County of Santa Cruz

Water Advisory Commission

701 Ocean Street, Room 312, Santa Cruz, CA 95060
(831) 454-2022 TDD/TTY -Call 711 www.scceh.com
EnvironmentalHealth@santacruzcounty.us

AGENDA

SANTA CRUZ COUNTY WATER ADVISORY COMMISSION

Wednesday December 3, 2025, 4pm

This meeting will be held in hybrid format. Commissioners are expected to attend in person.
In-Person: 701 Ocean Street; **5th Floor Redwood Room**
Remote via Teams: **Join the meeting now**
Meeting ID: 279 260 926 173 27 Passcode: Rh9h4tz9
Dial in by phone +1 831-454-2222 Phone conference ID: 234 666 628#

A. OPENING

1. Call to Order
2. Roll Call.

B. PUBLIC COMMUNICATIONS

Opportunity for the public to comment on items under the purview of the Water Advisory Commission but not on today's agenda.

C. CONSENT AGENDA

Items on the consent agenda are considered to be routine in nature and will be acted upon in one motion. Specific items may be removed by members of the advisory body or public for separate consideration and discussion. Routine items that will be found on the consent agenda are meeting minutes, drought response updates, and Groundwater Sustainability Agency updates.

1. Approval of Meeting Minutes for October 1, 2025
2. Update from Groundwater Sustainability Agencies
3. Drought Response Updates

D. COMMISSIONERS' REPORTS

Opportunity for Commissioners to provide brief updates

E. STAFF REPORTS AND ANNOUNCEMENTS

Opportunity for staff to provide brief updates

F. NEW BUSINESS

1. New County Well Database

Kevin Harmon will present the County's approach to updating and modernizing the geodatabase of wells that the County maintains. The new approach uses AI technology to read and extract data from the collection of well completion reports stored by the County and the Department of Water Resources.

Attachments: Draft Well Layer Update Methodology Report and Appendix.

2. 2025 Water Status Report

Review and provide comments on the 2025 Water Status Report update.

Attachments: Draft 2025 Water Status Report

G. UNFINISHED BUSINESS and UPDATES

None

H. CORRESPONDENCE

None

I. BOARD OF SUPERVISORS ACTION ON ITEMS AFFECTING WATER:

October 21, 2025, Items 9 and 10 regarding Measure Q Vision Plan and Funding Allocations.

J. ITEMS OF INTEREST

Big Basin Water Co. operations have stabilized, but long-term ownership remains elusive. Santa Cruz Sentinel, November 18 2025.

Loch Lomond reservoir poised to spill after atmospheric river storm drenches Santa Cruz County. Santa Cruz Sentinel, November 13, 2025

K. AGENDA ITEMS FOR FUTURE MEETINGS

Workplan

LAMP Updates

New Environmental Health Software (HS Gov Tech)

L. ADJOURNMENT

County of Santa Cruz

Water Advisory Commission

701 Ocean Street, Room 312, Santa Cruz, CA 95060
(831) 454-2022 TDD/TTY -Call 711 www.scceh.com
EnvironmentalHealth@santacruzcounty.us

Meeting Minutes

SANTA CRUZ COUNTY WATER ADVISORY COMMISSION

Wednesday October 1, 2025, 4pm

In-Person: 701 Ocean Street; **5th Floor Redwood Room**

A. **OPENING**

1. Call to Order
2. Roll Call. Welcome new Commissioner Sarah Easley Perez.

Commissioners in attendance: Cheap, Gillespie, Lockwood, Wilson, Perez
Commissioners absent with notification: Lego.
Staff in attendance: Sierra Ryan, Sean Abbey, David Reid, Clare Peabody

Commissioner Perez introduces herself to the Commission as the new District 5 Commissioner

B. **PUBLIC COMMUNICATIONS**

Opportunity for the public to comment on items under the purview of the Water Advisory Commission but not on today's agenda.

Becky Steinbruner expresses concern about battery storage. Dioxins and PFAS have been found by Blue Shirt Justice group on the Never Again Moss Landing website. She encourages the commission to comment on the 19 Minto road proposal. She says Central Water District wrote a letter in opposition to that project.

C. **CONSENT AGENDA**

Items on the consent agenda are considered to be routine in nature and will be acted upon in one motion. Specific items may be removed by members of the advisory body or public for separate consideration and discussion. Routine items that will be found on the consent agenda are meeting minutes, drought response updates, and Groundwater Sustainability Agency updates.

1. Approval of Meeting Minutes for August 6, 2025
2. Update from Groundwater Sustainability Agencies

3. Drought Response Updates

Moved by Lockwood with minor changes, seconded by Cheap. Perez abstained from item 1. Approved unanimously.

D. COMMISSIONERS' REPORTS

Opportunity for Commissioners to provide brief updates

Lockwood – Will be a flood MAR workshop hosted by DWR Nov 5 and 6. Some local folks like Mark Strudley will be presenting.

Gillespie – Scotts Valley Water (SVWD) and Santa Cruz Water have completed the pipeline on La Madrona for the Pipeline 1. They are working on the pump station at the top of La Madrona Drive next. Project is looking to wrap up in Spring of 2026. SVWD has drilled the Graceway well as a component of this project. They are moving on with the equipping phase now. Flow was tested at 500 GPM, but be able to reach 600. Depth is 970 feet.

E. STAFF REPORTS AND ANNOUNCEMENTS

Opportunity for staff to provide brief updates

- Water Resources staff have been interviewing the large water agencies to add that content to the Small Water System Interconnection and Consolidation Guidebook.
- On September 29th, Sierra attended one day of the CalMutuals meeting. She presented on the consolidation feasibility assessment, and guidebook development. Was invited to attend to panel with a person from LA County to discuss how counties can support small systems.
- Sierra will be presenting at Western Groundwater Congress on the Well Ordinance Update in early October

F. NEW BUSINESS

G. UNFINISHED BUSINESS and UPDATES

1. Update on Measure Q Planning and Implementation

Office or Response, Recovery, and Resiliency (OR3) Director David Reid will discuss the status of Measure Q, the Vision Plan, and the Grants program.

David Reid, Director of OR3 provided background on Measure Q. The Measure requires a Vision Plan, which is to be finalized on October 15th. The draft final Vision Plan will be reviewed and voted upon next week.

The Vision Plan covers 6 thematic areas of work being done in the County. Work in Water Resource Protection, Wildfire risk reduction, Parks, Coastal protections, agricultural and working lands protections, Wildlife and habitat protections.

Out of the 6 areas, grant making is being prioritized in three areas: Water Resource Protection, Wildflire and Forest Health, and Parks access. Other topics, particularly Wildlife habitat, were also brought up, but they felt that they were covered to a large extent by the work that would be done with Forest Health and Water. The Measure language also promotes leveraging additional funding such as state and federal grants, looking for multi-benefit projects, and equity. Other priorities include trying to be shovel ready. A lot of the conversations have been around the Vision Plan and Grant Program. The goal is to have the grant program Request for Proposals (RFP) out in late 2025 or early 2026. Still engaging stakeholders. The grant program will be iterative, it will not be perfect in year 1, but they can learn from recipients to improve it.

Stil determining things like how long the grant cycle should be. They want to get money out the door in early so taxpayers can see their dollars at work.

Cheap - Concerned that the Measure wants to be all things to all people, is the grant decision process public?
Director Reid - The two implementing agencies – Parks and OR3 – want a transparent process in terms of the applications and scoring criteria. It will be interesting to see how oversubscribed it is.

Gillespie – Doesn't seem like there is much of an avenue for the water districts to apply for those funds.
Director Reid – Horizontal infrastructure is expensive, only $1.2 mill per year which is not a lot to do things at a utility scale. It seems unlikely that utility infrastructure projects will score high. There is nothing in the language that prohibits agencies applying for that fund. Things like fire hardening around facilities could get funding.

Lockwood – during the stakeholder engagement period they did a good job pulling in information. Definitely appreciated the opportunity to weigh in.
Director Reid – there is a collaorative opportunity to think of ways to leverage funds and collaborate. Hopefully we can get money out the door for good things.

Lockwood – The Department of Conservation has given money for multi-benefit land repurposing and the collaboration helped get the already planned projects implemented. Hope that it helps get more funding in the future under Prop 4. For smaller projects that could be the future of land repurposing, there could be good crossover with these grant funds.

Director Reid encourages the water district staff to engage with their cities as they also get allocated funding. Climate resilience could be a priority for the cities.

In terms of the cities, Santa Cruz is allocated some of their $450k per year to hire two staff members. One parks maintenance worker and one urban forester that will do urban forestry and watershed lands work. Because of the budget cycle, the other cities should have made their plans for at least the first round of funds.

Public comment: Measure Q funding includes $300k for administration (one full, one partial FTE). June 10th Board of Supervisors agenda approved the administrative budget associated with Measure Q.

Director Reid: Parks and OR3 are the administrators, the person who is hired will be the Measure Q analyst and there will also be an external reviewer. They will do a first pass at scoring them. The staff and third-party reviewer will make a first pass and then refer to COAB

Eligibility is a non-profit or agency. Could still have a model with a fiscal sponsor.

2. Multi-Jurisdictional Hazard Mitigation Planning Process
   Clare Peabody from OR3 will discuss the results of the Multi-Jurisdictional Hazard Mitigation Planning process.
      i. Attachments: The website has a lot of content, and the draft MJHMP will be publicly posted there by September 29th: https://mitigatehazards.com/santacruzmjhmp/
      ii. Powerpoint presentation

This is the County's first multi-jurisdictional hazard mitigation plan. It's a 5-year planning process required by FEMA. It can lead to grants for proactive investment to reduce repetitive loss over time. The Plans are nationally standardized.

The process included 10 participating agencies including 3 cities and several water agencies. They went through the planning exercise together. Received FEMA funding to do this planning effort.

On a local level, the exercise to bring together fire districts and water districts in the same room led to some good conversations. The draft is available through October 19th. The plans are very long and have to meet specific requirements in terms of vulnerability assessments. The Executive summary provides a good overview. Had 8 or 9 hazards profiled.

For a small county like ours, we cannot afford to recover from a disaster on our own. If the Federal Government doesn't help us, we will repair fewer things and slower. Not every priority in the plan is going to be federally fundable through FEMA but they are still reflected in the Plan as priorities, which may help with funding for other grant priorities.

City of Santa Cruz had done a hazard mitigation plan on their own and was just starting their 5-year implementation cycle. San Lorenzo Valley Water District also did their own hazard mitigation process. Both had representation at the County planning process.

Public Comment: thinks the county should use komodo as a fire suppressant, it's a non-toxic alternative. Concerned in particular around the watershed around loch Lomond.

H. <u>CORRESPONDENCE</u>

None

I. <u>BOARD OF SUPERVISORS ACTION ON ITEMS AFFECTING WATER:</u>

None

J. <u>ITEMS OF INTEREST</u>

No Discussion

K. <u>AGENDA ITEMS FOR FUTURE MEETINGS</u>

Annual Water Status Report
LAMP updates
HS Gov Tech implementation
Guidebook update
AI Well Log work

L. <u>ADJOURNMENT 5:46</u>

County of Santa Cruz

Water Advisory Commission

701 Ocean Street, Room 312, Santa Cruz, CA 95060
(831) 454-2022 TDD/TTY -Call 711 www.scceh.com
EnvironmentalHealth@santacruzcounty.us

Subject: December 3, 2025 Water Advisory Commission Consent Agenda

Title: Groundwater Sustainability Agency Updates

Background

There are three groundwater basins in the County subject to the Sustainable Groundwater Management Act. The following updates come from the Groundwater Sustainability Agencies tasked with managing and monitoring those basins.

Pajaro Valley Water Management Agency

- Grant Funded Projects
    - Department of Conservation (DoC) Multibenefit Land Repurposing Grant, $8.89 million: Staff and consultants continue to meet monthly with DoC staff and the Statewide Support Entity (SSE); staff and consultant have formed a steering committee, the first meeting of which is scheduled to occur on December 11, 2025.
    - Department of Water Resources (DWR) Watershed Resilience Pilot Grant, $2 million: Staff and consultants hosted the fourth Pajaro River Watershed Resilience Workshop, focused on "Prioritizing Adaptation Strategies," on November 6, 2025, at Paicines Ranch; staff attended a DWR hosted "Cohort" meeting on November 14, 2025, to discuss the status of the pilot program; Staff submitted Invoice No. 5 in the amount of $414,421 to DWR on September 4, 2025 and is preparing Invoice No. 6 for submission by November 29, 2025.
- College Lake Integrated Resources Management Project
    - Water Treatment Plant & Intake Facility
        - Work continues on the Intake Facility and Water Treatment Plant; commissioning meetings are ongoing. PV Water's Operations Team completed the 7-day operational test in October; the Contractor continues to work on dialing in the treatment process.
    - Treated Water Pipeline
        - Project closeout activities are ongoing.

  - Environmental: Monitoring continues per permit requirements
  - Adaptive Management Plan (AMP): Fall 2025 vegetation surveys have been partially completed (some transects remain submerged) and Watsonville Wetlands Watch has started removing invasive species and vegetation management as per the Adaptive Management Plan.
  - Outreach Activities: Website and social media updates are ongoing.
- Watsonville Slough System Managed Aquifer Recharge & Recovery Projects
  - Permitting: Work on permit applications continues with staff and consultants corresponding with permitting agencies.
  - Environmental: The Board approved the addendum to the Environmental Impact Report on March 19, 2025.
  - Outreach: Communications are ongoing; staff and the support team continue to meet with property owners.
  - Property Rights: Four of eleven properties have closed escrow, others are pending.
- Sustainable Groundwater Management Act Well Monitoring Network Expansion
  - Permitting: Staff is working with a consultant to prepare a scope of work to support the preparation and submission of a coastal development permit for two wells in the coastal zone.
  - Property Rights: Staff is engaged in discussions with property owners for proposed wells that would be located on privately owned land.

    Environmental: Staff is engaged with consultants to prepare categorical exemptions in compliance with the California Environmental Quality Act.

Santa Cruz Mid-County Groundwater Agency

- The Agency is:
  - Continuing an evaluation of funding options for expenses associated with Sustainable Groundwater Management Act (SGMA) compliance. A draft of the funding options evaluation is expected in December 2025.
  - Considering applicants for two private well owner (PWO) representatives to fill positions that expire at the end of 2025. Six applicants were interviewed by an agency sub-committee. A recommendation for PWO representatives that will begin 4-year terms in 2026 is expected to be considered for Board approval in December 2025.

    - Continuing a study of increasing chloride concentrations in coastal monitoring wells in the Seascape area. Results of the study are expected in early 2026.
- The Agency held a coordination meeting with County Site Mitigation Program staff on November 12th to explore ways to better communicate on emerging groundwater contamination issues. The Agency is proposing an annual coordination meeting to ensure regular communications are occurring.
- The next regular meeting of the Agency is on December 11, 2025, at 6:00 pm.

Santa Margarita Groundwater Agency

- The Agency is:
- Seeking a private well owner (PWO) alternate representative position is currently vacant. More information is available at smgwa.org. Applications are due by December 15, 2025.
- Conducting a comprehensive periodic evaluation of the basin Groundwater Sustainability Plan (GSP). The Agency's consulting team indicated that addressing the Department of Water Resources' recommended corrective actions would likely require modifying sustainable management criteria for groundwater elevations and groundwater quality to a degree that a GSP amendment would be needed. The Agency will make a final decision on proceeding on a GSP amendment at its February 2026 meeting.
- The Agency held a coordination meeting with County Site Mitigation Program staff on November 12th to explore ways to better communicate on emerging groundwater contamination issues. The Agency is proposing an annual coordination meeting to ensure regular communications are occurring.
- The Agency participated in the 24th annual Environmental Town Hall on November 8th in Felton hosted by the Valley Women's Club.
- The next regular meeting of the Agency is on February 26, 2026, at 6:00 pm.

By: Sierra Ryan, Water Resources Program Manager with information from Rob Swartz and Brian Lockwood

County of Santa Cruz

Water Advisory Commission

701 Ocean Street, Room 312, Santa Cruz, CA 95060
(831) 454-2022 TDD/TTY -Call 711 www.scceh.com
EnvironmentalHealth@santacruzcounty.us

Subject: December 3, 2025, Water Advisory Commission

Title: Drought Response & Outreach Plan (DROP) Update

Background:
On September 23, 2021, Senate Bill (SB) 552 was signed into law. SB 552 requires that "a county shall establish a standing county drought and water shortage task force to facilitate drought and water shortage preparedness for state small water systems and domestic wells within the county's jurisdiction". The Water Advisory Commission voted to adopt the responsibility for implementing SB 552 and receives regular updates on the progress of implementation.

Updates:

- There are currently 200 applicants that have enrolled in the Regional Waterboards free well testing program. 177 wells have been tested and 23 are on the waiting list, down from 82 in October.
- Since the October 1st meeting, results were received for 44 wells and 9 had an exceedance of a drinking water standard.
    - One well with an exceedance for Hexavalent Chromium had two served households enroll in county services. The POU treatment installer is currently coordinating with those households.
- Staff provided guidance to several well-owners with E. coli exceedances and detections of Total Coliforms.
- County staff continues to work towards installing POU treatment systems at the Rountree Facility in South County, which is in exceedance for PFOA. Staff hope to have the units installed before the end of 2025.
- Sierra Ryan continues to represent interests of local government at the State Drought Response Interagency Partnership (DRIP) Collaborative.

By: Sean Abbey
Water Quality Specialist III

County of Santa Cruz

HEALTH SERVICES AGENCY
ENVIRONMENTAL HEALTH DIVISION
Water Resources Program
701 Ocean Street room 312, Santa Cruz, CA, 95060
Phone:(831) 454-7519 Fax:(831) 454-4770

AI-OCR Well Layer Update Project

DRAFT FINAL REPORT

Executive Summary:

Project Overview

The AI-OCR Well Layer Update Project modernized the County of Santa Cruz's GIS well layer—a foundational dataset for groundwater and water-resource management. Funded through the 2023 County Drought Resilience Planning Grant, the project applied AI-powered optical character recognition (OCR), custom Python scripts, GIS workflows, and structured quality-assurance (QA) methods to extract, clean, and integrate data from over 17,000 historical well completion files (representing approximately 10,000-11,000 unique wells). Many of these reports were decades-old scanned PDFs with handwritten or inconsistent entries, and the project's automated workflows brought them into a standardized, GIS-ready format to support more accurate analysis, planning, and long-term resilience.

Objectives

- Improve completeness and accuracy of historical well records.
- Standardize key fields including well ID, well type, depth, water levels, APN, and screen intervals.
- Align County, OCR derived, and Department of Water Resources (DWR) datasets for GIS integration.
- Reduce manual data-entry workload and significantly improve geocoding accuracy.

Approach

The project combined a custom-trained Microsoft Azure Document Intelligence model, six purpose-built Python scripts, fuzzy matching and aliasing logic, and multi-layered QA workflows to digitize, clean, classify, and integrate over 17,000 scanned well completion reports. ArcGIS-based workflows then geocoded wells to the highest achievable accuracy, with confidence flags for transparency.

Key Achievements

- **Records Processed:** 17,672 well completion reports (7,046 County-held; 10,626 DWR) processed in under one day—saving thousands of hours of manual effort.
- **Unique Wells Identified:** Created the most complete countywide inventory to date: 10,033 unique, deduplicated wells, including 713 previously unrecorded wells discovered in County archives.
- **Dataset Alignment:** Reconstructed 95% of DWR entries directly from scanned originals.
- **Attribute Completeness:**
    - OCR workflows filled 78% of missing drill depths.
    - Improved extraction of static water level, depths, yield, screen intervals, drilling methods, and locations.
    - Classified wells into production (6,092), destruction (2,193), monitoring, and additional types.
- **Geocoding Improvements:**
    - 91% of wells geocoded beyond PLSS centroid placement.
    - 5,799 used precise County-verified coordinates;
    - 3,288 mapped to parcel centroids;
    - 898 mapped using PLSS or manual fallback.
    - New “Exact” vs. “Approximate” vs “PLSS” confidence flags improve spatial transparency.

- **Duplicate Resolution:** Resolved 66 duplicate IDs and more than 6,000 redundant scans, improving dataset quality.

Quality Assurance Highlights

- Approximately 15% of records were flagged for review based on issues such as:
    - Conflicting depth measurements
    - Well type inconsistencies
    - Low-confidence APN/address matches
- Red-highlighted AI-derived fields, confidence scoring, and structured review queues ensured efficient and reliable manual verification.

Outcomes & Impact

The project produced the most accurate, complete, and spatially reliable well layer ever assembled for Santa Cruz County. Published to GISWeb in December 2025, the dataset now supports:

- Drought resilience planning
- Groundwater monitoring and basin assessment
- Well permitting and oversight
- Long-term water resource protection

The workflow is reproducible, scalable, and statewide-ready, offering a model for other counties and agencies seeking to modernize archival water datasets.

Conclusion

By integrating AI-OCR, Python automation, QA, and GIS, the County transformed decades of handwritten, inconsistent, and poorly geocoded records into a unified, high-confidence dataset. The project dramatically reduced manual workload, increased spatial and attribute completeness, and demonstrated how modern AI tools can convert legacy water records into actionable information for sustainable groundwater management.

The remainder of this report expands on these findings, outlining the methods, data processing steps, QA approach, and final outcomes in greater detail

Introduction

The AI-OCR Well Layer Update Project modernizes the County of Santa Cruz's GIS well layer, a critical tool for water resource management within the County, by automating data extraction and matching from well completion reports. Funded through the 2023 County Drought Resilience Planning Grant, the project combines AI-powered OCR, custom Python scripting, QA procedures, and GIS integration to significantly improve data accuracy and completeness. Initiated in 2024, this report outlines the project's methodology, including model training, six core data processing scripts, QA workflows, and GIS updates, as well as the resulting improvements.

Background and Challenge

Since its creation in 2005, through the merger of County and partner agency datasets, the County of Santa Cruz's GIS well layer has faced persistent issues with incomplete and inconsistent information. A 2009 integration with the DWR well layer provided some improvements; however, more than 75% of the 11,700 wells still lacked well type classifications, and a significant number of records were missing altogether. In addition, several wells were entered into the dataset more than once. These gaps and inconsistencies have limited the Water Resources Division and other agencies' ability to accurately identify production wells and assess their impacts.

The most recent well layer update project was designed to address these data gaps, and more, by applying AI-powered OCR to scanned well completion reports. Using custom scripts, the project extracts, cleans, and standardizes key fields—such as well ID, depth, water levels, APN, screen intervals, and address—and employs fuzzy matching to align these records with the existing DWR's well database This process enhances data completeness and consistency while preserving the County's existing spatial accuracy.

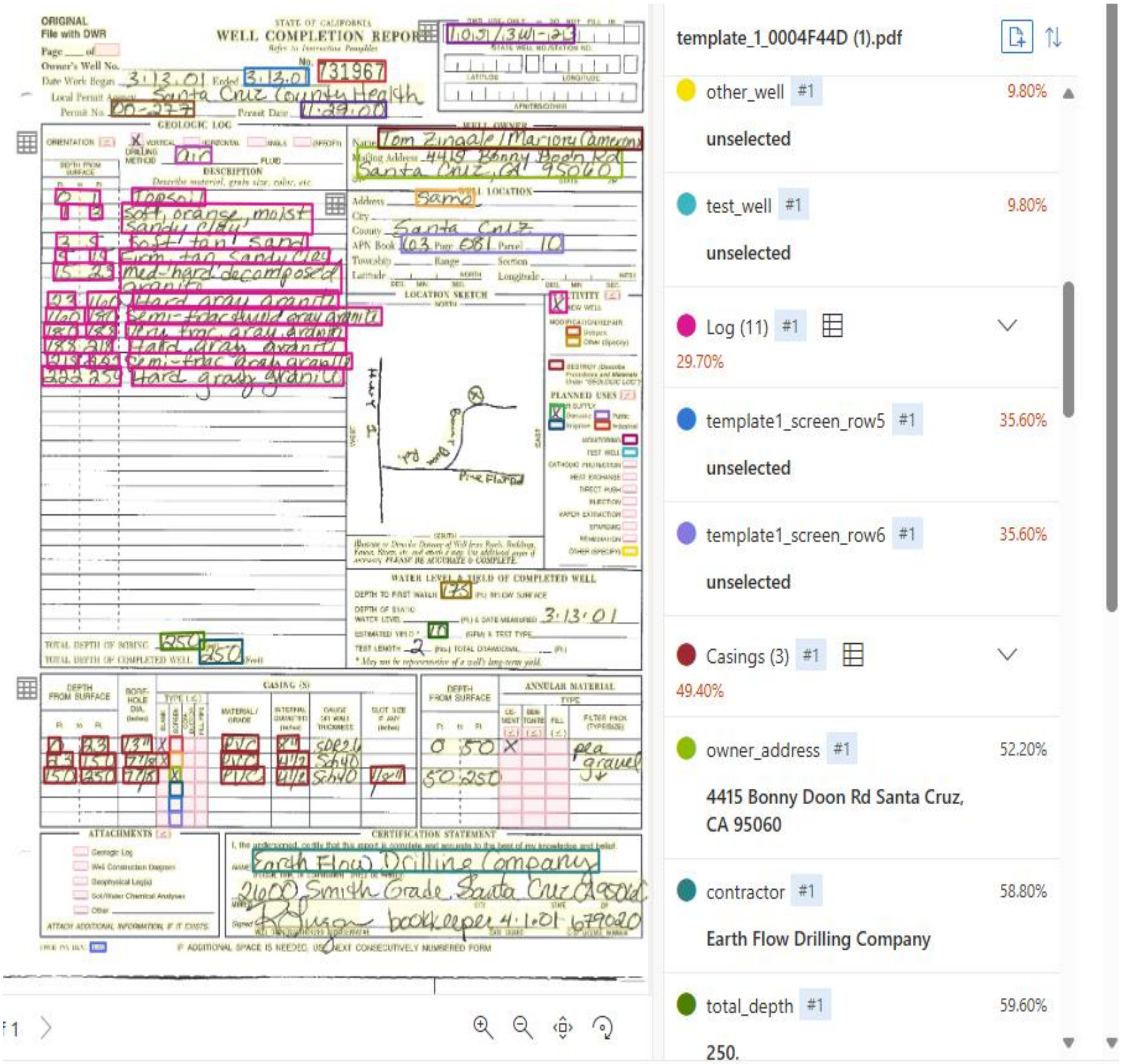
Methodology

The project employed Microsoft Azure Document Intelligence for OCR, supported by six custom Python scripts for data processing, continuous QA to validate outputs, and GIS integration to update and visualize the results. The overall workflow is summarized below, with a visual overview of the script sequence provided in the attached diagram (*Script Sequence Workflow,* Appendix Page 1).

Model Training

A custom Azure Document Intelligence model was trained using nine standardized DWR well completion report templates, dating back to the late 1940s to early 1950s. The model was designed to extract key fields such as well ID, total depth, water levels, completion dates, yields, well type, seals, lithology, screen intervals, APNs and other locational information. It was optimized to handle a wide range of report formats—including handwritten entries and low-resolution scans—ensuring consistent and accurate data extraction across standard California DWR reports.

An example of the model's digitized output, displaying color-coded bounding boxes for extracted data and a right-hand panel listing some of the parsed fields, is provided in the following Azure Document Intelligence screenshot:



Data Processing Scripts

Six custom Python scripts were developed to process, clean, classify, and enhance the data extracted through OCR. The scripts leverage AI-assisted coding tools (e.g., ChatGPT, Copilot) to accelerate prototyping and debugging. They utilize libraries including pandas, fitz, tempfile, numpy, os, json, logging, re, collections, rapidfuzz, azure-ai-formrecognizer, and openpyxl.

The scripts work sequentially to extract structured data from scanned well reports, clean and standardize fields, classify well types, match records across multiple datasets, integrate county records, and highlight AI-derived fields for quality control. The workflow for each script is summarized below.

1. OCR Extraction Script

Purpose: Convert raw PDF well completion reports into structured JSON data suitable for downstream processing.

Overview:
Processes scanned PDF well reports using Azure Document Intelligence for OCR, with fitz handling PDF manipulation and compression.

Key Functions:

- Compresses large PDFs (>4 MB) to ensure compatibility with Azure OCR.
- Extracts fields such as well_id, depth, APN, and Casings using the custom-trained Azure model.
- Captures page- and line-level content in nested JSON dictionaries.
- Logs extraction errors and skips failed PDFs.
- Outputs JSON files to specified directory.

2. Data Extraction and Classification Script

Purpose: Transform JSON outputs from OCR into structured Excel tables, with keyword detection to support well type classification.

Overview:
Processes JSON outputs from the OCR Extraction Script, extracting key fields and identifying destruction-related keywords. Outputs two Excel sheets: *Main Data* and *Casings*.

Key Functions:

- Parses JSON dictionaries, extracting fields like well_id, destruction_well, address, and APN.
- Detects destruction-related keywords (e.g., “abandonment,” “destroy,” “decommission”) using fuzzywuzzy (token_sort_ratio threshold 85), including common misspellings.
- Converts Casings lists into DataFrame rows with traceability via source_file and row_index.

- Skips invalid JSON structures and logs errors.
- Generates Excel file output to specified directory.

3. Data Cleaning and Matching Script

Purpose: Produce a cleaned, deduplicated dataset with standardized fields, enhanced classifications, and matched geospatial data.

Overview:
Processes the Excel output from extraction to clean, standardize, and enrich data. Handles OCR errors, refines well classifications, extracts perforation intervals, and prepares addresses/APNs for matching.

Key Functions:

- Cleans well_id using regex, handling OCR misreads.
- Classifes templates (1–9 or Unknown) via revision-based signatures.
- Refines destruction classifications using baseline keyword lists.
- Extracts perforation intervals and validates top/bottom values.
- Validates the values of Total Completed Depth and Total Drill Depth, ensuring that Total Completed Depth is always greater than Total Drill Depth.
- Detects duplicates using weighted scoring and retains the most complete record.
- Cleans addresses and other fields (e.g. City, Driller Name, Fluid Type, etc.) via dictionary-based aliases and regex;
- Searches for and extracts APNs from the document (which may be located in multiple places within the document).
- Derives Type of Work hierarchically from fields like keyword_destruction_well or new_well.
- Handles invalid numeric fields (NaN) and missing data.
- Generates Excel file output to specified directory.

4. Data Matching and Merging Script

Purpose: Align OCR-derived well records with the DWR database, enhance APN and address accuracy, and produce a comprehensive matched dataset.

Overview:
Integrates cleaned OCR data with DWR records and optionally checks Monterey

County datasets, generating multiple output files for matched and unmatched records.

Key Functions:

- Cleans well IDs and normalizes APNs (handling 8- and 9-digit formats).
- Performs primary matching on well IDs, validated by depth or water level.
- Multiple matches are resolved with a weighted scoring algorithm
- Refines APNs and addresses using parcel validation and fuzzy matching techniques.
- Validates DWR APNs against a dictionary mapping to parcel addresses. If an APN is invalid, it extracts APNs from the Well Location or other fields if present, or employs rapidfuzz for address matching. Final APNs are formatted to align with the parcel database standards used for GIS integration.
- Checks unmatched OCR records against Monterey dataset.
- Outputs:
    - unmatched_ocr_output.xlsx
    - unmatched_dwr_output.xlsx
    - matched_output.xlsx
- Maintains source tracking and logs matching details for QA.

5. County Database Matching Scripts

Purpose: Integrate county database records with DWR or unmatched OCR data, validating and enriching well records.

Overview:
Two scripts align county records with either DWR data or OCR data, producing matched and unmatched outputs.

Key Functions:

- Normalizes county and source IDs using regex (handling OCR misreads and variations).
- Matches county Log_Number to DWR Legacy Log Number/WCR Number or OCR well_id_cleaned.
- Prevents duplicate matches with sets and logs metadata.
- Produces three Excel outputs per script: matched records, unmatched county records, and unmatched source records.

6. Field Highlighting Script

Purpose: Visually highlight OCR-derived or non-validated fields in the final supplemented dataset for QA.

Overview:
Applies conditional formatting to the supplemented dataset Excel file using openpyxl, marking AI-sourced fields or unmatched values in red font.

Key Functions:

- Loads Excel workbook and maps columns dynamically.
- Highlights final_apn and final_address if match_source indicates non-validated or OCR-derived data.
- Highlights other fields based on a predefined field_map when source = "AI."
- Skips missing columns and logs errors.
- Saves the modified Excel file, overwriting the input for review.

The following image shows an example output of the field-highlighting script, with color-coded AI-generated values:

	A	B	E	H	I	N	P	AI	AK	AL	AM	AN	AS	BE
1	WCR Number	Legacy Log Number	City	Permit Date	Permit Number	Driller Name	Driller License Number	Total Drill Depth	tal Completed Dep	p Of Perforated Inter	m of Perforated In	Casing Diameter	Static Water Level	B118WellUse
92	WCR0153482	E0263622	Scotts Valley	3/19/15	SM-15-002	EXPLORATION GEOSERVICES INC (score=9	484288		30					Monitoring Well
93	WCR1992-018220	361482			92-136	POLLOCK'S DRILLING & PUMP SERVICE	379086	780	760	720	860		350	Domestic
94	WCR0274630	18373				C & N PUMP & WELL CO	68648							Domestic/Public Supply Well
95	WCR0290064	NN				FREEDOM DRILLING	unknown	150	145	66	150	10		
96	WCR0117565	NN	SANTA CRUZ	4/25/06	05-319	PIERCE & SON	unknown	580	220	370	510	12	99	
97	WCR0276070	NN	Watsonville	11/6/15	15-072	LITCHFIELD	unknown	415	362	0	470		60	
98	WCR0251594	327077	FELTON		8910189A	AQUA SCIENCE ENGINEERS INC	487000	83	81	79	81	2	79.4	Monitoring
99	WCR0177731	926182	LOS GATOS	2010-03-09 00:00:00	10-032	GREGG DRILLING & TESTING INC	485165	30	30	20	30	2	23	Monitoring Well
100	WCR0118327	261906	SANTA CRUZ		1928-28 AND 88-261	MAGGIORA BROS DRILLING INC	249957	7	17	7	17	2	9	Monitoring
101	WCR0029130	298476	FELTON			AQUA SCIENCE ENGINEERS INC	487000	60	59.7	57.7	59.7	2		Monitoring
102	WCR0046082	324356	FELTON		8910189A	JEFFREY D MORGAN WATER DEVELOPMEN	283326	334	334	329	334	2	314.2	Monitoring
103	WCR0025311	324357	FELTON		8910189A	PITCHER DRILLING CO	263085	326	305	295	305	1	267.05	Monitoring
104	WCR0277831	200034	S CRUZ			MAGGIORA BROS DRILLING INC	249957	15	14.5	4	14	4	6.5	Monitoring
105	WCR0123199	200037	S CRUZ			MAGGIORA BROS DRILLING INC	249957	14.8	14.8	4.3	14.3	4	6.1	Monitoring
106	WCR0021275	425649	WATSONVILLE	Na	325-25.01	R BARRICK FOR BAYLANDS	374152							Monitoring
107	WCR0315111	381860	WAT		9106	G E WEBER & ASSOCIATES INC	unknown	20	20	10	20	2	4	Monitoring
108	WCR0323923	55603				ROMAN WELL DRILLING	265064	415	365	325	355	12	10	Public Supply
109	WCR0032522	313021			1988-11-01 00:00:00	EARTH FLOW DRILLING	320331		300	220	300	5	240	Domestic
110	WCR0092975	261740			88-08	MAGGIORA BROS DRILLING INC	249957	16	15.1	4.7	14.7	4.5	12	Monitoring
111	WCR2010-011978	E0136967	CAPITOLA			R S I DRILLING	802334	5.5	5.5	4.5	5			Other Well
112	WCR0118334	550821	WATSONVILLE			EXPLORATION GEOSERVICES INC	484288	25	25	15	25	4	16.68	Monitoring
113	WCR0088754	365674		May 22, 1991	1991-05-20 00:00:00	PC EXPLORATION INC	265556							Monitoring
114	WCR2015-007806	E0285609	SANTA CRUZ	9/5/14	SM14-188	ALL WELL ABANDONMENT	848359							
115	WCR0276740	305149	CAPITOLA		88-000295	BAYLAND DRILLING CO	374152	30.5	27.8	8.3	27.8	2	13.7	Monitoring
116	WCR0305688	192402	SANTA CRUZ			ROGERS JOHNSON & ASSOCIATES	unknown	15	14.5	4.5	14.5	2	5.6	Monitoring
117	WCR0188616	192403	SANTA CRUZ			ROGERS JOHNSON & ASSOCIATES	unknown	16	15.5	5.5	15.5	2	5.8	Monitoring
118	WCR0246871	550823	WATSONVILLE			EXPLORATION GEOSERVICES INC	484288	25	25	15	25	2	17.44	Monitoring
119	WCR0304159	313867	WATSONVILLE		88-234	EXPLORATION GEOSERVICES INC	484288	33	30	9	30	2	16	Monitoring
120	WCR2015-007799	E0285622	SANTA CRUZ	9/5/14	SM14-192	ALL WELL ABANDONMENT	848359							
121	WCR0123198	423358	SOQUEL	2/5/91	91-034	WOODWARD DRILLING COMPANY	581639	36	36	21	36	4		Monitoring
122	WCR2010-011979	E0136966	CAPITOLA			R S I DRILLING	802334	10.5	10.5	9.5	10			Other Well
123	WCR2024-003460		Santa Cruz	2023-12-05 00:00:00	SR0015594	CONFLUENCE TECHNICAL SERVICES INC	1035255		8					Other

Sheet1

Note: See the Appendix for the code and detailed function descriptions.

Quality Assurance

The Well Layer Update data processing pipeline incorporates multiple layered quality assurance mechanisms to improve accuracy, reliability, and traceability. These mechanisms operate at the field-cleaning, matching, and validation stages and generate flags, scores, source tracking, and visual cues that guide manual review. Outputs from the Data Cleaning and Matching Script, Data Matching and Merging Script, County Database Matching Scripts, and Field Highlighting Script include dedicated QA columns, confidence metrics, and conditional formatting designed to highlight low-confidence or potentially erroneous values.

The key QA mechanisms are summarized below.

1. **Location Matching Confidence and Source Tracking**

   Purpose: Provide transparent, quantifiable confidence in the final APN and address used for geocoding, enabling targeted manual review of lower-quality locations.

   Implementation:

   - Matching follows a strict hierarchy: exact validated APN → exact validated address → fuzzy validated APN → fuzzy validated address → fallback methods.
   - Every record receives:

     – match_source / OCR_match_source: a human-readable string describing the match method and score (e.g., "ID match → APN validated", "Fuzzy address match (score 78.2%)")

     – Numeric fuzzy scores (0–100) from rapidfuzz when fuzzy address or APN parcel is used.

     – Alias/regex confidence scores when street-name or city dictionaries are applied during address cleaning.

     QA Role: Records with fuzzy scores <85, non-APN-based matches, or low alias/regex scores (<80%) are prioritized for manual review using GIS tools and source PDFs.

2. **AI Source Indicators for OCR-Derived Data**

   Purpose: Distinguish OCR-derived data from DWR or county-validated data.

   Implementation: Scripts track OCR-derived fields using source columns (e.g., apn_source, depth_source) marked "AI." The Field Highlighting Script applies red font to these cells (e.g., final_apn, final_address, seal_cleaned, template_id).

   QA Role: Red highlights visually flags potentially error-prone data, prompting analysts to verify against the original PDFs.

3. **Type Mismatches for Record Consistency**

   Purpose: Detect discrepancies between DWR and OCR-derived well type classifications.

   Implementation: The Record_Type_Mismatch field flags conflicts between DWR B118WellUse and OCR-derived types

   QA Role: Prioritize mismatched records and cross-reference source documents to resolve classification errors.

4. **Destruction Well Cross-Referencing**

   Purpose: Improve destruction well accuracy, especially when handwritten notes override checked boxes (common when a proposed well was abandoned after drilling began).

   Implementation: keyword_destruction_well refines the initial checkbox detection by comparing additional keyword counts against template-specific baselines.

   QA Role: Records where OCR and keyword indicators disagree and are flagged for manual confirmation.

5. **Total Depth Validation Flag (depth_flag / review_depths)**

   Purpose: Ensure consistency between Total Drill Depth and Total Completed Depth and catch common OCR or data-entry errors.

Implementation:

- Validates that Total Drill Depth ≥ Total Completed Depth (with minor tolerance).
- Sets depth_flag = 'review depths' when the relationship is violated or values appear swapped.

6. **Dictionary-Based Alias Cleaning with Regex Confidence Scoring**

   Purpose: Standardize inconsistent or misspelled categorical/text fields (City, Driller Name, Fluid, Test Type, Drilling Method, etc.) while quantifying confidence.

   Implementation:

   - Large alias dictionaries map historical and variant names to standardized values.
   - Two-stage cleaning: (1) dictionary lookup with rapidfuzz token-set ratio, (2) regex pattern fallback if no strong dictionary hit.
   - Each cleaned value receives an alias_match_score (0–100)

Manual Review Process

The QA outputs feeds into a structured review workflow:

- Red-highlighted AI-sourced fields (Field Highlighting Script)
- depth_flag = True or conflicting depth measurements
- Low dictionary/regex alias match scores (<80%) on City, Driller Name, Fluid Type, Drilling Method, etc.
- Record_Type_Mismatch or conflicting destruction indicators
- Fuzzy address/APN match scores <85 (especially non-APN or direct address matches)
- OCR_match_source indicating fallback or low-confidence location methods
- B118WellUse anomalies possibly caused by 'Confidential' stamp interferences

Verification Tools: ArcGIS Pro, original scanned PDFs, county parcel records and historical lists are utilized to update flagged record values.

GIS Workflow

The GIS workflow for the Well Layer Update emphasizes maximizing spatial accuracy of well locations by integrating updated database outputs with county-maintained spatial layers. It combines automated geocoding, join operations, and manual review to produce a high-confidence, GIS-ready dataset. The workflow prioritizes county-verified locations, leverages parcel-based centroids for unmatched (county) APN records, and applies PLSS grid centroids or manual location sketch plotting as fallbacks for records with missing or low-confidence location data.

Geocoding Process

1. Matching with County-Mapped Well Records:
Purpose: Preserve the accuracy of well locations already verified and mapped by county staff.

Implementation:

- Updated database records that match existing county-mapped wells are spatially joined based on Log_Number and well_id_cleaned.
- The county well layer geometry, established over years from DWR report records and maps, is preserved.
- County unique fields such as county_Staff are appended to matched records.

QA Role:

- High-confidence matches inherit county-verified locations, reducing the need for further location verification.

2. APN and Address-Based Geocoding with Parcel Layer

Purpose: Assign the most precise possible locations to wells using validated APNs or addresses, while distinguishing exact from approximate placements.

Implementation:

- Records with a validated APN or address (cross-referenced against the County's Parcel database) are spatially joined to the centroid of the corresponding parcel polygon.
- Each record receives a Location_Accuracy field with one of two values:
  - "Exact" – assigned when the match is based on an exact, validated APN or address match, or inheritance from an existing county-mapped well.
  - "Approximate" – assigned when the match relies on fuzzy APN or address matching with a fuzz score <100 but ≥80–85

<u>3. Fallback to PLSS Grid or Manual Review</u>

<u>Purpose</u>: Provide approximate locations for records that are lacking or have low confidence APN/address data.

<u>Implementation:</u>

- PLSS Grid: Wells are mapped to the centroid of their corresponding PLSS Section grid in the county.
- Manual Review with Location Sketches: Extract well positions (ongoing) from sketches in DWR reports into ArcGIS.
- Records with low (<80) OCR_match_source fuzzy scores (e.g., "Address matched (score 63.64)") are flagged for fallback or manual review processing before mapping to that address or PLSS.

<u>QA Role:</u>

- Wells without confident location data are mapped via PLSS centroids and are subject to future cross-checks with the source file to improve location data manually.

<u>GIS Summary:</u>

This GIS workflow ensures well locations in the Well Layer Update are geocoded with maximum accuracy by:

1. Prioritizing county-verified coordinates.
2. Leveraging parcel-based centroids for direct match or high-confidence APN records.

3. Applying existing DWR PLSS grid mapping for low-confidence or missing data.

Results and Metrics

The project processed 17,672 well completion reports (7,046 county, 10,626 DWR) in batch mode using the developed Python scripts, achieving improvements in data integration, accuracy, and efficiency. The following results and metrics highlight improvements across data processing, matching, geocoding, field supplementation, QA, and GIS integration.

Data Processing

- Total Records Processed: Successfully extracted data from 17,672 files, with nearly all files processed without errors, despite challenges with low-quality scans and illegible handwritten data. Note some of these files ended up not being well completion reports (i.e. e-logs, boring logs, etc.)

Matching Accuracy

- County-to-DWR Matching: 5,683 of the 6,698 county-scanned well completion reports (85%) were successfully matched to corresponding DWR records using a combination of exact well ID matching and fuzzy reconciliation techniques.
- The remaining ~15% (1,015 records) were unmatched due to either the record not existing in the DWR database or poor scan quality or illegible handwriting that prevented accurate well ID extraction.
- As of April 28, 2025, the DWR online system for WCR contained 9,406 well records geographically associated with Santa Cruz County. Around the same date, 10,651 PDF files were downloaded from the DWR website for processing.
- After the initial county-to-DWR matching, 3,723 DWR records remained unsupplemented by county scans (9,406 total DWR records – 5,683 county-matched). Of these, 3,070 were subsequently matched and enriched using OCR data extracted directly from the downloaded DWR-sourced PDFs

(following deduplication in the Data Cleaning and Matching Script). This left only 653 DWR records without OCR-derived supplemental data.

- Unique and Non-Well Records: 713 county-scanned records had no corresponding entry in the DWR database (true new records).

- ~530 county files were identified during QA review as non-well completion reports (e.g., boring logs, plan sets, or electric logs) or were illegible and excluded from final counts.

- The OCR extraction and processing pipeline, when applied to the 10,651 PDFs downloaded from DWR's OSWCR website, successfully recreated 95% of the 9,406 official DWR records for Santa Cruz County using only the digitized documents themselves. When combined with the 713 previously unrecorded wells identified from county-held scans, the project achieved 103% relative completeness compared to the official Santa Cruz County DWR database as of April 2025.

- The updated master database contains 10,033 unique well records. This final dataset is a synthesis of DWR records enriched with AI-OCR-extracted data, 713 previously unrecorded wells identified from county-held scans, and county records that were supplemented and validated using OCR.

Location Accuracy and Improvements

- Geocoding Success: Of the 10,033 wells in the updated dataset, 9,135 (91%) now have geocode locations ranging from exact well locations to approximate well locations—a significant improvement over the previous layer.
  - County-Mapped Wells: 5,799 matched records inherited precise county coordinates through GIS joins.
  - APN-Based Geocoding: 3,288 wells were mapped to parcel centroids, with 80% (2,644) validated as high-confidence (e.g. "APN matched") and the remaining 644 as approximate locations.

    - PLSS/Manual Fallback: 898 wells with low fuzzy-match scores (<80%) or missing location data were mapped to PLSS centroids, flagged for later QA review.
    - The remaining 48 inherited locations from DWR associated with GPS, Other, or Surveyed to Benchmark.

Well Type Improvements

The previous County well database had limited and inconsistently defined well type information. It often wasn't clear whether a well was a production well—or what type of production it served (domestic, irrigation, industrial, or municipal)—and it also did not reliably distinguish destruction wells from 'other' categories such as test, monitoring, or boring wells.

The updated database now classifies each well type based on the information available in the well completion reports, resulting in a more accurate and consistent countywide dataset. The updated counts include 6,092 production wells, 2,193 destruction wells, and 1,747 wells classified as other types.

Field Supplementation

- Critical Fields Filled: The Data Matching and Merging Script supplemented previously blank or unknown DWR fields using OCR and county data:

Field	Records Filled	Total Matched	Percent Filled
Well completion date	513	8,752	5.9%
Total drill depth	6,852	8,754	78.3%
Final well depth	405	8,754	4.6%
Static water level	545	8,754	6.2%
Well yield	1,074	8,754	12.3%
Well use/source	1,788	8,754	20.4%

Field	Records Filled	Total Matched	Percent Filled
Drilling method	896	8,754	10.2%
Screen intervals	670	8,754	7.7%

- Unique Fields: Additional fields (e.g., well seals, first water levels encountered, geologic materials) were extracted. Note: geologic materials were not processed or incorporated in any of the output datasets.

Duplicate Cleanup

- Duplicate Resolution: The Data Cleaning and Matching Script resolved 66 duplicate well IDs and 6000+ duplicate PDF scans, reducing redundancy by ~5% of the total dataset.

Time Savings

- Processing Efficiency: 17,672 reports were processed in under one day, saving significant amounts of time compared to manual extraction and geocoding, equivalent to thousands of hours saved if that would have been required to manually supplement or input the data.
- Script Runtime Breakdown:
    - OCR Extraction Script: ~70% of runtime
    - Post-Processing Scripts: ~30% hours, optimized by batch processing and indexing

Quality Assurance Metrics

QA-Flagged Records: Approximately 15% of all records (2,651 of 17,672) were flagged for manual review. Key flag categories included:

- ~2000 records affected by *"Confidential"* stamp interference (misread numbers, checkbox errors).
- 393 records with conflicting depth measurements (review_depths flag).
- 181 records with well type mismatches (often caused by *"Confidential"* stamp interference).
- 1,857 records with fuzzy match scores below 80% (from OCR_match_source).
- Roughly 5% of supplemented fields using the street-name alias dictionary generated low alias match scores (<80%), requiring manual refinement.

Discussion

The Well Layer Update project represents a transformative advancement in well data management, leveraging new and advanced AI tools to reducing manual effort from thousands of hours to under one day while enhancing the GIS well layer's quality. Key outcomes include:

- High match rates: 95% overall success across OCR records and DWR databse records
- Improved geocoding accuracy: 91% of wells geocoded to a location other than the centroid of the PLSS.
- Field supplementation: Critical fields filled (e.g., 78.3% of total drill depths)

Challenges and Solutions:

- Poor scan quality / illegible handwriting affected numerous reports, mitigated through data cleaning, OCR error logging, and QA.
- Template inconsistencies: Handled by template-specific cleaning logic
- OCR errors: Minimized using a combination of fuzzy matching, dictionary-based aliasing, standardized or pattern-based cleaning routines, and QA-flagged review.

Scalability

- Statewide-Ready Framework**:**

  The California DWR template-based processing, APN integration, and GIS workflows create a scalable framework that can be extended to other counties or applied to statewide DWR datasets.

Potential Enhancements

Model Training & Template Improvements

- Train additional documents to improve value extraction, especially for Template 2.
- Add training for all selection boxes across templates (equipment-type checkboxes, cathodic protection wells, injection wells, etc.) to prevent lesser-used boxes from interfering (common) with more critical selections located nearby.
- Additional training across all known templates to address missed or unrecognized templates.
- Retrain the AI-OCR model to better detect monitoring wells and other specialized well types.
- Refinement of template classifications.
- Improved template-specific logic for screen interval extraction, especially for Template 1.
- Template-specific well ID cleaning, such as enforcing expected ID ranges (e.g., Template 2).

OCR & Document Handling

- Improved OCR performance for low-quality or complex scans.
- Automated detection and removal of “Confidential” stamps.
- Enhanced parsing and handling of PDFs containing multiple well completion reports.
- Add logic to detect and classify documents that are not well completion reports (borings, e-logs, site plans, etc.).

Data Cleaning & Field Logic Enhancements

- Strengthen cleaning algorithms for Total Depth, including:
    - cross-checking Total Drill Depth against the bottom depth in the geologic log,
    - handling “Same” entries,
    - applying fallback logic to populate final depth from geologic log values when needed.
- Add logic and fields for detailed screen interval depths.
- Add lithology post-processing.
- Enhance screen interval logic (e.g., ensure the screen checkbox aligns with slot-size entries in the same row).
- Add a new field to classify wells that were attempted but ultimately dry (Files that contain:“Dry”, “Dry Hole”, “Dry Well”).
- Logic to classify wells with no “Proposed Use” selected as Production if a screen interval exists, with a QA flag to ensure they aren’t monitoring wells.
- More robust handling of duplicate well IDs.

Quality Assurance Enhancements

- Increased QA checks comparing AI-extracted data with DWR records to identify and correct errors
- Add QA checks to flag impossible or inconsistent values, such as:
    - static water level deeper than total drill depth,
    - bottom of perforated interval deeper than total drill depth.
- Additional QA checks to ensure fuzzy-matched low-score addresses remain within previously identified PLSS grids.
- Incorporate confidence scores in JSON outputs to guide QA and validation.

APN, Address, and Location Improvements

- Integrate historical APN datasets to account for subdivisions or APNs that have changed over time.
- Expand location-improvement logic, including using owner name information to fuzzy match against parcel databases and verify whether a corresponding parcel address aligns with the WCR-reported street/location.
- Add logic to split extracted TRS values into Township, Range, and Section fields for new or supplemented records.

Additional Functional Enhancements

- Utilize unredacted reports (when available) to access owner information.

Conclusion

The Well Layer Update project demonstrates the effectiveness of automated data processing and GIS integration in modernizing well record management, as well as the transformative potential of AI-powered OCR software. Recent advancements in AI OCR have significantly improved accuracy—particularly when handling complex layouts, mixed formatting, and handwritten text—making it possible to rapidly digitize and analyze large volumes of historical documentation.

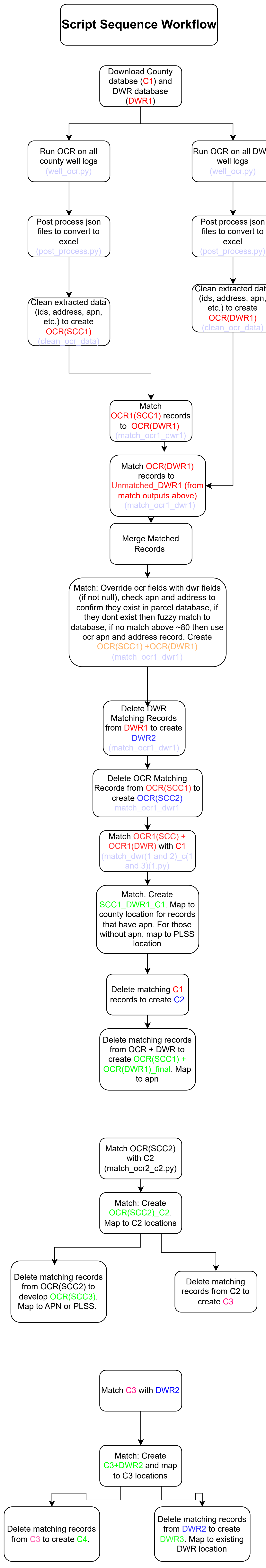
By processing 17,672 well completion reports in under a day, identifying missing records, supplementing key fields, and achieving 91% geocoding accuracy (tied to locations other than PLSS centroids), the project illustrates major gains in efficiency, completeness, and spatial reliability. These outcomes highlight the value of applying AI OCR to historical records across other County divisions, as well as in other counties or organizations with large archival datasets.

As AI technology continues to advance, accuracy and processing speed will further improve, expanding opportunities to modernize legacy datasets and integrate them into contemporary water management and planning workflows.

The updated well layer was integrated into GISWeb in December 2025, supporting drought resilience efforts and enhancing water resource decision-making. The project demonstrates how modern tools can transform previously inaccessible or labor-intensive data into a reliable, comprehensive resource for staff, analysts, and policymakers.

Appendix

Below is the script flowchart workflow as well as the primary scripts, with comments explaining the purpose of key functions.


Script Sequence Workflow
Download County databse (C1) and DWR database (DWR1)
Run OCR on all county well logs (well_ocr.py)
Run OCR on all DWR well logs (well_ocr.py)
Post process json files to convert to excel (post_process.py)
Post process json files to convert to excel (post_process.py)
Clean extracted data (ids, address, apn, etc.) to create OCR(SCC1) (clean_ocr_data)
Clean extracted data (ids, address, apn, etc.) to create OCR(DWR1) (clean_ocr_data)
Match OCR1(SCC1) records to OCR(DWR1) (match_ocr1_dwr1)
Match OCR(DWR1) records to Unmatched_DWR1 (from match outputs above) (match_ocr1_dwr1)
Merge Matched Records
Match: Override ocr fields with dwr fields (if not null), check apn and address to confirm they exist in parcel database, if they dont exist then fuzzy match to database, if no match above ~80 then use ocr apn and address record. Create OCR(SCC1) +OCR(DWR1) (match_ocr1_dwr1)
Delete DWR Matching Records from DWR1 to create DWR2 (match_ocr1_dwr1)
Delete OCR Matching Records from OCR(SCC1) to create OCR(SCC2) match_ocr1_dwr1
Match OCR1(SCC) + OCR1(DWR) with C1 (match_dwr(1 and 2)_c(1 and 3)(1.py)
Match. Create SCC1_DWR1_C1. Map to county location for records that have apn. For those without apn, map to PLSS location
Delete matching C1 records to create C2
Delete matching records from OCR + DWR to create OCR(SCC1) + OCR(DWR1)_final. Map to apn
Match OCR(SCC2) with C2 (match_ocr2_c2.py)
Match: Create OCR(SCC2)_C2. Map to C2 locations
Delete matching records from OCR(SCC2) to develop OCR(SCC3). Map to APN or PLSS.
Delete matching records from C2 to create C3
Match C3 with DWR2
Match: Create C3+DWR2 and map to C3 locations
Delete matching records from C3 to create C4.
Delete matching records from DWR2 to create DWR3. Map to existing DWR location

C:\Well Layer Update\Python\Azure\OCR Extraction Script secure.py

```python
import os
import json
import fitz
import tempfile
from azure.core.credentials import AzureKeyCredential
from azure.ai.formrecognizer import DocumentAnalysisClient

# Azure config
# Set these in your environment before running the script:
#   AZURE_FORMRECOGNIZER_ENDPOINT = https://your-resource.cognitiveservices.azure.com/
#   AZURE_FORMRECOGNIZER_KEY      = your-key-here


endpoint = os.getenv("AZURE_FORMRECOGNIZER_ENDPOINT")
key      = os.getenv("AZURE_FORMRECOGNIZER_KEY")
model_id = os.getenv("AZURE_FORMRECOGNIZER_MODEL_ID", "wcr11")

# Validate that required variables are present
if not endpoint or not key:
    raise EnvironmentError(
        "Please set the environment variables AZURE_FORMRECOGNIZER_ENDPOINT "
        "and AZURE_FORMRECOGNIZER_KEY before running this script."
    )

# Trim any accidental whitespace or trailing slashes
endpoint = endpoint.rstrip("/")

input_folder = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Well Completion Reports\Santa Cruz County\Failed"
output_folder = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\json files"
os.makedirs(output_folder, exist_ok=True)

# Azure client
client = DocumentAnalysisClient(endpoint=endpoint, credential=AzureKeyCredential(key))

#  Compress large image only PDFs
def compress_pdf_with_pymupdf(file_path, max_size_mb=4):
    file_size_mb = os.path.getsize(file_path) / (1024 * 1024)
    is_image_only = True

    try:
        doc = fitz.open(file_path)
        for page in doc:
            if page.get_text().strip():
                is_image_only = False
                break

```

```python
        if file_size_mb <= max_size_mb or not is_image_only:
            return file_path  # Skip compression

        print(f"🔧 Compressing large image-only PDF: {os.path.basename(file_path)}")
        temp_file = tempfile.NamedTemporaryFile(delete=False, suffix=".pdf")
        temp_file.close()  # Close so fitz can write to it
        new_doc = fitz.open()

        for page in doc:
            pix = page.get_pixmap(dpi=100)
            img_bytes = pix.tobytes("jpeg")
            img_doc = fitz.open("jpeg", img_bytes)
            rect = img_doc[0].rect
            pdf_page = new_doc.new_page(width=rect.width, height=rect.height)
            pdf_page.insert_image(rect, stream=img_bytes)

        new_doc.save(temp_file.name)
        new_doc.close()
        return temp_file.name

    except Exception as e:
        print(f"Could not compress PDF ({file_path}): {e}")
        return file_path

# Process each PDF
for dirpath, dirnames, filenames in os.walk(input_folder):
    for filename in filenames:  #iterate over the filenames list
        if filename.lower().endswith(".pdf"):
            original_path = os.path.join(dirpath, filename)  # Use dirpath to construct full path
            pdf_to_process = compress_pdf_with_pymupdf(original_path)

            print(f"\n📄 Processing file: {filename}")

            with open(pdf_to_process, "rb") as f:
                try:
                    poller = client.begin_analyze_document(
                        model_id=model_id,
                        document=f
                    )
                    result = poller.result()

                    result_dict = result.to_dict()
                    json_filename = os.path.splitext(filename)[0] + ".json"
                    json_path = os.path.join(output_folder, json_filename)

                    with open(json_path, "w", encoding="utf-8") as json_file:
                        json.dump(result_dict, json_file, indent=2)

                    print(f"Saved JSON: {json_path}")
```

```
                except Exception as e:
                    print(f"Error processing {filename}: {e}")

            # Clean up temp file if we compressed
            if pdf_to_process != original_path:
                try:
                    os.remove(pdf_to_process)
                    print(f"Deleted temporary file: {os.path.basename(pdf_to_process)}")
                except Exception as e:
                    print(f"Failed to delete temp file: {e}")
```

C:\Well Layer Update\Python\Azure\Data Extraction and Classification Script.py

```python
import os
import json
import pandas as pd
from fuzzywuzzy import fuzz
import re

# Paths for input JSON files and output Excel file
input_folder = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\json files"
output_excel = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\extracted_data_dwr.xlsx"

# Destruction related keywords (including common misspellings)
destruction_keywords = [
    "abandonment", "abandon", "abandone", "abandoment",
    "destroy", "destruction", "distruction", "demolish", "decommision", "decommission", "destroying"
]
destruction_keywords = [kw.lower() for kw in destruction_keywords]

# Extract raw text from JSON content for keyword search
def extract_raw_text(json_content):
    # Try to get the main content field
    raw_text = json_content.get("content", "").lower()
    if not raw_text:  # Fallback to concatenating line content from pages
        raw_text = " ".join(
            line["content"].lower()
            for page in json_content.get("pages", [])
            for line in page.get("lines", [])
            if "content" in line and isinstance(line["content"], str)
        )
    return raw_text

#Count destruction keywords using fuzzy matching to handle OCR errors
def count_destruction_keywords(text, keywords, threshold=85):
    words = re.findall(r'\b\w+\b', text)  # Split into words
    matched = []
    for word in words:
        scores = [fuzz.ratio(word, kw) for kw in keywords]
        max_score = max(scores, default=0)
        if max_score >= threshold:
            best_index = scores.index(max_score)
            best_match = keywords[best_index]
            matched.append((word, best_match))
    # Get unique original words
    unique_matched = sorted(set(word for word, _ in matched))
    return unique_matched

```

```python
# Initialize lists for data storage
main_data = []
casings_data = []

# Process JSON files to extract fields, generate destruction keywords, and structure data
for filename in os.listdir(input_folder):
    if filename.endswith(".json"):
        path = os.path.join(input_folder, filename)
        with open(path, "r", encoding="utf-8") as file:
            try:
                json_content = json.load(file)
                doc = json_content.get("documents", [{}])[0]
                fields = doc.get("fields", {})

                # Extract raw text for keyword search
                raw_text = extract_raw_text(json_content)
                matched_keywords = count_destruction_keywords(raw_text, destruction_keywords)

                # Build keyword string
                keywords_str = ', '.join(matched_keywords) if matched_keywords else ''

                # Main data: all top-level fields except Casings and Log, preserve existing destruction_well
                record = {
                    k: (v.get("content") if isinstance(v, dict) else v)
                    for k, v in fields.items()
                    if k not in ["Log", "Casings"]
                }
                record["source_file"] = filename
                record["additional_destruction_keywords"] = keywords_str

                main_data.append(record)

                # Casings data: parse each entry in the list
                if "Casings" in fields and fields["Casings"].get("value_type") == "list":
                    for idx, casing in enumerate(fields["Casings"]["value"], start=1):
                        row = {
                            "source_file": filename,
                            "row_index": idx
                        }
                        if casing.get("value_type") == "dictionary":
                            for key, val in casing["value"].items():
                                row[key] = val.get("content") if isinstance(val, dict) else val
                            casings_data.append(row)

            except Exception as e:
                print(f"Error processing {filename}: {e}")

# write extracted and classified data to Excel with separate sheets for Main Data and Casings
```

```
with pd.ExcelWriter(output_excel, engine="openpyxl") as writer:
    pd.DataFrame(main_data).to_excel(writer, index=False, sheet_name="Main Data")
    pd.DataFrame(casings_data).to_excel(writer, index=False, sheet_name="Casings")

print(f"Excel file created with 'Main Data' and 'Casings' at: {output_excel}")
```

C:\Well Layer Update\Python\Azure\Data Cleaning and Matching Script.py

```
import pandas as pd
import re
from rapidfuzz import fuzz, process

# === File Paths ===
input_excel = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\extracted_data_SCC_3.xlsx"
parcel_excel = r"C:\Well Layer Update\GIS\data_tables\Parcels.xlsx"
output_excel = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR(SCC1)_3.xlsx"
unique_excel = r"C:\Well Layer Update\Python\Scripts\unique_values.xlsx"

def clean_well_id(raw):
    if pd.isna(raw):
        return ""

    original = str(raw).strip()
    # Convert 'c' or 'C' to 'e' at the start
    if original.lower().startswith('c'):
        original = 'e' + original[1:]
    original = re.sub(r'[oO]', '0', original)
    cleaned = original.replace(" ", "").upper().strip()

    # Handle well IDs starting with E/e
    if original.lower().startswith('e'):
        temp = original.lower().replace('o', '0').upper()
        match = re.match(r"^(E)(\d{7,8})(.*)$", temp)
        if match:
            prefix = match.group(1)
            digits = match.group(2)
            return prefix + digits
        digits_after_e = re.sub(r"\D", "", temp[1:])
        if 6 <= len(digits_after_e) <= 8:
            return "E" + digits_after_e
        digits_match = re.match(r"^(E)(\d+)", temp)
        if digits_match:
            return digits_match.group(1) + digits_match.group(2)
        return temp

    # Preserve WCR-formatted IDs
    if re.match(r"^WCR\d{4}-\d{6}$", cleaned):
        return cleaned

    # Fallback cleanup for non-WCR values
    text = original.lower().replace(",", "").replace("no.", "").replace("/", " ")
    text = re.sub(r"[\s\n\r]+", " ", text).strip()

    # Check for date-like prefix (e.g., MM-DD-YYYY or MM.DD.YYYY)
```

```
    match = re.match(r"^(\d{1,2})[\.-](\d{1,2})[\.-](\d{2,4})(\d{5,7})$", text)
    if match:
        year_str = match.group(3)
        try:
            year = int(year_str)
            if year < 100:
                year += 1900 if year >= 35 else 2000
            if 1935 <= year <= 2025:
                return match.group(4)
        except ValueError:
            pass

    # Remove date-like prefix if present
    prefix_removed = re.sub(r"^\d{1,2}[\.-]\d{1,2}[\.-]\d{2,4}", "", text).strip()
    segments = re.split(r"[\s]+", prefix_removed)

    #Collect all digit sequences (5-8 digits) from segments
    all_segment_digits = []
    for segment in segments:
        segment_digits = re.findall(r"\d{5,8}", segment)
        all_segment_digits.extend(segment_digits)

    if all_segment_digits:
        # If multiple sequences, prefer the last one unless an earlier one i longer
        longest = max(all_segment_digits, key=len)
        last_segment_digits = [d for d in re.findall(r"\d{5,8}", segments[-1])] if segments else []
        if last_segment_digits and len(max(last_segment_digits, key=len, default="")) >= len(longest) - 1:
            return max(last_segment_digits, key=len)
        return longest

    # Global fallback for any remaining digit sequences
    digits = re.findall(r"\d{5,8}", text)
    if digits:
        return max(digits, key=len)

    return ""

# Load OCR Extracted Data
df = pd.read_excel(input_excel, dtype=str).fillna("")

# Load unique values for fuzzy matching
unique_df = pd.read_excel(unique_excel, header=None)
field_names = ['city', 'driller_name', 'drilling_method', 'fluid', 'test_type']
uniques_dict = {}
for idx, field in enumerate(field_names):
    col_data = unique_df[idx].dropna().tolist()
    if col_data:
        uniques_dict[field] = col_data[1:]  # Skip the first row (label)
```

```python

#Map to df column names (lowercase as per input)
col_map = {
    'city': 'city',
    'driller_name': 'contractor',
    'drilling_method': 'drilling_method',
    'fluid': 'fluid',
    'test_type': 'test_type'
}

# Clean each field using fuzzy matching to unique values, appending score for non-exact matches or '(no match)' if no match found
for field, col in col_map.items():
    if col in df.columns:
        def clean_val(val):
            if pd.isna(val) or not str(val).strip():
                return val
            val_str = str(val).strip().lower()
            choices = uniques_dict.get(field, [])
            if not choices:
                return val_str
            choices_lower = [str(c).strip().lower() for c in choices]
            choices_map = dict(zip(choices_lower, choices))
            if val_str in choices_lower:
                return choices_map[val_str]
            result = process.extractOne(val_str, choices_lower, scorer=fuzz.partial_ratio)
            if result:
                match_lower, score, _ = result
                if score >= 70:
                    return f"{choices_map.get(match_lower, match_lower)} (score={score})"
            return f"{val_str} (no match)"
        df[col] = df[col].apply(clean_val)

# Column names
APN_COL = "apn"
ADDR_COL = "address"
FALLBACK_ADDR_COL = "owner_address"
LOC_DESC_COL = "location_details"
TEMPLATE_ID_COL = "revision_cleaned"

# Define template signatures
TEMPLATE_SIGNATURES = {
    "Template 1": ['1197', '790', '0503'],
    "Template 2": ['965', '905', '968', '958', '969', '960', '246354', '354', '554', '154', '334', '384', '954', '304', '254', '454', '568', '934', '87649354', '87028354'],
    "Template 3": ['776', '1286', '1288', '775', '778', '7741'],
    "Template 4": ['46170', '46370'],
    "Template 5": ['115', '118', '113', '119'],
    "Template 6": ['246', '2467'],
    "Template 7": ['263', '253', '2634764'],
```

```python
    "Template 8": ['188'],
    "Template 9": ['1200', '12008', '112006', '12006', '1200', '12000', '12009', '10008', '40008']
}

DATE_SIGNATURES = {
    "Template 8": ['12/19/2017'],
    "Template 9": ['1/2006', '1/2008', '1/2000', '1/2005']
}

flat_signature_map = [(val, template) for template, vals in TEMPLATE_SIGNATURES.items() for val in vals]

flat_date_signature_map = [(val, template) for template, vals in DATE_SIGNATURES.items() for val in vals]

def preprocess_apn(apn):
    apn = str(apn).strip()
    if '\n' in apn:
        parts = [p.strip() for p in apn.split('\n') if p.strip()]
        if len(parts) == 3 and all(p.isdigit() for p in parts):
            first, second, third = parts
            if len(third) > len(second):
                second, third = third, second
            return '-'.join([first, second, third])
    return apn

def clean_apn_format(apn):
    if not isinstance(apn, str):
        return ""

    apn = preprocess_apn(apn)

    # Remove all non-digit characters
    digits = re.sub(r"\D", "", apn)

    # Expecting 8 digits (3+3+2 format)
    if len(digits) == 8:
        return f"{digits[:3]}-{digits[3:6]}-{digits[6:]}"

    # Sometimes 7-digit (e.g., missing leading 0 in first section)
    elif len(digits) == 7:
        return f"{digits[:2].zfill(3)}-{digits[2:5]}-{digits[5:].zfill(2)}"

    # Other cases invalid
    return ""

def clean_revision(value):
    if pd.isna(value) or not str(value).strip():
        return ""
    value = str(value).strip()
```

```
    # Check for date-like patterns
    date_patterns = [
        r"^\d{1,2}/\d{1,2}/\d{4}$",  # MM/DD/YYYY or M/D/YYYY
        r"^\d{1,2}-\d{1,2}-\d{4}$",  # MM-DD-YYYY or M-D-YYYY
        r"^\d{1,2}/\d{4}$",          # M/YYYY or MM/YYYY
        r"^\d{1,2}-\d{4}$"           # M-YYYY or MM-YYYY
    ]
    for pattern in date_patterns:
        if re.match(pattern, value):
            return value  # Return original string if it's a date
    # Otherwise, strip all non-digit characters and newlines
    digits = re.sub(r"\D", "", value)
    return digits

def identify_template(row):
    text = row.get(TEMPLATE_ID_COL, "")
    if not isinstance(text, str) or not text.strip():
        return "Unknown"
    # First check for date signatures (exact match)
    for val, template in flat_date_signature_map:
        if text == val:
            return template
    # Then check for numeric signatures (fuzzy match)
    match, score, _ = process.extractOne(
        text.strip(), [val for val, _ in flat_signature_map], scorer=fuzz.token_sort_ratio
    )
    if score >= 80:
        for val, template in flat_signature_map:
            if match == val:
                return template
    return "Unknown"

#Create revision_cleaned column before template identification
df["revision_cleaned"] = df["revision"].apply(clean_revision)
df["template_id"] = df.apply(identify_template, axis=1)

# Known baseline occurrences per template
known_baselines = {
    "Template 1": 1,  # DESTROY- (1)
    "Template 2": 2,  # destruction- (1) + Destroying- (1)
    "Template 3": 2,  # Destruction (1) + destruction- (1)
    "Template 6": 1,  # abandoned (1)
    "Template 9": 1,  # DESTROY- (1)
    "Template 4": 0,
    "Template 5": 0,
    "Template 7": 0,
    "Template 8": 0,
    "Unknown": 0
}

```

```python
# Baseline keywords per template (lowercase)
baseline_keywords = {
    "Template 1": ['destroy'],
    "Template 2": ['destruction', 'destroying'],
    "Template 3": ['destruction'],
    "Template 6": ['abandoned'],
    "Template 9": ['destroy'],
    "Template 4": [],
    "Template 5": [],
    "Template 7": [],
    "Template 8": [],
    "Unknown": []
}

# Function to process additional_destruction_keywords and set keyword_destruction_well
def enhance_destruction_detection(row):
    template = row['template_id']
    keywords_str = row.get('additional_destruction_keywords', '')
    baseline_set = set(baseline_keywords.get(template, []))
    destruction_well = row.get('destruction_well', ':unselected:').strip()

    if not keywords_str:
        row['additional_destruction_keywords'] = ''
        row['keyword_destruction_well'] = destruction_well
        return row

    # Split into list, strip spaces and lowercase for comparison
    keyword_list = [kw.strip().lower() for kw in keywords_str.split(',')]

    # Filter out 'instruction' or items starting with 'instruction'
    filtered_list = [kw for kw in keyword_list if not kw.startswith('instruction')]

    # Mark keywords based on template
    marked_list = []
    if template == "Unknown":
        marked_list = [f"{kw} (unknown)" for kw in filtered_list]
    else:
        for kw in filtered_list:
            if kw in baseline_set:
                marked_list.append(f"{kw} (baseline)")
            else:
                marked_list.append(kw)

    # Update the column to cleaned/marked string
    row['additional_destruction_keywords'] = ', '.join(marked_list)

    # Set keyword_destruction_well: default to destruction_well
    row['keyword_destruction_well'] = destruction_well

```

```
    # For known templates, override to :selected: if keyword count exceeds baseline and destruction_well is :unselected:
    if template != "Unknown":
        total_keyword_count = len(filtered_list)
        baseline_count = known_baselines.get(template, 0)
        if total_keyword_count > baseline_count and destruction_well == ":unselected:":
            row['keyword_destruction_well'] = ':selected:'
            # Unset other fields
            for field in ['deepen_well', 'recondition_other.', 'domestic_well', 'irrigation_well',
                          'public_well', 'monitoring_well', 'industrial_well', 'other_well',
                          'new_well', 'test_well']:
                row[field] = ':unselected:'

    return row

# Apply the enhancement after template_id is set
df = df.apply(enhance_destruction_detection, axis=1)

# Normalize other_well_no for Template 6 and merge duplicates
def normalize_other_well_no(val):
    if not isinstance(val, str):
        return ""
    return re.sub(r"[^\w]", "", val).upper()  # Remove all non-alphanumeric characters and uppercase

# Only process Template 6
template6_df = df[df["template_id"] == "Template 6"].copy()
template6_df["other_well_no_normalized"] = template6_df["other_well_no"].apply(normalize_other_well_no)

# Group by normalized other_well_no and merge fields
merged_template6 = (
    template6_df.groupby("other_well_no_normalized", as_index=False)
    .agg(lambda x: next((i for i in x if i.strip()), "") if x.dtype == "O" else x.iloc[0])
)

# Drop duplicates in original df
df = df[df["template_id"] != "Template 6"]

# Append merged template6 records back
df = pd.concat([df, merged_template6], ignore_index=True)

# Load Casings Sheet
casings_df = pd.read_excel(input_excel, sheet_name="Casings", dtype=str).fillna("")
casings_df["row_index"] = casings_df["row_index"].astype(int)
casings_df["top"] = casings_df["top"].astype(str)
casings_df["bottom"] = casings_df["bottom"].astype(str)

def clean_to_digits(value, is_top=False):
    if pd.isna(value) or not str(value).strip():
```

```
        return None
    # Extract the first sequence of digits before a non-digit character
    match = re.match(r"^\d+", str(value).strip())
    if match:
        cleaned = match.group(0)
        if not cleaned.isdigit():
            return None
        result = int(cleaned)
        # For top values > 999 ending with '0', strip the trailing '0'
        if is_top and result > 999 and cleaned.endswith('0'):
            result = int(cleaned[:-1])  # Remove the last digit ('0')
        return result
    return None

def extract_perforation_fields(row):
    template = row.get("template_id")
    source_file = row.get("source_file", "").strip()

    if template == "Template 1":
        selected_rows = []
        for i in range(1, 7):
            colname = f"template1_screen_row{i}"
            if row.get(colname, "").strip() == ":selected:":
                selected_rows.append(i)

        if not selected_rows:
            return pd.Series({"perforation_interval": "", "top_perforations": "",
"bottom_perforations": ""})

        matches = casings_df[
            (casings_df["source_file"] == source_file) &
            (casings_df["row_index"].isin(selected_rows))
        ].copy()

        if matches.empty:
            return pd.Series({"perforation_interval": "", "top_perforations": "",
"bottom_perforations": ""})

        matches["top_cleaned"] = matches["top"].apply(lambda x: clean_to_digits(x,
is_top=True))
        matches["bottom_cleaned"] = matches["bottom"].apply(clean_to_digits)

        top = matches["top_cleaned"].min()
        bottom = matches["bottom_cleaned"].max()

        if pd.isna(top) or pd.isna(bottom):
            return pd.Series({"perforation_interval": "", "top_perforations": "",
"bottom_perforations": ""})

        # Check if bottom is less than top
        if bottom < top:
```

```
            return pd.Series({
                "perforation_interval": "",
                "top_perforations": top,
                "bottom_perforations": ""
            })

        return pd.Series({
            "perforation_interval": f"{top} - {bottom}",
            "top_perforations": top,
            "bottom_perforations": bottom
        })

    elif template in {"Template 2", "Template 3", "Template 4", "Template 6", "Template 8", "Template 9", "Unknown"}:
        filtered = casings_df[casings_df["source_file"] == source_file].copy()

        if template in {"Template 8", "Template 9"}:
            if template == "Template 8":
                has_empty_type = (filtered["type"] == "").any()
                if has_empty_type:
                    # Misidentified Template 8: process all rows (like Templates 2, 3, 4, 6, Unknown)
                    pass
                else:
                    # True Template 8: only "Screen" rows
                    filtered = filtered[filtered["type"] == "Screen"]
            elif template == "Template 9":
                filtered = filtered[filtered["type"] == "Screen"]

        if filtered.empty:
            return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

        filtered["top_cleaned"] = filtered["top"].apply(lambda x: clean_to_digits(x, is_top=True))
        filtered["bottom_cleaned"] = filtered["bottom"].apply(clean_to_digits)

        filtered = filtered[
            (filtered["top_cleaned"].notnull()) &
            (filtered["bottom_cleaned"].notnull()) &
            (filtered["bottom_cleaned"] < 2000)
        ]

        if filtered.empty:
            return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

        top_row = filtered.loc[filtered["row_index"].idxmin()]
        bottom_row = filtered.loc[filtered["row_index"].idxmax()]

        top = top_row["top_cleaned"]
```

```
        bottom = bottom_row["bottom_cleaned"]

        if top is None or bottom is None:
            return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

        # Check if bottom is less than top
        if bottom < top:
            return pd.Series({
                "perforation_interval": "",
                "top_perforations": top,
                "bottom_perforations": ""
            })

        return pd.Series({
            "perforation_interval": f"{top} - {bottom}",
            "top_perforations": top,
            "bottom_perforations": bottom
        })

    elif template == "Template 5":
        raw_text = row.get("perforations_template5", "")
        if not raw_text.strip():
            return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

        ranges = re.findall(r"(\d{2,4})\s*[-–]\s*(\d{2,4})", raw_text)
        numbers = []

        for start, end in ranges:
            try:
                start_int = int(start)
                end_int = int(end)
                if 10 < start_int < 2000 and 10 < end_int < 2000:
                    numbers.extend([start_int, end_int])
            except ValueError:
                continue

        singles = re.findall(r"\b(\d{2,4})\b", raw_text)
        for val in singles:
            try:
                num = int(val)
                if 10 < num < 2000:
                    numbers.append(num)
            except ValueError:
                continue

        if numbers:
            top = min(numbers)
            bottom = max(numbers)
```

```
            # Check if bottom is less than top
            if bottom < top:
                return pd.Series({
                    "perforation_interval": "",
                    "top_perforations": top,
                    "bottom_perforations": ""
                })
            return pd.Series({
                "perforation_interval": f"{top} - {bottom}",
                "top_perforations": top,
                "bottom_perforations": bottom
            })

        return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

    elif template == "Template 7":
        raw_text = row.get("perforations_template7", "")
        if not raw_text.strip():
            return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

        ranges = re.findall(r"(\d{2,4})\s*[-–]\s*(\d{2,4})", raw_text)
        singles = re.findall(r"\b(\d{2,4})\b", raw_text)
        numbers = []

        for start, end in ranges:
            try:
                start_int = int(start)
                end_int = int(end)
                if 10 < start_int < 2000 and 10 < end_int < 2000:
                    numbers.extend([start_int, end_int])
            except ValueError:
                continue

        for val in singles:
            try:
                num = int(val)
                if 10 < num < 2000:
                    numbers.append(num)
            except ValueError:
                continue

        if numbers:
            top = min(numbers)
            bottom = max(numbers)
            # Check if bottom is less than top
            if bottom < top:
                return pd.Series({
                    "perforation_interval": "",
```

```
                    "top_perforations": top,
                    "bottom_perforations": ""
                })
            return pd.Series({
                "perforation_interval": f"{top} - {bottom}",
                "top_perforations": top,
                "bottom_perforations": bottom
            })

        return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

    return pd.Series({"perforation_interval": "", "top_perforations": "", "bottom_perforations": ""})

df[["perforation_interval", "top_perforations", "bottom_perforations"]] = df.apply(extract_perforation_fields, axis=1)

# Clean Well ID
df["well_id_cleaned"] = df.iloc[:, 0].apply(clean_well_id)

# Extract potential APN pattern from a text string
def extract_apn_from_text(text):
    if not isinstance(text, str):
        return ""
    text = text.upper().strip()

    text = text.replace("\n", " ").replace("APN#", "APN").replace("A.P.N.", "APN")

    match = re.search(r"\b(\d{2,3}-\d{3}-\d{2}[A-Z]?)\b", text)
    if match:
        return match.group(1)
    return ""

def fallback_apn_with_source(row):
    apn_raw = str(row.get(APN_COL, "")).strip()
    apn_digits = clean_apn_format(apn_raw)

    if apn_digits:
        return apn_digits.zfill(7), "Original APN field"

    apn_candidate = extract_apn_from_text(row.get("location_details", ""))
    source = "location_details"

    if not apn_candidate:
        apn_candidate = extract_apn_from_text(row.get("owner_well_id", ""))
        source = "owner_well_id"

    if apn_candidate:
        digits_only = clean_apn_format(apn_candidate)
        return digits_only.zfill(7), f"Extracted from {source}"
```

```python

    return "", "No APN found"

df[["apn_cleaned", "apn_cleaned_source"]] = df.apply(fallback_apn_with_source, axis=1,
result_type="expand")

# Define cleaning function for pump_test_length
def clean_pump_test_length(value):
    if pd.isna(value) or not str(value).strip():
        return ""

    value = str(value).strip()

    # Return empty string if slash is present
    if '/' in value:
        return ""

    # Remove any digits (and subsequent characters) after a non-digit, non-period character
    match = re.match(r'^[\d.]+', value)
    if not match:
        return ""

    cleaned = match.group(0)

    # Keep only digits and periods
    cleaned = re.sub(r'[^0-9.]', '', cleaned)

    # Ensure valid number format (remove multiple periods, trailing/leading periods)
    if cleaned.count('.') > 1:
        parts = cleaned.split('.')
        cleaned = parts[0] + '.' + ''.join(parts[1:])
    cleaned = cleaned.strip('.')

    # Return empty string if result is empty or invalid
    return cleaned if cleaned and cleaned != '.' else ""

numeric_columns = {
    "water_static": "water_static_cleaned",
    "depth_well": "depth_well_cleaned",
    "yield": "yield_cleaned",
    "seal": "seal_cleaned",
    "first_water": "first_water_cleaned",
    "total_depth": "total_depth_cleaned",
    "lat": "lat_cleaned",
    "long": "long_cleaned",
    "pump_test_length": "pump_test_length_cleaned"
}

def clean_numeric_field(value):
    if not isinstance(value, str):
```

```
        value = str(value)

    value = value.strip().lower()

    # Handle newline: keep only the first number before the newline
    if '\n' in value:
        value = value.split('\n')[0].strip()

    # Remove apostrophe and any digits following it
    value = re.sub(r"'\s*\d*(\"|''|’)?", "", value)

    value = re.sub(r'^[^\d]+', '', value)      #Remove non-digit prefix
    value = re.sub(r'[^\d]+$', '', value)

    range_pattern = r"^\s*(\d+(?:\.\d+)?)\s*(?:[-–]|to|t)\s*(\d+(?:\.\d+)?)\s*$"
    range_match = re.match(range_pattern, value)
    if range_match:
        try:
            num1 = float(range_match.group(1))
            num2 = float(range_match.group(2))
            avg = (num1 + num2) / 2
            return f"{avg:.2f}".rstrip('0').rstrip('.')
        except ValueError:
            pass

    value = re.sub(r'(?<=\d),(?=\d)', '.', value)
    value = re.sub(r'[^0-9.]', '', value)

    if value.count('.') > 1:
        parts = value.split('.')
        value = parts[0] + '.' + ''.join(parts[1:])

    return value

for colname, new_col in numeric_columns.items():
    if colname == "pump_test_length":
        df[new_col] = df[colname].apply(clean_pump_test_length)
    else:
        df[new_col] = df[colname].apply(clean_numeric_field)


# Initialize qa_flag column
df['depths_flag'] = ''

# Only check rows where both fields are non-empty
mask = (df['total_depth_cleaned'] != '') & (df['depth_well_cleaned'] != '')

# Apply logic: if total_depth < depth_well → flag 'review depths'
df.loc[mask, 'depths_flag'] = df.loc[mask].apply(
```

```
    lambda row: 'review depths' if float(row['total_depth_cleaned']) < float(row['depth_well_cleaned']) else '',
    axis=1
)


# Load and structure alias dictionary
alias_df = pd.read_excel(r"C:\Well Layer Update\GIS\data_tables\street_names_with_aliases.xlsx").fillna("")
alias_dict = {}
for _, row in alias_df.iterrows():
    canonical = row['canonical_name'].strip().upper()
    for alias in str(row['alias']).split(";"):
        alias = alias.strip().upper()
        if alias:
            alias_dict[alias] = canonical

for canonical in alias_df["canonical_name"].unique():
    canonical = canonical.strip().upper()
    alias_dict[canonical] = canonical

street_keywords = set(alias_dict.keys())

def extract_address_from_location(desc):
    if not isinstance(desc, str):
        return ""
    desc = desc.upper()
    desc = re.sub(r"[^\w\s]", " ", desc)
    desc = re.sub(r"\s+", " ", desc).strip()

    if re.search(r"\b(UTM|TOPO|SHEET|1[:;,\.]?62[,\.]?500|GRID|QUAD|COORDINATE|LAT|LONG|MAP)\b", desc):
        return ""

    if re.search(r"\b\d{1,5}\s*(FEET|FT|YARDS|MILES?)\b", desc):
        return ""

    tokens = desc.split()
    num_tokens = len(tokens)

    for i in range(num_tokens):
        for j in range(1, 5):
            if i + j > num_tokens:
                continue
            candidate_tokens = tokens[i:i + j]
            candidate = " ".join(candidate_tokens)
            if candidate in street_keywords:
                canonical = alias_dict.get(candidate, candidate)
                if i > 0 and re.fullmatch(r"\d{2,6}", tokens[i - 1]):
                    return f"{tokens[i - 1]} {canonical}"
                else:
```

```
                    return canonical

    for i in range(num_tokens):
        for j in range(1, 5):
            if i + j > num_tokens:
                continue
            candidate_tokens = tokens[i:i + j]
            candidate = " ".join(candidate_tokens)
            best_match, score, _ = process.extractOne(
                candidate,
                list(street_keywords),
                scorer=fuzz.token_sort_ratio
            )
            if score >= 85:
                canonical = alias_dict.get(best_match, best_match)
                if i > 0 and re.fullmatch(r"\d{2,6}", tokens[i - 1]):
                    return f"{tokens[i - 1]} {canonical}"
                else:
                    return canonical

    return ""

def resolve_address_with_source(row):
    addr = str(row.get(ADDR_COL, "")).strip()
    fallback = str(row.get(FALLBACK_ADDR_COL, "")).strip()
    loc_extracted = extract_address_from_location(row.get(LOC_DESC_COL, ""))

    fallback_keywords = ["same", "dame", "come", "came", "some", "sama", "same as above"]
    if any(fuzz.ratio(addr.lower(), word) >= 80 for word in fallback_keywords):
        return fallback, "Fallback address (approximate 'same')"

    if addr:
        return addr, "Direct address"
    if loc_extracted:
        return loc_extracted, "Extracted from location description"
    if fallback:
        return fallback, "Fallback address (owners)"
    return "", "No address available"

def strip_city_and_zip(address):
    if not address:
        return address

    cities = [
        "APTOS", "BEN LOMOND", "BOULDER CREEK", "CAPITOLA", "DAVENPORT",
        "LOS GATOS", "SANTA CRUZ", "SCOTTS VALLEY", "SOQUEL", "WATSONVILLE"
    ]
    address = address.upper()

    house_number_match = re.match(r"^\d+\s+", address)
```

```
    if not house_number_match:
        return address

    start = house_number_match.end()
    post_number = address[start:]
    tokens = post_number.split()
    retained_tokens = tokens[:1]
    remaining_tokens = tokens[1:]

    for idx, token in enumerate(remaining_tokens):
        joined = " ".join([*retained_tokens, *remaining_tokens[:idx + 1]])
        for city in cities:
            if city in joined:
                return address[:address.upper().find(city)].strip()

    return address

def clean_for_matching(addr):
    if not addr:
        return ""
    addr = addr.upper()
    addr = re.sub(r"\s*\n\s*", ", ", addr)
    addr = re.sub(r"\s+", " ", addr).strip()

    remove_keywords = [
        "P.O. BOX", "PO BOX", "O. BOX", "P.O.", "PO", "BOX", "PABOX", "DRAWER",
        "ABOVE", "OFF", "MAP", "MAPSHEET", "APN", "NO ADDRESS", "END OF", "ENDOF",
        "PART OF", "P/R", "SUITE", "C/O", "SHOPS", "NOTA", "WIR", "EX", "FEET NORTH OF"
    ]
    for keyword in remove_keywords:
        addr = re.sub(rf"\b{re.escape(keyword)}\b", "", addr)

    ocr_fixes = {
        r"\bARE\b": "AVE", r"\bSTH\b": "ST", r"\bDRI\b": "DR", r"5T": "ST",
        r"\bOAK STREE\b": "OAK ST", r"(\d)O(\d)": r"\1 0 \2",
        r"\bMOUNT\b": "MT", r"\bM\b": "MT", r"\bR\b": "RD",
        r"\bDRIE\b": "DR", r"\bQUINET\b": "QUINNETTE", r"\bJANE\b": "LANE",
        r"\bCEUK\b": "CREEEK", r"\bTHNY\b": "TWINY", r"\bELTON\b": "FELTON",
        r"\bBENLOMOND\b": "BEN LOMOND", r"\bBENLAMIND\b": "BEN LOMOND",
        r"\bBEN SOMAND\b": "BEN LOMOND", r"\bBEN HOWARD\b": "BEN LOMOND",
        r"\bBONNY DOON\b": "BONNY DOON", r"\bBUNNY DOON\b": "BONNY DOON",
        r"\bBROOKDALE\b": "", r"\bSANTACRUZ\b": "SANTA CRUZ",
        r"\bSANTACURZ\b": "SANTA CRUZ", r"\bCUPERTING\b": "CUPERTINO",
        r"\bROALND\b": "ROLAND", r"\bZIMT\b": "", r"\bHERMOURA\b": "HERMOSA",
        r"\bDALTON\b": "", r"\bSA\b": "CA", r"\bCALIFÓRNIA\b": "CA",
        r"\bCALLIL\b": "CA", r"\bCALIF\b": "CA", r"\bCALL\b": "CA", r"\bCAO\b": "CA",
        r"\bCAMILLA\b": "CAMPBELL", r"\bLETT?S VALLY\b": "LETTS VALLEY",
        r"\bENGINE GRADE\b": "EMPIRE GRADE", r"\bFORMAND\b": "LOMOND",
        r"\bSTREEL\b": "STREET", r"\bARE\b": "AVE", r"\bMª\b": "MA", r"\bª\b": "A"
    }
```

```
    for pattern, replacement in ocr_fixes.items():
        addr = re.sub(pattern, replacement, addr)

    addr = re.sub(r"^(WELL( SITE| HEAD| SA| LOCATION| HOUSE)?|ADDRESS|LOCATED AT|NEAR|NEXT TO|ADJACENT TO|BEHIND|ADJOINING)\s+", "", addr)

    replacements = {
        " STREET": " ST", " AVENUE": " AVE", " ROAD": " RD", " DRIVE": " DR",
        " BOULEVARD": " BLVD", " MOUNT ": " MT ", " HIGHWAY": " HWY",
        " LANE": " LN", " COURT": " CT", " PLACE": " PL", " TERRACE": " TER",
        " CIRCLE": " CIR", " TRAIL": " TRL", " CREEK": " CRK", " WAY": " WAY"
    }
    for k, v in replacements.items():
        addr = addr.replace(k, v)

    addr = re.sub(r"\s+[A-Z\s]*\sCA\s\d{5,6}$", "", addr)
    addr = re.sub(r"[^\w\s]+$", "", addr)
    addr = re.sub(r"\b\w\b$", "", addr).strip()
    addr = strip_city_and_zip(addr)
    return addr

df[["address_raw", "address_source"]] = df.apply(lambda row: pd.Series(resolve_address_with_source(row)), axis=1)
df["address_cleaned"] = df["address_raw"].astype(str).apply(lambda x: x.splitlines()[0] if x.strip() else "").str.strip().str.upper()
df["address_cleaned_for_matching"] = df["address_cleaned"].apply(clean_for_matching)

# Load Parcel Data and Prepare Matching Columns
parcels_df = pd.read_excel(parcel_excel, dtype=str).fillna("")
parcels_df.columns = parcels_df.columns.str.lower()
parcels_df["clean_apn"] = parcels_df.iloc[:, 0]
parcels_df["clean_address_parcel"] = parcels_df.iloc[:, 13].astype(str).str.strip().str.upper()
parcels_df["match_ready_address"] = parcels_df["clean_address_parcel"]

apn_lookup = parcels_df.set_index("clean_apn")[["clean_address_parcel"]].to_dict("index")

def fuzzy_address_lookup(addr):
    cleaned_input = clean_for_matching(addr)
    if not cleaned_input:
        return "", "", [], 0

    input_house_num = re.match(r"^(\d+)", cleaned_input)
    input_house_num = input_house_num.group(1) if input_house_num else ""

    matches = process.extract(
        cleaned_input,
        parcels_df["match_ready_address"].tolist(),
        scorer=fuzz.token_sort_ratio,
        limit=30
    )
```

```python

    candidates_log = []
    best_row, best_score, best_dist = None, 0, float("inf")

    for match_text, score, idx in matches:
        row = parcels_df.iloc[idx]
        match_addr = row["clean_address_parcel"]
        match_num = re.match(r"^(\d+)", match_addr)
        match_num = match_num.group(1) if match_num else ""
        dist = abs(int(input_house_num) - int(match_num)) if input_house_num and match_num else float("inf")
        candidates_log.append(f"{match_addr} (score={score}, dist={dist})")

        if score > best_score or (score == best_score and dist < best_dist):
            best_score, best_dist = score, dist
            best_row = row

    if best_row is not None:
        return best_row["clean_apn"], best_row["clean_address_parcel"], candidates_log, best_score

    return "", "", candidates_log, 0

# Normalize well_id_cleaned for duplicate comparison (strip leading zeros)
def normalize_well_id_for_comparison(well_id):
    if not isinstance(well_id, str) or not well_id.strip():
        return ""
    # Remove leading zeros, but preserve the rest of the ID (including 'E' prefix if present)
    if well_id.startswith('E'):
        return 'E' + well_id[1:].lstrip('0')
    return well_id.lstrip('0')

# Remove Duplicates Based on well_id_cleaned and multiple fields, including all-empty matches
def count_non_empty(row):
    """Count non-empty, non-whitespace fields in a row."""
    return sum(1 for val in row if isinstance(val, str) and val.strip() != "" or not pd.isna(val))

# Add duplicate details column
df['duplicate_exists'] = "No duplicates"

# Add a column to count non-empty fields for each row
df["non_empty_count"] = df.apply(count_non_empty, axis=1)

# Add a column for normalized well_id_cleaned for duplicate comparison
df["well_id_normalized"] = df["well_id_cleaned"].apply(normalize_well_id_for_comparison)

# Identify duplicates based on normalized well_id_cleaned and any field match or all fields empty
```

```
fields_to_check = [
    "total_depth_cleaned",
    "depth_well_cleaned",
    "water_static_cleaned",
    "yield_cleaned",
    "seal_cleaned",
    "first_water_cleaned",
    "permit_date",
    "permit_no",
    "date",
    "apn_cleaned"
]

# Filter for non-empty well_id_cleaned
non_empty_well_id_mask = df["well_id_cleaned"].notna() & (df["well_id_cleaned"] != "")
duplicates = pd.Series([False] * len(df), index=df.index)
for field in fields_to_check:
    # Mark duplicates where normalized well_id_cleaned and field match, and both are non-empty
    mask = (
        df.duplicated(subset=["well_id_normalized", field], keep=False) &
        (df[field].notna() & (df[field] != "")) &
        non_empty_well_id_mask
    )
    duplicates |= mask

# Add duplicates where all fields are empty for a given normalized well_id_cleaned
all_empty_mask = df[fields_to_check].apply(lambda x: x.isna() | (x == "")).all(axis=1)
all_empty_duplicates = (
    df[all_empty_mask & non_empty_well_id_mask].duplicated(subset=["well_id_normalized"], keep=False)
)
duplicates |= all_empty_duplicates

# Identify duplicate groups for logging
duplicate_groups = df[duplicates & non_empty_well_id_mask][["well_id_normalized", "well_id_cleaned", "total_depth_cleaned", "depth_well_cleaned"]].drop_duplicates()

print(f"Found {len(duplicate_groups)} groups of duplicates based on normalized well_id_cleaned and multiple fields or all fields empty")

# Process each duplicate group
records_to_keep = set()
for _, group in duplicate_groups.iterrows():
    well_id_norm = group["well_id_normalized"]
    total_depth = group["total_depth_cleaned"]
    depth_well = group["depth_well_cleaned"]

    # Get all records in this duplicate group (matching normalized well_id_cleaned and any field or all fields empty)
    group_records = df[
```

```python
        (df["well_id_normalized"] == well_id_norm) & (
            (df["total_depth_cleaned"] == total_depth) |
            (df["depth_well_cleaned"] == depth_well) |
            (df["water_static_cleaned"] == df[df["well_id_normalized"] == well_id_norm]["water_static_cleaned"].iloc[0]) |
            (df["yield_cleaned"] == df[df["well_id_normalized"] == well_id_norm]["yield_cleaned"].iloc[0]) |
            (df["seal_cleaned"] == df[df["well_id_normalized"] == well_id_norm]["seal_cleaned"].iloc[0]) |
            (df["first_water_cleaned"] == df[df["well_id_normalized"] == well_id_norm]["first_water_cleaned"].iloc[0]) |
            (df["permit_date"] == df[df["well_id_normalized"] == well_id_norm]["permit_date"].iloc[0]) |
            (df["permit_no"] == df[df["well_id_normalized"] == well_id_norm]["permit_no"].iloc[0]) |
            (df["date"] == df[df["well_id_normalized"] == well_id_norm]["date"].iloc[0]) |
            (df["apn_cleaned"] == df[df["well_id_normalized"] == well_id_norm]["apn_cleaned"].iloc[0]) |
            (
                df[fields_to_check].apply(lambda x: x.isna() | (x == "")).all(axis=1) &
                (df["well_id_normalized"] == well_id_norm)
            )
        ) & non_empty_well_id_mask
    ]

    if len(group_records) > 1:
        # Sort by non_empty_count (descending), then by source_file (prefer newer or cleaner files)
        most_complete = group_records.sort_values(
            by=["non_empty_count", "source_file"], ascending=[False, True]
        ).iloc[0]
        df.at[most_complete.name, 'duplicate_exists'] = "Duplicates deleted (matched depths, water_static, yield, seal, first_water, permit_date, permit_no, date, apn_cleaned, or all fields empty, including leading zero variations)"
        records_to_keep.add(most_complete.name)
    else:
        records_to_keep.add(group_records.index[0])

# Keep non-duplicate records and the most complete record from each duplicate group
non_duplicates = df[~duplicates]
duplicates_to_keep = df.loc[list(records_to_keep)]
df = pd.concat([non_duplicates, duplicates_to_keep], ignore_index=True)

# Ensure records with empty well_id_cleaned are marked as "No duplicates"
df.loc[df["well_id_cleaned"].isna() | (df["well_id_cleaned"] == ""), "duplicate_exists"] = "No duplicates"

# Flag distinct duplicates in the final DataFrame, excluding empty well_id_cleaned
well_id_counts = df[df["well_id_cleaned"].notna() & (df["well_id_cleaned"] != "")]["well_id_normalized"].value_counts()
duplicate_well_ids = well_id_counts[well_id_counts > 1].index
```

```
for well_id_norm in duplicate_well_ids:
    mask = df["well_id_normalized"] == well_id_norm
    for idx in df[mask].index:
        current = df.at[idx, 'duplicate_exists']
        if current == "No duplicates":
            df.at[idx, 'duplicate_exists'] = "Distinct duplicates exist (different depths, water_static, yield, seal, first_water, permit_date, permit_no, date, apn_cleaned, and not all empty, including leading zero variations)"
        elif current == "Duplicates deleted (matched depths, water_static, yield, seal, first_water, permit_date, permit_no, date, apn_cleaned, or all fields empty, including leading zero variations)":
            df.at[idx, 'duplicate_exists'] = "Duplicates deleted (matched depths, water_static, yield, seal, first_water, permit_date, permit_no, date, apn_cleaned, or all fields empty, including leading zero variations) and distinct duplicates exist"

# Drop temporary columns
df = df.drop(columns=["non_empty_count", "well_id_normalized"])

# Verify no duplicates remain
remaining_duplicates = df[
    (df["well_id_cleaned"].notna() & (df["well_id_cleaned"] != "") ) & (
        df.duplicated(subset=["well_id_cleaned", "total_depth_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "depth_well_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "water_static_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "yield_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "seal_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "first_water_cleaned"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "permit_date"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "permit_no"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "date"], keep=False) |
        df.duplicated(subset=["well_id_cleaned", "apn_cleaned"], keep=False) |
        (
            df[fields_to_check].apply(lambda x: x.isna() | (x == "")).all(axis=1) &
            df.duplicated(subset=["well_id_cleaned"], keep=False)
        )
    )
]
if not remaining_duplicates.empty:
    print("flag: some duplicates remain")
    print(remaining_duplicates[["well_id_cleaned", "total_depth_cleaned", "depth_well_cleaned", "water_static_cleaned", "yield_cleaned", "seal_cleaned", "first_water_cleaned", "permit_date", "permit_no", "date", "apn_cleaned", "source_file"]])
else:
    print("No duplicates remain.")

# Fuzzy Matching Loop
final_apn, final_address, match_source, fuzzy_logs = [], [], [], []

for i, (_, row) in enumerate(df.iterrows()):
    print(f"Processing record {i + 1} of {len(df)}")
```

```
    cleaned_apn = row["apn_cleaned"]
    cleaned_addr = row["address_cleaned"]
    source = row["address_source"]

    if cleaned_apn in apn_lookup:
        final_apn.append(cleaned_apn)
        final_address.append(apn_lookup[cleaned_apn]["clean_address_parcel"])
        match_source.append("APN matched")
        fuzzy_logs.append("")
    else:
        if source == "Extracted from location description":
            fuzzy_threshold = 75
        elif "Fallback address" in source:
            fuzzy_threshold = 85
        else:
            fuzzy_threshold = 62

        fuzzy_apn, fuzzy_addr, log, score = fuzzy_address_lookup(cleaned_addr)
        if fuzzy_apn and score >= fuzzy_threshold:
            final_apn.append(fuzzy_apn)
            final_address.append(fuzzy_addr)
            if "Fallback address" in source:
                match_source.append(f"Owners address (fallback) matched (score {score})")
            else:
                match_source.append(f"Address matched (score {score})")
        else:
            final_apn.append(cleaned_apn)
            final_address.append("")
            match_source.append("No match")
        fuzzy_logs.append("; ".join(log))

planned_use_columns = {
    "domestic_well": "Domestic Well",
    "irrigation_well": "Irrigation Well",
    "public_well": "Public Supply Well",
    "monitoring_well": "Monitoring Well",
    "industrial_well": "Industrial Well",
    "other_well": "Other Well",
    "test_well": "Test Well"
}

def determine_planned_use(row):
    selected = [
        label for col, label in planned_use_columns.items()
        if row.get(col, "").strip() == ":selected:"
    ]
    if len(selected) == 1:
        return selected[0]
    elif len(selected) > 1:
        return "Two or more uses/Possible Error from 'Confidential' Stamp"
```

```
    else:
        return "Other"

df["planned_use_former_use"] = df.apply(determine_planned_use, axis=1)

type_of_work_columns = {
    "new_well": "New Well",
    "keyword_destruction_well": "Well Destruction",
    "deepen_well": "Deepen Well",
    "replacement": "Replacement Well",
    "recondition_other.": "Well Reconditioning/Other"
}

def determine_type_of_work(row):
    selected = [
        label for col, label in type_of_work_columns.items()
        if row.get(col, "").strip() == ":selected:"
    ]
    if len(selected) == 1:
        return selected[0]
    elif len(selected) > 1:
        return "Two or more types"
    else:
        return "Unknown"

df["Type of Work"] = df.apply(determine_type_of_work, axis=1)

def map_record_type(type_of_work):
    if type_of_work == "New Well":
        return "WellCompletion/New/Production or Monitoring/NA"
    elif type_of_work == "Unknown":
        return "Unknown"
    elif type_of_work in ["Drill and Destroy", "Destroy"]:
        return "WellCompletion/Drill and Destroy/NA/NA"
    elif type_of_work == "Well Destruction":
        return "WellCompletion/Destruction/NA/NA"
    elif type_of_work in ["Deepen Well", "Replacement Well", "Well Reconditioning/Other"]:
        return "WellCompletion/Modification or Repair/Production or Monitoring/NA"
    else:
        return "Unknown"

df["Record Type"] = df["Type of Work"].apply(map_record_type)

df["final_apn"] = final_apn
df["final_address"] = final_address
df["match_source"] = match_source
df["fuzzy_candidates_log"] = fuzzy_logs

df.loc[df["template_id"] == "Template 8", "planned_use_former_use"] =
df.loc[df["template_id"] == "Template 8", "Planned Use"]
```

```
df.loc[df["template_id"] == "Template 8", "Type of Work"] = df.loc[df["template_id"] ==
"Template 8", "Activity"]
df["Record Type"] = df["Type of Work"].apply(map_record_type)

df.to_excel(output_excel, index=False)
print(f"Cleaned data written to: {output_excel}")
```

C:\Well Layer Update\Python\Azure\Data Matching and Merging Script.py

```python
import pandas as pd
import re
from rapidfuzz import process, fuzz
import logging
import numpy as np
from collections import defaultdict

# === File paths ===
file1_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR(SCC1).xlsx"
file2_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\DWR1.xlsx"
file_monterey = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\Well Completion Reports Monterey.xlsx"
matched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\matched_output.xlsx"
unmatched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched_output.xlsx"
unmatched_ocr_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched_ocr_output.xlsx"
unmatched_dwr_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched_dwr_output.xlsx"
supplemented_output_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\matched_output_supplemented.xlsx"
parcel_excel = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\Parcels.xlsx"


# file1_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR(DWR1).xlsx"
# file2_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched_dwr_output.xlsx"
# file_monterey = r"s:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\Well Completion Reports Monterey.xlsx"
# matched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched dwr files\matched_output.xlsx"
# unmatched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched dwr files\unmatched_output.xlsx"
# unmatched_ocr_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched dwr files\unmatched_ocr_output.xlsx"
# unmatched_dwr_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched dwr files\unmatched_dwr_output.xlsx"
# supplemented_output_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\unmatched dwr files\matched_output_supplemented.xlsx"
# parcel_excel = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\Parcels.xlsx"



#Load input datasets (OCR, DWR, Monterey, and parcel data) and preserve original match_source
```

```python
df1 = pd.read_excel(file1_path)
df2 = pd.read_excel(file2_path)
df3 = pd.read_excel(file_monterey)
parcels_df = pd.read_excel(parcel_excel)

if 'match_source' in df1.columns:
    df1['ocr_original_match_source'] = df1['match_source']

# Define helper function to clean and normalize well IDs
def clean_id(raw):
    if pd.isna(raw):
        return ""
    original = str(raw).strip()
    cleaned = original.replace(" ", "").upper().strip()
    # Convert 'c' to 'E' for IDs starting with 'c' or 'C'
    if cleaned.lower().startswith('c'):
        cleaned = 'E' + cleaned[1:]
    if re.match(r"^WCR\d{4}-\d{6}$", cleaned):
        return cleaned
    text = original.lower().replace(",", "")
    text = re.sub(r"no\.?", "", text)
    text = text.replace("/", " ")
    text = re.sub(r"[\s\n]+", " ", text)
    match = re.match(r"^(\d{1,2})[\.-](\d{1,2})[\.-](\d{2,4})(\d{3,8})$", text)
    if match:
        year_str = match.group(3)
        try:
            year = int(year_str)
            if year < 100:
                year += 1900 if year >= 35 else 2000
            if 1935 <= year <= 2025:
                return match.group(4)
        except ValueError:
            pass
    prefix_removed = re.sub(r"^\d{1,2}[\.-]\d{1,2}[\.-]\d{2,4}", "", text).strip()
    segments = re.split(r"[\s]+", prefix_removed)
    for segment in reversed(segments):
        segment_digits = re.findall(r"\d{3,8}", segment)
        if segment_digits:
            return segment_digits[-1].lstrip('0')
    digits = re.findall(r"\d{3,8}", text)
    if digits:
        return digits[-1].lstrip('0')
    # Return cleaned ID if it starts with 'E' and has digits
    if cleaned.startswith('E') and any(c.isdigit() for c in cleaned[1:]):
        return cleaned
    return ""

#Format APN strings to xxx-yyy-zz pattern for 8-digit APNs, consistent with SCC Parcel db
def format_apn_with_dashes(apn_str):
```

```python
    apn_str = str(apn_str).strip()
    clean_apn = re.sub(r'\D', '', apn_str)
    if len(clean_apn) == 8:
        return f"{clean_apn[:3]}-{clean_apn[3:6]}-{clean_apn[6:]}"
    else:
        return apn_str

# Normalize parcel APNs by removing non-digits and zero-padding to 8 digits
def normalize_parcel_apn(apn_str):
    return re.sub(r"\D", "", str(apn_str)).zfill(8)

# Generate APN signature by keeping non-zero digits
def apn_signature(apn):
    return ''.join([ch for ch in str(apn) if ch.isdigit() and ch != '0'])

#Clean address strings for consistent matching
def clean_for_matching(addr):
    if not isinstance(addr, str):
        addr = str(addr) if not pd.isna(addr) else ""
    addr = addr.upper()
    addr = addr.replace("(", "").replace(")", "")
    addr = re.sub(r"\bARE\b", "AVE", addr)
    addr = re.sub(r"\bSTH\b", "ST", addr)
    addr = re.sub(r"\bDRI\b", "DR", addr)
    addr = re.sub(r"\bMOUNT\b", "MT", addr)
    addr = re.sub(r"\bR\b", "RD", addr)
    addr = re.sub(r" STREET", " ST", addr)
    addr = re.sub(r" AVENUE", " AVE", addr)
    addr = re.sub(r" ROAD", " RD", addr)
    addr = re.sub(r" DRIVE", " DR", addr)
    addr = re.sub(r" LANE", " LN", addr)
    addr = re.sub(r" COURT", " CT", addr)
    addr = re.sub(r" PLACE", " PL", addr)
    addr = re.sub(r" TERRACE", " TER", addr)
    addr = re.sub(r" CIRCLE", " CIR", addr)
    addr = re.sub(r" TRAIL", " TRL", addr)
    addr = re.sub(r" CREEK", " CRK", addr)
    addr = re.sub(r" WAY", " WAY", addr)
    addr = re.sub(r"\s+[A-Z\s]*\sCA\s\d{5}$", "", addr)
    addr = re.sub(r"[^\w\s]+$", "", addr)
    addr = re.sub(r"\s+", " ", addr).strip()
    addr = re.sub(r"\b\w\b$", "", addr).strip()
    return addr

# Normalize APN strings by removing non-digits and handling specific cases
def normalize_apn(apn_str):
    digits = re.sub(r"\D", "", str(apn_str))
    if len(digits) == 9 and digits[-3] == '0':
        digits = digits[:-3] + digits[-2:]
    return digits.zfill(8)
```

```python

# Generate APN signature for matching by keeping non-zero digits
def get_apn_signature(apn_str):
    return "".join(d for d in str(apn_str) if d.isdigit() and d != "0")

# Extract APN from location text using regex
def extract_apn_from_location(location_str):
    match = re.search(r'\bAPN\s*(\d{2,3}[- ]?\d{2,3}[- ]?\d{2,3})\b', str(location_str), re.IGNORECASE)
    if match:
        return normalize_apn(match.group(1))
    return None

#Normalize year values for comparisons
def normalize_year(v):
    if pd.isna(v) or not str(v).strip():
        return None
    s = str(v).strip()
    if '/' in s:
        parts = s.split('/')
        if len(parts) >= 3 and parts[2].isdigit():
            y = int(parts[2])
            if len(parts[2]) == 2:
                y += 1900 if y >= 35 else 2000
            return y
    s = re.sub(r'\s', '', s)
    match = re.search(r'(\d{2,4})$', s)
    if match:
        y = int(match.group(1))
        if 0 <= y <= 99:
            y += 1900 if y >= 35 else 2000
        if 1935 <= y <= 2030:
            return y
    return None

# Find best matching DWR record based on well ID and field comparisons
def get_best_match(well_id, records, df1_row=None, fields_to_check=None, field_map=None, used_indices=None, target_df=None):
    if used_indices is None:
        used_indices = set()

    if not well_id.strip():
        return None, None, None, None, None, 0

    well_id_clean = str(well_id).strip().rstrip('M').replace('.0', '')
    if well_id_clean.startswith('E'):
        well_id_normalized = 'E' + well_id_clean[1:].lstrip('0')
    else:
        well_id_normalized = well_id_clean.lstrip('0')

```

```
    exact_matches = []
    for idx, (target_id, wcr_number, apn, addr, depth, water) in enumerate(records):
        if idx in used_indices:
            continue
        target_clean = str(target_id).strip().rstrip('M').replace('.0', '')
        if target_clean.startswith('E'):
            target_normalized = 'E' + target_clean[1:].lstrip('0')
        else:
            target_normalized = target_clean.lstrip('0')

        if well_id_normalized == target_normalized:
            exact_matches.append((target_id, wcr_number, apn, addr, depth, water, idx))

    if exact_matches:
        if len(exact_matches) == 1:
            target_id, wcr_number, apn, addr, depth, water, idx = exact_matches[0]
            logging.debug(f"Exact match: well_id_normalized='{well_id_normalized}', target='{target_id}' (single match)")
            return target_id, 100, apn, addr, idx, 0
        else:
            logging.debug(f"Multiple matches found for well_id_normalized='{well_id_normalized}': {len(exact_matches)} candidates")
            best_match = None
            best_match_score = -1
            best_match_idx = None
            best_matched_fields = []

            for match in exact_matches:
                target_id, wcr_number, apn, addr, depth, water, idx = match
                df2_row = df2.iloc[idx] if target_df is None else target_df.iloc[idx]
                match_score = 0
                matched_fields = []

                for df1_field in ["date", "depth_well_cleaned"] + [f for f in fields_to_check if f not in ["date", "depth_well_cleaned"]]:
                    df2_field = None
                    for key, (df1_col, _) in field_map.items():
                        if df1_col == df1_field:
                            df2_field = key
                            break
                    if not df2_field:
                        continue

                    df1_value = df1_row.get(df1_field) if df1_row is not None else np.nan
                    df2_value = df2_row.get(df2_field)

                    if isinstance(df1_value, str) and df1_value.strip() == '':
                        df1_value = np.nan
                    if isinstance(df2_value, str) and df2_value.strip() == '':
                        df2_value = np.nan

```

```
                    if df1_field == "date":
                        y1 = normalize_year(df1_value)
                        y2 = normalize_year(df2_value)
                        if y1 == y2 and y1 is not None:
                            match_score += 2
                            matched_fields.append(df1_field)
                            continue
                    elif df1_field == "depth_well_cleaned":
                        if not pd.isna(df1_value) and not pd.isna(df2_value):
                            if abs(df1_value - df2_value) <= 5:
                                match_score += 2
                                matched_fields.append(df1_field)
                            continue
                    elif df1_field in ["water_static_cleaned", "total_depth_cleaned", "yield_cleaned"]:
                        if not pd.isna(df1_value) and not pd.isna(df2_value):
                            if abs(df1_value - df2_value) <= 5:
                                match_score += 1
                                matched_fields.append(df1_field)
                            continue
                    elif df1_field == "apn_cleaned":
                        v1 = normalize_apn(df1_value) if df1_value else ''
                        v2 = normalize_apn(df2_value) if df2_value else ''
                        if v1 == v2 and v1:
                            match_score += 1
                            matched_fields.append(df1_field)
                            continue

                    if pd.isna(df1_value) and pd.isna(df2_value):
                        continue
                    elif isinstance(df1_value, str) and isinstance(df2_value, str):
                        if df1_value.strip().upper() == df2_value.strip().upper():
                            match_score += 1
                            matched_fields.append(df1_field)
                    elif df1_value == df2_value:
                        match_score += 1
                        matched_fields.append(df1_field)

                if match_score > best_match_score:
                    best_match = match
                    best_match_score = match_score
                    best_match_idx = idx
                    best_matched_fields = matched_fields

            if best_match:
                target_id, wcr_number, apn, addr, depth, water, idx = best_match
                logging.debug(f"Selected match for well_id_normalized='{well_id_normalized}' based on field matching (score={best_match_score}, matched_fields={best_matched_fields})")
                return target_id, 100, apn, addr, idx, best_match_score

```

```
            logging.debug(f"no match from field scoring for well_id_normalized='{well_id_normalized}' - falling back to depth/water")
            best_match = exact_matches[0]
            best_depth_match = False
            df1_depth = df1_row.get("depth_well_cleaned", np.nan) if df1_row is not None else np.nan
            df1_water = df1_row.get("water_static_cleaned", np.nan) if df1_row is not None else np.nan
            for match in exact_matches:
                target_id, wcr_number, apn, addr, depth, water, idx = match
                depth_match = (not pd.isna(df1_depth) and not pd.isna(depth) and abs(df1_depth - depth) <= 5)
                water_match = (not pd.isna(df1_water) and not pd.isna(water) and abs(df1_water - water) <= 5)
                if depth_match:
                    logging.debug(f"Selected match for well_id_normalized='{well_id_normalized}' due to depth match (df1_depth={df1_depth}, candidate_depth={depth})")
                    return target_id, 100, apn, addr, idx, 0
                elif water_match and not best_depth_match:
                    best_match = match
                    logging.debug(f"Selected match for well_id_normalized='{well_id_normalized}' due to water match (df1_water={df1_water}, candidate_water={water})")
                    best_depth_match = False
            target_id, wcr_number, apn, addr, depth, water, idx = best_match
            logging.debug(f"Fallback match: well_id_normalized='{well_id_normalized}', target='{target_id}' (index={idx})")
            return target_id, 100, apn, addr, idx, 0

    df1_depth = df1_row.get("depth_well_cleaned", np.nan) if df1_row is not None else np.nan
    df1_water = df1_row.get("water_static_cleaned", np.nan) if df1_row is not None else np.nan
    for idx, (target_id, wcr_number, apn, addr, depth, water) in enumerate(records):
        if idx in used_indices:
            continue
        target_clean = str(target_id).strip().rstrip('M').replace('.0', '')
        if target_clean.startswith('E'):
            target_normalized = 'E' + target_clean[1:].lstrip('0')
        else:
            target_normalized = target_clean.lstrip('0')

        is_substring_match = False
        if well_id_normalized in target_normalized and len(target_normalized) > len(well_id_normalized):
            is_substring_match = True
            match_type = f"Substring match (well_id in target): well_id_normalized='{well_id_normalized}', target='{target_clean}'"
        elif target_normalized in well_id_normalized and len(well_id_normalized) > len(target_normalized):
            is_substring_match = True
```

```
            match_type = f"Substring match (target in well_id): well_id_normalized='{well_id_normalized}', target='{target_clean}'"

        if is_substring_match:
            depth_match = (not pd.isna(df1_depth) and not pd.isna(depth) and abs(df1_depth - depth) <= 5)
            water_match = (not pd.isna(df1_water) and not pd.isna(water) and abs(df1_water - water) <= 5)
            if not (depth_match or water_match):
                logging.debug(f"Skipping substring match for well_id_normalized='{well_id_normalized}', target='{target_clean}' due to no depth/water match")
                continue

            logging.debug(match_type)
            return target_id, 80, apn, addr, idx, 0

    logging.debug(f"No match found for well_id_normalized='{well_id_normalized}'")
    return None, None, None, None, None, 0

# Perform fuzzy matching on addresses against parcel data
def fuzzy_address_lookup(addr):
    cleaned_input = clean_for_matching(addr)
    if not cleaned_input:
        return None, None, 0

    matches = process.extract(
        cleaned_input,
        parcels_df["parcel_address"].tolist(),
        scorer=fuzz.token_sort_ratio,
        limit=10
    )

    for match_text, score, idx in matches:
        if score >= 85:
            best_row = parcels_df.iloc[idx]
            return best_row["parcel_apn"], best_row["parcel_address"], score

    return None, None, 0

# Prepare DWR dataset (df2) for matching by cleaning IDs, APNs, and addresses
df2["clean_apn"] = df2.iloc[:, 27].apply(normalize_apn)
df2["clean_address"] = df2.iloc[:, 3].apply(clean_for_matching)
df2_depth = df2["Total Completed Depth"].astype(float).fillna(np.nan)
df2_water = df2["Static Water Level"].astype(float).fillna(np.nan)

df2_ids = []
for i in range(len(df2)):
    id1_raw = df2.iloc[i, 1]  # Legacy Log Number
    id0_raw = df2.iloc[i, 0]  # WCR Number
    id1 = clean_id(id1_raw)
```

```python
    id0 = clean_id(id0_raw)
    if id1 and re.match(r"^(WCR\d{4}-\d{6}|\d{3,8}|E\d+)$", id1):
        df2_ids.append(id1)
    else:
        df2_ids.append(id0)
df2_ids = pd.Series(df2_ids, index=df2.index)

df2_apn = df2.iloc[:, 27].astype(str).fillna('')
df2_addr = df2.iloc[:, 3].astype(str).fillna('')
df2_wcr = df2.iloc[:, 0].astype(str).fillna('')
df2_records = list(zip(df2_ids, df2_wcr, df2_apn, df2_addr, df2_depth, df2_water))

# Prepare Monterey dataset (df3) for secondary matching
df3.columns = df3.columns.str.strip()
df3_ids = []
for i in range(len(df3)):
    id1_raw = df3.iloc[i, 1]  # Legacy Log Number
    id0_raw = df3.iloc[i, 0]  # WCR Number
    id1 = clean_id(id1_raw)
    id0 = clean_id(id0_raw)
    if id1 and re.match(r"^(WCR\d{4}-\d{6}|\d{3,8}|E\d+)$", id1):
        df3_ids.append(id1)
    else:
        df3_ids.append(id0)
df3_ids = pd.Series(df3_ids, index=df3.index)
df3_wcr = df3.iloc[:, 0].astype(str).fillna('')
df3_apn = df3.iloc[:, 27].astype(str).fillna('')
df3_addr = df3.iloc[:, 3].astype(str).fillna('')
df3_depth = df3["Total Completed Depth"].astype(float).fillna(np.nan)
df3_water = df3["Static Water Level"].astype(float).fillna(np.nan)
df3_records = list(zip(df3_ids, df3_wcr, df3_apn, df3_addr, df3_depth, df3_water))

# Prepare parcel data for APN and address lookups
parcels_df["parcel_address"] = parcels_df.iloc[:, 13].astype(str).apply(clean_for_matching)
parcels_df["parcel_apn"] = parcels_df.iloc[:, 1].apply(normalize_parcel_apn)
parcel_apn_to_address = dict(zip(parcels_df["parcel_apn"], parcels_df["parcel_address"]))
signature_to_apns = defaultdict(list)
for apn in parcels_df["parcel_apn"]:
    sig = get_apn_signature(apn)
    signature_to_apns[sig].append(apn)
apn_lookup = parcels_df.set_index("parcel_apn")["parcel_address"].to_dict()

# Set up logging for debugging and tracking matches
logging.basicConfig(
    filename=r"C:\Well Layer Update\Python\Azure\Results\wc11\excel 3\fuzzy_match_debug.log",
    filemode="a",
    format="%(asctime)s [%(levelname)s] %(message)s",
    level=logging.DEBUG
)

```

```
# Log sample data for debugging
logging.debug(f"df1 well_id_cleaned sample: {df1['well_id_cleaned'].head(10).tolist()}")
logging.debug(f"df2_ids sample: {df2_ids.head(10).tolist()}")


# Define fields for validating duplicate matches
fields_to_check = [
    "total_depth_cleaned",
    "depth_well_cleaned",
    "water_static_cleaned",
    "yield_cleaned",
    "permit_date",
    "permit_no",
    "date",
    "apn_cleaned"
]

#Define mapping of DWR fields to OCR fields and their source columns
field_map = {
    "Owner Assigned Well Number": ("owner_well_id", "owner_assigned_source"),
    "Date Work Ended": ("date", "date_source"),
    "Permit Number": ("permit_no", "permit_number_source"),
    "Driller Name": ("contractor", "driller_name_source"),
    "Total Drill Depth": ("total_depth_cleaned", "total_drill_depth_source"),
    "Total Completed Depth": ("depth_well_cleaned", "total_completed_depth_source"),
    "Static Water Level": ("water_static_cleaned", "static_water_level_source"),
    "Well Yield": ("yield_cleaned", "well_yield_source"),
    "B118WellUse": ("planned_use_former_use", "b118welluse_source"),
    "Record Type": ("Type of Work", "record_type_source"),
    "Permit Date": ("permit_date", "permit_date_source"),
    "Driller License Number": ("contractor_license", "driller_license_number_source"),
    "Drilling Method": ("drilling_method", "drilling_method_source"),
    "Fluid": ("fluid", "fluid_source"),
    "Total Draw Down": ("total_drawdown", "total_drawdown_source"),
    "Test Type": ("test_type", "test_type_source"),
    "Pump Test Length": ("pump_test_length_cleaned", "pump_test_length_source"),
    "Other Observations": ("other_observations", "other_observations_source"),
    "City": ("city", "city_source"),
    "Ground Surface Elevation": ("surface_elevation", "surface_elevation_source"),
    "Elevation Determination Method": ("elevation_method", "elevation_method_source"),
    "Top Of Perforated Interval": ("top_perforations", "top_perforations_source"),
    "Bottom of Perforated Interval": ("bottom_perforations", "bottom_perforations_source"),
    "Decimal Latitude": ("lat_cleaned", "latitude_source"),
    "Decimal Longitude": ("long_cleaned", "longitude_source"),
    "APN": ("apn_cleaned", "apn_source")
}

#Define additional OCR fields to carry over to supplemented output
additional_fields = {
    "seal_cleaned": ("seal_cleaned", "seal_cleaned_source"),
```

```python
    "first_water_cleaned": ("first_water_cleaned", "first_water_cleaned_source"),
    "template_id": ("template_id", "template_id_source"),
    "OCR_match_source": ("ocr_original_match_source", "OCR_match_source_source")
}

#Initialize set to track used DWR indices for matching
used_dwr_indices = set()

# Match OCR records to DWR records and update APN/address
def match_apn_and_address(row):
    well_id = clean_id(row["well_id_cleaned"]).strip()
    if not well_id:
        return row["final_address"], str(row["final_apn"]).zfill(7), "No match", "No valid well ID", 0

    well_id = str(well_id).strip().lstrip('0').replace('.0', '')
    logging.debug(f"Processing well_id: {well_id}")

    match_id, score, apn_dwr, addr_dwr, dwr_idx, match_score = get_best_match(
        well_id, df2_records, df1_row=row, fields_to_check=fields_to_check, field_map=field_map, used_indices=used_dwr_indices, target_df=df2
    )
    if match_id and dwr_idx is not None:
        dwr_row = df2.iloc[dwr_idx]
        wcr_number = dwr_row["WCR Number"]
        dwr_apn = dwr_row["clean_apn"]
        dwr_address = dwr_row["clean_address"]
        logging.debug(f"Match found for well_id='{well_id}' to df2_id='{match_id}' (score={score}, duplicate_match_score={match_score})")
        if not dwr_apn or dwr_apn == '00000000':
            potential_apn = extract_apn_from_location(dwr_row["Well Location"])
            if potential_apn:
                dwr_apn = potential_apn
                logging.debug(f"Extracted APN '{dwr_apn}' from Well Location")

        if dwr_apn in parcel_apn_to_address:
            return parcel_apn_to_address[dwr_apn], dwr_apn, f"ID match (score {score}) → APN match", f"OCR ID '{well_id}' → DWR ID '{match_id}' | WCR Number: {wcr_number}", match_score

        matched_apn, matched_address, addr_score = fuzzy_address_lookup(dwr_address)
        if matched_apn and matched_address:
            return (
                matched_address,
                matched_apn,
                f"ID match (score {score}) → fuzzy address",
                f"OCR ID '{well_id}' → DWR ID '{match_id}' → Address (score: {addr_score}) | WCR Number: {wcr_number}",
                match_score
            )

```

```
        logging.debug(f"ID match for '{well_id}' to '{match_id}' (score: {score}) but no parcel match")
        return row["final_address"], str(row["final_apn"]).zfill(7), f"ID match (score {score}) with no parcel match", f"OCR ID '{well_id}' → DWR ID '{match_id}' | WCR Number: {wcr_number}", match_score

    location_details_raw = row.get("location_details", "")
    location_details_clean = clean_for_matching(location_details_raw)

    if location_details_clean:
        dwr_locations_raw = df2["Well Location"].fillna("").astype(str)
        dwr_locations_cleaned = dwr_locations_raw.apply(clean_for_matching).tolist()

        match_result = process.extractOne(location_details_clean, dwr_locations_cleaned, scorer=fuzz.token_sort_ratio)

        if match_result:
            best_match_text, score, match_index = match_result
            matched_row = df2.iloc[match_index]
            original_dwr_location = dwr_locations_raw.iloc[match_index]

            if score >= 80:
                legacy_log = str(matched_row.get("Legacy Log Number", "")).strip().upper()
                if legacy_log != "NN":
                    ocr_apn = str(row["final_apn"]).zfill(7)
                    ocr_match_source = str(row.get("match_source", ""))
                    if ocr_match_source == "No match" or ocr_apn not in parcel_apn_to_address:
                        ocr_apn = "0000000"
                    return row["final_address"], ocr_apn, "Fallback to OCR values", "Fuzzy match blocked: Legacy Log Number not 'NN'", 0

                dwr_apn = matched_row["clean_apn"]
                dwr_address = matched_row["clean_address"]

                if not dwr_apn or dwr_apn == '00000000':
                    potential_apn = extract_apn_from_location(matched_row["Well Location"])
                    if potential_apn:
                        dwr_apn = potential_apn

                if dwr_apn in parcel_apn_to_address:
                    wcr_number = matched_row["WCR Number"]
                    return parcel_apn_to_address[dwr_apn], dwr_apn, "Fuzzy location → APN match", f"Location '{location_details_raw}' → '{original_dwr_location}' | WCR Number: {wcr_number}", 0

                matched_apn, matched_address, addr_score = fuzzy_address_lookup(dwr_address)
                if matched_apn and matched_address:
                    wcr_number = matched_row["WCR Number"]
                    return (
                        matched_address,
```

```python
                        matched_apn,
                        "Fuzzy location → fuzzy address",
                        f"Location '{location_details_raw}' → '{original_dwr_location}' → Address: {dwr_address} (score: {score}) → Parcel (score: {addr_score}) | WCR Number: {wcr_number}",
                        0
                    )

                wcr_number = matched_row["WCR Number"]
                return row["final_address"], str(row["final_apn"]).zfill(7), "Fuzzy location match with no parcel match", f"Location '{location_details_raw}' → '{original_dwr_location}' (score: {score}) | WCR Number: {wcr_number}", 0

    ocr_apn = str(row["final_apn"]).zfill(7)
    ocr_match_source = str(row.get("match_source", ""))
    if ocr_match_source == "No match":
        ocr_apn = "0000000"
    return row["final_address"], ocr_apn, "Fallback to OCR values", "No match found", 0

# Apply matching logic to update APN and address fields in OCR dataset
df1_ids = df1["well_id_cleaned"].apply(clean_id)
df1_apn = df1["final_apn"].astype(str).fillna('')
df1_addr = df1["final_address"].astype(str).fillna('')

df1[["final_address", "final_apn", "match_source", "match_log", "duplicate_match_score"]] = df1.apply(
    lambda row: pd.Series(match_apn_and_address(row)),
    axis=1
)

#  Process matched and unmatched records, including Monterey dataset checks
matched_records = []
matched_dwr_ids = set()
unmatched_ocr_records = []

for i in range(len(df1)):
    source_id = df1_ids.iloc[i]
    match_source = df1.at[i, "match_source"]

    if match_source == "Fallback to OCR values":
        matched_in_monterey = "No"
        match_id_mon, score_mon, apn_mon, addr_mon, mon_idx, _ = get_best_match(source_id, df3_records, df1_row=df1.iloc[i], fields_to_check=fields_to_check, field_map=field_map, target_df=df3)
        print(f"NO MATCH: df1_id raw='{df1.at[i, 'well_id']}', cleaned='{source_id}'")
        if match_id_mon:
            matched_in_monterey = "Yes"
        unmatched_record = df1.iloc[i].copy()
        unmatched_record["Matched_in_Monterey"] = matched_in_monterey
        unmatched_ocr_records.append(unmatched_record)

```

```
    else:
        df1_depth = df1.at[i, "depth_well_cleaned"] if "depth_well_cleaned" in df1.columns else np.nan
        df1_water = df1.at[i, "water_static_cleaned"] if "water_static_cleaned" in df1.columns else np.nan
        match_id, score, apn_dwr, addr_dwr, dwr_index, match_score = get_best_match(source_id, df2_records, df1_row=df1.iloc[i], fields_to_check=fields_to_check, field_map=field_map, used_indices=used_dwr_indices, target_df=df2)
        if match_id and dwr_index is not None and dwr_index not in used_dwr_indices:
            logging.debug(f"[Row {i}] Match found for df1_id='{source_id}' to df2_id='{match_id}' (score={score}, duplicate_match_score={match_score})")
            record = df2.iloc[dwr_index].copy()
            record["final_address"] = df1.at[i, "final_address"]
            record["final_apn"] = df1.at[i, "final_apn"]
            record["match_source"] = match_source if match_source else f"ID match (score {score})"
            record["matched_df2_well_id"] = match_id
            record["matched_df1_well_id"] = source_id
            record["duplicate_match_score"] = match_score
            record["df1_row_index"] = i
            matched_records.append(record)
            matched_dwr_ids.add(match_id)
            used_dwr_indices.add(dwr_index)
        else:
            logging.debug(f"[Row {i}] No df2 match for df1_id='{source_id}' or df2 index {dwr_index} already used")
            unmatched_record = df1.iloc[i].copy()
            unmatched_record["Matched_in_Monterey"] = "No"
            unmatched_ocr_records.append(unmatched_record)

    if match_source in [
        "Fuzzy location → APN match",
        "Fuzzy location → fuzzy address",
        "Fuzzy location match with no parcel match"
    ]:
        location_raw = df1.at[i, "location_details"]
        location_clean = clean_for_matching(location_raw)
        dwr_locations_raw = df2["Well Location"].fillna("").astype(str)
        dwr_locations_clean = dwr_locations_raw.apply(clean_for_matching).tolist()

        match_result = process.extractOne(
            location_clean,
            dwr_locations_clean,
            scorer=fuzz.token_sort_ratio
        )

        if match_result:
            match_value, score, match_index = match_result
            if match_index not in used_dwr_indices:
                logging.debug(f"[Row {i}] Best fuzzy match='{match_value}' with score={score}")
```

```
                matched_row = df2.iloc[match_index].copy()
                matched_row["final_address"] = df1.at[i, "final_address"]
                matched_row["final_apn"] = df1.at[i, "final_apn"]
                matched_row["match_source"] = match_source
                matched_row["matched_df2_well_id"] = df2_ids.iloc[match_index]
                matched_row["matched_df1_well_id"] = source_id
                matched_row["duplicate_match_score"] = 0
                matched_row["df1_row_index"] = i
                matched_records.append(matched_row)
                matched_dwr_ids.add(df2_ids.iloc[match_index])
                used_dwr_indices.add(match_index)
            else:
                logging.debug(f"[Row {i}] Fuzzy match index {match_index} already used")
                unmatched_record = df1.iloc[i].copy()
                unmatched_record["Matched_in_Monterey"] = "No"
                unmatched_ocr_records.append(unmatched_record)
        else:
            logging.warning(f"[Row {i}] Fuzzy match expected but no location match found for location='{location_clean}'")
            unmatched_record = df1.iloc[i].copy()
            unmatched_record["Matched_in_Monterey"] = "No"
            unmatched_ocr_records.append(unmatched_record)

# Generate unmatched DWR records (those not matched to OCR)
unmatched_dwr_records = df2[~df2_ids.isin(matched_dwr_ids)]

# Save matched and unmatched records to Excel files
if matched_records:
    matched_df = pd.DataFrame(matched_records)
    matched_df["final_apn"] = matched_df["final_apn"].fillna("").astype(str)
    mask = (matched_df["match_source"] == "No match") & (
        matched_df["final_address"].isna() | (matched_df["final_address"].str.strip() == "")
    )
    matched_df.loc[mask, "final_apn"] = "0000000"

    fallback_lookup = df1.set_index("well_id_cleaned")["match_source"].to_dict()
    matched_df["fallback_match_source"] = ""

    for idx, row in matched_df.iterrows():
        if row.get("match_source") == "ID match with no parcel match":
            well_id = row.get("matched_df1_well_id")
            if well_id in fallback_lookup:
                matched_df.at[idx, "fallback_match_source"] = fallback_lookup[well_id]

    matched_df.to_excel(matched_output, index=False)
    print(f"Matched DWR records written to: {matched_output}")

if unmatched_ocr_records:
    unmatched_ocr_df = pd.DataFrame(unmatched_ocr_records)
    unmatched_ocr_df.to_excel(unmatched_ocr_output, index=False)
```

```
    print(f"Unmatched OCR records written to: {unmatched_ocr_output}")

if not unmatched_dwr_records.empty:
    unmatched_dwr_records.to_excel(unmatched_dwr_output, index=False)
    print(f"Unmatched DWR records written to: {unmatched_dwr_output}")

# Format APN fields in matched and unmatched OCR datasets
matched_df = pd.DataFrame(matched_records)
unmatched_ocr_df = pd.DataFrame(unmatched_ocr_records)

if not matched_df.empty:
    matched_df["final_apn"] = matched_df["final_apn"].apply(format_apn_with_dashes)

if not unmatched_ocr_df.empty:
    unmatched_ocr_df["final_apn"] = unmatched_ocr_df["final_apn"].apply(format_apn_with_dash-
es)

# Supplement matched dataset with OCR fields and source tracking
supplemented_df = matched_df.copy().reset_index(drop=True)

supplemented_df["duplicate_exists"] = ""
for _, source_col in list(field_map.values()) + list(additional_fields.values()):
    supplemented_df[source_col] = ""
for final_col, _ in additional_fields.items():
    supplemented_df[final_col] = np.nan

# create column for the OCR depths flag
supplemented_df["depths_flag"] = ""

columns_to_supplement = [
    (final_col, ocr_col, source_col)
    for final_col, (ocr_col, source_col) in field_map.items()
]
additional_columns = [
    (final_col, ocr_col, source_col)
    for final_col, (ocr_col, source_col) in additional_fields.items()
]

df1["well_id_cleaned"] = df1["well_id_cleaned"].apply(lambda x: str(x).strip().rstrip(".0")
if not pd.isna(x) else "")

for idx, row in supplemented_df.iterrows():
    df1_idx = row.get("df1_row_index")
    if pd.isna(df1_idx):
        continue
    ocr_row = df1.iloc[int(df1_idx)]
    supplemented_df.at[idx, "duplicate_exists"] = ocr_row.get("duplicate_exists", "")
    for match_col, ocr_col, source_col in columns_to_supplement:
        original_val = supplemented_df.at[idx, match_col]
        ocr_val = ocr_row.get(ocr_col, np.nan)
```

```
        if match_col == "B118WellUse":
            dwr_val_str = str(original_val).strip().upper() if isinstance(original_val, str) else ""
            ocr_val_str = str(ocr_val).strip()
            if dwr_val_str == "UNKNOWN" and ocr_val_str.upper() not in {"", "NAN"}:
                supplemented_df.at[idx, match_col] = ocr_val
                supplemented_df.at[idx, source_col] = "AI"
                continue
        if pd.isna(original_val) or (isinstance(original_val, str) and original_val.strip() == ""):
            if not pd.isna(ocr_val) and str(ocr_val).strip() != "":
                supplemented_df.at[idx, match_col] = ocr_val
                supplemented_df.at[idx, source_col] = "AI"
            else:
                supplemented_df.at[idx, source_col] = "DWR"
        else:
            supplemented_df.at[idx, source_col] = "DWR"
    for match_col, ocr_col, source_col in additional_columns:
        ocr_val = ocr_row.get(ocr_col, np.nan)
        if not pd.isna(ocr_val) and str(ocr_val).strip() != "":
            supplemented_df.at[idx, match_col] = ocr_val
            supplemented_df.at[idx, source_col] = "AI"
        else:
            supplemented_df.at[idx, match_col] = np.nan
            supplemented_df.at[idx, source_col] = "AI"


    # Pull the two source columns that belong to the depth fields
    completed_src = supplemented_df.at[idx, "total_completed_depth_source"]
    drill_src     = supplemented_df.at[idx, "total_drill_depth_source"]

    # If either source contains "AI" then copy OCR flag
    if (pd.notna(completed_src) and "AI" in str(completed_src).upper()) or \
       (pd.notna(drill_src)     and "AI" in str(drill_src).upper()):
        ocr_flag = ocr_row.get("depths_flag", "")
        supplemented_df.at[idx, "depths_flag"] = ocr_flag
    else:
        supplemented_df.at[idx, "depths_flag"] = ""   # blank when both are DWR (or empty)

# Flag record type mismatches between DWR and OCR datasets
for idx, row in supplemented_df.iterrows():
    df1_idx = row.get("df1_row_index")
    if pd.isna(df1_idx):
        continue
    ocr_row = df1.iloc[int(df1_idx)]
    dwr_record_type = str(row.get("Record Type", "")).strip().upper()
    ocr_record_type = str(ocr_row.get("Record Type", "")).strip().upper()
    if ocr_record_type in {"", "UNKNOWN"}:
        continue
    if dwr_record_type != ocr_record_type:
```

```
        supplemented_df.at[idx, "Record_Type_Mismatch"] = f"DWR: {dwr_record_type} ≠ OCR:
{ocr_record_type}"

#Add fallback match source for records with no parcel match
supplemented_df["fallback_match_source"] = ""
for idx, row in supplemented_df.iterrows():
    if row.get("match_source") == "ID match with no parcel match":
        df1_idx = row.get("df1_row_index")
        if not pd.isna(df1_idx):
            fallback_val = df1.iloc[int(df1_idx)].get("match_source", "")
            supplemented_df.at[idx, "fallback_match_source"] = fallback_val

# Save supplemented dataset to Excel
supplemented_df.to_excel(supplemented_output_path, index=False)
print("supplemented file written to:", supplemented_output_path)
```

C:\Well Layer Update\Python\Azure\County Database Matching Script.py

```python
import pandas as pd
import re
import logging




# df4_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\C1.xlsx"
# df5_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\excel 3\OCR(SCC1)_OCR(DWR1)2.xlsx"
# matched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\SCC1_DWR1_C1_2.xlsx"
# unmatched_county_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\C2.xlsx"
# unmatched_source_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR(SCC1)_OCR(DWR1)_Final.xlsx"

#c3 and dwr2
df4_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR2_C2\C3.xlsx"
df5_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\DWR2.xlsx"
matched_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\C3_DWR2\C3_DWR2.xlsx"
unmatched_county_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\C3_DWR2\C3.xlsx"
unmatched_source_output = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\C3_DWR2\DWR3.xlsx"

df4 = pd.read_excel(df4_path)
df5 = pd.read_excel(df5_path)


def clean_id(raw):
    if pd.isna(raw):
        return ""
    original = str(raw).strip()
    # Replace 'O' with '0' and 'e' with 'E'
    cleaned = original.replace("O", "0").replace("o", "0").replace("e", "E").upper().strip()
    if cleaned.lower().startswith('c'):
        cleaned = 'E' + cleaned[1:]
    if re.match(r"^WCR\d{4}-\d{6}$", cleaned):
        return cleaned
    text = original.lower().replace("O", "0").replace("o", "0").replace("e", "E").replace(",", "").replace("no.", "").replace("/", " ")
    text = re.sub(r"[\s\n]+", " ", text)
    match = re.match(r"^(\d{1,2})[\.-](\d{1,2})[\.-](\d{2,4})(\d{3,8})$", text)
    if match:
```

```python
        year_str = match.group(3)
        try:
            year = int(year_str)
            if year < 100:
                year += 1900 if year >= 35 else 2000
            if 1935 <= year <= 2025:
                return match.group(4)
        except ValueError:
            pass
    prefix_removed = re.sub(r"^\d{1,2}[\.-]\d{1,2}[\.-]\d{2,4}", "", text).strip()
    segments = re.split(r"[\s]+", prefix_removed)
    for segment in reversed(segments):
        segment_digits = re.findall(r"\d{3,8}", segment)
        if segment_digits:
            return segment_digits[-1].lstrip('0')
    digits = re.findall(r"\d{3,8}", text)
    if digits:
        return digits[-1].lstrip('0')
    if cleaned.startswith('E') and any(c.isdigit() for c in cleaned[1:]):
        return cleaned
    return ""


logging.basicConfig(
    filename=r"C:\Well Layer Update\GIS\master files\county_match_debug.log",
    filemode="a",
    format="%(asctime)s [%(levelname)s] %(message)s",
    level=logging.DEBUG
)


df4_ids = df4["Log_Number"].apply(clean_id)
df5_legacy_ids = df5["Legacy Log Number"].apply(clean_id)  # Apply clean_id for consistency
df5_wcr_ids = df5["WCR Number"].apply(clean_id) if "WCR Number" in df5.columns else pd.Series([""] * len(df5))


matched_records = []
unmatched_county_records = []
unmatched_source_records = []
used_df5_indices = set()

for i, county_id in enumerate(df4_ids):
    if not county_id.strip():
        continue  # Skip empty Log_Number
    county_id_clean = str(county_id).strip().rstrip('M').replace('.0', '')
    county_id_normalized = 'E' + county_id_clean[1:].lstrip('0') if county_id_clean.startswith('E') else county_id_clean.lstrip('0')

    # Debug specific case
```

```
    if county_id_clean.lower() == 'e0139534' or county_id_clean == 'E0139534':
        logging.debug(f"Debugging e0139534: county_id_clean='{county_id_clean}', county_id_normalized='{county_id_normalized}'")

    match_found = False
    matched_field = ""
    df5_row = None
    df4_row = df4.iloc[i]

    # try matching with Legacy Log Number
    for j, legacy_id in enumerate(df5_legacy_ids):
        if j in used_df5_indices:
            continue
        legacy_id_clean = str(legacy_id).strip().rstrip('M').replace('.0', '')
        legacy_id_normalized = 'E' + legacy_id_clean[1:].lstrip('0') if legacy_id_clean.startswith('E') else legacy_id_clean.lstrip('0')


        if county_id_normalized == legacy_id_normalized:
            logging.debug(f"Match found: county_id='{county_id}' to source_legacy_id='{legacy_id}'")
            df5_row = df5.iloc[j].copy()
            matched_field = "Legacy Log Number"
            used_df5_indices.add(j)
            match_found = True
            break

    #if no match on Legacy Log Number, try WCR Number
    if not match_found and "WCR Number" in df5.columns:
        for j, wcr_id in enumerate(df5_wcr_ids):
            if j in used_df5_indices:
                continue
            wcr_id_clean = str(wcr_id).strip().rstrip('M').replace('.0', '')
            wcr_id_normalized = 'E' + wcr_id_clean[1:].lstrip('0') if wcr_id_clean.startswith('E') else wcr_id_clean.lstrip('0')

            if county_id_normalized == wcr_id_normalized:
                logging.debug(f"Match found: county_id='{county_id}' to source_wcr_id='{wcr_id}'")
                df5_row = df5.iloc[j].copy()
                matched_field = "WCR Number"
                used_df5_indices.add(j)
                match_found = True
                break

    if match_found:
        # Append all df4 columns with 'county_' prefix
        df4_row_prefixed = df4_row.rename(lambda x: f"county_{x}")
        # Create record with df5 columns, df4 columns, and match metadata
        record = pd.concat([df5_row, df4_row_prefixed])
        record["matched_field"] = matched_field
```

```python
        record["cleaned_county_id"] = county_id
        record["source_id"] = legacy_id if matched_field == "Legacy Log Number" else wcr_id
        matched_records.append(record)
    else:
        unmatched_county_records.append(df4_row)


for j in range(len(df5)):
    if j not in used_df5_indices:
        df5_row = df5.iloc[j].copy()
        # Collect only unmatched df5 records
        unmatched_source_records.append(df5_row)
        logging.debug(f"Unmatched df5 record added: index={j}, source_id='{df5_row.get('Legacy Log Number', 'N/A') if df5_row.get('Legacy Log Number') else df5_row.get('WCR Number', 'N/A')}'")

# save matched
if matched_records:
    matched_df = pd.DataFrame(matched_records)
    matched_df.to_excel(matched_output, index=False)
    print(f"Matched records written to: {matched_output}")
else:
    print("No matched records to write to matched output.")

# save unmatched
if unmatched_county_records:
    unmatched_county_df = pd.DataFrame(unmatched_county_records)

    if 'IFILE' not in unmatched_county_df.columns:
        unmatched_county_df['IFILE'] = pd.NA
    unmatched_county_df.to_excel(unmatched_county_output, index=False)
    print(f"Unmatched county records written to: {unmatched_county_output}")
else:
    print("No unmatched county records with non-empty Log_Number.")

# save sources
if unmatched_source_records:
    unmatched_source_df = pd.DataFrame(unmatched_source_records)
    unmatched_source_df.to_excel(unmatched_source_output, index=False)
    print(f"Unmatched source records written to: {unmatched_source_output}")
else:
    print("No unmatched source records.")
```

C:\Well Layer Update\Python\Azure\Field Highlighting Script.py

```python
from openpyxl import load_workbook
from openpyxl.styles import Font


supplemented_path = r"S:\Water Resources\1_Water Resources Program\GIS\Well Layer Update\Python\Outputs\ocr11\master files 2\OCR(SCC1)_OCR(DWR1)_Final.xlsx"

# Updated field_map to include additional fields
field_map = {
    "Owner Assigned Well Number": "owner_assigned_source",
    "Permit Number": "permit_number_source",
    "Driller Name": "driller_name_source",
    "Total Drill Depth": "total_drill_depth_source",
    "Total Completed Depth": "total_completed_depth_source",
    "Static Water Level": "static_water_level_source",
    "Well Yield": "well_yield_source",
    "B118WellUse": "b118welluse_source",
    "Record Type": "record_type_source",
    "Permit Date": "permit_date_source",
    "Driller License Number": "driller_license_number_source",
    "Drilling Method": "drilling_method_source",
    "Fluid": "fluid_source",
    "Total Draw Down": "total_drawdown_source",
    "Test Type": "test_type_source",
    "Pump Test Length": "pump_test_length_source",
    "Other Observations": "other_observations_source",
    "City": "city_source",
    "Ground Surface Elevation": "surface_elevation_source",
    "Elevation Determination Method": "elevation_method_source",
    "Top Of Perforated Interval": "top_perforations_source",
    "Bottom of Perforated Interval": "bottom_perforations_source",
    "Decimal Latitude": "latitude_source",
    "Decimal Longitude": "longitude_source",
    "APN": "apn_source",
    "seal_cleaned": "seal_cleaned_source",
    "first_water_cleaned": "first_water_cleaned_source",
    "template_id": "template_id_source",
    "OCR_match_source": "OCR_match_source_source"
}

# Load workbook and sheet
wb = load_workbook(supplemented_path)
ws = wb.active


headers = {cell.value: idx for idx, cell in enumerate(ws[1], 1)}  # 1-based indexing

red_font = Font(color="FF0000")
```

```python

# Get indexes for special fields
final_apn_idx = headers.get("final_apn")
final_address_idx = headers.get("final_address")
match_source_idx = headers.get("match_source")

# Loop through rows starting from row 2 (excluding header)
for row in ws.iter_rows(min_row=2, max_row=ws.max_row):
    # Highlight special logic for final_apn and final_address
    match_source_val = row[match_source_idx - 1].value if match_source_idx else None

    if final_apn_idx:
        apn_cell = row[final_apn_idx - 1]
        if match_source_val not in ["DWR APN matched to parcel database", "Fuzzy DWR address match", "No match"]:
            apn_cell.font = red_font  # Mark as red if not in accepted match sources

    if final_address_idx:
        addr_cell = row[final_address_idx - 1]
        if match_source_val not in ["DWR APN matched to parcel database", "Fuzzy DWR address match"]:
            addr_cell.font = red_font

    #Apply red font to "AI"-sourced fields
    for field, source_col in field_map.items():
        col_idx = headers.get(field)
        source_idx = headers.get(source_col)

        if not col_idx or not source_idx:
            continue

        source_val = row[source_idx - 1].value
        if source_val == "AI":
            row[col_idx - 1].font = red_font  # Highlight the main field in red

# Save with highlights
wb.save(supplemented_path)
print("Field highlighting complete.")
```

Introduction

Santa Cruz County's water resources serve a critical role in providing municipal, domestic, and agricultural water supply, preserving fragile watersheds, providing resilient habitats, and supporting recreational and commercial activities. Nearly all water supplies are derived from local rainfall and captured through stream diversions and groundwater wells (Figure 1). County staff, local agencies, organizations, and the community continue to collaborate on long-term, adaptive solutions to sustain environmental quality and ensure safe, reliable, and affordable water resources for current and future needs.

Projects and planning efforts undertaken by regional agencies and non-profits are preparing for a future with a greater variability in precipitation and temperature than Santa Cruz experienced when our existing infrastructure was built, as well as a growing population. This report encompasses activities that took place during calendar year 2025 and reflects water use and rainfall from the 2025 water year which began October 1, 2024 and ended September 30, 2025.

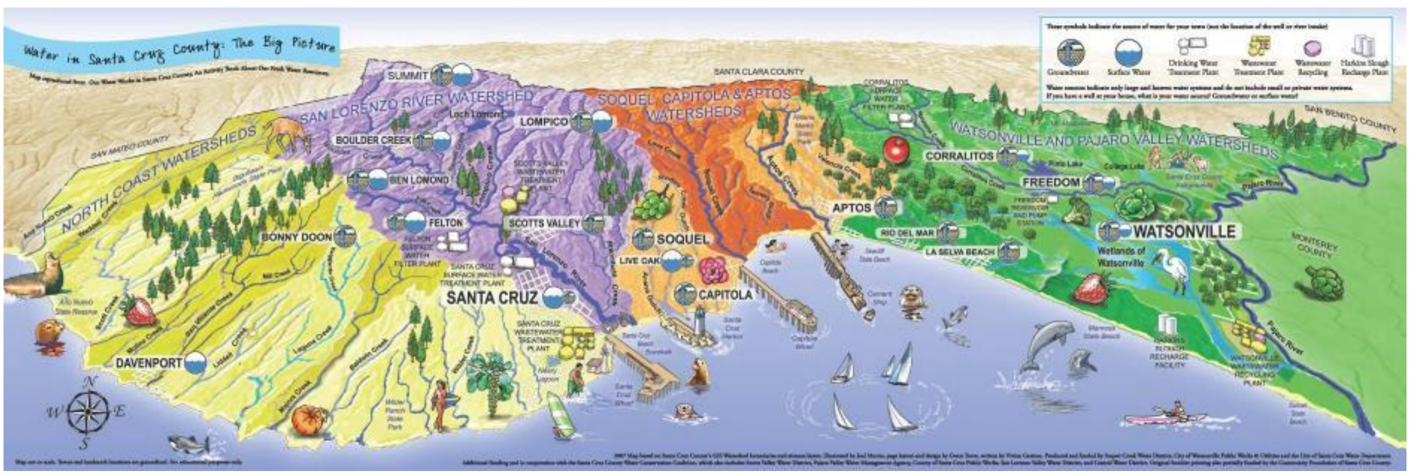


Figure 1: Local Water Supply Distribution.

Water Resource Management activities during 2025 were influenced by:

- Changing regulations: Many water systems in the County have been impacted by new water quality regulations that are being rolled out.
    - California has adopted a new Hexavalent Chromium (referred to as Chromium-6, best known from the movie *Erin Brockovich*) limit of 10 parts per billion. This is the lowest limit in the nation and possibly in the world. Mid- and South-Santa Cruz County have naturally occurring Chromium-6 above this new standard, prompting water systems to send notifications to customers and install expensive treatment systems.
    - Water systems are now required to monitor for per- and polyfluoroalkyl substances – known as PFAS. Currently no treatment is required, though notification to customers is required when established limits are exceeded. Drinking water limits will be enforced for all systems in 2029. The first treatment

plant in Santa Cruz County was successfully completed this year at the Buena Vista Migrant Center.

- Low Rainfall: Water year 2025 was average for California as a whole, but dry for Santa Cruz County. Rainfall averages at gauges in Ben Lomond and the City of Santa Cruz were just over 60% of the long-term average, and the lowest they have been since 2021 (Figure 2). Despite this, Loch Lomond, the only surface water reservoir in the County, was full at the end of the rainy season and had only dropped 7% below full at the end of the water year.

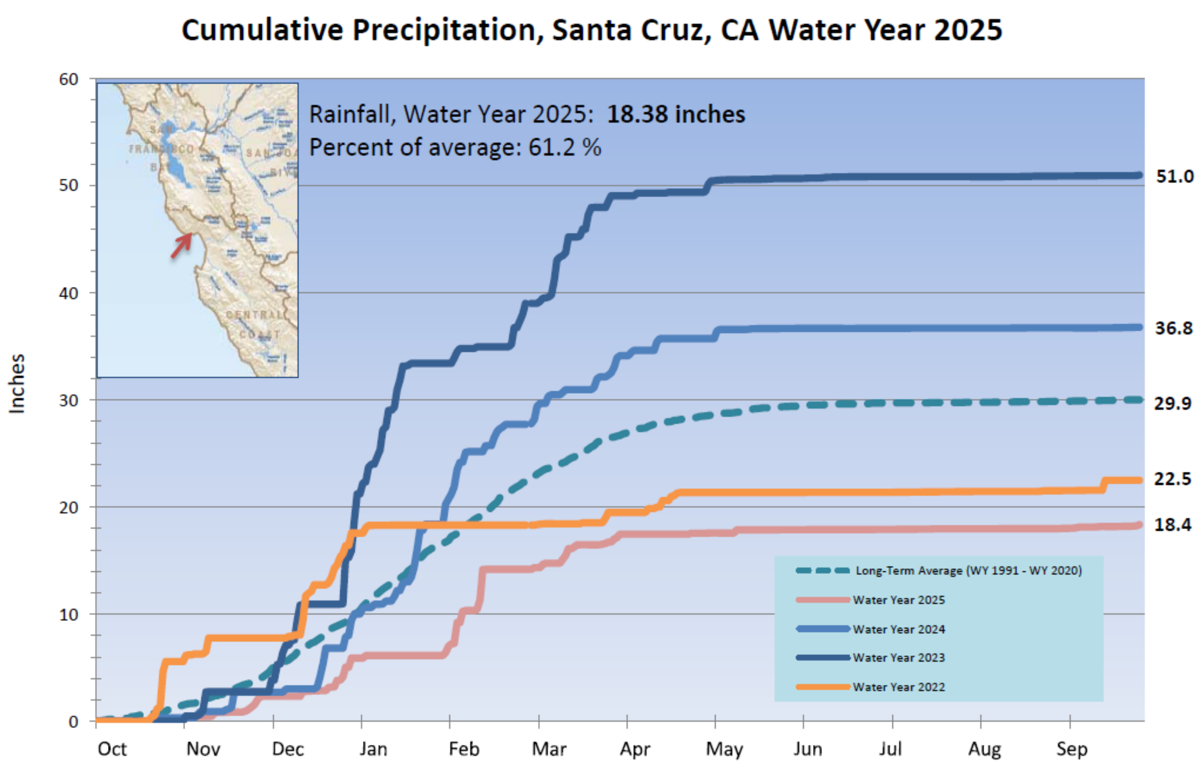


Figure 2: Rainfall in water years 2022-2025 compared to the long-term average, data from CIMIS, credit City of Santa Cruz Water Department.

- Low Temperatures: Despite the low rainfall, streamflow in the San Lorenzo River at Big Trees ended the year slightly above the long-term median (Figure 3). How can this be? Firstly, the waterways in the County have benefited from several wet years in a row, which make them more resilient to a single dry year. But secondly, and likely more important, Santa Cruz experienced a cool Summer.

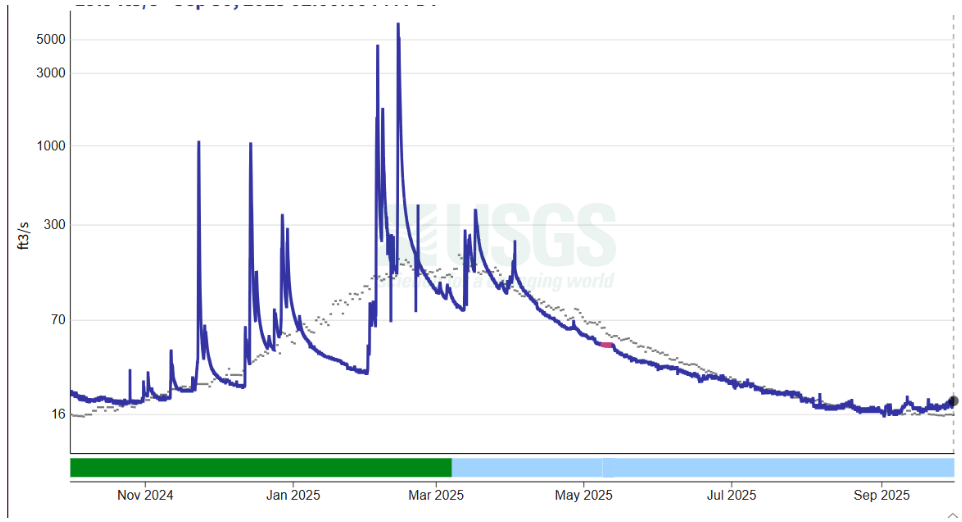


Figure 3: Streamflow in cubic-feet per second at Big Trees on the San Lorenzo River for Water Year 2025, shown in front of the long-term median. Credit USGS.

Year	**Maximum Average Temperature (°F) for June**
2025	60.2
2024	65
2023	61.04
2022	71.58
2021	69.13
2020	67.21

June is a good month to review when considering temperature impacts on water supplies. That is because the days in June are so long that high temperatures can have a dramatic impact on water resources through evapotranspiration from vegetation, and increased water demand for irrigation and other efforts to keep cool. Temperature records taken at the Watsonville Airport show that the maximum average temperature for June 2025 was 60.2 Fahrenheit (Table 1).

Table 1: Maximum Average Temperature for June 2025 at the Watsonville Airport, Weather Underground.

Key highlights for the year include:

- Municipal water use remains lower than it was when recent drought restrictions were in place, and is around 23% below the water use levels of the early 1980s.
- The Pajaro Valley Water Management Agency (PV Water) celebrated a ribbon cutting for the College Lake project.
- Implementation of Measure Q kicked off as the County and local partners hit the ground running. The Land Trust of Santa Cruz County was selected to be the land stewardship recipient of funds, the Citizens Oversite Advisory Board (COAB) was selected, and the Vision Plan adopted.
- The County completed the first Multi-Jurisdictional Hazard Mitigation plan.

- Groundwater basins continued to benefit from direct recharge and lower than average groundwater pumping due to ample surface water and cooler temperatures.
- The City of Santa Cruz and Scotts Valley Water District have nearly completed an intertie connecting their two water supplies. This intertie was first conceived of over a decade ago but the funding was not available until a grant from the State. The intertie is a key step for water supply resiliency and groundwater management.

Coordination between water practitioners is critical to regional resilience. This report was written by staff in the Santa Cruz County Water Resources Program in Environmental Health. Updates were provided by four County departments and by partners throughout the county including the San Lorenzo Valley Water District, Scotts Valley Water District, City of Santa Cruz Water Department, Soquel Creek Water District, Central Water District, City of Watsonville, Pajaro Valley Water Management Agency, Regional Water Management Foundation, Pajaro River Flood Management Agency, and the Resource Conservation District of Santa Cruz County.

Table of Contents

List of Figures

Section 1: Regional Water Supply Resilience

This section focuses on efforts by the County, municipal water providers, Groundwater Sustainability Agencies, and non-profit organizations to shore up existing water supplies and infrastructure, manage existing resources appropriately, and develop new water supplies.

Like much of California, Santa Cruz County is projected to face intensifying weather swings from extreme dry conditions to extreme wet conditions. These weather swings will be experienced as longer, more frequent, and more intense droughts, that are punctuated by more extreme rain events. We have already begun to see this change with the 2022-23 Atmospheric River events, which dropped record amounts of rain on our county, but had been preceded by three years of drought conditions. This change in rainfall patterns is likely to cause increasing stress on local water resources.

To meet this challenge, county residents and agencies have continued their efforts to limit water waste. In particular, the large water suppliers have been tremendously successful in reducing demand for water in the county. These agencies have achieved this through multiple methods, which include;

1. Utilizing smart metering, which can notify residents immediately when a leak occurs
2. Financially supporting residents to replace water intensive fixtures and irrigation, such as toilet and lawn replacement rebates.
3. Education campaigns to educate residents about water conservation, such as through the WaterSavingTips.org website.

The success of these efforts can be seen when comparing the number of water connections in the county to water production. Since 1984, the large water systems have increased the number of connections by approximately 32%, but their annual water production has decreased by 23% (see Figure 4). The decoupling of population growth and water demand began in the late 1990's and reflects higher standards for the water consumption of indoor fixtures and appliances. However, the trend took off after an exceptionally dry 2013 led to State of Emergency declaration. This resulted in many policy changes locally, and encouraged a mentality of water conservation among agencies and residents.

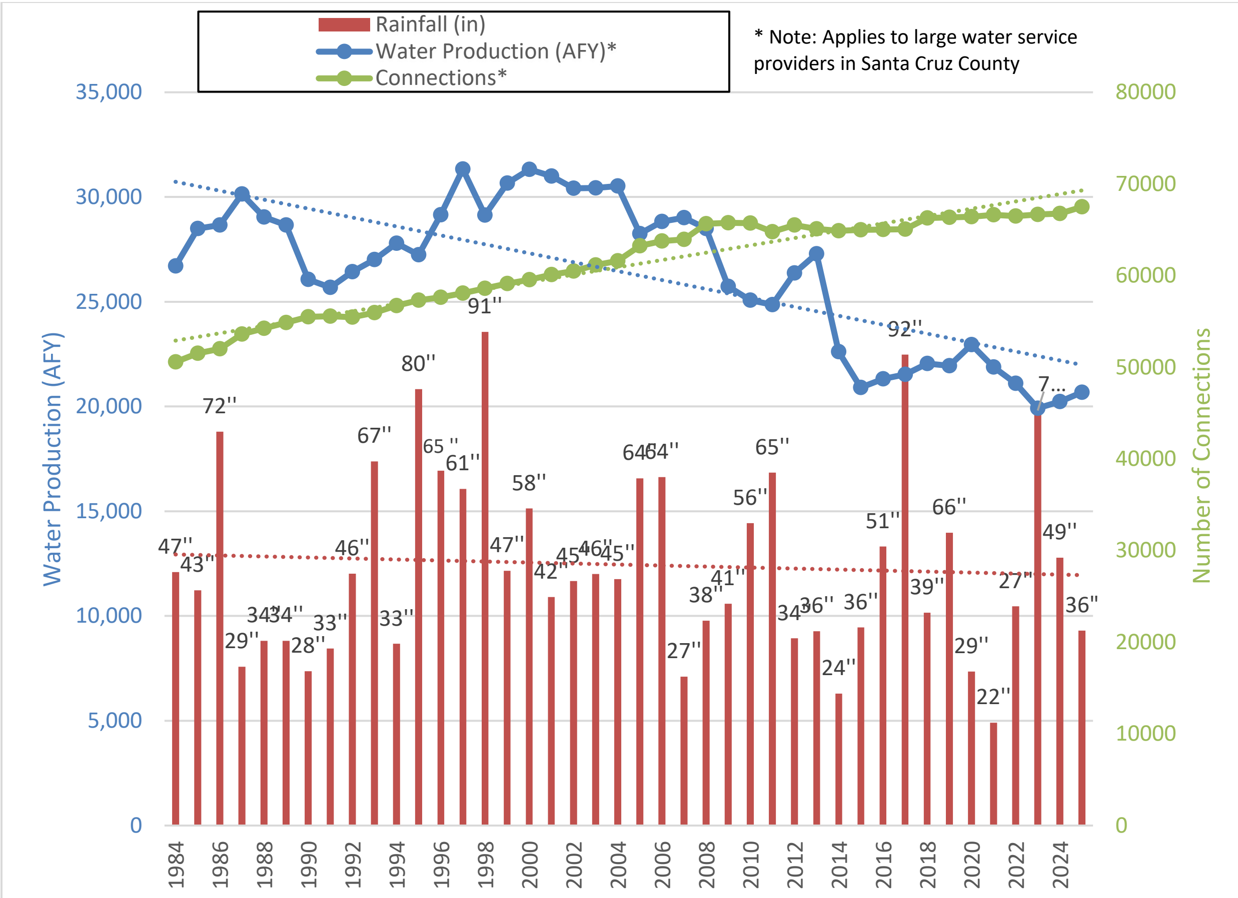


Figure 4: Water Production and Connections for Large Water Systems (1984-present), with Rainfall Data

Areas for Continued Improvement

As indoor fixtures and appliances have become increasingly efficient, outdoor water use continues to be large portion of residential water use. According to the California Department of Water Resources, outdoor water use accounts for roughly half of household usage[1]. Outdoor water use is extra challenging to our local water resources because it increases during drier times of the year, when water sources are most stressed. This is illustrated in Figure 5, which shows the 2024 water extraction data of small Community water systems (<200 connections) in Santa Cruz County. The peak water demand in August was double that of April, which can largely be attributed to increased outdoor water use. To reduce stress on local water resources, reducing outdoor water use should be a priority.

[1] Water Use in California's Communities - Public Policy Institute of California

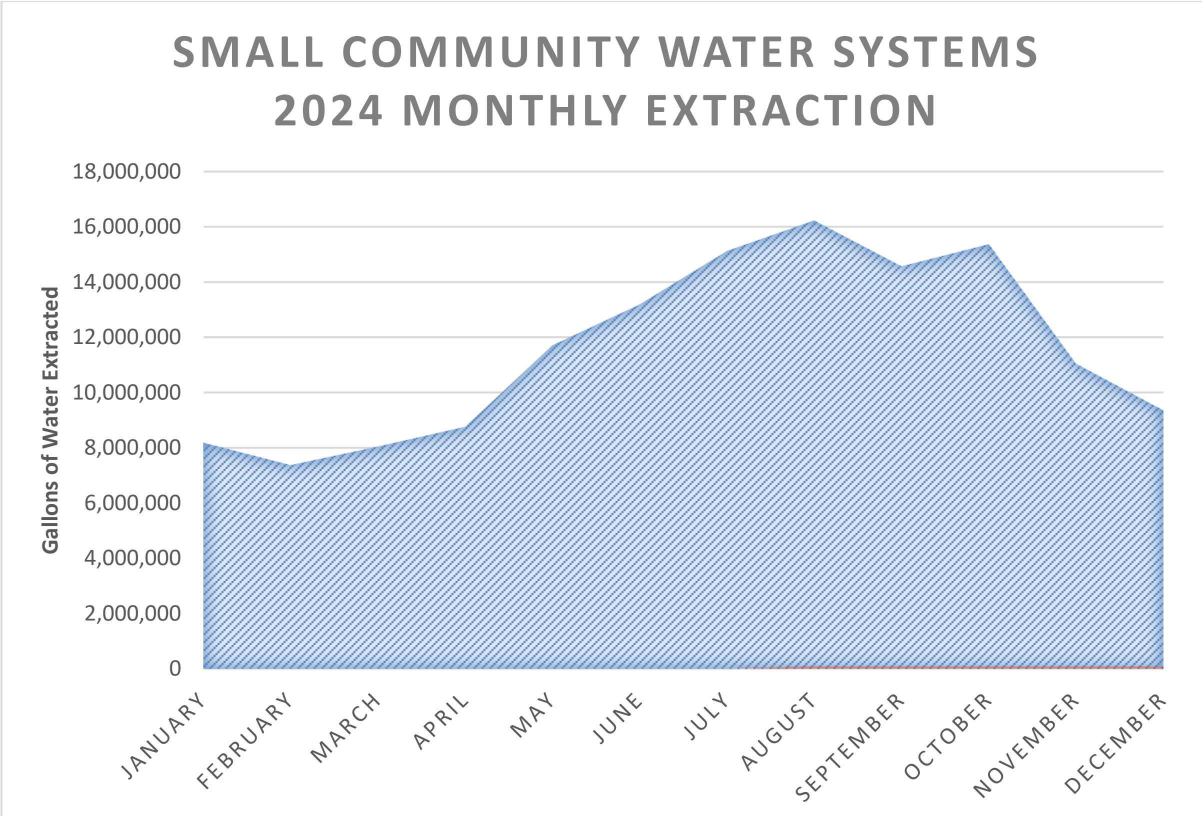


Figure 5: Small Community Water Systems 2024 Monthly Use, demonstrating an increase during the summer months.

It is important to note that these Small Community Water Systems are much more likely to serve larger parcels, with significant outdoor water use. The large water suppliers, which tend to serve more urbanized areas, water usage does not increase as much in the dry season. Figure 6 below shows the per capita daily usage of the customers of the county's five largest water suppliers and compares it with the average for Small Community Water systems. Not only do the large water suppliers have a lower winter baseline, but they also have significantly smaller increases in the summer. There are many factors that can affect per capita water usage[2], but a major driver is larger parcels with more irrigated space.

[2] State Water Resources Control Board: Factors that can affect per capita water

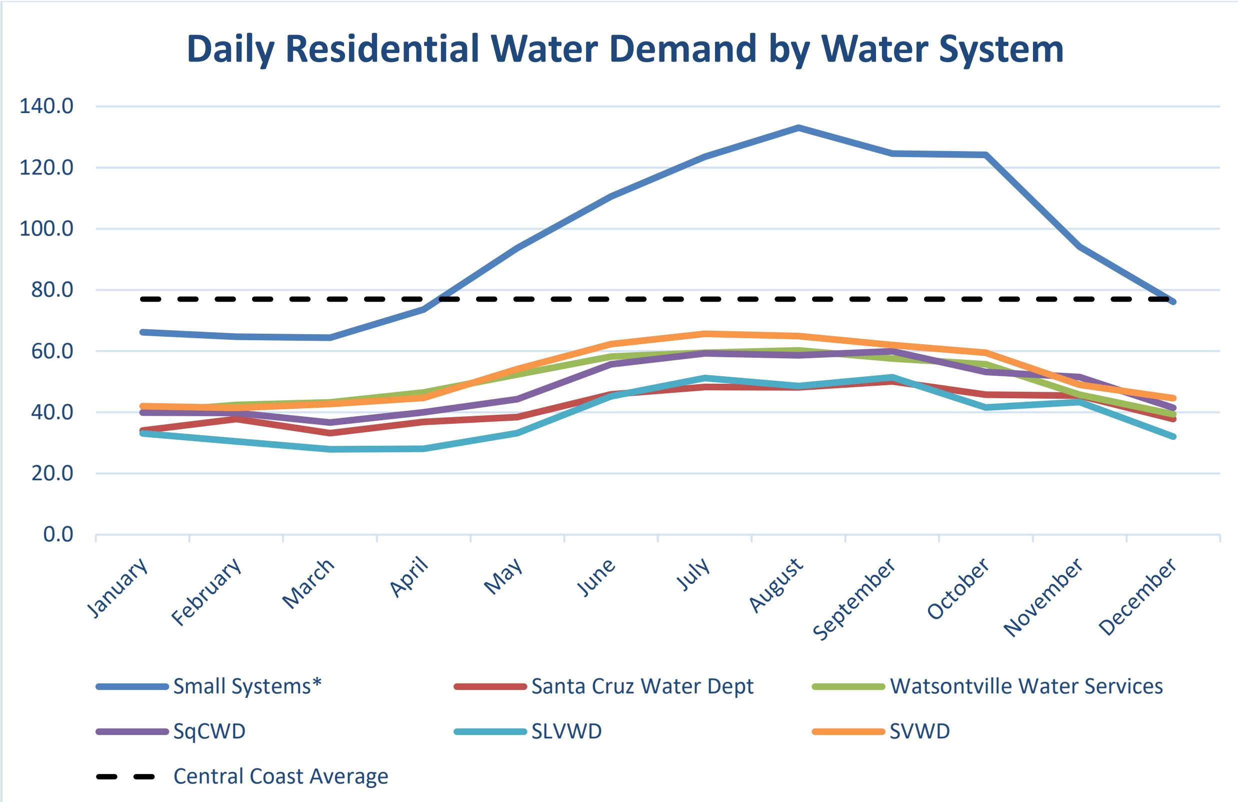


Figure 6: Per capita daily usage of the customers of the county's five largest water suppliers and compares it with the average for Small Community Water systems

The data presented here are averages and all systems will have more and less efficient users; however low-density development has greater potential for high per capita water demand. According to the SPUR report "Water for a Growing Bay Area"[3], increasing density can allow for more homes without using more water. The report identifies two reasons for this:

1. **Infill development often occupies space that was already irrigated.** Infill development, such as adding an accessory dwelling unit (ADU), converting a single-family home to a multiplex, or splitting a single lot into two will not increase the possible space irrigated, but can reduce it.
2. **New construction tends to be more efficient than older buildings.** New construction needs to comply with current building code standards and efficient landscaping requirements. Older buildings also tend to accumulate plumbing leaks over time.

The County continues to encourage more infill development, such as multi-family housing and ADUs, and through actions like the Housing Element Rezone Program[4].

For municipal water providers, water supply projects and activities encompass two primary categories. The first is using existing water sources efficiently by incentivizing low water use,

---

[3] Water for a Growing Bay Area | SPUR

[4] Housing Element Rezone Program

reducing leaks, upgrading infrastructure, and using new tools to reduce irrigation water needs. Santa Cruz County is one of the most efficient counties in the state when it comes to per capita water use, and incentives and messaging by the larger water suppliers continue to encourage efficiency. The second supply category is the creation of "new" water supplies through projects like groundwater recharge and wastewater recycling, as well as optimizing the timing of using existing surface and groundwater resources to increase water in storage.

Efficiency in Existing Water Supply – Municipal Suppliers

Water agencies are undertaking the following activities to improve the efficiency of current water supplies:

San Lorenzo Valley Water District (SLVWD)

- SLVWD adopted its 2025-27 budget and Capital Improvement Program and is working on implementing the projects. A number of pipeline and tank upgrades, fire prevention projects, and the construction of 3 new water tanks will occur in fiscal year 25/26. All project details are available on the SLVWD's website **here**.
- Construction of improvements necessary for consolidation of Bracken Brae and Forest Springs is ongoing pursuant to grant funding.
- The District maintains a Stage 1 water shortage designation due to unrestored water intakes and pipelines.
- SLVWD actively pursues incidents of water waste by investigating, recommending corrective action, and providing follow-up documentation of resolution. As of October 2023, ~50% of the meters have been upgraded. The new meters, combined with the Badger Eye on Water engagement portal, allow the customers to view hourly usage history, and set up leak detection alerts and high bill notifications.
- In Fiscal Year 2025/2026 SLVWD issued 32 rebates for Energy Star rated washing machines, low-flow toilets, and weather-based irrigation controller installations. SLVWD conducts a variety of public education activities such as a dedicated Water Use Efficiency Page on its website, e-Newsletters, billing inserts, Instagram and Facebook postings.
- SLVWD hired new General Manager Jason Lillion in July 2025.

Scotts Valley Water District (SVWD)

- SVWD continued assisting City of Scotts Valley in assessing the condition of the Scotts Valley Tertiary Treatment Plant and finding a mutually advantageous solution for wastewater operations, which provides the source of recycled water used throughout the City.
- SVWD utilized WaterSmart customer engagement portal for leak notification procedures and achieving continued reduction in the volume of water lost through leaks. Achieved 68% registration rate at WaterSmart in 2025.
- SVWD continued Think Twice Water Use Efficiency Program in response to the Stage 1 Water Condition. Program activities included 2x Turf Rebate, and Pool Cover Rebate.

- The SVWD is also working on rehabilitation of Bethany tank site, and completed design of structural modifications at Glenwood tank to prevent undermining of the tank foundation in reaction to a landslide that occurred below tank site during the winter storms of 2023.
- Completed construction of 1,450 LF 8" main on La Cuesta Dr. This main connected a new critical loop in SVWD's distribution system.

City of Santa Cruz Water Department (SCWD)

- In June 2025, the City received an order from the California State Water Resources Control Board approving its petitions for water rights changes and issuance of amended licenses and permits. The amended water rights add flexibility to the location of water diversions and broaden the places of use of the water. They also incorporate changes to instream flow requirements that are essential for the survival of coho salmon and steelhead trout, while also providing the operational flexibility needed to implement the City's Anadromous Salmonid Habitat Conservation Plan while maintaining long-term water supply reliability and supporting sustainable regional water resource management.
- Continuing planning for major infrastructure improvements at the Graham Hill Water Treatment Plant (GHWTP) under an innovative design-build framework. The GHWTP Facility Improvements Project is expected to begin construction in 2027. Ongoing construction of the Concrete Tanks Replacement Project as a predecessor project to the larger infrastructure improvements at the GHWTP, project is wrapping up, with construction expected to be completed in early 2026.
- Began construction of the Newell Creek Pipeline: Felton to Graham Hill Segment. The Newell Creek Pipeline provides the critical pipeline connection between Loch Lomond Reservoir and the Graham Hill Water Treatment Plant. The projects is anticipated to to have construction completed in the spring of 2027.
- Replacement of the City's University 4 Tank was initiated in 2025. Construction, which includes replacement of the existing 400,000 gallon tank, installation of a new maintenance tank, and other facility improvements, is expected to be completed in later 2026.
- Improvements at the City's Beltz 12 well facility to address water quality issues with native groundwater are also underway. The treatment system upgrades will allow for on-site treatment of ammonia by increasing chlorine contact time in a new pressure vessel. Construction is expected to be completed in 2026.

The Soquel Creek Water District (SqCWD)

- SqCWD continues a robust conservation program including: a large variety of indoor and outdoor rebates, a landscape budget tool for commercial landscapes, high water use diagnostics via phone, and free water saving devices like hose nozzles, faucet aerators, and low flow showerheads. The WaterSmart Customer Portal provides customers with their digital meter's daily and hourly water use, notifies them of potential leaks, and helps them diagnose the potential cause of high use. In addition to the WaterSmart Portal, staff assist customers by providing various tools (e.g., the Leak Guide, phone diagnostics, technician visits, etc.) to help them locate and resolve leaks. In 2025, average residential consumption was approximately 48 gallons per person per day.

- Work continued on several components of the $7.6 million Sustainable Groundwater Management Act Implementation Grant awarded to the Santa Cruz Mid-County Groundwater Agency that focus on SqCWD infrastructure and/or collaboration, including the:
    - Design and construction of a groundwater extraction well on Cunnison Lane in Soquel. The new well will improve redundancy and flexibility and help redistribute groundwater pumping further inland.
    - Completion of the Park Avenue transmission main/bottleneck improvements to increase system reliability and allow more flexibility to redistribute pumping inland away from coastal wells. - Completed at the end of 2024.
    - Regional Water Resources Optimization Study. In collaboration with the City of Santa Cruz, this project modeled and analyzed various implementations of select programs and management actions as identified in the Basin's Groundwater Sustainability Plan. - Completed in 2025.

Central Water District (CWD)

- The Central Water District is continuing its efforts to install a new well that will ensure water resilience for current and future needs.
- CWD is continuing to replace or update its water storage tanks located throughout the water district's boundaries.
- The District has actively been clearing plant overgrowth and debris for the purpose of reducing fire fuel loads at the District Office and at pertinent tank sites. The District has been working closely with the California Conservation Corp to complete these important projects.
- Central Water District customers continue to demonstrate commitment to ongoing conservation efforts by maintaining over 40% reduction in water consumption compared to the District's highest historical water usage.
- The District continues efforts to replace aging meters with new technology meters to allow customers more control over their water consumption. To date, over a quarter of the District's meters have been replaced, with plans to complete the remaining replacements within the next five years.

The City of Watsonville Public Works (CoW):

Figure 7: City of Watsonville educational brochure

- The City of Watsonville Water Division anticipates completion of a new 2.4MG water storage tank in December 2025. This tank will help the City maintain and continue to provide its customers with safe and reliable potable water. It will also provide emergency storage in the event of a catastrophic event such as an earthquake, drought, or in case of failure of the neighboring existing tank. The RWMF, in collaboration with, was awarded $5 million from California's Urban and Multibenefit Drought Relief Grant Program to fund the tank construction.
- CoW continues with construction of the pump station for its recently drilled well. This new well and pump station will supplement existing sources and maintain the water system's high level of reliability. This project is expected to be completed in Fall 2026.
- CoW continues to invest in its water main replacement program, achieving more than 2 miles of upgrades in 2024-25.
- CoW completed a rehabilitation of one of its existing welded steel water storage tanks. This included recoating the tank interior along with construction modifications and seismic retrofits to the supply piping, discharge piping, and overflow piping.
- City of Watsonville residents and businesses continue to receive water conservation education and outreach provided by the City's Outreach Team via the CoW's website, social media, newsletter, events, workshops and in-person presentations. CoW also continues to offer conservation devices and financial incentives to encourage conservation. See Figure 7.

Regional Project Funding

- The Regional Water Management Foundation (RWMF) is providing grant administration and acting as coordinator on two IRWM implementation grants awarded to the RWMF on behalf of the Santa Cruz Region.
- Proposition 1 IRWM Implementation Grant Program Round 1 award is funding three projects that collectively benefit water supply, water quality, watershed stewardship, stormwater and flood management and habitat restoration. All three projects were successfully completed in 2025.

Project Title	Lead Agency	Grant Award	Schedule
Countywide Sediment Reduction from Developed Parcels & Rural Roads	Resource Conservation District Santa Cruz County	$785,657	2020 – 2025
Davenport Water Supply Tank	County of Santa Cruz, Davenport County Sanitation District	$556,254	2020 – 2025

Watsonville Slough Farms Wetland Restoration	Resource Conservation District Santa Cruz County	$333,144	2021 – 2025

- Work on the Proposition 1 IRWM Implementation Grant Program Round 2 grant award began in Fall 2023. The projects provide benefits to water supply, water quality, watershed stewardship, stormwater and flood management, habitat restoration, and climate change response.

Project Title	Lead Agency	Grant Award	Schedule
Fire Hardening of Critical Water Supply Infrastructure	San Lorenzo Valley Water District	$305,000	2023 – 2027
Equalization Tank Replacement	County of Santa Cruz, Boulder Creek County Sanitation District	$405,312	2023 – 2026
Decision-Support Tool - Understanding Climate Influenced River Flooding	City of Santa Cruz, Department of Public Works	$179,375	2023 – 2025
Recreational Vehicle Sewage Disposal Station	City of Santa Cruz, Department of Public Works	$85,000	2023 – 2026
Atkinson Lane Integrated Flood Management and Watershed Restoration	City of Watsonville	$545,000	2023 – 2027
Drinking Water Treatment System & Secondary Water Source Rountree Facility	County of Santa Cruz, General Services	$800,000	2023 – 2026

Efficiency in Existing Water Supply – Agriculture

Ag Irrigation Efficiency Assistance

The RCD continues to assist growers with conserving water through improved irrigation efficiency and irrigation water use management, leveraging funds from PV Water, CDFA, CA FarmLink, and NRCS. Assistance with nitrogen management (related to water quality protection) is also often incorporated into the irrigation efficiency assistance. During 2025, RCD assisted 29 farming operations at 26 different sites to monitor and improve irrigation scheduling to achieve water conservation. The RCD also provided:

- irrigation system evaluations to identify operation and design improvement opportunities.
- irrigation system design recommendations to optimize irrigation efficiency.
- season-long monitoring of water applied, weather data, and soil moisture to inform growers of how the amount of water applied to their crops compares to the amount of water required by their crops,

- irrigator trainings in English and Spanish
- new online tools and resources available in English and Spanish targeted to provide growers with practical resources for use in the field.
- Limited, targeted outreach to increase grower participation in and awareness of water saving programs that are available.
- technical and financial assistance to implement more efficient water use practices.

Rebates for 2025 totaled approximately $45,000. Rebates and cost-share from PV Water helps growers purchase and install more efficient irrigation equipment such as lower flow sprinklers, sprinkler check valves, pressure regulators, pressure compensating drip tape, pipe retrofits, soil moisture sensors, irrigation monitoring equipment like flow meters and data loggers, and repairs of leaky pipe joints.

Managed Aquifer Recharge/ Recharge Net Metering Program

Managed Aquifer Recharge (MAR) is a landscape management strategy that can help support groundwater supply by capturing stormwater in an infiltration system (typically a strategically designed basin) where it can then infiltrate into the aquifer. Since 2016, the RCD, the University of California, Santa Cruz (UCSC) and private landowners have collaborated to implement three active MAR projects in the Pajaro Valley with funding from DWR, USDA NRCS, California Coastal Conservancy and State Water Resources Control Board. Monitoring results were received for 2 of 3 systems for the 2024 water year. The two systems infiltrated 82 af/yr and 142 af/year in the 2024 rain year. The other 1 system has been moved to a pre-programmed annual infiltration rate of 5 af/yr. Additionally, water quality monitoring indicates that these projects likely help to improve groundwater quality. Data indicate that water infiltrated in the MAR basins had lower Nitrate [NO3-N] levels than ambient groundwater.

The RCD, UCSC, PV Water and private landowners continue to collaborate to implement the Recharge Net Metering (ReNeM) Program in the Pajaro Valley. Funding from the Department of Conservation and the Department of Water Resources has been utilized to plan, permit and design a new infiltration basin to be implemented in 2026. Six new sites were assessed using subsurface imaging and boring (tTEM, CPT) to determine their likelihood for achieveing aquifer recharge. Those results are currently being reviewed to determine what future assessment is needed and where opportunities exist. Finally, RCD and UCSC partnered with Nature 4 Water (N4W) to complete a business plan to guide the future of RENEM program development and initiated philanthropic fundraising efforts to further grow the RENEM program in 2026.

Creation of New Water Supplies – Municipal

San Lorenzo Valley Water District (SLVWD):

- Conjunctive Use Project: Expect to issue a notice of intent (NOI) for an Environmental Impact Report in early 2026 for expanding surface water use within SLVWD's jurisdiction
- Loch Lomond Feasibility Study under District review to determine the best utilization of the SLVWD's 313 acre-foot Loch Lomond allotment.

Scotts Valley Water District (SVWD):

- Continued working with regional partners (City of Scotts Valley, City of Santa Cruz Water and Public Works) developing a strategic direction for maximizing wastewater utilization in the region and for the benefit of Santa Margarita Groundwater Basin.
- Completed construction on the Scotts Valley Transit Center LID Retrofit, Phase 2 Project. This project is funded by $1.5 million in Urban and Multibenefit Drought Relief Grant to construct.
- Construction is in progress on the regional intertie 1 project.
- Completed equipping of the Sucinto Well and placed this source online in Fall 2025.
- Completed construction of Grace Way Well. Site improvements are in progress in 2025, including installation of a motor control building, site fencing, and connection to the raw water main located on Scotts Valley Dr.

City of Santa Cruz Water Department (SCWD):

- Construction of the Intertie 1 Project is underway. Intertie 1 includes a pipeline and pump station to connect the City of Santa Cruz Water Department system with the Scotts Valley Water District water system which will allow for water transfer and exchanges between two water systems. The project is funded through a DWR grant awarded to the City of Scotts Valley. Construction is expected to be completed in March 2026.
- The Water Supply Augmentation Implementation Plan (WSAIP) adaptive roadmap was finalized in 2025, outlining strategies to meet the City's water supply reliability goal of having adequate supply to meet all customer demand under plausible, worst-case conditions. The WSAIP technical report will be completed in early 2026 and will be incorporated into the City's 2025 Urban Water Management Plan.
- Development of the Santa Cruz Water System Model was completed in coordination with University of Massachusetts, Amherst research group. This tool is being used to support ongoing water supply planning work.
- The Santa Cruz Mid-County Groundwater Agency (MGA) and its member agencies are advancing the five Components of the SGMA Implementation Grant. City components include completion of two aquifer storage and recovery wells at existing well sites, and groundwater modeling to support the evaluation of additional projects and management actions. Both components support the goals of the MGA as well as contributing toward the City's water supply augmentation needs.

- Aquifer Storage and Recovery (ASR):
    - Completed design for conversion of Beltz 8 and Beltz 12 wells to permanent ASR facilities through the aforementioned SGMA grant. Construction of Beltz 12 well ASR facility is underway and is scheduled to be completed in 2026. Construction of Beltz 8 well ASR facility is anticipated to begin in early 2026 with construction to be completed in late 2027.
    - Completed pilot testing of ASR at Beltz 9 well and will begin design of the conversion of Beltz 9 well to an ASR facility in early 2026.
    - To further advance ASR, a Business Case Evaluation (BCE) was conducted to identify a fourth ASR well to pilot and convert to an ASR well in July 2025. The BCE criteria included maximizing water supply, groundwater sustainability, operational flexibility/ease of use, costs, and implementation schedule. The most feasible well to perform pilot testing at next was identified as the City's Beltz 10 Well due to its operational flexibility/ease of use, lower costs, and shorter implementation schedule.

The Soquel Creek Water District (SqCWD)

- The Pure Water Soquel (PWS) Advanced Purified Groundwater Replenishment Project is expected to be operational in 2025. This project will recycle wastewater from the City of Santa Cruz's Wastewater Treatment Facility (SCWWTF) through an advanced water purification process and use it to recharge the critically overdrafted Mid-County Groundwater Basin and protect against seawater intrusion. SqCWD made the following progress on the major components of PWS (conveyance, treatment, and groundwater replenishment) in 2025:
    - Conveyance: The conveyance pipeline conveys a portion of secondary treated effluent wastewater from the SCWWTF to the Advanced Water Purification Facility (AWPF) for treatment and purified water to three (3) Seawater Intrusion Prevention (SWIP) wells for aquifer recharge. Construction of the conveyance pipeline was completed in 2025.
    - Purification Facility: The AWPF is a multi-step advanced water purification process involving ultrafiltration, reverse osmosis, and ultraviolet light with advanced oxidation in addition to a pre-treatment ozonation step and a post-treatment stabilization step. Construction of the AWPF, as well as start-up and commissioning activities were completed in 2025. See Figure 8
    - SWIP Wells: These wells deliver purified recycled water to the Mid-County Groundwater Basin. Construction of the SWIP Wells, as well as start-up and commissioning activities, is expected to be completed in December 2025.
    - Recycled Water Facility: As part of the overall PWS Project, SqCWD is constructing a facility at the SCWWTF to produce tertiary treated recycled water for on- and off-

site non-potable water uses. Construction of the recycled water facility continued in 2025.

- Funding: In 2025, SqCWD continued to work with the funding agencies (State Water Resources Control Board (Prop 1 Groundwater Grant and Seawater Intrusion Control Loan), Bureau of Reclamation (Title XVI Grant Program), and the Environmental Protection Agency (WIFIA Loan Program)).

Figure 8: Pure Water Soquel Treatment Steps including ultra filtration membranes, reverse osmosis, and ultraviolet light.

Groundwater Management

The Sustainable Groundwater Management Act of 2014 (SGMA) went into effect on January 1, 2015 and is a key driver for developing and implementing long-range plans for groundwater sustainability. SGMA required the formation of local Groundwater Sustainability Agencies (GSAs) to prepare Groundwater Sustainability Plans (GSPs) in all of the state’s high and medium priority groundwater basins. Upon submittal of a GSP, GSAs have a 20-year implementation timeframe to demonstrate basin sustainability based on meeting locally defined sustainable management criteria. SGMA also requires annual reporting on GSP implementation progress to the Department of Water Resources (DWR) and a comprehensive periodic evaluation of the GSP every five years.

Santa Cruz County has three basins that are subject to compliance under SGMA. For each of these basins, the associated GSAs and their activities towards implementing their respective GSPs in Water Year 2025 are described below.

Santa Margarita Groundwater Agency

Management of the Santa Margarita Basin is overseen by a Joint Powers Authority (JPA) consisting of the County of Santa Cruz (County), the Scotts Valley Water District (SVWD), and the San Lorenzo Valley Water District (SLVWD). This JPA is referred to as the Santa Margarita Groundwater Agency (SMGWA), which is the GSA for the basin. The SMGWA governing board includes two private well representatives, two representatives from each partner agency, and one representative each from the City of Scotts Valley, the City of Santa Cruz, and the Mount Hermon Association. The Santa Margarita Groundwater Basin has experienced a significant historical decline in groundwater levels, particularly in the southern part of the Basin near Scotts Valley and has likely also seen reductions in streamflow. While groundwater levels stabilized and are no longer declining, they have seen only modest recovery. A groundwater model analysis indicated the need to implement at least modest projects in order to maintain sustainability under future climate conditions. The GSP for Santa Margarita was adopted by the SMGWA Board in November 2021 and approved by DWR in April 2023.

In Water Year 2025, the SMGWA continued monitoring of its network of seven monitoring wells in areas of previous data gaps in the basin. Many of these wells are located near active stream gauges in the basin that are also monitored by SMGWA, which will help improve the understanding of the surface water-groundwater relationship in the basin as required by SGMA. SMGWA submitted its latest annual report to DWR for Water Year 2024 by the April 1, 2025 deadline. In August 2025, SMGWA began the required periodic evaluation of its GSP. The periodic evaluation is due to DWR by January 3, 2027.

Also during Water Year 2025, SMGWA tracked progress by the basin's water supply agencies as they continue to develop their respective projects needed for basin sustainability. Of note, SVWD was awarded grant funding to construct an intertie with the City of Santa Cruz. While the primary purpose of the intertie is to address water shortages during drought or emergency conditions, it can create opportunities for expanded conjunctive use to benefit the basin. Construction of the intertie project is nearing completion. SLVWD continued to take steps to conduct a feasibility analysis of the use of Loch Lomond Reservoir to expand conjunctive use in the basin. Consulting services to support the analysis were procured in 2024.

Santa Cruz Mid-County Groundwater Agency

Management of the Santa Cruz Mid-County Basin is overseen by a JPA consisting of the County, City of Santa Cruz, Soquel Creek Water District and Central Water District. This JPA is referred to as the Santa Cruz Mid-County Groundwater Agency (MGA), which is the GSA for the basin. The MGA governing board includes three private well representatives and two representatives from each member agency. The Mid-County Basin is designated by the State as being in a condition of critical overdraft due primarily to the risk of seawater intrusion into the aquifers. Despite significant improvement of coastal groundwater levels due to water conservation and pumping redistribution, groundwater modeling analyses indicate that additional projects will be necessary to achieve sustainability. The GSP was adopted by the Board in November 2019 and approved by DWR in June 2021.

In Water Year 2025, MGA continued monitoring of seven groundwater wells constructed by MGA and seven stream gauges to improve its understanding of surface water-groundwater interaction. The MGA continued implementation of a non-de minimis well registration, metering, and reporting program. Five wells were initially identified as being required to comply with the program. All of the applicable wells have been registered and the first annual report of groundwater extraction is due at the end of Water Year 2026.Also during the water year, MGA completed the first required periodic evaluation of its GSP. The evaluation was submitted on the January 30, 2025, deadline. Also during the water year, MGA began a funding options assessment for long-term funding sources for SGMA regulatory compliance, with a report on options expected in late 2025. Finally, MGA member agencies, Soquel Creek Water District and the City of Santa Cruz, continued an optimization study to identify combinations of projects to achieve sustainability in the basin and improve water supply reliability for consumers. Completion of the study is expected in late 2025.

Work continued on a $7.6 million Sustainable Groundwater Management Act Implementation (SGMI) grant awarded to the MGA by the Department of Water Resources Sustainable Groundwater Management program. The grant supports the implementation of high priority projects identified in the GSP. The individual member agencies are leading the management and implementation of their respective projects. Additional description is available in this report under the lead implementing agencies:

Project Title	Lead Agency	Grant Award	Status
Cunnison Lane Groundwater Well	Soquel Creek Water District	$1,675,000	Well constructed; treatment plant design underway.
Aquifer Storage & Recovery, Beltz Wellfield	City of Santa Cruz, Water Department	$1,650,000	Design completed; construction underway
Park Avenue Transmission Main Improvements	Soquel Creek Water District	$800,000	Completed
Technical Development of GSP Group 1 & 2 Projects	Soquel Creek Water District and City of Santa Cruz	$1,900,000	Underway
Sustainable Groundwater Management Evaluation & Planning	MGA and County of Santa Cruz	$1,575,000	Underway

Pajaro Valley Water Management Agency (PV Water)

The Pajaro Valley Water Management Agency is a special district created in 1984 by the California legislature and is the GSA for the Pajaro Valley Subbasin (Basin). PV Water's 2014 Basin Management Plan Update, Basin Management Plan: Groundwater Sustainability Update 2022 (GSU22), and several other key documents, serve as a GSP Alternative which aims to achieve groundwater sustainability by 2040. PV Water's efforts to achieve sustainability directly support beneficial users and uses including drinking water, agricultural irrigation, and many more. Groundwater routinely provides more than 90% of the basin's water supply with supplemental water sources such as recycled water, managed aquifer recharge water, and beginning in 2025, treated surface water from College Lake serving as the other major sources. PV Water's existing facilities, current projects, and management actions are designed to achieve multiple objectives including providing drought resilience, preserving beneficial uses of groundwater, and enhancing natural conditions. The two biggest uses of extracted groundwater are for domestic consumption and agricultural irrigation. As part of PV Water's GSU22, a well depth analysis of more than 1,150 domestic and agricultural wells was conducted to inform the development of sustainable management criteria to protect beneficial users of groundwater from significant and unreasonable negative impacts, as well as enhance the resiliency of drinking water and irrigation water supplies. The GSU22 is the most current version of PV Water's GSP Alternative, which will be updated approximately every five years following a periodic evaluation. PV Water is preparing the periodic update of its GSP Alternative, the Basin Management Plan: Groundwater Sustainability Update 2027 (GSU27). It is due to DWR in December 2026.

PV Water operates several existing water supply facilities and administers a series of programs to reduce groundwater extractions and help stop seawater intrusion. Supplemental water supply facilities reduce groundwater extractions through the production, distribution, and use of supplemental water supplies in-lieu of groundwater pumping. PV Water also funds and manages a comprehensive water conservation program that aims to improve use efficiencies for both agricultural and domestic water users. It also partners with University of California at Santa Cruz (UCSC), and the Resource Conservation District of Santa Cruz (RCD) on a program called "Recharge Net Metering," in which private landowners develop infiltration basins to capture and infiltrate rainwater runoff into the groundwater basin. PV Water's existing supplemental water supply facilities, the Recharge Net Metering Program, and water conservation program are described in greater detail below.

- **Coastal Distribution System (CDS):** The CDS is a distribution system composed of nearly 22 miles of pipeline used to deliver supplemental water supplies to farms in coastal areas of the Pajaro Valley. The area currently served by the CDS incorporates approximately 6,100 irrigated acres and is referred to as the Delivered Water Zone or the Delivered Water Service Area. Water delivered through the CDS replaces groundwater that would otherwise be pumped from coastal wells. Delivered water provides "in-lieu recharge" to the Pajaro Valley Basin; helping to eliminate the problems of groundwater overdraft and seawater intrusion, while helping to keep agriculture viable in the Pajaro Valley.

- **Harkins Slough Managed Aquifer Recharge and Recovery Facility (Harkins Slough Facility):** The Harkins Slough Facility diverts surface water from Harkins Slough and conveys it to a recharge basin where it percolates into a surficial aquifer of the San Andreas Terrace located near the coast. PV Water utilizes a series of wells to recover recharged water and deliver it to coastal farms through the CDS. The Harkins Slough

Facility commenced operations in 2002 and has recharged approximately 12,500 acre-feet through October 2025.

- **Watsonville Area Recycled Water Treatment Facility (RWF):** PV Water constructed the RWF and operates it in partnership with the City of Watsonville. Located adjacent to the Watsonville Wastewater Treatment Plant at the Water Resources Center, the RWF has the capacity to produce 4,000 acre-feet per year of tertiary treated disinfected recycled water. Recycled water is augmented with water from the Harkins Slough Facility, Supplemental Wells, the City of Watsonville's potable water system, and treated College Lake water to increase supply and improve the quality for agricultural irrigation needs. The RWF commenced operations in 2009 and has produced more than 43,350 acre-feet through September 2025.

- **Supplemental Wells:** In 2025 PV Water operated three production wells to augment the delivered water supply and improve water quality. The newest well was constructed as part of the College Lake Integrated Resources Management Project and put into service in August 2025. PV Water intends to bring a fourth supplemental well into service by summer 2027.

- **Recharge Net Metering (ReNeM):** PV Water, along with program partners from UCSC and the RCD, and participating private landowners, are implementing ReNeM to enhance recharge in the Pajaro Valley. The program incentivizes small scale recharge projects by providing rebates to landowners based on the volume of water infiltrated through infiltration systems. The rebates are intended to help offset maintenance and operation costs incurred by landowners. Currently, the program includes three infiltration basins, with additional sites under evaluation and design. The ReNeM program team is currently evaluating the water year 2025 performance of the three systems. During water year 2024, in which the annual total precipitation was equivalent to the long-term historical average, a combined total of approximately 229 acre-feet was infiltrated.

- **Water Conservation:** PV Water set a goal to achieve 5,000 acre-feet per year of water conservation when compared to the baseline period of 2006-2010. The program focuses on agricultural water conservation but also provides conservation services for domestic users. The agricultural conservation program leverages numerous technical partners including the RCD, the Natural Resources Conservation Service, the UC Cooperative Extension, the Resource Conservation District of Monterey County, and private consultants. The main components of the program are conservation outreach; partner collaboration, program coordination, demonstrations, rebates for efficient devices/materials; workshops and trainings; an irrigation efficiency program; and irrigation efficiency program evaluation. In March 2024, the PV Water Board of Directors approved a $1.37 million agreement to fund the agricultural conservation program support services through June 2027. Over the most recent evaluated rolling 5-year period (2020-2024), total agricultural water use was approximately 5,875 acre-feet less than the baseline period.

While the result of operating the existing facilities and administering these programs has been effective in helping to reduce overdraft and slow seawater intrusion, PV Water is working to construct and implement additional projects and management actions to achieve sustainable groundwater resources and provide resiliency. These additional efforts are described below.

- **College Lake Integrated Resources Management Project (College Lake Project):** The College Lake Project includes components required to store, treat, and deliver water from College Lake, for use as an irrigation supply in-lieu of pumped groundwater to reduce the rate of seawater intrusion while helping to preserve agriculture. The components include an adjustable weir structure designed to accommodate safe fish passage, intake pump-station, water treatment plant, a 6-mile conveyance pipeline, and two groundwater wells to support project operations. The weir is capable of raising the lake water level by 2.4 feet and increasing the total storage to approximately 1,800 acre-feet. An anticipated annual average of 1,800-2,300 acre-feet will be collected through a screened intake compliant with screening criteria for anadromous salmonids. Water is conveyed to the water treatment plant and then to the CDS where it is utilized in place of groundwater production. PV Water began construction of the College Lake Project in spring 2023 and began commissioning the facility and delivering water to customers in June 2025. Substantial completion of the project is anticipated to occur in fall 2025.

Figure 9: College Lake Weir System installed by the Pajaro Valley Water Management Agency.

- **Watsonville Slough System Managed Aquifer Recharge and Recovery Project (WSS-MARR):** WSS-MARR includes upgrades of the existing Harkins Slough Managed Aquifer

Recharge Facility (Harkins Slough Facility) and construction of the Struve Slough Project, a new managed aquifer recharge and recovery project. WSS-MARR includes project components to divert, convey, store, and recover surface water for use as an irrigation supply in-lieu of pumping groundwater. The components include upgrading the existing Harkins Slough Facility to install fisheries-compliant intake screens, upgrading the pump-station, development of a new recharge basin, and constructing series of recovery and monitoring wells. The Struve Slough Project includes a new screened intake on Struve Slough, a pumping-station to be located adjacent to the slough, as well as an approximate 7,150-foot conveyance pipeline. Collectively, WSS-MARR is designed to yield an estimated annual average of approximately 2,250 acre-feet for recharge and subsequent recovery. In March 2025, PV Water's Board of Directors certified an addendum to the certified environmental impact report for the project and adopted modifications to the project. PV Water has completed the 100% designs and is advancing efforts to obtain all necessary permits including a 4,000 AFY water right on Struve Slough and a Coastal Development Permit. The Agency is also in the process of procuring all necessary property rights for the project Construction of the Project is anticipated to begin in 2027.

- **Increased Recycled Water Deliveries:** PV Water continues efforts to increase recycled water deliveries to customers. PV Water is working to achieve this by increasing demand for recycled water and increasing storage to supply more water during periods of high demand. The goal is to increase demand by approximately 1,000 acre-feet per year and shoulder season demand by approximately 250 acre-feet per year over 2011 levels. Completed infrastructure improvements developed to increase recycled water deliveries included the construction of a 1.5-million-gallon storage tank, approximately 3.2 miles of additional CDS pipeline; an expanded RWF filter train; and improvements to the distribution pump station. PV Water continues to work closely with customers to maximize deliveries and increase recycled water use. In addition, condition and operational assessments of the RWF and the City of Watsonville Wastewater Treatment Plant have been conducted and will guide improvements in reliability and process performance in the future. These improvements began in 2025 with replacement of variable frequency drives and modifications to electrical equipment to enable connection to standby power during emergencies and planned shutdowns.

Guided by the GSU22 and future updates of the GSP Alternative, PV Water will continue efforts to achieve sustainable groundwater resources. Annual and periodic assessments every five years will evaluate basin conditions against sustainable management criteria established to provide a resilient and sustainable groundwater basin. The next major update and periodic evaluation is planned to begin in late 2025 and culminate in an updated GSP Alternative submitted by December 24, 2026.

County's Roles in Groundwater Management

Despite not managing significant water supply, or large augmentation projects, the County has served an important role in supporting groundwater management. This role includes oversight of wells, and groundwater data collection and synthesis, in addition to being a JPA signatory to both the MGA and SMGWA. The County has served as the lead in procuring and managing contracted services that leverage opportunities to strategically pool resources to benefit both basins. The County led a process to develop a regional data management system (DMS) to help the GSAs meet the requirements of SGMA, and additionally to collect and organize data collected by all of the water agencies in the County. The system can be viewed online at sccwaterdata.us/#/html/home. There are a few advantages to the regional system: it provides a robust storage system for critical historical data; it makes it easier to compare data across agencies; and the web portal makes it easy for interested parties to view results.

Wells GIS Layer Update (Nearing Completion 2025)

County staff developed an AI-powered optical character recognition workflow to automate data extraction from well completion reports, supplemented by custom Python scripts for post-processing, data cleaning, classification, and matching. Integrated with quality assurance checks and GIS workflows, the project is modernizing the County's wells layer by improving data accuracy, completeness, and spatial reliability. Key fields such as well type, depths, screen intervals, and location have been standardized, geocoded, and supplemented where previously incomplete or missing. The resulting well layer will provide a much-improved foundation for the County and partner agencies to conduct analysis, modeling, and decision-making. Because the AI model and scripts were trained on all California DWR well completion report templates, they are adaptable for statewide use, enabling rapid updates and modernization of other counties' well databases while providing a pathway for future AI-assisted data integration.

County of Santa Cruz Well Ordinance Update

Santa Cruz County Code (SCCC) Chapter 7.70 specifies measures for the siting, construction, and destruction of wells to protect groundwater resources and provide suitable water supply for the intended use. SCCC Chapter 7.73 specifies yield and water quality requirements for individual water systems that predominantly utilize wells. The last significant revisions of Chapter 7.70 and Chapter 7.73 were completed in 2009 and 1993, respectively. After a teo-years long process led by County Water Resources staff, the Board of Supervisors finalized the adoption of updated regulations to their well ordinance in early 2025. These changes went into effect in July 2025 after they were approved by the Coastal Commission.

The following are the significant changes to Chapter 7.70:

1. Additional measures are added to reduce impact of wells on groundwater resources, streams and associated public trust resources, karst areas, nearby wells, and designated groundwater extraction concern areas;
2. Different levels of review and protective measures for different types of wells are provided for, including discretionary review and potential for denial of Tier 4 wells;

3. Explicit provisions are added for review and comment on well applications by affected water agencies and groundwater sustainability agencies;
4. Provisions are added for regulation of soil borings and stormwater infiltration devices;
5. Metering of all newly installed non domestic wells will be required;
6. Penalties for code violations are added; and
7. Provisions are added for promulgation of specific policies for implementation of code requirements to allow more flexibility for implementation and adjustment of specific elements of effective policy.

The following are the significant changes proposed to Chapter 7.73:

1. More extensive water quality testing for individual water systems: Title 22 constituents, plus other constituents in water quality concern areas;
2. More stringent yield testing in known limited yield areas;
3. Recordation of a notice on the deed for new wells with limited yield or quality;
4. Individual Water System requirements also apply to non-domestic uses and additional testing is required for change or expansion of use;
5. Water quality testing and yield testing at the time of property transfer to inform the buyer.

Updates are provided through the website:
scceh.com/NewHome/Programs/WaterResources/WellOrdinanceUpdate.aspx

Small Water Systems and Domestic Wells

The Santa Cruz County Drinking Water Program oversees 106 active small water systems (SWSs), including water systems with 5-199 residential connections and systems serving at least 25 people per day for 60 or more days per year. These systems include housing developments and mutual water companies, in addition to facilities such as schools, office buildings, outdoor camps, and stores. SWSs can have greater water supply vulnerabilities than larger systems because they tend to have few sources, often just one well or spring, and a small population to bear the cost of repairs for their aging water sources and distribution systems.

The water quality and reliability of these systems is of critical importance to the County residents and visitors that depend on them. Recent extreme weather events since 2020 (e.g., the CZU Lightning complex Fire and winter storms of 2023) have exposed and heightened some of the vulnerabilities of these systems, such as lack of redundancy and aging infrastructure. Recent legislation such as SB 552 (drought planning for small water suppliers and rural communities) require SWSs, subject to funding availability, to implement specific resiliency measures such as joining a mutual aid network, obtaining a backup source of electricity, and securing additional water sources, if feasible.

While drinking water quality for SWS in Santa Cruz County is generally very good, water quality challengesexist for some systems. Nitrate contamination is a concern, and primarily affects areas in South County near agricultural land uses. Drinking Water Program staff work closely with a number of SWSs in this area which provide nitrate removal treatment to ensure water quality standards are met.

Another contaminant of local concern is Hexavalent Chromium, also known as Chromium-6. Chromium-6 occurs naturally in the Aromas Red Sands aquifer that is found in parts of Aptos and Watsonville. Drinking Water Program staff are working with 7 small water systems in this area with elevated levels of Chromium-6 to respond to the newly created Maximum Contaminant Level (MCL) of 10 micrograms per liter. Staff will be working with the affected systems to review proposed treatment solutions that will provide the affected residents with a source of water that meets the new requirements by October 1, 2028.

Per- and polyfluoroalkyl substances (PFAS), also known as "forever chemicals", are another emerging group of contaminants. These substances are found in many consumer products, including nonstick cookware and waterproof coatings, and end up concentrating in landfills. The EPA implemented a rule in 2024 requiring sampling for PFAS by 2027 and treatment by PFAS have been found in elevated levels in wells serving some SWSs adjacent to the Buena Vista Landfill. One of these facilities has an operational PFAS treatment system that has successfully removed the PFAS contaminants, and another has received a grant to address this issue as described below.

The County General Services Department has secured an $800,000 grant from the Department of Water Resources'(DWR) Proposition 1 Implementation Grant Program via the Integrated Regional Water Management Program to install a treatment system for Chromium-6 and PFAS, and to study options to improve source quality and reliability at the Rountree Facility in Watsonville. The system is served by a single well and is proactively working to address these emerging contaminants and improve the resiliency of the system.

County staff are also involved with coordinating several long-term projects to improve water supply reliability for SWSs. Renaissance High School is currently working to consolidate with the Soquel Creek Water District due to a lack of backup sources for its single supply well, and water quality concerns, including hexavalent chromium. The Crestwood Heights Water Association is working to consolidate with the City of Watsonville due to diminishing water supply from their source wells and a lack of funds to upgrade their system.

The County has also been awarded $97,800 in grant funding from DWR's Small Community Drought Relief Program for improvements to the Waterman Gap water system, a small water system at the northern edge of the County in Boulder Creek. The current stream source has declined in flow, and the system's backup wells have limited capacity. These funds will cover the cost of reconstructing a water line to an existing stream intake on Little Boulder Creek and installing four storage tanks to improve the system's ability to provide a reliable supply to residents.

Drinking Water Program staff continue to host Small Water Systems Forum meetings to provide regulatory updates to SWSs and encourage discussion and collaboration between these systems. One forum meeting was held in 2025, which was a joint meeting with the

Water Advisory Commission. This meeting sought feedback from the small water systems on the challenges with consolidation, while also providing results from a consolidation study completed by County staff.

Services Offered by the County

On December 1, 2021, the Water Advisory Commission (WAC) voted to take responsibility for implementing Senate Bill (SB) 552. SB 552 required the County to write a plan that includes potential water shortage risk analysis and proposed interim and long-term solutions for State Small Water Systems and domestic wells. This plan is now referred to as the Santa Cruz County Drought Response and Outreach Plan (DROP). The Water Quality Specialist and Water Resource Planner in the Water Resources Division have taken the lead on the implementation of the DROP.

Since the Board of Supervisors approved the DROP in December of 2022, staff created web portals for both Household Well Assistance and Small Water Systems that utilize the information gathered in creating the DROP. These pages are intended to act as a single repository for both informational and direct support resources. These webpages, incorporated feedback from private well owners and Small Water Systems regulated by the county.

The County secured a $600,000 grant from the California State Water Resources Control Board to offer water quality services to financially qualified households that are served by a household well. These services include testing well water and supplying alternative water sources when necessary, such as bottled water and in-home treatment systems. To identify households that could utilize these free services, Santa Cruz County began a partnership with the Central Coast Drinking Water Well Testing Program, a regional program from the Bay Foundation and Regional Water Quality Control Board that provides free well testing to all county residents. The goal of the Central Coast Program is to ensure that households are aware of their drinking water quality and improve understanding of groundwater quality on the Central Coast. By coordinating these two programs, Santa Cruz County was able to reach more residents and provide faster support to financially qualified residents. Contracts are in place to provide residents meeting income thresholds with the following services:

1. Emergency water hauling for wells that go dry due to drought conditions
2. Bottled water deliveries
3. Point of Use Treatment system installations
4. Additional water quality testing, including PFAS

Residents wishing to apply for these services can apply here: Water Quality Assistance Application (Solicitud de Asistencia para la Calidad del Agua).

To raise awareness of the resources above, and services already offered by the County such as well soundings, staff created and distributing mailers to parcels served by a domestic wells and completed in-person outreach in areas with water quality concerns that might serve residents meeting the income threshold.

In 2023, the County also received an additional $125,000 grant from DWR to focus on gaps in the DROP- specifically, a comprehensive wells GIS layer, a connection feasibility analysis for small water systems, and enhanced guidance on the consolidation process. As of 2025, both major deliverables are nearing completion.

Santa Cruz County Environmental Health was awarded funding through the SWRCB Safe and Affordable Funding for Equity and Resilience (SAFER) grant program and the DWR Urgent Drinking Water grant in 2021 and the contract was executed in 2023. This funding will be used to make progress to implement this plan, beginning with the tasks outlined in **Error! Reference source not found.** below (subject to change as the work progresses).

2025 DROP Implementaion Highlights
Connection Feasibility Analysis for Small Water Systems (Completed 2025): County staff completed a GIS-based analysis evaluating physical and managerial consolidation opportunities for State Small and Public Water Systems across the County. The physical consolidation assessment modeled pipeline infrastructure costs using key factors such as distance to large water mains, terrain slope, elevation differences affecting pressure, and widespread geotechnical constraints including landslides, liquefaction, expansive soils, high groundwater, and fault crossings. Distance emerged as the primary cost driver, with estimated pipeline costs averaging $3.4 million per system and ranging widely depending on route complexity. Notably, over half of the systems share potential pipeline paths with others, creating significant opportunities for cost-sharing through coordinated projects. The analysis also identified proximity-based connection opportunities for over 600 parcels served by individual domestic wells located within a feasible distance of large water system infrastructure. The managerial consolidation analysis, based on driving times between systems, revealed that most public water systems are within close proximity of one or more partners, supporting feasible administrative and operational integration as a lower-cost path to resilience.

Consolidation Process Brochure (In Progress): County staff are developing a consolidation guidance document for small water systems that may be considering consolidation. This guidance document will incorporate the connection feasibility analysis above, and will be structured around the feedback provided by the small water systems at the April meeting of the Water Advisory Commission. The intent of the guidance document will be to help struggling small water systems understand the steps needed to prepare for a consolidation with a large water system. County Staff intends to have the guidance document completed by the end of 2025.

Specific Water System Assistance

Big Basin Water Company (BBWC) is a privately-owned utility serving 540 households with drinking water and 30 parcels for wastewater management in the San Lorenzo Valley. BBWC was beset by years of financial mismanagement and lack of investment in critical infrastructure, leading to repeated service interruptions for customers, difficulty rebuilding for

CZU survivors, and litigation by the State Water Resources Control Board on the drinking water side and Regional Water Quality Control Board on the wastewater side. Since taking over in 2023, and with financial assistance from the State through a grant to the County of Santa Cruz, the court appointed Receiver has made significant strides in improving the system operations:

- Through a contract with Cypress Water Services, the drinking water system is functioning now with few water outages or boil water notices. The well, which previously ran 24/7 now only needs to run 8-10 hours per day to meet demand, due largely to improvements in leak repair.
- Rate increases were approved by the California Public Utilities Commission, and the billing system has been updated.
- A comprehensive needs assessment[5] was developed by Moonshot Missions that evaluates the further upgrades necessary. GEI Consulting is developing a water source feasibility study to determine what options there are for augmenting supply. They are also providing engineering designs to increase the capacity of the intertie with SLVWD.
- The Receiver has talked with several agencies and organizations about prospective long-term purchase/merging of the system.
- The wastewater system is now operated by the County of Santa Cruz CSA 7.
- County staff continue to regularly with the regulatory agencies as well as elected officials to work towards a sustainable resolution to the challenges of the BBWC.
- The Reciever and District 5 Supervisor held a Town Hall meeting on November 17, 2025 to provide updates to the community (Figure 10).

Figure 10: Big Basin Water Town Hall panel held in November 2025.

Waterman Gap is a State Small Water System at the Northwest corner of the County. Using a grant from the California Department of Water Resources (DWR), the County was able to support the water system in some critical upgrades. The project included reconstruction of a 9,500-foot, 1.5-inch diameter PVC supply line and diversion point in Little Boulder Creek and the installation of four (4) 4,995-gallon polyethylene tanks to supply the existing Waterman Gap water system (see Figure 11).

5 www.bigbasinwater.com/announcements/c64s18khew63q9snudmdzmq1cwft2y

Figure 11: New intake and tanks for Waterman Gap water system

Section 2: Water Quality of Santa Cruz County

As shown in Figure 12, several watersheds within Santa Cruz County have been identified by the State of California as having impaired waterbodies pursuant to Section 303(d) of the Federal Clean Water Act (CWA)[6]. By definition, 303(d) listings and adopted TMDLs are related to impacts on one or more beneficial uses and the need to control the source(s) of these impairments. The Regional Water Board has oversight over these waterbodies and manages water quality through implementing Total Maximum Daily Loads (TMDLs) that are incorporated into Basin[7] Plans, and the National Pollutant Discharge Elimination System (NPDES)[8] permit program, including the Storm water (MS4)[9] program. The County of Santa Cruz and the Cities of Santa Cruz, Capitola, Scotts Valley, and Watsonville conduct extensive water quality monitoring and there is ongoing collaboration to exchange data among the individual stakeholders.

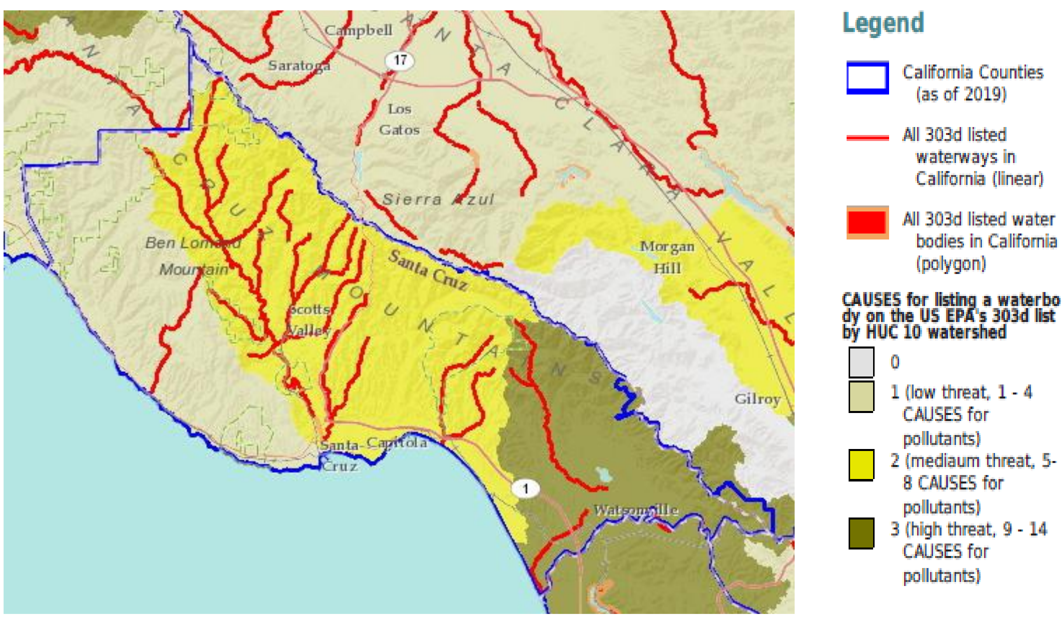


Figure 12: Map of watersheds with impaired water bodies in the County as identified by the Central Coast Regional Water Quality Control Board

Santa Cruz County Water Quality Program and Laboratory

---

[6] https://www.waterboards.ca.gov/water_issues/programs/tmdl/background.html
[7] https://www.waterboards.ca.gov/centralcoast/publications_forms/publications/basin_plan/
[8] https://www.waterboards.ca.gov/water_issues/programs/npdes/
[9] https://www.waterboards.ca.gov/water_issues/programs/stormwater/municipal.html

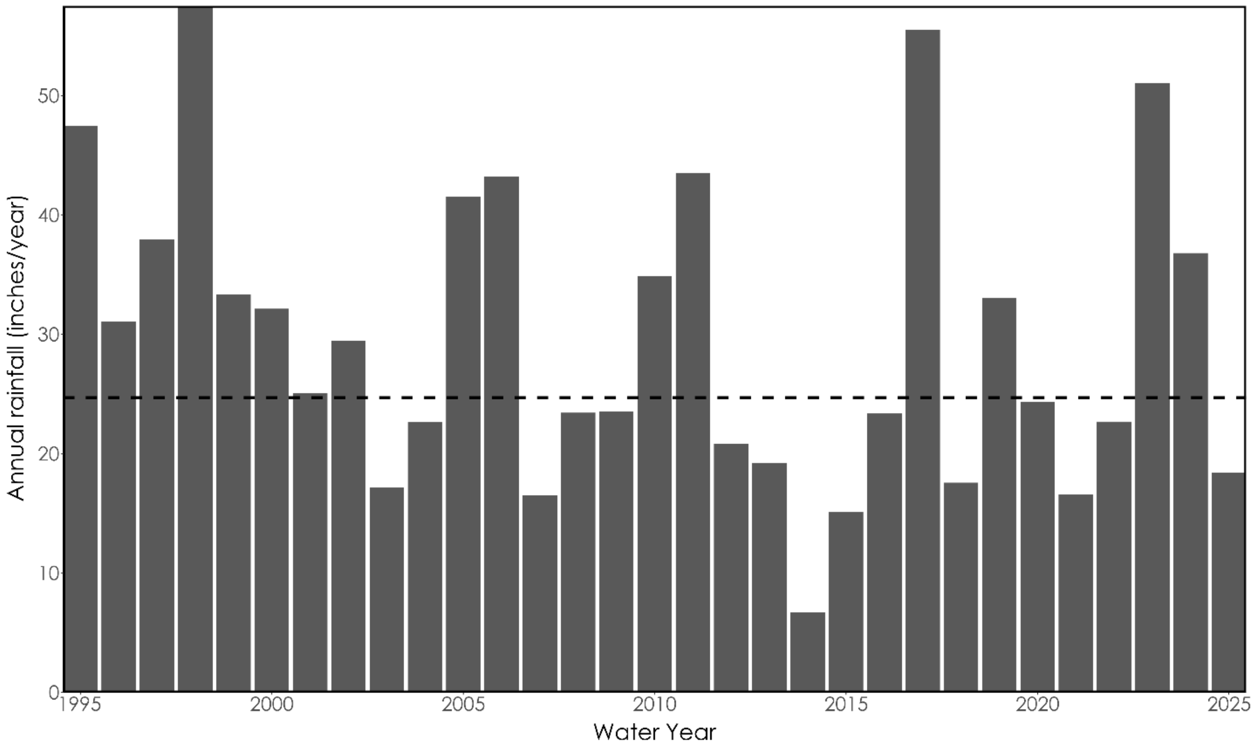


Figure 13: Cumulative rainfall for water years (starting October 1 and ending September 30) from CIMIS station 104 at De Laveaga. Dotted line indicates median over the thirty year period.

This year was a mild water year following two years of relatively high rainfall (Figure 13). The program routinely visits over 100 sites located in streams, rivers, lakes, and beaches throughout the county at weekly, monthly, and quarterly intervals depending on mandates, public health advisories, and logistical constraints (Figure 14). Over 1000 grab samples were collected this year as a part of routine monitoring efforts, with additional samples analyzed for special studies and fees for service. Coastal sites are monitored for fecal indicator bacteria in accordance with the California Beach Water Quality Program[10]. Freshwater sites are also monitored for fecal indicator bacteria along with other geochemical and physical parameters (e.g., Nitrate, ortho-phosphate, sediment, etc.).

While maintaining a standard suite of analytes, the lab is always improving and increasing capacity to run new constituents that improve understanding of water quality issues. The lab is currently adapting a protocol to measure chlorophyll a, an important metric correlated to the abundance of phytoplankton. Measurements of chlorophyll a will provide much needed insights into harmful algal bloom dynamics. Additionally, the lab has begun a pilot study to determine the sources of fecal indicator bacteria in the Watsonville Slough System using digital PCR. While the lab has used dPCR to assess the relative abundance of human DNA in water samples in the past, this is the first time that the program will use dPCR for TMDL management and assessment of best management practices. Furthermore, this ongoing

[10] https://mywaterquality.ca.gov/safe_to_swim/

microbial source tracking study will also determine the relative contributions of fecal indicator bacteria from cows, seagulls, and dogs. These data will help to differentiate between controllable and non-controllable sources of bacteria in Watsonville Slough and more accurately assess the status of TMDL attainment.

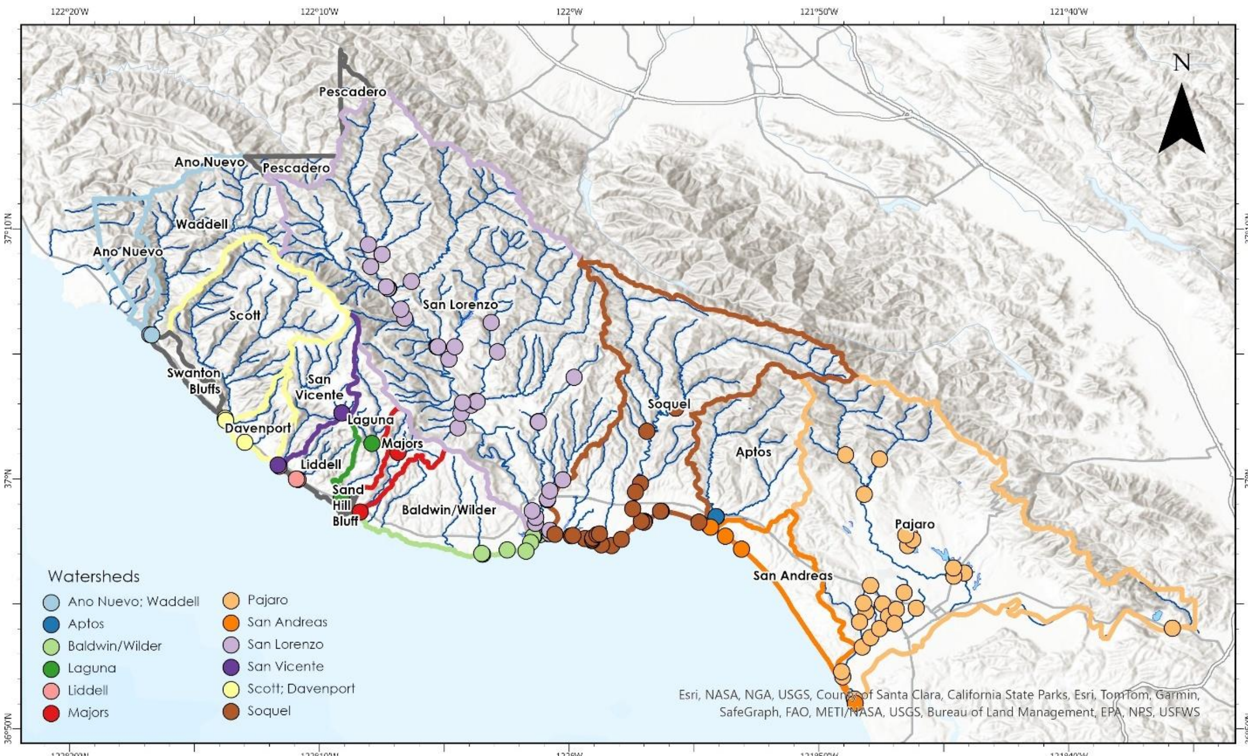


Figure 14: Santa Cruz Water Quality Monitoring Program routine monitoring site locations within respective Santa Cruz County watersheds. Each point represents the location where grab samples are collected, while lines represent watershed boundaries.

Harmful algal blooms

Harmful algal blooms (HABs) are caused by marine or freshwater microalgae. HABs are an ongoing issue worldwide and are projected to become more frequent and persistent due to anthropogenic impacts (i.e., climate change, nutrient enrichment). During a harmful algal bloom, microalgae produce toxins that can kill aquatic wildlife and are harmful to human health. In Santa Cruz County, freshwater HABs regularly occur during summer months in inland lakes and coastal lagoons. This year, blooms of *Microcystis* sp. producing the toxin microcystin were concentrated at Pinto and Kelly Lakes, both of which reached hazardous levels in August based on thresholds set by the EPA and guidance of the California Cyanobacterial and Harmful Algal Bloom Network[11] (Figure 15). Pinto Lake was closed on August 13, 2025 and remained closed through the end of this water year (Figure 16). While monitoring also occurs at Kelly Lake, this lake is private and the county has limited ability to

[11] https://mywaterquality.ca.gov/cyanohab/

close access, but provides results on microcystin toxin concentrations and public health guidelines to the community living on the lake on a weekly basis. Throughout the summer Corcoran, Moran, and Schwan Lagoons were also periodically sampled for the presence of nuisance species known to produce toxins. Microcystin was detected at all three lagoons during the summer and anatoxin was detected once at Schwan lagoon. However, levels never reached health advisory thresholds.

Figure 15: Cyanobacterial blooms at Pinto Lake (left) and Kelly Lake (right), summer 2025

The water quality laboratory monitors the presence and tracks the progress of blooms using both qualitative microscopy (Figure 17) and biochemical analyses that quantify toxin concentrations. Briefly, from early spring through fall, samples are collected regularly and checked for the presence of toxin producing genera. If visualized, then rapid plate kits are used to quantify toxin concentrations. The lab has capacity to measure microcystins, anatoxins, cylindrospermospin, saxitoxin, and domoic acid. Importantly, this approach allows for early detection and rapid response to the presence of toxin producing species. There is always site-to-site variability in the overall density of microorganisms/toxins and differences in the duration of the cyanobacterial bloom even within the same body of water. For instance, at Pinto Lake, the County Dock (PL16) generally has lower microcystin concentrations than other areas of the lake, likely due to prevailing winds and algal accumulation (Figure 16).

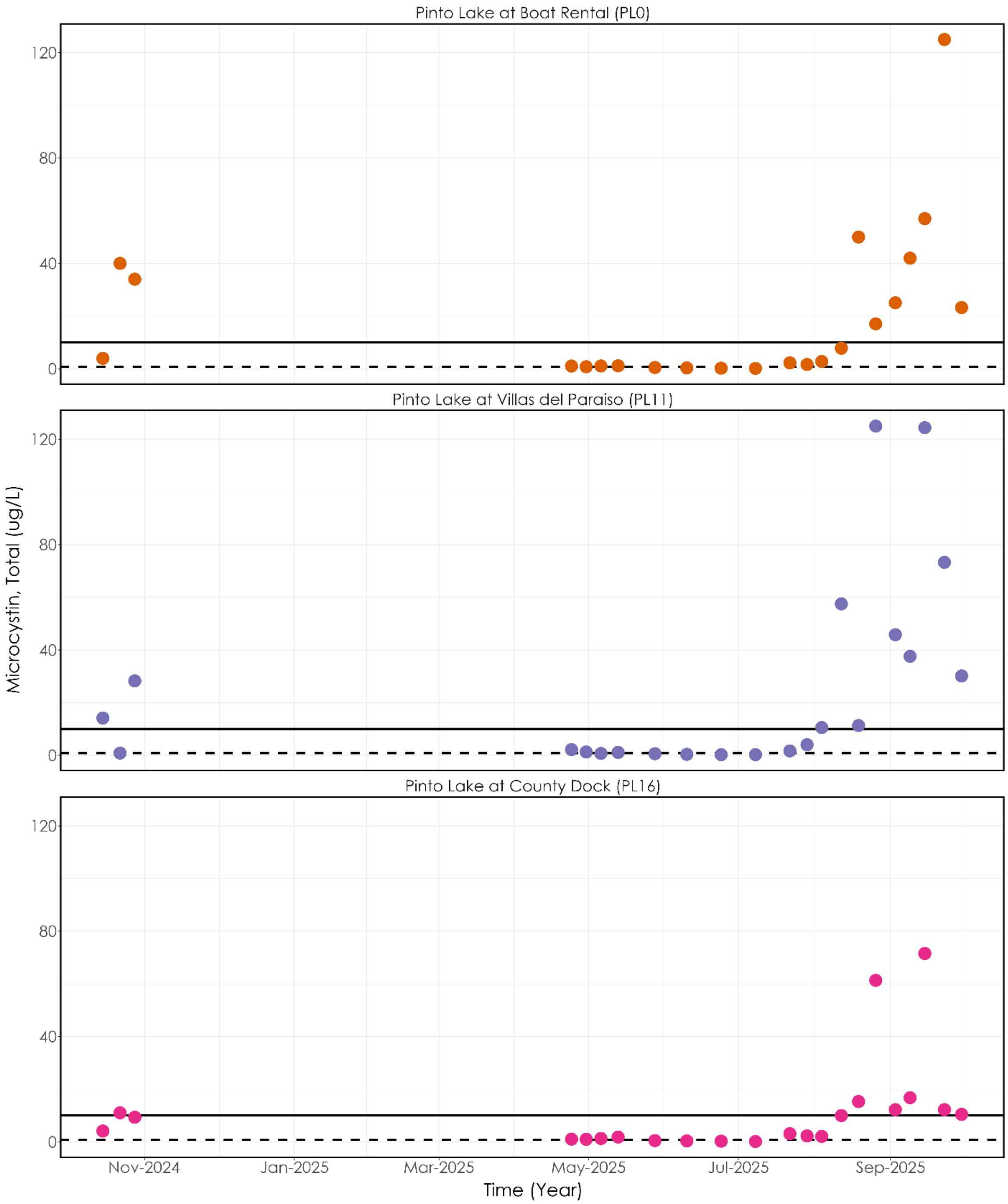


Figure 16: Total Microcystin concentrations (µg/L) at County Sites: PL0 (Pinto Lake at Boat Rental), PL11 (Pinto Lake at Villa del Paraiso), and PL16 (Pinto Lake at County Dock). Black dashed lines indicate 0.8 µg/L TMDL threshold. Solid black line indicates 10 µg/L, the threshold at which the County of Santa Cruz and City of Watsonville enact management actions for lake closure.

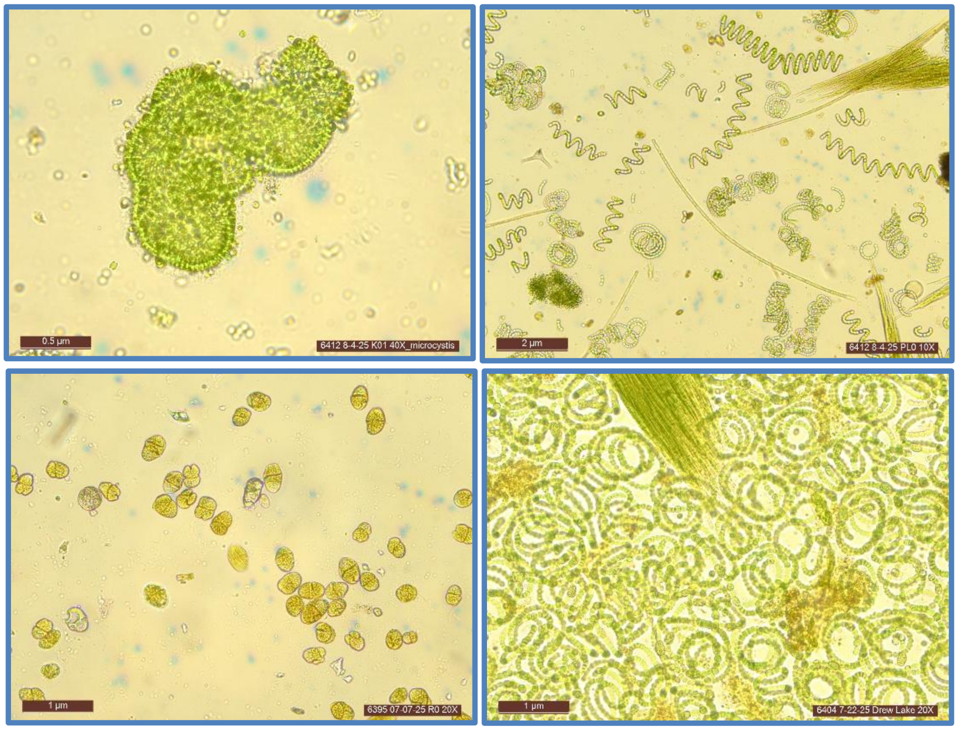


Figure 17: Photomicrographs close up of microcystin at Kelly Lake Dock (K01) upper left; phytoplankton community at Pinto Lake City Dock (PL0) upper right; Akashiwo sanguinea from Corcoran Lagoon (R0) bottom left; and Aphanizomenon (rod shaped alga) and Dolichospermum (circular alga) from Drew Lake (Drew-2) bottom right. Courtesy of Eric Baugher, County of Santa Cruz.

Appearance of a red tide (marine HAB) started on July 15^{th}, 2025. This event was short lived (~1 week). The most impacted beaches were New Brighton, Seacliff, Rio del Mar, and Platform. The dominant species of red tide phytoplankton found were Akashiwo sanguinea, Alexandrium, and Psuedo nitzschia (Figure 18). Elevated levels of domoic acid were measured in ocean water samples taken during the bloom. Domoic acid is responsible for marine mammal distress as the neurotoxin accumulates within the fish that they feed on and biomagnifies within their nervous system, affecting their brain and heart. Saxitoxin is another neurotoxin that is highly potent and causes paralytic shellfish poisoning in humans. Therefore, the ban on shellfish harvesting from May 1 to October 31 is always in place. It was more important to note the ban this year with the active red tide and phytoplankton responsible for paralytic shellfish poisoning.

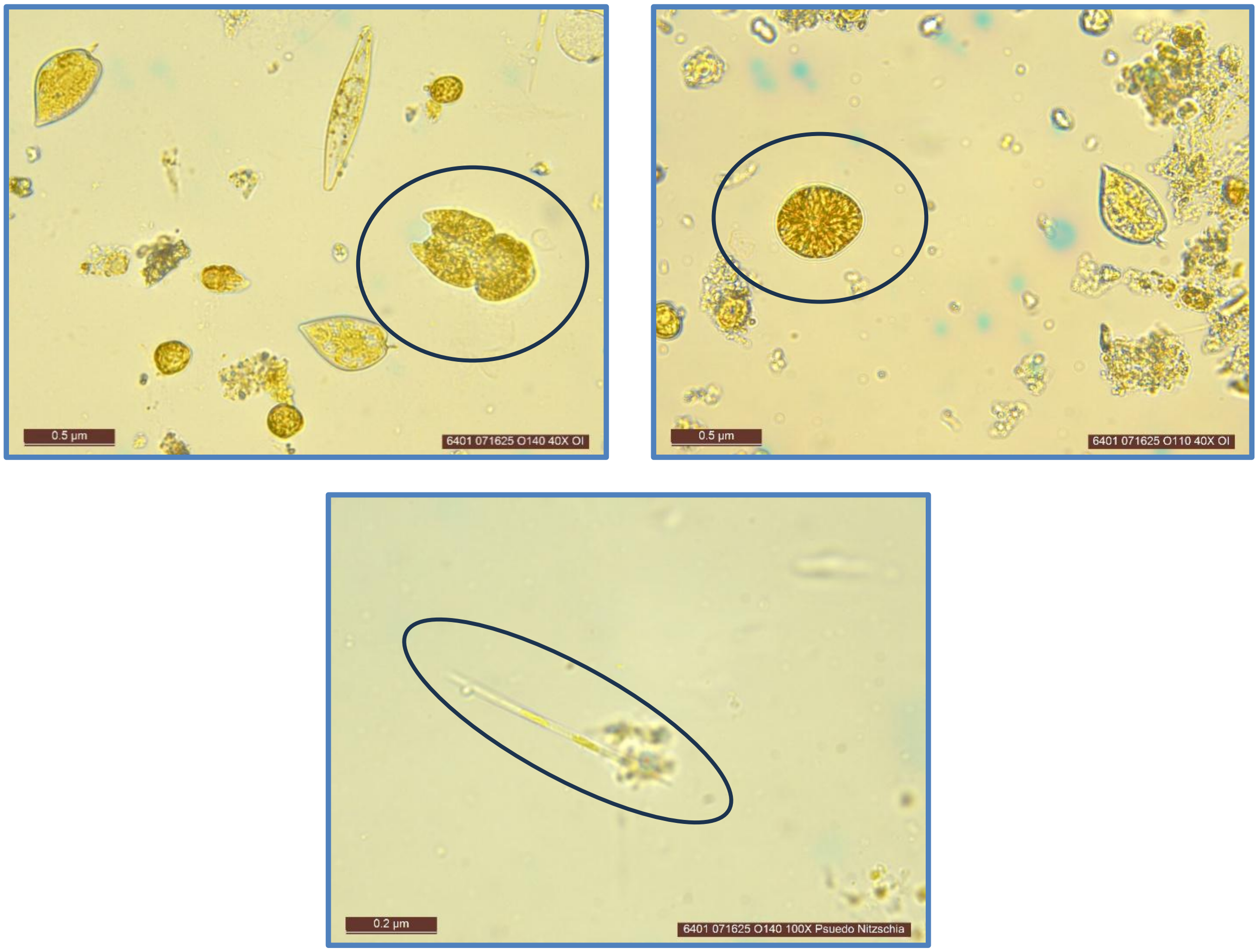


Figure 18: Photomicrographs close up of Akashiwo sanguinea at Seacliff Beach (O140) upper left; Alexandrium at Rio Del Mar Beach (O110) upper right; and Pseudo nitschia at Seacliff Beach (O140) bottom. Ellipses highlight alga specified. Courtesy of Eric Baugher, County of Santa Cruz

Beach Water Quality

Water quality in 2024-2025 at the County's beaches showed seasonal and episodic variability due to changes in rainfall and ocean conditions. 624 samples were collected across 47 sites spanning south county at the mouth of the Pajaro River to as far north as Waddell Creek Beach. Samples were generally collected before or 72 hours after measurable rainfall to avoid sampling when water quality is known to be degraded (Figure 19). Health advisories were posted to the Santa Cruz County Water Quality website when E. coli values exceeded 400 MPN/100mL, total coliforms exceeded 10,000 MPN/100 mL, and/or Enterococcus exceeded 104 MPN/100mL. At select high visitation beaches in the City of Santa Cruz, permanent water quality signs are also posted to alert the public to a health advisory on site. Public health messaging about rain events and water quality are permanently posted to the county's beach advisory website. Any elevated readings were re-tested to determine persistence of elevated fecal indicator bacteria and/or to remove the health advisory when levels fell below health advisory thresholds. In general, while fecal indicator bacteria are generally correlated, water quality thresholds are not usually exceeded for multiple parameters. For instance, at Cowell's beach (Figure 20a) health advisory thresholds for total coliforms and E. Coli were not exceeded this water year, however, the health advisory threshold for Enterococcus was exceeded twice (once in February and once in July). However, at Capitola beach at Soquel Creek, health advisories for all three fecal indicator bacteria were issued at the same time twice this water year (Figure 20b). This suggests that different factors may control fecal indicator bacteria at different locations.

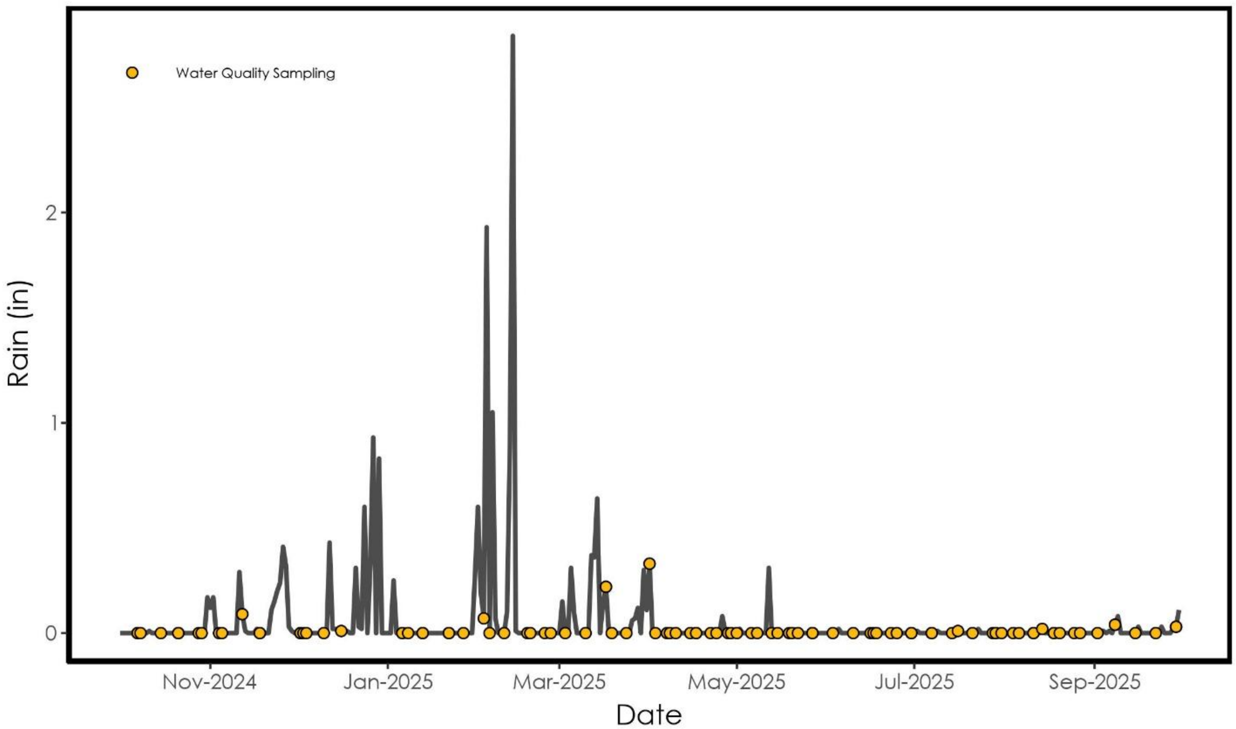


Figure 19: Cumulative daily rainfall for CIMIS station 104 at De Laveaga from October 1, 2024 through September 30, 2025

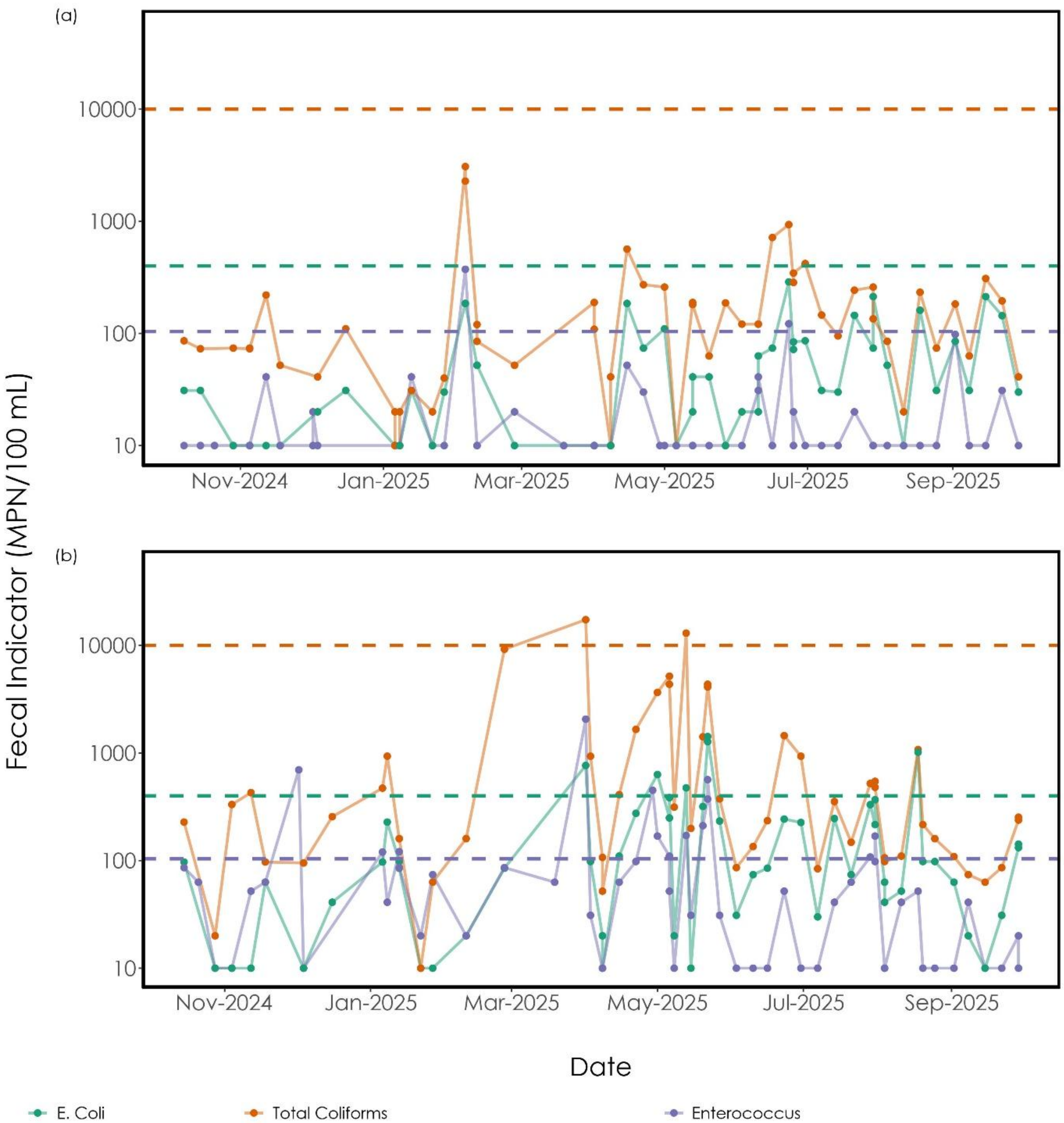


Figure 20: Timeseries of fecal indicator bacteria from (a) Cowell's Beach and (b) Capitola Beach at Soquel Creek during the 24-25 water year. Green indicates E.Coli, orange indicates Total Coliforms, and purple indicates Enterococcus. Dashed lines show health advisory.

Local Area Management Program (LAMP) 5-year assessment

A multi-year assessment of the LAMP program was conducted this year to fulfill requirements of the State Onsite Wastewater Treatment Systems (OWTS) Policy. Although OWTS exist throughout Santa Cruz County, the highest density of systems occurs in the Santa Cruz Mountains along the tributaries and mainstem of the San Lorenzo River. To complement an assessment of OWTS permitting data, the water quality monitoring program conducted analyses to determine whether changes in nutrients (nitrate) and pathogens (E. coli) have occurred over time in the San Lorenzo River watershed, presumably due to LAMP management practices. Nitrate and pathogens are two of the main surface water quality parameters that can be affected by onsite wastewater disposal and were therefore the main focus of this water quality assessment.

Robust statistical analysis of water quality monitoring data can be problematic due to unevenly spaced sampling and gaps in data collection that commonly occur due to a variety of logistical constraints. Therefore, a recent technique was developed to analyze these types of datasets by first using general additive models to fill in gaps in time series data and then utilize meta-regression to test for statistical trends. This approach was used to analyze trends in nitrate time series over long (30-years) and short (5-year) intervals. In general, data suggest that in the upper portions of the San Lorenzo River and the tributaries that empty into the San Lorenzo River, nitrate concentrations have declined over the past 30 years (Figure 21). In the mid-section of the San Lorenzo River and the tributaries that empty into the San Lorenzo River, nitrate concentrations have remained the same or increased over the past 30 years (Fig. 10). In the lower section of the San Lorenzo River and the tributaries that empty into the San Lorenzo River, nitrate concentrations have declined over the past 30 years (Figure 21). Nitrate concentrations have also generally declined over the past 30 years within the Carbonera and Branciforte Creeks that drain into the San Lorenzo River (Figure 21). Over most short time periods (i.e., 5 years) we were unable to detect significant trends in the data. This suggests that long-term data are necessary to assess the effectiveness of programs like LAMP which are most likely to result in small magnitude changes over time.

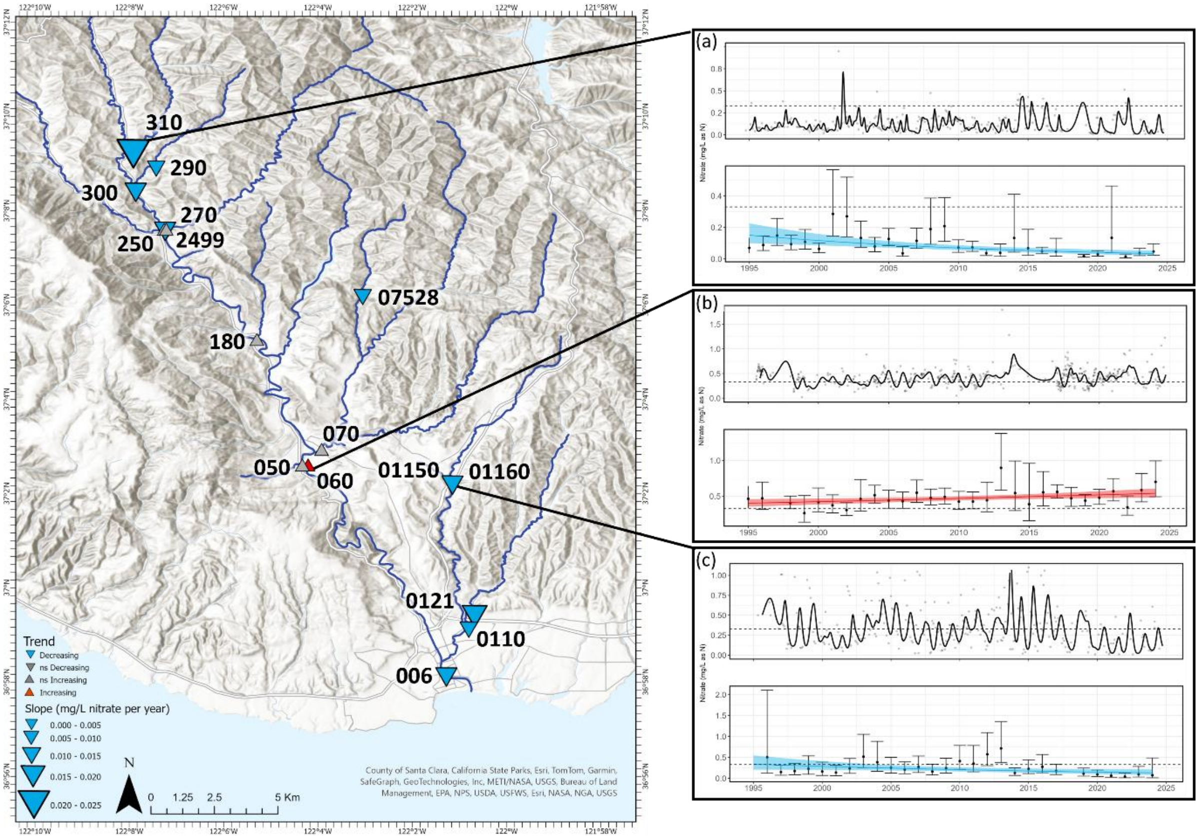


Figure 21: Trends in long-term (>28 years) nitrate concentrations during the dry season (June-September). Nitrate concentrations are declining in the northern portion of the watershed, represented by (a) Site 310, and in the Carbonera/Branciforte creeks regions, represented by (c) Site 01160. While nitrate concentrations are increasing at site 060, shown in (b). Upper panels show time series of nitrate concentrations over time (gray circles indicate grab samples, black line indicates GAM model fit). Lower panels show extracted dry season means and confidence intervals. Blue lines indicate declining nitrate concentrations over time, while red lines indicate increasing nitrate concentrations over time.

The two main criteria defined in the San Lorenzo River pathogens TMDL are: 1) 10% of pathogen samples within a 30 day period should not exceed the single sample health hazard threshold for E. Coli and Fecal Coliforms (i.e., 400 MPN/100mL) and 2) 30-day geometric mean of five or more E. Coli samples must be below 200 MPN/100mL. Unfortunately, due to logistical constraints, time series of geometric means that meet these criteria only exist for a handful of sites, but do not extend into 2024. Therefore, to evaluate current trends (2011-2024) in fecal indicator bacteria over time, we used logistic regression to assess how the probability of exceedance in E. Coli has changed over time. A probability of 0.1 is equivalent to 10% of samples exceeding the health advisory threshold. We did not detect significant trends in exceedances for any sites within the San Lorenzo River watershed and at only a single site (Aptos Creek at Mouth) across all datasets. Although pathogens do not appear to be declining over time, it is important to note that these data are not only highly dynamic but also exhibit a high degree of uncertainty (i.e., large error around each measurement). Therefore, it may be difficult to detect trends in these data at the current sampling frequency. Another issue with ongoing analysis of E. coli or Fecal Coliforms is that these bacteria can come from a variety of sources, some controllable (e.g., septic system management) and other non-controllable (e.g., wildlife). However, health risk often differs depending on the source of the bacteria, with human sources leading to higher risk of illness in humans. Therefore, efforts are underway to determine the relative abundance of pathogens that originate from human fecal material and whether certain environmental conditions can predict whether pathogens will be mostly of human origin vs other sources. These studies will provide critical information to help improve best management practices that aim to reduce bacterial loads and improve water quality.

Battery Fire Response

On January 16, 2025 the Vistra Moss Landing Power Plant in Monterey County caught fire releasing potentially harmful heavy metals into the atmosphere. In order to characterize the magnitude and extent of heavy metal deposition in Santa Cruz County lakes and estuaries the Water Quality Lab collected water and sediments at key lake and estuarine habitats closest to the south Santa Cruz county boundary. Three lakes (Pinto, Kelly, Drew) and one region of wetlands (Struve, Harkins, and Watsonville sloughs), hereafter referred to as Watsonville slough, were chosen to collect surface water samples due to their proximity to the fire, ecological and economic importance, and the presence of long-term water quality monitoring data. Due to a lack of historical data on heavy metals, surface water samples were also collected at three additional water bodies (Antonelli Pond, Forest Lake, Loch Lomond reservoir) in Santa Cruz County to serve as control sites (i.e., provide background levels of heavy metals unassociated with the battery fire) that are presumably outside the area of heavy metal deposition. In addition to surface water investigation, Pinto Lake and Watsonville slough were selected as ideal locations to conduct additional sediment sampling. Metals detected in surface water and sediment samples were below U.S. Environmental Protections Agency screening thresholds, with the exception of elevated manganese levels at Drew Lake, Watsonville Slough and Antonelli Pond (Figure 22). However, these elevated levels were not consistent with deposits from a battery fire and are likely due to natural or other anthropogenic sources. Importantly, control samples did not differ from sites close to the battery fire, except for manganese and lithium. Interestingly, lithium was actually higher at control locations, a result that is not consistent with excessive heavy metal deposition at sites close to the battery fire. Follow up testing for heavy metals may be useful in contextualizing the results from this sampling event or to provide baseline estimates for future events.

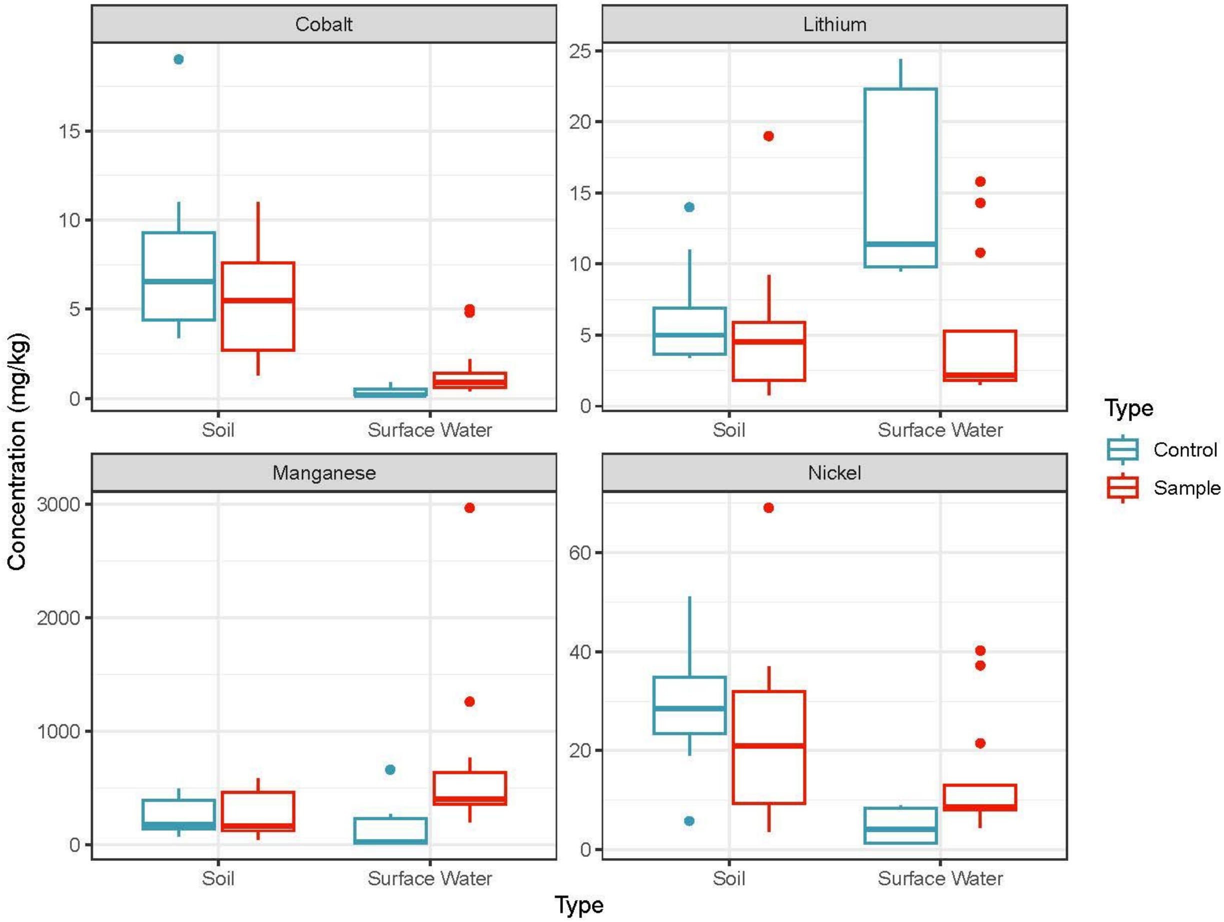


Figure 22: Boxplots of heavy metal concentrations in soil and surface water samples for control (blue) and sample (red) sites. Thick horizontal line depicts the median, box shows the interquartile range, and upper to lower whiskers show the %95 confidence interval with dots as outliers.

Drinking Water Quality and Supply Protection

San Lorenzo Valley Water District (SLVWD)

- Required water quality monitoring through the Unregulated Contaminant Monitoring Rule (UCMR5) was completed.

Scotts Valley Water District (SVWD)

- In support of the intertie 1 project, SVWD and SCWD staff is conducting a Disinfection By-Products formation potential study. Hydraulic modeling was conducted to determine approximate water age at extremities of the SVWD water system under the scenario of receiving SCWD water from the GHWTP. A bench test study of DBP formation potential is being conducted based on raw water supply of SCWD GHWTP water, and approximate age in SVWD's distribution system.
- Enhanced distribution system corrosion control monitoring program to include bi-weekly EPTDS pH/phosphate monitoring. Added 2 new distribution system sample sites to SVWD's quarterly corrosion control monitoring plan.

City of Santa Cruz Water Department (SCWD)

- Processed over 42,000 water samples testing for microbial, inorganic, organic, pesticides, herbicides, and radioactive contaminants.
- Continued ongoing monitoring of the Loch Lomond (4400912) and Santa Cruz (4410010) potable drinking water systems in compliance with drinking water permit regulations and bacteriological sample siting plan.
- Continued ongoing management and monitoring of Loch Lomond Reservoir for cyanobacteria and cyanotoxins.
- Executed the Santa Cruz Water Department's first nutrient-mitigation alum application at Loch Lomond Reservoir.
- Continued ongoing monitoring of raw source water and treated finished water for unregulated contaminants of emerging concern such as Per- and polyfluoroalkyl substances (PFAS)and pharmaceuticals and personal care products.
- Published the triennial 2025 Public Health Goals Report
- Published the 2024 Annual Water Quality Report in English and Spanish.
- Completed the initial monitoring requirements to comply with the final National Primary Drinking Water Regulation (NPDWR) for six PFAS.
- Completed the initial monitoring requirements to comply with the California State Water Resources Control Board Division of Drinking Water's (SWRCB-DDW) hexavalent chromium Maximum Contaminant Level (MCL).
- Maintained California State Environmental Laboratory Accreditation Program Certificate of Environmental Laboratory Accreditation for drinking water. Continued 6ppd monitoring in the San Lorenzo River in support of NMFS anadromous salmonid recovery efforts.

Soquel Creek Water District (SqCWD)

- Replacement of Soquel Creek Water District's Country Club well is part of a larger plan to build a water treatment plant at the location to treat 1,2,3 – Trichloropropane. Construction of the replacement well is complete and additional water quality investigation is ongoing.

- As part of the SqCWD's Well Master Plan and the Santa Cruz Mid-County Basin's Groundwater Sustainability Plan, SqCWD continues to redistribute groundwater pumping further inland where possible to reduce seawater intrusion at the coast.
- SqCWD maintains and continues to collect samples from a network of monitoring wells along the coastline to track groundwater levels and water quality to track seawater intrusion.
- SqCWD continues to make progress on compliance with the CA newly readopted standard for hexavalent chromium (Cr. 6). A pilot study to treat Cr.6 using reduction coagulation filtration (RCF) treatment to treat water from the Seascape, San Andreas, and Bonita Wells was completed last year. Design of the treatment system is now underway. Construction is targeted to begin in 2026, contingent on funding.
- Planning for construction of the Cunnison/Tannery Iron and Manganese Treatment Plant to support the new Cunnison well was underway in 2025.

The City of Watsonville Public Works (CoW):

- The City of Watsonville Environmental Sustainability Division is striving to implement multi-benefit projects that incorporate climate change, natural hazard mitigation, green infrastructure and habitat restoration benefits. The following grant funded projects are active:
  - Upper Struve Slough Wetland Enhancement and Public Access project funded by the Department of Water Resources in partnership with Watsonville Wetlands Watch is in implementation. This project will reduce localized flooding and improve habitat.
  - Watsonville Ramsay Park Phase IV California Natural Resource Agency grant will implement green infrastructure elements at the Nature Center. Improvements include green-permeable parking lots, bioswales, bioretention gardens, and reduction of heat island effects.
- CoW continues to monitor its groundwater quality. Six of the City's wells are impacted by the Chromium-6 maximum contaminant level (MCL) and will require treatment. The City has submitted a compliance plan to the Division of Drinking Water. This compliance plan outlines work completed to date, including treatment technology assessments, pilot testing, and preliminary design for the Chromium-6 impacted wells, as well as a schedule to implement the treatment design. The City continues its quarterly monitoring for both Chromium-6 and PFAS.

Non-Point Source Water Quality Programs

County Regulatory Programs

Community Development and Infrastructure (CDI) Public Works continue to review land development projects in the county and require stormwater mitigations for all projects that add or replace over 500 sq ft of impervious area, with quantitative mitigations for those over

5,000 sq ft. This will maintain—and in some cases improve—existing infiltration conditions, help reduce flash flooding, filter runoff from developed areas, and increase rainwater infiltration to mitigate the impacts of land development on groundwater resources.

Rural roads and home drainage

The RCD planned and facilitated a *Managing Erosion and Drainage on Rural Roads* technical training for the community. The 48 attendees included private property owners and public and private land managers. The RCD also conducted follow-up site visits to 13 unique properties who requested technical assistance related to rural roads or drainage and erosion issues around their home and/or streambanks. Recommendations were provided based on specific concerns. RCD discussed improvements that would improve infrastructure and also benefit water quality in local creeks. The RCD also fielded calls throughout the year from community members with questions regarding private roads maintenance, home drainage, and permitting.

Agricultural Water Quality

- In 2025, PV Water continued to operate multiple basin water quality monitoring programs in addition to operating water supply projects that achieve the dual purpose of augmenting water supply needs while helping to maintain or improve basin water quality. These programs are briefly summarized below:

    - **Surface Water Quality Monitoring Program:** PV Water staff routinely collects and analyzes water quality data from approximately 40 locations to provide information on the water quality conditions of streams, creeks, rivers, sloughs, and lakes in the Pajaro Valley. In addition, PV Water also maintains a large network of autonomous data loggers, and measures discharge to monitor hydrologic conditions with major focuses on the Watsonville Slough System, Pajaro River, and College Lake watersheds.

    - **Supplemental Water Quality Monitoring Program:** PV Water staff routinely collect and analyze water quality data from the CDS sources and points of delivery to provide information on changing water quality conditions, assess its suitability for agriculture customers, and to quantify nutrient and salinity inputs from water in the Delivered Water Service Area.

    - **Groundwater Quality Monitoring Program:** PV Water staff directly monitor groundwater quality from a network of over 180 public and private wells that staff routinely visit to obtain water level and water quality information. At minimum, staff visit these wells in the spring and fall of each year with a subset of wells monitored more frequently. PV Water’s groundwater monitoring network is supplemented by data collected by local water purveyors and other publicly available water quality datasets.

  - **Salt and Nutrient Management Plan:** PV Water developed and continues to implement the Salt and Nutrient Management Plan to ensure attainment of water quality objectives for protection of beneficial water uses and guide management of salts, nutrients, and other significant chemical compounds within the groundwater basin. As part of plan implementation, PV Water works with community partners, such as the RCD and U.C. Cooperative Extension, to continue to offer resources, rebates, and education opportunities that equip growers to efficiently irrigate and manage nutrient application.

- RCD Agricultural Program staff continued to work with NRCS, researchers, management agencies, funders and industry to get effective water quality best management practices developed, incentivized, and on the ground. RCD staff collaborated with California FarmLink and Kitchen Table Advisors to deliver a variety of technical assistance services targeting socially disadvantaged farmers and ranchers (SDFRs), including assistance with improved soil and nitrogen management for water quality regulatory compliance (Ag Order 4.0). Additionally, RCD continued to support growers to receive funding and implement projects through the California Department of Food and Agriculture (CDFA) State Water and Energy Efficiency Program (SWEEP) and the Healthy Soils Program (HSP). The SWEEP program offers growers the opportunity to apply for up to $200,000 for projects that improve water and energy use efficiency, and the HSP program offers growers up to $100,000 to help implement practices that improve soil health. In 2023 CDFA received additional funding and launched a pilot “block grant” program to support further distribution and implementation of SWEEP projects. The RCDSCC was awarded one of these block grants, in partnership with RCDMC and SMRCD. Through this block grant these three RCDs are currently supporting an additional pool of farming operations (35 total) for implementing water and energy conservation projects in the central coast region. The RCD also has secured funding to launch a compost cost-share program supporting growers to further the adoption of compost application to boost soil organic matter and increase soil water holding capacity.

Section 3: Natural Resources and Flood Management

Watershed management is a critical component of water supply and water quality. What happens on the landscape and in the watersheds impacts the availability and quality of water, and the damage caused by droughts and storms alike. Watershed health is also critical to the environmental users of water. The County, along with regional partners, has made deep commitments to fisheries resources and continues to implement multi-faceted approaches to ensure these species thrive. As these species are an indicator of overall watershed health, efforts to improve these populations have cascading benefits for all users of the watershed.

Fisheries Monitoring and Protection

County Water Resource Program and partner agency staff continue to implement various programs and projects to benefit steelhead and coho salmon habitat that is degraded due to historic and current land and water use. Coho salmon are listed as endangered under both the state and federal Endangered Species Act (ESA) and are critically endangered in Santa Cruz County. Steelhead are listed as threatened under the Federal ESA and continue to persist in most county streams at low to moderate population numbers. Current activities focus on improving dry season streamflow and habitat complexity, and wet season fish passage, in addition to ongoing monitoring efforts.

Juvenile Steelhead and Stream Habitat Monitoring:

- The Santa Cruz County Water Resource Program continues to partner with local water agencies and consulting fishery biologists to perform juvenile steelhead and habitat monitoring. This long-term, annual monitoring program measures the density of juvenile steelhead at monitoring sites in the San Lorenzo, Soquel, Aptos, and Pajaro watersheds. Population estimates are also made for Aptos and Pajaro Lagoons. The program also assesses habitat conditions for steelhead and coho salmon and helps inform conservation priorities throughout the County.

  During summer/fall of 2025, roughly 39 stream sites were sampled by electro-fishing (26 in San Lorenzo River, 9 in Soquel Creek, and 4 in Aptos Creek watersheds), and 2 lagoons were sampled by seining (Aptos and Pajaro Lagoon; Soquel Lagoon was also sampled through the City of Capitola). Streamflow was measured in several locations in each watershed. The field monitoring season ended in October 2025 and data are currently being analyzed. A few preliminary updates are provided below:

  - Adult steelhead spawned at all upper watershed sites, as indicated by young-of-the-year (YOY) presence at all sites.
  - Relative to water year 2024, water year 2025 had fewer stormflows to support adult spawning, and had below median baseflows which support juvenile growth rates. YOY steelhead growth rates appeared to be slower in 2025 than

2024, with fewer YOY reaching at least 75 mm Standard Length (SL), which is roughly the size of fish that are more likely to smolt (migrate to the ocean) during the upcoming spring season. Fish smaller than 75 mm are more likely to remain in the stream for another year and may or may not survive to smolt.

- Because YOY growth rates tend to be lower after milder winters compared to wet winters with higher baseflow and faster growth rates, site densities and production of larger juveniles (greater than or equal to 75 mm SL) are important metrics. Comparison of 2025 juvenile densities to long term averages at sampling sites, and juvenile production indices for larger juveniles, will be forthcoming after data analysis.
- The Aptos Lagoon juvenile steelhead population estimate was 85 in 2025. It was 312 in 2024. Most juveniles are much larger than 75 mm SL in Aptos Lagoon. Tidewater gobies were detected in the lagoon in 2025.
- No steelhead were detected in Pajaro Lagoon in 2025, as has been the case since monitoring began in 2012. One tidewater goby was captured during sampling in 2025.

- Fish Monitoring in North County Streams: Dr. Jerry Smith, Emeritus Professor, San Jose State University, performed annual fish monitoring at 5 sites in Waddell Creek during October 2025 (3 mainstem sites and one site on each of the east and west forks). Habitat conditions are relatively poor and have not yet improved substantially from the impacts of the CZU Fire and the heavy storms in 2022-2023. Spawning and rearing habitat is negatively impacted by sedimentation, particularly from the steep slopes on the west branch of the creek, which are largely unvegetated. Many pools downstream of this area are filled with sediment.

  Steelhead numbers were slightly improved compared to 2024, but were still extremely low. In 2025, there was a mean of 7.7 young of year (YOY) and 0.6 yearlings per 100 feet of stream. In 2024 there was an average of 4.8 YOY per 100 ft, and 1.8 yearlings and older per 100 ft.

  Five (5) juvenile coho were captured this year (0.6 / 100 ft) at 4 of the 5 sample sites. The last coho were captured in 2022 (1.9 per 100 ft). Since coho have a 3-year life cycle, the fish captured this year are of the same cohort.

  NOAA staff also conduct fish surveys in north Santa Cruz County watersheds and the Cities of Santa Cruz and Capitola also perform annual monitoring in County streams and lagoons. Results of those monitoring efforts for 2025 were not available at the time of this report.

Santa Cruz County Streamwood Program

- The goal of the County Water Resources Stream Wood Program is to increase the amount of stream wood (large woody material) in County streams. Stream wood provides essential habitat complexity for steelhead and coho salmon and helps support many other important watershed functions. The Stream Wood Program is guided by County policy that was adopted by the Board of Supervisors in 2009. Core program activities include public outreach, educating landowners and residents about the benefits of naturally recruited stream wood, and modifying (cutting or repositioning) stream wood only when necessary to protect property, infrastructure, habitat or safety. The program is run by County Environmental Health (Water Resources), in partnership with the Community Development and Infrastructure Department (Public Works). Program activities are overseen by resource management agencies including California Department of Fish and Wildlife (CDFW), NOAA Fisheries, and consulting experts in engineering, hydrology and geomorphology. Stream wood modified through the program is permitted by CDFW through its Lake and Streambed Alteration Agreement Program.

  This year the program received fourteen (14) requests for assistance which is a decrease compared to the past two water years. Stream wood was modified at a total of three sites, including one site where wood was successfully repositioned but not cut. In most cases, stream wood was preserved and property owners were amenable. See Figure 23 for an example.

Figure 23: Streamwood Complex at Big Creek in May 2025. These logs were left to provide habitat as they posed no immediate or direct threat.

  The Stream Wood Program was awarded a $500,000 grant from the California Wildlife Conservation Board which became active in July 2024. The funding augments the relatively small program annual budget of $40,000. This year grant funding was leveraged to make updates to [Stream Wood Program website](), develop new outreach materials, develop a new data management system, and to provide an annual training to CDI Public Works, Drainage Division staff. Program staff and consultants also began studies of stream reaches that are challenging from the perspective of stream wood management. These challenging stream sections have a high frequency of occurrence of stream wood that needs to be modified (and a disproportionate amount of program resources are directed to these areas). One product of the study will be hydrologic basemaps that will provide decision support when stream wood occurs in these areas. The maps can be referenced in the field to show the location of the stream wood, as well as the stream elevation and velocity under different flow scenarios. The tool will also show locations of buildings and topography to support management decision-making.

Fish Passage and Habitat Projects

- Water Resources Program staff continue to support efforts to maintain and improve fish passage in streams throughout the County.
    - The County of Santa Cruz Stream Crossing Inventory and Fish Passage Evaluation report identifies current priorities for fish passage among the County's road stream crossings. The 2022 update identifies locations on Casserly, Lompico and East Liddell Creeks where culvert replacement could improve passage for steelhead, aquatic and terrestrial animals. The study informs ongoing culvert replacement projects by Santa Cruz County CDI.
    - In 2014, County Water Resources staff completed an inventory and assessment of steelhead passage barriers on Branciforte Creek. With this report, RCDSCC has completed the removal of 3 barriers.
    - County staff continue to monitor and maintain fish ladders located in streams throughout the County.
    - Water Resources staff participate in the Caltrans FishPAC, a group dedicated to improving fish passage at state road crossings. In Santa Cruz County, Caltrans is actively working on 4 of the 6 highway crossings identified as high priority for replacement or remediation.
    - Water Resources staff continue to partner with RCD Santa Cruz County to outreach to property owners on the San Lorenzo River with the objective of identifying opportunities to remove fish passage barriers. This effort wis informed by the inventory completed by Santa Cruz County staff in November 2023, that identified historical, anthropogenic (human-built) structures on the San Lorenzo River that affect fish passage.
- The City of Santa Cruz Water Department received a Section 10(a)(1)(b) Federal Endangered Species Act permit related to their Anadromous Salmonid Habitat Conservation Plan for activities which may affect coho and steelhead. Additionally, they began implementation of an expanded juvenile salmonid monitoring program and conservation actions related to this permit including a "Non-Flow Conservation Fund" stream wood and riparian vegetation enhancement project on San Vicente Creek. This project is being completed in partnership with the RCD, BLM and other San Vicente Redwoods partners.
- Other City of Santa Cruz activities include:

- The City of Santa Cruz Water Department also received an award from the National Marine Fisheries Service for being a good partner in coho recovery.
- The City of Santa Cruz Water Department was successful in getting their water rights modified such that implementation of the aforementioned Section 10 permit and more sustainable regional water resource management is enabled.
- The City of Santa Cruz Water and Public Works Departments continue to seek funding for completion of the Branciforte Creek Flood Control Channel Fish Passage Enhancement Project design.
- The City of Santa Cruz Water Department and the RCD reinitiated planning for passage improvement at the City's historic Branciforte diversion site.
- The City of Santa Cruz Water Department initiated preliminary planning for Tait Diversion and Coast Pump Station improvements including fish passage and screening improvements.
- The City of Santa Cruz Water Department was a sponsor of lead tours and presented at the Salmonid Restoration Federation Conference.
- The City of Santa Cruz Water Department continued ongoing, routine monitoring related to fisheries, hydrology, water quality and rare terrestrial species relative to Water Department environmental regulatory compliance. Notably, no coho were observed in City water supply watersheds but they were observed in San Vicente Creek. Non-native juvenile chinook salmon were also observed in the San Lorenzo River, as was a (also non-native) spiny softshell turtle. Water Department staff also confirmed an earlier observation from the lower San Lorenzo River of a western pearlshell mussel – which had not been seen in the watershed in several decades.

- The Scott Creek Coastal Resiliency Project is a unique, integrated bridge replacement and ecological restoration project with multiple benefits. This project will improve community and highway resilience to climate change and sea level rise, will implement a major recovery action for endangered coho salmon and a suite of other listed species, and will improve public coastal access, amongst many other benefits. For nearly 10 years the Resource Conservation District of Santa Cruz County, the Santa Cruz County Regional Transportation Commission, and Caltrans, along with our state and federal resource agency partners, have collaborated to re-envision how to plan for major transportation infrastructure projects by focusing on first understanding the needs of the ecosystem, then designing infrastructure and restoration activities together to meet those needs. This year, with funding from the SHOPP program Caltrans worked with the RCD and partners to move the project into the environmental review phase. With funding from the Coastal Conservancy, the RCD continued active engagement of the Integrated Watershed Restoration Program

Technical Advisory Committee in the process, and advanced ecological components of the project in partnership with Caltrans.

- RCD watershed restoration program staff continued to advance restoration planning efforts along lower Scott Creek at Swanton Pacific Ranch with funding from the California Department of Conservation and State Coastal Conservancy. The project area is just upstream of the Scott Creek lagoon, which will be restored as part of the Scott Creek Coastal Resiliency Project. The project will eventually restore over 1 mile of creek to benefit listed species of fish, amphibians, and reptiles, and integrate with the lagoon restoration to significantly move the needle on species recovery. The project will also enhance the resiliency of working lands on the ranch and provide opportunities for student engagement. The project reached the 65% design milestone this year and implementation is currently anticipated to begin in 2027.

- In coordination with state and federal agencies and conservation partners, RCD watershed restoration program staff successfully implemented streamwood enhancement projects on San Vicente Creek (Cotoni Coast Dairies National Monument and San Vicente Redwoods preserve) and Aptos Creek (Nisene Marks State Park). The San Vicente Creek project also included significant removal of invasive plant species (primarily Cape Ivy and Clematis). The streamwood enhancement and invasive species removal work along San Vicente Creek was funded by the City of Santa Cruz's Non-Flow Conservation Fund (NFCF), which is associated with the City's Salmonid Habitat Conservation Plan. The RCD is co-managing the NFCF with the City and is the implementing partner for projects funded by the NFCF.

- RCD staff also implemented a floodplain enhancement project (invasive plant species removal and native plant revegetation) on the East fork of Soquel Creek (Soquel Demonstration State Forest) and the final phase of wetland enhancements at the Seascape Uplands Preserve to provide improved rearing habitat for the endangered Santa Cruz long-toed salamander.

- RCD staff continued to advance barrier removal projects in the Branciforte Creek watershed, a focal watershed for salmonid recovery. Specifically, in coordination with the City of Santa Cruz, RCD staff continued to help seek funding for the next design phase of the Branciforte Creek Flood Control Channel Fish Passage Project, as well as advanced planning and design for the removal of an old non-functional City flashboard dam, which has reached the 65% design milestone. In addition, RCD staff continued to work with the California Department of Parks and Recreation (State Parks) Santa Cruz District to develop restoration projects on State Parks lands impacted by illegal cannabis operations. Project examples include floodplain restoration, streamwood enhancement, decommissioning of legacy logging roads, and development of management plans for sensitive sandhills habitat.

- As construction of the College Lake Project continues in 2025, PV Water continues to implement a robust mitigation and monitoring program that was outlined in the College Lake Project EIR to protect environmental resources during construction. The project itself includes construction of a fish passage structure which will allow movement of fish into and out of College Lake. In prior conditions, fish did not have safe passage through College Lake including endangered steelhead.

Watershed Management Activities

- Office or Response, Recovery, and Resiliency (OR3) and County Parks have taken the lead on the implementation of Measure Q for the County. The Citizens Oversight Advisory Board (COAB) was seated in early 2025 and has been hard at work preparing for the first round of funding to be utilized. The Measure requires a Vision Plan, which was approved in October.

  The Vision Plan covers six thematic areas of work being done in the County: water resource protection, wildfire risk reduction, parks, coastal protections, agricultural and working lands protections, and wildlife and habitat protections.

  Out of the six areas, grant making is being prioritized in three areas: water resource protection, wildfire and forest health, and parks access. Other topics, particularly wildlife habitat, are covered to a large extent by the work that would be done with forest health and water. The Measure language also promotes leveraging additional funding such as state and federal grants, looking for multi-benefit projects, and equity. Other priorities include trying to be shovel ready. A lot of the conversations have been around the Vision Plan and Grant Program. The goal is to have the grant program Request for Proposals (RFP) out in late 2025 or early 2026. The grant program will be iterative, it will not be perfect in year 1, but they can learn from recipients how to improve it.

- CDI - Environmental Planning continues to protect aquatic habitat and riparian corridors through implementation of the County's Riparian Corridor and Wetlands Protection ordinance (16.30) by preserving, protecting, and restoring riparian corridors and wetlands to improve the extent and quality of riparian habitat. The polices outlined in Chapter 16.30, including limiting development activities within riparian corridors, safeguard protections for water quality, help prevent erosion, protect and enhance wildlife habitat and corridors, and protect aquatic habitat for sensitive species.

- CDI - Environmental Planning also protects aquatic habitats and wildlife that depend on those habitats for survival through implementation of the Sensitive Habitat Protection ordinance (16.32). By limiting development within sensitive habitats and working to restore the quality and extent of damaged sensitive habitats, including the restoration of native vegetation and natural drainages, Environmental Planning works to protect and enhance the functional capacity, productivity, and biological diversity of aquatic habitats throughout the County.

- The City of Santa Cruz Water Department continued ongoing management work including:
    - Review of well permits in potential karst protection zones,
    - Maintenance of watershed divide and stream crossing signs,
    - Continued onboarding of new forestry and fisheries consulting teams,
    - Revised the City's old growth tree policy to enable better protection of old growth trees and more rigorous forest management on their watershed lands,
    - Began development of a forest health project on the Newell Creek watershed lands,
    - Partnered with the RCD on a new fuel break across the Zayante watershed lands
    - Continued fuel management around Loch Lomond Reservoir, Laguna and Zayante watershed properties,
    - Continued invasive species control at Loch Lomond Reservoir,
    - Prohibited fish stocking at Loch Lomond Reservoir in the interest of preventing golden mussel from being introduced,
    - Collaborated with the County on emergency access planning in Lompico,
    - Provided regulatory support for operational emergencies,
    - Participated in San Loreno watershed-wide foothill yellow-legged frog surveys which confirmed their absence from the watershed,
    - Continued implementation of watershed education and interpretive programs, Continued assistance with San Lorenzo River lagoon and lower San Lorenzo River management,
    - Participated in several regional, large-scale restoration grant proposal efforts,
    - Pursued enforcement of illegal stream diversions and other unpermitted developments and other actions such as sandbar breaching that have potential water resources or fisheries implications,
    - Continued coordination with hazardous materials spill incident responders,
    - Continued implementation of the Low Effect Mount Hermon June beetle Habitat Conservation Plan and Operations and Maintenance Habitat Conservation Plan,
    - Continued surveillance and patrol of watershed lands to ensure protection of drinking water sources.
    - The City of Santa Cruz Water Department has also continued to participate heavily in the regional response to the houseless community and threats to drinking water sources including coordination on flood – related riparian camp

evacuations, patrols of key riparian areas along the San Lorenzo River and tributaries and support for camp cleanups.

- In 2025, PV Water continued to work with regulatory and technical experts such as aquatic ecologists, archeologists, biologists, Indigenous American monitors, and federal resource management agencies to support the operation of existing water supply facilities and guide construction of the College Lake Project. In addition, PV Water leveraged these experts to survey natural resources, guide development of the WSS-MARR project, and refine mitigation measures. The work performed included programs to observe and collect information on environmental and biological resources in and around the College Lake Project, the Harkins Slough Facility as well as the proposed Struve Slough Project locations with focused interest in cultural resources, waterfowl/nesting birds, South-Central California Coast Steelhead, and California Red Legged Frogs.

- PV Water also continues implementation of the College Lake Integrated Resources Management Project Adapted Management Plan 2022 which guides project operations and lake management. The plan provides adaptive management framework, including metrics, triggers, and management actions, to guide operations and maintenance of the project with a focus to mitigate impacts to the College Lake ecosystem. PV Water prepared a pre-operation Adaptive Management Plan annual report for water year 2024 and is conducting management and monitoring activities in accordance with the plan.

- As part of PV Water's broader basin management activities and planning under the Sustainable Groundwater Management Act (SGMA), it evaluated and considered groundwater dependent ecosystems (GDEs) when developing the GSU22 and adopting sustainable management criteria for interconnected surface waters. It was determined that at the time, there was minimal connection between surface water and groundwater and that there is no potential for significant and unreasonable depletions of interconnected surface water due to the existing disconnect. However, PV Water, through implementation of projects and programs to achieve sustainable groundwater resources, aims to increase the frequency and duration of hydraulic connectivity between groundwater and surface water where reasonably achievable. Enhanced connectivity would provide greater opportunity for groundwater dependent ecosystems to be restored, developed, expanded, and/or improved. PV Water is also currently planning to construct a series of eleven new wells throughout the basin to expand the available data on interconnected surface water and groundwater resources.

- PV Water has continued to improve, update, and expand the capabilities of the Pajaro Valley Hydrologic Model (PVHM) including the simulation of future scenario planning with climate change. The PVHM is one of the principal planning tools for the agency and in collaboration with the United States Geological Survey, it will be used to evaluate future basin conditions that support sustainable water resources for all beneficial uses including instream needs. Previous climate scenario modeling included modeling of future variable climate scenarios as well as an uncertainty analysis to support basin management planning. In 2025, PV Water worked to update the PVHM historical period through 2024 and subsequently will be conducting a validation of the most recent calibration and scenario modeling in the coming year.

- During 2025, PV Water continued to lead the Pajaro River Watershed Resilience Program planning project. The Pajaro River Watershed is one of five watersheds in the state to receive grant funding from DWR to develop resilience plans as part of the Watershed Resilience Panning Pilot Program. The program underscores the importance of watershed-based solutions, climate resilience, and equity through collaboration of local partners. The program builds on previous regional planning efforts such as the Pajaro River Watershed Integrated Regional Water Management Plan (IRMWP) and the Pajaro Compass Network. In 2025, PV Water, along with its support team, hosted five advisory group meetings and four watershed network workshops to guide development of the plan. Major work tasks completed include identifying and assessing existing regional networks; developing a watershed network; delineating the watershed area; developing a watershed resilience vision; assessing climate vulnerabilities and the state of the watershed; assessing vulnerabilities and risks; developing adaptation. The plan, which will be published in 2026, will describe priority strategies and actions for implementation based on community input to enhance resiliency of the Pajaro River Watershed and its water resource systems.

- During 2025, PV Water, along with program partners that include the Land Trust of Santa Cruz County, the RCD, Watsonville Wetlands Watch, the Pajaro Regional Flood Management Agency, UCSC, and the Nature Conservancy, continued to lead the Pajaro Valley Multibenefit Land Repurposing Program. The program is funded by an $8.89 million grant from the California Department of Conservation and seeks to repurpose agricultural land to reduce reliance on groundwater while providing community health, economic wellbeing, water supply, habitat, renewable energy, and climate benefits. In 2025, the program provided funding for implementation projects including approximately $2.375 million for the College Lake Project and approximately $800,000 for the Land Trust of Santa Cruz County's acquisition of agricultural lands impacted by flooding and brackish waters to repurpose as wetland and natural habitats. The program incorporates development of a multibenefit land repurposing plan to establish goals and a vision for multibenefit land repurposing in the Pajaro Valley; to develop future project prioritization and selection framework; and to develop monitoring and reporting framework on the program. In late 2025, the plan development process will begin with the first of four workshops of a steering committee.

Disaster Preparation and Response

County Hazard Mitigation and Flood Management Activities

- The County completed its first Multi-Jurisdictional Hazard Mitigation Plan (Figure 24). It's a 5-year planning process required by FEMA. It can lead to grants for proactive investment to reduce repetitive loss over time. The Plans are nationally standardized. The process included 10 participating agencies including 3 cities and several water agencies. They went through the planning exercise together. The Office of Response, Recovery, and Resiliency received FEMA funding to do this planning effort. https://mitigatehazards.com/santacruzmjhmp/

Figure 24: Hazard Mitigation Plan Cover

- As part of the County efforts to update the Local Hazard Mitigation Plan (LHMP) a detailed study of frequently flooded areas within the County, known as a repetitive loss area analysis (RLAA), was prepared. The RLAA improves the Counties' understanding of the causes of flood risk and the expanse of flooding in these regions. In September 2025 County staff from CDI, EH, and OR3 convened to discuss additional causes of flooding in frequently flooded areas. Potential mitigation and management options were discussed and incorporated into the updated LHMP.

- Results for the Community Rating System (CRS) 5-year cycle audit were received in October 2024. The CRS is a voluntary program the County participates in to improve floodplain management and increase development standards within the flood hazard areas. The County continues to maintain a class 8 rating for the additional floodplain management activities implemented by CDI, which affords residents a 10% reduction in their flood insurance premiums. Activities include, but are not limited to, documentation and protection of natural open spaces located in the FEMA designated Special Flood Hazard Area, higher regulatory standards for new development within the floodplain, and Stormwater and floodplain management planning activities.

- The Draft Zone 5 Drainage Master Plan was completed and accepted by the Zone 5 Board of Directors on January 30, 2024. Zone 5 covers, generally, the urban unincorporated areas of Soquel, Live Oak and the Pleasure point areas as well as the City of Capitola. The Zone 5 Master Plan updates included condition and capacity assessment of the large stormwater conveyances, 36" or larger in pipe diameter, within the Zone. Evaluation of the maintenance program and recommendation for improvements were included in the Zone 5 Master Plan update. Aside from assessing the condition and capacity of the larger storm water conveyances within the Zone, an additional goal of this Master Plan update is to generate detailed cost estimates for the current and proposed maintenance and Capital Improvement Program (CIP) upgrades for large drainage conveyances. The estimates are being utilized to seek additional sustainable funding sources for the improved maintenance and the CIP implementation from the benefiting property owners in the Zone. Once the funding study portion of the master planning efforts is completed the Zone 5 Drainage Master Plan will be finalized for acceptance by the Zone 5 Board of Directors.

Pajaro Regional Flood Management Agency (PRFMA)

- The Pajaro Regional Flood Management Agency (PRFMA) is a joint powers authority of the County of Santa Cruz, Santa Cruz County Flood Control and Water Conservation Zone No. 7, the County of Monterey, the Monterey County Water Resources Agency, and the City of Watsonville. Formed in 2021, the agency will plan, finance and implement projects and programs to reduce flood risk from the lower Pajaro River and its tributaries in Santa Cruz and Monterey Counties. Some of the PRFMA's accomplishments in 2025 include:

- The federal Pajaro River Flood Risk Management Project, now called the Pajaro River at Watsonville Project, is entering construction and earth moving for Reach 6 (Corralitos Creek between Green Valley Road and East Lake Avenue). Design of the remaining reaches (Reaches 5, 4, and 2/3 combined) is underway with the Army Corps, as well as early Right of Way activities for Reach 5. AB 876 and other strategic implementation programs will allow the project to shave years off of its total construction timeline. However, the project will still take 5-10 years to fully complete.

- PRFMA is part of a larger consortium recently awarded $71M from the NOAA Climate Resilience Regional Challenge grant. PRFMA will receive just over $10M as part of this grant to plan, design, and build additional flood risk reduction along the Santa Cruz County side of the Pajaro River upstream of the confluence with Salsipuedes Creek, where the USACE is not planning on introducing improvements as part of the Pajaro River at Watsonville Project. Project planning and implementation in this area will extend over the next five years or so. A contract was recently awarded to a consultant team to begin feasibility phase work for the project.

- PRFMA was also recently awarded $400,000 from the FEMA BRIC Program to start feasibility work to develop flood risk reduction solutions along what's called Reach 1 of the Pajaro River, extending downstream from the Highway 1 bridge to the ocean. It is anticipated that solutions developed will provide at least 100-yr flood protection, with implementation extending over the next 5-10 years. A consultant team will be selected in early calendar year 2026 to begin feasibility work.

- PRFMA has also partnered with AMBAG (the Association of Monterey Bay Area Governments) to receive a $2.5M grant from the California Department of Transportation to plan improvements to the Highway 1 crossing over the Pajaro River. Again, ultimate construction of any improvements will not likely be completed for 8-10 years, but this effort, combined with the progress mentioned above, will build capacity to the Pajaro River system where it is needed and outside of the Pajaro River at Watsonville Project. A consultant team, led by Mark Thomas Co., is working on initial planning phase elements of the project.

- PRFMA continues to synergistically collaborate with USACE's Engineering With Nature Program, California Department of Water Resources, other entities conducting Ecological Floodplain Inundation Potential modeling, the University of California, and California State University to examine how multibenefit habitat and groundwater recharge features can be incorporated into all projects. PRFMA is also working with UC Berkeley and the Vienna Research Center for Visual Computing to create a digital twin model of the lower watershed to explore project design and operational strategies to lower flood risk.

- PRMFA has also completed the rehabilitation of the upper 3000 feet of the Monterey County levee system on the Pajaro River to introduce more resilience to the levee system there, ahead of full reconstruction as part of the Pajaro River at Watsonville project. Furthermore, a large encampment clearing was conducted on both sides of the river system, removing over 200 tons of trash and debris from the river corridor as well as safety hazards that compromise levee and river maintenance operations.

- PRFMA is also a principal participant in the Pajaro River Watershed Resiliency Plan with Pajaro Valley Water Management Agency under a pilot program grant with CA DWR., is partnering with the Land Trust of Santa Cruz County in their Resilient Pajaro Estuary Program planning study, and has begun the development of a Local Hazard Mitigation Plan using a FEMA HMGP grant.

Forest and Fuels Management for Water Resource Protection and Watershed Health

- In 2025, the RCD partnered with NRCS to provide technical assistance and conservation planning to forestland owners throughout the county.
- The RCD implemented the Lockhart Gulch Shaded Fuel Break, a strategic wildfire resilience project spanning over 100 acres and 3.5 miles along a ridge between

Zayante and the City of Scotts Valley. This project was made possible through collaboration with multiple voluntary landowners, including the San Lorenzo Valley Water District, and the City of Santa Cruz Water Department. The shaded fuel break enhances wildfire preparedness for surrounding communities and protects critical water resources by reducing the risk of erosion and sedimentation following fire events. Through mindful implementation, the project promotes healthy watershed conditions and overall ecosystem function.

- The RCD also completed a 62-acre ecological restoration project with the San Lorenzo Valley Water District, removing invasive vegetation and improving sensitive Santa Cruz Sandhills habitat. This effort restored native plant communities to support rare and endemic wildlife and promote proper ecosystem function.
- Despite reduced capacity caused by delays in federal funding, with support from a private donor, Measure Q, and the CA State Coastal Conservancy, the RCD's no-cost chipping program continued to support defensible space creation for homes in the wildland-urban interface, serving over 440 individuals in the spring of 2025 and ## individuals in the Fall of 2025.

Attachment 1: Water Use in Santa Cruz County, 2025 (Data for smaller systems is from calendar year 2024)

Water Supplier	**Connections**	**Population**	**Water Use acre-feet/yr**	**Ground water**	**Surface Water**	**Recycled Water**	**Imported from Outside the County**
Santa Cruz City Water Dept.	25,001	94,626	7,616	10.00%	90.00%	0.00%	
Watsonville City Water Service	14,605	65,231	6,615	96.40%	3.60%		
Soquel Creek Water District	14,611	40,947	3,065	100.0%			
San Lorenzo Valley Water District	7,900	23,700	1,896	32.5%	67.5%		
Scotts Valley Water District	4,548	11,197	1,110	86.20%		13.8%	
Central Water District	826	2,726	375	100.00%			
Big Basin Water Company*	540	1,120	146	100.0%			
Mount Hermon Association	498	1,176	152	100.0%			
Forest Lakes Mutual Water Company	326	1,067	38	100.0%			
Smaller Water Systems (5-199 conn.)	2,565	7,978	1,321	86.0%	7.0%		8%
Individual Users*	8,000	21,000	2,350	95.0%	5.0%		
Pajaro Agriculture (SC Co only)**†			22,920	93.1%	1.6%	5.3%	
Mid- & North-County Agriculture*			2,400	90.0%	10.0%		
Totals	**79,420**	**270,768**	**50,004**	**79%**	**18%**	**2.7%**	**0.2%**
Summary by Water Source (acre-feet/year)				**39,355**	**9,189**	**1368**	**106**
Summary of Non-Agricultural Use (acre-feet/year)			**24,684**	**15,856**	**8,582**	**153**	**106**
Summary of Non-Agricultural Use (percent of total water use by source)				**40.3%**	**93.4%**	**11.2%**	**100.0%**

*Values are Estimates

** Includes a small number of water systems

† Recycled water source is the City of Watsonville

Attachment 2: Common Acronyms

AF	Acre Foot
AFY	Acre Foot per Year
BMP	Best Management Practices
CDI	Community Development and Infrastructure Department
CEQA	California Environmental Quality Act
CoW	City of Watsonville
CWD	Central Water District
DMS	Data Management System
DWR	Department of Water Resources
EIR	Environmental Impact Report
GSA	Groundwater Sustainability Agency
GSP	Groundwater Sustainability Plan
IRWM	Integrated Regional Water Management
JPA	Joint Powers Agreement
LAFCO	Local Agency Formation Commission
LID	Low Impact Development
MGA	Santa Cruz Mid-County Groundwater Agency
MGD	Million Gallons per Day
MGY	Million Gallons per Year
O&M	Operations and Maintenance
OR3	Office of Response, Recovery, and Resiliency
PPB	Parts Per Billion
PV Water	Pajaro Valley Water Management Agency
RCD	Resource Conservation District of Santa Cruz County
RWMF	Regional Water Management Foundation
SCWD	City of Santa Cruz Water Department
SGMA	Sustainable Groundwater Management Act
SLVWD	San Lorenzo Valley Water District
SMGWA	Santa Margarita Groundwater Agency
SqCWD	Soquel Creek Water District
SVWD	Scotts Valley Water District
UCSC	University of California, Santa Cruz

Attachment 3: Online Resources

County Water Resources Program	scceh.com/Home/Programs/WaterResources.aspx
County Water Quality Map	scceh.com/waterquality.aspx
County Steelhead Monitoring Program	scceh.com/steelhead.aspx
Santa Cruz County Office of Response, Recovery, and Resiliency	www.co.santa-cruz.ca.us/OR3.aspx
Central Water District	sites.google.com/view/centralwaterdistrict
City of Santa Cruz Water Department	www.cityofsantacruz.com/government/city-departments/water
City of Watsonville Public Works and Utilities	www.cityofwatsonville.org/590/Public-Works-Utilities
San Lorenzo Valley Water District (SLVWD)	www.slvwd.com/
Scotts Valley Water District (SVWD)	www.svwd.org/
Soquel Creek Water District (SqCWD)	www.soquelcreekwater.org/
Pajaro Valley Water Management Agency (PV Water)	www.pvwater.org/
Santa Cruz Mid-County Groundwater Agency (MGA)	www.midcountygroundwater.org/
Santa Margarita Groundwater Agency (SMGWA)	smgwa.org/
Resource Conservation District of Santa Cruz County (RCD)	www.rcdsantacruz.org/
Santa Cruz Integrated Regional Water Management Plan (IRWM)	www.santacruzirwmp.org/
Water Conservation Coalition of Santa Cruz County	watersavingtips.org/
Santa Cruz Countywide Data Viewer	sccwaterdata.us/#/html/home