Finite-element propagation of acoustic waves on a spherical shell

Chuck Sword, Jon Claerbout, and Norm Sleep

ABSTRACT

We model propagation of waves on the surface of the Earth, using a finite element approach. A spherical shell is covered by a fine triangular mesh, and propagation matrices are found that step an acoustic wave field on this mesh forward in time. The propagation matrices are determined by means of variational (finite-element) techniques. By varying the distribution of mass within the individual triangular faces, different dispersion curves are obtained. These curves can be made to conform, over a certain range of frequencies, to desired dispersion relations. Techniques developed in the course of this work may be applicable to the problem of migrating irregularly-sampled 3–D reflection seismic data.

IRREGULAR MESH

Modeling of dispersive surface waves

Surface waves on the Earth, including both Love waves and Rayleigh waves, are dispersive; that is, waves with different frequencies have different phase velocities. For shorter-wavelength (less than 1000 km) waves, this dispersive behavior can be attributed to changes in rock velocity with respect to depth. The modeling of surface waves is a problem of interest to many seismologists. Techniques used include normal-mode methods (Woodhouse and Dziewonski, 1984), variational methods (Tanimoto and Anderson, 1985), and ray tracing with Gaussian beams (Yomogida and Aki, 1985). We are proposing a different technique: to simulate propagation of surface waves on a spherical solid, by means of a spherical shell on which only acoustic waves propagate. To propagate waves on such a shell, we use the finite-element method in conjunction with an explicit time-differencing scheme.

Modeling on an irregular grid

Unfortunately, there is no way to tile the sphere densely with a regular mesh (a regular mesh

is one whose mesh points are all the same distance from adjacent points). There are only five ways to tile the sphere with a regular mesh; these tilings correspond to the tetrahedron, cube, octahedron, dodecahedron, and icosahedron. The densest of these, the icosahedron, has only 12 vertices (mesh points). We can achieve a sufficiently dense tiling by subdividing the 20 triangular faces of the icosahedron, but this denser tiling will no longer be in the form of a regular mesh. Distance to nearest neighbors will depend on the position of the mesh point. Because of this unavoidable irregularity, we have had to develop techniques for propagating waves on an irregular mesh. These techniques may be applicable to other situations, such as 3-D reflection seismic surveys, where irregularities in sampling are unavoidable. The problem of processing irregularly sampled data was previously discussed by Dellinger (1986).

Another possible approach to the modeling problem is to use an implicit method (Mitchell and Griffiths, 1980, pp. 200–209.). With this method we would be able to use grid points that were regularly spaced in latitude and longitude; these grid points would consequently be bunched near the north and south poles. Because of the density of grid points near the poles, an explicit differencing scheme on such a grid would be unstable.

TILING THE SPHERE

Subdividing the icosahedron

The icosahedron, a twenty-sided regular polyhedron, is used as the basis of our tiling schemes (see Figure 1). Each triangular face of the icosahedron is sub-divided into smaller triangles, and the sub-divided triangular surfaces are "inflated" so that the vertices of all the smaller, internal triangles lie on the surface of a sphere (Figure 2). It should be noted that these smaller triangles are not equilateral, nor are they identical to each other. The lengths of their sides, and the angles between their sides, vary. Each of these smaller triangles can be subdivided in turn, and after another "inflation" step, the sphere is tiled by yet more triangles. No matter how many times the subdivisions are carried out, at the original 12 vertices five edges will meet at one point. At all other vertices, six edges meet at one point. Another way of expressing this is that 12 original vertices alway have 5 nearest neighbors each, while all the others have 6. It can be shown that a sphere cannot be tiled such that all vertices have six nearest neighbors (Pugh, 1976, p.59).

We use a recursive algorithm to tile the sphere, since it is relatively simple in recursive schemes to keep track of nearest neighbors. This is better than trying to find nearest neighbors after the final step on the basis of the coordinates of the vertices. The icosahedron serves as the starting point of the recursion scheme. The icosahedron has the advantages of having triangular faces, and of having more faces than any other regular polyhedron. Details of the recursion scheme are given in Appendix A.

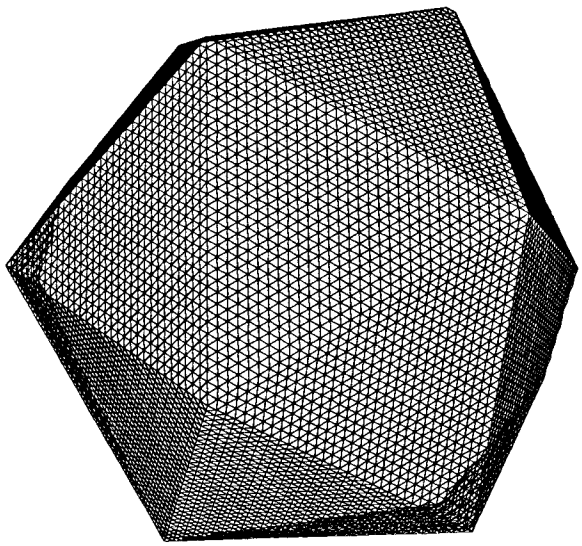


FIG. 1. Subdivided icosahedron. The artifacts at some of the edges are due to an imperfect hidden-line algorithm.

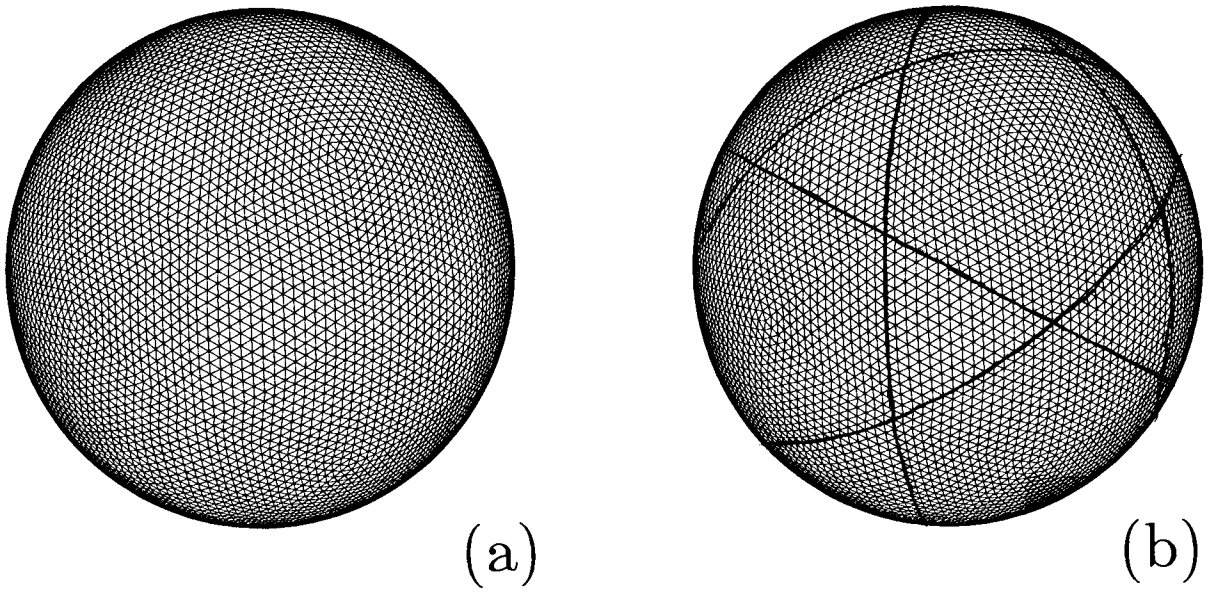


FIG. 2. Tiled sphere. This is the sort of sphere on which we will perform finite-element modeling. This is a level-5 recursion, with 10,242 vertices. It is shown with (b) and without (a) thickened lines to emphasize variations in the density of the grid.

There are some useful formulas that describe the number of vertices on the sphere after different levels of recursion. If the icosahedron, with 12 vertices, is considered the zero-th level of recursion, then the first level of recursion produces a grid made up of 42 vertices, the second level produces a grid made up of 162 vertices, and so on. In general, *number of vertices* $= 10 \times 4^k + 2$, where k is the level of the recursion. Other useful formulas are: *number of faces* $= 20 \times 4^k$, where k again represents the level of the recursion, and *number of edges* $= 30 \times 4^k$.

Regularizing the densities

Careful inspection of Figure 2 reveals systematic irregularities on the tiled sphere. There are well-defined areas with greater and lesser densities of vertices. The variation is about 6%; more important than the magnitude of this variation is its sharpness, which can create spurious reflections. To overcome this, we use a regularized sphere, on which some of these density variations have been smoothed out. The regularization technique is discussed below, in the section on numerical artifacts.

FINITE ELEMENTS

Derivation of 2-D acoustic wave equation

The derivation of the wave equation is given in almost any differential equations text (Berg and McGregor, 1966, pp. 181–185). For a one-dimensional string, with tension T and density (mass per unit length) ρ, the differential equation is:

$$\frac{T}{\rho}\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}, \tag{1}$$

where u is the vertical displacement of the string, t represents time, and x is the spatial coordinate.

In two dimensions there exists a similar wave equation, for a stretched membrane, with tension T and density (mass per unit area) ρ. The differential equation is now

$$\frac{T}{\rho}\nabla^2 u = \ddot{u}, \tag{2}$$

where $\ddot{u}$ is equivalent to d^2u/dt^2.

The irregular grid and finite element methods

Equation (2) is the equation we shall use to propagate waves on a sphere; note that it does not take sphericity of the membrane into account, since the membrane is assumed to be locally flat. We wish to use this continuous equation to obtain an equation that is discretized in time and space. We shall, for the moment, ignore the question of discretization in time; suffice it to say that we use a simple explicit differencing scheme. Spatial discretization is more complicated

since, as noted in the previous section, the sphere is tiled by an irregular grid. The problem, then, is as follows: given the value of u at a particular point and at its neighbors, find the value of the Laplacian ($\nabla^2 u$) at that point.

We would like to solve this problem by a finite-difference scheme. The difficulty is that if the Laplacian is not negative definite or negative semi-definite, the differencing scheme will be unstable. If some diagonal weighting element is added so that the Laplacian is guaranteed to be purely negative definite, the scheme will be stable but the wave will not propagate. What we need, then, is a way of finding a Laplacian that is negative semi-definite, meaning that it is on the edge of instability. The finite-element (variational) method allows us to find such a Laplacian.

There are two main approaches to solving a finite-element problem: the Galerkin method, and the Ritz-Rayleigh variational method (Zienkiewicz, 1971; Strang, 1986, pp. 430–431). Both of these methods give equivalent results, and for physically complicated problems, the Galerkin method is often the more straightforward. In fact, the Galerkin method was originally used to derive some of the results described below. We feel, however, that the Ritz-Rayleigh approach provides for simpler and more tutorial derivations.

The Lagrangian

Recall from elementary physics the concept of the Lagrangian, which is defined, for a particular system, as the difference between the kinetic and the potential energy:

$$\mathcal{L} \equiv \mathrm{KE} - \mathrm{PE}, \tag{3}$$

where PE represents the potential energy and KE represents the kinetic energy in the system. The equations of motion can be derived by minimizing the Lagrangian. This minimization can be accomplished by means of the Euler equations (Morse and Feshbach, 1953, pp. 276–278):

$$\frac{d}{dt}\left[\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right] = \frac{\partial \mathcal{L}}{\partial q_j}, \tag{4}$$

where q_j is the jth spatial coordinate, and $\dot{q}_j$ is the first time derivative of that coordinate. Simply plug the Lagrangian into equation (4), and the equations of motion will fall right out.

For the two-dimensional acoustic wave equation, the Lagrangian can be shown to be (Morse and Feshbach, 1953, p. 343):

$$\mathcal{L} = \frac{1}{2}\rho\left[\left(\frac{\partial u}{\partial t}\right)^2 - \frac{T}{\rho}|\nabla u|^2\right]. \tag{5}$$

Be sure to distinguish between time t and tension T.

Applying the Lagrangian to a triangular area

Now we are ready to apply our physical theory to the tiled sphere. Figure 3 shows a small portion of the grid near a particular point; this point has been designated point 0. Around point 0

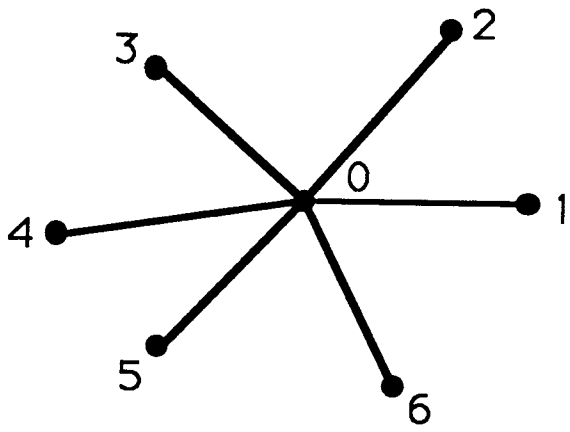


FIG. 3. A grid point and its nearest neighbors.

are its nearest neighbors, points 1 through 6. At each point i the function u takes on the value u_i. Now let the function u be constrained to vary linearly within each triangular face. The function u will be continuous, but its spatial derivatives will be discontinuous at the boundaries between triangular faces. This constraint that u be linear is a result of our choice of basis functions for u; a more complicated set of basis functions would allow us to specify a more complicated behavior for u. Three points determine a plane, so the values of u at the three corners of each triangle are sufficient to determine the values of u everywhere within that triangle.

Let us take a detailed look at the triangle which has points 0, 1, and 2 at its corners (see Figure 4). We want to find the equations of motion for this triangular face. Let $\dot{u} \equiv \partial u/\partial t$. We begin by finding $\int_s \mathcal{L}\, dx\, dy$, where $\int_s$ represents integration over the surface of the triangle. To do this, we must first find $\int_s \dot{u}^2\, dx\, dy$ and $\int_s |\nabla u|^2\, dx\, dy$. We are taking the density ρ and tension T to be constant within the triangle.

We find it convenient to first determine the gradient ∇u. It can be shown, after some algebra, that

$$\nabla u = \frac{1}{2A} \underset{\sim}{\mathbf{R}}_{90} \left[u_0(\mathbf{x}_2 - \mathbf{x}_1) + u_1(\mathbf{x}_0 - \mathbf{x}_2) + u_2(\mathbf{x}_1 - \mathbf{x}_0) \right]. \tag{6}$$

Here $\underset{\sim}{\mathbf{R}}_{90}$ is a coordinate-transform matrix that rotates vectors 90° in a counter-clockwise direction, A is the surface area of the triangle, and $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{x}_2$ are the coordinates of points 0, 1, and 2, respectively. We have not given the derivation of this equation, but we will mention some easily checked properties that suggest it is the correct one. First, it is linear in u_0, u_1, and u_2. Second, if $u_0 = u_1 = u_2$, the gradient is zero. Third, the scale factor $1/(2A)$ is correct. This last property can be checked by forming an isosceles right triangle, with point 1 placed at the 90° corner; let $u_1 = u_2$, and it can be seen that the equation degenerates properly into the 1-D case.

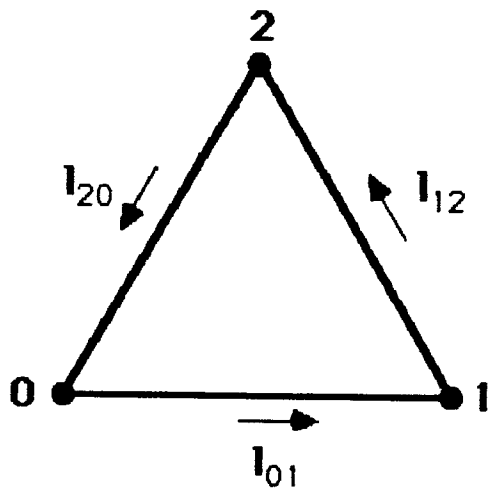


FIG. 4. Definition of $\mathbf{l}_{01}$, $\mathbf{l}_{12}$, and $\mathbf{l}_{20}$.

The gradient is constant within the triangle, so

$$\int_s |\nabla u|^2 \, dx\, dy = \tag{7}$$
$$\frac{1}{4A}(u_0^2 \mathbf{l}_{12}\cdot\mathbf{l}_{12} + u_1^2 \mathbf{l}_{20}\cdot\mathbf{l}_{20} + u_2^2 \mathbf{l}_{01}\cdot\mathbf{l}_{01} + 2u_0u_1 \mathbf{l}_{20}\cdot\mathbf{l}_{12} + 2u_0u_2 \mathbf{l}_{01}\cdot\mathbf{l}_{12} + 2u_1u_2 \mathbf{l}_{01}\cdot\mathbf{l}_{20}).$$

Here $\mathbf{l}_{01} \equiv \mathbf{x}_1 - \mathbf{x}_0$, $\mathbf{l}_{20} \equiv \mathbf{x}_0 - \mathbf{x}_2$, and $\mathbf{l}_{12} \equiv \mathbf{x}_2 - \mathbf{x}_1$ (see Figure 4).

The time derivative at any point $\mathbf{x}$ within the triangle is

$$\dot{u}(\mathbf{x}) = \dot{u}_0 + (\mathbf{x} - \mathbf{x}_0)\cdot\nabla\dot{u}. \tag{8}$$

Substitute (6) into (8) and integrate:

$$\int_s \dot{u}^2 \, dx\, dy = \frac{A}{6}(\dot{u}_0^2 + \dot{u}_1^2 + \dot{u}_2^2 + \dot{u}_0\dot{u}_1 + \dot{u}_0\dot{u}_2 + \dot{u}_1\dot{u}_2). \tag{9}$$

(This equation can be checked by letting $\dot{u}_0 = \dot{u}_1 = \dot{u}_2$ and noting that the 6's cancel.)

We can combine the results of equations (7) and (9) with equation (5), the definition of the Lagrangian:

$$\int_s \mathcal{L}\, dx\, dy = \frac{\rho A}{12}(\dot{u}_0^2 + \dot{u}_1^2 + \dot{u}_2^2 + \dot{u}_0\dot{u}_1 + \dot{u}_0\dot{u}_2 + \dot{u}_1\dot{u}_2) \tag{10}$$
$$-\frac{T}{8A}(u_0^2 \mathbf{l}_{12}\cdot\mathbf{l}_{12} + u_1^2 \mathbf{l}_{20}\cdot\mathbf{l}_{20} + u_2^2 \mathbf{l}_{01}\cdot\mathbf{l}_{01} + 2u_0u_1 \mathbf{l}_{20}\cdot\mathbf{l}_{12} + 2u_0u_2 \mathbf{l}_{01}\cdot\mathbf{l}_{12} + 2u_1u_2 \mathbf{l}_{01}\cdot\mathbf{l}_{20}).$$

Equation (10) gives us the equation for the Lagrangian integrated over a triangle. Now we can apply the variational principle, given in equation (4), to this formula, in order to find the equations of motion. To do this, however, we must specify q_j, the generalized spatial coordinate.

Below we will be discussing the equation of motion for point 0, based on the equations of motion of all the triangles surrounding that point. We will thus choose u_0 to be our generalized coordinate. Differentiating equation (10),

$$\int_s \frac{\partial \mathcal{L}}{\partial \dot{u}_0}\, dx\, dy = \frac{\rho A}{12}(2\dot{u}_0 + \dot{u}_1 + \dot{u}_2). \tag{11}$$

Similarly,

$$\int_s \frac{\partial \mathcal{L}}{\partial u_0}\, dx\, dy = -\frac{T}{8A}(2u_0\, \mathbf{l}_{12}^2 + 2u_1\, \mathbf{l}_{20}\cdot\mathbf{l}_{12} + 2u_2\, \mathbf{l}_{01}\cdot\mathbf{l}_{12}). \tag{12}$$

Thus, from the variational principle given in equation (4),

$$\frac{\rho A}{6}(\ddot{u}_0 + \frac{\ddot{u}_1}{2} + \frac{\ddot{u}_2}{2}) = -\frac{T}{4A}(u_0\, \mathbf{l}_{12}^2 + u_1\, \mathbf{l}_{20}\cdot\mathbf{l}_{12} + u_2\, \mathbf{l}_{01}\cdot\mathbf{l}_{12}). \tag{13}$$

As before, A represents the area of the triangle, ρ and T represent the density and tension within the triangle, and $\mathbf{l}_{01}$, $\mathbf{l}_{20}$, and $\mathbf{l}_{12}$ represent the vectors shown in Figure 4. As previously noted, T and ρ are taken to be constant within the triangle.

Combining the Lagrangians from adjacent triangles

Equation (13) shows the formula that is obtained by integrating over one triangular area; to obtain the true equations of motion, we must integrate over the hexagon made up of all the triangular faces surrounding u_0. This integration yields the result

$$\begin{aligned} &\frac{\ddot{u}_0}{6}\sum_{j=1}^{6}\rho_j A_j + \frac{1}{12}\sum_{j=1}^{6}(\rho_j A_j + \rho_{j-1}A_{j-1})\ddot{u}_j \\ &\quad = -\frac{u_0}{4}\sum_{j=1}^{6}\frac{T_j}{A_j}(\mathbf{l}_{j+1} - \mathbf{l}_j)^2 + \frac{1}{4}\sum_{j=1}^{6} u_j\left[\frac{T_j}{A_j}\mathbf{l}_{j+1}\cdot(\mathbf{l}_{j+1} - \mathbf{l}_j) + \frac{T_{j-1}}{A_{j-1}}\mathbf{l}_{j-1}\cdot(\mathbf{l}_{j-1} - \mathbf{l}_j)\right]. \end{aligned} \tag{14}$$

The terms used in this equation should be self-explanatory, but there is the possibility for confusion in the use of subscripts. For instance, j sometimes refers to the jth point, as with u_j and $\mathbf{l}_j$. At other times, it refers to the jth area, as with ρ_j, T_j, and A_j. Figure 5 is an attempt to clear up this ambiguity.

Equations similar to equation (14) can be written for all other u_i's. These simultaneous linear equations can be combined and written in matrix form:

$$\underset{\sim}{\mathbf{M}}\ddot{\mathbf{u}} = \underset{\sim}{\mathbf{T}}\mathbf{u}. \tag{15}$$

Here $\mathbf{u}$ represents the vector consisting of the values u_i at all points i, $\ddot{\mathbf{u}}$ represents the second derivative of $\mathbf{u}$ with respect to time, $\underset{\sim}{\mathbf{M}}$ is the "mass matrix" made up of terms similar to those on the left-hand side of equation (14), and $\underset{\sim}{\mathbf{T}}$ is the "tension matrix" made up of terms similar to those on the right-hand side of equation (14).

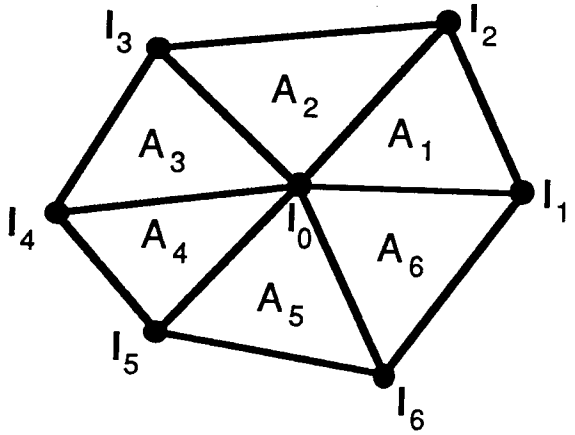


FIG. 5. Clearing up subscripts. There is an ambiguity when subscripts can refer to either a triangular face or a grid point. This Figure shows the convention used in this paper.

Showing that $\underset{\sim}{\mathbf{M}}$ and $\underset{\sim}{\mathbf{T}}$ are positive definite

Both $\underset{\sim}{\mathbf{M}}$ and $\underset{\sim}{\mathbf{T}}$ have the form of a "nearest-neighbor-plus-diagonal" matrix: on row i, the value in column j will be non-zero only if $j = i$ or if point j is a nearest neighbor of point i. We can show that $\underset{\sim}{\mathbf{M}}$ is positive definite and $\underset{\sim}{\mathbf{T}}$ is negative semi-definite, by giving the decomposition of $\underset{\sim}{\mathbf{T}}$ into a matrix product of the form $-\underset{\sim}{\mathbf{A}}^T\underset{\sim}{\mathbf{A}}$, and by giving the decomposition of $\underset{\sim}{\mathbf{M}}$ into the sum of a a matrix product of the form $\underset{\sim}{\mathbf{A}}^T\underset{\sim}{\mathbf{A}}$ plus a positive-definite diagonal matrix. We will not, however, give these decompositions in this paper.

Stepping forward in time

Now that a system of equations linking $\ddot{\mathbf{u}}$ and $\mathbf{u}$ has been found, $\mathbf{u}$ can be stepped forward in time. Let $\ddot{\mathbf{u}} = (\mathbf{u}_{t+1} - 2\mathbf{u}_t + \mathbf{u}_{t-1})/\Delta t_2$. Then

$$\mathbf{u}_{t+1} = \Delta t^2 \, \underset{\sim}{\mathbf{M}}^{-1} \underset{\sim}{\mathbf{T}} \mathbf{u}_t + 2\mathbf{u}_t - \mathbf{u}_{t-1}, \tag{16}$$

where $\underset{\sim}{\mathbf{M}}^{-1}$ is the inverse of $\underset{\sim}{\mathbf{M}}$. In order to find $\mathbf{u}_{t+1}$, then, it is necessary to find $\underset{\sim}{\mathbf{M}}^{-1}$. Keep in mind that $\underset{\sim}{\mathbf{M}}$ is large (perhaps 160,000 $\times$ 160,000 elements) and sparse (7 elements per row on average), while $\underset{\sim}{\mathbf{M}}^{-1}$ is large and non-sparse. The lack of sparsity in the inverse suggests that, rather than storing an impossibly large matrix $\underset{\sim}{\mathbf{M}}^{-1}$, we should solve a system of equations of the form $\underset{\sim}{\mathbf{M}}\mathbf{x} = \mathbf{b}$ at each time step, where $\mathbf{x} = \ddot{\mathbf{u}}$ and $\mathbf{b} = \underset{\sim}{\mathbf{T}}\mathbf{u}$. There are programs that can carry out this solution efficiently for sparse matrices (see Paige and Saunders (1975), for instance), but it still remains an expensive operation. There are approximations that can reduce the expense considerably, but first it is worth examining the frequency dispersion that is caused by our finite-element propagator matrices.

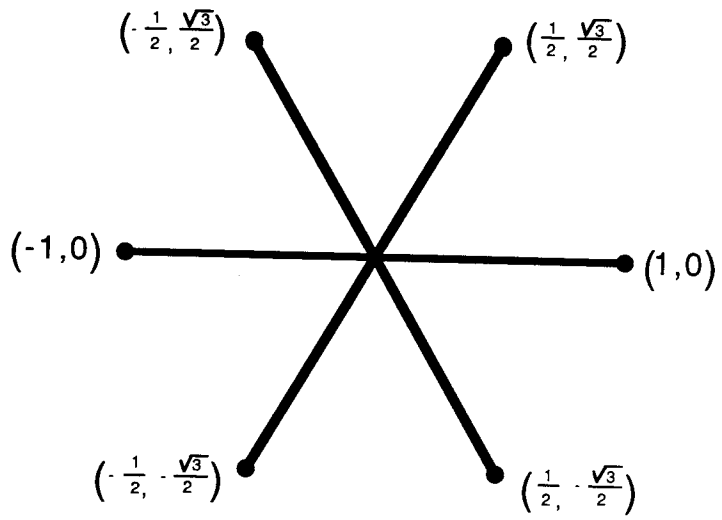


FIG. 6. Layout of points on a regular grid. The center point is assumed to be at coordinates (0., 0.).

DISPERSION RELATIONS

Definition of dispersion

In numerical models, as well as in the real earth, waves of different frequencies often travel at different speeds. As a result, a compact wave packet will grow less compact, that is, it will disperse, as it travels. This dispersion can be described quantitatively, by finding a function that gives velocity as a function of frequency. It is useful to examine the dispersion of the triangularly gridded model that we have described above.

Analyzing dispersion on a regular grid

We have used the finite-element method to find an equation for wave propagation, with the assumption that the grid is irregular. For the purpose of studying dispersion, let us instead assume that the grid is regular. In that case, equation (14) becomes

$$\ddot{u}_0 \rho A + \frac{\rho A}{6} \sum_{j=1}^{6} \ddot{u}_j = -\frac{3Tl^2}{2A} u_0 + \frac{Tl^2}{4A} \sum_{j=1}^{6} u_j, \tag{17}$$

where l is the length of a side of the triangle. Since the grid is regular, the triangles are all equilateral, and A, the area of each triangle, equals $\sqrt{3}l^2/4$. Recall that u_j equals the value of the wave field u at grid point j; Figure 6 shows how the grid points are arranged.

Equation (17) represents a convolutional operator; this operator can be Fourier transformed

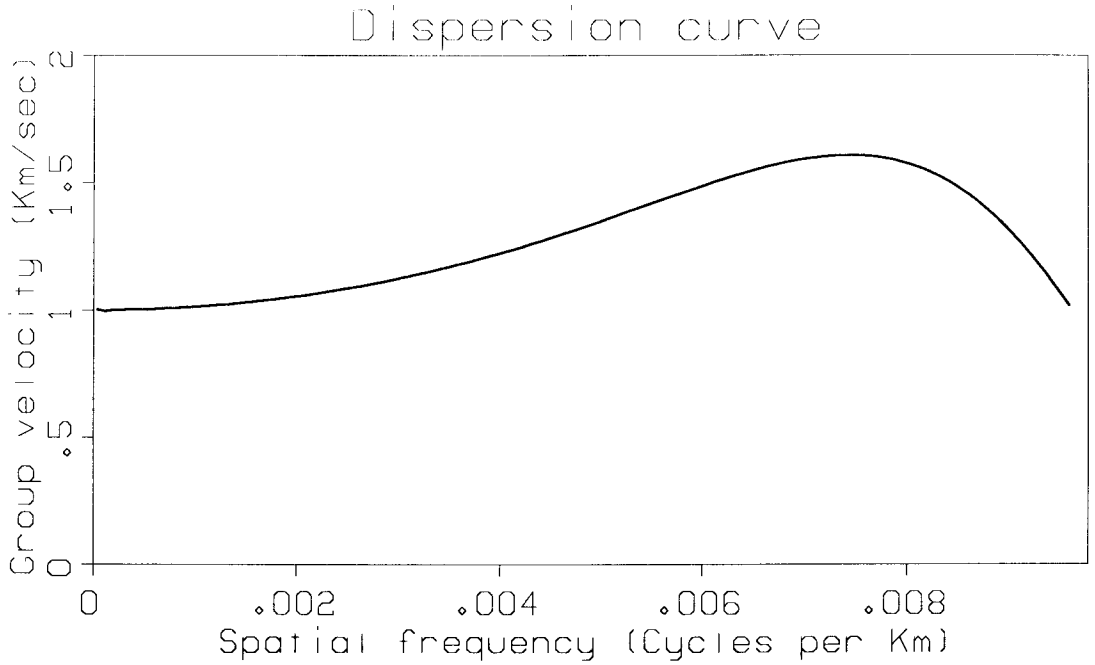


FIG. 7. Group velocity versus spatial frequency for original finite-element equation. Note that this curve is not flat, even at low spatial frequencies.

over x, y, and t with the result that

$$(6 + 2\,\mathrm{cosf}\,\mathbf{k}) \cdot (\cos \omega \Delta t - 1) = \frac{8T}{\rho} \frac{\Delta t^2}{l^2} (\mathrm{cosf}\,\mathbf{k} - 3), \tag{18}$$

where

$$\mathrm{cosf}\,\mathbf{k} \equiv \cos(k_x l) + \cos(\frac{1}{2} k_x l + \frac{\sqrt{3}}{2} k_y l) + \cos(-\frac{1}{2} k_x l + \frac{\sqrt{3}}{2} k_y l). \tag{19}$$

In this equation, ω is the temporal frequency, Δt is the discretization step in time, and $\mathbf{k} = (k_x, k_y)$, where k_x and k_y are the spatial frequencies in the x and y directions (see Figure 6).

Dispersion curves

Equation (18) may be easily solved (we won't solve it here) to give ω as a function of $\mathbf{k}$ (for the sake of simplicity, let us assume for the time being that $\mathbf{k} = k\hat{\mathbf{k}}$, where $\hat{\mathbf{k}}$ is a constant unit vector). We can speak of two sorts of velocity: phase velocity, defined as $v_p \equiv \omega(k)/k$, and group velocity, defined as $v_g \equiv d\omega/dk$ (Claerbout, 1976, pp. 13–16). Both definitions of velocity are useful at times, but let us focus our attention on group velocity, since this represents the velocity at which energy propagates. Figure 7 shows the plot of group velocity versus spatial frequency k for equation (18). This dispersion curve is not ideal. Unless we specify otherwise, we would like the dispersion curve to be as flat as possible, since a flat dispersion curve implies that waves of

different spatial frequencies all travel at the same velocity. The dispersion curve shown in Figure 7 is unsuitable; it is concave upwards, even at the lowest frequencies.

Why is the dispersion curve not flat? In our derivation of the equation of motion, we placed only one limitation on the model: the wave field u must vary linearly within each triangular area. This single restriction, however, as Fabio Rocca has pointed out (personal communication), is sufficient to explain why the dispersion curve tends to rise (at least at first) as spatial frequency increases. Recall from elementary physics the equation

$$v = \sqrt{\frac{\text{Rigidity}}{\text{Mass}}}\,. \tag{20}$$

At low frequencies (long wavelengths), propagating waves tend to notice only the effects of rigidity due to T, the tension on the membrane. As wavelengths shorten, however, the waves begin to be affected something else: the medium (think of it as an elastic membrane) can bend (change slope) only at the boundaries of the triangular areas. No bending can occur within the areas themselves, since u is constrained to vary only linearly within each area. Thus, as the wavelengths grow shorter, the waves begin to feel the effect of traveling in a medium composed of rigid plates. This apparent increase in rigidity with decreasing wavelength results in an increase in velocity, as predicted by equation (20).

Controlling the dispersion curve

How is it possible to compensate for this increase in apparent rigidity? Again, Fabio Rocca (personal communication) has suggested a physical justification for an approach that was developed by other means. The solution to the problem of increasing apparent rigidity is simple: provide an increase in apparent mass as compensation. Recall that the equations of motion (such as equation (18)) were derived under the assumption that the mass is evenly distributed over each triangular face. But suppose that the mass were not evenly distributed. Suppose, for instance, that the total mass within each triangular face remained unchanged, but that within each face, more of the mass were put at the corners of the triangle, and less at the center. This would result in an increase in angular momentum. For long-wavelength waves, the angular momentum would be unimportant, since such waves cause the slopes of the triangles to change only slowly with respect to time. Short-wavelength waves, on the other hand, cause the slopes to change rapidly. From the point of view of short-wavelength waves, then, an increase in angular momentum is equivalent to an increase in mass. Thus, from equation (20), shorter-wavelength waves, which feel an increase in apparent mass, will travel slower than longer-wavelength waves. If the angular momentum is chosen correctly, the effect of the apparent increase in mass will cancel out the effect of the apparent increase in rigidity.

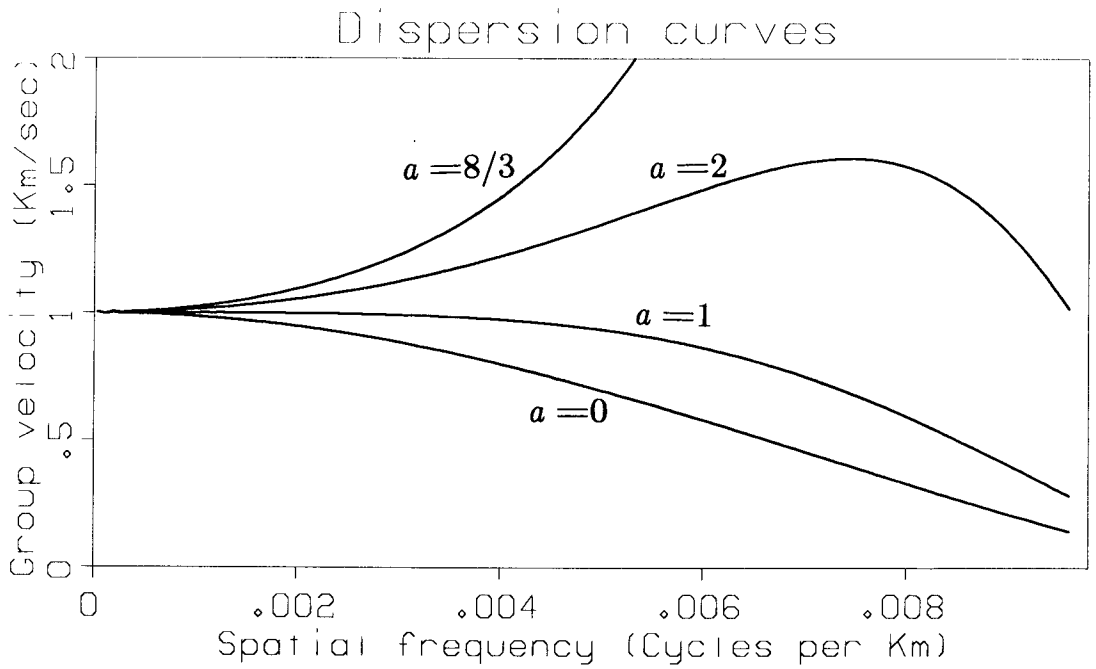


FIG. 8. Dispersion curves for the modified finite element equation. This Figure shows the dispersion curves resulting from various choices of a. The flattest curve (at low frequencies) results when $a = 1$.

Without presenting the derivation, we will give a generalized form of equation (18):

$$[12 + a(\text{cosf}\,\mathbf{k} - 3)] \cdot (\cos\omega\Delta t - 1) = \frac{8T}{\rho}\frac{\Delta t^2}{l^2}(\text{cosf}\,\mathbf{k} - 3), \tag{21}$$

where a is the parameter that determines the distribution of mass within the triangular faces. When $a = 2$, equation (18) results, meaning that the mass is evenly distributed over each face. When $a = 0$, the mass is concentrated at the corners of each face, and when $a = 8/3$, the mass is concentrated at the midpoint of each face. Figure 8 shows the dispersion curves associated with different values of a. It can be seen that the flattest curve (at low frequencies, anyway) is obtained when $a = 1$.

In general, the parameter a can be varied, along with T and ρ, to provide desired dispersion relations (a dispersion relation is a function describing how v_g or v_p varies with ω). Note that these parameters are not mathematical abstractions; each of them has a physical meaning, and each triangular face can have its own values for the three parameters. Unfortunately, these three parameters (or two, since ρ and T appear in the same term in equation (21)) are not sufficient to allow us to fit all dispersion curves. There is, however, an additional parameter that can help us.

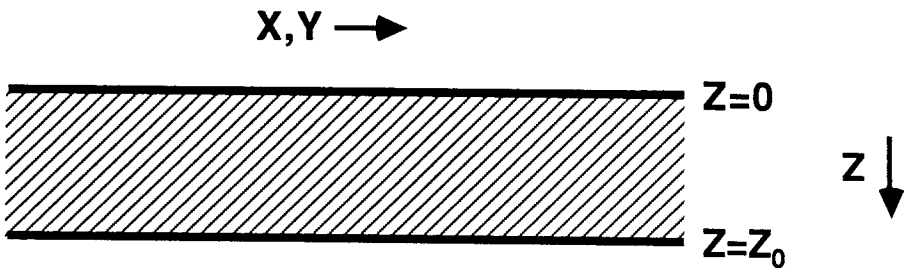


FIG. 9. Thick-plate model. Shown is a cross section through a thick plate.

The thick-plate model

Imagine that instead of propagating waves on a flat membrane, we were propagating them through a flat plate governed by the 3-D acoustic wave equation (see Figure 9). This equation is

$$\frac{\partial^2 u'}{\partial t^2} = \frac{\partial^2 u'}{\partial x^2} + \frac{\partial^2 u'}{\partial y^2} + \frac{\partial^2 u'}{\partial z^2}, \tag{22}$$

where u' is the wave field within the plate. Note that z is in the direction of thickness of the plate. Assume that the wavefield u' can be decomposed in the form

$$u'(x, y, z, t) = u(x, y, t)\,\Phi(z). \tag{23}$$

Substitute equation (23) into equation (22), and Fourier transform over z, with the result

$$\left(-\frac{\partial^2 u}{\partial t^2} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - k_z^2 u\right) \bar{\Phi} = 0, \tag{24}$$

where $\bar{\Phi}$ represents the Fourier transform of Φ. In general, $\bar{\Phi} \neq 0$, so

$$\left(\frac{\partial^2}{\partial t^2} + k_z^2\right) u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}. \tag{25}$$

Depending on the boundary conditions at $z = 0$ and $z = z_0$, the parameter k_z^2 can take on only certain values. However, by varying z_0 (the thickness of the plate), these allowed values of k_z^2 can be changed at will. The main consequence is that we have another adjustable parameter, since equation (25) resembles the two-dimensional acoustic wave equation, with a k_z^2 term added in. Equation (21), which gives the dispersion relation, can thus be rewritten:

$$[12 + a(\text{cosf}\,\mathbf{k} - 3)] \cdot (\cos\omega\Delta t - 1 + k_z^2) = \frac{8T}{\rho}\frac{\Delta t^2}{l^2}(\text{cosf}\,\mathbf{k} - 3). \tag{26}$$

Fitting dispersion curves

Now we are ready to discuss the problem of fitting the dispersion curves in our model to the dispersion curves observed in the earth. Suppose we know the phase velocity (v_p), as well as the

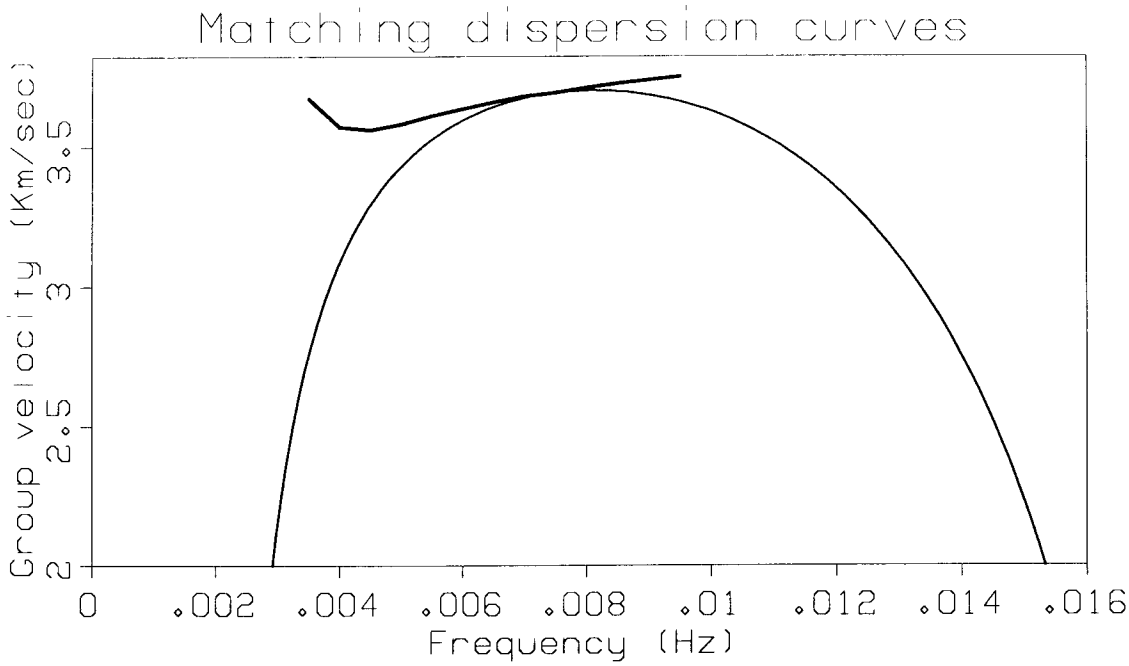


FIG. 10. Matching model dispersion to Earth dispersion. The thick line shows an observed group velocity function (PREM). The thinner line is an attempt to match it by varying certain parameters in the finite-element model. The slopes match, but the second derivatives of the curves do not.

first and second derivatives of phase velocity with respect to frequency ($dv_p/d\omega$ and $d^2v_p/d\omega^2$), at a given frequency. This is sufficient information for us to determine the value of the parameters T, a, and k_z^2. We will not give the equations for determining these parameters, but the general approach is to differentiate equation (26) repeatedly by ω, and to apply the relation $v_p = \omega/k$ and its derivatives.

Figure 10 shows the results of an attempt to match the model dispersion to the velocity dispersion found in the Earth. The Earth dispersion curve comes from Dziewonski and Anderson (1981) by way of Toshiro Tanimoto and Rob Clayton (personal communication); it is derived from a model known as PREM (for Preliminary Reference Earth Model). Note that the plot shows group velocity; matching v_p and its first and second derivatives is equivalent to matching group velocity (v_g) and its first derivative ($dv_g/d\omega$). The results shown in Figure 10 suggest that it might be advisable to find one more adjustable parameter, which would allow us to fit the second derivative of group velocity ($d^2v_g/d\omega^2$) as well.

As an alternative fitting method, instead of matching velocity and derivatives of velocity, it might be worthwhile to try and match several different velocities at several different frequencies, simultaneously. If there are three velocities, then all three can be fitted exactly (since there are

three adjustable parameters). If there are more than three velocities, then the problem must be solved by least-squares methods.

Approximations

So far we have not discussed the cost of our algorithms. As mentioned above, the system of equations

$$\ddot{\mathbf{u}} = \underset{\sim}{\mathbf{M}}^{-1}\,\underset{\sim}{\mathbf{T}}\mathbf{u} \tag{27}$$

requires that the inverse to $\underset{\sim}{\mathbf{M}}$ be found. Is it possible to find an approximation to $\underset{\sim}{\mathbf{M}}^{-1}$? Suppose $\underset{\sim}{\mathbf{M}}$ can be broken into two matrices, $\underset{\sim}{\mathbf{M}}_D$ and $\underset{\sim}{\mathbf{M}}_N$, such that $\underset{\sim}{\mathbf{M}}_D$ is diagonal and $\underset{\sim}{\mathbf{M}} = \underset{\sim}{\mathbf{M}}_D + \underset{\sim}{\mathbf{M}}_N$. Then

$$\underset{\sim}{\mathbf{M}}^{-1} \approx (\underset{\sim}{\mathbf{I}} - \underset{\sim}{\mathbf{M}}_D^{-1}\underset{\sim}{\mathbf{M}}_N)\,\underset{\sim}{\mathbf{M}}_D^{-1} \tag{28}$$

(recall that $\underset{\sim}{\mathbf{M}}_D$ is diagonal and thus easily inverted). The matrix inversion has been turned into multiplication by the sparse matrix $\underset{\sim}{\mathbf{M}}_N$.

This approximation can be expressed in terms of the dispersion relation. Let us examine equation (26), whose left-hand multiplier $[12 + a(\mathrm{cosf}(\mathbf{k}) - 3)]$ corresponds to the mass matrix $\underset{\sim}{\mathbf{M}}$. Which terms of this multiplier correspond to $\underset{\sim}{\mathbf{M}}_D$ and $\underset{\sim}{\mathbf{M}}_N$? The term involving a contains $\mathrm{cosf}(\mathbf{k})$, and is thus the only off-diagonal term. For small values of k, $\mathrm{cosf}(\mathbf{k}) \approx 3$, so this term can be considered small, especially if a is small as well. Therefore, the term containing a is both off-diagonal and small, a perfect choice for $\underset{\sim}{\mathbf{M}}_N$. The 12 corresponds to $\underset{\sim}{\mathbf{M}}_D$. By Taylor expansion,

$$[12 + a(\mathrm{cosf}\,\mathbf{k} - 3)]^{-1} \approx \frac{1}{12}\left[1 - \frac{a}{12}(\mathrm{cosf}\,\mathbf{k} - 3)\right]. \tag{29}$$

Thus we can rewrite equation (26) in the approximated form

$$(\cos\omega\Delta t - 1 + k_z^2) = \frac{2T}{3\rho}\frac{\Delta t^2}{l^2}\left[1 - \frac{a}{12}(\mathrm{cosf}\,\mathbf{k} - 3)\right]\cdot(\mathrm{cosf}\,\mathbf{k} - 3). \tag{30}$$

This equation shows, in an indirect way, how a matrix inversion can be replaced by a matrix multiplication. If the approximated inverse given in equation (28) is used in place of the exact inverse, then equation (30) should be used in place of equation (26) to compute the parameters T, a, and ${k_z}^2$ that best fit a given dispersion curve.

The final form of the finite-element equations

By now we have modified the finite-element equations quite a bit from their original form (equations (14) and (16)). Let us rewrite them. Equation (16) becomes:

$$\mathbf{u}_{t+1} = \Delta t^2\,(\underset{\sim}{\mathbf{I}} - \underset{\sim}{\mathbf{M}}_D^{-1}\underset{\sim}{\mathbf{M}}_N)\,\underset{\sim}{\mathbf{T}}\mathbf{u}_t + 2(\underset{\sim}{\mathbf{I}} - \underset{\sim}{\mathbf{K}}_{\mathbf{z}}^2)\mathbf{u}_t - \mathbf{u}_{t-1}, \tag{31}$$

where the Taylor-series approximation described above has been introduced. The adjustable parameter k_z^2 has been put into the form of a diagonal matrix, so that each grid point has a value

k_z^2 associated with it. It is convenient to define $\underset{\sim}{\mathbf{M}}_D$, $\underset{\sim}{\mathbf{M}}_N$, and $\underset{\sim}{\mathbf{T}}$ in a form similar to that of equation (14). Let us assume that $\mathbf{u}$ consists of only seven elements, and that these elements correspond to the points 0 through 7 shown in Figure 3. Then:

$$\underset{\sim}{\mathbf{M}}_D \mathbf{u} = \frac{u_0}{3} \sum_{j=1}^{6} \rho_j A_j, \tag{32}$$

$$\underset{\sim}{\mathbf{M}}_N \mathbf{u} = -\frac{u_0}{12} \sum_{j=1}^{6} a_j \rho_j A_j + \frac{1}{24} \sum_{j=1}^{6} u_j (a_j \rho_j A_j + a_{j-1} \rho_{j-1} A_{j-1}), \tag{33}$$

and

$$\underset{\sim}{\mathbf{T}} \mathbf{u} = -\frac{u_0}{4} \sum_{j=1}^{6} \frac{T_j}{A_j} (\mathbf{l}_{j+1} - \mathbf{l}_j)^2 + \frac{1}{4} \sum_{j=1}^{6} u_j \left[\frac{T_j}{A_j} \mathbf{l}_{j+1} \cdot (\mathbf{l}_{j+1} - \mathbf{l}_j) + \frac{T_{j-1}}{A_{j-1}} \mathbf{l}_{j-1} \cdot (\mathbf{l}_{j-1} - \mathbf{l}_j) \right]. \tag{34}$$

The notation is the same as in equation (14), except that the adjustable parameter a is now subscripted because each face has a value a associated with it. The matrix $\underset{\sim}{\mathbf{I}}$ is the identity matrix. The matrix $\underset{\sim}{\mathbf{M}}_D$ is diagonal, so $\underset{\sim}{\mathbf{M}}_D^{-1}$ is easily determined. It should be obvious how to define these matrixes when a large grid (rather than the 7-point grid shown here) is used.

Anisotropy of regular triangles

We have discussed how the velocity of wave propagation varies with frequency, and we have derived formulas for determining this velocity. These same formulas can be used to study anisotropy, that is, how velocity varies with direction of propagation. Recall that such equations as equation (30) can be used to determine phase or group velocity. These equations contain a term $\mathbf{k}$, which has been defined as $k\hat{\mathbf{k}}$, where k is the spatial frequency, and $\hat{\mathbf{k}}$ is a unit vector determining direction of propagation. Previous graphs, such as Figs. 7, 8, and 10, have been plotted under the assumption that $\hat{\mathbf{k}}$ is constant. Figure 11 shows what happens when different values of $\hat{\mathbf{k}}$ are used; the changes in $\hat{\mathbf{k}}$ correspond to changes in the direction of wave propagation.

Anisotropy of irregular triangles

Figure 11 shows that the problems caused by anisotropy are not too severe on a regular grid. These problems become more serious, however, as the grid becomes less regular. It is difficult to analyze dispersion relations on a wholly irregular grid, but it is possible to look at the effects of anisotropy on a grid that is partially irregular. Figure 12 shows one such semi-irregular grid. Notice that the hexagonal shape is no longer regular, but that it is made up of six identical triangles rotated with respect to each other. An entire plane may be regularly tiled with such a grid, making it possible to carry out analyses in the Fourier domain; the fact that the triangles are no longer regular allows us to study the effects of anisotropy in more detail. On the spherical grid there are many regions that are tiled by irregular triangles (see Figure 20 below).

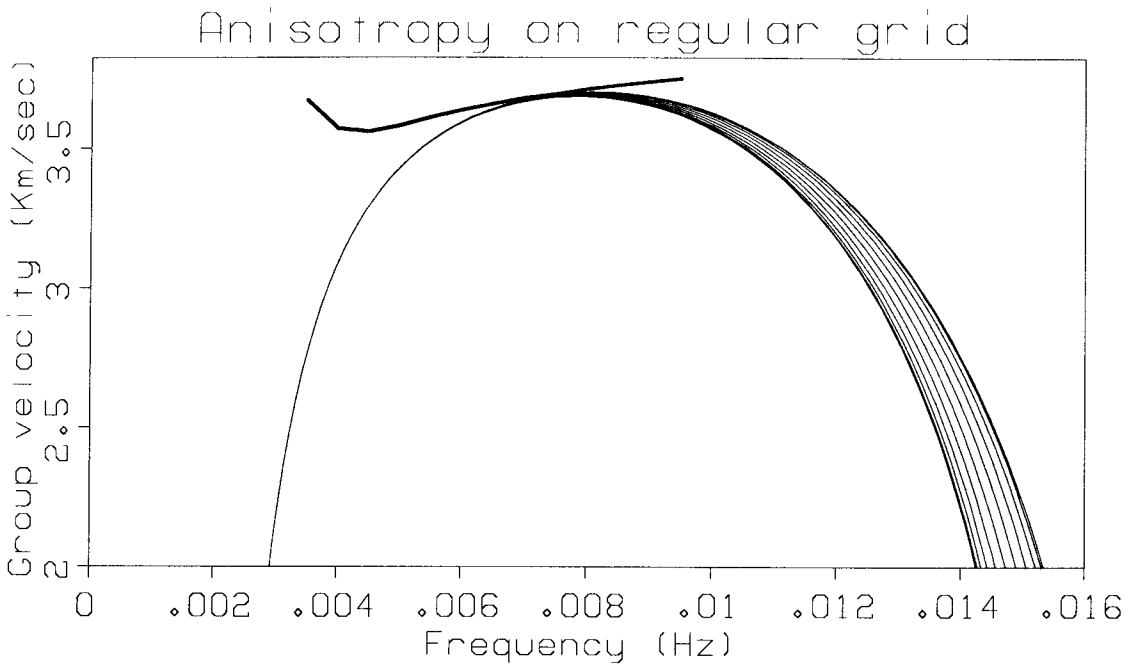


FIG. 11. Anisotropy on a regular grid. This plot is similar to Figure 10, except that the dispersion curve is plotted for waves propagating in a number of directions.

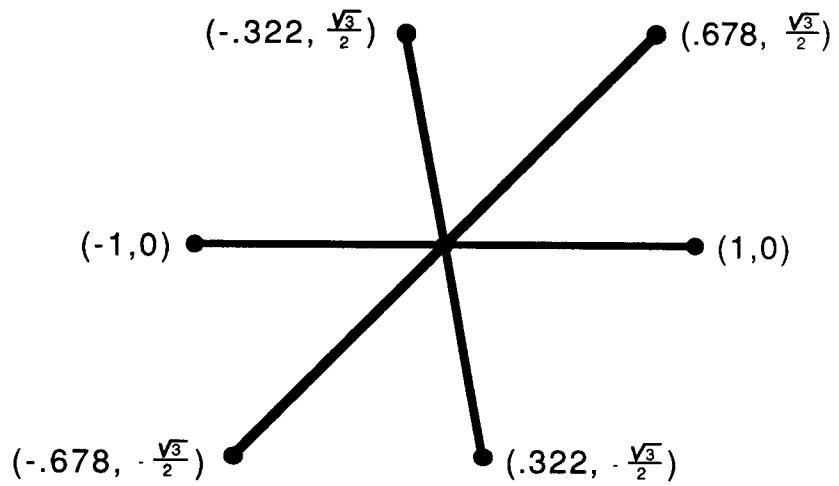


FIG. 12. Semi-irregular grid. This grid will be used to produce the curves shown in the next Figure (Figure 13).

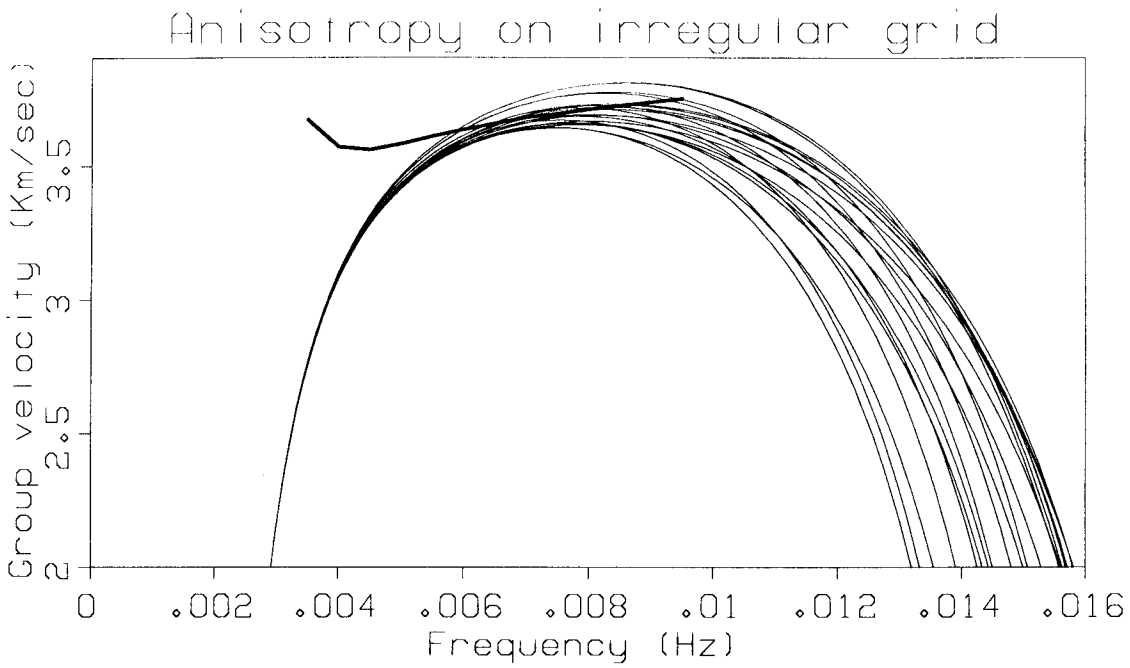


FIG. 13. Anisotropy on a semi-irregular grid. Shown here are the dispersion relations for various angles of propagation on the irregular grid shown in Figure 12. Note that this anisotropy is much greater than the anisotropy seen in Figure 11.

Figure 13 shows the dispersion curves that are obtained from the semi-irregular grid in Figure 12. Note that the anisotropy on this grid is much greater than on the regular grid. We should expect that anisotropic wave propagation will be a problem on the spherical grid, especially at higher spatial frequencies.

RESULTS

Propagation on a homogeneous sphere

We have run some numerical experiments to test the finite-element method discussed above. We first show the results of wave propagation on the surface of a homogeneous sphere, since in such a case it is easy to see artifacts. We used a regularized grid of recursion level 7, meaning that there were 163,842 grid points. This corresponds on the Earth's surface to a separation of 60 km between adjacent grid points. The propagation velocity was 4.0 km/sec, and the time steps were 5 seconds apart. The center frequency of the wavelet was .008 Hz; its half-width was .006 Hz. The parameters T, a, and k_z^2 were chosen so as to make the dispersion curve as flat as possible at the center frequency. From the figures for velocity and frequency, the dominant wavelength is

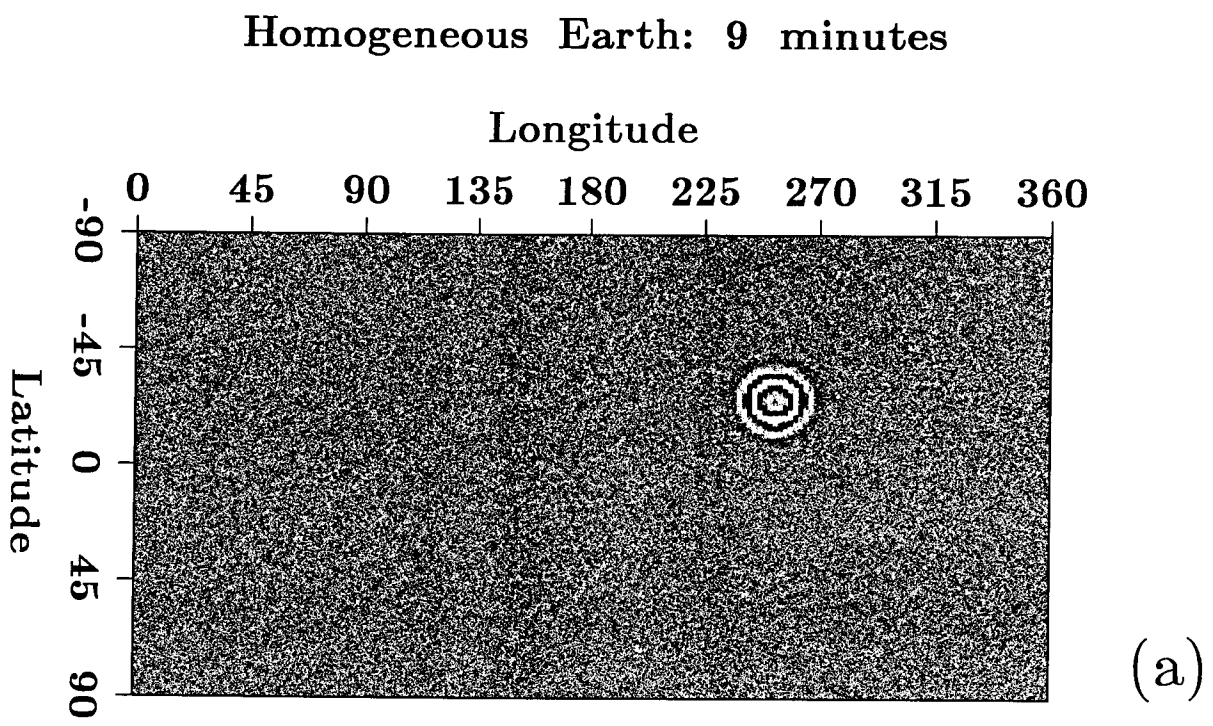
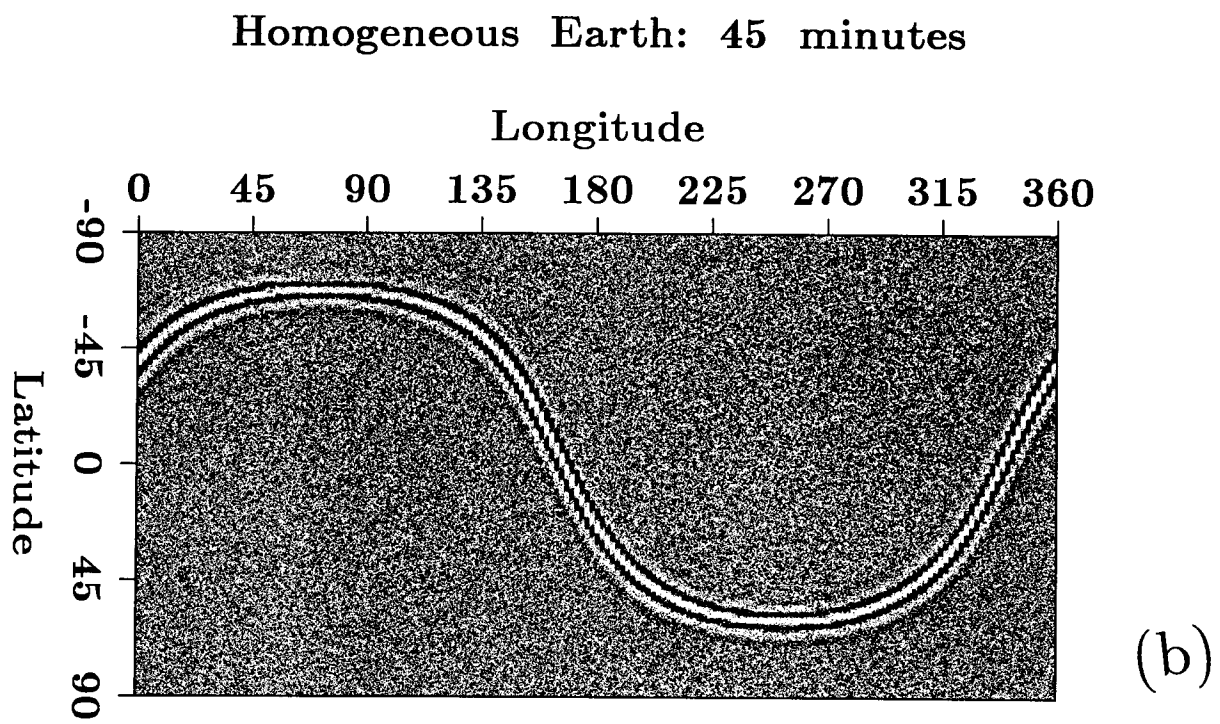


FIG. 14. Propagation on a homogeneous sphere. On this page and the following two pages are shown the results of propagating a wave on a homogeneous sphere.

500 km, meaning that there are 8 to 9 grid points per wavelength. The source was located on one of the twelve original vertices of the icosahedron; any artifacts would thus be expected to have five-fold symmetry. The modeling program simulated 6 hours of wave propagation (4320 time steps); it consumed 20 hours of CPU time on our Convex C-1 computer, and required about 25 Mbytes of RAM.

The results of the wave-propagation experiment are shown in Figure 14. This Figure shows a series of snapshots: the wave leaves the shot point (a), travels through the medium (b), arrives at the antipode (c), travels back towards the shot point (d), arrives back at the shot point (e), and leaves the shot point again (f). The final frame (g) is from a later time, after the wave has left the shot point for a third time. It is interesting to note the 90^o phase shifts between frames (b), (d), and (f); these phase shifts are due to the passing of the wave through a focus (the shot point or its antipode). The phase shifts are additive, so frame (f) is 180^o out of phase relative to frame (b).

By frame (g) some dispersion is visible. This is not too surprising; it is impossible to simulate completely flat dispersion curves. If we wanted to perform realistic modeling of seismograms, we would use a narrower frequency window, resulting in less dispersion and a wider (and less



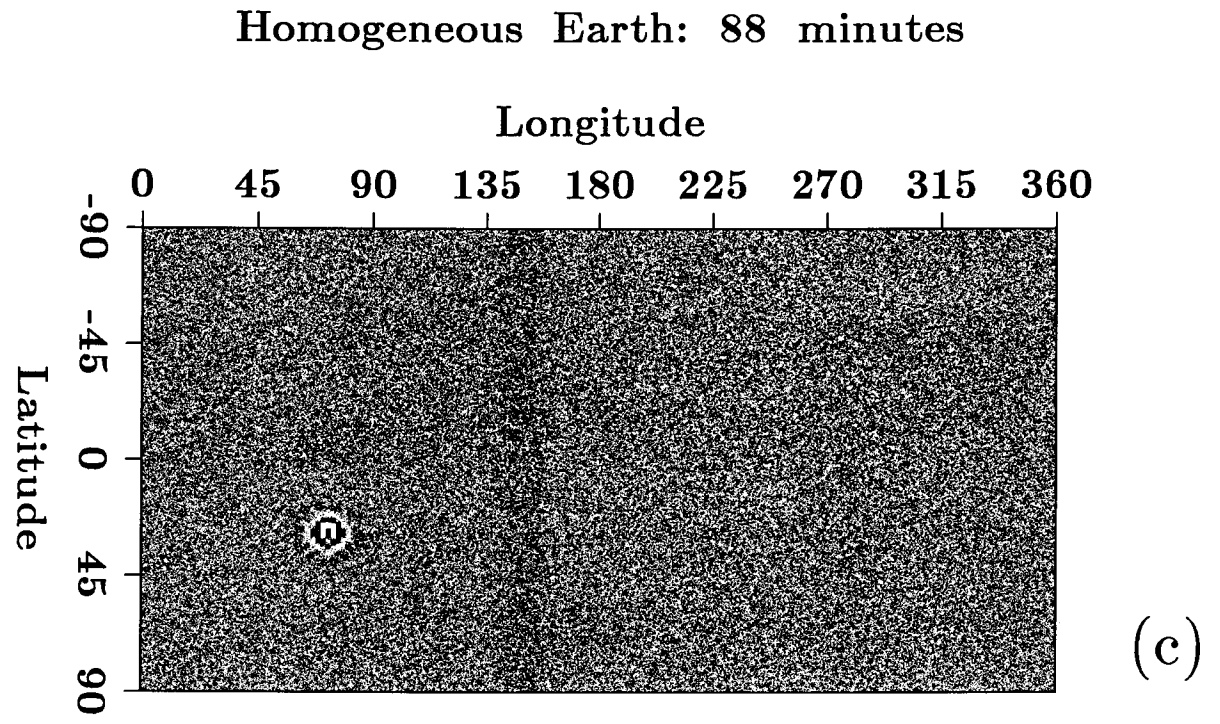
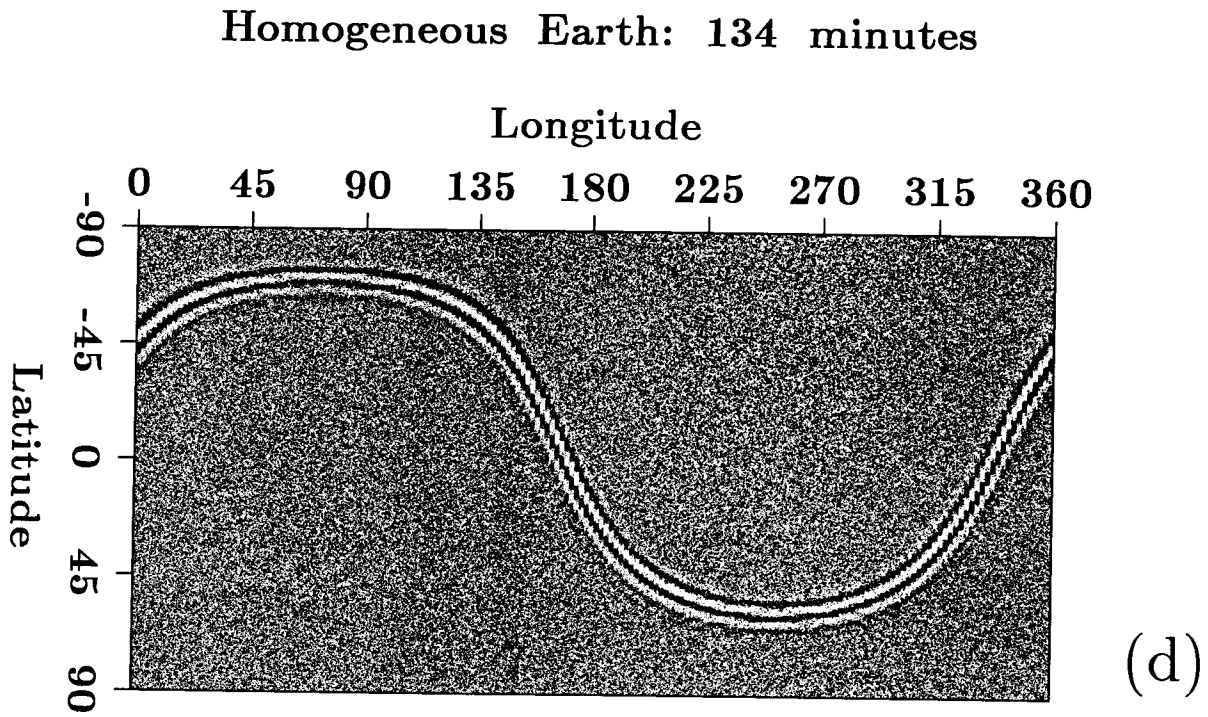


FIG. 14 continued.



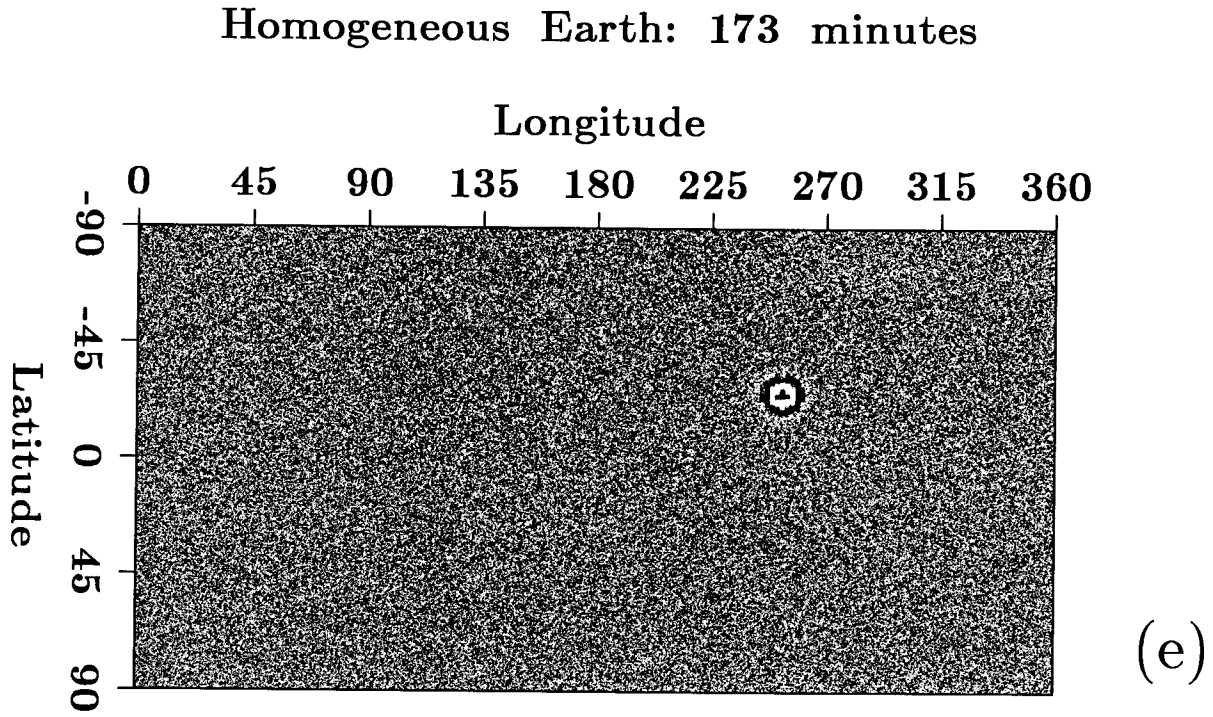
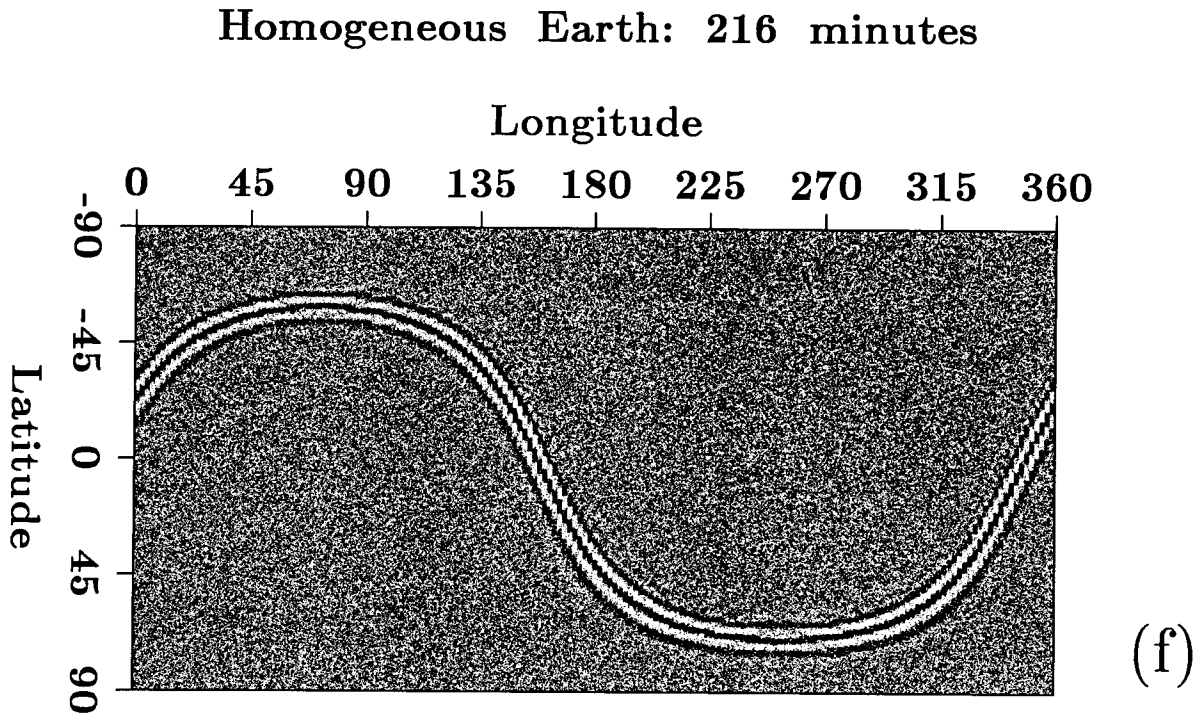


FIG. 14 continued.



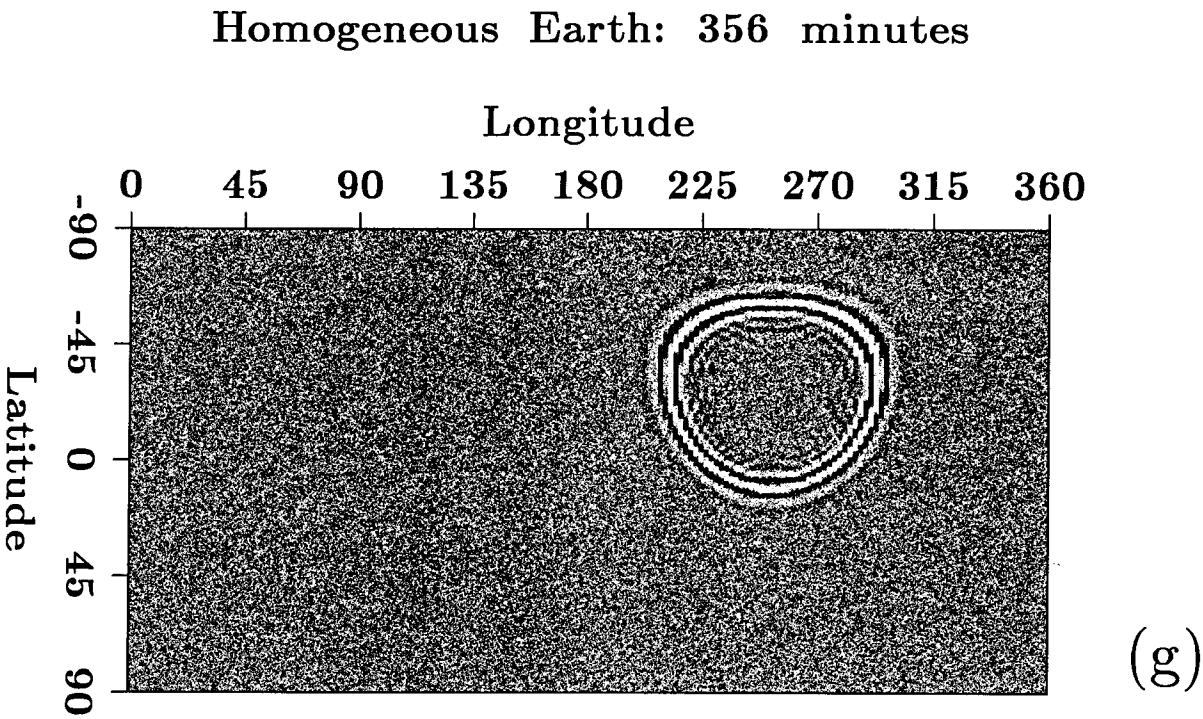


FIG. 14 continued.

photogenic) wavelet.

Propagation on an inhomogeneous sphere

We were encouraged by the previous results: on a homogeneous sphere, where artifacts should be most obvious to the eye, there are very few visible. We proceeded to model the propagation of waves on an inhomogeneous Earth, with a source located near the city of Semipalatinsk in Soviet Kazakhstan. The other parameters were the same as for the homogeneous Earth, an exception being that we used two sets of T, a, and ${k_z}^2$: one to give a group velocity of 5.0 km/sec at the center frequency, and the other to give a group velocity of 4.5 km/sec. All points on the Earth above -200 meters elevation were arbitrarily assigned the higher velocity, and all points below that depth were assigned the lower velocity. We must emphasize that this is not a geophysically correct model in two respects: surface wave velocities on the Earth actually vary by only a few percent from the average; and the velocity variations between different regions of the ocean are often greater than the variations between the average continental region and the average ocean region. Our choice of velocity models was necessitated mostly by convenience; we were recently provided a set of digital topography data courtesy of Bob Simpson at the USGS, and it was in a form suitable for our purposes.

The results of propagation in this "fast continents" model are shown in Figure 15; the sequence of snapshots should be self explanatory. Because of other demands on Convex memory and CPU time, we were able to model only the first 125 minutes of wave propagation. Apparent numerical artifacts near the tops and bottoms of some of the frames are illusory; they are due to the sparsity of the grid relative to the polar regions of the map projection used here.

Artifacts

Although few artifacts were visible to the eye in the homogeneous model, there are artifacts that reveal themselves upon closer inspection. In Figure 16 are shown some of the same frames as in Figure 14, but with the amplification boosted by a factor of 20. Two main sets of artifacts are visible. Frame (b) shows the most obvious ones: spurious diffractions off of the original 12 vertices of the icosahedron. These diffractions grow with time, due to a resonance that develops on the homogeneous sphere. We believe that on an inhomogeneous sphere, the resonance would be lost; in any case, these diffractions will probably be overwhelmed by diffractions due to any inhomogeneities.

Less obvious at first, but more disturbing, are the artifacts in frame (a). Besides the diffracted waves, a 5-pointed radial pattern is visible around the shot point. We are not sure of the cause of this artifact; it may be related to the anisotropy that will be discussed below. It is disturbing because, while it is not apparently a resonance effect, its strength increases with time. It becomes

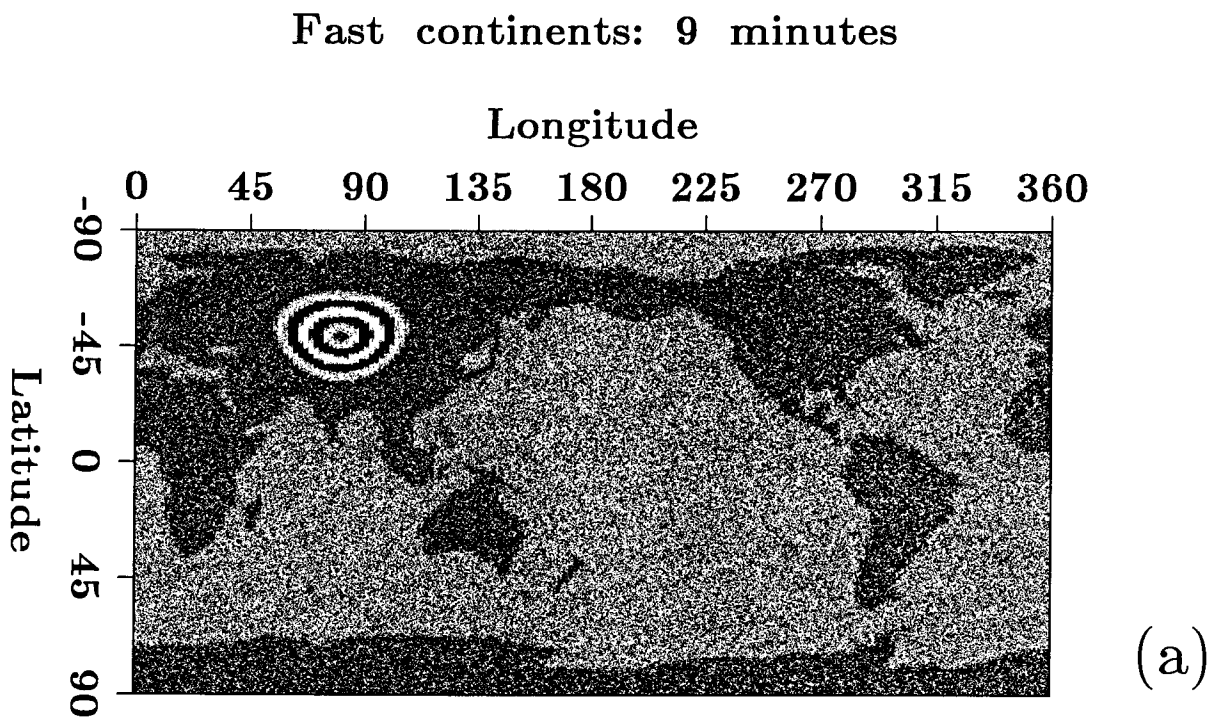


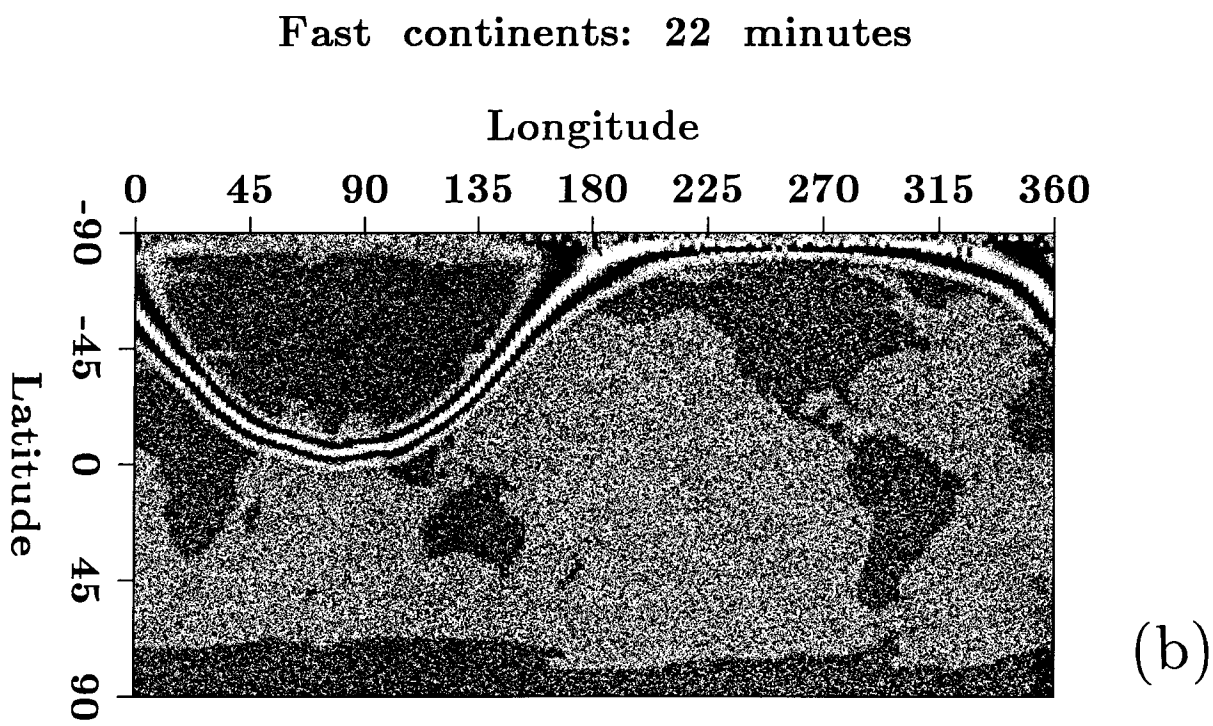
FIG. 15. Propagation on an inhomogeneous sphere. On this page and the following two pages are shown the results of propagating a wave on an inhomogeneous sphere. The velocity in the ocean is 10% lower than the velocity on the continents.

strong enough, in fact, to be noticed even on unamplified plots, if these plots are displayed on a device with more dynamic range than our laser printer.

Regularizing the grid

Systematic irregularities are visible on the recursively tiled sphere (Figure 2). There are well-defined areas with greater and lesser densities of vertices. These variations can be plotted, as in Figure 17. The brighter areas on this plot represent lower densities of vertices, while the darker areas represent higher densities. Another way of looking at this is to consider the bright regions to be where individual triangular faces have larger surface areas than average. These density contrasts present a problem. They represent irregularities in the gridding, and these irregularities can generate artifacts when the wave-field modeling is carried out. A significant problem is the sharpness of the density contrasts, since these sharp boundaries can produce unwanted reflections.

There are methods for removing these density contrasts. A least-squares problem was set up, with the goal of making the areas of the individual triangles as equal to each other as possible. The resulting regularized gridding scheme is shown in Figure 18. Note that density contrasts are



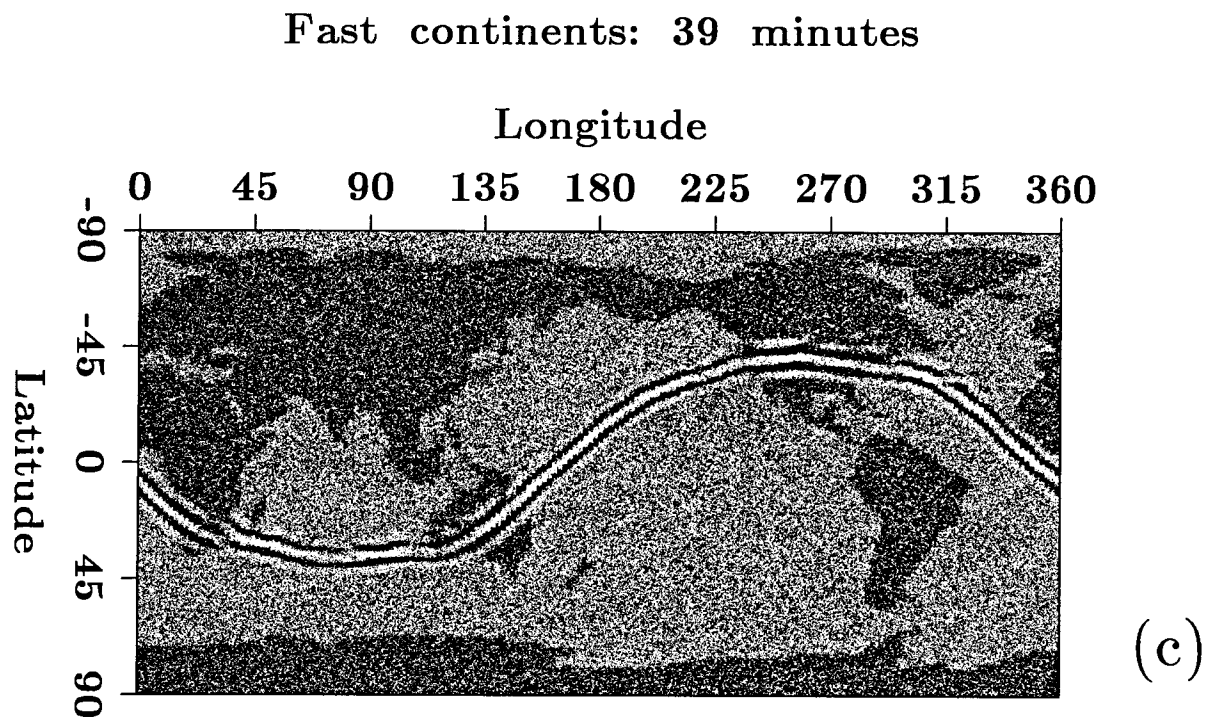
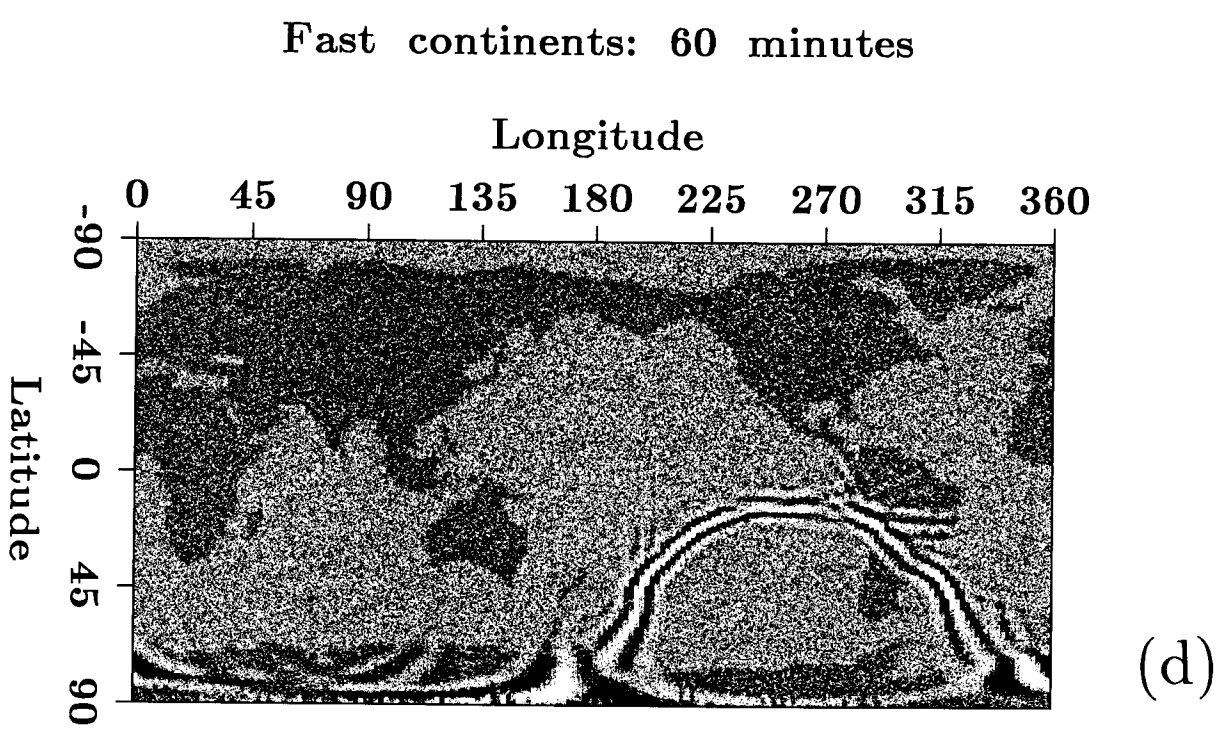


FIG. 15 continued.



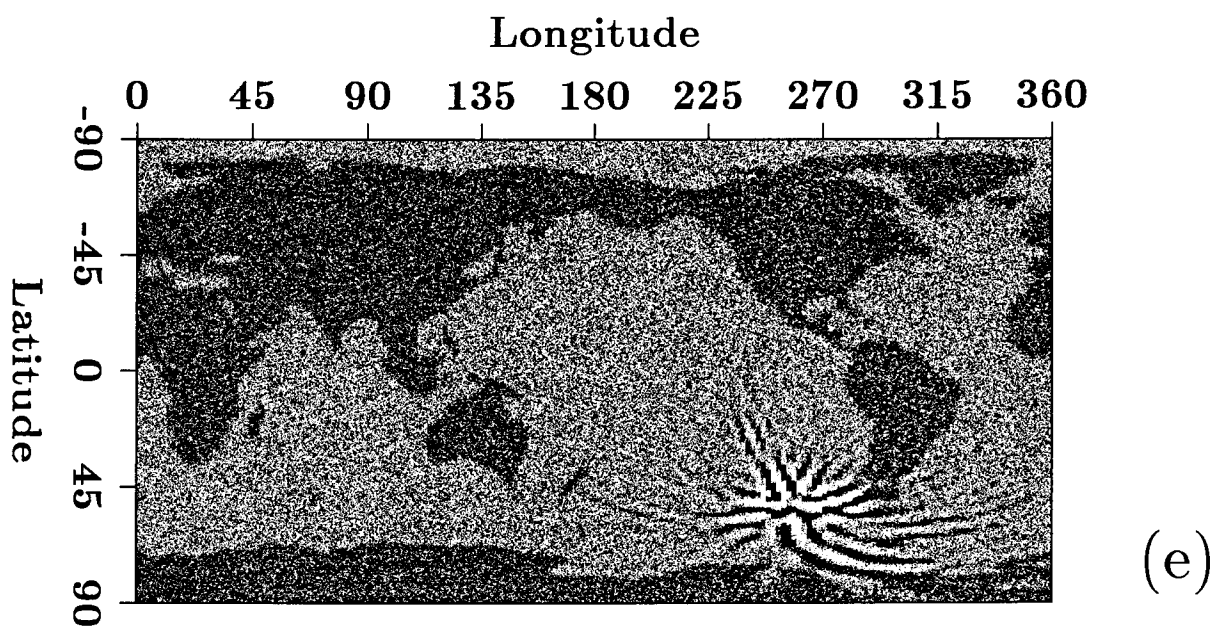
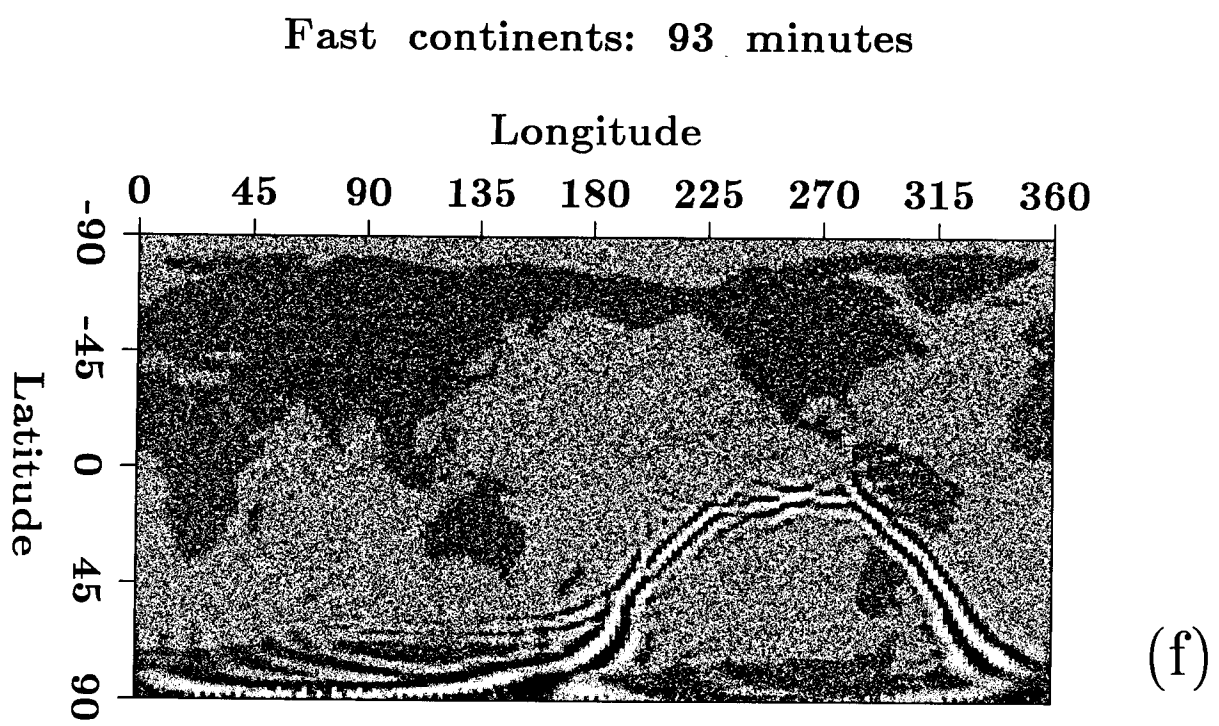
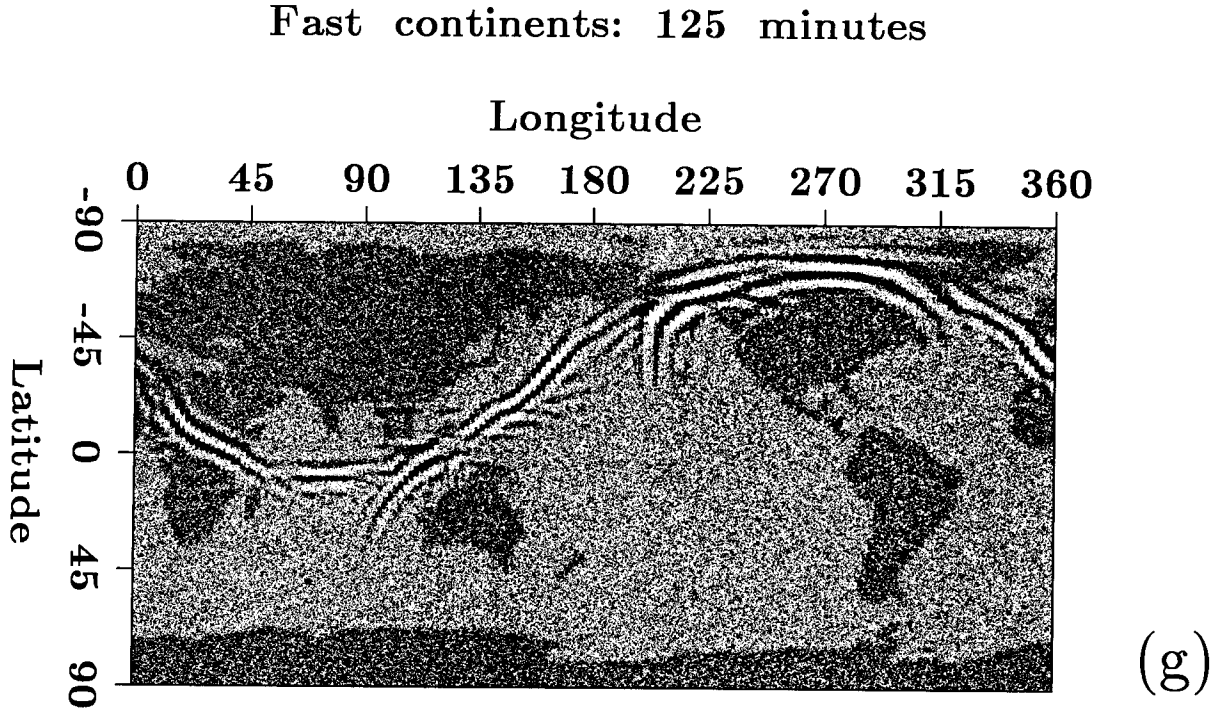


FIG. 15 continued.



(f)



(g)

FIG. 15 continued.

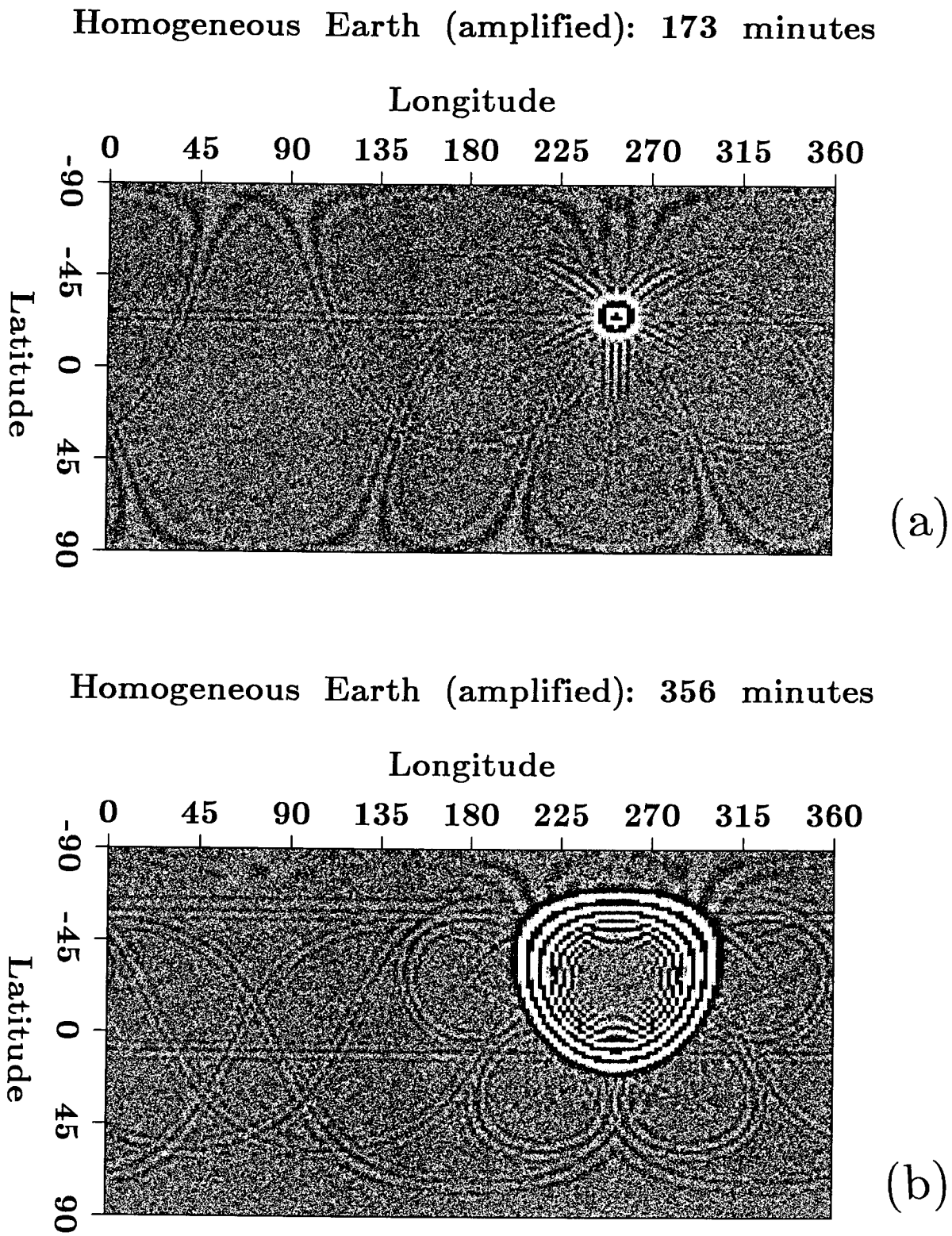


FIG. 16. Propagation on a homogeneous sphere, amplified. Two of the plots from Figure 14 have been amplified by a factor of 20 in order to make numerical artifacts visible.

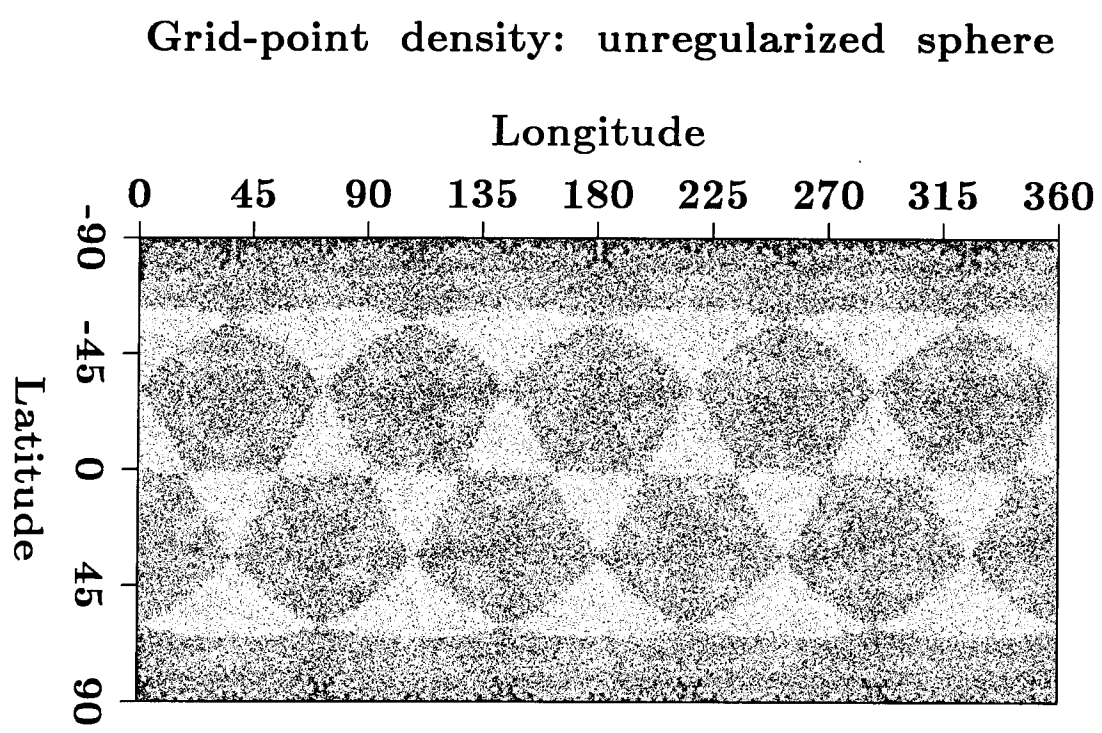


FIG. 17. Plot of area variation. Brighter regions show where the grid density is lower on the unregularized sphere.

no longer visible. If we were to produce a density plot similar to that of Figure 17, it would be uniformly gray.

Unfortunately, this regularization process produces side effects. In Figure 19 is presented a plot illustrating variations in edge length. In Figure 19 is plotted for each vertex the difference between the maximum and minimum distances to adjacent vertices. This plot may be considered to show the amount of local anisotropy. Figure 19 was computed from the original network, before the least-squares regularization process was applied. Figure 20 shows the same sort of plot, plotted at the same amplitude, for the network after regularization has been applied. Note that the regularization increased the measure of local anisotropy in some regions, while smoothing out sharp contrasts. The regions of increased anisotropy might be the cause of the artifacts noted in Figure 16.

FURTHER DIRECTIONS

Applications to 3-D migration

Reflection seismic data is often irregularly sampled. It is not always clear how to apply finite-difference methods, so Kirchoff methods are often applied instead. Some of the techniques

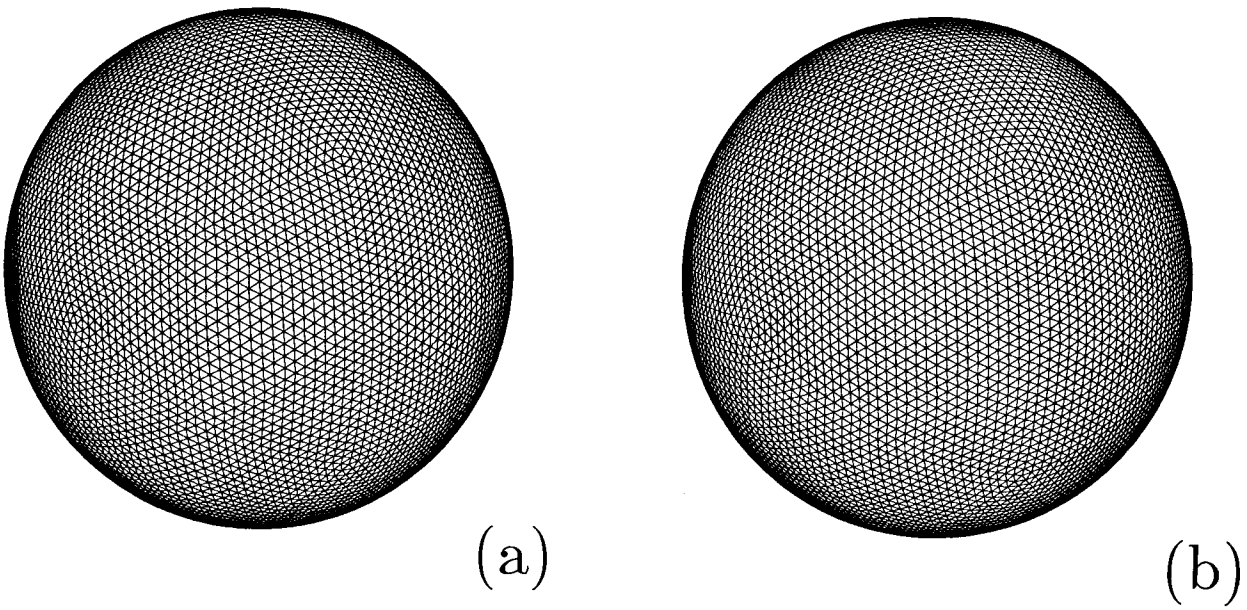


FIG. 18. Regularizing the icosahedral grid. The original icosahedral grid (a) can be regularized (b) to reduce sharp contrasts in the density of grid points.

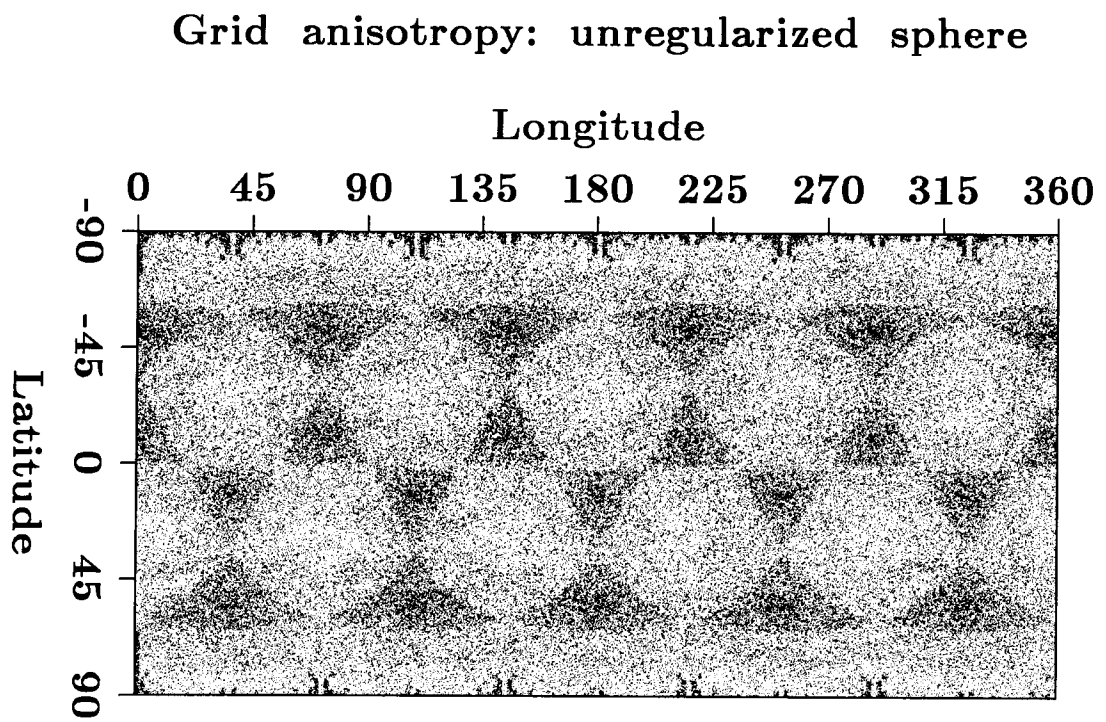


FIG. 19. Anisotropy on an unregularized grid. Brighter areas represent regions where grid points have nearest neighbors at varying distances. In darker areas, grid points have nearest neighbors that are about equally spaced.

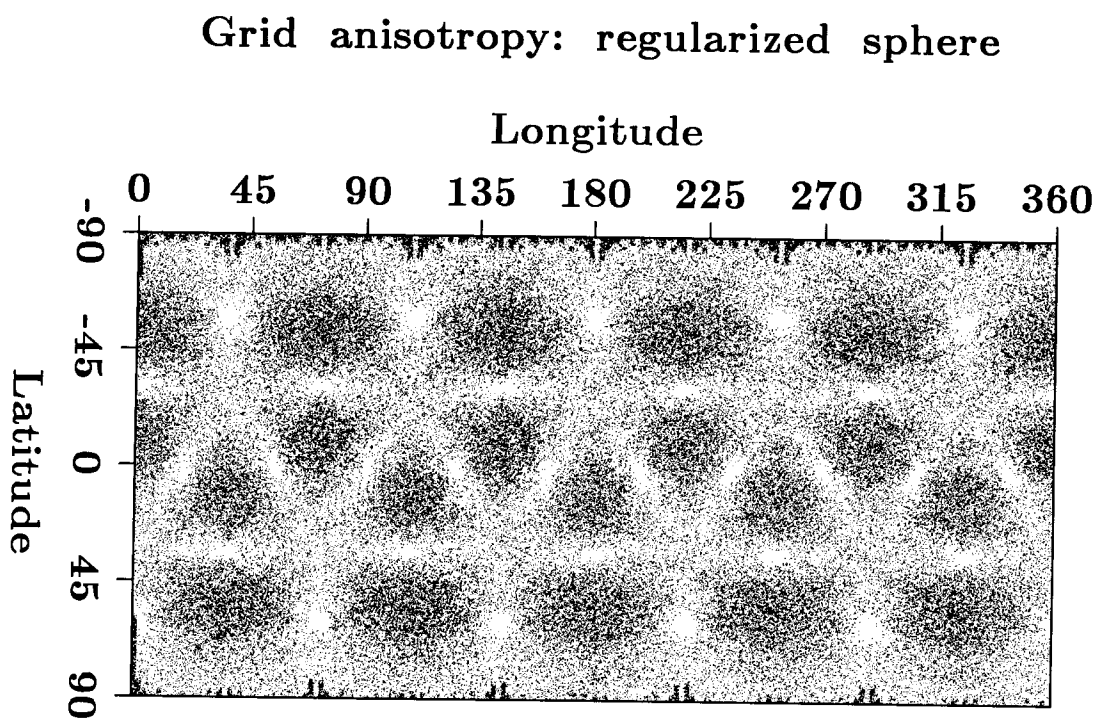


FIG. 20. Anisotropy on regularized grid. The brightness scale has the same meaning and amplification as for Figure 19.

developed in this paper might be applicable to the problem of performing finite-difference migration on irregularly sampled data. An unresolved question is whether it is better directly to process such data, or whether it is better to interpolate the data onto a regular grid, where conventional processing techniques can be applied.

It might be noted, by way of justifying the appearance of this paper in an SEP report, that the irregular-grid problem is best studied on a spherical shell, since there is then no need to worry about the complication of boundary conditions.

Fitting dispersion curves with 2-layer models

We noted previously that even with T, a, and $k_z{}^2$, we don't have enough adjustable parameters to fit observed dispersion curves very well. Perhaps we could obtain more parameters by going from a one-layer to a two-layer model. One potential difficulty is finding a coupling term between the two layers that is physically justifiable and numerically stable. Another is the possibility that such a model would allow the propagation of two modes simultaneously; this eventuality would not be desirable in a model that is supposed to emulate the propagation of only one mode.

ACKNOWLEDGEMENTS

We would like to thank the following people for useful ideas and discussions: Fabio Rocca, Rob Haar, Peter Mora, and Joe Dellinger. Bob Simpson at the US Geological Survey provided the topographic data, and Rick Ottolini loaded it onto our system. Toshiro Tanimoto and Rob Clayton provided the Preliminary Reference Earth Model. This research was performed on facilities made available through the generosity of the sponsors of the Stanford Exploration Project.

REFERENCES

Berg, P.W., and McGregor, J.L., 1966, Elementary partial differential equations: Holden-Day.

Claerbout, J.F., 1976, Fundamentals of geophysical data processing: with applications to petroleum prospecting: McGraw-Hill Book Company.

Dellinger, J., and Muir, F., 1986, Finite differencing with uneven spatial sampling: SEP–48, 269–275.

Dziewonski, A.M., and Anderson, D.L., 1981, Preliminary reference earth model (PREM): Phys. Earth Planet. Inter., **25**, 297–356.

Mitchell, A.R., and Griffiths, D.F., 1980, The finite difference method in partial differential equations: John Wiley & Sons.

Morse, P.M., and Feshbach, H., 1953, Methods of theoretical physics: McGraw-Hill Book Company.

Paige, C.C., and Saunders, M.A., 1975, Solution of sparse indefinite systems of linear equations: SIAM J. Numer. Anal., **12**, 617–629.

Pugh, A., 1976, Polyhedra: a visual approach: University of California Press.

Strang, G., 1986, Introduction to applied mathematics: Wellesley-Cambridge Press.

Tanimoto, T., and Anderson, D.L., 1985, Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s: Journal of Geophysical Research, **90**, 1842–1858.

Woodhouse, J.H., and Dziewonski, A.M., 1984, Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms: Journal of Geophysical Research, **89**, 5953–5986.

Yomogida, K., and Aki, K., 1985, Waveform synthesis of surface waves in a laterally heterogeneous Earth by the Gaussian beam method: Journal of Geophysical Research, **90**, 7665–7688.

Zienkiewicz, O.C., 1971, The finite element method in engineering science: McGraw-Hill.

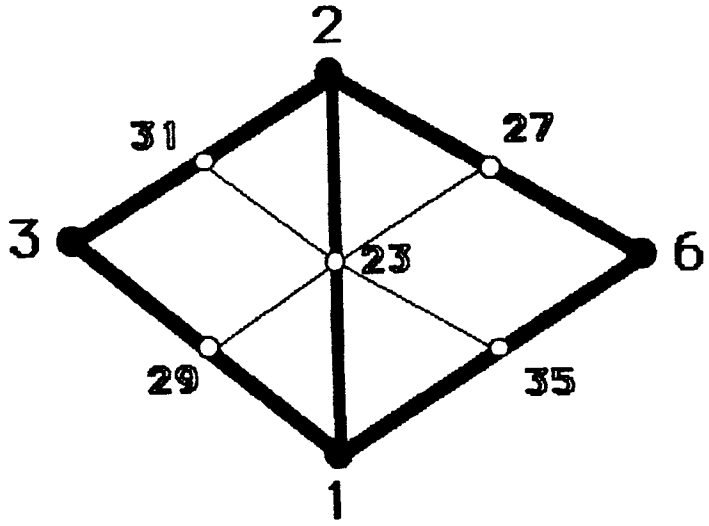


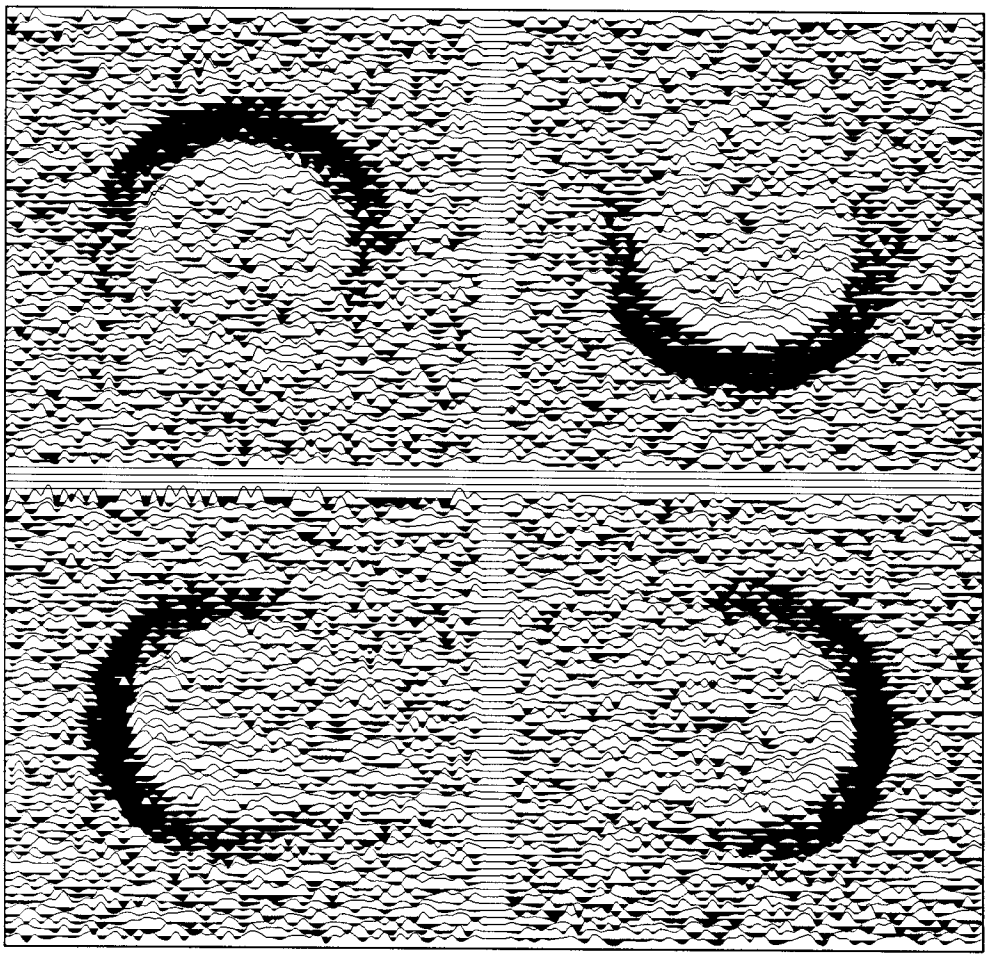
FIG. A-1. Recursion scheme. Shown is a hypothetical region of the grid. This Figure is used in the text to explain how the recursion scheme works.

APPENDIX A: CONSTRUCTING THE ICOSAHEDRAL GRID

We use a recursive algorithm to tile the sphere, since it is relatively simple in recursive schemes to keep track of nearest neighbors. It is much easier to keep track of nearest neighbors at each step of the recursion, rather than to try and find nearest neighbors afterwards on the basis of vertex coordinates. We use the icosahedron as the starting point of the recursion. The icosahedron has the advantages of having triangular faces, and of having more faces than any other regular polyhedron.

We begin with the result of the previous recursion (or with the original icosahedron, depending on which recursion level we are on). Each vertex on the grid is numbered, and there is a list giving the nearest neighbors of each vertex. As an example of how the recursion scheme works, let us look at two adjacent triangular faces (see Figure A-1). We place a new vertex on the midpoint of each edge (an edge is a line that connects two nearest-neighbor vertices, and a face is a triangle bounded by three such edges). Each of the new vertices is given its own identification number, and we can begin to make a new nearest-neighbor list.

First, we find new nearest neighbors for the old vertices. For instance, suppose, as in the Figure, that vertices 1 and 2 are nearest neighbors, and that on the edge connecting 1 and 2 we have placed a new vertex, number 23. Vertex 2 no longer has 1 as a nearest neighbor; vertex 1 has been replaced as a neighbor by vertex 23, and this should be noted on the new nearest-neighbor list. Similarly, 2 is replaced by 23 as a nearest neighbor of vertex 1. Thus, it is a straightforward matter to find the new nearest neighbors of old vertices. It is also easy to find two of the nearest neighbors of the new vertex: they are the two nearby old vertices. For instance, in Figure A-1, two of the nearest neighbors of vertex 23 are vertices 1 and 2.

Somewhat more complicated is the task of finding the other four nearest neighbors of the new vertices. To find these neighbors, it is necessary to refer to the old list of nearest neighbors. Assume we are looking for the new nearest neighbors of vertex 23 in the Figure. Before the subdivision took place, vertices 1 and 2 were both nearest neighbors of vertex 3, so that vertices 1, 2, and 3 laid at the corners of a triangular face. Similarly, vertices 1 and 2 were both nearest neighbors of vertex 6, thus implying the existence of a triangular face bounded by the edges connecting vertices 1, 2, and 6. Given the old nearest-neighbor list, what is the quickest way to determine that vertices 3 and 6 each have vertices 1 and 2 as nearest neighbors? Look at the list of the nearest neighbors of vertices 1 and 2. Only vertices 3 and 6 appear on *both* lists. The importance of vertices 3 and 6 should be obvious from the Figure: the new vertices that are constructed on edges 1-3, 1-6, 2-3, and 2-6 are the remaining nearest neighbors of vertex 23. Thus, by refering to the old nearest-neighbors list, and by keeping track of which new vertices are placed on which edges, it is not difficult to find the four remaining nearest neighbors of the new vertices.

A second list must be kept, in addition to the nearest-neighbors list. This second list contains the spatial coordinates of each point. It might seem that this list should be based on a spherical system of coordinates, since we are, after all, working with a sphere. We have found, however, that it is more convenient to work in a Cartesian (x-y-z) coordinate system, with the origin at the center of the sphere. Returning to the previous example, suppose we know the Cartesian coordinates of vertices 1 and 2; vertex 23 is then assigned the coordinates of the midpoint between vertices 1 and 2. This does not put vertex 23 on the surface of the sphere, so the coordinates of vertex 23 should be multiplied by a constant such that the vertex is relocated onto the surface. This multiplication by a constant should be done before the next recursion step; otherwise, subsequent new vertices will not be positioned as evenly as they could be.