Hazen *Technical Memorandum*

Valley Water

March 2, 2020

To: Samantha Greene, Ph.D.

From: Luke Wang
Jack Kiefer
Kinsey Hoffman
Leah Bensching

cc: Jing Wu, Metra Richert, Jessica Lovering

Technical Memorandum 3

Modeling Approach and Development

Introduction

Santa Clara Valley Water District (Valley Water) has developed a new model to forecast total water demand in Santa Clara County. Demand projections from the model will be used to support several planning initiatives and documents including:

- The 2021 Urban Water Management Plan (UWMP);
- Monitoring of and updates to the Water Supply Master Plan;
- Inputs to Valley Water's water supply planning model; and
- Evaluation of conservation programs and capital projects.

Valley Water manages a diverse portfolio of water supplies to provide water to Santa Clara County's 13 water supply retailers and non-retailer groundwater pumpers.[1] The majority of water users in Santa Clara County are customers of the water supply retailers. As a result, each retailer typically develops their own water demand forecasts. These forecasts are useful and have been used to inform Valley Water's prior UWMPs. However, Valley Water is responsible for County-wide water resource planning activities (e.g., groundwater management, treated water production, potable reuse development, surface water infrastructure management and development, and active conservation program implementation); collectively, these activities are better served by a consistent modeling approach and planning assumptions across the service area.

The purpose of this Technical Memorandum (TM 3) is to document the modeling approach selected to develop Valley Water's updated demand model. Major characteristics of the modeling approach include a statistical/econometric analytical framework, differentiation of rates of water use from drivers of growth, and model segmentation based on geography (e.g., retail agency), time of year, and water use sector. TM 3 also includes a summary of the statistical model fits and performance compared to historical

[1] Non-retail groundwater pumpers include private well owners that are outside of retailers' service areas.

Job no 20108-001

observations of water consumption. Discussions of model fits and performance are organized based on water use sector segmentation and includes the following sectors:

- Single family;
- Multifamily;
- Commercial, Industrial, and Institutional (CII); and
- Non-retailer groundwater pumpers.

The model sectors are designed to establish baseline demand projections without considering additional future water conservation. Projections of future conservation savings are generated separately by Valley Water’s water conservation model and then deducted from the baseline projections generated for the model sectors described herein.

Table of Contents

1. Modeling Approach

Valley Water's demand model is organized following the demand forecasting typology identified in TM 1.[2] This section provides a general overview of this approach to establish context for detailed discussions on model development in Sections 2 – 5 of this TM.

1.1 Model Segmentation

The demand model was segmented based on type of provider, i.e., retail agency or non-retail groundwater pumper. Within each provider type, the model was further segmented by geography, sector/billing classification, and time of year. For retail provided water, model geographies were based on each retail agency's service area within Santa Clara County. Billing classifications often differed among retail agencies necessitating standardization of billing classifications into common sectors (e.g., single family, multifamily, commercial, industrial, and institutional). Appendix A provides a detailed summary of the billing classifications for each retail agency, and the standardized sectors used for modeling; Valley Water directly solicited the retail agencies for input in standardizing billing classifications, particularly for classes that have the potential to span across multiple water use sectors (e.g., landscape irrigation and recycled water). Non-retail groundwater pumpers were organized geographically by groundwater basin charge zone, including W2 (representing the Santa Clara Plain sub-basin management area) and W5 (representing the Llagas sub-basin and Coyote Valley sub-basin management area). Water use classifications for non-retail groundwater pumpers are consistent across each charge zone and include agricultural, municipal, and domestic water use types. These water use classifications were ultimately organized into two model sectors, Municipal and Industrial (M&I) and Agricultural (Ag).

The retail agency demands were modeled using a monthly timestep, and non-retail groundwater pumper demands were modeled using an annual timestep. Non-retail groundwater pumper annual demands were then post-processed to monthly demands using a monthly distribution. Figure 1-1 further details the hierarchical structure of model segmentation.

[2] Technical Memorandum 1: Benchmark Analysis of Regional Demand Projection Models.

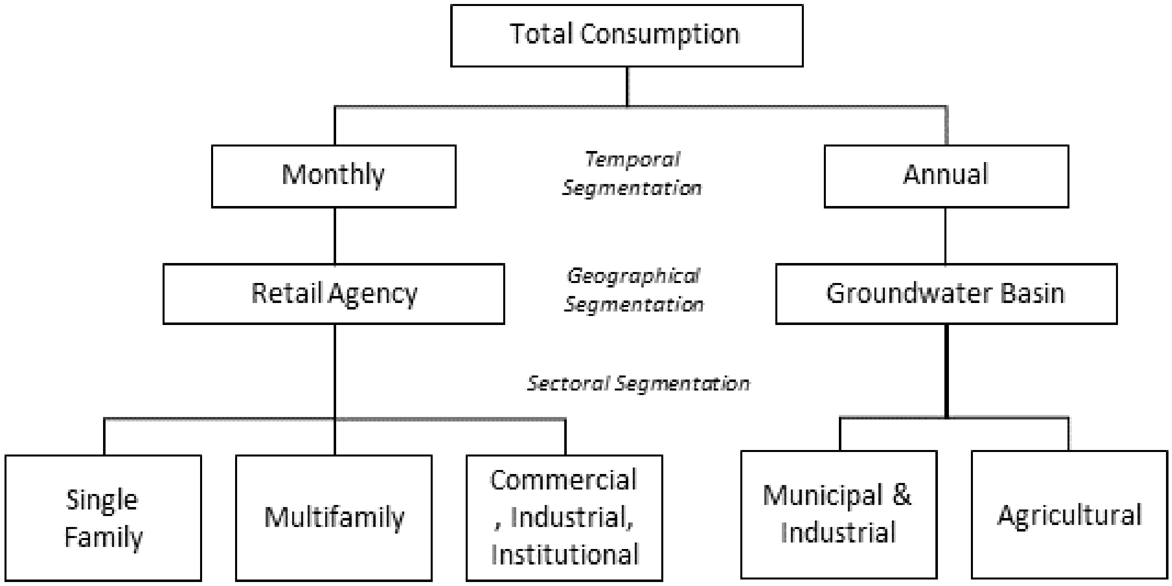


Figure 1-1: Hierarchy of Model Segmentation

1.2 Rate of Use Differentiation

Rate of use differentiation (i.e., characterizing consumption to reflect water using intensity) was applied in developing the retailer models. Rates of use were calculated given Equation (1) below, where for any given model sector Q reflects volumetric consumption, N is the count of driver units, and q is the rate of water use per driver unit.

$$Q \equiv N * \frac{Q}{N} \equiv N * q \qquad \textbf{(1)}$$

Rate of use differentiation requires a reliable and consistent historical driver unit dataset for model development and a corresponding future dataset representing projected driver unit counts. Consistent and reliable driver unit datasets for the retailer models were developed using data from the California Department of Finance (CADOF; historical data) and the Association of Bay Area Governments (ABAG; future projected data).[3] Corresponding driver units were not available for the non-retailer groundwater pumpers, so models were developed on a volumetric basis. Table 1-1 documents the driver units and corresponding rate of use for each retail model sector.

Table 1-1: Driver Units and Rate of Use for Each Retail Model Sector

Model Sector	Driver Unit (*N*)	Corresponding Rate of Use (*q*)
Single Family Multifamily	Housing units	Consumption per housing unit
CII	Employees	Consumption per employee
CII (Stanford)	Population	Consumption per capita

[3] Refer to Technical Memorandum 2: Data Collection and Review (TM 2).

1.3 Method / Statistical Approach

Valley Water collected historical consumption data from its retail agencies,[3] which generally spanned the period 2000-2018.[4] This dataset was sufficient from temporal, geographical, and sectoral perspectives (following sectoral standardization) to explore fitting customized statistical / econometric models identified in TM 1.[2] Development of historical econometric models provide a strong analytical benefit in forecasting demand, as they allow for the estimation of cause-effect relationships between weather, price, socioeconomic, and other factors that lead to variability in water demand. Quantifying these causal relationships allows for analysis of "what-if" scenarios that are uncertain, but important to consider for planning (e.g., climate change, development patterns, drought recovery).

Development of statistical / econometric models is an iterative process. Figure 1-2 and Table 1-2 outline the process used to fit the econometric models.

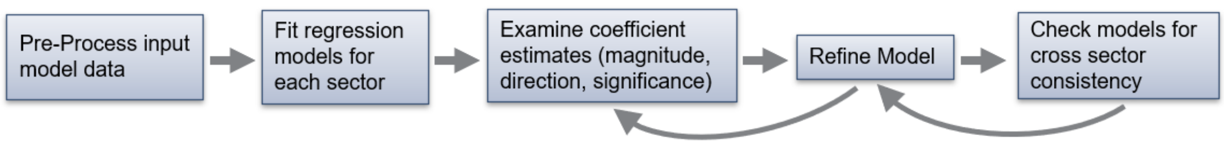


Figure 1-2: Process for Developing Statistical / Econometric Models

Table 1-2: Description of Model Fitting Procedures

Model Fitting Procedure	Description
Pre-process model input data[(a)]	Conduct necessary pre-processing calculations prior to model fitting, e.g.: • Geographical processing of driver units. • Calculate per-unit use. • Calculate natural logarithms of per-unit use and appropriate predictors. • Calculate departures from normal conditions for appropriate predictors (i.e., economic trend and weather). • Calculate any index, "dummy", or interacted parameters (e.g., seasonal cycle, geography, drought severity). • Smoothing monthly and bimonthly data to adjust for irregular billing cycles.
Fit regression models for each sector	Use statistical estimation software (e.g., R, SAS, EViews) to fit linear regression equations to per unit use with the initially selected predictor variables.
Examine coefficient estimates and measure of fit	Check measures of fit (e.g., R^2) and coefficient estimates for reasonable magnitude, direction/sign, and significance.
Refine model to improve measures of fit and coefficient estimates	If the model fit is poor or if coefficient estimates are illogical or insignificant, several actions can be taken, including but not limited to: • Identifying and removing outlier data points that have significant leverage on coefficient estimates. • Remove predictors with insignificant or illogical coefficient estimates from the regression equation. • Testing alternate specifications of predictor variables.
Check models for cross-sector consistency	Model fits and predictors are compared across sectors to judge estimates relative to prior expectations; e.g., testing if the relative effects of price and socioeconomic variables vary by sector in a logical way based on past experience.
[(a)] Model data pre-processing is detailed in TM 2.	

[4] Retail agencies submitted historical billing records of varying lengths. Sufficient retailers submitted records from 2000-2018 to establish model fits over the time period.

1.4 Summary of Model Predictors

Several model predictors were used to develop Valley Water's demand model. To be considered for use, potential predictors needed to pass the following conceptual criteria:

- Logical connection to explaining changes in water consumption;
- Historical record consistent with the time series of observed water consumption; and
- Availability of future projections consistent with the desired forecast horizon (i.e., 2020-2045) or a reasonable basis for assuming or generating projected values.

Initial selection of model predictors is discussed in detail in TM 2. However, during the model fitting process, derivatives of initial variables were also developed and included in subsequent model equations. One example is time lags on weather variables; supplementary variables were created from the temperature and precipitation time series at one to three-month lags. These lagged weather variables aimed to capture a delayed or persistent response in water use. A second example is an extended drought effect variable. The initial drought variables were directly calculated from historic water use restrictions. A supplemental drought variable was created that extended the last historic occurrence of mandatory water restrictions (2017) through the end of the historic dataset (2019); this "extended drought effect" variable was considered to represent inertia in behavioral changes in water use after the water use restrictions were no longer in place (i.e., delayed drought rebound). Table 1-3 details the predictors used to develop the demand models and identifies the expected sign and magnitude of the coefficient estimates resulting from the linear regression.

Table 1-3: Description of Demand Model Predictors

Predictor Variable	Log Transformed?	Expectations about Coefficient Estimates	Description
Departure from normal temperature[(a)]	Yes	Positive sign	Represents difference from long-term temperature. Higher than normal temperatures are associated with higher demands.
Departure from normal precipitation[(a)]	Yes	Negative sign	Represents difference from long-term precipitation. Higher than normal rainfall is associated with lower demands.
Seasonal index	No	Larger absolute magnitudes for agencies with greater seasonal peaking	Reflects the cyclical pattern in water use where demands a generally higher in the summer and lower in the winter. Represented in the model as a sine / cosine pair of variables.[(b)]
Price	Yes	Negative sign with absolute value between 0 and 1	Economic theory suggests negative correlation with demand.
Economic index	Yes	Positive sign	Several economic indices were explored as potential predictors[(c)] with the detrended Economic Cycles Research Institute (ECRI) selected as the index that produced the most reasonable coefficient estimates across model sectors. Water demand is positively correlated with economic fluctuations of the business cycle. The index is modeled in form of departures from long-term trend.
Housing density	Yes	Negative sign (commonly with absolute value between 0 and 1)	Housing density is negatively correlated with demand; on average, residences with more units per acre (or smaller parcel sizes) tend to use less water on outdoor uses.
Median income	Yes	Positive sign (commonly with absolute value between 0 and 1)	Economic theory suggests positive correlation of income with demand; generally geographical areas with higher median incomes tend to use more water.
Persons per household	Yes	Positive sign (commonly with absolute value between 0 and 1)	Positively correlated with demand; generally, residences with more people tend to use larger amounts of water.
Mix of Industries / economic activity[(d)]	Yes	N/A	The representation of industries / economic activity with a geographical area is related to the amount of water used within the CII sector. Fitted parameters for these variables are generally unique by utility, thus there is no generally accepted range of coefficient estimates.
Drought Severity	No	Negative sign	Reflects the effect of drought restrictions from the most recent drought (2014-2017, with extended restrictions though 2019) on water demand.[(e)] Defined as the presence of drought restrictions (represented as a binary) multiplied by the requested cutback (e.g. 0-30%).

(a) Lagged values of temperature and precipitation were also evaluated and included as model predictors as the influence of weather on water demand can persist several months.

(b) Most sectors have a single sine/cosine pair representing the seasonal cycle, except for Stanford. Stanford has two sine/cosine pairs to capture seasonal effects associated with the academic calendar. See Section 4.3 for additional discussion.

(c) Other economic indices explored as potential predictors are documented in TM 3.

(d) Detail on the derivation of specific predictors representing mix of industries / economic activity is documented in TM 3.

(e) A unique prediction variable was also evaluated for the 2008-2011 drought but was dropped during the model development process as the coefficient estimate was not statistically significant. The 2008-2011 drought overlapped with the severe economic downturn of the Great Recession which likely mutes its statistical significance.

2. Single Family Regression Development

This section reviews the development of the statistical regression for the single family residential sector.

2.1 Model Predictors and Fitted Coefficients

The fit for the final single family regression is presented in Table 2-1. Coefficient estimates are within the expected range for all explanatory variables.

Table 2-1: Single-Family Regression Predictors and Coefficients

Variable	Coefficient	Standard Error	t-Statistic	Probability
Intercept	3.821	0.324	11.776	<0.05
Seasonal index 1[a]	-0.283 (avg) -0.045 to -0.185	0.013 (avg) 0.008 to 0.026	-24.086 (avg) -7.379 to -24.086	<0.05
Seasonal index 2[a]	-0.262 (avg) -0.616 to -0.064	0.013 (avg) 0.008 to 0.026	-23.026 (avg) -44.960 to -3.786	<0.05
Departure from normal temperature	1.008	0.135	7.464	<0.05
Departure from normal temperature, 1-month lag	0.824	0.137	5.997	<0.05
Departure from normal temperature, 2-month lag	0.354	0.137	2.583	<0.05
Departure from normal temperature, 3-month lag	0.306	0.127	2.413	<0.05
Departure from normal precipitation	-0.008	0.003	-3.01	<0.05
Departure from normal precipitation, 1-month lag	-0.009	0.003	-3.649	<0.05
Departure from normal precipitation, 2-month lag	-0.004	0.003	-1.582	0.114
Price	-0.085	0.009	-9.942	<0.05
Economic index	0.945	0.101	9.316	<0.05
Housing density	-0.406	0.007	-60.745	<0.05
Median income	0.195	0.025	7.778	<0.05
Persons per household	0.473	0.04	11.907	<0.05
Drought severity, extended	-1.506	0.048	-31.109	<0.05
[a] Seasonal indices are unique to each retail agency.				

Variables with an increasing effect on water use (i.e., a positive coefficient) included temperature, economic index, median income, and persons per household. Variables with a decreasing effect on water use (i.e., a negative coefficient) included precipitation, price, housing density, and the extended drought effect.

2.2 Historical Model Performance

Figure 2-1 shows the observed and predicted per-unit use for the single family sector in gallons per unit per day (gpud) calculated as a unit-weighted average across all retail agencies. Performance of the single family regression is summarized in Table 2-2 which shows performance metrics for unit-weighted average County-wide demand. Visual inspection of the time series plot and review of the model fit parameters showed good performance at the County-wide level, including strong agreement with the observed seasonal cycle and ability to reproduce declining consumption during the Great Recession, recovery between the Great Recession and the recent drought, and the sharp decline and muted recovery following the most recent drought.

Historical performance of the single family regression was also strong at the retail agency-level. Model fit statistics calculated at the retail agency-level generally mirrored County-wide performance. Model fit statistics and time series plots for each retailer are presented in Appendix B.

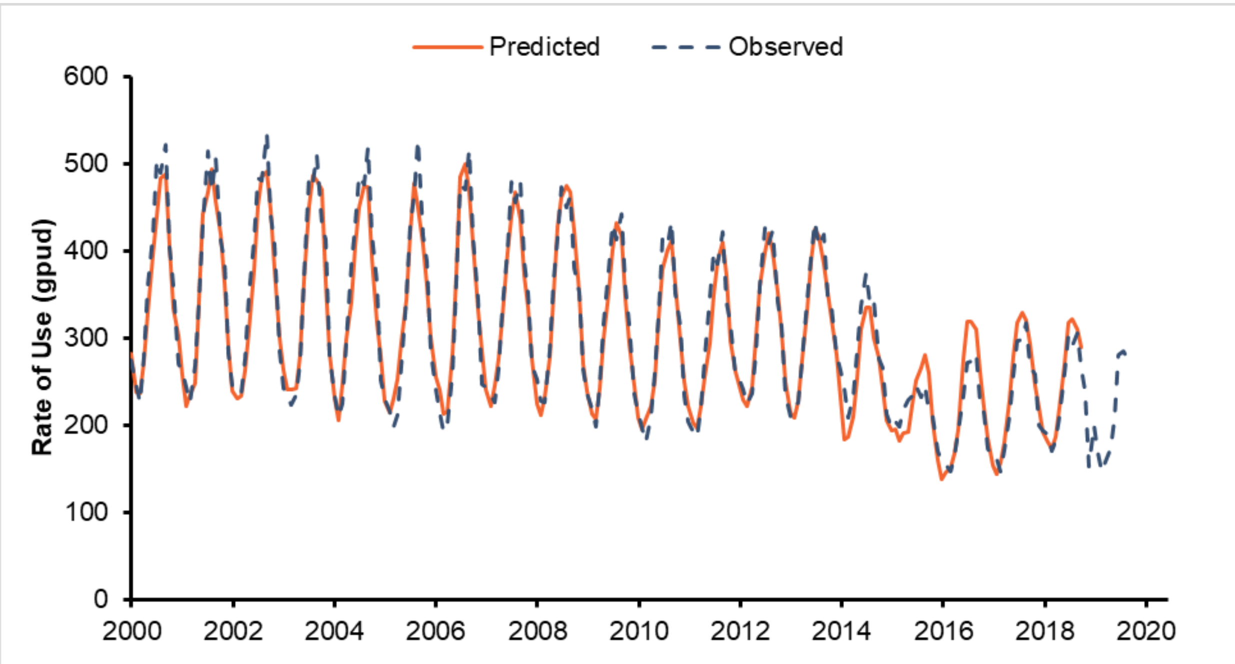


Figure 2-1: County-Wide Single-Family Observed and Predicted Per Unit Rate of Use

Table 2-2: County-Wide Single-Family Regression Performance Metrics

Regression Statistic[(a)]	Value
R-squared	0.95
Average Observed Value (gpud)	305.71
Mean Absolute Percentage Error	5.82%
Mean Bias	-1.13%

[(a)] Statistics calculated using County-wide unit-weighted average observations and predicted values from the regression fits.

3. Multifamily Regression Development

This section reviews the development of the statistical regression model for the multifamily residential sector.

3.1 Model Predictors and Fitted Coefficients

The fit for the final multifamily regression is presented in Table 3-1. Though most predictors are the same as the single family sector, several predictors (e.g., median income and 2-month lagged departure from precipitation) were dropped and certain predictors (e.g., the intercept term and drought severity) were allowed to vary by retail agency. These modifications to the model design resulted in stronger measures of fit and more reasonable coefficient estimates. Final coefficient estimates presented in Table 3-1 are within the expected range for all explanatory variables.

Table 3-1: Multifamily Regression Predictors and Coefficients

Variable	Coefficient	Standard Error	t-Statistic	Probability
Intercept	5.209	0.074	70.141	<0.05
Agency-specific intercepts[a]	-0.223 (avg) -0.719 to 0.280	0.013 (avg) 0.007 to 0.023	-31.555 (avg) -104.09 to 15.203	<0.05
Seasonal index 1[b]	-0.161 (avg) -0.372 to -0.056	0.012 (avg) 0.006 to 0.031	-16.311 (avg) -35.651 to -3.872	<0.05
Seasonal index 2[b]	-0.138 (avg) -0.255 to -0.056	0.012 (avg) 0.006 to	-13.943 (avg) -29.588 to -13.943	<0.05
Departure from normal temperature	0.488	0.098	4.974	<0.05
Departure from normal temperature, 1-month lag	0.514	0.100	5.155	<0.05
Departure from normal temperature, 2-month lag	0.397	0.094	4.226	<0.05
Departure from normal temperature, 3-month lag	0.194	0.092	2.101	<0.05
Departure from normal precipitation	-0.002	0.002	-1.127	0.260
Departure from normal precipitation, 1-month lag	-0.006	0.002	-2.954	<0.05
Price	-0.055	0.013	-4.347	<0.05
Economic index	1.568	0.091	17.226	<0.05
Housing density	-0.205	0.011	-18.105	<0.05
Persons per household	0.900	0.057	15.788	<0.05
Drought severity, extended[c]	-0.718	0.044	-16.294	<0.05

(a) Several agencies including San Jose Water Company, San Jose Municipal Water, Great Oaks Water Company, City of Gilroy, California Water Service, and the City of Sunnyvale were fitted with agency-specific intercept terms in order to optimize historical model performance.

(b) Seasonal indices are unique to each retail agency.

(c) Recorded drought severity coefficient estimate is for all agencies except San Jose Water Company, which was fitted an agency-specific drought severity coefficient.

Variables with an increasing effect on water use (i.e., a positive coefficient) included temperature, economic index, and persons per household. Variables with a decreasing effect on water use (i.e., a negative coefficient) included precipitation, price, housing density, and the extended drought effect.

3.2 Historical Model Performance

Figure 3-1 shows the observed and predicted per-unit use for the multifamily sector in gpud calculated as a unit-weighted average across all retail agencies.[5] Performance of the multifamily regression is summarized in Table 3-2 which shows performance metrics for unit-weighted average County-wide demand. Visual inspection of the time series plot and review of the model fit parameters showed good model performance at the County-wide level, including strong agreement with the observed seasonal cycle and ability to reproduce declining consumption during the Great Recession, recovery between the Great Recession and the recent drought, and the sharp decline and muted recovery following the most recent drought.

Historical performance of the multifamily regression was also strong at the retail agency-level. Model fit statistics calculated at the retail agency-level generally mirrored County-wide performance. Model fit statistics and time series plots for each retailer are presented in Appendix C.

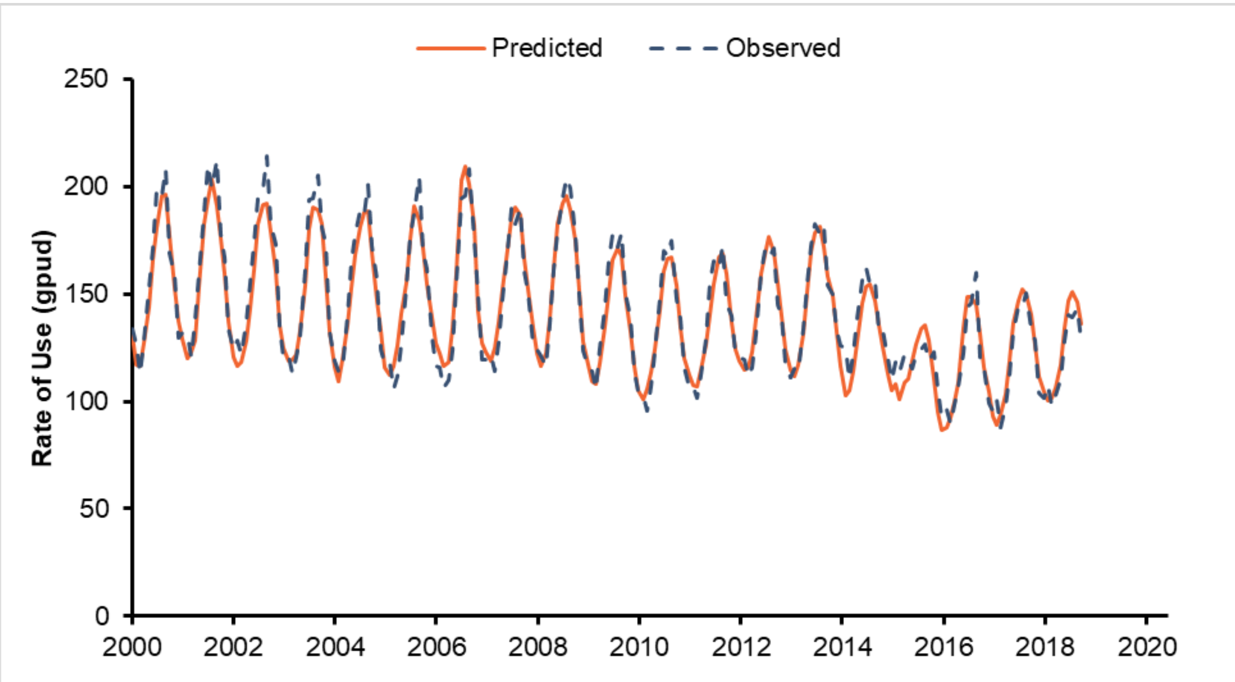


Figure 3-1: County-Wide Multifamily Observed and Predicted Per Unit Rate of Use

[5] Figure 3-1 excludes an outlier monthly observed datapoint for a single retail agency.

Table 3-2: County-Wide Multifamily Regression Performance Metrics

Regression Statistic[a]	Value
R-squared	0.94
Average Observed Value (gpud)	142.26
Mean Absolute Percentage Error	4.53%
Mean Bias	-0.87%

[a] Statistics calculated using County-wide unit-weighted average observations and predicted values from the regression fits.

4. CII Regression Development

This section reviews the development of the statistical regression for the CII sector. Distinct regressions representing the commercial, industrial, and institutional water use sectors[6] were initially considered. However, different billing classification schemes among retail agencies introduced definitional uncertainty in sectoral water use and driver units. For example, certain agencies lacked a distinct industrial billing classification while others combined commercial and institutional categories. Additional verification of water use at the account-level was not possible given the data constraints for this project.[7] In response to these constraints and uncertainties, total use within the commercial, industrial, and institutional sectors was consolidated into a single composite CII regression. The benefit of combining these sectors is a more parsimonious representation with respect to number of sectors, while providing a means to use the mix of industries to explain CII water use variability across retail agencies.

4.1 Model Predictors and Fitted Coefficients

Model predictors for the final CII regression equation along with their statistics are in Table 4-1. Note that understanding/quantifying the types of economic activity occurring within the County are important to understanding changes in CII consumption over time. Since individual regressions for the commercial, industrial, and institutional sectors were not developed, predictor variables representing the relative proportion of employment among different industry groupings was used in the CII regression. Proportional employment based on industry grouping is meant to reflect the relative mix of industries / economic activity within each retail agencies' service area. Most CII model predictors are similar to those used for the single family and multifamily sectors, however certain variables (e.g., 3-month lagged departure from normal temperature) were excluded during the regression refinement process. Final coefficient estimates presented in Table 4-1 are within the expected range for all explanatory variables.

[6] Refer to Appendix A for a summary of standardized sectors by retail agency.

[7] The finest spatial resolution of all consumption data was at the retail agency-level.

Table 4-1: CII Regression Predictors and Coefficients

Variable	Coefficient	Standard Error	t-Statistic	Probability
Intercept	-0.186	0.268	-0.695	0.49
Seasonal index 1[(a)]	-0.29 (avg) -0.41 to -0.17	0.02 (avg) 0.01 to 0.03	-20.79 (avg) -33.3 to -9.2	<0.05
Seasonal index 2[(a)]	-0.34 (avg) -0.53 to -0.10	0.02 (avg) 0.01 to 0.03	-23.34 (avg) -39.2 to -3.5	<0.05
Departure from normal temperature	1.037	0.158	6.580	<0.05
Departure from normal temperature, 1-month lag	0.912	0.161	5.657	<0.05
Departure from normal temperature, 2-month lag	0.370	0.158	2.340	<0.05
Departure from normal precipitation	-0.003	0.003	-0.997	0.32
Departure from normal precipitation, 1-month lag	-0.007	0.003	-2.312	<0.05
Departure from normal precipitation, 2-month lag	-0.002	0.003	-0.692	0.49
Price	-0.062	0.025	-2.453	<0.05
Economic index	0.963	0.140	6.881	<0.05
Proportion of total Employment (Retail)	0.142	0.032	4.430	<0.05
Proportion of total Employment (Professional Services)	0.499	0.031	16.065	<0.05
Proportion of total Employment (Information, Government, and Construction)	0.093	0.026	3.508	<0.05
Proportion of total Employment (Industrial)	0.351	0.026	13.249	<0.05
Proportion of total Employment (Health Education, and Recreational Services)	0.466	0.059	7.923	<0.05
Drought severity, extended	-1.424	0.070	-20.232	<0.05
[(a)] Coefficients vary by retailer.				

Variables with an increasing effect on water use (i.e., a positive coefficient) included temperature, economic index, and the mix of industries/economic activity ratios. Variables with a decreasing effect on water use (i.e., a negative coefficient) included precipitation, price, and the extended drought effect.

4.2 Historical Model Performance

Figure 4-1 shows the observed and predicted per-unit use for the CII sector in gallons per employee per day (gped) calculated as a unit-weighted average for across all retail agencies. Performance of the CII model is summarized in Table 4-2 which shows regression performance metrics for county wide demand. Visual inspection and performance metrics showed good model performance including the same seasonal cycle and quantities. The CII regression was also able to reproduce declining consumption during the Great Recession, recovery between the Great Recession and the recent drought, and the sharp decline and muted recovery following the most recent drought.

Historical performance of the CII regression was also strong at the retail agency-level. Model fit statistics calculated at the retail agency-level generally mirrored County-wide performance. Model fit statistics and time series plots for each retailer are presented in Appendix D.

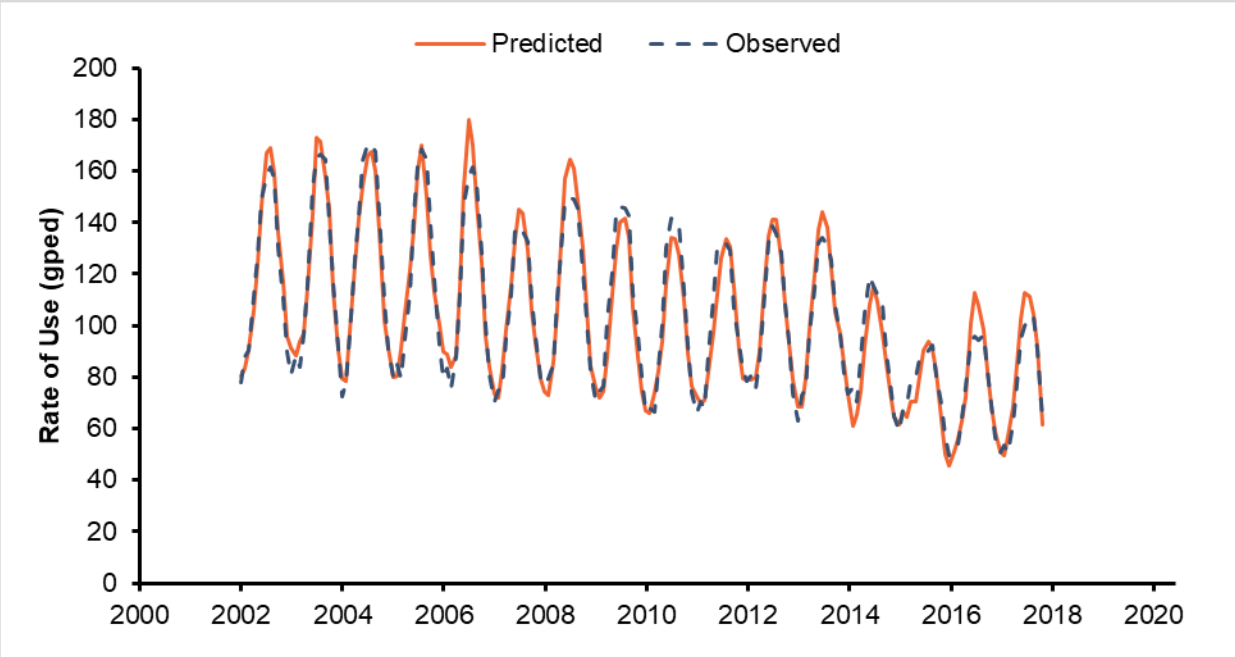


Figure 4-1: CII Observed and Predicted Rate of Use

Table 4-2: County-Wide CII Regression Performance Metrics

Regression Statistic[a]	Value
R-squared	0.96
Average Observed Value (gped)	103.89
Mean Absolute Percentage Error	5.08%
Mean Bias	-0.06%
[a] Statistics calculated using County-wide unit-weighted average observations and predicted values from the regression fits.	

4.3 Stanford University Regression Development

As an academic institution, Stanford University (Stanford) is considered part of the CII sector. However, an independent regression for Stanford was developed given its unique characteristics among retailers. Unlike other retail agencies, Stanford does not have accounts in the traditional sense as individual users are not billed. Additionally, employee water use as the sole driver unit (consistent with the CII sector for other retailers) is not appropriate for Stanford as students account for a significant portion of water use. This distinction informed the decision to use population (understood to be total faculty, staff, and students) as the driver unit for Stanford. Since the driver unit for the Stanford CII model was population, rather than jobs like the rest of the retailers' CII use, rate of use must be modeled separately. It is expected that the significant variables and/or magnitudes of coefficients would be different for Stanford than the other retailers' CII sectors due to the difference in driver units. A discussion of Stanford's regression predictors and fitted coefficients is presented in Appendix E. A summary of the Stanford's historical model performance is included in Appendix D.

5. Non-Retail Groundwater Pumper Regression Development

Historic water use for non-retail groundwater pumpers includes groundwater use by private well owners that are outside of retailers' service areas. Historic groundwater use was reported by groundwater basin and billing classification. The groundwater basins include Santa Clara Plain (referred to as charge zone "W2") as well as Coyote Valley sub-basin management area and the Llagas sub-basin and (referred to as charge zone "W5"). Water use was classified as either agricultural or municipal/industrial (M&I). M&I can include residential domestic water use.

Historical regression fits for non-retail groundwater pumpers were performed on annual water use. Agricultural water use was typically reported annually or semi-annually. M&I use was reported monthly or semi-annually. As a result, a monthly resolution for model fitting was not possible.

Further, historical model fits for non-retail groundwater pumpers were performed on a volumetric basis. Typical driver units for groundwater use, such as number of wells, did not support the "rate of use times driver" approach that was used for single family, multifamily, and CII model development.

Fitted models were only finalized for the M&I sector for the two groundwater basins. Agricultural use was often reported semi-annually (in January and July) and was estimated by a "table of averages" approach based on crop type, resulting in a lack of variability that could be modeled by predictor variables. Initial exploration of statistical/econometric model development showed that agricultural water use has been generally constant over the last twenty years and was not well-characterized by typical predictor variables.

5.1 Model Predictors and Fitted Coefficients

Model predictors for the non-retail groundwater pumpers M&I regression models along with their statistics are in Table 5-1. The two groundwater zones were modeled separately; a combined regression provided no improvement in the statistical significance of coefficients.

Table 5-1: Predictors for Non-Retail Groundwater Pumpers M&I Regression.

Basin	Variable	Coefficient	Std. Error	t-Statistic	Prob.
W2	Intercept	-0.59	4.08	-0.14	0.89
	Drought	-0.70	0.20	-3.54	<0.05
	Price	-0.81	0.06	-13.31	<0.05
	Temperature[a]	1.83	0.93	1.98	0.07
W5	Intercept	1.43	0.47	3.04	<0.05
	Number of Wells	0.19	0.04	5.56	<0.05
	Drought	-0.31	0.15	-2.09	0.06
	Price	-0.12	0.05	-2.41	<0.05
	Precipitation[a]	-0.09	0.02	-3.62	<0.05

[a] Temperature and precipitation for non-retail groundwater pumper models were in absolute terms, not departures from normal.

Variables with an increasing effect on water use (i.e., positive coefficient) included maximum temperature (used in the W2 model only) and number of wells (used in the W5 model only). Variables with a decreasing effect on water use (i.e., negative coefficient) included the extended drought effect, price, and precipitation (used in the W5 model only). Economic indices, density, and median income were not found to be statistically significant for the groundwater M&I regressions. Note that temperature was found to be statistically significant for the W2 charge zone but not for the W5 charge zone regression, while precipitation was found to be statistically significant for W5 but not W2.

5.2 Historical Model Performance

Performance of the groundwater M&I regressions is summarized in Table 5-2. Figure 5-1 and Figure 5-2 show the observed and predicted demand for the M&I sector for groundwater charge zone W2 and W5, respectively. The M&I W5 regression had a lower correlation coefficient than all other model fits described in this TM, likely due to the relatively constant annual average water use over the available period.

Table 5-2: Regression Performance Metrics for Groundwater M&I Models

Regression Performance Metric	M&I, W2	M&I, W5
R-squared	0.96	0.81
Average Observed Value (mgd)	7.81	7.68
Mean Absolute Percent Error	4.32%	3.54%
Mean Bias	-0.22%	-0.09%

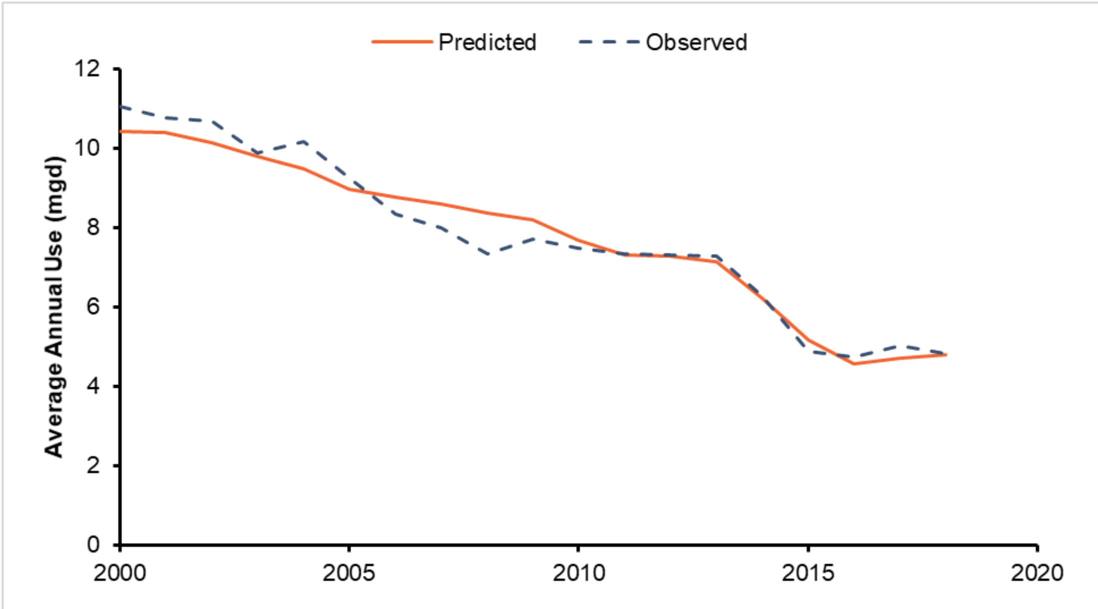


Figure 5-1: Observed and Predicted M&I Demand for Groundwater Basin W2

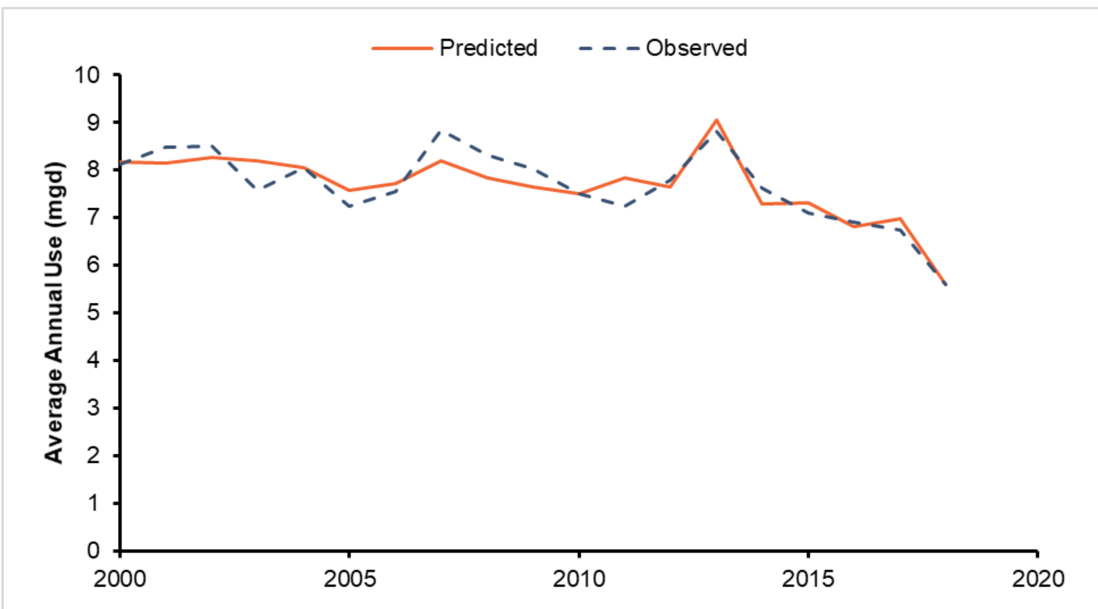
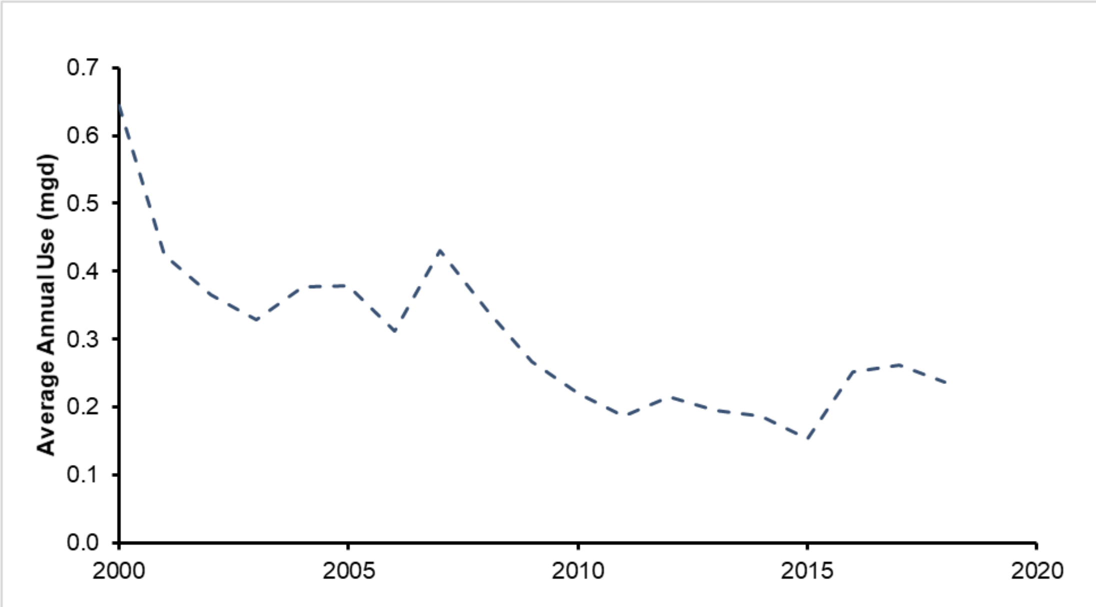


Figure 5-2: Observed and Predicted M&I Demand for Groundwater Basin W5

Figure 5-3 shows historic agricultural water use for the W2 and W5 charge zones. Agricultural water use in the W2 charge zone is less than 1 mgd and has been slightly declining over the last twenty years. Agricultural water use in the W5 charge zone has been generally constant over the last twenty years at approximately 23 mgd. Initial exploration of statistical/econometric model development showed that agricultural water use was not well-characterized by typical predictor variables. Agricultural water use in both charge zones would be well-represented by an average water use from a historical reference period that is then held constant into the future.



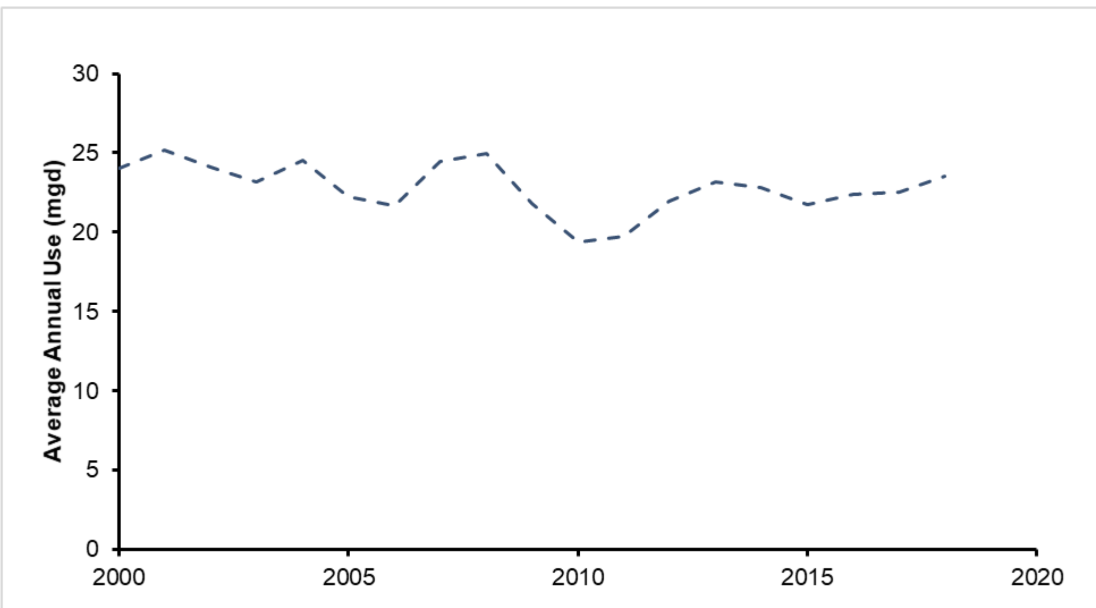


Figure 5-3: Observed Agricultural Demand for Groundwater Basin W2 (top) and W5 (bottom)

6. Summary / Conclusions

In summary, the statistical/econometric regressions presented in TM 2/4 show strong performance is explaining historical patterns of consumption over the last 20 years, including two major droughts and the Great Recession. All regressions had R-squared values of 0.81 or greater. The retailer-specific regressions, which represent the majority of water use in the County, had R-squared values of 0.94 or greater. None of the regressions demonstrated a large consistent bias. Based on this analysis, the regression reflect a suitable basis for forecasting.

The overall model approach allows for demand forecast scenario analysis based on varying assumptions of future conditions. Several forecast scenarios may be explored, including climate change-adjusted weather, alternate assumptions around the timing and magnitude of drought recovery, alternate assumptions around urban development, and/or different assumptions around future economic conditions. For any of these future scenarios, the model coefficients developed in this TM should be maintained as they reflect the best fitted estimates of causal relationships between external socioeconomic conditions and historical water demand given the available modeling data. Model scenarios can also be developed to address uncertainties in future predictor variables, such as housing / job growth and density. Future inputs in these scenarios could be conducted as a sensitivity analysis or be driven by alternate growth projections.

On a regular basis, overall model performance should be evaluated. Annually, forecasted consumption and input assumptions (e.g., driver unit counts, economic conditions, water rates, etc.) can be compared with observed conditions as data becomes available to monitor predictive performance. Less frequently (around every 5 years) model predictors should be revaluated using the process outlined in Figure 1-2. Major events, such as another drought or a severe economic recession may necessitate reexamination and/or refitting model coefficients and may cause changes in longer term expectations over the forecast period. As more data becomes available on the impacts of COVID-19 on County demographics and water use (e.g., potential shifts in CII to residential demand), reexamination of the underlying sectoral rates of water use as well as model coefficients should be conducted.