



Estimating Facility-Level Monthly Water Consumption of Commercial, Industrial, Municipal, and Thermoelectric Users in Virginia

Morgan McCarthy, Connor Brogan , Julie Shortridge , Robert Burgholzer , Joseph Kleiner , and Durelle Scott 

Research Impact Statement: Monthly data on facility-scale water consumption suggest that industrial and commercial consumption may be twice as high as previously estimated with significant seasonal variation.

ABSTRACT: Understanding water consumption is an important component of water management. However, water consumption data are limited and consumption coefficients do not account for variability through time and across users. This study combines federally maintained discharge data with state-maintained withdrawal data at monthly time steps to estimate facility-level and spatially aggregated water consumption in Virginia between 2010 and 2016. We evaluate (1) the feasibility of using discharge and withdrawal datasets to estimate sub-annual water consumption, (2) how these consumption estimates vary depending on the level of spatial aggregation, and (3) what patterns of seasonality exist in consumption estimates. We find that a combined process of text matching and geospatial analysis is effective in matching facilities and yielding monthly time-series of water consumption. Our results suggest that median consumption in industrial (17%) and commercial (19%) facilities may be higher than median consumption coefficients in the literature (10%). Consumption estimates also demonstrated more variability across facilities and seasons than aggregate coefficients in the literature suggest. Combining this approach with institutional knowledge can assist in quantifying issues such as inter-basin transfers and infiltration that impact consumption estimates, ultimately allowing for more accurate accounts of water use and availability.

(KEYWORDS: water use; consumption; data; withdrawal; discharge; NPDES.)

INTRODUCTION AND BACKGROUND

Knowledge of available water supply and how it is used is crucial for resource management, particularly under the pressures of climate change and population growth. An important key to comprehending the full water budget is estimating water consumption, generally defined as water that is removed from surface and groundwater resources and not returned due to

evaporation, transpiration, conveyance, incorporation into products, and consumption by animals and humans (Shaffer and Runkle 2007; Goldstein 2014). This is particularly true in humid regions where irrigation is not widespread; in these contexts, the largest withdrawals are generally associated with thermoelectric, industrial, and municipal uses (Averyt et al. 2013). This can suggest localized pockets of water stress based on withdrawals alone (Eldardiry et al. 2016), but the locations most at risk

Paper No. JAWR-21-0121-P of the *Journal of the American Water Resources Association* (JAWR). Received July 20, 2021; accepted June 1, 2022. © 2022 The Authors. Journal of the American Water Resources Association published by Wiley Periodicals LLC on behalf of American Water Resources Association. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. **Discussions are open until six months from issue publication.**

Department of Biological Systems Engineering (McCarthy, Brogan, Shortridge, Scott), Virginia Tech Blacksburg, Virginia, USA; and Office of Water Supply (Burgholzer, Kleiner), Virginia Department of Environmental Quality Richmond, Virginia, USA (Correspondence to Shortridge: jshortridge@vt.edu).

Citation: McCarthy, M., C. Brogan, J. Shortridge, R. Burgholzer, J. Kleiner, and D. Scott. 2022. "Estimating Facility-Level Monthly Water Consumption of Commercial, Industrial, Municipal, and Thermoelectric Users in Virginia." *JAWRA Journal of the American Water Resources Association* 1–19. <https://doi.org/10.1111/1752-1688.13037>.

of water shortages may be dramatically different if return flows are accounted for. This uncertainty compounds when considering future water allocations. While water withdrawals are declining nationwide (Dieter et al. 2018), trends in consumed water are less clear. Studies have observed increasing consumption trends in the industrial sector (Shaffer and Runkle 2007) but suggest that energy sector consumption is likely to decrease in the future, particularly under scenarios where renewable energy production expands (Peer and Sanders 2018; Khanal et al. 2020). This creates a key challenge for holistic water planning to balance competing demands for energy, industry, agriculture, water supply, and freshwater ecosystems. For instance, lack of data on water consumption was identified as a primary challenge in state water planning in Virginia (JLARC 2016). Lack of data on water consumption presents a key challenge in many current water resources research priorities as well, making it difficult to quantify the water footprint of many economically important activities (Mayer et al. 2016; Rushforth and Ruddell 2018) and characterize broad-scale coupled systems such as the water-energy-food nexus (Ruddell 2018).

Data on water consumption were last included in the United States Geological Survey (USGS) census of water use in 1982 and 1995 but were removed due to concerns about the validity and precision of these data (Ruddell 2018). In the face of limited empirical data, multiple approaches are used to calculate water consumption depending on the type of water use. Evapotranspiration models and remote sensing data can be employed to estimate consumption of water used for agricultural and landscape irrigation (e.g., Gates et al. 2012; Goldstein 2014; Dubinsky and Karunanithi 2017). In the thermoelectric sector, recent research has leveraged thermodynamic models to quantify the water consumed in electricity generation (Diehl and Harris 2014), as well as evaporative losses and thermal impacts associated with cooling water withdrawals (Chini et al. 2020). In water use sectors where consumption is not necessarily driven by evapotranspiration, other approaches are applied. Residential water consumption is sometimes based on water service and sewerage records (Azadeh et al. 2012; Altunkaynak and Nigussie 2017; Hastie et al. 2020), but this is only possible in instances where service areas are the same or household-level data are available. Because no centralized database of water utility data exists, this approach relies on data requests from individual utilities, making broad-scale, generalizable results challenging (Chini and Stillwell 2017). In other instances, residential consumption is estimated using the winter base rate method, where winter water use is assumed to be nonconsumptive indoor use and any additional use in

summer is assumed to be consumptive outdoor use (Ducnuigee et al. 2015; Li et al. 2017). However, this approach is not suitable in locations where winter irrigation occurs (Mini et al. 2014) or population changes seasonally due to factors such as tourism or student populations (Ducnuigee et al. 2015). Furthermore, public water supply systems also provide water for nonresidential commercial and industrial facilities that may consume water year-round (Shaffer 2009). Thus, applying the winter base rate method to municipal withdrawal data that serves a diversity of customers, rather than household use data, could lead to inaccuracies. Industrial and commercial users may also be self-supplied and reflected in separate withdrawal records. The diversity of water use processes in these sectors, combined with their small size and frequent exclusion from reporting requirements, leads to considerable uncertainty in estimates of water consumed by commercial and industrial facilities (Fanning and Trent 2009).

Given these challenges, studies and water managers often rely on consumption coefficients that provide a general estimate of the percentage of withdrawn water that is consumed for a certain water use sector (e.g., Mayer et al. 2016; Rushforth and Ruddell 2018). However, a review of published consumption coefficients across different sectors found that few studies calculated empirically based coefficients, and those that did were often limited in geographic scope and heterogeneity of users (Shaffer and Runkle 2007). This raises questions about the generalizability of these coefficients and the degree to which they are appropriate for use in other contexts. Furthermore, even facilities within the same sector will not use water in identical ways, meaning consumption rates will vary on a facility level. While industry-specific consumption coefficients can help with this issue, the most comprehensive compilation of consumption coefficients based on Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) codes is based almost entirely on data collected between 1954 and 1983, and thus may not be reflective of current practices (Shaffer and Runkle 2007; Ruddell 2018). Consumption also may vary due to seasonal water use patterns, (Shaffer 2009) climatic conditions (Talati et al. 2016; Li et al. 2017) or the introduction of new technologies (Shaffer and Runkle 2007). However, most consumption data available to calculate coefficients are snapshots of a single point in time or a low temporal resolution (1–5 years), masking this variability (Dunham et al. 2017).

A review concluded that for the majority of water uses evaluated (with the exception of irrigation and public water supplies), the most accurate method for estimating the volume of water consumed is to use

facility-level withdrawal and discharge records (Shaffer 2009). While this approach has been adopted in some studies (Shaffer 2009; Goldstein 2014; Dubinsky and Karunanithi 2017), a primary limitation in doing so is the availability of necessary data at appropriate temporal and spatial scales (Zhang and Balay 2014; Ducnuigee et al. 2015). While models of evapotranspiration and cooling water consumption have been applied to estimate consumption in the agricultural and thermoelectric sectors, respectively, this leaves a gap in understanding consumption in the industrial and commercial sectors where few alternatives for calculating consumption exist. Better industrial and commercial consumption estimates would greatly improve characterization of some of the largest water users in humid regions where irrigation is not widespread, but supply shortages and drought impacts are becoming more severe (JLARC 2016).

However, withdrawals and discharges of water are typically managed under separate reporting and regulatory programs. This means that even a facility that does withdraw and discharge water will likely report those volumes to different regulatory programs, and the required data format is unlikely to be consistent across them. Furthermore, in some cases, one-to-one matching between withdrawing and discharging facilities is inappropriate or infeasible. In this case, spatially aggregated consumption estimates could provide a way to identify areas likely to experience water stress due to consumptive withdrawals. However, various issues can confound discharge-based consumption estimates, such as leakage and inflow along distribution networks, the presence of stormwater in wastewater discharge, and differing customer bases (Shaffer 2009; Goldstein 2014; Ducnuigee et al. 2015). While these issues could introduce errors into discharge-based consumption estimates, their potential magnitude is not well understood.

More research that leverages discharge data could potentially provide a better characterization of water consumption across multiple users and time scales. This research aims to estimate water consumption by combining federally maintained, publicly available discharge data with state-maintained withdrawal data at monthly time steps to estimate facility-level and spatially aggregated water consumption across Virginia between 2010 and 2016. Our specific objectives are to (1) evaluate the feasibility of using publicly available discharge data to estimate sub-annual water consumption, (2) characterize how these consumption estimates vary depending on the level of spatial aggregation, and (3) assess what patterns of seasonality exist in consumption estimates across different sectors. We use geospatial and text-mining methods to match facilities across the withdrawal

and discharge datasets, demonstrating that this approach can account for the vast majority of water used in the state. We also develop spatially aggregated consumption estimates at the state, hydrologic unit code 6 (HUC-6), HUC-8, and HUC-10 levels. This results in an updated set of sector-specific, seasonally varying consumption estimates that can more precisely characterize water consumption across different facilities and time periods.

METHODOLOGY

Our process for using discharge records to calculate sector-specific, seasonally varying consumption estimates required several steps in data processing, quality assurance and quality control (QA/QC), and analysis. QA/QC procedures were first applied to identify erroneous discharge volumes and mislabeled outfalls among the discharge records. Text mining was then used to classify withdrawals and discharges into six water use categories: agriculture/irrigation, aquaculture, commercial, thermoelectric, industrial, or municipal. Withdrawals and discharges reporting from the same facility were then linked using geospatial analysis and textual analysis, creating a subset of data to calculate water consumption at a facility-level. Spatially aggregated consumptive rates were then ultimately calculated at the HUC-10, HUC-8, HUC-6, and statewide level, using all withdrawal and discharge records for the period of analysis. The details of this procedure are presented in the following sections.

Data Sources

Monthly surface and groundwater withdrawal records were obtained from the Virginia Department of Environmental Quality (DEQ) Annual Water Withdrawal Reporting database. The DEQ database includes approximately 1.35 million water withdrawal entries that represent monthly withdrawals from 4,132 facilities from 1982 to 2017. The data were trimmed to 2010 to 2016 to match the temporal coverage of the discharge dataset, resulting in 383,962 entries from 1,860 facilities.

Discharge monitoring reports (DMRs) were obtained from the U.S. Environmental Protection Agency's (USEPA) Enforcement and Compliance History Online (ECHO) database. This database houses point-source discharge records associated with the National Pollutant Discharge Elimination System (NPDES), which requires permits for any facility

discharging from a point source into the waters of the United States (U.S.). The DMRs contain reported values for each discharged monitored pollutant, including flow. Relevant facility attributes (FacilityID, Latitude/Longitude, and Permit Type) were obtained for all sites in Virginia, returning a list of 1,143 regulated facilities with an NPDES individual permit. DMR records associated with parameter code 50050 (flow in conduit or through treatment plant) between 2010 and 2016 were then obtained for all outfalls associated with these facilities. This time period was selected because data displayed in ECHO are only partially complete before the year 2009 (USEPA 2019) and because 2016 was the final year in which complete withdrawal data were available. In total, 25,849 discharge records from 734 facilities were extracted. While the majority (674) of facilities reported monthly discharge, some facilities reported annually (23 facilities), semi-annually (15 facilities), and quarterly (68 facilities). In these instances, monthly discharge was assumed to be constant over the reporting period. Outfall types were labeled using the “perm feature type code” attribute so that external outfalls tracking water that leaves the facility as discharge were separated from internal outfalls (which track water that is used within the facility) and MS4 outfalls (which track discharges from municipal separate storm sewer systems).

QA/QC Procedures

Direct linkages from water facility monitoring systems to state and federal reporting databases are rare. As a result, data quality can sometimes be impacted via transcription errors, with the most significant errors relating to unit conversion errors by reporting users (e.g., gallons per day reported as million gallons per day). As a result, using data from these systems requires QA/QC procedures to identify erroneous data, especially unit conversion errors. A summary of anomaly types observed in the ECHO dataset, and the corrections implemented for each type is presented in Table 1. Reported effluent values that are over 100 times greater than the outfall’s median discharge were flagged as potentially anomalous; this was the case for 25 of the 30,987 discharge entries. Flagged values were omitted from further analysis.

Because the parameter code 50050 includes internal and MS4 outfalls, these were removed from the analysis so that only discharge from external outfalls would be quantified. Although 1,078 of the 1,098 outfalls in ECHO were classified as external, a comparison of these outfalls against their VPDES permit ID numbers indicated that some had been mistakenly classified as such, resulting in excess reporting of

TABLE 1. A description of quality assurance and quality control issues present in Enforcement and Compliance History Online (ECHO) discharge records and methods adopted to resolve them.

Issue	Description	Resolution
Duplicate discharge monitoring report (DMR) entries	ECHO records sometimes included duplicate entries where an outfall’s discharge was present twice for a single monitoring period	Duplicate entries with identical reporting dates and discharge volumes were removed
Hydroelectric generating facilities	Hydropower withdrawals are not included in the Department of Environmental Quality (DEQ) annual reporting dataset due to a change in reporting requirements, but discharges are still present within the ECHO database	Hydropower facilities with individual National Pollutant Discharge Elimination System (NPDES) permits were removed from analysis
Anomalous discharge quantities	There are instances of suspected error in reported discharge values, with flows that are several orders of magnitude greater than the outfall’s median flow	Discharge records that exceeded the outfall median discharge by over two orders of magnitude were replaced with the outfall median
Mislabeled outfalls	Outfalls that track stormwater, internal processes, and MS4s are included in DMRs and, by default, are labeled as external in the ECHO system	VPDES permit numbers were reviewed and mislabeled outfall types were updated according to permit information
Sectoral classification	ECHO classifies all facilities as either industrial or municipal, whereas facilities in the DEQ dataset are classified according to 10 different sectors	Standard Industrial Classification (SIC) codes were used to classify ECHO facility when available. When SIC codes were unavailable, a semi-supervised text-mining approach was used to classify facilities

discharge. Mislabeled outfall types were updated according to VPDES permit information. Excluding internal outfalls decreased the dataset from 1,098 unique outfalls (31,812 DMR entries) to 729 unique outfalls (18,024 DMR entries).

Facility Level Classification

Facilities that report data to the DEQ withdrawal database classify themselves into 10 water use sectors: agriculture, irrigation, commercial, industrial, manufacturing, mining, fossil fuel power, hydropower, nuclear power, and public supply. All facilities in ECHO are categorized as either industrial (27%), municipal (68%), or NA (5%). To account for the discontinuity between the sectoral classification system across the two datasets, a semi-supervised learning approach was used to sort withdrawal and discharge facilities into one of six water sectors: agriculture/irrigation, aquaculture, commercial, thermoelectric, industrial, or municipal. In this analysis, the thermoelectric sector includes nuclear, coal, combination fossil fuel, and biomass electricity generation facilities. Hydropower was not included in the analysis because hydropower withdrawals were not reported after 2005, prior to the period when discharge records were available. Municipal facilities include public suppliers, water treatment, and wastewater treatment facilities. Commercial facilities are public centers that are not considered to be residential, including golf courses, hotels, schools, hospitals, government buildings, and prisons. Agricultural/irrigation facilities include farms, dairy farms, and nurseries. The vast majority of these facilities do not have a corresponding discharge permit associated with them, but aquaculture facilities do. Therefore, aquaculture was analyzed separately from agriculture. The industrial sector then covers manufacturing, mining, and industrial waste treatment from facilities.

The semi-supervised classification method leveraged sectoral labels from the withdrawal records and SIC codes available for some facilities in ECHO to initially classify the water sector for each discharging facility without an SIC code. Using guidance from domain experts and a training set of 1,793 withdrawing facilities, text mining techniques were used to identify the most frequent terms in facility names by water sector while excluding legal classifications such as “INC,” “LLC,” and “LLP.” This approach does have some limitations; for instance, many withdrawing agricultural facilities simply included the name of the owner, with no unique terms, in their facility name. Water treatment plants can also belong in the municipal or industrial sectors. Therefore, the rules created in this process involved domain expertise and several iterations of manual review in efforts to decrease classification error.

Facility Matching and Analysis

Calculating facility-level consumption requires identifying self-supplied water withdrawing facilities

that also discharge water, but not all withdrawal facilities will have discharge records as well. Irrigation and agriculture withdrawals rarely have corresponding discharge records, and municipal water and wastewater facilities generally cannot be matched due to the fact that municipal wastewater is often treated at a different location(s), and may be handled by different organizations than the potable water source. Nor are these linkages between water producers and wastewater processors delineated in public reporting databases. For this reason, only facilities in the aquaculture, commercial, thermoelectric, and industrial sectors were matched. Even in instances where a facility reports withdrawal and discharge, facility names and coordinates may differ slightly.

To overcome these challenges, we employed a mechanism that screens potential matches for discharging facilities first by distance and then by facility name. In the first step, the coordinates of each discharging facility were compared to all withdrawing facilities in that sector to identify withdrawals located within 10 km. The names of withdrawing facilities within 10 km were then compared to the discharging facility using fuzzy string matching, which quantifies the similarity between two character strings. A custom function was created to calculate the “string distance” between each discharging and withdrawal facility within 10 km of each other. String distance is equal to the minimum number of insertions, deletions, and substitutions of letters needed for the names to be identical. Prior to calculating the string distance, the function will slide blocks of text so they are consistent between the two strings, and trim spaces and unnecessary characters to reduce the string distance. This mechanism accounts for the fact that facility names across the two datasets are often inconsistent but share overlapping patterns or words. For instance, the facility name for the North Anna Nuclear Power Plant is “Dominion Generation: North Anna Nuclear Power Plant” in the DEQ dataset and “Dominion — North Anna Power Station” in ECHO; the string distance between these two names is 14.

This algorithm was applied to a total of 358 discharging facilities and conducted separately for each of the four included water use sectors. For each facility, a match was selected if the closest facility in terms of string distance was also the closest geographically. This algorithm resulted in 185 one-to-one matches between withdrawal and discharge facilities. These matches were subsequently reviewed by manually comparing additional attributes from the NPDES and DEQ databases, including owner, average withdrawal/discharge, address, and permit information to confirm or reject the matches.

Once matches were confirmed, facility-level consumptive use fractions (CU) were estimated by

summing monthly withdrawals (W) and discharges (D) from that facility and estimating consumptive use as the percentage of withdrawn water not returned via discharge. While some facilities were present in both datasets, they did not always have corresponding withdrawal and discharge records in all reported months. Thus, long-term withdrawal and discharge in million of cubic meters (MCM) per day were calculated using only the months where both withdrawal and discharge were reported and where reported values were within two orders of magnitude of each other to reduce the influence of missing or erroneous data. Because some matched facilities did not have overlapping periods of reported withdrawals and discharges, these facilities were excluded from the consumption calculations. These long-term totals were then used to calculate consumptive use as in Equation (1):

$$CU = \frac{(W-D)}{W}, \quad (1)$$

Seasonal consumptive rates for warm months (April–October) and cool months (November–March) were also calculated by summing withdrawals and discharges from those periods only.

In calculating consumptive use fractions in Equation (1), all withdrawals and discharges were summed regardless of the source or receiving water body. For instance, if a facility withdraws from groundwater and discharges the full withdrawal to surface water, this results in a CU value of 0. This definition of consumption is thus focused on water-use behavior at the facility level. Other definitions may be more appropriate if aimed at understanding the hydrologic impacts of water withdrawals, particularly in hydrologic systems with extensive interbasin transfers or little connectivity between different sources and return points. The use of alternative definitions of consumption can have significant impacts on the resulting estimates (Grubert and Sanders 2018); thus it is crucial to use a definition that is consistent with the overall goals of the analysis.

Spatial Aggregation Analysis

Under ideal conditions, there would be a one-to-one match between each NPDES permitted facility and its respective withdrawing permit to calculate water withdrawal, discharge, and consumption at the facility level. However, the disconnect in regulations, reporting, and data sources affects the quality and availability of data available in each dataset. Furthermore, there are some withdrawing facilities, particularly irrigation and municipal facilities, that do not have a single (or any) discharge facility associated with them.

Even in the industrial and commercial sector, some facilities may purchase water from public suppliers and discharge under their own permit, or withdraw their own water but discharge to public sewage systems. Thus, not all withdrawal facilities will have matching facility-level discharge records. Nevertheless, management and planning of water supply requires some understanding of water consumption across the multiple uses and facilities that cannot be matched. In these instances, spatially aggregated consumptive use estimates (for instance, at the watershed scale) can be useful in estimating the water balance across different areas and identifying locations with the greatest risk of water stress due to high consumption. This is reflected in the USGS baseline goals for water use reporting, which includes spatially aggregated consumption estimates for several use categories (USGS 2017). However, little research exists that systematically evaluates the impact and potential errors of these spatially aggregated estimates. Thus, we also calculated spatially aggregated consumptive rates to evaluate the suitability of using discharge data for this purpose and to assess how errors and inconsistencies in reported data behave at aggregate scales. The spatial scales considered in this study include HUC-10, HUC-8, HUC-6, and statewide. Thermoelectric facilities were excluded from the spatially aggregated totals because the large volume of water associated with thermoelectric use can mask the behavior of other sectors, and the vast majority of thermoelectric withdrawals (90.9%) and discharges (99.9%) can be accounted for in matched facilities. Spatially aggregated consumption was calculated by summing all monthly withdrawals and discharges within the watershed and using these to calculate a total consumption rate as in Equation (1). Discharge records for many smaller facilities are only present in ECHO for the year 2016; while discharges associated with these facilities are small enough to only have a minor impact on statewide estimates, they could potentially skew results at finer spatial scales. Thus, the watershed scale analysis was based on data from 2016 only, while the statewide assessment used data from the full period of analysis.

RESULTS

Facility Matching Results

Of the 358 facilities included in the matching procedure, 246 facilities had a same-sector withdrawal facility located within 10 km. The median number of withdrawal facilities within 10 km was 1 in the

aquaculture, commercial, and thermoelectric sectors, and 2 in the industrial sector (Table 2). This indicates that most discharge facilities only had a small number of withdrawals within 10 km. Ultimately, the algorithm identified a total of 185 one-to-one matches where the closest geographic match was also the closest match in terms of facility name. However, after manual review, 89 of these matches were determined to be false positives (i.e., a match was selected but determined to be inaccurate) based on discrepancies between the facility name, permit information, and water use data. In addition to the 96 true positives identified through the algorithm steps, 27 matches were determined via manual review, resulting in a final dataset of 123 matched facilities. The algorithm thus results in a false-negative rate of 16% (27/173) and a false-positive rate of 48% (89/185). Of these facilities, 24 did not have any overlapping periods of withdrawal and discharge records (i.e., there were no months where both withdrawal and discharge data from that facility were available), so the final dataset used for calculating facility-level consumption consisted of 99 facilities.

These matched facilities account for approximately 77% and 90% of the total reported annual withdrawal and discharge, respectively (Table 2). Although the matched facilities only account for a relatively small proportion of facilities, these matched facilities tend to be large water users. Thus, the procedure accounts for the majority of water withdrawn and discharged in the state. This is especially true in the thermoelectric sector, where matched facilities account for 90.9% of withdrawal and 99.9% of discharge, and industrial sector where matched facilities account for 58.4% of withdrawal and 98.3% of discharge.

Facility-level consumptive use was calculated for 99 matched facilities in the thermoelectric, industrial, commercial, and aquaculture sectors. Histograms showing variation in consumptive use fractions (Equation 1) across matched facilities in each sector are shown on the left side of Figure 1, with histograms of consumptive volumes [$W - D$] on the right side. A consumptive use value of 1.0 indicates that the withdrawal is entirely consumptive; that is, the discharge associated with that facility is zero. A consumptive fraction value of 0.0 would indicate entirely nonconsumptive use (discharges are equal to withdrawals) and a negative value indicates facilities reporting greater discharge than withdrawal. The industrial and commercial sectors both included a small number of facilities with discharges much greater than withdrawals. Because of the magnitude of this difference, these facilities exhibit strongly negative values of consumptive fractions (ranging to -80). However, the consumptive volume data on the right show that the volume of discharge exceeding withdrawal in these sectors is less than 200 and 15 m³ per day in the industrial and commercial sectors, respectively. Figure 2 shows that facilities that use large volumes of water typically have greater consumptive use values and fewer instances of discharge exceeding withdrawal.

The presence of facilities reporting discharge greater than withdrawal could be caused by data issues, such as reporting errors or mismatched records, facilities that receive some water from public supplies, or be reflective of infiltration into discharge pipes. NPDES requirements specify that “flow measurements should be within 10% of the inspector’s

TABLE 2. Facility matching results. Because the facility matching process was implemented separately for each sector, intermediate matching process results are only presented for individual sectors.

	All	Aquaculture	Commercial	Thermo-electric	Industrial
Facility data					
Withdrawal facilities	628	13	403	21	191
Discharge facilities	358	12	228	23	95
Withdrawal volume [MCM/day]	27.21	0.1287	0.0795	21.97	1.654
Discharge volume [MCM/day]	22.58	0.1249	0.053	19.56	0.6814
Matching process — intermediate steps					
NPDES facilities with at least one withdrawal facility within 10 km	NA	7	153	16	70
Median number of withdrawal facilities within 10 km (min-max)	NA	1 (0–2)	1 (0–10)	1 (0–2)	2 (0–9)
Median string distance of facilities within 10 km (min-max)	NA	14 (8–21)	16 (0–31)	13 (3–25)	11 (0–30)
Matching process — final results					
Number of matched facilities (algorithm)	185	7	108	17	53
Number of matched facilities after final review	123	9	54	14	46
Number of matched facilities with overlapping withdrawal and discharge records	99	8	39	12	40
Withdrawal volume [MCM per day] (% of total)	20.99 (77.1%)	0.1249 (97.1%)	0.0246 (31.0%)	19.98 (90.9%)	0.9653 (58.4%)
Discharge volume [MCM per day] (% of total)	20.26 (89.7%)	0.1060 (84.8%)	0.0208 (39.3%)	19.55 (99.9%)	0.670 (98.3%)

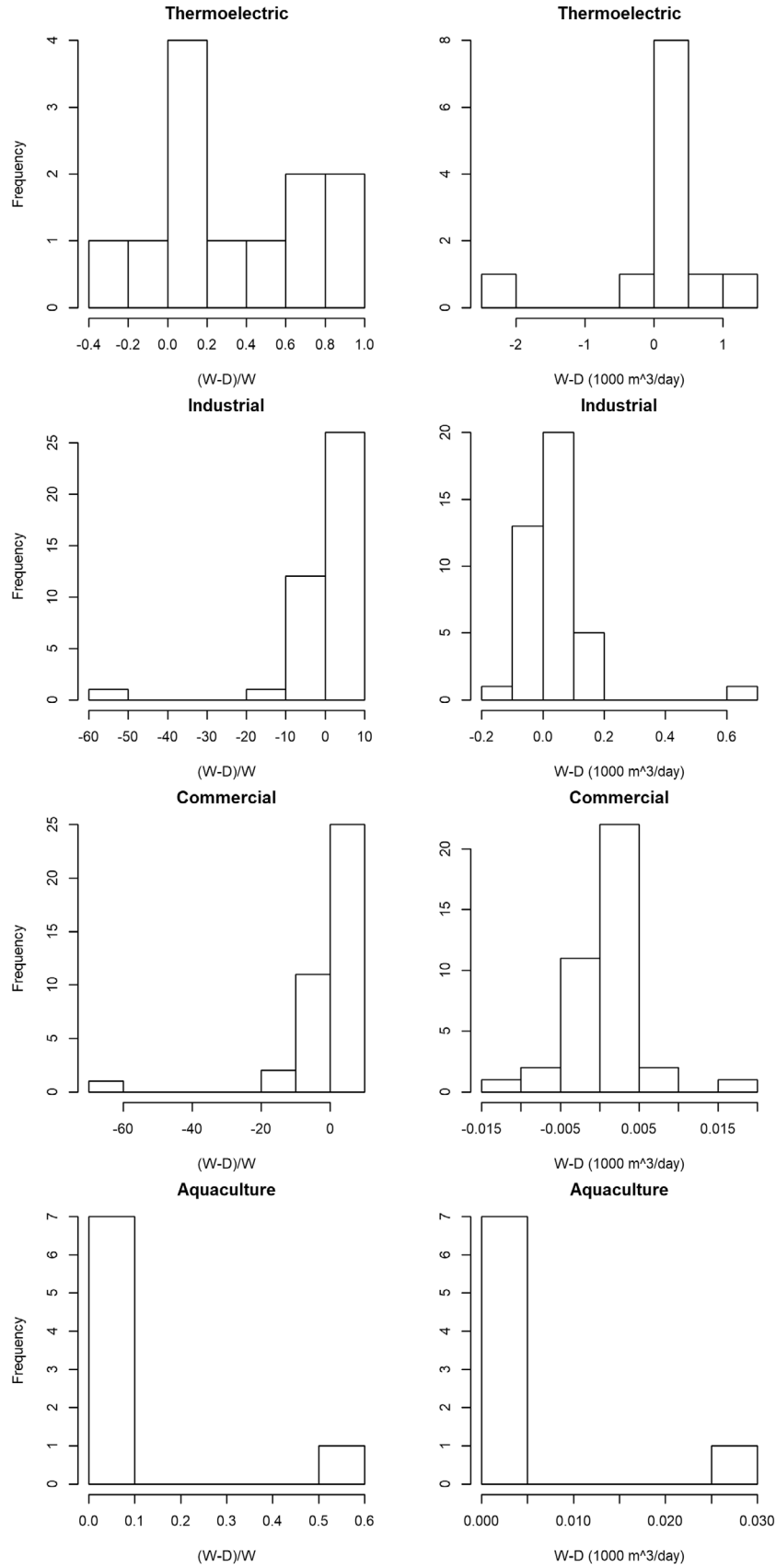


FIGURE 1. Histograms of facility-level consumptive fraction (left column) and consumptive volume (right column) estimates for the four sectors evaluated. Negative values indicate instances where discharge values exceeded withdrawals. *D*, discharge; *W*, withdrawal.

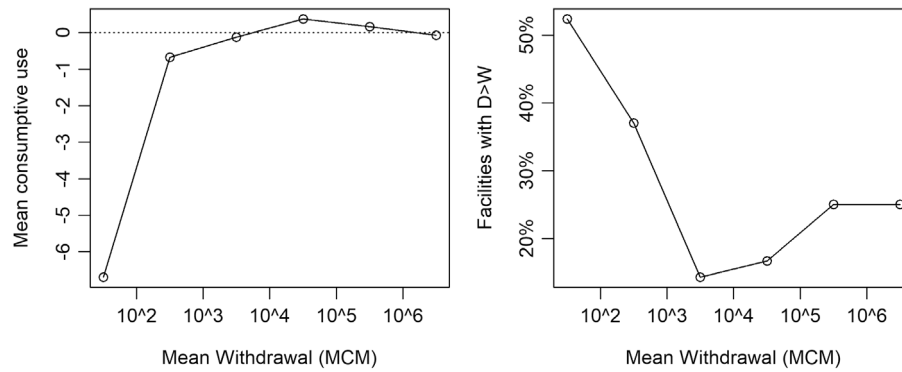


FIGURE 2. Mean consumptive use ($(W - D)/W$; unitless; left) and the percentage of facilities with discharge greater than withdrawal (right) for facilities of different withdrawal volumes. Negative consumptive use values indicate instances where discharge values exceeded withdrawals. MCM, million cubic meters.

TABLE 3. Summary of facility-level instances of discharge exceeding withdrawal ($D > W$) and exceeding withdrawal by over 10% ($D > W \times 110\%$) across all data and across warm (April–September) and cool (October–March) seasons.

Sector	Number of matched facilities	Facilities with discharge exceeding withdrawal (all records)		Facilities with discharge exceeding withdrawal (April–September)		Facilities with discharge exceeding withdrawal (October–March)	
		$D > W$	$D > W \times 110\%$	$D > W$	$D > W \times 110\%$	$D > W$	$D > W \times 110\%$
Thermoelectric	12	2	2	2	2	2	1
Industrial	40	14	9	14	9	12	9
Commercial	39	14	10	13	10	17	12
Aquaculture	8	0	0	0	0	0	0
Total	99	30	21	29	21	31	22

measurements to certify accurate flow measurement” (USEPA 2017); meter inaccuracies within this range could contribute to some instances of discharge exceeding withdrawal. Table 3 presents the number of facilities in each sector who reported discharge greater than withdrawal and over 10% greater than withdrawal when averaged across all records, and when averaged across only the warm and cool months. While 30 of the 99 facilities exhibited greater discharges than withdrawals, only 21 were over 10% greater than withdrawals, suggesting that some of these instances could be due to meter inaccuracy. Because rainfall and soil moisture levels tend to be highest in the winter months in Virginia, an increase in facilities reporting discharge greater than withdrawal in the winter would suggest that this is largely driven by infiltration into discharge pipes. This is the case for four facilities in the commercial sector, but none in the industrial or thermoelectric sectors.

Spatial Aggregation Analysis

To characterize consumption of water in sectors and facilities unaccounted for in the matched

facilities, spatially aggregated estimates of consumptive use across all nonthermoelectric facilities (i.e., matched and unmatched) and municipal facilities only were calculated across several watershed scales (Figure 3). Consumptive use was calculated as in Equation (1), but summing all withdrawals and discharges within the spatial unit prior to calculation. Thus, it represents the relative volume of withdrawn water not returned via discharge to any water body in that watershed. A value of 1.0 indicates no discharges, and a value less than 0.0 indicates that discharges exceed withdrawals. The left column presents results for all nonthermoelectric facilities. At the smallest (HUC-10) spatial scale, many locations exhibit discharges exceeding withdrawals, or very high consumption estimates (>0.75) where discharges are significantly smaller than withdrawn water. In some cases, discharges significantly exceed withdrawals. In certain locations, this may be an accurate reflection of water use. For instance, the city of Virginia Beach is supplied with an average of 130,000 m³ per day from Lake Gaston, located 76 miles away, while its wastewater is treated and discharged locally, resulting in local discharge values that exceed withdrawals. Accounting for these

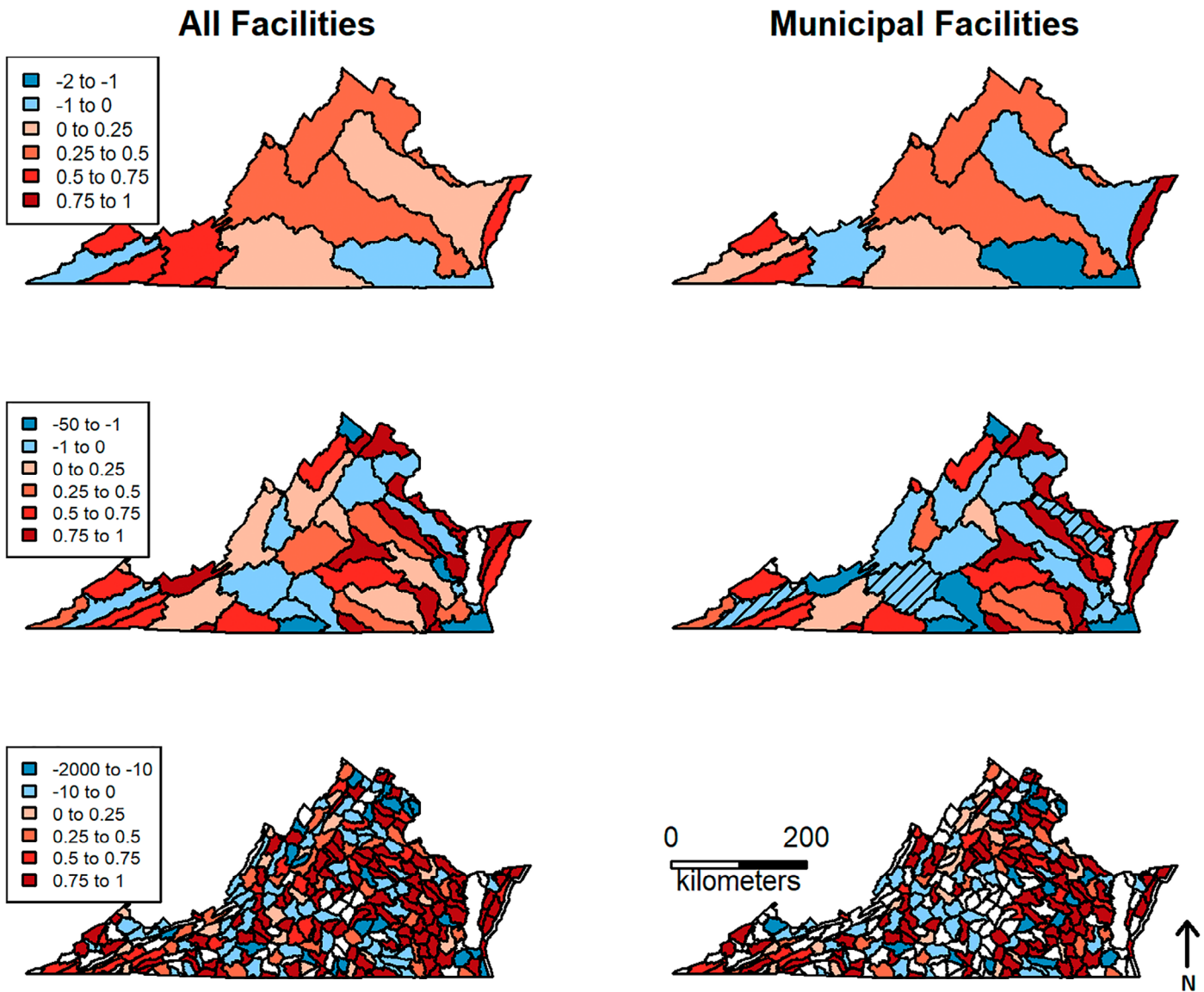


FIGURE 3. Spatially aggregated consumptive use ($[W - D]/W$; unitless) estimates for all nonthermoelectric withdrawals (left column) and municipal withdrawals (right column) in 2016. From top to bottom, results are presented for the hydrologic unit code 6 (HUC-6), HUC-8, and HUC-10 watershed scales. Empty (white) watersheds are those with no reported withdrawals. Negative values indicate instances where discharge values exceeded withdrawals. Hatched watersheds in the municipal HUC-8 map indicate case study watersheds shown in Figure 4 (from west to east, the Upper Clinch, Roanoke, and Lower Rappahannock rivers).

interbasin transfers could lead to more consistent consumption estimates. Similarly, the Eastern Shore peninsula of Virginia has one of the highest rates of irrigation in the state and high levels of withdrawal relative to point-source discharge in this location are expected. However, in other instances, this may reflect reporting errors that were also apparent in the facility-matching data.

The presence of negative or very high consumption becomes less frequent at larger spatial scales, with only two HUC-6 watersheds exhibiting discharge greater than withdrawal and none exhibiting consumption greater than 0.75. At larger scales, the

lower bound of consumptive use estimates also becomes less extreme (i.e., -2 at the HUC-6 scale compared to $-2,000$ at the HUC-10 scale). This could be due to the impact of interbasin transfers. While some long-distance transfers exist in the state (such as the Lake Gaston transfer described above), the majority are over relatively short distances. For instance, a municipal water supply and wastewater system could potentially withdrawal and discharge water across separate HUC-10 watersheds, but is less likely to do so across HUC-6 boundaries. Calculating consumption in the presence of these transfers introduces a level of complexity and can lead to instances

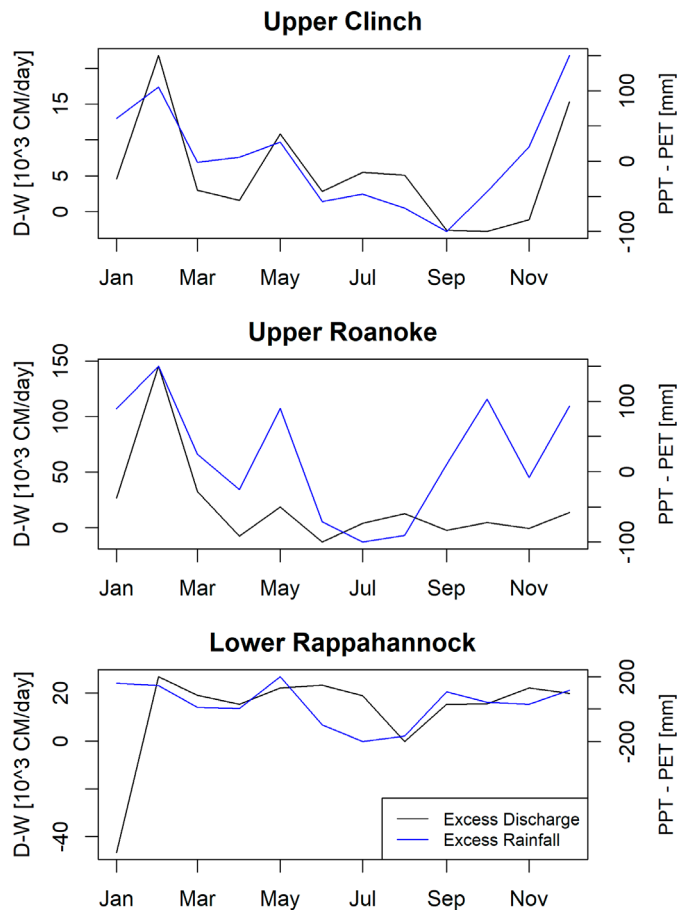


FIGURE 4. Time series of monthly excess discharge ($D - W$) and excess rainfall (precipitation – potential evapotranspiration) in three HUC-8 basins that exhibited greater discharge than withdrawal.

of both very high consumption (when water is withdrawn and transferred to a different watershed) as well as a net influx of water in receiving watersheds resulting in greater discharges than withdrawals.

The right column of Figure 3 presents spatially aggregated estimates for consumption in the municipal sector. Matching municipal withdrawal and discharge records is generally not appropriate due to differing customer bases and separate locations of water and wastewater treatment facilities. Spatially aggregated estimates of municipal consumption can address these issues, but may still be impacted by the occurrence of leakage and infiltration along extensive pipe networks and confounding issues such as combined stormwater systems. At finer spatial scales, these issues appear to be common, with many HUC-10 watersheds and counties exhibiting either discharge greater than withdrawal or very high (>75%) consumption rates. These issues appear to cancel each other out as the scale of spatial aggregation increases; while 32% of HUC 10 watersheds have

discharge greater than withdrawal, this occurs in only 28% and 18% of HUC 8 and HUC 6 watersheds, respectively (Figure 3).

To investigate potential drivers of discharge exceeding withdrawals at the watershed scale, three HUC-8 watersheds were investigated in more detail by comparing excess discharge ($D - W$) with excess rainfall (precipitation minus potential evapotranspiration). Excess rainfall in each location was calculated for each month in 2016 using gridded daily temperature and precipitation records from the PRISM dataset (Daly et al. 2008) with potential evapotranspiration estimated using the Thornthwaite method. Excess discharge in the Upper Clinch and Roanoke river basins tends to correlate with excess rainfall (Figure 4), suggesting that some portion of excess discharge could stem from infiltration of soil moisture into discharge infrastructure, as has been noted in other studies (Ducnuigeen et al. 2015; Hastie et al. 2020). However, interbasin transfers also play a role. This occurs in Virginia's Lower Rappahannock basin (3), where water supplies for Fredericksburg and Spotsylvania counties are withdrawn from multiple HUC-8 watersheds (the Rapidan/Upper Rappahannock, Lower Rappahannock, Lower Potomac, and Mattaponi), but are all discharged into the Lower Rappahannock below Fredericksburg.

It should also be noted that both withdrawal and discharge records exhibit a large amount of temporal variability; for instance, withdrawals in the Lower Rappahannock in January 2016 were about 60,000 m³ higher than they were during the remainder of the year to fill the Rocky Pen Run Reservoir. These variations point towards the benefit of integrating ECHO records with institutional knowledge about the water systems of interest to put the data in context. Even without this institutional knowledge, the impact of these outliers on overall water balance estimates decreases at larger scales, and points to the value of aggregate statistical measures (such as median values) that are robust to these outliers. Although it is beyond the scope of this study, systematic research that leverages the ECHO discharge records along with detailed data on municipal water supply and wastewater service areas could help identify the prevalence and driving factors associated with excess municipal discharge values.

Statewide Analysis

In addition to the watershed-scale consumption estimates presented above, statewide estimates of consumptive use were calculated at a monthly time step to characterize consumption behavior across different water use sectors. Figure 4 presents time

series of the total statewide volume of withdrawal and discharge for all nonthermoelectric facilities, as well as for the industrial and municipal sectors. These statewide withdrawal and discharge values were used to calculate a time series of consumptive use as in Equation (1), but summing monthly withdrawal and discharge across all facilities in the state

prior to calculating consumptive use. Time series for the aquaculture and commercial sectors were not included because discharge records in these sectors were only sparsely available prior to 2016, and time series for the thermoelectric sector is shown in the results for matched facilities in Figure 5. Statewide nonthermoelectric withdrawals exhibit a clear

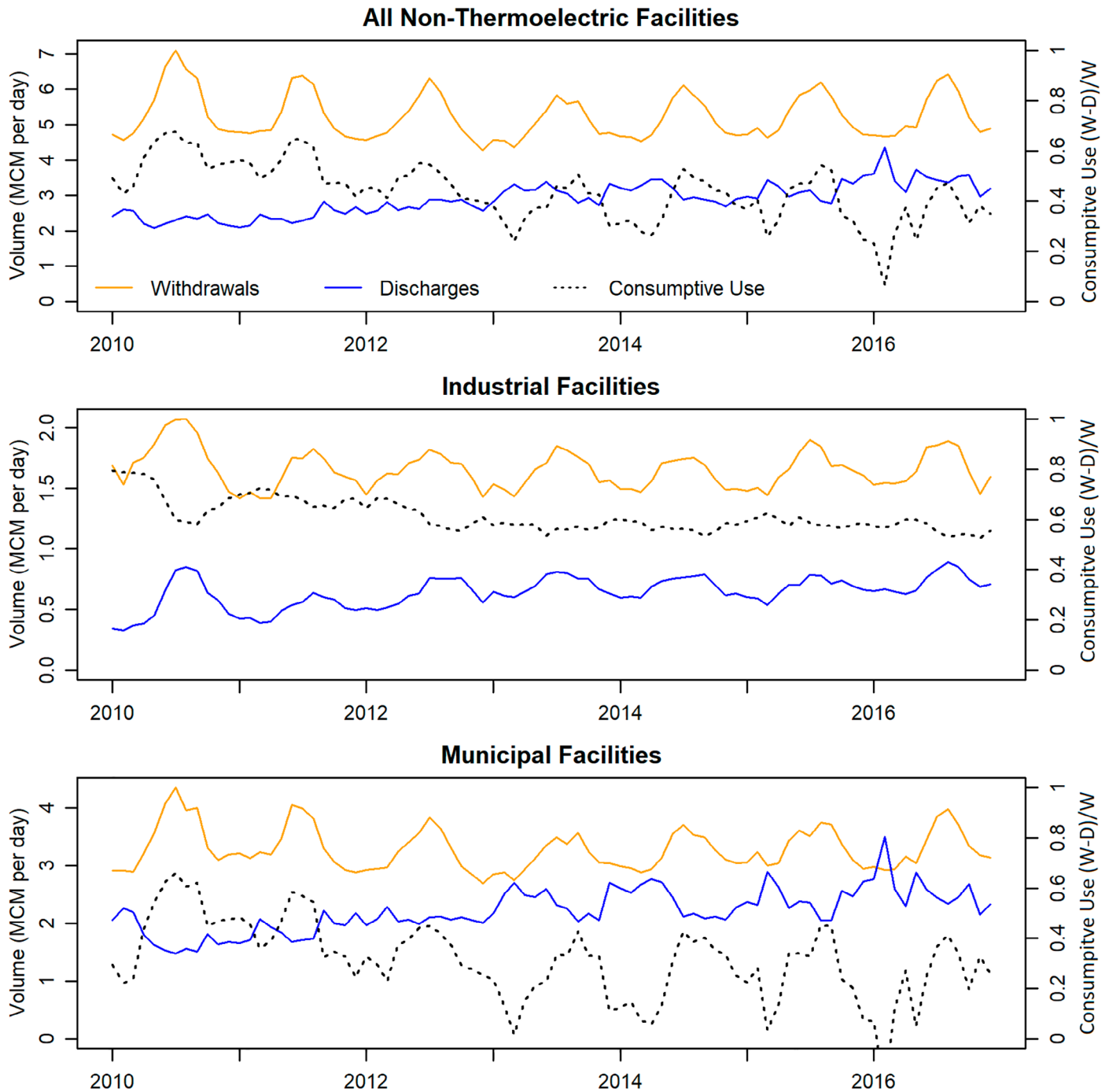


FIGURE 5. Time series of withdrawals, discharges, and consumptive use for all reporting nonthermoelectric facilities in the state, as well as results for the industrial and municipal sectors. Note that discharges in the commercial and aquaculture sectors were largely unreported prior to 2016, and are thus not included in the Figure.

seasonal pattern, but this seasonality is not present in discharge volumes. This results in consumption rates that peak in the summer months. In the industrial sector, withdrawal and discharge both peak in the summer months which results in relatively constant consumption rates ranging from 50% to 70%.

In the municipal sector, discharges do not exhibit the seasonality apparent in municipal withdrawals, leading to consumption peaking in the summer months. However, even in wintertime, consumption is generally still greater than zero, suggesting that not all water goes toward nonconsumptive indoor use. This could be because some municipal use in winter is consumptive, or it could be the result of issues such as water losses from pipe breaks and leaks. Consumption rates in the summer range from approximately 20%–60%. While discharges are not seasonal, they are still highly variable, which could potentially be associated with stormwater infiltration and discharge. However, these discharges only exceed withdrawal in one month of 2016. This period of discharge exceeding withdrawal was in a particularly wet winter (with a total of 340 mm of rainfall between December and February, compared to an average winter rainfall of 250 mm), which may have resulted in increased discharge of stormwater treated in combined sewer systems. Across all sectors,

discharge volumes exhibit a general increase over time, which could be due to increasing volumes of discharge, improved reporting, or improved data transfer to the ECHO system.

Figure 6 presents time series of withdrawals, discharges, and consumptive use for thermoelectric and industrial facilities with matched withdrawal and discharge records. In the thermoelectric sector, withdrawals and discharges both peak in summer months, resulting in relatively constant consumptive use. Discharge records occasionally exceed withdrawal records, possibly because errors in the reported data are of a comparable magnitude to the volume of water consumed, or could potentially be reflective of actual discharges exceeding withdrawal (for instance, if stormwater is sometimes combined with discharged process water). Industrial withdrawals and discharges exhibit similar seasonal behavior, with both increasing in summer months. This results in consumption that stays relatively constant at approximately 30%, compared to 50%–70% for unmatched industrial withdrawals and discharges. Despite some individual facilities reporting greater discharge than withdrawal, these facilities are typically fairly small-volume water users (Figures 1 and 2). Because consumption estimates for larger water users tend to be less erratic, statewide

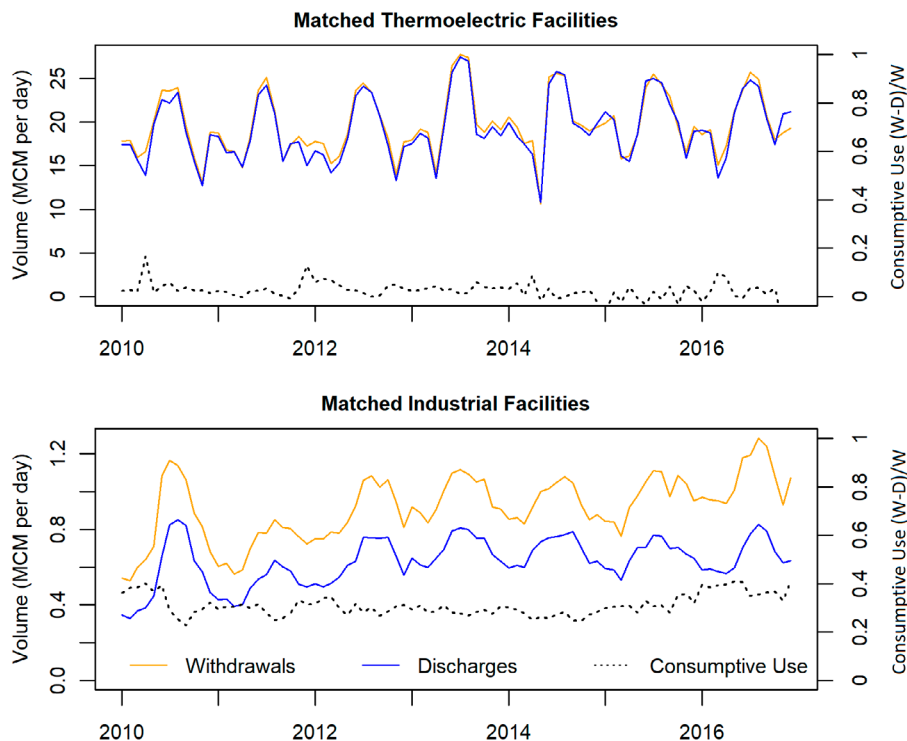


FIGURE 6. Time series of withdrawals, discharges, and consumptive use for matched facilities in the thermoelectric and industrial sectors. Note that discharges in the commercial and aquaculture sectors were largely unreported prior to 2016, and are thus not included in the figure. Municipal facilities were not included in the matching analysis.

totals are generally consistent with published consumption estimates.

Table 4 presents median values of annual, warm-season, and cool-season consumptive use (calculated as in Equation 1) across matched facilities (top half of table) and HUC-8 watersheds (bottom half of the table) in each water use sector using 2016 data. Matched facilities in the thermoelectric, industrial, and commercial sectors all exhibited higher consumption in warmer months than they did in the cooler months; however, these differences were only statistically significant in the commercial sector. Similar seasonal patterns are present in the HUC-8 results for the thermoelectric, commercial, and municipal sectors, as well as the total nonthermoelectric use. Confidence intervals around median values are wide in many sectors, likely reflecting a combination of the wide range of observed values among individual facilities and watersheds, as well as the small number of matched facilities in some sectors.

Table 5 presents percentile (25%, 50%, and 75%) values across for consumptive use estimates from 2016 across individual facilities (top half of table) and HUC-8 watersheds (bottom half of table). These values are compared to percentile values presented in Shaffer and Runkle (2007). Median consumptive use estimates based on withdrawal and discharge data at both the facility and watershed scale are higher (greater median values) and more widely distributed (larger interquartile range) compared to literature values. While 25th percentile values for facility-level consumptive use are below zero in the industrial and commercial sectors, most of these values fall within the 10% meter accuracy requirement in the NPDES guidelines (USEPA 2017). One exception is cool season consumptive use in commercial facilities (25th percentile value of -0.67). Because this value is much lower than warm season consumptive use, this could be the result of increase infiltration during winter months.

DISCUSSION

One of the objectives of this research was to evaluate the feasibility of using publicly available discharge data to estimate water consumption. While some QA/QC issues exist within the discharge records obtained from ECHO, the majority of these issues were addressed through fairly straightforward procedures. However, some institutional knowledge regarding NPDES permit numbers in Virginia was required to identify erroneously classified outfalls;

TABLE 4. Median estimates (with 90% confidence intervals in parenthesis) of facility-level and spatially aggregated (HUC-8) consumptive use ($(W - D)/W$; unitless) across different water use sectors for all withdrawals and discharges in 2016, as well as warm season (April–October) and cool season (November–December) withdrawals only. Negative values indicate instances where discharge values exceeded withdrawals. p -values in the lower rows present the results of Wilcoxon Signed Rank tests comparing the distribution of consumptive use values in the warm and cool seasons; values less than 0.05 indicate statistically significant differences in the values of warm and cool season consumption values.

	All nonthermoelectric	Thermoelectric	Industrial	Commercial	Aquaculture	Municipal
Facility level consumptive use						
Annual	NA	0.17 (0.01 to 0.75)	0.17 (0.01 to 0.44)	0.19 (−0.02 to 0.51)	0 (0 to 0.08)	NA
Warm season	NA	0.19 (0 to 0.76)	0.22 (0 to 0.44)	0.19 (0 to 0.55)	0 (0 to 0.03)	NA
Cool season	NA	0.15 (0.01 to 0.74)	0.13 (0.03 to 0.44)	0.02 (−0.06 to 0.25)	0 (0 to 0.13)	NA
Wilcoxon p -value	NA	0.97	0.41	0.01	0.18	NA
HUC 8 aggregate consumptive use						
Annual	0.46 (0.14 to 0.64)	0.07 (−0.92 to 0.84)	0.35 (0.29 to 0.83)	0.52 (0.13 to 0.71)	0 (−0.36 to 0.52)	0.28 (−0.09 to 0.58)
Warm season	0.55 (0.19 to 0.65)	0.04 (−0.62 to 0.85)	0.41 (0.27 to 0.87)	0.64 (0.23 to 0.77)	0.04 (−0.37 to 0.51)	0.35 (−0.01 to 0.65)
Cool season	0.18 (−0.04 to 0.55)	0.14 (−1.37 to 0.81)	0.43 (0.28 to 0.83)	0.4 (−0.05 to 0.62)	0 (−0.35 to 0.54)	0.18 (−0.2 to 0.58)
Wilcoxon p -value	<0.001	0.148	0.381	<0.001	0.855	<0.001

TABLE 5. Percentile values (25%, 50%, and 75%) of facility-level and spatially aggregated (HUC-8) consumptive use ($IW - D/W$; unitless) across different water use sectors for all withdrawals and discharges in 2016, as well as warm season (April–October) and cool season (November–December) withdrawals only. Negative values indicate instances where discharge values exceeded withdrawals. Literature values are taken from Shaffer and Runkle (2007).

	All nonthermoelectric			Thermoelectric			Industrial			Commercial			Aquaculture			Municipal		
	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th
Facility level consumptive use																		
Annual	NA	NA	NA	0.01	0.17	0.67	-0.07	0.17	0.67	-0.14	0.19	0.76	0.00	0.00	0.02	NA	NA	NA
Warm season	NA	NA	NA	0.00	0.19	0.61	-0.06	0.22	0.64	-0.09	0.19	0.81	0.00	0.00	0.01	NA	NA	NA
Cool season	NA	NA	NA	0.02	0.15	0.72	-0.06	0.13	0.68	-0.67	0.02	0.43	0.00	0.00	0.03	NA	NA	NA
HUC-8 aggregate consumptive use																		
Annual	-0.02	0.46	0.81	-0.21	0.07	0.59	0.24	0.35	0.91	-0.34	0.52	0.87	0.00	0.00	0.07	-0.30	0.28	0.73
Warm season	0.08	0.55	0.85	-0.15	0.04	0.63	0.23	0.41	0.96	-0.12	0.64	0.86	0.00	0.04	0.19	-0.25	0.35	0.76
Cool season	-0.17	0.18	0.76	-0.32	0.14	0.55	0.23	0.43	0.91	-1.37	0.40	0.78	-0.26	0.00	0.07	-0.32	0.18	0.74
Literature values																		
Annual	NA	NA	NA	0.01	0.02	0.03	0.06	0.10	0.13	0.08	0.10	0.13	NA	NA	NA	0.10	0.13	0.15

the degree to which this is necessary will vary across different states with different permitting programs. The NPDES records in ECHO are not limited to discharge, and in some cases include other parameters that could be important in estimating point-source impacts to water quality. However, our experience suggests that using these records for that purpose may still require QA/QC procedures and some institutional knowledge.

An important step in using discharge and withdrawal records to estimate consumption is matching facilities across the disparate datasets. The matching procedure developed for this purpose identified withdrawal facilities for 123 of the 358 discharging facilities evaluated; although the number of matched facilities was fairly low, these facilities account for a large percentage of the water withdrawn and discharged in the state. This is likely due to the fact that major discharging facilities have been a priority for inclusion in the ECHO discharge records, with many smaller facilities only included following the 2015 NPDES Electronic Reporting Rule (40 CFR 127). The procedure consisted of an automated geographic and text-matching algorithm which was followed up by manual review and validation. The algorithm had a false negative and positive rate of 16% and 48%, respectively. This implies that the algorithm captured the majority of final, true matches, but also included a large portion of erroneous matches that had to be removed through validation. This level of accuracy was suitable for our purposes because the relatively small number facilities allowed us to follow the algorithm with manual validation, and removing false matches is relatively easy compared to identifying missed matches.

Modification of the algorithm could reduce the false positive rate and allow the algorithm to function more effectively as a stand-alone procedure, which would likely be necessary for larger numbers of facilities. For instance, some facilities may have been matched simply because they only had one withdrawing facility within the 10 km search radius (and thus, the closest distance match and closest text match were assured to be the same withdrawing facility). Identifying threshold values of text distance or alternative text-mining procedures could assist in this regard; this could perhaps be accomplished by optimizing threshold values on a smaller subset of training data, and then applying this approach to remaining facilities. There are also situations where it may be desirable to create a more selective procedure (i.e., reduce the false-positive rate) even if it resulted in missing more true matches (an increase in the false-negative rate). For instance, water managers may require a less selective procedure that ensures all true matches are captured, even if it

requires manual removal of false matches. For many research purposes (e.g., analysis of water use behavior), identifying a large number of accurate matches in the algorithm may be sufficient, even if some potential matches are not captured.

While aggregate scale consumption values appeared reasonable, consumption estimates at finer scales (e.g., the facility-level and HUC-10) were more erratic and included many instances of discharges exceeding withdrawals. Some of these instances could be due to meter inaccuracy, as NPDES guidelines require flow measurement to be within 10% of inspector measurements (USEPA 2017). Many of the facilities where discharges exceeded withdrawal reported low withdrawal volumes and had greater instances of missing values, suggesting that these estimates may be the result of data quality issues that appear more prevalent in small water users. Because the volume of water withdrawn and discharged in these facilities is only a small portion of the total state-wide, state-wide estimates of consumptive use shown in Figures 4 and 5 are almost always positive. Within ECHO, major discharging facilities have been the priority with many smaller facilities only incorporated in 2016 following the 2015 NPDES Electronic Reporting Rule (40 CFR 127). The coverage and quality of discharge records included in ECHO are continually being improved, and a summary of the issues present in different states is updated regularly (USEPA 2019). Thus, it may become a more reliable source of discharge records for small facilities in the future. However, the presence of likely erroneous instances of reported discharge exceeding withdrawal presents a risk that using these data for planning and management could result in an overestimation of water availability if reported discharge exceeds true volumes. Because of this, the extent and magnitude of these potential errors should be evaluated prior to direct use of NPDES data for planning or policy purposes.

As has been observed in other studies of water use, the self-reported nature of both withdrawal and discharge volumes is a source of potential error, and data accuracy is likely to vary depending on regulatory requirements (Zhang and Balay 2014). In many contexts, reported water withdrawals and discharges are not actually required to be measured, and can be estimated instead. For instance, thermoelectric generating facilities are allowed to report estimated water use and consumption to the Energy Information Administration if measured data are unavailable (Averyt et al. 2013), and in some instances, these estimates have been shown to be thermodynamically implausible (Diehl and Harris 2014). A key implication of this research and the associated literature is the need for improved accuracy in self-reported water

use data, as well as the transfer and maintenance of data when multiple agencies are involved.

The need for more accurate and unified approaches to water use reporting has been widely acknowledged (Chini and Stillwell 2017; Dunham et al. 2017). However, this will likely require incentives or requirements for more accurate methods of water use quantification (e.g., metering requirements) which may be politically or institutionally challenging. Requiring that water users report the method by which they quantified water use and discharge and including these data along with the reported values could be more feasible and would allow data users to distinguish between measured and reported values. Systematic and thorough approaches for QA/QC of self-reported data could also be used to improve data accuracy, although it would require additional resources on behalf of agencies responsible for data collection. More detail about water-using facilities could also be helpful in contextualizing reported values. Particularly in the industrial and commercial sectors, water use will vary substantially based on the facility processes and characteristics. Industrial water use can include cooling water, food and beverage processing, incorporation into manufactured products, and even irrigation of turfgrass and landscaped areas. Improved facility-level data (for instance, consistent reporting of NAICS codes) would be useful in developing consumption coefficients that were specific to different types of industrial processes.

Our results have several important implications for planning and management. Even in instances where point-source discharge records are already used for water planning and modeling (as is the case in Virginia), these records are not always available for the same temporal period as withdrawals. In these instances, and in modeling water use under future conditions (such as increased withdrawals or climate change), consumptive use coefficients taken from the literature are often necessary to estimate the volume of withdrawn water expected to be returned via discharge. However, our facility-level results suggest that actual consumption in both the industrial and commercial sectors may be higher than literature values, particularly during the summer months. The inter-quartile range of facility-level consumption estimates in our analysis was also greater than the inter-quartile range of values from the literature. This implies that using consumption coefficients from the literature could result in consumption estimates that are lower and less variable than what is occurring in practice.

A key advantage of this approach is that it provides a longitudinal record of consumption that can be used to understand temporal dynamics of consumption, such as trends and seasonality. This is

useful since most existing data on consumption consists of snapshots that do not provide temporal information (Dunham et al. 2017), and even new efforts at collecting water consumption data would not result in a longitudinal record for many years (Ruddell 2018). Understanding seasonal patterns of water consumption is particularly important for drought planning, since withdrawals and consumption will likely peak during summer months when supplies are often lowest (Shaffer 2009; Goldstein 2014; Eldardiry et al. 2016). Shaffer (2009) discussed three potential types of seasonality in water consumption. The first type is constant withdrawal and consumption throughout the year, which would be expected in facilities where production rates and water use do not change seasonally. The second is constant consumption rates but seasonally varying withdrawals, which would be expected in industrial facilities that increase production in certain seasons but do not change water use. The behavior of the thermoelectric and industrial withdrawals shown in Figure 5 exhibit this seasonal behavior, suggesting that higher withdrawals in warm seasons are associated with increased production, rather than outdoor watering observed in other studies (Attaallah 2018). The third type of seasonality is when withdrawals and consumption rates both vary seasonally, which would be expected from a change in water end-uses, such as increased outdoor watering in summer. This behavior is apparent in the total nonthermoelectric withdrawals and municipal withdrawals shown in Figure 4.

The methodology demonstrated here opens up a number of avenues for additional research. A key advantage of this work is its scalability, which could be used to assess water consumption over a range of different users and locations as long as withdrawal data were available. Doing this in a systematic way would be valuable because existing data on water consumption have discrepancies in terms of reporting frequencies, sectors evaluated, and sample sizes, making comparisons between them difficult (Dunham et al. 2017). This approach could also be leveraged to develop regionally and sectoral-specific consumption estimates for industrial and commercial water use that have been highlighted as an important need in understanding economic implications of water use (Rushforth and Ruddell 2018), as well as predictive models that could estimate consumption based on facility characteristics. Multiple studies have demonstrated how water withdrawals vary depending on climate conditions (e.g., Das et al. 2018); having a longitudinal record of water consumption could characterize the relationship between climate and water consumption, improving our understanding of the impacts of drought

(Goldstein 2014) and climate change (Brown et al. 2013; Talati et al. 2016). Finally, additional research could lead to a refined understanding of the advantages, limitations, and differences that arise from different methods for estimating consumption. For instance, the challenges that leakage, infiltration, and stormwater discharge create for estimating municipal consumption based on discharge records are well known (Ducnuigeen et al. 2015; Hastie et al. 2020). Additional research leveraging these records could help to quantify whether discharge increases after rainfall or saturated soil conditions, providing insights into the prevalence or magnitude of these issues. Comparing discharge-based consumption estimates with other approaches (such as evapotranspiration modeling or winter-base rate methods), particularly if measurement data exist to validate consumption estimates, could lead to a better understanding of which consumption estimation methods are most appropriate in different contexts and geographic locations.

CONCLUSIONS

Data on water use are a key component to improving our understanding and management of water supply systems. However, limited data on water consumption create significant uncertainty surrounding estimates of water stress. Historic records of NPDES point-source discharges could help quantify water consumption and complement other consumption estimation methods, but its suitability for this purpose has not been previously evaluated. This research found that a combined process of text matching and geospatial analysis is effective in matching facilities across disparate withdrawal and discharge datasets, allowing for the development of facility-level water consumption estimates. While facility-level estimates of consumptive use varied widely, values for large water users and at larger spatial aggregation scales (HUC-8 and statewide) appeared more consistent with literature values. A key advantage of this approach is the ability to characterize seasonal patterns in water consumption, which is important in estimating water stress during summer months when supplies are lowest and both withdrawals and consumption are high. Our results suggest that median consumption in both the industrial (17%) and commercial (19%) sectors may be higher than consumption coefficients in the literature (median value of 10%). Our results also suggest more variability and seasonality in consumption compared to literature values. While our findings are specific to Virginia, the

approach we developed may be scaled up and applied elsewhere in additional comparative studies, furthering our overall understanding of water consumption.

DATA AVAILABILITY STATEMENT

All data and code generated and used during the study are available online: https://github.com/HARPGROUP/USGS_Consumptive_Use/tree/master/JAWRA_Files.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Geological Survey under Cooperative Agreement No. G17AC00322. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

AUTHOR CONTRIBUTIONS

Mary McCarthy: Data curation; formal analysis; investigation; methodology; writing – original draft. Connor Brogan: Data curation; formal analysis; investigation; methodology; writing – review and editing. Julie Shortridge: Conceptualization; funding acquisition; methodology; project administration; writing – review and editing. Robert Burgholzer: Conceptualization; funding acquisition; project administration; writing – review and editing. Joseph Kleiner: Methodology; resources; writing – review and editing. Durelle Scott: Conceptualization; funding acquisition; writing – review and editing.

LITERATURE CITED

Altunkaynak, A., and T.A. Nigussie. 2017. "Monthly Water Consumption Prediction Using Season Algorithm and Wavelet Transform-Based Models." *Journal of Water Resources Planning and Management* 143: 04017011.

Attaallah, N.A.M. 2018. "Demand Dissagregation for Non-Residential Water Users in the City of Logan, Utah, USA." M.S. thesis, Civil and Environmental Engineering, Utah State University, Logan, UT.

Averyt, K., J. Meldrum, P. Caldwell, G. Sun, S. McNulty, A. Huber-Lee, and N. Madden. 2013. "Sectoral Contributions to Surface Water Stress in the Coterminous United States." *Environmental Research Letters* 8: 035046.

Azadeh, A., N. Neshat, and H. Hamidipour. 2012. "Hybrid Fuzzy Regression-Artificial Neural Network for Improvement of Short-Term Water Consumption Estimation and Forecasting in

Uncertain and Complex Environments: Case of a Large Metropolitan City." *Journal of Water Resources Planning and Management* 138: 71–75.

Brown, T.C., R. Foti, and J.A. Ramirez. 2013. "Projected Freshwater Withdrawals in the United States under a Changing Climate." *Water Resources Research* 49: 1259–76.

Chini, C.M., and A.S. Stillwell. 2017. "Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database." *Journal of Water Resources Planning and Management* 143: 01816005.

Chini, C.M., L.H. Logan, and A.S. Stillwell. 2020. "Grey Water Footprints of U.S. Thermoelectric Power Plants from 2010–2016." *Advances in Water Resources* 145: 103733.

Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, G.H. Taylor, J. Curtis, and P.P. Pasteris. 2008. "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States." *International Journal of Climatology* 28 (15): 2031–64.

Das, P., J. Patskoski, and A. Sankarasubramanian. 2018. "Modeling the Irrigation Withdrawals Over the Coterminous US Using a Hierarchical Modeling Approach." *Water Resources Research* 54: 3769–87.

Diehl, T.H., and M.A. Harris. 2014. *Withdrawal and Consumption of Water by Thermoelectric Power Plants in the United States, 2010*. Washington, D.C.: U.S. Geological Survey.

Dieter, C.A., M.A. Maupin, R.R. Caldwell, M.A. Harris, T.I. Ivahnenko, J.K. Lovelace, N.L. Barber, and K.S. Linsey. 2018. "Estimated Use of Water in the United States in 2015." Report, Reston, VA.

Dubinsky, J., and A.T. Karunanithi. 2017. "Consumptive Water Use Analysis of Upper Rio Grande Basin in Southern Colorado." *Environmental Science & Technology* 51: 4452–60.

Ducnuigee, J., S.N. Ahmed, K.R. Bencala, H.L.N. Moltz, A. Nagel, and C.L. Schultz. 2015. "Interactive Geospatial Analysis Tool for Estimating Watershed-Scale Consumptive Use: Potomac River Basin Case Study." In *Advances in Watershed Science and Assessment*, edited by T. Younos and T.E. Parece, 141–69. Cham: Springer International Publishing.

Dunham, C., H. Fuchs, and H. Stratton. 2017. *Benefits of a National Survey on Water Demand: Existing Data and Reporting Recommendations*. Berkeley, CA: Energy Analysis and Environmental Impacts Division Lawrence Berkeley National Laboratory.

Eldardiry, H., E. Habib, and D.M. Borrok. 2016. "Small-Scale Catchment Analysis of Water Stress in Wet Regions of the U.S.: An Example from Louisiana." *Environmental Research Letters* 11: 124031.

Fanning, J.L., and V.P. Trent. 2009. "Water Use in Georgia by County for 2005; and Water-Use Trends, 1980–2005." USGS Numbered Series, U.S. Geological Survey, Reston, VA. <http://pubs.er.usgs.gov/publication/sir20095002>.

Gates, T.K., L.A. Garcia, R.A. Hemphill, E.D. Morway, and A. Elhaddad. 2012. *Irrigation Practices, Water Consumption, & Return Flows in Colorado's Lower Arkansas River Valley*. Fort Collins, CO: Colorado Water Institute.

Goldstein, R. 2014. *Evaluating Thermoelectric, Agricultural, and Municipal Water Consumption in a National Water Resources Framework*. Palo Alto, CA: Electric Power Research Institute.

Grubert, E., and K.T. Sanders. 2018. "Water Use in the United States Energy System: A National Assessment and Unit Process Inventory of Water Consumption and Withdrawals." *Environmental Science & Technology* 52: 6695–703.

Hastie, A.G., C.M. Chini, and A.S. Stillwell. 2020. "A Mass Balance Approach to Urban Water Analysis Using Multi-Resolution Data." *Journal of Industrial Ecology* 26 (1): 213–24.

JLARC (Joint Legislative Audit and Review Commission). 2016. "Report to the Governor and the General Assembly of Virginia:

- Effectiveness of Virginia's Water Resource Planning and Management." <http://jlarc.virginia.gov/pdfs/reports/Rpt486.pdf>.
- Khanal, S.P., S.M. Cohen, J.E. Macknick, A. Miara, M. Brown, and V.C. Tidwell. 2020. *Evolution of Electric Sector Water Use under Alternative Electricity Futures*. Golden, CO: National Renewable Energy Lab (NREL). <https://doi.org/10.1002/essoar.10502113.1>.
- Li, Y., Z. Tang, C. Liu, and A. Kilic. 2017. "Estimation and Investigation of Consumptive Water Use in Residential Area — Case Cities in Nebraska, U.S.A." *Sustainable Cities and Society* 35: 637–44.
- Mayer, A., S. Mubako, and B.L. Ruddell. 2016. "Developing the Greatest Blue Economy: Water Productivity, Fresh Water Depletion, and Virtual Water Trade in the Great Lakes Basin." *Earth's Future* 4: 282–97.
- Mini, C., T.S. Hogue, and S. Pincetl. 2014. "Estimation of Residential Outdoor Water Use in Los Angeles, California." *Landscape and Urban Planning* 127: 124–35.
- Peer, R.A.M., and K.T. Sanders. 2018. "The Water Consequences of a Transitioning US Power Sector." *Applied Energy* 210: 613–22.
- Ruddell, B.L. 2018. "HESS Opinions: How Should a Future Water Census Address Consumptive Use? (And Where Can We Substitute Withdrawal Data While We Wait?)." *Hydrology and Earth System Sciences* 22: 5551–58.
- Rushforth, R.R., and B.L. Ruddell. 2018. "A Spatially Detailed Blue Water Footprint of the United States Economy." *Hydrology and Earth System Sciences* 22: 3007–32.
- Shaffer, K.H. 2009. "Variations in Withdrawal, Return Flow, and Consumptive Use of Water in Ohio and Indiana, with Selected Data from Wisconsin, 1999–2004." Scientific Investigations Report, U.S. Geological Survey, Washington, D.C.
- Shaffer, K.H. and D.L. Runkle. 2007. "Consumptive Water-Use Coefficients for the Great Lakes Basin and Climatically Similar Areas." Scientific Investigations Report, U.S. Geological Survey, Washington, D.C.
- Talati, S., H. Zhai, G.P. Kyle, M.G. Morgan, P. Patel, and L. Liu. 2016. "Consumptive Water Use from Electricity Generation in the Southwest under Alternative Climate, Technology, and Policy Futures." *Environmental Science & Technology* 50: 12095–104.
- USEPA. 2017. "Chapter 6: Flow Measurement. Interim Revised National Pollutant Discharge Elimination System (NPDES) Compliance Inspection Manual. EPA Publication Number: 305-K-17-001." <https://www.epa.gov/sites/default/files/2017-03/documents/npdesinspect-chapter-06.pdf>.
- USEPA. 2019. "NPDES ERule Dashboard User Guide and Data Caveats|ECHO|US EPA." <https://echo.epa.gov/help/npdes-erule-dashboard/npdes-erule-dashboard-help>.
- USGS. 2017. "Water-Use Data and Research Program Baseline Standards." <https://water.usgs.gov/wausp/wudr/baseline-summary.html>.
- Zhang, Z. and J.W. Balay, 2014. "How Much Is Too Much?: Challenges to Water Withdrawal and Consumptive Use Management." *Journal of Water Resources Planning and Management* 140: 01814001.