# Test case for PEF estimation with sparse data II[a]

[a]Published in SEP report, 105, 117-122 (2000)

*Morgan Brown, Jon Claerbout, and Sergey Fomel*[1]

## INTRODUCTION

The two-stage missing data interpolation approach of Claerbout (1998) (henceforth, the *GEE approach*) has been applied with great success (Fomel et al., 1997; Clapp et al., 1998; Crawley, 2000) in the past. The main strength of the approach lies in the ability of the prediction error filter (PEF) to find multiple, hidden correlation in the known data, and then, via regularization, to impose the same correlation (covariance) onto the unknown model. Unfortunately, the GEE approach may break down in the face of very sparsely-distributed data, as the number of valid regression equations in the PEF estimation step may drop to zero. In this case, the most common approach is to simply retreat to regularizing with an isotropic differential filter (e.g., Laplacian), which leads to a minimum-energy solution and implicitly assumes an isotropic model covariance.

A pressing goal of many SEP researchers is to find a way of estimating a PEF from sparse data. Although new ideas are certainly required to solve this interesting problem, Claerbout (2000) proposes that a standard, simple test case first be constructed, and suggests using a known model with vanishing Gaussian curvature. In this paper, we present the following, simpler test case, which we feel makes for a better first step.

- **Model:** Deconvolve a 2-D field of random numbers with a simple dip filter, leading to a “plane-wave” model.
- **Filter:** The ideal interpolation filter is simply the dip filter used to create the model.
- **Data:** Subsample the known model randomly and so sparsely as to make conventional PEF estimation impossible.

We use the aforementioned dip filter to regularize a least squares estimation of the missing model points and show that this filter is ideal, in the sense that the model residual is relatively small. Interestingly, we found that the characteristics of the true model and interpolation result depended strongly on the accuracy (dip spectrum localization) of the dip filter. We chose the 8-point truncated sinc filter presented by Fomel (2000). We discuss briefly the motivation for this choice.

## METHODOLOGY

Claerbout (1998) presents a two-stage methodology for missing data interpolation. In the first stage of the so-called *GEE approach*, a prediction error filter (PEF) is estimated from

[1]**e-mail:** morgan@sep.stanford.edu, claerbout@stanford.edu, sergey@sep.stanford.edu

known data. In the second stage, the PEF is used in a least squares interpolation scheme to regularize the undetermined (missing) model points. Crawley (2000) extends the two-stage procedure to use nonstationary PEF's.

The first stage (PEF estimation) of the two-stage procedure consists of convolving the unknown filter coefficients with the known data, and adjusting the coefficients such that the residual is minimized. Conceptually, in the process of convolution, a filter template is slid past every point in the data domain. The GEE approach adheres to the following convention: unless every point in the filter template overlies known data, the regression equation for that output point is ignored, and will not contribute to the PEF estimation.

Unfortunately, when the known data is very sparsely distributed, all the regression equations may depend on missing data, making PEF estimation impossible. The motivation of this paper is not to present a new methodology for estimating a PEF from sparse data, but instead to create a very simple test case which fulfills the following criteria:

1. The known data is distributed so sparsely as to render the traditional GEE two-stage approach ineffective.
2. The underlying model is conceptually simple and stationary.
3. The ideal PEF for the underlying model is obtainable by common sense.

## The Test Case

Claerbout (2000) proposes a test case for which the Gaussian curvature of the model vanishes. In this paper, we present an even simpler test case. Given a 2-D random field, we deconvolve with a known dip (or steering) (Clapp et al., 1997) filter to obtain a "plane wave" model, as shown in Figure 1. To simulate collected "data", we sampled the model of Figure 1 at about 60 points randomly, and about two-thirds of the way down one trace in the center. The results are shown in Figure 2.

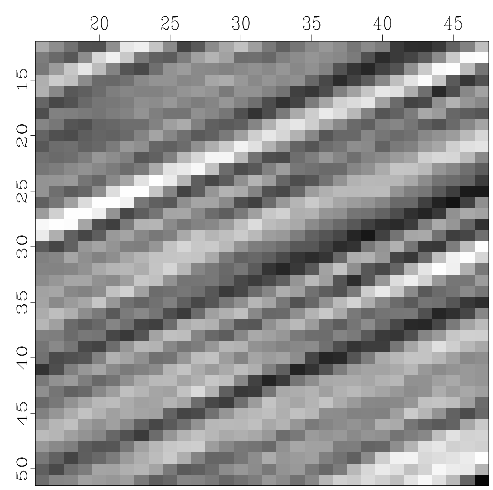


Figure 1: True model - plane waves dipping at 22.5°. test/ model

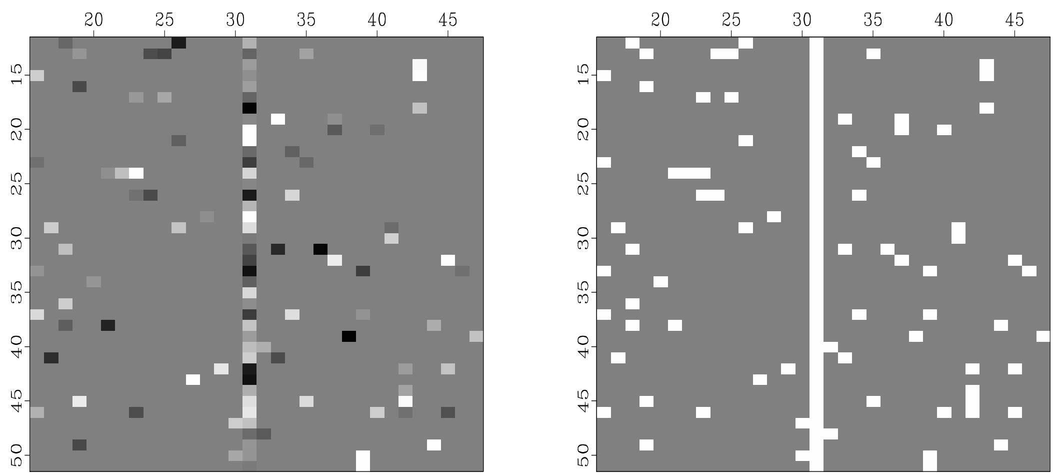

Figure 2: Left: Collected data - one known trace, about 60 randomly-selected known data points. Right: Known data selector. test/ data

## Digression: Accuracy of Dip Filters

Given a pure plane wave section, i.e., a wavefield where all events have linear moveout, designing a discrete multichannel filter to annihilate events with a given dip seems a trivial task. In fact, it is quite a tricky task; an exercise in interpolation. For many applications, accuracy considerations make the choice of interpolation algorithm critical. *Accuracy* here means localization of the filter's dip spectrum — ideally the filter should annihilate only the desired dip, or a narrow range of dips.

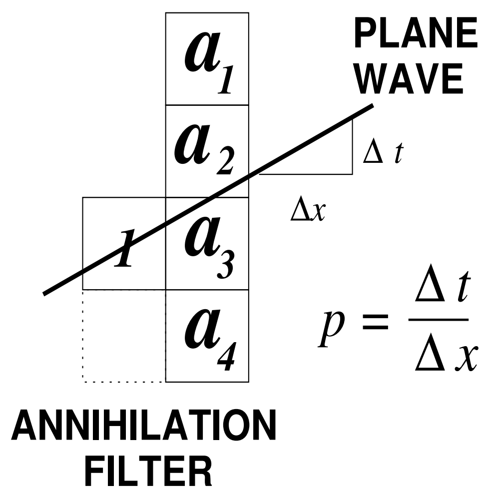


Figure 3: Steering filter schematic. Given a plane wave with dip $p$, choose the $a_i$ to best annihilate the plane wave. XFig/ steering

The problem is illustrated in Figure 3. Given a plane wave with dip $p$, we must set the filter coefficients $a_i$ to best annihilate the plane wave. Achieving good dip spectrum localization implies a filter with many coefficients, by the uncertainly principle (Bracewell, 1986). If computational cost was not an issue, the best choice would be a sinc function with as many coefficients as time samples. Realistically, however, a compromise must be found between pure sinc and simple linear interpolation. The reader is referred to (Fomel,

2000) paper, which discusses these issues much more thoroughly. The model of Figure 1 was computed using an 8-point tapered sinc function. Figure 4 compares the result of using, for the same task, dip filters computed via four different interpolation schemes: 8-point tapered sinc, 6-point local Lagrange, 4-point cubic convolution, and simple 2-point linear interpolation. As expected, we see that the more accurate interpolation schemes lead to increased spatial coherency in the model panel. Clapp (2000) has been successful in using as few as 3 coefficients in steering filters for regularizating tomography problems, so we see that the needed amount of steering filter accuracy is a problem-dependent parameter.

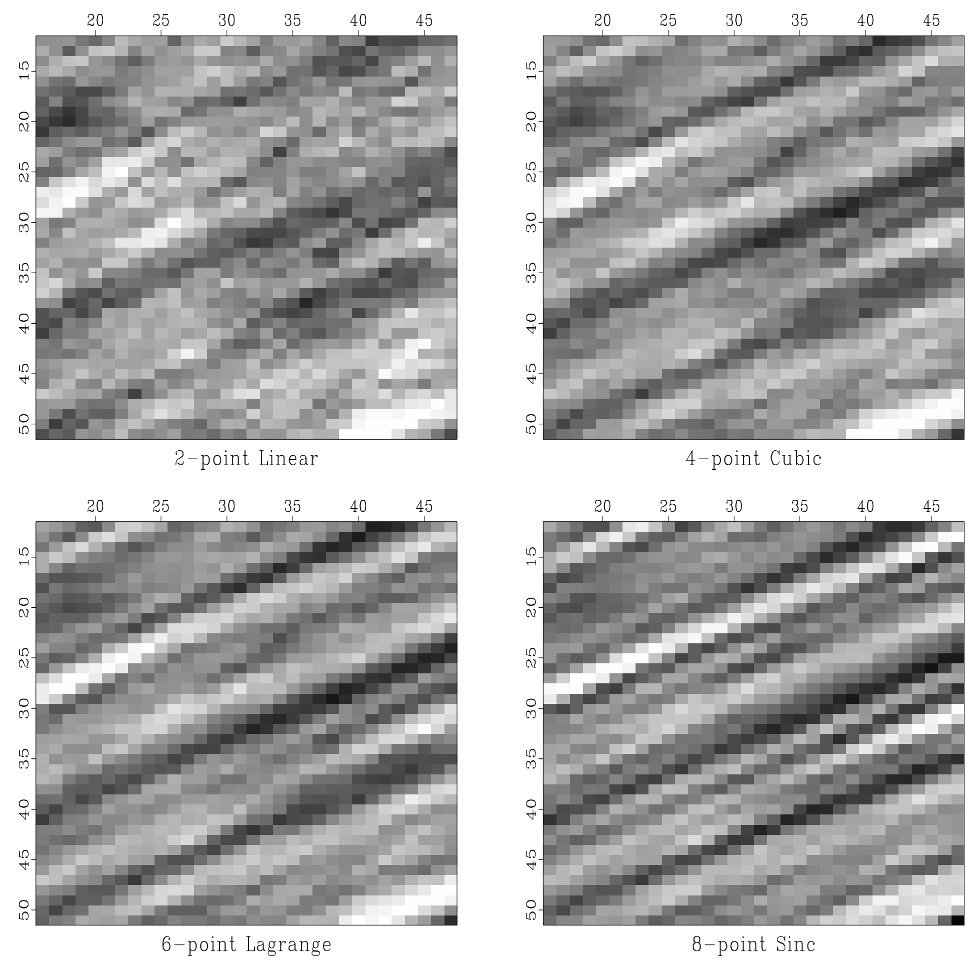


Figure 4: Interpolation schemes compared. Deconvolution of random image with labeled steering filters. test/ interp-comp

## INTERPOLATION RESULTS

The plane wave model of Figure 1 dips at $22.5°$, so we can easily design a filter to annihilate it. Using the GEE approach for interpolating missing data (Claerbout, 1998), we interpolate the data of Figure 2, using the 8-point tapered sinc steering filter discussed above. The results are shown in Figure 5. We see that the interpolation is quite good in the center region, where the filter can "see" one or more known data points, as evidenced by a nearly uncorrelated model residual. In the corners, the result is imperfect in regions in which no known data points exist along diagonal tracks. In order to suppress helix wraparound and other edge effects, we apply zero-padding around the edges of the study region.

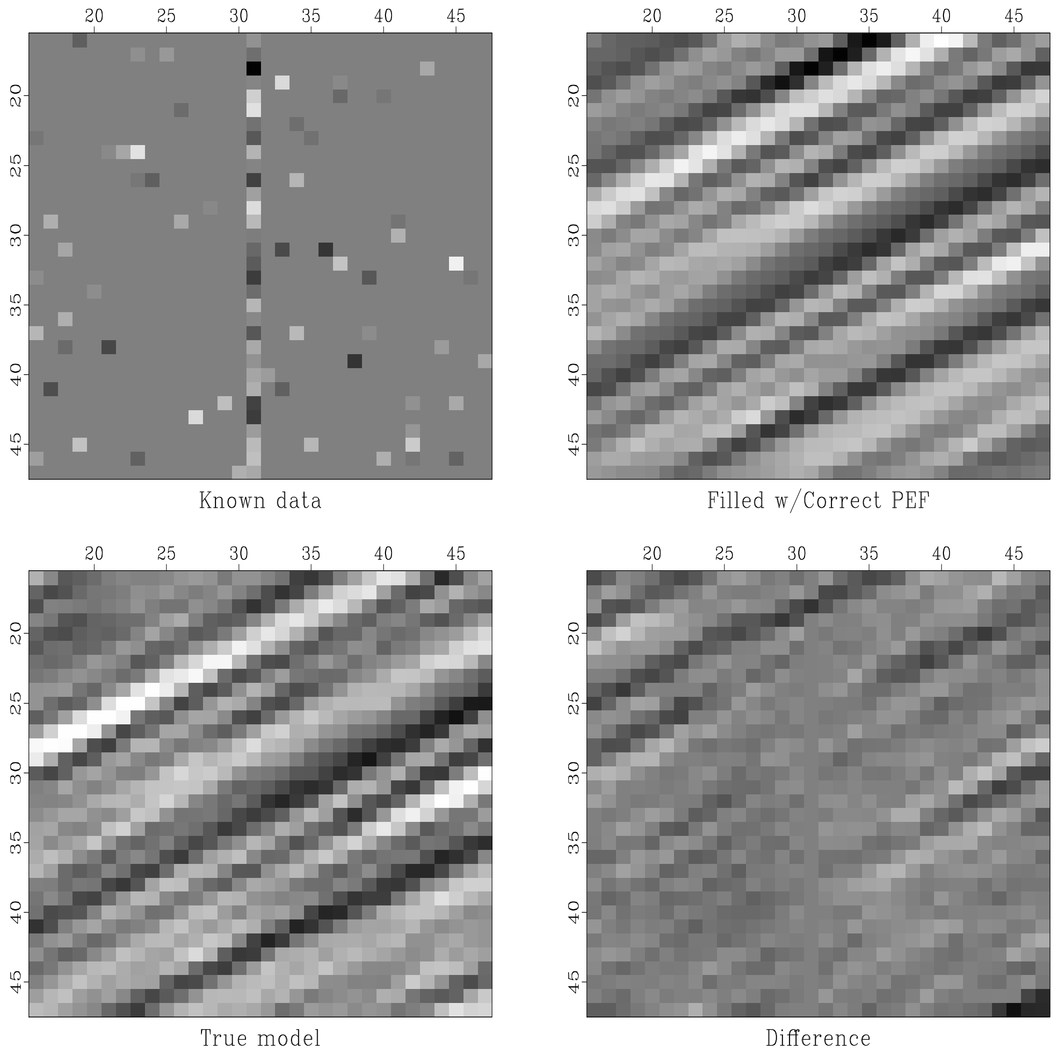


Figure 5: Clockwise from upper left: Known data; Interpolation regularized with 8-point tapered sinc steering filter; Difference between known model and interpolated result; known model. test/ correct

## CONCLUSIONS

We presented a 2-D test case for sparse data interpolation and give a good PEF with which to do it. The test case renders the traditional GEE two-stage interpolation scheme inapplicable. Claerbout (2000) suggests a nonlinear iteration, where filter and model are taken as unknown, but the best solution is still a subject of discussion among many SEP researchers. Regardless of the chosen interpolation strategy, the "correct" PEF and model are both known in this test case, so it should prove a useful starting point.

## REFERENCES

Bracewell, R. N., 1986, The Fourier Transform and its applications: McGraw-Hill.

Claerbout, J., 1998, Geophysical Estimation by Example: Environmental soundings image enhancement: `http://sepwww.stanford.edu/sep/prof/`.

———, 2000, Test case for PEF estimation with sparse data, *in* SEP-103: Stanford Exploration Project, 329–331.

Clapp, R. G., 2000, Regularizing tomography with non-stationary filters, *in* SEP-103: Stanford Exploration Project, 47–60.

Clapp, R. G., S. Fomel, and J. Claerbout, 1997, Solution steering with space-variant filters, *in* SEP-95: Stanford Exploration Project, 27–42.

Clapp, R. G., P. Sava, and J. F. Claerbout, 1998, Interval velocity estimation with a null-space, *in* SEP-97: Stanford Exploration Project, 147–156.

Crawley, S., 2000, Seismic trace interpolation with nonstationary prediction-error filters, *in* SEP-104: Stanford Exploration Project.

Fomel, S., 2000, Inverse B-spline interpolation, *in* SEP-105: Stanford Exploration Project, 79–108.

Fomel, S., R. Clapp, and J. Claerbout, 1997, Missing data interpolation by recursive filter preconditioning, *in* SEP-95: Stanford Exploration Project, 15–25.