Double decimation: an effective approach for surface-consistent computations

Charles Ursenbach, TGS

Summary

A land dataset can be decimated in such a way that the result naturally divides into even smaller subsets, suitable for surface-consistent computations. The underconstrained nature of these computations causes incongruence among solutions, but a simple constraint solves this issue. Combining solutions then yields an excellent approximation to the surface-consistent solution for the full dataset.

Introduction

Large land datasets present a robust challenge to seismic processing, particularly for surface-consistent computations. For these one must solve a problem in which all traces in the survey are present at once. One direction of research is to determine a method of separating the survey into portions that are tractable yet still produce accurate solutions.

This study describes a novel approach to decimation—referred to here as 'double decimation'—that is very simple to implement and produces small subsets that each cover the entire survey. These produce surface-consistent solutions with similar trends, and they can easily be combined to produce a solution very similar to that of the full dataset.

I describe the method below, along with comments on the underconstrained nature of surface-consistent problems. This is followed by application to a dense land data set.

Method

In a seismic context the term 'decimation' refers to the thinning of a dataset so that its size is considerably reduced, but in such a way that it is still representative of the full dataset in some key aspect. The meaning of the term 'double decimation' used here refers to decimation performed in such a way that the decimated result can be naturally divided into a collection of smaller independent subsets.

The essential approach is very simple. Assuming that sources and receivers have each been assigned a sequential index, proceed as follows:

1) Choose some integer N to determine the extent of decimation. The decimated result will contain $1/N$ as many traces as the original data, so there will be some practical limit to how large N can be.
2) For each trace I, compute the N-moduli of its source index S_I and receiver index R_I. If mod(S_I, N) = mod(R_I, N), the trace is retained in the decimated dataset; otherwise it is discarded. For some purposes, this can be the end of the decimation procedure.
3) One potential shortcoming of this method is that it is possible for some sources and/or receivers to not be represented in the decimated dataset if the modulus requirement is not satisfied by their indices. As shown in the example below, this is not expected to be a significant problem. However, if it is absolutely required in some instance to have every source and receiver represented, that can be dealt with at this point, as described below in the example.
4) Finally, each trace in the decimated dataset can be assigned a value $D_I \equiv$ mod(S_I or R_I, N). Based on this value the decimated dataset can be further separated into N independent datasets, each containing $1/N^2$ as many traces as the original dataset. They are independent in the sense required for surface-consistency, namely, that no source or receiver is referenced in more than one subset. Furthermore, if a typical scheme of indexing is used, then each subset will extend over the entire survey. Here each of these subsets will be referred to as 'doubly decimated'.

Note on underconstrained solutions

While the surface-consistent equations are overdetermined, they are also underconstrained (see, e.g., Taner et al., 1974; Wiggens et al., 1976; Morley & Claerbout, 1983; Cambois & Stoffa, 1992; Millar & Bancroft, 2006; van Vossen et al., 2006; Baek et al., 2019; Zhang et al., 2024). Thus the equations do not possess a unique solution. In the present context, this means that solutions obtained from various decimation schemes can give spuriously different results.

The underconstrained problem can be approached in a simple way. The equations we are solving are of the form

$$s_i + r_j + c_{k(i,j)} + h_{l(i,j)} = a_{ij},$$

where s_i is the contribution from the i^{th} source, r_j is the contribution from the j^{th} receiver, and c_k and h_l are CMP and offset bin contributions. These are parameters to be estimated, while a_{ij} is an observed property of the trace, such as log amplitude or static shift. The underconstrained nature of these equations can be recognized by noting that if a solution $\{s_i, r_j, c_k, h_l\}$ exists for all i, j, k, l, then we can add any four constants σ, ρ, χ, η to create $\{s_i + \sigma, r_j + \rho, c_k + \chi, h_l + \eta\}$, which will also be a solution as long as $\sigma + \rho + \chi + \eta = 0$. This last expression will determine the value of η if we choose σ, ρ, χ so as to constrain the source, receiver and CMP components to each be zero when summed over all traces. This is the course followed in this study to prevent spurious shifts between solution sets for various choices of decimation.

Example

The method has been tested on a land 3D dataset from West Texas, covering 85 mi^2 (220 km^2). This densely-acquired dataset contains 720 million traces, 40,000+ sources, and 40,000+ receivers, with offsets up to 10.5 mi (17 km). Its bin size (37.5 ft x 37.5 ft (11.4 m x 11.4 m)) is designed for a shallow target at ~2000-2500 ft (~600-750 m).

Surface-consistent amplitude scaling is performed to demonstrate the efficacy of the decimation method. An amplitude measure for each trace was determined in a typical way by computing the root-mean-square amplitude within a window of interest. The logarithm of the amplitude was computed, and the average log amplitude of the input was subtracted from each trace. Figure 1 shows a snapshot of a few of the source scalers that result when surface-consistent decomposition is applied to datasets of various sizes:

1) The full dataset (720 million traces (m.t.))
2) A dataset decimated using $N = 10$ (72 m.t.)
3) Subset A of 2), containing traces with $D_I = 1$ (7.2 m.t.)
4) Subset B of 2) with $D_I = 2$ and $D_I = 3$ (14.4 m.t.)
5) Subset C of 2) with $D_I = 4$-6 (21.6 m.t.)
6) Subset D of 2) with $D_I = 0$ and 7-9 (28.8 m.t.)

The agreement between the full and the decimated datasets, 1) and 2), is excellent. The doubly decimated subset, 3), and combinations of doubly decimated subsets, 4)-6), show more deviation, but the agreement is still very good. In Figure 2 we see a similar trend for receiver scalers.

In Figure 3 we see much less agreement among estimates of CMP scalers. This is not surprising, given that there are $O(10^2)$ as many CMP parameters as sources or receivers. Therefore, the fold determining each of them is much lower. Most processors rarely look at the CMP results, much less use them, and the disparity for different inputs seems not to have affected the source and receiver results significantly, which is what matters. We can decrease the observed CMP disparity by superbinning the CMPs to a sparser set, but it is not clear if this would provide a useful benefit.

In Figure 4 we have shown offset scalers for the full range of offset bins. This component follows an overall trend and is similar for all inputs. Greater deviation is seen at long offsets where fold is notably lower, as shown in the inset.

We turn next to map displays in Figures 5-10. In Figure 5 we show the logarithm of source scalers estimated from the full dataset. We do not show the result estimated from the decimated dataset, as it is visually indistinguishable from Figure 5. Instead, we show the difference between the two in Figure 6, multiplying by a factor of 20 to discern any behavior. The differences seem largely random.

We can also create a third set of scalers to map by combining the estimation results of inputs A, B, C, and D. Recall that these collectively used the same input traces as the decimated dataset but solved them in four separate smaller computations. We emphasize here how easy it is to recombine the results after the four surface-consistent decompositions, because no source or receiver is shared by any of them. Thus there is no weighting required to combine different results for the same source or receiver. We simply merge and sort the four sets of scalers into one. The result again is visually indistinguishable from Figure 5, and the difference x 20 is displayed in Figure 7. We see that the error, though small, is greater than in Figure 6, and in this case possesses some spatial trends. It is non-geological and presumably results from solution constraints which are slightly different than for the full and decimated datasets.

Figures 8-10 show analogous results for receiver scalers, with one instructive, behind-the-scenes difference: in the decimation process, one receiver was not included by the modulus criterion. This receiver had an index of 35917 and its receiver gather had only eleven traces, none of which had a source index ending in “7”. The dataset has a nominal receiver fold of nearly 17,000, so this is a rare outcome. If this receiver is required after decimation, there are five of the eleven traces with source index values ending in “9”, so we could set $D_I = 9$ for those five traces and discard the other six. Even though the receiver index has the “wrong” modulus for that group, it is not represented in any other group, so the independence of the groups would still be maintained. However, having said all that, in this case we chose to simply ignore the one missing receiver.

In Figure 8 we show the logarithm of receiver scalers estimated from the full dataset (minus index 35917, location indicated). As before, this is visually identical to the result for the decimated dataset and for the merge of results from subsets A-D. We display the differences x 20 in Figures 9 and 10 showing similar patterns as for the source scalers. Maps for individual subsets (not shown) are also similar to Figure 8 and to each other. Showing such consistency between subsets would be a valuable test for a larger dataset when the full result is not available for comparison.

Conclusions

I have described a ‘double decimation’ method for land datasets, suitable for surface-consistent computations. A simple approach to constraints ensures that solutions of partial and full datasets are congruent. Using these methods, subsets containing as little as 1% of the data can be solved independently, and their respective solutions combined to yield a very good approximation of the full solution. This can facilitate processing of increasingly large and dense land datasets.

Acknowledgements

The author expresses appreciation to TGS for permission to use this data, and to publish these results.

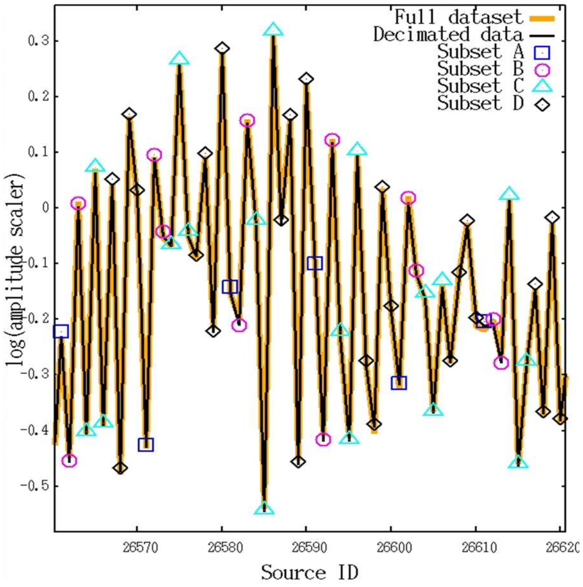

Figure 1: A few of the source scalers estimated by surface-consistent decomposition for several different input datasets. Agreement is generally excellent.

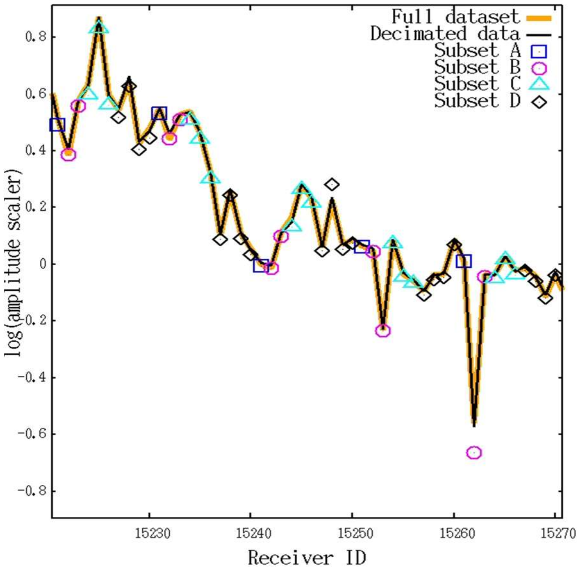

Figure 2: A few of the receiver scalers estimated by surface-consistent decomposition for several different input datasets. Agreement is generally excellent.

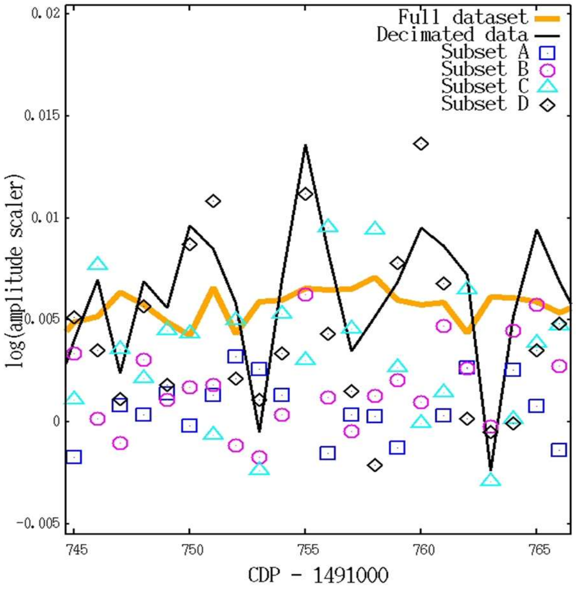

Figure 3: A few of the CMP scalers estimated by surface-consistent decomposition for several different input datasets. Agreement is worse than for Figures 1 and 2, as fold is much lower for CMPs.

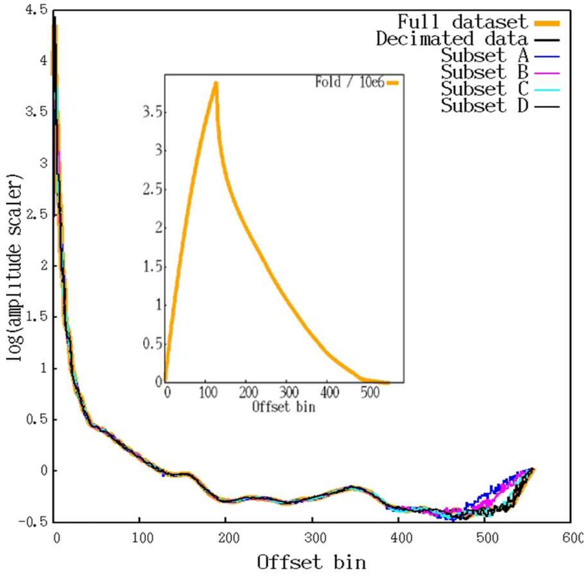

Figure 4: The full range of offset scalers estimated by surface-consistent decomposition for several different input datasets. Fold is shown in the insert. Agreement is worse for low-fold large offsets.

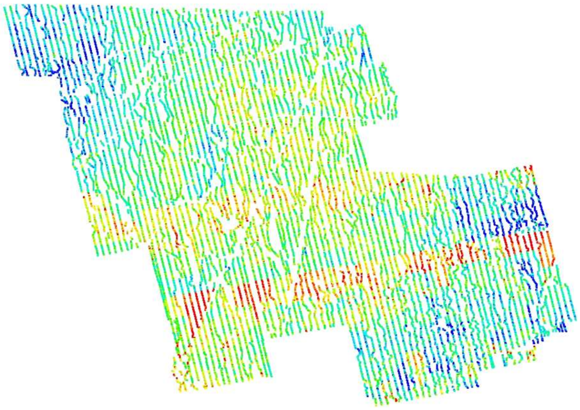

Figure 5: Map view of source scalers estimated by surface-consistent decomposition using the full dataset of this study. This is visually indistinguishable from similar maps for the decimated and the recombined doubly-decimated datasets. In Figures 5-10, green is zero, yellow and red are positive, and cyan and blue are negative.

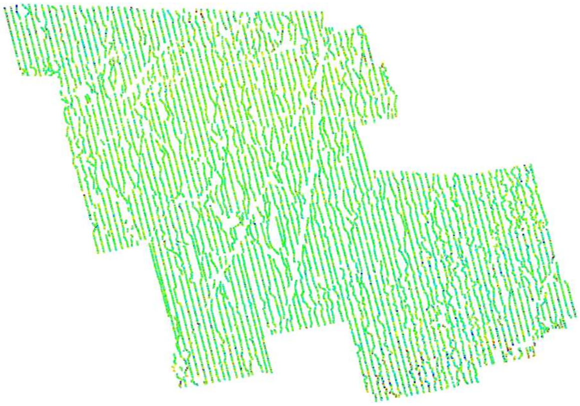

Figure 6: Twenty times the difference between source scalers estimated from the full and decimated datasets. The color limits are the same as for Figure 5. The differences appear small and random.

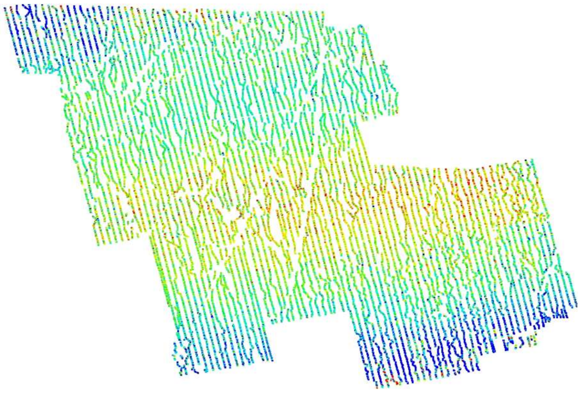

Figure 7: Twenty times the difference between source scalers estimated from the full and recombined datasets. The differences, greater than in Figure 6, are still relatively small but show some spatial trending due to a mathematical artefact.

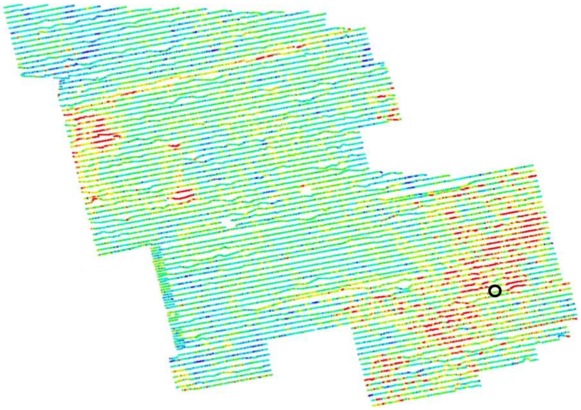

Figure 8: Map view of receiver scalers estimated by surface-consistent decomposition using the full dataset of this study. This is visually indistinguishable from similar maps for the decimated and the recombined datasets. The color bar limits are 50% larger than in Figure 5. The black circle shows the location of the missing receiver.

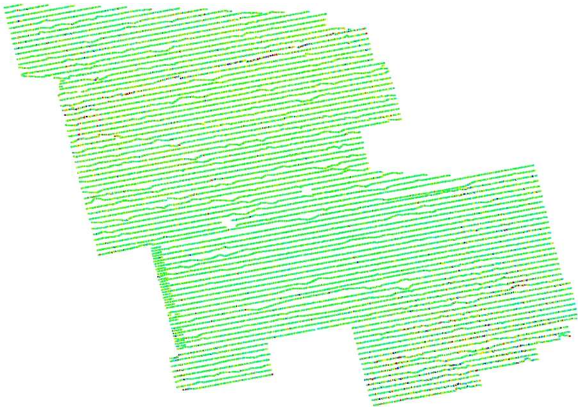

Figure 9: Twenty times the difference between receiver scalers estimated from the full and decimated datasets. The color limits are the same as for Figure 8. The differences appear small and random.

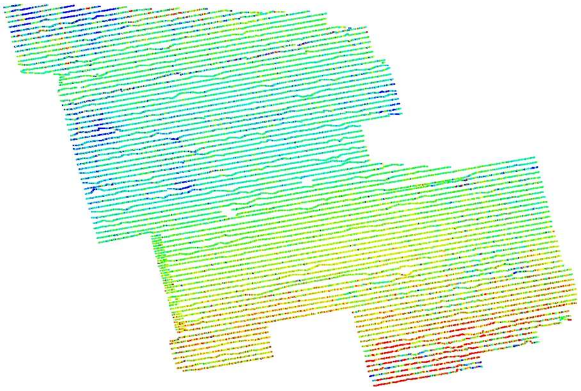

Figure 10: Twenty times the difference between receiver scalers estimated from the full and recombined datasets. The differences, greater than in Figure 9, are still relatively small but show some spatial trending due to a mathematical artefact.