UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Assessing characterization of large-scale groundwater quality with remote sensing

Permalink
https://escholarship.org/uc/item/46412362

Author
Gibbons, Aimee Christine

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Assessing characterization of large-scale groundwater quality with remote sensing

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Earth System Science

by

Aimée Christine Gibbons

Dissertation Committee:
Professor James S. Famiglietti, Chair
Professor Isabella Velicogna
Professor Kathleen Johnson

2016

DEDICATION

To my family, Gwendolyn, Gregory, and Conor,
and Elliot, Charlie, and Bagheera
for their continuous guidance, support, and patience.

TABLE OF CONTENTS

LIST OF FIGURES

vi

LIST OF TABLES

ACKNOWLEDGMENTS

CURRICULUM VITAE

Aimée Christine Gibbons

EDUCATION

Doctor of Philosophy in Earth System Science   **2016**
University of California, Irvine   *Irvine, California*

Master of Science in Earth System Science   **2013**
University of California, Irvine   *Irvine, California*

Bachelor of Science in Chemistry   **2011**
University of California, Irvine   *Irvine, California*

PROFESSIONAL EXPERIENCE

Graduate Research Assistant   **2011 − 2016**
University of California, Irvine   *Irvine, California*

Teaching Assistant   **2012 − 2016**
University of California, Irvine   *Irvine, California*

CONFERENCE PRESENTATIONS

Gibbons, A., Thomas, B. F., and Famiglietti, J. Empirical estimation of groundwater quality changes using remote sensing. *Poster, AGU Fall Meeting 2015.*

Gibbons, A., Thomas, B. F., and Famiglietti, J. Groundwater Quality Changes Using Remotely Sensed Groundwater Storage. *Poster, AGU Fall Meeting 2014.*

Gibbons, A., Thomas, B. F., and Famiglietti, J. Groundwater Quality Changes Using Remotely Sensed Groundwater Storage. *Talk, GSA Fall Meeting 2014.*

Gibbons, A., Thomas, B. F., and Famiglietti, J. Exploring the potential of remote sensing groundwater quality variations using GRACE. *Poster, AGU Fall Meeting 2013.*

Gibbons, A., Thomas, B. F., and Famiglietti, J. Correlating Dissolved Solid Concentration Changes with GRACE-based Changes in Water Table Depth. *Poster, AGU Fall Meeting 2012.*

ABSTRACT OF THE DISSERTATION

Assessing characterization of large-scale groundwater quality with remote sensing

By

Aimée Christine Gibbons

Doctor of Philosophy in Earth System Science

University of California, Irvine, 2016

Professor James S. Famiglietti, Chair

NASAs Gravity Recovery and Climate Experiment (GRACE) mission measures global gravity variability, which are converted into monthly groundwater storage variations in the world's largest watersheds. Large-scale groundwater storage variability is determined from GRACE and supplementary data at monthly and longer timescales for regions that are 150,000 km^2 and greater. Estimates of groundwater availability focus on quantity, but methods to infer groundwater quality have not yet been developed, in part due to a lack of spatially representative quality data. Natural contaminants dissolved in groundwater generally increase with depth due to continued dissolution of rock and soil material along flow paths. Anthropogenic contaminants are generally concentrated near the water table due to changes in frequency and location of recharge. These basic relationships between groundwater quality and depth provide the conceptual framework for the project research. This work aims to characterize relationships between observed total dissolved solid (TDS) concentrations and GRACE-derived subsurface storage anomalies for the High Plains aquifer in the central United States and the Central Valley aquifer in California. The relationship between observed water levels and contaminant concentrations are expected to vary based on physical parameters influencing spatial and/or temporal patterns of infiltration including dominant land use type, principle rock and soil types, and constituent. In this work, a database of publicly available *in situ* TDS concentrations in groundwater and groundwater levels is compiled

for each of the study areas and assessed for simplistic preliminary relationships, and methods of scaling point observations and large-scale gridded data are explored. Models estimating average TDS concentrations through time as a function of *in situ* groundwater levels and season are constructed and explore potential improvements by classifying models in terms of dominant lithology and land use, and by including GRACE-derived subsurface storage anomalies as a potential predictor. Finally, two spatial analysis approaches explore methods of TDS characterization on a subbasin scale and TDS variability in time and space on regional scales. Results of this work have implications on improving groundwater management practices by exploring potential methods of estimating groundwater quality on regional to global scales using remote sensing.

Chapter 1

Background and introduction

Groundwater is an important source of freshwater. It is a primary freshwater source to 2 billion people globally (Alley et al. [2002]), provides buffering capacity to populations relying primarily on surface water, and irrigates agriculture. Current demands on groundwater resources are expected to increase, particularly with a growing population and freshwater redistribution fueled by climate change (Solomon [2007], Wada et al. [2010], Gleeson et al. [2012]). Groundwater abstractions reapplied at the surface are also prone to redistribution by evapotranspiration (Lo and Famiglietti [2013]), providing positive feedback mechanisms enhancing shifts in water availability given likely changes in climate.

Developed and developing countries are currently facing groundwater depletion and contamination (Gleeson et al. [2011]), the two primary strains on groundwater resources (Wada et al. [2010]). To date, groundwater availability studies conservatively consider only quantity (Shiklomanov [2000], Vörösmarty et al. [2010], Gleeson et al. [2012]). Groundwater depletion rates are estimated in numerous studies (Rodell and Famiglietti [2002], Rodell et al. [2009], Wada et al. [2010], Famiglietti et al. [2011], Gleeson et al. [2012], Scanlon et al. [2012b,a], Voss et al. [2013], Döll et al. [2014], Richey et al. [2015b], Nanteza et al. [2016]), and only one study attempts to quantify remaining reservoir volumes (Richey et al. [2015a]). Estimating stress on groundwater resources accounts for availability, historically represented by renewable water cycle fluxes (WWAP [2003]) and more recently by withdrawal statistics

(Döll [2009], Wada et al. [2010, 2011]) or recharge flux (Richey et al. [2015b]). The recent approaches to quantifying groundwater stress are improvements upon historic methods, but continue to lack an explicit component of groundwater quality estimating the proportion of storage meeting water quality criteria for the intended use. To further address global groundwater availability, it is necessary to incorporate terms of quantity and quality that account for potability.

1.1 Historic groundwater quality

Depending on the intended use, the quality of groundwater is nearly as important as the quantity. This section gives a brief overview of the groundwater quality literature to provide context to the work in the following chapters.

1.1.1 Space and time scales of historic case studies

Classic groundwater quality studies generally focus on attributing a contamination event to a local source (e.g. Garrels and Mackenzie [1967], Todd et al. [1980], Presser and Ohlendorf [1987], Wagner and Gorelick [1987], Deverel and Gallanthine [1989], Meisinger et al. [1991], Alley [1993], Loftis [1996], Pebesma and De Kwaadsteniet [1997]). These studies are based on contamination point measurements, groundwater contamination transport models, and/or laboratory column experiments. Back-trajectories are estimated to isolate potential local point or non-point sources of contamination, such as from land-fills or agriculture (Meisinger et al. [1991], Kent and Landon [2013]). Studies falling under this category are generally on local spatial scales, as collecting and analyzing samples and running computational groundwater transport models can be expensive. While the literature in each of these fields is well established, initial conditions and parameterization are difficult where point measurements of groundwater quality are unavailable, limiting applicability of predictions to larger spatial

scales.

Basin-scale studies are important to understanding the quantity of freshwater within a watershed and can provide insight on the evolution of the water quality. Groundwater quality studies on basin scales are able to consider all sources of recharge within the watershed and whether the quality of recharge changes from the point of infiltration (Charles et al. [1993], Lashkaripour et al. [2005], Lashkaripour and Ghafoori [2011], Shamsudduha et al. [2015]).

Regional-scale groundwater quality studies are rooted in national monitoring assessment programs. National networks are generally implemented to identify sources of public health risks and inform management practices to prevent future exposure (Pebesma and De Kwaadsteniet [1997], Swartjes [1999], Broers and van der Grift [2004], Lashkaripour et al. [2005], Hossain [2006], Lashkaripour and Ghafoori [2011], Shamsudduha et al. [2015]).

Beginning in the early 1990s, groundwater quality research in the US turned toward assessing overall quality, as evidenced by the implementation of the National Water-Quality Assessment (NAWQA) Program by the United States Geological Survey (USGS) in 1991 (USGS NAWQA). The NAWQA program aims to provide "long-term consistent and comparable information on streams, rivers, groundwater, and aquatic systems" to help inform water management decisions from local to national levels using standardized sampling and analyzing protocol (Lapham et al. [1995]). The experiments in the following chapters would not be possible without the information collected and provided publicly by the NAWQA program.

Several studies compile data from the NAWQA program to evaluate changes in groundwater quality on regional spatial scales (Embrey and Runkle [2006], Anning et al. [2007], McMahon et al. [2007], McMahon and Chapelle [2008], DeSimone and Hamilton [2009], Gurdak et al. [2009], Lindsey et al. [2009], Ayotte et al. [2011], Lindsey and Rupert [2012], Chaudhuri and Ale [2014a]). It is challenging to obtain regularly-spaced samples over such large regions,

let alone within a small time frame, so such studies focus on aggregate sets of observations taken over a period of up to ten years and compare concentrations of constituents against either established health standards or a second set of observations taken over a different period. The median of the difference in total dissolved solid concentrations range from no significant change to significant increases in the range of 10-50 mg/L, depending on the land use network type and location in the Central Valley and no significant changes in the High Plains between 1988 and 2010 (Lindsey and Rupert [2012]). The median of the difference in chloride concentration ranges from no significant change to an increase between 1-20 mg/L in the Central Valley and significant chloride increases from 1-20 mg/L in the High Plains between 1988 and 2010 (Lindsey and Rupert [2012]). The median of the difference in nitrate concentration ranges from no significant change to an increase greater than 0.5 mg/L as N in both the Central Valley and the High Plains between 1988 and 2010 (Lindsey and Rupert [2012]).

1.1.2   Commonly studied groundwater quality parameters

Physical measures of water quality are those that quantify overall properties without specifically identifying constituent species present. Physical observations of water quality includes temperature, pH, turbidity, total suspended solids, electrical conductivity (E. C.), dissolved oxygen, salinity, and total dissolved solids. Parameters like temperature, pH, turbidity, and total suspended solids are useful for indicating habitability of surface water for fish and other macrofauna. Dissolved oxygen content is observed to determine whether an environment is oxic or anoxic and identify reduction-oxidation (redox) conditions, important in determining the direction of other chemical processes (e.g. denitrification or nitrification) (McMahon and Chapelle [2008], Landon et al. [2011]). Electrical conductivity is a cost-effective means of measuring the ionic content, and is closely linked to salinity and total dissolved solid concentrations (Deverel and Gallanthine [1989], Letey et al. [2002], Subramani et al. [2005],

4

Chaudhuri and Ale [2014a,b]). Total dissolved solid concentrations are a measure of all particles in a volume of water, and is a commonly used metric as an indicator of salinity (Alley [1993], Planert and Williams [1995], Loftis [1996], Anning et al. [2007], Lindsey and Rupert [2012]).

Regional geology is the expected source of natural groundwater contaminants in a given location. Local formations and underlying bedrock are the parent material source of natural contaminants, which can be dissolved into groundwater over time. Elements dissolved into groundwater related to common rock-forming materials include sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), silicon (Si), aluminum (Al), carbon (C), sulfur (S), and iron (Fe) (Alley [1993]). Other naturally occurring elements in groundwater are of potential concern due to toxicity in humans, including lead (Pb), arsenic (As), selenium (Se), mercury (Hg), cadmium (Cd), nitrogen (N), sulfur (S), uranium (U), manganese (Mn), copper (Cu), zinc (Zn), barium (Ba), and chromium (Cr) (Alley [1993], Jurgens et al. [2010]).

Arsenic is a groundwater contaminant commonly studied particularly because it is poisonous to humans at elevated concentrations, especially over long periods of time. Groundwater contamination by arsenic is intensely monitored in India and Bangladesh (Chatterjee et al. [1995], Bhattacharya et al. [1997], Acharyya et al. [2000], Chowdhury et al. [2000], Fazal et al. [2001], Hossain [2006], Shamsudduha et al. [2015]), where mass arsenic poisonings have occurred. Arsenic can infiltrate groundwater resources through certain fertilizer and pesticides, still used particularly in Asia. In most cases, however, arsenic is a naturally occurring contaminant eroded from natural deposits, currently thought to originate from pyrite oxidation triggered by lowering of the water table and/or natural reduction of oxy-hydroxide (Fazal et al. [2001], Hossain [2006]). Naturally occurring arsenic is also present in parts of China, Thailand, Vietnam, Mongolia, Nepal, Chile, Argentina, Mexico, and the southwest United States (Erban et al. [2013]).

Anthropogenic contaminants are those sourced from human practices. Common anthro-

pogenic contaminants include nitrogen, chloride, pesticides, heavy metals, and hydrocarbons (Walker et al. [1991], Böhlke [2002], Jurgens et al. [2010], Ayotte et al. [2011], Lindsey and Rupert [2012], Exner et al. [2014]). A common anthropogenic contaminant is nitrogen, which generally originates from non-point sources such as fertilizer, septic systems, and atmospheric deposition (Lindsey et al. [2009], Exner et al. [2014]). Nitrogen typically occurs in groundwater as nitrate and/or nitrite, a major component of inorganic fertilizers. Nitrate and nitrite are small, water-soluble ions which, when applied as fertilizer, provide biologically available nitrogen to increase crop yields. Liberal fertilizer use is practiced globally in agriculture, as it is a relatively cost-effective investment. However, when combined with regular irrigation increasing the frequency of recharge, the underlying groundwater is increasingly susceptible to nitrate contamination (Gurdak and Qi [2006], McMahon et al. [2007], McMahon and Chapelle [2008], Puckett et al. [2008], Lindsey et al. [2009], Landon et al. [2011], Kent and Landon [2013], Exner et al. [2014]).

1.1.3   Statistical tools used to describe water quality

Mapping observations of water quality is a common method of identifying potential patterns of contamination and identifying potential sources (Loftis [1996], Pebesma and De Kwaadsteniet [1997]). Spatial patterns of water quality are determined by spatially interpolating available observations. The most common spatial interpolation technique is kriging. Ordinary and simple kriging are the most common types, where ordinary kriging assumes an unknown mean value remains constant in unsampled locations, and simple kriging requires the mean value is known and is stationary over the domain (Tabios and Salas [1985], Cressie [1988, 1990]). Maps interpolated by kriging show the spatial distribution of groundwater quality concentrations, which can be used to identify contaminant sources or processes influencing water quality (McMahon et al. [2007], Landon et al. [2011], Shamsudduha et al. [2015]).

Detection of trends in water quality over time requires relatively frequent sampling. Trend detection methods and transformations for water quality concentrations are common. Selection of trend analysis techniques require familiarity with the data to select the best methods for the intended outcome (Lettenmaier [1976], Mosteller and Tukey [1977], Tukey [1977], Hirsch et al. [1982, 1991], Helsel and Hirsch [2002]). Transformations are applied to reshape the distribution of values and generally are log-transformed, although again familiarity with the data is necessary to selecting the best transformations (Lettenmaier [1976], Mosteller and Tukey [1977], Tukey [1977], Hirsch et al. [1982, 1991], Helsel and Hirsch [2002]). In general, non-parametric tests are implemented over parametric tests because the former do not assume that the spread of data have a specific distribution, often the normal distribution. To determine whether a potential trend through time exists, a rank correlation is often calculated between time and constituent concentration (Hirsch et al. [1991], Broers and van der Grift [2004], Lindsey et al. [2009], Kent and Landon [2013]). If the rank correlation is significant, then the magnitude of the trend is estimated using a Mann-Kendall test, also known as a Sen slope estimator (Hirsch et al. [1991], Loftis [1996], Broers and van der Grift [2004], Landon et al. [2011], Kent and Landon [2013]).

If temporally frequent samples are not available, it is common to define two or more periods for comparison, such as wet and dry seasons (e.g. Kent and Landon [2013], Bexfield and Jurgens [2014]) or different decadal periods (e.g. Lindsey and Rupert [2012], Chaudhuri and Ale [2014a], Shamsudduha et al. [2015]). Determining changes in groundwater quality parameters between two periods of time allows for collection and analysis of samples over larger spatial extents (Chaudhuri and Ale [2014a], Shamsudduha et al. [2015]). Potential differences between two periods are often determined using Wilcoxon rank-sum test, a seasonal Kendall test, or a Mann-Whitney test for median differences (Lettenmaier [1976], Hirsch et al. [1982, 1991], Loftis [1996], Helsel and Hirsch [2002], Lindsey et al. [2009], Lindsey and Rupert [2012]).

1.1.4 Variables attributed to water quality

Hydrogeologic properties determine characteristics of groundwater flow which logically affect groundwater quality patterns in time and space. The descriptive properties dictating 3-dimensional flow direction and velocity, such as hydraulic head gradients, porosity, specific yield, and hydraulic conductivity, determine the spatial patterns of potential constituent plumes and how the plume moves through the aquifer over time (Alley [1993], Fetter and Fetter [1999], Shamsudduha et al. [2015]). The presence or absence of confining rock or clay layers influence vertical movement within the aquifer. Geologic characteristics such as bedrock type or presence of natural deposits determine the natural constituent species expected due to dissolution. Hydrogeologic properties may be homogeneous or heterogeneous in space, determined by field surveys and tests, but a particular characteristic is often described as a range of values or as an average value over a given region.

Recharge is an important factor affecting groundwater quality, particularly in shallow groundwater. A well-established field of study, groundwater recharge is generally a function of climate, geomorphology, soil properties, and vegetation (Lerner et al. [1990], de Vries and Simmers [2002], Maxwell et al. [2016]). Natural freshwater recharge carrying contaminants in runoff can infiltrate an aquifer at the land surface or through surface water (Deverel and Gallanthine [1989], Lerner et al. [1990], Charles et al. [1993], de Vries and Simmers [2002], Gurdak and Qi [2006], Shamsudduha et al. [2015]). In arid climates where irrigated land cover is linked to soil salinization by evaporites, recharge by irrigation is a mechanism of constituent infiltration from soils to groundwater (Swartjes [1999], de Vries and Simmers [2002], Scanlon et al. [2005], Chaudhuri and Ale [2014a,b]).

Land use cover affects properties of recharge such as rates, frequency, and location, and therefore mechanisms of contaminant infiltration (Walker et al. [1991], Harbor [1994], Eckhardt and Stackelberg [1995], de Vries and Simmers [2002], Scanlon et al. [2005]). In areas

of natural land cover, the rate and frequency of recharge are related to regional climate and precipitation with infiltration related to properties of the lithology and geomorphology (Lerner et al. [1990], de Vries and Simmers [2002]). Land use cover changes from natural to anthropogenic uses alters properties of recharge and risk of exposure to constituents. The type of anthropogenic land use dictates the particular constituents introduced to a system deposited to the soil, leaving underlying groundwater reserves susceptible to contamination with future infiltration (Böhlke [2002], Scanlon et al. [2005], Kent and Landon [2013], Exner et al. [2014]). Paved urban land use reduces local recharge and redirects runoff with increased risk of exposure to contaminants such as hydrocarbons. Unpaved anthropogenic land use such as landfills and agriculture can recharge locally. Landfills can deposit chlorine, hydrocarbons, and heavy metals (Lisk [1991]). As discussed in section 1.1.2, agricultural land use is known to elevate concentrations of nitrate and TDS in soils and groundwater (Meisinger et al. [1991], Swartjes [1999], Böhlke [2002], Letey et al. [2002], Gurdak and Qi [2006], Lindsey et al. [2009], Landon et al. [2011], Lindsey and Rupert [2012], Kent and Landon [2013], Chaudhuri and Ale [2014b], Exner et al. [2014]).

1.2   Remote sensing and groundwater

Studies exploring remote sensing of water quality are currently limited to assessing water color or turbidity as an indicator of surface water quality variations (e.g. Mertes et al. [1993], Ritchie et al. [2003]). The Gravity Recovery and Climate Experiment (GRACE) mission measures changes in terrestrial water storage from which components of storage can be removed to give changes in groundwater storage, as described in the following sections. The large-scale behaviors can be used to evaluate groundwater dynamics over time, including rapid decreases in storage attributed to pumping.

9

1.2.1 GRACE terrestrial water storage anomalies

GRACE, a joint mission between the United States (NASA) and Germany (DLR), is a pair of satellites measuring time-variable gravity by orbiting Earth one after another in a pole-ward orientation, completing one global cycle approximately once a month (Tapley et al. [2004]). As GRACE passes over regions with larger mass, and therefore gravitational pull, the distance between the two satellites increases as the first satellite experiences the gravitational pull towards the surface mass before the other. The distance between them decreases when the second satellite also experiences the gravity of the larger mass on the surface. Moving away from the surface mass, the first satellite moves away from the surface, increasing the distance between the twin satellites. The distance between the GRACE satellites changes once more as the second satellite also moves away from the surface mass. By measuring the precise distance between the two satellites, changes in global gravity can be mapped at high vertical resolution.

The distance between the two satellites is constantly monitored by a microwave sensor, accurate to one micron ($1 \times 10^{-6}m$). North-south striping artifacts resulting from the pole-ward orbit path are removed, and gravitational effects of atmospheric pressure systems and glacial isostatic adjustments are subtracted from the time-variable gravity fields (Landerer and Swenson [2012]). The gravity fields are then spatially smoothed by applying a Gaussian filter with a 300 km radius in the RL05 level 3 solutions, then are inverted into mass variations (Wahr et al. [1998], Rodell and Famiglietti [1999, 2001], Velicogna et al. [2001], Swenson and Wahr [2002], Swenson et al. [2003], Landerer and Swenson [2012]). Orbiting at an altitude of 450 km, the monthly mass anomalies have a 300 km horizontal resolution at the Earth's surface, and a vertical accuracy of approximately 1-2 cm equivalent water height.

The resulting gridded monthly GRACE mass anomalies are attributed to total terrestrial water storage variations over land (Wahr et al. [1998], Rodell and Famiglietti [1999, 2001],

Velicogna et al. [2001], Swenson and Wahr [2002], Swenson et al. [2003]). GRACE measurements are holistic in that there is no differentiation between reservoir types (e.g. soil moisture, surface water, snow), providing the total integrated water storage anomalies in a column.

Because of filtering and truncation of spherical harmonics, the GRACE RL05 level 3 signal is attenuated. The amplitude is restored by applying the accompanying gridded scale factor to each month of storage anomalies (Landerer and Swenson [2012]). The scaling factor is generated by quantifying the differences between the filtered GRACE product and the Community Land Model (CLM) with the same filtering applied, provided by Swenson [2012]. Leakage errors, introduced by filtering and truncation of spherical harmonics, and measurement errors are both scaled to match the magnitude of scaled gridded GRACE data, with the total error of each grid cell equal to the square root of the sum of squared errors. For specific basins, the total error is calculated to account for correlation between neighboring grid cells.

1.2.2   GRACE and groundwater

GRACE is used to study total water storage and groundwater storage variations in many regions such as the Central Valley, High Plains, Northern India, Illinois, and the East African basin (Rodell and Famiglietti [2002], Yeh et al. [2006], Strassberg et al. [2007], Rodell et al. [2007, 2009], Strassberg et al. [2009], Famiglietti et al. [2011], Scanlon et al. [2012a,b], Voss et al. [2013], Richey et al. [2015a,b], Nanteza et al. [2016]). Because GRACE provides total integrated storage anomalies, groundwater anomalies are derived by subtracting the other regionally important hydrologic storage components from a water balance, as shown in equation 1.1 (Rodell and Famiglietti [2002]):

$$dS = dSurface + dSnow + dSoil + dGroundwater \qquad (1.1)$$

where d denotes anomalies, S is total terrestrial water storage, $Surface$ is surface water storage, $Snow$ is snow water equivalent storage, $Soil$ is soil moisture storage, and $Groundwater$ is groundwater storage. Surface water storage is generally the sum of all surface water such as rivers or lakes, as well as water stored in vegetation such as tree canopies. The relevant water storage terms vary by region. For example, ground-based groundwater level observations and modeled soil moisture are shown to match well with GRACE total water storage estimates in the High Plains aquifer in the central United States (Strassberg et al. [2007]).

Auxiliary data are necessary to isolate groundwater storage anomalies from the GRACE signal (Strassberg et al. [2007], Rodell et al. [2007, 2009], Strassberg et al. [2009], Famiglietti et al. [2011], Scanlon et al. [2012b], Richey et al. [2015a,b], Nanteza et al. [2016]). When available, *in situ* data are used to estimate regionally significant components of water storage variability, as demonstrated for snow water equivalent storage in Famiglietti et al. [2011]. When *a piori* data are unavailable, modeled storage estimates are used, as done for soil moisture by Famiglietti et al. [2011].

GRACE-derived groundwater depletion estimates indicate strong anthropogenic withdrawal impacts on quantity (Rodell and Famiglietti [2002], Yeh et al. [2006], Rodell et al. [2009], Famiglietti et al. [2011], Richey et al. [2015b]). Groundwater pumping, particularly in large volumes, alters hydrologic properties affecting flow mechanisms (Gurdak et al. [2009], McMahon et al. [2007]), which influences contaminant concentrations with depth by promoting vertical mixing. Mixing high and low quality water can result in an overall change in average water quality, where the direction of change depends on the relative volumes and concentrations of constituents of each water parcel.

Quantifying groundwater reserves has poignant implications for sustainable use and management; however such studies have not addressed freshwater availability as a factor of both quantity and quality. In the following work, the holistic perspective on groundwater quantity variations of GRACE are used with local point measurements of groundwater contaminant
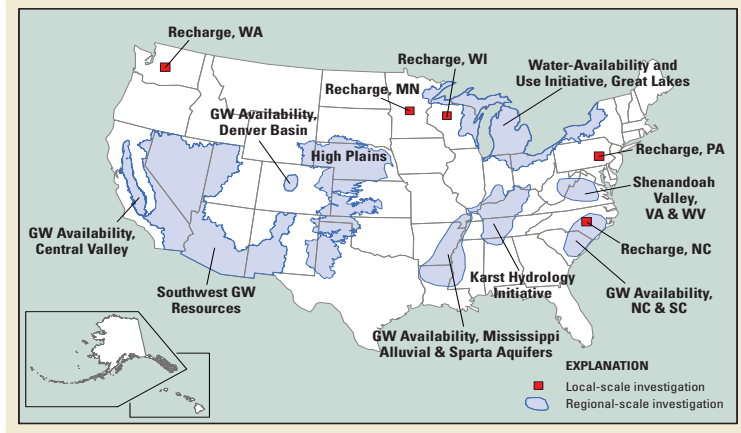
Figure 1.1: Map showing locations of case study areas in the US: California's Central Valley and the High Plains from Dennehy [2005].

concentrations to improve understanding of relationships between quantity and quality variations. To understand and develop these relationships, two case studies in the US serve as the focus: the High Plains aquifer and the California Central Valley aquifer, shown in Figure 1.1. These aquifers rank among the worlds largest, each having a contributing watershed area of $> 150,000 km^2$, important to reducing error of GRACE dS, and are both monitored on a local level by the USGS in coordination with relevant state agencies. National and state programs described in section 2.2.1 provide groundwater quality data and depth to water data for each aquifer, shown in Figure 2.1.

1.3 Hydrogeologic settings of study aquifers

1.3.1 High Plains

The High Plains aquifer lies under eight states in the central US, shown in Figure 1.1, including Wyoming, South Dakota, Nebraska, Kansas, Colorado, Oklahoma, New Mexico, and Texas. It extends from southern South Dakota in the north through the Texas panhandle in the south, and is bounded by the Rocky Mountains to the west, and extends into Nebraska, western Kansas, and the Oklahoma panhandle to the east (Whitehead [1996], Miller and

Appel [1997]). The elevation is highest in the west near the Rockies around 7,800 feet, and declines eastward to about 1,100 feet (Dennehy [2000]).

The High Plains has a dry continental semi-arid climate with frequent winds, low humidity, high evaporation rates, and moderate precipitation (Dennehy [2000], McMahon et al. [2007]). The mean annual air temperature is lowest in the northwest at approximately 5°C, and increases moving south to 18°C (Thornton et al. [1997], McMahon et al. [2007]). Precipitation on the High Plains is the main source of recharge to the aquifer (Whitehead [1996], Miller and Appel [1997], McGuire [2009], McGuire et al. [2012], McGuire [2014]). Precipitation reaches a maximum during late spring and early summer monsoons, with a mean annual precipitation reaching approximately 16 inches in the west, and 28 inches in the east (Dennehy [2000], McMahon et al. [2007], Gurdak et al. [2009], McGuire et al. [2012]). The aquifer discharges to surface water streams and to the atmosphere by evapotranspiration where the water table is near the land surface (Dennehy [2000]).

The principal geologic feature is the Ogallala formation underlying 80% of the High Plains (Dennehy [2000]). Some older formations are present in portions of the northern High Plains, the Arikaree and Brule, and some younger deposits are present on top of or adjacent to the Ogallala, namely alluvial deposits along river systems and the Nebraska Sand Hills (Whitehead [1996], Miller and Appel [1997], Dennehy [2000], McMahon et al. [2007], Qi and Christenson [2010]). The aquifer is generally unconfined, unconsolidated sand and gravel with a regional flow system, with an average specific yield of 0.15 (Whitehead [1996], Miller and Appel [1997], Dennehy [2000], McGuire [2009]).

Groundwater generally flows from west to east following the elevation gradient (Dennehy [2000], McMahon et al. [2007]). Groundwater levels have declined in some areas since pre-development (circa 1950), mostly occurring within the central and southern portions of the High Plains (McMahon et al. [2007], McGuire [2009], McGuire et al. [2012], McGuire [2014]). Pumping for irrigation, beginning around 1950, continue to reduce groundwater

storage (McGuire [2009], Qi and Christenson [2010], McGuire et al. [2012], McGuire [2014]). Contamination of recently recharged groundwater in the High Plains is generally attributed to interactions with rock material, redox reactions, mixing with water from other sources, and widespread irrigated cropland (McMahon et al. [2007]). Contaminants commonly studied in the High Plains aquifer include nitrate, dissolved solids, and pesticides, with high dissolved solid concentrations associated largely with bedrock types and in irrigated alluvial valleys, and nitrate and pesticides associated with agricultural land use (Whittemore [1995], Mehta et al. [2000], Whittemore et al. [2000], McMahon et al. [2007], Gurdak et al. [2009], Lindsey and Rupert [2012]).

The High Plains aquifer is the principal freshwater source in the region and supports one of the largest agricultural sectors in the country (Dennehy [2000]). Groundwater is the primary water source for approximately 80% of the 2 million people living within the High Plains boundaries (Dennehy [2000], McMahon et al. [2007]). Rangeland and agriculture are the dominant land cover uses in the area (McMahon et al. [2007], Homer et al. [2015]), with livestock being predominantly cattle and pigs (McMahon et al. [2007]). The region produces wheat, sorghum, and soybeans, and has recently increased production of corn, hay/alfalfa, cotton, and peanuts (McMahon et al. [2007]).

1.3.2 Central Valley

California's Central Valley aquifer extends through the middle of the state, as shown in Figure 1.1. It is almost entirely surrounded by mountain ranges with the Sierra Nevada Range to the east, the Coastal Ranges to the west, the Cascade and Klamath ranges to the north, and the Tehachapi Mountains to the south. The Sierra Nevada and Cascade ranges are primarily igneous formations, and the Coastal Ranges are marine sedimentary formations (Planert and Williams [1995]).

In the arid to semi-arid hot climate, surface water resources can be limited in the Central Valley due to low precipitation and relatively high reference evapotranspiration (Faunt [2009]). Mean annual precipitation between 1968 and 2003 was 11-16.8 inches, with a median of 13 inches (Faunt [2009]), falling mostly between November and April in average years. Before the 1960s, precipitation only accounted for a small portion of recharge, with streams carrying snow melt from the Sierra Nevada in the eastern valley and Cascade and Klamath mountains in the northern valley being the primary source of aquifer recharge (Planert and Williams [1995]). Currently, water users in the Valley rely heavily on a complex system of surface water diversions and Sierra Nevada melt water. In dry years, the underlying aquifer is pumped to meet the agricultural and urban needs (Faunt [2009], Scanlon et al. [2012b], Lo and Famiglietti [2013]). The water table is in steady decline as a result of decades of virtually unrestricted groundwater pumping (Faunt [2009], Famiglietti et al. [2011], Scanlon et al. [2012b]).

The valley-fill aquifer has a regional flow system and sits on top of the subsurface continuation of the Sierra Nevada, which has a slight slope westward (Faunt [2009]). The sediment fill originates from marine deposits, continental deposits, and deltaic deposits generally from the Sierra Nevada and Coastal ranges (Farrar and Bertoldi [1988], Faunt [2009]). Alluvial fans are present on all sides of the valley, with larger stream channels along the eastern boundary (Faunt [2009]). Approximately the top one hundred meters of the Central Valley aquifer is heterogeneous and unconfined, providing the primary source of freshwater, with conditions becoming increasingly confined with increasing depth due to the presence of clay beds (Planert and Williams [1995], Faunt [2009]). Saline water is present under the freshwater zone, but can be found at shallower depths due to upward migration of deeper water, estuary water trapped during sedimentation, or concentration of evaporites (Farrar and Bertoldi [1988], Faunt [2009]).

Groundwater contaminants from natural sources in the Central Valley include total dissolved

solids, chloride, boron, arsenic, and selenium (Deverel and Gallanthine [1989], Dubrovsky et al. [1993], Lindsey and Rupert [2012]). TDS concentrations in Central Valley groundwater can originate in streams, eroded from the parent material of the surrounding mountains. The igneous formations to the north and east have relatively low concentrations of TDS, particularly compared to the marine sedimentary formations to the west. TDS generally increases with depth, but is expected to be elevated in deeper wells of the southern and western portion of the San Joaquin Valley (Planert and Williams [1995]). Boron in the Central Valley is a result of leaching from rocks and soils containing borate and borosilicate minerals. Arsenic and selenium occur naturally in the San Joaquin Valley (Presser and Ohlendorf [1987], Deverel and Gallanthine [1989], Presser and Swain [1990], Dubrovsky et al. [1993], Welch et al. [2000], Letey et al. [2002]), and have been shown to be increasing in these areas (Presser and Swain [1990], Rezaie-Boroon and Fisher [2012]).

The Central Valley aquifer provides an alternative freshwater source in dry years when surface water deliveries cannot meet the needs of the increasing 3.8 million population and a large agricultural sector (Faunt [2009]). With a mostly flat valley floor and abundant sunshine, Central Valley farms produce a majority of the fruits and vegetables in the US with an estimated value of $17 billion per year (Faunt [2009]). Over 250 crop types are grown in the Central Valley, with the dominant types being cereal grains, hay, cotton, tomatoes, vegetables, citrus and tree fruits, nuts, and grapes (Faunt [2009]).

1.4   Research objectives and hypotheses

There is a strong literature history assessing groundwater quality in the US and abroad (e.g. Alley [1993], Litke [2001], McMahon et al. [2007], Gurdak et al. [2009], Lindsey and Rupert [2012], Shamsudduha et al. [2015]). The literature identifies point and non-point contaminant sources and processes influencing groundwater quality variability such as local

flow paths and mixing effects of pumping large volumes of groundwater (Charles et al. [1993], Lashkaripour et al. [2005], Schoups et al. [2005], Lashkaripour and Ghafoori [2011], Kent and Landon [2013], Bexfield and Jurgens [2014], Exner et al. [2014], Shamsudduha et al. [2015]). However, such classic studies are limited to relatively small spatial scales and/or step-wise changes of well clusters, with the intent to identify contaminant sources specific to one area. The work presented here aims to characterize groundwater quality on large spatial scales, and assess the role of water storage fluctuations on a more consistent temporal scale.

Total dissolved solid concentrations (TDS) are examined in this study for several reasons, many of which stem from the fact that it is a physical parameter. TDS is relatively easy to measure (e.g. compared to measuring total nitrogen), and is therefore expected to have a relatively large number of *in situ* observations. A sample of known volume is filtered for coarse material, dried at 180°C, and weighed as a dry mass per unit volume, typically reported in mg/L. TDS has a relatively straightforward mass balance, as there is a lack of potential physical, chemical, or biological transformations, making the methods developed here applicable to other constituents. The sources used in this work publishing TDS observations also publish measurements of other constituents of interest. TDS can encompass regionally-specific groundwater contaminants, as it is a physical parameter, but the relevance of TDS concentrations is not aquifer-specific, and so the methods detailed in this work can be applied to other regions.

The objective of this work is to explore paths to characterizing large-scale TDS concentrations. We first detail database compilation of *in situ* TDS concentrations and groundwater levels, explore approaches to scaling point and gridded data, and assess preliminary relationships between TDS and groundwater levels. Models to predict average TDS concentrations through time are constructed as a function of groundwater levels and season. We hypothesize that constructing models by classification of dominant lithology and dominant land use add characteristics of groundwater recharge which improve prediction of TDS, and that including

GRACE-derived subsurface storage anomalies as a potential explanatory variable improves accuracy of TDS predictions. Lastly, two methods of spatial analysis are explored, one to characterize TDS on subbasin scales, and the other to characterize regional TDS variability in space and time. We hypothesize that spatial patterns of TDS on subbasin and regional scales can be related to spatial patterns of GRACE-derived subsurface storage anomalies over the GRACE period. Methods of characterizing TDS on large scales developed in the following studies lay the groundwork for characterizing large-scale groundwater quality using regionally important constituents, and ultimately predicting fluctuations in groundwater quality using remote sensing.

Chapter 2

Database compilation and evaluating scaling approaches of well data

2.1 Introduction

Groundwater resources provide the primary source of freshwater for domestic and agricultural use for over 2 billion people globally (Alley et al. [2002]). The demand on groundwater is likely to increase with a growing population. Climate change, groundwater contamination, and natural variability in groundwater quality with depth further complicate projected demands on groundwater. Climate change is driving a redistribution of precipitation, a primary source of groundwater recharge, from mid-latitudes to high and low latitudes (Solomon [2007]), which will likely increase dependence on groundwater sources as surface water resources retreat in arid and semi-arid regions. Groundwater contamination by both natural and anthropogenic sources are the subjects of many studies (see section 1.1), but typically apply to local spatial scales and relatively short time periods. The heavy reliance on groundwater reserves and the complex nature of groundwater quality point to the importance of characterizing groundwater quantity and quality.

NASAs Gravity Recovery and Climate Experiment (GRACE) mission monitors monthly groundwater storage variations in some of the world's largest surface and groundwater

basins (Rodell and Famiglietti [2002], Swenson and Wahr [2006], Landerer and Swenson [2012], Swenson [2012]). Large-scale groundwater storage variations can be determined from GRACE and supplementary data at monthly and longer timescales for regions that are 150,000 km^2 and greater. Estimates of groundwater availability focus on quantity (Rodell and Famiglietti [2002], Rodell et al. [2009], Famiglietti et al. [2011], Gleeson et al. [2012], Scanlon et al. [2012a,b], Voss et al. [2013], Döll et al. [2014], Richey et al. [2015a,b]), but methods to infer groundwater quality are not yet developed, in part due to a lack of spatially representative quality data.

Groundwater quality observations are not as readily available as groundwater level observations. The limited availability of samples is likely rooted in the costs associated with field campaigns and sample analysis. Groundwater generally moves slowly relative to surface water, and so it is thought that groundwater quality changes also occur slowly. Where observations are available, the necessary metadata attributing location and date of sampling are not always provided and therefore cannot be incorporated in this analysis. To maximize the available groundwater quality data, we focus on total dissolved solid concentrations.

The total dissolved solid concentration (TDS) of a sample is the mass of dried material from a unit of volume, typically reported in milligrams per liter (mg/L). TDS is a commonly measured physical parameter used as a general description of water quality. The mass balance of TDS is relatively straightforward because there are inherently no chemical or physical changes. The significance of TDS is not regionally specific and is therefore an ideal test parameter for developing methods to characterize large-scale groundwater quality.

The High Plains aquifer has a total area of 450,000 km^2 under portions of eight states in the center of the United States, stretching from the southern edge of South Dakota to the panhandle of Texas. The High Plains aquifer is a primary freshwater source to a large agricultural sector as well as most rural and urban inhabitants (McMahon et al. [2007]). The Central Valley watershed has a total area of 160,000 km^2, which includes the underlying

aquifer having a total area of $52,000km^2$ (Faunt [2009], Famiglietti et al. [2011]). It supplies water for irrigation to a highly productive agricultural industry, and to a growing urban population (Faunt [2009]). The High Plains and Central Valley aquifers are among the largest and most heavily studied aquifers in the US. The large areal extents and relatively rich recorded histories make the two aquifers prime candidates for this work.

Satellite and regional model output products are improving and becoming unique tools to assess many earth system processes, particularly in the face of climate change (Solomon [2007], Taylor et al. [2012], Stocker et al. [2013]). Remote sensing and modeling products such as GRACE are invaluable, but output accuracy must be validated with real-world observations (Strassberg et al. [2007, 2009], Henry et al. [2011], Scanlon et al. [2012b]). Bridging spatial scales between *in situ* point observations and large-scale regional data is increasingly necessary as the latter becomes more readily available.

This chapter details database compilation, discusses scaling approaches, specifically using point observations and gridded products, and explores preliminary assessment of the compiled *in situ* database.

2.2 Data sources and pre-processing

2.2.1 *In situ* groundwater data

A major obstacle to modeling large-scale groundwater quality is the limited availability of spatially and temporally consistent *in situ* observations, as discussed in section 1.1. Spatially consistent data is necessary to ensure the entire study area is represented, and temporally consistent data is necessary to avoid signal aliasing. This section details data sources, database compilation, and pre-processing of ground-based observations to ensure quality control across sources and remove potential duplicate records while maintaining as much

spatial and temporal representation as possible in each study area.

Total dissolved solid concentrations and groundwater level observation sources include the USGS National Water Quality Assessment (NAWQA) Program (USGS NAWQA), the California Groundwater Ambient Monitoring Assessment (GAMA) program (California State Water Resources Control Board/Division of Water Quality), and the USGS Historical Water Quality Data for the High Plains Regional Ground-Water Study Area (Litke [2001]). Nondetect TDS values are included and prescribed to be the half of the detection limit of 10 mg/L (Hirsch et al. [1991], Lapham et al. [1995]). Supplementary groundwater levels are sourced from the USGS National Water Information System (NWIS) (USGS) and from the California Statewide Groundwater Elevation Monitoring (CASGEM) Program courtesy of Brian F. Thomas (California Department of Water Resources [a]). All data are downloadable online with the exception of Litke [2001], which is available on CD-ROM upon request.

Study area	Variable	No. sites	No. observ.	Date range
Central Valley	TDS	18,356	69,390	May 1923 - Jan 2015
	WL	40,127	1,230,770	May 1916 - Feb 2015
High Plains	TDS	11,155	18,130	Nov 1931 - May 2015
	WL	35,736	1,656,695	Jan 1905 - Aug 2015

Table 2.1: Summary of compiled database for the Central Valley and High Plains study areas. Note that some sites provide both TDS and water level (WL) data.

All *in situ* records include associated relevant metadata, namely well location coordinates, date of observation, site identification number, and measurement units. Data and metadata contain both numeric and alphanumeric information, downloaded as text or csv files. Some sources provide site metadata in a separate file, associated to individual observations by an alphanumeric site identification number. One row corresponds to one observation, and each column corresponds to the measurement and its metadata.

For each of the above *in situ* data sources, the relevant file is imported and associated with relevant metadata. Because the files are not strictly numeric, each file is read in by line. Observations are discarded if there is no value recorded for TDS and/or water level, the

measurement date is incomplete, or if outside the study area of interest. Each observation is associated with site metadata by a site identification number, if provided in separate files. If necessary, units are converted to be uniform across data sources (e.g. site coordinates to decimal degrees, observation dates to a uniform string order format, TDS to mg/L, water levels to cm below the land surface). This filtered data for each source contains the following columns: site name, site identification number, observation date, latitude, longitude, water level value, and TDS value. Placeholders are used where a water level or TDS value exists without the other. For each study area, all data from each source is combined.

(a) Central Valley   (b) High Plains
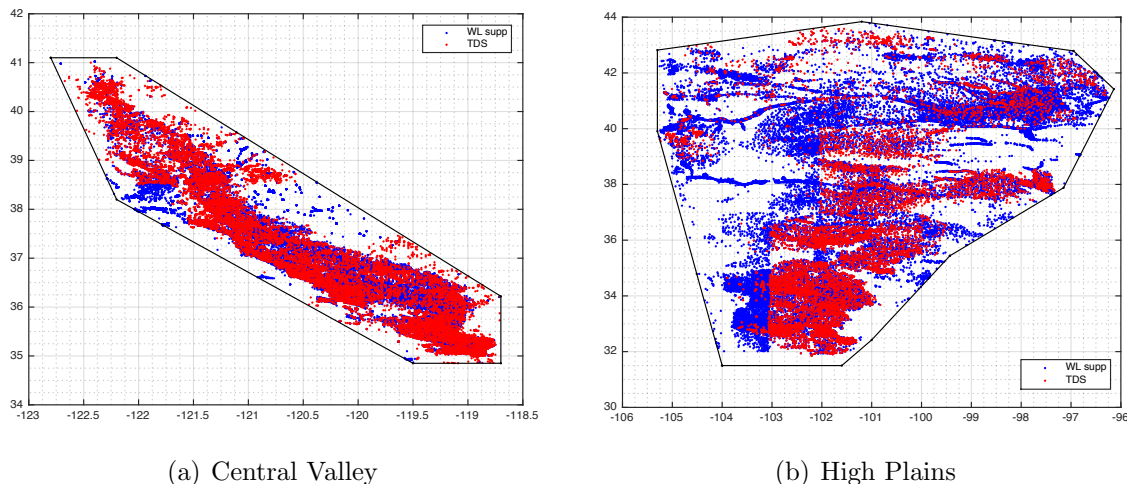
Figure 2.1: Locations of *in situ* well sites and bounding polygon defining area of interest in the Central Valley and High Plains. Sites providing supplementary groundwater levels are denoted in blue and sites providing total dissolved solid concentrations in red. Note that relative sizes of study areas are not to scale.

In each study area, the raw data from all sources are compiled preserving variable columns and row indices, and checking that the data format and reporting convention within each column are uniform. Water level observations are converted from feet below the land surface to centimeters below the land surface. Note that water levels are most frequently reported as positive values below the surface in the data sources above, so larger values denote deeper *in situ* water levels and smaller values denote shallower water levels. Sample date strings are converted into date numbers. Original date strings are maintained in a new column for
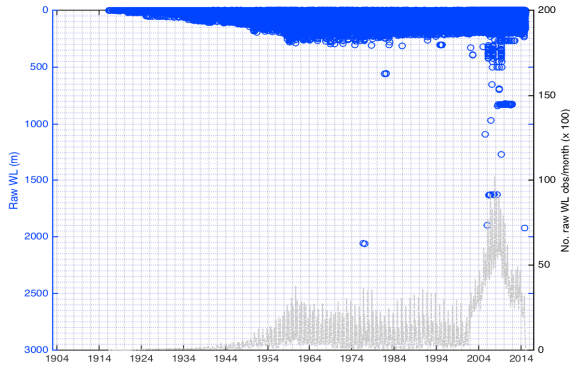
(a) Central Valley WL   (b) High Plains WL
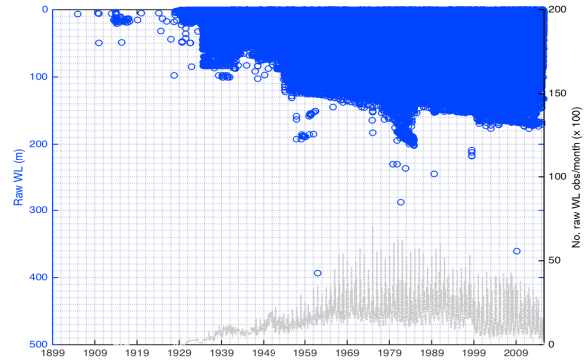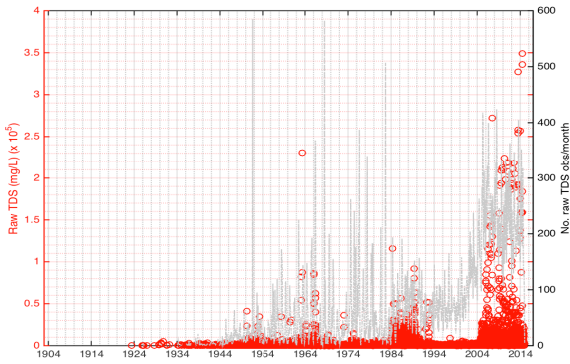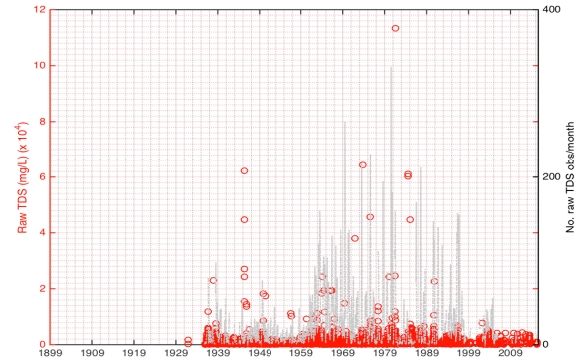
(c) Central Valley TDS   (d) High Plains TDS

Figure 2.2: Complete raw record of *in situ* groundwater levels (WL) in m below the land surface (2.2(a), 2.2(b)) and total dissolved solid concentrations (TDS) in mg/L (2.2(c), 2.2(d)) for the Central Valley (2.2(a), 2.2(c)) and the High Plains (2.2(b), 2.2(d)). Number of observations per calendar month for each variable are denoted in gray.

reference.

To define spatial boundaries for each study area, the smallest possible area enclosing the maximum number of well sites is determined by connecting as few vertices as possible into a convex polygon, as shown in Figure 2.1. Vertices are defined in decimal coordinates. Data falling outside of the polygon are removed if necessary. The convex polygon coordinates are stored for later use.

Unique pairs of well site identification number and observation date are identified to remove multiple or duplicate data. If there are more site identification, date pairs than unique

observation values, and the site coordinates are identical or in very close proximity (< 1 km), the mean of the values is taken. We then check for remaining multiples of site identification number, date pairs that have coordinates reported to a different number of decimal places, generally a minimum of 5. These cases are assumed to be taken from the same location, and are averaged. Finally, the observation dates are grouped by calendar month. Again, the original dates are preserved but set aside for reference if necessary. The resulting TDS and water level observations and associated metadata are referred to as the raw data set, summarized in Table 2.1 and locations shown in Figure 2.1.

2.2.2 GRACE and auxiliary data

GRACE CSR RL05 terrestrial water storage anomalies (dS) are obtained from NASA's TELLUS portal (Swenson and Wahr [2006], Landerer and Swenson [2012], Swenson [2012]). As described in section 1.2.1, the level 3 GRACE data is a $1° \times 1°$ gridded product of vertically integrated total water storage anomalies and accompanying CLM4 300 km scaling factor, scaled measurement error, and scaled leakage error. The scaling factor is applied to restore amplitude to the attenuated storage anomaly signal resulting from filtering and processing of gridded data (Landerer and Swenson [2012]). The Central Valley and High Plains aquifer study areas are each isolated from the global data by masking out data outside of the relevant study area using longitude and latitude coordinates. Values are reported in centimeters water equivalent height unless otherwise noted.

Gridded dS values are converted to subsurface storage anomalies ($dSubsurface$) using a rearranged water balance:

$$dSubsurface = dS - (dRR + dCan + dSWE) \tag{2.1}$$

where dRR is routed river storage anomalies, $dCan$ is canopy storage anomalies, and $dSWE$

26

is snow water equivalent anomalies. Snow water equivalent and canopy storage are from the GLDAS suite model outputs, specifically CLM, VIC, and NOAH $1° \times 1°$ monthly products. Monthly routed river storage from CLM4 at $0.5° \times 0.5°$ resolution, courtesy of Sasha Richey and MinHui Lo, are linearly interpolated to $1° \times 1°$ resolution to match GRACE and other GLDAS model outputs (Rodell et al. [2007], Perry and Niemann [2007], Scanlon et al. [2012b], Richey et al. [2015a,b]). Each storage variable from each model is converted from monthly absolute storage to monthly anomalies. Storage anomalies are then averaged across models for each variable, and used to calculate subsurface storage anomalies as in equation 2.1. Note that $dSubsurface$ includes both groundwater and soil moisture storage anomalies.

The error of $dSubsurface$ is equal to the square root of the sum of squared errors for each term in equation 2.1. The error for dS includes the scaled measurement errors, scaled leakage errors, and the error associated with correlation between neighboring grid cells for each masked basin (Swenson [2012], Richey et al. [2015a]). The error for routed river storage is assumed to be 50% (Richey et al. [2015a]). Canopy storage and snow water equivalent errors are calculated as the standard deviation of storage anomalies of the GLDAS suite (Richey et al. [2015a]).

Soil moisture is not readily measured in the US. Observations that are available generally exist on short-term and local spatial scales (Perry and Niemann [2007], Strassberg et al. [2009], Longuevergne et al. [2010]) and therefore are not representative at the GRACE scale. Previous studies removing soil moisture storage from GRACE often utilize modeled soil moisture (Rodell et al. [2009], Tiwari et al. [2009], Famiglietti et al. [2011], Henry et al. [2011], Scanlon et al. [2012b]). In this work, soil moisture storage is not explicitly removed to maintain a conservative estimate of subsurface storage anomalies because of the large increase in error associated with modeled soil moisture variability within and across models (Richey et al. [2015a,b]).

2.3 Scaling approaches
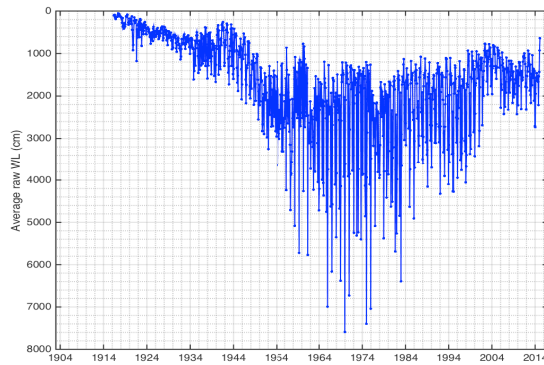
(a) Central Valley average raw WL
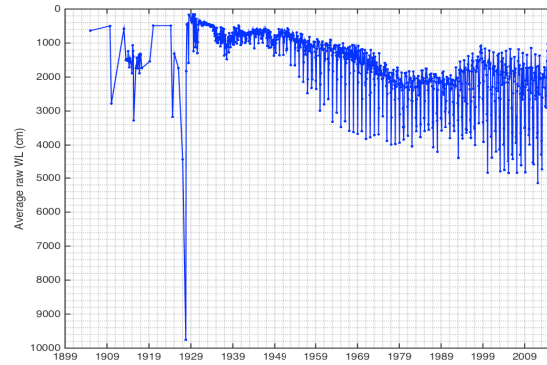
(b) High Plains average raw WL
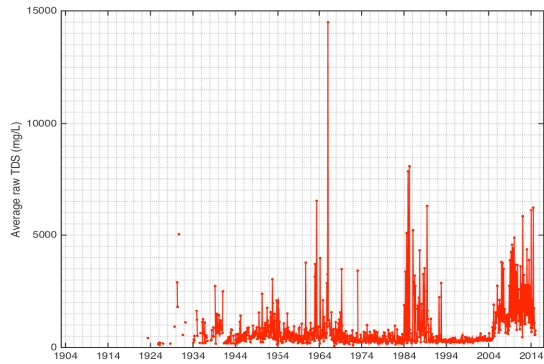
(c) Central Valley average raw TDS
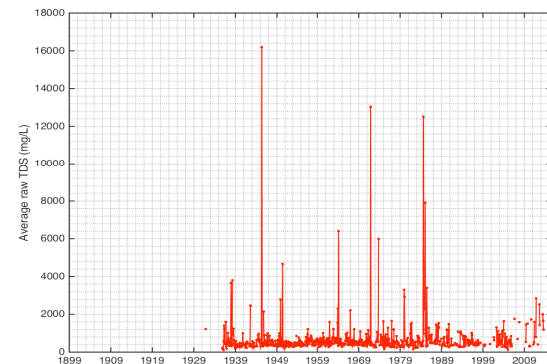
(d) High Plains average raw TDS

Figure 2.3: Monthly average of raw groundwater levels (WL) in cm below the land surface (2.3(a), 2.3(b)) and TDS in mg/L (2.3(c), 2.3(d)) for the Central Valley (left) and High Plains (right).

In this section we evaluate select common scaling approaches. The goal is to have both point observations and relatively coarse grid cells on a common spatial scale, necessary for combined analyses. The focus of this evaluation is on monthly point observations and monthly uniform $1° \times 1°$ grids matching the GRACE product resolution, but can be applied at other temporal and/or spatial scales.

Note that kriging of point data is intentionally excluded from this analysis. Accurate analysis and interpretation of kriging requires additional rigorous assumptions of the *in situ* data be met, depending on the type of kriging method. The most common versions are ordi-

nary and simple kriging. To accurately interpolate by kriging, the observed process must have some degree of stationarity (Tabios and Salas [1985], Cressie [1988, 1990]). Ordinary kriging assumes an unknown mean value is constant over the unsampled domain (Cressie [1988]). Simple kriging requires a known mean and assumes it to be stationary over the domain (Cressie [1988, 1990]). Groundwater TDS behavior on large-scale domains is not well-known, making it difficult to accurately interpret results of TDS kriging. In excluding kriging approaches, we aim to preserve the observed TDS recorded dynamics.

To gain some insight into the raw average monthly values, the mean of all raw observations within each calendar month is calculated to generate Figure 2.3. Note that this averaging does not consider any spatial areas or weighting. The averages are susceptible to outliers or extreme values, but can be used as a guideline for comparing basin averages from the spatial interpolation approaches.

2.3.1 Averaging points within $1° \times 1°$ grids

The first approach to represent a larger area with point observations is to average values falling into each grid cell as defined by the GRACE spatial resolution. This is straightforward to implement, and requires no further treatment of GRACE data. In each study area, ground-based water levels and TDS within a $1° \times 1°$ grid cell are compiled into monthly maps. All values of TDS and water levels are averaged within each GRACE $1° \times 1°$ grid cell for each calendar month. Groundwater constituent sampling at each well site, and therefore within a grid cell, is often temporally inconsistent for periods exceeding 2-3 months. To avoid data processing artifacts, interpolations to estimate a continuous monthly time series for point data are not performed in this relatively coarse upscaling.

The simplistic averaging method does not account for the spatial distribution of well sites in a grid cell for a given month with respect to the grid cell center or to other wells. Because
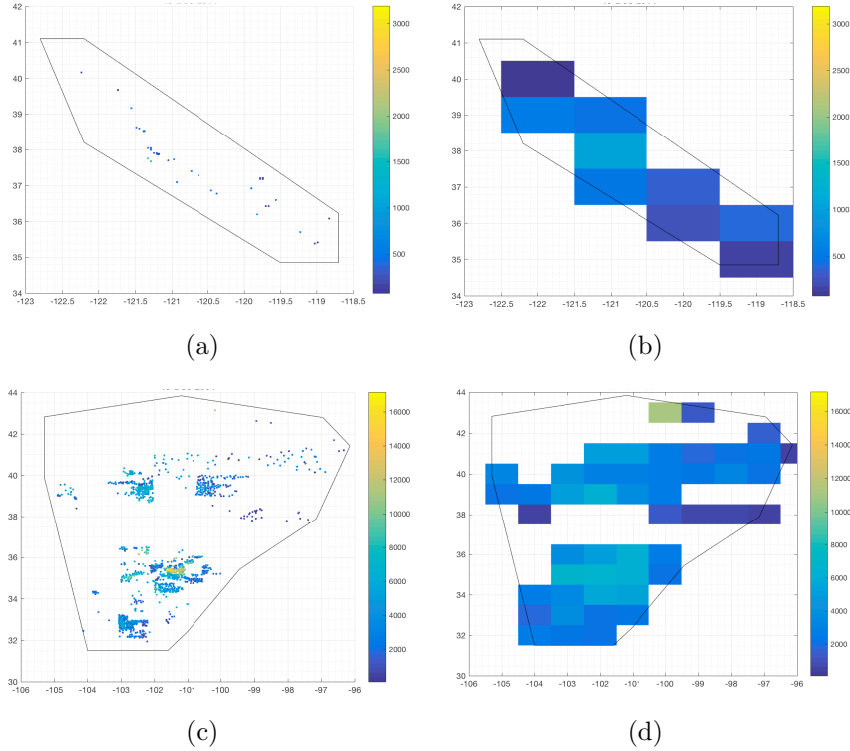
Figure 2.4: Examples showing *in situ* observations averaged within 1° × 1° grid cells for December 2014. TDS observations in mg/L are shown for the Central Valley (top). Groundwater level observations in cm below the land surface are shown for the High Plains (bottom).

the GRACE level 3 product processing employs Gaussian smoothing, giving more weight to *in situ* values nearer the grid cell center may improve agreement. An observation near the perimeter of a grid cell bears the same weight to that grid cell value as one at the center, with no impact on values to the nearest neighboring grid cell value. In reality, wells within or across grid cells may or may not be hydrologically connected. However, wells clustered in close proximity (few to tens of kilometers) can share a hydrologic link or at a minimum environmental setting characteristics.

Figure 2.4 shows two examples of one month (December 2014) of *in situ* observations and the resulting map when averaging by 1° × 1° grid cells, with TDS in the Central Valley and groundwater levels in the High Plains. Note that the example shown for Central Valley TDS during the GRACE period has one of the largest numbers of *in situ* observations within the

(a) Central Valley WL (b) High Plains WL

(c) Central Valley TDS (d) High Plains TDS
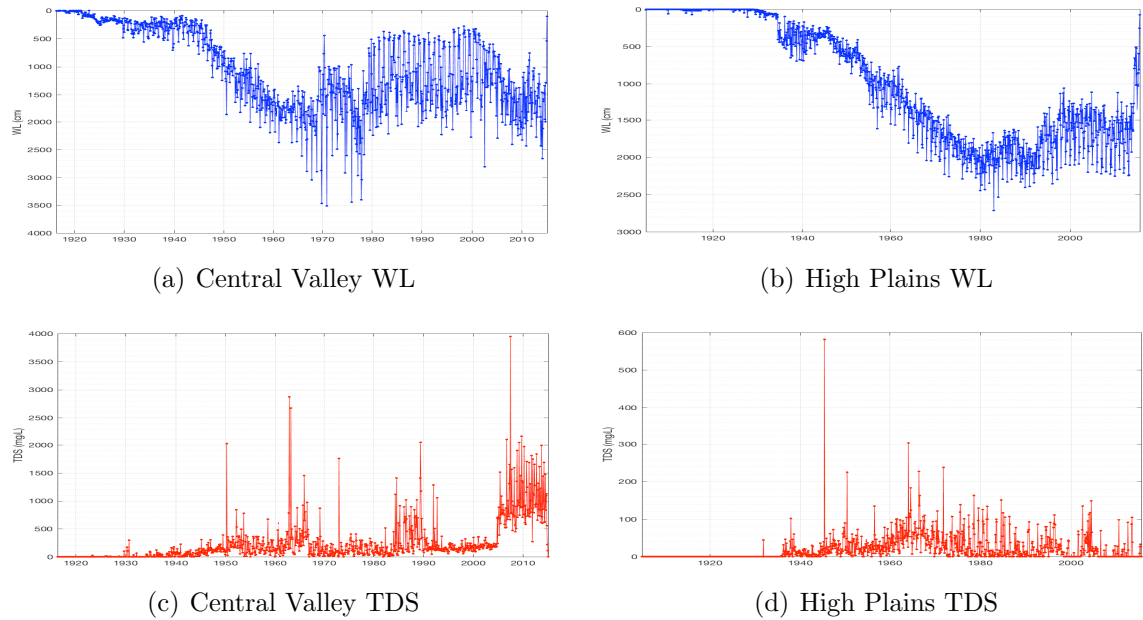
Figure 2.5: Monthly area-weighted basin average time series of groundwater levels in cm below the land surface (2.5(a), 2.5(b)) and TDS in mg/L (2.5(c), 2.5(d)) averaged within $1° \times 1°$ grid cells for the Central Valley (left) and High Plains (right).

compiled database (see Figure 2.2). The areal extent is generally preserved in both areas, as each monthly grid cell value is defined as the average of any well locations falling within said grid cell. However, the inconsistency of point observations from month to month and in spatial extent mean that this simple upscale approach is vulnerable to applying a single value to an entire grid cell, which can easily go unnoticed if not explicitly checked. The monthly basin average time series of upscaled groundwater levels and TDS for this method are noisy (Figure 2.5), and the relative magnitudes of peaks do not match the raw basin averages particularly well, likely due to the inconsistency of *in situ* observations in space and time.

This approach to upscaling is crude and susceptible to attributing a small number of values to the larger pixel resolution, particularly if ground-based observations are sparse in space and/or time. While the spatial extent is generally preserved in this upscaling method, the accurate representation of observed groundwater quality dynamic changes and associated

uncertainty is not likely preserved.

2.3.2 Linear interpolation to $1° \times 1°$ grids

(a) Central Valley WL

(b) High Plains WL

(c) Central Valley TDS
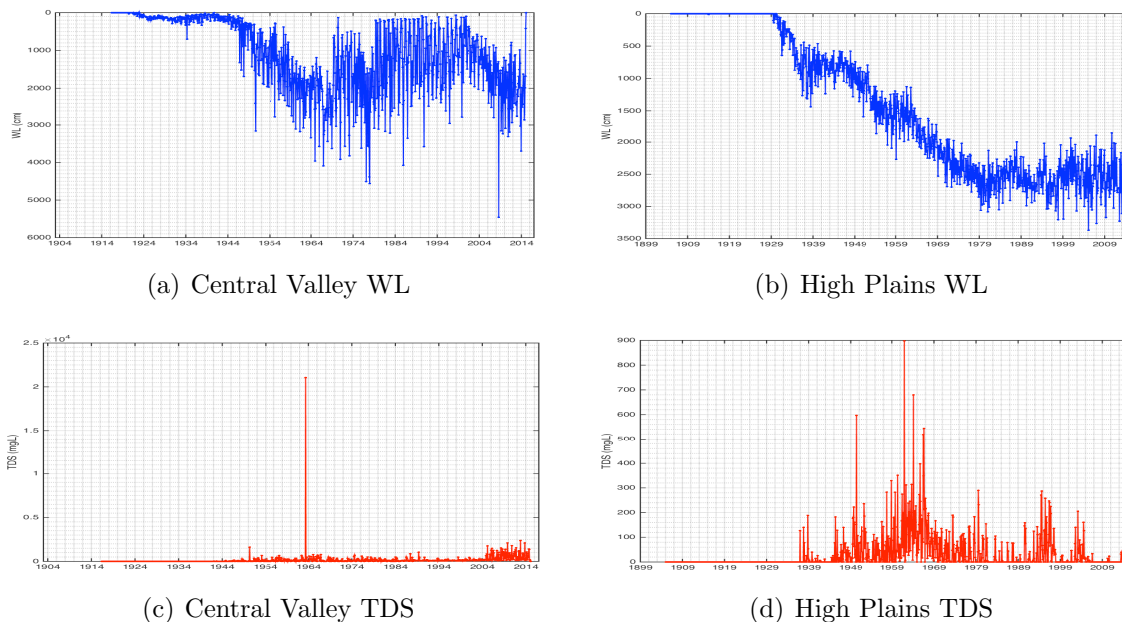
(d) High Plains TDS

Figure 2.6: Monthly area-weighted basin average time series of groundwater levels in cm below the land surface (2.6(a), 2.6(b)) and TDS in mg/L (2.6(c), 2.6(d)) using linear interpolation for the Central Valley (left) and High Plains (right).

The linear interpolation approach is perhaps the most common method of scaling between different grid resolutions (e.g. Helsel and Hirsch [2002], Rodell et al. [2007], Perry and Niemann [2007], Scanlon et al. [2012b], Harris et al. [2014], Richey et al. [2015a,b]). First, a uniform grid domain defining the query points is created to match the $1° \times 1°$ GRACE domain for the area of interest. Next, the well coordinates and data values (TDS or water level) of raw point observations are used to calculate an interpolant surface at each time step. The interpolant surface is then evaluated at the query points for each month, producing monthly gridded products of TDS and groundwater levels.

(a) Observed TDS, Dec. 2014.   (b) Linear interpolation.   (c) Inverse distance weighting.
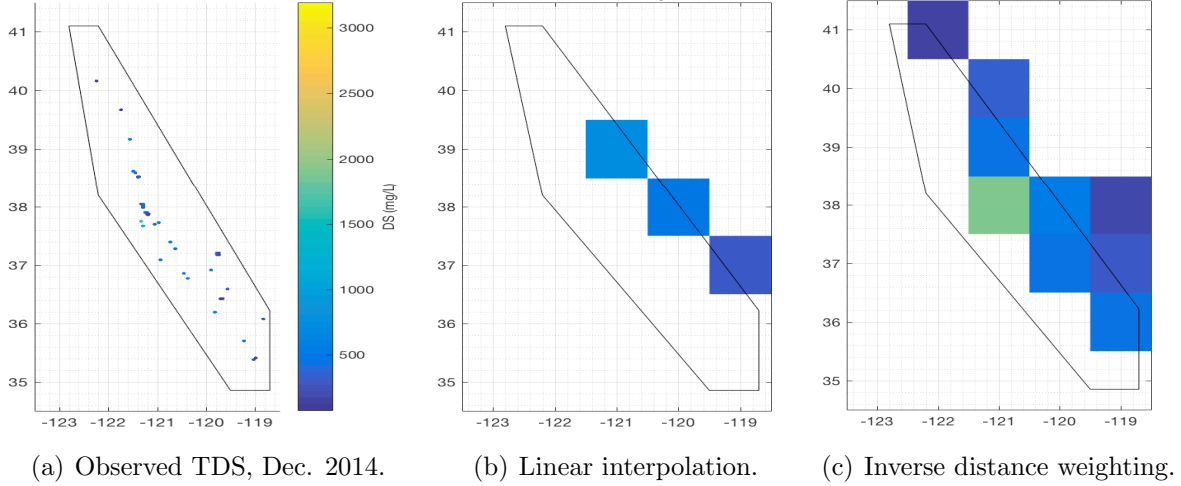
Figure 2.7: Locations of TDS point observations for December 2014 in the Central Valley (2.7(a)) upscaled to $1° \times 1°$ grids by linear interpolation (2.7(b)), and by inverse distance weighting (2.7(c)). TDS values are indicated by color in mg/L. Note the coordinate axes and color bar apply to all panels.

2.3.3   Inverse distance weighting interpolation to $1° \times 1°$ grids

Inverse distance weighting is another common interpolation method (e.g. Richey et al. [2016], Sahoo et al. [2016], Thomas et al. [2016]) that does just what the name suggests: weights observations by proximity to the desired interpolation point. The weighting function is equal to the reciprocal of the distance, d, between a known point, x_i, and the query point, x, raised to a power parameter, p, for u_i observations, as shown for one dimension in equations 2.2 and 2.3. More weight is given to observations closer to the interpolated point for values of p larger than the number of dimensions. In this analysis, 3 spatial dimensions and one temporal dimension are considered, thus a conservative value of $p = 5$ is applied. Large values of p produce an interpolated surface increasingly towards Thiessen polygons (Teegavarapu and Chandramouli [2005], Chang et al. [2006]), described in section 2.3.4.

$$u(x) = \frac{\Sigma(w_i \times u_i)}{\Sigma w_i} \tag{2.2}$$

(a) Central Valley WL

(b) High Plains WL

(c) Central Valley TDS
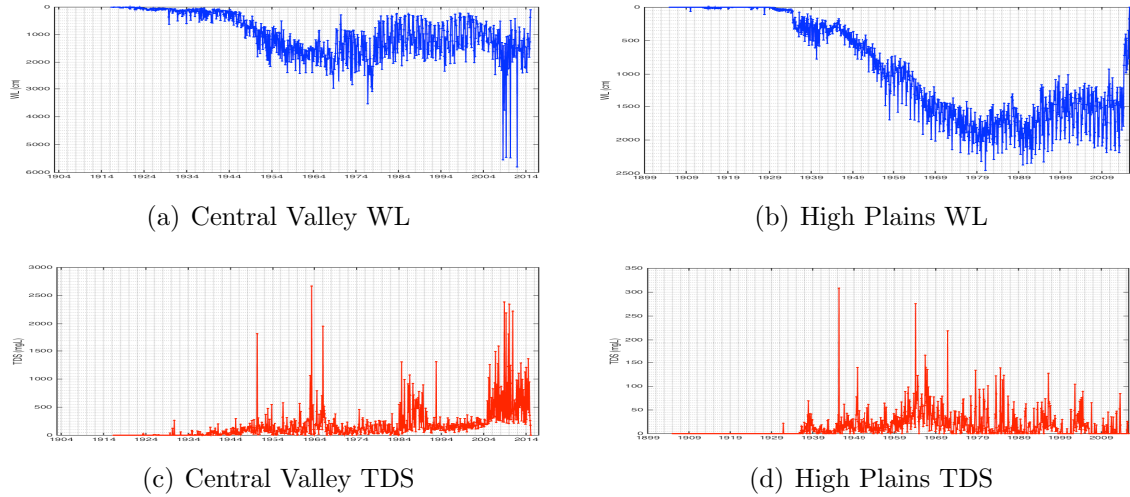
(d) High Plains TDS

Figure 2.8: Monthly area-weighted basin average time series of groundwater levels in cm below the land surface (2.8(a), 2.8(b)) and TDS in mg/L (2.8(c), 2.8(d)) using inverse distance weighting interpolation for the Central Valley (left) and High Plains (right).

$$w_i = \frac{1}{d(x, x_i)^p} \tag{2.3}$$

Similar to the linear interpolation approach, a query grid domain is first defined to match the $1° \times 1°$ GRACE domain for the area of interest. Next, the raw point observations occurring at each monthly time step are identified, and the data values are weighted based on the proximity of each well site to the query points. A fixed radius of $0.5°$ is used in this analysis to reduce the influence of values outside the bounds of each $1° \times 1°$ grid cell. Inverse distance weighting of points to a $1° \times 1°$ grid performs better than linear interpolation, retaining much of the areal extent, variability, and spatial pattern of the point observations, as shown for one month of TDS in the Central Valley in Figure 2.7 with basin averages for TDS and groundwater levels shown in Figure 2.8 for the Central Valley and High Plains.

(a) Central Valley	(b) High Plains
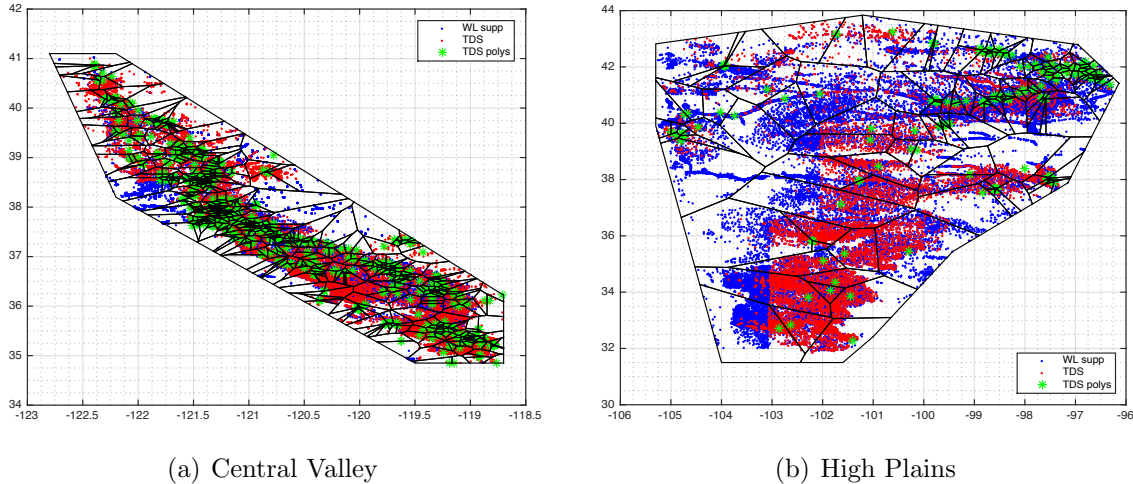
Figure 2.9: All well locations and Thiessen polygons in the Central Valley and High Plains. TDS well locations used to construct polygons are denoted by green markers. Note that relative sizes of study areas are not to scale.

2.3.4 Thiessen polygons

Thiessen polygons are a version of nearest-neighbor interpolation where the relative distances between a set of points determine delineated areas (Thiessen [1911]). Each dividing line is equidistant to two points, and each node is equidistant to 3 or more points. The partitioned areas can then be represented by the respective points. By name, Thiessen polygons are specific to meteorology and geophysics, but the delineations are also known as voronoi polygons, Dirichlet tessellations, and Wigner Seitz unit cells in other applications.

Thiessen polygons are constructed using locations of wells providing TDS observations at relative temporal consistency (Thiessen [1911], Rodell et al. [2007], Houborg et al. [2012]). From the full raw data for each study area, unique well sites are identified by site identification number and longitude and latitude coordinates. In instances of two sites having the same coordinates but different identification code names, the shorter of the two records is discarded. The number of observations through time for each location is determined. Each variable column is then distributed into a matrix organized by unique well site and date of observation. Three matrices of equal size are created, one storing values of each *in situ*

variable of interest (TDS and water level) and one storing the associated observation dates, where rows indicate a unique well site and columns indicate different observation dates. Placeholders are used where necessary. Each site row is sorted by observation date. Multiple observations at a well site within a calendar month are averaged, if necessary. A two-column matrix is created with rows corresponding to unique well sites storing longitude and latitude coordinates.

(a) Central Valley WL

(b) High Plains WL

(c) Central Valley TDS
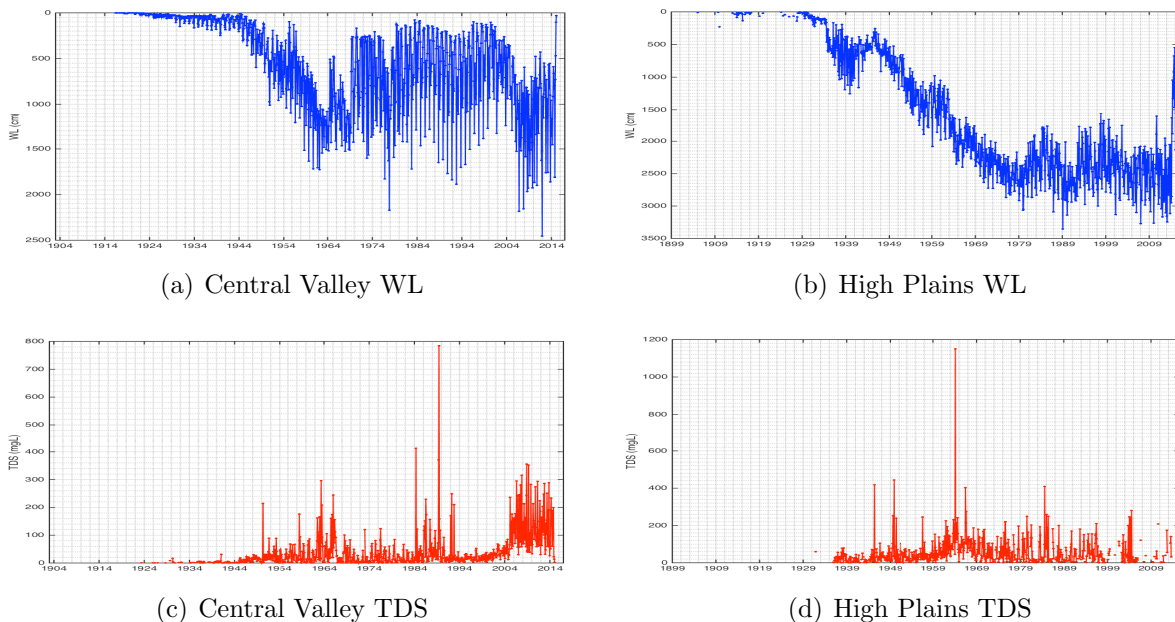
(d) High Plains TDS

Figure 2.10: Full monthly basin-average record of *in situ* groundwater levels (WL) in cm below the land surface and total dissolved solid concentrations (TDS) in mg/L in the Central Valley and High Plains calculated using Thiessen polygons.

Well sites providing at least 3 months of TDS observations and 4 or fewer months between successive TDS observations are identified. Sites with 3 or more observations are considered to represent active wells and thus more likely to represent long-term characteristics of groundwater constituent concentrations (Rodell et al. [2007], Kent and Landon [2013]). Sites with 4 or fewer consecutively missing months of observations are considered to represent seasonal variability (Rodell et al. [2007], Houborg et al. [2012]). At well sites meeting the above criteria, monthly TDS values are interpolated linearly up to 3 months (3 values), and are identified as locations around which Thiessen polygons are constructed. Exterior polygons

36
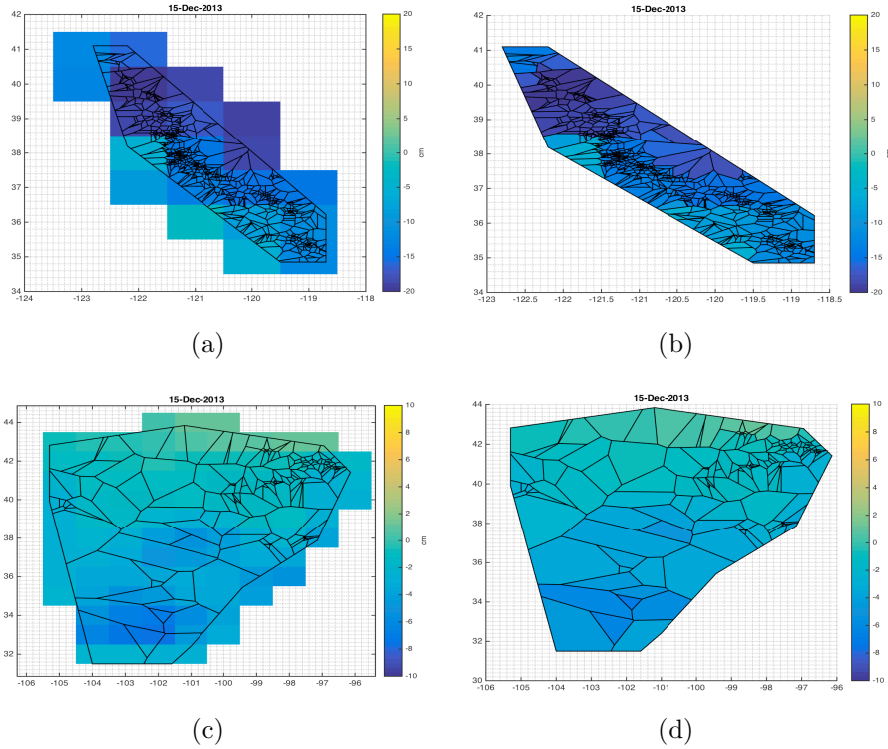
Figure 2.11: Examples showing gridded GRACE subsurface storage anomalies (cm) re-distributed into constructed polygons in the Central Valley and High Plains for December 2013.

are bounded by the lines connecting the convex vertices defining the area of interest, as described in section 2.2.1. The vertices of each constructed Thiessen polygon are stored and saved.

The area of individual Thiessen polygons are calculated as a fraction of the total enclosed area of interest. The total area of the study region is calculated using the surface area of the polygon vertices on a unit sphere with a radius equal to the average radius of the Earth, $6,371km$. The total areas are $96,489km^2$ and $750,429km^2$ for the Central Valley and High Plains, respectively. The area of individual Thiessen polygons are then calculated in km^2 using the fractional areas of the total. The values of the total area and individual polygon areas are stored and saved.

Constructed Thiessen polygons are chosen to be static through time to maintain spatial

(a) Central Valley dS.

(b) High Plains dS.

(c) Central Valley $dSubsurface$.

(d) High Plains $dSubsurface$.
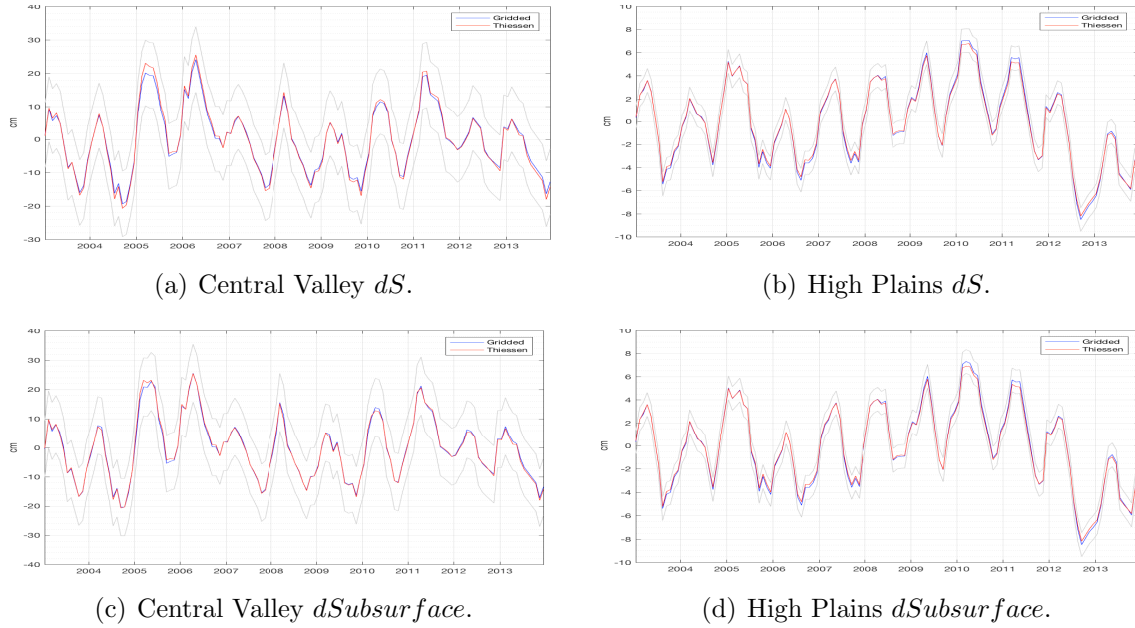
Figure 2.12: Monthly basin-averages of GRACE-derived dS and $dSubsurface$ in cm calculated by $1° \times 1°$ grids (blue) and by polygons (red) in the Central Valley and High Plains. Errors are indicated in gray.

representation of each study area in its entirety, whether or not a value is reported in each polygon at every time step. If internal polygons change based on available observations for each month, only fractions of each study area would be represented at a time, and it would be difficult to draw robust conclusions on areas of constantly varying locations and spatial extents.

TDS data from selected Thiessen polygon sites are temporally interpolated to fill in data for periods of up to 3 months. To avoid losing areal extent and drawing results from too small a database, all well records with at least 4 months of observations and any length of missing observation months supplement Thiessen polygon monthly averages, but are not interpolated through time. Well sites and constructed Thiessen polygons for the High Plains and the Central Valley are shown in Figure 2.9, where red and light blue markers indicate TDS and water level wells meeting the selection criteria, and green and dark blue markers indicate wells providing supplementary TDS and water level data.

38

Gridded GRACE terrestrial water storage and GRACE-derived subsurface storage anomalies are spatially redistributed into Thiessen polygons for comparison with *in situ* observations in the High Plains and the Central Valley study regions. The monthly GRACE-derived value of a polygon is calculated as an area-weighted average of the overlapping $1° \times 1°$ grid cells, as shown in Figure 2.11. Area-weighted basin average time series calculated using the polygon redistribution approach closely match time series calculated using the conventional gridded averaging approach in both study areas, as shown in Figure 2.12. The close match of the different averaging approaches indicates that the GRACE signal is not apparently perturbed in polygon redistribution processing on a basin scale.

2.4 Discussion

2.4.1 Database compilation and preliminary assessment

The ground-based observational records in the Central Valley and High Plains study areas begin long before the GRACE period. Data recorded before the satellite launch in 2002 cannot be used in an analysis with GRACE. However, historical data can provide context for the long-term TDS and groundwater levels in each study region. Ultimately, the goal of this work is to characterize TDS in groundwater as a function of explanatory variables such that average TDS can be predicted given a set of known variables. Although it is most likely that TDS variability cannot be explained by only one or two variables, namely time and groundwater levels, this section explores the potential of the compiled raw database to explain observed TDS.

Raw TDS and water level data in both the Central Valley (Figure 2.2(a), 2.2(c)) and the High Plains (Figure 2.2(b), 2.2(d)) begin roughly in the 1930s. The raw time series of TDS and water levels are noisy in both study areas, and it is difficult to determine whether the

apparent patterns are truly representative, or if they are an artifact of regular sampling schedules. Observation counts are low at the beginning of each record, and so are not likely to be representative of large areal extents. The number of observations increase by the 1950s.

The number of observations makes the raw data difficult to read, but the trends of the extremes are identifiable. From the beginning of the record in the Central Valley until roughly the 1960s, the deepest groundwater levels become deeper. The California State Water Project began surface water deliveries to the Central Valley for irrigation in 1960, which alleviated reliance on groundwater. California experienced record flooding late in 1955 to early 1956, which is reflected in the groundwater level record shown in Figure 2.2(a). TDS in the Central Valley appears to be highest in the 2000s, reaching larger maximum observed values than in the previous 60 years of the record, perhaps pointing to the salinization of soil in the San Joaquin Valley (Dubrovsky et al. [1993]). In the High Plains, in Figure 2.2(b), groundwater levels appear to steadily deepen for the majority of the observed record, particularly from 1954 to the end of the record in 2014. TDS in the High Plains appear to have high values throughout the record until the early 1990s, however it is unclear if TDS truly declines during this period or if it is an artifact of the number of TDS observations, which drop off drastically in the 2000s (Figure 2.2(d)).

The scatter plots in Figure 2.13 are to specifically examine potential relationships between raw TDS concentrations and groundwater levels. Note that only observations that have both a water level and TDS value are included, which total to 885 in the Central Valley between December 1949 and July 2014, and 1,527 in the High Plains between November 1931 and May 2015. If a linear increase in TDS with increasing depth is expected, as in a natural setting, then TDS is expected to increase as groundwater levels deepen. However, there is no such linearity in either Figure 2.13(a) or 2.13(b). The wide range of TDS values in shallow groundwater suggests loading at the surface for both regions, indicative of fertilizer application (McMahon et al. [2007]). There is no apparent long-term trend in the relationship

40

(a) Central Valley   (b) High Plains
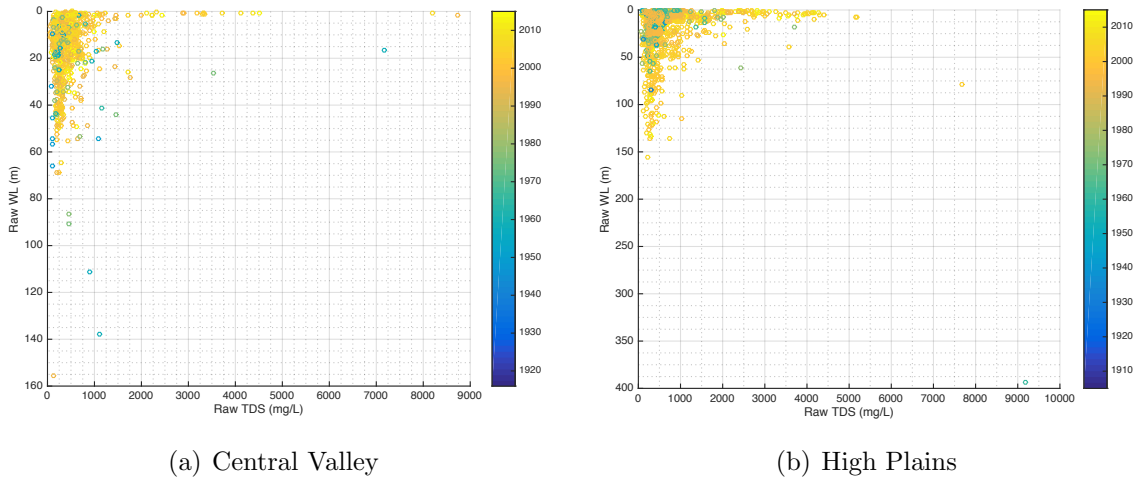
Figure 2.13: Scatter plots of raw TDS (mg/L) versus groundwater level (m below surface) for the Central Valley and High Plains. Colors denote year of observation. Note the difference in the y-axes.

between TDS and groundwater levels in either study area, as there is no discernible pattern by year, denoted by colors in Figure 2.13.

Preliminary regression on the raw ground-based data is performed to determine if any simplistic relationships can be quantified. While linear relationships are not expected, simple regressions may indicate directions of potential trends. To investigate simple relationships over the entire length of each *in situ* record, a linear regression is performed where time, as the number of days since January 0, 0000, is the only independent variable and TDS and water level are each a response variable. Then a linear regression is performed using groundwater levels as the sole predictor of TDS concentrations. As is often the case in regression, intercept values do not always carry physical meaning, but can not be excluded without meaningful justification.

In the Central Valley, groundwater levels became shallower on average over the full record, as denoted by the negative slope in equation 2.4. As a reminder, larger values of groundwater levels indicate deeper below the land surface. Although it is clear in Figure 2.2(a) that the deepest observations of groundwater levels get deeper until about 1960, there appears to be

some recovery beginning in the 1980s. The number of observations before 1980 are fewer than those after 1980, effectively making the average water level response after 1980 carry more weight in this simplistic regression. TDS concentration increases with time, having a positive slope associated with the time component, as shown in equation 2.5. Again, the larger number of observations in the more recent past are effectively weighted over the earlier portions of the record with fewer observations, and the extreme values are larger for the end of the record (Figure 2.2(c)).

$$WL_{CV} = -0.0394t + 3.0658e4 \tag{2.4}$$

$$TDS_{CV} = 0.0481t - 3.3903e4 \tag{2.5}$$

In the High Plains, the positive slope in equation 2.6 suggests groundwater levels deepened on average over the full record, albeit small. As previously discussed for Figure 2.2(b), the extreme water levels appear to steadily deepen beginning in the mid-1950s. The positive TDS slope with time in equation 2.7 suggests a very small increase in concentration over the full record.

$$WL_{HP} = 0.0749t - 5.2065e4 \tag{2.6}$$

$$TDS_{HP} = 0.0022t - 962.0253 \tag{2.7}$$

Raw TDS is estimated as a function of raw groundwater levels. Note again that only observations providing a value for both TDS and water level can be used here. The raw TDS

values are regressed using raw water levels as the predictor, shown in equations 2.8 and 2.9 for the Central Valley and High Plains. The respective Pearson's correlation values, R, of -0.1584 and -0.0356 are small and further support the hypothesis that groundwater levels alone can not explain the variability of TDS, as expected. The negative slope values of groundwater levels support the scatter plots in Figure 2.13 suggesting TDS loading in shallow groundwater.

$$TDS_{CV} = -0.0674WL_{CV} + 608.6928 \tag{2.8}$$

$$TDS_{HP} = -0.0103WL_{HP} + 690.4019 \tag{2.9}$$

Monthly box-and-whisker plots of the raw data in Figure 2.14(a) and 2.14(c) suggest seasonal cycles of groundwater levels. In the High Plains, median groundwater levels gradually shallow during the summer monsoon months from March to October, as shown by smaller values in Figure 2.14(c). This is expected, as the summer monsoon precipitation is the main source of recharge to the aquifer (McMahon et al. [2007]). However, there is no apparent seasonality of TDS in Figure 2.14(d), nor is there a suggestion that TDS is directly influenced by the groundwater level seasonal cycle.

In the Central Valley, median groundwater levels beneath the surface become shallower between March and October, as denoted in Figure 2.14(a). While counter-intuitive for an arid/semi-arid climate characterized by warm and dry summers, the timing of this seasonality likely reflects the availability of surface water from spring snow melt and deliveries to the Central Valley. Snow melt from the Sierra Nevada mountains runs into the Valley through natural river systems beginning in March or April (Miller et al. [2003]). Surface water deliveries via concrete-lined aqueducts generally peak in volume between June and August

(a) Central Valley WL

(b) Central Valley TDS

(c) High Plains WL

(d) High Plains TDS
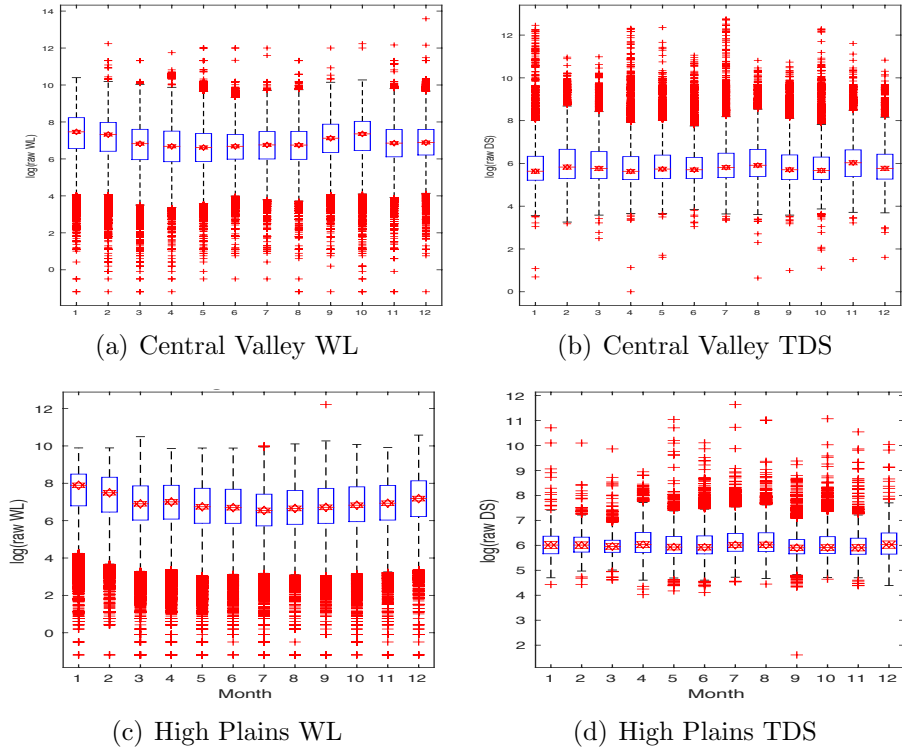
Figure 2.14: Quartiles of natural log of raw *in situ* groundwater levels (WL) in cm below the surface and total dissolved solid concentrations (TDS) in mg/L grouped by calendar month. The median of each month is denoted by the red line in the middle of each box, with the top and bottom of each box indicating the 25th and 75th percentiles, respectively. Values more than 1.5 times the interquartile range are denoted by red markers.

and decline sharply in September and October (California Department of Water Resources [b]). California water laws currently determine surface water allocation volumes based on use in the previous year, essentially requiring end users to use as much of their allotment as possible each year to ensure a comparable amount in the future (Shupe et al. [1989]). When surface water is readily available during the spring and summer months, users in the Central Valley rely less on groundwater. When surface deliveries decline in late summer and fall, a period also characterized by relatively low precipitation, users turn to groundwater to meet water needs. As in the High Plains, there is no apparent seasonal cycle of TDS concentrations in the Central Valley or indication that TDS is directly related to the seasonality of groundwater levels.

The *in situ* dataset compiled here provides a long-term picture of total dissolved solid concentrations in groundwater in the Central Valley and High Plains aquifers. Preliminary exploratory data analysis suggests a long-term decline in groundwater levels in the Central Valley (Figure 2.2(a)) and a long-term slight recovery in the High Plains (Figure 2.2(b)). Seasonality of raw groundwater level data are subtle without further noise reduction, but do indicate responses to surface water deliveries and/or precipitation (Figures 2.14(a) and 2.14(c)). Total dissolved solid observations are much more limited in number, and the variability is not as easily explained by known surface water processes. Preliminary analysis suggests a long-term increase in groundwater TDS over the full record in both study areas (Figures 2.2(c) and 2.2(d)), but there is no apparent seasonality or direct response to fluctuations in groundwater levels because both systems are likely well-mixed rather than stratified as expected in a natural system (Figures 2.14(b) and 2.14(d), equations 2.8 and 2.9). As expected, TDS concentrations can not be characterized by time and groundwater levels alone, particularly in regions with high agricultural productivity that may be susceptible to surface loading potentially from fertilizer use (Figure 2.13).

2.4.2   Spatial scaling

In both the Central Valley and High Plains, the magnitude of average water level depth is dampened using linear interpolation and inverse distance weighting compared to the Thiessen polygons. The dampening is likely due to the area weighting of the linearly interpolated basin time series. It is expected that the magnitude of area weighted average values differ from the rough raw monthly average, as the latter has no consideration of areas. More importantly, patterns of average groundwater levels over time appear to be represented by linear interpolation, inverse distance weighting, and Thiessen polygon scaling approaches. The long-term patterns in groundwater levels seen in the raw data for both study areas (Figure 2.2(a), 2.2(b)) are identifiable in the basin averages for each method. The Kendall's

tau rank correlation values are calculated for each monthly basin average against the raw monthly average to compare how well the patterns are captured by each method, shown in Table 2.4.2. In both locations, inverse distance weighting results in the highest correlation, closely followed by Thiessen polygon scaling, with linear interpolation having the lowest correlation for groundwater levels. The large number of groundwater level observations, particularly at the end of each record in both basins is an obvious advantage to all scaling approaches. Groundwater level observations in both study areas are an order of magnitude or two larger than the TDS counterparts.

As with groundwater levels, the relative magnitudes of TDS within each basin are dampened compared to the raw monthly averages. The overall patterns of the raw data are not reflected in the linear interpolation for either study area likely due to too few observations for accurate spatial representation in a $1° \times 1°$ grid cell, with tau correlations of 0.2755 and 0.1599 in the Central Valley and High Plains, respectively. Average TDS patterns are better captured by inverse distance weighting and Thiessen polygon approaches. In the High Plains, inverse distance weighting results in only slightly larger correlation values than the Thiessen polygons, at 0.3330 versus 0.3160. It is worth noting that the end of the High Plains raw TDS record becomes sparse in the early 2000s, which is reflected in the polygon basin average but not in the inverse distance weighting. In the Central Valley, the Thiessen polygon method results in a very slightly larger rank correlation of 0.4492 compared to inverse distance weighting at 0.4452. The relative sizes of TDS concentrations by Thiessen polygons appear to more closely match the raw monthly average than inverse distance weighting, such as the increase between 1984 and 1994 relative to the increase between 2004 and 2014. Inverse distance weighting is also slightly noisier in the Central Valley than Thiessen polygons, for example during the period of relatively low concentrations between 1994 and 2004.

Study area	Variable	Linear	IDW	Thiessen
Central Valley	TDS	0.2755	0.4453	0.4492
	WL	0.5806	0.6359	0.6100
High Plains	TDS	0.1599	0.3330	0.3160
	WL	0.5947	0.6409	0.6347

Table 2.2: Kendall's Tau rank correlations of monthly raw mean values against area weighted basin averages calculated by linear interpolation, inverse distance weighting, and Thiessen polygon scaling approaches. All values are statistically significant at the 1% level.

Linear interpolation of points to $1° \times 1°$ grid cells does not perform well compared to inverse distance weighting interpolation. The example in Figure 2.7(a) shows observed TDS in December 2014, a month with relatively large sample size and spatial extent of TDS, reaching from approximately $35.5°$ N to over $40°$ N, and from $-119°$ W to over $-122°$ W. After linear interpolation, the original areal extent is reduced to only three $1° \times 1°$ grid cells between $36.5°$ to $39.5°$ and $-118.5°$ to $-121.5°$. The variability of TDS values is dampened and the pattern of values is misconstrued using linear interpolation. In the example shown in Figure 2.7(a) for the Central Valley, *in situ* point values range from approximately 200 mg/L in the northern and southern most locations to 2,000 mg/L around the middle-west ($-121.3°$ W, $37.75°$ N), with one observation reaching over 3,000 mg/L. Linear interpolation reduces the range of values for December 2014 to be between roughly 250 mg/L and 1,250 mg/L at best. The pattern of raw TDS values for this particular month show a general increase in TDS from the northeastern valley boundary towards the southwestern boundary which is reflected by the inverse distance weighting pattern, whereas the linearly interpolated pattern instead suggests TDS increases towards the northwest.

Inverse distance weighting is the most successful approach to spatially scaling points to $1° \times 1°$ grid cells. Taking into account the proximity of point observations to the desired grid cell

query point is clearly advantageous to preserving as much of the original point information as possible. Unlike linear interpolation, the inverse distance weighted interpolation grid cells match the spatial extent of the point observation locations as well as maintaining an accurate representation of the observed spatial pattern (Figure 2.7).

(a) Central Valley WL.

(b) High Plains WL.

(c) Central Valley TDS.

(d) High Plains TDS.
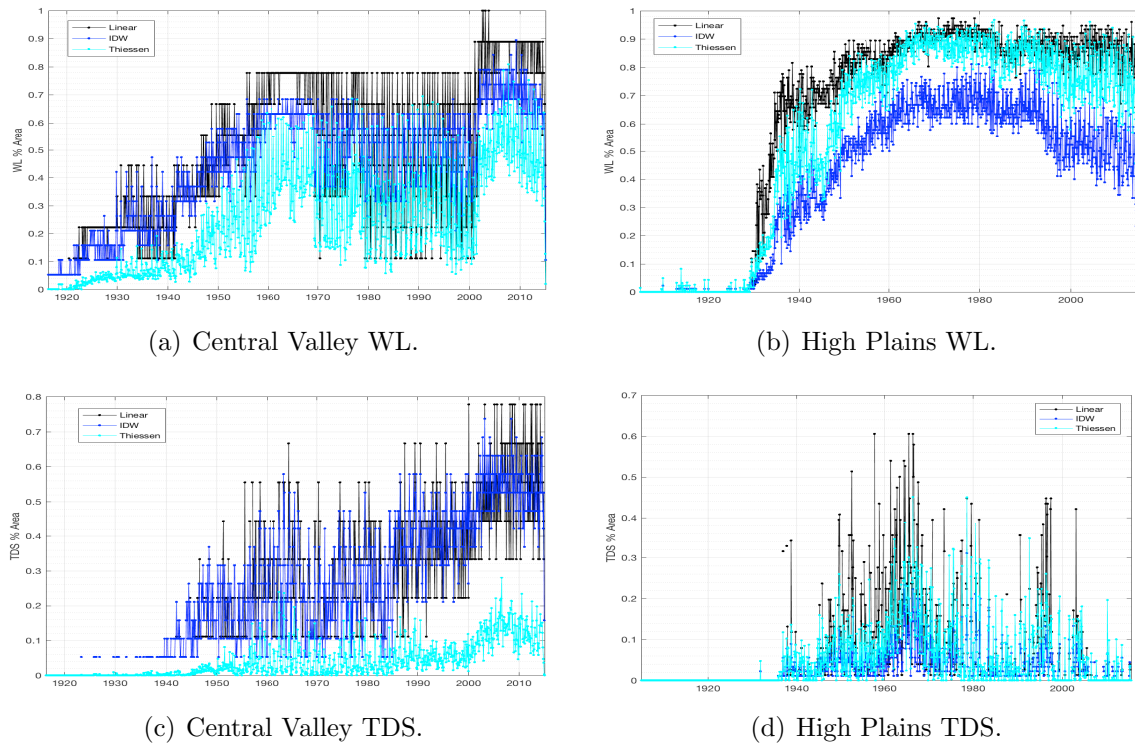
Figure 2.15: Percent of total area represented at each monthly time step by linear interpolation, inverse distance weighting, and Thiessen polygon scaling approaches.

Thiessen polygons constructed around selected TDS well sites are temporally static to force consistent areal representation of each region. As with scaling to $1° \times 1°$ grid cells, months with no observations in a polygon do occur. Individual polygons are represented by monthly time series, providing insight to average TDS concentrations in groundwater and average groundwater levels based on all data available for each month.

Applying constructed Thiessen polygons to the compiled database can be used to identify patterns of groundwater levels and TDS concentrations. For example, Figure 2.16 shows the median TDS concentration for individual polygons in each study area. Sub-regions

with higher median TDS can be identified and potentially attributed to a specific event or phenomenon, such as the soil salinization in the San Joaquin Valley in the southern half of the Central Valley (Dubrovsky et al. [1993]).

(a) Central Valley
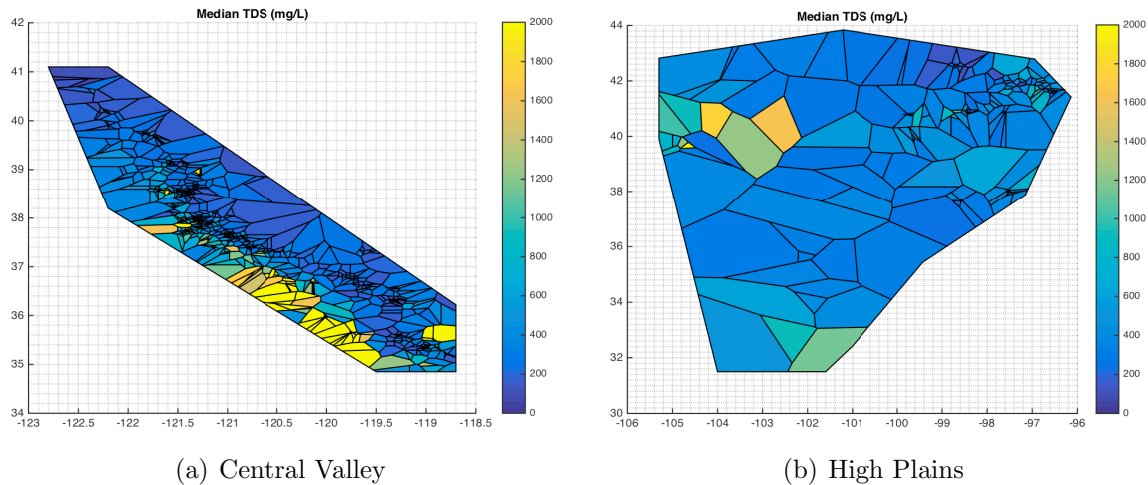
(b) High Plains

Figure 2.16: Median TDS (mg/L) by Thiessen polygon for the Central Valley and High Plains.

Individual Thiessen polygons can be used to study ground-based observations for smaller areas. In most cases, individual polygons are smaller than $1° \times 1°$ grid cells, the latter of which are on too large a scale to have meaningful implications from a local water municipality standpoint. Polygons constructed here represent spatial scales closer to cities or counties and may be a useful local groundwater management tool. For example, the monthly average groundwater level and TDS concentration for a polygon can be generated to study the long-term record.

2.5   Conclusions

Although imperfect, lacking comprehensive metadata such as screened well depth and hydrogeologic formation, the compiled data set is as complete a compilation of TDS in groundwater as is publicly available from several national and state agencies all following the same

sampling protocol. Steps are taken to ensure duplicate information is removed, and that all information across sources is reported in uniform format or units. This database is the basis from which we aim to characterize responses to various hydroclimatic conditions (e.g. droughts, wet periods, land use).

The long-term *in situ* data set provides context of large-scale TDS concentrations in groundwater and groundwater levels leading up to and during the GRACE period in the Central Valley and High Plains, but using GRACE to characterize TDS likely requires supplemental hydroclimatic and/or hydrogeological information. Future work using data from the planned GRACE follow-on mission (GRACE-FO) could potentially provide additional records necessary to characterize long-term TDS responses to groundwater variability.

Of the scaling approaches attempted, the inverse distance weighting and Thiessen polygons methods are best able to represent the variability, spatial extent, and pattern of available *in situ* observations. TDS patterns and variability are best represented by these two spatial interpolation methods, suggesting the observed TDS dynamics are likely preserved. To characterize large-scale TDS, the two following chapters will focus on using Thiessen polygons and inverse distance weighting to bridge the disparity of spatial resolutions of *in situ* observations and $1° \times 1°$ GRACE grid cells.

Chapter 3

Lithology and land use regime models using GRACE time-variable gravity data

3.1 Introduction

California's Central Valley aquifer extends through the middle of the state bounded by mountain ranges. In the arid/semi-arid climate, surface water resources can be limited due to low precipitation, a median of 13 inches between 1968 and 2003, which generally falls between October and March (Faunt [2009]). Water users in the Valley rely heavily on a complex system of surface water diversions and Sierra Nevada melt water controlled by local, state, and federal institutions. In dry years, the underlying aquifer is pumped to meet the agricultural and urban needs. The water table is in steady decline as a result of decades of virtually unrestricted groundwater pumping (Faunt [2009], Famiglietti et al. [2011], Scanlon et al. [2012b]).

The High Plains aquifer lies under 8 states in the central US. Precipitation reaches a maximum during late spring and early summer monsoons, with a mean annual precipitation reaching approximately 16 inches in the West, and 28 inches in the East (McMahon et al. [2007], Gurdak et al. [2009], McGuire et al. [2012]). Beginning around 1950, groundwater pumping for irrigation reduced storage and saturated thickness in the High Plains (Rodell

and Famiglietti [2002], Strassberg et al. [2007], McGuire et al. [2012]).

Agriculture production in both study areas rely on groundwater as a source for irrigation (Rodell and Famiglietti [2002], Strassberg et al. [2007], Faunt [2009], Famiglietti et al. [2011], Scanlon et al. [2012b], McGuire et al. [2012]). While studies have investigated instances of groundwater contamination in the Central Valley (Lindsey and Rupert [2012]) and the High Plains (McMahon et al. [2007], Gurdak et al. [2009], Lindsey and Rupert [2012], McGuire et al. [2012]), it is currently unclear if or how the decline of the water table affects the quality of water produced by the aquifer. The quality of freshwater produced has implications on the accessibility and cost of use, should treatment be necessary.

A common metric to assess freshwater quality is the total dissolved solid concentration. Total dissolved solid concentrations (TDS) are relatively easy to measure (e.g. compared to measuring total nitrogen), and therefore expected to have a relatively large number of *in situ* observations. A sample of known volume is filtered for coarse material, dried at 180°C, and weighed as a dry mass per unit volume, typically reported in mg/L. Because there are no physical or chemical transformations, TDS has a relatively straightforward mass balance making the methods developed here a potential template to apply to other constituents. TDS can encompass regionally-specific groundwater contaminants, but the relevance of TDS concentrations is not aquifer-specific, and so the methods detailed in this work can be applied to other regions.

In order to use point *in situ* observations and relatively coarse $1° \times 1°$ gridded GRACE data in one analysis, the spatial resolution disparity must be addressed. Thiesesn polygons are a nearest-neighbor weighting method for defining areas of interest, as described in the previous chapter (Thiessen [1911], section 2.3.4). In an attempt to keep processing errors low for *in situ* observations, criteria are defined for selecting well sites around which Thiessen polygons are constructed. The constructed polygons are considered to be units representative of an area larger than point observations and smaller than $1° \times 1°$ grid cells.

Lithology and land use cover affect properties of groundwater recharge, and therefore mechanisms of constituent infiltration. Lithology acts as a measure of texture affecting recharge rates as porosity and specific retention in the soil matrix (e.g. clay versus sand). Land use cover types also affect rates and frequencies of recharge (Charles et al. [1993]), and can indicate whether anthropogenic constituents should be anticipated (e.g. grassland versus developed). Data maps of lithology and land use cover are available at relatively high spatial resolutions, but are generally a snapshot in time. We hypothesize using lithology and land use cover types, physical characteristics controlling recharge and likely contaminant infiltration, to define categories is useful for comparing differences in the average effects on recharge of constituents to groundwater.

Collecting and measuring data *in situ* can be expensive and time consuming, and it is difficult to accurately represent large spatial extents. Recent groundwater availability studies estimate large-scale aquifer depletion rates and aquifer stress using monthly water storage variations from NASA's Gravity Recovery and Climate Experiment (GRACE) mission (Famiglietti et al. [2011], Scanlon et al. [2012a], Richey et al. [2015a,b]). To further evaluate available groundwater resources, assessing potability of groundwater is necessary. This is especially true when assessing groundwater stress (Richey et al. [2015b]) which accounts for availability. In Richey et al. [2015b], availability is represented by groundwater recharge flux. In reality, availability is a function of the recharge flux and groundwater storage of freshwater that meets water quality criteria for its proposed use. We hypothesize that GRACE as an additional predictor variable significantly improves models characterizing average TDS concentrations in groundwater on relatively large spatial scales.

3.2 Data

3.2.1 *In situ* water level, total dissolved solid observations

In situ water level and total dissolved solid (TDS) observations are compiled from well records in the Central Valley and the High Plains study area, as described in section 2.2.1. In the Central Valley, water level below the land surface and TDS observations are compiled from the USGS National Water Quality Assessment (NAWQA) Program and California's Department of Water Resources Groundwater Ambient Monitoring Assessment (GAMA) program. In the High Plains, *in situ* observations are compiled from the USGS National Water Quality Assessment (NAWQA) Program and the USGS Historical Water Quality Data for the High Plains Regional Ground-Water Study Area (Litke [2001]). In both study areas, supplementary water levels are compiled from the USGS National Water Information System (NWIS).

Water level records are more readily available in both space and time than TDS records in the Central Valley and High Plains study areas. In general, more TDS data are available in the Central Valley than in the High Plains, especially during the GRACE period.

3.2.2 GIS maps: Lithology and land cover

Conterminous US geology classification maps are available in digitized form (Ludington et al. [2007], Stoeser et al. [2005]), compiled by state by the USGS Mineral Resources Program ($http://minerals.usgs.gov/$). The digitized maps have spatial resolutions ranging from 1:100,000 to 1:1,000,000, based on the resolution of the original map. Information regarding geologic formations and parent material are useful to identifying expected natural constituent species dissolved in groundwater. However, TDS encompasses all dissolved material natural and anthropogenic. Thus the intended use of digitized geologic maps in this work focus on

identifying the dominant lithology. The dominant lithology type describes physical charac-
teristics of the rock unit in an area, such as texture and grain size, which have implications
on groundwater recharge rates and therefore infiltration of constituents.

The USGS National Land Cover Database (NLCD) provides land cover maps at 30 meter
spatial resolution for 2001, 2006, and 2011 (Homer et al. [2015]). The NLCD maps provide
the spatial extent and distribution of 16 discrete land cover categories over the contiguous
US. For the purposes of this work, only 2011 is used. The 16 categories are consolidated to
9 broader groups: water and ice, developed, barren, forest, shrub and scrub, grassland and
herbaceous, pasture and hay, cultivated crops, and wetlands.

3.2.3   GRACE and auxiliary data

Gridded GRACE CSR RL05 terrestrial water storage anomalies (dS) are obtained from
NASA's TELLUS portal (Rodell and Famiglietti [2002], Swenson and Wahr [2006], Landerer
and Swenson [2012], Swenson [2012]) and processed as in section 2.2.2. Subsurface storage
anomalies ($dSubsurface$) are defined as the sum of soil moisture and groundwater storage
anomalies (Richey et al. [2015a,b]), derived from total water storage anomalies by subtracting
routed river, canopy, and snow water equivalent storage anomalies from modeled output
(see section 2.2.2). Gridded subsurface storage anomalies are redistributed into constructed
Thiessen polygons as described in section 2.3.4 (see Figures 2.11, 2.12).

Specific yield is the ratio of the volume of water drained by gravity (V_g) to the total volume of
water (V_{total}) in porous media, and is related to soil porosity (n) and specific retention (S_r), as
in equation (3.1). Specific yield can be used to convert between groundwater storage changes
(ΔGWS) and water table changes (ΔWT) over a known areal extent (A). Described further
in section 3.3.3, $dSubsurface$ is divided by specific yield. Maps of specific yield produced

by USGS studies are available online (Faunt [2009], McGuire et al. [2012]).

$$S_y = \frac{V_g}{V_{total}} = n - S_r \tag{3.1}$$

$$\Delta GWS = A \times S_y \times \Delta WT \tag{3.2}$$

3.3   Methods

3.3.1   *In situ* point observations

As described in sections 2.2.1 and 2.3.4, compiled well data are processed by well site co-ordinate pair and code identification name. Well sites providing TDS data with 4 or more collective months of observations, and 3 or fewer consecutive months of absent observations are selected to construct temporally static Thiessen polygons, shown in Figure 2.9. Well sites having 4 or more collective months of observations are considered to represent active wells and thus more likely to represent long-term characteristics of groundwater constituent concentrations. Sites having 3 or fewer consecutive months of absent observations are considered to represent seasonal variability (Rodell et al. [2007]). Sites with fewer than 4 months of observations and/or 3 or more consecutively missing months of observations are used to supplement monthly polygon time series.

Monthly water level and TDS time series are calculated for each Thiessen polygon by averaging all *in situ* observations occurring within each calendar month. If no observations in a month are available in a polygon, the break in the record is reflected in that polygon time series as a missing value. Area weighted average timeseries for the Central Valley and High Plains are shown in Figure 2.10.

The *in situ* TDS data is the smallest subset of the compiled database (see Table 2.1). Preliminary work used all available TDS observations to construct models. However, using all available data for model construction creates an issue of model validation. The lack of independent data for model validation makes it difficult to assess the ability of each model to predict TDS given new predictor variable values. Instead of using all available TDS data to construct each model, a Monte Carlo cross validation sub-sampling is implemented to improve understanding of model variability.

Training and validation subsets are created from the matrix of *in situ* observations organized by well site. The training subset is selected as 75% of all observation time indexes, TDS and water level, by generating random integers from one to the maximum observation-holding index. The validation subset is defined as the remaining 25% of *in situ* observation values. Thiessen polygon and basin averages are calculated as described in section 2.3.4 for the training and validation subsets separately. Each model construction proceeds as described in section 3.3.4 using only the training subset. The validation subset polygon averages of water levels are then used as "new" predictor values, and the predicted TDS values from each model are evaluated against the validation TDS subset using summary statistics. Monte Carlo re-sampling into training and validation subsets is performed 1,000 times in each study area.

3.3.2 Lithology, land cover regimes

GIS data maps of lithology (Ludington et al. [2007], Stoeser et al. [2005]), land cover (Homer et al. [2015]), and specific yield (Faunt [2009], McGuire et al. [2012]) for both study areas are considered to be one snapshot in time. Using ArcGIS software, all spatial data maps are partitioned into constructed polygons. A histogram of categorical data is generated and used to calculate dominant lithology, dominant land cover, and average specific yield for each polygon.

Regime number	Lithology	Land use
1	Alluvium	WaterIceSnow
2	Alluvium	Developed
3	Alluvium	Barren
4	Alluvium	Forest
5	Alluvium	ScrubShrub
6	Alluvium	GrasslandHerbaceous
7	Alluvium	PastureHay
8	Alluvium	Crops
9	Alluvium	Wetlands
10	Andesite	Developed
11	Andesite	Barren
12	Andesite	Forest
13	Andesite	ScrubShrub
14	Dune sand	WaterIceSnow
15	Dune sand	Developed
16	Dune sand	Barren
17	Dune sand	PastureHay
18	Dune sand	Crops
19	Dune sand	Wetlands
20	Gabbro	Developed
21	Gabbro	Forest
22	Gabbro	ScrubShrub
23	Granodiorite	Developed
24	Granodiorite	Forest
25	Granodiorite	ScrubShrub
26	Intermediate volcanic rock	Forest
27	Mafic volcanic rock	Developed
28	Mafic volcanic rock	Forest
29	Mudstone	Developed
30	Mudstone	Forest
31	Mudstone	ScrubShrub
32	Peridotite	Developed
33	Peridotite	Forest
34	Rhyolite	Developed
35	Sandstone	Developed
36	Sandstone	Barren
37	Sandstone	Forest
38	Sandstone	ScrubShrub
39	Sandstone	GrasslandHerbaceous
40	Sandstone	Crops
41	Schist	Developed
42	Schist	Forest
43	Schist	ScrubShrub
44	Slate	Forest
45	water	Developed
46	water	Forest

Table 3.1: Dominant lithology, dominant land use cover, and associated regime number for the Central Valley.

Regime number	Lithology	Land use
1	Arkose	Developed
2	Arkose	GrasslandHerbaceous
3	Clay or mud	Developed
4	Clay or mud	Forest
5	Clay or mud	ScrubShrub
6	Clay or mud	GrasslandHerbaceous
7	Clay or mud	Wetlands
8	Fine grained mixed clastic rock	Developed
9	Gravel	WaterIceSnow
10	Gravel	Developed
11	Gravel	Forest
12	Gravel	GrasslandHerbaceous
13	Gravel	Wetlands
14	Limestone	Developed
15	Limestone	GrasslandHerbaceous
16	Mudstone	Developed
17	Mudstone	Forest
18	Mudstone	GrasslandHerbaceous
19	Mudstone	Wetlands
20	Sand	Developed
21	Sand	ScrubShrub
22	Sand	GrasslandHerbaceous
23	Sandstone	Developed
24	Sandstone	Forest
25	Sandstone	ScrubShrub
26	Sandstone	GrasslandHerbaceous
27	Sandstone	Wetlands
28	Shale	Developed
29	Shale	Forest
30	Shale	ScrubShrub
31	Shale	GrasslandHerbaceous
32	Silt	Developed
33	Silt	GrasslandHerbaceous
34	Siltstone	GrasslandHerbaceous
35	water	GrasslandHerbaceous

Table 3.2: Dominant lithology, dominant land use cover, and associated regime number for the High Plains.

(a) Central Valley.
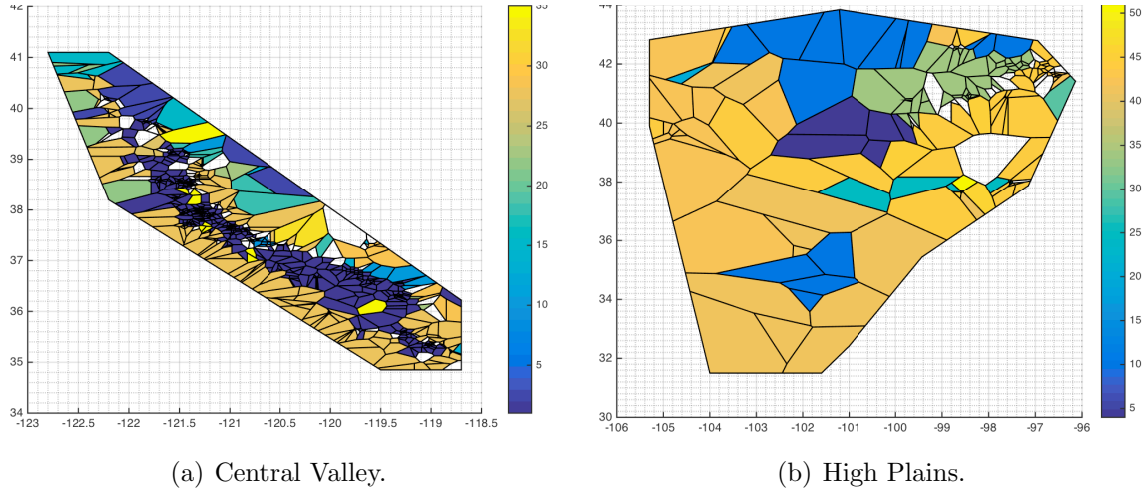
(b) High Plains.

Figure 3.1: Dominant lithology type for Thiessen polygons in the Central Valley and High Plains. Note that number types do not translate between study areas.

As discussed in section 3.2.2, only the land use cover map for 2011 is considered here. Note that the dominant land cover for the large majority of polygons by area in both the Central Valley and High Plains did not change between the previous map years (2001 and 2006).

Dominant lithology and dominant land cover pairs are used to categorize polygons into regimes with similar physical characteristics potentially affecting recharge rates and groundwater quality. Maps of lithology and land cover are partitioned into constructed polygons in ArcGIS in both the Central Valley and High Plains. A histogram of all lithology and land cover classifications is generated for each polygon, and the dominant lithology and land cover types are determined. In cases where there are two or more types of lithology or land use considered dominant for a particular polygon, if any of the neighboring polygons share a dominant categorization type, the common type is chosen as dominant for the polygon of interest. Unique pairs of dominant lithology and dominant land cover are considered a lithology, land cover regime. As mentioned in section 3.2.2, land cover types are consolidated from 16 to 9 categories, substantially reducing the maximum possible number of dominant lithology, land cover pairs. The classification process results in 46 total regimes in the Central Valley (Figure 3.3(a), Table 3.1) and 35 regimes in the High Plains (Figure 3.3(b), Table

60

(a) Central Valley.   (b) High Plains.
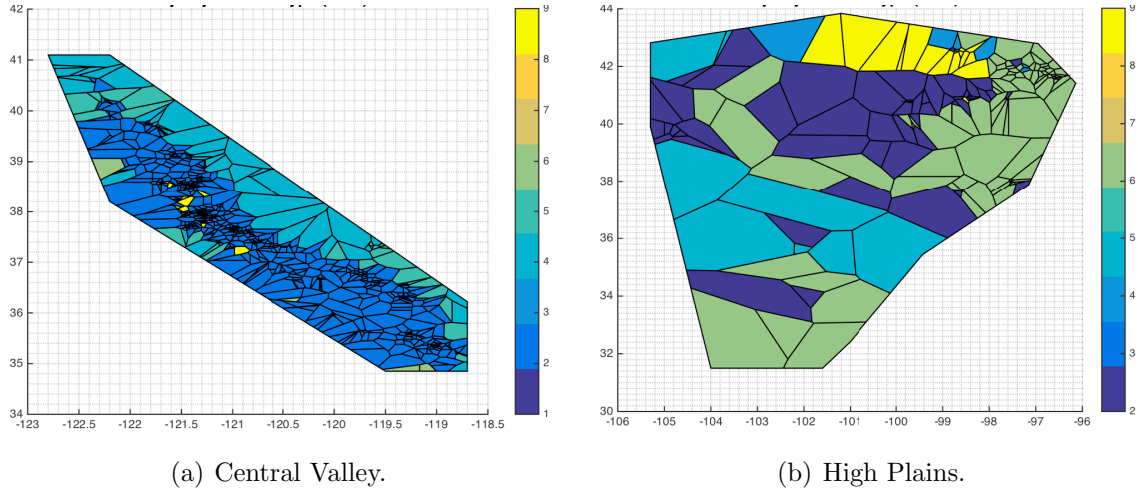
Figure 3.2: Dominant land use cover for Thiessen polygons in the Central Valley and High Plains.

3.2).

The spatial distribution of dominant lithology and dominant land cover are expected to generally occur in specific areas, due to the presence of large geologic formations or the prominence of a particular land cover. Lithology, land cover regimes are not restricted by location in that all polygons within a particular regime are not required to be adjacent, as shown in Figure (3.3). Lithology, land cover regimes do not necessarily translate between the Central Valley and High Plains study areas (Tables 3.1, 3.2). While the 9 land cover categorizations are easily translated across regions, it is not surprising that dominant lithology types are mostly different. Only 6 regimes occur in both areas: mudstone/developed, mudstone/forest, sandstone/developed, sandstone/forest, sandstone/shrub and scrub, sandstone/grassland and herbaceous, and sandstone/grassland.

Average time series for each lithology, land cover regime are calculated as an area-weighted average of regime polygons for each month.

61

(a) Central Valley
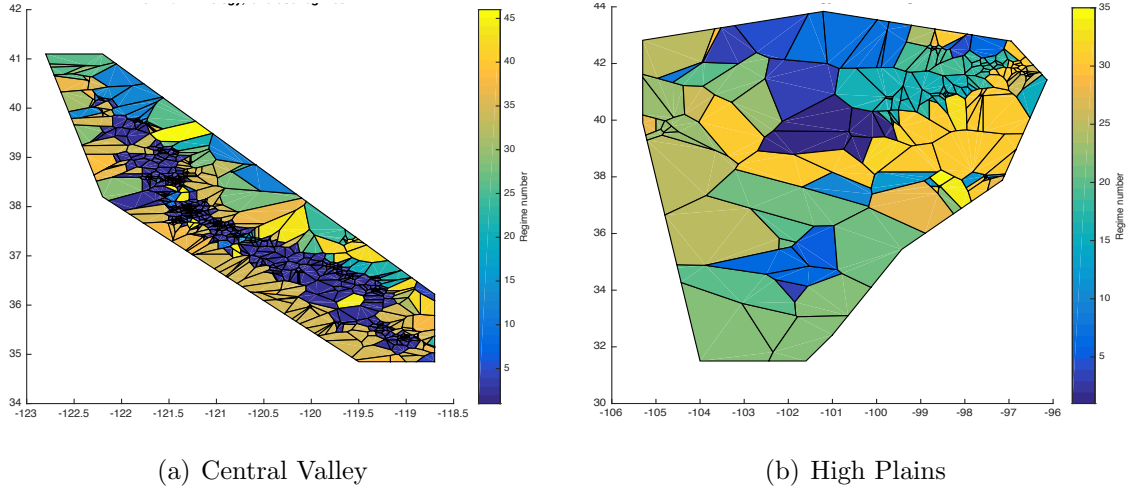
(b) High Plains

Figure 3.3: Locations of lithology, land use regimes by Thiessen polygon in the Central Valley and High Plains. Note that numbers assigned to regimes do not translate across study regions.

3.3.3 GRACE and specific yield

As described in section 2.3.4, monthly GRACE $dSubsurface$ grids are redistributed into Thiessen polygons, and the average specific yield is applied to incorporate a measure of porosity. For each polygon, an area-weighted average of the overlapping $1° \times 1°$ grid cells is calculated for each month. Basin averaged subsurface storage anomalies calculated by Thiessen polygon closely match gridded basin average anomalies (Figure 2.12), suggesting GRACE storage anomalies are not significantly perturbed in polygon redistribution at the basin scale.

Specific yield maps are partitioned into polygons in ArcGIS, and a histogram of values is used to calculate the weighted average S_y for each polygon. Following equation 3.2, $dSubsurface$ is divided by average specific yield for each polygon. To distinguish between GRACE-derived subsurface storage anomalies resulting from equation 2.1 and the specific yield-applied subsurface storage anomalies resulting from equation 3.2, the latter is referred to as $dSubsurface_{S_y}$ for the remainder of this work. Note again that this $dSubsurface_{S_y}$ does not explicitly exclude soil moisture storage anomalies.

As with *in situ* data, time series for each lithology, land cover regime are calculated as an area-weighted average of regime polygons for each month. This approach presents a caveat, as aquifer properties and individual GRACE grid cells are not independent in space. Specific yield and hydraulic conductivity are generally lognormally distributed in space. Post-processing de-striping, Gaussian filtering, and truncation, and the associated spatial covariance of GRACE data is well-documented (Wahr et al. [1998], Swenson and Wahr [2006], Wahr et al. [2006], Landerer and Swenson [2012]). Constructed polygons function as irregularly shaped units, rather than uniformly shaped grid cells. While lithology, land use cover regime polygons are generally grouped together in space, it was not a specified requirement. Furthermore, no account of neighboring polygons within or across regimes are considered. So the relevant polygons used to calculate the area-weighted average time series of a given regime may not always be adjacent to polygons of the same regime. The categorization by regime polygons is not likely affected by the potential covariance of *in situ* data, as the spatial scales of polygons are large by comparison. However, the spatial extent of polygons are generally smaller than a $1° \times 1°$ GRACE grid cell, with few exceptions in the northeastern, central, and south High Plains (see Figure 2.11). The approach used here effectively ignores the spatial covariance inherent to the gridded GRACE product by assuming polygons act as individual units for GRACE-derived $dSubsurface_{S_y}$.

3.3.4 Generalized linear models and model construction

Generalized linear models (GLMs) are a distinguished set of regression models in which the distribution of response variable values can be non-normal. Distributions in the GLM family include normal, Poisson, inverse Gaussian, binomial, and gamma. The response variable is related to the predictor variables by a canonical link function, and term coefficients are denoted as β.

Measures of water quality often have skewed-right distributions (Helsel and Hirsch [2002],

Shamsudduha et al. [2015]). Preliminary work assumed a Poisson distribution of TDS values, as all TDS observation values are inherently positive. However, the Poisson distribution is a discrete probability distribution, and TDS and other water quality constituent observations are continuous. The gamma distribution, a positive continuous distribution, better matches possible expected values of TDS and other constituents and is also able to deal with the issue of non-detect values (Shamsudduha et al. [2015]). All models in this work assume a gamma distribution with a logarithmic canonical link (Shamsudduha et al. [2015]).

Each generalized linear model is calculated using area weighted monthly average time series of $in\ situ$ TDS as the response variable, and monthly average time series of $in\ situ$ water levels, x_1, categorical season (wet or dry), x_2, and when applicable $dSubsurface_{S_y}$, x_3, as predictor variables. Area-weighted monthly average time series for TDS, water levels, and $dSubsurface_{S_y}$ are calculated as described in sections 3.3.1 and 3.3.3, respectively. A categorical variable for season is included as an indicator of wet versus dry months based on climatological precipitation for each region. The most precipitation occurs from approximately October to March in the arid/semi-arid Central Valley (Faunt [2009]), and from March to August in the High Plains summer monsoons (McMahon et al. [2007]).

$$\log TDS = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 \quad (3.3)$$

Quadratic, linear, linear interaction, and intercept terms are considered following Shamsudduha et al. [2015]. While included in the full possible equation of predictor terms (equation 3.3), x_2 is categorical, so the squared term $\beta_{x_2^2}$ is not considered. The minimum form for all models is the constant model, so the constant β_0 occurs in 100% of all resampling runs for both study areas. As is often true in regression analyses, intercept terms do not necessarily carry physical meaning, but cannot be excluded without meaningful justification. GLMs are first constructed including all possible prediction terms, as shown in equation 3.3, then are adjusted step-wise to include only predictor terms that significantly improve the model. A

64

term is added to the model if the sum of squared errors (SSE) is less than 0.05. A term is removed from the model if the p-value of the F-statistic is larger than 0.10. To prevent potential continuous adding and removing of the same terms, the number of adjustments is limited to 20.

Several model forms are constructed to test each of the two hypotheses. In total, there are 4 model categories: (1) the most basic models, $model_0$ for the entire *in situ* record and $model_0Gt$ for the GRACE period, (2) the basic GRACE model, $modelG_0$, (3) the basic regime models, $lith/land_0$ for the full *in situ* record, and $lith/land_0Gt$ for the GRACE period, and (4) the full regime models, $lith/land$.

Model type	Time period	Regimes	$dSubsurface_{S_y}$
$model_0$	full	-	-
$model_0Gt$	GRACE	-	-
$modelG_0$	GRACE	-	y
$lith/land_0$	full	y	-
$lith/land_0Gt$	GRACE	y	-
$lith/land$	GRACE	y	y

Table 3.3: Summary of model categories. The time period refers to whether or not the model is limited to the GRACE period. Regimes refer to polygon categorization by lithology, land use regime, where no regimes (-) indicates one basin-wide model. $dSubsurface_{S_y}$ refers to inclusion (y) or exclusion (-) of GRACE $dSubsurface_{S_y}$ as a potential predictor variable.

Each model type has specific set of constraints, summarized in Table 3.3. The two most basic models, $model_0$ and $model_0Gt$, exclude GRACE $dSubsurface_{S_y}$ as a predictor variable and ignore regime classifications, resulting in two basin-wide models over the full historic record and the GRACE record, respectively. The basic GRACE model, $modelG_0$, is also basin-wide as it ignores regime classifications. The basic regime models, $lith/land_0$ and $lith/land_0Gt$, are constructed for each lithology, land use regime excluding GRACE as a potential predictor for the full and GRACE periods, respectively. Finally, the full regime models, $lith/land$, are constructed for each lithology, land use regime and include GRACE $dSubsurface_{S_y}$ as a potential predictor.

Models including GRACE $dSubsurface_{S_y}$ as a potential predictor variable can have up to 10 terms, as shown in equation 3.3, and are limited to the GRACE period (2003-2014). Models excluding GRACE can have up to 6 terms, as in equation 3.4, and are not necessarily limited to the GRACE period. Models categorized using lithology, land use regimes can begin as either 3.3 or 3.4, depending on whether GRACE $dSubsurface_{S_y}$ is included as a predictor variable or not.

$$\ln TDS = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 \qquad (3.4)$$

To test the hypothesis that lithology, land use regimes are valuable for comparing differences in average recharge effects on TDS to groundwater, we compare the most basic single model, $model_0$, against the basic regime models, $lith/land_0$. The most basic single model, $model_0$, uses only water level and season as predictor variables of TDS, and is basin-wide. The basic regime models also only use water level and season as predictor variables of TDS, but a model is constructed for each lithology, land use regime. GRACE $dSubsurface_{S_y}$ terms are intentionally excluded from this portion of the analysis to test only the success or failure of the regime classification. Because $dSubsurface_{S_y}$ is ignored here, there is no restriction to the GRACE period, so the full length of *in situ* observation records for the Central Valley and High Plains are included.

To test the hypothesis that GRACE $dSubsurface_{S_y}$ is a valuable predictor variable, we first assess cases in which at least one $dSubsurface_{S_y}$ term is included in the final step-wise model. Then we compare a model with $dSubsurface_{S_y}$ as a potential predictor, $modelG_0$, against a model excluding $dSubsurface_{S_y}$, $model_0Gt$, over the GRACE period. The lithology, land use regimes are ignored to test only the impact of GRACE $dSubsurface_{S_y}$ as a TDS predictor, and so are basin-wide.

Finally, we examine models characterized by lithology, land use regimes that also include

GRACE as a potential predictor. Because $dSubsurface_{S_y}$ is included in this analysis, the time period is restricted to the GRACE period. Again, all models initially include all possible predictor terms, then terms are removed and/or added based on improvement of the model SSE.

3.4   Results

All models are reconstructed based on the training subset for each Monte Carlo resampling. In one resampling run, three basin-wide models and three sets of regime models are calculated, totaling to 141 models for the Central Valley and 108 models for the High Plains. Time and computational restraints limit the amount of data output that can be saved for each resampling run. Instead of saving every resampling set run, summary statistics are generated.

For each model and resampling run, 31 variables are saved, which can be grouped by model term coefficients, model fit, and model prediction. Once each model is constructed with the training subset, the calculated coefficients (β) and associated p-values of relevant predictor terms are saved. Model fit statistics are calculated using the training TDS data (y) and the expected TDS values ($\hat{y}$) from the model. Model fit statistics include the number of observation months (N), correlation and p-value of the training data and the expected TDS ($R_{y,\hat{y}}$, $pval_{R_{y,\hat{y}}}$), root mean squared error of the expected TDS ($rmse_{y,\hat{y}}$), standard deviation of the training TDS data (σ_y) and the expected TDS ($\sigma_{\hat{y}}$), mean model bias ($meanbias_{y,\hat{y}}$), dispersion, deviance, and F-statistic versus a constant model and p-value ($Fstat$, $pval_{Fstat}$).

Model prediction statistics are generated using predicted TDS ($\hat{y}_p$) and the independent validating TDS data (y_v). Predicted TDS $\hat{y}_p$ is a new model estimate produced by evaluating the constructed model with the validation groundwater level subset. Categorical season is applied as necessary in the same manner as done for the training subset. Model predic-

tion statistics include correlation and p-value of predicted and validating TDS data ($R_{y_v,\hat{y}_p}$, $pval_{R_{y_v,\hat{y}_p}}$), root mean squared error of the validating TDS ($rmse_{y_v,\hat{y}_p}$), standard deviation of the validating TDS data (σ_{y_v}) and the predicted TDS ($\sigma_{\hat{y}_p}$), and mean bias ($meanbias_{y_v,\hat{y}_p}$).

In the following sections, each set of model comparisons are discussed first by model fit statistics, followed by term coefficients, and finally by model prediction statistics. Each of the statistics listed is discussed in terms of the mean or median value from all 1,000 Monte Carlo resampling runs. When appropriate, the standard deviation or the interquartile range (iqr) of the statistic from all resampling is noted. Unless otherwise stated, statistical significance is reported at the 1% level.

3.4.1   Regime models

Unfortunately, the lack of temporally consistent TDS data limits model construction for several regimes. For months when a TDS value occurs in a regime, a corresponding *in situ* water level is not always available, or vice versa. The lack of available concurrent *in situ* data within the same month leads to several regime models that can not be constructed, despite meeting criteria for the *in situ* database and even Thiessen polygon construction. The regimes affected include 4, 17, and 41 in the Central Valley and 2, 8-13, 24, 29, 30, 32, 34, and 35 in the High Plains. While it may seem like a lot of models, particularly in the High Plains, the percent of total area affected is small, totaling to 0.69% (665.78 km^2) in the Central Valley and 5.30% (39,772.8 km^2) in the High Plains. Because the affected regime models can not be constructed given the available data, they will be omitted from the remaining analysis.

Model	N* (σ_N)	$R_{y,\hat{y}}$ (%smallp)	$rmse_{y,\hat{y}}$	Mean bias$_{y,\hat{y}}$	F-stat (σ_{F-stat})
CV $model_0$	920.469 (3.226)	0.3539 (1)	52.76	-4.304	33.675 (4.911)
CV $lith/land_0$ 2	885.208 (3.28)	0.133 (1)	96.01	-1.29	9.53 (1.73)
CV $lith/land_0$ 35	757.077 (5.15)	0.256 (1)	169.61	-12.05	17.84 (2.66)
HP $model_0$	711.475 (4.10)	0.2329 (1)	67.55	-4.29	12.62 (3.08)
HP $lith/land_0$ 25	102.42 (3.18)	0.2569 (0.524)	71.55	-2.48	4.99 (3.19)
HP $lith/land_0$ 31	356.45 (5.14)	8.81e-17 (0.024)	189.32	-1.3e-12	2.76 (1.11)

Table 3.4: Summary of model fit statistics for the basic model, $model_0$, and the two $lith/land_0$ models covering the largest area in each case for the full *in situ* record. Reported values are a median of the 1,000 Monte Carlo resampling runs, unless denoted by an asterisk (*) indicating the mean value is reported.

3.4.1.1 Full *in situ* record (excludes $dSubsurface_{S_y}$)

To test the usefulness of polygon categorization by lithology, land use regimes, the most basic model, $model_0$, is compared to the basic regime models for the full *in situ* record.

The model fit statistics for the basin wide $model_0$ and two $lith/land_0$ models covering the largest fractional area in each location are summarized in Table 3.4. Central Valley $lith/land_0$ regimes 2, alluvium/developed, and 35, sandstone/developed, cover the largest fractional areas of 27.46% and 23.28% respectively. High Plains $lith/land_0$ regimes 25, sandstone/scrub, shrub, and 31, shale/grassland, herbaceous, cover the largest areas at 12.51% and 15.84% of the total area respectively. The $lith/land_0$ models summarized in Table 3.4 are selected by fractional area because larger spatial extents are likely to encompass a larger number of ground-based observations for model construction.

After $lith/land_0$ regimes 2 and 35, Central Valley $lith/land_0$ model 45 has the largest average number of months of observations at 235.61 ($\sigma_N = 5.25$), despite only covering 1.42% of the total area. Ten models (9, 12, 13, 15, 29, 32, 37, 38, 39, 43) have reasonably large numbers of observations between 103 and 153.73 average months. The remaining thirty models have moderate to low average months of observations between 1.98 and 91.7. Ten models (3, 10, 11, 14, 16, 18, 19, 25, 27, 33) have high median $R_{y,\hat{y}}$ values between 0.341 (26.3% significance rate) and 0.99999 (81.4%), but have low average number of observations between 3.82 and 73.39. Models 29, 34, and 38 have the highest median $R_{y,\hat{y}}$ values and corresponding significance rates of 0.246 (59.4%), 0.284 (90.2%), and 0.313 (79.1%). Fifteen models have moderate $R_{y,\hat{y}}$ values but low significance rates. The remaining fifteen models (1, 5, 7, 9, 12, 13, 15, 24, 2, 30, 31, 32, 36, 40, 44) have very small $R_{y,\hat{y}}$ values on the order of 10^{-15} to 10^{-17} with low significance rates.

Seven Central Valley $lith/land_0$ models (11, 13, 25, 26, 28, 33, 43) have median $rmse_{y,\hat{y}}$ values lower than $model_0$ between 0.595 for model 11 and 52.7 for model 25. The Central Valley $lith/land_0$ models with the lowest median $rmse_{y,\hat{y}}$ are 11, 13, 26, 28, and 33. However, models 11 and 33 have very low numbers of observations relative to the full period at 3.82 and 18.54 months, respectively. Seventeen models (2, 6, 8, 10, 12, 14, 15, 18, 19, 20, 22, 23, 27, 31, 34, 42, 44) have moderate $rmse_{y,\hat{y}}$ values between 53.7 for model 14 and 98.8 for model 22. The remaining nineteen models (1, 3, 5, 7, 9, 16, 21, 24, 29, 30, 32, 35, 36, 37, 38, 39, 40, 45, 46) have large $rmse_{y,\hat{y}}$ values between 101.7 for model 37 and 1.14e3 for model 40. Only four Central Valley $lith/land_0$ models have small F-statistic standard deviations and relatively high significance rates, two of which are models 2 and 35 which both have F-statistic significance rates of 100%, summarized in Table 3.4. The other two are models 8, with $\sigma_F = 5.86$ at 92.2%, and 34, with $\sigma_F = 3.16$ at 61.1% significance.

In the High Plains, $lith/land_0$ models 20, 21, and 22, follow area coverage of models 25 and 31 covering 5%, 9%, and 12.04% of the total area, respectively. Although $lith/land_0$ 20-22

are 3,000+ km^2 smaller than $lith/land_0$ 25, they have averages of 211.05 ($\sigma_N = 4.20$), 214.63 ($\sigma_N = 4.43$), and 287.38 ($\sigma_N = 4.47$) months of *in situ* observations over all resampling runs. Models 1, 6, 16, and 28 have 148.38 ($\sigma_N = 3.76$), 129.18 ($\sigma_N = 3.18$), 114.29 ($\sigma_N = 3.08$), and 110.03 ($\sigma_N = 4.05$) months of observations of the full record respectively, with the remaining thirteen models having averages between 3.94 and 99.72 months. Models 17 and 27 have the highest median $R_{y,\hat{y}}$ values of 0.997 (62.2% significance rate) and 0.964 (16.8%), but have very low average number of months of observations of only 6.57 and 3.94 respectively. Models 1, 6, 7, and 22 have the highest median $R_{y,\hat{y}}$ values and corresponding significance rates of 0.424 (99.9%), 0.258 (69.4%), 0.285 (43.8%), and 0.261 (99.7%). Six $lith/land_0$ models have moderate $R_{y,\hat{y}}$ values between 0.173 and 0.305, but have either a low number of observations or low rates of significance over Monte Carlo resampling. The eight remaining $lith/land_0$ models have very small median $R_{y,\hat{y}}$ values on the order of 10^{-16} to 10^{-18} with low significance rates.

The two High Plains $lith/land_0$ models (17, 27) with the lowest median $rmse_{y,\hat{y}}$ are again those with very few months of observations. Seven models (6, 7, 14, 16, 18, 19, 25) have reasonably large numbers of monthly observations and the lowest median $rmse_{y,\hat{y}}$ between 44.27 mg/L for model 7 and 90.75 mg/L for model 6. Thirteen models (1, 3-5, 15, 20-23, 26, 28, 31, 33) have moderate to high values of $rmse_{y,\hat{y}}$ between 103.07 mg/L and 840.54 mg/L. Models 1, 21, and 22 have the highest rates of significant F-statistics with reasonably low F-statistic standard deviations, with 99.9% significance and $\sigma_F = 10.21$ for model 1, 71.9% and $\sigma_F = 2.55$ for model 21, and 98.9% and $\sigma_F = 5.89$ for model 22.

In the Central Valley, β_0 is statistically significant for 100% of runs for $model_0$, $lith/land_0$ 2, $lith/land_0$ 35, and twenty-eight other $lith/land_0$ models (3, 7-10, 12, 13, 15, 20- 26, 28, 29, 31-34, 37- 39, 43- 46). However, three models (9, 23, 39) out of those twenty-eight have a iqr_{β_0} value larger than the value of β_0 itself on the order of 10^{76} to 10^{78}. Eight models (1, 6, 14, 16, 18, 19, 27, 40) have high significance rates between 83.2% and 99.7% with β_0

Table 3.5: Summary of coefficient terms and interquartile ranges (iqr) for $model_0$ and the two $lith/land_0$ models covering the largest fractional area in each case. Coefficient values are reported as a median of all resampling runs. Note all terms including $dSubsurface_{S_y}$ are excluded.

Term	Central Valley			High Plains		
	$model_0$	$lith/land_0$ 2	$lith/land_0$ 35	$model_0$	$lith/land_0$ 25	$lith/land_0$ 31
β_0 (iqr)	2.96 (0.055)	3.81 (0.064)	3.75 (0.098)	2.55 (0.15)	4.62 (0.11)	4.52 (0.085)
β_{x_1} (iqr)	1.4e-3 (6.4e-4)	3.28e-4 (7.4e-4)	1.3e-3 (1.37e-4)	1.6e-3 (2.2e-4)	8.24e-5 (1.8e-5)	-1.3e-3 (1.1e-3)
β_{x_2} (iqr)	-1.16 (0.039)	-0.875 (0.078)	-0.78 (0.07)	0.257 (0.069)	0.195 (0.036)	0.315 (0.072)
$\beta_{x_1^2}$ (iqr)	-6.06e-7 (9.6e-8)	4.53e-7 (5.6e-8)	-3.22e-7 (6.5e-8)	-4.19e-7 (6.87e-8)	3.83e-8 (9.4e-9)	5.41e-7 (8.3e-8)
$\beta_{x_1 x_2}$ (iqr)	6.99e-4 (1.2e-4)	6.64e-4 (8.5e-5)	3.99e-4 (7.8e-5)	2.22e-4 (4.02e-5)	-1.14e-4 (3.4e-5)	6.62e-4 (1.5e-4)

values between 3.02 ($iqr_{\beta_0} = 0.81$) for model 11 and 9.53 ($iqr_{\beta_0} = 0.036$) for model 36. The remaining five models (5, 11, 30, 36, 42) have moderate to low significance rates between 0% for model 42 and 68.9% for model 36.

Only $model_0$ and $lith/land_0$ 35 have β_{x_1} significant for 100% of runs in the Central Valley, with the term occurring in 3% and 18% of runs respectively. Three models (8, 19, 23) have significance rates from 80.6% to 87.7%, but models 19 and 23 have $iqr_{\beta_{x_1}}$ values larger than β_{x_1}. Model 8 has an median β_{x_1} of 0.0033 with $iqr_{\beta_{x_1}}$ of 8.5e-47, present in 89.9% of runs. Twenty-three other $lith/land_0$ models (1, 2, 3, 6, 7, 9, 11, 13, 14, 16, 20, 21, 24, 26, 29, 32, 33, 36, 37, 39, 40, 43, 45) have $iqr_{\beta_{x_1}}$ values larger than β_{x_1}. Three models (8, 10, 12) have the next highest significance rates for β_{x_1} between 61.1% and 87.7% occurring between 18.5% and 89.9% of runs. The remaining fifteen models (13, 15, 20, 24-26, 28-31, 33, 34, 40, 44, 45) have relatively low significance rates topping out around 25% of runs. Models 5 and 42 do not have a β_{x_1} term for any runs.

Central Valley $model_0$ and $lith/land_0$ models 2 and 35 have β_{x_2} terms significant for 100%

of runs occurring in 3%, 21.9%, and 18% of runs respectively, with values as reported in Table 3.5. Three other $lith/land_0$ models (23, 27, 39) have high significance rates between 71.4% and 82.3% but only model 27 has a reasonable β_{x_2} and $iqr_{\beta_{x_2}}$ values of -2.46 and 1.2 respectively while models 23 and 39 have values on the order of 10^{77} to 10^{78} with relatively low occurrences between 10.5% and 27.7%. Two models (20, 34) have moderate significance rates of 39.98% and 64.4% occurring in 91.3% and 97.8% of runs. The remaining models have a combination of low significance rates, and/or low occurrence. Models 5 and 42 do not have a β_{x_2} term for any runs.

The $\beta_{x_1^2}$ term in Central Valley $model_0$ is significant at 37.5%, occurring in 97.8% of runs. In $lith/land_0$ 2 and 35, $\beta_{x_1^2}$ is significant at 0% and 95.8% occurring in 5.3% and 99.7% of runs respectively. Ten other $lith/land_0$ models (3, 6, 9, 10, 12, 21, 22, 23, 38, 39) have significance rates for $\beta_{x_1^2}$ higher than $model_0$ between 42.9% and 79.1%, but only models 3, 10, 12, and 22 have reasonable $\beta_{x_1^2}$ values and smaller $iqr_{\beta_{x_1^2}}$ values. Models 5, 30, 42, and 45 do not have a $\beta_{x_1^2}$ term for any runs.

The $\beta_{x_1 x_2}$ in Central Valley $model_0$ is significant at 12.7%, occurring in 97% of runs. In $lith/land_0$ 2 and 35, $\beta_{x_1 x_2}$ is significant at 0% and 7.3% occurring in 78.1% and 82% of runs respectively. $\beta_{x_1 x_2}$ generally has low frequency of occurrence in the Central Valley $lith/land_0$ models, below 30% for twenty nine models generally with low significance rates. Six models (6, 21, 23, 25, 27, 39) have moderate to high occurrence between 36.7% and 82.9%, low to moderate significance rates between 1.9% and 60%. Six models (5, 14, 30, 36, 42, 45) had no occurrence of $\beta_{x_1 x_2}$.

In the High Plains, β_0 is statistically significant for 100% of runs for $model_0$ and fifteen $lith/land_0$ models (3, 6, 7, 16, 18-23, 25, 26, 28, 31, 33), all of which have iqr_{β_0} one order of magnitude smaller than the value of β_0, between 3.999 ($iqr_{\beta_0} = 0.598$) for model 3 and 6.017 ($iqr_{\beta_0} = 0.073$) for model 20. Five $lith/land_0$ models (1, 4, 5, 14, 17) have high β_0 significance rates between 92.4% and 99.9%. The remaining two models (15, 27) have significance rates

of 56.7% and 60.2% and relatively large iqr_{β_0} compared to β_0.

The β_{x_1} term occurs in 94.7% of runs, significant at 100% for $model_0$. Only three High Plains $lith/land_0$ models (1, 17, 22) have high significance rates and moderate to high occurrence of β_{x_1} between 90.8% (88.3% occurrence) and 92.9% (56% occurrence), with β_{x_1} values between 2.45e-4 ($iqr_{\beta_{x_1}}$ = 6.04e-5) for model 22 and 0.0021 ($iqr_{\beta_{x_1}}$ = 3.5e-5) for model 17. The remaining $lith/land_0$ models all have significance rates of β_{x_1} below 50%. Four of the remaining models (3, 6, 7, 25) have high occurrence rates between 76.1% and 88.2%. Eleven models have low occurrence of β_{x_1} between 2% and 55%. Four models (20, 23, 31, 33) have no β_{x_1} term for any runs.

High Plains $model_0$ has a β_{x_2} term present in 92.8% of runs, with a significance rate of only 39.1%. $lith/land_0$ model 26 has the highest β_{x_2} significance rate of 57.6%, occurring in only 3.3% of runs, with a value of 1.63 ($iqr_{\beta_{x_2}}$ = 0.0016). Model 6 has the highest occurrence of 77%, but is only significant for 18.4% of runs, with a β_{x_2} value of 0.1869 ($iqr_{\beta_{x_2}}$ = 0.044). Models 25 and 28 have moderate occurrence of a β_{x_2} term of 61.6% and 64.8% respectively, but low significance rates of 4.6% and 21.5%. All remaining $lith/land_0$ models have significance rates below 30% and occurrence rates below 50%. Five models (1, 4, 5, 18, 33) have no β_{x_2} term for any runs.

The $\beta_{x_1^2}$ term in $model_0$ occurs in 100% of runs, significant at 99.9% in the High Plains. Only $lith/land_0$ model 3 has a high occurrence of 93.1% and relatively high significance rate of 63.3% for $\beta_{x_1^2}$, with a value of -4.81e-7 ($iqr_{\beta_{x_1^2}}$ = 1.17e-7). Models 6 and 17 have moderate significance rates of 24.9% and 54.7% and occurrence rates of 68.3% and 9.5% respectively, with $\beta_{x_1^2}$ values of 3.40e-8 ($iqr_{\beta_{x_1^2}}$ = 9.0e-9) and -4.6e-4 ($iqr_{\beta_{x_1^2}}$ = 2.25e-5). The remaining models have low significance and low occurrence rates, both below 50%.

The $\beta_{x_1 x_2}$ in $model_0$ occurs in 5.3% of runs, significant at 5.7%. High Plains $lith/land_0$ model 26 has the highest occurrence of $\beta_{x_1 x_2}$ in 91.1% of runs, and the highest significance rate of all

Table 3.6: Summary of model prediction statistics for basin-wide $model_0$ and select $lith/land_0$ models over the full *in situ* record. Reported values are a median of the 1,000 Monte Carlo resampling runs. Units for $rmse_{y_v,\hat{y}_p}$ and $meanbias_{y_v,\hat{y}_p}$ are mg/L.

Statistic	Central Valley			High Plains		
	$model_0$	$lith/land_0$ 2	$lith/land_0$ 35	$model_0$	$lith/land_0$ 25	$lith/land_0$ 31
$R_{y_v,\hat{y}_p}$ (% small p)	0.2420 (1)	0.092 (0.548)	0.189 (0.988)	0.191 (0.99)	0.2015 (0.107)	-4.9e-17 (0.006)
$rmse_{y_v,\hat{y}_p}$	38.12	82.5	121.99	54.36	51.69	111.17
$meanbias_{y_v,\hat{y}_p}$	5.42	12.66	8.34	11.64	-1.16	23.77

$lith/land_0$ models at 54% with a $\beta_{x_1x_2}$ of 0.0026 ($iqr_{\beta_{x_1x_2}}$ = 6.7e-4). Models 15 and 21 have moderate to low significance and moderate occurrence rates of 44.3% (40.6% occurrence) and 16.4% (65.2% occurrence), respectively, with $\beta_{x_1x_2}$ values of -0.0055 ($iqr_{\beta_{x_1x_2}}$ = 0.0021) and -2.72e-4 ($iqr_{\beta_{x_1x_2}}$ = 8.3e-5). The remaining models have significance and occurrence rates below 40%.

In the Central Valley, the median correlation $R_{y_v,\hat{y}_p}$ for $model_0$ is 0.2420 significant for 100% of runs. The median $rmse_{y_v,\hat{y}_p}$ is 38.12 mg/L, and the median $meanbias_{y_v,\hat{y}_p}$ is 5.42 mg/L. Only three $lith/land_0$ models (2, 25, 35) in the Central Valley have comparable prediction statistics to $model_0$. Model 25 has the highest $R_{y_v,\hat{y}_p}$ of 0.5620, but is only significant for 38.2% of runs, with $rmse_{y_v,\hat{y}_p}$ of 62.69 mg/L and a low $meanbias_{y_v,\hat{y}_p}$ of 4.35 mg/L. Model 2 has the lowest $R_{y_v,\hat{y}_p}$ of 0.092 significant for 54.8% of runs, with $rmse_{y_v,\hat{y}_p}$ of 82.5 mg/L, and a relatively low $meanbias_{y_v,\hat{y}_p}$ of 12.66 mg/L. Model 35 has a $R_{y_v,\hat{y}_p}$ of 0.189 significant for 98.8% of runs, with a moderately high $rmse_{y_v,\hat{y}_p}$ of 121.99 mg/L, and a low $meanbias_{y_v,\hat{y}_p}$ of 8.34 mg/L. Prediction statistics for the entire set of $lith/land_0$ models in the Central Valley are described below.

Five Central Valley $lith/land_0$ models (2, 25, 27, 34, 35) have the largest magnitude median $R_{y_v,\hat{y}_p}$ and highest accompanying significance rates over resampling with values of 0.092 (54.8%), 0.562 (38.2%), 0.799 (40.8%), 0.293 (32.3%), and 0.189 (98.8%) respectively. Six models (8, 19, 20, 29, 33, 38) have moderately large $R_{y_v,\hat{y}_p}$ magnitudes but low significance

rates, from 0.119 at 0.6% for model 33 to 0.218 at 10.1% for model 20. Ten models (3, 10, 21, 22, 23, 28, 37, 39, 43, 45) have moderately low $R_{y_v,\hat{y}_p}$ magnitudes between -0.0044 for model 21 and 0.0985 for model 23, with significance rates up to 3.7%. Sixteen models (6, 7, 9, 12, 13, 14, 15, 16, 24, 26, 31, 32, 40, 42, 44, 46) have very small $R_{y_v,\hat{y}_p}$ positive or negative values on 10^{-16} to 10^{-18} orders of magnitude, with significance rates at less than 1% or 0%. Models 1, 5, 11, 18, 30, and 36 have $R_{y_v,\hat{y}_p}$ values of 0, also with significance rates at less than 1% or 0%.

Only two $lith/land_0$ models (13, 28) have median $rmse_{y_v,\hat{y}_p}$ values smaller than $model_0$ of 25.66 and 31.43 respectively. Twenty-one models (2, 5, 6, 10-12, 14, 15, 20, 23, 25-27, 31-34, 37, 42-44) have moderate $rmse_{y_v,\hat{y}_p}$ values under 100 mg/L, on the same order of magnitude as $model_0$. The remaining twenty $lith/land_0$ models (1, 3, 7, 8, 9, 16, 18, 19, 21, 22, 24, 29, 30, 35, 36, 38, 39, 40, 45, 46) have high median $rmse_{y_v,\hat{y}_p}$ values between 100 mg/L and 1.36e3 mg/L.

Nine Central Valley $lith/land_0$ models (6, 10, 12, 13, 22, 25, 33, 34, 45) have a lower magnitude $meanbias_{y_v,\hat{y}_p}$ value than $model_0$ between -0.083 for model 34 and 4.43 for model 10. Nine models (1, 14, 18, 20, 23, 28, 32, 35, 37) have relatively low magnitude $meanbias_{y_v,\hat{y}_p}$ between 5.83 for model 1 and -9.60 for model 18. Nineteen models (2, 3, 5, 7, 8, 9, 11, 15, 19, 24, 26, 29, 31, 38, 39, 42, 43, 44, 46) have moderate $meanbias_{y_v,\hat{y}_p}$ magnitudes under 100, and the remaining six models (16, 21, 27, 30, 36, 40) have high $meanbias_{y_v,\hat{y}_p}$ magnitudes between 100 and 1.19e3 for model 40.

In the High Plains, the median correlation $R_{y_v,\hat{y}_p}$ for $model_0$ is 0.191 significant for 99% of runs. The median $rmse_{y_v,\hat{y}_p}$ is 54.36 mg/L, and the median $meanbias_{y_v,\hat{y}_p}$ is 11.64 mg/L.

Four $lith/land_0$ models, 1, 6, 22, and 25 in the High Plains have comparable prediction statistics to $model_0$. Model 22 has a $R_{y_v,\hat{y}_p}$ of 0.2179 with the highest significance rate of 64.7% and a relatively low $meanbias_{y_v,\hat{y}_p}$ of -18.77, but has a large value for $rmse_{y_v,\hat{y}_p}$ of

317.58 mg/L. Model 6 has the lowest $R_{y_v,\hat{y}_p}$ of 0.195 significant for only 16.8% of runs, but has a moderately low $rmse_{y_v,\hat{y}_p}$ of 87.17 mg/L, and the lowest $meanbias_{y_v,\hat{y}_p}$ of -0.5737 mg/L. Model 25 has $R_{y_v,\hat{y}_p}$ of 0.2015 significant for only 10.7% of runs, but has relatively low $rmse_{y_v,\hat{y}_p}$ of 51.69 mg/L, and a small $meanbias_{y_v,\hat{y}_p}$ of -1.16 mg/L. Model 1 has the highest $R_{y_v,\hat{y}_p}$ of 0.2526 significant for 43.6% of runs, with a moderate $rmse_{y_v,\hat{y}_p}$ of 118.98 mg/L, and a moderately low $meanbias_{y_v,\hat{y}_p}$ of 38.52 mg/L. Prediction statistics for the entire set of $lith/land_0$ models in the High Plains are described below.

The High Plains $lith/land_0$ models 1, 6, 22, 25, and 27 have the highest median $R_{y_v,\hat{y}_p}$ in the High Plains at 0.2526 (43.6% of runs significant), 0.1954 (16.8%), 0.2179 (64.7%), 0.2015 (10.7%), and 0.4883 (0.3%) respectively. Five models (3, 7, 21, 26, 28) have moderate $R_{y_v,\hat{y}_p}$ magnitudes between -0.0267 (1.8% significant) for model 21 and 0.176 (3.6%) for model 7. The remaining thirteen models have very small (10^{-16}) or 0 values for $R_{y_v,\hat{y}_p}$, with significance rates at less than 1% or 0%.

Five $lith/land_0$ models (7, 17, 18, 25, 27) have median $rmse_{y_v,\hat{y}_p}$ values lower than that of $model_0$ between 7.78 for model 17 and 51.69 for model 25. Four models (6, 14, 16, 19) have moderate $rmse_{y_v,\hat{y}_p}$ values between 64.13 and 87.17. The remaining thirteen models have high $rmse_{y_v,\hat{y}_p}$ values between 111.17 and 937.54. Eight models (4, 5, 6, 7, 25, 27, 28, 33) have smaller magnitude median $meanbias_{y_v,\hat{y}_p}$ than $model_0$ between -0.344 for model 5 and 11.33 for model 27. Ten models (1, 3, 14, 15, 16, 18, 19, 20, 22, 31) have moderate $meanbias_{y_v,\hat{y}_p}$ magnitudes between 15.05 for model 14 and 47.5 for model 15. The remaining four models (17, 21, 23, 26) have high $meanbias_{y_v,\hat{y}_p}$.

3.4.1.2 GRACE-period record, including $dSubsurface_{S_y}$

To assess whether including regime categorizations improves upon a basin-wide model including GRACE $dSubsurface_{S_y}$ as a predictor, the basic GRACE $dSubsurface_{S_y}$ model,

$modelG_0$ is compared against the full regime models, $lith/land$, including $dSubsurface_{S_y}$.

Central Valley $modelG_0$ has an average N of 132 months, the total possible for the GRACE period, with a median $R_{y,\hat{y}}$ of 0.4307 significant for 100% of runs. The median $rmse_{y,\hat{y}}$ for $modelG_0$ is 63.08 mg/L and median $meanbias_{y,\hat{y}}$ of 58.61. The median F-statistic versus the constant model is 8.65 significant for 100% of runs, with a σ_{F-stat} of 2.51.

Central Valley models 2 and 35 are the only two $lith/land$ regimes to have N over 100. Nine models (8, 9, 12, 15, 37-39, 43, 45) have at least half of the possible N, and thirteen models (1, 5, 11, 14, 16, 18, 19, 22, 30, 31, 36, 40, 42) have small average N below 10. Models 2 and 35 have $R_{y,\hat{y}}$ significance rates of 100%, but have smaller values of $R_{y,\hat{y}}$ at 0.4029 and 0.3691 respectively. Nine models (1, 8, 11, 16, 18, 19, 21, 27, 33) have $R_{y,\hat{y}}$ values larger than $modelG_0$ between 0.4991 (97.4% significance) and 1 (65.8% significance), but only model 8 has moderately large N. Twenty models (6, 7, 10, 12, 14, 15, 20, 23, 25, 28, 31, 32, 34, 37, 38, 40, 43-46) have moderate $R_{y,\hat{y}}$ and significance, and the remaining twelve $lith/land$ models (3, 5, 9, 13, 22, 24, 26, 29, 30, 36, 39, 42) have very small $R_{y,\hat{y}}$ values and low significance rates.

Eleven $lith/land$ models (2, 11, 12, 13, 25, 26, 27, 28, 33, 37, 43) have a median $rmse_{y,\hat{y}}$ smaller than $modelG_0$ between 0.6325 for model 11 and 57.07 for model 37. Sixteen models (1, 5, 6, 8, 10, 14, 15, 18, 19, 20, 31, 32, 34, 42, 44, 45) have moderate $rmse_{y,\hat{y}}$ values between 64.8 for model 31 and 119.58 for model 5. The remaining sixteen models have large $rmse_{y,\hat{y}}$ values. Thirty $lith/land$ models (1-3, 5-8, 10, 12-6, 18-20, 24-26, 28, 29, 31-34, 37, 42-45) have median $meanbias_{y,\hat{y}}$ magnitudes smaller than $modelG_0$ between 2.09e-11 for model 1 and -58.84 for model 3. Nine models (11, 21, 22, 27, 30, 35, 36, 40, 46) have moderate to high median $meanbias_{y,\hat{y}}$ magnitudes between -61.45 and 3.74e3. The remaining four models (9, 23, 38, 39) have unrealistically high $meanbias_{y,\hat{y}}$ magnitudes on the order of 10^{76}.

Only $lith/land$ model 2 has a significant F-statistic for 100% of runs, with an F-statistic

comparable to $modelG_0$ of 8.02 and a smaller σ_{F-stat} than $modelG_0$ of 1.699. Sixteen other $lith/land$ models (3, 10, 13, 20, 24, 25, 26, 28, 29, 32, 35, 37, 43, 44, 45, 46) have F-statistic values over 1.5, σ_{F-stat} smaller than the respective F-statistic value and than that of $modelG_0$ between 0.87 and 2.42, but only models 35 and 43 have moderate to high significance rates of 99.7% and 69.8% respectively. Three models (8, 14, 34) have σ_{F-stat} larger than $modelG_0$ but smaller than the F-statistic value itself. The remaining models have either significance rates of 0% or σ_{F-stat} larger than the F-statistic value.

In the High Plains, $modelG_0$ has an average N of 55.015 months of the 132 total possible for the GRACE period, with a median $R_{y,\hat{y}}$ of 0.3091 significant for 40.7% of runs. The median $rmse_{y,\hat{y}}$ for $modelG_0$ is 56.19 mg/L and median $meanbias_{y,\hat{y}}$ of 22.26. The median F-statistic versus the constant model is 3.47 significant for 27.9% of runs, with a σ_{F-stat} of 289.2.

The number of observations during the GRACE period in the High Plains are low, with average N between 1 and 33.7 out of 132 possible months. Three $lith/land$ models (25, 31, 33) have larger median $R_{y,\hat{y}}$ and higher significance rates than $modelG_0$ of 0.601 (52.8% significance), 0.531 (63.1% significance), and 0.916 (52.9% significance) respectively. Another six models (3, 16, 18, 22, 26, 27) have larger $R_{y,\hat{y}}$ between 0.464 and 0.975, but lower significance rates than $modelG_0$. Model 6 has a lower $R_{y,\hat{y}}$ value of 0.192 but a higher significance rate of 46.3% than $modelG_0$. The remaining twelve models have smaller median $R_{y,\hat{y}}$ and significance rates than $modelG_0$. Four $lith/land$ models (7, 16, 18, 27) have smaller median $rmse_{y,\hat{y}}$ values than $modelG_0$ between 3.56 mg/L for model 27 and 50.68 mg/L for model 18. Eight models (6, 14, 17, 19, 22, 25, 28, 33) have moderate median $rmse_{y,\hat{y}}$ values between 58.31 mg/L and 95.03 mg/L. The remaining ten models (1, 3, 4, 5, 15, 20, 21, 23, 26, 31) have large $rmse_{y,\hat{y}}$ values between 103.1 and 1.04e3 mg/L.

Seven $lith/land$ models (7, 14, 18, 19, 22, 23, 27) have median $meanbias_{y,\hat{y}}$ magnitudes smaller than that of $modelG_0$ between 1.36e-10 for model 27 and 20.3 for model 23. Six

models (1, 6, 16, 20, 21, 25) have moderate $meanbias_{y,\hat{y}}$ magnitudes between -31.67 and -46.86, with models 16 and 25 having positive magnitudes. The remaining nine models (3, 4, 5, 15, 17, 26, 28, 31, 33) have large $meanbias_{y,\hat{y}}$ values between -92.4 and 1.37e4. Four High Plains $lith/land$ models (18, 23, 25, 31) have larger median F-statistics and smaller σ_{F-stat} than $modelG_0$, but only models 18 and 25 have σ_{F-stat} values smaller than the F-statistic value itself of 4.404 with σ_{F-stat} of 3.38 significant for 9.8% of runs for model 18, and 4.77 with σ_{F-stat} of 2.067 significant for 21.6% of runs. Model 14 has a σ_{F-stat} smaller than the F-statistic value, but is still 2 orders of magnitude larger than $modelG_0$. The remaining models have σ_{F-stat} values larger than the respective F-statistic by at least two orders of magnitude, as well as being larger than σ_{F-stat} for $modelG_0$.

The median term coefficient and associated interquartile range for Central Valley $modelG_0$ are summarized in Table 3.8. All $lith/land$ models have significance rates of β_0 and β_{x_1} lower than $modelG_0$. Four $lith/land$ models (18, 22, 23, 39) have higher β_{x_2} significance rates than $modelG_0$ between 70.7% and 100%. Twenty-six models (1, 2, 6, 7, 9-12, 15, 16, 18, 19, 21, 23, 25, 27, 31, 33, 35, 37, 38-40, 43-45) have higher β_{x_3} significance rates than $modelG_0$ between 4.0% and 99.57%. Thirty-one models (1, 3, 6-12, 15, 18-21, 23, 25-28, 31-36, 38, 39, 43-46) have higher $\beta_{x_1^2}$ significance rates than $modelG_0$ between 6.12% and 97.8%. Twenty-nine models (1, 3, 6-9, 11, 12, 15, 16, 18-21, 23, 25, 27, 29, 31-35, 38-40, 43, 45, 46) have higher $\beta_{x_3^2}$ significance rates than $modelG_0$ between 1.47% and 98.3%. Twenty-eight models (1, 3, 6-10, 12, 15, 19-28, 31-35, 38, 39, 44-46) have higher $\beta_{x_1 x_2}$ significance rates than $modelG_0$ between 3.8% and 95.1%. Thirty-one models (1-3, 6-12, 15, 16, 18-21, 23, 25-28, 31-35, 38, 39, 43, 44, 46) have higher $\beta_{x_1 x_3}$ significance rates than $modelG_0$ between 5.17% and 93.98%. Twenty-three models (1-3, 6-9, 12, 15, 19-21, 23, 25, 27, 31, 33-35, 38, 39, 45, 46) have higher $\beta_{x_2 x_3}$ significance rates than $modelG_0$ between 8.67% and 94.82%.

The average term coefficient and associated standard deviation for High Plains $modelG_0$ are summarized in Table 3.8. Seventeen models (3, 6, 7, 14-23, 25-27, 33) have higher β_0

Table 3.7: Summary of model prediction statistics for basin-wide $modelG_0$ and select $lith/land$ models over the GRACE period, including $dSubsurface_{S_y}$ as a potential predictor. Reported values are a median of the 1,000 Monte Carlo resampling runs. Units for $rmse_{y_v,\hat{y}_p}$ and $meanbias_{y_v,\hat{y}_p}$ are mg/L.

Statistic	Central Valley			High Plains		
	$modelG_0$	$lith/land$ 2	$lith/land$ 35	$modelG_0$	$lith/land$ 25	$lith/land$ 31
$R_{y_v,\hat{y}_p}$ (% small p)	0.2644 (0.76)	0.2756 (0.811)	0.2579 (0.746)	0.004675 (0.033)	0.148 (0.006)	0.2529 (0.076)
$rmse_{y_v,\hat{y}_p}$	59.7	34.45	189.25	484.2	94.3	110.8
$meanbias_{y_v,\hat{y}_p}$	62.3	32.46	132.1	176.99	30.8	73.1

significance rates than $modelG_0$ between 48.4% and 99.8%. Ten $lith/land$ models (3, 6, 14, 15, 19, 20, 23, 26, 27, 28) have higher β_{x_1} significance rates than $modelG_0$ between 21.93% and 79.63%. Eleven models (3, 6, 16, 18, 22, 23, 25, 26, 28, 31, 33) have higher β_{x_2} significance rates than $modelG_0$ between 8.6% and 84.9%. Two models (26, 31) have higher β_{x_3} significance rates than $modelG_0$ of 71.8% and 84.4% respectively. Two $lith/land$ models (26, 28) have higher $\beta_{x_1^2}$ significance rates than $modelG_0$ of 66.5% and 49.1% respectively. Eleven models (3, 6, 16, 18, 22, 23, 26, 27, 28, 31, 33) have higher $\beta_{x_3^2}$ significance rates than $modelG_0$ between 9.64% and 68.0%. Eight models (3, 6, 23, 25, 26, 28, 31, 33) have higher $\beta_{x_1 x_2}$ significance rates than $modelG_0$ between 12.97% and 72.3%. Eight models (3, 6, 16, 18, 23, 26, 28, 33) have higher $\beta_{x_1 x_3}$ significance rates than $modelG_0$ between 17.1% and 61.42%. Four models (3, 6, 26, 31) have higher $\beta_{x_2 x_3}$ significance rates than $modelG_0$ between 15.5% and 66.3%.

In the Central Valley, the basin wide model $modelG_0$ has a $R_{y_v,\hat{y}_p}$ value of 0.2644 with a 76% significance rate, a median $rmse_{y_v,\hat{y}_p}$ of 59.73 mg/L, and a median $meanbias_{y_v,\hat{y}_p}$ of 62.33 mg/L. The only two $lith/land$ models that have similar prediction statistics are models 2 and 35. The $lith/land$ 2 model has the highest $R_{y_v,\hat{y}_p}$ value of the three of 0.2756 significant for 81.1% of runs, a low $rmse_{y_v,\hat{y}_p}$ of 34.45 mg/L, and a low $meanbias_{y_v,\hat{y}_p}$ of 32.46 mg/L. Regime model 35 has a comparable $R_{y_v,\hat{y}_p}$ value to $modelG_0$ of 0.2579 with a significance

rate of 74.6%, a larger $rmse_{y_v, \hat{y}_p}$ of 189.25 mg/L, and a larger $meanbias_{y_v, \hat{y}_p}$ of 132.13 mg/L.

Three Central Valley $lith/land$ models (2, 27, 33) have larger $R_{y_v, \hat{y}_p}$ values than $modelG_0$, but models 27 and 33 have relatively low significance rates of 36% and 4.9% respectively. Five models (19, 20, 34, 35, 43) have comparable $R_{y_v, \hat{y}_p}$ magnitudes, but have significance rates below 14% with the exception of model 35 at 74.6% significance. All other $lith/land$ models have low $R_{y_v, \hat{y}_p}$ magnitudes and significance rates.

Seven $lith/land$ models (2, 12, 13, 26, 28, 37, 43) have smaller $rmse_{y_v, \hat{y}_p}$ values than $modelG_0$ between 25.53 for model 13 and 49.57 for model 37. Thirteen models (5, 10, 14, 15, 20, 25, 31, 32, 33, 34, 42, 44, 45) have moderate $rmse_{y_v, \hat{y}_p}$ values larger than that of $modelG_0$, but under 100 mg/L. Nineteen models (1, 3, 6, 7, 8, 11, 16, 18, 19, 21, 22, 24, 27, 29, 30, 35, 36,40, 46) have moderate to large $rmse_{y_v, \hat{y}_p}$ values between 110.04 for model 18 and 1.64e3 for model 21. The remaining four models (9, 23, 38, 39) have very large values on the order of 10^{76}.

Twenty-nine Central Valley $lith/land$ models (1-3, 5-8, 10-15, 18, 20, 24-26, 28, 29, 31-34, 37, 42-45) have smaller magnitude $meanbias_{y_v, \hat{y}_p}$ than that of $modelG_0$ between 0.0624 for model 13 and 56.44 for model 29. Model 46 has a moderate $meanbias_{y_v, \hat{y}_p}$ on the same order of magnitude as for $modelG_0$ of -86.57. Nine models (16, 19, 21, 22, 27, 30, 35, 36, 40) have higher magnitude $meanbias_{y_v, \hat{y}_p}$ than $model_0$ between 132.13 for model 35 and 4.05e3 for model 27, with the remaining four models (9, 23, 38, 39) on the order of 10^{76}.

In the High Plains, the basin wide model $modelG_0$ has a $R_{y_v, \hat{y}_p}$ value of 0.004675 with a 3.3% significance rate, a median $rmse_{y_v, \hat{y}_p}$ of 484.19 mg/L, and median $meanbias_{y_v, \hat{y}_p}$ of 176.99 mg/L.

Three High Plains $lith/land$ models (26, 28, 31) have higher $R_{y_v, \hat{y}_p}$ values and higher significance rates than $modelG_0$ of 0.0372 (3.9% significance), 0.0782 (14.8% significance), and 0.2529 (7.6% significance) respectively. Five additional models (22, 23, 25, 27, 33) have

higher $R_{y_v, \hat{y}_p}$ values than $modelG_0$ between 0.1010 and 0.4965, but have low significance rates of 2% or lower.

Nineteen $lith/land$ models (1, 4, 5, 6, 7, 14, 15, 16, 17, 18, 19, 21, 22, 23, 25, 27, 28, 31, 33) have smaller median $rmse_{y_v, \hat{y}_p}$ values than $modelG_0$ between 21.08 for model 27 and 357.6 for model 23. Three models (3, 20, 26) have larger $rmse_{y_v, \hat{y}_p}$ values than $modelG_0$. Nineteen models (1, 4, 5, 6, 7, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 31, 33) have smaller magnitude $meanbias_{y_v, \hat{y}_p}$ than $modelG_0$ between 6.15 for model 27 and 131.5 for model 15. Only three $lith/land$ models (3, 26, 28) have larger median $meanbias_{y_v, \hat{y}_p}$ magnitudes than $modelG_0$.

3.4.2 GRACE $dSubsurface_{S_y}$ as a predictor variable

3.4.2.1 Basin-wide models over the GRACE period record (excludes regimes)

To test the TDS predictability added by including GRACE $dSubsurface_{S_y}$ as a predictor, the most basic model limited to the GRACE period, $model_0Gt$, is compared to the basic GRACE model, $modelG_0$. Both models cover the same time period, exclude regime classifications (basin-wide), and include ground-based water level observations and categorical season as predictors. The only difference is that the basic GRACE model $modelG_0$ includes $dSubsurface_{S_y}$ as a possible predictor variable. Note that the total number of months in the GRACE period is 132.

In the Central Valley, including $dSubsurface_{S_y}$ as a potential predictor modestly improves model fitting of average TDS during the GRACE period. Both basin-wide models have *in situ* observations for each month in the GRACE period (132), but the estimated TDS ($\hat{y}$) from $modelG_0$ is more correlated with TDS training data than $model_0Gt$, with median values of $R_{y, \hat{y}}$ of 0.4307 and 0.3783, respectively. It is worth noting that for every Monte Carlo

resampling run in the Central Valley, $R_{y,\hat{y}}$ is statistically significant at $pval_{R_{y,\hat{y}}} < 0.01$. The median $rmse_{y,\hat{y}}$ are 64.432 and 63.08 mg/L for $model_0Gt$ and $modelG_0$ respectively, suggesting that estimation accuracy is improved by including $dSubsurface_{S_y}$, albeit moderately. The median F-statistics versus the respective constant models are 12.327 for $model_0Gt$ and 8.654 for $modelG_0$, which are statistically significant at the 1% level for every resampling run in the Central Valley. While the larger median F-statistic for $model_0Gt$ suggests more strongly against the respective constant model, the standard deviations the F-statistics of 2.967 for $model_0Gt$ and 2.514 for $modelG_0$ over all resampling runs indicate that the model fit variability is reduced by including $dSubsurface_{S_y}$.

The High Plains case study is much more uncertain than the Central Valley during the GRACE period due to the substantial reduction in TDS observations beginning in the late 1990s, as shown in Figure 2.2(d). Although characterizing average TDS in the High Plains during the GRACE period poses a particular challenge given the current availability of observations, assessing model statistics may still provide some insight into the potential of GRACE $dSubsurface_{S_y}$ as a potential predictor. Both $model_0Gt$ and $modelG_0$ have an average of 55.015 ($\sigma_N = 0.93$) months of *in situ* data out of 132 GRACE-period months. With ground-based values for fewer than half of the period of interest, including $dSubsurface_{S_y}$ as a predictor improves $R_{y,\hat{y}}$ and the occurrence of significance at the 1% level, with median values of 0.1356 for $model_0Gt$, statistically significant for only 13.3% of resampling runs, and 0.3091 for $modelG_0$, significant for 40.7% of resampling runs. The median $rmse_{y,\hat{y}}$ of 56.19 mg/L for $modelG_0$ compared to 55.88 mg/L for $model_0Gt$ are very close. However, $dSubsurface_{S_y}$ terms may introduce variability, as suggested by a $meanbias_{y,\hat{y}}$ of 22.26 for $modelG_0$ compared to 2.16 for $model_0Gt$. The F-statistic and the percent occurrence of statistical significance improve from 2.467 significant for 3.4% of resamples for $model_0Gt$ to 3.470 significant for 27.9% of resamples for $modelG_0$, but the F-statistic standard deviations of 1.459 and 289.216, respectively indicate that including $dSubsurface_{S_y}$ introduces substantial variability to the model fit.

Table 3.8: Summary of coefficient terms and interquartile ranges for $model_0Gt$ and $modelG_0$. Coefficient values are reported as a median of all resampling runs.

Term	Central Valley		High Plains	
	$model_0Gt$	$modelG_0$	$model_0Gt$	$modelG_0$
β_0 (iqr)	3.982 (0.090)	3.97 (0.11)	3.78 (20.18)	17.899 (21.3)
β_{x_1} (iqr)	7.520e-4 (1.04e-4)	7.69e-4 (1.1e-4)	-0.0179 (5.2e-3)	-0.0175 (8.1e-3)
β_{x_2} (iqr)	-0.415 (0.081)	-0.370 (0.088)	0.647 (0.082)	0.7799 (0.14)
β_{x_3} (iqr)	-	1.76e-4 (2.8e-5)	-	-0.0213 (7.1e-3)
$\beta_{x_1^2}$ (iqr)	-5.14e-7 (2.22e-7)	-4.87e-7 (9.5e-8)	3.86e-6 (1.05e-6)	4.12e-6 (1.3e-6)
$\beta_{x_3^2}$ (iqr)	-	-1.98e-7 (2.0e-8)	-	-2.35e-4 (4.8e-4)
$\beta_{x_1x_2}$ (iqr)	6.02e-4 (1.42e-4)	6.90e-4 (1.5e-4)	-2.4e-3 (5.3e-4)	-0.0025 (6.2e-4)
$\beta_{x_1x_3}$ (iqr)	-	-4.33e-7 (7.65e-8)	-	-1.98e-5 (8.1e-5)
$\beta_{x_2x_3}$ (iqr)	-	-3.75e-4 (7.41e-5)	-	-0.0124 (0.037)

In the Central Valley, β_0 for $model_0Gt$ and $modelG_0$ are statistically significant for 100% of resampling runs. The form of $model_0Gt$ varies minimally, as suggested by relatively small interquartile ranges shown in Table 3.8. Inclusion of β_{x_1} and β_{x_2} each occur in 95.5% of runs, with statistical significance at 99.9% and 89.5% respectively, when present. The squared and interaction terms $\beta_{x_1^2}$ and $\beta_{x_1x_2}$ are likely negligible for $model_0Gt$, occurring for only 2.5% and 4.1% of runs, with 0% and 2.44% significance rates. The form of $modelG_0$ is much more variable than $model_0Gt$ in the Central Valley, with each possible predictor term occurring in the course of resampling. Linear terms β_{x_1}, β_{x_2}, and β_{x_3} occur in 39.8%, 86.9%, and 11.6% of runs, having significance rates of 100%, 63.2%, and 3.4% respectively. The two interaction terms including $dSubsurface_{S_y}$ occur more frequently in runs than the interaction term excluding it. $\beta_{x_1x_3}$ is included in 59.5% of runs and 11.1% include $\beta_{x_2x_3}$, with 4.37% and 4.5% significance rates, respectively. The squared term $\beta_{x_1^2}$ occurs in a small 7.6% of runs,

Table 3.9: Summary of model prediction statistics for basin-wide $model_0Gt$ and $modelG_0$ over the GRACE period, excluding and including $dSubsurface_{S_y}$ as a potential predictor respectively. Reported values are a median of the 1,000 Monte Carlo resampling runs. Units for $rmse_{y_v,\hat{y}_p}$ and $meanbias_{y_v,\hat{y}_p}$ are mg/L.

	Central Valley		High Plains	
Statistic	$model_0Gt$	$modelG_0$	$model_0Gt$	$modelG_0$
$R_{y_v,\hat{y}_p}$ (% small p)	0.2558 (0.637)	0.2644 (0.76)	-6.4e-17 (0.007)	0.0047 (0.033)
$rmse_{y_v,\hat{y}_p}$	58.76	59.73	44.7	484.2
$meanbias_{y_v,\hat{y}_p}$	39.9	62.3	27.5	176.99

with 5.26% significance. The two terms with the lowest rates of inclusion, $\beta_{x_1x_2}$ in 1.9% of runs and $\beta_{x_3^2}$ in only 0.4%, also both have significance rates of 0%.

In the High Plains, β_0 are significant for 59% and 37.5% of runs for $model_0Gt$ and $modelG_0$, respectively. Groundwater levels appear to be more informative to average TDS than categorical season in the High Plains version of $model_0Gt$. The linear and squared terms β_{x_1} and $\beta_{x_1^2}$ occur in 46.8% and 45.3% of resampling runs, with significance rates of 14.1% and 16.78% respectively. The categorical terms β_{x_2} and $\beta_{x_1x_2}$ are included in 1.9% and 4.8% of runs, both with significance rates of 0%. Including $dSubsurface_{S_y}$ as a predictor in $modelG_0$ gives the illusion of improving the model form compared to $model_0Gt$, with increased inclusion and significance of β_{x_1}, β_{x_2}, $\beta_{x_1^2}$, and $\beta_{x_1x_2}$. All possible $dSubsurface_{S_y}$ terms occur in the course of resampling, most notably β_{x_3} occurring in 81.3% of runs with 48.46% significance. However, as previously mentioned for the High Plains, including $dSubsurface_{S_y}$ introduces variability to the model, as demonstrated by the large standard deviations of coefficient values for $modelG_0$ compared to $model_0Gt$ in Table 3.8.

Predicted TDS $\hat{y}_p$ is moderately more correlated with the independent validation TDS subset y_v with median $R_{y_v,\hat{y}_p}$ values of 0.2558 for $model_0Gt$ and 0.2644 for $modelG_0$ in the Central Valley, and -6.40e-17 for $model_0Gt$ and 0.004675 for $modelG_0$ in the High Plains. The percent of resampling runs with statistically significant correlations also increases for $modelG_0$ by 12.3% in the Central Valley and 2.6% in the High Plains. The median $rmse_{y_v,\hat{y}_p}$ are very

close between the two models in the Central Valley, at 58.76 mg/L for $model_0Gt$ and 59.73 mg/L for $modelG_0$, but are an order of magnitude larger in the High Plains when including $dSubsurface_{S_y}$, at 44.70 mg/L for $model_0Gt$ and 484.19 mg/L for $modelG_0$. The median $meanbias_{y_v,\hat{y}_p}$ over all runs are 39.91 mg/L for $model_0Gt$ and 62.33 mg/L for $modelG_0$ in the Central Valley, and 27.48 mg/L for $model_0Gt$ and 176.99 mg/L for $modelG_0$ in the High Plains.

3.4.2.2 Regime models excluding and including $dSubsurface_{S_y}$

It is also of interest to test whether adding GRACE $dSubsurface_{S_y}$ as a predictor improves lithology, land use regime models. We compare basic regime models, $lith/land_0Gt$, against the full regime models, $lith/land$, excluding and including $dSubsurface_{S_y}$ as a predictor, respectively. Again, both sets of models are limited to the GRACE period so differences can be accurately attributed to the inclusion (or exclusion) of $dSubsurface_{S_y}$ as a predictor.

In the Central Valley, the average number of observation months N and σ_N are the same between $lith/land_0Gt$ and $lith/land$ for all regimes. Models 2 and 35 are the only two regimes to have N over 130. Nine models (8, 9, 12, 15, 37-39, 43, 45) have at least half of the possible N between 66.3 ($\sigma_N = 2.89$) and 80.1 ($\sigma_N = 2.72$). Thirteen models (1, 5, 11, 14, 16, 18, 19, 22, 30, 31, 36, 40, 42) have small average N below 10. Model 35 has a higher median $R_{y,\hat{y}}$ in $lith/land$ of 0.3691 than in $lith/land_0Gt$ of 0.3571, both significant for 100% of runs. Twenty-one $lith/land$ models (1-3, 6-9, 12, 14, 15, 19, 21, 25-27, 31, 33, 37, 43, 45, 46) have higher median $R_{y,\hat{y}}$ and higher significance rates than the $lith/land_0Gt$ versions. Models 40 and 44 have higher $R_{y,\hat{y}}$ values for $lith/land$, but lower significance rates. Five models (16, 20, 24, 28, 34) have lower $R_{y,\hat{y}}$ for $lith/land$, but higher significance rates. The remaining fourteen models have $lith/land$ median $R_{y,\hat{y}}$ and significance rates that are less than or equal to the $lith/land_0Gt$ versions.

Twenty Central Valley *lith/land* models (1, 2, 5, 7, 8, 12, 15, 18, 19, 21, 22, 26-28, 30, 33, 35, 43, 45, 46) have smaller median $rmse_{y,\hat{y}}$ values than in $lith/land_0Gt$ between 30.9 for model 33 and 1.03e3 for model 36. Model 36 $rmse_{y,\hat{y}}$ is the same for $lith/land_0Gt$ and *lith/land*. The remaining twenty-two models have a higher $rmse_{y,\hat{y}}$ for *lith/land* than for $lith/land_0Gt$, with four models (9, 23, 38, 39) having unrealistically large values for *lith/land* on the order of 10^{76}. Only three models (5, 6, 10) have smaller magnitude $meanbias_{y,\hat{y}}$ for *lith/land* than for $lith/land_0Gt$ of -28.89, 1.13, and 12.37 respectively. Model 36 maintains the same $meanbias_{y,\hat{y}}$ for $lith/land_0Gt$ and *lith/land*. The remaining thirty-nine models have higher magnitude $meanbias_{y,\hat{y}}$ medians for *lith/land*, with four models (9, 23, 38, 39) having unrealistic values on the order of 10^{76}.

Ten *lith/land* models (3, 9, 10, 25, 26, 28, 32, 34, 43, 46) in the Central Valley have smaller σ_{F-stat} and higher F-statistic significance rates than the $lith/land_0Gt$ versions. Six additional *lith/land* models (6, 7, 21, 23, 39, 40) have smaller σ_{F-stat} and higher F-statistic significance rates than the $lith/land_0Gt$ versions, but the σ_{F-stat} values are larger than the median F-statistic values. Model 35 has a smaller σ_{F-stat} for *lith/land* than for $lith/land_0Gt$ with a high F-statistic significance rate of 99.7% for both versions. Five models (8, 13, 20, 24, 29) have larger or equal σ_{F-stat} values and smaller or equal significance rates for *lith/land* versus $lith/land_0Gt$. Twelve models (1, 2, 14-16, 22, 27, 31, 33, 37, 44, 45) have larger σ_{F-stat} for *lith/land* than for $lith/land_0Gt$, but only four of those models (2, 14, 44, 45) have σ_{F-stat} values smaller than the F-statistic in both cases. The nine remaining models (5, 11, 12, 18, 19, 30, 36, 38, 42) have σ_{F-stat} values at least two orders of magnitude larger than the median F-statistic value in both *lith/land* and $lith/land_0Gt$ versions, with the exception of $lith/land_0Gt$ model 12.

In the High Plains, the average number of observation months N and σ_N are the same between $lith/land_0Gt$ and *lith/land* for all regimes. Again, the number of observations during the GRACE period in the High Plains are low, with average N between 1 and 33.7

out of 132 possible months. Seven $lith/land$ models (3, 16, 18, 25, 27, 31, 33) have higher median $R_{y,\hat{y}}$ values and significance rates than the $lith/land_0Gt$ counterparts. Ten $lith/land$ models (6, 14, 15, 17, 19, 20-23, 26) have smaller $R_{y,\hat{y}}$ and lower significance rates compared to the $lith/land_0Gt$ versions. Two of the remaining models, 7 and 28, have lower $lith/land$ $R_{y,\hat{y}}$ values with significance rates that stay the same and increase, respectively. The remaining three models (1, 4, 5) have an average N between 1 and 1.76 months, and therefore a value of $R_{y,\hat{y}}$ can not be quantified. These three models will be ignored for this portion of the analysis.

Nine High Plains $lith/land$ models (3, 16, 18, 22, 23, 25, 27, 31, 33) have smaller median $rmse_{y,\hat{y}}$ values compared to $lith/land_0Gt$, with reductions ranging from 0.42 ($lith/land$ $rmse_{y,\hat{y}} = 3.56$) for model 27 to 123.1 ($lith/land$ $rmse_{y,\hat{y}} = 892.6$) for model 3. The other thirteen $lith/land$ models (1, 4-7, 14, 15, 17, 19-21, 26, 28) have increased $rmse_{y,\hat{y}}$ compared to the $lith/land_0Gt$ versions by between 1.4 for model 7 to 834.6 for model 20. Only two models (20, 27) have smaller median $meanbias_{y,\hat{y}}$ magnitudes for $lith/land$ of -41.03 and 1.36e-10 respectively than for the $lith/land_0Gt$ versions. The remaining twenty models have larger magnitude $meanbias_{y,\hat{y}}$ for $lith/land$ with values between 4.06 for model 14 and 1.37e4 for model 3.

High Plains model 25 is the only model that has a larger F-statistic value and smaller σ_{F-stat} of 4.77 ($\sigma_{F-stat} = 2.07$) in $lith/land$ versus 3.49 ($\sigma_{F-stat} = 2.22$) in $lith/land_0Gt$, suggesting more strongly against the constant model with lower variability of model fit. The F-statistic and σ_{F-stat} increase from the basic $lith/land_0Gt$ for model 18 from 3.96 ($\sigma_{F-stat} = 2.74$) to 4.40 ($\sigma_{F-stat} = 3.38$) for $lith/land$. All other models have a σ_{F-stat} value larger than the F-statistic value in both $lith/land_0Gt$ and $lith/land$ versions.

In the Central Valley, five $lith/land$ models (6, 16, 22, 25, 40) have higher β_0 significance rates than in $lith/land_0Gt$ between 95.0% and 100%. Twenty-seven $lith/land$ models (1, 3, 6-13, 16, 18, 20, 22, 23, 25, 26, 28, 29, 31-33, 35, 38-40, 43) have higher β_{x_1} significance

rates than in $lith/land_0Gt$ between 1.99% and 100%. Twenty-three $lith/land$ models (1-3, 6, 8-12, 16, 18-23, 28, 31-34, 39, 40) have higher β_{x_2} significance rates than in $lith/land_0Gt$ between 2.33% and 100%. Nineteen $lith/land$ models (1, 3, 6-8, 10, 18, 20, 23, 25, 26, 28, 31-34, 39, 43, 44) have higher $\beta_{x_1^2}$ significance rates than in $lith/land_0Gt$ between 6.12% and 97.8%. Nineteen $lith/land$ models (1, 3, 6, 8, 10, 12, 15, 20, 23, 25, 27, 31, 32, 33, 34, 35, 38, 39, 44) have higher $\beta_{x_1x_2}$ significance rates than in $lith/land_0Gt$ between 7.81% and 95.08%.

The intention of comparing $lith/land_0Gt$ models to $lith/land$ models is to determine whether including $dSubsurface_{S_y}$ terms improves TDS characterization models. Only Central Valley $lith/land$ model 2 has a β_{x_3} value larger than $iqr_{\beta_{x_3}}$ of 0.0015 and 2.5e-4 respectively, and a high significance rate of 77.1% occurring in 14% of runs. No other $lith/land$ models have β_{x_3} values on reasonable magnitudes or that are larger than $iqr_{\beta_{x_3^2}}$ and a 70% significance rate. Models 9, 23, 38, and 39 have high $\beta_{x_3^2}$, $\beta_{x_1x_3}$, and $\beta_{x_2x_3}$ significance rates for $lith/land$, but all interquartile ranges are over 10^{70} in magnitude. Models 19 and 21 have $\beta_{x_2x_3}$ significance rates over 70% with values of -0.0359 ($iqr_{\beta_{x_2x_3}} = 0.0064$) and 0.0016 ($iqr_{\beta_{x_2x_3}} = 4.05e-4$) respectively.

In the High Plains, ten $lith/land$ models (6, 7, 14, 15, 17, 19, 20, 21, 22, 27) have higher β_0 significance rates than in $lith/land_0Gt$ between 51.3% and 93.8%. Fifteen $lith/land$ models (6, 7, 14-16, 19-21, 23, 25-28, 31, 33) have higher β_{x_1} significance rates than in $lith/land_0Gt$ between 14.9% and 79.6%. Seven $lith/land$ models (6, 16, 23, 25, 26, 28, 31) have higher β_{x_2} significance rates than in $lith/land_0Gt$ between 8.6% and 65.5%. Nine $lith/land$ models (6, 16, 18, 23, 25, 26, 28, 31, 33) have higher $\beta_{x_1^2}$ significance rates than in $lith/land_0Gt$ between 16.2% and 66.5%. Six $lith/land$ models (3, 6, 18, 25, 28, 33) have higher $\beta_{x_1x_2}$ significance rates than in $lith/land_0Gt$ between 6.9% and 41.2%.

Only High Plains $lith/land$ models 26 and 31 have β_{x_3} terms with significance rates of at least 70% of -0.0351 ($iqr_{\beta_{x_3}} = 0.0166$) and -0.05 ($iqr_{\beta_{x_3}} = 0.0187$) respectively. No other $dSubsurface_{S_y}$ terms have significance rates of at least 70% in the High Plains $lith/land$

Table 3.10: Summary of model prediction statistics for $lith/land_0Gt$ and $lith/land$ regime models over the GRACE period, excluding and including $dSubsurface_{S_y}$ as a potential predictor respectively. Reported values are a median of the 1,000 Monte Carlo resampling runs. Units for $rmse_{y_v,\hat{y}_p}$ and $meanbias_{y_v,\hat{y}_p}$ are mg/L.

Statistic	Central Valley				High Plains	
	$lith/land_0Gt$ 2	$lith/land$ 2	$lith/land_0Gt$ 35	$lith/land$ 35	$lith/land_0Gt$ 31	$lith/land$ 31
$R_{y_v,\hat{y}_p}$ (% small p)	0.2159 (0.446)	0.2756 (0.811)	0.279 (0.751)	0.2579 (0.746)	0 (0.005)	0.2529 (0.076)
$rmse_{y_v,\hat{y}_p}$	33.8	34.45	184.7	189.25	97.6	110.8
$meanbias_{y_v,\hat{y}_p}$	28.6	32.46	79.8	132.1	41.1	73.1

models.

In the Central Valley, fourteen $lith/land$ models (2, 3, 6, 12, 15, 21, 26, 27, 33, 37, 39, 44-46) have larger median $R_{y_v,\hat{y}_p}$ values and larger significance rates than the $lith/land_0Gt$ versions. Three models (19, 23, 25) have larger $R_{y_v,\hat{y}_p}$ values for $lith/land$ than $lith/land_0Gt$, but have smaller significance rates. Twelve models (1, 7, 9, 13, 16, 20, 28, 29, 31, 34, 40, 43) have larger $R_{y_v,\hat{y}_p}$ values but smaller significance rates for $lith/land$ versions. Eight models (8, 10, 18, 24, 32, 35, 36, 38) have smaller or equal $R_{y_v,\hat{y}_p}$ values and smaller or equal significance rates for $lith/land$ compared to $lith/land_0Gt$. The five remaining models (5, 14, 22, 30, 42) have significance rates of $R_{y_v,\hat{y}_p}$ of 0 for both $lith/land$ and $lith/land_0Gt$ versions.

Seven Central Valley $lith/land$ models (16, 22, 26, 30, 37, 40, 46) have smaller median $rmse_{y_v,\hat{y}_p}$ values than the $lith/land_0Gt$ versions albeit small, with the largest reduction for model 22 of 15.15. Model 36 has the same median $rmse_{y_v,\hat{y}_p}$ for $lith/land$ and $lith/land_0Gt$ of 1.2e3 mg/L. The remaining thirty-five models (1-3, 5-15, 18-21, 23-25, 27-29, 31-35, 38, 39, 42-45) have larger $rmse_{y_v,\hat{y}_p}$ for $lith/land$ than $lith/land_0Gt$, with four models (9, 23, 38, 39) having unrealistic $rmse_{y_v,\hat{y}_p}$ for $lith/land$ on the order of 10^{76}. Ten models (3, 5, 7, 10, 11, 13, 16, 31, 34, 37) have smaller $meanbias_{y_v,\hat{y}_p}$ magnitudes for $lith/land$ than $lith/land_0Gt$, with improvements between 0.52 for model 31 and 24.3 for model 5. Model 3 is the only model with an improved $meanbias_{y_v,\hat{y}_p}$ to switch signs, going from 43.61 in $lith/land_0Gt$ to

-14.15 in $lith/land$. The remaining thirty-two models have larger $meanbias_{y_v,\hat{y}_p}$ magnitudes for $lith/land$ than $lith/land_0Gt$, with four models (9, 23, 38, 39) having unrealistic values for $lith/land$ on the order of 10^{76}.

In the High Plains, only models 26 and 31 have median $R_{y_v,\hat{y}_p}$ values improve from 0.0015 (2.8% significance) and 0 (0.5% significance) for $lith/land_0Gt$ to 0.0372 (3.9% significance) and 0.2529 (7.6% significance) for $lith/land$. The remaining models have weakened $R_{y_v,\hat{y}_p}$ magnitudes for $lith/land$ versus $lith/land_0Gt$. Only two models (15, 22) have reduced $rmse_{y_v,\hat{y}_p}$ from $lith/land_0Gt$ to $lith/land$, with model 15 decreasing from 280.0 to 235.84, and model 22 from 254.88 to 214.64. The remaining twenty models have increased $rmse_{y_v,\hat{y}_p}$ from $lith/land_0Gt$ to $lith/land$. Nine models (1, 7, 14, 15, 21, 22, 23, 25, 27) have reduced median $meanbias_{y_v,\hat{y}_p}$ magnitudes from $lith/land_0Gt$ to $lith/land$ between 6.15 for model 27 and 131.5 for model 15. Thirteen models (3-6, 16-20, 26, 28, 31, 33) have increased median $meanbias_{y_v,\hat{y}_p}$ magnitudes from $lith/land_0Gt$ to $lith/land$ between 17.24 for model 19 and 6.9e7 for model 28.

3.5 Discussion

It is clear that some models are unrealistic with median statistics on very large orders of magnitude, specifically the Central Valley $lith/land_0$ regime models 9, 23, and 39, and $lith/land$ regime models 9, 23, 38, and 39. These regimes represent small proportions of the total area, between 0.246% for model 23 and 2.4% for model 38, but are not the smallest areal representation compared to other regimes. Model 23 has relatively few average months of *in situ* observations of 17.55 during the GRACE period, and models 9, 38, and 39 have a moderate number of average months of *in situ* observations between 66.27 and 80.14, so it is likely that the available data are not sufficient to accurately constrain average TDS. These models will be excluded from the remaining analysis.

3.5.1   Regime models

To determine whether regime categorization is advantageous, we first compare the basic $model_0$ against the basic regime $lith/land_0$ models for the full *in situ* record.

In the Central Valley, $model_0$ and $lith/land_0$ 35 have the highest significance rates of correlation between validating TDS subsets and predicted TDS over Monte Carlo resampling, but the median $R_{y_v,\hat{y}_p}$ value for $model_0$ is larger than $lith/land_0$ 35 suggesting the basin-wide model better captures variability. Other $lith/land_0$ models have higher median $R_{y_v,\hat{y}_p}$ values than $model_0$ in the Central Valley such as $lith/land_0$ 25 and 27, but with much lower significance rates it is prudent to scrutinize the predictive capabilities of such models. Overall, the median $R_{y_v,\hat{y}_p}$ values of $model0$ and $lith/land_0$ models are generally low, indicating additional processes or parameters are important to average TDS characterization, as is expected.

The basic $model_0$ has a small median $rmse_{y_v,\hat{y}_p}$ between validating TDS and predicted TDS, with only $lith/land_0$ 13 and 28 having smaller median values, suggesting $model_0$ is more accurately constrained than most $lith/land_0$ models. This is supported by the large average N for $model_0$ compared to $lith/land0$. The median $meanbias_{y_v,\hat{y}_p}$ magnitude for $model_0$ is positive and small, indicating a small overestimation of average TDS. Models 13 and 25 have smaller median $meanbias_{y_v,\hat{y}_p}$ magnitudes than $model_0$, and models 28 and 35 are relatively small on the same order of magnitude, all of which are positive.

For the full Central Valley *in situ* record, the basin-wide $model_0$ is better constrained than the $lith/land_0$ models, and so more accurately predicts average TDS with only ground-based groundwater levels and categorical season as explanatory variables. In the overwhelming majority of runs, the model form of Central Valley $model_0$ is a function of the intercept, x_1^2, and x_1x_2 terms. The intercept is significant at the 1% level for all runs, and $\beta_{x_1^2}$ and $\beta_{x_1x_2}$ are significant at the 5% level for most runs.

In the High Plains, $model_0$ has the highest significance rate of correlation between validating TDS subsets and predicted TDS over Monte Carlo resampling. Two $lith/land_0$ models, 1 and 22, have higher median $R_{y_v,\hat{y}_p}$ values than $model_0$, but even with the highest significance rates of the High Plains $lith/land_0$ models, they are much lower than the 99% significance rate of $model_0$. Two other $lith/land_0$ models, 6 and 25, have higher median $R_{y_v,\hat{y}_p}$ values than $model_0$, but again have relatively low significance rates. As with the Central Valley, the overall low median $R_{y_v,\hat{y}_p}$ values suggest additional information is necessary to characterize average TDS.

Basin-wide $model_0$ in the High Plains has a moderately small median $rmse_{y_v,\hat{y}_p}$ between validating TDS and predicted TDS. All five of the $lith/land_0$ models with smaller median $rmse_{y_v,\hat{y}_p}$ values have relatively few months of observations, with $lith/land_0$ 17 and 27 having fewer than 10 and models 7, 18, and 25 with average N between 77 and 103. Similarly, the eight models with smaller median $meanbias_{y_v,\hat{y}_p}$ magnitudes than $model_0$ have relatively small N, with $lith/land_0$ 4, 27, and 33 having average N fewer than 50, and $lith/land_0$ 5, 6, 7, 25, and 28 having average N between 77 and 130 over the entire record. The two $lith/land_0$ models with the highest average N in the High Plains, 22 and 31, have high $rmse_{y_v,\hat{y}_p}$ compared to $model_0$, but comparable $meanbias_{y_v,\hat{y}_p}$ magnitudes.

For the full High Plains *in situ* record, basin-wide $model_0$ has at least twice as many months of observations than any $lith/land_0$ model, and is therefore better constrained than the $lith/land_0$ models given *in situ* water levels and categorical season as predictor variables. In the large majority of runs, the High Plains $model_0$ equation is a function of the intercept, x_1, x_2, and x_1^2 terms. The intercept term is significant at the 1% level for all runs. The β_{x_1} term is significant at the 1% level for all runs in which the term is present (present in 94.7% of runs). The β_{x_2} term is significant at the 5% level for most (90.4%) runs when present (present in 92.8% of runs). The $\beta_{x_1^2}$ term is present in every run, and is significant at the 1% level for all runs.

Next, we compare basin-wide $modelG_0$ against the $lith/land$ models to determine whether regime categorization improves models which include GRACE $dSubsurface_{S_y}$.

In the Central Valley, the two $lith/land$ models (2, 35) covering the largest fractional areas have prediction statistics comparable to those of $modelG_0$. The $lith/land$ 35 model has a comparable median $R_{y_v,\hat{y}_p}$ value and significance rate to $modelG_0$, but larger median $rmse_{y_v,\hat{y}_p}$ and median $meanbias_{y_v,\hat{y}_p}$. The $lith/land$ 2 model prediction statistics are all improved over the basin-wide model.

It is difficult to determine whether there are improvements in average TDS in the High Plains $modelG_0$ compared to $lith/land$ models with so few observations during the GRACE period. The $modelG_0$ prediction statistics are not particularly high, with very small median $R_{y_v,\hat{y}_p}$ value and significance rate, and relatively large median $rmse_{y_v,\hat{y}_p}$ and median $meanbias_{y_v,\hat{y}_p}$.

Over the entire $in\ situ$ record, regime categorization does not appear to improve upon the basin-wide models in either the Central Valley or High Plains with only groundwater levels and categorical season as predictors. When focusing on the GRACE period and including $dSubsurface_{S_y}$ as a potential predictor, categorizing by regime results in Central Valley $lith/land$ model 2 more accurately predicting TDS than the basin-wide $modelG_0$, having all model prediction statistics improve. The $lith/land$ model 35 is comparable to $modelG_0$, predicting TDS with slightly less accuracy. Central Valley regime 2 is alluvium/developed, covers the largest percent area, and is centrally located running through the center of the study area. Regime 35 is sandstone/developed, covers the second-largest percent area in the Central Valley, and is mostly located on the southwestern border with some exceptions, always adjacent to at least one other sandstone/developed polygon. It is likely that some combination of these factors explain the prediction improvement. Covering the largest percent areas likely increases the number of $in\ situ$ observations available to constrain the model. Dominant lithology types alluvium and sandstone imply a generally larger grain size, which suggests a larger porosity and permeability and therefore hydraulic connectivity to

the surface. The developed dominant land use may suggest a link to anthropological runoff.

3.5.2 GRACE $dSubsurface_{S_y}$ as a predictor

The effects of including $dSubsurface_{S_y}$ as a possible predictor of large-scale TDS are isolated by comparing basin wide models $model_0Gt$ and $modelG_0$, each ignoring lithology, land use regime classifications.

Due to the sharp reduction of *in situ* TDS observations in the High Plains leading up to and during the GRACE period, it is difficult to accurately constrain the basic $model_0Gt$, let alone quantify improvement or degradation attributed to including $dSubsurface_{S_y}$ in $modelG_0$. Instead, we use the opportunity to compare relative variability. While the fit statistics of $model_0Gt$ suggest a moderate fit considering there are fewer than half of the possible months with a ground-based value, inconsistency of significant predictor terms selected and the poor prediction statistics values indicate the inability to accurately constrain the basic $model_0Gt$ in the High Plains given the available data. Including $dSubsurface_{S_y}$ as a potential predictor only increases model variability. Summarized in Table 3.8, the differences in average coefficient terms and relative standard deviations between $model_0Gt$ and $modelG_0$ succinctly demonstrate the additional variability introduced by including $dSubsurface_{S_y}$. With the exceptions of the categorical season and $dSubsurface_{S_y}$ linear terms, the term coefficient values and the standard deviations of $modelG_0$ are many orders of magnitude larger than the $model_0Gt$ counterparts.

Including $dSubsurface_{S_y}$ as an additional predictor in the Central Valley suggests an improvement in characterizing basin-averaged TDS. The fit and prediction statistics of $modelG_0$ are generally an improvement over those of $model_0Gt$. The $modelG_0$ form is somewhat more variable in which terms are present, but the average coefficient values of shared terms in $modelG_0$ are within one standard deviation of the estimate in $model_0Gt$, as shown in Table

3.8. It is also worth noting that the coefficient standard deviations in $modelG_0$ are each one order of magnitude smaller than the average coefficient estimate. Small standard deviations of coefficient values and a smaller standard deviation of F-statistic values for $modelG_0$ suggest low variability of model fit in the Central Valley.

To determine whether including $dSubsurface_{S_y}$ as a predictor of TDS improves lithology, land use regime models, the basic regime models limited to the GRACE period but excluding $dSubsurface_{S_y}$, $lith/land_0Gt$, are compared to the full regime models including $dSubsurface_{S_y}$ as a potential predictor, $lith/land$.

Again, the limited availability of High Plains TDS observations during the GRACE period pose an additional challenge to determining whether or not including $dSubsurface_{S_y}$ to regime models improves TDS prediction. Only High Plains $lith/land$ models 26 and 31 have improved correlations between TDS prediction and validating TDS, but the low significance rates indicate these improvements are not likely reliable. Only models 15 and 22 have smaller median $rmse_{y_v,\hat{y}_p}$ and median $meanbias_{y_v,\hat{y}_p}$ when including $dSubsurface_{S_y}$ as a possible predictor in the High Plains.

Including $dSubsurface_{S_y}$ as a potential predictor in Central Valley $lith/land$ models improves TDS prediction accuracy for models 37 and 46, with all model prediction statistics improving over the $lith/land_0Gt$ versions. Model 16 had smaller median $rmse_{y_v,\hat{y}_p}$ and median $meanbias_{y_v,\hat{y}_p}$ values and a larger $R_{y_v,\hat{y}_p}$, but the reduced significance rate of $R_{y_v,\hat{y}_p}$ for the $lith/land$ version implies the correlation is less reliable. Central Valley models 37 and 46 cover 4.34% and 1.63% of the total area, respectively, and have moderate to low average numbers of observation months of 69.97 and 36.2 respectively. The two models covering the largest area in the Central Valley (2, 35) have larger $R_{y_v,\hat{y}_p}$ and larger or comparable corresponding significance values for the $lith/land$ versions, but slightly larger median $rmse_{y_v,\hat{y}_p}$ values and larger median $meanbias_{y_v,\hat{y}_p}$ values. Although the TDS predictions improve compared to validating TDS data for models 2 and 35, the increased $rmse_{y_v,\hat{y}_p}$ and $meanbias_{y_v,\hat{y}_p}$

suggest that including $dSubsurface_{S_y}$ introduces additional variability to the models.

3.6   Conclusions

Over the full *in situ* record and excluding GRACE $dSubsurface_{S_y}$, the basin-wide $model_0$ is better constrained than the $lith/land_0$ models in the Central Valley, having the largest number of observational months for model construction, the highest correlation of predicted and validating TDS with the most consistent statistical significance, and among the lowest $rmse_{y_v, \hat{y}_p}$ and $meanbias_{y_v, \hat{y}_p}$. This is expected, as there is more data available to constrain the basin-wide $model_0$ than the $lith/land_0$ models.

During the GRACE period, the centrally located Central Valley alluvium/developed model including $dSubsurface_{S_y}$ more accurately predicts TDS concentrations than the basin-wide $modelG_0$. Covering the largest proportion of the study area, the alluvium/developed model likely has more *in situ* observations for model construction than the other $lith/land$ models. The developed land use category includes developed open space (largely lawn grasses), low, medium, and high intensity as defined by the National Land Cover Database, with impervious surfaces accounting for less than 20% to 100% of the total cover (Homer et al. [2015]). The specific regime categorization implies hydraulic connectivity between the surface and subsurface, and suggests a link between average TDS and anthropogenic runoff in these areas.

Including GRACE-derived $dSubsurface_{S_y}$ as a potential predictor of TDS in groundwater introduces additional variability into nearly all regime model predictions. This may be due to regimes areas well below the recommended spatial extent of gridded GRACE products (Landerer and Swenson [2012]). However, a GRACE product can provide additional information to characterizing TDS in groundwater on a large spatial scale. Including $dSubsurface_{S_y}$ as a potential predictor of TDS in the basin-wide Central Valley $modelG_0$ improves accuracy

of TDS prediction compared to the basin-wide model excluding $dSubsurface_{S_y}$, with low variability of model fit.

Generalized linear models can be powerful tools of characterizing groundwater quality, as in Shamsudduha et al. [2015]. But extensive, readily available information on hydrogeologic settings and a robust groundwater quality monitoring network are necessary to accurately constrain models and accurately attribute constituent events.

Future directions of this work may include studying TDS in groundwater for specific regime types, which may confirm (or reject) the possible explanations for improved model prediction given here. Future work may also pursue characterizing TDS in time as well as space, rather than just through time as in this work.

Chapter 4

Evaluation of two spatial analysis approaches on well and GRACE data

4.1   Introduction

In the previous chapter, Thiessen polygons in the Central Valley and High Plains aquifers are categorized by dominant lithology and dominant land use cover pairs to build regime models of average TDS concentrations. As noted in section 3.3.3, lithology, land use regimes are not restricted to adjacent Thiessen polygons. So redistribution of $dSubsurface_{S_y}$ into polygons and the subsequent categorization into regimes ignore the spatial covariance inherent in the GRACE gridded product (see section 1.2.1).

This chapter explores two approaches to address the spatial covariance caveat of the previous chapter. The first approach is based on a step of pre-processing gridded GRACE data which greatly increases the spatial covariance. This approach is referred to as the GRACE footprint. The second approach, empirical orthogonal function (EOF) decomposition, is a well-documented method of analyzing spatial and temporal variability (Bjornsson and Venegas [1997], Perry and Niemann [2007], Schmidt et al. [2008], Becker et al. [2010], Crossley et al. [2012], Scanlon et al. [2012b]).

4.2 GRACE footprints

4.2.1 Theoretical basis

Pre-processing of the GRACE CSR RL05 level 3 gridded product includes Gaussian filtering of total water storage anomalies, dS, and the accompanying CLM4 scaling factor for each grid cell with a radius of 300 km (Landerer and Swenson [2012]). Gaussian filtering effectively influences the value of one grid cell as dependent on all other values within the 300 km radius, with the values nearest the grid cell center of interest having the strongest influence. $1° \times 1°$ grid cells are approximately 111 km $\times$ 111 km at the equator, so the 300 km radius of influence for a $1° \times 1°$ grid cell includes the area covered by several neighboring grid cells depending on the latitude. We refer to the 300 km area of influence for a grid cell as the GRACE footprint.

If each GRACE $1° \times 1°$ grid cell value is influenced by all values within a 300 km radius, it stands to reason that *in situ* observations falling within a grid cell footprint are inherently part of that GRACE grid cell value. The goal of this experiment is to characterize average TDS within GRACE footprints as a function of *in situ* groundwater levels and dS in defining subbasin areas of interest.

Subbasin	Grid cell center	TDS sites	WL sites
North CV	(-122, 40)	10,409	25,717
South CV	(-120, 36)	16,456	32,768
North HP	(-101, 41)	2,650	14,412
Central HP	(-101, 37)	4,813	10,434
South HP	(-102, 34)	6,339	12,359

Table 4.1: Summary of GRACE $1° \times 1°$ selected grid cell center coordinates in the Central Valley (CV) and High Plains (HP) subbasins and the number of *in situ* well sites within each footprint over the full *in situ* record.

4.2.2   Methods: Footprint of a grid cell

(a) GRACE dS grid cells.

(b) Central Valley well sites.

(c) GRACE dS grid cells.

(d) High Plains well sites.
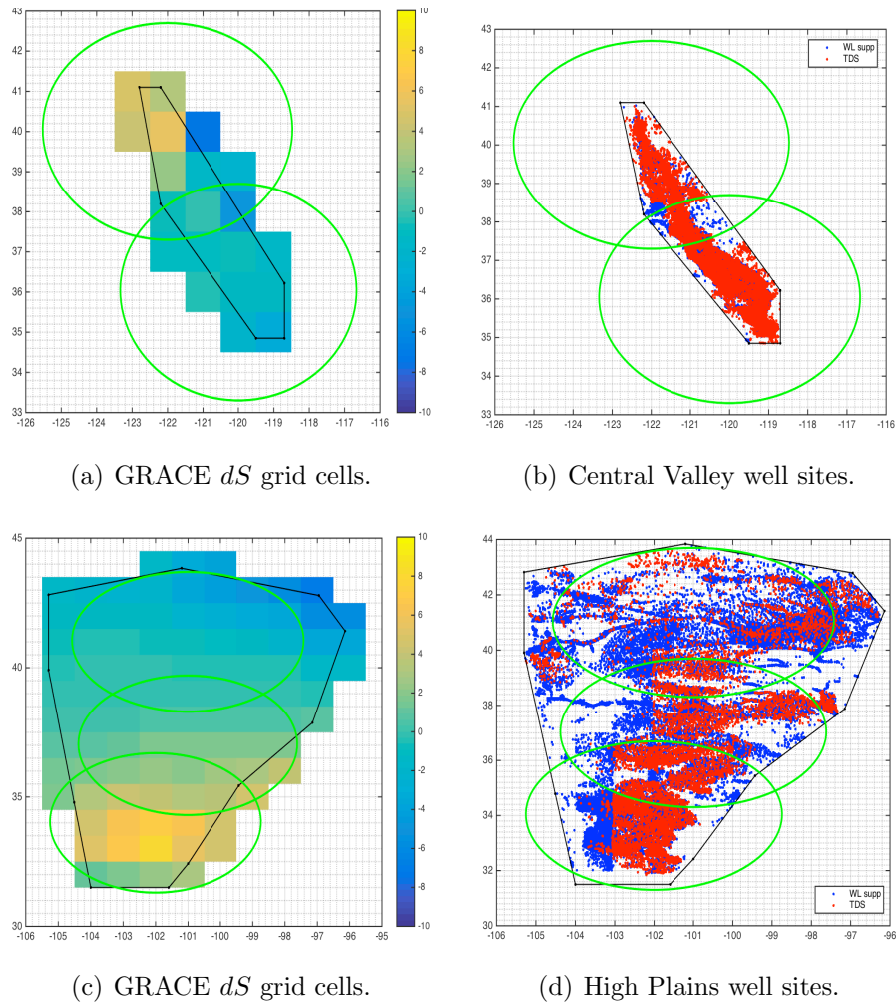
Figure 4.1: Maps of $1° \times 1°$ GRACE dS for January 2003, footprint areas, and *in situ* well sites for each subbasin in the Central Valley and High Plains.

The High Plains and Central Valley are separated into subbasins based on physical characteristics of local aquifer dynamics and climate (McMahon et al. [2007], Faunt [2009], Scanlon et al. [2012a]). The High Plains is divided into three subbasins, north, central, and south. The Central Valley covers a smaller area, divided into two subbasins in the north and south. A $1° \times 1°$ grid cell nearest the center of each subbasin is selected. The selected grid cells are chosen to minimize the footprint area overlapping neighboring subbasins and areas outside of the *in situ* database aquifer boundaries, to the extent possible. The selected grid cell

center coordinates are considered the center of each footprint area, summarized in Table 4.1. The area within a 300 km radius from each grid cell center is determined, shown in Figure 4.1.

Figure 4.1 depicts all grid cells for the Central Valley and High Plains for one month to illustrate the number of adjacent grid cells within a footprint radius. The dS time series for each GRACE footprint is simply the calculated dS value of the selected grid cell through time. All *in situ* well locations within each footprint are used to calculate monthly values for that footprint.

The *in situ* TDS and groundwater level sites located within each footprint area are identified by longitude and latitude decimal coordinates. For each footprint area, the mean and median of all ground-based observations occurring in a given calendar month are calculated for the full available record. Because there are no spatial delineations within each footprint to calculate area-weighted averages, the median value of *in situ* observations is taken to reduce the influence of potential outliers.
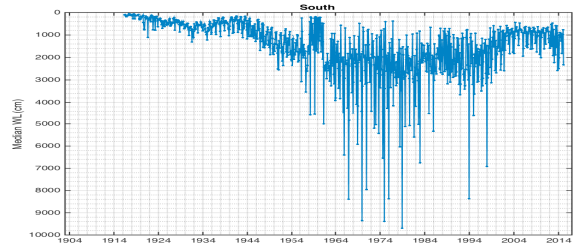
4.2.3   Results and discussion

The full *in situ* period records of the GRACE footprint areas are shown in Figures 4.2 and 4.3 for the Central Valley and High Plains respectively. The differences in groundwater levels between footprint subbasins is particularly clear, with the y-axis scales in both north footprints being substantially smaller than those further south. As expected from the climate gradients from north to south in both the Central Valley and High Plains, groundwater levels in the southern portions of each area reach deeper than those in the north.
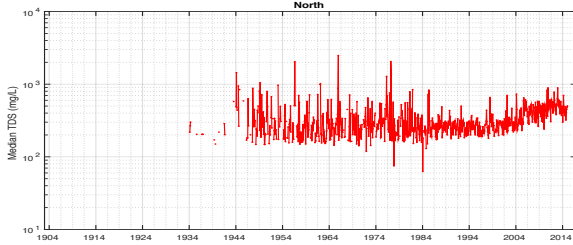
The differences in TDS between subbasins is less clear than groundwater levels. Again, the northern footprints in both areas have generally smaller TDS concentrations over the full record, with smaller y-axis scales than the subbasins further south. In the Central Valley,
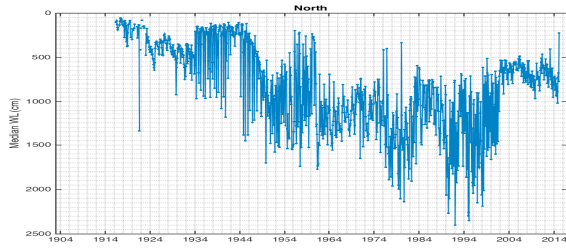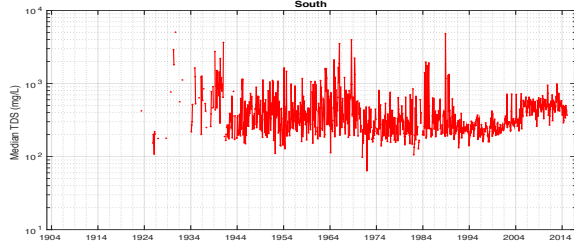
(a) N Central Valley WL   (b) S Central Valley WL
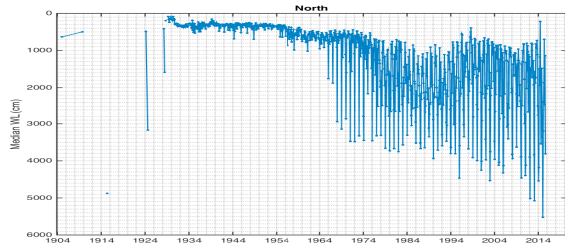
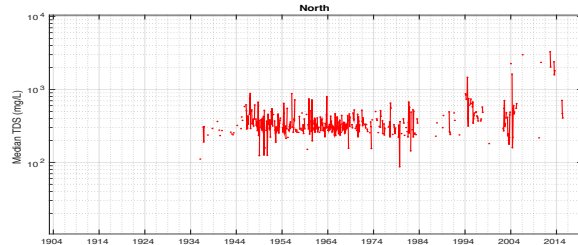(c) N Central Valley TDS   (d) S Central Valley TDS

Figure 4.2: Complete records of median *in situ* groundwater levels (WL) in cm below the land surface (4.2(a), 4.2(b)) and median total dissolved solid concentrations (TDS) in mg/L (4.2(c), 4.2(d)) for the north and south GRACE footprints in the Central Valley. Note the different y-axes scales for groundwater levels, and the semi-log axes for TDS.

median TDS in the north and south footprints both increase between 2004 and 2014, peaking under 1,000 mg/L in both subbasins. In the High Plains, median TDS in the central footprint appears to have the largest peak values occurring around 1964 and 1973. While there are a relatively large number of basin-wide observations during these periods for the High Plains (see Figure 2.2), the number of TDS observations in one month over the full record is generally under 200 with few exceptions, so it is possible that the median TDS peaks are the result of only a few observations. It is worth noting that there is some overlap between each of the subbasin footprint areas and therefore well observations in both locations.

Over the full record, the Kendall tau rank correlations between median TDS and median groundwater levels are small in magnitude, as summarized in Table 4.2. The correlations between TDS and groundwater levels for both the north and south Central Valley footprints over the full record are significant at the 1% level. Negative tau values suggest TDS concentrations increase with shallowing groundwater levels, likely indicating the influence of soil

104

(a) N High Plains WL

(b) N High Plains TDS
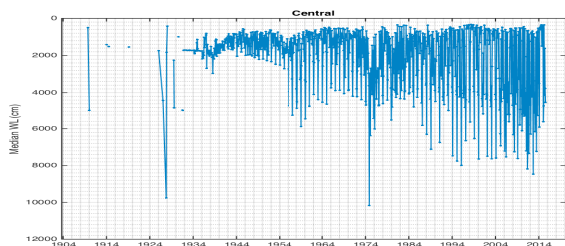
(c) C High Plains WL
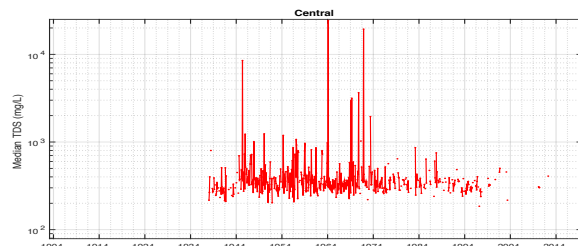
(d) C High Plains TDS
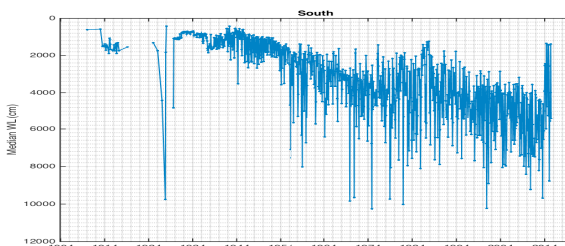
(e) S High Plains WL
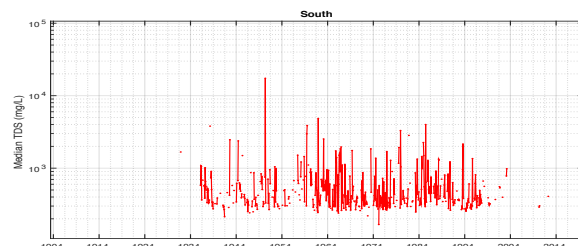
(f) S High Plains TDS

Figure 4.3: Complete records of median *in situ* groundwater levels (WL) in cm below the land surface (4.3(a), 4.3(c), 4.3(e)) and median total dissolved solid concentrations (TDS) in mg/L (4.3(b), 4.3(d), 4.3(f)) for the north, central, and south GRACE footprints in the High Plains. Note the different y-axes scales for groundwater levels, and the semi-log axes for TDS.
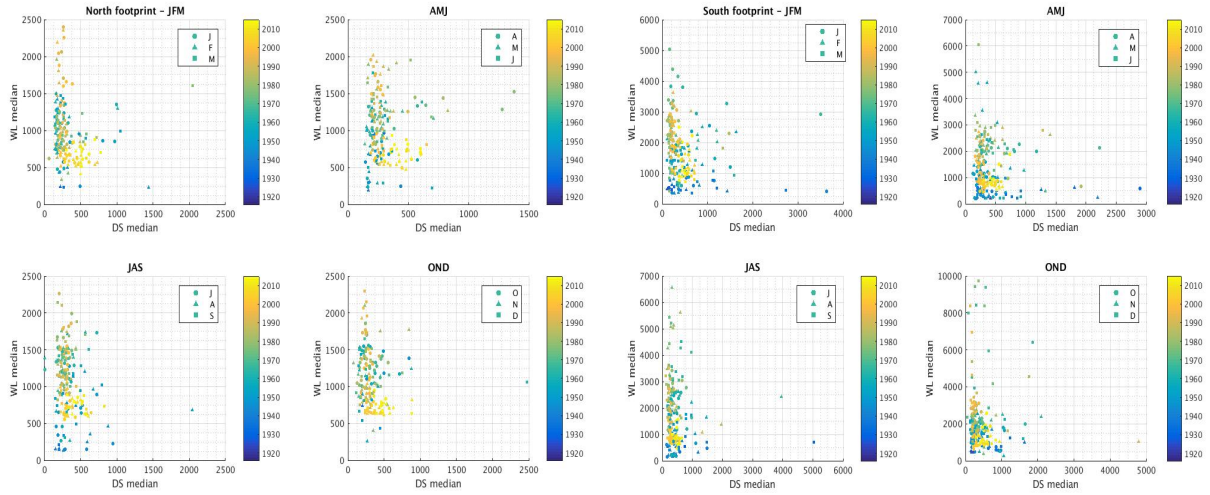
(a) N Central Valley   (b) S Central Valley

Figure 4.4: Scatter plots of median TDS in mg/L versus median groundwater level (WL) in cm below the land surface for the north and south GRACE footprints in the Central Valley separated by 3 month season. Colors indicate year, shapes indicate month. Note the different axes scales.
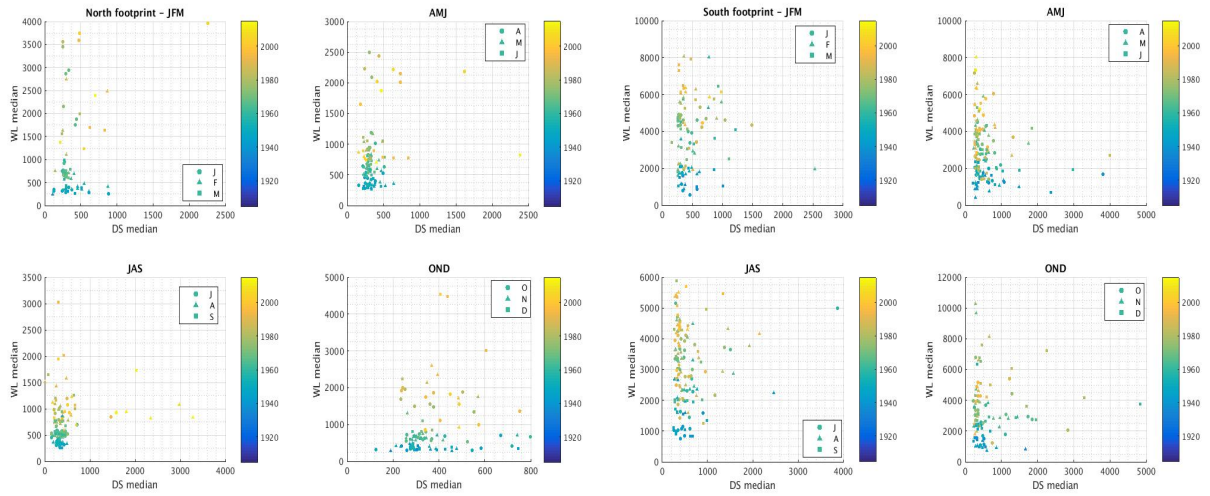
(a) N High Plains   (b) S High Plains

Figure 4.5: Scatter plots of median TDS in mg/L versus median groundwater level (WL) in cm below the land surface for the north and south GRACE footprints in the High Plains separated by 3 month season. Colors indicate year, shapes indicate month. Note the different axes scales.

salinization and irrigation (Deverel and Gallanthine [1989], Dubrovsky et al. [1993], Bexfield and Jurgens [2014]).

Correlations for the High Plains north and south footprints are significant at the 5% level. Although small in magnitude, the relative signs of tau for the north and south footprints are indicative of the hydrogeologic setting of each subbasin. The positive tau in the northern High Plains suggests TDS concentrations in groundwater increase, albeit slightly, with increasing groundwater level depth below the surface, as would be expected of a system relying more heavily on precipitation and surface water than pumping groundwater (McMahon et al. [2007], Gurdak et al. [2009]). As in the Central Valley, the negative sign of tau in the southern High Plains suggests shallower groundwater has a higher concentration of TDS, likely due to salinization, fertilizer application, and irrigation (McMahon et al. [2007], Chaudhuri and Ale [2014a,b]).

To determine whether any seasonal relationships exist, scatter plots of median TDS versus median groundwater level for the north and south footprint areas are shown in Figures 4.4 and 4.5, where each panel is a three-month season and colors indicate year. If a seasonal relationship existed, a pattern of median TDS against median groundwater level may be expected, and could change over time. In the Central Valley, there are no apparent patterns that could suggest a relationship as each of the panels have a spread of median groundwater levels and a generally limited range of median TDS concentrations. Nor is a change in relationship over time apparent, as there is no discernible pattern of color.

In the High Plains, both the north and south footprint areas show a general deepening of median groundwater levels over time, with more yellow points associated with larger magnitude median groundwater levels (in cm below the land surface) on the y-axes. An analogous pattern for median TDS over time is not apparent in most panels shown in Figure 4.5, but rather tends to stay within the same range as median groundwater levels deepen over time. In the summer months of July, August, and September (JAS), the north High

107

Plains footprint shows about seven points occurring after the year 2000 which have median TDS concentrations exceeding 1,000 mg/L. This appears to be a deviance from the general upper bound median concentration for the northern High Plains, with a few exceptions in the winter (JFM) and spring (AMJ) panels also occurring after the year 2000. The deviance from prior summer seasons suggests either a potential change in the median TDS-median groundwater level relationship, or more likely a specific event.

Subbasin	$T_{TDS,WL}$ (p-val)	$T_{TDS,dS}$ (p-val)	$T_{WL,dS}$ (p-val)
North CV	-0.1304 (5.97e-8)	0.091 (0.123)	-0.419 (2.6e-7)
South CV	-0.1475 (2.5e-11)	0.059 (0.32)	-0.0997 (0.09)
North HP	0.0798 (0.02)	-0.083 (0.56)	0.059 (0.31)
Central HP	0.0044 (0.88)	0.067 (1)	0.012 (0.84)
South HP	-0.0594 (0.05)	0.20 (0.82)	-0.062 (0.29)

Table 4.2: Kendall tau rank correlations and associated p-values of median monthly *in situ* data for GRACE footprints. $T_{TDS,WL}$ is the correlation for the median TDS and median WL for the full period, $T_{TDS,dS}$ is for the median TDS and dS for the GRACE period, and $T_{WL,dS}$ is for the median WL and dS for the GRACE period.

(a) N Central Valley   (b) S Central Valley
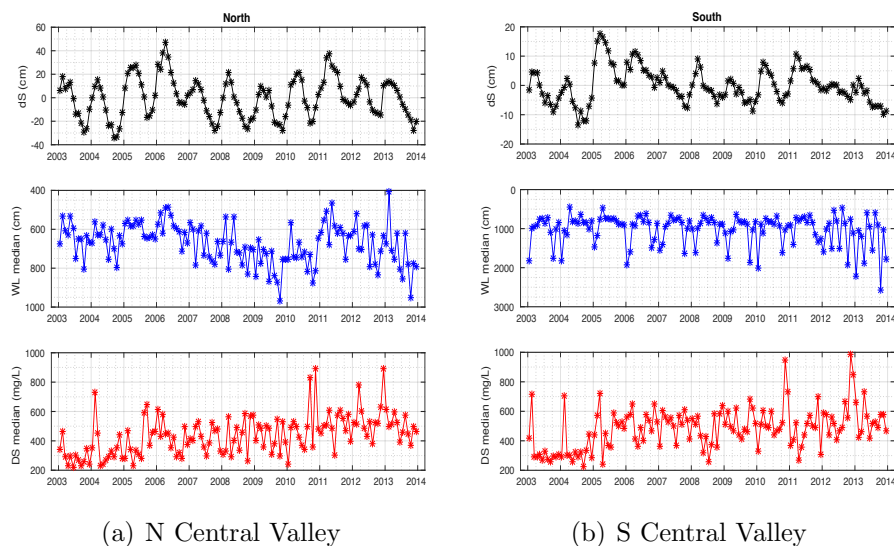
Figure 4.6: GRACE-period records of dS in cm, median *in situ* groundwater levels (WL) in cm below the land surface, and median total dissolved solid concentrations (TDS) in mg/L for the north and south GRACE footprints in the Central Valley. Note the different y-axes scales.

Focusing on the GRACE period between January 2003 and December 2014, as expected, there are no statistically significant correlations between median TDS and GRACE dS, as
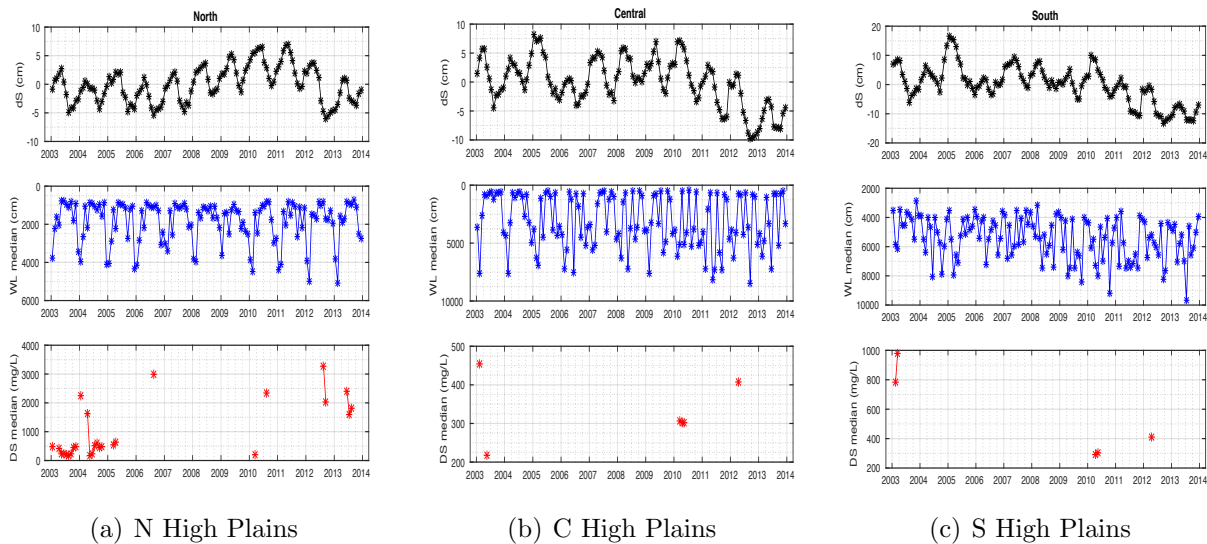
Figure 4.7: GRACE-period records of *dS* in cm, median *in situ* groundwater levels (WL) in cm below the land surface, and median total dissolved solid concentrations (TDS) in mg/L for the north, central, and south GRACE footprints in the High Plains. Note the different y-axes scales.

shown in Table 4.2. Median TDS appears to increase gradually over the GRACE period in both subbasins of the Central Valley, but as expected are not directly related to total storage anomalies. As discussed in section 2.2.1 and shown in Figure 2.2, TDS observations in the High Plains decline dramatically before the GRACE period. The central and south High Plains footprints have only about five months of TDS values between 2003 and 2014 despite having more TDS well sites than the north footprint (see Table 4.1). Although reported in Table 4.2, drawing conclusions about the relationship between median TDS and groundwater levels or *dS* during the GRACE period is impractical.

There appears to be some agreement between *dS* and median groundwater levels in the northern Central Valley during the GRACE period in Figure 4.6(a) beginning around 2006, supported by a correlation significant at the 1% level. California experienced periods of droughts beginning around 2007, with decreased precipitation, and therefore snow melt and stream flow. As expected, an increased reliance on pumping groundwater in response to drought conditions result in groundwater level declines during this period, and are reflected

in dS in the north Central Valley. The southern Central Valley dS fluctuates less than the north with less precipitation and stronger dependence on surface water deliveries. The effects of drought on dS in the southern Central Valley are discernible, particularly in 2007, 2009, and 2012, but there does not appear to be a similar fluctuation in median groundwater levels in Figure 4.6(b) until 2014 when both dS and groundwater levels decline. A notable pattern of groundwater levels in the southern Central Valley is a sharp decline near the end of each calendar year. This is very likely due to the reduction of surface water deliveries, which generally taper off in volume in September or October (California Department of Water Resources [b]), driving a return to groundwater pumping.

There are no apparent correlations between median groundwater level and dS in the High Plains subbasins during the GRACE period, with small tau magnitudes and large associated p-values. Median groundwater levels in the north and central High Plains do not appear to reflect any patterns shown in dS, but rather fluctuate with some regularity, likely due to stronger dependence on summer monsoon precipitation and available surface water resources in these subbasins. Median groundwater levels in the southern High Plains are generally deeper than in the north and central subbasins, with the shallowest median level around 3,000 cm below the land surface. The monthly median groundwater level in the southern High Plains (Figure 4.7(c)) is noisy, but the upper and lower limits appear to deepen particularly after 2011, when the value of dS is almost consistently negative.

4.3 EOF analysis

4.3.1 Theoretical basis: Spatial and temporal analysis

Empirical orthogonal function (EOF) decomposition is a statistical analysis method which can isolate both temporal and spatial patterns. EOF analysis is a well-documented approach

in the Earth sciences (e.g. Deverel and Gallanthine [1989], Bjornsson and Venegas [1997], Perry and Niemann [2007], Taschetto and England [2009], Park and Sohn [2010], Dewitte et al. [2012]), performed on GRACE total water storage anomalies (Schmidt et al. [2008], Becker et al. [2010]), and has been used to bridge spatial scales between ground observations and GRACE data (Crossley et al. [2012], Scanlon et al. [2012b]).

EOF analysis is performed on TDS, groundwater levels, and GRACE $dSubsurface$ each as individual fields, rather than on the coupled variability of two fields to better understand large-scale TDS variability and identify potential similarities in groundwater levels and $dSubsurface$. Coupled field analyses are challenging to interpret because only modes of variability in which the two fields are strongly coupled are identified (Bjornsson and Venegas [1997]). Coupled field analysis of TDS with groundwater levels or $dSubsurface$ may misdirect or entirely overlook large-scale variability patterns of TDS in space and time, so this work will focus on analysis of individual fields.

The goal of this experiment is to determine whether similarities in variability patterns of ground-based TDS in groundwater and GRACE $dSubsurface$ can be identified using EOF decomposition, which may be indicative of potential underlying physical relationships. EOF analysis requires that observations are available at the same locations for every time step (Bjornsson and Venegas [1997], Perry and Niemann [2007]). As discussed in section 2.2.1 and shown in Figure 2.2, the spatial and temporal availability of the compiled TDS data is much more variable in the High Plains than the Central Valley. The remaining analysis will therefore focus only on the Central Valley. Unfortunately, results of an EOF decomposition are domain-specific, so any potential pattern similarities or physical relationships between TDS and GRACE $dSubsurface$ can not necessarily be translated to other areas of interest.

4.3.2 Methods

To meet the requirement that observations are present at the same locations for every time step, the *in situ* TDS and groundwater level data are spatially interpolated to $1° \times 1°$ grid cells using inverse distance weighting, as described in section 2.3.3 (Bjornsson and Venegas [1997], Perry and Niemann [2007], Sahoo et al. [2016], Richey et al. [2016], Thomas et al. [2016]). The gridded TDS and groundwater level data are trimmed to the GRACE period, January 2003 to December 2013, which fortunately is the period when the spatial extent of the gridded TDS interpolant is consistently largest. Gridded TDS and water level data are linearly interpolated through time only if three or fewer consecutive months are absent for a given grid cell. The data are then trimmed in space to maintain only grid cells that have a value for each of the 132 possible GRACE months. The gridded TDS and groundwater level data are then converted into spatial anomalies. GRACE $dSubsurface$ is calculated as described in section 2.2.2 for the Central Valley. Anomalies of *in situ* TDS, groundwater levels, and $dSubsurface$ are each regridded to two-dimensions, time and space, on which the EOF analysis is performed. The resulting principal components and eigenvector spatial patterns are then converted back to the original spatial and temporal dimensions. We perform the analysis for the first three modes, which by definition account for the most variability.

4.3.3 Results and discussion

Figure 4.8 shows the first three modes of principal components of TDS, groundwater levels, and $dSubsurface$, with Figure 4.9 showing the corresponding first three modes of eigenvector spatial patterns. The eigenvector spatial patterns represent a standing oscillation with the principal components representing that oscillation through time for the corresponding mode (Bjornsson and Venegas [1997]). The areas in Figure 4.9 with the largest magnitude values are where the standing oscillation fluctuates most for that mode.
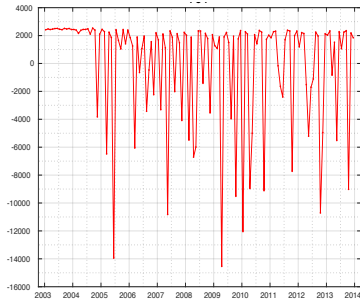
(a) TDS PC 1.
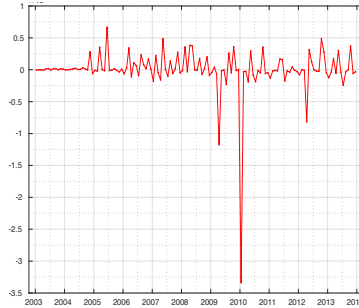
(b) TDS PC 2*.
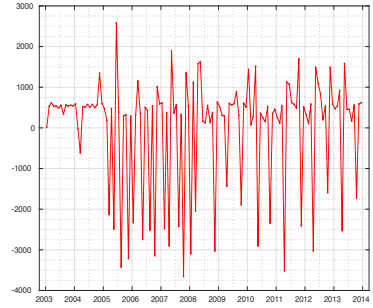
(c) TDS PC 3.
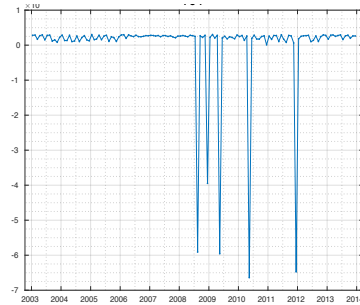
(d) WL PC 1*.
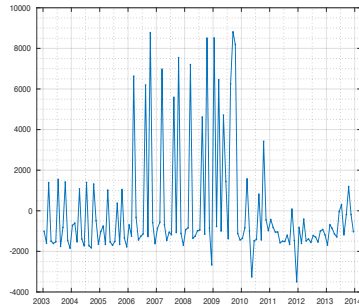
(e) WL PC 2.
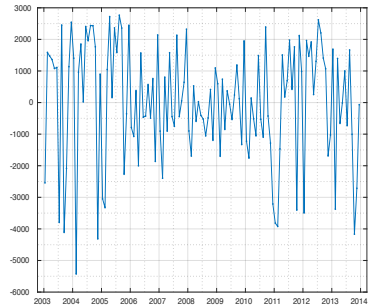
(f) WL PC 3.
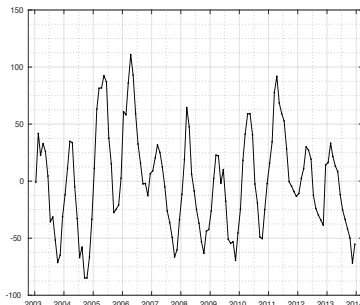
(g) *dSubsurface* PC 1.
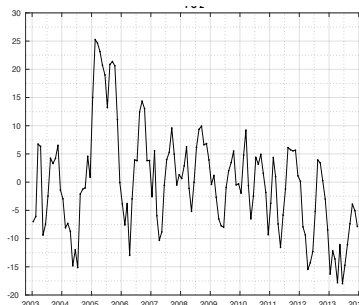
(h) *dSubsurface* PC 2.
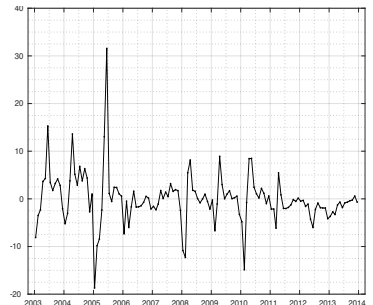
(i) *dSubsurface* PC 3.

Figure 4.8: Principal components of the first 3 modes of gridded *in situ* TDS anomalies, groundwater levels, and GRACE *dSubsurface* in the Central Valley. Note that PCs denoted by * have y-axes on the order of 10^4.

(a) TDS EOF 1. (b) TDS EOF 2. (c) TDS EOF 3.

(d) WL EOF 1. (e) WL EOF 2. (f) WL EOF 3.

(g) *dSubsurface* EOF 1. (h) *dSubsurface* EOF 2. (i) *dSubsurface* EOF 3.
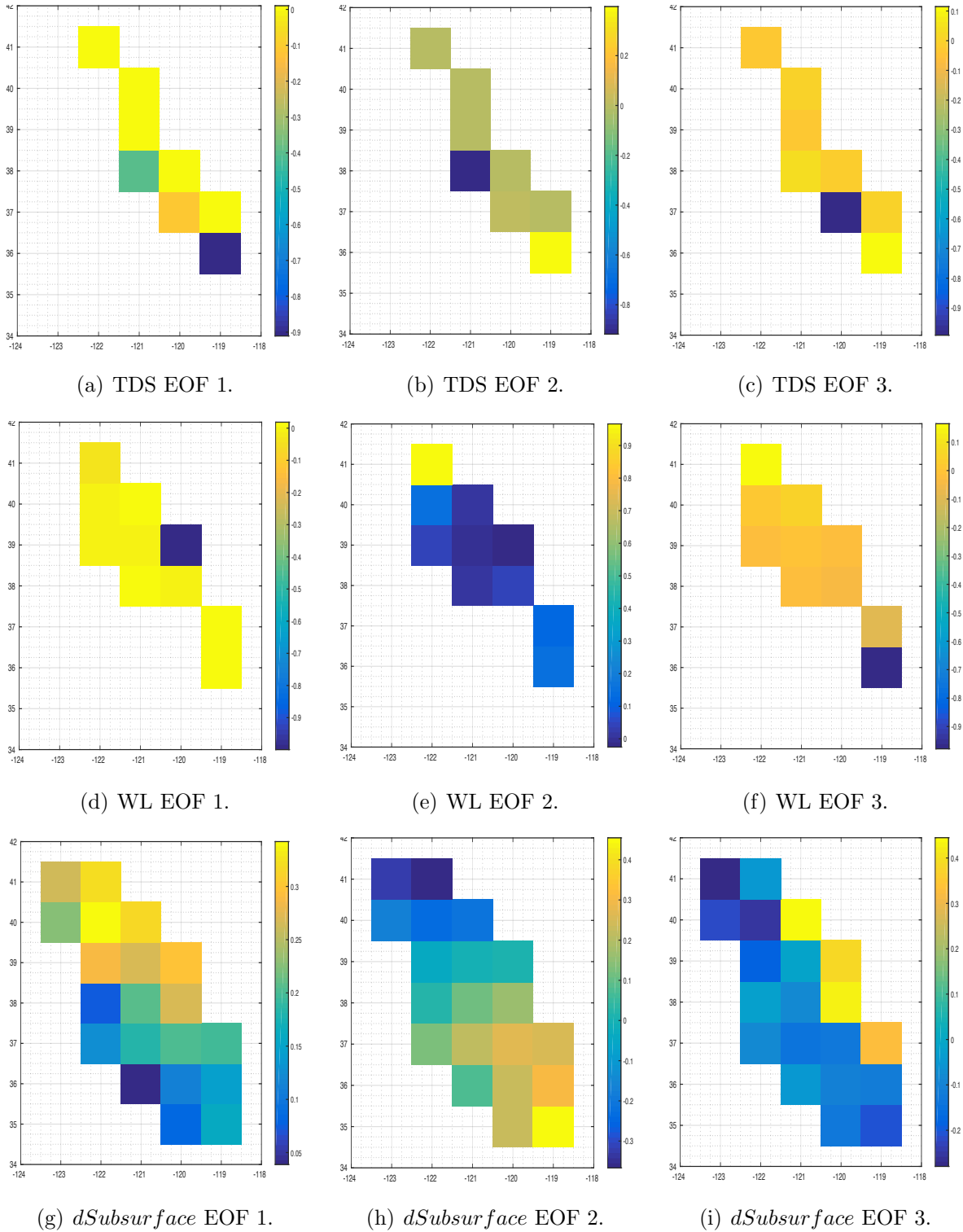
Figure 4.9: Eigenvectors of the first 3 modes of gridded *in situ* TDS anomalies, groundwater level anomalies, and GRACE *dSubsurface* in the Central Valley.

(a) TDS mode 1. (b) TDS mode 2. (c) TDS mode 3.

(d) WL mode 1. (e) WL mode 2. (f) WL mode 3.
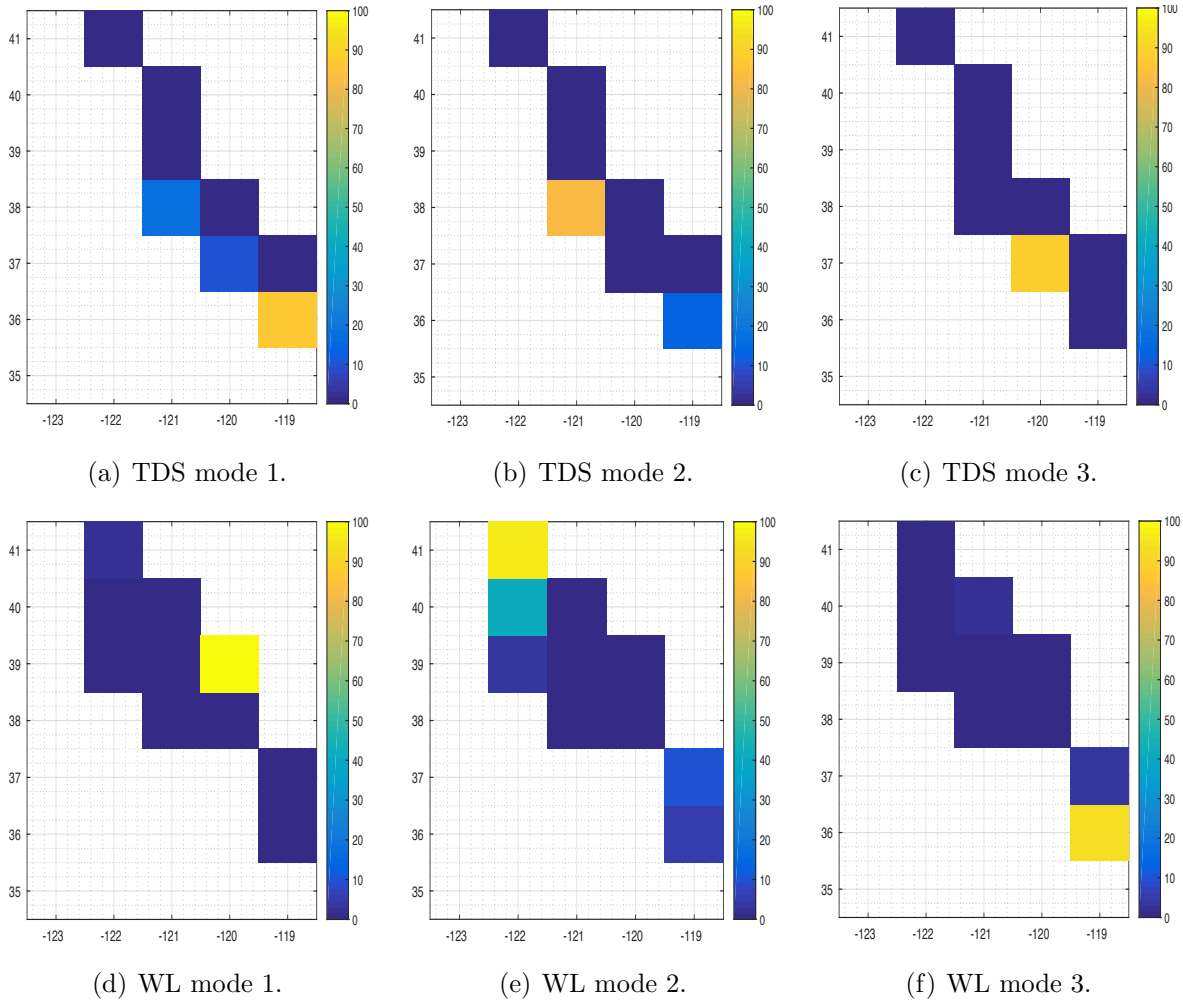
Figure 4.10: Spatial distribution of homogeneous variance for the first 3 modes of TDS and groundwater levels in the Central Valley. Colors represent percent (%) of variance accounted for by the relevant mode.

Over 90% of the total variance of TDS is captured in the first two modes for the GRACE period. The first mode of EOF decomposition explains 51.4%, 92.6% and 96.1% of total variance for *in situ* TDS, *in situ* groundwater levels, and *dSubsurface* respectively. The second mode explains 42.4%, 5%, and 2.6% of total variance respectively. The third mode explains 6.1%, 2.2%, and 1.2% of the total variance respectively.

Homogeneous maps of variance in Figure 4.10 show the spatial distribution of the percent of variance accounted for by the respective mode (Bjornsson and Venegas [1997]). The variance is calculated as the squared correlation of the principal components with each grid cell of

(a) WL mode 1.

(b) WL mode 2.

(c) WL mode 3.

(d) *dSubsurface* mode 1.

(e) *dSubsurface* mode 2.

(f) *dSubsurface* mode 3.
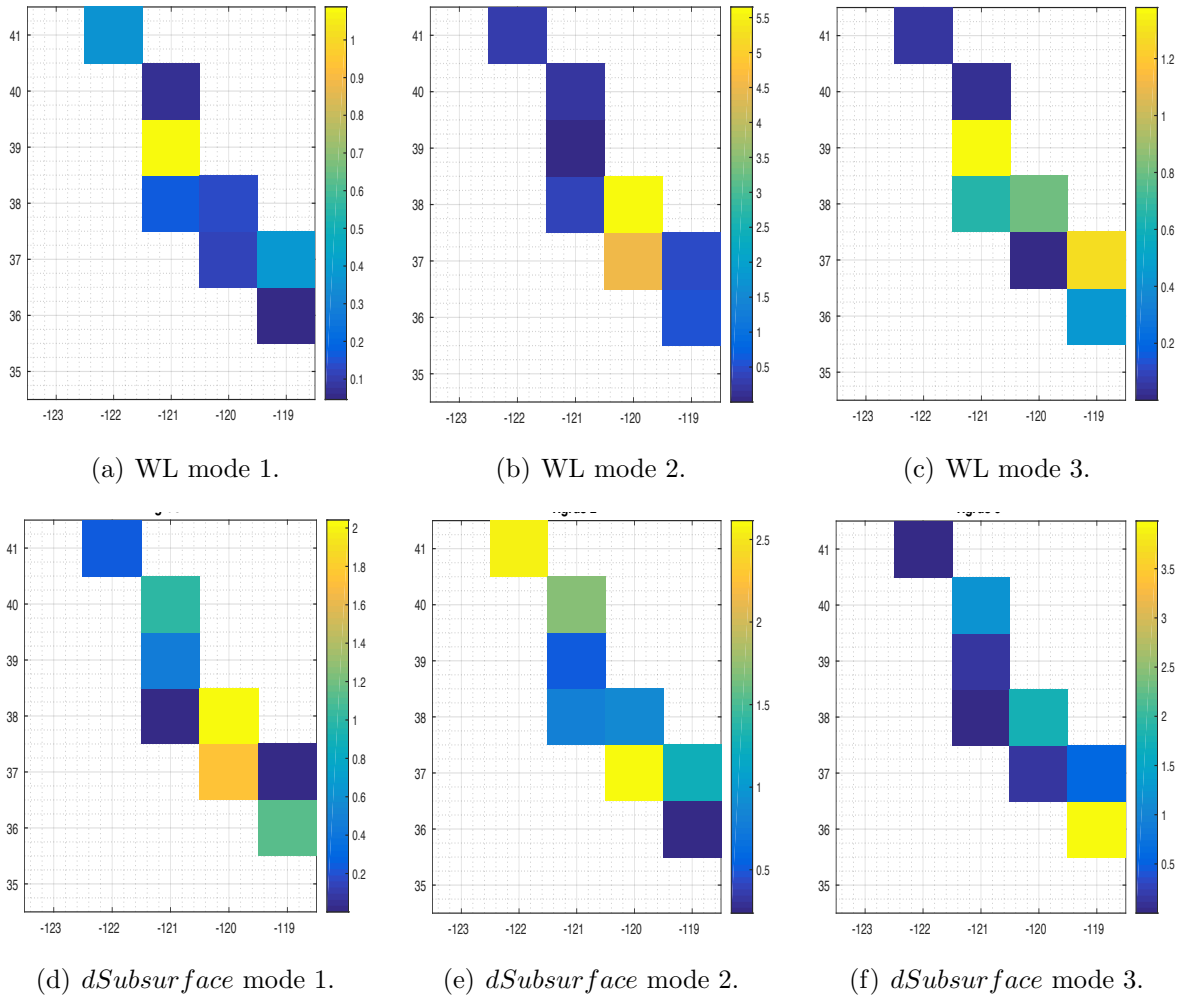
Figure 4.11: Heterogeneous variance of TDS predictable by the first three modes of groundwater levels and *dSubsurface* in the Central Valley. Colors represent percent (%) variance. Note different color scales.

values within the same field, reported as a percent. For example, for the first mode of TDS, the correlation between the first mode of principal components and each TDS grid cell value over time is calculated and squared to produce the percent variance explained by the first mode shown in Figure 4.11(d).

As Figures 4.11(d), 4.11(e), and 4.11(f) show, only three grid cells account for the most variance of TDS in the Central Valley during the GRACE period in the first three modes. The three grid cells accounting for the most variance of TDS are the southwestern-most grid cells of the study area along the San Joaquin Valley. This suggests the majority of

116

TDS variability during the GRACE period is likely linked to the soil salinization in the San Joaquin Valley (Deverel and Gallanthine [1989], Andrews et al. [2002], Letey et al. [2002], Pitman and Läuchli [2002], Schoups et al. [2005]) and selenium derived from the marine sedimentary Coastal Ranges (Presser and Ohlendorf [1987], Deverel and Gallanthine [1989], Presser and Swain [1990], Dubrovsky et al. [1993], Letey et al. [2002]).

As with TDS, very few grid cells account for most of the variance of groundwater levels in the Central Valley during the GRACE period, as shown in Figures 4.11(a), 4.11(b), and 4.11(c). The variance of the first mode, which as previously stated accounts for 92.6% of the total variance, is almost entirely explained by one grid cell on the eastern boundary near the Sierra Nevada ranges, which experience high annual variability of precipitation (Anderson [2016]). Groundwater level variability in this area are likely highly sensitive to reductions in snow melt water. The northern and southern most grid cells account for the majority of variance in the second and third modes, respectively. The northern Central Valley experiences higher average annual precipitation and runoff, likely reflected in the second mode. The southern Central Valley is heavily reliant upon surface water imports which are often reduced in dry years, likely reflected in the third mode.

Heterogeneous variance maps, shown in Figure 4.11, are similar to the homogeneous versions, but calculate variance as a function of the principal components of one field and the grid cell values of a different field to estimate how well the values of the second field can be predicted if the principal components of the first field are known (Bjornsson and Venegas [1997]). The maps in Figure 4.11 show the percent variance (%) of TDS that can be predicted by the first three principal components of groundwater levels and $dSubsurface$. TDS concentrations are unlikely to be accurately predicted by any of the first three modes of either groundwater levels or $dSubsurface$, as indicated by the very small percent values in Figure 4.11, all well below 10%.

4.4 Conclusions

The GRACE footprint areas show potential for characterizing TDS with a GRACE product. The complex and heavily managed hydrologic system in the Central Valley makes it difficult to accurately quantify a relationship between TDS, groundwater levels, and total water storage anomalies. The moderately less complex High Plains subbasins may be more suitable for future work, provided a substantial volume of groundwater quality observations in time and space. The database compiled for this work provide large-scale and long-term context for groundwater levels and TDS concentrations in the Central Valley and High Plains leading up to the GRACE period.

EOF decomposition is a powerful method of identifying and analyzing dominant variability patterns in time and space in complex systems. The analysis on TDS suggests the majority of TDS variability can be explained by the first two modes of decomposition over the GRACE period, localized in the southern Central Valley known to have selenium infiltration and soil salinization. Heterogeneous variance maps show the first three decomposition modes of groundwater levels and *dSubsurface* are not likely to predict TDS in the Central Valley during the GRACE period.

Large-scale TDS characterization is challenging, particularly during the relatively short GRACE period. However, the methods described in this chapter have potential in future work using a comprehensively monitored groundwater quality parameter with the planned GRACE follow on mission (GRACE-FO) to characterize large-scale groundwater quality.

Chapter 5

Conclusions

Groundwater quality has a strong literature background with studies in the US and abroad. Point and non-point sources of contaminants and processes controlling variability of groundwater quality, such as influences from agricultural land use or pumping large volumes of groundwater leading to vertical mixing, are widely studied (e.g. Alley [1993], Charles et al. [1993], Litke [2001], Lashkaripour et al. [2005], McMahon et al. [2007], Lindsey and Rupert [2012], Shamsudduha et al. [2015]). The groundwater quality literature generally focuses on relatively small spatial scales and/or step-wise changes in constituent concentrations between two periods. The work presented here lays the groundwork for characterizing large-scale groundwater quality on a more consistent temporal scale and assessing the role of water storage fluctuations.

The total dissolved solid (TDS) concentration of a groundwater sample of known volume is the weight of the dried, filtered material. The physical parameter is relatively easy to measure so a large number of readily available observations is expected compared to constituent parameters with more intensive measurement procedures such as total nitrogen. TDS has a straightforward mass balance as there are no potential physical, chemical, or biological transformations, so the methods described in this work can be applied to other constituents of interest. TDS concentrations are not region-specific, making the methods explored in this work applicable to other locations. TDS can include regionally-specific constituents,

which can be highlighted using the analysis approaches explored here, such as selenium in the southern Central Valley.

This work explores approaches to characterizing large-scale TDS concentrations in groundwater, with the ultimate goal of predicting fluctuations in groundwater quality using remote sensing. First, a database of *in situ* TDS and groundwater level observations are compiled for the Central Valley aquifer in California and the High Plains aquifer in the central US, and preliminary relationships and scaling approaches are evaluated. Next, models predicting average TDS concentrations over time are constructed as a function of *in situ* groundwater levels and categorical season, assessing whether dominant lithology and land use classification and/or inclusion of GRACE-derived subsurface storage anomalies, $dSubsurface_{S_y}$, as a potential predictor variable make quantifiable improvements to TDS predictions. Finally, two spatial analysis approaches are explored, the first examining TDS characterization on subbasin spatial scales, and the second evaluating regional TDS variability in space and time.

The compiled *in situ* TDS database is as comprehensive as possible using publicly available observation records from various national and state agencies following the same sampling procedures. The database is imperfect as it lacks metadata such as screened well depth and hydrogeologic information that is not readily available from every sampling source, but it does provide context of large-scale TDS in groundwater over an extended period. The long-term *in situ* TDS and groundwater level data are used throughout this work to assess TDS characterization and potential influences of various hydroclimatic conditions.

To characterize TDS on large spatial scales, with a particular interest in using NASA's GRACE product to explore the role of water storage fluctuations, four scaling approaches are assessed to bridge point observations with a $1° \times 1°$ grid. The scaling approaches evaluated include an average of all points within a $1° \times 1°$ grid cell, linear interpolation of points to a $1° \times 1°$ grid, inverse distance weighting of points to a $1° \times 1°$ grid, and scaling *in situ* points and $1° \times 1°$ grid cells to constructed Thiessen polygons. The inverse distance

weighting and Thiessen polygon approaches best represent the variability, spatial extent, and patterns of observed TDS, suggesting the dynamics of TDS are preserved. Redistribution of $1° \times 1°$ gridded GRACE total water storage anomalies and subsurface storage anomalies into Thiessen polygons does not significantly perturb the basin-wide signal (see Figure 2.12).

TDS models with Thiessen polygons categorized by dominant lithology and dominant land use regimes do not improve prediction of TDS through time compared to basin-wide models, both excluding GRACE-derived $dSubsurface_{S_y}$ as a potential predictor. The effects of fewer observations to constrain regime models outweigh the potential improvements of dominant regime classification. Future work may include more comprehensive classifications incorporating secondary or tertiary lithology and/or land use.

Over the GRACE period, the centrally-located alluvium/developed regime model covering the largest fractional area in the Central Valley more accurately predicts TDS than the basin-wide model over the same period, both including $dSubsurface_{S_y}$ as a potential predictor. The regime model improvement over the basin-wide model is likely due to some combination of central location, large area and therefore large number of observations for model constraint, the implied hydraulic connectivity between the surface and subsurface due to the dominance of alluvium cover, and the suggested link between average TDS and anthropogenic runoff. Future work may consider further investigating the potential explanations for improved TDS prediction for the alluvium/developed regime model in the Central Valley compared to the basin-wide model offered here.

TDS models including $dSubsurface_{S_y}$ as a potential predictor generally introduces additional variability to regime model predictions. The additional variability may be due to the relatively small regime areas compared to the area recommended for reduced error of gridded GRACE products. However, the basin-wide model in the Central Valley better predicts TDS when including $dSubsurface_{S_y}$ as a predictor compared to the basin-wide model excluding $dSubsurface_{S_y}$ over the same period with low variability of model fit, suggesting the

additional information improves model constraint on the basin-scale. Future work may incorporate data from the planned GRACE-FO mission to improve robustness of a basin-wide TDS model in the Central Valley.

The first spatial analysis approach are the GRACE footprints. GRACE footprints are defined by the Gaussian filtering radius of pre-processing level 3 gridded data (Landerer and Swenson [2012]). The footprint areas show potential for characterizing TDS on a subbasin scale using a GRACE product. The heavily managed hydrology of the Central Valley makes TDS difficult to characterize as a function of groundwater levels and total water storage anomalies. The hydrologic system of the High Plains are more amenable to TDS characterization by footprint areas, but is currently limited by the availability of *in situ* TDS observations during the GRACE period. Future work may include data from the GRACE-FO mission, provided increased availability of *in situ* TDS observations in time and space.

The second spatial analysis approach is single-field EOF decomposition. EOF analysis on TDS interpolated by inverse distance weighting suggests the majority of variability is explained by the first two modes of decomposition during the GRACE period, localized in the southern Central Valley where selenium infiltration and soil salinization are well-documented. Heterogeneous variance analysis maps show the first three decomposition modes of groundwater levels and GRACE-derived $dSubsurface$ are not likely to predict TDS in the Central Valley during the GRACE period. Future work incorporating GRACE-FO data may confirm that decomposition modes of $dSubsurface$ alone are not likely to accurately predict TDS. Further work may explore heterogeneous variance analysis of a GRACE product with TDS, or other constituent of interest, in another aquifer of interest.

The GRACE period alone is likely too short to accurately draw robust conclusions regarding groundwater fluctuations and potential TDS responses. However, the long-term *in situ* database over the Central Valley and High Plains study areas give context to large-scale TDS concentrations and groundwater levels. Characterization of TDS concentrations in ground-

water is difficult, particularly during the relatively short GRACE period between 2003 and 2014, but the methods described in this work show potential for future investigations of large-scale groundwater constituent concentration characterization with the planned GRACE-FO mission, provided ground-based observations of groundwater quality parameters are readily available.

Freshwater availability and sustainability assessments must consider both quantity and quality of resources, particularly as global groundwater dependency is likely to increase with population growth and freshwater redistribution is driven by climate change. To achieve the goal of predicting variability in large-scale groundwater quality using remote sensing, we must first be able to accurately characterize and quantify constituent fluctuations. Comprehensive *in situ* observations of groundwater constituents are crucial to accurately characterize fluctuations in groundwater quality. Successful quantification of large-scale groundwater potability will improve estimation accuracy of groundwater availability and sustainability.

Bibliography

S. K. Acharyya, S. Lahiri, B. C. Raymahashay, and A. Bhowmik. Arsenic toxicity of groundwater in parts of the Bengal basin in India and Bangladesh: the role of quaternary stratigraphy and holocene sea-level fluctuation. *Environmental Geology*, 39(10):1127–1137, 2000. ISSN 1432-0495. doi: 10.1007/s002540000107. URL http://dx.doi.org/10.1007/s002540000107.

W. M. Alley. *Regional ground-water quality*. John Wiley & Sons, 1993.

W. M. Alley, R. W. Healy, J. W. LaBaugh, and T. E. Reilly. Flow and storage in groundwater systems. *science*, 296(5575):1985–1990, 2002.

M. Anderson. California's most significant droughts: Comparing historical and recent conditions, 2016.

S. S. Andrews, J. P. Mitchell, R. Mancinelli, D. L. Karlen, T. K. Hartz, W. R. Horwath, G. S. Pettygrove, K. M. Scow, and D. S. Munk. On-farm assessment of soil quality in California's Central Valley. *Agronomy Journal*, 94(1):12–23, 2002.

D. W. Anning, N. J. Bauch, S. J. Gerner, M. E. Flynn, S. N. Hamlin, S. J. Moore, D. H. Schaefer, S. K. Anderholm, and L. E. Spangler. Dissolved solids in basin-fill aquifers and streams in the southwestern united states. Technical report, Geological Survey (US), 2007.

J. D. Ayotte, J. M. Gronberg, and L. E. Apodaca. *Trace elements and radon in groundwater across the United States, 1992-2003*. US Department of the Interior, US Geological Survey, 2011.

M. Becker, W. LLovel, A. Cazenave, A. Güntner, and J.-F. Crétaux. Recent hydrological behavior of the East African great lakes region inferred from GRACE, satellite altimetry and rainfall observations. *Comptes Rendus Geoscience*, 342(3):223–233, 2010.

L. M. Bexfield and B. C. Jurgens. Effects of seasonal operation on the quality of water produced by public-supply wells. *Groundwater*, 52(S1):10–24, 2014. ISSN 1745-6584. doi: 10.1111/gwat.12174. URL http://dx.doi.org/10.1111/gwat.12174.

P. Bhattacharya, D. Chatterjee, and G. Jacks. Occurrence of arsenic-contaminated groundwater in alluvial aquifers from delta plains, Eastern India: Options for safe drinking water supply. *International Journal of Water Resources Development*, 13(1):79–92, 1997. doi: 10.1080/07900629749944. URL http://dx.doi.org/10.1080/07900629749944.

H. Bjornsson and S. Venegas. A manual for EOF and SVD analyses of climatic data. *CCGCR Report*, 97(1):112–134, 1997.

J.-K. Böhlke. Groundwater recharge and agricultural contamination. *Hydrogeology Journal*, 10(1):153–179, 2002.

H. P. Broers and B. van der Grift. Regional monitoring of temporal changes in groundwater quality. *Journal of Hydrology*, 296(1–4):192 – 220, 2004. ISSN 0022-1694. doi: http://dx. doi.org/10.1016/j.jhydrol.2004.03.022. URL `http://www.sciencedirect.com/science/article/pii/S0022169404001891`.

California Department of Water Resources. California Statewide Groundwater Elevation Monitoring Program, a. URL `http://www.water.ca.gov/groundwater/casgem/`.

California Department of Water Resources. Bulletin 132 Management of the California State Water Project, b. URL `http://www.water.ca.gov/swpao/bulletin_home.cfm`.

California State Water Resources Control Board/Division of Water Quality. Groundwater Ambient Monitoring and Assessment Program. URL `http://www.swrcb.ca.gov/gama/`.

C.-L. Chang, S.-L. Lo, and S.-L. Yu. The parameter optimization in the inverse distance method by geneticalgorithm for estimating precipitation. *Environmental Monitoring and Assessment*, 117(1):145–155, 2006. ISSN 1573-2959. doi: 10.1007/s10661-006-8498-0. URL `http://dx.doi.org/10.1007/s10661-006-8498-0`.

E. G. Charles, C. Behroozi, J. Schooley, and J. L. Hoffman. A method for evaluating groundwater-recharge areas in New Jersey. *New Jersey Geological Survey Geological Survey Report GSR-32*, 1993.

A. Chatterjee, D. Das, B. K. Mandal, T. R. Chowdhury, G. Samanta, and D. Chakraborti. Arsenic in ground water in six districts of West Bengal, India: The biggest arsenic calamity in the world. i: Arsenic species in drinking water and urine of the affected people. *Analyst*, 120(3):643–650, 1995.

S. Chaudhuri and S. Ale. Long term (1960–2010) trends in groundwater contamination and salinization in the Ogallala aquifer in Texas. *Journal of Hydrology*, 513:376 – 390, 2014a. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/j.jhydrol.2014.03.033. URL `http://www.sciencedirect.com/science/article/pii/S0022169414002182`.

S. Chaudhuri and S. Ale. Temporal evolution of depth-stratified groundwater salinity in municipal wells in the major aquifers in Texas, {USA}. *Science of The Total Environment*, 472:370 – 380, 2014b. ISSN 0048-9697. doi: http://dx.doi.org/10.1016/j.scitotenv.2013.10.120. URL `http://www.sciencedirect.com/science/article/pii/S0048969713012643`.

U. K. Chowdhury, B. K. Biswas, T. R. Chowdhury, G. Samanta, B. K. Mandal, G. C. Basu, C. R. Chanda, D. Lodh, K. C. Saha, S. K. Mukherjee, et al. Groundwater arsenic contamination in Bangladesh and West Bengal, India. *Environmental health perspectives*, 108(5):393, 2000.

N. Cressie. Spatial prediction and ordinary kriging. *Mathematical Geology*, 20(4):405–421, 1988. ISSN 1573-8868. doi: 10.1007/BF00892986. URL `http://dx.doi.org/10.1007/BF00892986`.

N. Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990. ISSN 1573-8868. doi: 10.1007/BF00889887. URL `http://dx.doi.org/10.1007/BF00889887`.

D. Crossley, C. De Linage, J. Hinderer, J.-P. Boy, and J. Famiglietti. A comparison of the gravity field over Central Europe from superconducting gravimeters, GRACE and global hydrological models, using EOF analysis. *Geophysical Journal International*, 189 (2):877–897, 2012.

J. J. de Vries and I. Simmers. Groundwater recharge: an overview of processes and challenges. *Hydrogeology Journal*, 10(1):5–17, 2002. ISSN 1435-0157. doi: 10.1007/s10040-001-0171-7. URL `http://dx.doi.org/10.1007/s10040-001-0171-7`.

K. F. Dennehy. High Plains regional ground-water study. Technical report, Geological Survey (US), 2000.

K. F. Dennehy. Ground-water resources program. Technical report, 2005.

L. A. DeSimone and P. A. Hamilton. *Quality of water from domestic wells in principal aquifers of the United States, 1991-2004*. US Department of the Interior, US Geological Survey, 2009.

S. Deverel and S. Gallanthine. Relation of salinity and selenium in shallow groundwater to hydrologic and geochemical processes, Western San Joaquin Valley, California. *Journal of Hydrology*, 109(1):125 – 149, 1989. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/0022-1694(89)90011-5. URL `http://www.sciencedirect.com/science/article/pii/0022169489900115`.

B. Dewitte, J. Vazquez-Cuervo, K. Goubanova, S. Illig, K. Takahashi, G. Cambon, S. Purca, D. Correa, D. Gutiérrez, A. Sifeddine, et al. Change in El Nino flavours over 1958–2008: Implications for the long-term trend of the upwelling off Peru. *Deep Sea Research Part II: Topical Studies in Oceanography*, 77:143–156, 2012.

P. Döll. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. *Environmental Research Letters*, 4(3):035006, 2009. URL `http://stacks.iop.org/1748-9326/4/i=3/a=035006`.

P. Döll, H. Müller Schmied, C. Schuh, F. T. Portmann, and A. Eicker. Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, 50(7):5698–5720, 2014. ISSN 1944-7973. doi: 10.1002/2014WR015595. URL `http://dx.doi.org/10.1002/2014WR015595`.

N. Dubrovsky, S. Deverel, and R. Gilliom. Multiscale approach to regional ground-water quality assessment: Selenium in the San Joaquin Valley, California. *Regional Ground-Water Quality, Van Nostrand Reinhold*, pages 537–562, 1993.

D. A. V. Eckhardt and P. E. Stackelberg. Relation of ground-water quality to land use on long island, new york. *Ground Water*, 33(6):1019–1033, 1995. ISSN 1745-6584. doi: 10.1111/j.1745-6584.1995.tb00047.x. URL `http://dx.doi.org/10.1111/j.1745-6584.1995.tb00047.x`.

S. S. Embrey and D. L. Runkle. *Microbial quality of the Nation's ground-water resources, 1993-2004*. US Department of the Interior, US Geological Survey, 2006.

L. E. Erban, S. M. Gorelick, H. A. Zebker, and S. Fendorf. Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping-induced land subsidence. *Proceedings of the National Academy of Sciences*, 110(34):13751–13756, 2013. doi: 10.1073/pnas.1300503110. URL `http://www.pnas.org/content/110/34/13751.abstract`.

M. E. Exner, A. J. Hirsh, and R. F. Spalding. Nebraska's groundwater legacy: Nitrate contamination beneath irrigated cropland. *Water Resources Research*, 50(5):4474–4489, 2014. ISSN 1944-7973. doi: 10.1002/2013WR015073. URL `http://dx.doi.org/10.1002/2013WR015073`.

J. S. Famiglietti, M. Lo, S. L. Ho, J. Bethune, K. J. Anderson, T. H. Syed, S. C. Swenson, C. R. de Linage, and M. Rodell. Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, 38(3):n/a–n/a, 2011. ISSN 1944-8007. doi: 10.1029/2010GL046442. URL `http://dx.doi.org/10.1029/2010GL046442`. L03403.

C. Farrar and G. Bertoldi. Region 4, central valley and pacific coast ranges. *Hydrogeology. The Geological Society of North America, Boulder Colorado. 1988. p 59-67. 4 fig, 28 ref.*, 1988.

C. C. Faunt. Groundwater availability of the Central Valley Aquifer, California. *U.S. Geol. Surv. Prof. Pap.1766*, page 173, 2009.

M. A. Fazal, T. Kawachi, and E. Ichion. Extent and severity of groundwater arsenic contamination in bangladesh. *Water International*, 26(3):370–379, 2001. doi: 10.1080/02508060108686929. URL `http://dx.doi.org/10.1080/02508060108686929`.

C. W. Fetter and C. Fetter. *Contaminant hydrogeology*, volume 500. Prentice hall New Jersey, 1999.

R. M. Garrels and F. T. Mackenzie. Origin of the chemical compositions of some springs and lakes. *Equilibrium concepts in natural water systems*, 67:222–242, 1967.

T. Gleeson, L. Marklund, L. Smith, and A. H. Manning. Classifying the water table at regional to continental scales. *Geophysical Research Letters*, 38(5):n/a–n/a, 2011. ISSN 1944-8007. doi: 10.1029/2010GL046427. URL `http://dx.doi.org/10.1029/2010GL046427`. L05401.

T. Gleeson, Y. Wada, M. F. Bierkens, and L. P. van Beek. Water balance of global aquifers revealed by groundwater footprint. *Nature*, 488(7410):197–200, 2012.

J. J. Gurdak and S. L. Qi. Vulnerability of recently recharged ground water in the High Plains aquifer to nitrate contamination. Technical report, U. S. Geological Survey, 2006.

J. J. Gurdak, P. B. McMahon, K. Dennehy, and S. L. Qi. Water quality in the High Plains aquifer, Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming, 1999–2004. *US Geological Survey Circular*, 1337:63, 2009.

J. M. Harbor. A practical method for estimating the impact of land-use change on surface runoff, groundwater recharge and wetland hydrology. *Journal of the American Planning Association*, 60(1):95–108, 1994. doi: 10.1080/01944369408975555. URL `http://dx.doi.org/10.1080/01944369408975555`.

I. Harris, P. Jones, T. Osborn, and D. Lister. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, 34(3):623–642, 2014. ISSN 1097-0088. doi: 10.1002/joc.3711. URL `http://dx.doi.org/10.1002/joc.3711`.

D. R. Helsel and R. M. Hirsch. Statistical methods in water resources. In *Techniques of Water Resources Investigations Book 4*, chapter A3, page 510. U. S. Geological Survey, 2002.

C. M. Henry, D. M. Allen, and J. Huang. Groundwater storage variability and annual recharge using well-hydrograph and GRACE satellite data. *Hydrogeology Journal*, 19(4): 741–755, 2011. ISSN 1435-0157. doi: 10.1007/s10040-011-0724-3. URL `http://dx.doi.org/10.1007/s10040-011-0724-3`.

R. M. Hirsch, J. R. Slack, and R. A. Smith. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18(1):107–121, 1982. ISSN 1944-7973. doi: 10.1029/WR018i001p00107. URL `http://dx.doi.org/10.1029/WR018i001p00107`.

R. M. Hirsch, R. B. Alexander, and R. A. Smith. Selection of methods for the detection and estimation of trends in water quality. *Water resources research*, 27(5):803–813, 1991.

C. G. Homer, J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5):345–354, 2015.

M. Hossain. Arsenic contamination in Bangladesh—an overview. *Agriculture, Ecosystems and Environment*, 113(1–4):1 – 16, 2006. ISSN 0167-8809. doi: http://dx.doi.org/10.1016/j.agee.2005.08.034. URL `http://www.sciencedirect.com/science/article/pii/S0167880905004329`.

R. Houborg, M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, 48(7):n/a–n/a, 2012. ISSN 1944-7973. doi: 10.1029/2011WR011291. URL `http://dx.doi.org/10.1029/2011WR011291`. W07525.

B. C. Jurgens, M. S. Fram, K. Belitz, K. R. Burow, and M. K. Landon. Effects of groundwater development on uranium: Central valley, california, usa. *Ground Water*, 48(6):913–928, 2010. ISSN 1745-6584. doi: 10.1111/j.1745-6584.2009.00635.x. URL `http://dx.doi.org/10.1111/j.1745-6584.2009.00635.x`.

R. Kent and M. K. Landon. Trends in concentrations of nitrate and total dissolved solids in public supply wells of the Bunker Hill, Lytle, Rialto, and Colton groundwater subbasins, San Bernardino County, California: Influence of legacy land use. *Science of The Total Environment*, 452–453:125 – 136, 2013. ISSN 0048-9697. doi: http://dx.doi.org/10.1016/j.scitotenv.2013.02.042. URL `http://www.sciencedirect.com/science/article/pii/S0048969713002246`.

F. W. Landerer and S. C. Swenson. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48(4):n/a–n/a, 2012. ISSN 1944-7973. doi: 10.1029/2011WR011453. URL `http://dx.doi.org/10.1029/2011WR011453`.

M. K. Landon, C. T. Green, K. Belitz, M. J. Singleton, and B. K. Esser. Relations of hydrogeologic factors, groundwater reduction-oxidation conditions, and temporal and spatial distributions of nitrate, central-eastside san joaquin valley, california, usa. *Hydrogeology Journal*, 19(6):1203–1224, 2011. ISSN 1435-0157. doi: 10.1007/s10040-011-0750-1. URL `http://dx.doi.org/10.1007/s10040-011-0750-1`.

W. W. Lapham, F. D. Wilde, and M. T. Koterba. Ground-water data-collection protocols and procedures for the national water-quality assessment program: Selection, installation, and documentation of wells, and collection of related data. Technical report, Geological Survey (US), 1995.

G. Lashkaripour and M. Ghafoori. The effects of water table decline on the groundwater quality in aquifer of Torbat Jam Plain, Northeast Iran. *International Journal of Emerging Sciences*, 1(2):153, 2011.

G. R. Lashkaripour, A. Asghari-Moghaddam, and M. Allaf-Najib. The effects of water table decline on the groundwater quality in Marand Plain, Northwest Iran. *Iranian Int. J. Sci*, 6(1):47–60, 2005.

D. N. Lerner, A. S. Issar, and I. Simmers. *Groundwater recharge: a guide to understanding and estimating natural recharge*, volume 8. Heise Hannover, 1990.

J. Letey, C. Williams, and M. Alemi. Salinity, drainage and selenium problems in the western san joaquin valley of california. *Irrigation and Drainage Systems*, 16(4):253–259, 2002. ISSN 1573-0654. doi: 10.1023/A:1024812826664. URL `http://dx.doi.org/10.1023/A:1024812826664`.

D. P. Lettenmaier. Detection of trends in water quality data from records with dependent observations. *Water Resources Research*, 12(5):1037–1046, 1976. ISSN 1944-7973. doi: 10.1029/WR012i005p01037. URL `http://dx.doi.org/10.1029/WR012i005p01037`.

B. D. Lindsey and M. G. Rupert. Methods for evaluating temporal groundwater quality data and results of decadal-scale changes in chloride, dissolved solids, and nitrate concentrations in groundwater in the United States, 1988–2010. *US Geological Survey Scientific Investigations Report*, 5049(2012):46, 2012.

B. D. Lindsey, M. P. Berndt, B. G. Katz, A. F. Ardis, and K. A. Skach. Factors affecting water quality in selected carbonate aquifers in the united states, 1993-2005. Technical report, 2009.

D. J. Lisk. Environmental effects of landfills. *Science of The Total Environment*, 100:415 – 468, 1991. ISSN 0048-9697. doi: http://dx.doi.org/10.1016/0048-9697(91)90387-T. URL `http://www.sciencedirect.com/science/article/pii/004896979190387T`.

D. W. Litke. Historical water-quality data for the High Plains Regional Groundwater study area in Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming, 1930-98. *U.S. Geological Survey Water-Resources Investigations Report, 00-4254*, page 65, 2001.

M.-H. Lo and J. S. Famiglietti. Irrigation in California's Central Valley strengthens the southwestern US water cycle. *Geophysical Research Letters*, 40(2):301–306, 2013.

J. C. Loftis. Trends in groundwater quality. *Hydrological processes*, 10(2):335–355, 1996.

L. Longuevergne, B. R. Scanlon, and C. R. Wilson. GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research*, 46(11):n/a–n/a, 2010. ISSN 1944-7973. doi: 10.1029/2009WR008564. URL `http://dx.doi.org/10.1029/2009WR008564`. W11517.

S. Ludington, B. C. Moring, R. J. Miller, P. A. Stone, A. A. Bookstrom, D. R. Bedford, J. G. Evans, G. A. Haxel, C. J. Nutt, K. S. Flyn, et al. Preliminary integrated geologic map databases for the United States. *Western States: California, Nevada, Arizona, Washington, Oregon, Idaho, and Utah. Version*, 1, 2007.

R. M. Maxwell, L. E. Condon, S. J. Kollet, K. Maher, R. Haggerty, and M. M. Forrester. The imprint of climate and geology on the residence times of groundwater. *Geophysical Research Letters*, 43(2):701–708, 2016. ISSN 1944-8007. doi: 10.1002/2015GL066916. URL `http://dx.doi.org/10.1002/2015GL066916`. 2015GL066916.

V. McGuire. Water-level changes in the high plains aquifer, predevelopment to 2007, 2005-06, and 2006-07. *Publications of the US Geological Survey*, page 17, 2009.

V. McGuire. Water-level changes and change in water in storage in the High Plains Aquifer, predevelopment to 2013 and 2011–2013, Scientific Investigations Report 2014–5218. *US Geological Survey*, 2014.

V. L. McGuire, K. D. Lund, and B. K. Densmore. Saturated Thickness and Water in Storage in the High Plains Aquifer, 2009, and Water-Level Changes and Changes in Water in Storage in the High Plains Aquifer, 1980 to 1995, 1995 to 2000, 2000 to 2005, and 2005 to 2009. *US Geological Survey Scientific Investigations Report*, 5177:28, 2012.

P. McMahon and F. Chapelle. Redox processes and water quality of selected principal aquifer systems. *Ground Water*, 46(2):259–271, 2008. ISSN 1745-6584. doi: 10.1111/j.1745-6584. 2007.00385.x. URL `http://dx.doi.org/10.1111/j.1745-6584.2007.00385.x`.

P. B. McMahon, K. F. Dennehy, B. W. Bruce, J. J. Gurdak, and S. L. Qi. Water-quality assessment of the High Plains Aquifer, 1999–2004. *U.S. Geological Survey Professional Paper 1749*, page 136, 2007.

S. Mehta, A. E. Fryar, and J. L. Banner. Controls on the regional-scale salinization of the Ogallala aquifer, Southern High Plains, Texas, {USA}. *Applied Geochemistry*, 15(6):849 – 864, 2000. ISSN 0883-2927. doi: http://dx.doi.org/10.1016/S0883-2927(99)00098-0. URL `http://www.sciencedirect.com/science/article/pii/S0883292799000980`.

J. Meisinger, W. Hargrove, R. Mikkelsen, J. Williams, and V. Benson. Effects of cover crops on groundwater quality. *Cover Crops for Clean Water. Soil and Water Conservation Society. Ankeny, Iowa*, 266:793–799, 1991.

L. A. Mertes, M. O. Smith, and J. B. Adams. Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images. *Remote Sensing of Environment*, 43(3):281–301, 1993.

J. A. Miller and C. L. Appel. *Ground water atlas of the United States: Segment 3, Kansas, Missouri, Nebraska*. Number 730-D. U. S. Geological Survey, 1997.

N. L. Miller, K. E. Bashford, and E. Strem. Potential impacts of climate change on california hydrology1. *JAWRA Journal of the American Water Resources Association*, 39(4):771– 784, 2003. ISSN 1752-1688. doi: 10.1111/j.1752-1688.2003.tb04404.x. URL `http://dx.doi.org/10.1111/j.1752-1688.2003.tb04404.x`.

F. Mosteller and J. W. Tukey. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.

J. Nanteza, C. R. de Linage, B. Thomas, and J. Famiglietti. Monitoring groundwater storage changes in complex basement aquifers: An evaluation of the GRACE satellites over East Africa. *Water Resources Research*, pages n/a–n/a, 2016. ISSN 1944-7973. doi: 10.1002/ 2016WR018846. URL `http://dx.doi.org/10.1002/2016WR018846`.

H.-S. Park and B. Sohn. Recent trends in changes of vegetation over East Asia coupled with temperature and rainfall variations. *Journal of Geophysical Research: Atmospheres*, 115 (D14), 2010.

E. J. Pebesma and J. De Kwaadsteniet. Mapping groundwater quality in the Netherlands. *Journal of Hydrology*, 200(1):364–386, 1997.

M. A. Perry and J. D. Niemann. Analysis and estimation of soil moisture at the catchment scale using EOFs. *Journal of Hydrology*, 334(3–4):388 – 404, 2007. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/j.jhydrol.2006.10.014. URL `http://www.sciencedirect.com/science/article/pii/S0022169406005518`.

M. G. Pitman and A. Läuchli. *Global Impact of Salinity and Agricultural Ecosystems*, pages 3–20. Springer Netherlands, Dordrecht, 2002. ISBN 978-0-306-48155-0. doi: 10.1007/0-306-48155-3_1. URL `http://dx.doi.org/10.1007/0-306-48155-3_1`.

M. Planert and J. S. Williams. *Ground Water Atlas of the United States: Segment 1, California, Nevada*. Number 730-B. U. S. Geological Survey, 1995.

T. S. Presser and H. M. Ohlendorf. Biogeochemical cycling of selenium in the San Joaquin Valley, California, USA. *Environmental Management*, 11(6):805–821, 1987. ISSN 1432-1009. doi: 10.1007/BF01867247. URL `http://dx.doi.org/10.1007/BF01867247`.

T. S. Presser and W. C. Swain. Geochemical evidence for se mobilization by the weathering of pyritic shale, san joaquin valley, california, u.s.a. *Applied Geochemistry*, 5(5):703 – 717, 1990. ISSN 0883-2927. doi: http://dx.doi.org/10.1016/0883-2927(90)90066-E. URL `http://www.sciencedirect.com/science/article/pii/088329279090066E`.

L. J. Puckett, C. Zamora, H. Essaid, J. T. Wilson, H. M. Johnson, M. J. Brayton, and J. R. Vogel. Transport and fate of nitrate at the ground-water/surface-water interface. *Journal of Environmental Quality*, 37:1034–1050, 2008. doi: 10.2134/jeq2006.0550.

S. Qi and S. Christenson. Assessing groundwater availability in the High Plains aquifer in parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming. *US Geological Survey Fact Sheet*, 3008(4), 2010.

M. H. Rezaie-Boroon and J. B. Fisher. Linking groundwater quality and quantity: An assessment of satellite-based groundwater storage anomalies from GRACE against ground measurements of contaminants in California. *Journal of Environmental Science and Engineering. B*, 1(11B):1271, 2012.

A. S. Richey, B. F. Thomas, M. Lo, J. S. Famiglietti, S. Swenson, and M. Rodell. Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework. *Water Resources Research*, 51(7):5198–5216, 7 2015a. ISSN 1944-7973. doi: 10.1002/2015WR017351. URL `http:https://dx.doi.org/10.1002/2015WR017351`.

A. S. Richey, B. F. Thomas, M.-H. Lo, J. T. Reager, J. S. Famiglietti, K. Voss, S. Swenson, and M. Rodell. Quantifying renewable groundwater stress with GRACE. *Water Resources Research*, 51(7):5217–5238, 2015b. ISSN 1944-7973. doi: 10.1002/2015WR017349. URL `http://dx.doi.org/10.1002/2015WR017349`.

A. S. Richey, B. F. Thomas, M.-H. Lo, J. T. Reager, J. S. Famiglietti, K. Voss, S. Swenson, and M. Rodell. Reply to comment by sahoo et al. on "Quantifying renewable groundwater stress with GRACE". *Water Resources Research*, 52(5):4188–4192, 2016. ISSN 1944-7973. doi: 10.1002/2015WR018329. URL `http://dx.doi.org/10.1002/2015WR018329`.

J. C. Ritchie, P. V. Zimba, and J. H. Everitt. Remote sensing techniques to assess water quality. *Photogrammetric Engineering & Remote Sensing*, 69(6):695–704, 2003.

M. Rodell and J. Famiglietti. The potential for satellite-based monitoring of groundwater storage changes using GRACE: the High Plains aquifer, central US. *Journal of Hydrology*, 263(1–4):245 – 256, 2002. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/S0022-1694(02)00060-4. URL `http://www.sciencedirect.com/science/article/pii/S0022169402000604`.

M. Rodell and J. S. Famiglietti. Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resources Research*, 35 (9):2705–2723, 1999. ISSN 1944-7973. doi: 10.1029/1999WR900141. URL `http://dx.doi.org/10.1029/1999WR900141`.

M. Rodell and J. S. Famiglietti. An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, 37(5):1327–1339, 2001. ISSN 1944-7973. doi: 10.1029/2000WR900306. URL `http://dx.doi.org/10.1029/2000WR900306`.

M. Rodell, J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *HYDROGEOLOGY JOURNAL*, 15(1):159–166, FEB 2007. ISSN 1431-2174. doi: {10.1007/s10040-006-0103-7}.

M. Rodell, I. Velicogna, and J. S. Famiglietti. Satellite-based estimates of groundwater depletion in India. *NATURE*, 460(7258):999–U80, AUG 20 2009. ISSN 0028-0836. doi: {10.1038/nature08238}.

S. Sahoo, T. Russo, and U. Lall. Comment on "Quantifying renewable groundwater stress with GRACE" by Alexandra S. Richey et al. *Water Resources Research*, 52(5):4184–4187, 2016. ISSN 1944-7973. doi: 10.1002/2015WR018085. URL `http://dx.doi.org/10.1002/2015WR018085`.

B. R. Scanlon, R. C. Reedy, D. A. Stonestrom, D. E. Prudic, and K. F. Dennehy. Impact of land use and land cover change on groundwater recharge and quality in the southwestern US. *Global Change Biology*, 11(10):1577–1593, 2005. ISSN 1365-2486. doi: 10.1111/j.1365-2486.2005.01026.x. URL `http://dx.doi.org/10.1111/j.1365-2486.2005.01026.x`.

B. R. Scanlon, C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the national academy of sciences*, 109(24): 9320–9325, 2012a.

B. R. Scanlon, L. Longuevergne, and D. Long. Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research*, 48(4):n/a–n/a, 2012b. ISSN 1944-7973. doi: 10.1029/2011WR011312. URL `http://dx.doi.org/10.1029/2011WR011312`. W04520.

R. Schmidt, S. Petrovic, A. Güntner, F. Barthelmes, J. Wünsch, and J. Kusche. Periodic components of water storage changes from GRACE and global hydrology models. *Journal of Geophysical Research: Solid Earth*, 113(B8), 2008.

G. Schoups, J. W. Hopmans, C. A. Young, J. A. Vrugt, W. W. Wallender, K. K. Tanji, and S. Panday. Sustainability of irrigated agriculture in the san joaquin valley, california. *Proceedings of the National Academy of Sciences*, 102(43):15352–15356, 2005. doi: 10. 1073/pnas.0507723102. URL `http://www.pnas.org/content/102/43/15352.abstract`.

M. Shamsudduha, R. G. Taylor, and R. E. Chandler. A generalized regression model of arsenic variations in the shallow groundwater of Bangladesh. *Water Resources Research*, 51(1):685–703, 2015. ISSN 1944-7973. doi: 10.1002/2013WR014572. URL `http://dx. doi.org/10.1002/2013WR014572`.

I. A. Shiklomanov. Appraisal and assessment of world water resources. *Water International*, 25(1):11–32, 2000. doi: 10.1080/02508060008686794. URL `http://dx.doi.org/10.1080/ 02508060008686794`.

S. J. Shupe, G. D. Weatherford, and E. Checchio. Western water rights: the era of reallocation. *Nat. Resources J.*, 29:413, 1989.

S. Solomon. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press, 2007.

T. Stocker, D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, B. Bex, and B. Midgley. IPCC, 2013: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. 2013.

D. B. Stoeser, G. N. Green, L. C. Morath, W. D. Heran, A. B. Wilson, D. W. Moore, and B. Gosen. Preliminary integrated geologic map databases for the United States. *US Geological Survey, Open-File Report*, 1351, 2005.

G. Strassberg, B. R. Scanlon, and M. Rodell. Comparison of seasonal terrestrial water storage variations from GRACE with groundwater-level measurements from the High Plains Aquifer (USA). *Geophysical Research Letters*, 34(14):n/a–n/a, 2007. ISSN 1944-8007. doi: 10.1029/2007GL030139. URL `http://dx.doi.org/10.1029/2007GL030139`. L14402.

G. Strassberg, B. R. Scanlon, and D. Chambers. Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States. *Water Resour. Res.*, 45(5), 2009.

T. Subramani, L. Elango, and S. R. Damodarasamy. Groundwater quality and its suitability for drinking and agricultural use in chithar river basin, tamil nadu, india. *Environmental Geology*, 47(8):1099–1110, 2005. ISSN 1432-0495. doi: 10.1007/s00254-005-1243-0. URL `http://dx.doi.org/10.1007/s00254-005-1243-0`.

F. A. Swartjes. Risk-based assessment of soil and groundwater quality in the Netherlands: Standards and remediation urgency. *Risk Analysis*, 19(6):1235–1249, 1999. ISSN 1573-9147. doi: 10.1023/A:1007003332488. URL `http://dx.doi.org/10.1023/A:1007003332488`.

S. Swenson and J. Wahr. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research: Solid Earth*, 107(B9):ETG 3–1–ETG 3–13, 2002. ISSN 2156-2202. doi: 10.1029/2001JB000576. URL `http://dx.doi.org/10.1029/2001JB000576`. 2193.

S. Swenson, J. Wahr, and P. C. D. Milly. Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, 39(8):n/a–n/a, 2003. ISSN 1944-7973. doi: 10.1029/2002WR001808. URL `http://dx.doi.org/10.1029/2002WR001808`. 1223.

S. C. Swenson. GRACE monthly land water mass grids NETCDF release 5.0. Ver. 5.0. PO.DAAC,CA,USA., 2012. URL `http://dx.doi.org/10.5067/TELND-NC005`.

S. C. Swenson and J. Wahr. Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters*, 33(8):n/a–n/a, 2006. ISSN 1944-8007. doi: 10.1029/2005GL025285. URL `http://dx.doi.org/10.1029/2005GL025285`. L08402.

G. Q. Tabios and J. D. Salas. A comparative analysis of techniques for spatial interpolation of precipitation1. *JAWRA Journal of the American Water Resources Association*, 21(3): 365–380, 1985. ISSN 1752-1688. doi: 10.1111/j.1752-1688.1985.tb00147.x. URL `http://dx.doi.org/10.1111/j.1752-1688.1985.tb00147.x`.

B. D. Tapley, S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins. GRACE Measurements of Mass Variability in the Earth System. *Science*, 305(5683):503–505, 2004. ISSN 0036-8075. doi: 10.1126/science.1099192. URL `http://science.sciencemag.org/content/305/5683/503`.

A. S. Taschetto and M. H. England. El Niño Modoki impacts on Australian rainfall. *Journal of Climate*, 22(11):3167–3174, 2009.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485, 2012.

R. S. Teegavarapu and V. Chandramouli. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312(1–4):191 – 206, 2005. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/j.jhydrol.2005.02.015. URL `http://www.sciencedirect.com/science/article/pii/S0022169405000880`.

A. H. Thiessen. Precipitation averages for large areas. *Monthly weather review*, 39(7):1082–1089, 1911.

B. F. Thomas, F. W. Landerer, D. N. Wiese, and J. S. Famiglietti. A comparison of watershed storage trends over the eastern and upper Midwestern regions of the United States, 2003–2015. *Water Resources Research*, 52(8):6335–6347, 2016. ISSN 1944-7973. doi: 10.1002/2016WR018617. URL `http://dx.doi.org/10.1002/2016WR018617`.

P. E. Thornton, S. W. Running, and M. A. White. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, 190(3):214 – 251, 1997. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/S0022-1694(96)03128-9. URL `http://www.sciencedirect.com/science/article/pii/S0022169496031289`.

V. M. Tiwari, J. Wahr, and S. C. Swenson. Dwindling groundwater resources in northern India, from satellite gravity observations. *Geophysical Research Letters*, 36(18), 2009. doi: 10.1029/2009GL039401. URL `http://dx.doi.org/10.1029/2009GL039401`.

D. K. Todd, L. W. Mays, et al. Groundwater hydrology. 1980.

J. W. Tukey. Exploratory data analysis. page 506, 1977.

USGS. USGS Water Data for USA: National Water Information System. URL `http://waterdata.usgs.gov/nwis`.

USGS NAWQA. National Water Quality Assessment Program. URL `http://water.usgs.gov/nawqa/`.

I. Velicogna, J. Wahr, and H. Van den Dool. Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals? *Journal of Geophysical Research: Solid Earth*, 106(B8):16415–16434, 2001. ISSN 2156-2202. doi: 10.1029/2001JB000228. URL `http://dx.doi.org/10.1029/2001JB000228`.

C. J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, et al. Global threats to human water security and river biodiversity. *Nature*, 467(7315):555–561, 2010.

K. A. Voss, J. S. Famiglietti, M. Lo, C. Linage, M. Rodell, and S. C. Swenson. Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water resources research*, 49 (2):904–914, 2013.

Y. Wada, L. P. H. van Beek, C. M. van Kempen, J. W. T. M. Reckman, S. Vasak, and M. F. P. Bierkens. Global depletion of groundwater resources. *Geophysical Research Letters*, 37(20), 2010. ISSN 1944-8007. doi: 10.1029/2010GL044571. URL `http://dx.doi.org/10.1029/2010GL044571`. L20402.

Y. Wada, L. P. H. van Beek, D. Viviroli, H. H. Dürr, R. Weingartner, and M. F. P. Bierkens. Global monthly water stress: 2. water demand and severity of water stress. *Water Resources Research*, 47(7):n/a–n/a, 2011. ISSN 1944-7973. doi: 10.1029/2010WR009792. URL `http://dx.doi.org/10.1029/2010WR009792`. W07518.

B. J. Wagner and S. M. Gorelick. Optimal groundwater quality management under parameter uncertainty. *Water Resources Research*, 23(7):1162–1174, 1987. ISSN 1944-7973. doi: 10.1029/WR023i007p01162. URL `http://dx.doi.org/10.1029/WR023i007p01162`.

J. Wahr, M. Molenaar, and F. Bryan. Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research: Solid Earth*, 103(B12):30205–30229, 1998. ISSN 2156-2202. doi: 10.1029/98JB02844. URL `http://dx.doi.org/10.1029/98JB02844`.

J. Wahr, S. Swenson, and I. Velicogna. Accuracy of GRACE mass estimates. *Geophysical Research Letters*, 33(6):n/a–n/a, 2006. ISSN 1944-8007. doi: 10.1029/2005GL025305. URL `http://dx.doi.org/10.1029/2005GL025305`. L06401.

G. R. Walker, I. D. Jolly, and P. G. Cook. A new chloride leaching approach to the estimation of diffuse recharge following a change in land use. *Journal of Hydrology*, 128(1):49 – 67, 1991. ISSN 0022-1694. doi: http://dx.doi.org/10.1016/0022-1694(91)90131-Z. URL `http://www.sciencedirect.com/science/article/pii/002216949190131Z`.

A. H. Welch, D. Westjohn, D. R. Helsel, and R. B. Wanty. Arsenic in ground water of the united states: Occurrence and geochemistry. *Ground Water*, 38(4):589–604, 2000. ISSN 1745-6584. doi: 10.1111/j.1745-6584.2000.tb00251.x. URL `http://dx.doi.org/10.1111/j.1745-6584.2000.tb00251.x`.

R. Whitehead. *Ground Water Atlas of the United States: Segment 8, Montana, North Dakota, South Dakota, Wyoming.* Number 730-I. U. S. Geological Survey, 1996.

D. Whittemore, M. Tsou, C. McElwee, et al. Arkansas river salinity and contamination of the High Plains aquifer. In *Challenges facing irrigation and drainage in the new millennium. Proceedings US Committee on Irrigation and Drainage, Fort Collins, Colorado, USA, June 2000.*, pages 225–246. US Committee on Irrigation and Drainage, 2000.

D. O. Whittemore. Geochemical differentiation of oil and gas brine from other saltwater sources contaminating water resources: Case studies from Kansas and Oklahoma. *Environmental Geosciences*, 2(1):15–31, 1995.

U. WWAP. United nations world water assessment programme. the world water development report 1: Water for people, water for life, 2003.

P. J.-F. Yeh, S. C. Swenson, J. S. Famiglietti, and M. Rodell. Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, 42(12):n/a–n/a, 2006. ISSN 1944-7973. doi: 10.1029/2006WR005374. URL `http://dx.doi.org/10.1029/2006WR005374`. W12203.