Check for updates

Water Resources Research

TECHNICAL REPORTS: DATA

10.1029/2020WR027751

Reanalysis of Water Withdrawal for Irrigation, Electric Power, and Public Supply Sectors in the Conterminous United States, 1950–2016

Ryan A. McManamay[1], Binita KC[2,3], Melissa R. Allen-Dumas[4], Shih-Chieh Kao[2], Christa M. Brelsford[5], Benjamin L. Ruddell[6], Jibonananda Sanyal[4], Robert N. Stewart[5], and Budhendra L. Bhaduri[5]

[1]Department of Environmental Science, Baylor University, Waco, TX, USA, [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [3]Goddard Earth Sciences Data and Information Services Center, NASA Godard Space Flight Center /ADNET, Greenbelt, MD, USA, [4]Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [5]National Security Emerging Technologies Division, Oak Ridge, National Laboratory, Oak Ridge, TN, USA, [6]School of Informatics Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

Key Points:

- Spatiotemporally comprehensive water use data are needed to characterize water resource availability and model long-term hydrologic changes
- We present an annual (1950–2016) data set on US agriculture, electric power, and public supply water use at the county level
- The data set provides new spatiotemporally rich information compiled using consistent reanalysis methods to inform multi-sectoral research

Supporting Information:

- Supporting Information S1

Correspondence to:

R. A. McManamay,
Ryan_McManamay@baylor.edu

Citation:

McManamay, R. A., KC, B., Allen-Dumas, M. R., Kao, S. C., Brelsford, C. M., Ruddell, B. L., et al. (2021). Reanalysis of water withdrawal for irrigation, electric power, and public supply sectors in the conterminous United States, 1950–2016. *Water Resources Research*, *57*, e2020WR027751. https://doi.org/10.1029/2020WR027751

Received 18 APR 2020
Accepted 10 JAN 2021

© 2021. American Geophysical Union. All Rights Reserved.

Abstract

Accurately measuring water use by the economy is essential for developing reliable models of water resource availability. Indeed, these models rely on retrospective analyses that provide insights into shifting human population demands and adaptions to water shortages. However, accurate, methodologically consistent, empirically authentic, and spatiotemporally comprehensive historical datasets for water withdrawals are scarce. Herein, we present a reanalysis of annual resolution (1950–2016) historical data set on irrigation, electric power, and public supply water withdrawal within the conterminous United States (US) at the county-level, and, for power plants, at the site-level. To estimate electric power water use, we synthesized a historically comprehensive list of generators and historic patterns in generation across fuels, prime movers, and cooling technologies. Irrigation water use estimation required building a crop-demand model that utilized historical information on irrigated acreage for crops and golf courses, stage-specific crop water demand, and climate information. To estimate public water supply use, we developed a random forest model constructed from information on population, infrastructure, climate, and land cover. These estimates generally agree with total county and state water use information provided by the US Geological Survey (USGS) water use circular and estimates generated from independent studies for specific years. However, we also observed discrepancies between our estimates and USGS data that appear to be caused by inconsistencies in the methods used by the USGS's primary data sources at the state level over decades of data collection, highlighting the importance of reanalysis to yield spatiotemporally consistent and intercomparable estimates of water use.

1. Introduction

As global population expands, many nations have increasingly experienced water stress due to changes in climate and increasing water demands (Vörösmarty et al., 2000). Indeed, 49 countries are currently classified as water stressed (FAO, 2016), and in over 50% of basins worldwide, water use exceeds water availability at least during a portion of the year, leading to ecological degradation and societal vulnerability (Hoekstra & Mekonnen, 2012). Unfortunately, past adaptation strategies to avoid water stress, such as water supply expansion and increasing water use efficiency, are unlikely to ameliorate future water severity issues compounded by population growth and climate change (Brown et al., 2019). In the future, management, including more drastic adaptations to water demands, will require accurate water use science, informed by accurate water use data.

Studies requiring consistent and accurate information on historical water use are limited by the spatiotemporal resolution and precision of archived water use records (Devineni et al., 2015). Water resource planning commonly relies on retrospective historical analyses, which provides insights into patterns of human population growth and societal demands on water resources (e.g. Brown et al., 2019). Specifically, spatiotemporally variant and historically comprehensive measurements of water use are important for understanding changes in regional water resource availability and modeling long-term changes in hydrology.

However, obtaining accurate and methodologically consistent information on water usage is difficult, owing to unknown and possibly large methodological inconsistencies between measurement methods in different places and times, yielding wide ranges of uncertainty on water use as estimates are updated and refined (e.g., Averyt et al., 2013; Devineni et al., 2015; Rushforth & Ruddell, 2018). A severe challenge of compiling water use data over many years and many jurisdictions is the integration of data from disparate sources and methods (Allen et al., 2018), which may vary in spatiotemporal scale, variable definitions, and accuracy of source information. The endeavor of harmonizing inconsistent historical data is sometimes called reanalysis (e.g., Mesinger et al., 2006).

The United States (US) has among the world's highest per capita water footprint supporting consumption of goods and services (Hoekstra & Mekonnen, 2012). Over 90% of total water withdrawals in the US are comprised from three economic sectors: thermoelectric generation, irrigation, and public supply (Dieter et al., 2018). Projected increases in population, decreases in water availability, and increases in water temperature could render these sectors highly vulnerable to climate change (Barnett & Pierce, 2008; Elliott et al., 2014; Van Vliet et al., 2012). Recent evidence suggests, however, that water withdrawals in the US has decreased due to advancements in water efficiency and regulation (Dieter et al., 2018; B. H. Harris et al., 2014; Maupin et al., 2017). From 2005 to 2015, withdrawals for thermoelectric power declined 37.5% due to decommissioning of coal-fired plants and more-efficient cooling technologies and withdrawals for public supply decreased 12% despite continued population increases (Dieter et al., 2018; B. H. Harris et al., 2014; Maupin et al., 2017). Nevertheless, between 2020 and 2060, the U. S. population is expected to increase over 21% (U. S. census bureau, 2019), and electricity generation is projected to increase 35%, primarily supported through fossil or nuclear technologies (69% of generation) (AEO, 2019). Within the agriculture industry, total planted acreage is expected to remain constant by 2028, whereas total crop production is expected to increase, on average, 8.4% (USDA, 2019). This suggests that the future agriculture industry will involve more intensive agriculture practices and eventually increased water demand, unless irrigation efficiency dramatically improves. Furthermore, increased irrigation water withdrawal is at odds with suggested adaption strategies to avoid water shortages, which include transfers of irrigation water back to other sectors (Brown et al., 2019). We need better data to assess these trends and make plans.

The most spatiotemporally comprehensive compilation of primary sources of water use data in the US is provided by the US Geological Survey (USGS) Water Use Circular Series (WUCS) (USGS, 2019). The WUCS estimates and reports water usage within eight major sectors and numerous sub-sectors every five years (USGS, 2019). Limitations of the USGS WUCS include inconsistencies in the sectors and subsectors reported across time and methods used to derive those estimates, variable spatial resolutions of reported estimates and low temporal resolution of estimates (see Supporting Information 2 for full accounting of variables). Specifically, the USGS WUCS data set provides quinquennial information at the county level from 1985 to 2015. Prior to 1985, water usage across a limited number of sectors is provided only at the state level. In many cases, water use estimates for specific sub-sectors (e.g., crop and golf irrigation) or sources of water (e.g., saline vs. fresh) are reported for only a subset of years. Water use by sector is estimated through numerous published methodologies, which may vary as methods improve (e.g., Dickens et al., 2011; M. A. Harris & Diehl, 2017); however, the USGS also relies heavily on state-reported aggregate and county-level use by different entities, who use various methods or definitions in determining water withdrawals (Macknick et al., 2011; Maupin et al., 2017).

As one example, thermoelectric water use has been reported by two federal agencies, the USGS and the US Energy Information Administration (EIA). Estimates of thermoelectric water withdrawals between these agencies are based on different methods have been shown to vary substantially (Peer et al., 2016, 2019; M. A. Harris & Diehl, 2017). In both cases, water use at the power plant level is only provided for individual years or only recent years (post-2000) and is estimated for only a subset of power plants ($n < 1400$ as of 2010). However, according to the EIA, over 8,000 power plants were operating in the US in 2010, many of which use water for operations besides thermoelectric cooling (Macknick et al., 2011; R. A. M. Peer et al., 2019). Comprehensive historical data on water use for electricity production technologies could prove valuable for modeling, especially evaluating situations of water shortages.

Herein we present annual timescale (1950–2016) and spatially comprehensive estimates of water use for electric power production, agricultural irrigation, and public supply at the county or site-level (electric power plants) within the conterminous US, using a spatiotemporally consistent reanalysis technique. We focus on these three sectors, as they represent the largest reported water use of the US economy. Although we rely on the USGS WUCS divisions in sectors (Dieter et al., 2018), we fully recognize that sectoral divisions can be subjective and substantially overlap. For instance, electricity production is only one life cycle of the US energy system and fuel production and refining, reported under mining and industrial sectors by the USGS WUCS, represent considerable users of water as well. Borrowing terminology from the climate sciences, our "reanalysis" was aimed at developing a consistently derived historical record data product analogous to the USGS WUCS estimates while addressing major data gaps and limitations of those efforts. Specifically, we use consistent estimation methods to provide 67 years of annual county-level (or finer granularity) water use estimates by source and subsector to prevent spatial or temporal bias.

To estimate the water use of power generation, we synthesized a historically comprehensive list of generators and historic patterns in generation across fuels, prime movers, and cooling technologies. To estimate the water use of irrigation, we compiled historical information on crop and golf acreage and climate information to use in a crop-demand model that considered crop type and water use per growth stage. To estimate the water use of public water supply, we developed a random forest model constructed from information on population, infrastructure information, climate, and land cover. As a validation exercise, we compare our estimates to those provided by the USGS WUCS and other independent studies for specific years. Additionally, we seek to determine the cause of any clear discrepancies between this method and the USGS data. We also provide example applications to showcase the utility of the data set to multi-sector research.

2. Methods

2.1. Overview and Scope

For our analysis, we adopted the same definition for "water use" as the USGS WUCS, where withdrawal is termed the "total amount of water removed from the water source for a particular use," and consumption is termed, "the amount of water that is not readily available for another use because it is evaporated, transpired, incorporated into products, consumed by livestock or humans, or otherwise removed from the immediate water environment" (Dieter et al., 2018). Generally, our study focuses on the withdrawal of fresh surface and ground water removed from the "immediate water environment" of a river, lake, or aquifer, as these sources of water could be more reliably estimated. These sources dominate uses for irrigation and public supply; However, for electrical power generation, we include estimates of both withdrawal and consumption and include additional water sources: saline surface, saline groundwater, reclaimed waste discharge, and mixed sources, because information on consumption and sources of water were more readily available and these water sources constitute a larger fraction of total water use for that sector. This definition of water use leaves some ambiguity (Ruddell, 2018), but it is coherent with the USGS's methods, which are the primary focus of our reanalysis.

All variables from the data products we developed in this study are provided in Table 1. For the three sectors, we include variables depicting the source of water (e.g., fresh, saline) and the temporal and spatial resolutions of our analysis. With some exceptions, data are provided at the county resolution at an annual timestep from 1950 to 2016. Because irrigation estimates were derived using daily hydrometeorological information, we provide daily irrigation water use estimates for individual crops (although we urge caution in over-reliance on daily values due to uncertainties in other factors, for example, crop planting dates). Water use for electricity production is provided at county, power-plant, and electricity generating unit (EGU) resolutions. Power plants may be comprised of multiple EGUs, in which case, water use for all EGUs operating at a plant were summed. In many cases, the raw data we used in building estimates of water use was reported less frequently than an annual basis (e.g., crop land acreages and sources of water). To fill in these data gaps, we used linear and spline interpolation methods of the raw datasets.

For irrigation and electric power production, we provide multiple scenarios of varying estimates. For electric power production, scenarios are based on average, minimum, and maximum water withdrawal and consumption coefficients provided by Macknick et al. (2011, 2013) and Averyt et al. (2013). Irrigation scenarios varied according to different estimates of irrigated acreage and whether precipitation soil moisture

10.1029/2020WR027751

Table 1
Overview of Water-Use Related Data Products Provided by the Study for the Three Sectors

Sector	Categories	Time step	Spatial resolution	Water sources	Scenarios
Electric power generation	Total withdrawal, total consumption, total generation, individual generation, withdrawal, consumption estimates for all technologies listed in Table 2	Annual	County, state, power plant	Fresh surface water, fresh ground water, saline surface water, saline groundwater, reclaimed discharge, mixed sources, other sources	Coefficients reported in Macknick et al. (2012): (1) Medium, (2) Minimum, (3) Maximum
Irrigation	Total irrigation, crop irrigation, golf irrigation, individual irrigation withdrawal for 43 crop types in Table 3, time-dependent irrigation efficiencies	Daily, annual	County, state	Fresh surface water, fresh ground water	(1) No climate, low acreage, (2) No climate, high acreage, (3) Climate-adjusted, low acreage, (4) Climate-adjusted, high acreage
Public supply	Total withdrawal	Annual	County, state	Fresh surface water, fresh ground water	One scenario

was considered. Because irrigation can vary immensely by crop type, we estimated irrigation separately for over 40 crops and for golf courses.

2.2. Electricity Production Water Withdrawal and Consumption

Water use for electricity production in this study only considered the energy technologies reported by Macknick et al. (2012) and Averyt et al. (2013), which primarily consisted of off-stream power uses (Table 2). These included technologies reliant on thermoelectric cooling but also other technologies that use water for operations and maintenance of electricity production (e.g., cleaning solar photovoltaic panels or wind turbine blades). We did not elect to include hydropower due to significant uncertainty and challenges in deriving those estimates. Similar to other technologies, hydroelectric dams use water to generate electricity, although consumption of water occurs primarily through evaporative losses within reservoirs, not directly associated with power generation (Mekonnen & Hoekstra, 2012; Grubert et al., 2016). It is possible to estimate hydropower water use and consumption into equivalent effects of a municipality or farmer's water use on the "immediate water environment" (Ruddell, 2018). However, complexities arise when estimating hydropower water use, particularly whether use is calculating only during generation, whether it should include all flow passed through a dam or only flow passed across generating units, and how this relates to reservoir volumes and evaporative loss (i.e. consumption). Furthermore, hydroelectric power is usually only one of many purposes of a dam and reservoir and is commonly a lower-priority of water allocation compared with recreation, navigation, flood control, and water supply (Bonnet et al., 2015; Uria-Martinez et al., 2018). Finally, reservoir volumes are adaptively managed and reallocated, which depart significantly from established reservoir operation rule curves (Doyle & Patterson, 2019). To accurately estimate hydro-

Table 2
Types of Fuels, Cooling Technologies, and Prime Movers Used to Characterize Electric Generating Units (EGU) for Estimating Water Withdrawal and Consumption

Fuels	Cooling technologies	Prime movers
Biomass, Coal, Geothermal, Hydropower, Pumped Storage, Natural Gas, Nuclear, Oil, Solar, Wind	Cooling pond, Dry cooled, Once-through, Recirculating, None	Binary cycle, Combined cycle, Combustion turbine, Dry stream, Enhanced geothermal system, Fuel Cell, Photovoltaic, Trough, Steam, None

Note. Fuels, cooling technologies, and prime mover combinations resulted in 35 EGU types. EGU technology combinations and water withdrawal and consumption estimates per megawatt hour (MWh) for all each technology are provided in Tables S1 and S2.

10.1029/2020WR027751

power water use (equivalent to other technologies presented in our study), reservoir operations, specifically time-variant water volumes used solely for hydropower generation should be proportioned by allocated reservoir volumes and associated evaporation rates. Due to the significant challenges in assembling the required information for the entire period of record, we did not consider hydropower water use in our study.

To extend the record of county-level electric power water use and water consumption backward to 1950, we followed a "coefficient approach" used by Averyt et al. (2013) to obtain water withdrawal and consumption at the EGU scale according to a combination of fuel type, cooling system type, and prime mover technologies (Table 1, Figure S1). By calculating coefficients on a per-energy production basis for each EGU technology, historical water use can then be extrapolated by estimating historical electricity generation. An overview of our approach to assign water use per EGU technology is provided as a flow chart in Figure S1. We first developed a master list of all power plants and generators in operation at any time as recorded in the EIA Form 860 Annual Electric Generator Report (EIA 2019a) or Form 923 Power Plant Operations Report (EIA 2019b). The list included 8928 plants and 23697 generators in operation as early as 1891 and as late as 2016. Of the list of generators, 18,039 (or 76% of the total) were in operation as of 2016. EIA form 860 includes information at the generator level including nameplate capacity (MW), capacity factors (%), ownership, age, fuel use, status of operation (e.g., operating, retired, out-of-service, installation postponed), month and year of initial operation, month and year of retirement, boiler type and efficiency, and cooling system types. EIA form 923 provides geospatial locations of power plants and information on monthly and annual generation at the generator and power plant level from 2001 to 2017.

Combinations of fuel type, cooling systems, and prime mover technologies resulted in 35 different EGU typologies (Tables S1 and S2). Averyt et al.'s database reported water withdrawal and consumption for the majority of EGUs operating in 2008 using water use estimates per technology provided Macknick et al. (2011; 2012). Based on the EIA generator identifier, we joined Averyt et al's database to our list of EGUs and calculated average, minimum, and maximum water use and water consumption per MWh of energy produced (based on 2008 generation estimates, as this mirrored Avery et al.'s data) (Tables S1 and S2). A total of 9398 EGUs in our master data set were not found in Averyt et al.'s database. Of these, however, only 1,422 required water for electricity production and hence, needed an appropriate technology code.

For missing records, we used a similar approach as Averyt et al. (2013) to associate cooling technologies to each EGU (Figure S1). As Averyt et al. noted, each plant may include multiple different types of cooling infrastructure and boilers, multiple boilers may be associated with each generator, and each cooling structure may be associated with multiple boilers; however, EIA provides no explicit linkage between each cooling structure and each generator. EIA does, however, provide cooling operation types at the power plant level. Of the 1422 generators missing from Averyt et al.'s database that required cooling, 892 had no cooling technology information reported for their respective plants in the EIA data set. Additionally, 457 generators had one cooling technology reported for their respective power plant whereas 73 generators had multiple cooling technologies reported. The 892 EGUs with missing cooling technologies had information on fuel type and prime mover, which we summarized into fuel-prime mover combinations (i.e., "New Codes," Tables S1 and S2). We calculated water withdrawal and consumption per MWh for these New Codes based on averages across all EGU typologies sharing the same fuel and prime mover type combinations (Tables S1 and S2). The 457 generators were assigned the same cooling technology as that listed for the power plant. For the 73 generators with multiple cooling technologies reported for their respective power plants, we summarized the proportion of total intake capacity for a given cooling type at each plant. We then proportionalized the generating capacity of each generator based on the total nameplate capacity of each plant. For each individual record, we assigned cooling technologies to generators by matching the approximate intake proportions of cooling types to generating capacity proportions. We also considered whether other generators at the same plant had cooling technologies already assigned to them (i.e., not missing data).

After compiling water use per MWh for all EGU typologies, we then applied these rates to annual generation for each EGU to estimate water withdrawal and consumption for every year since 1950. This required developing a temporally comprehensive data set of annual generation, which required compiling values from three sources. Source 1: The EIA 923 form provided annual generation at the generator level for a subset of generators from 2008 to 2016. Source 2: The EIA 923 form also provided annual generation from 2001 to 2016 for entire power plants operating within that time frame. Based on the proportion of each generator's nameplate capacity relative to the total nameplate capacity for each respective power plant, we proportionalized generation from plants to the generator level. Source 3: In cases where generation information was missing, we estimated annual generation by multiplying nameplate capacities with capacity factors and constraining those estimates to only years (or partial years) when each generator was operating (based on initial month-year and retirement month-year). Capacity factor estimates were obtained from Averyt et al. (2013) or, if unavailable, calculated by dividing average net reported generation (2001–2016) by the total potential generation assuming plants were operating continuously; these capacity factors were applied to each generator. Based on initial operating year and retirement year, we compiled generation from 1950 to 2016 prioritizing data from source 1 first, source 2 second, and finally source 3 if the previous sources were unavailable.

We applied water use coefficients per technology code to each EGU's annual generation to calculate water use for all years. To partition water use for each EGU by the source of water, we first relied on the sources reported in EIA's thermoelectric cooling water data for 2014 through 2016 (EIA 2019c). EIA provides 13 categories of water sources, which for simplicity sake, we reduced into seven sources: fresh surface, fresh groundwater, saline surface, saline groundwater, reclaimed discharge, mixed sources, and other. Water sources reported in 2014 were assumed constant for all previous years. The EIA data only provided water sources for 831 of the >8000 power plants in our data set. To partition water sources for the remaining power plants, we relied on USGS water use estimates at the county-level; however, these were only available for four water sources (fresh surface, fresh groundwater, saline surface, saline groundwater) because other sources were intermittently reported. We used a nonlinear moving-average interpolation approach to backcast historical water use by source for every 5 years. Using USGS county-level data from 1985 to 2015, we developed linear regressions of source-specific water use versus year to estimate source-specific water use in 1950. We started the interpolation by deriving 1980 values using the average between 1950 estimates (from regression) and empirical values for 1985 and 1990. In turn, 1975 values were estimated by interpolating 1950 estimates with 1980 and 1985 values, and so on until 1955 values were interpolated. This type of moving-average estimation weights values more heavily based on known values in empirical data than estimated values. Water use by source was then converted into proportional values.

In some cases, our approach yielded water use for electricity production for counties in which the USGS reported no usage; hence, there was no ability to backcast or interpolate water sources for these instances. To develop estimates for these counties, we used Spatially Constrained Multivariate Clustering in ArcGIS (10.3) to develop regionally affiliated clusters of similar water use patterns (Figure S2). Using total proportional water use by source (across all sectors) for 2015, we developed spatial clusters using a *K*th nearest neighbor approach based on Euclidean distances. We selected 100 clusters for the CONUS and a neighborhood radius of 20 counties (Figure S2). Based on counties where thermoelectric water use by source was reported from the USGS, we averaged the proportion of water use by source for each cluster for each year (1950–2015). Those proportions were then applied to counties with missing data falling within the same cluster. All proportions were multiplied by total electricity production water use values to partition use by the four water sources.

The above approach yielded water use by source for every 5 years from 1950 to 2016. To generate annual estimates, we used a nonlinear spline interpolation within the impute TS package in the R programming environment (Moritz & Bartz-Beielstein, 2017).

2.3. Irrigation Water Withdrawal

Irrigation for crops is the predominant driver of irrigated water practices worldwide. In some US counties, however, golf irrigation is the largest use of irrigated water (Ivahnenko, 2009; Maupin et al., 2017). In addition, in years 2005–2015, the USGS WUCS differentiated irrigation water use into only crop and golf sub-sectors. To compare our estimates to that of the USGS required that we differentiate between these sub-sectors as well. For our purposes, we term both of these irrigation practices as "Agriculture," as crop irrigation is dedicated to field crops, vegetables, and fruit and nut orchards, whereas the golf irrigated is dedicated to turf grass, another cultivated product.

10.1029/2020WR027751

2.3.1. Crop Irrigation

Water use for crop irrigation was estimated using historical agricultural irrigated acreage, crop water demand models, and irrigation loss by crop types. Irrigation withdrawals for a given crop are generally estimated using Equation 1 (Dickens et al., 2011)

$$W = \frac{(A \times C)}{L}, \tag{1}$$

where W is the irrigation withdrawals in acre-feet for a crop; A is the irrigated acreage of each crop in the specified state, in acres; C is the irrigation water requirement for each individual crop in feet; and L is the potential water loss while irrigating in decimal fraction.

The first necessity and challenge lay in obtaining historically comprehensive irrigated acreage estimates for different crops. Historical total land acreage and irrigated land acreage from 1950 to 2017 by agriculture product at county level was reported at 4- to 5-years increments by the agriculture (Ag) census conducted by the United States Department of Agriculture (USDA). The compiled census results were obtained from the Inter-university Consortium for Political and Social Research (ICPSR) (Haines et al., 2016) and the USDA National Agricultural Statistics Service (NASS) (USDA NASS 2020). Data were available only for the following years: 1950, 1954, 1959, 1964, 1969, 1972, 1978, 1982, 1987, 1992, 1997, 2002, 2007, 2012, and 2017. Agriculture products and the naming convention of agriculture products reported by the Ag Census varied immensely across years and required manually compiling variables across all years available. Considerable individual attention was required to ensure consistent assemblage of land acreage by crop types across the years. Counties with missing crop land acreage for individual years, indicated by no values, were imputed using classification and regression trees method in "mice" package in R (van Buuren & Groothuis-Oudshoorn, 2011; van Buuren, 2018). Imputation methods use values where data is available to predict missing observations. For each variable with missing values for counties, predictive models are developed using all available observations as a response variable, whereas other variables in the data set, such as total cropland harvested and crop production estimates for each county, as predictor variables. Following imputation, land acreage estimates were only available for 15 years of the entire 67-year period of record. To estimate annual fluctuations in crop land acreage values, we used the same nonlinear spline interpolation reported earlier (ImputeTS package in R). Examples of land-acreage interpolation results are provided in Figure S3.

Following imputation and interpolation, total land acreage estimates were available for 44 agricultural products, whereas estimates of irrigated land acreage were available for only 17 crops and 4 major crop categories (Table 3). Since not all land planted for a given crop is irrigated, total land acreage would overestimate irrigated water use. This required that we estimate irrigated acreage for the remaining 27 crops. We multiplied total land acreage estimates of agricultural products for each county by the proportions of land irrigated according to the respective major crop categories (cropland, berries, fruit and nut orchards, and vegetables) (Table 3).

Dickens et al. (2011) suggests that irrigated lands estimated by the Ag census may under-estimate the actual land irrigated for agriculture. Given the potential for underestimation and uncertainty in our estimates of irrigated land acreages, we developed two scenarios of low and high irrigated lands. For the "low" scenario, we used irrigated land acreages directly reported from the Ag census for the 17 crops and then applied irrigated acreage proportions to only agriculture products with missing irrigated land coverages. For the "high" scenario, we applied irrigated land proportions for the major crop categories to all agricultural products and then selected the maximum irrigated land acreage estimate for each crop.

Crop water requirement *(C)* is largely dependent upon crop water needs (ET_{crop} in mm/day) in relation to background evapotranspiration rates across the country, which was calculated using the following equation:

$$ET_{crop} = ET_o \times K_c, \tag{2}$$

where ET_o is the reference evapotranspiration, and K_c is the crop factor. Evapotranspiration was calculated using temperature and daylight hours using the Blaney–Criddle (1962) equation:

10.1029/2020WR027751

Table 3
Agriculture Products With Land Acreage Provided by US Department of Agriculture (via ICPSR[1]) Used to Estimate Agriculture Irrigation

Plant type	Agriculture products
Crops	Alfalfa*, Barley*, Buckwheat*, Corn for grain*, Corn for silage*, Cotton*, Hay* (besides alfalfa), Irish potatoes*, Oats*, Peanuts*, Rice*, Rye, Sorghum for grain*, Sorghum for Silage*, Soybeans for beans*, Sugarbeets, Sweet potatoes*, Tobacco*, Wheat*
Fruits and vegetables	Cantaloupes, Cucumber, Hot peppers, Lettuce, Snapbeans, Spinach, Sweet peppers, Tomatoes, Watermelons, Other vegetable category (besides those listed)
Fruit and nut orchards, vineyards, and berries	Almonds, Apples, Apricots, Avocados, Blueberries, Cherries, Citrus, Grapes, Olives, Peaches, Pears, Plums and prunes, Raspberries, Strawberries, Other berries (besides those listed)
Aggregated agriculture lands for estimating irrigated lands	Berries*, Cropland harvested*, Fruit and orchard*, Vegetables (all) *

Note. *Indicates Crops With Irrigated Land Estimates, Whereas for all Other Crops, Only Total Land was Available and Irrigated Acreage Required Estimation
*https://www.icpsr.umich.edu/index.html.

$$\mathrm{ET}_o = p\left(0.457 \times T + 8.128\right), \tag{3}$$

where, ET_o is the daily reference evapotranspiration (mm day^{-1}). T is the daily temperature (°C) given as and p is the daily percentage of daytime hours. The Blaney–Criddle (BC) equation is often used because it only relies on air-temperature and day-light hours datasets, which are widely available. For the same reason, we elected to use the BC method; however, we note that the BC represents potential ET, not reference ET, and can be inaccurate relative to Penman-Monteith, which requires significantly more information (Brouwer & Heibloem, 1986). Daily ET_o was calculated for each county from 1950 to 2016 using T and p values provided via Daymet gridded data (Thornton et al., 2018).

K_c depends on the type of crop, the growth stage of the crop, geography, and the climate. Generally, K_c coefficients are represented as nonlinear, seasonally variant curves that vary based on planting dates, total length of the growing season and the lengths of the various growth stages, which vary by climate zones (Figure S4). The total growing period starts from sowing, transplanting, or, in the case of perennial crops, the bloom date, to the last day of the harvest and depends on the type of crop, and the climate. The duration of each growth stage for various field crops and their corresponding K_c values were obtained from Brouwer and Heibloem (1986). Planting date and bloom dates for each crop were acquired from the USDA NASS (USDA 2010). The report provides the begin and end date of planting/blooming and harvesting for each crop by state. For consistency, we determined the plantation date or bloom date as 15 days prior to the provided end dates for planting and blooming. Using the total duration of growth stages, we calculated the number of days in a month associated with the stage of growth for a given crop. Hence, for each crop, K_c for that month was calculated by simply multiplying the crop factor for that growth stage with number of days in a month that are in that stage of growth (Figure S4). If there are multiple growth stages in a month, then the arithmetic mean of the K_c values were taken. Depending on the crop type, USDA provides the planting dates and bloom dates only for some states whereas acreage of crops for each year commonly occur in states without documented planting and bloom dates. To calculate the crop factor for states with missing values, the same value for crop factors were applied to those states according to average values for crop hardiness zones. State-specific and/or zone-specific K_c values were multiplied by ET_o to obtain ET_c by county.

ET_c provides an estimate of crop water demand; however, this does not take effective precipitation and soil water availability into account (Hoekstra, 2019), and thus, ET_c is an overestimate of irrigation; We then can use ET_c as an upper ceiling of future irrigation, assuming irrigation is halted when soil availability meets crop water demands. When estimating daily irrigated water use for crops (IWU_c), daily soil water availability is taken into account using Equation 4:

$$\mathrm{IWU}_c = \mathrm{ET}_c - \left(P - Q_s\right) - \Delta S, \tag{4}$$

where P is daily precipitation, Q_s is daily surface runoff, and (ΔS) is the daily change in soil moisture. Precipitation, surface runoff, and soil moisture were obtained from Livneh et al. (2015) and Livneh et al. (2015), respectively, and were summarized by county at the daily time step. This suggests that IWU_c is blue water consumption (evaporated by crops), whereas IWU_c—ET_c is green water use (rainfall) by crops.

Water loss during irrigation (conveyance loss) depends on the efficiency of the type of irrigation system used. Dickens et al. (2011) provide a range of estimates for L, reported as field irrigation efficiencies among various methods. We summarized these efficiencies by three major categories presented in Table 3. County-level estimates of acres irrigated by sprinklers, surface flooding, or micro-drip irrigation were available in 1950 (Haines et al., 2016) and from 1985 to 2015 (Maupin et al., 2017). To estimate acreage irrigated by each of the methods from 1955 to 1980, we used the same nonlinear moving-average interpolation mentioned previously in the thermoelectric analysis. Because micro-drip technologies were largely unavailable prior to the 1960s (Taylor et al., 2014), we presumed acreage irrigated via micro-drip prior to 1960 were 0. Using the range of values in Table 3, we derived time-variant conveyance loss estimates assuming that efficiency increased via technology updates from minimum values in 1950 to 75th percentile values in 2016. To calculate L for each county for each year, we calculated a weighted average of conveyance efficiency based on the proportion of acreage in each county irrigated using each of three major categories.

We developed four scenarios irrigation demand or withdrawal w based on the two estimates of irrigated land acreage described earlier and using estimates of crop water demand and withdrawal, ET_c versus IWU_c, respectively. Here, we replace term C in Equation 1 for ET_c or IWU_c, depending on the scenario. For each scenario, daily crop water demand or irrigation requirements were aggregated to annual estimates and summed across all agricultural products. Using a similar approach to the thermoelectric analysis, we calculated proportions of total irrigation water use into groundwater and surface water factions by relying on USGS estimates where available (Maupin et al., 2017) and using the combination of regression and interpolation methods mentioned previously. For each scenario, we then multiplied total crop water use requirements to these proportions.

2.3.2. Golf Course Irrigation

Estimating golf course irrigation followed a similar approach to crop irrigation, except that there were no consistent datasets on golf course acreage in the US from 1950 to 2016. To estimate acreage, we obtained the US Golf Course Database (2019), which provided latitude and longitude locations of golf courses, the number of holes, and the year built. We reconstructed a historical data set of golf courses by summarizing the number of holes per county per year. Numbers of golf course holes have been previously used to estimate golf course size and irrigated acreage with high accuracy (Ivahnenko, 2009). Indeed, our data showed strong association ($r^2 = 0.88$) between number of course holes and golf course acreage reported by the USGS (Figure S5). Using USGS estimates of irrigated golf course acreage for years 2005, 2010, and 2015, we calculated acreage-per-hole coefficients for each county or for each state if county estimates were not available. We then used the coefficients to estimate irrigated golf course acreage per county per year.

We used the same approach for estimating golf irrigation as crop irrigation (Equations 1–4). We assembled data on the types of turf grass grown in different states or zones, their growing season, and associated kc values. Similar as above, states or zones with missing values were assigned average values of nearest neighboring zones or states. We presumed all irrigation was achieved via sprinkler systems and only applied time-variant conveyance loss estimates for sprinklers (coefficients provided in Table 4). In contrast to crop irrigation, only two scenarios of golf course irrigation were developed using ET_c and IWU_c as we only had a single estimate of golf acreage for each year.

2.4. Public Supply Water Withdrawal

According to Dieter et al. (2018), public supply from municipal sources accounts for 14% of total water use in the US. We focused on municipal supplies, rather than domestic self-supply, which accounts for 3.8% of total water use. The USGS uses a variety of methods and data for estimating municipal public-supply

Table 4
Estimates of Field Irrigation Efficiencies (L) Among Three Major Irrigation Methods

Irrigation method	Average	Min	Max	75th percentile
Spray	0.819	0.650	0.950	0.900
Micro-drip	0.867	0.850	0.900	0.875
Surface flooding	0.729	0.650	0.850	0.775

Note. Data summarized from Dickens et al. [11].

withdrawals, which include collecting data from individual state water regulatory agencies, surveys, or estimating use based on population served (Kenny et al., 2009). Estimating the population served by public supply is not straightforward and requires gathering information from multiple sources including EPA SDWIS (2014), U.S. Census (1996; 2019), and public suppliers (Kenny et al., 2009). For the entire time period under consideration for our analysis, compiling information from these sources to estimate public supply was impractical, or impossible (due to missing data). In contrast, Worland et al. (2018) used a Bayesian hierarchical regression approach to model municipal water use in counties across the US and the authors reported relatively good model performance.

Likewise, we adopted a statistical modeling approach to model public water supply. Specifically, we utilized a random forest statistical modeling approach trained to match USGS observations for public water supply using a suite of predictor variables. Random forests are a form of ensemble machine learning where many decision trees (>500) are iteratively constructed and fit to a data set using random subsets of observations and random subsets of variables available to each tree node (Breiman, 2001). Each tree is fit to the training data and then predictions are combined from all trees. The unused observations, termed the out-of-bag sample, are used in cross-validation procedures to estimate error and variable importance (Breiman, 2001; Cutler et al., 2007). Similar to Worland et al. (2018), we compiled a suite of time-variant predictor variables at the county-level hypothesized to be important to estimating public water supply use. These included: county-level population estimates, land use, climate, runoff, dam storage, aquifer permeability, and spatial autocovariates (Table S3). These variables serve as surrogates of complex mechanisms that we were unable to model in a mechanistic fashion. For instance, dam storage serves as a surrogate of water supply availability and proximity to dense populations, as we cannot characterize water distribution infrastructure for the entire US. As another example, regions of high aquifer permeability are indicative of groundwater source reliance and thus, permeability is used as predictor variables for partitioning ground and surface water sources. One difference in our approach to that of Worland et al. is that we did not account for economic and behavioral information (e.g., Gini index, voting partisan metrics) that might be important for predicting shifts in water use efficiency over time. However, many of these variables are unavailable for our entire period of record. Our assessment of land cover, in conjunction with spatial autocovariates, may serve as surrogates of efficiency since per-capita water use increases with dense urbanization and higher efficiency tends to follow north-south gradients and urban-rural gradients (Sankarasubramanian et al., 2017).

2.4.1. Population Estimates

Annual county-level population data was tabulated and estimated based on two major data sets: 1) the National Cancer Institute Surveillance, Epidemiology and End Results (NCI SEER, 2019) Program (1969–2016) and 2) U.S. Census Bureau Population of States and Counties of the United States (1790–1990), Part III (U.S. Census, 1996). The NCI SEER data set provides information on intercensal shifts in population due to major human migrations (e.g., hurricane Katrina) and changes in county boundaries or development of new counties. Methods for accounting for shifts in county boundaries or identification codes are provided in Text S1. We estimated intercensal populations using the Das Gupta (1981) method formula:

$$P_t = \left[P_{3652} \times \left(\frac{t}{3652}\right)\right] + \left[P_0 \times \left(\frac{3652 - t}{3652}\right)\right],$$

with P_t = population estimate at time t, P_{3652} = ten-year from reference year decadal census count, t = time in days elapsed since reference year decadal census count, and P_0 = reference year census count.

2.4.2. Other Predictors

Besides population size, public supply water use estimates are likely influenced by the extent of urban development in conjunction with other land uses, infrastructure for supply and delivery (such as reservoirs), climate and water availability, and patterns related to spatial dependencies (Table S3). To provide a temporally comprehensive assessment of land cover from 1950 to 2010, we used two derived datasets of historical (1936–2005) and contemporary (2005–2016) land cover maps produced at annual time steps by the USGS using a land use and land cover (LULC) model (T. Sohl et al., 2016; 2014, respectively). LULC maps were generated at 250-m resolution with 14 land cover classes (T. Sohl et al., 2016). Historical LULC maps were developed using a suite of historical datasets (e.g., Census of Agriculture) (T. Sohl et al., 2016) whereas contemporary LULC maps were based suitability surfaces derived from historical land cover and land use change (T. L. Sohl et al., 2014). We used zonal statistics in ArcMap 10.2 to tabulate the area occupied by each LC type in each county per year. For public supply, the primary focus was on temporal changes in developed land cover; however, other land cover types (barren, forested, and agriculture) provide geographic context of how urban areas interact with the landscape across the country.

To account for potential public water supply volumes in reservoirs, we summarized the storage available per county per year for different types of reservoirs. Although the Congressionally authorized purpose of reservoirs was provided by the US Army Corps of Engineers National Inventory of Dams (NID), there are many US reservoirs serving as sources of public water supply that were not originally authorized for that purpose. We used the National Anthropogenic Barrier Data set (Ostroff et al., 2013), which provides accurate coordinates of dams, to summarize reservoir storage in each county. Using ownership, purpose, and year of construction, we summarized the total NID storage available each year for all reservoirs, federally owned reservoirs, state/local/utility owned reservoirs, and reservoirs authorized for water supply.

Demand for public water supply is driven by population size and urban area extent, but also potential reductions in demand due to limited water availability due to geographical differences in hydroclimatic regions or temporal variance, such as drought. We compiled 800-m grid monthly precipitation datasets (1950–2010) from PRISM (Daly et al., 2008) and summarized these values at the county level using zonal statistics (ArcMap 10.2). Monthly precipitation was summarized as annual averages and total annuals. Runoff per county was derived using the USGS WaterWatch estimates computed for hydrologic unit code (HUC) eight watersheds (USGS, 2018). HUC eight boundaries were overlaid with county boundaries and weighted averages (based on area) were used to derive county level runoff estimates. Generally, groundwater use varies considerably among aquifers, which are characterized by different lithologies with varying levels of bedrock permeability (Maupin & Barber, 2005). Wolock et al. (2004) translated bedrock lithologies into permeability classes for small watersheds across the US, with "1" being lowest permeability and "7" being the highest. As an indication of propensity for groundwater use, we summarized permeability classes into counties.

Patterns in public water use in a given county may also reflect usage in neighboring counties due to complex physical or social infrastructures, unobvious municipal boundaries, or water resources that share county boundaries. Because these are difficult to map, variables depicting spatial dependencies among counties can provide a surrogate of these unobserved factors. Using 2015 USGS county-level water use estimates, we derived inverse distance-weighted autocovariates using the spdep package (Spatial Dependence: Weighting Schemes, Statistics and Models) in the R programming environment (Bivand et al., 2013; Bivand & Piras, 2015). Spatial autocovariates were generated for 2015 public supply groundwater, surface water and total water estimates separately. The autocov_dist function was used to predict each of these estimates based on coordinates from the centroid of each county and neighborhoods within a 20-county radius.

Random forest models were calibrated using a subset of the USGS public water supply use estimates available every 5 years at the county level from 1985 to 2015. Random forests were implemented using the randomForest package in R (Liaw & Wiener, 2002) using 500 trees for each training session. Separate models were developed to predict total groundwater public supply use, total surface water, and total ground and surface water combined. Response variables and a set of predictor variables were log (x+1) transformed prior to analysis. Variable importance of random forest models is provided in Table S4. All county-level USGS data were used for training model development, except 2002, which only considered freshwater.

10.1029/2020WR027751

2.5. Data Overview and Validation

Our analysis provides annual (and daily for irrigation) water use estimates by water source for the three sectors at the US county-level and state-level for every year from 1950 to 2016 (Table 1). Multiple scenarios are typically included to provide ranges of water use values. Examples of spatial variation for individual years for water withdrawal within irrigation and electricity production sectors are provided in Figure 1 and for the public water supply sector in Figure 2. Additionally, we provide annual water withdrawal estimates for electricity production at the power-plant level (Figure 3, consumption is available but not pictured).

For all three sectors, 67 years of data provides an unprecedented ability to examine long-term regional and sub-regional shifts in water use patterns at high spatial resolution (Figure 4). We split the conterminous US into nine state-groupings following USGS Interior Regions to examine region-specific behaviors in water use over time (Figure 4). Using the entire period of record, we calculated slopes of water use magnitudes over time for each county and time-variant spline curves, representing a central tendency in water use behavior among all counties within a region (Figure 4). With some exceptions, overall trends indicate that water withdrawals have predominantly increased across all sectors over the 67-years period, although irrigation and electricity production has shown more evidence of increases and decreases. While this long-term examination of public water supply withdrawals is agnostic to recent efficiency upgrades, estimates of withdrawal for other sectors take socioeconomic shifts into account. For instance, many counties across all regions show evidence of decreases in withdrawals for electricity production in the last decade due to the decommissioning of large thermoelectric power plants, primarily coal (Figure 4). Likewise, irrigation considers increases in conveyance efficiency along with increasing shifts toward sprinklers and micro-drip systems, away from flood irrigation. Irrigation water use was estimated separately for 43 individual crops and for golf courses, which provides rich data to examine sub-sector specific spatiotemporal trends and isolate predominant shifts in water use among those sub-sectors at the county-level (Figure 5).

2.6. Validation

The USGS provides the most spatially and temporally comprehensive information on water use in the US to support a comparison with our results. Water use for each sector and source were compared to USGS estimates at both the county and state levels for the periods 1985–2015 and 1950–2015, respectively, when available (Table 5, S4–S7). However, data availability for comparison varies depended on sub-sector. For instance, separate estimates of crop and golf irrigation by the USGS were only available from 2005 to -2015. We observed numerous cases of no water use reported for the USGS, whereas our estimates yielded > 0 water usage (Figure S6). Therefore, we conducted two separate comparisons of our data to USGS estimates for each sector and source (except public supply): one including all data and another excluding zero values. All data, except public supply, were log (x+1) transformed prior to analysis and compared using linear regression.

We also compared our results to that of recent studies conducted for individual years where data were available. For example, Peer et al. (2016) analyzed cooling water for thermoelectric power plants in 2010 based on operator-reported EIA data; however, M. A. Harris and Diehl (2017) suggested that some reported estimates are thermodynamically implausible and developed a separate model for estimating water use at thermoelectric power plants operating in 2010. We compared our water use estimates at individual power plants to those of USGS-modeled estimates reported in M. A. Harris and Diehl (2017) and Peer et al. (2016) for the year 2010. At a national level, E. A. Grubert and Sanders (2018) analyze water use according to water source for year 2014 within all life cycles and technologies of the US energy system. To provide a comparison, we isolated Grubert and Sander's water withdrawal and consumption estimates within the "conversion" or "production" stages of the energy life cycle relevant to electricity production, particularly power plant cooling or washing solar panels and wind turbine blades. We then compared our total 2014 water withdrawal and consumption estimates according to water source with that reported by E. A. Grubert and Sanders (2018). Based on the Agriculture Census Farm and Ranch Irrigation Survey, the USDA National Agriculture Statistics Service provides irrigation water (in acre feet) applied to farms at the state level for 2013 (USDA 2020). Additionally, Marston et al. (2018) provides green and blue water footprint estimates of many commodities including irrigation for crop production at the county-level for year 2010. We compared

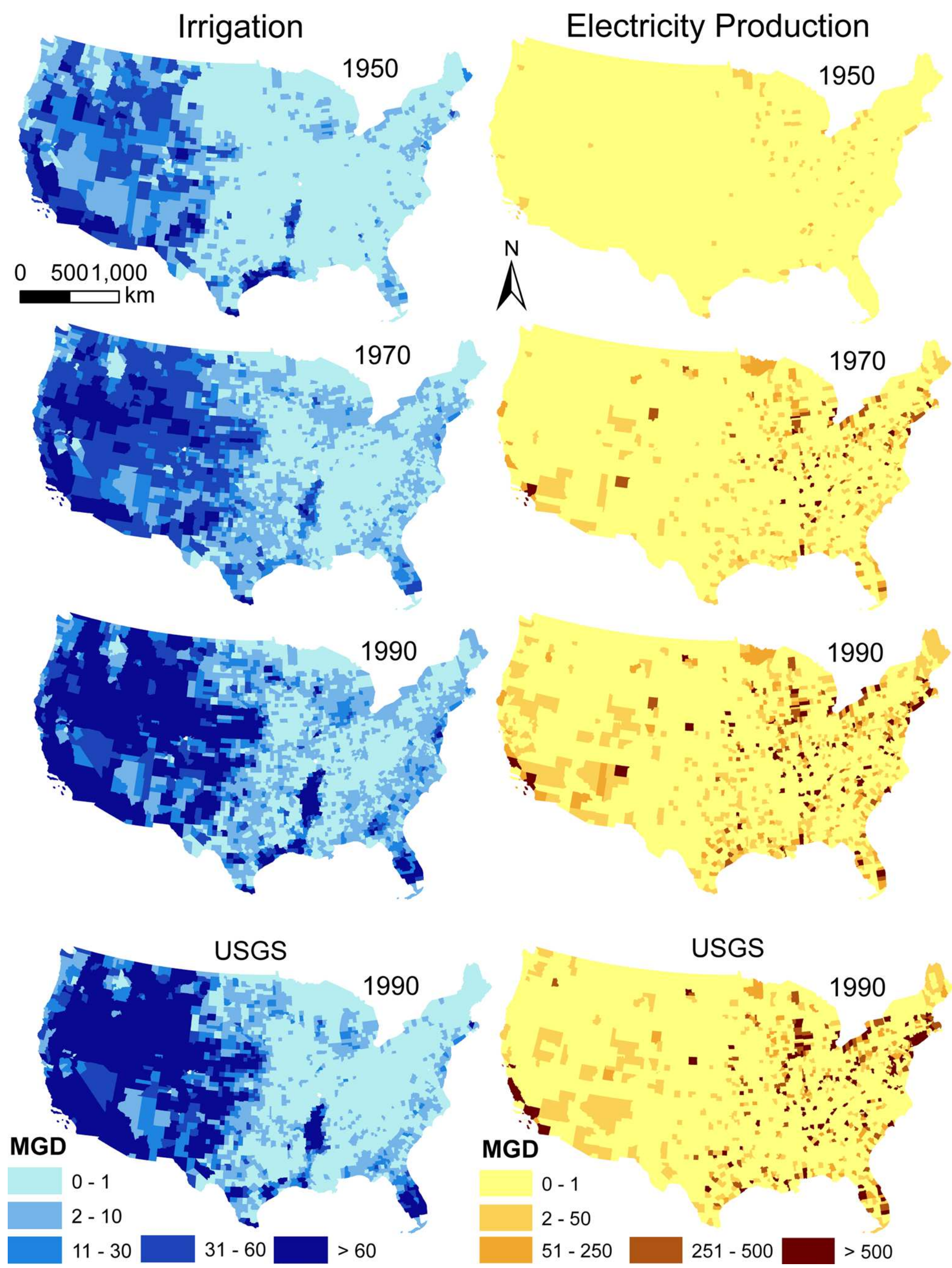


Figure 1. County-level total water use estimates from this study for irrigation (crop and golf, low acreage, climate adjusted scenario) and electric power production (medium coefficient scenario) for selected years. For comparison, USGS irrigation and thermoelectric water use estimates are shown for 1990. MGD, million gallons per day; USGS, US Geological Survey.

our crop irrigation estimates to these data sources at state and county levels, respectively, for the respectively years. Finally, Worland et al. (2018) used a statistical model calibrated from USGS data to estimate of public supply water use per household (*wh*) in counties for the year 2010. The best model from that study explained 66% of variation in USGS data. To determine our model's performance for 2010, we compared total public water supplies and *wh* calculated from our estimates to that of the USGS.

Generally, our data show agreement with USGS estimates at both state and county scales, with some exceptions (Figure 6, Tables 4, S5–S8). Random forests predicting USGS public water supply use values displayed

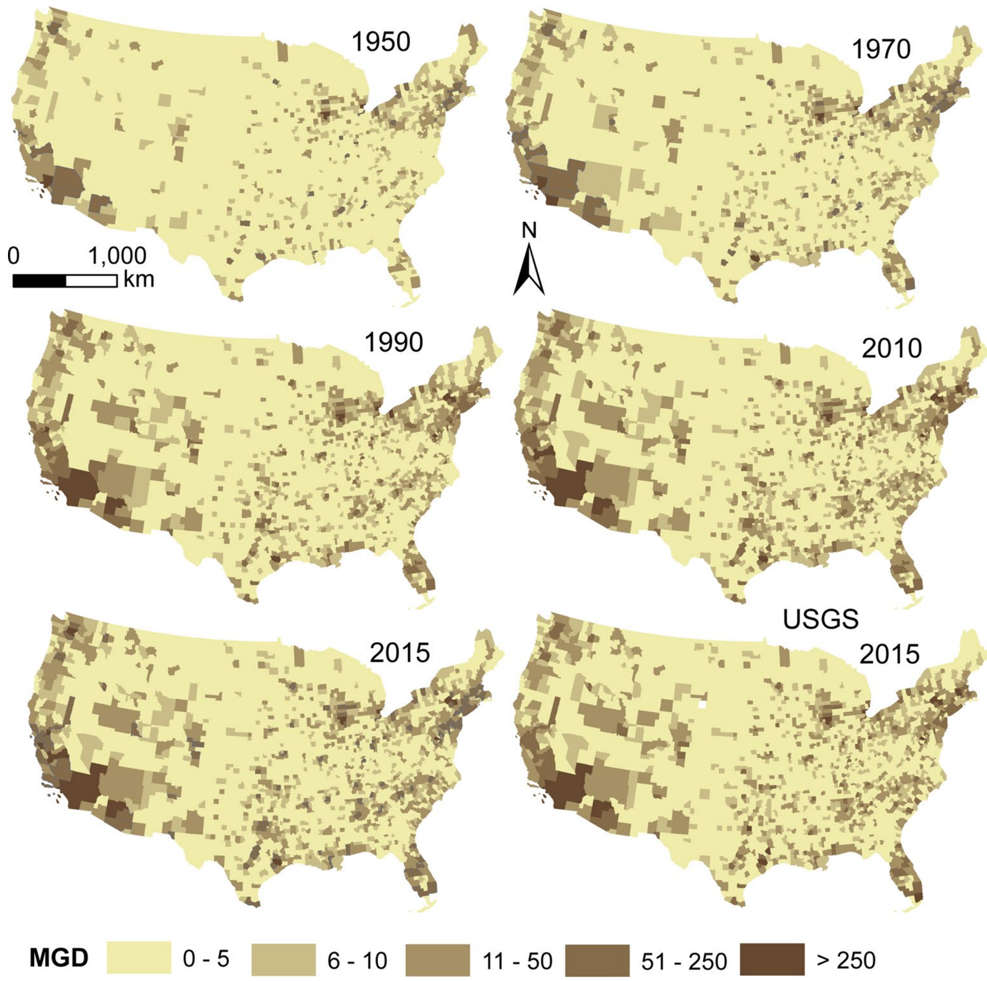


Figure 2. County-level total public supply water use estimates from this study for selected years. For comparison, USGS public supply water use estimates are shown for 2015. MGD, million gallons per day; USGS, US Geological Survey.

strong performance ($r^2 > 0.90$). Population and developed land cover were the most important variables for all three models (Table S4). Water supply reservoir storage was important in predicting surface water public supply whereas aquifer permeability was important in discerning groundwater sources of public supply (Table S4). As we expected, spatial autocovariates were also important to differentiating water sources.

Estimates of irrigation water use were strongly associated with that of USGS total water use estimates ($r^2 > 0.80$), whereas our state and county electricity production water use estimates displayed weaker associations with USGS data (state, $r^2 = 0.55$; county, $r^2 = 0.63$). Some discrepancies between our estimates and USGS WUCS are expected because our approach estimated water use for a wider variety of electricity production technologies (solar PV, wind) and our analysis includes a more comprehensive list of power plants than the thermoelectric facilities reported by the USGS. Partitioning water use by source (e.g., groundwater, surface water) and by sub-sector (e.g., crop, golf irrigation) did not necessarily lead to weaker associations between our estimates and USGS data but showed mixed results. In some cases, such as thermoelectric use, estimates for fresh and saline surface water usage showed stronger agreement with USGS estimates than total usage, whereas partitioning water use estimates into fresh and saline ground water sources showed less agreement with USGS data (Tables 5, S6). Excluding zeros from the analysis only marginally improved the strength of agreement between our estimates and that of USGS, except for crop and golf irrigation. Following removal of zero values, we observed considerable improvement in r^2 values for both crop and golf water use estimates, both as totals and partitioned by source (Tables 5, S8).

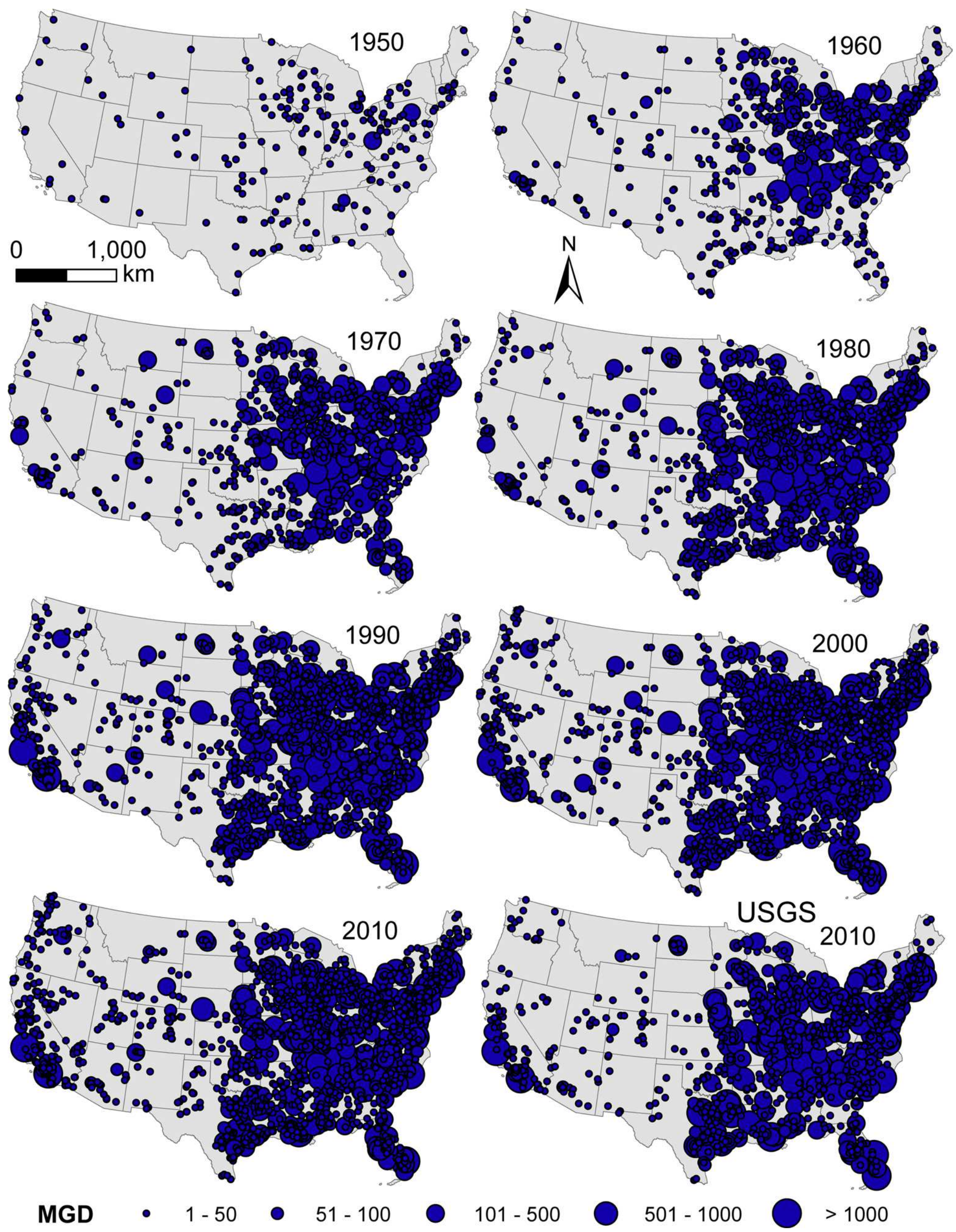
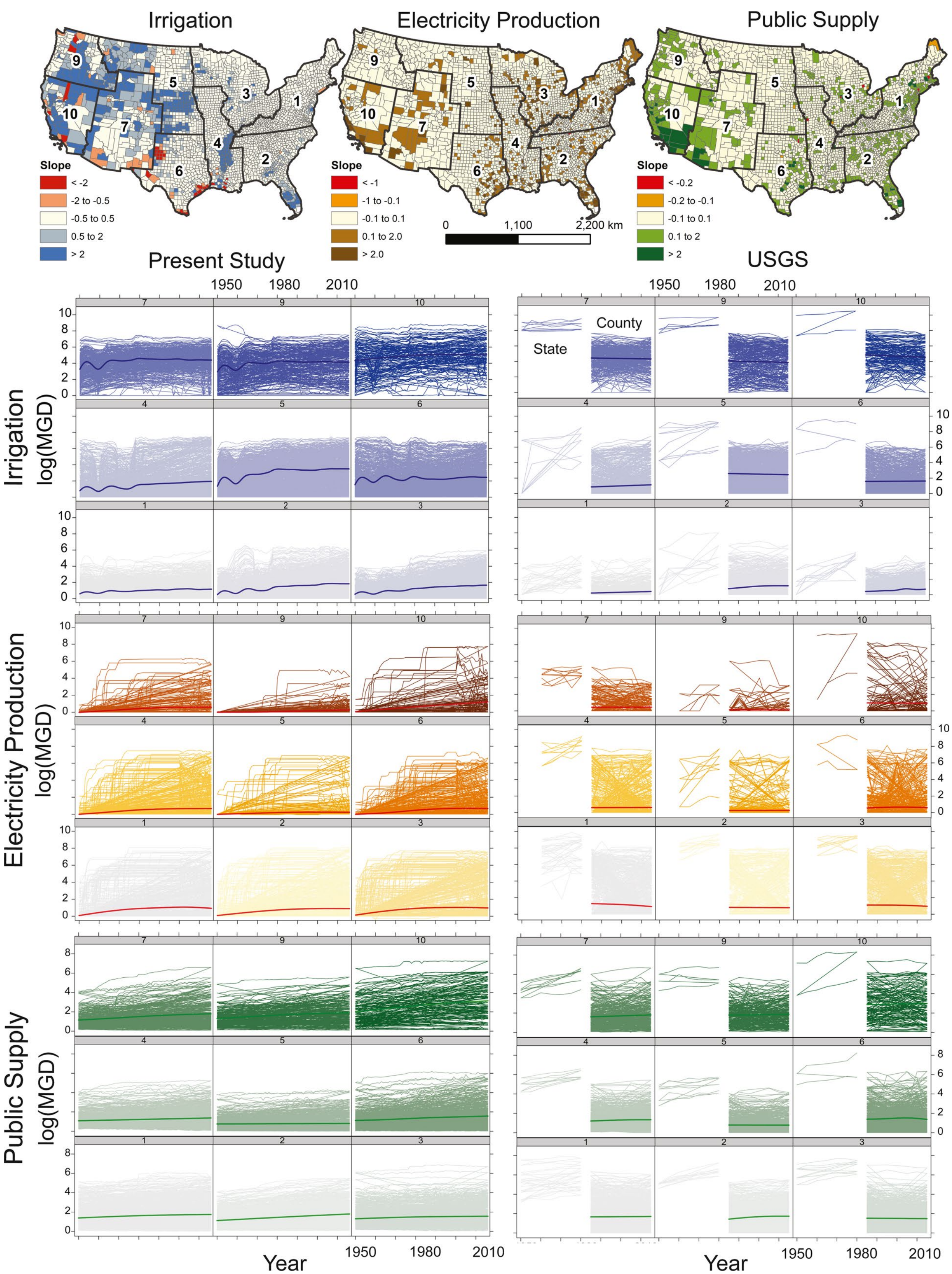


Figure 3. Estimated water use at power plants for selected years from this study based on the medium coefficient scenario. For comparison, data from the USGS (M. A. Harris & Diehl 2017) are provided. MGD, million gallons per day; USGS, US Geological Survey.

At the power plant level, we observed more agreement between our data and water use estimates provided by M. A. Harris and Diehl (2017) and Peer et al. (2016) (Figure 7). Average, minimum, and maximum water use estimated for power plants showed relatively strong agreement with Harris and Diehl's modeled estimates ($r^2 = 0.73$, $r^2 = 0.76$, $r^2 = 0.56$, respectively, Figure 7a). Likewise, our water withdrawal and consumption estimates for power plants were strongly related to Peer et al. (2016) ($r^2 = 0.72$, $r^2 = 0.53$, respectively, Figure 7b). However, these strong relationships only correspond to power plants represented in both our analysis and that of M. A. Harris and Diehl (2017) and Peer et al. (2016). We documented a total of 1,208 and 5,603 power plants operating in 2010 with water usage requirements that were absent from M. A. Harris and Diehl (2017) and Peer et al. (2016) (Table S9). Most of these plants were wind, natural gas, solar, biomass,


Irrigation
Electricity Production
Public Supply
Slope
< -2
-2 to -0.5
-0.5 to 0.5
0.5 to 2
> 2
< -1
-1 to -0.1
-0.1 to 0.1
0.1 to 2.0
> 2.0
< -0.2
-0.2 to -0.1
-0.1 to 0.1
0.1 to 2
> 2
0
1,100
2,200 km
Present Study
USGS
State
County
Irrigation log(MGD)
Electricity Production log(MGD)
Public Supply log(MGD)
Year
1950
1980
2010

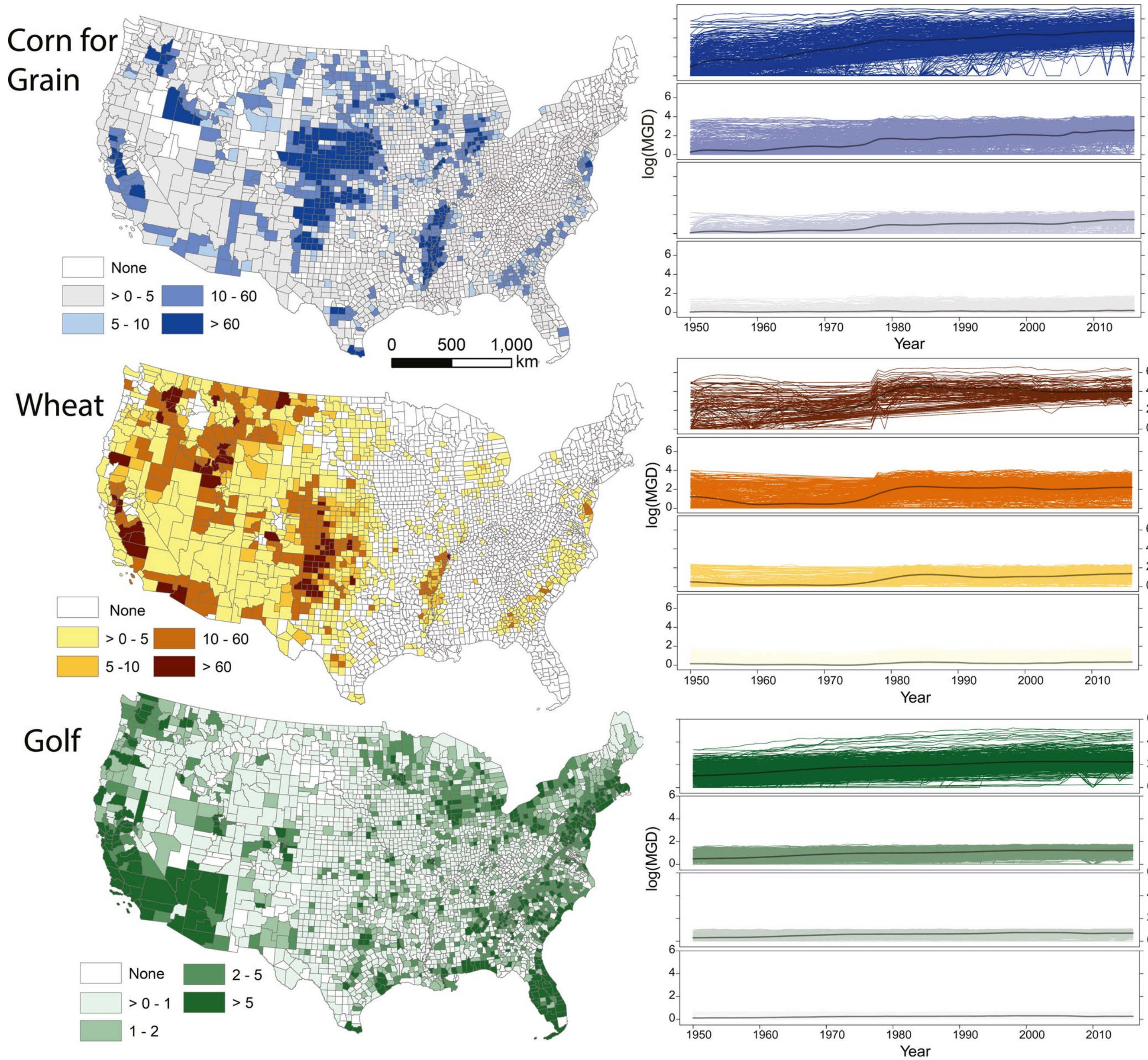


Figure 5. Example of spatiotemporally rich data provided by the study using irrigation water use estimates for two agricultural crops and for golf courses. Maximum value ranges are provided in US county-level maps on left. Color spectra for each map are associated with the color spectra of temporal trends within panels on the right. MGD, million gallons per day.

and coal, but only represent 2% of the total estimated water usage from the electric power production sector (Table S9).

We also observed general agreement between electricity production water use in our study and that of national estimates provided by E. A. Grubert and Sanders (2018) (Figures S7–S9). Our nationwide water withdrawal and consumption estimates for electricity production mirrored those of Grubert and Sanders across all fuels, except water use for wind energy (Figure S7). Likewise, our estimates of withdrawal and

Figure 4. Spatiotemporal analytics of sectoral water use patterns within regions of the United States. Maps display average slopes of long-term (1950–2016) changes in water use per county (top). Panels display temporal trends of annual sectoral water use estimates within counties from this study (left) and quinquennial water use estimates within states and counties from USGS Water Use Circular Series (right). County-level data from the USGS are only available post-1985, prior to which only state data is available. USGS, USGS, US Geological Survey.

Table 5
Adjusted r^2 Values Examining Agreement Between Water Use Estimates From This Study to Those Reported by the USGS

	States		Counties	
Water use variable	All	Remove 0s	All	Remove 0s
	Public supply			
Total water withdrawal (WW)	0.91	–	0.96	–
Surface WW	0.84	–	0.93	–
Ground WW	0.94	–	0.96	–
Electricity production[b]				
Total water withdrawal (WW)[b]	0.55	0.52	0.63	0.44
Surface fresh WW[b]	0.68	0.66	0.61	0.53
Ground fresh WW[b]	0.32	0.31	0.26	0.28
Surface saline WW[b]	0.61	0.60	0.67	0.67
Ground saline WW[b]	0.02	0.02	0.16	0.23
Total water consumption[b]	0.07	0.21	0.40	0.21
Irrigation				
Total water withdrawal (WW)	0.83	0.82	0.84	0.84
Surface fresh WW	0.83	0.82	0.84	0.84
Ground fresh WW	0.84	0.83	0.87	0.86
Total crop WW[a]	0.14	0.91	0.36	0.80
Surface fresh WW[a]	0.24	0.91	0.45	0.82
Ground fresh WW[a]	0.18	0.91	0.38	0.82
Total golf course WW[a]	0.30	0.67	0.53	0.67
Surface fresh WW[a]	0.23	0.26	0.48	0.57
Ground fresh WW[a]	0.22	0.32	0.44	0.56

Note. Comparisons for thermoelectric water use are based on water use estimated using medium coefficients reported by Averyt et al. (2013) (see Table 1). Comparisons for irrigation are based on water withdrawal from the the low acreage, climate-adjusted scenario (IWUc, see Table 1). USGS data are reported every 5 years. With some exceptions, comparisons include 13 years of state-level data spanning 1950–2015 and 6 years of county-level data spanning 1985–2015 for counties. Comparisons were conducted using all data and then after removing 0 values. Data were log(x+1) transformed prior to analysis. Results of comparisons between the USGS data and all scenarios from this study are provided in supporting information, Tables S4–S7.
[a]State and county data only available from 2005 to 2015.
[b]State level data only available from 1960 to 2015.

consumption by water source and fuel type agreed generally with that of Grubert and Sanders, with some exceptions (Figures S8–S9). Grubert and Sanders partitioned water source into two nonmutually exclusive classifications: (1) freshwater, brackish or saline and (2) groundwater, surface water, or reuse. These are analogous, but not directly translatable, into the seven mutually exclusive water source categories we report for electricity production (Table 1). Because our analysis does not discriminate brackish and saline, these were combined into one category (saline). In addition, for comparison, we presume that all reclaimed discharge (i.e., reuse) reported by Grubert and Sanders was also classified as either brackish or saline in their study. Generally, our withdrawal and consumption estimates for all fuel technologies showed agreement with Grubert and Sanders, except Solar PV, where our estimates included saline and surface water sources, whereas Grubert and Sanders reported none (Figures S8–S9). Additionally, for some fuels, our saline estimates showed divergence from those of Grubert and Sanders. This could be related to differences in how water sources were defined, particularly "mixed" sources reported in our study (unreported in Grubert and Sanders). These included combinations of surface and groundwater and/or combinations of fresh and

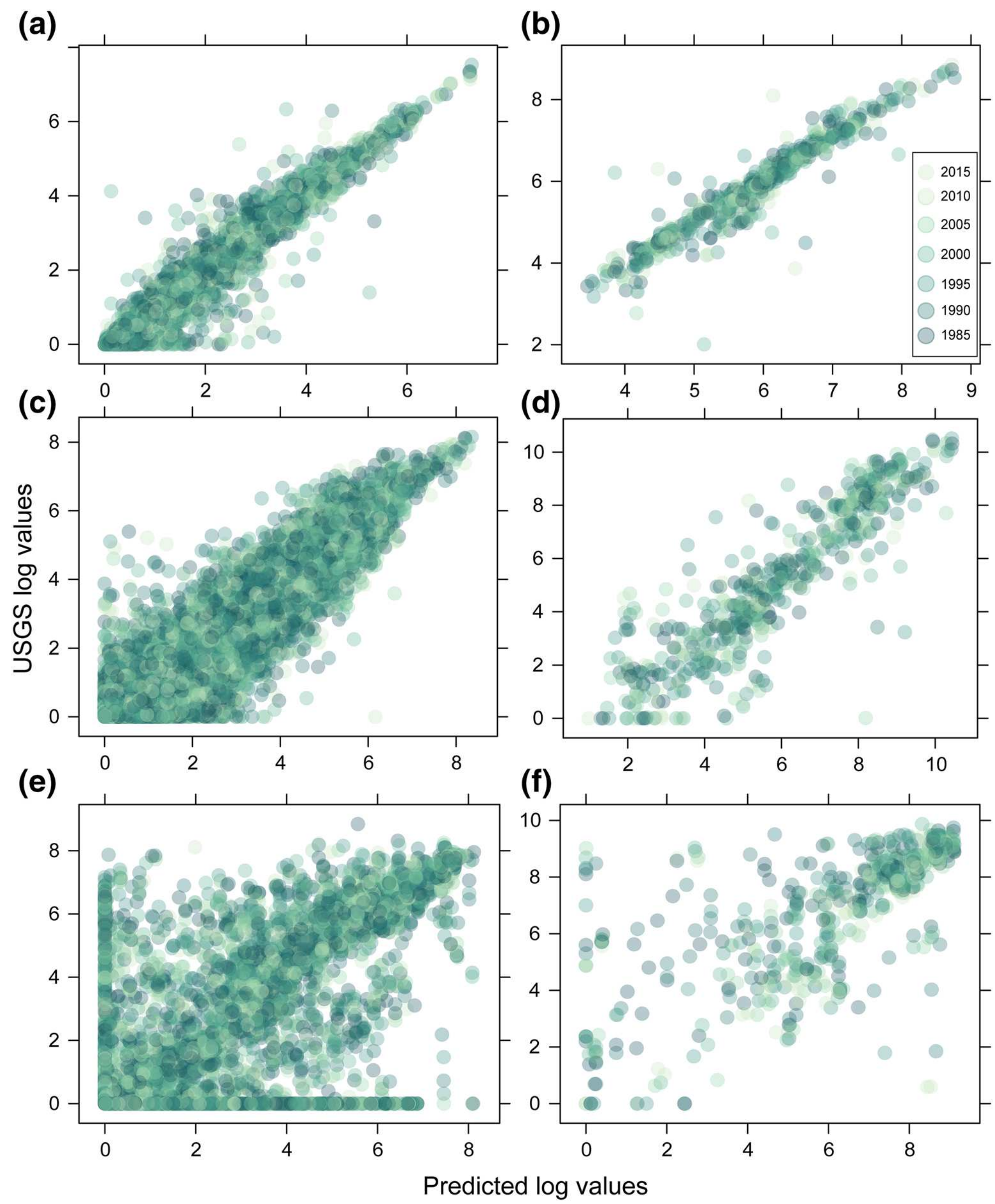


Figure 6. Comparisons of water use estimates (log transformed 10^6 g d^{-1}) from this study (predicted) versus those provided by the USGS. Comparisons include public water supply water use at the (a) county and (b) state levels, total irrigation water use (based on low acreage, climate-adjusted scenario) at the (c) county and (d) state levels, and electricity production water use (based on medium coefficient estimates) at the (e) county and (f) state levels. USGS, US Geological Survey.

saline water used for thermoelectric cooling, which were likely reported as brackish or saline in Grubert and Sanders.

Data from our study also aligned well with independent studies on irrigation and public supply. Our crop irrigation estimates were strongly associated with values reported by the USDA Agriculture Census at the state level (r^2 = 0.93, Figure S10) and with values reported by Marston et al. (2018) at the county level (Figure 8). Comparisons with Marston et al.'s study included crop irrigation estimates for total irrigation (blue) water use (r^2 = 0.84), irrigation use from groundwater sources (r^2 = 0.88), irrigation from surface water (r^2 = 0.84), and rainfall use by irrigated crops (green) (r^2 = 0.76). Because Worland et al. (2018) used USGS estimates of public supply in 2010 for their model, we compared our data to that of USGS for the same year.

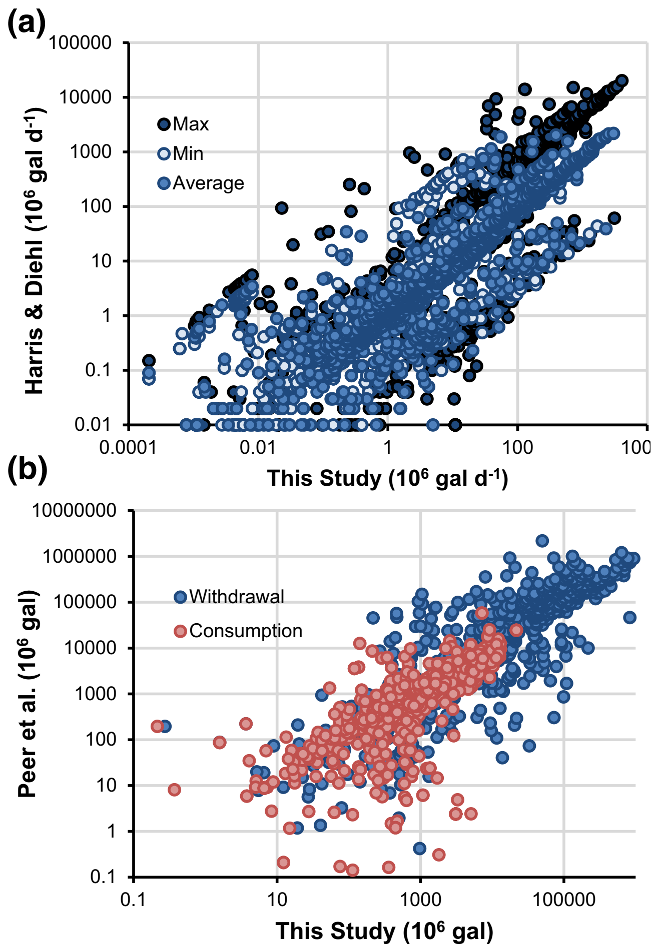


Figure 7. Comparison of electricity production water use estimates for individual power plants between our study and other studies, including (a) water withdrawal estimates from our study and M. A. Harris & Diehl (2017) and (b) average water withdrawal and consumption estimates from our study and Peer et al. (2016).

In addition, Worland et al.'s model predictions were not readily available. Our total public supply estimates and calculated *wh* values displayed strong associations with USGS 2010 county-level estimates ($r^2 = 0.97$ and $r^2 = 0.87$, respectively) (Figure S11).

Differences between our estimates and that of the USGS or other studies could arise for multiple reasons, the most obvious being slight differences in methods and sources of information. For example, related to electricity production, M. A. Harris and Diehl (2017) compare three different methods employed by federal agencies for estimating thermoelectric water usage in the US: (1) a highly detailed model approach (USGS) (Diehl et al., 2013), (2) reported estimates by power companies (EIA, 2019a; 2019b), and (3) a compilation approach heavily reliant on withdrawal coefficients for different fuel-cooling system combinations (Maupin et al., 2017). The study found that for over 50% of plants, the maximum estimated withdrawal was at least twice the magnitude of the minimum estimate. The USGS detailed model approach likely represents the most accurate withdrawal estimates, as it ensures estimates are thermodynamically plausible. However, this approach requires significant effort, including heat-and-water budgets and fuel consumption in relation to cooling technologies and local climate information (Diehl et al., 2013). Hence, this level of effort would be impractical for all the years included in our analysis 1950–2016. In contrast, our electricity production water-use estimates relied on a coefficient-type approach similar to that of approach 3 listed above. This compilation approach is most commonly used to generate county-level estimates reported by the USGS (M. A. Harris & Diehl, 2017). Interestingly, our estimates showed more agreement with modeled estimates at the power plant level than the USGS county and state compilations.

Our irrigation estimates differed from that of the USGS most likely because of differences in sources of information, primarily related to irrigated acreage. Estimates of irrigated acreage range widely among different sources, such as the USGS National Water Use Information Program (NWIP), the USDA Farm and Rach Irrigation Survey, and the USDA Census of Agriculture (Dickens et al., 2011). USGS NWIP typically has the highest estimates of irrigated agriculture from remote sensing. Our irrigation estimates rely on acreage from the Census of Agriculture, as these data provided the most temporally comprehensive period available. Regardless of these differences, our irrigation water use was strongly associated to USGS (Figure 6c and Table 5), Marston et al. (2018) (Figure 8), and that of the USDA Farm and Ranch Irrigation Survey (Figure S7).

2.7. Limitations

Users of our data should be aware of several limitations of our approach. An important consideration is that our data are constructed for 67 years and intended to be used for long-term analyses at county-to-regional scales; therefore, the data are not suitable for single-year analysis or detailed assessments of water use at individual sites, such as power plants or individual farms. Partitioning water sources for each sector was limited by available information. Although EIA provided water sources for individual power plants, water source information was missing for many power plants and required using proportions of water sources from county-level data. For irrigation and power plants lacking water source information, we relied on backcasting proportions of use by water sources reported by the USGS, which were only available post-1985. In the least, our interpolation approach can detect and emulate long-term shifts in use among different water sources. Another caveat is that some of the raw data used in our study (e.g., crop acreage, water sources)

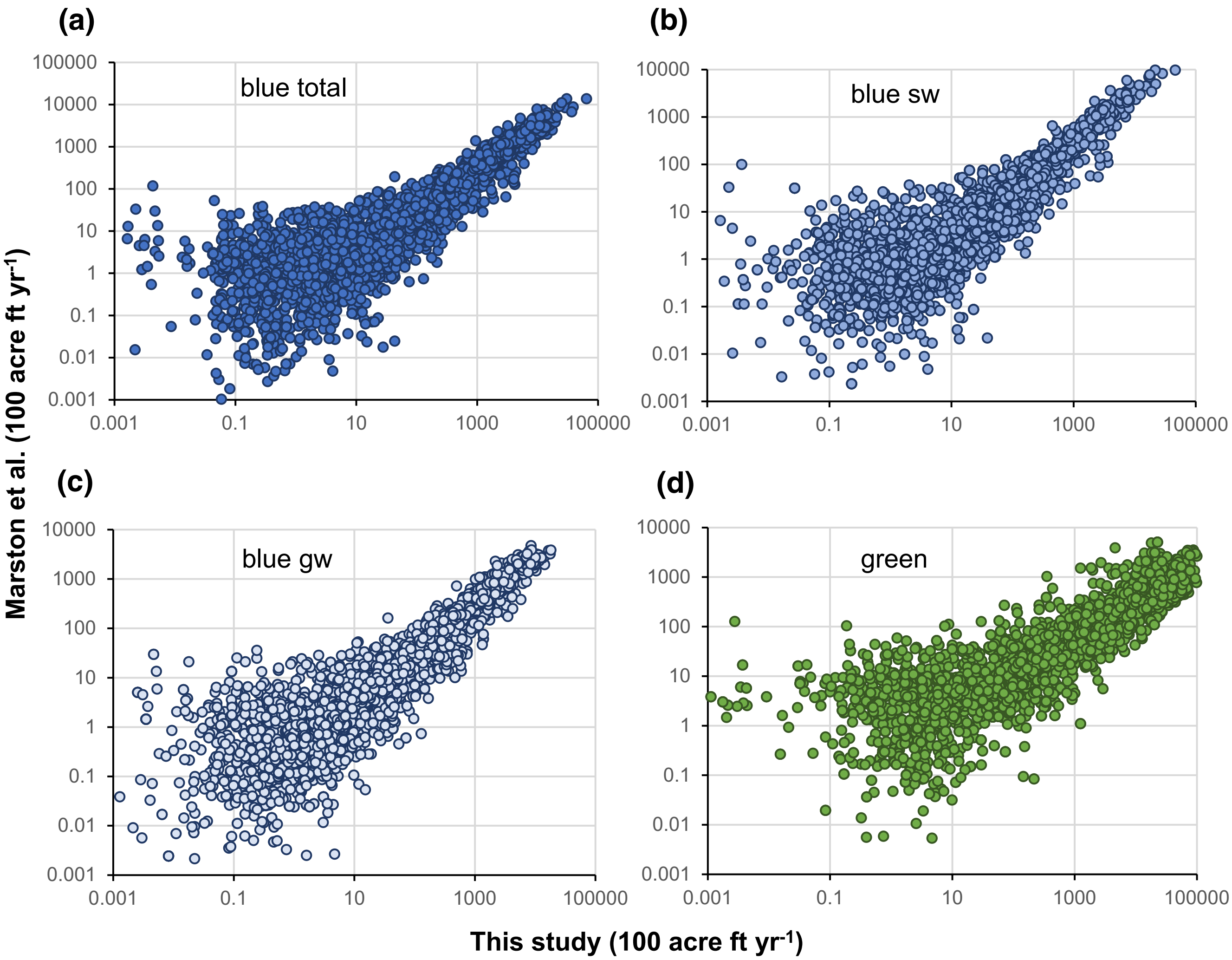


Figure 8. Comparison of 2010 county-level crop irrigation water use estimates between our study and that of Marston et al. (2018). Comparisons include total crop irrigation for (a) total blue water (fresh surface and groundwater), (b) blue surface water, (c) blue ground water, and (d) green water use (rainfall use by crops).

were either missing years of information or only available in 4 or 5-year increments. Developing annual estimates required that we interpolate between those incremental periods (e.g., Figure S3). Although spline interpolation can mimic the natural fluctuations in values, this approach may miss extreme episodic events out of the norm.

Electricity production water use estimates were limited by lack of information on generation at the EGU or power plant level for the pre-2000 period. Prior to 2000, we assumed capacity factors for a given EGU were constant (average of 2000–2016 values); hence, generation (MWh) was primarily based on EGU nameplate capacity and the documented month-years of operation within the period of 1950–2000. This coarse approach will miss periods in which entire EGUs are out of operation due to maintenance or longer curtailment periods. Although recent shifts in capacity factors have been noted for coal, nuclear, and natural gas technologies, most of these changes in capacity factors were less than 10% across the entire 2000–2016 period (Logan et al., 2017). Additionally, we presume that water use coefficients for EGUs (taken from Averyt et al.) have remained constant over time. Of course, this is not an accurate portrayal of increases in water use efficiency over time for different fuels, prime movers, and cooling technologies. Hence, our historical estimates, at times, may under-estimate actual withdrawal and consumption rates.

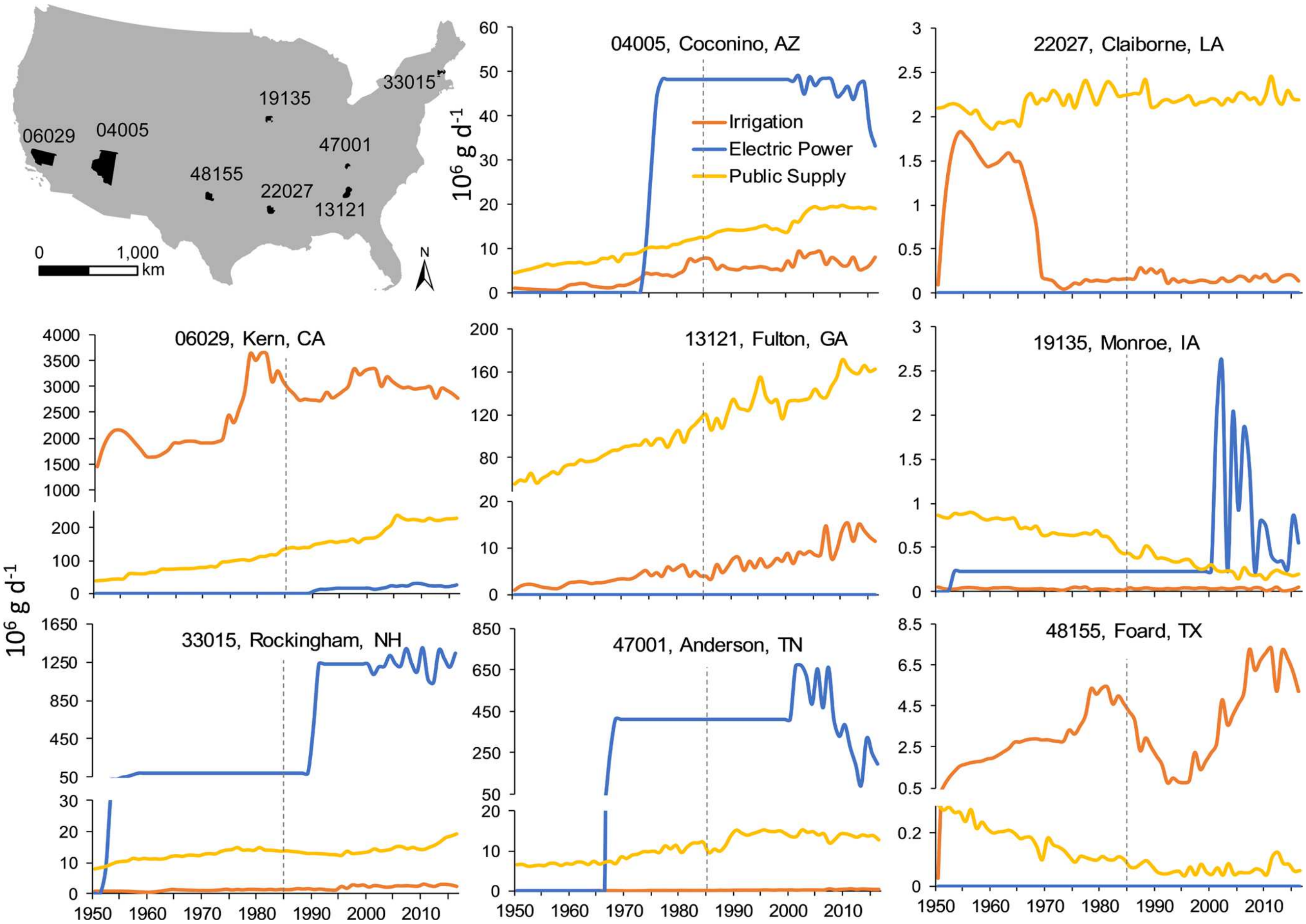


Figure 9. Example of the utility of this study's water use data in examining spatially and temporally rich patterns in water use among sectors for specific counties. The dashed line for year 1985 indicates the earliest date at which USGS data are available at the county level. Breaks in *y*-axis are used in some panels order to examine patterns amongst all sectors, which vary significantly in magnitude. USGS, US Geological Survey.

Whereas our irrigation and electric power water use relied on mechanistic model approaches, our public water supply use relied on a statistical model approach. While the statistical model was highly accurate, it embodies a black box approach agnostic to shifts in social behavior and efficiency that influence usage rates (Worland et al., 2018). Variables indicative of increasing infrastructure intensity, such as urban land cover and urban-rural gradients (represented by autocovariates), should provide surrogacy for efficiency; however, future efforts should focus toward building more mechanistic representations of public water supply that account for human agents.

2.8. Example Applications of the Data

We explored the data in a few ways to give potential users an idea of the versatility of the spatially rich, long-term product. Specifically, we highlight the advantages of our data in revealing patterns that would otherwise be un-noticeable through the USGS WUCS or data from single-year periods. We suggest our data are advantageous in that it: (1) extends further back in time at the county-level than USGS WUCS records, (2) is able to discern annual events invisible to quinquennial surveys, and (3) provides a new data product to support exploratory analytics for use in multi-sector dynamics research. In part, the benefits of such spatio-temporally rich data can be clearly seen in Figure 4. For a given sector or sub-sector, our data offers roughly 10X the volume of data available from the USGS Water Circular series.

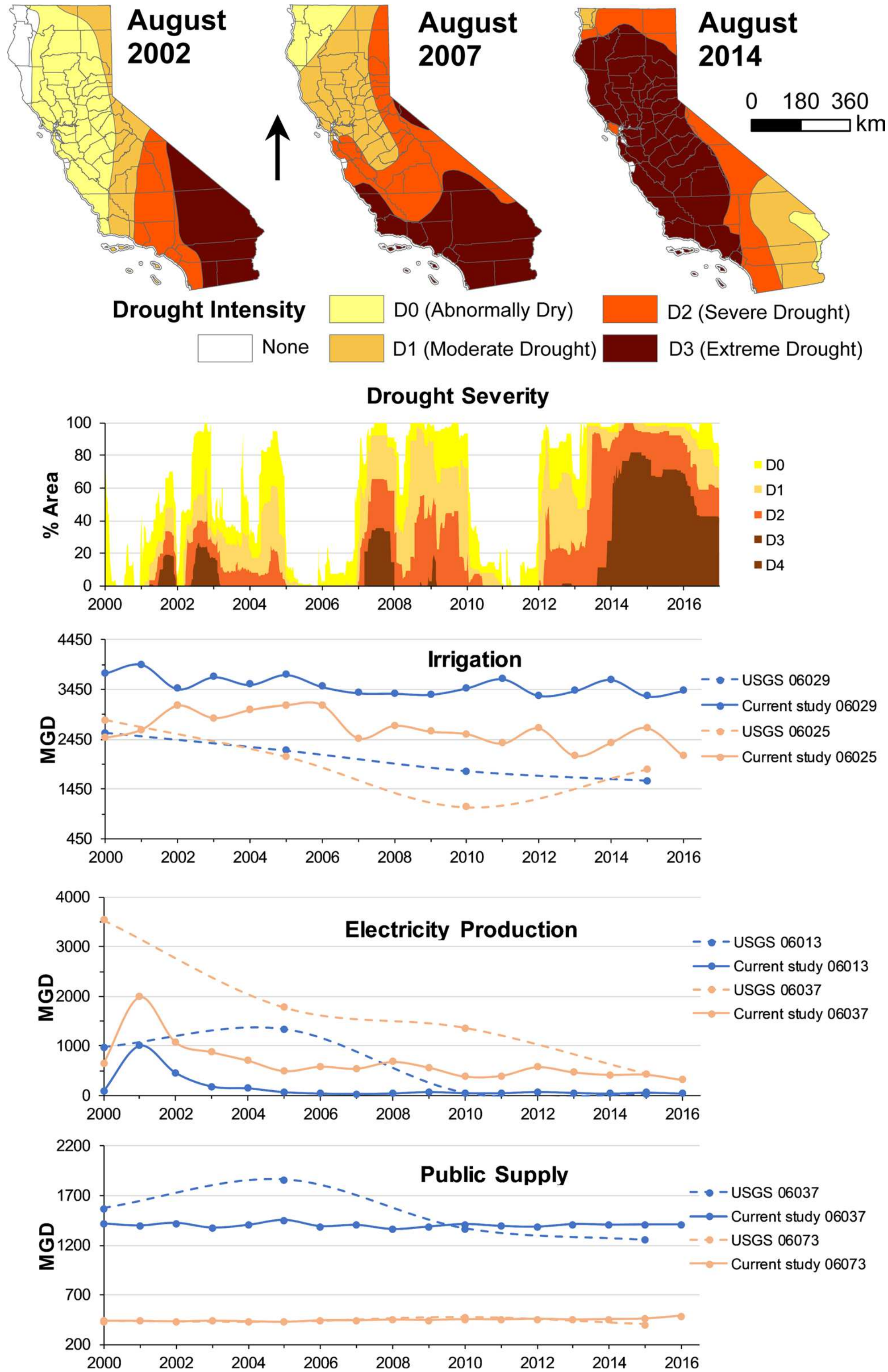


Figure 10. Utility of this study's water use data in providing high resolution temporal information, such as examining 16 years of water use during periods of drought. Drought intensity maps and spatial coverage data were obtained from The National Drought Mitigation Center (NDMC 2020). Data from our study provide more temporal granularity (annual) than that of USGS and display distinct patterns indicative of situations where water stress estimated from our results will be very different from that estimated from USGS data. USGS, US Geological Survey.

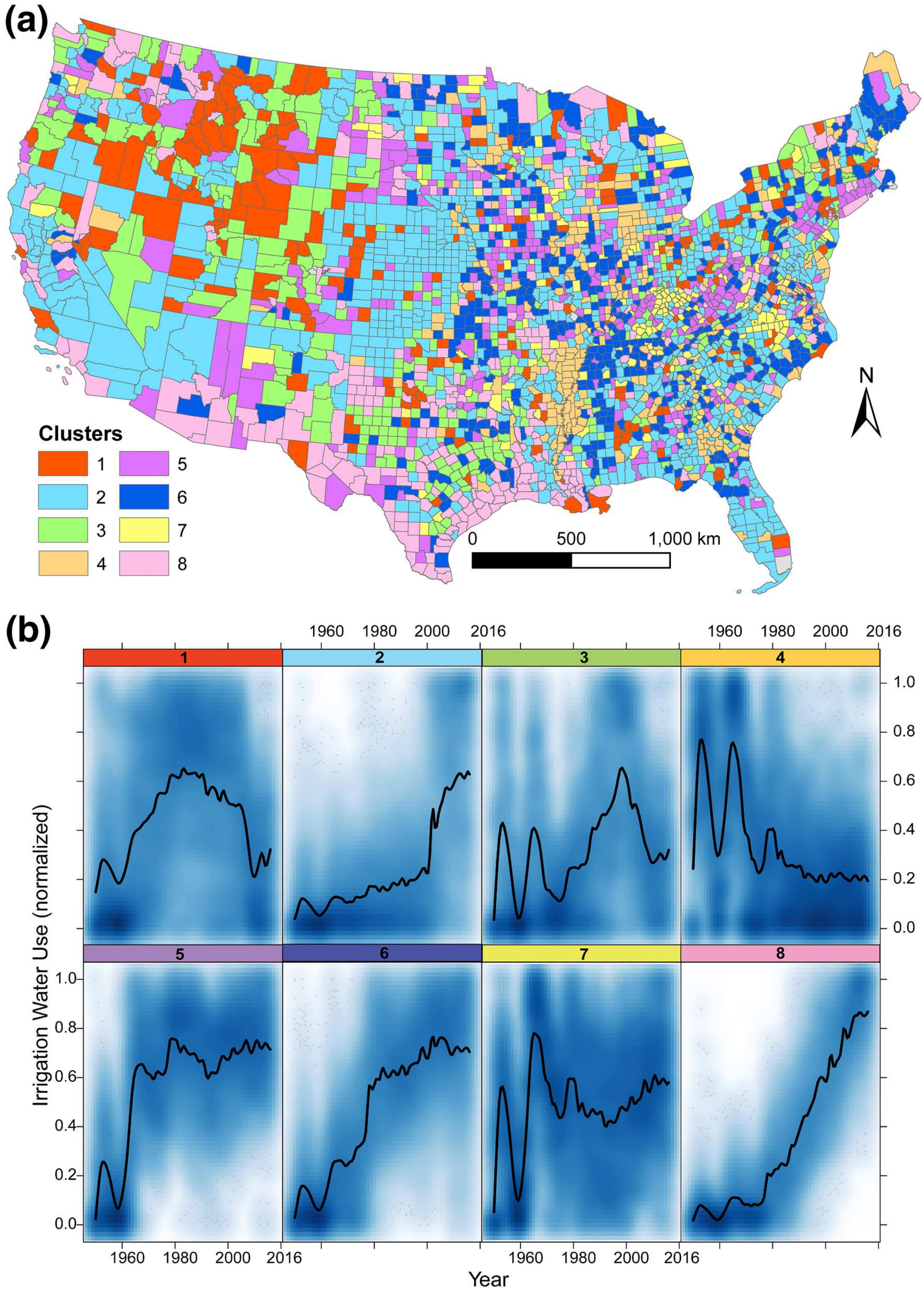


Figure 11. Example of the utility of this study's data set in examining nationwide patterns in irrigation water use. Dynamic time warping was applied to total irrigation water use to generate a distance matrix based on similarities in the chronology of temporal behaviors. The matrix was used in a clustering algorithm resulting in (a) clusters or groups of counties sharing similar water use behavior. (b) Mean behavior tendencies (bold lines) and heat maps of individual county behaviors falling into each cluster are provided.

As one example application, we selected counties that differed widely in their multisector water use behavior (Figure 9). We explored annual trends in county-level water withdrawals pre- and post-1985 as an indication of the rich information that our analysis adds to the prevailing sectoral water use data available only post-1985 from the USGS WUCS (Figure 9). The immediate observation of our data set is the high degree of inter-annual fluctuations, supporting the need for high temporal resolution of water use datasets. These trends provide a deeper perspective of shifting patterns in water use, or even tradeoffs in water use among sectors over time. Additionally, differential water use varies greatly among different counties, especially prior to 1985. For example, public water supply use continues to increase with expansive urbanization and population increases in Fulton County, GA (Atlanta), whereas public water supply use has consistently decreased in Monroe County, IA where population has declined since early 1900s due to reductions in coal mining operations. Comparison of pre- and post-1985 periods indicated that pre-1985 time periods, at times, displayed very different trends than the last 30 years. For instance, in Coconino County, AZ, electric power production water use was virtually nonexistent prior to 1976, before Navajo Coal-fired Generating plant became operational. Likewise, irrigation appears relatively stable post-1985 in Kern County, CA; however, irrigation peaked in the mid-1970s, before which it had been steadily increasing.

As another application of our data set, we explored how our data could be used in studies examining the compounded effects of widespread climatic stress, such as drought, on water use required for sustaining socioeconomic demands. Droughts in California have placed stress on agricultural communities to maintain levels of production resulting in compounded stresses on limited water availabilities (Marston & Konar, 2017). We obtained monthly drought intensity records for California from 2000 to 2016 from the NDMC (2010) and compared these patterns to sectoral water withdrawals reported by the USGS WUCS and our annual estimates for a few selected counties (Figure 10). Generally, relative magnitudes and long-term directionality (increasing/decreasing) for sector water withdrawals among counties showed consistencies with that of the USGS WUCS. However, our irrigation estimates for Kern County (06029, Central valley) and Imperial County (06025, Southeastern extreme) were very different from estimates reported by the USGS WUCS (Figure 10). Despite significant drought from 2012 to 2014, Marston and Konar (2017) report 3% increases in irrigation in California's Central Valley due to increased crop water requirements from higher temperatures. In agreement with their study, our data show short-term increases in irrigated water use across those years, at least for Kern county, and a delayed increase in Imperial county (Figure 10). Our irrigation estimates suggest water use in the face of shortages were more severe in drought situations than the USGS WUCS estimates indicate. Furthermore, our estimates tend to align well with irrigation magnitudes reported by California Department of Food and Agriculture (CDFA 2019). Interestingly, the opposite is true for electricity production, where our numbers suggest lower withdrawals than that reported by the USGS. However, our electricity production water use estimates are based on actual generation, which was directly obtained from EIA during this period. Many instances of reduced regional electricity production in response to drought have been documented (Harto et al., 2012), and our data provide the temporal resolution needed to explore these relationships.

As a final example, we use exploratory analytics to examine divergent long-term patterns in irrigation water withdrawal (i.e. behaviors) from 1950 to 2016 in counties across the CONUS. To examine similarities and differences in behaviors we employed Dynamic Time Warping (DTW), which finds optimal alignment between two time series through similarity measurements between chronological points in datasets (Kruskal & Liberman, 1983). Time series data may display similarities in chronological trends, but small differences in shifts along the temporal axis will result in misalignment and low similarity in conventional distance measurements (e.g., Euclidean distance). DTW overcomes this challenge by developing a “warping” path along the temporal axis, from which distance measures are minimized to align chronological patterns among different entities (Berndt & Clifford, 1994). We applied DTW to total irrigation water usage using the WSTAMP package in R (Piburn et al., 2017), which calculates a distance matrix based on dissimilarities in time-series data. We standardized water use data for each county from 0 to 1 so that trends would be purely based on behavior and agnostic to magnitudes. We then used hierarchical agglomerative clustering (Ward, 1963) to group counties based on minimal distances.

Clusters were highly divergent in their behavior regarding irrigation water use (Figure 11). For instance, irrigation water use increased dramatically during the entire period for members of cluster 1, whereas irrigation did not increase substantially until post-1990 and post-2000 in members of clusters 5 and 8, respectively (Figure 11). Other clusters displayed a plateauing with time (e.g., Clusters 2 and 6), whereas others showing erratic behavior pre-1980 (e.g., Clusters 4 and 7). While some clusters displayed geographical affiliation (e.g., Clusters 2 and 4), others were spread across the entire CONUS (e.g., Clusters 1, 5, and 6) (Figure 11). This information and type of analysis can be useful for examining groups of entities displaying similarities in long-term tradeoffs among water usage sectors or human adaptation strategies to water shortages.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability Statement

The data set developed in this study is freely available from the U.S. Department of Energy Integrated Multi-Sector, Multi-Scale Modeling (IM3) Data Hub: https://doi.org/10.25584/data.2020-12.1644/1735756.

Acknowledgments
This manuscript has been authored by employees of UT-Battelle, under contract DE-AC05-00OR22725 with the US Department of Energy. This research was supported by the U.S. Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. This work was also conducted as a part of the "Reanalyzing and Predicting U.S. Water Use using Economic History and Forecast Data; an experiment in short-range national hydro-economic data synthesis" Working Group supported by the John Wesley Powell Center for Analysis and Synthesis, funded by the USGS. We thank Casey Burleyson and Chris Vernon for posting the reanalysis data set to the IM3 website. We are also appreciative to five anonymous reviewers who provided comments on previous versions of this manuscript.

References

AEO Annual Energy Outlook (2019). *U.S. Energy information administration*. Retrieved from https://www.eia.gov/outlooks/aeo/pdf/aeo2019.pdf

Allen, M. R., Zaidi, S. M. A., Chandola, V., Morton, A. M., Brelsford, C. M., McManamay, R. A., et al. (2018). A survey of analytical methods for inclusion in a new energy-water nexus knowledge discovery framework. *Big Earth Data*, *2*(3), 197–227.

Averyt, K., Macknick, J., Rogers, J., Madden, N., Fisher, J., Meldrum, J., & Newmark, R. (2013). Water use for electricity in the United States: an analysis of reported and calculated water use information for 2008. *Environmental Research Letters*, *8*(1), 015001.

Barnett, T. P., & Pierce, D. W. (2008). When will Lake Mead go dry? *Water Resources Research*, *44*(3), W03201. https://doi.org/10.1029/2007WR006704

Berndt, D. J., & Clifford, J. (1994). *Using dynamic time warping to find patterns in time series. KDD Workshop* (Vol. 10, pp. 359–370). Retrieved from https://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf

Bivand, R., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods. *Geographical Analysis*, *45*(2), 150–179.

Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, *63*(18), 1–36. https://doi.org/10.18637/jss.v063.i18

Blaney, H. F., & Criddle, W. D. (1962). *Determining consumptive use and irrigation water requirements (No. 1275)*. US Department of Agriculture.

Bonnet, M., Witt, A., Stewart, K., Hadjerioua, B., & Mobley, M. (2015). *The economic benefits of multipurpose reservoirs in the United States – federal Hydropower Fleet*. Oak Ridge, TN: Oak Ridge National Laboratory.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Brouwer, C., & Heibloem, M. (1986). Irrigation water management: Irrigation water needs. *Training Manual*, *3*, Chapter 3: Crop Water Needs. Retrieved from http://www.fao.org/docrep/s2022e/s2022e07.htm

Brown, T. C., Mahat, V., & Ramirez, J. A. (2019). Adaptation to future water shortages in the United States caused by population growth and climate change. *Earth's Future*, *7*(3), 219–234.

Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783–2792.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *28*(15), 2031–2064. Retrieved from http://prism.oregonstate.edu

Das Gupta, P. (1981). *Intercensal estimates for the states, unpublished memorandum*. February 26, 1981. Cited in United States. Bureau of the Census. (1992). Current Population Reports: Population estimates and projections. Series P-25 (Issues 1091-1108) US Department of Commerce, Bureau of the Census.

Devineni, N., Lall, U., Etienne, E., Shi, D., & Xi, C. (2015). America's water risk: Current demand and climate variability. *Geophysical Research Letters*, *42*(7), 2285–2293.

Dickens, J. M., Forbes, B. T., Cobean, D. S., & Tadayon, S. (2011). *Documentation of methods and inventory of irrigation data collected for the 2000 and 2005 US geological survey estimated use of water in the United States, comparison of USGS-compiled irrigation data to other sources, and recommendations for future compilations*. US Geological Survey Scientific Investigations Report.

Diehl, T. H., Harris, M. A., Murphy, J. C., Hutson, S. S., & Ladd, U. S. (2013). *Methods for estimating water consumption for thermoelectric power plants in the United States, U.S. Geological Survey Scientific Investigations Report (No. 2013–5188, p. 78)* US Geological Survey. https://doi.org/10.3133/sir20135188

Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., et al. (2018). *Estimated use of water in the United States in 2015*. U.S. Geological Survey Circular. https://doi.org/10.3133/cir1441

Doyle, M. W., & Patterson, L. A. (2019). Federal decentralization and adaptive management of water resources: Reservoir reallocation by the U.S. Army Corps of Engineers. *Journal of the American Water Resources Association*, *55*, 1248–1267.

EIA U.S. (2019a). *Electricity information administration. Form EIA-860*. Retrieved from https://www.eia.gov/electricity/data/eia860/

EIA U.S. (2019b). *Electricity information administration. Form EIA-923*. Retrieved from https://www.eia.gov/electricity/data/eia923/

EIA U.S. (2019c). *Energy Information Administration. Thermoelectric cooling water data*. Retrieved from https://www.eia.gov/electricity/data/water/

Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., & Eisner, S. (2014). Constraints and potentials of future irrigation water availability on agricultural production under climate change. *Proceedings of the National Academy of Sciences*, *111*(9), 3239–3244.

FAO Food and Agriculture Organization of the United Nations (2016). *FAOSTAT database FAO*. Retrieved from http://www.fao.org/faostat

Grubert, E. A. (2016). Water consumption from hydroelectricity in the United States. *Advances in Water Resources*, *96*, 88–94.

Grubert, E. A., & Sanders, K. T. (2018). Water use in the United States Energy System: A national assessment and unit process inventory of water consumption and withdrawals. *Environmental Science and Technology*, *52*, 6695–6703.

Haines, M., Fishback, P., & Rhode, P. (2016). *United States agriculture data, 1840-2012. Study No. ICPSR35206-v3*. Inter-university Consortium for Political and Social Research.

Harris, M. A., & Diehl, T. H. (2017). A comparison of three federal datasets for thermoelectric water withdrawals in the United States for 2010. *Journal of the American Water Resources Association*, *53*(5), 1062–1080.

Harris, B. H., Hershbein, B., & Kearney, M. S. (2014). *Tidal wave or drop in the bucket? Differences in water use across the United States*. The Hamilton Project.

Harto, C. B., Yan, Y. E., Demissie, Y. K., Elcock, D., Tidwell, V. C., Hallett, K., et al. (2012). *Analysis of drought impacts on electricity production in the Western and Texas interconnections of the United States*. Argon National Laboratory ANL/EVS/R-11/14. https://doi.org/10.2172/1035461

Hoekstra, A. Y. (2019). Green-blue water accounting in a soil water balance. *Advances in Water Resources*, *129*, 112–117.

Hoekstra, A. Y., & Mekonnen, M. M. (2012). The water footprint of humanity. *Proceedings of the National Academy of Sciences*, *109*(9), 3232–3237.

Ivahnenko, T. (2009). *Estimated Colorado golf course irrigation water use, 2005 (No. 2008-1267)*. US Geological Survey.

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., & Maupin, M. A. (2009). *Estimated use of water in the United States in 2005 (No. 1344)*. US Geological Survey.

Kruskal, J., & Liberman, M. (1983). *The symmetric time warping problem: From continuous to discrete*. In Time warps, string edits and macromolecules: The theory and practice of sequence comparison (pp. 125–161). Reading, MA: Addison-Wesley.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. *Scientific Data*, *2*, 150042. https://doi.org/10.7289/v5x34vf6

Logan, J., Marcy, C., McCall, J., Flores-Espino, F., Bloom, A., Aabakken, J., et al. (2017). *Electricity generation baseline report. NREL/TP-6A20-67645*. Golden, CO National Energy Technology Laboratory

Macknick, J., Newmark, R., Heath, G., & Hallett, K. C. (2011). *Review of operational water consumption and withdrawal factors for electricity generating technologies (No. NREL/TP-6A20-50900)*. Golden, CO (United States): National Renewable Energy Lab. (NREL).

Macknick, J., Newmark, R., Heath, G., & Hallett, K. C. (2012). Operational water consumption and withdrawal factors for electricity generating technologies: A review of existing literature. *Environmental Research Letters*, *7*(4), 045802.

Marston, L., & Konar, M. (2017). Drought impacts to water footprints and virtual water transfers of the Central Valley of California. *Water Resources Research*, *53*(7), 5756–5773. https://doi.org/10.1002/2016wr020251

Maupin, M. A., & Barber, N. L. (2005). *Estimated Withdrawals from Principal Aquifers in the United States, 2000 (Vol. 1279)*. Geological Survey (USGS). Retrieved from https://pubs.usgs.gov/circ/2005/1279

Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2017). *Estimated use of water in the United States in 2010*. US Geological Survey.

Marston, L., Ao, Y., Konar, M., Mekonnen, M. M., & Hoekstra, A. Y. (2018). High-resolution water footprints of production of the United States. *Water Resources Research*, *54*(3), 2288–2316. https://doi.org/10.1002/2017wr021923

Mekonnen, M. M., & Hoekstra, A. Y. (2012). The blue water footprint of electricity from hydropower. *Hydrology and Earth System Sciences*, *16*, 179–187.

Mesinger, F., DiMego, G., Eugenia Kalnay, E., Kenneth Mitchell, K., PerryShafran, C. P. C., et al. (2006). North American regional reanalysis. *Bulletin of the American Meteorological Society*, *87*, 343–360.

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207–218. https://doi.org/10.32614/RJ-2017-009

NASS & USDA (2010). Field Crops: Usual Planting and Harvesting Dates (Vol. 628). Agricultural Handbook: USDA National Agricultural Statistics Service. Retrieved from https://downloads.usda.library.cornell.edu/usda-esmis/files/vm40xr56k/dv13zw65p/w9505297d/planting-10-29-2010.pdf

NASS & USDA (2020). *United States department of agriculture, national agricultural statistics service*. Retrieved from https://www.nass.usda.gov/index.php

NCISEER (2019). *National cancer Institute surveillance, Epidemiology and end results Program U.S. Population data - 1969-2017*. Retrieved from https://seer.cancer.gov/popdata/download.html

NDMC (The National Drought Mitigation Center) (2020). *United States drought monitor*. Retrieved from https://droughtmonitor.unl.edu/CurrentMap.aspx

Ostroff, S., Wieferich, D. J., Cooper, A., & Infante, D. (2013). *National anthropogenic barrier dataset (NABD)*. Denver, CO: US Geological Survey—Aquatic GAP Program.

Peer, R. A. M., Grubert, E., Kelly, T., & Sanders (2019). A regional assessment of the water embedded in the US electricity system. *Environmental Research Letters*, *14*(8), 084014.

Peer, R. A. M., & Sanders, K. T. (2016). Characterizing cooling water source and usage patterns across US thermoelectric power plants: a comprehensive assessment of self-reported cooling water data. *Environmental Research Letters*, *11*(12), 124030.

Piburn, J., Stewart, R., Myers, A., Sorokine, A., Axley, E., Anderson, D., et al. (2017). *The world spatiotemporal analytics and mapping project (WSTAMP): Further progress in discovering, exploring, and mapping spatiotemporal patterns across the world's largest open source data sets*. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences (Vol. 4).

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*, 5551–5558. https://doi.org/10.5194/hess-22-5551-2018

Rushforth, R. R., & Ruddell, B. L. (2018). A spatially detailed blue water footprint of the United States economy. *Hydrology and Earth System Sciences*, *22*, 3007.

Sankarasubramanian, A., Sabo, J. L., Larson, K. L., Seo, S. B., Sinha, T., Bhowmik, R., et al. (2017). Synthesis of public water supply use in the United States: Spatio-temporal patterns and socio-economic controls. *Earth's Future*, *5*(7), 771–788. https://doi.org/10.1002/2016EF000511

Sohl, T., Reker, R., Bouchard, M., Sayler, K., Dornbierer, J., Wika, S., et al. (2016). Modeled historical land use and land cover for the conterminous United States. *Journal of Land Use Science*, *11*(4), 476–499.

Sohl, T. L., Sayler, K. L., Bouchard, M. A., Reker, R. R., Friesz, A. M., Bennett, S. L., (2014). Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States. *Ecological Applications*, *24*(5), 1015–1036.

Taylor, R., Parker, D., & Zilberman, D. (2014). Contribution of university of California cooperative extension to drip irrigation. *ARE Update*, *18*(2), 5–8.

Thornton, M. M., Thornton, P. E., Wei, Y., Mayer, B. W., Cook, R. B., & Vose, R. S. (2018). *Daymet: Annual climate summaries on a 1-km grid for north America Version 3*. Oak Ridge, TN, USA: ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1343

Uria-Martinez, R., Johnson, M., O'Connor, P., Samu, N., Witt, A., Battey, H., et al. (2018). *2017 hydropower market report, United States*. Retrieved from https://www.energy.gov/eere/water/downloads/2017-hydropower-market-report

US Golf Course Database (2019). Retrieved from http://www.coursedatabase.com/license/

U.S. Census Bureau (1996). *Population of states and counties of the United States: 1790 to 1990.* Retrieved from https://www.census.gov/population/www/censusdata/pop1790-1990.html

U.S. Census Bureau (2019). *Population projections*. Retrieved from https://www.census.gov/programs-surveys/popproj.html

USDA United States Department of Agriculture (2019). *Agricultural projections to 2028 interagency agricultural projections committee*. Retrieved from https://www.ers.usda.gov/webdocs/publications/92600/oce-2019-1.pdf?v=3630.9

USGS United States Geological Survey (2019). *Water use in the United States*. Retrieved from https://www.usgs.gov/mission-areas/water-resources/science/water-use-united-states?qt-science_center_objects=0#qt-science_center_objects

USGS (U.S. Geological Survey) (2018). *WaterWatch. Computed Runoff in hydrologic units*. Retrieved from https://waterwatch.usgs.gov/?id=romap3

van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and Hall/CRC.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Vliet, M. T., Yearsley, J. R., Ludwig, F., Vögele, S., Lettenmaier, D. P., & Kabat, P. (2012). Vulnerability of US and European electricity supply to climate change. *Nature Climate Change*, *2*(9), 676.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *Science*, *289*(5477), 284–288.

Ward, J. H., Jr (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Wolock, D. M., Winter, T. C., & McMahon, G. (2004). Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental Management*, *34*(1), S71–S88.

Worland, S. C., Steinschneider, S., & Hornberger, G. M. (2018). Drivers of variability in public-supply water use across the contiguous United States. *Water Resources Research*, *54*(3), 1868–1889. https://doi.org/10.1002/2017wr021268