Application of random forest algorithm to detect snowfall from ATMS measurements

Justinas Kilpys[1], Huan Meng[2], Yalei You[3], Jun Dong[3], Ralf Ferraro[2]

[1] *Institute of Geosciences, Vilnius University, Vilnius, Lithuania*
[2] *NOAA/NESDIS/Center for Satellite Applications and Research (STAR), College Park, Maryland, USA*
[3] *University of Maryland/ESSIC/Cooperative Institute for Satellite Earth System Studies (CISESS), College Park, Maryland, USA*

Introduction

Snowfall estimation from space using passive microwave measurements continues to have lower accuracy compared to rainfall detection (Kongoli et. al., 2018). In this study, Random Forest (RF) algorithm was applied to detect the snowfall using Advanced Technology Microwave Sounder (ATMS) brightness temperature measurements. RF algorithm was applied, because it is considered to be insensitive to over-fitting and efficient in handling high data dimensionality and multicolinearity (Belgiua & Drăguţ, 2016).

Data and methodology

Global *in situ,* ATMS data from the years 2015 and 2018 were used for RF snowfall detection model training and validation. Ground observations retrieved from NOAA's NCDC were matched with ATMS observations within 90 min interval and 50 km buffer of the particular ATMS beam. In the next step, information from NOAA's Global Forecast System (GFS) analysis was collocated with the same ATMS *in situ* data pair. Collocated ATMS and GFS variables were used as input predictors in Random Forest clasification and *in situ* observations were used as "ground truth". The variable importance measurement provided by the random forest classifier has been used to identify the most relevant ATMS channels and GFS variables for the snowfall detection. Three different RF algorithms were generated and analyzed in detail:

1. $\mathbf{RF}_{\mathbf{ATMS}}$ – random forest algorithm using 88-183 GHz ATMS data (Channels 16-20).
2. $\mathbf{RF}_{\mathbf{ATMS+GFS}}$ – random forest algorithm tuned for snowfall in warmer weather regime (≥ 10 °C). Predictors: ATMS 165-183 GHz data (Ch. 17-19); GFS variables RH800, and TMP850.
3. $\mathbf{RF}_{\mathbf{ATMS+GFS(cold)}}$ – random forest algorithm tuned for snowfall in colder weather regime (< 10 °C). Predictors: ATMS 165.6 and 183.3±4.5 GHz data (Ch. 17 and 19), GFS variables CWAT, RH800 and TMP900.

Results

Table 1. Validation scores of three different RF snowfall detection methods. POD – Probability of Detection; FAR – False Alarm Rate; HSS – Heidke Skill Score. Scores in blue indicate validation results for colder (< -10 °C) weather regime.

Method	Predictors	Fraction correct	POD	FAR	HSS
Hybrid (Kongoli et al. 2018)	ATMS ch. 16-22 (88-183 GHz) + 6 GFS var.	0.79	0.72	0.17	0.55
RF_{ATMS}	ATMS ch. 16-20 (88-183 GHz)	0.80 0.59	0.90 0.93	0.25 0.45	0.61 0.19
$RF_{ATMS+GFS}$	ATMS ch. 17, 18 and 19 (165-183 GHz) + 2 GFS var.	0.83 0.65	0.90 0.86	0.21 0.39	0.66 0.30
$RF_{ATMS+GFS(cold)}$	ATMS ch. 17 and 19, (165-183 GHz) + 3 GFS var.	0.83 0.71	0.89 0.82	0.21 0.33	0.65 0.41

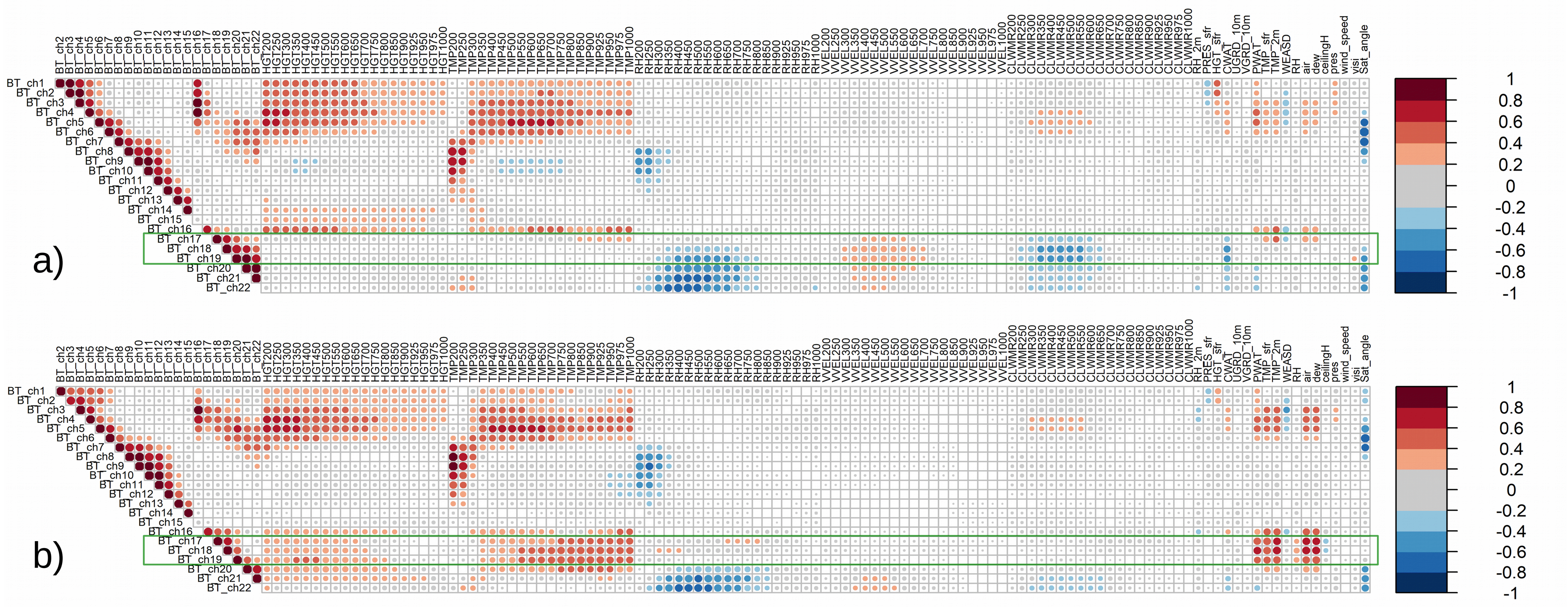

Figure 1. Correlation matrices between ATMS brightness temperatures and GFS variables during the snowfall events: a) when air temperature is between 5 °C and -10 °C; b) when air temperature is between lower than -10 °C.

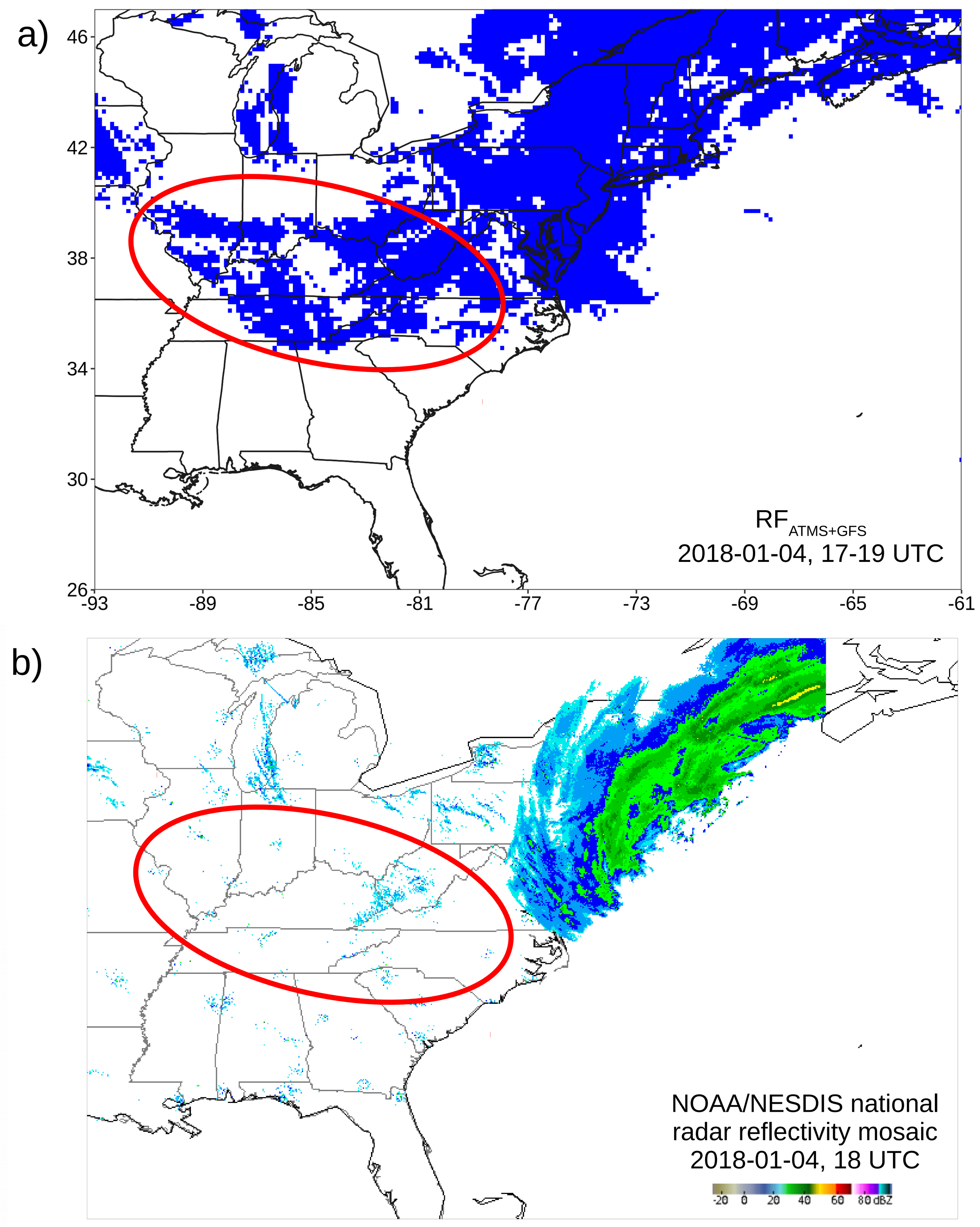

Figure 2. Snowfall during the 4th January 2018: a) snowfall based on $RF_{ATMS+GFS}$ algorithm; b) precipitation intensity based on radar reflectivity.

For all analyzed RF algorithms best predictors for snowfall detection were high-frequency ATMS channels 17, 18 and 19 (165-183 GHz). Using only ATMS channels (RF_{ATMS}) probability of detection (POD) was 0.90, but the false alarm rate was also high (FAR = 0.25) (Table 1). Using $RF_{ATMS+GFS}$ method it was possible to reduce FAR to 0.21 and improve overall accuracy to 0.83. $RF_{ATMS+GFS}$ algorithm evaluation showed that in cold weather conditions (air temperature < -10 °C) ATMS data is inadequate to distinguish snowfall from no precipitation cases. Wang et. al. (2013) reported that at bellow -10 °C the amount of liquid water in the clouds are very dependent to cloud type and presence of supercooled could mask snowfall signal in ATMS data. In this study it was determined that in colder conditions ATMS 165-183 GHz brightness temperatures correlates better with surface parameters rather than with variables in middle troposphere (Fig. 1). $RF_{ATMS+GFS(cold)}$ algorithm was tuned for colder weather and it was determined that atmospheric cloud water content (CWAT) and air temperature at lower troposphere (800-900 hPa) are better snowfall predictors than ATMS brightness temperatures.

Reference:

Belgiua M. and L. Drăguţ (2016). Random forest in remote sensing: A review of applications and future directions. *J. Photogr. and Rem. Sens.*, 114: 24-31.

Kongoli C., H. Meng, J. Dong, and R. Ferraro (2015). A snowfall detection algorithm over land utilizing high-frequency passive microwave measurements—Application to ATMS. *J. Geophys. Res. Atmos.*, 120: 1918–1932.

Kongoli C., Meng H., Dong J., Ferraro R. (2018). A hybrid snowfall detection method from satellite passive microwave measurements and global forecast weather models. *Q. J. R. Meteorol. Soc.*, 144: 120–132.

Wang Y., Liu G., Seo E.K. and Fu Y. (2013). Liquid water in snowing clouds: Implications for satellite remote sensing of snowfall. *Atmospheric Research*, 131: 60–72.

Conclusions

Although snowfall detection in the cold weather conditions using ATMS sensor data is still challenging, random forest classification has demonstrated improvement over the logistic regression and hybrid algorithms previously used for snowfall detection (Kongoli et. al., 2015; Kongoli et. al., 2018).