Baseline asymmetry and tau-space projection

Technical report THEM2001-02
February 2001

Daniel Lemire, Ph.D.
(514) 524-3517
Report funded by **THEM Geophysics Inc.**

Keywords: Tau domain, splines, wavelets

Introduction

This report assumes that the reader is familiar with **THEM Geophysics Inc**. THEM system is an innovative EM system appropriate for rough terrain mineral exploration. It was developed over the last 4-5 years and is still considered a prototype.

One of the unresolved issue has been the noise level. It is only one of several issues of course, but signal-to-noise ration seems to be a convenient way to measure the quality of an EM system and is of significant commercial interest. Unfortunately, there isn't a standardized measure of the noise level because processing software and hardware vary greatly from one system to another.

At this point in time, it seems that processing the signal itself (and not subproducts such as channels) might be a priority because there are qualified consultants able to prepare the various subproducts in a satisfying manner for the industry. Two problems have been identified over time : is baseline correction is sufficient and how to remove the high frequency noise in the signal?

Note: This report is a follow-up on THEM2001-01 and some of the text remains the same.

Data under consideration

We will studying the file bob3.dat provided by THEM Geophysics Inc. The file is 217 Mbytes in size and dated from April 21st 1999. X and Z coils are available but only Z coil will be considered here. This analysis may not apply to all collected data by THEM EM system and the reader should be aware that there might be components in the signal which are specific to this particular file. It is however believed to be a representative signal set by **THEM Geophysics Inc.** It is assumed that the emitter was set at 30 Hz so that we have 60 half-cycles (or *frames*) per second and 512 samples per half-cycles (for a sampling rate of 30kHz) over both the X and Z coils. For this data, the scaling factor for conversion in PPM that 1 million divided by 8 times the standard deviation which is around 30 which means that the smallest non-zero value measured by the system is 30 PPM. Notice that the amplitude of the signal is estimated using the standard deviation and through a simple max - min formula which is more sensitive to changes in the morphology of the signal and noise during the on-time.

Definition of the problem

Current EM processing software extract all of their information from the off-time of the signal, that is, the delay during which the emitter is turned off and the signal is decaying (more or less according to an exponential curve). It is also where the apparent signal-to-noise ratio is lower (it can be as low as 1:25).

Baseline correction is achieved through the SIMn algorithm using $n = 3$ and a late window of 48 samples (see [1]). Whenever stacking is not explicitly specified we stacked each two subsequent half-cycles (simply averaging one with the other after inverting the signs). This type of data will be refered to as "raw" since no denoising was used on it.

Our goal is to increase substantially the signal-to-noise ratio and make the data more accurate. The current implementation of THEM software achieve noise reduction through stacking[1] over 6 frames or more and using an experimental wavelet denoising scheme referred to as *DeSpike* (to remove noise due to sudden and brief atmospheric discharges). This wavelet denoising scheme won't be discussed here even though it is important to handle atmospheric discharges in any future processing software.

Are pulses symmetrical? Is the baseline flawed?

It will be verified that odd and even pulses are asymmetrical in the processed data. Amplitude and morphology are slightly different. In what follows, no stacking was used (baseline correction was applied : SIMn algorithm with parameters $n = 3$ and $k_2 = 48$).

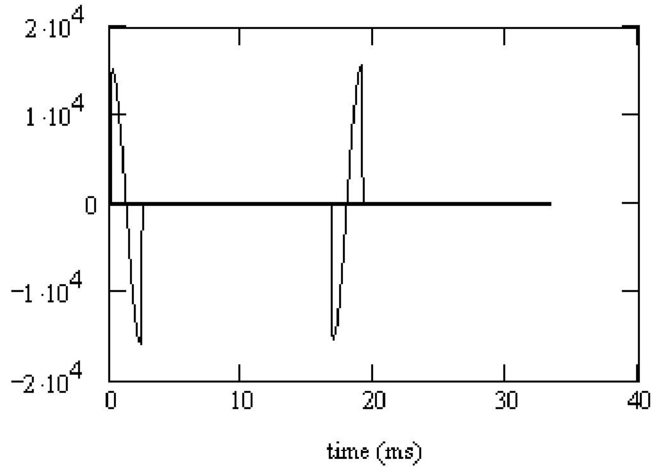
1 Stacking means that we average the responses over n half-cycles (switching signs) and then use this averaged response instead of the x half-cycles (in effect, downsampling the signal by a factor of n). We can also replace the average with the median.

While this analysis was done only on a sample of data and the exact nature of the asymmetry may vary from flight to fight, we can still conclude that unlikely what was previously assumed (by the SIMn algorithm in particular) even and odd half-cycles are not symmetrical.

A. Morphological comparison

The morphology of a full cycle should look like what is presented in Fig. 1. While the scale may vary, the first half of a cycle should match the second half with signs reversed. We will note the first half of the frame F_1 and the second half F_2 with the understanding that $F_1(i)$ refers to the ith sample of frame F_1 (in our test data, each frame has 512 samples). We therefore have that $F_1(i) \approx -F_s(i)$. Of course, there is a small time differential between the two half-cycles but at 30 Hz, the motion over the ground is unlikely to exceed 1 m which shouldn't change much the geophysical response given that the emitter itself is about 10 m in diameter.

Figure 1. General morphology of a full cycle (simulated data)



We are therefore free to assume that $F_1(i) + F_s(i) \approx 0$ for any given cycle and that the remaining is noise-related or perhaps, software-related.

Clearly, we don't have symmetry. It is clear that the peak we can see in half the half-cycles is not baseline dependent and is a system artifact. However, we still need to verify that the baseline correction satisfy some basic symmetry.

B. Invariance of the SIMn algorithm under sign change and translation

Let's note the raw signal $S(i)$ as a function of sample i. We will note T_{512} the translation of the signal by 512 samples (one half-cycle) so that $T_{512}(S)(\mathrm{i}) = S(\mathrm{i} + 512)$. Finally, let's note B the baseline correction so that $B(S)(i)$ gives you the baseline corrected signal. It can be shown that the SIMn algorithm is invariant under sign change : changing the sign of the data, computing the baseline and then changing the signs again amount to a normal baseline correction so that $-B(S)(i) = B(-S)(i)$ for every i.

In order to show that SIMn doesn't introduce any asymmetry, we need to show that $B(T_{512}(S))(i) = T_{512}(S)(i)$ for every i. It is best checked numerically... What we get considering the first frames of the Bob2 signal is that while the first frame we compare is different, the others are virtually identical. Therefore symmetry is preserved by the SIMs algorithm.

Table. Root Mean Square values of the error in selected frames. Values are in PPM so they have been scaled by $10^6/(\max S - \min S)$.

	$B(T_{512}(S))(i)$ - $T_{512}(S)(i)$
R.M.S. of error in first frame	480 PPM
R.M.S. of error second frame	7 PPM
R.M.S. of error in all other frames	0 PPM

C. Does symmetry matter?

If instead of having $F_1(i) \approx -F_s(i)$ we have $F_1(i) \approx -F_s(i)\, a(i)$ where $a(i)$ is not constant over i, then stacking over a cycle will give us $F_1(i)$ - $F_s(i) \approx (1 + a(i))\, F_s(i)$. Since the factor $(1 + a(i))$ can be assumed to have no geophysical meaning, we are actually loosing or corrupting information. We can then ask whether or not $a(i)$ is significantly different from a constant. It should be noted that as long as $a(i)$ is more or less constant, it won't be a problem, so that, for example, the mean of $a(i)$ is irrelevant. Please note that $a(i)$ is not necessarily purely system dependent function and we expect a(i) to vary over time during a single flight.

One way to estimate $a(i)$ is to first estimate F_1 and then F_2 through stacking or averaging over a large number of cycles (we chose cycles 252 to 277 in the data file). We also scale the values in PPM (using a factor of $10^6/(\max F_k - \min F_k)$ where $k = 1$ or 2 respectively). We got a standard deviation of about 30 for $a(i)$ with a mean value of 0.127 : we would want a standard deviation close to 0 and might have expected a mean value near 1).

Table 1. F1 and F2 as computed by stacking using cycles 252 to 277 (using a conversion factor of 28 PPM/system unit)

	F1 (PPM)	***F2 (PPM)***	***F1/F2 = a***
Mean	-,065	-,005	,127
Standard deviation	1,28E+005	1,33E+005	29,8
Standard deviation of the second half	292	271	22,5

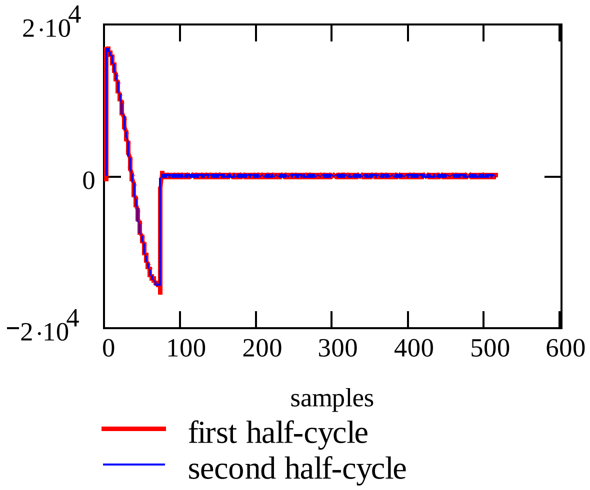
D. Conclusion regarding the baseline correction algorithm

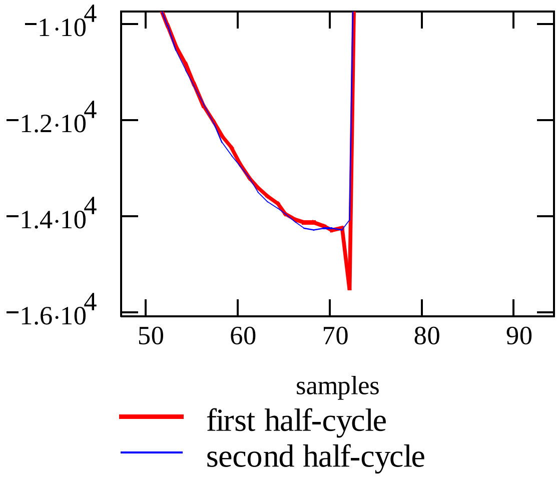
From A, B and C above, we can deduce quite convingly that the baseline algorithm cannot introduce any significant asymetry in the signal such as asymetrical odd and even pulses. This asymetry is caused by the system itself and it is therefore recommended that stacking be done over full cycles when processing the data.

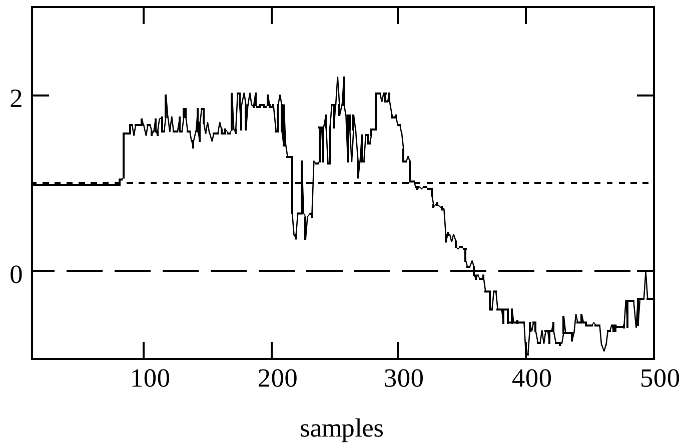
Signal Modeling

We use a model because it is the only approach that allows us to know exactly what the real signal is. In experimental data, one can only guess usually using a model, what the real signal is. Using a model from the start and polluting us allows us to evaluate scientifically our algorithms. It should be noted that we assume from this point forward that the baseline has been properly corrected. Notice in particular that this model doesn't take into account asymmetry between odd and even half-cycles.

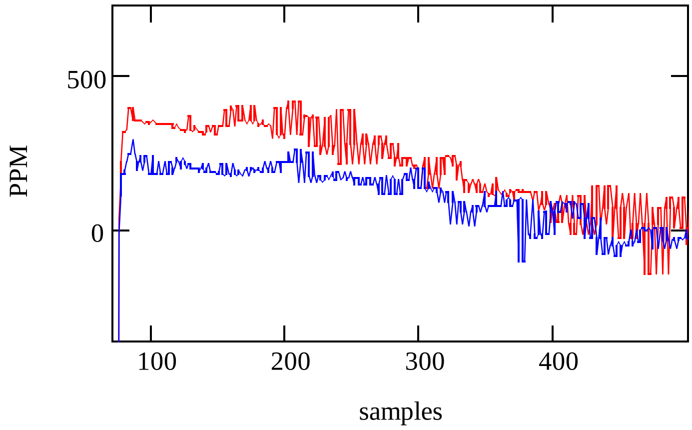
GARBAGE



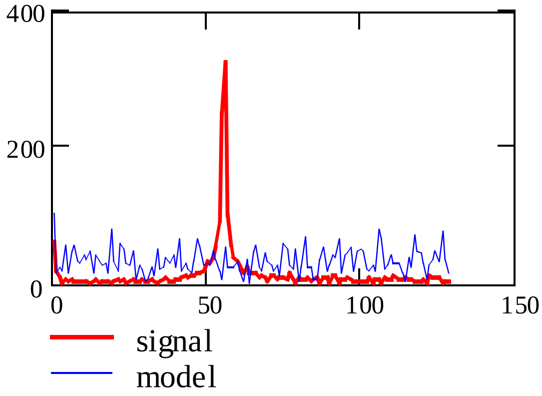




(am avec medsmooth 21)

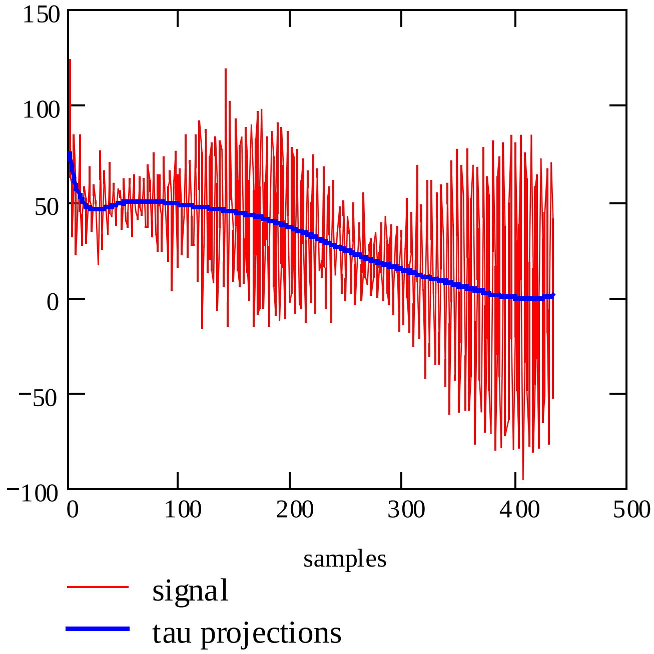


F1(on top) and F2 (below) both converted in PPM (a factor of 28 was used for both F1 and F2 for convertion to PPM)



Comparison between stacked F2 noise during off-time (last 256 samples) and a model based on white noise

Tau projection applied to F1



Conclusion

Stacking half-cycles simply doesn't achieve the noise reduction we need. It is quite possible that the signal is polluted by hardware deficiencies such as bad shielding. Better software approaches are available and needed. More extensive testing is needed (over several files of data) and atmospheric discharges should be processed as well.

The author would like to thank THEM Geophysics Inc. for its support and funding.

References

[1] Daniel Lemire, *Off-time denoising : is stacking the solution?*, Technical Report THEM2001-01, Montreal, January 12th 2000.
[2] Daniel Lemire, *Rapport sur un nouvel algorithme de correction de la ligne de base pour THEM Geophysics Splines, interpolation des moyennes et moindres carrés (SIM)*, Technical Report, Montreal, June 12 th 1999.

Daniel Lemire is a R&D consultant with a Ph.D. in Engineering mathematics. He lives in Montreal and can be found on the web at www.ondelette.com.