STATE BOARD MONITORING SPECIAL STUDY
Technical Workgroup Meeting #9 – Data Integration
December 12, 2023
9:30 am – 11:30 am

MEETING NOTES

Attendees (listed alphabetically)

In Person

- Eli Ateljevich/DWR
- Manny Bahia/State Water Contractors
- Erika Britney, ICF
- Bill McLaughlin/DWR
- Jacob McQuirk/DWR
- Jenna O'Neill/ICF
- Nicky Sandhu/DWR
- Teresa Trinh/DWR
- Karen Tolentino/DWR

Via Teams

- Bryan Barnhart/DWR
- Lauren Beaudin/State Water Resources Control Board
- Tom Boardman/Westlands Water District
- Ching-Fu Chang/Contra Costa Water District
- Chandra Chilmakuri/State Water Contractors
- John Collins/State Water Resources Control Board
- Jared Frantzich/DWR
- Bryant Giorgi/DWR
- Jelena Hartman/State Water Resources Control Board
- Tracy Hinojosa/DWR
- Dave Huston/DWR
- Michelle Leinfelder/UC Extension
- Hans Kim/DWR
- Parviz Nader-Tehrani/DWR
- Jane Tannous/DWR
- Grace Windler/USBR
- Zhenlin Zhang/DWR
- Denise/Captioner

Agreements and Action Items

- **Kalman Filter.** Ching-Fu Chang expressed a preference for the response method because of concerns about the nudging component of the Kalman Filter and DWR agreed to focus on this method. He also has ideas about how to bypass the built-in nudging in the Kalman filter. DWR has previously presented a formulation of the Kalman filter that does not include state nudging. Participants agreed DWR and CCWD can work together on these ideas at some point; however, the topic seems moot as long as the focus is on the response methods.
- **Flow and the relationship between flow and salinity.** Tom Burke was concerned about the role of flow in slide 9 and Jelena Hartman was interested in using models to explore the role

of variations in flow in salinity levels. DWR agreed flow and barriers both contribute to the water quality. The source of mass and consequences of omitting it in modeling was the focus of the present discussion, but the implications of the results on the larger study are apparent. The full dynamics of conditions shown on this slide, as well as a revised graphic including gates and flow conditions and exploration using other flows, will be discussed extensively as part of the main study.

- **Salt Load Map-DCD vs inferred at CCWD intake location.** Ching-Fu Chang wanted to know where the source near the CCWD intake comes from (Salt Load Map, slide 21) – is it inferred or is it DCD? Zhenlin said we can confirm that it is DCD, but we need to confirm the source is not from our inferred sources. Zhenlin is working on plots that delineate the two. She will send the revised plots to Ching-Fu and see if he agrees. Even if the sources are not produced by the data assimilation process, some revision of DCD may be warranted — in some areas with limited information the sources and sinks of applied water are generic.
- **Tom Paine Slough-flows and salinity.** Tom Burke asked for an offline discussion about the high salinity in Tom Paine Slough and the flow patterns into Paradise Cut and offered to help us investigate whether there is a discernable relationship between the EC of diversions at Tom Paine and returns on Paradise Cut.
- **Distinguishing between DCD vs. inferred/assimilated sources/sinks.** Twice in the presentation it seemed that separating DCD-generated sources and sinks from assimilated sources of mass will make both easier to interpret.
- **Peer review on technical method.** DWR will be submitting the write-up of methods to a peer reviewed journal to make sure that our study jives with this body of literature.

Welcome

Teresa Trinh opened the meeting. This is the last MSS meeting of 2023. The topic for this meeting will be data integration and data assimilation and will be led by our DWR modeling team, Eli Ateljevich and Zhenlin Zhang. This will be an interactive and collaborative workshop, so we will be taking questions and comments along the way instead of holding them until the end.

Data Assimilation and Data Integration Workshop

The purpose of this workshop is informational and to help everyone understand data assimilation and data related to the MSS. Many stakeholders expressed an interest in understanding what data assimilation is and how it fits into the project.

A summary of discussion during the workshop is organized by slide number and the exchanges between participants and DWR staff.

Discussions related to Slide 9

Tom Burke, Hydraulic Systems for SDWA. Is there a version of this slide that shows flow? He was concerned that just looking at salinity in isolation would not provide the whole story, that we may be distracted a bit if we ignore flow in this process. For example, local

contributions may appear high if the flow [at Vernalis] was very low, then the local contributions to salinity may have a greater impact. That doesn't mean they're necessarily driving the salinity levels except when Vernalis itself is extremely low. Further, he agrees the salt must be interior generated or generated from inflow. But I'm not sure it's a good idea to separate flow from salinity because it could be that the low flows at Vernalis during those periods just was not flushing the interior salts out.

Eli Ateljevich, DWR: Clarified that the missing piece is that Tom would like to see is a second plot below [the ones on Slide 9] that show flows, whether it happens with low flows. He agreed flows are an important element to the overall circumstances, indicating that ultimately modeling will need to factor in salinity, flows, and barriers. Such a graphic will be generated in the main study. Eli clarified that the modeling team is currently trying to come up with the ingredients needed to do good modeling and in this segment of the project the mass sources of salinity. Ultimately, the modeling will look at the holistic effect of South Delta sources, e.g., inflow at Vernalis, the influence of the barriers and salinity fluxes. Those three variables are conspiring to create the concentrations shown in the slide. You won't be able to model it without all three.

Ching-Fu Chang, Contra Costa Water District. Pointed the following observations from this graph [on Slide 9]:

- It is the hybrid period (not the Delta source period) where we see the different patterns between OLD and SJR;
- Although there are fluctuations, there appears to be a decreasing trend in the OLD pattern from the beginning of the hybrid period towards the end of the Delta source period. It appears salinity is decreasing on a steady trend.
- At SJR there is a sudden drop in salinity from the end of hybrid period to the beginning of Delta source period.
- Some processes are happening on the Vernalis side, but not on the OLD side where salinity at OLD is experiencing a steady linear decrease overall.

Eli Ateljevich: While salinity deceases at Vernalis during the hybrid period, it could be best described as ups and downs and salinity at OLD and SRJ never detach completely from Vernalis. The signal-to-noise ratio is so disadvantageous he is concerned whether detecting a trend is possible. If anything, salinity goes down everywhere.

Zhenlin Zhang, DWR (via chat): Provided a DMS Datastore link that participants can look at for Vernalis flow at Old and Tom Paine Slough. This is an interface [on the site] that will allow you combine things together into one plot and take a look over a longer period and see whether [a trend] is persistent or not.
https://dwrbdodatastore.azurewebsites.net/repoui?sdate=2020-12-02T00%3A26&edate=2023-12-

12T00%3A26&repo_level=%5B%27screened%27%5D&selections=old%7C%7Cflow%2Cold%7C%7Cec%2Cver%7C%7Cec%2Cvns%7C%7Cflow.

Discussions related to Slide 19

Tom Burke, Hydraulic Systems for SDWA: Asked for clarification on whether the model inferences include both flow and concentration or is it just concentration that's being imposed on the model at that location. He reiterated that in order to ensure that you're selecting appropriate concentration for the source inflow, you need to know the flow rate.

Eli Ateljevich, DWR: The inferences only include concentration (which will be interpreted as mass of salt). Flow is taken from the modeling assumptions document, as implemented through adjustments to DeltaCD. Salinity numbers are best interpreted as mass, as this is a less sensitive parameter. The calculations and sensitivities will be described in the report. An exception may be Paradise Cut where in order to match gauges, you need to get the right concentration to show the influence on Old River.

Ching-Fu Chang, Contra Costa Water Agency: Which method will be used? If not nudging, are both Response-Based Method and the Kalman Filter Method being used? My preference would be the Response-Based Method, or a hybrid with the Kalman Filter. He asked DWR to elaborate on the built-in nudging of the Kalman Filter and indicated that he has ideas for how to by-pass the built-in nudging in the Kalman Filter.

Eli Ateljevich, DWR: DWR modeling team's (we/our) preference is for our preference is for the Response-Based Method. We originally introduced it because we thought the response method would be the easiest one to describe, but we got good results and open to using it and Response-Based Method. The tradeoffs are:

- The Response Based Method identify does not require pre-identification of a small number of candidate sources.
- Kalman Filter entertain slightly nonlinear relationships. It is kind of useful if you want to incorporate flow uncertainty directly into it but lead to a lot of unknowns being inferred. At the moment this advantage goes unused -- we would rather leave things as having been data-driven in the flow department and infer the mass sources.

Zhenlin Zhang, DWR: Added that she ran one Kalman Filter run which removed the nudging component. The results were very similar to the ones with nudging, the inferred sources were almost the same. However, using Kalman Filter built-in nudging seems to make the model more stable, but she did not know why. The classic explanation is that correcting the state prevents inferring sources to make up for accumulated modeling error.

Ultimately, DWR plans to use the Response-Based Method and it appears.

Lauren Beaudin, SWRCB: Pointed out that reserve stations and the validation point are a good flag, but asked how it will be determined which stations will be used for the data assimilation process verses the validation process? She agreed the "leave-one-out" approach described by Eli below would be helpful.

Eli Ateljevich: We are needing to take an iterative approach based on how leaving out a station affects accuracy of inference, for example:

- There is a station upstream of the Five Points area that is very important to the inference. If you have that station in, you can clearly infer the presence of the Tracy Wastewater Treatment Plant influence. If you remove that station, you see the same mass inferred over a wider area, and it gets mis-ascribed to other potential sources. When we put it back in, it sharpens up and you see that influence. Most other single stations can be removed without doing that kind of harm to the inference. What monitoring stations we need to have is tied to the question of which current stations we can afford to hold back without compromising inference.
- As a test, we can pull stations out one at a time and see how well they're inferred by all the others. Do that sequentially and we come up with the equivalent of a leave-one-out-cross-validation.

Does that sound reasonable if we do the leave-one-out approach and we do them all? We are open to suggestions about which stations to remove.

Nicky Sandhu/DWR: Pointed out there will always be uncertainty. In terms of sources, that's the highest uncertainty in terms of the mass modeling from them. He asked that we keep in mind that there are ranges of uncertainty for everything, including things that we measure.

Discussions related to Slide 21

Ching-Fu Chang, CCWD: Your explanation of the method, indicated that the issue with the superposition response-based fingerprinting method is an undetermined regression problem. What kind of method did you use for the superposition to come up with this result? It will be good to include the details of the method in the report.

Eli Ateljevich, DWR: Agreed. There are two regularization methods applied:

1. A penalty on sharp derivatives in the time series (things that are sort of weird mathematical artifacts they don't look like a real signal). This is a smoothness assumption.
2. A sparsity assumption, which means that there is a parameter controlling the number of source locations. When a larger number of sources are included, you tend to get overfit.

DWR plans to get this [study & description of methods] published in a peer reviewed journal so people who are used to using these methods will review it. Eli confirmed methods will also be described in the final report.

Ching-Fu Chang, CCWD: In the top three panels [of Slide 21], you see a huge red source right next to a fairly big blue sink. What would be the physical explanation of this?

Zhenlin Zhang: The salt load is composed of two components:

1. Delta CD, our base model, that's the model that we've used in the past historically to model the EC in this field in the South Delta.
2. The source being inferred.

For the Delta CD it shows up as a sink because more mass is taken out than going in. Our data assimilation methods do not infer sinks. Net withdrawal of water from a node will show up as a sink, that's not an inferred salt load, the model is using data from the DeltaCD. The information on sources will be in the final report.

Tom Burke, Hydraulic Systems for SDWA: Were sinks inferred, or were some of them from DCD also?

Eli Ateljevich: Sinks are almost entirely from DCD and are the byproduct of null zones. Some sources are from DCD, but the big ones are inferred. The final report will separate which salt load is from DCD and which is from the inferred source, as this was an obvious interest of many participants in the meeting.

Tom Burke, Hydraulic Systems for SDWA: Do you have a way to double check your inferences with field data to confirm whether it makes sense to have an inference at that location?

Eli Ateljevich: We can compare to detailed spatial data, but we can't get specific location point sources and, in most cases, don't have data. The method should get things right regionally, but it's unlikely that we could identify individual sources in the field (like a pipe or a discharge) and there are limitations on how accurately DSM2 can depict the locations of sources. We will try to demonstrate that we can match the patterns of the data that we see in the surface transects, plus in the withheld data (see also response concerning cross-validation).

Jelena Hartman, SWRCB: Wanted to be sure that we are looking at the "big picture" context, and we are not just seeing these results because of the low flows during the model period. This does fall within the period where Vernalis flows at 100 or 200 CFS. The connection to flow when looking at these sources and sinks of the load matter, because when the flow is low we have maybe 100 or 200 at Vernalis. But then in applying the model, can we get a sense at which kind of flow or conditions did these loads start to not matter?

Eli Ateljevich: Pointed out that studying this is exactly what the main study is intended to do in a nutshell. That is the end point that we want to get to.

Parviz Nader-Tehrani, DWR initiated a discussion about Tom Paine Slough and whether any water there could circulate back to Paradise Cut. DWR did a comparison and did not see much of a co-variance between the two. But this doesn't mean that they're not connected. There could be a link, but we don't know yet. Tom Burke pointed out that there's actually very little irrigation water that comes into Tom Paine Slough. There is a strong correlation between the water withdrawn from Tom Paine Slough and the discharge into Paradise Cut. Tom is unsure why Tom Payne Slough has such high salt concentration, because almost no irrigation water goes into it. It's primarily a source for diversion. Tom also pointed out that there is a correlation with groundwater. By the time water gets to Paradise Cut, it's not only the water that came out of Tom Paine Slough but it's mixed with a lot of groundwater that's discharged into Paradise Cut.

Eli Ateljevich: We would love to work with you more on this. We did not see the relationship you would expect from this story line. It is the storyline as we understand it as well. But you don't see the fluctuations that originate from Tom Paine showing up at some delayed fashion at Paradise Cut. It's a bit confusing and indicates that there's other sources or something else going on. [Note: follow-up using simple mass balance indicates that some source of salinity is needed to explain elevated levels of salinity in Tom Paine]

Ching-Fu Chang, CCWD: Identified an issue regarding a large source at the node closest to the CCWD intake and asked DWR to clarify or look into what is going on in the model that it would appear that way. He asked that DWR provide a summary of which sources (on the salt load maps) are from DCD and which are inferred and that DWR share a report of DCD if there is one.

Zhenlin Zhang, DWR: This means there is a large flux that goes into that location, it does not mean that EC there is high. The dot represents EC x Flow, the product of the two is high. We will follow concerning whether this is inferred or DCD.

Discussion and Q&A

- Eli Ateljevich, *DWR* mentioned that they are available by email if there are questions or comments that you want to bring up to them outside of the meeting.
- Tom Burke*, Hydraulic Systems for SDWA* would like an off-line discussion about Tom Paine Slough and the flow patterns to Pescadero, as well as the model's sources for Tom Paine.
- Ching-Fu, *CCWD* acknowledged that details of Response-Based Method will be in the final report but said that it seems like all discussion has been Kalman-filter based. He asked to be able to review this before the final report comes out. Tom would like a copy of this too.
- Manny Bahia, *SWC* asked what is the timeline for the report and the peer review? Eli outlined the schedule: the final report is due December 2024. To make this deadline, we

need to get a draft out for review for several months by mid-summer 2024. The data assimilation methodology needs to be put out for peer review within the modeling community by April. Draft of that earlier. We will also have draft loadings available to use so that they can be rolled out for the other projects.

- Jacob Mcquirk, *DWR* encouraged everyone to attend in-person meetings. If we don't have the attendance to warrant them, then we don't waste the resources and go back to virtual meetings.

Closing & next steps

- The MSS is in the process of ending data collection and moving into the Final MSS report preparation. The target is to submit the report to the Water Board by the end of 2024.
- The MSS website is up and running; this is where all meeting information like notes and presentations will be posted.
- The next meeting will be around March. Until then, you may contact Teresa Trinh or Bill McLaughlin about what you would like to hear at the next public meeting or any questions that anyone may have.