Visualization of Seismic Data in Geophysics and Astrophysics

Joanna Leng[1], John Brooke[1], Terry Hewitt[1] and Huw Davies[2]

[1] CSAR, Manchester Computing, University of Manchester
[2] Jane Herdman Laboratories, Dept. Earth Sciences, University of Liverpool
email: j.leng@man.ac.uk, j.m.brooke@man.ac.uk, W.T.Hewitt@man.ac.uk, davies@liverpool.ac.uk

Abstract

Seismic data obtained from surface observations of the Earth and the Sun has enormous importance in increasing our knowledge of the internal structure of these objects. By making special assumptions (e.g. spherical symmetry) it is possible to use analytic tools or relatively simple numerical calculations to analyze the tomographic data. In both cases, however, the real problem is inherently three dimensional and to resolve detailed structure enormous amounts of data are required. We show in this paper how visualization can aid the process of interpretation. The spherical geometry of such objects, and the huge variations in physical quantities with depth, present formidable challenges to current visualization systems. We present a major case study using AVS/Express to visualize some of the largest geophysical simulations ever created and show how we intend to extend these techniques to astrophysics.

1 Seismic Data in Geophysics and Astrophysics

1.1 Importance of seismic data

Seismic data is of great importance in both geophysics and astrophysics, as it constitutes one of the main methods of deducing the physical structure and composition of the interior of a planet or star. Seismic tomography utilizes information obtained by recording the arrival times of seismic waves at different points on the Earth's surface from a known shock epicenter, such as an earthquake. By making the assumption that conditions in the Earth vary with radius only, it is possible to build a model of the interior of the Earth from seismic data that can show features such as the boundary between the Earth's core and mantle and between the solid inner core and fluid outer core. It is a remarkable intellectual achievement that this was done as far back as the 1930's, without the aid of electronic computers, by using the physical insights of an elegant theory. However for studying variations from spherical symmetry that give rise to, for example, plate tectonics, it is necessary to apply tomography in a fully 3-dimensional simulation.

Another, more recent application of seismology has been the use of helioseismology to deduce the rotational structure of the sun's convection zone. Helioseismology is made possible by advances in astronomical instrumentation and techniques that can detect tiny changes in the emission or absorption spectral line due to a given ion, caused by the arrival of waves at the solar surface.

The problem of deducing conditions throughout the interior of a volume from measurements obtained at the surface is a classic example of an inverse problem. Inverse problems are typically ill conditioned in both a mathematical and physical sense. This means that tiny error bars in the surface measurements lead to large uncertainties about conditions in the interior, and that the reconstruction depends on certain physical assumptions which are also uncertain and lead to large variations in the structures deduced from observations. In the case of helioseismology, for example, the convective mixing length is completely uncertain and this is an important parameter in reconstructing the solar interior.

1.2 Why visualization is both important and difficult for seismic data

Given the complexity of the tomography problem, the physical insight and intuition of the modeller can very often be crucial. Visualization can play an important role, since it allows the representation of large numerical data sets in 3D in a manner related to the geometry of the object being studied. In both geoseismology and helioseismology this geometry is that of a sphere, or spherical shell. This poses severe challenges to visualization software. There are complex effects due to curvature and the convergence of points equally spaced in latitude and longitude as the radial coordinate decreases from the surface to the center. We discuss these issues next in the context of a particular case study.

2 Case Study: Seismic Tomography and Convection Modeling of The Earth's Mantle by The Terra Group

The UK base of the Terra Consortium is at the University of Liverpool; they run computational simulations of the Earth mantle's (the layer between the crust and core), which run on 512 processors of the Cray T3E for up to 12 hours. The study of mantle currents are important for understanding both continental drift as well as volcanic and seismic activity.

The mantle is a relatively thick layer, which extends nearly halfway to the centre of the Earth. The spherical shell being modeled is much thicker than the shells used for ocean or atmospheric modeling, it is 2900 km compared to the 100 km of the atmosphere (the radius of the Earth is 6370 km). The mantle is modelled as a thick viscous liquid, usually by finite element analysis, the ocean and atmosphere on the other hand are usually modelled by finite difference methods. The mantle is crucially different from the ocean and atmosphere in that it is a thick shell. Thus the effects of curvature are fundamental to the visualization. The shape and structure of this spherical shell makes it closer to that of astrophysical simulations of stars or planets rather than ocean or atmosphere modelling.

Seismograms from just over 3000 seismic stations resulting from around 10,000 earthquakes have been collected. The travel times of these seismic observations are inverted by velocity tomography so the group can compare the lateral variations in seismic velocity produced with the lateral temperature perturbations of the mantle modeling. In this way they can verify the simulation, by comparing theory with observation. The problems can be classified into three main areas, modelling this spherical geometry, visualizing it in 3D and comparing observational data with the result of simulations.

2.1 Modeling This Spherical Shell

The solution to the inversion of seismic data dates back to the 1930's. Since the variations are dominantly radial the model is effectively reduced to 1D, depth is the only factor. Given this history it has been quite common to analyse the data by 2D projections (Fig 1).

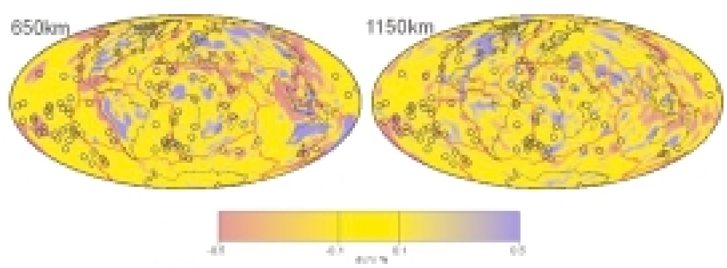

Fig. 1: 2D projections for two shells of given depth.

However academic peers are now demanding 3D analysis of what are truly 3D problems (both the tomography and the simulation).

2.2 3D Visualization

We have concentrated on visualizing the tomography data initially. The group converts the simulation data to the tomography format, and hence can visualize both types of data. We will look to visualize the unconverted simulation data in future work. The data mesh for the tomography is made of cells. The cells are not defined by array data as they are for atmospheric modeling but by a structure. The data cells are defined in a spherical coordinate system so implicitly they have curved surfaces. Visualization systems use a Cartesian coordinate system and straight lines to define all surfaces. To transform the simulation cells into renderable objects they must be resampled in Cartesian space so they tessellate and fit the original curvature.

The complexity of the data and of visualization systems mean that this software is best produced by a graphics expert rather than a geophysics researcher. The number of cells required to visualize the data mean the results will be slow to manipulate unless specialist graphics hardware is used.

3 Technical aspects of the visualization

3.1 Special Challenges of Visualizing Thick Spherical Shells

There are several reasons why this type of data is difficult to visualize.

- Graphics renderers only handle straight lines, spheres have few straight lines, are defined by spherical coordinates and to get well formed tessellated cells extra cells must be introduced.
- The number of cells needed to visualize this data means special data reduction/handling techniques are required.
- Placement of reference information is important since this allows the researcher to relate features inside the sphere to surfical features. The user needs "volcanic" hot spots (e.g. Hawaii, Iceland), tectonic plate boundaries and coastlines to be present in each image.
- The shell is thick and located close to the centre of the sphere, the angular component of the data is more important than it would be for atmospheric data.
- Perceptual problems with understanding a spherical shell:
  - A sphere looks the same from every angle so it is difficult to orientate.
  - Perspective has little effect.
  - Colour, transparency and shading sometimes work to make the pseudo-colour difficult to interpret. There is stronger darker band of colour around the outer circumference of the sphere.
- Problems with depth:
  - A 2D projection of one layer gives no depth information, is distorted and does not link around on itself.
  - The original visualization only related to depth slices. When seen in 3D many depths are seen at the same time, the depth layers need a new normalization function to make 3D images valid.
  - A cut plane through the sphere shows how the data varies with depth but a plane is 2D and loses 3D information i.e. data with an angular component is lost.
  - Internal data with an angular component may be self-obscuring, difficult to position because of problems with perspective and there is ambiguity in its relation to the reference data.

3.2 Visualization of The Earth's Mantle As A Spherical Shell

A C structure defines the cells in spherical coordinates. The organization of data cells is shown in Fig. 2, which shows the cells are placed in rings around each latitude band. The bands at this resolution are $5°$ and the number of cells in each band increase from the pole to the equator; the volume of each cell is equal. There are higher resolution tomography data sets with 1° latitude bands, which has 25 as many data cells as the 5° data. The convection simulations give finer resolution, approaching the equivalent of 0.4° bands. The coordinates for each cell are defined in a spherical coordinate system (latitude, longitude and radius) and so have implicitly curved edges. There are three cells at the pole and these are in the shape of curved triangular prisms while the rest of the cells are curved hexahedrons (rectangular prisms).

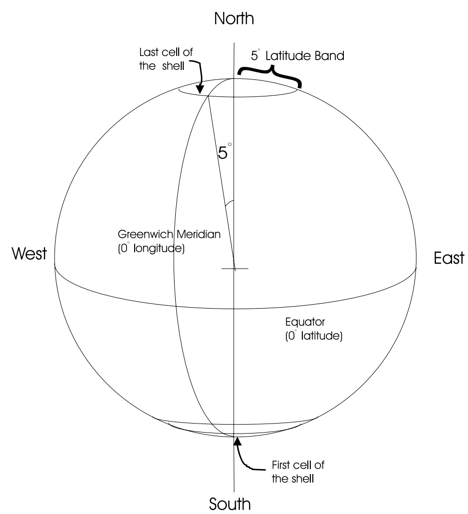

Fig. 2: A schematic diagram to show how one shell of tomography cells are arranged.

AVS/Express was used for this application. A number of modules were developed in C and C++ with the Developers Edition of the software. User interfaces were designed for these modules and application networks built so they could be used at the Terra group's site in Liverpool using the Visualization Edition of AVS/Express.

Cells used by a visualization system are defined as being of a particular type, e.g., prism or hexahedron. A coordinate is given for each apex of the cell, then the connections between each coordinate are specified and finally a data value is given for each cell (cell data) or for each apex (node data).

Initially the data was mapped to just the cells explicitly defined by the C structure (Fig. 3). Coordinates were calculated for the apex of each cell in turn even though neighboring cells would have common apex locations. The coordinate of each apex would appear at least twice in the coordinate array, once for each cell with a shared apex. Not surprising the resulting cells did not tessellate, gaps appeared between most latitude bands. More interestingly there was a banding effect across the surface caused by light shading. The shading had two causes. Between latitude bands not only were there gaps but there were cell overlaps which stopped the surface being smooth. Secondly there was a shading effect between cells of the same band, this was caused by two adjacent cells with shared apexes referencing different coordinates on the coordinates array although they should be the same point in 3D space. The reason for this seems to be the inaccurate nature of floats and coordinates are stored as floats. The result was very unsatisfactory; the solution was to increase the number of cells but to keep the number of coordinates as low as possible.

3.3 Resampling The Cell Sets

The resampling method removed all duplicate 3D Locations from the coordinates array. Coordinates were calculated for the cells of a whole latitude band so neighbouring cells with common apexes would never reference difference points on the coordinate array for the same point in 3D space. This removed the shared apexes within a latitude band from the coordinate array but some others remained. The first cell on each band begins at the same longitude and will share an apex with the bands above and below. The bands that lie on each side of the equator share all their adjoining apexes. It is a default throughout that when there are two possible coordinates that the one from the upper band will be used.

3.4 Neighbouring Bands With A Harmonic Number Of Cells

Occasionally adjacent bands would both have a number of component cells that would be divisible by the same small integer, e.g., in the low resolution data the second band has 3 cells and the third band has 12 cells, both are divisible by 3. This effect causes an incidentally shared apex with its known shading problems. In these cases the coordinate from the upper band is selected.

3.5 Resampling The Poles

In all resolutions of the data the latitude band at the poles consists of 3 triangular prisms and the next band consists of 9 hexahedron. For each pole the cells were resampled to give 9 triangular prisms, 3 for each of the original prisms. The coordinates were used from the neighbouring band whether upper or lower.

3.6 Resampling The Hexahedrons In All Other Latitude Bands

Each cell from the other latitude bands should consist of one curved hexahedron. This does not tessellate because it is impossible to give it curved edges they must be straight. Each shell has the same number of cells, they tile up on each other so extra cells only need to be added within each shell. The solution must be scalable so it can be extended to higher resolution data.

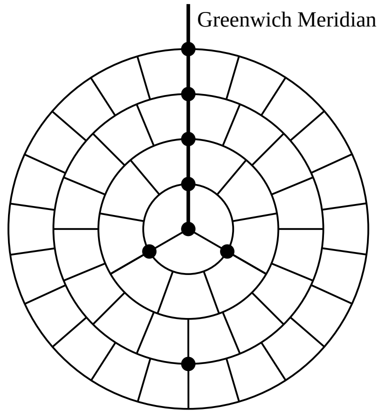

Fig. 3: Shows the first four latitude bands, viewed from the South Pole, the dots show where there are duplicate 3D locations in the coordinates array.

The resampling solution was to map one hexahedron for all of the original simulated data cells but not to map it to the full extents of the cell and to add triangular prisms to fill in the gaps. There were gaps being caused by cells in the band immediately above or below not having apexes exactly in line with the cell being looked at. The number of cells increases in each band from the South Pole to the equator and then decreases up to the North Pole. A solution is symmetrical about the equator so now we can just consider the geometry below the equator.

Each data cell consists of one hexahedron and a number of triangular prisms. The hexahedron is cut so it changes from its initial rectangular prism shape to that of a trapezoid prism and triangular prisms are added to give the cell a curvature that tessellates with the cells in the latitude bands above and below. All hexahedrons will keep four of their original apexes but some may keep six.

Imagine you are looking at the surface of a hexahedron so it looks like a rectangle. Now imagine the cells in the band above, they are smaller and there may be one or two which lie across the top of our cell. Our hexahedron must curve at the apex of each of the cells above if it is to tessellate. To do this we cut our cell from the apex of the cell above to the bottom left apex of our cell giving us one triangular prism and a trapezoid prism. If there are two apexes above we cut twice and get two triangular prisms and one trapezoid prism. The cells in the band below are larger and also do not align with our cell. We must make our trapezoid prism curve to tessellate with the cells below. This time we cut the cell from the last apex above to the first apex below, we keep a trapezoid prism but adding another triangular prism.

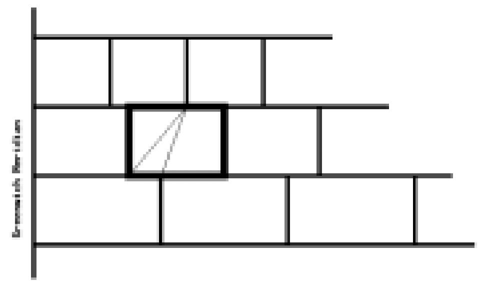

Fig. 4: A diagram showing the resampling scheme. It is the surface of a group of cells in 3 latitude bands projected in 2D. The cells appear rectangular; they are below the equator so each band has progressively more cells. We are looking at the cell with the thickened border; the thin line shows how it is cut into two triangular prisms and a trapezoid prism to make it curve with the bands above and below it.

3.7 Reading Cell Sets For Better Display And Data Management

The cell sets produced by the method given above gave a smooth evenly shaded sphere. Unfortunately the whole data had been resampled so it contained about 3 times the number of cells. It stopped being interactive and problems with visualization methods become clearer.

One of the most common techniques used in visualization is the bounding box, it shows the extents of the data. For a spherical shell the bounds are not easy to detect, the extents of the data are clear when it is seen as the sphere's surface but one of the extents are completely embedded within the other, i.e. the innermost shells is entirely within the outermost one.

All the shells are needed for visualization but removing cells as early in the pipeline as possible increases interactivity e.g. when the sphere is cut the data values across the plane must be calculated but after cutting any data cells that are completely internal can be removed. Separating the outermost shell and innermost shell from the rest is helpful as they can then be used to mark the data's bounds.

To achieve both of these objectives the innermost shell, the outermost shell and the rest of the shells were placed in separate groups of cell sets. Each of these three groups could then be separated or kept together and passed down different visualization pipelines to produce the most appropriate images.

3.8 Reference Data, Pseudo-Colour and Data Normalization

The tomographic data of the Earth's mantle is strongly related to other information, volcanic "hot spots", tectonic plate boundaries and coastlines. It

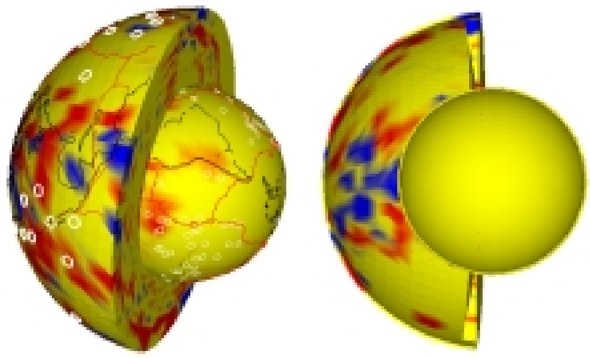

Fig. 5: These figures use the three groups of cell sets one for the inner most shell, the outermost one and one for all the other shells. The picture on the left shows how these can be used to create a visualization while the picture on the right shows that the internal data was removed early in the visualization pipeline to make it more interactive.

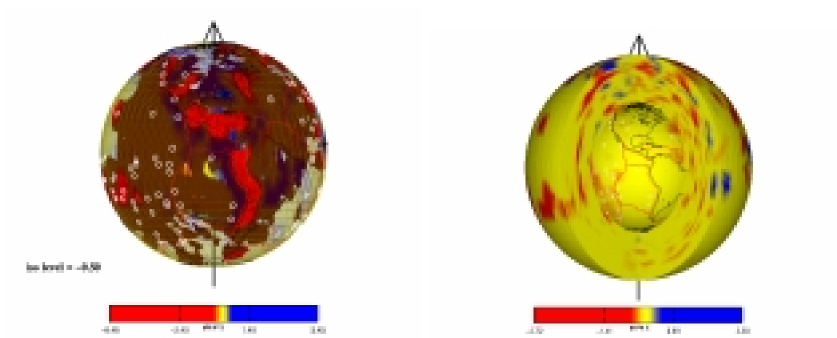

Fig. 6: Two further visualizations. The picture on the left is of an isosurface, the outer most shell is semi-transparent while the inner most shell is opaque and shows the bounds of the data. The picture on the right shows how the inner most shell is used to display reference data that would otherwise be cut away.

is vital that this reference data is always visible and unambiguously displayed.

For tectonic plate boundaries and coastlines a simple polyline could be used to join the appropriate coordinates into a line. The "volcanic" hot spots are traditionally viewed as a ring centering on the location of the hot spot on the Earth's surface. To achieve this an equation was used to calculate 12 points around each hot spot. All of the reference data is stored in spherical coordinates and then converted to Cartesian space. The equations used to ring the hot spot works in the spherical domain, there are two equations to do this. Unfortunately each equation has discontinuities at certain spherical coordinates, the problem areas, either around the pole or at the equator. A patch is needed between equations to make the rings approximately circular at all locations on the Earth's surface.

The colour used for all the data, the simulated data and the reference data and the background adheres to the standard colour representations of the field of mantle studies. The tomography data is given as percentage velocity perturbation, which means the data is tightly clustered around the zero value. Many cells have a zero; i.e. have a zero perturbation or are at the average value. The scientists interest is in the data close to zero either just faster or slower than average so a saturated colour scale is used at the extents of the data values that rapidly changes around zero.

A new normalization function was added to give the root mean square of data values for each layer. The colour map has been designed to highlight data values close to but not at zero. The lower layers have data values more closely clustered around zero than the outer layers. The normalization function makes the variation in the inner most layers as prominent as the outer most one while still using the same colour map for all layers.

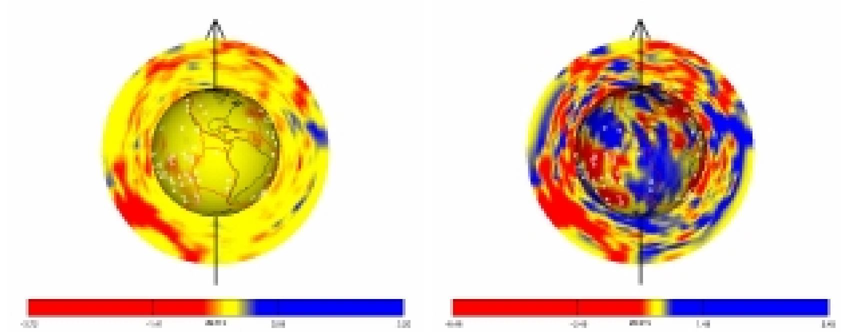

Fig. 7: The image on the left uses data, which is not normalized while the one on the right uses normalized data.

4 Special Problems of Large and Complex Datasets

4.1 General Issues

There is increasing realization that high-end computing brings special challenges with regard to the storage, movement and archiving of the very large data sets that simulations using hundreds of processors can produce. One can view such data sets as a challenge in several ways: order of magnitude terms, volume of data, numbers of files, cataloguing issues etc... However the output of these simulations also raises problems in relation to the extraction of information and insight, which cannot be so easily evaluated.

Our contention is that visualization techniques can help with the process of the extraction of information from large-scale simulations. This in turn can help with decisions about what data needs to be stored and to how the simulations are set up and run. We are optimizing the simulation process by providing tools to harness the physical insight of the scientists running the simulation. The geometry of a spherical shell presents novel challenges, which are interesting both in terms of mathematics and also human perception. It is known that to envisage global movement on a spherical surface is challenging, this becomes more so when questions of depth and varying radius come into play. We have drawn attention in section 3.1 to questions of normalization. Many physical quantities (temperature, pressure, velocity of sound) are strongly radially dependent in spherical objects and this presents challenges to the visualization software. It is important to provide solutions which can be readily adapted by the end-user, questions about appropriate normalizations, use of symbolic referencing information (e.g. colour, symbols, superimposed outlines) needs to be decided by the scientists for the detailed interpretation of their simulation results.

How the data from a simulation is then stored depends on many factors. If a group has sufficient storage they may decide to store all data for analysis over an extended period. However, it will be more usual that some trade-off needs to take place between the resource implications of storing data and the desire to extract from it as much scientific benefit as possible. It is here that the ability to make informed decisions can be potentially advantageous. While some aspects of data reduction can be automated, e.g. by compression techniques, it is not currently possible to automate the intuition and judgment of expert scientists.

4.2 Choice of a Visualization Tool

The tomography data is cell-based. As discussed previously this places special demands on the visualization software. Our choice of AVS/Express was determined partly by the flexibility it provides in the handling of different types of data, with complex geometrical reference. Such data is not unique to tomography, most finite element simulations also produce complex data where spatial connectivity is an important issue. Such data sets are also large and will grow larger. The demand is for finer resolution simulations that capture more detailed physics because they keep up with the rate of increase in processing power (both parallel and serial). Whereas a highly trained scientific imagination can handle the problems of tomography with radial variance, this becomes impossible in 3D with the wealth of detail produced by modern simulations.

We have become aware in our case study that visualization methods for data in a spherical geometry are relatively underdeveloped. Nearly all visualization techniques relate to a Cartesian coordinate system. Also there are complex problems of perception of geometrical objects within a sphere rather than the usual implied bounding cuboid appropriate to many visualization techniques.

Seismic data in a sphere brings these challenges into sharp focus. However the use of a flexible and programming tool such as AVS allows some fascinating experiments. For instance one can visualize the mantle looking from the core upwards and from the surface downwards. Isosurfaces extend deep into the sphere and the effects of curvature and a diminishing horizontal length scale can be studied. These opportunities challenge and develop the scientific intuition in ways which 2D slices cannot.

4.3 Further Work

The classical use of the term "debugging" relates to the process of finding the errors in a program. There are increasingly sophisticated techniques for examining, monitoring and altering the values of variables as the program is being run. However, with the scale and complexity of modern simulations it is often impossible to do this effectively because there is so much numerical data. Visualization can help because anomalous geometric and physical effects can be seen directly. This can give clues as to where the simulation is going wrong. In a multi-physics simulation it may be possible via visualization to localize the problem to particular physics modules for example. When visualization systems are linked to the code so they can halt execution, change variables and parameters the simulation is steered and the physics experimented with. This extends the concept of debugging to include a stage where the code is tested to see that it produces results, which agree with experiment or with other firmly established theoretical results.

We intend to generalize the techniques described here to other areas. Spherical geometry is clearly of vital importance in the study of the interiors of both planets and stars. We will continue our work by studying the convection zone in the sun [6].

Acknowledgements. We thank SGI for the loan of an Octane workstation, as part of SGI's ongoing contribution to providing high-end computational facilities to UK research [7].

References

1. Davies, JH and HP Bunge, "Seismically 'fast' geodynamic mantle models", submitted to Geophys. Res. Lett.
2. Bunge HP and JH Davies, "Tomographic images of a mantle circulation model", submitted to Geophys. Res. Lett.
3. Rhodes, M and JH Davies, "Global tomographic detection of mantle plumes in the uppermost lower mantle", submitted to Geophys. J. Int.
4. http://mantle.esc.liv.ac.uk/davies/AGU99/
5. Advanced Visual Systems Inc., "AVS/Express Developer's Reference, Release 3.0"
6. Wilson, P.R., Solar and Stellar Activity Cycles, Cambridge University Press, 1994
7. http://www.csar.cfs.ac.uk/