Stanford Exploration Project, Report 115, May 22, 2004, pages 347–357

Regularizing Madagascar: PEFs from the data space?

William Curry[1]

ABSTRACT

The Madagascar seasat dataset presents a problem where data are collected along crossing tracks. These tracks are not straight, and appear to be irregular in the model space. Previous methods assumed that the data were regularly sampled in the model space coordinate system. I warp the model space to look more like the data space, so that prediction-error filters can be estimated in the more regularly-sampled data space. I try two different approaches to this problem on the warped space, one where the data is single-valued and fixed, while the other is multivalued and allowed to vary. The former method works very well, while the second one works well.

INTRODUCTION

The Madagascar seasat sea level dataset is a collection of two passes (ascending and descending) of the GEOSAT satellite over a region of the Southwest Indian Ridge in the Indian Ocean. There is a densely-acquired region of the dataset in the south, which ranges from 40 to 70 degrees (E) longitude and 30 to 40 degrees (S) latitude, while the latitude of the sparsely-acquired data ranges from 20 to 40 degrees (S) latitude.

The satellite tracks are much like feathered marine geophone cables, sail lines, or shot lines in a 3D seismic survey. Any method that hopes to succeed on 3D seismic data should be able to deal with this toy problem.

There are several issues to account for when dealing with this data. The crossing tracks of the dataset result in single locations with multivalued data, which can be substantially different due to tidal variations, weather patterns, and currents. In addition, there is spiky noise throughout the dataset, which is infamous for its effect on least-squares based methods (Claerbout and Muir, 1973; Guitton and Symes, 2003). Finally, the only tracks in the northern half of the survey area have a very wide spacing, so an interpolation problem also arises.

Previous work on this dataset at SEP (Ecker and Berlioux, 1995; Lomask, 1998, 2002) has mainly dealt with the systematic errors present in the dense dataset (Ecker and Berlioux, 1995), or with ways in which to use information in the dense portion of the data to regularize the missing bins in the northern, sparse portion of the data (Lomask, 1998, 2002).

Some previous methods of dealing with sparse data rely on creating proxy data, either by

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

an initial guess of a prediction-error filter (PEF) (Claerbout, 1999) or by using coarser scales of the data (Curry, 2002). I propose a method where we solve the interpolation problem in the data space coordinate system. The model space is warped to fit the data location and two different approaches are applied. In the first approach te data are preprocessed so that they are single valued. In the second, two sets of co-located tracks are used to create a map.

BACKGROUND

The Madagascar regularization problem has been approached using the following fitting goals (Lomask, 2002):

$$\begin{aligned} \mathbf{W}\frac{d}{dt}[\mathbf{Lm}-\mathbf{d}] &\approx \mathbf{0} \\ \epsilon\mathbf{Am} &\approx \mathbf{0}. \end{aligned} \tag{1}$$

In these fitting goals: $\mathbf{W}$ corresponds to a weight for ends of tracks and spikes in the data, $\frac{d}{dt}$ is a derivative along each track used to eliminate low frequency variations along each track, $\mathbf{L}$ is a linear interpolation operator that moves from values on a regular grid to the data points, $\mathbf{m}$ is the desired gridded model, $\mathbf{d}$ are the data points along the tracks, $\mathbf{A}$ is a regularization operator, and ϵ is a trade-off parameter between the two fitting goals.

The regularization operator ($\mathbf{A}$) typically is a Laplacian, a prediction-error filter (PEF), or a non-stationary PEF (Crawley, 2000). When using a PEF, it first must be estimated on some training data, using a least-squares fitting goal,

$$\mathbf{W}(\mathbf{DKf}+\mathbf{d}) \approx \mathbf{0}, \tag{2}$$

in which $\mathbf{W}$ is a weight to exclude equations with missing data, $\mathbf{D}$ is convolution with the data, $\mathbf{K}$ constrains the first filter coefficient to 1, $\mathbf{f}$ is the unknown filter, and $\mathbf{d}$ is a copy of the data.

In order to set a benchmark for how effective a prediction-error filter can be as a regularization operator, a PEF is estimated on the densely-sampled portion of the Madagascar dataset and is then used to interpolate the sparse tracks in the same area. The data are interpolated by using the following fitting goals:

$$\begin{aligned} \mathbf{K_{data}m} &\approx \mathbf{m_k} \\ \epsilon\mathbf{Am} &\approx \mathbf{0}. \end{aligned} \tag{3}$$

Here $\mathbf{K_{data}}$ is a mask for known data, $\mathbf{m_k}$ are the data, $\mathbf{m}$ is the model, and $\mathbf{A}$ is the regularization operator. In this case, the input data is the output of the fitting goals in equation (1), with ϵ set to zero. Results for using a Laplacian, a PEF estimated on well-sampled data, and a non-stationary PEF estimated on dense data as the regularization operator $\mathbf{A}$ are all shown in Figure 1. As we can see, as the complexity of the regularization operator increases, the interpolated result improves. This is because the Laplacian assumes isotropic behavior in the data while the PEF and non-stationary PEF are based on statistical information in the data. The PEF assumes statisical stationarity in the data while the non-stationary recognizies and accounts for the non-stationary nature of the data. These results set an upper benchmark for what the best possible interpolation could be using these methods for this type of dataset. In this example we are benefiting from the fully sampled nature of the input.

Now that the best-case scenarios are out of the way, we can see what we can accomplish without cheating and using the dense data. By using only the sparse tracks, we are not able to capture nearly as much information about the model as we have in the previous case. The results for interpolating with a Laplacian as well as stationary and non-stationary PEFs estimated solely on the sparse tracks with the multi-scale method (Curry and Brown, 2001; Curry, 2002, 2003) are shown in Figure 2.

In this case, there are regions of the data where the Laplacian gives the best result, and regions where the non-stationary PEF gives the best result. One example of this is the spreading ridge. The non-stationary PEF is able to interpolate some fine features of the ridge that Laplacian interpolation is incapable of. However, in other regions, the non-stationary PEF performs poorly, and the far simpler Laplacian interpolation gives a more reasonable result. This is because the PEFs estimated using the multi-scale approach do not always properly characterize the data, as local information is destroyed in the multi-scale averaging process.

Some of the problems encountered by the non-stationary PEF are due to the spatial distribution of the data. The multi-scale approach used in Figure 2 blindly rescales the data without regard to their location. By simply making the grid cells larger, very coarse bins must be used before the data become contiguous, and many data points along the tracks fall into these bins. We are clearly in need of a method that will take into account that these data are collected along curved tracks that are at an angle.

PEFS IN THE DATA SPACE

The known data points in the model space are distributed along curved crossing tracks, making it very difficult to estimate a PEF in this space. However, in the data space of the fitting goals in equation (1) of the previous section, the data are sampled in a regular space: a series of regularly-sampled tracks. By transforming the coordinates of the data and model, we can keep the data space regularly-sampled while warping the model space so that a PEF estimated in the data space can be used in the model space.

The existing data are collected in a series of ascending and descending crossing tracks. The goal is to transform the model space so that it more closely resembles the pattern in which the data was collected, which is in two sets of crossing tracks. This can be done by first stretching the longitudinal axis so that the tracks are orthogonal, subtracting the mean values of the coordinates of the data, rotating the data so that the tracks are aligned North-South and East-West, applying two parabolic shifts along a rotated axis, and then finally breaking the data up into ascending and descending tracks and griding the data. This process is shown in Figure 3.

The method used to warp the model space is not especially important. It would perhaps be better to use specific coordinates of the data space, but this warping accomplishes the same

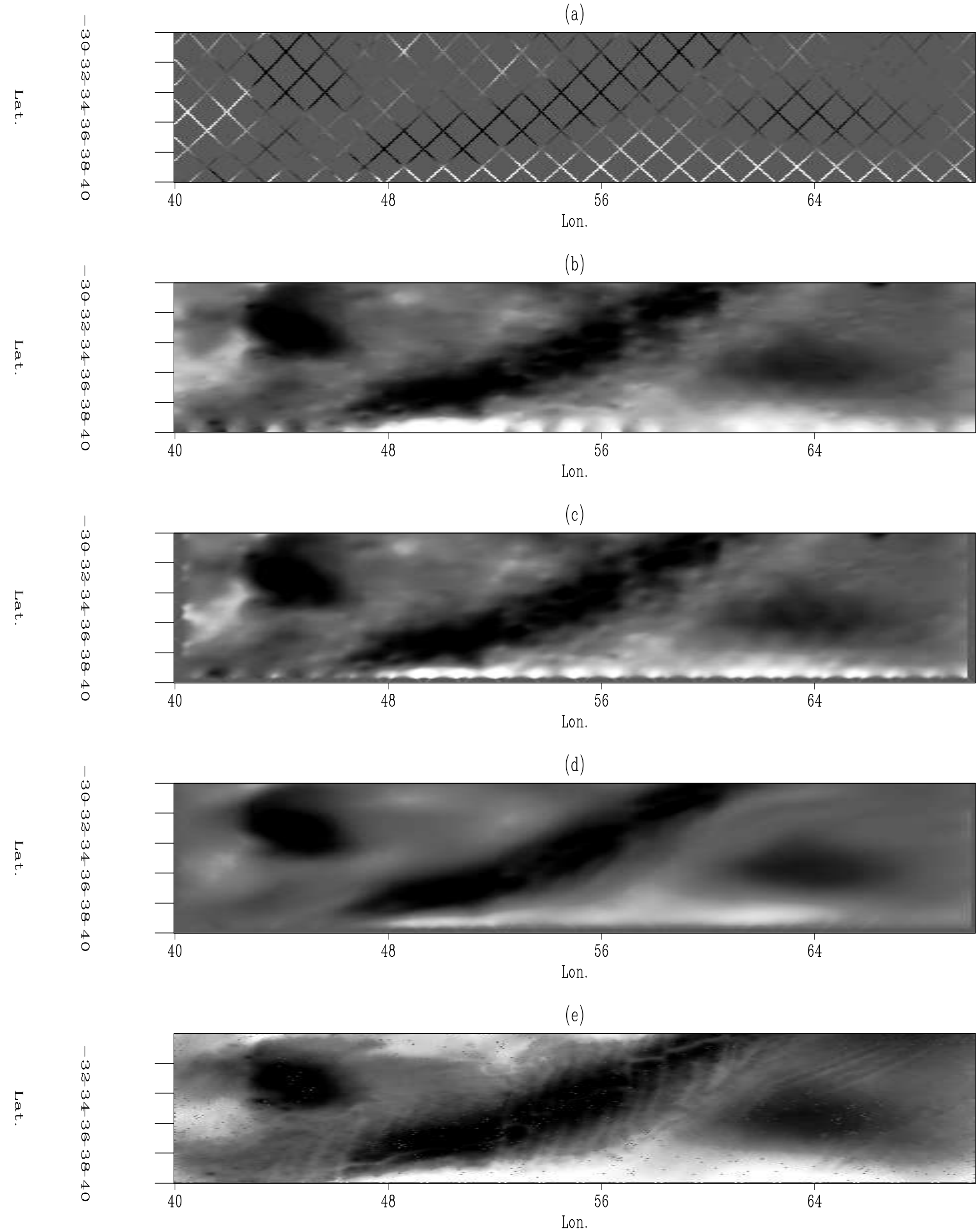

Figure 1: When we know the answer, from top to bottom: (a) The sparse tracks on the lower half of the data set; (b) Those tracks interpolated with Laplacian regularization; (c) The sparse tracks interpolated with a PEF estimated on the co-located dense tracks; (d) the same as (c), but using a non-stationary PEF; (e) the dense tracks from the same area. bill1-sparsedata [ER]

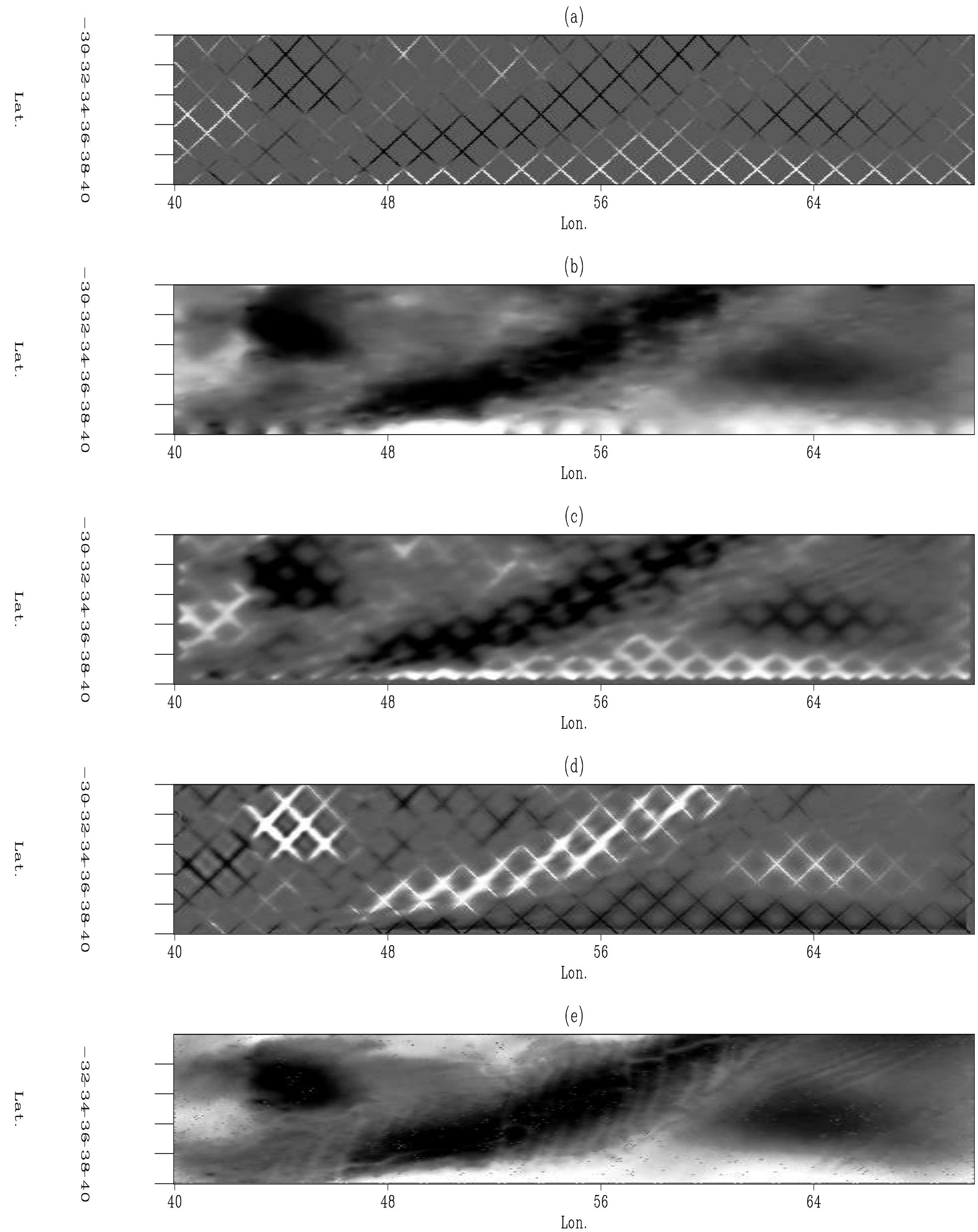

Figure 2: When we don't know the answer, from top to bottom:(a) The sparse tracks on the lower half of the data set; (b) Those tracks interpolated with Laplacian regularization; (c) The sparse tracks interpolated with a PEF estimated on the same sparse tracks; (d) the same as (c), but using a non-stationary PEF; (e) the dense tracks from the same area. bill1-unrotpef [ER]

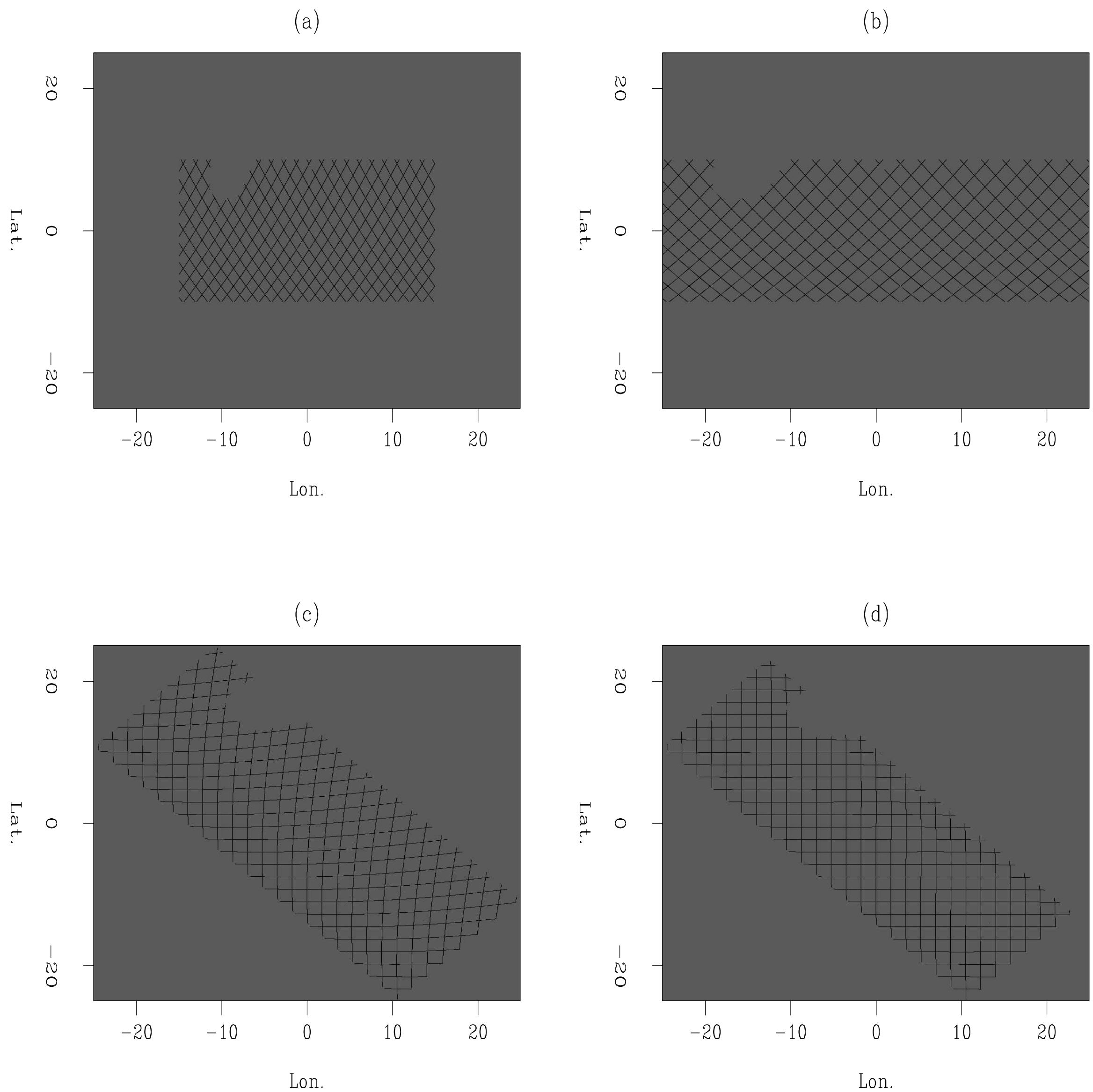

Figure 3: The steps used in warping the coordinate system.(a): Original model; (b): Stretching and centering; (c): Rotation; (d): Parabolic shifts. bill1-warping [ER]

end goal of having straightened orthogonal tracks. We can use the same warped grid for our model space, allowing us to use a similar methodology that we used for the regularly-sampled data problem. The two sets of orthogonal tracks with missing data is very similar to a problem shown by Claerbout (1999), where a pair of 1D PEFs are used to fill in missing data. This can be expressed as

$$\begin{aligned} \mathbf{K_{data}m} &\approx \mathbf{m_k} \\ \epsilon\mathbf{A_a m} &\approx \mathbf{0} \\ \epsilon\mathbf{A_d m} &\approx \mathbf{0}, \end{aligned} \tag{4}$$

where $\mathbf{K}$, $\mathbf{m}$ and $\mathbf{m_k}$ are the same as above, and $\mathbf{A_a}$ and $\mathbf{A_d}$ are 1D PEFs that are estimated on the ascending and descending tracks in the data space, respectively. This approach assumes that the known data is single-valued, and fixes it so that it does not change. The two sets of tracks are reduced to a single map with the fitting goals in equation (1), with ϵ set to 0. This becomes the data, and the obtained model is shown in Figure 4. The results of this approach

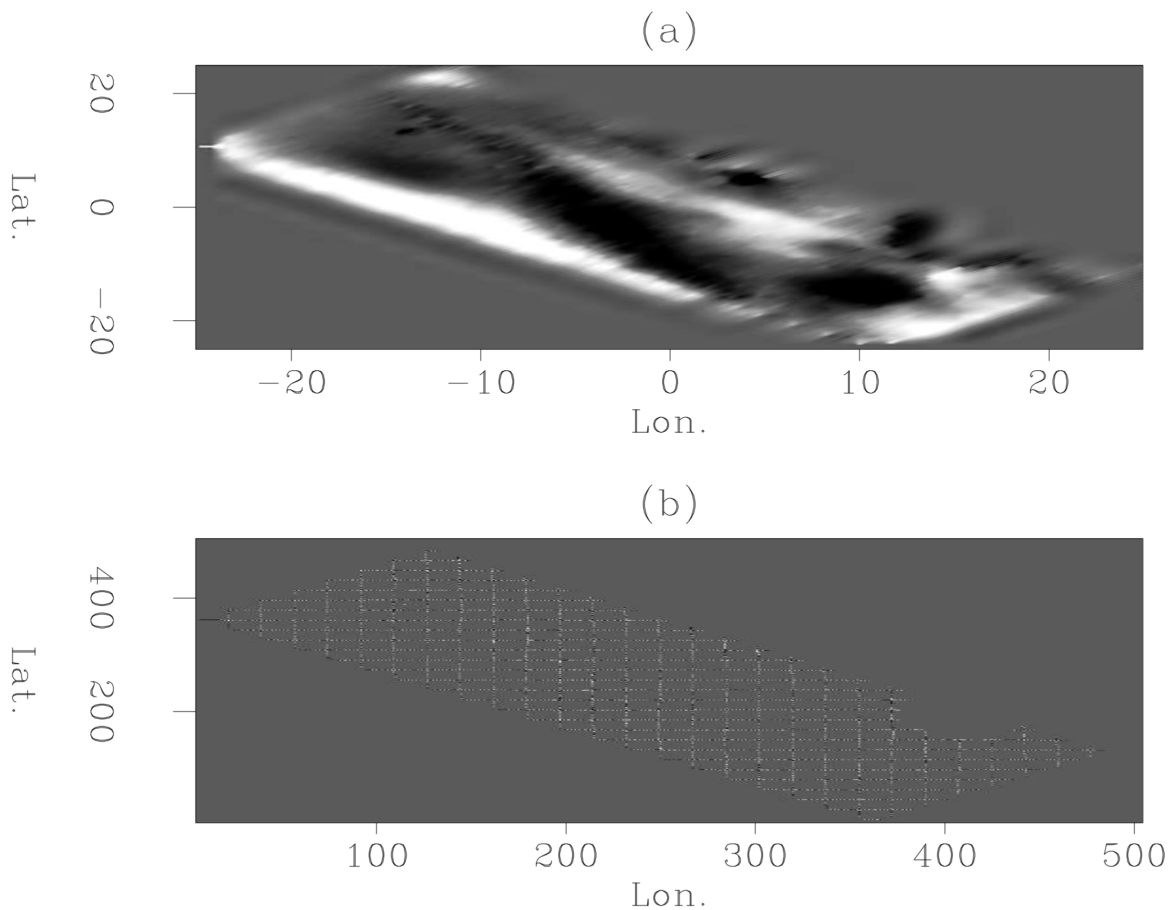

Figure 4: Results in a warped space using the fitting goals in equation (4). Above: model obtained after 500 iterations. Below: residual. bill1-warpfill1 [ER,M]

are good. The general trend of the data is interpolated, and the spreading ridge is clearly present in the model, which is not the case with Laplacian interpolation. Since this method only deals with the merged tracks, another method which uses the two track sets separately is tested next.

In this second method, the following fitting goals are solved:

$$\begin{aligned} \mathbf{A_a}(\mathbf{m} - \mathbf{d_a}) &\approx \mathbf{0} \\ \mathbf{A_d}(\mathbf{m} - \mathbf{d_d}) &\approx \mathbf{0}, \end{aligned} \tag{5}$$

where $\mathbf{A_a}$ and $\mathbf{A_d}$ are again 1D PEFs that are estimated in the data space, which is now two panels containing separately gridded ascending and descending tracks, represented by $\mathbf{d_a}$ and $\mathbf{d_d}$. The results for this method are shown in Figure 5.

The residual space is now twice the size as that for the fitting goals in equation (4), since the ascending and descending tracks are now dealt with separately. This approach appears to also be successful, although some effects of the PEF are present in the data, making the result look like a roughened result. The spreading ridge can be easily delineated. The residual space is interesting to look at, as the influence of the two sets of tracks can be seen in the different portions of the residual. The imprint of the data has disappeared from one half of the residual and has appeared in the other.

CONCLUSIONS AND FUTURE WORK

By estimating a pair of 1D PEFs in a warped data space, the sparse, curved tracks of the Madagascar dataset were successfully interpolated. Other methods that require a starting guess or rescaled proxy data were much less successful, as they were estimated in a domain where the data are not optimally distributed.

The tracks in the Madagascar dataset have several similarities with the tracks seen in 3D seismic data, such as sail lines, receiver cables, and cut lines for sources. When estimating PEFs on this data, the choice of doing it in the model space and the data space needs to be investigated.

To obtain a better result for this data, the next obvious step is to use non-stationary PEFs, as the character of the Madagascar data changes greatly with position.

ACKNOWLEDGMENTS

I would like to thank Jon Claerbout, Jesse Lomask and Antoine Guitton for several fruitful discussions.

REFERENCES

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, no. 05, 826–844.

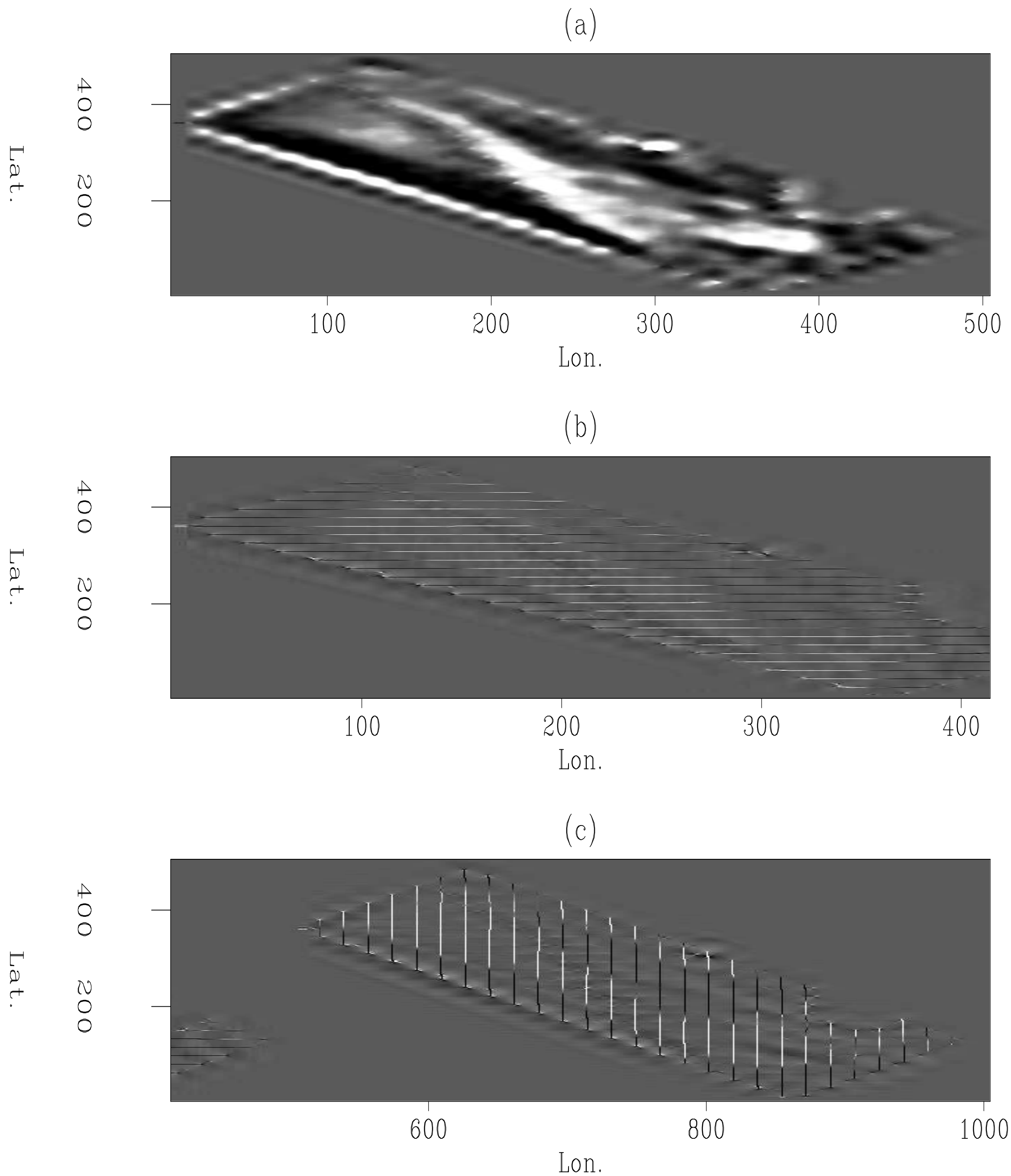

Figure 5: Results in a warped space using the fitting goals in equation (5). (a): model obtained after 150 iterations. (b),(c): two parts of the residual space. The data space contains the two sets of tracks, and is twice the size of the model space. bill1-warpfill2a [ER,M]

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.

Curry, W., 2003, More fitting equations for PEF estimation on sparse data: SEP–**114**, 171–176.

Ecker, C., and Berlioux, A., 1995, Flying over the ocean southeast of Madagascar: SEP–**84**, 295–306.

Guitton, A., and Symes, W., 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, no. 4, 1310–1319.

Lomask, J., 1998, Madagascar revisited: A missing data problem: SEP–**97**, 207–216.

Lomask, J., 2002, Madagascar satellite data: An inversion test case: SEP–**111**, 337–349.