RESEARCH ARTICLE

The potential for uncertainty in Numerical Weather Prediction model verification when using solid precipitation observations

Samuel T. Buisán[1] | Craig D. Smith[2] | Amber Ross[2] | John Kochendorfer[3] |
José Luís Collado[1] | Javier Alastrué[1] | Mareile Wolff[4] | Yves-Alain Roulet[5] |
Michael E. Earle[6] | Timo Laine[7] | Roy Rasmussen[8] | Rodica Nitu[9]

[1]Delegación Territorial de AEMET (Spanish State Meteorological Agency) en Aragón, Zaragoza, Spain

[2]Environment and Climate Change Canada, Saskatoon, Saskatchewan, Canada

[3]National Atmospheric and Oceanic Administration, Air Resources Laboratory, Atmospheric Turbulence and Diffusion Division, Oak Ridge, Tennessee

[4]Norwegian Meteorological Institute, Oslo, Norway

[5]Meteoswiss, Payerne, Switzerland

[6]Environment and Climate Change Canada, Dartmouth, Nova Scotia, Canada

[7]Finnish Meteorological Institute, Helsinki, Finland

[8]National Center for Atmospheric Research, Boulder, Colorado

[9]Environment and Climate Change Canada, Toronto, Ontario, Canada

Correspondence
Samuel T. Buisán, Delegación Territorial de AEMET (Spanish State Meteorological Agency) en Aragón, Paseo del Canal 17, Zaragoza 50007, Spain.
Email: sbuisans@aemet.es

Abstract

Precipitation forecasts made by Numerical Weather Prediction (NWP) models are typically verified using precipitation gauge observations that are often prone to the wind-induced undercatch of solid precipitation. Therefore, apparent model biases in solid precipitation forecasts may be due in part to the measurements and not the model. To reduce solid precipitation measurement biases, adjustments in the form of transfer functions were derived within the framework of the World Meteorological Organization Solid Precipitation Inter-Comparison Experiment (WMO-SPICE). These transfer functions were applied to single-Alter shielded gauge measurements at selected SPICE sites during two winter seasons (2015–2016 and 2016–2017). Along with measurements from the WMO automated field reference configuration at each of these SPICE sites, the adjusted and unadjusted gauge observations were used to analyze the bias in a Global NWP model precipitation forecast. The verification of NWP winter precipitation using operational gauges may be subject to verification uncertainty, the magnitude and sign of which varies with the gauge-shield configuration and the relation between model and site-specific local climatologies. The application of a transfer function to alter-shielded gauge measurements increases the amount of solid precipitation reported by the gauge and therefore reduces the NWP precipitation bias at sites where the model tends to overestimate precipitation, and increases the bias at sites where the model underestimates the precipitation. This complicates model verification when only operational (non-reference) gauge observations are available. Modelers, forecasters, and climatologists must consider this when comparing modeled and observed precipitation.

KEYWORDS
NWP verification, snow, solid precipitation, SPICE

1 | INTRODUCTION

Winter precipitation forecasts are needed to help address meteorological and hydrological hazards; however, solid precipitation measurements available for data assimilation and forecast verification are affected by potentially large undercatch errors (Goodison et al., 1998; Rasmussen et al., 2012). These errors propagate directly into precipitation forecasts, and affect model climatology, data assimilation, and nowcasting.

Brun et al. (2013) determined that the total amount of ERA-Interim precipitation during December, January, and February in northern Eurasia exceeded the uncorrected measurement from standard precipitation gauges by 18%. Lespinas et al. (2015) demonstrated that a strong negative bias in observed winter precipitation makes the assimilation of such observations difficult in the Canadian Precipitation Analyses product (CaPA; Fortin et al., 2018), and determined that winter season precipitation could not be used for verification because of the large observational errors. Vionnet et al. (2015) analyzed the performance of the Global Environmental Multiscale Model (GEM) in complex terrain during the winter, but excluded winter precipitation. To minimize false verifications due to known undercatch, Schirmer et al. (2015) attempted to verify Numerical Weather Prediction (NWP) model forecasts in mountainous terrain using observations from ultrasonic snow depth measurements and snow pillows instead of gauge measurements. Likewise, Lopez et al. (2013) discarded precipitation gauge measurements during snowfall events to avoid errors in the experimental 4D-Var assimilation of SYNOP rain gauge data at the European Centre for Medium Range Weather Forecast (ECMWF).

Discrepancies between the magnitude of measured and actual solid precipitation hinder the use of automated gauge measurements in winter for NWP assimilation and verification. For this reason, procedures, methodologies and related studies attempting to adjust the bias in solid precipitation measurements are vital to the scientific and operational communities using NWP precipitation data. In the framework of the WMO-SPICE (Nitu et al., 2019), a set of transfer functions was derived for adjusting the wind-induced undercatch of solid precipitation measurements recorded by weighing gauges (Kochendorfer et al., 2017a, 2017b, 2018).

Recently, a unique precipitation dataset consisting of high-quality post-SPICE data recorded at selected sites (Smith et al., 2019a) became available for the 2015–2016 and 2016–2017 winter seasons. This new dataset is independent from the data used for the SPICE transfer function development, and it is used in this study to analyze the uncertainty in winter precipitation verification associated with the wind-induced undercatch.

The goal of this work is to demonstrate the potential for uncertainty when solid precipitation measurements from automated gauges are used for NWP verification. This is accomplished by evaluating the biases between a Global NWP Model and each of the following: the reference precipitation measurement from a double-fence shielded gauge; the measured (unadjusted) precipitation amount from automated gauges; and the measured precipitation amount adjusted using the SPICE transfer functions. This evaluation is conducted using measurements from the SPICE sites in different climate regimes that operated a WMO Double Fence Automated Reference (DFAR; Nitu et al., 2019) serving as the reference configuration. The potential magnitude of verification uncertainty, their dependency on the precipitation gauge and wind shield configuration, and their relation to the model forecast accuracy is assessed for each site.

For this analysis, the "true bias" is the observed bias between the reference (DFAR) and the model, the "apparent bias" is the bias between the unadjusted operational automated precipitation gauge and the model (i.e., the configuration that can lead to verification uncertainty), and the "adjusted bias" is the bias between the adjusted precipitation (from the operational gauge) and the model. We tested the hypothesis that adjusting the operational gauge measurements using the SPICE transfer functions provides a more accurate NWP verification in the absence of a reference. Transfer function limitations are also evaluated and discussed (Kochendorfer et al., 2017b; Smith et al., 2019b). The overall assessment and characterization of the bias in the NWP precipitation product is not within the scope of this study.

2 | DATA AND METHODS

The precipitation data used in this analysis were obtained from post-SPICE observations during the 2015–2016 and 2016–2017 winter seasons at the CARE, Bratt's Lake, Marshall, Haukeliseter, Sodankylä, Weissfluhjoch, and Formigal-Sarrios SPICE sites. A map is provided in Figure 1, with details pertinent to the present analysis in Table 1. A detailed layout and description of each site is included in the WMO-SPICE site commissioning reports (see link to reports in reference section) and in the WMO-SPICE final report (Nitu et al., 2019). The in situ observations used in this analysis were derived from either the DFAR (reference gauge), which consisted of a Geonor T-200B3 or OTT Pluvio^2 automated gauge with a single-Alter shield inside an octagonal double fence, or a single-Alter shielded Geonor T-200B3 or OTT Pluvio^2 automated gauge (operational gauge). The Haukeliseter (Norway) site is shown in Figure 1 as an example.

FIGURE 1 Map of the SPICE sites included in this study. © Google Maps/Google Earth. Inset photo is an image of the Haukeliseter site showing the operational gauge (OTT Pluvio^2 weighing gauge with a single alter shield) in the foreground and the reference configuration (DFAR) in the background. Similar configurations exists at each of the sites in this study

TABLE 1 List of SPICE sites used in this study, including their climate zone and elevation (geographical and nearest grid point from the model)

Site	Acronym	Country	Climate zone	Elevation (m) geographical /nearest grid point	DFAR (mm) D/A/W (%)	Pearson correlation
Centre for Atmospheric Research Experiments (CARE)	CAR	Canada	Humid continental subject to lake effect	251/242	780.19 32/97/50	0.77**
Formigal-Sarrios	FOR	Spain	Alpine climate with Atlantic influence	1,800/1,919	1,772.9 41/92/20	0.87**
Haukeliseter	HKL	Norway	Mountains, well above the tree line	991/1,071	1,451.04 36/97/65	0.90**
Marshall	MAR	United States	Continental	1,742/1,646	540.61 15/71/45	0.88**
Sodankyla	SOD	Finland	Northern Boreal	179/204	583.27 29/100/0	0.86**
Weissfluhjoch	WFJ	Switzerland	Alpine	2,537/2,130	1,495.36 33/92/28	0.83**
Bratt's Lake	XBK	Canada	Continental	585/583	100.74 16/100/61	0.81**

Note: The sixth column shows the accumulated precipitation in the reference (DFAR) during the study period and the percentage of days from total with precipitation (D) (accumulation > 1 mm), percentage of these days that were adjusted (A), and percentage of these days with wind conditions higher than $U > 4$ m/s that were adjusted (W). The last column shows the Pearson correlation coefficient between daily observed (DFAR) and model forecasted precipitation **p-value < .01.

The precipitation measurement data were processed as described in Smith *et al.* (2019b); outliers (e.g., from gauge servicing) were removed from the 1-min time series using range checks and a jump filter, and high-frequency noise was reduced using a Gaussian filter. Precipitation amounts were then identified using a signal aggregation processing technique (Pan *et al.*, 2016; Smith *et al.*, 2019b) for all weighing gauges (Geonor T-200B3 or Pluvio^2) in the following configurations: Double Fence Automated Reference (DFAR) and Single-Alter shielded (SA). Finally, the 1-min time series were resampled to produce 30-min accumulation datasets. The daily accumulations were calculated based on the sum of all 30-min data both before and after adjustment for wind bias.

The 30-min data from the SA precipitation gauges were adjusted using the WMO-SPICE single-Alter eq. 3 from Kochendorfer *et al.* (2017b):

$$CE = e^{-0.0348(U)(1-\tan^{-1}(1.366(T_{air}))+0.779)}$$

where (CE) is the catch efficiency calculated for each 30-min period, U is wind speed in m/s and Tair is air temperature in degrees Celsius. Both U and Tair were 30-min means. Finally, the adjusted precipitation is calculated as follows:

$$\text{Adjusted precipitation} = \frac{\text{Observed precipitation at operational gauge}}{CE}$$

Because the transfer function requires both wind speed (U) and air temperature (Tair), if either of these are missing for a 30-min period, the adjustment cannot be made. This impacts the amount of data available for intercomparison. Days with missing 30-min precipitation, wind speed, or air temperature measurements were excluded from the intercomparison, but this comprised fewer than 5% of the available days. The length of the winter period also varied depending on the site, so the number of intercomparison days was different for each site. However, a minimum of 150 days for each winter season were available for analysis at each site.

The 24-hr forecasted accumulations were retrieved at the nearest grid point of each SPICE site from the high-resolution operational ECMWF model run at 00 UTC. The elevation of the nearest grid point (Table 1) is, in general, close to that of each site, with the exception of Formigal and Weissfluhjoch that are both located in alpine environments. The assumption that the in-situ precipitation measurement was comparable to the NWP gridded output was based on the following: relative to

other seasons, winter precipitation is usually more spatially homogeneous, as it is not typically caused by convective activity, and the longer 24-hr accumulation period reduced the uncertainty associated with shorter forecasted periods for precipitation accumulation.

3 | RESULTS

Table 1 shows the percentage of days from total (D) with precipitation (accumulation > 1 mm), percentage of these days that were adjusted (A), and percentage of these days with wind conditions higher than $U > 4$ m/s that were adjusted (W). At all sites, more than 90% of the days with precipitation during the winter were adjusted, with the exception of Marshall (70%), where winter seasons were characterized by a larger proportion of liquid precipitation events. Haukeliseter and Bratt's Lake experienced more windy adjusted days than the other sites, which was in agreement with SPICE results (Nitu et al., 2018).

Table 1 also shows that significant precipitation events were captured consistently by the model at all sites, as illustrated by the Pearson correlation (r) between daily ECMWF forecasted and DFAR observed precipitation amounts. For all sites, there was a significant statistical correlation (p-value < .01), with maximum values of r close to .90 at Haukeliseter and minimum values of 0.77 at CARE. The relatively high correlations may demonstrate the overall quality of the model forecast, but does not address the bias, whether real or related to the observation bias.

Figure 2 shows the time series of accumulated precipitation measurements and the accumulated ECMWF modeled precipitation at the CARE and Weissfluhjoch sites. As shown in the figure, the true bias was site
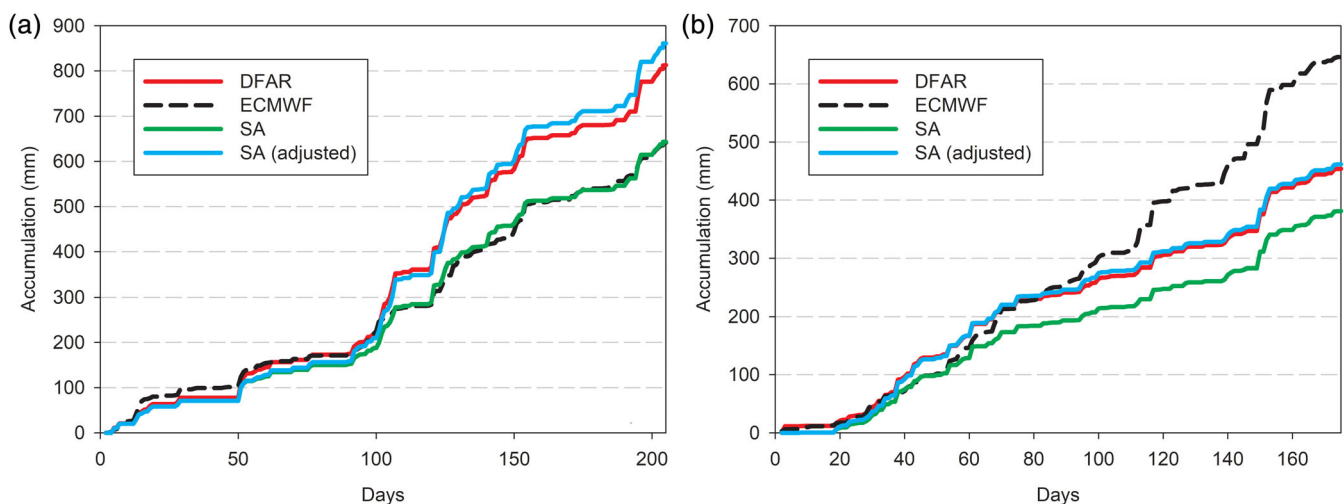
FIGURE 2 Seasonal accumulation of precipitation 2016/2017 as forecasted by ECMWF (blue), and measured by DFAR (red), SA (orange) and adjusted precipitation using eq. 3 (adjusted) at two selected sites (a) Weissflujoch and (b) CARE

FIGURE 3 Daily bias (%) between the DFAR precipitation, the SA precipitation, and the adjusted precipitation SA (adjusted) compared to the ECMWF forecasted precipitation
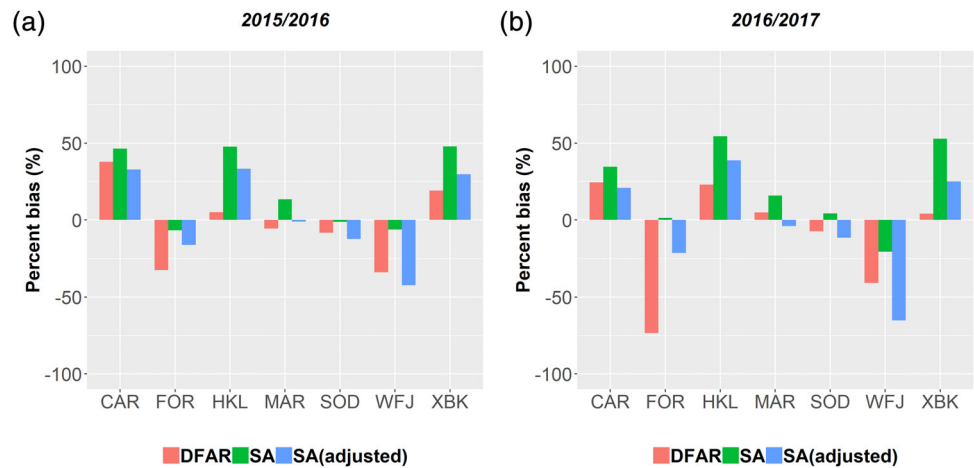
dependent, as illustrated by the differences between the ECMWF total accumulation and the DFAR-measured accumulation. Weissfluhjoch shows a negative true bias (150 mm or 20% of total precipitation) in ECMWF total precipitation relative to the DFAR. However, the apparent bias is near zero; without the DFAR measurements or an adjustment, this would incorrectly validate the model performance. When the universal transfer function was applied, the adjusted bias more accurately assessed the model performance at this site. CARE shows a large positive true bias (200 mm or 45% of total precipitation) in ECMWF total precipitation relative to the DFAR; the model overestimates the winter precipitation at this site. The apparent bias is even higher, so without DFAR measurements or an adjustment, this would lead to a larger error (250 mm or 55%). At CARE, the adjusted bias agrees with the DFAR measurements and with the true bias. Following the same procedure, the behavior and potential uncertainty in the verification results for all of the SPICE sites shown in Table 1 were analyzed.

There were significant differences in total accumulation among the SPICE sites, which were higher for alpine sites and lower for continental and boreal sites (Table 1). Measurement differences among the sites are summarized in Figure 3, which shows biases (true, apparent, and adjusted) for each site. To avoid the impact of the site-specific amount of precipitation accumulation on the site inter-comparison, instead of absolute bias we use percentage bias (%). This parameter was calculated as the difference between the model and the observation relative to the model, and was used to demonstrate the potential for verification uncertainty across sites. The apparent bias in the model (i.e., as compared to the unadjusted gauge) varied substantially by site. These differences may be related to regional differences in model performance, but may also be highly dependent on the catch efficiency of the gauge (which in turn is dependent on wind speed and precipitation type). The effects of the adjustment on the bias were

also site-dependent, partially due to the inherent differences in the model biases across sites, and partially due to the appropriateness of the universal transfer function for a specific site (Kochendorfer et al., 2017b; Smith et al., 2019b).

For measurements from CARE, the model significantly overestimated the precipitation in both seasons, and the adjustment performed well (the adjusted SA accumulation was closer to the reference amount). At Formigal, the model significantly underestimated the precipitation in both seasons, but to a lesser degree (the model amount was closer to the reference amount) during 2015–2016. The universal adjustment for Formigal is too mild, and therefore only partially corrects the undercatch. The unadjusted SA amounts and modeled precipitation estimates agreed with each other, which could lead to erroneous conclusions on the model performance because undercatch was not accounted for. At Haukeliseter the model showed good agreement with the DFAR during 2015–2016, but verification against unadjusted measurements showed an apparent over-estimation. Verification against adjusted measurements still resulted in a large positive bias, because the universal transfer function under-adjusted the SA at this site (as in Formigal), although the adjusted bias was still smaller than the apparent bias. Similarly, during 2016–2017 the model overestimated precipitation compared to DFAR, and the precipitation overestimation was even larger when compared to unadjusted measurements. Using the adjusted SA precipitation, as expected, the positive bias was reduced compared to the unadjusted measurements, but was still larger than for the DFAR (since the transfer function under-adjust at this site).

At Marshall, during the 2015–2016 winter season, the model agreed with the DFAR, and the adjustment slightly underestimated the precipitation amount relative to that measured by the DFAR. During the 2016–2017 winter season, the model continued to agree with the DFAR, and the adjustment slightly overestimated the

seasonal accumulation. At Sodankylä, the model slightly underestimated the precipitation for both seasons. The adjusted Sodankylä SA precipitation was slightly over-estimated compared to the DFAR, but in general, the agreement was quite good. Because the wind speed was generally quite low at Sodankylä, the percent differences between the SA, the DFAR, and the adjusted SA measurements were smaller than at the windier sites. At Weissfluhjoch, the model underestimated the precipitation for both seasons as compared to the DFAR. The universal adjustment for Weissfluhjoch was too strong, and there-fore over-adjusted the precipitation for both winter sea-sons when compared to the DFAR. The operational unadjusted gauge showed good agreement with the model, especially during 2015–2016, and this could influence verification interpretation. At Bratt's Lake, the model overestimated the precipitation amount relative to the DFAR, and the universal adjustment under-adjusted the precipitation, especially during the 2016–2017 winter season.

At CARE, Haukeliseter and Bratt's Lake, the apparent bias was positive and greater than the true bias, but it was improved when the precipitation was adjusted (adjusted bias), increasing the measured amounts such that they compared better with the model. At Haukeliseter, a significant difference between the true and apparent bias was evident, which was attributed to the low catch ratio of the SA relative to the DFAR, the universal adjustment is however too mild for Haukeliseter, and leads to under-adjustment, so that the adjusted bias remains largely positive (indicating an over-forecast). At Formigal, Weissfluhjoch, and Sodankylä, the apparent bias was negative and smaller in absolute value than the true bias. Because the model underestimated precipitation at these three sites, the bias actually worsened after increasing the SA precipitation measurements with the adjustment (adjusted bias). This change was noticeable at Formigal, but almost negligible at Sodankylä. In the case of Weissfluhjoch, where the SA measurements were over-adjusted, the adjusted bias was worse than the true bias. Finally, at Marshall, the over-adjustment produced a negative adjusted bias, whereas the apparent bias was positive (in 2016/2017).

Figure 3 shows that the highest apparent bias (and largest difference between this apparent and true bias) is at Haukeliseter and Bratt's Lake. Since Haukeliseter is the site with the highest mean 30-min wind speed during snowfall events (6.7 m/s), with values up to 20 m/s (Kochendorfer et al., 2017b), there were numerous events for which the SA catch ratio was very small or even zero. The adjustment of a very small amount results in high uncertainty in the adjusted value and the adjustment when the catch is zero is not possible. The finding also

applies to Bratt's Lake, which had the second highest mean wind speed (4.4 m/s). For other sites with higher rates of precipitation, such as Formigal, Weissfluhjoch, and CARE, the apparent bias was similar or lower than for Haukeliseter and Bratt's Lake.

To evaluate the impact of wind on the results, the data were filtered to include only days with daily average wind speeds lower than 4 m/s. Figure 4 shows that for the sites that exhibited a higher apparent bias (Haukeliseter and Bratt's lake), this bias was reduced for events occurring at lower wind speeds, producing results similar to the other sites where the apparent bias was similar to Figure 3. This result indicates that transfer functions are difficult to apply under high wind conditions because the catch ratios can be low (and potentially zero), such that errors and biases can be significantly augmented by the large adjustment that is required. This is consistent with the large uncertainties noted for windy sites with single-Alter or unshielded gauges by Kochendorfer et al. (2017b) and Smith et al. (2019b), and an uncertainty analysis performed using several different wind shields (Kochendorfer et al., 2018).

4 | DISCUSSION AND CONCLUSIONS

This work illustrates the complexity of NWP model fore-cast precipitation verification for winter precipitation using automated gauge measurements. The main conclu-sions are:

i The adjustment of SA-shielded gauge measurements always resulted in precipitation amounts that were closer to the DFAR measurements.
ii At sites where the model overestimated precipitation as compared to the DFAR, such as CARE, Haukeliseter, and Bratt's Lake, the adjusted precipi-tation reduced the bias.
iii At sites where the model underestimated precipitation, such as Weissfluhjoch, Formigal, Sodankylä, and Mar-shall, the adjusted precipitation increased the bias.
iv The universal SA transfer function performance was variable (Smith et al., 2019b), as it under-adjusted certain sites and over-adjusted others. This intro-duces additional uncertainty in model verification results. However, adjusted measurements were still more accurate (i.e., closer to the reference) than unadjusted measurements.
v DFAR observations, which provide more accurate measurements of solid precipitation, are necessary to evaluate model-forecasted precipitation estimates. In the absence of a DFAR, adjusting gauge measure-ments of winter precipitation with transfer functions

FIGURE 4 Daily bias (%) between the DFAR precipitation, the SA precipitation, and the adjusted precipitation SA (adjusted)compared to the ECMWF forecasted precipitation for days where the daily average winds were below 4 m/s (scale is maintained relative to Figure 3)
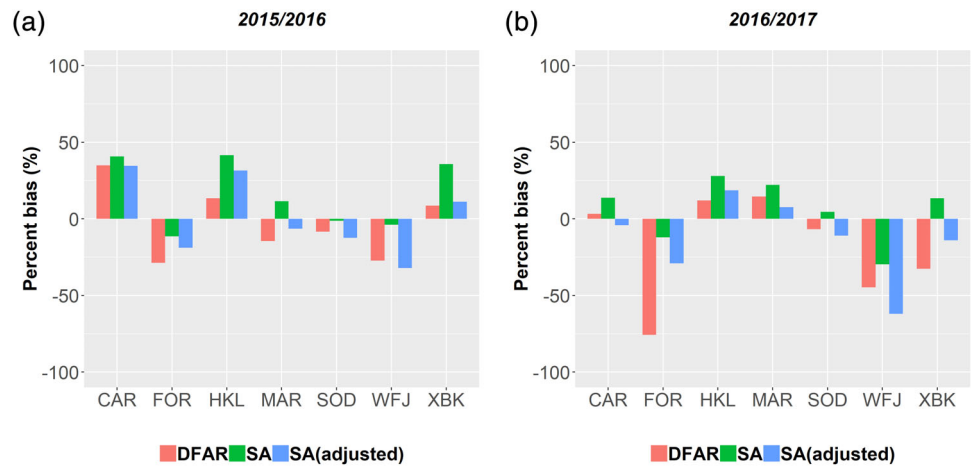
is critical for a better assessment of the model precipitation bias, but the limitations and uncertainty of transfer functions cannot be quantified without reference (DFAR) precipitation measurements.

vi For the verification of modeled winter precipitation, data from locations characterized by high wind should be used with caution, as the unadjusted catch efficiency may be low or zero, as was the case for Bratt's Lake (Smith *et al.*, 2019b), hampering the efficacy of the adjustment.

The verification of winter precipitation with operational gauges in the absence of a DFAR will only provide an estimate of the apparent bias between the observation and the model, and may therefore result in potential verification misinterpretation. The application of transfer functions to adjust measurements from operational gauges allows for the estimation of the adjusted bias, which, depending on the specific site, may be more representative of the bias between the operational gauge and NWP forecasted amounts.

Other transfer functions found in the literature (Goodison *et al.*, 1998; Wolff *et al.*, 2015; Buisán *et al.*, 2017; Colli *et al.*, 2018; Kochendorfer *et al.*, 2018) can be applied to weighing gauges in different configurations, manual gauges, or even tipping buckets for model verification, but the potential verification uncertainty would still persist.

These results, as well as those from Smith *et al.* (2019b), demonstrate varied performance of the SPICE transfer functions at different sites, suggesting that further work is required to develop and test site- or climate-specific transfer functions, for example, by using a multi-year analysis. Additionally, more work is needed to evaluate the performance of the adjustment function in areas without DFARs, but with similar climatic conditions. Hydrometeor characteristics and fall velocity have also been shown to affect catch efficiency (Thériault *et al.*, 2012) and should be considered in further development and application of

transfer functions. This could also lead to recommendations on improving ancillary measurements at operational sites such that solid precipitation measurements and adjustments are improved for better NWP verification.

These conclusions and recommendations were based on the intercomparison with the ECMWF model, with a simple verification method, but the same conclusions on the potential verification uncertainty could apply to any other model evaluation. Irrespective of the details of the model verification approach employed, in the absence of a DFAR or other reference observation, winter precipitation verification is complex and subject to uncertainty. The amplitude of the undercatch as well as the performance of the undercatch adjustment could potentially change from year to year and from site to site, but in general, the problems associated with winter verification will persist. Although the number of SPICE sites is limited, the solid precipitation data collected at these sites is of the highest quality, and can be used to continue to inform the development, refinement, and application of transfer functions in various climate regimes for applications such as model verification.

These findings should be considered by modelers, forecasters, and climatologists to avoid misinterpreting verification results between modeled and observed precipitation.

ORCID
Samuel T. Buisán https://orcid.org/0000-0002-9551-9918

REFERENCES
Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F. and Morin, S. (2013) Simulation of northern Eurasian local snow depth, mass and density using a detailed snowpack model and meteorological reanalyses. *Journal of Hydrometeorology*, 14, 203–219.
Buisán, S.T., Earle, M.E., Collado, J.L., Kochendorfer, J., Alastrué, J., Wolff, M., Smith, C.D. and López-Moreno, J.I. (2017) Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network.

Atmospheric Measurement Techniques, 10, 1079–1091. https://doi.org/10.5194/amt-10-1079-2017.

Colli, M., Stagnaro, M., Lanza, L., Rasmussen, R. and Thériault, J. M. (2018) Using snowfall intensity to improve the correction of wind-induced undercatch in solid precipitation measurements. *Hydrology and Earth System Sciences*, 2018–447. https://doi.org/10.5194/hess-2018-447.

Fortin, V., Roy, G., Stadnyk, T., Koenig, K., Gasset, N. and Mahidjiba, A. (2018) Ten years of science based on the Canadian precipitation analysis: a CaPA system overview and literature review. *Atmosphere-Ocean*, 56(3), 178–196. https://doi.org/10.1080/07055900.2018.1474728.

Goodison, B.E., Louie, P.Y.T. and Yang, D. (1998) *WMO solid precipitation measurement intercomparison*. WMO Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp.

Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M.E., Reverdin, A., Wong, K., Smith, C.D., Yang, D., Roulet, Y.-A., Buisan, S., Laine, T., Lee, G., Aceituno, J.L.C., Alastrué, J., Isaksen, K., Meyers, T., Brækkan, R., Landolt, S., Jachcik, A. and Poikonen, A. (2017b) Analysis of single-alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. *Hydrology and Earth System Sciences*, 21, 3525–3542. https://doi.org/10.5194/hess-21-3525-2017.

Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M.E., Reverdin, A., Wong, K., Smith, C.D., Yang, D., Roulet, Y.-A., Meyers, T., Buisan, S., Isaksen, K., Brækkan, R., Landolt, S. and Jachcik, A. (2018) Testing and development of transfer functions for weighing precipitation gauges in WMO-SPICE. *Hydrology and Earth System Sciences*, 22, 1437–1452. https://doi.org/10.5194/hess-22-1437-2018.

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M.E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R. and Leeper, R. (2017a) The quantification and correction of wind-induced precipitation measurement errors. *Hydrology and Earth System Sciences*, 21, 1973–1989. https://doi.org/10.5194/hess-21-1973-2017.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P. and Stadnyk, T. (2015) Performance evaluation of the Canadian precipitation analysis (CaPA). *Journal of Hydrometeorology*, 16(5), 2045–2064.

Lopez, P. (2013) Experimental 4D-var assimilation of SYNOP rain gauge data at ECMWF. *Monthly Weather Review*, 141, 1527–1544.

Nitu, R., Roulet, Y.-A., Wolff, M., Earle, M., Reverdin, A., Smith, C., Kochendorfer, J., Morin, S., Rasmussen, R., Wong, K., Alastrué, J., Arnold, L., Baker, B., Buisán, S., Collado, J.L., Colli, M., Collins, B., Gaydos, A., Hannula, H.-R., Hoover, J., Joe, P., Kontu, A., Laine, T., Lanza, L., Lanzinger, E., Lee, G. W., Lejeune, Y., Leppänen, L., Mekis, E., Panel, J.-M., Poikonen, A., Ryu, S., Sabatini, F., Theriault, J., Yang, D., Genthon, C., van den Heuvel, F., Hirasawa, N., Konishi, H., Nishimura, K. and Senese, A. (2018) *WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012 – 2015)*, WMO-IOM Report No. 131.

Nitu, R., Roulet, Y.-A., Wolff, M., Earle, M., Reverdin, A., Smith, C., Kochendorfer, J., Morin, S., Rasmussen, R., Wong, K., Alastrué, J., Arnold, L., Baker, B., Buisán, S., Collado, J.L., Colli, M., Collins, B., Gaydos, A., Hannula, H.-R., Hoover, J., Joe, P., Kontu, A., Laine, T., Lanza, L., Lanzinger, E., Lee, G.W., Lejeune, Y., Leppänen, L., Mekis, E., Panel, J.-M., Poikonen, A., Ryu, S., Sabatini, F., Theriault, J., Yang, D., Genthon, C., van den Heuvel, F., Hirasawa, N., Konishi, H., Nishimura, K. and Senese, A. (2019) *WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012 – 2015)*, WMO-IOM Report No. 131.

Pan, X., Yang, D., Li, Y., Barr, A., Helgason, W., Hayashi, M., Marsh, P., Pomeroy, J. and Janowicz, R.J. (2016) Bias corrections of precipitation measurements across experimental sites in different ecoclimatic regions of western Canada. *The Cryosphere*, 10, 2347–2360. https://doi.org/10.5194/tc-10-2347-2016.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A.P., Black, J., Thériault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K. and Gutmann, E. (2012) How well are we measuring snow: the NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society*, 93, 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1.

Schirmer, M. and Jamieson, B. (2015) Verification of analysed and forecasted winter precipitation in complex terrain. *The Cryosphere*, 9, 587–601. https://doi.org/10.5194/tc-9-587-2015.

Smith, C.D., Ross, A., Kochendorfer, J., Earle, M.E., Wolff, M., Buisan, S., Roulet, Y.-A. and Laine, T. (2019a) The post-SPICE (2015/2016 and 2016/2017) winter precipitation intercomparison data. *PANGAEA*. https://doi.org/10.1594/PANGAEA.907379.

Smith, C.D., Ross, A., Kochendorfer, J., Earle, M.E., Wolff, M., Buisan, S., Roulet, Y.-A. and Laine, T. (2019b) Evaluation of the WMO-SPICE transfer functions for adjusting the wind bias in solid precipitation measurements. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-2019-313 in review.

Thériault, J.M., Rasmussen, R., Ikeda, K. and Landolt, S. (2012) Dependence of snow gauge collection efficiency on snowflake characteristics. *Journal of Applied Meteorology and Climatology*, 51, 745–762. https://doi.org/10.1175/JAMC-D-11-0116.1.

Vionnet, V., Bélair, S., Girard, C. and Plante, A. (2015) Wintertime subkilometer numerical forecasts of near-surface variables in the Canadian Rocky Mountains. *Monthly Weather Review*, 143, 666–686.

Wolff, M.A., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T. and Brækkan, R. (2015) Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19, 951–967. https://doi.org/10.5194/hess-19-951.