Snowfall Model Validation Using Surface Observations and an Optimal Estimation Snowfall Retrieval

FRANZISKA HELLMUTH,^a BJØRG JENNY KOKKVOLL ENGBAHL,^{b,a} TRUDE STORELVMØ,^a ROBERT O. DAVID,^a
AND STEVEN J. COOPER^c

^a *Department of Geosciences, University of Oslo, Oslo, Norway*

^b *Norwegian Meteorological Institute, Oslo, Norway*

^c *University of Utah, Salt Lake City, Utah*

(Manuscript received 24 November 2020, in final form 9 July 2021)

ABSTRACT: In the winter, orographic precipitation falls as snow in the mid- to high latitudes where it causes avalanches, affects local infrastructure, or leads to flooding during the spring thaw. We present a technique to validate operational numerical weather prediction model simulations in complex terrain. The presented verification technique uses a combined retrieval approach to obtain surface snowfall accumulation and vertical profiles of snow water at the Haukelisetter test site, Norway. Both surface observations and vertical profiles of snow are used to validate model simulations from the Norwegian Meteorological Institute's operational forecast system and two simulations with adjusted cloud microphysics. Retrieved surface snowfall is validated against measurements conducted with a double-fence automated reference gauge (DFAR). In comparison, the optimal estimation snowfall retrieval produces +10.9% more surface snowfall than the DFAR. The predicted surface snowfall from the operational forecast model and two additional simulations with microphysical adjustments (CTRL and ICE-T) are overestimated at the surface with +41.0%, +43.8%, and +59.2%, respectively. Simultaneously, the CTRL and ICE-T simulations underestimate the mean snow water path by −1071.4% and −523.7%, respectively. The study shows that we would reach false conclusions only using surface accumulation or vertical snow water content profiles. These results highlight the need to combine ground-based in situ and vertically profiling remote sensing instruments to identify biases in numerical weather prediction.

KEYWORDS: Snowfall; Radars/Radar observations; Cloud parameterizations; Numerical weather prediction/forecasting; Parameterization; Mountain meteorology; Cloud microphysics; Complex terrain; Forecasting techniques; Freezing precipitation; In situ atmospheric observations; Measurements; Model comparison; Precipitation; Remote sensing; Snow; Winter/cool season

1. Introduction

Forecasting precipitation quantitatively is challenging, especially in complex terrain where the evaluation of forecast models is difficult due to the sparse distribution of precipitation gauges (Barstad and Smith 2005). Furthermore, coarse-resolution numerical weather prediction (NWP) models are less skillful in simulating orographic precipitation, especially over narrow topography (Gowan et al. 2018). This difficulty is unfortunate, as the forecasts of avalanches, glacier mass budgets, and flash floods depend on accurate models.

NWP models with ≤ 4 -km resolution are now increasingly capable of better representation of terrain, which is critical for predicting orographic precipitation (Colle and Mass 2000; Colle et al. 2005; Garvert et al. 2005; Schwartz 2014). More importantly, the higher model resolution allows for representing small-scale phenomena such as convective dynamics (Gowan et al. 2018), thus enabling more accurate simulations of the magnitude and location of the precipitation. Improved microphysical parameterizations can influence the distribution of latent heat release, hence generating potential vorticity,

which can affect storm evolution (Joos and Wernli 2012). Sensitivity tests varying the microphysical scheme of the Consortium for small-scale Modeling model showed that storm development depended on the correct vertical placement of the precipitation inside a modeled storm (Joos and Wernli 2012). Therefore, accurate precipitation observations at the surface and in the vertical are necessary to validate snowfall predictions from models.

However, measuring precipitation, especially snow, is a well-known challenge (Thériault et al. 2012; Wolff et al. 2013; Colli et al. 2015). Winter precipitation measurements can show more than 100% difference between different observing networks where the exact values depend upon regional snowfall characteristics (Kochendorfer et al. 2017). At high wind speeds, precipitation loss can be severe. Additionally, blowing surface snow can enter the gauge and artificially enhance the reported precipitation (Nitu et al. 2018). A double fence automated reference (DFAR) gauge provides a more accurate estimate of snow accumulation than single fence gauges during high wind speeds (Wolff et al. 2015; Kochendorfer et al. 2017). Nevertheless, the DFAR still has an undercatch depending on the wind speeds (Nitu et al. 2018). Such uncertainties must be considered when using gauge observations to evaluate model forecasts and remote sensing algorithm estimates of snowfall (Wolff et al. 2015).

Quantitative snowfall estimates from radar-based remote sensing techniques rely on deriving a snowfall rate (S) from a radar reflectivity (Z). However, studies have shown that the estimated value S depends critically upon the choice of

Denotes content that is immediately available upon publication as open access.

Corresponding author: Franziska Hellmuth, franziska.hellmuth@geo.uio.no

microphysical assumptions required for the inversion (Kulie and Bennartz 2009). Subsequent studies incorporated scene-specific snowflake microphysical property information into the retrieval scheme to better match environmental conditions (Wood et al. 2015). Cooper et al. (2017) used in situ observations of snowflake particle size distribution (PSD) and habit from the Multi-Angle Snow Camera (MASC) to improve radar-based estimated snowfall at Barrow, Alaska. This study exploited an optimal estimation snowfall retrieval (OESR) approach that combined radar reflectivities, in situ microphysical property estimates, and environmental information into a common retrieval framework to provide estimates of snowfall rate.

Schirle et al. (2019) modified the Cooper et al. (2017) algorithm to the instrumentation and environmental conditions at the Haukeliseter Test Site (HTS) as part of the High-Latitude Measurement of Snowfall (HiLaMS) campaign. This scheme is used in this work to estimate both surface snowfall rates and vertical profiles of snow water content (SWC).

To estimate forecast uncertainty, modeling centers utilize high-resolution ensemble prediction. For example, the Meteorological Cooperation on Operational Numerical Weather Prediction (MetCoOp) produces an Ensemble Prediction (MEPS) by perturbing the initial state for the deterministic run of the Applications of Research to Operations at Mesoscale (AROME)-MetCoOp model (Frogner et al. 2019).

Early versions of MEPS produced too much cloud ice with its default ICE3 cloud microphysics scheme. Thus, Müller et al. (2017) changed the ICE3 microphysics scheme to improve the representation of fog, low-level, and cirrus clouds. More recently, Engdahl et al. (2020b) showed that even with the improvements, the ICE3 microphysics scheme depleted supercooled liquid water too quickly and produced a surplus of snow and graupel. For this reason, Engdahl et al. (2020b) introduced a series of changes to the ICE3 scheme based on the parameterization development by Thompson et al. (2004, 2008). The modified microphysics include updated ice nucleation, riming/accretion parameterization, and changes to the rain class PSD. Idealized 1D experiments showed that the modified scheme prolonged the existence and produced higher amounts of supercooled liquid water (Engdahl et al. 2020b). Many studies with adjusted microphysical schemes, such as Liu et al. (2011), have validated surface precipitation forecasts to observations but not the vertical distribution of hydrometeors. A comparison of six microphysical parameterization schemes showed most parameterization schemes resemble each other in the upper troposphere but differ by a factor of 3 close to the surface. The Lin et al. (1983) scheme (similar to ICE3) shows low snow and cloud water, yet high graupel amounts in the atmosphere. Combined with high surface precipitation amounts, this indicates a high precipitation efficiency. Conversely, the Thompson et al. (2008) and Morrison et al. (2009) schemes have low graupel production. For this reason, the atmospheric snow and cloud liquid water contents are several times larger than in other simulations.

This study demonstrates the added value of validating snowfall in NWP models using surface observations and vertical profiles of snowfall obtained by the advanced techniques of Cooper et al. (2017) and Schirle et al. (2019) during winter

2016/17. We use DFAR snow observations, radar-based snowfall retrievals (Cooper et al. 2017; Schirle et al. 2019), and high-resolution forecast data from MEPS at the HTS. The instrumental setup at HTS is a unique opportunity to apply the radar snowfall retrieval by Schirle et al. (2019) and verify snowfall in the operational MEPS and sensitivity simulations with adjusted cloud microphysics conducted by Engdahl et al. (2020a,b).

We structure the remainder of this paper as follows: Section 2 describes the study location, observations, and methodology. Section 3 describes the snowfall regimes, followed by validating the OESR retrieved solid surface accumulated precipitation compared to the DFAR. Next, we discuss the seasonal bias in the accumulated solid surface precipitation and the vertical snowfall distribution of the NWP model simulations (MEPS, CTRL, and ICE-T) relative to the DFAR and OESR. Finally, section 4 presents the conclusions.

2. Methodology

a. Haukeliseter test site

The World Meteorological Organization (WMO) Haukeliseter test site (HTS), shown in Fig. 1, is situated on a mountain plateau at 991 m above sea level in Telemark County, Norway (59.81°N, 7.21°E). The Norwegian Meteorological Institute (MET Norway), operates the WMO measurement site for snow accumulation since winter 2010/11. The HTS is shielded from passing storm systems to the west by mountain peaks that extend up to 500 m above the site. Meanwhile, winds from the east can reach the HTS almost unobstructed (see Fig. 1b).

The temperature, precipitation, and wind are measured and recorded every minute at the HTS. Snowfall accounts for approximately 50% of the annual precipitation at the site, and the snow depth typically reaches 2–3 m in winter (Wolff et al. 2015). Therefore, to represent the measurement of the 2-m air temperature, this analysis uses the temperature from the mast at 4.5 m. An anemometer mounted at 10 m, typically 8 m, above the snow surface provides the wind speed and direction. The dominant wind directions for snowfall are westerly and southeasterly, with typical wind speeds below 20 and 12 m s⁻¹, respectively (see Fig. 2). At HTS, the DFAR consists of a precipitation-weighting gauge (Geonor T-200B3) encircled by a double fence to reduce any impacts of high winds and blowing snow on the precipitation measurements (Goodison et al. 1998). An overview of the instrumentation at HTS, including the DFAR and the meteorological mast, is shown in Fig. 1c.

During the HiLaMS field campaign, which took place at HTS during the 2016/17 winter, three additional instruments, a MASC, a particle imaging package, and a micro rain radar (MRR), were installed to study snowfall (see Fig. 1d). In this study, the OESR algorithm uses only the MASC and MRR. The MASC (Fig. 1d, left) consists of three cameras, three flashes, and two near-infrared sensors pointing at a ring center. The near-infrared sensors trigger the flashes and cameras to obtain a three-dimensional image of the hydrometeor and detect the hydrometeors as they pass through the ring. Hydrometeor fall speed is determined from the time it takes to fall between the

FIG. 1. (a) Representation of the topography in MEPS and the MEPS model domain. HTS is located in the mountainous region in Southern Norway. The contours and shading present the elevation of the $2.5 \text{ km} \times 2.5 \text{ km}$ grid cells. (b) Topographic map around HTS. From the DTM 10 terrain model of Geonorge (2018). HTS is surrounded by 500 m higher mountains to the west and more open to the southeast. (c) DFAR, unprotected precipitation gauges, and meteorological mast at HTS. The arrow indicates the westerly wind direction. The panel is adapted from Wolff et al. (2015). (d) Additional instruments installed during HiLaMS during winter 2016/17: Multi-Angle Snowflake Camera (MASC), Precipitation Imaging Package, and Micro Rain Radar (MRR) shown from left to right, respectively.

two vertically arranged infrared sensors (Garrett et al. 2012). In addition, the MASC provides estimates of the snow crystal habit, PSD, and near-surface snowfall velocity, which were used to determine the OESR assumptions (see section 2c).

Meanwhile, the MRR (Fig. 1d, right) operates at 24 GHz and was used to determine particle reflectivity and Doppler velocity aloft, thus providing the vertical macrophysical structure of snow events. The radar-based OESR retrieved profiles every minute at a vertical resolution of 100 m between 300 and 3000 m.

b. Snowfall regime analysis

An analysis of the 10-m wind (Fig. 2) indicates that most snowfall occurred in two distinct wind regimes (west and east). Furthermore, the MRR reflectivities suggested that the vertical structure of precipitation differed depending on the wind regime, with the westerly events consisting of pulses of high reflectivities ($>25 \text{ dBZ}$; see Fig. 3a). In contrast, during the easterly wind regime, the vertical reflectivity structure was associated with light precipitation ($\leq 25 \text{ dBZ}$; see Fig. 3b). Therefore, the snowfall events were separated by wind regime, namely, a westerly ($202.5^\circ\text{--}22.5^\circ$) and an easterly ($22.5^\circ\text{--}202.5^\circ$) regime, similarly to Schirle et al. (2019).

The wind regime was assigned for each minute observation when the surface temperature was less than 2°C depending on the observed wind direction during that minute. However, to omit local topographically induced turbulence on the wind regime assignment, a given wind regime was only assigned if it lasted for more than 10 min. Thus, we allow wind shifts lasting less than 10 min to be assigned to the encompassing wind direction. Once we determined the wind regime to the minute data, the respective OESR was used to calculate the surface accumulated snowfall and the vertical SWC. We only considered days with continuous hourly observations (24 h) in the analysis, both from the DFAR and OESR. Furthermore, only days are used where the DFAR and OESR observed more

than 0.25 mm day^{-1} precipitation to exclude measurement noise from the analysis.

The DFAR measurements were summed over each hour and assigned to the instantaneous 10-m wind speed and direction at the end of the hour as measured by the anemometer at HTS to make the observations and simulations comparable. Similarly, the OESR surface accumulation was summed each hour and assigned to the instantaneous wind speed and direction at the end of the hour, regardless of the wind regime used for

FIG. 2. Wind rose of the 10-m wind during snowfall events when 24-h accumulation $\geq 25 \text{ mm day}^{-1}$ and 2-m temperature $< 2^\circ\text{C}$ measured at HTS, during winter 2016/17. Colors indicate the 10-m wind speed categories used in this study.

FIG. 3. Examples of MRR reflectivity during the two typical snowfall regimes. (a) The time series for a westerly snowfall regime case during high wind speed on 24–25 Dec 2016. The westerly snowfall regime was associated with interchanging patterns of high and low reflectivities during snowfall. (b) An example during an easterly snowfall regime with low wind speeds and light precipitation of convective nature, on 9–10 Jan 2017.

the OESR within the hour. We discuss the two distinct precipitation regimes with their specific meteorology, MRR reflectivities (Fig. 3), and associated snowfall in section 3a.

c. Optimal estimation retrieval algorithm

Schirle et al. (2019) describe the OESR technique used to estimate the vertical profile of snow properties. Use of the flexible OESR approach allows a means to combine radar reflectivities, in situ observations, atmospheric temperature profiles, and a priori information into a common retrieval framework to provide an estimate of snowfall properties consistent with each. The scheme retrieves an exponential PSD for each range bin of the MRR that then can be converted to SWC using particle model size–mass dimensional relationships. The SWC is then transformed into snowfall water equivalent at the surface through use of fall speed observations.

During the HiLaMS campaign, Schirle et al. (2019) explored the impact of different microphysical assumptions in the OESR on retrieved surface snowfall. They found best-case differences between total seasonal retrieved surface accumulations and HTS DFAR observations of +9% and +16% during easterly and westerly snowfall regimes, respectively. However, the combination of measurements that maximized retrieval performance differed for these snowfall regimes. The best OESR results used observations of PSD and fall speed from the MASC for the relatively low-wind easterly events. For the high-wind westerly snowfall events, near-surface turbulence and blowing snow dictated the use of a temperature–PSD relationship and fall speeds from the MRR for best results. However, the most important assumption for the first-order accuracy of any snowfall retrieval scheme is the selection of a particle model that is well-matched to scene environmental conditions (Cooper et al. 2017; Schirle et al. 2019). At HTS, partially rimed aggregates dominated the MASC images during both wind regimes (Fig. 4). As such, Schirle et al. (2019) used a snow particle aggregate model developed for the

CloudSat mission that produces a high reflectivity per unit mass relationship consistent with aggregates entrained in high liquid water content aloft (Wood et al. 2015).

In this work, we use these assumptions to compare surface snowfall accumulations in the OESR with DFAR measurements for a slightly different classification scheme for snowfall events than those used in the Schirle et al. (2019).

Good agreement between retrieved and observed snowfall values at the surface provides confidence in the retrieved SWC values aloft. At HTS, Schirle et al. (2019) showed that using a rimed aggregate particle model produced retrieved snowfall values that well-matched the MET Norway DFAR observations. Given that surface snowfall characteristics are dictated by snowfall processes aloft, there should be a strong physical correlation between the aggregate particles at the surface and those the MRR measures above. The OESR scheme also uses a PSD–temperature relationship that allows PSD to change with height for more accurate in-cloud retrievals of SWC (Wood 2011). Current research focuses on using airplane in situ microphysical observations to evaluate retrieved vertical profiles of snow water. Such work, however, is beyond the scope of this paper.

d. Operational weather forecast model—MEPS

In this study, we make use of the archived MEPS surface forecasts, the control (CTRL) run, and a version of the CTRL with the modified microphysics scheme (ICE-T) from Engdahl et al. (2020a).

MEPS is the operational ensemble forecast system used at MET Norway (Frogner et al. 2019). It is based on HARMONIE-AROME (version 40h1.1), a mesoscale non-hydrostatic, convection-permitting NWP model (The MetCoOp Team 2017), which, in turn, is based on the model developed by Meteo-France, AROME (Seity et al. 2011; Bengtsson et al. 2017). HARMONIE-AROME uses the single-moment ICE3 bulk microphysics scheme (Caniaux et al. 1994; Pinty and

FIG. 4. (a) Typical snowflake habits observed during events classified in the easterly snowfall regime. (b) Examples of large precipitating crystals observed during the westerly snowfall regime. This figure is adapted from Schirle et al. (2019).

Jabouille 1998) to represent cloud microphysics. ICE3 simulates mass mixing ratios of cloud water and ice, rain, snow, and graupel (Cohard and Pinty 2000a,b).

Within MEPS, HARMONIE-AROME runs at a horizontal resolution of 2.5 km with 65 hybrid levels in the vertical. Figure 1a shows the MEPS model domain and simulated elevation with a domain center at 63°N, 15°E. MEPS consists of ten HARMONIE-AROME ensemble members. We use the ensemble output from the MEPS archive with initialization at 0000 UTC, with 3-hourly cycling for data assimilation. The control run has initial and lateral boundary conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF) High-Resolution forecast. The ensemble is created by giving members one through nine perturbed initial and lateral boundary conditions based on the scaled lagged average forecasting method (Køltzow 2017). In this study, we average the ensemble members (the ensemble mean) to derive the reported MEPS variables in the following figures—a detailed

description of the MEPS operational setup can the interested reader find in Frogner et al. (2019).

Engdahl et al. (2020b) pointed out a coding error, corrected in the CTRL run, and therefore, the CTRL microphysics deviates slightly from the archived MEPS simulations. In a follow-up study, Engdahl et al. (2020a) ran 3D simulations with both the bug-fixed microphysics scheme, CTRL, and ICE-T, respectively, for December 2016–February 2017. An overview of the differences between the two microphysical schemes in CTRL and ICE-T is found in Table 1 and Engdahl et al. (2020b). The Engdahl et al. (2020a) model setup is as follows: The domain has the exact horizontal and vertical resolution as MEPS but is placed further to the west to provide additional spinup time for weather systems from the west. CTRL and ICE-T are initialized every day at 0000 UTC. The initialization uses the initial and lateral boundaries from the operational ECMWF-Integrated Forecast System but with no surface or upper air data assimilation. Therefore, clouds and

TABLE 1. List of microphysical processes which are altered in CTRL and ICE-T. Table is taken from Engdahl et al. (2020b).

Expt	Process altered	Previous	New
CTRL	Heterogeneous ice nucleation	Code mistake	Bugfix
ICE-T	Autoconversion (cloud to rain)	Khairoutdinov and Kogan (2000)	Berry and Reinhardt (1974)
	Rain accreting cloud water	Variable efficiency (Müller et al. 2017)	Variable efficiency (Thompson)
	Heterogeneous ice nucleation	Meyers et al. (1992)	Cooper (1986)
	Freezing of water drops	None	Bigg (1953)
	Graupel collecting cloud water	Dry growth: Ferrier (1994); Wet growth: Musil (1970) and Nelson (1983)	Cober and List (1993)
	Snow collecting cloud water	Farley et al. (1989)	Wang and Ji (2000)
	Rain collecting snow/graupel	Ferrier (1994); Eff = 1.0	New variable: Collection efficiency
	Hydrometeor properties	Locatelli and Hobbs (1974)	Thompson et al. (2008)
	Rain inverse exponential y-intercept parameter	$8 \times 10^6 \text{ m}^{-4}$ (Marshall–Palmer)	Variable intercept parameter (Thompson et al. 2004)

FIG. 5. We reduced the known 10-m wind bias (Frogner et al. 2019) according to the correlation equation between 10-m wind speeds observed at the DFAR and (a) MEPS, (b) CTRL, and (c) ICE-T. The red line indicates the linear correlation between DFAR and OESR. The gray dashed lines represent the line of equality where model simulations are equal to observations.

precipitation are not available at the beginning of the simulation. ICE-T led to a general increase in supercooled liquid water, a shift in hydrometeor distribution from graupel to snow, and a shift in the precipitation pattern with more precipitation spilling over to the lee side of mountain barriers.

e. Model validation

To validate the model, we use the closest grid point to HTS, which is located 0.9 km from the site in HARMONIE-AROME and has a similar altitude of 1041 m above sea level. Identical to Engdahl et al. (2020a), we exclude the first 12 h of each simulation and then analyze the following 24 h to account for the model spinup time of clouds and precipitation in MEPS, CTRL, and ICE-T. The DFAR and OESR data are compared to the same periods as the model simulations. Hence a day of interest starts at 1200 UTC and ends the next day at 1200 UTC. Finally, the total accumulated precipitation difference within an hour is calculated for the observations, the retrieved values, and the model simulations.

Additionally, the instantaneous 2-m temperature of the observations and the model had to remain below 2°C to ensure only snowfall is present due to the OESR setup and resulting surface snowfall accumulation and SWC from the OESR. The separation into snowfall regimes takes the instantaneous 10-m wind direction from MEPS, CTRL, and ICE-T. Engdahl et al. (2020a) verified CTRL and ICE-T with 177 WMO stations in Norway for the three winter months simulation and found the 10-m mean error in the wind speed of approximately +0.7 m s⁻¹. However, the complex terrain around the HTS and its representation in MEPS, CTRL, and ICE-T leads to a greater wind bias simulation (see Fig. 5). Therefore, we analyzed the simulated and observed wind speeds, and we found a simulated bias of +2.92, +2.44, and +2.55 m s⁻¹ for MEPS, CTRL, and ICE-T, respectively (see Fig. 5). Hence, we corrected the simulated wind speed in the postprocessing to compare the solid surface accumulation and vertical precipitation more accurately to the corresponding observed wind speeds.

The MEPS archive does not provide the vertical parameters to calculate the SWC profile for all ensemble members. Hence, we did not carry out a vertical validation for MEPS simulations. Instead, we validate the instantaneous amount of snowfall in the vertical from the simulations CTRL (which we assume is close to the deterministic forecast in MEPS) and ICE-T to the OESR vertical profiles of SWC.

In CTRL and ICE-T, the SWCs were derived by summing the mass mixing ratios of cloud ice, snow, and graupel and then converting to grams per cubic meter (g m⁻³) [Eq. (1)] by using the vertically resolved air density [Eq. (2)] of MEPS as follows:

$$\text{SWC}(\sigma) = \rho(\sigma)[m_{\text{snow}}(\sigma) + m_{\text{graupel}}(\sigma) + m_{\text{cloudice}}(\sigma)] \times 10^6 \text{ (g m}^{-3}\text{)}, \quad (1)$$

where

$$\rho(\sigma) = \frac{p(\sigma)}{R_d T(\sigma)} \text{ (kg m}^{-3}\text{)}. \quad (2)$$

The SWC was calculated at each σ -pressure level, but for validation, the SWC is only compared to the corresponding heights of the retrieved snowfall values from the MRR. Additionally, we compare the total snow water path (SWP) from the instantaneous values of the OESR, the CTRL, and ICE-T simulations to validate the integrated ice mass from 400 to 3000 m over HTS. Then the SWP is calculated at each hourly instantaneous output of the model and corresponding time in the OESR with Simpson's rule [Eq. (3)]:

$$\begin{aligned} \text{SWP} &= \int_{h=400\text{m}}^{h=3000\text{m}} \text{SWC}(h) dh \\ &= \frac{h_{3000\text{m}} - h_{400\text{m}}}{3} \left[\text{SWC}(h_{400\text{m}}) + \text{SWC}(h_{3000\text{m}}) \right. \\ &\quad \left. + 4 \times \text{SWC}\left(\frac{h_{400\text{m}} + h_{3000\text{m}}}{2}\right) \right] \text{ (g m}^{-2}\text{)}. \end{aligned} \quad (3)$$

3. Results

a. Snowfall regimes

The DFAR observed the most surface snowfall accumulation during the westerly snowfall regime, which accounted for 73% (146.5 mm, see Fig. 6a) of the total precipitation during the 2016/17 winter. In this snowfall regime, 20% of the precipitation occurred at wind speeds higher than 12 m s^{-1} and sometimes exceeding 20 m s^{-1} (Fig. 2). Figure 3 demonstrates MRR reflectivities during the two distinct snowfall regimes. Westerly snowfall events consisted of intermittent periods of heavy ($>25 \text{ dBZ}$) and light ($<15 \text{ dBZ}$) snowfall with a duration of approximately 30 min (Fig. 3a). The observed precipitation pattern is consistent with HTS being located to the lee of the higher mountains to the west (see Fig. 1b). Previous studies found that latent heat release on the windward slope of mountain barriers leads to pulsed precipitation on the leeward side (Sinclair et al. 1997; Kaplan et al. 2009).

In contrast, the easterly snowfall regime was associated with light precipitation of 54.0 mm (Fig. 6b) and winds of less than 12 m s^{-1} (Fig. 2). The MRR reflectivities for the easterly snowfall regime were consistent with continuous moderate precipitation with values near 15–20 dBZ (see Fig. 3b). Moderate easterly winds likely enhanced the snowfall as the low-level easterly flow impinged the mountain barrier, causing enhanced lift and orographic precipitation. The precipitation during the easterly snowfall regime was dominated by rimed aggregates, as shown in Fig. 4b. The riming is likely due to the formation of a low-level feeder cloud, which acted as a source of additional condensate where the snowfall gained mass by riming (Borys et al. 2003; Lowenthal et al. 2016; Ramelli et al. 2021b). Indeed, Fig. 3b shows an increase in reflectivity around 1000 m above the surface between 1800 and 0000 UTC. The increase is likely due to the precipitation growth in the low-level feeder cloud.

b. Retrieval validation

During the 2016/17 winter, a difference of 10.9% between retrieved (OESR) and DFAR total surface accumulations was observed (Table 2). When separating by snowfall regime, the retrieval overestimated the accumulated surface precipitation by +7.3% and +20.5% during the westerly and easterly snowfall regimes, respectively (see Fig. 6a and Table 2). Furthermore, when comparing the OESR and DFAR snowfall rates, a strong correlation between the measurements of $R^2 = 0.81$ and 0.71 for the westerly and easterly snowfall regimes, respectively, was observed (see Fig. 7)—indicating that the microphysical a priori selected for the OESR are well-matched for the HTS conditions. Additionally, these results largely agree with Schirle et al. (2019), who found a +16% and +9% difference between OESR and DFAR observed seasonal snowfall accumulations during westerly and easterly events. Slight differences between the studies are likely due to the minor changes in snowfall regime classification in this study.

Notably, the difference between the retrieved and measured precipitation could be due to the undercatch of the DFAR at high wind speeds (Thériault et al. 2015; Nitu et al. 2018; Colli et al. 2020). Nitu et al. (2018) showed that the DFAR has an undercatchment up to 10% for wind speeds below 9 m s^{-1} .

Meanwhile, they estimated that the DFAR has an undercatch of 20% during wind speed between 10 and 20 m s^{-1} . When accounting for the influence of wind speed on the DFAR measurements, the difference between the DFAR and OESR surface accumulation becomes -6.8% and $+9.2\%$ for the westerly and easterly regimes, respectively (see Table 2). Thus, the DFAR undercatchment can explain the observed difference between the OESR and DFAR during westerly events. In contrast, during the easterly snowfall regime, accounting for potential undercatchment does not explain the difference between the OESR and DFAR. As a result, the cumulative difference between the DFAR and OESR, without correction, for the easterly snowfall regime over the entire season is only 11 mm.

As discussed in section 2c, the similarity between the retrieved surface snowfall amount from the OESR and the DFAR gives confidence in the retrieved snow water in the vertical over the HTS. These values will be used as a reference to evaluate the vertical profiles of snow water from the CTRL and ICE-T simulations, as discussed in section 3d.

c. Validation of surface snowfall

Following the technique described in section 2e, we compare the simulated accumulated surface snowfall of MEPS, CTRL, and ICE-T to the DFAR measurements over 27 days. Comparing MEPS, CTRL, and ICE-T reveals an overestimation of +41.0%, +43.8%, and +59.2%, respectively (Table 2). MEPS relies on an ensemble approach that more accurately predicts the dynamical evolution of precipitation events, leading to better performance (Frogner et al. 2019). The CTRL simulation outperforms ICE-T as ICE-T simulates more precipitation over HTS than the CTRL. The increase in simulated surface snowfall in ICE-T is likely related to increased generation of snow and reduced production of graupel, as snow has lower fall velocity than graupel. This means the condensate can remain in the atmosphere longer and thereby be advected over further distances. Engdahl et al. (2020a) showed that the precipitation pattern had changed with ICE-T, leading to less precipitation up-slope of the mountains and more on the lee side, where HTS is situated [see Fig. 7 in Engdahl et al. (2020a)]. The total surface precipitation within the domain was reduced from ICE-T to CTRL (Engdahl et al. 2020a), suggesting a lower precipitation efficiency. Likewise, Liu et al. (2011) found that the Lin et al. (1983) based scheme had less condensate in the atmosphere, combined with an excess of snowfall precipitation, and suggested a high precipitation efficiency due to a high fraction of fast-falling graupel.

Engdahl et al. (2020b) changed the coefficients in the mass-diameter relation [$m(D) = aD^b$] and terminal fall velocity [$v(D) = cD^d(\rho_{00}/\rho_{\text{dref}})^{0.4}$] within ICE-T. All coefficients (a , b , c , d) were changed for snow, while only c and d got changed for graupel, which leads to graupel falling more slowly and snow falling slower when $D < 600 \mu\text{m}$ and faster for $D > 600 \mu\text{m}$. Yet, the size distribution of snow is shifted toward smaller particles, so the net fall speed of snow is reduced.

The shift in precipitation pattern is evident in the results. Indeed, when comparing the simulated accumulated surface snowfall by snowfall regime, the westerly snowfall regime events

FIG. 6. Surface snowfall accumulation for DFAR observations, OESR, MEPS, CTRL, and ICE-T, separated into (a) westerly and (b) easterly snowfall regimes for 27 precipitation days during winter 2016/17. The sum (Σ) over the 27 days of total precipitation accumulation in each snowfall regime is presented in the figure legend in (a) and (b). According to the correlation equation in Fig. 5, the separation into wind speed regimes is done for the corrected simulated wind for MEPS, CTRL, and ICE-T. (c),(d) The event hours observed at the

TABLE 2. Surface snowfall accumulations for 27 precipitation days during winter 2016/17. Wind regimes are separated into westerly and easterly snowfall regimes as described in section 2b. “Diff” defines the percentage difference in cumulative snowfall accumulation between OESR, MEPS, CTRL, ICE-T, and the DFAR. The last five rows represent the adjustment to DFAR undercatch depending on the wind speed, where 10% undercatch has been applied to winds below 10 m s^{-1} and 20% undercatch for winds above 10 m s^{-1} .

	West (mm)	Diff (%)	East (mm)	Diff (%)	Total (mm)	Diff (%)
DFAR	146.5		54.0		200.5	
OESR	157.2	+7.3	65.1	+20.5	222.3	+10.9
MEPS	233.9	+59.7	48.8	−9.7	282.7	+41.0
CTRL	233.3	+59.3	55.0	+1.8	288.3	+43.8
ICE-T	262.4	+79.2	56.7	+5.0	319.2	+59.2
DFAR	168.7		59.6		228.3	
OESR	157.2	−6.8	65.1	+9.2	222.3	−2.63
MEPS	233.9	+38.6	48.8	−18.1	282.7	+23.8
CTRL	233.3	+38.3	55.0	−7.7	288.3	+26.3
ICE-T	262.4	+55.6	56.7	−4.8	319.2	+39.8

show an overestimation at the surface of +59.7%, +59.3%, and +79.2%, for MEPS, CTRL, and ICE-T, respectively (see Table 2).

Meanwhile, the simulated easterly snowfall regime events show to be within the observation uncertainty of snowfall at the surface of −9.7%, +1.8%, and +5.0% for MEPS, CTRL, and ICE-T, respectively (see Table 2). Therefore, the shift in precipitation to the lee of the mountain barrier dominates the model’s overestimation during westerly events. Snowfall shift is especially the case for the ICE-T simulation, which shows the most significant overestimation during the westerly snowfall regime.

When limiting the westerly snowfall to wind speed below 12 m s^{-1} , the overestimation in accumulated snowfall by MEPS and ICE-T falls to +9.6% and +13.3%, respectively. Simultaneously, CTRL underestimates the observed accumulated snowfall by −0.2%. In contrast, above 12 m s^{-1} , the surface snowfall accumulation during the westerly snowfall regime is greatly overestimated by +510.7%, +494.0%, and +557.6% for MEPS, CTRL, and ICE-T, respectively (see Fig. 6a). Thus, the models are simulating too much snowfall at high wind speeds. Additionally, when accounting for the number of event hours simulated by the model between 12 and 16 m s^{-1} , the models also simulate twice as many event hours as observed by the DFAR (see Fig. 6c). At wind speeds above 16 m s^{-1} , the simulated event hours are closer to the observations. Thus, most of the surface snowfall overestimation in the simulations stems from too many events simulated between 12 and 16 m s^{-1} during the westerly regime. Additionally, the snowfall amount per event hour is overestimated at wind speeds above 12 m s^{-1} , suggesting that models simulate too much spillover when higher wind speeds are simulated (see Fig. 6e). Indeed, higher cross-mountain wind speeds increase spillover precipitation (Chater and Sturman 1998; Kaplan et al. 2012). Regardless of the mechanisms responsible,

the overestimation in simulated surface snowfall occurs in both the models that rely on an ensemble of NWP models (here presented as the ensemble mean of MEPS) or have altered microphysical schemes (CTRL, ICE-T).

In contrast to the westerly snowfall regime, there is no clear dependence on accumulated snowfall bias at different wind speeds during the easterly snowfall regime (Fig. 6b). However, when looking at the simulated amount of accumulated snowfall per event hour, the simulations have a significant bias at wind speeds above 8 m s^{-1} (see Fig. 6f). This overestimation may be associated with enhanced orographic lifting at higher wind speeds or due to fewer event hours in the simulations. Nevertheless, the exact mechanism for this bias is beyond the scope of this study.

As previously mentioned, the DFAR is prone to underestimating precipitation at high wind speed. Therefore, some of the models’ overestimation may be due to this, particularly as the largest overestimations occurred at wind speeds greater than 12 m s^{-1} for the westerly snowfall regime. Applying an undercatch error of 10% ($<9 \text{ m s}^{-1}$) and 20% ($>10 \text{ m s}^{-1}$) leads to a reduction in the overestimation of snowfall by the models at the surface over all events during the study period of +23.8%, +26.3%, and +39.8% for MEPS, CTRL, and ICE-T, respectively (see Table 2). Although accounting for the undercatchment reduces the difference between the simulated and measured surface accumulated precipitation, it does not completely explain the simulated overestimation in this study. Koltzow et al. (2020) discussed how wind-induced undercatch of surface snowfall could impact the verification of precipitation forecasts in Norway. The study used MEPS forecasts and found that wind-induced undercatch of snowfall has a considerable impact on the NWP verification, especially for single fence precipitation gauges. The use of the DFAR in this study reduces the potential biases due to undercatch and

←

DFAR and OESR and the simulated precipitation hours in the NWPs. The total sum (Σ) of the precipitation hours is presented in the figure legend for the 27 precipitation days. Further, the snowfall accumulation is divided by the event hours in (e) westerly and (f) easterly snowfall regimes.

FIG. 7. Surface snowfall validation of OESR of the hourly snowfall accumulation, with DFAR observations on the x axis and OESR values on the y axis. The hourly accumulated snowfall amount is separated into (a) westerly and (b) easterly snowfall regimes. The red line indicates the linear correlation between the DFAR and the OESR surface snowfall accumulation. We counted 121 events at the 0.1 mm h^{-1} bin during westerly snowfall, while we counted only 45 events at the 0.1 mm h^{-1} bin during easterly snowfall. The gray dashed lines represent the line of equality where the OESR is equal to the observations.

substantially increases confidence in the verification results. Furthermore, if the DFAR undercatch significantly biased the observations, the difference between OESR and DFAR would be more pronounced, especially at high wind speeds.

On the contrary, at wind speeds above 16 m s^{-1} during the westerly snowfall regime, the DFAR observed more snowfall than estimated by the OESR, suggesting that blowing snow may have artificially elevated the accumulated snowfall measured by the DFAR at these high wind speeds. Therefore, undercatchment by the DFAR is not significantly biasing the observations. Thus, the models' overestimation in simulated snowfall is an actual bias, as the models produce too much snow at higher wind speeds (see Figs. 6c,d).

The following section investigates if the overestimation in snowfall in the simulations extends into the vertical.

d. Validation of the vertical distribution of snowfall

The vertical distribution of SWC, as retrieved by the MRR using the OESR, is used to validate CTRL and ICE-T to identify if there is a connection between the overestimation of surface precipitation and the distribution of condensates in the atmosphere.

The MRR retrieved SWC averaged during winter 2016/17 indicates precipitation formation near cloud top and subsequent growth of hydrometeors as the precipitation falls through the cloud (see Fig. 8a). At 800 m, the mean SWC reaches its maximum of 0.13 g m^{-3} and then begins to decrease as the precipitation continues to fall (see Fig. 8a). As the retrieved SWC depends on the observed reflectivity from the MRR, which is highly dependent on the falling hydrometeors' size, the reduction in mean SWC below 800 m is likely due to a general decrease in the precipitation size. A

decrease in hydrometeor size can occur either because of snow crystal fragmentation or sublimation. Layers of strong wind shear are often present in mountain valleys, leading to fragmentation and subsequently reduced reflectivity of precipitation (Ramelli et al. 2021a). Additionally, Ramelli et al. (2021a) showed that due to upstream topography, a relatively dry boundary layer could exist in the lee of mountains where falling precipitation can sublimate, reducing the observed radar reflectivity. Although it is unclear whether the decrease in MRR reflectivity and calculated SWC is primarily due to sublimation or fragmentation, the MASC images (see Fig. 4) support that fragmentation was not frequently observed at HTS. Thus, the maximum mean SWC is assumed to correspond to the mean cloud base height. Compared to the modeled SWC, both the CTRL and ICE-T simulations also capture a decrease in SWC at $\sim 600 \text{ m}$ above the surface (see Fig. 8a), indicating that they correctly account for the vertical evolution of hydrometeors as they fall to the surface.

Separating the SWC by wind regime, during the westerly snowfall regime (see Figs. 8b–f), the maximum in mean SWC of 0.24 g m^{-3} occurs in the OESR between 800 and 1000 m, especially at winds between 8 and 16 m s^{-1} (Figs. 8d,e). During the easterly events (Figs. 8g–i), the mean SWC reaches a maximum of 0.19 g m^{-3} between 600 and 1000 m at 4 to 8 m s^{-1} (see Fig. 8h). Thus, the westerly events tend to produce higher maxima in mean SWCs, consistent with a moister westerly air mass and subsequent precipitation as observed by the DFAR (see Fig. 6a and Table 2). Furthermore, the lowering of the maximum SWC during easterly snowfall regimes suggests that the cloud base is lower during easterly events. The cloud base lowering is consistent with the topography around HTS, with higher mountains to the west. Similarly, both CTRL and ICE-T

FIG. 8. Mean (solid) and standard deviation (dashed) of instantaneous SWC separated into (a)–(f) west and (g)–(i) east snowfall regimes for days fulfilling the analysis requirements during winter 2016/17. The SWC is sorted into snowfall regimes with the adjusted wind speed correction from the correlation equation in Fig. 5b for CTRL and Fig. 5c for ICE-T.

correctly represent higher SWC during the westerly snowfall regime. However, only ICE-T reproduces the lowering of the SWC maximum during the easterly snowfall regime.

Comparing the vertical profiles of SWC from the simulations to the OESR, it is clear that the simulations underestimate the maximum in SWC. Furthermore, the simulations show significantly less variability in SWC with height, indicating that they are not correctly representing the microphysical processes. In particular, the total mean SWC simulated by CTRL is almost constant with height and has a maximum value around 0.05 g m^{-3} but decreases below

600 m (Fig. 8a). Similarly, the total mean SWC in ICE-T increases linearly as the precipitation falls and is higher relative to the CTRL simulation. However, the seasonal mean SWC is still too low in ICE-T, with a maximum value of 0.06 g m^{-3} at 1200 m (Fig. 8a). Thus, in principle, the CTRL and ICE-T simulations account for the sublimation below the cloud base but produce an overestimation in surface accumulation yet an underestimation in vertical SWC.

The microphysical adjustments within ICE-T lead to more SWC in the column above HTS, reducing the bias compared to the mean SWC retrieved with the MRR. Indeed, Engdahl et al. (2020a)

FIG. 9. The seasonal mean of instantaneous SWP for OESR, CTRL, and ICE-T forecasts separated into (a) westerly and (b) easterly snowfall regimes for 27 precipitation days during winter 2016/17. The separation into wind speed is done with the corrected wind speed bias from Fig. 5b for CTRL and Fig. 5c for ICE-T. The black line indicates the standard deviation ($\pm\sigma$) in each snowfall wind regime. The mean and standard deviation in each wind speed bin is calculated individually; hence, the result is dependent on the event hours. The figure labels show the total mean of SWP over all wind speeds and all precipitation event hours.

demonstrated that the ICE-T scheme simulates an increase in the vertical distribution of snow and a slight reduction in graupel relative to CTRL above the HTS. Therefore, the ICE-T scheme allows the modeled solid hydrometeors to remain in the cloud longer due to the slower fall velocity of snow than graupel, thereby increasing the SWC in the vertical. Nevertheless, ICE-T still simulates too much surface snowfall compared to the DFAR.

Furthermore, the simulated sublimation of frozen precipitation below 1200 m is too strong compared to the OESR over HTS (see Fig. 8a). However, note that for the comparison between the MRR retrieved and modeled SWC, the SWC profiles are only examined to 400 m above the HTS. Thus, an upwind grid box could advect simulated precipitation to the HTS, where the low-level SWC is higher and missed in this analysis.

For a better comparison of SWC based on the snowfall regime, the integrated value of SWC, SWP, is used. Recall that the accumulated snowfall at the surface for wind speeds below 12 m s^{-1} agreed well between DFAR, CTRL, and ICE-T. However, at higher wind speeds, the simulations overestimated snowfall accumulation. The overestimation was mainly observed during the westerly snowfall regime (Fig. 6a). The mean SWP divided into snowfall regimes and wind speeds could indicate if the vertical bias is similar for snowfall. In contrast to the surface accumulation, the simulated instantaneous SWP, averaged over the 27 days, is significantly underestimated relative to the OESR, which was in good agreement with the DFAR for surface snowfall (+10.9% and see Fig. 7). When analyzing the average simulated versus retrieved SWP averaged over all event hours, the CTRL and ICE-T have a deficit of 1071.4% and 523.7%, respectively. Separating the SWP by snowfall regimes reveals

too little SWP irrespective of the snowfall regime (see Fig. 9). Comparing the SWP between CTRL and ICE-T, the ICE-T scheme reduced the deficit by a factor of 2 (1.96 for west and 2.08 for east). However, the underestimation in SWP was not consistent at all 10-m wind speeds. Specifically, during the westerly snowfall regime, the OESR observed more SWP at wind speeds below 16 m s^{-1} , while the simulations overestimated SWP at higher wind speeds. The microphysical adjustments reduced the error in SWP in the westerly regime from -55.9% (CTRL) to -40.1% in ICE-T, below 16 m s^{-1} (see Fig. 9a). At westerly wind speeds higher than 16 m s^{-1} , both simulations show an overestimation above $+300\%$ in SWP mean, mainly because the MRR-retrieved SWP was close to zero at wind speeds higher than 16 m s^{-1} . The overestimation of mean SWP at high wind speeds in CTRL and ICE-T is similar to the surface precipitation accumulation comparison. Still, it occurs at 16 m s^{-1} instead of 12 m s^{-1} during the westerly snowfall regime. During the easterly snowfall regime, there was no clear transition of underestimating SWP based on wind speed. Nevertheless, the models simulated more SWP at wind speeds above 8 m s^{-1} , likely related to fewer precipitation events (see Fig. 6d). As the simulated surface accumulated snowfall and SWP show similar biases based on wind speed, we suggest that it may be due to the same reason—too high wind speeds and subsequent spillover estimated in MEPS and its counterparts (CTRL and ICE-T) (Müller et al. 2017; Frogner et al. 2019).

Regardless, the simulations overestimated the surface snowfall accumulation when comparing all event hours while significantly underestimating the SWP. The underestimation may be due to several factors, including a poor representation of the terrain in the model, an overzealous conversion of liquid to ice, and subsequent snow growth in the microphysics schemes.

Indeed, the simulated location of the HTS was 51 m higher and closer to the barrier crest in the model domain. Thus, the modeled precipitation would be susceptible to more spillover precipitation at high wind speeds, as was observed. Additionally, with its reduced conversion of snow to graupel, ICE-T led to an increase in SWC and SWP over HTS. Therefore, it represented the vertical distribution of snowfall more accurately relative to the OESR. ICE-T gives more snowfall than the CTRL in the atmosphere and at the surface. Snow can stay longer in the atmosphere because of its lower fall velocity. Still, at the same time, the hydrometeors in ICE-T can have more mass but a lower diameter leading to more snow accumulation at the surface.

Nevertheless, an in-depth analysis of why these biases occur is required to improve regional models, especially in the vertical. We want to point out that this study only investigates the effects of changes to the ICE3 cloud microphysical scheme in the operational forecast model and the DFAR and MRR retrieval approach to validate forecast models. We are aware that differences between the model simulations and OESR could also be due to other aspects of the model, such as kinematic and thermodynamic structures. However, MET Norway continuously validates the operational model with radiosondes within the model domain. Engdahl et al. (2020b) also showed good agreement between modeled and observed atmospheric temperature profiles for a drizzle test case.

Furthermore, there exists a broad range of microphysical schemes. Most one-moment microphysical schemes calculate the particle mass explicitly and diagnose the number concentration based on the particle mass. In contrast, two-moment schemes, like Thompson et al. (2008) and Morrison et al. (2009), calculate the number concentration explicitly for some hydrometeors. Higher-order schemes like bin and three-moment schemes calculate the number concentration explicitly within bins of size intervals or simulate radar reflectivities. While bin and three-moment schemes seem more applicable in this study, they are also computationally expensive to run operationally (Morrison et al. 2009).

Combining the surface and vertical validation presented in this study provides a new technique for validating operational forecast models with point measurements. For example, one might assume that microphysical adjustments should reduce the model's surface precipitation amount at high wind speeds without the vertical information. However, corrections based on the surface observations alone would likely lead to further underestimating the vertical SWC, leading to other errors in simulated parameters. Thus, using both vertical and surface observations during future model development should help when adjusting microphysics schemes.

4. Conclusions

Here we present a new method for validating NWP model simulations in complex terrain with state-of-the-art observations. Specifically, we investigated how the model simulations from MET Norway's ensemble forecast product MEPS and two additional simulations with modified cloud microphysics schemes (CTRL and ICE-T) compare to surface observations and retrieved snowfall values. This study evaluated the model

performance for 27 precipitation days by comparing simulated accumulated snowfall, SWC, and SWP to measurements and retrieved values at the HTS in Southern Norway.

An OESR algorithm, based on reflectivity profiles from an MRR and a priori assumptions that included ice crystal habit information provided by a MASC, was used to obtain vertical profiles of the SWC and SWP over the HTS (Schirle et al. 2019). The retrieved surface snowfall accumulation was compared to snowfall measurements from the HTS DFAR gauge to validate the OESR. The validation was conducted for two distinct snowfall regimes, where different a priori assumptions were used in the OESR, according to the estimates made by Schirle et al. (2019). Once shown to compare well with the DFAR surface accumulation, the OESR retrieved SWC and SWP served as a reference to evaluate the vertical profiles of simulated SWC and SWP in the CTRL and ICE-T simulations.

We found that MEPS, CTRL, and ICE-T overestimated the surface snowfall accumulation by validating model simulations with differing microphysical schemes compared to the DFAR during winter 2016/17. The adjustments by Engdahl et al. (2020b) in the ICE-T scheme increased the overestimation of snowfall at the surface from 41.0% (MEPS) to 59.2% relative to the DFAR. Additionally, we find that the overestimation is more pronounced during the westerly snowfall regime (MEPS: 59.7%, CTRL: 59.3%, ICE-T: 79.2%). The significant overestimation of snowfall at the surface is related to the increase in simulated snow in the region of HTS, as discussed by Engdahl et al. (2020a). The undercatchment of the DFAR can partially explain the overestimation of surface accumulation at 10-m wind speeds higher than 10 m s^{-1} . However, undercatchment alone was not able to resolve the overestimation of snowfall at the surface. Furthermore, MEPS, CTRL, and ICE-T simulated more high wind speed event hours than observed. This likely contributed to the seasonal overestimation in the simulations due to more spillover precipitation reaching HTS than observed.

In contrast, the model simulations produced too little snowfall throughout the vertical column relative to the OESR retrievals. The vertical representation of SWC in ICE-T improved relative to CTRL, likely due to the microphysical adjustments that maintained supercooled liquid water longer in the column (Engdahl et al. 2020b). This adjustment reduced the mean SWP bias by a factor of 2 from CTRL (-1071.4%) to ICE-T (-523.7%). Nevertheless, too little snowfall in the vertical in CTRL and ICE-T is likely related to the generation of hydrometeors which are too large in the model simulations. These larger and heavier hydrometeors fall out too quickly ultimately, reducing the SWC while generating too much snow at the surface.

As shown here, comparisons of vertical profile estimates of snow water derived from a combined DFAR and MRR retrieval approach can identify potential biases in snow water production within MEPS, CTRL, and ICE-T. However, using either surface accumulation values or vertical SWC profiles would lead to misleading conclusions regarding potential biases in snowfall production in the studied simulations at HTS. Therefore, we recommend that future model validation studies combine ground-based in situ and vertically profiling remote sensing instruments, where possible. In addition, we

recommend the continued pursuit of event and site-specific observations of particle habit and fall speed to improve radar-based OESRs for locations with differing meteorological influences to contextualize our proposed explanations for the model biases in MEPS, CTRL, and ICE-T. Furthermore, liquid water path estimates from microwave radiometers and advanced lidar would help develop additional tools for validating NWP models' ability to correctly predict the vertical distribution of microphysical processes.

Finally, this study shows that the combination of radar and surface precipitation measurements provides the necessary information about the vertical evolution of precipitation to inform future improvements in microphysical schemes in NWP models.

Acknowledgments. We want to thank Richard Moore from MET Norway for initiating this study. BJKE work is part of the WISLINE project funded by the Norwegian Research Council, Grant 244106/E10. FH, TS, and ROD would also like to acknowledge support from the European Research Council (ERC) through Grant StG 758005. Furthermore, SJC work is part of the HiLaMS project funded by the National Science Foundation Grant 1531690.

Data availability statement. Historical real time observations can be accessed from the klima Met Norway interface. Three month simulations by Engdahl et al. (2020a) can be found at https://thredds.met.no/thredds/catalog/metusers/bjorgjike-3mnd_ws/catalog.html and archived MEPS via <https://thredds.met.no/thredds/catalog/meps25epsarchive/catalog.html>. The interested reader can contact the authors directly for the MRR reflectivity, OESR surface snowfall, and SWC from the HiLaMS campaign.

REFERENCES

- Barstad, I., and R. B. Smith, 2005: Evaluation of an orographic precipitation model. *J. Hydrometeor.*, **6**, 85–99, <https://doi.org/10.1175/JHM-404.1>.
- Bengtsson, L., and Coauthors, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Mon. Wea. Rev.*, **145**, 1919–1935, <https://doi.org/10.1175/MWR-D-16-0417.1>.
- Berry, E. X., and R. L. Reinhardt, 1974: An analysis of cloud drop growth by collection Part II. Single initial distributions. *J. Atmos. Sci.*, **31**, 1825–1831, [https://doi.org/10.1175/1520-0469\(1974\)031<1825:AAOC DG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<1825:AAOC DG>2.0.CO;2).
- Bigg, E. K., 1953: The supercooling of water. *Proc. Phys. Soc.*, **66B**, 688–694, <https://doi.org/10.1088/0370-1301/66/8/309>.
- Borys, R. D., D. H. Lowenthal, S. A. Cohn, and W. O. J. Brown, 2003: Mountaintop and radar measurements of anthropogenic aerosol effects on snow growth and snowfall rate. *Geophys. Res. Lett.*, **30**, 1538, <https://doi.org/10.1029/2002GL016855>.
- Caniaux, G., J.-L. Redelsperger, and J.-P. Lafore, 1994: A numerical study of the stratiform region of a fast-moving squall line. Part I: General description and water and heat budgets. *J. Atmos. Sci.*, **51**, 2046–2074, [https://doi.org/10.1175/1520-0469\(1994\)051<2046:ANSOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<2046:ANSOTS>2.0.CO;2).
- Chater, A. M., and A. P. Sturman, 1998: Atmospheric conditions influencing the spillover of rainfall to lee of the Southern Alps, New Zealand. *Int. J. Climatol.*, **18**, 77–92, [https://doi.org/10.1002/\(SICI\)1097-0088\(199801\)18:1<77::AID-JOC218>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1097-0088(199801)18:1<77::AID-JOC218>3.0.CO;2-M).
- Cober, S. G., and R. List, 1993: Measurements of the heat and mass transfer parameters characterizing conical graupel growth. *J. Atmos. Sci.*, **50**, 1591–1609, [https://doi.org/10.1175/1520-0469\(1993\)050<1591:MOTHAM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<1591:MOTHAM>2.0.CO;2).
- Cohard, J.-M., and J.-P. Pinty, 2000a: A comprehensive two-moment warm microphysical bulk scheme. I: Description and tests. *Quart. J. Roy. Meteor. Soc.*, **126**, 1815–1842, <https://doi.org/10.1002/qj.49712656613>.
- , and —, 2000b: A comprehensive two-moment warm microphysical bulk scheme. II: 2D experiments with a non-hydrostatic model. *Quart. J. Roy. Meteor. Soc.*, **126**, 1843–1859, <https://doi.org/10.1002/qj.49712656614>.
- Colle, B. A., and C. F. Mass, 2000: The 5–9 February 1996 flooding event over the Pacific Northwest: Sensitivity studies and evaluation of the MM5 precipitation forecasts. *Mon. Wea. Rev.*, **128**, 593–617, [https://doi.org/10.1175/1520-0493\(2000\)128<0593:TFFEOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0593:TFFEOT>2.0.CO;2).
- , M. F. Garvert, J. B. Wolfe, C. F. Mass, and C. P. Woods, 2005: The 13–14 December 2001 IMPROVE-2 event. Part III: Simulated microphysical budgets and sensitivity studies. *J. Atmos. Sci.*, **62**, 3535–3558, <https://doi.org/10.1175/JAS3552.1>.
- Colli, M., R. Rasmussen, J. M. Thériault, L. G. Lanza, C. B. Baker, and J. Kochendorfer, 2015: An improved trajectory model to evaluate the collection performance of snow gauges. *J. Appl. Meteor. Climatol.*, **54**, 1826–1836, <https://doi.org/10.1175/JAMC-D-15-0035.1>.
- , M. Stagnaro, L. G. Lanza, R. Rasmussen, and J. M. Thériault, 2020: Adjustments for wind-induced undercatch in snowfall measurements based on precipitation intensity. *J. Hydrometeorol.*, **21**, 1039–1050, <https://doi.org/10.1175/JHM-D-19-0222.1>.
- Cooper, S. J., N. B. Wood, and T. S. L'Ecuyer, 2017: A variational technique to estimate snowfall rate from coincident radar, snowflake, and fall-speed observations. *Atmos. Meas. Tech.*, **10**, 2557–2571, <https://doi.org/10.5194/amt-10-2557-2017>.
- Cooper, W. A., 1986: Ice initiation in natural clouds. *Precipitation Enhancement—A Scientific Challenge*, Meteor. Monogr., No. 21, Amer. Meteor. Soc., 29–32.
- Engdahl, B. J. K., B. E. K. Nygaard, V. Losnedal, G. Thompson, and L. Bengtsson, 2020a: Effects of the ICE-T microphysics scheme in HARMONIE-AROME on estimated ice loads on transmission lines. *Cold Reg. Sci. Technol.*, **179**, 103139, <https://doi.org/10.1016/j.coldregions.2020.103139>.
- , G. Thompson, and L. Bengtsson, 2020b: Improving the representation of supercooled liquid water in the HARMONIE-AROME weather forecast model. *Tellus*, **72A**, 1–18, <https://doi.org/10.1080/16000870.2019.1697603>.
- Farley, R. D., P. E. Price, H. D. Orville, and J. H. Hirsch, 1989: On the numerical simulation of graupel/hail initiation via the riming of snow in bulk water microphysical cloud models. *J. Appl. Meteor.*, **28**, 1128–1131, [https://doi.org/10.1175/1520-0450\(1989\)028<1128:OTNSOG>2.0.CO;2](https://doi.org/10.1175/1520-0450(1989)028<1128:OTNSOG>2.0.CO;2).
- Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280, [https://doi.org/10.1175/1520-0469\(1994\)051<0249:ADMMPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2).
- Frognier, I.-L., A. T. Singleton, M. O. Körtzow, and U. Andrae, 2019: Convection-permitting ensembles: Challenges related to their design and use. *Quart. J. Roy. Meteor. Soc.*, **145** (Suppl.), 90–106, <https://doi.org/10.1002/qj.3525>.
- Garrett, T. J., C. Fallgatter, K. Shkurko, and D. Howlett, 2012: Fall speed measurement and high-resolution multi-angle photogrammetry of hydrometeors in free fall. *Atmos. Meas. Tech.*, **5**, 2625–2633, <https://doi.org/10.5194/amt-5-2625-2012>.

- Garvert, M. F., C. P. Woods, B. A. Colle, C. F. Mass, P. V. Hobbs, M. T. Stoelinga, and J. B. Wolfe, 2005: The 13–14 December 2001 IMPROVE-2 event. Part II: Comparisons of MM5 model simulations of clouds and precipitation with observations. *J. Atmos. Sci.*, **62**, 3520–3534, <https://doi.org/10.1175/JAS3551.1>.
- Geonorge, 2018: DTM 10 Terrengmodell (UTM33)–Kartverket–Kartkatalogen. Geonorge, accessed 28 May 2018, <https://kartkatalog.geonorge.no/metadata/kartverket/dtm-10-terrengmodell-utm33/dddbb667-1303-4ac5-8640-7ec04c0e3918>.
- Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp.
- Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Wea. Forecasting*, **33**, 739–765, <https://doi.org/10.1175/WAF-D-17-0144.1>.
- Joos, H., and H. Wernli, 2012: Influence of microphysical processes on the potential vorticity development in a warm conveyor belt: A case-study with the limited-area model COSMO. *Quart. J. Roy. Meteor. Soc.*, **138**, 407–418, <https://doi.org/10.1002/qj.934>.
- Kaplan, M. L., C. S. Adaniya, P. J. Marzette, K. C. King, S. J. Underwood, and J. M. Lewis, 2009: The role of upstream midtropospheric circulations in the Sierra Nevada enabling leeside (spillover) precipitation. Part II: A secondary atmospheric river accompanying a midlevel jet. *J. Hydrometeorol.*, **10**, 1327–1354, <https://doi.org/10.1175/2009JHM1106.1>.
- , R. K. Vellore, P. J. Marzette, and J. M. Lewis, 2012: The role of windward-side diabatic heating in Sierra Nevada spillover precipitation. *J. Hydrometeorol.*, **13**, 1172–1194, <https://doi.org/10.1175/JHM-D-11-06.1>.
- Khairoutdinov, M., and Y. Kogan, 2000: A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus. *Mon. Wea. Rev.*, **128**, 229–243, [https://doi.org/10.1175/1520-0493\(2000\)128<0229:ANCPPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0229:ANCPPI>2.0.CO;2).
- Kochendorfer, J., and Coauthors, 2017: Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. *Hydrol. Earth Syst. Sci.*, **21**, 3525–3542, <https://doi.org/10.5194/hess-21-3525-2017>.
- Køltzow, M., 2017: MetCoOp Ensemble Prediction System (MEPS). Accessed April 2020, https://drive.google.com/file/d/0B-SaEtrDE91WWEJoNkJiUm5TNzg/view?usp=sharing&usp=embed_facebook.
- , B. Casati, T. Haiden, and T. Valkonen, 2020: Verification of solid precipitation forecasts from numerical weather prediction models in Norway. *Wea. Forecasting*, **35**, 2279–2292, <https://doi.org/10.1175/WAF-D-20-0060.1>.
- Kulie, M. S., and R. Bennartz, 2009: Utilizing spaceborne radars to retrieve dry snowfall. *J. Appl. Meteor. Climatol.*, **48**, 2564–2580, <https://doi.org/10.1175/2009JAMC2193.1>.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).
- Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations. *Mon. Wea. Rev.*, **139**, 3533–3553, <https://doi.org/10.1175/MWR-D-11-00009.1>.
- Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *J. Geophys. Res.*, **79**, 2185–2197, <https://doi.org/10.1029/JC079i015p02185>.
- Lowenthal, D., and Coauthors, 2016: Isotopic fractionation in wintertime orographic clouds. *J. Atmos. Oceanic Technol.*, **33**, 2663–2678, <https://doi.org/10.1175/JTECH-D-15-0233.1>.
- Meyers, M. P., P. J. DeMott, and W. R. Cotton, 1992: New primary ice-nucleation parameterizations in an explicit cloud model. *J. Appl. Meteor. Climatol.*, **31**, 708–721, [https://doi.org/10.1175/1520-0450\(1992\)031<0708:NPINPI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1992)031<0708:NPINPI>2.0.CO;2).
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>.
- Müller, M., and Coauthors, 2017: AROME-MetCoOp: A Nordic convective-scale operational weather prediction model. *Wea. Forecasting*, **32**, 609–627, <https://doi.org/10.1175/WAF-D-16-0099.1>.
- Musil, D. J., 1970: Computer modeling of hailstone growth in feeder clouds. *J. Atmos. Sci.*, **27**, 474–482, [https://doi.org/10.1175/1520-0469\(1970\)027<0474:CMOHGI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1970)027<0474:CMOHGI>2.0.CO;2).
- Nelson, S. P., 1983: The influence of storm flow structure on hail growth. *J. Atmos. Sci.*, **40**, 1965–1983, [https://doi.org/10.1175/1520-0469\(1983\)040<1965:TIOFSF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1965:TIOFSF>2.0.CO;2).
- Nitu, R., and Coauthors, 2018: WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012–15). IOM Tech. Rep. 131, WMO, Geneva, Switzerland, 1445 pp.
- Pinty, J.-P., and P. Jabouille, 1998: A mixed-phased cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation. Preprints, *Conf. on Cloud Physics*, Everett, WA, Amer. Meteor. Soc., 217–220.
- Ramelli, F., and Coauthors, 2021a: Influence of low-level blocking and turbulence on the microphysics of a mixed-phase cloud in an inner-Alpine valley. *Atmos. Chem. Phys.*, **21**, 5151–5172, <https://doi.org/10.5194/acp-21-5151-2021>.
- , and Coauthors, 2021b: Microphysical investigation of the seeder and feeder region of an Alpine mixed-phase cloud. *Atmos. Chem. Phys.*, **21**, 6681–6706, <https://doi.org/10.5194/acp-21-6681-2021>.
- Schirle, C. E., S. J. Cooper, M. A. Wolff, C. Pettersen, N. B. Wood, T. S. L'Ecuyer, T. Ilmo, and K. Nygård, 2019: Estimation of snowfall properties at a mountainous site in Norway using combined radar and in situ microphysical observations. *J. Appl. Meteor. Climatol.*, **58**, 1337–1352, <https://doi.org/10.1175/JAMC-D-18-0281.1>.
- Schwartz, C. S., 2014: Reproducing the September 2013 record-breaking rainfall over the Colorado front range with high-resolution WRF forecasts. *Wea. Forecasting*, **29**, 393–402, <https://doi.org/10.1175/WAF-D-13-00136.1>.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, <https://doi.org/10.1175/2010MWR3425.1>.
- Sinclair, M. R., D. S. Wratt, R. D. Henderson, and W. R. Gray, 1997: Factors affecting the distribution and spillover of precipitation in the Southern Alps of New Zealand—A case study. *J. Appl. Meteor.*, **36**, 428–442, [https://doi.org/10.1175/1520-0450\(1997\)036<0428:FATDAS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<0428:FATDAS>2.0.CO;2).
- The MetCoOp Team, 2017: The MetCoOp ensemble MEP. *ALADIN-HIRLAM Newsletter*, No. 10, CNRM, France, 98–103, <http://www.umr-cnrm.fr/aladin/spip.php?article260>.
- Thériault, J. M., R. Rasmussen, K. Ikeda, and S. Landolt, 2012: Dependence of snow gauge collection efficiency on snowflake

- characteristics. *J. Appl. Meteor. Climatol.*, **51**, 745–762, <https://doi.org/10.1175/JAMC-D-11-0116.1>.
- , —, E. Petro, J.-Y. Trépanier, M. Colli, and L. G. Lanza, 2015: Impact of wind direction, wind speed, and particle characteristics on the collection efficiency of the double fence intercomparison reference. *J. Appl. Meteor. Climatol.*, **54**, 1918–1930, <https://doi.org/10.1175/JAMC-D-15-0034.1>.
- Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2).
- , P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- Wang, P. K., and W. Ji, 2000: Collision efficiencies of ice crystals at low–intermediate Reynolds numbers colliding with supercooled cloud droplets: A numerical study. *J. Atmos. Sci.*, **57**, 1001–1009, [https://doi.org/10.1175/1520-0469\(2000\)057<1001:CEOICA>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<1001:CEOICA>2.0.CO;2).
- Wolff, M. A., K. Isaksen, R. Brækkan, E. Alfnes, A. Petersen-Øverleir, and E. Ruud, 2013: Measurements of wind-induced loss of solid precipitation: Description of a Norwegian field study. *Hydrol. Res.*, **44**, 35–43, <https://doi.org/10.2166/nh.2012.166>.
- , —, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan, 2015: Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: Results of a Norwegian field study. *Hydrol. Earth Syst. Sci.*, **19**, 951–967, <https://doi.org/10.5194/hess-19-951-2015>.
- Wood, N. B., 2011: Estimation of snow microphysical properties with application to millimeter-wavelength radar retrievals for snowfall rate. Ph.D. thesis, Colorado State University, 248 pp., <http://hdl.handle.net/10217/48170>.
- , T. S. L'Ecuyer, A. J. Heymsfield, and G. L. Stephens, 2015: Microphysical constraints on millimeter-wavelength scattering properties of snow particles. *J. Appl. Meteor. Climatol.*, **54**, 909–931, <https://doi.org/10.1175/JAMC-D-14-0137.1>.