# A Machine-Learning Approach to Mitigate Ground Clutter in the GPM Combined Radar-Radiometer Algorithm (CORRA) Precipitation Estimates

Mircea Grecu,[a,b] Gerald M. Heymsfield,[a] Stephen Nicholls,[a,c] Stephen Lang,[a,c] William S. Olson,[a,d]

[a] *NASA Goddard Space Flight Center, Greenbelt, Maryland*

[b] *Morgan State University, Baltimore, MD*

[c] *Science Systems and Applications, Inc., Greenbelt, Maryland*

[d] *University of Maryland, Baltimore County, Baltimore,Maryland*

*Corresponding author*: Mircea Grecu, email:mircea.grecu-1@nasa.gov

ABSTRACT: In this study, a machine-learning based methodology is developed to mitigate the effects of ground clutter on precipitation estimates from the Global Precipitation Mission Combined Radar-Radiometer Algorithm. Ground clutter can corrupt and obscure precipitation echo in radar observations, leading to inaccuracies in estimates. To improve upon the previous work, this study introduces a more general ML approach that enables a more systematic investigation and a better understanding of uncertainties in clutter mitigation. To allow for a less restrictive exploration of conditional relations between precipitation above the lowest clutter free bin and surface precipitation, reflectivity observations above the clutter are included in a fixed-size set of predictors along with the precipitation type, surface type, and freezing level to estimate surface precipitation rates, and several ML-based estimation methods are investigated. A Gradient Boosting Model (GBM) is ultimately identified as the best candidate for systematic evaluations, as it is computationally fast to train and apply while effective in applications. The GBM appears effective in providing unbiased estimates; however, it is not much more effective in reducing random errors in the estimates relative to a simple bias correction approach. The fact that other ML approaches such as the k nearest neighbor method, the random forest method, and feed forward neural network approach showed similar performance in initial evaluations suggests that the inability of the GBM to achieve much more than bias removal is the result of indeterminacy in the data rather than limitations in the ML approach.

SIGNIFICANCE STATEMENT: Ground clutter can obscure and corrupt the precipitation echo in the reflectivity observations by space borne radar, leading to inaccuracies and biases the surface precipitation estimates. In this study, a machine learning approach is developed to mitigate the effects of ground clutter on precipitation estimates from the Global Precipitation Mission (GPM) Combined Radar-Radiometer Algorithm (CORRA). The approach is shown to be effective in removing the biases associated with the simplest ground clutter mitigation approach and reducing the random errors associated with more complex climatologically based bias-removal approaches.

## 1. Introduction

In radar meteorology, the echo originating in power emitted by the radar and reflected by the ground is called ground clutter. Ground clutter has as negative impact on observations collected by Dual Frequency Precipitation Radar (DPR) of the NASA Global Precipitation Measurement (GPM) mission (Skofronick-Jackson et al. 2017), as it may obscure or corrupt radar signal associated with precipitation. The extent of ground clutter in space-borne radar observations increases with incidence angle (Kubota et al. 2016). Shown in Fig. 1 is a single scan representation of the Ku-band reflectivity observed by the DPR from GPM orbit 50853 on 9 February 2023. The enhanced reflectivity values at ranges close to (and larger than) 170 are contaminated by ground clutter. The lowest bins that are deemed clutter-free by the DPR algorithm (Iguchi et al. 2021) are indicated by the black line in the plot. As apparent from the figure, the number of bins affected by clutter is quite significant for observations near the edge of the swath. Relative to the sea-level, 20 range bins (bin width of 125 m) affected by clutter at the maximum scan incidence angle of 17° are equivalent to a clutter height of about 2.4 km at the edge of the DPR swath. While the assumption that the precipitation flux does not change significantly with height may be reasonable in some situations, it is likely to result in significant biases in the surface precipitation estimates in weather systems with freezing levels close to the ground. This is because ice processes such as riming and depositional growth can result in significant flux changes.

To mitigate such biases, statistical correction methodologies, akin to those used to estimate the surface reflectivity from ground-based precipitation radar observations, may be used. Specifically, in ground-based radar, as the height of horizontally scanning radar beams increase with range, the lowest-elevation reflectivity observations may be significantly elevated above the ground at large

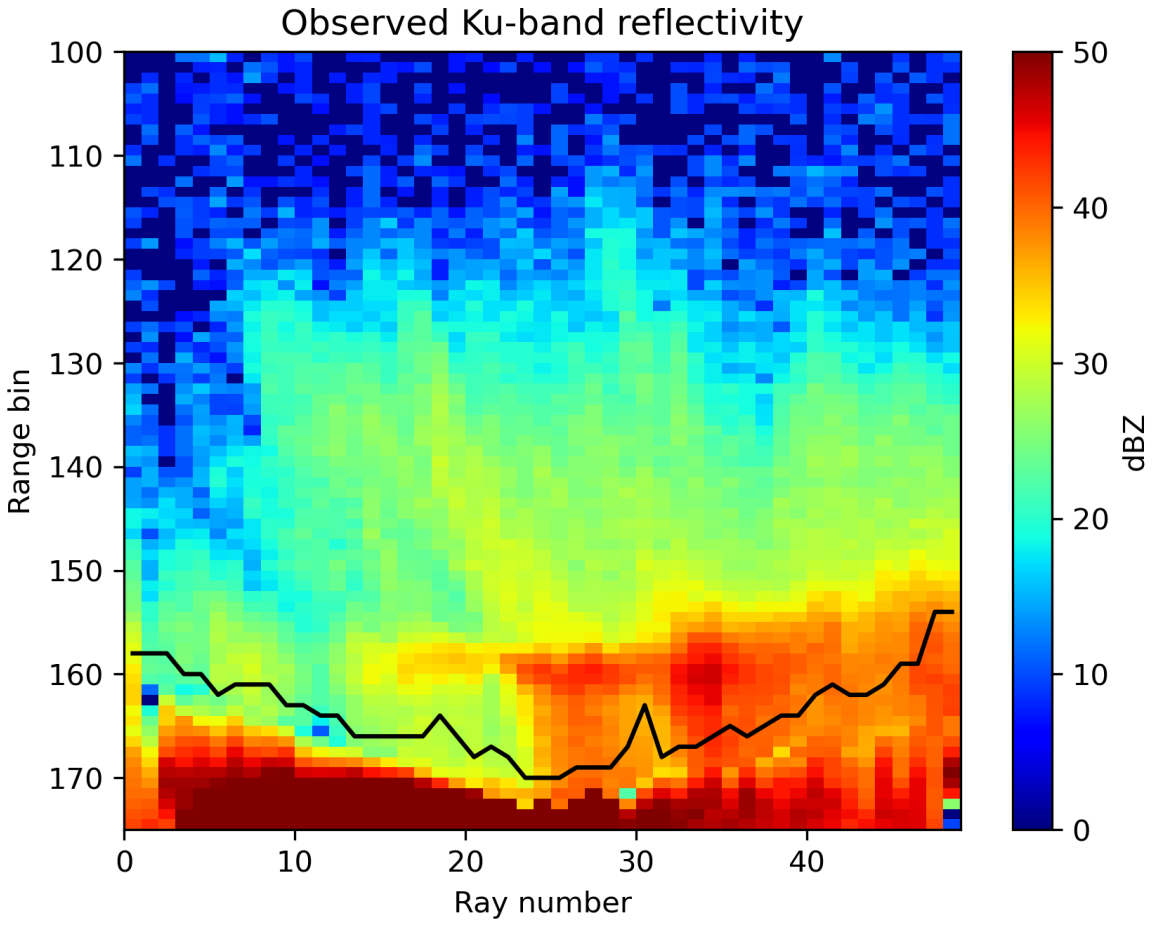



FIG. 1. Cross-track section of observed Ku-band reflectivity field. Range bin spacing is 125 m and the black line indicates the lowest clutter-free bins. Bin 175 corresponds to the Earth ellipsoid.

ranges. For example, the beam center height is about 1500 m for an elevation angle of 0.5° and a range of 100.0 km (NWS 2023). Traditionally, to estimate the surface reflectivity from the lowest elevation angle reflectivity observations of ground radars, short range reflectivity observations from multiple elevation angle scans are used to derive statistical relationships between surface reflectivities and reflectivities aloft (Koistinen 1991). A similar approach can be applied to mitigate ground clutter in space-borne radar observations, with the difference being that the relationships between surface precipitation rates and precipitation rates aloft are derived from near-nadir space-borne radar observations and associated precipitation estimates that are minimally impacted by ground clutter. This approach has already been applied by Hirose et al. (2021) to refine the GPM DPR surface precipitation estimates.

In this study, we present a machine learning (ML) based methodology (Géron 2022) to investigate and mitigate ground clutter effects on precipitation estimates from the GPM combined radar-radiometer algorithm (CORRA). While conceptually similar to the approach of Hirose et al. (2021), our methodology is different in several key aspects and provides additional insight into ground-clutter-related uncertainties in the surface precipitation estimates and the best strategies to mitigate them. Unlike Hirose et al. (2021), we use reflectivity profile observations (rather than profiles of estimated precipitation rates) in the derivation of relationships between the precipitation rate in the lowest clutter free bin and the surface precipitation rate. The benefit of using reflectivity rather than precipitation profiles is that it enables the development of more physically consistent estimates. That is, radar profiling algorithms (Iguchi et al. 2021; Grecu et al. 2016) require assumptions regarding precipitation structure in the clutter to accurately incorporate estimates of the path integrated attenuation (PIA) from the Surface Reference Technique (SRT) to correct for attenuation down to the surface. However, if the clutter mitigation technique requires precipitation estimates, it can only be applied after the radar estimation process is complete. This may result in inconsistencies between the assumptions regarding the attenuation due to precipitation in the clutter and the actual precipitation estimates. While such inconsistencies may be addressed through iterative procedures, they result in a more computationally intensive retrieval process. In contrast, a clutter mitigation technique that uses reflectivity observations directly to derive relations between information above the clutter and precipitation in the clutter can be explicitly incorporated into the attenuation correction and precipitation estimation process, and this eliminates the need for iterative procedures to ensure the consistency of results. It should be mentioned, however, that the benefit (if any) of estimating the reflectivity in the clutter is limited in deep convection, because there are large uncertainties in the attenuation correction process both above and in the clutter. In this case, additional uncertainties caused by physical inconsistencies may not matter. Another distinction relative to Hirose et al. (2021), is that our methodology is based on ML, which is beneficial from the feature engineering perspective (Zheng and Casari 2018). Specifically, machine learning models can effectively extract relevant information from the data without having to resort to the explicit identification of features (defined as numerical attributes uniquely derived through a computational procedure applied to input data), thereby reducing the need for manual feature engineering, which can be time-consuming and error-prone for human experts. For example, the precipitation gradient

**Lines 94-98: You mention one of the benefits of using ML techniques are the feature engineering capabilities of these systems, yet this isn't really discussed afterward. What did the RF feature importance scores tell you about the predictors? Further, you could also include an examination of the SHAP values from the NN to see if you get similar results. It would be interesting to see this discussed more, and then features related back to physical processes/ relationships one might expect to see. Incorporating other atmospheric variables like temperature, humidity etc. in future work might also provide some useful/interesting context.**

with respect to radar range is an intuitively-derived feature in the surface precipitation estimation approach of Hirose et al. (2021). While features that make intuitive sense are valuable, questions regarding their optimality are difficult to objectively address without tedious investigation. From this perspective, ML procedures that do not require explicit features are worth considering. In addition, the organization of data in a format that facilitates the development of an ML model automatically facilitates the model's evaluation.

The paper is organized as follows. In Section 2, we present the ML methodology used to estimate the surface precipitation rate from reflectivity observations not affected by clutter as well as additional information such as the precipitation and surface type and the zero degree isotherm height. In Section 3, we present the results of the application of the ML methodology to the GPM CORRA precipitation estimates. In Section 4, we offer some conclusions from the study.

## 2. Methodology

### *a. General considerations*

The simplest method to estimate the precipitation rate at a given height above the sea level (and for a given precipitation type $PT$, surface type $ST$, and freezing level ($FL$)) from a precipitation rate at a higher level is to re-scale the higher level value by the ratio of the climatological mean precipitation rates at the two levels. Mathematically, this may be written as

$$P_{rate}(H_1, PT, ST, FL) = P_{rate}(H_2, PT, ST, FL)\frac{< P_{rate}(H_1, PT, ST, FL) >}{< P_{rate}(H_2, PT, ST, FL) >} \quad (1)$$

where $P_{rate}$ is the precipitation rate, $H_1$ is the height where the estimate is needed, but for which no direct measurement is available, $H_2$ is lowest clutter free height ($H_2 > H_1$) where a radar measurement is available, and operator $< \square >$ denotes the climatological mean over a large dataset characterized by the same freezing level, surface and precipitation type.

While simple in form, the challenge in applying a clutter correction methodology based on Eq. (1) is the derivation of the correction factors $\frac{<P_{rate}(H_1)>}{<P_{rate}(H_2)>}$ for all possible ($H_1$,$H_2$) pairs. Nevertheless, because the ground clutter depth is a function of the scanning incidence angle, estimates of the

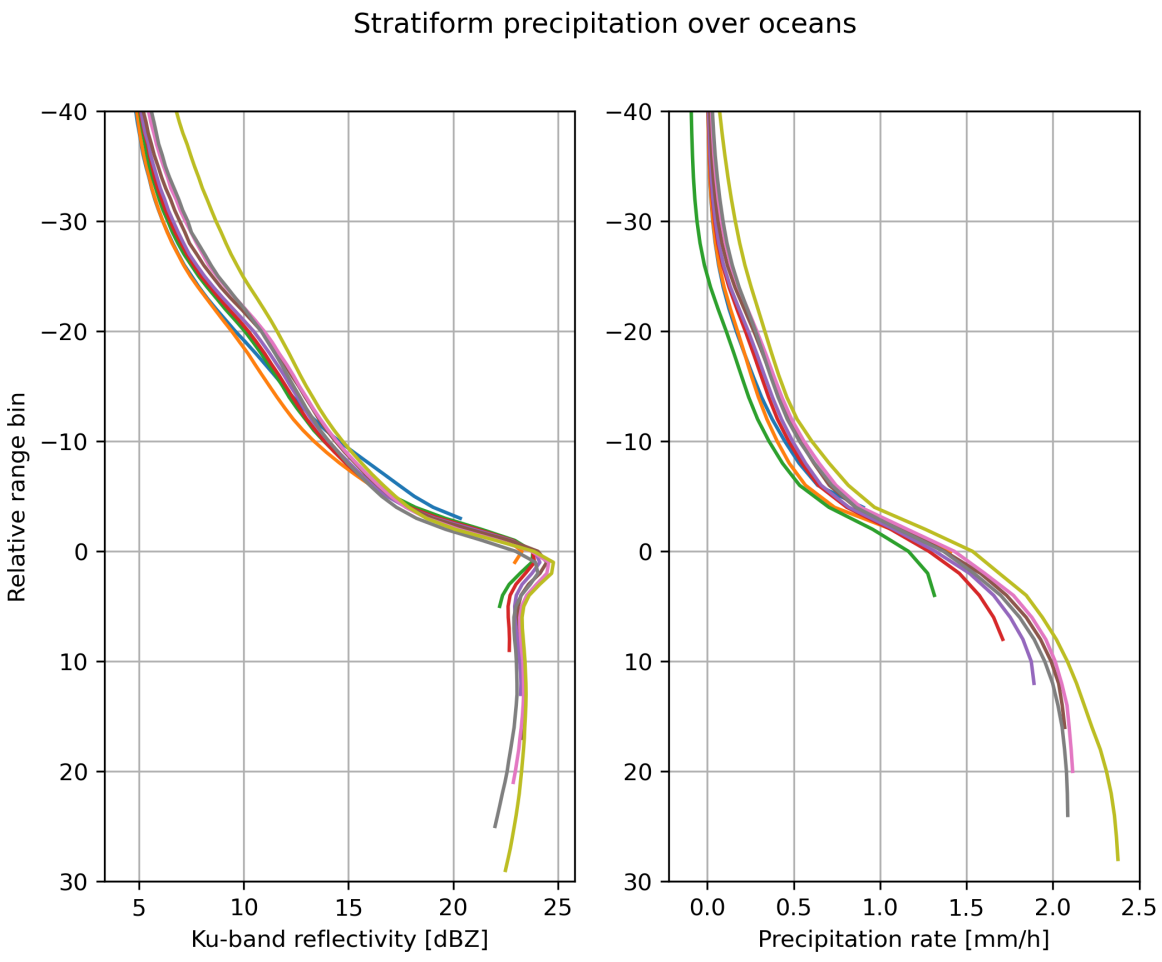



FIG. 2. Conditional mean reflectivity and precipitation rate profiles over oceans for stratiform precipitation with various freezing level heights.

**Figure 2: What reflectivity units do you use to calculate the mean profiles here? Are the means produced in Z or dBZ space? Further, what reflectivity units are fed into the different ML models? Finally, what do the different colored lines represent? Perhaps a legend is needed here.**

climatological correction factor derived from near-nadir reflectivity observations and precipitation estimates may be used to mitigate the clutter near the edges of the swath. Shown in Fig. 2 are over-oceans conditional mean reflectivity and precipitation rate profiles from the GPM CORRA algorithm (Grecu et al. 2016) for stratiform precipitation with various freezing level heights. The profiles are plotted relative to the 0°C bin to emphasize similarities rather than differences due to temperature-dependent processes. One year's worth (i.e. 2018) of DPR observations and associated GPM CORRA retrievals characterized by fewer than eight bins affected by clutter are selected and used in calculations of the mean profiles. The data are partitioned based on the freezing level height in 12 distinct subsets, with the freezing level heights of each subset within 125

m of 1.875+$k$*0.25 km with $k$ varying from 0 to 11, resulting in 12 conditional mean profiles. As shown in the figure, the mean reflectivity and the associated precipitation profiles tend to align with one another. This behaviour may be used to mitigate the impact of clutter, even in near-nadir DPR observations that are affected by clutter at relatively low altitudes that make direct precipitation rate estimation at or near the surface impossible. Specifically, the data in Fig. 2 suggests that

$$\frac{< P_{rate}(H_1, PT, ST, FL) >}{< P_{rate}(H_2, PT, ST, FL) >} \approx \frac{< P_{rate}(H_1 + dFL, PT, ST, FL + dFL) >}{< P_{rate}(H_2 + dFL, PT, ST, FL + dFL) >} \tag{2}$$

where dFL is the difference between two distinct freezing level heights (FLH). The veracity of Eq. (2) is supported by the fact that in plots relative to the 0°C isotherm, the conditional mean precipitation profiles in Fig. 2 look very similar to profiles characterized by higher FLH and extending to greater depths below the 0°C. Here, the conditional mean precipitation rate refers to the mean precipitation rate in situations where precipitation is occurring in the LCFB (non-zero). The selection of profiles (with a maximum of eight radar bins impacted by ground clutter) results in a minimum value of $H_1$ of 1,000m (for a climatology derived from profiles with at most six bins affected by clutter). However, one can use Eq. (2) to approximate $\frac{<P_{rate}(0,PT,ST,FLH)>}{<P_{rate}(H_2,PT,ST,FLH)>}$ as $\frac{<P_{rate}(1{,}000m,PT,ST,FLH+1{,}000m)>}{<P_{rate}(H_2+1{,}000m,PT,ST,FLH+1{,}000m)>}$.

Shown in Fig. 3 are conditional mean reflectivity and precipitation rate profiles from CORRA for stratiform precipitation with various freezing level heights over land. The conditional mean precipitation profiles over land exhibit more variability than over oceans. However, this may be a consequence of precipitation retrieval artifacts rather than differences in temperature-dependent physical processes. Specifically, given that the SRT PIA estimates are noisier and less reliable over land, their impact on precipitation estimates may be less systematic, which could result in a larger spread of conditional mean estimates. Nevertheless, Eq. (2) is still a reasonable assumption.

The mean reflectivity profiles shown in Figs. 2 and 3 are stratified by precipitation type (stratiform), freezing level and surface type only, but it is conceivable that features that further separate the relationships between the reflectivity observations and the final precipitation estimates exist. As previously mentioned, Hirose et al. (2021) use the precipitation slope to stratify the database of near-nadir precipitation supporting their precipitation refinement process. In the current study, we

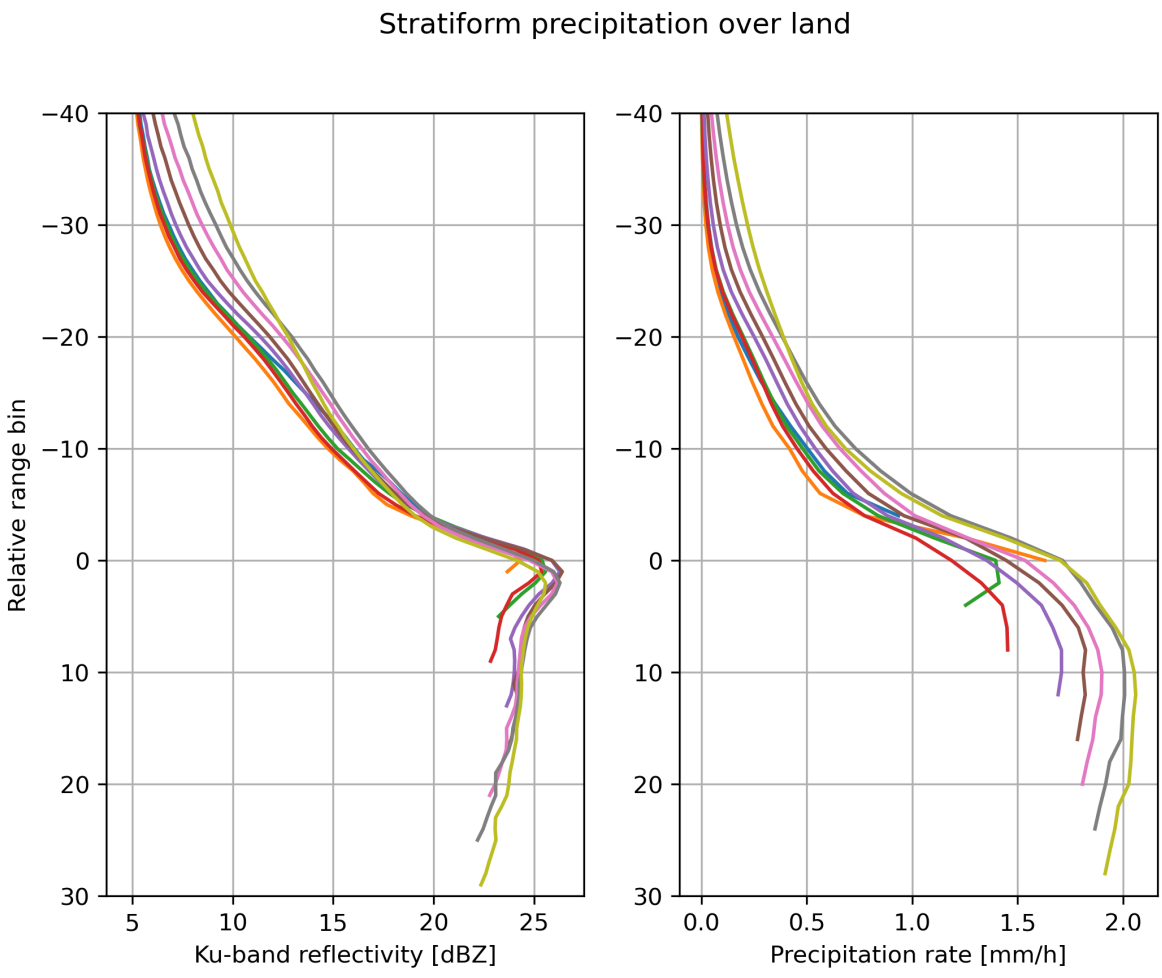



FIG. 3. Same as 2 but over land.

also investigate the slope of the reflectivity profile below the freezing level as a feature potentially useful for predicting the surface precipitation rate from the lowest clutter free precipitation rate. Specifically, the slope of the Ku-band reflectivity observations in the first six radar bins below the bright-band bottom Iguchi et al. (2021) is used to stratify the observations into five categories. The resulting mean reflectivity profiles and the associated mean precipitation profiles are shown in Fig. 4 for three of these categories, the other two (i.e. associated with slopes with absolute values larger than 0.75dB/bin) accounting for less than 10% of the total number of profiles. As seen in the figure, distinct mean reflectivity profiles result in distinct mean precipitation profiles. This behavior may used to derive more accurate surface precipitation estimates than those derived from Eq. (1). However, to make effective use of the reflectivity slope and other such features, questions regarding the optimal strategy to calculate the slopes and partition them by value, especially when

the ground clutter extends close to or above the freezing level, need to be addressed. As there is no obvious strategy to address such questions, we resort to a ML approach.

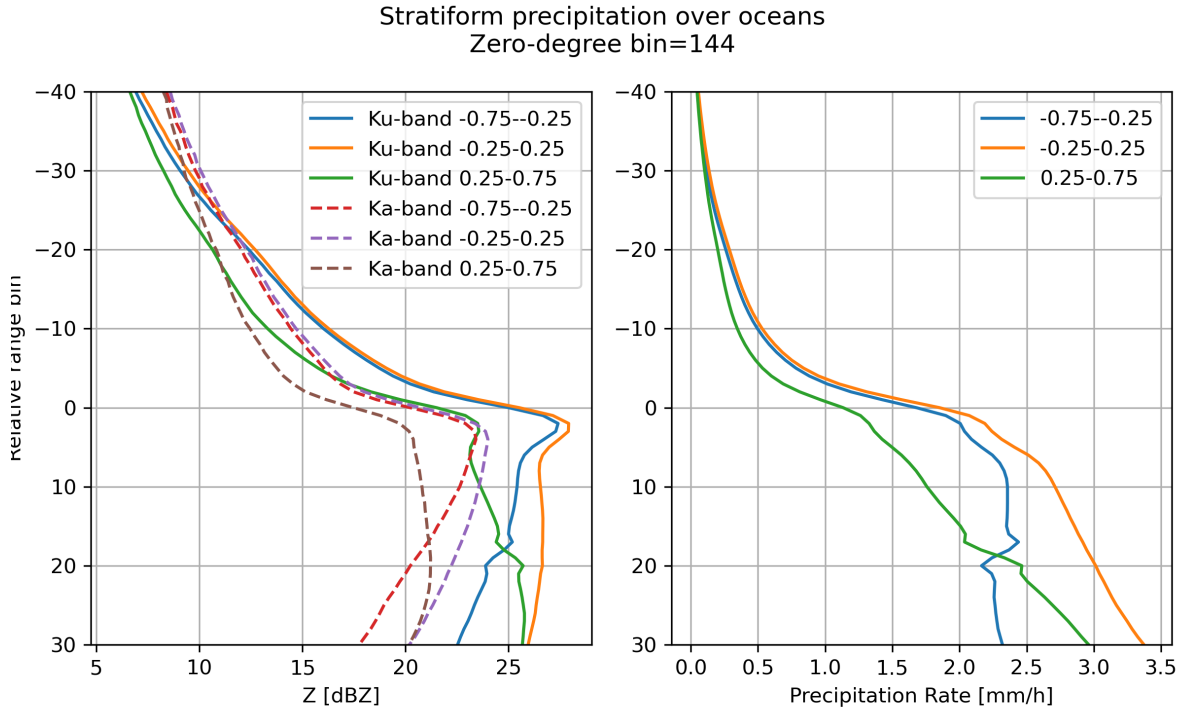



FIG. 4. Same as 2 but for a zero-degree bin of 144 and stratified by reflectivity slopes. The dashed lines in the left-hand side panel of the figures indicate the conditional mean reflectivity profiles at Ka-band associated with the three classes.

As previously mentioned, ML approaches do not require the explicit use of Eq. (1) and the stratification of the dataset by manually designed and optimized features. Instead, they require the organization of the dataset into a design matrix and a response matrix (Bishop and Nasrabadi 2006). In machine learning, the concepts of design and response matrices are borrowed from regression analysis, with the design matrix representing the array of predictor variables, while the response matrix representing the array of predicted variables. Each row of the design matrix corresponds to a single observation or data point, while each column represents a different predictor variable or feature.

In our study, the design matrix is an array of reflectivity observations and associated information, with each row containing the reflectivity values from a fixed-size portion of an observed profile. In addition to the reflectivity information, the zero degree bin, the position of the lowest clutter-free

**Line 188: A table summarizing the data going in/coming out of the models would be really helpful to keep track their shapes, data types, origins, units, number of samples, reference papers etc.**

bin (LCFB) bin relative to the zero degree bin, the position of the surface relative to the zero degree bin, and the LCFB precipitation rate are included in the design matrix. To make the ML models computationally efficient, the number of reflectivity observations above the LCFB is set to 30. Larger numbers of reflectivity observations above the LCFB were tested, but did not result in improved results.

**1. Line 196-197: For clarity, where does the surface precipitation rate come from, and what is its spatio-tempor resolution? Is this one of the derived GPM products? It would be interesting to see these results relate to in sit surface rates at a series of different stations over land, but that would likely be beyond this project.**

The response matrix is one-dimensional, i.e. a vector, and it contains the associated surface precipitation rates. As explained above, the profiles in the training/evaluation dataset are characterized by at most eight bins affected by clutter. Although minimally affected by clutter, there are no surface estimates in the profiles of the training dataset, initially, and the precipitation estimates associated with the clutter-free observations need to be extended to the surface. To achieve this, for every profile with a FL height greater than 1.5 km above sea level, we regress the precipitation rate against range using the estimates associated with the lowest four clutter-free bins and employ the resulting regression to estimate the precipitation down to the sea level (next eight bins). It should be mentioned that this extrapolation does not eliminate the need for more comprehensive methodologies (or make them superfluous), because at higher incidence angles, the ground clutter has a more significant and complex impact on profiles than the eight bins contaminated by ground-clutter in the training dataset. When the FL is below 1.5 km, the precipitation slope may not be reliably derived, as the possible existence of ice-phase and melting precipitation in the four lowest clutter-free bins may significantly affect the vertical distribution of precipitation rates. For such profiles, we use a k-Nearest Neighbor (k-NN) (Friedman 2001) approach to extend the precipitation estimates into the clutter. Specifically, given that bin $n_1$ in a profile with freezing level height $FLH_1$ is roughly characterized by the same temperature as bin $n_1 - dn$ in a profile with freezing level height $FLH_2$, where $dn$ is the integer part of $\frac{FLH_2 - FLH_1}{125\,m}$, we search for the k nearest neighbors of a profile with $FLH_1 < 1.5\,km$ among profiles with $FLH_2 = FLH_1 + 1.0\,km$. The proximity is evaluated using the Euclidean distance in a system of reference relative to the zero degree bin. Then we use the $k$ nearest neighbors to fill information in the clutter region, which is possible because bins affected by clutter for profiles with a given freezing level height $FLH_1$ are clutter-free in profiles with a freezing level height greater by 1.0 km, i.e $FLH_2 = FLH_1 + 1.0\,km$. While neither the slope-based extrapolation nor the reflectivity-based extension are error free, they nevertheless provide a reasonable methodology to extrapolate nearest surface precipitation rate rates in the

**Lines 200-230: It would also be helpful if you provided a simple model diagram/flow chart showing how the inputs are used to predict the surface rates across the different ML models versus the kNN interpolation or simple linear extrapolation techniques.**

training data down to the sea level. Alternative methodologies based on cloud-resolving models (CRMs) and ground-based radar observations are possible. However, these methodologies are not necessarily bias- or complication-free, because CRMs may exhibit microphysical biases, while the derivation of CRM simulations and the collection of ground observations representative of the global distribution of precipitation events are extremely laborious processes. While CRMs and ground observations may be able to eventually provide better (more complete) datasets to develop methodologies to mitigate ground clutter in space-borne radar observations, the approach in this study is still needed because the number of bins affected by ground-clutter is significantly greater than eight (which is the number of clutter-affected bins in the training dataset in our study).

The structured organization of the dataset makes it possible to explore multiple ML models with minimum effort and select the optimal one. While ML models are generally physics-agnostic in the sense that they do not explicitly make use of physical laws, they can exploit physical causality embedded in the dataset. For example, if slopes of the reflectivity profiles above the clutter are reliable predictors of the precipitation rate at the surface relative to the lowest clutter-free precipitation rate(as suggested by Fig.4), then an ML model based on the k-NN (Friedman 2001) will be able to exploit this causality because similar reflectivity profiles in the design matrix result in similar slopes. However, potentially more accurate or computationally more efficient ML models may exist, and so in addition to the k-NN model, we also consider a Gradient Boosting (GB) model (Friedman 2001), and a random forest (RF) model (Ho 1995). Both the GB and RF models are based on decision trees (DTs) that are built through a process that constructs a tree-like structure by recursively splitting the dataset based on feature conditions (Bishop and Nasrabadi 2006). However, while the GB model starts with a weak DT and iteratively adds DTs to minimize residuals, the RF model derives an ensemble of DTs and uses their average for prediction. The k-NN and RF model implementations used in the study are based on the scikit-learn library (Pedregosa et al. 2011), while the GB model is based on the efficient implementation of Ke et al. (2017). In addition to the three scikit-learn based models, we consider a neural network (NN) model (Goodfellow et al. 2016; Géron 2022) based on the TensorFlow library (Abadi et al. 2016).

The scikit-learn library provides a convenient interface to train and test ML models. As such, the definition of the scikit-learn ML models requires minimal specifications. They include the number of neighbors for the k-NN model and the number of trees for the random forest model,

. Additionally, how were the hyperparameters on lines 255-260 determined? Did you use Bayesian search, hyperband, etc.? A parameter sweep might be beneficial and bring the NN closer to the performance of the LGBM. For the RF/LGBM models, you mention default parameters. Does this mean max_depth was set to None? If so, there may be overfitting concerns, and adjusting these values could help. In general, I feel that a more in-depth examination of the ML architectures (and performing a few additional tests) while addressing these specific questions would provide valuable context when comparing to simpler linear extrapolation techniques.

and the number of estimators for the gradient boosting model. Similarly, the Light Gradient Boosting Model (LGBM) implementation of Ke et al. (2017) requires minimal specifications. We set the number of neighbors to 20 and use the default values for the RF and LGBM models. The TensorFlow library, on the other hand, requires a more specific definition of the model architecture. We use a simple fully connected feed forward neural network with two hidden layers. The number of neurons is set to 32 in both hidden layers. The activation function is the rectified linear unit (ReLU) (Nair and Hinton 2010) and the output layer is a linear layer. The loss function is the mean squared error (MSE), and the optimizer is the Adam optimizer (Kingma and Ba 2014). The TensorFlow model is included in the study to provide insight into how the complexity of the ML model affects performance. Specifically, while the scikit-learn models and the LGBM are relatively simple, the NN model is more complex (Goodfellow et al. 2016) and may be able to capture more complex relationships between the reflectivities above the clutter and the surface precipitation rate.

Shown in Fig. 5 is the cumulative distribution of number of bins affected by clutter for rays in the DPR's outer swath. As apparent in the figure, more than eight bins are affected by clutter for the vast majority of the DPR outer swath profiles over land, with about half of profiles characterized by more than 15 bins affected by clutter and 10% of profiles characterized by more than 26 clutter-affected bins. The dataset of observations and precipitation profiles minimally affected by clutter and extended to sea-level, as explained above, is used to develop and test the different ML model approaches. To simulate clutter effects, $n_c$ bins are assumed affected by clutter, where $n_c$ is a random integer uniformly distributed between 1 and 26. The upper limit 26 was chosen based on the results shown in Fig. 5. While about 10% of profiles exhibit more than 26 bins affected by clutter, a larger upper limit might result in biases in the estimation because the distribution of the number of pixels affected by clutter is not uniform. A sampling strategy consistent with the cumulative distribution function in Fig. 5 may be used, but it would unnecessarily increase the size of the training dataset, because the sample size necessary to mitigate noise for $n_c > 26$ would result in significant oversampling for $n_c \leq 26$. However, given that the LGBM and the NN models have good extrapolation capabilities, and the number of profiles with $nc > 26$ is relatively small, deriving ML models for $n_c \leq 26$ and applying them for $n_c > 26$ is not necessarily a poor choice. Moreover, an additional set of models trained exclusively for $n_c > 26$ may be derived. However, it is beneficial to first systematically investigate the performance and behavior of ML for $nc \leq 26$.

Lines 269-280: I feel this description regarding the selection of the upper limit 26 is a bit verbose, and can likely be condensed.

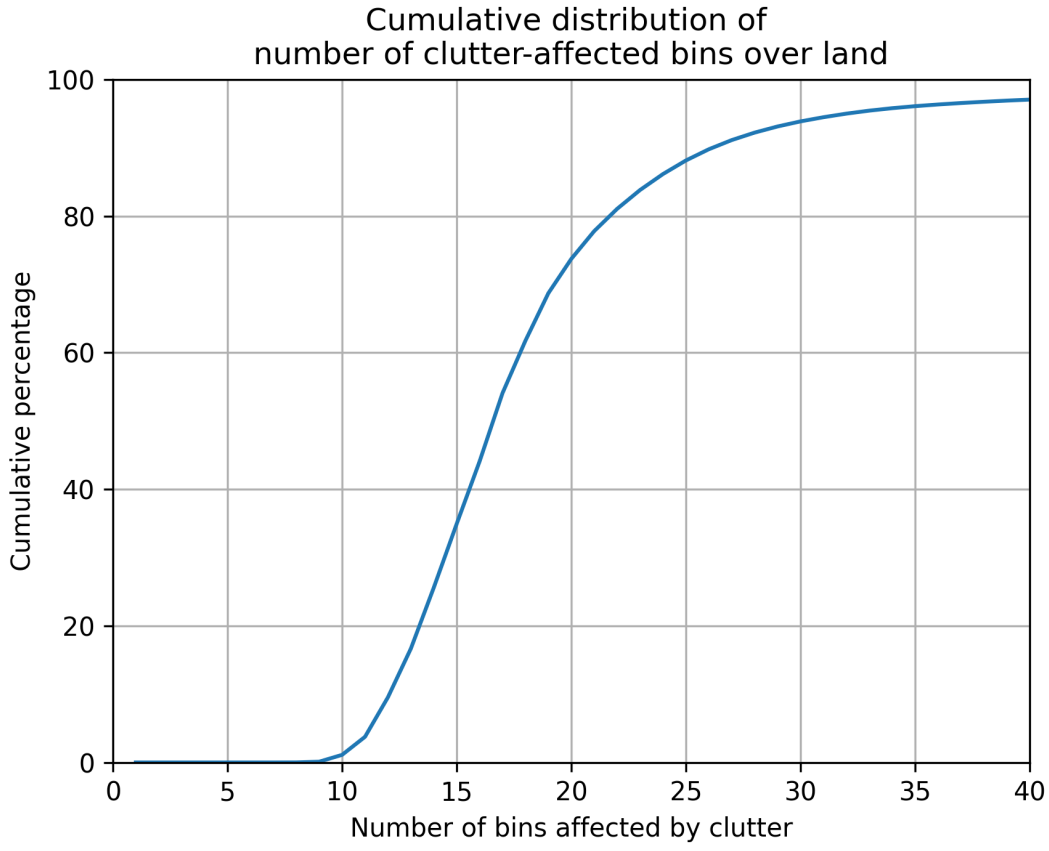



FIG. 5. The cumulative distribution of number of bins affected by clutter for rays in the DPR's outer swath (defined as the portion of the swath within 12 rays from the edges) over land.

**Another major comment is regarding the sample size and data splitting. It wasn't totally clear to me, but was only one year of GPM data being used for training/testing? While one year of GPM data is substantial, there is much more available, and I feel incorporating some of this could potentially impact performance/results. More sophisticated techniques like the NN or LGBM often strongly benefit from a larger dataset by capturing additional complex nonlinear features missed by linear techniques, potentially reducing the noise in the first three columns of Fig. 10. This is especially true for NN's with multiple hidden layers. Expanding to multiple years might boost NN performance and improve predictions over different seasons as weather patterns shift year-to-year. Additionally, when you state, "the DPR dataset is randomly split into a training and a testing dataset with 70% of profiles in the training dataset and the remaining 30% in the testing dataset," what do you mean by profile? If you mean the vertical reflectivity bins for a single observation from the DPR, I would recommend not splitting randomly by the vertical profiles (e.g., training on profile at timestep X and testing on timestep X+1 may lead to overfitting). Instead, I would recommend a paradigm of training on the first 80% of a month, holding the next 10% of days in a validation dataset, and the last 10% in the test dataset (or something equivalent). These types of splitting methods are much less prone to overfitting, as the atmosphere decouples much more strongly between training/testing datasets.**

To evaluate the performance of the ML models, we use a cross-validation approach. Specifically, the DPR dataset is randomly split into a training and a testing dataset with 70% of profiles in the training dataset and the remaining 30% in the testing dataset. The training dataset is used to optimize the ML models, while the testing dataset is used to evaluate them. The evaluation is based on calculations of the correlation coefficient and bias between the predicted and observed surface precipitation rates.

## 3. Results

The reason for considering several ML model architectures is to ensure that there is no latent information in the input data that is not properly captured. The inclusion of multiple ML models reduces the likelihood of such a possibility, as the models are based on different statistical modeling paradigms. However, in our initial model testing, no particular ML model emerged as significantly better than the others. This outcome, which is not totally surprising, may be an indication that the relations between the surface precipitation rate and the precipitation rate at a given height above the surface depend on a multitude of factors that cannot be directly observed or do not have a clear signature in the reflectivity observations. Nevertheless, some models are preferable to others. Specifically, the k-NN model is rather slow in applications, as it requires searches through its supporting database. The NN model on the other hand is slow in training and prone to overfitting (i.e. it tends to produce smaller errors in training, but large biases in the evaluation process). The LGBM and RF models have similar performances, with significantly lower computational costs associated with the LGBM model. Therefore, based on this initial testing, we choose the LGBM as the best option, and instead of exploring additional methodologies or carrying out further tuning, we focus on characterizing its performance, especially in relation to a simple estimation methodology.

### *a. Stratiform precipitation over land*

Before describing the performance of the different ML estimation methods, we will first examine the persistence solution as a benchmark. In this simple solution, the precipitation rate at the LCFB is assumed to be the same as the surface precipitation rate. Shown in the left-hand side panel of Fig. 6 is the correlation coefficient between the actual surface precipitation rate and the LCFB precipitation rate up to 26 bins above the surface for stratiform precipitation events over land. The vertical axis is the difference between the surface bin and the LCFB, and the horizontal axis represents the zero-degree bin. As seen in the panel, the correlation decreases with the position of the LCFB above the surface. The bins marking a more significant correlation decrease (from above 0.8 to below 0.75) generally occur in the mixed and ice phase. The biases associated with the LCFB-derived precipitation rate relative to the surface precipitation rate are shown in the right-hand side panel of Fig. 6. This type of estimation is referred to as persistence in the figure

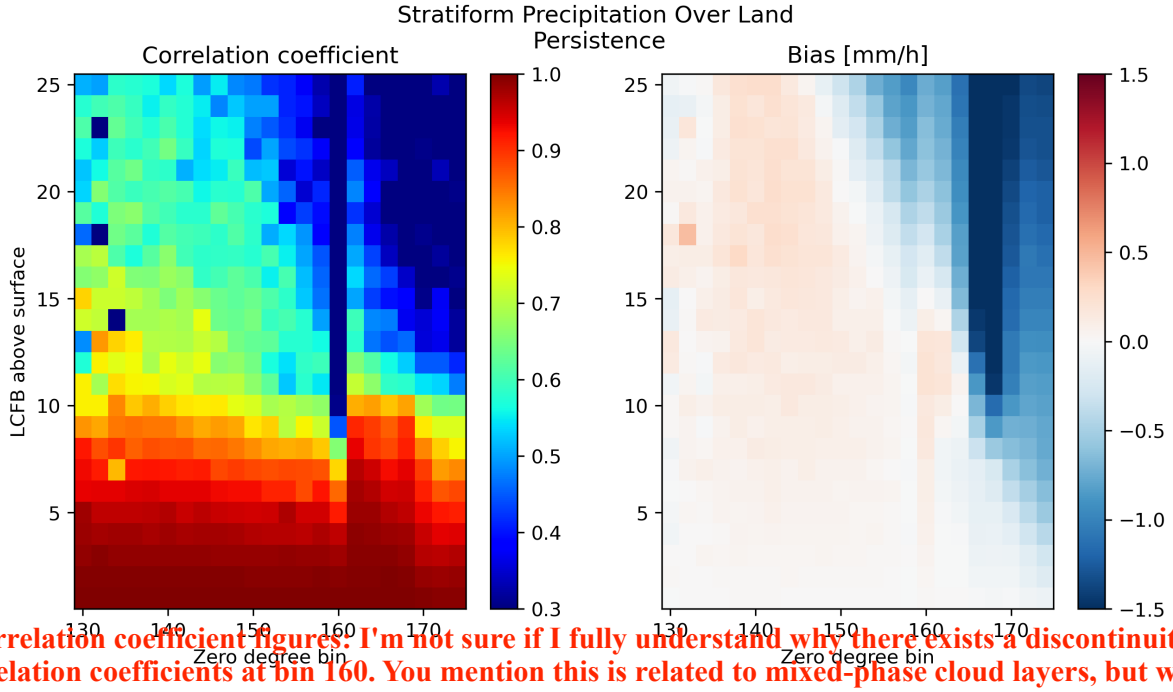
**Line 315: It might be useful to provide a sentence or two explaining the importance of the zero-degree bin in the context of the surface precipitation rate, as it isn't totally clear as to why this is such an important variable to consider from the outset of the paper.**

and henceforth. Similar to the correlation coefficient, the largest biases occur when the LCFB is in the ice phase. The correlation coefficients exhibit a discontinuous distribution for profiles with a zero degree bin near 160, while the biases exhibit a relatively continuous distribution for profiles with a zero degree bin near 167. This behavior is likely a consequence of the fact that precipitation estimates in the mixed layer may be biased and noisy, and this may impact the procedure used to fill in the precipitation estimates in the eight bins affected by clutter in the database. At the same time, the DPR detection capabilities deteriorate for profiles with only snow above the clutter or if the melting layer is close to the clutter.

Unlike persistence-based estimates, surface precipitation estimates based on Eq. (1) would be bias-free (assuming that the precipitation climatology is bias-free in the training dataset). However, the distribution of correlation coefficients between the estimates and the true surface values would not be different from that shown in Fig. 6. In other words, systematic errors are zero in estimates based on Eq. (1), but the random differences remain largely the same. An estimation superior to bias removal would also show an improvement in the distribution of the correlation coefficients and an overall reduction in the root mean squared error (RMSE). Shown in Fig. 7 are results for the LGBM method. As seen in the figure, the correlation coefficients increase slightly relative to those shown in Fig. 6, while biases are almost zero. In particular, the biases in the ice phase associated with the persistence-based estimates are largely removed. However, the marginal (at best) improvement in the correlations between the estimated surface precipitation rates and those in the databases suggests that there is significant variability of precipitation profiles in the clutter that cannot be reliably predicted from observations in the clutter-free portion of reflectivity profile.

*b. Convective precipitation over land*

Shown in Fig. 8 are the distributions of correlation coefficients and biases of the persistence-based estimator of surface convective precipitation over land. Results are qualitatively similar to those obtained for stratiform precipitation over land, but with larger biases when the LCFB is in the ice phase. Some positive biases for bins in the mixed phase are also obvious. These biases are most likely the consequence of artifacts in the precipitation estimates across the melting layer




**Correlation coefficient figures. I'm not sure if I fully understand why there exists a discontinuity in the correlation coefficients at bin 160. You mention this is related to mixed-phase cloud layers, but would this layer not shift slightly in its vertical position over time? Further, why is there a bimodality in the coefficients, with a spike in high correlations near the surface and then another spike around bin 170? What is the physical intuition behind this distribution?**

FIG. 6. Performance of a persistence-based clutter mitigation method for stratiform precipitation over land. The left-hand-side panel shows the correlation coefficient of the surface precipitation rates with the precipitation rates in the LCFB (which serves as the surface estimate in the persistence-based scheme), while the right-hand-side panel shows mean differences between the surface precipitate rates and the LCFB precipitation rates. Values are plotted for different surface vs. LCFB bin differences (vertical axis) and for different zero-degree bins (horizontal axis).

due to use of different reflectivity/precipitation lookup tables. The distributions of correlation coefficients and biases associated with the LGBM for convective precipitation over land are shown in Fig. 9. As seen in the figure, both the correlation coefficient and the bias improve relatively to results in Fig. 8. However, the bias distribution exhibits more variability around zero than the bias associated with stratiform precipitation over land. This is most likely a consequence of convective precipitation exhibiting more vertical variability while being about five times less frequent than stratiform precipitation. This makes the statistics of convective precipitation profiles in the training dataset noisier than those of stratiform precipitation. Noise can be mitigated by extending the dataset through inclusion of DPR observations and CORRA estimates from other periods, but it is likely that a portion of the noise is caused by artifacts due to multiple scattering and non-uniform

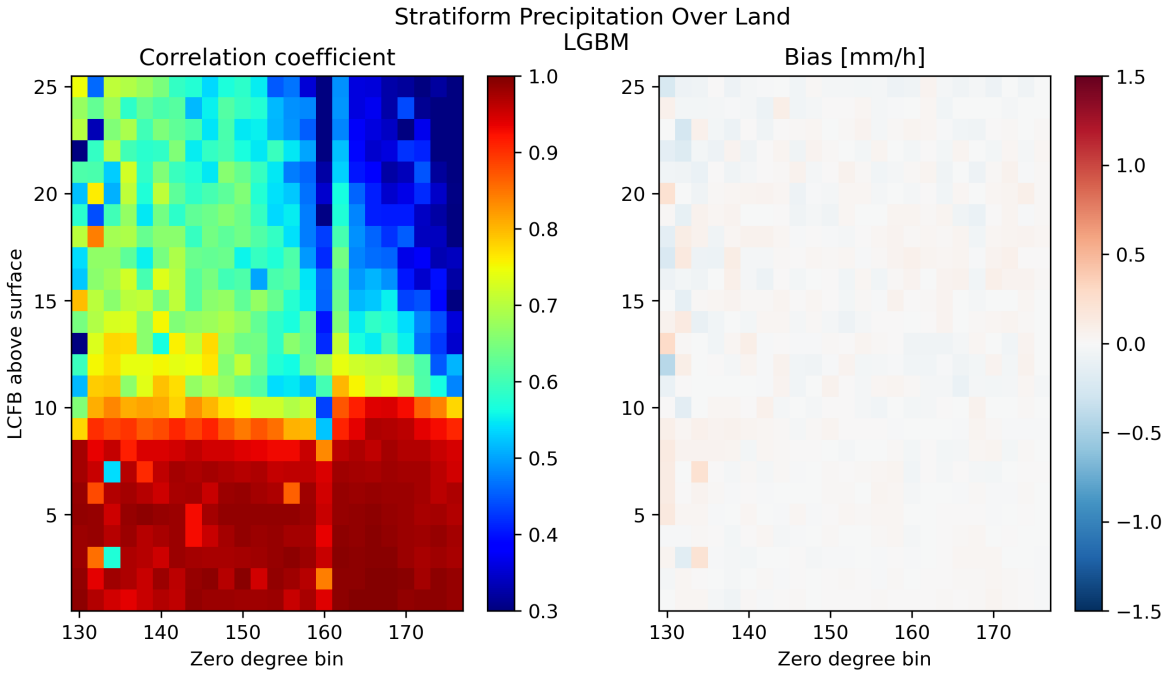



FIG. 7. Performance of the LGBM clutter mitigation method for stratiform precipitation over land. That is, same as Fig. 6 but for LGBM surface precipitation rate estimates instead of the persistence-based estimates.

beam filling in the precipitation estimation procedure. From this perspective, it is beneficial that convective dataset extension be considered at the same time with or after a refinement of the convective precipitation estimation methods in CORRA.

### *c. Precipitation over oceans*

The statistics for precipitation over oceans are qualitatively similar to those over land; see Figs. 10 and 11. The most significant difference is that, as suggested by Figs. 3 and 4, the mean precipitation profiles have different shapes, with the oceanic precipitation generally exhibiting more systematic increases with range below the freezing level than precipitation over land. However, the LGBM clutter correction schemes exhibit behaviors similar to those over land for both stratiform and convective precipitation types. This is shown in Fig. 10 for stratiform precipitation and in Fig. 11 for convective precipitation. The LGBM model for stratiform precipitation appears rather noisy

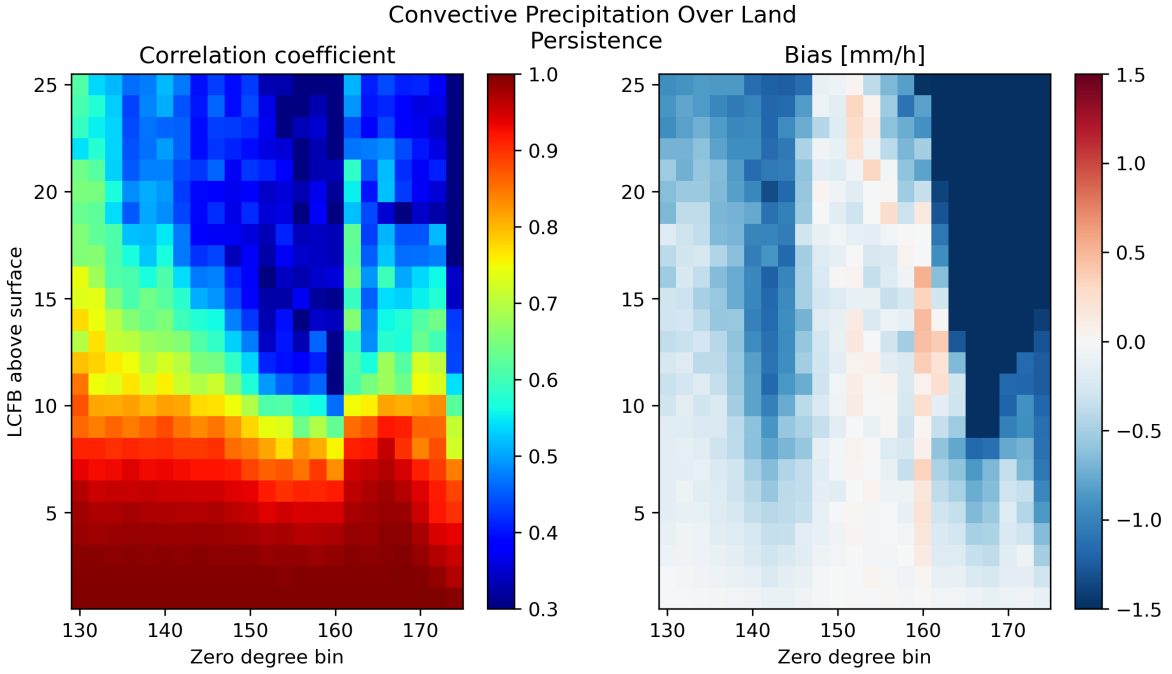



FIG. 8. Same as Fig. 6 but for convective precipitation over land.

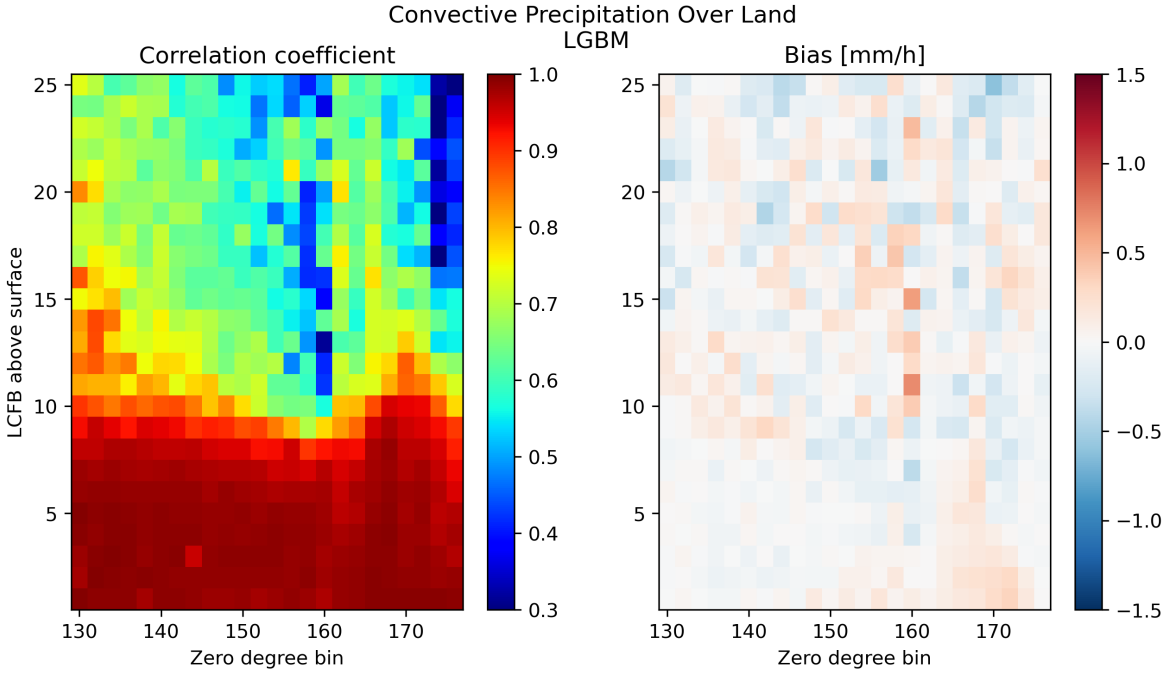



FIG. 9. Same as Fig. 7 but for convective precipitation over land.

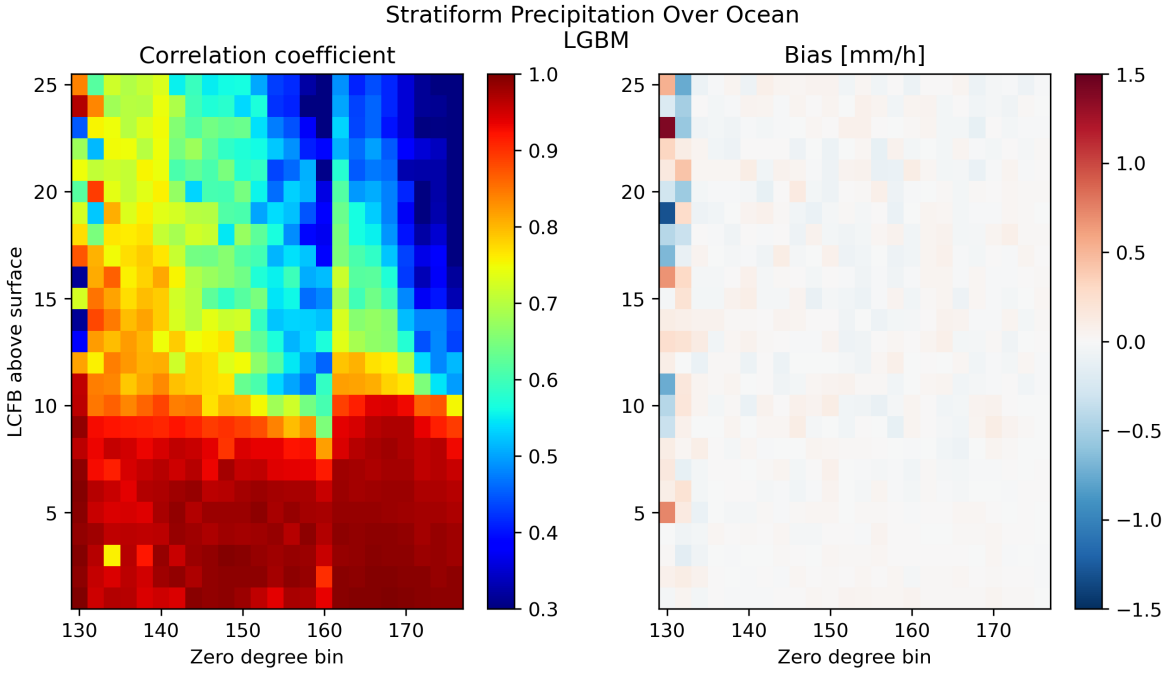



FIG. 10. Performance of the LGBM clutter mitigation method for stratiform precipitation over oceans.

for profiles with a zero degree bin less than 133 and an LCBF more than 4 bins above the surface, which is most likely a consequence of the relatively small number of such profiles in the training dataset. On the other hand, the LGBM model appears noisy in general for convective precipitation, which is consistent with its behavior over land.

### *d. Evaluation of the random errors in the correction*

In the previous section, the LGBM method was shown to be effective in removing the biases associated with the persistence-based estimates. However, the random errors in the estimates were not necessarily reduced. Specifically, the correlation coefficients between LGBM surface precipitation estimates and actual surface precipitation estimates did not appear to be improved relative to those associated with the persistence-based estimates.

To investigate this quantitatively, we calculate the relative RMSE associated with both a climatological scaling correction based on Eq. (1) and the LGBM estimates. The relative RMSE involves

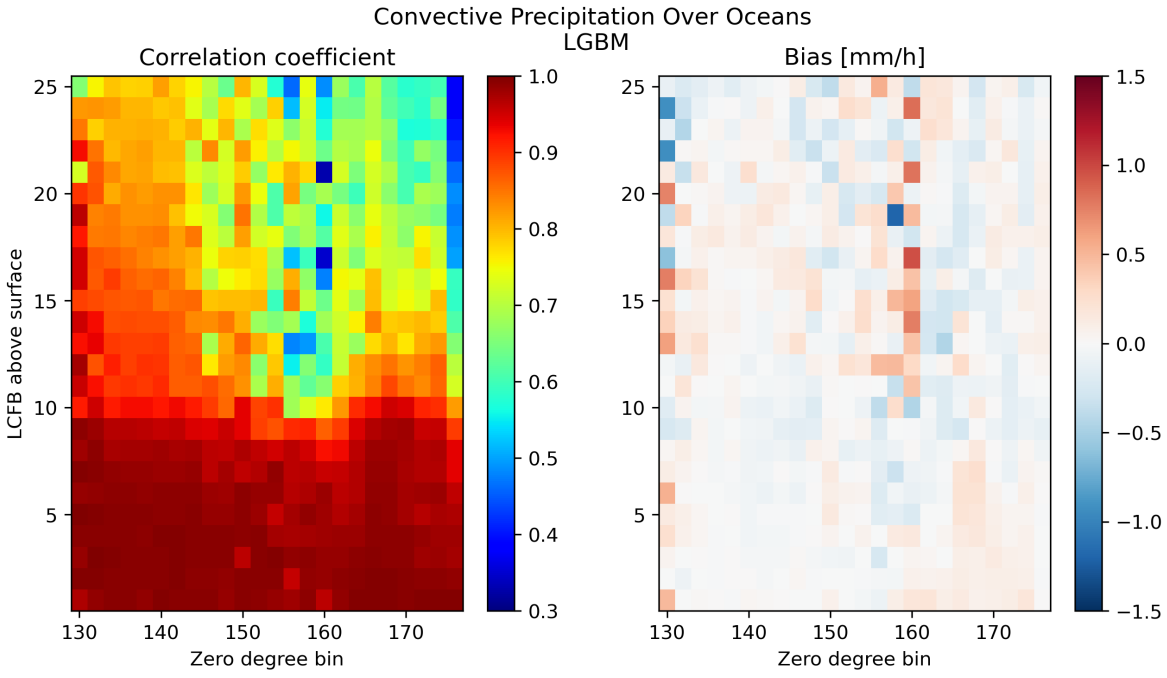
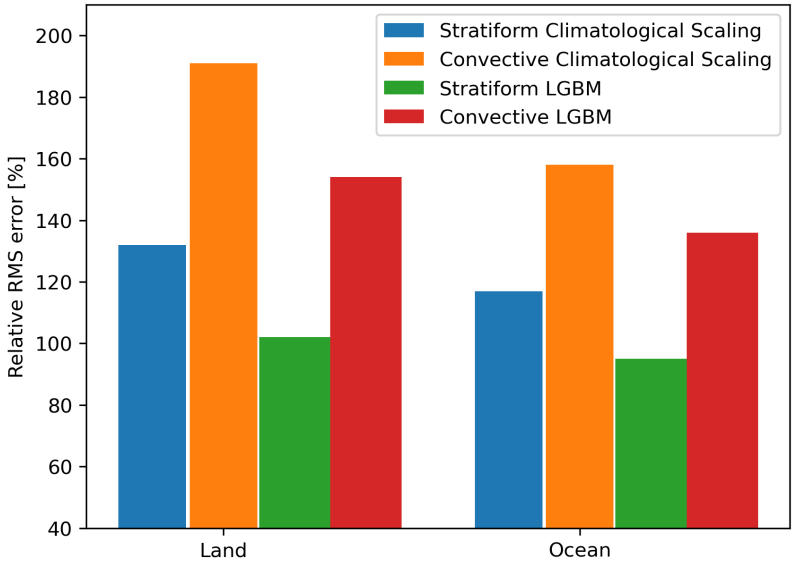



FIG. 11. Same as in Fig. 10 but for convective precipitation.

normalization by the standard deviation of the conditional surface precipitation rates. Results are shown in Fig. 12. Here, the LGBM estimates do not exhibit RMSEs that are much smaller than those of the climatological scaling estimates. This suggests that a simple bias-removal methodology based on Eq. (1) in section a is likely to be satisfactory in many respects. Nevertheless, the application of the LGBM method results in some RMSE reduction. As expected, the relative RMSEs are greater in convective than in stratiform precipitation and greater over land than over oceans. The fact that the LGBM method (which is representative of a broader class of one-dimensional clutter mitigation ML techniques) does not result in significant improvements relative to the simple bias correction provided by Eq. (1) is not necessarily an indication that ML techniques offer no benefit to the clutter mitigation problem. One potential advantage of the ML techniques is that they can incorporate radiometer observations, which may yield a significant benefit in the estimation of light precipitation over oceans. Also, the correction methods explored in this study as well as in the previous work of Hirose et al. (2021) make exclusive use of profile-level information. However, modern deep learning architectures such as U-Nets (Siddique et al. 2021) can readily process 3D information that may be useful for identifying the impacts of phenomena such as the wind shear on

**Lines 405-409: This is an interesting application and I would recommend making reference to a recent paper by King et al., 2024 that looks at this topic in more detail (i.e., using a U-Net with spaceborne radar and additional atmospheric variables to reconstruct blind zone reflectivity profiles).**




410 FIG. 12. Relative RMSE for both the persistence and LGBM methods as a function of precipitation and surface
411 type.

408 reflectivity observations and use this kind of information to more accurately predict the distribution
409 of precipitation in the clutter. These topics will be explored in future studies.

## 412 4. Application to GPM CORRA precipitation estimates over the Continental US in the cold
413 season

414 To investigate the impacts of clutter mitigation on the estimation of precipitation over the Conti-
415 nental US (CONUS) in the cold season, we apply the LGBM method to all GPM CORRA retrievals
416 over CONUS from 1 December 2021 to 28 February 2022. While the same type of analysis can
417 be applied to the entire GPM domain over all seasons, given that the focus of this study is on
418 fundamental benefits and limitations of profile level corrections rather than their climatological
419 impact, we limit our focus to a single region and season and defer more extensive analyses to future
420 studies. Only profiles with freezing levels below 1250 m are considered in the analysis because
421 they are given to the largest corrections (and errors in the absence of any correction), as the LCFB
422 may be associated with temperatures below freezing, while the surface precipitation may be rain.
423 Shown in Fig. 13 is the mean Ku-band reflectivity conditioned on the observed profiles being
424 classified as precipitating. The means are conditioned on the associated profiles being classified

**Line 419: Have you looked at cases where the model performed exceptionally/poorly? Specifically, I am curious about extreme precipitation events? Typically, these models prefer to stay safe with their estimates and regress towards the mean value, and I am curious if the associated problems are evident. I am curious if you see that pattern here as well?**

**Have you examined how this model performs with orographic precipitation? It might be good to mention something about this (even just as follow-up work), as it is a very active area of research.**

as precipitating. As seen in the figure, the region contaminated by clutter (characterized by large reflectivity values) increases in height with the incidence angle. Some artifacts related to the processing of the received power to mitigate sidelobe clutter (Kubota et al. 2016) are also apparent in the figure. Specifically, while some enhanced echo is visible above 4.0 km, a slight reduction in the reflectivities is apparent near the center of the swath (roughly from ray 20 to ray 30). The reduction is more significant below the average height of the LCFB (blue line in the figure), but that reduction does not directly impact the precipitation estimation, as the pixels associated with it are classified as clutter.

Shown in the top panel of Fig. 14 are the conditional near-surface precipitation estimated by CORRA and the surface precipitation predicted by the LGBM method. As seen in this figure, the clutter mitigation methodology has a significant impact on the precipitation estimates, with the impact increasing from center towards the edges of the swath. This behavior is, most likely, a consequence of the fact that the DPR's detection capabilities deteriorate near the edges of the swath for precipitation systems with low FLH. The detected profiles are fewer but are characterized by more intense (and deeper) precipitation that results in reflectivity observations that can be reliably distinguished from clutter. This hypothesis is consistent with the distribution of the number of precipitation profiles as a function of ray, shown in the bottom panel of Fig. 14. However, the clutter correction technique does not result in artificial increases of intensity with distance from the center of the swath in the overall (unconditional) precipitation rate. This point is illustrated in Fig. 15. Instead, the opposite effect, i.e. a reduction of the unconditional precipitation rate estimates with distance from the swath center (consistent with the DPR precipitation detection capabilities near the edges of the swath), is apparent in the figure. The overall impact of the LGBM surface precipitation rate estimation procedure is significant for precipitation systems with freezing level heights below 1250 m over CONUS. This is an indication that significant precipitation growth processes such as water vapor deposition and riming occur in the clutter region.

## 5. Summary and Conclusions

In this study, a new method for mitigating ground clutter effects in precipitation estimates derived from the GPM mission's CORRA algorithm is developed. CORRA combines data from the DPR

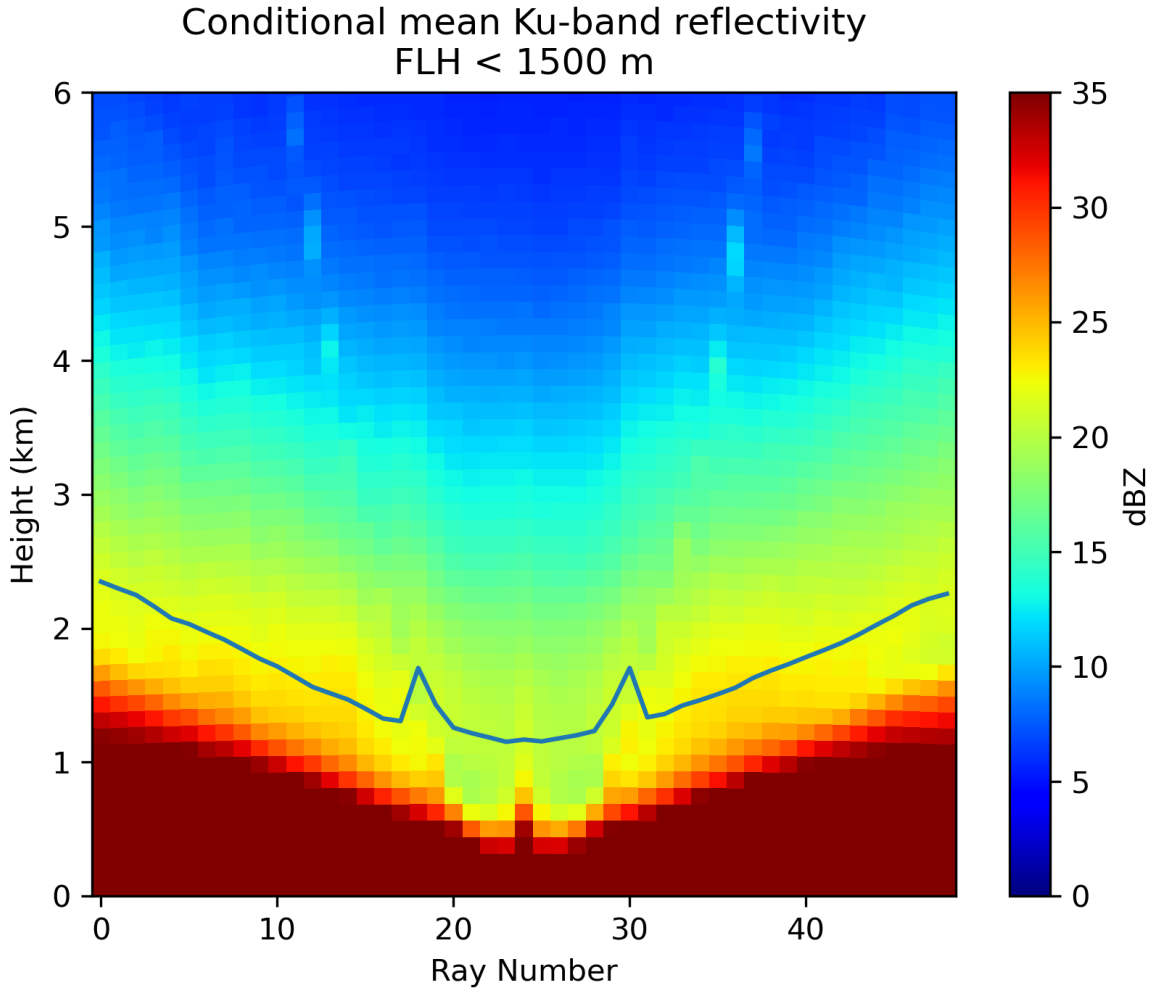



FIG. 13. Mean Ku-band reflectivity conditioned on the observed profile being classified as precipitating. The blue line indicates the average height of the LCFB.

and GMI on the GPM core satellite to estimate precipitation rate, and ground clutter is a significant problem for spaceborne radar observations, as it can obscure or corrupt the signal associated with precipitation. An approach to mitigate ground clutter using statistical relationships based on precipitation estimates from near-nadir scans has already been developed (Hirose et al. 2021) and applied to precipitation estimates from the DPR algorithm (Iguchi et al. 2021). However, the study of Hirose et al. (2021) did not fully explore the benefits and limitations of statistical methods to mitigate clutter in the DPR reflectivity observations.

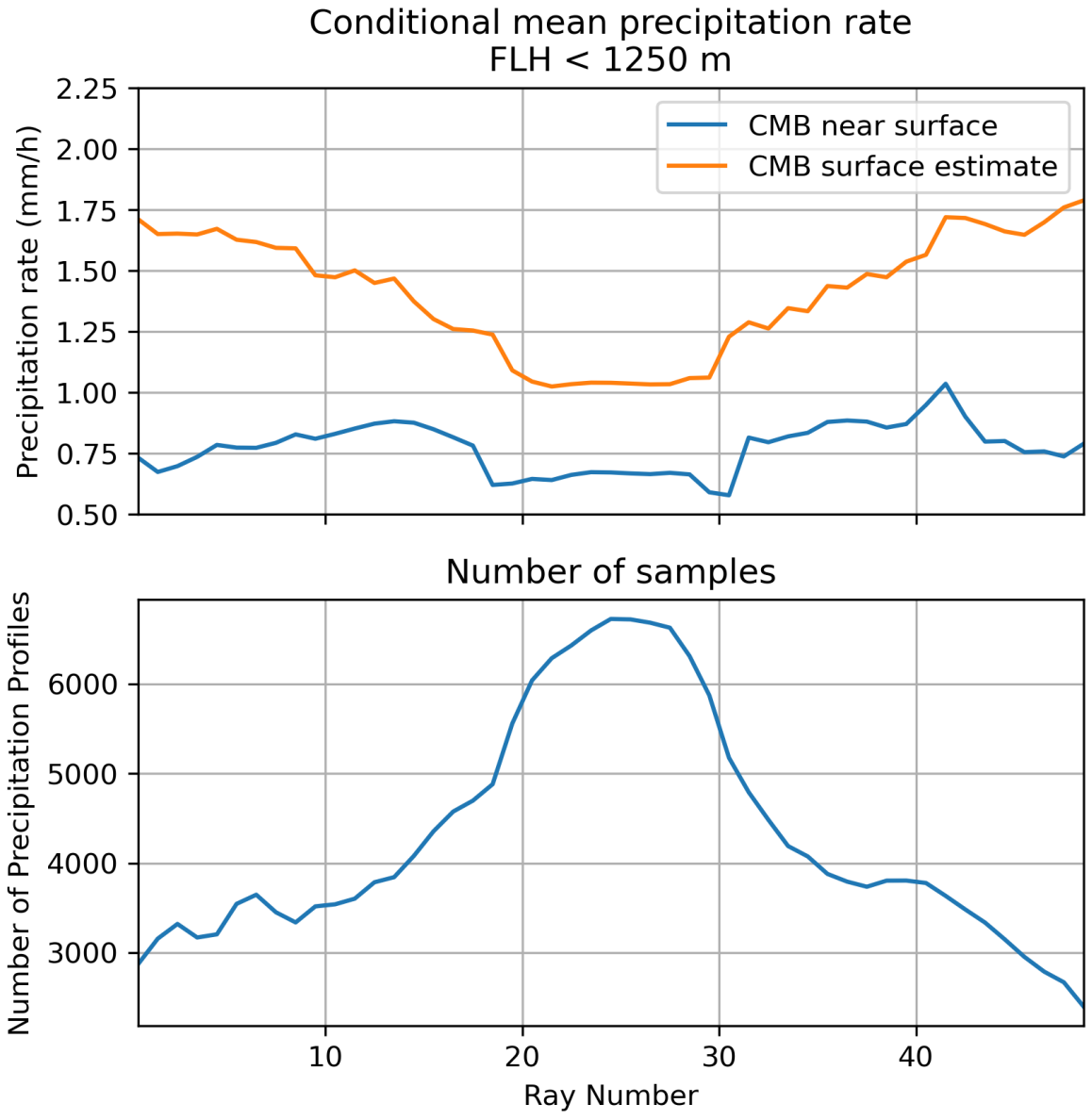



FIG. 14. Top panel: Conditional near-surface mean precipitation rate from the CORRA and the surface mean precipitation rate predicted by the LGBM method. Bottom panel: Number of detected precipitation profiles as a function of ray index.

To build upon the previous work, ML approaches are investigated to gain further insight into the uncertainties of surface precipitate rates derived from information in the portion of the reflectivity profile not affected by clutter. The ML model uses reflectivity observations, along with additional information such as precipitation type, surface type, and freezing level, to estimate the surface

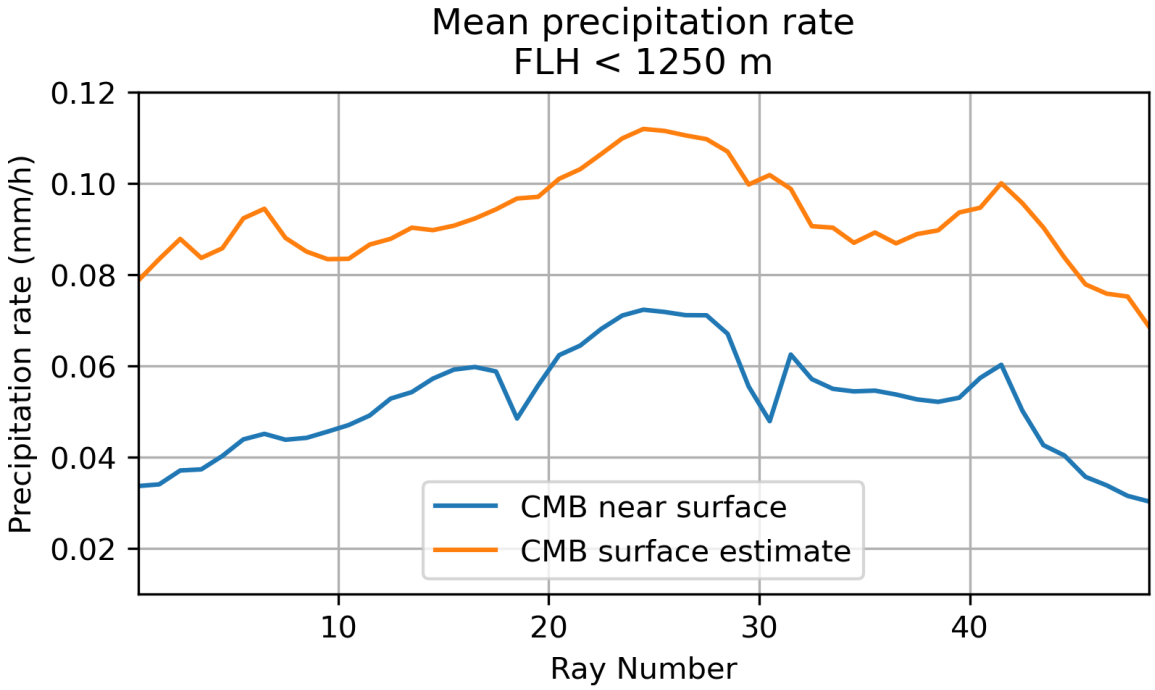



FIG. 15. Top panel: Near-surface mean precipitation rate from CORRA and the surface mean precipitation rate predicted by the LGBM method.

precipitation rate. The benefits of this approach include the use of ML models efficient at leveraging existing features and capturing complex relationships within the data without relying on explicit feature engineering, and systematic evaluation of estimates is also facilitated. Specifically, various machine learning architectures are investigated to automatically extract information from the data without resorting to subjective efforts. A preliminary evaluation suggests that no architecture offers significantly better performance than the others, and so we select the Light Gradient Boosting Model (LGBM) of Ke et al. (2017) as the best candidate for further systematic evaluations, since it is computationally fast to train and deploy while effective in application.

The database used in the training and evaluation of the ML models in the study is derived from one year of DPR near-nadir observed reflectivity profiles that are minimally affected by clutter. A minor deficiency of this database is that while the number of bins affected by clutter is small, one still needs to resort to statistical models and assumptions to derive the surface precipitation estimates. For profiles with a FLH greater than 1.5 km, characterized by a sufficiently large number of observations not affected by clutter in the liquid phase, we simply use linear extrapolation to estimate the surface

precipitation rate. For all the other profiles, a k-NN method is used. Specifically, the k-nearest neighbors of a profile are sought among profiles characterized by the same displacement of the LCFB relative to the zero degree bin as that of the surface bin relative to the zero degree bin for the profile in question. The proximity is evaluated using the Euclidean norm in a system of reference relative to the zero degree bin. While the estimation of surface precipitation in the database construction may introduce non-negligible random errors, and even biases, it is most likely preferable to limiting the clutter mitigation to only clutter that extends greater than 1.0 km above the surface (the extent of clutter at nadir view). Nevertheless, this postulate needs to be evaluated in further studies. Over land, high-quality ground radar precipitation estimates adjusted by rain-gauges such as those provided by the MRMS product (Zhang et al. 2016) may be used for this purpose. The evaluation is likely to be more challenging over oceans, as data useful for direct validation of estimates is very limited.

Estimates that use the LCFB as a proxy for surface precipitation rate are systematically different from the surface precipitation rates in the training data, prompting an assessment of the LGBM model within this context. Specifically, using the LCFB's precipitation rate as an estimate of surface precipitation rate yields biased results, and so the LGBM model's capacity to produce unbiased estimates is scrutinized. The LGBM model demonstrates effectiveness in providing unbiased estimates, yet does not appear more capable of mitigating random errors than a basic climatological scaling method. Furthermore, the performances of other machine learning techniques like k-nearest neighbor, random forest, and feedforward neural networks mirror that of the LGBM model in initial assessments, implying that the LGBM model's inability to improve upon bias removal stems from the nature of the problem rather than inherent limitations of the approach.

## 6. Acknowledgments.

This work was supported by the NASA Global Precipitation Measurement Mission (PMM) project. The authors thank Drs. Tsengdar Lee and Will McCarty (NASA Headquarters) for their support of this effort.

## 7. Data availability statement.

The version 7 of GPM DPR and CORRA data can be accessed online (https://https://arthurhouhttps.pps.eosdis.nasa.gov/gpmdata/).

## References


Abadi, M., and Coauthors, 2016: Tensorflow: a system for large-scale machine learning. *Osdi*, Savannah, GA, USA, Vol. 16, 265–283.

Bishop, C. M., and N. M. Nasrabadi, 2006: *Pattern recognition and machine learning*, Vol. 4. Springer.

Friedman, J. H., 2001: Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Géron, A., 2022: *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*." O'Reilly Media, Inc.".

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep learning*. MIT press.

Grecu, M., W. S. Olson, S. J. Munchak, S. Ringerud, L. Liao, Z. Haddad, B. L. Kelley, and S. F. McLaughlin, 2016: The gpm combined algorithm. *Journal of Atmospheric and Oceanic Technology*, **33 (10)**, 2225–2245.

Hirose, M., S. Shige, T. Kubota, F. A. Furuzawa, H. Minda, and H. Masunaga, 2021: Refinement of surface precipitation estimates for the dual-frequency precipitation radar on the gpm core observatory using near-nadir measurements. *Journal of the Meteorological Society of Japan. Ser. II*, **99 (5)**, 1231–1252, https://doi.org/10.2151/jmsj.2021-060.

Ho, T. K., 1995: Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, Vol. 1, 278–282.

Iguchi, T., S. Seto, R. Meneghini, N. Yoshida, J. Awaka, M. Le, V. Chandrasekar, and T. Kubota, 2021: Gpm/dpr level-2 algorithm theoretical basis document. *NASA Goddard Space Flight Center*.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, **30**.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koistinen, J., 1991: Operational correction of radar rainfall errors due to the vertical reflectivity profile. *Preprints, 25th Int. Conf. on Radar Meteorology, Paris, France, Amer. Meteor. Soc*, Vol. 91, 94.

Kubota, T., T. Iguchi, M. Kojima, L. Liao, T. Masaki, H. Hanado, R. Meneghini, and R. Oki, 2016: A statistical method for reducing sidelobe clutter for the ku-band precipitation radar on board the gpm core observatory. *Journal of Atmospheric and Oceanic Technology*, **33 (7)**, 1413–1428.

Nair, V., and G. E. Hinton, 2010: Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814.

NWS, 2023: National weather service beam property calculator. URL https://training.weather.gov/wdtd/tools/beamwidth/.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, **12**, 2825–2830.

Siddique, N., S. Paheding, C. P. Elkin, and V. Devabhaktuni, 2021: U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, **9**, 82 031–82 057.

Skofronick-Jackson, G., and Coauthors, 2017: The global precipitation measurement (gpm) mission for science and society. *Bulletin of the American Meteorological Society*, **98 (8)**, 1679–1695.

Zhang, J., and Coauthors, 2016: Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97 (4)**, 621–638.

Zheng, A., and A. Casari, 2018: *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".