Performance Evaluation of Polarimetric Radio Occultation Measurements in Detecting Precipitation

Author: Ignacio Cordova Pou
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisors: Ramon Padullés (CSIC), Joan Bech (UB)
(Dated: June 16, 2022)

Abstract: For the first time, ROHP-PAZ mission is using polarimetric radio occultations measurements to estimate precipitation. This paper evaluates the performance of the averaged vertical profiles of the integrated differential phase shift $\Delta\Phi$ in detecting precipitation above a certain threshold. Using IMERG precipitation products as target, I developed an algorithm that computes the F1 Score for several averages between different heights h_o and h_f. I found that the optimal range for computing these averages starts at $h_o \lesssim 1\ km$ and reaches h_f between 5 km and 10 km. These averaged profiles achieve F1 Scores ranging from 0.4 to 0.6 depending on the precipitation threshold. Moreover, the results show that the optimal range of heights for computing the average shifts to higher altitudes when studying tropical regions. This could be due to higher levels of noise registered near the surface at low latitudes and to the vertical structure of convective precipitation which is predominant in these regions.

I. INTRODUCTION

Evaluating the performance of any remote sensing technique is of vital importance. A properly quantified evaluation can serve as judgment to the success of an entire mission and will give scientists the insights necessary to allocate resources of time and money in order to improve results, and that is precisely the objective of this paper. A Global Navigation Satellite System (GNSS) radio occultation (RO) experiment is taking place in the Spanish Low Earth Orbiter (LEO) satellite PAZ. The Radio Occultation and Heavy Precipitation experiment aboard PAZ (ROHP-PAZ mission) is the first to use the GNSS Polarimetric RO (PRO) technique (see FIG. 1), designed to test the capability of PRO to detect heavy rain events and other atmospheric phenomena [1].

RO is a technique for obtaining the vertical gradient of the atmospheric refractive index, from which we can derive thermodynamic properties such as temperature, pressure and water vapour. This is achieved by using a LEO to collect the signals transmitted by the GNSS systems once they have traveled through the atmosphere. The acquisition takes place when the LEO sets behind the horizon (occultation). This technique has been widely used and provides highly accurate atmospheric data [2]. The ROHP-PAZ mission not only provides these vertical thermodynamic profiles but also, for the first time, collects polarimetric information, which is the point of interest of this study. GNSS systems transmit circularly polarized light that is collected using two orthogonal linearly polarized antennas (horizontal H and vertical V). In the case of heavy rain events, large raindrops are flattened out along their horizontal dimension due to air friction. These raindrops cause depolarization between the H and V components which are affected differently when propagating through a rainy atmosphere. PROs allow us to measure the differential phase shift

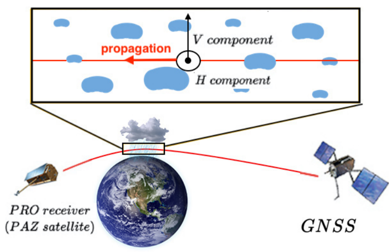

FIG. 1: GNSS signals experience depolarization in the presence of large raindrops and the PAZ receiver measures it at different heights. Image credit: Ramon Padullés.

$\Delta\phi$ between the H and V components. One of the objectives of this mission is to use the measured $\Delta\phi$ as a precipitation estimator and this paper evaluates the performance of $\Delta\phi$ on detecting precipitation above a certain threshold. By averaging the $\Delta\phi$ vertical profiles I obtained a scalar from which to retrieve a True/False class for precipitation. This True/False class extracted form the average d will be compared to our ground truth defined in section II. This study evaluates the performance of different averaging limits and provides the optimal range of heights $[h_o, h_f]$ to compute the averaged $\Delta\Phi$ that best detects precipitation. The optimal range $[h_o, h_f]$ will depend on the intensity of the precipitation we aim to detect and, as found in this study, on the region where the PRO took place. The performance of each average $[h_o, h_f]$ is evaluated using Precision-Recall curves which have been proven to be suitable for binary classifiers on imbalanced datasets [3]. I expect the average $[h_o, h_f]$ that best predicts the True/False target to be somewhere between 0.1 km and 10.0 km since that is where the presence of large raindrops and other hydrometeors will cause the most depolarization between the H and V components. Furthermore, I also performed a region segmentation (section III C.) to the dataset expecting to see a difference in performance between tropical and extratropical regions due to major differences in cloud formations and types of precipitation.

II. DATA

The main observable of the experiment is the differential phase shift $\Delta\phi$ as a function of height:

$$\Delta\phi(h) = \phi_H(h) - \phi_V(h). \tag{1}$$

The dataset contains over 85.000 measurements distributed globally providing vertical profiles (heights from 0.1 km to 40 km) of $\Delta\phi$ in 100 m intervals. To properly understand our data two PROs are shown in FIG. 2.

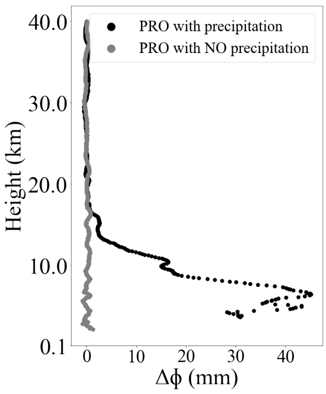

FIG. 2: In gray PRO registered at -53° 52' N -64° 32' W with start time on May 10th 2018 21:39:06 UCT and an average surface precipitation of 0.0 mm/h. In black PRO registered at 04° 21' N 82° 03' W with start time on May 17th 2018 at 01:38:31 UCT and an average surface precipitation of 9.5 mm/h

To properly evaluate the performance, a target must be specified. The target (ground truth) used in this paper was obtained from the NASA's IMERG "final" precipitation dataset. Surface rain in mm/h is obtained by co-location of the PAZ radio occultation profiles with IMERG rain products. It corresponds to the IMERG average across the area of the rays below 6 km, projected onto the Earth surface. The co-location approach is detailed in [4, Fig. 3]. The goal is to evaluate and quantify how well does the vertically averaged $\Delta\phi$ perform in detecting precipitation above a certain threshold regarded as *true precipitation.* This way, our target is defined as a binary True/False variable being True for precipitation above the percentile of choice and False for precipitation below it. This study covers targets built using the 90^{th} percentile (moderate precipitation) to the 99^{th} percentile (heavy precipitation) of our dataset.

For this work, I built a Python module with useful functions and algorithm implementations. The data processing tools I used are Python packages Pandas, Numpy, Matplotlib and Scikit-learn. These functions allow the user to perform data-cleaning and data-preprocessing, calculate averages of $\Delta\phi$ between two altitudes, define ground truths based on the percentile of interest, find the average with the optimal performance and many other useful actions. Please refer to the documentation and code which can be found at https://github.com/ignaciocordova/final_thesis.

III. METHODOLOGY

Each PRO provides a vertical profile. Given the array $\Delta\Phi$ containing the depolarization $\Delta\phi$ at each height (0.1 km - 40 km), the first step is to convert it into a scalar so that it can be compared to the target *true precipitation.* This is achieved by computing the average between two heights $[h_o, h_f]$ as:

$$\langle\Delta\Phi\rangle^{(j)}_{h_o\text{-}h_f} = \frac{1}{h_f - h_o} \sum_{i=h_o}^{h_f} \Delta\phi_i^{(j)}. \tag{2}$$

If the average is computed for each one of the j=1,...,m different PROs we obtain:

$$\left(\langle\Delta\Phi\rangle^{(1)}_{h_o\text{-}h_f}, \quad \dots \quad , \langle\Delta\Phi\rangle^{(m)}_{h_o\text{-}h_f} \right). \tag{3}$$

Now it is time to evaluate the performance of the average $[h_o, h_f]$ using the ground truth (4). This process will be repeated for different combinations of h_o and h_f using the algorithm explained in subsection III B.

$$(p_1, \quad \dots \quad , p_m) \tag{4}$$

A. Precision Recall curves

When studying the performance of a binary classifier on imbalanced classes, it is not enough to use *Accuracy* because it does not prevent the "all-false" cheating. This could happen by imposing a very restrictive threshold for the averages $\langle\Delta\Phi\rangle_{h_o\text{-}h_f}$ so that all result in *False.* Our target $(p_1, ..., p_m)$ will be mostly composed of no precipitation (False) achieving this way a high accuracy with zero predictive skill. That is why *Precision* (P) and *Recall* (R) must be introduced as:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN} \tag{5}$$

where TP is the n° of *True Positives* and FP and FN are the n° of *False Positives* and *False Negatives* respectively. We demand a high P and a high R for a model to be skillful. Using the F1 Score (6) one can obtain the binarization threshold that optimizes P and R, which are equally weighted as:

$$F1 = \frac{2PR}{P + R}. \tag{6}$$

B. Algorithm to compute F1 Scores

The main objective of this study is to find the averaging limits $[h_0, h_f]$ that result in the $\langle\Delta\Phi\rangle_{h_o\text{-}h_f}$ that best detects surface precipitation. I developed an algorithm (FIG. 3) for computing all F1 Scores iterating through a large set of different pairs $[h_0, h_f]$. I start by selecting a pair of heights, calculating the averages for all the PROs and normalizing them. The next step is to obtain the Precision and Recall curves using the normalized averages against the True/False target. These curves are obtained by calculating P and R using various thresholds to binarize the averages. These values can be represented in a PR curve (see FIG. 3) and the largest F1 Score is saved into a matrix that will contain the largest F1 Score for each one of the averages $[h_0, h_f]$ (see FIG. 4).

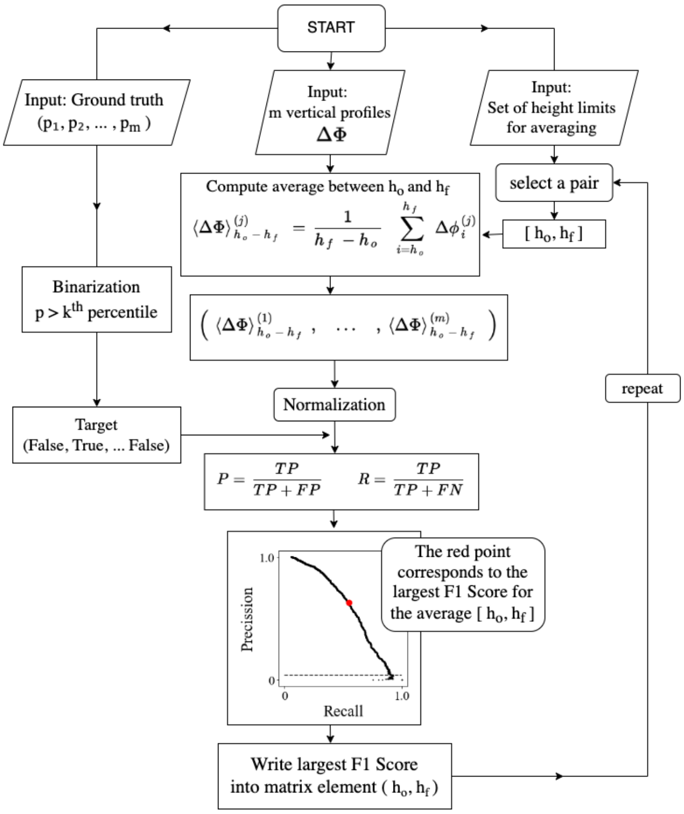

FIG. 3: Algorithm used to find the averaging range $[h_o, h_f]$ with the best F1 Score for a given percentile k.

Once all the averages $[h_o, h_f]$ have been compared against our target we can select the averaging range with the highest F1 Score. Note that values where $h_f < h_o$ have no physical meaning thus represented in white (in FIG. 4) as empty values.

FIG. 4 shows the performance of different averages $\langle\Delta\Phi\rangle_{h_o\text{-}h_f}$ as precipitation detectors. In this particular case, I found that the optimal range for averaging the vertical profiles of $\Delta\Phi$ corresponds to $h_o = 0.6\ km$ and $h_f = 9.3\ km$ with an optimal F1 Score of 0.534 (brightest point in FIG. 4). A close look to FIG. 4 also shows

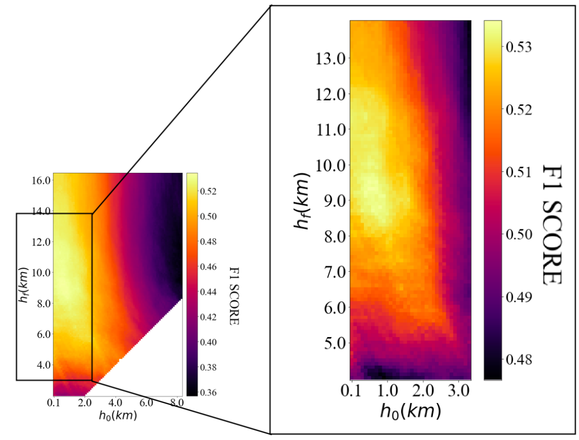

FIG. 4: Highest F1 Score for each average $[h_o, h_f]$ using a target of 95^{th} percentile as True precipitation.

that the average should be computed starting from values $h_o \lesssim 2km$, meaning that useful information to detect precipitation is obtained from signals traveling below this height. In other words, as an arbitrary example: if we were to use $\langle\Delta\Phi\rangle_{3km\text{-}9.3km}$ instead of $\langle\Delta\Phi\rangle_{0.6km\text{-}9.3km}$ as a precipitation estimator, the performance would be worse.

C. Region segmentation

There are many factors playing an important role in determining the performance of this remote sensing technique. This section aims to analyze the changes in performance of PROs registered in tropical regions. The main differences between tropical and extratropical regions are cloud formations, vertical structure, rain intensity and even the raindrops' shape. I expect the combination of these factors to result in a difference in performance depending on where the PRO took place. The metadata contains information of latitude and longitude which I used to segment the dataset into two: one of PROs collected in latitudes $< 30^o$ for the tropical PROs and one in latitudes $> 30^o$ for the extratropical PROs. Following the steps described in the previous section, I proceeded to find the averaged depolarization $\langle\Delta\Phi\rangle_{h_o\text{-}h_f}$ that best accounts for precipitation. The results showing different performances and optimal averaging ranges are discussed in the next section.

IV. RESULTS AND DISCUSSION

Applying the algorithm described in the previous section I was able to obtain the optimal range of heights for computing the average of the vertical profiles $\Delta\Phi$. The results presented in TABLE 1 show the best performing average for each percentile. Since the Target changes from one percentile to another, these results are not meant to serve as comparison between them. Instead, they provide information about the expected performance when trying to detect precipitation above a certain threshold.

As expected, the optimal averages contain depolarization measures on heights ranging between $h_o \sim 0.1$ km and $h_f \sim$ from 5.0 km to 10.0 km. This shows a high correlation between surface precipitation (target) and not only raindrops but also other hydrometeors present in these altitudes, as shown in [5].

Precipitation percentile	True Precipitation threshold (mm/h)	Optimal range $[h_o, h_f]$ (km)	True/False $\langle\Delta\Phi\rangle$ threshold (mm)	F1 Score	Precision	Recall
99	4.53	0.1-8.2	7.56	0.396	0.327	0.504
98	3.11	0.6-7.8	6.69	0.449	0.434	0.466
97	2.64	0.1-5.6	6.12	0.494	0.452	0.544
96	2.03	0.7-9.3	3.84	0.510	0.448	0.593
95	1.78	0.6-9.3	3.82	0.534	0.519	0.550
94	1.54	0.1-10.2	3.10	0.544	0.515	0.577
93	1.37	0.4-8.2	3.22	0.556	0.518	0.600
92	1.27	0.1-6.7	3.33	0.569	0.525	0.620
91	1.20	0.2-9.0	2.44	0.578	0.526	0.642
90	1.00	0.2-8.6	2.40	0.591	0.544	0.646

TABLE I: Best performing average for detecting precipitation above the *True Precipitation threshold (mm/h).*

The contingency table in FIG. 5 provides a detailed analysis of the performance of the optimal average $\langle\Delta\Phi\rangle_{0.7km\text{-}9.3km}$ on detecting surface precipitation above 2.03 mm/h. The table shows how the binarization performed by my algorithm correctly classifies the majority of the False class (no precipitation event). On the one hand, False Positives (upper right box in FIG 5.) indicate that some other phenomenon that is not represented by the target used here (surface precipitation) is causing depolarization, thus showing a "Yes" although surface precipitation is not above 2.03 mm/h. On the other hand, False Negatives (bottom left box in FIG. 5) show that surface precipitation above 2.03 mm/h was detected by IMERG, but the average depolarization registered by the LEO PAZ was below 3.84 mm, thus not detecting the above mentioned precipitation.

While the technique works well in the vast majority of cases, I try to present an explanation for the missed classifications. Regarding the False Negatives, one important aspect of the target used to evaluate the performance is the temporal resolution. While the IMERG surface precipitation products have a 30-minute time resolution, our PROs take less than two minutes to register a complete vertical profile. This can result in the IMERG detecting precipitation at some point in time when the GNSS signal had already passed (or the PRO had not even begun). As for the False Positives, they could be reduced by using a different, more complex target, that includes the vertical structure of precipitation and the presence of other hydrometeors, for example, the use of spaceborne radars, although then the difficulty arises from the low number of coincident measurements between PAZ-LEO and the radar observations. Another possible contribution to the number of missed classifications (both False Negatives and False Positives) is the miss-co-location due to uncertainties in the PRO location [4].

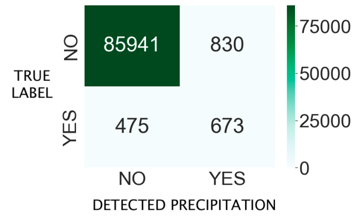

FIG. 5: Contingency table for detecting surface precipitation above 2.03 mm/h using the best performing average $\langle\Delta\Phi\rangle_{0.7km\text{-}9.3km}$

A. Performance in tropical regions.

After performing region segmentation, the optimal ranges for averaging the vertical profiles to detect surface precipitation above different thresholds are shown in TABLE II. It is very interesting to see how the optimal range for computing the average in tropical regions is completely different to the one obtained in TABLE I. Whereas in the general performance I obtained optimal ranges starting below ~ 1 km, in the case of tropical regions the optimal averages start at heights of around $h_o \sim 3.5$ km. This result can be attributed to the fact that the polarimetric measures near the surface present higher levels of noise in the tropical regions [4]. The rejection of measures with high levels of noise (0 km - 3 km) results in an increase in the performance.

Precipitation percentile	True Precipitation threshold (mm/h)	Optimal range $[h_o, h_f]$ (km)	True/False $\langle\Delta\Phi\rangle$ threshold (mm)	F1 Score	Precision	Recall
99	4.63	6.6-13.1	9.44	0.500	0.630	0.415
98	3.30	4.5-9.0	9.30	0.518	0.528	0.509
97	2.60	1.1-10.0	6.22	0.534	0.498	0.575
96	2.19	0.3-12.9	4.78	0.545	0.541	0.549
95	1.91	3.4-14.3	3.85	0.579	0.582	0.577
94	1.73	3.6-18.2	2.80	0.598	0.650	0.554
93	1.52	3.3-18.4	2.47	0.593	0.638	0.554
92	1.34	4.5-13.9	2.50	0.605	0.562	0.654
91	1.21	4.5-14.2	2.42	0.607	0.596	0.619
90	1.10	3.1-13.2	2.45	0.619	0.589	0.651

TABLE II: Results for tropical regions.

Another important result is the fact that the optimal averaging range extends to much higher altitudes h_f. TABLE II shows $h_f \sim 13.7km$ which implies that depolarization at these altitudes appears to be useful to detect surface precipitation. This can be due to the fact that tropical regions present predominantly convective rain with precipitation sometimes extending to very high altitudes, as found in [6], where it is shown that regions with precipitation above 15 km are found mostly over tropical land and the West Pacific Warm Pool (tropical oceanic region).

Finally, Figure 6 shows that the detection of precipitation is achieved with higher performance in tropical regions. This means that the F1 Score of the best tropical average is, for any given percentile of interest, higher than the F1 Score corresponding to the best extratropical average. This can be caused by the fact that the IMERG precipitation products are obtained using radiometers from geostationary satellites (IR) which perform better when detecting convective precipitation, predominant on the tropics.

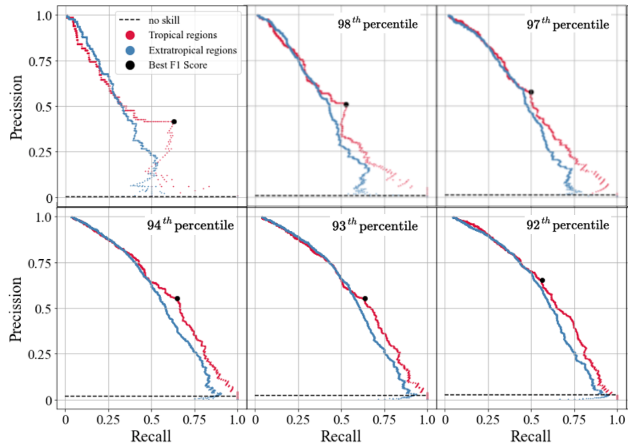

FIG. 6: Precision-Recall curves of the optimal $\langle \Delta\Phi \rangle_{h_o\text{-}h_f}$ for different percentiles of interest. Each curve shows the performance of the best performing average.

V. CONCLUSIONS

This paper evaluates the performance of using vertically averaged polarimetric radio occultation measurements to detect surface precipitation. The performance is evaluated using the F1 Score achieved by the averages when detecting precipitation above a certain threshold. As ground truth I have used the IMERG rain products averaged across the area of rays below 6 km, projected onto the Earth surface. I found that the optimal range for computing the average is between heights of $h_o \lesssim 1\ km$ and h_f ranging from values of 5 km to 10 km. Moreover, when studying specifically tropical regions, I show how the optimal range of heights shifts to values of $h_o \sim 3.5\ km$ and $h_f \sim 14\ km$ due to higher levels of noise at low altitudes and the presence of convective precipitation at high altitudes. The performance obtained (F1 Score) presents values ranging from 0.4 for very imbalanced classes to 0.6 for less imbalanced. This performance largely surpasses that of a random model and shows very promising results for this new technique. Some ways to improve the performance obtained include:

1. Trying to identify what is causing the False Positives. I suggest a search of False Positives to then find coincident radar measurements to analyze if the IMERG is failing to provide accurate precipitation data on those measurements.

2. Incorporating the use of targets that include vertical structure of precipitation and the presence of other hydrometeors, for example, data from space and ground based radars. The problem would be the low number of coincident measures between PAZ and the radar observations which won't allow for statistical studies but only for analysis of individual measurements.

3. Increasing the quality of the polarimetric measurements at low altitudes (below 1-2 km) which present a considerable amount of "empty" values near the surface.

VI. ACKNOWLEDGEMENTS

I thank my advisor Ramon Padullés for his guidance and commitment. He is undoubtedly the kind of mentor any aspiring scientist would want to work with. I also thank Joan Bech for his advice and expertise and Gonzalo Córdova for all our insightful discussions.

[1] Cardellach, E., Tomas, S., Oliveras, S., Padullés, R., Rius, A., De la Torre-Juárez, M., Turk, F.J., Ao, C.O., Kursinski, E.R., Schreiner, B., Ector, D., Cucurull, *Sensitivity of PAZ LEO Polarimetric GNSS Radio-Occultation Experiment to Precipitation Events, IEEE Transactions on Geoscience and Remote Sensing, 2014, 53(1), doi:10.1109/TGRS.2014.2320309*

[2] Anthes, R. (2011). Exploring Earth's atmosphere with radio occultation: Contributions to weather, climate and space weather, Atmos. Meas. Tech., doi:10.5194/amt-4-1077-2011

[3] Saito, T., Rehmsmeier, M., *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015 Mar 4;10(3):e0118432. doi:10.1371 /journal.pone.0118432. PMID: 25738806; PMCID: PMC4349800.*

[4] Padullés, R., Ao, C. O., Turk, F. J., de la Torre Juárez, M., Iijima, B., Wang, K. N., and Cardellach, E., Calibration and validation of the Polarimetric Radio Occultation and Heavy Precipitation experiment aboard the PAZ satellite, Atmos. Meas. Tech., 13, 1299–1313, 2020, doi:10.5194/amt-13-1299-2020

[5] Padullés, R., Cardellach, E., Turk, F.J., Ao, C.O., de la Torre Juarez, M., Gong, J., and Wu, D.L. Sensing Horizontally Oriented Frozen Particles With Polarimetric Radio Occultations Aboard PAZ: Validation Using GMI Coincident Observations and Cloudsat a Priori Information. 2022. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-13.

[6] Liu, C., and E. J. Zipser (2015), The globaldistribution of largest, deepest, andmost intense precipitation systems, Geophys. Res. Lett.,42, 3591–3595,doi:10.1002/2015GL063776.