DISPERSED BY WIND PROJECT: A MOTION ANALYSIS OVER WIND FIELDS

Ángel M. FELICÍSIMO
Ingeniería Cartográfica, Geodesia y Fotogrametría
Escuela Politécnica
10071 Cáceres SPAIN
amfeli@unex.es

Jesús MUÑOZ
Real Jardín Botánico
Plaza de Murillo, 2
28014 Madrid SPAIN
jmunoz@ma-rjb.csic.es

ABSTRACT

Remote sensing data in which each pixel stores vectorial information represent a challenge for most image processing and GIS applications. For an ongoing project which aim is to test if wind is the dispersal vehicle of several plant groups in the Southern Hemisphere, we have used scatterometer wind data (QuikSCAT) to generate anisotropic cost-distance surfaces. Because of the vectorial nature of the QuikSCAT raw data, we found several problems in the data processing and implementation within GIS. The aim of this paper is to present the problems we found and to propose some solutions. Main problems were related to (1) lack of complete daily coverage, (2) lack of circular statistical tools within GIS packages, (3) reprojection of vectorial data, (4) maintenance of spatial continuity, and (5) restoration of direction values after reprojection of vectorial data. The emphasis of this study is to present the solutions we took for each of the previous problems, which in the end allowed generating the cost-distance surfaces with commercial GIS after programming some of the operations.

KEY WORDS
GIS, anisotropic cost-distance, vectorial data management, long-distance dispersal

1 INTRODUCTION

Hypothesis of difficult experimental demonstration are raised in some, if not many, scientific fields. Biological sciences are not an exception, and there are theories which foundations are more intuitions o beliefs than facts empirically checkable. The descriptive nature of such theories derives in many cases of the lack of instruments or technology to check the hypothesis behind. Fortunately, the oncoming technology occasionally allows testing such hypothesis to accept, reject or revise them.

As an example, the wind as been invoked to explain the strong floristic affinities shared among landmasses in the Southern Hemisphere. Long-distance dispersal by wind is favored by some authors (e.g. [1] for bryophytes, [2] for pteridophytes) and rejected by others who consider that present distributions in the Southern Hemisphere rather reflect old Gondwanan connections back to the Cretaceous (vicariance hypothesis, e.g. [3] for a general view, [4] for bryophytes).

Consequently, we can see that although wind is periodically presented as a dispersion vehicle, its role is not widely accepted perhaps because experimental confirmation was difficult. Certainly, the lack of reliable data on wind azimuth and speed at Earth scale hampered the empirical validation of the wind hypothesis until recently, when satellite scatterometers started to provide daily near-global coverage on wind properties with high statistical accuracy.

In this context, the general aim of Dispersed by Wind Project (DWP) is to test the hypothesis of long-distance dispersal by wind of several cryptogamic plant groups in the Southern Hemisphere (mosses, pteridophytes, lichens and liverworts). To this end, we have used data from the SeaWinds scatterometer (on board the QuikSCAT satellite) to model wind connectivity among landmasses and to test how well these "wind highways" explain the observed floristic similarities. In synthesis, we interpreted DWP as an example of multidisciplinary studies joining remote sensing, GIS and statistical analyses to model and check a biological hypothesis, a new and promising way in biological studies.

2 OBJECTIVES

The general aim of DWP is to test the hypothesis of long-distance dispersal by wind of several cryptogamic plant groups in the Southern Hemisphere. In order to reach this objective, we checked the degree of association between the floristic similarity index and wind connectivity 'similarity' matrices.

Consequently, the core of the DWP consists in verifying if the "winds paths" can explain similarities in flora composition. In the present work we will take care of the problems related to GIS and remote sensing processing, a set of questions that must be outstanding because they reflects pitfalls which solution would highly improve GIS operability.

3 MATERIALS AND DATA

3.1 THE QUIKSCAT DATA

Most data currently used to monitor the biosphere derive from Earth-orbiting satellites. A large part of those data are commercial products managed by private or
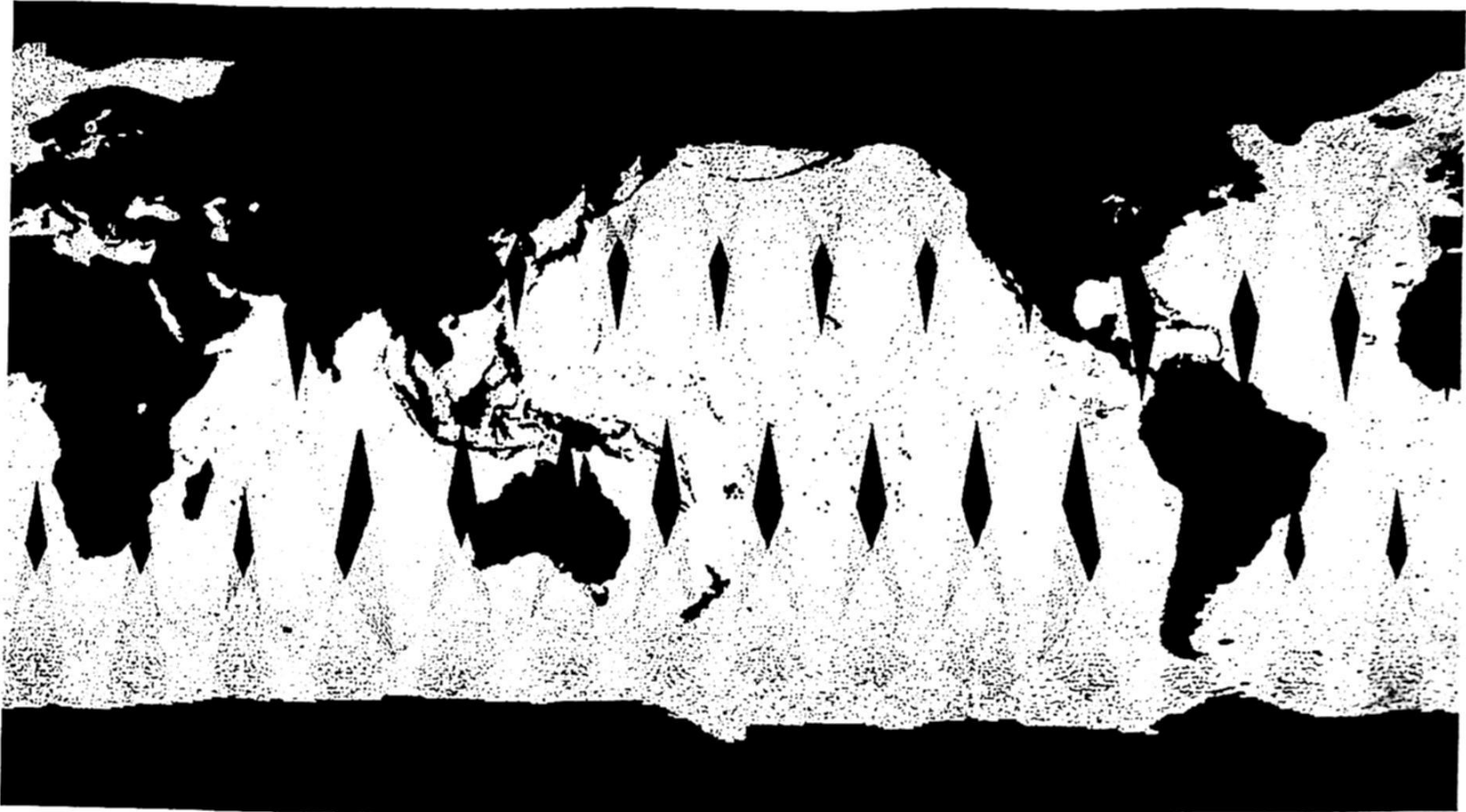

Figure 1. The image shows a typical coverage for one day of SeaWINDS on QuikSCAT Level 3 data.

governmental agencies (SPOT Image, European Space Agency, NASA, etc.). Others are freely available, and are an invaluable source of information for research at continental or global scale, as the budget needed for such studies using commercial products would be beyond most research projects budgets. Examples of studies benefited from such free datasets are those listed in [5], or in the World Climate Research Program, which long-term data sets are generated from observations of operational satellites [6]. Further studies using such sensors are [7] with SeaWiFS[1] or [8] with MODIS[2].

The SeaWinds on QuikSCAT scatterometer was developed by NASA JPL to measure the speed and direction of ocean surface winds. The QuikSCAT satellite

was launched into a sun-synchronous, 803-kilometer, circular orbit on 19 June 1999. The SeaWinds instrument is a specialized microwave radar (scatterometer) that measures near-surface wind speed and direction under all weather and cloud conditions over Earth's oceans [9].

All QuikSCAT standard products are in Hierarchical Data Format (HDF). HDF is a data format developed by the National Center for Supercomputing Applications (NCSA) at the University of Illinois[3]. The QuikSCAT HDF files contain a set of 16 Scientific Data Sets (SDS). An HDF SDS is a fixed dimensional array. The QuikSCAT SDS are a simple 0.25° rectangular grid that contains 1440 pixels from east to west and 720 pixels from south to north (JPL, 2001b) in a standard "plate carrée" projection (Figure 1).

---

[1] http://seawifs.gsfc.nasa.gov/SEAWIFS.html
[2] http://modis.gsfc.nasa.gov/

---

[3] http://hdf.ncsa.uiuc.edu

239

The data used in our study include daily data from June 1999 to March 2003, and were obtained from PO-DAAC (Physical Oceanography Distributed Active Archive Center).

3.2 BIOLOGICAL DATA

For DWP we have recorded distributions of mosses, liverworts, lichens, and pteridophytes for 27 localities in the Southern Hemisphere, including oceanic islands and continental areas (Figure 2).

The recorded data are four data sets: mosses (601 species), liverworts (461), lichens (597), and pteridophytes (192), totaling 1851 species. Only species shared by at least two localities were included.

3.3 SOFTWARE

Software includes IDL 5.4 (Research Systems, Inc., Boulder, CO, USA), ArcInfo 7.1.2 and ArcView 3.2 (Environmental Systems Research Institute, Redlands, CA, USA). We must say that there exist many software applications to work with remote sensing and, more generally, spatial data. At present, the limits between remote image processing applications and geographic information systems are more and more blurring. The

types of software applications, very different in their early days, have suffered a "convergent evolution" and presently they share many of the previously exclusive tools.

All current applications allow working with conventional remote sensing images, in which each pixel stores a reflectance value in a fixed wavelength interval. Data from active sensors (radar) also can be treated with a more restricted set of applications. All are scalar data, in which each pixel has assigned a single integer (digital value).

In the last few years we have assisted to the bloom of a new generation of data essentially different form the previous. In them each pixel stores information about a vector, not on a scalar as in the previous, general case. We will see how this type of data generates problems which solutions are not easy through current image processing applications and geographic information systems. It is not our aim to pillory GIS packages, but rather to highlight some pitfalls which solution would highly improve GIS operability, especially by incorporating tools for vectorial data analysis. To this end we will point out some problems we experienced and the solutions we adopted in the progress of the DWP, a project that uses wind vectorial data.

Figure 2. Study areas included in the Dispersed by Wind Project.

need to work simultaneously with remote sensing images and digital cartographic information has caused that both

4 METHODOLOGY

Statistical methods are based in the comparison of similarity/distance matrices. Our aim is to check the degree of association between floristic similarity and wind connectivity matrices.

4.1 DISTANCES O SIMILARITY MATRICES

4.1.1 Floristic similarity

Floristic similarity between each locality pair was calculated using the Ochiai index [10]. This index is based in concordance of species presence/absence and varies from 0 (no similarity) to 1 (same floristic composition). The results are four 27x27 floristic similarity matrices, one for taxonomic group.

4.1.2 Wind distances or connectivity

Anisotropic Cost Analysis (ACA) is a procedure to calculate the minimum accumulative-travel cost from a source to each cell location on a grid or raster model. The analysis calculates the accumulative cost over the field of wind vectors, considering such field as a friction surface. ACA is based in the measured wind-vectors fields. Connectivity extent between two points depends upon two factors: dispersion of the vector set, and the value of their modules (wind speed). Two well connected points will be joined by a vector chain without sinuosities. If those vectors have large modules the cost-path will be low.

ACA tools allow computing cost values from georreferenced data including the coordinates of the origin and end points, and the direction and module of the intervening vectors. We also should define the angular tolerance allowed as deviation from the main vector direction. This function assigns a minimum cost for the movement in the exact vector azimuth, and incrementally penalizes angular deviations, generating less strict models. In order to perform the anisotropic cost analysis, some conditions must be fulfilled, namely spatial topologic continuity, as well as distance and azimuth correctness. The result will be a surface in which the value of each pixel equals the cost of connecting it with the pixel origin. In our study, the above conditions were not fulfilled by the original data, and the GIS process were not negligible, as it is exposed next (Section 0).

Anisotropic cost analysis is not available in all GIS packages. Among common ones, only ArcInfo 7 (ESRI Inc., Redlans, CA, USA), and Idrisi 32 (ClarkLabs, Worcester, MA, USA) have this tool. Surprisingly, the tool is not implemented in ArcInfo 8 (formerly ArcGIS). ArcInfo tool (*costpath* command) is considerably more complete that its counterpart in Idrisi (*varcost* command), and this accounts for its use in the DWP.

To run the cost-path analysis requires (1) to determine the origin, namely, in DWP, each of the 27 sample points, (2) to define the wind field based in wind azimuth and speed, where azimuth defines minimum-cost direction and speed is considered the inverse of resistance, and (3) to define the angular tolerance, or the increment of cost corresponding to incremental deviations from the minimum-cost azimuth. As an example of the cost-path analysis, Figure 3 shows the results of the analysis for the April 1-10, 2000 period for Bouvet Island (54°26'S, 3°24'E).

Figure 3. Anisotropic cost surface from Bouvet Island (arrow) for period April 1-10, 2000 (increasing cost from white to dark in arbitrary units).

4.2 WIND CONNECTIVITY PROCESSING

Scalar data are simple data measured in a linear scale (e.g. reflectance values in a remote sensing image), and can be computed by conventional statistical methods. However, QuikSCAT pixels store vectorial information. A vector is a structure that combines both metric and topologic properties. Metric properties are the azimuth or angle respect a reference system, and the module or vector length. The topologic property is the vector direction (from-to). Under the convention adopted for QuikSCAT data, a wind direction of 0° implies a flow toward (not from) the North. The QuikSCAT vectors are stored as their u and v components in separate SDS of the HDF file.

While processing the original wind data to generate the cost surfaces needed to compare with the floristic similarities we found several problems:

4.2.1 First problem: Lack of daily data and circular statistics

QuikSCAT daily cover is not complete. Figure 1 shows empty areas result of the orbital trajectory. These areas must be removed because they break the spatial continuity necessary for anisotropic-path analysis. As those empty areas change daily, the adopted solution was to generate 10-day synthesis. This solution guarantees that each pixel is covered by several measurements, commonly between 5 and 7.

Most GIS software packages implement map-algebra calculations. However, wind-vectors cannot be analyzed throughout conventional statistics because they are distributed in a circular scale. A simple vectorial mean cannot be calculated straightforward, but the process must be decomposed into several elementary steps to use the appropriate circular data statistics [11, 12].

241

For example, it is not possible to calculate the average wind azimuth directly, but operating with u and v components, which are themselves linear variables. Decomposition of a vector with module S and azimuth γ is simple: $u = S \cdot \sin \gamma$ and $v = S \cdot \cos \gamma$.

Although such decomposition is a common procedure in circular statistics, it is not included in any GIS package, and it must be programmed. The average azimuth for a given dataset is then calculated from those u and v components. The first step involves calculating the arithmetic mean of the u and v components separately for each pixel using the 10-day SDS syntheses. The azimuth is calculated later with the expression $\gamma = \arctan(\bar{v}/\bar{u})$. Special care deserve pixels where $u=0$ and $u=v=0$ (no wind). Average wind speed follows the expression: $S = sqrt(u \cdot u + v \cdot v)$.

Figure 4. Three-day sequence of satellite data over Australasia showing the displacement of empty areas. 10-day syntheses guarantee complete area coverage.

4.2.2 Second problem: Projection properties and spatial continuity

QuikSCAT data could not be used directly in our project because the original simple cylindrical projection ('plate carrée') breaks the spatial continuity at Greenwich meridian. Under such conditions, two points next to each other but separated by Greenwich meridian would be, in practice, unconnected. The best solution would be also the simplest: to work directly with unprojected data in spherical coordinates (latitude/longitude). However, GIS packages only allow spatial analysis of data projected in a planar coordinate system. This is a severe pitfall of GIS packages that should be addressed.

The adopted solution was to project the data to Universal Polar Stereographic (UPS), a planar perspective projection where one pole is viewed from the other pole. The tangent point was the South Pole. Under such conditions this is a conformal projection with true-direction values from the South Pole. Figure represents average wind speed in the study area after transformation to UPS projection, which guarantees continuity over Greenwich meridian. A clear but thin belt of strong winds (pale tones) surrounds Antarctica, after which a wide calm belt (dark) is evident.
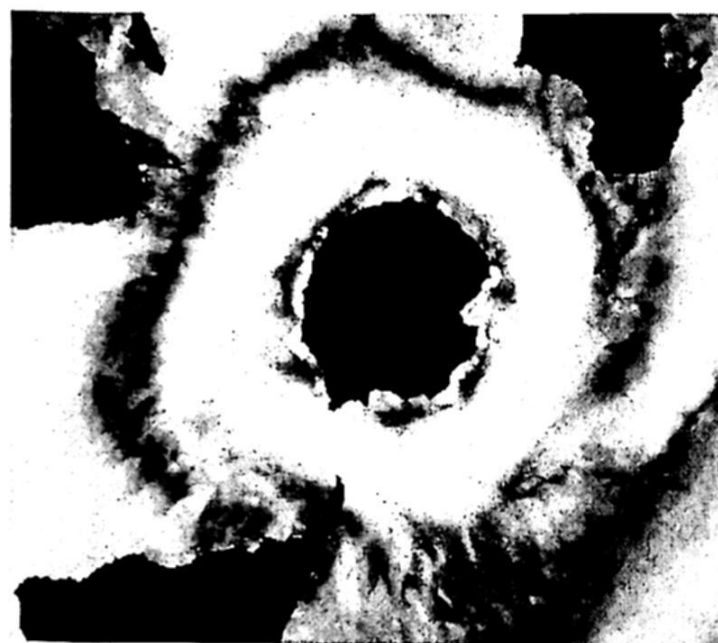

Figure 5. Example of SeaWinds on QuikSCAT data in UPS projection. Variable is average wind speed (increasing speed from dark to pale: 0-30 m s^{-1}).

4.2.3 Third problem: azimuth values restoration under UPS projection

Transformation from "plate carrée" to UPS projection guarantee spatial continuity, but it does not transform azimuth data to the new projection. This is another pitfall of GIS packages: raster data are projected without considering that some data are projection-dependent. For example, a pixel with azimuth value of 45 (wind towards the North-East) will keep this value after projection. Once in UPS projection this value will be correct only if the pixel location is exactly in the upper half of the

Greenwich meridian of Figure 5, otherwise information will be wrong.

In DWP we have developed a procedure that solves the above problems in the particular case of UPS projection. It takes two steps (Figure 6).

1. We define a new global Cartesian system in which the origin of coordinates (0,0) is the South Pole. Next we calculate the azimuth of the vector pointing towards the true North in the new system. The azimuth of a pixel i with coordinates $[x(i), y(i)]$ in the new system follows the expression: *true North azimuth* = arctan $[y(i)/x(i)]$. This *true North azimuth* is the deviation generated in the projection process, and will be employed to calculate the azimuth in UPS projection. Depending on the place of the pixel on the projected space, its *true North azimuth* will be distributed in the 0-360° range.

2. Denoting "*true azimuth*" to the azimuth values in the original "plate carrée" projection, calculation of azimuth in the UPS projection will follow the expression:

UPS azimuth = (true North azimuth + true azimuth) · modulus 360

This solution is not general but specific for UPS projection. However, it shows a problem that only can be solved after an important modification of projection routines, considering the vectorial or scalar nature of the raster data.

Table 1 (see end of paper) shows the summary of operations performed from SeaWinds HDF data reading to calculation of cost-path. Although the programming language is indicated, the processes are independent of software applications; they are shown here just as an example.
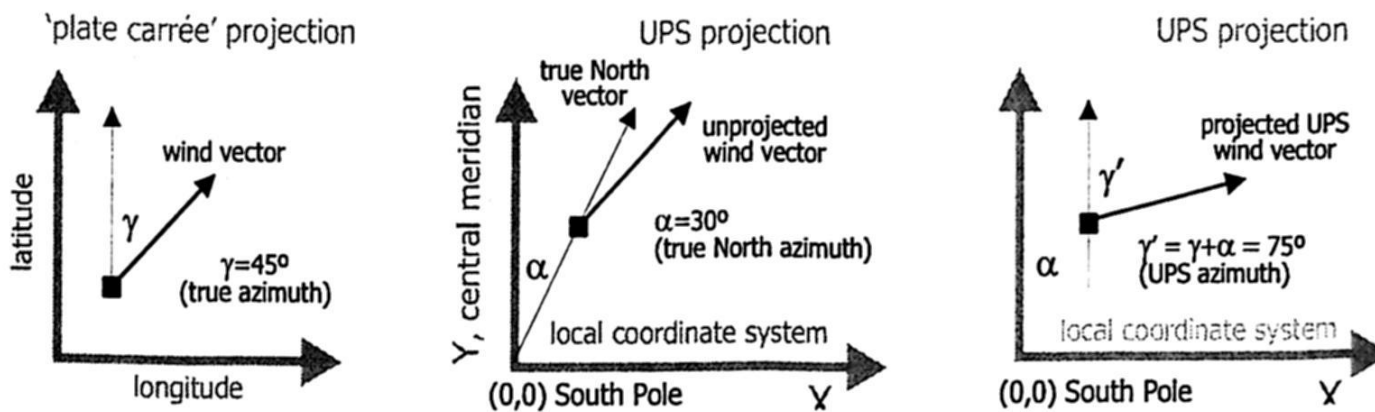

Figure 6. Calculation of azimuth vector in UPS projection.

5 CONCLUSION

In this paper we raise some questions arisen during DWP progress that are common to any work with vectorial data. Encountered problems are not negligible, because vectors appear often as spatial datasets (e.g.: aspect, dip in geological maps, animal dispersal movements, or sea currents, besides wind analysis as in the present study). Analysis with such data is the field of GIS. However, it seems evident that management and analysis of vectorial data with the available tools can cause troubles.

A downside is the impossibility of working with bidimensional (latitude, longitude) or three-dimensional (latitude, longitude, distance from origin) spherical coordinates. Occasionally, a GIS can *represent* the data from spherical coordinates, but the analyses *must* be done on projected data. This is not a severe problem when the study area is small. However, in global scale studies distortion in metric properties or rupture of topologic properties can prevent some analyses.

Further shortcomings come from the nature of the data employed in the DWP. Generally, GIS are not designed to manage vectorial data, and it will be necessary

programming of operations that, for scalar data, would be straightforward. For example, an arithmetic mean or standard deviation of a dataset would be calculated as a mean direction and a circular standard deviation [12]. Their formulation is simple, but such calculation tools are not included in any GIS package.

Finally, in current GIS to project data means to change the location of spatial elements while maintains constant theirs attribute values. This is reasonable with scalar data (temperature, reflectance, etc.), but it produces severe errors when treating azimuths or other directional values, because the reference system changes.

How to explain the lack of literature on this type of problems? There is a wealth of publications on wind data, but they mainly deal with data capture algorithms [5, 15], sensor calibration and influential factors [16], quality data analysis [17], or the foreseeable impact of the new data in previous climatic models [18]. Perhaps the limited use of such data outside those fields resides [13] in that they are available only recently and in the prevalence of local over global studies. Even some of such global studies does not experience the problems found in the DWP because they are not affected by metric or topological constraints. For example, Chen [19] studied the results of models generated from QuikSCAT and TOPEX/POSEIDON

243

(altimetry) data over a large Pacific Ocean extension. In this case the only limitation was that the data were superimposable, and consequently he used the original "plate carrée" projection. The same happens in [20] studying sea surface temperature or [21] comparing temperature and QuikSCAT data over the tropics.

In the Dispersion by Wind Project we have performed analysis of vectorial data through a commercial GIS. Although the adopted solutions are appropriate for our project, we would like to attract developers' attention to improve GIS tools. Otherwise, management as well as analyses and operations with vectorial data, increasingly common with the launch of new satellites, can result prevented.

ACKNOWLEDGEMENT

REFERENCES

[1] B.O. van Zanten & T. Pócs, Distribution and dispersal of Bryophytes. *Advances Bryol.*, *1*, 1981, 479-562.

[2] P.J. Brownsey, New Zealand's pteridophyte flora - plants of ancient lineage but recent arrival? *Brittonia*, *53*, 2001, 284-303.

[3] C.J. Humphries & L.R. Parenti, *Cladistic biogeography: Interpreting patterns of plant and animal distributions* (Oxford: Oxford University Press. 1999).

[4] W. Frey, M. Stech & K. Meissner, Chloroplast DNA-relationship in palaeoaustral Lopidium concinnun (Hypopterigiaceae, Musci). An example of stenoevolution in mosses. Studies in austral temperate rain forest bryophytes 2. *Plant Systematics and Evolution*, *218*, 1999, 67-75.

[5] F.J. Wentz & M. Schabel, Precise climate monitoring using complementary satellite data sets. *Nature*, *403*, 2000, 414-416.

[6] G. Ohring & A. Gruber, Climate monitoring from operational satellites: accomplishments, problems, and prospects. *Advances in Spatial Research*, *28*, 2001, 207-219.

[7] M.J. Behrenfeld, et al., Biospheric Primery Production during an ENSO transition. *Science*, *291*, 2001, 2594-2597.

[8] C.O. Justice, et al., An overview of MODIS Land data processing and product status. *Remote Sensing of Environment, in press*, 2002.

[9] J.P. Laboratory, *QuikSCAT Science Data Product. User's Manual. Version 2.2*, JPL Document 18053, California Institute of Technology, CA.

[10] P. Legendre & L. Legendre, *Numerical Ecology* (Amsterdam: Elsevier. 1998).

[11] K.V. Mardia, *Statistics of directional data* (New York, USA: Academic Press. 1972).

[12] N.I. Fisher, *Statistical analysis of circular data* (Cambridge, UK: Cambridge University Press. 1993).

[13] J.F. Commandeur & W.J. Heiser, *Mathematical dérivations in the proximity scaling (PROXSCAL) of symmetric data matrices*, Technical Report RR-93-03, Leiden University. Leiden, NL.

[14] D.A. Jackson, PROTEST: a PROcrustean Randomization TEST of community environment concordance. *Ecoscience*, *2*(3), 1995, 297-303.

[15] D.-J. Kim & W.M. Moon, Estimation of sea surface wind vector using RADARSAT data. *Remote Sensing of Environment*, *80*, 2002, 55-64.

[16] W.T. Liu, Progress in scatterometer application. *Journal of Oceanography*, *58*, 2002, 121-136.

[17] A. Stoffelen, M. Portabella & A. Voorrips. Towards the assimilation of QuikSCAT winds. *Fifth International Winds Workshop Proceedings*. Lorne, Australia. 2000.

[18] Y. Tahara. The preliminary study of the impact of QuikSCAT/SeaWinds ocean surface wind data to the JMA global model. *Fifth International Winds Workshop Proceedings*. Lorne, Australia. 2000.

[19] D. Chen, Applying satellite remote sensing to predicting 1999-2000 La Niña. *Remote Sensing of Environment*, *77*, 2001, 275-278.

[20] F.J. Wentz, et al., Satellite measurements of sea surface temperature through clouds. *Science*, *288*, 2001, 847-850.

[21] C. Day, Two satellites observe details of ocean-atmosphere coupling. *Physics Today*, *56*, 2000, 23-24.