Statistical Science
2000, Vol. 15, No. 4, 342–356

Regression Techniques in Plate Tectonics

Ted Chang, Daijin Ko, Jean-Yves Royer and Jiandong Lu

Abstract. We discuss a linearized model to analyze the errors in the reconstruction of the relative motion of two tectonic plates using marine magnetic anomaly data. More complicated geometries, consisting of several plates, can be analyzed by breaking the geometry into its stochastically independent parts and repeatedly applying a few simple algorithms to recombine these parts. A regression version of Welch's solution to the Behrens–Fisher problem is needed in the recombination process.

The methodology is illustrated using data from the Indian Ocean. Through a historical perspective we show how improving data density and improving statistical techniques have led to more sophisticated models for the Indo-Australian plate.

We propose an influence-based regression diagnostic for tectonic data. A generalization of the standardized influence matrix of Lu, Ko and Chang is applied to study the influence of a group of data points on a subparameter of interest. This methodology could also be used in treatment-block designs to analyze the influence of the blocks on the estimated treatment effects.

Key words and phrases: Nonlinear regression, spherical regression, plate tectonics, regression diagnostics, influence function.

1. INTRODUCTION

Many of Nature's most spectacular geological phenomena, such as earthquakes, volcanoes and mountain formation, result from interactions between rigid tectonic plates in motion relative to one another at the Earth's surface. An understanding of the past positions of the tectonic plates is vital to any understanding of the Earth's history. In this paper we will discuss the types of data used to reconstruct the past positions of tectonic plates and show how many questions of geological interest can be answered by using modifications of tools previously developed for linear regression.

Ted Chang is Professor, Department of Statistics, University of Virginia, Halsey Hall, P.O. Box 400135, Charlottesville, Virginia 29904-4135. Daijin Ko is Professor, Department of Biostatistics, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0032. Jean-Yves Royer is Doctor, Institut Universitaire Europen de la Mer, Centre National de la Recherche Scientifique, Universit de Bretagne Occidentale, Place Nicolas Copernic, 29280 Plouzane, France. Jiandong Lu is Doctor, Merck Company, Inc., P.O. Box 4, WP39-169, West Point, Pennsylvania 19486.

When two tectonic plates diverge at a midoceanic ridge, molten crustal material is extruded from the ridge and carried away from the ridge on both plates. A useful mental picture is to imagine the plates as conveyer belts moving away from the ridge. The molten crustal material solidifies immediately after extrusion. It follows that if the material on both plates that was extruded at a specific time in the past (e.g., ten million years before the present) could be identified, we would have two congruent curves, one on each plate. The shape of the curves would be identical to the shape of the ridge at the time of their extrusion (see Figure 1), which is not necessarily the shape of the present-day ridge.

The molten crust has a high iron content and, as it solidifies, acquires a magnetization based upon the prevailing Earth's magnetic field at the time of the crust's extrusion. At known times in the past, the Earth's magnetic field reversed itself (the North magnetic pole flipped to point southward). Hence, the ocean floor is made of stripes of material with alternating reversed and normal polarity magnetization. This pattern of stripes will be symmetric relative to the ridge where the material emplaced. A research vessel towing a magnetometer over these stripes will record a series of positive and negative

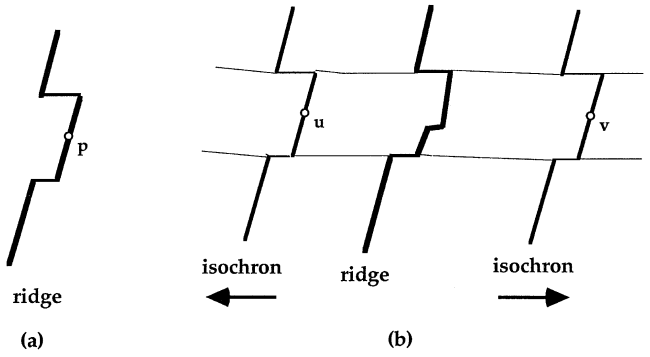

FIG. 1. (a) *Shape of mid-ocean ridge at past time when the Earth's magnetic field was reversed;* (b) *Present-day shape of ridge. Arrows represent direction of motion of the plates relative to the present-day ridge axis. Using magnetic means one can detect crustal material which was extruded at the time of* (a). *This results in two congruent* isochrons, *one on each side of the current ridge. Crustal material extruded from point p is now located at u and v. Since the isochrons are congruent, a theorem of Euler asserts the existence of a three-dimensional rotation matrix A which takes the left side lineation into the right side lineation; that is,* $v = Au$.

Fracture zones are shown as horizontal lighter lines. The roughly vertical segments are the magnetic anomaly lineations. *The isochrons are modeled as piecewise great circle segments with offsets at the fracture zones.*

magnetic anomalies superimposed on the present-day Earth's magnetic field. Thus from such magnetic anomaly profiles, one can map and date paired *magnetic anomaly lineations*, one set on each side of the ridge (see Figure 2).

For example, if we can identify the anomaly 5 lineation (ten million years before the present), on both sides of the ridge we would have two congruent curves. A theorem due to Euler states that there is a 3×3 matrix A which takes one curve into the other. In other words, if u is a point on one lineation, v a point on the opposing lineation so that u and v extruded from the same point on the ridge, then $v = Au$. A must satisfy the identities $AA^T = I$ and $\det(A) = 1$. The collection of these matrices, standardly denoted by $\mathcal{S}O(3)$, form a three-dimensional submanifold of the vector space of 3×3 matrices (which we can identify with Euclidean nine-dimensional space R^9). Physically, any such A represents a rotation of three-dimensional space.

This A is the parameter of interest and is said to be the *reconstruction of the u-plate to the v-plate* (in a coordinate system fixed in the v-plate). We note that both plates move. It follows that A is a relative reconstruction, not a reconstruction with respect to some absolute coordinate system. The issue of moving coordinate systems is quite confusing and is discussed in detail in the Appendix.

The magnetic anomaly lineations appear to be piecewise straight with offsets at the so called *fracture zones*. Fracture zones appear like cliffs on the ocean floor. We will use the term *isochron* for the combined curve consisting of a magnetic anomaly lineation together with its offsetting fracture zones since each isochron consists of points of the same age.

The data (see Figure 3) consists of identified points on the isochrons. Magnetic anomaly profiles

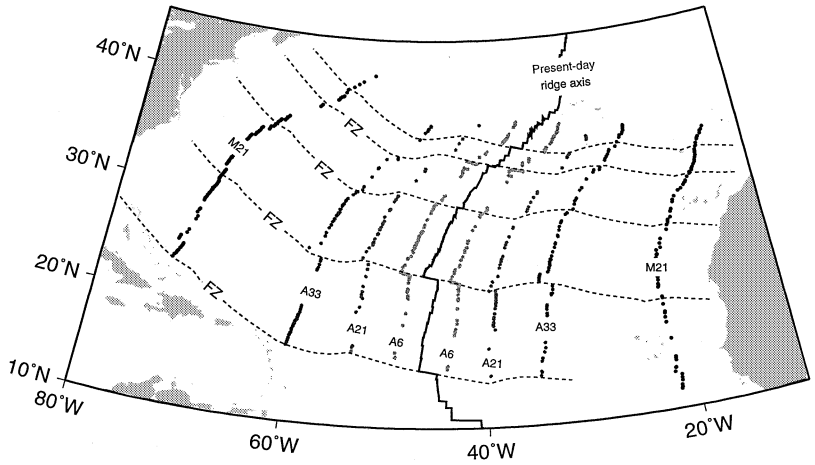

FIG. 2. *Selected Central Atlantic fracture zones and magnetic anomaly lineations: Present location mid-Atlantic ridge–solid line trending North to South. Fracture zones (FZ)—dashed lines trending East to West (fracture zones are, from North to South, Oceanographer, Hayes, Atlantis, Southern Kane,* 15d20). *Identified locations on magnetic anomaly lineations–small circles* [*magnetic anomaly lineations are anomaly A*6 (*lightest,* 20 *million years*), *A*21 (50 *million years*), *A*33 (80 *million years*), *M*21 (*darkest,* 150 *million years*)]. *Data courtesy Kim Klitgord, U.S. Geological Survey, Woods Hole, MA. GMT mapping software courtesy of Wessel and Smith* (1995).

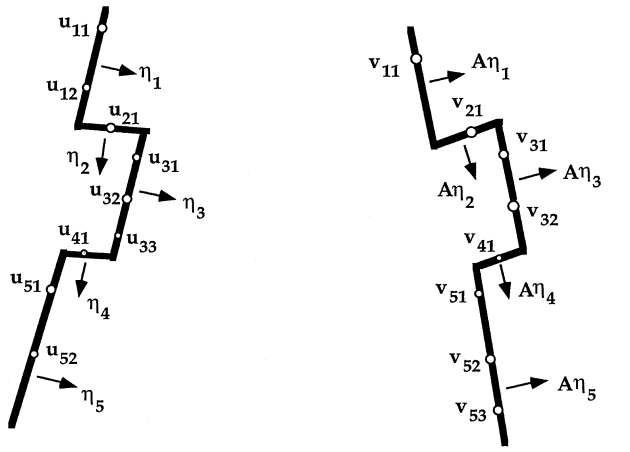

FIG. 3. *Stylized diagram of data on two opposing isochrons. Identified points on the isochrons are labelled u_{ij} and v_{ik}. Points with $i = 1, 3, 5$ correspond to identified locations of points on segments of magnetic anomaly lineations, points with $i = 2, 4$ to identified locations of points on fracture zone segments. Each segment is modeled as a great circle: the normal vectors to these great circles are labelled η_i on one isochron and $A\eta_i$ on the opposing isochron. The η_i are unknown and represent nuisance parameters. A is an unknown three-dimensional rotation matrix, and is the parameter of interest.*

from different parts of the world are remarkably similar and by picking a characteristic point on the profile, one can determine where the ship crossed a particular magnetic anomaly lineation. Points on the fracture zones are determined by measurement of the depth of the ocean floor. *Note that it is highly unlikely that the trajectory of a research vessel will cross matching points on the opposing isochrons and hence the identified points are not homologous; that is, the data does not come in pairs (u, v) with $v = Au$.*

In an earlier survey paper, Chang (1993) discussed an idealized model of geophysical data and the insights it yields on how the shape of the plate boundaries affects the statistical properties of an estimated reconstruction. The present paper will not revisit these questions and will concentrate on more data analytic problems.

A second paper, Kirkwood, Royer, Chang and Gordon (1999), aimed at the geophysics audience, outlines the use of the techniques discussed here in Sections 2 to 5. The present paper, in addition to discussing more recent developments, takes a historical perspective. Through a case study of the Indian Ocean region, we will show how improving geophysical data, combined with improving statistical techniques, have resulted in more sophisticated and detailed plate motion models.

2. LINEARIZED STATISTICAL ANALYSIS

Hellinger (1981) proposed to model the isochron as a union of great circle segments. If η is the normal to a great circle segment on one isochron, $A\eta$ is the normal to its mate on the opposing isochron. Thus if there are s segments, we have a total of $3 + 2s$ parameters: 3 parameters for the rotation A and $2s$ nuisance parameters for the section normals $\eta_1, \ldots, \eta_s$. Let u_{ij} represent identified points on the ith section of one isochron and let v_{ik} represent identified points on the ith section of the opposing isochron. Hellinger proposed to estimate A and the η_i by minimizing the objective function

$$(1) \qquad r(A, \eta) = \sum_i \left[\sum_j \frac{(u_{ij}^T \eta_i)^2}{\sigma_{ij}^2} + \sum_k \frac{(v_{ik}^T A \eta_i)^2}{\tilde{\sigma}_{ik}^2} \right].$$

Here σ_{ij} and $\tilde{\sigma}_{ik}$ are known error constants which must be supplied by the scientist. They are related to the indeterminancy of the magnetic anomaly profile and the navigation system in use at the time the data was collected.

In particular, the magnetic data collected consists of a squiggle which represents the strength of the magnetic field along the trajectory of the vessel. After eliminating the dominant effects of the current Earth's magnetic field and correcting for latitude, the *remnant* magnetic trace has a remarkably similar appearance in the different oceans of the world. Each magnetic anomaly has a characteristic shape in the remnant magnetic trace and a skilled marine geophysicist is able to interpret the trace and locate the approximate point on the trace corresponding to the specific anomaly of interest. The level of uncertainty in locating the correct point on the trace is due primarily to the distinctiveness of the characteristic trace of the anomaly. Once the point is located on the trace, it has to be correlated with the trajectory of the ship to determine its geographic location on the Earth's surface. This process is time-consuming. We note that before the advent of satellite navigation, the position of the ship was determined from stellar navigation, which was especially unreliable during overcast weather.

The sources of error are discussed at length in Kirkwood et al. (1999). From our experience, the errors can reasonably be modeled as normally distributed. Kirkwood et al. (1999) and Shaw and Cande (1990) include standard exploratory data analysis plots to evaluate normality.

A is typically referred to as the reconstruction to the time of a specific magnetic reversal, that is, a "reconstruction to time A5," as opposed to a "reconstruction to 10 million years before present" and

hence any errors in the dating of the reversal are not germane.

The errors in plate tectonic data range typically from 2 to 20 km., which is miniscule compared to the circumference of the Earth (approximately 40,000 km.). Accordingly, to understand the asymptotic distribution of the estimator $\widehat{A}$ of A determined by minimizing (1), Chang (1988) applied a "local flat earth" approximation, as follows.

Let $F(\mu, \kappa_0)$ denote the Fisher–von Mises–Langevin distribution on Ω_3, the unit sphere in Euclidean three-dimensional space, whose density (with respect to surface measure on the sphere) is

$$f_{\mu,\kappa_0}(u) = c(\kappa)\exp(\kappa_0 u^T\mu), \qquad c(\kappa_0) = \frac{\kappa_0}{4\pi \sin h\kappa_0}. \tag{2}$$

Here u and μ are on the sphere. μ is called the *modal* vector and κ_0 the *concentration* parameter. Write $u = (u^T\mu)\mu + x$, where x is in $\mu^\perp$, the vectors perpendicular to μ. Then

$$\begin{aligned} f_{\mu,\kappa_0}(u) &= c(\kappa_0)\exp(\kappa_0\sqrt{1-|x|^2}) \\ &= c(\kappa_0)\exp(\kappa_0(1-\tfrac{1}{2}|x|^2)) + O(|x|^4) \end{aligned} \tag{3}$$

so that as $\kappa_0 \to \infty$, $\kappa_0^{1/2}x$ approaches a $N(0, I-\mu\mu^T)$ distribution. We see that, intuitively, $\kappa_0^{-1/2}$ plays the role of a scale parameter.

Large κ_0 approximations are essentially "local flat earth" approximations. For $\mu \in \Omega_3$, let $\phi_\mu : \mu^\perp \to \Omega_3$ be the map defined by

$$\phi_\mu(x) = \sqrt{1-|x|^2}\mu + x$$

for $x \in \mu^\perp$. We note that $\mu^\perp$ is the tangent space at μ to Ω_3 and that

$$\cos(|x|)\mu + \sin(|x|)\frac{x}{|x|} = \phi_\mu(x) + o(|x|^3). \tag{4}$$

The left-hand side of (4) is the equation for the (azimuthal equidistant) polar projection centered at $\mu \in \Omega_3$. It is the bane of cartographers that maps from $R^2 \to \Omega_3$ distort somewhere. If we want to map Antarctica, we should choose a polar projection centered at the South pole, not a Mercator projection. Similarly our choice of replacing u by $x \in \mu^\perp$ satisfying $u = \phi_\mu(x)$ is asymptotically, for large κ_0, nondistorting (near μ).

We assume, that for unknown κ, for $A \in \mathscr{SO}(3)$ and for $\eta_i,\ \alpha_{ij},\ \tilde{\alpha}_{ik} \in \Omega_3$,

$$\begin{aligned} u_{ij} &\sim F(\alpha_{ij}, \kappa\sigma_{ij}^{-2}), \qquad & v_{ik} &\sim F(\tilde{\alpha}_{ik}, \kappa\tilde{\sigma}_{ik}^{-2}), \\ \alpha_{ij}^T\eta_i &= 0, & \tilde{\alpha}_{ik}^T A\eta_i &= 0. \end{aligned} \tag{5}$$

(These assumptions are a bit too strong; see Chang, 1988 for slightly weaker assumptions.) Formally we take asymptotics as $\kappa \to \infty$. Note, however, that usually the error constants σ_{ij} and $\tilde{\sigma}_{ik}$ supplied by geophysicists are quite accurate so that $\kappa \approx 1$. An error in u_{ij} of 20 km corresponds to $\kappa\sigma_{ij}^{-2} \approx 2 \times 10^5$. We note that the $\kappa\sigma_{ij}^{-2}$ and $\kappa\tilde{\sigma}_{ik}^{-2}$ play the role of the concentration parameter κ_0 in (2). Thus large κ_0 asymptotics are justified. Therefore, we write $\kappa \to \infty$ to indicate mathematically that we are using an asymptotic approximation which is accurate when all the $\kappa\sigma_{ij}^{-2}$ and $\kappa\tilde{\sigma}_{ik}^{-2}$ are large.

Let $x_{ij} = u_{ij}^T\eta_i$ and $y_{ik} = v_{ik}^T A\eta_i$. Here x_{ij} and y_{ik} are scalar and $O_p(\kappa^{-1/2})$ (so that as $\kappa \to \infty$, the distribution of $\kappa^{1/2}x_{ij}$ approaches that of a nondegenerate random variable).

Represent the estimator $\widehat{\eta}_i$ of η_i obtained from (1) as $\widehat{\eta}_i = (\widehat{\eta}_i^T\eta_i)\eta_i + \xi_i$, where ξ_i is a vector constrained to the two-dimensional space $\eta_i^\perp$ and is zero when $\widehat{\eta}_i = \eta_i$. We note that $\widehat{\eta}_i$ is constrained to lie in the two-dimensional space Ω_3. Thus to represent the error in $\widehat{\eta}_i$ we need to choose a projection from a two-dimensional Euclidean space into Ω_3. The inverse image ξ_i of $\widehat{\eta}_i$ under this projection represents the error in $\widehat{\eta}_i$ as an estimate of η_i.

If h is a vector in R^3, define

$$M(h) = \begin{bmatrix} 0 & -h_3 & h_2 \\ h_3 & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{bmatrix},$$

$$\exp(M(h)) = I + M(h) + \frac{1}{2!}[M(h)]^2 + \cdots.$$

It is well known in Lie group theory that any element of $\mathscr{SO}(3)$ can be written in the form $\exp(M)$ for some skew-symmetric matrix M. Because $\mathscr{SO}(3)$ is a group under matrix multiplication and both A and $\widehat{A}$ are members of $\mathscr{SO}(3)$, there exists t in R^3 for which $\widehat{A} = A\exp(M(t))$. Heuristically, $\widehat{A}$ should be close to A, so that $A^T\widehat{A}$ is a small rotation (that is, represents a rotation with a small angle). Thus we express $\widehat{A} = A\exp(M(t))$ so that t represents the error of $\widehat{A}$ as an estimate of A. We note that because $\mathscr{SO}(3)$ is a group under matrix multiplication, it is more natural to use a multiplicative (rather than additive) description of error.

Then t and the ξ_i will be $O_p(\kappa^{-1/2})$. Then (1) becomes

$$\begin{aligned} r(\widehat{A}, \widehat{\eta}_i) = \sum_i \Bigg[&\sum_j \frac{(x_{ij} + \alpha_{ij}^T\xi_i)^2}{\sigma_{ij}^2} \\ &+ \sum_k \frac{(y_{ik} + \tilde{\alpha}_{ik}^T A\xi_i - \tilde{\alpha}_{ik}^T AM(\eta_i)t)^2}{\tilde{\sigma}_{ik}^2} \Bigg] \\ &+ O_p(\kappa^{-3/2}). \end{aligned} \tag{6}$$

This is the error sum of squares of the following (weighted) regression:

$$\begin{aligned}\sigma_{ij}^{-1}x_{ij} &= -\sigma_{ij}^{-1}\alpha_{ij}^T\xi_i + \varepsilon_{ij},\\ \tilde{\sigma}_{ik}^{-1}y_{ik} &= -\tilde{\sigma}_{ik}^{-1}\tilde{\alpha}_{ik}^T A\xi_i + \tilde{\sigma}_{ik}^{-1}\tilde{\alpha}_{ik}^T AM(\eta_i)t + \tilde{\varepsilon}_{ik},\end{aligned} \tag{7}$$

where ε_{ij} and $\tilde{\varepsilon}_{ik}$ are (univariate) $N(0, \kappa^{-1})$. In this regression, the response variables are x_{ij} and y_{ik} (which are unobservable since η_i and A are unknown), the "unknown" parameters are ξ, and t (whose "true" values are zero) and the design matrix is determined from α_{ij}, $\tilde{\alpha}_{ik}$, η_i, and A (all of which need to be estimated from the data). While this might seem mysterious, as discussed below it is fully consistent with how nonlinear regressions are linearized.

The general approach is as follows:

1. Calculate $\widehat{A}$ and $\widehat{\eta}_i$ to minimize $r(A, \eta)$ as defined in (1).
2. Estimate the design matrix of the approximating regression (7).
3. Estimate κ by $\widehat{\kappa}$, where $1/\widehat{\kappa} = \text{MSE} = r(\widehat{A}, \widehat{\eta})/(N-2s-3)$. N is the total number of data points, and s is the number of sections. We note that estimating t subtracts three degrees of freedom and estimating each ξ_i subtracts two degrees of freedom from the error sum of squares of (7).
4. Because A is unknown, we do not have an estimate of t. Nevertheless, we can estimate the covariance matrix of t from the estimated design matrix and the estimate $\widehat{\kappa}$. Thus a confidence region for A can be calculated as
$$\begin{aligned}\mathscr{C} = \{A : A = \widehat{A}\exp(M(t)),\\ \text{where } t^T\,\text{Cov}(t)t < 3F_\alpha(3, N-2s-3)\}.\end{aligned}$$
5. The usual extra error sum of squares tests for nested models apply, using $r(\widehat{A}, \widehat{\eta})$ as an error sum of squares.

Although our model (5) cannot be put into the form of the usual weighted nonlinear regression model,

$$y_i = f(x_i, \theta) + \varepsilon_i, \tag{8}$$

where, for known constants w_i, the $\sqrt{w_i}\varepsilon_i$ are i.i.d. with $E(\varepsilon_i) = 0$ and $\text{Var}(\sqrt{w_i}\varepsilon_i) = \sigma^2$, our basic approach, steps 1–5 above, nevertheless resembles what is done in nonlinear regression. That is, given a typical nonlinear weighted regression model (8), one estimates θ by minimizing $\text{SSE}(\theta) = \sum_{i=1}^n w_i(y_i - f(x_i, \theta))^2$. Because $\text{SSE}(\widehat{\theta}) \approx \sum_i w_i(y_i - f(x_i, \theta) - (\partial f/\partial\theta)(x_i, \theta)(\widehat{\theta}-\theta))^2$, one then considers the linearized regression model

$$y_i - f(x_i, \theta) = \left[\frac{\partial f}{\partial\theta}(x_i, \theta)\right](\widehat{\theta} - \theta) + \varepsilon_i. \tag{9}$$

Analysis of this linearized regression model now corresponds to our approach: the response variable $y_i - f(x_i, \theta)$ is unobservable, the design matrix whose rows are $(\partial f/\partial\theta)(x_i, \theta)$ is unknown and must be estimated using $(\partial f/\partial\theta)(x_i, \widehat{\theta})$, and the parameter $\widehat{\theta}-\theta$ cannot be estimated from the sample, although its covariance matrix is estimable from the approximating regression model (9) (see Table 1).

Usually, the nonlinearity of the regression is assessed by considering the quadratic terms in the expansion

$$\begin{aligned}f(x_i, \widehat{\theta}) \approx f(x_i, \theta) + \frac{\partial f}{\partial\theta}(x_i, \theta)(\widehat{\theta} - \theta)\\ + \frac{1}{2}\frac{\partial^2 f}{\partial\theta^2}(x_i, \theta)(\widehat{\theta} - \theta)^2.\end{aligned}$$

If the next term of the expansion (6) is calculated, it can be shown that, for the magnitudes of the x_{ij}, y_{ik}, ξ_i and t commonly encountered, these nonlinear terms can be safely ignored relative to the linear terms. We note that the nonlinearity in (1) arises only from the nonlinearity of the data and parameter spaces. In particular, the eigenvalues of $\text{Cov}(t)$ are small (in our experience 10^{-4} or less) so that the curvature of $\mathscr{S}O(3)$ does not create difficulties.

The unimportance of the "curvature" effects is related to the parameterizations $\widehat{\eta}_i = (\widehat{\eta}_i^T\eta_i)\eta_i + \xi_i$ and $\widehat{A} = A\exp(M(t))$ of Ω_3 and $\mathscr{S}O(3)$. These parameterizations are centered around the true values η_i and A. If instead $\widehat{\eta}_i$ were represented, for example, in terms of latitude and longitude, the curvature effects would not be ignorable whenever η_i gets too close to the North and South poles. Our choices of parameterization at η_i and at A greatly ameliorate the curvature effects and also result in substantially simpler formulas.

Essentially, as discussed above, using ξ_i to parameterize $\widehat{\eta}_i$ is "polar projection" centered at η_i and hence is asymptotically nondistorting in the region of interest: namely, close to η_i. Similarly, our parameterization of $\mathscr{S}O(3)$ is a "polar projection" centered at A and is asymptotically nondistorting in a neighborhood of A. The reader is referred to Chang, Stock and Molnar (1990) for a longer discussion of this point. In addition see Kirkwood et al. (1999) for an illustrative simulation of the curvature effects of parameterization in a geophysical context.

3. NEW GEOMETRIES FROM OLD: THE INDIAN OCEAN ACCORDING TO ROYER AND CHANG

The data discussed in the previous section occurs only when two plates diverge at a midoceanic ridge. At other plate junctions, the relative motion of the

TABLE 1
Analogs between nonlinear regression model (8) *and Hellinger plate tectonics model* (5)

	Nonlinear regression	**Hellinger model**
Parameter	θ	A: rotation η_i: section normal
Response variable	y_i	(none)
Data		u_{ij}, v_{ik}: observed locations
Independent variables	x_i	True positions (unobserved) α_{ij}: u-side, $\tilde{\alpha}_{ik}$: v-side
"Scaled" error variance	σ^2	κ^{-1}
Weights	w_i	σ_{ij}^{-2}, $\tilde{\sigma}_{ik}^{-2}$
	$SSE(\theta)$	$r(A, \eta)$
Parameter for linearized regression	$\hat{\theta} - \theta$	t: error in rotation ξ_i: error in section normal
Response for linearized regression	$y_i - f(x_i, \theta)$	Distance from sections $x_{ij} = u_{ij}^T \eta$, $y_{ik} = v_{ik}^T A \eta_i$
Design matrix for linearized regression	$\partial f / \partial \theta(x_i, \theta)$	$-\alpha_{ij}^T$ $-\tilde{\alpha}_{ik}^T A$, $\tilde{\alpha}_{ik}^T A M(\eta_i)$

two plates must be inferred by composing a chain of diverging plate boundaries. Thus Stock and Molnar (1983) determine the relative motion of Pacific to North America by composing a chain going around the world in the opposite direction: Pacific to Antarctica, Antarctica to Indo-Australia, Indo-Australia to Africa and Africa to North America.

To combine two rotations, we can use the following algorithm: Suppose $A^T \hat{A} = \exp(M(t_A))$, $B^T \hat{B} = \exp(M(t_B))$ and let $C = BA$ and $\hat{C} = \hat{B}\hat{A}$. Let

$$\begin{aligned}\exp(M(t_C)) = C^T \hat{C} &= A^T B^T \hat{B} A A^T \hat{A} \\ &= A^T \exp(M(t_B)) A \exp(M(t_A)) \\ &= \exp(M(A^T t_B)) \exp(M(t_A)) \\ &= \exp(M(A^T t_B + t_A)) + o(|t_B|, |t_A|).\end{aligned}$$

Therefore $t_C = A^T t_B + t_A + o(|t_B|, |t_A|)$, yielding the estimate

$$\widehat{\mathrm{Cov}}(t_C) = \hat{A}^T \, \widehat{\mathrm{Cov}}(t_B) \hat{A} + \widehat{\mathrm{Cov}}(t_A).$$

We note that t_A and t_B are independent because the data used to calculate $\hat{A}$ and $\hat{B}$, having come from different ridges, are independent.

Royer and Chang (1991) used these statistical techniques to calculate the relative motion of Australia to India. Although at that time the commonly accepted geometry involved a unitary Indo-Australian plate, previous authors had noted that the fit of the data would be improved by splitting the Indo-Australian plate. In addition, undulations in the ocean basement as well as seismic activity provided evidence of relative motion between India and Australia. Royer and Chang established that Australia and India move relative to each other by calculating a confidence region for their relative motion and finding that the identity is not included.

Assuming that India and Australia move separately, the reconstruction A of India to Somalia can be calculated across the Carlsberg Ridge and the reconstruction B of Somalia to Australia can be calculated across the Central Indian Ridge (see Figure 4). Thus, as above, the reconstruction BA of India to Australia can be estimated together with a statistical estimate of its error.

However, data along the Central Indian Ridge was, at that time, rather sparse. The Australian, Somalian, and Antarctic plates meet at a triple junction of three diverging plates with the Southwest Indian Ridge separating Somalia from Antarctica and the Southeast Indian Ridge separating Australia from Antarctica. Let C reconstruct Australia to Antarctica and D reconstruct Antarctica to Somalia. Then DCB reconstructs Somalia to itself and hence is the identity. The triple junction can be analyzed in a method similar to the previous section using a combined regression with 6 degrees of freedom for the rotation parameters, $2s$ degrees of freedom for the sections, and $N - 2s - 6$ degrees of freedom for the error, where N is the total number of data points along the three ridges and s is the total number of sections.

In this way, the data along the Southwest Indian Ridge (separating Somalia from Antarctica) and the Southeast Indian Ridge (separating Australia from Antarctica) is used to augment the data along the Central Indian Ridge to estimate reconstruction of Somalia to Australia.

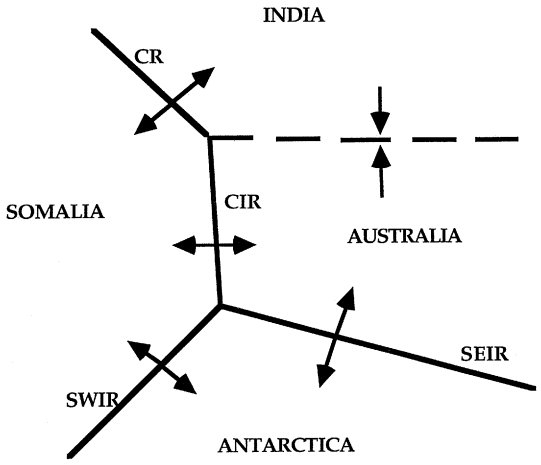

FIG. 4. *Stylized figure of Somalia/Antarctica/India-Australia geometry proposed by Royer and Chang* (1991). *A more realistic map of these boundaries is shown in Figure* 7. *Arrows indicate direction of relative motion of the plates. Spreading* (*diverging arrows*) *occurs and data exists along the Carlsburg Ridge* (*CR*), *Central Indian Ridge* (*CIR*), *Southwest Indian Ridge* (*SWIR*) *and Southeast Indian Ridge* (*SEIR*). *Let A, B, C, D represent the reconstructions of India to Somalia* (*using data from CR*), *Somalia to Australia* (*using data from CIR*), *Australia to Antarctica* (*using data from SEIR*), *and Antarctica to Somalia* (*using data from SWIR*), *respectively. The SOM/AUS/ANT triple junction closure condition assures DCB is the identity. Data does not exist along the AUS/IND boundary* (*broken line with converging arrows*); *however, the reconstruction of India to Australia is BA.*

Since data only exists at diverging plate boundaries, very complicated geometries can be analyzed in two steps:

1. Break up the geometry into its constituent stochastically independent parts (simple boundaries, triple junctions and two linked triple junctions; see Figure 5).
2. Repeatedly use three algorithms to combine rotations (see Figure 6).

These algorithms are:

1. Combine two independently estimated rotations, as described above.
2. Invert a rotation.
3. Combine two independent estimates of the same rotation.

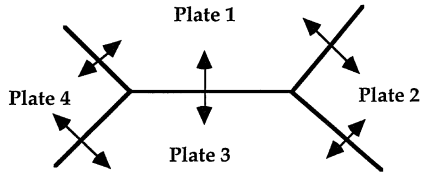

FIG. 5. *Two linked triple junctions. Spreading occurs on all five boundaries and hence all five boundaries have data.*

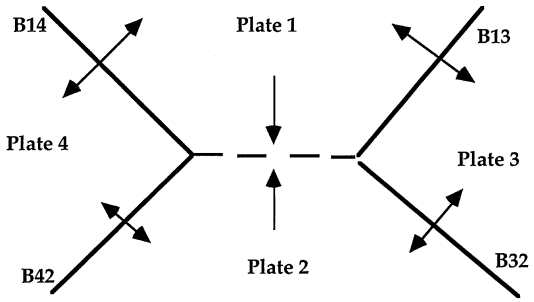

FIG. 6. *Hypothetical* 4 *plate geometry. Spreading occurs and data exists on plate boundaries shown with solid lines. Plates* 1 *and* 2 *are converging and no data exists on their boundary, shown with broken line. Let the* 3×3 *rotation matrices A, B, C, D reconstruct Plate* 1 *to Plate* 3 (*using data from boundary B*13), *Plate* 3 *to Plate* 2 (*data from B*32), *Plate* 1 *to Plate* 4 (*data from B*14), *and Plate* 4 *to Plate* 2 (*data from B*42), *repsectively. Then BA reconstructs Plate* 1 *to Plate* 2 *and is a combination of two independently estimated rotation matrices* (*A and B*). *Similarly DC reconstructs Plate* 1 *to Plate* 2 *and is also a combination of two independently estimated rotations. Because DC uses data from boundaries B*14 *and B*42 *and BA uses data from boundaries B*13 *and B*32, *DC and BA are independent. Thus DC and BA are independent estimates of the same reconstruction from Plate* 1 *to Plate* 2.

To invert a rotation, we note that if $\widehat{A} = A\exp(M(t))$, then $\widehat{A}^{-1} = \exp(M(-t))A^{-1} = A^{-1}\exp(M(-At))$. Writing $\widehat{A}^{-1} = A^{-1}\exp(M(t'))$, we have $t' = -At$, so

$$\widehat{\mathrm{Cov}}(t') = \widehat{A}\,\widehat{\mathrm{Cov}}(t)\widehat{A}^T.$$

To combine two or more independent estimates $\widehat{A}_i$ of the same rotation A, let $\widehat{A}_i = A\exp(M(t_i))$ and let $P_i = \mathrm{Cov}(t_i)^{-1}$. Let

$$t = (P_1 + \cdots + P_k)^{-1}(P_1t_1 + \cdots + P_kt_k)$$

be the standard (multivariate) weighted average of $t_1, \ldots, t_k$. Then

$$\mathrm{Cov}(t) = (P_1 + \cdots + P_k)^{-1}$$

and $\widehat{A} = A\ \exp(M(t))$ is the desired combined estimate.

Of course A is unknown and so are all the t_i. However, note that $\sum_i P_i(t_i - t) = 0$. Thus we define $\widehat{A}$ to be the solution obtained by iteratively solving the $k+1$ equations

$$\sum_i \widehat{P}_i u_i = 0,$$

$$\widehat{A}^T\widehat{A}_i = \exp(-M(t))A^TA\exp(M(t_i)) \approx \exp(M(t_i - t)) = \exp(M(u_i))$$

for the unknowns $\widehat{A}$ and $u_i = t_i - t$.

Figure 6 shows a plate geometry which would lead to independent estimates of the same rotation.

Royer and Chang, however, used the algorithm for combining independent estimates of the same rotation in a different manner. Geophysical evidence indicates that if differential motion between Australia and India occurs, it would have started within the last seven million years. It follows that all reconstructions for isochrons older than seven million years estimate the same India to Australia motion. In other words, all reconstructions from isochrons older than seven million years can be combined to estimate the motion of India to Australia over the most recent seven million years.

The basic message of this section and the preceeding section is that asymptotically, as $\kappa \to \infty$, standard tools from linear statistics can be adapted to the plate tectonics problem. This is not surprising: both the parameter space $\mathcal{S}O(3) \times \Omega_3 \times \cdots \times \Omega_3 = \{\theta = (A, \eta_1, \ldots, \eta_s)\}$ and the data space Ω_3^N are differentiable manifolds. This means that, locally, they can be linearly approximated by their tangent hyperplanes. Intuitively, any asymptotics (large sample or large κ) of a consistent estimator $\widehat{\theta}$ of $\theta = (A, \eta_1, \ldots, \eta_s)$, flattens the parameter space. However, large κ asymptotics concentrates the data ever closer to its "true" values and hence also flattens the data space (locally around the true value of each data point). Hence for spherical data, large κ asymptotic expansions are much simpler than large N asymptotics because the latter flattens only the parameter space. They are also more realistic, because, as discussed at beginning of Section 2, the errors in tectonic data are the order of 10 km and hence the curvature of the earth is not locally significant.

Note also that, although large κ asymptotics *locally* flattens both the parameter and the data space, neither gets *globally* flattened. Global flattening would occur, for example, if one were to take a fixed parameterization such as latitude and longitude for Ω_3 or Euler angles for $\mathcal{S}O(3)$. As discussed previously, the use of local coordinates, centered around the true values of the parameters η_i and A, has greatly ameliorated the curvature effects of the nonlinear regression.

Chang (1993) has an example of the disastrous consequences of an attempted global flattening of $\mathcal{S}O(3)$ in the calculation of the errors of a combination of two independently estimated rotations. On the other hand, the curvature of the earth is highly relevant globally because, as mentioned above, the Pacific plate must be reconstructed to the North American plate by using a chain of diverging plate boundaries going around the world in the opposite direction. In other words, although it is important to use local coordinates to calculate the distribution of the deviation $A^T\widehat{A}$, it is equally important to think of the problem in terms of rotations of the sphere as opposed to Euclidean motions in the plane. Mathematicians have always used this approach, of formulating theorems and definitions globally, but using local coordinates for calculation, when working with differentiable manifolds.

4. THE BEHRENS–FISHER PROBLEM REARS ITS UGLY HEAD

Consider the following regression problem:

$$Y_1 = X_1\beta_1 + \varepsilon_1,$$
$$Y_2 = X_2\beta_2 + \varepsilon_2,$$

where Y_i is $n_i \times 1$, X_i is $n_i \times p$ and ε_i is multivariate normal $N_{n_i}(0, \sigma_i^2 I_{n_i})$. If $\widehat{\beta}_i$ denotes the usual least squares estimate of β_i, and we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we can take a pooled estimate s^2 of σ^2 so that

$$\frac{1}{p}\left(\widehat{\beta}_1 + \widehat{\beta}_2 - \beta_1 - \beta_2\right)^T \left(s^2(X_1^TX_1 + X_2^TX_2)\right)^{-1} \left(\widehat{\beta}_1 + \widehat{\beta}_2 - \beta_1 - \beta_2\right)$$

is distributed $F(p, n_1 + n_2 - 2p)$. If $\sigma_1^2 \neq \sigma_2^2$, however, we will say that a "generalized" Behrens–Fisher problem arises. We refer the reader to Scheffé (1970) for a discussion of approaches to the usual Behrens–Fisher problem.

Consider now the problem of combining two rotations $\widehat{A}$ and $\widehat{B}$. Suppose these two rotations are estimated from two diverging boundaries. Recalling the regression formulation of Section 2 and that $\kappa^{-1/2}$ in that formulation plays the role of a scale parameter, if $\kappa_A \neq \kappa_B$ we obtain a "generalized" Behrens–Fisher problem of this form. We note that if different data sets are used to estimate A and B, the weights σ_{ij} and $\tilde{\sigma}_{ik}$ might not be consistently assigned between the two data sets, and hence one might suspect that $\kappa_A \neq \kappa_B$.

Alternatively, consider the simultaneous estimation of a triple junction. After linearization, this problem has the following form: $Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i$, where i indexes the three arms of the triple junction, β is a six-dimensional parameter corresponding to the two independent rotations, γ_i is a parameter corresponding to the sections on the ith arm and ε_i is distributed $N_{N_i}(0, (1/\kappa_i)I_{N_i})$ where N_i is the number of data points on the ith arm of the triple junction. If the κ_i are not all equal, we have another, different generalized Behrens–Fisher problem.

In Royer and Chang (1991), the triple junction generalized Behrens–Fisher problem did not arise,

TABLE 2
Selected data from Royer and Chang (1991)[1]

	N	s	df	$1/\hat{\kappa}$
SOM/IND	24	4	13	0.085
SOM/AUS/ANT tj	136	24	82	0.34

[1]Plates labeled as in Figure 4.

but for some isochrons, the combining rotations generalized Behrens–Fisher problem did arise. In particular, consider the calculations in Table 2 from that paper (at chron 13 or 35.5 million years before the present).

We apply the usual F-test for equality of variances in a weighted regression context. This test is highly sensitive to normal distribution assumptions. However, because the $\kappa\sigma_{ij}^{-2}$ and $\kappa\tilde{\sigma}_{ik}^{-2}$ are on the order of 10^5, the discussion following (3) yields that the Fisher–von Mises–Langevin distribution is well approximated by a normal distribution. The value of the test statistic is 0.34/.085, and comparing this to a $F_{82,13}$ distribution, we conclude that $\kappa_{\mathrm{SOM/IND}}$ is not the same as $\kappa_{\mathrm{SOM/AUS/ANT}}$.

Johansen (1980) extended Welch's approximate solution for the classical Behrens–Fisher problem to the regression context. Restating Johansen's result slightly, let

$$Y = X\beta + \varepsilon,$$

$$\mathrm{Cov}(\varepsilon) = \text{ block diag } \left[\sigma_1^2\Sigma_1, \ldots, \sigma_k^2\Sigma_k\right]$$

where the Σ_i are $n_i \times n_i$ known matrices, but the σ_i^2 are unknown. Write $Y = \left[Y_1^T \cdots Y_k^T\right]^T$ and $X = \left[X_1^T \cdots X_k^T\right]^T$ where Y_i and X_i are $n_i \times 1$ and $n_i \times p$, respectively. Let s_i^2 be the standard (Σ_i^{-1}-weighted) regression estimate of σ_i^2 for the model $Y_i = X_i\beta + \varepsilon_i$. Johansen considers the weighted extra error sum of squares statistic Q_W, with weights estimated by s_i^2, for testing a linear hypothesis on β and approximates its distribution by a distribution of the form $d \cdot F_{\nu_1,\,\nu_2}$ where d and ν_2 are estimated from the data to match the mean and variance of Q_W with those of $d \cdot F_{\nu_1,\,\nu_2}$ to terms of order $o(\min(n_1, \ldots, n_k)^{-1})$.

Kirkwood and Chang (1998) apply Johansen's work to the plate tectonics context. The chief mathematical difficulty is that Johansen's results are asymptotic as the sample size $N \to \infty$, whereas plate tectonics is done with $\kappa \to \infty$ asymptotics. They show that if $N, \kappa \to \infty$ so that $N/\kappa \to 0$, then both types of asymptotics can be simultaneously used. These results have been implemented in the algorithms for combining two independent rotations, for combining multiple independent estimates of the same reconstruction and for fitting a triple junction.

It is well known that in the classical Behrens–Fisher problem the distribution of $(Y_1 + Y_2 - (\theta_1 + \theta_2))/\sqrt{s_1^2 + s_2^2}$ can be conservatively approximated by a t-distribution with $\min(\nu_1, \nu_2)$ degrees of freedom where ν_i is the degrees of freedom in the distribution of s_i^2. Kirkwood and Chang (1998) also generalize this result to the plate tectonics context. Whereas, for example, applying the Johansen–Welch approximation to a chain of diverging plate boundaries requires the full data set for each boundary, the conservative approximation requires only summary statistics. This makes it especially suitable for combining reconstructions calculated by different groups of geophysicists (because they tend to specialize in different parts of the world).

5. THE INDIAN OCEAN ACCORDING TO ROYER AND GORDON (1997)

By 1997, the data density in the Indian Ocean had improved sufficiently to cast doubt upon the adequacy of the plate model proposed by Royer and Chang (1991). In particular, with more data, Royer and Gordon proposed that a small piece, which they called Capricorn (for the Tropic of Capricorn which passes through it), of the previously defined Australian plate moves separately from the main Australian plate. Figure 7 illustrates their proposed geometry.

To establish the inadequacy of the 1991 model, Royer and Gordon used two extra error sum of squares tests. These tests were based upon the analysis in Section 2. We recall that the $r(A, \eta)$ of (1) is asymptotically the error sum of squares of a suitably defined regression. These extra error sum of squares tests are asymptotically valid as long as the reduced model parameter space is a submanifold of the full parameter space.

The first test Royer and Gordon applied was a test of the "closure" of the Australia/Somalia/Antarctica triple junction. Assuming the correctness of the 1991 geometry (as shown in Figure 4), the reconstruction of Australia to Somalia followed by the reconstruction of Australia to Antarctica should reconstruct Australia to Antarctica.

To construct the extra error sum of squares test statistic, the full model assumes no relationship between these three plate boundaries. Thus if Australia is reconstructed to Somalia across the Central Indian Ridge, yielding a value r_1 in (1), Somalia to Antarctica across the Southwest Indian Ridge yielding r_2, and Australia to Antarctica across the Southeast Indian Ridge yielding r_3, the error sum of squares for the full model is $r_F = r_1 + r_2 + r_3$ with

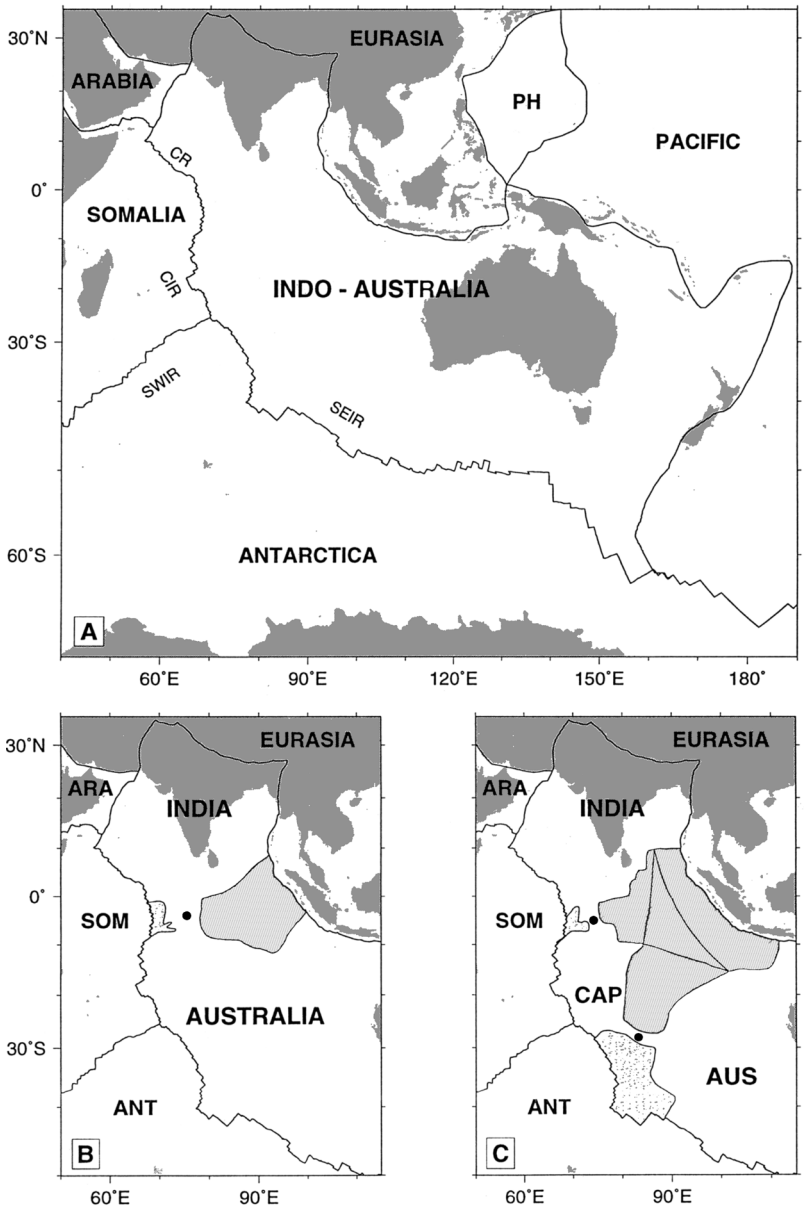

FIG. 7. *Plate geometries in the Indian Ocean area. Continuous lines in* (a) *show the plate boundaries between the major plates* (*PH = Philippine plate*). *Within the Indian Ocean, the plate boundaries are narrow midoceanic spreading ridges: the Carlsberg Ridge* (*CR*), *the Central Indian Ridge* (*CIR*), *the Southwest Indian Ridge* (*SWIR*) *and the Southeast Indian Ridge* (*SEIR*). (a) *Traditionally the Indian and Australian plates were considered as a single rigid plate;* (b) *Observed deformation of the Earth's crust* (*within the stippled area*) *and plate reconstructions suggested that the Indian and Australian were two distinct plates, separated by a wide and diffuse boundary* (*Wiens et al.,* 1985). *In this model, Australia rotates counter-clockwise relative to India about a pivot* (*solid circle*) *located within the plate boundary* (*Royer and Chang,* 1991; *DeMets, Gordon and Vogt,* 1994). *Darker stippled area is area of convergence, lighter stippled area is area of divergence;* (c) *Further analysis of plate reconstruction data as well as deformation patterns in the ocean basin showed that the former Indo-Australian plate is a composite plate made of three rigid components: the Capricorn* (*CAP*), *the Indian and the Australian plate, and of wide diffuse boundaries between them* (*stippled areas*) *where the plates are deforming. In this model, Australia rotates counter-clockwise relative to Capricorn about a pivot* (*solid circle*) *located within the plate boundary* (*Royer and Gordon,* 1997).

$N - 2s - 9$ degrees of freedom. Here N and s are the total number of data points and sections among the three boundaries.

The reduced model uses a triple junction fit, yielding a combined error sum of squares of r_R with $N - 2s - 6$ degrees of freedom. Assuming equal κ's all around, the usual F-test would compare $(r_R - r_F)/r_F$ with critical values from the $3(N - 2s - 9)^{-1}F_{3,\ N-2s-9}$ distribution.

In fact, Royer and Gordon compared $r_R - r_F = 39.9$ with χ^2_3. This is analagous to the correct (weighted) regression test when all the ε_i are known to have variance σ^2/w_i with w_i known and $\sigma^2 = 1$. Because all fitted $\widehat{\kappa}$ were less than 1, this approach would be expected to be conservative.

The second test was a test of the rigidity of the Australia and Antarctic plates across the Southeast Indian Ridge (SEIR). Royer and Gordon suspected that if a plate boundary exists it would intersect the SEIR at around $80^\circ E$. Thus the full model fitted Australia to Antarctica separately using data on the SEIR northwest of $77^\circ E$ and data on the SEIR southeast of $84^\circ E$. Let the error sum of squares of these two fits be r_1 and r_2, respectively, yielding a full model error sum of squares of $r_F = r_1 + r_2$. The reduced model fitted Australia to Antarctica using all the data on the SEIR (except that between $77^\circ E$ and $84^\circ E$), yielding an error sum of squares of r_R. Again Royer and Gordon eschewed the usual F test and conservatively compared $r_R - r_F = 21.0$ with χ^2_3.

Thus both tests emphatically contradicted the 1991 geometry.

To test the consistency of the proposed new geometry, Royer and Gordon tested the closure of the presumed Capricorn/Antarctica/Somalia triple junction. This is the same test as the earlier test of closure described above that Royer and Gordon applied to the Australia/Somalia/Antarctica triple junction, except that some of the data from the old Australia plate is attributed to the Capricorn plate in the newly defined geometry. In this test, the correctness of the new geometry is the null hypothesis (a situation similar to goodness of fit tests). The conservative χ^2 test, yielded $r_R - r_F = 5.0$ (p-value = 0.17) whereas the more conventional F-test yielded $F = 2.196$ (p-value = 0.09). This led Royer and Gordon to accept the correctness of the new geometry.

Thus between the 1991 and 1997 studies, the original Indo-Australian plate has been split into three components: India, Australia and Capricorn. As Royer and Gordon comment, "In our view the central tenet of plate tectonics is that plate interiors can be usefully approximated as being rigid. In that sense, our model is consistent with plate tectonics as long as the Indian, Capricorn and Australian components, and not the Indo-Australian composite, are each considered to be a plate." Nevertheless, they note the boundaries between these three plates are more diffuse than commonly encountered and the relative rates of motion between them are slow. Examining Figure 7, the new model shows Capricorn and Australia behaving as two arms of a scissor with a pivot point on their boundary. A similar model holds for India and Capricorn. Royer and Gordon conclude that "it makes sense to regard the Indo-Australian 'plate' as a higher level plate tectonic unit, which we refer to as a 'composite plate' " with India, Australia and Capricorn as its components.

6. INFLUENCE DIAGNOSTICS FOR TECTONIC DATA

Data is collected at great expense through oceanographic or aeromagnetic survey, often in remote parts of the globe. In addition, as perhaps may be evident from the description of the sources of error given above, conversion of the data, as collected, to identified locations on the magnetic anomaly lineations is also time-consuming. As one of the authors once commented to the other early in our collaboration, "If you knew how much work went into each data point, you would stop complaining about it."

For these reasons, it is important to be able to isolate possibly erroneous data or identify subsets of the data which might explain an unexpected reconstruction. In this section we describe modifications to standard regression diagnostics and illustrate their use for this purpose.

An alternative approach to test the closure of the Capricorn/Antarctica/Somalia triple junction would be to calculate the reconstructions $\widehat{A}_{\mathrm{CAP/ANT}}$, $\widehat{A}_{\mathrm{ANT/SOM}}$, and $\widehat{A}_{\mathrm{SOM/CAP}}$ using the data from the respective three arms of the triple junction and to test whether or not the combined reconstruction

$$\widehat{A}_{\mathrm{CAP/ANT}}\widehat{A}_{\mathrm{ANT/SOM}}\widehat{A}_{\mathrm{SOM/CAP}} \tag{10}$$

is significantly different from the identity. This test yielded a p-value of 0.076 (as opposed to 0.09 for the corresponding F-test).

Both this test and the extra error sum of squares test of closure performed by Royer and Gordon assume that the κ's on the three arms of the triple junction are equal. The data is given in Table 3.

The F-test for the equality of two κ's described in Section 4 would contradict this assumption.

Using the unequal κ approach outlined in Section 4, the p-value to test if (10) is significantly different from the identity is 0.021. This casts substantial

TABLE 3
Selected data from Royer and Gordon (1997)[1]

	N	s	df	$1/\widehat{\kappa}$
CAP/ANT	30	7	13	0.4104
ANT/SOM	56	11	31	0.4305
SOM/CAP	176	23	127	0.8754

[1]Plates labeled as in Figure 7c

doubt on the conclusion by Royer and Gordon (1997) that their proposed new geometry is correct. Visually, it appears that the new geometry is correct, and to investigate whether or not this unfortunate p-value is due to a few anomalous bad data points, we have used a modification of the influence matrix ideas proposed by Lu, Ko and Chang (1997).

The influence function $IF(x^*; \beta, F)$ of a statistical functional $\beta(F)$ is defined to be (see Hampel, Ronchetti, Rousseeuw and Stahel, 1986)

$$IF(x^*; \beta, F) = \frac{d}{d\varepsilon}\beta((1-\varepsilon)F + \varepsilon\delta_{x^*})\Big|_{\varepsilon=0}, \tag{11}$$

where F is the cumulative distribution function of a random variable X and δ_{x^*} is the cumulative distribution function for the random variable which always takes the value x^*. If β is a p-dimensional vector, so is IF.

For the linear regression model $y = x\beta + \varepsilon$ with normal errors, if we envision that (x, y) has a joint distribution F,

$$\beta(F) = E_F(x^T x)^{-1} E_F(x^T y)^{-1}, \tag{12}$$

where E_F denotes expected value under the distribution F. Note that with the definition (12), the least squares estimate of β is $\widehat{\beta} = \beta(\widehat{F})$ where $\widehat{F}$ is the empirical cumulative distribution. In this case, (11) works out to be

$$IF(x^*, y^*; \beta, F) = E_F(x^T x)^{-1} x^{*T}(y^* - x^*\beta(F)). \tag{13}$$

When evaluated at $\widehat{F}$, (13) becomes

$$IF(x^*, y^*; \beta, \widehat{F}) = n\left[\sum_i x_i^T x_i\right]^{-1} x^{*T}(y^* - x^*\widehat{\beta}). \tag{14}$$

Several points are immediately obvious from definition (11). First, if x^* happens to be a data point, $IF(x^*; \beta, F)$ measures the effect on $\widehat{\beta}$ of the infintesimal deletion of x^* as opposed to Cook's D which measures, in the regression context, the effect of complete deletion. Second, if β takes values in a manifold, $IF(x^*; \beta, F)$ is a tangent vector to that manifold at $\beta(F)$, because the derivative $c'(0)$ of a curve $c(\varepsilon)$ is always tangent to the curve at $c(0)$.

If $f\colon R^p \to R^q$, its derivative $Df(0)$ is a linear transformation $R^p \to R^q$ even if f is nonlinear. Thus if $\{x_1, \ldots, x_k\}$ is a collection of points, we can define the influence of this group of points on β by

$$IF(w; \beta, F) \equiv \frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\beta\left((1-\varepsilon)F + \varepsilon\sum_i w_i\delta_{x_i}\right),$$

where each w_i is a case weight for x_i, $i = 1, \ldots, k$. The linearity of derivatives (and specifically the chain rule) implies that

$$IF(w; \beta, F) = \sum_i w_i IF(x_i; \beta, F). \tag{15}$$

Although $(1-\varepsilon)F + \varepsilon\sum_i w_i\delta_{x_i}$ is a distribution function only if $\sum_i w_i = 1$ and $w_i \geq 0$ for each i, we can use the right-hand side of (15) to extend the definition of $IF(w; \beta, F)$ to arbitrary weight vectors $w = (w_1, \ldots, w_k)^T$.

Now suppose $\theta = [\beta \quad \psi]^T$, where θ is the full parameter and ψ is a nuisance parameter, and decompose the Fisher information matrix, evaluated at $\theta(F)$, as

$$H = \begin{bmatrix} H_{\beta\beta} & H_{\beta\psi} \\ H_{\psi\beta} & H_{\psi\psi} \end{bmatrix}.$$

We can use H to standardize $IF(w; \beta, F)$,

$$IF(w; \beta, F)^T\left[H_{\beta\beta} - H_{\beta\psi}H_{\psi\psi}^{-1}H_{\psi\beta}\right]IF(w; \beta, F).$$

Using (15), this is $w^T\mathrm{PSIM}w$ where PSIM is, by slight extension of the terminology in Lu et al (1997), the partial standardized influence matrix, whose $(i, j)-$th entry is

$$IF(x_i; \beta, F)^T\left[H_{\beta\beta} - H_{\beta\psi}H_{\psi\psi}^{-1}H_{\psi\beta}\right]IF(x_j; \beta, F).$$

When $\theta = \beta$, $F = \widehat{F}$ and $\{x_1, \ldots, x_k\}$ are the data points, Lu, Ko and Chang (1997) show that the dominant eigenvalue of the PSIM is, to first approximation, the local influence as defined by Cook (1986).

In the tectonic context, it is of interest to study the influence of a group of points, namely those points that constitute a section on the parameter of interest: the rotation. In other words we propose to consider the influence of the ith section $\{u_{ij}, v_{ik}\}$ on the estimator $\widehat{\beta}$ defined by $A^T\widehat{A} = M(\widehat{\beta})$. In this case the nuisance parameters represent the section normals.

We have noted above that influence statistics measure the effects of infintesimal deletion. It is meaningful to talk about the partial deletion of a section of points and perhaps it is more natural to talk about the infinitesimal deletion of a section of points than of a single point. We also note that the essential features of this discussion are that the data breaks up into groups (sections) whose influence on a parameter of interest β is to be studied. These features occur elsewhere in statistics: in treatment-block designs, we might be interested

in the influence of each block on the estimated treatment effects.

We will define the *partial sectional influence* of a subcollection $\{x_1, \ldots, x_k\}$ of the data to the largest eigenvalue λ of the PSIM evaluated at $\{x_1, \ldots, x_k\}$ and $F = \widehat{F}$. This subcollection of points can be influential due to misfit, placement of the points $\{x_1, \ldots, x_k\}$ or size k of the subcollection. To remove the effect of the size of the subcollection, we will define the *normalized partial sectional influence* to be $(\lambda/k - 1)\sqrt{k/2}$. The rationale for this normalization can be found in Lu, Ko and Chang (1997).

Referring to Table 3, the misfit is worst between Somalia and Capricorn. We have applied these ideas to examine the influence of the sections on this boundary to the reconstruction $\widehat{A}_{\mathrm{SOM/CAP}}$. Figure 8 shows this data, where the labelling is by sections. Figure 9 shows the partial sectional influence versus the normalized partial sectional influence.

Examining Figure 9, we see that section W is clearly flagged under both of these criteria. Section V is less clearly flagged, but the graph suggests that the influence of Section V is due primarily to misfit or location, and not to its size. Because V and W are in the same geographic location, close to the triple junction, it seems appropriate to examine both V and W.

Recall, that an unequal κ test,

$$\widehat{A}_{\mathrm{CAP/ANT}}\widehat{A}_{\mathrm{ANT/SOM}}\widehat{A}_{\mathrm{SOM/CAP}}$$

is significantly different from the identity that has p-value of 0.021. If W is removed, that p-value increases to 0.091 and if both V and W are removed, the p-value jumps to 0.312. We note that W has 11 points (representing 6.25% of the SOM/CAP data and 4.20% of the complete data) and V has 5 points (representing 2.54% and 1.91% of the SOM/CAP and complete data, respectively).

Mindful of the admonition to keep requests for data reexamination to a minimum, we suggested that the data on sections V and W be rechecked. Five points, two on section V and three on section W, were found to be misidentified. When these points were deleted, but the remainder of sections V and W retained, the unequal κ-test statistic has a p-value of 0.080.

The analysis suggests that the new Royer–Gordon geometry is consistent with the data. Because the complete deletion of V and W has such a dramatic effect on the p-value, the analysis further suggests the possibility of some slight nonrigidity near the locations of sections V and W, that is, on the Central Indian Ridge (SOM/CAP boundary) near the triple junction.

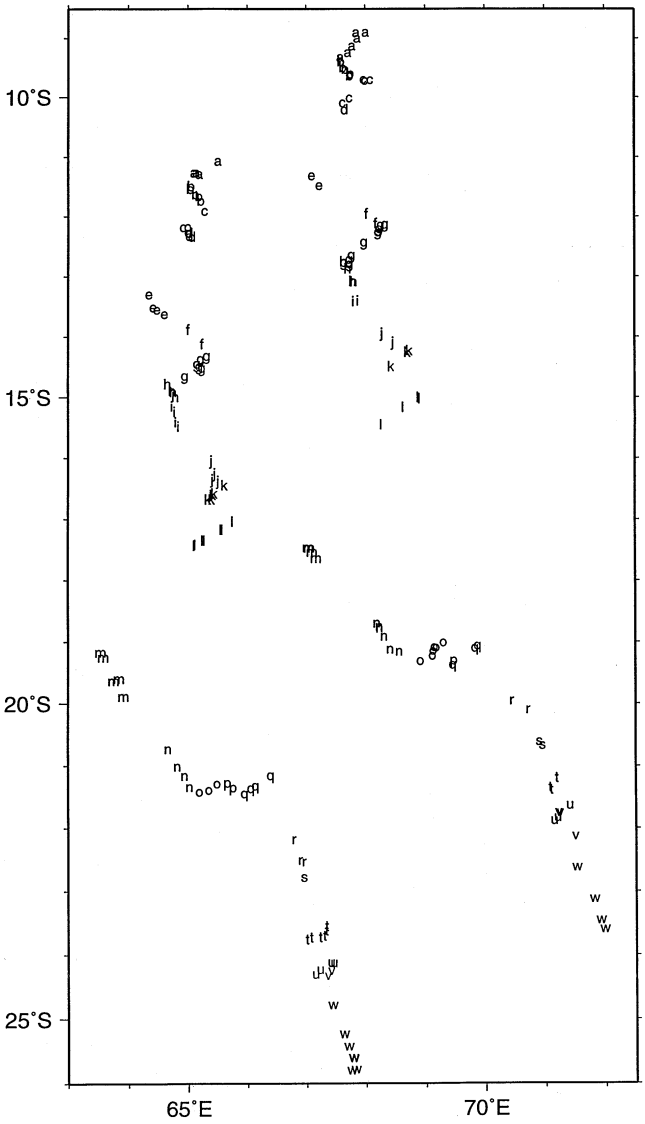

FIG. 8. *SOM/CAP data, labeled by section.*

7. CONCLUSIONS

This paper discusses how linear regression techniques have been applied to a statistical analysis of tectonic plate reconstructions. The problem is intrinsically nonlinear, but curvature effects can be minimized by using a local parameterization of the rotation group and separate "local flat earth" approximations at each data point. This approach is fully consistent with standard mathematical approaches for dealing with differentiable manifolds.

Further research is needed both to accomodate new types of data and to develop improved statistical techniques as increasing data density allows detection of deviation from the basic models considered here. For example, in some parts of the world, the data density is sufficient to support a model

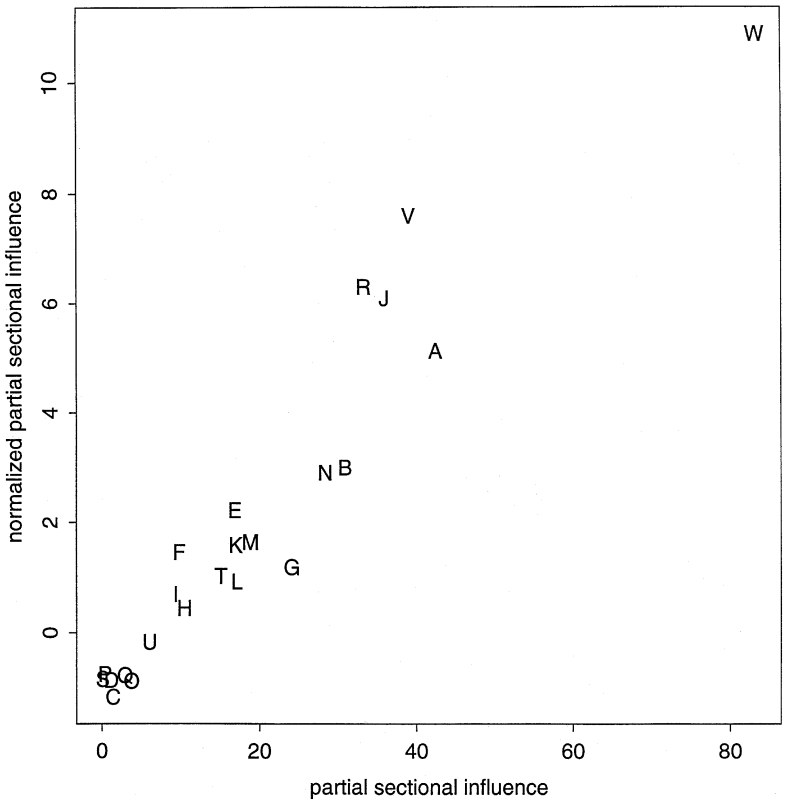

FIG. 9. *SOM/CAP data: partial sectional influence versus normalized partial sectional influence, letters refer to section name of the data shown in Figure* 8. *Sections which most influence the reconstruction have large partial sectional influence. This influence can be due to sections with a large number of data points; normalized partial sectional influence corrects for this effect. We see that Section* W *is clearly labelled as influential, even after normalization. This can be due to the orientation of the section, errors in the data, or local nonrigidities. Section* V *is geographically close to Section* W *and also appears to be somewhat influential.*

of boundaries that are not piecewise straight, but rather exhibit some slight curvature. Chang, Hendriks, and Yang (1999) discuss a suggested approach for this situation.

Royer and Gordon (1997) propose a hierarchical model of composite plates with components which are rigid but move slightly relative to each other. This suggests that in the future, as data becomes more dense and of higher quality, statistical techniques to analyze plate motions which exhibit mild nonrigidities will be needed. Nothing (of which the authors are aware) has been done on this problem.

APPENDIX: MOVING COORDINATE SYSTEMS

We will denote by $\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3$ a hypothetical fixed coordinate system. In practice, since all the plates move, $\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3$ is unobservable. Suppose the lineation was extruded at time $t_0 < 0$, where $t = 0$ denotes the present. Let $\mathbf{U}_i(t)$, respectively, $\mathbf{V}_i(t)$, be the evolution of $\mathbf{e}_i$ in t in the U and V plates, respectively. Thus $\mathbf{U}_i(t_0) = \mathbf{e}_i = \mathbf{V}_i(t_0)$.

Let $p = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$ be a point extruded at time t_0. p moves along the trajectory $u(t) = u_1\mathbf{U}_1(t) + u_2\mathbf{U}_2(t) + u_3\mathbf{U}_3(t)$ on the U plate and the trajectory $v(t) = u_1\mathbf{V}_1(t) + u_2\mathbf{V}_2(t) + u_3\mathbf{V}_3(t)$ on the V plate. Notice that in the coordinate system $\mathbf{U}_1(t),\ \mathbf{U}_2(t),\ \mathbf{U}_3(t)$ which is fixed to the U plate, the coordinates of $u(t)$ are $[u_1\ \ u_2\ \ u_3]^T$. Elementary matrix algebra establishes that, in the coordinate system $\mathbf{U}_1(t),\ \mathbf{U}_2(t),\ \mathbf{U}_3(t)$, $v(t) = A(t)[u_1\ \ u_2\ \ u_3]^T$ where $A(t) = [\mathbf{U}_1(t)\ \ \mathbf{U}_2(t)\ \ \mathbf{U}_3(t)]^T[\mathbf{V}_1(t)\ \ \mathbf{V}_2(t)\ \ \mathbf{V}_3(t)]$.

Thus $A(0)$ gives the motion of the V plate relative to the U plate, over the time period $(t_0, 0)$, in a coordinate system fixed in the U plate. Alternatively $A(0)^T$ reconstructs the past position of the V plate, relative to the U plate, with the U plate fixed in its present location. Similarly, $A(0)^T$ gives the motion of the U plate relative to the V plate in a coordinate system fixed in the V plate and $A(0)$ reconstructs the past position of the U plate, relative to the V plate, with the V plate fixed in its present location.

Note that $[\mathbf{V}_1(t)\ \ \mathbf{V}_2(t)\ \ \mathbf{V}_3(t)][\mathbf{U}_1(t)\ \ \mathbf{U}_2(t)\ \ \mathbf{U}_3(t)]^T$ is the motion of the V plate, relative to the U plate, in the fixed and unobservable coordinate system $\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3$.

REFERENCES

CHANG, T. (1988). On reconstructing tectonic plate motion from ship track crossings. *J. Amer. Statist. Assoc.* **83** 1178–1183.

CHANG, T. (1993). Spherical regression and the statistics of tectonic plate reconstructions. *Internat. Statist. Rev.* **61** 299–316.

CHANG, T., HENDRIKS, H. and YANG, R. (1999). An application to plate tectonics of regression estimation for matched Ornstein–Uhlenbeck processes. Unpublished manuscript.

CHANG, T., STOCK, J. and MOLNAR, P. (1990). The rotation group in plate tectonics and the representation of uncertainties of plate reconstructions. *Geophys. J. Int.* **101** 649–661.

COOK, R. D. (1986). Assessment of local influence. *J. Roy. Statist. Soc. Ser. B* **48** 133–169.

DEMETS, C., GORDON, R. G. and VOGT, P. (1994). Location of the Africa-Australia-India triple junction and motion between the Australian and Indian plates: results from an aeromagnetic investigation of the Central Indian and Carlsberg ridges. *Geophys. J. Int.* **119** 893–930.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HELLINGER, S. J. (1981). The uncertainties of finite rotations in plate tectonics. *J. Geophy. Res.* **86B** 9312–9318.

JOHANSEN, S. (1980). The Welch–James approximation to the distribution of the residual sums of squares in a weighted linear regression. *Biometrika* **67** 85–92.

KIRKWOOD, B. H. and CHANG, T. (1998). Combining estimates of tectonic plate rotations: an extension of Welch's method to spherical regression. *J. Multivariate Anal.* **65** 71–108.

KIRKWOOD, B. H., ROYER, J.-Y., CHANG, T. and GORDON, R. (1999). Statistical tools for estimating and combining finite rotations and their uncertainties. *Geophys. J. Int.* **137** 408–428

LU, J., KO, D. and CHANG, T. (1997). The standardized influence matrix and its applications. *J. Amer. Statist. Assoc.* **92** 1572–1580.

ROYER, J.-Y. and CHANG, T. (1991). Evidence for relative motions between the Indian and Australian plates during the last 20 Myr from plate tectonic reconstructions. Implications for the deformation of the Indo-Australian plate. *J. Geophys. Res.* **96** 11779–11802.

ROYER, J.-Y. and GORDON, R. G. (1997). The motion and boundary between the Capricorn and Australian plates. *Science* **277** 1268–1274.

SCHEFFÉ. H. (1970). Practical solutions of the Bherens–Fisher problem. *J. Amer. Statist. Assoc.* **65** 1501–1508.

SHAW, P. R. and CANDE, S. C. (1990). High-resolution inversion for South Atlantic plate kinematics using joint altimeter and magnetic anomaly data. *J. Geophys. Res.* **95** 2625–2644.

STOCK, J. and MOLNAR, P. (1983). Some geometrical aspects of uncertainties in combined plate reconstructions. *Geology* **11** 697–701.

WESSEL, P. and SMITH, W. H. F. (1995). New version of the Generic Mapping Tools released. *EOS, Trans. Amer. Geophys. Un.* **76** 329.

WIENS, D. A., DEMETS, D. C., GORDON, R. G., STEIN, S., ARGUS, D., ENGELN, J. F., LUNDGREN, P., QUIBLE, D., STEIN, C., WEINSTEIN, S. and WOODS, D. F. (1985). A diffuse plate boundary model for Indian Ocean tectonics. *Geophys. Res. Lett.* **12** 429–432.