

What's new in TSG8, 11 April 2019

TSG build 8.0.4.2

The Summary screen's new Difference plot gets a mini documentation spot here, and the new Scalar Script Builder tool has its own separate documentation. The rest get medium-to-short descriptions below. As always, a log of changes is shown in <https://research.csiro.au/static/tsg/tsglog1619.docx>.

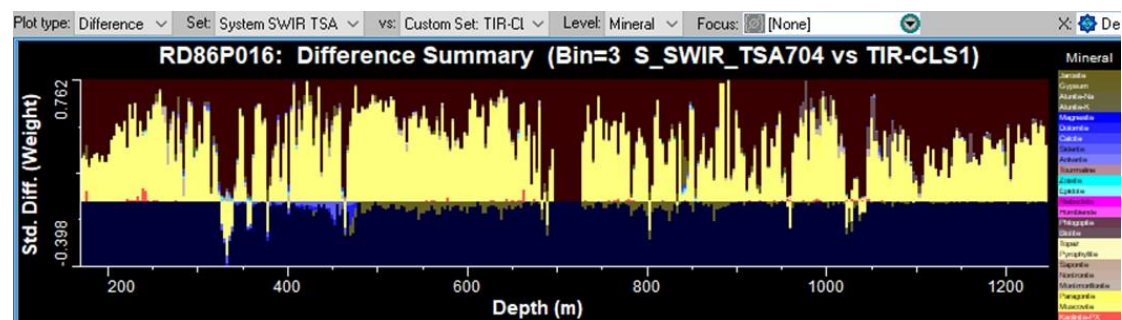
Summary screen difference plot

Reminder: The summary screen is not just for TSA unmixing result sets – it can work with any scalar set that *looks* like a TSA set. This includes CLS unmixing results, external CLS / TSA unmixing results, and user scalar sets¹.

There is now a third kind of Summary-screen plot – ‘**Difference**’. It gives a useful overview of differences between two unmixing result sets.

The Difference plot can be used on any two sets that have something in common. (‘In common’ is simply determined by **item name matching**.) So for example if you plot System SWIR TSAPlus minerals² against the System TIR jCLST minerals then the plot will just deal with the minerals that are found in both, leaving out the TIR-only anhydrous & other minerals and the SWIR-only minerals.

The plot



- The **Plot type** list at top left now has options Overview, Spatial and **Difference**.

¹ Yes you can create a ‘user scalar set’ that *looks* like a TSA unmixing set. You have to provide the mineral, weight, and optionally group and error scalars yourself, but if you have these scalars handy (typically as import-type scalars) then the game is on. Some people use this mechanism to bring externally-calculated unmixing results into TSG. There is more information in the big old ‘what’s new’ document, starting on page 166.

² Things can get a little confusing when comparing minerals versus groups. For example there is a ‘Silica’ group in both SWIR and TIR results, but no ‘Quartz’ mineral in a SWIR result.

- Once you select **Difference**, the on-screen controls are rearranged to make room for a new list called “**Vs.**”.
- Select the first result set from the **Set** list, as usual, and the second from the new **Vs.** list.

The plot’s Y=0 line is somewhere around the middle rather than along the bottom. Any bars plotted above Y=0 show things that are more in the first result set than the second, and any plotted below Y=0 show things that are more in the second than the first. A bar is made up of coloured tiles corresponding to different items (minerals or groups).

More info

In most respects the plot works like a Spatial one.

You get a cross-hair cursor with read-out. As with the Spatial plot, the cursor navigates Floaters with plot-bar granularity and there is no current-sample marker. You can select mineral or group, scope through a focus mask or class scalar, change the X-axis scalar, adjust the X bin size, annotate below the plot with a ribbon, and perform subsetting (but only on weight and error thresholds). If you go into the **right-click menu**, you will find that the **Spatial options** sub-tree is active. You’ll also find that the **Histogram total** sub-tree is active, allowing you to select count or weight totals for the plot.

What is Y?

Ah yes, modern philosophy. This section gets quite introspective – feel free to skip it.

The plot involves binning – *aggregation*. Unless you are working with a dataset that’s small enough to let you drive things with **X = Index** and **bin size = 1**, there will generally be more than one sample contributing to a bin. Also, the number of samples per bin will vary due to ‘**Focus**’ masking, varying X (possibly), and NULL unmixing results. As you will see the bars are standardised for their number of contributors – if they weren’t then the plot would be hard to judge overall. Each bin has a total for each common mineral (or group). These totals are the coloured tiles³ that you see in a bin’s bar. Also, TSG has already worked out the items (minerals or groups) that are common to both result sets. These are the ‘active items’; all other items are ignored.

So then.

- Say Histogram total is set to **Weight** and to keep it simple we have turned off weight and error Subsetting⁴. Say we’re looking at a bin for which there should be 10 contributing samples, but 3 have been knocked out by the focus

³ Normally a single bar will have just a few different tiles as you don’t usually get all minerals showing up everywhere through a dataset.

⁴ Error ‘Subsetting’ is for filtering out high-error samples, so that only the more reliable results contribute to the plot. Weight ‘Subsetting’ is for discarding small bar totals to tidy up the plot, and is commonly used in Spatial plots. Think twice before using either of these (especially weight) in a Difference plot.

mask. So there are (at most) 7 active samples for this bin. For each of these 7 samples:

- a. We work with the first result set, going through the mixture level⁵. If we have an 'active' item then we take the corresponding weight and add it to the item's total for the bin.
 - b. We do the same thing with the second result set, only the loop is over *its* mixture level and the weights are subtracted rather than added.
 - c. If the **Weight by recovery rate** menu option is turned on then all the weights in a and b are scaled by the sample's recovery rate⁶.
 - d. We build up a real sample count for the bin. If *either*⁷ result set has at least one active item then we add 1 to the count. (If *both* result sets have nothing but NULLS then we leave the count alone. Such a sample is effectively a dud.)
- We divide each of the bin's item totals by the real sample count. This standardises the bin to a sample count of 1.
 - Non-zero (plus or minus) item totals get a showing in the bin's histogram bar. Positive items have their tiles shown above Y=0 and negative items below. The biggest possible bar, given well-behaved⁸ weights:
 - i. All items in the first result set were active and none of them were in the second result set. This gives the bar a maximum of 1.
 - ii. Also, all items in the second result set were active and none of them were in the first result set. This gives the bar a minimum of -1.

Count or weight?

If Histogram total is set to **Count** rather than **Weight** then uniform weights are made up. Each unmixing result at each mixture level gets a weight of 1 (optionally scaled by recovery rate). This gives a plot of a different flavour. An active item in a result set counts the same, as its (original) unmixing weight is now ignored. E.g., if a sample's result set A has muscovite at 1.0 and result set B has it at 0.15 (or even 0.01 say) then the calculation says: 'both result sets have muscovite so it cancels out'. There is a strong assumption in play when using **Count**: *you have done your level best in preparing the unmixing results*. All of the unmixing calculations in TSG have a 'minimum weight' setting⁹ that you can and *should* use to stop the algorithm from reporting tiny phases that it honestly can't judge properly.

⁵ e.g., Maybe this is a mix3 result set and so we are looping by 3.

⁶ You can calculate a 'recovery rate' scalar for a HyLogging-compatible dataset. For such a dataset, recovery rate is [depth along hole] / [distance spanned in the measurement space, i.e., in the core tray]. It often varies from 1.0 in real HyLogging datasets. Here, recovery rate is used to make the binning work as if the measurement space were depth-along-hole. This option is only available when you have Depth driving the X axis.

⁷ If one result set is all NULL for a sample while the other isn't then this is definitely something that we'd like the difference plot to show.

⁸ The weights in TSG's TSA and CLS result sets are 'well behaved': the weights for a single sample's results always add up to 1 (or are all NULL).

⁹ The default minimum weight in SWIR TSA is 15% and you can get at TSA settings via **File -> Settings [TSA]**. When creating a CLS or external-TSA result set, the scalar settings page includes a minimum weight field.

Deeper down the rabbit hole

To finish off I'll go through a few examples to clarify the plot's workings. Before we take the plunge, remember that a TSA or CLS unmixing result is in a world of its own. If it gives you a non-NULL result then the result's unmixing weights will add to 1. The weights are *relative*, not absolute abundance. For example a SWIR result may say 0.53 muscovite + 0.37 chlorite even though the sample may actually contain 0.98 (98%) quartz. Some minerals (like quartz) are basically invisible to the SWIR.

1. In the first instance, say we are dealing with two SWIR TSA mineral result sets, one a mix2 set and the other a mix3. This is nice – the result sets have the same mineral classes so no classes are 'inactive' here.
Say we are looking at a sample for which set A has [muscovite:0.57, kaolinite:0.33] and B has [muscovite:0.57, kaolinite:0.18, chlorite-fe:0.15]. In the plot's calculations:
 - The muscovites have the same weights (0.57) and so 'cancel out'. They are not different - no difference contribution.
 - There is more kaolinite in A than B. The bin's kaolinite total gets an increase of +0.15.
 - There is more chlorite-fe in B than A. (A actually has none.) The bin's chlorite-fe total gets a decrease of -0.15.
2. Same situation but we are looking at the next sample and it contributes to the same bin. Set A has [muscovite:0.51, chlorite-fe:0.49] and B has [muscovite:0.51, chlorite-fe:0.34, kaolinite:0.15]
 - The muscovites cancel out again.
 - There is more kaolinite in B than A. The bin's kaolinite total gets a decrease of -0.15.
 - There is more chlorite-fe in A than B. The bin's chlorite-fe total gets an increase of +0.15.

After these two samples we are left with no differences. Although the result sets *are* different, they cancel one another out in the binning process. If this trend continued then the plot would report no differences. I would argue that this is a meaningful outcome in a spatially continuous dataset.

3. Now say we are looking at a SWIR TSA mix3 mineral result-set versus a TIR jCLST mix4 one. The plot only considers minerals that are common to both sets – all others are invisible. So we have a sample for which set A has [muscovite:0.57, kaolinite:0.18, chlorite-fe:0.15]. B has [quartz:0.60, muscovite:0.20, montmorillonite:0.15, kaolinite:0.05] but quartz is invisible here so B's set is [muscovite:0.20, montmorillonite:0.15, kaolinite:0.05].
 - The bin's muscovite total gets an increase of $0.57 - 0.20 = +0.37$.
 - The bin's kaolinite total gets an increase of $0.18 - 0.05 = +0.13$
 - The bin's chlorite-fe total gets an increase of $0.15 - 0 = +0.15$
 - The bin's montmorillonite total gets a decrease of $0 - 0.15 = -0.15$.
 You can see that TSG did not readjust the weights for set B after removing the 0.60 quartz (which is not active in the plot). Not fair? The author's

current view is that it *is* fair. The TIR 'says' to the SWIR: "Hey smarty-pants! There's more to the world than hydrous minerals you know." The plot will convey such findings. The meaning of the difference plot is more complicated with heterogeneous result sets, but maybe more interesting.

4. To end this section on a... well a note of some kind or other, say we're looking at the same sample as 3 but have histogram total set to **Counts** rather than Weights. So set A has [muscovite, kaolinite, chlorite-fe] and set B has [muscovite, montmorillonite, kaolinite].
 - Muscovite cancels out. There is no difference.
 - Kaolinite cancels out. There is no difference.
 - Set A has chlorite-fe but set B doesn't. The bin's chlorite-fe total gets an increase of 1.
 - Set B has montmorillonite but set A doesn't. The bin's montmorillonite total gets a decrease of -1.

Scalar script builder

You can now get scalar batch scripts without writing them, or even knowing TSG's scalar script language. This release brings a new tool that generates a script from a scalar in your dataset. If you give the tool a scalar (for example an Arithmetic Expression scalar) that depends on various other scalars then it will chase down the dependencies and include them in the script.

The tool is launched from the **Edit -> Scalar script builder** menu.

The documentation can be downloaded from here:

https://research.csiro.au/static/tsg/tsg_scalar_script_builder.pdf

Scalar sets

TSG has offered 'scalar set' functionality for some years now. A scalar set is a way of organising a collection of scalars that form something like an unmixing result. TSA unmixing results are held in system-controlled scalar sets. CLS unmixing results built on system libraries and CLS / TSA results on external libraries are held in less-formal scalar sets. You can also put together your own 'user scalar sets'.

Copying

Edit -> Scalar sets -> Copy a scalar set

You can now spin off a copy of an existing scalar set, e.g., a TSA unmixing result. You get a new **user** scalar set containing copies of most¹⁰ of the original scalars. The copied scalars are import-type scalars.

The copied set can be used in TSG where sets are supported, e.g., plotted in the Summary screen and used for scoping a Scatter sub-screen. The Mineral and Group scalars can be modified by one of TSG's interactive class editors.

Select a scalar set to copy (its scalars are copied too)

DS1	CLS1	Rank=3	Scalars=7
DS1	U_VNIR_TSA704	Rank=2	Scalars=9
DS1	S_VNIR_TSA704	Rank=2	Scalars=9
DS1	S_SWIR_TSA704	Rank=2	Scalars=11
DS1	U_SWIR_TSA705	Rank=2	Scalars=13
DS2	U_TIR_TSA703	Rank=3	Scalars=12
DS2	S_TIR_TSA703	Rank=3	Scalars=12

Call it:

Reasons you might want to make a copy of a scalar set could include:

- To *thoroughly* prevent a TSA result set from being recalculated;
- To preserve a 'before' version of a TSA result set that can be compared against a recalculated result (e.g., using the Summary screen's Difference plot);
- To hand-edit individual unmixing results using one of TSG's class editors;
- Soon to come: to reassign individual samples' unmixing results using a TSA Floater in 'multiple result' sub-mode.

Aside from a set selection list, the dialog (above) has a field for the name of the copied set. You can provide your own name or just go with the default. The name can be 31 characters long.

Deleting

Edit -> Scalar sets -> Delete a scalar set

In the past this dialog just let you delete a set definition (*user sets only*). Afterwards, you had to go and delete the set's scalars if you had had enough of them too. The dialog now has a new button **Delete set and scalars** that will delete the set definition *and* its scalars on one go.

¹⁰ Mineral, Group, Weight and Error scalars are copied. Specialist scalars like 'Unbound water sTAS' are not copied.

Floater feature frequency chart

I've been *waiting* for this. I can now call this option 'new and improved' without expecting too much ridicule. It is mostly *quite* new after all. As regards the 'improved' part, here are some dots:

- The control panel is modeless. You can do other things in TSG while it is shown.
- The Scatter and Grid plots have current-sample markers.
- The scatter and Grid plots can be rotated by 90 degrees. This is useful if you have (drillhole) depth on 'X' and want it to go down the vertical axis. (Yes Depth goes down rather than up the vertical axis.)
 - Also, the floater's **Vertical lines** can be displayed on a rotated plot.
- The Depth feature filters work on absolute values. (This was done mostly for 'MinMax' features – see below.) For Scatter and grid plots the **Max** depth is not used for feature filtering anymore but (scatter plots) is instead used for colour clipping when colouring by feature depth.
- Scatter plot colouring has been opened up. You can now colour the plot by feature depth (the only option in the past), order¹¹, or an auxiliary scalar.
- The X and Y range-subsetting fields are now complemented by scrollbars. If you subset the X (or Y) range then a scrollbar appears under the range fields, allowing you to move your sub-range 'window' around.
- Handling of the 'MinMax'¹² feature-extraction algorithm's results is richer.
 - You get two new checkboxes **Troughs** and **Peaks**, allowing you to have either or both active in the plot.
 - If both are active in a scatterplot that's coloured by depth then a 'dead centre' rainbow LUT is used for colouring (to emphasise the strong peaks *and* troughs).

Aux attach

Most of the work has been 'under the hood' to make the handling of external unmixing¹³ more consistent. A small visible change is that the **File -> Settings [Aux]** dialog is brought up automatically after an aux attach, mainly to bring aux wavelength subsetting to your attention. Users with dataset pairs (e.g., paired VNIR_SWIR + TIR datasets) will find *two* lots of aux settings here – one for each dataset.

¹¹ In the feature-extraction calculation (which is normally done on import), TSG currently calculates features one at a time. The first one is normally the deepest, the second one the second-deepest, and so on. Feature order is generally a kind of 'non parametric depth'. Only a handful of spectra may have a feature as deep as 0.8, say, but every spectrum will have a 'feature #1'. So it can be interesting to colour the plot by feature order.

¹² If you go into **File -> Settings [FeatEx]**, you will find that TSG offers a number of feature-extraction algorithms. One of them is called '**MinMax**'. It doesn't return widths and its depths are loosely defined, but it does both peaks and troughs at the same time. Peaks get positive 'depths' and troughs negative. As it does both, it is a popular choice for TIR spectra.

¹³ External TSA or CLS unmixing is done using a suitable aux-attached library.

Miscellaneous

- **Testrocks** dataset identification has been improved.
- The '**PUCKWCAL**' wavelength-calibration module has been made robust to missed features, and its (automatic) integration in regular and QC TestRocks imports has been tidied up.
- Tray screen **Rockmarks** editor: There's an option to **export to CSV**. Along with Rockmarks tag-type and description, this CSV includes columns for Hole, Tray, Section, SecDist, depth.
- **Encrypted datasets**: Up to 128 dataset passwords are saved. A high-restriction encrypted dataset can now be used for external unmixing (it couldn't in the past).
- **Checklist**: There's an option to delete the journal.
- The **ASCII spectral import** now works according to the '**Euro CSV**' setting in **File -> Settings [System]**. (It's for when you have a **comma instead of a dot** in floating-comma numbers.)
This handling may be done more automatically (according to your computer's regional settings) in future, but given the global nature of our line of work an interactive option will remain.
- **Summary screen**: There's a new option in **File -> Settings [System]** to have raw bin totals instead of percent in the exported CSVs. Also, full TSA mineral names are now shown on the Overview plot's X axis if the '**Use TSA abbreviations**' *colourbar* setting is turned off.
- **File -> Save as** now **copies layouts too**.
- **Corescan import**: There's a basic QC filter on input spectra. A spectrum must have a reflectance range within [0,1] to be accepted.
- **SDS import**: There's '**de-step**' option in the SDS import's detailed settings page (top left), and (QC users take note) there's a similar option in the 'Defaults for SDF and SDS' dialog. When enabled, TSG does a VNIR – SWIR step correction as best it can, similar to the one that's normally done by the HyLogger operating software. *It should only be enabled for measurements that have been made with the HyLogger's step correction turned off. It is here to assist in diagnosing potential issues around the VNIR – SWIR join.*