Appendix I: Algorithm of *NeoKinema*

by Peter Bird, UCLA, 2005

Overview

Geodetic studies over the past century have shown that velocities of benchmarks near the most active faults are not steady because of cycles of elastic strain accumulation and release in earthquakes and/or creep events. Extrapolating this result to faults with mean slip rates of 1 mm/a or less, we expect that velocities adjacent to such faults might vary significantly if averaged over less than 10^4 years. At longer time-scales, plate tectonic models based on marine magnetic anomalies show that large plates change their velocities on a scale of 10^7 years due to the birth and death of spreading ridges, subducting slabs, and other plate-boundary faults. The smaller plates within complex "orogens" [*Bird*, 2003] might be expected to show important velocity variations on a scale of 10^6 years because less relative advection of faults is needed to significantly change the shape of a small plate. However, it is reasonable to expect that, if surface velocities could be measured over scales of 10^4 to 10^6 years, they would be stable in most regions. This is the "long-term average" velocity field that we seek to estimate with program *NeoKinema.*

To first order, the strain rates and fault slip rates obtained from derivatives of the long-term average velocity field should be free of elastic strain contributions, and result instead from permanent strain mechanisms such as frictional sliding in the upper crust and dislocation creep in the lower crust. Therefore, it is also reasonable to expect that long-term average strain rates in the upper lithosphere should be proportional to long-term average seismic moment production (in N m^{-1} s^{-1}). The necessary conversion factors are the elastic shear modulus (which is well known) and the "coupled lithosphere thickness" contributing to seismicity, which has been estimated by *Bird & Kagan* [2004] based on 20^{th}-century seismicity. Thus, results from *NeoKinema* lead directly to stationary models of long-term average seismicity and seismic hazard. Stationary models have value in the design of zoning and building code ordinances, which cannot be expected to change rapidly in response to time-dependent seismic hazard forecasts. Furthermore, better knowledge of the long-term average seismicity map contributes to basic science studies of the time-dependence of seismicity by defining a basic process relative to which positive and negative seismicity anomalies of 10^2 to 10^3 year duration can be measured, using existing catalogs supplemented by historical records and archeology.

Models of the long-term average velocity field can be either "forward," "dynamic" models (based on the momentum equation, and assumed rheologies) or "inverse," "kinematic" models (based on observations, with additional constraints to increase realism). Dynamic models contribute more to theoretical understanding of tectonophysics, because suites of model experiments can elucidate the effects of rheologic and boundary parameters. But kinematic models are more reliable estimators of seismic hazard, because they fit available data better in particular actual cases. *NeoKinema* is kinematic. The data sets it fits include (1) geodetic data from time/space windows without major earthquakes; (2) long-term average fault slip rates from geologic data; (3) principal stress directions; and (4) velocity boundary conditions from plate tectonic models. The assumptions it employs to increase realism are (a) microplate tectonics: anelastic strain rates in unfaulted continuum lithosphere should be minimized; and (b) isotropy: principal strain rate axes in unfaulted continuum lithosphere should coincide in direction with principal stress directions from data.

Objective Function

In many inverse problems the data are discrete, because they come from measurements at distinct points. Assume that all data that constrain velocity or strain rate at particular points have been transformed to scalar rate estimates r_k. (Subscript $k = 1, \ldots, K$ identifies the scalar datum, which is typically one horizontal component of long-term average velocity derived from a geodetic benchmark velocity.) Let the corresponding scalar rate predictions derived from the velocity field of the model be called p_k. Assume that each scalar rate r_k has an uncertainty that can be approximated by a Gaussian probability distribution with standard deviation σ_k, and assume <u>temporarily</u> that the errors in these rates are independent. Then the natural logarithm of the density of the joint probability that the model matches all the data is formed from the individual probability densities (Φ) as:

$$S \equiv \ln\left\{\prod_{k=1}^{K}\Phi(p_k = r_k)\right\} = \sum_{k=1}^{K}\ln\left[\Phi(p_k = r_k)\right] = -\sum_{k=1}^{K}\left[\frac{(p_k - r_k)^2}{2\sigma_k^2} + \ln(\sigma_k) + \ln\sqrt{2\pi}\right] \qquad (1)$$

and the part of this which is variable (with respect to changes in the model) is the familiar weighted-squares-of-prediction-errors criterion

$$S' \equiv -\sum_{k=1}^{K}\frac{(p_k - r_k)^2}{2\sigma_k^2} \qquad (2)$$

which is to be maximized.

However, in the *NeoKinema* algorithm we also consider some constraints (geologic slip rates) to apply all along the trace of a fault, and other constraints (minimization of strain rate, and isostropy) to apply all across the area of unfaulted continuum. There is no natural way of "counting" these constraints as discrete data (or pseudo-data), and no natural, "correct" weighting of these constraints against point data in the objective function. Instead, we leave this choice to the user of the program, by introducing parameters called "reference length" L_0 and "reference area" A_0 which are used to maintain non-dimensionality in a generalized objective function that includes both line- and area-integrals:

$$S'' \equiv -\sum_{k=1}^{K}\frac{(p_k - r_k)^2}{2\sigma_k^2} - \frac{1}{L_0}\sum_{m=1}^{M}\int_{\text{length}}\frac{\left(p_m - r_m\right)^2}{2\sigma_m^2}\,\mathrm{d}\ell - \frac{1}{A_0}\sum_{n=1}^{3}\iint_{\text{area}}\frac{\left(p_n - r_n\right)^2}{2\sigma_n^2}\,\mathrm{d}a \qquad (3)$$

where $m = 1,\ldots,M$ enumerates the target rates r_m associated with fault-slip degrees of freedom, and $n = 1, 2, 3$ enumerates the 3 target rates r_n associated with the 3 components of strain-rates at each continuum point. The first term of this objective function includes the target velocities derived from geodetic benchmark-velocity data, the second term includes the targets derived from geologic slip-rate data, and the third term includes the targets derived from stress-direction data (and the stiff microplate assumption). Therefore, L_0 and A_0 can be considered as dimensional tuning parameters to be adjusted, by trial-and-error or systematic search, to equalize the fit of *NeoKinema* models to all 3 classes of data. If a calculation is performed with no geodetic data ($K = 0$), then the solution will depend only on a single dimensional tuning parameter: the ratio A_0/L_0 .

Finite Element Approximation

It is only necessary to estimate the horizontal components of the long-term average velocity, and only necessary to do this on the planet's surface. Therefore, we divide the area of the model into spherical-triangle finite elements [*Kong & Bird,* 1995] and solve for the horizontal components of velocity at each node. Long-term average velocities at other points are determined by interpolation, and long-term anelastic strain rates are determined by differentiation. (Where these elements are small, the surface of the sphere is locally almost flat, and the nodal functions of such elements are very close to those of plane-triangle "constant-strain" finite elements.)

On the surface of a spherical planet with radius *R*, define a coordinate system of colatitude (θ) measured southward from the North Pole, and longitude (ϕ) measured eastward from the prime meridian. The unknowns in each velocity solution are the horizontal θ-components and ϕ-components of the velocity of the surface. All predicted rates p_k , p_m , p_n can be expressed as different linear combinations of the velocity components *v* (Southward) and *w* (Eastward) at each of the *J* nodes of a finite element grid:

$$p_k = c_k + \sum_{j=1}^{J} \left(f_{kj} v_j + g_{kj} w_j \right). \qquad (4)$$

(There are also 2 similar equations where *k* is replaced by *m* or *n*; however, in these equations the coefficients f_{mj} , g_{mj} , f_{nj} , g_{nj} are considered to be functions of position along a fault trace or across the model area, rather than constants.)

System of Linearized Equations

With these linear relations between nodal velocities and model predictions, S'' is a quadratic form in the nodal-velocity-component values v_j and w_j , so it is maximized by finding the single stationary point in multi-dimensional velocity space where

$$\frac{\partial S''}{\partial v_i} = 0 = \frac{\partial S''}{\partial w_i}; \qquad i = 1, \ldots, J \,. \tag{5}$$

Algebraically, this leads to a $2J \times 2J$ linear system, which can be thought of as being partitioned into 4 submatrices times two subvectors equaling two subvectors:

$$\left[\begin{array}{c|c} A_{ij} & B_{ij} \\ \hline C_{ij} & D_{ij} \end{array}\right]\left[\begin{array}{c} v_j \\ \hline w_j \end{array}\right] = \left[\begin{array}{c} E_i \\ \hline F_i \end{array}\right] \tag{6}$$

using the abbreviations

$$A_{ij} = \sum_{k=1}^{K} \frac{f_{ki} f_{kj}}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{f_{mi} f_{mj}}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_o} \sum_{n=1}^{3} \iint_{\text{area}} \frac{f_{ni} f_{nj}}{\sigma_n^2} \mathrm{d}a \tag{7a}$$

$$B_{ij} = \sum_{k=1}^{K} \frac{f_{ki} g_{kj}}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{f_{mi} g_{mj}}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_o} \sum_{n=1}^{3} \iint_{\text{area}} \frac{f_{ni} g_{nj}}{\sigma_n^2} \mathrm{d}a \tag{7b}$$

$$C_{ij} = \sum_{k=1}^{K} \frac{g_{ki} f_{kj}}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{g_{mi} f_{mj}}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_o} \sum_{n=1}^{3} \iint_{\text{area}} \frac{g_{ni} f_{nj}}{\sigma_n^2} \mathrm{d}a \tag{7c}$$

$$D_{ij} = \sum_{k=1}^{K} \frac{g_{ki} g_{kj}}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{g_{mi} g_{mj}}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_o} \sum_{n=1}^{3} \iint_{\text{area}} \frac{g_{ni} g_{nj}}{\sigma_n^2} \mathrm{d}a \tag{7d}$$

$$E_i = \sum_{k=1}^{K} \frac{f_{ki}\left(r_k - c_k\right)}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{f_{mi}\left(r_m - c_m\right)}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_0} \sum_{n=1}^{3} \iint_{\text{area}} \frac{f_{ni}\left(r_n - c_n\right)}{\sigma_n^2} \mathrm{d}a \tag{7e}$$

$$F_i = \sum_{k=1}^{K} \frac{g_{ki}\left(r_k - c_k\right)}{\sigma_k^2} + \frac{1}{L_0} \sum_{m=1}^{M} \int_{\text{length}} \frac{g_{mi}\left(r_m - c_m\right)}{\sigma_m^2} \mathrm{d}\ell + \frac{1}{A_0} \sum_{n=1}^{3} \iint_{\text{area}} \frac{g_{ni}\left(r_n - c_n\right)}{\sigma_n^2} \mathrm{d}a \tag{7f}$$

(However, in practice it is efficient to reorder the row equations so that the unknown velocity components appear in the order $v_1, w_1, v_2, w_3, \ldots, v_J, w_J$. If we then renumber the nodes so as to minimize the maximum difference between indices of nodes connected by one finite element, the linear system will have reduced bandwidth and can be solved in less computer time and memory.)

Boundary Conditions

The equations stated above could be singular in the absence of boundary conditions, if there are no geodetic data, or if the geodetic velocity reference frame is free-floating. (This latter case will be discussed below.) In such cases, some edge(s) of the model must be fixed (or moved in a predetermined way) to define a velocity reference frame. To implement a velocity boundary condition, we replace the row equations that state that S'' is stationary (with respect to variations in those nodal velocity components) with simpler equations stating the desired values of these components. (Depending on the type of solver used, it may also be desirable to weight these constraint equations so that their coefficients are comparable to the eigenvalues of the unmodified matrix of coefficients.)

Even if the system is not singular, application of boundary conditions will often be desirable, to take advantage of the velocity information provided by plate tectonic models describing the relatively rigid portions of the large plates that lie outside orogens.

Only velocity boundary conditions are possible in *NeoKinema*. Stress is described only by orientation (but not magnitude) within the model domain, so "stress" (traction) boundary conditions are not available. However, if no velocity boundary condition is prescribed along a model edge, the effects will be similar to those of "traction-free" boundary conditions found in dynamic models. Such treatment would be appropriate if the model domain were limited to the overriding plate in a subduction zone, for that part of the model boundary running along the trench.

Continuum Stiffness: the Microplate Constraint

An essential context for all the fault-related geologic data showing locally intense straining is that they should compete with an *a-priori* assumption that in other places the strain-rate is close to zero. An appropriate formalism is to assign a zero target strain-rate, with a statistical uncertainty. A larger standard deviation could be attached to this null target rate in complex or poorly-studied regions where unknown faults might be buried and overlooked.

Referring to equation (3) above, the first ($n=1$) continuum constraint is expressed by:

$$\frac{\left(p_1-r_1\right)^2}{\sigma_1^2}=\frac{p_1^2}{\sigma_1^2}\equiv\frac{\dot{\varepsilon}_{\theta\theta}^2+\dot{\varepsilon}_{\theta\theta}\dot{\varepsilon}_{\phi\phi}+\dot{\varepsilon}_{\phi\phi}^2+\dot{\varepsilon}_{\theta\phi}^2}{\mu^2} \quad (8)$$

where μ is a scalar measure (such as the mean absolute value of the largest principal value) of a typical continuum strain-rate (in s^{-1}) in a particular application. One approach is to estimate μ from the off-fault seismicity of the model region, if a catalog with accurate locations is available. Alternatively, *NeoKinema* can be run repeatedly to estimate μ by the boot-strap method. We have found that (in realistic, non-degenerate problems) the strain rate of the continuum is largely determined by fault incompatibilities, fault discontinuities, discrepancies between adjacent fault slip rates, and/or discrepancies between geologic and geodetic data. Thus it often depends only weakly on the μ initially assumed, and convergence is rapid if a typical continuum strain-rate from the last calculation is input as the new value of μ.

NeoKinema also accepts distinct (positive) values for μ in each finite element, if desired. If any these values are zero or are not provided, a default value read from the input parameter file is used in that element.

The particular scalar function of the strain rate tensor that is used in (8) has the effect of causing unfaulted areas to behave as Newtonian-viscous sheets of lithosphere in a state of plane stress. The *NeoKinema* algorithm will result in velocities that minimize the area integral of squared strain-rates for the unfaulted elements; this is exactly the result one would obtain by deriving a dynamic FE algorithm from the momentum equation (in the absence of horizontal boundary tractions or body forces), adopting a linear rheology, and solving for velocity with inhomogeneous boundary conditions.

The 2×2 strain-rate tensor $\dot{\tilde{\varepsilon}}$ on the spherical surface is calculated by summing spatial derivatives of nodal functions multipled by nodal velocities. The nodal functions that we use were introduced by *Kong & Bird* [1995] and shown to satisfy the requirements of horizontality, continuity, and completeness:

$$\begin{bmatrix} v(\theta,\phi) \\ w(\theta,\phi) \end{bmatrix} = \sum_{j=1}^{J} \begin{bmatrix} G_{1,1}^{j}(\theta,\phi) & G_{2,1}^{j}(\theta,\phi) \\ G_{1,2}^{j}(\theta,\phi) & G_{2,2}^{j}(\theta,\phi) \end{bmatrix} \begin{bmatrix} v_j \\ w_j \end{bmatrix}. \quad (9)$$

In this notation, the superscript *j* on the vector nodal function $\vec{G}_x^j$ or nodal function component $G_{x,y}^j$ identifies the node that has unit velocity (all other nodes having zero velocity in this particular nodal function). Subscript $x = 1$ indicates the nodal function associated with unit southward velocity *v*; subscript $x = 2$ indicates the nodal function associated with unit eastward velocity *w*. Subscript $y = 1$ indicates the southward or θ-component of the vector nodal function $\vec{G}_x^j$, and subscript $y = 2$ indicates the eastward or ϕ-component.

The contribution to the coefficients of the linear system is $\Delta E_i = \Delta F_i = 0$ and

$$
\begin{aligned}
\Delta A_{ij} &= \frac{1}{A_0 R^2}\left\{\iint_a \frac{1}{\mu^2}\left[\begin{array}{l} 2\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{1,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{1,2}^j}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\phi}\frac{\partial G_{1,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}G_{1,1}^j + G_{1,1}^i\frac{\partial G_{1,1}^j}{\partial\theta}\right) + \\ 2\left(\csc\theta\frac{\partial G_{1,2}^i}{\partial\phi} + \frac{G_{1,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{1,2}^j}{\partial\phi} + \frac{G_{1,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{1,1}^i}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\theta} - \frac{G_{1,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{1,1}^j}{\partial\phi} + \frac{\partial G_{1,2}^j}{\partial\theta} - \frac{G_{1,2}^j}{\tan\theta}\right) \end{array}\right] da\right\} \\
\Delta B_{ij} &= \frac{1}{A_0 R^2}\left\{\iint_a \frac{1}{\mu^2}\left[\begin{array}{l} 2\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{2,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\phi}\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}G_{2,1}^j + G_{1,1}^i\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \\ 2\left(\csc\theta\frac{\partial G_{1,2}^i}{\partial\phi} + \frac{G_{1,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{G_{2,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{1,1}^i}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\theta} - \frac{G_{1,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,1}^j}{\partial\phi} + \frac{\partial G_{2,2}^j}{\partial\theta} - \frac{G_{2,2}^j}{\tan\theta}\right) \end{array}\right] da\right\} \\
\Delta D_{ij} &= \frac{1}{A_0 R^2}\left\{\iint_a \frac{1}{\mu^2}\left[\begin{array}{l} 2\frac{\partial G_{2,1}^i}{\partial\theta}\frac{\partial G_{2,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{2,1}^i}{\partial\theta}\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{\partial G_{2,2}^i}{\partial\phi}\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{2,1}^i}{\partial\theta}G_{2,1}^j + G_{2,1}^i\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \\ 2\left(\csc\theta\frac{\partial G_{2,2}^i}{\partial\phi} + \frac{G_{2,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{G_{2,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{2,1}^i}{\partial\phi} + \frac{\partial G_{2,2}^i}{\partial\theta} - \frac{G_{2,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,1}^j}{\partial\phi} + \frac{\partial G_{2,2}^j}{\partial\theta} - \frac{G_{2,2}^j}{\tan\theta}\right) \end{array}\right] da\right\}
\end{aligned} \quad (10)
$$

In practice, area integrals are treated as the sum of integrals over individual (unfaulted) finite elements; within each element, integrals are performed numerically using 7 Gauss points with associated weights [*Zienkiewicz*, 1977]. Continuum strain-rate will also be minimized in elements which contain faults, but this will be part of a different algorithm, described below.

Use of Stress Directions: the Isotropy Constraint

One principal stress direction must always be perpendicular to the free surface of a planet, or approximately vertical. Thus, the orientation of the stress tensor is well described by the azimuth (γ; measured clockwise from North) of the most-compressive horizontal principal stress ($\hat{\sigma}_{1\mathrm{h}}$). These directions are tabulated in data sets such as the World Stress Map.

Unfortunately, these data are very noisy. Variance in stress direction does not approach zero as pairs of data points are selected closer and closer together. Another problem is that the uncertainties assigned to individual directions are mostly generic estimates, not the result of repeated measurements at one point. A third problem is that there are spatial gaps in the data sets, such that many finite elements in a fine grid won't contain any data. To handle all these problems, we first interpolate observed stress directions to the center of each finite element, using an algorithm by *Bird & Li* [1996]. Specifically, we use the algorithm variant with pre-averaging of clustered data. This algorithm provides an estimate, $\delta\gamma$, of the standard deviation (in radians) of the azimuth γ of the interpolated direction $\hat{\sigma}_{1\mathrm{h}}$. The uncertainties $\delta\gamma$ from this pre-clustering algorithm are larger than those from the independent-data variant of the algorithm, but we believe these larger uncertainties to be more realistic.

To use this information about stress in *NeoKinema* models, we approximate the lithosphere as horizontally isotropic, so that the principal directions of the strain rate tensor in unfaulted continuum elements should be the same as the principal directions of stress. There may be an error of up to 35° associated with this assumption if and where the lithosphere contains unrecognized weak faults. Even so, the solutions will be more accurate and reasonable than ones which ignore stress data and leave the orientations of continuum strain rates completely unconstrained. (Unconstrained models often show sinistral simple-shear straining adjacent to dextral strike-slip faults, and extensional continuum straining adjacent to thrust faults. Such local reversals of stress are implausible and should be suppressed for a realistic simulation.)

Once we know the azimuth of $\hat{\sigma}_{1\mathrm{h}}$, we use this as the direction of a new local horizontal axis $\hat{\alpha}$, and also define a perpendicular horizontal axis $\hat{\beta}$ (right-handed: $\hat{\alpha} \times \hat{\beta} = \hat{r}$). In these coordinates, the requirement that $\hat{\alpha}$ is the most-compressive horizontal principal strain-rate direction can be stated in two equations: $\dot{\varepsilon}_{\alpha\beta} = 0$ and $\dot{\varepsilon}_{\alpha\alpha} < \dot{\varepsilon}_{\beta\beta}$. In terms of the global coordinate system, the former constraint, which is the $n = 2$ constraint in (3), becomes

$$\dot{\varepsilon}_{\theta\phi} \cos(2\gamma) + \frac{\dot{\varepsilon}_{\theta\theta} - \dot{\varepsilon}_{\phi\phi}}{2} \sin(2\gamma) = 0 . \tag{11}$$

In terms of derivatives of velocity, this is

$$\frac{1}{2R}\left\{\left(\csc\theta\frac{\partial v}{\partial\phi} + \frac{\partial w}{\partial\theta} - \frac{w}{\tan\theta}\right)\cos(2\gamma) + \left(\frac{\partial v}{\partial\theta} - \csc\theta\frac{\partial w}{\partial\phi} - \frac{v}{\tan\theta}\right)\sin(2\gamma)\right\} = 0 , \tag{12}$$

so the coefficients of the linear system can be computed from the factors

$$\begin{aligned} f_{2j} &= \frac{1}{2R}\left\{\left(\csc\theta\frac{\partial G_{1,1}^{j}}{\partial\phi} + \frac{\partial G_{1,2}^{j}}{\partial\theta} - \frac{G_{1,2}^{j}}{\tan\theta}\right)\cos(2\gamma) + \left(\frac{\partial G_{1,1}^{j}}{\partial\theta} - \csc\theta\frac{\partial G_{1,2}^{j}}{\partial\phi} - \frac{G_{1,1}^{j}}{\tan\theta}\right)\sin(2\gamma)\right\} \\ g_{2j} &= \frac{1}{2R}\left\{\left(\csc\theta\frac{\partial G_{2,1}^{j}}{\partial\phi} + \frac{\partial G_{2,2}^{j}}{\partial\theta} - \frac{G_{2,2}^{j}}{\tan\theta}\right)\cos(2\gamma) + \left(\frac{\partial G_{2,1}^{j}}{\partial\theta} - \csc\theta\frac{\partial G_{2,2}^{j}}{\partial\phi} - \frac{G_{2,1}^{j}}{\tan\theta}\right)\sin(2\gamma)\right\} \end{aligned} \tag{13}$$

if we use $c_2 = 0$ and a target rate $r_2 = 0$. It is necessary to decide what standard deviation σ_2 to associate with this constraint $\dot{\varepsilon}_{\alpha\beta} = 0$, since we have transformed the constraint from one concerning an angle to one concerning a strain-rate component. When the calculation is first started and no strain-rates are known, a purely arbitrary small strain rate uncertainty (ξ) must be assigned as σ_k. However, once strain rate estimates are available from the previous iteration of the solution, it is better to use

$$\sigma_2 = 2(\delta\gamma)\sqrt{\dot{\varepsilon}_{\theta\phi}^2 + \frac{1}{4}\left(\dot{\varepsilon}_{\theta\theta} - \dot{\varepsilon}_{\phi\phi}\right)^2} . \quad (14)$$

This requires that the velocity solution be iterated.

The latter requirement was the inequality $\dot{\varepsilon}_{\alpha\alpha} < \dot{\varepsilon}_{\beta\beta}$. During the later iterations of the solution, *NeoKinema* evaluates the strain rates $\dot{\varepsilon}_{\alpha\alpha}$ and $\dot{\varepsilon}_{\beta\beta}$ to see if this is true. If not, then in future iterations *NeoKinema* imposes an additional continuum constraint, $n = 3$ in (3), that $\dot{\varepsilon}_{\beta\beta} = \dot{\varepsilon}_{\alpha\alpha} + \xi$, where ξ is a small (positive) strain rate difference which must be chosen as an input parameter. In terms of the global coordinates, this becomes

$$\left(\dot{\varepsilon}_{\phi\phi} - \dot{\varepsilon}_{\theta\theta}\right)\cos(2\gamma) + 2\dot{\varepsilon}_{\theta\phi}\sin(2\gamma) = \xi . \quad (15)$$

This can be expressed in terms of velocity components as

$$\frac{1}{R}\left\{\left(\csc\theta\frac{\partial w}{\partial\phi} + \frac{v}{\tan\theta} - \frac{\partial v}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial v}{\partial\phi} + \frac{\partial w}{\partial\theta} - \frac{w}{\tan\theta}\right)\sin(2\gamma)\right\} = \xi , \quad (16)$$

so the coefficients of the linear system can be computed from the factors

$$\begin{aligned} f_{3j} &= \frac{1}{R}\left\{\left(\csc\theta\frac{\partial G_{1,2}^j}{\partial\phi} + \frac{G_{1,1}^j}{\tan\theta} - \frac{\partial G_{1,1}^j}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial G_{1,1}^j}{\partial\phi} + \frac{\partial G_{1,2}^j}{\partial\theta} - \frac{G_{1,2}^j}{\tan\theta}\right)\sin(2\gamma)\right\} \\ g_{3j} &= \frac{1}{R}\left\{\left(\csc\theta\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{G_{2,1}^j}{\tan\theta} - \frac{\partial G_{2,1}^j}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial G_{2,1}^j}{\partial\phi} + \frac{\partial G_{2,2}^j}{\partial\theta} - \frac{G_{2,2}^j}{\tan\theta}\right)\sin(2\gamma)\right\} \end{aligned} \quad (17)$$

and $c_3 = 0$ if we create the target rate $r_3 = \xi$. The same value of ξ is used to set the standard deviation for this constraint as $\sigma_3 = (0.83)\xi$, so that the implied Gaussian distribution of ($\dot{\varepsilon}_{\beta\beta} - \dot{\varepsilon}_{\alpha\alpha}$) (required by our weighted least-squares method) will approximate the desired Heaviside distribution for small positive values.

In cases where stress-direction data are very sparse, it may be desirable or necessary to use active fault segments as additional stress-direction indicators, assuming $\hat{\sigma}_{1h}$ perpendicular to thrusts, *etc*. Such an option has been provided in *NeoKinema*, but it should be used with caution.

The difficulty is that only the first phase of movement on a fault should be used to indicate stress, because in later tectonic phases the fault remains an inherited plane of weakness, even though stress fields may rotate. Yet, one can rarely be certain that all the faults in a given problem area are new.

Use of Fault Slip Rate Data

NeoKinema solves for only the horizontal components of velocity at the surface, so a fault is treated as a surface discontinuity in horizontal velocity. The offset-rate parameter of greatest interest is the heave rate, which is the horizontal component of the slip rate. For convenience, and to reduce errors, 7 fault types have been predefined, so that all fault offset rates can be entered with positive numbers (and in conventional units of mm/a). For the first 5 fault types, the heave rate is directly specified. Types R and L (for Right-lateral and Left-lateral, respectively) have heave-rate vectors parallel to the fault trace. Type D (for Divergent or Detachment) has heave-rate at right angles to the trace, with an opening or spreading sense, and is used to describe mid-ocean spreading ridges, low-angle detachment faults, convex-upward "rolling-hinge" detachment faults, concave-upward listric normal faults, and rotating sets of planar "bookshelf" normal faults. Type P (for thrust Plate, or naPpe) has heave-rate at right angles to the trace, with a shortening or convergent sense. Type S (for Subduction) is treated the same as type P within *NeoKinema*, but the special fault type S is passed to output files for other programs (such as *Long_Term_Seismicity*, which treats subduction zones differently from other convergent boundaries). For the 2 remaining fault types, the throw rate (vertical component of slip rate) is entered: type T represents planar Thrusts, and type N represents planar Normal faults. For these last 2 types, a fault dip must be assumed so that *NeoKinema* can convert throw rates to heave rates. The dip angles currently programmed are 20° for Thrusts, and 55° for Normal faults, which are consistent with the dips assumed in the seismicity calibration study of *Bird & Kagan* [2004]. (Other dips could be used, but there would be risk of confusion and inconsistency if *NeoKinema* output were used as input to *Long_Term_Seismicity*, which assumes the *Bird & Kagan* dip values.)

Geologic studies of offset surfaces and/or offset piercing points, supported by radiometric or stratigraphic dating, establish long-term average offset rates (slip rates and/or throw rates and/or heave rates) for many faults. In a few cases, these rates are determined in a specific small region, and could be treated as point constraints with known Gaussian probability density functions. However, most cases are more difficult: (a) Some offsets are so large that the rate has to be interpreted as the mean rate between the offset piercing points, rather than the rate at a point. (b) Many studies establish only upper and/or lower bounds on the slip rate, not a preferred value. Then, rate constraints from different locations have to be merged to estimate a preferred slip rate, and its residual uncertainty, for the fault as a whole. (c) Rates determined over time windows of less than 10^4 years, or more than 10^6 years, must be treated with caution, as rates are expected to be less stable outside this time window. (d) Many slip-rates quoted in the literature are second- or third-hand restatements of tentative rates that have not been peer-reviewed; these must also be treated with caution. (e) Some authors who compiled offset rates have assigned uncertainties to be a fixed fraction of the estimated slip rate; such uncertainties are often seriously underestimated and in need of revision.

For all these reasons, we decided that users of *NeoKinema* should merge available offset rate information for each fault, by the editorial process of their choice, and then input only one preferred rate (r_m) and the standard deviation (σ_m) best appromating the actual PDF, for each component (parallel and perpendicular) of the heave rate of each fault. If firm lower and upper bounds ($r^{(l)}$ and $r^{(u)}$, respectively) are available on the rate, then $r_m = \left(r^{(l)} + r^{(u)}\right)/2$ and $\sigma_m = \left(r^{(u)} - r^{(l)}\right)/4$ might be reasonable choices. If no information is available, one can set $r_m = 0$ and $\sigma_m \to \infty$, which leaves that fault free to slip in any way to optimizes the fit to other types of data. Long faults (like the San Andreas) which have multiple intersections with other faults are best treated by dividing them into sections that have distinct rates and uncertainties.

Fault slip rate is, in general, a two-component vector. If both the dip-slip and the strike-slip components of the slip rate are known, *NeoKinema* treats these as two distinct scalar constraints along the same fault trace. When only the dip-slip rate is known, *NeoKinema* provides an option to permit limited strike-slip in proportion to the amount of dip-slip. (This is useful because otherwise a thrust fault with a complex trace could not slip without deforming its hanging wall and/or footwall, and such deformation would be strongly resisted by the continuum stiffness constraint discussed above.) There is no corresponding provision for limited dip-slip on known strike-slip faults, because strike-slip faults are modeled as vertically-dipping, and thus any dip-slip component would not affect the horizontal velocity components estimated by *NeoKinema*.

When a fault is long enough to cross several finite elements, *NeoKinema* attempts to impose the same offset rate in all elements. In the case of rigid-microplate tectonics, where each fault connects to other faults at triple-junctions, this method is reasonably accurate. (The only difficulty occurs where there are rapid relative rotations of adjacent microplates, but this can be handled by segmenting the faults and varying the target rates along the strike of each fault.) The other end-member is the case where no faults connect, but all terminate within the domain. In that case, each fault might be expected (on the basis of crack theory for linear materials) to have an ellipsoidal profile of slip rate versus length. Such "elliptical" faults would have a mean offset rate which is only 79% ($\pi/4$) of their maximum offset rate. Thus, *NeoKinema* might overstate fault-related strain-rates by 27% in some cases where faults do not connect and where the geologic offset rates reported are all maxima along their respective traces. However, if the geologic offset rates were determined at random points of convenience, then once again no systematic error is expected.

The simplest way to impose fault slip rates would be to use each offset rate as a constraint on the relative velocities of adjacent nodes on opposite sides of the fault. This approach would require a finite element grid that conforms to all fault traces, providing matched nodes on opposite sides of each fault, and triple nodes at fault intersections. However, the number of faults in many applications is so great that such customized grids are very time-consuming to prepare; they may also require unreasonable amounts of computer time and memory to solve. Thus, we have developed a more general approach, loosely based on the substructuring method from engineering finite elements. Our new method allows any number of faults to cross a given finite element.

For each finite element containing one or more fault traces, there are four steps: (a) Form the target strain-rate tensor for that element as the sum of the strain-rate tensors implied by all the active fault segments cutting that element; (b) Form the matrix of covariances of the strain-rate components in that element as the sum of the covariances added by all the active fault segments, plus the small covariance of the strain-rate in the continuum around them; (c) Diagonalize the covariance matrix to find its three principal axes (in strain-rate space) along which the uncertainties are independent, and rotate the target strain-rates into this new coordinate system; (d) Add these 3 independent targets as scalar data with known uncertainties in the global system of equations.

The strain-rate tensor in the horizontal plane, $\dot{\tilde{\varepsilon}}$, is a second-rank tensor of size 2×2. We simplify the notation by treating the three independent components of the strain-rate tensor ($\dot{\varepsilon}_{\theta\theta} = \dot{\varepsilon}_{NS}, \dot{\varepsilon}_{\theta\phi} = \dot{\varepsilon}_{SE}, \dot{\varepsilon}_{\phi\phi} = \dot{\varepsilon}_{EW}$) as a one-subscript vector ($\dot{\varepsilon}_q;\ q = 1,2,3$), permitting us to write the covariance of strain-rates as a 3×3 matrix. If all the active fault segments that cut (even part-way) through one finite element are numbered $z = 1,\ldots,Z$, then we express the target strain-rate vector in the element as a linear combination of their scalar slip-rates s_z:

$$\dot{\varepsilon}_q = \sum_{z=1}^{Z} H_{zq}\, s_z; \qquad q = 1,2,3. \tag{18}$$

The covariance matrix of the strain-rate components is composed of two parts: the continuum compliance common to all parts of the lithosphere (see "Continuum Stiffness: the Microplate Constraint"), and the terms arising from the standard deviations δs_z of the scalar slip-rates s_z:

$$\tilde{V} = \mu^2 \begin{bmatrix} 4/3 & 0 & -2/3 \\ 0 & 1 & 0 \\ -2/3 & 0 & 4/3 \end{bmatrix} + \sum_{z=1}^{Z} \left(\delta s_z\right)^2 \left[\vec{H}_z^{\mathrm{T}} \vec{H}_z\right] \tag{19}$$

To find $\vec{H}_z$ (the partial derivative of element strain-rate with respect to slip-rate of one active fault), we impose a rule that no node may lie exactly on a fault. Also, we incrementally straighten the trace of any fault that crosses the same element boundary more than once, until the number of crossings is reduced to 1 or 0. Then, each fault segment (with its projected extensions, if necessary) must separate one node of the element from the other two. Let u_z be the index number of the isolated node. If node u_z is on the right side of the fault segment (when looking along its azimuth γ_z, measured clockwise from North), then we define the variable η_z as +1; otherwise, it is -1. Let κ_z be the fraction of the width of the element that is cut by the fault segment: $0 < \kappa_z \leq 1$.

In the case of a strike-slip fault, the scalar-slip rate s_z is defined as the right-lateral heave rate. (Left-lateral rates are negative right-lateral rates.) Then

$$\vec{H}_z = \frac{\eta_z \kappa_z}{R} \begin{pmatrix} \dfrac{\partial G_{1,1}^{u_z}}{\partial \theta} \cos\gamma_z - \dfrac{\partial G_{2,1}^{u_z}}{\partial \theta} \sin\gamma_z , \\ \dfrac{1}{2}\left(\dfrac{\partial G_{1,1}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} - \dfrac{\partial G_{2,1}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{\partial G_{1,2}^{u_z}}{\partial \theta} \cos\gamma_z - \dfrac{\partial G_{2,2}^{u_z}}{\partial \theta} \sin\gamma_z - \dfrac{G_{1,2}^{u_z} \cos\gamma_z - G_{2,2}^{u_z} \sin\gamma_z}{\tan\theta} \right), \\ \dfrac{\partial G_{1,2}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} - \dfrac{\partial G_{2,2}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{G_{1,1}^{u_z} \cos\gamma_z - G_{2,1}^{u_z} \sin\gamma_z}{\tan\theta} \end{pmatrix} . \quad (20a)$$

In the case of dip-slip faulting, it is most convenient to define s_z as the net horizontal extension rate perpendicular to the fault trace. (Thrusting is considered to be negative extension.) In the case of detachment faulting, net horizontal extension is the distance from the breakaway fault in the foot-wall to the tip of the hanging-wall (reconstructed if necessary), regardless of whether the fault slipped at a low angle or, alternatively, slipped at a high angle and then rotated during further extension. In the more common case of dip-slip faulting without horizontal-axis rotation of foot-wall or hanging-wall, net horizontal extension rate is the relative vertical offset (throw) rate times the cotangent of the fault dip. Our convention is that normal and detachment faulting have positive s_z and thrust faults have negative values. Then,

$$\vec{H}_z = \frac{\eta_z \kappa_z}{R} \begin{pmatrix} \frac{\partial G_{1,1}^{u_z}}{\partial \theta} \sin\gamma_z + \frac{\partial G_{2,1}^{u_z}}{\partial \theta} \cos\gamma_z, \\ \frac{1}{2}\left(\frac{\partial G_{1,1}^{u_z}}{\partial \phi} \frac{\sin\gamma_z}{\sin\theta} + \frac{\partial G_{2,1}^{u_z}}{\partial \phi} \frac{\cos\gamma_z}{\sin\theta} + \frac{\partial G_{1,2}^{u_z}}{\partial \theta} \sin\gamma_z + \frac{\partial G_{2,2}^{u_z}}{\partial \theta} \cos\gamma_z - \frac{G_{1,2}^{u_z} \sin\gamma_z + G_{2,2}^{u_z} \cos\gamma_z}{\tan\theta} \right), \\ \frac{\partial G_{1,2}^{u_z}}{\partial \phi} \frac{\sin\gamma_z}{\sin\theta} + \frac{\partial G_{2,2}^{u_z}}{\partial \phi} \frac{\cos\gamma_z}{\sin\theta} + \frac{G_{1,1}^{u_z} \sin\gamma_z + G_{2,1}^{u_z} \cos\gamma_z}{\tan\theta} \end{pmatrix}. \quad (20b)$$

The next step is to find the 3 positive eigenvalues (λ_h; $h = 1,2,3$) of $\widetilde{V}$ and their corresponding unit eigenvectors (Λ_{hq}). These eigenvectors indicate strain-rate patterns whose uncertainties are uncorrelated and independent; they have target amplitudes of $r_h = \sum_{q=1}^{3} \dot{\varepsilon}_q \Lambda_{hq}$ and standard deviations of $\sigma_h = \sqrt{\lambda_h}$, respectively. Each of the three targets is now imposed as a scalar datum in the global system of equations. The corresponding coefficients of the nodal velocities are

$$\begin{aligned} f_{mj} &= \frac{1}{R}\left[\frac{\partial G_{1,1}^j}{\partial \theta},\quad \frac{1}{2}\left(\csc\theta \frac{\partial G_{1,1}^j}{\partial \phi} + \frac{\partial G_{1,2}^j}{\partial \theta} - \frac{G_{1,2}^j}{\tan\theta} \right),\quad \csc\theta \frac{\partial G_{1,2}^j}{\partial \phi} + \frac{G_{1,1}^j}{\tan\theta} \right] \begin{bmatrix} \Lambda_{h1} \\ \Lambda_{h2} \\ \Lambda_{h3} \end{bmatrix} \\ g_{mj} &= \frac{1}{R}\left[\frac{\partial G_{2,1}^j}{\partial \theta},\quad \frac{1}{2}\left(\csc\theta \frac{\partial G_{2,1}^j}{\partial \phi} + \frac{\partial G_{2,2}^j}{\partial \theta} - \frac{G_{2,2}^j}{\tan\theta} \right),\quad \csc\theta \frac{\partial G_{2,2}^j}{\partial \phi} + \frac{G_{2,1}^j}{\tan\theta} \right] \begin{bmatrix} \Lambda_{h1} \\ \Lambda_{h2} \\ \Lambda_{h3} \end{bmatrix} \end{aligned}. \quad (21)$$

(Note that *m* now equals *h* plus an integer that counts how many other fault-related target rates have previously been incorporated into the linear system.)

Now, the substructuring method in engineering finite elements (our guiding metaphor) involves three steps: (1) Condense the stiffness of the substructure into a simpler element that can represent it; (2) Compute the global solution; and (3) Perfom a local solution to distribute displacements and strains within the substructure. Although our method is kinematic rather than dynamic, there are close parallels. Above, we described how the target strain rates (and their uncertainties) from an arbitrary number of faults are reduced to an equivalent 3-DOF model. Below, we show how a local maximum-likelihood solution distributes the total strain rate into its component parts.

Once the global velocity solution has been found, *NeoKinema* performs a local optimization calculation within each faulting element to find the *predicted* (model) rates p_z at which each fault ($z = 1, \ldots, Z$) is slipping, as well as the residual strain-rate $\dot{\varepsilon}_q^{\mathrm{c}}$ which is due to deformation of the continuum around the faults. (In the enumeration of *Z*, a fault with both dip-slip and strike-slip components is considered as "two faults" that happen to have the same trace.) The total strain-rate of the element must be the sum of the continuum and the fault contributions:

$$\dot{\varepsilon}_q^{\mathrm{c}} + \sum_{z=1}^{Z} H_{zq} p_z = \dot{\varepsilon}_q \,. \tag{23}$$

This problem is different from the global problem because the $\dot{\varepsilon}_q$ vector is now known. Because of this constraint, it is convenient to use the Lagrange multiplier method with three temporary weight variables (ς_1, ς_2, ς_3). We define the local objective function (in one element) that is to be optimized as:

$$S''' \equiv -\sum_{z=1}^{Z} \frac{L_z}{L_0} \frac{\left(p_z - s_z\right)^2}{(\delta s_z)^2} - \frac{A}{A_0} \frac{\left(\dot{\varepsilon}_1^{\mathrm{c}2} + \dot{\varepsilon}_1^{\mathrm{c}} \dot{\varepsilon}_3^{\mathrm{c}} + \dot{\varepsilon}_3^{\mathrm{c}2} + \dot{\varepsilon}_2^{\mathrm{c}2}\right)}{\mu^2} - \sum_{q=1}^{3} \varsigma_q \left(\dot{\varepsilon}_q^{\mathrm{c}} + \sum_{z=1}^{Z} H_{zq} p_z - \dot{\varepsilon}_q \right) \tag{24}$$

where L_z is the length of each fault segment, δs_z is the standard deviation of its offset rate (according to the input data), and *A* is the area of the element. To find a local solution that has all fault rates as close as possible to their goals, while the continuum strain-rate is close to zero, and the total strain-rate is correct, we find the stationary point of S''' with respect to variations in the p_z, the $\dot{\varepsilon}_m^{\mathrm{c}}$, and the ς_m in turn, leading to a small linear system of equations with a positive-definite coefficient matrix.

Once all local substructure solutions are completed, an average offset rate for each fault is also computed, as the average of the rates p_z in all the elements the fault passes through, with averaging weights proportional to the segment lengths. Both measures of predicted fault offset rate are saved to files, and our plotting software (*NeoKineMap*) can display either the average or the element-specific model offset rate components.

Use of Geodetic Data

Equations (1) through (3) already provide for the incorporation of geodetic velocity components at benchmarks, but only in certain ideal cases. Three practical difficulties often arise: (a) The two velocity components at one benchmark and/or the velocities at different benchmarks have correlated uncertainties. (b) The relation between the velocity reference frame for the geodetic velocities and that of the velocity boundary conditions may be uncertain. (This occurs when all, or almost all, of the benchmarks used in the geodetic velocity solution are located within an orogen, and few or none are outside the orogen on rigid plates.) (c) Geodetic velocities at benchmarks near active faults do not represent long-term average velocities because the faults remain locked, or else suddenly slip by large amounts, during the period of observation.

Correlated uncertainties in geodetic velocity components (problem a) violate the assumption of independence used to obtain the simple objective function in (2). Therefore, coordinates must be rotated to new variable space of the same dimensionality, in which the uncertainties are independent, and prediction errors should be evaluated in those new coordinates. It is well-known that (2) should be replaced by:

$$S' \equiv -\frac{1}{2}\sum_{j=1}^{K}\sum_{k=1}^{K}\left(p_k - r_k\right)N_{jk}\left(p_j - r_j\right) \tag{25}$$

where the "normal matrix" $\tilde{N}$ is the inverse of the covariance matrix $\tilde{C}$ of the observed velocity components $\vec{r}$. Since $\tilde{C}$ is a positive-definite matrix, and (25) is still a quadratic form like (2), this presents no problems except for expansions of algebra, computational effort, and computer memory. We see from (4) above that model predictions $\vec{p}$ are each sums of 6 terms (concerning the 2 horizontal velocity components at each of the 3 nodes in the finite element surrounding the benchmark). After extensive algebra based on (25) (omitted here) these nodal degrees of freedom become linked to those of any other finite element containing a geodetic benchmark (whose uncertainties are correlated with those of the first). If the correlations are only local (*e.g.,* between the N-S and E-W velocity components at each benchmark) then $\tilde{C}$ is block-diagonal, and $\tilde{N}$ is block-diagonal, and the linear system (6) of *NeoKinema* retains any banded nature that may have been achieved by the intelligent ordering of velocity degrees of freedom. However, if $\tilde{C}$ is a general full matrix, then the resulting linear system is no longer banded, and both computer memory and solution time burdens increase. (In this case, we see the important benefit of working in only 2 space dimensions on the surface of the planet, where 5,000 to 10,000 nodes, or 10,000 to 20,000 degrees of freedom, still permit a reasonably fine grid of nodes and elements.)

If the velocity reference frame of the geodetic data is unclear (problem b above), then this is handled in *NeoKinema* by adding 3 large eigenvalues to $\tilde{C}$, corresponding to eigenvectors representing uniform steady rotations of the entire geodetic network around each of 3 orthogonal axes through the center of the planet. In the normal matrix $\tilde{N}$, the corresponding (inverted) eigenvalues become nearly zero, mapping any systematic prediction error in the velocity reference frame to infinitesimal contributions to the objective functions S' and S''. (Note that, in this case, the specification of velocity boundary conditions at the model edges is mandatory.) However, the condition number of $\tilde{C}$ should not be made too large, or the quality of its numerical inverse $\tilde{N}$ will suffer. In the current version of *NeoKinema* we add rotational eigenvalues with magnitude 10°/Ma to $\tilde{C}$, and use 64-bit arithmetic in the inversion to obtain $\tilde{N}$.

The third problem (c above) is geodetic benchmarks close to faults, at which the observed velocity is not the long-term average. Before using these velocities as constraints in the global optimization, *NeoKinema* corrects them to estimated long-term average velocities by adding the estimated long-term average rates of coseismic displacement due to all faults in the model. These coseismic displacements are computed using analytic solutions by *Mansinha & Smylie* [1967, 1971] for the effects of rectangular patches of uniform dislocation in a uniform elastic halfspace. We assume uniform Poisson's ratio of 0.25 in the halfspace. We fix uniform upper (shallow) and lower (deep) extents of the dislocation patches with input parameters, for example 1 km and 12 km, respectively, for most crustal faults, but 14 km and 40 km, respectively, for subduction zones [*Bird & Kagan,* 2004]. We provide two alternate methods for determining the long-term average slip rate of the faults (for purposes of correcting the geodetic velocities): a conservative "geologic" estimate using the data set of fault slip rates which are the targets p_m for *NeoKinema*, or a "self-consistent" estimate in which fault slip rates are taken from the previous iteration of the solution. The self-consistent method is preferred unless it leads to an instability.

If any fault is creeping steadily (such as the central segment of the San Andreas fault in California), this fault is flagged with a logical switch on input, and corrections to geodetic velocities at benchmarks will not include any coseismic contribution from that fault. We do not currently have any algorithm to handle intermediate cases of combined fault creep and coseismic offset on the same fault.

The underlying assumption of this correction method is that no important earthquakes have occurred during the time window of geodetic data collection. If they have, then we prefer to edit out (exclude) geodetic velocities affected by these earthquakes. This can be done by redetermining the velocities using only preseismic observations, or by simply omitting such benchmarks.

Experience has shown that it is also necessary to exclude benchmarks very close to fast-moving faults, regardless of their seismic history, for three reasons. First, errors in digitizing fault traces occasionally cause a benchmark to be placed (in the virtual world of the *NeoKinema* model) on the wrong side of a fault. This causes an entirely spurious prediction error equal to 100% of the slip rate, which will systematically bias any weighted-least-squares algorithm. Second, if the benchmark is located closer to the fault trace than the nearest nodes, the interpolated model velocity p_k at that benchmark will be skewed by nodal-function interpolation toward the velocity of the adjacent block; such errors can be almost as large. Third, we have found that the *Mansinha & Smylie* dislocation solutions are ill-conditioned (for Fortran-based evaluation) at points very close to the edges of the dislocation patches, and give noisy corrections. For all these reasons, we apply automated deletion of benchmarks less than 2 km from fast-moving faults as a data pre-processing step. Consistent with this guideline, we recommend hand-editing of the finite element grid to create "corridors" of narrow elements along fast-moving faults, with width not more than 4 km. (However, this editing is not necessary along minor faults whose slip rates are expected to be comparable to, or less than, the uncertainties in benchmark velocities.)

Iterative Improvement

At 4 points in the algorithm described above, we referred to the use of model estimates from a previous iteration to improve the solution. All of these iterations are combined and performed simultaneously, with 10 to 40 iterations per run. Here we discuss some details of this iteration which can affect the stability and precision of *NeoKinema* solutions if not properly handled.

While the interpolation of principal stress directions to all finite elements is only performed once, the conversion of the resulting azimuth uncertainty $\delta\gamma$ to the strain rate uncertainty σ_2 in equation (14) requires knowledge of the strain rate, and is iterated. If $\delta\gamma$ is small (because the stress direction is well-known) and the strain rate also becomes small in the same element, then σ_2 can become very small, leading to unreasonably large eigenvalues and an undesirably large condition number in the global linear system (6). This causes random numerical noise to propagate through the solver to the inferred velocities of all nodes. To keep such random numerical fluctuations down to an acceptable level, we limit the eigenvalues of the linear system on the high side by arbitrarily reducing the weight on the $n = 2$ ($\dot{\varepsilon}_{\alpha\beta} = 0$) constraint at such elements. This reduction in weighting occurs when $\sigma_2 < \xi$.

The other component of enforcing principle strain rate axes for unfaulted continuum elements is to check whether the sense of strain rate is correct. Equations (15)-(17) are employed only in cases of incorrect sense (such as N-S extension where there should be N-S compression). Experience shows, however, that a bounded oscillation of period 2 ("recidivism") often occurs if the constraint is removed after an iteration resulting in the desired sense of strain rate. Therefore, *NeoKinema* never removes this constraint from any element where it has once been imposed. To avoid applying the constraint to more elements than necessary, the program postpones adding this constraint at any element until the latter half of the set of planned iterations. This allows other iterated components of the solution to stabilize before this irreversible change is imposed.

In the section on use of fault slip rate data, it was noted that an option is provided to allow limited amounts of strike-slip on nominally dip-slip faults for which no strike-slip component data are available. The assumed uncertainty in strike-slip rate is proportional to the current dip-slip rate, so that the slip vector of the nominally dip-slip fault is confined to lie within a specified angle about the direction normal to the trace. The feedback between dip-slip and strike-slip during iteration can lead to a bounded oscillation of period 2 for certain fault geometries. To suppress this, the update of the uncertainty in strike-slip rate is slightly damped, by always averaging the new uncertainty with the previous uncertainty.

Finally, it was mentioned that the correction of observed geodetic velocities to estimated long-term average velocities requires knowledge of fault slip rates. Where the velocity of a geodetic benchmark is unusually sensitive to the slip rates of adjacent faults (*e.g.,* a benchmark above two shallow-dipping conjugate thrust faults) this feedback during iteration can combine with errors in the elastic dislocation correction to cause an unbounded, exponentially-growing instability. *NeoKinema* provides two alternative remedies: If the "self-consistent" current fault slip rates are used for the geodetic corrections, then changes in those fault rates are slightly damped (for this purpose only) by averaging each new rate with the previous rate. In case this might not be sufficient for stability, another option is provided to use the "conservative" adjustment of geodetic velocities which is based on input fault slip rate targets, rather than current model rates.

With these precautions, *NeoKinema* solutions typically converge to RMS velocity changes (between adjacent iterations) of 10^{-3} of the overall RMS velocity, or better. This is acceptable for purposes of seismic hazard estimation. We believe that most of the residual noise is due to numerical error in the solution of linear systems, which potentially could be further reduced with 64-bit arithmetic, or tighter constraints on the range of eigenvalues of the coefficient matrix in the linear system.

Idealized Test Cases Performed

Many finite element grids used for testing overlap the North pole and the international date line. Codes which contain algebraic errors in the formulas for nodal functions, strain rates, or coefficients of the linear system will often display irregularities in these regions, but none were seen in these tests. Note that there are some gaps in the numbering of the tests described here, because some tests (Test07, Test12), while successful, had limited value for displaying how *NeoKinema* works.

TEST01: No input data, other than a zero velocity imposed at two boundary nodes: The solution is static, as expected.

TEST02: No input data, other than boundary conditions of one fixed node, and one other boundary node that rotates around the first: The solution is rigid-body Eulerian rotation of the model domain on the sphere, with internal strain rates orders of magnitude less than the rotation rate.

TEST03: No input data, other than boundary velocities set along two opposite sides of a rectangular domain so as to enforce quasi-uniform extension. (Exactly uniform extension is not possible on a sphere.) Strain rates were quasi-uniform across the model domain, and horizontal shortening and vertical shortening were each half of the rate of horizontal extension, as expected for a uniform sheet of Newtonian-viscous material in "plane stress."

TEST04: Same grid and boundary conditions as in Test03. Stress-direction data divide the rectangular model domain into two provinces: one province with "expected" $\hat{\sigma}_{1h}$ perpendicular to boundary velocities, and an "anomalous" province with $\hat{\sigma}_{1h}$ parallel to boundary velocities. In this case, fitting the stress-direction data requires increasing the strain rates in some unfaulted continuum elements, so the result depends on μ. If μ is initially small (5×10^{-17} s^{-1}) compared to the mean overall strain rate enforced by the boundary conditions (5×10^{-16} s^{-1}), continuum errors are large (mean = 10.43 σ, RMS = 11.36 σ) and stress-direction errors are also large (mean = 1.36 σ, RMS = 2.01 σ), and the strain rate does not fit the specified stress directions well, because *NeoKinema* is trying hard not to exacerbate continuum errors which are already of order 10 σ. However, when the "bootstrap method" is used to reset μ to the empirical value (5×10^{-16} s^{-1}), continuum errors become far smaller as a direct consequence (mean = 1.01 σ, RMS = 1.25 σ) and stress errors are also improved (mean = 0.48 σ, RMS = 0.65 σ) as an indirect consequence, and the $\hat{\dot{\varepsilon}}_{1h}$ directions align much better with specified stress directions.

TEST05: Same grid, boundary conditions, and $\mu = 5\times10^{-16}$ s^{-1} as in Test04. No stress-direction data. Three faults of unknown slip rate ($\sigma_m = 10^4$ mm/a) make up a 3-segment plate boundary with divergent, strike-slip, and divergent fault types, respectively. The strike-slip fault is parallel to the boundary velocities, but the detachment faults strike at ~70° to the boundary velocities. The logical switch that allows faults to imply stress directions is off. The option for limited strike-slip on dip-slip faults is off. Results are quasi plate-like, except that one plate "tears away" from the boundary velocities and rotates rapidly (at the cost of internal deformation) so as to achieve exact, pure dip-slip on the detachment faults. This behavior shows the need for an allowance for limited strike-slip components on nominally dip-slip faults.

TEST06: Same as Test05, except that ±20° flexibility was added to the directions of the slip vectors of dip-slip faults (relative to the expected trace-normal direction). Results are more plate-like; internal deformation of plates is nearly eliminated, and the heave-rate plot shows substantial amounts of strike-slip on the divergent faults, added by the new option.

TEST08: A strike-slip fault following a small circle on the sphere, with unknown slip-rate, is driven indirectly by imposed velocity on one nearby node. Essentially rigid-plate rotation is the result. Only minor deformation occurs in the unfaulted continuum elements near the fault.

TEST09: Tests the option for "type-4" velocity boundary conditions, in which only the plate affinity of each boundary node is provided to *NeoKinema* as a 2-letter code, and *NeoKinema* computes the appropriate boundary velocity from Euler poles (using the table from *Bird*, 2003). This also resulted in the desired quasi-rigid-plate behavoir.

TEST10: First test incorporating geodetic data. Model domain fixed at only two boundary nodes. The logical switch that lets the geodetic velocity reference frame float is off. The geodetic data are highly artificial: approximately one benchmark per finite element, with velocities computed from a constant Euler vector with respect to the fixed boundary nodes. The result is that almost all of the model domain moves with the geodetic velocity field, except for the two corners which were pinned by velocity boundary conditions.

TEST11: Similar to Test10, except that 10°/Ma of velocity reference frame loosening is allowed. Result: All nodal velocities < 0.0004 mm/a, as expected.

TEST13: Tests the "unlocking" correction of geodetic velocities by addition of (estimated) long-term-average coseismic slip, using the self-consistent method. This test uses a very artificial case of a straight, N-S dextral fault which is locked down to 100 km during the period of geodetic data collection, but slipping at unknown long-term average rate. The boundary conditions enforce 50 mm/a of simple dextral shear on N-S planes. A single line of 34 geodetic benchmarks, spaced 35 km apart, is placed across the fault at right angles and the interseismic (locked-fault) velocities of these benchmarks are simulated with the inverse-tangent formula to create the artificial test data. The reference area $A_0 = 6.2 \times 10^9$ m^2, which is also the finite element size. Computed long-term average velocity, after correction of the geodetic velocities, has the form of rigid-plate motion, as expected.

TEST14: Two slightly "harder" variants of Test13: (a) Same test, but with inhomogeneous boundary conditions changed to type-0 (free) along one margin, so that geodesy is now the only data requiring movement, and program must “discover” the rigid-plate solution by iteration. It reaches dV/V = 0.001 in 10 iterations, with dV/V decreasing by a factor of 2 each time. (b) Same test as 14(a) above, after 60° rotation of the finite element grids and the geodetic data. Also successful; the rotation had no effect, as the equations are not orientation-dependent. This test and the previous test show that the vertical-fault, strike-slip dislocation code (based on *Mansinha & Smilie*, 1967) is working properly.

TEST15: It is harder to test the dip-slip dislocation code (based on *Mansinha & Smilie*, 1971) because we do not have a simple analytic solution to use in simulating the geodetic data. Instead, we performed a practical test over the Cascadia subduction zone, using actual geodetic data from the WUSC solution of *Bennett et al.* [1999], to see if the apparently transient velocities in the forearc could be "corrected" by *NeoKinema*. The solution converged to dV/V = 0.0002, and showed long-term average velocities in the Cascadia forearc reduced to order 3 mm/a, with a pattern of N-S shortening, which is believed to be correct because it agrees with stress-direction data (not input for this test). The NE-SW shortening that dominated the raw geodetic velocity solution was identified as an elastic transient and removed.

Note on Versions

This Appendix describes *NeoKinema* version 2.0 of December 2004. The earlier version 1 used a different, less transparent method to set the relative weights on geodetic, geologic, and continuum constraints in the objective function. These weights were pre-programmed and partially dependent on the sizes of finite elements; this defect has been corrected in version 2. (Version 1 was never published or distributed, and results based on version 1 have only been presented in the form of abstracts. We mention this distinction only because readers might otherwise assume that this Appendix describes version 1.)

References Cited

Bennett, R. A., J. L. Davis, and B. P. Wernicke (1999) Present-day pattern of Cordilleran deformation in the western United States, Geology, 27(4), 371-374.

Bird, P. (2003) An updated digital model of plate boundaries, Geochemistry Geophysics Geosystems, 4, 1027, doi:10.1029/2001GC000252.

Bird, P., and Y. Y. Kagan (2004) Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in 7 tectonic settings, Bull. Seismol. Soc. Am., 94(6), 2380-2399, plus electronic supplement.

Bird, P., and Y. Li (1996) Interpolation of principal stress directions by nonparametric statistics: Global maps with confidence limits, J. Geophys. Res., 101, 5435-5443.

Kong, X., and P. Bird (1995) SHELLS: A thin-plate program for modeling neotectonics of regional or global lithosphere with faults, J. Geophys. Res., 100, 22,129-22,131.

Mansinha, L., and D. E. Smylie (1967) Effects of earthquakes on the Chandler wobble and the secular polar shift, J. Geophys. Res., 72, 4731.

Mansinha, L., and D. E. Smylie (1971) The displacement fields of inclined faults, Bull. Seismol. Soc. Am., 61, 1433-1440.a

Zienkiewicz, O. C. (1977), The Finite Element Method, McGraw-Hill, New York.