AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2015JB012544

Estimation and propagation of uncertainties associated with paleomagnetic directions

David Heslop[1] **and Andrew P. Roberts**[1]

[1] Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia

Key Points:
- Paleomagnetic directions are placed in a probabilistic framework
- Directional uncertainties can be estimated readily from demagnetization data
- Directional uncertainties can be propagated through a paleomagnetic reconstruction

Correspondence to:
D. Heslop,
david.heslop@anu.edu.au

Citation:
Heslop D., and A. P. Roberts (2016), Estimation and propagation of uncertainties associated with paleomagnetic directions, *J. Geophys. Res.*, *121*, 2274–2289, doi:10.1002/2015JB012544.

Received 22 SEP 2015
Accepted 6 APR 2016
Accepted article online 11 APR 2016
Published online 27 APR 2016

Abstract Principal component analysis (PCA) is a well-established technique in paleomagnetism and provides a means to estimate magnetic remanence directions from univectorial segments of stepwise demagnetization data. Derived directions constrain past geomagnetic field behavior and form the foundation of chronological and tectonic reconstructions. PCA of isolated remanence segments relies on estimates of the segment mean and covariance matrix, which can carry large uncertainties given the relatively small number of demagnetization data points used to characterize individual specimens. Traditional PCA does not, however, lend itself to quantification of these uncertainties, and inferences drawn from paleomagnetic reconstructions suffer from an inability to propagate uncertainties from individual specimens to higher levels, such as in calculations of paleomagnetic site mean directions and pole positions. In this study, we employ a probabilistic reformulation of PCA that represents the unknowns involved in the data fitting process as probability density functions. Such probability density functions represent our state of knowledge about the unknowns in the fitting process and provide a tractable framework with which to rigorously quantify uncertainties associated with remanence directions estimated from demagnetization data. These uncertainties can be propagated readily through each step of a paleomagnetic reconstruction to enable quantification of uncertainties for all stages of the data interpretation sequence, removing the need for arbitrary selection/rejection criteria at the specimen level. Rigorous uncertainty determination helps to protect against spurious inferences being drawn from uncertain data.

1. Introduction

Paleomagnetic remanence directions preserved in rocks and sediments provide key paleomagnetic and geological information. A given paleomagnetic specimen may, however, contain multiple superimposed remanence components that were acquired over different geological time periods. A key task in paleomagnetism is to separate, quantify, and attribute these remanence components to specific geomagnetic or geological events and processes [*Irving*, 1964; *McElhinny*, 1973; *Tarling*, 1983]. Stepwise demagnetization of a paleomagnetic specimen, typically with thermal or alternating field (AF) techniques, is used to isolate different remanence components [*As and Zijderveld*, 1958; *Collinson et al.*, 1967]. Specifically, within a demagnetization sequence, sets of consecutive data points that lie along a straight line (i.e., a univectorial trend) are attributed to the removal of a single remanence component. The estimated direction of an identified remanence component can then be used for paleomagnetic and geological reconstructions. Given the utility of paleomagnetic directions, it is essential that they can be estimated from demagnetization data in a rigorous manner and where possible their uncertainty should be determined. Properly quantified uncertainties are key to a quantitative understanding of geomagnetic field variability and constraining tectonic reconstructions.

Kirschvink [1980] proposed the use of principal component analysis (PCA) to estimate the best fit line to a sequence of stepwise acquired demagnetization data for an individual specimen. Identification of remanence components is typically performed interactively through inclusion and removal of demagnetization points in the PCA analysis. Simple statistics (see section 2) can then be used to compare the different candidate models. Two sources of uncertainty limit the utility of PCA when analyzing demagnetization data. First, experimental noise can make individual specimen directions highly uncertain. The maximum angular deviation (MAD) statistic was developed to quantify this uncertainty and provides a basis to reject specimens with poor demagnetization behavior [*Kirschvink*, 1980]. However, the MAD is limited in its utility because it cannot be readily propagated. Thus, the directional uncertainty represented by the MAD is typically ignored above the

specimen level. Such a failure to propagate specimen level uncertainties through the paleomagnetic data interpretation sequence inhibits quantification of the uncertainty associated with higher-level statistics, such as paleomagnetic site mean directions. In turn, this issue will persist through the reconstruction of virtual geomagnetic pole (VGP) positions, apparent polar wander paths, etc. Second, quantification of remanence directions is typically based on a small number of demagnetization steps along a selected segment of a larger data set ($\leq$10 measurement points in most studies). To provide reliable results, PCA requires many more cases (the number of demagnetization steps) than the dimension (three for paleomagnetic directions) of the problem [*Guadagnoli and Velicer*, 1988]. This small sample uncertainty is not represented by currently used statistics, such as the MAD, and its influence on paleomagnetic reconstructions remains unconstrained, but must be considered.

A number of studies have extended the approach of *Kirschvink* [1980] to facilitate objective model selection and to define confidence regions for paleomagnetic directions. *Schmidt* [1982] developed an extension to the PCA-based approach of *Kirschvink* [1980] termed linearity spectrum analysis (LSA). Instead of considering specimens in isolation, LSA is based on analysis of groups of specimens from a homogenous source to isolate demagnetization segments that exhibit maximum linearity. Through analysis of specimens in combination, LSA can readily reduce a collection of demagnetization data into its geologically relevant parts while also providing information concerning interspecimen consistency of demagnetization behavior. The shortcomings of existing PCA-based analysis were identified by *Kent et al.* [1983], who considered the analysis of demagnetization data from a geometrical perspective. The LINEFIND algorithm developed by *Kent et al.* [1983] performed a collection of hypothesis tests to assess the linearity of all possible segments within a demagnetization sequence. The set of segments deemed to be linear is then thinned into a subset of nonoverlapping segments. In a similar manner, a subset of segments with planar demagnetization behavior is also constructed. Within the statistical framework developed by *Kent et al.* [1983] this search process provides an estimate of the linear and planar segments within a demagnetization sequence and their associated confidence limits. *McFadden and Schmidt* [1986] compared how measurement errors were considered in the LSA and LINEFIND approaches and provided a statistical framework to estimate overall mean directions and their cones of confidence based on these errors. It is not apparent why the methods of *Schmidt* [1982], *Kent et al.* [1983], and *McFadden and Schmidt* [1986] have not been employed more widely by the paleomagnetic community. Instead, the PCA-based approach of *Kirschvink* [1980] combined with interactive identification of remanence components has become the default analysis technique. More recently, *Khokhlov and Hulot* [2016] developed an approximation with which the PCA fitting statistics provided by *Kirschvink* [1980] could be transformed into the framework of *Fisher* [1953] probability distributions. Much of the existing paleomagnetic data interpretation sequence is based on *Fisher* [1953] statistics; therefore, an ability to place PCA of specimen demagnetization data within this framework may prove to be fruitful.

As paleomagnetic studies address ever more complex problems, there is a pressing need to quantify directional uncertainties through each stage of a data interpretation sequence in a paleomagnetic reconstruction. Such uncertainty propagation is essential to quantitative studies. It provides a rigorous basis for handling and understanding uncertainties, and it protects against spurious inferences being drawn from uncertain data. In this paper, we show how PCA can be placed within a probabilistic framework that enables quantification of directional uncertainties that result from experimental noise and small data sets. These uncertainties are represented in such a way that they can be propagated readily to higher levels in a data interpretation sequence in paleomagnetism (e.g., when calculating site mean directions and VGPs). Additionally, uncertainties can be separated to provide inclination- and declination-specific errors and to allow probabilistic assessment of indicative properties, such as whether a characteristic remanent magnetization (ChRM) direction will intercept the origin of a vector demagnetization diagram [*Zijderveld*, 1967].

2. Principal Component Analysis

When interpreting stepwise demagnetization data for a paleomagnetic specimen, it is necessary to identify segments of univectorial behavior that are considered to carry geomagnetic or geological information [*Zijderveld*, 1967; *Dunlop*, 1979]. Once a segment of N measurements along a univectorial demagnetization path has been selected, its intensity, F, inclination, I, and declination, D, can be converted to a Cartesian

coordinate system:

$$t_1(i) = F(i)\cos I(i)\cos D(i), \tag{1}$$

$$t_2(i) = F(i)\cos I(i)\sin D(i), \quad \text{and} \tag{2}$$

$$t_3(i) = F(i)\sin I(i), \tag{3}$$

where i corresponds to the ith measurement in the segment. If the mean of these points is calculated:

$$\bar{\mathbf{t}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{t}(i), \tag{4}$$

then the data can be centered by determining the deviations:

$$\delta\mathbf{t}(i) = \mathbf{t}(i) - \bar{\mathbf{t}}, \tag{5}$$

and the covariance matrix of the demagnetization data can be defined:

$$\mathbf{S} = \frac{1}{N}\begin{pmatrix} \sum \delta t_1(i)\delta t_1(i) & \sum \delta t_1(i)\delta t_2(i) & \sum \delta t_1(i)\delta t_3(i) \\ \sum \delta t_2(i)\delta t_1(i) & \sum \delta t_2(i)\delta t_2(i) & \sum \delta t_2(i)\delta t_3(i) \\ \sum \delta t_3(i)\delta t_1(i) & \sum \delta t_3(i)\delta t_2(i) & \sum \delta t_3(i)\delta t_3(i) \end{pmatrix}. \tag{6}$$

PCA considers the squared error between a measured remanence, $\mathbf{t}(i)$, and its optimal linear reconstruction, $\hat{\mathbf{t}}(i)$, to minimize $\sum_i \parallel \mathbf{t}(i) - \hat{\mathbf{t}}(i) \parallel$ over the selected demagnetization segment. The direction of the best fit (least squares) vector through the demagnetization data is found by solving the eigenvalue problem:

$$\mathbf{S}\mathbf{u}_i = \lambda_i\mathbf{u}_i, \quad i = 1, \ldots, 3, \tag{7}$$

where $\mathbf{u}_i$ and λ_i are the eigenvectors and eigenvalues of $\mathbf{S}$, respectively (reordering of the eigenvalues and eigenvectors may be necessary to ensure that $\lambda_1 \geq \lambda_2 \geq \lambda_3$). The eigenvectors are normalized to unit length, and I and D of the paleomagnetic vector are given by the leading eigenvector $\mathbf{u}^*$:

$$I = \arcsin\left(\mathbf{u}_3^*\right), \quad \text{and} \tag{8}$$

$$D = \arctan\left(\mathbf{u}_2^*/\mathbf{u}_1^*\right). \tag{9}$$

For univectorial demagnetization, deviations of data from the fitted vector are expected to be minor and are attributable to the random walk produced by the demagnetization of an ensemble of magnetic particles [*Khokhlov and Hulot*, 2016] or to instrument noise; therefore, $\lambda_1 \gg \lambda_2 \simeq \lambda_3$. *Kirschvink* [1980] proposed MAD as a measure of directional uncertainty, which treats demagnetization data as if they are drawn from a multivariate Gaussian distribution with a mean of $\bar{\mathbf{t}}$ and covariance matrix that describes a prolate ellipsoid (i.e., $\lambda_1 \gg \lambda_2 = \lambda_3$). The MAD represents the conical angle between the leading eigenvector and the minor cross section of the prolate ellipsoid projected onto the surface of a unit sphere:

$$\mathrm{MAD} = \arctan\left(\frac{\sqrt{\lambda_2^2 + \lambda_3^2}}{\lambda_1}\right). \tag{10}$$

An example of a least squares PCA fit to demagnetization data is given in Figure 1. Once a sequence of remanence directions that corresponds to a paleomagnetic component of interest has been selected, it is simple to calculate the direction of the best fit vector and corresponding MAD. While fitted directions must carry some level of uncertainty (represented by a nonzero MAD), it is challenging to quantify this uncertainty in a useful manner. For example, the MAD can act as a sample acceptance/rejection criterion based on a suitably chosen cutoff [e.g., *Leonhardt et al.*, 2004]; however, it does not provide an obvious means to propagate the uncertainty that it represents to higher levels, such as when calculating paleomagnetic site mean directions.

Kirschvink [1980] proposed that when a demagnetization segment is considered to represent a ChRM, the PCA solution can be centered on the origin of the Cartesian coordinate system. This is termed an

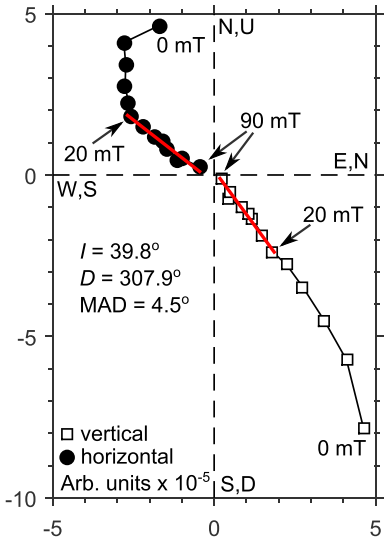

Figure 1. Orthogonal vector demagnetization diagram with alternating field demagnetization results for sample eba24a [from *Tauxe*, 1998]. The sample contains a readily demagnetized viscous component (0–20 mT), and the ChRM is defined over the 20–90 mT demagnetization interval. PCA of the ChRM interval (red lines) provides a least squares estimate of *I*, *D*, and the MAD.

"anchored fit" and will ensure that the fitted vector passes through the origin of the vector demagnetization diagram of *Zijderveld* [1967]. In practice this can be achieved by removing the $\bar{\mathbf{t}}$ term from equation (5) and calculating the covariance matrix from the data directly rather than the deviations. Although this provides a simple approach to force a fitted vector through the origin, it can produce a detrimental elongation in the covariance structure of the data that will usually result in a spurious reduction of the overall uncertainty (Figure 2). When MADs are reported for anchored fits, the extent to which they have been reduced by artificial elongation of the covariance structure is not apparent, making it difficult to assess the true directional uncertainty.

McFadden and Schmidt [1986] considered the direction of the best fit vector as a random draw from a *Fisher* [1953] distribution. This allowed them to estimate an approximate uncertainty expressed as a $p \times 100\%$ confidence region for the true paleomagnetic direction that is defined by a cone with semiangle (a_p):

$$a_p = \arccos\left(1 - \frac{\lambda_2 + \lambda_3}{\lambda_1}\left[\left(\frac{1}{1-p}^{\frac{1}{N-1}}\right) - 1\right]\right). \tag{11}$$

The fidelity of a PCA-derived direction and its associated MAD depend on **S** being representative of both the true paleomagnetic vector and the measurement errors. Paleomagnetic directions are typically defined

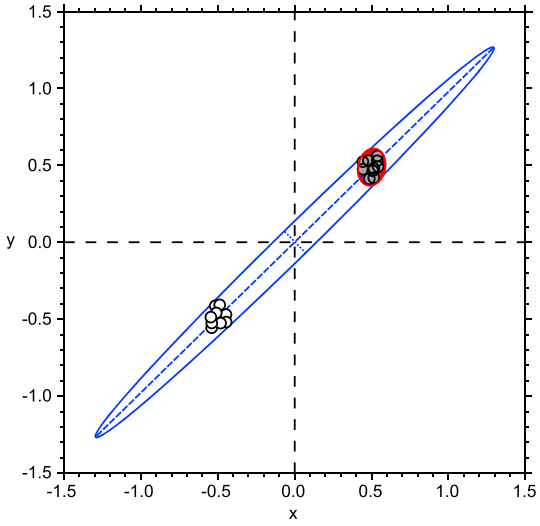

Figure 2. Two-dimensional example of the covariance distortion that can occur when data are not centered. The data (shaded circles) have a covariance ellipse (red) that is almost circular, which indicates that they are close to being isotropic. When the covariance is calculated without centering the data, the ellipse becomes strongly elongated (blue), which artificially lengthens the leading principal component (dashed line) with respect to the second principal component (dotted blue line). It is simple to illustrate why this elongation occurs. Using noncentered data to define the covariance structure is analogous to creating a second set of data points (open circles) on the opposite side of the coordinate system origin. The noncentered covariance structure is effectively based on both sets of points; thus, the ellipse is artificially elongated rather than being approximately isotropic.

using between 4 and 10 demagnetization points; however, the eigenstructure of **S** can be distorted when the number of dimensions (three) is close to *N* [*Dempster*, 1969]. Thus, it is necessary to consider uncertainties that originate from measurement errors and from working with a small *N* and how, in combination, they can be propagated through each step in a paleomagnetic data processing sequence.

3. A Probabilistic PCA Framework

The MAD ellipsoid concept of *Kirschvink* [1980] can be extended such that demagnetization data can be viewed as a realization from a constrained three-dimensional Gaussian distribution that is composed of signal and noise components. For the signal component, univectorial demagnetization data should fall on a line that passes through three-dimensional Cartesian space. If the direction of the line is represented by the three-element vector, **W**, then the covariance matrix that describes the signal is $\mathbf{W}\mathbf{W}^T$, which has a rank of 1 (i.e., the data are univectorial). When combined with the mean, μ, of the noise-free remanence measurements, the signal can be represented by a multivariate Gaussian distribution $\mathcal{N}(\mu, \mathbf{W}\mathbf{W}^T)$.

Kent et al. [1983] proposed a specific noise model for demagnetization data, whereby the error variance depends on the magnitude of the remanence. To estimate such a noise model requires several repeated remanence measurements at each demagnetization step. Here we follow the simpler approach of *Kirschvink* [1980], where noise in demagnetization data is assumed to be Gaussian, isotropic, and independent of the remanence magnitude. If noise has standard deviation, σ, it corresponds to a three-dimensional Gaussian distribution $\mathcal{N}(0, \sigma^2\mathbf{I})$, where **I** is a 3 by 3 identity matrix.

The convolution of the signal and noise distributions is also Gaussian and is defined by the parameters μ, **W**, and σ^2:

$$\mathbf{t} \sim \mathcal{N}(\mu, \mathbf{C}), \tag{12}$$

where $\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$. Equation (12) defines a constrained Gaussian model that forms the basis of probabilistic PCA (PPCA) [*Tipping and Bishop*, 1999]. The PPCA representation allows maximum likelihood estimates of μ, **W**, and σ to be found, but the uncertainties associated with these parameters remain unquantified. To overcome this hurdle, the PPCA model can be placed within a Bayesian framework that estimates posterior probability distributions for μ, **W**, and σ^2. These parameters are represented by probability distributions, rather than by point estimates, so their uncertainty can be readily quantified and propagated to higher levels of the paleomagnetic data interpretation sequence. We adopt a variational Bayes approach developed by *Bishop* [1999a], which is computationally efficient and provides Gaussian approximations to the posterior probability distributions (see Appendix A for details). Specifically, the approximate posterior probability distribution of μ is given by $Q(\mu)$, which is a three-dimensional Gaussian distribution with mean $\mathbf{m}_\mu$ and covariance $\mathbf{\Sigma}_\mu$, i.e., $\mathcal{N}(\mathbf{m}_\mu, \mathbf{\Sigma}_\mu)$. Similarly, the direction is also represented by a three-dimensional Gaussian distribution, $Q(\mathbf{W})$, with mean $\mathbf{m}_\mathbf{W}$ and covariance $\mathbf{\Sigma}_\mathbf{W}$, i.e., $\mathcal{N}(\mathbf{m}_\mathbf{W}, \mathbf{\Sigma}_\mathbf{W})$.

An important source of uncertainty in paleomagnetic directions is the correct identification of linear segments within demagnetization data [*McFadden and Schmidt*, 1986]. Although earlier algorithms automated this process [*Schmidt*, 1982; *Kent et al.*, 1983], they have not been adopted widely and most studies are based on interactive data analysis within vector demagnetization diagrams [*Zijderveld*, 1967]. A great advantage of variational Bayes as an approximate inference algorithm is its speed [*Bishop*, 2006]. In future, this speed may be exploited to facilitate rapid comparison between candidate models to provide an objective means of identifying linear demagnetization segments.

4. Uncertainty Determination for a Single Specimen

Once $Q(\mu)$ and $Q(\mathbf{W})$ have been estimated for a given remanence vector (Appendices A and B), two key questions can be addressed. First, what is the directional uncertainty of the remanence vector? Second, for a hypothesized ChRM, do the vector and the origin coincide within uncertainty? Directional uncertainty is represented by $Q(\mathbf{W})$, and random draws from this multivariate Gaussian distribution will provide realizations of the direction of the paleomagnetic vector. To test if a ChRM vector passes through the origin of the Cartesian space, it is necessary to consider $Q(\mu)$ and $Q(\mathbf{W})$ in combination. While $Q(\mathbf{W})$ defines the distribution of possible vector directions, $Q(\mu)$ defines a distribution that represents the true (i.e., noise-free) center of the data points through which the vectors must pass. Random draws from $Q(\mu)$ and $Q(\mathbf{W})$ define vectors that pass

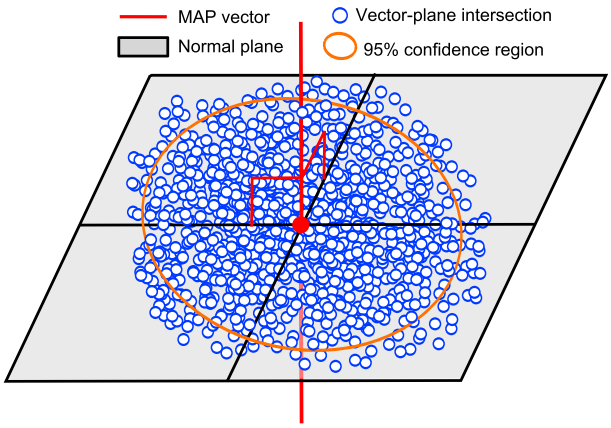

Figure 3. Schematic demonstration of the determination of a confidence region around a given point on a maximum a posteriori probability (MAP) vector. Points of intersection between random vectors generated from $Q(\mu)$ and $Q(\mathbf{W})$ and a plane (shaded) normal to the MAP vector (red) are used to define a circular confidence region (orange). For example, the 95% confidence region is defined by a circle (orange) that bounds 95% of the intersection points of the random vectors and is centered on the MAP intersection point (red symbol).

through the three-dimensional measurement space, which can be compared to the maximum a posteriori probability (MAP) vector that is defined by $\mathbf{m}_\mu$ and $\mathbf{m}_\mathbf{W}$ and that corresponds to the modes of $Q(\mu)$ and $Q(\mathbf{W})$.

A confidence region can be defined around the MAP vector at any point along its length by defining a normal plane that intersects the MAP vector at that point. The intersection of the plane with vectors obtained via random draws from $Q(\mu)$ and $Q(\mathbf{W})$ then provides a collection of points with a distribution of distances along the plane from the selected point on the MAP vector (Figure 3). Determination of the 95th percentile of the distances between the intersection points and the MAP vector then defines a 95% confidence region for the intersection of the true paleomagnetic vector with the defined plane. This process can be repeated for a collection of positions along the MAP vector to define a 95% confidence region through the measurement space (Figure 4a).

To test if the origin of the Cartesian space is contained within the 95% confidence region of the true vector, it is necessary to consider the path of vectors defined by $Q(\mu)$ and $Q(\mathbf{W})$. In combination, random draws $\mathbf{R}_\mu$ and

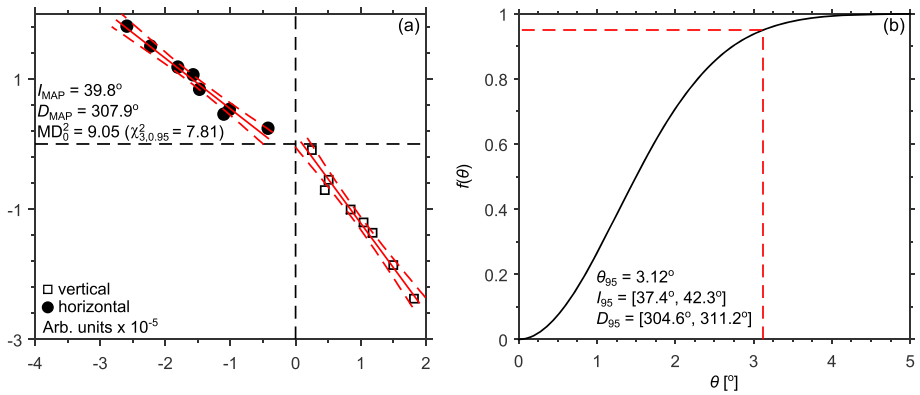

Figure 4. (a) Orthogonal vector demagnetization diagram with alternating field demagnetization results for sample eba24a [from *Tauxe*, 1998], which includes the maximum a posteriori probability (MAP) vector (solid lines) and 95% confidence region for the true vector (dashed lines). Given that $MD_0^2 > \chi^2_{3,0.95}$, the 95% confidence region of the identified ChRM does not contain the origin of the diagram and must be interpreted within this constraint. (b) The multivariate Gaussian distribution $Q(\mathbf{W})$ calculated for sample eba24a defines a distribution of off-axis angles (black line, equation (15)). The 95th percentile of the distribution defines a 95% confidence region (θ_{95}, red dashed line) around the MAP vector for the true direction of the isolated remanence component in sample eba24a. This distribution can be marginalized [*Love and Constable*, 2003] to provide separate uncertainties on the ChRM inclination (I_{95}) and declination (D_{95}).

$\mathbf{R_W}$ from the distributions $Q(\mu)$ and $Q(\mathbf{W})$, respectively, define a vector that passes through Cartesian space. The position, $\mathbf{Z}$, along the vector that is closest to the origin is given by

$$\mathbf{Z} = -\mathbf{R}_\mu \mathbf{R}_\mathbf{W}^{+} \mathbf{R}_\mathbf{W} + \mathbf{R}_\mu, \tag{13}$$

where $^+$ is the Moore-Penrose pseudoinverse [*Penrose*, 1955]. If this process is repeated for a large number of random draws, it is possible to define an empirical distribution for $\mathbf{Z}$ and estimate its mean ($\bar{\mathbf{Z}}$) and covariance matrix ($\mathbf{S_Z}$). Assuming that $\mathbf{Z}$ is approximately multivariate Gaussian, the squared *Mahalanobis* [1936] distance between the origin and $\bar{\mathbf{Z}}$ is given by

$$\mathrm{MD}_0^2 = \bar{\mathbf{Z}}^T \mathbf{S_Z}^{-1} \bar{\mathbf{Z}}. \tag{14}$$

If the origin falls within the 95% confidence region for the true paleomagnetic vector, then based on a χ^2 distribution with three degrees of freedom, $\mathrm{MD}_0^2 \leq \chi^2_{3,0.95} = 7.81$ [*Rousseeuw and van Zomeren*, 1990] (Figure 4a). It is important to note that visual inspection of confidence bands within an orthogonal vector demagnetization diagram can be misleading. It is intuitive to assume that if the origin lies within the two plotted confidence bands, then the origin must be located within the 95% confidence region of the true paleomagnetic vector. However, the two projections of a vector in an orthogonal plot depend on each other and a given point on one projection corresponds to a single point on the second projection. Therefore, it is insufficient for the two confidence regions to contain the origin; instead, locations within the confidence region that correspond to the same position along the paleomagnetic vector must contain the origin simultaneously. A test indicating that a vector is not coincident with the origin within uncertainty does not imply that a given specimen should be rejected as an unreliable recorder of the paleomagnetic field. Instead, the vector direction should be considered within the context of the statistical test result and geological interpretations should take this factor into consideration. As discussed in section 2, anchoring a PCA fit through the origin of the vector demagnetization diagram will elongate the multivariate Gaussian distribution toward the origin and artificially reduce the MAD value. Our proposed approach enables probabilistic evaluation of whether a paleomagnetic vector will intercept the origin within the uncertainties of the data and, thus, removes the requirement for anchoring.

Estimation of the directional uncertainty of the remanence vector is straightforward. The distribution of angles, θ, between draws from $Q(\mathbf{W})$ and the MAP direction given by $\mathbf{m_W}$ is a Gaussian off-axis angular distribution, with a probability density function (pdf) [*Love and Constable*, 2003]:

$$\begin{aligned} p(\theta \mid \boldsymbol{\Sigma}_\mathbf{W}, \|\mathbf{m_W}\|) = \frac{1}{2}\sin\theta \left[-\frac{1}{2}\frac{\|\mathbf{m_W}\|^2}{\boldsymbol{\Sigma}_\mathbf{W}}\right] \\ \times \left\{ \left[1 + \frac{\|\mathbf{m_W}\|^2}{\boldsymbol{\Sigma}_\mathbf{W}}\cos^2\theta\right] \exp\left[\frac{1}{2}\frac{\|\mathbf{m_W}\|^2}{\boldsymbol{\Sigma}_\mathbf{W}}\cos^2\theta\right] \right. \\ \left. \times \left[1 + \mathrm{erf}\left[\frac{1}{\sqrt{2}}\left(\frac{\|\mathbf{m_W}\|}{\sqrt{\boldsymbol{\Sigma}_\mathbf{W}}}\right)\cos\theta\right]\right] + \sqrt{\frac{2}{\pi}}\left(\frac{\|\mathbf{m_W}\|}{\sqrt{\boldsymbol{\Sigma}_\mathbf{W}}}\right)\cos\theta \right\}, \end{aligned} \tag{15}$$

where $\|\mathbf{m_W}\|$ is the magnitude of the $\mathbf{m_W}$ vector. A given quantile of this distribution then defines a circular confidence region on a unit sphere centered at $\mathbf{m_W}/\|\mathbf{m_W}\|$ and within which the true direction is expected to occur with a given probability. For example, the 95th percentile defines a 95% circular confidence region, which we term θ_{95} (Figure 4b). The boundary of this confidence region can be represented as a sequence of inclination/declination pairs and can, therefore, be illustrated on stereoplots or transformed to provide a confidence region for VGPs, etc. When considering records of geomagnetic secular variation, it may be desirable to consider the uncertainty in inclination and declination separately. *Love and Constable* [2003] provided marginal density functions for both inclinations (their Appendix F) and declinations (their Appendix G), from which separate confidence intervals for inclinations and declinations can be calculated (Figure 4b). It is important to note that because of their fundamentally different forms, θ_{95} and MADs cannot be compared directly.

The performance of variational Bayesian PCA (VBPCA) for estimating uncertainties, as introduced above in section 3, was assessed using a numerical simulation. A given number of data points, N, were created along a straight line to represent univectorial demagnetization data. Isotropic Gaussian pseudorandom numbers were then added to the points to simulate noise. PCA and VBPCA were then performed on the simulated data, and for each approach the angle between the estimated and true vectors was calculated. In the case of the PCA solution, equation (11) was solved to find the value of p where a_p is equal to the angle between the

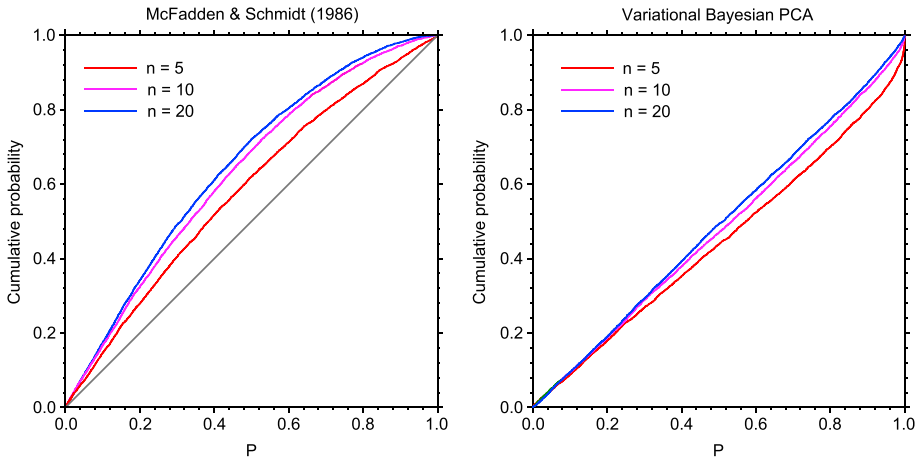

Figure 5. Empirical cumulative distribution functions of *p* values for uncertainty estimates obtained from the approximation of (left) *McFadden and Schmidt* [1986] and (right) VBPCA. Bias for a given number of data points (*N*) is expressed as a deviation from a line with 1:1 slope (gray). The VBPCA approach gives rise to less bias than the approximation of *McFadden and Schmidt* [1986], and the bias decreases with increasing *N*.

least squares best fit and true vector. For the VBPCA solution, the equivalent *p* value for the angle between the MAP and true vector was determined from a cumulative Gaussian off-axis angular distribution defined by $\|\mathbf{m_W}\|$ and $\mathbf{\Sigma_W}$ [*Love and Constable*, 2003, equation E8]. This process was repeated 5000 times with noise levels producing cases with MADs between 4° and 19°. If a given technique produces unbiased estimates of the uncertainty associated with an estimated vector direction, the observed *p* values should be distributed uniformly in the interval [0, 1]; therefore, their empirical cumulative distribution function will have a 1:1 slope between 0 and 1. Empirical distributions of *p* values are shown in Figure 5, with the VBPCA approach giving rise to less bias than the approximate uncertainty of *McFadden and Schmidt* [1986]. Additionally, VBPCA bias decreases with increasing *N*, while the bias in the approximate uncertainty estimate of *McFadden and Schmidt* [1986] increases. This demonstrates the suitability of VBPCA for estimating paleomagnetic directions and their uncertainties from demagnetization data.

An ability to estimate uncertainties at the specimen level is crucial for evaluating records of polarity reversals or excursions, where weak transitional fields may result in degraded paleomagnetic recording. To demonstrate this approach, an inclination record originally presented by *Heslop et al.* [2013] from marine sediment core MD00-2361 (offshore of Western Australia) is shown in Figure 6. The presented core interval spans the Matuyama-Brunhes boundary; however, the specimens were not oriented azimuthally and absolute

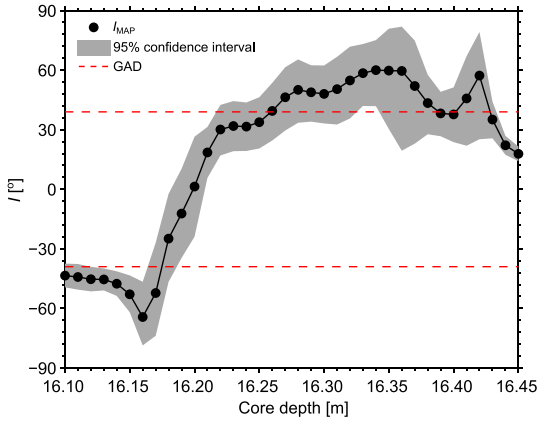

Figure 6. Estimated inclinations (closed symbols) and 95% confidence interval (shaded) through a record of the Matuyama-Brunhes boundary from marine sediment core MD00-2361 from offshore of Western Australia [*Heslop et al.*, 2013]. U-channel samples were AF demagnetized [*Weeks et al.*, 1993], and the 50–100 mT interval was used to represent the ChRM. Geocentric axial dipole (GAD) inclinations are shown for reference (red dashed lines). While the inclination record looks smooth, the inclination uncertainty at each specimen depth can be as much as ±15°.

declination information is not available. After fitting the specimen ChRMs using the outlined approach, uncertainties associated with the estimated ChRM inclinations were determined using the Gaussian marginal density function for inclination [*Love and Constable*, 2003, equation F1].

5. Uncertainty Propagation

Once defined rigorously at the specimen level, it is possible to propagate uncertainties through subsequent paleomagnetic data analyses, such as when calculating paleomagnetic site mean directions. Given that $Q(\mathbf{W})$ corresponds to a Gaussian distribution for a given specimen, it would be tempting to simply convolve the distributions for collections of specimens to produce another Gaussian distribution that represents the mean paleomagnetic direction. In the paleomagnetic context, however, $Q(\mathbf{W})$ defines a collection of directions and this property must be respected. The convolution of two Gaussian off-axis angular distributions does not produce another Gaussian off-axis angular distribution [*de Sa*, 2007]. Therefore, we propagate directional uncertainties numerically as described below.

5.1. Calculation of Site Mean Directions

In situations where paleomagnetic records, such as those from lava flows, provide a snapshot of the field, or where a long sequence of flows is thought to be representative of the average field over a given time interval, it is desirable to calculate a mean direction and its α_{95} uncertainty using the framework developed by *Fisher* [1953]. This is a common approach in paleomagnetic studies; however, it does not consider the directional uncertainties at the specimen level.

The framework introduced in section 4 provides a way to incorporate specimen uncertainties into calculation of a mean. However, as discussed above it is necessary to propagate directional uncertainties numerically. To achieve this, consider an ensemble of N^s specimens collected from a site. If $Q(\mathbf{W})$ has been determined for each specimen, then a unit vector that corresponds to a realization of the paleomagnetic direction of the ith specimen is given by

$$\tilde{\mathbf{W}}_i = \frac{\mathbf{W}_i}{\|\mathbf{W}_i\|}. \tag{16}$$

This is repeated for all N^s specimens, and the length, R, of the resultant vector produced by summing the sample of unit vectors is calculated. Assuming that the specimen directions originate from a *Fisher* [1953] distribution, a realization of the site mean paleomagnetic direction, $\mathbf{x^s}$, can be calculated by permuting the direction of the $\tilde{\mathbf{W}}$ resultant vector with a random angle, ϕ, given by [*Fisher*, 1953]

$$\phi = \arccos\left(1 - \frac{N^s - R}{R}\left[\left(\frac{1}{P}\right)^{\frac{1}{N^s-1}} - 1\right]\right), \tag{17}$$

where P is a uniformly distributed random number between 0 and 1. The process of drawing a random direction for each specimen, combining these directions and permuting by a random angle, can then be repeated a large number of times, N_b (typically $\sim 10^4$), to produce an ensemble of site mean directions, $\mathbf{X^s}$, that incorporates the uncertainty in individual specimen directions. If the confidence region for the true site paleomagnetic mean direction is assumed to be circular, it can be defined by finding the 95th percentile ($\beta_{95-\text{circular}}$) of the angles between the cases of $\mathbf{X^s}$ and the mean of $\mathbf{X^s}$. Alternatively, if a circular confidence region is not assumed, a nonparametric 95% confidence region for the true site mean direction can be defined by kernel density estimation. It is necessary to determine the maximum density that defines a contour on the surface of a unit sphere that encloses 95% of the points in $\mathbf{X^s}$. For a given point, $\mathbf{x_0}$, on the surface of a unit sphere, an unnormalized estimate of the density, $\hat{f}$, based on a Fisher distribution kernel is given by *Hall et al.* [1987]

$$\hat{f}(\mathbf{x_0}, \tilde{\kappa}) = \sum_{i=1}^{N_b} \exp\left(\tilde{\kappa}\mathbf{X_i^{s\mathsf{T}}}\mathbf{x_0}\right), \tag{18}$$

where $\tilde{\kappa}$ represents the precision parameter of a kernel based on a Fisher distribution [*Fisher*, 1953]. A suitable value of $\tilde{\kappa}$ can be estimated as [*Garcia-Portugués*, 2013]

$$\tilde{\kappa} = \sqrt[3]{\frac{\hat{\kappa}[(1+4\hat{\kappa}^2)\sinh(2\hat{\kappa}) - 2\hat{\kappa}\cosh(2\hat{\kappa})]N_b}{8\sinh^2(\hat{\kappa})}}, \tag{19}$$

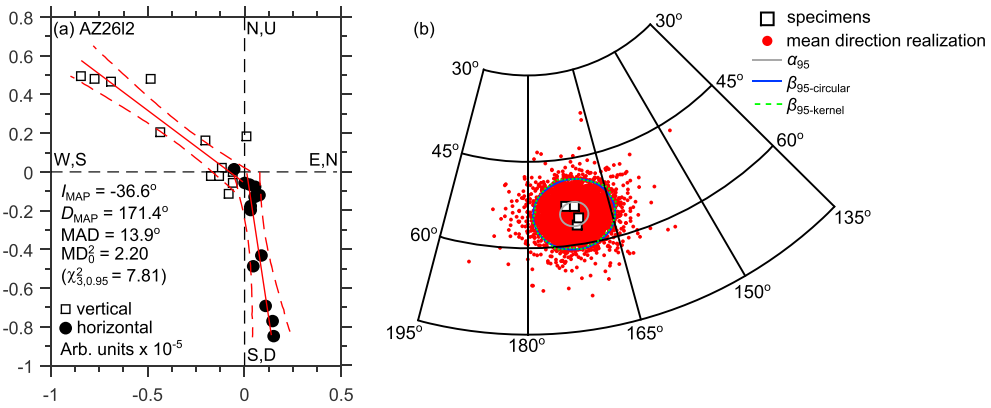

Figure 7. (a) Demagnetization data, for example specimen l2 from site AZ26 on São Miguel Island, Azores [*Johnson et al.*, 1998]. Solid red lines indicate the maximum a posteriori probability (MAP) vector, and dashed red lines represent the 95% confidence region. (b) When four specimens (open symbols) are averaged, an α_{95} can be calculated (gray line), which does not account for the directional uncertainty carried by individual specimens (as shown in Figure 7a). Numerical simulation of an ensemble of mean directions (red dots) allows estimation of a 95% confidence region for the true mean based on an angle that defines a circle that bounds 95% of the ensemble ($\beta_{95-\text{circular}}$, blue line) or by kernel density estimation ($\beta_{95-\text{kernel}}$, green dashed line).

where $\hat{\kappa}$ corresponds to an estimate of the precision parameter of the points in $\mathbf{X^s}$. If equation (18) is used to estimate the density at each location in $\mathbf{X^s}$, then the 5th percentile of these densities corresponds to a density, which when mapped on the surface of a unit sphere will produce a contour, $\beta_{95-\text{kernel}}$, enclosing 95% of the points in $\mathbf{X^s}$.

When specimen directions carry no uncertainty (i.e., $\mathbf{\Sigma_W} = 0$), the β_{95} region will correspond to the α_{95} value commonly used in paleomagnetic studies. Given that real specimen directions will always carry an uncertainty, the β_{95} region will be larger than that defined by an α_{95} and should allow more robust inferences to be drawn concerning site mean directions. As with the single specimen case discussed in section 4, the limits of the site mean confidence region are defined simply by a collection of directions (i.e., points on the surface of a unit sphere) and can be readily converted into other representations, such as VGPs.

An example of propagating specimen uncertainties into a site mean is given in Figure 7, which is based on data from *Johnson et al.* [1998] who studied lavas from São Miguel Island, Azores. *Johnson et al.* [1998] applied strict specimen acceptance/rejection criteria (e.g., ChRM MADs must be <5°) before calculating site means. To demonstrate uncertainty propagation, we selected four specimens (a1, a3, j2, and l2) from site AZ26 that yield similar directions but have MADs >12°. Thus, our example is intended to be illustrative of situations involving small numbers of specimens with large individual uncertainties, rather than a reassessment of the stringent analysis of *Johnson et al.* [1998]. In a traditional PCA approach, these four specimens have similar ChRM directions and their average yields $\alpha_{95} = 2.1°$ (Figure 7). If, however, the individual ChRM uncertainties are propagated into the average, the uncertainty increases by a factor of ~3, reaching $\beta_{95-\text{circular}} = 6.2°$. In this example the specimens have similar MADs and were assumed to carry equal weight in the calculation of the site mean direction. In cases where the specimens from a site have a wide range of uncertainties, it would be feasible to calculate a weighted site mean direction with individual specimen weights derived from their uncertainties.

5.2. Combining Directional Records

To demonstrate a more complex example, we consider the combination of uncertainties in paleomagnetic directions during a period of excursional field geometry. Based on a compilation of sedimentary records from the North Atlantic Ocean and South China Sea, *Laj et al.* [2006] hypothesized that the Iceland Basin excursion (~190 ka) was dominated by changes in the dipole component of the field. *Laj et al.* [2006] showed an additional VGP path reconstructed by *Roberts et al.* [1997] from Ocean Drilling Program (ODP) Site 884. This path contains a more complex double loop behavior; however, age control at ODP Site 884 was limited to an indirect chronology, so *Laj et al.* [2006] excluded it from their final compilation. Subsequently, *Roberts* [2008] provided additional chronological evidence that relates the ODP Site 884 record to the Iceland Basin excursion and here we perform a detailed analysis of uncertainties associated with its VGP path.

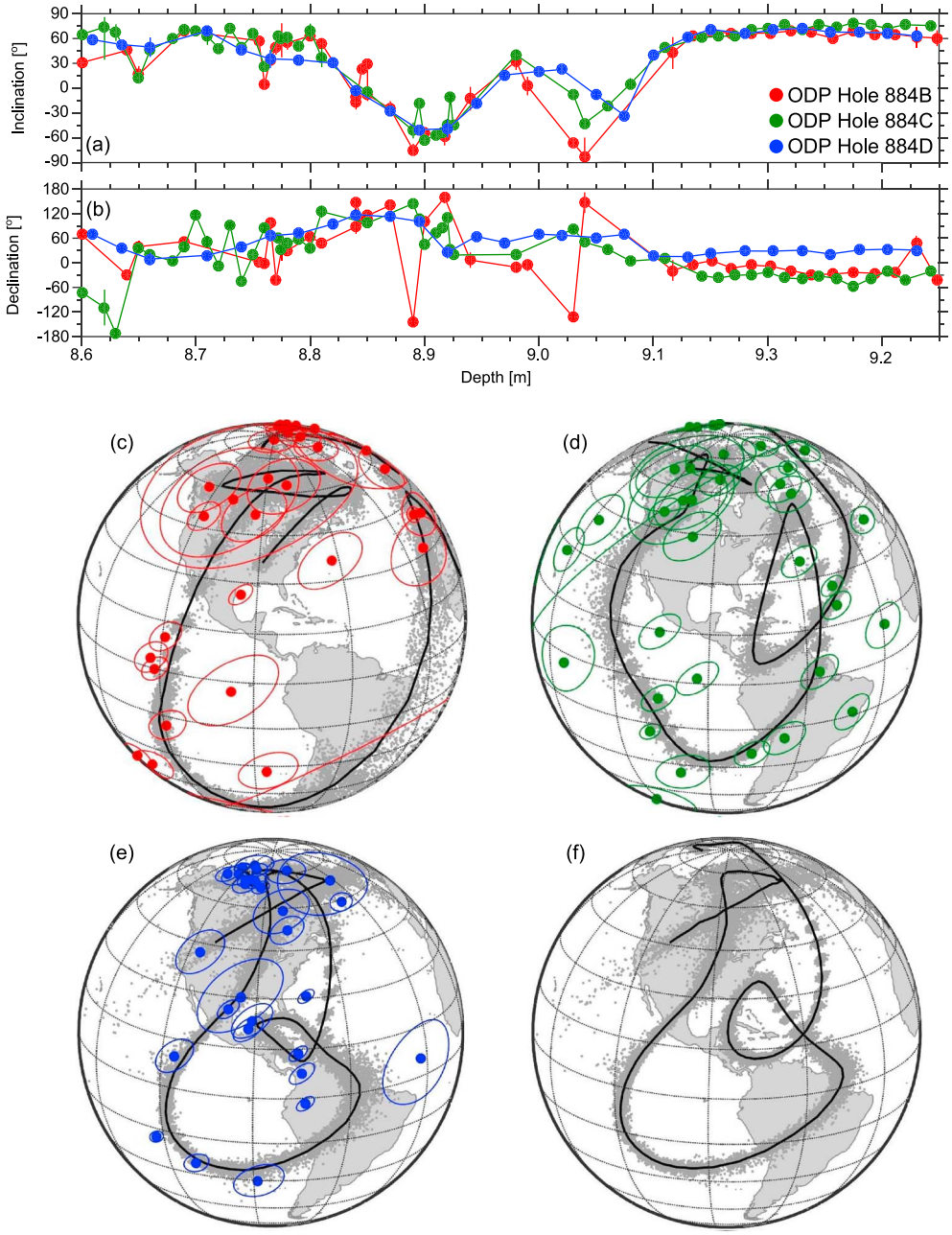

Figure 8. Paleomagnetic (a) inclination and (b) declination for the Iceland Basin excursion from three parallel sediment cores. Vertical bars indicate the marginalized uncertainties associated with each direction (in most cases the bars are narrower than the width of the symbol). Mean VGP paths (black line) based on weighted averages of specimen VGPs (solid colored symbols) and their uncertainties (colored ellipses) for (c) ODP Hole 884B, (d) ODP Hole 884C, and (e) ODP Hole 884D. Gray symbols represent weighted mean directions described in the text. (f) Composite VGP path and weighted mean directions obtained by pooling data from ODP Holes 884B, 884C, and 884D.

Roberts [2008] produced three parallel excursion records based on AF demagnetization of discrete specimens from ODP Holes 884B, 884C, and 884D with an additional record based on u-channel samples from Hole 884D. Using magnetic susceptibility records from these three holes, data for individual specimens were transferred onto a single depth scale based on ODP Hole 884D [*Roberts*, 2008]. The three discrete specimen records are shown in Figures 8a and 8b, with inclination and declination uncertainties calculated for stepwise AF demagnetization data using the procedure introduced in section 4. VGP paths for the individual holes

were calculated using a moving average with weights drawn from a Gaussian distribution. For a given depth, weights were drawn from a Gaussian pdf with a mean corresponding to the depth of interest and a standard deviation of 2 cm (selected as a balance between the specimen sampling interval and the smoothness of the resulting VGP path). A random unit vector was then drawn from $Q(\mathbf{W})$ for each specimen and was multiplied by its corresponding weight, with the resulting vectors then averaged according to the procedure of *Fisher and Hall* [1989] to provided a weighted mean direction. This process was repeated 300 times at each depth to provide a distribution of weighted mean directions that were subsequently transformed into VGPs (Figures 8c–8e).

The VGP records from ODP Holes 884C (Figure 8d) and 884D (Figure 8e) are characterized by lower directional uncertainties, and both records have a double clockwise loop structure. In contrast, specimens from ODP Hole 884B generally carry larger directional uncertainties and yield a single clockwise loop (Figure 8c). The absence of an inner initial loop in the ODP Hole 884B record may be a sampling artifact rather than a genuine paleomagnetic feature and could result from an apparent gap in the record between ~9.05 and 9.15 m (Figures 8a and 8b). At face value, the three records of the Iceland Basin excursion appear similar, but it is useful to constrain statistically the limits of palaeomagnetic interpretations. Few such multiple records of the same excursion exist in the literature, and it is important to assess the extent to which inconsistencies among the records limit overall interpretation of excursional field behavior.

In many instances uncertainties associated with individual directions in the respective holes from ODP Site 884 are smaller than the directional differences among the holes. Such inconsistencies cannot be attributed to measurement noise alone. Instead, it is apparent that some of the disagreement among the holes must be attributed to complexities associated with paleomagnetic signal acquisition in sediments [e.g., *Heslop et al.*, 2014] or to factors associated with coring such as local field deviations associated with the bottom hole assembly of the ODP coring system. Given the slight depth offsets between specimens from different holes, it cannot be assumed that given specimens experienced exactly the same field; therefore, it is not appropriate to employ *Fisher* [1953] statistics to estimate a site mean VGP path. Instead, a composite VGP path was produced using the same weighted moving average procedure, with specimens from the three holes pooled together (Figure 8f). The double clockwise VGP loop structure persists within this combined record, with ODP Hole 884B playing little role in defining the initial inner loop because of the apparent sampling gap between ~9.05 and 9.15 m. The VGP path passes over the North Atlantic Ocean, South America, the Pacific Ocean, and North America, with VGP uncertainties typically <5° throughout the excursion, which indicates that the ODP Site 884 record is inconsistent with the compilation of *Laj et al.* [2006]. If the records of *Roberts* [2008] correspond to the Iceland Basin excursion, the excursional field geometry must have been more complex than the dipole-dominated structure hypothesized by *Laj et al.* [2006].

6. Conclusions

Paleomagnetic directions estimated from demagnetization experiments are inherently uncertain. This has implications not only for constraining past geomagnetic field behavior but also for geological reconstructions based on paleomagnetic directions. Thus, it is essential that paleomagnetic reconstructions include fully quantified uncertainties to ensure robust interpretation and to protect against spurious inferences. VBPCA provides a means to readily estimate directional uncertainty, even for the small data sets that are typical of most demagnetization experiments. This uncertainty takes the form of a Gaussian off-axis angular distribution, which allows confidence regions to be defined for single specimens and numerical propagation of uncertainty to subsequent analysis of paleomagnetic directions, such as when determining site mean directions. In turn, these uncertainties can be readily transformed to provide confidence regions for representations of VGPs or paleomagnetic poles and apparent polar wander paths.

Appendix A

Given an observed data set, D, composed of N vectors, $\mathbf{t}(1), \ldots, \mathbf{t}(N)$, the log-likelihood of the PPCA model parameters is given by [*Tipping and Bishop*, 1999]

$$\mathcal{L}(\mu, \mathbf{W}, \sigma^2) = -\frac{N}{2}\left\{3\ln(2\pi) + \ln|\mathbf{C}| + \mathrm{tr}(\mathbf{C}^{-1}\mathbf{S})\right\}, \tag{A1}$$

where

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{t}(i)-\mu)(\mathbf{t}(i)-\mu)^T, \tag{A2}$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}, \tag{A3}$$

and μ, $\mathbf{W}$, and σ^2 follow the definitions given in section 3.

To estimate the best fit line through demagnetization data and, importantly, the uncertainty associated with the line, it is necessary to estimate distributions for μ, $\mathbf{W}$, and σ^2. This can be achieved with a Bayesian approach based on defining a prior distribution $P(\mu, \mathbf{W}, \sigma^2)$ for the model parameters, which when multiplied by the likelihood function (equation (A1)) for the data set, D, and normalized, yields the posterior distribution $P(\mu, \mathbf{W}, \sigma^2|D)$. This provides the predictive density [*Bishop*, 1999b]:

$$P(\mathbf{t}|D) = \int\int\int P(\mathbf{t}|\mu, \mathbf{W}, \sigma^2)P(\mu, \mathbf{W}, \sigma^2|D)\mathrm{d}\mu d\mathbf{W}\mathrm{d}\sigma^2. \tag{A4}$$

The variational Bayes approach to PCA developed by *Bishop* [1999a] introduces a distribution, Q, to approximate the true posterior distribution. Furthermore, it is assumed that Q can be factorized over its component variables. This factorization means that for the described Bayesian PCA model [*Bishop*, 1999a],

$$Q(\mathbf{X}, \mathbf{W}, \mu, \tau) = Q(\mathbf{X})Q(\mathbf{W})Q(\mu)Q(\tau), \tag{A5}$$

where $\mathbf{X}$ corresponds to principal component scores and $\tau = \sigma^{-2}$. This approach is computationally efficient and provides an approximation to equation (A4) for between 1 and $d-1$ principal components, where d is the dimension of the data set. Full justification of this approach is provided by *Bishop* [1999a], who outlines an iterative procedure to estimate the factors in the variational model (Appendix B). The parameters μ and $\mathbf{W}$, which are required to represent a paleomagnetic vector, take the form

$$Q(\mu) = \mathcal{N}(\mu \mid \mathbf{m}_\mu, \mathbf{\Sigma}_\mu) \quad \text{and} \tag{A6}$$

$$Q(\mathbf{W}) = \mathcal{N}(\mathbf{W} \mid \mathbf{m}_\mathbf{W}, \mathbf{\Sigma}_\mathbf{W}), \tag{A7}$$

where $\mathbf{m}$ and $\mathbf{\Sigma}$ correspond to a mean and covariance matrix, respectively, and $\mathbf{m}_\mu$, $\mathbf{\Sigma}_\mu$, $\mathbf{m}_\mathbf{W}$, and $\mathbf{\Sigma}_\mathbf{W}$ are estimated by the variational algorithm (Appendix B).

Appendix B

The variational Bayesian PCA approach of *Bishop* [1999a] is designed to fit between 1 and $d-1$ principal components, where d is the dimensionality of the data set. When q is the maximum number of components to be considered, the model includes an additional parameter, $\alpha = \alpha_1, \ldots, \alpha_q$, with each value of α controlling the inverse variance of the corresponding component of $\mathbf{W}$. Each term in α is represented by a Gamma probability density function (pdf). Large values in α_i will reduce the influence of the ith component; thus, α helps to resolve the question of how many components are needed to describe a data set [*Bishop*, 1999a].

When analyzing of demagnetization data, it is assumed that a segment selected via interactive point selection within vector demagnetization diagrams will be linear and only the leading principal component carries meaningful directional information. Thus, if a demagnetization segment is considered to be linear, it is expected that $\alpha_1 \ll \alpha_2 \simeq \alpha_3$. For simplicity, it could be assumed that $q = 1$ and the model will consist of only a single principal component and a noise contribution. As discussed in section 3, however, future studies may focus on objective identification of linear segments within demagnetization data and the use of α could play a role in determining if a given demagnetization interval is univectorial. Thus, α is included in the outlined approach, which is appropriate for $q = [1, d-1]$ components.

The Bayesian approach requires definition of prior distributions for μ, α, and σ^2. Little is known about these parameters; therefore, they are assigned broad priors:

$$P(\mu) = \mathcal{N}(\mu|\mathbf{0}, \beta^{-1}\mathbf{I}), \tag{B1}$$

$$P(\alpha) = \prod_{i=1}^{q}\Gamma(\alpha_i|a_\alpha, b_\alpha), \tag{B2}$$

and

$$P(\tau) = \Gamma(\tau|a_\tau, b_\tau), \tag{B3}$$

where $\tau = \sigma^{-2}$, $\beta = 10^{-3}$, $a_\alpha = b_\alpha = a_\tau = b_\tau = 10^{-3}$, $\mathbf{I}$ is an identity matrix, and the Gamma distribution $\Gamma(x|a, b)$ is given by

$$\Gamma(x|a, b) = \frac{b^a x^{a-1} \exp(-bx)}{\Gamma(a)}. \tag{B4}$$

As discussed in Appendix A, the variational approach assumes that the true joint posterior distribution can be approximated using a factorized form:

$$Q(\mathbf{X}, \mathbf{W}, \alpha, \mu, \tau) = Q(\mathbf{X})Q(\mathbf{W})Q(\alpha)Q(\mu)Q(\tau), \tag{B5}$$

where the individual components take the form

$$Q(X) = \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{x}_n|\mathbf{m}_\mathbf{x}^{(n)}, \mathbf{\Sigma}_\mathbf{x}\right), \tag{B6}$$

$$Q(\mu) = \mathcal{N}\left(\mu|\mathbf{m}_\mu, \mathbf{\Sigma}_\mathbf{w}\right), \tag{B7}$$

$$Q(\mathbf{W}) = \prod_{k=1}^{d} \mathcal{N}\left(\tilde{\mathbf{w}}_k|\mathbf{m}_\mathbf{w}^{(k)}, \mathbf{\Sigma}_\mathbf{w}\right), \tag{B8}$$

$$Q(\alpha) = \prod_{i=1}^{q} \Gamma\left(\alpha_i|\tilde{a}_\alpha, \tilde{b}_{\alpha i}\right), \tag{B9}$$

and

$$Q(\tau) = \Gamma\left(\tau|\tilde{a}_\tau, \tilde{b}_\tau\right). \tag{B10}$$

Distribution parameters are initialized based on maximum likelihood estimates obtained by probabilistic PCA [*Tipping and Bishop*, 1999]. An iterative cycle is then performed, with the distributions $Q(\mathbf{X})$, $Q(\mathbf{W})$, $Q(\alpha)$, $Q(\mu)$, and $Q(\tau)$ being updated based on the following expressions:

$$\mathbf{m}_\mathbf{x}^{(n)} = \langle\tau\rangle\mathbf{\Sigma}_\mathbf{x}\langle\mathbf{W}^\mathsf{T}\rangle(\mathbf{t}_n - \langle\mu\rangle), \tag{B11}$$

$$\mathbf{\Sigma}_\mathbf{x} = (\mathbf{I} + \langle\tau\rangle\langle\mathbf{W}^\mathsf{T}\mathbf{W}\rangle)^{-1}, \tag{B12}$$

$$\mathbf{m}_\mu = \langle\tau\rangle\mathbf{\Sigma}_\mu \sum_{n=1}^{N}(\mathbf{t}_n - \langle\mathbf{W}\rangle\langle\mathbf{x}_n\rangle), \tag{B13}$$

$$\mathbf{\Sigma}_\mu = (\beta + N\langle\tau\rangle)^{-1}\mathbf{I}, \tag{B14}$$

$$\mathbf{m}_\mathbf{w}^{(k)} = \langle\tau\rangle\mathbf{\Sigma}_\mathbf{w} \sum_{n=1}^{N}\langle\mathbf{x}_n\rangle(t_{nk} - \langle\mu_k\rangle), \tag{B15}$$

$$\mathbf{\Sigma}_\mathbf{w} = \left(\mathrm{diag}\langle\alpha\rangle + \langle\tau\rangle \sum_{n=1}^{N}\langle\mathbf{x}_n\mathbf{x}_n^\mathsf{T}\rangle\right)^{-1}, \tag{B16}$$

$$\tilde{a}_\alpha = a_\alpha + \frac{d}{2}, \tag{B17}$$

$$\tilde{b}_{\alpha i} = b_\alpha + \frac{\langle\|\mathbf{w}_i\|^2\rangle}{2}, \tag{B18}$$

$$\tilde{a}_\tau = a_\tau + \frac{Nd}{2}, \tag{B19}$$

and

$$\tilde{b}_\tau = b_\tau + \frac{1}{2}\sum_{n=1}^{N}\left\{\|\mathbf{t}_n\|^2 + \langle\|\mu\|^2\rangle + \mathrm{Tr}(\langle\mathbf{W}^\mathsf{T}\mathbf{W}\rangle\langle\mathbf{x}_n\mathbf{x}_n^\mathsf{T}\rangle) + 2\langle\mu^\mathsf{T}\rangle\langle\mathbf{W}\rangle\langle\mathbf{x}_n\rangle - 2\mathbf{t}_n^\mathsf{T}\langle\mathbf{W}\rangle\langle\mathbf{x}_n\rangle - 2\mathbf{t}_n^\mathsf{T}\langle\mu\rangle\right\}. \tag{B20}$$

Note that the update rules incorporate the terms that define the prior distributions (β, a_α, b_α, a_τ, and b_τ) into the appropriate Q distributions (the priors are fixed; therefore, these terms do not change as part of the updates). Expected values are given by $\langle\cdot\rangle$, where for a d-dimensional Gaussian distribution $\mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma})$, the expectations $\langle\mathbf{x}\rangle$ and $\langle\mathbf{x}\mathbf{x}^\mathsf{T}\rangle$ are given by *Šmídl and Quinn* [2006]

$$\langle\mathbf{x}\rangle = \mu \tag{B21}$$

and

$$\langle\mathbf{x}\mathbf{x}^\mathsf{T}\rangle = d\mathbf{\Sigma} + \mu\mu^\mathsf{T}. \tag{B22}$$

Acknowledgments
This work was supported by the Australian Research Council (grant DP120103952). We are grateful to Cor Langereis and two anonymous reviewers for providing critical comments that helped to improve the manuscript. Software to perform the variational Bayesian procedure is available from the corresponding author's web page.

References

As, J. A., and J. D. A. Zijderveld (1958), Magnetic cleaning of rocks in palaeomagnetic research, *Geophys. J. R. Astron. Soc.*, *1*, 308–319.
Bishop, C. M. (1999a), Variational principal components, in *Proceedings of the Ninth International Conference on Artificial Neural Networks, ICANN99*, pp. 509–514, IEEE.
Bishop, C. M. (1999b), *Advances in Neural Information Processing Systems*, MIT Press, vol. 11, chap. Bayesian PCA, pp. 382–388, San Francisco, Calif.
Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, 738 pp., Springer, New York.
Collinson, D. W., K. M. Creer, and S. K. Runcorn (1967), *Methods in Palaeomagnetism*, 609 pp., Elsevier, Amsterdam.
Dempster, A. P. (1969), *Elements of Continuous Multivariate Analysis*, 388 pp., Addison Wesley Longman, Reading, Mass.
de Sa, J. P. M. (2007), *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*, 505 pp., Springer, Berlin.
Dunlop, D. J. (1979), On the use of Zijderveld vector diagrams in multicomponent paleomagnetic studies, *Phys. Earth Planet. Inter.*, *20*, 12–24, doi:10.1016/0031-9201(79)90103-1.
Fisher, N. I., and P. Hall (1989), Bootstrap confidence regions for directional data, *J. Am. Stat. Assoc.*, *84*, 996–1002, doi:10.1080/01621459.1989.10478864.
Fisher, R. A. (1953), Dispersion on a sphere, *Proc. R. Soc. A*, *217*, 295–305, doi:10.1098/rspa.1953.0064.
Garcia-Portugués, E. (2013), Exact risk improvement of bandwidth selectors for kernel density estimation with directional data, *Electron. J. Stat.*, *7*, 1655–1685, doi:10.1214/13-EJS821.
Guadagnoli, E., and W. F. Velicer (1988), Relation of sample size to the stability of component patterns, *Psychol. Bull.*, *103*, 265–275, doi:10.1037/0033-2909.103.2.265.
Hall, P., G. S. Watson, and J. Cabrera (1987), Kernel density estimation with spherical data, *Biometrika*, *74*, 751–762, doi:10.1093/biomet/74.4.751.
Heslop, D., A. P. Roberts, L. Chang, M. Davies, A. Abrajevitch, and P. De Deckker (2013), Quantifying magnetite magnetofossil contributions to sedimentary magnetizations, *Earth Planet. Sci. Lett.*, *382*, 58–65, doi:10.1016/j.epsl.2013.09.011.
Heslop, D., A. P. Roberts, and R. Hawkins (2014), A statistical simulation of magnetic particle alignment in sediments, *Geophys. J. Int.*, *197*, 828–837, doi:10.1093/gji/ggu038.
Irving, E. (1964), *Palaeomagnetism and Its Application to Geological and Geophysical Problems*, 399 pp., John Wiley, New York.
Johnson, C. L., J. R. Wijbrans, C. G. Constable, J. Gee, H. Staudigel, L. Tauxe, V.-H. Forjaz, and M. Salgueiro (1998), $^{40}Ar/^{39}Ar$ ages and paleomagnetism of São Miguel lavas, Azores, *Earth Planet. Sci. Lett.*, *160*, 637–649, doi:10.1016/S0012-821X(98)00117-4.
Kent, J. T., J. C. Briden, and K. V. Mardia (1983), Linear and planar structure in ordered multivariate data as applied to progressive demagnetization of palaeomagnetic remanence, *Geophys. J. R. Astron. Soc.*, *75*, 593–621, doi:10.1111/j.1365-246X.1983.tb05001.x.
Khokhlov, A., and G. Hulot (2016), Principal component analysis of palaeomagnetic directions: Converting a Maximum Angluar Deviation (MAD) into an α_{95} angle, *Geophys. J. Int.*, *204*, 274–291, doi:10.1093/gji/ggv451.
Kirschvink, J. L. (1980), The least-squares line and plane and the analysis of palaeomagnetic data, *Geophys. J. R. Astron. Soc.*, *62*, 699–718, doi:10.1111/j.1365-246X.1980.tb02601.x.
Laj, C., C. Kissel, and A. P. Roberts (2006), Geomagnetic field behavior during the Iceland Basin and Laschamp geomagnetic excursions: A simple transitional field geometry?, *Geochem. Geophys. Geosyst.*, *7*, Q03004, doi:10.1029/2005GC001122.
Leonhardt, R., C. Heunemann, and D. Krása (2004), Analyzing absolute paleointensity determinations: Acceptance criteria and the software ThellierTool4.0, *Geochem. Geophys. Geosyst.*, *5*, Q12016, doi:10.1029/2004GC000807.
Love, J. J., and C. G. Constable (2003), Gaussian statistics for palaeomagnetic vectors, *Geophys. J. Int.*, *152*, 515–565, doi:10.1046/j.1365-246X.2003.01858.x.
Mahalanobis, P. C. (1936), On the generalized distance in statistics, *Proc. Natl. Inst. Sci.*, *2*, 49–55.
McElhinny, M. W. (1973), *Palaeomagnetism and Plate Tectonics*, 359 pp., Cambridge Univ. Press, Cambridge.
McFadden, P. L., and P. W. Schmidt (1986), The accumulation of palaeomagnetic results from multicomponent analysis, *Geophys. J. R. Astron. Soc.*, *86*, 965–979, doi:10.1111/j.1365-246X.1986.tb00670.x.
Penrose, R. (1955), A generalized inverse for matrices, *Proc. Cambridge Philos. Soc.*, *51*, 406–413, doi:10.1017/S0305004100030401.
Roberts, A. P. (2008), Geomagnetic excursions: Knowns and unknowns, *Geophys. Res. Lett.*, *35*, L17307, doi:10.1029/2008GL034719.
Roberts, A. P., B. Lehman, R. J. Weeks, K. L. Verosub, and C. Laj (1997), Relative paleointensity of the geomagnetic field over the last 200,000 years from ODP Sites 883 and 884, North Pacific Ocean, *Earth Planet. Sci. Lett.*, *152*, 11–23, doi:10.1016/S0012-821X(97)00132-5.
Rousseeuw, P. J., and B. C. van Zomeren (1990), Unmasking multivariate outliers and leverage points, *J. Am. Stat. Assoc.*, *85*, 633–651, doi:10.1080/01621459.1990.10474920.
Schmidt, P. W. (1982), Linearity spectrum analysis of multi-component magnetizations and its application to some igneous rocks from south-eastern Australia, *Geophys. J. R. Astron. Soc.*, *70*, 647–665, doi:10.1111/j.1365-246X.1982.tb05978.x.
Šmídl, V., and A. Quinn (2006), *The Variational Bayes Method in Signal Processing*, 228 pp., Springer, Berlin.
Tarling, D. H. (1983), *Palaeomagnetism: Principles and Applications in Geology, Geophysics and Archaeology*, 380 pp., Chapman and Hall, London.
Tauxe, L. (1998), *Paleomagnetic Principles and Practice*, 301 pp., Kluwer Acad., Netherlands.

Tipping, M. E., and C. M. Bishop (1999), Probabilistic principal component analysis, *J. R. Stat. Soc. B*, *21*, 611–622, doi:10.1111/1467-9868.00196.

Weeks, R., C. Laj, L. Endignoux, M. Fuller, A. P. Roberts, R. Manganne, E. Blanchard, and W. Goree (1993), Improvements in long-core measurement techniques: Applications in palaeomagnetism and palaeoceanography, *Geophys. J. Int.*, *114*, 651–662, doi:10.1111/j.1365-246X.1993.tb06994.x.

Zijderveld, J. D. A. (1967), A.C. demagnetization of rocks: Analysis of results, in *Methods in Palaeomagnetism*, pp. 254–286, Elsevier, Amsterdam.