

Review article: Modern methods of analysis for three-dimensional orientational data

Joshua R. Davis^{1*}, Sarah J. Titus²

¹Department of Mathematics and Statistics, Carleton College; jdavis@carleton.edu

²Department of Geology, Carleton College; stitus@carleton.edu

Abstract

Structural geology studies commonly include data about orientations of objects in space. By “orientation” we mean not just a single direction, such as a foliation pole or the long axis of an ellipsoid, but a complete three-dimensional orientation of a body such as a foliation-lineation pair, a fold, an ellipsoid, etc. Over the past four decades, researchers in various fields have developed theory and algorithms for dealing with such data. In this paper, we explain how to apply orientation statistics to common geologic data types. We review plotting systems, measures of location and dispersion, inference (confidence/credible regions and hypothesis tests) for population means, and regression. We pay special attention to methods that work for small sample sizes and widely dispersed data. Our original contributions include a concept of Kamb contouring for orientations, a technique for handling anisotropy in confidence/credible regions, and large-scale numerical experiments on the performance of various inference methods. We conclude with a detailed study of foliation-lineations from the western Idaho shear zone, using statistical results to argue that the data are not consistent with a published model for them.

Keywords: rotations, orientation statistics, visualization

1. Introduction

Orientation statistics is a set of concepts and techniques for analyzing rotations of three-dimensional space, and objects whose orientations in space can be described by such rotations. The subject has been developed and applied over the past four decades, in diverse fields such as statistics (e.g., Downs, 1972; Prentice, 1986), medical imaging (Fletcher et al., 2004), biomechanics (Rancourt et al., 2000), robotics (Krakowski et al., 2007), computer vision (Govindu, 2004), and astronomy (Mardia and Jupp, 2000, Chapter 13). In mineralogy, orientation statistics is used to analyze crystallographic orientations (e.g., Matthies et al., 1988; Bachmann et al., 2010), while in seismology it is used to analyze earthquake focal mechanisms (?).

Our opinion is that orientation statistics is an essential tool for structural geology as well. For example, here are four geologic problems that cannot be answered satisfactorily without such a theory (Fig. 1).

- Giorgis and Tikoff (2004) modeled the deformation of the western Idaho shear zone as homogeneous monoclinic transpression along a NS-striking, vertical shear plane. Among their data were the foliation-lineation pairs of Fig. 1A. Are these data consistent with the foliation-lineation orientations predicted by the proposed deformation? (With 95% credibility, the answer is no. See Section 7.)

- Fig. 1B shows axial planes and hinge lines for a synthetic data set of cylindrical folds. Are there any outliers in the data? (Yes, although they are not apparent in this plot. See Section 3.)

- Suppose that a geologist wishes to perform numerical experiments about how rock fabric develops. Her plan is to begin with a set of ellipsoidal clasts of equal shape but random orientation (Fig. 1C), and then simulate their deformation in various flow types. How does she generate the uniformly random initial orientations? How does she quantify the distribution of orientations after deformation? (See Section 4.4.)

- Fig. 1D shows rotations deduced from paleomagnetic data from the Troodos massif, Cyprus, colored from west (red) to east (magenta). The approximate rainbow pattern suggests that rotation depends systematically on easting. How do we quantify that tendency? Is it statistically significant or merely a result of random variation? (Yes, it is significant. See Sections 3.4 and 6.3.)

To address these and other geologic questions about orientations, this paper summarizes and compares techniques for several fundamental statistical problems: data visualization (Section 3), describing the location and dispersion of a data set (Section 4), confidence regions and hypothesis tests (Section 5), and regression (Section 6). Special attention is paid to non-asymptotic methods suitable

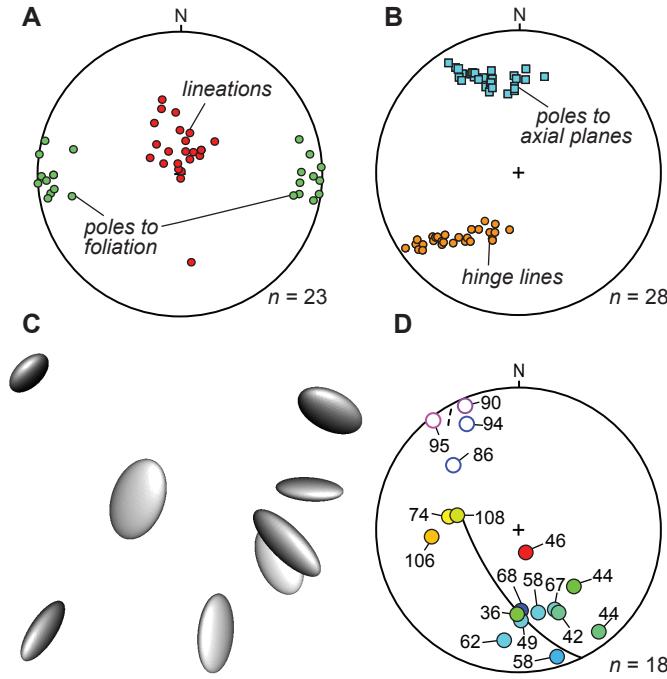


Figure 1: Example situations where orientation statistics applies. A. Western Idaho shear zone foliation-lineations (Giorgis and Tikoff, 2004). B. Fold axial planes and hinge lines, with a hidden outlier. C. Randomly generating ellipsoid orientations. D. Rotations inferred from paleomagnetic data, shown as axes of rotation and right-handed angles of rotation about those axes, in degrees (Titus et al., 2015). They are colored by distance from a graben. The curve is a proposed fit to this geographic tendency.

for the small, dispersed data sets typical of geologic field studies. Our original contributions to the methodology include a concept of Kamb contouring for orientational data, a technique for handling anisotropy in percentile confidence/credible regions, and numerical experiments on the performance of various inference methods, for data sets of varying size and dispersion.

In synthetic and natural examples, we apply orientation statistics to a wide variety of geologic data: foliation-lineation pairs, folds, slickensides, earthquake focal mechanisms, tectonic plate motions, etc. We conclude with one substantial application, in which we evaluate a published model of deformation in the western Idaho shear zone (Section 7). An extensive appendix summarizes the mathematical and statistical background material. Free software tools, including interactive versions of the figures, are available from the first author.

2. Geologic data as rotations

2.1. Motivation

Geologic field studies commonly include measurements of foliations as strike-dip pairs. In principle such data could be analyzed as pairs of numbers using two-dimensional statistics, but many complications arise. For example, strikes 000° and 360° are numerically different

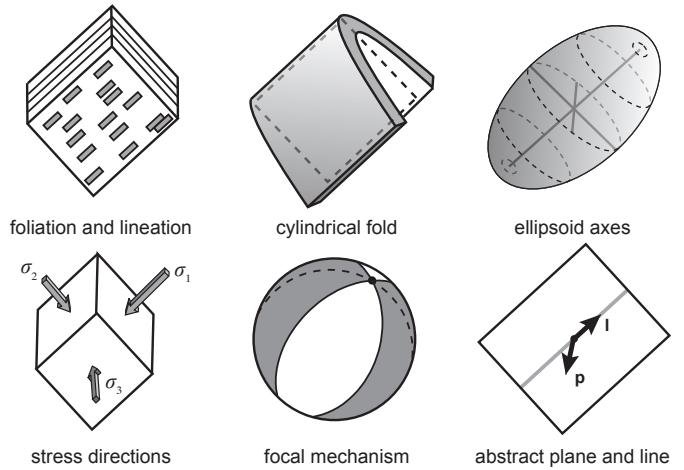


Figure 2: Five examples of orientational data types that can be viewed abstractly as lines in planes.

but physically identical. Two foliations with strikes 040° and 125° are more similar if both have dip 05° than if both have dip 80° . For these and other reasons, foliation planes are typically expressed as poles, which can be analyzed using methods of directional statistics such as the Watson and Bingham distributions (e.g., Mardia and Jupp, 2000; Borradaile, 2003; Tauxe, 2010).

When foliations are paired with lineations, it is tempting to treat each of these data types using directional statistics, but this approach leads to various problems (e.g., Borradaile, 2003, Section 10.4). For example, the average of a set of foliation poles may not be perpendicular to the average of their corresponding lineations. Also, plotting foliations and lineations separately can obscure important features of a data set (Section 3.1). Fundamentally, the fact that foliations and their associated lineations are perpendicular implies that they are correlated.

Our approach is to convert foliation-lineation pairs, and similar geologic data types, into rotations. The rotations are then treatable by orientation statistics, which enjoys a large literature. This approach is complicated somewhat by the fact that many geologic data types are not simply rotations but rather sets of rotations exhibiting certain symmetries. In this section, after discussing several classes of examples, we give an overview of how orientation statistics is applicable to such data types.

2.2. Lines within planes

Our first class of examples includes data types characterized by a plane and a line within that plane, such as foliation-lineation pairs (Fig. 2). Another example is cylindrical folds, whose orientations can be expressed as axial planes and hinge lines within those planes. A third example is triaxial ellipsoids, where the directions of the long, intermediate, and short axes are known, but their lengths are unknown or can be ignored. The complete orientation of such an ellipsoid can be specified as the plane containing the long and intermediate axes, and the long axis within

that plane. The same idea applies to stress orientations, if the principal stress directions and their order are known. A final example is earthquake focal mechanisms (Kagan, 2013). Given a focal mechanism, consider the line along which its nodal planes intersect, as well as the plane containing that line that bisects the extensional regions. In this way, a focal mechanism corresponds to a plane and a line within that plane.

(In each of these examples we have adopted a convention for abstracting geologic data into plane-line pairs. The particular convention chosen is not important, as long as it establishes a smooth one-to-one correspondence between the data type and plane-line pairs, and it is applied consistently.)

For a given plane and line in that plane, let \mathbf{p} be the unit pole to the plane and \mathbf{l} a unit vector along the line. Working within some chosen $[x \ y \ z]^\top$ coordinate system, form a rotation matrix \mathbf{R} whose rows are \mathbf{p} , \mathbf{l} , and $\mathbf{p} \times \mathbf{l}$ respectively. Geometrically, \mathbf{R} is the rotation that takes the plane to the y - z -plane and the line to the y -axis.

Inconveniently, there is some ambiguity in the conversion from lines-within-planes to rotations. Because \mathbf{p} can be negated without changing the plane that it describes, the matrix with rows $-\mathbf{p}, \mathbf{l}, -\mathbf{p} \times \mathbf{l}$ is an equally suitable representative rotation. Because \mathbf{l} can also be negated, so is the matrix with rows $\mathbf{p}, -\mathbf{l}, -\mathbf{p} \times \mathbf{l}$, and so is the matrix with rows $-\mathbf{p}, -\mathbf{l}, \mathbf{p} \times \mathbf{l}$. For $k = 1, 2, 3$, let \mathbf{J}_k be the rotation about the k th coordinate axis by $\pi = 180^\circ$:

$$\mathbf{J}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{J}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \text{ etc.}$$

Let \mathbf{I} be the 3×3 identity tensor and $G = \{\mathbf{I}, \mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3\}$. Then

$$GR = \{\mathbf{I}, \mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3\}\mathbf{R} = \{\mathbf{R}, \mathbf{J}_1\mathbf{R}, \mathbf{J}_2\mathbf{R}, \mathbf{J}_3\mathbf{R}\}$$

is the set of four rotations representing the given plane-line pair.

Fig. 3A-D depicts these rotations for a synthetic example. In each row of the figure, a plane-line pair rotates from the same observed orientation into the same reference orientation, along a different steady rotational path. The axis of rotation is shown, scaled by the total amount of rotation required about that axis.

In other words, G is a group of symmetries that acts on rotations without changing their meaning as plane-line pairs. It is analogous to a crystallographic point group, which rotates and reflects a crystal lattice into orientations that are physically indistinguishable from a given starting orientation. In fact, this G coincides with the rotational part of the point group of olivine.

2.3. Rays within planes

Similarly, consider a planar fault, where the direction of slip along the fault is known (Fig. 4A). The slip direction could be inferred from slickenside striae or estimates

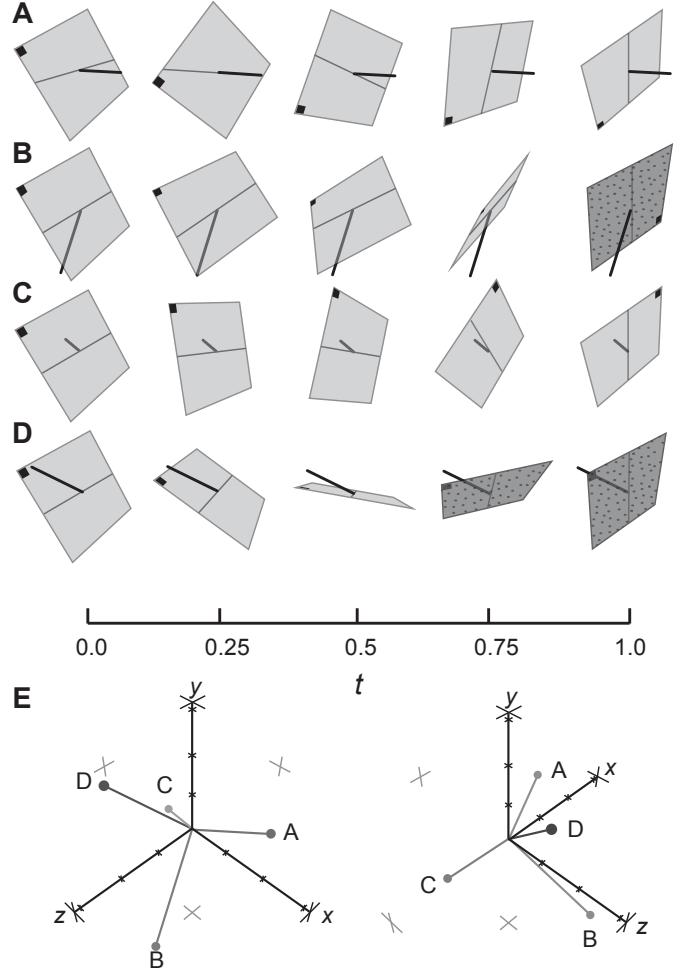


Figure 3: The four-to-one correspondence between lines-within-planes and rotations. A-D. Each row is a five-frame animation of a steady rotation that takes the given plane-line pair to the reference orientation. The axis of the rotation is superimposed, with length proportional to the total amount (angle) of rotation. E. Two views of the angle-axis plot of those four rotations (Section 3.2).

of offset on multiple outcrop faces. Geometrically, this situation can be described as a fault plane and a vorticity vector within that plane. For example, in coordinates $[x \ y \ z]^\top$ aligned with east, north, and up, a vertical, EW-striking strike-slip fault has fault pole $\mathbf{p} = \pm[0 \ 1 \ 0]^\top$ and vorticity vector $\mathbf{w} = [0 \ 0 \ -1]^\top$ for dextral motion or $\mathbf{w} = [0 \ 0 \ 1]^\top$ for sinistral motion.

(In geologic data sets, it is more common to express this kind of orientation as a fault pole \mathbf{p} and a movement direction \mathbf{h} of the hanging wall, for example through strike, dip, and rake. However, this convention is undefined for vertical faults. When presented with a fault orientation in terms of \mathbf{p} and \mathbf{h} , we first ensure that \mathbf{p} is downward-rather than upward-pointing, and then compute $\mathbf{w} = \mathbf{h} \times \mathbf{p}$. See Fig. 4A.)

Given a unit pole \mathbf{p} to the fault plane and a unit vorticity vector \mathbf{w} in that plane, form a matrix \mathbf{R} with rows \mathbf{p} , \mathbf{w} , and $\mathbf{p} \times \mathbf{w}$. Notice that \mathbf{p} can be negated without changing

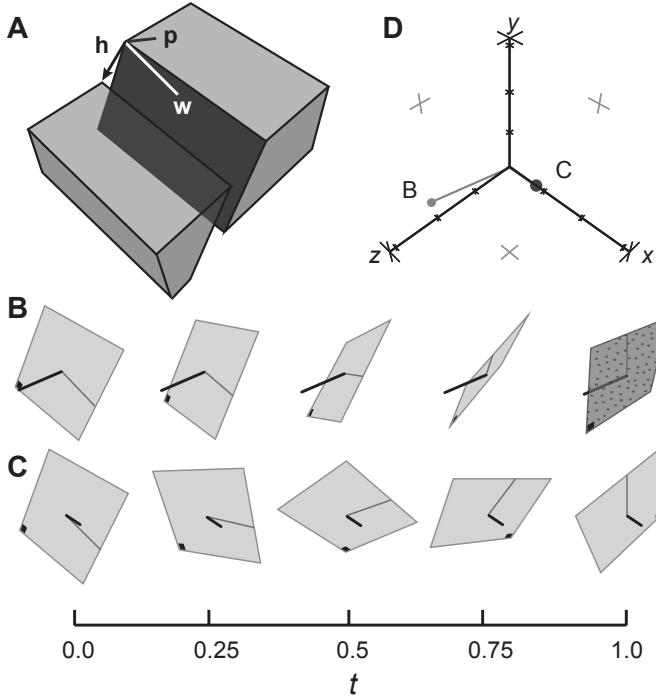


Figure 4: Faults with slip directions exhibit ray-in-plane symmetry. A. Block diagram of an oblique-normal fault, with pole \mathbf{p} , hanging wall motion direction \mathbf{h} , and vorticity direction \mathbf{w} . B-C. Two animations of the plane-vorticity pair rotating into a reference orientation. D. The angle-axis plot of those two rotations (Section 3.2).

its physical meaning, but \mathbf{w} cannot. Thus the fault-with-slip-direction amounts to a plane and a ray (directed line) within that plane. The two rotations that represent this plane-ray pair are \mathbf{R} and $\mathbf{J}_2\mathbf{R}$, and the symmetry group is $G = \{\mathbf{I}, \mathbf{J}_2\}$.

Fig. 4BC shows a synthetic example. In each row, a plane-ray pair rotates from the same observed orientation (that of Fig. 4A) into the same reference orientation, along two different steady paths. As in Fig. 3A-D, the rotation axis is shown, scaled by the amount of rotation about that axis.

2.4. Crystallographic orientations

Microstructural data often include crystallographic orientations, perhaps obtained through U-stage microscopy or electron backscatter diffraction (EBSD). Each orientation corresponds to a set of rotations related by the crystallographic point group of the mineral.

For example, the orientation of a quartz lattice can be described using three orthonormal vectors equivalent to Miller indices [100], [010], [001]. Placing these vectors into the rows of a matrix produces a rotation \mathbf{R} , which is subject to trigonal trapezohedral symmetry. Let

$$\mathbf{K} = \begin{bmatrix} \cos(120^\circ) & -\sin(120^\circ) & 0 \\ \sin(120^\circ) & \cos(120^\circ) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

be rotation about the z -axis of space through $2\pi/3 = 120^\circ$. In this notation the trigonal trapezohedral point group is

$G = \{\mathbf{I}, \mathbf{K}, \mathbf{K}^2, \mathbf{J}_1, \mathbf{KJ}_1\mathbf{K}^2, \mathbf{K}^2\mathbf{J}_1\mathbf{K}\}$, and the six rotations $G\mathbf{R}$ are equivalent to \mathbf{R} .

2.5. Rotations themselves

In some instances, geologists wish to analyze actual rotations, rather than orientations of rigid objects. Unlike in the preceding examples, there is no four-, two- or six-fold symmetry relating the data to rotations. The symmetry group $G = \{\mathbf{I}\}$ is trivial.

For example, in Section 6.4 we analyze the relative motion of the Farallon and Pacific tectonic plates (Engebretson et al., 1984), which is expressed as a sequence of rotations about varying Euler poles. Also, in Sections 3.4 and 6.3 we analyze a set of rotations inferred from paleomagnetic directions combined with dike orientations. Finally, some data sets can be treated as rotational, even though they are subject to symmetry, because the data are so tightly concentrated that the symmetry has negligible effect (Section 2.7).

2.6. Non-examples

To clarify the scope of this paper, we now list some data types that are not treatable by its methods. A key concept is the number of degrees of freedom, which equals the dimension of the underlying sample space. Rotations of three-dimensional space have three degrees of freedom, such as three Euler angles or the three components of a vorticity vector (Appendix A). They form a three-dimensional mathematical space called the *special orthogonal group* and denoted $\text{SO}(3)$ (Appendix B). If the geologic data under consideration do not have exactly three degrees of freedom, then they cannot be viewed as rotations.

For example, foliations have two degrees of freedom, such as the strike and dip of the foliation plane or the trend and plunge of its pole vector. Foliations are well treated by directional statistics on their pole vectors (e.g., Mardia and Jupp, 2000). The same remarks apply to lineations. On the other hand, foliations paired with lineations have three degrees of freedom, such as strike, dip, and rake, and are easily viewed as rotations.

Similarly, paleomagnetic directions have two degrees of freedom, such as trend and plunge, and are subject to directional statistics (Borradaile, 2003; Tauxe, 2010). There are infinitely many rotations \mathbf{R} that return a paleomagnetic direction to a reference direction. If one wishes to determine rotations from paleomagnetic directions with greater specificity, then additional information is required. For example, if sedimentary bedding is present, then one might require \mathbf{R} to restore that bedding to horizontal. If dikes are present, then one might require \mathbf{R} to restore them to vertical (Allerton and Vine, 1987). Or one might simply assume that the smallest possible rotation has occurred. See Sections 3.4 and 6.3.

Several concepts in structural geology are expressed as ellipsoids: finite strain, shape preferred orientation,

anisotropy of magnetic susceptibility, and so on. Ellipsoids have six degrees of freedom. Intuitively, three of them are contained in an ellipsoid's orientation, which is subject to line-in-plane symmetry (Section 2.2). However, the orientation is well-behaved only when the ellipsoid is triaxial. When two of the axes are equal in length, the ellipsoid's orientation is determined by the direction of the other axis, which has only two degrees of freedom. Therefore orientation statistics is poorly suited to analysis of ellipsoid data sets that contain nearly spheroidal ellipsoids. The other three degrees of freedom in an ellipsoid arise from its shape: the lengths of its three axes, or the volume, octahedral shear strain E_s , and Lode's parameter ν . Despite their having the correct number of degrees of freedom, ellipsoid shapes are not sensibly treatable as rotations. (Mathematically, the sample space is a three-dimensional space unrelated to $SO(3)$.)

Finally, rotations of two-dimensional space have only one degree of freedom: the angle of rotation about an axis perpendicular to the space. They are treatable by directional statistics in two dimensions, which is also called *circular statistics* (Mardia and Jupp, 2000). This paper treats rotations of three-dimensional space only, although some of the methods have been developed in all dimensions (e.g., Downs, 1972).

2.7. Strategy

By converting our data into rotations, we gain access to an enormous body of techniques already developed in statistics and other disciplines. However, some care is required in applying these techniques to structural geology data, as we now explain.

Orientation statistics is necessarily a blend of two mathematical influences, geometry and probability, although some authors or disciplines emphasize one over the other (Mardia and Jupp, 2000; Stanfill et al., 2013). In differential geometry of curved spaces, distance is measured along special curves called *geodesics*. For example, on a sphere the geodesics are the great circles, and the distance between two points is measured along a great circle. In the space $SO(3)$ of rotations, the geodesic between two rotations \mathbf{R} and \mathbf{Q} is a steady progressive rotation, starting at \mathbf{R} and ending at \mathbf{Q} , and the distance $d(\mathbf{R}, \mathbf{Q})$ is the rotation angle of the net finite rotation \mathbf{QR}^\top (Appendix C). At the same time, the space of rotations is equipped with a well-behaved notion of integration (Appendix D), in terms of which probability distributions can be defined. For example, there are several competing notions of normal distribution (e.g., Matthies et al., 1988; Nikolayev and Savelyeva, 1997; Bunge, 2013; Qiu et al., 2014). These concepts of distance and integration, and hence many the statistical methods that depend on them, can be extended to orientations with any symmetry group G .

In general statistics, large data sets are often easier to analyze than small ones, because they allow the use of asymptotic methods such as the central limit theorem. In orientation statistics, asymptotic methods also make

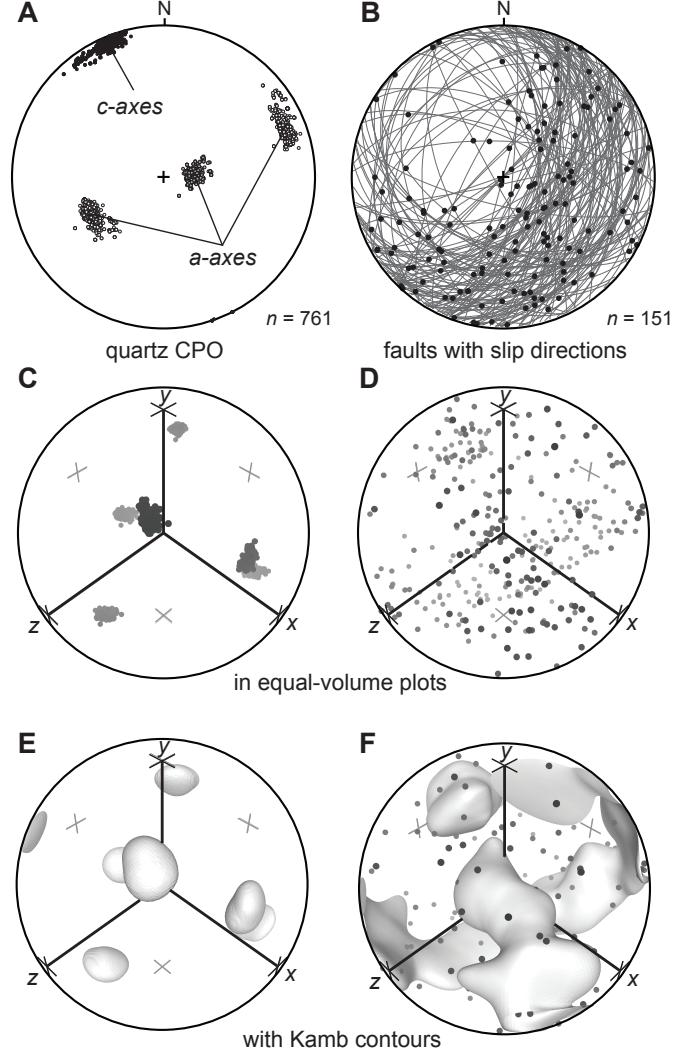


Figure 5: Structural geology data sets may be smaller and more widely dispersed than orientation data sets in other fields, such as mineralogy. On the left, 761 crystallographic orientations from a single quartz grain. On the right, 151 fault-with-slip orientations. From top to bottom, the plot types are equal-area, equal-volume (Section 3.3), and equal-volume with 6σ Kamb level surface (Section 3.5).

tightly concentrated data sets easier than widely dispersed ones. Unfortunately, structural geology data sets are often small and dispersed, compared to orientation data sets in other disciplines. For example, Fig. 5A shows 761 crystallographic orientations measured by EBSD in a single quartz grain from the Moine thrust zone, Scotland (Strine and Wojtal, 2004; Michels et al., 2015). The orientations are tightly concentrated, partially because a grain boundary was declared wherever adjacent EBSD measurements differed by at least 10° in orientation. For comparison, Fig. 5B shows 151 slickenside orientations that we have measured in the Troodos ophiolite, Cyprus. Fig. 5CD show the same data in the equal-volume rotation plot described below (Section 3.3), with each quartz orientation shown six times and each slickenside shown twice, because

of symmetry. The equal-area plots suggest, and the equal-volume plots make clear, that the slickensides are much more widely dispersed than the quartz orientations. The foliation-lineation pairs of Fig. 1A and the paleomagnetic rotations of Fig. 1D are also few and dispersed. Guided by such examples, this paper pays special attention to methods that work for small sample sizes and wide dispersion.

In fact, some data sets are so tightly concentrated that they can be analyzed as rotations, ignoring symmetry, without any practical harm. For example, in the quartz orientations of Fig. 5AC, the distances within each symmetric copy are much smaller than the distances among the symmetric copies. Consequently, distance-based methods effectively operate within one of the symmetric copies, as if the orientations were simply rotations. Similarly, if one imagines a probability distribution inferred from each symmetric copy, then the tails of these distributions overlap negligibly. Isolating one symmetric copy and truncating its corresponding distribution does not appreciably affect a distribution-based method. While describing the analysis of EBSD data, Bachmann et al. (2010) made similar remarks before concluding that symmetry can often be ignored. In contrast, there is no obvious way to isolate the symmetric copies of the slickensides in Fig. 5BD. Therefore this paper’s plots, hypothesis tests, and other techniques are designed to work with symmetry by default. In examples, we ignore symmetry only after checking that doing so is reasonable. However, we often explain statistical concepts in the simplified context of pure rotations without symmetry, for clarity of exposition. Details are left to Appendix J.

3. Plots

3.1. Motivation

Returning to an example from Section 1, recall that Fig. 1B shows a synthetic data set of 28 fold orientations. The same data are shown on the left side of Fig. 6A, with each line drawn on its corresponding plane. The data seem to form an elongated cloud with no outliers. To investigate further, we convert each fold orientation into a set of four rotations, following Section 2.2. Each rotation is expressible as three Euler angles: a rotation through an angle α about the x -axis, followed by a rotation through β about the z -axis and then a rotation through γ about the x -axis. (This is one of several conventions for Euler angles. All have similar theoretical properties. See Appendix A.) The left side of Fig. 6B is a still image of an interactive three-dimensional plot of these Euler angle triples. An outlier is now apparent.

Where is this outlier in Fig. 1B? Ignoring the great circles, the left side of Fig. 6A is the same plot, with some foliation-lineation pairs shaded white, some shaded gray, and one shaded black. The black datum is revealed as an outlier, because its axial plane is similar to the white planes but its hinge line is similar to the gray lines. What

makes this one fold unusual is neither its hinge line nor its axial plane but rather the relationship between them. This relationship is neglected entirely in Fig. 1B and obscured greatly in the left side of Fig. 6A.

The fundamental problem here is that rotations and orientations have three degrees of freedom and hence form a three-dimensional space (Section 2.6 and Appendix B). Any two-dimensional plot of such a space must suffer a loss of information. Interactive three-dimensional plots can help the geologist visualize her data more faithfully.

However, even three-dimensional plots must make compromises. Perhaps the Euler angle plot’s greatest defect is the one called *gimbal lock* (e.g., Koks, 2006, p. 180). When the middle Euler angle β is a multiple of $\pi = 180^\circ$, the rotation depends not on α and γ separately, but rather on $\alpha \pm \gamma$. Consequently, entire lines along the $\beta = 0$ and $\beta = \pi$ boundary planes in the Euler angle plot represent a single rotation, and the plot is extremely distorted near these lines. The right side of Fig. 6B shows an example with a synthetic data set of 17 tightly concentrated quartz orientations. Each orientation appears six times, due to trigonal trapezohedral symmetry (Section 2.4). Four of the symmetric copies of the data appear tightly concentrated, as is correct. However, the other two copies fall near the $\beta = 0$ plane and appear “smeared out” along the gimbal lock lines. Despite such well-known flaws (e.g., Frank, 1988), the Euler angle plot is common in the literature (e.g., Krieger Lassen et al., 1994; Randle and Engler, 2000).

Designing a plotting system for rotational data is therefore a subtle problem. The ideal plot would be three-dimensional, with a smooth, one-to-one correspondence between rotations and points in the plot. Information about rotations and the orientations that they represent would be easily readable. Unlike the Euler angle plot, the plot would distort the space of rotations minimally. It would have an equal-angle property for accurately depicting the shapes of data sets, like the equal-angle hemispherical projection. It would also have an equal-volume property for accurately depicting the density of data, like the equal-area hemispherical projection.

No one plotting system can fulfill all of these desiderata. In our opinion, the equal-angle and equal-volume rotation plots described in this section come closest to the ideal for structural geology applications. They are also easily understandable as higher-dimensional analogues of the equal-angle and equal-area hemispherical plots. In this section, we also develop a notion of Kamb density contouring for rotational data.

3.2. Angle-axis plot

We begin by developing a precursor to the equal-angle and equal-volume plots. Any rotation of three-dimensional space can be expressed as a rotation about an axis \mathbf{u} through an angle α , where \mathbf{u} is a unit vector and $0 \leq \alpha \leq \pi = 180^\circ$ (Appendix A). The scaled vector $\alpha\mathbf{u}$ describes both the axis and the angle. We plot this vector $\alpha\mathbf{u}$ as a

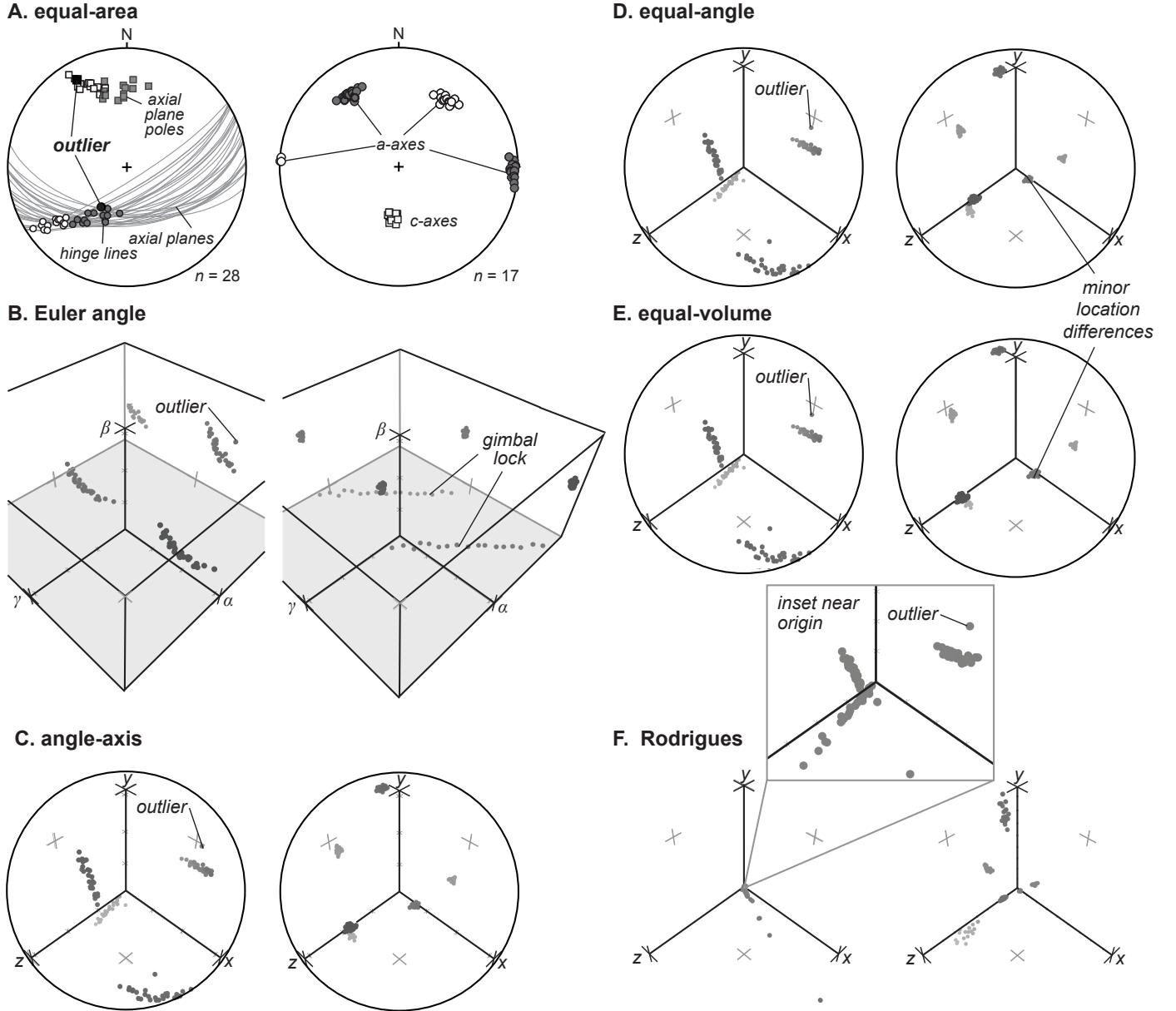


Figure 6: Comparison of orientation plotting systems. In each panel, the left side shows the synthetic data set of 28 fold orientations from Fig. 1B. The right side shows a synthetic data set of 17 quartz orientations. Six styles of plot are shown: equal-area, Euler angle, angle-axis, equal-angle rotational, equal-volume, and Rodrigues.

three-dimensional point in the usual manner, by placing the tail at the origin and plotting where the head falls. This system for plotting rotations is commonly called the *angle-axis plot* (Frank, 1988; Altmann, 2005; Bachmann et al., 2010; Jabłoński, 2011).

For example, Fig. 3E is two views of an angle-axis plot, showing the four rotations from Fig. 3A-D. The left viewpoint is identical to the viewpoint used in A-D. In essence, we have simply moved the four rotation axes to the origin of a new plot and recorded their endpoints. Similarly, Fig. 4D shows the two rotations from Fig. 4BC. For two more examples, Fig. 6C shows the same fold orientations and quartz orientations as in Fig. 6AB. On the left, the

outlier is clearly visible. On the right, all six symmetric copies of the quartz orientations appear nearly correctly concentrated.

In the angle-axis plotting system, the space $\text{SO}(3)$ of rotations plots as a solid ball of radius π . Each point in the interior of the ball corresponds to one and only one rotation (α, \mathbf{u}) . The center of the ball corresponds to the identity rotation \mathbf{I} . Points on the boundary of the ball are in two-to-one correspondence with rotations, because (π, \mathbf{u}) and $(\pi, -\mathbf{u})$ represent the same finite rotation, for any \mathbf{u} .

This situation should seem familiar to geologists, because similar remarks apply to hemispherical plots of direc-

tional data. These plots depict the space of lines through the origin in three-dimensional Euclidean space as a two-dimensional disk, with the understanding that antipodal boundary points of the disk are identical. By a mathematical coincidence, $\text{SO}(3)$ can be depicted as a three-dimensional ball with antipodal points identified. Intuitively, therefore, the angle-axis plot is much like a hemispherical plot in one higher dimension.

Information about rotations can be read from the angle-axis plot easily. Each point corresponds to a rotation. The vector from the origin to the point is the axis of rotation, and its length is the angle of rotation. Alternatively, the vector can be read as vorticity. For rotations can be regarded as steady homogeneous deformations with no distortional component (Appendix A), and $\alpha\mathbf{u}$ is exactly half the vorticity vector \mathbf{w} of such a deformation. Suppose that the rotation arises as the finite result of a steady progressive rotation, that begins at \mathbf{I} at time $t = 0$ and ends at \mathbf{R} at $t = 1$. In the angle-axis plot, the progressive rotation begins at the origin at $t = 0$ and moves toward the point $\alpha\mathbf{u}$ with constant velocity, reaching that point at $t = 1$. If we let the rotation proceed beyond $t = 1$, it continues out to the boundary point $\pi\mathbf{u}$, reappears at the opposite boundary point $-\pi\mathbf{u}$, and proceeds back to the origin.

3.3. Three variations

The angle-axis plot can be radially distorted to produce other plots with various properties. Instead of plotting the rotation (α, \mathbf{u}) at the point $\alpha\mathbf{u}$, we plot it at the point $\rho\mathbf{u}$, where ρ is a certain increasing function of α . Here are three popular examples.

First, choosing $\rho = \tan(\alpha/4)$ produces the *equal-angle plot* of radius 1 (Rowenhorst et al., 2015). It is the direct analogue of the equal-angle hemispherical plot of directional data. Because of its equal-angle property (Appendix C.1), it is the plot to be used when the shape of a data set is important. For example, Fig. 6D shows equal-angle plots of the same fold and quartz data sets from earlier in the figure.

Second,

$$\rho = \left(\frac{3}{4\pi^2} (\alpha - \sin \alpha) \right)^{1/3}$$

produces the *equal-volume* or *homochoric plot* of radius $(3/(4\pi))^{1/3} \approx 0.62$ (e.g., Frank, 1988; Roșca et al., 2014; Rowenhorst et al., 2015). It is the direct analogue of the equal-area hemispherical plot. Because of its equal-volume property (Appendix D), it is the plot to be used when the density of a data set is important. Fig. 5CD and Fig. 6E are examples of the equal-volume plot.

Third, $\rho = \tan(\alpha/2)$ produces the *Rodrigues plot*, which has infinite radius (e.g., Frank, 1988; Morawiec and Field, 1996; Randle and Engler, 2000). See Fig. 6F for examples. This plot is preferred in some fields because of its geometric properties. For example, all steady rotations, not just those through \mathbf{I} , plot as straight lines. On the

other hand, distances between points are greatly distorted — much more so than in the equal-angle or equal-volume plots. When all symmetric copies of an orientation are shown, some may plot very far from the origin, leaving those near the origin too visually compressed to be clearly distinguished (Fig. 6F and its inset). Many authors solve this problem by showing just one symmetric copy near the origin (e.g., Kagan, 2013). That tactic suffices for tightly concentrated data, where there is a clear separation among symmetric copies (Section 2.7). For widely dispersed data, we recommend beginning with the equal-angle and equal-volume plots instead.

In each of these plots, the interior points correspond one-to-one with rotations. In the equal-angle and equal-volume plots, antipodal boundary points are identified. In the Rodrigues plot, there is no boundary, and rotations through $\pi = 180^\circ$ are simply not plotted. In all three plots, the axis of rotation is easily read from the plotted point. However, the radial distortion means that the angle of rotation cannot be read precisely. A steady progressive rotation starting from \mathbf{I} still plots along a line through the origin, but it does not travel that line with constant speed.

3.4. Example: Cyprus paleomagnetic rotations

The Troodos ophiolite in Cyprus includes two orthogonal structures — the NS-striking Solea graben and the EW-striking, dextral Arakapas fault — that form a ridge-transform intersection (MacLeod et al., 1990). In Fig. 7A, the approximate location of this intersection is near the town of Mandria. Sheeted dikes and gabbros are preserved in both the inside (east, off the map) and outside (west) corners, providing a view of mid-crustal deformation in the system.

One common approach to characterizing deformation in the ophiolite is to combine paleomagnetic analysis and field measurements of dike orientations (e.g., Allerton and Vine, 1987; Bonhommet et al., 1988; Hurst et al., 1992; Morris and Maffione, 2016). Following Allerton and Vine (1987), we assume that the dikes intruded vertically, and that the pre-deformation magnetization was aligned with the Troodos magnetization vector \mathbf{t} shown in the inset in Fig. 7A (Clube and Robertson, 1986).

Titus et al. (2015) presented data from 18 stations within the outside corner of the system, northwest of Mandria (Fig. 7A). At each station, they measured 4–21 dike orientations and collected 5–12 specimens for paleomagnetic analysis. We use directional statistics (e.g., Mardia and Jupp, 2000; Borradaile, 2003; Tauxe, 2010) to compute the average dike pole \mathbf{d} and paleomagnetic vector \mathbf{m} at each station (Fig. 7BC). For simplicity of exposition, we do not consider the uncertainty in these means, which may be large, especially where only a few measurements were taken. (See Morris et al. (1998); Morris and Maffione (2016) for examples of incorporating errors into rotational analysis.)

Working in coordinates where x , y , and z point east, north, and up, respectively, we wish to deduce the ro-

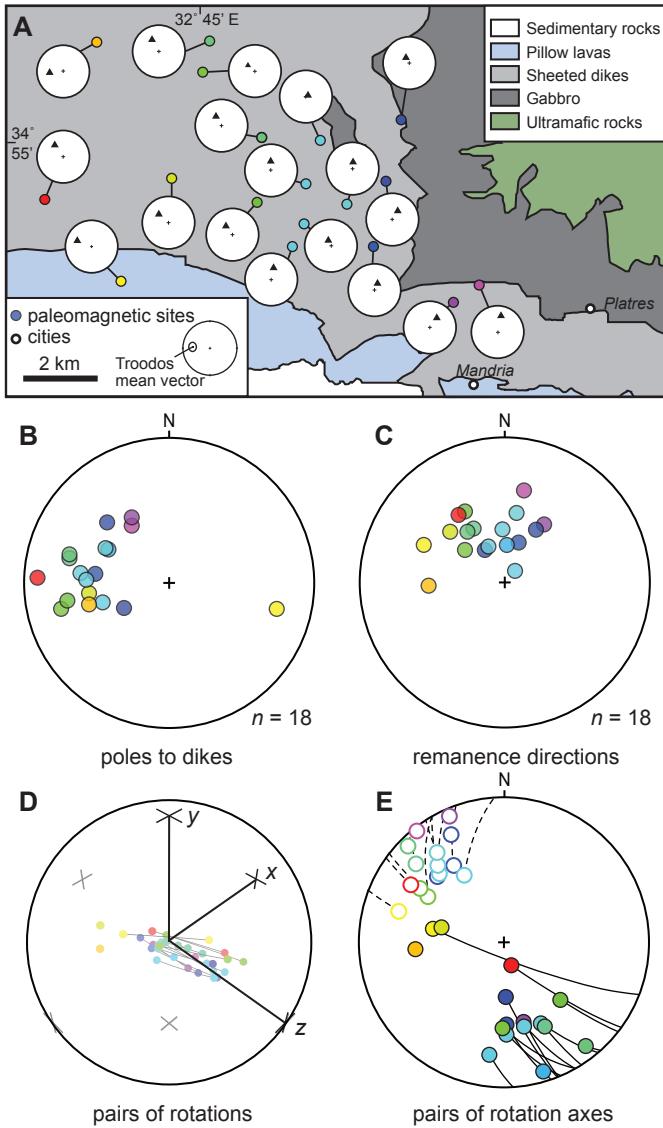


Figure 7: Dike orientation and paleomagnetic data from the Troodos ophiolite, Cyprus. In all panels, data are colored from red (west) to magenta (east). A. Map of the field area, showing the 18 stations. B. Dike poles. C. Paleomagnetic directions, all of which point down. D. Equal-angle rotation plot of the deduced rotations. At stations where there are two solutions, they are joined by a geodesic curve. E. Equal-area plot of the rotation axes labeled with the angles of rotation about them.

tation \mathbf{R} experienced by the rocks at each station. The paleomagnetic data impose the constraint $\mathbf{R}\mathbf{t} = \mathbf{m}$, while the dikes require that $\mathbf{R}^\top \mathbf{d}$ be horizontal or as close to horizontal as possible. See Appendix K for details.

Fig. 7DE show the resulting rotations. At 13 of the stations there are two solutions, which we join with a geodesic curve. Each curve gives an idea of the effect of the vertical dike assumption, in that every point \mathbf{R} along the curve solves $\mathbf{R}\mathbf{t} = \mathbf{m}$, but only the two endpoints make $\mathbf{R}^\top \mathbf{d}$ horizontal as well. At three of the other stations, the best \mathbf{R} results in $\mathbf{R}^\top \mathbf{d}$ diverging from horizontal by roughly 1° . At the other two stations $\mathbf{R}^\top \mathbf{d}$ diverges dramatically from

horizontal: by 46° and 71° . They are two of the westernmost stations, corresponding to the unpaired orange points on the left sides of Fig. 7DE.

The data are colored by easting, from red (west) to magenta (east). Notice that the colors do not appear completely random or jumbled. Rather, they form rainbow patterns, suggesting that rotation varies systematically with distance from the graben. We explore that geographic tendency in Section 6.3.

3.5. Kamb contouring

Kamb contouring is a tool for visualizing the density of a directional data set in an equal-area hemispherical plot (e.g., Vollmer, 1995). The user specifies a smoothing factor k , which equaled 3 in the original paper of Kamb (1959), and another number m . Based on k and the data, the algorithm computes numbers r and σ . It defines the “density” of any point on the sphere to be the number of data points within distance r of that point. It then draws contours into the equal-area plot, consisting of those points on the sphere that have density $m\sigma$. In this section, we construct the analogous concept for rotational data.

Our algorithm begins with a data set $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ of rotations, as well as the user-specified numbers k and m . The density of any rotation \mathbf{Q} is defined to be the number of \mathbf{R}_i in the ball of radius r centered at \mathbf{Q} . The choice of r is motivated by the question, “What is the difference between the sample and the uniform distribution?” Under the uniform distribution, the probability p that a datum lies in a ball equals the volume of the ball divided by the volume of the whole space. In our case, the ball of radius r has volume $(r - \sin r)/\pi$, and $\text{SO}(3)$ has volume 1 (Appendix D). So the probability that a datum lies in any given ball of radius r is $p = (r - \sin r)/\pi$. The number of data within the ball is binomially distributed with probability p . So the expected value is $\mu = np$ and the standard deviation is $\sigma = \sqrt{np(1-p)}$. The algorithm chooses r such that $\mu = k\sigma$, which implies that $p = k^2/(n + k^2)$. Therefore n and k determine p , which determines σ and r . There is no closed-form expression for r , but approximations (Roşca et al., 2014) and iterative algorithms such as Newton’s method can find r quickly.

Once σ and r are known, the algorithm searches through $\text{SO}(3)$, seeking points \mathbf{Q} that have density $m\sigma$ and declaring those points to form the $m\sigma$ Kamb level surface. This search can be viewed as a special case of a more general problem: Given a constant c and a real-valued function f on $\text{SO}(3)$, plot the level surface consisting of all points \mathbf{Q} where $f(\mathbf{Q}) = c$. In the Kamb context, $c = m\sigma$ and f is

$$f(\mathbf{Q}) = \sum_{i=1}^n h(d(\mathbf{Q}, \mathbf{R}_i)),$$

where h is defined piecewise by $h(\alpha) = 1$ if $\alpha \leq r$ and $h(\alpha) = 0$ if $\alpha > r$. Appendix E.1 describes an adaptive mesh algorithm for approximating the level surface as a union of polygons.

The discontinuity of h at $\alpha = r$ can lead to jaggedness in the plotted level surface. Smoother results can be obtained by using smoother functions h in the definition of f . Intuitively, such an h not only counts the data \mathbf{R}_i near \mathbf{Q} , but also weights each datum according to its proximity to \mathbf{Q} . For computational speed one can define h using a polynomial of small degree (Vollmer, 1995). Appendix E.2 explains how to compute the coefficients of polynomials of degree three or less.

Fig. 5EF are two examples of 6σ Kamb level surfaces, with $k = 3$ and degree-3 weighting polynomials. In each plot, the surface seems to end abruptly at the boundary of the plot, but always continues seamlessly on the other side of the plot. Each symmetric piece of the surface has the same shape in $\text{SO}(3)$ but appears distorted in the figure, because of the shape distortion of the equal-volume plot.

4. Location and dispersion

4.1. Motivation

When presented with any data set, the first statistics that we compute are often the mean and the standard deviation, which measure the location and dispersion of the data, respectively. In the rotational context, the definition and computation of well-behaved means turns out to be a subtle problem, much like the design of well-behaved plots.

In two dimensions, any rotation can be described as a single angle. However, one cannot reliably average rotations simply by averaging their angles. For example, two finite rotations through 3° and 359° should average to a rotation through 1° , but the arithmetic mean of 3° and 359° is 181° . The fundamental issue is that the relationship between rotations and their angles is discontinuous at 0° and 360° . By adopting a different convention for angles, we can shift this *branch cut* elsewhere in the space of rotations, but we can never eliminate it.

In three dimensions, averaging Euler angles does not produce a well-behaved notion of mean rotation (e.g., Krieger Lassen et al., 1994). The branch cut and gimbal lock (Section 3.1) are two of the problems. Another issue is that the Euler angles describe a rotation as a sequence of three rotations about the coordinate axes of space, and these rotations do not commute. The order of the sequence matters.

Alternatively, each three-dimensional rotation can be described as a special orthogonal matrix \mathbf{R} (Appendix A). However, arithmetically averaging such matrices usually produces a matrix that is not a rotation at all. We return to this idea in Section 4.3.

Another strategy is to work with “infinitesimal” rotations (e.g., Altman, 2005, Chapter 4). Given rotations $\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n\}$, we compute their principal matrix logarithms $\log \mathbf{R}_i$ (Appendix A), average those logarithms, and then compute the exponential of the average matrix. This procedure does produce a rotation matrix, and it dodges the commutativity issue that plagues Euler angles,

but it is still subject to branch cut problems along the rotations through $\pm\pi = \pm180^\circ$. We return to this idea in Section 4.2.

In fact, every method for representing rotations suffers from drawbacks (Appendix A), which complicate even simple calculations such as the mean. In this section, we summarize two approaches to quantifying the mean of a data set and four approaches to quantifying dispersion. We demonstrate these methods in two applications.

4.2. Fréchet mean and variance

In elementary statistics, the variance of a data set $\{x_1, \dots, x_n\}$ about a point x is $\Psi(x) = n^{-1} \sum (x - x_i)^2$, and the sample mean $\bar{x} = n^{-1} \sum x_i$ is the point x where the variance is minimized. Similarly, for any rotation \mathbf{R} , define the *Fréchet variance* of the data $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ about \mathbf{R} as

$$\Psi(\mathbf{R}) = \frac{1}{2n} \sum_{i=1}^n d(\mathbf{R}, \mathbf{R}_i)^2$$

(e.g., Moakher, 2002; Krakowski et al., 2007; Fletcher, 2013; Hartley et al., 2013). When the \mathbf{R}_i are tightly concentrated about \mathbf{R} , $\Psi(\mathbf{R})$ is near zero. Because distances in $\text{SO}(3)$ never exceed π , Ψ never exceeds $\pi^2/2 \approx 4.93$.

Define the *Fréchet mean* $\bar{\mathbf{R}}$ of the data as the \mathbf{R} that minimizes Ψ . This mean can be computed using general numerical optimization methods, but Appendix C.2 gives a concrete algorithm, that is an iterative version of the logarithmic method mentioned in Section 4.1. Fig. 8A shows the mean of a synthetic data set of 10 rotations. The plot includes a geodesic from the mean to each datum. The variance is the sum of the squared lengths of these geodesics.

The variance $\Psi(\bar{\mathbf{R}})$ of a data set about its mean is one measure of the dispersion of the data set. For example, in Fig. 5C and Fig. 8A, $\Psi(\bar{\mathbf{R}})$ equals 0.007 and 0.172, respectively. In Fig. 5D, $\Psi(\mathbf{R}) = 1.453$, but the equal-volume plot and Kamb level surfaces suggest that the data set has multiple regions of high density, making the sample mean and the variance about that mean unsuitable as summary statistics.

4.3. Projected arithmetic mean

Recall that the arithmetic mean $\tilde{\mathbf{R}} = n^{-1} \sum \mathbf{R}_i$ is usually not a rotation matrix. However, suppose that the \mathbf{R}_i are concentrated about some mean rotation (rather than girdled, concentrated about two distinct means, etc.). Then $\tilde{\mathbf{R}}$ is close to that mean in the space of 3×3 matrices. So it is not unreasonable to define a *projected arithmetic mean* $\bar{\mathbf{R}}$ as the rotation closest to $\tilde{\mathbf{R}}$ (e.g., Moakher, 2002). Fig. 8F is a cartoon illustration of this idea. The projected arithmetic mean is usually close to, but not equal to, the Fréchet mean of Section 4.2. In this paper we denote both means as $\bar{\mathbf{R}}$, distinguishing between them only when the distinction is important.

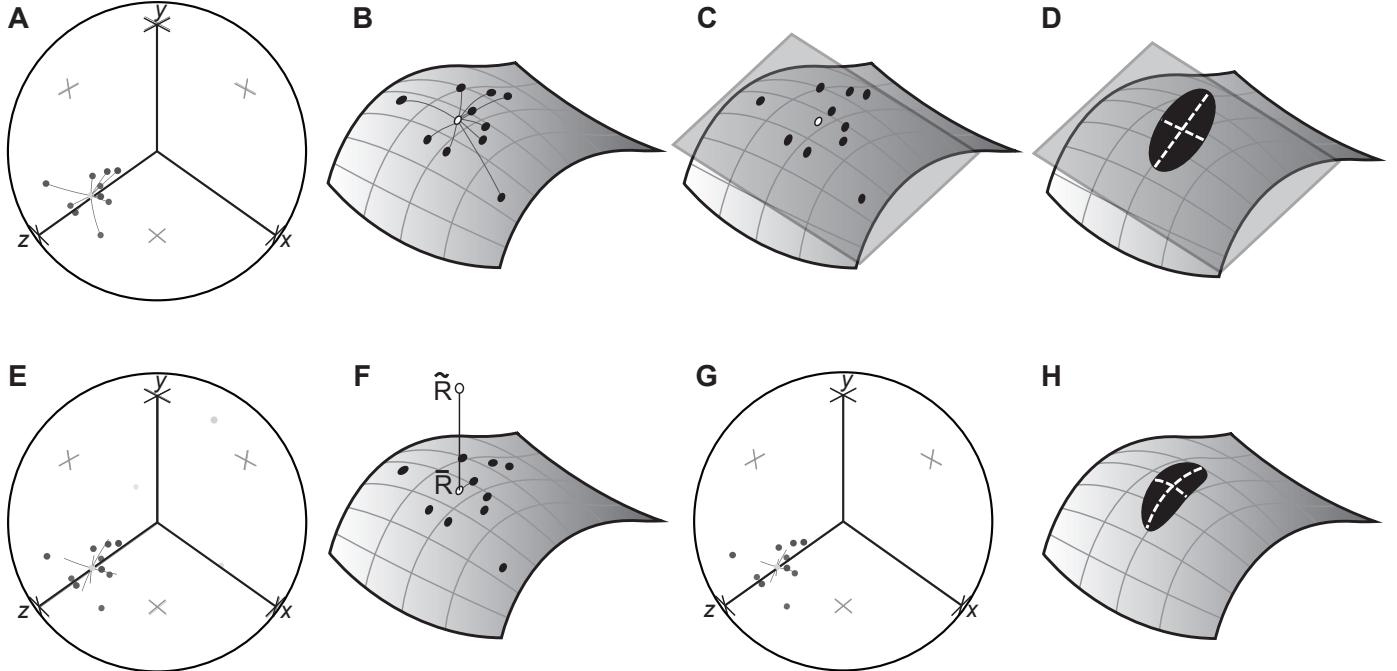


Figure 8: Mean and dispersion of a synthetic data set of 10 rotations. A. Equal-angle rotation plot of the Fréchet mean as the minimizer of the Fréchet variance. B. Cartoon depicting $\text{SO}(3)$ as a two- rather than three-dimensional, with Fréchet mean and variance. C. The data transferred to the tangent space at the mean. D. Principal component analysis in the tangent space. E. Equal-angle rotation plot of projected arithmetic mean, including geodesics toward the three scatter points. F. Cartoon of the projected arithmetic mean. G. Equal-angle rotation plot of principal geodesic analysis. H. The principal components from panel D, transferred back to $\text{SO}(3)$.

If $\det \tilde{\mathbf{R}} > 0$, which is typical for data concentrated about a mean, then the projected arithmetic mean is

$$\bar{\mathbf{R}} = \tilde{\mathbf{R}}(\tilde{\mathbf{R}}^\top \tilde{\mathbf{R}})^{-1/2}$$

(Downs, 1972). To compute $(\tilde{\mathbf{R}}^\top \tilde{\mathbf{R}})^{-1/2}$, one can diagonalize and then raise each diagonal entry to the $-1/2$ power. Related constructions include the singular value decomposition $\tilde{\mathbf{R}} = \mathbf{U}\Sigma\mathbf{V}^\top$, in which $\tilde{\mathbf{R}} = \mathbf{U}\mathbf{V}^\top$ (Krieger Lassen et al., 1994; Lay, 1994), and the polar decomposition $\tilde{\mathbf{R}} = \bar{\mathbf{R}}\mathbf{H}$, where $\mathbf{H} = \mathbf{V}\Sigma\mathbf{V}^\top$ is symmetric and positive-definite (Elliott, 1970; Moakher, 2002).

The projected arithmetic mean can be viewed in a different way, which leads to a notion of dispersion about the mean. Rotations correspond to antipodal pairs of unit quaternions $\mathbf{q} = [q_0 \ q_1 \ q_2 \ q_3]^\top$ (Appendix A), or equivalently to lines in four-dimensional space, which are subject to directional statistics (Moran, 1975; Prentice, 1986). Let \mathbf{q}_i be the unit four-dimensional vector corresponding to \mathbf{R}_i , and form the 4×4 *scatter matrix*

$$\mathbf{T} = \frac{1}{n} \sum_{i=1}^n \mathbf{q}_i \mathbf{q}_i^\top.$$

Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ be the eigenvalues of \mathbf{T} . They are non-negative and sum to 1. The associated unit eigenvectors can be viewed as unit quaternions and hence rotations. For data concentrated about a mean, $\lambda_1 + \lambda_4 > 0.5$ and the rotation corresponding to λ_1 equals the projected arithmetic mean. Each of the other eigenvalues measures

the dispersion of the sample in the direction of its associated rotation (Bachmann et al., 2010).

Fig. 8E shows an example. The eigenvalues of \mathbf{T} are 0.92, 0.05, 0.02, and 0.01. The four corresponding rotations are shown in gray. A geodesic curve is drawn from the mean toward each of the other three rotations, with length equal to the square root of the corresponding eigenvalue.

4.4. Example: Fabric from rigid clasts

Suppose that we wish to study how fabric develops in rock containing rigid ellipsoidal clasts. If the host rock is linearly viscous, the deformation is slow, and the clasts are adequately separated in space, then the rotation of the clasts is described well by the dynamical theory of Jeffery (1922). In general, its differential equations are too complicated to be solved exactly. Many authors have analyzed special cases, such as plane-strain deformation or spheroidal clasts (e.g., Passchier, 1987; Jezek et al., 1996). In this subsection, as an illustration of Section 4.3, we study the fabrics that develop in monoclinic transpression of rigid ellipsoids with axial ratios 1 : 2 : 3.

Fig. 9A shows 1,000 randomly distributed ellipsoid orientations, chosen by sampling from the uniform distribution on $\text{SO}(3)$ (Appendix D.2). Notice that the orientations appear equally dense in all regions of this plot, because of the plot's equal-volume property. We subject these initial ellipsoids to homogeneous monoclinic transpression (Fossen and Tikoff, 1993). This kind of deforma-

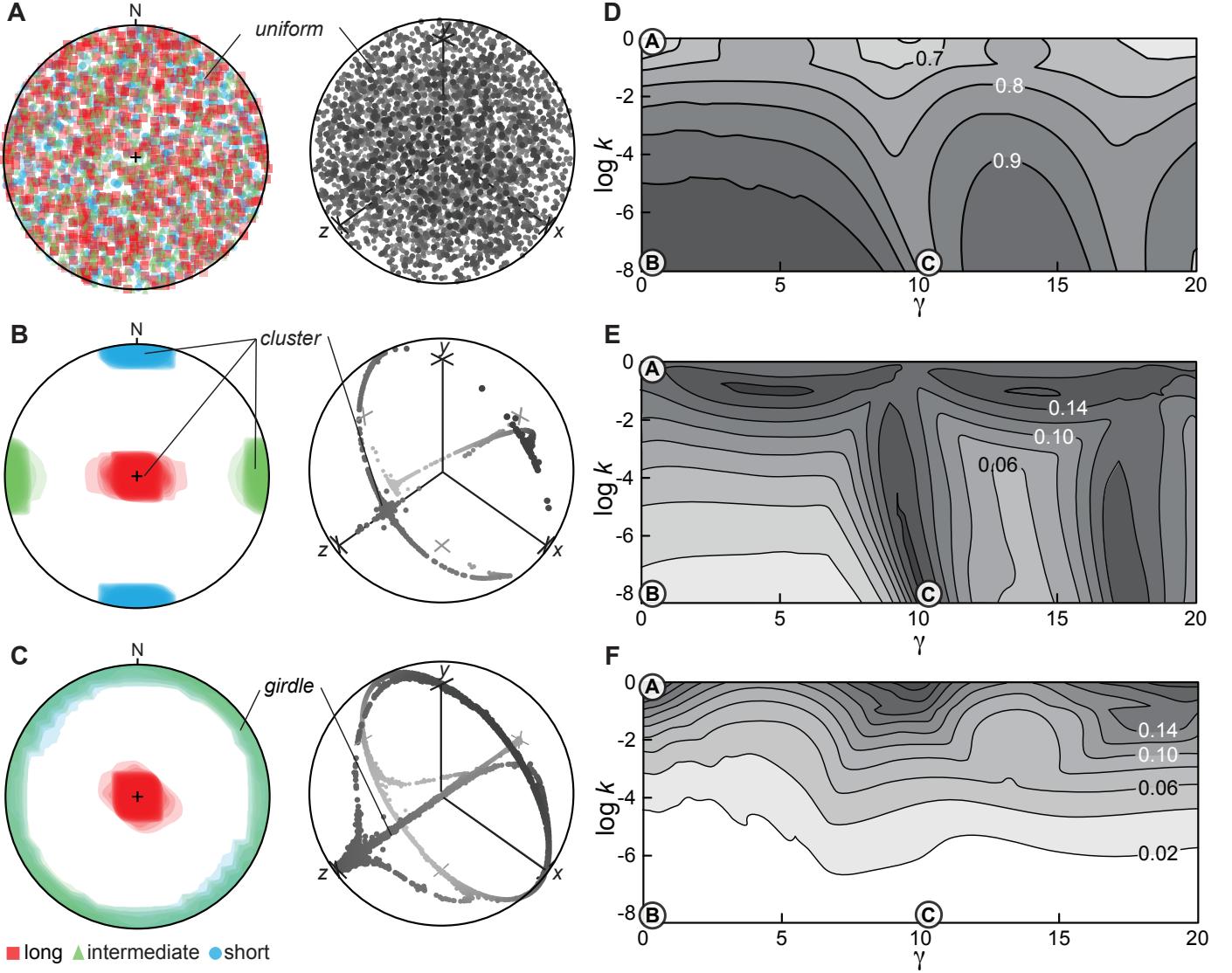


Figure 9: Numerical simulation of rigid ellipsoids. A. Equal-area and equal-volume plots of 1,000 uniform initial orientations. B. Final orientations after the $(0, -8)$ transpression. The orientations are concentrating at an attractor. C. Final orientations after the $(10.25, -8)$ transpression. The orientations form a girdle. D. The landscape of homogeneous monoclinic transpressions, with contours of λ_1 . E. Contours of λ_2 . F. Contours of $\lambda_3 + \lambda_4$.

tion is parametrized by pairs of numbers $(\gamma, \log k)$, where γ is the magnitude of simple shear and $\log k$ describes the shortening across the shear zone. For any given $(\gamma, \log k)$, we simulate the rotation of the ellipsoids into their final orientations using the Lie group Runge-Kutta method of Davis et al. (2013, Section 4). We compute the Fréchet mean final orientation and one of its representative rotations. For each final orientation, we select the representative rotation that is closest to the mean rotation. Then, for that set of rotations, we compute the tensor \mathbf{T} and its eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$. Fig. 9D-F are contour plots of λ_1 , λ_2 , and $\lambda_3 + \lambda_4$ as functions of $(\gamma, \log k)$. To explain the meaning of these plots, we focus on three points of interest.

First, at $(\gamma, \log k) \approx (0, 0)$, $\lambda_1 \ll 1$ and the other $\lambda_i \gg 0$. The final ellipsoid orientations are widely dispersed

throughout $\text{SO}(3)$. This makes sense, because such a small deformation alters the initially uniform fabric of Fig. 9A only slightly.

Second, at $(0, -8)$, which corresponds to 99.97% coaxial shortening, $\lambda_1 \approx 1$ and the other $\lambda_i \approx 0$. These eigenvalues indicate that the final ellipsoids are tightly clustered about a mean orientation, suggesting a fabric attractor. Fig. 9B confirms that most of the clasts have approached a particular orientation. (If the deformation is allowed to continue, then the rest of the clasts are also attracted there.)

Third, at $(10.25, -8)$, $\lambda_1 \ll 1$, $\lambda_2 \gg 0$, and $\lambda_3 \approx \lambda_4 \approx 0$. The ellipsoids have not approached an attractor, because their orientations are widely dispersed in one direction in $\text{SO}(3)$. That is, they form a girdle, which is depicted in Fig. 9C.

Beyond these three examples, the contour plots help us detect patterns across the entire landscape of transpressions. For example, the plots indicate either attraction or girdling throughout the bottom half of the landscape. Therefore, if a geologist observes some other kind of wide dispersion in a field fabric, then she surmises that $\log k$ is small.

Other ellipsoid axial ratios and other shear zone geometries, such as triclinic transpression (Jones and Holdsworth, 1998; Lin et al., 1998), can be analyzed similarly. Also, instead of the eigenvalues of \mathbf{T} from Section 4.3, one could measure dispersion using the principal component magnitudes from Section 4.5 or the eigenvalues of $\hat{\mathbf{K}}$ from Section 4.7.

4.5. Tangent space approximation

In this subsection we summarize two related approaches to describing anisotropy of dispersion: principal geodesic analysis and Mahalanobis distance. These methods are appropriate only for tightly concentrated data, such as EBSD data from a single grain (e.g., Bachmann et al., 2010) or the results of certain simulations (Section 5.5). Geologic field data may be too widely dispersed for these methods. Although the computations are not complicated, we leave all details to Appendix C.2.

Recall from geometry and calculus that any point on a smooth curve possesses a tangent line, which is the line that best approximates the curve near that point. Similarly, at any point on a smooth surface there is a tangent plane, which approximates the surface near that point. Analogously, at any point \mathbf{R} in the curved, three-dimensional space $\text{SO}(3)$, there is a flat three-dimensional tangent space, which approximates $\text{SO}(3)$ near the point of tangency.

This subsection's strategy for analyzing rotational data $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ is to compute their Fréchet mean $\bar{\mathbf{R}}$ and then to transfer the data from $\text{SO}(3)$ to the tangent space at $\bar{\mathbf{R}}$. Fig. 8BC are cartoons illustrating this idea. The tangent space is a computationally convenient setting, because it is a Euclidean vector space, where a wide range of standard multivariate statistical techniques can be applied (e.g., van den Boogaart and Schaeben, 2002a, p. 481). The transference distorts the shape of the data set and hence the statistical conclusions drawn from it, especially for points far from $\bar{\mathbf{R}}$. This is why tangent space approximation is appropriate only for tightly concentrated data.

In Euclidean statistics, the technique of *principal component analysis* is used to quantify anisotropy of dispersion. The first principal component captures the direction and magnitude of greatest spread. The second principal component is perpendicular to the first, and captures the next-greatest spread. Other principal directions are defined similarly, up to the dimensionality of the data. Fletcher et al. (2004) adapted principal component analysis into *principal geodesic analysis* on $\text{SO}(3)$, by performing principal component analysis in the tangent space (Fig. 8D) and transferring the results back to $\text{SO}(3)$ as geodesics

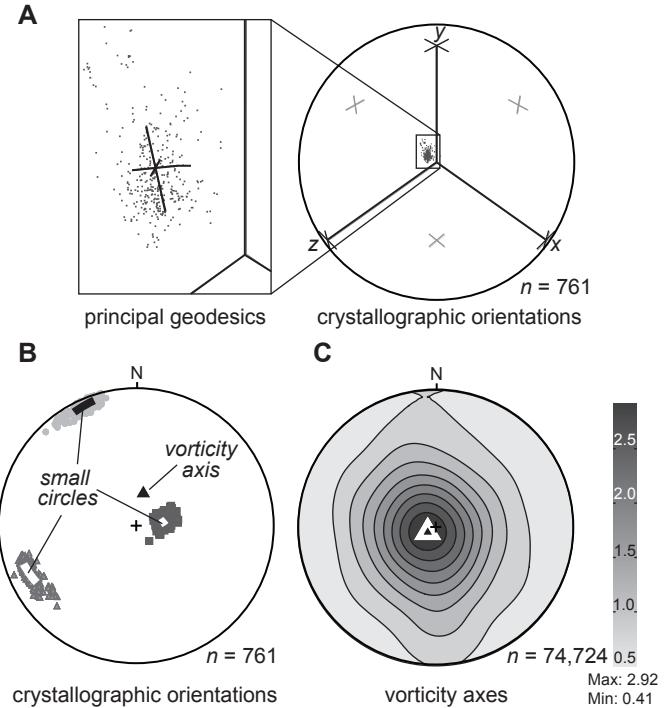


Figure 10: Inferring vorticity of deformation from dispersion of crystallographic orientations. A. Equal-angle plot of 761 quartz orientations from a single grain, with principal geodesics superimposed. B. Equal-area lower-hemispherical plot of the same data set, with the deduced vorticity vector and small circles about that vector fitting the dispersed axes. C. Adapted from Michels et al. (2015). Density contours (multiples of uniform density, de la Vallée Poussin kernel) of vorticity vectors inferred from 74,724 such grains.

(Fig. 8H). Fig. 8G shows the principal geodesics for the synthetic data set.

Mahalanobis distance is closely related to principal component analysis. It is an anisotropic notion of distance, that measures how far from $\bar{\mathbf{R}}$ any given rotation \mathbf{Q} is, compared to the data $\{\mathbf{R}_i\}$. For example, if the data are widely spread in a certain direction from $\bar{\mathbf{R}}$, then \mathbf{Q} can be far from $\bar{\mathbf{R}}$ in that direction, without being considered far by this notion of distance. The set of rotations \mathbf{Q} that are Mahalanobis distance 1 from $\bar{\mathbf{R}}$ forms an ellipsoid, whose axes are exactly the principal geodesics (Fig. 8H). In this paper, Mahalanobis distance helps us interpret the results of bootstrap and Markov chain Monte Carlo simulations in Section 5.5.

Other examples of tangent space approximation include the inference method of Rancourt et al. (2000) (see Section 5.2) and the wrapped trivariate normal distribution used by Qiu et al. (2014) (see Section 5.4).

4.6. Example: Crystallographic vorticity axis

Bestmann and Prior (2003) and Reddy and Buchan (2005) argued that as a rock deforms its mineral grains generally rotate about the vorticity vector of the deformation, so that their crystallographic axes are dispersed along small circles with the vorticity as their common pole.

Therefore, after deformation the vorticity direction should be readable from the dispersion of crystallographic axes observed in thin section.

To develop this idea, Michels et al. (2015) used EBSD to measure quartz orientations in three deformed quartzites. The left side of Fig. 5 shows their data from a single quartz grain from the Moine thrust zone (Strine and Wojtal, 2004), in equal-area and equal-volume plots. Because the data set is tightly concentrated, it is reasonable to ignore the symmetry and apply a tangent space approximation. Fig. 10A shows the same data in an equal-angle rotation plot. The inset shows the three principal geodesics. We can conceptualize the first principal geodesic as a steady rotation $\mathbf{Q}(t)$, about some unit vector \mathbf{u} , that passes through the mean orientation at $t = 0$. Among all such steady rotations, the first principal geodesic is the one that best fits the orientations. When $\mathbf{Q}(t)$ is dissected along its rows, we obtain three curves of unit vectors, which are small circles about \mathbf{u} (Michels et al., 2015, Supplemental Material). Fig. 10B shows these small circles fitting their respective data well. In this sense, \mathbf{u} is a fitted vorticity axis for this grain.

Each grain produces a vorticity axis in this manner. Fig. 10C shows density contours for vorticity axes inferred from the 74,724 grains in the Moine thrust zone sample. The vector at the point maximum is taken to be the bulk vorticity axis for the sample. Michels et al. (2015) showed, in all three samples tested, that the inferred bulk vorticity axis agreed with other methods of estimating vorticity direction.

4.7. Maximum likelihood estimation

The *matrix Fisher distribution* is the most widely studied analogue of the normal distribution on $\text{SO}(3)$ (e.g., Downs, 1972; Khatri and Mardia, 1977; Prentice, 1986; Wood, 1993; Mardia and Jupp, 2000; Chikuse, 2003; Kume et al., 2013). Recall that the normal distribution on the real numbers has two parameters: The mean μ describes the center of the distribution, and the variance σ^2 describes the dispersion. Similarly, the multivariate normal distribution is parametrized by a mean vector μ and a symmetric, positive-definite covariance matrix Σ . Analogously, the matrix Fisher distribution on $\text{SO}(3)$ is parametrized by a matrix \mathbf{M} describing the center of the distribution and a symmetric, positive-definite matrix \mathbf{K} describing the dispersion. Large eigenvalues in \mathbf{K} indicate tight concentration around \mathbf{M} , while the distribution goes to uniform as $\mathbf{K} \rightarrow \mathbf{0}$. See Appendix G for details, including an efficient method for sampling from this distribution (Kent et al., 2013).

Given a data set $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ of rotations, one can assume that they are drawn independently from a single matrix Fisher distribution, and then ask which values of \mathbf{M} and \mathbf{K} maximize the probability of the data given \mathbf{M} and \mathbf{K} . Those values, called the *maximum likelihood estimates* $\hat{\mathbf{M}}$ and $\hat{\mathbf{K}}$, quantify the mean and dispersion of the sample. For data concentrated about a mean, $\hat{\mathbf{M}}$ equals the

projected arithmetic mean of Section 4.3 (Downs, 1972). Calculation of $\hat{\mathbf{K}}$ is more difficult (Appendix G).

For example, the data set of Fig. 8 is drawn from a matrix Fisher distribution where \mathbf{K} has eigenvalues 1, 7, 7. Our maximum likelihood estimation software reports $\hat{\mathbf{K}}$ with eigenvalues 0.000001, 6, 13. The discrepancy is unsurprising for such a small data set ($n = 10$). For larger data sets drawn from this distribution, it is more typical for $\hat{\mathbf{K}}$ to have eigenvalues 1, 6, 8 (when $n = 100$) or 1, 7, 7 (when $n = 1,000$).

Beyond this small example, maximum likelihood estimation and the matrix Fisher distribution play a supporting role throughout this paper. They appear in parametric bootstrapping (Section 5.3) and the selection of inference methods (Sections 5.8 and 7.3).

5. Inference about the mean

5.1. Motivation

Often we wish to extrapolate from a data set, to make inferences about the larger population that it represents. For example, we cannot measure foliation and lineation at every point in a field area. So we measure 40 foliation-lineations, assume that they are representative of the area, and try to draw conclusions from them. In particular, the mean of a data set estimates the true population mean, but with some uncertainty. Hypothesis tests and confidence regions are two common strategies for handling this uncertainty.

A *hypothesis test* begins with the statement of a null hypothesis and an alternative hypothesis. In this paper the null hypothesis is usually of the form “The population mean is \mathbf{Q} ” for a given rotation \mathbf{Q} , and the alternative hypothesis is “The population mean is not \mathbf{Q} .” The tester also chooses a *significance level* α , such as 0.05 or 0.01. The test assumes that the null hypothesis is true and computes a *p-value*, which is the probability of observing the data, or data more “extreme” than those observed, given the null hypothesis. If the *p-value* is less than the significance level, then the test rejects the null hypothesis. Otherwise, the test is inconclusive. In particular, large *p*-values do not confirm the null hypothesis (e.g., Kline, 2004; Goodman, 2008).

For any hypothesis test with significance level α , there is a corresponding $100(1 - \alpha)\%$ *confidence region*, consisting of those \mathbf{Q} that would not be rejected if the test were applied to them. For example, a 95% confidence region corresponds to a test with $\alpha = 0.05$. Like hypothesis tests, confidence regions are easily misinterpreted. One can say that the method of computing the 95% confidence region, if repeated for many such data sets drawn from that population, would produce regions containing the true population mean for 95% of those data sets (e.g., Wolf, 1974; Efron and Tibshirani, 1993). It is not valid to say that a 95% confidence region has a 95% probability of containing

the population mean. However, in the Bayesian framework (Section 5.4) there is an analogous notion of *credible region* that does have this interpretation (Gregory, 2005).

In this section, we summarize six approaches to constructing confidence/credible regions and performing hypothesis tests for the mean \mathbf{Q} of a population of rotations. Three of the approaches can be applied to two-sample problems, where the quantity of interest is the difference in mean of two populations. All of our confidence/credible regions are ellipsoids, which can be described verbally or graphically. We also present numerical experiments, which evaluate the suitability of these approaches for geologic data.

5.2. Three asymptotic methods

The orientation statistics literature offers various asymptotic confidence regions and hypothesis tests. Compared to the simulation-based approaches of the following subsections, these closed-form methods are fast to compute. However, they may produce poor results for data that are few in number or widely dispersed. Further, they are not obviously adaptable to situations where the rotational data are subject to symmetry (Section 2). Therefore these methods may not be widely applicable in geology.

Nevertheless, we have tried three such methods. All of them assume that the data are *independent and identically distributed*, meaning that they are independently drawn from a single population. First, Downs (1972, Eq. 5.7, $p = 2$ case) described confidence regions for the mean of a version of the matrix Fisher distribution (Appendix G), assuming both large sample size and a tightly concentrated sample. Second, Prentice (1986) converted rotations to quaternions (Appendix A) and then used sampling theory of the sample moment-of-inertia matrix to construct a confidence region, assuming large sample size. Third, Rancourt et al. (2000) used Euclidean statistical methods in the tangent space at the mean. Like other tangent space approximations (Section 4.5), this method assumed tightly concentrated data. Section 5.6 investigates the performance of these methods relative to each other and the following non-asymptotic approaches.

5.3. Bootstrapping

Bootstrapping is an approach to inference based on numerical simulation (Efron and Tibshirani, 1993). Compared to the closed-form methods just described, it is slow to compute. However, it is easy to implement, even when the rotational data are subject to symmetry, and it can work even for widely dispersed data. In this section we summarize bootstrapping of the (Fréchet or projected arithmetic) mean $\bar{\mathbf{R}}$ of rotational data $D = \{\mathbf{R}_1, \dots, \mathbf{R}_n\}$, assumed to be independent and identically distributed.

Concretely, we sample the data set with replacement n times to produce a synthetic data set, called a *bootstrap sample*. The bootstrap sample has the same size as the original data set, but some of the original data are

omitted and others are repeated. Then we compute the mean of the bootstrap sample. We repeat this procedure $b = 10,000$ times, say, to produce a set of b bootstrapped means, which we denote $\bar{\mathbf{R}}^1, \bar{\mathbf{R}}^2, \dots, \bar{\mathbf{R}}^b$. Their dispersion characterizes the uncertainty in $\bar{\mathbf{R}}$ as an estimate of the population mean.

Conceptually, the data set is an empirical approximation to the population's true probability distribution. A bootstrap sample is a set of n rotations drawn independently from this empirical distribution. The dispersion of the set $\{\bar{\mathbf{R}}^1, \dots, \bar{\mathbf{R}}^b\}$ approximates the dispersion of the sampling distribution for the mean, which is the crux of inference.

Because the data set D probably does not contain any points deep in the tails of the true population distribution, it tends to underrepresent the dispersion of that distribution, especially when n is small (e.g., Tauxe, 2010, Section 12.3). An alternative approach is the *parametric bootstrap* (Efron and Tibshirani, 1993, Chapter 21): Fit a matrix Fisher (or other) distribution to the data by estimating its parameters \mathbf{M} and \mathbf{K} (Section 4.7), and build bootstrap samples by drawing from this distribution rather than the original data set D .

5.4. Markov chain Monte Carlo

Another common simulation technique is *Markov chain Monte Carlo (MCMC)*; e.g., Gregory, 2005; Robert and Casella, 2013). See Appendix H for an overview of the key concepts and algorithms. In this subsection, we merely summarize how the work of Qiu et al. (2014) can be applied to our inference problems. Again we assume that the data $D = \{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ are independent and identically distributed.

Like the matrix Fisher distribution, the *wrapped trivariate normal distribution* is an analogue of the normal distribution (Appendix F.3). It is parametrized by a mean rotation \mathbf{M} and a scalar κ that measures concentration isotropically about \mathbf{M} . Unlike the matrix Fisher distribution, this distribution does not permit maximum likelihood estimation.

Within the Bayesian interpretation of statistics, \mathbf{M} and κ can be viewed as random variables, whose uncertainty reflects our incomplete knowledge of the system under consideration. The MCMC simulation produces a large set of pairs $\{(\mathbf{M}, \kappa)^1, (\mathbf{M}, \kappa)^2, \dots, (\mathbf{M}, \kappa)^b\}$, where $b = 10^7$, say. This set approximates the probability distribution $\rho(\mathbf{M}, \kappa | D)$ of the parameters given the data. Discarding the κ -values amounts to marginalizing over κ , so that the remaining \mathbf{M} -values approximate the distribution of \mathbf{M} given the data. In this way, $\mathbf{M}^1, \dots, \mathbf{M}^b$ capture the uncertainty in the estimated population mean.

5.5. Processing simulation results

We have not yet described confidence/credible regions and hypothesis tests based on the simulations of the preceding sections. To that end, let $\mathbf{S}^1, \dots, \mathbf{S}^b$ be either

the bootstrapped means $\bar{\mathbf{R}}^1, \dots, \bar{\mathbf{R}}^b$ of Section 5.3 or the wrapped trivariate normal means $\mathbf{M}^1, \dots, \mathbf{M}^b$ of Section 5.4. Let $\bar{\mathbf{S}}$ be their mean.

The basic idea is to construct a region, centered on $\bar{\mathbf{S}}$, that contains 95% of the \mathbf{S}^j (Efron and Tibshirani, 1993, Chapter 13). Bingham et al. (2009); Qiu et al. (2014) constructed a circular region around each of the rows (or columns) of $\bar{\mathbf{S}}$, based on angular differences with the rows of the \mathbf{S}^j . We prefer to treat the \mathbf{S}^j holistically rather than row-by-row, and we find that the \mathbf{S}^j often form a cloud whose shape is ellipsoidal but not spherical. We handle this anisotropy using Mahalanobis distance (Section 4.5).

Concretely, we first inspect $\mathbf{S}^1, \dots, \mathbf{S}^b$ in the equal-angle rotation plot (Section 3.3), to verify that they form a tight ellipsoidal cloud. If so, then we compute their mean $\bar{\mathbf{S}}$, their Mahalanobis distances to $\bar{\mathbf{S}}$, and the 95th percentile q of those distances. The set of all rotations \mathbf{Q} whose Mahalanobis distance equals q forms an ellipsoid, centered at $\bar{\mathbf{S}}$, that encloses 95% of the \mathbf{S}^j . The region bounded by this ellipsoid is the 95% confidence/credible region. A null hypothesis \mathbf{Q} is rejected if it lies outside the region — that is, if its Mahalanobis distance to $\bar{\mathbf{S}}$ exceeds q .

The confidence/credible region can be reported in various ways. It can be displayed as an ellipsoid in the equal-angle or equal-volume plot. Dissected along the rows or columns of its constituent matrices, it can be displayed as three ellipses in an equal-area or equal-angle hemispherical plot. Alternatively, the region can be summarized verbally by stating its center, the directions of its vertices, and the distance to those vertices, which are computable by principal geodesic analysis (Section 4.5). Similar summaries are used in Bingham statistics of directional data (e.g., Borodaaile, 2003, Section 10.6.2). Depending on the application, the region can also be described as a range of strikes, dips, and rakes, a range of trends, plunges, and angles of rotation, etc. See Sections 5.8 and 7.3 for examples.

5.6. Numerical experiments

We have summarized three asymptotic methods and three simulation-based methods for constructing 95% confidence regions and credible regions for the mean: Downs (1972), Prentice (1986), Rancourt et al. (2000), non-parametric bootstrapping, parametric bootstrapping, and MCMC simulation. This subsection summarizes a series of numerical experiments on the accuracy of these methods, comprising several years of computer processor time (completed in a few weeks, using up to 100 processors at a time). See Appendix I for more detail about the procedures and results.

Briefly, we generate $N \geq 2,000$ synthetic data sets, for which the true population mean \mathbf{M} is known, and ask whether the confidence/credible regions produced by the six methods contain that \mathbf{M} . The proportion of trials in which a method's region contains \mathbf{M} is called the *coverage rate* of the method. One expects 95% confidence regions to produce coverage rates around 0.95. If a rate exceeds 0.95, then the regions are too large and the method is too

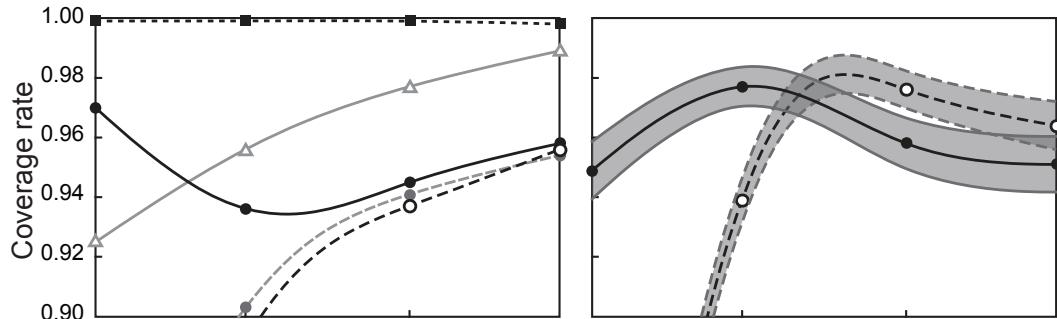
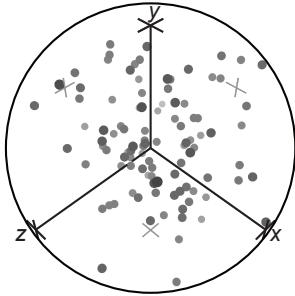
conservative in its inferences. A rate less than 0.95 indicates that the regions are too small and the method tends to over-reject null hypotheses. Although Bayesian credible regions operate on a different logic, it is common to investigate their performance in the same way (e.g., Bingham et al., 2009; Qiu et al., 2014).

The synthetic data sets are drawn from matrix Fisher distributions (Section 4.7) where the concentration \mathbf{K} has prescribed eigenvalues. Fig. 11A summarizes experiments using eigenvalues 1, 1, 1. Such data sets are very widely dispersed, and we view them as a “worst-case scenario” for geologic data. Fig. 11BC summarize experiments based on eigenvalues 1, 1, 7 and 1, 7, 7, respectively. Because the eigenvalues are not all equal, these data sets are anisotropically distributed about their means. Fig. 11D summarizes experiments about isotropic data sets produced by eigenvalues 7, 7, 7. These data sets are moderately dispersed, and we view them as more typical of geologic field studies.

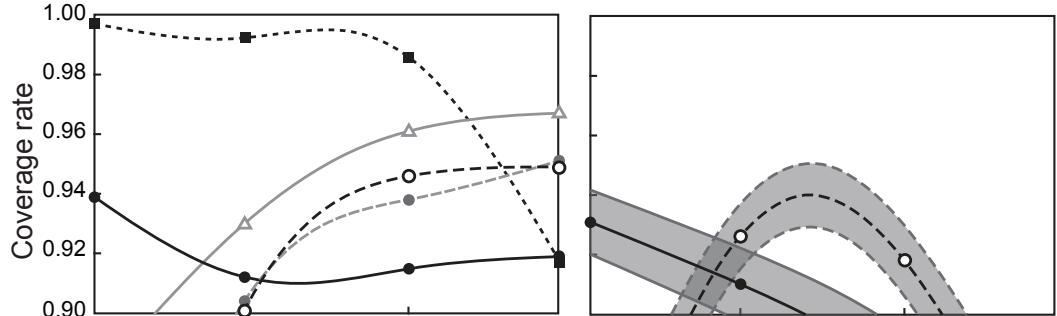
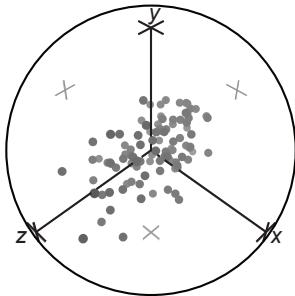
Our first experiment tests the accuracy of all six methods for rotational data without symmetry. The results, shown in the middle column of Fig. 11, largely corroborate the theoretical statements made in earlier subsections. The coverage rates produced by the method of Downs (1972) are much greater than 0.95 in all but one of our tests. In other words, that method's confidence regions are too large, for sample sizes and concentrations as small as ours. The Prentice (1986) confidence regions are too small for $n \leq 30$ but fairly accurate for $n \geq 100$. The Rancourt et al. (2000) confidence regions become more accurate as the data become more concentrated, but are too small even in our most concentrated tests. Like Prentice (1986), non-parametric bootstrapping produces coverage rates that are too small for small n but accurate for larger n . The parametric bootstrap confidence regions are consistently larger than the non-parametric ones. They are variously too small, accurate, and too large. Overall, MCMC is the most consistently accurate of the six methods, in that its 95% credible regions approach 0.95 coverage for most sample sizes and concentrations. However, its regions are conspicuously too small in the eigenvalue-(1, 1, 7) tests of Fig. 11B. The reason may be that those data sets, being anisotropic, are poorly fit by the wrapped trivariate normal distribution, which is isotropic. Nevertheless, the method performs well on the anisotropic data sets of Fig. 11C.

Our second experiment is similar to the first, but now the rotations are subject to line-in-plane symmetry (Section 2.2). We test only two methods: MCMC and non-parametric bootstrapping. MCMC is chosen because it fares well in the first experiment. Non-parametric bootstrapping fares less well there, but it may be attractive to practitioners because of its ease of implementation. Both methods are easily adapted to handle symmetry. The coverage rates are shown in the right column of Fig. 11. The bootstrap confidence regions are inaccurate for small sample sizes and small concentrations. MCMC simulation is quite accurate in most of the tests. However, this

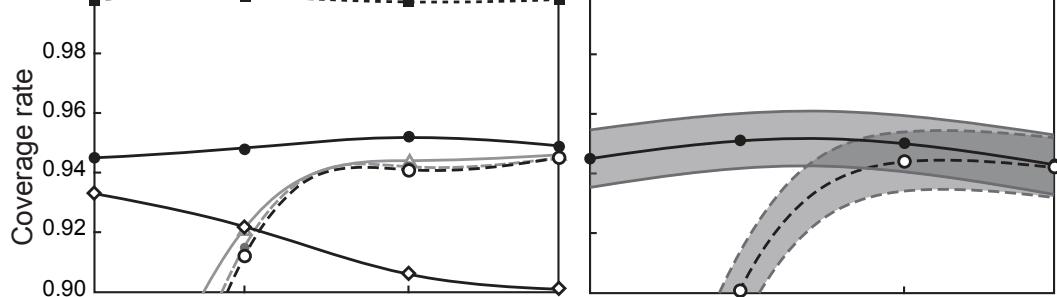
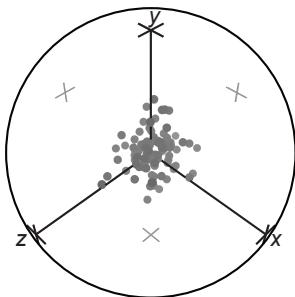
A eigenvalues 1, 1, 1



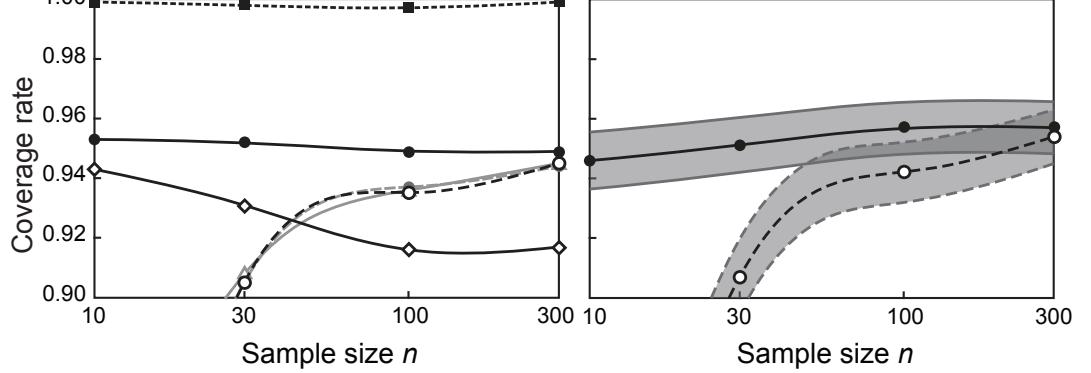
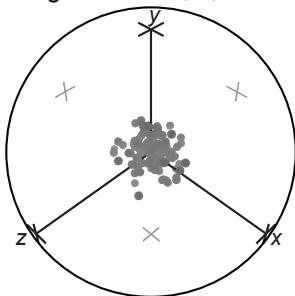
B eigenvalues 1, 1, 7



C eigenvalues 1, 7, 7



D eigenvalues 7, 7, 7



KEY	---□---	Downs (1972)	-○-	non-parametric bootstrapping
	—◇—	Rancourt et al. (2000)	—△—	parametric bootstrapping
	—●—	Prentice (1986)	—●—	Markov chain Monte Carlo

Figure 11: Coverage rates for methods of inference at the 95% confidence/credible level. A. Data sets drawn from a matrix Fisher distribution where \mathbf{K} has eigenvalues 1, 1, 1. Equal-volume plot of an example data set. Coverage rates of six methods for rotations (first experiment). Coverage rates of two methods for orientations with line-in-plane symmetry (second experiment), with their 95% margins of error. B. The same plots for eigenvalues 1, 1, 7. C. The same plots for eigenvalues 1, 7, 7. D. The same plots for eigenvalues 7, 7, 7. In the first two plots of the first experiment, the Rancourt et al. (2000) method plots below the bottom of the graph.

method again performs poorly on eigenvalue-(1, 1, 7) data sets (right side of Fig. 11B).

Our third experiment is identical to the second, but focusing on one specific case: $n = 23$ and eigenvalues $0.000003, 11, 33$. The 95% coverage rates are 0.887 ± 0.013 for non-parametric bootstrapping and 0.926 ± 0.011 for MCMC. Therefore the 95% confidence/credible regions are somewhat too small. The 99% confidence/credible regions produce coverage rates of 0.956 ± 0.008 and 0.977 ± 0.006 . This experiment informs our analysis of foliation-lineation pairs from the western Idaho shear zone in Section 7.

5.7. Two-sample inference

All of the inference techniques discussed thus far attempt to understand the mean of a single population based on a sample from that population. However, there are situations in which a geologist has two data sets, one from each of two populations, and wants to know whether the two populations are different. There are two common cases.

In some two-sample problems the data are paired. For example, suppose that we have n field stations, each with one measurement of shape-preferred orientation and one measurement of anisotropy of magnetic susceptibility. It is natural to ask whether the two kinds of ellipsoid have the same orientation, or differ in a systematic way, or differ unsystematically. So let $\mathbf{R}_i, \mathbf{Q}_i$ be the orientations of the two ellipsoids at the i th station, and let $\mathbf{S}_i = \mathbf{Q}_i \mathbf{R}_i^\top$ be the difference between the two orientations. A one-sample confidence/credible region for the mean of the \mathbf{S}_i describes the difference between the two kinds. A hypothesis test rejects the null hypothesis that the two populations have the same mean, if \mathbf{I} is outside the confidence/credible region. !!Same symmetry group (Appendix J).

In other problems the data are not paired. For example, suppose that we have n foliation-lineations $\mathbf{R}_1, \dots, \mathbf{R}_n$ observed in one rock unit and m foliation-lineations $\mathbf{Q}_1, \dots, \mathbf{Q}_m$ observed in an adjacent unit. We wish to know whether the two populations are different, and by how much. Downs (1972) gave an asymptotic method for unpaired two-sample rotation problems, but his method could not handle symmetries. Fortunately, bootstrapping and MCMC can be adapted to such problems easily. We assume that both data sets are independent and identically distributed, and further that the two data sets are independent of each other. We apply our chosen simulation method to the first data set, to generate a large set of rotations $\mathbf{S}_1^1, \dots, \mathbf{S}_1^b$. Then we apply the same method to the second data set, to generate rotations $\mathbf{S}_2^1, \dots, \mathbf{S}_2^b$. Finally we compute the difference rotation-by-rotation: $\mathbf{S}^j = \mathbf{S}_2^j (\mathbf{S}_1^j)^\top$. These \mathbf{S}^j characterize the uncertainty in the estimated difference between the means of the two populations. They can be processed as in Section 5.5 to construct confidence/credible regions and to perform hypothesis tests.

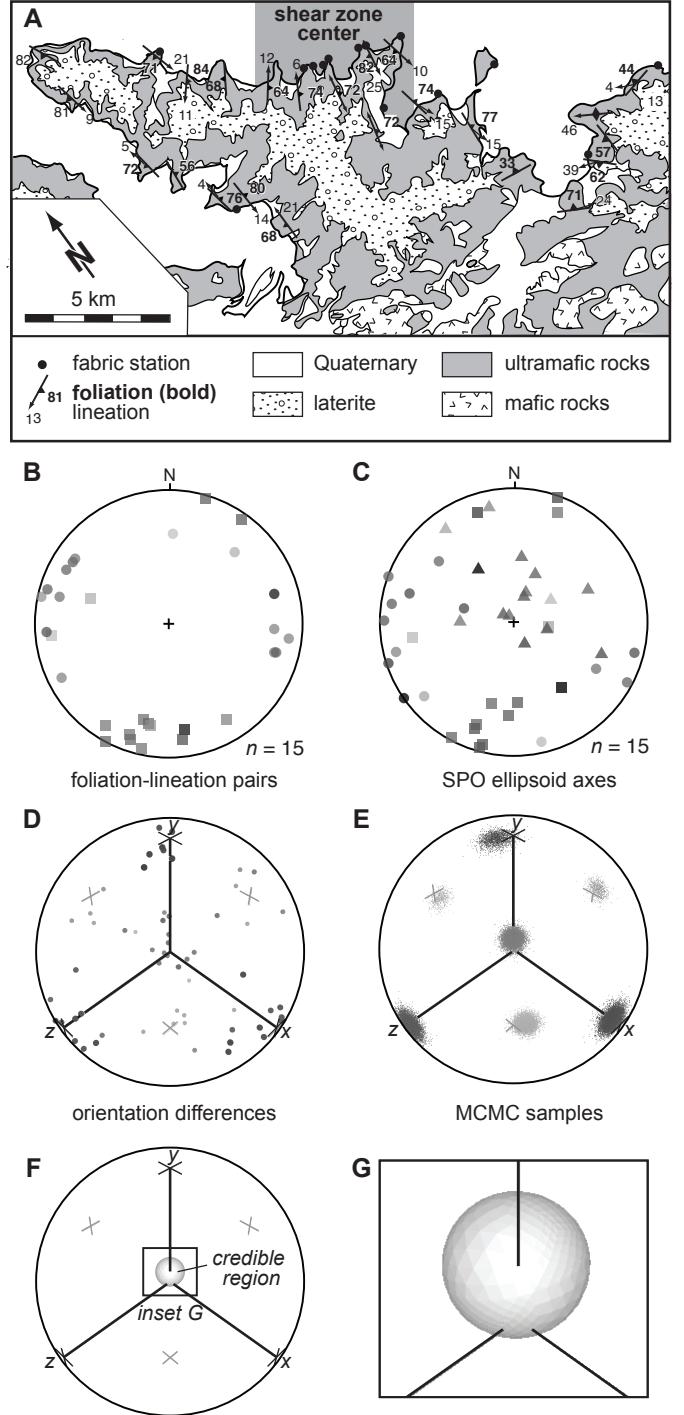


Figure 12: Comparing two kinds of fabric data from the Bogota Peninsula shear zone, New Caledonia. A. Simplified map of the field area, showing the 15 stations (modified from Titus et al. (2011)). B. Foliation-lineation pairs, shaded by position along the transect, from dark (northwest) to light (southeast). C. Orientations of SPO ellipsoids, with matching shading. D. The difference in orientation between SPO and foliation-lineation. E. A representative subset of 10,000 MCMC samples, forming a tight ellipsoidal cluster. F. The ellipsoidal 95% credible region, engulfing the origin. G. The credible region, magnified.

5.8. Example: New Caledonia fabrics

The Massif du Sud in New Caledonia exposes the mantle section of an ophiolite, which was emplaced onto

the Norfolk Ridge during the Eocene (Cluzel et al., 2006). Prinzhöfer and Nicolas (1980) identified a major shear zone within the ophiolite on the Bogota Peninsula (Fig. 12A), which they interpreted as a paleotransform fault. This inference was based on increasing fabric strength and the rotation of foliations over a 30 km-wide region, from shallow NW-striking outside the shear zone to subvertical NS-striking inside the shear zone. Those authors also noted symmetric fabric rotation centered on a 3 km-wide high-strain core.

Titus et al. (2011) presented field data across this shear zone, including from the central ~ 15 km reproduced in Fig. 12. Data from 15 stations included the orientation of field fabrics, i.e., foliations and lineations (Fig. 12B), as well as the shape preferred orientation (SPO) of macroscopic orthopyroxene grains (Fig. 12C). The SPO ellipsoids came from outcrop tracings on multiple faces and were calculated following the method of Robin (2002); Launeau and Robin (2005).

One might expect the two orientations to agree: lineations with the long axes of SPO and foliation poles with the short axes. This idea can be tested using a paired two-sample inference. At each station we compute the rotation $\mathbf{Q}_i \mathbf{R}_i^\top$ taking the foliation-lineation \mathbf{R}_i to the SPO orientation \mathbf{Q}_i (Fig. 12D). If these orientation differences formed a cluster centered on \mathbf{I} , which is at the origin of the equal-volume plot, then that would corroborate our expectation that the two kinds of orientation agree. In this case, they instead form a loose cluster near the origin but not clearly engulfing it. The plot is ambiguous.

Fig. 12BC are shaded by distance along the transect. If these lines formed a clear dark-to-light pattern, then that would indicate systematic spatial variation. In this case, we see no clear pattern. Nor do we see patterns when we shade Fig. 12BCD according to other quantities, such as distance from the center of the shear zone (not shown). So it is reasonable to proceed under the assumption that the orientation differences are independent and identically distributed.

After choosing representative rotations near the origin, we estimate the matrix Fisher parameters (Section 4.7) and find that $\hat{\mathbf{K}}$ has eigenvalues 1.4, 5.0, 5.4. Among the experiments performed in Section 5.6, this situation most closely resembles the eigenvalue-(1, 7, 7), $n = 10$ case shown on the right side of Fig. 11C. In that case, MCMC simulation performs well, while bootstrapping produces confidence regions that are much too small. So we use MCMC to generate 10^7 possible means for the population of orientation differences. Fig. 12E shows a representative 10,000 of them. Because they form a tight ellipsoidal cloud, inference based on percentiles of Mahalanobis distance can proceed. Fig. 12FG shows one symmetric copy of the 95% credible region.

The credible region contains the origin. As a null hypothesis, \mathbf{I} is assigned a p -value greater than 0.1. (For reasons of efficiency, our software does not report precise

p -values when they exceed 0.1.) So we do not reject the hypothesis that the two kinds of fabric have the same orientation. The test is inconclusive. Nevertheless, the credible region offers some insight into the relationship between the two fabrics. For example, the entire credible region is within distance $0.73 = 42^\circ$ of \mathbf{I} . So, with 95% credibility, the two kinds of fabric differ by no more than 42° in orientation.

6. Regression

6.1. Motivation

The inference methods of Section 5 assume that the observed data are independently drawn from a single population (or, in Section 5.7, two populations). However, geologic data often exhibit spatial, temporal, or other dependencies that violate this assumption (e.g., Borradaile, 2003; Wackernagel, 2003). Such dependencies may require treatment by kriging, time series, or other methods. In lieu of those topics, this section introduces another basic statistical tool that can be used for exploring how rotations and orientations depend on other quantities: regression.

In elementary statistics, a common task is to fit a line $y = mx + b$ to some data points $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where the x_i are certain but the y_i are uncertain (e.g., Sen and Srivastava, 1990). The advantage of restricting one's attention to lines is that, when a dependency is found, it is easy to interpret: For each unit change in x , y changes by m units. The disadvantage of using lines is that they are capable of describing only the simplest, least subtle dependencies.

Analogously, in this section we assume data $\{(s_i, \mathbf{R}_i), \dots, (s_n, \mathbf{R}_n)\}$, where the s_i are certain scalars and the \mathbf{R}_i are uncertain rotations. Our first method, geodesic regression, fits the simplest kind of curve: a steady progressive rotation. Our second method, kernel regression, fits potentially complicated curves that are not easily interpreted. Both methods let us quantify the dependence of \mathbf{R} on s , ascertain whether it is statistically significant, and extrapolate from the data to predict missing values.

6.2. Geodesic regression

Given data $\{(s_i, \mathbf{R}_i)\}$, we seek the geodesic curve

$$\mathbf{R} = \exp(\mathbf{M}s)\mathbf{B}$$

(Appendix A and Appendix C.1) that best expresses the dependence of the \mathbf{R}_i on the s_i . That is, we seek the rotation matrix \mathbf{B} and anti-symmetric matrix \mathbf{M} that minimize the sum of squared residuals

$$E = \frac{1}{2n} \sum_{i=1}^n d(\mathbf{R}_i, \exp(\mathbf{M}s_i)\mathbf{B})^2. \quad (1)$$

These matrices \mathbf{M} and \mathbf{B} can be computed using numerical optimization methods. See Appendix C.3 for details.

As in Euclidean linear regression, the amount of variance explained by the curve can be measured by a statistic

$$R^2 = 1 - E/\Psi(\bar{\mathbf{R}}).$$

This statistic satisfies $0 \leq R^2 \leq 1$, with $R^2 = 1$ indicating a perfect fit to the data. Fletcher (2013) described a hypothesis test for whether the direction \mathbf{M} is different from zero — that is, whether the tendency discovered by the regression is statistically significant. Intuitively, the relationship between the s_i and the \mathbf{R}_i is meaningless if a random permutation of the s_i yields just as large of an R^2 statistic. Therefore, we perform many (say, 1,000 or 10,000) trials, each of which proceeds by randomly permuting the s_i , performing the regression, and computing R^2 . Let p be the fraction of these R^2 values that are greater than the original R^2 value. Small p -values indicate statistical significance.

6.3. Example: Cyprus paleomagnetic rotations revisited

We return now to the data set of Section 3.4, which consists of rotations derived from paleomagnetic data and dike measurements from the Troodos ophiolite in Cyprus. Recall from Fig. 7 that the rotations seem to exhibit a systematic dependence on easting. We can quantify this tendency using regression. Compared to the treatment of Section 6.2, there is one complication: Most of the rotations \mathbf{R}_i are actually pairs of candidate rotations \mathbf{R}_i and \mathbf{R}'_i , with neither clearly to be preferred over the other. To resolve this ambiguity, we measure the fitness of a geodesic using whichever representative of the pair is closer to the geodesic. In other words, the misfit $d(\mathbf{R}_i, \exp(\mathbf{M}s_i)\mathbf{B})$ in Eq. (1) is replaced by

$$\min(d(\mathbf{R}_i, \exp(\mathbf{M}s_i)\mathbf{B}), d(\mathbf{R}'_i, \exp(\mathbf{M}s_i)\mathbf{B})).$$

Intuitively, this notion of misfit seeks the simplest and most consistent pattern of rotation. The scalars $s_i < 0$ are eastings in km relative to the town of Mandria and the Solea graben.

Fig. 13A displays the results of the regression. Only the rotation chosen to represent each pair is shown. The variance of these rotations about their Fréchet mean is $\Psi(\bar{\mathbf{R}}) = 0.44$. The best-fit geodesic is $\mathbf{R}(s) = (\exp \mathbf{M}s)\mathbf{B}$, where

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} 0 & -0.153 & -0.163 \\ 0.153 & 0 & -0.052 \\ 0.163 & 0.052 & 0 \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} -0.276 & -0.611 & 0.742 \\ -0.759 & -0.335 & -0.558 \\ 0.590 & -0.717 & -0.372 \end{bmatrix}. \end{aligned}$$

The tensor \mathbf{M} describes instantaneous rotation about

$$\mathbf{u} = [0.052 \quad -0.163 \quad 0.153]^\top.$$

That is, for each km of easting the regression predicts $|\mathbf{u}| = 0.230 = 13^\circ$ of rotation about an axis with trend-plunge

$(162^\circ, -42^\circ)$. The station farthest from the graben has experienced 158° of this rotation.

The quality of the fit is measured by $R^2 = 0.54$. A permutation test with 10,000 trials yields 11 trials with larger R^2 . Therefore $p = 0.001$ with ± 0.001 as a 95% margin of error. The result is significant at the 95% confidence level.

We can use the regression curve $\mathbf{R}(s)$ to predict rotations throughout the region northwest of Mandria. For example, the predicted rotation at the graben is $\mathbf{R}(0) = \mathbf{B}$. The equation $\mathbf{B}\mathbf{t} = \mathbf{m}$ implies that \mathbf{m} has trend-plunge $(16^\circ, 44^\circ)$. Similarly, for any vector \mathbf{d} , $\mathbf{B}\mathbf{d}$ is horizontal if and only if \mathbf{d} is perpendicular to

$$\mathbf{B}^\top \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.942 \\ 0.327 \\ -0.080 \end{bmatrix}.$$

Future measurements of paleomagnetic directions \mathbf{m} and dike poles \mathbf{d} at the graben could be compared to these predictions.

Concerned that the dike assumption performs poorly in the western part of the study area, we repeat the analysis with the four westernmost stations omitted. In brief, the regression predicts 11° of rotation per km of easting, about an axis with trend-plunge $(172^\circ, -33^\circ)$. The goodness of fit is $R^2 = 0.31$, and the result is significant with $p = 0.022 \pm 0.009$ based on 1,000 permutations. Extrapolating to the westernmost station in the study area, the regression predicts 134° relative to the graben.

6.4. Kernel regression

Another technique for fitting a curve to data, called *kernel regression*, was adapted to $\text{SO}(3)$ by Davis et al. (2010) (see also Jabłoński (2011)). The data $\{(s_1, \mathbf{R}_1), \dots, (s_n, \mathbf{R}_n)\}$ are assumed to be drawn independently from some unknown probability distribution on the space of all pairs (s, \mathbf{R}) . For any fixed s , there is an unknown conditional distribution for \mathbf{R} given that value of s . Kernel regression uses the data to estimate the expected value \mathbf{R} of this conditional distribution. As various values s are tried, various rotations \mathbf{R} are produced. In this way, the regression curve is described as a sequence of points $\mathbf{R} = \mathbf{R}(s)$.

See Appendix C.3 for details. Here we mention only that the smoothing of the curve depends on a number $h > 0$, called the *bandwidth*. This number can be chosen by the user, but Davis et al. (2010) described an algorithm for choosing h automatically.

As for geodesic regression, one can define a statistic $R^2 = 1 - E/\Psi(\bar{\mathbf{R}})$ in terms of the residual sum of squares

$$E = \frac{1}{2n} \sum_{i=1}^n d(\mathbf{R}_i, \mathbf{R}(s_i))^2.$$

Significance of the result can be tested using a permutation test. On each trial, the s_i are permuted, the bandwidth h is computed, the rotations $\mathbf{R}(s_i)$ are predicted at the permuted s_i , and R^2 is computed from them. The p -value

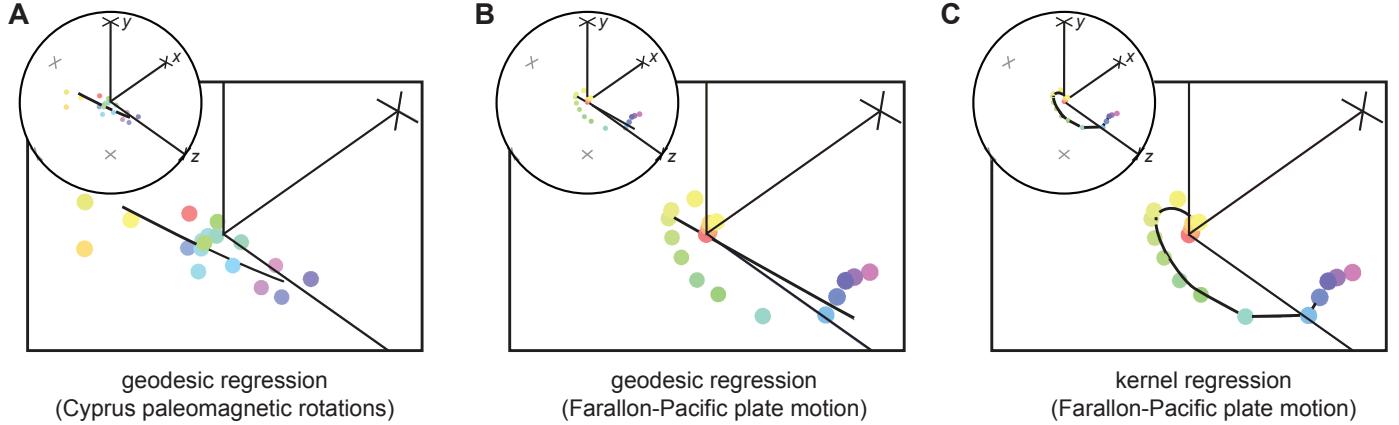


Figure 13: Regression. A. Equal-angle rotation plot of geodesic regression of the Cyprus paleomagnetic rotations, showing only the rotations that were selected by the curve. B. Equal-area plot, showing rotation axes labeled with angles of rotation about those axes. C. Equal-angle plot of Farallon-Pacific relative plate motions, colored by time (present in red). The geodesic regression curve is superimposed. D. With kernel regression superimposed.

of such a test is the proportion of trials that produce R^2 greater than the R^2 for the un-permuted data.

Kernel regression is capable of expressing dependencies that are much more complicated than geodesics. The disadvantage of this kind of regression is that the result is difficult to summarize verbally. One cannot make statements such as, “For each unit change in s , this much rotation occurs,” because such statements describe only geodesic curves. One cannot report best-fit parameter values, because there are no parameters. However, the method is still useful for interpolating among data, for example.

6.5. Example: Farallon and Pacific plates

Engebretson et al. (1984) reported data on the motions of tectonic plates over the past 163 million years. Prentice (1987) analyzed their data for the Farallon-Pacific cumulative relative motion, using a curve-fitting technique derived from directional statistics. These data are shown in Fig. 13B, colored by time. The superimposed curve is the best-fit geodesic. This regression produces $R^2 = 0.56$, and a 1,000-permutation test produces $p = 0.000$. So the geodesic captures some of the tendency in the data, with statistical significance.

However, Fig. 13B suggests that the geodesic does not really capture the smooth curve of the data well. Fig. 13C shows the result of kernel regression with the same data. The kernel is the standard normal density, and the bandwidth $h = 0.040$ is chosen by the Davis et al. (2010) algorithm. This curve fits the data much more closely than does the geodesic. The regression produces $R^2 = 0.98$, and a 1,000-permutation test produces $p = 0.000$.

7. Application: Idaho foliation-lineations

In this section we work a detailed example, which is intended to demonstrate how the techniques of this paper

could be applied to a problem of current research. Occasionally we make side remarks or compute extraneous statistics, not because they are essential to the application, but merely as further illustration of points made earlier.

7.1. Problem

The western Idaho shear zone follows the boundary between accreted terranes and the North American craton (Fig. 14A). This NS-striking shear zone developed solid-state fabrics during the Late Cretaceous (Giorgis et al., 2008). The shear zone fabrics were reactivated during Basin and Range extension in the Miocene (Tikoff et al., 2001).

Fig. 14B shows foliation-lineation pairs measured by Giorgis and Tikoff (2004) at 23 stations in the western Idaho shear zone. The measurements are from granodiorites of the Little Goose Creek Complex. These field fabrics were corrected for Miocene-recent extension, resulting in nearly vertical foliations, parallel to the shear zone boundaries, and steep lineations.

Using these field fabric orientations and consistent shear sense indicators, Giorgis and Tikoff (2004) modeled deformation in this shear zone as dextral, monoclinic transpression along a vertical, NS-striking shear plane. In geographic coordinates aligned with east, north, and up, the velocity gradient tensor is

$$\mathbf{L} = \begin{bmatrix} \log k & 0 & 0 \\ \gamma & 0 & 0 \\ 0 & 0 & -\log k \end{bmatrix},$$

where $\gamma > 0$ is the magnitude of dextral simple shear and $\log k < 0$ describes the shortening (Fossen and Tikoff, 1993). This kind of deformation predicts foliation approximately parallel to the shear plane. The lineation direction depends on the size of γ relative to the quantity

$$\frac{-(\log k)(k+1)\sqrt{k^2+1}}{k}$$

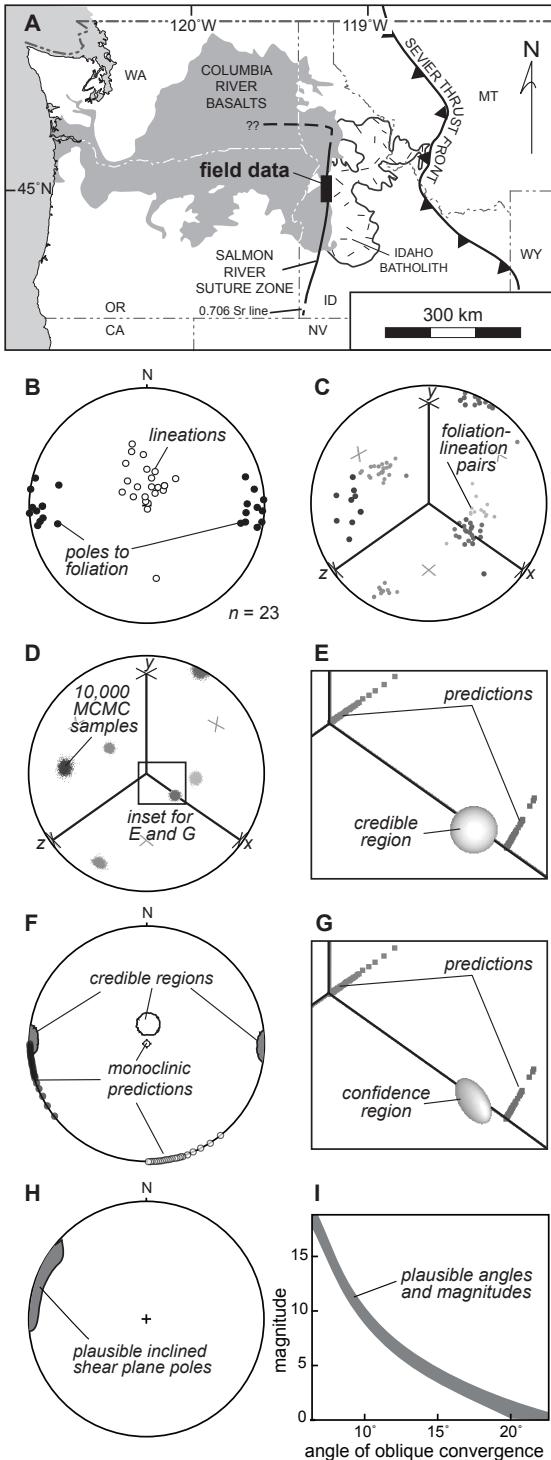


Figure 14: Foliation-lineation pairs from the western Idaho shear zone. A. Map of the surrounding region, modified from Giorgis and Tikoff (2004). B-C. The 23 observed foliation-lineation pairs, in equal-area and equal-volume plots. D. A representative 10,000 samples from the MCMC simulation. They form a tight, ellipsoidal cloud. E-F. The 95% credible region in equal-volume and equal-area. The predictions of the monoclinic model are superimposed. G. The 95% confidence region from bootstrapping. H. Plausible inclined shear zone poles. I. Plausible inclined angles of oblique convergence and magnitudes.

(Appendix L). If γ is less than that quantity, then lineation is vertical. If γ is greater, then lineation is horizontal and approximately parallel to the shear plane. If they are equal, then the finite strain ellipsoid is an oblate spheroid and lineation is ill-defined.

Because their observed lineations were approximately vertical, Giorgis and Tikoff (2004) deduced that the monoclinal transpression in the western Idaho shear zone involved relatively little simple shear. Of course, the lineations are not exactly vertical, if only because field data always contain some error. However, these lineations seem to be clustered about a non-vertical direction, which may indicate some other deformation geometry.

So we are led naturally to a geologic question: Could these data have arisen from the proposed model of deformation? A statistical interpretation of this question is: Can the data be explained by random variation about an orientation predicted by the proposed model? In other words, can any predicted foliation-lineation be the mean of the population from which the data were drawn?

7.2. Preliminary analysis

Fig. 14C shows the same data in an equal-volume plot, with each foliation-lineation shown four times due to line-in-plane symmetry (Section 2.2). There is one outlier, which corresponds to the outlier lineation in Fig. 14B. Although the data are dispersed enough that Mahalanobis distance should not be regarded as a precise measure of anisotropy of dispersion, the outlier can be detected automatically using this distance. The outlier's distance is 3.9, while the other data have distances at most 2.9. Lacking any particular reason to disregard this outlier, we keep it in the data set.

As in Sections 5.8 and 6, we plot the foliation-lineation data colored according to various combinations of easting and northing (not shown here). We find no clear geographic patterns. Also, the proposed deformation model is homogeneous and therefore predicts constant foliation-lineation throughout the shear zone. So we proceed under the assumption that the data are independent and identically distributed.

The Fréchet mean of the data set is the foliation-lineation with strike-dip-rake ($001^\circ, 89^\circ, 77^\circ$) and lineation trend-plunge ($5^\circ, 77^\circ$). For each foliation-lineation, we choose the representative rotation that is closest to this mean. The projected arithmetic mean of these rotations is half a degree away from the Fréchet mean. Maximum likelihood estimation of the matrix Fisher parameters produces a concentration matrix $\hat{\mathbf{K}}$ with eigenvalues 0.000003, 11, 33. This concentration informs our inferences in the next subsection.

!Foliation-lineations are subject to orthorhombic bipyramidal symmetry, for which the critical threshold is $\pi/2 \approx 1.57$ (Appendix J). The diameter of the data is 1.70 (or 0.92 without the outlier). So the data are not tightly clustered enough that symmetry can be ignored in computations such as the Fréchet mean. On the other hand, the

diameter is much smaller than the theoretical maximum of $2\pi/3 \approx 2.09$, so our numerical methods should perform well.

7.3. Credible region

Working under the assumption that the foliation-lineations are independent and identically distributed, we wish to understand the population from which they were sampled. Among the inference methods evaluated in Section 5.6, MCMC performs best. So we proceed using that method, keeping in mind that its credible regions appear somewhat too small in the third experiment.

MCMC simulation produces $b = 10^7$ wrapped trivariate normal distribution means $\mathbf{M}^1, \dots, \mathbf{M}^b$. Fig. 14D shows 10,000 of them, chosen at regular intervals. Because they form a tight ellipsoidal cloud, inference based on Mahalanobis distance can proceed. Fig. 14EF shows the 95% credible region in equal-volume and equal-area plots.

The center of the credible region is nearly identical to the Fréchet mean of the data. The region is nearly spherical, with all six vertices approximately 8° away from the center. Over the credible region, foliation strikes range from 173° to 189° and from 354° to 009° . Foliation dips are 81° - 90° , and lineation rakes are 69° - 84° and 96° - 111° . The reader should bear in mind that this kind of summary over-represents the size of the credible region by enclosing it in a box-like region of $\text{SO}(3)$ that is easier to describe. For example, strike-dip-rake ($009^\circ, 81^\circ, 111^\circ$) falls within these ranges, but that foliation-lineation is outside the credible region with $p = 0$.

As an aside, we also compute a 95% confidence region using non-parametric bootstrapping with $b = 10,000$ bootstrap samples (Fig. 14G). This ellipsoid is more eccentric than the MCMC ellipsoid, with radii 4° , 7° , and 11° . We also compute an MCMC credible region and bootstrap confidence region for the data treated as pure rotations, ignoring their line-in-plane symmetry. These regions are similar to the ones computed with symmetry in mind.

7.4. Testing the proposed model

The credible region and Mahalanobis distances computed above let us quickly perform a range of hypothesis tests, for transpressions with $0 \leq \gamma \leq 50$ and $-50 \leq \log k \leq 0$. For each pair $(\gamma, \log k)$, we compute the predicted foliation-lineation orientation \mathbf{Q} and a p -value for the null hypothesis “The mean foliation-lineation is \mathbf{Q} .”

The foliation-lineation predicted by each $(\gamma, \log k)$ lies outside the 95% credible region (Fig. 14EF). The largest p -value attained is $p = 0.0008$. Consequently each test rejects its null hypothesis. In other words, the variation observed in the foliation-lineations cannot arise from random noise (shaped like the wrapped trivariate normal distribution) about a foliation-lineation predicted by homogeneous monoclinic transpression. The data are inconsistent with every proposed deformation.

The third numerical experiment of Section 5.6 suggests that nominal 95% credible regions achieve something

closer to 93% credibility on data sets like this one. Fortunately, the p -values here are so small that the proposed deformations land outside the nominal 99% credible region, which enjoys approximately 98% credibility in the third experiment. So we can still reject the proposed deformations as highly improbable given these data. (The bootstrap confidence region produces similar results, with $p = 0.0001$ being the largest p -value attained. Therefore data like these are improbable under any of the proposed deformations.)

7.5. Alternative models

If the proposed model is inconsistent with the data, then which model would be consistent? In this subsection we consider inclined transpressions (Jones and Holdsworth, 1998; Lin et al., 1998), which can be specified by four parameters: the strike σ and dip δ of the shear zone, the trend τ of the direction along which the blocks approach each other, and a scalar m describing the magnitude of deformation or the rate of approach. Monoclinic transpression is the special case where $\delta = 90^\circ$. We call $\tau - \sigma$ the angle of oblique convergence.

The rotation matrix

$$\mathbf{R} = \begin{bmatrix} \sin \sigma & -\cos \sigma \sin \delta & -\cos \sigma \cos \delta \\ \cos \sigma & \sin \sigma \sin \delta & \sin \sigma \cos \delta \\ 0 & -\cos \delta & \sin \delta \end{bmatrix}$$

describes the relationship between geographic coordinates $[x \ y \ z]^\top$ and local coordinates $[\tilde{x} \ \tilde{y} \ \tilde{z}]^\top$ aligned with the shear plane. Its columns are the vectors pointing along strike, down pole, and up dip, respectively, in geographic coordinates. The scaled approach direction

$$\mathbf{u} = m \begin{bmatrix} \sin \tau \\ \cos \tau \\ 0 \end{bmatrix}$$

transforms to

$$\tilde{\mathbf{u}} = \mathbf{R}^\top \mathbf{u} = m \begin{bmatrix} \cos(\tau - \sigma) \\ -\sin(\delta) \sin(\tau - \sigma) \\ -\cos(\delta) \sin(\tau - \sigma) \end{bmatrix}$$

in local coordinates. The velocity gradient tensor is

$$\tilde{\mathbf{L}} = \begin{bmatrix} 0 & \tilde{u}_1 & 0 \\ 0 & \tilde{u}_2 & 0 \\ 0 & \tilde{u}_3 & -\tilde{u}_2 \end{bmatrix}$$

in local coordinates and $\mathbf{L} = \mathbf{R} \tilde{\mathbf{L}} \mathbf{R}^\top$ in geographic coordinates.

We can now explore which inclined transpressions are consistent with the observed foliation-lineations, by trying combinations of parameters and recording those that predict foliation-lineation within the 95% credible region (Fig. 14HI). We find combinations with strikes ranging from 355° to 043° , dips from 78° to 89° , angles of oblique convergence from 7° to 22° , and m from 0.1 to 18.1. Among these combinations, strike increases with angle of

oblique convergence and decreases with magnitude. The strikes close to the observed 000° imply small angles of oblique convergence and large magnitudes.

For example, one of the plausible transpressions is $\sigma = 000^\circ$, $\delta = 89^\circ$, $\tau - \sigma = 13^\circ$, and $m = 5.2$. This shear plane is NS-striking and nearly vertical. The shortening is 69% and the kinematic vorticity is $w_k = 0.91$. Considering that the present shear zone width is 4 km, this transpression predicts an original width of 13 km and 38 km of dextral offset.

Similarly, one could use the credible region to constrain more complicated models, such as homogeneous triclinic transpression with inclined extrusion (Fernández and Díaz-Azpiroz, 2009) or heterogeneous triclinic transpression (Jaeger, 1962; Robin and Cruden, 1994). We return to the issue of choosing models in Section 8.3.

8. Discussion

8.1. Introductory survey

We have focused on a few of the most elementary tasks in descriptive and inferential statistics: plotting data, computing the mean of a data set, confidence regions for the population mean, and so on. In each case we have described two or three methods, to give an impression of the approaches available. However, the reader should not take this survey as comprehensive or complete, even in the few topics that it covers.

For example, the Kamb contouring of Section 3 is an example of a general problem known as *kernel density estimation* (e.g., Mardia and Jupp, 2000, Section 12.6). The MTEX software package (Schaeben et al., 2007; Hielscher and Schaeben, 2008) performs several kinds of kernel density estimation. Wide-ranging theoretical results on estimating a density from an orientational data set can be found, for example in Hendriks (1990).

In regard to location and dispersion (Section 4), Penne (2009); Stanfill et al. (2013) studied the robustness of the Fréchet and projected arithmetic means and the analogous notions of median. Sommer et al. (2014) defined an exact principal geodesic analysis, to which the method of Section 4.5 is an approximation.

Efron and Tibshirani (1993); Fisher et al. (1996) described improvements to the basic bootstrapping and percentile confidence regions used in Section 5. Bingham et al. (2009); Qiu (2013) performed MCMC simulation for distributions other than the wrapped trivariate normal distribution. Chang and Rivest (2001); Bhattacharya and Patrangenaru (2005); Stanfill et al. (2015) developed still other inference methods for means.

Compared to geodesic regression of Section 6, Prentice (1987); Hinkle et al. (2012) fit more sophisticated curves capable of expressing more subtle dependencies. Prentice (1989); Chang and Rivest (2001) developed regression of paired rotations ($\mathbf{Q}_i, \mathbf{R}_i$), fitting a relationship of the form $\mathbf{R} = \mathbf{A}\mathbf{Q}\mathbf{B}^\top$ for constant \mathbf{A} and \mathbf{B} . Other techniques for

correlation, regression, and analysis of variance may be adaptable from directional statistics (Mardia and Jupp, 2000, Chapter 11).

Although much of the statistical theory is developed for rotations, various disciplines treat orientations with symmetries like those in structural geology. The MTEX software package (Schaeben et al., 2007; Hielscher and Schaeben, 2008) supports a wide variety of symmetries. Seismology uses several of the symmetries from Section 2 (Tape and Tape, 2012; Kagan, 2013). See also Arnold and Jupp (2013) for a detailed statistical treatment of line-in-plane symmetry.

In short, we regard this paper as review of, or an introduction to, some basic concepts of statistics, and a demonstration of how those concepts can inform geologic studies. It is not an authoritative manual of best practices, but rather an invitation to explore further techniques.

8.2. Proposed geologic workflow

While many of this paper's examples illustrate particular statistical concepts in isolation, real geologic problems encountered by practicing geologists will usually require a combination of methods. Our application to Idaho foliation-lineation pairs (Section 7) demonstrates a process of inquiry, starting with exploratory data analysis, leading to the framing of a testable question, the computation and interpretation of answers, and the formulation of new questions. With the help of Fig. 15, we now make this proposed workflow more explicit.

When confronting a data set, the geologist first considers the type of data: scalars, vectors, directions, orientations, ellipsoids, etc. Frequently data sets contain multiple types. This paper focuses narrowly on orientational data, so that it can develop the statistics in more depth.

Once the data are recognized as orientations, the next issue is symmetry. Consider, for example, the orientations of sandstone beds containing ripples. At first glance, it may seem that a bed is a plane and a flow direction is a line in that plane. The symmetry group is then the $\{\mathbf{I}, \mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3\}$ of Section 2.2. However, if the ripples are asymmetric, then the flow direction has a preferred sense, and the symmetry reduces to the $\{\mathbf{I}, \mathbf{J}_2\}$ of Section 2.3. If the younging direction of the bed is also known, then the bedding plane has a preferred pole, and the symmetry reduces to the trivial $\{\mathbf{I}\}$ of Section 2.5. If the younging is known but the ripples are symmetric, then the symmetry is $\{\mathbf{I}, \mathbf{J}_1\}$, which is not among the examples treated in Section 2.

The next step is to plot the data set to get an impression of its most important features. Do the data contain multiple regions of high density, as in Fig. 5D? The equal-volume plot and Kamb contouring (Section 3.5) can help detect them. Are the orientations independent and identically distributed? The geologist plots them colored by other variables, such as location and rock unit, to get an idea of whether they might depend on those variables

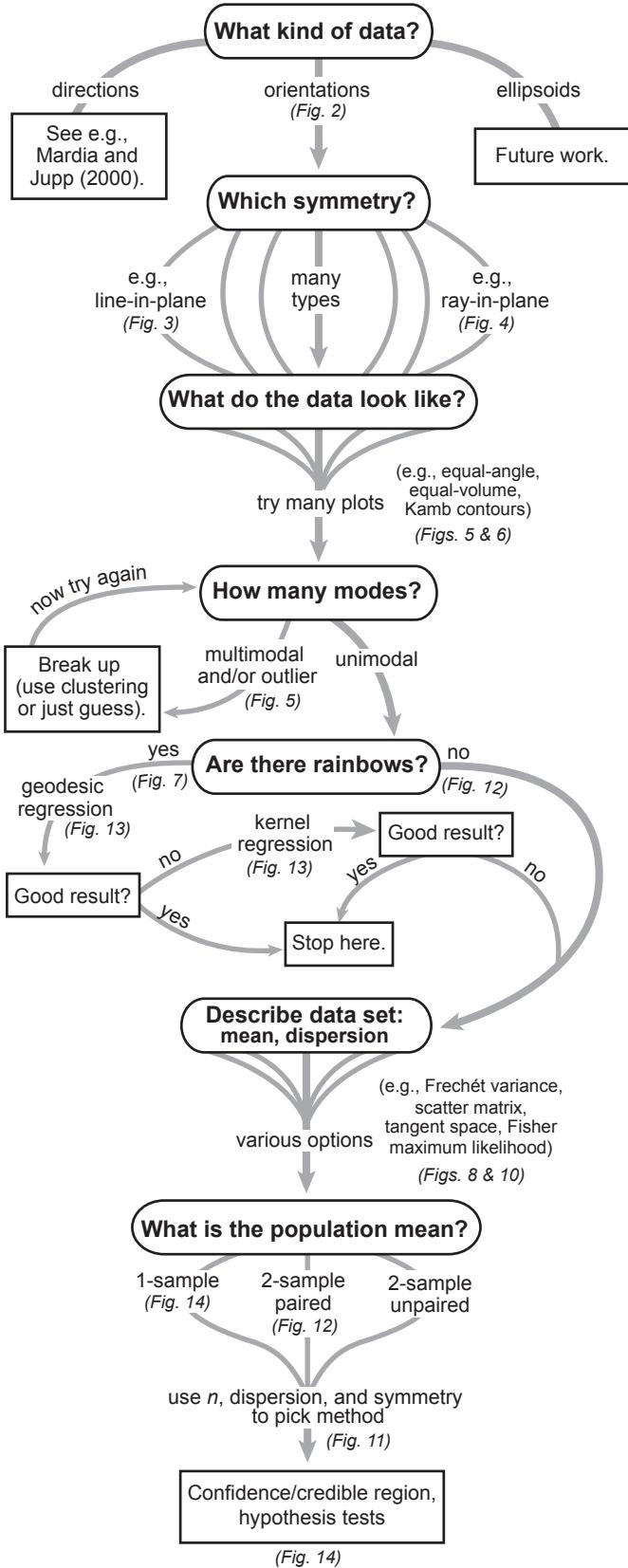


Figure 15: Proposed workflow using the techniques of this paper.

(e.g., Section 3.4). Are the orientations tightly concentrated or widely dispersed? The answer might affect the performance of subsequent analyses (Section 5.6).

When plotting the data, the geologist tries various kinds of plot, because they possess differing strengths. On one hand, an equal-area plot such as Fig. 14B shows foliation and lineation directions in an easily interpretable way. These directions are much harder to read from an abstract rotation plot such as Fig. 14C. On the other hand, the rotation plots are truer to the geometry of the space of orientations. Equal-volume plots portray outliers and multiple modes accurately (Figs. 5 and 6). Equal-angle plots depict the shape of simulation results faithfully (Figs. 12 and 14). Both rotation plots depict entire orientations, not just partial views of certain aspects (e.g., Fig. 7E). Geologists should practice and learn these rotation plots, just as they have learned the direction plots already. The rotation plots complement, rather than replace, the direction plots as tools in the geologist's toolkit.

If the plots show multiple regions of high density or "clumping" of colors, then the data might represent more than one population. The geologist can dissect the data set, either by hand or by clustering algorithms (Appendix C.4), to isolate subsets for further analysis. If the coloration varies smoothly throughout the plot, then that tendency can be assessed using geodesic or kernel regression (Section 6).

Often the variation in the data cannot be explained by multiple populations or other variables. Rather, the variation appears to be random "noise" about a central location. In that case, the data can be summarized using the sample mean and dispersion (Section 4). This paper gives several options, none of which is obviously to be favored as the best practice. At a minimum, some scalar measures of dispersion, such as the eigenvalues of \mathbf{T} or $\hat{\mathbf{K}}$, should be reported, because orientation statistics methods are sensitive to the amount of dispersion in data (Section 5.6). A more detailed report would also include the directions of those dispersions.

In that situation, where the data appear to be independently drawn from a single population, the geologist may wish to infer the mean of that population. The choice of inference method is informed by the symmetry, the sample size n , and the dispersion (Fig. 11). In most cases, MCMC simulation seems preferable. However, in cases resembling our eigenvalue-(1, 1, 7) tests, any one of MCMC, parametric bootstrapping, non-parametric bootstrapping, and Prentice (1986) might be warranted. The output of the computation is a confidence region or credible region and a p -value for any chosen null hypothesis about the population mean. When performing a hypothesis test, the geologist does not simply report rejection or failure to reject. Instead, she reports her chosen significance level, the p -value, and the confidence/credible region, so that the reader can assess the significance of the result and the size of the effect. She also explains her choice of inference method. An overly conservative method can be trusted

when it rejects a null hypothesis, but it rejects less often than it should. When an overly aggressive method rejects a null hypothesis, that result is accompanied by a caveat.

A hypothesis test is informative only when it rejects the null hypothesis. If the geologist's favored explanation for the data is rejected, then perhaps the explanation should be discarded and a new one sought out. If the favored explanation is not rejected, then the geologist can continue developing it, with the understanding that it still might be falsified by other data in the future. In this way, statistical hypothesis testing is a quantitative manifestation of the scientific method (e.g., Gregory, 2005, p. 1).

Finally, in the past it has been common to publish geologic data graphically, for example as strikes and dips marked on maps. As computational tools mature and disseminate, it will be increasingly valuable to have access to a variety of well-curated data sets. At a minimum, authors could publish their data as numerical tables in appendices. However, data could be more conveniently accessed if they were published in electronic databases, in open file formats, with thorough documentation and complete metadata (e.g., Mookerjee et al., 2015; Walker et al., 2015).

8.3. Statistics versus geology

When we perform statistical calculations on data, the question of whether the results are statistically significant may have little to do with whether they are geologically meaningful or appropriate.

For example, in Section 6.3 we discover a significant dependence of paleomagnetic rotations on easting. This result raises several questions. First, is there any geologic reason why such a tendency should arise in the Troodos ophiolite? How would it happen, physically? If it were produced by domino-style block rotation, which would require a sequence of normal faults, then is there any evidence for these faults in the field? Second, the geodesic regression assumes a constant rotation per km of easting. Is that assumption reasonable, or do we expect an increase or decrease in the amount of rotation, or a change in the axis of rotation? Ideally the regression model would be informed by a dynamical theory of the system. Third, the regression is based on data from sheeted dikes outside the Solea graben, but we use it to extrapolate into the gabbros inside the graben (Fig. 7A). Is that extrapolation reasonable, or should other kinematics take over in the gabbros or the graben?

For another example, in Section 7.5 we exclude monoclinic transpressions as being inconsistent with the Idaho foliation-lineations. Then we catalogue inclined transpressions that are consistent with those data. However, is it plausible, based on field observations and the tectonic context, for the shear zone to be inclined? Some of our transpressions incline the shear zone very little, but others deviate far from verticality. How much deviation is reasonable? What range of strikes is reasonable, and how much

dextral offset is reasonable? Should we worry that our assumption about foliation-lineation aligning with the finite strain ellipsoid might be unrealistic? Should we worry that homogeneous deformations might be too simplistic? What other models should we try?

Statistics is a computational tool that complements, rather than replaces, geologic intuition and judgment. It can help falsify some hypotheses, by showing that the data are incompatible with those hypotheses even after uncertainty is taken into account. However, it does not tell the geologist which hypotheses to test. Determining the appropriate assumptions, model, and constraints remains part of the geologist's creative art.

8.4. Future work

There are several concrete ways in which the methods of this paper could be enhanced and expanded, to improve their utility for structural geologists.

First, MCMC simulation works best when based on a distribution that fits the data. All of this paper's MCMC simulations are based on the wrapped trivariate normal distribution. That kind of MCMC performs well for data drawn from that distribution (Qiu et al., 2014), other isotropic distributions (Fig. 11AD), and even some anisotropic distributions (Fig. 11C). However, it performs poorly for some kinds of anisotropic data (Fig. 11B). Quantile-quantile plots of density vs. distance from the mean (e.g., Tape and Tape, 2012; Kagan, 2013) can help assess goodness of fit, but only for isotropic distributions. Other measures of fitness, and MCMC techniques based on anisotropic distributions such as the matrix Fisher distribution, would allow this paper's methods to be applied to a wider variety of geologic data sets.

Second, throughout this paper we use plots and regressions to assess spatial, temporal, and other heterogeneity in data. However, the geostatistics literature (e.g., Wackernagel, 2003) offers other methods, some of which are adaptable to orientations. Mardia and Jupp (2000, Section 11.5) summarized time series for directions, which apply to rotations via quaternions (Section 4.3 and Appendix A). van den Boogaart and Schaeben (2002b); Guimiaux et al. (2003) developed kriging of directions and rotations. Spatially aware methods seem especially valuable for structural geology studies, because structural data are often irregularly sampled in space. In fact, the sampling is influenced by the availability of outcrops, which is itself influenced by the very structures under study.

Finally, a model of deformation can be informed by various kinds of orientational data, including foliation-lineation pairs (Section 7), crystallographic lattices (e.g., Lister et al., 1978; Hansen et al., 2014), and systems of fault slips (Marrett and Allmendinger, 1990; Cladouhos and Allmendinger, 1993; Gapais et al., 2000, e.g.). However, a geologic data set often contains other kinds of data constraining the same deformation. For example, our analysis in Section 7 ignores accompanying shape preferred orientation ellipsoids of feldspar clasts, which were actually

the focus of Giorgis and Tikoff (2004). In previous work, we have demonstrated how to integrate multiple data types into the computation of least-squares best-fit deformations (Davis and Titus, 2011; Davis et al., 2013; Davis and Giorgis, 2014). However, those studies did not rigorously assess the uncertainty in their results. Because of the highly nonlinear nature of the models, there was no readily available theory giving confidence regions for the parameters, and it was not possible to propagate large uncertainties in the data through the fitting process.

We proposed a practical way forward in Davis et al. (2015). The least-squares best fit can be recast as a maximum likelihood estimate. This likelihood-based formulation, when combined with a suitable *prior distribution* for the parameters (Appendix H), is amenable to Bayesian MCMC simulation. The advantage of MCMC is that it produces not just a single best fit, but rather a probability distribution of parameter values. From this distribution, credible regions can be constructed and hypothesis tests performed, much as in Sections 5 and 7. Further, the Bayesian logic behind the MCMC lets us compare the relative probabilities of differing models (Gregory, 2005, Section 12.7). For example, when choosing between a simple model and a more complicated model that describes the data better, one can evaluate whether the better description is worth the extra complication. We intend to develop and apply this approach in future work.

9. Conclusion

This paper surveys concepts and techniques of orientation statistics, with explicit applications for the practicing structural geologist. Except for a few modest contributions — Kamb contouring, anisotropic confidence/credible regions, and numerical experiments — all of the methods discussed are already available in the literatures of various disciplines. We hope that summarizing this body of knowledge in the structural geology literature will speed its adoption there.

Structural studies often make claims such as “the faults are differently oriented on the two sides of the fold” or “the foliations and lineations steepen with proximity to the center of the shear zone.” Such claims can be supported or refuted with statistical argumentation. The fact that geologic data sets are often small and dispersed only increases the importance of applying statistics. In some situations it may not be possible to produce rigorous conclusions with an acceptable level of certainty. However, a culture should be fostered, in which rigor is expected by default.

10. Acknowledgments

We thank Vasileios Chatzaras, Cameron Davidson, Bridget Garnier, Andrew Horst, Zachary D. Michels, Nicolas Roberts, Tor Stetson-Lee, and Basil Tikoff for helpful

geologic discussions, Joseph Pedtke, Samir Rachid Zaim, Alex Rusciano, Katie St. Clair, John Stromme, and Bill Titus for helpful statistical discussions, and Michael Tie for computing support. This research was funded by NSF-EAR 1119606 (Davis) and NSF-EAR 1151851 (Titus). Computing resources were funded in part by a grant from the Howard Hughes Medical Institute to Carleton College.

11. References

- Allerton, S., Vine, F.J., 1987. Spreading structure of the Troodos ophiolite, Cyprus: Some paleomagnetic constraints. *Geology* 15, 593–597.
- Altmann, S.L., 2005. Rotations, quaternions, and double groups. Courier Corporation.
- Arnold, R., Jupp, P., 2013. Statistics of orthogonal axial frames. *Biometrika* 100, 571–586.
- Asgharbeygi, N., Maleki, A., 2008. Geodesic k-means clustering, in: 19th International Conference on Pattern Recognition, IEEE. pp. 1–4.
- Bachmann, F., Hielscher, R., Jupp, P.E., Pantleon, W., Schaeben, H., Wegert, E., 2010. Inferential statistics of electron backscatter diffraction data from within individual crystalline grains. *Journal of Applied Crystallography* 43, 1338–1355.
- Bestmann, M., Prior, D.J., 2003. Intragranular dynamic recrystallization in naturally deformed calcite marble: diffusion accommodated grain boundary sliding as a result of subgrain rotation recrystallization. *Journal of Structural Geology* 25, 1597–1613.
- Bhattacharya, R., Patrangenaru, V., 2005. Large sample theory of intrinsic and extrinsic sample means on manifolds: II. *Annals of statistics* 33, 1225–1259.
- Bingham, M.A., Vardeman, S.B., Nordman, D.J., et al., 2009. Bayes one-sample and one-way random effects analyses for 3-D orientations with application to materials science. *Bayesian Analysis* 4, 607–629.
- Bonhommet, N., Roperch, P., Calza, F., 1988. Paleomagnetic arguments for block rotations along the Arakapas fault (Cyprus). *Geology* 16, 422–425.
- van den Boogaart, K.G., Schaeben, H., 2002a. Kriging of regionalized directions, axes, and orientations I: directions and axes. *Mathematical Geology* 34, 479–503.
- van den Boogaart, K.G., Schaeben, H., 2002b. Kriging of regionalized directions, axes, and orientations II: orientations. *Mathematical Geology* 34, 671–677.
- Borradaile, G.J., 2003. Statistics of Earth Science Data. Springer-Verlag, Berlin.
- Bunge, H.J., 2013. Texture analysis in materials science: mathematical methods. Elsevier.
- Chang, T., Rivest, L.P., 2001. M-estimation for location and regression parameters in group models: A case study using Stiefel manifolds. *Annals of statistics* 29, 784–814.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Chikuse, Y., 2003. Statistics on Special Manifolds. Springer, New York.
- Cladouhos, T.T., Allmendinger, R.W., 1993. Finite strain and rotation from fault-slip data. *Journal of Structural Geology* 15, 771–784.
- Clube, T.M.M., Robertson, A., 1986. The palaeorotation of the Troodos microplate, Cyprus, in the late Mesozoic-early Cenozoic plate tectonic framework of the eastern Mediterranean. *Surveys in Geophysics* 8, 375–437.
- Cluzel, D., Meffre, S., Maurizot, P., Crawford, A.J., 2006. Earliest Eocene (53 Ma) convergence in the southwest Pacific: evidence from pre-obduction dikes in the ophiolite of New Caledonia. *Terra Nova* 18, 395–402.
- Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2010. Population

- shape regression from random design data. *International Journal of Computer Vision* 90, 255–266.
- Davis, J.R., Giorgis, S., 2014. An inverse approach to constraining strain and vorticity using rigid clast shape preferred orientation data. *Journal of Structural Geology* 68, 337–346.
- Davis, J.R., Titus, S.J., 2011. Homogeneous steady deformation: A review of computational techniques. *Journal of Structural Geology* 33, 1046–1062.
- Davis, J.R., Titus, S.J., Giorgis, S.D., Horsman, E.M., 2015. Markov chain Monte Carlo simulation to assess uncertainty in models of naturally deformed rock, AGU Fall Meeting. T33E-2971.
- Davis, J.R., Titus, S.J., Horsman, E., 2013. Non-steady homogeneous deformations: Computational techniques using Lie theory, and application to ellipsoidal markers in naturally deformed rocks. *Journal of Structural Geology* 56, 142–155.
- De Paor, D.G., 1996. Computation of orientations for GIS—the “roll” of quaternions. *Computer Methods in the Geosciences* 15, 447–456.
- Diaconis, P., Shahshahani, M., 1987. The subgroup algorithm for generating uniform random variables. *Probability in the Engineering and Informational Sciences* 1, 15–32.
- Downs, T., 1972. Orientation statistics. *Biometrika* 59, 665–676.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Elliott, D., 1970. Determination of finite strain and initial shape from deformed elliptical objects. *Geological Society of America Bulletin* 81, 2221–2236.
- Engebretson, D.C., Cox, A., Gordon, R.G., 1984. Relative motions between oceanic plates of the Pacific basin. *Journal of Geophysical Research: Solid Earth (1978–2012)* 89, 10291–10310.
- Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise, in: *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press. pp. 226–231.
- Fernández, C., Díaz-Azpiroz, M., 2009. Triclinic transpression zones with inclined extrusion. *Journal of Structural Geology* 31, 1255–1269.
- Fisher, N.I., Hall, P., Jing, B.Y., Wood, A.T., 1996. Improved pivotal methods for constructing confidence regions with directional data. *Journal of the American Statistical Association* 91, 1062–1070.
- Fletcher, P.T., 2013. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision* 105, 171–185.
- Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging* 23, 995–1005.
- Fletcher, P.T., Venkatasubramanian, S., Joshi, S., 2009. The geometric median on Riemannian manifolds with application to robust atlas estimation. *Neuroimage* 45, S143–S152.
- Fossen, H., Tikoff, B., 1993. The deformation matrix for simultaneous simple shearing, pure shearing and volume change, and its application to transpression-transstension tectonics. *Journal of Structural Geology* 15, 413–422.
- Frank, C., 1988. Orientation mapping. *MRS Bulletin* 13, 24–31.
- Gapais, D., Cobbold, P.R., Bourgeois, O., Rouby, D., de Urreiztieta, M., 2000. Tectonic significance of fault-slip data. *Journal of Structural Geology* 22, 881–888.
- Giorgis, S., McClelland, W., Fayon, A., Singer, B.S., Tikoff, B., 2008. Timing of deformation and exhumation in the western Idaho shear zone, McCall, Idaho. *Geological Society of America Bulletin* 120, 1119–1133.
- Giorgis, S., Tikoff, B., 2004. Constraints on kinematics and strain from feldspar porphyroblast populations, in: Alsop, I., Holdsworth, R., McCaffery, K. (Eds.), *Transport and flow processes in shear zones*. Geological Society, London. volume 224, pp. 265–285.
- Goodman, S., 2008. A dirty dozen: twelve p-value misconceptions, in: *Seminars in Hematology*, Elsevier. pp. 135–140.
- Govindu, V.M., 2004. Lie-algebraic averaging for globally consistent motion estimation, in: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE. pp. I–684.
- Gregory, P.C., 2005. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, Cambridge.
- Grove, K., Karcher, H., Ruh, E.A., 1974. Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Mathematische Annalen* 211, 7–21.
- Gumiaux, C., Gapais, D., Brun, J., 2003. Geostatistics applied to best-fit interpolation of orientation data. *Tectonophysics* 376, 241–259.
- Hansen, L.N., Zhao, Y.H., Zimmerman, M.E., Kohlstedt, D.L., 2014. Protracted fabric evolution in olivine: Implications for the relationship among strain, crystallographic fabric, and seismic anisotropy. *Earth and Planetary Science Letters* 387, 157–168.
- Hartley, R., Trumpf, J., Dai, Y., Li, H., 2013. Rotation averaging. *International Journal of Computer Vision* 103, 267–305.
- Hendriks, H., 1990. Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *The Annals of Statistics* 18, 832–849.
- Hielscher, R., Schaeben, H., 2008. A novel pole figure inversion method: specification of the MTEX algorithm. *Journal of Applied Crystallography* 41, 1024–1037.
- Hinkle, J., Muralidharan, P., Fletcher, P.T., Joshi, S., 2012. Polynomial regression on Riemannian manifolds, in: *Computer Vision–ECCV 2012*. Springer, pp. 1–14.
- Hurst, S.D., Verosub, K.L., Moores, E.M., 1992. Paleomagnetic constraints on the formation of the Solea graben, Troodos ophiolite, Cyprus. *Tectonophysics* 208, 431–445.
- Jabłoński, G., 2011. Robust nonparametric regression with output in $\text{SO}(3)$, in: *Computer Recognition Systems 4*. Springer, pp. 127–136.
- Jaeger, J., 1962. *Elasticity, Fracture, and Flow*. John Wiley & Sons, New York.
- Jeffery, G., 1922. The motion of ellipsoidal particles immersed in a viscous fluid. *Proceedings of the Royal Society of London A102*, 161–179.
- Jezek, J., Schulmann, K., Segeth, K., 1996. Fabric evolution of rigid inclusions during mixed coaxial and simple shear flows. *Tectonophysics* 257, 203–221.
- Jones, R.R., Holdsworth, R.E., 1998. Oblique simple shear in transpression zones, in: Holdsworth, R.E., Strachan, R.A., Dewey, J.F. (Eds.), *Continental Transpressional and Transtensional Tectonics*. Geological Society, London. volume 135, pp. 35–40.
- Kagan, Y.Y., 2013. Double-couple earthquake source: symmetry and rotation. *Geophysical Journal International* 194, 1167–1179.
- Kamb, W.B., 1959. Ice petrofabric observations from Blue Glacier, Washington, in relation to theory and experiment. *Journal of Geophysical Research* 64, 18911909.
- Karcher, H., 1977. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics* 30, 509–541.
- Kent, J.T., Ganeiber, A.M., Mardia, K.V., 2013. A new method to simulate the Bingham and related distributions in directional data analysis with applications. arXiv preprint arXiv:1310.8110 .
- Khatri, C.G., Mardia, K.V., 1977. The von Mises-Fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* , 95–106.
- Kline, R.B., 2004. What's wrong with statistical tests—and where we go from here, in: *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. American Psychological Association, Washington, DC, USA, pp. 61–91.
- Koks, D., 2006. *Explorations in Mathematical Physics: The Concepts Behind an Elegant Language*. SpringerLink: Springer e-Books, Springer.
- Krakowski, K.A., Hüper, K., Manton, J.H., 2007. On the computation of the Karcher mean on spheres and special orthogonal groups, in: *Proc. Workshop Robot. Math. (RoboMat07)*.
- Krieger Lassen, N.C., Juul Jensen, D., Conradsen, K., 1994. On the statistical analysis of orientation data. *Acta Crystallographica Section A: Foundations of Crystallography* 50, 741–748.
- Kume, A., Preston, S., Wood, A.T., 2013. Saddlepoint approxima-

- tions for the normalizing constant of Fisher–Bingham distributions on products of spheres and Stiefel manifolds. *Biometrika* 100, 971–984.
- Launeau, P., Robin, P.Y.F., 2005. Determination of fabric and strain ellipsoids from measured sectional ellipsesimplementation and applications. *Journal of Structural Geology* 27, 2223–2233.
- Lay, D.C., 1994. *Linear Algebra and Its Applications*. Addison Wesley, Reading, Massachusetts.
- Lin, S., Jiang, D., Williams, P.F., 1998. Transpression (or transtension) zones of triclinic symmetry: natural example and theoretical modelling. Geological Society, London, Special Publications 135, 41–57.
- Lister, G.S., Paterson, M.S., Hobbs, B.E., 1978. The simulation of fabric development in plastic deformation and its application to quartzite: the model. *Tectonophysics* 45, 107–158.
- MacLeod, C.J., Allerton, S., Gass, I., Xenophontos, C., 1990. Structure of a fossil ridge–transform intersection in the Troodos ophiolite. *Nature* 348, 717–720.
- Manton, J.H., 2004. A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups, in: Control, Automation, Robotics and Vision Conference, 2004., IEEE. pp. 2211–2216.
- Mardia, K.V., Jupp, P.E., 2000. *Directional Statistics*. Wiley, New York.
- Marrett, R., Allmendinger, R.W., 1990. Kinematic analysis of fault-slip data. *Journal of Structural Geology* 12, 973–986.
- Matthies, S., Muller, J., Vinet, G., 1988. On the normal distribution in the orientation space. *Textures and Microstructures* 10, 77–96.
- Michels, Z.D., Kruckenberg, S., Davis, J.R., Tikoff, B., 2015. Determining vorticity axes from grain-scale dispersion of crystallographic orientations. *Geology* 43, 803–806.
- Miles, R.E., 1965. On random rotations in R^3 . *Biometrika* 52, 636–639.
- Moakher, M., 2002. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications* 24, 1–16.
- Monahan, J.F., 2011. *Numerical methods of statistics*. Cambridge University Press.
- Mookerjee, M., Vieira, D., Chan, M., Gil, Y., Pavlis, T., Spear, F., Tikoff, B., 2015. Field data management: Integrating cyberscience and geoscience. *Eos* 96, 18–21.
- Moran, P., 1975. Quaternions, Haar measure, and the estimation of a paleomagnetic rotation, in: Perspectives in Probability and Statistics. Academic Press, San Diego, pp. 295–301.
- Morawiec, A., Field, D., 1996. Rodrigues parameterization for orientation and misorientation distributions. *Philosophical Magazine A* 73, 1113–1130.
- Morris, A., Anderson, M.W., Robertson, A.H., 1998. Multiple tectonic rotations and transform tectonism in an intraoceanic suture zone, SW Cyprus. *Tectonophysics* 299, 229–253.
- Morris, A., Maffione, M., 2016. Is the Troodos ophiolite (Cyprus) a complete, transform fault–bounded Neotethyan ridge segment? *Geology* 44, 199–202.
- Nikolayev, D.I., Savolova, T.I., 1997. Normal distribution on the rotation group $SO(3)$. *Textures and Microstructures* 29, 201–234.
- Passchier, C.W., 1987. Stable positions of rigid objects in non-coaxial flow; a study in vorticity analysis. *Journal of Structural Geology* 9, 679–690.
- Pennec, X., 2009. Statistical computing on manifolds: from Riemannian geometry to computational anatomy, in: Emerging Trends in Visual Computing. Springer, pp. 347–386.
- Prentice, M.J., 1986. Orientation statistics without parametric assumptions. *Journal of the Royal Statistical Society. Series B (Methodological)* 48, 214–222.
- Prentice, M.J., 1987. Fitting smooth paths to rotation data. *Applied Statistics* 36, 325–331.
- Prentice, M.J., 1989. Spherical regression on matched pairs of orientation statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 51, 241–248.
- Prinzhofer, A., Nicolas, A., 1980. The Bogota Peninsula, New Caledonia: a possible oceanic transform fault. *The Journal of Geology* 88, 387–398.
- Provost, A., Buisson, C., Merle, O., 2004. From progressive to finite deformation and back. *Journal of Geophysical Research* 109, doi:10.1029/2001JB001734.
- Qiu, Y., 2013. Isotropic Distributions for 3-Dimension Rotations and One-Sample Bayes Inference. Ph.D. thesis. Iowa State University. Ames, Iowa.
- Qiu, Y., Nordman, D.J., Vardeman, S.B., 2014. A wrapped trivariate normal distribution and Bayes inference for 3-D rotations. *Statistica Sinica* 24, 897–917.
- Rancourt, D., Rivest, L.P., Asselin, J., 2000. Using orientation statistics to investigate variations in human kinematics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49, 81–94.
- Randle, V., Engler, O., 2000. *Introduction to Texture Analysis: Macrotecture, Microtexture and Orientation Mapping*. CRC Press.
- Reddy, S.M., Buchan, C., 2005. Constraining kinematic rotation axes in high-strain zones: a potential microstructural method? Geological Society, London, Special Publications 243, 1–10.
- Robert, C., Casella, G., 2013. *Monte Carlo Statistical Methods*. Springer Science & Business Media.
- Robin, P.Y., Cruden, A.R., 1994. Strain and vorticity patterns in ideally ductile transpression zones. *Journal of Structural Geology* 16, 447–466.
- Robin, P.Y.F., 2002. Determination of fabric and strain ellipsoids from measured sectional ellipses - theory. *Journal of Structural Geology* 24, 531–544.
- Rosca, D., Morawiec, A., De Graef, M., 2014. A new method of constructing a grid in the space of 3d rotations and its applications to texture analysis. *Modelling and Simulation in Materials Science and Engineering* 22, 075013.
- Rowenhorst, D., Rollett, A., Rohrer, G., Groeber, M., Jackson, M., Konijnenberg, P.J., De Graef, M., 2015. Consistent representations of and conversions between 3D rotations. *Modelling and Simulation in Materials Science and Engineering* 23, 083501.
- Schaeben, H., Hielscher, R., Fundenberger, J.J., Potts, D., Prestin, J., 2007. Orientation density function-controlled pole probability density function measurements: automated adaptive control of texture goniometers. *Journal of Applied Crystallography* 40, 570–579.
- Sei, T., Shibata, H., Takemura, A., Ohara, K., Takayama, N., 2013. Properties and applications of Fisher distribution on the rotation group. *Journal of Multivariate Analysis* 116, 440–455.
- Sen, A., Srivastava, M., 1990. *Regression Analysis: Theory, Methods, and Applications*. Springer-Verlag, New York.
- Sommer, S., Lauze, F., Nielsen, M., 2014. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics* 40, 283–313.
- Spivak, M., 1979. *A Comprehensive Introduction to Differential Geometry*. volume 1. Publish or Perish, Houston, Texas.
- Stanfill, B., Genschel, U., Hofmann, H., 2013. Point estimation of the central orientation of random rotations. *Technometrics* 55, 524–535.
- Stanfill, B., Genschel, U., Hofmann, H., Nordman, D., 2015. Non-parametric confidence regions for the central orientation of random rotations. *Journal of Multivariate Analysis* 135, 106–116.
- Strine, M., Wojtal, S.F., 2004. Evidence for non-plane strain flattening along the Moine thrust, Loch Srath nan Aisinn, North-West Scotland. *Journal of Structural Geology* 26, 1755–1772.
- Tape, W., Tape, C., 2012. A geometric setting for moment tensors. *Geophysical Journal International* 190, 476–498.
- Tauxe, L., 2010. *Essentials of Paleomagnetism*. University of California Press.
- Tikoff, B., Kelso, P., Manduca, C., Markley, M., Gillaspy, J., 2001. Lithospheric and crustal reactivation of an ancient plate boundary: The assembly and disassembly of the Salmon River suture zone, Idaho, USA. *Geological Society, London, Special Publications* 186, 213–231.
- Titus, S.J., Maes, S.M., Benford, B., Ferr, E.C., Tikoff, B., 2011. Fabric development in the mantle section of a paleotransform fault and its effect on ophiolite obduction, New Caledonia. *Lithosphere* 3, 221–244, doi: doi: 10.1130/L122.1.

- Titus, S.J., Wagner, C., Alexander, S., Scott, C., Davis, J.R., 2015. Styles of deformation on either side of a ridge-transform intersection, Troodos ophiolite, Cyprus, AGU Fall Meeting, T41E-2947.
- Vollmer, F.W., 1995. Program for automatic contouring of spherical orientation data using a modified Kamb method. Computers & Geosciences 21, 3149.
- Wackernagel, H., 2003. Multivariate Geostatistics. Springer Science & Business Media, Berlin. 3rd edition.
- Walker, J., Tikoff, B., Newman, J., Good, J., Ash, J., Michels, Z., Andrew, J., Williams, R., Moeller, A., Richard, S., 2015. Cyberinfrastructure for structural geology and tectonics, GSA Fall Meeting, p. 820.
- Wolf, F.L., 1974. Elements of Probability and Statistics. McGraw-Hill. 2nd edition.
- Wood, A.T., 1993. Estimation of the concentration parameters of the Fisher matrix distribution on $\text{SO}(3)$ and the Bingham distribution on S_q , $q \geq 2$. Australian Journal of Statistics 35, 69–79.

Appendix A. Representations of rotations

There are various formalisms for describing or representing rotations of three-dimensional space, with varying strengths and weaknesses. In this section we summarize many of the formalisms, including formulas and algorithms for converting among them. For brevity, we ignore special cases in the conversions. In this paper, we use the angle-axis and matrix representations most frequently. Angles and axes are geometrically intuitive, while matrices are often algebraically convenient.

Appendix A.1. Matrix

A rotation is a particular kind of linear transformation of space. After a choice of coordinates on the space, the transformation can be written as a matrix \mathbf{R} . The transformation must preserve lengths of vectors and angles between vectors. In terms of \mathbf{R} , this condition amounts to $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$. That is, $\mathbf{R}^{-1} = \mathbf{R}^\top$. Such matrices are called *orthogonal*. Some have determinant 1, and are called *special orthogonal* or simply *rotation* matrices. We ignore the others, which have determinant -1 and involve reflection of space as well as rotation.

Appendix A.2. Angle and axis

Any rotation can be described by an axis and an angle of rotation about that axis. To be precise and consistent, let us agree that the axis is a unit vector $\mathbf{u} = [u_1 \ u_2 \ u_3]^\top$ and the angle α expresses rotation about that axis in a right-handed sense. To convert an angle-axis representation (α, \mathbf{u}) to a rotation matrix \mathbf{R} , first form the matrix

$$\mathbf{U} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}.$$

Then

$$\mathbf{R} = \mathbf{I} + (\sin \alpha)\mathbf{U} + (1 - \cos \alpha)\mathbf{U}^2. \quad (\text{A.1})$$

To convert \mathbf{R} back to an angle-axis representation, let $\alpha = \arccos((\text{tr } \mathbf{R} - 1)/2)$ and $\mathbf{U} = (2 \sin \alpha)^{-1}(\mathbf{R} - \mathbf{R}^\top)$, and extract \mathbf{u} from \mathbf{U} . As an aside, we note that (α, \mathbf{u}) and

$(-\alpha, -\mathbf{u})$ represent the same rotation, while $(-\alpha, \mathbf{u})$ and $(\alpha, -\mathbf{u})$ represent the inverse. Notice also that α between 0 and π suffice to describe all finite rotations.

Appendix A.3. Infinitesimal rotation

The *matrix exponential* function is defined as

$$\exp \mathbf{A} = \sum_{k=0}^{\infty} \mathbf{A}^k / k! = \mathbf{I} + \mathbf{A} + \mathbf{A}^2/2 + \mathbf{A}^3/6 + \dots$$

(Provost et al., 2004; Davis and Titus, 2011). A matrix \mathbf{W} is *anti-symmetric* if $\mathbf{W}^\top = -\mathbf{W}$. The exponential of any anti-symmetric matrix \mathbf{W} is a rotation matrix $\exp \mathbf{W}$. It can be computed efficiently by setting $\alpha = \sqrt{\text{tr}(\mathbf{W}^\top \mathbf{W})}/2$ and $\mathbf{U} = \mathbf{W}/\alpha$, and proceeding as in Appendix A.2. That is,

$$\begin{aligned} \exp \mathbf{W} &= \mathbf{I} + (\sin \alpha)\mathbf{U} + (1 - \cos \alpha)\mathbf{U}^2 \\ &= \mathbf{I} + \frac{\sin \alpha}{\alpha}\mathbf{W} + \frac{1 - \cos \alpha}{\alpha^2}\mathbf{W}^2. \end{aligned}$$

Conversely, given a rotation \mathbf{R} , any anti-symmetric \mathbf{W} such that $\exp \mathbf{W} = \mathbf{R}$ is called a *logarithm* of \mathbf{R} . Logarithms are not unique, but one can compute the *principal logarithm* of \mathbf{R} , denoted $\log \mathbf{R}$, as follows. Let $\alpha = \arccos((\text{tr } \mathbf{R} - 1)/2)$ and $\mathbf{U} = (2 \sin \alpha)^{-1}(\mathbf{R} - \mathbf{R}^\top)$, and then set $\log \mathbf{R} = \alpha \mathbf{U}$.

The tensor \mathbf{W} represents an infinitesimal rotation. It is an element of the Lie algebra $\mathfrak{so}(3)$, which is the tangent space to $\text{SO}(3)$ at the identity \mathbf{I} . Like $\text{SO}(3)$, the tangent space is three-dimensional, as evidenced by the fact that there are three degrees of freedom in \mathbf{W} , because it is anti-symmetric. The three degrees of freedom can be repackaged as a vector

$$[W_{32} \ W_{13} \ W_{21}]^\top = \alpha \mathbf{u} = \mathbf{w}/2,$$

where (α, \mathbf{u}) is the angle-axis representation and \mathbf{w} is the vorticity vector from Appendix A.4.

Appendix A.4. Vorticity vector

In structural geology, deformations are often decomposed into their distortional and rotational aspects (e.g., Davis and Titus, 2011). A steady homogeneous deformation with velocity gradient tensor \mathbf{L} has stretching tensor $\mathbf{D} = (\mathbf{L} + \mathbf{L}^\top)/2$ and vorticity tensor $\mathbf{W} = (\mathbf{L} - \mathbf{L}^\top)/2$. The finite deformation $\exp \mathbf{L}$ is the simultaneous superposition of the finite distortion $\exp \mathbf{D}$ and the finite rotation $\exp \mathbf{W}$. The rotational aspect is alternatively expressed as the vorticity vector

$$\mathbf{w} = [2W_{32} \ 2W_{13} \ 2W_{21}]^\top.$$

From this point of view, a rotation is simply a deformation with no distortion: $\mathbf{D} = \mathbf{0}$, and $\mathbf{L} = \mathbf{W}$ is anti-symmetric. The vorticity vector \mathbf{w} relates to an angle-axis representation (α, \mathbf{u}) of the rotation via the equation $\mathbf{w} = 2\alpha \mathbf{u}$.

Appendix A.5. Euler angles

Geologists frequently describe rotations in terms of three Euler angles. There are various conventions for these angles, all of which have similar theoretical properties. For example, any rotation can be expressed as a rotation through an angle α_1 about the x_1 axis, followed by a rotation through α_2 about the x_2 axis, followed by a rotation through α_3 about the x_3 axis (all in the right-handed sense). These angle-axis rotations can be expressed as a rotation matrices $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ as above, and the total rotation matrix is then $\mathbf{R} = \mathbf{R}_3 \mathbf{R}_2 \mathbf{R}_1$. We omit the procedure for converting a given rotation matrix \mathbf{R} to its Euler angles, because it is complicated and it depends on the particular convention in use.

Appendix A.6. Quaternion

Angle-axis representations are closely related to quaternions (e.g., Moran, 1975; Prentice, 1986; De Paor, 1996). A *quaternion* is a quantity of the form $a_0 + a_1 i + a_2 j + a_3 k$, where the a_ℓ are real numbers and i, j, k are distinct imaginary units obeying the relations

$$i^2 = j^2 = k^2 = ijk = -1.$$

Multiplication of quaternions is not commutative, but all other basic rules of algebra hold. From these relations follow many others, including

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik.$$

By a common abuse of notation, one can write a quaternion as the formal sum of a scalar and a vector:

$$\begin{aligned} a_0 + a_1 i + a_2 j + a_3 k &\leftrightarrow a_0 + [\begin{array}{ccc} a_1 & a_2 & a_3 \end{array}]^\top \\ &\leftrightarrow a_0 + \mathbf{a}. \end{aligned}$$

In this notation, quaternion multiplication is

$$(a_0 + \mathbf{a})(b_0 + \mathbf{b}) = (a_0 b_0 - \mathbf{a} \cdot \mathbf{b}) + (\mathbf{a} \times \mathbf{b} + a_0 \mathbf{b} + b_0 \mathbf{a}).$$

The norm of a quaternion $a_0 + \mathbf{a}$ is

$$|a_0 + \mathbf{a}| = \sqrt{a_0^2 + |\mathbf{a}|^2} = \sqrt{a_0^2 + a_1^2 + a_2^2 + a_3^2}.$$

A *unit* quaternion is a quaternion of norm 1.

For any unit $a_0 + \mathbf{a}$, the function that takes a vector \mathbf{v} to the vector part of the quaternion

$$\begin{aligned} &(a_0 + \mathbf{a})(0 + \mathbf{v})(a_0 - \mathbf{a}) \\ &= 0 + (-2(\mathbf{a} \times \mathbf{v}) \times \mathbf{a} + 2a_0(\mathbf{a} \times \mathbf{v}) + \mathbf{v}) \end{aligned}$$

is a rotation. This correspondence between unit quaternions and rotations is two-to-one, because $a_0 + \mathbf{a}$ and $-[a_0 + \mathbf{a}]$ (meaning $-a_0 + -\mathbf{a}$) represent the same rotation. More explicitly, the angle-axis rotation (α, \mathbf{u}) corresponds to the two unit quaternions $\pm[(\cos \alpha/2) + (\sin \alpha/2)\mathbf{u}]$. Conversely, the unit quaternions $\pm[a_0 + \mathbf{a}]$ both correspond to the angle-axis rotation $(2 \arccos a_0, \mathbf{a}/\sqrt{1-a_0^2})$. Via this correspondence, $a_0 + \mathbf{a}$ can be converted into other representations, including the rotation matrix

$$\mathbf{I} + 2 \begin{bmatrix} -a_2^2 - a_3^2 & a_1 a_2 - a_0 a_3 & a_0 a_2 + a_1 a_3 \\ a_1 a_2 + a_0 a_3 & -a_1^2 - a_3^2 & a_2 a_3 - a_0 a_1 \\ a_1 a_3 - a_0 a_2 & a_0 a_1 + a_2 a_3 & -a_1^2 - a_2^2 \end{bmatrix}.$$

Appendix B. Lie group

The set of all rotation matrices is called the *special orthogonal group*, denoted $\text{SO}(3)$. The term *group* here indicates that four conditions are satisfied:

- Closure under multiplication: If \mathbf{Q} and \mathbf{R} are rotation matrices, then so is the product \mathbf{QR} .
- Associativity: If \mathbf{Q} , \mathbf{R} , and \mathbf{S} are rotation matrices, then $\mathbf{Q}(\mathbf{RS}) = (\mathbf{QR})\mathbf{S}$.
- Identity: The identity matrix \mathbf{I} is a rotation matrix.
- Closure under inverse: If \mathbf{Q} is a rotation matrix, then so is the inverse \mathbf{Q}^{-1} , which satisfies $\mathbf{QQ}^{-1} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{I}$.

In fact, $\text{SO}(3)$ constitutes not just a group but a *Lie group*. Roughly, this term means that $\text{SO}(3)$ is a smooth, curved “hypersurface” (a *manifold*, in the jargon) within the nine-dimensional space of all 3×3 matrices, and that matrix multiplication and inversion amount to smooth transformations on this smooth space. That is, $\text{SO}(3)$ is a realm where multiplication and geometry both work well, and well with each other. (On the other hand, $\text{SO}(3)$ is not closed under addition and scaling, which is why the arithmetic mean of rotation matrices is often not a rotation matrix.)

The space $\text{SO}(3)$ is three-dimensional, which means that three numbers are required to specify a rotation: three Euler angles, the three components of the vorticity vector \mathbf{w} , the three independent entries of an antisymmetric \mathbf{W} , etc. Topologically, $\text{SO}(3)$ is equal to the *real projective space* \mathbb{RP}^3 , which is the space of lines through the origin in four-dimensional Euclidean space. This space can be constructed by identifying antipodal points on the boundary of a three-dimensional ball. This fact is the basis for the angle-axis plot in Section 3.

Appendix C. Geometry

Appendix C.1. Theory

On the space $\text{SO}(3)$ of rotations there is a canonical notion of differential geometry, called the *bi-invariant Riemannian metric* (e.g., Spivak, 1979, Vol. 1, Ch. 10), which implies well-behaved notions of geodesic curves, distance between points, etc. In this somewhat technical subsection, we review some key aspects of this metric, to support the claims of Section 3 and Appendix C.2. The uninterested or unprepared reader may safely skip to Appendix C.2.

Let $T_{\mathbf{R}}\text{SO}(3)$ denote the tangent space to $\text{SO}(3)$ at a point \mathbf{R} . This tangent space consists of all matrices of the form \mathbf{RW} where \mathbf{W} is anti-symmetric. The metric in $T_{\mathbf{R}}\text{SO}(3)$ is half of the Euclidean metric on the ambient 9-dimensional space of matrices:

$$\langle \mathbf{RW}_1, \mathbf{RW}_2 \rangle = \frac{1}{2} \text{tr} ((\mathbf{RW}_1)^\top \mathbf{RW}_2) = \frac{1}{2} \text{tr} \mathbf{W}_1^\top \mathbf{W}_2.$$

and the *mean* of the \mathbf{R}_i is defined to be the \mathbf{Q} that minimizes Ψ . If the \mathbf{R}_i are widely spread, then there may be multiple local minima for Ψ . The global minimum is called the *Fréchet mean*, *Riemannian center of mass*, or *geodesic L^2 mean* (Karcher, 1977; Hartley et al., 2013). There has been extensive work on algorithms for computing these means (e.g., Govindu, 2004; Manton, 2004; Krakowski et al., 2007). The algorithm of Manton (2004) iteratively improves a seed value \mathbf{S} as follows (with $\mathbf{t}(\cdot)$ denoting \top).

```

Input R1, ..., Rn, S, epsilon
Q = S
W = (log(t(Q).R1) + ... + log(t(Q).Rn)) / n
errorSquared = tr(t(W).W)
While errorSquared >= epsilon^2
    Q = Q.exp(W)
    W = (log(t(Q).R1) + ... + log(t(Q).Rn)) / n
    errorSquared = tr(t(W).W)
Output Q

```

If the \mathbf{R}_i are all contained in an open ball of radius $\pi/2$, then Ψ has a single local minimum on that ball (Grove et al., 1974, Theorem 3.7), and the algorithm just given finds it from $\mathbf{S} = \mathbf{R}_1$. More generally, one can run the algorithm with multiple choices of \mathbf{S} and choose the result that yields the minimum variance overall.

Similarly, the *median* of the $\{\mathbf{R}_i\}$ can be defined as the rotation \mathbf{R} that minimizes $\sum d(\mathbf{R}, \mathbf{R}_i)$ (e.g., Fletcher et al., 2009). We do not use the median in this paper, but see Stanfill et al. (2013) for a comparison of means and medians in $\text{SO}(3)$.

For any rotation \mathbf{R} , let $\mathbf{v} = [v_1 \ v_2 \ v_3]^\top$ be the corresponding tangent vector at $\bar{\mathbf{R}}$:

$$\begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} = \log(\mathbf{R}\bar{\mathbf{R}}^\top).$$

The *covariance* Σ of a data set $\{\mathbf{R}_i\}$ with mean $\bar{\mathbf{R}}$ is defined as the 3×3 tensor

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^\top.$$

Let $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq 0$ be the square roots of the eigenvalues of Σ and $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ the corresponding unit eigenvectors. The *principal geodesic analysis* of Fletcher et al. (2004) essentially outputs these α_j and \mathbf{u}_j . The α_j are analogous to standard deviations. The j th principal geodesic is the curve $\mathbf{R}(t)$ of rotations corresponding to the line of tangent vectors $\mathbf{v}(t) = t\alpha_j \mathbf{u}_j$, with t running from -1 to 1 . The *Mahalanobis distance* of any rotation \mathbf{R} relative to the data set is $\sqrt{\mathbf{v}^\top \Sigma^{-1} \mathbf{v}}$, where \mathbf{v} corresponds to \mathbf{R} .

As an aside, we now describe the distance from a point to a geodesic curve. Let \mathbf{Q} be a rotation and $\mathbf{R}(\alpha) = (\exp \alpha \mathbf{U}) \mathbf{B}$ a geodesic, where \mathbf{U} corresponds to a unit vector \mathbf{u} as in Appendix A. Optimizing the distance between

\mathbf{Q} and \mathbf{R} is tantamount to optimizing

$$\begin{aligned} & \text{tr } \mathbf{Q}^\top (\exp \alpha \mathbf{U}) \mathbf{B} \\ &= \text{tr } (\mathbf{I} + (\sin \alpha) \mathbf{U} + (1 - \cos \alpha) \mathbf{U}^2) \mathbf{B} \mathbf{Q}^\top \\ &= \text{tr } (\mathbf{I} + \mathbf{U}^2) \mathbf{B} \mathbf{Q}^\top \\ &\quad + (\sin \alpha) \text{tr } \mathbf{U} \mathbf{B} \mathbf{Q}^\top - (\cos \alpha) \text{tr } \mathbf{U}^2 \mathbf{B} \mathbf{Q}^\top. \end{aligned}$$

with respect to α . The derivative equals zero when

$$\tan \alpha = -\text{tr } \mathbf{U} \mathbf{B} \mathbf{Q}^\top / \text{tr } \mathbf{U}^2 \mathbf{B} \mathbf{Q}^\top.$$

The two solutions correspond to the two points on the geodesic $\mathbf{R}(\alpha)$ where the distance from \mathbf{Q} is maximized and minimized.

Appendix C.3. Regression

Suppose that we are given data packaged as pairs $\{(s_1, \mathbf{R}_1), \dots, (s_n, \mathbf{R}_n)\}$, where the s_i are exact scalars and the \mathbf{R}_i are inexact rotations. We seek the geodesic curve

$$\mathbf{R} = \exp(\mathbf{M}s) \mathbf{B}$$

(Appendix A and Appendix C.1) that best expresses the dependence of the \mathbf{R}_i on the s_i . That is, we seek the rotation matrix \mathbf{B} and anti-symmetric matrix \mathbf{M} that minimize the residual sum of squares

$$E = \frac{1}{2n} \sum_{i=1}^n d(\mathbf{R}_i, \exp(\mathbf{M}s_i) \mathbf{B})^2,$$

where d denotes the distance between two rotations (Appendix C.2).

Unlike Euclidean linear regression, the minimization of E does not enjoy any simple linear-algebraic solution. To find the optimal parameter values, we must resort to iterative numerical methods. So that these methods may operate in a Euclidean space, we parametrize \mathbf{B} as $\mathbf{B} = (\exp \mathbf{W}) \mathbf{Q}$, where \mathbf{W} is an anti-symmetric matrix and \mathbf{Q} is the \mathbf{R}_i such that s_i is closest to zero. The unknowns are then the six non-redundant entries of \mathbf{M} and \mathbf{W} .

Similarly, the best-fit surface of the form

$$\mathbf{R} = \exp(\mathbf{M}s + \mathbf{L}t) \mathbf{B}$$

can be found by minimizing

$$E = \frac{1}{2n} \sum_{i=1}^n d(\mathbf{R}_i, \exp(\mathbf{M}s_i + \mathbf{L}t_i) \mathbf{B})^2$$

over anti-symmetric \mathbf{M}, \mathbf{L} and rotations \mathbf{B} . For $\mathbf{B} = (\exp \mathbf{W}) \mathbf{Q}$, the unknowns are the nine non-redundant entries of $\mathbf{M}, \mathbf{L}, \mathbf{W}$. It is worth noting that the curve regression is a special case of the surface regression, in which it is assumed that $\mathbf{L} = \mathbf{0}$. Hence the surface regression must produce at least as good a fit as the curve regression does. Further, the Fréchet mean of section Section 4.2 is a special case of the curve regression, in which $\mathbf{M} = \mathbf{0}$.

For kernel regression, let K be a *kernel* — that is, a probability density function on the real line, such as the

standard normal density $K(s) = (2\pi)^{-1/2}e^{-s^2/2}$. Let $h > 0$ be the *bandwidth* and let $K_h(s) = K(s/h)/h$ be K scaled by h .

For any scalar s and rotation \mathbf{R} , define

$$E_h(D, s, \mathbf{R}) = \frac{\sum_D K_h(s - s_i) d(\mathbf{R}_i, \mathbf{R})^2}{\sum_D K_h(s - s_i)},$$

where the sums are over the pairs (s_i, \mathbf{R}_i) in D . Let $\mathbf{R}(s)$ denote the value of \mathbf{R} that minimizes $E_h(D, s, \mathbf{R})$. This $\mathbf{R}(s)$ is the rotation predicted for the value s of the independent variable. It can be computed using iterative numerical optimization.

Typically the kernel K is chosen by the user. The bandwidth h can also be chosen manually, but Davis et al. (2010) described a least squares cross-validation algorithm for choosing h , as follows. Let $D^{(j)}$ denote D with the j th pair (s_j, \mathbf{R}_j) removed. For any h , let $\mathbf{R}_h^{(j)}$ denote the \mathbf{R} that minimizes $E_h(D^{(j)}, s_j, \mathbf{R})$. Then let h be the positive number that minimizes

$$\sum_{j=1}^n d(\mathbf{R}_j, \mathbf{R}_h^{(j)})^2.$$

Computing h in this manner requires a nested numerical optimization and hence potentially much more computer time than the rest of the regression.

Appendix C.4. Clustering

In a *clustering* problem, the goal is to partition a given data set into clusters, such that the data within each cluster are more similar to each other than to the data in the other clusters. Because there are many notions of what makes data similar, there are many clustering algorithms, which work well in different situations. In this subsection we summarize two methods that are easily adapted to rotational data.

The key concept of *k-means clustering* is that each datum should be closer to its cluster's mean than to the mean of any other cluster. Asgharbeygi and Maleki (2008) described two *k*-means clustering algorithms, the first of which is:

```
Input R1, ..., Rn, gamma0, epsilon, bound
gamma1 = gamma0
change = epsilon + 1
b = 0
While b < bound and change >= epsilon
    b += 1
    gamma0 = gamma1
    For h = 1, ..., n
        find the l in 1, ..., k
            that minimizes euc(gamma0, h, 1)
        move h in gamma1 to gamma1[1]
    change = abs(wgd(gamma1) - wgd(gamma0))
Output gamma1
```

Here, $\gamma = (\gamma_1, \dots, \gamma_k)$ is the list of clusters, where each cluster γ_ℓ is a list of integers. We require that the integers $1, \dots, n$ are partitioned among the clusters and no cluster is empty. The algorithm takes an initial γ as input, and improves this γ until a specified precision ϵ or number of steps *bound* is reached. The subroutine *wgd*(γ) computes

$$\sum_{\ell=1}^k \sum_{i,j \in \gamma_\ell} d(\mathbf{R}_i, \mathbf{R}_j)^2.$$

The subroutine *euc*(γ, h, ℓ) computes

$$\frac{2}{|\gamma_\ell|} \sum_{j \in \gamma_\ell} d(\mathbf{R}_h, \mathbf{R}_j)^2 - \frac{1}{|\gamma_\ell|^2} \sum_{i,j \in \gamma_\ell} d(\mathbf{R}_i, \mathbf{R}_j)^2,$$

where $|\gamma_\ell|$ is the size of the set γ_ℓ .

An important disadvantage of this approach is that the number k of clusters must be specified ahead of time rather than discovered by the algorithm. In fact, the user must supply a seed clustering, which the algorithm iteratively refines until it cannot be improved further. The resulting clustering is a locally optimal k -means clustering. A different local optimum could be reached by starting from a different seed, so the user may want to try several seeds.

In contrast, *DBSCAN clustering* partitions the data into contiguous regions of high density (Ester et al., 1996). The user supplies a radius ϵ and a positive integer m . A datum is called “dense” if there are at least m data points within radius ϵ of that point (including the point itself). Two data are declared to be in the same cluster if their distance is less than ϵ and at least one of the two data is dense. One consequence of this criterion is that some isolated points are not incorporated into any cluster.

Appendix D. Volume

Appendix D.1. Haar measure

To carry out probability and statistics on rotations, we must be able to integrate functions over the space $\text{SO}(3)$. That is, we require a notion of volume dV , so that for any (reasonable) real-valued function f on $\text{SO}(3)$, the integral $\int_{\text{SO}(3)} f \, dV$ can be computed. In mathematics, a notion of volume is called a *measure*. Lie groups carry natural left- and right-invariant measures, called *Haar measures*.

The Haar measure on $\text{SO}(3)$ is both left- and right-invariant and explicitly computable. In the angle-axis representation (α, \mathbf{u}) , let (ϕ, θ) be spherical coordinates for \mathbf{u} , where $0 \leq \phi \leq \pi$ is the co-latitude and $0 \leq \theta \leq 2\pi$ is the azimuth, so that

$$\mathbf{u} = [\sin \phi \cos \theta \quad \sin \phi \sin \theta \quad \cos \phi]^\top.$$

In these coordinates (α, ϕ, θ) on $\text{SO}(3)$, Haar measure is

$$dV = \frac{1}{4\pi^2} (\sin \phi)(1 - \cos \alpha) \, d\alpha \, d\phi \, d\theta \tag{D.1}$$

(Miles, 1965).

The coordinates (α, ϕ, θ) are computationally convenient because they incorporate α , which is the distance from the point in question to \mathbf{I} . For any $0 \leq r \leq \pi$, the ball B_r of radius r centered at \mathbf{I} has volume

$$\begin{aligned} & \int_{B_r} 1 \, dV \\ &= \int_0^{2\pi} \int_0^\pi \int_0^r \frac{1}{4\pi^2} (\sin \phi)(1 - \cos \alpha) \, d\alpha \, d\phi \, d\theta \\ &= \frac{1}{4\pi^2} \int_0^r (1 - \cos \alpha) \, d\alpha \int_0^\pi \sin \phi \, d\phi \int_0^{2\pi} 1 \, d\theta \\ &= (r - \sin r)/\pi. \end{aligned}$$

For example, the volumes of $B_{\pi/2}$ and $B_{2\pi/3}$ are approximately 0.18 and 0.39. Left- and right-invariance imply that the same statements hold for balls centered at other points in $\text{SO}(3)$. Also, $B_\pi = \text{SO}(3)$ has volume 1.

In Section 3 we consider alternative coordinates on $\text{SO}(3)$, in which α is distorted. Let

$$\rho = \left(\frac{3}{4\pi^2} (\alpha - \sin \alpha) \right)^{1/3}.$$

This ρ is an increasing function of α , with $\rho(0) = 0$ and $\rho(\pi) = (3/(4\pi))^{1/3}$. The derivative is

$$\begin{aligned} \frac{d\rho}{d\alpha} &= \frac{1}{3} \left(\frac{3}{4\pi^2} (\alpha - \sin \alpha) \right)^{-2/3} \frac{3}{4\pi^2} (1 - \cos \alpha) \\ &= \frac{1}{4\pi^2} \rho^{-2} (1 - \cos \alpha). \end{aligned}$$

Consequently, $\frac{1}{4\pi^2} (1 - \cos \alpha) \, d\alpha = \rho^2 \, d\rho$, and the Haar measure is

$$dV = \rho^2 (\sin \phi) \, d\rho \, d\phi \, d\theta.$$

This dV is the usual (Lebesgue) volume element in spherical coordinates on three-dimensional Euclidean space. Therefore, the coordinates (ρ, ϕ, θ) establish a volume-respecting map between $\text{SO}(3)$ and the ball of radius $(3/(4\pi))^{1/3} \approx 0.62$ in Euclidean space. This observation is the basis for the equal-volume plot of Section 3.3.

Appendix D.2. Uniform distribution

The normalized Haar measure is a translation-invariant probability measure, and hence the natural concept of uniform distribution on the space of rotations.

Diaconis and Shahshahani (1987) summarized a variety of methods for generating uniformly random rotations. Here we describe only their *subgroup algorithm*. Choose a vector \mathbf{v} uniformly on the unit sphere, for example using the algorithm above. Let $\mathbf{x} = [1 \ 0 \ 0]^\top - \mathbf{v}$, and let $\mathbf{J} = \mathbf{I} - 2\mathbf{x}\mathbf{x}^\top$. Choose an angle β uniformly on $[0, 2\pi]$, and form the rotation matrix

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix}.$$

Then the desired uniformly random rotation is $\mathbf{R} = -\mathbf{J}\mathbf{Q}$.

The equal-volume plot of Section 3.3 yields another rejection sampling algorithm for generating uniformly random rotations. Let $r = (3/(4\pi))^{1/3}$ be the radius of the equal-volume plot. Repeatedly choose x , y , and z uniformly on $[-r, r]$, until $\mathbf{x} = [x \ y \ z]^\top$ satisfies $|\mathbf{x}| \leq r$. (The expected number of tries is $(8r^3)/(4\pi r^3/3) \approx 1.9$.) Let $\rho = |\mathbf{x}|$ and $\mathbf{u} = \mathbf{x}/\rho$. Finally, compute α from ρ , by inverting the definition of ρ . There is no closed-form expression for α in terms of ρ , so an iterative numerical algorithm such as Newton's method is required.

Sei et al. (2013) described the following test of uniformity (see also Mardia and Jupp (2000, Ch. 13)). Given a sample $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$, compute the arithmetic mean $\tilde{\mathbf{R}} = n^{-1} \sum \mathbf{R}_i$ and the Rayleigh statistic $S = 3n \operatorname{tr} \tilde{\mathbf{R}}^\top \tilde{\mathbf{R}}$. If the sample comes from the uniform distribution, this S is asymptotically distributed as χ^2 with 9 degrees of freedom. Thus low values of $p = 1 - \int_0^S f \, dx$, where f is the density of the χ^2 distribution, indicate non-uniformity.

Appendix E. Plotting functions

This section describes how to visualize level sets of real-valued functions f on $\text{SO}(3)$. The Kamb contouring algorithm of Section 3.5 is a special case.

Appendix E.1. Adaptive mesh algorithm

Let f be a real-valued function on $\text{SO}(3)$. Graphing such a function is not possible, because the resulting plot would be four-dimensional. However, for any constant c , the level surface $f(\mathbf{Q}) = c$ can be drawn into the equal-angle or equal-volume plot using the following algorithm. In addition to the function f and number c , the user specifies two integers $N_1, N_2 \geq 0$.

The algorithm begins by dividing the plot into seven “boxes”: one around the center of the plot, and one attached to each of that box’s sides. For each box, the value of f is recorded at each vertex. The algorithm then begins a process of non-adaptive refinement, whose goal is to construct a fine enough mesh on $\text{SO}(3)$ that none of the major features of the level surface are missed. The algorithm halves each box in three directions, to produce eight sub-boxes. It splits each of those sub-boxes into eight smaller boxes, and splits those, and so on, always recording the value of f at every vertex.

After N_1 of these refinements, the algorithm enters an adaptive stage, whose goal is to improve the detail of the level surface. In this stage, only the boxes that intersect the level surface are sub-divided. Assuming that the non-adaptive refinement was fine enough, such a box is detectable because it has one vertex where $f \geq c$ and another where $f \leq c$.

After N_2 of these adaptive refinements, the final stage of the algorithm approximates the level surface within each box that intersects it. The value of f is linearly interpolated along each edge of the box, to approximate the points

where the level surface intersects that edge. These points are taken as vertices of a not-necessarily-planar polygon, which is drawn on the computer's screen or exported to a file for use by other software. The examples shown in 5EF use five non-adaptive refinements and one adaptive refinement.

Appendix E.2. Kamb weighting functions

Any density weighting function h employed in the Kamb contouring algorithm of Section 3.5 must integrate to $(r - \sin r)/\pi$ over the ball B_r of radius r . Using the Haar measure of Appendix D, we compute

$$\begin{aligned} & \int_{B_r} h(\alpha) dV \\ &= \int_0^{2\pi} \int_0^\pi \int_0^r h(\alpha) \frac{1}{4\pi^2} (\sin \phi)(1 - \cos \alpha) d\alpha d\phi d\theta \\ &= \frac{1}{4\pi^2} \int_0^r h(\alpha) (1 - \cos \alpha) d\alpha \int_0^\pi \sin \phi d\phi \int_0^{2\pi} 1 d\theta \\ &= \frac{1}{\pi} \int_0^r h(\alpha) (1 - \cos \alpha) d\alpha. \end{aligned}$$

Now suppose that h is defined on B_r as

$$h(\alpha) = c_3 \alpha^3 + c_2 \alpha^2 + c_1 \alpha + c_0,$$

for some unknown coefficients c_i . Repeated integration by parts yields

$$\begin{aligned} & \frac{1}{\pi} \int_0^r (c_3 \alpha^3 + c_2 \alpha^2 + c_1 \alpha + c_0) (1 - \cos \alpha) d\alpha \\ &= \pi^{-1} d_3 c_3 + \pi^{-1} d_2 c_2 + \pi^{-1} d_1 c_1 + \pi^{-1} d_0 c_0, \end{aligned}$$

where

$$\begin{aligned} d_3 &= (6r - r^3) \sin r + (6 - 3r^2) \cos r + r^4/4 - 6, \\ d_2 &= (2 - r^2) \sin r - 2r \cos r + r^3/3, \\ d_1 &= -r \sin r - \cos r + r^2/2 + 1, \\ d_0 &= r - \sin r. \end{aligned}$$

In this way, the equation $\int_{B_r} h(\alpha) dV = (r - \sin r)/\pi$ amounts to one linear constraint on the c_i . We can therefore impose additional desiderata before solving for the c_i . Here are three examples.

Imposing $c_3 = c_2 = c_1 = 0$ produces $h(\alpha) = 1$. This constant h corresponds to not weighting the points at all. Because all weighting functions are zero for $\alpha > r$, this h has the undesirable property of being discontinuous at $\alpha = r$.

A linear h satisfies $c_3 = c_2 = 0$. Continuity at $\alpha = r$ requires $h(r) = 0$. Imposing these conditions results in

$$c_0 = \frac{2r^2 - 2r \sin r}{2 \cos r + r^2 - 2}, \quad c_1 = -c_0/r.$$

A cubic h with differentiable behavior at $\alpha = r$ and $\alpha = 0$ satisfies $h(r) = 0$, $h'(r) = 0$, and $h'(0) = 0$. Imposing these conditions results in

$$c_0 = \frac{2r^4 - 2r^3 \sin r}{12r \sin r + 24 \cos r + r^4 - 24},$$

$$c_1 = 0, \quad c_2 = -3c_0/r^2, \quad c_3 = 2c_0/r^3.$$

Appendix F. Uniform-angle, random-spin distributions

There is a particularly important class of probability distributions on the space $\text{SO}(3)$ of rotations, called *uniform-axis, random-spin* distributions (e.g., Bingham et al., 2009; Qiu et al., 2014). This class includes the uniform distribution and several competing notions of normal distribution (e.g., Downs, 1972; Nikolayev and Savyolova, 1997; Bunge, 2013). Each uniform-axis, random-spin distribution has two parameters: a rotation \mathbf{S} , which describes the center of the distribution, and a number $\kappa > 0$, which describes how concentrated the distribution is about the center. The distribution is radially symmetric about the center, in that its density depends only on the distance α from \mathbf{S} .

More precisely, each uniform-axis, random-spin distribution has an angular density $g(\alpha|\kappa)$ on the interval $[-\pi, \pi]$. This g is symmetric in α , in that $g(\alpha|\kappa) = g(-\alpha|\kappa)$. (Equivalently, g can be defined on $[0, \pi]$.) To sample a rotation from the distribution, sample an angle α from the distribution on $[-\pi, \pi]$ defined by g . Also, sample a unit vector \mathbf{u} from the uniform distribution on the unit sphere, for example by choosing θ uniformly on $[0, 2\pi]$ and z uniformly on $[-1, 1]$, and then computing $r = \sqrt{1 - z^2}$ and

$$\mathbf{u} = [r \cos \theta \quad r \sin \theta \quad z]^\top.$$

Once α and \mathbf{u} are chosen, compute the rotation matrix \mathbf{R} for the angle-axis rotation (α, \mathbf{u}) . The uniform-axis, random-spin rotation is then \mathbf{SR} .

Appendix F.1. Uniform distribution

For the uniform distribution (Appendix D.2), Eq. (D.1) implies that the axis \mathbf{u} of rotation is uniformly distributed on the unit sphere and the angle α has density $g(\alpha|\kappa) = (1 - \cos \alpha)/(2\pi)$ on $[-\pi, \pi]$. That is, the uniform distribution is trivially a uniform-axis, random-spin distribution, in which any rotation \mathbf{S} is the center and g does not depend on κ .

Appendix D.2 gives two algorithms for sampling from the uniform distribution, but here is a rejection sampling algorithm based on the uniform-axis, random-spin framework. Repeatedly choose α uniformly on $[-\pi, \pi]$ and y uniformly on $[0, 1/\pi]$, until they satisfy $y \leq (1 - \cos \alpha)/(2\pi)$. The expected number of tries is 2.

Appendix F.2. Isotropic matrix Fisher distribution

The *isotropic matrix Fisher distribution* (Matthies et al., 1988; Qiu et al., 2014) on $\text{SO}(3)$ is a special case of the matrix Fisher distribution of Appendix G, in which the concentration \mathbf{K} is a scalar matrix $(\kappa^2/2)\mathbf{I}$. It is a uniform-axis, random-spin distribution with angular density

$$g(\alpha|\kappa) = \frac{1 - \cos \alpha}{2\pi} \frac{e^{\kappa^2 \cos \alpha}}{I_0(\kappa^2) - I_1(\kappa^2)},$$

where I_0 and I_1 are modified Bessel functions. As $\kappa \rightarrow 0^+$, this distribution approaches the uniform distribution.

Sampling the matrix Fisher distribution with $\mathbf{K} = (\kappa^2/2)\mathbf{I}$ is an efficient way to sample this distribution. For completeness, here is a method for sampling the distribution as a uniform-axis, random-spin distribution. First define

$$f(\alpha) = (1 - \cos \alpha)e^{\kappa^2 \cos \alpha}.$$

If $\kappa \leq 2^{-1/2}$, then the maximum value of f on $[-\pi, \pi]$ is $2e^{-\kappa^2}$. So repeatedly choose α uniformly on $[-\pi, \pi]$ and y uniformly on $[0, 2e^{-\kappa^2}]$, until they satisfy $y \leq (1 - \cos \alpha) \exp(\kappa^2 \cos \alpha)$. The expected number of tries depends on κ , but never exceeds two. On the other hand, if $\kappa \geq 2^{-1/2}$, then the maximum of f is $\kappa^{-2} \exp(\kappa^2 - 1)$. So repeatedly choose α uniformly on $[-\pi, \pi]$ and y uniformly on $[0, \kappa^{-2} \exp(\kappa^2 - 1)]$, until they satisfy $y \leq (1 - \cos \alpha) \exp(\kappa^2 \cos \alpha)$. In this case, cursory tests suggest that the expected number of tries is approximately 1.8κ . So this method is not efficient for large κ .

The density of the isotropic matrix Fisher distribution, with respect to the Haar measure on $\text{SO}(3)$, is

$$f(\mathbf{R}|\mathbf{S}, \kappa) = \frac{\exp(\kappa^2 \cos d(\mathbf{S}, \mathbf{R}))}{I_0(\kappa^2) - I_1(\kappa^2)}.$$

So the likelihood for a set $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ of independent observations is

$$\mathcal{L}(\mathbf{S}, \kappa) = \frac{\exp(\kappa^2 \sum \cos d(\mathbf{S}, \mathbf{R}_i))}{(I_0(\kappa^2) - I_1(\kappa^2))^n}.$$

Maximum likelihood estimates for \mathbf{S} and κ can be computed by numerical optimization. Chang and Rivest (2001) described M -estimation methods for the mean, in a class of distributions that includes the isotropic matrix Fisher distribution.

Appendix F.3. Wrapped trivariate normal distribution

Qiu et al. (2014) compared various uniform-axis, random-spin distributions. They argued that the *isotropic Gaussian distribution* enjoys a version of the central limit theorem, and that the more tractable *wrapped trivariate normal distribution* approximates the isotropic Gaussian distribution well. We focus on the latter. Its angular density is

$$g(\alpha|\kappa) = \frac{\kappa^3}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} (2m\pi - \alpha)^2 e^{-\kappa^2(2m\pi - \alpha)^2/2}.$$

Qiu et al. (2014) gave two efficient methods for sampling from this distribution. In the first method, a number w is chosen from the χ^2 distribution with three degrees of freedom, b is chosen to be zero or one with equal probability, α is computed as

$$\alpha = (-1)^b |\kappa^{-1}\sqrt{w} - \pi[\kappa^{-1}\sqrt{w}/\pi]|,$$

and the rest of the uniform-axis, random-spin sampling proceeds as usual. In the second method, three numbers w_1, w_2, w_3 are drawn from the normal distribution $N(0, \kappa^{-2})$ with mean zero and variance κ^{-2} . The wrapped trivariate normal distribution rotation is then

$$\mathbf{S} \exp \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}.$$

The density of the wrapped trivariate normal distribution, with respect to the Haar measure on $\text{SO}(3)$, is

$$f(\mathbf{R}|\mathbf{S}, \kappa) = \frac{2\pi}{1 - \cos \alpha} g(\alpha|\kappa),$$

where $\alpha = d(\mathbf{S}, \mathbf{R})$. This density is singular at $\mathbf{R} = \mathbf{S}$ (Qiu et al., 2014). Consequently, maximum likelihood estimates of \mathbf{S} and κ are impossible.

Appendix G. Matrix Fisher distribution

Downs (1972) defined the *matrix Fisher distribution* on $\text{SO}(3)$ (also called *matrix von Mises-Fisher* and *matrix Langevin*) to be the probability distribution whose density is

$$f(\mathbf{R}; \mathbf{M}, \mathbf{K}) = a \exp(\text{tr } \mathbf{K} \mathbf{M}^\top \mathbf{R})$$

with respect to the uniform distribution. Here, a is a normalizing constant, \mathbf{M} describes the center of the distribution, and \mathbf{K} is a positive-definite, symmetric matrix that describes the spread of the distribution. Large eigenvalues in \mathbf{K} indicate tight clustering around \mathbf{M} , while the distribution goes to uniform as $\mathbf{K} \rightarrow \mathbf{0}$. There are two ways to fill in the details (Sei et al., 2013). Both versions appear in the literature, and their results are not identical, so readers must use them with care.

On the one hand, the first two columns (or rows) of a rotation matrix uniquely determine the third. So $\text{SO}(3)$ can be identified with the set of 3×2 matrices \mathbf{R} such that $\mathbf{R}^\top \mathbf{R} = \mathbf{I}_2$. Such matrices form the *Stiefel manifold* $V_2(\mathbb{R}^3)$ (Downs, 1972; Mardia and Jupp, 2000). In this setting, \mathbf{M} is a 3×2 matrix in $V_2(\mathbb{R}^3)$ and \mathbf{K} is 2×2 .

On the other hand, the *orthogonal group* $O(3) = V_3(\mathbb{R}^3)$ is the set of 3×3 matrices \mathbf{R} such that $\mathbf{R}^\top \mathbf{R} = \mathbf{I}_3$. This group consists of two connected components, one of which is $\text{SO}(3)$. An \mathbf{M} in $O(3)$ and $3 \times 3 \mathbf{K}$ determine a matrix Fisher distribution on $O(3)$. If \mathbf{M} is in $\text{SO}(3)$ and we restrict the distribution to $\text{SO}(3)$ and renormalize, then we obtain a distribution on $\text{SO}(3)$ centered on \mathbf{M} .

As Sei et al. (2013) explained, the first notion of matrix Fisher distribution on $\text{SO}(3)$ is a strict submodel of the second. That is, any distribution of the first form can be expressed in the second form, but not *vice-versa*. Except for the confidence region of Downs (1972) mentioned in Section 5.6, this paper always uses the second version.

The normalizing constant $a = a(\mathbf{K})$ is rather complicated (Khatri and Mardia, 1977). Wood (1993) showed that a^{-1} simplifies to

$$a^{-1} = \int_0^1 I_0((\Phi_{11} - \Phi_{22})u) \cdot I_0((\Phi_{11} + \Phi_{22})(1-u)) \cdot \exp(\Phi_{33}(1-2u)) du, \quad (\text{G.1})$$

where $\Phi_{11} \geq \Phi_{22} \geq \Phi_{33} \geq 0$ are the eigenvalues of \mathbf{K} and I_0 is the modified Bessel function of degree 0. (That paper's expression for a^{-1} includes an extraneous π^2 factor, which we have removed.) Sei et al. (2013) also derived a series expansion for a^{-1} .

Kent et al. (2013) described an efficient rejection sampling method for generating rotations from this distribution. The method relies on a relationship, due to Prentice (1986), between the Bingham distribution on the unit quaternions (Appendix A) and the matrix Fisher distribution on $\text{SO}(3)$. To prepare for generating a sample, diagonalize $\mathbf{K} = \mathbf{V}\Phi\mathbf{V}^\top$ such that \mathbf{V} is a rotation and $\Phi_{11} \geq \Phi_{22} \geq \Phi_{33} \geq 0$. Let Λ be the 4×4 diagonal matrix with diagonal entries 0, $2(\Phi_{22} + \Phi_{33})$, $2(\Phi_{11} + \Phi_{33})$, $2(\Phi_{11} + \Phi_{22})$. Let b be the greatest solution to

$$\sum_{i=1}^4 \frac{1}{b + 2\Lambda_{ii}} = 1,$$

and let $\Omega = \mathbf{I} + 2\Lambda/b$. Also define $m^* = (4/b)^2 \exp(b/2-2)$, $f^*(\mathbf{q}) = \exp(-\mathbf{q}^\top \Lambda \mathbf{q})$, and $g^*(\mathbf{q}) = (\mathbf{q}^\top \Omega \mathbf{q})^{-2}$ for any 4-vector $\mathbf{q} = [q_0 \ q_1 \ q_2 \ q_3]^\top$. Once these quantities are computed, any number of rotations can be generated quickly. To generate each rotation \mathbf{R} , draw \mathbf{x} from the four-dimensional normal distribution $N_4(\mathbf{0}, \Omega^{-1})$ with mean $\mathbf{0}$ and covariance Ω^{-1} . Let $\mathbf{q} = \mathbf{x}/|\mathbf{x}|$. Also draw w from the uniform distribution on $[0, 1]$. Repeat these choices until $w < f^*(\mathbf{q})/(m^* g^*(\mathbf{q}))$. Then convert the quaternion \mathbf{q} to a rotation matrix \mathbf{Q} and compute $\mathbf{R} = \mathbf{M}\mathbf{Q}\mathbf{V}^\top$. In cursory tests, about 2.2 choices of \mathbf{x} and w are needed to generate each \mathbf{R} . This success rate is consistent with the 45% efficiency claims of Kent et al. (2013).

Given a data set $\{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ in $\text{SO}(3)$, we now summarize the maximum likelihood estimates of \mathbf{M} and \mathbf{K} (Sei et al., 2013). Let $\tilde{\mathbf{R}} = n^{-1} \sum \mathbf{R}_i$ be the arithmetic mean of the sample (which may not be orthogonal or of positive determinant). Compute the “signed” singular value decomposition $\tilde{\mathbf{R}} = \mathbf{Q}\mathbf{G}\mathbf{V}^\top$, where \mathbf{Q} and \mathbf{V} are rotations, \mathbf{G} is diagonal, and $|G_{11}| \geq G_{22} \geq G_{33} \geq 0$ (Prentice, 1986). The maximum likelihood estimate of \mathbf{M} is simply

$$\hat{\mathbf{M}} = \mathbf{Q}\mathbf{V}^\top.$$

The maximum likelihood estimate of \mathbf{K} is

$$\hat{\mathbf{K}} = \mathbf{V} \begin{bmatrix} \hat{\Phi}_{11} & 0 & 0 \\ 0 & \hat{\Phi}_{22} & 0 \\ 0 & 0 & \hat{\Phi}_{33} \end{bmatrix} \mathbf{V}^\top,$$

where the $\hat{\Phi}_{ii}$ are chosen to maximize

$$\Phi_{11}G_{11} + \Phi_{22}G_{22} + \Phi_{33}G_{33} - \log a^{-1}(\Phi_{11}, \Phi_{22}, \Phi_{33}).$$

The maximum can be computed using Eq. (G.1) and any numerical optimization method, but see Sei et al. (2013) for a more sophisticated approach.

Appendix H. Markov chain Monte Carlo

Appendix H.1. General concepts

References for this topic include Chib and Greenberg (1995); Gregory (2005); Monahan (2011); Robert and Casella (2013). Let $D = \{\mathbf{R}_1, \dots, \mathbf{R}_n\}$ be a rotational data set. Suppose that our statistical model depends on some parameters

$$\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_m]^\top.$$

We could find the value of θ that best fits the data D , but here we want to obtain more information beyond simply the best fit, by understanding the probability distribution of θ given D . By Bayes' theorem,

$$\rho(\theta|D) = \rho(D)^{-1} \rho(\theta)\rho(D|\theta). \quad (\text{H.1})$$

The first factor on the right side of this equation is independent of θ and hence amounts to a normalizing constant, which has no practical effect on our calculations. We now explain the other two factors.

The second factor is the *prior*, which represents our knowledge of the parameters θ before taking the data into account. Commonly we have no prior knowledge and wish to make no *a priori* assumptions, so we construct an *uninformative* prior. For a location parameter such as the mean, we might use the uniform distribution. For a positive real parameter we might use a *Jeffreys prior*, which is independent of the scale used to measure the parameter. The details vary from case to case.

The third factor is the *likelihood*, which captures the probability of the data given the parameters. Assuming that the data are drawn independently from the distribution $\rho(\mathbf{R}|\theta)$, the likelihood is

$$\rho(D|\theta) = \prod_{i=1}^n \rho(\mathbf{R}_i|\theta).$$

The left side of Eq. (H.1) is the *posterior* probability distribution. We sample from this distribution using the *Metropolis-Hastings algorithm*, which is an example of a Markov chain Monte Carlo (MCMC) method. The algorithm relies on some tuning parameters, which are described below. Starting from a seed value θ^0 for the parameters θ , it produces new values $\theta^1, \theta^2, \theta^3, \dots$ by computing θ^{j+1} from θ^j as follows:

1. Choose a candidate θ^{j+1} by perturbing θ^j slightly. Commonly θ^{j+1} is chosen from a normal distribution whose mean is θ^j and whose dispersion is a tuning parameter.

2. Compute the ratio

$$r = \frac{\rho(\theta^{j+1}|D)}{\rho(\theta^j|D)} = \frac{\rho(\theta^{j+1})\rho(D|\theta^{j+1})}{\rho(\theta^j)\rho(D|\theta^j)}$$

of posteriors. Notice that the normalizing constant $\rho(D)^{-1}$ cancels itself in this ratio and thus has no effect on the algorithm.

3. Choose a random number x uniformly on the interval $[0, 1]$. If $x > r$, then set $\theta^{j+1} = \theta^j$.

Thus the new value θ^{j+1} is either the old value θ^j or the candidate chosen by perturbing θ^j . When the candidate is better than the old value, as measured by the ratio r , then it is taken as the new value. When the old value is better than the candidate, then either one may be chosen, with probability determined by r . In this way the algorithm seeks out parameter values θ that fit the data well, while also considering values that fit the data less well.

The first 10^6 (say) iterations of the algorithm are regarded as a *burn-in phase*, during which the algorithm explores the parameter space, possibly ranging far from the seed value θ^0 . Periodically we evaluate the *acceptance rate*, which is the proportion of iterations in which the candidate is chosen over the old value. We adjust the tuning parameters to keep the acceptance rate between 30% and 40% (e.g., Qiu et al., 2014). See Gregory (2005); Robert and Casella (2013) for other performance heuristics.

After burn-in, the algorithm enters a *collection phase* of 10^7 iterations (say), in which every value θ^j is recorded in memory as it is generated. The output of the collection phase is a large sample of parameter values θ , whose density in the parameter space approximates the posterior. Various auxiliary quantities can be computed from this MCMC sample. In particular, if one of the parameters in θ is the mean of the probability distribution, then we can marginalize over the other parameters to understand the distribution of the mean given the data.

Appendix H.2. Wrapped trivariate normal distribution

As an example, we summarize the MCMC method of Qiu et al. (2014) for the wrapped trivariate normal distribution, which has parameters \mathbf{M} and κ (Appendix F.3). For an uninformative prior, we assume that \mathbf{M} and κ are independent: $\rho(\mathbf{M}, \kappa) = \rho(\mathbf{M})\rho(\kappa)$. For $\rho(\mathbf{M})$ we use the uniform distribution on $\text{SO}(3)$ (Appendix D.2). For $\rho(\kappa)$ we use the Jeffreys prior

$$J(\kappa) = \kappa \sqrt{-\frac{9}{\kappa^2} + \frac{\kappa^5}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \frac{H(\alpha, \kappa, 4)^2}{H(\alpha, \kappa, 2)} d\alpha},$$

where

$$H(\alpha, \kappa, c) = \sum_{m=-\infty}^{\infty} (2m\pi - \alpha)^c e^{-\frac{1}{2}\kappa^2(2m\pi - \alpha)^2}.$$

The likelihood is

$$\rho(D|\mathbf{M}, \kappa) = \prod_{i=1}^n \frac{g(d(\mathbf{M}, \mathbf{R}_i)|\kappa)}{2 - 2 \cos d(\mathbf{M}, \mathbf{R}_i)},$$

where $g(\alpha|\kappa)$ is the angular density of the distribution. Maximum likelihood estimates are not possible because the likelihood goes to infinity as \mathbf{M} approaches any \mathbf{R}_i .

To sample from the posterior distribution $\rho(\mathbf{M}, \kappa|D)$, Qiu et al. (2014) used a variant of Metropolis-Hastings called *Metropolis-Hastings-within-Gibbs*. The algorithm operates on $\eta = -\log \kappa$ instead of κ itself, with tuning parameters κ_F and γ controlling the size of the perturbations to \mathbf{M} and η . Given \mathbf{M}^j and η^j , the next values \mathbf{M}^{j+1} and η^{j+1} are generated as follows.

1. Choose a candidate \mathbf{M}^{j+1} from the isotropic matrix Fisher distribution (Appendix F.2) with parameters \mathbf{M}^j and κ_F .
2. Compute $r = \frac{\rho(\mathbf{M}^{j+1}, \eta^j|D)}{\rho(\mathbf{M}^j, \eta^j|D)}$.
3. Choose x uniformly on $[0, 1]$. If $x > r$, then set $\mathbf{M}^{j+1} = \mathbf{M}^j$.
4. Choose a candidate η^{j+1} from the normal distribution with mean η^j and variance γ^2 .
5. Compute $r = \frac{\rho(\mathbf{M}^{j+1}, \eta^{j+1}|D)}{\rho(\mathbf{M}^{j+1}, \eta^j|D)}$.
6. Choose x uniformly on $[0, 1]$. If $x > r$, then set $\eta^{j+1} = \eta^j$.

In our implementation, we perturb \mathbf{M}^j to \mathbf{M}^{j+1} by drawing from a wrapped trivariate normal distribution with concentration $\kappa = e^{-\nu}$. Sampling from this distribution is easier to implement and faster to execute than sampling from the isotropic matrix Fisher distribution. Suggested values for the tuning parameters ν and γ are given in Table H.1. That table also shows typical values of the variance Ψ (Section 4.2). When presented with a data set, we first compute n and Ψ . Then we use the table to guess η . Then we use the table to guess ν and γ , which are further tuned over the course of the simulation. The collected means $\{\mathbf{M}^j\}$ approximate the distribution of \mathbf{M} given the data. Frequently they form a tight ellipsoidal cloud, which can then be summarized using percentiles of Mahalanobis distance as in Section 5.5.

Appendix H.3. Other distributions

The Metropolis-Hastings-within-Gibbs approach just outlined can be applied similarly to other uniform-axis, random-spin distributions (Appendix F) with only two changes. First, the Jeffreys prior for κ must be computed. Second, the likelihood must be altered to use the appropriate angular density. For example, Bingham et al. (2009) applied this approach to the von Mises uniform-axis random-spin distribution. Qiu (2013) lists Jeffreys priors for many uniform-axis, random-spin distributions and describes software for their MCMC simulation.

	$n = 10$	30	100	300	1,000
$\eta = 1$					
0		1.3607, -1.30, 0.50	1.4532, -1.75, 0.13	1.4468, -2.35, 0.08	
-1	0.1879, -1.95, 0.40	0.1983, -2.45, 0.23	0.2017, -3.05, 0.13	0.2017, -3.60, 0.07	0.2026, -4.20, 0.04
-2	0.0245, -2.75, 0.40	0.0268, -3.40, 0.23	0.0274, -4.10, 0.13		
-3	0.0034, -4.10, 0.40	0.0036, -4.45, 0.23	0.0037, -5.05, 0.13		
-4	0.0004, -4.95, 0.40	0.0005, -5.45, 0.23			

Table H.1: Tuning the MCMC sampler to achieve an acceptance rate of approximately 35%, based on synthetic data sets drawn from the wrapped trivariate normal distribution. For each combination of sample size n and spread $\eta = -\log \kappa$, three numbers Ψ , ν , γ are given. The first number is the variance Ψ of a typical sample about its mean. The second number ν controls the perturbation of \mathbf{M}^j to \mathbf{M}^{j+1} . Similarly, γ controls the perturbation of η^j to η^{j+1} . Tunings for some combinations of n and η are missing from the table because they could not be found consistently across multiple data sets.

For the matrix Fisher distribution on $\text{SO}(3)$ with parameters \mathbf{M} and \mathbf{K} (Appendix G) a similar approach should work in theory. It might be advantageous to operate on the matrix logarithm $\log \mathbf{K}$, which inhabits a vector space, rather than \mathbf{K} itself. The likelihood would require computation of the normalizing constant, which is slow but tractable. The most difficult aspect could be the choice of prior for \mathbf{K} or $\log \mathbf{K}$. To our knowledge the Jeffreys prior does not appear in the literature.

Appendix I. Experimental results

This section gives greater detail about the procedures and results of our numerical experiments on coverage rates of confidence/credible regions. See Section 5.6 for context and discussion.

Appendix I.1. First experiment: rotations

In the first experiment, for each combination of sample size n and concentration parameters $\kappa_1, \kappa_2, \kappa_3$, we perform the following procedure N times. Choose a uniformly random \mathbf{M} and a symmetric \mathbf{K} with eigenvalues $\kappa_1, \kappa_2, \kappa_3$ and uniformly random orientation. Generate a synthetic data set of size n by sampling from the matrix Fisher distribution with mean \mathbf{M} and concentration \mathbf{K} . Use the six methods, with $b = 10,000$ bootstrap samples and $b = 1,000,000$ MCMC samples, to construct six confidence/credible regions. Because the trial is automated, do not visually inspect the equal-angle or equal-volume plot to make sure that the simulation results form a tight ellipsoidal cloud. For each method's region, record whether the true population mean \mathbf{M} is in that region.

Once all N trials are complete, we compute the proportion of trials in which \mathbf{M} was in each of the six regions. These proportions are shown in Table I.2, for both 95% and 99% confidence/credible regions. The 95% results are also shown graphically in Fig. 11A-D.

The interested reader can estimate the uncertainty in each proportion from the observed proportion \hat{p} and the number N of trials. As long as $N\hat{p} \geq 10$ and $N(1 - \hat{p}) \geq 10$, the 95% confidence interval for the proportion is well-approximated by $\hat{p} \pm 2SE$, where $SE = \sqrt{\hat{p}(1 - \hat{p})/N}$. For example, $\hat{p} = 0.965$ and $N = 2,050$ produces 0.965 ± 0.008 .

Some of the methods occasionally fail on small and widely dispersed data sets. The Rancourt et al. (2000) method fails in 0.7% of the $n = 10, \kappa_1 = \kappa_2 = \kappa_3 = 1$ trials, because the sample mean is more than $\pi/2$ away from the true mean, rendering the Rancourt et al. (2000) tangent space approximation invalid. Our MCMC sampler fails to tune in 3.0% of the $n = 10, \kappa_1 = \kappa_2 = \kappa_3 = 1$ trials, 0.1% of the $n = 10, \kappa_1 = \kappa_2 = 1, \kappa_3 = 7$ trials, and 0.1% of the $n = 30, \kappa_1 = \kappa_2 = \kappa_3 = 1$ trials. Table I.2 and Fig. 11 ignore these failing trials, showing only the number of successful trials and their coverage rates.

Appendix I.2. Second experiment: Orthorhombic bipyramidal symmetry

The second experiment concerns rotations up to line-in-plane symmetry (Section 2.2). We use only two methods: non-parametric bootstrapping and MCMC. In all other aspects this experiment is identical to the first experiment.

The results are shown in Table I.3, with their 95% confidence intervals. In five cases the MCMC simulation sometimes fails to tune. Again failures are excluded from the trial count N . The 95% coverage rates are also shown graphically in the right side of Fig. 11.

Appendix I.3. Third experiment: For Idaho foliation-lineations

The third experiment is similar to the second, in that we apply non-parametric bootstrapping and MCMC to rotations with line-in-plane symmetry. However, only one combination of sample size and concentration is used: $n = 23, \kappa_1 = 0.000003, \kappa_2 = 11, \kappa_3 = 33$. These parameters are derived from the western Idaho shear zone foliation-lineation data of Giorgis and Tikoff (2004).

Based on $N = 2,266$ trials, the bootstrapping coverage rates are 0.887 ± 0.013 for 95% confidence and 0.956 ± 0.008 for 99% confidence. The MCMC coverage rates are 0.926 ± 0.011 and 0.977 ± 0.006 . The MCMC simulation never fails to tune.

As an aside, during the preparation of this paper we have run two other special cases of this experiment. For $n = 23, \kappa_1 = 0.000006, \kappa_2 = 11$, and $\kappa_3 = 33, N = 2,011$ trials produce bootstrapping coverage rates of 0.888 ± 0.014 and 0.952 ± 0.010 and MCMC coverage rates of 0.925 ± 0.012 .

These observations let us extend many of the concepts and methods for SO(3) to the new setting of SO(3)/G.

For example, the Fréchet variance (Section 4.2) of n points $G\mathbf{R}_1, \dots, G\mathbf{R}_n$ about a point $G\mathbf{R}$ in SO(3)/G is

$$\Psi(G\mathbf{R}) = \frac{1}{2n} \sum_{i=1}^n \min_{\mathbf{J} \in G} d(\mathbf{JR}, \mathbf{R}_i)^2.$$

In the Fréchet mean algorithm of Appendix C.2, we change the line

$$W = (\log(t(Q).R1) + \dots + \log(t(Q).Rn)) / n$$

to

$$W = (\log(t(Q).R1') + \dots + \log(t(Q).Rn')) / n$$

where \mathbf{R}'_i is the representative of $G\mathbf{R}_i$ closest to \mathbf{Q} . If the data are sufficiently concentrated, then this algorithm always uses a single set of representative rotations for them. More precisely, define the *separation* of G to be the minimum distance between any two of its elements:

$$\text{sep}(G) = \min_{\mathbf{J}, \mathbf{K} \in G} d(\mathbf{J}, \mathbf{K}) = \min_{\mathbf{J} \in G} d(\mathbf{J}, \mathbf{I}).$$

Also define the *diameter* of a set of rotations to be the maximum distance between any two of them:

$$\text{diam}(\{\mathbf{R}_1, \dots, \mathbf{R}_n\}) = \max_{i,j} d(\mathbf{R}_i, \mathbf{R}_j).$$

Suppose that $\text{diam}(\{\mathbf{R}_1, \dots, \mathbf{R}_n\}) < \text{sep}(G)/2$. Then for any $\mathbf{J} \neq \mathbf{I}$, the distance from any \mathbf{R}_i to any \mathbf{JR}_j is greater than $\text{sep}(G)/2$ by the triangle inequality. !!So the algorithm never uses \mathbf{JR}_j to represent \mathbf{R}_j . Similar remarks apply to diameter, tangent space approximation, non-parametric bootstrapping of the Fréchet mean, geodesic regression, kernel regression, k-means clustering, and DBSCAN clustering.

The projected arithmetic mean and \mathbf{T} tensor of Section 4.3 are not obviously extendable to SO(3)/G. Neither are the maximum likelihood estimate of Section 4.7, the parametric bootstrap based on it, and the inference methods of Downs (1972); Prentice (1986). However, when the data are tightly concentrated, the tails of the underlying distribution are small enough that symmetry can be neglected without any practical harm. The inference method of Rancourt et al. (2000) assumes tight concentration always. Our Markov chain Monte Carlo simulation uses a symmetrized wrapped trivariate normal distribution. However, the Jeffreys prior for the concentration κ is naively left unchanged.

bachmannetal2010 do iterative projected arithmetic mean

in two-sample inference, same symmetry group for null hypothesis

or work in a primitive domain (subset not quotient)? bingham and qiu seem to. and kagan. what does MTEX do? makes sense if data are tightly clustered, so that the flips don't interact with each other much, and tails of distribution are thin

plotting: show every representative of every datum
Kamb: slight change to algebra

huckemannetal2010 does geodesic PCA on quotients of manifolds. prentice1982 does antipodally symmetric matrix Fisher, which on SO(3) has four modes, which I think have exactly the same symmetry as foliation-lineations

some results of Arnold and Jupp (2013) are applicable to orientations with line-in-plane symmetry: testing and fitting various distributions, an approximate sample mean, one- and two-sample inference, application to earthquake focal mechanisms

Appendix K. Rotation from paleomagnetic vector and initially vertical dike

In this section, we describe how to determine one or two rotations \mathbf{R} from the Troodos mean vector \mathbf{t} , an observed paleomagnetic vector \mathbf{m} , and an observed dike pole \mathbf{d} (Section 3.4). The method is essentially that of Allerton and Vine (1987), but recast in linear-algebraic notation.

First, requiring \mathbf{R} to rotate \mathbf{t} into alignment with \mathbf{m} leads to the vector equation $\mathbf{R}\mathbf{t} = \mathbf{m}$. This equation does not determine \mathbf{R} uniquely. For example, if \mathbf{R} is any solution and \mathbf{S} is any rotation about \mathbf{m} , then the matrix \mathbf{SR} also solves the equation: $\mathbf{SR}\mathbf{t} = \mathbf{Sm} = \mathbf{m}$. So we additionally assume that the dike's initial orientation was vertical, or as close to vertical as possible. This requirement leads to the scalar equation $(\mathbf{R}^\top \mathbf{d})_3 = 0$. Typically there are either zero or two rotations \mathbf{R} that solve these equations. In the former case, we seek the unique \mathbf{R} that minimizes $(\mathbf{R}^\top \mathbf{d})_3^2$.

Choose a matrix \mathbf{T} such that the first row of \mathbf{T} is \mathbf{t} and the rows of \mathbf{T} form a right-handed orthonormal basis. Then \mathbf{T} is a rotation matrix. Similarly, form a rotation matrix \mathbf{M} whose first row is \mathbf{m} . Also, let

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \sigma & -\sin \sigma \\ 0 & \sin \sigma & \cos \sigma \end{bmatrix},$$

where σ is unknown. Then $\mathbf{M}^\top \mathbf{STt} = \mathbf{m}$, so $\mathbf{M}^\top \mathbf{ST}$ is a solution to $\mathbf{R}\mathbf{t} = \mathbf{m}$. In fact, the entire ambiguity in the solution amounts to the choice of σ in \mathbf{S} .

We now use the dike assumption to select σ . Some algebra shows that

$$(\mathbf{T}^\top \mathbf{S}^\top \mathbf{Md})_3 = \alpha \cos \sigma + \beta \sin \sigma + \gamma,$$

where the coefficients α, β, γ depend on $\mathbf{T}, \mathbf{M}, \mathbf{d}$. After the Weierstrass substitution $\cos \sigma = (1 - s^2)/(1 + s^2)$, $\sin \sigma = 2s/(1 + s^2)$, the equation $(\mathbf{R}^\top \mathbf{d})_3 = 0$ amounts to $as^2 + bs + c = 0$, for new coefficients a, b, c . There are the usual three cases:

- If $b^2 - 4ac > 0$, then there are two solutions s , and hence two equally valid values for σ , \mathbf{S} , and $\mathbf{R} = \mathbf{M}^\top \mathbf{ST}$.

- In the rare case when $b^2 - 4ac = 0$, there is one solution s and hence one \mathbf{R} .
- If $b^2 - 4ac < 0$, then there is no solution s . It is tempting to set $s = -b/(2a)$, because that is where the parabola $y = as^2 + bs + c$ comes closest to $y = 0$. However, this choice of s does not typically produce the \mathbf{R} that makes $\mathbf{R}^\top \mathbf{d}$ closest to horizontal. In fact, the \mathbf{R} produced depends on the particular \mathbf{T} and \mathbf{M} chosen.

When we find ourselves in the third case, we instead optimize $(\mathbf{T}^\top \mathbf{S}^\top \mathbf{M} \mathbf{d})_3$ with respect to σ . The derivative equals zero only when $\tan \sigma = \beta/\alpha$. The two solutions for σ imply two values for \mathbf{R} . We select the one that makes $(\mathbf{R}^\top \mathbf{d})_3$ smaller in absolute value.

Appendix L. Monoclinic transpression

In this section we summarize some theory of monoclinic transpression/transtension (Fossen and Tikoff, 1993). The velocity gradient tensor is

$$\mathbf{L} = \begin{bmatrix} 0 & \gamma & 0 \\ 0 & \log k & 0 \\ 0 & 0 & -\log k \end{bmatrix},$$

and the solution is $\mathbf{x}(t) = (\exp t\mathbf{L})\mathbf{x}(0)$ for any initial condition $\mathbf{x}(0)$ (e.g., Davis and Titus, 2011). Because $t\mathbf{L}$ is linear in γ and $\log k$, there is a time-intensity tradeoff. For example, running a deformation with a given $(\gamma, \log k)$ for one unit of time produces the same finite deformation as running $(3\gamma, 3\log k)$ for one third as much time. When one envisions the landscape of monoclinic transpressions as a $(\gamma, \log k)$ -plane, as in Fig. 9, each finite deformation plots at a particular point. As any given steady transpression $(\gamma, \log k)$ proceeds through time, its cumulative finite deformation traces out a straight trajectory from the origin to the point $(\gamma, \log k)$. The angle $\arctan(-(\log k)/\gamma)$ between this trajectory and the γ -axis is the angle of oblique convergence.

The position gradient tensor is

$$\mathbf{F} = \exp \mathbf{L} = \begin{bmatrix} 1 & \frac{\gamma(k-1)}{\log k} & 0 \\ 0 & k & 0 \\ 0 & 0 & \frac{1}{k} \end{bmatrix}.$$

The finite strain axes are the eigenvectors of

$$\mathbf{FF}^\top = \begin{bmatrix} 1 + \frac{\gamma^2(k-1)^2}{(\log k)^2} & \frac{\gamma k(k-1)}{\log k} & 0 \\ \frac{\gamma k(k-1)}{\log k} & k^2 & 0 \\ 0 & 0 & \frac{1}{k^2} \end{bmatrix}.$$

There is always a vertical eigenvector with eigenvalue k^{-2} . The other two eigenvalues are

$$\frac{-b \pm \sqrt{b^2 - 4k^2}}{2},$$

where

$$b = -1 - k^2 - \frac{\gamma^2(k-1)^2}{(\log k)^2}.$$

Their eigenvectors are horizontal. The eigenvalue associated to the vertical eigenvector equals the larger of the other two eigenvalues when

$$\frac{1}{k^2} = \frac{-b + \sqrt{b^2 - 4k^2}}{2}.$$

This equation implies that

$$\gamma^2 = \frac{(\log k)^2(k+1)^2(k^2+1)}{k^2}.$$

For dextral transpression, where $\gamma > 0$ and $0 < k < 1$,

$$\gamma = \frac{-(\log k)(k+1)\sqrt{k^2+1}}{k}. \quad (\text{L.1})$$

When γ and $\log k$ satisfy this equation, \mathbf{FF}^\top can be checked to have eigenvalues k^{-2} , k^{-2} , k^4 . That is, the finite strain ellipsoid is indeed an oblate spheroid.

Similar calculations can be used to derive or check the following statements. When γ is less than the right side of Eq. (L.1), the vertical eigenvector has the greatest eigenvalue, so lineation is vertical. When γ is greater than the right side, lineation is horizontal. Now regard Eq. (L.1) as defining a function $\gamma = \gamma(\log k)$. The tangent line at $\log k = 0$ has equation $\gamma = -\sqrt{8}\log k$, and it is the steepest line through the origin that does not exceed the function. This line corresponds to a steady monoclinic transpression with angle of oblique convergence $\arctan(8^{-1/2}) \approx 19.5^\circ$. Steady transpressions with larger angles of oblique convergence produce vertical lineation at all times t . Steady transpressions with smaller angles of oblique convergence produce horizontal lineation at small t and vertical lineation at large t .