Classifying precipitation from GEO Satellite Observations: Diagnostic Model

Shruti A. Upadhyaya[1,*], Pierre-Emmanuel Kirstetter[2,3,4,5,*], Robert J. Kuligowski[6], Maresa Searls[2]

[1] Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

[2]School of Meteorology, University of Oklahoma, Norman, Oklahoma

[3]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, Oklahoma

[4]Advanced Radar Research Center, University of Oklahoma, Norman, Oklahoma

[5]NOAA/National Severe Storms Laboratory, Norman, Oklahoma

[6]NOAA/NESDIS/Center for Satellite Applications and Research, College Park, Maryland

[*]*Corresponding authors:* Pierre-Emmanuel Kirstetter (pierre.kirstetter@noaa.gov);

Shruti A. Upadhyaya (shruti.a.upadhyaya-1@ou.edu)

Funding Information:

Funding for this research was provided by the GOES-R Series Risk Reduction program, which provided support to the Cooperative Institute for Mesoscale Meteorological Studies at the University of Oklahoma under Grant NA16OAR4320115. P. Kirstetter acknowledges support from NASA Global Precipitation Measurement Ground Validation program under Grant NNX16AL23G and Precipitation Measurement Missions program under Grant 80NSSC19K0681

Key Points:

- A diagnostic analysis on a machine learning precipitation classification model is performed
- Model uses a comprehensive set of predictors derived from GOES-16 satellite observations and numerical weather predictions (NWP).
- Brightness temperature textures and inter-band differences are efficient predictors.
- Environmental predictors such as CAPE, lapse rate, relative humidity, and precipitable water, bring complementary information

Abstract:

Improvements in remote sensing capability and improvements in artificial intelligence have created significant opportunities to advance understanding of precipitation processes. While highly advanced Machine Learning (ML) techniques improve the accuracy of precipitation retrievals, how these observations contribute to our understanding of precipitation processes remains an underexplored research question. In a companion manuscript, a precipitation type prognostic ML model is developed by deriving predictors from the Advanced Baseline Imager (ABI) sensor onboard Geostationary Observing Environmental Satellite (GOES)-16. In this study, these predictors are linked to different precipitation processes. It is observed that satellite observations are important in separating Rain and No-Rain areas. For stratiform precipitation types, predictors related to atmospheric moisture content, such as relative humidity and precipitable water, are the most important predictors, while for convective types, predictors such as 850-500hPa lapse-rate and Convective Available Potential Energy (CAPE) are more important. The diagnostic analysis confirms the benefit of spatial textures derived from ABI observations to improve the classification accuracy. It is recommended to combine the heritage water vapor channel T6.2 with the IR T11.2 channel for improved precipitation classification. Overall, this study provides guidance to atmospheric and remote sensing scientists on a large array of predictors that can be used from geostationary satellites and multispectral sensors for precipitation studies.

Keywords: GOES-16, Numerical Weather Prediction, Precipitation, Machine Learning, Classification, Geostationary Satellites, Interpretable Machine Learning

1. Introduction

Mapping precipitation from space has been well recognized for over five decades. The high spatial and temporal resolutions and the improved spectral information from the new generation of Geostationary Earth Orbit (GEO) satellites provide opportunities to improve our understanding of clouds and precipitation processes, particularly the characterization of convective processes that are at the core of severe and extreme weather (National Academies of Sciences, Engineering, and Medicine 2018). This follows priorities identified in the 2017-2027 decadal survey for Earth Sciences and Applications from Space (ESAS 2017) by the Earth science community. Identifying these processes also helps improve precipitation retrieval accuracy from GEO sensors (Grams et al., 2016; Thies et al., 2008).

Several studies attempted to identify precipitation processes from space with previous GEO sensors. Yet, significant challenges exist due to the indirectness in the information related to cloud top heights obtained from the Visible (VIS)/Infrared (IR) regions of the electromagnetic spectrum (Kidder and Vonder Haar; 1995). Some of the early studies used a single VIS channel with the hypothesis that clouds producing rain have higher optical thickness and appear brighter than non-raining clouds in VIS images (Follansbee, 1973; Kilonsky and Ramage, 1976). More studies focused on using IR channels since they are available during both day and night. Rain-producing clouds are often associated with cold cloud tops in IR brightness temperature images (BT; Arkin, 1979). However some clouds do not substantiate this hypothesis, e.g. stratus clouds appear bright in VIS images but do not produce as much rain as convective systems, and high-level cirrus clouds appear cold in IR image but do not produce rain (Kidder and Vonder Haar, 1995). Bi-spectral techniques involving both VIS and IR channels were designed to identify different precipitation systems (Lovejoy and Austin, 1979; Tsonis and Isaac, 1985). To adapt bi-spectral techniques to day and night retrievals, the water vapor (WV) absorption channel is used in current algorithms (Upadhyaya et al., 2014; Tao et al., 2018). For a summary of early techniques of satellite precipitation detection and quantification, readers are referred to Barret and Martin (1981). Over the last three decades, the quality, resolution, and information captured by GEO sensors have significantly improved; e.g. from two channels with the VISSR (Visible-Infrared Spin Scan Radiometer) sensor onboard GOES-1 to sixteen channels with the ABI (Advanced Baseline Images) onboard the latest-generation GOES-R satellite. Satellite precipitation algorithms (SPA) have evolved accordingly, with improved ability to identify precipitation processes through the use of multiple spectral channels. Yet, current SPA falls short of using the full set of available GEO IR observations and likely underutilize the information content.

A popular use of multiple spectral channels involves deriving spectral and textural predictors from one or a combination of channels (Note: a texture represents spatial characteristics that account for neighbourhood information). The most common approach for combining channels is to take the difference between brightness temperatures (further referred to as BTD). For example, Tjemkes et al. (1997) showed that the difference between the IR window channel (11.2 μm in ABI) and the water vapor (WV) absorption channel (6.2 μm in ABI) can be used to separate overshooting cloud tops and cirrus clouds. Radiative transfer simulations revealed that the difference between two infrared channels (e.g., 11.2 μm and 8.4 μm) provides information about cloud phase (Baum and Platnick, 2006; Giannakos and Feidas, 2013). This cloud phase detection can be further improved by comparing two difference indices (i.e., difference of BTDs: D-BTD). The most commonly used D-BTD index involves the difference between the 8.5 μm-11.2μm and 11.2μm-12.3μm BTD values (So and Shin, 2018). Single channel indices include textures as the representations of the visual characteristics of a surface (Mohanaiah et al., 2013). Texture indices derived from several individual channels are found to be useful at all stages of the precipitation retrieval (Ba and Gruber, 2001; Kuligowski, 2016; Hong et al., 2004; Giannakos and Feidas, 2013; Tian et al., 1999). To complement the cloud top information coming from GEO sensors, alternate sources of information can come from Numerical Weather Prediction (NWP) model data and topographic information (Grams et al., 2014; Min et al., 2018; Upadhyaya et al., 2016). More indices can be derived and analysed, such as combinations of BTDs and D-BTDs and textures from BTDs and D-BTDs.

In a companion manuscript (Upadhyaya et al., 2021; hereafter referred to as Part I) of this study, a comprehensive set of indices from GOES-16 ABI observations were derived and matched with surface precipitation types from the Ground Validation Multi-Radar/Multi-Sensor (GV-MRMS) system (Zhang et al., 2016; Kirstetter et al., 2018) across the conterminous United States (CONUS). A machine learning (ML) based Random Forest (RF) model is built to explore various new indices and prognose the identification of precipitation types. In this study (hereafter referred to as Part II), the focus is on peering into the developed ML model and its interpretation. Major identified research gaps and motivating research questions are discussed as follows.

1. While several categories of indices are already proposed in the literature, many more that can be derived have not been examined. The operational products, such as the Self-Calibrating Multivariate Precipitation Retrieval (SCaMPR; Kuligowski et al. 2016), and Precipitation Information from Remotely Sensed Information using Artificial Neural Networks - Cloud Classification System (PERSIANN-CCS; Hong et al., 2004), do not make exhaustive use of these indices. For the first time to our knowledge, a framework is proposed to perform consistent and systematic analyses on satellite-based indices for precipitation detection and classification of types. In order to make recommendations for science and operational use, the first research question under investigation is: **what is the impact of the different satellite predictors on classification accuracy?**

2. It has been long recognized that the indirectness in the information from IR sensors in detecting and retrieving precipitation can be complemented by environmental predictors from NWP models. For example, operational products, such as SCaMPR, use environmental predictors such as Relative Humidity (RH) to mitigate the overestimation (underestimation) of GEO retrieved rainfall in dry (wet) environments. In part I, additional environmental predictors, such as Convective Available Potential Energy (CAPE), vertical lapse rate of temperature, and wind shear are utilized. This study investigates the significance of environmental predictors with motivation to address the research question: **What is the relative impact of satellite-based predictors compared to environmental predictors?**
3. The global operational satellite precipitation products, such as PERSIANN-CCS, and Integrated Multi-satellitE Retrievals for the Global Precipitation Mission (IMERG; Huffman et al., 2015), utilize only one channel from GEO sensors. As GEO satellites uniquely provide the longest period of global precipitation observations (over more than four decades), there is a need to assess the accuracy they allow for reanalyses, along with highlighting the progress made possible with recent sensors. This motivates to investigate the research question: **How do the multi-spectral channels from the new generation of GEO sensors compare to historical benchmarks that use only legacy channels?**
4. With improvement in computational power and growth in artificial intelligence, big data from GEO sensor observations can feed the latest generation ML methods (Tao et al., 2018; Min et al., 2018; Meyer et al., 2016) to model their complex interactions. While highly advanced ML techniques improve the overall accuracy of precipitation retrievals, these data-driven models do not connect predictors with processes. How much each predictor contributes to our understanding of precipitation processes

remains an underexplored research question. The complex nature of ML algorithms, however, makes it challenging to physically interpret these models and indices. Recently, several tools have been developed to peer into these models, thus making it possible to address this major gap in understanding: **Which predictors are contributing to different precipitation types?**

Section 2 of this paper describes the data sets used and how they were pre-processed prior to use by the RF model. Section 3 outlines the RF model and the experiments that were conducted, and Section 4 describes the results of these experiments, followed by concluding remarks in Section 5.

2. Data and Pre-processing

2.1. GV-MRMS

Data for the Global Precipitation Measurement (GPM) Ground Validation (GV-MRMS; Kirstetter et al., 2012, 2014, 2018) based on MRMS (Zhang et al., 2016) is used as a reference. The study period is over summer 2018 (June, July, August, and September) and the study area is the conterminous United States (CONUS) with latitude bounds [25°N 50°N] and longitude bounds [125°E 67°E]. The spatial and temporal resolutions of GV-MRMS are 0.01°× 0.01° and 30 min, respectively. The reference product is the surface precipitation type derived from MRMS. Precipitation types relate to different precipitation processes and drive the MRMS precipitation quantification. Identifying precipitation types is also key for quantification from the GOES ABI sensor.

The precipitation types as classified in GV-MRMS are 1) Warm Stratiform rain, 2) Cool Stratiform rain, 3) Convective rain, 4) Tropical Stratiform/Mix, 5) Tropical Convective/Mix, 6) Hail, 7) Snow, and 8) No-Precipitation. This empirical classification is based on several radar and NWP based environmental variables with adaptable thresholding parameters (Zhang et al., 2016). Part I of the study discussed the potential and limitations in re-creating the same classification from ABI observations, and the advantages of a probabilistic classification (provided by a Random Forest machine learning approach) over a deterministic classification.

Note that radar estimates have their own uncertainties, such as beam blockage, non-precipitation echoes, and bright bands (Zhang et al., 2016), which impacts precipitation classification. In order to use only reliable observations as reference, a Radar Quality Index (RQI) is used for quality control (QC) purposes, with a threshold of 98% for most precipitation types and a lower threshold of 90% for Hail due to lower sample sizes. Due to very limited sample size in summer, the Snow type is not considered in this analysis. For further details on RQI, the readers are referred to Zhang et al. (2011, 2016). Also note that only ~35% of total observations satisfy the RQI criteria across CONUS, out of which 72% are from Eastern CONUS and the remaining 28% are from Western CONUS (again highlighting the significance of GOES-16 observations over complex terrain).

2.2. ABI observations and derived predictors

This study uses the five parallax adjusted GOES-16, ABI channels (Channel 8: 6.2μm, 10: 7.3μm, 11: 8.5μm, 14: 11.2 μm, and 15: 12.3 μm) used by SCaMPR (Kuligowski et al., 2016). The parallax adjustment is based on trigonometry. Based on the known satellite location (in 3 dimensions) and Earth's radius, the cloud-top height is estimated by using the ABI band 14 temperature as a proxy for the cloud-top temperature (a reasonable assumption for precipitating clouds) and by comparing it to the spatially interpolated temperature-height profile from GFS model forecasts.

Six categories of predictors are derived from these observations and are listed in Table 1. Category 1, BT, is brightness temperatures from the individual ABI channels. Category 2 includes the differences between two channel BTs, called BTDs. Category 3 corresponds to the difference between BTDs, represented as D-BTDs. Category 4 consists of five types of Textures (Te), namely "mean", "variance", "homogeneity", "contrast", and "entropy", that are derived from the Grey Level Co-occurrence Matrix (GLCM; Haralick et al., 1973) for all predictors from Categories 1 - 3. Category 5 is the Satellite Zenith Angle (Ze). In total, 249 predictors are derived from satellite observations. These categories are discussed in Section 1. More details are provided in Section 4, and interested readers are also referred to Part I.

The spatial and temporal resolutions of ABI are, respectively, 2 km at nadir and 15 min for the full disk (reduced to 10 min after the study period). To match the datasets, GV-MRMS is aggregated to the ABI spatial resolution. Two additional QCs are applied to the data. First, a minimum percent coverage of precipitating pixels from GV-MRMS within one coarser ABI grid cell (Prain) is applied to mitigate the influence of partially precipitating grids on accuracy (following Upadhyaya et al., 2020). Only grids with Prain greater than 95% or less than 5% are used for analysis. Secondly, only grids with homogeneous precipitation types are targeted; i.e. grids with at least 98% of the same precipitation type are used for analysis for most types, except 90% for convective and Tropical Convective/Mix precipitation types, and 80% for Hail. The sample size of the data set after all the QC is given for each month in Table 2. Each monthly dataset is broken down into training (70%) and testing (30%) for representative samples from the entire summer season. The independence between the training and testing datasets with respect to storms being placed in both was checked, and its impact on the results was found to be negligible. It is also ensured that events in the training and testing datasets have good spatial and temporal representations. Since the training data is highly unbalanced across precipitation types, under sampling is applied on the most populated types (e.g. Warm Stratiform) to create a balanced dataset and optimize the RF model training. For evaluation, the testing dataset is used.

2.3. Numerical Weather Prediction (NWP) model based environmental predictors

To complement the cloud top information from ABI observations and provide information about mid- and low-level environmental conditions, NWP-based predictors are used. The next-generation hourly updated assimilation and model forecast cycle Rapid Refresh (RAP), part of the NOAA/National Centers for Environmental Prediction (NCEP) operational suite since May 2012, is used (Benjamin et al., 2016). In total 19 predictors adopted from Grams et al. (2014) are used (Table 3) along with the previously-described satellite-based predictors.

3. Summary of Prognostic Modelling and Proposed Experiments for Diagnostic Analysis

3.1. Summary Part I: Prognostic Model

As mentioned in previous sections, Part I of this article focuses on prognostic modelling; i.e., the design, training, and assessment of a machine learning based model. Comprehensive sets of predictors are derived and tested, many of which are derived for this particular application for the first time in the literature. The main objective is to evaluate the potential for the new generation of GEO ABI satellite observations to discriminate precipitation processes or types, and to quantify the accuracy that can be achieved for each precipitation type. The important question is addressed: Can the model be applied to real case events? The overall results in the form of a normalized contingency matrix is shown in Table 4. The analysis showed that the ML model has an overall classification accuracy greater than 75%, with particularly good ability at delineating Precipitation from No-Precipitation. It also displays good accuracy in terms of detecting precipitation types, such as Cool Stratiform, Warm Stratiform, and Hail. Tropical types, Tropical Stratiform/Mix and Tropical Convective/Mix, and Convective type are more challenging.

RF is by design a probabilistic classifier. RF computes the probability of a sample belonging to each precipitation type before the dominant probability class is assigned in a deterministic way. Part I highlights the need to use probabilities to objectively handle precipitation types identified with various levels of certainty from a user perspective. In particular, an "uncertain" type can be defined using the predicted probabilities. RF models also compute feature/predictors importance. Overall, it is shown that environmental predictors have higher importance than satellite predictors. Feature importance can be used to design more parsimonious models. The analysis indicates how the number of predictors can be reduced from 260 to 68 without significantly compromising accuracy. The present study focuses on understanding the significance of different predictors for each precipitation type by analyzing the structure of the RF model.

3.2. Experimental set-ups for diagnostic modelling

An RF model was built and evaluated by using the scikit-learn framework (https://scikit-learn.org/). The list of experiments carried out to diagnose the RF model and understand the impact of different predictors is given in Table 5. The need for and background of these analyses is discussed in their respective results sections. Note that all these experiments are trained with 70% balanced data as given in Table 2 and statistics are computed using the remaining 30% of validation data.

Analyses 1-3 are designed similarly to Part I but each with different sets of predictors. The first analysis runs experiments using various satellite predictors as listed in Table 1 to get a deeper understanding on the significance of BT, BTD, D-BTD, Te, and Ze. For the second analysis, three experiments involve satellite predictors only, environmental predictors only, and satellite and environmental predictors altogether. For the third analysis, RF experiments with predictors based on only one IR channel (T11.2) and two channels (T11.2, T6.2) that simulate historical sensors are compared with estimation using three channels (T11.2, T6.2, T12.3) and all five channels.

Analysis 4 focuses on estimating important predictors for each precipitation type by extracting the contribution of all predictors and by ranking them according to their contribution to estimated probability. The Treeinterpreter python library (Andosa, 2015) is used to interpret the RF models. For a given sample and precipitation type, it extracts the estimation path of the forest from root to the leaf and the contribution of each predictor.

As an example, a hypothetical sample output with few predictors and decisions is shown in Fig. 1. Assuming that the hypothetical sample follows the red line path, then the predicted probability for the sample to belong to the tree class is shown in the equation in Fig. 1, where Mean is the value at each root node and other variables are contributions from subsequent tree nodes. The predicted probability of a sample belonging to the tree class is 0.83, with the highest contribution coming from predictor "$BT_1 - BT_2$". Similar estimates and contributions can be retrieved for each estimated class (precipitation types), which adds up the probability of a sample to belong to each type to a total of 1. Inherently, RF models build these probability estimation forests for each precipitation type, and the sample is assigned deterministically to the dominant class. In the present study, Treeinterpreter is run separately for each precipitation type with randomly selected sub-samples from the validation dataset that are correctly classified. Since it is impractical to show the contribution from each of the 260 predictors, the highest contributing predictors in a particular precipitation type are identified by ranking their average contributions to the tested sample. Then, the distribution of contribution of these high contributing predictors from all samples is analysed.

Quantifying and analysing the significant predictors for each precipitation type (Analysis 4 in Table 5) allows an in-depth assessment of falsely classified precipitation types.

4. Results and Discussions

4.1. Significance of different categories of satellite predictors

RF models are developed with individual categories of satellite predictors (Table 1) to understand which categories of predictors are significant and for what precipitation types. Figure 2 shows accuracy for each precipitation type and for each of the categories of satellite predictors.

It can be observed that all categories of satellite predictors are valuable for the identification of different precipitation types. Overall, models built on texture based predictors (Te) show higher accuracy than other models built on other individual categories, especially for the Warm Stratiform type. For the Hail and No-Precipitation types, all models show similar performances, except for the model built with the Zenith Angle predictor. The accuracy of the Zenith Angle model ranges from 20-25%, which is generally lower than for the other precipitation type models. Yet, for the Tropical-Convective Mix and the Cool Stratiform types, the Zenith Angle only model shows similar skill to the models that use the satellite and / or environmental predictors. This should not be entirely surprising given that the climatology of occurrence of some of these precipitation types has a strong spatial dependence; for instance, the Cool Stratiform type is generally observed in the northern and western portions of the ABI viewing area whereas Tropical-Convective Mix is much more prevalent closer to the GOES-16 subpoint. However, it should also be noted that spatial climatology of precipitation classes are a part of the MRMS classification system and thus may also be contributing to the strength of the relationship between GOES-16 zenith angle and precipitation class. BTD and D-BTD based models generally show very similar accuracy with marginally higher scores for D-BTD. Since the best results are consistently achieved using models combining all satellite predictors, it appears that the precipitation type classification benefits somewhat from their inclusion.

Except for No-Precipitation, the improvement is in the range of 5-10% for all other precipitating types.

4.2. Comparative performance analysis of satellite based predictors and environmental predictors

In Part I of the manuscript, it is observed that environmental predictors display an overall higher feature importance than satellite predictors. This section aims at comparing the significance of satellite based predictors to environmental predictors for each precipitation type. It can be observed from Fig. 3 that models based on satellite predictors show higher accuracy for convective types (Hail, Convective and Tropical Convective/Mix) and No-Precipitation, while classification based on environmental predictors display higher performances for stratiform types (Cool Stratiform, Warm Stratiform, and Tropical Stratiform/Mix). It confirms results from previous studies that convective precipitation can be detected well by GEO sensors, whereas the identification of shallow clouds can be improved with numerical model fields (e.g. Ebert et al., 2007). As in the previous section, models combining satellite and environmental predictors improve the accuracy, especially for convective types. It can possibly be inferred from Fig. 3 that for the stratiform classes, the satellite data are not contributing any significant useful information beyond that contained in the environmental variables from the numerical weather model. However, it must also be kept in mind that the additional information content in the satellite data that improves the skill for the convective classes can indirectly improve the skill for the stratiform classes by e.g., preventing a convective pixel from being incorrectly classified as stratiform.

4.3. Benchmarking the precipitation typology from historical GEO sensors to new generation GEO sensors

Since the launch of geostationary satellites in the early 1970s until now, the IR channel at ~11 μm has been the legacy channel with almost five decades of routine global data. These channel observations are still used in almost all operational GEO precipitation retrieval algorithms and other merged satellite precipitation products (e.g., IMERG: Huffman et al., 2015; Tropical Amount of Precipitation with an Estimate of ERrors / TAPEER: Roca et al., 2010; PERSIANN: Sorooshian et al., 2000). During the late 1970s, the WV absorption channel at 6.2 μm from geostationary orbit was introduced in the Meteosat-1 Meteosat Visible and Infrared Imager, making it also a legacy channel with 3-4 decades of observations. Although its importance has been established in many precipitation retrieval studies (Ba and Gruber, 2001; Upadhyaya and Ramsankaran, 2014, 2016; Kuligowski et al., 2016), this channel is still not very commonly used in operational merged products. A third channel which showed significance in precipitation retrievals, in particular to separate water and ice phase clouds, is the channel at 12.3 μm (So and Shin, 2018; Kuligowski et al., 2016; Kühnlein et al., 2014; Thies et al., 2008; Behrangi et al., 2009). Due to its sensitivity to water vapor content in the atmosphere, this channel is considered a "dirty" IR channel. In this section, a benchmark is set up on the accuracy that can be achieved with the historical legacy channels and with the additional new-generation channels. It sets the stage to revisit climate data records for improved precipitation products using observations from more than one channel. Figure 4 shows the classification accuracy obtained with each precipitation type by using only the 11.2 μm channel derived predictors, the two legacy channels 11.2 μm and 6.2 μm, the three channels 11.2 μm, 6.2 μm and 12.3 μm, and all five channels. Each of these models include all possible BTDs, D-BTDs, and textures from the channels they contain.

A consistent improvement is observed in the classification performance by introducing additional channels. Specifically, a significant jump in accuracy is observed in most classes by adding the WV channel T6.2 to the IR T11.2, with about a 5% gain for the No-Precipitation and Convective types, and more than 10% with all other precipitation types. The addition of the IR T12.3 observations results in more modest improvements in the range of 2-4% for most types. The highest accuracy is obtained with five channels, indicating the need to test additional channels from the ABI . Finally, environmental predictors are also significantly important, with around 10% improvement in stratiform precipitation types (Warm Stratiform, Tropical Stratiform/Mix, and Cool Stratiform). It suggests that current operational products could incorporate more environmental predictors in addition to the mean-layer Relative Humidity used currently in SCaMPR (the operational NOAA algorithm for ABI: Kuligowski et al., 2016).

4.4. Important predictors for each precipitation type

This section identifies the most important predictors contributing to each precipitation type. This question addresses a significant gap of knowledge in the use of GEO sensors for precipitation characterization. The Treeinterpreter is implemented as explained in Section 3.2.

Figure 5a shows the box-plot distributions of contributions to the identification of the No-Precipitation type, computed with Treeinterpreter from the validation samples. The fifteen highest contributing predictors are displayed on the horizontal axis, with their contributions normalized to ease the inter-comparison between different predictors. For example, the highest contributing predictor to the identification of the No-Precipitation type is the D-BTD satellite predictor (T7.3 – T11.2) - (T8.5 – T12.3), with a contribution of 5-6% to the overall predicted probability of correctly classified No-Precipitation samples. This cumulative contribution line shows that the fifteen predictors together are responsible for more than 50% of the total contribution. Figure 5b shows an example of distributions for the same D-BTD predictor across the precipitation types, which illustrates the predictor ability in separating precipitation types. The distributions of predictors values from the D-BTD predictor are significantly different across types, with large negative values for the No-Precipitation type, higher values for stratiform types (Cool Stratiform, Warm Stratiform, and Tropical/Stratiform Mix), and contributions close to zero for convective types (Tropical Convective/Mix, Convective, and Hail).

Figure 6 shows the contribution distributions of the highest contributing predictors for all other precipitation types. As noted in Part 1, environmental predictors provide the highest contributions for most precipitation types. The highest contributing satellite predictors are identified in Figure S1. Figures 7 and 8 display the distributions of predictors to different precipitation types from the most important environmental and satellite predictors, respectively.

No-Precipitation Type:

From Fig. 5a, it can be observed that the first fifteen highest contributing predictors are all satellite-based, with accumulated contribution of more than 50% of the total contribution. It indicates the importance of satellite observations in separating Rain and No-Rain areas. The most important predictor is the D-BTD (T7.3 – T11.2) - (T8.5 – T12.3), which displays significantly different values with No-Precipitation than with precipitating types (Fig. 5b). Among other predictors, D-BTDs and textures of D-BTDs show higher contributions (Fig. 5a). Within D-BTDs predictors, one can notice the frequent combination of WV – IR and IR – IR channels. From Fig. 8k, D-BTDs involving WV – IR and WV – IR combinations display lesser separation of No-Precipitation from other types.

Hail Type:

Figure 6a indicates that the highest contributing predictors for the identification of Hail include both environmental and satellite predictors, and that their combined contributions add up to more than 60% of the total. Hail is generally associated with a high lapse rate (Fig. 7l), relatively warmer temperature near the surface (Fig. 7c), lower relative humidity at 500 hPa (Fig. 7f), and higher surface CAPE compared to other precipitation types (Fig. 7k). In earlier studies, RH and / or PW are used as predictors to provide information about low-level environmental conditions (Ba and Gruber, 2001; Vicente et al., 1998; Kuligowski et al., 2016), but the present analysis highlights the significance of other environmental parameters. However, it should be noted that there is a significant overlap in the values of the environmental predictors with the No-Precipitation type, which explains why that type has much greater reliance on satellite data than do the others.

Regarding satellite based-predictors, again D-BTDs generally make the greatest contributions. There is a noticeable structure in these D-BTDs; i.e., they largely consist of differences between two WV channels and/or the difference between one WV channel and the "dirty" IR channel T8.5, which is also sensitive to WV content. D-BTDs generally show smaller departures from 0 with Hail compared to other precipitation types (Fig.8i, j, k). In terms of textures, Contrast and Entropy tend to display higher values compared to other types (Fig.8b, c), and, regarding BTs, the Hail type is consistently associated with the coldest cloud-top temperatures (Fig. 8a, d, e).

From the distribution analysis (Fig. 7, 8), it can be observed that Hail characteristics have significant overlap with those associated with Convective precipitation types, which explains why Hail is often incorrectly classified as one of the Convective types (Table 4).

Convective Type:

From Figure 6b, Zenith Angle is the only satellite predictor in the fifteen highest contributing predictors, which cumulatively contributes to slightly more than 50% of the overall estimated probability. The remaining contribution is mainly from satellite predictors.

As it is observed, Hail, Lapse Rate, surface-based CAPE, Relative Humidity, and surface potential temperature show higher importance (Fig. 6b). Boxplots of both environmental and satellite predictors (Fig. 7 and 8, respectively) display significant overlap with other precipitation types except Cool Stratiform and No-Precipitation. It explains the lower accuracy obtained with the Convective precipitation type and the misclassification with Hail, Tropical Convective/Mix and Warm Stratiform (Table 4).

In terms of satellite predictors, D-BTDs display higher differences (Figs. 8i, j, k) between Convective types and Stratiform types . Similarly, different signatures are observed for the T6.2 Textures predictors (Figs. 8b, c); e.g. higher entropy and contrast with convective than with stratiform types. Overall, separating Convective types from other precipitation types is challenging.

Tropical Convective/Mix (TCM):

Like the Convective type, TCM displays most contributions (~80% total) from environmental predictors (Fig. 6c). The only satellite contributor is Zenith angle with the third highest contribution (>5%). It is also interesting to observe that the 850-500 hPa lapse rate contributes more than 10% in all three convective types, and it is one of the top contributing predictors in separating convective types from stratiform types. TCM is associated with higher precipitable water (Fig.7i) and Wet Bulb Temperature (WBT; Fig.7g). Both TCM and Tropical Stratiform/Mix have similar model predictors taking higher values; e.g., the 500hPa temperature (Fig.7a) and Relative Humidity profiles (Fig.7e, f). This explains the misclassification between both tropical classes (Table 4).

Regarding satellite predictors (Fig. 8), Zenith angle may indirectly capture the preferred location of occurring tropical precipitation types in the southeastern CONUS. From Fig. 8, there is overlap between TCM and Tropical Stratiform/Mix (TSM), while TCM has similar characteristics to the Convective type. These similarities in their respective predictor values make it challenging to separate these two types (Table 4).

Tropical Stratiform/Mix (TSM):

From Figure 6f and similar to TCM (Fig.6c), environmental predictors and Zenith Angle contribute around 60% of the total information content for identifying TSM. Notably, in stratiform types, predictors related to atmospheric moisture content such as relative humidity and precipitable water have higher contributions, while for convective types, the 850-500 hPa lapse-rate and CAPE consistently show higher contributions. Stratiform types show lower lapse rates than convective types in general (Fig. 7l). As reported in the TCM discussion, RH and PW are both high for the tropical classes (Fig. 7i, e, f).

In terms of satellite predictors (Fig. 8), BTs are generally highest for No-Precipitation, lower for Stratiform types (with Tropical Stratiform/Mix being colder than Warm Stratiform), and lower yet with Convective types where the coldest cloud tops are found with Hail (Fig 8a,d), which is perfectly consistent with their acting as a rough proxy for cloud-top temperature. The distribution of the values of BTs and other indices indicate that TSM characteristics range between TCM and Warm Stratiform, which in turn explains why TSM is often misclassified as one of these other two classes.

Warm Stratiform:

For Warm Stratiform (Fig. 6e), the highest contributing predictors are PW, humidity and other temperature-based environmental predictors, along with satellite zenith angle. The height of the 0°C isotherm is lower than for the other precipitation types. The distribution of values of RH, PW, and WBT are slightly lower than for the Convective types and TSM (Fig

7i, g), but with a large degree of overlap. This explains the mis-classification of Warm Stratiform as Convective types or TSM.

The satellite predictors exhibit a shift in BT signatures and other predictors (Fig. 8), from No-Precipitation to Stratiform to Convective types, but with considerable overlap. This again highlights the need to identify precipitation types probabilistically rather than deterministically.

Cool Stratiform:

For the Cool Stratiform type, almost 90% of the information content comes from the first fifteen predictors (Fig. 6d). Most of these predictors are temperature-based as expected, with the highest contribution coming from WBT which is also used in MRMS to separate Cool Stratiform from other precipitation types. The environmental temperature values are generally lower for Cool Stratiform than any other precipitation types (Fig. 7). Consistently, the same trend is observed for the height of the 0°C isotherm (Fig. 7j) and the 850-500hPa lapse rate. Other environmental predictors, such as RH, do not separate Cool Stratiform well from other precipitation types.

While the low-level environment is colder with Cool Stratiform than other precipitation types, Fig. 8 shows that ABI cloud top temperatures are warmer than Warm Stratiform (Fig. 8a, d, e) with some overlap.

4.5. Why incorrect estimates?

The causes of misclassification can be explained well for most precipitation types by the overlap of predictor distributions across different precipitation types (section 4.4; Figs. 7 and 8). However, the reason for overestimating the rain area (5% of No-Precipitation is misclassified as Warm Stratiform, as shown in Table 4), is not explained by the analysis thus far since it is observed that No-Precipitation is well separated from other classes for several predictors. The overestimation of rain area is further analyzed by plotting the distributions of the important predictors for the No-Precipitation type for both training and testing dataset separately, and by highlighting in which class they are (mis)classified. Figure 9 shows a representative example with the D-BTD predictor (T7.3 – T11.2) - (T8.5 – T12.3) which is the highest contributing predictor for No-Precipitation (see Fig. 5b).

As expected, the misclassified No-Precipitation samples are associated with different distributions of the predictor than the training No-Precipitation samples, in particular Warm Stratiform, which explains the large misclassification of No-Precipitation in this type. In Part I, the misclassified No-Precipitation samples are observed to generally occur along the edges of rainy areas with low RF estimated probabilities. From Figure 9, the characteristics of such misclassified samples are closer to the distributions of other precipitation types, which explains why the RF models tend to misclassify such samples. This is attributed to the sub-pixel rainfall variability and possible surface contribution along the edges of rainy areas associated with the satellite sampling resolution (i.e. non uniform beam filling (NUBF) as reported in Kirstetter et al., 2012, 2013 and Upadhyaya et al., 2020). Other sources of uncertainty can arise from the spatio-temporal matching between ABI and GV-MRMS, and possibly from internal MRMS procedures to avoid virga (Zhang et al., 2016). We also observe a similar behaviour with other precipitation type misclassifications (Figure S2).

5. Conclusions

The specific objective of this study is to understand the relationship between satellite and NWP environmental predictors and MRMS-classified precipitation types and processes through *interpreting* the developed ML based classification model. For the first time to our knowledge, a consistent and systematic analysis is performed on GEO satellite-based indices for precipitation detection and classification of types. The motivating research questions and major conclusions are indicated below:

Analysis 1: What is the impact of different categories of satellite predictors on classification accuracy?

- An improvement in the range of 5-20% is observed, with highest accuracy improvement for Warm Stratiform when compared to models developed with only Brightness Temperatures of 5 channels. Specifically, texture-based predictors significantly improve the classification accuracy.

Analysis 2: What is the relative impact of satellite-based predictors and environmental predictors?

- Except for Hail and No-Precipitation, the detection scores improve in the range of 10-20% by adding environmental predictors along with satellite predictors. Hail and No-Precipitation types achieve maximum accuracy with satellite predictors.

Analysis 3: How does the new generation of GEO sensors compare to historical benchmarks with legacy channels?

- Classification accuracy improves for all precipitation types by adding the WV channel T6.2 to the IR T11.2 channel, with a gain of around 5% for No-Precipitation and Convective types, and more than 10% with all other precipitation types.
- The highest accuracy with satellite predictors is obtained with all five channels used in this study, suggesting the need to test additional channels.

Analysis 4: Which predictors are contributing to different precipitation types?

- In terms of satellite predictors, BTs are empirically confirmed to be consistent with the physical understanding of BTs as a proxy for cloud-top temperatures: they display the warmest values for No-Precipitation, lower values for Stratiform types (with Tropical

Stratiform/Mix being colder than Warm Stratiform), and a further drop with Convective types, where the coldest cloud tops are found with Hail.

- Satellite observations are important in separating Rain and No-Rain areas. Of particular importance for precipitation detection are the D-BTDs predictors containing combinations of WV – IR and IR – IR channels.
- In stratiform types, predictors related to atmospheric moisture content such as relative humidity and precipitable water have the highest contribution, while for convective types, predictors 850-500 hPa lapse-rate and CAPE consistently showed the highest contribution.

Recommendations from the study are:

1. It is advantageous to derive predictors from the satellite brightness temperatures (e.g., texture, inter-band differences) instead of only using single-pixel, single-channel values.
2. Environmental predictors from NWP, such as CAPE, lapse rate, relative humidity, and precipitable water, bring complementary information and are therefore recommended to be included in retrieval algorithms
3. When possible, it is recommended to include the heritage channel T6.2 in operational precipitation retrieval algorithms and for precipitation reanalyses.

The conclusions and recommendations from this study will ultimately aid towards improved precipitation characterization and retrievals from space. In future work, more channels will be considered, as well as similar satellite platforms such as Himawari, Geostationary - Korea Multi-Purpose Satellite (GEO-KOMPSAT), Indian National Satellite (INSAT), Meteosat and FengYun (FY) series.

Acknowledgements

We are very much indebted to the teams responsible for the GOES-R, MRMS and SCaMPR products. Datasets for this research are available in these in-text data citation references: Kuligowski et al. (2016), Kirstetter et al. (2012; 2014), Benjamin et al. (2016). Authors acknowledge Y. Derin for helping with downloading and extracting environmental predictors. We also thank the editor and reviewers for their suggestions and comments that helped to improve the manuscript. Funding for this research was provided by the GOES-R Series Risk Reduction program, which provided support to the Cooperative Institute for Mesoscale Meteorological Studies at the University of Oklahoma under Grant NA16OAR4320115. P. Kirstetter acknowledges support from NASA Global Precipitation Measurement Ground Validation program under Grant NNX16AL23G and Precipitation Measurement Missions program under Grant 80NSSC19K0681. The contents of this paper are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of NOAA or the U.S. Government.

Reference

Andosa, Saabas. "Andosa/Treeinterpreter." *GitHub*, 2015, github.com/andosa/treeinterpreter.

Arkin, P. A. (1979). The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array. *Monthly Weather Review*, *107*(10), 1382-1387.

Ba, M. B., & Gruber, A. (2001). GOES multispectral rainfall algorithm (GMSRA). *Journal of Applied Meteorology*, *40*(8), 1500-1514.

Barret, E. C., and D. W. Martin (1981), *The Use of Satellite Data in Rainfall Monitoring*, 340 pp., Academic, London, U. K.

Baum, B. A., & Platnick, S. (2006). Introduction to MODIS cloud products. In *Earth science satellite remote sensing* (pp. 74-91). Springer, Berlin, Heidelberg.

Behrangi, A., Hsu, K. L., Imam, B., Sorooshian, S., Huffman, G. J., & Kuligowski, R. J. (2009). PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. *Journal of Hydrometeorology*, *10*(6), 1414-1429.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., ... & Lin, H. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, *144*(4), 1669-1694.

Ebert, E. E., Janowiak, J. E., & Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, *88*(1), 47-64.

Follansbee, W. A. (1973). Estimation of average daily rainfall from satellite cloud photographs.

Giannakos, A., & Feidas, H. (2013). Classification of convective and stratiform rain based on the spectral and textural features of Meteosat Second Generation infrared data. *Theoretical and applied climatology*, *113*(3-4), 495-510.

Grams, H. M., Kirstetter, P. E., & Gourley, J. J. (2016). Naïve Bayesian precipitation type retrieval from satellite using a cloud-top and ground-radar matched climatology. *Journal of Hydrometeorology*, *17*(10), 2649-2665.

Grams, H. M., Zhang, J., & Elmore, K. L. (2014). Automated identification of enhanced rainfall rates using the near-storm environment for radar precipitation estimates. *Journal of Hydrometeorology*, *15*(3), 1238-1254.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

Hong, Y., Hsu, K. L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*(12), 1834-1853.

Kidder, S. Q., and T. H. Vonder Haar, 1995: *Satellite Meteorology: An Introduction*. Academic Press.

Kidder, S. Q., KIDDER, R. M., & Haar, T. H. V. (1995). *Satellite meteorology: an introduction*. Gulf Professional Publishing.

Kilonsky, B. J., & Ramage, C. S. (1976). A technique for estimating tropical open-ocean rainfall from satellite observations. *Journal of Applied Meteorology*, *15*(9), 972-975.

Kirstetter, P. E., Hong, Y., Gourley, J. J., Cao, Q., Schwaller, M., & Petersen, W. (2014). Research framework to bridge from the Global Precipitation Measurement Mission core satellite to the constellation sensors using ground-radar-based national mosaic QPE. *Remote sensing of the terrestrial water cycle*, 61-79.

Kirstetter, P. E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., ... & Amitai, E. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar–based National Mosaic QPE. *Journal of Hydrometeorology*, *13*(4), 1285-1300.

Kirstetter, P. E., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*, *144*, 191-205.

Kühnlein, M., Appelhans, T., Thies, B., & Nauß, T. (2014). Precipitation estimates from MSG SEVIRI daytime, nighttime, and twilight data with random forests. *Journal of Applied Meteorology and Climatology*, *53*(11), 2457-2480.

Kuligowski, R. J., Li, Y., Hao, Y., & Zhang, Y. (2016). Improvements to the GOES-R rainfall rate algorithm. *Journal of Hydrometeorology*, *17*(6), 1693-1704.

Lovejoy, S., & Austin, G. L. (1979). The delineation of rain areas from visible and IR satellite data for GATE and mid-latitudes. *Atmosphere-ocean*, *17*(1), 77-92.

Meyer, H., Kühnlein, M., Appelhans, T., & Nauss, T. (2016). Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmospheric research*, *169*, 424-433.

Min, M., Bai, C., Guo, J., Sun, F., Liu, C., Wang, F., ... & Dong, L. (2018). Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(5), 2557-2570.

Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International journal of scientific and research publications*, *3*(5), 1.

National Academies of Sciences, Engineering, and Medicine (2018). *Thriving on our changing planet: A decadal strategy for Earth observation from space*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24938

So, D., & Shin, D. B. (2018). Classification of precipitating clouds using satellite infrared observations and its implications for rainfall estimation. *Quarterly Journal of the Royal Meteorological Society*, *144*, 133-144.

Tao, Y., Hsu, K., Ihler, A., Gao, X., & Sorooshian, S. (2018). A two-stage deep neural network framework for precipitation estimation from bispectral satellite information. *Journal of Hydrometeorology*, *19*(2), 393-408.

Thies, B., Nauß, T., & Bendix, J. (2008). Precipitation process and rainfall intensity differentiation using Meteosat second generation spinning enhanced visible and infrared imager data. *Journal of Geophysical Research: Atmospheres*, *113*(D23).

Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Haar, T. H. V., & Reinke, D. L. (1999). A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, *10*(1), 138-151.

Tjemkes, S. A., Van de Berg, L., & Schmetz, J. (1997). Warm water vapour pixels over high clouds as observed by Meteosat. *Beitrage zur Physik der Atmosphare-Contributions to Atmospheric Physics*, *70*(1), 15-22.

Tsonis, A. A., & Isaac, G. A. (1985). On a new approach for instantaneous rain area delineation in the midlatitudes using GOES data. *Journal of climate and applied meteorology*, *24*(11), 1208-1218.

Upadhyaya, S. A., Kirstetter, P. E., Gourley, J. J., & Kuligowski, R. J. (2020). On the propagation of satellite precipitation estimation errors: from passive microwave to infrared estimates. *Journal of Hydrometeorology*, *21*(6), 1367-1381.

Upadhyaya, S. A., Kirstetter, P. E., Kuligowski, R. J., & M. Searls (2021). Classifying precipitation from GEO Satellite Observations: Prognostic Model. *Journal of Geophysical Research: Atmosphere* (Submitted).

Upadhyaya, S., & Ramsankaran, RAAJ. (2016). Modified-INSAT Multi-Spectral Rainfall Algorithm (M-IMSRA) at climate region scale: Development and validation. *Remote Sensing of Environment*, *187*, 186-201.

Upadhyaya, S., & Ramsankaran, RAAJ. (2014). Multi-index rain detection: a new approach for regional rain area detection from remotely sensed data. *Journal of Hydrometeorology*, *15*(6), 2314-2330.

Vicente, G. A., Scofield, R. A., & Menzel, W. P. (1998). The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, *79*(9), 1883-1898.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., ... & Arthur, A. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621-638.

Zhang, J., Qi Y. ,Howard K. ,Langston C. ,and Kaney B. (2011). Radar Quality Index (RQI)— A combined measure of beam blockage and VPR effects in a national network. *Proc.*

Eighth Int. Symp. on Weather Radar and Hydrology, Exeter, United Kingdom, Royal Meteorological Society.

List of Tables:

Table 1: Categories and example of predictors used in study

Category	***Predictor Type***	***Example***
1	BT (Brightness Temperature)	T7.3
2	BTD (Brightness Temperature Difference)	T8.5 – T11.2
3	D-BTD (Difference of BTDs)	(T6.2* – T7.3) – (T8.5 – T11.2)
4	Te (GLCM Textures)	T11.2 mean
5	Ze (Zenith Angle)	Ze
6	Environmental Predictors (NWP)	Details in Table 3

*T6.2 is read as brightness temperature of ABI channel 6.2μm

Table 2. Quality controlled sample size across MRMS-GV different precipitation types

	Convec	***Cool_Strat***	***Hail***	***NoPrecip***	***Trp_ConvMix***	***Trp_StratMix***	***WarmStart***
Total	114,749	131,539	40,774	4,118,832	67,866	883,276	14,723,928
Train (70%)	87,919	98,677	31,612	2,966,493	51,842	651,764	10,694,772
Test (30%)	26,830	32,862	9,162	1,152,339	16,024	231,512	4,029,156
Balanced training Sample sized							
Balanced Train	31,612	31,612	31,612	31,612	31,612	31,612	31,612

Table 3: Environmental predictors derived from RAP Model

Sl No	**Environmental Variable**
1	Vertically integrated precipitable water (kg/m2)
2	1000-700-hPa mean relative humidity (%)
3–6	Relative humidity (%) at 900-hPa, 850-hPa, 700-hPa and 500-hPa
7	Surface equivalent potential temperature (K)
8	Surface-based convective available potential energy (CAPE) (J/kg)
9	Surface temperature (C)
10–12	Temperature (K) at 850-hPa, 700-hPa and 500-hPa
13	Height of 0C isotherm (km)
14 – 16	**Wind shear (m/s) from surface to 850 hPa, 700 hPa and 500 hPa**
17– 18	**Lapse rate (K/km) at 850-500-hPa and 850-700-hPa**
19	**Wet Bulb Temperature**

* Note: The **bold rows** are predictors derived from other RAP output fields and other predictors are directly available from RAP output.

Table 4. Normalized Contingency Matrix (in %), overall classification accuracy and Kappa coefficient. Blue highlighted cells are class accuracy statistics (Probability of Detection).

		RF: Predicted classes								Overall Accuracy
		Convec	**Cool_Strat**	**Hail**	**NoPrecip**	**Trp_ConvMix**	**Trp_StratMix**	**WarmStart**	**Total**	
MRMS: Precip Types	**Convec**	**47**	**0**	**14**	**1**	**11**	**9**	**19**	**100**	**75.90**
	Cool_Strat	**0**	**97**	**0**	**0**	**0**	**0**	**3**	**100**	
	Hail	**26**	**0**	**70**	**0**	**2**	**0**	**1**	**100**	
	NoPrecip	**0**	**0**	**1**	**93**	**0**	**0**	**5**	**100**	
	Trp_ConvMix	**10**	**0**	**6**	**0**	**55**	**26**	**3**	**100**	
	Trp_StratMix	**4**	**0**	**0**	**0**	**12**	**64**	**20**	**100**	
	WarmStart	**10**	**1**	**2**	**2**	**2**	**11**	**72**	**100**	

Table 5. List of experiments/analysis and their objectives

Analysis	Objective
1	Understand the significance of different categories of satellite predictors
2	Accuracy assessment of classifier with satellite only, environmental predictors only and overall predictors
3	Benchmarking accuracy that can be achieved with historical GEO sensors operating with one or two channels compared to the new generation of satellite sensors
4	Feature/Predictor importance for each precipitation type individually
5	Why incorrect estimates?

List of Figures:

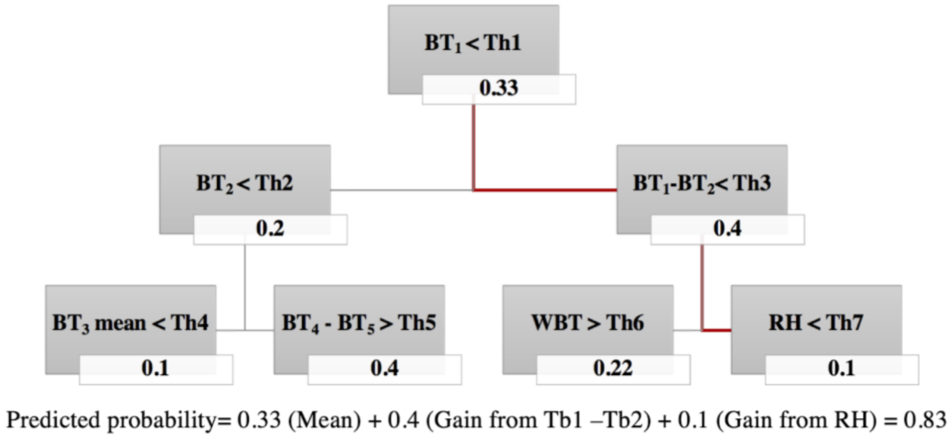

Figure 1. Treeinterpreter output for a hypothetical sample with Threshold (Th), Brightness Temperature (BT), Wet Bulb Temperature (WBT), and Relative Humidity (RH). Red lines indicate the decision path taken by the algorithm for the sample under consideration. Each grey box indicates the decision function at each node and “Th” represents the decision threshold determined during the training phase. The boxes with numbers located at the bottom of each grey box represent the contribution from each node to the final estimated probability.

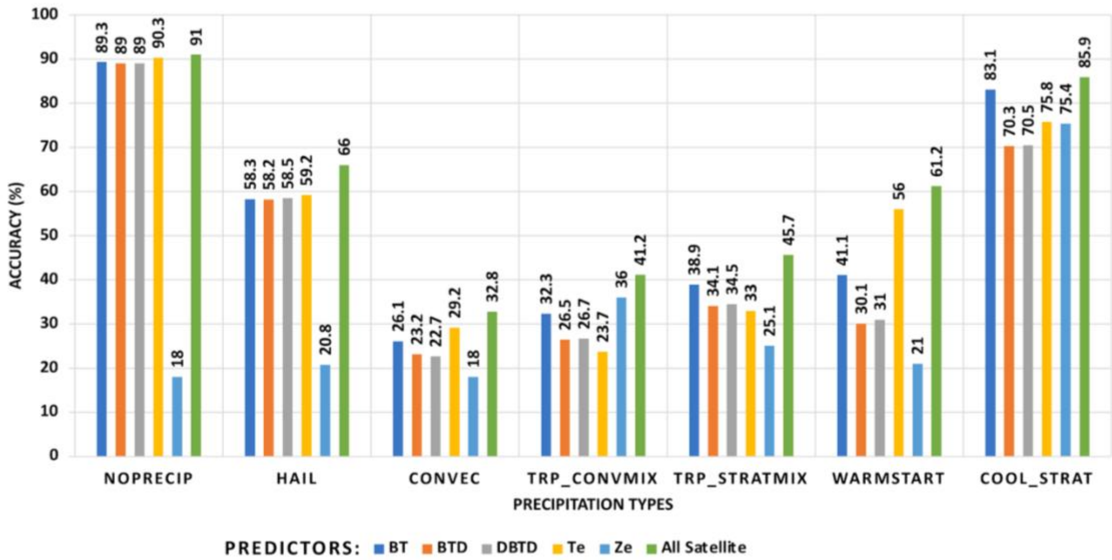

Figure 2. Classification accuracy for each precipitation type obtained with RF models developed with separate categories of satellite predictors: Brightness Temperatures (BT), Brightness Temperature Difference (BTD), Difference of BTD (D-BTD), Texture (Te), and Zenith Angle (Ze).

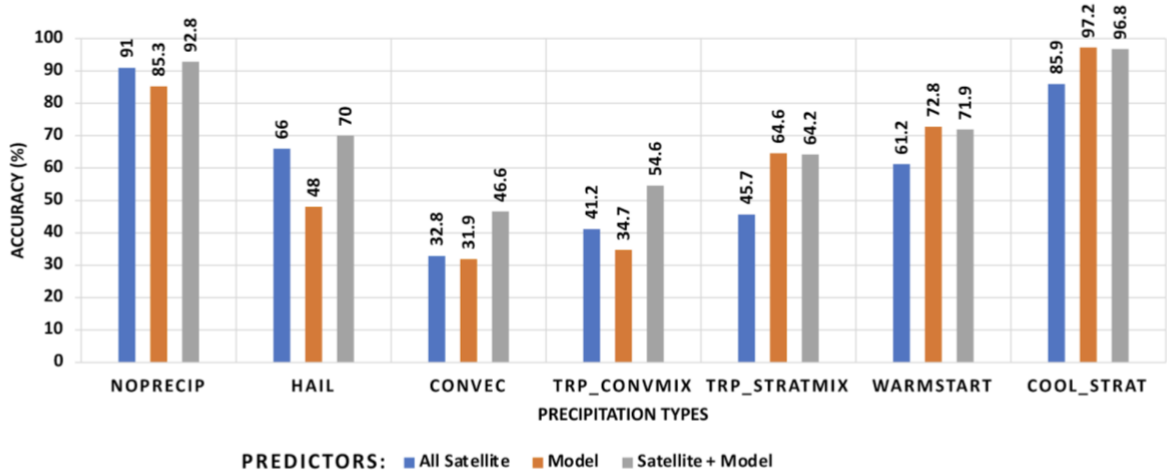

Figure 3. Classification accuracy for each precipitation type from satellite predictors, environmental predictors, and both.

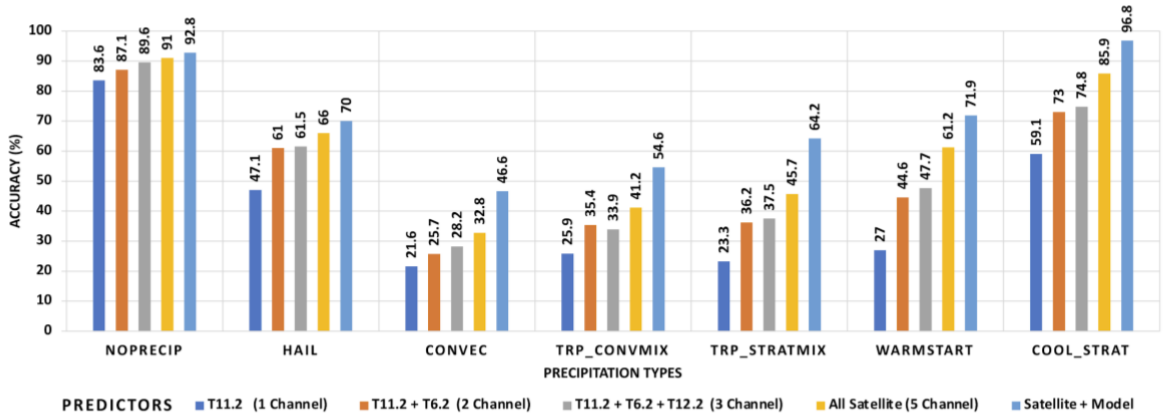

Figure 4. Classification accuracy for each precipitation type obtained with RF models developed with one channel (11.2 µm), two (11.2 and 6.2 µm), three (11.2, 6.2, 12.3 µm), and all five channel satellite predictors.

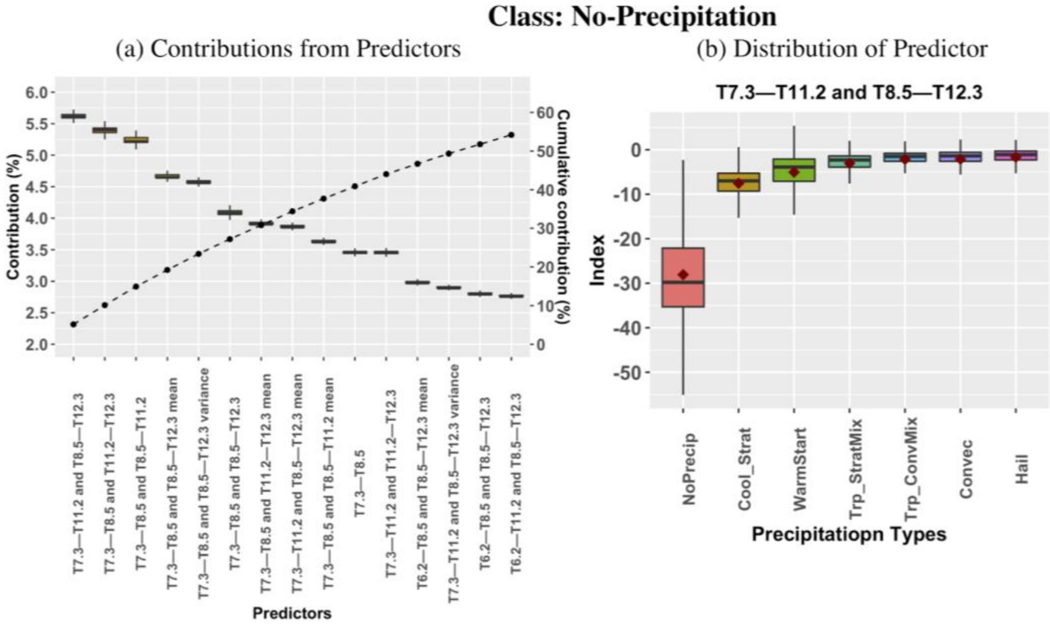

Figure 5. Examples of most important predictors identification: (a) Box plots of contributions to the identification of the No-Precipitation type for the first fifteen predictors, along with cumulative mean contribution (dashed line); (b) Distributions of predictor values (D-BTD=(T7.3-T11.2) - (T8.5-T12.3)) and mean (red dots) to the identification of different precipitation types.

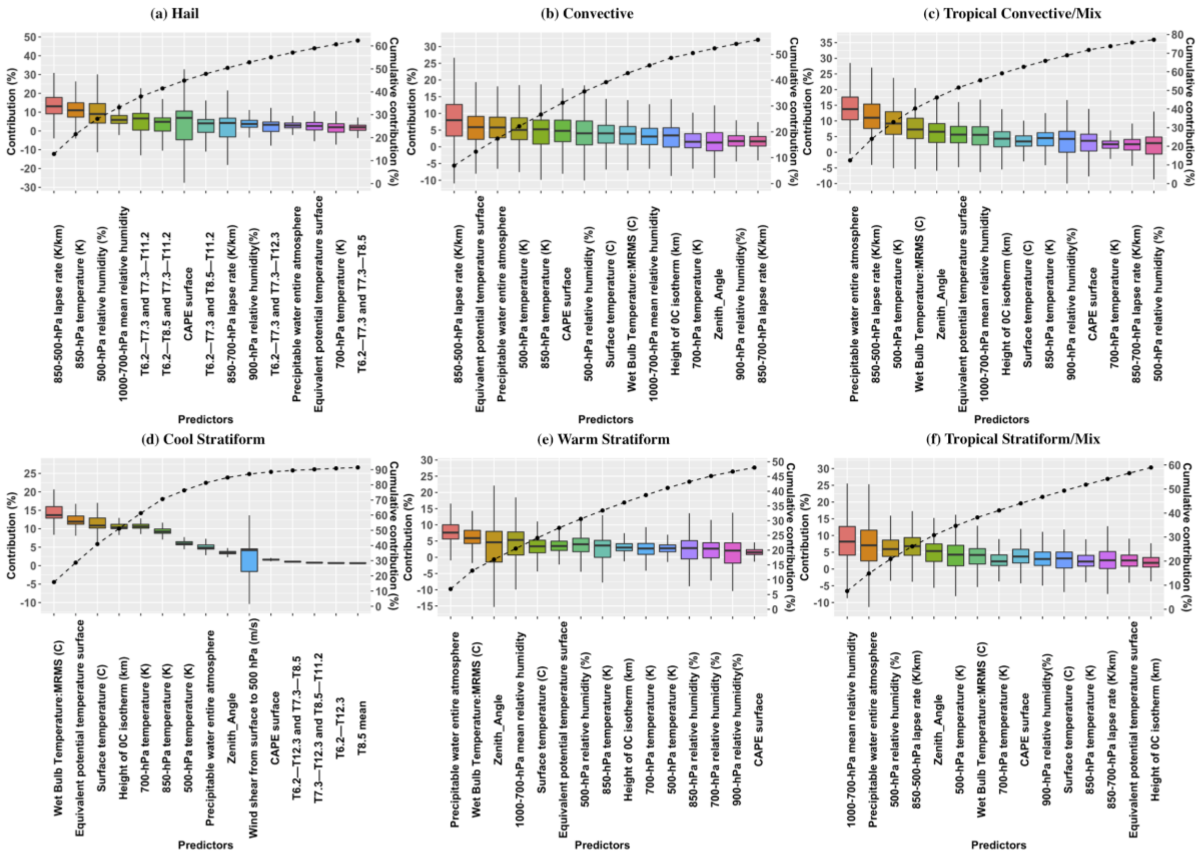

Figure 6. Same as Figure 5a but for all other precipitation types.

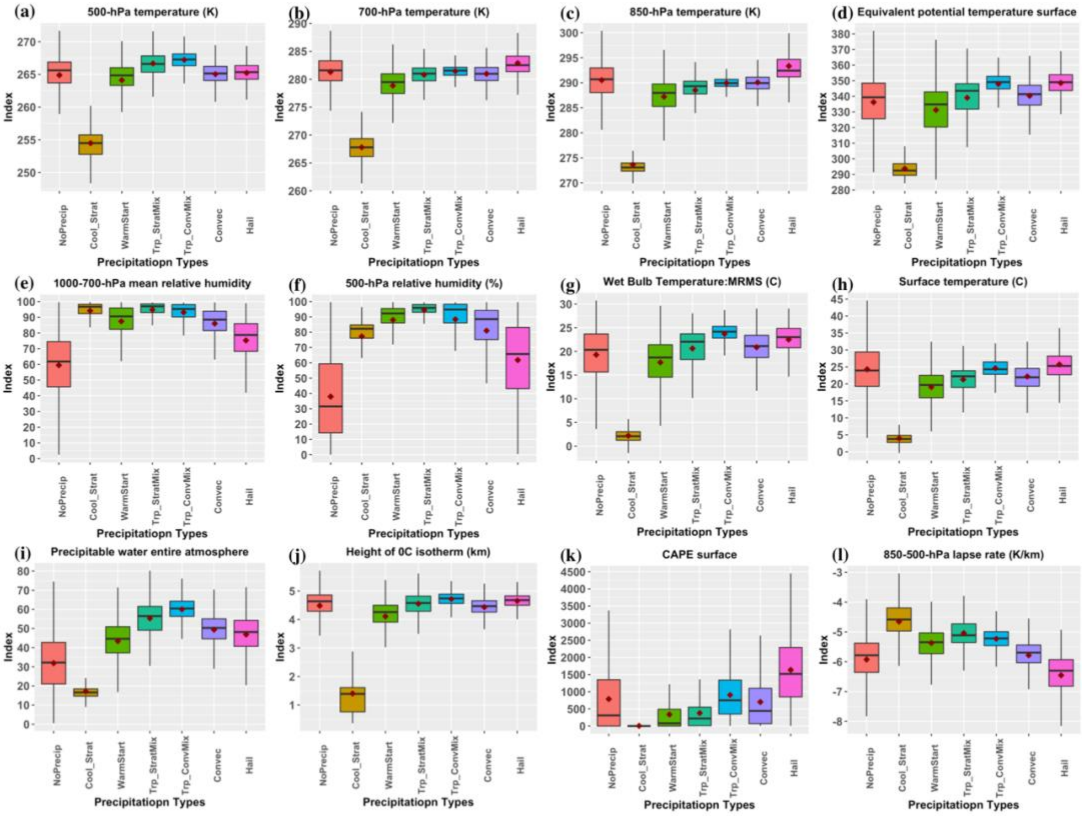

Figure 7. Distributions of the most significant environmental predictors across different precipitation types

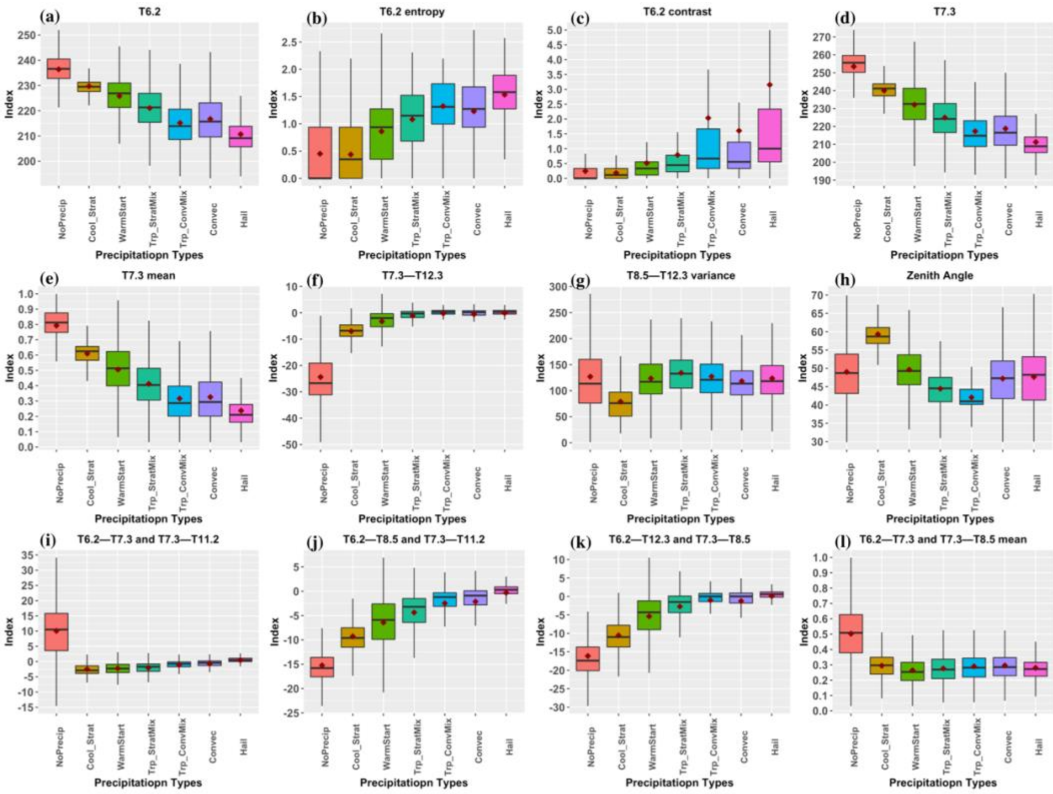

Figure 8. Distribution of the most significant satellite predictors across different precipitation types

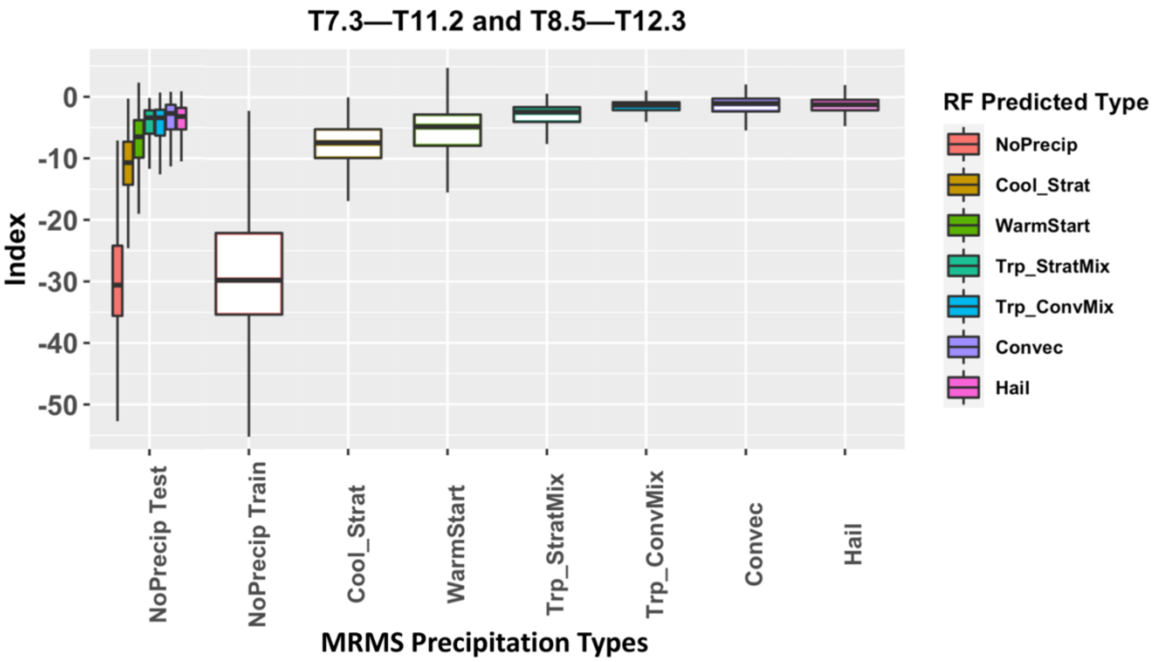

Figure 9. Same as Figure 5b, and for No-Precipitation type distributions separated for both training and testing data across RF predicted precipitation types. The smaller boxes on the left are MRMS-classified No-Precipitation testing data separated according to RF estimated precipitation types. The distributions associated with other MRMS-classified precipitation types are also given for comparison purposes.