

Spring 2024 DMAG Meeting Minutes

Attendance:

Casey Lanham, Amy Mager, Richard Tanabe, Jayde Alderman, Chris Rogers, Greg Beachley, Amanda Cole, Nichole Miller, Katy Blades, Aaron Pina, Tim Sharac, Dana Grabowski, Mark Kuether, Zac Najacht, Jean Steele

WELCOME

We introduced Jean Steele, an APhL fellow who is helping build automation to increase efficiency. She has been learning how to do precipitation review, NTN review, and helping us with some of our new processes.

DATA REVIEW AND REPORTING

Zac started with a review of the current network status (see agenda for details). Mark provided an update of the posted data, as well as his continued efforts to complete the Data Manager SOPs.

DATA REVIEW UPDATES

Zac then discussed updates to the data review process. Last fall, based on an audit recommendation, NADP moved some of its spreadsheet based processes into the Data Review application. This included automating reviewer flags, qualifiers, and comments instead of manually looking them up, copying and pasting into the application.

One of the efficiency issues is related to the number of samples to review. Zac has traditionally needed to review about 1000 samples every month. The current improvement process for data review continues efforts to change the serial nature of the process into a more parallel process. Additionally, efforts for reducing the number of samples reviewed are also being made. These efforts involve changing roles and tasks within the sample receiving and data teams.

The latest update involves reviewing the specific aspects that may possibly change a sample's QR code or induce a notes code. The receiving aspects are the same. These involve checking and logging in the sample, as well as double data entry with compare reporting. Once the original check-in occurs, the sample is processed in a batch process using these steps:

- Prepare monthly dataset for review by getting sampling dates/sample numbers, loading chemistry data, applying SL (screening level) and running history.
- Identify and correct date/time overlaps and gaps (> 60 minutes) in the data set.
- Review analytical data for missing chemistry - address any missing.
- Check for missing precip and reload or correct precip data as needed.
- Identify and invalidate bulk or undefined samples.

- Check for and confirm coding for sample qualifiers based on lab comments (receiving or analytical staff).
- Identify remaining issues based on field comments.

Jean has worked through this process for the August data set, and reports she had to review 375 out of ~ 1000 samples to complete the review. Jean has also been developing SQL queries and Python scripts to automate some of these steps further.

Amanda Cole asked about how we will monitor the effectiveness of this process change, and what the breakdown of the A's, B's, and C's were. Mark and Jean will perform final review on two or three months of samples together as a cross check. We can also check monthly and historic value of the relative frequency of the A's B's and C's, as well as the valcodes, invalcodes, and sample type classifications. If we see a change, it can investigate further to determine what was missed.

Chris Rogers asked about the breakdowns on the A's, B's, and C's. Zac felt ballpark numbers were about 25%, 50%, and 25%.

(5 Year historic A: 24%, B: 58%, C: 18%)

(Calendar Year 2023 up to AUGUST A: 18%, B: 61%, C: 21%)

DATA REVIEW FUTURE PLANS

For future plans of data processing, Mark will be investigating the feasibility of using Google Document AI, with OCR technology, to read the hand written FORFS from NTN. The goal is to determine if we can use that as one of our initial data entry steps. The initial step is testing out the mechanical aspects of the process. If that pans out, then we can plan how this will be integrated into the workflow.

Amy raised the question of whether anyone else had worked with Google Document AI technology before. If anyone has any experience, please contact Mark to help on this goal.

Prior future plans also involved performing statistical analysis to identify samples that did not need to be reviewed. With the above changes, we put this effort on hold. We can revisit it later to see if it would provide additional benefits.

MLN UPDATE

Zac indicated that the team is still having conversations on the types of note codes, the quality ratings, and how to apply those for MLN samples. The 2022 data set has been completed and reported out, but does not have notes codes or quality ratings (QR) at this time. We hope to be able to apply those to the 2023 data set.

AMON UPDATES

Amon continues to be dynamic, but the team still intends to shift this to a static table. This is needed so final checks, like other networks, can be easily performed. The block for this has been Mark's understanding of the network and the overall process. When trying to document a formal process, his confusion became evident. Mark and Zac will iron out the process together. After they do this, Mark should be able to shift the data to a static table and process it.

Amanda Cole wanted the group to know that her group has been studying how measurement corrections made by taking pressure and aperture temperatures into account impact the overall AMON measurements. They have studied Canadian sites so far, but may start studying US sites as well. The group felt this would be an ideal topic to discuss during the Fall meeting.

WEB UPDATES

Mark has published the formal Data Request form. If you need Mark to generate any type of report, please use this form so we can track the frequency and types of requests that we are receiving.

Casey, Mark, and Richard will be meeting with the web developers (UW Madison IT department) next week for a status update. They are working on our test web site, correcting previous errors, and developing the Precipitation pages.

Casey gave an update on exposing an API or automated accessible download mechanism. His department had a large move so he did not have the chance investigate this further. He relayed that the IT department is weary of automated access types like this because of potential Denial of Service (DOS) attacks (https://en.wikipedia.org/wiki/Denial-of-service_attack). He relayed that there has already been an attempted DOS attack on the NADP web site before any talk of automated access. Casey will continue to investigate this as a possible data access tool.

WEB FUTURE PLANS:

The team will continue with efforts to produce the precipitation network, correct existing issues, and investigate the API / automated access question.

MAP UPDATES

Mark will need to update his legacy ArcGIS software to ArcGIS Pro. He will need to ensure the scripts used to produce the maps still work with the ArcGIS Pro. Greg Beachley volunteered his assistance if needed, as he had just upgraded and can help in places where he had troubles.

The topic of updates to the AMON quarterly charts did not garner a lot of additional feedback. Mark takes this as a sign that it is not a high priority with the audience. Mark will not be making additional enhancements at this time. However, in the future, Mark would like to investigate one

earlier suggestion made to produce a quarterly chart over multiple years to compare multiple years by quarter.

Chris Rodgers asked why we had a gap in the 2022 maps, where sites in ND and SD failed to make the map criteria. Dana indicated there were a lot of 'v' note codes due to mountain sites efficiency performing poorly during snowy weather. Additionally, some sites suffered from long running mechanical problems. The NADP PO will be implementing some ways to monitor the criteria of sites throughout the year so we can focus on those in danger of developing coverage issues earlier during the year. One will be a basic report, while another may be a map illustrating risk. Tim Sharac has already proposed one possible map for this purpose. Richard will be discussing this further at the NOS meeting.

Tim advised he had a SQL Script he used to monitor site performance from Castnet data. He agreed to share the script as a possible resource, but advised it is long and complex.

One of the issues raised in other meetings is how we treat samples with a 'v' notes code. There was some question of whether the executive committee had voted on an action that is not being adhered to. Dana will be discussing this further in the NOS meeting this spring for discussion.

Tim asked a follow up question on whether the wind levels could have had an impact on the efficiency. Dana confirmed that in winter months this is especially true. Many sites have wind shields on the precipitation gauges, but not on the collector.

MAP FUTURE PLANS:

After the scripts are running, Mark would like to correct the 2016 and 2017 NH4 rasters (grid files), as well as the precipitation maps. The precipitation maps from 2016 forward. It was discovered last fall that the precipitation map did not include the NADP precip data, starting from 2016 onward. Mark proposes to update the raster file download page to append the updated maps next to the existing maps so former data will remain available.

DATA PRODUCT UPDATES:

Mark has published the formal data request form. Please use this for future requests. It can be accessed at this site link:

(https://uwmadison.col.qualtrics.com/jfe/form/SV_80TTeHYtBkdIxW6)

Mark has also published the change log spreadsheets for the NTN and MDN networks. These are network specific and are located in the Data Notices section on each network's web page (<https://nadp.slh.wisc.edu/networks/national-trends-network/>), (<https://nadp.slh.wisc.edu/networks/mercury-deposition-network/>). There is a metadata file in the Metadata section that described the fields used in the spreadsheet. The goal of this tool is to allow researchers to see when changes have occurred, what time frame the changes affect, what

was modified, and possibly why. This includes both site ID's as well as sample ID's where feasible. These files will be occasionally be updated.

In addition to these, Mark discussed updates to sample data due to time clock offset issues. This occurs when the NADP PO is unaware that the clock for a sites data logger has changed from GMT to local time in the past. In this case, the precipitation data becomes offset from the sample data which can cause samples to become invalid and measurements to be incorrect. The News section on the web site has a description of this issue and how the NADP PO has been dealing with it. Additionally, Mark will discuss this briefly in the NOS meeting.

ADDITIONAL DISCUSSIONS:

Tim updated the group on the wind rose project. This was discussed in QAAG and the group was nearly unanimous in its voting on how to handle sites where the wind is not predominately westerly. Time has also compiled two spreadsheets which contained a list of all sites with infractions, and a list of all sites overall. We need to post these to the NADP website, linked to either the QAAG or NOS committees. Richard will work with Tim to post these spreadsheets.

DMAG GOING FORWARD:

There were no motions raised to bring forward to Executive for the Spring 2024 meeting.

Zac and Mark will continue to co-chair going forward.