emp-data

Catarina Pien & Lilly McCormick

# Goal

Examine USBR stations listed in the Water Quality Control Plan to determine whether we are meeting requirements.

## Notes

**Communications from Ted**

* The station contact is Melanie Lowe at USBR’s California-Great Basin Regional Office
* When we last emailed in December, she said “our office staff clean and calibrate the stations monthly when we are able to access them, and Central Valley Office coordinates with Tracy Office regarding issues w/access or stations going down/not reporting. However, I believe Tracy Office is currently cleaning/calibrating/maintaining the Collinsville and Ripon stations routinely.”
* C19 discontinued due to safety concerns and because the city of Vallejo no longer uses the intake on Cache Slough. Ted was not able to find any documentation that this was ever officially communicated to SWRCB.

**Requirements in Table 5 of the WQCP**

* **For CLL, SAL, CNT, UNI, PCT, CCS, EMM**: Continuous recording (every 15 minutes) of water temperature, electrical conductivity (EC), and/or dissolved oxygen. For municipal and industrial intake chloride objectives, EC can be monitored and converted to chloride concentration.
* **For DMC**: Continuous, multi-parameter monitoring (recording every 1 to 15 minutes with telemetry capabilities) includes the following variables: water temperature, EC, pH, dissolved oxygen, turbidity, chlorophyll a fluorescence, tidal elevation, and meteorological data (air temperature, wind speed and direction, solar radiation).

**What is on CDEC**

* C2 – Sacramento River at Collinsville (CLL): EC (100 - event), WT (25 - hourly)
* C4 – SJR at San Andreas Landing (SAL): EC (100 - event), WT (25 - hourly)
* C5 – Contra Costa Canal at Pumping Plant #1 (CNT): EC (100 - event), WT (25 - hourly)
* C8 – Old River near Middle River (UNI): EC (100 - event), WT (25 - hourly)
* C14 – Sacramento River at Port Chicago (PCT): EC (sensor 100), WT (25 - event)
* \*C19 – Cache Slough at City of Vallejo Intake (CCS): discontinued 1/1/2015
* D22 – Sacramento River at Emmaton (EMM): EC (100 - event), WT (25 - hourly)
* DMC1 – Delta-Mendota Canal at Jones Pumping Plant (DMC): EC (100 - event), WT (25 - hourly)

**Conclusions**

* We are unsure if we are supposed to be collecting WT, EC, AND DO at the 15-minute interval. If so, we are only collecting EC at the frequency specified by Table 5 for water temperature.
  + Ted: In some cases, D-1641 has explicit compliance requirements related to one or the other parameters (e.g., EC at D10 and C14 as described in Table 4). This is also true for DO at P8, although that’s spelled out in the Bay-Delta Water Quality Control Plan rather than D-1641.
  + Scott/Ted: Since 2007, the Continuous Environmental Monitoring Program (CEMP) has been capturing real-time data every 15 minutes for all water quality constituents, including SpC (EC), DO, and WT, from our 15 monitoring stations. / Scott’s decision to have his group move to fully-equipped instruments (pH, EC, Temp, DO, Chl-a, and Turbidity sensors) for all of their baseline/compliance stations even when not explicitly required by D-1641 came partially from internal DWR requests for that data but also just to avoid any confusion…generally our thought is calibrating six sensors on an EXO2 doesn’t really cost all that much more compared to just temp and EC (in either equipment or staff time) and the data is often quite valuable.
* We are not collecting all the parameters specifed by Table 5 for DMC.
* Most stations seem to be collected relatively well for 2024, except CNT which was missing a few months of data.
* CLL and DMC also have >10% data outside of suitable range/not collected.

**Next Steps**

* We could get the QA/QC protocols from Melanie’s group. All we did here was basic range filtering, and not 100% sure of appropriate EC range.
* We could check if extra data are collected that are not reported on CDEC.
* We could check why certain parameters are collected at frequencies different from what has been specified in the Water Quality Control Plan.
* Ted: the new Bay-Delta Plan POI would really benefit from a separate table for continuous water quality that specifies exactly what parameters they want measured at each station. The footnote thing and differentiating between “continuous recorder” and “multiparameter” stations is really confusing

# Analysis

## Electrical Conductivity

### Download data

### Examine range of values

ggplot(ec %>% filter(!is.na(month))) +   
 geom\_boxplot(aes(month,value, fill = month))+  
 facet\_wrap(~station\_id, scales = "free\_y", ncol = 2) +  
 labs(y = "EC (uS/cm)")+  
 theme\_bw() +   
 theme(legend.position = "bottom")

|  |
| --- |
| Boxplot of EC (uS/cm) values in 2024. |

Filtered to suitable ranges, with 1000 uS/cm as upper limit for CNT, DMC, SAL, UNI, and no upper limit for other stations.

### Histogram for filtered values

ggplot(ec %>%filter(range == 1)) + geom\_histogram(aes(value), color = "lightblue4")+  
 facet\_wrap(~station\_id, scales = "free\_x", ncol = 2) +  
 labs(x = "EC (uS/cm)")+  
 theme\_bw()

|  |
| --- |
| Histogram of EC values. |

## Water Temperature

### Download Data

### Examine range of water temperature values

ggplot(wt %>% filter(!is.na(month))) +   
 geom\_boxplot(aes(month, value, fill = month))+  
 facet\_wrap(~station\_id, scales = "free\_y", ncol = 2)+  
 labs(y = "Water Temperature (F)")+  
 theme\_bw()+  
 theme(legend.position = "bottom")

|  |
| --- |
| Boxplot of WT (deg F) values in 2024. |

Filtered to suitable ranges, with 30F as lower limit and 100F as upper limit. Not sure what the deal with DMC is in February, CLL in August.

### Filtered data histogram

ggplot(wt %>%filter(range == 1)) + geom\_histogram(aes(value), color = "lightblue4")+  
 facet\_wrap(~station\_id, ncol = 2) +  
 labs(x = "Water Temperature (F)")+  
 theme\_bw()

|  |
| --- |
| Histogram of WT values. |

### Annual Summary

prop\_ec\_year <- ec %>%  
 group\_by(station\_id) %>%  
 summarize(n = n(),  
 max = max(value),  
 min=min(value),  
 sum\_good = sum(range, na.rm = TRUE),  
 prop\_good = round(sum\_good/n,3), # data within range/data collected  
 prop\_all = round(sum\_good/(24\*4\*366),3))%>%#data within range/data that should have been collected if every time step was collected  
 mutate(sensor = "EC")

prop\_wt\_year <- wt %>%  
 group\_by(station\_id) %>%  
 summarize(n = n(),  
 max = max(value),  
 min=min(value),  
 sum\_good = sum(range, na.rm = TRUE),  
 prop\_good = round(sum\_good/n,3),  
 prop\_all = round(sum\_good/(24\*366),3)) %>%  
 mutate(sensor = "Temperature")

prop\_data <- bind\_rows(prop\_ec\_year, prop\_wt\_year)

# Summary

ggplot(prop\_data) +   
 geom\_col(aes(station\_id, prop\_all, fill = sensor), position = position\_dodge2()) +  
 scale\_fill\_manual(values = c("orange2", "steelblue4")) +  
 labs(x = "Station Code", y = "Proportion Data Within Reasonable Range") +   
 scale\_y\_continuous(breaks = seq(0, 1, by = 0.1)) +   
 theme\_bw()

|  |
| --- |
| Proportion of EC and WT Data Collected and within Range in 2024. |

library(kableExtra)  
kableExtra::kable(prop\_data)

Sample Sizes of Data and Proportion of Good Data. Prop\_good = Proportion of data that was collected that was within range. Prop\_all = Proportion of data that should have been collected (if all time points were available) that was within range.

| station\_id | n | max | min | sum\_good | prop\_good | prop\_all | sensor |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CLL | 34044 | 16177.0 | 0.0 | 34044 | 1.000 | 0.969 | EC |
| CNT | 24287 | 1836.0 | 1.0 | 24285 | 1.000 | 0.691 | EC |
| DMC | 34067 | 10028.0 | 0.0 | 34066 | 1.000 | 0.970 | EC |
| EMM | 34463 | 8588.0 | 0.0 | 34463 | 1.000 | 0.981 | EC |
| PCT | 33388 | 28175.0 | 0.0 | 33388 | 1.000 | 0.950 | EC |
| SAL | 34947 | 699.0 | 1.0 | 34947 | 1.000 | 0.995 | EC |
| UNI | 34937 | 758.0 | 0.0 | 34937 | 1.000 | 0.994 | EC |
| CLL | 8510 | 5914.0 | 8.5 | 7546 | 0.887 | 0.859 | Temperature |
| CNT | 6097 | 788.0 | 6.1 | 6087 | 0.998 | 0.693 | Temperature |
| DMC | 8520 | 205.0 | 46.4 | 7679 | 0.901 | 0.874 | Temperature |
| EMM | 8618 | 72.3 | 13.6 | 8548 | 0.992 | 0.973 | Temperature |
| PCT | 8348 | 73.6 | 49.5 | 8348 | 1.000 | 0.950 | Temperature |
| SAL | 8738 | 78.6 | 47.8 | 8738 | 1.000 | 0.995 | Temperature |
| UNI | 8739 | 88.3 | 33.1 | 8739 | 1.000 | 0.995 | Temperature |

# Individual Station Plots

## Electrical Conductivity

plot\_ec("CLL")

|  |
| --- |
| Proportion of EC values collected and within range by day at CLL. White boxes are missing data for the entire day. |

plot\_ec("CNT")

|  |
| --- |
| Proportion of EC values collected and within range by day at CNT. White boxes are missing data for the entire day. |

plot\_ec("SAL")

|  |
| --- |
| Proportion of EC values collected and within range by day at SAL White boxes are missing data for the entire day. |

plot\_ec("UNI")

|  |
| --- |
| Proportion of EC values collected and within range by day at UNI White boxes are missing data for the entire day. |

plot\_ec("PCT")

|  |
| --- |
| Proportion of EC values collected and within range by day at PCT. White boxes are missing data for the entire day. |

plot\_ec("EMM")

|  |
| --- |
| Proportion of EC values collected and within range by day at EMM. White boxes are missing data for the entire day. |

plot\_ec("DMC")

|  |
| --- |
| Proportion of EC values collected and within range by day at DMC. White boxes are missing data for the entire day. |

## Water Temperature

plot\_wt("CLL")

|  |
| --- |
| Proportion of WT values collected and within range by day at CLL. White boxes are missing data for the entire day. |

plot\_wt("CNT")

|  |
| --- |
| Proportion of WT values collected and within range by day at CNT. White boxes are missing data for the entire day. |

plot\_wt("SAL")

|  |
| --- |
| Proportion of WT values collected and within range by day at CNT. White boxes are missing data for the entire day. |

plot\_wt("UNI")

|  |
| --- |
| Proportion of WT values collected and within range by day at UNI. White boxes are missing data for the entire day. |

plot\_wt("PCT")

|  |
| --- |
| Proportion of WT values collected and within range by day at PCT. White boxes are missing data for the entire day. |

plot\_wt("EMM")

|  |
| --- |
| Proportion of WT values collected and within range by day at EMM. White boxes are missing data for the entire day. |

plot\_wt("DMC")

|  |
| --- |
| Proportion of WT values collected and within range by day at DMC. White boxes are missing data for the entire day. |