 gives an overview of all parameters and summarizes the optimal setting for the ΔSNOW.MODEL. A discussion of the best-fitted values and of the model sensitivity to parameter changes can be found in Sect. 4.

The minimal RMSE between all SWE observations used for calibration (SWE_{cal}) and the respective modeled values is $30.1\,\text{kg}\,\text{m}^{-2}$. It is reached for new snow density $\rho_0 = 81\,\text{kg}\,\text{m}^{-3}$, maximum density $\rho_{\text{max}} = 401\,\text{kg}\,\text{m}^{-3}$, "viscosity parameters" $\eta_0 = 8.5 \times 10^6\,\text{Pa}\,\text{s}$ and $k = 0.030\,\text{m}^3\,\text{kg}^{-1}$, threshold deviation $\tau = 2.4\,\text{cm}$, and "overburden parameters" $c_{\text{ov}} = 5.1 \times 10^{-4}\,\text{Pa}^{-1}$ and $k_{\text{ov}} = 0.38$.[18] ~~A graphical analysis of the model sensitivity to parameter changes is shown in Fig. 2. SWE_{pk} was chosen to indicate sensitivity, since it is probably the most important quantity for snow-hydrology and other fields where snow mass is a key variable (see Introduction).~~

3.1.1 ~~New Snow Density ρ_0~~

~~Being aware of both – the huge possible variations of new snow density ρ_0 depending on meteorological conditions during snowfalls and the possible cruciality of this parameter for SWE simulation by the ΔSNOW.MODEL – ρ_0 was chosen to be a constant in the framework of the model. For $\rho_0 = 81\,\text{kg}\,\text{m}^{-3}$ the minimal root mean square differences/errors (RMSE) between all SWE observations used for calibration (SWE_{cal}) and the respective modeled values was reached. This value clearly lies within the broader frame of possible new snow densities and quite closely to Sturm and Holmgren (1998)'s $75\,\text{kg}\,\text{m}^{-3}$ (The ΔSNOW.MODEL could be seen as an extended combination of the Sturm and Holmgren (1998) and Martinec and Rango (1991) approaches.), but it is found in the lower part for "typical" new snow densities (e.g., Helfricht, 2018). A possible explanation could be that the SWE measurement records used for the calibration tend to underrepresent late winter and spring conditions. Regular (weekly, biweekly) observations capture the short melt seasons worse than the (much) longer accumulation phases.~~

[17] Most content of subsections 3.1.1 to 3.1.5 of the submitted manuscript was transferred to the Discussion section during revision. The segmentation 3.1.1-3.1.5 was omitted.

[18] All the optimized values (except ρ_{max}) where rounded to two significant figures for the revised manuscript.

Therefore, SWE records might be biased towards early and mid winter new snow densities, which are lower (e.g., Jonas et al., 2009). Still, there are also some indications that using e.g., $100\,\mathrm{kg\,m^{-3}}$ as constant new snow density when modeling SWE results in an overestimation of precipitation (up to 30% according to Mair et al., 2016). The calibrated value for ρ_0 can be regarded as a reasonable result, even more when only considering it as a model parameter but not as a physical constant.

The sensitivity analysis illustrated in Fig. 2 confirms the importance of a good choice of ρ_0. Increasing ρ_0 quite fast leads to a decrease of the relative bias of seasonal SWE maxima (SWE_{pk}). (Note the definition of the relative bias in Fig. 2's caption.) In absolute values: too small ρ_0 cause too small SWE_{pk}, using higher values leads to an overestimation of SWE_{pk}. This behavior supports above-mentioned tendency to overestimate precipitation when choosing $100\,\mathrm{kg\,m^{-3}}$ as new snow density. As expected, the new snow density is the most crucial parameter of the ΔSNOW.MODEL (cf. Tab 3). The median relative bias of SWE_{pk} changes by -0.46% per $+1\,\mathrm{kg\,m^{-3}}$, if the whole calibration range of ρ_0 is considered to calculate the sensitivity ($50-200\,\mathrm{kg\,m^{-3}}$). This means a median change in SWE_{pk} of $+0.37\,\mathrm{kg\,m^{-2}}$ when ρ_0 is risen by $+1\,\mathrm{kg\,m^{-3}}$. If the limits are chosen tighter around the optimal value, the gradient is even steeper: -0.62% and $+0.50\,\mathrm{kg\,m^{-2}}$ per $+1\,\mathrm{kg\,m^{-3}}$, respectively, when the gradient is approximated for the range $70-90\,\mathrm{kg\,m^{-3}}$. Widely-used $\rho_0 = 100\,\mathrm{kg\,m^{-3}}$, consequently, causes a median overestimation of SWE_{pk} of about 12% in the ΔSNOW.MODEL (same for daily SWE). Users should be aware of this. The suggestion clearly is to either use the best-fitted parameters of this study or recalibrate *all* parameters (with appropriate SWE data), but not adjusting only single parameters. As the calibration data of this study are spread across various climates and altitudes, users can be quite confident to get good results if using $\rho_0 = 81\,\mathrm{kg\,m^{-3}}$. This value seems to be a good compromise – at least at alpine areas. However, for (very) maritime, very dry, polar or tundra regions the optimized ρ_0 should be used with caution; if possible, recalibration is recommended.

3.1.2 Maximum Density ρ_{max}

Of course, the maximum bulk snow density of a snowpack changes from year to year and site to site. For the ΔSNOW.MODEL simplicity and independence from meteorological variables outweigh precision. Even more so, when there are good arguments for the existence of a "typical" maximum bulk density ρ_{max}. Put simply, not too old seasonal and also ephemeral snowpacks melt away when they get water saturated. Before that, there is limited time for dry densification; dry winter snow's bulk density is widely described as staying below about $350\,\mathrm{kg\,m^{-3}}$ (e.g., Paterson, 1998; Sandells et al., 2012). Accounting for the fact that volumetric liquid water content of about 10% marks the funicular mode of liquid distribution in old, coarse-grained snow (Denoth, 1982; Mitterer et al., 2011), this leads to the rough estimate of a typical maximum bulk density of about $\frac{9}{10}\cdot 350 + \frac{1}{10}\cdot 1000 = 415\,\mathrm{kg\,m^{-3}}$. Convincingly, the fittest value for ρ_{max} in the ΔSNOW.MODEL turns out to be $401\,\mathrm{kg\,m^{-3}}$, which is close to that value and well situated within the range given in the literature (Table 3). Moreover this is virtually the same value like the median maximum seasonal density of the SWE_{val} data records ($400\,\mathrm{kg\,m^{-3}}$, see box plot in Fig. 3) – another indication why ρ_{max} could be regarded as a typical seasonal maximum of ρ_{b}.

Figure 2 illustrates the similarity between ρ_0 and ρ_{max} regarding their influence on SWE simulations. Keeping the other six ΔSNOW.MODEL parameters constant but increasing ρ_{max} leads to increased SWE_{pk} and vice versa. This is not surprising, however reasonable. The ΔSNOW.MODEL is not as sensitive to changes in ρ_{max} than to changes in ρ_0: Raising ρ_{max} by

$+1\,\mathrm{kg\,m^{-3}}$ leads to a mean decrease of the relative bias of SWE_{pk} of -0.06%, which corresponds to an increase in absolute SWE_{pk} of $+0.24\,\mathrm{kg\,m^{-2}}$ per $+1\,\mathrm{kg\,m^{-3}}$. The same argumentation like for ρ_0 in Sect. 3.1.1 lets users of the ΔSNOW.MODEL be quite sure when taking $\rho_{\mathrm{max}} = 401\,\mathrm{kg\,m^{-3}}$, the best-fitted value according to this study's calibration. Be aware that solely changing parameter ρ_{max} for an application of the ΔSNOW.MODEL elsewhere, without proper recalibration of the other parameters, might lead to significant changes in the results for SWE. For the data set of this study, e.g., using $\rho_{\mathrm{max}} = 500\,\mathrm{kg\,m^{-3}}$ (instead of $401\,\mathrm{kg\,m^{-3}}$) would lead to a median overestimation of SWE_{pk} of about $24\,\mathrm{kg\,m^{-2}}$.

3.1.3 Viscosity Parameters η_0 and k

Equation (2) represents the settlement and densification function of the ΔSNOW.MODEL. Viscosity of layer i at time t is defined as $\eta(i,t) = \eta_0 \cdot e^{k \cdot \rho(i,t)}$. Two parameters η_0 and k act as adjustment screws and have to be calibrated. Literature places η_0 in the order of some $10^6\,\mathrm{Pa\,s}$, k is supposed to be in the range of 0.01 to $0.2\,\mathrm{m^3\,kg^{-1}}$ (Table 3). The latter was varied over the mentioned range during the calibration process and – satisfactorily – it's optimized value is $k = 0.030\,\mathrm{m^3\,kg^{-1}}$, which is very close to the ones used in widely accepted, sophisticated, thermodynamic models (Jordan et al. (2010), Vionnet et al. (2012); see also Sect. 2.3). The range over which η_0 was varied during the calibration was 1 to $20 \times 10^6\,\mathrm{Pa\,s}$. Best-fitted value is $8.5 \times 10^6\,\mathrm{Pa\,s}$, pretty close to other studies' values (Table 3).

As far as the ΔSNOW.MODEL's sensitivity to changes in the "viscosity parameters" η_0 and k is concerned, Fig. 2 shows that an isolated rise of the model snow viscosity – either by enhancing η_0 or k – increases the relative bias of SWE_{pk}, which means a decrease in absolute values of SWE_{pk}. This behavior is consistent, since higher viscosity reduces the densification rate and the model snowpack tendentially stays deeper. Consequently, increases in observed snow depth tend to bring less new snow in the ΔSNOW.MODEL (New Snow Module). Finally, simulated SWE_{pk} is reduced when η_0 or k are increased and vice versa.

3.1.4 Discrepancy Parameter τ

The ΔSNOW.MODEL's parameter to cope with uncertainties in snow depth is τ. It is supposed to be in the order of a few centimeters (at maximum). In particular, it should avoid excessive production of snow mass in the model through too frequent simulation of new snow events (see Sect. 2.2). τ is kind of a peculiarity of the ΔSNOW.MODEL and therefore no bounds can be found in literature. It was generously accepted to range between 1 and $20\,\mathrm{cm}$ for calibration and turned out to be optimal at $\tau = 2.4\,\mathrm{cm}$ (Table 3). Given the wide range of possible values, this is very close to what it would be expected to be as a measure for HS_{obs} accuracy.

Model sensitivity to changes in τ turns out to be quite low for values in the order of a few centimeters, but the influence on simulated SWE_{pk} is strongly increasing if τ is chosen greater than about $5\,\mathrm{cm}$ (Fig. 2). This result makes a lot of sense, if τ is seen as a measure of observation accuracy, because this is very likely to be better than $5\,\mathrm{cm}$. Like changes in η_0 and k, changes in τ are indirect proportional to changes in SWE_{pk} – for a closely related reason: The bigger τ the more often (small) new snow events are not counted as such because the *Scaling module* is more frequently activated at the cost of the New Snow Module (see Sect. 2.2). Mass gains are tendentially modeled less frequently and, as a consequence, snow water equivalents stay smaller.

3.1.5 Overburden Parameters c_{ov} and k_{ov}

Aside τ, there are two more parameters that are peculiar to the ΔSNOW.MODEL. They are needed to simulate unsteady compaction by overburden load of new snow (Eq. (3)). Because of their (presumed) uniqueness in the snow model spectrum there is no information available on how to choose them (see Sect. 2.3 for more). However, the calibration produces $c_{ov} = 5.10 \times 10^{-4}\,\mathrm{Pa}^{-1}$ and $k_{ov} = 0.379$ as fittest values (Table 3).

As outlined in Sect. 2.2.1, the implementation of overburden strain in the ΔSNOW.MODEL is supposed to be an important aspect of the model. Still, the sensitivity of modeled SWE_{pk} to changes in either c_{ov} or k_{ov} – without changing the respective other – are quite minor. (See Fig. 2) for c_{ov}. k_{ov} is not shown, because it is comparable, but with opposite sign.) The reason for this relative insensitivity of the model to changes in c_{ov} and k_{ov} could be the contradicting effects of these two "overburden parameters": Higher c_{ov} push overburden strain ϵ towards 1.0 (Eq. (3)), which increases the role of overburden snow. hs^* and HS^* of Eq. (4 are reduced and, consequently, the new layer thickness (and mass) is increased (Eq. (5)). Higher c_{ov}, therefore, lead to higher SWE and SWE_{pk}. For k_{ov} it is the opposite, higher values of k_{ov} cause lower SWE.

3.1 Validation and Comparison to other modelsIllustration

[19]As pointed out, the observations of snow water equivalent where divided in two sets, one for calibration (SWE_{cal}) and one for validation (SWE_{val}). The question arises, how accurate the manual SWE observations are. The two data sources do not address this point (Gruber, 2014; Marty,2017). In general, this is not easy to be answered since SWE measurements made with snow sampling cylinders mostly are used as references in comparison studies, without addressing *their* accuracy (e.g., Sturm et al., 2010; Dixon and Boon, 2012; Kinar and Pomeroy, 2015; Leppänen et al., 2018a). The majority of SWE_{cal} and SWE_{val} comes from the Hydrographic Service of Tyrol, Austria, where snow sampling cylinders ($500\,\mathrm{cm}^3$) are used (Sect. 2.3.1). The repeatability of this kind of measurement is estimated at $\pm 4\%$ for glacier mass balance studies, with filling height having the largest influence on accuracy (R. Prinz, Univ. of Innsbruck, Austria; pers. comm.). According to Leppaenen et al. (2018b), who compared various density samplers, the "variability among the replications of each sampler and variability among the samplers [is] both less than 10% [. . .] However, the uncertainty introduced by using different samplers was higher than the uncertainty caused by observer error". Roughly interpreting these density(!) measurement "variabilities" as relative observation errors for SWE, the results for absolute accuracy would typically spread across the wide range of about 2 to $50\,\mathrm{kg\,m^{-2}}$.

[20]Table 4 provides an overview of model uncertainties for SWE. Vionnet et al. (2012) find a root mean square error and bias of $39.7\,\mathrm{kg\,m^{-2}}$ and $-17.3\,\mathrm{kg\,m^{-2}}$, respectively, comparing 1722 manual samplings at Col de Porte (Chartreuse Mountains, France) and Crocus. These seem to be quite pessimistic values, since root mean square differences found by Langlois et al. (2009) are significantly lower. (However, they base on much fewer data.) Roughly summarized, SWE observations as well as "first-class" snow models' SWE simulations are associated with comparable uncertainties; RMSEs might be favorably approximated in the order of 10 to $20\,\mathrm{kg\,m^{-2}}$.

[19]The remarks on SWE observation accuracy were moved from the beginning of the Validation and Illustration section of the initially submitted manuscript to the Methods section of the revised manuscript. Content-related changes of this paragraph are minor.

[20]The paragraph about SWE accuracy of thermodynamic snow models was moved to Sect. 4.7 during the revision.

In this study no quantitative comparison with thermodynamic snow models was performed, since they need further meteorological data and the focus was on data records constrained to snow depths. However, the ΔSNOW.MODEL was thoroughly evaluated against ERMs. Figure 2 and Table 4 show the results. Even though ERMs do not need meteorological data, it is not straight forward to calibrate them for new sites and applications. From the vast number of ERMs (cf. Avanzi et al., 2015) the ones of Pistocchi (2016) and Guyennon et al. (2019) were chosen to be fitted to SWE_{cal}~~(standard: Pi16, Gu19; calibrated: Pi16cal, Gu19cal)~~. These models are ~~(~~quite~~)~~ new and easy to calibrate. Additionally, an approach simply using a constant bulk snow density at every point in time was calibrated to fit this study's data. ~~Interestingly,~~ $278\,\text{kg}\,\text{m}^{-3}$ turned out to be the optimal value minimizing root mean square errors of all SWE_{cal}~~SWE~~ values~~(ρ_{278})~~. Moreover, Jonas et al. (2009) and Sturm et al. (2010) were used for comparison~~(Jo09, St10)~~. Unfortunately, calibration of these powerful models would have needed much more data than the 780 SWE-HS-pairs of the SWE_{cal} data set~~, because this would have needed much more data these powerful models could not be calibrated with SWE_{cal}, because this would have needed much more SWE-HS-pairs than those about 1500 having been processed in this study~~. Therefore, Jonas et al. (2009) and Sturm et al. (2010) were used with their standard parameters, but for Jonas et al. (2009) it was distinguished between regions (see Fig. 2's caption)~~: Region 6 (Jo09R6) having the highest and Region 7 (Jo09R7) having the lowest "region-specific offset", respectively~~. Other contemporary approaches had to be ignored, mostly because of the problematic transferability of regional parameters (e.g., McCreight and Small, 2014, or Mizukami and Perica, 2008).

~~Figure 4's upper left panel indicates the decent performance of the ERMs at the mean when applied to the validation-half of the data set.~~ The bias of modeled SWE (lower left panel in Fig. 2) is quite low and tendentially positive, meaning SWE is often slightly overestimated by the ERMs.~~(Distinct values are given in Table 4.)~~ The ΔSNOW.MODEL ~~– on the contrary –~~ slightly underestimates SWE on average, the median bias is $-3.0\,\text{kg}\,\text{m}^{-2}$~~$-4.0\,\text{kg}\,\text{m}^{-2}$~~. The overall good results for the ERMs is not particularly surprising, since they are dedicated to perform well *on average*. The specially calibrated versions of Pistocchi (2016) and Guyennon et al. (2019) show a significantly smaller bias than their originals. The model of Jonas et al. (2009) has the smallest~~(actually a negative)~~ bias for their "Region 7", encompassing the dry, inneralpine Engadin as well as parts of the Southern Alps and the very East of Switzerland (Samnaun), which is partly influenced by orographic precipitation from Northwesterly flows. In terms of heterogeneity in precipitation climate "Region 7" is comparable to the region where the SWE data of this study comes from. ~~Sturm et al. (2010) assess the bias for their model (with their "alpine" data set) at $+29\,\text{kg}\,\text{m}^{-2}$ with a standard deviation of $57\,\text{kg}\,\text{m}^{-2}$, and they outline that "in a test against extensive Canadian data, 90% of the computed SWE values fell within $\pm 80\,\text{kg}\,\text{m}^{-2}$ of measured values". This is a much more conservative estimation than the results for this study would suggest.~~

The other three indicators illustrated in Fig. 2 and summarized in Table 4~~– bias of seasonal snow mass maximum SWE_{pk} (upper right panel) and the root mean square errors of individual SWE (lower left panel) as well as of SWE_{pk} (lower right panel) –~~ signify the better performance of the ΔSNOW.MODEL compared to ERMs:[21] The latter~~ERMs~~ are intrinsically tied to snow depth (see Sect. 1.2) and are systematically forced to overestimate SWE_{pk}. Developers of ERMs are well aware of this, nevertheless, ~~this is a pity since~~ SWE_{pk} is probably the most-wanted snowpack feature in hydrology, climatology, and

[21] The line break of the discussion paper was omitted in the revised manuscript.

extreme value analysis. ~~The ΔSNOW.MODEL proofs to perform much better here.~~ Note, the maximal SWE of a winter season not necessarily equals the highest measured SWE, because measurements are only taken weekly or biweekly. In the vast majority of the SWE records used for this study, the highest seasonal observation is followed by at least one lower SWE reading. Sometimes real SWE might be higher after the highest measurement of a winter season was taken, but a thorough data check revealed, this is of minor importance here. It is sufficiently precise to assume that measured seasonal maximum SWE equals SWE_{pk}. The ΔSNOW.MODEL's bias of SWE_{pk} is very minor, only $+0.3\,\mathrm{kg\,m^{-2}}$~~$+2.3\,\mathrm{kg\,m^{-2}}$ at median~~. ~~ERMs typically suffer from (far) too high SWE_{pk} simulations. The reasoning was given in the Introduction: "ERMs calibrated for good estimates of mean-SWE (necessarily) fail to model SWE_{pk} sufficiently well", since they are "overregulated" by the snow depth.~~ Moreover, the ΔSNOW.MODEL works better for the timing of SWE_{pk} (not shown in Fig. 2 and Table 4)~~does not only work well for the bias in SWE and SWE_{pk}, but also for the bias in the timing of SWE_{pk} (not shown in Fig. 2, but in the last column of Table 4)~~. ERMs tend to model SWE_{pk} some days too early, because~~ –~~ the date of modeled SWE_{pk} is shifted~~pulled~~ towards the date of highest HS (cf. Fig. 4)~~. The ΔSNOW.MODEL virtually reduces this bias to zero.~~

Another satisfactory validation result for the ΔSNOW.MODEL is shown in Fig. 2's upper~~lower~~ panels. RMSEs for all SWE values are constantly lower than if modeled with ERMs: a RMSE~~a median error~~ of $30.8\,\mathrm{kg\,m^{-2}}$~~$23.9\,\mathrm{kg\,m^{-2}}$~~ (ΔSNOW.MODEL) contrasts RMSEs~~faces median errors~~ between 39.1 and $50.9\,\mathrm{kg\,m^{-2}}$ (ERMs). Calibrating the models of Pistocchi (2016) and Guyennon et al. (2019) results in some improvement, at least they perform much better than the "constant density approach" ~~ρ_{278} approach~~ after the calibration. The model of Jonas et al. (2009) does a decent job also without recalibration, which is remarkable. Sturm et al. (2010)'s method is~~probably suffers from the handicap of being~~ calibrated with data from the Rocky Mountains. For this comparison the "alpine" parameters of Sturm et al. (2010) were taken, however, conditions might differ ~~(~~too~~)~~ much from the European Alps. Absolute errors in SWE increase with increasing SWE. For snowpacks lighter than $75\,\mathrm{kg\,m^{-2}}$ ΔSNOW.MODEL RMSE is $17\,\mathrm{kg\,m^{-2}}$, between $75\,\mathrm{kg\,m^{-2}}$ and $150\,\mathrm{kg\,m^{-2}}$ it is $26\,\mathrm{kg\,m^{-2}}$, and for snowpacks heavier than $150\,\mathrm{kg\,m^{-2}}$ it increases to $43\,\mathrm{kg\,m^{-2}}$.

~~Not least as a consequence of the mentioned drawbacks of the ERMs when dealing with distinct values, the advantages of the ΔSNOW.MODEL get most obvious when modeling SWE_{pk} (Fig. 4, lower right; Table 2, last three columns).~~ The ΔSNOW.MODEL also has~~manages to have~~ a small RMSE~~range of RMSEs, with a median~~ of $36.3\,\mathrm{kg\,m^{-2}}$~~$23.1\,\mathrm{kg\,m^{-2}}$~~ when modeling SWE_{pk} (Fig. 2, upper right; Table 4, last column)~~ and 75% of the errors staying below $34.4\,\mathrm{kg\,m^{-2}}$~~. Also the SWE_{pk}-RMSEs for the different SWE classes~~These values for SWE_{pk}~~ are very close to those for ~~(daily)~~ SWE, which emphasizes the ΔSNOW.MODEL's ability to model all individual SWEs comparably~~equally~~ well. The evaluated ERMs have much higher, mostly at least doubled~~doubled to tripled~~ errors in simulated SWE_{pk}~~and rather big spreads~~. Remarkably, the simple ρ_{278} approach is performing relatively well. In case the Jonas et al. (2009) model is suitably adjusted to regional specialties, it performs better than the other ERMs, but still significantly worse than the ΔSNOW.MODEL.~~The Jonas et al. (2009) model – if suitable adjusted to regional specialties – performs best, still, well beaten by the ΔSNOW.MODEL.~~

The ΔSNOW.MODEL ~~root mean square errors – slightly lower than $25\,\mathrm{kg\,m^{-2}}$, also for *certain values* like SWE_{pk} – are higher than the errors of the thermodynamic snow models (roughly estimated at 10 to $20\,\mathrm{kg\,m^{-2}}$), which is not surprising given the latter's demands on input data and computational power. However, the ΔSNOW.MODEL~~ outperforms empirical regression models. This can be argued on base of this study~~the study in hand~~ (especially Fig. 2), but even more when looking at the ERM studies themselves: Jonas et al. (2009) provide RMSEs between 50.9 and 53.2 $\mathrm{kg\,m^{-2}}$ for their standard model, which are quite high values compared to the findings of the study in hand (39.4 $\mathrm{kg\,m^{-2}}$ for their Region 7~~ca. 30 $\mathrm{kg\,m^{-2}}$ for SWE~~, see Table 4). One explanation could be that Jonas et al. (2009)~~,~~ as well as other ERM studies~~,~~ rely on a huge amount of, but still diverse measurements in terms of record length, observations per season etc.~~very diverse measurements (Still, lots of them!).~~ The ΔSNOW.MODEL-study~~ΔSNOW.MODEL study~~ only consists of data from selected stations with long and regular SWE readings, where also ERMs seem to work better. Guyennon et al. (2019) summarize their and other studies' validation results using MAE, the mean absolute error~~, and Sturm et al. (2019) publish values for SWE-BIAS~~. Sturm et al. (2010) assess the bias for their "alpine" model at $+29\,\mathrm{kg\,m^{-2}}$ with a standard deviation of $57\,\mathrm{kg\,m^{-2}}$, and they outline that "in a test against extensive Canadian data, 90% of the computed SWE values fell within $\pm 80\,\mathrm{kg\,m^{-2}}$ of measured values". Table 4 provides an overview and shows~~the overview, and again it gets obvious~~, that ERMs generally perform better with this study's data~~,~~ than with their original~~Guyennon et al. (2019)'s~~ data.

3.2 Illustration

[22]Figure 1 schematically shows the functioning of the ΔSNOW.MODEL. A practical example is ~~now~~ provided in Fig. 3, based on the optimal calibration parameters found during this study~~ and using the same colors as in Fig. 1~~. Kössen, the station shown, is situated in the Northern Alps at 590 m above sea level (cf. Fig. A1). Although it is a low-lying place it is known to be snowy, which is, firstly, due to intense orographic enhancement of precipitation associated with Northwesterly to Northeasterly flows in the respective region (Wastl, 2008) and, secondly, comparably frequent inflow of cold continental air masses from Northeast. Showing Kössen~~ – rather than a high altitude station –~~ should emphasize the versatile usability of the ΔSNOW.MODEL:~~.~~ It is not only designed for high areas with deep, long-lasting snowpacks, but also for, e.g., valleys with shallow, ephemeral snowpacks. Winter 2008/09 was chosen because the ΔSNOW.MODEL shows a rather typical performance in terms of RMSE and BIAS ~~RMSEs etc.~~ in Kössen then (see Table 4, values in brackets) and because some important, model-intrinsic features can be addressed and discussed:

Late November 2008 brought the first, however transient snowpack of the season (Fig. 3). The ΔSNOW.MODEL identifies two days with snow fall (purple markings) and models two respective snow layers, which can be distinguished by the thin black line in Fig. 3. After about a week the snowpack starts to melt, the snow layers reach $\rho_{\max}$ very fast (the blue shading gets dark), and finally all the snow was converted to runoff (green markings). In the second half of December there were three days with new snow, followed by a strong decline in snow depth. In the frame of the ΔSNOW.MODEL this HS decrease is only possible, if the layers "get wet" and ~~–~~the *Drenching module* is activated (marked in brown). The layers get denser, starting at the top. However, the decrease was "manageable" by only increasing the two uppermost layer densities to $\rho_{\max}$ and making the third layer just a bit denser. Not all layers got to $\rho_{\max}$ ("saturated") and no runoff was modeled. The ΔSNOW.MODEL conserves the two dense layers until the end of the winter, which can clearly be seen in Fig. 3. ~~(~~One could interpret the layers as consisting of melt forms or a refrozen crust. However, interpretations like that require caution, because modeling such detailed layer features is not the

[22]The Illustration subsection was newly introduced during the revision.

intention of the ΔSNOW.MODEL!~~)~~ During January Fig. 3 shows a phase where ~~mnothing special happens. M~~odeled values and observations agree to a high extent and only the *Scaling module* 8~~i~~d=general]is activated formakes small adjustments (white markings). Small "stretching events" can be recognized, e.g. on January 2nd and 3rd, where model snow layers are set less dense in order to avoid too frequent mass gains. (This model behavior was thoroughly described in Sect. 2.2.2.) During continuous snowfalls in February the successive darkening of the blue layer shadings illustrates a phase of consequent compaction, which actually lasts until March, when strong decreases in HS_{obs} ~~(and, thus, in HS) already~~ start to activate the *Drenching module*. Still, runoff is not yet produced. Only in the second half of March the whole model~~-~~snowpack reaches ρ_{max} ("saturation"). The ablation phase is clearly distinguishable and lets the snowpack vanish quite fast until about April 10th, 2009.

The snow depth record of Kössen from 2008/09 was also used to compare different ERMs and the ΔSNOW.MODEL to SWE observations (Fig. 4). These measurements (light blue circles) are part of the SWE_{val} sample and were manually made with snow sampling cylinders; one after the December 2008 snowfall, and another nine on a nearly weekly base between late January and late March 2009. Figure 4 also provides various model results and some respective key values are given in Table 4. Not surprisingly, thus evidently, the ERMs' SWE curves "follow" the snow depth curve (black dashed line). The ΔSNOW.MODEL (red line) does not get the first four measurements decently correct, the ERMs perform better in this illustrative case.~~ (By the way, the "jumpy behavior" of the model of Jonas et al. (2009), criticized by Pistocchi (2016), does not play a role.)~~ But~~,~~ after the stronger snowfalls of February~~,~~ the picture changes indisputably in favor of the ΔSNOW.MODEL. This is a typical pattern, nothing special for Kössen 2008/09, ~~(~~albeit it is quite pronounced in this ~~illustrative~~ example~~)~~: The ERMs are too strongly tied to snow depth and, therefore, mostly (1) overestimate SWE_{pk}, (2) model its occurrence too early, and (3) – most importantly – force modeled SWE to reduce during pure compaction phases after snowfalls.~~ All these points were discussed in detail earlier, Fig. 4 visualizes them. At the same time, the ΔSNOW.MODEL does a good job in modeling mean and maximum SWE, not only in Kössen 2008/09 but also on average (cf. Fig. 2 and Table 4).~~ Evidently, the ability of the ΔSNOW.MODEL to "conserve" mass during the phases with dry metamorphism is its strongest point, not only in Kössen 2008/09 but also on average (cf. Fig. 2 and Table 4).

4 Discussion and Outlook

Model results clearly depend on the parameters. Their optimal values are subject to~~of~~ calibration. The choice of the best-fitted values is rated and discussed in the following Sects. 4.1 to 4.5. Sections 4.6 to 4.8 cover possible future developments, accuracy issues, and the ΔSNOW.MODEL's applicability in remote sensing.[23]

4.1 New Snow Density ρ_0

Being aware of both – the huge possible variations of new snow density depending on meteorological conditions during snowfalls and the possible cruciality of this parameter for SWE simulation by the ΔSNOW.MODEL – ρ_0 was chosen to be a constant in the framework of the model. $\rho_0 = 81\,\mathrm{kg\,m^{-3}}$ turned out to be the best choice after calibration with SWE_{cal}. This

[23] Sects. 4.1 to 4.5 basically mirror Sects. 3.1.1 to 3.1.5 of the initially submitted manuscript. Only minor changes were made during the revision.

value clearly lies within the broader frame of possible new snow densities (Table 3) and quite closely to Sturm and Holmgren (1998)'s $75\,\mathrm{kg\,m^{-3}}$, but it is found in the lower part for "typical" new snow densities (e.g., Helfricht et al., 2018). A possible explanation could be that the SWE measurement records used for the calibration tend to underrepresent late winter and spring conditions. Regular (weekly, biweekly) observations capture the short melt seasons worse than the longer accumulation phases. Therefore, SWE records might be biased towards early and mid winter new snow densities, which are lower (e.g., Jonas et al., 2009). Still, there are also some indications that using, e.g., $100\,\mathrm{kg\,m^{-3}}$ as constant new snow density when modeling SWE results in an overestimation of precipitation (up to 30% according to Mair et al., 2016). The calibrated value for ρ_0 can be regarded as a reasonable result, even more when only considering it as a model parameter but not as a physical constant.

The sensitivity analysis illustrated in Fig. 5 confirms the importance of a good choice of ρ_0. Increasing ρ_0 quite fast leads to a decrease of the relative bias of seasonal SWE maxima (SWE_{pk}). Note the definition of the relative bias in Fig. 5's caption. In absolute values: too small ρ_0 cause too small SWE_{pk}, using higher values leads to an overestimation of SWE_{pk}. This behavior supports above-mentioned tendency to overestimate precipitation when choosing constant $100\,\mathrm{kg\,m^{-3}}$ as new snow density. As expected, the new snow density is the most crucial parameter of the ΔSNOW.MODEL (cf. Table 3). The median relative bias of SWE_{pk} changes by -0.46% per $+1\,\mathrm{kg\,m^{-3}}$, if the whole calibration range of ρ_0 is considered to calculate the sensitivity ($50-200\,\mathrm{kg\,m^{-3}}$). This means a median change in SWE_{pk} of $+0.37\,\mathrm{kg\,m^{-2}}$ when ρ_0 is risen by $+1\,\mathrm{kg\,m^{-3}}$. If the limits are chosen tighter around the optimal value, the gradient is even steeper: -0.62% and $+0.50\,\mathrm{kg\,m^{-2}}$ per $+1\,\mathrm{kg\,m^{-3}}$, respectively, when the gradient is approximated for the range $70-90\,\mathrm{kg\,m^{-3}}$. Widely-used $\rho_0=100\,\mathrm{kg\,m^{-3}}$, consequently, causes a median overestimation of SWE_{pk} of about 12% in the ΔSNOW.MODEL. Daily SWE show the same behavior (not shown). Users should be aware of this. The suggestion clearly is to either use the best-fitted parameters of this study or recalibrate *all* parameters with appropriate SWE data, but not adjusting only single parameters. As the calibration data of this study are spread across various climates and altitudes, users can be quite confident to get good results if using $\rho_0=81\,\mathrm{kg\,m^{-3}}$. This value seems to be a good compromise, at least at alpine areas. However, for very maritime, very dry, polar or tundra regions the optimized ρ_0 should be used with caution; if possible, recalibration is recommended.

4.2 Maximum Density ρ_{max}

Of course, the maximum bulk snow density of a snowpack changes from year to year and site to site. For the ΔSNOW.MODEL simplicity and independence from meteorological variables outweigh precision. Even more so, when there are good arguments for the existence of a "typical" maximum bulk density ρ_{max}. Put simply, (not too old) seasonal and also ephemeral snowpacks melt away when they get water saturated. Before that, there is limited time for dry densification; dry winter snow's bulk density is widely described as staying below about $350\,\mathrm{kg\,m^{-3}}$ (e.g., Paterson, 1998; Sandells et al., 2012). Accounting for the fact that volumetric liquid water content of about 10% marks the funicular mode of liquid distribution in old, coarse-grained snow (Denoth, 1982; Mitterer et al., 2011), this leads to the rough estimate of a typical maximum bulk density of about $\frac{9}{10}\cdot 350+\frac{1}{10}\cdot 1000=415\,\mathrm{kg\,m^{-3}}$. Convincingly, the fittest value for ρ_{max} in the ΔSNOW.MODEL turns out to be $401\,\mathrm{kg\,m^{-3}}$, which is close to that value and well situated within the range given in the literature (Table 3). Moreover, this is virtually the same value like the median maximum seasonal density of the SWE_{val} data records ($400\,\text{kg}\,\text{m}^{-3}$, see box plot in Fig. 6), another indication why ρ_{max} could be regarded as a typical seasonal maximum of ρ_{b}.[24]

Figure 5 illustrates the similarity between ρ_0 and ρ_{max} regarding their influence on SWE simulations. Keeping the other six ΔSNOW.MODEL parameters constant but increasing ρ_{max} leads to increased SWE_{pk} and vice versa – just like ρ_0. This is not surprising, however reasonable. The ΔSNOW.MODEL is not as sensitive to changes in ρ_{max} than to changes in ρ_0: Raising ρ_{max} by $+1\,\text{kg}\,\text{m}^{-3}$ leads to a mean decrease of the relative bias of SWE_{pk} of -0.06%, which corresponds to an increase in absolute SWE_{pk} of $+0.24\,\text{kg}\,\text{m}^{-2}$ per $+1\,\text{kg}\,\text{m}^{-3}$. The same argumentation like for ρ_0 in Sect. 4.1 lets users of the ΔSNOW.MODEL be quite sure when taking $\rho_{\text{max}} = 401\,\text{kg}\,\text{m}^{-3}$, the best-fitted value according to this study's calibration. Be aware that solely changing parameter ρ_{max} for an application of the ΔSNOW.MODEL elsewhere, without proper recalibration of the other parameters, might lead to significant changes in the results for SWE.

4.3 Viscosity Parameters η_0 and k

Equation (2) represents the settlement and densification function of the ΔSNOW.MODEL. Two parameters η_0 and k act as adjustment screws and have to be calibrated. In this study best-fittet η_0 is $8.5 \times 10^6\,\text{Pa}\,\text{s}$ and the optimized value for k is $0.030\,\text{m}^3\,\text{kg}^{-1}$. Both values are close to other studies' results and suggestions (Table 3).

As far as the ΔSNOW.MODEL's sensitivity to changes in the "viscosity parameters" η_0 and k is concerned, Fig. 5 shows that an isolated rise of the model snow viscosity – either by enhancing η_0 or k – increases the relative bias of SWE_{pk}, which means a decrease in absolute values of SWE_{pk}. This behavior is consistent, since higher viscosity reduces the densification rate and the model-snowpack tendentially stays deeper. Consequently, increases in observed snow depth tend to bring less new snow while the *New Snow module* is run (Sect. 2.2.1). Finally, simulated SWE_{pk} is reduced when η_0 or k are increased and vice versa.

4.4 Threshold Deviation τ

The ΔSNOW.MODEL's parameter to cope with uncertainties in snow depth is τ. It is supposed to be not bigger than a few centimeters. In particular, it should avoid excessive production of snow mass in the model through too frequent simulation of new snow events (see Sect. 2.2). τ is kind of a peculiarity of the ΔSNOW.MODEL and therefore no bounds can be found in literature. It was generously accepted to range between 1 and $20\,\text{cm}$ for calibration and turned out to be optimal at $\tau = 2.4\,\text{cm}$ (Table 3). Given the wide range of possible values, this is very close to what it would be expected to be as a measure for HS_{obs} accuracy.

Model sensitivity to changes in τ turns out to be quite low for values in the order of a few centimeters, but the influence on simulated SWE_{pk} is strongly increasing if τ is chosen greater than about $5\,\text{cm}$ (Fig. 5). This result makes a lot of sense, if τ is seen as a measure of observation accuracy, because this is very likely to be better than $5\,\text{cm}$. Like changes in η_0 and k, changes in τ are indirect proportional to changes in SWE_{pk}, for a closely related reason: The bigger τ the more often small new snow events are not counted as such because the *Scaling module* (Sect. 2.2.2) is more frequently activated at the cost of the *New*

[24]Figure 3 of the initially submitted manuscript is Fig. 6 of the revised version.

Snow module (Sect. 2.2.1). Mass gains are tendentially modeled less frequently and, as a consequence, snow water equivalents get smaller.

4.5 Overburden Parameters c_{ov} and k_{ov}

Aside τ, there are two more parameters that are peculiar to the ΔSNOW.MODEL. They are needed to simulate unsteady compaction by overburden load of new snow. Because of their presumed uniqueness in the snow model spectrum there is no information available on how to choose them (see Sect. 2.3). However, the calibration produces $c_{\text{ov}} = 5.1 \times 10^{-4}\,\text{Pa}^{-1}$ and $k_{\text{ov}} = 0.38$ as fittest values (Table 3).

As outlined in Sect. 2.2.1, the implementation of overburden strain in the ΔSNOW.MODEL is supposed to be an important aspect of the model. Still, the sensitivity of modeled SWE_{pk} to changes in either c_{ov} or k_{ov} are quite minor. (See Fig. 5 for c_{ov}. k_{ov} is not shown, because it is comparable, but with different sign.) The reason for this relative insensitivity of the model to changes in c_{ov} and k_{ov} could be the contradicting effects of these two "overburden parameters": Higher c_{ov} push overburden strain ϵ towards 1.0 (cf. Eq. (3)), which increases the role of overburden snow. hs^* and HS^* of Eq. (4) are reduced and, consequently, the new layer thickness and mass are increased (Eq. (5)). Higher c_{ov}, therefore, lead to higher SWE and SWE_{pk}. For k_{ov} it is the opposite, higher values of k_{ov} cause lower SWE.

4.6 Incorporating Rain-on-Snow and other possible improvements

In principle, the ΔSNOW.MODEL could deal with rain-on-snow events.~~Some questions remain. The treatment of rain-on-snow events surely is one of them. The ΔSNOW.MODEL can – in principle – deal with it.~~ Unsteady compaction due to overburden load, for example, is not restricted to new snow. It could also be triggered by the mass of rain water – in nature, but also in the framework of the ΔSNOW.MODEL. Still, the respective feature is not implemented~~coding is not finalized~~ at the moment, because identifying criteria for rain-on-snow events based on pure snow depth records is very problematic~~a problem~~, and its resolving ~~(if at all possible)~~ is beyond the scope of this paper. In case ~~some~~ meta information on~~, e.g.,~~ rain climate ~~(maybe for a stochastic "rain generator")~~ or on precipitation type and amount~~(no matter if liquid or solid) combined with information on the snowfall line~~ is available, it could ~~quite easily~~ be incorporated in the ΔSNOW.MODEL. Given the relative success of the ΔSNOW.MODEL in its current version~~ – especially the reduction of the RMSE of SWE_{pk} by 50-70% compared to ERMs and *maybe* even down to the SWE error range of thermodynamic models –~~ the probably very costly, but potentially often only very minor improvements when including rain-on-snow should be considered.

Another~~ discussion point and~~ eventual future development is the refinement of the density parameters ρ_0 and ρ_{max} since, firstly, the ΔSNOW.MODEL reacts quite sensitive on their changes and, secondly, some relations are well known, e.g., ρ_0's dependence on the climatic aridness or ρ_{max}'s tendentious increase for aging~~(very) old~~ snow. Setting ρ_{max} to a fixed value at about $400\,\text{kg}\,\text{m}^{-3}$ actually disqualifies the ΔSNOW.MODEL for snow older than estimated 200 days. Additional calibrations could be performed for very maritime, very dry, polar, or tundra regions as well as for very long-lasting snowpacks. Note, however, all of these adaptions introduce more parameters to the ΔSNOW.MODEL and reduce its generality. Benefits should be evaluated critically, and probably this evaluation should start with~~. Probably against~~ the overburden load treatment of the

ΔSNOW.MODEL. I~~, since i~~t is possible that refining the density parameters is more valuable than the special treatment of unsteady compaction due to overburden loadsΔSNOW.MODEL.~~. . .~~

4.7 *SWE* Accuracy

Table 4 provides an overview of ~~model~~ uncertainties for SWE, also for thermodynamic models:~~.~~ Vionnet et al. (2012) find a root mean square error and bias of $39.7\,\mathrm{kg\,m^{-2}}$ and $-17.3\,\mathrm{kg\,m^{-2}}$, respectively, comparing 1722 manual samplings at Col de Porte (Chartreuse Mountains, France) and Crocus. Wever et al. (2015) and Sandells et al. (2012) come up with RMSEs of about $39.5\,\mathrm{kg\,m^{-2}}$ (SNOWPACK) and $30-49\,\mathrm{kg\,m^{-2}}$ (SNOBAL), respectively. Langlois et al. (2009) find more optimistic values, however, based on much fewer data. On the contrary, Egli et al. (2009) give reason to expect higher RMSEs, but their study exclusively bases on data from snowy, high altitude station Weissfluhjoch (Switzerland), which intrinsically promotes higher absolute errors. Essery et al. (2013)'s comprehensive simulation experiment results in a RMSE-range of $23-77\,\mathrm{kg\,m^{-2}}$.~~ These seem to be quite pessimistic values, since root mean square differences found by Langlois et al. (2009) are significantly lower. (However, they base on much fewer data.) Roughly summarized, SWE observations as well as "first-class" snow models' SWE simulations are associated with comparable uncertainties; RMSEs might be favorably approximated in the order of 10 to $20\,\mathrm{kg\,m^{-2}}$.~~

As a synopsis of the study in hand, absolute SWE accuracies could be estimated as follows: (1) 2 to $50\,\mathrm{kg\,m^{-2}}$ for manual measurements, which are widely used as reference, (2) 30 to $40\,\mathrm{kg\,m^{-2}}$ for thermodynamic models, and (3) 40 to $50\,\mathrm{kg\,m^{-2}}$ for empirical regression models. In this respect, it is striking to find the ΔSNOW.MODEL's RMSE at $30.8\,\mathrm{kg\,m^{-2}}$. ~~"Maybe" is emphasized in the last paragraph because it leads to another discussion and outlook point: An intensive, multi-year model comparison should be performed; at some benchmark sites with fully equipped snow stations and – very important – different methods of SWE measurements, including regular manual observations (with sampling cylinders etc.). Some of those data sets do already exist, however, a comprehensive comparison of techniques and methods to measure and, in particular, model SWE is lacking. Often the "target variable" is bulk density (not SWE), and relative (not absolute) numbers are the only information on observation accuracy one can get, although recently efforts are undertaken (e.g., in the framework of the ESSEM COST Action ES1404; Leppänen et al., 2018; López-Moreno et al., 2020,. . .) – at least concerning measurements. Shouldn't there be more studies, that also comprehensively quantify the abilities of various, especially thermodynamic snow models to simulate SWE? A top-quality comparison between the ΔSNOW.MODEL and thermodynamic snow models is actually difficult to achieve since hardly any numbers for SWE accuracy of thermodynamic models are available. *Maybe* they perform worse than the generously estimated RMSEs of 10 to $20\,\mathrm{kg\,m^{-2}}$ of this study?~~

4.8 Application to Remote Sensing Data

L~~Last but not least, l~~ooking at current developments in deriving SWE from snow depths~~, that are~~ monitored with lidar and photogrammetry, the ΔSNOW.MODEL should be considered as one of the "potential [. . .] other snow density models" (Smyth et al., 2019) that should be included in respective future research.~~ – following Smyth et al. (2019) – one of the "potentially [applicable] other snow density models"~~. Lidar and photogrammetry have errors in the order of $10\,\mathrm{cm}$ (Smyth et al., 2019), typically corresponding to SWE errors of 20 to $40\,\mathrm{kg\,m^{-2}}$. This is in the order ~~or somewhat more than the error~~ of the ΔSNOW.MODEL errors. Remote sensing derived snow depth data are discontinuous through time. The ΔSNOW.MODEL would have to be adapted to that, but for the benefit of upgrading the ΔSNOW.MODEL from a point model to a computationally fast distributed model. ~~(which might not be a big issue), though, for the big benefit of being independent from meteorological data and models – and their errors.~~

5 Conclusions

A new method to simulate snow water equivalents (SWEs) is presented. It exclusively needs snow depths and their temporal changes as input, which is its major advantage compared to many other snow models. It is shown that basic snow physics, smartly implemented in a layer model, suffice to better calculate SWE than snow models relying on empirical regressions. ~~Consequently, the study's null hypothesis (Sect. 1.3) is rejected.~~

Regular~~Gapless~~ snow depth records are used to stepwise model the evolution of seasonal snowpacks, focusing on their mass (i.e. SWE) and respective load. Snow compaction is assumed to follow Newtonian viscosity, unsteady stress for underlying snow layers by the overburden load of new snow is regarded separately, melted mass is distributed from upper to lower layers, and – eponymous for the model – the measured change in snow depth between two observations is used as a precious corrective, though by accounting for measurement uncertainties. ~~The model steps are rather simple, however tricky in details, and all is frankly revealed in this article.~~

The ΔSNOW.MODEL mainly bases on Martinec and Rango (1991) and Sturm and Holmgren (1998), and transforms them to a modern R-code, which is available through https://cran.r-project.org/package=nixmass ~~(*nixmass* package)~~. Aside snow depth, meteorological ~~Other meteorological (aside HS)~~ and also geographical input is consequently avoided in the framework of the ΔSNOW.MODEL. Still, calibration of seven parameters is needed. To provide an optimal setting and utmost applicability, data from 14 climatologically different places in the Swiss and Austrian Alps are utilized. This is challenging, since calibration needs multi-year SWE observations as well as consecutive (e.g. daily) snow depth readings from the same places. The ΔSNOW.MODEL is calibrated with 67 ~~the help of 71~~ winters. The validation data set consists of another 71~~73~~ independent winters. Whereas calibration is rather complex, the application of the ΔSNOW.MODEL is cheap in terms of computational effort: Deriving a one-year SWE record from 365 snow depth values, e.g., only takes a few seconds with today's standard desktop CPUs and can certainly be speeded up significantly.

In this study it~~It~~ is argued that the ΔSNOW.MODEL is situated between sophisticated "thermodynamic snow models", necessitating lots of meteorological and other input, and modest "empirical regression models" (ERMs), relying on statistical relations between SWE and snow depth, date, altitude, and region. These key qualities of the ΔSNOW.MODEL are:

- ~~This "interposition" is true in terms of model~~ low complexity: The ΔSNOW.MODEL is a semi-empirical multi-layer model with seven parameters. It only needs regular HS records as input. In some respect it is even less demanding than ERMs, because no information on date, altitude, or region is required. ~~but only needs one input variable, which is HS.~~

– Still, the ΔSNOW.MODEL is even less demanding than ERMs: It exclusively needs HS records, though gapless ones. No information on date, altitude, or region is required.

– In terms of high universality: The ΔSNOW.MODEL simulates individual SWE values – like the important seasonal maximum SWE_{pk} – comparably well as SWE averages., this shifts the ΔSNOW.MODEL close to thermodynamic models, and because the ΔSNOW.MODEL simulates individual SWE values – like the important seasonal maximum SWE_{pk} – as good as SWE averages, it can compete with thermodynamic models at the, e.g., daily level as well (which is not reasonable for ERMs).

– high accuracy: TNot least, the ΔSNOW.MODEL's performance in modeling SWE and SWE_{pk} is comparable to thermodynamic models and superior to ERMs.lies between thermodynamic models and ERMs, albeit close to the very sophisticated ones: Root mean square errors for SWE_{pk} are $36.3\,\text{kg}\,\text{m}^{-2}$ for the ΔSNOW.MODEL and about 70 to $> 100\,\text{kg}\,\text{m}^{-2}$ for ERMs. (cf. Table 4) are at ca. $23\,\text{kg}\,\text{m}^{-2}$ (ΔSNOW.MODEL), at about 60 to $90\,\text{kg}\,\text{m}^{-2}$ (ERMs), and somewhere between 10 and $40\,\text{kg}\,\text{m}^{-2}$ (thermodynamic models).

Given these promising results, the ΔSNOW.MODEL's ancestors Sturm and Holmgren (1998)'s argument, whereby "snow load plays a more limited role in determining the compaction behavior in seasonal snow than grain and bond characteristics and temperature", might be disproved.

The development of the ΔSNOW.MODEL is application-driven. It is therefore not surprising that this study provides no significant new findings in snow physics. Still, the ΔSNOW.MODEL seems to be the first model (since long) that takes well known basic snow principles and arranges them in a physically consistent way, while consequently ignoring all potential information except snow depth. Not particularly innovative, but remarkably successful. Nevertheless, the synopsis of the ΔSNOW.MODEL and measured data gives significant hints on two important snowpack features (at least for the Alps): Typical mean density for new snow (24 h) seems to be clearly below often assumed $100\,\text{kg}\,\text{m}^{-3}$ and a characteristic average maximum bulk density for seasonal snow (also including ephemeral snowpacks from low-lying places) can rather be found around $400\,\text{kg}\,\text{m}^{-3}$ than at often cited 500 to $600\,\text{kg}\,\text{m}^{-3}$, which might be biased by "too alpine" snowpacks for many applications. The ΔSNOW.MODEL is widely usable, but first of all it can attribute snow water equivalents to all longterm and historic snow depth records, which are so valuable for climatological studies and extreme value analysis for risk assessment of natural hazards .

Code availability. R-code of the ΔSNOW.MODEL and some empirical regression models: https://cran.r-project.org/package=nixmass. (Phyton-code of the ΔSNOW.MODEL, ported by M. Theurl (Univ. of Graz, Austria): https://bitbucket.org/atraxoo/snow_to_swe.)https://cran.r-project.org/package= (nixmass package)

Appendix A

A map with the stations used for calibration and validation of the ΔSNOW.MODEL is shown in Fig. A1. Table A1 provides details on the stations and the data.

Appendix B

The *Scaling module* (Sect. 2.2.2) recalculates the “viscosity parameter” η_0. This temporary $\eta_0^*(i,t)$ does not only depend on the point in time t (whenever the *Scaling module* is activated), but is also different for each layer i. The reason is described in the following.

The *Scaling module* aims for the condition, that today's model snow depth $HS(t)$ equals today's observed snow depth $HS_{\text{obs}}(t)$.

$$HS(t) = \sum_{i=1}^{ly(t)} hs(i,t) \stackrel{!}{=} HS_{\text{obs}}(t)$$

It follows from Eq. (2) and substituting $x(i,t) = \Delta t \cdot \widehat{\sigma}(i,t) \cdot e^{-k \cdot \rho(i,t)}$:

$$\sum_{i=1}^{ly(t)} hs(i,t) = \sum_{i=1}^{ly(t)} \frac{\eta_0^*(t) \cdot hs(i,t-1)}{\eta_0^*(t) + x(i,t)} \stackrel{!}{=} HS_{\text{obs}}(t), \tag{B1}$$

which is a rational function f of the form

$$f(\eta) = \sum_{i=1}^{N} \frac{\eta \cdot h_i}{\eta + x_i}$$

Because $f(\eta)$ has poles at $-x_1, \ldots, -x_N$, the equation $f(\eta) = HS_{\text{obs}}$ has multiple solutions. Consequently, this approach – with $\eta_0^*(t)$ being independent from layer i – shows a clear non-physical behavior making it necessary to calculate different $\eta_0^*(i,t)$ for each layer i based on Eq. (B1):

$$\eta_0^*(i,t) = \frac{x(i,t) \cdot hs(i,t)}{hs(i,t-1) - hs(i,t)}$$

The solution of this issue in the *Scaling module* of the ΔSNOW.MODEL bases on the assumption, that observed compaction between $t-1$ and t can be approximated linearly for each layer:

$$\frac{hs(i,t)}{hs(i,t-1)} \stackrel{!}{\approx} \frac{HS_{\text{obs}}(t)}{HS_{\text{obs}}(t-1)}$$

The layer-individual viscosities can be calculated as

$$\eta_0^*(i,t) = \frac{x(i,t) \cdot HS_{\text{obs}}(t)}{HS_{\text{obs}}(t-1) - HS_{\text{obs}}(t)}$$

Substituting those values for η_0^* in Eq. (B1) fulfills its precondition, and the modeled equals the observed snow depth. The newly calculated $\eta_0^*(i,t)$ are different for each layer – in contrast to the fixed η_0 defined in Sect. 2.1, which is valid for the whole snowpack (outside the *Scaling module*). Note~~, that~~ these new viscosities are only used temporarily in the *Scaling module*. They have no analog in reality and can also have negative values, but they are mathematically sound.

Appendix C: Example of Application – Snow Load Map of Austria

In this section an example is given how the ΔSNOW.MODEL can be used to attain a map of snow loads in Austria~~; snow load $SL = SWE \cdot g$~~. European Standards (e.g., European Committee for Standardization, 2015) define the "characteristic snow load" s_k as the weight of snow on the ground with an annual probability of exceedance of 0.02, i.e. a snow load that – on average – is exceeded only once within 50 years. Unfortunately, SWE is not measured on a regular basis at a reasonable number of sites in Austria (and most other countries). The ΔSNOW.MODEL, however, can provide longterm Austrian SWE series from widely available HS series, which can in turn be used for a spatial extreme value model. No other snow model is capable of this in a comparable manner, since either SWE_{pk} is poorly modeled (ERMs) or more meteorological input would be needed (thermodynamic models). Among several possibilities to spatially model snow depth extremes like max-stable processes (see e.g. Blanchet and Davison, 2011), the *smooth modeling* approach of Blanchet and Lehning (2010) can be used when marginals instead of spatial extremal dependence is in focus.

C1 Smooth Modeling

Extremes following a generalized extreme value distribution (GEV; Coles, 2001) with parameters μ, σ and ξ can be modeled in space by considering linear relations for the three parameters of the form

$$\eta(x) = \alpha_0 + \sum_{k=1}^{m} \alpha_k y_k(x) \tag{C1}$$

at location x, where η denotes one of the GEV parameters, $y_1, \ldots, y_m$ are the considered covariates as smooth functions of the location, and $\alpha_0, \ldots, \alpha_m \in \mathbb{R}$ are the coefficients. Assuming spatially independent stations, the log-likelihood function then reads as

$$l = \sum_{k=1}^{K} \ell_k \left(\mu(x_k), \sigma(x_k), \xi(x_k)\right), \tag{C2}$$

where l only depends on the coefficients of the linear models for the GEV parameters. This approach was termed *smooth modeling* by Blanchet and Lehning (2010). A smooth spatial model for extreme snow depths in Austria was already presented in Schellander and Hell (2018), using longitude, latitude, altitude, and mean snow depth at 421 stations. Considering the strong correlation between snow depth and snow water equivalent, it would be natural to spatially model SWE extremes in the same manner.

C2 Fitting a Spatial Extreme Value Model

For this application 214 stations with regular~~gapless~~ snow depth observations in and tightly around Austria of the National Weather Service (ZAMG) and the Hydrological Services are used. The dataset has undergone quality control by the maintaining institutions and covers altitudes between 118 and 2290 m.~~118 and 2290 m a.s.l.~~ The records have lengths of 43 years and cover winters from 1970/71 to 2011/2012.

In a first step the ΔSNOW.MODEL was applied to these snow depth series to achieve 214 data series of SWE across Austria. ~~(This indeed is the great strength of the ΔSNOW.MODEL and can hardly be done with other methods!)~~ Then the linear models for the three GEV parameters according to Sect. C1 were defined via a model selection procedure. For that purpose a generalized linear regression was performed between the parameters and the covariates longitude, latitude, altitude, and mean snow depth, which were added in a stepwise manner. Using the Akaike information criterion (AIC; Akaike, 1974), the best linear model between a given full model ($\mu \sim$ all covariates) and a null model ($\mu \sim 1$) with the smallest AIC was selected. Using these models and the covariates of the 214 stations, a smooth spatial model for the yearly maxima of the SWE values was fitted.

C3 Return Level Map of 50-year Snow Load in Austria

The spatial extreme value model developed in the previous section was applied to a grid provided by the SNOWGRID climate analysis (Olefs et al., 2013). It offers the necessary covariates longitude, latitude, altitude, and~~ with~~ yearly mean snow depths from 1961 to 2016. The grid features a horizontal resolution of $1 \times 1\,\mathrm{km}$. Some minor SNOWGRID pixels have unrealistically large mean snow depth values, arising from a poor implementation of lateral snow redistribution at high altitudes (18 pixels, i.e. 0.02% with values between 5 and $65\,\mathrm{m}$). They are masked for the calculation of SWE return level maps. The return level map for a return period of 50 years can be seen in Fig. A2.

As expected, due to the strong correlation of the SWE maxima with mean snow depth, the largest snow loads are located in the mountainous areas of Austria. Although the unrealistic mean snow depth values of SNOWGRID are masked, the model produces a number of 59 (0.06%) unrealistic snow load values larger than $25\,\mathrm{kN\,m^{-2}}$ in an altitude range between 1500 and 3700 m~~a depth range between 1500 and 3700 m a.s.l.~~. For a model that would be seriously used e.g. in general risk assessment or structural design, this problem could possibly be tackled with a non-linear relation between SWE maxima and mean snow depth or altitude. This is, however, beyond the scope of this study. Note, that in the actual Austrian standard (Austrian Standards Institute, 2018) there are no normative snow load values defined above $1500\,\mathrm{m}$ altitude.

All but two locations of the Austrian SWE measurement series that were used for calibration *and* validation of the ΔSNOW.MODEL (see Sect. 2.3.1) are included in the dataset used to fit the spatial model in Sect. C2. Those two stations, Holzgau and Felbertauern with 14 years of SWE observations each, are used to qualitatively compare (1) the spatial model fitted in Sect. C2, (2) SWE extremes modeled from daily snow depths with the ΔSNOW.MODEL, and (3) extremes computed "directly" from (ca. weekly) observed SWE values. Figure A3 gives an idea of the model performance at stations Holzgau and Felbertauern (see Figs. A1 and A2 for their locations)~~(see Fig. A2 for their locations)~~. For the lower-lying station Holzgau (1100 m)

~~(1100 m a.s.l.)~~ all three variants overlap very well. The 50-year return level is $4.65\,\mathrm{kN\,m^{-2}}$ for the smooth spatial model, $4.72\,\mathrm{kN\,m^{-2}}$ for the ΔSNOW.MODEL, and $4.8\,\mathrm{kN\,m^{-2}}$ for the observations. Note, that the latter stem from weekly observations and, therefore, not necessarily reflect the true yearly maxima, which naturally must be equal or slightly higher. By the way, the corresponding value of s_k from the Austrian snow load standard for Holzgau is $6.3\,\mathrm{kN\,m^{-2}}$ (Austrian Standards Institute (2018); accessible online at eHORA (2006)).

For the higher station Felbertauern $(1650\,\mathrm{m})$~~(1650 m a.s.l.)~~ the agreement between SWE from the ΔSNOW.MODEL and observed values is again very good. However, their GEV fits differ significantly. While the fit to the observations shows a negative shape parameter of $\xi = -0.1$, the fit to the values modeled with the ΔSNOW.MODEL gives a positive shape parameter of $\xi = 0.1$, leading to much larger return levels for higher recurrence times~~due to the Fréchet-like distribution~~. It should be pointed out that the GEV fits based on ΔSNOW.MODEL simulations and observations are unreliable, given the short data sample of only 14 yearly maxima. Indeed, by using a sample size of 43 years and borrowing strength from neighboring stations, the spatial model provides the best fit to observations as well as modeled SWE values. The 50-year snow load return values are $6.4\,\mathrm{kN\,m^{-2}}$ for the spatial model, $6.8\,\mathrm{kN\,m^{-2}}$ for the ΔSNOW.MODEL, and $5.7\,\mathrm{kN\,m^{-2}}$ for the fit to the observations. No normative value is defined for Felbertauern because it is situated higher than $1500\,\mathrm{m}$~~1500 m a.s.l.~~ (Austrian Standards Institute, 2018).

Author contributions. MW rose and led the project, structured and managed it. He was a key figure in developing and designing the snow model, and he did most of the writing. HS developed, coded and calibrated the model. He wrote the R-package and helped writing the paper, particularly the application example in the appendix. SG developed early versions of the model and its code.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors want to acknowledge T. Hell (Dept. of Mathematics, Univ. of Innsbruck, Austria) who substantially helped with the *Scaling module* (Sect. 2.2.2 and Appendix B). Thanks also go to the Hydrographic Service of Tyrol (Austria) which provided part of the data. A. Radlherr and J. Staudacher (ZAMG, Austria) is thanked for vivid discussions and proof-reading. Not least, we want to acknowledge M. Theurl and J. Abermann from the Univ. of Graz (Austria) for porting the ΔSNOW.MODEL's R-code to Phyton.

Financial support. This work was embedded in the project "Schneelast.Reform", funded by the Austrian Research Promotion Agency (FFG) and the Austrian Economic Chamber (WKO), in particular by their Association of the Austrian Wood Industries (FV Holzindustrie).

Review statement. This paper was edited by Markus Weiler and reviewed by three anonymous referees. The authors highly appreciate their comments and suggestions since they significantly improved the manuscript.

References

Akaike, H.: A new look at the statistical model identification, IEEE Transactions on Automatic Control, 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.

Armstrong, R. L. and Brun, E.: Snow and climate: physical processes, surface energy exchange and modeling, Cambridge University Press, Cambridge, 2010.

Austrian Standards Institute: ÖNORM B 1991-1-3:2018-12-01, 2018.

Avanzi, F., De Michele, C., and Ghezzi, A.: On the performances of empirical regressions for the estimation of bulk snow density, Geografia Fisica e Dinamica Quaternaria, pp. 105–112, https://doi.org/10.4461/GFDQ.2015.38.10, 2015.

Blanchet, J. and Davison, A.: Spatial modeling of extreme snow depth, The Annals of Applied Statistics, 5, 1699–1725, https://doi.org/10.1214/11-AOAS464SUPP, 2011.

Blanchet, J. and Lehning, M.: Mapping snow depth return levels: smooth spatial modeling versus station interpolation, Hydrology and Earth System Sciences, 14, 2527–2544, https://doi.org/10.5194/hess-14-2527-2010, 2010.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific Computing, 16, 1190–1208, https://doi.org/10.1137/0916069, 1995.

Coles, S.: An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer-Verlag, London, 2001.

De Michele, C., Avanzi, F., Ghezzi, A., and Jommi, C.: Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model, The Cryosphere, 7, 433–444, https://doi.org/10.5194/tc-7-433-2013, 2013.

Deems, J. S., Painter, T. H., and Finnegan, D. C.: Lidar measurement of snow depth: a review, Journal of Glaciology, 59, 467–479, https://doi.org/10.3189/2013JoG12J154, 2013.

Denoth, A.: The Pendular-Funicular Liquid Transition and Snow Metamorphism, Journal of Glaciology, 28, 357–364, https://doi.org/10.3189/S0022143000011692, 1982.

Dietz, A., Kuenzer, C., Gessner, U., and Dech, S.: Remote Sensing of Snow – a Review of available methods, International Journal of Remote Sensing, 33, 4094–4134, https://doi.org/10.1080/01431161.2011.640964, 2012.

Dixon, D. and Boon, S.: Comparison of the SnowHydro snow sampler with existing snow tube designs, Hydrological Processes, 26, 2555–2562, https://doi.org/10.1002/hyp.9317, 2012.

Egli, L., Jonas, T., and Meister, R.: Comparison of different automatic methods for estimating snow water equivalent, Cold Regions Science and Technology, 57, 107–115, https://doi.org/https://doi.org/10.1016/j.coldregions.2009.02.008, 2009.

eHORA: Natural Hazard Overview & Risk Assessment Austria, https://hora.gv.at/, [Online; accessed 13-March-2020], 2006.

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

European Committee for Standardization: EN 1991-1-3:2003/A1:2015, 2015.

Fierz, C., Armstrong, R., Durand, Y., Etchevers, P., Greene, E., McClung, D., Nishimura, K., Satyawali, P., and Sokratov, S.: The International Classification for Seasonal Snow on the Ground, IHP-VII Technical Documents in Hydrology, 83, http://unesdoc.unesco.org/images/0018/001864/186462e.pdf, 2009.

Garvelmann, J., Pohl, S., and Weiler, M.: From observation to the quantification of snow processes with a time-lapse camera network, Hydrology and Earth System Sciences, 17, 1415–1429, https://doi.org/10.5194/hess-17-1415-2013, 2013.

Goodison, B. E., Ferguson, H. L., and McKay, G. A.: Measurement and data analysis, in: Handbook of Snow: Principles, Processes, Management and Use, edited by Gray, D. M. and Male, D. H., pp. 191–274, Pergamon Press, Toronto, Canada, 1981.

Gruber, S.: Modelling snow water equivalent based on daily snow depths, Master's thesis, University of Innsbruck, 2014.

Guyennon, N., Valt, M., Salerno, F., Petrangeli, A. B., and Romano, E.: Estimating the snow water equivalent from snow depth measurements in the Italian Alps, Cold Regions Science and Technology, 167, https://doi.org/10.1016/j.coldregions.2019.102859, 2019.

Haberkorn, A.: European Snow Booklet – an Inventory of Snow Measurements in Europe, https://doi.org/10.16904/envidat.59, 2019.

Heilig, A., Schneebeli, M., and Eisen, O.: Upward-looking ground-penetrating radar for monitoring snowpack stratigraphy, Cold Regions Science and Technology, 59, 152–162, https://doi.org/10.1016/j.coldregions.2009.07.008, 2009.

Helfricht, K., Hartl, L., Koch, R., Marty, C., and Olefs, M.: Obtaining sub-daily new snow density from automated measurements in high mountain regions, Hydrology and Earth System Sciences, 22, 2655–2668, https://doi.org/10.5194/hess-22-2655-2018, 2018.

Hill, D. F., Burakowski, E. A., Crumley, R. L., Keon, J., Hu, J. M., Arendt, A. A., Wikstrom Jones, K., and Wolken, G. J.: Converting snow depth to snow water equivalent using climatological variables, The Cryosphere, 13, 1767–1784, https://doi.org/10.5194/tc-13-1767-2019, 2019.

International Organization for Standardization: ISO 4355:2013-12-01, 2013.

Johnson, J. B., Gelvin, A. B., Duvoy, P., Schaefer, G. L., Poole, G., and Horton, G. D.: Performance characteristics of a new electronic snow water equivalent sensor in different climates, Hydrological Processes, 29, 1418–1433, https://doi.org/10.1002/hyp.10211, 2015.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, Journal of Hydrology, 378, 161–167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Jordan, R.: A One-Dimensional Temperature Model for a Snow Cover: Technical documentation for SNTHERM.89, Tech. rep., Corps of Engineers, U.S. Army Cold Regions Research & Engineering Laboratory, 1991.

Jordan, R., Albert, M., and Brun, E.: Physical Processes within the snow cover and their parametrization, in: Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling, pp. 12–69, Cambridge University Press, Cambridge, 2010.

Keeler, C.: Some Physical Properties of Alpine Snow, Tech. rep., Corps of Engineers, U.S. Army Cold Regions Research & Engineering Laboratory, 1969.

Kinar, N. J. and Pomeroy, J. W.: Measurement of the physical properties of the snowpack, Reviews of Geophysics, 53, 481–544, https://doi.org/10.1002/2015RG000481, 2015.

Koch, F., Henkel, P., Appel, F., Schmid, L., Bach, H., Lamm, M., Prasch, M., Schweizer, J., and Mauser, W.: Retrieval of Snow Water Equivalent, Liquid Water Content, and Snow Height of Dry and Wet Snow by Combining GPS Signal Attenuation and Time Delay, Water Resources Research, https://doi.org/10.1029/2018WR024431, 2019.

Kojima, K.: Densification of Seasonal Snow Cover, in: Physics of Snow and Ice : proceedings, vol. 1 of 2, pp. 929–952, Sapporo, Japan, http://hdl.handle.net/2115/20351, 1967.

Langlois, A., Kohn, J., Royer, A., Cliche, P., Brucker, L., Picard, G., Fily, M., Derksen, C., and Willemet, J. M.: Simulation of Snow Water Equivalent (SWE) Using Thermodynamic Snow Models in Québec, Canada, Journal of Hydrometeorology, 10, 1447–1463, https://doi.org/10.1175/2009JHM1154.1, 2009.

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., and Zimmerli, M.: SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations, Cold Regions Science and Technology, 30, 145–157, https://doi.org/10.1016/S0165-232X(99)00022-1, 1999.

Lehning, M., Bartelt, P., Brown, R., Fierz, C., and Satyawali, P.: A physical SNOWPACK model for the Swiss Avalanche Warning Services. Part II: Snow Microstructure, Cold Regions Science and Technology, 35, 147–167, 2002.

Leppänen, L., Kontu, A., and Pulliainen, J.: Automated Measurements of Snow on the Ground in Sodankylä, Geophysica, 53, 45–64, http://www.geophysica.fi/pdf/geophysica_2018_53_leppanen.pdf, 2018.

López-Moreno, J. I., Leppänen, L., Luks, B., Holko, L., Picard, G., Sanmiguel-Vallelado, A., Alonso-González, E., Finger, D. C., Arslan, A. N., Gillemot, K., Sensoy, A., Sorman, A., Ertaş, M. C., Fassnacht, S. R., Fierz, C., and Marty, C.: Intercomparison of measurements of bulk snow density and water equivalent of snow cover with snow core samplers: Instrumental bias and variability induced by observers, Hydrological Processes, 34, 3120–3133, https://doi.org/10.1002/hyp.13785, 2020.

Mair, E., Leitinger, G., Della Chiesa, S., Niedrist, G., Tappeiner, U., and Bertoldi, G.: A simple method to combine snow height and meteorological observations to estimate winter precipitation at sub-daily resolution, Hydrological Sciences Journal, 61, 2050–2060, https://doi.org/10.1080/02626667.2015.1081203, 2016.

Marks, D. G., Kimball, J. S., Tingey, D., and Link, T. E.: The Sensitivity of Snowmelt Processes to Climate Conditions and Forest Cover during Rain-on-Snow: A Case Study of the 1996 Pacific Northwest Flood, Hydrological Processes, 12, 1569–1587, 1998.

Martinec, J.: Zimni prognosy s pouzitim radioisotopu (Winter forecasts with the use of radioisotopes), Vltavska kaskada (The Vltava reservoir system), VUV Praha-Podbab, pp. 45–60, 1956.

Martinec, J.: Expected snow loads on structures from incomplete hydrological data, Journal of Glaciology, 19, 185–195, https://doi.org/10.3189/S0022143000029270, 1977.

Martinec, J. and Rango, A.: Indirect evaluation of snow reserves in mountain basins, in: Snow, Hydrology and Forests in High Alpine Areas, 205, pp. 111–120, 1991.

Marty, C.: GCOS SWE data from 11 stations in Switzerland, https://doi.org/10.16904/15, type: dataset, 2017.

McCreight, J. L. and Small, E. E.: Modeling bulk density and snow water equivalent using daily snow depth observations, The Cryosphere, 8, 521–536, https://doi.org/10.5194/tc-8-521-2014, 2014.

Mitterer, C., Hirashima, H., and Schweizer, J.: Wet-snow instabilities: comparison of measured and modelled liquid water content and snow stratigraphy, Annals of Glaciology, 52, 201–208, https://doi.org/10.3189/172756411797252077, 2011.

Mizukami, N. and Perica, S.: Spatiotemporal Characteristics of Snowpack Density in the Mountainous Regions of the Western United States, Journal of Hydrometeorology, 9, 1416–1426, https://doi.org/10.1175/2008JHM981.1, 2008.

Nash, J. C.: On Best Practice Optimization Methods in R, Journal of Statistical Software, 60, 1–14, http://www.jstatsoft.org/v60/i02/, 2014.

Olefs, M., Schöner, W., Suklitsch, M., Wittmann, C., Niedermoser, B., Neururer, A., and Wurzer, A.: SNOWGRID – A New Operational Snow Cover Model in Austria, International Snow Science Workshop Grenoble – Chamonix Mont-Blanc – October 07-11, 2013, pp. 38–45, https://arc.lib.montana.edu/snow-science/item/1785, 2013.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., and Winstral, A.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, Remote Sensing of Environment, 184, 139–152, https://doi.org/10.1016/j.rse.2016.06.018, 2016.

Parajka, J., Haas, P., Kirnbauer, R., Jansa, J., and Blöschl, G.: Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale, Hydrological Processes, 26, 3327–3337, https://doi.org/10.1002/hyp.8389, 2012.

Paterson, W. S. B.: The physics of glaciers, Butterworth/Heinemann, Oxford ; Woburn, MA, 3rd edn., 1998.

Pistocchi, A.: Simple estimation of snow density in an Alpine region, Journal of Hydrology: Regional Studies, 6, 82–89, https://doi.org/10.1016/j.ejrh.2016.03.004, 2016.

Powell, M.: The BOBYQA algorithm for bound constrained optimization without derivatives, Report DAMTP 2009/NA06, University of Cambridge, https://www.nag.co.uk/IndustryArticles/bound_optimization_quadratic_approximation.pdf, 2009.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2019.

Rohrer, M. and Braun, L.: Long-Term Records of Snow Cover Water Equivalent in the Swiss Alps, Nordic Hydrology, 25, 65–78, https://doi.org/10.2166/nh.1994.0020, 1994.

Sandells, M., Flerchinger, G. N., Gurney, R., and Marks, D. G.: Simulation of snow and soil water content as a basis for satellite retrievals, Hydrology Research, 43, 720–735, https://doi.org/10.2166/nh.2012.028, 2012.

Schattan, P., Köhli, M., Schrön, M., Baroni, G., and Oswald, S. E.: Sensing Area-Average Snow Water Equivalent with Cosmic-Ray Neutrons: The Influence of Fractional Snow Cover, Water Resources Research, 55, 10 796–10 812, https://doi.org/10.1029/2019WR025647, 2019.

Schellander, H. and Hell, T.: Modeling snow depth extremes in Austria, Natural Hazards, 94, 1367–1389, https://doi.org/10.1007/s11069-018-3481-y, 2018.

Seibert, P., Frank, A., and Formayer, H.: Synoptic and regional patterns of heavy precipitation in Austria, Theoretical and Applied Climatology, 87, 139–153, https://doi.org/10.1007/s00704-006-0198-8, 2007.

Smith, C. D., Kontu, A., Laffin, R., and Pomeroy, J. W.: An assessment of two automated snow water equivalent instruments during the WMO Solid Precipitation Intercomparison Experiment, The Cryosphere, 11, 101–116, https://doi.org/10.5194/tc-11-101-2017, 2017.

Smyth, E. J., Raleigh, M. S., and Small, E. E.: Particle Filter Data Assimilation of Monthly Snow Depth Observations Improves Estimation of Snow Density and SWE, Water Resources Research, 55, 1296–1311, https://doi.org/10.1029/2018WR023400, 2019.

Sturm, M. and Holmgren, J.: Differences in compaction behavior of three climate classes of snow, Annals of Glaciology, 26, 125–130, 1998.

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., and Lea, J.: Estimating Snow Water Equivalent Using Snow Depth Data and Climate Classes, Journal of Hydrometeorology, 11, 1380–1394, https://doi.org/10.1175/2010JHM1202.1, 2010.

Valt, M., Romano, E., and Guyennon, N.: Snowcover density and snow water equivalent in the Italian Alps, in: Proceedings ISSW 2018, International Snow Science Workshop, Innsbruck, Austria, 2018.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J. M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geoscientific Model Development, 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Wastl, C.: Klimatologische Analyse von orographisch beeinflussten Niederschlagsstrukturen im Alpenraum, Phd thesis, Ludwig-Maximilians-Universität München, http://nbn-resolving.de/urn:nbn:de:bvb:19-95453, 2008.

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., and Lehning, M.: Verification of the multi-layer SNOWPACK model with different water transport schemes, The Cryosphere, 9, 2271–2293, https://doi.org/10.5194/tc-9-2271-2015, 2015.

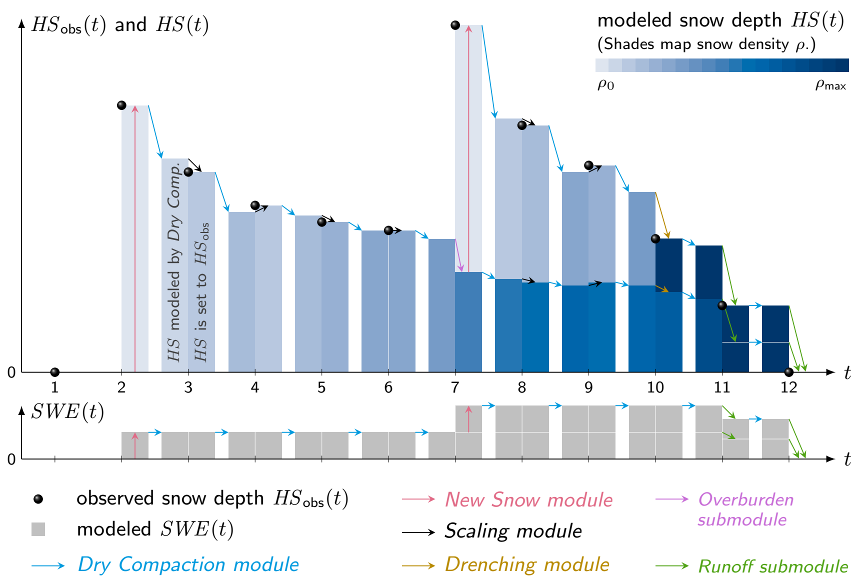

Figure 1. Schematic figure of the ΔSNOW.MODEL's principles. See text for more details. ~~At time $t = 1$ no snow is observed (black bullet at $HS_{\mathrm{obs}}(1) = 0$) and – consequently – no snow is modeled. At $t = 2$ initially no snow is modeled. However, snow is observed ($HS_{\mathrm{obs}}(2) > 0$) and, thus, model snow depth $HS(2)$ is set to the observed value by the ΔSNOW.MODEL's *New Snow Module* (pink arrow) and a certain SWE (green boxes) is assigned to this new snow layer. For $t = 3$ densification by the *Dry Metamorphism Module* (Dry Met.) is modeled (light blue arrow). Snow density (ρ, shown as bluish shades) is enhanced, but SWE stays constant. Still, the model and observation slightly disagree ($|\Delta HS| < |\tau|$). The *Scaling Module* solves this issue (black arrow): At $t = 3$ (and also at $t = 4, 5, 6, 8$, and 9) modeled $HS(t)$ is scaled to equal $HS_{\mathrm{obs}}(t)$ with respective consequences for the snow density, but not altering SWE. $HS_{\mathrm{obs}}(7)$ is way higher than $HS(7)$, the *New Snow Module* builds up a new layer and raises SWE accordingly. The *New Snow Module* also treats the unsteady and strong compaction of the underlying layer(s) due to overburden snow load (purple arrow). At $t = 10$ observed snow depth is significantly smaller than $HS(10)$, the *Drenching Module* (brown arrows) wettens the layers from top to bottom until ρ_{max} is reached. Figuratively, the layers get water-saturated, however, at $t = 10$ not all layers reach ρ_{max}: No mass loss is requested by the model, and SWE stays constant. At $t = 11$ the Drenching Module necessitates mass loss by runoff (cyan arrows) as all layer densities would have to exceed ρ_{max} to fulfill $HS(t) = HS_{\mathrm{obs}}(t)$, but this is not possible in the ΔSNOW.MODEL. All layers are set to ρ_{max} and they "get cut" by an appropriate amount of depth and mass, respectively, depending on their thickness: Thick layers contribute more to the mass loss than thin ones. In the end, at $t = 12$, no snow is observed anymore and final runoff is modeled.~~

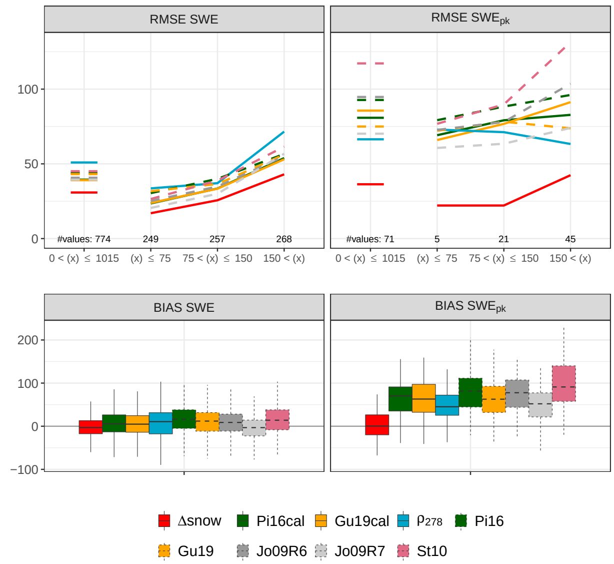

Figure 2. Root mean square errors (RMSE) and biases (BIAS) between the ΔSNOW.MODEL and different empirical regression models and the SWE_{val} observations.~~Validation results for the model biases and root mean square errors (RMSE). The plots show the results for the models applied to the SWE_{val} data.~~ The ΔSNOW.MODEL, Pistocchi (2016)'s and Guyennon et al. (2019)'s models, as well as~~Pistocchi (2016), Guyennon et al. (2019), and~~ the "constant density approach" were calibrated with SWE_{cal} data (Δsnow, Pi16cal, Gu19cal, ρ_{278}; upper panels, solid lines). Dashed lines indicate the Pistocchi (2016), the Guyennon et al. (2019), the Jonas et al. (2009), and the Sturm et al. (2010) models with their standard parameters (Pi16, Gu19, Jo09R6, Jo09R7, and St10). Jo09R6 and Jo09R7 together illustrate the maximum possible spread of the Jonas et al. (2009) model since Region 6 (R6) and Region 7 (R7) are characterized by the highest and lowest "region-specific offset", respectively. The upper left panel shows RMSEs for all SWE_{val} values (short horizontal lines) as well as for three SWE classes: $SWE \leq 75$, $SWE > 150$, and intermediate. Analogously for SWE_{pk} (upper right panel). The boxes for the biases (lower panels) encompass 774~~761~~ values (left panel~~s~~, SWE) and 71~~73~~ values (right panel~~s~~, SWE_{pk}) and spread from the 25%- to the 75%-quantile, the whiskers indicate 1.5 times the interquartile range.~~minima and maxima~~. ~~The ΔSNOW.MODEL does not behave significantly better for the bias of all SWE (upper left), however, its performance for seasonal maxima SWE_{pk} (upper right) as well as for the mean errors (lower panels) is very convincing. (Note the different y-axes scalings.)~~ Units are $\text{kg}\,\text{m}^{-2}$.

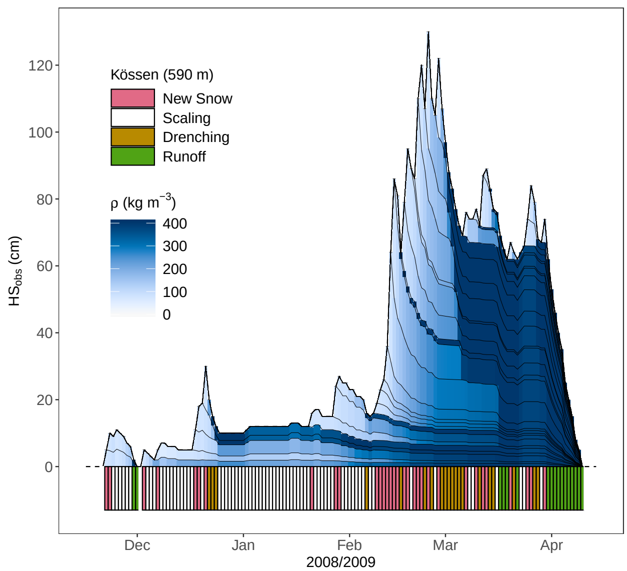

Figure 3. Winter of 2008/09 in Kössen (Northern Alps, Austria) ~~The winter of 2008/09 in Kössen, a low-lying but snowy station at the Northern Alps,~~ portrays density evolution ~~(shades)~~ as simulated by the ΔSNOW.MODEL. The regular~~gapless~~, daily snow depth record is used as only model input. The *New Snow module*, the *Scaling module* and the *Drenching module*, as well as the *Runoff submodule* ~~Three (out of four) modules of the ΔSNOW.MODEL~~ are depicted in colors at the bottom, whenever activated~~: Drenching, New Snow, and Scaling (The Dry *Compaction module* is activated at every point in time.) Runoff is a subcategory of the *Drenching module*~~. Note, the ΔSNOW.MODEL is not intended to simulate individual layers, but to calculate daily SWE, SWE_{pk}, and *mean* daily bulk density as accurate as possible. ~~This figure illustrates what happens during the modeling. However, the aim of the ΔSNOW.MODEL is to get daily SWE and SWE_{pk} right – i.e. *mean* daily bulk density, not layer-individual densities.~~ Descriptions and discussions of some features are given in the text.

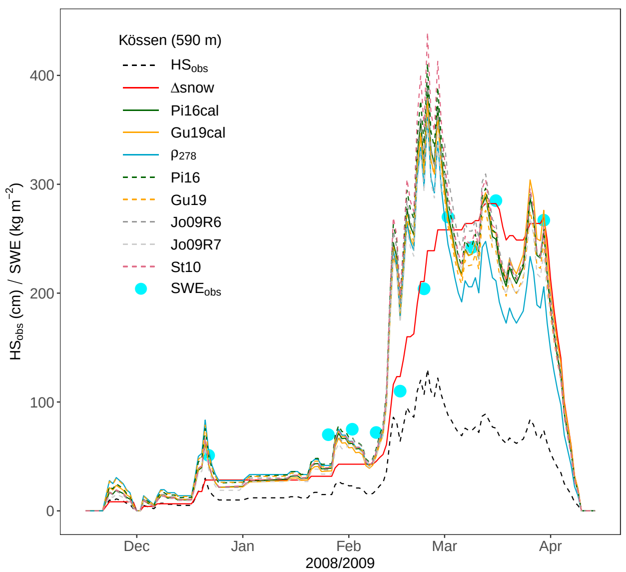

Figure 4. SWE simulations and observations (~~mostly weekly;~~ SWE_{obs}) for the winter 2008/09 in Kössen (cf. Fig. 3). Details and ~~Model~~ abbreviations are given in the text (Sect. 3.2) and summarized in Fig. 2.~~ See Table 4 for values, and consider the note in its caption. This plot is an illustration of the ΔSNOW.MODEL performance during a distinct winter and outlines important features, which are addressed in the text~~.

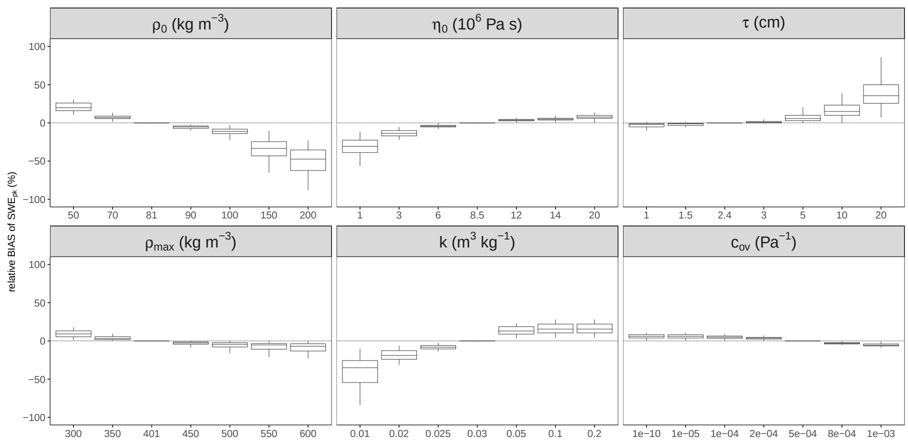

Figure 5. Sensitivity of SWE_{pk} to changes in model parameters. The "relative BIAS of SWE_{pk}"~~relative bias of SWE_{pk}~~ is defined as the difference between SWE_{pk} with best-fitted values and SWE_{pk} with changed parameters (while all others are kept unchanged), divided by the best-fitted SWE_{pk}. The boxes comprise SWE_{pk} of all stations and all years of the validation data set SWE_{val} (71~~73~~ values) and display medians as well as 25% and 75% percentiles, the whiskers indicate 1.5 times the interquartile range.~~minimum and maximum biases. (Parameter k_{ov} behaves unremarkably – similar to c_{ov} – and is not shown here.)~~ Details and analysis see text.

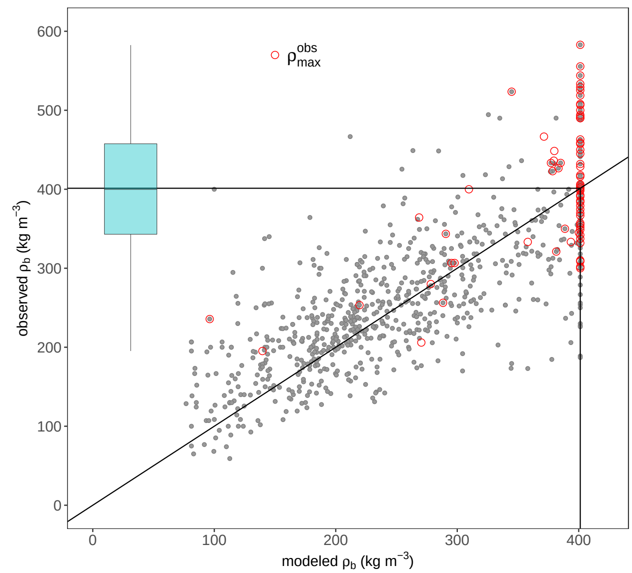

Figure 6. Scatter plot of all modeled bulk snow densities ρ_b versus all observed ρ_b from the validation data set. (SWE_val, 767 data pairs. Seven observations, which are higher than $600\,\mathrm{kg\,m^{-3}}$, were ignored due to implausibility.) Red c~~C~~ircles reflect the 71~~73~~ observed yearly maxima ($\rho^\mathrm{obs}_\mathrm{max}$), most of them occur when also modeled snowpack is~~all modeled snow layers are saturated~~ at $\rho_\mathrm{max} = 401\,\mathrm{kg\,m^{-3}}$. The box plot shows the distribution of $\rho^\mathrm{obs}_\mathrm{max}$ with median, 25% and 75% percentiles, and whiskers at 1.5 times the interquartile range.~~: The median is at $400\,\mathrm{kg\,m^{-3}}$. (This round value is somewhat fortuitously and should not be taken too seriously.) The horizontal line compares it to the ΔSNOW.MODEL's maximum density at ρ_max.~~

Table 1. Different types of SWE models, categorized by their essential input. TD, SE, and ERMs are abbreviations for thermodynamic, semi-empirical, and empirical regression models, respectively.

essential input	TD	SE	ERMs
HS (single values)			x
HS (regular records)	x[a]	x	
one or more atmospheric variable(s)	x		
date	(x)[b]		x
location parameters[c]	(x)[b]		x

[a]or another precipitation input
[b]only essential in some cases, e.g. for parameterizations
[c]altitude, regional climate, etc.

Table 2. Summary of compaction processes and processes forcing mass changes that are integrated in the modules and submodules of the ΔSNOW.MODEL, and of processes that are ignored.

module	process
New Snow	significant rise of HS, enhanced compaction due to overburden load (*Overburden submodule*)
Dry Compaction	significant decline of HS due to dry metamorphism[a] and/or deformation[a]
Drenching	significant decline of HS due to wet metamorphism[a], runoff through melt (*Runoff submodule*)
Scaling	adjustments to small changes of HS within threshold deviation τ
ignored:	snow drift compaction[a] and mass changes due to:
	rain-on-snow, runoff during snowfalls, wind drift, small snowfalls, sublimation and deposition

[a]terminology follows Jordan et al. (2010)

Table 3. The seven parameters of the ΔSNOW.MODEL. The last column depicts model sensitivity to changes in the density parameters. The respective gradients are means over the whole calibration ranges. ~~Detailed information is found in the text.~~

Parameter par	unit	optimal value	calibration range	literature range	sensitivity $\frac{\delta SWE_{\text{pk}}\,[\mathrm{kg\,m^{-2}}]}{\delta par}$
ρ_0	$\mathrm{kg\,m^{-3}}$	**81**	50-200	75[a], 10-350 (70-110)[b]	+0.37 (+0.50[†])
ρ_{max}	$\mathrm{kg\,m^{-3}}$	**401**	300-600	450[c],217-598[d], 400-800[e]	+0.24
η_0	10^6 Pa s	**8.5**~~8.51~~	1-20	8.5[a], 6[f], 7.62237[g]	not calc.
k	$\mathrm{m^3\,kg^{-1}}$	**0.030**~~0.0299~~	0.01-0.2	0.011-0.08[a], 0.185[h], 0.023[f,g], 0.021[i]	not calc.
τ	cm	**2.4**~~2.36~~	1-20	-	not calc.
c_{ov}	$10^{-4}\,\mathrm{Pa^{-1}}$	**5.1**~~5.10~~	0-10	-	not calc.
k_{ov}	-	**0.38**~~0.379~~	0.01-10	-	not calc.

[a]Sturm and Holmgren (1998), [b]Helfricht et al. (2018) with range for means in brackets, [c]Rohrer and Braun (1994), [d]Sturm et al. (2010), [e]Paterson (1998), [f]Jordan et al. (2010), [g]Vionnet et al. (2012), [h]Keeler (1969), [i]Jordan (1991). See Sect. 2.3 for more details. [†]The value in brackets is the gradient taken from the smaller window between 70 and 90 $\mathrm{kg\,m^{-3}}$ (cf. Sect. 4.1).

Table 4. Overview on SWE accuracies of different models~~methods~~ and studies. ~~The numbers of this study are the median values, which are also depicted in Fig. 3.~~ The numbers in brackets represent the results for the example portrayed in Figs. 3 and 4 from station Kössen in 2008/09. ~~Note, the performance of the ΔSNOW.MODEL of the example is quite ordinary, while other models do better on average.~~ Units are $\mathrm{kg\,m^{-2}}$ ~~except last column, which is in days,~~ TD is short for thermodynamic snow models. Model abbreviations see caption of Fig. 2.

source	model (version)	SWE BIAS	SWE RMSE	SWE MAE	SWE_{pk} BIAS	SWE_{pk} RMSE	~~SWE_{pk} BIAS [d]~~
this study	ΔSNOW.MODEL	−3.0 ~~−4.0~~	30.8~~23.9~~ (21)	21.9~~19.5~~	0.3~~2.3~~ (−3)	36.3~~23.1~~	~~0~~
	Gu19cal	4.8~~4.0~~	39.1~~31.3~~ (43)	27.6~~24.4~~	63.0~~67.3~~ (93)	85.6~~70.8~~	~~−6~~
	Pi16cal	5.6~~5.3~~	39.4~~32.9~~ (47)	28.1~~25.6~~	70.3~~71.0~~ (106)	80.8~~72.2~~	~~−6~~
	Jo09R7	−3.2 ~~−2.0~~	39.4~~26.6~~ (41)	27.3~~21.9~~	52.0~~56.5~~ (74)	70.2~~58.1~~	~~−2~~
	St10	14.0~~17.6~~	45.1~~37.1~~ (57)	32.6~~28.5~~	91.1~~95.1~~ (154)	117.2~~95.7~~	~~−11~~
	ρ_{278}	10.6~~14.8~~	50.9~~38.1~~ (51)	36.3~~31.2~~	45.2~~47.7~~ (77)	66.4~~53.5~~	~~−16~~
Guyennon et al. (2019)	Gu19			49.2			
	Pi16cal			50.6			
	Jo09cal			48.5			
	St10cal			51.0			
Jonas et al. (2009)	Jo09		50.9 – 53.2				
Sturm et al. (2010)	St10 (“alpine”)	29 ± 57					
Vionnet et al. (2012)	Crocus	−17.3	39.7				
Langlois et al. (2009)	Crocus	−7.9 to −5.4	10.8 – 12.5				
	SNTHERM	9 to 18.1	18.3 – 19.3				
	SNOWPACK	−0.1 to 5.6	7.4 – 14.5				
Egli et al. (2009)	SNOWPACK		56				
Wever et al. (2015)	SNOWPACK		ca. 39.5				
Sandells et al. (2012)	SNOBAL		30 – 49			17 – 44[a]	
Essery et al. (2013)	various TD[b]		23 – 77				

[a]This is not RMSE of SWE_{pk}, but RMSE “from establishment of snowpack to SWE_{pk}”. [b]See Essery et al. (2013)’s Table 10: RMSE for up to 1700 uncalibrated and calibrated simulations.

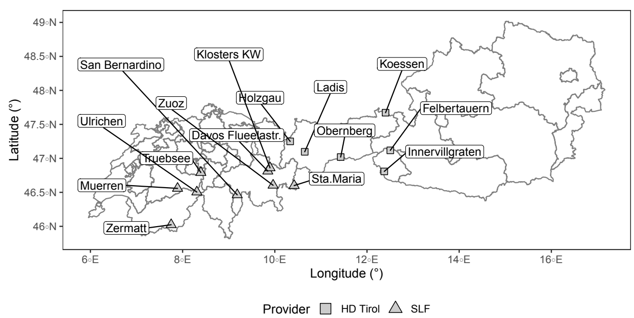

Figure A1. Locations of the stations used for calibration and validation. Austrian stations are operated by the Hydrographic Service of Tyrol (HD Tirol), the Swiss stations by the WSL Institute for Snow and Avalanche Research SLF. See Table A1 and text for more details.

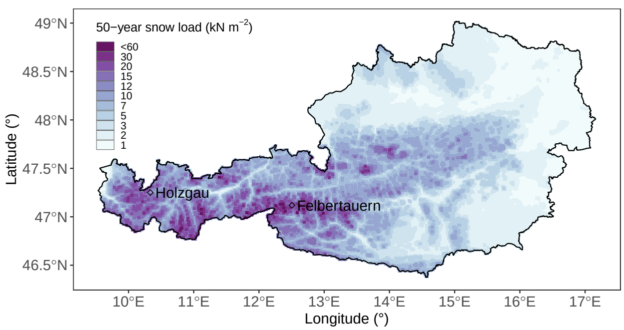

Figure A2. 50-year return levels of snow load in Austria. Two stations with SWE observations are outlined for a qualitative validation. This map bases on 214 snow depth records, ΔSNOW.MODEL derived SWE, and smooth spatial modeling of their extremes. ~~See text for details.~~

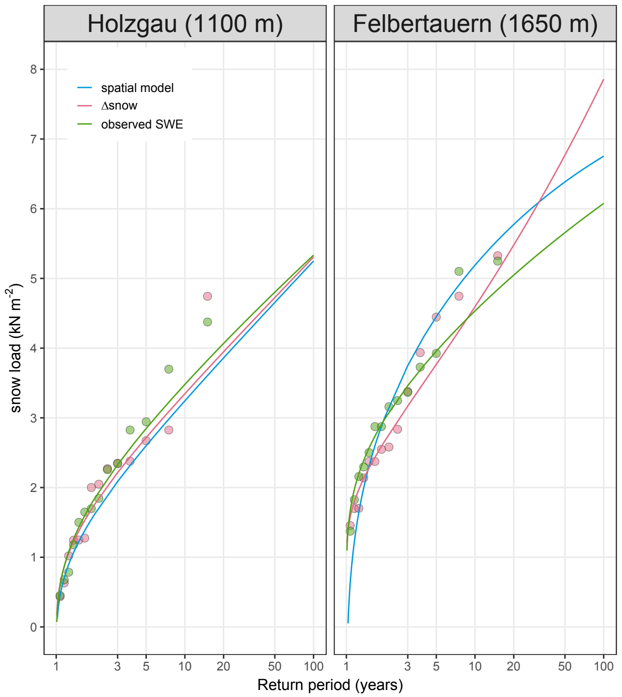

Figure A3. Return levels of snow load at stations Holzgau (left) and Felbertauern. Return periods in years are shown on the logarithmic x-axis. The blue line shows return levels obtained with the spatial extreme value model, pink bullets and lines depict yearly maxima and the GEV fit of SWE values modeled from daily snow depths with the ΔSNOW.MODEL, and green colors represent yearly SWE maxima and the corresponding GEV fit from (ca. weekly) observations.

Table A1. Overview of stations with daily snow depths record and about weekly/biweekly (Austria/Switzerland) manual SWE observations which were used for calibration and validation. $\#_{\text{cal}}^{SWE}$ and $\#_{\text{val}}^{SWE}$ give the numbers of respective manual SWE observations. Stations #1 to #6 are located in the Austrian province of Tyrol, #5 and #6 are in the sub-province of Eastern Tyrol; all operated by the Hydrographic Service of Tyrol. Swiss stations #7 to #15 are operated by the WSL Institute for Snow and Avalanche Research SLF. Compare Fig. A1. The data sources are Gruber (2014) and Marty (2017).

#	station name	lon [°]	lat [°]	alt [m]	$\#_{\text{cal}}^{SWE}$	$\#_{\text{val}}^{SWE}$	calibration seasons[a]	valibration seasons[a]
1	Holzgau	10.333300	47.25000	1100	116	100	7 odd in 1999-2011	7 even in 1998-2010
2	Ladis	10.649200	47.09690	1350	83	66	7 odd in 1999-2011	6~~7~~ even in 1998-2010[b]
3	Obernberg	11.429200	47.01940	1360	105	88	7 odd in 1999-2011	7 even in 1998-2010
4	Koessen	12.402800	47.67170	590	87	70	7 odd in 1999-2011	6~~7~~ even in 1998-2010[b]
5	Felbertauern	12.505600	47.11810	1650	126	114	7 odd in 1999-2011	7 even in 1998-2010
6	Innervillgraten	12.375000	46.80830	1400	96	115	7 odd in 1999-2011	7 even in 1998-2010
7	Muerren	7.890193	46.55818	1650	37	27	2009,2012,2015,2017	2006,2011,2014,2016
8	Truebsee	8.395291	46.79121	1780	4	11	2016	2015,2017
9	Ulrichen	8.308283	46.50461	1350	24	23	2009,2013,2015,2017	2007,2011,2014,2016
10	Zermatt	7.751165	46.02340	1600	47	76	1961,1963 and 7 even in 2004-2016	3 even 1960-1964, 7 odd in 2005-2017
11	Davos Flueelastr.	9.848163	46.81255	1560	8	19	2012	2008,2017
12	Klosters KW	9.895973	46.86058	1200	12~~0~~	22~~10~~	1999	1998,2017
13	San Bernardino	9.184634	46.46326	1640	11	14~~13~~	2007	2006,2014
14	Sta.Maria	10.419344	46.59981	1415	0	8	-	1969
15	Zuoz	9.962676	46.60433	1710	24	21	2011,2013,2015,2017	2006,2012,2014,2016
$\sum$					780	774	67	71

[a]Indicated years mark the start of respective winter seasons. [b]2006 is missing.