kongoli Cezar (Orcid ID: 0000-0002-9484-0064)

A Hybrid Snowfall Detection Method from Satellite Passive Microwave Measurements and Global Forecast Weather Models

Cezar Kongoli[12], Huan Meng[2], Jun Dong[2], Ralph Ferraro[2]

[1]University of Maryland/ Earth System Science Interdisciplinary Center/Cooperative Institute for Climate and Satellites-Maryland

[2] National Oceanic and Atmospheric Administration/ National Environmental Satellite, Data, and Information Service

Corresponding author e-mail: cezar.kongoli@noaa.gov

Keywords: Satellite Remote Sensing, Passive Microwave Measurements, Snowfall Detection, Global Weather Prediction

Running Head: A Hybrid Snowfall Detection Method

Abstract

Despite significant progress made in snowfall estimation from space, methods utilizing passive microwave measurements continue to be plagued by low detectability compared to those that estimate rainfall. This paper presents a hybrid snowfall detection algorithm that combines the output from a statistical algorithm utilizing satellite passive microwave measurements with the output from a statistical algorithm trained with in-situ data that uses meteorological variables derived from a global forecast model as predictors. The satellite algorithm computes the probability of snowfall over land using logistic regression and the principal components of the high frequency brightness temperature measurements at AMSU/MHS and ATMS channel frequencies 89 GHz and above. In a separate investigation, analysis of modelled data derived from NOAA's Global Forecast System (GFS) showed that cloud thickness and relative humidity at 1- to 3-km height were the best predictors of snowfall occurrence. A statistical logistical regression model that combined cloud thickness, relative humidity and vertical velocity was selected among statistically significant variants as the one with the highest overall classification accuracy. Next, the weather-based and satellite model outputs were combined in a weighting scheme to produce a final probability of snowfall output, which was then used to classify a weather event as "snowing" or "no snowing" based on an a-priori threshold probability. Statistical analysis indicated that a scheme with equal weights applied to the weather-based and

2

satellite model significantly improved satellite snowfall detection. Example applications of the hybrid algorithm over continental US demonstrated the improvement for a major snowfall event and for an event dominated by lighter snowfall.

1. Introduction

Snowfall prediction remains one of the most difficult challenges in weather forecasting. Current precipitation forecasts rely on observations from gauges, radars, and satellites which augment those made by numerical weather prediction (NWP) models. For all types of precipitation, each of these data sources has its limitations which become even more serious when the precipitation falls in the form of snow. Additionally, different NWP models often provide conflicting information regarding snowfall location, extent, and intensity. This fact makes it especially important for weather forecasters to have reliable observations to confirm model predictions and to infer more accurate storm information for the forecasting areas.

With broad and unobstructed coverage, satellite observations provide an excellent alternative to ground based gauge and radar data. Routine satellite snowfall detection from visible or infrared measurements has a major limitation in that it is very difficult to distinguish between precipitating and non-precipitating high-latitude clouds during winter. Progress has only been made recently (Staelin and Chen, 2000; Chen and Staelin, 2003; Kongoli et al., 2003; Skofronick-Jackson et al., 2004; Surussavadee et al., 2012; Liu and Seo, 2013; Kongoli et al.,

3

2015) when high frequency passive microwave instruments such as the Advanced Microwave Sounding Unit (AMSU) began flying on board polar orbiting satellites. Kongoli et al. (2003) presented the first operational snowfall detection algorithm using AMSU measurements (Ferraro et al., 2005). Detection of snowfall was based on a decision tree classification algorithm using measurements in the microwave window, water vapor and oxygen absorption regions. The underlying assumption was that for warmer and more opaque atmospheres, scattering by snowfall-sized ice particles decreases the brightness temperatures at the high frequency channels (89 GHz and above), and this scattering signature can be detected using opaque (oxygen and water vapor absorption) and window channels in combination. For colder and less opaque atmospheres, retrievals were considered too noisy due to surface effects, and thus were not performed.

Recent advances in space-borne radar systems and in particular the Cloud Profiling Radar (CPR) on board NASA's CloudSat satellite led to new insights into winter precipitation and the development of new and improved satellite passive microwave snowfall detection algorithms. Liu and Seo (2013) developed a probability-based statistical method from MHS channel measurements using Cloudsat radar data as "ground truth". A new finding was that on most occasions, the brightness temperatures are higher under snowfall than no snowfall conditions, likely due to emission by cloud liquid water which masks the scattering signal. Motivated by the results of this study, Kongoli et al. (2015) analyzed snowfall signatures using measurements from the Advanced Technology Microwave Sounder (ATMS) onboard Suomi National Polar-

4

orbiting Partnership (SNPP) and in-situ data as ground truth. They also found that brightness

temperatures at ATMS high frequency channels are significantly higher for snowfall, but only in

colder weather conditions. In warmer weather, brightness temperatures typically decreased for

snowfall compared to no-snowfall, likely due to the scattering signal dominating the response.

A two-step probability-based snowfall detection algorithm was developed and trained with in-

situ data for a colder and a warmer weather regime, using the AMSU or ATMS limb corrected

oxygen absorption channel brightness temperature at 53.6 GHz (referred to hereafter as TB53L)

as temperature proxy to define each weather regime. The colder snowfall regime retrievals

extended to conditions associated with TB53L down to 240 Kelvin, which corresponds to a near

surface temperature at about -15 ^{0}C. For TB53L less than 240 Kelvin retrievals were deemed

indeterminate due to the lack of a sufficiently large and representative training sample to train

and evaluate the algorithm in these very cold weather conditions. Figure 1 presents an example

using the threshold-based Kongoli et al., 2003 and the probability-based Kongoli et al., 2015

snowfall detection algorithms. Shown on the map are snowfall rates retrieved using a 1DVAR-

based physical inversion scheme (Meng et al., 2017) applied over snowfall detected areas. The

new snowfall detection algorithm retrieves a larger fraction of snowfall including over

"indeterminate" colder areas (denoted as white) and at a lower false alarm.

Despite the advances mentioned above, space-based estimation continues to be a highly

difficult problem in modern hydrometeorology (Levizzani et al., 2011). Specifically, there are

inherent limitations to achieving a high snowfall detectabilty from satellite passive microwave

5

measurements. Evaluation of the new algorithm of Kongoli et al. (2015) has revealed its

performance can degrade for light snowfall, snowfall generated from shallow clouds and

significant snowfall associated with a weak scattering signal. One strategy to improve satellite

retrievals is development of algorithms that exploit information from modeled meteorological

variables. In this paper, a hybrid snowfall detection scheme is presented that explicitly blends

satellite snowfall output with that estimated from ancillary meteorological data derived from a

global forecast system. To this end, an algorithm using forecast parameters is trained with in-

situ data to compute the probability of snowfall, which is then combined with that of the satellite

in a weighting scheme to produce the final output. The paper is organized as follows. Section 2

describes the satellite, weather-based and the proposed hybrid algorithm. Section 3 details the

data used: in-situ, satellite and modeled data, as well as the collocation methodology. Section 4

presents and discusses the results for the weather-based, satellite and the hybrid model outputs.

Finally, summary and conclusions are provided in section 5.

2. Methods

2.1 Satellite Snowfall Detection

The satellite snowfall detection algorithm (hereafter referred to as "satellite SD algorithm") is

described in detail in the study of Kongoli et al. (2015). Here we provide a general outline. The

satellite SD algorithm computes the probability of snowfall over land using logistic regression

6

and the principal components (PCs) of the high frequency brightness temperature measurements

at AMSU/MHS and ATMS channel frequencies 89 GHz and above.

The probability of snowfall is expressed with the logistic regression function:

$$\text{Ln}(P/(1-P)) = a0 + a1*PC1 + a2*PC2 + a3*PC3 + a4*cosLZA \tag{1}$$

Where P is probability of snowfall, *cosLZA* is the cosine of the zenith angle, and *PC1*, *PC2* and

PC3 are the first three PCs. Each *PC* is computed as a linear combination of the brightness

temperatures multiplied by fixed weights derived in algorithm training. The probability of

snowfall P is computed as:

$$P = \text{Exp}(B) / (1 + \text{Exp}(B)) \tag{2}$$

where B is the expression on the right side of Eq. 1. The oxygen absorption channel at 53.6 GHz

(AMSU-A channel 5 and ATMS channel 6) is utilized as temperature proxy to define two

retrieval regimes: a colder and a warmer one. To optimize retrievals, the principal component

weights and partial logistic regression coefficients are trained with in-situ station observations of

snowfall and no-snowfall occurrence and pre-computed separately for the warmer (TB53L

between 244 and 252 Kelvin) and colder (TB53L between 240 and 244 Kelvin) weather regimes.

In the operational application of the algorithm, NWP-derived land surface temperature (LST) is

used to filter out liquid precipitation in the form of rain. This is consistent with the algorithm

training against only snowfall and no-precipitation cases during the winter. This algorithm is a

major advancement compared to the previous version of Kongoli et al. (2003) in that it allows

7

snowfall retrievals in colder environments down to near surface temperatures at about -15^0 C. In addition, the statistical probabilistic approach is a more robust method than the previous decision tree approach. A snowfall rate algorithm (Meng et al., 2017) is applied to snowfall-detected scenes to retrieve cloud ice water path and particle effective diameter, which are then used to estimate surface snowfall rate (Figure 2). These cloud properties are retrieved using an inversion method with an iteration algorithm and a two-stream Radiative Transfer (RT) Model (Yan et. al, 2008). Next, snow particle terminal velocity is computed (Heymsfield and Westbrook, 2010) and snowfall rate by numerically solving a complex integral.

2.2 Weather-based Snowfall Detection

A similar probabilistic logistic regression approach was adopted for snowfall detection using forecast meteorological variables as predictors (hereafter referred to as "weather-based SD algorithm"). The practical utility for choosing to develop a statistical algorithm from weather forecast data is to improve satellite model output (from any current or future satellite) in operational applications by blending it explicitly with a weather-based model output in a simple weighting scheme. Experience with satellite passive microwave instruments and theoretical investigations (e.g., Munchak and Johnson, 2013) indicate that the detectability of precipitation over snow cover surfaces deteriorates significantly compared to bare land. In addition, snowfall analysis and estimation from physical parameters would be desirable in and of itself, for providing alternative retrieval strategies and insights into the relative importance of these parameters in snowfall processes. Note that this type of investigation has mostly focused on

8

rainfall. For instance, the relationship between cloud top temperature and rainfall has been widely used in the development of rain rate retrieval algorithms from satellite infrared measurements (e.g., Arkin et al., 1987; Vinzente et al., 1998). Cloud top temperature is a proxy for cloud thickness, the latter more physically related to rain rate. Yu et al. (2016) explored the influence of several environmental parameters on both rain and snowfall estimations and found that relative humidity and vertical velocity are related to the occurrence of snowfall, more so than that of rainfall. On the other hand, experience with the AMSU/MHS and ATMS satellite algorithms have shown that cloud thickness computed from forecast data is an efficient filter in reducing false alarms (Meng et al., 2017). Therefore, to develop the algorithm, the forecast variables considered include relative humidity at 2-m, 1-km, 2-km and 3-km height, cloud thickness, cloud top height, cloud top temperature, and vertical velocity at 1-km, 2-km, and 3-km height.

2.3 Hybrid Algorithm

The rationale for the hybrid algorithm is to compute an output as a weighted average of outputs from the satellite and the weather-based SD algorithms:

$$PR_{hyb} = W_{sat} * PR_{sat} + W_{wea} * PR_{wea} \qquad (3)$$

9

Where *PR* refers to the probability of snowfall, *W* refers the weight, and *hyb*, *sat* and *wea* refer to the hybrid, satellite and the weather-based SD algorithms, respectively. Note that $W_{sat} + W_{wea} = 1$. Equation 3 may therefore be written as:

$$PR_{hyb} = f*PR_{sat} + (1-f)* PR_{wea} \qquad (4)$$

Where $f = W_{sat}/W_{wea}$. Presented with a set of brightness temperatures and ancillary data, for a specific weighting parameter *f*, snowfall probabilities for the satellite and the weather model are computed based on Eq.4. Next, the resulting probability value (PR_{hyb}) is assigned to "snowfall" if it is greater than an *a priori* threshold probability.

3. Datasets and Collocation Methodology

3.1 In-situ Ground Truth Data

In-situ ground truth data for algorithm training and evaluation were obtained from the Quality Controlled Local Climatology Data (QCLCD) product distributed by NOAA's National Climate Data Center (NCDC, www.ncdc.noaa.gov). The data consists of hourly, daily, and monthly summaries for approximately 1,600 U.S. locations. Data are available beginning January 1, 2005 and continue to the present. In this study the hourly dataset was used, which includes measurements of 2-m surface temperature, wet bulb temperature, pressure, relative humidity, visibility and present weather. The present weather contains information on the type of precipitation, which was used for (ground truth) classification of cases into falling snow or no-

10

precipitation. Reported in this dataset is also hourly snowfall accumulation in liquid water equivalent.

3.2 Satellite Data

The satellite SD algorithm utilizes measurements from ATMS or the AMSU-A/MHS pair. Presented in this study are quantitative analysis and results for the algorithm based on ATMS and only a qualitative assessment of the AMSU-A/MHS-based model.

ATMS is the successor of AMSU-A and MHS. It is aboard the Suomi National Polar-orbiting Partnership (S-NPP) satellite. It will also be the only microwave radiometer aboard the future Joint Polar Satellite System (JPSS) satellites. Table I presents ATMS channel characteristics, including center frequencies, noise equivalent temperature (NEDT), horizontal spatial resolution and corresponding AMSU channels. ATMS observes millimeter-wave spectra at 22 frequencies. Channels 1–15 observe 53-GHz oxygen and 23-GHz water vapor absorption bands and resemble AMSU-A channels. Channels 16–22 observe near and below the 183-GHz water vapor resonance and resemble AMSU-B (MHS) channels. As shown, ATMS's horizontal spatial resolution at nadir is 75 km for channels 1 and 2, 33 km for channels 3 – 16, and 15 km for channels 17 – 22. AMSU's spatial resolution at nadir is 50 km for channels below 60 GHz and 15 km otherwise. The improvements of ATMS over AMSU include 1) about 400 km wider swath reducing gaps between orbits; 2) the addition of observation at 51.76 GHz providing more information about surface, stratiform precipitation, and tropospheric temperature profiles; 3) the addition of observation at 183.31±1.8 and ±4.5 GHz providing more information about water

11

vapor and precipitation; 4) better spatial resolution for 50-GHz channels, that is, from ~50 to ~33 km at nadir; and 5) Nyquist spatial sampling for channels below 90 GHz, which enables discretionary image sharpening. ATMS surface channels near 23.8, 31.4, and 88.2 GHz have lower spatial resolution than those of AMSU.

3.3 GFS Data

Ancillary geophysical data were obtained from the Global Forecast System (GFS), the global operational NWP model of NOAA's National Centers for Environmental Prediction (NCEP). GFS model is run four times daily and generates output out to 192-hour at 3-hour increments. For this application, the 1 degree spatial resolution GFS output was utilized. Variables obtained include layer relative humidity at 2 m, 1 km, 2 km and 3 km height, cloud thickness, cloud top and base heights, cloud top and base temperatures, and vertical velocity at 1 km, 2 km, and 3 km height. Cloud thickness was computed utilizing GFS relative humidity information. An atmospheric level is considered to have cloud if relative humidity is at or above 89%. This threshold is consistent with the critical relative humidity value used in the GFS microphysics scheme for one-degree resolution at mid-latitude locations (Meng et al., 2017).

3.4 Data Matching Methodology

A dataset was compiled matching satellite, in-situ ground truth and GFS data. The ground truth in-situ data contained information on observed snowfall occurrence (snowfall vs. no-snowfall) that was needed for algorithm training and evaluation. Periods during the winter

12

seasons between 2014 and 2016 were selected to sample snowfall events over continental United States. First, ATMS swath satellite brightness temperature data were matched to GFS data to a. compute the probability of snowfall output by application of the satellite SD algorithm, and b. obtain GFS variables for the development of the weather-based SD algorithm. Next, the matched satellite and GFS data were collocated with in-situ hourly station observations to classify each case as "falling snow" or "no snowfall". Maximum time offset between the satellite and station-collocated pairs was set at 30 minutes with satellite time stamp following station time. Maximum separation distance between satellite footprint centroid and station location was set at 25 km. Only the closest station within the 25 km distance from the satellite footprint centroid was matched. Another important consideration in collecting the ground truth snowfall sample from in-situ observations was the selection of snowfall cases in colder weather conditions, as indicated by station 2 m surface temperature or the ATMS or AMSU-A limb-corrected brightness temperature at 53.6 GHz. Examination of snowfall observations revealed that lighter snowfall with hourly liquid water equivalent accumulations reported as "trace" was a significant fraction. The minimum liquid water equivalent rate measured and reported is 0.25 mm hr^{-1} and thus rates below this value are reported as "trace". For colder weather corresponding to 53.6 GHz limb-corrected brightness temperatures 244 K and below, most of the snowfall sampled consisted of "trace" snowfall cases. A case with "trace" snowfall accumulation was retained in the snowfall sample when it was also flagged as "snowing" at the time of observation. "No-snowing" cases selected were only those that had no snow or rain

13

reported at observation time and the liquid water equivalent rate reported was zero. In other words, a case with "trace" amounts of hourly accumulation that was not flagged as "snowing" at observation time (present weather was other than "snow" or "rain") was not selected in the ground truth snowfall sample. However, a "snowing" case (present weather was flagged as "snowing") that had zero accumulations of hourly snowfall was retained in the ground truth snowfall sample. The zero-precipitation "snowing" cases were a significant fraction that if removed would have resulted in a much smaller ground truth snowfall sample. "Snowing" cases with 2-m surface temperature at 2°C and above were removed from the snowfall sample due to a higher likelihood of being a rainfall case. These criteria were applied as additional quality control checks to minimize ambiguity in ground truth data as much as possible. The resulting match-up sample contained over 43,000 cases in total, with 37% consisting of "snowing" cases.

4. Results and Discussion

4.1 Weather-based SD Algorithm

Logistic regression results for the modeled weather variables as predictors and ground truth snowfall occurrence as the response indicate that cloud thickness gives the highest overall classification accuracy of snowfall versus no snowfall at 77%, followed by 1 km and 2 km relative humidity at 76%, 3 km relative humidity at 73% and cloud top height at 70%. Overall accuracy was computed as the ratio between the correct number of cases classified (snowfall + no snowfall) and the total count of cases considered, for a threshold probability of 0.5. Vertical velocity and cloud top temperature were significant (P-value less than 0.001) but each gives a

14

lower overall classification rate. Interestingly, cloud top temperature has the highest (negative) correlation with modeled snowfall rate (greater than 0.1 mm hr[-1]) retrieved from ATMS (R =-0.25) and measured surface precipitation rate greater than 0.25 mm hr[-1] (R =-0.11). Cloud thickness comes in second in the magnitude of the correlation value (positively correlated with modeled and measured snowfall rate).

Figure 3 displays the distribution of cloud thickness for the snowfall and the no snowfall sample. Consistent with regression results, one can visually distinguish the marked difference in the shape of the distributions, e.g., the much larger fraction of snowfall cases associated with thicker clouds compared to the no-snowfall sample. This also results in a large difference in the mean and median values of cloud thickness between the snowfall and the no snowfall sample. Figure 4 displays histograms of the relative humidity distribution at 1 km, 2 km and 3 km height for the snowfall and no snowfall sample. The snowfall sample is associated with a larger fraction of high relative humidity than the no-snowfall sample. Figure 5 displays same distributions of relative humidity as Figure 4, but for thinner clouds (less than 3000 m in thickness). For the 3 km relative humidity, the shape of the distributions between the snowfall and no snowfall sample becomes closer than for the 1 km and 2 km relative humidity, suggesting that for thin clouds 1 and 2 km relative humidity parameters are more important in distinguishing between snowfall and no snowfall. Figure 6 shows the histograms of cloud top temperature for the snowfall and the no snowfall sample. Cloud top temperature mean value for the snowfall sample is -18.2 ^{0}C and the distribution is nearly bell-shaped. On the other hand, the mean value

15

for the no-snowfall sample is higher than that of the snowfall, and the distribution is also nearly

bell shaped except for the peak of the distribution centered at 1 ^{0}C.

Table II gives logistic regression model results using the stepwise elimination method.

Shown in the table are the variables selected, partial regression coefficients, the Z-test statistic

values and the significance (P-values). The selected model variables are cloud thickness, the 2

m, 1 km, and 3 km level humidity, and the 2 km and 3 km level vertical velocity. Other

combinations of variables and models were considered in logistic regression, with performance

results not better than the one presented here. Table III presents performance statistics of the

selected model: the confusion matrix, the probability of snowfall detection (POD), the false

alarm rate (FAR), the overall classification rate, and the Heidke Skill Score (HSS).   POD was

computed as the fraction of correctly predicted snowfall cases (out of all the snowfall cases),

FAR as the fraction of incorrectly predicted no-snowfall cases (out of all the no-snowfall cases).

Heidke Skill Score (HSS; Wilks, 2011) was computed using the following:

HSS = 2(ad-bc)/[a+c)(c+d) + (a+b)(b+d)]   *(5)*

Where *a* is the number of events correctly forecasted, in our case, snowfall, *d* is the number of

no-events correctly forecasted, in our case, no-snowfall, *c* is the number of no-events incorrectly

forecasted, in our case, no-snowfall, and **d** is the number of events incorrectly forecasted, in our

case, snowfall.   A negative HSS value means that the model does worse than a chance forecast,

and a positive value between zero (no skill) and 1 (perfect skill) indicates progressively better

skill than the chance forecast. Overall accuracy and HSS are moderately high at 79% (with a

16

threshold classification probability of 0.5) and 0.55, respectively. POD, FAR, overall

classification rate and HSS depend on the selected threshold probability value: A higher

threshold probability will always lower both FAR and POD, but may or may not lower overall

classification or HSS. A threshold probability of 0.6, for instance, gives the following: POD =

0.62, FAR = 0.12, overall classification = 0.78 and HSS = 0.52.

4.2 Hybrid Model and Comparisons with the Satellite SD Algorithm

An important parameter in the hybrid algorithm is the weighting coefficient f in eq. 4,

which measures the relative uncertainty information of the satellite output with respect to the

model forecast. A rigorous analysis and estimation of this parameter is not the subject of this

study. Monitoring of the hybrid algorithm performance suggests that a weighting scheme with f

varying between 0.4 and 0.6 produces improved results *over continental US* across all satellite

platforms carrying ATMS (NPP) or the AMSU/MHS pair (NOAA-18, -19, and Metop). Table

IV presents statistical results of the ATMS satellite and the hybrid algorithm, the latter using a

weighting coefficient f equal to 0.5, i.e., when GFS-based and satellite outputs have the same

uncertainty information. Overall classification rate is improved from 70% to 75% and HSS from

0.23 to 0.44. It is important to emphasize that the statistical results reported here are computed

from comparisons with in-situ measurements *where light snowfall cases* (reported as "trace" or

with zero surface accumulation) *were dominant*.

17

Figures 7 and 8 demonstrate the improvement of the hybrid model for a snowfall event on February 5, 2014 and more recently on 31 January, 2017 (Figures 9 and 10). The satellite algorithm parameters (PC scores and partial logistic regression coefficients) for ATMS (Figure 7) and AMSU-MHS pair (Figure 9) have been pre-computed separately from training with in-situ data, whereas the weather based algorithm applied to each satellite instrument is the same as the one reported here. As shown on Figure 7, a considerable snowfall area of a significant event was missed by the satellite algorithm but captured by the hybrid algorithm. This was a major snowfall event heading Northeast of US and moving over antecedent snow-cover. Snow cover maps generated from NOAA's Interactive Snow and Ice Mapping System (IMS; Helfrich et al., 2007) before (top images) and after (bottom image) 5 February, 2014 are shown in Figure 8. Visual inspection of these maps suggests that areas hit by the snowstorm on February 5 were already snow-covered on 4 February, 2017.

The snowfall event depicted in Figure 9 was less extensive, consisting mostly of lighter snow falling over snow-covered land as suggested by IMS snow cover maps over continental US prior to and following the event (Figure 10). Visual comparison between the satellite snowfall detected area and that from the radar reveals a slight displacement: missed satellite snowfall on one edge and false alarm on the other. These satellite mis-classification errors could have occurred right after the start (missed snowfall) and after the end of a moving snowstorm (false new snow on the ground), and both these areas are correctly classified by the hybrid algorithm.

Summary and Conclusions

18

Snowfall prediction remains one of the most difficult challenges in weather forecasting. This paper presents a hybrid snowfall detection algorithm that weights the outputs from a statistical algorithm utilizing satellite passive microwave measurements and from a statistical algorithm based on ancillary weather data derived from a global forecast model. This approach is taken to improve the detectability of satellite snowfall and the utilization of satellite methods in weather and hydrological forecasting. The satellite algorithm computes the probability of snowfall over land using logistic regression and the principal components of the high frequency brightness temperature measurements at AMSU/MHS and ATMS channel frequencies 89 GHz and above. This algorithm is a major advancement compared to a previous version in that it allows snowfall retrievals in colder environments, down to near surface temperatures at about -15^0 C. Algorithm parameters consisting of principal component weights and partial regression coefficients have been pre-computed in previous studies using in-situ data for training. A similar logistic regression method is applied to train a weather-based algorithm, with meteorological variables derived from NOAA's Global Forecast System (GFS) as predictors. In-situ data are obtained from the Quality Controlled Local Climatology Data (QCLCD) hourly product distributed by NOAA's National Climate Data Center. Specifically, GFS-derived variables considered were cloud thickness, cloud top height, cloud top temperature, atmospheric humidity at 2-m, 1-km, 2-km, and 3-km height, and vertical velocity at 1-km, 2-km, and 3-km height. Logistic regression analysis showed that cloud thickness and relative humidity were the parameters with the largest impact on classification accuracy. A statistical logistical regression

19

model that combined cloud thickness, humidity and vertical velocity was selected among different statistically significant combinations as the one maximizing accuracy, with an overall classification rate at 79% and a Heidke Skill Score of 0.55. Next, weather-based and satellite model output expressed as probability of snowfall were combined in a simple weighting scheme to produce a final probability of snowfall output, which was then used to classify an event as "snowing" based on a a priori threshold probability. Statistical analysis indicated that a scheme with equal weights applied to the weather-based and satellite model significantly improved satellite snowfall detection. Example applications of the hybrid algorithm over continental US demonstrated the improvement for a major snowfall event and for an event dominated by lighter snowfall. Future work is expected to extend retrievals over very cold snowfall conditions, e.g, in areas over Alaska that has temperatures lower than in Continental US, as well as to refine the weighting scheme.

5.   References

Arkin PA., Ardanuy PE. 1989. Estimating climatic-scale precipitation from space: a review. *J. Climate*, 2, 1229-38, DOI: 10.1175/1520-0442(1989)002<1229:ECSPFS>2.0.CO;2.

Chen FW., Staelin DH. 2003. AIRS/AMSU/HSB precipitation estimates.*IEEE Trans. Geo. Rem. Sens.*, 41(2), 410-417, DOI**:** 10.1109/TGRS.2002.808322.

Ferraro R.R., Weng F., Grody NC., Zhao L., Meng H., Kongoli C., Pellegrino P., Qiu S., Dean C. 2005. NOAA operational hydrological products derived from the Advanced Microwave Sounding Unit, IEEE Transactions on Geoscience and Remote Sensing, 43, 1036, DOI: 10.1109/TGRS.2004.843249.

Helfrich SR, McNamara D, Ramsay BH, Baldwin T, Kasheta T. 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS), *Hydrological Processes*, 21:1576–1586, DOI:10.1002/hyp.6720.

Heymsfield AJ., Westbrook CD. 2010. Advances in the estimation of ice particle fall speeds using laboratory and field measurements. *J. Atmos. Sci.*, 67, 2469-2482, DOI: 10.1175/2010JAS3379.1.

Meng H., Ferraro RR., Yan B., Zhao L., Dong J., Kongoli C., Wang N-Y, Zavodsky B. A 1DVAR snowfall rate retrieval algorithm for passive microwave radiometers. *Journal of geophysical Research-Atmospheres*, 122(12), 6520-6540. **DOI:** 10.1002/2016JD026325.

Munchak SJ., Jackson GS.2013. Evaluation of precipitation detection over various surfaces from passive microwave imagers and sounders, *Atmospheric Research*, 131, 81-94, DOI:10.1016/j.atmosres.2012.10.011.

21

Kongoli, C, Pellegrino P., Ferraro RR., Grody NC., Meng H. 2003. A new snowfall detection algorithm over land using measurements from the Advanced Microwave Sounding Unit (AMSU*), Geophys. Res. Lett*., 30(4), 1756-1759, DOI**:** 10.1029/2003GL017177.

Kongoli C.,Meng H, Dong J, Ferraro RR. 2015. A Snowfall detection algorithm over land utilizing high-frequency passive microwave measurements – Application to ATMS. *J. Geophys. Res. – Atmospheres*, 120(5), DOI: 10.1002/2014JD022427.

Laviola S., Dong J., Kongoli C., Meng H., Ferraro RR., Levizzani V. 2015. An intercomparison of two passive microwave snowfall detection algorithms over Europe, *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, Milan, Italy, DOI: 10.1109/IGARSS.2015.7325878.

Levizzani V., Laviola S., Cattani E. 2011. Detection and Measurement of Snowfall from Space, *Remote Sens*, 3, 145-166, DOI:10.3390/rs3010145.

Liu G. and Seo E.2013. Detecting snowfall over land by satellite high-frequency microwave observations: The lack of scattering signature and a statistical approach.*Journal of Geophysical Research – Atmospheres,* 118(3), 1376-1387, DOI: 10.1002/jgrd.50172.

22

Skofronick-Jackson, G., Kim MJ, Weinman J, Chang DE.2004. A physical model to determine snowfall over land by microwave radiometry.*IEEE Trans. Geosci. Rem. Sens.*, 42**,** 1047-1058, DOI:10.1109/TGRS.2004.825585

Staelin DH., Chen FW. 2000. Precipitation observations near 54 and 183 GHz using the NOAA-15 satellite.*IEEE Trans. Geo.Rem. Sens.*, 38(5), 2322-2332. DOI**:** 10.1109/36.868889.

Surussavadee C., Blackwell WJ., Entekhabi D., Leslie RV. 2012.A global precipitation retrieval algorithm for Suomi NPP ATMS, Geoscience and Remote Sensing Symposium (IGARSS), p. 1924 – 1927, DOI: 10.1109/IGARSS.2012.6351128, 22-27 July 2012, Munich, Germany.

Vicente, GA.,Scofield RA, Menzel WP. 1998. The operational GOES infra-red estimation technique. *Bull. Amer. Meteor. Soc.,* 79, 1883-98, DOI: 10.1175/1520-0477(1998)079<1883:TOGIRE>2.0.CO;2.

Wilks DS. 2011. Statistical Methods in the Atmospheric Sciences. Elsevier

Yan B., Weng, F., Meng H.2008. Retrieval of Snow Surface Microwave Emissivity under All Weather Conditions from AMSU Window Channel Brightness Temperatures, *J. Geophys. Res.,* 113, D19206. DOI:10.1029/2007JD009559.

Yalei Y., Wang N-Y, Ferraro RR., Meyers P. 2016. A Prototype precipitation retrieval algorithm over land for ATMS, Journal *of Geophysical Research-Atmospheres*, DOI: 10.1175/JHM-D-15-0163.1.

List of Figures

Figure 1. ATMS snowfall rate (top left) applied to the Kongoli et al. (2003) version of the snowfall detection algorithm (left) and the new more advanced probability-based version of Kongoli et al. (2015) (top right). The bottom image is the radar-based precipitation type over US.   The "white" coded area on the top left image denotes weather associated with TB53L less than 245 Kelvin over which snowfall retrievals were not performed. The new algorithm extends retrievals over these areas and performs better overall.

Figure 2 (from Laviola et al.,2015). Left: new ATMS SD algorithm, and right: Inversion of Radiative Transfer (RT) module in the snowfall rate algorithm of Meng et al., (submitted). Limp-corrected channel 6 ATMS brightness temperature (TB53L) is used as proxy for surface temperature. The colder regime extension (TB53L down to 240 K or approx. -15 ^{0}C surface temperature) allows colder and lighter snowfall retrievals than the heritage.

Figure 3. Distribution of GFS cloud thickness (m) over the no snowfall and snowfall samples. The vertical red and green lines represent the median and mean values of cloud thickness, respectively.

Figure 4. Distribution of GFS relative humidity over the no snowfall and snowfall samples. The vertical red and green lines represent the median and mean values of relative humidity, respectively.

Figure 5. Distribution of GFS relative humidity over the no snowfall and snowfall samples for cloud thickness less than 3000 m. The vertical red and green lines represent the median and mean values of relative humidity, respectively.

Figure 6. Distribution of GFS cloud top temperature over the no snowfall and snowfall samples. The vertical red and green lines represent the median and mean values of cloud top temperatures, respectively.

Figure 7. Snowfall rate from ATMS satellite (top left) and the hybrid (top right) snowfall detection algorithms, the latter driven by GFS derived variables, during a major snowfall event in February 2014. The bottom image is the near coincident radar reflectivity. The noted oval areas show legitimate snowfall that was missed by the satellite algorithm but captured by the hybrid algorithm.

Figure 8. Snow on the ground over continental US from the NOAA's Interactive Multi-Sensor and Ice mapping System (IMS) before (top images) and after (bottom image) February 5, 2014.

Figure 9. Snowfall rate from satellite (top left) and the hybrid (top right) snowfall detection algorithm from AMSU-MHS pair on NOAA-19 satellite and GFS data, respectively. The bottom image is the near coincident radar reflectivity.

Figure 10. Snow on the ground over continental US from the NOAA's Interactive Multi-Sensor and Ice mapping System (IMS) before (top) and after (bottom) January 31, 2017.

List of Tables

Table I: ATMS and AMSU/MHS channel characteristics

Table II: Logistic regression coefficients of the weather model

Table III. Performance statistics of the weather, satellite and the hybrid snowfall detection model

Table IV. Performance statistics of the satellite and the hybrid snowfall detection model for a threshold probability = 0.5

Table I: ATMS and AMSU/MHS channel characteristics

ATMS				AMSU[+]	
Ch.	Frequencies (GHz)	Predicted NEΔT(K)[*]	Nadir (km)	Ch.	Measured NEΔT(K)
1	23.80	0.28	75	1	0.21
2	31.40	0.35	75	2	0.26
3	50.30	0.42	33	3	0.22
4	51.76	0.31	33	-	-
5	52.80	0.32	33	4	0.14
6	53.596±0.115	0.35	33	5	0.15
7	54.40	0.32	33	6	0.15
8	54.94	0.32	33	7	0.13
9	55.50	0.35	33	8	0.14
10	f_0−57.290344	0.49	33	9	0.24
11	f_0±0.217	0.67	33	10	0.25
12	f_0±0.3222±0.048	0.70	33	11	0.28
13	f_0±0.3222±0.022	1.06	33	12	0.40
14	f_0±0.3222±0.010	1.45	33	13	0.54
15	f_0±0.3222±0.045	2.40	33	14	0.91
16	88.2	0.29	33	16[#]	0.35
17	165.6	0.44	15	17[#]	0.76
18	183.31±7.0	0.34	15	20	0.55
19	183.31±4.5	0.39	15	-	-
20	183.31±3.0	0.48	15	19	0.68
21	183.31±1.8	0.49	15	-	-
22	183.31±1.0	0.62	15	18	0.98

Center frequencies for channels 16 and 17 are 89 and 157 GHz respectively.

*Integration times for ATMS channels 1-16 are one ninth those of AMSU.

+AMSU resolution at nadir is 50 km below 60 GHz ad 15 km otherwise.

Table II: Logistic regression coefficients of the weather model

Variable	Estimate	Std. Err.	Zstat	P-value
Cthick	0.000634	1E-05	-43.52	<0.0001
Hum1	0.032457	3E-05	25.19	<0.0001
Hum3	-0.005273	9E-04	35.95	<0.0001
V2	0.584000	8E-04	-6.601	<0.0001
V3	-0.717000	6E-05	10.22	<0.0001
Hum	0.027847	6E-05	-11.41	<0.0001
Intercept	-6.380459	0.002	17.24	<0.0001

Table III. Performance statistics of the weather, satellite and the hybrid snowfall detection model

Model/Statistics	FAR	POD	Overall	Heidke
Weather1	0.17	0.72	0.79	0.55
Weather2	0.12	0.62	0.78	0.52

1 Threshold probability = 0.5
2 Threshold probability = 0.6

Table IV. Performance statistics of the satellite and the hybrid snowfall detection model for a threshold probability = 0.5

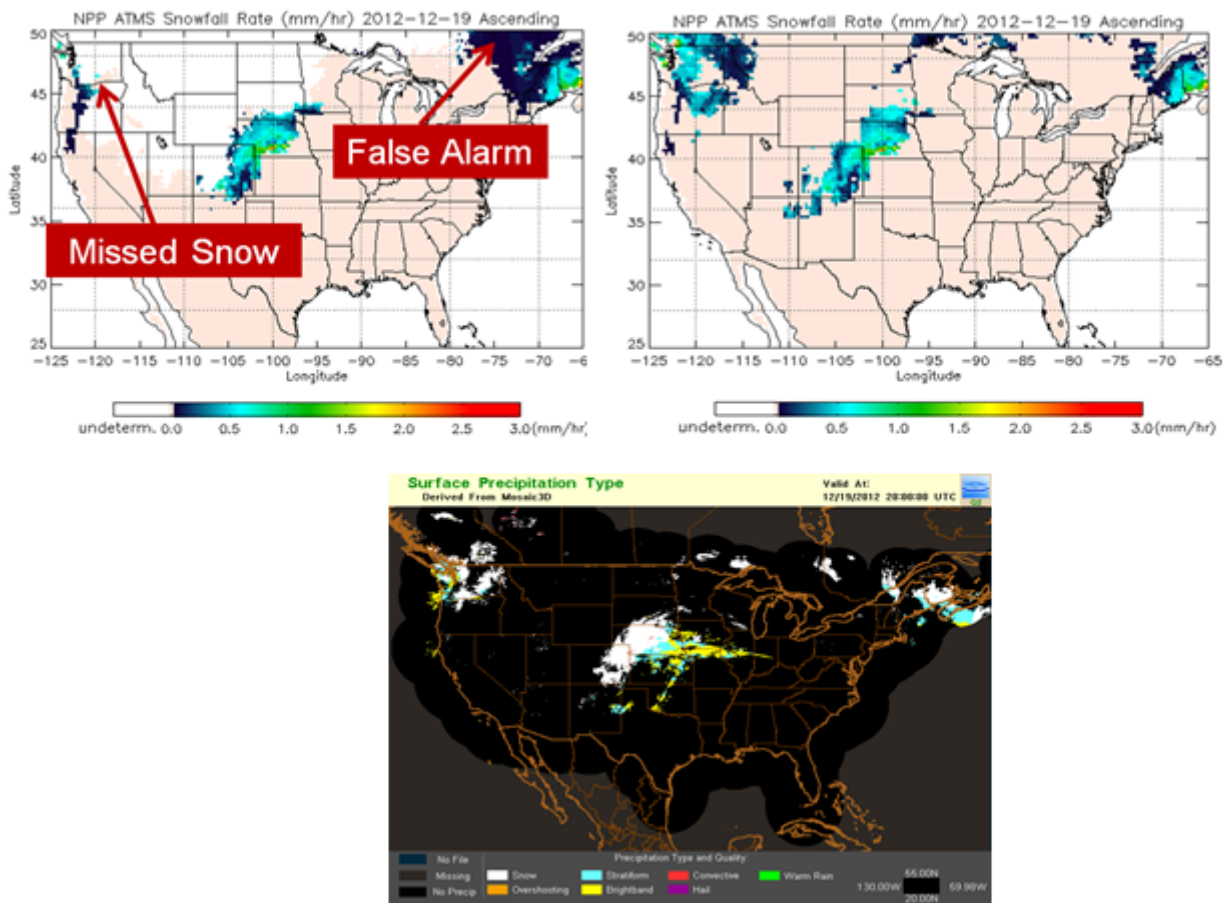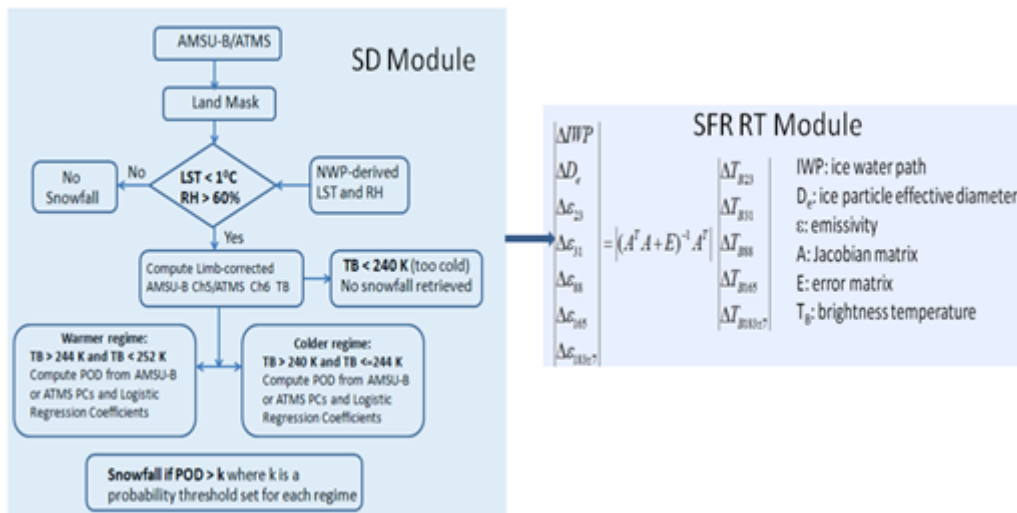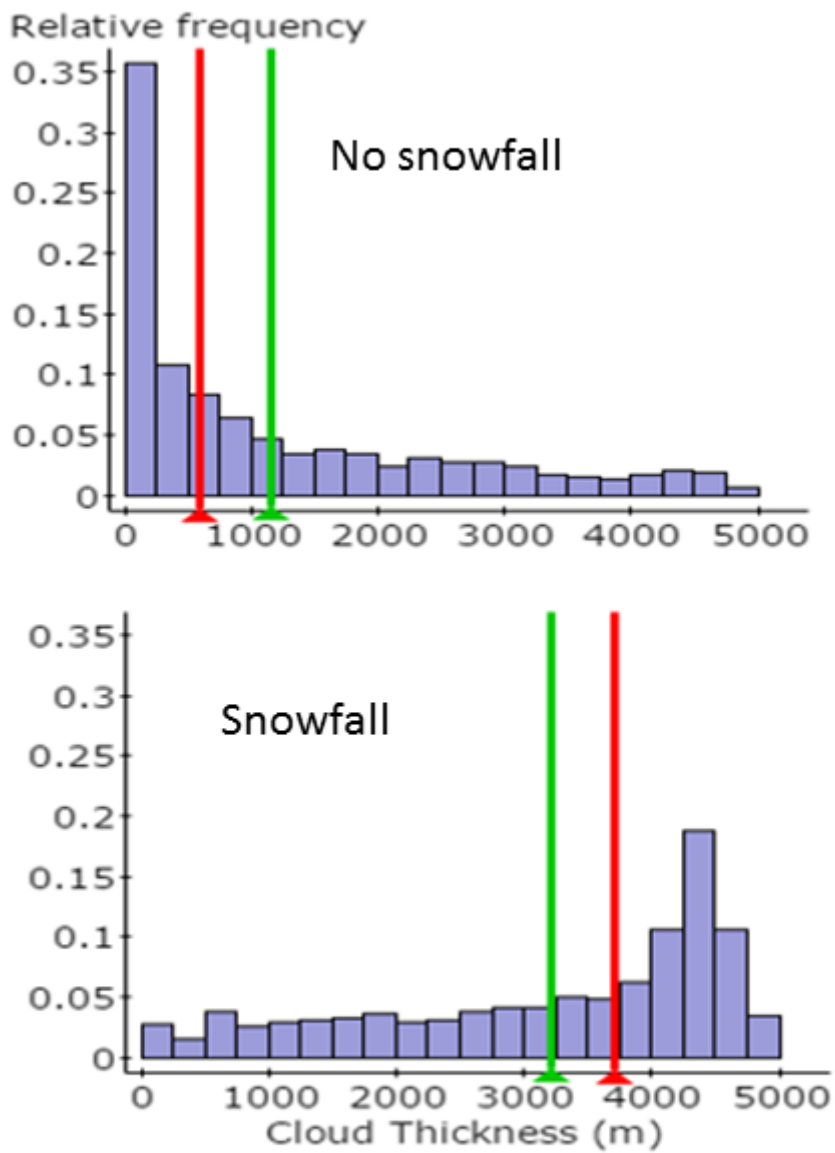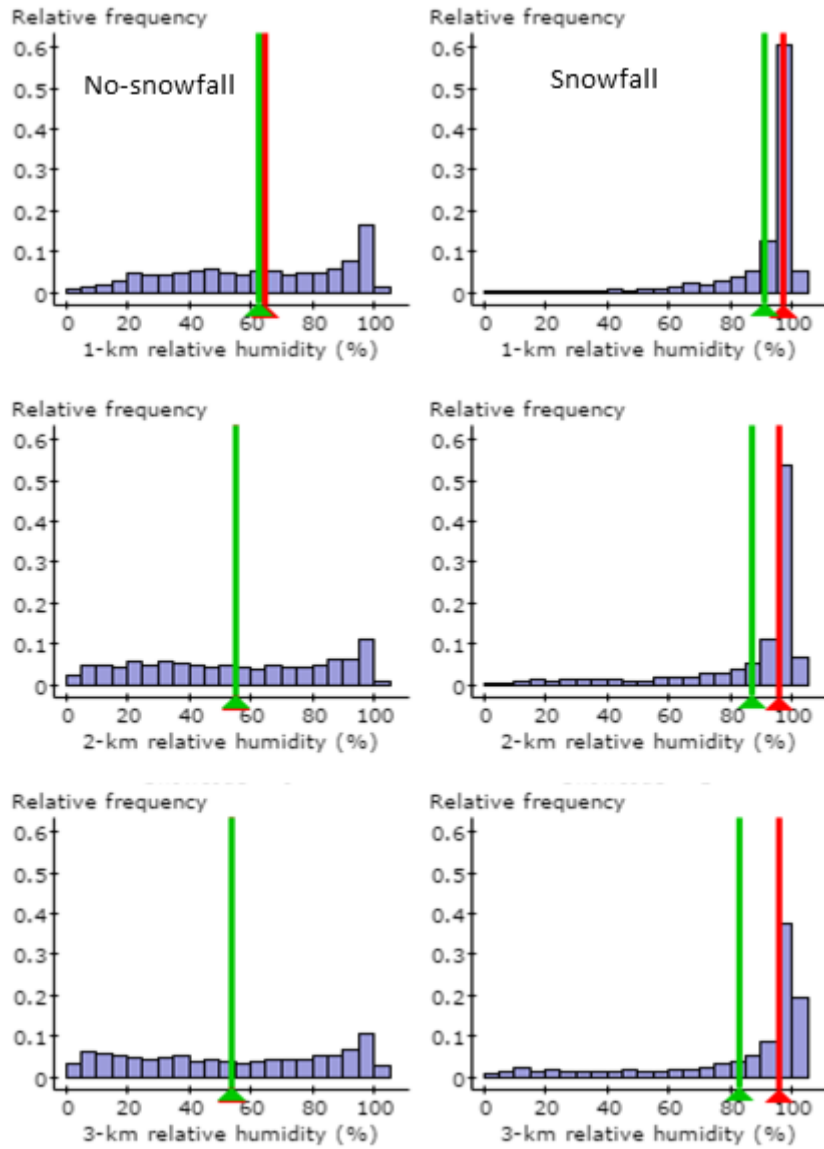| Model/Statistics	FAR	POD	Overall	Heidke
Satellite	0.18	0.41	0.70	0.23
Hybrid	0.11	0.52	0.75	0.44

Figure1.tif
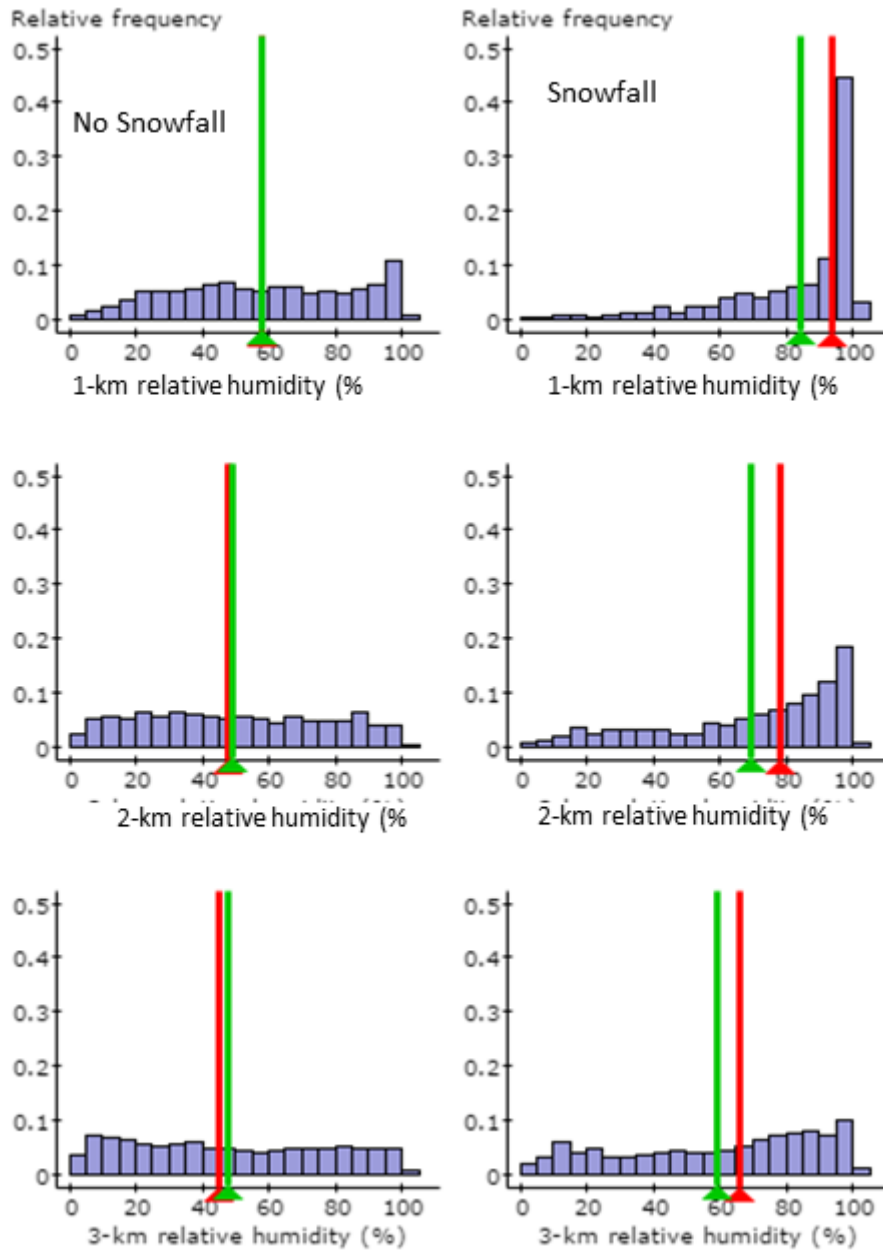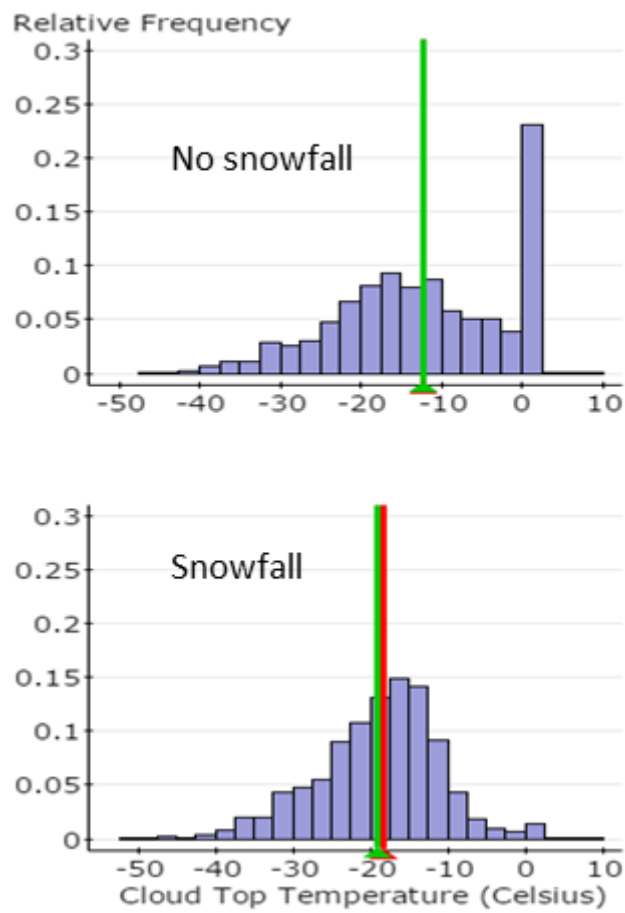
Figure2.tif

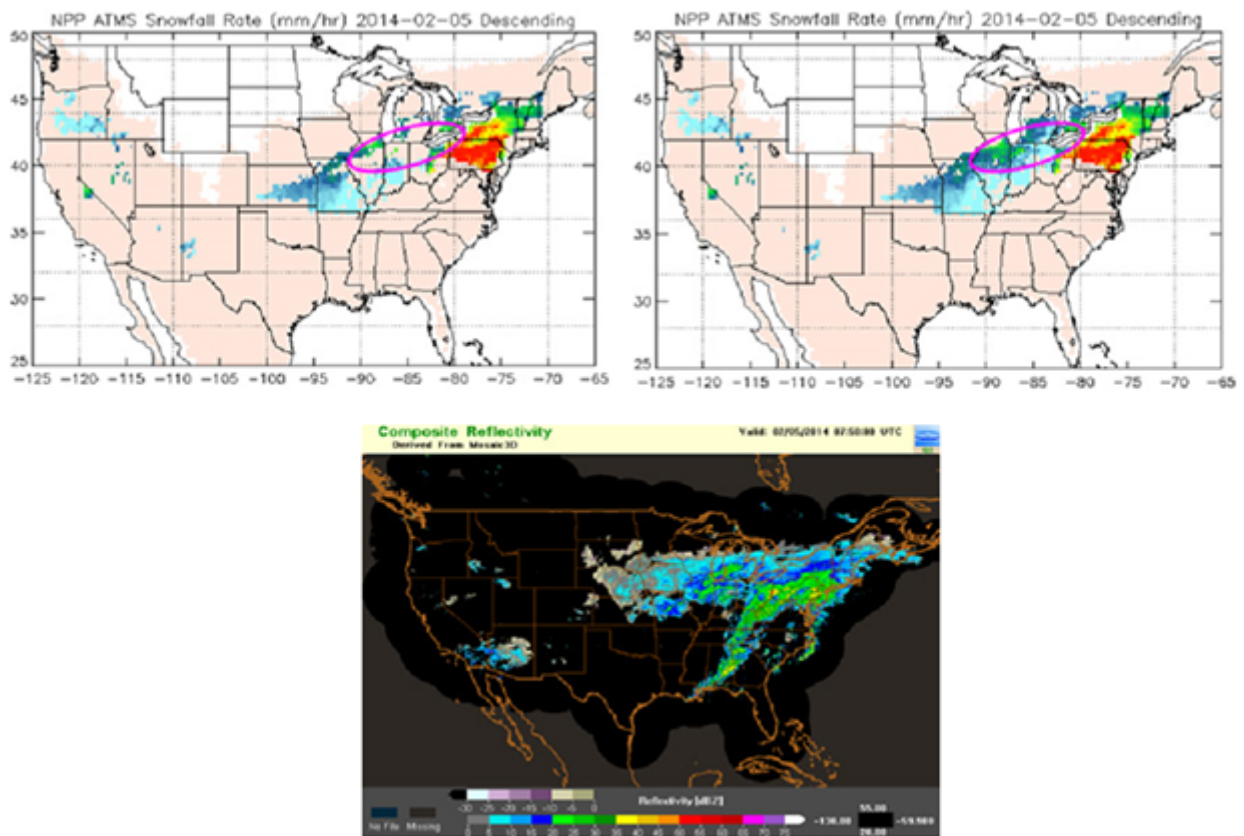Figure3_updated.tif

Figure4_updated.tif
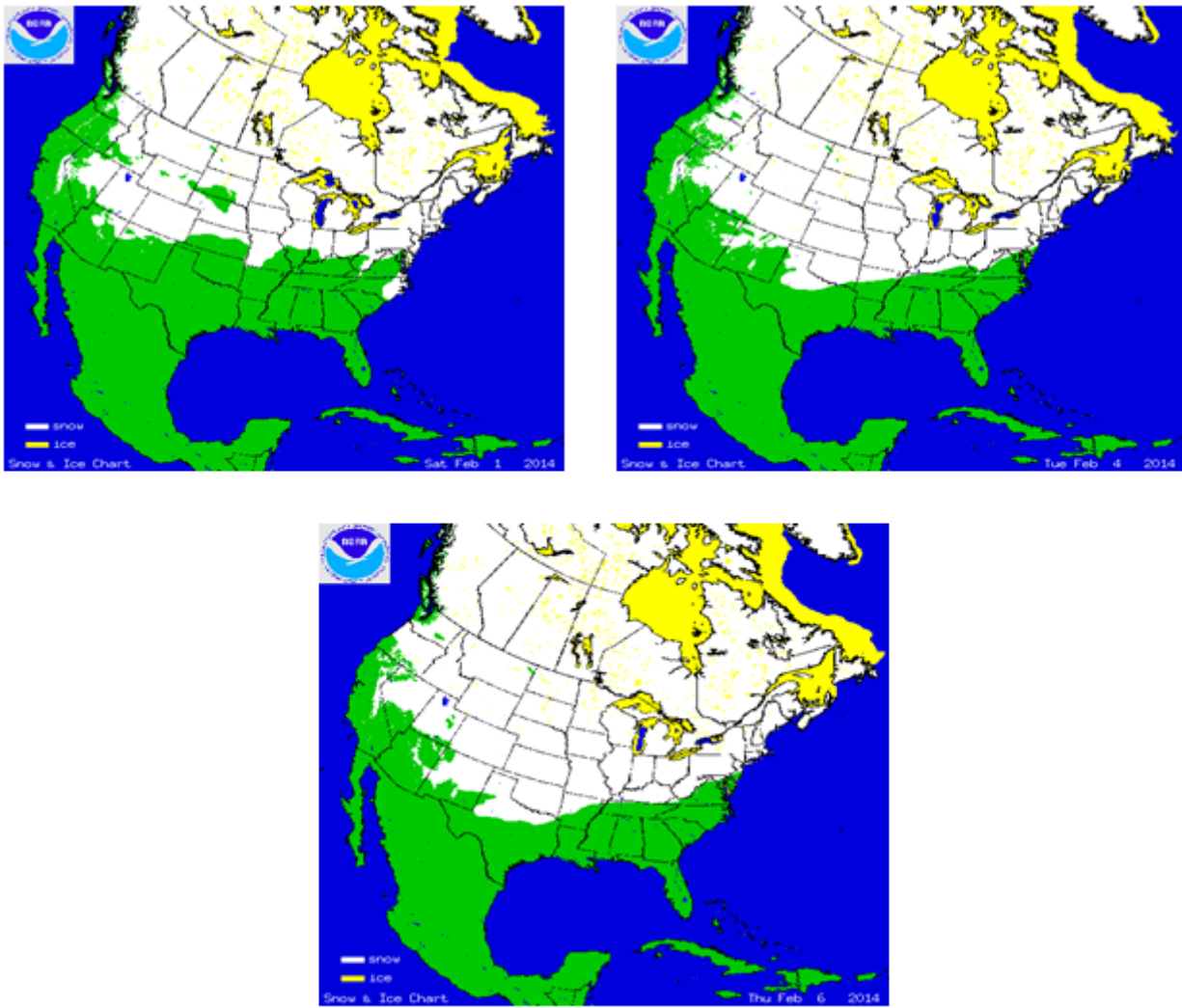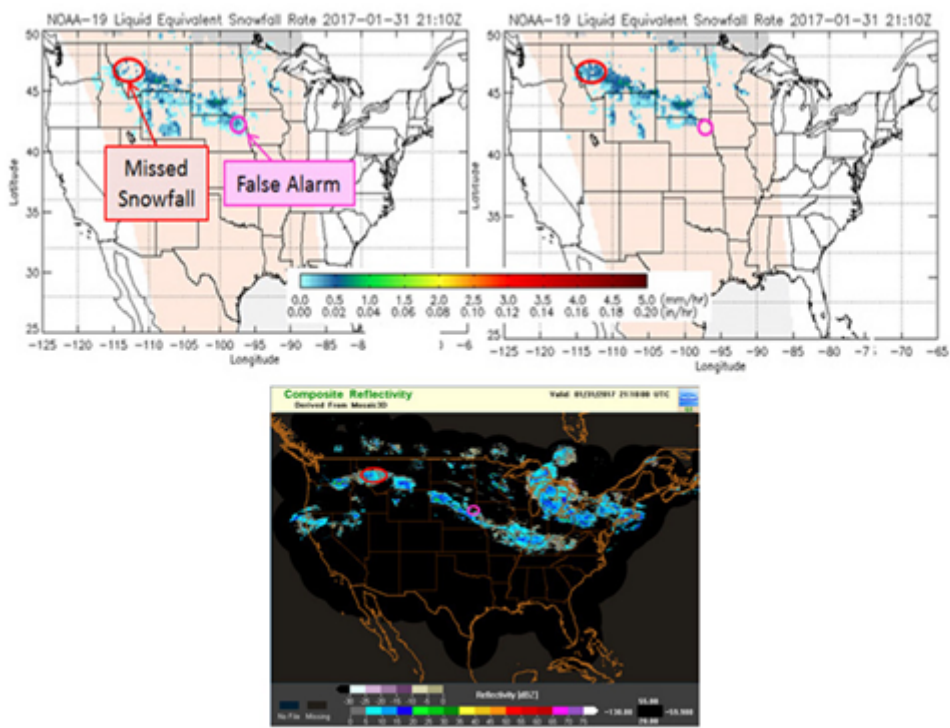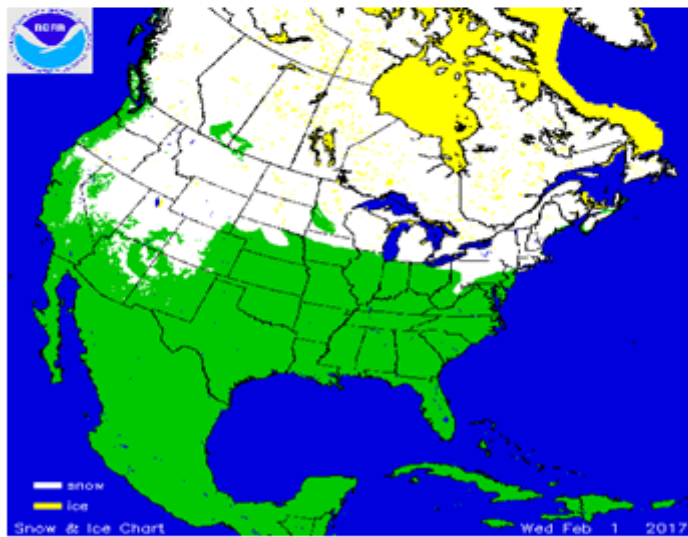
Figure5_updated.tif

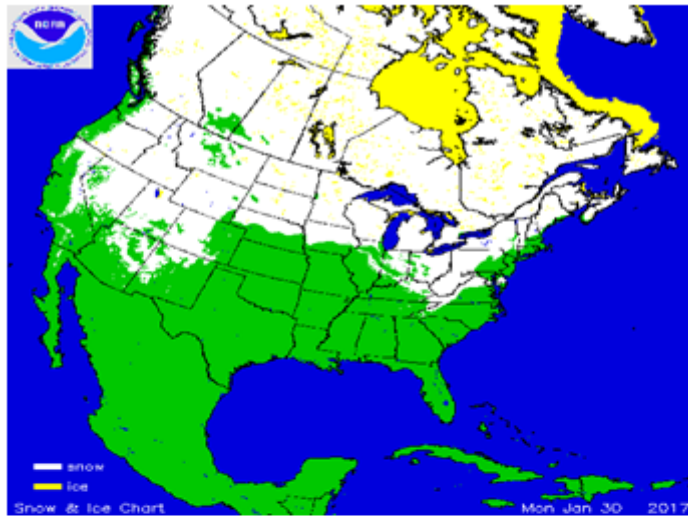Figure6_updated.tif

Figure7.tif

Figure8.tif

Figure9.tif

Figure10.tif

A Hybrid Snowfall Detection Method from Satellite Passive Microwave Measurements and Global Forecast Weather Models

*Cezar Kongoli, Huan Meng, Jun Dong, Ralph Ferraro

➤ The paper presents a hybrid approach to satellite snowfall detection that can improve the performance of satellite-based methods and increase their utility in operational weather and hydrological forecasting;

➤ It presents an analysis and new insights into modeled meteorological variables that are related to snowfall occurrence;

➤ It presents a weather-based snowfall detection algorithm using modeled data from a global weather forecast system.
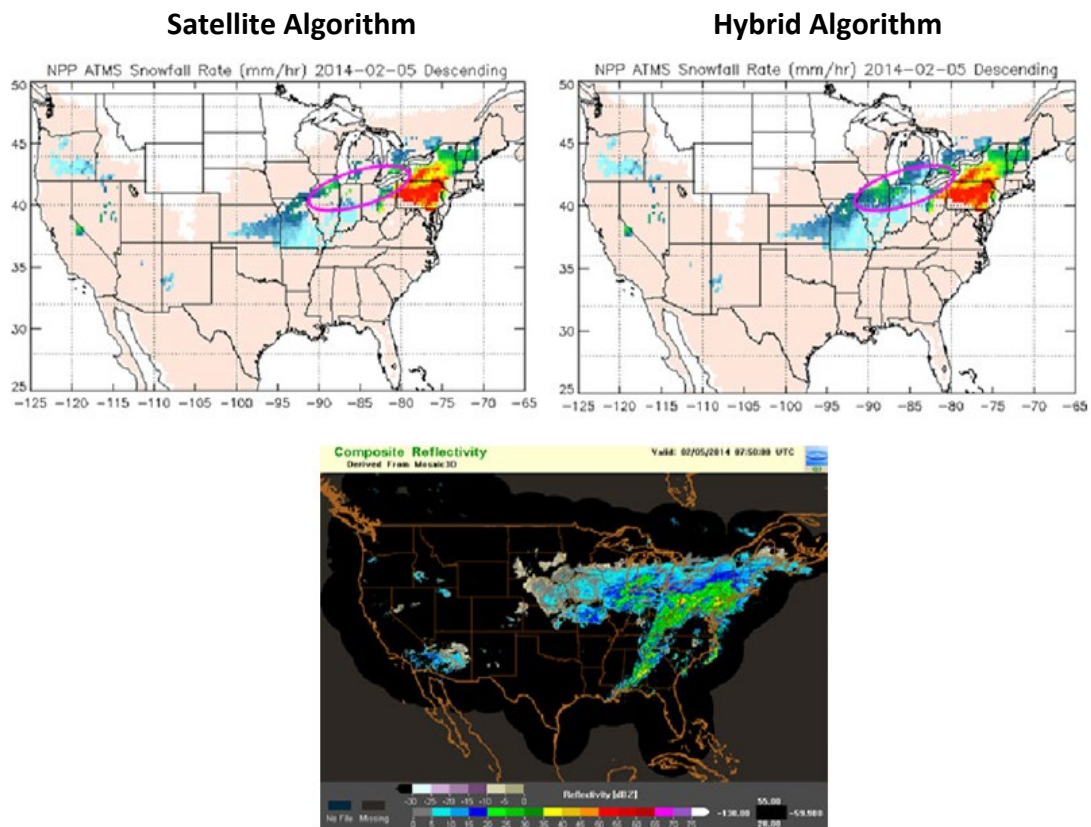
Satellite Algorithm   **Hybrid Algorithm**

Table I: ATMS and AMSU/MHS channel characteristics

ATMS				AMAU[+]	
Ch.	Frequencies (GHz)	Predicted NE"T (K)*	Nadir (km)	Ch.	Measured NE"T (K)*
1	23.80	0.28	75	1	0.21
2	31.40	0.35	75	2	0.26
3	50.30	0.42	33	3	0.22
4	51.76	0.31	33	-	-
5	52.80	0.32	33	4	0.14
6	53.596±0.115	0.35	33	5	0.15
7	54.40	0.32	33	6	0.15
8	54.94	0.32	33	7	0.13
9	55.50	0.35	33	8	0.14
10	f_0=57.290344	0.49	33	9	0.24
11	f_0±0.217	0.67	33	10	0.25
12	f_0±0.3222±0.048	0.70	33	11	0.28
13	f_0±0.3222±0.022	1.06	33	12	0.40
14	f_0±0.3222±0.010	1.45	33	13	0.54
15	f_0±0.3222±0.045	2.40	33	14	0.91
16	88.2	0.29	33	16[#]	0.35
17	165.6	0.44	15	17[#]	0.76
18	183.31±7.0	0.34	15	20	0.55
19	183.31±4.5	0.39	15	-	-
20	183.31±3.0	0.48	15	19	0.68
21	183.31±1.8	0.49	15	-	-
22	183.31±1.0	0.62	15	18	0.98

Center frequencies for channels 16 and 17 are 89 and 157 GHz respectively.

*Integration times for ATMS channels 1-16 are one ninth those of AMSU.

+AMSU resolution at nadir is 50 km below 60 GHz ad 15 km otherwise.

Table I: ATMS and AMSU/MHS channel characteristics

	ATMS			AMSU[+]	
Ch.	Frequencies (GHz)	Predicted NEΔT(K)[*]	Nadir (km)	Ch.	Measured NEΔT(K)
1	23.80	0.28	75	1	0.21
2	31.40	0.35	75	2	0.26
3	50.30	0.42	33	3	0.22
4	51.76	0.31	33	-	-
5	52.80	0.32	33	4	0.14
6	53.596±0.115	0.35	33	5	0.15
7	54.40	0.32	33	6	0.15
8	54.94	0.32	33	7	0.13
9	55.50	0.35	33	8	0.14
10	f_0−57.290344	0.49	33	9	0.24
11	f_0±0.217	0.67	33	10	0.25
12	f_0±0.3222±0.048	0.70	33	11	0.28
13	f_0±0.3222±0.022	1.06	33	12	0.40
14	f_0±0.3222±0.010	1.45	33	13	0.54
15	f_0±0.3222±0.045	2.40	33	14	0.91
16	88.2	0.29	33	16[#]	0.35
17	165.6	0.44	15	17[#]	0.76
18	183.31±7.0	0.34	15	20	0.55
19	183.31±4.5	0.39	15	-	-
20	183.31±3.0	0.48	15	19	0.68
21	183.31±1.8	0.49	15	-	-
22	183.31±1.0	0.62	15	18	0.98

Center frequencies for channels 16 and 17 are 89 and 157 GHz respectively.

*Integration times for ATMS channels 1-16 are one ninth those of AMSU.

+AMSU resolution at nadir is 50 km below 60 GHz ad 15 km otherwise.

Table II: Logistic regression coefficients of the weather model

Variable	Estimate	Std. Err.	Zstat	P-value
Cthick	0.000634	1E-05	-43.52	<0.0001
Hum1	0.032457	3E-05	25.19	<0.0001
Hum3	-0.005273	9E-04	35.95	<0.0001
V2	0.584000	8E-04	-6.601	<0.0001
V3	-0.717000	6E-05	10.22	<0.0001
Hum	0.027847	6E-05	-11.41	<0.0001
Intercept	-6.380459	0.002	17.24	<0.0001

Table III. Performance statistics of the weather, satellite and the hybrid snowfall detection model

Model/Statistics	FAR	POD	Overall	Heidke
Weather1	0.17	0.72	0.79	0.55
Weather2	0.12	0.62	0.78	0.52

1 Threshold probability = 0.5
2 Threshold probability = 0.6

Table IV. Performance statistics of the satellite and the hybrid snowfall detection model for a threshold probability = 0.5

Model/Statistics	FAR	POD	Overall	Heidke
Satellite	0.18	0.41	0.70	0.23
Hybrid	0.11	0.52	0.75	0.44