Enhancing Municipal Water System Planning and Operations Through Climate-Sensitive Demand Estimates

Ryan C. Johnson[1], Steven J. Burian [1,2], Carlos A. Oroza [3], James Halgren [1,4], Trevor Irons [3,5] Emily Baur [3] Danyal Aziz[2], Daniyal Hassan [3], Jiada Li[6], Tracie Kirkham[7], Jessie Stewart[7], Laura Briefer[7]

[1]Alabama Water Institute, University of Alabama, Tuscaloosa, Alabama, USA

[2]Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, Alabama, USA

[3]Civil and Environmental Engineering, University of Utah, Salt Lake City, Utah, USA

[4]Lynker Technologies, Leesburg, Virginia, USA

[5]Montana Technical University, Butte, Montana, USA

[6]Department of Civil and Environmental Engineering, College of Engineering, Colorado State University, Fort Collins, Colorado, USA

[7]Salt Lake City Department of Public Utilities, Salt Lake City, Utah, USA

Contents of this file

1. Description of CSD-WDM

1.1. Demand data

Per-capita demands are decomposed into indoor and outdoor water uses. This is done by calculating the average November through March demands for each year to determine the mean per-capita indoor water demand (D_I). Monthly outdoor water use (D_{Om}) is determined by subtracting the year's (y) respective mean indoor use ($\overline{D}_{Iy}$) from each irrigation month's (m) total per-capita demand (D_{Tm}), see equation S1.

$$D_{Om} = D_{Tm} - \overline{D}_{Iy} \tag{1}$$

1.2. Model Inputs

Precipitation and temperature data leverage the North American Land Data Assimilation System's (NLDAS) one-hour temporal and one-eighth degree spatial resolution climate forcing estimates (Xia et al., 2012). While local weather stations are within the municipal service area, a single national source broadens this framework's generalizability to other western water systems. Monthly mean temperature (^{o}C) and cumulative precipitation (mm) metrics are calculated from March through October, one month prior and extending till the completion of the irrigation season. Monthly hydro-climate metrics are not included as November to March model inputs, assuming demands remain independent of weather as there is no irrigation.

Mountain streamflow functions as a hydro-climate supply metric, capturing the complex interactions among mountainous topography (snowdrift, aspect, and microclimates), variable winter precipitation patterns (global climate oscillations and Great Salt Lake influences), snowmelt (timing, duration, and quantity), and unique mountain hydrology (groundwater and baseflow) that contribute to supply availability which precipitation and air temperature alone do not (Bales et al., 2006, Ahl, Woods, & Zuuring, 2008). The four supply streams' daily discharge measurements are collected (1980-2017) from the United States Geological Survey (USGS) and Salt Lake County at the canyon mouths prior to extensive water diversion. The few missing values are spatially interpolated as a function of up or downstream measurements (Hughes & Smakhtin, 1996). From this data mean monthly streamflow metrics ($\overline{Q_{cfs}}$) are calculated for the supply creeks.

Supply watershed snowfall completes the hydro-climate metrics, where November to April monthly and seasonal snowfall (S_{mm}) is retrieved from the Alta Guard station at the headwaters of Little Cottonwood Creek. This metric bridges the gap between climate conditions and surface water supply, and also uses the 76 years of continuous observations to define the frequency and magnitude of the climate scenarios.

To account for the evolving urbanization dynamics during the study period, population and housing data is acquired from the U.S. Census (Census, 2012). Linear interpolation between decadal census observations provides continuous population (p) and housing (H) data, and combined with SLCDPU's service area, population and housing densities are determined (p/km^2, H/km^2).

Long-term conservation trends provide the final input metrics in the analysis. In the year 2000, the Utah Division of Water Resources established statewide per-capita water-use goals for public community water systems to be at least 25% by 2050, and then in 2014 amended the target to 2025 after substantial progress had been achieved (Utah Department of Natural Resources, 2014). To recognize these policy influences on demand, the variability in each month's 1980-2000 mean per-capita water use forms the baseline to project Utah's 25% reduction in water use by 2025. This conservation metric (C_m) applies each month's unique conservation rate to each month spanning 2000 to 2025. Equation 2 details this function where m refers to the month of interest, $\overline{D}_m$ is that month's mean per-capita water use, and y is the number of years past the goals' implementation.

$$C_m = \overline{D}_m - \frac{\overline{D}_m * 25\%}{25yrs} * y \tag{2}$$

Adopting a linear conservation goal is representative of western U.S. water conservation policy (Utah Department of Natural Resources, 2019, Hertzbern, 2018, Friedman, 2018, Colorado Water Conservation Board, 2015, Southern Nevada Water Authority, 2019) and presents a predictor of anticipated long-term trends that could improve seasonal forecasting accuracy. Municipalities that do not use a constant rate typically use a constant percent rate reduction, which would form this metric (U.S. EPA, 1998). *Utah's 2019 Regional Municipal & Industrial Water Conservation Goals* supports SLCDPU's long-term per-capita reduction, albeit with significant year-to-year variability. While a linear conservation goal exhibits characteristics of stationarity, at an annual resolution, demands demonstrate non-stationarity by exiting their historical range of observations. No additional policy changes occurred during the study period, however, this metric can be updated if non-linear conservation is anticipated in the future.

The initial indoor demand features include conservation goals, population, and urbanization metrics. Climate and supply features are omitted as predictors because indoor use has been shown to be influenced by the number of people per household, appliance type, and associated water-use efficiencies rather than climate conditions (Jacobs & Haarhoff, 2007). These residential water use drivers are not included as potential inputs into CSD-WDM due to this study's spatial scale and multi-sector composition of municipal-produced water demands. Irrigation season months include these metrics, plus current and antecedent monthly mean air temperature, precipitation, supply streamflows, and snowfall metrics. Antecedent metrics begin in March, prior to the irrigation season start, and incrementally increase per month until the season's termination in October. For example, October's inputs include March through October hydro-climate and supply metrics.

1.3. Variable Selection

The CSD-WDM is a Python-based (v3.8.5) demand forecasting model that automatically selects key demand drivers and optimizes hyperparameters to accurately forecast mean monthly per-capita water use. The model uses each month's complete set of possible demand influencing metrics as inputs into a three-phase feature selection process: 1) feature correlation to demand, 2) collinearity removal, and 3) driver selection. The term driver is designed to inform that a feature/metric is selected to be a key demand predictor. Phase one evaluates each feature's Pearson correlation coefficient with each month's demands. This value doubles as a threshold parameter, permitting features meeting or exceeding the threshold to pass to the next phase. Phase two evaluates for collinearity between features and is also an adjustable parameter. This process eliminates the lesser demand-correlated feature, should collinearity between two features be greater than the threshold, resulting in demand-correlated features with acceptable levels of collinearity (< 10) to the final phase of variable selection (Song & Kroll, 2011). Phase three uses Scikit-learn 0.24.1 recursive feature elimination (RFE) to select the optimal monthly outdoor water demand drivers (Pedregosa et al., 2011). RFE is an efficient, effective, and model-specific feature selection algorithm that identifies drivers that are most relevant to predicting monthly demands. Using the GridSearchCV function, the algorithm assigns feature importance weights and recursively prunes the number of features over five-fold cross-validation until the optimal drivers are selected. Figure SS1 illustrates CSD-WDM's automated workflow using July as an example.

Scikit-learn OLS regression serves as the CSD-WDM regression algorithm due to its driver-target interaction interpretability. The algorithm is calibrated on each month's demand drivers, undergoes a final five-fold cross-validation, and is fitted without a y-intercept. During calibration, an exhaustive grid search function evaluates root-mean-squared-error (*RMSE)* over the correlation (0-0.7 in 0.05 increments) and collinearity (0.65-0.90 in 0.05 increments) parameters. This process delivers each month's optimal demand drivers, coefficient weights, and modeling error. CSD-WDM's regression base aids in interpretability, see Equation 3 where m is the month of interest, β is the coefficient weight, and x is the selected driver.

$$lpcd_m = \beta_1 x_1 + \beta_2 x_2 \ + \ \ldots \ + \ \beta_n x_n \tag{3}$$

The CSD-WDM model calibration results are presented in several ways to improve user interpretability. This includes driver coefficients, per-capita, and acre-feet prediction units, data-frames for additional analysis, and figures to illustrate performance at monthly to annual temporal resolutions. The variety of results allows the user to select the method most appropriate for their respective needs and decision-making.

The CSD-WDM identified Utah's conservation goal metric as the optimal indoor demand driver. No urbanization metrics (population, housing, density, etc) were identified as statistically significant or improving indoor predictive performance. During the irrigation season, a total of eighteen hydro-climate and supply metrics are identified as demand drivers, emphasizing antecedent monthly air temperatures from June through October. Table SS1 displays April to October predictors and their respective coefficient weights. Preliminary feature development included the municipal fraction of irrigated, impervious, developed, residential, and urban land uses.

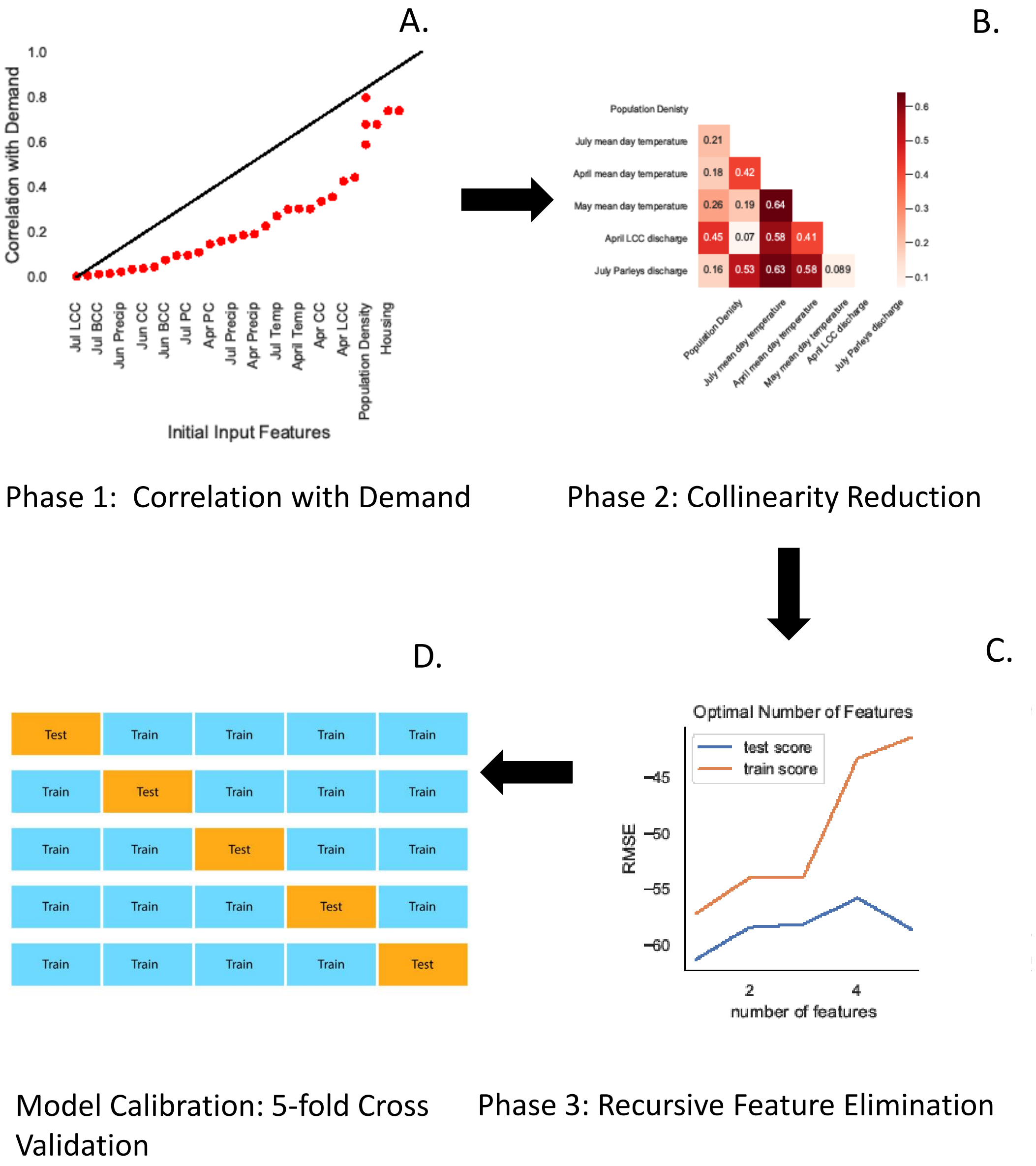

Figure S1. CSD-WDM's feature selection process

1.4. CSD-WDM Calibration

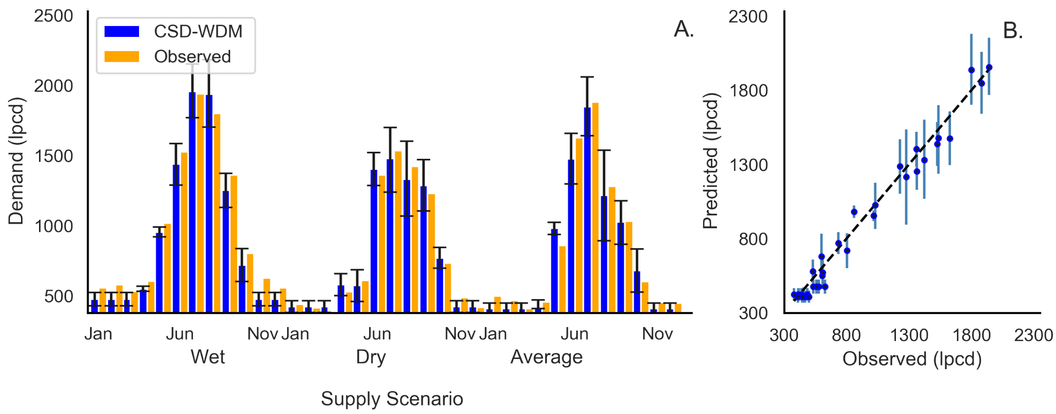

Figure S2. The CSD-WDM captures water use dynamics in response to drought, average, and surplus supply scenarios.

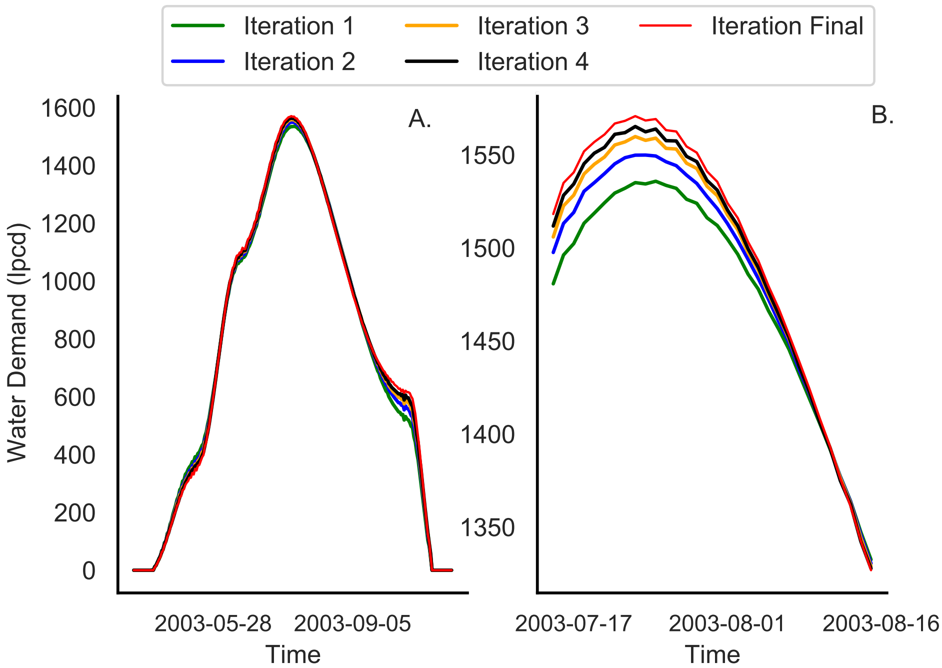

Figure S3. An iterative process down scales monthly demand values to a daily time step to provide a continuous demand time series (A). A key aspect of the iterative process is maintaining the mean monthly demands, performed by increasing or decreasing the spline value for each month (B).

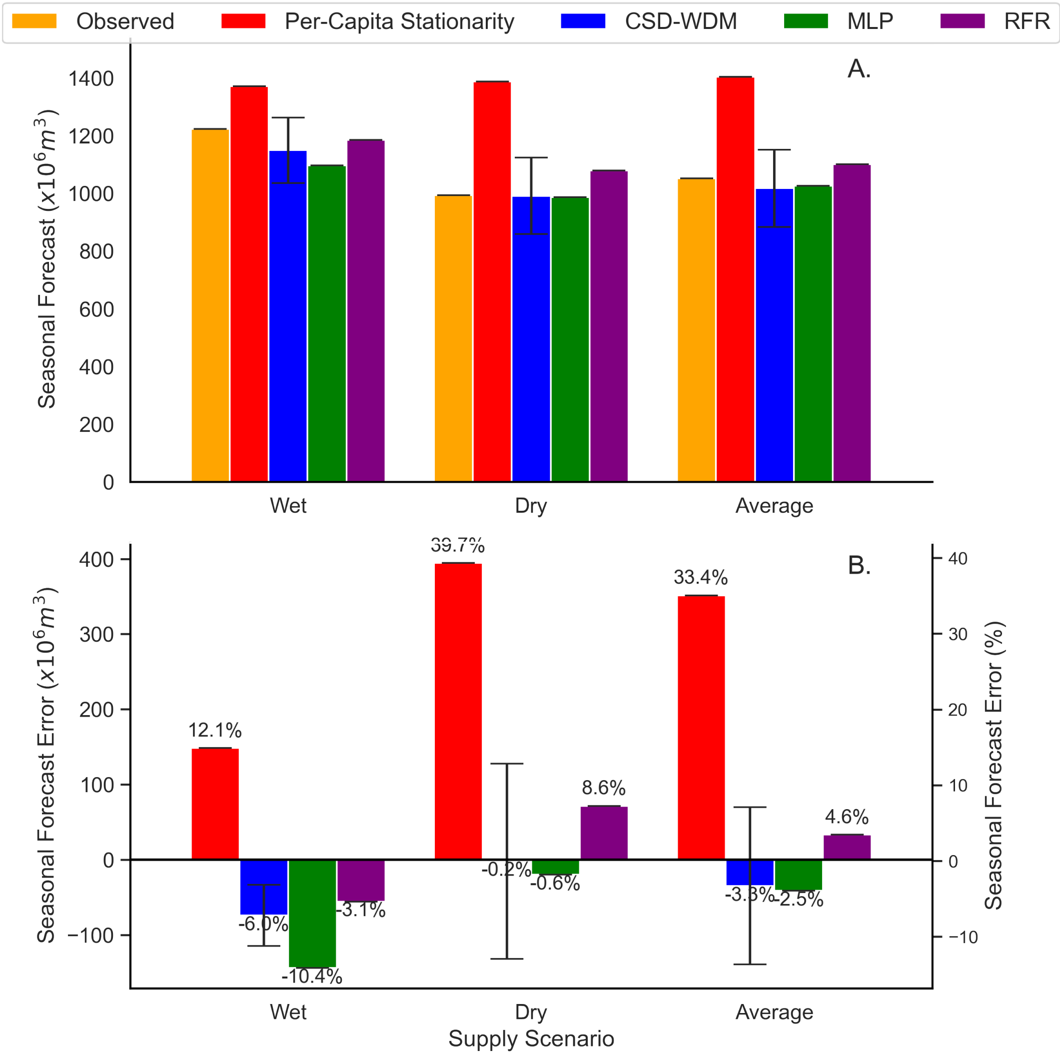

Figure S4. Seasonal demand forecasting methods relying on stationarity can significantly over-predict water demands (A), leading to high forecasting error (B).

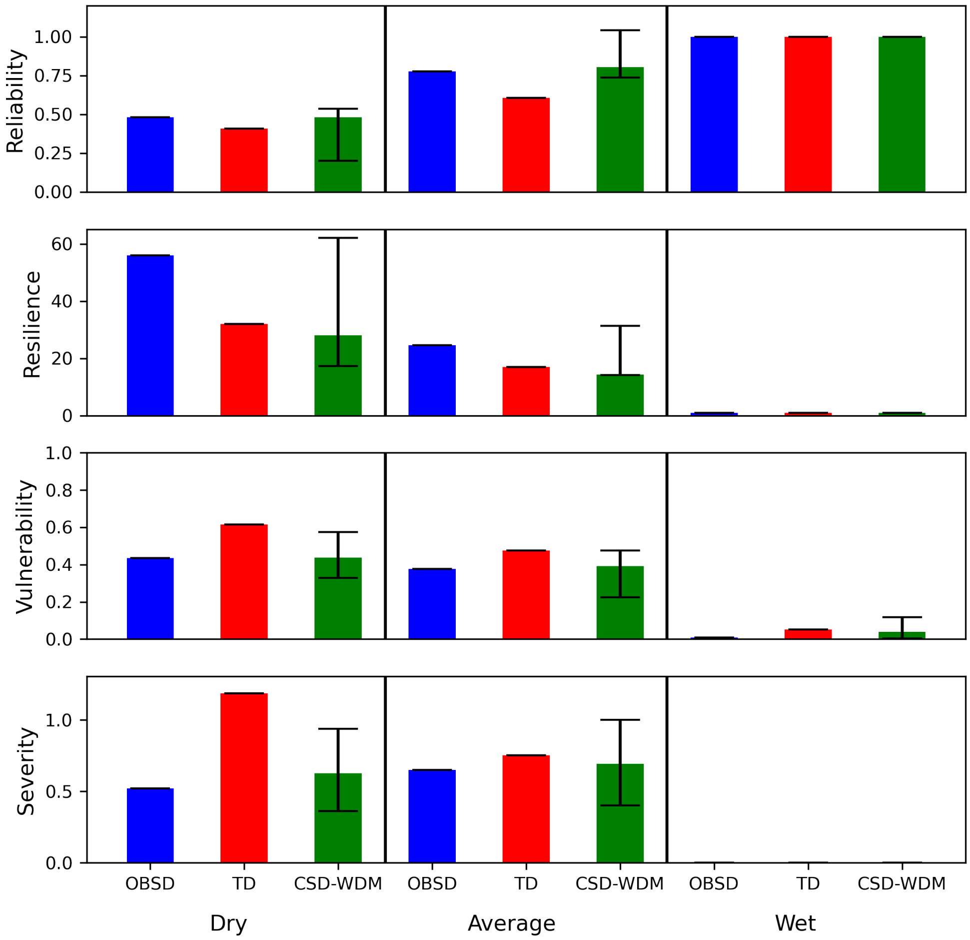

Figure S5. Daily RRV for observed (OBSD), stationary traditional (TD), and non-stationary dynamic (CSD-WDM) water demand simulations. The non-stationary dynamic demand simulations mirror the observed results while stationary traditional methods indicate reduced reliability and greater vulnerability relative to the observed. The error bars in the CSD-WDM predictions communicate the forecast's uncertainty to a 95% confidence interval, a missing component of the traditional demand forecast.

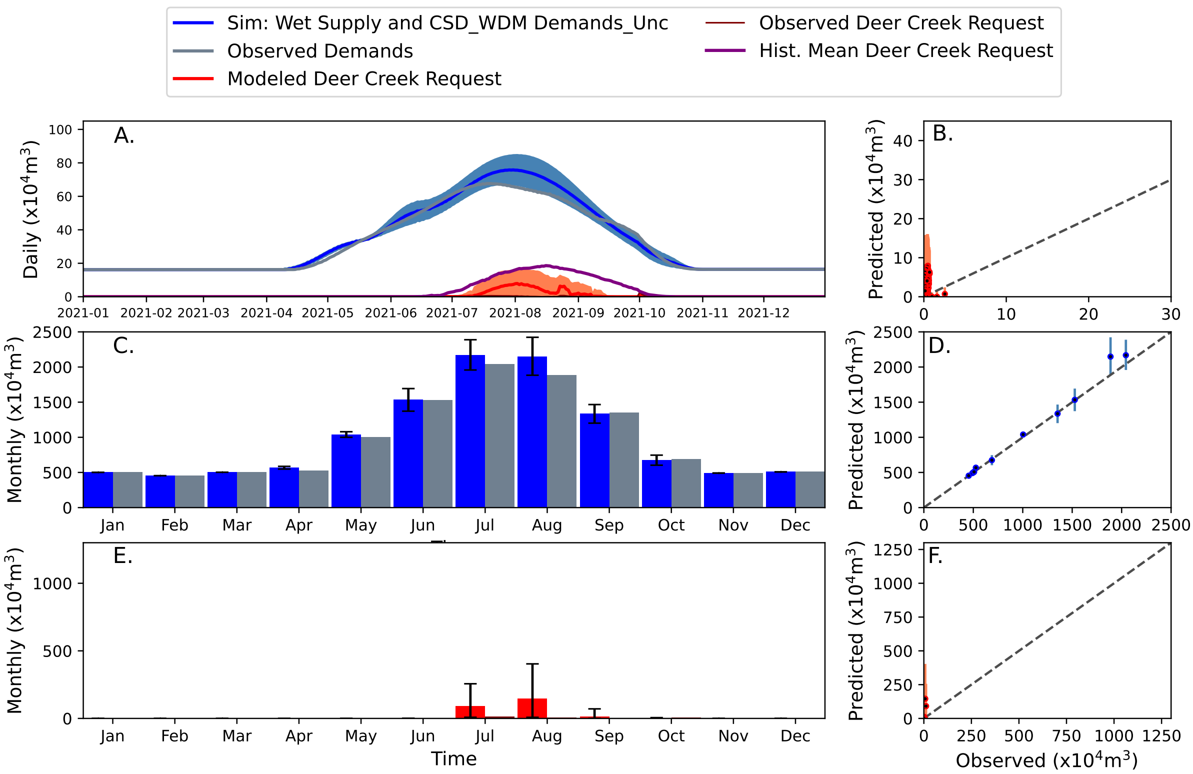

Figure S6. Water System Performance during wet supply conditions and CSD-WDM forecasted demands

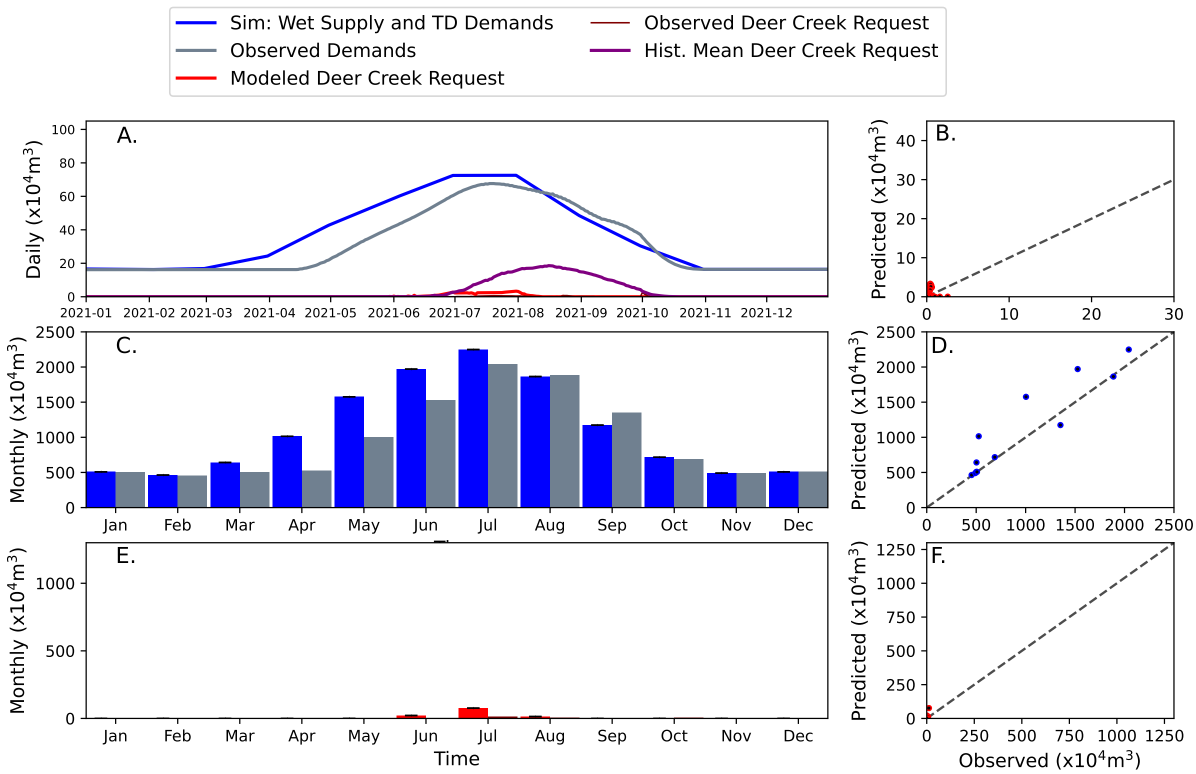

Figure S7. Water System Performance during wet supply conditions and traditional forecasted demands

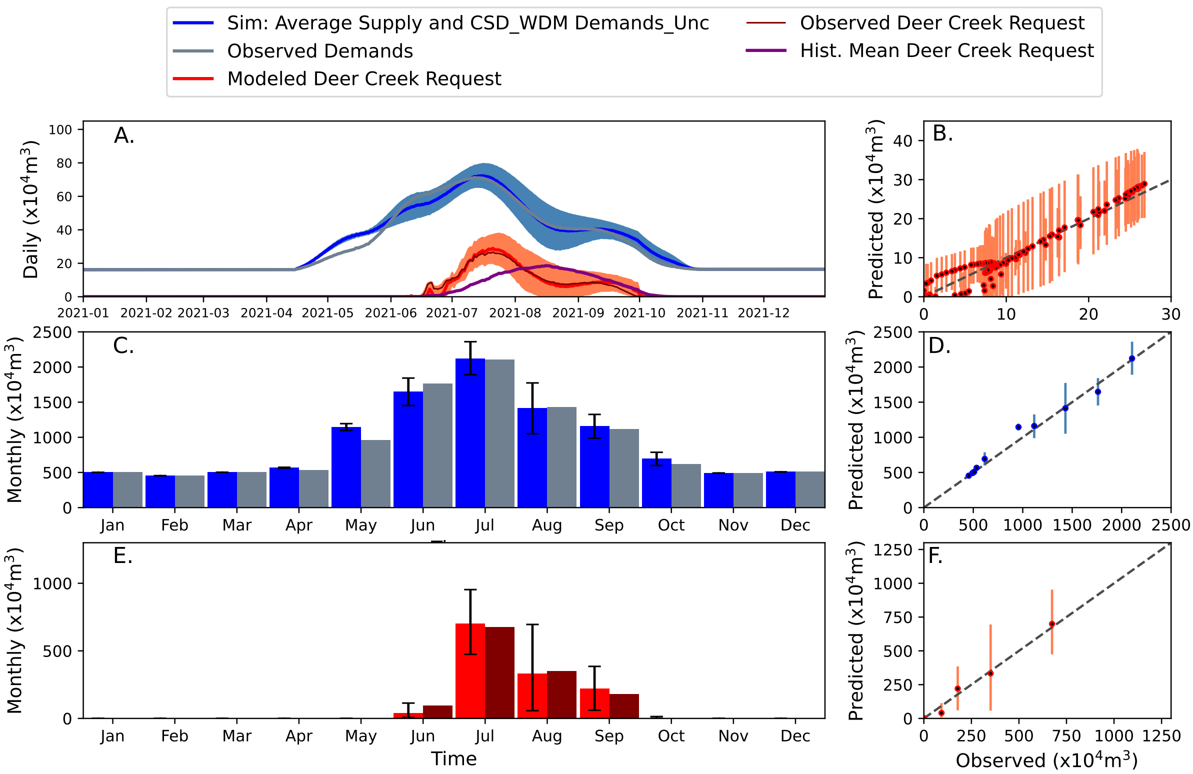

Figure S8. Water System Performance during average supply conditions and CSD-WDM forecasted demands

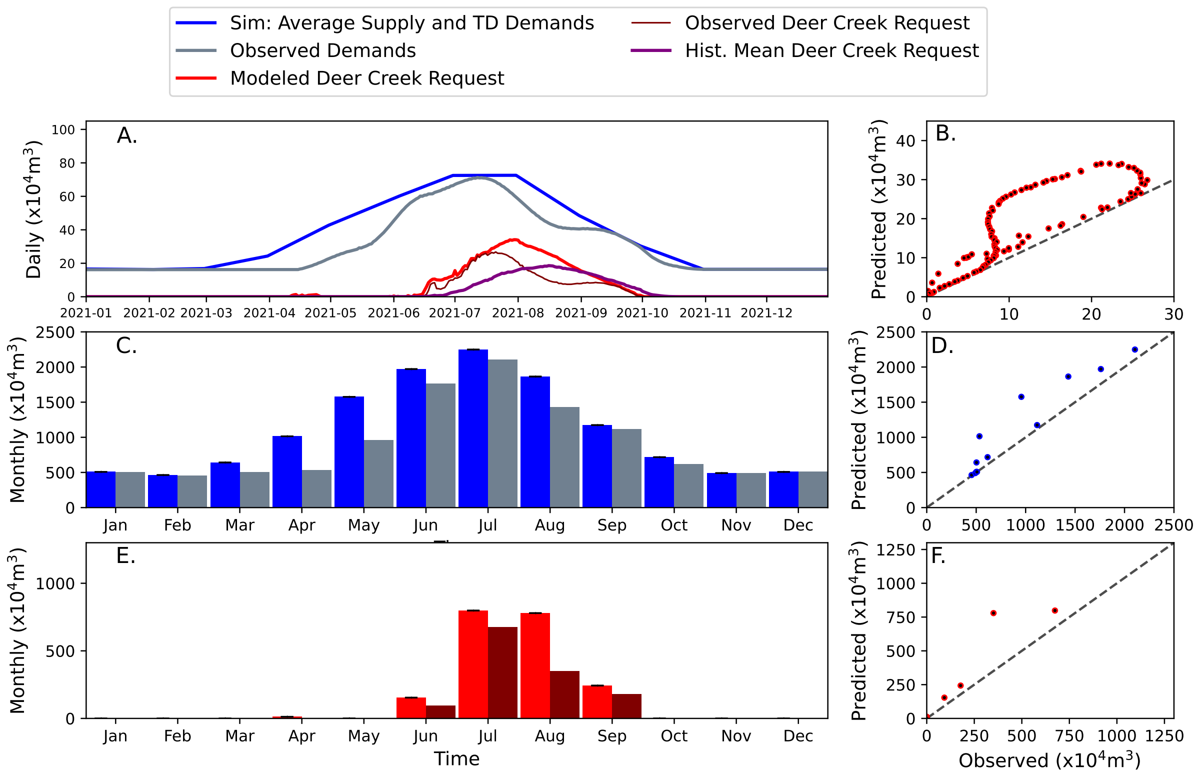

Figure S9. Water System Performance during average supply conditions and traditional forecasted demands

Table S1. Water demand drivers and CSD-WDM coefficient weights.

Predictor	Apr	May	Jun	Jul	Aug	Sep	Oct
Population Density[1]				-0.30	-.11		
Mar LCC Streamflow[2]	1				57.0		
Apr LCC Streamflow[2]	1				1		
May LCC Streamflow[2]					1		
May BCC Streamflow[2]					43.9		
Season Snowfall[3]			0.45				
Apr Mean Temperature[4]	21.3		33.9	-14.8	11.4	16.0	
Apr Precipitation[5]	-1.17				-0.04		
May Mean Temperature[4]		54.1	56.4	36.1	-1.10	12.5	-2.02
May Precipitation[5]		-3.97					
Jun Mean Temperature[4]					5.45	-14.4	
Jun Precipitation[5]			-14.4		4.62		
Jul Mean Temperature[4]				118	43.5	-27.6	
Aug Mean Temperature[4]					23.9	45.2	
Aug Precipitation[5]					-6.90		
Sep Mean Temperature[4]						30.7	
Sep Precipitation[5]						-3.63	
Oct Mean Temperature[5]							21.7

[1] change in demand per $persons/km^2$

[2] change in demand per cms of streamflow (x10^{-3})

[3] change in demand per mm of snow

[4] change in demand per °C

[5] change in demand per mm of liquid precipitation

Table S2. Modeled daily RRV. The stationary traditional demand forecasting methods exhibit greater error (value in parenthesis) from the observed compared to the non-stationary dynamic demand (CSD-WDM) simulations. Furthermore, the stationary demand simulations are deterministic and do not communicate prediction uncertainties. In comparison, the novel approach leveraging the CSD-WDM's internal demand prediction error characterizes the range of uncertainty (Uncertainty Low/High) to a 95% confidence interval, providing a foundation to enhance operational decision making.

Metric	Climate Scenario (snowpack)	Observed Demands	Stationary Demands	Non-Stationary Demands	Non-Stationary Uncertainty (Lo/Hi)
Reliability	Dry	0.48	0.41 (-15%)	0.48 (2%)	0.43-0.76
	Average	0.78	0.61 (-22%)	0.80 (1%)	0.57-0.87
	Wet	1.0	1.0 (0%)	1.0 (0%)	1.0
Resilience*	Dry	56	32 (42%)	28 (50%)	17-62
	Average	25	17(32%)	14 (44%)	14-31
	Wet	1	1 (0%)	1 (0%)	1
Vulnerability	Dry	0.44	0.61 (39%)	0.44 (0%)	0.33-0.57
	Average	0.38	0.48 (26%)	0.39 (3%)	0.22-0.48
	Wet	0.01	0.05 (400%)	0.04 (300%)	0-0.12
Peak Severity	Dry	0.52	1.19 (129%)	0.63 (21%)	0.36-0.94
	Average	0.65	0.75 (15%)	0.69 (6%)	0.40-1.0
	Wet	0.0	0.0 (0%)	0.0 (0%)	0.0
Vulnerability Class	Dry	High	Extreme	High	High - Very High
	Average	High	Very High	High	Medium-High
	Wet	Low	Low	Low	Low-Medium
Peak Severity Class	Dry	High	Extreme	High	Medium-High
	Average	High	Very High	High	Medium - Very High
	Wet	Low	Low	Low	Low

*units in days

Table S3. Daily water system RRV percentage (%) difference from each the historical mean.

Metric	Climate Scenario (snowpack)	Observed Demands	Traditional Demands	CSD-WDM Demands
Reliability	Below Average	-25	-35	-23
	Average	23	-3	28
	Above Average	59	59	59
Resilience	Below Average	-1750	-960	-830
	Average	-710	-460	-370
	Above Average	67	67	67
Average Vulnerability	Below Average	54	117	54
	Average	32	68	38
	Above Average	-97	-82	-87
Peak Severity	Below Average	79	307	116
	Average	123	158	138
	Above Average	-100	-100	-100

Table S4. Monthly water system RRV percentage (%) difference from the historical mean.

Metric	Climate Scenario (snowpack)	Observed Demands	Traditional Demands	CSD-WDM Demands
Reliability	Below Average	-55	-100	-55
	Average	-9	-32	-9
	Above Average	59	36	59
Resilience	Below Average	-98	-164	-98
	Average	34	-45	-34
	Above Average	67	67	67
Average Vulnerability	Below Average	71	140	69
	Average	20	83	24
	Above Average	-87	-52	-79
Peak Severity	Below Average	88	338	125
	Average	100	165	113
	Above Average	-100	-97	-100

Table S5. The range of daily water system RRV differences by climate conditions.

Varying Climate	Observed Demands	Traditional Demands	CSD-WDM Demands	Range
Reliability	0.52	0.44	0.52	0.52
Resilience*	55	31	27	55
Average Vulnerability	0.43	0.60	0.40	0.60
Peak Severity	0.65	1.19	0.69	1.19
Supply Range**	157%	157%	157%	157%
Varying Demand	**Below-Average**	**Average**	**Above-Average**	**Range**
Reliability	0.07	0.12	0.0	0.12
Resilience*	28	11	0	28
Average Vulnerability	0.17	0.1	0.04	0.17
Peak Severity	0.67	0.1	0.0	0.67
Demand Range**	28%	28%	28%	28%

*units in days.

**function of the seasonal percent of historically observed.

Table S6. The range of monthly water system RRV differences by climate conditions and demand forecast.

Varying Climate	Observed Demands	Traditional Demands	CSD-WDM Demands	Range
Reliability	0.71	0.86	0.71	0.86
Resilience*	5	7	5	7
Average Vulnerability	0.45	0.54	0.42	0.54
Peak Severity	0.58	1.27	0.66	1.27
Supply Range**	157%	157%	157%	157%
Varying Demand	**Below-Average**	**Average**	**Above-Average**	**Range**
Reliability	0.29	0.14	0.14	0.29
Resilience*	2	2	0	2
Average Vulnerability	0.19	0.18	0.1	0.19
Peak Severity	0.73	0.19	0.01	0.73
Demand Range**	28%	28%	28	28%

*units in months.

**function of the seasonal percent of historically observed.

Table S7. Seasonal SLCDPU supply and demand as a ratio of average historical values. For this UWS, there is greater variability in supply (158%) than demand (28%).

Hydroclimate Scenario	Observed Demands	Stationary (Traditional) Demands	Non-Stationary (CSD-WDM) Demands	Range in Demand Per Hydroclimate Scenario	Streamflow
Dry	1.03	1.31	1.03	0.28	0.53
Average	1.06	1.31	1.09	0.25	0.62
Wet	1.12	1.31	1.18	0.19	2.11
Range by Climate.	0.09	0	0.15		1.58

References

Ahl, R., Woods, S., & Zuuring, H. (2008, 10). Hydrologic calibration and validation of swat in a snow-dominated rocky mountain watershed, montana, u.s.a.1. *JAWRA Journal of the American Water Resources Association*, *44*, 1411 - 1430. doi: 10.1111/j.1752-1688.2008.00233.x

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western united states. *Water Resources Research*, *42*(8). doi: 10.1029/2005WR004387

Census. (2012). *Population and housing unit counts* (Tech. Rep.). U.S. Department of Commerce. Retrieved from `https://www.census.gov/prod/cen2010/cph-2-46.pdf`

Colorado Water Conservation Board. (2015). *Colorado's water plan* (Tech. Rep.). Denver, Colorado: Colorado Water Conservation Board.

Friedman. (2018). *Senate bill no. 1668, water management planning.* Retrieved from `$https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB1668$`

Hertzbern. (2018). *Senate bill no. 606.* Retrieved from `$https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB606$`

Hughes, D. A., & Smakhtin, V. (1996). Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. *Hydrological Sciences Journal*, *41*(6), 851-871. Retrieved from `https://doi.org/10.1080/02626669609491555` doi: 10.1080/02626669609491555

Jacobs, H., & Haarhoff, J. (2007, 12). Prioritisation of parameters influencing residential water use and wastewater flow. *Journal of Water Supply Research and Technology-aqua - J WATER SUPPLY RES TECHNOL-AQ*, *56*. doi: 10.2166/aqua.2007.068

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others

(2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(10), 2825-2830.

Song, P., & Kroll, C. (2011, 05). The impact of multicollinearity on small sample hydrologic regional regression. *Water Resources Research*, *49*, 3713-3722. doi: 10.1061/41173(414)389

Southern Nevada Water Authority. (2019). *Joint water conservation plan* (Tech. Rep.). Las Vegas, Nevada: Southern Nevada Water Authority.

U.S. EPA. (1998). *Water conservation plan guidelines* (Tech. Rep.). Washington, D.C.: U.S. Environmental Protection Agency.

Utah Department of Natural Resources. (2014). *Utah's municipal and industrial water conservation plan: Investing in the future.* Utah Division of Water Resources. Retrieved from $https://water.utah.gov/wp-content/uploads/2019/01/MIConservation_Revision_2012.pdf$

Utah Department of Natural Resources. (2019). *Utah's regional m & i water conservation goals* (Tech. Rep.). Utah Department of Natural Resources. Retrieved from https://water.utah.gov/wp-content/uploads/2019/12/Regional-Water-Conservation-Goals-Report-Final.pdf

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., ... Lohmann, D. (2012). Continental-scale water and energy flux analysis and validation for north american land data assimilation system project phase 2 (nldas-2): 2. validation of model-simulated streamflow. *Journal of Geophysical Research: Atmospheres*, *117*(D3). doi: 10.1029/2011JD016051