

PAPER • OPEN ACCESS

Performance of Principal Component Analysis to Classify Precipitation Type from Raindrop Size Distribution Data at Kototabang, Indonesia

To cite this article: Marzuki *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **303** 012054

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Performance of Principal Component Analysis to Classify Precipitation Type from Raindrop Size Distribution Data at Kototabang, Indonesia

Marzuki¹, H Hashiguchi², M Vonnisa¹, Harmadi¹, and T Shimomai³, M Yoseva¹ and E Saufina¹

¹Department of Physics, Andalas University, Padang, Indonesia.

²Research Institute for Sustainable Humanosphere (RISH), Kyoto University, Japan.

³Interdisciplinary Faculty of Science and Engineering, Shimane University, Japan.

marzuki@sci.unand.ac.id

Abstract. This study examines the use of principal component analysis (PCA) to classify the RDSD data at Kototabang, Indonesia. In addition to PCA with 6 attributes (hereinafter called PCA6) that had been developed by a previous researcher, this study also examines PCA with 7 attributes (PCA7) by adding radar reflectivity factor. The PCA is applied to the RDSD that had been classified by a wind profiler into Stratiform (S), deep convective (DC), shallow convective (SHC) and mixed stratiform/ convective (MSC). The number of unclassified data is much smaller than that reported by previous study in which it is about 33-47% with PCA6 and 29-44% with PCA7. While the PCA classifies the same group for different rain type from wind profiler, especially for Group I (moderate D_0 and large N_w) and II (small D_0 and N_w), some differences are observed. Each rain type classified by wind profiler has different dominant group in which Group II is dominant for S, Group I and V (large D_0 and low N_w) are dominant for DC, Group I, IV (small D_0 and large N_w) and VI (small D_0 and very large N_w) are dominant for SHC and Group I is also dominant for MSC type.

Keywords : principal component analysis, MSC, deep convective, raindrop.

1. Introduction

Rain type classification is very important because each rain type undergoes different physical processes. The most common type of rain is stratiform and convective. The cloud dynamics of these two rain types are different so that the microphysical processes that affect the growth of hydrometeor in both clouds are also different. Convective cloud is characterized by vertical air motion which is much stronger than that associated with stratiform rain. Hydrometeor in convective core grows through riming or accretion processes which generate large and dense raindrops. On the other hand, hydrometeor in stratiform rain grows through vapor deposition and aggregation processes, lead to smaller and less dense hydrometeors. Once melted, hydrometeor in stratiform produce more small-sized drops [1]. The difference in raindrop size distribution (RDSD) between convective and stratiform led to the importance of separation of these two types of rain for weather radar data conversion [2].

In general, classification of precipitation can be done in two ways, namely using radar and RDSD data. Precipitation can be classified using the texture of the radar reflectivity fields from weather radar



[3]. The precipitation type can be also deduced by analysing the vertical structure of reflectivity, velocity, and spectral width derived from the vertical beam of wind profiler observation [4]. Rain classification based on radar data is very accurate and it does not depend on rainfall intensity. However, radar observation particularly wind profiler is limited to several locations. As an alternative, classification can be done using the characteristics of RDSO. Rain classification deduced by analysing the RDSO can be based on the relationship between RDSO parameters and rainfall intensity. For example, the relationship between median mean diameter (D_0) and normalized intercept parameter of gamma distribution (N_w) given by the equation of $\log_{10} N_w = -1.6D_0 + 6.3$ separates RDSO into two clusters (convective-stratiform) [5]. Hereinafter, this method is called as the BR09 method. Recently, $N_w = 3.85$ is recommended as a threshold between convective and stratiform, and hereinafter this method is abbreviated as TH15 [6].

Rain classification which is based on the RDSO parameters depends on rainfall intensity. Precipitation with rainfall rate more than 10 m h^{-1} is always classified as convective rain and vice versa, precipitation with rainfall rate less than 5 m h^{-1} is classified as stratiform. This classification is not wrong because convective and stratiform rains are generally characterized by high and low rainfall intensity, respectively. However, not all rains with low intensity are stratiform type. Precipitation with low rainfall rate sometimes also occurs in convective rain [7]. To overcome this limitation, it has recently been developed rain classification using principle component analysis (PCA) [8].

The PCA is a technique which is commonly used to observe the variability of complex and large number of data. It uses linear regression to determine the main mode of variability in data. The PCA does not require the assumption of rainfall rate. However, previous study found that the PCA did not classify large number of data (70%) into any group [8]. Therefore, this work examined the application of PCA to classify the RDSO data in Indonesia. The performance of PCA was examined using the RDSO with known type of rain which has been classified using wind profiler observation [9].

2. Data and Method

This work applies the PCA to RDSO data collected by a 2D-Video Disdrometer (2DVD) at Kototabang, West Sumatra (100.32° E , 0.20° S), over eight consecutive years (2003-2010). An overview of 2DVD observation at Kototabang can be found in two publications [9,10]. The RDSO data have a temporal resolution of 2 minutes, adopting 0.2 bin interval from 0.4 to 10.25 mm. We only consider the RDSO with rainfall rate larger than 0.1 mm h^{-1} and having four consecutive bins with non-zero value. The RDSO data have been classified into stratiform (S), deep convective (DC), shallow convective (SHC) and mixed stratiform/convective (MSC) rains, using a 1.3 GHz wind profiler [9].

The PCA requires $M \times N$ array as input where M is the number of attributes that describe the characteristics of RDSO and N is the number of data points. In this study, we examined the performance of PCA with $M = 6$ and $M = 7$. For $M = 6$, the attributes consist of rainfall rate (R), liquid water content (LWC), mass-weighted mean diameter (D_m), total number of raindrops (N_t), normalized intercept parameter of gamma distribution (N_w) and the normalized standard deviation of the mass spectrum (σ_m), which are expressed as:

$$R = 6\pi \cdot 10^{-4} \int_0^{\infty} D^3 v(D) N(D) dD \quad (1)$$

$$LWC = \frac{\pi \cdot 10^{-3}}{6} \int_0^{\infty} D^3 N(D) dD \quad (2)$$

$$D_m = \frac{\int_0^{\infty} D^4 N(D) dD}{\int_0^{\infty} D^3 N(D) dD} \quad (3)$$

$$N_w = \frac{4^4}{\pi \rho_w} \frac{LWC}{D_m^4} \quad (4)$$

$$\sigma_m = \sqrt{\frac{\int_0^{D_{\max}} (D - D_m)^2 D^3 N(D) dD}{\int_0^{D_{\max}} D^3 N(D) dD}} \quad (5)$$

The parameters of N_w , N_t , R , and LWC are included in the PCA in log scale. For $M = 7$, the value of Z is added into the array, which is expressed as

$$Z = \int_0^{\infty} D^6 N(D) dD. \quad (6)$$

The value of Z is the most widely used integral rainfall parameter because this parameter is directly provided by weather radar. The parameter of Z is also included in the PCA in log scale.

The PCA produces a set of vectors called empirical orthogonal functions (EOFs). This work only uses two EOFs, namely, EOF-1 and EOF-2, as in previous study [8]. EOF-1 is a vector that represents the largest amount of variance and EOF-2 describes the largest amount of variance after the first vector is removed. The PCA is only applied to the RDSD with shape parameter (μ) more than -3 and less than 15. The shape parameter is expressed by

$$\mu = \left(\frac{\sigma_m}{D_m} \right)^2 - 4. \quad (7)$$

3. Result and Discussion

Figure 1 (a) shows the distribution of PC1 and PC2 for all 2DVD data at Kototabang obtained using 6 attributes (PCA6). The dashed line shows the 1.5 threshold as used in the previous study [8]. Data were grouped into 6 groups: Group I (positive PC1 but ambiguous PC2), Group II (negative PC1), Group III (positive PC2 but ambiguous PC1), Group IV (negative PC2 and ambiguous PC1), Group V (positive PC1 and positive PC2) and VI (positive PC1 and negative PC2). Of 8-year data, about 58% of the data are classified into these 6 groups and 48% data are unclassified. The unclassified data in the current study is smaller than that obtained by the previous study ($\sim 77\%$) [8]. EOF1 and EOF 2 explain about 59% and 30% of the variability, respectively.

Although the name of the group in this study is the same as in previous study [8], the physical meaning of each group is different. In this study, Group I means moderate D_0 and large N_w , Group II indicates small D_0 and N_w , Group III means moderate D_0 and small N_w , Group IV indicates small D_0 and large N_w , Group V means large D_0 and low N_w , and Group VI indicates small D_0 and very large N_w (figure 2). Kototabang RDSD has the largest number of data point in Group I and II about 16% and 20%, respectively. Group I is classified as convective rain and Group II is classified as stratiform by TH15 method.

Figures 1b-e show the distribution of 6 DSD groups in 4 types of rain classified by wind profiler. The stratiform rain has a large number of point with negative PC1 and deep convective rain has a large number of points in positive PC1. This characteristic is also observed in mixed stratiform/convective and shallow convective rains. Moreover, shallow convective has a larger number of points with negative PC2 than other rain types. Number of unclassified data for stratiform, deep, shallow and mixed stratiform/convective is 47, 45, 34, and 33%, respectively (table 1). In stratiform rain, the largest number of data is observed in Group II and the largest number of data is in Group I and V for deep convective rain. For shallow convective rain, the largest percentage of data points lie in Group I, followed by Group IV, II and VI. In mixed stratiform/convective, the data are mostly concentrated in Group I. Thus, the largest number of data point for convective rain lie in Group I, while it is in Group II for stratiform rain. This is consistent with the DSD classified by TH15 method.

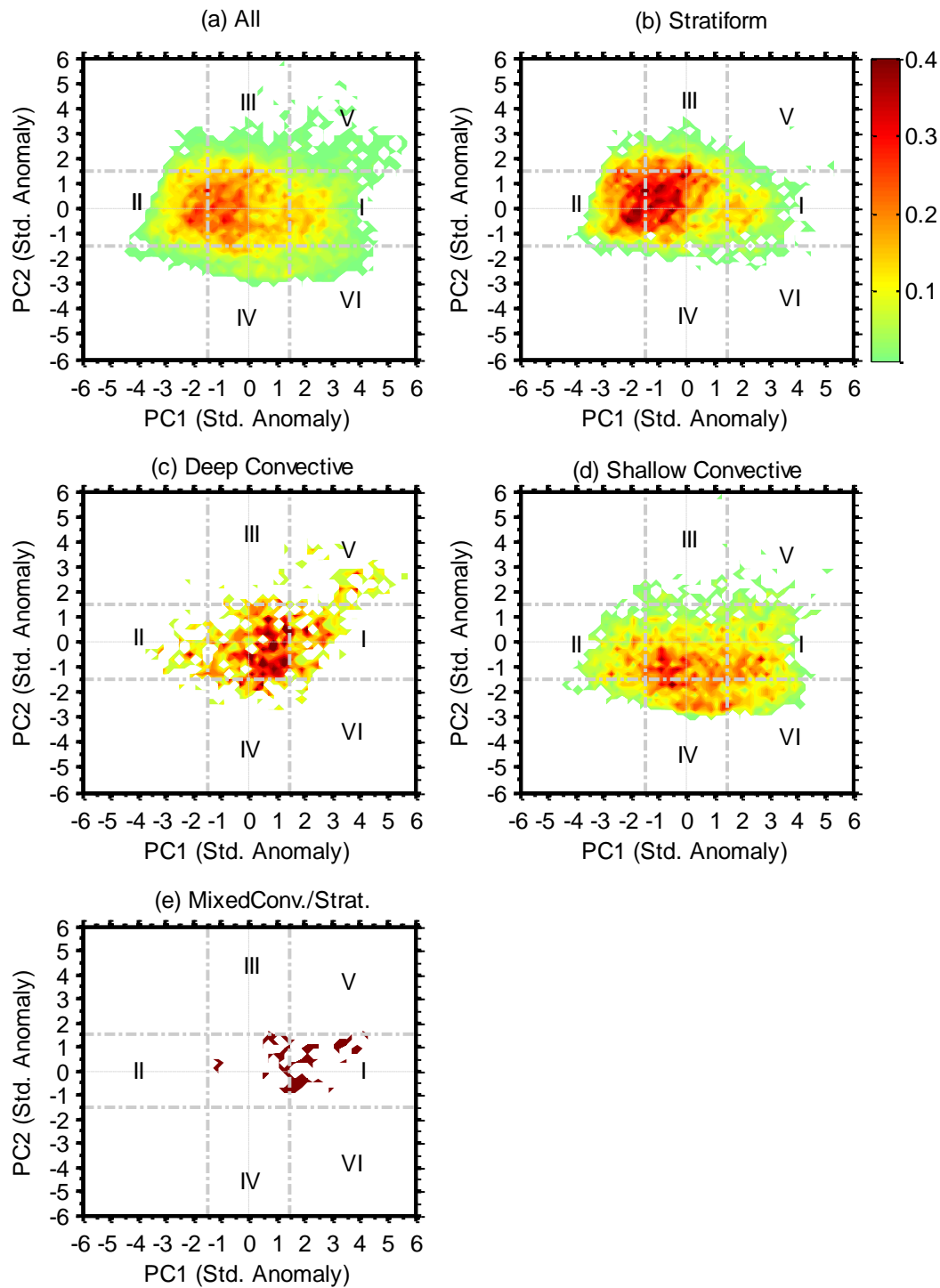


Figure 1. Distribution of normalized frequency of PC1 and PC2 estimated using 6 parameters. Dashed gray lines indicate $\sigma_{1.5}$ thresholds. All (a) indicates all data without rain classification, stratiform (b), deep convective (c), shallow convective (d) and mixed convective/stratiform (e) indicate raindrop spectra classified by wind profiler.

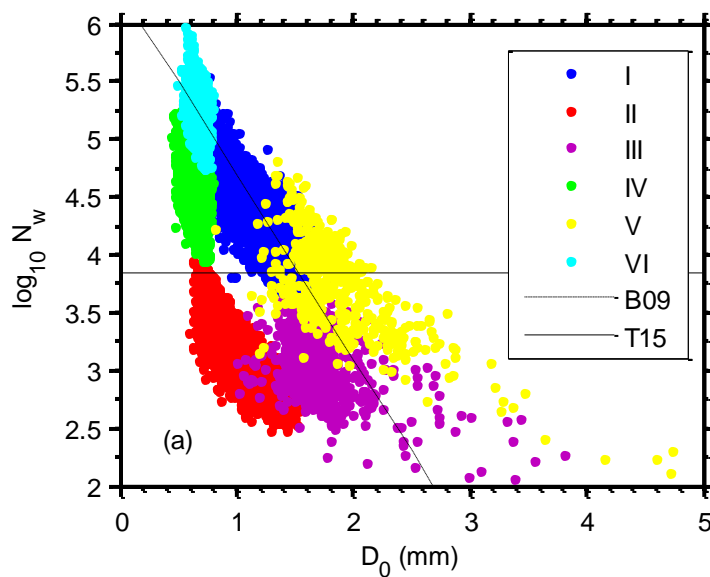


Figure 2. Two dimensional distribution of each group for $\log_{10}N_w - D_0$. Lines from BR09 and TH15 methods are also given.

Table 1. Percentage of data for each DSD group. U indicates unclassified data.

| Type | Data Number | 6 Parameters (%) | | | | | | 7 Parameters (%) | | | | | | | |
|------|-------------|------------------|----|-----|----|----|----|------------------|----|----|-----|----|----|----|----|
| | | I | II | III | IV | V | VI | U | I | II | III | IV | V | VI | U |
| All | 13991 | 16 | 20 | 7 | 8 | 3 | 4 | 42 | 16 | 20 | 8 | 9 | 4 | 5 | 38 |
| ST | 7499 | 12 | 27 | 10 | 2 | 1 | 1 | 47 | 12 | 25 | 13 | 2 | 2 | 2 | 44 |
| DC | 1415 | 18 | 9 | 5 | 8 | 13 | 2 | 45 | 24 | 12 | 4 | 9 | 12 | 2 | 38 |
| SHC | 4828 | 20 | 13 | 3 | 17 | 3 | 10 | 34 | 17 | 17 | 3 | 19 | 3 | 10 | 31 |
| Mix | 249 | 43 | 8 | 7 | 0 | 7 | 2 | 33 | 49 | 8 | 5 | 1 | 6 | 3 | 29 |

Figure 3 shows the distribution of PC1 and PC2 for all 2DVD data classified by PCA with 7 attributes (PCA7). EOF1 and EOF2 explain about 61% and 30% variance, respectively. In general, the distribution of data is not much different from figure 1. However, we can see some differences. The number of unclassified data in figure 3 is smaller than Figure 1 and it is also observed for all rain types classified by wind profiler. The number of unclassified data decreases about 3-7% from figure 1. In deep convective and mixed stratiform/convective, there is an increase of data number in Group I. There is a decrease the data number in Group II and an increase in Group III for stratiform rain. The number of data in Group III that exceeds the BR09 line is also smaller in figure 3 than that in figure 1. In addition, the percentage of data in Group I distributed below the TH15 line also increases (figure 4). From figures 2 and 4, we can see some limitation of BR09 and TH15 in classifying the RDSD at Kototabang. BR09 does not provide a clear separation between stratiform and convective in which BR09 line lies in the middle of Group I, III, V and VI. TH15 method seems better because it only splits one group namely, Group V.

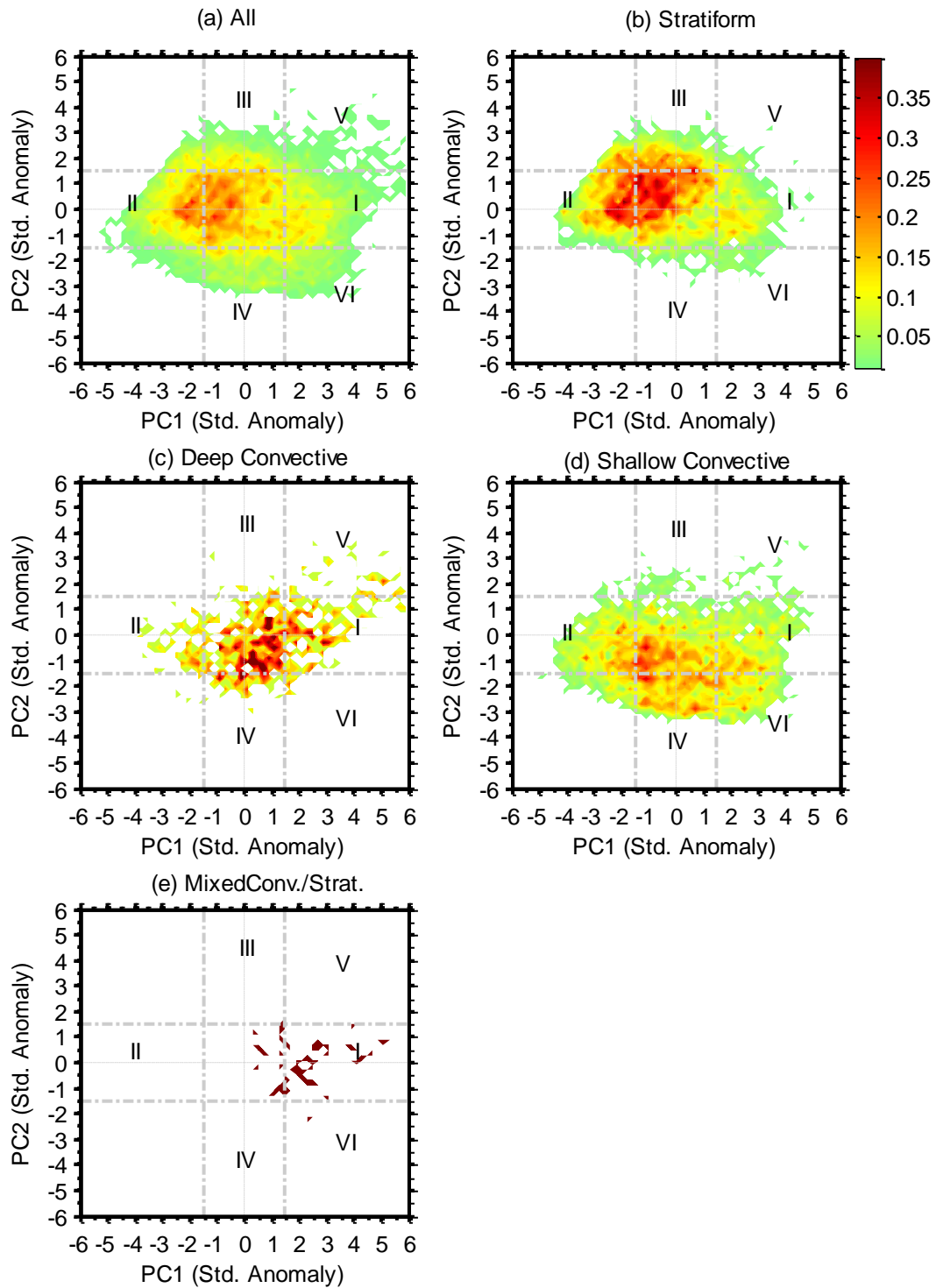


Figure 3. Same as figure 1 but for PCA with 7 attributes.

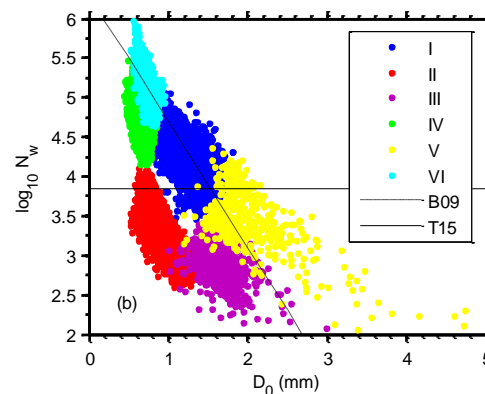


Figure 4. Same as figure 2 but for PCA with 7 attributes.

4. Conclusion

The results of this study show that the PCA can be used to classify precipitation types through RSD observation in Indonesia. The PCA with 7 attributes provides a smaller number of unclassified data than that of 6 attributes. In general, characteristics of RSD for each group which is classified by the PCA are consistent with the characteristics of DSD based on wind profiler classification scheme. The current study only analyzed data at one location. Data from several locations in Indonesia are being analyzed and will be published in another article. Moreover, RSD in Indonesia has significant diurnal, intra-seasonal and regional variations. Application of PCA to observe such variability is also being examined.

Acknowledgments

This work was supported the Ministry of Research, Technology and Higher Education of the Republic of Indonesia under the 2017 and 2018 International Joint Collaboration and Scientific Publication grant (Contract no. 02/UN.16.1.17/PP.KLN/LPPM/2017 and 07/UN.16.17/PP.PKLN.K/LPPM/2018).

References

- [1] Houze Jr R A 1993 *Cloud Dynamics* (California: Academic Press)
- [2] Tokay A and Short D A 1996 Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds *J. Appl. Meteor.* **35** pp 355–71
- [3] Steiner M, Houze Jr R A and Yuter S E 1995 Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data *J. Appl. Meteor.* **34** pp 1978–2007
- [4] Williams C R., Ecklund W L and Gage K S 1995 Classification of precipitating clouds in the tropics using 915-MHz wind profilers *J. Atmos. Oceanic Technol.* **12** pp 996–1012
- [5] Bringi V N, Williams C R, Thurai M and May P T 2009 Using dual-polarized radar and dual-frequency profiler for DSD characterization: a case study from Darwin, Australia *J. Atmos. Oceanic Technol.* **26** pp 2107–22
- [6] Thompson E J, Rutledge S A, Dolan B and Thurai M 2015 Drop size distributions and radar observations of convective and stratiform rain over the equatorial Indian and West Pacific Oceans *J. Atmos. Sci.* **72** pp 4091–25
- [7] Tokay A, Short D A, Williams C R, Ecklund W L and Gage K S 1999 Tropical rainfall associated with convective and stratiform clouds: intercomparison of disdrometer and profiler measurements *J. Appl. Meteor.* **38** pp 302–20
- [8] Dolan B, Fuchs B, Rutledge S A, Barnes E A and Thompson E J 2018 Primary modes of global drop size distributions *J. Atmos. Sci.* **75** pp 1453–76
- [9] Marzuki, Hashiguchi H, Kozu T, Shimomai T, Shibagaki Y, Takahashi Y 2016 Precipitation microstructure in different Madden–Julian Oscillation phases over Sumatra *Atmos. Res.* **168** pp 121–38
- [10] Marzuki, Randeu W L, Kozu T, Shimomai T, Hashiguchi H and Schoenhuber M 2013 Raindrop axis ratios, fall velocities and size distribution over Sumatra from 2D-video disdrometer measurement in advances in precipitation science ed Michaelides S *Atmos. Res.* **119** pp 23–37