Corsi, Steven <srcorsi@usgs.gov>

The data

13 messages

Corsi, Steven <srcorsi@usgs.gov> Wed, Feb 19, 2014 at 11:45 AM
To: Brian Pellerin <bpeller@usgs.gov>

Hi Brian

A couple questions on the data from the January download:

1. There are many negative values in the Fluorescence data (have a look at the peak B data in the summary variables). How do you folks deal with this?

2. The files that you sent in January include the vectorized data from July 2013 to Jan 2014. I think there was some corrections needed to the vectorized data previous to that download that I have not received yet for the GLRI data. True? The summary data looks to have the corrected values as there are graphs of comparisons with the re-processed data. Maybe I just didn't get the new version of the vectorized data. Or could I be missing a file that you sent on this somewhere?

Thanks
Steve

--
==========
Steven R. Corsi Phone: (608) 821-3835
Research Hydrologist email: srcorsi@usgs.gov
U.S. Geological Survey
Wisconsin Water Science Center
8505 Research Way
Middleton, WI 53562
==========

Pellerin, Brian <bpeller@usgs.gov> Wed, Feb 19, 2014 at 4:29 PM
To: "Corsi, Steven" <srcorsi@usgs.gov>
Cc: Jacob Fleck <jafleck@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Hi Steve. I'm cc'ing Angela and Jacob since they can correct me if I'm wrong. As for negatives, I think we'd probably need to treat these as any other BDL sample. (They can be a bit negative at those ex/em's because of the very low signals and corrections for blanks, etc.). I think you could either treat these as zeros or some fraction of the BDL (like half?). I recall Jacob having an opinion on this.

As for the second question, was it related to this earlier email string (pasted below)? I'm guessing not, since that was about absorbance data - but Angela, can you confirm the corrections were made? Steve, can you confirm that you're talking about EEMs data and I'll try to dig up the issue we were dealing with? Thanks.

Brian

On Mon, Jan 13, 2014 at 1:22 PM, Corsi, Steven <srcorsi@usgs.gov> wrote:

> Hi Brian
>
> Somehow this email from you got buried and I just found it.... Sorry for the delay, but if you folks can make the correction and resend, that would great.
>
> On another note, I am presenting some of the results from storm sewer sampling the week after next and would like to include the optical data for those sampling events. Do you know when those will be ready to send off to me? I will need

> the full data sets so I can run a few contour graphs for presentation.
>
> Happy New Year!
>
> Steve

[Quoted text hidden]

--
Brian A. Pellerin
USGS, CA Water Science Center
6000 J Street, Placer Hall
Sacramento, CA 95819

bpeller@usgs.gov
(916) 278-3167

[Quoted text hidden]

--
Brian A. Pellerin
USGS, CA Water Science Center
6000 J Street, Placer Hall
Sacramento, CA 95819

bpeller@usgs.gov
(916) 278-3167

Fleck, Jacob <jafleck@usgs.gov> Wed, Feb 19, 2014 at 4:55 PM
To: "Pellerin, Brian" <bpeller@usgs.gov>
Cc: "Corsi, Steven" <srcorsi@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Hi all,

yes, I am opinionated on this topic :)

regarding interpretation, yes, it is a "<" result. handle it as any other "<". Helsel has numerous opinions on this matter.

I perform a post-processing censorship of data for my projects. I use 2 approaches depending on the number of field blanks a project has. I like to follow the >7 rule.

1) If there are sufficient blanks run, I create a minimum reporting limit (MRL) for the project using the blanks' vectorized data (avg+3xstdev)

2) if there are insufficient blanks to create a robust MRL for the project, then I use the MDL or LRL (2xMDL) from the lab. if this has not been supplied to you, we can send it on to you. just let us know.

i then replace the "<" data with whatever approach you prefer (MRL, 1/2 MRL, "0", "<"...)

I also like to pull the summary data from the vectorized data again myself and cross-check to make sure things agree. there was a period when the B peak was double blank corrected that resulted in lots of negatives - but only in the summary table, not the vectorized data. It should have been corrected on the files sent out but you should confirm this.

we are working towards supplying a more "finished" results package but I have kept most of this type of data handling on the project-level at this time.

cheers,
Jacob

Jacob Fleck
Research Hydrologist
USGS CAWSC
6000 J st, Placer Hall
Sacramento, CA 95819
jafleck@usgs.gov
916-278-3063

[Quoted text hidden]

Corsi, Steven <srcorsi@usgs.gov> Wed, Feb 19, 2014 at 4:56 PM
To: "Pellerin, Brian" <bpeller@usgs.gov>
Cc: Jacob Fleck <jafleck@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

OK, you are correct that it was the abs and not the EEMs that I was referring to.

So the other thing I found in our email correspondence was this info on negative value treatment:

"Hi Steve. Good question. I just spoke to Angela and Jacob about it, and they have a vectorized file that is 3x the standard deviation of the long-term instrument blanks. You could use this as a MDL for the full EEMs, below which you can decide how to handle the data (e.g. call it a zero, 1/2 the MDL, ND, etc.). I'm not sure what you should do, although it seems like lots of folks avoid calling numbers below the MDL "zero".

Jacob also mentioned that you could use the field blanks that you sent us and generate a similar average EEMs dataset that might be a more accurate MDL (since our blanks are an instrument blank only). You should have a fair number of blanks if you choose to do that."

And this from Brian B:
"I typically adjust the baseline so there are no zero values. If you see large zero values, there is a problem as zeros arise when the water blank matrix is subtracted from the sample matrix."

So my question really is: How have you done it for your data analysis and published work in the past? I would like to be consistent with historical methods. The field blanks also have negative values, so won't really serve as a good MDL. Maybe the best solution is to use the lab-generated values as Jacob mentions above.

OK, I think there is a chance of loosing my sanity during this process due to data overload...

Thanks
Steve
[Quoted text hidden]

Corsi, Steven <srcorsi@usgs.gov> Wed, Feb 19, 2014 at 4:57 PM
To: "Fleck, Jacob" <jafleck@usgs.gov>
Cc: "Pellerin, Brian" <bpeller@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Thanks Jacob

Our emails crossed in the cyber. I will have a look at your suggestions now.

Steve
[Quoted text hidden]

Corsi, Steven <srcorsi@usgs.gov> Wed, Feb 19, 2014 at 5:24 PM
To: "Fleck, Jacob" <jafleck@usgs.gov>
Cc: "Pellerin, Brian" <bpeller@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Thanks for all of the input on this. We can try option #1 to correct for blanks from the project level. I will assume that we can use blanks from multiple sites together and then we will have plenty for this. So, just to be clear, you are creating an MRL for each individual EEMs pair generating ~6400 MRLs, correct? I will probably assign them 1/2 the MRL in the final data to allow the option of log transformations in the data analysis.

... and then the same process for abs...

Also to be clear--the vectorized versions that you are sending have not been lab-blank-corrected at all before I get them, correct?

Thanks
Steve

On Wed, Feb 19, 2014 at 4:55 PM, Fleck, Jacob <jafleck@usgs.gov> wrote:
[Quoted text hidden]
[Quoted text hidden]

Hansen, Angela <anhansen@usgs.gov> Wed, Feb 19, 2014 at 6:52 PM
To: "Corsi, Steven" <srcorsi@usgs.gov>
Cc: "Pellerin, Brian" <bpeller@usgs.gov>, Jacob Fleck <jafleck@usgs.gov>

Hi Steve,

In response to your absorbance question... I looked into this and found that approximately 25 grNums did not have the scattering correction applied to the Full Absorbance scan. Our processing routine failed to apply the correction to samples that required a dilution in the lab. I can provide the corrected scans for you tomorrow.

Thanks,
Angela
[Quoted text hidden]
--
****
Angela Hansen
U.S. Geological Survey
California Water Science Center
6000 J Street
Sacramento, CA 95819-6129

Lab: 916.278.3204
Cell: 916.715.0225

Fleck, Jacob <jafleck@usgs.gov> Wed, Feb 19, 2014 at 7:48 PM
To: "Corsi, Steven" <srcorsi@usgs.gov>
Cc: "Pellerin, Brian" <bpeller@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Hi Steve,

you are correct, I apply the censorship to the entire vectorized data pairs. same for abs wavelengths.

the vectorized data are actually already instrument blank corrected as part of the processing used to raman normalize and remove the rayleigh peaks. this correction is based on the instrument performance for the day. the lab MDLs that are generated are based on those instrument blank corrected data to address the variability across different instrument runs, cuvette handling, etc. Because these data are blank corrected, we evaluate the lab MDL as 3xstdev only as the avg should be very close to 0 based on this correction. When evaluating the field/method blanks, i include the average to account for the other sources of contamination and establish the MRL as avg+3xstdev to be conservative about what data are really above all the noise from collection through analysis.

best,
Jacob

Jacob Fleck
Research Hydrologist
USGS CAWSC
6000 J st, Placer Hall
Sacramento, CA 95819
jafleck@usgs.gov
916-278-3063

[Quoted text hidden]

Corsi, Steven <srcorsi@usgs.gov> Thu, Feb 20, 2014 at 9:14 AM
To: "Fleck, Jacob" <jafleck@usgs.gov>

Cc: "Pellerin, Brian" <bpeller@usgs.gov>, Angela Hansen <anhansen@usgs.gov>

Great--that all makes logical sense. Thanks for all of the input. Much appreciated!

Steve
[Quoted text hidden]

Corsi, Steven <srcorsi@usgs.gov> Thu, Feb 20, 2014 at 9:21 AM
To: Michael Tate <mttate@usgs.gov>, Peter Lenaker <plenaker@usgs.gov>

Mike

I have been in contact with CA about how to work with the EEMs and abs data. There are negative values that show up occasionally, and it is likely just analytical noise at the low end even after daily blank corrections. He suggests using the field blanks to make a final MRL to apply. This is something we should probably include in the python scripts as a normal routine. First, to generate the MRLs from the field blanks, and second, to apply them to the results. It would make most sense to do this on the 3-D array formatted EEMS, but could be done directly in 2-D array format for the abs data. Do you already have something similar or is this a whole new script to generate?

We should be able to generate the 3-D array from California WSC to use in your python scripts, so regardless of where the data is coming from, the scripts can be used.

We can talk more about details if needed also. I will be in this afternoon and tomorrow if you need some clarification.

Thanks
Steve
[Quoted text hidden]

Michael Tate <mttate@usgs.gov> Thu, Feb 20, 2014 at 9:46 AM
To: "Corsi, Steven" <srcorsi@usgs.gov>
Cc: Peter Lenaker <plenaker@usgs.gov>

Steve,

I have already written some scripts that deal with analytical blanks so I think that I can easily integrate something like this into the scripts. I have a running list of modifications and additions to the scripts that I would like to add.

I have cursorily looked at the scripts that the CA folks sent out. I think that Pete and I have basically arrived at a similar data processing method as them.

On my list of things to do is to write a script that will apply QA flags to data based on analytical blanks (and possibly field blanks). It might be most appropriate to apply QA based on field blanks at the data analysis stage.

Here is my pecking order for feature to add to the scripts:
1. add a few basic statical outputs for assessing tea standard (% difference and relative coefficient of variance - depending on N tea samples) - needs a little more work
2. make text output a bit more flexible as output options are added to the EEMs and Abs library - basically finished
3. assessing duplicate analysis - almost done with this one
4. plotting linear and diagonal slices of EEM at regular excitation wavelength intervals (just getting started on this one)
5. peak hunting - have an idea on how to do this. I recently saw a thread in a user group where the researchers were using peak searching to identify voltage maxima across a 3d field. I would like to try using their methods.

I will stop by sometime this afternoon. Kids school is delayed so I won't be in until about 1030. It would be good to catchup to make sure we are all on the same page.

Mike

Michael Tate
US Geological Survey
8505 Research Way
Middleton, WI 53562

Office: 608-821-3830
Cell: 608-669-4202
mttate@usgs.gov

[Quoted text hidden]

Hansen, Angela <anhansen@usgs.gov> Thu, Feb 20, 2014 at 2:24 PM
To: "Corsi, Steven" <srcorsi@usgs.gov>
Cc: "Fleck, Jacob" <jafleck@usgs.gov>, "Pellerin, Brian" <bpeller@usgs.gov>

Hi Steve,

Please find attached your corrected absorbance scans. In total, 28 samples were affected. Please let me know if you need anything else.

Thank you,
Angela
[Quoted text hidden]

--
****
Angela Hansen
U.S. Geological Survey
California Water Science Center
6000 J Street
Sacramento, CA 95819-6129

Lab: 916.278.3204
Cell: 916.715.0225

AbsFullScans_Corr_022014.xlsx
226K

Corsi, Steven <srcorsi@usgs.gov> Mon, Feb 24, 2014 at 8:25 AM
To: "Hansen, Angela" <anhansen@usgs.gov>

Hi Angela

Thanks for the help on this!

Best
Steve
[Quoted text hidden]