Vector Correlation: Review, Exposition, and Geographic Application

Brian Hanson, Katherine Klink, Kenji Matsuura, Scott M. Robeson, and Cort J. Willmott[1]

Department of Geography, University of Delaware, Newark, DE 19716,
FAX 302/831-6654, e-mail byy00920@udelvm.udel.edu.

Abstract. **Correlation provides one of the most useful and widely employed techniques of data analysis in geography and in the larger scientific community. Most correlation coefficients describe relations among scalar data. Vector-valued data have received less attention despite their obvious geographic importance: vectors can describe motions across the landscape and relations between locations. After a brief review of the meaning and past applications of vector correlation, we present a vector correlation measure for comparing two-dimensional vectors. This vector correlation describes the goodness-of-fit of a relationship between two sets of vectors that includes translation, scaling, and either rotational or reflectional dependency. Linear regression parameters additionally emerge from the computation. We illustrate the properties of the vector correlation using wind velocity data.**

Key Words: vector correlation, statistics, bootstrap, wind.

QUANTITATIVE comparisons among two or more variables are well known in the geographic literature. Most of us are familiar with Pearson's product-moment correlation between two variables, and for decades geographers have used that correlation to describe the degree of association between two variables (Taylor 1977; Slocum 1990). Other measures of correlation, including rank-order correlations, have been used when required by the data (Slocum 1990). Despite our familiarity with and use of correlation, geographers have made little use of bivariate correlation between those variables, like the movement of people and goods, whose observations may be expressed as vectors. Geographers' use of univariate directional statistics spans at least two decades (Clark 1971, 1972; Gaile and Burt 1980), but it too has been limited (Gaile 1990). Geologists have made more frequent use of directional statistics (cf. Fisher et al. 1985; Davis 1986). Our use of whole-vector (nonunit) univariate statistics is even less common (Odland et al. 1989). More importantly, the "vector" correlations that geographers have proposed and used (Costanzo and Gale 1984) have not accounted for magnitude. One could even argue that our correlations of directional data have been inadequate (see the exchange between Costanzo and Gale 1985, and Klink and Willmott 1985). With few exceptions (e.g., Tobler 1978), correlations among vector-valued spatial fields fail to appear in the geographic literature. Our purpose is to introduce vector correlation among vector-valued spatial fields into the geographic literature and to explicate a particularly useful vector correlation coefficient.

Principles of Vector Correlation

Some geographical data can be represented by vectors or line segments characterized by their length and their direction. Some vectors have a constant, unit magnitude and are called directional or unit vectors. Vectors with nonunit magnitudes are termed whole vectors or simply vectors. Common examples of vectors include topographic slopes, wind velocities, ocean currents, and migration rates of human populations. Examples of directional data include orientations of cirque basins and directions of a road network. Any point along a path can be assigned a sense of direction, but magnitude cannot be assigned unless a rate of motion or a distance along the path is apparent.

Annals of the Association of American Geographers, 82(1), 1992, pp. 103–116

A useful vector correlation measure should possess several desirable properties. The measure should have a bounded magnitude, preferably such that the maximum magnitude indicates a perfect functional relationship and a zero magnitude indicates complete independence of the two vector-valued variables. Simple linear transformations of either of the variables should not influence the correlation. A simple linear transformation may be any combination of scaling ($\mathbf{z}'_j = k\mathbf{z}_j$, where k is a scalar constant, $\mathbf{z}_j$ is the jth observation of a vector variable, and $\mathbf{z}'_j$ the transformed observation, Fig. 1a, b) and translation ($\mathbf{z}'_j = \mathbf{z}_j + \mathbf{a}$, where $\mathbf{a}$ is a constant vector). In contrast, the measure should account for both rotational and reflectional dependencies between two vector-valued variables, $\mathbf{z}_j$ and $\mathbf{w}_j$. Vector-valued variables are rotationally dependent when the pairwise difference between the angular components of two variables is constant: $\angle\mathbf{z}_j - \angle\mathbf{w}_j = \Theta =$ constant. Rotational dependence indicates that the two vector series turn clockwise or counterclockwise together (Fig. 1a, c, d). The vectors $\mathbf{z}_j$ and $\mathbf{w}_j$ have a reflectional dependence when the pairwise sum of their angles is constant: $\angle\mathbf{z}_j + \angle\mathbf{w}_j = \Phi =$ constant. With reflection, the vectors in one variable tend to progress clockwise while those in the other variable progress counterclockwise (Fig. 1a, e, f). Vector correlation should be insensitive (in magnitude) to rotation or reflection in either variable (in other words, its magnitude should be rotation- and reflection-invariant), although the correlation ought to provide information on the angle of rotation (Θ) or axis of reflection ($\Phi/2$).

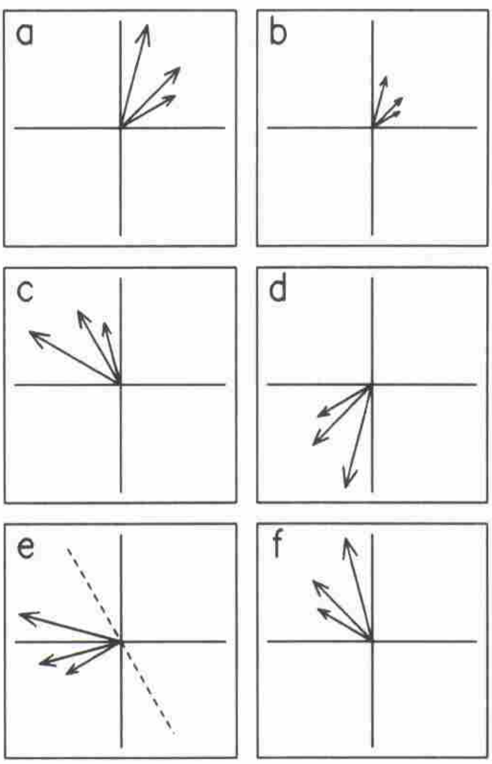

Figure 1. Examples of vector scaling, rotation, and reflection. The arrows in panel (a) show a set of three vectors that undergo transformations, as a set, in the subsequent panels. Panel (b) shows the three vectors scaled at 0.5, (c) a counterclockwise rotation of 75°, (d) a 180° rotation, (e) a reflection about an axis 30° counterclockwise from vertical, and (f) the reflection about the vertical axis of the diagram.

Vector Correlation Literature

A number of directional and vector correlation measures have been proposed, including both nonparametric and parametric coefficients. One of the earliest contributions on the expression of the correlation between vectors appeared more than fifty years ago (Masuyama 1939). Most measures have been developed for directional data, but some could be extended to accommodate vector observations (Table 1).

Nonparametric correlation coefficients for directional data are based on the ranks of directional variables. Measures proposed by Hillman (1974) and Fisher and Lee (1982) range from -1 for a reflectional relationship between the variables to $+1$ for a rotational dependence. Uncorrelated variables result in a value of zero. Mardia (1975) defined a slightly different rank correlation coefficient with a range of zero to one. His coefficient is defined as the maximum of a reflection and a rotation measure, allowing both types of dependence to be distinguished. Perfect rotational or reflectional dependence occurs when the measure is one and uncorrelated variables yield a value of zero. Fisher and Lee (1982) extended Mardia's coefficient to range from -1 to $+1$. All of the above measures are invariant under rotation or reflection, but they do

Table 1. Comparison of Directional and Vector Correlation Measures

Author	Properties	Range
Masuyama (1939)	V, P, R, F	[−1,1]
Epp et al. (1971)	D, P	[−1,1]
Downs (1974)	D, P, R, F, A	[−1,1]
Hillman (1974)	D, N, R, F	[−1,1]
Mardia (1975)	D, N, R, F	[0,1]
Kundu (1976)	V, P, R, A	[0,1]
Johnson & Wehrly (1977)	D, P, R, F, A	[0,1]
Mardia & Puri (1978)	D, P, R, F	[0,4]
Stephens (1979)	D, P, R, A	[0,1]
Jupp & Mardia (1980)	D, P, R, F	[0,c][a]
Batschelet (1981)	D, P, R	[0,1]
Fisher & Lee (1982)	D, N, R, F	[−1,1]
Rivest (1982)	D, P, R, F	[−1,1]
Fisher & Lee (1983, 1986)	D, P, R, F	[−1,1]
Wylie et al. (1985)	V, P, R	[0,1]
Breckling (1989)	V, P, R, F, A	[0,1]
Hanson et al. (this paper)	V, P, R, F, A	[−1,1]

[a] See text.
Properties indicated as follows:
D: Limited to directional data.
V: Usable with whole vector data.
P: Parametric.
N: Nonparametric.
R: Invariant under rotation of the original data.
F: Invariant under reflection of the original data.
A: Angle of rotation or axis of reflection is given.

not provide information on the axis of reflection or angle of rotation. Since vectors can be ranked by either magnitude or direction, rank correlation measures cannot be extended unambiguously to whole-vector variables.

Parametric directional measures suggested by Mardia and Puri (1978) and Jupp and Mardia (1980) give values that range from zero to a positive upper bound. The coefficient defined by Mardia and Puri is zero for uncorrelated directions, two for perfect rotational or reflectional dependence, and reaches four when the two variables are coincident. Jupp and Mardia (1980) proposed a coefficient whose limit is based on the rank of the covariance matrix of the trigonometric components of the directional variables. Their coefficient is zero for uncorrelated data and reaches its maximum when the variables have a perfect rotational or reflectional relationship. Both measures are invariant under rotation or reflection, but neither conveys information on the angle of rotation or axis of reflection.

Other parametric directional coefficients are within the range −1 to +1. The coefficient proposed by Epp et al. (1971) reaches −1 when the directional variables are 180° out of phase, is zero for a phase shift of ±90°, and reaches +1 when the variables are coincident. Their coefficient also is zero for uncorrelated data. Downs's (1974) directional coefficient also varies between −1 and +1, where −1 denotes a perfect reflectional dependence and +1 indicates a perfect rotational relationship. The axis of reflection or angle of rotation can be obtained from the measure. Coefficients developed by Rivest (1982) and Fisher and Lee (1983, 1986) also range from −1 to +1. Once again, −1 indicates perfect reflectional and +1 perfect rotational dependence. Except for the Epp et al. (1971) measure, all the coefficients described here are invariant under rotation or reflection. Only Downs's (1974) statistic contains information on the axis of reflection or angle of rotation.

Parametric measures that range from zero to one include a coefficient developed by Johnson and Wehrly (1977). Their correlation reaches one when one variable is a constant rotation, or a rotation with a reflection, of the other variable. Stephens (1979) and Batschelet (1981) describe measures that attain their maximum (+1) for a rotational dependence and are zero for uncorrelated data. Stephens's formulation also gives the mean rotation angle between the two variables. Johnson and Wehrly's coefficient is invariant for rotation or reflection in either variable. Both Stephens's (1979) and Batschelet's (1981) coefficients are invariant only for rotation and do not detect reflectional relationships.

Whole-vector correlation measures have been proposed by Kundu (1976), Wylie et al. (1985), and Breckling (1989). Kundu's (1976) coefficient ranges from zero to one and is maximized by perfect rotational dependence between the variables. The angle of rotation is easily found from the measure. The coefficients discussed by Stephens (1979) and Batschelet (1981) are essentially unit-vector reductions of Kundu's measure. Wylie et al. (1985) base their coefficient on the individual correlations of trigonometric vector components. Rotational dependence results in a correlation of one, but the angle of rotation is not provided. In each case, the measure is rotation-invariant and is zero when the variables are uncorrelated. Neither measure is sensitive to reflectional relationships, and they will vary in magnitude under reflection of either variable. Breckling (1989) described a coefficient that is very similar to Downs's (1974) measure, with an extension to accommodate vector data. His coefficient attains a value of one for either a perfect rotational or reflectional dependence, is zero for uncorrelated variables, and is invariant under rotation or reflection. The angle of rotation or axis of reflection also can be determined. Breckling's correlation is synonymous with our development below, but is not straightforward computationally nor does it allow easy interpretation of rotational or reflectional qualities.

Rotary (vector) cospectra (Mooers 1973) also can (in theory) lead to correlation expressions equivalent to ours for both rotational and reflectional correlation. As with cross correlation functions derived from scalar cospectra, short data series or nonstationarity can produce unreliable estimates.

Vector Correlation in a Regression Context

Simple Regression with Vector-valued Variables

Correlation measures how strongly some underlying function relates one variable to another. For a simple underlying function, the extraction of a correlation coefficient may closely follow the estimation of the function's parameters from data; parametric correlation and regression are inextricably linked.

Pearson's product-moment correlation is easily understood as a goodness-of-fit measure associated with a least-squares linear regression model. Most of us visualize correlation by the degree to which an ellipse of data points approximates a regression line drawn through the data. The vector correlation explicated here is a direct analogue to the scalar product-moment correlation.

Consider the correlation between n pairs of two dimensional vectors, represented here as complex numbers $\mathbf{z}_j$ and $\mathbf{w}_j$. The observations are

$$\mathbf{z}_j = x_j + iy_j; \quad j = 1, \ldots, n \tag{1a}$$

and

$$\mathbf{w}_j = u_j + iv_j; \quad j = 1, \ldots, n \tag{1b}$$

where $i = \sqrt{-1}$, and x_j and y_j are the components of the two-dimensional vector $\mathbf{z}_j$ and so forth. Variance and covariance then can be defined as

$$\sigma_z^2 = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{z}_j - \overline{\mathbf{z}})^*(\mathbf{z}_j - \overline{\mathbf{z}}) = \sigma_x^2 + \sigma_y^2 \tag{2}$$

and

$$\boldsymbol{\sigma}_{zw} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{z}_j - \overline{\mathbf{z}})^*(\mathbf{w}_j - \overline{\mathbf{w}})$$
$$= (\sigma_{xu} + \sigma_{yv}) + i(\sigma_{xv} - \sigma_{yu}) \tag{3}$$

where $\mathbf{z}_j^* = x_j - iy_j$ is the complex conjugate of $\mathbf{z}_j$. An overbar indicates the arithmetic mean. Variance of $\mathbf{z}_j$ is just the sum of its components' variances while the covariance is a vector in its own right (here represented as a complex number). The real portion (first component) of $\boldsymbol{\sigma}_{zw}$ is the sum of the covariances between corresponding elements of the variables. The imaginary portion (second component) measures the "twisting" of one vector's components into the opposite components of the other vector. Twisting implies a direction (from one vector into another) so vector covariance contains the asymmetry $\boldsymbol{\sigma}_{zw} = \boldsymbol{\sigma}_{wz}^*$ not found in scalar covariance.

A vector correlation measure defined as

$$\boldsymbol{\rho}_{zw} = \frac{\boldsymbol{\sigma}_{zw}}{\sigma_z \sigma_w} \tag{4}$$

forms a direct analogue to scalar correlation. Interpretation of this measure arises naturally from its relationship to a scalar regression, but in this case involving complex numbers.

Consider the regression equation

$$\mathbf{w}_j = \boldsymbol{\beta} \mathbf{z}_j + \boldsymbol{\alpha} + \boldsymbol{\epsilon}_j \tag{5}$$

with the complex regression coefficients $\boldsymbol{\beta} = b_0 + ib_1$ and $\boldsymbol{\alpha} = a_0 + ia_1$. Fit error associated with the jth observation is $\boldsymbol{\epsilon}_j$. A least-squares regression then requires that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ minimize $\sum_{j=1}^{n} \boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_j$. Expressions for the regression coefficients are

$$\boldsymbol{\beta} = \boldsymbol{\rho}_{zw} \frac{\sigma_w}{\sigma_z} \tag{6}$$

and

$$\boldsymbol{\alpha} = \overline{\mathbf{w}} - \boldsymbol{\beta} \overline{\mathbf{z}}, \tag{7}$$

and they are again exact analogues to the scalar case. Furthermore the squared magnitude of the correlation, $\boldsymbol{\rho}_{zw}\boldsymbol{\rho}_{zw}^*$ is the proportion of the variance in $\mathbf{w}_j$ explained by the regression equation. The magnitude of the correlation is in the range zero to one and is unaffected by linear transformations of either $\mathbf{z}_j$ or $\mathbf{w}_j$.

Correlation Under Rotation

Equation (5) implies a transformation that involves translation, rotation, and scaling of the vector $\mathbf{z}_j$. The "intercept" $\boldsymbol{\alpha}$ defines a translation of coordinates. Rotation and scaling appear clearly when $\boldsymbol{\beta}$ and $\mathbf{z}_j$ are recast in their polar forms, $\boldsymbol{\beta} = Be^{i\Theta}$ and $\mathbf{z}_j = Z_j e^{i\theta_j}$, where the magnitude and angle are

$$B = \sqrt{b_0^2 + b_1^2} \tag{8}$$

and

$$\tan \Theta = \frac{b_1}{b_0}. \tag{9}$$

Equation (5) can be rewritten

$$\mathbf{w}_j = BZ_j e^{i(\Theta+\theta_j)} + \boldsymbol{\alpha} + \boldsymbol{\epsilon}_j \tag{10}$$

where it is apparent that B is the scaling factor and Θ is the rotation angle from $\bar{\mathbf{z}}$ to $\bar{\mathbf{w}}$.

Vector correlation $\boldsymbol{\rho}_{zw}$ is related to the regression "slope" $\boldsymbol{\beta}$ by a real scalar (equation 6). Hence, the phase angle Θ can be extracted directly from the correlation,

$$\tan \Theta = \frac{\mathrm{Im}\ \boldsymbol{\rho}_{zw}}{\mathrm{Re}\ \boldsymbol{\rho}_{zw}} = \frac{\sigma_{xv} - \sigma_{yu}}{\sigma_{xu} + \sigma_{yv}}, \tag{11}$$

which is equvalent to equation (9). The phase Θ extracted from $\boldsymbol{\rho}_{zw}$ is the angle required to go from the orientation of $\bar{\mathbf{z}}$ to the orientation of $\bar{\mathbf{w}}$. Switching the order of subscripts requires the same angular travel, but in the opposite direction.

Vector correlation $\boldsymbol{\rho}_{zw}$ is itself a two-dimensional vector. While correlation and regression must be computed in Cartesian form, they are easier to understand in polar form. The magnitude of the correlation $|\boldsymbol{\rho}_{zw}|$ is a single number with an interpretation quite similar to Pearson's scalar correlation. It ranges from zero for no correlation to one for perfect correlation, and its square is a ratio of explained variance to a total variance. Neither the correlation magnitude nor the size scaling factor B retain a sign. Phase conveys information similar to the sign of a scalar correlation, but drawn from a much richer set of possibilities.

Correlation Under Reflection

Calculations discussed above can be slightly modified to obtain a correlation that describes a reflectional relationship between two sets of vectors. Covariance and correlation constructed on the conjugate of the independent variable consequently provide measures of the reflectional relationship. Taking the complex conjugates of a set of vectors produces no change in the variances of the set, $\sigma_z^2 = \sigma_{z^*}^2$. Covariance involving complex conjugates becomes

$$\boldsymbol{\sigma}_{z^*w} = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{z}_j^* - \bar{\mathbf{z}}^*)^*(\mathbf{w}_j - \bar{\mathbf{w}})$$

$$= (\sigma_{xu} - \sigma_{yv}) + i(\sigma_{xv} + \sigma_{yu}). \tag{12}$$

Equation (4) produces the reflectional correlation $\boldsymbol{\rho}_{z^*w}$ when $\boldsymbol{\sigma}_{z^*w}$ is used instead of $\boldsymbol{\sigma}_{zw}$. Reflectional correlation and covariance have the symmetry $\boldsymbol{\sigma}_{z^*w} = \boldsymbol{\sigma}_{w^*z}$. Changing the sign in the exponent of equation (10) and computing new regression parameters produces

$$\mathbf{w}_j = B'Z_j e^{i(\Phi-\theta_j)} + \boldsymbol{\alpha}' + \boldsymbol{\epsilon}_j. \tag{13}$$

(Obtaining the complex conjugate in polar coordinates merely requires changing the sign of the azimuth without modifying the magnitude.) Constructing the regression of $\mathbf{w}_j$ on $\mathbf{z}_j^*$ thus produces the coefficients $\boldsymbol{\beta}' = B'e^{i\Phi}$ and $\boldsymbol{\alpha}'$ of equation (13). As with the rotation, reflectional correlation magnitude expresses goodness-of-fit, and the magnitude of $\boldsymbol{\beta}'$ is the scaling factor. The azimuth of $\boldsymbol{\beta}'$ (Φ) indicates twice the axis of reflection—the direction of the line about which variations in the paried observations are approximately symmetric. Owing to the symmetry of the reflectional correlation, Φ keeps the same value regardless of whether $\mathbf{z}_j$ or $\mathbf{w}_j$ is the independent vector.

Comparison of Reflection and Rotation

Whether rotation or reflection is the better model can be determined by calculating both coefficients and comparing the results. A slightly simpler comparison may be derived from the correlation magnitudes. We define a rotation/reflection index

$$\xi = \sigma_{xu}\sigma_{yv} - \sigma_{xv}\sigma_{yu}. \tag{14}$$

When ξ is positive, rotation is a better model, whereas if ξ is negative, reflection is better. The relationship of the index to the two models is apparent in their explained-variance forms:

$$|\rho_{zw}|^2 = \frac{(\sigma_{xu})^2 + (\sigma_{yv})^2 + (\sigma_{xv})^2 + (\sigma_{yu})^2 + 2\xi}{\sigma_z^2\sigma_w^2} \quad (15)$$

and

$$|\rho_{z^*w}|^2 = \frac{(\sigma_{xu})^2 + (\sigma_{yv})^2 + (\sigma_{xv})^2 + (\sigma_{yu})^2 - 2\xi}{\sigma_z^2\sigma_w^2}\,. \quad (16)$$

While calculating this index a priori nets only a negligible decrease in the total number of calculations, the form of ξ illuminates some differences between reflection and rotation. Rotation, for instance, dominates when covariances among like components (i.e., real-to-real and imaginary-to-imaginary) are of the same sign while unlike components are of opposite sign.

A more geometric interpretation of the rotation/reflection index is possible when it is written as the determinant of the covariance tensor:

$$\xi = \det \boldsymbol{\sigma}_{ij} = \det \begin{pmatrix} \sigma_{xu} & \sigma_{xv} \\ \sigma_{yu} & \sigma_{yv} \end{pmatrix}. \quad (17)$$

The pattern of signs required to produce a positive or negative value of ξ corresponds to the pattern of signs in the rotation and reflection matrices

$$\begin{pmatrix} \cos\Theta & \sin\Theta \\ -\sin\Theta & \cos\Theta \end{pmatrix} \text{ and } \begin{pmatrix} \cos\Phi & \sin\Phi \\ \sin\Phi & -\cos\Phi \end{pmatrix}. \quad (18)$$

Covariances represent the inner product (dot product) of two n-dimensional vectors of deviations from the mean. Another interpretation of the inner product is as the product of the absolute values of the vectors (in this case, n times the standard deviations) and the cosine of the angle between them (in the n-dimensional space). It is thus no coincidence that the pattern of signs of the covariances required for ξ to indicate rotation or reflection is the same as the pattern of signs in the matrices of (18).

Similarity between equations (15) and (16) allows us to define a correlation measure as

$$\rho = \left(\frac{\xi}{|\xi|}\right) \sqrt{\frac{(\sigma_{xu})^2 + (\sigma_{yv})^2 + (\sigma_{xv})^2 + (\sigma_{yu})^2 + 2|\xi|}{\sigma_z^2\sigma_w^2}}\,. \quad (19)$$

Equation (19) selects ρ as the largest correlation of (15) and (16), and it gives the correlation a negative sign if the reflection model is superior. This correlation ranges between -1 and $+1$, where $+1$ implies a perfect rotational relationship, -1 implies a perfect reflectional relationship, and zero implies completely uncorrelated data with respect to both types of relationship. Our correlation also is invariant under scaling, rotation, reflection, and translation of the constituent vectors. Its square is the proportion of variance explained by the regression model.

Correlations among Upper-Air Wind Fields over the U.S.

To illustrate the use of vector correlation, we compare upper-level wind fields over the coterminous U.S. Wind data were drawn from the National Meteorological Center (NMC) Northern Hemisphere octagonal grid analyses. The NMC data then were interpolated from the octagonal grid to a 2° of latitude by 2° of longitude grid using a 16-point distance-weighted interpolation scheme yielding 208 grid points. A computer program for calculating the interpolated field was provided along with the data by the National Center for Atmospheric Research. Owing to the relatively smooth nature of upper-air wind fields, the interpolated field should be quite accurate. Our first example considers the correlation between the 850-mb and 500-mb winds at 1200 UTC (Universal Time Coordinate) on 9 February 1989. Our second case study examines winds at these same levels for 0000 UTC on 12 February 1987.

Reflectional Example

Upper-air wind conditions at 1200 UTC on 9 February 1989 were dominated by a deep low-pressure system over Baffin Island in northeastern Canada and a weaker center of low pressure just off the west coast of the U.S. Strong westerly and northwesterly winds in the eastern two-thirds of the country resulted from the steep pressure gradient and deep trough surrounding the polar low. Cold advection occurred throughout the region. In the western U.S., the secondary low-pressure system yielded warm advection from a more southerly and southeasterly wind regime. High pressure near the surface and centered over Texas led to noticeable veering (clockwise turning) of winds in the 850-mb–500-mb layer west of the system, and backing (counterclockwise turning) to its east.

Vector correlation between these 850-mb and 500-mb wind fields is, on the average, reflectional (Fig. 2). Approximately 60 percent of the variance is explained by the reflection model ($\hat{\rho} = -0.77$) and the axis of reflection is 129° (clockwise from North). The most influential patterns are (1) the clockwise turning of the 500-mb winds from the 850-mb winds in the western and north-central parts of the country, and (2) the counterclockwise turning in the East, particularly in the Southeast. A moderate, positive correlation among the speed components of the velocity fields is apparent (Fig. 3a) while the directional components are spatially segregated (Fig. 3b). Over the central and eastern portions of the country, the 500-mb wind directions are predominantly from the west or northwest. While there is more variability within the 850-mb field in these regions, the 850-mb winds also are from the west and northwest. Covariance between the 850-mb and 500-mb wind directions over the central and eastern portions of the U.S. produces a dense concentration of points on a direction scatterplot (Fig. 3b). A relationship between the southerly and southeasterly winds (at both levels) over the western part of the country also results in a characteristic but elongated swarm of points on the direction scatterplot. This swarm, however, is based on fewer grid points as well as on weaker and more variable winds; therefore it is less distinctive than the pattern over the central and eastern U.S. It is noteworthy that vector correlations can be interpreted only partially through a graphical examination of scalar components (Klink and Willmott 1989).

Rotational Example

Conditions at 0000 UTC on 12 February 1987 were controlled by a sharp ridge of high pressure centered over the northern Rocky Mountains and a strong low-pressure system over New Brunswick, Canada. The complementary positions of the ridge and the low center yielded strong northwesterly winds from the Great Lakes into the Northeast. In the West, upper-air winds were split around the ridge into two branches. A southwesterly wind regime was evident in the western part of the country. To the south, a westerly flow occurred. Only slight veering of winds in the 850-mb–500-mb layer was evident.

Vector correlation between the 850-mb (Fig. 4a) and 500-mb (Fig. 4b) winds on February 12 exhibits a strong rotational character ($\hat{\rho} = 0.84$). The 500-mb field is rotated on the average 16° clockwise from the 850-mb field. Strong westerly and northwesterly winds (at both levels) in the eastern half of the country dominate the correlation. Correspondence between high speed in the East and lower speeds in the West also is evident in a speed-component scatterplot (Fig. 5a). The contribution of the directional components is not as clear; but a relatively dense grouping of points (centered at about 90° on the direction scatterplot, Fig. 5b) suggests an important covariance among westerly and near-westerly winds at the two levels. Such maps and graphics are essential to an informed interpretation of vector correlation.

It is noted that the rotation is not necessarily produced by a consistent thermal wind pattern. A thermal wind that varied only slightly over the map would produce, in all cases, a high rotational correlation. High rotational correlation, conversely, does not always imply a consistent thermal wind pattern. Any correlation between wind vectors at different pressure levels will show a rotational correlation if the area examined is not large enough to contain regions of both veering and backing winds. Negative correlation, implying reflection, would thus indicate a situation of synoptic complexity.

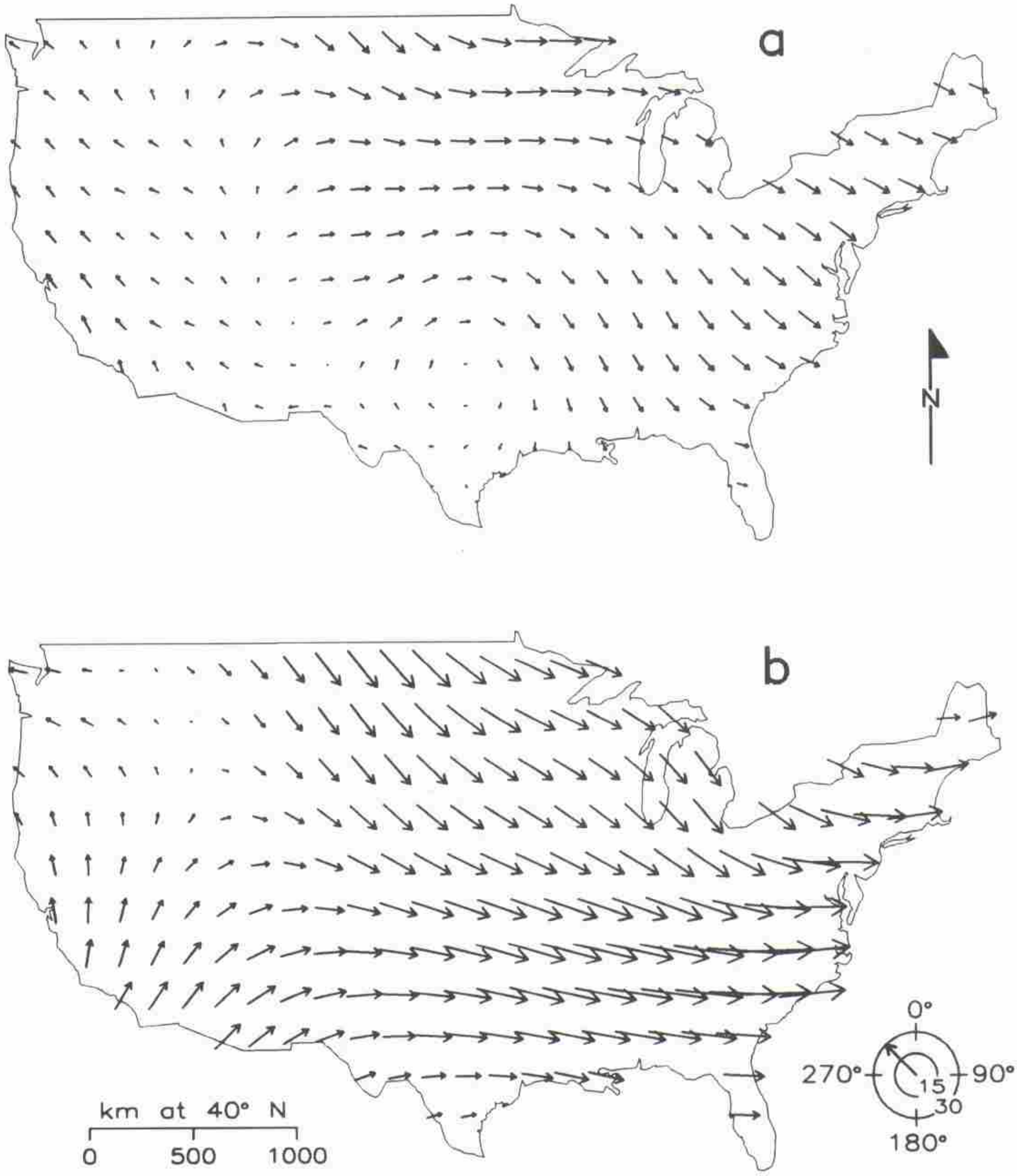

Figure 2. Upper-air wind vectors over the coterminous U.S. for 1200 UTC 9 February 1989 (Mercator projection). Each vector (arrow) is plotted over a point on the 2° by 2° grid. Arrow lengths are proportional to the wind-speed component (m s^{-1}) according to the scale at lower right: (a) winds at 850 mb, (b) winds at 500 mb.

Evaluating Reliability

Bootstrap methods (Efron and Tibshirani 1986) provide a means to estimate the natural variability expected in ρ. An empirical probability distribution for ρ can be generated by choosing n_B random samples of size n (with replacement) from our wind-field data set, also of size n. The correlation is then reevaluated for each of the n_B samples. An empiri-

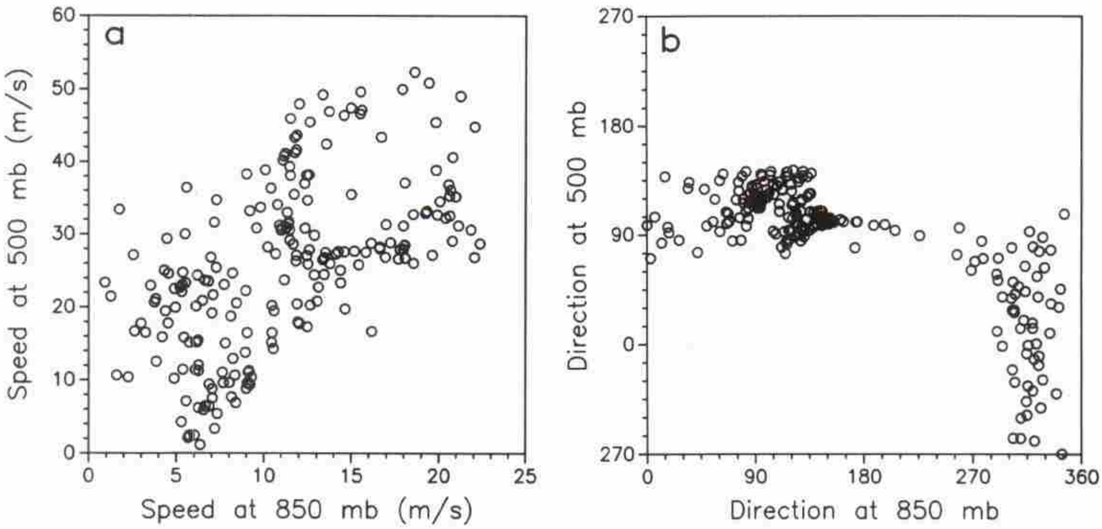

Figure 3. Scatterplots comparing polar components of the wind velocities shown in Figure 2: (a) speeds, (b) directions (in the meteorological coordinate system: angles are clockwise from North).

cally derived frequency distribution can be formed and evaluated using the n_B estimates of ρ. Usually, n_B is large.

Following Willmott et al. (1985a), 1000 bootstrap estimates of ρ were computed from $n_B = 1000$ random samples, each of size $n = 208$. A scatterplot of the bootstrap-derived vector correlation values for the 9 February 1989 fields (Fig. 6) indicates that $\hat{\rho}$ is centered about the initial correlation of −0.77. Magnitudes of our bootstrapped $\hat{\rho}$s vary little (±0.06) as do the axes of reflection (±8°). These bootstrap estimates suggest that our original evaluation of ρ is reliable. One thousand bootstrap estimates of ρ were similarly obtained for the 12 February 1987 fields (Fig. 7). Once again, the estimated values are centered roughly about the original correlation of 0.84. Magnitudes range from about 0.79 to 0.89 (±0.05) and rotation angles from about 10° to 22° (±6°).

Whether two data sets are derived from random sampling, simulation, or interpolation (e.g., to a regular grid), the correlation between them is a useful algebraic description of their similarity. When the data have not been randomly sampled, as is the case here, the usual statistical interpretations of reliability, confidence, or significance are inappropriate. A degree of spatial autocorrelation also exists within our (and most geographic) data, and this constitutes another interpretational difficulty since spatial autocorrelation can bias reliability estimates (Cliff and Ord 1981). The small ranges of our bootstrapped vector correlations then may imply greater reliabilities than actually exist.

Summary and Conclusions

A bivariate vector correlation measure has been developed, described, and applied to wind velocity fields to illustrate its use. This correlation characterizes the strength of a linear relationship between two variables whose observations are two-dimensional vectors. It takes into account the magnitudes of the vector observations as well as rotational and reflectional relationships between the two variables of interest. A few comparable correlations have appeared in the wider scientific literature, but they have not been adapted to geographic problems. A bootstrap method was used in conjunction with the wind-field examples to demonstrate a way to examine the correlation's reliability.

Geographers regularly examine direction, flow, and other vector quantities, but they rarely subject them to appropriate (vector) statistical methods. Our limited use of directional

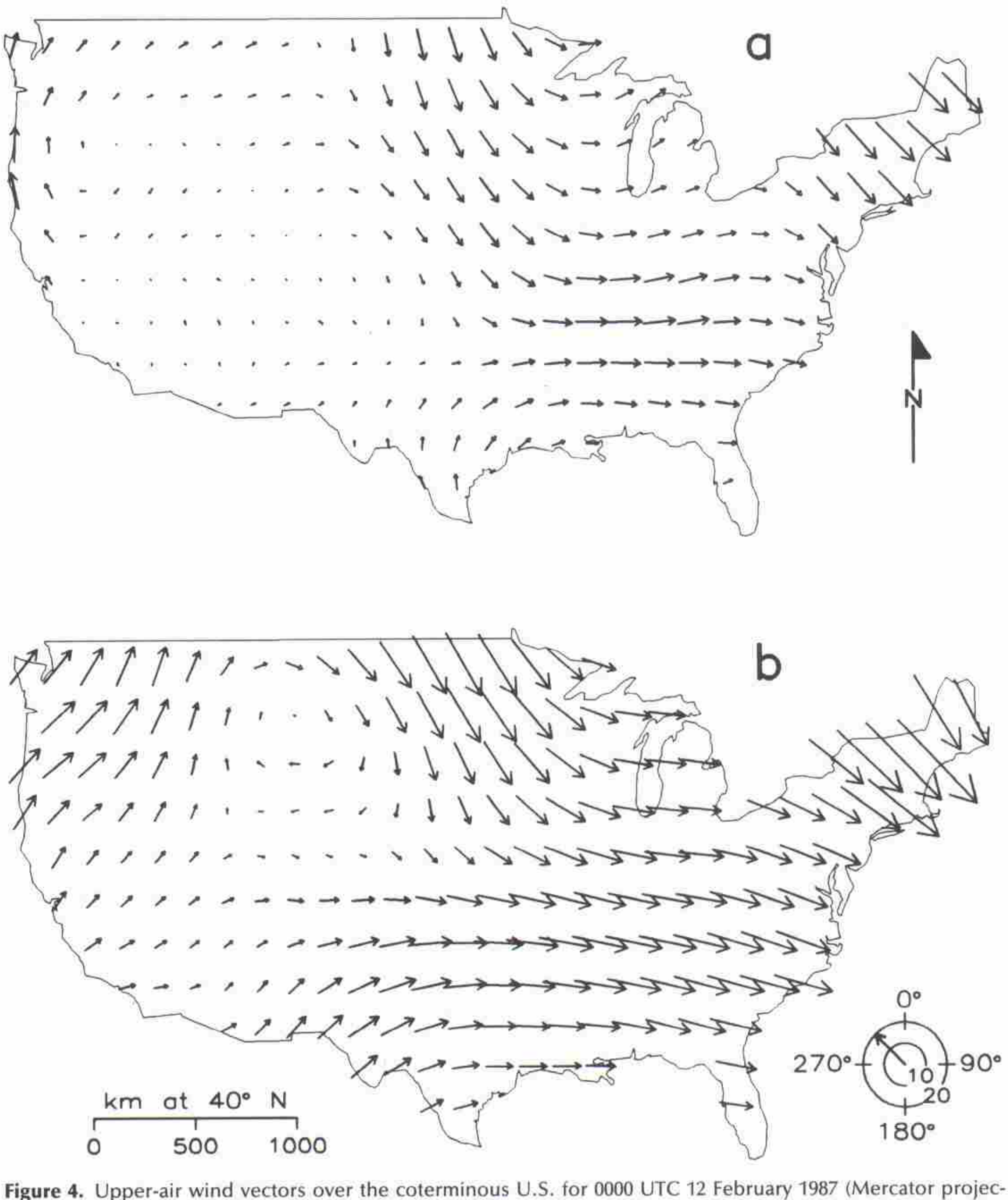

Figure 4. Upper-air wind vectors over the coterminous U.S. for 0000 UTC 12 February 1987 (Mercator projection). The same vector conventions are used as in Figure 2, except at a different magnitude scale: (a) winds at 850 mb, (b) winds at 500 mb.

and vector statistics reveals more about the relative unavailability of such methods than about the number of geographic problems that could make appropriate use of such statistics. Witness the lack of treatment of vector statistics and data within our standard geographical statistics texts and the unavailability of the necessary algorithms within standard statistical packages. A primary purpose of this paper has been to present vector correlation

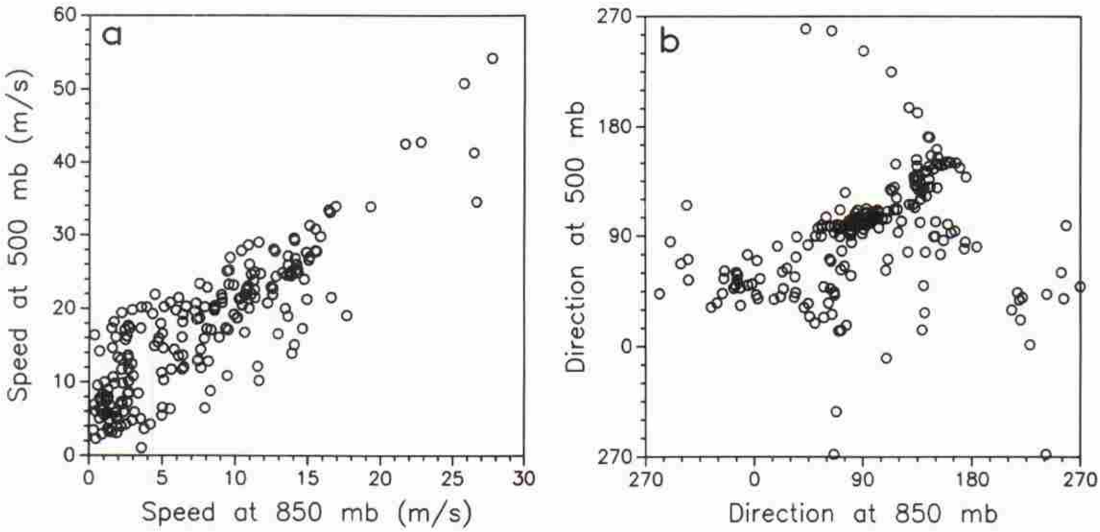

Figure 5. Scatterplots comparing polar components of the wind velocities shown in Figure 4: (a) speeds, (b) directions.

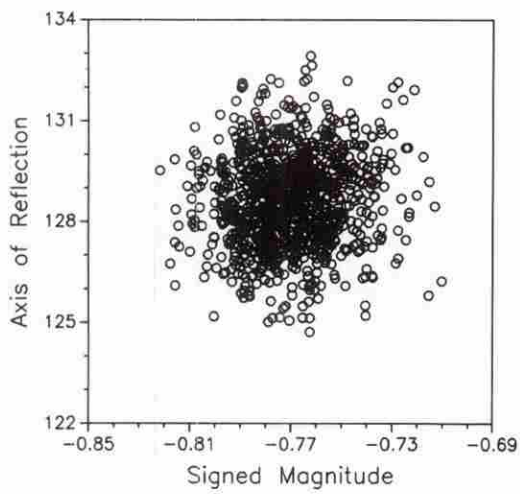

Figure 6. Bootstrapped estimates of the vector correlation between 850-mb and 500-mb winds for 1200 UTC 9 February 1989. Each of 1000 bootstrapped correlations is located according to signed magnitude (ρ) and reflection axes ($\Phi/2$).

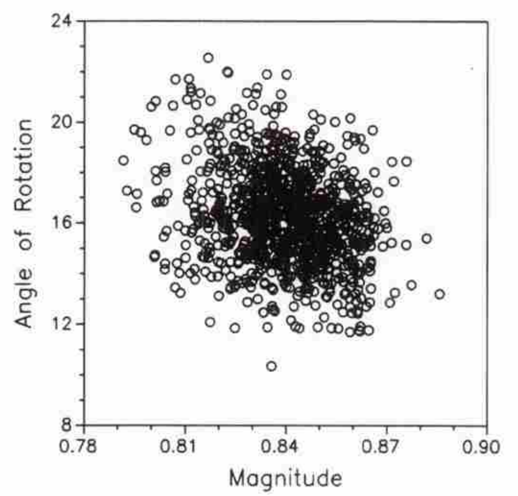

Figure 7. Bootstrapped estimates of the vector correlation between 850-mb and 500-mb winds for 0000 UTC 12 February 1987, similar to Figure 6 except the angles here represent rotation angles (Θ).

in such a way that it can be readily applied to a large number of geographic problems. An algorithm is given in the Appendix; a computer program is also available (see Appendix).

Vector correlation should find nearly as many uses within geography as scalar correlation has found. Consider that flows across the landscape can be represented as two-dimensional vectors as can the spatial gradients associated with most geographic fields.

Potential applications include comparisons of residential mobility patterns, where movers' origins and destinations define the endpoints of data vectors, and evaluations of the earth's magnetic field over time or space where the magnetic imprints that are 'frozen' in igneous rock at the time of solidification can be represented as vectors. Imagined journeys from one location to another on mental maps also could be represented as vectors and analyzed using vector correlation. Vectors additionally are used to represent the first harmonic of any climate variable's seasonal cycle (e.g., Willmott et al. 1985b) and could be used to compare interannual or spatial variability in the seasonal cycle.

Many useful extensions of bivariate vector correlation also may be conceived. Our presentation of vector regression could be extended easily to multiple regression using the same least-squares criterion. Typical data transformations applied in scalar regression may be challenged by the much fuller palette of transforms provided by complex functions. Geographers additionally may find it useful to develop and apply cross- and autocorrelation functions or eigenvector techniques based on vector correlation. The best extensions will help solve geographic problems that previously could not be analyzed. We are confident that geographers have many such problems in mind.

Acknowledgments

Annette Grilli, University of Delaware, provided considerable insight on and technical help with the visualization of vector data and correlation. James O'Donnell, University of Connecticut, assisted with the initial evaluation of several directional correlation statistics. Scott Webber, University of Delaware, assisted with upper-air wind data and the grid-interpolation program. The NMC octagonal grid data and associated interpolation program were obtained from the National Center for Atmospheric Research.

Appendix: Calculation of the Coefficients

Calculation of the various regression and correlation coefficients described on pp. 106–08 requires one pass through the data to collect the required sums. Component sums are obtained from

$$S_x = \sum_{j=1}^{n} x_j \ ; \quad S_y = \sum_{j=1}^{n} y_j \ ;$$

$$S_u = \sum_{j=1}^{n} u_j \ ; \quad S_v = \sum_{j=1}^{n} v_j \ . \tag{A.1}$$

The sums of squares are

$$S_{xx} = \sum_{j=1}^{n} x_j^2 \ ; \quad S_{yy} = \sum_{j=1}^{n} y_j^2 \ ;$$

$$S_{uu} = \sum_{j=1}^{n} u_j^2 \ ; \quad S_{vv} = \sum_{j=1}^{n} v_j^2 \ ; \tag{A.2}$$

while the between-vector cross-products are

$$S_{xu} = \sum_{j=1}^{n} x_j u_j \ ; \quad S_{xv} = \sum_{j=1}^{n} x_j v_j \ ;$$

$$S_{yu} = \sum_{j=1}^{n} y_j u_j \ ; \quad S_{yv} = \sum_{j=1}^{n} y_j v_j \ . \tag{A.3}$$

Intermediate calculations yield the variances:

$$\sigma_z^2 = \frac{S_{xx} + S_{yy}}{n} - \frac{S_x^2 + S_y^2}{n^2} \tag{A.4}$$

$$\sigma_w^2 = \frac{S_{uu} + S_{yy}}{n} - \frac{S_u^2 + S_v^2}{n^2} \tag{A.5}$$

and the four component covariances, σ_{xu}, σ_{xv}, σ_{yu}, and σ_{yv}, from

$$\sigma_{xu} = \frac{S_{xu}}{n} - \frac{S_x S_u}{n^2} \tag{A.6}$$

and so on. Users of computerized statistical packages should note that all of the preceding calculations can be replaced by treating the components x, y, u, and v as four independent variables and generating their means and their variance/covariance matrix.

The reflection/rotation coefficient ξ can be calculated using equation (14). A special sign variable, $s \equiv \text{sign}(\xi) = \xi/|\xi|$, helps to automate the remaining calculations:

$$\rho = s\sqrt{\frac{(\sigma_{xu})^2 + (\sigma_{yv})^2 + (\sigma_{xv})^2 + (\sigma_{yu})^2 + 2s\xi}{(\sigma_x^2 + \sigma_y^2)(\sigma_u^2 + \sigma_v^2)}} \tag{A.7}$$

$$B = s\rho\sqrt{\frac{\sigma_u^2 + \sigma_v^2}{\sigma_x^2 + \sigma_y^2}}$$

and

$$\Theta = \arctan\left(\frac{\sigma_{xv} - s\sigma_{yu}}{\sigma_{xu} + s\sigma_{yv}}\right). \tag{A.9}$$

Here B is the scale factor for either reflection or rotation and similarly Θ represents either Θ or Φ, depending on whether rotation or reflection dominates.

When coefficients for predicted values ($\hat{w}_j = \hat{u}_j + i\hat{v}_j$) are required, they are

$$b_0 = B \cos \Theta \tag{A.10}$$

$$b_1 = B \sin \Theta \tag{A.11}$$

$$a_0 = \frac{S_u - b_0 S_x + s b_1 S_y}{n} \tag{A.12}$$

$$a_1 = \frac{S_v - s b_0 S_y - b_1 S_x}{n} \tag{A.13}$$

$$\hat{u}_j = a_0 + b_0 x_j - s b_1 y_j \tag{A.14}$$

$$\hat{v}_j = a_1 + s b_0 y_j + b_1 x_j . \tag{A.15}$$

If one wishes to choose rotation or reflection regardless of which dominates, one may set $s = +1$ for rotation and $s = -1$ for reflection, regardless of the value of ξ. Equations (A.7) onward then will produce the desired result. A Fortran subroutine for making these calculations, and a sample program, can be obtained by writing to Professor Hanson, or preferably via electronic mail to Hanson at aey16590@udelvm on Bitnet or aey16590@udelvm.udel.edu on Internet.

Notes

1. The author list is given alphabetically to indicate an equal contribution from each author.

References

Batschelet, E. 1981. *Circular statistics in biology.* London: Academic Press.

Breckling, J. 1989. *The analysis of directional time series: Applications to wind speed and direction.* Lecture Notes in Statistics, Vol. 61. New York: Springer-Verlag.

Clark, W. A. V. 1971. A test of directional bias in residential mobility. In *Perspectives in geography 1: Models of spatial variation,* ed. H. McConnell and D. Yaseen, pp. 1–27. DeKalb, IL: Northern Illinois University Press.

———. 1972. Some vector representations of intraurban residential mobility. *International Geography,* (Proceedings of the I.G.U. Congress, Montreal, Canada) 2:796–98.

Cliff, A. D. and Ord, J. K. 1981. *Spatial processes: Models and applications.* London: Pion Limited.

Costanzo, C. M., and Gale, N. 1984. Evaluating the similarity of geographic flows. *The Professional Geographer* 36:182–87.

——— and ——— 1985. Flows, statistics and maps: A rejoinder. *The Professional Geographer* 37:58–59.

Davis, J. C. 1986. *Statistics and data analysis in geology,* 2nd ed. New York: John Wiley and Sons.

Downs, T. D. 1974. Rotational angular correlations. In *Biorhythms and human reproduction,* ed. M. Ferin et al., pp. 97–104. New York: John Wiley and Sons.

Efron, B., and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1:54–77.

Epp, R. J.; Tukey, J. W.; and Watson, G. S. 1971. Testing unit vectors for correlation. *Journal of Geophysical Research* 76:8480–83.

Fisher, N. I. et al. 1985. Spatial analysis of two-dimensional orientation data. *Mathematical Geology* 17:177–94.

——— and Lee, A. J. 1982. Nonparametric measures of angular-angular association. *Biometrika* 69:315–21.

——— **and** ———. 1983. A correlation coefficient for circular data. *Biometrika* 70:327–32.

——— **and** ———. 1986. Correlation coefficients for random variables on a unit sphere or hypersphere. *Biometrika* 73:159–64.

Gaile, G. L. 1990. Whither spatial statistics? *The Professional Geographer* 42:95–100.

——— **and Burt, J. E.** 1980. *Directional Statistics.*Norwich, CT: Geo Books (Concepts and Techniques in Modern Geography, no. 25).

Hillman, D. C. 1974. Correlation coefficients for ranked angular variates. In *Chronobiology,* ed. L. E. Scheving and F. Halberg, pp. 723–30. Tokyo: Igaku Shoin.

Johnson, R. A., and Wehrly, T. 1977. Measures and models for angular correlation and angular-linear correlation. *Journal of the Royal Statistical Society, Series B* 39:222–29.

Jupp, P. E., and Mardia, K. V. 1980. A general correlation coefficient for directional data and related regression problems. *Biometrika* 67:163–73.

Klink, K., and Willmott, C. J. 1985. Comments on "Evaluating the similarity of geographic flows." *The Professional Geographer* 37:56–58.

——— **and** ———. 1989. Principal components of the surface wind field in the United States: A comparison of analyses based upon wind velocity, direction, and speed. *International Journal of Climatology* 9:293–308.

Kundu, P. K. 1976. Ekman veering observed near the ocean bottom. *Journal of Physical Oceanography* 6:238–42.

Mardia, K. V. 1975. Statistics of directional data (with discussion). *Journal of the Royal Statistical Society, Series B* 37:349–93.

——— **and Puri, M. L.** 1978. A spherical correlation coefficient robust against scale. *Biometrika* 65:391–95.

Masuyama, M. 1939. Correlation between tensor quantities. *Proceedings, Physico-Mathematical Society of Japan, 3rd Series* 21:638–47.

Mooers, C. N. K. 1973. A technique for the cross spectrum analysis of pairs and complex-valued time series, with emphasis on properties of polarized components and rotational invariants. *Deep-Sea Research* 20:1129–41.

Odland, J.; Golledge, R. G.; and Rogerson, P. A. 1989. Mathematical and statistical analysis in human geography. In *Geography in America*, ed. G. L. Gaile and C. J. Willmott, pp. 719–45. Columbus, OH: Merrill Publishing Company.

Rivest, L.-P. 1982. Some statistical methods for bivariate circular data. *Journal of the Royal Statistical Society, Series B* 44:81–90.

Slocum, T. A. 1990. The use of quantitative methods in major geographical journals, 1956–1986. *The Professional Geographer* 42:84–94.

Stephens, M. A. 1979. Vector correlation. *Biometrika* 66:41–48.

Taylor, P. J. 1977. *Quantitative methods in geography: An introduction to spatial analysis.* Boston: Houghton Mifflin Co. (Reprinted in 1983 by Waveland Press, Inc.)

Tobler, W. R. 1978. Comparison of plane forms. *Geographical Analysis* 10:154–62.

Willmott, C. J., et al. 1985a. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90(C5):8995–9005.

———; **Rowe, C. M., and Mintz, Y.** 1985b. Climatology of the terrestrial seasonal water cycle. *Journal of Climatology* 5:589–606.

Wylie, D. P., et al. 1985. Autocorrelation of wind observations. *Monthly Weather Review* 113:849–57.

Submitted 9/90, revised 2/91, accepted 5/91.