Title
Characterizing the spatial-temporal distribution of California’s agricultural water utilization using a water footprint analysis in R

Author
Booth, Lorenzo

Publication Date
2018

University of California, Merced

Characterizing the spatial-temporal distribution of California's agricultural water utilization using a water footprint analysis in R

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Environmental Systems

by

Lorenzo Ade Booth

Committee in charge:
Prof. Joshua Viers
Prof. Thomas Harmon
Prof. Shawn Newsam

2018

The format of this thesis drew heavy inspiration from the "Masters/Doctoral Thesis LaTeX Template Version 2.1 (2/9/15)"[1]

[1] V2: vel@latextemplates.com, Johannes Böttcher. V1: Steven Gunn, Sunil Patel

The thesis of Lorenzo Ade Booth is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Thomas Harmon

Shawn Newsam

Joshua Viers (Chair)

University of California, Merced
2018

To my wives, Lenovo Thinkpad X201t (2013-2016) and Lenovo Thinkpad T430 (2014-2018).

Abstract

Different activities utilize water resources at different rates, rely on different water sources, and thus have a different "footprint" in regard to their water demand. However, the water footprint of these activities differs depending on the region and the time when the activity occurs. In popular literature, the footprint metric has seen some success in the ability to represent the large magnitude of human resource consumption, to more relatable per-capita quantities. However, these depictions are often based on global surveys that mask regional variability in water use. This thesis presents a framework for agricultural water footprint assessment, implemented using free and open source software. It explores the variability of the water footprint of select agricultural activities in the State of California from 2008 through 2015. The results describe a diverse landscape of water use variability driven primarily by crop choice. The study reveals that crop specific water footprints can have significant, inter-annual variability, independent of climatic conditions. Supported by water distribution infrastructure, agricultural activities can be found across the state. Water footprint assessments can be used by growers and resource managers, who wish to maximize utility per unit of water allocated, planners who wish to understand national risks and strengths, or the informed citizen who wishes to evaluate the sustainability of their consumptive activities.

Acknowledgements

I am deeply grateful for the encouragement and patience of my supervisor Dr. Joshua Viers. I also thank Dr. Leigh Bernachi deeply for the support and encouragement. I am thankful for the advice and time investment of my committee. I also owe a large debt of gratitude to all of the people who have ever turned a helpful ear in my direction, or who have gone out of the way to offer me assistance, praise, or criticism. Friends and family alike, thank you.

Most importantly, I am thankful for all of the developers who wrote the free/libre software that enabled this research. I'm also grateful for the wealth of public information maintained by the State of California. I hope this study will benefit at least one person in the future.

I also would like to acknowlege support from the University of California Office of the Persident's Multi-Campus Research Programs and Initiatives (MR-15-328473) through UC Water, the University of California Water Security and Sustainability Research Initiative.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Since 2014, the World Economic Forum has identified water crises in the top-five global risks to society in terms of impact, based on a broad survey of risk perception among participants in business, academia, international organizations, and government (World Economic Forum, 2018). Where food and freshwater are scarce, they often become central to social stability. Because fluid nature, freshwater systems often cross political boundaries, allocation of this scarce resource has been a source of numerous conflicts throughout history (Gleick, 1993).

At the margin, our thirst for this resource becomes flexible and yields goods, services, and other productive activities (Zetland, 2014). When these activities take place at times and scales far removed from their input resources, it can be useful to quantitatively evaluate the inputs, outputs, and environmental impacts of a product or activity, over its entire life cycle. Life-cycle assessment (LCA) can refer to both the standardized method of evaluating these impacts, or the result of the method (Guinée and Heijungs, 2017; ISO/TC 207/SC 5, 2006). Related to LCA are a family of "footprints", sustainability indicators that relate economic activities to physical quantities of resources consumed (see Chapter 2 for more details). Carbon footprints are a popular concept, that describe quantities of greenhouse gas emissions (e.g. tonnes of CO^2 equivalent) required to enable a consumptive activity (Minx et al., 2009). Ecological footprints relate consumptive activities to the biosphere's regenerative capacity, which can be expressed in terms of the annual productive capacity of a unit of land area (e.g. arable land, forest, or fishing grounds) (Wackernagel et al., 1999). Most industrial activities, are unable to be entirely supported by the natural resources of adjacent lands, and the remainder of the ecological footprint is met through appropriations from other regions. Global ecological footprint assessments have calculated that the rate of current human resource consumption exceeds the rate of natural production by over 1.5 times (Ewing et al., 2010). This description does not forecast when the rate of overshoot will deplete global stocks, but it does serve a benchmark for evaluating the sustainability of human activities.

Water footprints (WF) were inspired by ecological footprints and describe the quantities of water (e.g. gallons of water) required to enable different consumptive activities (Hoekstra, 2017; Hoekstra et al., 2011). Water footprints are closely related to the virtual water concept, which describes the presence of virtual volumes of water traded between nations in the form of water embodied in goods or services (Allan, 2003; Allan, 1997). Alan described the option of importing virtual water as a means for water scarce nations in the Middle East to alleviate pressure on domestic water resources. The water footprint concept extends this metaphor with an allusion to ecological footprints, by describing the volumes of water required to enable different consumptive activities, or societies as a whole (A.K. Chapagain and A.Y. Hoekstra, 2004).

In popular literature, the footprint metric has seen some success in the ability to relate the large magnitude of human resource consumption, to more relatable per-capita quantities (Kim, 2015a; Kim, 2015b). It has also earned criticism outside of the water resources field from the aesthetic perspective of over-burdening the footprint metaphor and from the practical perspective of having questionable utility for policy support (Safire, 2008; Wichelns, 2004). More generally, the family of life-cycle sustainability indicators are derivative metrics, which accumulate uncertainty from the numerous observations and estimations that are incorporated into the final assessment (Heijungs, R. and Huijbregts, M.A.J., 2004; Lloyd and Ries, 2007). Recent efforts aimed at establishing, standards for data quality indicators (DQI) and other scoring criteria are driven in part by a desire to properly account for sources of uncertainty in life-cycle assessments (Cooper and Kahn, 2012; Edelen and Ingwersen, 2016). Similar desires have been expressed towards water footprint assessments. As described by Hoekstra (2017): "The field has to mature still in terms of calibrating model results against field data, adding uncertainties to estimates and inter-model comparisons as done in the field of climate studies". Additionally, researchers now rely on computational methods to synthesize the large quantity of environmental data and observations that are characteristic of studies conducted at large temporal or regional scales. It is still uncommon for data and computational methods to be published along with the completed studies, which obstructs the reproducibility of many hydrologic studies (Hutton et al., 2016). These later reasons motivated the form of this study—an elementary water footprint analysis decomposed into a reproducible framework.

As a case study in resource sustainability, the State of California presents a unique combination of agricultural and economic activities, resource constraints, and environmental monitoring efforts. Among the United States, California has the greatest population, greatest total farm sales (cash receipts), and if considered separately, would rank as the fifth largest economy in the world, by gross domestic product (Press, 2018). Nine out of California's one hundred million acres contain irrigated agriculture; which requires 30 million acre feet of irrigation in an average year, accounting for 80% of the state's water use (California Department of Water Resources, 2014; Joel Kimmelshue, Mica Heilmann, and Land IQ, 2017). This freshwater requirement is met in part from a vast network of water storage and conveyance infrastructure, which transfer water from the northern third of California, where 2/3 of the precipitation and runoff occurs, to the southern two-thirds, where 3/4 of the anthropogenic water demands are located (Dettinger et al., 2011).

Management of California's freshwater resources are constrained by dynamic availability on one side and strong, persistent demands on the other. Seasonal variations in precipitation affect the availability of freshwater resources in California The state has recently endured a 5-year drought from 2011-2016, marked by a period from 2012 to 2014 that had the worst drought severity in the past millennium [1] (Griffin and Anchukaitis, 2014). On the other side, California's water resources underpin its standing as one of

[1] As defined by cumulative precipitation defect and cumulative Palmer Drought Severity Index (PDSI), see Griffin and Anchukaitis, Geophys. Res. Lett. (2014)

the most productive agricultural exporters in the world and as an important component of the nation's food security. In 2015, California produced more than 99 percent of the United States' almonds, pistachios, walnuts, grapes, peaches, and pomegranates (California Department of Food and Agriculture, 2017b). In the same year, international exports accounted for approximately 26 percent of the state's agricultural production by volume, adding up to 44 percent of the total agricultural sales by value. California is the sole national exporter of many valuable commodities, including almonds, walnuts, and pistachios, which all lie in the top five of the state's agricultural exports by value (California Department of Food and Agriculture, 2017a).

Unpredictable seasonal availability and uncertain international appetite makes it difficult to predict the nature of future constraints and pressures on California's water resources. There is no guarantee that future climatic, economic, or resource environments will accommodate all of the things that societies value: healthy produce, delicious animal foods, verdant natural vistas, thriving native wildlife, and the autonomy that comes from regional food security.

The current attention placed in life-cycle sustainability indicators demonstrates an awareness of the desire to maintain environmental, social, and economic systems without limiting the ability of future generations to meet their needs (World Commission on Environment and Development, 1987; Finkbeiner et al., 2010). When coupled with scenario analysis, these indicators can support strategic decisions to ensure the security of natural resource supplies. Water footprint assessments have been used to quantify the impact of lifestyles on California's water resources (Fulton, Cooley, and Gleick, 2012) and have been proposed as policy support tools (Fulton, Cooley, and Gleick, 2014). Additionally, these assessments have been used to describe the effect of California water resource challenges on international trade networks (Marston and Konar, 2017). While water footprint assessments align with the resource sustainability challenges of California, water scarcity is a problem shared by many nations globally (United Nations World Water Assessment Programme, 2015). Therefore, reproducible sustainability assessments are useful in their ability to be applied and compared between different environmental and economic systems.

1.2 Research Objective

The primary goal of this thesis is to conduct a reproducible water footprint assessment of field and row crop agricultural production in the State of California. This study was motivated in part by popular "gallon-per-almond" depictions of water use in the press, which prompted attention toward the hidden water demands of modern lifestyles. While a useful communication tool, these metrics had limited utility as a decision-support tool, as uncertainty in the underlying measurements was not always explicitly conveyed in the final derived metric. This limitation is shared by many hydrologic studies of water use in California, where magnitudes of uncertainty make it difficult to assess whether mitigation strategies are having their intended effect of improving the state's water security. One way to improve this knowledge defect is to increase the exposure and scrutiny of hydrologic analyses, by ensuring that studies are able to be reproduced and improved upon by as many individuals as possible.

This thesis describes a water footprint assessment implemented using completely Free, Libre and Open Source Software (FLOSS), expressed as a series of documents using the literate programming paradigm (Knuth, 1984). The objectives that guided this research were as follows:

- To examine how the WF of crops varies across the recent record of environmental observations
- To create a tool that can simulate how the WF metric is sensitive to different data sources, including different climate and land use scenarios
- To create a reproducible hydrologic analysis implemented in an accessible form

Chapter 2

Methods

2.1 Overview

This study will begin with an overview of the water footprint method and definitions of terminology and concepts which support the method. The core of this study is a water balance model that calculates crop water requirements on a gridded basis using crop coefficients. Next, an application of this model will be described as implemented to calculate water footprints in a case study of select California agricultural activities.

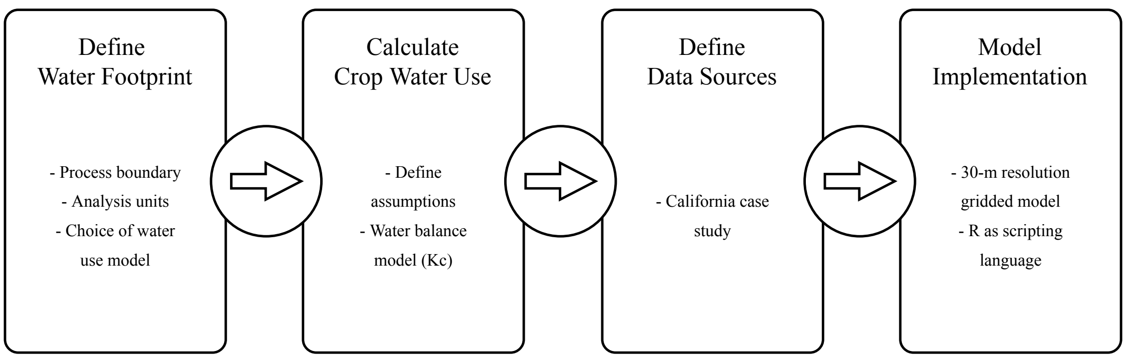

FIGURE 2.1: Overview of the topics covered in this study.

2.2 Terminology

This study examines the direct water footprint of plant cultivation in California in the water years 2008 to 2015. The direct water footprint refers to the water consumed during the irrigation periods within each water year. The Assessment Manual distinguishes between the WF of different processes along the supply chain of a product. For produce, one can imagine water use by the grower, the food processor, the retailer, and the ultimate customer, who may use additional volumes of water in the course of food preparation. ISO 14046:2014 suggests consolidating environmental impacts based on the control of unit processes or equity share. This study treats food cultivation as a "facility" boundary, in the life-cycle of a food product. This study does not consider these end-use impact WFs of food production, nor does it consider the volumes of water "consumed" by assimilating pollutants into freshwater bodies during food production (termed, the grey WF).

"Water consumption" merits additional clarification. Here, "consumption" is defined as the volumes of water transpired by cultivated plants, within a given water year. The Assessment Manual defines "blue water" consumption as the "loss of water from the available ground-surface water body in a catchment area", and "green water" consumption as "rainwater insofar as it does not become run-off" (Equation 2.1). In this study, counties and other hydrologic regions are considered, instead of catchment boundaries. It is assumed that a parcel of water, once transpired, is no longer available for use within a given region of interest.

$$WF_{proc} = WF_{proc,green} + WF_{proc,blue} \cancel{+WF_{proc,grey}} \tag{2.1}$$

In accordance with the Assessment Manual, production water footprints in this study are expressed per unit of product—as in, volumes of water per unit mass of harvested crop, with base dimensions $[L^3M^{-1}]$. Units used in this study include US customary units (gal/lb$_m$) and SI units (m^3/tonne). Through dimensional analysis, the WF can also be represented in terms of another product unit, such as food calories (gal/kcal) or dollars.

2.3 Calculation of the Water Footprint

In accordance with the Assessment Manual, the blue and green components of the water footprint are calculated as the blue and green components of crop water use, divided by crop yield (WF, $[L^3M^{-1}]$) (Equation 2.2).

$$\begin{aligned} WF_{blue} &= \frac{CWU_{blue} \times A/\cancel{A}}{Y/\cancel{A}} = \frac{CWU_{blue} \times A}{Y} \\ WF_{green} &= \frac{CWU_{green} \times A/\cancel{A}}{Y/\cancel{A}} = \frac{CWU_{green} \times A}{Y} \end{aligned} \tag{2.2}$$

Crop water use is defined by the amount of water used by a crop in a given year. The volume of water used by a crop can be considered as a depth of water $[CWU, [L]]$ expressed over a cultivated area $[A, L^2]$. To produce a volumetric quantity, CWU must be multiplied by the area of crop grown (to produce m^3/ha, for example). Crop yield (Y, $[ML^{-2}]$) is defined as the mass of food product produced in a given year. This is expressed as a mass quantity per unit area of crop grown (e.g., tonne/ha).

2.3.1 Calculation of Crop Water Use

This study assumes that crop water use is equivalent to the crop's water requirement multiplied by the process irrigation efficiency. The crop water requirement (CWR, $[L]$) is the amount of water a crop would require to grow under "standard conditions", achieving the maximum rate of evapotranspiration (Equation 2.3). "Standard" growth conditions are described in FAO Irrigation and Drainage Paper No. 56 as "disease-free, well-fertilized crops, grown in large fields, under optimum soil water conditions, and achieving full production under the given climatic conditions" (Allen et al., 1998). The irrigation efficiency (E_i) is defined the ratio of irrigation water beneficially used, to irrigation water actually applied to the crop[1] (Equation 2.4) (Martin and Gilley, 1993).

$$CWU \equiv CWR \times E_i \tag{2.3}$$

$$E_i = \frac{\text{volume of irrigation water beneficially used}}{\text{volume of irrigation water supplied}} \times 100\% \tag{2.4}$$

Simulations run with an irrigation efficiency of 100% (such as this study) can be considered as a lower-bound for real-world crop water use.

2.3.2 Calculation of the Crop Water Requirement

In accordance with the Assessment Manual, the blue and green components of the crop water requirement (CWR) are calculated by accumulating daily crop evapotranspiration over the complete growing period (Equation 2.5). By canceling units, this can be represented a depth of water (CWU, $[L]$) expressed used over the cultivated area.

$$CWR = 10 \times \sum_{t}^{lgp} \left(ET_{c,\ blue} + ET_{c,\ green} \right) \tag{2.5}$$

Daily crop evapotranspiration (ET_c, $[LT_{-1}]$) is defined by the amount of water transpired in a single day, by an unstressed crop grown under standard conditions (refer to 2.3.2 for "standard conditions"). This is represented as a depth of water (e.g. *mm*) transpired by a unit crop, per day. The length of the growing period (lgp, $[T]$) is defined in units of days, from the first day to the last day of applied irrigation.

The irrigated, freshwater component of crop evapotranspiration ($ET_{c,blue}$) is calculated as the difference between daily crop evapotranspiration (ET_c) and daily effective precipitation (P_{eff}). The rainwater rainwater component of crop evapotranspiration ($ET_{c,green}$) is calculated as minimum of the daily effective precipitation (P_{eff}) and daily

[1] Application efficiency (E_a) and application adequacy can be distinguished from overall irrigation efficiency when designing an irrigation system or an irrigation schedule.

crop evapotranspiration (Equation 2.6). If there is more precipitation than crop evapotranspiration, then the irrigated component is equal to zero ($ET_{c,blue} = 0$), and the rainwater component is equal to the daily crop evapotranspiration ($ET_{c,green} = ET_c$).

$$\begin{aligned} ET_{c,\,blue} &= \max\left(0, ET_c - P_{eff}\right) \\ ET_{c,\,green} &= \min\left(ET_c, P_{eff}\right) \end{aligned} \tag{2.6}$$

2.3.3 Calculation of Crop Evapotranspiration

Evapotranspiration (*ET*) describes all of the physical processes by which water is lost to the atmosphere in phase conversions to vapor at or near terrestrial surfaces (Dingman, 2015). It incorporates evaporation, the free conversion of water to vapor from soil and plant surfaces, and transpiration, which is the plant-driven conversion of water to vapor that generally occurs inside of plant stomata. Over a growing season, the amount of water lost to ET is orders of magnitude greater than the amount of water physically embodied in plant tissue. Equation 2.5 assumes that the volume of water physically embodied in plant tissue is negligible compared to the volume of water transpired over the length of a crop's growing season.

This study simulates daily crop evapotranspiration under standard conditions using the "single crop coefficient" method described in FAO Irrigation and Drainage Paper No. 56 (Allen et al., 1998; Doorenbos and Pruitt, 1977). This method is shared by other crop evapotranspiration tools developed for California, including Basic Irrigation Scheduling (BIS) and California Simulation of Evapotranspiration of Applied Water (Cal-SIMETAW) (Orang et al., 2013; Snyder et al., 2007). In this method, ETc is derived by multiplying the evapotranspiration of a reference crop (ET0) by a crop-specific experimentally-derived factor (K_c) (Equation 2.7).

$$ET_c = ET_0 \times K_c \tag{2.7}$$

2.3.4 Definition of Reference Evapotranspiration

Evapotranspiration is constrained by physical factors, including meteorological conditions (temperature, humidity, wind speed, solar radiation), soil characteristics, and the physiological conditions of the given plant. By using a reference crop, grown under "standard conditions" (see 2.3.1), it is possible to define a "reference evapotranspiration" (ET_0), as a function of solely meteorological observations. This "reference evapotranspiration" would reflect an ambient "evaporative demand" that is independent of crop type, management practices, or soil factors (Allen et al., 1998).

In accordance with FAO-56, this study uses a reference ET defined as: "*A hypothetical reference crop with an assumed height of 0.12 m, a fixed surface resistance of 70 s m^{-1}, and albedo of 0.23*". While this is described as a grass reference surface, alfalfa is also used in some regions as a reference crop (with a height of 0.50 meters for a full cover of alfalfa) (Walter et al., 2001). A standard method of calculating ET_0 uses the FAO Penman-Monteith equation; the ASCE method for ET_0 is nearly identical, and differs only in the values used for hourly surface resistance (Allen et al., 2005).

2.3.5 Definition of Crop Coefficients

In order to calculate the specific crop evapotranspiration (ET_c), the reference evapotranspiration is scaled by an empirical factor that incorporates all of the factors that distinguish the specific crop from the reference. These factors include physiological differences which constrain transpiration (crop height, leaf albedo, age, canopy roughness, and stomatal properties) and cultivation differences which constrain soil evaporation and transpiration (irrigation frequencies, soil type, planting density) (Allen et al., 1998; Doorenbos and Pruitt, 1977). K_c values can be calculated as the ratio between ET_o and ET_c at any given point in a crop's growth period (Equation 2.8, from Equation 2.7).

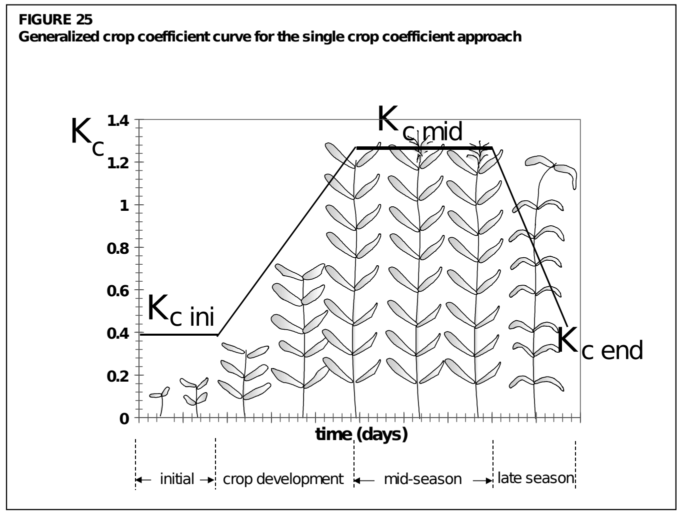

FIGURE 2.2: Crop coefficient curve for a hypothetical annual crop. The crop coefficient curve begins on the first day of irrigation, which is typically the day of planting for most annual plants (Allen et al., 1998).

$$K_c = \frac{ET_c}{ET_0} \tag{2.8}$$

In accordance with FAO-56, this study uses a daily crop coefficient (K_c), computed from a curve, anchored at reference values that describe the initial, mid-season, and late-station crop coefficients. Daily values are calculated by interpolating between these defined values according to the duration of different growth stages (Figure 2.2, also discussed further in section 2.3.3). K_c values are specific to the type of reference crop used, and can be converted between grass and alfalfa references with a scaling factor.

2.4 California Case Study

The methodology described above was used to evaluate the water footprint of orchards and field crops in California, from 2008 through 2015 water years. This case study incorporates a combination of surveyed observations and modeled simulations to both calculate the water footprint and compute daily crop evapotranspiration. Specifically, the entire state was divided into a 30-meter resolution grid and crop water requirements were computed on a daily time step. This was combined with county-level surveys of crop yield to produce county-level water footprint calculations.

2.4.1 Reference Evapotranspiration (Spatial CIMIS)

This study used the California Irrigation Management Information System (CIMIS) to obtain daily reference evapotranspiration (ET_0) observations across the state. Specifically, the Spatial CIMIS data product was used to obtain raster (gridded) representations of daily ET_0 at a 4 km spatial resolution. This data was upscaled to 30 meters, using bilinear interpolation (see section 2.7). The original data is housed and maintained by the California Department of Water Resources (DWR), and can be accessed through the CIMIS web interface [2].

CIMIS comprises a network of over 100 automated weather stations that measure the different meteorological parameters at urban and rural sites throughout California. The system was originally established as a project of DWR and the University of California, Davis in 1982 (*CIMIS*; Snyder and Pruitt, 1992). Each station is sited away from buildings and trees, on a bed of healthy grass that is: "well maintained, properly irrigated and fertilized and mowed or grazed frequently to maintain a height between 10 to 15 centimeters (4 to 6 inches)" (Eching and Moellenberndt, December 1998). Hourly weather observations are transmitted nightly to Sacramento, where the data are used to compute an average daily evapotranspiration of the reference grass surface underneath each station, using a modified version of the 1977 FAO Penman-Monteith ET_0 equation (PM). The CIMIS Equation differs in its use of a wind function and a method of calculating net radiation from mean hourly solar radiation (Dong et al., 1992; Eching and Moellenberndt, December 1998).The ET_0 observations are made publicly available with the primary purpose of aiding agricultural growers develop irrigation schedules.

While the CIMIS network provides station-specific ET_0 calculations, the Spatial CIMIS data product produces a continuous daily ET_0 calculation across the entire state. This is accomplished by using raster observations from the National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environmental Satellite (GOES) system as inputs to the ASCE-Penman-Monteith (ASCE-PM) ET equation (Hart et al., 2009). Spatial CIMIS also interpolates temperature and wind measurements from CIMIS stations, to serve as inputs to the ASCE-PM equation (Equation A.1, also see Appendix A).

Radiative inputs to the ASCE-PM equation are derived from a clear sky factor that is directly related to cloud cover, as observed by GOES satellite data. Specifically, Spatial CIMIS uses GOES visible imagery (visible radiance) to derive a clearness parameter that is directly related to cloud cover in a given grid cell. This is combined with a clear sky solar radiation model developed for the Heliosat-II model (Rigollier, Bauer, and Wald, 2000). Heliosat-II is a software commissioned by the Solar Radiation Data (SoDa) project, with the purpose of converting images acquired by geostationary meteorological satellites into maps of global solar irradiation, received at ground level (Lefèvre, Albuisson, and Wald, 2004). The model incorporates a seasonal turbidity factor, which describes atmospheric attenuation of light due to aerosols and gases. Additional description of inputs to the Spatial CIMIS implementation of the ASCE-PM equation can be found in Appendix A).

Spatial CIMIS has a weakness in estimating solar radiation in scenarios where changes in the surface albedo can be mistaken for cloud cover. This typically occurs in regions that have snowfall and persistent fog, both common winter conditions for some regions

[2] http://www.cimis.water.ca.gov/SpatialData.aspx

in California. Grid cells that contain snowcover and/or fog that persist for greater than 14 days lead to an underestimation of cloud cover and an over-prediction of net radiation during cloudy days Hart et al., 2009. Depending on the location in California, some studies have found good agreement between Spatial CIMIS ET_0 and other methods, while others have used Spatial CIMIS after applying correction factors (Figures 2-4 in Cahn and Farrara, 2012, Figure 6 in Howes, 2017, Figures 3,4 in Orang et al., 2013).

2.4.2 Crop Coefficients (Basic Irrigation Scheduling)

This study used crop coefficients from Basic Irrigation Scheduling (BIS) to scale Spatial CIMIS ET_0 into crop-specific estimations of evapotranspiration ETc. K_c values for 45 unique crops were selected from the BIS software. These values were supplemented with K_c values from the Consumptive Use Program Plus (CUP+) for garlic and oranges and values from the University of California Division of Agriculture and Natural Resources (UCANR) for some orchard crops. K_c values for peppermint and unspecified caneberries were selected from the AgriMet crop coefficients, which were assembled by the United States Bureau of Reclamation (USBR), Pacific Northwest region. K_c values for unstressed Pomegranites were obtained from a study conducted at the Ben-Gurion University of the Negev, Israel.

BIS is an application implemented in Microsoft Excel that is used for the planning of irrigation schedules for crops in California (Snyder et al., 2007). The software was developed as a collaboration between the University of California, Davis, the California Department of Water Resources, and the University of California Cooperative Extension. The program is currently hosted by the UC Davis Biometerology Group and can be accessed at the BIS home page [3].

Among other uses, BIS is used to determine irrigation schedules, irrigation timings, and maximum allowable soil water depletion for 66 unique crop types. It accomplishes this by estimating crop evapotranspiration given mean climate data for a particular region. BIS partitions evapotranspiration into the component of water evaporated from spoil and plant surfaces (E) and the component transpired by leaves (T). As the crop matures, the ratio of T to ET increases, until the transpiration component dominates crop ET. To account for the variable ET_c, BIS defines: K_c values at different stages in a crop's life cycle, typical planting and harvest days, and the proportion of the growing period dedicated to each growth stage. These coefficients are defined according to the FAO-56 "single crop coefficient" method, which assigns values according to 4 growth stages of a typical crop: initial growth, crop development, mid-season, and late-season (Allen et al., 1998; Doorenbos and Pruitt, 1977). These growth stages characterize a crop's daily K_c function, a curve that describes how the values vary as a function of the time in the crop's growing period.

BIS distinguishes between four main crop types. Type-1 crops describe field and row crops that have a period of senescence and defoliation (Figure 2.3). They are characterized by crop coefficients with three inflection points, at 10% ground shading, 75% ground shading, and the onset of senescence. Some type-1 crops such as peas and lettuce, are harvested before their period of senescence. They are characterized by two inflection points, at 10% ground shading and 75% ground shading. Type-2 crops have

[3] `http://biomet.ucdavis.edu/irrigation_scheduling/bis/BIS.htm`

K_c values that are essentially fixed for most of the season. These include alfalfa, pasture, and most types of turfgrass. Shading of soil by dormant grass may cause an overprediction of soil evaporation and total ETc, however the error may be slight due to the lower overall ETo during the cold winter season Richard L. Snyder, 2014. Type-3 crops do not have a water requirement prior to shoot and leaf growth in the spring (e.g. deciduous trees and vines) and can be characterized by a K_c curve with two inflection points. Type-4 crops represent orchard crops that have fixed K_c values throughout their growing season—similar to type-2 crops. Type-4 crops include subtropical orchards (avocado, citrus, and olives) (Snyder et al., 2007).

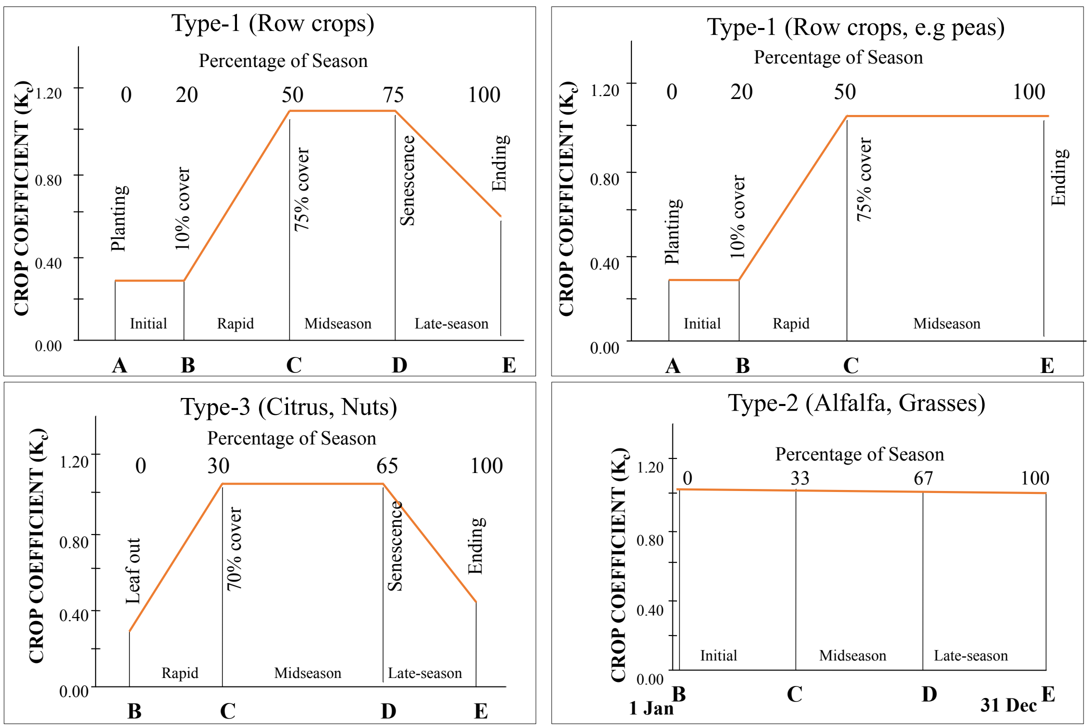

FIGURE 2.3: Crop coefficient curves expressed as a function of time since the onset of irrigation. Modified from (Snyder et al., 2007)

The daily crop coefficient values are derived by the length of each growth period and the value of its endpoint. The beginning and end dates of a given crop's growing period are defined by a "Planting" month and day (signifying the first date of irrigation) and a "Harvest" month and day (signifying the last day of irrigation). The values used for crop coefficients can be found in Appendix B and a description of the function used to generate daily K_c values can be found in Appendix I.

BIS crop coefficients were aggregated across a range of studies and specifications; the provenance is described in a note within the BIS program (reproduced in Appendix B). Many of these values were reproduced in the Consumptive Use Program Plus, an application implemented in Microsoft Excel that also estimates crop evapotranspiration through a method of crop coefficients identical to BIS Orang, Matyac, and Snyder, 2011. CUP+ crop coefficients can be found within the CUP+ application, which can be accessed through the DWR "Land and Water Use" webpage [4]. When K_c values differed between BIS and CUP+, values from CUP+ were chosen (see Appendix B).

K_c values for some orchard crops were obtained from a orchard irrigation reference, published UCANR (Schwankl et al., 2007). K_c values for peppermint and caneberries were obtained from AgriMet, a program from the USBR Pacific Northwest region, which includes an evapotranspiration modeling effort for the Columbia river basin (Bureau of Reclamation, 2016). These values were reported for use with an alfalfa reference, and were converted to the grass ET_0 according to the recommended conversion factors in FAO-56. AgriMet coefficients are maintained by USBR and can be found on the AgriMet "Crop Water Use Information" webpage [5]. Coefficients for pomegranates were obtained from (Bhantana and Lazarovitch, 2010). K_c values were reconciled to fit the growth season partitions used in BIS and CUP+ (see Appendix B).

2.4.3 Land cover (USDA-NASS Cropland Data Layer)

This study assigned K_c values to individual grid cells according to the crop cover, as observed in the Cropland Data Layer (CDL). The United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) has produced land cover raster image products for major agricultural regions since 1970, and for the 48 conterminous states since 2009 (Boryan et al., 2011; Han et al., 2012). Annual CDL images can be viewed through CropScape, a web GIS application maintained by USDA-NASS and the Center for Spatial Information Science and Systems at George Mason University [6]. CDL rasters can be downloaded from the CropScape web service, or at the National Resources Conservation Service Geospatial Data Gateway [7].

The CDL was first created by the USDA NASS Research and Development Division, Geospatial Information Branch, Spatial Analysis Research Section (USDA-NASS, 2018). It was based on an image processing and acreage estimation software named Peditor, written in the 1970s and maintained through 2006 (Boryan et al., 2011). The stated goal of the NASS CDL program is to provide commodity acreage estimates to the Agricultural Statistics Board and other agricultural stakeholders. CDL rasters use standard land cover categories, with an emphasis on agricultural land covers. Records for the State of California begin in the 2007 calendar year; CDL products have a 56-meter spatial resolution from 2007-2009, and a 30-meter spatial resolution from 2009-present.

Currently, the CDL is primary constructed from the supervised classification of remotely sensed satellite imagery, from the Advanced Wide Field Sensor (AWiFS) onboard the Indian Remote Sensing (IRS) satellite, RESOURCESAT-1 (Boryan et al., 2011). This is supplemented with imagery from land imaging sensors[8] onboard the United States Geological Survey Landsat satellites and 16-day Normalized Difference Vegetation Index (NDVI) composites, from the National Aeronautics and Space Administration (NASA)

[4] http://wdl.water.ca.gov/landwateruse/models.cfm

[5] https://www.usbr.gov/pn/agrimet/h2ouse.html

[6] https://nassgeodata.gmu.edu/CropScape/

[7] https://datagateway.nrcs.usda.gov/

[8] Specifically, the Thematic Mapper (TM) on Landsat 4-5, the Enhanced Thematic Mapper (ETM+) on Landsat 7 and the Operational Land Imager (OLI) on Landsat 8.

moderate-resolution imaging spectroradiometer (MODIS). A table of sensor specifications can be found in Appendix 3.

The primary source of ground truth observations for the CDL products is the USDA Farm Service Agency (FSA) Command Land Unit (CLU) program (Boryan et al., 2011). The FSA CLU comprises digitized polygon boundaries of "semi-permanent 'fields'" and is a confidential NASS-internal data set. Auxiliary input data sources include the USGS National Elevation Data set (NED), the Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Dataset (NLCD).

Prior to 2006, classification was performed using a maximum likelihood classifier in the NASS-internal Peditor program, an image processing software written in Pascal and FORTRAN (Boryan et al., 2011). Beginning in 2006, Rulequest Research's See5.0 software was used to create a decision tree classifier. This is applied to the remotely sensed imagery using the MRLC NLCD Mapping Tool and ERDAS Imagine.

Accuracy reports are presented in state-level metadata files each annual CDL survey. For supervised classification, ground-truthed observations are defined as polygons, and are subsequently buffered inward by 30 meters. This was done in part to reconcile differences between the different spatial resolutions of the remotely sensed imagery (see Appendix C). Prior to 2016, this method of inward-buffering was used for validation and the construction of accuracy reports. However, this excluded edge pixels (locations near boarder of different landcover types) from the accuracy reports. This resulted in a somewhat inflated accuracy assessments. Starting in 2017, only "unbuffered" accuracy assessments are reported. In 2016, the CDL metadata included both "buffered" and "unbuffered" accuracy reports (elements reproduced in Appendix C). Overall accuracy for California FSA crops tend to range between 80 and 90 percent.

2.5 Precipitation (PRISM)

In order to determine the proportion of daily crop water requirements that were met by direct rainfall, this study used precipitation observations from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) climate mapping system. Specifically, 800-m daily precipitation rasters were upscaled to 30 meters, using bilinear interpolation (see section 2.4). The PRISM Climate Group at the Northwest Alliance for Computational; Science and Engineering (NACSE) at Oregon State University maintains daily 800-m and 4-km raster datasets of precipitation across the 48 conterminous states, spanning back to 1981. The group also maintains raster datasets of temperature (mean, minimum, and maximum), dewpoint temperature, vapor pressure deficit (minimum and maximum), and 30-year annual "normals" (climatological averages). PRISM rasters are freely available on the PRISM climate group homepage[9], with the exception of 800-m monthly and daily data, which must be ordered.

PRISM precipitation rasters were commissioned by USDA through the Natural Resources Conservation Service (NRCS) to serve as the official spatial climate data sets of the USDA (Daly et al., 2008). PRISM rasters are created at a 30-arcsecond (800-m) spatial resolution and are also available at a 2.5-arcminute resolution (4km), matching previous USDA-NRCS 1961-1990 climate data sets developed in the 1990s.

[9] http://prism.oregonstate.edu/

At its core, PRISM is is an interpolation technique that reproduces the spatial climate patterns of the United States, with a particular emphasis on the effect of elevation and slope on precipitation (Daly et al., 2002). The method was originally developed by Dr. Christopher Daly of Oregon State University in an attempt to reproduce the process that climatologists used to construct climate maps of the United States (Daly et al., 2008; Daly and Bryant, 2013). At its core, the model incorporates data from surface weather stations (13,000 for precipitation and 10,000 for temperature). PRISM utilizes a linear climate-elevation relationship, rather than a multiple regression model due to difficulties in predicting "complex relationships between multiple independent variables and climate". Instead, weather station observations are weighted by distance, elevation, coastal proximity, topographic facet, vertical layer, topographic position, and effective terrain (Daly et al., 2008).

Accuracy estimates using single-deletion jackknife cross validation, leave-one-out cross validation, and a 70% prediction interval have been performed on various revisions to the PRISM method (Daly et al., 2008; Daly, Smith, and Olson, 2015). Regional mean absolute error between predicted and observed precipitation and temperature tend to be similar overall and higher in the physiographically complex western United States. A 2008 evaluation of PRISM for the central California coast saw good agreement between PRISM, WorldClim, and Daymet temperature observations for the central valley of California (Daly et al., 2008).

2.6 Yield

Per crop water footprints were calculated by aggregating daily crop water use over each water year and dividing by the reported yield in each calendar year. Reported yields were obtained from the annual County Agricultural Commissioner's (CAC) reports, aggregated statewide by USDA-NASS (*California County Agricultural Commissioners' Annual Crop Report Manual* 2012). Over 300 unique commodities are recorded in the CAC reports, which includes animal and processed agricultural goods. These categories are reconciled with the crop cover categories observed in the CDL using a pairing similar to the relation described in an earlier water footprint assessment of California (Fulton, Cooley, and Gleick, 2012). The relation used in this study can be found in Appendix D.1. CAC commodities were only included for consideration if they appeared in every year in the study period (2008-2015). The CAC commodities not included in this study can be found in Appendix D.2.

To assess broader trends in agricultural commodities, CDL land cover classes were further assigned a broad commodity label according to the Indicative Crop Classification 1.1 (ICC), a taxonomy of agricultural, developed by the FAO for the 2010 World Programme for the Census of Agriculture (WCA). A description of the ICC-commodity relations used in this study can be found in Appendix 4C.

2.7 Model Framework

The core analytic method employed by this study can be divided into three components: 1. a gridded daily estimation of crop water requirement, 2. a gridded daily estimation of the proportions of crop water requirement that are satisfied by rainwater and irrigation respectively, 3. a regional aggregation to reconcile the modeled crop data with county-level surveyed yields.

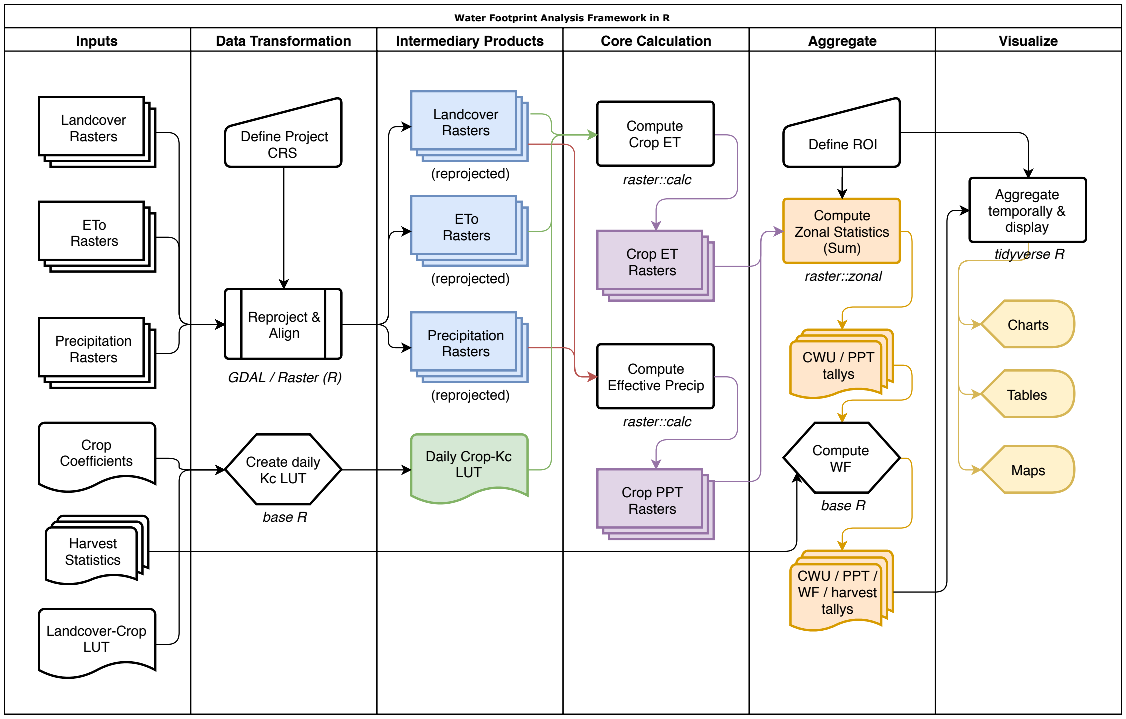

FIGURE 2.4: Architectual overview of the framework

Raster data sets are scaled and reprojected into a common coordinate reference system using R (a statistical analysis software) as a wrapper for utilities found within the Geospatial Data Abstraction Library (GDAL) (GDAL Development Team, 2018; R Core Team, 2017). Crop water requirements were calculated on a cell-wise basis as raster algebra expressions in R using the 'raster' package (Hijmans et al., 2017). Zonal statistics functions in the 'raster' R package were used to aggregate CWU in order to match the spatial and temporal resolution of the yield surveys. Summary statistics were calculated and visualized using a variety of functions in base R and the 'tidyverse' family of packages Wickham, Chang, and RStudio, 2016; Wickham and RStudio, 2017.

The entire analytic workflow is written in the literate programming style, as R markdown notebooks (Allaire et al., 2018; Knuth, 1984). Notebooks are separated according the scripts which comprise the water balance model and the scripts which aggregate and visualize the outputs. For the California case study, a 24-thread workstation was used to run the water balance on a daily interval, across the state (excluding islands), at a 30-meter spatial resolution (1,071,543,084 simulated observations per raster). The notebooks are reproduced in Appendix I.

Chapter 3

Results

3.1 Proximate analysis – Meteorological conditions

The 2008 through 2015 water years extended across 20-year extremes in mean annual precipitation and mean annual temperature (Figure 3.1). Noteworthy years included the span from 2011 to 2016, which contained a 5-year drought marked by warm summertime temperatures. This span also includes the coldest year since 2000 (2011). The wettest year since 2000 also lies within the 2008-2015 study period.

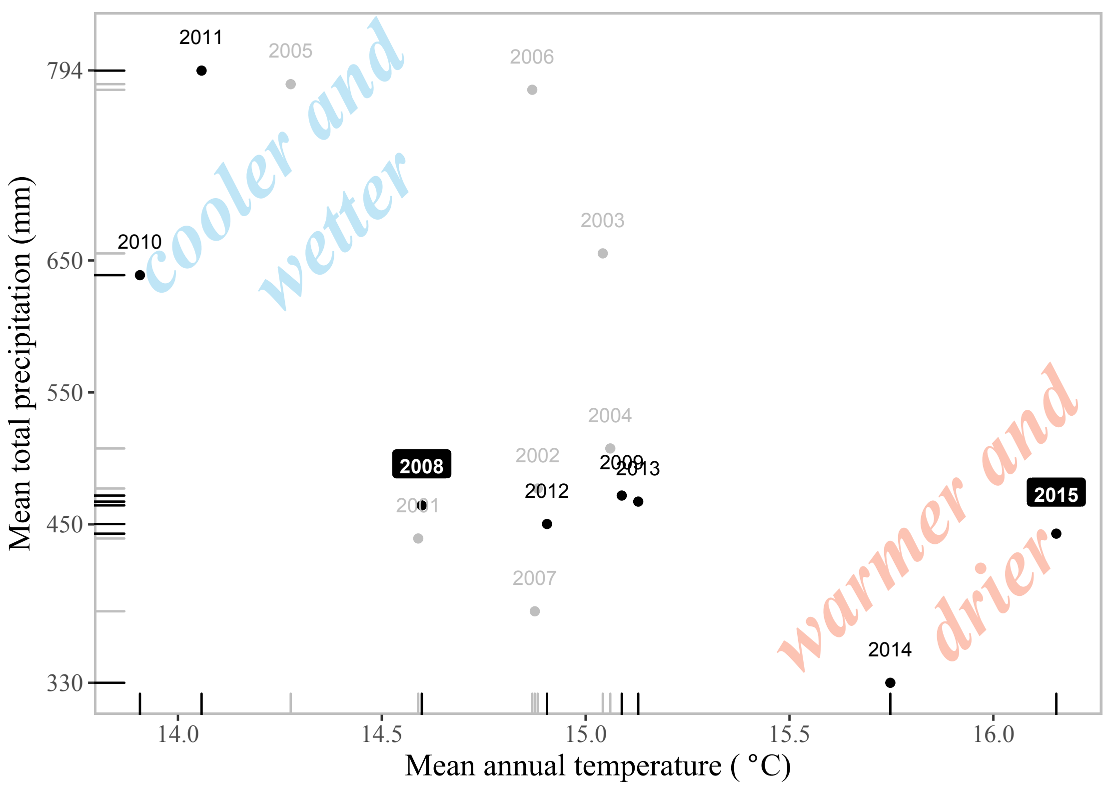

FIGURE 3.1: Mean annual precipitation and mean annual temperature since water year 2000, as depicted in the PRISM daily rasters. The period modeled in this study spans WY 2008-2015.

Average annual statewide temperature was summarized from the 800-m PRISM data sets and compared to the 4-km mean temperature and precipitation products. Aggregations from both datasets produced identical values, demonstrating that the analysis could be replicated (with more coarse temporal resolution) with the precipitation data sets that are freely available from the PRISM web page.

3.2 Land use and distribution

Among the fruits, nuts, and vegetables examined in this study, a total of 54 unique crops were observed in the CDL from 2008 through 2015. Counties with the greatest variety of crops included Fresno (31 unique crops), Riverside (29), and Kern (27), while counties with the least crop diversity included Alpine, Del Norte, Humboldt, Mariposa, Mendocino, Nevada, and Tuolumne (1 unique crop). (Table E.1) Overall, the number of unique crops followed a positive trend, year over year.

The area of harvested acres also followed a positive trend until the 2012 water year, whereupon the number of harvested acres declined from a peak of nearly 7 million acres, to approximately 5 million acres in 2015 (Figure E.1). The majority of harvested acres in any given year are found in the counties within the Sacramento and San Joaquin Valleys (Figure E.3). The top 10 counties by average harvested acres were: Fresno, Tulare, Kern, San Joaquin, and Stanislaus (Table E.2).

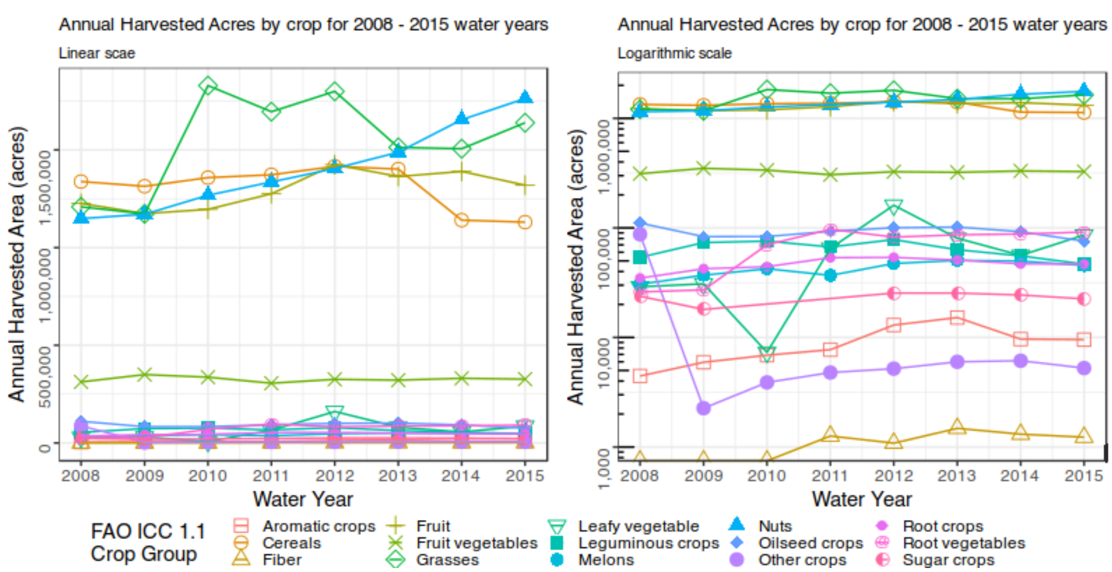

FIGURE 3.2: Total annual harvested acres, aggregated by FAO ICC 1.1 Group. Individual crops are visualized in Figure E.2

Of the modeled crops, the only crop categories that displayed a decrease in irrigated area were cereals, leguminous crops, oilseed crops, and the "other crops" category. While nuts and grasses displayed the largest magnitude of increase irrigated area, the general categories of leafy vegetables and root vegetables experienced the largest proportional increases from 2008 to 20115 (Table E.3).

3.3 Crop water use and evapotranspiration

Consistent with the water footprint assessment manual, crop evapotranspiration was partitioned into a rain-fed ET and irrigation-fed ET (Refer to section 2.3). In order to calculate the blue and green components of crop ET, volumes of precipitated water were tallied over irrigated acres. While there was an observed declining trend of precipitation over irrigated crops, the crop water requirements remained fairly constant, with a positive trend across the state (Figure 3.3, Table F.1).

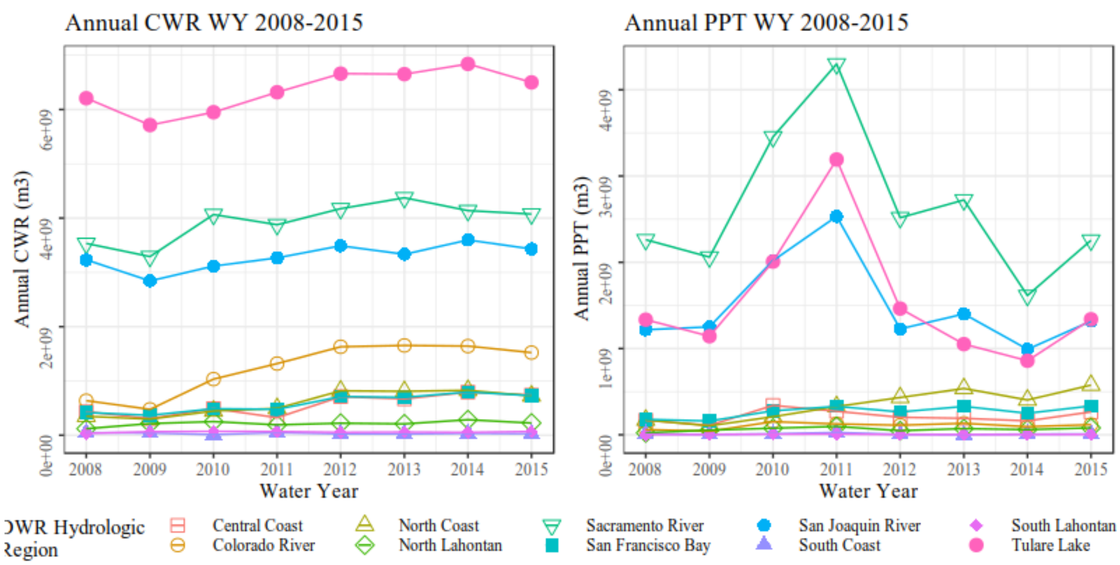

FIGURE 3.3: Annual CWR and precipitation by DWR hydrologic region.

After accounting for the precipitation component of ETc, similar regional patterns persist as with the crop water requirement, with the exception of counties in the Tulare Lake hydrologic region, which had the highest magnitude of rain-fed ET for most years in the study period (Figure 3.4).

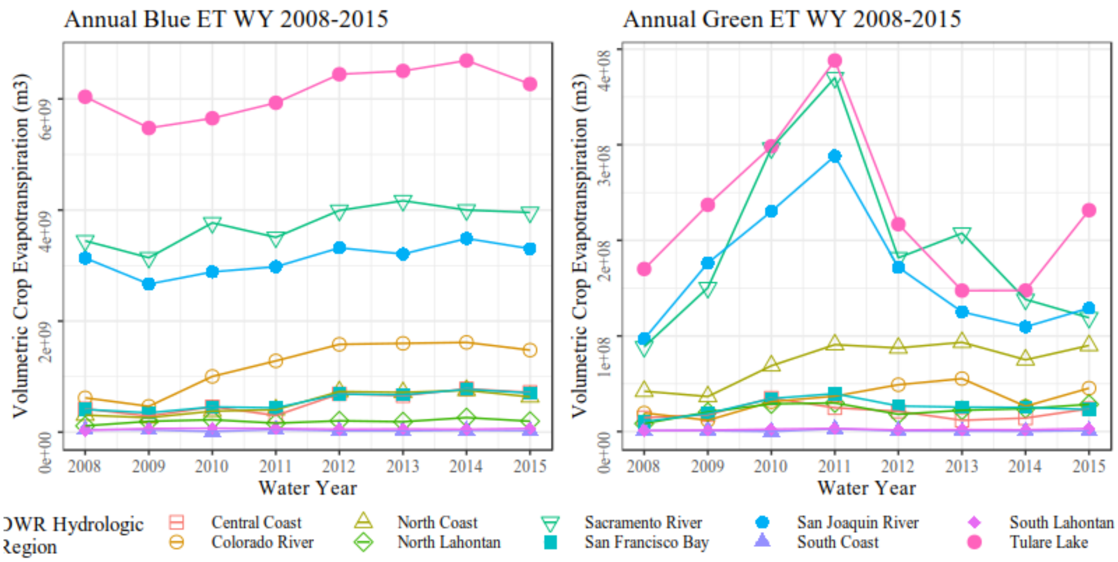

FIGURE 3.4: Annual CWR and precipitation by DWR hydrologic region.

Statewide, the modeled irrigated crops in the 95th percentile of crop water requirement included alfalfa, almonds, walnuts, grapes and miscellaneous pasture grasses (other hay/non-alfalfa). Overall, trends of CWR were consistent for most modeled crops (Figure 3.5, Table F.2). The distribution of precipitation across different crops was similar to the distribution of annual CWR among crops (both are directly related to the number of irrigated acres, for a given crop). Modeled crops that received the most precipitation (95th percentile) included: alfalfa, almonds, rice, and grapes (Figure F.3).

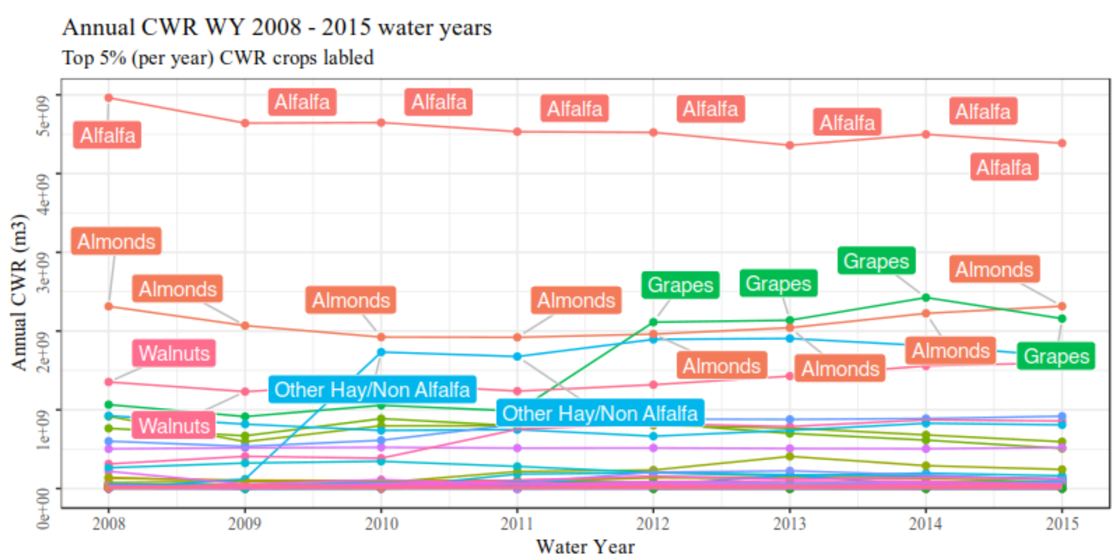

FIGURE 3.5: Annual crop water requirement by CDL crop. Crops in the top 5% of observations foe each year are labeled. This effectively highlights the top 3 crops per year.

Spatially, the majority of crop ET is located in the central hydrologic regions of California. The majority of CWR is located in the southern regions (Tulare Basin) and the majority of precipitation over cultivated acres occurs in the northern regions (Sacramento River). Maps which describe these spatial trends can be found in Appendix F, Figures F.4 and F.5.

3.4 Green and blue water footprints

The water footprint incorporates the effect of yields on crop water use. Assuming negligible losses of water, the crop water requirement assumed to be equivalent to the actual crop water use. The resultant water footprints can be considered a "best-case scenario", as inefficiencies in water distribution and application can only increase the actual crop water use, increasing the blue component of the water footprint. Water footprints are expressed in units of cubic meter of water per metric ton of harvested product. From a resource management perspective, the WF of applied (blue) water is most valuable for regions that are predominantly reliant on surface water resources. Total WF figures are presented in Appendix 5, sections G-H. From 2008 to 2015, the blue WF was always orders of magnitude higher than the green WF (Figure 3.6), further demonstrating the minor role of direct rainfall toward satisfying crop water requirements in California (Figure F.1).

Across large regional extents, the overall water footprint for most hydrologic regions does not vary much year to year, with the exception of isolated fluctuations driven by changes in reported yield (Figure 3.6). These fluctuations are also visible in the crop-specific annual totals (Figure 3.7). For example, low mint yields in Shasta County in 2013 inflate the 2013 water footprint for mint (Figure 3.7). This fluctuation is also visible in the regional WF statistic—due to the low intensity of agriculture in the North Lahontan region (Figure 3.4), the low yield bias on the WF is visible at the regional scale.

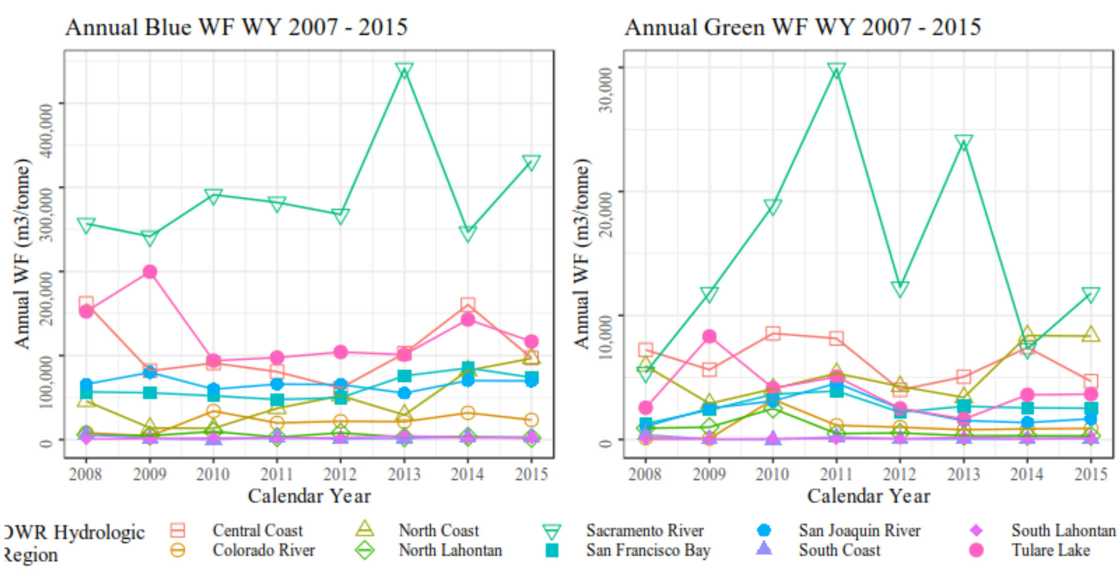

FIGURE 3.6: Annual blue and green water footprint by hydrologic region, expressed as cubic meters of water per metric ton of harvested product.

Most commodity groups experience an increase in blue WF, with fiber crops seeing a nearly 500% increase between 2008 and 2015. Nuts (-20%) and oilseed crops (-40%) experienced a decrease in blue WF over the same period (Table G.3, see also Figure G.1).

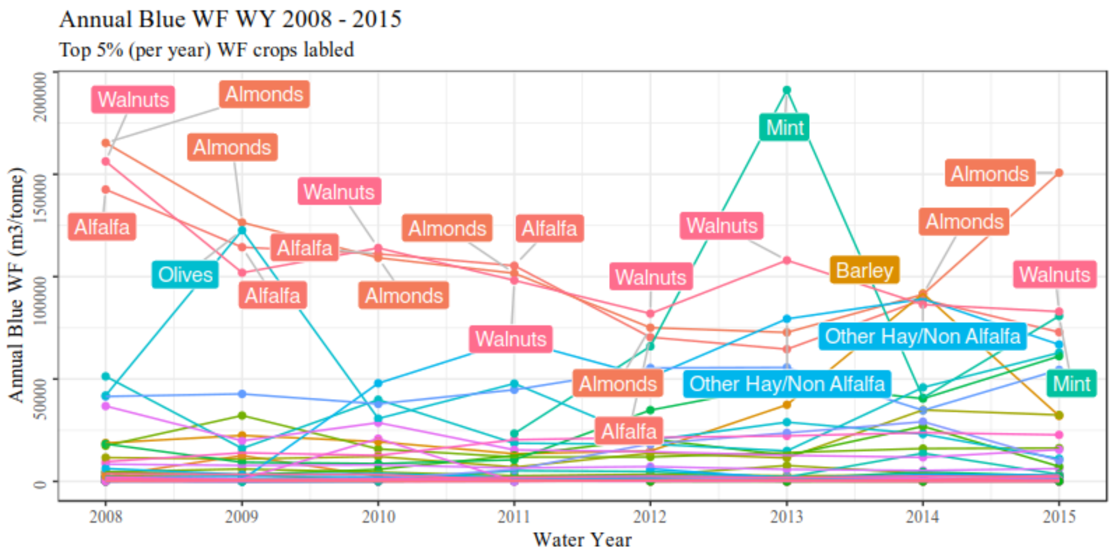

FIGURE 3.7: Annual blue water footprint by CDL crop. Crops in the top 5% of observations foe each year are labeled. This effectively highlights the top 3 crops per year.

Chapter 4

Discussion

4.1 Validation of crop ET model

A 2018 study by the University of California, Davis compared the consumptive use of water by crops in the Sacramento-San Joaquin Delta of California using seven different crop evpotranspiration models. This "Delta ET" study included methods that were based on crop coefficients and methods which are reliant on remotely sensed satellite measurements. Monthly crop ET values were published along with the region of interest, for the 2015 and 2016 water years. Overall crop ET observations from this study were compared to the monthly mean of the seven ET models from the Delta ET study.

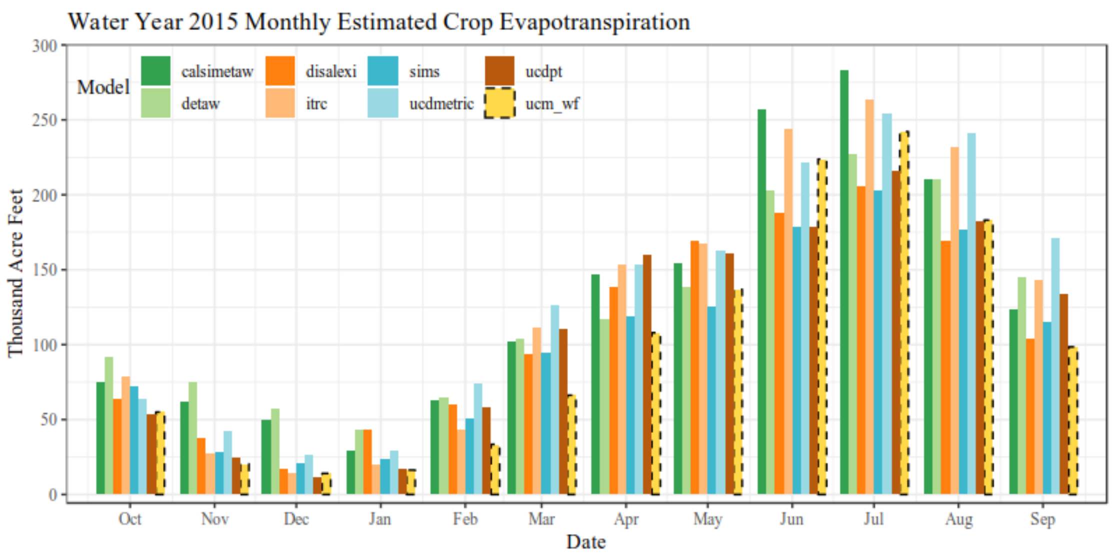

FIGURE 4.1: Monthly total crop ET modeled in the Delta Service Region for the 2015 water year. Calculations from this study labeled "ucm_wf". Data from (Medellín-Azuara et al., 2018).

Overall, there was general agreement between this study and the methods detailed in the Delta ET study. This study tended to underestimate crop ET each month by no greater than 50% (see Table H.1 and figure 4.1). The highest proportion of underestimation occurred during the winter months. However, due to the small magnitude of wintertime ET, this only resulted in a 22% cumulative underestimation (see Figure

H.2). The Delta ET models include some land cover classes that this study does not model. The SIMS ET model implemented by the Delta ET study does not model semi-agricultural/right-of-way and wet herbaceous/sub-irrigated pasture. The results of this study closely match the monthly results from SIMS within 1%.

4.2 Distribution of water utilization

The water footprint can be thought as a measure of the effectiveness of a unit application of water, given yields as the test for effectiveness. Regionally, it is expected that the highest proportion of crop water use would occur in the intensively-cultivated central valley region (encompassing the Sacramento River, San Joaquin River, and Tulare Lake hydrologic regions). Compared to regions less suitable for agriculture (or dominated by urban land use). these regions are exceptional in their overall water use. However, they are not exceptional in the water footprint of agricultural activities (Figure 4.2). For example, Monterey county contained the largest overall average water footprint of agricultural production, in spite of the possessing a small proportion of overall agricultural water requirement.

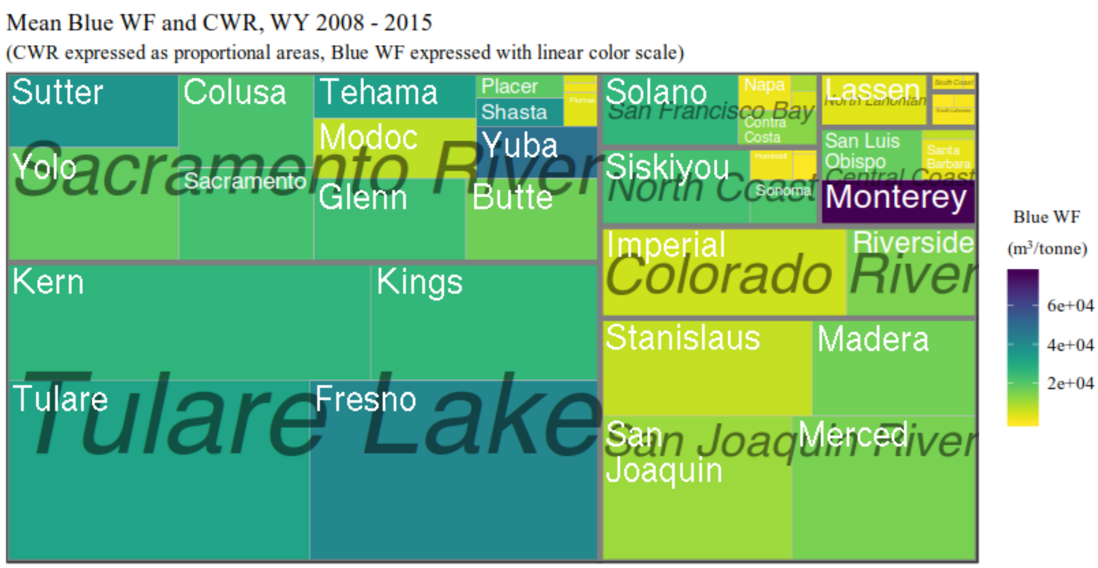

FIGURE 4.2: Treeplot of mean annual blue water footprint expressed on a linear color scale and mean annual crop water requirement expressed as a proportion of overall average CWR. Counties are further grouped by DWR hydrologic region.

Among agricultural commodities, average water footprints agree with other assessments in terms of rank order of water footprint and overall crop water requirement. For example, nuts and grasses both have a large water footprint and large crop water requirement, compared to other crops modeled in this study (Figure 4.3). The large proportional crop water requirement could be function of crop-specific ET characteristics, or it could be an artifact of a large overall cultivated area. However, when compared the proportion of harvested acres, lower WF crops make up a slightly larger portion of cultivated acres than fruits and nuts (Figure G.4).

4.3 Resource availability impacts

Droughts can be used to study the effects of reductions the overall amount of water available in a distribution system. For the drought period starting in 2012, reductions in harvested acres were observed, especially with grasses and some specialty crops. Crop water requirements either remain constant or increase for some crops (as the warm temperatures of the 2011-16 drought drive higher rates of potential evapotranspiration). The equivalence of crop water use and crop water requirement was a central assumption in this study. The response of the water footprint under deficit irrigation can be an important topic for future work, as reductions in water use may be less effective from a footprint perspective if yields are dramatically affected.

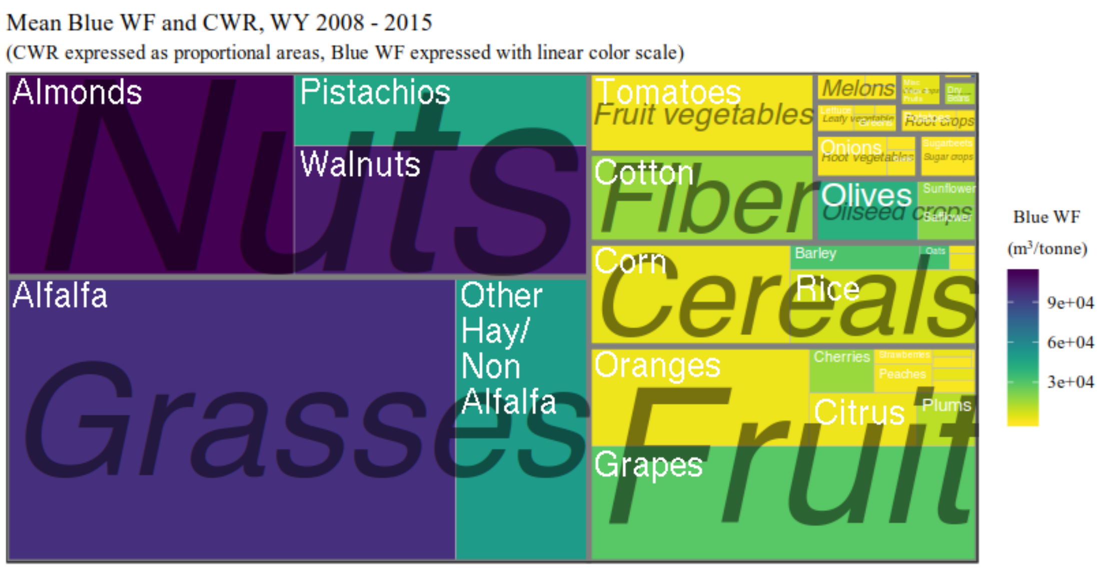

FIGURE 4.3: Treeplot of mean annual blue water footprint expressed on a logarithmic color scale and mean annual crop water requirement expressed as a proportion of overall average CWR. Crops are further grouped by ICC group.

Some water footprint assessments use the proportion of the blue and green water footprint to draw conclusions regarding a region's reliance on a particular type of water resource (Johansson et al., 2016). All regions in California have a significantly larger blue WF than green wf, and these footprints are not necessary to identify the state's reliance on surface water resources. Regionally, the counties with the highest overall CWU tended to contain crops with higher water footprints than those with lower CWU (Figure 4.4).

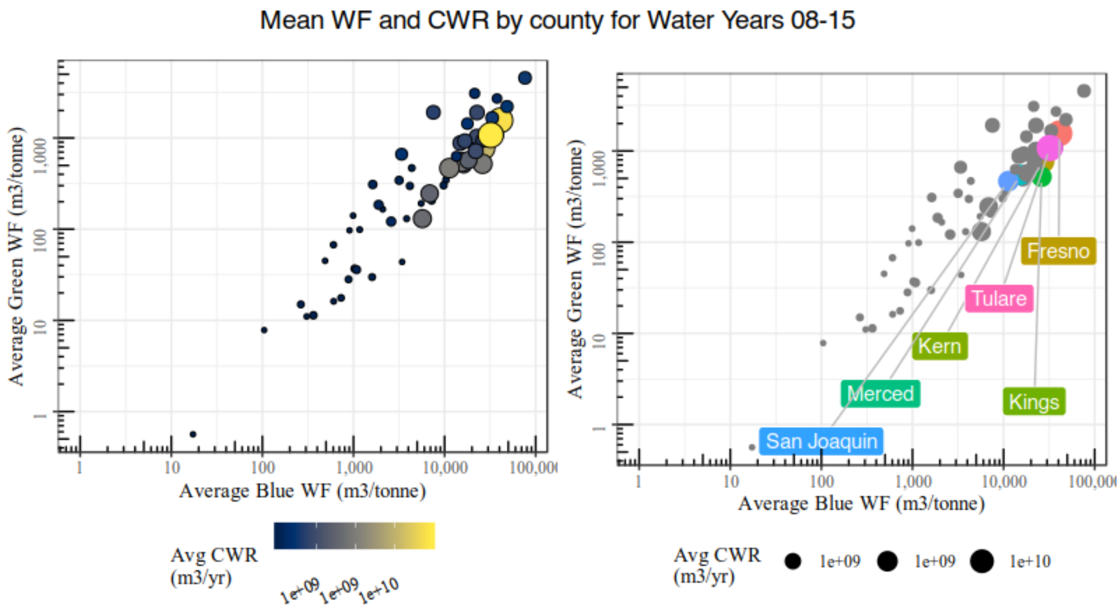

FIGURE 4.4: Average water footprint and crop water requirement. For panel 2, counties in the top 10% of total state CWU are highlighted.

Chapter 5

Summary and Conclusions

This study explored the distribution of water footprints across the State of California regionally, across different commodities, and across a 7 year period, marked by wet and dry climatic extremes. A model of crop water use was coupled with surveyed observations of precipitation, harvest statistics, and a land cover model. Findings from this study revealed an overall insensitivity of the water footprint to climatic extremes and significant inter-annual variability in the metric (by orders of magnitude at times).

As a highly derivative metric, the water footprint accumulates errors from all of the data sources used in its calculation. Unreliable yield reports can dramatically change the water footprint, due to the power-law relationship (multiplicative inverse) between the water footprint and crop yield. By quantifying the uncertainty of this metric, the water footprint could become even more useful as a decision support tool. However, even exploring the relative proportions of water footprints are useful in defining the conceptual extent of the water-use for a given territory or commodity. Future studies can conduct sensitivity analysis of the metric, to examine which input parameters (aside from yield) have the greatest effect on water footprint variability.

In the course of this study, a framework was created and implemented in R that allows this analysis to be replicated and run with different inputs. This framework can be utilized in future analyses to compare the footprint metric with the ever improving agricultural methodologies found in California, from modeling irrigation efficiencies, to using improved land use surveys and methods of modeling crop evapotranspiration. The framework can also be applied to different regions, provided that there are harvest and crop ET models which adequately characterize the region.

An understanding of the water footprint of agricultural production can provide information to the grower who wishes to maximize the economic return of a given volume of water, the state planner who wishes to maximize utility per unit of water allocated, the national administrator who wishes to understand national risks and strengths, or the informed citizen who wishes to align their consumptive activities with a vision for the conditions conferred to the next generation. This information is a critical component of the continuous motivation to characterize relationships between society and natural resource systems, with the ultimate goal of creating sustainable and resilient social and natural systems.

Appendix A

Spatial CIMIS

A.1 ASCE-ET Equation as used by Spatial CIMIS

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{C_n}{T_m + 273} U_2(e_s - e_a)}{\Delta + \gamma(1 + C_d U_2)} \tag{A.1}$$

A.2 CIMIS weather station sensors used for ground station interpolation

CIMIS Sensor	Model	Sensitivity
Total solar radiation (pyranometer)	Li-Cor LI200S	±5%
Air temperature	Fenwal UUT5J1 (HMP35)	±0.1 ° C
Relative humidity	Vaisala Humicap (HMP35)	±2–4% RH
Wind direction (wind vane)	Met-One 024A	±5%
Wind speed (anemometer)	Met-One 014A	±1.5%
Precipitation (tipping bucket)	TI TE525M	±1% at 5 cm/h

TABLE A.1: CIMIS weather station sensors, via (Hart et al., 2009)

A.3 Spatial CIMIS ASCE-PM parameters derivation

CIMIS sensors	ASCE-PM Parameter	Interpolation Method
Dew Point (T_{dewp})	actual vapor pressure (e_a)	average of TG, RST
Min air temp (T_{min})	daily max air temperature (T_x)	average of TG, 2-D RST
Max air temp (T_{min})	daily min air temperature (T_n)	average of TG, 2-D RST
Avg wind speed (U_2)	wind speed (U_2)	3-D RST

TABLE A.2: Spatial CIMIS ASCE-PM station parameters and ASCE-PM derivations, via (Hart et al., 2009)

Note: T_x, T_n, and T_{dewp} represent temperature at 1.5 m above ground level. U_2 represents wind peed at 2 m above ground level.
Note: T_{dewp} is calculated as a function of relative humidity and air temperature and is

returned as a station parameter by the CIMIS system.
Note: Truncated Gaussian (TG) interpolation is calculated according to the method in (Thornton, Running, and White, 1997).
Note: Regularized Spline with Tension (RST) interpolation is computing using 3D splines in the `v.vol.rst` GRASS module (GRASS Development Team, 2017).

GOES sensor	ASCE-PM Parameter	Derivation Method
GOES Visible & Linke turbidity & Solar Almanac	Solar radiation (R_n, R_{ns}, R_{nl}, R_s, R_so)	Heliosat-II
GOES Visible & Linke turbidity & Solar Almanac	Clear sky factor (K)	Similar to Heliosat-II

TABLE A.3: Spatial CIMIS ASCE-PM satellite parameters and derivation, via (Hart et al., 2009)

A.4 Spatial CIMIS ASCE-PM ETo framework

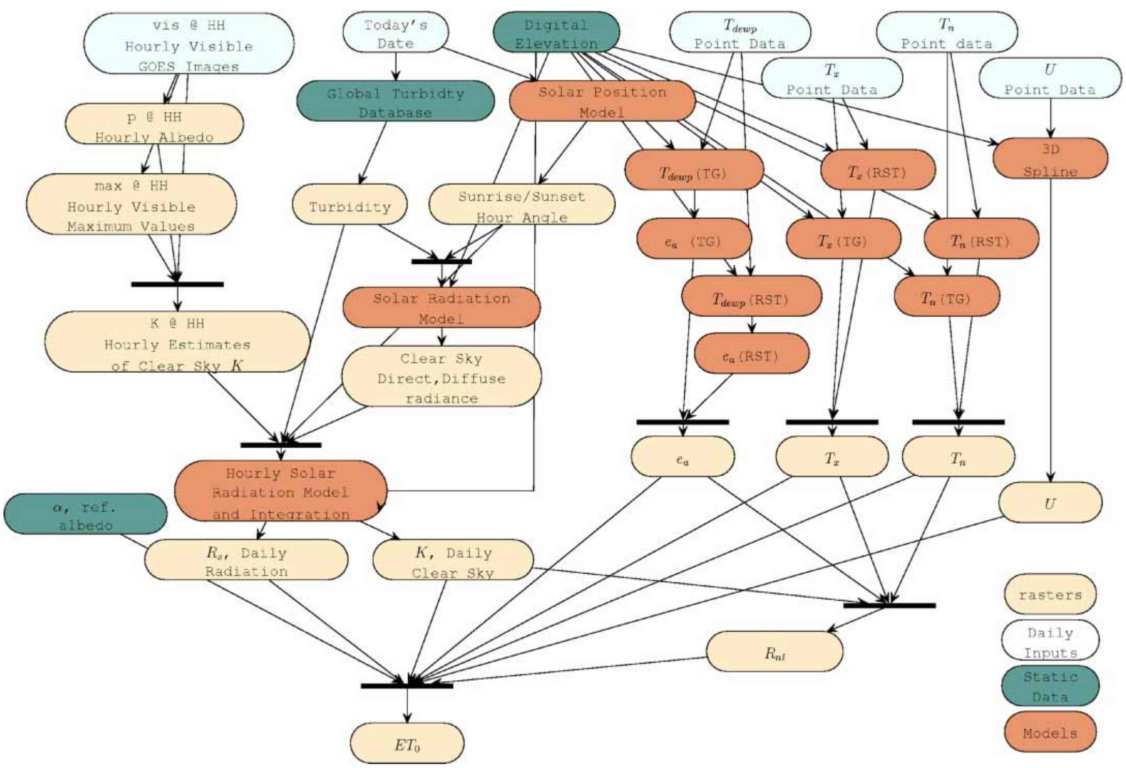

FIGURE A.1: Spatial CIMIS processing flowchart via (Hart et al., 2009)

Appendix B

Crop Coefficients

B.1 BIS/CUP+ Type-1 crops (field and row) modeled in this study

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
1.03	Asparagus	12	25	95	0.25	1	0.25	1	1	12	31
1.04	Barley	20	45	75	0.7	1.1	0.15	11	1	5	31
1.06	Beans (dry)	24	40	91	0.2	1	0.1	6	15	9	30
1.1	Broccoli	20	50	83	0.3	1	0.8	3	15	7	1
1.11	Cabbage	25	63	88	0.3	1	0.85	8	1	11	15
1.12	Carrots	20	50	83	0.85	0.95	0.8	1	15	5	15
1.13	Celery	15	40	90	0.8	0.95	0.95	9	15	1	15
1.15	Corn (silage)	20	45	100	0.2	1	1	5	1	8	15
1.16	Cotton	15	25	85	0.35	0.95	0.5	5	15	10	15
1.17	Cucumber	19	47	85	0.8	0.85	0.85	3	15	6	15
1.18	Eggplant	23	54	85	0.8	0.9	0.85	4	1	11	15
1.19	Flax	17	45	80	0.2	1.1	0.25	4	1	7	31
C.1.2	Grains (small)	20	45	75	0.33	1.1	0.15	11	1	5	31
C.1.21	Grains (winter)	20	45	75	0.33	1.05	0.15	11	1	5	31
C.1.22	Garlic	25	73	92	0.55	1.3	0.2	10	25	6	12
1.23	Lentil	24	40	91	0.2	1	0.1	6	15	9	30
1.24	Lettuce	25	65	90	0.8	0.8	0.8	3	15	7	15
1.25	Melon	21	50	83	0.8	0.95	0.75	4	1	11	15
1.26	Millet	14	36	75	0.3	1	0.3	11	1	5	31
1.27	Mustard	25	63	88	0.3	1	0.85	8	1	11	15
1.28	Oats	20	45	75	0.33	1.1	0.15	11	1	5	31
1.29	Onion (dry)	10	26	75	0.55	1.2	0.55	3	1	10	1
1.31	Peas	20	47	83	0.2	1	1	3	1	5	31
1.32	Peppers	20	45	85	0.8	1	0.85	3	1	8	31
1.33	Rice	24	37	86	1.2	1.05	0.8	5	15	9	30
1.34	Radishes	20	45	85	0.8	0.85	0.75	4	1	5	1
1.35	Potatoes	20	45	78	0.8	1.1	0.7	4	15	8	15
1.36	Safflower	17	45	80	0.2	1.05	0.25	4	1	7	31
1.38	Sorghum	16	42	75	0.2	1.05	0.5	4	1	11	15
1.4	Squash	20	50	80	0.52	0.9	0.7	1	15	4	15
1.41	Strawberries w/mulch	15	45	80	0.2	0.7	0.7	5	1	9	30
1.42	Sugarbeet	15	45	80	0.2	1.15	0.95	3	15	9	30
1.43	Sugarcane	17	44	78	0.4	1.25	0.75	4	1	12	27
1.44	Sunflower	20	45	80	0.2	1.1	0.4	5	1	9	10
1.45	Sweet Potatoes	20	45	78	0.8	1.1	0.7	4	15	8	15
1.46	Tomatoes	25	50	80	0.3	1.1	0.65	4	1	8	31
1.47	Vegetables	33	67	92	0.8	0.9	0.9	3	1	8	31
1.48	Wheat	20	45	75	0.33	1.1	0.15	11	1	5	31
1.49	Watermelon	20	50	75	0.8	1	0.75	4	1	11	15

TABLE B.1: Data from (Snyder et al., 2007). Garlic, grains (small and winter), and crop numbering from Orang, Matyac, and Snyder, 2011.

B.2 BIS/CUP+ Type-2 crops (grass and pasture) modeled in this study

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
2.01	Alfalfa (annual)	25	50	75	1	1	1	1	1	12	31

TABLE B.2: Data from (Snyder et al., 2007). Crop numbering from Orang, Matyac, and Snyder, 2011.

B.3 BIS/CUP+ Type-3 crops (deciduous orchards and vines) modeled in this study

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
A.3.01	Almonds	0	47	80	0.54	0.94	0.7	3	16	11	15
3.02	Apple	0	50	75	0.55	1.15	0.8	4	1	11	15
C.3.04	Table Grapes	0	25	75	0.45	1.05	0.35	4	1	11	1
A.3.07	Stone fruit	0	33	80	0.55	0.87	0.68	3	1	10	31
A.3.08	Walnuts	0	40	73	0.12	1.14	0.28	3	16	11	15
3.09	Peach	0	50	90	0.55	1.2	0.65	4	1	10	15
A3.10	Prunes	0	15	62	0.62	0.96	0.57	4	1	10	31
A.3.12	Pistachios	0	36	64	0.7	1.17	0.35	4	1	11	15

TABLE B.3: Data from (Schwankl et al., 2007). Apples and peaches from (Snyder et al., 2007). Table grapes from (Orang, Matyac, and Snyder, 2011).

B.4 BIS/CUP+ Type-4 crops (subtropical orchards) modeled in this study

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
4.06	Olives	0	33	67	0.8	0.8	0.8	1	1	12	31
4.07	Orange	0	33	67	1	1	1	1	1	12	31
A.4.10	Pears	7	29	79	0.55	0.87	0.65	5	16	12	31

TABLE B.4: Data from (Snyder et al., 2007). Pears from (Schwankl et al., 2007).

B.5 Agrimet crop coefficients

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
PPMTcc	Peppermint	10	57	76	0.19	0.95	0.8	4	30	8	1
TBERcc	Caneberries	10	48	67	0.19	1	0.8	11	1	9	30

TABLE B.5: Data for peppermint from Canyon County (Idaho) Extension Office, 1976. Caneberries from USBR Mid-Pacific Region, 1975 (Bureau of Reclamation, 2016).

B.6 Other crop coefficients

Crop Number	Crop Name	%season B	%season C	%season D	Kc AB	Kc CD	Kc E	Planting Month	Planting Day	Harvest Month	Harvest Day
BL 2010	Pomegranates	14	43	43	0.16	0.64	0.2	4	30	10	15

TABLE B.6: Data from (Bhantana and Lazarovitch, 2010)

B.7 BIS crop coefficient notes

Reproduced from (Snyder et al., 2007).

```
   Kc data marked in blue were derived in work by T.C. Hsiao and former students
at UC Davis.
   The Kc for corn was derived by Steduto and Hsiao (1998) maize canopies uhnder
two soil water regimes II. Seasonal trends of evapotranspiration, carbon dioxide
assimilation and canopy conductance, and as related to leaf area index.  Agric.
and forest Meteorol.  89:185-200.  The Kc =1.05 for cotton is based on work
by Held and Hsiao The Kc = 1.00 for sorghum is based on work by Held and Hsiao.
Millet and For tomato a Kc = 1.10 was selected based on unpublished data from
Snyder and Cahn and on expeiments by Held & Hsiao.  The kc values reported by
Held and Hsiao were slightly higher, but the tomatoes were full canopy (not
in beds, which is the normal practice).  The data from Snyder and Cahn were
typical for California practices.  The data for sunflower were based on data
from Hsiao (personal communication)
   Kc data marked in green were derived from several sources.  The assumption
is that corn has a Kc = 1.00 for ETo calculated using the Pruitt and Doorenbos
(1977) hourly ETo equation that is used by the California Irrigation Management
Information System CIMIS
   Snyder and Pruitt (1992) Evapotranspiration Data Management in California
Irrigation & Drainage Session Proceedings/Water Forum '92, EE,HY,IR,WR Div/ASCE,
Baltimore, MD/August 2-6, 1992.  pp128-133..
   Relative Kp values for alfalfa for the crops marked in green were selected
from Wright (1982) New Evapotranspiration Crop Coefficients.  Presented at Irrigation
and Drainage Specialty Conference, ASCE, July 17-20, Albuquerque, New Mexico.
pp 57-74.
   The peak Kp values were corn = 0.95, alfalfa = 1.0, beans = 1.0, potatoes
= 0.8, sugar beets = 1.0, peas = 0.9, and cereals = 1.0.  Because the equation
for ETo was not available at that time, the Kp values cannot be used directly.
However, assuming the Kc = 1.00 is correct for corn, then the approximate peak
Kc values for the other crops are found by dividing the Kp by 0.95.  The peak
Kc values for a grass ETo are corn = 1.00, alfalfa = 1.05, beans = 1.00, potatoes
= 0.85, sugar beets = 1.05, peas =0.95, and cereals = 1.05.
   The rice Kc = 1.1 more current data on typical ETo from CIMIs and the article
   Lourence and Pruitt (1971) Energy balance and water use of rice grown in
the Central Valley of California.  Agron.  J. 63:827-832.
```

Lourence and Pruitt found ET of rice to be about 4-5% higher than lysimeter measured grass in Davis. Rice ET was measured by Bowen ratio about 25 miles north of Davis. The postulated that the ETo would be less in the rice growing region because of higher humidity. As a result, they recommend a Kc = 1.20 to 1.25. However, CIMIS data indicates that the ETo is only about 5% higher in Davis than at the Nicolas site and in Colusa, which are near the rice growing region. As a result, we would recommend a Kc = 1.05 x 1.05 = 1.10 to estimate ETrice from ETo estimated at a CIMIS station in the rice growing region.

Appendix C

Crop Coefficients

C.1 Satellite imaging sensor specifications

	Landsat TM	IRS AWiFS
Altitude	705 km	817 km
Equatorial crossing time	9:45 +/- 15 min	10:30 +/- 5 min
Temporal resolution	16 days	5 days
Spatial resolution	30×30 m (reflective), 120×120 m (thermal)	56×56 m
Radiometric resolution	8-bit (256)	10-bit (1024)
Spectral resolution	6 (B, G, R, NIR, SWIR, MIR) + Thermal IR	4 (G, R, NIR,SWIR)
Swath width	185 km	740 km
Scene size	184×170 km	370×370 km

TABLE C.1: Data from .(Boryan et al., 2011).

C.2 Selection from 2016 Cropland Data Layer accuracy report

	Correct	Total	Accuracy	Error	Kappa
Overall	1696086	1919814	88.30%	11.70%	0.871
FSA Crops	891603	995949	89.50%	10.50%	0.887
Principal Crops	313515	361429	86.70%	13.30%	0.844
Tilled Crops	523619	606242	86.40%	13.60%	0.848
Forage	159414	178712	89.20%	10.80%	0.699
Vegetables	81947	105093	78.00%	22.00%	0.679
Orchards	208570	239815	87.00%	13.00%	0.817
Berries	110	210	52.40%	47.60%	0.272

TABLE C.2: Data from 2016 "Buffered" CDL validation. Available at CDL web site [1].

	Correct	Total	Accuracy	Error	Kappa
Overall	1512346	1864389	81.10%	18.90%	0.79
FSA Crops	707938	940610	75.30%	24.70%	0.728
Principal Crops	219953	335109	65.60%	34.40%	0.589
Tilled Crops	346070	521238	66.40%	33.60%	0.615
Forage	166276	214196	77.60%	22.40%	0.477
Vegetables	24976	53307	46.90%	53.10%	0.426
Orchards	195592	266572	73.40%	26.60%	0.631
Berries	134	409	32.80%	67.20%	0.058

TABLE C.3: Data from 2016 "Buffered" CDL validation. Available at CDL web site [2].

C.3 Description of CDL inputs and methods

<table>
<tr><td>Year</td><td>1997</td><td>1998</td><td>1999</td><td>2000</td><td>2001</td><td>2002</td><td>2003</td><td>2004</td><td>2005</td></tr>
<tr><td>Field Data</td><td colspan="9">Field level training data collected from June Agricultural Surveys</td></tr>
<tr><td>Satellite Sensor</td><td colspan="9">Landsat 4/5/7</td></tr>
<tr><td>Other inputs</td><td colspan="9"></td></tr>
<tr><td>MODIS</td><td colspan="8"></td><td>AwiFS</td></tr>
<tr><td>AwiFS</td><td colspan="9"></td></tr>
<tr><td>Landsat</td><td colspan="9"></td></tr>
<tr><td>Deimos-1 / UK-DMC 2</td><td colspan="9"></td></tr>
<tr><td>National Land Cover Dataset</td><td colspan="9"></td></tr>
<tr><td>Cultivated Mask Data</td><td colspan="9"></td></tr>
<tr><td>Classification Method</td><td colspan="9">Maximum likelihood classifier</td></tr>
<tr><td>Vector Processing</td><td colspan="9" rowspan="2">USGS PEDITOR</td></tr>
<tr><td>Raster Processing</td></tr>
</table>

<table>
<tr><td>Year</td><td>2006</td><td>2007</td><td>2008</td><td>2009</td><td>2010</td><td>2011</td><td>2012</td><td>2013</td><td>2014</td><td>2015</td></tr>
<tr><td>Field Data</td><td colspan="10">Farm Service Agency Command Land Unit</td></tr>
<tr><td>Satellite Sensor</td><td colspan="7">IRS – P6 Resourcesat-1</td><td colspan="3">Landsat 8</td></tr>
<tr><td>Other inputs</td><td colspan="10"></td></tr>
<tr><td>MODIS</td><td colspan="10">MODIS NDVI product</td></tr>
<tr><td>AwiFS</td><td colspan="8">AwiFS Orthorectified by GeoEye</td><td colspan="2"></td></tr>
<tr><td>Landsat</td><td colspan="3"></td><td colspan="5">Landsat TM / ETM</td><td colspan="2"></td></tr>
<tr><td>Deimos-1 / UK-DMC 2</td><td colspan="5"></td><td colspan="5">DCM</td></tr>
<tr><td>National Land Cover Dataset</td><td colspan="3"></td><td colspan="7">NLCD non-agricultural data</td></tr>
<tr><td>Cultivated Mask Data</td><td colspan="2"></td><td colspan="8">Coverage for entire USA</td></tr>
<tr><td>Classification Method</td><td colspan="10">Decision-tree based classifier</td></tr>
<tr><td>Vector Processing</td><td colspan="10">ESRI ArcGIS</td></tr>
<tr><td>Raster Processing</td><td colspan="10">ERDAS Imagine</td></tr>
</table>

TABLE C.4: Data from Boryan et al., 2011; Han et al., 2012; Kimberly Panozzo, 2016

Appendix D

Agricultural Commodity Relations

D.1 Commodities and Commodity Codes as reported by the County Agricultural Commissioners' Reports

CDL Landcover Class	CAC Commodity
Almonds	Almonds All
Apples	Apples All
Apricots	Apricots All
Asparagus	Asparagus Fresh Market, Asparagus Unspecified
Barley	Barley Feed, Barley Seed, Barley Unspecified
Dry Beans	Beans Blackeye (Peas), Beans Dry Edible Unspecified, Beans Garbanzo, Beans Kidney Red, Beans Lima Baby Dry, Beans Lima Green, Beans Lima Large Dry, Beans Lima Unspecified, Beans Pink, Beans Seed, Peas Dry Edible
Peas	Beans Fava, Beans Fresh Unspecified, Beans Snap Fresh Market, Beans Snap Processing, Beans Snap Unspecified, Peas Edible Pod (Snow), Peas Green Unspecified, Peas Seed
Caneberries	Berries Blackberries, Berries Boysenberries, Berries Bushberries Unspecified, Berries Loganberries, Berries Raspberries
Strawberries	Berries Strawberries Fresh Market, Berries Strawberries Processing, Berries Strawberries Unspecified
Broccoli	Broccoli Food Service, Broccoli Fresh Market, Broccoli Processing, Broccoli Unspecified
Cabbage	Cabbage Chinese & Specialty, Cabbage Head
Carrots	Carrots Food Service, Carrots Fresh Market, Carrots Processing, Carrots Unspecified
Celery	Celery Food Service, Celery Fresh Market, Celery Unspecified
Cherries	Cherries Sweet
Citrus	Citrus By-Products Misc., Citrus Unspecified, Lemons All, Limes All, Tangerines & Mandarins
Greens	Collard Greens, Endive All, Salad Greens Misc.
Corn	Corn Grain, Corn Seed, Corn Silage, Corn Sweet All, Corn White
Pop or Orn Corn	Corn Popcorn
Cotton	Cotton Lint Pima, Cotton Lint Unspecified, Cotton Lint Upland, Cotton Seed Planting, Cottonseed
Cucumbers	Cucumbers
Eggplants	Eggplant All
Small grains (average)	Food Grains Misc., Rice Wild, Rye Grain
Misc Vegs & Fruits	Fruits & Nuts Unspecified, Vegetables Oriental All, Vegetables Unspecified
Garlic	Garlic All

Grapes	Grapefruit All, Grapes Raisin, Grapes Table, Grapes Unspecified, Grapes Wine
Mustard	Greens Turnip & Mustard
Alfalfa	Hay Alfalfa, Seed Alfalfa, Sprouts Alfalfa & Bean
Other Hay/Non Alfalfa	Hay Grain, Hay Green Chop, Hay Other Unspecified, Hay Sudan, Hay Wild, Silage, Straw
Lettuce	Lettuce Bulk Salad Products, Lettuce Head, Lettuce Leaf, Lettuce Romaine, Lettuce Unspecified
Cantaloupes	Melons Cantaloupe
Honeydew Melons	Melons Honeydew, Melons Unspecified
Watermelons	Melons Watermelon
Mint	Mint
Nectarines	Nectarines
Oats	Oats Grain, Oats Seed
Olives	Olives
Onions	Onions
Oranges	Oranges Navel, Oranges Unspecified, Oranges Valencia
Peaches	Peaches Clingstone, Peaches Freestone, Peaches Unspecified
Pears	Pears Asian, Pears Bartlett, Pears Prickly, Pears Unspecified
Peppers	Peppers Bell, Peppers Chili Hot
Pistachios	Pistachios
Plums	Plumcots, Plums, Plums Dried
Pomegranates	Pomegranates
Potatoes	Potatoes All, Potatoes Seed
Sweet Potatoes	Potatoes Sweet
Pumpkins	Pumpkins
Radishes	Radishes
Rice	Rice Milling, Rice Seed
Safflower	Safflower, Safflower Seed Planting
Sorghum	Sorghum Grain, Sorghum Silage
Herbs	Spices And Herbs
Squash	Squash
Sugarbeets	Sugar Beets
Sunflower	Sunflower Seed Planting
Tomatoes	Tomatoes Cherry, Tomatoes Fresh Market, Tomatoes Processing, Tomatoes Unspecified
Triticale	Triticale
Walnuts	Walnuts Black, Walnuts English
Wheat all (average)	Wheat All, Wheat Seed

TABLE D.1: Relations between the USDA-NASS Cropland Data Layer land cover classes used in this study and corresponding commodity classes in the County Agricultural Comissioners' annual report. This paring follows a similar pairing made for an earlier water footprint assessment of California agriculture and trade (Fulton, Cooley, and Gleick, 2012).

Due to the length, this appendix does not contain the CDL and CAC classes that are not used in this study. Both raw descriptors are available at the home pages of the California "County Ag Commissioners' Data Listing"[1] and the "Cropland Data Layer"[2].

[1] USDA-NASS: https://www.nass.usda.gov/Statistics_by_State/California/Publications/AgComm/Detail/

[2] USDA-NASS: https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php

D.2 CAC commodities not present in all years

APIARY PRODUCTS BEES NUCLEI	VEGETABLES BABY
BARLEY SEED	WALNUTS BLACK
BEANS KIDNEY RED	WATERCRESS
BEANS LIMA GREEN	ASPARAGUS FRESH MARKET
BEANS PINK	CHIVES
BEANS SNAP FRESH MARKET	CORN SEED
BERRIES BOYSENBERRIES	FISH SHELL
BIOMASS FOR ENERGY	FLOWERS CHRYSANTHEMUM CUT POM.
CATTLE CALVES EXCLUDED UNSPECIFIED	FLOWERS LILACS CUT
CHERIMOYAS	FOOD GRAINS MISC.
CHESTNUTS	HIDES SKIN & FUR
CHICKENS CHICKS BROILER	MELONS CRENSHAW
CHICKENS HENS SPENT	MOHAIR
CORN CRAZY	NURSERY HORTICULTRAL SPECIMIN MISC.
CORN POPCORN	PEAS SEED
CORN WHITE	QUAIL
EGGS CHICKEN HATCHING	RHUBARB
FLOWERS ANTHURIUMS CUT	SEED VETCH
FLOWERS GARDENIAS CUT	SHEEP BREEDING STOCK
FLOWERS ORCHIDS CUT ALL	SOYBEANS
JOJOBA	BERRIES LOGANBERRIES
MACADAMIA NUTS	CHAYOTES
NURSERY FLOWER BULBS/CORMS/RHIZOMES	EGGS DUCK ALL
NURSERY FLOWER PROPAGATIVE MATERIALS	TARO ROOT
NURSERY FRUIT/VINE/NUT NON-BEARING	BEANS SNAP PROCESSING
NURSERY GERANIUMS	FLOWERS ROSES CUT STANDARD
OATS SEED	PEARS PRICKLY
PEANUTS ALL	CATTLE CALVES EXCLUDED UNSPECIFIE
PEAS DRY EDIBLE	GOAT CHEESE
PHEASANTS	GUAR
POTATOES SEED	NURSERY FLOWER BULBS/CORMS/RHIZOM
SEED CLOVER UNSPECIFIED	NURSERY FLOWER PROPAGATIVE MATERI
TOMATILLO	NURSERY FRUIT/VINE/NUT NON-BEARIN
TOMATOES CHERRY	NURSERY HORTICULTRAL SPECIMIN MIS
TOMATOES GREENHOUSE	SPINACH PROCESSING
TURKEYS TOMS & HENS	

TABLE D.2: County Agricultural Comissioners' commodity categories that were not present in all years. These commodities were excluded from yield consideration for the study years (2008-2015). This list was constructed by itemizing all unique commodities across the State of California, for each calendar year. The intersection of each annual list was computed (representing the set containing elements common to all annual lists). The remaining elements are displayed in the above table. The derivation is documented in Appendix I, Notebook 9 (yield prep), Section 4.

D.3 FAO Indicative Crop Classification

In order to assess broader trends in agricultural commodities, this study utilizes a taxonomy of agricultural commodities, named the Indicative Crop Classification (ICC 1.1) The ICC was developed by the Food Agriculture Organization of the United Nations (FAO) for the 2010 World Programme for the Census of Agriculture (WCA) (Food and Agriculture Organization of the United Nations, 2017). The WCA provides assistance and a standardized framework for countries to conduct national agricultural censuses. ICC 1.1 is specified in Annex 4, "Classification of Crops" in the *World Programme for the Census of Agriculture 2020*. Originally developed for WCA 2010, the 1.1 incorporates commodity classifications from the UN Central Product Classification 2.1, which includes categories based on "the nature of the product and industry of origin". A translation between crop cover classes in the Cropland Data Layer is presented below.

ICC 1.1 Group	CDL Crop
Cereals	Corn, Rice, Sorghum, Sweet Corn, Pop or Orn Corn, Barley, Durum Wheat, Spring Wheat, Winter Wheat, Other Small Grains, Oats, Millet, Triticale
Fiber crops	Cotton, Flaxseed
Oilseed crops	Sunflower, Safflower, Mustard, Olives
Fiber	Mint
Grasses	Alfalfa, Other Hay/Non Alfalfa
Sugar crops	Sugarbeets, Sugarcane
Leguminous crops	Dry Beans, Lentils, Peas
Root crops	Potatoes, Sweet Potatoes
Other crops	Other Crops, Misc Vegs & Fruits, Herbs
Vegetables and melons	Watermelons, Onions, Cucumbers, Tomatoes, Carrots, Asparagus, Garlic, Cantaloupes, Honeydew Melons, Broccoli, Greens, Lettuce, Cabbage, Celery, Radishes, Eggplants
Fruit and nuts	Caneberries, Cherries, Peaches, Apples, Grapes, Citrus, Almonds, Walnuts, Pears, Pistachios, Prunes, Oranges, Pomegranates, Nectarines, Plums, Strawberries, Squash, Apricots, Pumpkins
Stimulant, spice and aromatic crops	Peppers

TABLE D.3: Data from manual reconciliation between WCA2020 and CDL land cover categories..

Appendix E

Extended results, plots, and tabulations: Harvest statistics

E.1 Unique crop cover observations

County	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alameda	1	2	3	3	3	3	3	3	**3**
Alpine	NA	NA	1	1	1	1	1	1	**1**
Amador	3	3	3	3	3	3	3	3	**3**
Butte	10	10	10	10	10	10	10	10	**11**
Calaveras	1	2	4	4	4	4	4	3	**4**
Colusa	10	10	10	11	12	12	12	12	**13**
Contra Costa	7	7	9	10	11	10	10	9	**13**
Del Norte	NA	NA	1	1	1	NA	NA	1	**1**
El Dorado	NA	NA	1	5	4	3	3	5	**7**
Fresno	25	26	27	29	31	29	31	27	**34**
Glenn	12	11	13	12	14	14	15	15	**16**
Humboldt	NA	NA	1	1	1	1	1	1	**1**
Imperial	7	7	6	9	17	17	17	17	**17**
Inyo	NA	1	1	1	2	2	1	1	**2**
Kern	17	15	18	20	22	20	21	20	**27**
Kings	15	15	17	16	16	16	15	14	**20**
Lake	2	1	1	2	2	2	2	2	**2**
Lassen	4	5	5	2	2	2	2	2	**5**
Los Angeles	2	2	1	4	3	4	5	4	**5**
Madera	14	14	14	14	14	14	14	14	**16**
Marin	1	1	2	2	2	2	2	2	**2**
Mariposa	1	1	NA	NA	1	1	1	1	**1**
Mendocino	1	NA	NA	1	1	1	1	1	**1**
Merced	16	17	17	16	18	19	17	16	**19**
Modoc	3	4	5	6	6	6	NA	NA	**6**
Mono	2	2	2	2	2	2	1	1	**2**
Monterey	13	11	11	14	16	14	15	13	**20**
Napa	2	2	3	3	2	2	2	3	**3**
Nevada	1	1	1	1	1	1	1	1	**1**
Orange	1	2	1	1	NA	NA	1	1	**2**
Placer	3	4	5	6	7	7	7	6	**7**
Plumas	1	1	2	2	2	2	2	2	**2**
Riverside	7	6	9	13	23	23	23	25	**29**
Sacramento	9	9	12	12	12	13	12	13	**17**
San Benito	8	7	7	9	9	8	8	8	**11**
San Bernardino	3	3	4	4	4	5	5	6	**7**
San Diego	2	2	2	4	3	1	1	2	**7**
San Joaquin	15	17	20	20	22	23	23	21	**24**
San Luis Obispo	10	6	7	11	11	11	10	10	**13**
San Mateo	1	1	2	2	2	1	2	3	**3**
Santa Barbara	6	3	4	9	10	9	5	3	**11**

Santa Clara	9	11	10	11	11	12	13	10	**16**
Santa Cruz	2	2	2	3	4	2	2	2	**4**
Shasta	2	3	4	5	5	5	5	5	**5**
Sierra	1	1	2	2	2	2	2	2	**2**
Siskiyou	4	4	5	7	7	7	7	7	**7**
Solano	12	13	16	12	14	13	13	13	**18**
Sonoma	3	3	4	3	3	3	4	4	**4**
Stanislaus	16	15	17	18	19	19	18	13	**20**
Sutter	11	11	13	12	13	14	14	12	**15**
Tehama	7	7	8	8	9	9	9	9	**9**
Trinity	NA	NA	1	1	2	2	NA	NA	**2**
Tulare	18	17	18	18	20	20	21	20	**22**
Tuolumne	NA	NA	1	1	1	1	1	1	**1**
Ventura	4	2	2	7	6	8	8	7	**12**
Yolo	8	8	9	9	11	11	11	11	**11**
Yuba	5	6	7	12	7	7	7	7	**13**
Total	**42**	**41**	**41**	**44**	**52**	**52**	**52**	**51**	**54**

TABLE E.1: Counts of unique harvested acres, by county. The "All Years" column reflects all of the unique crops observed in the county across all years (not a sum). The "Total" row reflects all of the unique crops observed across all counties (statewide, not a sum).

E.2 Tabulations of annual total harvested acres

County	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alameda	495	2669	8520	8124	8969	7439	6324	5265	**47805**
Alpine	NA	NA	122	122	150	200	250	250	**1094**
Amador	4370	4158	4089	4057	4656	4485	4590	5292	**35697**
Butte	184967	187306	178926	193601	193850	201037	181045	195052	**1515784**
Calaveras	800	1600	2040	2240	2125	2115	2018	1890	**14828**
Colusa	242810	249640	252330	251634	263465	273166	231506	221384	**1985935**
Contra Costa	19168	15661	19772	22816	22904	22115	19419	22270	**164125**
Del Norte	NA	NA	2500	2430	2600	NA	NA	3430	**10960**
El Dorado	NA	NA	216	3282	2458	2456	2446	2728	**13586**
Fresno	793714	802837	821311	842423	926798	810135	829450	838490	**6665158**
Glenn	165310	183890	191691	189255	209960	215460	202056	200782	**1558404**
Humboldt	NA	NA	3523	10898	10600	10600	10600	10600	**56821**
Imperial	207654	186966	177470	168820	249197	232607	244150	297291	**1764155**
Inyo	NA	3200	3200	3280	4985	4825	2420	0	**21910**
Kern	552386	466046	568866	629747	664117	653081	710300	637097	**4881640**
Kings	271455	219582	325388	284861	300483	279992	262936	258067	**2202764**
Lake	10654	2600	3100	11400	11450	11550	11770	11810	**74334**
Lassen	64974	66708	65956	27000	28903	28903	28903	28000	**339347**
Los Angeles	6098	7145	53	10396	10702	9153	9862	9797	**63206**
Madera	247400	232810	263640	274520	280020	282850	275550	286050	**2142840**
Marin	195	193	4526	4156	3766	2985	3325	3445	**22591**
Mariposa	100	90	NA	NA	104	110	77	73	**554**
Mendocino	16400	NA	NA	16700	16800	16800	16900	16900	**100500**
Merced	364514	386420	373260	353240	361521	392613	387214	397120	**3015902**
Modoc	36725	40947	70950	72006	72006	72006	NA	NA	**364640**
Mono	16000	16000	16000	16000	15480	15180	8100	0	**102760**
Monterey	78459	74346	67197	64421	121974	62325	61968	57875	**588565**
Napa	43114	43238	43867	44065	43824	44370	44250	44170	**350898**
Nevada	402	248	236	343	312	424	342	352	**2659**
Orange	77	1187	761	856	NA	NA	658	545	**4084**
Placer	14269	16788	19440	16001	17817	18872	16681	12711	**132579**
Plumas	5500	6105	10000	10000	10000	10000	10000	8410	**70015**
Riverside	67158	53033	69990	75616	118133	130868	129724	123244	**767766**

Sacramento	78080	86120	98208	102478	107793	108657	108906	108866	**799108**
San Benito	11139	12639	26488	34258	32545	22785	20931	21179	**181964**
San Bernardino	10770	11045	13522	17996	17026	11216	14744	13512	**109831**
San Diego	7253	6700	6707	12716	7032	4930	3850	6815	**56003**
San Joaquin	460975	446557	569117	557527	610927	592610	609174	585641	**4432528**
San Luis Obispo	50149	49221	63759	71798	72308	60171	62487	71235	**501128**
San Mateo	400	330	693	842	815	423	554	718	**4775**
Santa Barbara	36672	12102	18647	53670	53693	53558	32203	29672	**290217**
Santa Clara	5777	9989	10446	11745	12214	11319	12043	10391	**83924**
Santa Cruz	3899	3802	3927	7424	10936	3976	3954	3776	**41694**
Shasta	3940	4120	24310	25130	24360	24830	24850	25950	**157490**
Sierra	2000	160	3550	3550	3550	3550	3440	1920	**21720**
Siskiyou	64846	55128	57629	47578	68960	57958	67767	63160	**483026**
Solano	76959	80369	98301	91945	105974	112797	118849	107193	**792387**
Sonoma	56037	56998	72569	74040	71083	72555	71436	69585	**544303**
Stanislaus	312403	364254	416074	404743	408018	389387	378462	352622	**3025963**
Sutter	172240	180825	193966	191191	214061	221544	189271	188254	**1551352**
Tehama	36604	40203	47065	45752	60310	59440	60321	60471	**410166**
	NA	NA	550	550	664	664	NA	NA	**2428**
Tulare	576412	570857	767545	764970	785709	625683	653328	926969	**5671473**
Tuolumne	NA	NA	300	300	NA	NA	NA	NA	**1980**
Ventura	26288	13608	300	360	360	360	33380	15554	**181491**
Yolo	178347	173891	189830	202280	207440	207170	201400	175710	**1536068**
Yuba	48218	51590	56841	55837	63547	66415	66345	69236	**478029**
Total	**5634576**	**5501921**	**6322744**	**6412672**	**6938988**	**6569985**	**6452889**	**6609179**	**50442954**

TABLE E.2: Sums of harvested acres, by county. This table was constructed by tallying harvested acres as they appeared in the annual NASS Ag Commissioners' Reports, *provided that there were* ***one or more*** *observations of the crop in the CDL raster in a given county, in a given year*. Crops that have an entry in the Ag Comissioners' report, but are not observed in the CDL *are not* counted.

CDL Name	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alfalfa	1165225	1129997	1040060	950020	1041451	949816	890016	866566	**8033151**
Almonds	756861	772152	840160	875861	904130	940070	1038417	1108242	**7235893**
Apples	3956	5179	5444	6256	5091	4712	2990	2599	**36227**
Apricots	6333	9525	10242	9018	8838	8780	9079	NA	**61815**
Asparagus	15024	11282	7310	7450	9380	8280	8890	2820	**70436**
Barley	41967	44894	53402	53045	67685	33869	22139	27802	**344803**
Broccoli	NA	NA	NA	NA	83060	23480	21470	22263	**150273**
Cabbage	NA	NA	NA	NA	3025	1707	2390	1220	**8342**
Caneberries	NA	NA	NA	NA	4183	NA	1460	NA	**5643**
Cantaloupes	24735	31821	34770	32162	27145	27587	25500	22570	**226290**
Carrots	15504	NA	1862	17322	18430	20369	23015	24252	**120754**
Cherries	32837	33697	39909	38457	39931	41384	39668	39597	**305480**
Citrus	45472	23197	29858	33290	34200	54080	66922	58791	**345810**
Corn	728252	667073	701585	726573	750525	764181	629081	623698	**5590968**
Cucumbers	NA	NA	NA	NA	NA	2140	NA	NA	**2140**
Dry Beans	51604	72811	75022	61209	73126	58086	51966	45010	**488834**
Eggplants	NA	NA	NA	NA	349	406	260	220	**1235**
Garlic	NA	17200	21700	27590	18820	22060	22410	23148	**152928**
Grapes	788579	749033	751832	846602	951599	878907	897978	868988	**6733518**
Greens	13976	19563	NA	NA	9240	5980	3230	2830	**54819**
Herbs	1277	NA	NA	105	863	1155	1249	1597	**6246**
Honeydew Melons	NA	NA	NA	NA	11363	13067	13653	11979	**50062**
Lettuce	NA	0	0	57628	56158	40934	20119	57020	**231859**
Mint	NA	NA	NA	1485	1085	1480	1310	1230	**6590**
Misc Vegs & Fruits	86137	2261	3888	4678	4326	4839	4882	3657	**114668**
Nectarines	33380	31603	28470	25320	25385	25184	22075	21120	**212537**
Oats	4692	8248	5905	4371	3193	3283	7285	7381	**44358**
Olives	30210	34174	37403	38602	40840	42336	42846	38005	**304416**
Onions	10505	9968	46591	52042	45156	44009	42538	43889	**294698**
Oranges	187310	178512	179943	174411	161417	164020	169924	160514	**1376051**
Other Hay/ Non Alfalfa	43225	42925	789403	745708	758451	563191	616352	772646	**4331901**
Peaches	68165	67517	68587	58222	62742	60589	56301	53133	**495256**
Pears	NA	NA	NA	NA	7203	6276	6689	6780	**26948**
Peas	2306	622	717	5791	5025	5128	3543	1641	**24773**
Peppers	4449	5941	6890	7735	12978	15191	9653	9552	**72389**
Pistachios	142447	137320	145288	170378	188852	210038	256779	291167	**1542269**
Plums	42667	40439	40670	46939	86495	82857	79535	71890	**491492**
Pomegranates	6870	9543	14570	11465	12820	9552	10350	9460	**84630**
Potatoes	20138	24799	26310	35322	37562	32452	28070	27380	**232033**
Pumpkins	NA	NA	NA	292	3419	4167	3633	2801	**14312**
Radishes	NA	NA	NA	NA	NA	NA	NA	249	**249**
Rice	539462	582225	587929	575860	578811	578870	458293	445117	**4346567**
Safflower	61307	22485	25705	28313	29254	22576	18297	19798	**227735**
Sorghum	20908	9310	7055	11960	11832	12600	13100	16084	**102849**
Squash	1292	NA	2444	4551	2122	2096	1288	120	**13913**
Strawberries	10534	25739	26517	25833	25005	27667	26959	26596	**194850**
Sugarbeets	23773	18022	NA	NA	25400	25400	24400	22500	**139495**
Sunflower	19208	26605	20350	26158	30200	36865	31210	17690	**208286**
Sweet Potatoes	14660	17586	17930	18180	16270	18440	18870	19250	**141186**
Tomatoes	311095	349887	334273	300925	319699	312646	325752	323482	**2577759**
Triticale	2833	2205	2370	1120	3820	9150	10600	10270	**42368**
Walnuts	249462	261351	282790	289733	312247	338223	359793	363705	**2457304**
Watermelons	5939	5210	7590	4690	8787	9810	10660	10860	**63546**
Total	**5634576**	**5501921**	**6322744**	**6412672**	**6938988**	**6569985**	**6452889**	**6609179**	50442954

TABLE E.3: Sums of harvested acres, by CDL crop. This table was constructed by tallying harvested acres as they appeared in the annual NASS Ag Commissioners' Reports, *provided that there were* ***one or more*** *observations of the crop in the CDL raster in a given county, in a given year*. Crops that have an entry in the Ag Comissioners' report, but are not observed in the CDL *are not* counted.

FAO ICC 1.1 Group	% Difference	Difference
Aromatic crops	115	5103
Cereals	-16	-207762
Fiber	NA	1230
Fruit	8	93365
Fruit vegetables	5	14236
Grasses	36	430762
Leafy vegetable	197	57153
Leguminous crops	-13	-7259
Melons	48	14735
Nuts	53	614344
Oilseed crops	-32	-35232
Other crops	-94	-82160
Root crops	34	11832
Root vegetables	252	65529

TABLE E.4: Data from Table E.3, expressed as a percent difference and actual difference (between 2015 and 2008) in harvested acres, aggregated by FAO Indicative Crop Classification 1.1 (see Appendix D.3).

E.3 Visual representations of annual total harvested acres

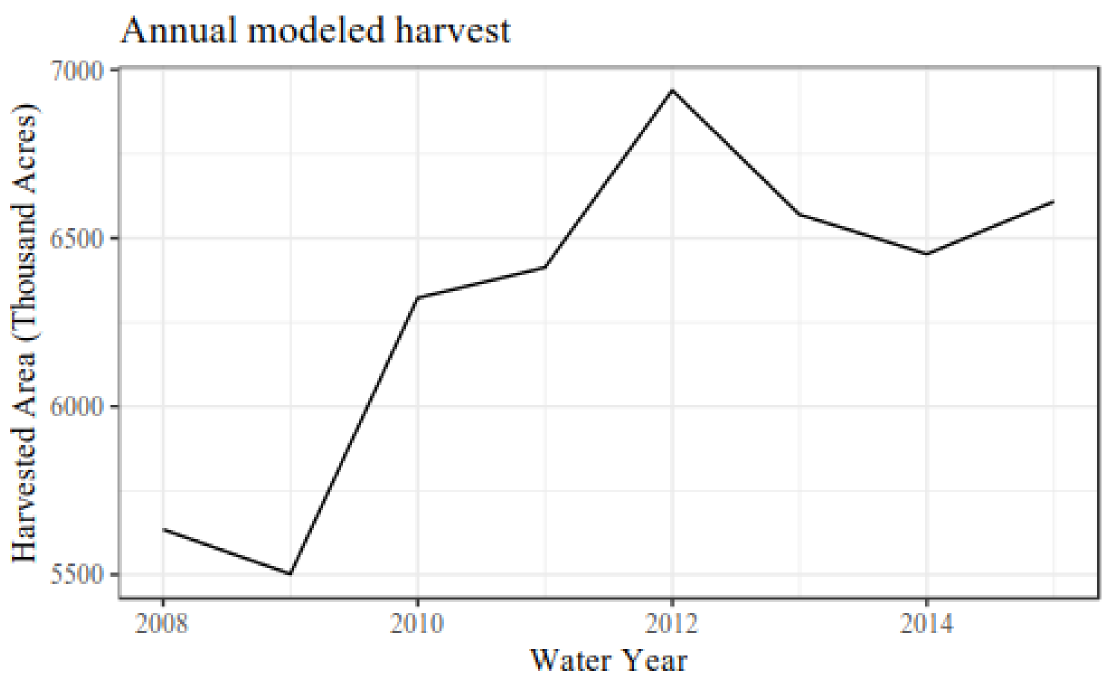

FIGURE E.1: Total annual harvested acres, across all modeled crops and counties.

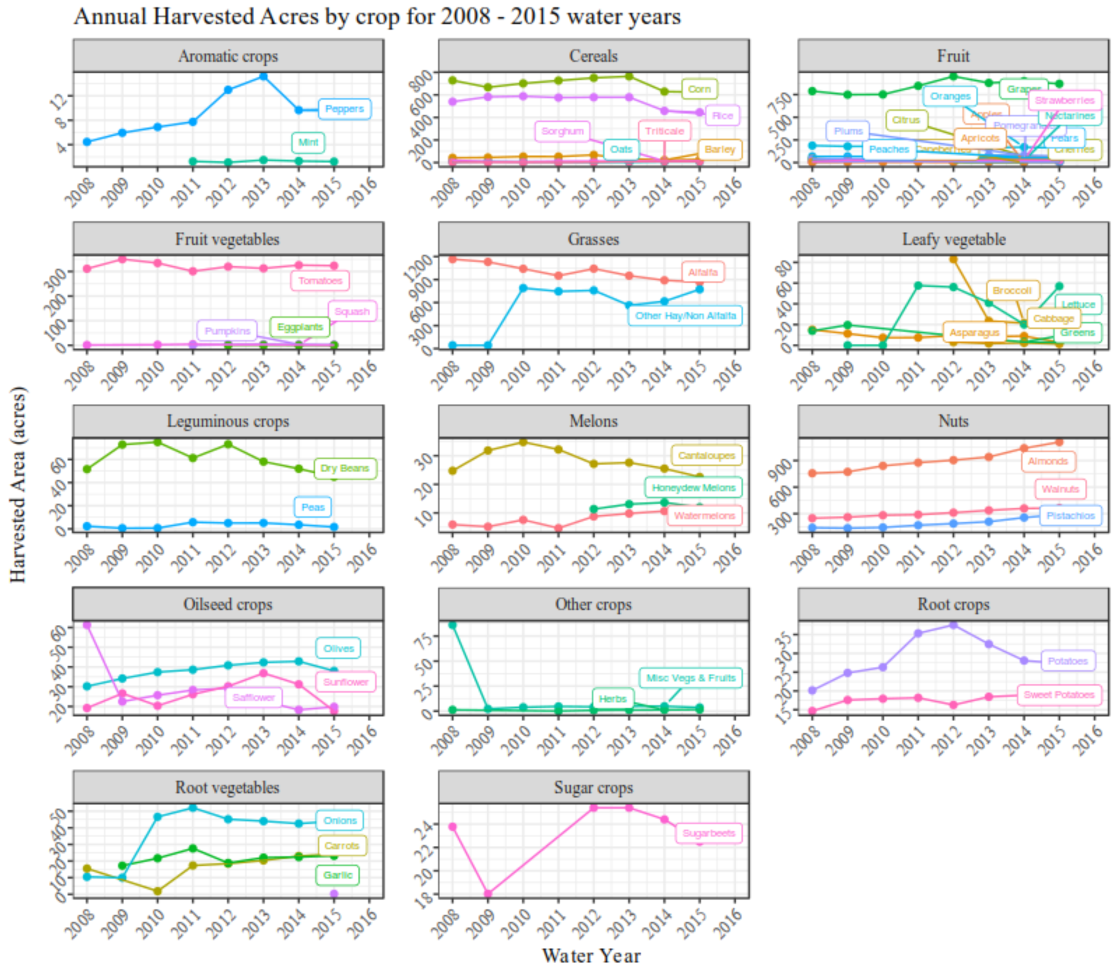

FIGURE E.2: Annual sums of harvested acres, by CDL crop and by FAO ICC 1.1. Harvested acres are reported by the County Ag Comissioners' reports and are included in the study only if there was an observation in the CDL for a given year.

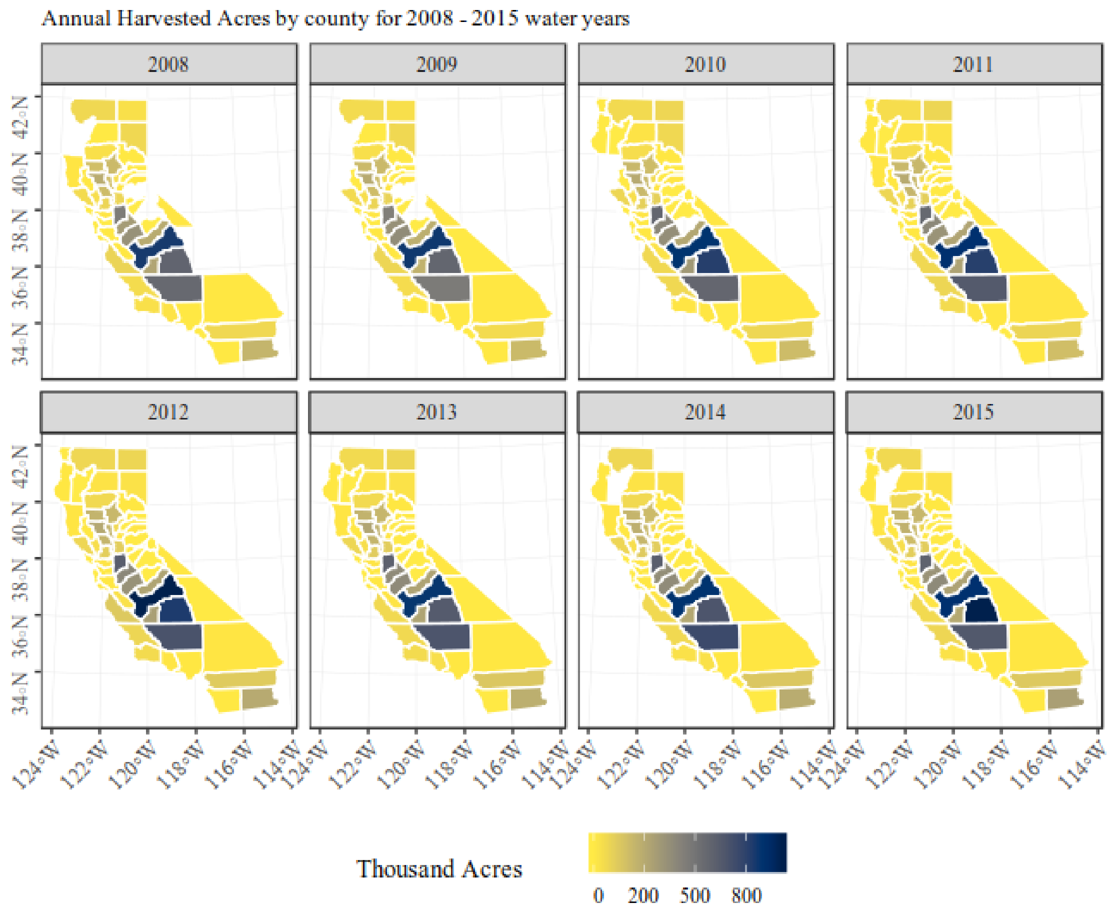

FIGURE E.3: Annual sums of harvested acres, by county, as reported by the County Ag Comissioners' reports, provided that there is an observation in the CDL for that year.

Appendix F

Extended results, plots, and tabulations: Crop water requirement

F.1 Tabulations of annual modeled crop water requirement

County	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alameda	4.86E+06	5.90E+06	6.10E+06	5.76E+07	4.12E+07	2.21E+07	1.05E+07	6.86E+06	**1.55E+08**
Alpine	NA	NA	8.08E+05	5.45E+05	1.78E+06	9.88E+05	1.82E+06	9.17E+05	**6.86E+06**
Amador	5.41E+06	3.66E+06	4.49E+06	3.76E+06	1.75E+07	1.47E+07	7.15E+06	5.18E+06	**6.19E+07**
Butte	4.08E+08	3.67E+08	3.81E+08	3.91E+08	3.91E+08	4.20E+08	4.37E+08	4.36E+08	**3.23E+09**
Calaveras	1.69E+06	1.46E+06	4.19E+06	1.86E+06	6.31E+06	4.38E+06	2.82E+06	2.15E+06	**2.49E+07**
Colusa	4.37E+08	4.41E+08	4.29E+08	4.28E+08	4.56E+08	4.91E+08	4.95E+08	4.94E+08	**3.67E+09**
Contra Costa	8.09E+07	6.65E+07	9.64E+07	9.53E+07	1.34E+08	1.05E+08	1.46E+08	1.28E+08	**8.51E+08**
Del Norte	NA	NA	9.74E+05	1.51E+05	2.35E+04	NA	NA	2.20E+07	**2.32E+07**
El Dorado	NA	NA	2.43E+05	8.36E+04	4.26E+06	4.92E+06	3.94E+05	2.51E+05	**1.02E+07**
Fresno	1.93E+09	1.76E+09	1.86E+09	1.90E+09	1.95E+09	1.94E+09	1.95E+09	1.83E+09	**1.51E+10**
Glenn	4.14E+08	4.37E+08	4.35E+08	4.42E+08	4.94E+08	4.97E+08	4.78E+08	4.71E+08	**3.67E+09**
Humboldt	NA	NA	2.26E+07	5.81E+07	6.03E+07	5.23E+07	9.24E+07	1.47E+07	**3.00E+08**
Imperial	2.92E+08	2.67E+08	6.80E+08	8.45E+08	1.09E+09	1.17E+09	1.12E+09	1.03E+09	**6.48E+09**
Inyo	NA	3.13E+06	8.22E+06	1.62E+07	1.75E+07	1.25E+07	1.42E+07	1.84E+07	**9.02E+07**
Kern	1.41E+09	1.25E+09	1.42E+09	1.55E+09	1.67E+09	1.63E+09	1.72E+09	1.58E+09	**1.22E+10**
Kings	9.04E+08	8.40E+08	8.86E+08	9.29E+08	1.02E+09	1.05E+09	1.04E+09	1.04E+09	**7.71E+09**
Lake	4.91E+06	6.50E+05	1.95E+05	3.91E+05	3.83E+07	4.24E+07	8.76E+06	8.52E+06	**1.04E+08**
Lassen	1.22E+08	2.15E+08	2.51E+08	1.91E+08	2.20E+08	2.08E+08	2.84E+08	2.24E+08	**1.72E+09**
Los Angeles	3.16E+07	3.97E+07	6.50E+04	3.95E+07	2.09E+07	1.55E+07	1.38E+07	2.61E+07	**1.87E+08**
Madera	6.68E+08	6.57E+08	6.08E+08	6.56E+08	5.45E+08	4.77E+08	4.30E+08	4.91E+08	**4.53E+09**
Marin	9.47E+03	1.89E+03	5.90E+05	9.94E+04	3.41E+05	4.74E+06	3.76E+06	1.77E+06	**1.13E+07**
Mariposa	1.00E+05	8.89E+03	NA	NA	2.83E+05	1.36E+05	3.85E+05	2.94E+05	**1.21E+06**
Mendocino	6.89E+05	NA	NA	1.32E+06	5.58E+07	6.72E+07	1.22E+07	8.86E+06	**1.46E+08**
Merced	9.16E+08	8.57E+08	9.19E+08	9.24E+08	1.04E+09	9.86E+08	1.11E+09	1.00E+09	**7.75E+09**
Modoc	1.96E+08	1.61E+08	4.70E+08	4.48E+08	3.95E+08	4.22E+08	NA	NA	**2.09E+09**
Mono	6.05E+06	1.26E+07	1.50E+07	1.46E+07	2.27E+06	2.12E+07	1.46E+07	2.33E+07	**1.10E+08**
Monterey	2.28E+08	1.67E+08	1.96E+08	1.41E+08	3.34E+08	3.12E+08	3.83E+08	3.52E+08	**2.11E+09**
Napa	3.74E+06	1.01E+06	1.12E+07	2.08E+06	1.24E+08	1.35E+08	1.64E+08	1.31E+08	**5.72E+08**
Nevada	6.03E+03	1.78E+03	8.94E+02	8.50E+02	1.03E+06	1.29E+06	7.19E+04	4.38E+03	**2.40E+06**
Orange	4.05E+04	1.43E+05	1.91E+05	1.33E+05	NA	NA	1.47E+05	1.29E+04	**6.67E+05**
Placer	6.56E+07	6.53E+07	8.68E+07	7.76E+07	8.03E+07	8.16E+07	9.32E+07	9.25E+07	**6.43E+08**
Plumas	3.69E+07	2.94E+07	4.77E+07	4.82E+07	2.81E+07	4.01E+07	6.70E+07	5.10E+07	**3.48E+08**
Riverside	3.45E+08	2.14E+08	3.54E+08	4.75E+08	5.44E+08	4.87E+08	5.27E+08	4.92E+08	**3.44E+09**
Sacramento	3.52E+08	2.75E+08	4.94E+08	4.36E+08	5.15E+08	4.90E+08	5.63E+08	5.50E+08	**3.67E+09**
San Benito	1.96E+07	5.12E+06	1.44E+07	2.49E+07	4.53E+07	3.20E+07	2.26E+07	1.46E+07	**1.79E+08**
San Bernardino	2.87E+07	5.05E+07	4.86E+07	3.79E+07	3.36E+07	2.49E+07	2.38E+07	2.77E+07	**2.76E+08**
San Diego	1.73E+06	5.29E+05	3.33E+04	2.93E+04	3.84E+05	9.68E+04	1.01E+05	1.84E+05	**3.08E+06**
San Joaquin	9.99E+08	7.90E+08	9.38E+08	9.88E+08	1.02E+09	1.01E+09	1.16E+09	1.10E+09	**8.01E+09**

San Luis Obispo	1.58E+08	1.16E+08	1.65E+08	1.34E+08	2.34E+08	2.21E+08	2.35E+08	2.70E+08	**1.53E+09**
San Mateo	2.13E+04	1.73E+03	3.21E+05	6.67E+04	1.71E+04	1.07E+05	2.51E+05	8.79E+04	**8.74E+05**
Santa Barbara	2.46E+07	1.81E+07	1.16E+08	2.43E+07	9.44E+07	1.02E+08	1.52E+08	1.03E+08	**6.35E+08**
Santa Clara	2.41E+07	1.78E+07	9.80E+06	1.43E+07	2.24E+07	1.81E+07	1.93E+07	1.15E+07	**1.37E+08**
Santa Cruz	2.34E+06	1.04E+06	3.34E+05	9.90E+04	3.07E+06	1.68E+06	2.34E+06	1.37E+06	**1.23E+07**
Shasta	3.74E+07	1.18E+07	8.43E+07	1.07E+08	1.12E+08	9.80E+07	1.30E+08	1.20E+08	**7.00E+08**
Sierra	6.01E+06	5.21E+06	1.36E+07	1.03E+07	6.16E+06	1.26E+07	2.17E+07	1.19E+07	**8.76E+07**
Siskiyou	3.32E+08	2.89E+08	4.06E+08	4.26E+08	5.06E+08	4.78E+08	4.74E+08	4.68E+08	**3.38E+09**
Solano	3.06E+08	2.77E+08	3.64E+08	3.08E+08	3.91E+08	4.12E+08	4.53E+08	4.51E+08	**2.96E+09**
Sonoma	1.67E+07	7.74E+06	1.56E+07	9.29E+06	1.95E+08	2.12E+08	2.49E+08	2.11E+08	**9.16E+08**
Stanislaus	6.41E+08	5.34E+08	6.43E+08	6.94E+08	8.62E+08	8.37E+08	8.88E+08	8.29E+08	**5.93E+09**
Sutter	4.66E+08	4.27E+08	4.10E+08	4.13E+08	4.62E+08	4.86E+08	5.23E+08	4.97E+08	**3.68E+09**
Tehama	2.46E+08	2.58E+08	3.06E+08	2.44E+08	2.47E+08	3.02E+08	2.47E+08	2.75E+08	**2.12E+09**
Trinity	NA	NA	1.13E+06	4.58E+05	4.37E+05	1.92E+05	NA	NA	**2.21E+06**
Tulare	1.96E+09	1.87E+09	1.78E+09	1.94E+09	2.01E+09	2.04E+09	2.14E+09	2.04E+09	**1.58E+10**
Tuolumne	NA	NA	4.49E+03	1.10E+03	1.97E+04	2.23E+04	4.35E+05	6.58E+04	**5.49E+05**
Ventura	1.36E+07	1.22E+06	1.17E+07	8.87E+06	4.59E+06	8.61E+06	2.08E+07	2.29E+06	**7.18E+07**
Yolo	6.53E+08	5.88E+08	6.71E+08	6.13E+08	7.36E+08	7.50E+08	8.27E+08	8.10E+08	**5.65E+09**
Yuba	2.10E+08	2.26E+08	2.36E+08	2.19E+08	2.07E+08	2.35E+08	2.47E+08	2.58E+08	**1.84E+09**
Total	**1.50E+10**	**1.36E+10**	**1.59E+10**	**1.64E+10**	**1.85E+10**	**1.85E+10**	**1.90E+10**	**1.80E+10**	**1.35E+11**

TABLE F.1: Annual modeled crop water requirement (CWR) by county, for cultivated areas as identified in the CDL.

CDL Name	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alfalfa	4.96E+09	4.64E+09	4.65E+09	4.53E+09	4.52E+09	4.36E+09	4.50E+09	4.39E+09	**3.65E+10**
Almonds	2.31E+09	2.07E+09	1.92E+09	1.92E+09	1.96E+09	2.04E+09	2.22E+09	2.32E+09	**1.68E+10**
Apples	5.30E+06	3.81E+06	3.87E+06	4.93E+06	4.91E+06	3.28E+06	4.56E+05	4.84E+05	**2.70E+07**
Apricots	1.47E+07	4.76E+07	1.34E+07	4.85E+06	1.05E+07	4.13E+06	5.82E+06	NA	**1.01E+08**
Asparagus	1.46E+07	1.22E+07	1.61E+06	9.46E+06	1.11E+07	1.17E+07	9.34E+06	3.18E+06	**7.33E+07**
Barley	1.31E+08	1.05E+08	1.00E+08	1.03E+08	1.41E+08	1.49E+08	1.47E+08	1.18E+08	**9.93E+08**
Broccoli	NA	NA	NA	NA	7.06E+06	1.02E+07	1.27E+07	5.98E+06	**3.59E+07**
Cabbage	NA	NA	NA	NA	1.18E+06	1.04E+06	1.23E+06	6.66E+05	**4.12E+06**
Caneberries	NA	NA	NA	NA	9.42E+05	NA	9.10E+05	NA	**1.85E+06**
Cantaloupes	5.98E+07	3.17E+07	7.15E+07	7.49E+07	6.15E+07	3.84E+07	3.91E+07	4.65E+07	**4.24E+08**
Carrots	4.79E+05	NA	1.33E+04	2.21E+07	2.95E+07	1.93E+07	3.17E+07	2.32E+07	**1.26E+08**
Cherries	1.40E+08	9.22E+07	8.82E+07	7.13E+07	1.38E+08	1.26E+08	1.26E+08	8.33E+07	**8.65E+08**
Citrus	7.56E+07	8.86E+07	7.53E+07	2.15E+08	2.34E+08	4.08E+08	2.90E+08	2.42E+08	**1.63E+09**
Corn	9.17E+08	5.94E+08	7.96E+08	7.97E+08	8.03E+08	7.67E+08	6.80E+08	5.94E+08	**5.95E+09**
Cotton	7.64E+08	6.69E+08	8.87E+08	7.96E+08	8.28E+08	6.99E+08	6.14E+08	5.09E+08	**5.77E+09**
Cucumbers	NA	NA	NA	NA	NA	6.04E+05	NA	NA	**6.04E+05**
Dry Beans	1.97E+07	1.61E+07	3.45E+07	4.18E+07	4.19E+07	4.15E+07	4.88E+07	2.18E+07	**2.66E+08**
Eggplants	NA	NA	NA	NA	3.41E+05	9.09E+04	2.04E+05	2.11E+05	**8.47E+05**
Garlic	NA	1.62E+07	1.43E+07	3.80E+07	3.64E+07	3.98E+07	4.05E+07	3.39E+07	**2.19E+08**
Grapes	1.06E+09	9.14E+08	1.06E+09	9.84E+08	2.11E+09	2.14E+09	2.43E+09	2.16E+09	**1.29E+10**
Greens	1.87E+07	1.68E+07	NA	NA	3.01E+07	1.91E+07	4.47E+07	6.32E+07	**1.92E+08**
Herbs	2.50E+06	NA	NA	4.63E+05	1.82E+06	2.09E+06	8.06E+05	4.32E+06	**1.20E+07**
Honeydew Melons	NA	NA	NA	NA	1.41E+07	1.46E+07	1.52E+07	1.68E+07	**6.07E+07**
Lettuce	NA	3.09E+03	5.14E+07	6.46E+07	6.38E+07	5.22E+07	4.91E+07	2.11E+07	**3.02E+08**
Mint	NA	NA	NA	1.42E+06	1.59E+06	2.47E+06	1.60E+06	2.12E+06	**9.19E+06**
Misc Vegs & Fruits	1.82E+07	3.59E+05	3.45E+06	9.05E+07	7.78E+07	5.89E+07	8.10E+07	8.29E+07	**4.13E+08**
Nectarines	2.34E+07	5.02E+06	1.12E+07	2.84E+07	1.33E+07	5.61E+06	2.85E+07	9.85E+06	**1.25E+08**
Oats	2.51E+07	2.17E+07	2.85E+07	2.59E+07	3.29E+07	1.93E+07	3.47E+07	3.65E+07	**2.25E+08**
Olives	2.63E+08	3.23E+08	3.45E+08	2.81E+08	2.02E+08	1.74E+08	1.60E+08	1.23E+08	**1.87E+09**
Onions	1.16E+07	9.76E+05	7.71E+06	1.83E+08	2.00E+08	1.63E+08	1.95E+08	1.62E+08	**9.23E+08**
Oranges	9.23E+08	8.17E+08	7.38E+08	7.48E+08	6.64E+08	7.39E+08	8.28E+08	8.09E+08	**6.27E+09**

Other Hay/ Non Alfalfa	8.29E+05	1.19E+08	1.73E+09	1.68E+09	1.89E+09	1.91E+09	1.81E+09	1.68E+09	**1.08E+10**
Peaches	5.59E+07	4.07E+07	7.17E+07	4.45E+07	4.57E+07	1.67E+08	3.89E+07	2.12E+07	**4.86E+08**
Pears	NA	NA	NA	NA	1.99E+07	1.90E+07	2.01E+07	1.82E+07	**7.72E+07**
Peas	7.82E+04	7.26E+02	4.33E+02	6.85E+05	1.10E+06	9.17E+05	7.50E+05	1.24E+05	**3.66E+06**
Peppers	4.27E+06	3.08E+05	2.88E+06	3.66E+06	1.68E+07	1.06E+07	8.13E+06	1.43E+07	**6.10E+07**
Pistachios	6.01E+08	5.35E+08	6.12E+08	8.16E+08	8.82E+08	8.78E+08	8.88E+08	9.18E+08	**6.13E+09**
Plums	3.84E+07	2.97E+07	6.80E+07	6.77E+07	2.07E+08	2.24E+08	1.74E+08	1.04E+08	**9.13E+08**
Pomegranates	2.44E+07	1.65E+07	2.90E+07	1.92E+07	2.14E+07	1.66E+07	1.75E+07	2.30E+07	**1.68E+08**
Potatoes	1.75E+07	2.97E+07	4.96E+07	5.97E+07	7.45E+07	6.11E+07	6.93E+07	7.31E+07	**4.34E+08**
Pumpkins	NA	NA	NA	1.43E+05	4.99E+03	8.70E+05	1.35E+06	4.31E+05	**2.79E+06**
Radishes	NA	NA	NA	NA	NA	NA	NA	5.32E+04	**5.32E+04**
Rice	5.04E+08	5.17E+08	5.24E+08	5.13E+08	5.15E+08	5.08E+08	5.05E+08	5.20E+08	**4.10E+09**
Safflower	2.18E+08	5.50E+07	1.12E+08	5.37E+07	5.25E+07	4.42E+07	3.63E+07	4.72E+07	**6.19E+08**
Sorghum	8.86E+06	8.13E+05	7.52E+06	1.73E+07	2.20E+07	2.10E+07	2.81E+07	2.47E+07	**1.30E+08**
Squash	1.13E+04	NA	9.41E+02	5.61E+04	2.66E+05	1.96E+05	5.09E+04	3.26E+02	**5.82E+05**
Strawberries	3.15E+07	2.66E+07	4.31E+07	1.98E+07	4.22E+07	2.44E+07	6.14E+07	4.65E+07	**2.95E+08**
Sugarbeets	2.36E+07	6.95E+05	NA	NA	1.58E+08	1.18E+08	1.41E+08	1.26E+08	**5.67E+08**
Sunflower	1.70E+07	5.01E+07	2.85E+07	5.69E+07	8.68E+07	9.46E+07	9.34E+07	4.33E+07	**4.71E+08**
Sweet Potatoes	2.87E+07	2.63E+07	2.48E+07	8.88E+06	3.64E+07	2.51E+07	3.61E+06	7.76E+06	**1.62E+08**
Tomatoes	3.12E+08	4.08E+08	3.85E+08	7.52E+08	8.25E+08	7.87E+08	8.72E+08	8.57E+08	**5.20E+09**
Triticale	5.00E+06	6.76E+06	5.15E+06	1.82E+06	7.45E+06	1.11E+07	1.79E+07	2.87E+07	**8.39E+07**
Walnuts	1.35E+09	1.23E+09	1.33E+09	1.24E+09	1.32E+09	1.43E+09	1.56E+09	1.60E+09	**1.11E+10**
Watermelons	3.36E+06	4.55E+07	4.54E+06	5.78E+06	2.32E+07	2.98E+07	2.52E+07	2.27E+07	**1.60E+08**
Total	**1.50E+10**	**1.36E+10**	**1.59E+10**	**1.64E+10**	**1.85E+10**	**1.85E+10**	**1.90E+10**	**1.80E+10**	**1.35E+11**

TABLE F.2: Annual modeled crop water requirement (CWR) by crop, for cultivated areas as identified in the CDL.

F.2 Visual representations of annual modeled crop water requirement

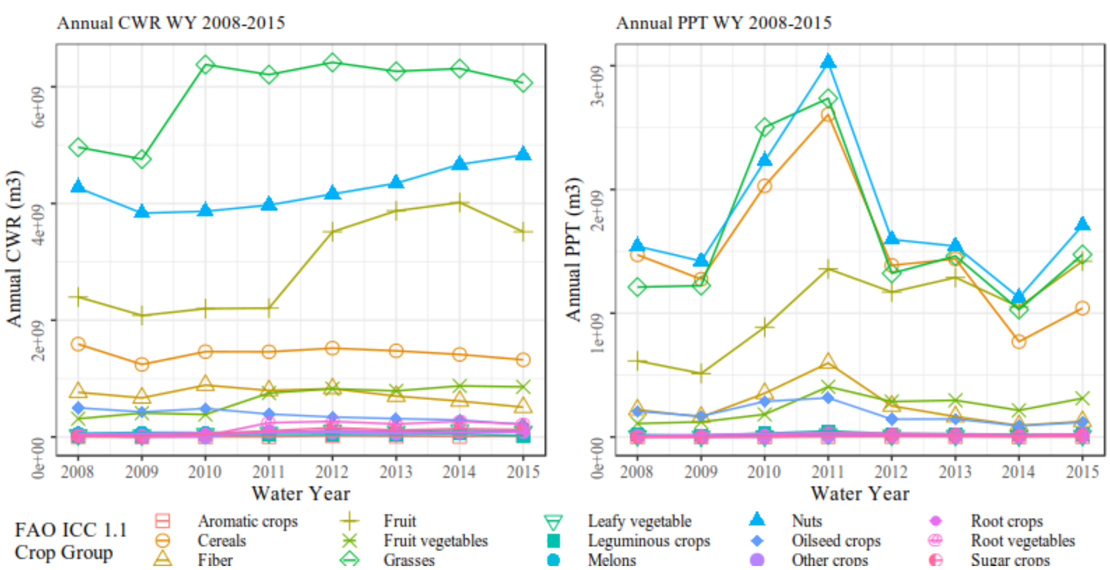

FIGURE F.1: Annual sum of modeled crop water requirement and observed precipitation over irrigated lands, aggregated by ICC 1.1 crop group

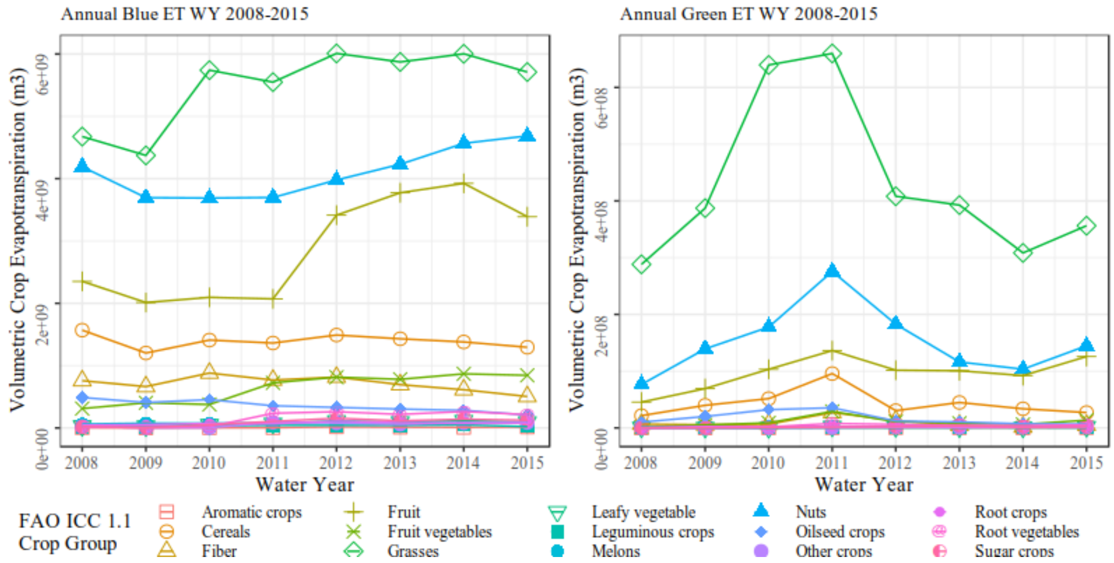

FIGURE F.2: Annual sum of modeled green and blue crop evapotranspiration, aggregated by ICC 1.1 crop group. Note the order of magnitude difference in the scale of the y-axis between green and blue ET.

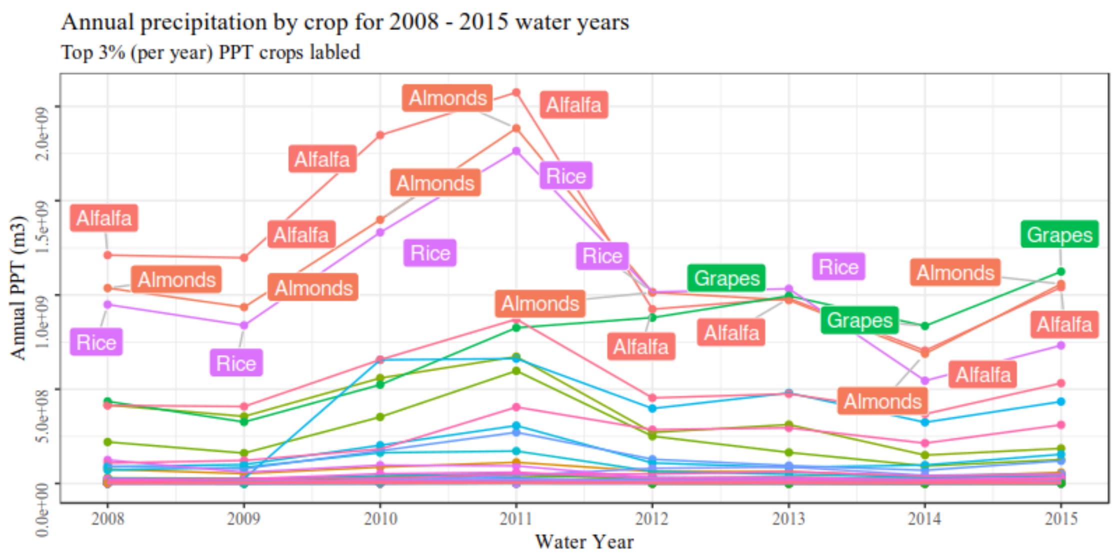

FIGURE F.3: Annual sum of precipitation over CDL irrigated crops. Observations in the 95% percentile of overall annual precipitation are highlighted.

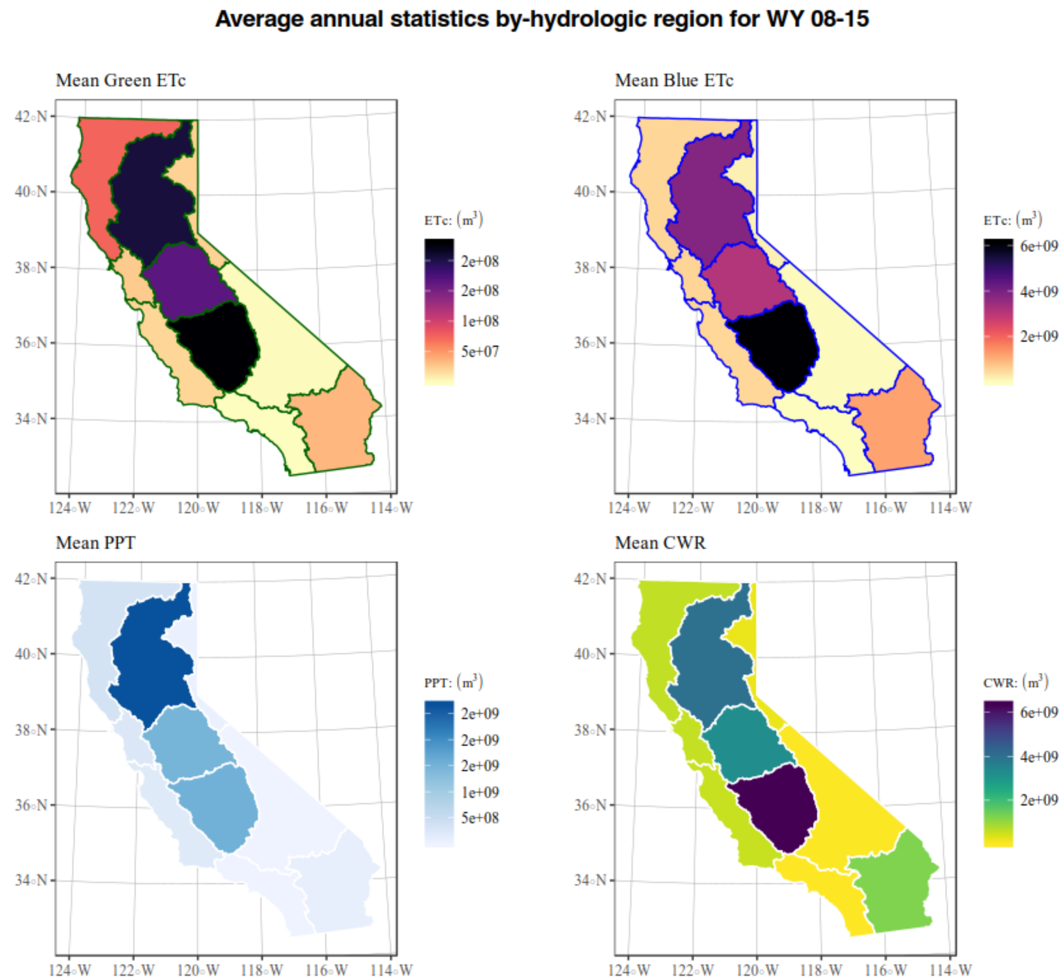

FIGURE F.4: Spatial distribution of mean annual crop evapotranspiration, precipitation (over irrigated acres), and crop water requirement, by DWR hydrological region.

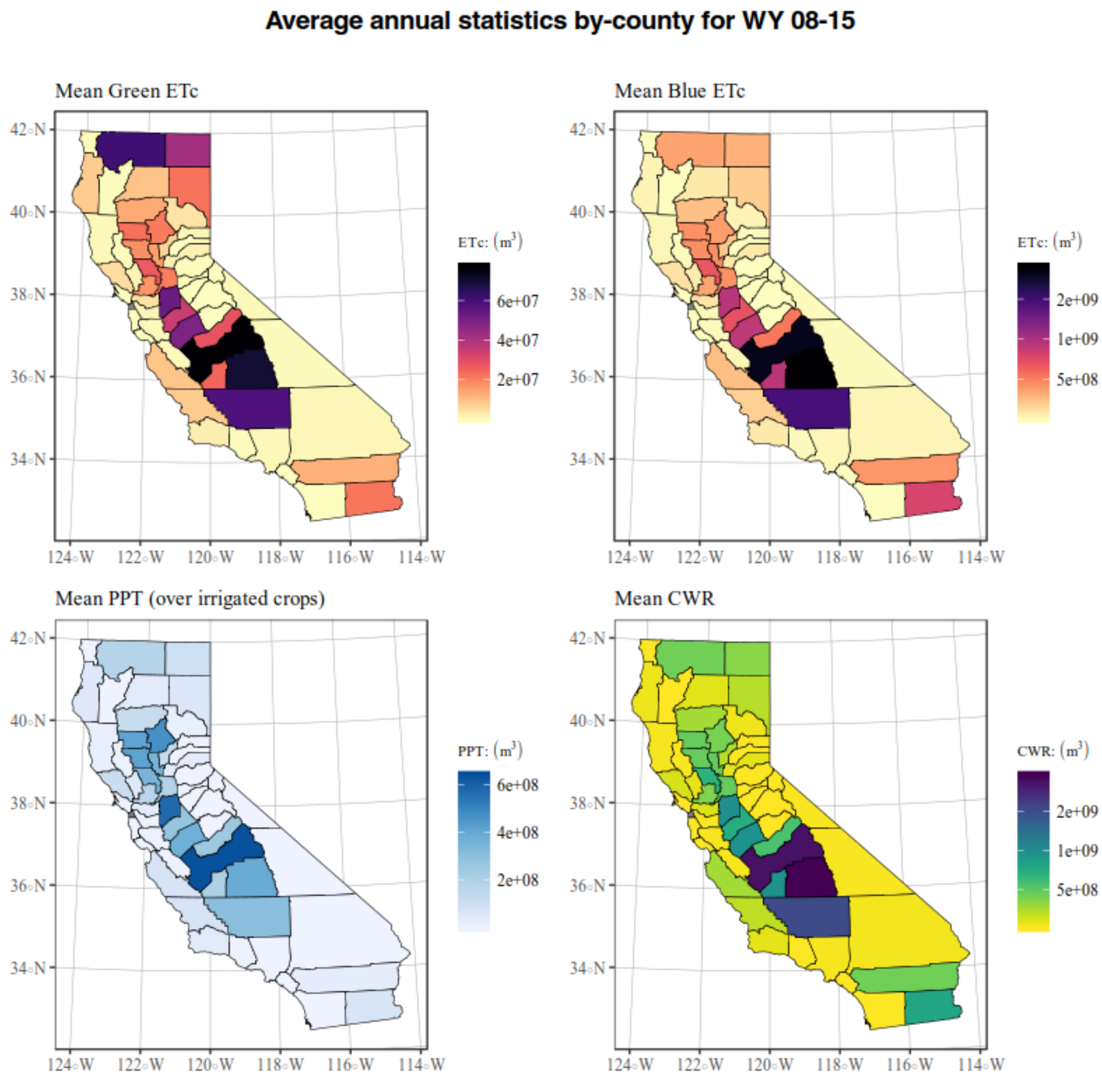

FIGURE F.5: Spatial distribution of mean annual crop evapotranspiration, precipitation (over irrigated acres), and crop water requirement, by county.

Appendix G

Extended results, plots, and tabulations: Water footprint

G.1 Tabulations of annual modeled water footprints

County	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alameda	2.03E+03	4.38E+03	1.20E+03	9.67E+03	5.97E+03	3190	3580	5.63E+03	**35700**
Alpine	NA	NA	2.73E+03	1.71E+03	6.53E+03	1090	1000	5.05E+02	**13600**
Amador	8.27E+03	6.68E+03	1.31E+04	1.09E+04	1.35E+04	8630	12400	1.21E+04	**85500**
Butte	2.74E+04	2.29E+04	1.79E+04	1.70E+04	1.15E+04	12700	15700	1.58E+04	**141000**
Calaveras	4.23E+03	2.67E+03	5.39E+03	2.88E+03	5.82E+03	4220	4070	2.41E+03	**31700**
Colusa	1.49E+04	2.53E+04	2.95E+04	2.25E+04	2.36E+04	25300	23200	1.72E+04	**182000**
Contra Costa	9.18E+03	1.44E+04	1.53E+04	1.30E+04	1.30E+04	11600	21000	1.55E+04	**113000**
Del Norte	NA	NA	1.72E+02	2.58E+01	3.83E+00	NA	NA	2.48E+03	**2680**
El Dorado	NA	NA	9.36E+02	8.11E+02	2.33E+03	1240	644	1.64E+03	**7610**
Fresno	5.20E+04	9.29E+04	2.68E+04	2.59E+04	2.79E+04	28300	43700	4.35E+04	**341000**
Glenn	2.27E+04	2.69E+04	1.97E+04	3.01E+04	2.33E+04	25600	24600	2.59E+04	**199000**
Humboldt	NA	NA	7.49E+02	1.61E+03	2.03E+03	2370	4190	6.65E+02	**11600**
Imperial	1.66E+03	1.77E+03	1.55E+03	4.32E+03	6.89E+03	12400	11200	6.88E+03	**46700**
Inyo	NA	1.66E+02	4.36E+02	8.38E+02	1.61E+03	1020	997	1.31E+03	**6380**
Kern	3.29E+04	2.05E+04	2.11E+04	2.90E+04	3.04E+04	27800	40400	1.79E+04	**220000**
Kings	3.45E+04	3.59E+04	2.28E+04	1.99E+04	2.20E+04	23100	24200	3.04E+04	**213000**
Lake	7.97E+03	5.48E+02	1.60E+02	3.01E+02	1.36E+03	1140	1250	3.25E+02	**13100**
Lassen	6.78E+03	5.59E+03	9.44E+03	1.75E+03	2.07E+03	1980	2750	1.95E+03	**32300**
Los Angeles	7.26E+02	9.29E+02	1.44E+02	1.93E+03	6.93E+02	687	1830	1.60E+03	**8540**
Madera	1.94E+04	3.89E+04	1.08E+04	1.72E+04	1.19E+04	10700	12300	1.28E+04	**134000**
Marin	5.25E+01	8.49E+00	4.61E+02	3.72E+01	9.26E+02	17000	11900	1.53E+04	**45800**
Mariposa	7.46E+02	9.80E+01	NA	NA	3.02E+03	1180	7190	8.53E+03	**20800**
Mendocino	1.66E+01	NA	NA	2.54E+01	8.66E+02	951	217	1.70E+02	**2250**
Merced	1.37E+04	1.91E+04	1.50E+04	2.13E+04	1.69E+04	13200	13700	1.45E+04	**127000**
Modoc	4.23E+03	5.29E+03	1.34E+04	1.36E+04	8.68E+03	11500	NA	NA	**56700**
Mono	1.35E+02	2.05E+02	2.41E+02	2.38E+02	4.62E+01	559	306	5.08E+02	**2240**
Monterey	1.40E+05	7.33E+04	8.21E+04	7.47E+04	4.81E+04	73200	100000	5.63E+04	**648000**
Napa	2.98E+03	2.55E+03	3.83E+03	1.04E+03	9.59E+02	1050	2390	1.78E+03	**16600**
Nevada	5.59E+00	7.59E-01	1.34E+00	5.27E-01	1.32E+03	1140	60.1	7.32E+00	**2540**
Orange	4.72E+01	3.02E+01	9.04E+00	6.18E+00	NA	NA	13.7	1.09E+00	**107**
Placer	7.27E+03	7.43E+03	9.64E+03	3.90E+04	2.26E+04	20200	33100	2.23E+04	**161000**
Plumas	2.46E+03	1.77E+03	3.47E+03	3.79E+03	2.70E+03	3320	5450	5.18E+03	**28100**
Riverside	6.60E+03	9.31E+05	3.57E+04	1.66E+04	1.59E+04	9740	21500	1.74E+04	**1050000**
Sacramento	1.22E+04	6.55E+03	3.87E+04	2.49E+04	2.26E+04	41600	22700	1.94E+04	**189000**
San Benito	1.12E+04	2.71E+03	4.95E+03	2.84E+03	2.57E+03	2400	8440	2.43E+04	**59400**
San Bernardino	7.07E+02	8.39E+02	1.37E+03	6.41E+02	6.67E+02	3130	862	7.38E+02	**8960**
San Diego	4.09E+03	2.39E+01	1.72E+00	5.37E+01	1.03E+01	10.8	14.5	5.96E+01	**4260**
San Joaquin	1.21E+04	9.31E+03	1.27E+04	1.27E+04	1.10E+04	12000	11900	1.23E+04	**94100**

San Luis Obispo	1.46E+04	1.04E+04	6.32E+03	8.54E+03	9.93E+03	28400	56600	1.92E+04	**154000**
San Mateo	9.79E+01	1.11E+01	1.12E+03	1.11E+02	3.45E+01	134	5360	1.14E+03	**8020**
Santa Barbara	2.42E+03	8.71E+02	5.99E+03	2.76E+03	3.67E+03	2810	1710	1.48E+03	**21700**
Santa Clara	1.85E+04	1.61E+04	1.42E+04	7.87E+03	3.67E+03	4410	3950	1.17E+04	**80500**
Santa Cruz	1.35E+03	1.31E+02	1.94E+02	2.80E+01	1.10E+03	890	733	5.47E+02	**4980**
Shasta	8.43E+03	3.23E+03	3.37E+03	2.33E+04	3.34E+04	197000	7830	6.28E+04	**339000**
Sierra	1.38E+03	1.71E+04	3.36E+03	2.54E+03	1.58E+03	3280	5660	3.77E+03	**38700**
Siskiyou	2.66E+03	2.39E+03	5.26E+03	3.11E+04	4.58E+04	19400	44400	4.57E+04	**197000**
Solano	2.52E+04	2.06E+04	1.97E+04	1.99E+04	2.73E+04	41100	39500	2.54E+04	**219000**
Sonoma	4.91E+04	1.44E+04	9.49E+03	9.14E+03	6.37E+03	9550	41700	5.60E+04	**196000**
Stanislaus	8.57E+03	5.66E+03	6.09E+03	5.55E+03	5.91E+03	6860	9510	8.64E+03	**56800**
Sutter	3.82E+04	3.68E+04	5.65E+04	2.51E+04	4.16E+04	31400	38100	3.55E+04	**303000**
Tehama	2.97E+04	3.52E+04	3.13E+04	3.98E+04	4.44E+04	43600	23300	3.22E+04	**280000**
Trinity	NA	NA	1.86E+03	7.55E+02	1.41E+03	507	NA	NA	**4530**
Tulare	3.56E+04	5.87E+04	2.73E+04	2.78E+04	2.63E+04	23500	38100	2.85E+04	**266000**
Tuolumne	NA	NA	7.17E+00	1.76E+00	3.15E+01	26.2	513	9.59E+01	**675**
Ventura	1.59E+03	7.76E+01	4.51E+01	1.21E+03	1.65E+02	890	1000	1.01E+03	**5990**
Yolo	2.13E+04	1.58E+04	2.19E+04	1.42E+04	2.06E+04	18100	19200	1.95E+04	**151000**
Yuba	6.39E+04	4.85E+04	6.00E+04	7.16E+04	1.83E+04	28400	33100	8.17E+04	**405000**
Total	**7.85E+05**	**1.65E+06**	**6.96E+05**	**7.14E+05**	**6.63E+05**	**875000**	**860000**	**8.50E+05**	**7090000**

TABLE G.1: Annual modeled total water footprint of cultivated agriculture by county, for cultivated areas as identified in the CDL.

cdl.name	2008	2009	2010	2011	2012	2013	2014	2015	**All Years**
Alfalfa	1.50E+05	1.23E+05	1.23E+05	118000	7.55E+04	68600	92800	7.73E+04	**8.27E+05**
Almonds	1.68E+05	1.32E+05	1.15E+05	110000	7.91E+04	75200	93800	1.54E+05	**9.28E+05**
Apples	9.87E+02	5.73E+02	7.69E+02	275	3.59E+02	49.8	16.2	1.14E+01	**3.04E+03**
Apricots	2.49E+03	1.29E+04	2.56E+03	595	5.89E+02	524	1040	NA	**2.07E+04**
Asparagus	2.87E+03	3.47E+03	4.23E+02	981	9.82E+02	1890	1340	4.31E+02	**1.24E+04**
Barley	2.09E+04	2.57E+04	2.29E+04	16300	1.71E+04	41100	98500	3.50E+04	**2.78E+05**
Broccoli	NA	NA	NA	NA	8.76E+01	257	256	1.13E+02	**7.13E+02**
Cabbage	NA	NA	NA	NA	4.54E+01	51.6	40.2	3.50E+01	**1.72E+02**
Caneberries	NA	NA	NA	NA	7.43E+01	NA	100	NA	**1.75E+02**
Cantaloupes	3.04E+02	4.37E+02	6.48E+02	1110	1.22E+03	874	1250	8.37E+02	**6.67E+03**
Carrots	1.21E+00	NA	2.50E-01	106	1.06E+02	126	97.1	2.06E+02	**6.42E+02**
Cherries	1.17E+04	1.12E+04	1.27E+04	7480	1.41E+04	11700	35600	3.33E+04	**1.38E+05**
Citrus	9.17E+02	8.51E+02	8.29E+02	2430	3.16E+03	7970	4010	2.34E+03	**2.25E+04**
Corn	4.67E+03	6.12E+03	4.59E+03	2850	3.44E+03	2780	2990	2.74E+03	**3.02E+04**
Cotton	1.76E+04	3.24E+04	1.60E+04	12200	1.22E+04	14300	16100	1.64E+04	**1.37E+05**
Cucumbers	NA	NA	NA	NA	NA	35.6	NA	NA	**3.56E+01**
Dry Beans	4.22E+03	3.25E+03	5.87E+03	12700	2.07E+04	12800	27400	7.29E+03	**9.42E+04**
Eggplants	NA	NA	NA	NA	9.25E+01	15.1	49.5	5.59E+01	**2.13E+02**
Garlic	NA	1.14E+02	8.09E+01	323	5.92E+02	436	503	6.87E+02	**2.73E+03**
Grapes	1.85E+04	9.56E+03	9.03E+03	11200	3.56E+04	48900	41200	6.29E+04	**2.37E+05**
Greens	1.61E+02	6.54E+02	NA	NA	1.42E+03	1140	4690	2.73E+03	**1.08E+04**
Herbs	6.81E+02	NA	NA	659	4.93E+02	486	216	5.39E+02	**3.07E+03**
Honeydew Melons	NA	NA	NA	NA	3.15E+02	573	759	4.46E+02	**2.09E+03**
Lettuce	NA	5.45E-03	9.74E+01	657	1.79E+02	131	200	1.32E+02	**1.40E+03**
Mint	NA	NA	NA	41400	6.92E+04	205000	42000	8.75E+04	**4.45E+05**
Misc Vegs & Fruits	1.86E+03	5.86E+01	4.97E+02	5260	4.86E+03	2080	14100	3.48E+03	**3.22E+04**
Nectarines	2.35E+03	1.56E+02	2.66E+02	393	7.51E+02	510	534	1.91E+02	**5.16E+03**
Oats	5.79E+04	9.48E+05	4.83E+04	24800	2.14E+04	17100	52800	6.76E+04	**1.24E+06**
Olives	4.38E+04	1.30E+05	3.34E+04	54100	2.11E+04	30200	24100	1.17E+04	**3.48E+05**
Onions	4.83E+01	3.90E+00	4.19E+01	1310	1.65E+03	1060	1150	1.48E+03	**6.75E+03**
Oranges	6.36E+03	3.58E+03	4.05E+03	2030	1.03E+03	1270	1530	1.58E+03	**2.14E+04**
Other Hay/ Non Alfalfa	4.30E+01	8.41E+02	5.34E+04	78000	5.46E+04	83900	93300	7.07E+04	**4.35E+05**
Peaches	1.97E+03	2.55E+03	2.32E+03	1720	2.76E+03	2070	902	3.42E+02	**1.46E+04**

Pears	NA	NA	NA	NA	6.16E+03	1530	3430	3.21E+03	**1.43E+04**
Peas	4.76E+01	3.58E-01	1.15E-01	2280	2.02E+03	1000	1480	1.73E+02	**7.00E+03**
Peppers	2.88E+02	2.01E+01	1.04E+02	109	4.96E+02	404	163	3.00E+02	**1.88E+03**
Pistachios	4.20E+04	4.37E+04	3.93E+04	46600	5.66E+04	56600	35000	5.55E+04	**3.75E+05**
Plums	9.87E+02	8.68E+02	1.13E+03	6110	1.91E+04	24300	29600	1.03E+04	**9.25E+04**
Pomegranates	1.44E+03	5.84E+02	1.48E+03	787	6.10E+02	2090	1570	5.67E+02	**9.13E+03**
Potatoes	9.34E+01	3.95E+02	3.78E+02	561	5.20E+02	600	493	4.63E+02	**3.50E+03**
Pumpkins	NA	NA	NA	69.8	1.25E-01	31.4	83.9	8.01E+00	**1.93E+02**
Radishes	NA	NA	NA	NA	NA	NA	NA	4.04E+01	**4.04E+01**
Rice	8.42E+03	7.88E+03	8.17E+03	6930	7.23E+03	5940	5380	6.27E+03	**5.62E+04**
Safflower	3.69E+04	2.04E+04	3.00E+04	16700	1.47E+04	13000	11800	1.57E+04	**1.59E+05**
Sorghum	2.01E+03	1.24E+03	2.13E+04	114	1.99E+02	110	146	8.39E+02	**2.60E+04**
Squash	2.27E+00	NA	1.31E-01	24.8	4.74E+02	153	21.2	3.65E-01	**6.75E+02**
Strawberries	6.21E+02	3.58E+02	3.98E+02	190	1.83E+02	132	1540	7.58E+02	**4.18E+03**
Sugarbeets	2.88E+01	9.40E-01	NA	NA	1.47E+02	110	147	1.41E+02	**5.75E+02**
Sunflower	9.64E+03	1.41E+04	1.28E+04	21000	2.15E+04	22500	23900	2.28E+04	**1.48E+05**
Sweet Potatoes	2.82E+02	1.48E+02	1.26E+02	83.5	2.64E+02	154	61.4	7.45E+01	**1.19E+03**
Tomatoes	1.69E+03	7.49E+02	1.26E+03	1960	2.52E+03	1870	2610	2.21E+03	**1.49E+04**
Triticale	4.31E+03	4.21E+03	4.54E+03	831	7.08E+02	545	686	3.30E+03	**1.91E+04**
Walnuts	1.58E+05	1.05E+05	1.18E+05	103000	8.40E+04	110000	87500	8.44E+04	**8.50E+05**
Watermelons	6.26E+01	5.93E+02	4.76E+01	115	6.98E+02	760	453	4.12E+02	**3.14E+03**
Total	**7.85E+05**	**1.65E+06**	**6.96E+05**	**714000**	**6.63E+05**	**875000**	**860000**	**8.50E+05**	**7.09E+06**

TABLE G.2: Annual modeled total water footprint of cultivated agriculture by crop, for cultivated areas as identified in the CDL.

FAO ICC 1.1 Group	% Difference	Difference
Aromatic crops	4	12
Cereals	18	17546
Fiber	491	86291
Fruit	139	67240
Fruit vegetables	35	583
Grasses	-1	-1979
Leafy vegetable	14	413
Leguminous crops	75	3198
Melons	362	1329
Nuts	-20	-73854
Oilseed crops	-44	-40144
Other crops	58	1482
Root crops	43	162
Root vegetables	4776	2364

TABLE G.3: Data from Table G.2, expressed as a percent difference and actual difference (between 2015 and 2008) in total water footprint, aggregated by FAO Indicative Crop Classification 1.1 (see Appendix D.3).

G.2 Visual representations of annual water footprints

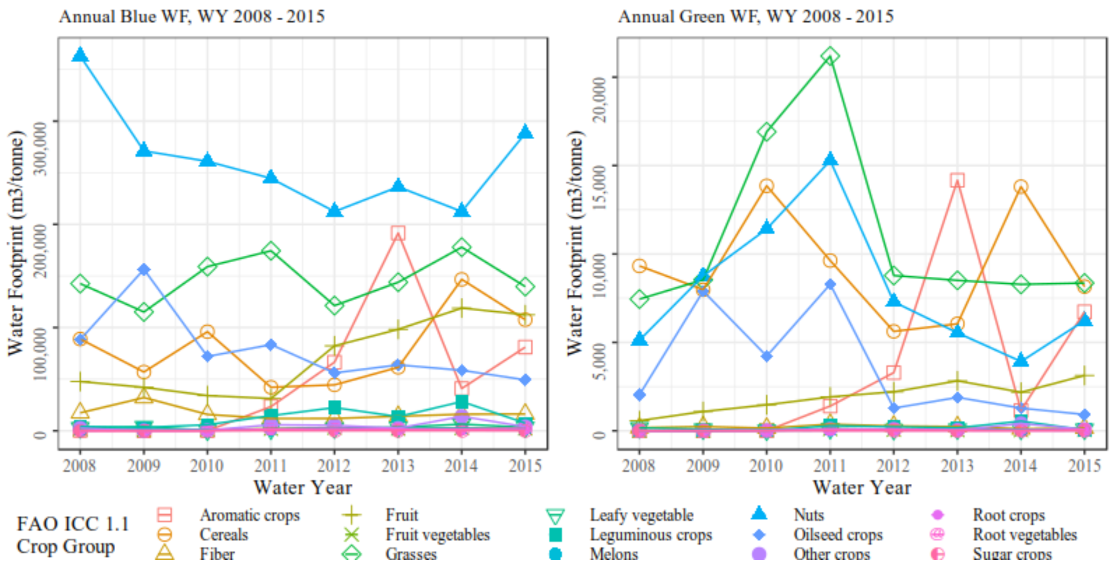

FIGURE G.1: Annual blue and green water footprint by ICC group, expressed as cubic meters of water per metric ton of harvested product.

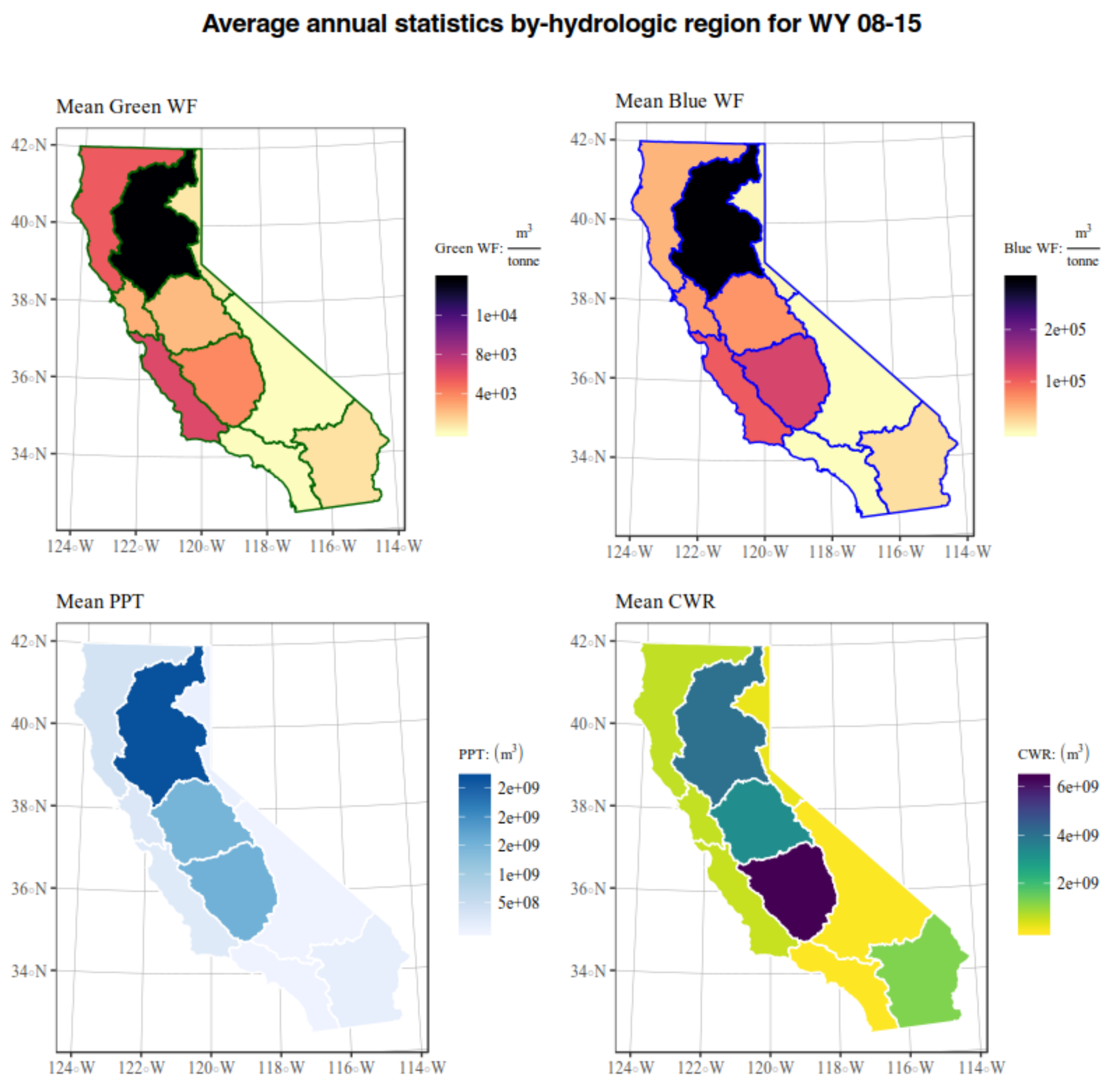

FIGURE G.2: Spatial distribution of mean annual water footprint, precipitation (over irrigated acres), and crop water requirement, by DWR hydrological region.

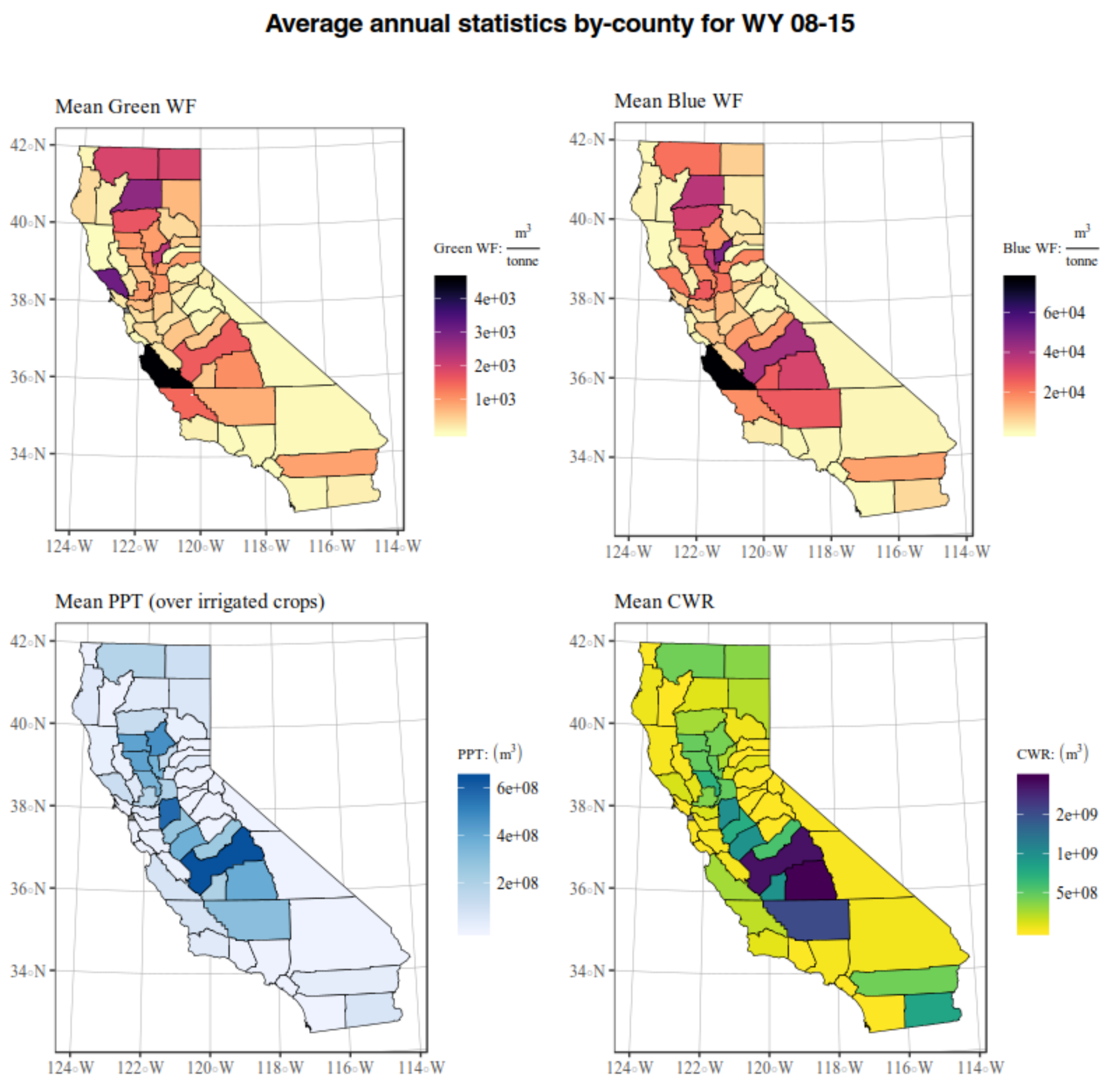

FIGURE G.3: Spatial distribution of mean annual water footprint, precipitation (over irrigated acres), and crop water requirement, by county.

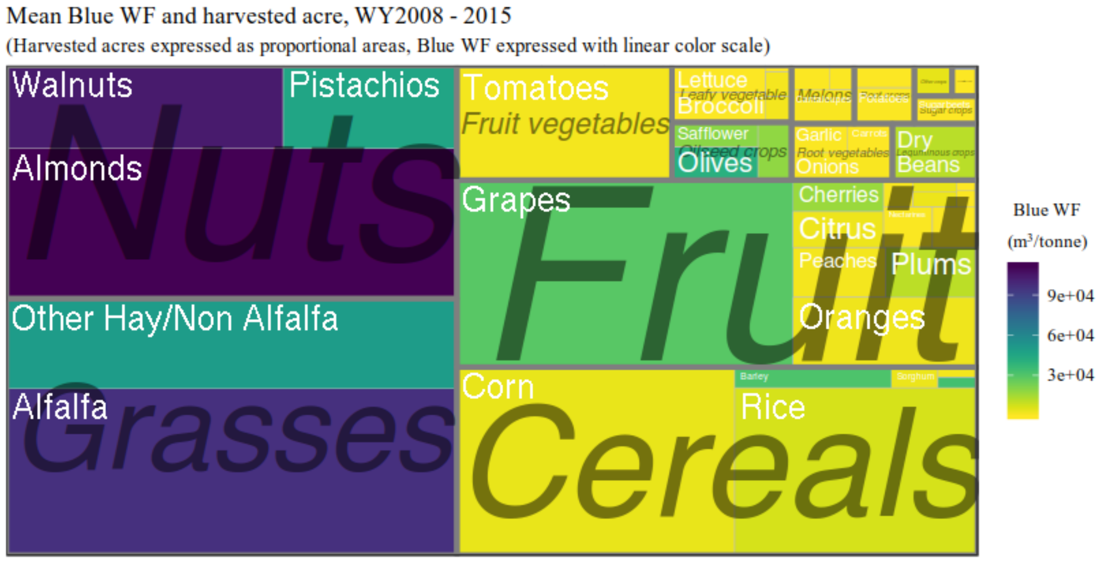

FIGURE G.4: Treeplot of mean annual blue water footprint expressed on a logarithmic color scale and mean annual harvested acres expressed as a proportion of the total statewide harvested acres. Crops are further grouped by ICC group.

Appendix H

Validation - Delta Crop ET Comparative Study

H.1 Background Information

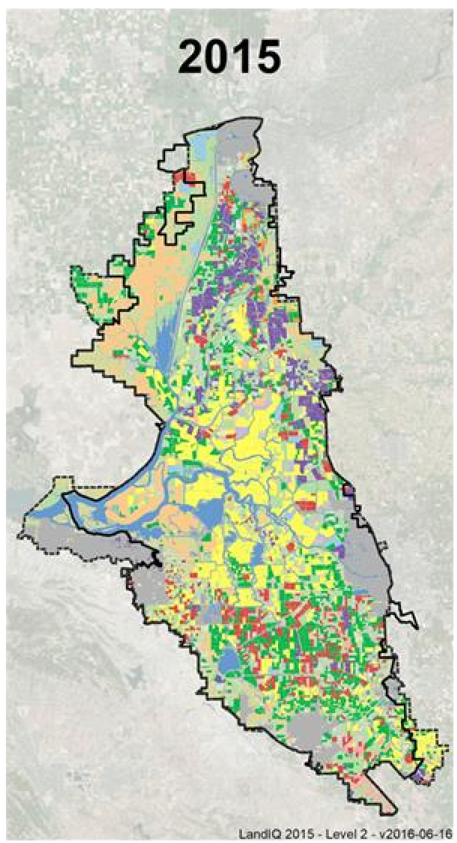

FIGURE H.1: Region modeled in the 2015 Delta Crop ET Comparative Study. The modeled "Delta Service Region" is marginally larger than the legal boundaries, which do not extend as far to the West. Map from (Medellín-Azuara et al., 2018)

	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	**Cuml**
TAF (ucm)	55	20	14	16	33	66	108	127	224	242	183	99	**1198**
% diff	36.6	56.5	55.5	49.9	48.6	44	31.5	21	4.33	6.38	18.7	35.7	**22.2**
ratio	63.4	43.5	44.5	50.1	51.4	56	68.5	79	95.7	93.6	81.3	64.3	**77.8**

TABLE H.1: Monthly difference between Delta ET and this study. Data from (Medellín-Azuara et al., 2018).

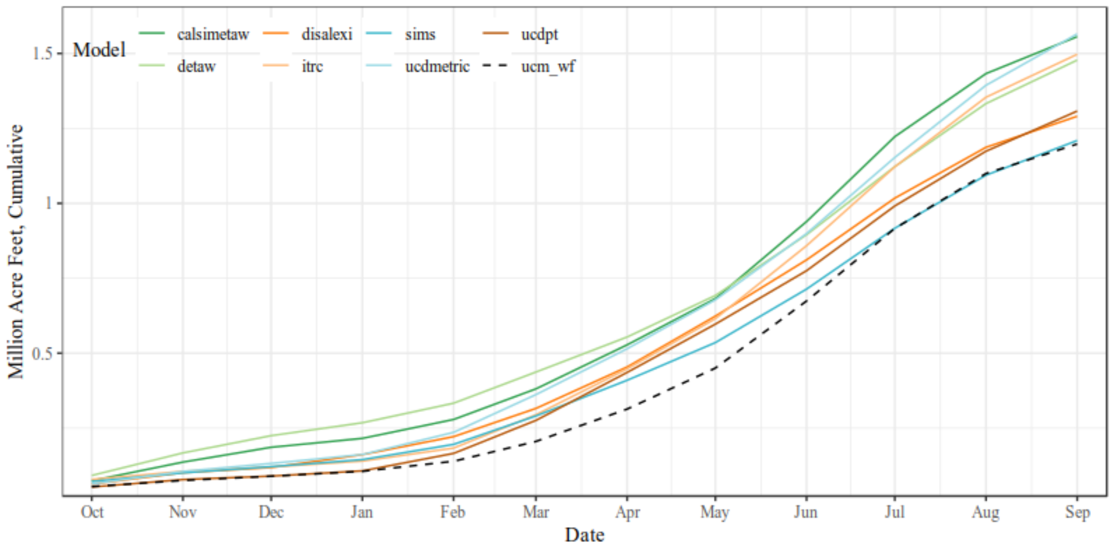

FIGURE H.2: Cumulative crop ET modeled in the Delta Service Region for the 2015 water year. Calculations from this study labeled "ucm_wf". Data from (Medellín-Azuara et al., 2018).

Appendix I

Analysis Scripts

The analytic portions of this study were spread across 9 individual files. Aggregation and data transformation operations (for map and chart creation) are not included in this appendix. Script purposes are as follows:

- `2.kc_prep.Rmd` creates daily crop coefficients and encodes the relation between crop coefficients and landcover classes
- `3.Lc_prep.Rmd` reconciles the landcover rasters to the project coordinate reference system (CRS)
- `4.eto_prep.Rmd` reconciles the reference et rasters to the project CRS
- `5.cwu_calcs.Rmd` executes daily, cell-wise crop ET calculation
- `6.ppt_prep.Rmd` reconciles the precipitation rasters to the project CRS
- `7.zone_sums.Rmd` computes precipitation over irrigated crops and crop water requirements, per crop, per region of interest
- `8.zone_prep.Rmd` disentangles the result of the previous script, assigns metadata, and performs unit conversions as necessary
- `9.yield_prep.Rmd` reconciles yield records with the crop classes used in the crop ET model
- `10.wf_calc.Rmd` computes the water footprint and aggregates harvested acreage from the yield records

The scripts begin at #2. There is no script prefixed with 1.

2.kc_prep.Rmd

In this file, we create the Kc rasters using Kc values compiled for the CUP+ model. Out of 70 crops that we have Kc values for,

We also import Kc data compiled by Snyder, Orang, Bali, Eching, and Zaccaria 2000 (revised 2014). The data was compiled for the `Basic Irrigation Scheduling` model (BISe), an Excel application that estimates irrigation water requirement and crop ET using the ASCE-PM equation. Kcs are found on the `CropRef` worksheet of the BISe application. Remarks are as follows:

> Kc data marked in blue were derived in work by T.C. Hsiao and former students at UC Davis.
>
> The Kc for corn was derived by Steduto and Hsiao (1998) maize canopies uhnder two soil water regimes II. Seasonal trends of evapotranspiration, carbon dioxide assimilation and canopy conductance, and as related to leaf area index. Agric. and forest Meteorol. 89:185-200. The Kc =1.05 for cotton is based on work by Held and Hsiao The Kc = 1.00 for sorghum is based on work by Held and Hsiao. Millet and For tomato a Kc = 1.10 was selected based on unpublished data from Snyder and Cahn and on expeiments by Held & Hsiao. The kc values reported by Held and Hsiao were slightly higher, but the tomatoes were full canopy (not in beds, which is the normal practice). The data from Snyder and Cahn were typical for California practices. The data for sunflower were based on data from Hsiao (personal communication)
>
> Kc data marked in green were derived from several sources. The assumption is that corn has a Kc = 1.00 for ETo calculated using the Pruitt and Doorenbos (1977) hourly ETo equation that is used by the California Irrigation Management Information System CIMIS
>
> Snyder and Pruitt (1992) Evapotranspiration Data Management in California Irrigation & Drainage Session Proceedings/Water Forum '92, EE,HY,IR,WR Div/ASCE, Baltimore, MD/August 2-6, 1992. pp128-133..
>
> Relative Kp values for alfalfa for the crops marked in green were selected from
>
> Wright (1982) New Evapotranspiration Crop Coefficients. Presented at Irrigation and Drainage Specialty Conference, ASCE, July 17-20, Albuquerque, New Mexico. pp 57-74.
>
> The peak Kp values were corn = 0.95, alfalfa = 1.0, beans = 1.0, potatoes = 0.8, sugar beets = 1.0, peas = 0.9, and cereals = 1.0. Because the equation for ETo was not available at that time, the Kp values cannot be used directly. However, assuming the Kc = 1.00 is correct for corn, then the approximate peak Kc values for the other crops are found by dividing the Kp by 0.95. The peak Kc values for a grass ETo are corn = 1.00, alfalfa = 1.05, beans = 1.00, potatoes = 0.85, sugar beets = 1.05, peas =0.95, and cereals = 1.05.
>
> The rice Kc = 1.1 more current data on typical ETo from CIMIs and the article
>
> Lourence and Pruitt (1971) Energy balance and water use of rice grown in the Central Valley of California. Agron. J. 63:827-832.
>
> Lourence and Pruitt found ET of rice to be about 4-5% higher than lysimeter measured grass in Davis. Rice ET was measured by Bowen ratio about 25 miles north of Davis. The postulated that the ETo would be less in the rice growing region because of higher humidity. As a result, they recommend a Kc = 1.20 to 1.25. However, CIMIS data indicates that the ETo is only about 5% higher in Davis than at the Nicolas site and in Colusa, which are near the rice growing region. As a result, we would recommend a Kc = 1.05 x 1.05 = 1.10 to estimate ETrice from ETo estimated at a CIMIS station in the rice growing region..

1: Prepare lookup table

We multiply coefficients by 100 to get integer values between 0 and 100. This will allow us to reclassify the landcover rasters and retain the 8-bit grayscale color depth. If we were to crate rasters with decimal values, the values would be stored as 4-byte floating point and we'd end up with a huge 32-bit image that would be very large on disk, large in memory, and would process at a glacial pace. Of course, it means sacrificing a bit of precision (which we will quantify later).

```
CDL.Kc.LUT <- read_csv("input/TABLES/cup+_kc_cdl.csv")

CDL.Kc.LUT[["start.date.1"]] <- as.Date(paste(CDL.Kc.LUT[["Planting Month"]],
    CDL.Kc.LUT[["Planting Day"]], 2001, sep = "-"), format = "%m-%d-%Y")
CDL.Kc.LUT[["end.date.1"]] <- as.Date(paste(CDL.Kc.LUT[["Harvest Month"]],
    CDL.Kc.LUT[["Harvest Day"]], ifelse(CDL.Kc.LUT[["Harvest Month"]] <
        CDL.Kc.LUT[["Planting Month"]], 2002, 2001), sep = "-"),
    format = "%m-%d-%Y")
CDL.Kc.LUT[["start.date.2"]] <- as.Date(paste(CDL.Kc.LUT[["2nd Planting Month"]],
    CDL.Kc.LUT[["2nd Planting Day"]], 2001, sep = "-"), format = "%m-%d-%Y")
CDL.Kc.LUT[["end.date.2"]] <- as.Date(paste(CDL.Kc.LUT[["2nd Harvest Month"]],
    CDL.Kc.LUT[["2nd Harvest Day"]], ifelse(CDL.Kc.LUT[["2nd Harvest Month"]] <
        CDL.Kc.LUT[["2nd Planting Month"]], 2002, 2001), sep = "-"),
    format = "%m-%d-%Y")

CDL.Kc.LUT[["is.doublecrop"]] <- !is.na(CDL.Kc.LUT[["2nd Planting Day"]])

# TEST: Remove unrepresented crops and print TODO: Move to
# unit tests
warning("The following landcover classes do not have matching coefficients, and will be removed: ",
    paste(unique(unlist(CDL.Kc.LUT[!complete.cases(CDL.Kc.LUT[["Kc AB"]]),
        "cdl_name"])), " "))
CDL.Kc.LUT <- CDL.Kc.LUT[complete.cases(CDL.Kc.LUT[["Kc AB"]]),
    ]

# Scale coeff by 100
CDL.Kc.LUT[, c("Kc AB", "Kc CD", "Kc E", "2nd Kc AB", "2nd Kc CD",
    "2nd Kc E")] <- CDL.Kc.LUT[, c("Kc AB", "Kc CD", "Kc E",
    "2nd Kc AB", "2nd Kc CD", "2nd Kc E")] * 100

# Calculate length of growing period
CDL.Kc.LUT[["lgp.1"]] <- CDL.Kc.LUT$end.date.1 - CDL.Kc.LUT$start.date.1
CDL.Kc.LUT[["lgp.2"]] <- CDL.Kc.LUT$end.date.2 - CDL.Kc.LUT$start.date.2

# Calculate dates of Kc inflection points TODO: Clean up with
# a function; DRY
CDL.Kc.LUT <- cbind(CDL.Kc.LUT, t_a.1 = CDL.Kc.LUT[["start.date.1"]],
    t_b.1 = CDL.Kc.LUT[["start.date.1"]] + (CDL.Kc.LUT[["lgp.1"]] *
        CDL.Kc.LUT[["% season B"]] * 0.01), t_c.1 = CDL.Kc.LUT[["start.date.1"]] +
        (CDL.Kc.LUT[["lgp.1"]] * CDL.Kc.LUT[["% season C"]] *
            0.01), t_d.1 = CDL.Kc.LUT[["start.date.1"]] + (CDL.Kc.LUT[["lgp.1"]] *
        CDL.Kc.LUT[["% season D"]] * 0.01), t_e.1 = CDL.Kc.LUT[["end.date.1"]],
    t_a.2 = CDL.Kc.LUT[["start.date.2"]], t_b.2 = CDL.Kc.LUT[["start.date.2"]] +
        (CDL.Kc.LUT[["lgp.2"]] * CDL.Kc.LUT[["% season B"]] *
            0.01), t_c.2 = CDL.Kc.LUT[["start.date.2"]] + (CDL.Kc.LUT[["lgp.2"]] *
        CDL.Kc.LUT[["% season C"]] * 0.01), t_d.2 = CDL.Kc.LUT[["start.date.2"]] +
```

```
        (CDL.Kc.LUT[["lgp.2"]] * CDL.Kc.LUT[["% season D"]] *
            0.01), t_e.2 = CDL.Kc.LUT[["end.date.2"]])
```

Next we create the actual look-up table that contains daily Kc values, computed according to the heuristic from CUP+. We're using if-else tests for clarity, but the following could be vectorized for a marginal speedup.

TODO: Vectorize with logical array (http://kitchingroup.cheme.cmu.edu/blog/2013/02/23/Vectorized-piecewise-functions/)

```
# Plot and inspect plot(seq.Date(as.Date('2001-01-01'), by =
# 'day', length.out = 730),
# calculateKcDaily(seq.Date(as.Date('2001-01-01'), by =
# 'day', length.out = 730), as.data.frame(CDL.Kc.LUT)[4,]))

# Create daily Kc values for 2 years.
CDL.Kc.LUT.daily <- do.call("rbind", by(CDL.Kc.LUT, 1:nrow(CDL.Kc.LUT),
    function(row) rbind(calculateKcDaily(seq.Date(as.Date("2001-01-01"),
        by = "day", length.out = 730), row))))

# TEST: See if we can wrap the coefficients of the second
# year into the beginning of the first without overlap DRY:
# This test should occur when CDL.Kc.LUT is imported.  TODO:
# Move to unit tests
if (any(c(by(CDL.Kc.LUT.daily, 1:nrow(CDL.Kc.LUT.daily), function(row) any(row[1:365] !=
    0 & row[366:730] != 0))))) warning("oops, you STILL may have double crops with overlapping growing
    paste(which(c(by(CDL.Kc.LUT.daily, 1:nrow(CDL.Kc.LUT.daily),
        function(row) any(row[1:365] != 0 & row[366:730] != 0)))),
        collapse = " "))

# Wrap into single year and label
CDL.Kc.LUT.daily <- pmax(CDL.Kc.LUT.daily[, 1:365], CDL.Kc.LUT.daily[,
    366:730])
CDL.Kc.LUT.daily <- CDL.Kc.LUT.daily[, -(366:730)]
colnames(CDL.Kc.LUT.daily) <- 1:365
CDL.Kc.LUT.daily <- cbind(value = CDL.Kc.LUT[["VALUE"]], cdl_name = CDL.Kc.LUT[["cdl_name"]],
    as.data.frame(CDL.Kc.LUT.daily))

# TODO: Evaluate if it's more readable to set NA earlier in
# logic
CDL.Kc.LUT.daily[CDL.Kc.LUT.daily == 0] <- NA

# Plot and inspect par(mfrow = c(7, 7)) for (row in
# 1:nrow(CDL.Kc.LUT.daily)){
# plot(1:365,CDL.Kc.LUT.daily[row,3:367]) }
```

This final structure is not "tidy", in the sense of variables forming columns and observations forming rows. It's a 3-dimensional lookup table project into two.

```
saveRDS(CDL.Kc.LUT.daily, "output/tables/CDL_Kc_LUT_daily.rds")
write.csv(CDL.Kc.LUT.daily, "output/tables/CDL_Kc_LUT_daily.csv")
```

2. Create special LUT for dual-cropped regions for later disentanglement

```
# Subset to only dual-cropped regions
CDL.LUT.dual <- CDL.Kc.LUT[CDL.Kc.LUT[["is.doublecrop"]] == TRUE,
    ]
# Create daily Kc values for 2 years.
CDL.LUT.dual <- do.call("rbind", by(CDL.LUT.dual, 1:nrow(CDL.LUT.dual),
    function(row) rbind(tagKcDaily(seq.Date(as.Date("2001-01-01"),
        by = "day", length.out = 730), row))))

CDL.LUT.dual <- pmax(CDL.LUT.dual[, 1:365], CDL.LUT.dual[, 366:730])
CDL.LUT.dual <- CDL.LUT.dual[, -(366:730)]
colnames(CDL.LUT.dual) <- 1:365
CDL.LUT.dual <- cbind(value = CDL.Kc.LUT[CDL.Kc.LUT[["is.doublecrop"]] ==
    TRUE, ][["VALUE"]], cdl_name = CDL.Kc.LUT[CDL.Kc.LUT[["is.doublecrop"]] ==
    TRUE, ][["cdl_name"]], as.data.frame(CDL.LUT.dual))
CDL.LUT.dual[CDL.LUT.dual == 0] <- NA

saveRDS(CDL.LUT.dual, "output/tables/CDL_LUT_dualtagged.rds")
write.csv(CDL.LUT.dual, "output/tables/CDL_LUT_dualtagged.csv")
```

3.Lc_prep

There are different export options in the USDA, NASS, CropScape and Cropland Data Layers tool. Users can export subsets of the national data set under different regional masks, such as state boundaries, however all inherit the default `WGS 84 / Lon Lat` CRS, reproduced below as OGC WKT:

```
GEOGCS["WGS 84",
    DATUM["WGS_1984",
        SPHEROID["WGS 84",6378137,298.257223563,
            AUTHORITY["EPSG","7030"]],
        AUTHORITY["EPSG","6326"]],
    PRIMEM["Greenwich",0,
        AUTHORITY["EPSG","8901"]],
    UNIT["degree",0.0174532925199433,
        AUTHORITY["EPSG","9122"]],
    AUTHORITY["EPSG","4326"]]
```

```r
# ca_boundary defines the boundary for the area/(s) of
# analysis (includes counties) ca_boundary <-
# readOGR('input/MAPS/cb_2014_CA_5m.shp',
# layer='cb_2014_CA_5m')
ca_boundary.path <- "Input/SHAPEFILES/cnty24k09/cnty24k09_state_poly_s100.shp"
lc.dir <- "input/CDL_CA_WGS84"
lc.crs <- CRS("+init=epsg:4326")
```

1. Collect landcover rasters and inspect

First we collect all of the landcover rasters in our input directory and organize them into a table with relevant metadata, e.g. resolution and extent. Rasters are tagged as **anomalous** if their extent or resolution is different from the most common values of the group (criteria `mode`).

```r
# TODO: Generalize as function
lc.paths <- list.files(path = lc.dir, pattern = ".(tif)$", full.names = T,
    recursive = TRUE)

lc.table <- data.frame(abs_path = lc.paths, source = sapply(strsplit(file_path_sans_ext(lc.paths),
    "/"), "[[", 2), product_name = sapply(strsplit(file_path_sans_ext(lc.paths),
    "/"), "[[", 4), date = as.Date(sprintf("%s-01-01", regmatches(lc.paths,
    regexpr("(?<=CA_WGS84\\/CDL_)(.{4})(?=_clip)", lc.paths,
        perl = TRUE))), format = "%F"), stringsAsFactors = FALSE)

# HACK: Clean up ugly product name

lc.table[["product_name"]] <- paste0("CDL_", 2007:2016)

# Tag (x,y)[min,max,res], ncell, and nlayers if applicable
lc.table <- rtTagParam(lc.table)
# Tag anomalous parameters according to metadata parameters
# specified
lc.table["anoml"] <- tagAnomalous(lc.table, c("xmin", "xmax",
    "ymin", "ymax", "ncell", "xres", "yres"), modal)
rm(lc.paths)
```

2. Clean anomalous rasters and re-inspect

2.1 Upscale coarse landcover rasters

At this point, we visually inspect anom.indicies and perform manual cleaning depending on what the problem is. For CDL, the resolution of older products is 56, so we must upscale to match the later years. **NOTE:** Here, the criterion for cleaning is `anoml == TRUE`, since there are no other anomalous features, but this probably won't be the same for different data sources!

`resample_cdl` is a function that resamples each raster that you feed into it to match the dimensions of *the last raster in the landcover file table.* It can be run sequentially on one thread (commented out below), or in parallel on multiple threads using `parallel`.

```
# anom.table = makeAnomtable(lc.table)

# Upscale with nearest neighbor interpolation (to preserve
# categorical variable) to match resolution of most recent
# CDL (56 meters)
resample_cdl <- function(lc.table, abs_path, source, product_name) {
    dir.create(paste0("output/cleaned_inputs/", source, "_30m/"),
        recursive = TRUE, showWarnings = FALSE)
    outpath <- paste0("output/cleaned_inputs/", source, "_30m/",
        product_name, ".tif")
    resample(raster(abs_path), raster(lc.table[["abs_path"]][nrow(lc.table)]),
        method = "ngb", filename = outpath, format = "GTiff",
        prj = TRUE, progress = "text", datatype = "INT1U", overwrite = TRUE)
}

# TODO: Figure out why apply returns extra row, creating NA
# directory Non parallel version
# apply(lc.table[lc.table['xres'] == 56,], 1, function(x)
# resample_cdl(lc.table, x['abs_path'], x['source'],
# x['product_name']))

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(min((detectCores() - 1), sum(lc.table["anoml"] ==
    TRUE)))
clusterExport(cl, list("lc.table", "resample_cdl"))
clusterEvalQ(cl, library(raster))
parRapply(cl, lc.table[lc.table["anoml"] == TRUE, ], function(x) resample_cdl(lc.table,
    x["abs_path"], x["source"], x["product_name"]))
stopCluster(cl)
```

```
project_cdl <- function(lc.table, abs_path, source, product_name) {
    dir.create(paste0("output/cleaned_inputs/", source, "_30m/"),
        recursive = TRUE, showWarnings = FALSE)
    outpath <- paste0("output/cleaned_inputs/", source, "_30m/",
        product_name, ".tif")
    projectRaster(raster(abs_path), crs = CRS("+init=epsg:4326"),
        method = "ngb", filename = outpath, format = "GTiff",
        prj = TRUE, progress = "text", datatype = "INT1U", overwrite = TRUE)
}

cl <- makeCluster(min((detectCores() - 1), nrow(lc.table)), outfile = "debug.txt")
```

```
clusterExport(cl, list("lc.table", "project_cdl"))
clusterEvalQ(cl, library(raster))
parRapply(cl, lc.table, function(x) project_cdl(lc.table, x["abs_path"],
    x["source"], x["product_name"]))
stopCluster(cl)
```

2.2 Import and re-inspect upscaled rasters

```
# Load cleaned files into file table TODO: DRY
lc.paths.cleaned <- list.files(path = "output/cleaned_inputs/CDL_CA_30m",
    pattern = ".(tif)$", full.names = T, recursive = TRUE)

# TODO: remove hardcoded gsub
lc.table.cleaned <- data.frame(abs_path = lc.paths.cleaned, source = gsub(".{4}$",
    "", sapply(strsplit(file_path_sans_ext(lc.paths.cleaned),
        "/"), "[[", 3)), product_name = sapply(strsplit(file_path_sans_ext(lc.paths.cleaned),
    "/"), "[[", 4), date = as.Date(sprintf("%s-01-01", regmatches(lc.paths.cleaned,
    regexpr("(?<=CA_WGS84_30m\\/CDL_)(.*)(?=.tif)", lc.paths.cleaned,
        perl = TRUE))), format = "%F"), stringsAsFactors = FALSE)

# Merge cleaned into file table and redo checks TODO: DRY
# TODO: Add test to ensure that all of 'anoml' column ==
# FALSE
lc.table[lc.table[["date"]] %in% lc.table.cleaned[["date"]],
    ] <- lc.table.cleaned[lc.table.cleaned[["date"]] %in% lc.table[["date"]],
    ]

# Tag (x,y)[min,max,res], ncell, and nlayers if applicable
lc.table <- rtTagParam(lc.table)
# Tag anomalous parameters according to metadata parameters
# specified
lc.table["anoml"] <- tagAnomalous(lc.table, c("xmin", "xmax",
    "ymin", "ymax", "ncell", "xres", "yres"), modal)

rm(lc.paths.cleaned, lc.table.cleaned)
```

2.3 Warp Lc

We use gdalwarp to allign all of our datasets, with the same extent and same number of rows/columns.

```
warp_lc <- function(abs_path, cut.shapefile, source, date) {
    dir.create(paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date)), recursive = TRUE,
        showWarnings = FALSE)
    outpath <- paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date), "/", format(date,
            format = "%Y_%m_%d"), ".tif")
    gdalwarp_wrapper("bin/gdal/apps/gdalwarp.exe", "-s_srs EPSG:4326 -t_srs EPSG:3310 -tr 30 30 -r near
        cut.shapefile, abs_path, outpath)
}
```

```r
# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(1, outfile = "debug.txt")
clusterExport(cl, list("lc.table", "warp_lc", "ca_boundary.path",
    "gdalwarp_wrapper"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, lc.table, function(x) warp_lc(x[["abs_path"]],
    ca_boundary.path, x[["source"]], x[["date"]]))
stopCluster(cl)
```

4.eto_prep.Rmd (CWU Calculations)

```
# ca_boundary defines the boundary for the area/(s) of
# analysis (includes counties) ca_boundary <-
# readOGR('Input/SHAPEFILES/cb_2014_CA_5m.shp',
# layer='cb_2014_CA_5m')
ca_boundary.path <- "Input/SHAPEFILES/cnty24k09/cnty24k09_state_poly_s100.shp"
eto.dir <- "input/SPATIALCIMIS"
eto.crs <- CRS("+init=epsg:3310")
```

To perform operations on collections of daily ETo layers, we can stack each raster file

```
eto.paths <- list.files(path = eto.dir, pattern = ".(asc)$",
    full.names = T, recursive = TRUE)

eto.table <- data.frame(abs_path = eto.paths, source = sapply(strsplit(file_path_sans_ext(eto.paths),
    "/"), "[[", 2), product_name = sapply(strsplit(file_path_sans_ext(eto.paths),
    "/"), "[[", 6), date = as.Date(regmatches(eto.paths, regexpr("(?<=SPATIALCIMIS)(.*)(?=ETo.asc)",
    eto.paths, perl = TRUE)), format = "/%Y/%m/%d/"), stringsAsFactors = FALSE)

eto.table["w.year"] <- as.Date(waterYearlt(eto.table$date))

eto.table["minval"] <- sapply(eto.table[["abs_path"]], function(x) minValue(setMinMax(raster(x))))
eto.table["maxval"] <- sapply(eto.table[["abs_path"]], function(x) maxValue(setMinMax(raster(x))))

eto.table[c("xres", "yres", "minval", "maxval")] <- sapply(eto.table[["abs_path"]],
    function(x) {
        x <- setMinMax(raster(x))
        return(c(xres(x), yres(x), minValue(x), maxValue(x)))
    })

# Tag (x,y)[min,max,res], ncell, and nlayers if applicable
eto.table <- rtTagParam(eto.table)
# Tag anomalous parameters according to metadata parameters
# specified
eto.table["anoml"] <- tagAnomalous(eto.table, c("xmin", "xmax",
    "ymin", "ymax", "ncell", "xres", "yres"), modal)

rm(eto.paths)
```

```
anom.table <- makeAnomtable(eto.table)
anom.indicies <- unique(which(anom.table == TRUE, arr.ind = TRUE)[,
    1])
anom.table <- cbind(eto.table[anom.indicies, ], anom.table[anom.indicies,
    ])

# At this point, we visually inspect anom.indicies and
# perform manual cleaning depending on what the problem is

# (x,y)[res] 2012-06-25 is the only one with an anomalous
# resolution (500m vs 2000m)'

# (x,y)[min,max] The rest appear to be missing a couple of
```

```r
# rows or columns worth of data

# Limit our analysis to 2007-2016
eto.table <- eto.table[!(year(eto.table[["date"]]) < 2007), ]
seq(ymd("2007-01-01"), ymd("2016-12-31"), by = "day")
missing.date.index <- !(seq(ymd("2007-01-01"), ymd("2016-12-31"),
    by = "day") %in% eto.table[["date"]])
missing.dates <- seq(ymd("2007-01-01"), ymd("2016-12-31"), by = "day")[missing.date.index]
warning(paste0("Missing dates: "), missing.dates, "will be set to the last complete observation.")

filled.obs <- as.data.frame(lapply(eto.table[eto.table[["date"]] ==
    missing.dates[1] - 1, ], rep, length(missing.dates)))
filled.obs[["date"]] <- missing.dates
eto.table <- rbind(eto.table, filled.obs)
rm(filled.obs, missing.dates, missing.date.index)
```

As there is nothing major amiss, the resolution and minor extent issues can be resolved by `resample`-ing, which is required for upscaling anyway.

Round ETo

This is a very quick step. Completes in under 10 minutes on 22 threads.

```r
scale_eto <- function(abs_path, eto.crs, source, date) {
    dir.create(paste0("output/cleaned_inputs/", source, "_scaled/",
        year(date)), recursive = TRUE, showWarnings = FALSE)
    outpath <- paste0("output/cleaned_inputs/", source, "_scaled/",
        year(date), "/", format(date, format = "%Y_%m_%d"), ".tif")
    calc(raster(abs_path, crs = eto.crs), fun = function(x) {
        round(x * 100)
    }, format = "GTiff", progress = "text", datatype = "INT2U",
        overwrite = TRUE, filename = outpath)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(min((detectCores() - 2), nrow(eto.table)),
    outfile = "debug.txt")
clusterExport(cl, list("eto.table", "scale_eto"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, eto.table, function(x) scale_eto(x[["abs_path"]],
    CRS("+init=epsg:3310"), x[["source"]], x[["date"]]))
stopCluster(cl)
```

2.2 Import and re-inspect rounded rasters

```r
# Load cleaned files into file table TODO: DRY
eto.paths.scaled <- list.files(path = "output/cleaned_inputs/SPATIALCIMIS_scaled",
```

```
    pattern = ".(tif)$", full.names = T, recursive = TRUE)

# TODO: remove hardcoded gsub
eto.table.scaled <- data.frame(abs_path = eto.paths.scaled, source = sapply(strsplit(file_path_sans_ext
    "/"), "[[", 3), product_name = "ETo", date = as.Date(sapply(strsplit(file_path_sans_ext(eto.paths.s
    "/"), "[[", 5), format = "%Y-%m-%d"), stringsAsFactors = FALSE)

rm(eto.paths.scaled, eto.table)
```

Warp ETo

```
warp_eto <- function(abs_path, cut.shapefile, source, date) {
    dir.create(paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date)), recursive = TRUE,
        showWarnings = FALSE)
    outpath <- paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date), "/", format(date,
            format = "%Y_%m_%d"), ".tif")
    gdalwarp_wrapper("bin/gdal/apps/gdalwarp.exe", "-s_srs EPSG:3310 -t_srs EPSG:3310 -tr 30 30 -r bili
        cut.shapefile, abs_path, outpath)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(20, outfile = "debug.txt")
clusterExport(cl, list("eto.table.scaled", "warp_eto", "ca_boundary.path",
    "gdalwarp_wrapper"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, eto.table.scaled, function(x) warp_eto(x[["abs_path"]],
    ca_boundary.path, x[["source"]], x[["date"]]))
stopCluster(cl)
```

Performance note: Started at 16:11 finished at 23:01. 7 hrs on 20 threads, smt

5_cwu_calcs.Rmd (CWU calulations)

```
eto.dir <- "output/cleaned_inputs/SPATIALCIMIS_scaled_projected"
lc.dir <- "output/cleaned_inputs/CDL_CA_projected"
kc.lut <- readRDS("output/tables/CDL_Kc_LUT_daily.rds")
# TODO: Remove leap year hack (`366` = kc.lut[['365']])
kc.lut <- cbind(index = 1:nrow(kc.lut), kc.lut, `366` = kc.lut[["365"]])
global.crs <- CRS("+init=epsg:3310")
```

Note: We have already alligned and reprojected all data sets in previous steps.

1.1 Import rounded eto

```
# Load cleaned files into file table TODO: DRY
eto.paths <- list.files(path = eto.dir, pattern = ".(tif)$",
    full.names = T, recursive = TRUE)

# TODO: remove hardcoded gsub
eto.table <- data.frame(abs_path = eto.paths, source = sapply(strsplit(file_path_sans_ext(eto.paths),
    "/"), "[[", 3), product_name = "ETo", date = as.Date(sapply(strsplit(file_path_sans_ext(eto.paths),
    "/"), "[[", 5), format = "%Y-%m-%d"), stringsAsFactors = FALSE)

rm(eto.paths, eto.dir)
```

1.2 Import rounded landcover

```
# Load cleaned files into file table TODO: DRY
lc.paths <- list.files(path = lc.dir, pattern = ".(tif)$", full.names = T,
    recursive = TRUE)

lc.table <- data.frame(abs_path = lc.paths, source = sapply(strsplit(file_path_sans_ext(lc.paths),
    "/"), "[[", 3), product_name = "CDL", date = as.Date(sprintf("%s-01-01",
    sapply(strsplit(file_path_sans_ext(lc.paths), "/"), "[[",
        4)), format = "%F"), stringsAsFactors = FALSE)

rm(lc.paths, lc.dir)
```

1.3 Append data to eto table

```
eto.table[["lc_path"]] <- sapply(year(eto.table[["date"]]), function(x) lc.table[year(lc.table[["date"]]
    x, ][["abs_path"]])
```

2 Perform CWR/ETc calculation day-wise

TODO: Verify that `global.crs` is not needed

```
make_etc <- function(abs_path, lc_path, global.crs, date) {
    dir.create(paste0("output/cwr_calcs/", year(date)), recursive = TRUE,
        showWarnings = FALSE)
    outpath <- paste0("output/cwr_calcs/", year(date), "/", format(date,
        format = "%Y_%m_%d"), ".tif")
    day.of.year <- yday(date)
    overlay(raster(abs_path), raster(lc_path), fun = function(eto.value,
        lc.value) {
        return(kc.lut[match(lc.value, kc.lut[["value"]]), as.character(day.of.year)] *
            eto.value)
    }, format = "GTiff", progress = "text", datatype = "INT4U",
        overwrite = TRUE, filename = outpath, forcefun = TRUE)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(min((detectCores() - 2), nrow(eto.table)),
    outfile = "debug.txt")
clusterExport(cl, list("eto.table", "make_etc", "kc.lut"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, eto.table, function(x) make_etc(x[["abs_path"]],
    x[["lc_path"]], CRS("+init=epsg:4326"), x[["date"]]))
stopCluster(cl)
```

Performance notes:

duration 32 hrs w/ 20 threads, smt 34h:27m

6.ppt_prep.Rmd (CWU calulations)

```
# ca_boundary defines the boundary for the area/(s) of
# analysis (includes counties)
ca_boundary.path <- "input/SHAPEFILES/cnty24k09/cnty24k09_state_poly_s100.shp"
ppt.dir <- "input/PRISM"
ppt.crs <- CRS("+init=epsg:4269")
```

To perform operations on collections of daily ppt layers, we can stack each raster file

```
ppt.paths <- list.files(path = ppt.dir, pattern = ".(tif)$",
    full.names = T, recursive = TRUE)

ppt.table <- data.frame(abs_path = ppt.paths, source = sapply(strsplit(file_path_sans_ext(ppt.paths),
    "/"), "[[", 2), product_name = "ppt", date = as.Date(regmatches(ppt.paths,
    regexpr("(?<=prism_ppt_us_30s_)(.*)(?=.tif)", ppt.paths,
        perl = TRUE)), format = "%Y%m%d"), stringsAsFactors = FALSE)

# ppt.table['w.year'] <- as.Date(waterYearlt(ppt.table$date))

# ppt.table['minval'] <- sapply(ppt.table[['abs_path']],
# function(x) minValue(setMinMax(raster(x))))
# ppt.table['maxval'] <- sapply(ppt.table[['abs_path']],
# function(x) maxValue(setMinMax(raster(x))))

# ppt.table[c('xres', 'yres', 'minval', 'maxval')] <-
# sapply(ppt.table[['abs_path']], function(x) { x <-
# setMinMax(raster(x))
# return(c(xres(x),yres(x),minValue(x),maxValue(x))) })

# Tag (x,y)[min,max,res], ncell, and nlayers if applicable
# ppt.table <- rtTagParam(ppt.table) Tag anomalous parameters
# according to metadata parameters specified
# ppt.table['anoml'] <- tagAnomalous(ppt.table, c('xmin',
# 'xmax', 'ymin', 'ymax', 'ncell', 'xres', 'yres'), modal)

rm(ppt.paths)
```

At least for `PRISM`, the data are clean.

```
any(ppt.table$anoml)
# Limit our analysis to 2007-2016
ppt.table <- ppt.table[!(year(ppt.table[["date"]]) < 2007), ]

missing.date.index <- !(seq(ymd("2007-01-01"), ymd("2015-12-31"),
    by = "day") %in% ppt.table[["date"]])
any(missing.date.index)
rm(filled.obs, missing.dates, missing.date.index)
```

As there is nothing major amiss, the resolution and minor extent issues can be resolved by `resample`-ing, which is required for upscaling anyway.

Round ppt

This is a very quick step. Completes in under 6 minutes on 20 threads.

```
scale_ppt <- function(abs_path, ppt.crs, source, date) {
    dir.create(paste0("output/cleaned_inputs/", source, "_scaled/",
        year(date)), recursive = TRUE, showWarnings = FALSE)
    outpath <- paste0("output/cleaned_inputs/", source, "_scaled/",
        year(date), "/", format(date, format = "%Y_%m_%d"), ".tif")
    calc(raster(abs_path, crs = ppt.crs), fun = function(x) {
        round(x * 100)
    }, format = "GTiff", progress = "text", datatype = "INT2U",
        overwrite = TRUE, filename = outpath)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(min((detectCores() - 5), nrow(ppt.table)),
    outfile = "debug.txt")
clusterExport(cl, list("ppt.table", "scale_ppt"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, ppt.table, function(x) scale_ppt(x[["abs_path"]],
    CRS("+init=epsg:4269"), x[["source"]], x[["date"]]))
stopCluster(cl)
```

2.2 Import and re-inspect rounded rasters

```
# Load cleaned files into file table TODO: DRY
ppt.paths.scaled <- list.files(path = "output/cleaned_inputs/PRISM_scaled",
    pattern = ".(tif)$", full.names = T, recursive = TRUE)

# TODO: remove hardcoded gsub
ppt.table.scaled <- data.frame(abs_path = ppt.paths.scaled, source = sapply(strsplit(file_path_sans_ext
    "/"), "[[", 3), product_name = "ppt", date = as.Date(sapply(strsplit(file_path_sans_ext(ppt.paths.s
    "/"), "[[", 5), format = "%Y-%m-%d"), stringsAsFactors = FALSE)

rm(ppt.paths.scaled, ppt.table)
```

Warp ppt

Warp took 6 hrs on 19 threads (12 hyperthreaded cores) Started: 23:35 Ended: 05:17

```
warp_ppt <- function(abs_path, cut.shapefile, source, date) {
    dir.create(paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date)), recursive = TRUE,
        showWarnings = FALSE)
    outpath <- paste0("E:/Users/lbooth/Documents/wfar/output/cleaned_inputs/",
        source, "_projected/", year(date), "/", format(date,
```

```
            format = "%Y_%m_%d"), ".tif")
    gdalwarp_wrapper("bin/gdal/apps/gdalwarp.exe", "-s_srs EPSG:4269 -t_srs EPSG:3310 -tr 30 30 -r bili
        cut.shapefile, abs_path, outpath)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(20, outfile = "debug.txt")
clusterExport(cl, list("ppt.table.scaled", "warp_ppt", "ca_boundary.path",
    "gdalwarp_wrapper"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, ppt.table.scaled, function(x) warp_ppt(x[["abs_path"]],
    ca_boundary.path, x[["source"]], x[["date"]]))
stopCluster(cl)
```

7.zone_sums (County Aggregations)

```r
# ca_boundary defines the boundary for the area/(s) of
# analysis (includes counties) ca_boundary <-
# readOGR('input/SHAPEFILES/cb_2014_CA_5m.shp',
# layer='cb_2014_CA_5m')
ca.counties.path <- "output/ca_counties.tif"
ca.counties <- readOGR("input/SHAPEFILES/cnty24k09_poly/cnty24k09_poly_s100.shp")
cwr.dir <- "output/cwr_calcs"
ppt.dir <- "output/cleaned_inputs/PRISM_scaled_projected"
lc.dir <- "output/cleaned_inputs/CDL_CA_projected"
index.dir <- "output/intermediaries/county_lc_index/"
```

```r
cwr.paths <- list.files(path = cwr.dir, pattern = ".(tif)$",
    full.names = T, recursive = TRUE)

cwr.table <- data.frame(abs_path = cwr.paths, source = sapply(strsplit(file_path_sans_ext(cwr.paths),
    "/"), "[[", 2), product_name = "cwr", date = as.Date(sapply(strsplit(file_path_sans_ext(cwr.paths),
    "/"), "[[", 4), format = "%Y-%m-%d"), stringsAsFactors = FALSE)

cwr.table["w.year"] <- as.Date(waterYearlt(cwr.table$date))

ppt.paths <- list.files(path = ppt.dir, pattern = ".(tif)$",
    full.names = T, recursive = TRUE)

ppt.table <- data.frame(abs_path = ppt.paths, source = sapply(strsplit(file_path_sans_ext(ppt.paths),
    "/"), "[[", 3), product_name = "ppt", date = as.Date(sapply(strsplit(file_path_sans_ext(ppt.paths),
    "/"), "[[", 5), format = "%Y-%m-%d"), stringsAsFactors = FALSE)

ppt.table["w.year"] <- as.Date(waterYearlt(ppt.table$date))

lc.paths <- list.files(path = lc.dir, pattern = ".(tif)$", full.names = T,
    recursive = TRUE)

lc.table <- data.frame(abs_path = lc.paths, source = sapply(strsplit(file_path_sans_ext(lc.paths),
    "/"), "[[", 3), product_name = "CDL", date = as.Date(sprintf("%s-01-01",
    sapply(strsplit(file_path_sans_ext(lc.paths), "/"), "[[",
        4)), format = "%F"), stringsAsFactors = FALSE)

index.paths <- list.files(path = index.dir, pattern = ".(tif)$",
    full.names = T, recursive = TRUE)

index.table <- data.frame(abs_path = index.paths, date = as.Date(sprintf("%s-01-01",
    sapply(strsplit(file_path_sans_ext(index.paths), "/"), "[[",
        4)), format = "%F"), stringsAsFactors = FALSE)

## TODO: Add input validation logic
rm(ppt.dir, cwr.dir, ppt.paths, cwr.paths, lc.paths, lc.dir,
    index.dir, index.paths)
```

Tangle aggregation units (2 units)

```
# raster(raster()) uses the first landcover raster as a
# template, discarding values 22858 seconds
ca.counties.r <- rasterize(ca.counties, raster(raster(lc.table[1,
    1])), "NUM", fun = "last", filename = "output/ca_counties.tif")
```

```
tangle_lc_boundaries <- function(lc_path, region_path, date) {
    dir.create("output/intermediaries/county_lc_index", recursive = TRUE,
        showWarnings = FALSE)
    outpath <- paste0("output/intermediaries/county_lc_index/",
        year(date), ".tif")
    overlay(raster(lc_path), raster(region_path), fun = function(x,
        y) {
        return(szudzik_pair(x, y))
    }, format = "GTiff", progress = "text", datatype = "INT4U",
        overwrite = TRUE, filename = outpath, forcefun = TRUE)
}

# Make only as many clusters as necessary, bound by available
# cores
cl <- makeCluster(min((detectCores() - 2), nrow(lc.table)), outfile = "debug.txt")
clusterExport(cl, list("lc.table", "tangle_lc_boundaries", "ca.counties.path",
    "szudzik_pair"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
parRapply(cl, lc.table, function(x) tangle_lc_boundaries(x[["abs_path"]],
    ca.counties.path, x[["date"]]))
stopCluster(cl)
# 529 seconds
```

```
make_zone_sum <- function(abs_paths, index_path) {
    z.table = zonal(raster(abs_paths), raster(index_path), "sum",
        na.rm = TRUE, progress = "text")
    return(z.table)
}

product_name = "cwr"

# Make only as many clusters as necessary, bound by available
# cores
exe.starttime <- Sys.time()
cl <- makeCluster(min((detectCores() - 2), nrow(cwr.table)),
    outfile = "debug.txt")
clusterExport(cl, list("cwr.table", "index.table", "make_zone_sum",
    "product_name"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))
```

```
for (year.index in unique(year(cwr.table[["date"]]))) {

    clusterExport(cl, "year.index")
    z.table <- clusterMap(cl, make_zone_sum, abs_paths = cwr.table[year(cwr.table[["date"]]) ==
        year.index, ][["abs_path"]], MoreArgs = list(index_path = index.table[year(index.table[["date"]]
        year.index, ][["abs_path"]]), RECYCLE = TRUE, SIMPLIFY = FALSE)

    dir.create("output/summaries/", recursive = TRUE, showWarnings = FALSE)
    saveRDS(z.table, paste0("output/summaries/", product_name,
        "_", year.index, ".rds"))

    z.table <- Reduce(function(dtf1, dtf2) cbind(dtf1, dtf2[,
        2]), z.table)
    z.table <- cbind(szudzik_unpair(z.table[, 1]), z.table[,
        -1])
    colnames(z.table) <- c("crop", "county", seq(1, length(colnames(z.table)) -
        2))
    write.csv(z.table, file = paste0("output/summaries/", product_name,
        "_", year.index, ".csv"))
}
stopCluster(cl)
exe.stoptime <- Sys.time()
print(paste("Began calculation at", exe.starttime, "and completed at",
    exe.stoptime))
print(exe.stoptime - exe.starttime)
```

Performance notes: 2 days 12 hrs on 20 threads, smt 8.5 days on 6 threads, smt

```
make_zone_sum <- function(abs_paths, index_path) {
    z.table = zonal(raster(abs_paths), raster(index_path), "sum",
        na.rm = TRUE, progress = "text")
    return(z.table)
}

product_name = "ppt"

# Make only as many clusters as necessary, bound by available
# cores
exe.starttime <- Sys.time()
cl <- makeCluster(min((detectCores() - 2), nrow(ppt.table)),
    outfile = "debug.txt")
clusterExport(cl, list("ppt.table", "index.table", "make_zone_sum",
    "product_name"))
clusterEvalQ(cl, {
    library(raster)
    library(lubridate)
})
clusterEvalQ(cl, rasterOptions(progress = "text", time = TRUE))

for (year.index in unique(year(ppt.table[["date"]]))) {

    clusterExport(cl, "year.index")
    z.table <- clusterMap(cl, make_zone_sum, abs_paths = ppt.table[year(ppt.table[["date"]]) ==
        year.index, ][["abs_path"]], MoreArgs = list(index_path = index.table[year(index.table[["date"]]
```

```
        year.index, ][["abs_path"]]), RECYCLE = TRUE, SIMPLIFY = FALSE)

    dir.create("output/summaries/", recursive = TRUE, showWarnings = FALSE)
    saveRDS(z.table, paste0("output/summaries/", product_name,
        "_", year.index, ".rds"))


    z.table <- Reduce(function(dtf1, dtf2) cbind(dtf1, dtf2[,
        2]), z.table)
    z.table <- cbind(szudzik_unpair(z.table[, 1]), z.table[,
        -1])
    colnames(z.table) <- c("crop", "county", seq(1, length(colnames(z.table)) -
        2))
    write.csv(z.table, file = paste0("output/summaries/", product_name,
        "_", year.index, ".csv"))
}
stopCluster(cl)
exe.stoptime <- Sys.time()
print(paste("Began calculation at", exe.starttime, "and completed at",
    exe.stoptime))
print(exe.stoptime - exe.starttime)
```

(2 days, 6 hrs)

8.zone_prep.Rmd (Preparation of zone sums for WF calculation)

This worksheet leans heavily on `tidyverse` packages.

Introduction

In the previous step, we computed zonal sums for crop water demand and precipitation respectively. Each unique zone was created as a pair of a county index and a landcover (i.e. crop) index. This worksheet unpairs the zone index and recovers the crop and county information.

`output/summaries/{cwr,ppt}_year.rds` contains the raw zone sums, stored as large `lists`, with each list representing a year of observations. Each daily observation is stored as a 2-D matrix, with one column representing the zone indicies, and the other column representing the zone sums.

Some landcover indices, represent regions that are dual-cropped. That is, for part of the year, the region contains one crop, and for another part of the year, it contains a different crop. For these regions, we replace the dual-crop landcover index with the appropriate single-crop index. The lookup table for these substitutions is found in `output/tables/CDL_LUT_dualtagged.rds`.

Crop and landcover index pairings are found in `output/tables/CDL_Kc_LUT_daily.rds`. Modeled crop indicies are found in `output/tables/CDL_Kc_LUT_daily.rds`.

```
calc.dir <- "output/summaries"
kc.lut <- readRDS("output/tables/CDL_Kc_LUT_daily.rds")
cdl.table <- read_csv("input/TABLES/cdl_classes_all.csv")
CDL.LUT.dual <- readRDS("output/tables/CDL_LUT_dualtagged.rds")
```

```
calc.paths <- list.files(path = calc.dir, pattern = ".(rds)$",
    full.names = T, recursive = TRUE)

calc.table <- data.frame(abs_path = calc.paths, product_name = sapply(strsplit(basename(file_path_sans_e
    "_"), "[[", 1), date = as.Date(sprintf("%s-01-01", sapply(strsplit(basename(file_path_sans_ext(calc
    "_"), "[[", 2)), format = "%F"), stringsAsFactors = FALSE)

## TODO: Add input validation logic
rm(calc.dir, calc.paths)
```

Split double-cropped categories into respective crops

There are some categories that represent two crop's worth of CWU in a respective year. Setting aside the assumptions that we made for modeling this split CWU, we now must split the year depending on what we expect to be growing in the region on the particular day of year. We make use of `CDL_LUT_dualtagged`, a table that associates a dual-crop category with the expected crop for a given day-of-year.

First, we prepare the look-up-table for the dual-crop categories. Second, we create a function that changes re-assigns a dual cropped zone to the appropriate planted crop depending on the day of year. We call this function in the following block.

```
# SMALL HACK: Resolve for leap years by adding a 366'th day
# that has crop parameters equal to the 365th Instead of
```

```r
# modeling Feb 29, we model Dec 31st twice TODO: We can make
# this more accurate by modeling Feb 28th twice, or adding
# logic to extend the growing season by one day during leap
# years.  Nevertheless, late December and February conditions
# are typically similar.
CDL.LUT.dual <- cbind(CDL.LUT.dual, `366` = CDL.LUT.dual[["365"]])
CDL.LUT.dual <- gather(CDL.LUT.dual, "day_of_year", "real_value",
    3:368, -cdl_name, convert = TRUE)

reassign_dualcrop <- function(crop, date) {
    real_value <- left_join(data.frame(crop = crop, date = yday(date)),
        CDL.LUT.dual, by = c(crop = "value", date = "day_of_year"))[["real_value"]]
    return(ifelse(is.na(real_value), crop, real_value))
}
```

Read in and prep files (`rds` import)

TODO: This function was written before I moved everything north of `gather` (`Reduce` et al.) into the region-aggregation worksheet (`7.county-aggregations.Rmd`), When you re-generate the `.rds` files, remove the `Reduce()` and `szudzik_unpair()` logic from `assemble_zonesums`.

There are a lot of entries to process in this chunk, but it can be done on a laptop. Expect 4M entries to take a minute or two on a modern laptop (sandy bridge or newer, 8GB RAM or more).

```r
assemble_zonesums <- function(calc.table, parameter){
  calc.table <- calc.table[calc.table[["product_name"]] == parameter,]
  master.table <- data.frame()
  for (rownum in 1:nrow(calc.table)){

  master.table <- rbind(master.table,                              # Concat to placeholder dataframe
    readRDS(calc.table[rownum,"abs_path"]) %>%                     # Read in list of daily zone-sums
    map(as.data.frame) %>%                                         # Convert all entries in list from matri
    Reduce(function(x, y) full_join(x, y, by = "zone"), .) %>% # Combine to data frame, each col is a d
    {cbind(szudzik_unpair(.[,1]), .[,-1])} %>%                     # Unpair "zone" into "crop" and "county"
    `colnames<-`(c("crop", "county", seq(1,ncol(.)-2))) %>%        # Name wide dataframe: "crop", "county",
    gather(date, zsum, -crop, -county) %>%                          # Gather into narrow table
    mutate(date = as.Date(as.numeric(date) - 1,
                          origin = paste0(year(calc.table[rownum,"date"]),"-01-01"),
                          format="%Y-%m-%d"))) %>%                 # Change day-of-year to date
    mutate(crop = reassign_dualcrop(crop, date))                   # Unpair dual-crop categories
  }
  return(master.table)
}

# WARNING: Hardcoded paths
cwr.master <- assemble_zonesums(calc.table, "cwr")
saveRDS(cwr.master, "output/cwr_master.rds")

ppt.master <- assemble_zonesums(calc.table, "ppt")
saveRDS(ppt.master, "output/ppt_master.rds")
```

Note that the line that uses `purrr::map()` could be replaced with: `lapply(., as.data.frame)`

`TODO`: There are a few elements of the above logic that can be improved: * Rather than binding the individual daily counts into a wide table, then collapsing into key-value pairs with `gather`, our data are *already* in key-value pairs from the previous step! (key = zone, value = zonesum). Turning it into a wide table is convenient for labeling dates, but if we added, some date metadata from the previous step, then we could use `dplyr:mutate()` to assign a date early on.

Clean zone sums to remove non-crop counts

When we computed zone sums in the last step, we used the raw landcover rasters, which included land cover classifications that we were not interested in (urban, grassland, forest, shrubland, water). We are only interested in the landcover classifications that correspond to the crops that we modeled. There are other crop classifications present in the landcover raster that we did not model, due to a lack of information or otherwise.

It would be useful to: 1. remove the non-crop landcover category zone sums, and 2. identify the crops that had a landcover classification, but were not modeled. In order to do this, we simply compare the unique landcover categories from our zone sum table (below, we're using the `cwr` table) to the crop categories in the crop coefficient lookup table (found in `output/tables/CDL_Kc_LUT_daily.rds`, wherein the crop categories have already set to use the same index as the landcover categories).

`TODO`: This logic should be moved to the kc-prep worksheet. `NOTE`: Some of these tests seem kind of silly, but they've saved me twice already. Tests are good.

```
# List landuse-zones present in our zonesums that are NOT
# present in our crop model table These zones are either
# other non-crop categories, or crops that we did not model
(not.counted <- sort(unique(cwr.master$crop)[!(unique(cwr.master$crop) %in%
    kc.lut$value)]))

# Test that the values that went into `not.counted` are the
# same for the ppt and cwr tables Should be all TRUE, since
# the crop-zone layer was the same for both cwr and ppt
# aggregation
stopifnot(unique(cwr.master$crop)[!(unique(cwr.master$crop) %in%
    kc.lut$value)] == unique(ppt.master$crop)[!(unique(ppt.master$crop) %in%
    kc.lut$value)])

# Test that the values in `not.counted` are in the list of
# landcover indicies NOT present in the kc table This should
# also be TRUE, since the CDL table contains the metadata for
# all of the land-use-zones However, this CDL table also
# contains many NA entries that are simply not present in the
# CDL raster at all TODO: Remove or something, this is
# superfluous and confusing
stopifnot(not.counted %in% unique(cdl.table[["VALUE"]])[!(unique(cdl.table[["VALUE"]]) %in%
    kc.lut$value)])

# Subset cdl.table by not.counted so that we can see the
# description name, and reassgn to not.coun#ted
not.counted <- cdl.table[cdl.table[["VALUE"]] %in% not.counted,
    ]

# Remove these values from both datasets WARNING: Hardcoded
# paths
cwr.master <- cwr.master %>% filter(!(crop %in% not.counted[["VALUE"]])) %>%
    write_rds("output/cwr_master_cleaned.rds", compress = "gz")
```

```r
ppt.master <- ppt.master %>% filter(!(crop %in% not.counted[["VALUE"]])) %>%
    write_rds("output/ppt_master_cleaned.rds", compress = "gz")

# Check that the values were removed.  Now, there should not
# be any zonesums present that are not in our lookup table
# Should be TRUE
stopifnot(is_empty(sort(unique(cwr.master[["crop"]])[!(unique(cwr.master[["crop"]]) %in%
    kc.lut[["value"]])])))
```

Crops present in dataset but not modeled in our analysis are as follows (after removing non-irrigated-crop landcover categories):

```r
(not.counted <- not.counted[not.counted[["VALUE"]] %in% c(5,
    26, 27, 31, 38, 71, 74, 231, 232, 233, 234, 238, 242, 244,
    247, 250), ])
# WARNING: Hardcoded path
write.csv(not.counted, file = "output/excluded_crops.csv")
```

Convert raw values into volumes

Reverse scaling factors

For different reasons, we applied multiplication factors to individual cell values in order to avoid working with floting point values (wikipedia actually has an article on this if you want to learn more https://en.wikipedia.org/wiki/Scale_factor_%28computer_science%29). Now would be a good time to undo those earlier scaling factors before we are unable to (ie. once we start adding and subtracting terms).

We applied scaling factors of 100 to the evapotranspiration and precipitation rasters, AND to the crop coefficients. If z_1 is our cumulative scaling factor for ET:

$$\begin{array}{c}\mathrm{CWR} \approx \mathrm{ETo} \times \mathrm{Kc} \\ \underline{\mathrm{CWR} \times z_1 = (100 \cdot \mathrm{ETo}) \times (100 \cdot \mathrm{Kc})} \\ z_1 = (100 \times 100) = 100^2\end{array}$$

Likewise, our scaling factor for PPT (z_2) is:

$$\begin{array}{c}\underline{\mathrm{ppt} \times z_2 = \mathrm{ppt} \times 100} \\ z = 100\end{array}$$

Note that there is no error here. If you think

$$z$$

for precipitation should be

$$100^2$$

, think about it a bit harder.

`TODO`: I'd like to figure out a more elegant and automated way of keeping track of these scaling factors.

```r
# WARNING: Hardcoded scaling factors
z1 = 100 * 100
z2 = 100

cwr.master <- cwr.master %>% mutate(zsum = zsum/z1)
```

```r
ppt.master <- ppt.master %>% mutate(zsum = zsum/z2)
```

Convert depths into volumes

`{cwr,ppt}.master` both represent the sum of all of the daily depths of precipitation and crop evapotranspiration 'observed' in all of the grid cells of each zone. We can turn this value into an actual volume of water by multiplying the depth of water in each cell by the area of each cell. Since we use a uniform grid for the entire state (!!! see below), each cell has the same area. There are a few ways to think of the following operation:

1. Dimensional analysis: wherein you multiply depths of ET by a cell-conversion-factor

$$x \text{ mm ET} \times \frac{30 * 30 \ (m^2)}{cell} \times \frac{\text{m}}{1000 \text{ mm}} = \frac{0.9 \cdot x \ m^3 ET}{cell}$$

2. Distribution: wherein we perform the above operation on each cell. Multiplying this cell-conversion-factor by a zone sum is the same as multiplying every cell by the conversion factor, and summing the cells. If our coversion factor is y, and we have cells a through d in a particular zone, then:

$$(z \cdot a + z \cdot b + z \cdot c + z \cdot d) = z \cdot \underbrace{(a + b + c + d)}_{zone \ sum}$$

`WARNING/TODO`: Early on, we made a simplifying assumption regarding the grid used to model california. This means that the grids in the northern parts of the state over-estimate how much water precipitated/transpired. This can be fixed with a more appropriate choice of map projection.

```r
xy.res <- 30
z.unit <- 0.001  #Units are in milimeters = 0.001 meters

cwr.master <- cwr.master %>% mutate(vol = zsum * (xy.res^2) *
    z.unit)

ppt.master <- ppt.master %>% mutate(vol = zsum * (xy.res^2) *
    z.unit)
```

Compute IRW and Green-water ET

In order to partition the water footprint into a rain-fed and irrigated water component, we first need to calculate the irrigation water requirement.

```r
# We don't have PPT observations for the year of 2016, so
# let's drop them from the CWR table
cwr.master <- cwr.master %>% filter(date < as.Date("2016-01-01"))

# WARNING: Hardcoded path
cwu.master <- cwr.master %>% select(crop, county, date) %>% mutate(cwr = cwr.master[["vol"]]) %>%
    mutate(ppt = ppt.master[["vol"]]) %>% mutate(et.b = pmax(0,
    (cwr - ppt))) %>% mutate(et.g = pmin(cwr, ppt)) %>% write_rds("output/cwu_master.rds",
    compress = "gz")
```

9.yield_prep.Rmd (County Aggregations)

Contents

This worksheet uses `tidyverse` packages, heavily.

NOTE Make sure to set your working directory to the project root. In RStudio, this means set the `Knit Directory` to "Project Directory".

Meta-data (units)

In our `YIELDS` directory, we inport county-level crop yield data, collected from the County Agricultural Commissioners' (USDA/NASS). Hereafter, this dataset will be referred to as `yield.master`, since this table contains yields, and the yields could be sourced from anywhere (not just USDA/NASS/CAC).

The data used in this worksheet can be found at the following URL, (as of June 2017): https://www.nass.usda.gov/Statistics_by_State/California/Publications/AgComm/Detail/

From the USDA/NASS PDF reports: > Values are recorded for all products grown during the calendar year, regardless of when they are marketed.

From the USDA/NASS reports, *most* production units are expressed in mass units, specifically (short) Tons US (2000 lbs). When tonnes (metric) are used in NASS reorts, I've seen them explicitly specified as 'metric tons'. Other production units include: "Lbs/LBS/lbs" (pounds US), "Cwt/CWT/cwt" (hundredweight, aka 100 pounds US), "Col/COL/col" (colonies, for apiary production), "Each/EACH/each" (units, for certin products such as bee nuclei, queen bees, and turkey eggs for hatching), "Cord/CORD/cord" (volumetric unit for timber) and "Thou/THOU/thou" (thousands of units, for nursery plants).

Area units are in *US survey acres*, based on the square survey foot (NBS Special Publication 447, NIST Handbook 44 - 2012). Differences between the survey acre and the international acre are <2ppm (wikipedia::acre, NIST Handbook 44 - 2012). When hectares are used in NASS reports, I've seen them explicitly specified as 'hectares'.

Yield units are reported as (production mass)/(harvested acre). Specifically, yields are expresed as short tons per acre.

All of the above♪ is also detailed in USDA/ERS Agricultural Handbook Number 697 *"Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products"*. This document also provides conversion factors for unusual commodity weights and measures (eg. carton, bushel, sack, barrel, lug) that we (luckily) don't have to use.

CDL Quirks

Some quirks that you may notice in the CDL data set:

- Inconsistent capitalization in the `Unit` field (LBS vs Lbs vs lbs)
- Typos in the `County` field ("State Total" vs "State Totals" and "San Luis Obisp" vs "San Luis Obispo")
- Typos in the `Commodity Name` field ("CATTLE CALVES EXCLUDED UNSPECIFIE" vs "CATTLE CALVES EXCLUDED UNSPECIFIED" and "NURSERY HORTICULTRAL SPECIMIN MIS" vs "NURSERY HORTICULTRAL SPECIMIN MISC." and others)

Import data and set constants

```
yield.dir <- "input/YIELDS"
kc.index <- readRDS("output/tables/CDL_Kc_LUT_daily.rds")[, 1:2]
kc.lut <- readRDS("output/tables/CDL_Kc_LUT_daily.rds")
cdl_nass <- read_csv("output/tables/CDL_NASS.csv")
counties <- readOGR("input/SHAPEFILES/cnty24k09_poly/cnty24k09_poly_s100.shp")
counties.index <- read_csv("output/ca_counties_attributes.csv")
```

If you end up doing lots of unit conversions, it may make sense to use a dedicated library/package, but luckily, the harvested crops that we are interested in are only reported in US tons (short tons). It's easier for me to verify the calculations on paper if everything is in SI units, so let's define some conversion factors here, for later use.

All conversion factors are from the USDA/ERS Agricultural Handbook Number 697 *"Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products"*.

I really like the `udunits2` and `units` packages, along with the UNIDATA library that they depend on.

```
# TODO: Use units/udunits2/UNIDATA Short tons (US) to tonnes
# (metric tons)
conv <- list(ton = list(ton = 1, kg = 907.18474, tonne = 0.90718474),
    lb = list(lb = 1, kg = 0.45359237, tonne = 0.00045359237),
    cwt = list(cwt = 1, lb = 100, kg = 45.35924, tonne = 0.045359237))
```

Import and prepare yield and county data

Note that by default, the attribute table is read with R data frame conventions, thatis, coercing to factors whenever possible. You can use `readr::type_convert` to run `readr`'s type-guessing logic on existing data frames.

```
yield.paths <- list.files(path = yield.dir, pattern = ".(csv)$",
    full.names = T)

yield.master <- do.call("rbind", lapply(yield.paths, read_csv,
    col_types = cols(`Harvested Acres` = col_number(), Value = col_number())))

rm(yield.dir, yield.paths)
```

```r
# Make the `Units` column lowercase to make things consistent
yield.master[["Unit"]] <- tolower(yield.master[["Unit"]])
```

Explore unique values in harvest dataset

Unique crop categories

Here, we create `common.crops`, which is a character vector of the crop names that have an entry in *all* of the yearly crop reports. In other words, if there's a crop name category that is present in 2005-2011, but not present in 2012, then we will ignore that particular crop.

We're only checking this becaue the length of the unique crop names differs from year to year, suggesting that some categories are present some years, and some aren't present other years. That said, having an entry does not guarentee that there were any harvested acres that year. In other words, a year may have an entry for a crop that did not have any harvested acres (which, is expected).

```r
# Calc number of unique crop categories each year
yield.master %>% group_by(Year) %>% summarize(unique = length(unique(`Crop Name`)))

# Find the crop categories that are present in all years
# (Inner join across all categories)
yield.master %>% group_by(Year) %>% distinct(`Crop Name`)
```

The key operation is `dplyr::intersect()` (find which members of "Crop Name" are in common) which can be thought of as a SQL inner join. First, we produce a list of unique (`dplyr::distinct`) crop names by-year, then we intersect the list of names (performing the inner join).

```r
# TODO: Use dplyr::inner_join() instead of Reduce(intersect)
common.crops <- yield.master %>% group_by(Year) %>% distinct(`Crop Name`) %>%
    summarize(cropname = list(`Crop Name`)) %>% dplyr::select(cropname) %>%
    lapply(function(x) {
        (x)
    })

common.crops <- Reduce(intersect, common.crops[["cropname"]])
all.crops <- unique(unique(yield.master[["Crop Name"]]))
```

Note the crops that were removed. We may want to look at these in more detail later. epecially for categories like "CORN WHITE", or "BEANS SNAP PROCESSING", and "TARO ROOT" (who knew that CA produced taro).

```r
# TODO: Investate the mass of the crops that weren't
# harvested every year
all.crops[!(all.crops %in% common.crops)]
```

Unique ROI categories

We also have to allign the regions of interest (ROI) in our landcover dataset to our harvest dataset. Remember that we partitioned the CWU calculation into ROIs based on counties, since harvest yields are reported at the county-scale.

Now, we have to make sure that the county categories that we used for the spatial partitioning (from our political border shapefile) allign with the county categories used in the harvest yield reporting. We will also construct a look-up table.

Luckily, the CAC county names match up perfectly with the shapefile metadata, aside from a typo or two.

```
# Which counties in the harvest dataset ARE NOT present in
# the ROI index?
unique(yield.master$County)[which(!(unique(yield.master$County) %in%
    counties.index$NAME_PCASE))]

# Fix the typos WARNING: HARDCODED VALUES
yield.master <- yield.master %>% mutate(County = replace(County,
    County == "San Luis Obisp", "San Luis Obispo")) %>% mutate(County = replace(County,
    County == "State Total", "State Totals"))
```

Associate yield ROIs and crop categories with landcover ROIs and crop categories

Associate counties from harvest record with landcover ROIs

Here, we use `dplyr::inner_join` to join the county identifiers, (`counties.index["r_index"]`s to the CAC master table, by matching county names. TODO: Replace `r_index` with `NUM` once we update `ca_counties.tif`

```
# Merge county identifiers
yield.master <- yield.master %>% left_join(counties.index[c("NAME_PCASE",
    "r_index")], by = c(County = "NAME_PCASE")) %>% rename(roi.index = r_index) %>%
    mutate(roi.index = replace(roi.index, County == "Sum of Others",
        991)) %>% mutate(roi.index = replace(roi.index, County ==
    "State Totals", 999))

# TODO: TEST: Did the join operation work properly?
# setdiff(t1[sample(nrow(t1), 100),][c('County',
# 'roi.index')], counties.index[c('NAME_PCASE', 'NUM')] )
# t1[sample(nrow(t1), 100),][['roi.index']] %in%
# counties.index[['NUM']]
```

Associate harvested crops with CDL crops

Summarize harvested acres statewide for comparison with landcover areas

Here, we generate a list harvested acres for all years, for all counties (`County Code == 999`). We save this file as `NASS_acres_state.csv`

The key operation is `dplyr::outer_join()`, an operation that can be thought of as the union set.

Why did we care about harvested acres? It serves as a point of comparison, and allows us to see if the acres harvested align with the acres planted, as identified by CDL. We do this comparison in the `21.validation` worksheet.

NOTE: We are fortunate that the CAC data set includes harvested acres. Don't expect all yield reports from other regions to include harvested acres.

NOTE: I dropped the typos in Crop Name when manually reconciling CDL and CAC categories. TODO: Maybe create logic for flagging instances when a commodity code is mapped to more than one commodity name (which is what happened for "CATTLE CALVES…", "NURSERY FLOWER…" and etc.).

Perform crop-category association

For now, we **manually** reconciled the USDA/NASS production categories to the CDL classes, using a pairing similiar to that used by Fulton et al. 2012 (Appendix 2), where they related the County Agricultural Commissioners' Production Data Commodity Descriptions to "PI Codes" (Where PI likely refers to "Photo Identifier"). This manually-annotated table is found under `output/tables/CDL_NASS.csv`. Let's import it and create a lookup table.

Here, we use `dplyr::inner_join` to join the `cdl_values` and the `cdl_names` to the CAC master table, by matching `Commodity Codes`.

```
# Import if it hasn't been done already WARNING: Hardcoded
# path
cdl_nass <- read_csv("output/tables/CDL_NASS.csv")

yield.lc.master <- yield.master %>% left_join(cdl_nass[c("Commodity Code",
    "cdl_value", "cdl_name")], by = "Commodity Code")

yield.lc.harvested <- yield.lc.master %>% filter(!is.na(cdl_value) &
    !is.na(Production))

yield.lc.excluded <- yield.lc.master %>% filter(is.na(cdl_value) |
    is.na(Production))
```

We just excluded crops that had a missing landcover index (`cdl_values`) **and** no reported yield (`Production`). Therefore, the entries that we excluded had either missing `cdl_values` **or** missing `Production`.

Clean up unit names (CAC-specific)

At least for the CAC harvest dataset, the `Unit` field had some formatting inconsistency. Removing the non-harvested crops seems to have left only one type of production unit

```
unique(yield.lc.harvested["Unit"])
```

`NA` values are found in the entries for `State Totals` or `Sum of Others`, which are expressed in the same units as the entries for each county. Let's just coerce all units to tons (after a quick sanity check).

```
# WARNING: Hardcoded units
yield.lc.harvested <- yield.lc.harvested %>% mutate(Unit = "ton")
```

Summarize harvest data by regions of intereste

Now, let's summarize harvest data by categories and prepare for exporting into our next step. Here is where one of the larger assumptions are made:

- Assumption: our mapping of crops from the harvest data to our landcover classes and crop coefficient classes is somewhat accurate. For example, it's reasonable to map `BERRIES BLACKBERRIES` and `BERRIES RASPBERRIES` into a general *'Caneberries'* category, since we can expect them to have similar growing characteristics, and the landcover dataset does not discriminate between different types of caneberries.

Ideally, we would like to be able to discriminate between different caneberries *and* have a crop growth model for each type of plant.

- Assumption: There is no "double-counting" of harvested crop mass between similar crop categories. For example, if we combine `PEACHES CLINGSTONE`, `PEACHES FREESTONE`, and `PEACHES UNSPECIFIED` into a general *'Peaches'* category, then we wouldn't expect any double counting. What about `CELERY FOOD SERVICE`, `CELERY FRESH MARKET`, and `CELERY UNSPECIFIED`? Well, if a region reports celery harvested for the fresh market, it wouldn't also be reported food service sale. The same goes for other commodities. Cotton has a few categories:`COTTON LINT PIMA`, `COTTON LINT UNSPECIFIED`, `COTTON LINT UPLAND`, `COTTON SEED PLANTING`, and `COTTONSEED`. We wouldn't expect the same biomass to get counted multiple times as lint, planting seed, and cottonseed.

Summarize harvested weight

So, now let's add up all of the production amounts within each county, by commodity, by year. For the county tallies,we drop the categories that represent aggregrations of counties. For the statewide tally, we condition on the `State Totals`category.

```
# WARNING: Hardcoded path
yield.lc.bycounty <- yield.lc.harvested %>%
  group_by(Year, cdl_name, cdl_value, roi.index, County) %>%
  filter(`County Code` != 991 & `County Code` != 999) %>%
  summarise(prod.tons = sum(Production), val.usd = sum(Value), hvst.acres = sum(`Harvested Acres`)) %>%
  mutate(prod.tonne = `prod.tons` * conv[["ton"]][["tonne"]]) %>%
  rename(cdl.value = cdl_value, cdl.name = cdl_name) %>%              # TODO: Replace with consistent na
  mutate_all(as.character) %>%                                          # TODO: Remove with readr:write_cs
  write_csv(., "output/yields/NASS_summarized_bycounty.csv")

yield.lc.bystate <- yield.lc.harvested %>%
  group_by(Year, cdl_name, cdl_value, roi.index, County) %>%
  filter(`County Code` == 999) %>%
  summarise(prod.tons = sum(Production), val.usd = sum(Value), hvst.acres = sum(`Harvested Acres`)) %>%
  mutate(prod.tonne = `prod.tons` * conv[["ton"]][["tonne"]]) %>%
  rename(cdl.value = cdl_value, cdl.name = cdl_name) %>%              # TODO: Replace with consistent na
  mutate_all(as.character) %>%                                          # TODO: Remove with readr:write_cs
  write_csv(., "output/yields/NASS_summarized_bystate.csv")
```

10.wf_calc.Rmd (County Aggregations)

Contents

This worksheet uses `tidyverse` packages, heavily.

NOTE Make sure to set your working directory to the project root. In RStudio, this means set the `Knit Directory` to "Project Directory".

```
# # TODO: Remove readr::type_convert(), once CDL_Kc_LUT_daily
# isnt written with factors kc.index <-
# readRDS('output/tables/CDL_Kc_LUT_daily.rds')[,1:2] %>%
# mutate_all(as.character) %>% type_convert

cdl.index <- read_csv("input/TABLES/cdl_crops.csv")
cwu.master <- readRDS("output/cwu_master.rds")
yield.master <- read_csv("output/yields/NASS_summarized_bycounty.csv")
```

Perform the WF calculation

By calendar year

TODO: Create `../wfs/` if it does not exist already

```
cwu.master.cyear <- cwu.master %>% group_by(crop, county, year = year(date)) %>%
    summarize_at(vars(cwr, ppt, et.b, et.g), funs(sum(., na.rm = TRUE)))

wf.master.cyear <- yield.master %>% inner_join(cwu.master.cyear,
    by = c(cdl.value = "crop", roi.index = "county", Year = "year")) %>%
    mutate(wf.b = (et.b/prod.tonne)) %>% mutate(wf.g = (et.g/prod.tonne)) %>%
    write_csv("output/wfs/wf_total_cyr.csv") %>% write_rds("output/wfs/wf_total_cyr.rds",
    compress = "gz")
```

By water year

```
cwu.master.wyear <- cwu.master %>%
  group_by(crop, county, year = year(waterYearlt(date))) %>%        # Create water year grouping
  filter(n_distinct(date) == 365 | n_distinct(date) == 366) %>%     # Filter for only complete years
  summarize_at(vars(cwr, ppt, et.b, et.g), funs(sum(., na.rm = TRUE)))

wf.master.wyear <- yield.master %>%
  inner_join(cwu.master.wyear, by = c("cdl.value" = "crop", "roi.index" = "county", "Year" = "year")) %
  mutate(wf.b = (et.b / prod.tonne)) %>%
  mutate(wf.g = (et.g / prod.tonne)) %>%
```

```
write_csv("output/wfs/wf_total_wyr.csv") %>%
write_rds("output/wfs/wf_total_wyr.rds", compress = "gz")
```

Save results

Note that we are explicitly casting categorical variables into the proper data type. We do this now, since we are going to be passing these modeling results to the visualization and analysis routines, which can perform specialized operations on

Inspecting discrepancies

Confused why the number of observations in the wf table is so much dramatically smaller (by ~50%) than the yield observations table? Well, there are many entries for crops harvested in counties that were not observed in the landcover dataset for that particular year. This is simply due to deficiencies in the land cover data set (in this case, the CDL).

For example, the CAC records report lettuce harvests for every year from 2007 through 2016. However, the CDL data set does not report any pixels of lettuce landcover for 2007 and 2009 (including the double-crop lettuce landcover categories). There are many other entries with discrepancies, either due to:

- Missing observations in the yield or landcover datasets respectively. `Note`: The CDL dataset is the result of supervised classification (specifically, a decision tree classifier), whereas the CAC dataset is the result of surveys. We can assume that the CAC dataset is more reflective of what is actually harvested and grown within a region.
- Misallignments between the yield and landcover datasets. For example, in Imperial county, sweet corn does not appear in the landcover datasets until 2011, however it is present in the harvest records every year, and unspecified corn is present in the landcover datasets every year. One attempt to reconcile data sets would be to aggregate all types of corn into one category.

To inspect this a bit deeper, we can use `dplyr::anti_join()`, which only displays entries in x (`yield.master`) that **don't** have a matching entry in y (`cwu.master...`).

Bibliography

A.K. Chapagain and A.Y. Hoekstra (Nov. 2004). *Water Footprints of Nations*. Volume 1: Main Report 16. Delft, the Netherlands: UNESCO-IHE Institute for Water Education, p. 80.

Allaire, J. J. et al. (Mar. 2018). *Rmarkdown: Dynamic Documents for R*.

Allan, John Anthony (1997). *'Virtual Water': A Long Term Solution for Water Short Middle Eastern Economies?* School of Oriental and African Studies, University of London.

— (Mar. 2003). "Virtual Water - the Water, Food, and Trade Nexus. Useful Concept or Misleading Metaphor?" en. In: *Water International* 28.1, pp. 106–113. ISSN: 0250-8060, 1941-1707. DOI: 10.1080/02508060.2003.9724812.

Allen, R. G. et al. (1998). *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. FAO irrigation and drainage paper 56. Rome: Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-104219-9.

Allen, Richard G. et al. (Jan. 2005). *The ASCE Standarized Reference Evapotranspiration Equation*. en. Tech. rep. American Society of Civil Engineers.

Bhantana, Parashuram and Naftali Lazarovitch (May 2010). "Evapotranspiration, Crop Coefficient and Growth of Two Young Pomegranate (Punica Granatum L.) Varieties under Salt Stress". en. In: *Agricultural Water Management* 97.5, pp. 715–722. ISSN: 03783774. DOI: 10.1016/j.agwat.2009.12.016.

Boryan, Claire et al. (Aug. 2011). "Monitoring US Agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program". In: *Geocarto International* 26.5, pp. 341–358. ISSN: 1010-6049. DOI: 10.1080/10106049.2011.562309.

Bureau of Reclamation (Jan. 2016). *About AgriMet Crop Coefficients*.

Cahn, Michael and Barry Farrara (Sept. 2012). *ETgage(R) Can Provide Accurate Estimates of Reference Evapotranspiration*.

California County Agricultural Commissioners' Annual Crop Report Manual (2012). Tech. rep. Sacramento, CA: United States Department of Agriculture, National Agricultural Statistics Service, California Field Office.

California Department of Food and Agriculture (2017a). *California Agricultural Exports, 2015-2016*. Tech. rep. Sacramento, CA: California Department of Food and Agriculture.

— (2017b). *California Agricultural Statistics Review, 2015-2016*. Tech. rep. Sacramento, CA: California Department of Food and Agriculture.

California Department of Water Resources (2014). *Water Portfolios, Chapter 10, Volume 5, California Water Plan Update 2013*. Tech. rep. Volume 5. Sacramento, CA, USA: State of California.

CIMIS. http://www.cimis.water.ca.gov/Default.aspx.

Cooper, Joyce Smith and Ezra Kahn (May 2012). "Commentary on Issues in Data Quality Analysis in Life Cycle Assessment". en. In: *The International Journal of Life Cycle*

Assessment 17.4, pp. 499–503. ISSN: 0948-3349, 1614-7502. DOI: 10.1007/s11367-011-0371-x.

Daly, Christopher and Kirk Bryant (June 2013). *The PRISM Climate and Weather System – An Introduction*.

Daly, Christopher, Joseph I. Smith, and Keith V. Olson (Oct. 2015). "Mapping Atmospheric Moisture Climatologies across the Conterminous United States". en. In: *PLOS ONE* 10.10. Ed. by Robert Guralnick, e0141140. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0141140.

Daly, Christopher et al. (Sept. 2002). "A Knowledge-Based Approach to the Statistical Mapping of Climate". In: *Climate Research* 22.2, pp. 99–113. DOI: 10.3354/cr022099.

Daly, Christopher et al. (Dec. 2008). "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States". en. In: *International Journal of Climatology* 28.15, pp. 2031–2064. ISSN: 08998418, 10970088. DOI: 10.1002/joc.1688.

Dettinger, Michael D. et al. (Mar. 2011). "Atmospheric Rivers, Floods and the Water Resources of California". en. In: *Water* 3.2, pp. 445–478. DOI: 10.3390/w3020445.

Dingman, S. Lawrence (2015). *Physical Hydrology*. Waveland press.

Dong, A. et al. (May 1992). "Estimation of Daytime Net Radiation Over Well-Watered Grass". en. In: *Journal of Irrigation and Drainage Engineering* 118.3, pp. 466–479. ISSN: 0733-9437, 1943-4774. DOI: 10.1061/(ASCE)0733-9437(1992)118:3(466).

Doorenbos, J. and W. O. Pruitt (1977). *Guidelines for Predicting Crop Water Requirements*. Rev. FAO irrigation and drainage paper ; 24. Rome: Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-100279-7.

Eching, Simon and David Moellenberndt (Decamber 1998). *Technical Elements of Cimis*. Tech. rep. Sacramento, CA: California Department of Water Resources, p. 57.

Edelen, Ashley and Wesley Ingwersen (June 2016). *Guidance on Data Quality Assessment for Life Cycle Inventory Data*. en. Tech. rep. EPA/600/R-16/096. Washington, DC: Life Cycle Assessment Research Center, Systems Analysis Branch/ Sustainable Technology Division, National Risk Management Research Laboratory, U.S. Environmental Protection Agency.

Ewing, Brad et al. (2010). *The Ecological Footprint Atlas 2010*. en. Tech. rep. Oakland, CA: Global Footprint Network, p. 113.

Finkbeiner, Matthias et al. (Oct. 2010). "Towards Life Cycle Sustainability Assessment". en. In: *Sustainability* 2.10, pp. 3309–3322. ISSN: 2071-1050. DOI: 10.3390/su2103309.

Food and Agriculture Organization of the United Nations (May 2017). *World Programme for the Census of Agriculture 2020*. Food and Agriculture Organization of the United Nations. ISBN: 978-92-5-108865-4.

Fulton, Julian, Heather Cooley, and Peter H. Gleick (2012). *California's Water Footprint*. Tech. rep. Oakland, CA: Pacific Institute for Studies in Development, Environment, and Security.

— (2014). "Water Footprint Outcomes and Policy Relevance Change with Scale Considered: Evidence from California". In: *Water resources management* 28.11, pp. 3637–3649.

GDAL Development Team (2018). *GDAL - Geospatial Data Abstraction Library, Version 2.2.3*.

Gleick, Peter H. (1993). "Water and Conflict: Fresh Water Resources and International Security". In: *International Security* 18.1, p. 79. ISSN: 01622889. DOI: 10.2307/2539033.

GRASS Development Team (2017). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*.

Griffin, Daniel and Kevin J. Anchukaitis (Dec. 2014). "How Unusual Is the 2012–2014 California Drought?" en. In: *Geophysical Research Letters* 41.24, 2014GL062433. ISSN: 1944-8007. DOI: 10.1002/2014GL062433.

Guinée, Jeroen and Reinout Heijungs (2017). "Introduction to Life Cycle Assessment". In: *Sustainable Supply Chains*. Ed. by Yann Bouchery et al. Vol. 4. Cham: Springer International Publishing, pp. 15–41. ISBN: 978-3-319-29789-7 978-3-319-29791-0. DOI: 10.1007/978-3-319-29791-0_2.

Han, Weiguo et al. (June 2012). "CropScape: A Web Service Based Application for Exploring and Disseminating US Conterminous Geospatial Cropland Data Products for Decision Support". In: *Computers and Electronics in Agriculture* 84, pp. 111–123. ISSN: 0168-1699. DOI: 10.1016/j.compag.2012.03.005.

Hart, Q. J. et al. (Mar. 2009). "Daily Reference Evapotranspiration for California Using Satellite Imagery and Weather Station Measurement Interpolation". In: *Civil Engineering and Environmental Systems* 26.1, pp. 19–33. ISSN: 1028-6608. DOI: 10.1080/10286600802003500.

Heijungs, R. and Huijbregts, M.A.J. (2004). "A review of approaches to treat uncertainty in LCA". null. In: *Proceedings of the 2nd Biennial Meeting of iEMSs, Complexity and integrated resources management, 14-17 June 2004, Osnabrück, Germany*, pp. 332–339.

Hijmans, Robert J. et al. (Nov. 2017). *Raster: Geographic Data Analysis and Modeling*.

Hoekstra, Arjen Y. (Mar. 2017). "Water Footprint Assessment: Evolvement of a New Research Field". en. In: *Water Resources Management*. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-017-1618-5.

Hoekstra, Arjen Y. et al. (2011). *The Water Footprint Assessment Manual: Setting the Global Standard*. London ; Washington, DC: Earthscan. ISBN: 978-1-84971-279-8 1-84971-279-4.

Howes, Daniel J. (June 2017). *Appendix F. Irrigation Training and Research Center Mapping Evapotranspiratio at High Respolutojn with Internalized Calibration (ITRC-METRIC)*. Tech. rep. California Polytechnic State University San Luis Obispo: Irrigation Training and Research Center, p. 28.

Hutton, Christopher et al. (Oct. 2016). "Most Computational Hydrology Is Not Reproducible, so Is It Really Science?: REPRODUCIBLE COMPUTATIONAL HYDROLOGY". en. In: *Water Resources Research* 52.10, pp. 7548–7555. ISSN: 00431397. DOI: 10.1002/2016WR019285.

ISO/TC 207/SC 5 (2006). *ISO 14040:2006 Environmental Management – Life Cycle Assessment – Principles and Framework*.

Joel Kimmelshue, Mica Heilmann, and Land IQ (June 2017). *2014 Statewide Land Use Mapping*. Tech. rep. Sacramento, CA: California Department of Water Resources.

Johansson, Emma Li et al. (Oct. 2016). "Green and Blue Water Demand from Large-Scale Land Acquisitions in Africa". en. In: *Proceedings of the National Academy of Sciences* 113.41, pp. 11471–11476. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1524741113.

Kim, Kyle (Apr. 2015a). "From Steak to Mangoes, Here Are Some Water-Hogging Foods". en-US. In: *Los Angeles Times*. ISSN: 0458-3035.

— (Apr. 2015b). "Water Leaves a 'footprint' in Our Food; Here's How It Works". en-US. In: *Los Angeles Times*. ISSN: 0458-3035.

Kimberly Panozzo (Aug. 2016). "A Validation of NASS Crop Data Layer in the Maumee River Watershed". PhD thesis. The University of Toledo.

Knuth, D. E. (Feb. 1984). "Literate Programming". en. In: *The Computer Journal* 27.2, pp. 97–111. ISSN: 0010-4620, 1460-2067. DOI: 10.1093/comjnl/27.2.97.

Lefèvre, Mireille, Michel Albuisson, and Lucien Wald (Feb. 2004). *Description of the Software Heliosat-2 for the Conversion of Images Acquired by Meteosat Satellites in the Visible Band into Maps of Solar Radiation Available at Ground Level*. en. Tech. rep. hal-00867218. Sophia Antipolis, Valbonne, France: Armines / Ecole des Mines de Paris.

Lloyd, Shannon M. and Robert Ries (Jan. 2007). "Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches". en. In: *Journal of Industrial Ecology* 11.1, pp. 161–179. ISSN: 1530-9290. DOI: 10.1162/jiec.2007.1136.

Martin, Derrel L. and James R. Gilley (Sept. 1993). "Irrigation Water Requirements - Chapter 2". In: *National Engineering Handbook Part 623 - Irrigation*. Soil Conservation Service, U.S. Department of Agriculture.

Medellín-Azuara, Josué et al. (Feb. 2018). *A Comparative Study for Estimating Crop Evapotranspiration in the Sacramento-San Joaquin Delta*. Report for the Office of the Delta Watermaster. Davis, CA: Center for Watershed Sciences, University of California, Davis.

Minx, J.C. et al. (Sept. 2009). "Input-Outpoot Analysis and Carbon Footprinting: An Overview of Applications". en. In: *Economic Systems Research* 21.3, pp. 187–216. ISSN: 0953-5314, 1469-5758. DOI: 10.1080/09535310903541298.

Orang, Morteza N., J. Scott Matyac, and Richard L. Snyder (2011). *CUP PLUS (CONSUMPTIVE USE PROGRAM PLUS) MODEL - Version 6.1*.

Orang, Morteza N et al. (Aug. 2013). "California Simulation of Evapotranspiration of Applied Water and Agricultural Energy Use in California". In: *Journal of Integrative Agriculture* 12.8, pp. 1371–1388. ISSN: 2095-3119. DOI: 10.1016/S2095-3119(13)60742-X.

Press, The Associated (May 2018). "California Now World's 5th Largest Economy, Surpassing UK". en-US. In: *The New York Times*. ISSN: 0362-4331.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Richard L. Snyder (Mar. 2014). *Irrigation Scheduling: Water Balance Method*. Tech. rep.

Rigollier, Christelle, Olivier Bauer, and Lucien Wald (Jan. 2000). "On the Clear Sky Model of the ESRA — European Solar Radiation Atlas — with Respect to the Heliosat Method". In: *Solar Energy* 68.1, pp. 33–48. ISSN: 0038-092X. DOI: 10.1016/S0038-092X(99)00055-9.

Safire, William (Feb. 2008). "Footprint". In: *New York Times*.

Schwankl, Lawrence J. et al. (2007). *Understanding Your Orchard's Water Requirements*. Soil, Water, & Irrigation 8212. Davis, CA: University of California, Division of Agriculture and Natural Resources.

Snyder, Richard L. et al. (2007). *Basic Irrigation Scheduling (BIS)*. Davis, CA: University of California, Davis.

Snyder, RL and WO Pruitt (1992). "Evapotranspiration Data Management in California". In: *Irrigation and Drainage: Saving a Threatened Resource—In Search of Solutions*. ASCE. Baltimore, Maryland, United States, pp. 128–133.

Thornton, Peter E., Steven W. Running, and Michael A. White (Mar. 1997). "Generating Surfaces of Daily Meteorological Variables over Large Regions of Complex Terrain". en. In: *Journal of Hydrology* 190.3-4, pp. 214–251. ISSN: 00221694. DOI: 10.1016/S0022-1694(96)03128-9.

United Nations World Water Assessment Programme (2015). *The United Nations World Water Development Report 2015: Water for a Sustainable World*. Vol. 1. Paris: UNESCO Publishing.

USDA-NASS (Jan. 2018). *USDA National Agricultural Statistics Service Cropland Data Layer*. https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php.

Wackernagel, Mathis et al. (June 1999). "National Natural Capital Accounting with the Ecological Footprint Concept". In: *Ecological Economics* 29.3, pp. 375–390. ISSN: 0921-8009. DOI: 10.1016/S0921-8009(98)90063-5.

Walter, Ivan A. et al. (May 2001). "ASCE's Standardized Reference Evapotranspiration Equation". en. In: American Society of Civil Engineers, pp. 1–11. ISBN: 978-0-7844-0499-7. DOI: 10.1061/40499(2000)126.

Wichelns, Dennis (Apr. 2004). "The Policy Relevance of Virtual Water Can Be Enhanced by Considering Comparative Advantages". en. In: *Agricultural Water Management* 66.1, pp. 49–63. ISSN: 03783774. DOI: 10.1016/j.agwat.2003.09.006.

Wickham, Hadley, Winston Chang, and RStudio (Dec. 2016). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.

Wickham, Hadley and RStudio (Nov. 2017). *Tidyverse: Easily Install and Load the 'Tidyverse'*.

World Commission on Environment and Development, ed. (1987). *Our Common Future*. Oxford paperbacks. Oxford ; New York: Oxford University Press. ISBN: 978-0-19-282080-8.

World Economic Forum (2018). *The Global Risks Report 2018*. Tech. rep. Geneva, Switzerland: World Economic Forum.

Zetland, David (2014). *Living with Water Scarcity*. Aguanomics Press.