Standard deviation calculations/estimation for airborne TEM in Aarhus Workbench

With version 5.8.0 of Aarhus Workbench we changed the default data and data STD calculation.

For SkyTEM xyz data we have long had the option to import the raw data STD values and used those to calculate the total average data STD values. This has not been an option for SkyTEM skb data. Instead we have there had to rely on a total average data STD based on the distribution of the raw data within the averaging trapeze.

With version 5.8.0 of Aarhus Workbench it has become possible, and indeed the new default, to calculate the raw data STD for SkyTEM skb data from the individual shots as they are stacked together to form the raw data. And as a result, we can now also use the total average data STD based on the raw data STD values for SkyTEM skb data.

This gets a bit technical, so let us get a few definitions out of the way.

Definitions:

STD:
Standard deviation.

Raw data:
A stack of single transient decay curves that typically ensures that power line harmonics are stacked out. The stack size is normally dictated by measurement scheme/instrument and the user does not have access to the single transient decay curves.

Average data:
Data soundings after further stacking of the raw data in the processing module define by the user. The average data is the data that enters the inversion scheme.

STD_{Raw_data}:
The STD estimate for the raw data stacks on the individually gates. This STD can be 1) stated in the import xyz-data files for each gate, for airborne TEM data, or 2) calculated during import from the single transient sounding curves that enters the raw data. This is only possible with SkyTEM skb and tTEM skb data, and only applied if the *CalculateRawDataSTD* setting is set to 1 in the geometry file.

$STD_{Uniform}$:
A uniform STD for all raw data gates. This is the minimum STD a gate can have. The **$STD_{Uniform}$** is defined by the *UniformDataSTD* string in the geometry file.

STD_{Gex_gate}:
A gate specific STD that can be defined by an optional *GateSTDs* string in the geometry file. This is rarely used.

STD$_{Raw\_user}$:
STD added by the user on a raw data gate during data processing. Added by selecting some raw sounding curves/gates and then increasing the STD.

STD$_{Average\_user}$:
STD added by the user on an average data gate during data processing. Added by selecting some average sounding curves/gates and then increasing the STD.

STD$_{Average\_stack}$:
STD calculated from the stacking of raw data to average data. This STD is only use if the *CalculateRawDataSTD* setting is set to 0 in the geometry file.

STD calculations/estimates

The total raw data STD for each individual gate (**STD$_{Raw\_total}$**) displayed in the program are calculated as:

$$STD_{Raw\_total} = \sqrt{STD_{Raw\_data}^2 + STD_{Raw\_user}^2 + STD_{Uniform}^2}$$

If the **STD$_{Raw\_data}$** has not been estimated, **STD$_{Raw\_data}$** is set to zero.

The total average data STD for each individual gate (**STD$_{Average\_total}$**) displayed in the program and used in the inversions are calculated as:

$$STD_{Average\_total} = \sqrt{STD_{Average}^2 + STD_{Average\_user}^2 + STD_{Uniform}^2}$$

where the contribution from the averaging (**STD$_{Average}$**) depend on the *CalculateRawDataSTD* setting.

If *CalculateRawDataSTD* = 0 then:

$$STD_{Average} = \sqrt{STD_{Average\_stack}^2 + \frac{1}{N}\sum_i STD_{Raw\_user,i}^2 + STD_{Gex\_gate}^2}$$

If *CalculateRawDataSTD* = 1 and the **STD$_{Raw\_data}$** exist in the database, then:

$$STD_{Average} = \sqrt{\frac{1}{N^2}\sum_i STD_{Raw\_data,i}^2 + \frac{1}{N}\sum_i STD_{Raw\_user,i}^2 + STD_{Gex\_gate}^2}$$

Where the sum is over the number of raw gates N within the averaging width for a given gate time.

Note how the contribution from the STD_{Raw_data} values will decrease with $1/\sqrt{N}$ for a larger data stack. That was not the case before as $STD_{Average\_stack}$ was purely based on the distribution of raw data within the averaging trapeze.

Mean value calculation

With version 5.8.0 of Aarhus Workbench, we changed the calculation of the **Average_data** from a simple average of the **Raw_data** to a weighted average with the STD_{Raw_total}.

If *CalculateRawDataSTD* = 0, then a simple average is used:

$$Average\_data = \frac{1}{N}\sum_i Raw\_data, i$$

If *CalculateRawDataSTD* = 1, then a weighted average is used:

$$Average\_data = \frac{1}{\sum_i (1/STD_{Raw\_total,i})^2}\sum_i (1/STD_{Raw\_total,i})^2 \cdot Raw\_data, i$$

Where the sum again in both cases is over the raw gates within the averaging width for a given gate time.

Examples

As a result of these changes one will observe slightly different average data STD values for SkyTEM data. This is most apparent for SkyTEM skb data near quick data level changes, such as the ones from quick altitude changes. Here the average data STD used to be dominated by the contribution from the STD based on the distribution of the raw data within the averaging trapeze and this resulted in much larger STD values than ideal.

As an example, see figure 1 and 2 below. Both show high moment data as raw in the top and then as average in the bottom. The brown line is the flight altitude. Figure 1 is data with raw data STD values and figure 2 is data without raw data STD values. Notice the difference in STD values near the quick altitude change at 13:57:20 and near the late times noisy gates at 13:58:25.

The average data largely stays the same. The weights will for most of the data be dominated by the contribution from the uniform STD value set during the import. The change from simple averages to

weighted averages will be most apparent in the noisy part of the data where the STD based on the raw data STD values are large, but that is also where we want to suppress the noisiest data. It is possible to add raw user STD to selected individual gates, but as such contributions will dominate the weights and suppress the selected individual gates, it is not recommended to do so. Adding data dependent STD is a version of raw user STD, but it is different as it is added to all gates with a given time, rather than selected individual gates. It is only when the raw user STD changes quickly within the averaging trapeze that it potentially can become problematic.

It is possible to force data to be imported without doing the calculation of the raw data STD for SkyTEM skb data, but it is then not possible to later recalculate the raw data STD values without doing a completely new import. To do so simply set the *CalculateRawDataSTD* value in the geometry file to 0.

A geometry file with the *CalculateRawDataSTD* value set to 0 can on the other hand always be reapplied to change the calculation of the data and data STD values of data already imported with the raw data STD calculation. The raw data STD will then simply be ignored, and the data and data STD will then be calculated as it was done before version 5.8.0 of Aarhus Workbench.
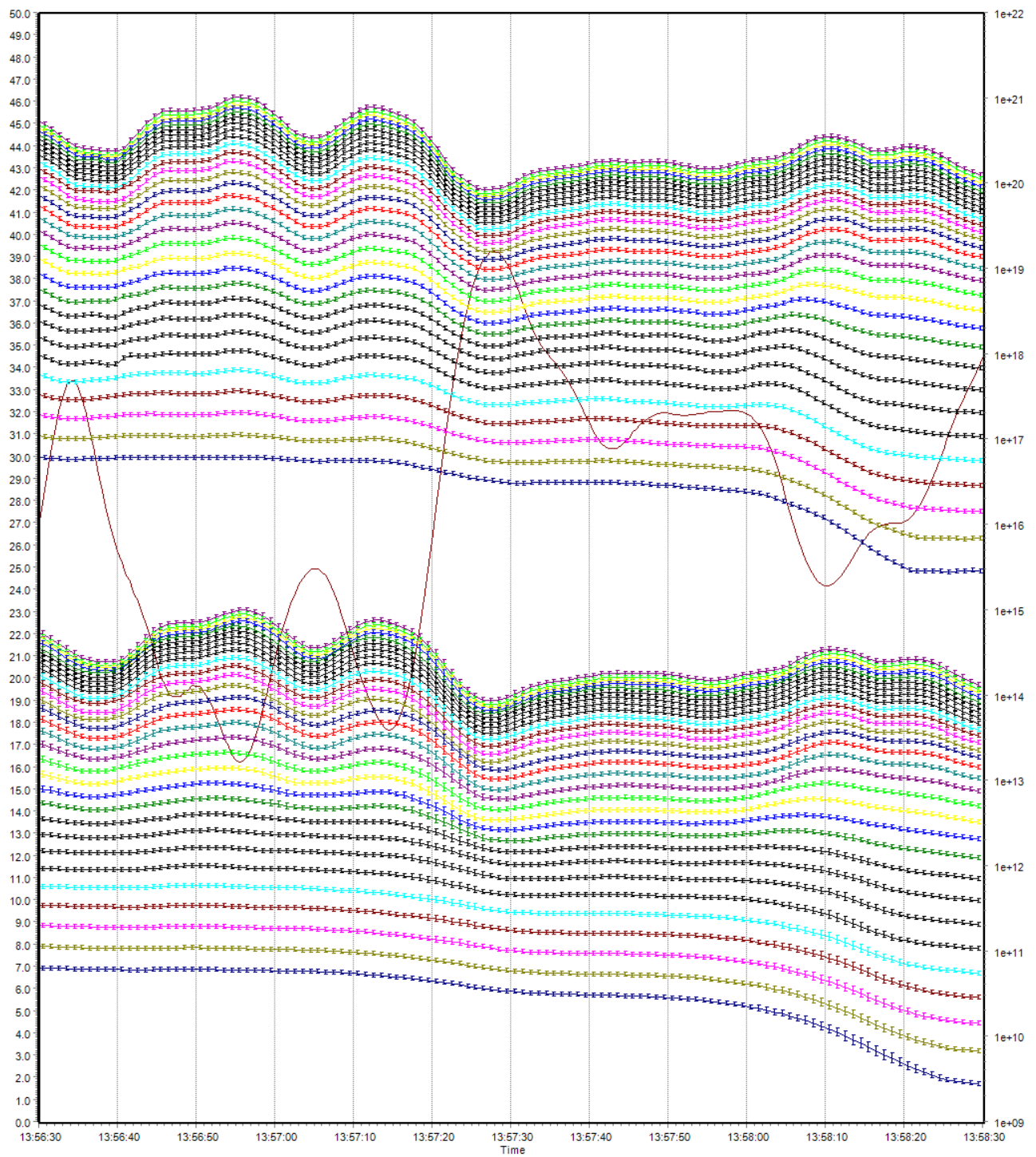
Figure 1. High moment data as raw in the top and then as average in the bottom. The brown line is the flight altitude. Here the data was imported without raw data STD.
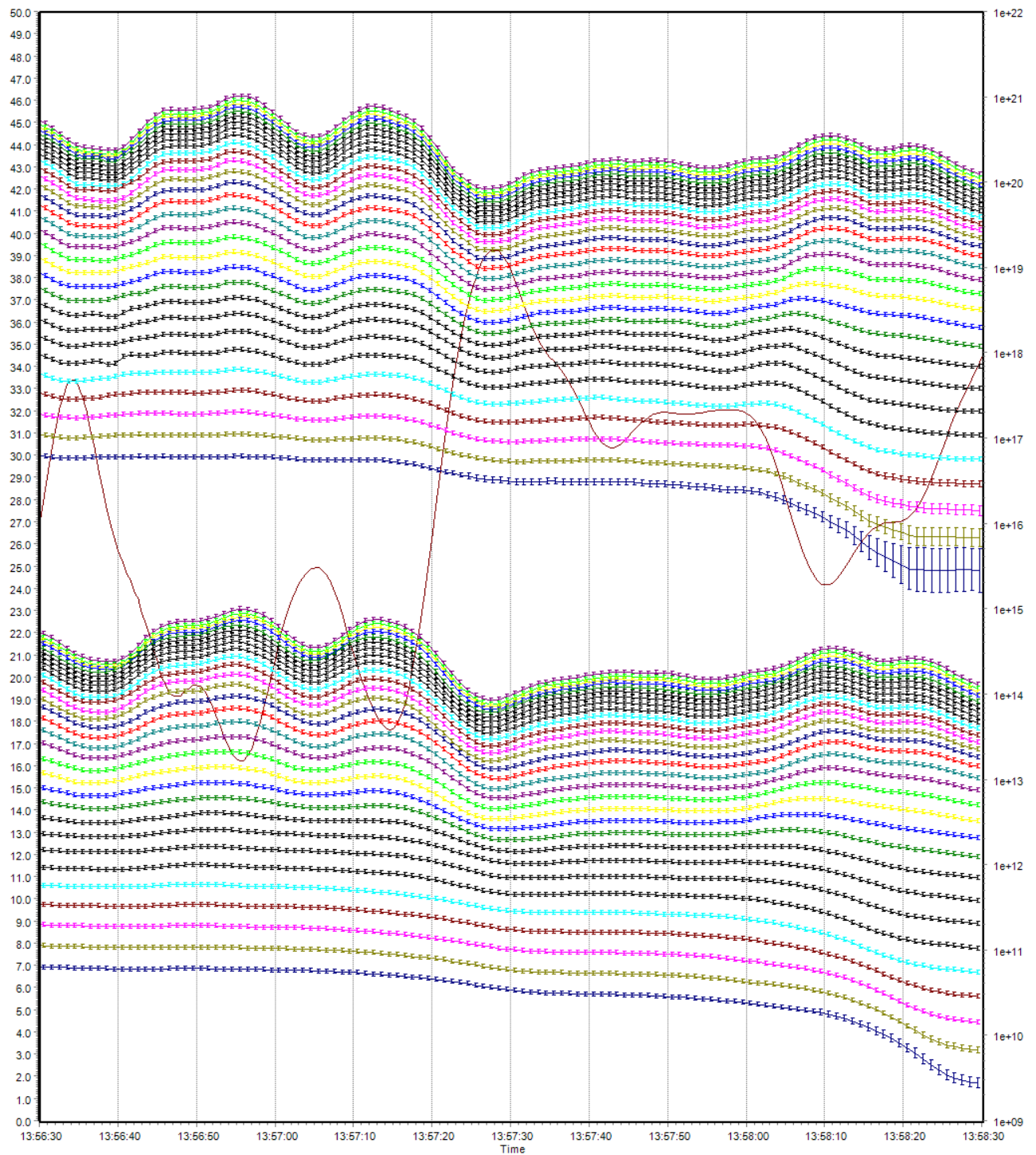
Figure 2. High moment data as raw in the top and then as average in the bottom. The brown line is the flight altitude. Here the data was imported with raw data STD.