VecStatGraphs2D, A Tool for the Analysis of Two-Dimensional Vector Data: An Example Using QuikSCAT Ocean Winds

P. G. Rodríguez, M. E. Polo, A. Cuartero, Á. M. Felicísimo, and J. C. Ruiz-Cuetos

Abstract—**Circular or directional data are used in disciplines such as meteorology, geomatics, biology, and geology. The analysis of angular data requires special methods that are available in some statistical packages. However, these tools analyze only the angular values and do not include the vector modules, assuming unit vectors in all cases. In this letter, an open-source graphic and statistical package, i.e., VecStatGraphs2D, is described. It works in the R environment and provides statistics and graphics for modules (linear) and azimuths (circular), as well as graphics for the joint analysis of modules and azimuths. QuikSCAT satellite wind data are used to demonstrate some features of the package. QuikSCAT data are non-unit-length vectors, where both azimuth and magnitude (speed) are derived from u and v vector components (vector projections over the x- and y-axes). The example is used to show the seasonal change of winds in the Intertropical Convergence Zone, a key area in the ocean bird migration from the North to South Atlantic oceans.**

Index Terms—**Circular statistics, directional data, QuikSCAT, R project, statistical graphics.**

I. Introduction

CIRCULAR statistics, which is also known as directional statistics, deal with angular data, such as axes or vectors, to analyze the azimuth and magnitude in two dimensions.

A significant amount of literature exists on this topic [1]–[4]. Circular statistics has been applied in a variety of disciplines, including biology (e.g., bird migrations and animal tracking), meteorology (wind analysis) [5], geography (anisotropic cost analysis and travel) [6], geology (tectonics and magnetic fields), remote sensing [7], and laser scanner error analysis [8].

The statistics involving vectors can be used to analyze azimuths (angular) and magnitudes (linear), either jointly or independently. Some analyses focus on only azimuths (the "angular scale" [9]), assuming that the elements are unit vectors. However, many applications, such as mapping, geodesy, remote sensing, or photogrammetry, require the joint study of directional and linear components.

Unfortunately, this type of analysis has been incorporated into few of the currently available packages. The VecStatGraphs2D package has been developed to fill some of these gaps, and it also offers some specialized graphics that are not available in other packages. The R environment (http://cran.r-project.org/) is a statistical programming tool available as free software under the terms of the GNU General Public License.

II. Objectives

This letter aims to describe the principal features of VecStatGraphs2D, a new R package for the statistical analysis of 2-D directional non-unit-length vectors. QuikSCAT satellite wind data are analyzed to present the statistics and graphics features of the package.

III. Area, Data, and Material

This section describes the study location, the data, and the available software for directional statistics.

A. Study Area and Data

We used a wind data set for the Atlantic Ocean for a statistical analysis example. The wind data come from the SeaWinds scatterometer on the QuikSCAT satellite. This instrument generates daily vector wind data over the ocean's surface (http://manati.orbit.nesdis.noaa.gov/quikscat/) that have been used in a lot of applications [10]–[12]. Data are given as the projections u and v of the wind vector over the x-axis (longitude) and y-axis (latitude), respectively. These vectors can be used to estimate azimuth (the direction that wind blows toward) and vector module (wind speed) with a 25 km $\times$ 25 km spatial resolution.

We selected several subsets of 1920 wind vectors from an oceanic area called the Intertropical Convergence Zone (ITCZ). The ITCZ is a very interesting zone due to the seasonal wind changes, which act as a temporal switch that drives when the migrations of some oceanic birds begin [13].

Winds can be a barrier or a contributing force to the flights of oceanic wanderers. In the ITCZ, both calms and westerly winds prevent the transit over the migratory "gate," which is located at approximately 0°–10° N between Africa and South America. In this example, we used wind samples from a rectangular zone

Manuscript received July 9, 2013; revised August 6, 2013; accepted September 6, 2013.

P. G. Rodríguez, A. Cuartero, and J. C. Ruiz-Cuetos are with the Politechnic School, University of Extremadura, 10003 Cáceres, Spain (e-mail: pablogr@unex.es; acuartero@unex.es; jcruizcue@gmail.com).

M. E. Polo and Á. M. Felicísimo are with the University Center of Mérida, University of Extremadura, 06800 Mérida, Spain (e-mail: mepolo@unex.es; amfeli@unex.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2281840

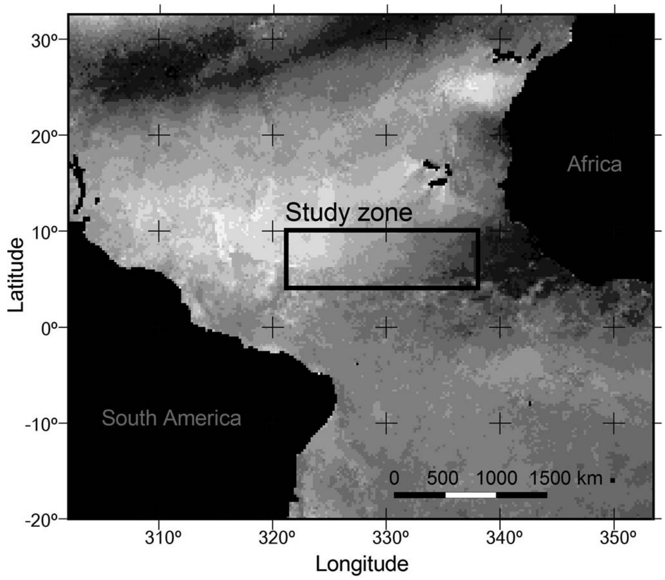

Fig. 1. Study area is a 650 km × 2200 km rectangle located in the ITCZ between 4° and 10° N latitude.

within a 4°–10° North latitude and 320°–340° longitude box (see Fig. 1).

The samples represent six periods corresponding to days 225, 250, 275, 300, 315, and 350 of the year 2002. This period provides a good example of the seasonal changes in westerly winds that drive ocean bird migrations.

B. Current Software and VecStatGraphs2D Features

Currently, the 2-D circular statistical free software available are the R packages Circular [14] and CircStats [15], which provide a large variety of functions for statistical analysis and some graphics representation of circular data. The commercial software available are Oriana [16] and CircStat (MATLAB toolbox). These packages generate a broad set of statistics, including circular descriptions (mean, median, circular standard deviation, circular variance, etc.), tests of uniformity, and single direction or von Mises distributions. These programs can also be used to create graphics, such as individual vectors, circular histograms, and mean direction vectors, but they only work with angular data and assume that the vectors are unit vectors.

The features of the VecStatGraphs2D package are divided into three sections: 1) the analysis of vector modules by means of linear statistics; 2) the analysis of azimuths using circular statistics; and 3) the graphical joint analysis of linear and angular data. The first two sections include descriptive statistics, such as the mean, variance, and standard deviation for linear and angular data and the circular dispersion, kurtosis, skewness, and von Mises parameters for circular data. The third section generates graphics, some of which are conventional and others that are less so. All the graphics can be generated in Scalable Vector Graphics (SVG) format, which provides the maximum quality for publications and allows for subsequent editing of any property (e.g., size, color, position) of any element (e.g., text, line, circle) by means of an SVG editor (e.g., Inkscape).

IV. VecStatGraphs2D Statistics and Graphics

A. Input Data Format

The input files are text files (.txt) in tuples of two or four tab-delimited elements.

The format of the input data in 2-D can be of three types:

- Cartesian coordinates: four-tuple, the first pair defining the node origin and the second pair defining the end node.
- Polar coordinates: two-tuple, where the first element is the module and the second is the azimuth.
- Incremental data (increasing x and y): two-tuple, where the first element is increasing x (Δx) and the second increasing y (Δy).

Only the first option allows the absolute location of each vector to be defined in a reference system; the remaining options only define magnitude and azimuth (θ). Angles are sexagesimal, and the default reference system is geographic: 0° to North (y-axis), incrementing clockwise.

B. Linear and Circular Statistics

Linear statistics are used to analyze the module or magnitude of the vectors as scalars. The routines provide descriptive statistics, including the number of elements, maximum and minimum values, arithmetic mean, standard error, standard deviation, variance, skewness, kurtosis, and the Rayleigh and Rao test of normality.

Circular statistics are used to analyze the azimuths and provide information such as the mean azimuth, mean module (by means of vector addition), circular standard deviation, circular variance, von Mises parameter, circular dispersal, skewness, and kurtosis (see expressions in [1]–[3]).

C. Uniformity Analysis

The uniformity of angular data is analyzed by means of the Rayleigh [1] and Rao tests [17], which provide the P-value of the hypothesis of uniformity.

D. Graphical Analysis

The package includes routines to plot the azimuth distribution (in both ungrouped raw data and circular histograms) and the joint modules and azimuth distribution. It can also generate a density map applied to the end nodes of the vectors, which allows for the evaluation of more complex distribution properties and the detection of outliers.

VecStatGraphs2D generates the following graphics:

- Distribution plot (Q–Q plot) for the graphical evaluation of the azimuth uniformity. In this graphic, each azimuth is plotted against the value expected for the vector given the uniform distribution. If the data are uniform, the points should plot on a 45° line.
- Ungrouped azimuth data distribution: each azimuth is plotted as a stacked point within a unit circle. If the sample size is large, each point may represent more than one piece of data. Red lines show the mean azimuth and confidence interval.

TABLE I
LINEAR STATISTICAL ANALYSIS OF WIND VECTORS
(* UNITS OF VECTOR MODULES ARE IN METERS PER SECOND;
** DIMENSIONLESS COEFFICIENT; $N = 1920$)

DAY	*MEAN	*STER	*SDEV	**SKEW	**KURT
225	3.89	0.04	1.52	-0.08	-0.13
250	5.34	0.03	1.20	-0.35	1.51
275	3.56	0.04	1.63	0.15	-0.31
300	5.58	0.04	1.54	-0.04	-0.53
325	7.19	0.04	1.85	-0.12	-0.19
350	6.43	0.04	1.66	-0.23	-0.39

DAY: Julian day; MEAN: arithmetic mean; STER: mean standard error; SDEV: standard deviation; SKEW: skewness coefficient; KURT: kurtosis coefficient; N: sample size.

TABLE II
CIRCULAR STATISTICAL ANALYSIS OF WIND VECTORS
(* UNITS OF ANGULAR VALUES ARE IN DEGREES;
** DIMENSIONLESS COEFFICIENT; $N = 1920$)

DAY	*MAZ	*MMD	*CSD	*CV	**VMIS	**SKEW	**KURT
225	14.5	0.63	0.96	0.37	1.65	0.09	0.71
250	21.5	0.79	0.69	0.21	2.71	-0.10	-1.24
275	267.9	0.44	1.28	0.56	0.99	-0.18	0.63
300	265.7	0.97	0.23	0.03	19.94	-0.72	1.14
325	245.0	0.94	0.36	0.06	8.04	1.00	0.00
350	233.8	0.95	0.33	0.05	9.92	-1.52	1.21

DAY: Julian day; MAZ: mean azimuth; MMD: mean module (from the vector sum); CSD: circular standard deviation; CV: circular variance; VMIS: Von Mises parameter; SKEW: skewness coefficient; KURT: kurtosis coefficient; N: sample size.

- Circular histogram: a polar-area diagram, where sectors have the same angular width, and the radius is proportional to the relative frequency of azimuths in the sector.
- Vector distribution: each vector is plotted as a line from the common (0, 0) origin with the corresponding module and azimuth. The mean vector is represented by a red arrow.
- End node density map: a graphic where the local density of the end nodes is represented a color ramp. The local density values are calculated by a moving window that calculates the count of end nodes per unit area.

Each chart type has some options that are controlled by parameters. The most notable is the option to export the plot to an SVG file, generating a high-quality and fully editable graphic.

V. RESULTS

Here, we show the results of the analysis of the QuikSCAT wind data, as explained in Section III-A.

A. Linear and Angular Statistical Analysis

Table I summarizes the linear statistics for data between day 225 and day 350 (minimum and maximum values, range, variance, and population statistics are not included in this table).

Table II summarizes the results of the circular statistics for the azimuths of unit vectors. The variation of the mean azimuth allows the detection of strong changes in wind direction during the study period. Complementary, i.e., mean module and circular standard deviation, and additional parameters show the progressive reduction of wind variability (see graphics), from variable N–NE winds to more consistent W–SW winds at the end of year. The large values of the mean modules and von Mises parameters suggest that the data are focused, as the uniformity tests confirm.

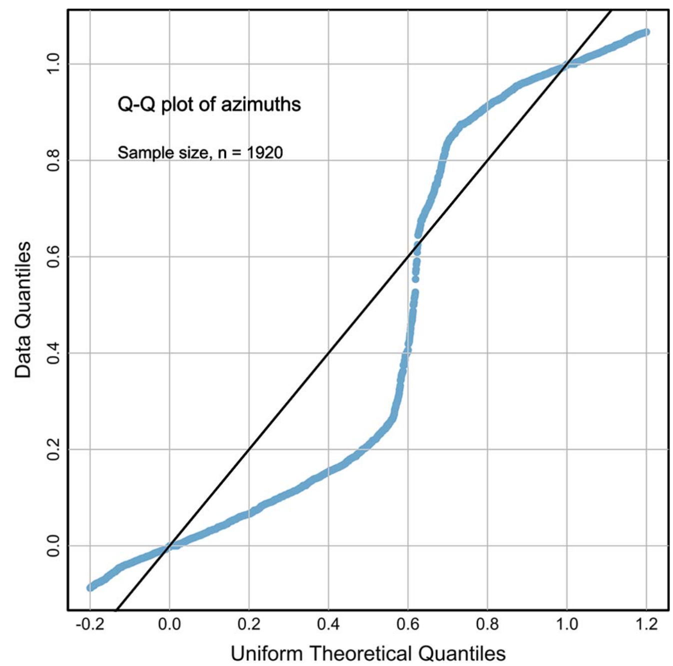

Fig. 2. Q–Q plot of azimuth values (day 225), showing that the data differ from a uniform circular distribution (diagonal line).

B. Uniformity Analysis

The uniformity analysis of the data distribution is performed using the Rao and Rayleigh tests. Both tests provide P-values < 0.01 in all cases and, consequently, the hypothesis of angular uniformity is rejected.

C. Graphical Statistical Plots

The graphical analysis includes a set of graphics that shows some examples for the wind data. Fig. 2 shows an azimuth quantile plot (Q–Q plot) for the first period (day 225), which allows a graphical test of the data adjusted to a uniform circular theoretical distribution.

Fig. 3 shows the data as stacking points in the second period (day 250), with the mean azimuth and 95% confidence arc indicated. Due to the sample size and distribution, each point in this graphic represents two observations. Winds in the first and second period blow mostly to the NE, acting as an obstacle to north-to-south bird migration.

Fig. 4 shows a circular histogram of relative frequencies for grouped data. The bin width is 15° in this example, but this is a customizable parameter. In this example showing the third period (day 275), we observe that the winds change compared with the previous days, now blowing to the west.

The previous graphics consider the wind data as unit vectors. Fig. 5 shows a graphical representation of the real modules and azimuths of the wind vectors (fourth period, i.e., day 300). The units of this graphic are in meters per second (wind speed), and

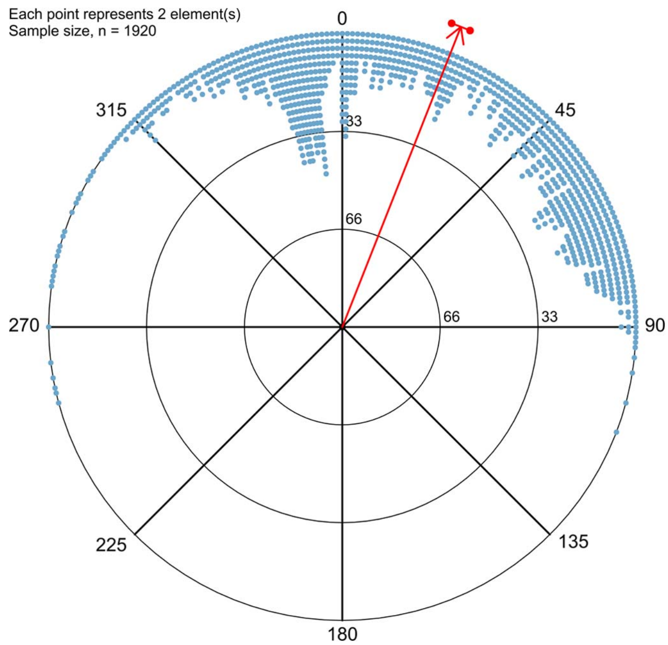

Fig. 3. Stacked data for the second period (day 250), where each dot represents two observations.

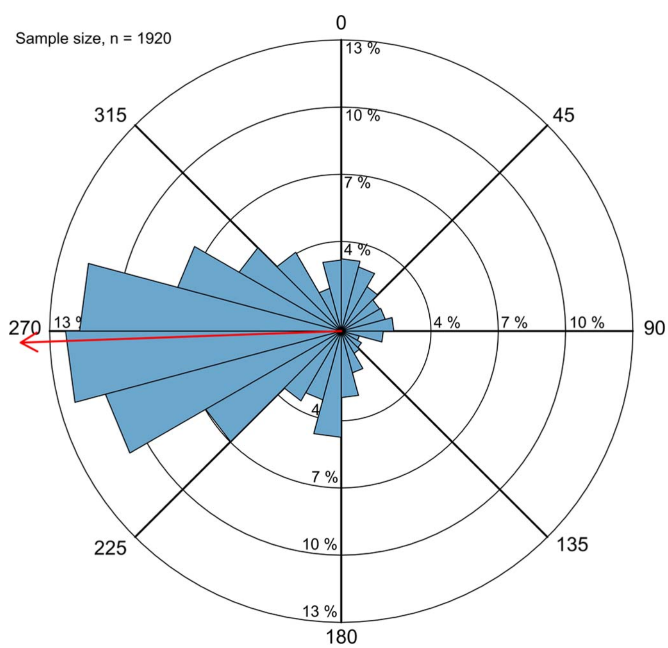

Fig. 4. Circular histogram with 15° sectors, showing the relative frequencies of winds in the third period (day 275).

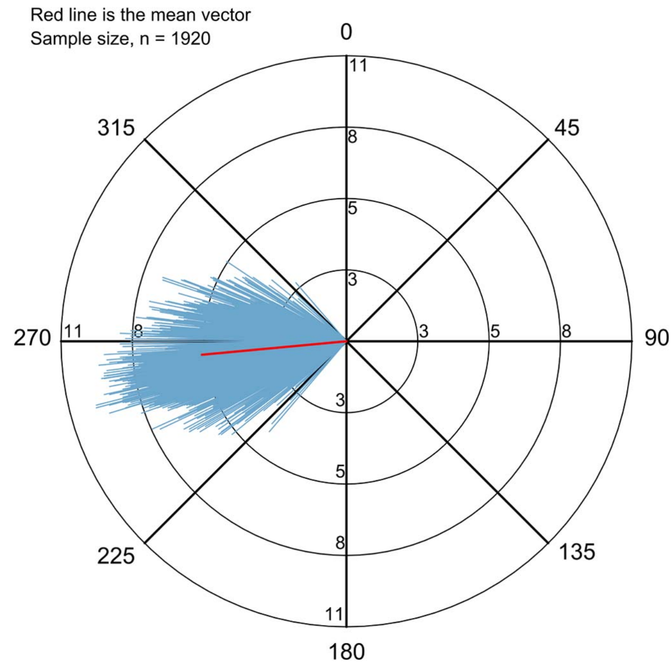

Fig. 5. Vector distribution of the fourth period (day 300). The units of the axis are in meters per second (wind speed). The red segment is the mean vector.

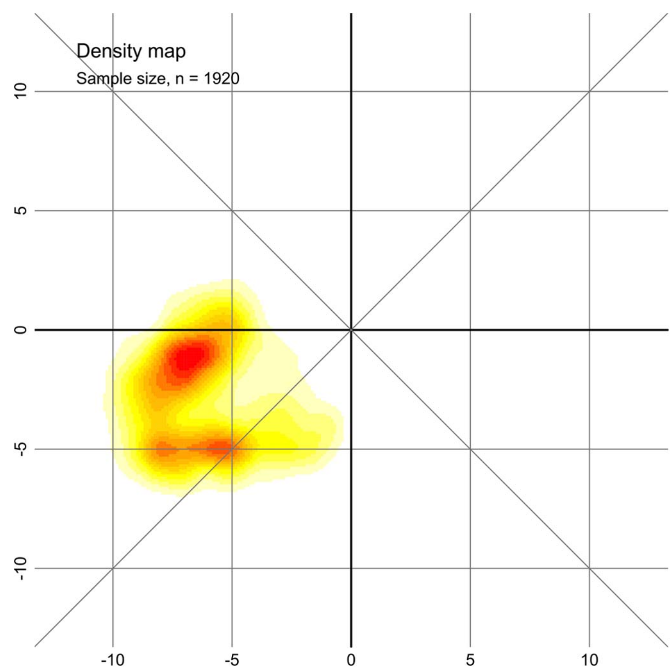

Fig. 6. Density map of the vector representation of the real azimuth and module values for the fifth period (day 325). The units of the axis are in meters per second (wind speed).

the red segment shows the mean vector generated by the vector addition (not from the arithmetic mean of the modules). The winds remain stable, blowing to the W.

Fig. 6 shows a density map of the vector end nodes, highlighting the concentration of vectors in both azimuth and module (fifth period, i.e., day 325, with winds drifting to the SW, similar to in the sixth period, i.e., day 350).

This plot can incorporate the end nodes as points and can detect the outliers as a percentage of the data. This detection is performed from the higher value of the modules or from the harmonic mean of a moving window. This plot is the best way of analyzing multimodal distributions or detecting complex patterns that are not revealed by the statistics.

Fig. 7 synthesizes the wind changes in the six periods, showing the mean wind vectors (see Fig. 5) and drifting from the NE (first two periods) to the SW (last two periods).

VI. Discussion and Conclusion

In this letter, we have presented an R package developed to analyze vector data in 2-D space. This package includes

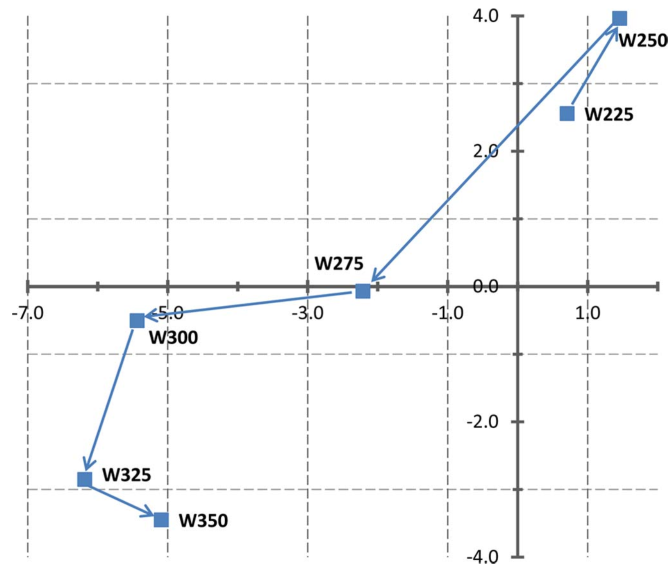

Fig. 7. Drift of the winds from the first period (day 225) to the sixth period (day 350), showing the changes in wind speed and direction. The units of the axis are in meters per second (wind speed). Wind acts as a spatial–temporal switch driving the bird migration [13].

conventional linear and angular descriptive statistics, as well as some inferential tests for circular distributions, but the most innovative contribution is the set of routines to generate high-quality plots. Some of these plots include features that are not available in other similar packages (complete vector representation and density maps). The plots can be saved as SVG graphics that are fully editable with vector graphics editors as Inkscape (open source).

The conventional statistics for circular data assume unit vectors $(\text{module} = 1)$, and the azimuths are analyzed based on that premise. QuikSCAT winds are an example of data whose analysis requires treatment as non-unit vectors, and they present an opportunity to illustrate the graphic capabilities of the VecStatsGraphs2D package. Graphics can help analyze this type of data and help understand some of its properties and spatial patterns.

REFERENCES

[1] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[2] S. R. Jammalamadaka and A. SenGupta, *Topics in Circular Statistics*. Singapore: World Scientific, 2001.

[3] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Chichester, U.K.: Wiley, 2000.

[4] J. H. Zar, *Biostatistical Analysis*, 4th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[5] J. A. Bowers, I. D. Morton, and G. I. Mould, “Directional statistics of the wind and waves,” *Appl. Ocean Res.*, vol. 22, no. 1, pp. 13–30, Feb. 2000.

[6] J. Corcoran, P. Chhetri, and R. Stimson, “Using circular statistics to explore the geography of the journey to work,” *Regional Sci.*, vol. 88, no. 1, pp. 119–132, Mar. 2009.

[7] A. Cuartero, Á. M. Felicísimo, M. E. Polo, A. Caro, and P. G. Rodríguez, “Positional accuracy analysis of satellite imagery by circular statistics,” *Photogramm. Eng. Remote Sens.*, vol. 76, no. 11, pp. 1275–1286, Nov. 2010.

[8] M. E. Polo and Á. M. Felicísimo, “Full positional accuracy analysis of spatial data by means of circular statistics,” *Trans. GIS*, vol. 14, no. 4, pp. 421–434, Aug. 2010.

[9] P. Berens, “CircStat: A MATLAB toolbox for circular statistics,” *J. Statist. Softw.*, vol. 31, no. 10, pp. 1–21, Sep. 2009.

[10] H. S. Anderson and D. G. Long, “Sea ice mapping method for SeaWinds,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 647–657, Mar. 2005.

[11] R. Royer, K. Goita, J. Kohn, and D. De Sève, “Monitoring dry, wet, and no-snow conditions from microwave satellite observations,” *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 670–674, Oct. 2010.

[12] R. Kumar, A. Chakraborty, A. Parekh, and R. Sikhakolli, “Evaluation of Oceansat-2-derived ocean surface winds using observations from global buoys and other scatterometers,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2571–2576, May 2013.

[13] A. M. Felicisimo, J. Munoz, and J. Gonzalez-Solis, “Ocean surface winds drive dynamics of transoceanic aerial movements,” *PLoS ONE*, vol. 3, no. 8, p. e2928, 2008.

[14] U. Lund and C. Agostinelli, *Circular*, ver. 0.4-3, 2011, R package.

[15] U. Lund and C. Agostinelli, *CircStats: Circular Statistics*, ver. 0.2-4, 2012, R package.

[16] K. C. Service, *Oriana for Windows*, ver. 4.01, Kovach Computing Service, Pentraeth, Wales, U.K., 2012.

[17] E. Batschelet, *Circular Statistics in Biology*. London, U.K.: Academic, 1981.