Evaluation of Precipitation Instrument Performance

Background

Accurate rainfall is crucial to a wide array of applications including radar observations, modeling of climate and hydrological processes, and urban impacts. Numerous types of instruments, ranging from optical to mechanical systems, are available for measuring precipitation, each having their own benefits and deficiencies. However, not all sensors measure precipitation the same way and these differences in methods can lead to substantial differences in the measured accumulations and rates. A recommendation from a World Meteorological Organization (WMO) report that assessed the performance of twenty-six types of rain gauges was that the development of a "composite working reference" was needed to calculate a best estimate (Lanza et al 2009). The WMO report also concluded that instruments need to be improved to reduce the uncertainty of measurements at intensities below 20 mm/hr.

The Atmospheric Radiation Measurement user facility (ARM) Southern Great Plains observatory (SGP) houses over 50 precipitation instruments, utilizing approximately 10 different methods for measurement (Table 1). Each of the instruments listed in Table 1 are deployed at the SGP central facility with duplicates of the video disdrometer, impact disdrometer, and laser disdrometer. These instruments are deployed relatively close to one another in the "Central Cluster" of the SGP central facility (Fig. 1). This effort employed simple clustering techniques on these data to calculate a precipitation best estimate which in turn was used to develop a weighting scheme for each instrument type across rain rates that can be reused in the future for different configurations of instrument deployments.

Table 1. Table of the various types of precipitation sensors deployed at the ARM SGP site.

Instrument Type	ARM System	Install Year
NovaLynx Tipping Bucket	MET/PRECIPMET	1993/2017
Texas Electronics Tipping Bucket	STAMP	2016
Vaisala Present Weather Detector	MET	2010
Optical Scientific Optical Rain Gauge	ORG/MET	2007
Distromet Impact Disdrometer	DISDROMETER	2006
Joanneum Research Video Disdrometer	VDIS	2011
Vaisala WXT-520	AOSMET/MWR3C/PRECIPMET	2016/2011
OTT Hydromet Parsivel2 Disdrometer	PARS2/LD	2016
OTT Hydromet Pluvio2 Weighing Bucket	WBPLUVIO2	2016

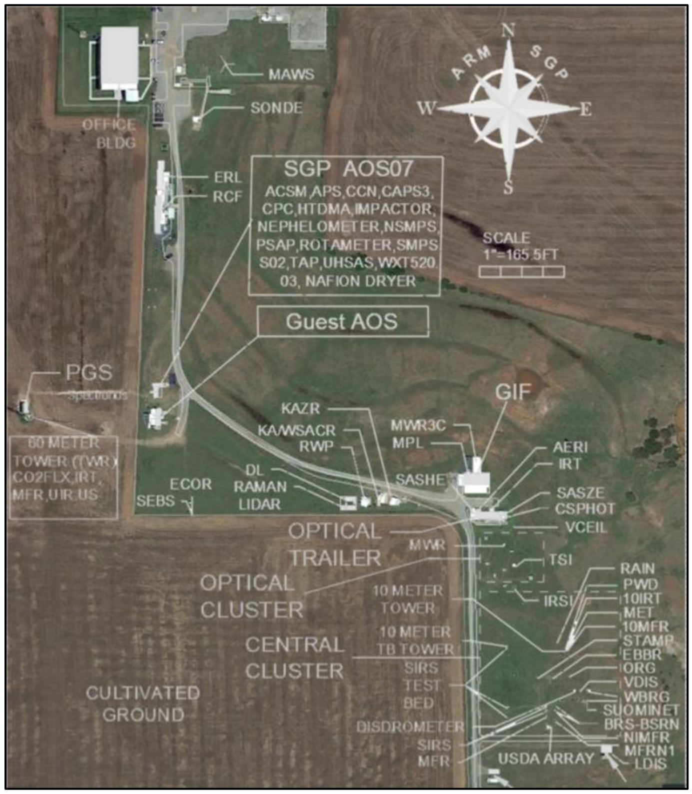

Figure 1. Instrument layout at the ARM SGP central facility. All the precipitation sensors, with the exception of the WXT deployed as part of AOSMET, are deployed in the "Central Cluster" in the lower right corner.

Methodology

Quality Control

While ARM provides machine readable data quality reports (DQRs) that could be used to flag data, they can sometimes flag too much or too little data and only the DQRs that flagged the data as incorrect were used to exclude data from the analysis. They did provide information on a problem with the optical rain gauge that occurred throughout a majority of analysis period which caused it to overreport precipitation. This provided a good opportunity to implement a broad quality control mechanism that's consistent across all instruments.

The Parsivel and 2-dimensional video disdrometer provide measurements of drop size and fall speed. PyDSD was used to filter the Parsivel data and recalculate the rain rates (Hardin and Guy 2017). Additionally, a script created by Bobby Jackson, was used to filter the 2-D video disdrometer data to do similar filtering of the data. This removed drops that were ±50% from the expected terminal velocity for the drop diameter. The data for both instruments were also filtered based on the drop diameter, removing drops outside the range of 0.2 mm to 5 mm.

Data from all instruments were read in to an xarray object along with additional environmental information such as temperature, wind speed and direction, relative humidity, atmospheric pressure, liquid water path, and precipitable water vapor using the Atmospheric data Community

Toolkit (Theisen et al. 2020). These data were converted to rain rate and then resampled to 1-minute averages to ensure a standard timeframe. Solid precipitation is difficult to measure and introduces complications in the analysis, so all data where temperatures were less than 2ºC were removed.

In order to quality control the data in an automated method, each day was processed separately to remove erroneously high precipitation rates. Daily accumulations outside of 1 standard deviation from the mean of all instruments were subjected to additional quality control processing. If this upper limit was greater than 200 mm/hr then the limit was adjusted to be 2 standard deviations from the median value. If this threshold was above 100 mm/hr it was adjusted to be 3 standard deviations from the median. This additional quality control sets data outside the 0.999 quantile to 0. In order to improve processing and data storage efficiency, only times during which precipitation was recorded were written to file.

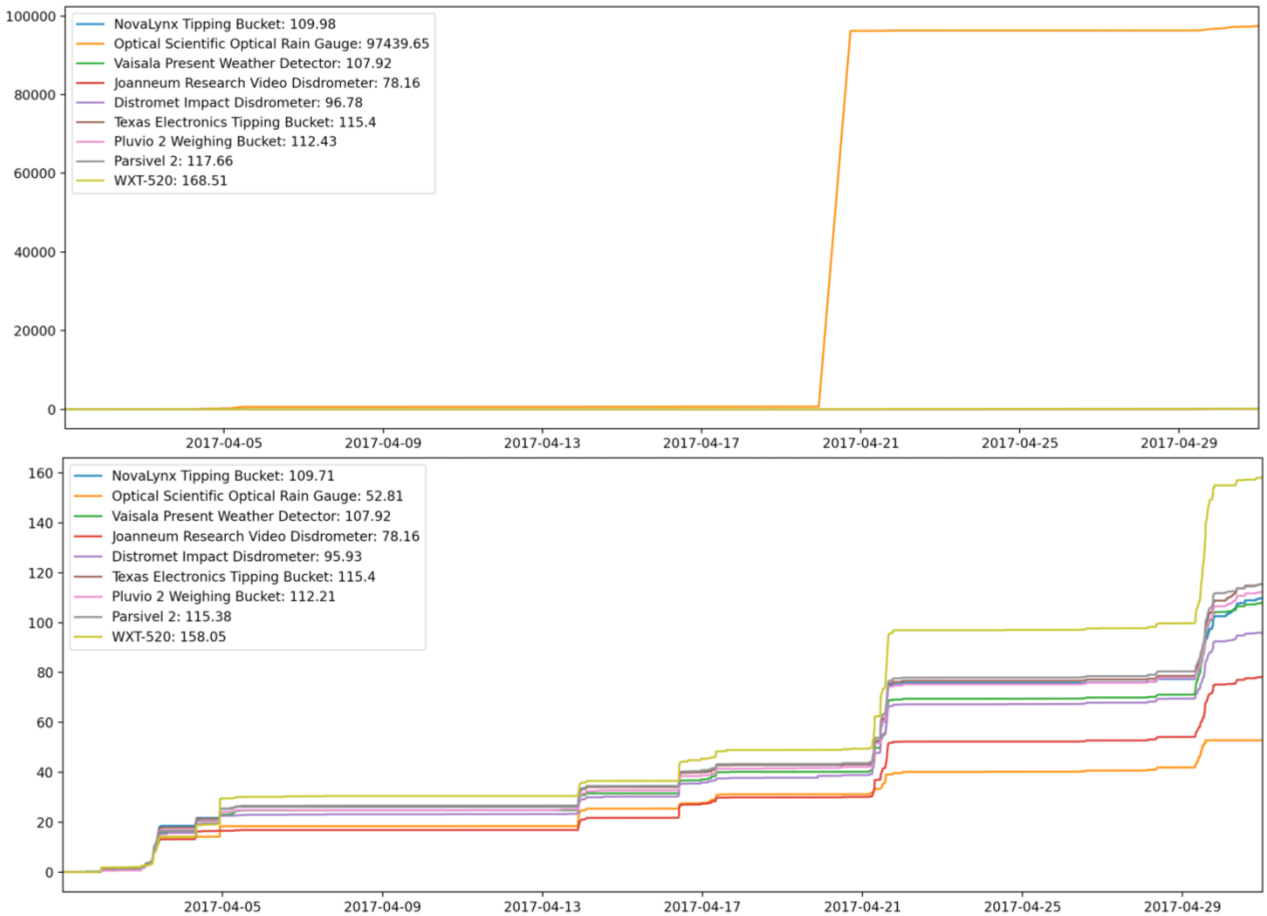

Figure 2. Precipitation accumulations from April 2017 before (top) and after (bottom) the percentile quality control method was applied.

Additionally, a problem with the present weather detector was discovered and reported to the ARM program. Starting in late October/earl-November 2017, the PWD never reported rain rates above 10 mm/hr. Data for the PWD from November 1, 2017 onward were excluded in the analysis.

Clustering

As noted earlier, some of the instruments had duplicates deployed on site. Using multiple instruments of the same instrument could bias the clustering results. Data were averaged for the duplicate systems; Parsivels, WXTs, 2D video disdrometers, and impact disdrometers.

K-means clustering algorithm from the scipy python library was used to calculate clusters of rain rates at each time step using rain rates greater than 0 mm/hr. The number of clusters was varied from 2 to 6 in order to determine the response of the precipitation best estimate with increasing number of clusters. The mean of the cluster with the most members is used as the best estimate. The greater the number of instruments reporting similar rain rates, the higher the confidence is that the cluster is representative of the actual truth. The frequency in which the different instruments are included in the dominant cluster and the corresponding mean rain rate is recorded in a 2-dimensional histogram. The rain rate bins have increasing widths with increasing rates (Table 2).

Table 2. Size of frequency histogram bins across varying rain rates.

Rain Rates (mm/hr)	Bin Width (mm/hr)
0 – 25	0.5
25 – 50	1.0
50 – 75	2.5
75 – 125	5.0

Instrument Weights

The frequency of occurrence was normalized by the total number of samples across instruments for each rain rate. The frequency of occurrence was used as a weight for each instrument and stored as a look up table for varying numbers of clusters in order to determine which clustering weights performed best when applied to the data. Additionally, a generic weighting scheme which applied an 80% weight to the average of the large catchment devices (tipping and weighing buckets) and 20% to the remaining devices was calculated as a comparison.

Results

With an increasing number of clusters, the dominant cluster included less and less members which resulted in a decreasing total accumulation across the 3 years of the analysis (Table 3). The tipping bucket and weighing bucket rain gauges are generally regarded as the gold standards for precipitation measurements and they agree within 4 mm of each other during this entire period which lends credit to that argument. The results from the clustering show that when only 2 clusters are used, the resulting total accumulation is closest to these 2 gauges.

Table 3. Total accumulations for each instrument and the k-means clustering from 2-6 for 2017-2019.

Gauge	2007-2009 Accumulation (mm)
NovaLynx Tipping Bucket Rain Gauge	2506.71
Optical Scientific Optical Rain Gauge	1569.36
Vaisala Present Weather Detector	648.15
Joanneum Research Video Disdrometer	1840.88
Distromet Impact Disdrometer (Average)	1595.00
Texas Electronics Tipping Bucket	2372.47
OTT Hydromet Pluvio 2 Weighing Bucket	2510.89
OTT Hydromet Parsivel 2 Disdrometer	2551.03
Vaisala WXT-520/536	2088.53
K-Means Cluster 2	2518.26
K-Means Cluster 3	2464.0
K-Means Cluster 4	2416.32
K-Means Cluster 5	2328.33
K-Means Cluster 6	2138.67

The resultant frequency plot for using 2 clusters shows some distinct features (Fig. 3). The tipping bucket gauges aren't used at precipitation rates below their minimum detectable amount and the optical gauges are used more often at the lower precipitation rates and the large catchment gauges (tipping and weighing bucket) are used more often at the higher precipitation rates. This aligns with expectations and suggests that the k-means clustering algorithm can be useful for calculating average rain rates across instruments. Investigating the frequency diagram in from Figure 3 in a different manner can provide more insight into each instruments use (Fig. 4).

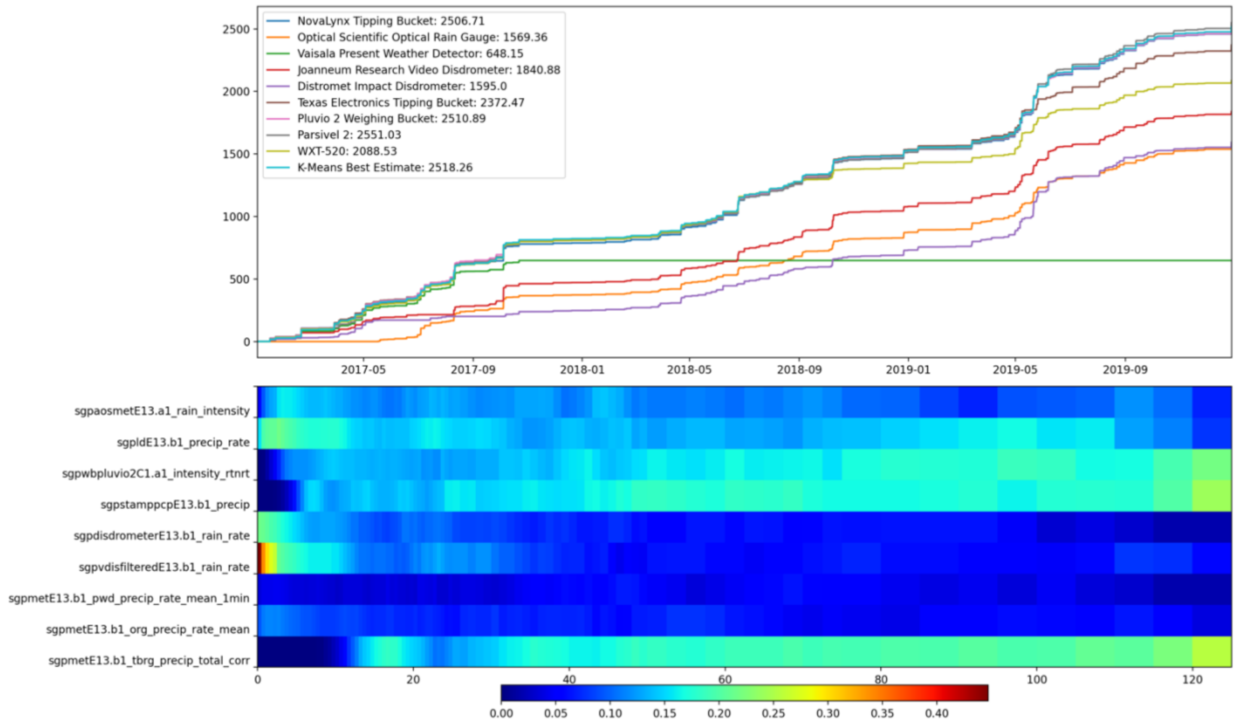

Figure 3. Accumulations for all gauges and the K-means cluster of 2 (top) and the frequency of use in the dominant cluster (bottom).

Figure 4 provides better detail on how each instrument is used across rain rates. The video disdrometers were used very heavily during low precipitation rates and then decreased with heavier rates. This lines up with the known operation of the instrument as it generally failed during larger precipitation events. The most notable features are that of the two tipping bucket rain gauges. They both show a slight bump at the rain rate corresponding to 1 tip. It would be expected that there should be similar features around the other tips as well as the tipping buckets won't measure rain rates that are outside these ranges as accurately. This is exactly what is seen once the number of clusters are increased (Fig. 5). While a cluster of 2 may provide comparable rain rates, the features in the frequency plots do not match the physical operations of the instrument as well which is needed in order to use the frequency information as a weighting scheme.

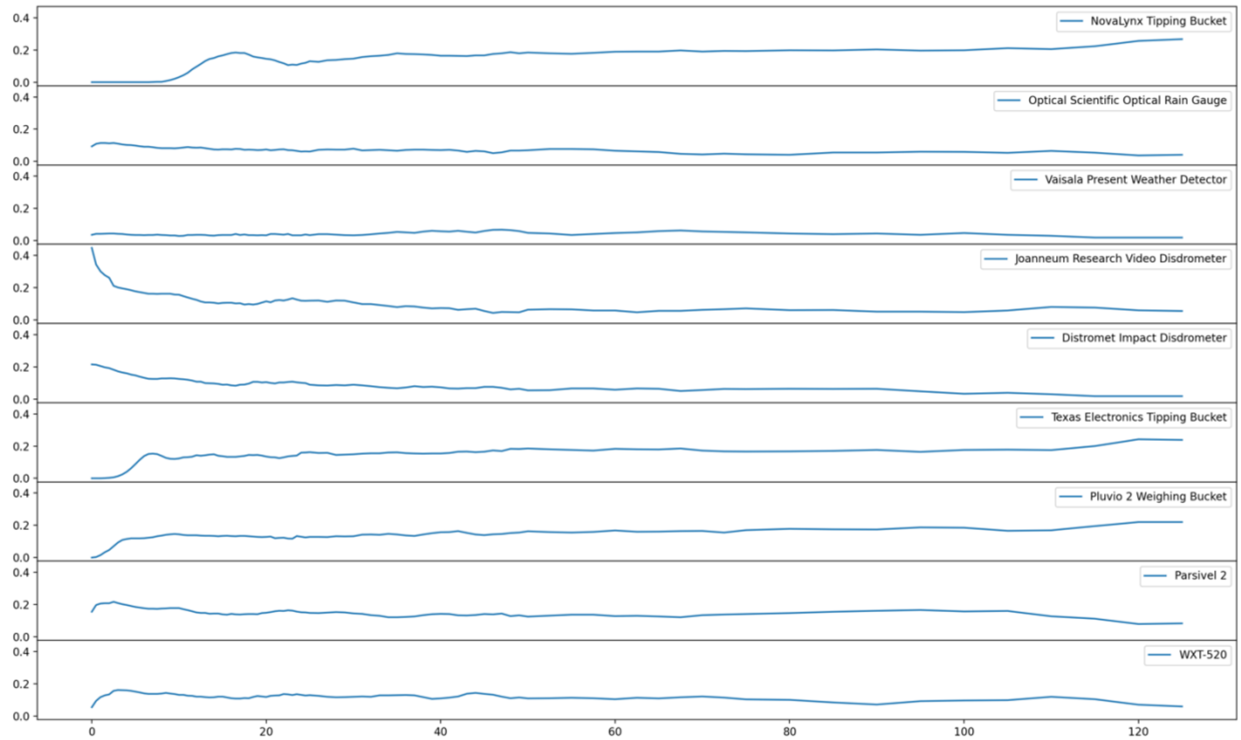

Figure 4. The frequency diagram from Figure 3 as line plots for each instrument. Values are normalized by the sum of samples at each rain rate across all instruments.

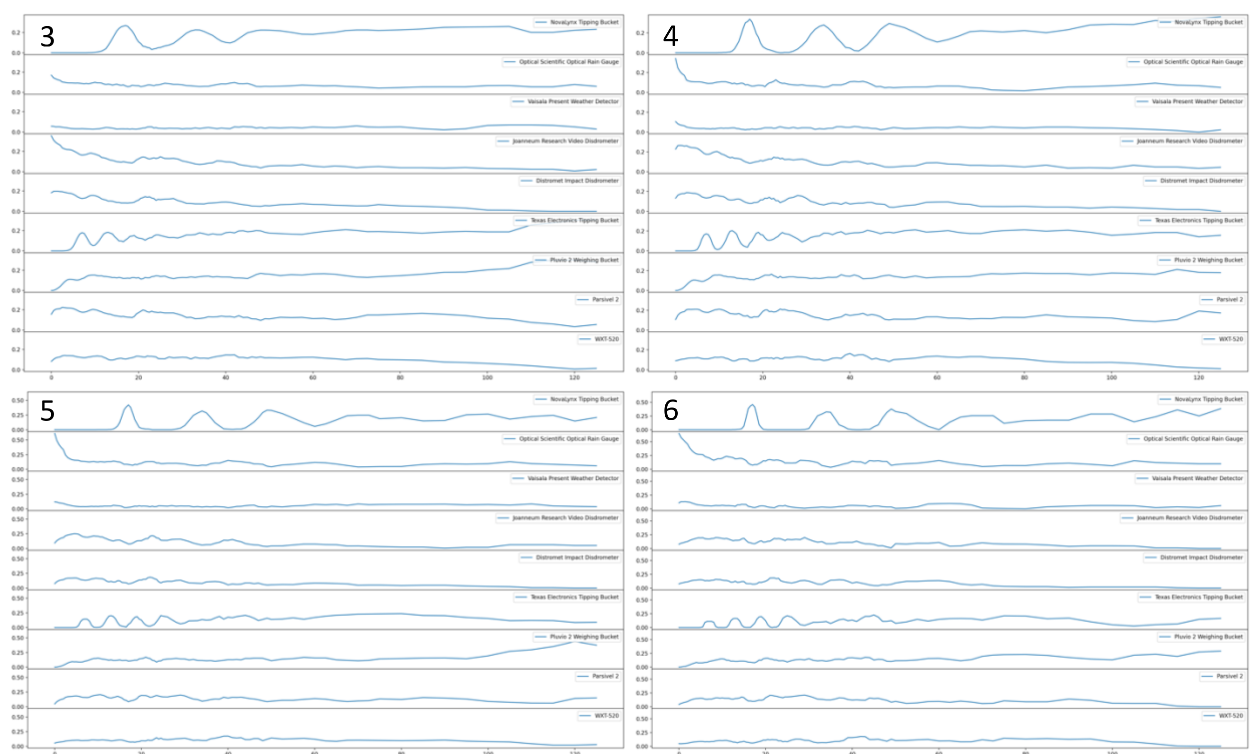

Figure 5. Corresponding frequency plots for clusters of 3 (top left), 4 (top right), 5 (bottom left) and 6 (bottom right).

These normalized frequencies were used as a weighting scheme and applied to all the instruments, including the duplicates. The values were then normalized by the sum of all the weights that were applied. Weights using a cluster of 5 provided the closest comparable accumulation to the k-means clustering results using 2 clusters (2518.26 mm; Table 4).

Table 4. Total accumulation for 2017-2019 using the weights generated from different number of clusters.

Clusters	Total Accumulation (mm)
2	2619.79
3	2603.67
4	2570.99
5	2515.02
6	2462.12

In order to ensure reproducibility of these results, data from SGP for 2020 was processed using the k-means clustering method and the weighting method with the weights derived from the 2017-2019 data. The initial results showed that the k-means clustering algorithm (2 clusters; 530 mm) performed well compared to the weighting method (632 mm) as the weighting method was over-estimating precipitation. Upon further inspection of the data, there were a number of problematic periods of data that did not get screened out with the quality control methods. In order to filter out some of these problematic periods, a threshold of 5 instruments (of the 11 total) were required to be reporting precipitation before it was considered valid. This reduced the weighted precipitation total from a total accumulation of 632 mm to 537 mm. The 80/20 methodology yielded an accumulation of 529 mm which is similar, albeit slightly less, to these other more complicated methods.

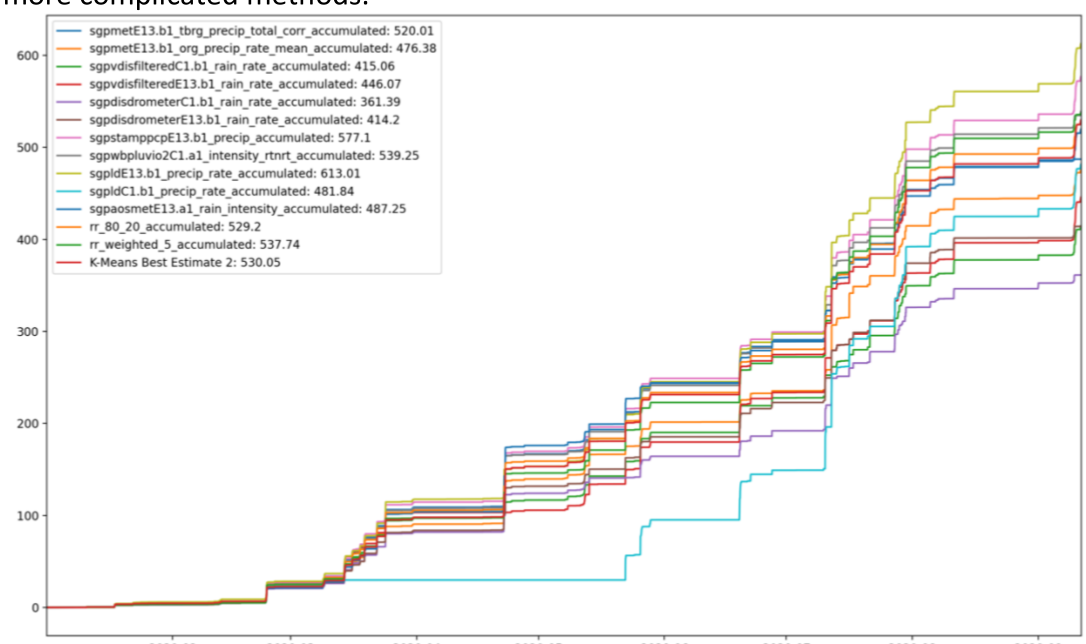

Figure 6. 2020 precipitation accumulation using weights and methods from the 2017-2019 analysis.

Additionally, the weights were used to process data from ARM's Eastern North Atlantic (ENA) site over 2019-2020, which see predominantly stratiform precipitation and drizzle. ENA is not as heavily instrumented with precipitation instrumentation as SGP with 5 instruments being used in the analysis. Like SGP, there were problematic periods of data and a threshold of 2 or more instruments were required to consider the precipitation valid. The present weather detector suffered from the same problem that was discovered during the SGP analysis and was excluded from the analysis. Interestingly, the weighing bucket rain gauge recorded more precipitation than the optical methods that would have been thought to perform better in light precipitation or drizzle (Table 5). Using similar weights as SGP (created using 5 clusters), the total accumulation (1064.44 mm) was similar to the 80/20 calculation (1068.3 mm). The K-means method using a cluster of 2 provided an accumulation of 1008.83 mm. There was a broad variety of accumulations, so more effort is needed to further investigate the ENA data (Fig. 7).

Table 5. ENA precipitation accumulations from 2019-2020

Instrument	Accumulation (mm)
Optical Scientific Optical Rain Gauge	841.01
Joanneum Research Video Disdrometer	1019.58
OTT Hydromet Pluvio 2 Weighing Bucket	1107.55
OTT Hydromet Parsivel 2 Disdrometer	1038.11
Vaisala WXT-520/536	526.87

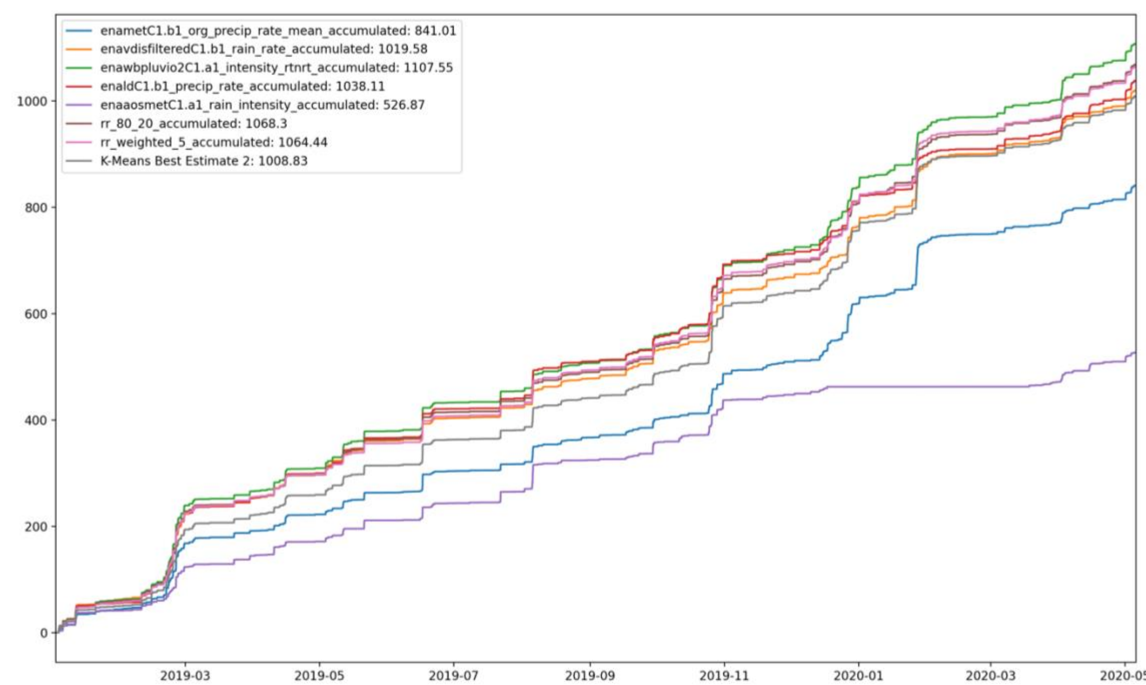

Figure 7. 2019-2020 rain accumulations from ARM's ENA site.

Conclusions

This study looked at simple machine learning methods (clustering) for the calculation of a best estimate precipitation product. Problematic datasets did provide some challenges in using these methods to get a best estimate, but after additional quality control, the clustering techniques did show promising results. The clustering results also lead to the creation of a weighting scheme for each instrument type. The creation of this weighting scheme across rain rates was one of the main goals of this effort and will be available for future use through the online github repository. Additionally, the data will be submitted to ARM as an evaluation PI product. The weights, when applied to SGP data from 2020 did yield results that closely aligned with what is considered one of the gold standards. Applying the same weights and methods to data from the ENA site did not yield as promising of results and more analysis is needed to understand the measurements and the potential data problems.

References

Hardin, J. and N. Guy. (2017, December). PyDSD, doi: http://doi.org/10.5281/zenodo.9991

Theisen, A., Collis, S., Jackson, R., Sherman, Z., Hickmon, N. L., Kehoe, K. E., Godine, C., Sockol, A. J., King, A., Kyrouac, J. K., and Giansiracusa, M. T.: ARM-DOE/ACT: Version 0.4.0 (Version v0.4.0), Zenodo, https://doi.org/10.5281/zenodo.3985272, 2020.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020. a