

Analysis of the DBZH Reflectivity of a Radar and the Precipitation in Bogota Using Functional Statistics and Data Mining Techniques

Ángel Villabona-Ortiz^{a*}, Rodrigo Ortega-Toro^b, Yessica Ramirez-Yara^c

^aChemical Engineering Department, Process Design and Biomass Utilization Research Group (IDAB), Universidad de Cartagena, Cartagena, Colombia.

^bFood Engineering Department, Food Packaging and Shelf-Life research group (FP&SL), Universidad de Cartagena, Cartagena, Colombia.

^cAGH University of Krakow. Faculty of Geo-data Science, Geodesy and Environmental Engineering. Land Subsidence and Hazard Mitigation Research Group. PHD student.

avillabonao@unicartagena.edu.co

This study aims to identify the sweep angle of the dual polarimetric weather radar located in Bogotá, owned by the District Institute for Risk Management and Climate Change (IDIGER), which presents the greatest statistical variability and spatial coverage, in terms of the horizontal component of reflectivity, using functional data analysis, in order to take this reflectivity sweep angle as a reference for correlation analysis between the data recorded by the radar and the accumulated precipitation measured on the surface. For this purpose, a precipitation event that caused emergencies in the city was identified and selected as a case study to identify the degree of correlation between reflectivity and precipitation. Following this, database, spatial and information management analyses were carried out in radial format, in order to abstract from the radar data only the information that is in spatial intersection with the IDIGER monitoring stations, using data mining using Python and RStudio. Finally, correlation analyses were performed to identify patterns, relationships and trends between the data set measured by different sensors using the functional statistics approach.

1. Introduction

Precipitation is water formed in the Earth's atmosphere that then falls on the surface after the air is saturated with moisture, and is measured cumulatively for a given time period and area. Thus, remote sensors, including weather radar, are used to measure and monitor atmospheric water phenomena in the form of rain, hail and snow (Guo et al., 2024). A weather radar is equivalent to the use of hundreds of rain gauges distributed throughout the radar coverage area, which transmit the information in real time; in addition, it performs cloud volume studies, at different slices or sections, as well as tracking and studying severe phenomena such as hurricanes, providing much better coverage in space and also in time (Sokol et al., 2021). Radars provide spatially distributed reflectivity data; however, it is necessary to convert them to rainfall rates. Due to the Spatio-temporal variability of precipitation, as well as the attenuation of the signal through raindrops, it is very difficult to find a suitable relation to transform reflectivity measurements into rainfall rate. The statistics of intermittent rainfall rates are non-Gaussian and show a typical asymmetric distribution that is limited to zero. This restricts the use of well-established stochastic models that assume parametric statistics. Thus, a common solution is to introduce a suitable data transformation to approximate a normal distribution. Currently, techniques are used for the interpretation of data obtained from weather radars, such as Data Mining, which is used in the extraction of patterns from large heterogeneous data sets related to the prediction of meteorological phenomena (Sodunke et al., 2025). For this, pySTEPS assumes a log-normal distribution of rainfall rates by applying the logarithmic transformation. Its advantages are: simplifying the estimation of the distribution parameters, and decomposing the transformed rainfall fields, defines a multiplicative cascade, where multiplications are replaced with additions

in the transformed space. The key idea is to decompose the rainfall field into a multiplicative cascade, where the cascade levels represent different spatial scales and treat them separately in the immediate diffusion model. This study presents functional statistical analyses of the horizontal component to identify the angle of measurement of the IDIGER weather radar, located in Bogotá, which presents the greatest statistical variability and spatial coverage in order to use the radar reflectivity information taken at this angle to perform correlation analysis with the accumulated precipitation measured on the surface, and in this way, lay the foundations for using the values recorded by the weather radar to perform nowcasting and transform the data forecast there into rainfall. One of the purposes of using functional data is to analyse a set of functions instead of discretised data, decreasing the volume of the data considerably in order to discover and evaluate the behaviour and patterns of the curves over time. It is for this reason that the need arises to develop methodological strategies based on statistical analysis for the estimation of the intensity and amount of precipitation using weather radar and ground station measurements together.

2. Materials and Methods

2.1 Study area

For the statistical analysis of the correlation between reflectivity and precipitation, an EEC WeatherTech X-band IDIGER weather radar (Alabama, USA), model DWSR-2001X Dual-Polarization Weather Radar with a coverage radius of 60 km, installed at 2570 m above sea level, and located in the west of Bogotá, at the geographical coordinates $\varphi = 4.67516$ and $\lambda = -74.11397$. Accumulated precipitation information recorded by IDIGER stations is required to perform correlation analyses. Data were taken from 46 stations that record precipitation values in the city of Bogotá by means of climatological (rainfall, temperature and humidity), pluviometric (amount of rainfall in a range of time), hydrological (levels in water bodies) and hydrometeorological (precipitation, temperature, humidity and levels in water bodies) stations. The precipitation data were taken on January 25, 2020, to monitor a precipitation event in the city with a temporality of 4 h and 30 min, taking approximately 6 minutes (35 images), starting at 13:39 and ending at 17:09. Similarly, the information was corroborated with measurements from IDIGER stations in Bogotá, which will present high precipitation measurements.

2.2 Temporal identification for analysis

The precipitation data measured by meteorological radars allow capturing at the same time the spatial information of the rainfall coverage and its intensity, given its resolution, for very short periods of time, thus generating large volumes of information. For this reason, reflectivity and precipitation values are correlated for those moments in time when precipitation occurs. The IDIGER weather radar has a temporal resolution of 6 minutes on average for all angles, thus generating large volumes of information, not only on the horizontal reflectivity of precipitation, but also on all the polarimetric variables available. According to the information observed for Bogotá in the radar images and in the information hosted on the IDIGER servers, related to precipitation for the months of December 2019, January and February 2020, it was found that on January 25, 2020 there was a significant precipitation event in the city with values recorded by all stations for the entire day of approximately 1000 mm, for the period between 13:00 and 17:00 h. **Figure 1** shows the radar image for 2:41 p.m. showing precipitation events in the eastern fringe of the city with average reflectivity values of 26 dBZ, as well as some rainfall in the municipalities surrounding the city. The time zone used in this study is -0500 UTC. The radar data used had an analysis time window of 3 hours and 36 minutes, updated approximately every 6 minutes (36 radar images for each shooting angle stored in CfRadial format), starting at 18:39 h and ending at 22:15 h. Similarly, data measured by IDIGER stations in the same time range were taken, finding high precipitation measurements.

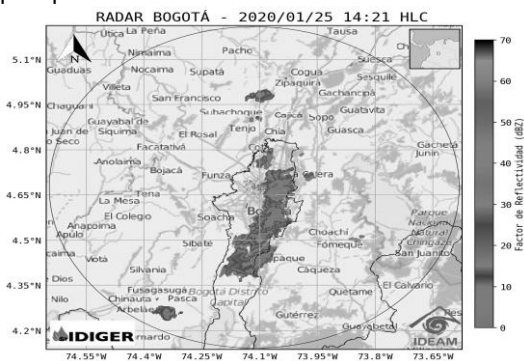


Figure 1. Precipitation event for the city of Bogotá on 25 January 2019. Image obtained from IDEAM servers

2.3 Correlation analysis between estimated reflectivity and precipitation measured by stations

The first step is to identify how many significant variations are present in the recorded reflectivity and precipitation data in the time window of the analysis (knots), in order to establish groups that may be equally spaced to determine significant changes in the behaviour of the variables. Following the methodology of Aguilera & Aguilera-Morillo, (2013), code was implemented in RStudio to identify which of the function bases generates the lowest mean square error for each radar sweep angle and precipitation, considering the roughness coefficient such that it minimises the generalised cross-validation. To construct the functional objects, it was identified that the reflectivity and precipitation values present different behaviours approximately every 12 minutes, so 18 knots were established for these variables. There are different methods for smoothing the functional data. Among the best known are the function bases, which are a set of known functions Φ_k independent of each other and whose linear combination can approximate any function. If the discrete data to be modelled is periodic, Ramsay & Silverman, (2005), recommend using Fourier function bases; otherwise, BSplines function bases should be used. Spline functions are piecewise polynomial functions on which constraints are imposed at the joint points. Because it is important to choose the roughness level of the curves, generalised cross-validation is used as a tool to obtain the roughness penalty coefficient. The information taken for the statistical analyses corresponds to data recorded by the stations on a minute basis, so it was necessary to accumulate the precipitation values as a function of the radar sweep hours (approximately every 6 minutes for all angles), procedures carried out in Python. For the hours of 19:34, 19:40, 19:46, 19:52, 19:58, 20:05, 20:11 and 20:17; accumulated precipitation values greater than zero were recorded for more than 20 IDIGER stations, indicating that, for those precipitation events, both radars and stations were constantly recording the event that caused emergencies in the city.

3. Results and Discussion

3.1 Horizontal analysis

The IDIGER radar has five angles of inclination to make sweeps over the city (table 1). Due to anthropic intervention, there are obstacles in the sweeps made by the radar, such as bridges, buildings, among others; therefore, the ideal angle for reflectivity analysis will be the one where the greatest amount of hydrometeors of interest in the atmosphere can be measured, thus representing an optimal spatial coverage in the city in terms of the stations located on the surface. Table 1 shows the basic descriptive statistics of the horizontal reflectivity gate at the angle of 4.35°, where there was the intersection with the monitoring stations every 6 min, analysed by hours, obtaining that for the border hours, the standard deviation of the horizontal reflectivity is slightly wider, due to the fact that for those hours, the precipitation cloud has been dissolving; while the hours where the greatest magnitude in the reflectivity values is presented, are those that generated the strongest and most constant precipitation, due to the displacement presented by the precipitation cloud as the storm evolves. As for the maximum value of accumulated reflectivity for the hours of analysis, the highest value was at 19:28, while the hour with the lowest accumulated values was at 22:02. The above analyses were carried out only in those gates where there is spatial overlap with the IDIGER stations.

Table 1. Descriptive statistics of horizontal reflectivity

	0.5°	2.5°	3.0°	4.35°
Spatial representativeness	0	14	0	20
σ DBZH	14.750	14.333	13.306	14.645
DBZH Media	12.795	12.743	14.780	12.040
% Amount of data	18.59%	38.52%	33.09%	40.27%

Table 1 shows the overall descriptive statistics for the radar scanning angles. There it is observed in the line of spatial representativeness, the number of times in which, for the scan angle, the highest number of stations recording precipitation data given a spatial intersection was obtained (the sum does not correspond to 36 since, for the times of 18:44, 18:51, 19:34, 20:35, 20:54, 21:06; for some scan angles the same number of stations was counted), being the angle 4.35° was the one where the radar images showed the greatest special coverage, followed by the 2.5° and 4.7° angles. The 0.5° and 3.0° scan angles did not show, for the hours of analysis, greater spatial coverage than the other angles. The mean and standard deviation of the reflectivity values are similar for the measurement angles, with the 3.0° angle having the lowest standard deviation and the highest mean, while the highest standard deviation is for the 0.5° angle. The highest reflectivity values stored in the radar images are for the scanning angle of 4.7°. The angle 4.35° is the one with the greatest amount of information (40.27%) in relation to the spatial intersection between the radar images and the 46 IDEGER stations in the time frame of the analysis carried out.

3.2 Smoothing of functional data.

To find the best configuration for the creation of the functional objects, curves were estimated using Fourier and BSplines function bases with and without roughness coefficient in order to find the best idealisation that generates the lowest mean square error considering the roughness coefficient that minimises the generalised cross-validation. From Table 2, it is found that, for all sweep angles and for the accumulated precipitation, the best parameterisation for the creation of the functional objects correspond to the use of BSplines function bases with roughness coefficient employing 18 knots, producing the lowest mean square errors MSE. This indicates that the mean horizontal reflectivity for all angles and the accumulated precipitation per unit of time does not exhibit sinusoidal behaviour, which is why, when the exercise is carried out using Fourier function bases, they show MSEs with higher values than the MSEs calculated with BSplines. Therefore, the functional data for the angle reflectivity and cumulative precipitation means will be constructed following BSplines function bases with roughness.

Table 2. Root mean square error of the mean reflectivity curves by angle

Angle	0.5	2.5	3	4.35	4.7	Precipitation
Bsplines	13.032	24.144	8.199	4.453	13.032	0.0060
Bsplines λ	2.908	2.121	2.324	1.725	2.909	0.0015
Fourier	12.945	14.744	6.197	4.870	12.946	0.0079
Fourier λ	9.283	9.063	3.948	2.598	9.284	0.0017

3.3 Functional descriptive analysis of reflectivity and precipitation

Once the parameters for the construction of the functional curves were determined, average curves were obtained for each reflectivity angle and the accumulated precipitation in the Hilbert-Schmidt space constructed through the combination of BSplines, reducing the dimensionality of the information, thus allowing to appreciate the real behavior of both reflectivity and accumulated precipitation. The sweep angle of 0.5° is the angle that records the least amount of information according to the spatial location of the IDIGER stations, with a total value of 18.5%, where it was also observed that the available information is not totally constant over time. The mean reflectivity curve shows the highest amount of trend, showing a constant behavior around 12.795 DBZH with a σ of 14.750. The curve for the 2.5° sweep angle showed a higher amount of precipitation information compared to the 0.5° functional curve, with a data percentage of 38.5% and a mean curve with higher variability. For the 2.5° angle it shows a constant overall behavior with a mean of 12.743 DBZH and a σ of 14.333. Regarding the mean reflectivity curve for the 3.0° angle, a smoothed behavior was obtained compared to the 2.5° angle due to the fact that for the 3.0° angle there is less information: 33.09% representativeness, with a mean of 14.780 DBZH and a σ of 13.306. For the 4. 3° shows an increasing behavior during the first minutes of the storm until hour 18:44 and then decreases until hour 18:57; after that, it shows an increasing behavior until hour 19:21 and then decreases until hour 20:23; After that, it shows constant and decreases from hour 21:43 until the end of the storm. Finally, for the radar the 4.7° scan angle shows a decreasing behavior at the beginning of the storm until 18:39. From Figure 2 of the reflectivity averages for all angles, it can be observed that, as a whole, the functional mean curves show the same trend, with the highest reflectivity values recorded on average between 19:22 and 20:24, the time window where the storm was most intense in the city, producing emergencies. The values at the beginning of the storm vary according to the angle of measurement, with the angles of 2.5° and 4.35° being those where the storm starts from less to more intense. At the end of the storm, the 4.7° angle shows a marked decreasing behavior, with negative values further away from zero.

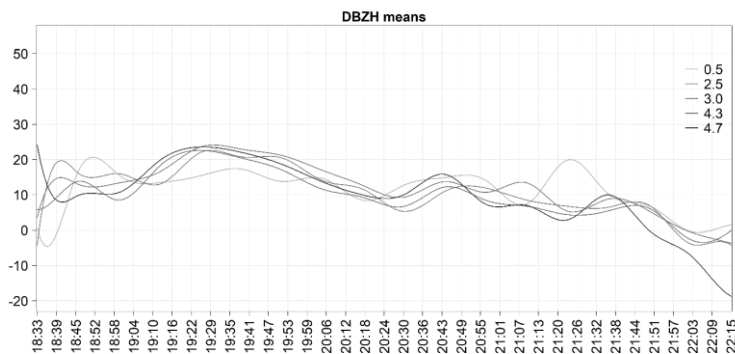


Figure 2. Reflectivity functional means curves for all IDIGER radar angles

In relation to the values of the accumulated precipitation, every 6 minutes measured on the surface by the IDIGER stations (Figure 3), the curves of the values measured by the stations are shown in grey, and the average functional curve of the accumulated precipitation is shown in black. There the behaviour is similar to the reflectivity curves since this curve shows a constant behaviour equal to zero from the beginning of the storm until 19:03, showing an increase in the accumulated values until 19:09 and then decreasing until 19:21; from there, the values increase until 19:40, and then slowly decay until the end of the storm.

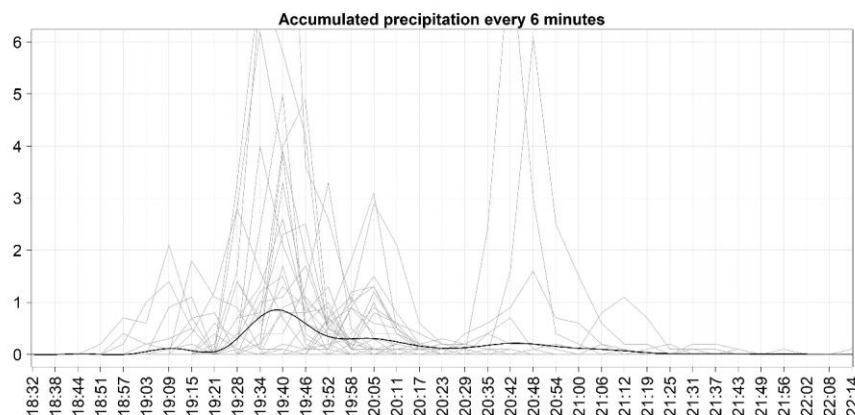


Figure 3. Precipitation functional mean curve

3.4 Functional correlation

In order to carry out the correlation analysis the statistical research carried out by Sepúlveda Berrío, (2016), was taken as a reference, where he states that, for the estimation of the coefficients of the Marshall equation, information on the configuration of the radar and the climatological conditions of the area must be taken into account. The author estimates coefficients to the ratio $Z_e = aRR^b$, which work only under the parameterisation of the radar of the Metropolitan Area of the Aburrá Valley. It is not the objective of this research to estimate the coefficients of the Marshall equation (the reader is informed about the basis and implications of the estimation of the equation's parameters) but rather to find statistical relationships to define which sweep angle has the greatest spatial and statistical coverage between reflectivity and accumulated precipitation. In order to perform the functional correlation analyses, a sample of stations with the greatest spatial representativeness with the radar images was taken for the hours in which they recorded accumulated precipitation values, thus forming pairs of data, which are not constant over time, and then, by means of the FDA eval.fd library function, obtaining the results summarized in Table 3.

The value that accompanies the name of the station corresponds to the number of records greater than zero of the accumulated precipitation of the total of the temporality of analysis. In order to select the angle with the best correlation between the reflectivity and precipitation variables, Pearson's correlation will be taken into account, plus the number of available pairs between reflectivity and precipitation and the greater amount of radar data, remembering that the greater the amount of data in the sample, the closer it will be to the real behaviour of the variable. Regarding the IDIGER station, the highest correlation value in absolute value is presented with the measurement angle of 4.7° . For this angle, the available pair information between reflectivity and accumulated precipitation is 5. For the angle of 2.5° , a Pearson correlation is visualized with a value of 0.422, followed by the angle of 3.0° with a value of 0.333, followed by the angle of 0.5° with 0.247 and finally, the angle of 4.35° with a correlation of 0.086 presents the lowest value, indicating that the values of reflectivity and accumulated precipitation do not present statistical behaviours in common. For the Juan Rey station, the highest correlation value corresponds to the angle of 4.7° , followed by the angles 3.0° , 4.35° , 3.0° and 0.5° , where the angles with the highest number of available pairs with precipitation were for the angles of 2.5° , 3.0° and 4.35° . The stations of La Fiscala and Vitelma present different behaviours to those mentioned above, where the highest value of correlation is given at the angle of 0.5° , followed by the angle of 4.35° , 4.7° , 2.5° and 3.0° .

Table 3. Correlation matrix between reflectivity and cumulative precipitation

	IDIGER (12)	Juan Rey (13)	La Fiscala (9)	Vitelma (7)
DBZH05	-0.247	0.125	0.553	0.724
Radar data	13	4	12	8
Pairs with precipitation	8	3	8	4
DBZH25	0.422	0.317	0.109	-0.395
Radar data	25	18	17	17
Pairs with precipitation	10	11	9	7
DBZH3	0.333	0.665	-0.006	-0.107
Radar data	8	17	15	16
Pairs with precipitation	6	11	9	7
DBZH435	0.086	0.595	-0.326	0.364
Radar data	24	19	18	17
Pairs with precipitation	10	11	9	7
DBZH47	-0.628	0.698	-0.240	0.392
Radar data	6	17	16	17
Pairs with precipitation	5	10	9	7

4. Conclusions

By smoothing the discretised data through the expansion of function bases through the analysis of functional data, it is possible to consolidate the information closest to the functional mean, removing outliers that may alter the analyses and results. The expansion by B-splines function bases smoothed the totality of treated curves, being the determinant of selection, the basis with the smallest quadratic error, using as validator the generalised cross-validation, which computationally is the best method to minimise the GVC parameter. In order to carry out statistical correlation analyses of the reflectivity measured by the radar and the accumulated precipitation recorded by the surface monitoring stations; it is advisable to use the data measured by the radar angle of 4.35° since it is at this inclination where it has the greatest spatial coverage in relation to the fixed location of the IDIGER stations. Precipitation data measured by the stations localise precipitation in a timely manner, so it would be optimal to use additional sensors to those established by IDIGER to perform statistical validations and comparisons of precipitation and reflectivity.

Nomenclature

DBZH – Radar horizontal reflectivity factor

IDIGER – District Institute for Risk Management and Climate Change (in Spanish)

σ – Standard deviation of reflectivity, dBZ

GVC – Gradient of Vertical Reflectivity Change

References

- Aguilera, A. M., & Aguilera-Morillo, M. C. (2013). Comparative study of different B-spline approaches for functional data. *Mathematical and Computer Modelling*, 58(7–8), 1568–1579. <https://doi.org/10.1016/J.MCM.2013.04.007>
- Guo, R., Fan, X., Zhou, H., & Liu, Y. (2024). Multi-Sensor Precipitation Estimation from Space: Data Sources, Methods and Validation. *Remote Sensing* 2024, Vol. 16, Page 4753, 16(24), 4753. <https://doi.org/10.3390/RS16244753>
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed.). Springer New York. <https://doi.org/10.1007/B98888>
- Sepúlveda Berrío, J. (2016). *Estimación cuantitativa de precipitación a partir de la información de Radar Meteorológico del Área Metropolitana del Valle de Aburrá* [Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/58046>
- Sodunke, M. A., Ojo, J. S., Lawal, Y. B., Ojo, O. L., Owolabi, G. A., & Olateju, A. I. (2025). Application of machine learning models for rainfall prediction and estimation of rain-induced attenuation for satellite communication in a tropical region. *Journal of Atmospheric and Solar-Terrestrial Physics*, 268, 106443. <https://doi.org/10.1016/J.JASTP.2025.106443>
- Sokol, Z., Szturc, J., Orellana-Alvear, J., Popová, J., Jurczyk, A., & Céleri, R. (2021). The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review. *Remote Sensing*, 13(3), 351. <https://doi.org/10.3390/RS13030351>