Refining Virginia's Water Supply Estimates with State and Federally Maintained Data

STAT 5525 Final Project*

Morgan McCarthy
Biological Systems Engineering
Blacksburg, Virginia
mccartma@vt.edu

CCS CONCEPTS

• **Information systems** → *Information retrieval*; • **Applied computing** → *Physical sciences and engineering*;

KEYWORDS

Consumptive Use, Water Supply, Water Resources, Withdrawal, Discharge

1 INTRODUCTION

Unlike the arid regions in the Southwest, Virginia has historically been a water plentiful state. With rainfall averaging 43 inches per year and a total combined flow of 22.5 billion gallons per day in all freshwater streams, the Commonwealth has been able to meet and exceed water demands for its constituents [25]. Withdrawals from surface water (including rivers, streams, and reservoirs) and groundwater (aquifers) are the two sources that supply Virginia with the water needed to meet the full range of human and environmental purposes [6].

However, the drought of 1999-2002 sparked a major shift in water management perception due to record low flows, several public water systems on the brink of failure, and widespread wildfire [23]. Along with this unexpected drought, the Virginia Department of Environmental Quality (VDEQ) predicts a 32% increase in water demand through 2040, while having gaps in data on ground and surface water use across various economic sectors. This ultimately framed the urgency for improving water resource modeling and management, which lead to the creation of Virginia's first State Water Resources Plan in October of 2015.

The plan was created by the VDEQ and serves as the primary planning mechanism for sustainable water use in Virginia [23]. The effectiveness of this plan was the studied and evaluated by the Joint Legislative Audit and Review Commission (JLARC) in 2016 [6]. One of the mains challenges noted by the JLARC is a lack of knowledge on the impact of consumptive water use on water supply. Where consumption is defined as water that is removed from a water system and not returned [12]. This includes losses through evaporation, transpiration, conveyance, and consumption by animals and humans.

Consumption is especially important for water management because large water users will not necessarily impact supply if the withdrawn water is returned to its source. For example, power producing facilities withdraw large volumes of water to cool their equipment, and then discharge it to a surface water source after once-through or re-circulation [12]. Loss through consumption has the largest impact on water availability; however, current state regulations do not require information on consumptive use. Consequently, this means current assessments of water supply depend solely on the analysis of water withdrawals and not return flows.

Accurate estimates of consumptive water use in Virginia can aid the DEQ's capacity to improve permitting, plan for water scarce scenarios, and model stream flows and water availability. This can be accomplished by combining monthly withdrawal data maintained by the VDEQ and monthly discharge data maintained by the EPA. Specifically, the characterization of consumption over several spatial scales (e.g. State, HUC 6, and County) and water sectors (e.g. Energy, Industrial, Municipal, Commercial, and Agriculture/Irrigation) will be investigated.

1.1 Response to Proposal Comments

In response to the comments provided on the submitted proposal, each task is elaborated on in separate section below. Data Pre-Processing and Model Building is presented before Data Exploration/Visualization to show the results of anomaly detection, data cleaning, and classification techniques. Anomaly detection and classification in this analysis relies heavily on domain expertise and manual exploration of entries. Anomalies are considered to be monthly withdrawal and discharge quantities that exceed expected values in its facility or water sector. Both of the anomaly detection and classification take aspects from multiple concepts discussed throughout the semester. Where anomalies were found and then resolved using a decision tree, and classification uses text mining and semi-supervised rule ordering.

2 DATA DESCRIPTION

2.1 Withdrawal Data

For the state of Virginia, monthly withdrawals from surface and groundwater are regulated by the DEQ Office of Water Supply. Monthly withdrawals (in MGM) are self reported by users under the Virginia Water Withdrawal Permitting Program (VWP), which is submitted once a year to represent withdrawals from the previous

*The data and analysis completed in this report overlap with Morgan McCarthy's Thesis work for her fulfillment of requirements for the degree of Master of Science in Biological Systems Engineering

STAT 5525, December 2018, Blacksburg, Virginia
2018.

calendar year. Reporting is required for those that exceed 10,000 gal/day on average, or agricultural users that exceed 1 MGM in a single month [25]. Approximately half of surface water users are exempt from permitting under grandfather law and other exemptions [6]. Nevertheless, exempt water users are encouraged to voluntarily report their water withdrawals. Ultimately, the monthly withdrawals are stored in the Virginia Water Use Data System (VWUDS), which is not publicly available but can be accessed through cooperative efforts. Additional attributes in the VWUDS include a unique facility ID, withdrawing source ID, owner, facility name, date (yyyy-mm-dd), water withdrawn (MGM), facility status, economic sector, lat/long, and the source type (well vs. surface water).

In terms of data completeness and quality, the self-reported nature of the data introduces noise to the monthly withdrawal data set. As stated before, a vast majority of surface water withdrawals are unpermitted (82%) [6]. For those that are in the database, there are average monthly withdrawals in MGM from 1982-2017 (Table 1). The main sources of noise include missing/incorrect coordinates on a withdrawing point level and the self-categorization of economic sectors for the facilities.

Table 1: Summary of Withdrawal Data

Facilities	Sources	Mean Withdrawal (MGD)	Entries
1,866	5,186	1.76	384,646

2.2 Discharge Data

Section 402 of the Clean Water Act-requires all point source dischargers to obtain National Pollutant Discharge Elimination System (NPDES) permit [8]. In Virginia, NPDES permits are issued by the DEQ, while the Environmental Protection Agency (EPA) ultimately maintains authority over them. This requires the Commonwealth to have regulations comparable to the EPA's standards under their Virginia Discharge Elimination System Program (VPDES). Therefore, discharge monitoring reports are available to the public on the EPA's Enforcement and Compliance History Online System (ECHO). Data can be retrieved using Representational State Transfer (RESTful) web services from the ICIS NPDES database located in ECHO. Where REST uses HTTP as its underlying protocol to perform API-based (Application Program Interface) tasks.

ECHO is as-reported from original source databases and only represents a snapshot of the data as it existed during extraction. Therefore, data is more complete for major permittees, while data for smaller facilities may vary widely. Discharge Monitoring Reports (DMRs) are routinely updated based on when data is entered for a facility. Various flow statistics are available, however, monthly averages (in MGD) are used in this analysis. Unfortunately, complete DMR data is only available starting at 2010. Therefore, analysis of consumptive use is limited to between 2010-2017 (Table 2).

The main source of noise originates from the various permits and outfalls located in each DMR. Even though "flow in-conduit" was queried for, it was found that there was a presence of additional

Table 2: Summary of Discharge Data

Facilities	Sources	Mean Discharge (MGD)	Entries
696	696	22.5	24,896

permits (e.g. general, associated, etc.) and outfalls that track storm water and cumulative in-facility processes. This grossly overestimated discharge estimates. There are also multiple suspected unit conversion errors (most likely human input error) in the reported monthly discharges, which are addressed in data pre-processing methods. Along with this, ECHO does not provide a complete list of the water sector that each facility belongs to. Rather, it generally lists facilities as either municipal or industrial, which is not correct (and is addressed through classification and association analysis techniques).

2.3 Spatial Data frames

The spatial datasets for Virginia Counties and HUC watersheds were obtained from the Census Bureau and the USDA Geospatial Data Gateway of Watershed Boundary Dataset, respectively.

3 DATA PRE-PROCESSING

3.1 Discharge Data: ECHO REST Services

When retrieving discharge monitoring reports from the EPA's ECHO system, there are countless features that can dramatically increase dimensionality, which leads to a larger computational expense. ECHO offers two web services to obtain discharge data;*Facility Search* and *Effluent Charts* [1] [2]. In the *Facility Search* service, there are compliance and enforcement data for over 800,000 regulated facilities. For each facility, there are 299 attributes (known as qcolumns) that include information ranging from unique IDs, location, permit type, etc [1]. In efforts to decrease dimensionality of 800,000 * 299 possible returned entries, only relevant parameters (e.g. State) and qcolumns (e.g. FacilityID, Latitude/Longitude, Permit Type) were selected and included in the request URL. Using R, the customized request URL was downloaded, parsed (to obtain a query ID), and returned to obtain a list of 1,147 regulated facilities with an NPDES permit in Virginia.

With this list, the *Effluent Chart* REST service is used to obtain all tables of permitted effluent limits, releases, and violations over a defined date range for each facility. Each effluent chart provides information on detailed discharge limits, Discharge Monitoring Reports, and NPDES Violations. In each report, there can be up to 300 CWA parameter codes that represent effluent pollutants defined by the EPA and 183 units of measure [2]. It was found that not every facility acquired from the *Facility Search* service has discharge monitoring reports within the specified date range. This is due to the presence of permits that were terminated or expired before the year 2010. Pre-Processing efforts here include querying for *Flow, in conduit or thru treatment plant* (parameter code = 50050), outfalls ending in 001, monthly averages (in MGD), and setting the date range to 01/01/2010-12/31/2017. This results in a data frame of 24,896 monthly average discharges from 696 facilities between 01/01/2010-12/31/2018 (Table 2).

3.2 Withdrawal Data: Missing/Incorrect Coordinates

Raw withdrawal data was obtained from collaborators within Virginia's Department of Environmental Quality's Office of Water Supply. There are approximately 1.35 million entries that represent monthly withdrawals from 4,132 facilities and 7,119 sources from 01/01/1982-12/31/2017. The data is reduced to fit within 01/01/2010-12/31/2017 to match the range of the discharge data (384,646 entries).

The data is further pre-processed by aggregating facilities fitting within spatial references (e.g. HUC 6 watersheds and counties). Unfortunately, the self-reporting nature of the withdrawal data accrues areas of noise. This includes 1,521 instances of missing coordinates (-9999,9999 or NA,NA) or incorrect coordinates that dramatically exceed the physical boundaries of the State (-83 to -75 for Longitude, 36 to 39.5 for Latitude). Google Maps and the use of associated information (County Name, Facility Name, Source Name, etc.) were used to locate the exact coordinates of the missing withdrawal locations. Latitudes and Longitudes provided by Google Maps are recorded with an accuracy of six digits. If exact coordinates could not be found (due to lack of sufficient information), the center of the reported city or county was used. This allows for spatial analysis to be performed on the withdrawal data.

4 MODEL BUILDING

4.1 Flagging Anomalous Discharge Data

4.1.1 Flagging Criteria. Once the discharge and withdrawal data were collected and pre-processed, aggregated monthly flows (MGD) over time were plotted to visualize anomalous areas that may require quality assurance and control measures (QA/QC). It was found that the withdrawal data is vetted by VDEQ staff and does not display any blatant anomalies; while the discharge data have instances of effluent values (in MGD) that are orders of magnitude higher than expected. Domain experts, Robert Burgholzer and Joseph Kleiner from the VDEQ's Office of Water Supply, were consulted to identify typical ranges for both the withdrawal and discharge data. After deliberation, it is suspected that unit conversion errors from gallons per day to million gallons per day, are the main source of the peculiar discharge values. Therefore, criteria were created to automate flagging suspicious/anomalous entries in the discharge data.

Criteria were generated using brute force association analysis with guidance from domain expertise. Entries were sorted from high to low reported effluent values and manually examined for common patterns. It was found that anomalous entries typically have effluent values that exceed 100 times its median outfall discharge, as well as 1,000,000 times. Records with effluent values that exceed its facility's design flow, or report a design flow equal to zero are also noted as suspicious. Using these four criteria/association rules, 1,740 out of the 24,896 entries were flagged as anomalous.

4.1.2 Resolve Anomalies with Weighted Decision Tree. The flagged entries were then subsetted and examined to indicate which criteria best detects an anomaly (or which does not). This required domain expertise and manual exploration of the discharge records. The top ten flagged entries with the highest difference between its reported effluent and the median flow for its respective outfall were investigated. It became clear that entries with multiple flags (>1) were the most suspicious, especially, with the presence of exceeding facility design flow. Since each criteria has its own level of urgency, a decision tree was created to decide how to resolve the suspicious entries.

Instead of treating each flagging criteria as a node on a decision tree, a singular weighted flag is used as the splitting attribute. Each criteria is assigned a subjective weight from 0 to 1 (where 1 indicates the most error potential), and summed to equal an overall weighted value (Table 3). This continuous attribute is then used to split the flagged entries into two classes: high concern and low concern. If the weighted flag value is greater than or equal to 1, it is labeled as high concern. Otherwise, it is labeled as a low concern. Of the 24,896 discharge entries, 1,740 were flagged, and approximately 61% of the flags were labeled as a high concern and replaced with its median outfall discharge (Figure 1). The remaining 39%, labeled as low concerns, were left intact.

4.2 Classification of Water Sectors

As described in the *Data Description* section, the classification of economic sectors for each facility is another large source of noise in both the withdrawal and discharge data. The self-reporting nature of the data causes a disconnect in the definitions of each sector. Specifically, all of the facilities in the discharge data were categorized as Industrial (24%), Municipal (72%), or NA (4%). This general classification assumes energy facilities are in the industrial sector, which skews results when comparing quantities of water between clustered withdrawal and discharge data. The classifications of facilities in the withdrawal data is more realistic, but is overwhelming with 13 different sectors. Therefore, a semi-supervised learning approach involving text mining and rule-based classification is used to sort facilities into one of five water sectors: Agriculture/Irrigation, Commercial, Energy, Industrial, or Municipal.

In this analysis, power stations, fossil power, and nuclear power facilities are assumed to be in the Energy Sector [12]. Hydropower facilities are not included, due to a change in reporting requirements on a State level. Municipal facilities include residential areas, sewage waste, waste water treatment facilities, and sewage treatment facilities. Commercial facilities are buildings that are not considered to be residential [24]. This includes golf courses, hotels, schools, prisons, etc. Agricultural/Irrigation facilities include farms, dairy farms, nurseries, fish hatcheries, and aquaculture. The vast majority of these facilities will not have a corresponding discharge permit associated with them. Most studies assume water used for agriculture and irrigation of crops is completely lost through evapotranspiration [11]. The industrial sector then covers manufacturing, mining, and industrial waste treatment from facilities.

Table 3: Subjective Weights Assigned to Flagging Criteria to Represent Node on Decision Tree

Design Flow = 0	DMR Value > 100*Median	DMR Value > 106*Median	DMR Value > Design Flow	Weighted Flag
0.25	0.4	0.4	1	$\sum(Flags)$

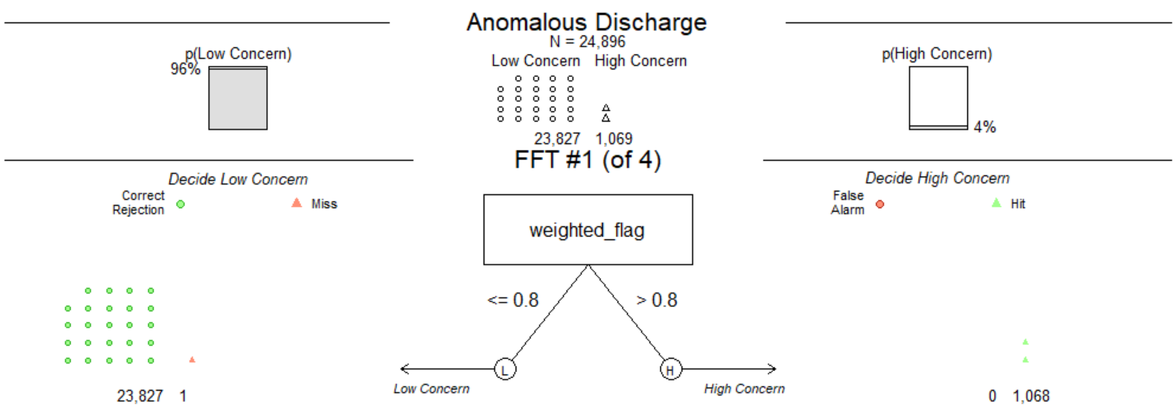


Figure 1: Decision Tree to Classify Anomalous DMR Entries

Using guidance from domain experts and a small training set from the withdrawal data, text mining techniques were used to find the most frequent terms in facility names by water sector. This approach is considered to be semi-supervised, because it uses a smaller training set of facilities where the labeled water sector is considered to be "correct", to classify the unlabeled records. The challenge in this approach is finding unique terms in facility names that encapsulate a water sector, but are not prevalent in other sectors. For example, the term "INC." is present in a vast majority of Industrial facilities, however it is also present in commercial and agricultural facilities. Therefore, INC. and other legal terms (e.g. LLC. and LLP.) are not a true indication of facility type. Using facility names as a basis of classification has its limits, and those limits should be taken into consideration. For instance, there are many withdrawing agricultural facilities that simply include 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 in efforts to decrease training error.

Rules were created using the *grep()* function in R, which searches for terms that match the defined character vector. In essence, the classification rule for the Energy Sector includes the following terms that act as argument patterns:

```
if(length(grep('\\bPOWER\\b',Wateruse$Facility_Name[i]))>0|
     length(grep('\\bPOWER_STATION\\b',Wateruse$Facility_Name[i]))>0|
     length(grep('\\bCOMBUSTION\\b',Wateruse$Facility_Name[i]))>0|
     length(grep('\\bCLINCH_RIVER_PLANT\\b',Wateruse$Facility_Name[i]))>0|
     length(grep('\\bENERGY\\b',Wateruse$Facility_Name[i]))>0){
  Wateruse$Reclass_Use_Type[i]<-'Energy'
```

Since terms in facility names may trigger more than one rule, all rules are ordered according to their priority, with Energy on top and Industrial near the bottom. Energy facilities withdraw and discharge the most water, therefore their correct classification is the most important as to not skew the distribution of water quantities in the other sectors. Industrial is at the bottom of the rule set due to its lack of unique terms relative to the other sectors. If none of the rules are fired, the class is ultimately assigned to its original classification (which can be referred to the default class).

5 DATA EXPLORATION/VISUALIZATION

After both the monthly and discharge data were pre-processed, cleaned, and classified in terms of water sectors, they were examined at different spatial scales. Figure 2 shows statewide discharges (in blue) versus withdrawals (in red) through time in MGD. This figure includes all water sectors, where the energy sector is making up the majority of the summed withdrawals and discharges over time (>80%) [25]. Consumption is in light grey and ranges from 0 to 100% on the right axis. There are apparent seasonal variations, where higher consumption and water use rates increase in the Summer months and dip in the winter.

Figure 3 then shows how characterization of consumption differ when applied to a spatial reference. Here, the energy sector (on top) is separated from the other sectors (on bottom) to show the distribution of power generating facilities, which are mainly concentrated along the James River and near Richmond. A key point here is that there are likely errors in the data that are canceling each other out in the statewide estimates, that become visible at finer spatial scales. For instance, looking at energy it appears that there are unmatched facilities in southwest that are making consumption look really high (0.75-1.0). There is then suspicious behavior in the York/Rappahanock (Energy figure) and in the Chowan (Non-Energy figure) making consumption negative. This may indicate a potential error in the discharge data or a consequence of using self-reported

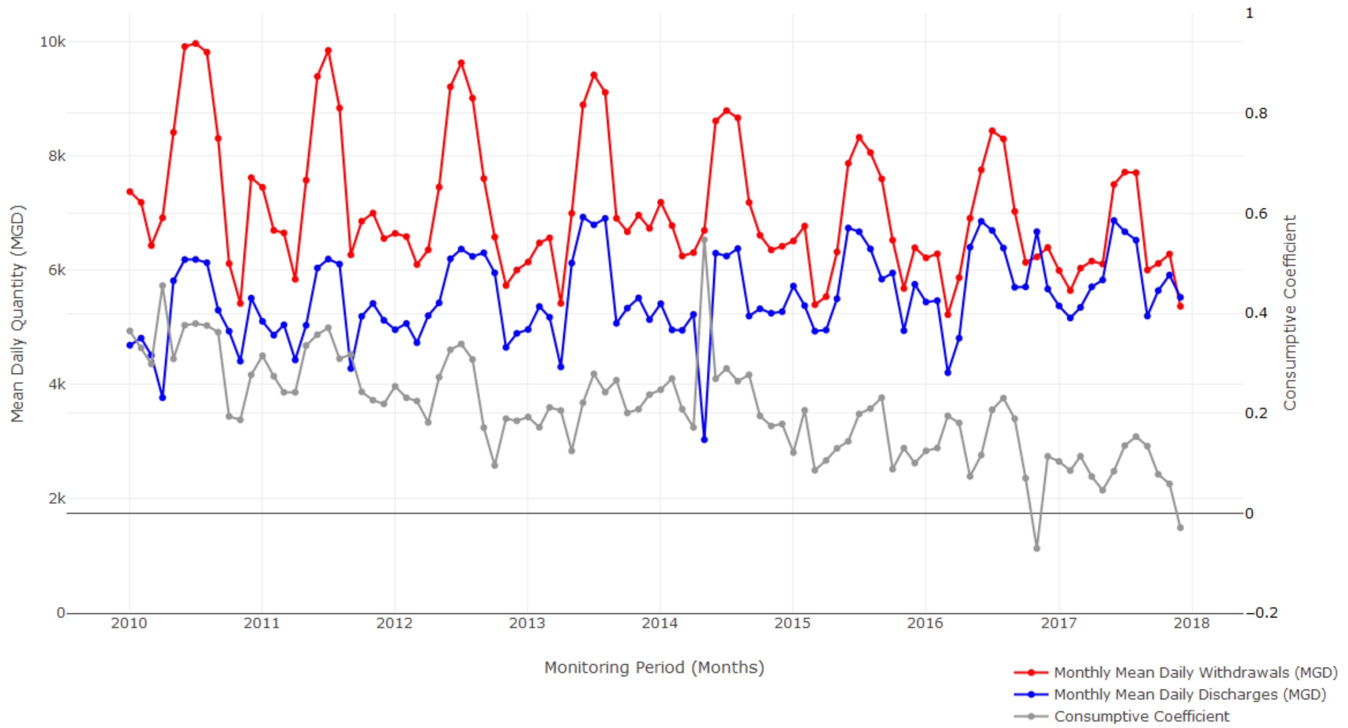


Figure 2: Statewide Mean Daily Withdrawals vs. Discharges (MGD) in Virginia

coordinates.

As the spatial scale becomes finer, more is revealed. In Figure 4, specific regions and facilities are highlighted if consumption is either very negative or close to 1. Despite the noisy areas, emerging patterns can be extracted from this figure. For example, counties with high consumption along the eastern shore and the Shenandoah region, correspond to the dense agriculture that is located in those regions. Another interesting pattern is the distribution of negative consumptive counties in the non-energy sectors. They seem to follow along major river ways, perhaps suggesting geophysical boundaries may need to be considered during consumptive analysis as well.

6 MODEL EVALUATION

In order to evaluate the results of this analysis, withdrawal results are compared to yearly reports drafted by the Virginia Department of Environmental Quality and consumptive estimates are compared to typical coefficients from climatically similar areas. Table 4 shows the comparison of statewide withdrawals in MGD calculated during this analysis and values reported in the annual water resources status reports to Virginia's governor. All values are within a 3% error, which is within reason but suggests results should communicated with domain experts.

In terms of consumption estimates, Table 5 shows the calculated statewide coefficients for each water use sector form 2010 to 2017, and then an overall average. Consumption coefficients from 15 climatically similar areas and studies were then collected and averaged, to evaluate the results from the analysis performed in this report (Table 6).

Table 4: Comparison of Reported Statewide Withdrawals (MGD) per Year [25].

	2010	2011	2012	2013	2014	2015	2016	2017
Results	7,794	7,418	7,246	7,195	7,160	6,757	6,741	6,389
VDEQ	7,621	7,559	7,100	7,047	6,984	6,558	6,613	6,313
Error	2.27%	1.87%	2.06%	2.10%	2.52%	3.03%	1.94%	1.20%

In relation to the results in Table 5, consumption coefficients for Agriculture/Irrigation and Municipal are relatively close to estimates from climatically similar areas (Table 6). Consumption estimates for the Commercial sector are approximately a factor of ten higher than expected values. This is most likely due to a low number of facilities being classified as commercial in the discharge data (116 facilities), relative to the withdrawal data (427 facilities). This is a strong indication that the classification model used may not be suitable for a high level of specificity.

Consumption rates in the energy sector show a decrease from about 30% to 1% from 2010 to 2017. Even though the overall average consumption rate for the sector is 15%, this emerging decline may suggest the effect of legislation and the switch to once-through cooling for power producing facilities. As for the Industrial sector,

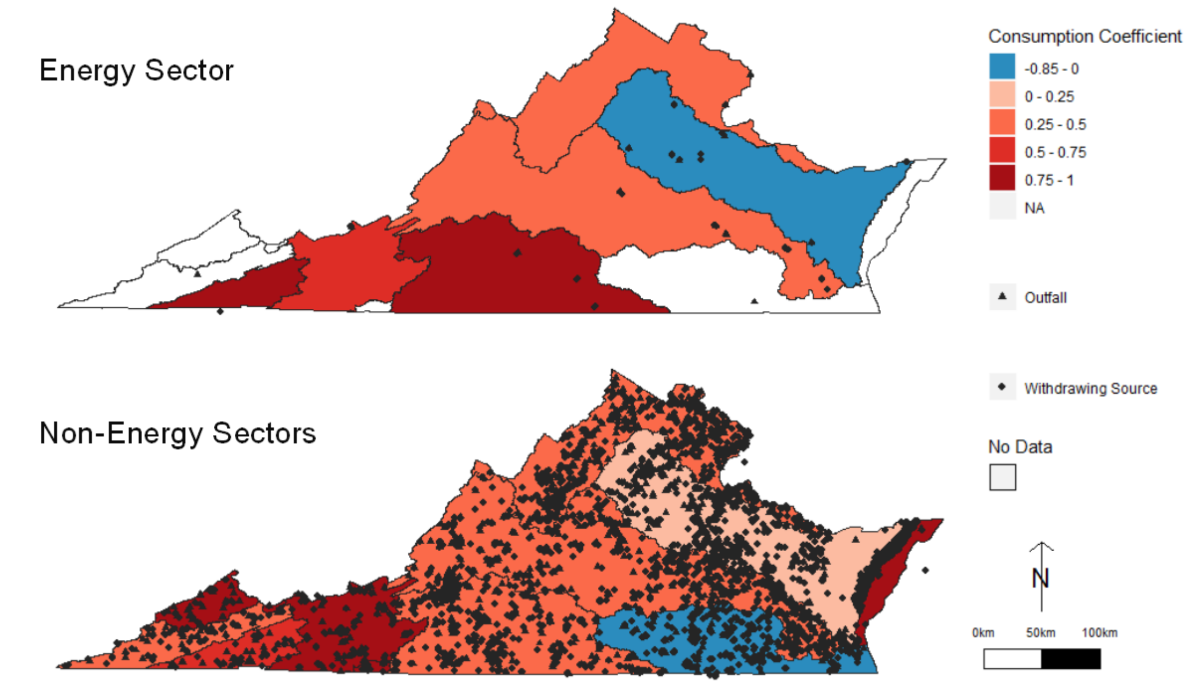


Figure 3: Long Term Average (2010-2017) Consumption in HUC6 Watersheds in Virginia (Energy vs. Non-Energy)

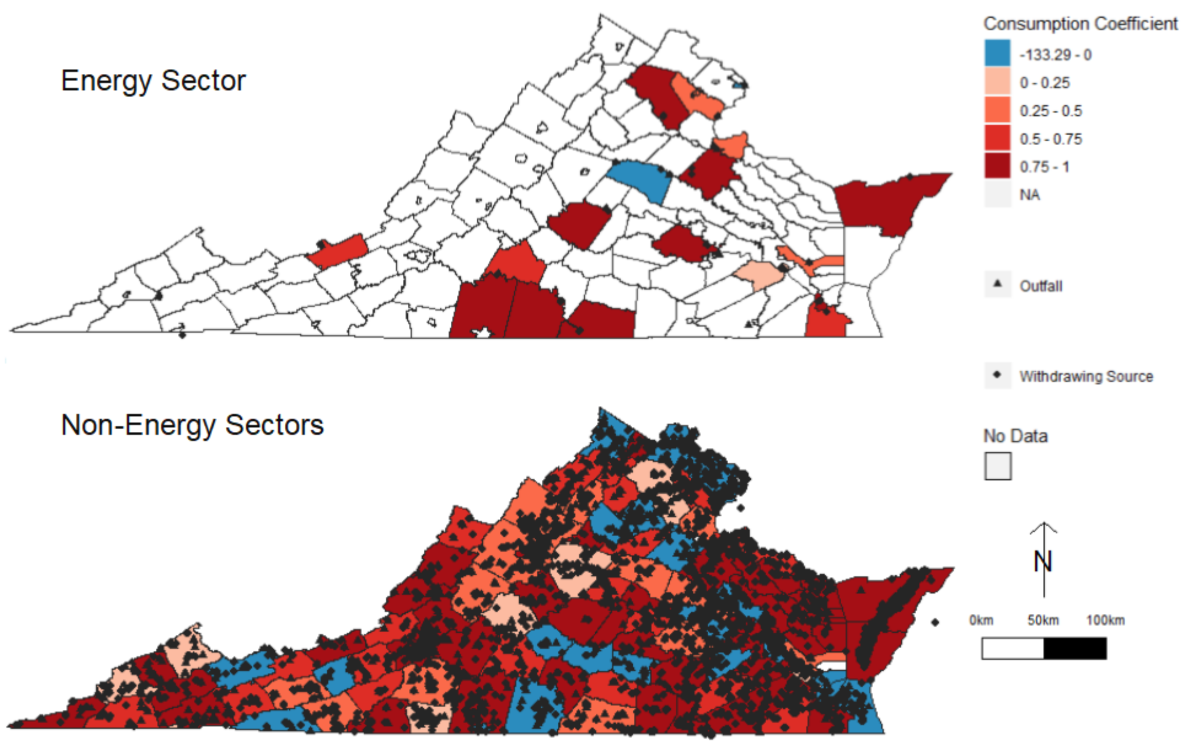


Figure 4: Long Term Average (2010-2017) Consumption in Virginia Counties (Energy vs. Non-Energy)

Table 5: Statewide Consumption Coefficients for Classified Sectors (2010-2017)

Sector	2010	2011	2012	2013	2014	2015	2016	2017	Average
Agriculture/Irrigation	1.00	1.00	1.00	1.00	1.00	0.78	0.79	0.77	**0.92**
Commercial	0.99	0.91	0.91	0.87	0.89	0.79	0.78	0.65	**0.85**
Energy	0.26	0.24	0.19	0.17	0.19	0.08	0.08	0.01	**0.15**
Industrial	0.69	0.69	0.66	0.59	0.60	0.55	0.53	0.49	**0.60**
Municipal	0.45	0.38	0.28	0.21	0.23	0.24	0.10	0.26	**0.27**

Table 6: Consumption Coefficients from Climatically Similar Areas

Geographical Area	Agriculture/Irrigation	Commercial	Energy	Industrial	Municipal
Ohio [5]	-	-	0.1	6.0	20
Delaware [15]	-	-	1	4	10
Massachusetts [14]	-	10	-	10	15
Great Lakes [19]	75	-	0.385	7.85	20
Connecticut [20]	-	-	-	-	20
Tennessee [16]	100	11	0.5	11	24
Wisconsin [4]	100	-	1	10	-
Great Lakes [7]	-	-	2	10	-
Europe [10]	80	-	5	20	20
Rhode Island [13] 100	8	-	4	15	
Illinois [17] 80	9.6	-	32.8	28.4	
Pennsylvania [15]	100	4.6	4	8.4	37
Rhode Island [21]	100	10	-	10	15
Pennsylvania [9]	100	10	1.23	8.1	10
World [3]	70	-	-	11	12
	90.5	**9.03**	**1.7**	**10.9**	**18.95**

consumption appears to decrease from 70% in 2010, but lingers around 50% throughout the years. Whether this is a product of noisy data, or a revelation about industrial water use in the state, requires further consultation with VDEQ domain experts.

7 REAL-WORLD INSIGHTS

Responsibility for managing water resources in Virginia is divided among the State Water Commission, the State Water Control Board, and the VDEQ Office of Water Supply. These three entities coordinate legislative activity, develop regulations, and implement water resource planning, respectively [6]. The Virginia DEQ (VDEQ) collects and analyzes surface and ground water usage data to understand the impact of permitted withdrawals on statewide supply [6]. They direct water supply management for the Commonwealth through three programs including groundwater characterization, water supply planning, and water withdrawal permitting [18].

Water Resource planning is designed as a statewide partnership, which enables regional and local entities to identify their specific future water needs with support and oversight from the State. There are 48 local plans that the DEQ reviewed (as of 2015) to aid in the development of a statewide plan. The first State Water Resources Plan was published in October 2015 and serves as the primary planning mechanism for Virginia sustainable water use through the year 2040 [6]. Water use is considered sustainable when withdrawals can indefinitely meet the full range of human and environmental water needs without resulting in adverse effects on other users and the environment [6]. The plan was created by VDEQ staff and is intended to assist state and local policy makers to develop more informed water supply and management policies. It includes information regarding existing water use, sources of supply, population and water demand projections, anticipated deficits, potential new supply sources, and current efforts supporting efficient water use [22].

This plan and future plans will benefit from this analysis because discharge data has not been considered when analyzing water use in the state of Virginia. This analysis serves to highlight areas that require more attention, whether that's reaching out to users for better reporting or large users that are consuming more than they are returning. In particular, industrial and energy facilities could be focused on because they use the most water and have higher than typical consumptive use rates compared to climatically similar areas. There's also the question of which spatial scale should be used to inform management decisions. Water supply management requires a level of specificity that aids stakeholders to move towards more sustainable practices, but isn't too specific that it isn't applicable to everyone. With this being said, policy makers can leverage the figures created in this analysis to influence more specific water supply planning and management techniques than what is currently available.

8 LESSONS LEARNED

This analysis has made it clear that self-reported "big"-data can be very noisy. Unlike research where the individual is collecting data and can pinpoint what errors look like, data analytics is almost like a guessing game. In terms of this project, domain experts were highly relied on to understand typical ranges and values of water use throughout the years. Moving forward with this research, I want to move towards analyzing consumption on a facility level.

I have created a script that uses nearest neighbor logic and fuzzy pattern matching to locate links between facilities in the monthly discharge and withdrawal data sets. This in itself is tricky because I have learned that using an attribute like facility name can only get you so far in analysis. There is also a certain level of trust with the self reported coordinates provided by water users. That is why it was not included in this particular report. However, I believe matching facilities and mapping consumptive use on a facility level will be useful in the investigation of whether errors stem from missing/unmatched facilities or from errors in facility-level data. In the end, this project has taught me the meaning of patience, documentation, and to stray away from tempting rabbit holes of trying to perfect code and analysis.

REFERENCES

[1] 2018. ECHO CWA REST Services: Facility Search - Water. https://echo.epa.gov/tools/web-services/facility-search-water
[2] 2018. ECHO EFF REST Services: Effluent Charts-Discharge Monitoring Reports. https://echo.epa.gov/tools/web-services/effluent-charts
[3] Shiklomanov A. 2003. *World Water Resources at the Beginning of the Twenty-First Century.* The Press Syndicate of The University of Cambridge, Cambridge. 25 pages.
[4] James T. Krohelski B.R. Ellefson, Kraig S. Rury. 1987. *Water Use in Wisconsin.* Technical Report. USGS. https://doi.org/87-699
[5] E. Downey Brill, Glenn E. Stout, Robert W. Fuessle, Randoplh M Lyon, and Keith E. Wojnarowski. 1977. *Issues Realted to Water Allocation in the Lower Ohio River Basin.* Technical Report. Office of Energy, Minerals and Industry USEPA. 87 pages. https://doi.org/USEPAR-804821
[6] Joint Legislative Audit Comissions and Review. 2017. *Effectiveness of Virginia ' s Water Resource Planning and Management.* Technical Report October 2016. Joint Legislative Audit and Review Commissions, Richmond.
[7] Iternational Joint Commission. 1985. *Great Lakes Diversions and Consumptive Uses.* Technical Report. Interantional Joint Commission. 96 pages.
[8] Melanie D. Davenport. 2014. *VPDES Permit Manual.* Technical Report. Virginia Department of Environmental Quality, Richmond. 524 pages. file:///G:/MyDrive/GRA/LiteratureReviews&Methodologies/VPDES_Program.pdf%0D
[9] H O W Do, W E Use, H O W Much, Water Is, I N Pennsylvania, and W E Get. 2000. *Estimated Water Withdrawals and Use in Pennsylvania , 1995.* Technical Report February. USGS. 4 pages. https://doi.org/USGSFactSheet174-99
[10] European Environment Agency. 2005. *European environment outlook.* Technical Report 4. European Environment Agency, Copenhagen. 92 pages. https://doi.org/ISSN1725-9177
[11] Julia L. Fanning and Victoria P. Trent. 2009. *Water Use in Georgia by County for 2005 and Water-Use Trends , 1980 – 2005.* Technical Report. USGS, Reston. 192 pages. https://doi.org/ScientificInvestigationReport2009-5002
[12] R. Goldstein. 2013. *Evaluating Thermoelectric, Agricultural, and Municipal Water Consumption in a National Water Resources Framework.* Technical Report. Electric Power Research Institute, Palo Alto. https://doi.org/3002001154
[13] By M A Horn, P A Croft, and Lisa Bratton. 1994. *ESTIMATION OF WATER WITHDRAWAL AND DISTRIBUTION , WATER USE , AND WASTEWATER COLLECTION AND RETURN FLOW IN CUMBERLAND , RHODE ISLAND , 1988.* Technical Report. USGS, Providence. https://doi.org/Water-ResourcesInvestigationsReport93-4023
[14] Ma Horn. 2000. Method for Estimating Water Use and Interbasin Transfers of Freshwater and Wastewater in an Urbanized Basin. 1 (2000). http://pubs.usgs.gov/wri/wri994287/
[15] Susan S. Hutson, Kristin S. Linsey, Russell A. Ludlow, Betzaida Reyes, and Jennifer L. Shourds. 2015. *Estimated Use of Water in the Delaware River Basin in Delaware, New Jersey, New York, and Pennsylvania, 2010.* Technical Report. National Water Census. https://doi.org/ScientificInvestigationsReport2015-5142
[16] Jaala M. Brooks John A. Robinson. 2010. *Public Water-Supply Systems and Associated Water Use in Tennessee, 2005.* Technical Report. US Deaprtment of the Interior and USGS.
[17] John K. LaTour. 1991. *Determination of Water Use in Rockford and Kankakee Areas, Illinois.* Technical Report. USGS, Urbana, Illinois. https://doi.org/90-4166
[18] John A Kitzhaber and Marcus Wah. 2017. *Status of Virginia's Water Resources: A Report on Virginia's Water Resources Management Activities (2017).* Technical Report 503. Virginia Department of Environmental Quality, Richmond.
[19] Christine Manninen and Roger Gauthier. 1999. *Living with the Lakes.* Technical Report. US Army Corps of Engineers Detroit District. 40 pages. http://www.watershedcouncil.org/waterresources/greatlakes/threats-to-the-great-lakes/great-lakes-water-levels/other-opinions-on-low-water-levels/files/livingwiththelakes.pdf
[20] John R. Mullaney. 2004. *Water Use , Ground-Water Recharge and Availability , and Quality of Water in the Greenwich Area , Fairfield County , Connecticut and Westchester County , New York , 2000 – 2002 Water Use , Ground-Water Recharge and Availability , and Quality of Water in t.* Technical Report. USGS, East Hartford. 71 pages. https://doi.org/Water-ResourcesInvestigationsReport03-4300
[21] Mark T Nimiroski, Emily C Wild, Gale A Norton, and U S Geological Survey. 2005. *Water Use and Availability in the Woonasquatucket and Moshassuck River Basins , North-Central Rhode Island Scientific Investigations Report 2005-5031.* Technical Report. USGS, Reston. https://doi.org/ScientificInvestigationsReport2005-5031
[22] Virginia Department of Environmental Quality. 2015. *Status of Virginia's Water Resources: A Report on Virginia's Water Resources Management Activities (2015).* Technical Report. Virginia Department of Environmental Quality, Richmond. 45 pages.
[23] Virginia Department of Environmental Quality. 2016. *Status of Virginia's Water Resources: A Report on Virginia's Water Resources Management Activities (2016).* Technical Report. Virginia Department of Environmental Quality, Richmond. 50 pages.
[24] Kimberley H Shaffer and Donna L Runkle. 2007. Consumptive Water-Use Coefficients for the Great Lakes Basin and climatically similar areas. (2007), 191.
[25] Virginia Department of Environmental Quality. 2018. *Status of Virginia's Water Resources: A Report on Virginia's Water Resources Management Activities (2018).* Technical Report. Virginia Department of Environmental Quality, Richmond. 49 pages.

Consumptive Water Use

Refining State Water Supply Estimates with Discharge and Withdrawal Data

Morgan McCarthy

Stat 5525: Final Project
12/05/2018

1999-2002

Record Low Mean Daily Flow Levels[2]

32%

Predicted Net Increase in Mean Daily Water Demand[1]

Data Gaps

In Ground and Surface Water Use across Certain Sectors[2]

Urgency for Improved Water Resource Modeling & Management in Virginia

Data Description

	Withdrawals	Discharges/Return Flows
Agency Source	DEQ Virginia Department of Environmental Quality	ECHO Enforcement and Compliance History Online
Data Source	Users under the Virginia Water Withdrawal Permitting Program	Users under the Virginia Pollution Discharge Elimination System (VPDES) Program
Reporting Requirements	Withdrawals > 10,000 gal/day[3] Agricultural Withdrawals > 1 MGM[3]	All Point source discharges to waters of U.S[4]
Statistics	696 Facilities 696 Outfalls Mean Discharge: 22.5 MGD Entries: 24, 896	1,866 Facilities 5,186 Outfalls Mean Withdrawal: 1.76 MGD Entries: 384,646

Pre-Processing

Use EPA's ECHO REST Services

Reduce 800,000 * 299 possible entries to 1,147 VPDES permitted facilities.

Query for specific parameter codes to reduce computational

Query using RESTful Services

Subset Withdrawal Data

Subsetted Withdrawals to match range of monthly discharges

1.35 Million entries -> 384,646

1982-2017 -> 2010-2017

1,521 Flagged Coordinates in VWUDS Database

Google Maps to Locate Withdrawing Sources

If facility could not be found, City Coordinates were used

Missing/Incorrect Coordinates

Model Building

Flag Anomalous Discharge Data

Flagging Criteria

> 100 * Median Outfall Discharge
> 1M * Median Outfall Discharge
> Design Flow
Design Flow = 0

Resolve Anomalies

Weighted Decision Tree

If weighted flag ≥ 1→ Replace with Median Discharge

If weighted flag ≤ 1→ No Change

Classify Water Use Sectors

Text Mining & Association Analysis

Aggregate into Common Sectors

Use Most Frequent Terms in Facility Names to Classify Unknown Records

Model Building: Anomaly Detection

Criteria were generated using brute force association analysis with guidance from domain expertise. Suspicious entries were sorted from high to low reported effluent values and manually examined for common patterns. 1,740 out of the 24,896 entries were flagged as anomalous.

Design Flow = 0	DMR Value > 100*Median	DMR Value > 10^6*Median	DMR Value > Design Flow	Weighted Flag	Resolution
0.25	0.4	0.4	1	Σ(Flags)	if weighted flag>=1 high concern, else low concern

After investigating the top ten flagged entries, it became clear some flagging criteria are better at detecting anomalous entries than others. Therefore, weights were assigned and resolutions were determined from an overall summed weighted flag.

Model Building: Classification

Text Mining Techniques were used to find most frequent terms in facility names in small training set sampled from the withdrawal data.

Municipal:
1. Service
2. Area
3. **WTP**
4. Water
5. System
6. Town
7. **Park**
8. County
9. Public
10. Estates
11. Home
12. Lake
13. Creek
14. River
15. Mobile
16. Point
17. City

Commercial:
1. Golf
2. Club
3. Course
4. Country
5. **Park**
6. Resort
7. Center
8. School
9. **Company**
10. **INC**
11. **Farms**
12. **Hills**

Industrial:
1. **Plant**
2. **INC**
3. Quarry
4. Prep
5. **WTP**
6. Mine
7. **Company**
8. Industries
9. PLT

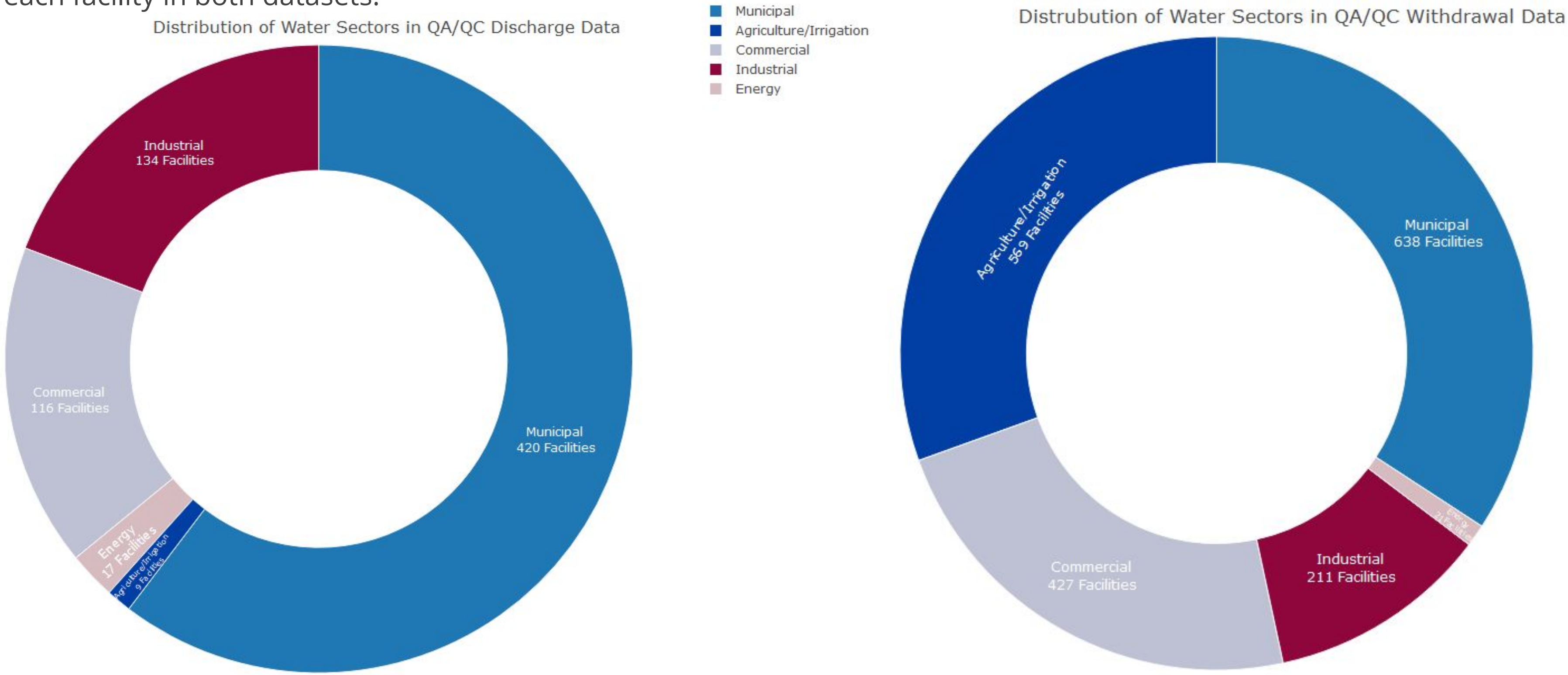
Energy:
1. Power
2. Station
3. **Plant**
4. Combustion
5. Turbine
6. Energy

Agriculture/Irrigation:
1. Farm
2. Farms
3. **INC**
4. Dairy
5. Nursery
6. **LLC**
7. Hill
8. Hog
9. Production
10. Nurseries
11. Bain

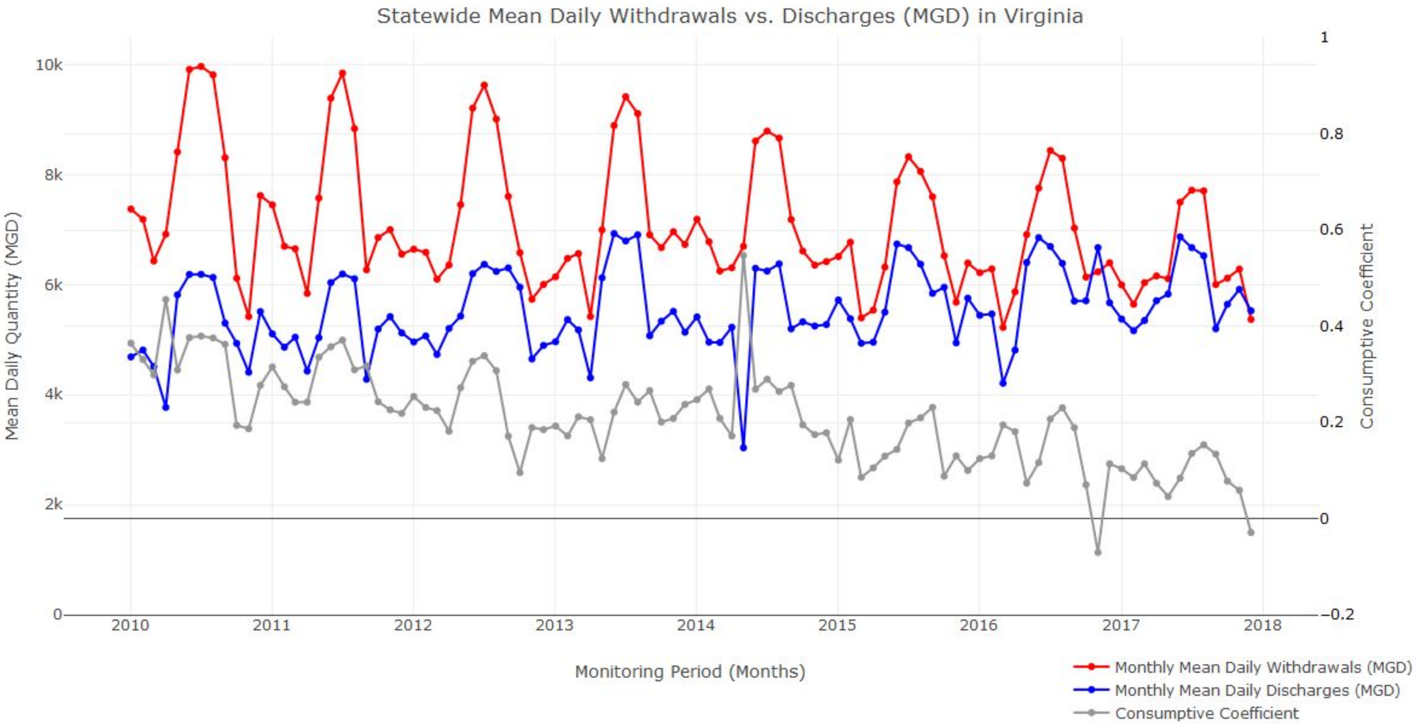
Challenge: finding unique words that capture the water sector but don't dip into the others.

Model Building: Classification

With these terms, rules were created and ordered to classify the economic sector for each facility in both datasets.



Data Exploration/Visualization



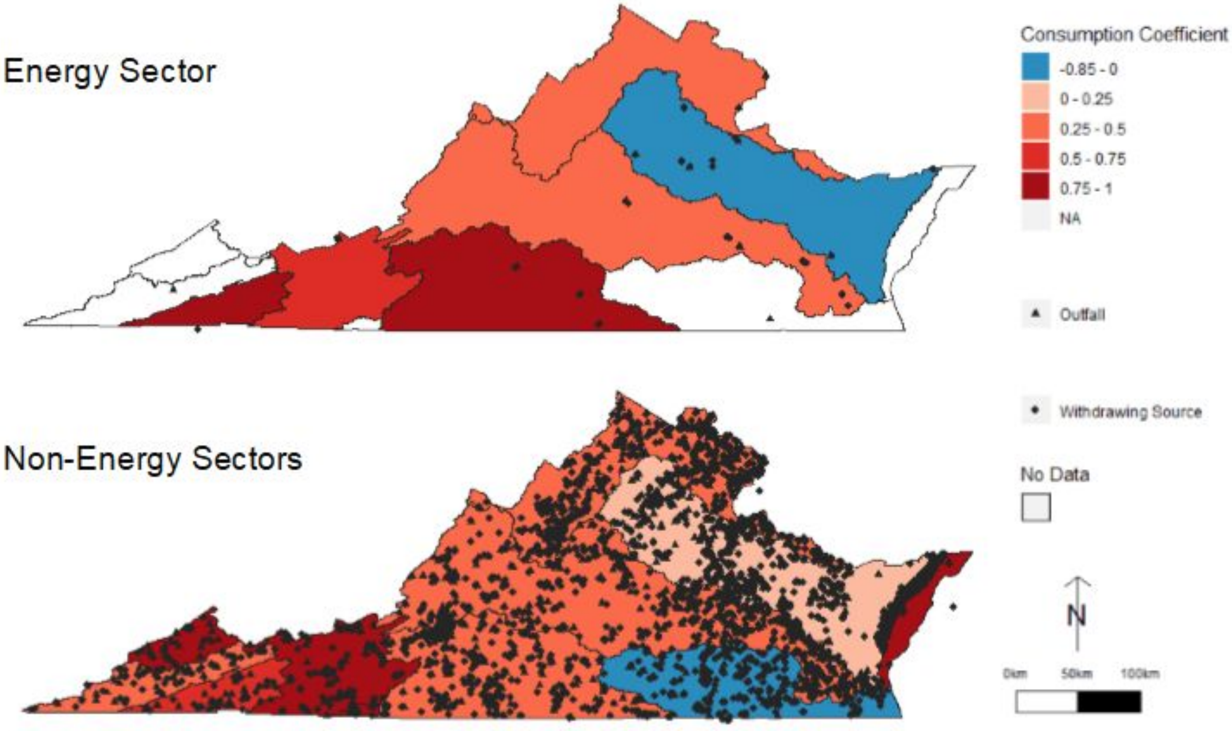
Withdrawals | Discharges

Consumptive Water Use

$$\frac{\sum_{i=1}^{n}(Withdrawals) - \sum_{i=1}^{n}(Discharges)}{\sum_{i=1}^{n}(Withdrawals)}$$

- Seasonal Variation in withdrawals, discharges, and consumption over time.
- Gradual decrease in consumption rates.

Data Exploration/Visualization

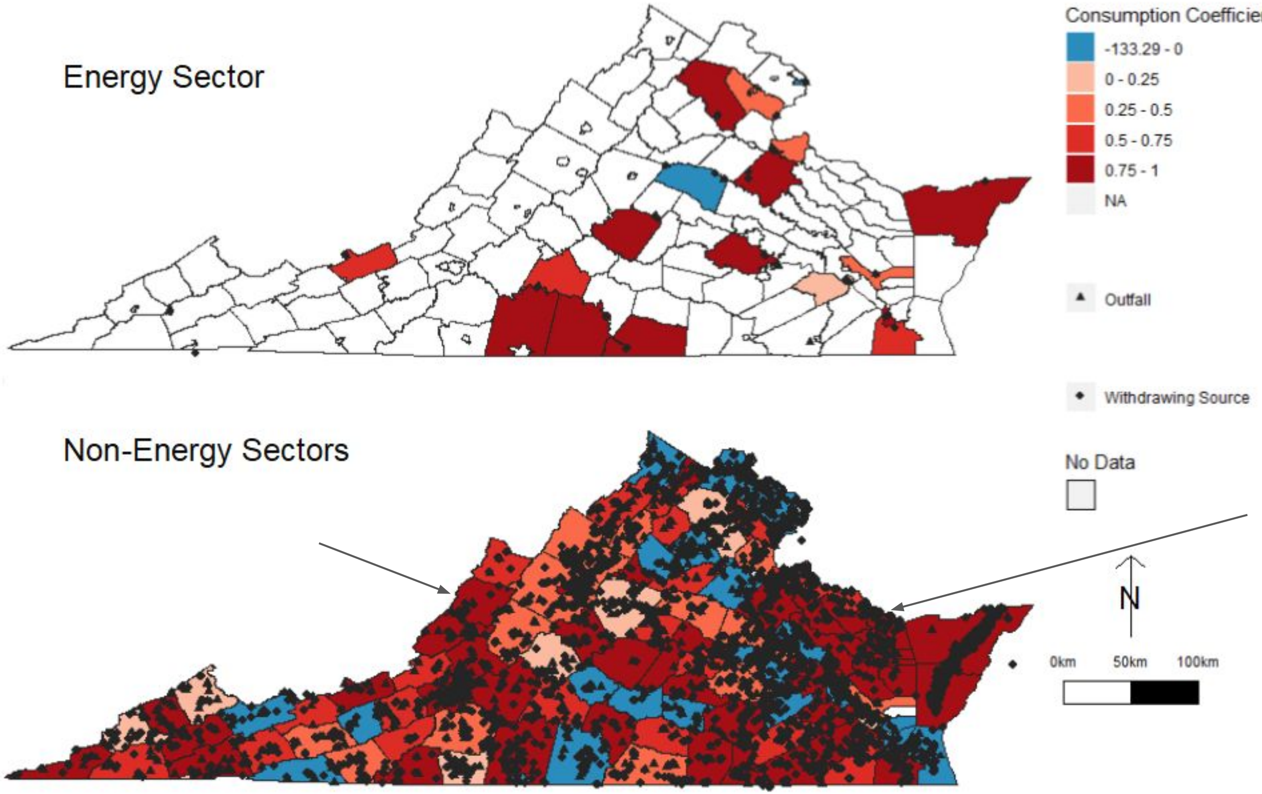


- Errors in the data that are canceling each other out in the statewide estimates, that become visible at finer spatial scales

-Unmatched energy facilities lead to perception of high consumption in Southwest VA

- Distribution of permitted and reporting facilities are located near highly populated areas of Richmond and Northern Virginia.

Data Exploration/Visualization



- Finer scale pinpoints facilities that are reporting more discharge than what is being withdrawn

- High consumption along the eastern shore and the Shenandoah region, correspond to the dense agriculture that is located in those regions.

-Negative consumption following major river ways

Model Evaluation

Comparison of Statewide Withdrawals (MGD) per Year from Analysis and VDEQ Reported Values								
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Report	7,794	7,418	7,246	7,195	7,160	6,757	6,741	6,389
VDEQ	7,621	7,559	7,100	7,047	6,984	6,558	6,613	6,313
Error	2.27%	1.87%	2.06%	2.10%	2.52%	3.03%	1.94%	1.20%

Statewide Consumption Coefficients for Classified Sectors									
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**Average**
Agriculture	1.00	1.00	1.00	1.00	1.00	0.78	0.79	0.77	**0.92**
Commercial	0.99	0.91	0.91	0.87	0.89	0.79	0.78	0.65	**0.85**
Energy	0.26	0.24	0.19	0.17	0.19	0.08	0.08	0.01	**0.15**
Industrial	0.69	0.69	0.66	0.59	0.60	0.55	0.53	0.49	**0.60**
Municipal	0.45	0.38	0.28	0.21	0.23	0.24	0.10	0.26	**0.27**

- Results are continuously evaluated through collaboration with VDEQ. Best indication is to compare to withdrawal analysis performed by VDEQ every year.
- Error is within reason, but could be lower. Further collaboration with domain experts is needed to understand discrepancy.
- Energy, Commercial, and Industrial averages are higher than typical values (~4%, ~10%, ~11%)
- Energy consumption has decreased and may be a result of legislation

Real-World Insights

Water Availability & Use Science Program

Provide Assessment of the Water Resources of the U.S.

Water Use Data & Research Program

Support State Water Resource Agencies in Collecting and Reporting Water Use Data

Office of Water Supply

Collects and Analyzes State Water Use Data. Directs Water Supply Management and Planning

The State can use improved estimates of water supply to fuel more informed water supply management and planning. This will help the USGS provide more accurate assessments of the water resources in the U.S as well.