Simultaneous Pre- and Post-NMO Deconvolution

Jon F. Claerbout

ABSTRACT

Wave theory justifies both pre- and post-NMO deconvolution, but the filters should be estimated *simultaneously*, not *sequentially*. A conjugate gradient procedure estimates the two filters simultaneously. The method also handles spherical divergence independently from statistical weighting. Test cases demonstrate the expected interaction between NMO and deconvolution. The tests were inadequate to establish that simultaneous estimation is superior to sequential estimation.

THE ART OF DIVIDING BY ZERO

That our data is limited in bandwidth is generally ascribed to a convolutional wavelet resulting from the source and its reverberation in the weathered zone. Deconvolution is the art of removing this "surface wavelet." The surface wavelet is unknown and is generally estimated from the data itself. The earth possesses reflectivity at all scales. The basic model for estimating the surface wavelet relies on the assumption that the spectrum of the earth's reflection coefficients is white. This assumption is imperfect, but experience generally shows it to be better than none. With this assumption, the spectrum of the surface wavelet is the spectrum of the seismogram itself. A zero-phase estimate of the surface wavelet is obtained from the spectrum of a seismogram (or average spectrum of a group of seismograms) by taking the square root, and inverse Fourier transforming. Deconvolution may be performed by dividing the Fourier transform of the data by the square root of the spectrum (or average spectrum) and inverse transforming. Since the seismic bandwidth drops off into noise at high frequencies, deconvolution may be called "the art of dividing by zero". To control the problem of noise amplification at high frequencies, the result is often bandpass filtered with a display filter.

Most deconvolution filters are estimated and applied in the time domain. The time domain is naturally suited to tailoring the deconvolution filter to be causal and of limited time duration. This in itself controls noise somewhat. A filter $f_t = (1, f_1, f_2, \cdots, f_n)$ is defined as a function of its coefficients f_t. By means of least squares, these coefficients are adjusted so as to minimize the power in the filter's output. It is shown in FGDP that this procedure whitens the output spectrum as does the Fourier method previously described. This filter turns out to be minimum phase and is the inverse of the surface wavelet if it happens to be minimum phase. This time-domain method does not escape the high-frequency difficulty of the Fourier method. But in the time domain, there are alternate means of attacking the problem. One means, to be exploited in this paper, is to constrain the early filter coefficients to be zero. The early coefficients $f_1, f_2, \cdots, f_{gap}$ determine the slow variations in the spectrum. (For example $f_1 \exp i\omega\Delta t$ is slowly variable in the Fourier domain). By constraining these coefficients to be zero, the deconvolution cannot amplify the high frequency band with respect to the central signal band. Further, the gap has a simple physical interpretation. Gapped time-domain deconvolution does not attempt to convert reflections to impulses, but instead to convert reflections to short wavelets the length of the gap. So it attempts the possible, not the impossible.

MECHANISMS AT THE SOURCE

In geophysics we see a historical struggle between those who interpret everything they see as manifestations of the *heterogeneity* of the earth, and those who interpret everything they see as manifestations of the *anisotropy* of the earth. *Deconvolution* is also a potential battleground for this struggle. *Surface-consistent* methods can be classified as being based on the *heterogeneity* of the earth. Slant stack and radial trace deconvolutions can be classified as being based on *anisotropy* of the radiated waves.

In a homogeneous medium, a point source of energy can radiate the same waveform in all directions. If the source lies near any strong vertical *inhomogeneity* the distant waveform becomes angle dependent (anisotropic). The simplest example is the free-surface ghost. This ghost generally follows the first arrival so soon, that it is difficult to distinguishable from it. The time delay becomes shorter with increasing angle, giving spectral changes with angle.

Bubble and multiple reflection as points on a continuum

Consider the two models: (1) bubble generated wavelets and (2) reverberations of multiple reflections. These seem to be very different models. However, the bubble could be modeled as a particular form of near-surface inhomogeneity. Imagine very shallow, very slow, very strong velocity stratification, atop a homogeneous medium. In the mathematical limit, any minimum phase reverberation might result from an impulse at the surface. In this model any wave in the subsurface must propagate vertically at the near surface. So a wave from the surface travels vertically from the surface, to the subsurface where it diverges into a spreading spherical wave. So, near-surface reverberation mimics the isotropy of a bubble. There is no divergence until the wave escapes the near surface, then the divergence is spherical. So, allowing for the time delay through the near-surface, spherical divergence correction of the pseudobubble also mimics the real bubble.

NMO to bring reverb to wider offsets

Suppose a dereverberation filter is known for the zero-offset trace. Consider a family of hyperbolic multiple reflections in a water layer. The arrivals are uniformly spaced at zero offset but compress together at wider offsets. If the zero-offset filter is applied at other offsets, the deconvolution will deteriorate with offset. This deterioration with angle can be diminished if the seismograms are corrected for normal moveout. Experimental confirmation is found later in the paper.

Normal moveout correction transforms the time axis to the axis of travel-time depth via $t^2 = \tau^2 + x^2/v^2$. So it is natural to refer to decon before NMO as t-decon, and decon after NMO as τ-decon. The method to be developed here simultaneously estimates the pre- and post-NMO deconvolution filters. It will be called *$t\,\tau$–decon*, *simultaneous* decon or *anisotropic* decon. It may be argued that any decon after NMO is anisotropic. This is true after inverse NMO. Simultaneous $t\,\tau$ decon is anisotropic before and after NMO.

Iterative cascades

Before tackling the simultaneous estimation problem, let us consider a repetition of familiar sequential estimates that converges to something roughly like the simultaneous estimate. A single deconvolution filter could be estimated for a field profile. The field profile could be deconvolved with this filter (i.e. t-decon). Then NMO could be done and another filter estimated and applied (i.e. τ-decon); Then *inverse* NMO could be done and the output could be deconvolved again (i.e. t-decon again), ad infinitum. In the log spectral domain, t-decon removes the ω-spectral average (over channels) and τ-decon removes the k_τ-spectral average. The averaging directions are neither perpendicular nor parallel, so when you remove one average, you affect the other. Except for noise problems, the result of the infinite cascade should be like the *simultaneous* filter estimation technique to be described (i.e. $t\ \tau$-decon).

ESTIMATION METHOD

No need for the Levinson recursion

The time has come to abandon the Levinson recursion. The computer economies do not warrant the loss of flexibility. The economies never really were that significant, because the effort to determine the autocorrelation needed by the Levinson recursion is much greater than the cost of the recursion itself. In fact, the cost of gathering the autocorrelation typically matches the cost of an ordinary, non-Levinson solution.

Simultaneous equation solving

There are some excellent exact methods of solving simultaneous equations, but I prefer iterative methods. The conjugate gradient method is an iterative method that obtains the exact solution in precisely n steps, where n is the number of unknowns. So it offers the same advantages of "exact methods". I generally prefer to do less than the full number of iterations. In this study I always did 10 iterations for a filter of 30 points. Wang and Treitel [1973], used conjugate gradients for classical deconvolution. They reported normalized mean square errors of 10^{-3} for this filter size and iteration count. Geophysical errors are generally much larger. Wang and Treitel solved the classical square system $(A^T A)^{-1}$, whereas here we solve the overdetermined system directly. I have not yet studied the effect of the number of iterations. Simple theoretical considerations suggest it may be better to quit early. In the early iterations, the prediction error power drops rapidly. In the late iterations, the filter develops complexity, while hardly dropping the error power. Then it is fitting the noise instead of the signal. It seems that limiting the number of iterations is a natural way of limiting the number of degrees of freedom in the filter. It is a natural way of having a long filter without having a lot of degrees of freedom.

I used a particular dialect of the conjugate gradient method called the Paige-Saunders [1982] method. The Paige-Saunders program requires that you provide subroutines to compute

$$y \leftarrow y + A\ x$$

and

$$x \leftarrow x + A^T\ y$$

The matrix A should not change from one iteration to the next, as would happen if a non-linear problem were being relinearized. In that case, it is appropriate to restart the conjugate gradient method whenever relinearization occurs, or else use the *partan* method (Luenberger [1973]).

Simultaneous estimate of bubble and reverb filters

Simultaneous estimation of pre- and post-NMO decon filters ($t\,\tau$-decon) is based on the following model:

$$data \approx \frac{1}{1-bub}\ \frac{1}{t}\ NMO^{-1}\ \frac{random}{1-reverb} \tag{1}$$

First, random white noise is divided by a reverberation filter. (An example of *reverb* expressed in Z-transform notation is $c\ Z^n$ where c is the sea-floor reflection coefficient, and Z^n is the delay operator for vertical travel time). The result of the division is thrown out into hyperbolas via inverse NMO; a spherical divergence multiplier $1/t$ is applied; and the final result is convolved with a bubble wavelet to give the synthetic data. In this model the bubble wavelet is really $bubble = 1/(1-bub)$. So *bub* is not a physical variable at all, but a convenient mathematical function that we need for the optimization. In this model the same bubble wavelet is applied to all offsets, i.e. the bubble is presumed to be an isotropic radiator. Anisotropy is in the *reverb /NMO* interaction.

As an equation for modeling seismograms, (1) is very simple. But inversion is more difficult than modeling and (1) provides a more detailed physical model than ordinarily underlies deconvolution. The model could be further generalized to reflection coefficient as a function of angle by changing NMO to a generalized radial trace moveout. The experimental study at the end of the paper suggests that the most significant factor not incorporated in (1) is the different moveout velocities of different events.

Invert (1) by premultiplying in turn by the inverse of each of the operators.

$$(1-reverb)\ NMO\ \ t\ \ (1-bub)\ data = random \tag{2}$$

Both *bub* and *reverb* are unknowns to be solved for. Linearize (2) by neglecting the product of these two unknowns.

$$NMO\ t\ data\ -\ random\ =\ reverb\ NMO\ t\ data\ +\ NMO\ t\ bub\ data \tag{3}$$

To implement (3), the unknowns, *bub* and *reverb* are taken to be filters with a gap separating them from the unit pulse, i.e. in Z-transform notation, they are typically of the form:

$$bub \text{ or } reverb\ =\ a_{15}Z^{15} + a_{11}Z^{11} + \ \cdots\ + a_{45}Z^{45} \tag{4}$$

If we have difficulty with the linearization, we increase the gap 15 thereby decreasing the predictability hence decreasing *bub* and *reverb* and their non-linear product. Whenever the bubble is estimated, it can be removed and the process can be repeated with a smaller nonlinear product. Thus the 15 can gradually be reduced (theoretically) to a smaller gap. In practice I haven't done this. I have been more concerned with testing many data sets than with crafting the best job on a single data set.

Equation (3) is representable as an output power minimization merely by asking sum squared of *random* to be as small as possible. Thus *random* is the prediction error. By the usual techniques of least squares analysis, this may be written as an overdetermined system.

$$NMO\ t\ data\ \approx\ reverb\ NMO\ t\ data\ +\ NMO\ t\ bub\ data \tag{5}$$

This is not so simple a thing as a two-channel least-squares prediction problem. We will need to use an ordinary simultaneous equation solver such as conjugate gradients method. Let us see how (5) can be represented with matrices. Define $d = data$, $d' = NMO\ t\ data$, r a dereverberation filter, and b a debubble prediction filter. Let $*$ denote convolution. The regression (5) becomes

$$d'\ \approx\ r * d' + NMO\ t\ b * d \tag{6}$$

As matrices in (6), t is a diagonal matrix with the divergence correction t along its diagonal. Likewise, the *NMO* matrix is a square matrix, identical to a (z, t)-plane containing all zeros except an interpolation operator centered along the hyperbolic trajectory $v^2t^2 = z^2+x^2$.

The convolution $b * d$ is representable as a matrix times a vector in either of two ways. First, like FGDP equation 7-1-1, the *seismogram* enters the matrix repeatedly, as a succession of down-shifted columns. Alternately, the *filter* could enter the matrix as a succession of down-shifted columns. When I look at (6) I think of the *filter* being in the matrix. When I implement (6) in an optimization program to solve for the unknown filters r and b, I think of the *data* being in the matrix, and the filters r and b being column vectors, that is, I rewrite (6) as

$$d' \approx [d' *]\, r + NMO\ t\ [d\ *]\, b \tag{7}$$

and think of $[d\ *]$ as one of those down-shifted column matrices. The prediction gap enters by the downshift of the data within the matrix. The regression (7), is the main point of this paper. All my field data tests were implemented with (7).

The regression (7) applies to divergence-corrected moveout-corrected data. Alternately the regression my be formulated on the raw data. Operating on (7) with $t^{-1}\, NMO^{-1}$ yields

$$d \approx \frac{1}{t}\, NMO^{-1}\, [(NMO\ t\ d)\ *]\, r + [d\ *]\, b$$

$$d \approx [d\ *]\, b + \frac{1}{t}\, NMO^{-1}\, [d'\ *]\, r \tag{8}$$

Wedge weights

Multiple channels are incorporated by placing matrix equations like (7) [or (8)] beneath one another. Additionally, the conjugate-gradient method allows systems like (7) to be multiplied by weights. Weighting functions in the (t, x)-plane could be anything. The most basic weights are simply windows, i.e. functions whose values are either zero or one. Typically these windows are wedge shaped like mute windows. They select portions of the data that can be expected to have a common dereverberation filter, typically the space between two separated radial traces. So, a weight function for a typical x is a box-car function of t. This box-car function may be placed on the diagonal of a matrix and premultiplied onto (7). Placing systems like (7) beneath one another to allow many channels, places many different box-car functions end to end along the diagonal weighting operator. With these interpretations in mind, for a weighted multichannel estimation we write

$$W\ \{\ d' \approx [d'\ *]\, r + NMO\ t\ [d\ *]\, b\ \} \tag{9}$$

Now let us generalize the method still further. Since reflection coefficient is a function of angle we may want a different reverberation filter in different wedge shaped windows in the (t, x)-plane. But the problems in different windows are not independent, because the bubble filter should be the same. For two fitting windows we have the two overdetermined systems to solve simultaneously:

$$W_1\ \{\ d' \approx [d'\ *]\, r_1 + NMO\ t\ [d\ *]\, b\ \} \tag{10a}$$

$$W_2\ \{\ d' \approx [d'\ *]\, r_2 + NMO\ t\ [d\ *]\, b\ \} \tag{10b}$$

The set (10a,b) is a ready extension to my computer program. Since the number of iterations is always taken to be 10 regardless of the number of parameters being sought, the cost of adding on (10b) increases the cost by less than a factor of two. But I haven't done it, because I am unsure which of many generalizations will be most useful. The answer seems largely data dependent.

Exotic domains

In a program for the regression (10a,b), normal moveout correction can be changed to radial-trace moveout correction. This is a more natural direction for the prediction, lowers the cost of (10a,b), but complicates the issue of missing channels. Likewise we could go to the slant-stack moveout domain to acquire some well known theoretical advantages (FGDP) when several velocities are present simultaneously.

These "exotic" domains often introduce practical difficulties transforming back to the original physical domain. Nobody likes a lot of artifacts in their data. The trick here is to return from the exotic domain not with the prediction error but with the prediction itself. Since the prediction power should always be smaller than the data power, the miscellaneous errors of transforming to and from the exotic domain are also weakened. This is a reason to base the regression on (8) instead of (7). The regression (8) might seem to be more difficult because it involves inverse NMO instead of NMO, but I don't think this really makes it significantly more difficult. Probably transpose NMO would work as well as inverse NMO anyway. Had I considered these factors at the time, I would have based my program on (8) instead of (7).

Surface consistent decon

One way to implement surface-consistent deconvolution is as a "round robin" of surface deconvolutions. A more elegant approach is for computations within the construction of the decon filters to be surface consistent. For example, the familiar Levinson and Burg methods sum numerators and denominators (which divide to give the reflection coefficients). The sums ordinarily include only a single trace, but they could be extended in a surface consistent way.

There is a filter xg(s) for every s and a filter xs(g) for every g. Let $[A\,(s\,,\,g\,)\,*]$ denote convolution over t for each s and g. Likewise $[*\,A\,(s\,,\,g\,)]$ denotes the transpose to convolution. The inner part of the conjugate gradient iterations would look like:

```
loop on conjugate gradient iterations
    loop on geophone
        loop on shot
            if not transposed
                yg(s) = yg(s) + [A(s,g) *] xg(s)
                ys(g) = ys(g) + [A(s,g) *] xs(g)
            if transposed
                xg(s) = xg(s) + yg(s) [* A(s,g)]
                xs(g) = xs(g) + ys(g) [* A(s,g)]
```

EXAMPLES

Hard bottom data from Canada

Figure 0 shows a marine profile from Canada (Yilmaz and Cumro #27) after t^2 gain and NMO. The center panel was also deconvolved before the NMO. The deconvolution process estimated a single filter for the whole profile. The estimation process used portions of five traces shown in the right hand panel. From the right hand panel, you can see that the portions used were all in a small propagation angle wedge, and that 4 out of 5 traces were ignored in the estimation process (to speed the program). The filter is displayed aside the deconvolved profile, and you can see the 60ms gap between the leading spike and the regression coefficients. Subjectively, the results of the deconvolution are excellent near zero offset, but they deteriorate with offset. Objectively, the power in the output is substantially less for the inner traces after deconvolution. At late times we notice that the deconvolution has also worked somewhat beyond its fitting interval, which is encouraging.

Figure 1 repeats the calculation of figure 0, but the deconvolution was done *after* NMO. The results are of the same general quality within the fitting wedge, but the validity outside the fitting interval now extends much further. This is really encouraging!

Figures 2 and 3 repeat the experiments of figures 1 and 2, but the fitting is centered about a large angle. We see the same general results with the quality of the deconvolution now deteriorating toward vertical incidence.

The experiments of figures 1-4 showed that for this data, the results of dereverberation were more impressive than the results of debubbling. In figure 5, I chose a favorable set of weights for dereverberation. Figure 6 sets up the same weights for simultaneous debubble and dereverberation. We see the debubble filter is slightly smaller than the dereverberation filter. It is disappointing that the simultaneous deconvolution is not perceptibly better than simple dereverberation.

In the time remaining before the report deadline, a number of other experiments were tried, though none of them were particularly conclusive. In figure 6, the fitting region was passed all the way through the first arrivals into the region where the NMO stretch is extreme. This didn't seem to hurt anything. The output power is clearly reduced near the first arrivals, as it should be.

The Backus filter may be expressed as $(1 + cZ)^2$, the square refering to reverberation at both the shot and the geophone. In figure 7, I doubled the length of the dereverberation filter, to test the idea that it had been too short. The resulting filter does indeed resemble the Backus function $(1 + 2cZ + c^2Z^2)$, which is encouraging, but the quality of the deconvolution has not been subjectively improved from previous figures.

A variety of other ideas were tested, but not shown, because the results were not conclusive. For example, the appearance of the data before NMO suggested that the bubble function is really long, perhaps 1.5sec. On the test I made, I found that subjectively, the bubble in the data appeared to be *longer* after I removed it than before. This is the result of too many degrees of freedom. Some kind of spatial filtering incorporated in the deconvolution estimate might help.

I still feel optimistic about achieving a convincing extraction of bubble and dereverberation from this data. The remarkable flexibility of the conjugate gradient approach enables us to construct the more complicated models that seem to be required for this data set. For example, surface consistent thinking suggests that this single profile should be dereverbed with two filters, a shot filter which is constant across the gather, as in this study, and also another filter that is allowed to vary from trace to trace, because successive geophones see a different sea floor. Of course with all these extra filters, there are potentially a lot of extra degrees of freedom, but I think there are realistic ways of limiting this difficulty.

None of the deconvolutions was good enough to fully suppress the peglegs at 1.7sec and at 2.4sec. I attribute this to a significant feature of this Canadian marine data that was not incorporated in the model. There are clearly observable differences between the velocity of primaries, peglegs, and seafloor multiple reflections. Studying the peglegs that were not removed completely, you can observe that the reverberation period lengthens by a half wavelength at wide offset. So a more complicated model is called for.

Other data

Seven other field profiles were tested, none giving such clear results as the Canadian data, so the results are not shown. One was John Toldi's Chevron Central Valley data, and six were taken from the Yilmaz and Cumro catalog. These were

01	South Texas	vibrator
08	Central America	dynamite
09	Alaska	vibrator
13	Offshore Louisiana	airgun
14	Offshore Texas	Aquapulse
39	Middle East	Geoflex

Deconvolution, unlike migration, is mainly a study of the earth's surface, which is extremely variable from place to place, so although positive results were not obtained, I do not feel that good results cannot be obtained, only that more work will be required and results will be more modest, or that results may be good, but not self-evidently so. For example, the land data sets generally have ground roll and I did not have time to assure the ground roll was excluded from the fitting region.

REFERENCES

FGDP refers to *Fundamentals of geophysical data processing* by Jon F. Claerbout: first published in 1973 by McGraw-Hill and reissued in 1985 by Blackwell Scientific Publications.

Luenberger, D. G., 1973, Introduction to linear and nonlinear programming: Addison-Wesley.

Paige, C.C. and Michael A. Saunders, 1982 LSQR: An algorithm for sparse linear equations and sparse least squares, ACM transactions on Mathematical Software, vol 8, no 1, p 43-71

Paige, C.C. and Michael A. Saunders, 1982 Algorithm 583, LSQR: Sparse linear equations and least squares problems, ACM transactions on Mathematical Software, vol 8, no 2, p 195-209

Wang R.J., and Treitel, S., 1973, The determination of digital Weiner filters by means of gradient methods: Geophys., **38,** 310-326.

Yilmaz, O., and Cumro, D., 1983, Worldwide Assortment of Field Seismic Records, released by Western Geophysical Company of America, Houston

Ziolkowski, A., 1984, *Deconvolution,* International Human Resources Development Corporation, Boston.

Fig. 3. Single filter/profile deconvolutions.

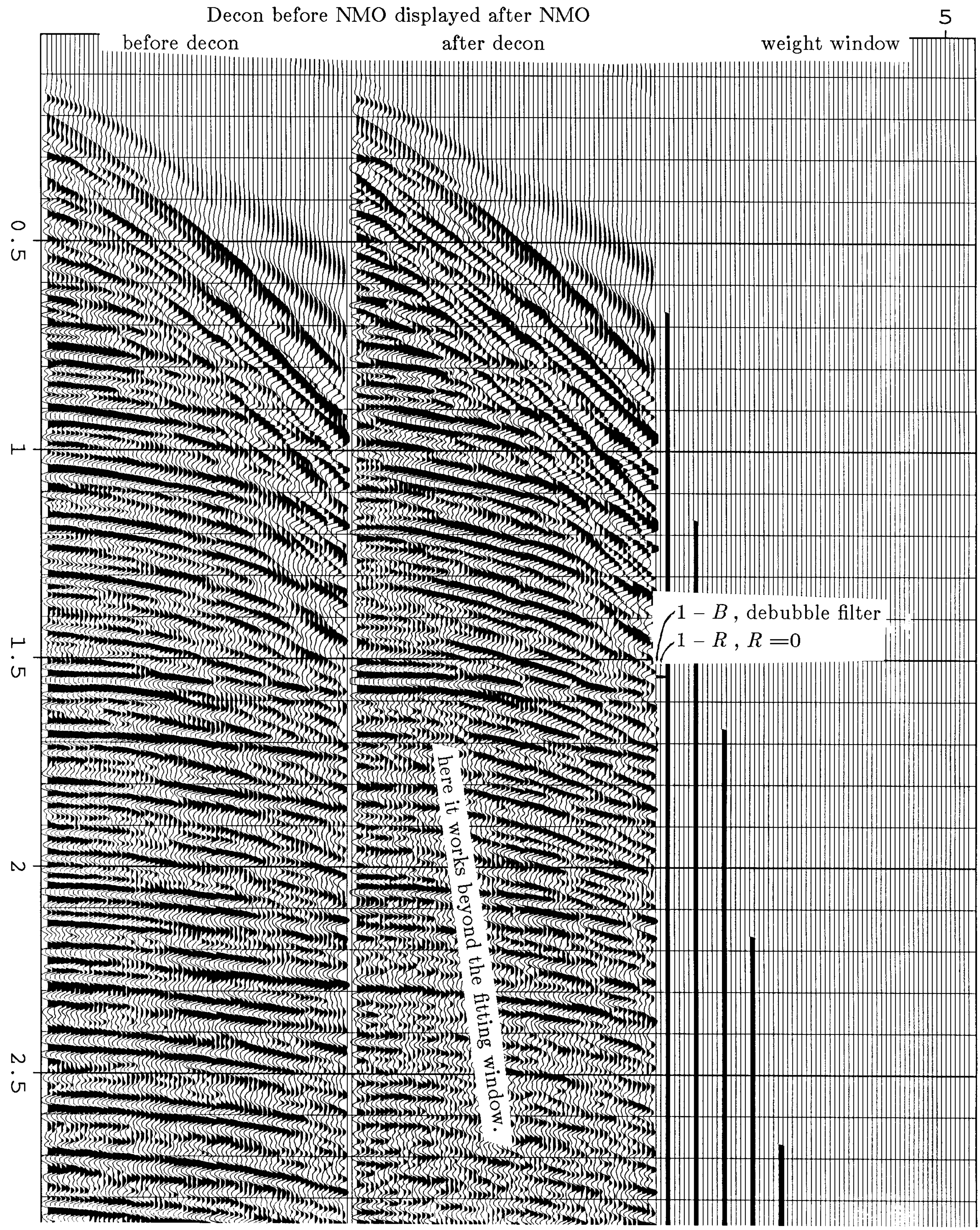


Fig. 4. Decon after NMO displayed after NMO.

Fig. 5. Debubble weighted away from zero offset.

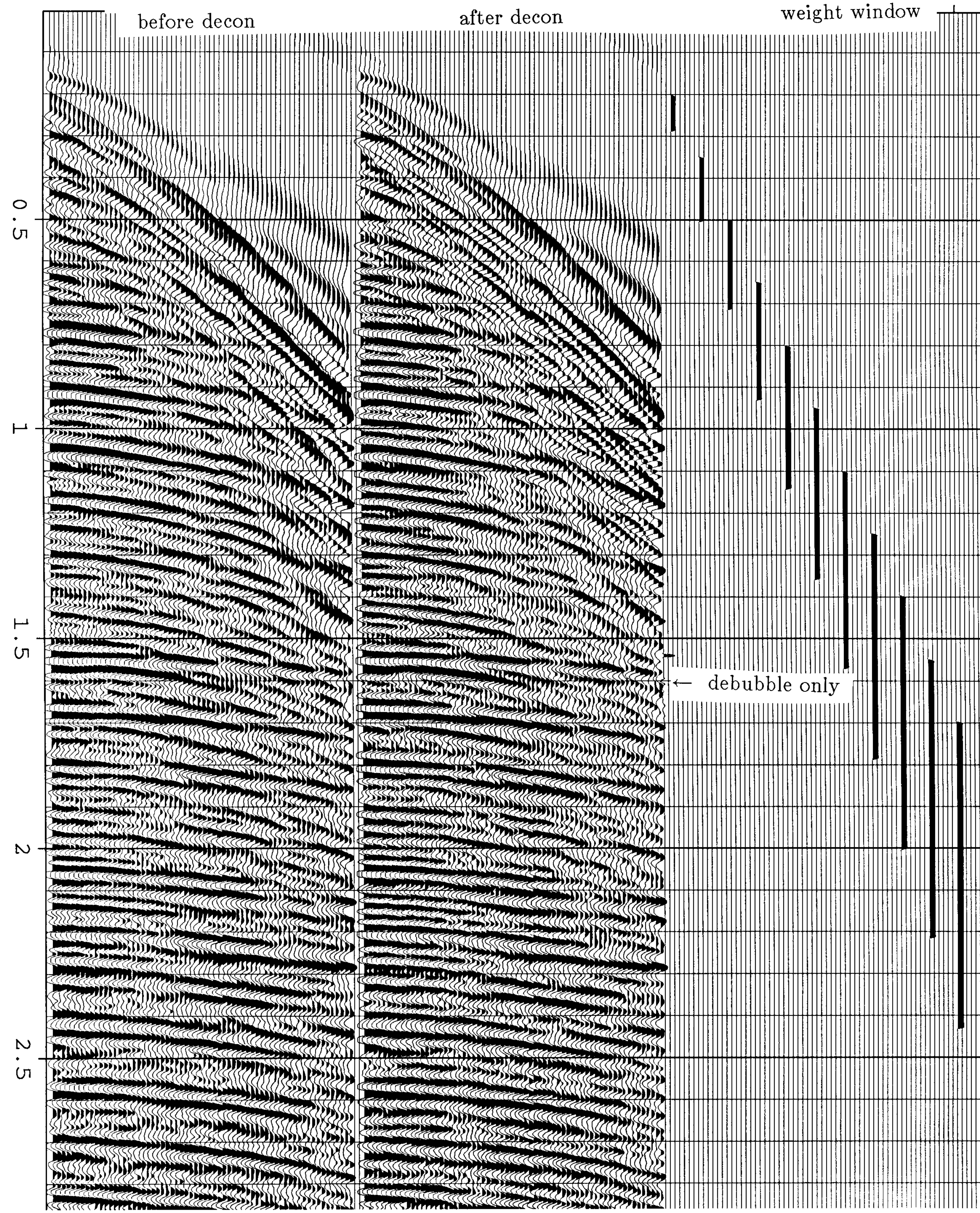


Fig. 6. Dereverb weighted away from zero offset.

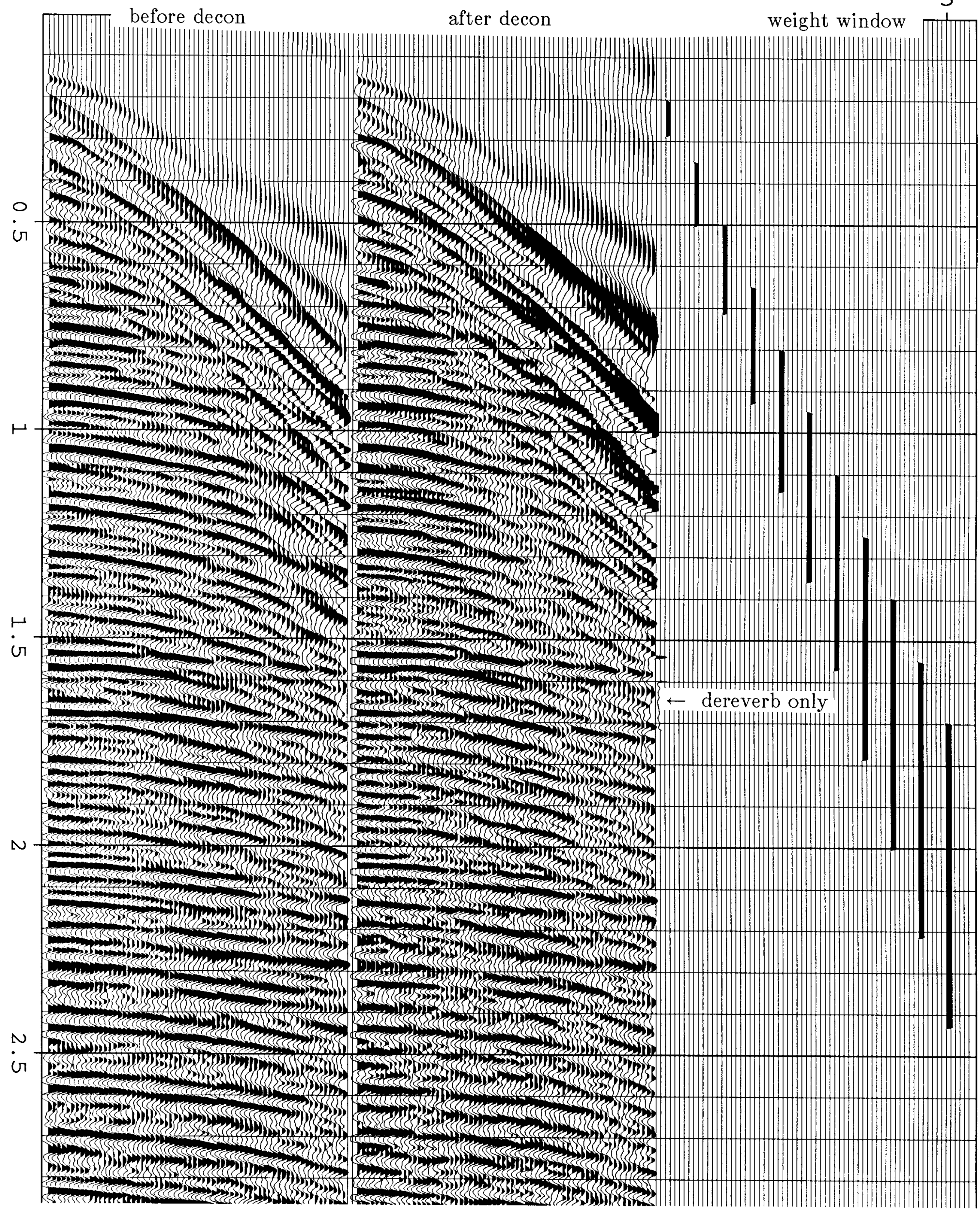


Fig. 7. Dereverb with a larger fitting window.

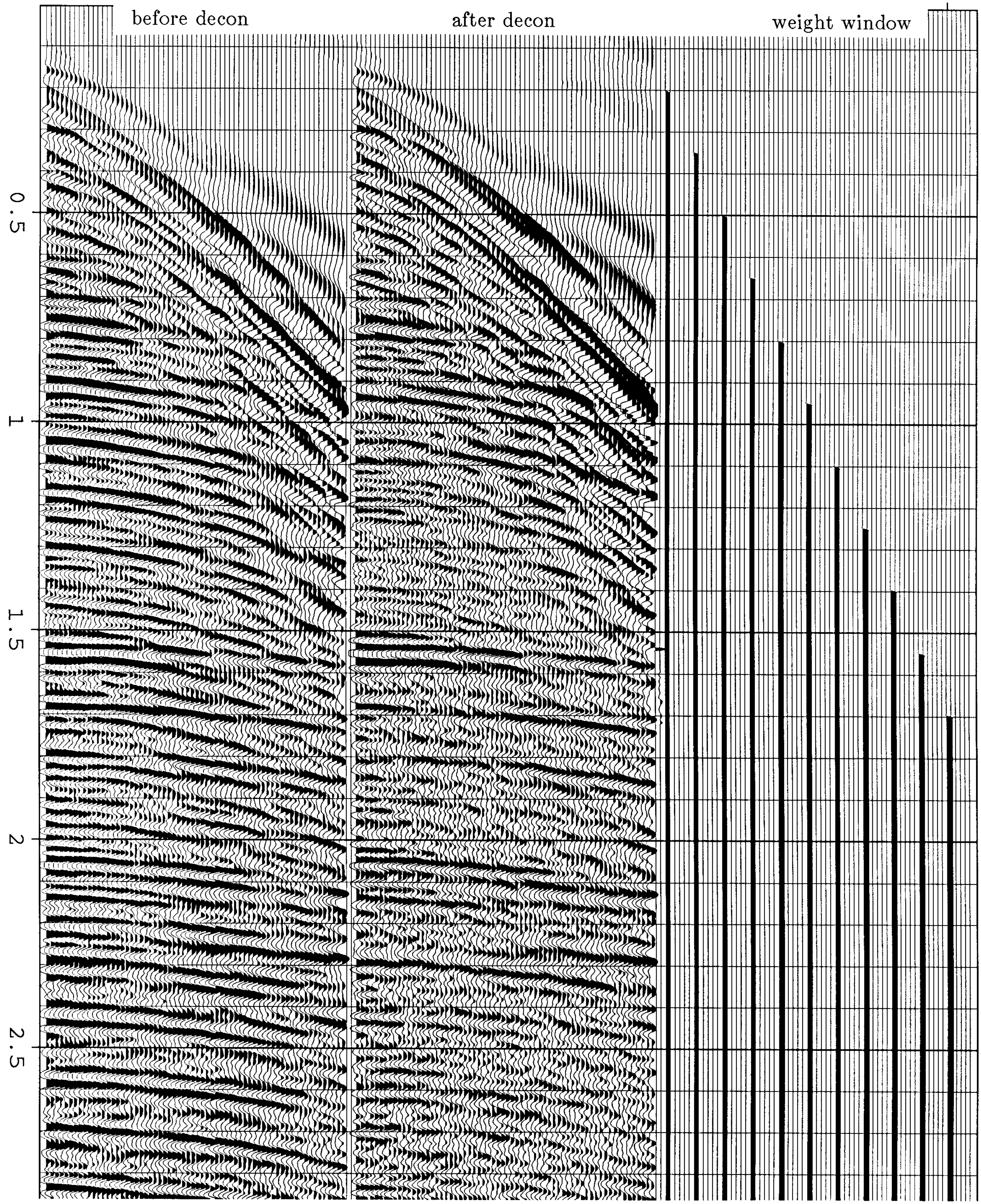


Fig. 8. Simultaneous dereverb and debubble.

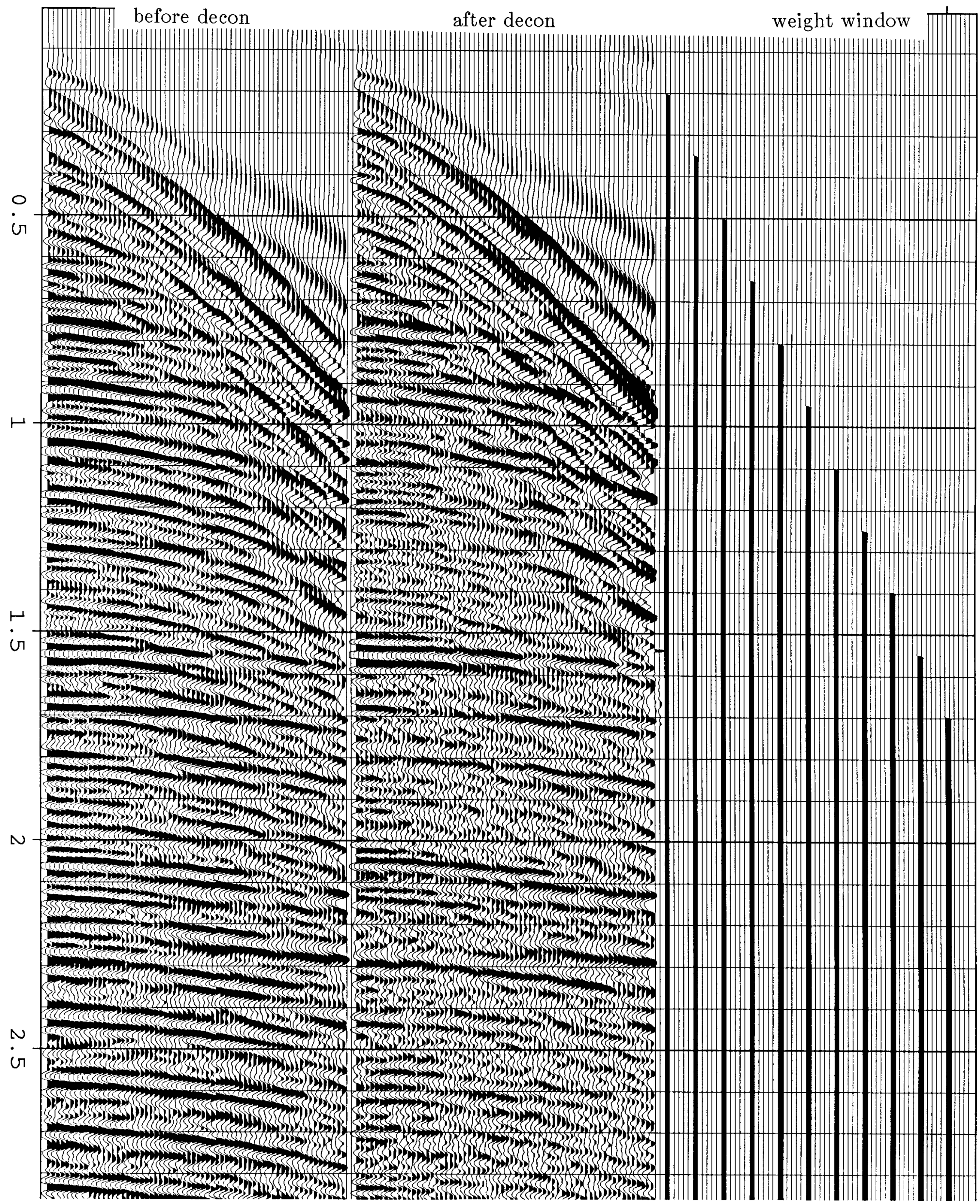


Fig. 9. More anisotropic decon, fitting everywhere.

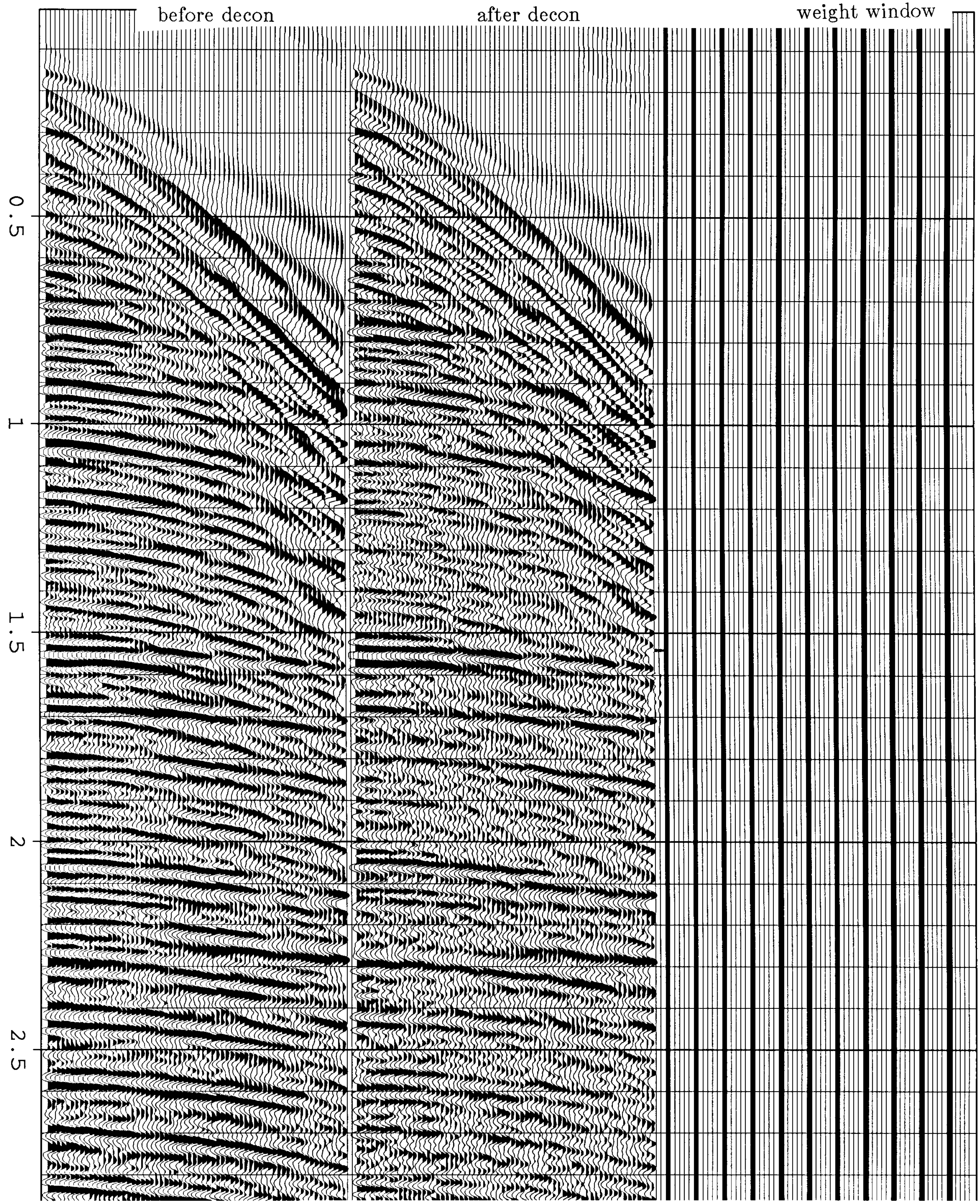


Fig. 10. More anisotropic decon, double length dereverb filter.

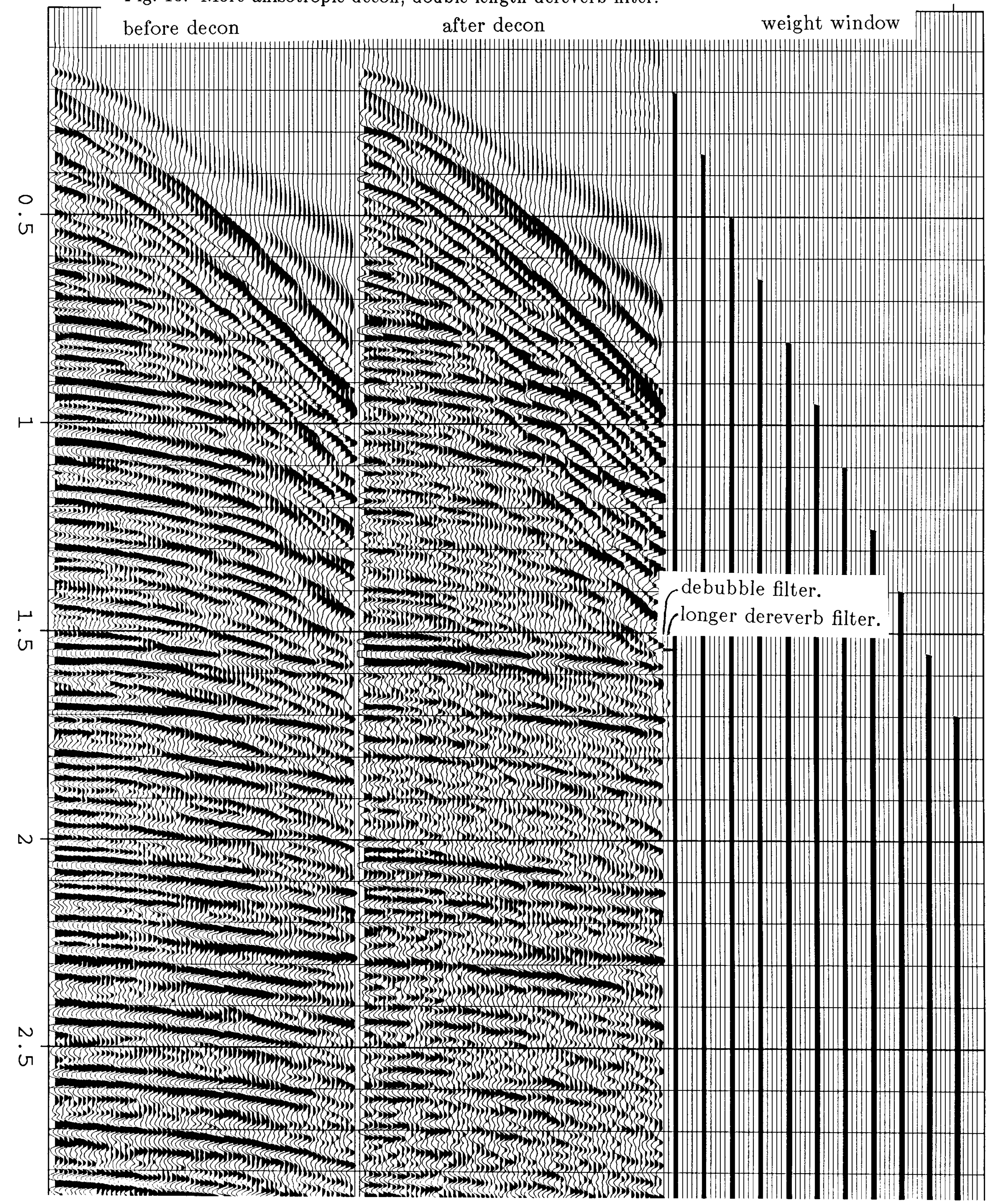


Fig 0. Single filter/profile deconvolutions

Decon before NMO displayed after NMO

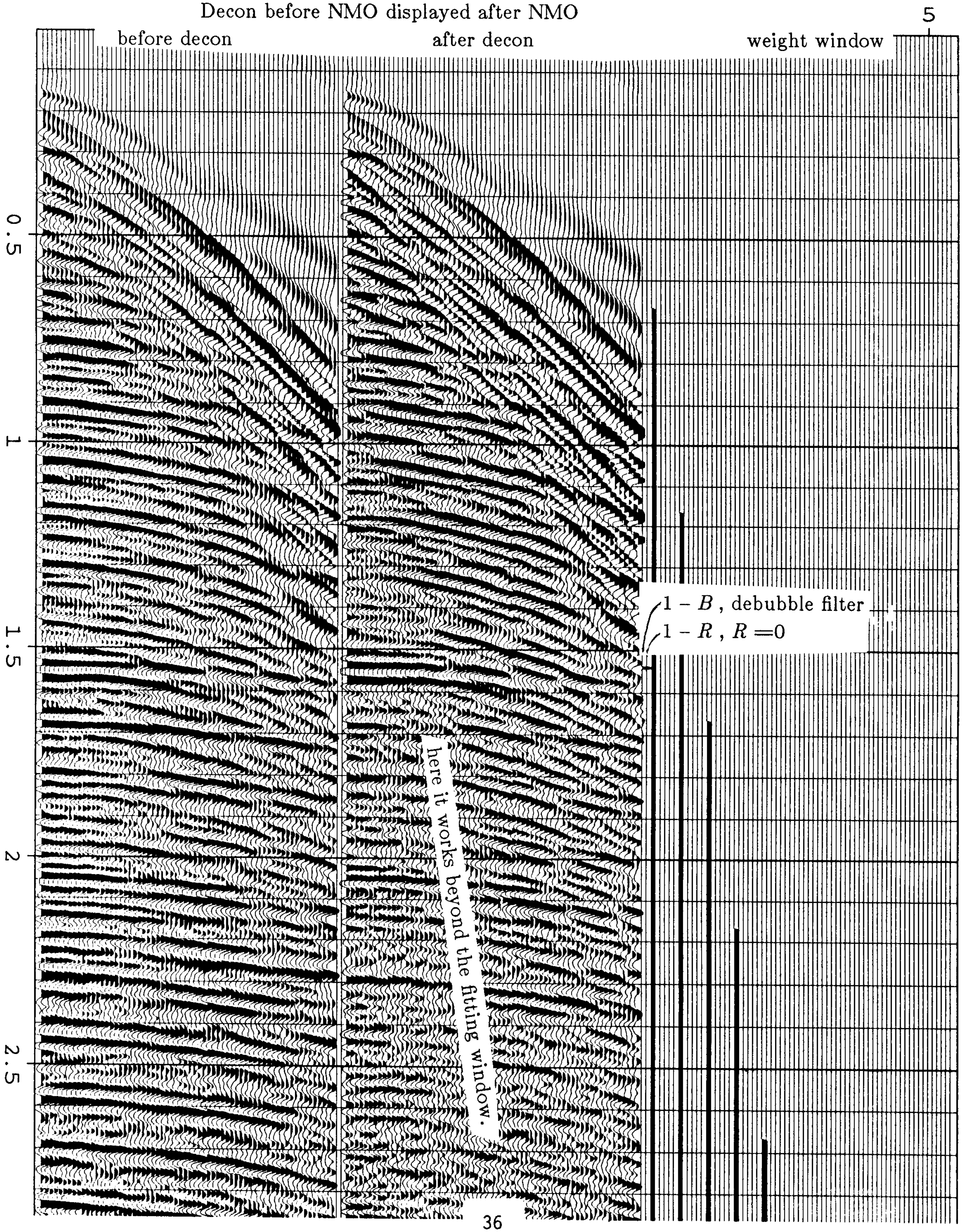


Fig 1. Decon after NMO displayed after NMO

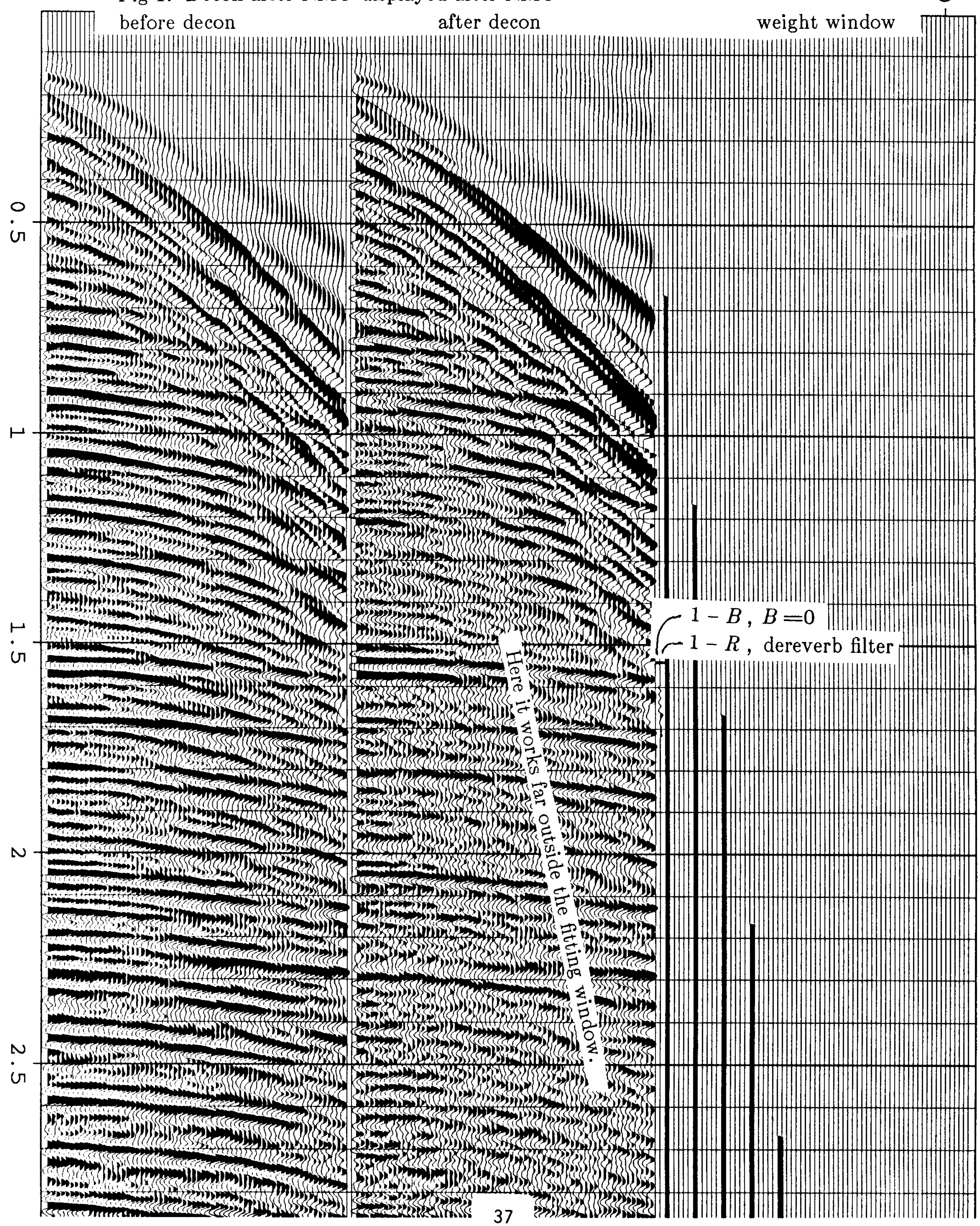


Fig 2. Debubble weighted away from zero offset.

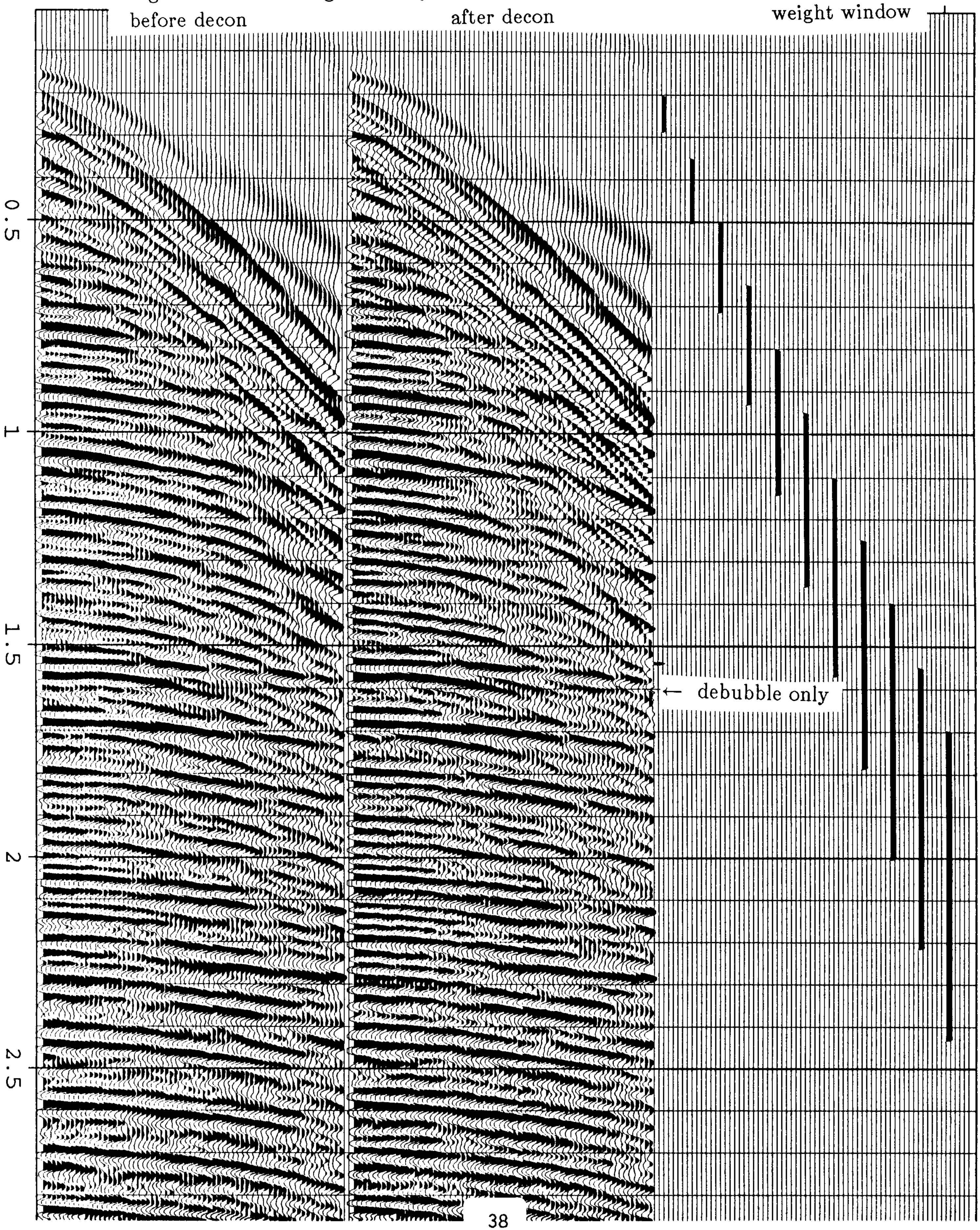


Fig 3. Dereverb weighted away from zero offset.

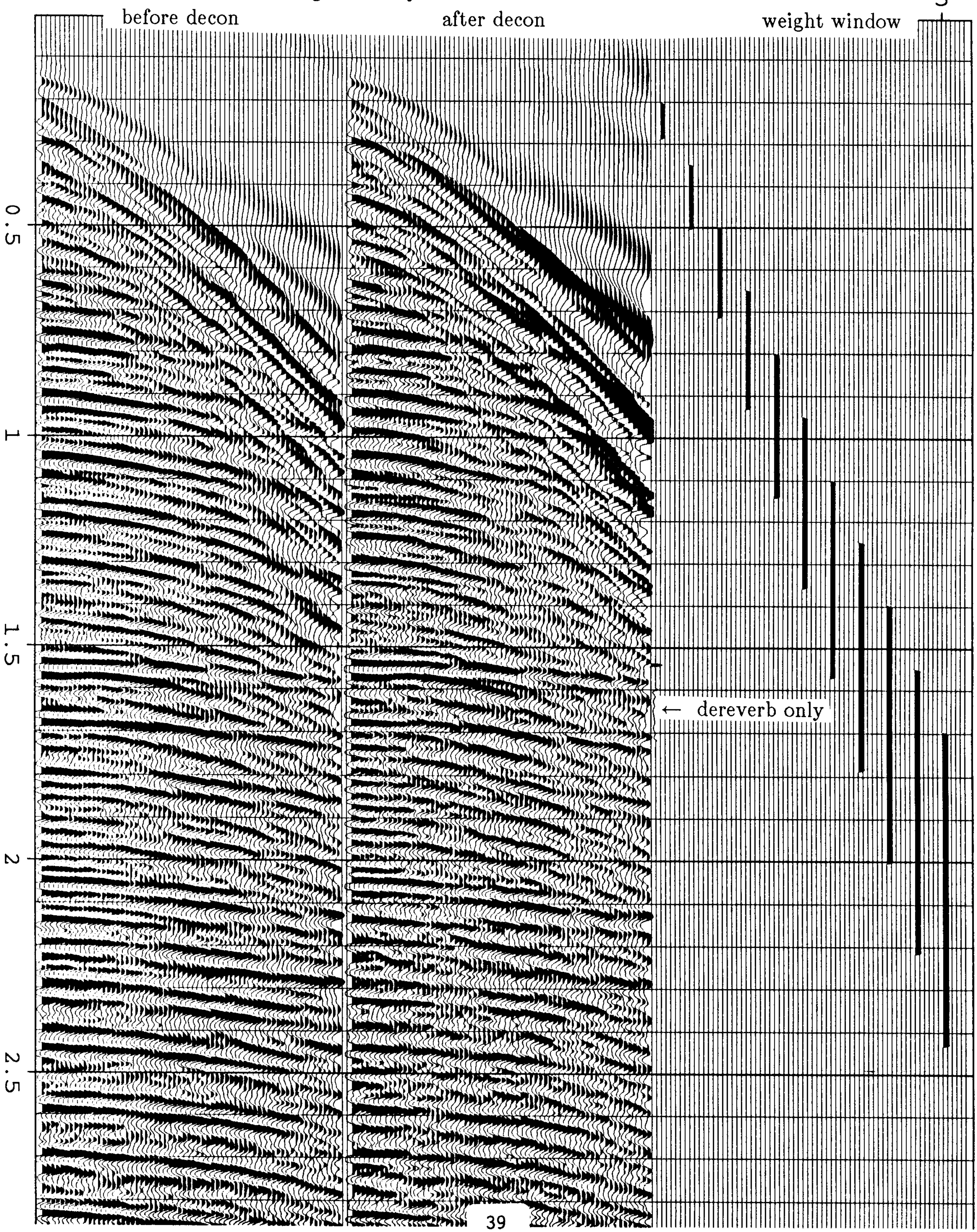


Fig 4. Dereverb with a larger fitting window.

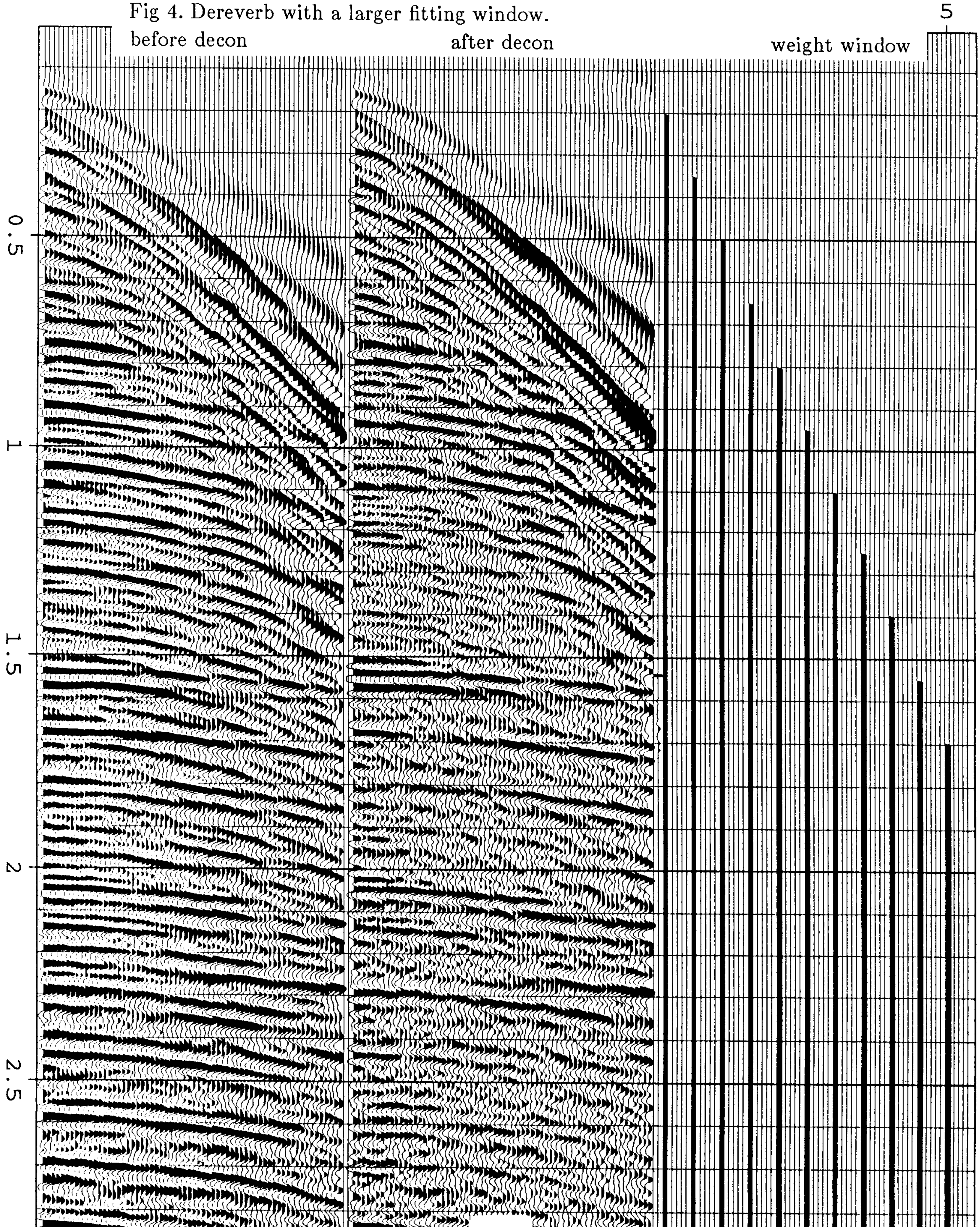


Fig 5. Simultaneous dereverb and debubble.

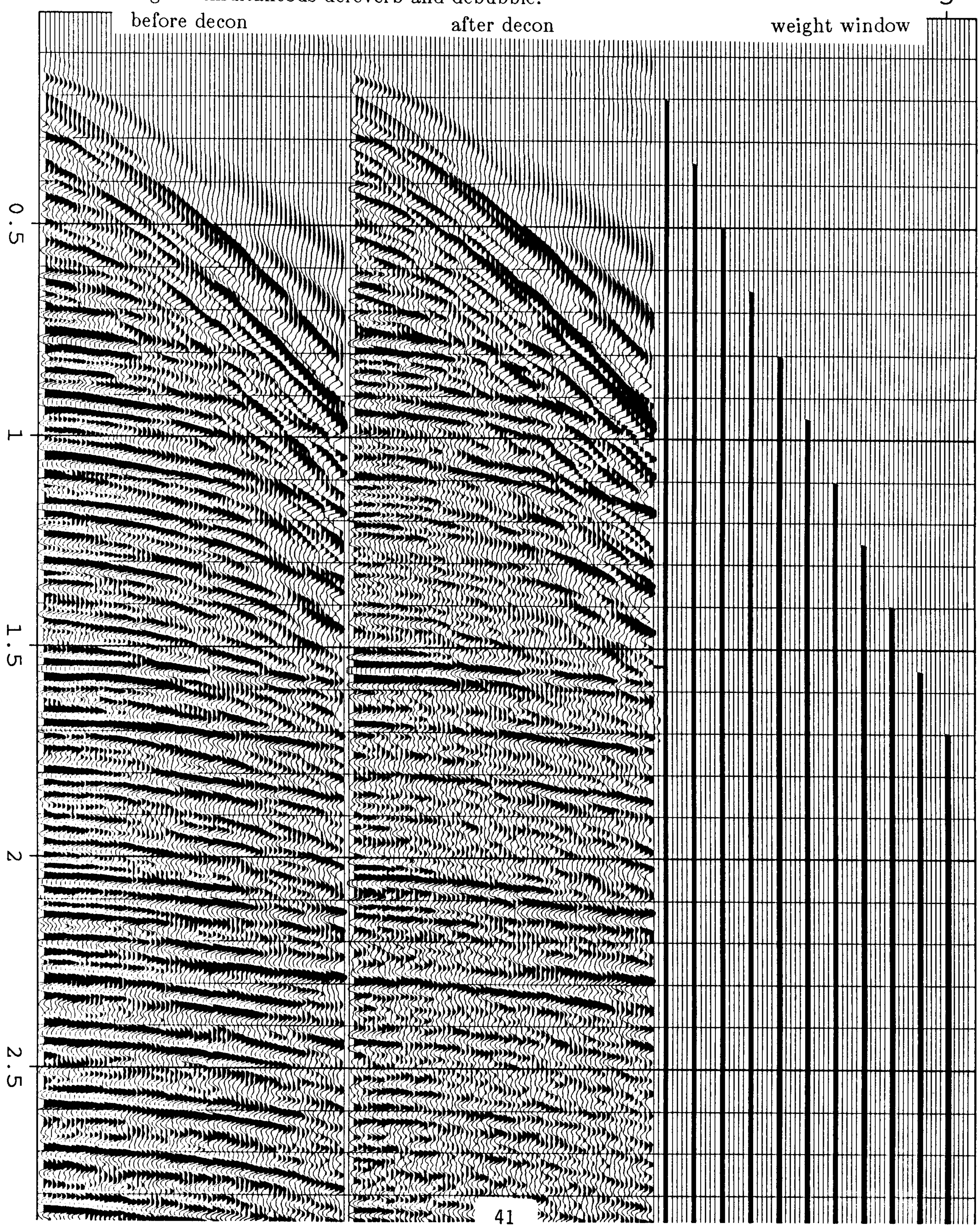


Fig 6. More anisotropic decon, fitting everywhere.

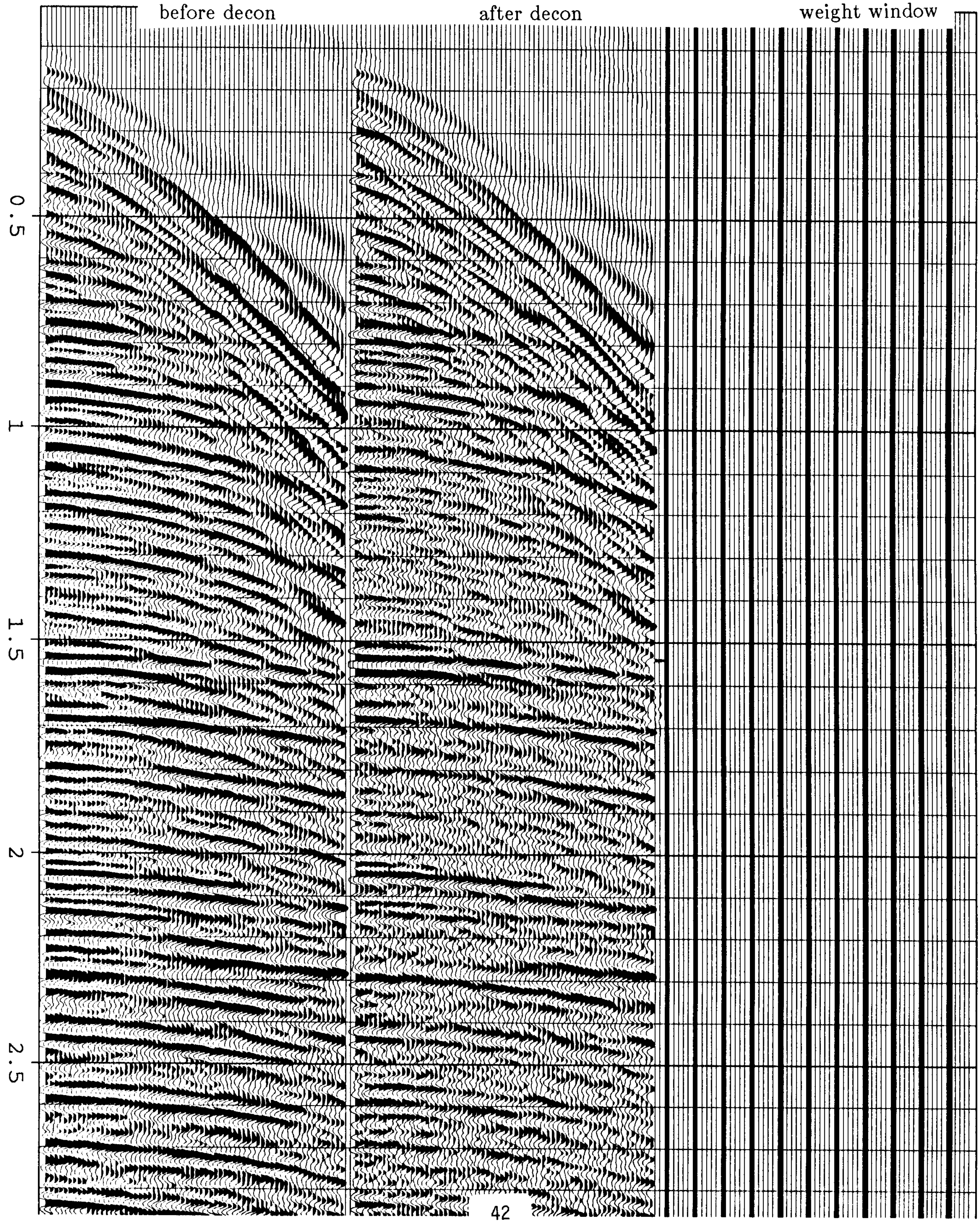
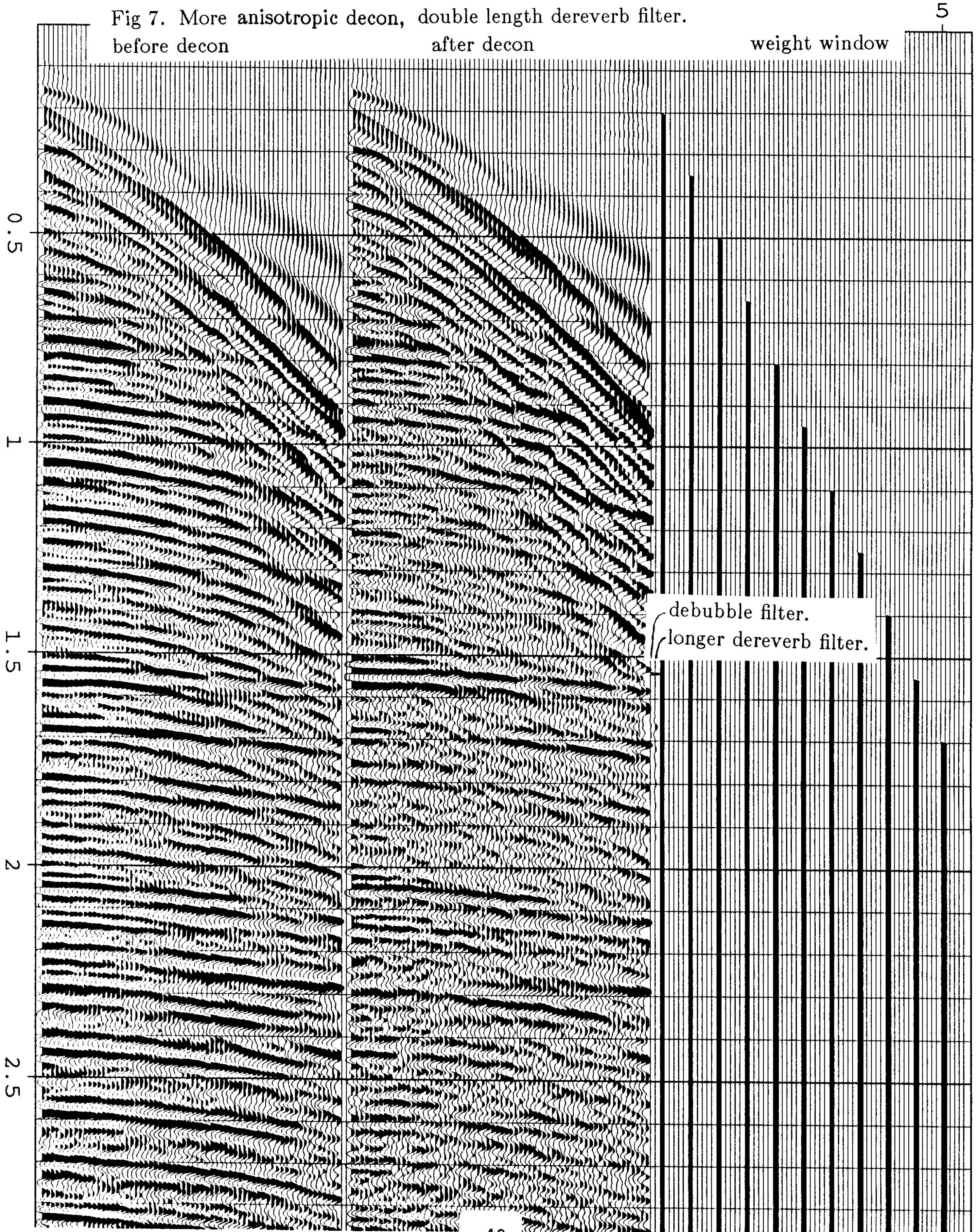


Fig 7. More anisotropic decon, double length dereverb filter.



Soviet Industry Grows Its Own Potatoes

By Vladimir Kontorovich
And Vladimir Shlapentokh

The most consistent pessimists on the prospects for the Soviet economy are to be found not in the West, but among many Soviet managers and economists themselves. They gloomily attribute economic stagnation to the vested interests of the bureaucracy, which resists any serious organizational change and seems itself impossible to dislodge. However, dynamic forces in the "informal economy" may yet at least stem the decline.

This informal economy must be distinguished both from the official economy, where everything is done in accordance with the law, and from the "second economy," the underground sector in which participants are absorbed in personal enrichment. The informal economy is the sphere of economic activity in which Soviet managers seek to achieve official, legal goals—above all, the fulfillment of the plan—through illegal means.

The informal Soviet economy is, to some degree, reminiscent of the European absolute monarchies of several hundred years ago. The then-emerging capitalist class employed thousands of tricks to combat the countless royal rules that blocked economic growth. Today, Soviet managers similarly seek to outsmart the authorities in order to increase production and implement technological innovations. Some of the most significant procedures are those used to cope with procurement problems, the devices used to prevent production stoppages resulting from the caprices of numerous suppliers.

As is commonly known, Soviet industries are plagued by the problems of inadequate and late supplies. Factories and workers often sit idle, or function well below capacity, due to the unavailability of raw materials or replacement parts. In consequence, the pace of work is frequently erratic, with work forces scurrying furiously to get out one month's production in a few days, after awaiting the arrival of some needed supplies. Despite these difficulties, managers are still held accountable for meeting production plans, which often results in fabricated reports showing that plans have been fulfilled.

The informal economy helps to solve these problems in two, essentially opposite, ways. One is the trade (usually barter) between enterprises. This serves to correct planners' errors, as enterprises exchange things they have received but do not need for things they need but were allocated to others. A market mechanism, pushed out of sight, thus helps the planned economy to function.

Another no less widespread method of informal decentralization is more unorthodox. While conventional entrepreneurs see the division of labor as natural, Soviet managers boldly defy Adam Smith's prescriptions and tend to diminish this division as much as possible.

Enterprise directors cut their dependence on unreliable outside suppliers by setting up in-house production of everything imaginable. Plants, both large and small, design and produce equipment for their own use. Moreover, they manufacture most standard parts for their own equipment, such as castings. For example, a Volga automobile plant not only makes robots for its own use, but also produces integrated circuits for them. A shoe plant could be mistaken for a machine factory or a chemical plant, since it makes its own equipment and glue.

This enterprise autarky is most widespread in the repair and servicing of equipment and the production of spare parts, functions traditionally ignored by central planners. But the construction of buildings may be done by enterprises themselves, and transportation by their own fleets of trucks. The management of a refrigerator railroad car repair depot in one city is planning to build an electric generating station, since the state power network suffers from frequent blackouts.

Autarky is spreading even in the computer industry. As a recent article in Literaturnaya Gazeta suggests, computer centers rely less and less on the assistance of specialized firms and create their own programs, as well as repair computers themselves or with the aid of free-lance specialists. The director of one of the most prestigious computer centers in the country, at the Siberian section of the Academy of Science, calls for all of his colleagues to move toward "self-service."

Many enterprises are also increasingly producing goods and providing services for their own workers. In urban "agricultural shops," workers and engineers grow vegetables and raise cattle that they consume themselves. In the industrial region of Kemerovo, some 40% of potatoes are produced by industrial enterprises.

Of course, these in-house facilities also need materials and equipment. These are taken out of planned allocations for other purposes, or are obtained through barter with other enterprises. Indeed, in-house facilities may even produce goods for the purpose of barter. In this way, the informal market and informal autarky reinforce each other.

The informal economy proves that Soviet managers, officially merely employees executing orders from above, possess a great deal of entrepreneurial spirit. The level of innovation undertaken by many of these managers is, in the Soviet context, quite inspiring.

Soviet planning officials and political leaders, however, condemn enterprise autarky, for it contradicts one of the fundamental principles of economics: The advantages of the division of labor. They cite standard metalworking products, 70% of which are produced in specialized plants in the U.S., while only 3% to 4% are so produced in the U.S.S.R. The costs of in-house manufacturing have repeatedly been found to be much higher than those of specialized plants. Equipment in the "auxiliary" shops is also often poorly utilized. Thus, it is argued, as the division of labor increases productivity, the trend toward autarky lowers it.

In reality, the condemnation of enterprise autarky by Soviet leaders is purely ritualistic. They are fully aware of the beneficial effects of these practices and therefore do not seek to suppress them and in some cases may even encourage them. The Soviet media even depict managers who establish in-house production as taking good care of their enterprises.

There is indeed a case in favor of enterprise self-sufficiency. The greater the division of labor, the greater the need for the coordination of specialized producers, and this coordination is not without its costs. The larger the number of producers, and the number of goods produced, the greater the burden on central planners and the higher the level of disruptions of supply and production.

In our opinion, the increased production costs of autarky allow the Soviet economy to avoid still greater losses from inadequate supplies. The fact that the degree of plant specialization in the U.S. is much higher only means that the cost of coordination in market economies is much lower than in planned ones.

Mr. Kontorovich lives in Haverford, Pa., where he is a researcher on the Soviet economy. Mr. Shlapentokh is a professor of sociology at Michigan State University in East Lansing.