

A CloudSat-Informed Convective Cloud Identification Algorithm (CICCA) for Adding a Time Dimension to CloudSat Observations

Abstract

This paper introduces a method for augmenting the poor sampling characteristics of spaceborne active sensors by using them to guide the interpretation of imagery from passive sensors. The method is illustrated by training a new infrared convection identification algorithm using inferred locations of convective cores within high resolution CloudSat observations. A dataset featuring collocated space-borne radar observations from CloudSat and high-resolution cloud top temperature (CTT) output from PATMOS-x between 2007 and 2010 over the tropical North Atlantic is used for algorithm training. The resulting CloudSat-informed convective cloud identification algorithm (CICCA) synthesizes high-sensitivity CloudSat precipitation flags and PATMOS-x cloud top temperature (CTT) output through a quadratic discriminant analysis (QDA). When tested using independent observations, CICCA displays improved ability to approximate CloudSat convective cloud detection relative to simple thresholding approaches, adding a time dimension to CloudSat observations of maritime convection. For example, CICCA exhibits a Peirce Skill Score (PSS) 15.6% higher than 235 Kelvin (K) CTT threshold at identifying instances of deep convection. CICCA is used to examine the diurnal cycle of maritime convection through a CloudSat-like lens and assess the impact of CloudSat's twice-daily sampling estimates of convective area fraction. The analysis corroborates the dominance of diurnal harmonic in almost all regions and seasons with an overall average amplitude of 70% for both land-dominated and ocean-dominated pixels. However, land-dominated pixels show dramatically higher seasonal variability ranging from 113% in summer to 11% in winter. Subsampling the CICCA results to only the hours corresponding to 1:30 and 13:30 Cloudsat overpass, leads to a bias of $\sim 2\%$ in estimated convective fraction across the tropical Atlantic.

(1) Introduction

The improved understanding and parameterization of deep moist convection has an integral role in characterizing severe weather phenomena. The puzzling nature of convection frequency, with its varying radiative and hydrological implications, continues to bedevil cloud-resolving models (CRMs) and general circulation models (GCMs) and muddle our knowledge of future climate (Nesbitt and Zipser, 2003; Stevens and Bony, 2013; Bony et al., 2015). In particular, the seasonal, regional and diurnal variation of deep convection remains poorly understood. Increased understanding of deep convection and its variation is a vital step for understanding and representing our climate and future climate.

The hindered study of deep convection and its variability is perhaps most evident in maritime locations with the absence of ground-based radar. The limited observations in these regions pose a challenge for quantifying seasonal and diurnal variation of convection and the mechanisms behind it. One attempt to fill this void is use of remote sensing techniques. Passive radiometers reporting the brightness temperature, or longwave irradiance, of cloud tops have been used in the past to replace traditional ground-based radar (Meisner and Arkin, Arkin 1979, 1987; Maddox 1980; Augustine 1984; Albright et al. 1985; Hartmann and Recker 1986; Fu et al. 1990; Hendon and Woodberry 1993; Janowiak et al. 1994; Cairns 1995; Chang et al. 1995; Anagnostou et al. 1999, Imaoka and Spencer 2000; Yang and Slingo 2001). For identifying deep convection, authors choose varying brightness temperature thresholds to “see” all pixels with cloud top brightness temperatures lower than threshold temperature as deep convection or convectively related cloud. Specifically, 215 K in Fu et al. (1990), 219 K in Maddox (1980), 230 K in Yang and Slingo (2001), 235 K in both Meisner and Arkin, Arkin (1987,1979) and 253 K in Anagnostou et al. (1999). Some literature further uses the brightness temperatures to estimate rainfall, according to brightness temperature-rain rate relationships (e.g. Chang et al. 1995; Anagnostou et al. 1999; Imaoka and Spencer 2000; Yang and Slingo 2001). With ample temporal resolution typical of passive radiometers, cold cloud techniques can approximate diurnal variation of convection, rainfall or emissivity.

We look to expand the observation of deep convection, especially maritime, via a collaboration of aforementioned cloud top brightness temperature and active space-borne radar, namely CloudSat. This approach parallels the collocation of two similar datasets provided by Tropical Rainfall Measuring Mission (TRMM), which launched in late 1997. TRMM’s satellite provides both an elliptically scanning microwave radiometer and a three-dimensional precipitation radar (Kummerow et al. 1998, 2000). This paper is novel in relation to past literature in its use of CloudSat’s space-borne radar instead of TRMM’s precipitation radar to quantify deep convection (e.g., Nesbitt and Zipser 2003; Kikuchi and Wang 2008; Yang and Smith 2008). CloudSat provides high sensitivity discrimination between convective cores and stratiform cloud regions and thus is a well-suited tool for the current

application (Fig. 1; Im et al., 2005). CloudSat's performance has been demonstrated by various authors (e.g. Haynes et al., 2007,2009; Stephens et al., 2008,2017). The use of CloudSat as an alternative to TRMM may accrue promising results because of the increased sensitivity of CloudSat.

Sampling from CloudSat is inherently limited from its narrow relectivity curtain and long repeat cycle; thus, to fully deploy CloudSat-derived information it will be used in conjunction with an overlapping pixel provided from geostationary satellite brightness temperature hourly data to train an algorithm to input only geostationary data and output convective cloud fraction, known as CloudSat-informed convective cloud identification algorithm (CICCA). The algorithm is meant to provide mock CloudSat output; thus, allowing continuous hourly observation of maritime regions with sophisticated interpolation and expansion of CloudSat's utility. Expanding CloudSat's utility with this algorithm allows for continuous diurnal analysis that doesn't project longer scale variability (e.g. weekly or monthly variations as a result of disjointed CloudSat or TRMM analysis at a given location). CICCA also allows a unique opportunity to further quantify the ability of CloudSat to provide climatologies with bi-daily scanning (e.g. Liu and Zipser, 2008). Results show CloudSat and CICCA comparisons as well as CloudSat-derived diurnal analysis.

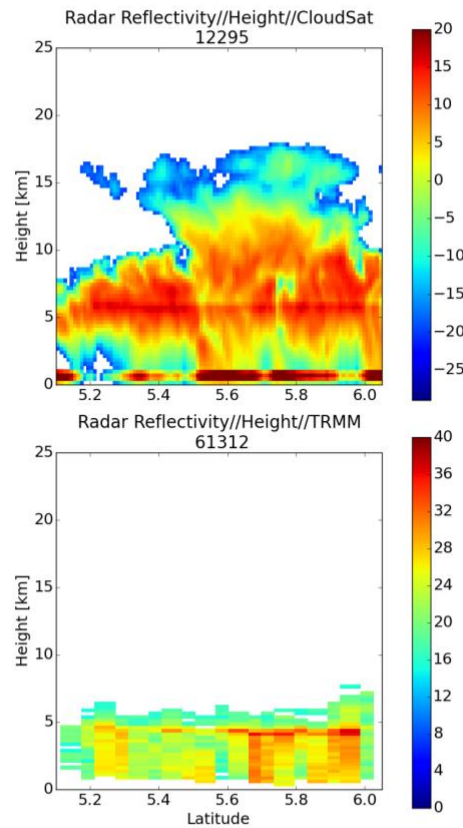


Figure 1: Comparison of CloudSat's CPR reflectivity (dBZ) and TRMM's PR reflectivity (dBZ) for crossover match in 2008.

2 Instruments

2.1 CloudSat

Cloudsat is part of the A-train satellite constellation and was launched in 2006 (L'Ecuyer and Jiang, 2010). CloudSat, along with the other A-train satellites, flies in a sun synchronous, polar orbit at an altitude of 705 km and an inclination of 98° with equatorial crossing times at approximately 1:30 and 13:30 local time (LT) and a 16 day repeat cycle, about 1.5-3 times as fast as TRMM's PR repeat cycle (Stephens et al., 2002). CloudSat carries a nadir-viewing, high power, W-band pulsed radar, Cloud Profiling Radar (CPR). CPR has a vertical, cross-track and along-track resolution of .5 km, 1.4 km and 1.7 km respectively. The instrument operates at 94 gigahertz (GHz) with a sensitivity of -30 dBZ more than adequate for distinguishing the stratiform and convective regions of storms and distinguishing raining and non-raining clouds (Im et al., 2005; Lebsock et al., 2011; Haynes et al., 2009; Deng et al., 2010; Igel et al., 2014; Sauter et al., 2019). Various comparisons of CloudSat to passive microwave and Ku-band radar observations showcase the enhanced sensitivity of CloudSat (Berg et al., 2010). Additionally, CloudSat is used as a calibration tool for notable programs such as NASA-JAXA GPM (Stephens et al., 2018).

CloudSat's CPR is used to measure reflectivity as a means of informing whether deep convection is occurring. Specifically, CloudSat's 2C-Column-Integrated Precipitation Retrieval Product (2C-PRECIP-COLUMN) expresses one of four precipitation flags: deep convection, stratiform, shallow precipitation or no precipitation. Liquid precipitation acutely attenuates the W-band reflectivity profile; thus, 2C-PRECIP-COLUMN provides a very accurate classification of precipitation mode (Haynes et al., 2009). The product algorithm relies on a distinctive signature of increasing reflectivity profiles with height from surface to above the freezing level, indicative of raindrops lofted by convective updraft to identify deep convection. Stratiform precipitation is distinguished from deep convection when the reflectivity curtain is comprised primarily of frozen hydrometeors above the freezing level and a clear transition to liquid hydrometeors 0.5 km below the freezing level. Shallow convection is distinguished by increasing reflectivity with height occurring below the freezing level (Bouniol et al., 2016; Haynes et al., 2009, 2014).

2.2 PATMOS-x

Geostationary brightness temperature used in conjunction with CloudSat is provided by PATMOS-x. PATMOS-x cloud products from the Geostationary Operational Environmental Satellite East (GOES-East) and Meteosat Second Generation (MSG) datasets are used in tandem. PATMOS-x is a satellite-based climate data record generated by the National Oceanic and Atmospheric

Administration (NOAA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) (Heidinger et al., 2012; Schmetz et al., 2002). Combining these two satellites, PATMOS-X provides a view of the Atlantic Ocean in our region of interest, 5N – 15N and 10W - 80W. GOES East spans laterally from 40W- 80W, while MSG covers the remaining extent of the interest region, 10W-40W.

PATMOS-x dataset supplies cloud information from calibrated Advanced Very High Resolution Radiometer (AVHRR) observations and selected ancillary data that have been made from various NOAA satellite instruments since 1978. AVHRR visible and infrared radiances provide information for detecting clouds on a global scale at a spatial resolution of 1.1 km at nadir over the GOES-East swath (Heidinger et al., 2014; Sun et al., 2015). MSG satellites utilize Spinning Enhanced Infrared Imager (SEVIRI) for continuous observation of the Earth. SEVIRI consists of 12 visible and infrared channels to detect cloud properties. PATMOS-x then converts brightness temperatures from satellite products into cloud top temperature (CTT) in degrees Kelvin. The resulting spatial resolution is $0.1^\circ \times 0.1^\circ$ and temporal resolution is 1 hour (Heidinger et al., 2012). The application of this paper will refer only to PATMOS-x for CTT (K) and cloud top height (CTH; km).

3 The CloudSat-informed Convective Cloud Identification Algorithm (CICCA)

The synthesis of these two datasets follows the methodology suggested in Sauter et al. (2017, 2019). Explicitly, along a CloudSat granule every grid point is collocated with the nearest neighbor PATMOS-x grid point. The nearest neighbor PATMOS-x pixel is considered the center pixel. The collocated data is categorized according to the 2C-PRECIP-COLUMN convective flag. If CloudSat grid point reports a deep convection flag, the nearest PATMOS-X grid point is categorized as a deep convection grid point. In addition to the PATMOS-x center pixel CTT (CTT_{CP}), average CTT is computed for a range of progressively larger areal footprints surrounding the pixel in the PATMOS-x grid, ranging from 36-58 km in diameter (GOES-East: 3-5 pixels; MSG: 6-11 pixels). The smallest areal footprint average CTT is deemed ring 1 average CTT (CTT_1) and the numerical assignments increase as size increases.

To train PATMOS-x CTT data with CloudSat data, more care is needed to further synchronize these two datasets. The collocated pixels are grouped such that data is only recorded if a pixel group is homogenous (e.g. all pixel flags in the group reflect deep convection) and are capped on either side by a no precipitation flag. This provides a more reflective description of a given pixel group between a heterogeneous pixel group or a single isolated deep convection pixel. This filter is imperative as the footprint technique records characteristics of adjacent pixels as well as the center pixel. An illustrative figure is included to help visualize this methodology, note the change in average temperature in heterogeneous reject group versus the homogeneous record group, 246 K to 225 K (Fig. 2). After this grouping, collocated pixel groups are filtered out if reported CloudSat and PATMOS-x cloud top heights (CTHs) are not within 1 km of each other. If this criterion is not

satisfied it is assumed that the satellites aren't recording the same scene. This CTH filter accounts for the fact that CloudSat and PATMOS-x data can be as much as 30 minutes detached. The unfiltered pixels are compiled according to the CloudSat flag: deep convection, stratiform, shallow and non-precipitating pixel.

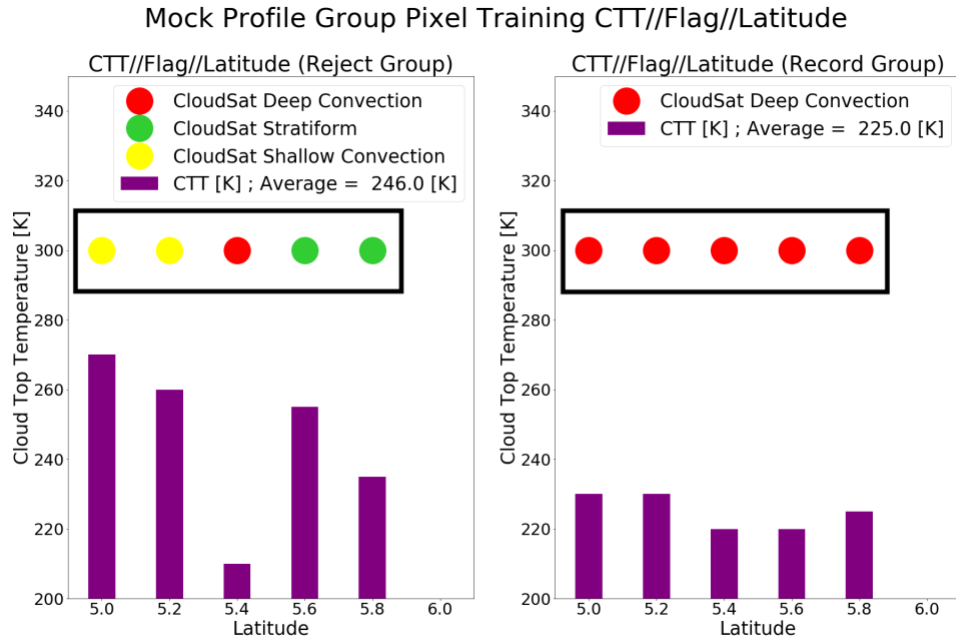


Figure 2: An illustrative description of grouping the pixels according to homogeneity with the dots representing CloudSat flag and the Cloud Top Temperature (CTT) shown below in purple. (Left) a representation of a rejected pixel group. (Right) a representation of a recorded pixel group.

The database of points is significantly reduced from filtering but what is left is a robust set of vectors (data couplets) representative of 4 years of collocated satellite data (2007-2010). This process is repeated across the tropical Atlantic (5° N - 15° N) for 10° x 10° regions denoted from East to West (right to left) as Region 1,2,n... (Fig. 2). Also listed is the number of vectors for each specific region (Fig. 3). Data derived from both CloudSat alone and the collocation dataset are shown in the results.















Tropical Atlantic Regional Boxes (10x10 degrees)						
						
Region 7 5N-15N 70W-80W	Region 6 5N-15N 60W-70W	Region 5 5N-15N 50W-60W	Region 4 5N-15N 40W-50W	Region 3 5N-15N 30W-40W	Region 2 5N-15N 20W-30W	Region 1 5N-15N 10W-20W
						

Figure 3: A geospatial representation of the paper's test region, the tropical North Atlantic. Listed are the regions' title, latitude, longitude and the number of representative vectors generated. (Have to regenerate with vector number.

We integrate this trained data in an algorithm that identifies convective cloud fraction, effectively interpolating CloudSat observation. Figure 3 is shown to help describe the algorithm. The algorithm requires an input vector and a reference vector, the reference vector being the trained PATMOS-x data. In practice, PATMOS-x data provides a latitude/longitude grid of CTT (Fig. 4b). Every grid point will contain two features CTT_{CP} and CTT_1 . Sole use of CTT_{CP} threshold is illustrated; specifically, a 235 K threshold used by Arkin (1979), Arkin and Meisner (1987) is shown in figure 4a. These features are translated to a vector with two components, CTT_1 and ratio of CTT_1 to CTT_{CP} . This choice provides a vector representative of both the magnitude of the CTT as well as the ratio of the presumed decay of CTT moving radially outward from the center reflective of deep moist convection (e.g. Yuan et al., 2011; Sauter et al., 2019). This data vector is the input vector. The input vector is then compared to the specific regional database informed by CloudSat, figure 4d shows an example collocation between PATMOS-x and CloudSat, comprising the regional database (Fig. 4d). The regional database generates identically formatted vectors to geostationary input. The algorithm then applies quadratic discriminant analysis (QDA). QDA is a multivariate pattern recognition tool, or an extension of Fischer's linear discriminant analysis (LDA). QDA involves assuming Gaussian distribution of multiple classes in vector space, with no assumptions on covariance, and Bayesian retrieval of said classes (Hastie et al., 2008). QDA generates a reference vector concerning all database vectors pertaining to a certain precipitation flag. The QDA method compares the input to the reference flags and then decides whether the grid point reflects deep convection or other flags, as well as a percentage of certainty. For purposes of this paper, the algorithm guess will be presented as deep convection (yes) and the remaining flags (no). A pixel is only considered a deep convective pixel if the percentage of certainty is greater than 50% (Fig. 4c). The simplicity of this methodology is not meant to predict where blips of the most intense convection or overshooting tops occur, but rather the whole convective cloud area. In fact, this cold cloud technique should cover a significantly larger area than just where CloudSat identifies deep convection (i.e. the ratio of the area of maxima in convective updraft vs. the entire resulting convective cloud top).

The nature of identifying deep convection demands caution when using supervised learning techniques. Most importantly, the dataset of usable deep convection points is very small, diminishing the utility of statistical or machine learning implementation. It follows; the most important thing is to increase the data pool of deep convection points. Fortuitously, data vectors for deep convection occupy very similar space throughout all regions, which is perhaps not surprising. In consideration of the dearth of deep convection data and transregional similarities, deep convection data is pooled for all regions to use. As an example, the reference vector for Region 3 utilizes deep convection vectors from Regions 1-7. Thus, the learner has 280 deep convection vectors to implement. Stratiform, shallow convection and non-precipitating pixels have ample data for each individual region and don't require data pooling, as these phenomena are much more common. In summary, QDA in every region refers to 280 ($24 + 16 + 37 + 39 + 50 + 50 = 280$; Regions 1-7) global data vectors for deep convection and 280 regionally unique vectors for the remaining CloudSat flags ($280+280+280+280 = 1120$ vectors for every region). An example regional vector space plot is shown in Results.

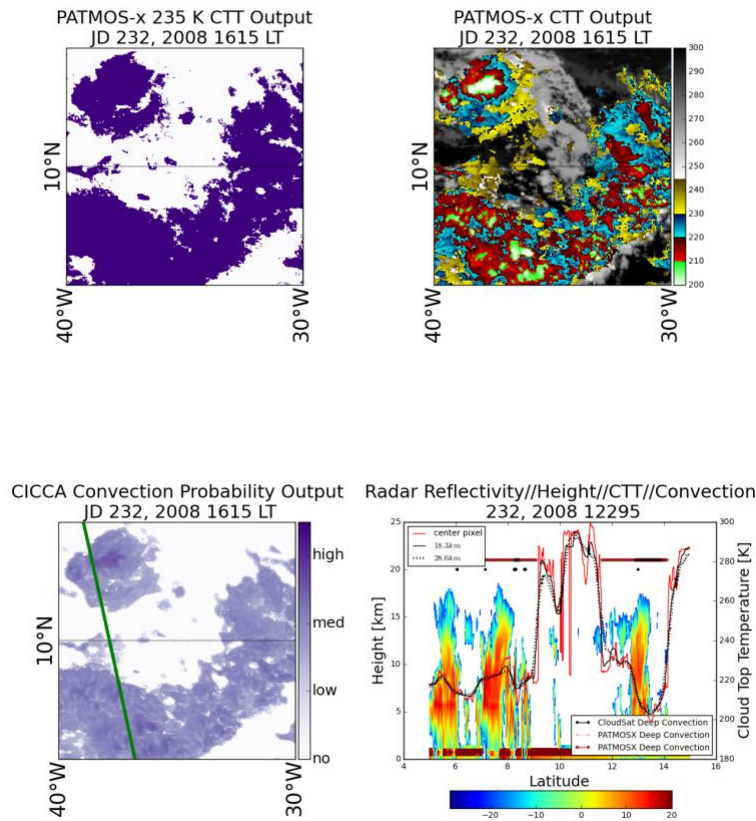


Figure 4: (Top left) a geospatial representation of CTT_{CP} that exceed 235 K. (Top right) a geospatial representation of CTT, denoted by color fill. (Bottom left) CICCA provided probabilistic output, denoted by purple shade. Green stripe passing through this picture represents collocated CloudSat path. (Bottom right) curtain representation of CloudSat reflectivity with CTT also projected on image.

3.1 Skill Score/Cross Validation of Algorithm

To validate the agreement of both this paper's algorithm and CTT temperature thresholds with CloudSat data on deep convection occurrence, a skill score methodology satisfactory to the unique criterion of this validation is necessary. The algorithm refers to a database of points compiled in the time period 2007-2010. To separate the training data and validation data, days in which reference data on deep convection were recorded are omitted from the validation set, a simple form of cross validation.

The methodology is based on a contingency table, typically used to compare forecasts to actual occurrence. In this analysis, the proposed algorithm seeing or not seeing deep convection is the forecast and the collocated CloudSat data seeing or not seeing deep convection is considered the actual occurrence. For each validation step, there are four different outcomes. For this specific application, three considerations are suggested

- 1) Deep convection events are comparatively rare in this data set; thus, excess weight should be put on validation steps where CloudSat sees deep convection (Yes/Yes and No/Yes).
- 2) In relation to most severe weather events, false alarms are considered less harmful than misses; thus, a miss will be weighted more heavily than false alarms for invalidating algorithm skill.
- 3) Similar to (1) Yes/Yes validation is more important than No/No validation; thus, a Yes/Yes occurrence will be weighted more heavily for validating algorithm skill.

To satisfy the unique criteria of this application the Peirce Skill Score (PSS) is used (Eq. 1-2). PSS, originally proposed by Peirce in 1884, can be understood as the hit rate subtracted by the false alarm rate (Wilks, 2011). PSS is constrained to be unbiased, adjusting importance of any contingent outcome to the relative frequency of that outcome. PSS thus provides a skill score that adapts to the rarity of the proposed event, while still considering both misses and false alarms. Additionally, misses are penalized more heavily because of their relative rarity to false alarms. Other literature assessing skill of detection algorithms (real-time and nowcasting) use probability of detection (POD), false alarm rate (FAR), critical success index (CSI) and bias (e.g. Cai et al., 2008; Donovan et al., 2008; Kessinger, 2017). The nature of this paper's validation data sets makes those options not as attractive (or applicable) because of the rarity of deep convection.

$$PSS = \frac{(a + d)/n - [(a + b)(a + c) + (b + d)(c + d)]/n^2}{1 - [(a + c)^2 + (b + d)^2]/n^2}$$

$$= \frac{ad - bc}{(a + c)(b + d)}$$

Regions	Ring 1, 50% PSS (CICCA)	Ring 2, 50% PSS	235 PSS	219 PSS	215 PSS	Cloudy pixels n=
1	.744	.787	.753	.520	.321	60945
2	.767	.787	.643	.303	.137	168709
3	.901	.893	.687	.224	.202	205478
4	.729	.790	.742	.183	.144	153983
5	.858	.907	.653	.281	.140	147540
6	.855	.866	.740	.273	.180	96171
7	.702	.792	.589	.140	.123	91538
Mean	.794	.842	.687	.275	.178	All Pixels

Table 1: Resulting PSS for every algorithm and the regional result as well as the sample population number.

The skill of the algorithm is assessed between three temperature thresholds that are used in literature (Table 1). Additionally, the skill was tested with an algorithm considering an additional dimension, CTT₁/CTT₂. This paper's 1-ring algorithm did a better job identifying where CloudSat saw deep convection occur, and thus a better job at identifying convective cloud than the best performing temperature threshold (235 K), 15.6% better. As a side remark, the 2-ring algorithm shows a better skill score but the probabilistic output is very suspect as compared to 1-ring output. Illustratively, the edges of the cloud shield show a very high probability of reflecting deep convection, which is an errant result and suggests invalidity of the 2-ring algorithm (Fig. 5). Presumably this will be the outcome in all cases that aren't completely isolated convection. It should be further noted that more sophisticated methods exist for detection, such as subtracting various IR channels and getting a brightness temperature difference (BTD) or including other indicative measurements like total flash rate (e.g. Schemtz et al., 1997; Donovan et al., 2008; Bedka et al., 2010; Kessinger, 2017). These methods are more readily used in nowcasting and forecasting detection.

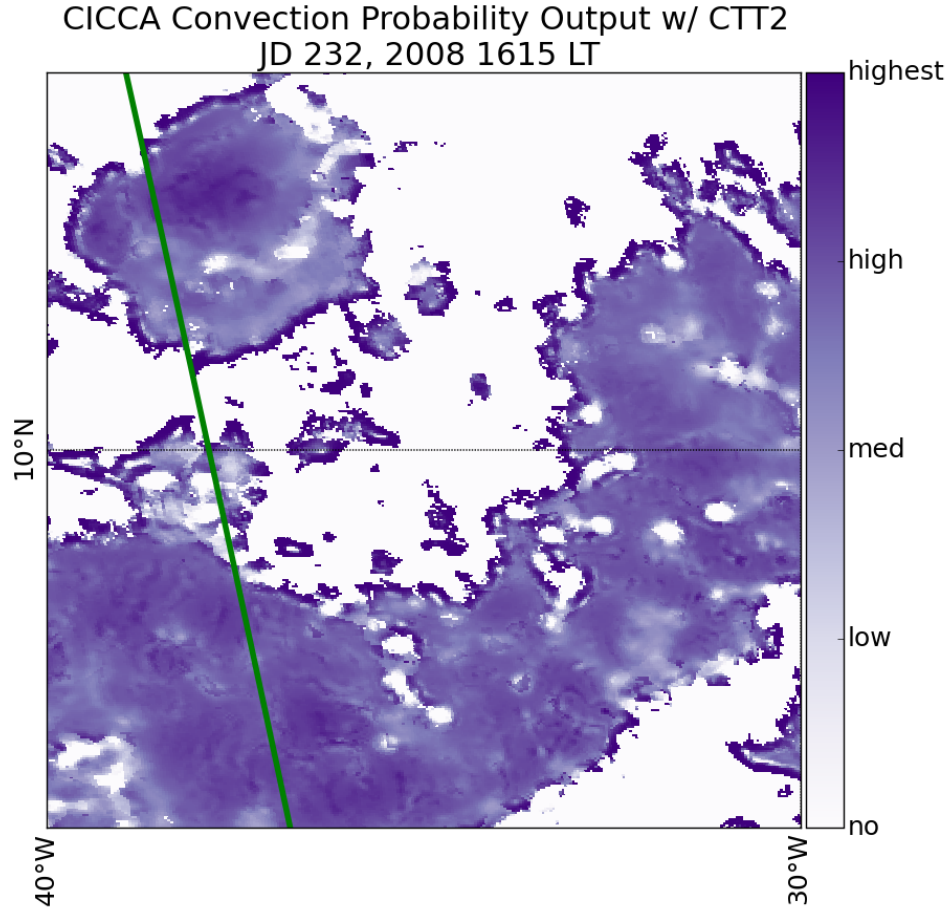


Figure 5: A geospatial representation of CICCA when including CTT₂ data. Purple shade represents probabilistic outcome from QDA, approximating increasing confidence that deep convection is occurring.

3.2 Algorithm Reference Vector

The frame of this algorithm can be shown more explicitly by scrutinizing the two input dimensions: the CTT₁ and CTT₁ / CTT_{CP}. As previously stated, the algorithm considers all the deep convection points spanning all regions; thus, the deep convection points on all 7 regions will occupy the same space. It is useful to represent the vector space, as QDA and its eventual decision-making are graphical in nature. The quadratic in QDA refers to the fact that the decision-making plane doesn't have to be linear. The deep convection vectors typically occupy a cluster indicative of lower CTT temperatures and a higher ratio of CTT₁ / CTT_{CP}, that increases as CTT₁ increases (Fig. 6b). It's apparent that the vast range and bimodal nature of CTT_{CP} makes thresholding deep convection by temperature difficult (Fig. 6a). In all regions, stratiform pixels share plenty of vector space with deep convection pixels; however, stratiform pixels show a

much larger distribution across this two-dimensional space. Finally, in all regions shallow pixels occupy a very unique space indicative of warmer CTT_{CP} (Fig. 6b).

The synthesis of both CTT_{CP} and CTT_1 allows for simple understanding of precipitation pixels and the applicability of categorization. Despite the overlap, it is clear precipitation types occupy unique parts of the vector space. The clustering of precipitation types encourages the use of statistical and machine learning. This clustering is likely why CICCA is able to identify deep convection more consistently than a temperature threshold. Conversely, the overlap of various precipitation types always results in some errant identification. For all of the regions, the ability for CICCA to discriminate the proper precipitation type, considering the reference vectors, is only about 80%. Typically, the issue in identification is between stratiform pixels and deep convective pixels. In consideration of the current skill and faults of CICCA, the algorithm could be further improved with increased dimensionality to avoid overlap.

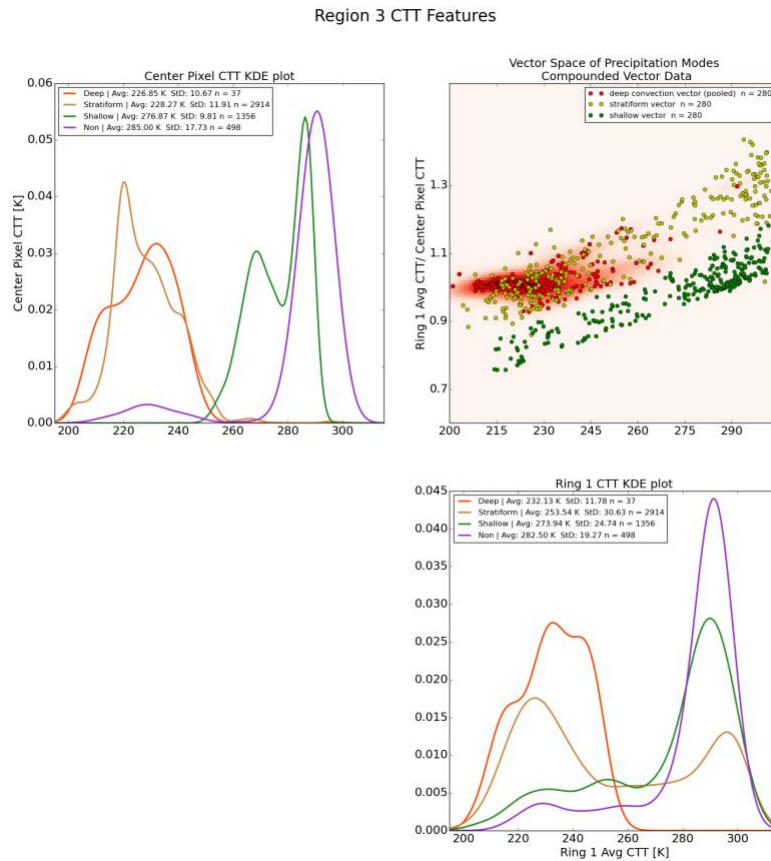


Figure 6: (Top Left) a kernel density function showing frequency of CTT_{CP} by precipitation flag types. red: deep convection, yellow: stratiform, green: shallow convection, purple: non-precipitating. (Bottom Right) as in top left but with CTT_1 . (Top right) The vector space of the data couplets synthesized from CTT_{CP} and CTT_1 density functions with same color representation and shaded red emphasizing deep convection clustering.

4 Results

4.1 CloudSat compared to CICCA

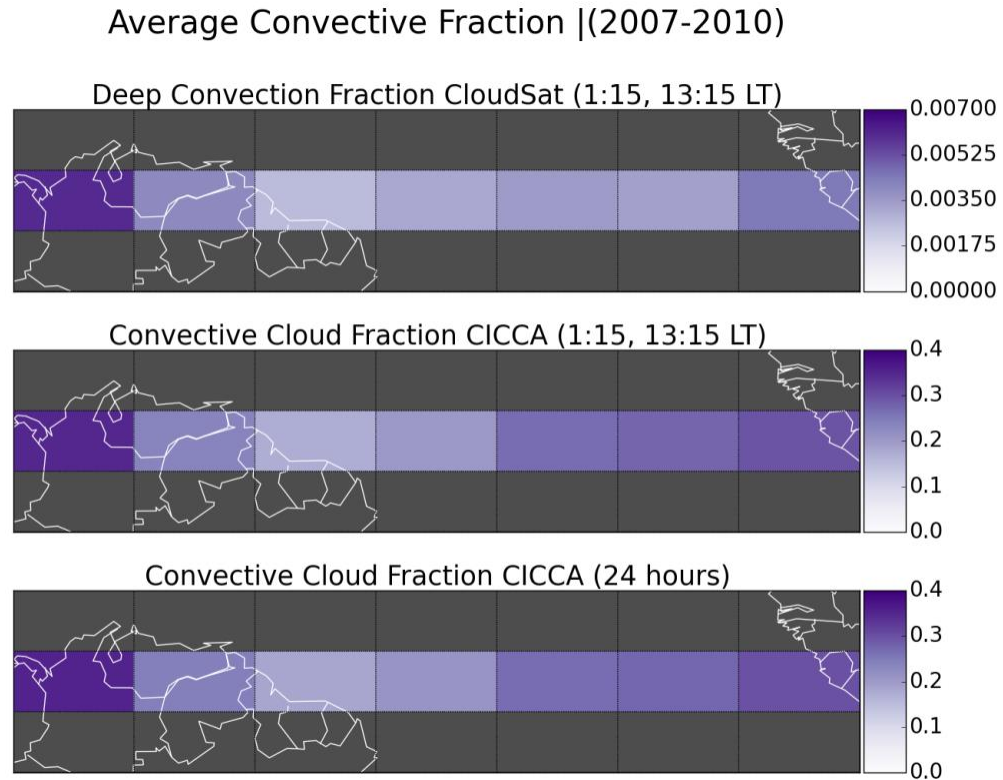


Figure 7: (Top) average deep convection fraction for entire study period by region from CloudSat. (Middle) average convective cloud fraction by region from CICCA for only CloudSat times. (Bottom) as in middle but for all 24 hours of the day.

CloudSat deep convection fraction is presented across all regions (Fig. 7a). These statistics only reflect CloudSat overpass times; CloudSat scans roughly around 1:30 and 13:30 local time (LT) every day. That said, this data suggests the general magnitude of convection but not, for example, the seasonal or diurnal cycle. Not surprisingly, the regions with the highest percentage of deep convection are the mostly land-mixed regions (regions 1,6,7). The oceanic pixels have a lower incidence of deep convection hovering around 0.3% of all CloudSat pixels (Regions 2-5, Fig. 7 b-e).

Results from CICCA for 2007-2010 show the ability of CICCA to mock CloudSat at 1:15 and 13:15 LT (e.g. average scan time ~ 15 minutes different) and interpolate convective cloud fraction for the remaining times not directly observed by CloudSat. Neglecting other factors, there should be a correlation to CloudSat deep convection fraction and CICCA convective cloud fraction as they vary in time and space (as

previously stated, these two values will be off by an order of magnitude). The regional variance generally matches up between CloudSat deep convection fraction and algorithm-identified convective cloud fraction at CloudSat times, the linear coefficient of association (R^2) is .77 (Fig. 6 a-b). Mixed land-ocean regions show the highest and second highest convective fractions (Regions 1 and 7). Conversely, Region 5 shows the lowest CloudSat deep convection and convective cloud fraction. In consideration of CloudSat bi-daily scan times, CICCAs convective cloud fraction averaged over the whole day and just during CloudSat times are very similar. Specifically, over the seven regions CICCAs results at Cloudsat times convective cloud fraction is only 2.1% lower than CICCAs results at all hours (Fig. 7 b-c). There are however some discrepancies between CloudSat and CICCAs, for example: Region 6 shows higher CloudSat deep convection fraction than Regions 2 and 3 but shows a smaller CICCAs convective cloud fraction (Fig. 7 a-b).

4.2 Interpolating CloudSat for Hourly observation w/ CICCAs

CICCAs time series interpolates CloudSat observations to hourly resolution and discriminates by season, providing results inaccessible to CloudSat alone. Results include: hourly-binned convection fraction averages, hourly-binned convection intensity averages, and diurnal harmonic analysis (diurnal and inner-diurnal). Examples are shown for one region (Region 3) in this paper, remaining regions are shown in supplementary data.

CICCAs results show an average daily fluctuation of convective cloud fraction separated by seasons. As far as seasonality, JJA and SON show higher convective cloud fraction than DJF and MAM (Fig. 8). The average amplitude (peak to trough as percentage of maximum) throughout regions ranges, quite dramatically, from 20% - 50%. As far as phase, the minimum in phase between regions converges between 10:15 and 15:15 local time (LT). The maximum in phase varies considerably more between region and season, but generally occurs between late afternoon and early morning.

The probabilistic nature of this algorithm allows for a novel analysis of the average percentage of certainty that a given pixel reflects deep convection. Nesbitt and Zipser (2003) quantifies the daily time series of convective intensity with two variables: minimum 85-GHz percentage and maximum height of 30-dBZ echo. Contrastingly, this paper will use the average convective probability of all convective pixels (over 50%) as a proxy for convective intensity. To confirm this proxy, post-analysis is done to compare high probability and low probability CICCAs pixels (80-90%, 50-60%). To quantify the difference, the maximum attenuation value over 5 km is recorded from 2007-2010; collocations are from the previously used validation dataset. Considering attenuation at heights greater than 5 km is an attempt to reduce noise from attenuation measurements that are not related to deep convection (mostly to reduce noise from ground attenuation and shallow

convection attenuation). Hourly averages of convective intensity don't vary nearly as much as the variance of convective cloud fraction presented previously (Fig. 8). The average convective intensity ranges from about 70% to 80% across all regions and seasons. The daily fluctuation of convective intensity appears to tell a heterogeneous story between individual regions, especially in comparison with overall convective cloud fraction. Further, post-analysis shows higher average attenuation values for high probability CICC pixels (Table 2).

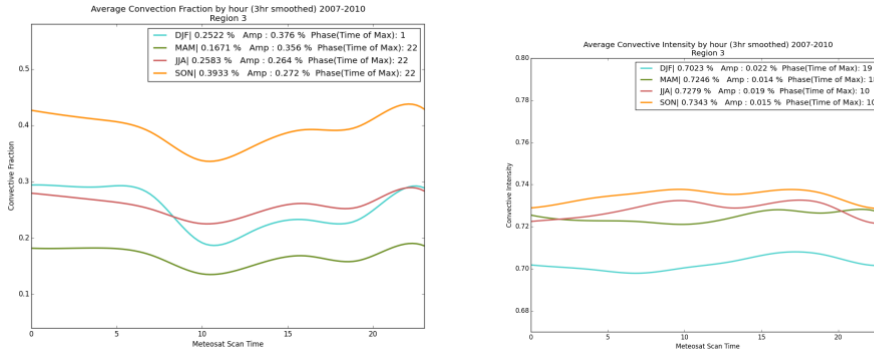


Figure 8: (Left) a 3-h smoothed curve of hourly convective cloud fraction from CICC over entire study period. Corresponding colors are as follows: blue: DJF, green: MAM, red: JJA, orange: SON. (Right) as in left but the average convective intensity.

Region	80-90% Max att. Height > 5km (dBz)	50-60% Max att. Height > 5km (dBz)	t-test result	df
1	10.77	7.18	p-value < .001	1551
2	10.44	7.83	p-value < .001	2089
3	11.56	7.82	p-value < .001	1283
4	9.12	8.76	p-value = .013	1436
5	9.15	8.62	p-value = .001	1127
6	9.70	8.04	p-value < .001	1294
7	9.90	8.96	p-value < .001	2402

Table 2: maximum attenuation values above 5 km are recorded by CICC probabilities. One tailed t test are performed for every probability group by region. Resulting p values and degrees of freedom are presented.

4.3 Diurnal Characteristics

The contributions to the overall convective cloud time-series variance are quantified for 24-h variance as well as the 12-h and 8-h harmonics. An example of both the appended season-specific time series and a power series are shown in figure 9. The power series shows a continuous spectral plot and a dotted line denoting 95% confidence of a statistically significant harmonic peak (Fig. 9b). Not surprisingly, every region and season shows a strong 24-h signal (Table 3). Further analysis explores the relationship between land/ocean pixels as well as inter-seasonal differences (for analysis only predominant regions are chosen e.g. land regions = 6,7 and ocean regions = 2,3,4). One-tailed t-test shows land regions show a higher contribution from the diurnal frequency to the overall variation (land = 80%, ocean = 71%, p-value = .06). A one-tailed t-test between summer (JJA) regions and winter (DJF) regions shows a much stronger contribution to overall variance from 24-h signal in the northern hemisphere summer (land = 85%, ocean = 73%, p-value = .01).

Fast Fourier Transform (FFT) is performed on convective cloud fraction time series, particularly in respect to the 24-h wave (diurnal variation), as informed by the comparatively large contribution of diurnal frequency to the overall variation of the convective cloud fraction. Both 24-h amplitude (of convective cloud fraction) and phase are presented; results are discriminated by region and season (Fig. 10). Important to observe is the dampening amplitude over land from northern hemisphere summer to winter; the 24-h wave over land transitions from 113% amplitude to 11% (Fig. 10). The land variance accounts for most of the seasonality. Despite the extreme variance over land, the net amplitudes for the 24-h wave averaged across all 4 seasons tend to be the same for land and ocean, around 70%. As for the 24-h frequency phase, the overall delay is about 2 hours between land and ocean. However, like amplitude there is a high degree of variance between seasons.

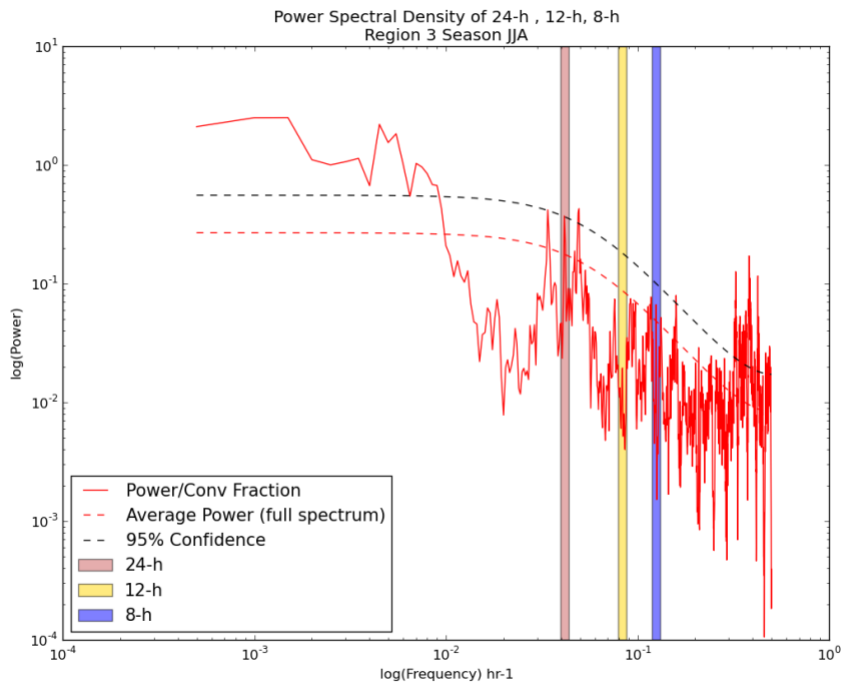
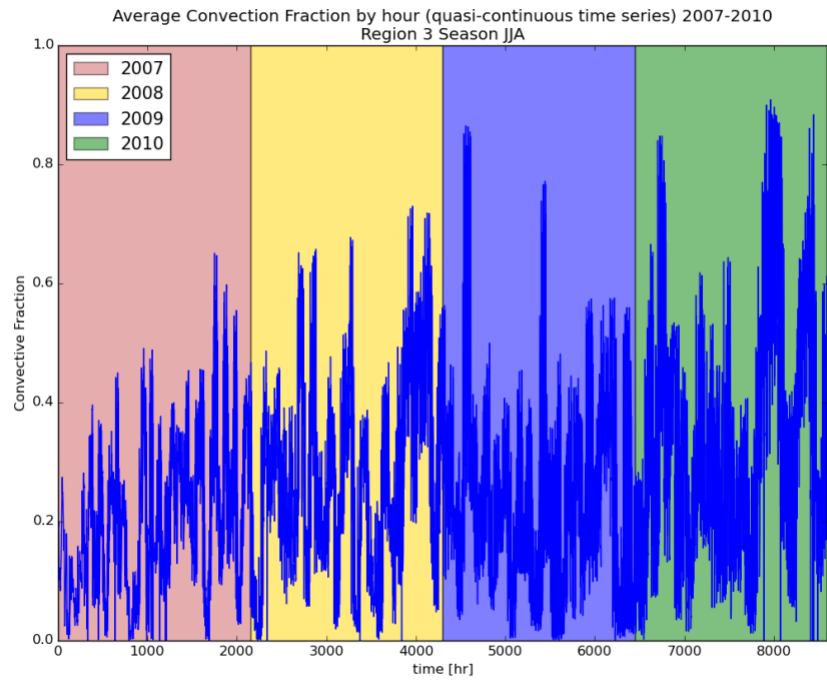


Figure 9: (Top; a) a quasi-continuous function of CICCA convective cloud fraction by hour over the entire time period for JJA. Different shading pattern divides the different years. (Bottom; b) a discrete representation of the power spectral density of the JJA time series highlighting certain frequencies: 24-h, 12-h and 8-h. A plot of average power and confidence interval are plotted in tandem with dotted lines.

Regions	24 h /DJF Cont	24 h /MAM Cont	24 h /JJA Cont	24 h /SON Cont	(8-h + 12-h) /24-h /DJF Cont	(8-h + 12-h) /24-h /MAM Cont	(8-h + 12-h) /24-h /JJA Cont	(8-h + 12-h) /24-h /SON Cont
1	71.1	84.0	86.3	89.0	15.3	6.00	6.15	6.50
2	83.6	56.5	78.5	67.4	12.6	32.7	11.1	21.9
3	75.4	62.6	74.1	77.4	18.8	35.7	21.8	15.0
4	64.5	95.8	77.4	85.2	31.0	14.4	17.7	10.5
5	66.3	43.5	90.4	85.9	36.6	*84.5	6.98	14.4
6	81.6	76.6	96.0	88.8	14.4	20.0	3.49	11.0
7	71.5	58.2	91.7	76.4	28.7	36.2	4.80	25.1

CICCA Convection Fraction 24 hr Amplitude and Phase (2007-2010)

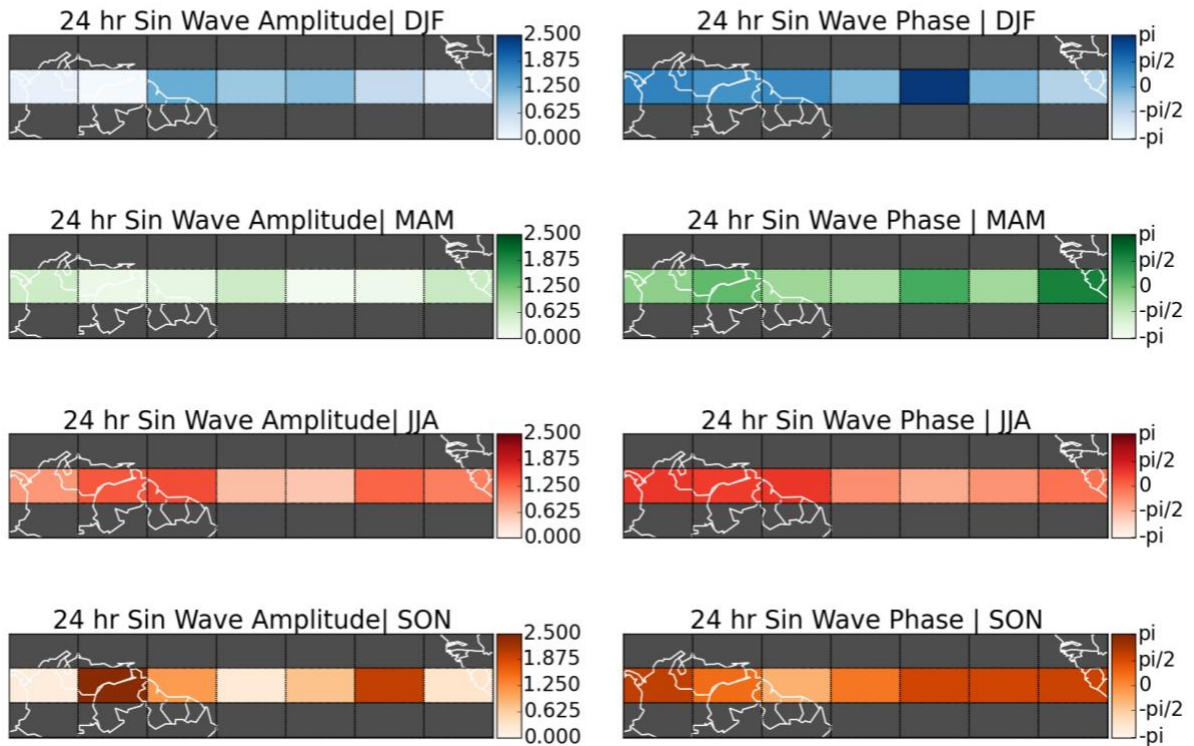


Figure 10: (Left column) the 24-h sin wave amplitude is plotted by region and shading patterns correspond to different seasons. Corresponding colors are as follows: blue: DJF, green: MAM, red: JJA, orange: SON. (Right column) 24-h sin wave phase is plotted. Colors as in left column.

Table 3: The contribution to the overall variation from the 24-h sin wave is listed for every region and season. Further, the fraction of the inner diurnal variations to the 24-h wave is reported.

5 Discussion

Results suggest three separate metrics demonstrating the ability of CICCAs to emulate CloudSat: comparable variance of convective fraction, CICCAs probabilistic outcome discerning between attenuation values and enhanced skill score from thresholding techniques. First, there was general agreement between CloudSat and CICCAs at CloudSat times. Additionally, the difference between overall magnitude of CICCAs for all hours of the day and CICCAs at CloudSat times suggests an average bias of around 2%. This suggests CloudSat scan times do a good job at capturing the overall (all seasons) convective cloud fraction. This implies CloudSat-derived information is well suited for climatic analysis at or higher than yearly timescale. Further, CICCAs probabilistic outcome demonstrates a statistically significant ability to increase confidence of convective cloud occurrence as CloudSat attenuation values increase. This suggests that CICCAs can coarsely interpolate CloudSat attenuation too. Despite these findings, there should be some mention as to the discrepancy between overall convective fraction hierarchy in CloudSat's deep convection fraction and CICCAs's convective cloud fraction. Holding other variables constant, the convective cloud hierarchy should mirror CloudSat. One such explanation could be an affinity for more cold cloud outflow in the regions closer to the African coast. This result is not an artifact of CICCAs alone as analysis of 235 K cloud fraction shows a similar result. A thorough quantification or explanation of that result is beyond the scope of this paper. Last, despite the difference in magnitude, the skill score results suggest an improved ability to identify deep convection from brightness temperature alone. Though the discrepancies in these two values (CloudSat deep convection and CICCAs convective cloud fraction) still exist, the skill score results suggest improved accuracy in representing deep convection at higher spatial and temporal resolution.

New insights are provided from emulated CloudSat time series. The hourly-binned averages of convective cloud fraction from CICCAs show the average hourly convection fraction summing all contributions (e.g. insolation and transient convective systems). In agreement with other literature the LEEM (late evening early morning) maxima was predominant in all regions (e.g. Chen and Houze 1997, Yang and Slingo 2001, Nesbitt and Zipser 2003). It is important to note the maximum in cold cloud coverage may follow the precipitation peak by 2-3 hours, the slight tardiness of maximum in the predominantly land regions could be explained by this (e.g. Slingo et al., 2004). It is important to emphasize that the amplitudes of these binned averages do not represent the diurnal amplitude average, rather the sum of all variations. In that vein, these patterns show general magnitude and the huge seasonal disparity in the tropical North Atlantic, omitted from most previous analyses of this type as well as sole CloudSat output.

The hourly-binned averages of convection intensity from CICCAs show the average hourly convection intensity, summing contributions from all wave frequencies. Previous literature unanimously agrees variation in convective intensity has little effect on the overall diurnal cycle; rather the convective

frequency is more telling (e.g. Nesbitt and Zipser 2003, Dai et al. 2007). A simple multiplication of the hourly-binned convective cloud fraction and convective intensity confirms this, as the pattern mirrors convective cloud fraction pattern. Nonetheless, the pattern of intensity has importance in other applications and the ability to partition by intensity (or probability) is a skill of CICCAs. A possible explanation: different times show increased probability of reflecting larger scale organization than just isolated convection. Larger convective systems would show a higher maximum attenuation value and such discrimination between small and large cycles have been explored in previous literature (e.g. Mapes and Houze Jr. 1993, Nesbitt and Zipser 2003). MCS generation can have favorable times and might be reflected in convective intensity binned averages or reflect a preferential time for a transitioning system to migrate into a particular region (Houze Jr., 2004).

Harmonic analysis of CICCAs shows interpolated CloudSat observations that are reasonably consistent with previous literature. One theory explaining diurnal cycle is the thermal destabilization of the boundary layer. This has been touted as the explanation for land-based diurnal variability but also a contributor to ocean-based variability (Yang and Smith, 2006). In analyzing time-series of a particular region and season, it can be reasonably assumed that insolation, the cause of thermal destabilization, is a 24-h phenomenon. The results from this paper's correlation analysis suggest a statistically significant stronger contribution from this 24-h wave over land and in northern hemisphere summer. This is to be expected because of more responsive terrestrial thermal exchange. Oceanic thermal exchange involves a heating element with much higher thermal capacity and upwelling that can hamper the thermal destabilization of the boundary layer (Yang and Smith, 2006). It does, however, seem that insolation is a very high contributor to overall variation even in oceanic pixels, a result that may have been previously overlooked, but this could be an artifact of this dataset encompassing a region with much stronger insolation as compared to the rest of the planet. Additionally, seasons and regions with smaller contributions from the 24-hr signal may have different mechanisms contributing. Yang and Slingo (2001) reports the contribution from semi-diurnal and 8-hr to typically be 40% and 15% of the diurnal harmonic (Yang and Slingo 2001). This paper finds the average ratio of inner-diurnal to diurnal to be around 20%, on the whole much less but some regions do exhibit a substantial inner diurnal contribution. For example, Region 5 may be an interesting area to analyze further to isolate inner-diurnal contributions to convective variability as it is well above the typical inner-diurnal contribution. Also, the FFT-informed 24-h periodic amplitude and phase show consistency with previous literature. Two caveats to comparing this 24-h convective cloud amplitude are: the limited regional scope of this analysis and the direct comparison of this paper's convective cloud fraction to rain rate. In time period (NDJFM), Sakaeda et al. (2017) find a diurnal amplitude, derived from TRMM 3b42 rain rate, of about 50% in the Eastern Atlantic and about 20% over land which is consistent with this paper's convective cloud fraction in the same season (Sakaeda et al. 2017). Nesbitt and Zipser (2003) finds TRMM 3b42 rain rate amplitude of 30% for oceanic pixels and 125% for land pixels (Nesbitt and Zipser, 2003). The discrepancy likely has to do with broad regional reporting and lack of explicit

seasonality within their analysis. Dai et al. analyze the diurnal amplitude of multiple rainfall products (TRMM 3B42, CMORPH, PERSIANN and MI) and find amplitudes ranging from 70-140% over this paper's land regions and 20-60% in oceanic regions in JJA (amplitude percentages harmonized because reported amplitude as mean to peak e.g doubled; Dai et al. 2007). Dai et al. also mention a 2-hour phase change between central Atlantic regions and coastal Atlantic regions (Dai et al. 2007). General agreement with past literature regarding 24-h variation suggests validity of using CICCAs to emulate CloudSat at hourly resolution.

6 Conclusions

This paper presents the use of CICCAs for generating hourly resolution of mock CloudSat observations as well as quantitative and qualitative validation. The implementation of CICCAs allows analysis of convective cloud fraction inaccessible to CloudSat alone and shows results in consonance with analogous metrics in previous literature. In considering that, suggestions are given for continually improving and implementing CICCAs as an extension of CloudSat.

Three possible improvements to CICCAs are presented: higher resolution dataset, increased dimensionality and improved processing of data. The obvious limitation of CICCAs in its current form is both the spatial and temporal resolution afforded by CloudSat, MSG and GOES-East. The expansion of CICCAs beyond the Atlantic is just a matter of creating a new reference database for new regions. One way to do this is by making a new collocation dataset with GOES-16 data and CloudSat. Advanced Baseline Imager (ABI) operating in default "flex" mode will allow scanning every 15 minutes, a marked improvement from GOES installation in this paper (GOES-16 citation <https://www.goes-r.gov/spacesegment/abi.html>). ABI continuous disk mode will allow full disk images up to 5-minute resolution. The impact of improved resolution means more collocations between CloudSat and GOES, higher correspondence and better validation of skill (i.e. the two images can only be out of phase by as much as 7.5 minutes – 2.5 minutes). CICCAs can also be improved by increasing the dimensionality of the representative vector used as input in QDA. The redundancy between convective pixels and stratiform pixels is an obvious crutch to CICCAs and adding characteristic variables that differentiate those two precipitation modes is necessary for further improvement. One possible consideration is involving lightning flash rate. Donovan et al. (2008) shows the addition of lightning mappers improved skill of convective identification algorithms (Donovan et al., 2008). In considering GOES-16, the satellite is equipped with Geostationary Lightning Mapper (GLM) that measures total lightning occurrence at spatial resolution of 10 km (GOES-16 citation <https://www.goes-r.gov/spacesegment/abi.html>). This could easily be integrated into CICCAs. Another vector member could be cloud optical depth (COD), which is integrated into identification algorithm used in Sauter et al. 2019. One downfall of the integration of COD is the sophistication of deriving COD; use of future passive radiometers might not output this data as PATMOS-x does (Heidinger et al., 2012). Lastly, QDA may not

be ideally suited for the learning aspect of this algorithm. Further research could show other statistical or machine learning techniques or an ensemble of techniques to be more effective with the same input.

CICCA could have plenty of applications in industry as well as science and climate applications. Detection products such as Cloud Top Height (CTOP or CTH), Cloud Classification (CC) and Global Convective Diagnostic (GCD) exist that utilize visible and infrared imagery to detect hazardous convective events (Donovan et al., 2008). These products need improvement, especially in curtailing false alarms (Donovan et al., 2008). Convective Diagnosis Oceanic Algorithm (CDO) is now the state-of-the-art product; CDO incorporates GCD and CTH that generally consider BTD (Kessinger, 2017). CICCA in its current form is by no means fit to replace these detection algorithms, but possible future adaptations could be implemented in CDO. This considering the unprecedented training by CloudSat, areal footprint technique and probabilistic output, not present in other detection products (Donovan et al., 2008; Kessinger, 2017). CICCA is well suited to address existing problems in climate models and storm development. CRMs would benefit from CICCA-provided high-resolution validation of convective cloud fraction. Future work can be done in comparing diurnal and seasonal convective cloud variation between CICCA and existing climate models. Specifically, expanding the current scope of this paper's results for direct comparison to an array of CRMs and GCMs. Another aspect not mentioned previously is using CICCA in storm tracking and observing the life cycle of storms. The probabilistic outcome of CICCA can presumably measure strengthening and decaying as well as tracking of convective cores within large cloud shields, an application that is currently poorly represented by now casting and short-term forecasting products like NCAR's Convective Nowcasting Oceanic (CNO) products (Cai et al, 2008). More work can be done on observing storms within this CICCA dataset, especially in the context of environmental variables (e.g. Igel et al. 2015; Sauter et al., 2019). Understanding of storm development within environmental context will further improve the understanding of storm morphology and lead to more informed detection and prediction.