Memorandum

DATE: June 18, 2014

TO: Richard Meyerhoff, CV-SALTS TPM

COPY TO: CV-SALTS Phase II Project Committee

SUBJECT: Phase II Conceptual Model - Task 3: Groundwater Data Refinements and Updates

This memorandum is being submitted on behalf of the LWA team and fulfills the Task 3 requirement of the CV-SALTS Phase II Conceptual Model Workplan (Workplan). The purpose of this memorandum is to discuss revisions and additional data gathered to update the groundwater quality database that was developed as part of the CV-SALTS Phase I Initial Conceptual Model (ICM) workplan. The ICM adopted the use of 22 Initial Analysis Zones, or IAZs, covering the Central Valley floor for purposes of the conceptual scale water, salt, and nitrate balance calculations. The data collected for the Phase I ICM groundwater quality database included not only the 22 IAZs, but the entirety of the Central Valley Regional Water Quality Control Board (RWQCB) Region 5 jurisdiction. Updates to the database have focused on the entirety of Region 5 for all time periods.

The data used in the CV-SALTS groundwater quality database originate from five sources:

- RWQCB Waste Discharge Requirements (WDR) data per the Dairy CARES program (Dairy);
- California Department of Public Health (CDPH);
- Department of Water Resources (DWR);
- the United States Geological Survey’s (USGS) National Water Information System (NWIS) program; and
- Geotracker Groundwater Ambient Monitoring and Assessment (GAMA) program.

1 The LWA Team consists of the following firms: Larry Walker Associates, Luhdorff and Scalmanini Consulting Engineers, Kennedy/Jenks Consultants, PlanTierra, Systech Water Resources, Carollo Engineers, Giorgos Kourakos (independent consultant), and Formation Environmental, LLC.
These source databases contain publicly available salinity and nitrate test data from wells throughout Region 5. Updates to the groundwater quality database included re-acquiring Geotracker GAMA data (to correct systematic errors discovered in the raw data gathered from the Geotracker GAMA online database, identified in Phase I\(^2\)), and updating database sources (USGS NWIS, DWR, and CDPH) with data current to 2014. This included adding any additional test results that the source databases contained, current to 2014. The RWQCB WDR Dairy data were not updated\(^3\).

The groundwater quality database consists of nitrate (as nitrogen) (NO\(_3\)-N) and total dissolved solids (TDS) test results of groundwater samples from wells. Types of groundwater wells include domestic, public supply, industrial, monitoring, irrigation, and stock wells. Records that provided test results of these exact analyte results were preferentially selected; however, test results of nitrate (as nitrate) (NO\(_3\)-NO\(_3\)), electrical conductivity (EC), and specific conductivity (SC) were also acquired. NO\(_3\)-NO\(_3\) was converted to NO\(_3\)-N by dividing NO\(_3\)-NO\(_3\) by 4.4268, and EC or SC was converted to TDS by multiplying the results by 0.64, a commonly used multiplier\(^4\).

**QA/QC**

A systematic approach was used for quality assurance/quality control (QA/QC) of assembled groundwater quality database to filter out any erroneous records, identify non-detections, and filter out duplications of test results between and within databases. The sources of data report test results differently and various qualifiers, remarks, and comments are used to provide additional information. When assembling the database, care was taken to include as many tests as possible; however, any conversions and assumptions have been noted to aid in using the database appropriately.

**Non-Detections**

A non-detection indicates when a laboratory method was not able to detect a concentration for a particular analyte. The detection limit of a constituent is not always reported or clear. In the compiled database for CV-SALTS, non-detections for nitrate are assigned the value of 0.225 mg/L NO\(_3\)-N (2 mg/L NO\(_3\)-NO\(_3\), which is the “Detection Limit for Purposes of Reporting (DLR)” used by CDPH and the SWRCB\(^5\)). A value of 10 mg/L TDS was used for non-detections of TDS and EC/SC.

**Estimated Concentrations**

Some records in the source databases indicated the reported concentration is an estimate. In the USGS NWIS database, an “E” for estimated is used as a remark code, or a “value extrapolated at low end” qualifier is used to indicate the reported concentration is an estimate. In these cases, the

---

\(^2\) A large number of well tests for nitrate and TDS were discovered to have reported in the wrong units (nitrate as N instead of nitrate as nitrate, and mg/L TDS instead of µg/L TDS). See CV-SALTS Initial Conceptual Model (ICM) Phase I, Tasks 7 and 8 Final Report (2013) section 3.1 for more detail.

\(^3\) The RWQCB WDR Dairy dataset was not scoped to be updated for the entirety of Region 5 in Phase II, due to the amount of effort involved in acquiring, compiling, determining well locations, and performing quality control/quality assurance. Instead, the previous RWQCB WDR Dairy dataset established for Dairy CARES was used.

\(^4\) EC and SC data were collected and transformed to TDS using the ratio TDS = EC*0.64 for wells without TDS data (Tchobanoglous and Burton.1991.). It should be noted that this ratio is a common approximation, and can change depending on local conditions. Converted SC/EC tests are indicated in the database in the comment field.

\(^5\) GROUNDWATER INFORMATION SHEET: Nitrate http://www.waterboards.ca.gov/gama/docs/coc_nitrate.pdf
estimated value was kept in the database as is (or converted to NO$_3$-N or TDS if appropriate). Some USGS records do not provide a concentration but state: “presence verified but not quantified”. These records were not kept as the description is too vague to assign a concentration value. In the USGS NWIS and Geotracker GAMA (specifically the Electronic Data File (EDF) portion) databases, some test concentrations are reported with a “>” sign, indicating that the actual concentration is greater than the reported concentrations. The USGS NWIS database also uses a qualifier code of “d” meaning a “diluted sample: method hi range exceeded”. This can happen when the laboratory method determines a concentration above the limit of the method used. It is assumed that the actual concentration is higher. In these cases, the reported value was kept in the database as is (or converted to NO$_3$-N or TDS if appropriate).

**Other Data Qualifiers**

The USGS uses a qualifier “V” indicating a “value affected by contamination”. This occurred in a few instances only for TDS concentrations. While contamination of a sample could be very important for some analytes, it was assumed that the results of a potentially contaminated sample analyzed for TDS would not result in a significantly different value; therefore, these records were kept.

Records with a qualifier stating that the results were inaccurate or questionable were not kept. Additional characters appear in the remark or data qualifier fields which do not have a definition and are assumed to have been entered in error. Characters such as “-” and “.” in CDPH’s database are not defined, so there is no compelling reason to remove them from the dataset, therefore, these records were accepted as is.

**Data Quality Field**

A field was added to the groundwater quality database to indicate the quality of the data entered for each water quality sample entry. If the reported concentration had no qualifiers indicating that the result was inaccurate, or no assumptions were made about the test result, it was given the highest data quality value. If a test was reported as a non-detection and a concentration was assumed, or if the concentration was an estimated value, contaminated sample, or the concentration exceeded the limits of the laboratory method used, that entry received a lower quality value ranking. **Table 1** shows how these categories were assigned a 1, 2, or 3.
Table 1. Definitions of Data Quality Value Field in Database

<table>
<thead>
<tr>
<th>Data Quality Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Concentration is reported without a qualifier that indicated that the result was inaccurate. Concentration kept as is or converted to NO₃⁻-N or TDS if needed.</td>
</tr>
<tr>
<td>2</td>
<td>Concentration is reported as ND, concentration = 0, or qualifier = &lt;; 10 mg/L TDS or 0.225 NO₃⁻-N mg/L assumed. The only exception is for USGS test results for TDS as a summation of its constituents that contained a “&lt;” qualifier.</td>
</tr>
<tr>
<td>3</td>
<td>Value is reported with “&gt;” sign (indicating that the concentration exceeded the laboratory method and that the concentration is higher than what is reported), or that the concentration was estimated (for example by extrapolation of a diluted sample), or the sample may have been contaminated (only applies to TDS results). These records were included as is or converted to NO₃⁻-N or TDS if necessary.</td>
</tr>
</tbody>
</table>

WELL LATITUDE/LONGITUDE LOCATION REFINEMENTS

Precise locations for wells can greatly improve assessments of ambient conditions, especially at the local scale. Unfortunately, many of the publicly available online databases do not provide exact coordinates for wells due to confidentiality/security and/or lack of data. For example, DWR locates their wells using the State Well Number (SWN) which, at best, can only be located within a quarter-quarter section. CDPH wells, reported within the Geotracker GAMA database, are purposefully obfuscated up to a mile in any direction for security reasons. For large scale studies, such as the Phase I ICM work, imprecise locations are not as important when performing analyses over very large regions. However, for localized studies that attempt to refine ambient conditions on a higher resolution, the imprecise accuracy of the well locations may affect how ambient conditions are understood at a local scale. The following describes the data and methods used to assign latitude/longitude coordinates to wells.

The USGS, Geotracker GAMA, and previous RWQCB WDR Dairy datasets all contained latitude and longitude values for each well with salt and nitrate data. The USGS reports accurate latitude and longitude values for each of their reported wells, but the USGS does not provide well types or water uses. Geotracker GAMA only provides accurate locations for monitoring wells at regulated sites; all other locations for supply wells are approximated to within one-mile of the actual location for CDPH wells or within ½ mile of the actual location for GAMA wells (reported by SWRCB, Lawrence Livermore National Laboratory (LLNL), or USGS). The CDPH database does not include well locations. Therefore, a hierarchy of methods was used to assign locations to CDPH wells. The methods for assigning spatial locations to CDPH wells are listed below, with priority given to methods highest on the list when possible.

1. Use CDPH systems in Geotracker GAMA (locations obfuscated 1-mile in any direction).
2. Use systems in Phase I CV-SALTS database that had been geo-located previously.
3. Use an average of other wells in the same water system that have latitude/longitude locations.

USGS parameter code 70301 is used for TDS results, where the reported concentration is the sum of the measured constituents. If at least one of the constituents measured resulted in a non-detect, than the “<” qualifier is carried through for the TDS result. This does not mean that there was a non-detection for TDS, but only for 1 or more of the constituents. In this case, the original concentration was kept, but was assigned the data quality value of 2.
4. Use the latitude/longitude of the centroids of the water system boundaries provided by the Environmental Health Investigations Branch (EHIB)\(^7\) of CDPH.

5. Use the address of the Water System Headquarters and geocode the location based on street address, then city (if no street provided).
   a. This had to be checked to make sure the water system headquarters was in the same county as the primary station code indicates (first two numbers of primary station code e.g., the water system headquarters might be listed as somewhere in the Bay area, but the well primary station code indicates the well is located in a county somewhere in the Sierra Nevada mountains).

6. Additional systems that were unable to be located by the above methods were located using an online search engine for the water system name and attempts to locate the area of the well through various websites.

DWR wells were provided with latitude/longitude, but upon investigation it appeared that some wells were located incorrectly. A systematic approach was used to re-assign locations to DWR wells. Section and quarter-quarter shapefiles available from the U.S. Bureau of Land Management (BLM)\(^8\) were used to check locations. The BLM shapefile of township/range/section/quarter-quarter does not cover the entire Region 5 area. The list below shows the priority of actions used to assign locations for wells with salt and nitrate data from DWR.

1. If the provided latitude/longitude placed the well in the correct quarter-quarter section, the latitude/longitude entry was unchanged.

2. If the provided latitude/longitude placed the well in the incorrect quarter-quarter section, but the quarter-quarter section is available on the BLM shapefile, the well was moved to centroid of the correct quarter-quarter section.

3. If the provided latitude/longitude places it in the correct section, and no quarter-quarter exists in the BLM shapefile, the latitude/longitude entry was unchanged.

4. If the provided latitude/longitude placed the well in a different section, and no quarter-quarter exists in the BLM shapefile, the well was moved to the centroid of the correct section.

**SUMMARY**

The types and numbers of wells collected for each source database is shown in Table 2, and Figure 1 shows the number of wells by source and decade. The data record spans from 1909 to 2014, however, the majority of the well test data are from the 1950s to present. The earlier data (1940s-1970s) are largely from the DWR and USGS databases, with small amounts of data from the other sources. The later time period (1980s to present) contains data from all five sources, however, the GAMA and CDPH databases make up the majority. Well test data from the RWQCB WDR Dairy database is primarily in the 2000s, with a small amount of data in 1990s. Figure 2 is a map showing the spatial coverage (locations) of each well in the entire groundwater quality dataset compiled for CV-SALTS Phase II Conceptual Model, Task 3 by data source.

---

\(^7\) CDPH Environmental Health Investigations Branch (EHIB) http://www.ehib.org/page.jsp?page_key=762

\(^8\) Public Land Survey System (PLSS) http://www.geocommunicator.gov/Geocomm/lsis_home/home/index.htm
The updated database contains an additional 2,743 wells from DWR (314 within the Central Valley Floor), and an additional 2,252 wells from CDPH (142 within the Central Valley Floor). Additional duplicate wells between DWR and USGS were identified due to higher scrutiny of the data. This resulted in the identification of 10,675 duplicate wells between DWR and USGS NWIS, and a net reduction of 2,198 wells from the USGS NWIS database. In the Geotracker GAMA database, test results from the USGS NWIS database and CDPH database are identified. Instead of attempting to match-up well codes, between USGS NWIS, CDPH, and Geotracker GAMA (as was performed in the Phase I ICM database\(^9\)) all records within the Geotracker GAMA database that were indicated to have come from the USGS NWIS and CDPH source databases were removed. The assumption was made that any records from CDPH and USGS NWIS that were included in the Geotracker GAMA database would be included in the databases obtained directly from the CDPH and USGS NWIS databases. This resulted in a large refinement of the Geotracker GAMA data, with a net reduction of 7,554 wells overall, but only a reduction of 620 wells within the Central Valley Floor.

For the database as a whole, there was a net reduction of 4,538 wells. Within the Central Valley Floor, there was only a net reduction of 912 wells. This is due to the identification of duplicate wells through increased scrutiny of the data, and by using only the most up-to-date databases from original sources.

\(^9\) It was assumed that CDPH and USGS NWIS databases contained the latest and most accurate data. Often times well IDs can change through time, depending on reporting methods, updates to databases, changes in management, etc. Therefore the identification of duplicate records can often lead to incomplete identification of duplicate records and wells between the databases. For this reason, the original CDPH and USGS NWIS sources were considered the most up-to-date and were used in lieu of CDPH and USGS NWIS records contained within the Geotracker GAMA database.
### Table 2. Well Types and Number of Wells for each Data Source

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Types of Wells</th>
<th>Entire Database</th>
<th>Within IAZs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWR&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Domestic, Industrial, Public Supply, Agricultural, Monitoring/Observation/Test</td>
<td>17,150</td>
<td>13,452</td>
</tr>
<tr>
<td>USGS</td>
<td>Not Reported</td>
<td>7,850</td>
<td>3,620</td>
</tr>
<tr>
<td>Geotracker</td>
<td>Public Supply, Domestic, Monitoring</td>
<td>7,293</td>
<td>5,482</td>
</tr>
<tr>
<td>GAMA</td>
<td>Public Supply</td>
<td>9,806</td>
<td>5,682</td>
</tr>
<tr>
<td>RWQCB</td>
<td>Monitoring, Domestic, Agricultural</td>
<td>4,179</td>
<td>4,157</td>
</tr>
<tr>
<td>RWQCB (WDR Dairy Data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td><strong>45,940</strong></td>
<td><strong>32,393</strong></td>
</tr>
</tbody>
</table>

#### Number of Wells Per Decade by Source

![Number of Wells Per Decade by Source](image)

**Figure 1. Number of Wells With a Nitrate or TDS Test by Decade and by Source**

<sup>10</sup> At the time this Technical Memorandum was written, additional data from DWR are expected to be incorporated into the database upon receipt.
Figure 2. Map of Well Locations by Data Source
DATA SOURCES

<table>
<thead>
<tr>
<th>DATA SOURCE</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geotracker GAMA</td>
<td><a href="http://geotracker.waterboards.ca.gov/gama/">http://geotracker.waterboards.ca.gov/gama/</a></td>
</tr>
<tr>
<td>USGS NWIS</td>
<td><a href="http://nwis.waterdata.usgs.gov/ca/nwis/qwdata">http://nwis.waterdata.usgs.gov/ca/nwis/qwdata</a></td>
</tr>
<tr>
<td>DWR Water Data Library</td>
<td><a href="http://www.water.ca.gov/waterdatalibrary/waterquality/index.cfm">http://www.water.ca.gov/waterdatalibrary/waterquality/index.cfm</a></td>
</tr>
<tr>
<td></td>
<td>(and request from Eric Senter <a href="mailto:eric.senter@water.ca.gov">eric.senter@water.ca.gov</a>)</td>
</tr>
<tr>
<td>CDPH</td>
<td><a href="http://www.cdph.ca.gov/certlic/drinkingwater/Pages/EDTlibrary.aspx">http://www.cdph.ca.gov/certlic/drinkingwater/Pages/EDTlibrary.aspx</a></td>
</tr>
<tr>
<td>RWQCB WDR Dairy Data</td>
<td>Personal Communication with the Central Valley Regional Board</td>
</tr>
</tbody>
</table>

REFERENCES

