Verification of Solid Precipitation Forecasts from Numerical Weather Prediction Models in Norway

MORTEN KØLTZOW,^a BARBARA CASATI,^b THOMAS HAIDEN,^c AND TERESA VALKONEN^a

^a Norwegian Meteorological Institute, Oslo, Norway; ^b Environment and Climate Change Canada, Toronto, Ontario, Canada; ^c ECMWF, Reading, United Kingdom

(Manuscript received 20 April 2020, in final form 12 August 2020)

ABSTRACT: Assessing the quality of precipitation forecasts requires observations, but all precipitation observations have associated uncertainties making it difficult to quantify the true forecast quality. One of the largest uncertainties is due to the wind-induced undercatch of solid precipitation gauge measurements. This study discusses how this impacts the verification of precipitation forecasts for Norway for one global model [the high-resolution version of the ECMWF Integrated Forecasting System (IFS-HRES)], and one high-resolution, limited-area model [Applications of Research to Operations at Mesoscale (MEPS)]. First, the forecasts are compared with high-quality reference measurements (less undercatch) and with more simple measurement equipment commonly available (substantial undercatch) at the Haukelisetter observation site. Then the verification is extended to include all Norwegian observation sites: 1) stratified by wind speed, since calm (windy) conditions experience less (more) undercatch; and 2) by applying transfer functions, which convert measured precipitation to what would have been measured with high-quality equipment with less undercatch, before the forecast–observation comparison is performed. Results show that the wind-induced undercatch of solid precipitation has a substantial impact on verification results. Furthermore, applying transfer functions to adjust for wind-induced undercatch of solid precipitation gives a more realistic picture of true forecast capabilities. In particular, estimates of systematic forecast biases are improved, and to a lesser degree, verification scores like correlation, RMSE, ETS, and stable equitable error in probability space (SEEPS). However, uncertainties associated with applying transfer functions are substantial and need to be taken into account in the verification process. Precipitation forecast verification for liquid and solid precipitation should be done separately whenever possible.

KEYWORDS: Forecast verification/skill; Forecasting; Numerical weather prediction/forecasting; Model errors; Model evaluation/performance

1. Introduction

Accurate precipitation forecasts are important for weather warnings, as an integral part of the hydrological cycle, and for people's everyday life. However, all liquid-equivalent precipitation measurements, whether it is by precipitation gauges, ultrasonic measurements, snow pillows, radar, satellite, or by other equipment, operated by professionals or measured by citizen weather stations, have weaknesses (e.g., Sun et al. 2018; Tapiador et al. 2017; Nitu et al. 2018; Goble et al. 2020). This makes it difficult to quantify the true precipitation, and hence the true quality of weather forecasts. This study discusses how the wind-induced undercatch of solid precipitation by precipitation gauge measurements (e.g., Rasmussen et al. 2012; Nitu et al. 2018) impacts the verification process of weather forecasts.

Precipitation gauges are widely used for verification of short- and medium-range precipitation forecasts (e.g., Rodwell et al. 2010; Cookson-Hills et al. 2017; Frogner et al. 2019), as the basic input to many gridded precipitation products, and as ground truth for calibrating or verifying precipitation estimates from radar and satellite (e.g., Sun et al. 2018; Sebastianelli et al.

2013). The main advantages of precipitation gauges are that the observation data are easily accessible, measure precipitation directly and include long time series for a number of places. However, some disadvantages are spatial and temporal representativity and observational errors. The challenges of the high spatial and temporal variability of precipitation in the verification process have been addressed the last few decades with the emergence of high-resolution numerical weather prediction (NWP) models (e.g., Dorninger et al. 2018; Gilleland et al. 2010). However, the observational error part has received less attention. In cold regions and periods (e.g., high latitudes, elevations, and winter) the wind-induced undercatch of solid precipitation is the dominant source of observational errors. In the presence of wind the airflow around the opening of the precipitation gauge is disturbed and hydrometeors may not fall into the gauge (e.g., Rasmussen et al. 2012; Nitu et al. 2018; Colli et al. 2016). The magnitude of the error depends on, among others, the wind speed, the shape of the gauge, artificial and natural shielding of the gauge, precipitation phase, and shape and fall velocity of the hydrometeors (e.g., Kochendorfer et al. 2017; Rasmussen et al. 2012; Nitu et al. 2018). Wetting and evaporation errors, capping on

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

Corresponding author: Morten Køltzow, famo@met.no

DOI: 10.1175/WAF-D-20-0060.1

© 2020 American Meteorological Society

This article is licensed under a Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

top of the precipitation gauge and blowing snow are some other features that may also have an impact on the quality of the observations (Rasmussen et al. 2012).

The World Meteorological Organization (WMO) has facilitated several projects to improve precipitation gauge measurements (Goodison et al. 1998; Sevruk et al. 2009; Nitu et al. 2018). This includes the recently finished WMO Solid Precipitation Intercomparison Experiment (SPICE) where at a number of locations (with various climates) a set of different measurement equipment has been tested and compared. A reference measurement setup has been agreed on, the double-fence automated reference (DFAR), which minimizes the wind-induced undercatch and to which other measurement setups can be compared (e.g., Rasmussen et al. 2012; Wolff et al. 2013). Such reference setups are not feasible to be used widely as they require more space, are more expensive than simpler configurations, and may not follow local legal restrictions. However, based on controlled experiments, so-called transfer functions (TFs) have been developed to estimate what would have been measured by the DFAR, given measurements with a simpler configuration.

In many verification studies, the effect of wind-induced undercatch has been ignored (e.g., Frogner et al. 2019; Haiden et al. 2012), or used to explain results without an attempt toward quantification of the effect (e.g., Gowan et al. 2018; Wang et al. 2019). However, the undercatch has been taken into account in a quantitative way in some studies; Buisán et al. (2020) compared verification results obtained using WMO SPICE DFAR observations and more commonly used measurement equipment prone to wind-induced undercatch (both unadjusted and adjusted with TFs). They demonstrated that in the absence of DFAR or other reference observations, solid precipitation verification is complex and subject to uncertainty, but that applying TFs provide more reliable verification scores although the results then still contain the limitations and uncertainty associated with the TFs, Cookson-Hills et al. (2017) applied a TF when the daily mean air temperature was below 0°C, Bromwich et al. (2018) used gridded precipitation datasets that had undergone quality-control procedures to better handle the wind-induced undercatch, while Schirmer and Jamieson (2015) verified forecast winter and mountain precipitation against observations from ultrasonic snow depth measurements and snow pillows. In the latter study a systematic underestimation in forecast winter and mountain precipitation was found, as opposed to the forecast overestimation found in studies using precipitation gauges prone to undercatch. Similarly, a change from overestimation to underestimation of precipitation was found for four NWP systems when corrected precipitation measurements were used in an Arctic region (Køltzow et al. 2019). These studies show that not only is the magnitude of the precipitation errors uncertain in cold regions, but sometimes also the sign of the bias.

The aim of this work is 1) to study the impact of wind-induced undercatch of observed solid precipitation on forecast verification, 2) contribute to the development of best practices for using precipitation gauges in the verification and 3) assess the true quality of solid precipitation forecasts in Norway from the high-resolution version of the global ECMWF Integrated Forecasting System (IFS-HRES), and the control run from the operational Nordic regional convection-permitting system,

MetCoOp Ensemble Prediction System (MEPS). This study extends previous work by putting additional effort in estimating the uncertainty in the verification introduced by wind-induced undercatch, by extending the time period, domain, and the verification metrics studied. This work also uses the outcomes of the recently finished WMO SPICE in new applications.

The observations and NWP systems used in this study are described in section 2, while the applied transfer functions are described and discussed in section 3. In section 4, precipitation observations, adjusted observations, and forecasted precipitation are compared for one of the WMO SPICE supersites, Haukeliseter in Norway. In section 5, hourly precipitation forecasts in Norway are verified stratified by wind speed and by making adjustments of observations by transfer functions. In section 6, a method to adapt the adjustments on hourly precipitation to daily precipitation observations are suggested and tested to increase the data sample. A discussion on where wind-induced undercatch of solid precipitation is important is done in section 7. Finally, a summary and conclusions are made in section 8.

2. Observation and model data

a. Observation data

In this study we use quality controlled observations from the Norwegian Meteorological Institute (eklima.met.no) of hourly and daily precipitation, instantaneous 2-m air temperature (T2m) and 10-m wind speed (WS10). The quality control system consists of both automatic and human quality control routines to flag or remove suspicious or erroneous observations (Kielland 2005). However, erroneous precipitation observations due to wind-induced undercatch are not removed or flagged in the quality control process. All hourly precipitation observations are measured with Single-Alter shielded Geonor (SA Geonor) precipitation gauges while the equipment for daily precipitation varies. In total, there are 76 stations with hourly precipitation (SA Geonor), T2m, and WS10 measurements. In addition there are 206 stations where daily precipitation is measured by Norwegian/Swedish gauges for which 179 have a Norwegian version of the Nipher windshield and 27 are unshielded. It should be noted that the automated SA Geonor measurements are built to maintain homogeneity of the precipitation records in Norway when replacing the manual gauges and have been shown to be very similar in what they measure (Bakkehøi et al. 1985; Jacobi et al. 2019; M. Wolff 2020, personal communication). For the Norwegian observations the wind-induced undercatch is the dominant error source, but also wetting and evaporation errors, capping on top of the precipitation gauge and blowing snow may play a substantial role (Førland et al. 1996). The spatial distribution of observation sites is shown in Fig. 1.

The Haukeliseter observation site was a supersite in the recently finished WMO SPICE where targeted studies were done on wind-induced undercatchment of solid precipitation. Haukeliseter is located in the Norwegian mountains (59.8118°N, 7.2143°E, 991 m MSL, Fig. 1) and described in detail in Wolff et al. (2013, 2015). In this study we make use of the Haukeliseter collocated Geonor precipitation observations with SA and DFAR shielding, WS10 and T2m measurements. It should be noticed that also the DFAR Geonor measurements experience wind-induced

FIG. 1. (a) The MEPS integration domain shown with model topography, location of the Haukeliseter observation site (used in section 4) is given as a blue circle. (b) The 76 Norwegian observation sites measuring hourly precipitation, used in section 5, (c) 123 daily precipitation with long records to calculate climatology, and (d) all 282 daily precipitation observation sites used in section 6.

undercatch of solid precipitation, estimated to be 5%–6% (Nitu et al. 2018), but considerably less than by other measurement equipment. Kochendorfer et al. (2017) showed that averaged over the eight WMO SPICE sites the wind-induced undercatch was 24% for SA shieldings and 34% for completely unshielded precipitation gauges.

b. Model data

The two NWP systems verified are the operational versions of 1) the high-resolution version of the global ECMWF Integrated Forecasting System (IFS-HRES; Buizza et al. 2017), and 2) the control run from the Nordic regional convection-permitting system, MetCoOp Ensemble Prediction System (MEPS; Müller et al. 2017; Bengtsson et al. 2017; Frogner et al. 2019).

IFS-HRES employs a 4D-Var assimilation scheme for upper air fields, and optimal interpolation for near-surface variables. The system has approximately 9-km grid spacing and 137 vertical layers. Cloud and large-scale precipitation processes are described by prognostic equations for cloud liquid water, cloud ice, rain, snow and a gridbox fractional cloud cover. The basic design of the scheme with respect to the prognostic cloud fraction and sources/sinks of all cloud variables due to the major microphysical generation and destruction processes follows Tiedtke (1993). However, liquid and ice water contents are independent, allowing a more physically realistic representation of supercooled liquid water and mixed-phase cloud than the original scheme. Rain and snow precipitate with a determined terminal fall speed and can be advected by the three-dimensional wind. A multidimensional implicit solver is used for the numerical solution of the cloud and precipitation prognostic equations (Forbes et al. 2011).

MEPS is an operational regional convection-permitting ensemble prediction system for Norway, Sweden and Finland, based on the HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed (HARMONIE)–Applications of Research to Operations at Mesoscale (AROME). The system employs a 3D-Var assimilation scheme for upper air and optimal interpolation for surface assimilation. The system has 2.5-km grid spacing and 65 vertical layers. Lateral boundary conditions are taken from 6-h old IFS-HRES forecasts, and the regional domain is shown in Fig. 1. In MEPS, the cloud

microphysics is handled by the ICE3 scheme (Cohard and Pinty 2000a,b), with some additions known as the OCND2 option (Müller et al. 2017). ICE3 is a single-moment scheme with explicit calculations for the mass of cloud water, rain, cloud ice, snow, and graupel. Shallow convection is parameterized by the EDMFm scheme (Bengtsson et al. 2017).

The model data are paired to the observation data by applying nearest neighbor interpolation. The choice of interpolation method may have an impact on the verification results, but Kjøltzow et al. (2019) showed that for a winter period in Norway the difference in root-mean-square error applying near neighbor and bilinear interpolation is less than 2% for precipitation. The two NWP systems do not use precipitation measurements in their assimilation process and the verification is therefore done with independent observation data.

The main verification period in this study is the 3-yr-long period from January 2017 to December 2019 in which both model systems have gone through upgrades. In the verification of operational systems this will always be the case unless the systems are verified for shorter periods. To make the statistics more robust longer periods are required and we do believe the model upgrades have minor impact on how undercatch in precipitation observations influences the verification. Also, in the period chosen, there were only minor changes in the microphysics scheme in both the MEPS and the IFS. In the section on verification of daily precipitation we present time series of the evolution of forecast errors during the period.

3. Transfer functions

Transfer functions aim to transfer measured precipitation exposed to wind-induced undercatch to what would have been measured with equipment less prone to wind-induced undercatch, e.g., transfer SA Geonor measurements to what would have been measured by DFAR Geonor. Many different TFs are reported in the literature (e.g., Goodison et al. 1998; Sevruk et al. 2009; Smith 2009; Wolff et al. 2015; Kochendorfer et al. 2017; Buisán et al. 2017), and they may vary in input parameters, sites (characterized by different climatologies) used for development, wind sheltering, measurement equipment and accumulation periods, to mention some aspects. We apply

TABLE 1. Universal coefficients a , b , and c for 10-m height wind measurements, for Eq. (1) based on all WMO SPICE sites (single-Alter shielded and unshielded) and a cutoff wind speed (insufficient data for deriving CE above this limit) from Table 2 in Kochendorfer et al. (2017). Same coefficients and wind speed cutoff based on individual WMO SPICE sites for Single-Alter shieldings obtained from J. Kochendorfer (2020, personal communication).

TF coefficients	a	b	c	Max WS
Universal SA	0.0281	1.628	0.837	9 m s^{-1}
Universal unshielded	0.0623	0.776	0.431	9 m s^{-1}
CARE SA	0.025 44	2.48	0.4059	$6, 5 \text{ m s}^{-1}$
Haukeliseter SA	0.050 68	0.9753	0.3435	12 m s^{-1}
Sodankyla SA	0.040 05	0.061 13	2.85×10^{-10}	$3, 5 \text{ m s}^{-1}$
Caribou Creek SA	0.020 34	29.32	0.7738	5 m s^{-1}
Weissfluhjoch SA	0.0289	0.1051	3.0×10^{-5}	9 m s^{-1}
Formigal SA	0.118	0.2376	1.41×10^{-9}	$6, 5 \text{ m s}^{-1}$
Marshall SA	0.040 09	0.5159	0.2679	6 m s^{-1}
Bratt's Lake SA	0.035 48	1.955	0.8036	7 m s^{-1}

the TF developed during the recently finished WMO SPICE described in Kochendorfer et al. (2017):

$$\text{CE} = \exp\{-aU[1 - \arctan(bT_{\text{air}} + c)]\}. \quad (1)$$

The TF [Eq. (1)] states that the catch efficiency (CE) of observed precipitation is a (negative) exponential function of the wind speed U with a coefficient in the exponent that varies with the arcus tangens of air temperature T_{air} . Universal values of the three coefficients a , b , and c are determined based on observation data from 30-min precipitation events from eight globally distributed WMO SPICE sites for unshielded and SA shielded measurements, for wind measurements at 10 m and gauge height (Kochendorfer et al. 2017).

A substantial case-to-case variability in the observed CE for the same wind speed and temperature exists (e.g., Fig. 4 in Wolff et al. 2015), implying that additional factors must be of importance, e.g., precipitation characteristics and other local conditions. The site dependency is seen in Table 1 with differing a , b , and c coefficients when these are estimated from the eight individual WMO SPICE sites (coefficients provided by J. Kochendorfer 2020, personal communication). Plotting the CEs with the different site coefficients (Fig. 2) underlines the large uncertainty in the adjustment of the observed precipitation. For example, with an air temperature of -2.5°C and 3 m s^{-1} wind speed the CE varies from 0.88 to 0.52 depending on the choice of coefficients (sites). In addition, for all coefficient sets there is a maximum wind speed (max WS, right column of Table 1), beyond which the CE is kept constant, since above this threshold the observational data are scarce and the coefficients are difficult to be estimated. In Fig. 2 the dashed lines show the extrapolated CE without such a cutoff beyond max WS; for these wind speeds the CE should be used very cautiously. A suggestion on how the observation uncertainty can be taken into account in the verification process is made and tested in section 5.

4. Observed, adjusted, and forecasted precipitation at Haukeliseter

At Haukeliseter precipitation is observed with both the commonly applied SA Geonor (prone to wind-induced undercatch) and the WMO reference equipment DFAR (less prone

to wind-induced undercatch). Accumulated precipitation from DFAR and the unadjusted SA Geonor measurements, MEPS and IFS-HRES forecasts (lead time from +6 to +30 h) at Haukeliseter from 15 December 2017 to 31 March 2018 is shown in Fig. 3. In addition, adjusted precipitation accumulated after applying TFs [Eq. (1) with different sets of coefficients from Table 1] on the SA Geonor measurements are shown. Notice that this period is not included in the period used to determine the TF coefficients for Haukeliseter (Kochendorfer et al. 2017) and can therefore be used for independent verification. The SA Geonor at Haukeliseter measures 57% (accumulated 225 mm) of the precipitation amount in the DFAR Geonor (393 mm), illustrating the problem of wind-induced undercatch of solid precipitation. The undercatch also results in less accurate observations in terms of variability, because the amount of undercatch depends on wind speed, demonstrated by a linear correlation between SA and DFAR Geonor measurements of 0.88 (e.g., SA Geonor measurements are not always able to identify which events had the most precipitation).

By applying the TF with universal coefficients, the accumulated precipitation is 306 mm, which is 36% more than measured by SA Geonor, but only 78% of what was measured by the DFAR Geonor. Applying the individual coefficients give a large span in estimates of accumulated precipitation, from 250 mm (Sodankyla coefficients) to 444 mm (Formigal coefficients), i.e., a span from 64% to 113% of DFAR Geonor. The correlation between TFs estimated precipitation and DFAR Geonor measurements varies from 0.88 to 0.93, i.e., a small improvement over the SA/DFAR correlation. The best match after applying TFs and DFAR measurements is, as expected, found when using the coefficients from Haukeliseter itself, with estimated accumulated precipitation to be 372 mm (95% of DFAR Geonor) and 0.93 in correlation. In comparison, the more complex transfer function developed by Wolff et al. (2015) for Haukeliseter gives a similar result (2% overestimation and correlation of 0.93) for the investigated period.

The validity of the different sets of coefficients for the TF are limited by the max WS10 in Table 1. For wind speeds higher than these thresholds there was insufficient data to estimate the CE from the WMO SPICE sites. In Fig. 3, we have kept a constant CE, determined by the max WS10, beyond these

FIG. 2. Parameterized catch efficiency (CE) for a single alter installation as a function of 10-m wind speed (x axis) for three different temperatures, (left) 0°C , (center) -2.5°C , and (right) -5°C , calculated with Eq. (1) with coefficients a , b , and c estimated from the different WMO SPICE sites (Table 1): CARE (red), Haukelisetter (green), Sodankyla (blue), Caribou Creek (cyan), Weissfluhjoch (pink), Formigall (yellow), Marshall (gray), and Bratt's Lake (black). The CE based on universal coefficients estimated by pooling all sites is the dotted black line.

thresholds. However, since Haukelisetter is a windy site, it can be argued that this approach gives a too high CE (resulting in a too mild adjustment) for coefficient sets with a low maximum wind speed threshold. We therefore also applied a threshold on 12 m s^{-1} for all coefficient sets, i.e., applied the dashed lines in Fig. 2 up to 12 m s^{-1} . With the exception of the Haukelisetter-based coefficients (already a threshold on 12 m s^{-1}) this gave a relatively modest increase in estimated precipitation in the range between 3% and 27% (4% for the universal coefficients)

at Haukelisetter; however, the main features of Fig. 3 are kept (not shown).

At Haukelisetter, MEPS (479 mm) and IFS-HRES (521 mm) have a substantial overestimation of SA Geonor measured precipitation, with 213% and 232% of the observed precipitation, respectively. They also overestimate, but to a much smaller degree, compared to DFAR Geonor measurements with 122% (MEPS) and 133% (IFS-HRES). Also in terms of correlation the numbers are improved, 0.74 (both models)

FIG. 3. Accumulated precipitation at Haukelisetter (15 Dec 2017–31 Mar 2018) from hourly observations with SA Geonor (black dotted) and DFAR Geonor (black solid) both marked with black circles. Estimated precipitation with TFs applied on SA Geonor observations are given in dashed lines applying the universal coefficients (black marked with asterisks) and from the individual WMO SPICE sites: CARE (red), Haukelisetter (green), Sodankyla (blue), Caribou Creek (cyan), Weissfluhjoch (pink), Formigall (yellow), Marshall (gray), and Bratt's Lake (black). In addition are the accumulation of the forecasted precipitation shown for MEPS (solid blue marked with filled circles) and IFS-HRES (solid red marked with filled circles).

FIG. 4. Liquid (red) and solid (blue) precipitation forecasts divided by observed precipitation (SA Geonor measurements) for (a) MEPS and (b) IFS-HRES, stratified by observed wind speed. Shaded areas are 95% confidence intervals calculated by bootstrapping.

against DFAR Geonor, while the correlation against SA Geonor is 0.65 (IFS-HRES) and 0.66 (MEPS). However, the correlation between forecasts and TFs are only slightly higher compared to SA Geonor (increase between 0.01 and 0.04, with the Haukelisetra coefficients giving the largest increase). This indicates that applying TFs improves estimates of forecast bias (i.e., the correspondence between the mean forecast and mean observation) but to a more limited degree the correlation.

5. Verification of hourly precipitation forecasts

a. Stratification on wind speed

By stratifying the precipitation verification with respect to wind speed, more weight can be put on the results obtained during calm conditions and less during windy conditions exposed to wind-induced undercatch. The additive forecast bias of hourly accumulated liquid ($T2m > +2^{\circ}\text{C}$) and solid precipitation ($< -2^{\circ}\text{C}$) from MEPS and IFS-HRES when compared to uncorrected SA Geonor measurements, stratified by wind speed are shown in Figs. 4a and 4b. The observations used are from the 76 sites observing hourly precipitation, T2m and WS10 described in section 2a. For MEPS, the liquid precipitation bias is insensitive to wind speed, while the bias for solid precipitation increases with wind speed. We believe that (at least part of) this positive bias is artificial and due to the measurements wind-induced undercatch. A similar behavior is also seen for solid precipitation from IFS-HRES. However, IFS-HRES also has an increase in precipitation bias with wind speed for liquid precipitation that cannot be explained by observational issues.

In addition to continuous metrics, skill scores based on categorical forecasts are commonly used for precipitation verification. The Equitable Threat Score (ETS) is shown for liquid and solid precipitation in calm ($< 2.5 \text{ m s}^{-1}$) and windy conditions ($> 4 \text{ m s}^{-1}$) for a variety of precipitation thresholds in Figs. 5a and 5b. The ETS is a measure evaluated from the threat score = hits/(hits + false alarms + misses), which then is modified for hits obtained by a random forecast [see Wilks (2011, chapter 8) or Jolliffe and Stephenson (2012) for definition and more details]. The ETS therefore measures the fraction of observed and/or forecast events that were correctly predicted, relative to chance and is one of the most widely used

scores in the verification of precipitation forecasts. ETS is a positively oriented score, i.e., a larger score indicates a better forecast, with a perfect score being 1 and a forecast with no skill achieving the value 0. During calm conditions the ETS is quite similar for liquid and solid precipitation, but during windy conditions there is a large difference depending on the precipitation phase, apparently showing low skill for solid precipitation, which we argue is (at least partially) due to wind-induced observation errors.

The verification of precipitation stratified by wind speed clearly indicates that when solid precipitation is present the results may be misleading if the wind-induced undercatch is not considered. However, we do not suggest applying such an approach operationally: a weakness of applying the wind stratification approach to assess the forecast quality is that many precipitation events happen during windy conditions and the trustful part of the verification is then based on a much smaller data sample. Such an approach would also put more weight on wind-sheltered sites and generally include fewer cases from windy areas such as coastal and mountainous regions. In addition, only including calm conditions may provide a skewed sample in terms of types of precipitation systems.

b. Applying transfer functions

In this section, we use the TF developed by Kochendorfer et al. (2017) and described in section 3, to adjust the 76 hourly SA Geonor observations (described in section 2a and also used in section 5a) before used in verification. The large spread in CE depending on the choice of coefficient set (Table 1 and Figs. 2 and 3) underlines the high uncertainty associated with the choice of transfer function. Ideally, the most appropriate TF and coefficients should be found for each observation site (e.g., sites similar to Haukelisetra should use coefficients from Haukelisetra etc). However, there is no straightforward and objective way to do this. Therefore, we present for all Norwegian observation sites that observe hourly precipitation, WS10, and T2m, precipitation biases applying the universal coefficients, accompanied by the spread resulting from applying all sets of coefficients in Table 1. The latter is an attempt to quantify the uncertainty in the choice of TFs. For this purpose, we argue that extrapolation beyond the maximum wind speed thresholds in Table 1 (dashed curves in Fig. 2) gives a better

FIG. 5. Equitable threat score as function of precipitation thresholds for (a) calm conditions ($<2.5 \text{ m s}^{-1}$) and (b) for windy conditions ($>4.0 \text{ m s}^{-1}$). Liquid precipitation is shown in red and solid precipitation is shown in blue. MEPS is shown in solid lines and IFS-HRES is shown in dashed lines. Shaded areas are 95% confidence intervals calculated by bootstrapping.

representation of the uncertainty than applying an unrealistic cutoff threshold (constant CE beyond WSmax).

Liquid and solid precipitation biases for all Norwegian sites for MEPS and IFS-HRES for a 3-yr-long period (from January 2017 to December 2019) are shown in Fig. 6. An underestimation of solid precipitation is seen at coastal stations (marked with blue), which is even more pronounced after applying TFs. Opposite to this, a general overestimation in solid precipitation

is seen in the mountains (stations situated at more than 700 m MSL marked with red) before corrections, but reduced to a neutral or less pronounced positive bias after applying TFs. The uncertainty (spread between different choices of TF coefficients) is large for the mountain sites, which is expected due to higher wind speeds. The average MEPS bias is positive for both liquid (forecast/measured = 1.05) and solid (1.15) precipitation before corrections. However, applying the universal

FIG. 6. Forecasted precipitation from (a) MEPS and (b) IFS-HRES divided by observed precipitation per station for the period 1 Jan 2017–31 Dec 2019. Liquid ($>+2^{\circ}\text{C}$) precipitation compared to SA Geonor as triangles, solid ($<-2^{\circ}\text{C}$) precipitation compared to SA Geonor as asterisks, solid ($<-2^{\circ}\text{C}$) precipitation compared to SA Geonor adjusted with universal coefficients as green dots, and adjusted with all coefficients from different WMO sites as boxplots. Coast stations are plotted in blue, mountain stations in red, and stations are plotted from left to right from the highest underestimation toward the highest overestimation, as for the boxplot median values.

coefficients give an underestimation of solid precipitation (forecast/adjusted observations = 0.87). To estimate the uncertainty in the average bias we assume that all coefficient sets in Table 1 are equally likely to be representative for a site. Then the total bias is calculated by combining random draws of which TF coefficients to be used at each station, e.g., a random draw of coefficients on site 1 is combined with a random draw of coefficients on site 2 and so on until the last site and an average bias can be calculated. This process is repeated 1000 times and the spread between 1000 calculated average biases is used as a measure of the uncertainty, ranging from 0.71 to 0.88, with median 0.82 (forecast/adjusted observation).

Many of the same features found for MEPS are also valid for IFS-HRES. For IFS-HRES, an overestimation of liquid precipitation (forecasted/measured = 1.26) is present, while the solid precipitation bias change from an overestimation (1.21) to an underestimation (0.92) after adjusting the observations with an uncertainty range from 0.75 to 0.93 and median 0.86. After adjustments of the observations it is evident that both model systems overestimate liquid precipitation, but underestimate the true solid precipitation. It is important to understand the reasons for this model behavior, but it requires further investigations and is out of the scope of this paper. Also verification metrics like normalized standard deviation of the error (NSDE), root-mean-square error (RMSE), correlation, and ETS were calculated with and without applying TFs for solid precipitation. However, only small changes in these metrics were found when applying raw and adjusted observations (not shown).

The large local spatial variability in precipitation biases seen for both model systems, e.g., between the three mountain stations Klevavatn (underestimation), Midtstova (overestimation), and Finse (neutral/underestimation), situated less than 20 km apart, indicate that local effects and the observation site sampling play an important role for the verification results. Moreover, when averaging over all sites above they are all equally weighted, i.e., areas with a denser observation network are given more weight. It would therefore be beneficial to also verify precipitation against daily accumulated precipitation that then would include many more observation sites and make the statistics more robust and possibly more geographical representative.

6. Verification of daily precipitation forecasts

Precipitation observations are becoming more and more automated and thereby also available for a range of accumulation periods (e.g., as short as an hour). However, a large number of precipitation measurements are still done manually, e.g., daily accumulated precipitation at Norwegian climate stations. To be able to include these observations in the verification process will make the results more robust (increase the amount of observations) and make it possible to create long-term verification time series.

To apply TFs to adjust daily observations for wind-induced undercatch requires some additional considerations; 1) information on the time of day when the precipitation mainly accumulated on a given day would be needed to use the best possible WS10 and T2m to calculate the CE. To overcome this

issue we follow the approach of Vormoor and Skaugen (2013) and distribute the observed daily precipitation in time by using the temporal patterns of the forecasts. If the forecast has no precipitation, the observed precipitation is distributed equally over the 24 h; 2) observations of WS10 and T2m may not be available. To overcome this we use WS10 and T2m from the forecasts, which was shown to be a useful approach by Masuda et al. (2019) using reanalysis; 3) a mixture of observation equipment/sheltering exists and hence different TFs are ideally needed for different stations. Norwegian stations observing daily precipitation deploy SA Geonor, Norwegian/Swedish gauges with a Nipher wind shield and Norwegian unshielded precipitation gauges. However, the SA Geonor measurements and the manual gauges used in Norway have been shown to be very similar in what they measure (Bakkehoi et al. 1985; Jacobi et al. 2019; M. Wolff 2020, personal communication). Therefore, we apply Eq. (1) with the universal SA Geonor coefficients (Table 1) on all sites with windshield and the coefficients (Table 1) for the unshielded gauges for the unshielded observations.

To test steps 1 and 2 in the described approach, we compare the adjusted daily precipitation at observation sites where these adjustments can be done both based on WS10, T2m, and precipitation timing from forecasts and observations. For the winter months December, January and February 2017–19, applying observations-based adjustments (observed T2m, WS10, and timing of precipitation) give an increase in total precipitation of 17%, while applying MEPS and IFS-HRES adjustments (forecasted T2m, WS10, and timing of precipitation) give 22% and 20% increase, respectively. A general positive wind speed bias in MEPS leads to an overestimation of the wind-induced undercatch. On the other hand, the correlation between adjustments calculated by observed values and MEPS is 0.87, while 0.79 for IFS-HRES. We attribute the lower correlation for IFS-HRES to less spatial detail and an underestimation of wind speed in some regions (e.g., in the mountains). This comparison is done mainly on sites equipped with SA Geonor. In the rest of this section we calculate the CE based on observations where observations of WS10 and T2m exist and based on MEPS forecasts for the other sites.

The daily bias for different seasons, before and after measurement adjustments are shown in Fig. 7. A substantial change in the biases is present during winter (DJF). For MEPS the bias changes, on average, from an almost neutral bias to an underestimation of almost 0.8 mm day^{-1} ($\sim 20\%$). A clear change in bias for MEPS is also seen during spring (MAM) and autumn (SON). Without any adjustment of the observations IFS-HRES show a consistent overestimation of $0.4\text{--}0.8 \text{ mm day}^{-1}$ ($10\text{--}25\%$) during all seasons. However, in winter this is changed to a negative bias ($\sim 0.2 \text{ mm day}^{-1}$, $\sim 5\%$) after measurement adjustments. Also in spring and autumn the measurement adjustments clearly reduce the IFS-HRES overestimation. The interpretation of these results should take into account the added uncertainty in the adjustment process of daily accumulated precipitation described above. For example, since the adjustment with T2m and WS10 from MEPS had a tendency to adjust too much, the impact seen on the verification results can be looked at as an upper limit of expected impact. However, these results are qualitatively similar to what was found for

FIG. 7. Daily precipitation and forecast biases averaged over 282 Norwegian stations for December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON). Raw measurements are shown in gray and adjusted measurements with TF [Eq. (1) and universal coefficients] are shown in black. Biases are shown for MEPS (red) and IFS-HRES (blue) against raw measurements (vs RAW) and against adjusted measurements (vs ADJ).

hourly observations and adjustments. Although a large impact is found in systematic biases when adjusting the observations, only minor impact is again seen on metrics like correlation, NSDE, RMSE, and ETS (not shown).

In the following we apply the “stable equitable error in probability space” metric, hereafter named SEEPS (Rodwell et al. 2010). SEEPS divides the precipitation into “dry,” “light precipitation,” and “heavy precipitation” based on the local climatology. In our use we have chosen “dry” to be when measured daily accumulated precipitation is less than 0.5 mm and the 1/3 highest precipitation amounts (excluding dry events) are defined as “heavy.” This means that “heavy precipitation” events are not necessarily high-impact or extreme precipitation events. The SEEPS score is chosen because it is one of the headline verification scores at ECMWF and MET-Norway and encourages refinement, discourages hedging, can be combined for different climatic regions, is less sensitive to sampling uncertainty, is equitable, and measures the error in probability space and is therefore less sensitive to observation and representativeness errors (Rodwell et al. 2010). The SEEPS score can be decomposed, and different aspects of the precipitation forecast error can be quantified (Haiden et al. 2012). A disadvantage is that SEEPS requires knowledge about local precipitation climatology, i.e., it is necessary to have long time series of precipitation observations available, which should be adjusted for wind-induced observation errors. However, this is not always possible (i.e., WS10 and T2m are not available for all sites). Instead we take advantage of the fact that SEEPS measures the error in probability space and therefore we adjust the forecasts to what should be expected to be measured ($CE \times \text{forecasted precipitation}$) in windy conditions. The adjusted forecasts are then compared with the (unadjusted) observations and the SEEPS score can be calculated.

For both NWP systems SEEPS scores are slightly worse after adjustments are made, and on average 3%–4% worse for

the winter months (Fig. 8). A simple test by only calculating SEEPS for sites with elevations higher than 300 m MSL (e.g., increasing the fraction of solid precipitation in the dataset) shows slightly higher differences (not shown). Furthermore, the change in SEEPS score due to adjustments is less than the difference between the two NWP systems verified. MEPS and IFS-HRES have quite similar performance in 2019, but in certain periods (e.g., autumn) in 2018 and 2017 IFS-HRES scores better. The decomposition of the SEEPS score (Fig. 9) shows that these differences originate from cases where forecasted MEPS precipitation is less intense than the observed precipitation. An annual cycle with larger errors in summer than winter for MEPS indicates that part of this is related to double-penalty issues, which is more pronounced with the higher horizontal resolution in MEPS. IFS-HRES has a general overestimation of the frequency of light precipitation, and heavy precipitation during summer, but a weaker annual cycle of the error. However, MEPS shows a better frequency bias for all the three precipitation categories and lower errors during dry conditions. Hence, the SEEPS decomposition clearly identifies different strengths and weaknesses in the two NWP systems. However, it is not within the scope of this paper to discuss these intramodel differences in detail.

The SEEPS diagnostics also show how the adjustment of precipitation impacts the total score. The adjustments lead to a minor improvement, in both systems, during winter in bias frequency for dry conditions. In addition, the winter bias frequency of heavy precipitation events in IFS-HRES is improved by reducing the overestimation. For MEPS, a winter overestimation of the heavy precipitation events is turned into an underestimation of the same events. For the error components the adjustments give a higher contribution from the category “forecasted light and observed heavy,” but a lower error contribution from “forecasted light and observed dry.”

FIG. 8. Time series of daily SEEPS skill score (1-SEEPS) for MEPS (red) and IFS-HRES (blue) compared with unadjusted (dashed lines) and adjusted (solid lines) precipitation measurements from 123 stations. A 3-month running mean is applied when plotting.

7. Which areas are exposed to wind-induced undercatch of solid precipitation?

The impact of wind-induced observation errors on forecast verification, as calculated by TFs, depend on the amount of solid precipitation, wind speed and temperature. The impact therefore varies in time and with season and region. In Scandinavia, the solid precipitation is a substantial part of the total precipitation during winter, as shown by forecasted precipitation from MEPS for December 2018–February 2019 (Figs. 10a–c). The fraction of solid precipitation increases inland and northward, as expected, and liquid precipitation is mostly limited to coastal/ocean regions or inland areas south of 60°N. The areas with the highest precipitation amounts, in the Norwegian mountains, are dominated by solid precipitation (>600 mm). The spatial distribution of the wind-induced undercatch is estimated in Fig. 10d, i.e., the fraction of the total precipitation that we would expect to observe with SA Geonor measurements given that the model precipitation equals what would be measured with DFAR (hereafter named model-estimate of the SA Geonor fraction). The model-based CE is calculated by applying MEPS forecasts of WS10 and T2m in Eq. (1) with the universal coefficients from Table 1. Then the model-estimate of the SA Geonor fraction can be estimated as the sum of hourly model-based CE $\times$ forecasted precipitation divided by the sum of the forecasted precipitation (assumed to equal DFAR measured).

The model-estimated SA Geonor fraction varies with temperature (e.g., higher over the sea, but reduced northward), and wind speed (e.g., lower in the mountains, but higher for lower elevated inland regions), and shows substantial geographical variability. While estimated SA Geonor fraction in Finland is relatively homogeneous (~ 0.7 – 0.8 , decreasing northward), the spatial pattern in Norway is more complex, but follows the Norwegian topography and coastline with a minimum estimated SA Geonor fraction in the mountains (<0.5),

but higher fractions in more wind sheltered valleys and lower elevation inland regions (>0.80). It should also be noticed that the SA Geonor fraction along the Norwegian coast decreases from the south (>0.9) to north (<0.6). In summary, the model-estimated SA Geonor fraction is less than 0.9 for most of the domain with the exception of the south east part covering Denmark and the southwest coast of Norway. Wind-induced observation errors can therefore possibly impact the verification results for most parts of the MEPS domain.

8. Conclusions

The aim of this study is to investigate how the wind-induced undercatch of solid precipitation measurements impacts weather forecast verification. Precipitation gauges are a key component for verification of weather forecasts, as input to gridded precipitation products, and as ground truth for precipitation estimates from radar and satellite products. Even with moderate wind speeds the wind-induced undercatch of solid precipitation by precipitation gauges is substantial (e.g., Rasmussen et al. 2012; Nitu et al. 2018) and makes it difficult to quantify the true precipitation, and hence the true quality of weather forecasts.

This study takes transfer function coefficients for calculating the gauge catch efficiency (CE) that were developed at eight sites at various parts of the world during the WMO SPICE project and applies them to 3-yr gauge measurements collected in Norway. Typical operational SA measurements before and after the CE adjustments and more reliable reference DFIR Geonor measurements were compared to the precipitation forecasts from a global model (high-resolution ECMWF IFS-HRES) and high-resolution regional model [Applications of Research to Operations at Mesoscale (MEPS)]. In addition, the difference between liquid precipitation (less prone to undercatch) and solid precipitation verification has been studied. The main findings can be summarized as follows;

FIG. 9. Time series of bias frequencies (diagonal) and error contributions (larger number means larger error) in the calculation of SEEPS. MEPS (red) and IFS-HRES (blue) compared with unadjusted (dashed lines) and adjusted (solid lines) precipitation at 123 stations. A 1-month running mean is applied when plotting.

- The wind-induced undercatch of solid precipitation introduces observational errors that can have substantial impact on NWP verification results. Verification at the Haukeliseter supersite shows that more reliable observations, i.e., those from DFAR Geonor, result in a substantial improvement in forecast errors compared to the standard SA Geonor equipment.
- Applying TFs to adjust for wind-induced undercatch of solid precipitation provides useful information and gives a more realistic picture of the true forecast capabilities. In particular,

estimates of systematic forecast biases are improved and, to a lesser degree, verification metrics like correlation, NSDE, RMSE, and ETS.

- It is possible to make use of WS10, T2m, and the temporal distribution of precipitation from forecasts to adjust daily observations of precipitation, but this adds an extra layer of uncertainty and the interpretation of the verification results should be made with caution. An alternative, and a more consistent way, is to evaluate forecasted precipitation reduced by the model-estimated precipitation undercatch with raw measurements.

FIG. 10. Accumulated (a) liquid and (b) solid precipitation for 1 Dec 2018–28 Feb 2019 from MEPS. (c) The fraction of solid precipitation, solid/(liquid + solid), is shown. (d) The expected fraction of total precipitation that would be observed with SA Geonor measurements based on MEPS precipitation, WS10, and T2m.

- Applying TFs is associated with uncertainty, which should be taken into account in the verification process and the interpretation should be done with caution. Further work on reducing the uncertainty in the TFs is needed. Open issues are how to find the most appropriate TF for a given observational site, and how best to include the (remaining) observational uncertainty in the verification process.
- Due to the uncertainty associated with applying TFs it is recommended to complement the precipitation verification with different types of precipitation measurements independent of precipitation gauges (e.g., snow pillows) when available.
- The interpretation of precipitation verification is easier if the evaluation for liquid and solid precipitation is done separately.
- Model skill is dependent on types of precipitation (liquid versus solid), location (mountain versus coastal), seasons, and precipitation rate (dry, light, and heavy). For the particular region, period, and NWP systems in this study some specific findings are: 1) an impression of overestimation of winter precipitation was changed to an underestimation after adjustments of observations; 2) the adjustment of observations increases the underestimation of coastal precipitation and reduces the overestimation of precipitation in the mountains when compared to raw observations (most pronounced in MEPS); 3) due to the limitations of existing TFs it remains to provide reliable estimates of verification metrics like correlation, RMSE, ETS, and SEEPS. Only verifying solid precipitation in calm conditions reduces

the problem of wind-induced undercatch but has other disadvantages.

Existing TFs have been evaluated recently (Pierre et al. 2019; Smith et al. 2019) and the associated uncertainty has been documented. The performance varies substantially by site, but in general the TFs underadjust at windier sites (also seen at Haukelisetter in this study), while a net overadjustment is found for less windy sites (Smith et al. 2019). Their results also suggest that TFs are useful but should be applied with caution similar to what we found in our study. Furthermore, the results suggest that local meteorology and conditions (e.g., natural shielding through vegetation) are important for understanding the undercatch and that assigning appropriate TFs to all observation sites is therefore a difficult task (Pierre et al. 2019).

To further improve our understanding of how wind-induced undercatch affects verification results there are a number of possibilities. Here we identify at least three ways forward; 1) Improve TFs and how they are used in the verification process; 2) Perform precipitation verification for global models on all WMO SPICE sites against DFAR, SA and unshielded gauges to better understand the impact (e.g., Buisán et al. 2020); 3) Perform complementary forecast verification against a variety of solid precipitation observation types, when available (e.g., Schirmer and Jamieson 2015).

Acknowledgments. We thank Mareille Wolff (MET Norway) and John Kochendorfer for important input and discussions on precipitation observation errors and John Kochendorfer for providing the transfer function coefficients based on the individual WMO SPICE sites given in Table 1. The authors would also like to thank the three anonymous reviewers, whose comments, questions, and suggestions helped to improve this article. The work described in this paper has received funding from the European Union's Horizon 2020 Research and Innovation programme through Grant 727862 APPLICATE. The content is the sole responsibility of the author(s) and it does not represent the opinion of the European Commission, and the Commission is not responsible for any use that might be made of information contained. This study was also supported by the Norwegian Research Council Project 280573 "Advanced models and weather prediction in the Arctic: Enhanced capacity from observations and polar process representations (ALERTNESS)." This work is a contribution to the Year of Polar Prediction (YOPP), a flagship activity of the Polar Prediction Project (PPP), initiated by the World Weather Research Programme (WWRP) of the World Meteorological Organization (WMO). The scientific exchange and discussion within the Year of Polar Prediction verification team have significantly contributed to the work described in this paper.

REFERENCES

- Bakkehøi, S., K. Øien, and E. Førland, 1985: An automatic precipitation gauge based on vibrating-wire strain gauges. *Hydrol. Res.*, **16**, 193–202, <https://doi.org/10.2166/nh.1985.0015>.
- 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>.
- Bromwich, D. H., and Coauthors, 2018: The Arctic system reanalysis, version 2. *Bull. Amer. Meteor. Soc.*, **99**, 805–828, <https://doi.org/10.1175/BAMS-D-16-0215.1>.
- Buisán, S. T., M. E. Earle, J. L. Collado, J. Kochendorfer, J. Alastrué, M. Wolff, C. D. Smith, and J. I. López-Moreno, 2017: Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network. *Atmos. Meas. Tech.*, **10**, 1079–1091, <https://doi.org/10.5194/amt-10-1079-2017>.
- , and Coauthors, 2020: The potential for uncertainty in numerical weather prediction model verification when using solid precipitation observations. *Atmos. Sci. Lett.*, **21**, e976, <https://doi.org/10.1002/asl.976>.
- Buizza, R., and Coauthors, 2017: IFS Cycle 43r3 brings model and assimilation updates. *ECMWF Newsletter*, No. 152, ECMWF, Reading, United Kingdom, 18–22, <https://www.ecmwf.int/en/newsletter/152/meteorology/ifs-cycle-43r3-brings-model-and-assimilation-updates>.
- 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/10.1256/smsqj.56613>.
- , 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/10.1256/smsqj.56614>.
- Colli, M., L. G. Lanza, R. Rasmussen, and J. M. Thériault, 2016: The collection efficiency of shielded and unshielded precipitation gauges. Part I: CFD airflow modeling. *J. Hydrometeorol.*, **17**, 231–243, <https://doi.org/10.1175/JHM-D-15-0010.1>.
- Cookson-Hills, P., D. J. Kirshbaum, M. Surcel, J. G. Doyle, L. Fillion, D. Jacques, and S. Baek, 2017: Verification of 24-h quantitative precipitation forecasts over the Pacific northwest from a high-resolution ensemble Kalman filter system. *Wea. Forecasting*, **32**, 1185–1208, <https://doi.org/10.1175/WAF-D-16-0180.1>.
- Dorninger, M., E. Gilleland, B. Casati, M. P. Mittermaier, E. E. Ebert, B. G. Brown, and L. J. Wilson, 2018: The setup of the MesoVICT project. *Bull. Amer. Meteor. Soc.*, **99**, 1887–1906, <https://doi.org/10.1175/BAMS-D-17-0164.1>.
- Forbes, R. M., A. M. Tompkins, and A. Untch, 2011: A new prognostic bulk microphysics scheme for the IFS. ECMWF Tech. Memo. 649, 30 pp.
- Førland, E. J., and Coauthors, 1996: Manual for operational correction of Nordic precipitation data. DNMI KLIMA Rep. 24/96, 66 pp.
- Frogner, I.-L., A. T. Singleton, M. Ø. Kjøltzow, and U. Andrae, 2019: Convection-permitting ensembles: Challenges related to their design and use. *Quart. J. Roy. Meteor. Soc.*, **145** (Suppl. 1), 90–106, <https://doi.org/10.1002/qj.3525>.
- Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, <https://doi.org/10.1175/2010BAMS2819.1>.
- Goble, P. E., N. J. Doesken, I. Durre, R. S. Schumacher, A. Stewart, and J. Turner, 2020: Who received the most rain today?: An analysis of daily precipitation extremes in the contiguous United States using CoCoRaHS and COOP reports. *Bull. Amer. Meteor. Soc.*, **101**, E710–E719, <https://doi.org/10.1175/BAMS-D-18-0310.1>.
- 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, 318 pp., <https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-67-solid-precip/WMOtd872.pdf>.

- 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>.
- Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, **140**, 2720–2733, <https://doi.org/10.1175/MWR-D-11-00301.1>.
- Jacobi, H.-W., K. Ebell, S. Schoger, and M. A. Wolff, 2019: Multi-instrument approach for the correction of observed precipitation in the Arctic. Poster presentation, *Svalbard Science Conf. 2019*, Oslo, Norway, Svalbard Science Forum, Abstract 1024, <https://www.forskningsradet.no/contentassets/f464e19d364c40b59170a1956a98e747/book-of-abstracts-ssc2019.pdf>; final poster is available at <https://drive.google.com/file/d/14SZfIRVcMtkAIQzW-ZTBqZ-2sfMuc2e/view?usp=sharing>.
- Jolliffe, I. T., and D. B. Stephenson, Eds., 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Sciences*. Wiley-Blackwell, 292 pp.
- Kielland, G., 2005: KVALOBS—The quality assurance system of Norwegian Meteorological Institute observations. Instruments and Observing Methods Rep. 82, WMO/TD-1265, 5 pp., [https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3\(13\)_Norway_Kielland.pdf](https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3(13)_Norway_Kielland.pdf).
- 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., B. Casati, E. Bazile, T. Haiden, and T. Valkonen, 2019: An NWP model intercomparison of surface weather parameters in the European Arctic during the year of polar prediction special observing period Northern Hemisphere 1. *Wea. Forecasting*, **34**, 959–983, <https://doi.org/10.1175/WAF-D-19-0003.1>.
- Masuda, M., A. Yatagai, K. Kamiguchi, and K. Tanaka, 2019: Daily adjustment for wind-induced precipitation undercatch of daily gridded precipitation in Japan. *Earth Space Sci.*, **6**, 1469–1479, <https://doi.org/10.1029/2019EA000659>.
- 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>.
- Nitu, R., and Coauthors, 2018: SPICE: WMO Solid Precipitation Intercomparison Experiment (2012–2015). Final Report, Instruments and Observing Methods Rep. 131, WMO, 1445 pp., <https://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html>.
- Pierre, A., S. Jutras, C. Smith, J. Kochendorfer, V. Fortin, and F. Anctil, 2019: Evaluation of catch efficiency transfer functions for unshielded and single-alter-shielded solid precipitation measurements. *J. Atmos. Oceanic Technol.*, **36**, 865–881, <https://doi.org/10.1175/JTECH-D-18-0112.1>.
- Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, <https://doi.org/10.1175/BAMS-D-11-00052.1>.
- Rodwell, M. J., D. S. Richardson, T. D. Hewson, and T. Haiden, 2010: A new equitable score suitable for verifying precipitation in numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **136**, 1344–1363, <https://doi.org/10.1002/qj.656>.
- Schirmer, M., and B. Jamieson, 2015: Verification of analysed and forecasted winter precipitation in complex terrain. *Cryosphere*, **9**, 587–601, <https://doi.org/10.5194/tc-9-587-2015>.
- Sebastianelli, S., F. Russo, F. Napolitano, and L. Baldini, 2013: On precipitation measurements collected by a weather radar and a rain gauge network. *Nat. Hazards Earth Syst. Sci.*, **13**, 605–623, <https://doi.org/10.5194/nhess-13-605-2013>.
- Sevruk, B., M. Ondras, and B. Chvila, 2009: The WMO precipitation measurement intercomparisons. *Atmos. Res.*, **92**, 376–380, <https://doi.org/10.1016/j.atmosres.2009.01.016>.
- Smith, C. D., 2009: The relationship between snowfall catch efficiency and wind speed for the Geonor T-200B precipitation gauge utilizing various wind shield configurations. *Proc. 77th Western Snow Conf.*, Canmore, Alberta, Canada, 115–121.
- , A. Ross, J. Kochendorfer, M. E. Earle, M. Wolff, S. Buisán, Y.-A. Roulet, and T. Laine, 2019: Evaluation of the WMO-Solid Precipitation Intercomparison Experiment (SPICE) transfer functions for adjusting the wind bias in solid precipitation measurements. *Hydrol. Earth Syst. Sci.*, **24**, 4025–4043, <https://doi.org/10.5194/hess-24-4025-2020>.
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, <https://doi.org/10.1002/2017RG000574>.
- Tapiador, F. J., and Coauthors, 2017: Global precipitation measurements for validating climate models. *Atmos. Res.*, **197**, 1–20, <https://doi.org/10.1016/j.atmosres.2017.06.021>.
- Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3061, [https://doi.org/10.1175/1520-0493\(1993\)121<3040:ROCILS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2).
- Vormoor, K., and T. Skaugen, 2013: Temporal disaggregation of daily temperature and precipitation grid data for Norway. *J. Hydrometeorol.*, **14**, 989–999, <https://doi.org/10.1175/JHM-D-12-0139.1>.
- Wang, G., X. Zhang, and S. Zhang, 2019: Performance of three re-analysis precipitation datasets over the Qinling-Daba Mountains, Eastern Fringe of Tibetan Plateau, China. *Adv. Meteor.*, **2019**, 1–16, <https://doi.org/10.1155/2019/7698171>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wolff, M., 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>.