Stanford Exploration Project, Report 114, October 14, 2003, pages 169–175

Short Note

More fitting equations for PEF estimation on sparse data

William Curry[1]

INTRODUCTION

Prediction-error filters (PEFs) can be used to interpolate data (Spitz, 1991; Claerbout, 1999; Crawley, 2000). In order to generate a PEF, a least-squares problem is solved where regularly-sampled data is convolved with an unknown PEF. When estimating a PEF with missing data, the equations that contain missing data can be eliminated from the inversion. When dealing with sparse data, however, all equations may contain missing data, so a PEF cannot be estimated.

This problem has been addressed by regridding the data to multiple different scales, and then introducing those coarser copies of the data to the problem to gain more fitting equations (Curry and Brown, 2001; Curry, 2002). When creating coarser copies of the data, there is freedom in choosing the method used to regrid the data, as well as the choice of the coarser grid itself.

Existing methods use multiple different grid sizes in order to extract more information from the sparse data. More fitting equations can be generated by not only varying the size of the grid, but also by varying the positioning of that grid, i.e., the location of the origin of the grid. As the grid becomes coarser, the number of possible grid positions increases, and as the dimensionality of the problem increases, that number increases further.

By varying the grid location, a PEF can be more accurately determined on sparse data. 2D and 3D examples are shown for non-stationary PEF estimation, with a noticeable improvement in the interpolated result.

BACKGROUND

A PEF can be estimated by solving the minimization problem where known data (**d**) is convolved (**D**) with an unknown PEF (**f**), so that

$$\mathbf{W}(\mathbf{DKf}+\mathbf{d}) \approx \mathbf{0}, \tag{1}$$

[1]**email:** bill@sep.stanford.edu

where **W** is a weight for missing data and **K** constrains the first PEF coefficient to be 1.

When all of the equations contain missing data, **W** is 0 everywhere, and the problem cannot be solved. In this case, rescaled copies of the data can be substituted for the original data in equation (1), resulting in

$$\mathbf{W}\left(\begin{bmatrix}\mathbf{D_0}\\ \mathbf{D_1}\\ \mathbf{D_2}\\ \dots\\ \mathbf{D_n}\end{bmatrix}\mathbf{Kf}+\begin{bmatrix}\mathbf{d_0}\\ \mathbf{d_1}\\ \mathbf{d_2}\\ \dots\\ \mathbf{d_n}\end{bmatrix}\right)\approx\mathbf{0}. \tag{2}$$

In this case, $\mathbf{d_i}$ represents the various different rescaled copies of the data, $\mathbf{D_i}$ is convolution with that rescaled data, and **W** is now a weight for all scales of data. The data is rescaled by taking the original finely gridded data, and transforming it to a series of points located at the center of cells with data. Adjoint linear interpolation is then performed to move the data points onto the new grid.

For the case of a non-stationary PEF, the equations remain largely the same, except that the PEF varies with position. When convolving different sizes of data with a non-stationary PEF, the PEF must be sub-sampled so that the spatial dimensions of the non-stationary PEF and the data match. This is accomplished by the introduction of a sub-sampling operator $\mathbf{P_i}$, so

$$\mathbf{W}\left(\begin{bmatrix}\mathbf{D_0}\\ \mathbf{D_1}\\ \mathbf{D_2}\\ \dots\\ \mathbf{D_n}\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ \mathbf{P_1}\\ \mathbf{P_2}\\ \dots\\ \mathbf{P_n}\end{bmatrix}\mathbf{Kf}+\begin{bmatrix}\mathbf{d_0}\\ \mathbf{d_1}\\ \mathbf{d_2}\\ \dots\\ \mathbf{d_n}\end{bmatrix}\right)\approx\mathbf{0}. \tag{3}$$

In addition, since the model space has now increased substantially, a second fitting goal must be added,

$$\epsilon\mathbf{Af}\approx\mathbf{0}, \tag{4}$$

that ensures that the PEF will vary smoothly over space. In equation (4), **A** is a regularization operator (in this paper, a spatial Laplacian), and ϵ is a scale factor.

GRID SHIFTING

Instead of simply taking multiple scales of data ($\mathbf{d_i}$), multiple different grid locations can be used for each scale, as there is freedom in how they are chosen. For the case of a grid twice as coarse as the original fine grid, there would be 4 possible grid locations for a 2D case, shown in Figure 1. This number is inversely proportional to the size of the bins, and increases exponentially with dimension. However, in the case of irregular traces, there is no point in shifting the grid along the time axis, as the number of equations would be unaffected. In the case of a non-stationary PEF, extra bookkeeping is required, as the different grid locations will correspond to different regions of the non-stationary PEF. This does make using these extra versions of the data more expensive, as the weight **W** and sub-sampler **P** are not the same at one scale and need to be recomputed.

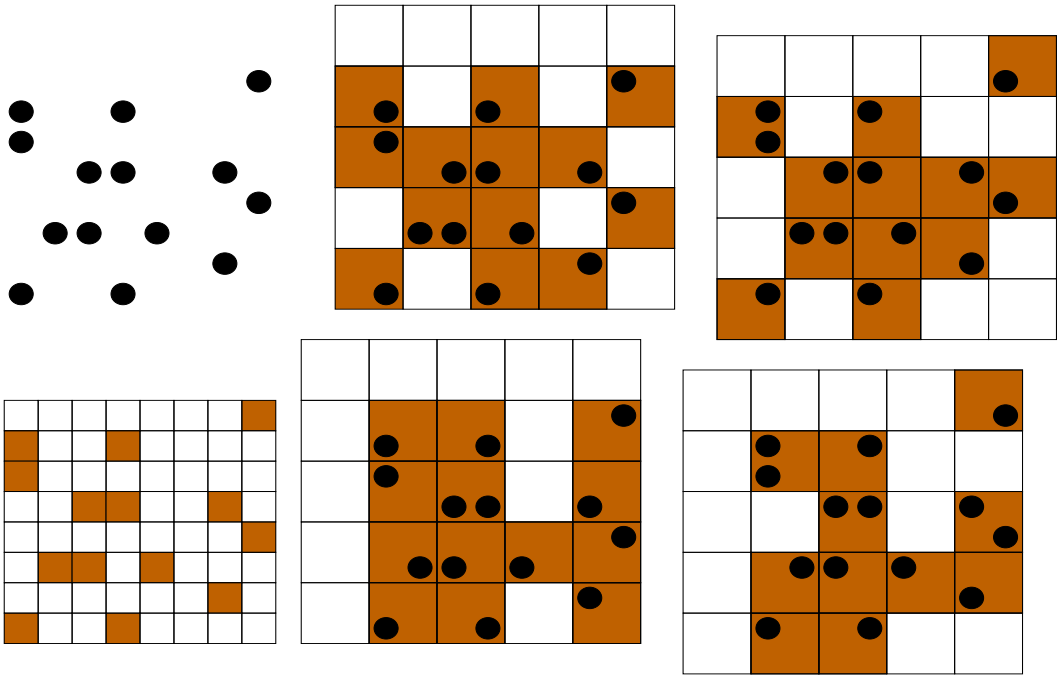

Figure 1: Shifting grids, clockwise from lower left: original fine-scale data; data points extracted from that fine grid; and four different grid orientations for a single coarser scale. Dots signify data, colored bins contain data, and white bins are empty. bill1-shift [NR]

EXAMPLES

The CASC93 dataset (Nabelek and Zandt, 1993) is a passive seismic experiment where arrivals from distant earthquakes were recorded by an irregular 2D array of seismometers in central Oregon. As we can see in Figure 2, the interpolated result with the PEF estimated from the shifted data looks better than the original multiscale result. In this case 7 different scales were used for (b), while for (c) 10 additional versions of the data were used. The difference is most pronounced in the area with coarser sampling, since the different grid positions would return more unique information than in the densely-sampled area.

Next, a 3D synthetic case is examined. The "quarter-dome" model (Claerbout, 1993) has been sub-sampled randomly so that only 10 percent of traces remain. It was then interpolated with both a standard multiscale PEF as well as a multiscale PEF estimated with different grid positions. The results are shown in Figure 3. The number of versions of the data varies even more in this case, where 5 different scales were used in the multiscale estimation, while 39 additional versions of the data were used when grid position was varied. The results are improved, although nearly 8 times as many copies of data were used. However most of those versions of the data were at the coarser scales.

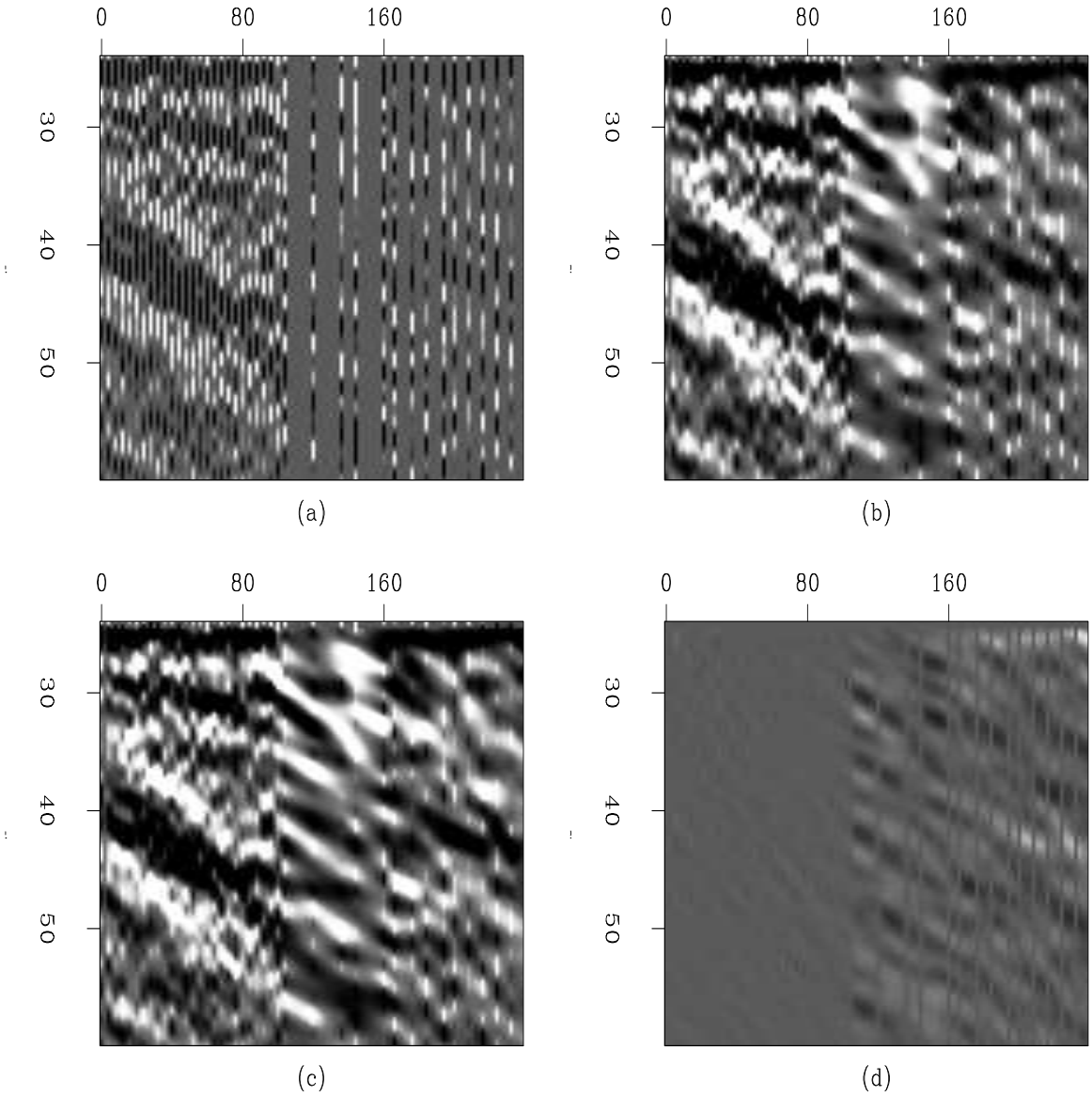

Figure 2: CASC93 data set. (a) original data; Interpolated with multiscale PEF with (b) no shifting and (c) with shifting; (d) difference between (b) and (c). bill1-casc [ER]

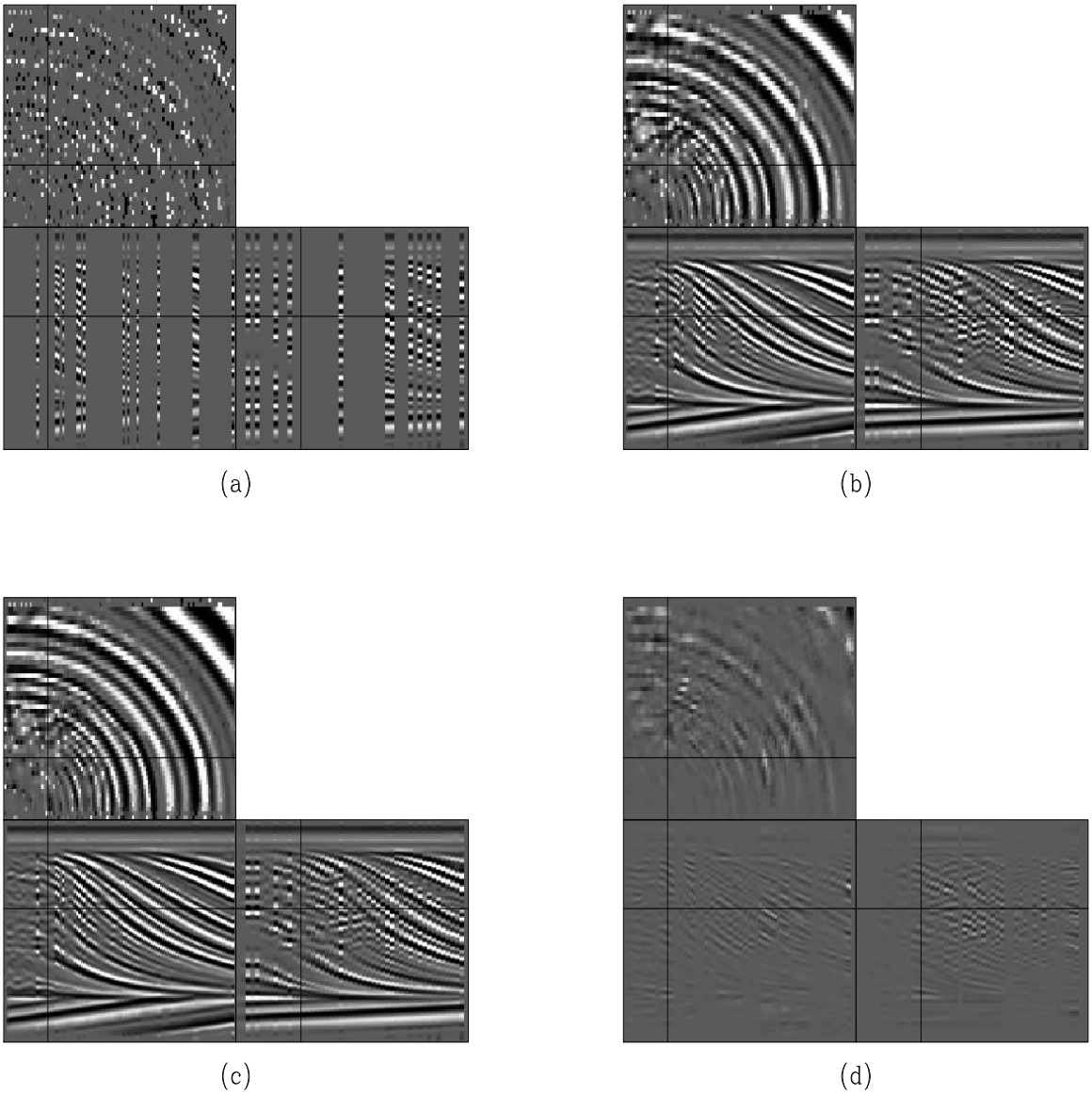

Figure 3: The quarter-dome model: (a) original data; (b) interpolated with multiscale PEF; (c) interpolated with multiscale PEF with grid position; (d) difference between (b) and (c). bill1-qdome [CR]

CONCLUSIONS

By altering the position of a grid as well as the coarseness of a grid, additional information can be introduced to the problem of PEF estimation. This information can be used for both stationary and non-stationary PEFs in any number of dimensions. As the dimensionality of a problem increases, the number of possible grid positions increases exponentially, whereas the number of scales of data is unaffected by the dimensionality of the problem. The benefits of this method increase with dimension, as shown in the differences between the 2D and 3D examples. The method also shows greater effectiveness around larger gaps in the data. However, the cost of regridding the data many times as well as the associated bookkeeping is problematic.

This method can be used to glean more information from the data using existing techniques. The question of how to best regrid the data remains, as does the question of whether it is best to manipulate the filter or the data when attempting to estimate a PEF on sparse data. Overall, this method provides one more means of manipulating the data to better constrain the PEF, and shows improvement over existing methods.

REFERENCES

Claerbout, J. F., 1993, 3-D local monoplane annihilator: SEP–**77**, 19–25.

Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

Crawley, S., 2000, Seismic trace interpolation with nonstationary prediction-error filters: Ph.D. thesis, Stanford University.

Curry, W., and Brown, M., 2001, A new multiscale prediction-error filter for sparse data interpolation: SEP–**110**, 113–122.

Curry, W., 2002, Non-stationary, multi-scale prediction-error filters and irregularly sampled data: SEP–**111**, 327–337.

NabelekJ., X.-Q. L. S. A. J. B. A. F. B. L. A. M. T., and Zandt, G., 1993, A high-resolution image of the cascadia subduction zone from teleseismic converted phases recorded by a broadband seismic array: Eos Trans. AGU, **74(43)**, 431.

Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Seismic trace interpolation in the f-x domain:, Soc. of Expl. Geophys., Geophysics, 785–794.