BI-DIMENSIONAL VECTOR DATA ANALYSIS OF POSITIONAL ACCURACY OF LANDSAT-8 IMAGE WITH PYCIRCULARSTATS

A. Cuartero[1], Mercedes E. Paoletti[2], Senior Member, IEEE, A. Rey Presas[3], Juan M. Haut[4] Senior Member, IEEE,

[1] Department of Graphic Expression, University of Extremadura, Cáceres, Spain.
[2] Department of Computer Architecture, Complutense University, Madrid, Spain.
[3] Computer Graphics and Data Engineering Group (COGRADE), Centro Singular de Investigación en Tecnoloxías Intelixentes (CITIUS), University of Santiago de Compostela, Spain.
[4] Department of Technology of Computers and Communications, University of Extremadura, Spain.

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883588

ABSTRACT

Analyzing directional data, in particular circular data, requires methods that are being available in libraries with a well-known prestige as Python including SciPy, NumPy or SciKit-Learn libs. An open-source library has been implemented to be executed by the Python interpreter, called PyCircularStats. Source code:
https://github.com/mhaut/pycircularstats
The potential of PyCircularStats is shown with an example of analyzing two-dimensional data using circular statistics. The practical case chosen is the positional accuracy analysis of a satellite image of LandSat-8 in Cáceres, Spain, with 99 control points taken with GNSS systems.
In this work, the possibilities of two-dimensional data analysis using circular statistics using the PyCircularStats tool with the results of this case of use is presented.

Index Terms— *Analysis data, circular graphical statistics, geospatial big data, remote sensing*

1. INTRODUCTION

Directional data is characterized by being defined on a circular or spherical domain, i.e. data belongs to angular measurements, made up of an orientation or angle. Circular data [1] is measured in angles or directions on the unit circle and refers to the orientation in the plane, following a Von Mises distribution [2] and being the simplest kind of directional data.

There are many examples of circular data in field of Earth Sciences [3] as varied as Geology [4], Biology [5], Oceanographic [6], or Meteorology [7] areas. Applications such as the measurement of winds [8], [9], migratory movements or the cations such as the measurement of winds, migratory movements or the analysis of striae orientation on fault planes, although directional data can be found in others fields too, as image processing [10], [11], Physics [12], Medicine [13], [14] or even in machine learning methods [15].

A particular case to analyze the positional accuracy of satellite image of LandSat-8 in Caceres, Spain, is showed with this graphical circular statistic. Additionally, this paper performs a comparative between PyCircularStats and VecStatGraphs2D, other vector analysis using graphical and analytical methods developed in R.

This work is organized as follows: In section 1 a brief introduction of circular data and its applications in multiple fields has been made. Section 2 presents an analysis tool: PycircularStat. Section 3 shows an example of the results of a case of use showing a comparison of the results with another analysis tool: VecStarTGraph2D. Finally, in Section 4 we present a summary of the work and conclusions.

2. METHODOLOGY

PyCircularStats is an open-source library, based on the VecStatGraphs2D package, but which it has been developed in Python and integrates the same methods and operations of circular statistics, and, with and an extended graphical analysis allowing a better a more flexible use. Also, to introduce the operations of this package and demonstrate its great potential within the field of statistics, a graphical interface has been developed to facilitate its use.

Evaluation and implementation of PyCircularStats algorithm were performed by using python programming language, more concretely, python 3.6.9. and the libraries: numpy, scipy, matplotlib, scikit-learn, pyqt. All scripts, documentation and non-sensitive data are available on the GitHub: https://github.com/mhaut/pycircularstats

In addition, a graphical interface has been developed to facilitate analysis of vector data. This interface called Circular Statistics Studio (see Figure 1).

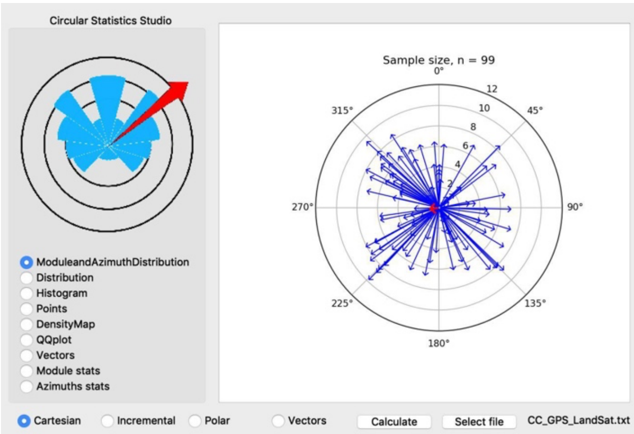

Figure 1. Interface Circular Statistics Studio of the PyCircularStats.

3. PRACTICAL APPLICATIONS. CASE STUDY

A practical example of the usefulness of graphical circular statistical analysis is presented. The evaluation of geometric accuracy of high-resolution satellite images (HRSIs) has been increasingly analyzed in recent years, therefore the practical application presented has been about this topic. The spatial data used in this practical case consists of both, one Landsat-8 satellite image whose planimetric positional accuracy is to be analyzed and 99 independent ground control points (IGCP) will be obtained to evaluate the positional quality of the original image. The coordinate differences of both data sets allow the analysis of planimetric positional accuracy using circular statistics.

3.1. Spatial imagen analyzed

The spatial images analyzed is Landsat-8, its spatial resolution is 15 meters. Each scene is composed of 11 GeoTIFF files one per spectral band, a metadata file (MTL) and one additional file with scene quality assessment (QA) information. The analyzed images belong to band 8, the panchromatic band, since it has better spatial resolution than the rest of bands. The xy-coordinates pair for each pixel is guaranteed to be located inside the pixel limits. Each image corresponds to a surface of approximately 180 km from North to South and 190 km from East to West, and it has been obtained from an orbital flight of height 705 km. The image used in the experiment dates from 09/15/2018 (see Figure 2).

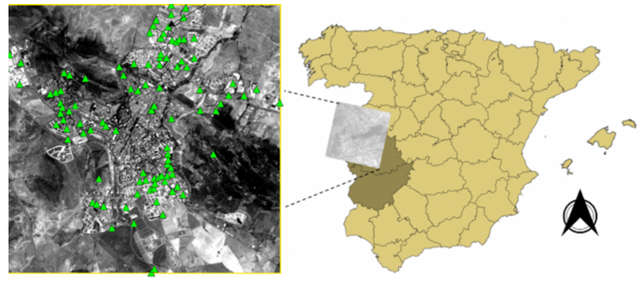

Figure 2. Study area: The geographical location of the Landsat 8 spatial image, in Cáceres, Spain (right image) and distribution of all IGCPs green triangles (left image).

3.2. Positional control data used

The Independent Ground Control Point (IGCP) used for analyses last Landsat spatial imaged, have been obtained using DGPS, so has centimetric accuracy. The set of IGCPs available in this study belongs to the Geodesic Polygonal Network of Cáceres and it consists of 99 points distributed through the urbanizing area of the municipality of Cáceres, Spain (see Figure 2).

3.3. Methods and results

In this case study, analyzes were performed with both libraries, PyCircularStats and VecStatGraphs2D, in order to compare linear and circular statistical results. Table 1 shows results obtained for linear and circular statistics and uniformity test in both libraries. Additionally, Figure 3 shows a map made with PyCircularStats where each vector is shown in its correct spatial position of the image. In other words, this map shows the vectors without concentrating on the center of a circle, as has been done in the graphs of Figure 4 as circular statistical analysis.

In Figure 4 different graphics of circulars statistics in PyCircularStats are showed: draw module and azimuth distribution, draws of histogram, distribution and points, and

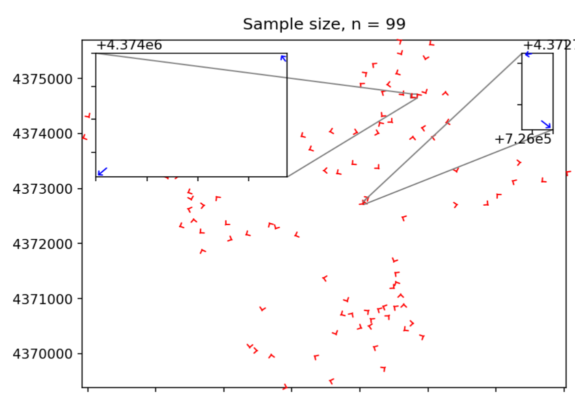

Figure 3. Graphical Maps of 99 vectors in LandSat-8 of Cáceres, Spain, generated by PyCircularStat.

others graphical information as density map and Q-Q plot of azimuth.

Table 1. Results of linear and circular statistics of accuracy positional LandSat-8 image.

	PyCircularStats

Linear Statistics-Modules (values in Meters)	`NUMBER OF ELEMENTS = 99`
	`MIN VALUE = 0.2062`
	`MAX VALUE = 10.0419`
	`RANGE = 9.8358`
	`ARITHMETIC MEAN = 5.5165`
	`MEAN STANDARD ERROR = 0.2354`
	`STANDARD DEVIATION = 2.3418`
	`VARIANCE = 5.4839`
	`POPULATION STANDARD DEVIATION = 2.3299`
	`POPULATION VARIANCE = 5.4285`
	`SKEWNESS COEFFICIENT = -0.3259`
	`KURTOSIS COEFFICIENT = -0.7888`

Circular Statistics-Azimuts (values in Degrees)	`NUMBER OF ELEMENTS = 99`
	`MEAN AZIMUTH = 262.5862`
	`MEAN MODULE = 0.1794`
	`CIRCULAR STANDARD DEVIATION = 1.8536`
	`CIRCULAR VARIANCE = 0.8206`
	`CIRCULAR DISPERSAL = 14.0481`
	`VON MISES PARAMETER = 0.3648`
	`SKEWNESS COEFFICIENT = 0.1173`
	`KURTOSIS COEFFICIENT = 0.0558`

Uniformity Test	Rao Test. the hypothesis of uniformity is accepted for P = 0.01 Rayleigh Test: P-value for the hypothesis of uniformity = 0.041
	VecStatGraph2D

Linear Statistics-Modules (values in Meters)	`NUMBER OF ELEMENTS =  99"`
	`MIN VALUE =  0.2062"`
	`MAX VALUE =  10.0419"`
	`RANGE =  9.8358"`
	`ARITHMETIC MEAN =  5.5165"`
	`MEAN STANDARD ERROR =  0.2354"`
	`STANDARD DEVIATION =  2.3418"`
	`VARIANCE =  5.4839"`
	`POPULATION STANDARD DEVIATION =  2.3299"`
	`POPULATION VARIANCE =  5.4285"`
	`SKEWNESS COEFFICIENT =  -0.3259"`
	`KURTOSIS COEFFICIENT =  -0.7888"`

Circular Statistics-Azimuts (values in Degrees)	`NUMBER OF ELEMENTS = 99"`
	`MEAN AZIMUTH = 262.5862"`
	`MEAN MODULE = 0.1794"`
	`CIRCULAR STANDARD DEVIATION = 1.8536"`
	`CIRCULAR VARIANCE = 0.8206"`
	`CIRCULAR DISPERSAL = 14.0481"`
	`VON MISES PARAMETER = 0.3648"`
	`SKEWNESS COEFFICIENT = 0.1173"`
	`KURTOSIS COEFFICIENT = 0.0558"`

Uniformity Test	Rao Test: the hypothesis of uniformity is accepted for P = 0.01" "Rayleigh Test: P-value for the hypothesis of uniformity = 0.041"

Table 1, Figure 3 and Figure 4 summarize linear and circular analysis of the results about de positional accuracy of this Landsat 8 image. In Table 1 we can see that the results obtained with both tools are the same. Therefore, it can be considered that the tools work correctly.
On the other hand, the graphical analysis includes a set of graphics that it allows us to assess the results from a graphic perspective. Summarized to this analysis it can be concluded that the positional accuracy of the image analyzed is very good for several reasons: 1) a good linear

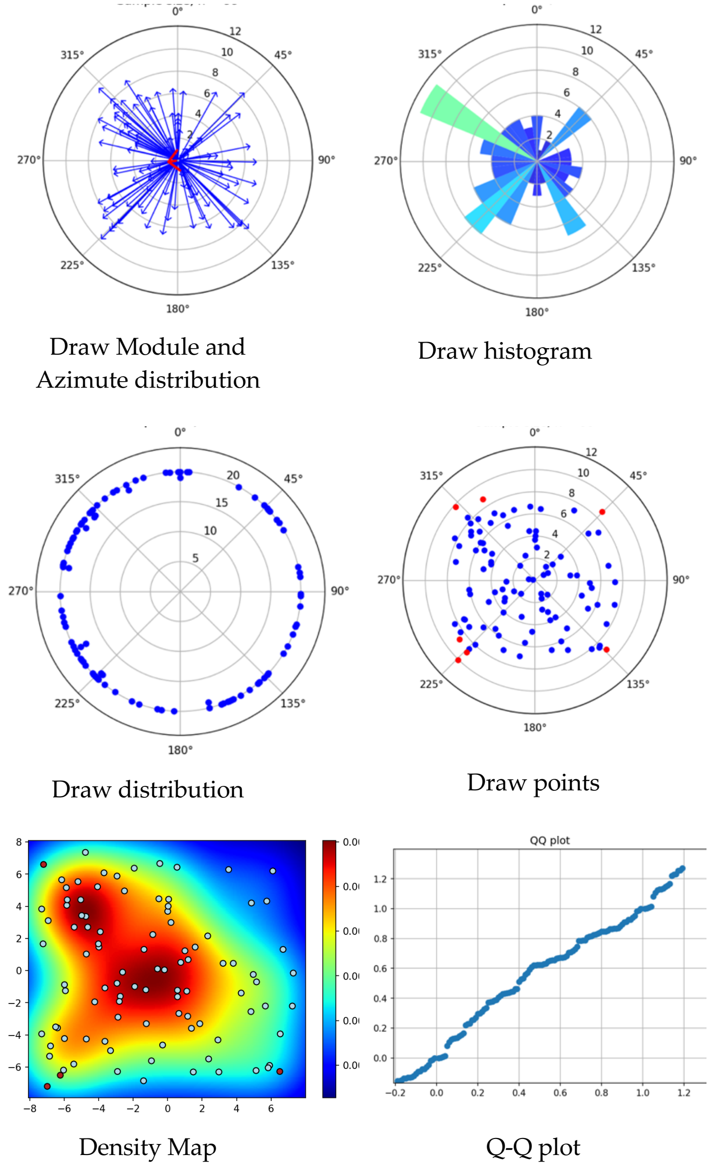

Figure 4. Graphics of circulars statistics for a positional accuracy analysis of one LandSat-8 of Cáceres, Spain.

value of the positional accuracy of LandSa-8 image is 5.5 m ± 2.3 m and the maximum value (10 m) does not exceed the spatial resolution of the image (15 m); 2) All graphics show a good distribution data and 3) uniformity tests were accepted.

4. CONCLUSIONS

This article presents an example of analyzing circular data using PyCircularStats, as an alternative to other existing ones such as VecStatGraph2D. Results of analysis of the example shown that both tools are the same statistic.

This PyCircularStats tool includes linear analysis as well as descriptive angular statistics, it also includes graphic analysis and density maps. These graphics are easily editable since it has been complemented with a graphic interface.

In this work, a practical example of use has been demonstrated when analyzing the planimetric positional precision of a spatial image, LandSat-8 from Cáceres, Spain, and with 99 ground control points. The positional accuracy of LandSa-8 image is 5.5 m ± 2.3 m. This result is similar to that obtained in [16] where this same image is analyzed using only conventional statistics (5.22 ± 1.95 m).

This work shows the possibility of performing more complete two-dimensional data analysis using circular statistics and knowing the tools that develop them.

ACKNOWLEDGMENT

The authors wish to acknowledge the funding received from the Consejería de Economía, Ciencia y Agenda Digital of the Junta de Extremadura, the European Regional Development Fund (ERDF) of the European Union through grant reference GR21040 and the Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, Ministerio de Economía y Competitividad (PID2019-105221RB-C42) AEI /10.13039/501100011033. We also wish to thank Cáceres city Council for transfer of the IGCPs data for putting available so many free data of great positional accuracy.

11. REFERENCES

[1] N. I. Fisher, *Statistical analysis of circular data*, Cambridge. 1993.

[2] M. Gordon, L.; Hudson, *A Characterization of the Von Mises Distribution. The Annals of Statistics.* 1977.

[3] G. J. Borradaile, *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation*, Springer-V. Berlin Heidelberg, 2003.

[4] J. C. Davis, *Statistics and Data Analysis in Geology*, 2nd ed. New York, NY, USA, 1990.

[5] E. Batschelet, *Circular Statistics in Biology*. Academic Press, 1981.

[6] P. G. Rodríguez, M. E. Polo, A. Cuartero, Á. M. Felícisimo, and J. C. Ruiz-Cuetos, "VecStatGraphs2D, A tool for the analysis of two-dimensional vector data: An example using QuikSCAT ocean winds," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 921–925, 2014.

[7] Y. Chikuse, *Statistics on Special Manifolds*. New York, NY, USA: Springer, 2003.

[8] A. M. Herráez, D.P.; Sevilla, M.I.A.; Canals, L.F.; Barbero, J.M.C.; Rodríguez, "A GIS-based fire spreadsimulator integrating a simplified physical wildland firemodel and a wind fieldmodel.," *Int. J. Geogr. Inf. Sci.*, vol. 31, pp. 2142–2163, 2017.

[9] B. Yusof, N.; Zurita-Milla, R.; Kraak, M.J.; Retsios, "Interactive discovery of sequential patterns in time series of wind data," *Int. J. Geogr. Inf. Sci.*, vol. 30, pp. 1486–1506, 2016.

[10] A. Hanbury, "Circular statistics applied to colour images," in *8th Computer Vision Winter Workshop*, 2003, vol. 91, no. 1, pp. 53–71.

[11] N. Nikolaidis and I. Pitas, "Nonlinear processing and analysis of angular signals," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3181–3194, 1998.

[12] E. van Doorn, D. B., Sreenivasan K.R., and M. Casella, "Statistics of wind direction and its increments," *Phys. Fluids*, vol. 12, pp. 1529–1534, 2000.

[13] N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience: the dangers of double dipping," *Nat. Neurosci.*, pp. 535 – 540, 2009.

[14] C. T. Le, P. Liu, B. R. Lindgren, K. A. Daly, and G. Scott Giebink, "Some statistical methods for investigating the date of birth as a disease indicator," *Stat. Med.*, vol. 22, pp. 2127–2135, 2003.

[15] Suvrit Sra, "Directional Statistics in Machine Learning: a Brief Review," 2016.

[16] M. Sánchez, A. Cuartero, M. Barrena, and A. Plaza, "A new method for positional accuracy analysis in georeferenced satellite images without independent ground control points," *Remote Sens.*, vol. 12, no. 24, pp. 1–13, 2020.