VULNERABILITY OF RECENTLY RECHARGED GROUNDWATER IN THE CALIFORNIA COASTAL BASINS TO NO_3^- CONTAMINATION

A thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
The requirement for
The degree

Master of Science
In
Geosciences

by

Gabriela Eliai Geyer

San Francisco, California

August 2014

CERTIFICATION OF APPROVAL

I certify that I have read **Vulnerability of Recently Recharged Groundwater in the California Coastal Basins to NO3– Contamination** by Gabriela Eliai Geyer, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Science in Earth and Climate Sciences at San Francisco State University.

Jason J. Gurdak
Assistant Professor of Earth and Climate Sciences

Leonard Sklar
Associate Professor of Earth and Climate Sciences

Leora Nanus
Adjunct Professor of Earth and Climate Sciences

VULNERABILITY OF RECENTLY RECHARGED GROUNDWATER IN THE CALIFORNIA COASTAL BASINS TO NO_3^- CONTAMINATION

Gabriela Eliai Geyer
San Francisco State University
2014

Nitrate (NO_3^-) is the most widespread non-point source (NPS) contaminant in groundwater. The California Coastal Basin (CCB) aquifers are vulnerable to contamination because of the heavy agricultural and industrial practices that contribute to NPS NO_3^-. Predicting the probability of non-point source (NPS) NO_3^- contamination in groundwater can be a valuable tool for managing water resources. Previous studies have identified important controls of NO_3^- contamination and developed logistic regression models that have successfully predicted the probability of NO_3^- exceeding background concentrations. However, there has not been a successfully calibrated model that represents the vulnerability in the CCB. My primary research objectives are to develop a model that better represents the vulnerability of groundwater in the CCB to NPS NO_3^- contamination. I will identify important controls of NO_3^- contamination in different parts of the CCB. My findings will provide useful information for resource managers and policy makers when making decisions about California's water resources.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee

8-12-2014
Date

ACKNOWLEDGMENTS

I would like to thank Jason Gurdak for his unwavering guidance and support throughout my time as a graduate student at San Francisco State University. He encouraged me to stay focused and work hard in my classes and in my research. Jason's dedication to his students is truly inspiring. I would also like to thank Leonard Sklar and Leora Nanus for their unbelievable patience, support, and confidence in me. They always led me in the direction I needed to go. Special thanks to Tendai Chitewere, Lisa White, and Vanessa Sorrell. Tendai put the idea in my head that I could be a scientist, Lisa helped me become a good scientist, and Vanessa helped me stay sane. And finally, I would like to thank my family and friends who always had faith in me that I could finish.

Partial funding for this research was provided by Research Institute for Humanity and Nature under the Human-Environmental Security in Asia-Pacific Ring of Fire: Water-Energy-Food Nexus. Funding was also provided by the Robert W. Maxwell Memorial Scholarship and the COAST Graduate Student Research Award. The support from these projects and scholarships paved my way to finishing this thesis. Thank you.

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF FIGURES vii
1.0 .... Introduction
.... 1
2.0 .... Study Area
.... 4
3.0 .... Methods
.... 5
3.1 Well selection 5
3.2 Groundwater-quality data 6
3.3 Compilation of explanatory variables in GIS 7
3.4 Logistic regression modeling 8
4.0 .... Results
.... 11
4.1 Univariate relations between STA factors and NO_3^- 11
4.2 Multivariate relations between STA variables and NO_3^- 15
4.3 Validation of the multivariate models 18
5.0 .... Discussion
.... 19
5.1 Univariate models 19
5.2 Multivariate models 19
5.3 Validation Models 25
6.0 .... Conclusions
.... 26
7.0 .... References
.... 28
TABLES 31
FIGURES 41

LIST OF TABLES

Table Page

1. Summary of nitrate (NO3–) concentrations in wells used for calibration and validation of the logistic regression models vi1
2. Description of source, transport, and attenuation explanatory variables 32
3. Results of the univariate logistic regression analysis 35
4. Model calibration and goodness-of-fit for the multivariate logistic regression models of NO3– greater than or equal to the relative background concentration in recently recharged groundwater.. 37
5. Parameters for logistic regression models for NO3– above relative background concentration (2 mg/L) in recently recharged groundwater of sub-regions of the California Coastal Basin (CCB) aquifer system 38
6. Model validation and goodness-of-fit for the multivariate logistic regression models of nitrate greater than or equal to the relative background concentration in recently recharged groundwater.. 39
7. Validation of parameters for logistic regression models for nitrate greater than the background concentration (2 mg/L) in recently recharged groundwater of sub-regions of the California Coastal Basin (CCB) aquifer system 40

LIST OF FIGURES

Figure Page

1. Map of California Coastal Basin aquifers and locations of wells 41
2. Map of calibration and validation wells used in Gurdak and Qi (2012) groundwater vulnerability study 45
3. Nitrate concentrations of all well sites in the CCB 46
4. Nitrate concentrations of calibration well sites of the CCB 47
5. Nitrate concentrations of validation well sites in the CCB 48
6. Map of ranges of nitrate concentrations of all well sites in the CCB 49
7. Map of ranges of nitrate concentrations of calibration well sites in the CCB .. 50
8. Map of ranges of nitrate concentrations of validation well sites in the CCB ... 51
9. Sales of nitrogen fertilizer in California, 1945-2008 52

1.0 Introduction

All groundwater resources are vulnerable to nonpoint source (NPS) contamination. Interest in predicting NPS contamination in groundwater has increased in recent years because of widespread detection of NPS contaminants and the implications for human and aquatic health and resource sustainability (Gurdak, 2008). Models that predict the vulnerability of groundwater to NPS contamination can provide an inexpensive tool to water-resource managers during policy decisions regarding groundwater protection, monitoring, and remediation strategies.

Along the California coast are more than 100 basins that comprise the California Coastal Basin (CCB) aquifers (Figure 1), which is one of 62 principal aquifers (PAs) in the U.S. PAs are regionally extensive aquifers and aquifer systems of national significance because of their high productivity and are critically important sources of potable water (Lapham et al., 2005). The CCB aquifers are important sources of water for drinking supplies, agriculture irrigation, and industry (California Department of Water Resources, 2003). In California, 43% of the population relies solely on domestic self-supply (public) wells that generally receive no water treatment prior to use (California Department of Water Resources, 2003). The dominant land-use/land-cover (LULC) across the CCB is urban and agriculture, which are well-known source of NPS contaminants in groundwater.

Nitrate (NO_3^-) is the most ubiquitous NPS contaminant of groundwater worldwide (Spalding et al., 1993). Intensive agricultural practices in California have contributed to chronic NO_3^- loading to groundwater due to the use of synthetically produced nitrogen fertilizers (Harter et al., 2012). Leaking septic tanks, sewage, and erosion of soils containing natural sources of NO_3^- also contribute to higher levels of NO_3^- in groundwater (Focazio et al., 2002). Drinking water with elevated NO_3^- concentrations is potentially toxic for humans because the oxygen-carrying hemoglobin is converted into methemoglobin and inhibits the movement of oxygen through the body (California Department of Water Resources, 2003). Infants that consume water with high levels of NO_3^- can develop methemoglobinemia (blue baby syndrome) that can lead to brain damage or death (Fewtrell, 2004). In response to these potential negative health effects, the U.S. Environmental Protective Agency (USEPA) has set the maximum contaminant level (MCL) of safe drinking water at 10 mg/L NO_3^- as nitrogen (N) (USEPA, 2012).

Previous studies have used logistic regression models to predict the likelihood of NO_3^- contamination to the CCB aquifers by using explanatory variables that include spatial and soil characteristics (Gurdak and Qi, 2012; Nolan, 2002). Gurdak and Qi (2012) developed a model for the entire CCB system using explanatory variables that represent the source, transport, and attenuation (STA) factors of NPS NO_3^-, but the model had poor predictive ability and may have been over-fit to the calibration data.

Although NO_3^- concentrations are highly variable in most groundwater systems, recently recharged groundwater is often the most vulnerable component of the flow system to NPS contamination (Gurdak and Qi, 2012). Recently recharged groundwater is defined here as recharge during the last 60 years, which coincides with widespread development of the CCB aquifer and expansion of urban and agricultural land use. The quality of most recently recharged groundwater has likely experienced human influence to some degree (Gurdak and Qi, 2012). Therefore, establishing a threshold based on NO_3^- concentration that distinguishes between natural and anthropogenic processes is useful when evaluating the effects of STA factors on NPS NO_3^- contamination in groundwater. I use 2.0 mg/L (as N) as the threshold between background (natural) and anthropogenic processes, which is consistent with studies that report background NO_3^- concentrations of 2.0 to 2.3 mg/L beneath forest, rangeland, and pasture areas of the U.S. (Mueller and Helsel, 1996; Nolan and Hitt, 2003).

The primary research objective of this study is to create a better calibrated model of the CCB with improved predictive capability by determining scale-dependent controls on NPS NO_3^- contamination (defined here as concentrations greater than background) in the most vulnerable part of the groundwater flow system. The primary objective is motivated by my hypothesis that some STA factors that control NPS NO_3^- contamination have scale-dependent relations at the sub-regional CCB aquifer scale. I also hypothesize that a larger and more representative compilation of well data will help develop better calibrated models with higher-degree of predictive ability. To test these hypotheses, I will develop univariate and multivariate logistic regression models for the northern CCB aquifers, central CCB aquifers, southern CCB aquifers, and for the entire CCB aquifer system. This study will provide fundamental information about the scale-dependent relations of STA factors in the CCB aquifer, which can be used by decision makers to develop better management strategies and policies that protect California's groundwater resources from NPS NO_3^- contamination.

2.0 Study Area

The CCB is comprised of more than 100 basins located along the coast of California. Faulting and folding in the tectonically active area created mainly northwest trending structural troughs that parallel the coastline. The intermontane basins consist of unconsolidated and semi-consolidated marine sediment and sedimentary deposits. Temperatures in the coastal regions of California are moderated by the ocean and therefore the temperature range in a day only varies by about 20 degrees Fahrenheit. The CCB is classified as having a Mediterranean climate where winters are cool and summers are warm. Runoff and precipitation are directly related in the CCB. Although coastal areas in California receive moderate to abundant precipitation, the demand for water throughout the state has led to substantial network of canals, aqueducts, and reservoirs built to accommodate the transportation of water. For this study, the CCB was divided into three sub-regions (north, central, and south) that have different hydrogeologic characteristics, LULC, and groundwater-quality issues. The CCB was also evaluated as a whole system and therefore incorporates the characteristics of each of the sub-regions.

3.0 Methods

3.1 Well selection

The spatial distribution of the 132 wells used by Gurdak and Qi (2012) in their previous CCB vulnerability model was sparse in some sub-regions and dense in other sub-regions of the CCB aquifer (Figure 2). To improve the spatial distribution of wells that intercept recently recharged groundwater, I selected additional wells with NO_3^- concentration data from the USGS National Water Information System (NWIS) database. Similar to Gurdak and Qi (2012), I used tritium (^3H) as the primary selection criterion for wells that intercept recently recharged groundwater. Before atmospheric testing of thermonuclear bombs began in the early 1950's, the tritium content of precipitation across the U.S. was approximately 8 Tritium Units (TU) (Thatcher et al., 1962). Tritium is radioactive, with a half-life of 12.43 years. Therefore, groundwater that is derived completely from precipitation that fell before the early 1950's atmospheric testing would contain less than 0.3 tritium units (TU) in 2014. Tritium values greater than 0.3 TU indicate that groundwater samples contain at least a portion of water that was recharged during the last 60 years and was used as selection criteria for wells that intercept recently recharged groundwater. If time-series groundwater-quality samples were collected at an individual well, the most recent sample was used in this study. This selection process removed variability caused by changes in NO_3^- concentrations over time.

After well selection, the number of wells in each sub-region of the CCB aquifer was randomly filtered using a geographic information system (GIS). The filtering removed neighboring wells from high-density spatial clusters to create a more uniform and consistent spatial density across each sub-region of the CCB aquifer. The filtering also eliminated double counting of GIS-based explanatory variables within a 500-m buffer radius around each well, and reduced the potential for spatial autocorrelation (Johnson and Belitz, 2009; Worrall and Kolpin, 2004).

The well selection and filtering steps resulted in 135 wells across the CCB aquifer, which includes 45 wells in each of the north, central, south sub-regions (Table 1). For purposes of developing logistic regression models in each CCB sub-region and for the entire CCB (as described below), two-thirds of the wells were randomly selected for model calibration and the remaining one-third for model validation (Table 1).

3.2 Groundwater-quality data

Concentrations of NO_3^- from the 135 wells ranged from 0.04 to 40.0 mg/L, with a median concentration of 1.45 mg/L (Table 1). Of the three sub-regions, the north CCB has the lowest NO_3^- concentrations (median, 0.20 mg/L and maximum 11.05 mg/L) (Table 1). The distribution of NO_3^- concentrations is somewhat similar in the central (median, 3.30 mg/L and maximum, 40 mg/L) and south (median, 2.19 mg/L and maximum, 34 mg/L) sub-regions (Table 1). Of the 135 selected wells, 57 wells have NO_3^- above the threshold of 2 mg/L (38 calibration wells [North = 7; Central = 18; South = 13] and 19 validation wells [North = 2; Central = 7; South = 10]) (Figures 3-8).

3.3 Compilation of explanatory variables in GIS

Following methods from Gurdak and Qi (2012), I applied a 500-m circular buffer around each of the 135 wells to extract GIS-based STA factors (Table 2) that were tested as explanatory variables in the logistic regression modeling. A 500-m circular buffer is commonly used to delineate land-use/land-cover (LULC) that potentially affects groundwater of concern (Eckhardt and Stackelberg, 1995; Nolan et al., 2002).

The GIS-based explanatory variables represent STA factors that may control NPS NO_3^- levels above the relative background concentration (Gurdak and Qi, 2012). Source variables represent NO_3^- loading (farm fertilizer, manure from confined animal feeding operations, land use/land cover including cultivated crops and irrigated cropland, atmospheric NO_3^- and total N, population density, and aqueous geochemical indicators in groundwater); transport variables represent NO_3^- mobilization in the soil, unsaturated zone, and saturated zone to the well (water inputs from precipitation and runoff, hydrologic and geochemical properties of soil and aquifer material, depth to the water table, depth to the well screen below the water table, recharge rates and selected management practices); and attenuation variables represent denitrification and/or dilution of NO_3^- (Table 2). Reduction/oxidation (redox) conditions are an important control on groundwater vulnerability to NO_3^- contamination (Gurdak and Qi, 2012). Dissolved oxygen (DO) is an important factor to consider in NO_3^- vulnerability assessments because under anoxic conditions, denitrification occurs. When oxygen levels are depleted, microorganisms prefer to use NO_3^-, a process in which NO_3^- is reduced to nitrogen gas (N_2), the more stable and less harmful form of nitrogen (McMahon and Chapelle, 2008). The presence of dissolved and particulate organic carbon is significant in the redox process as well because carbon is the most common electron donor available in groundwater systems (Thurman, 1985).

3.4 Logistic regression modeling

I developed logistic regression models using JMP® statistical software and the GIS-based explanatory variables (STA factors) to predict the vulnerability of groundwater to NPS NO_3^- above the background concentration (2 mg/L). Logistic regression has been widely used in groundwater vulnerability assessments because it predicts the probability of a binary response using a threshold that is meaningful for specific management problems (Gurdak, 2008). Logistic regression is applicable for non-parametric and dichotomous (binary) data, which often characterize environmental and groundwater-quality data. The response variable is established using a binary threshold, which is commonly set at a drinking-water standard, laboratory detection level, or relative background concentration (Nolan, 2002; Rupert, 1998; Tesoriero et al., 1997).

I used a number of statistical parameters that are calculated during logistic regression to help evaluate how well the overall model works, how important each of the explanatory variables are in the overall model, and if the form of the model appears to be correct (Menard, 2002). I also evaluated the predictive ability of the overall logistic regression model. I used the log-likelihood ratio (LLR) to measure the success and significance of the logistic regression model as a whole by comparing observed with predicted values (Hosmer and Lemeshow, 1989). The highest LLR indicates the most significant model, taking into account the degrees of freedom (number of explanatory variable) in the model. The p-values of the LLR indicate model significance of the model coefficients.

I used the partial likelihood ratio, percent correct (PC) responses, model sensitivity, and area under the ROC (Receiver Operating Characteristic) curve to evaluate the logistic regression model-fitting (Hosmer and Lemeshow, 1989; Menard, 2002). The partial likelihood ratio is similar to the LLR, but is evaluated to determine the significance of adding one or more new variables to an existing multivariate logistic regression model (Helsel and Hirsch, 1992). A model with the addition of one new variable is more significant than the original model if the partial likelihood ratio is greater than the value of the chi-square distribution with degrees of freedom equal to one (Helsel and Hirsch, 1992). The partial likelihood ratio was used exclusively during the iterative processes of the multivariate logistic regression analysis to select the explanatory variables that produce the best fitting model. Because of the large number of iterations, partial likelihood ratios and corresponding preliminary multivariate models are not listed in this thesis. The overall rate of correct classification, or percent correct (PC) responses, is the number of observed exceedances predicted by the model as exceedances, plus the number of observed nonexceedances predicted as nonexceedances, divided by the combined number of observed exceedances and nonexceedances (Hosmer and

Lemeshow, 1989). Sensitivity is defined as the number of observed exceedances predicted as exceedances divided by the total number of observed exceedances. Higher values of PC and sensitivity indicate better fitting models. The area under the ROC curve, represented by the c statistic, is a measure of the model's ability to discriminate between groundwater samples with NO_3^- greater than or equal to the background concentration and those samples that do not. Hosmer and Lemeshow (1989) suggest that 0.7 less than c statistic less than 0.8 is acceptable discrimination.

I evaluated model calibration using the degree of correspondence between the predicted probabilities of NO_3^- exceeding the threshold and the actual NO_3^- concentrations exceeding the threshold. The Hosmer-Lemeshow (HL) goodness-of-fit test statistic was used to evaluate the model calibration. The null hypothesis of the HL test is that the model fits the data; therefore, a higher HL p-value indicates a well-calibrated model (Hosmer and Lemeshow, 1989).

Because the explanatory variables were reported in different units, I standardized the coefficients after final model selection using the standardization technique outlined by Menard (2002). The advantage to standardized coefficients is that the relative impact and magnitude of effect of the explanatory variables can be directly compared.

Problems with the model may arise if strongly correlated explanatory variables are included in a multivariate logistic regression model (Hosmer and Lemeshow, 1989). Multicollinearity, or strong correlations between two or more explanatory variables, may inflate the variance of the parameter estimates and cause a lack of statistical significance of individual explanatory variables, even though the overall model may be strongly significant (Hosmer and Lemeshow, 1989). Incorrect conclusions about relations in the model may be drawn if multicollinearity is present because an unrealistic model coefficient sign or unstable slope coefficients may result (Hosmer and Lemeshow, 1989). To detect multicollinearity, Pearson correlation coefficients and multicollinearity diagnostic statistics were examined during model development and selection. A Pearson's correlation coefficient greater than 0.7 indicates there is a strong correlation between two explanatory variables.

4.0 Results

4.1 Univariate relations between STA factors and NO_3^-

Univariate relations between NO_3^- concentration greater than or equal to 2 mg/L and explanatory variables were evaluated and are summarized in Table 3. The coefficients listed in Table 3 indicate the nature of the univariate relation; coefficient values greater than zero indicate positive relations, and coefficient values less than zero indicate inverse relations with NO_3^- greater than or equal to 2 mg/L. An alpha level of 0.2 was chosen as the inclusion criteria for selecting explanatory variables into the multivariate analysis rather than the more traditional alpha level of 0.10. Hosmer and Lemeshow (1989) suggest that an alpha level of 0.10 has failed to identify variables known to be important during some multiple logistic regression analyses.

Results of the univariate analysis indicate that 15 explanatory variables are statistically significant in the north CCB aquifer (Table 3). The three significant source variables in the north CCB include farm fertilizer, atmospheric NO_3^-, and calcium concentration in groundwater. The significant source variables all had positive coefficients (Table 3), which indicates a positive relation between increases in N sources and increases in the probability of NO_3^- exceeding the relative background concentration. Of the nine significant transport variables in the north CCB, five have negative coefficients and four have positive coefficients (Table 3). The variables that have negative coefficients represent processes that impede transport and include recharge, precipitation, soil thickness, soil group B, and annual runoff. The variables that have positive coefficients represent processes that promote transport in the north CCB and include soil group D, upper soil erodibility, soil wind erodibility, and fresh surface water withdrawal. The significant attenuation variables include the positively correlated DO and temperature, as well as the negatively correlated tritium and manganese concentration in groundwater (Table 3).

Results of the univariate analysis indicate that 15 explanatory variables are statistically significant in the central CCB aquifer (Table 3). The significant source variables include the positively correlated open space (recreational vegetation, such as lawn grass), as well as the negatively correlated shrubs and wetland. Of the eight significant transport variables, soil slope, soil permeability, soil bulk density, and soil sand have positive coefficients, while soil<No. 200 sieve, soil >No. 10 sieve, and soil loss tolerance factor have negative coefficients. The significant attenuation variables include the positively correlated DO, oxic redox, and tritium, as well as the negatively correlated iron concentration and manganese concentration.

Results of the univariate analysis indicate that 22 explanatory variables are statistically significant in the south CCB aquifer (Table 3). The significant source variables include medium intensity development, high intensity development, and atmospheric NO_3^-, all of which have positive coefficients. Most of the 16 transport variables have positive coefficients, including soil group B, soil erodibility, upper soil erodibility, soil available water capacity, soil<No 4 sieve, soil<No. 200 sieve, soil<No. 10 sieve, and soil silt. Transport variables that have a negative coefficient are altitude, recharge, precipitation, soil group C, soil bulk density, soil sand, soil loss tolerance factor, and annual runoff. The significant attenuation variables include DO, seasonally high water table, and manganese concentration. DO has a direct relation to NO_3^- while the latter two are inversely related to NO_3^-.

Results of the univariate analysis indicate that 22 explanatory variables are statistically significant across the entire CCB aquifer (combined north, central, and south) (Table 3). The significant source variables include farm fertilizer and medium intensity development, which are positively related to NO_3^- concentration, as well as shrubs, grassland, and crops, which are inversely related to NO_3^-. Of the 13 significant transport variables, Hortonian overland flow, upper soil erodibility, soil<No. 4 sieve, soil<No. 10 sieve, and fresh surface water withdrawal have positive coefficients. The remaining significant transport variables have negative coefficients and include altitude, recharge,

precipitation, soil thickness, soil group C, soil loss tolerance factor, soil wind erodibility, and annual runoff. The significant attenuation variables are DO and temperature, which have positive coefficients, and iron concentration and manganese concentration, which have negative coefficients.

The univariate analysis helps distinguish the scale-dependent relations between some STA factors and NPS NO_3^- concentrations greater than background. Some variables were significant only at the sub-regional scale or were significant in multiple sub-regions. Of the source variables, only atmospheric NO_3^- was significant in multiplesub-regions (north and south) (Table 3). Transport variables that were significant in more than one sub-region include recharge, precipitation, soil group B, upper soil erodibility, soil bulk density, soil<No. 200 and 10 sieve, soil sand, soil loss tolerance factor, and annual runoff. Of these variables, all exhibited the same sign (- or +) among the sub-regions in which they were significant, except for soil group B, soil bulk density, soil<No. 200 and 10 sieve, and soil sand. From the attenuation variables, DO is significant in all four regions while manganese concentration is significant in every region except the central CCB.

Some variables were also significant at the sub-regional and regional scales. The significant variables in the entire CCB univariate models overlapped with 11 in the north, five in the central, and 12 in the south sub-regions (Table 3). The significant variables specific to just one sub-region are also significant in the entire CCB univariate modeling. The north CCB's common variables with the entire CCB model are: farm fertilizer (S), soil thickness (T), soil wind erodibility (T), fresh surface water withdrawal (T), and temperature (A). The central CCB only had shrubs (S) and iron concentration (A) in common with the entire CCB model. The south CCB shared medium intensity development (S), altitude (T), soil group C (T), and soil No. 4 sieve (T) with the entire CCB univariate model. Interestingly, only grassland (S), and crops (S) are significant in only the regional scale (CCB only) models and not significant at the sub-regional scale.

4.2 Multivariate relations between STA variables and NO_3^-

Of the 78 explanatory variables, 41 were initially carried forward for multivariate analyses based on the alpha level of 0.2. However, all explanatory variables were evaluated later using the partial likelihood ratio during multivariate analyses. The variable selection for multivariate model development required too many iterative steps to list in this thesis. Details of the final multivariate logistic regression models are presented in Tables 4 and 5.

The log-likelihood p-value (<0.001) for the north CCB model indicates high statistical significance and the overall model fit was excellent (HL p-value = 0.944) (Table 4). Variables that comprised the best multivariate model for the north CCB are farm fertilizer (S), DO (A), and soil thickness (T) (Table 5). Farm fertilizer and DO have positive coefficients, representing a direct relation with the probability of NO_3^- exceeding the background concentration, while soil thickness has an inverse relationship. Pearson correlation coefficient statistics indicate that each of the variables in the best north CCB multivariate model are not strongly correlated to each other.

The log-likelihood p-value (0.013) for the central CCB indicates good statistical significance, but the overall model fit was poor (HL p-value = 0.295) (Table 4). A best model for the central CCB was found that includes open space (S) and soil group D (T) (Table 4). The open space variables has a positive coefficients, representing a direct relation to NO_3^- concentration, but the soil group D has a negative coefficient and represents an inverse relation to NO_3^- concentration. Pearson correlation coefficient statistics indicate that the two variables are independent of one another. A second model was found for the central CCB that only included DO (A) as a significant variable (Table 5). DO has a positive correlation to NO_3^- concentration in the central CCB. The log-likelihood p-value (p-value <0.001) indicates high statistical significance and the overall model fit was excellent (HL p-value = 0.993) (Table 4).

Similar to the central CCB multivariate model, the log-likelihood p-value (<0.001) for the south CCB indicates high statistical significance in the model, but overall poor model fit (HL p-value = 0.375) (Table 4).The best south CCB multivariate model included low intensity development (S), crops (S), and DO (A) (Table 5). All of the variables in the best model have positive coefficients and are positively related to NO_3^-concentrations. Pearson correlation coefficient statistics of the variables in the southern region indicate that the variables in the model are not strongly correlated to each other.

The log-likelihood p-value (<0.001) for the entire CCB indicates high statistical significance in the model and very good overall model fit (HL p-value = 0.759) (Table 4). The best multivariate model found for the entire CCB consists of DO (A), soil thickness (T), soil available water capacity (T), and farm fertilizer (S) (Table 5). Soil thickness has a negative coefficient in the model, therefore an increase in soil thickness decreases the probability of NO_3^- exceeding the threshold. Farm fertilizer, soil thickness, and soil available water capacity all have a positive correlation. Pearson correlation coefficient statistics indicate that each of the variables in the best multivariate model of the entire CCB are not strongly correlated to each other.

4.3 Validation of the multivariate models

The multivariate models (Tables 4 and 5) were validated to evaluate predictive ability of the probability of NO_3^- greater than the background concentration at the validation wells. The explanatory variables that were determined to be the most significant in each of multivariate models (north CCB, central CCB, south CCB, and CCB) were used in the validation process. The probability of each validation exceeding the NO_3^- threshold (2 mg/L) was calculated using logistic regression in JMP. (Table 6). Coefficients from the logistic regression models were used in the equation along with the explanatory variables' values (Table 6).

The validation indicates poor predictive ability of the north CCB model (Table 7). Although each explanatory variable coefficient had the same signs (+/-) observed in the calibration models, the p-values were very high (Table 7). The overall model is poor (LLR p-value = 0.912).

Similarly, the validation indicates poor predictive ability of the central CCB model (Table 7). The explanatory variables had the same relation to NO_3^- concentration, but the p-values were very high (Table 7). The overall model is poor (LLR p-value = 0.889). However, the validation indicates good predictive ability of the model B in the central CCB (Table 7). DO, the only parameter in the model, is statistically significant (p-value = <0.001). The overall validation model is significant (LLR p-value = <0.001) and the overall fit was excellent (HL = 0.997).

The validation indicates good predictive ability of the south CCB model (Table 7). Each explanatory variable had the same relation to NO_3^- concentrations as was observed in the calibration model. The statistical significance of crops (S) and DO (A) was fairly good (p-value = 0.174 and 0.107, respectively). The overall south CCB validation model is significant (LLR p-value = <0.001) and the overall fit was excellent (HL = 0.926).

The validation indicates moderately good predictive ability of the CCB model (combined north, central, and south) (Table 7). The relation of each explanatory variable to NO_3^- concentrations is the same as the relations observed in the calibration model. Soil thickness (T) and DO (A) are statistically significant (p-value = 0.182 and 0.098, respectively). Farm fertilizer (S) and soil available water capacity (T), however, were not statistically significant. The overall model was not significant (LLR p-value = 0.295).

5.0 Discussion

5.1 Univariate models

The central CCB model did not have one single expected NO_3^- source. In fact, the central model had very bad p-values for N sources that are expected to correlate to N concentrations. The N loading variables that were significant in the univariate modeling of the central CCB (shrubs, open space, and wetland) do not account for sources of NO_3^-. 11 out of 20 of the well sites that exceeded the relative NO_3^- background concentration had a zero percentage of farm fertilizer application, confined manure application, crops, and irrigated cropland.

Iron was only important in central CCB and the entire CCB. There may be some implication about how much a sub-region like the central CCB affects the overall model for the whole CCB.

A Pearson coefficient correlation analysis of the variables making up the best model of the central CCB revealed strong correlations between DO and 7 of the other 14 significant univariate variables.

5.2 Multivariate models

The north CCB multivariate analysis resulted in explanatory variables that make conceptual sense. Farm fertilizer (S) and DO (A) are both positively correlated to NO_3^- concentration, therefore an increase in farm fertilizer or DO correlates to a higher concentration of NO_3^-. Farm fertilizer usually contains high amounts of NO_3^- so high amounts of fertilizer application should reflect high NO_3^- concentrations. Similarly, the presence of DO in a system impedes denitrification (McMahon and Chapelle, 2008) and thus NO_3^- cannot reduce to the less harmful molecular nitrogen (N_2). Soil thickness is inversely related to NO_3^- in the model, suggesting that an increase in soil thickness restricts the transportation of NO_3^- into groundwater. Overall, the explanatory variables and their relation to NO_3^- concentration are acceptable conceptually.

In model A of the central CCB, the percentage of open space (S) is positively related to NO_3^- concentrations (Table 5). A greater percentage of open space indicates greater concentrations of NO_3^-. Open space refers to areas characterized by a mixture of constructed material and vegetation. The majority of the vegetation in these areas is lawn grass or vegetation planted for recreational purposes or erosion control. Fertilizer applied to lawns has been linked to NO_3^- contamination (source) so we expect to see a positive relation to NO_3^-. Soil group D (T) is inversely correlated with NO_3^-. Thus, lower NO_3^- concentrations in the central CCB correlated to a higher percentage of soil group D. Soil group D is characterized as having very slow infiltration and transmission rates when thoroughly wet and thus acts to impede the transport of NO_3^- to the groundwater.

Model B of the central CCB only includes DO (A) as an explanatory variable (Table 5) and has an excellent model (Table 4). A positive correlation between NO_3^- and DO is well documented (McMahon and Chapelle, 2008) and therefore the significance of the variable in the model is to be expected. However, in the multivariate analysis, including other variables along with DO made the model unstable. It is probable that DO is the leading control in the sub-region and trumps other variables from showing any significance. However, it may also be possible that the proper variables to adequately describe NPS NO_3^- contamination above the background in the central CCB were identified in the study. Many of the source variables in the central CCB have negative coefficients, inversely relating NO_3^- concentration to NO_3^- input. However, more than half of the calibration well sites were above the background concentration threshold and the sub-region contained the highest concentration of NO_3^- recorded within the data set. According to the Groundwater Vulnerability Study (2010) of Santa Clara County, NO_3^- concentrations are beginning to decline in the Santa Clara valley of the central CCB sub-region (Todd Engineers, Kennedy/Jenks Consultants, 2010). The observed trend in land use (2001 to 2009) of the area includes a gradual shift from agricultural land to suburban housing and a decrease in the number of feedlots in the area. The NO_3^- concentrations observed in the central CCB may therefore be a reflection of past land use and so the data used to assess land use (collected in 2001) may not be representative of historical NO_3^- sources. It may also be probable that a 500-m buffer extraction is not sufficient to represent the contributing area for each well, however further investigation is needed to support this theory.

The variables in the best model for the south CCB include crops (S), low development (S), and DO (A) (Table 5). Crop is characterized by the area designated for production of annual crops. The area of crop production is directly related to NO_3^- concentrations, a reflection of the nitrogen input due to fertilizer application (crops and farm fertilizer have a high Pearson correlation in the south CCB). Low development is defined as an area with a mixture of constructed materials and vegetation with single family homes being common attributes. The direct relation between low development areas and NO_3^- concentration is likely due to leaking septic tanks from single family homes and (or) the application of lawn fertilizers. Similar to the other sub-regions' multivariate models, DO is also positively correlated to NO_3^- concentration in the south CCB.

The CCB (north, central, and south) best model includes farm fertilizer (S), soil thickness (T), available water capacity (T), and DO (A) (Table 5). Farm fertilizer has a positive coefficient, representing a positive correlation to NO_3^- concentration. Farm fertilizer was an important explanatory variable in the north CCB as well, indicating that the significance of the STA factor is scale invariant (significant across sub-regional and regional scales). Crop production, an explanatory variable in the south CCB model, is highly correlated to farm fertilizer in the CCB data set. Therefore, farm fertilizer as an explanatory variable of the whole CCB system may also be representing crop production important in the south CCB. Soil thickness was only important in the north CCB and the entire CCB models. The north CCB and CCB models also share DO as an explanatory variable. The only variable that is specific to the CCB model is available water capacity. Available water capacity (AWC) has a positive coefficient in the model, thus an increase in AWC corresponds to an increase in NO_3^- concentration and likely represents increased transport of NO_3^- to the groundwater.

DO was significant in every best multivariate model, which is not surprising given the well-known relation between DO and NO_3^-. Similar to findings by previous studies (Tesoriero et al., 1997; Nolan et al., 2002 and 2003; and Gurdak and Qi, 2006 and 2012), DO is a major control on NO_3^- concentrations. The presence of DO in the well samples of this study proved to be the most significant control on NO_3^- concentrations in the CCB. Future groundwater vulnerability studies of the CCB should therefore consider DO as a major control on NO_3^- levels. Likewise, water management agencies should implement measuring DO during water quality sampling as a standard practice. During the collection of water quality data for this study, it was found that DO is not regularly collected during water sampling. Considering the importance of DO on denitrification and the relative inexpensive tools used to measure it, DO should be regularly monitored.

Although DO is the most important variable in the models, it should be noted that a major goal of studies such as these is to be able to predict NO_3^- concentrations where wells do not exist. However, DO can only be measured where a well is present. It may therefore be beneficial to conduct a study where DO concentrations are predicted first using logistic regression models, then apply the findings to NO_3^- vulnerability models. Pearson correlations might also provide clues as to what non-geochemical variables are associated with DO concentrations.

Future multivariate models of the CCB should also take climate change into consideration. The vulnerability model I created for the CCB uses variables that were collected in the past, meaning my model is temporal. If the goal of a vulnerability model is to be able to predict nitrate concentrations by finding what variables control concentrations, climate change scenarios that can affect these variables need to be considered. Climate change can affect geochemical variables, and possibly land use practices. Increasing temperature and extreme weather events can have negative effects on crop production, decreasing yields. Increased temperatures due may also increase the demand for irrigation (Anderson et al. 2008), possibly further decreasing yields. The risk of decreasing yields may lead farmers to shift to more profitable (and less water-demanding) crops. If these crops require different amounts of nitrogen fertilizer, we may increase or decrease the vulnerability of nitrate contamination in the CCB.

The role of CO_2 as a fertilizer should also be considered in NO_3^- groundwater vulnerability models. A global climate model created by Salmon-Monviola et al. (2013) resulted in an increase in denitrification when CO_2 concentrations in soils were high. However, the denitrification process produces nitrous oxide (N_2O), the fourth largest greenhouse gas contributing to climate change (EPA, 2010). Since N_2O is a greenhouse gas, the potential to further exacerbate climate change effects could occur, creating a positive feedback. Initially, an increase in denitrification could be seen as a benefit to increased CO_2 in the atmosphere because the increase in denitrification implies that less NO_3^- will leach into groundwater. However, solely focusing on denitrification can be misleading because the quantity of NO_3^- fertilizer application is increasing (Figure 9). Therefore, a groundwater vulnerability model of the CCB would benefit from incorporating the positive feedback associated with the deposition of CO_2 in soil and its effects on climate change.

5.3 Validation Models

The validation model for the north CCB is poor. Because the north CCB and CCB (north, central, and south) models shared similar explanatory variables, it is likely that the explanatory variables are still significant in the north. A possible explanation for the poor statistical significance in the validation model is that only 2 of the 15 validation wells are above NO_3^- threshold (2 mg/L). The imbalance in wells above and below the threshold does not support a robust model. The median NO_3^- concentration in the north of all the wells is 0.2 mg/L, relatively far from the 2 mg/L threshold set in the logistic regression modeling. A different threshold in the north should therefore be used in future models because the low concentrations in the north cannot have a normal distribution of data with the relatively high threshold set for the entire CCB.

The validation model for the central CCB (model A) is poor. The explanatory variables in the model, open space (S) and soil group D (T), are not intuitively correlated to NO3-. Therefore, it is possible that the statistical significance of the variables observed in the calibration model A is just noise and does not necessarily represent the important controls in the system. Conversely, model B had very good statistical significance in the validation analysis. DO appears as an important explanatory variable in each of the models, so the probability of it being noise is not likely. The high correlation between DO and NO_3- concentrations in the central CCB is evident; however, there may be other factors that control NO3- in the sub-region. During the calibration analysis, the inclusion of other variables along with DO created an unstable model. It is probable that DO alone is very important, yet reasoning as to why it is the only explanatory variable in the model is still unclear.

The south CCB validation analysis had moderately good results. Crops (S) and DO (A), were both statistically significant (p-value <0.2). Although low intensity development (S) was not statistically significant, the overall model was excellent (LLR p-value = <0.001; HL = 0.926).

Overall, the validation models were poor. It is likely that there were not enough validation wells for each sub-region so the validation models were not robust. Also, the difference in NO3- concentrations between sub-regions, most notably the north, could account for the poor validation models. However, the poor validation models should not be used to evaluate the goodness of the calibration models because they do not disprove that the significant variables found in the calibration models predict NO3- concentrations in the CCB.

6.0 Conclusions

The variables identified as controlling factors on NO3- concentrations in the CCB can help water management agencies identify areas of the aquifer that are vulnerable to NO3- contamination. Different factors were important in different sub-regions of the CCB, therefore scale is shown to be very important in this study. A major finding is that factors can be scale invariant. Land-use is scale dependent because of the difference in land management practices between the sub-regions. DO, farm fertilizer, and soil thickness are scale invariant because they are important factors both regionally and sub-regionally. DO is the most important controlling factor of this study and should be monitored on a regular basis so that future studies have a wider and more complete data set with which to work.

7.0 References

Anderson, J., Chung, F., Anderson, M., Brekke, L., Easton, D., Ejeta, M., Peterson, R., Snyder, R., 2008, "Progress on incorporating climate change into management of California's water resources." *Climate Change*, vol. 89, pp 91-108.

California Department of Water Resources, 2003, *California's Groundwater: Bulletin 188*, 14 p.

Fewtrell, L., 2004, "Drinking-water nitrate, methemoglobinemia, and global burden of disease: A discussion," *Environmental Health Perspectives*, vol. 112, no 14, 1371-1374 p.

Eckhardt, D.A.V.; Stackelberg, P.E., 1995, "Relation of ground-water quality to land use on Long Island, New York," *Ground Water*, vol. 33, no. 6, 1019-1033 pp.

Fewtrell, L., 2004, "Drinking-water nitrate, methemoglobinemia, and global burden of disease: a discussion. *Environ. Health Perspectives*, vol. 112, no. 14, 1371–1374.

Focazio, M. J.; Reilly, T. E.; Rupert, M. G.; Helsel, D. R., 2002, *Assessing ground-water vulnerability to contamination — providing scientifically defensible information for decision makers*, Circular 1224; U.S. Geological Survey: Reston, VA, 33 p.; http://pubs.usgs.gov/circ/2002/circ1224/.

Gurdak, J. J. *Ground-water vulnerability: Nonpoint-source contamination, climate variability, and the High Plains aquifer*, 2008, VDM Verlag Publishing: Saarbrucken, Germany, ISBN: 978-3-639-09427-5, 223 p.

Gurdak, J.J., and S. L. Qi, 2012, "Vulnerability of Recently Recharged Groundwater in Principle Aquifers of the United States to Nitrate Contamination," *Environmental Science & Technology*, vol. 46, no. 11, 6004–6012. doi:10.1021/es300688b.

Harter, T., J. R. Lund, J. Darby, G. E. Fogg, R. Howitt, K. K. Jessoe, G. S. Pettygrove, J. F. Quinn, J. H. Viers, D. B. Boyle, H. E. Canada, N. DeLaMora, K. N. Dzurella, A. Fryjoff-Hung, A. D. Hollander, K. L. Honeycutt, M. W. Jenkins, V. B. Jensen, A. M. King, G. Kourakos, D. Liptzin, E. M. Lopez, M. M. Mayzelle, A. McNally, J. Medellin-Azuara, and T. S. Rosenstock, 2012, "Addressing Nitrate in California's Drinking Water With A Focus on Tulare Lake Basin and Salinas Valley Groundwater," Report for the State Water Resources Control Board

Report to the Legislature, Center for Watershed Sciences, University of California, Davis, 87p.

Helsel, D. R.; Hirsch, R. M. *Statistical methods in water resources*, 1992, Elsevier Science Publishing Company, Inc., New York, NY, 522 p.

Hosmer, D. W.; Lemeshow, S. Applied logistic regression; 2000, John Wiley & Sons, Inc.: New York, 373 p.

Johnson, T. D.; Belitz, K. 2009, Assigning land use to supply wells for the statistical characterization of 636 regional groundwater quality: Correlating urban land use and VOC occurrence. *J. Hydrol.,* 637 370(1–4), 100–108.

JMP®, Version 10, SAS Institute Inc., Cary, NC, 1989-2007.

Lapham, W. W.; Hamilton, P. A.; Myers, D. N. 2005, National water quality assessment program Cycle II Regional assessments of aquifers, Fact Sheet 2005-3013; U.S. Geological Survey: Reston, VA,; 4 p. http://pubs.usgs.gov/fs/2005/3013/ (accessed May 22, 2012).

McMahon, P. B.; Chapelle, F. H. 2008, Redox processes and water quality of selected principal aquifer systems. *Ground Water* 46(2), 259–271.

Menard, S. *Applied logistic regression analysis*, 2002, Sage Publications: Thousand Oaks, CA, 111 p.

Mueller, D. K.; Helsel, D. R. *Nutrients in the Nation's waters – Too much of a good thing?*, 1996, Circular 1136; U.S. Geological Survey: Reston, VA, 24 p.

Nolan, B.T.; Hitt, K.J.; Ruddy, B.C., 2002, "Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States," *Environmental Science Technology*, vol. 40, no. 24, 7834-7840 pp.

Nolan, B. T.; Hitt, K. J., 2003, *Nutrients in shallow ground waters beneath relatively undeveloped areas of* 641 *the conterminous United States*, Water-Resources Investigations Report 02-4289; U.S. Geological 642 Survey: Reston, VA, 2003, 17 p.

Rosenstock, T.S.; Liptzin, D.; Six, J.; Tomich, T.P., 2013, Nitrogen fertilizer use in California: Assessing the data, trends and a way forward, *California Agriculture* 67(1):68-79

Ruddy, B.C.; Lorenz, D.L.; Mueller, D.K., 2006, *County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001*: U.S. Geological Survey Scientific Investigations Report 2006-5012, 17 p.

Rupert, M. G. *Probability of detecting atrazine/desethyl-atrazine and elevated concentrations of nitrate (NO_2+NO_3-N) in ground water of the Idaho part of the upper Snake River Basin*, 1998, Water-Resources Investigations Report 98–4203; U.S. Geological Survey: Boise, ID, 32 p.; http://id.water.usgs.gov/PDF/wri984203/.Salmon-Monviola, J., Moreau, P., Benhamou, C., Durand, P., Merot, P., Oehler, F., Gascuel-Odoux, C., 2013, Effect of climate change on increased atmospheric CO2 on hydrological and nitrogen cycling in an intensive agricultural headwater catchment in western France, *Climatic Change*, vol. 120, pp.433-447.

Spalding, R. F.; Exner, M. E. 1993, Occurrence of nitrate in groundwater – A Review. *J. Environ. Qual*., 22(3), 392–402.

Tesoriero, A. J.; Voss, F. D. 1997, Predicting the probability of elevated nitrate concentrations in the Puget Sound Basin – Implications for aquifer susceptibility and vulnerability. *Ground Water* 35(6), 1029–1039.

Thatcher, L. L., 1962, The distribution of tritium fallout in precipitation over North America. *Bull. Internal. Assoc. Sci. Hydrol.* 7(2), 48–58.

Thurman, E.M., 1985, Organic Geochemistry of Natural Waters. Martinus Nijhoff/DR W. Junk: Boston, Massachusetts, 489 p.

Todd Engineers, Kennedy/Jenks Consultants, 2010, Groundwater Vulnerability Study Santa Clara Valley Water District.

U.S. Environmental Protection Agency, 2012, "Drinking water contaminants," Accessible at http://epa.gov/drink/contaminants/index.cfm

Worrall, F.; Kolpin, D., 2004, Aquifer vulnerability to pesticide pollution – combining soil, land-use and 639 aquifer properties with molecular descriptors. *J. Hydrol.*, 293(1–4), 191–204.

TABLES

Table 1. Summary of nitrate (NO_3^-) concentrations in wells used for calibration and validation of the logistic regression models.

	North	Central	South	CCB-All
Study area (km^2)	4614	9552	12263	26429
NO_3^- threshold (mg/L)	2	2	2	2
Subset for model calibration				
Number	30	30	30	90
Maximum	11.05	40	25.9	40
75th percentile	1.64	5.79	6.88	4.78
50th percentile	0.25	3.34	1.64	1.09
25th percentile	0.04	0.16	0.09	0.06
Minimum	0.03	0.05	0.04	0.03
Subset for model validation				
Number	15	15	15	45
Maximum	3.99	28.6	34	34
75th percentile	1.02	5.80	5.63	3.92
50th percentile	0.19	3.11	2.67	1.71
25th percentile	0.04	0.08	1.66	0.11
Minimum	0.026	0.05	0.79	0.03
All data (calibration and validation)				
Number	45	45	45	135
Maximum	11.05	40	34	40
75th percentile	1.26	5.70	6.10	4.31
50th percentile	0.20	3.30	2.19	1.45
25th percentile	0.04	0.08	0.56	0.07
Minimum	0.026	0.05	0.04	0.03
[mg/L, milligrams per liter]				

Table 2: Description of source, transport, and attenuation explanatory variables.

Explanatory Variable	Description	Source
Nitrogen Source		
confined manure	Average Nitrogen from confined animals in 500 m buffer (kg/km^2 of cropland) (average of 1982, 1987, 1992, and 1997)	Ruddy et al., 2006
farm fertilizer	Average Nitrogen application from farms in 500 m buffer (kg/km^2 of cropland) (averaged from 1987-2006)	Ruddy et al., 2006
open water	Open water in 500 m buffer (%)	LaMotte, 2008
open space	Developed land cover (open space) in 500 m buffer (%)	LaMotte, 2008
low intensity development	Developed land cover (low intensity) in 500 m buffer (%)	LaMotte, 2008
medium intensity development	Developed land cover (medium intensity) in 500 m buffer (%)	LaMotte, 2008
high intensity development	Developed land cover (high intensity) in 500 m buffer (%)	LaMotte, 2008
barren land	Barren land (includes rock, sand, and clay) in 500 m buffer (%)	LaMotte, 2008
deciduous forest	Deciduous forest in 500 m buffer (%)	LaMotte, 2008
evergreen forest	Evergreen forest in 500 m buffer (%)	LaMotte, 2008
mixed forest	Mixed forest (deciduous and evergreen are each less than 75 percent of cover) in 500 m buffer (%)	LaMotte, 2008
shrubs	Shrublands in 500 m buffer (%)	LaMotte, 2008
grassland	Grasslands/herbaceous in 500 m buffer (%)	LaMotte, 2008
pasture	Pasture/hay in 500 m buffer (%)	LaMotte, 2008
crops	Cultivated crops in 500 m buffer (%) (Includes annual and perennial crops and actively tilled land)	LaMotte, 2008
woody wetland	Woody wetlands in 500 m buffer (%)	LaMotte, 2008
wetland	Emergent herbaceous wetlands (%)	LaMotte, 2008
atmospheric NO_3^-	Average atmospheric NO_3^- wet deposition in 500 m buffer (2005 to 2009) (kg/km2)	Bingham, 2011
calcium concentration	Calcium concentration in groundwater (mg/L)	NWIS & NAWQA
chloride concentration	Chloride concentration in groundwater (mg/L)	NWIS & NAWQA
sodium concentration	Sodium concentration in groundwater (mg/L)	NWIS & NAWQA
TDS	Total dissolved solids (TDS) concentration in groundwater (mg/L)	NWIS & NAWQA
atmospheric N	Average atmospheric nitrogen (N) deposition (1985 to 2001) (kg/yr)	Ruddy et al., 2006
irrigated cropland	Irrigated lands in 500 m buffer (%)	Pervez et al., 2010
population density	Average population density in 500 m buffer (1990 U.S. Census) (people/km2)	Nolan et al., 2006
Transport		
altitude	Altitude of well at land surface above mean sea level (m)	NWIS & NAWQA
top of screen	Depth of top of well screen below land surface (m)	NWIS & NAWQA
bottom of screen	Depth of bottom of well screen below land surface (m)	NWIS & NAWQA

Explanatory Variable	Description	Source
Transport (cont'd)		
water level	Depth to water below land surface measured at well during water quality sampling (m)	NWIS & NAWQA
recharge	Average groundwater recharge in 500 m buffer (mm/yr)	Wolock, 2003
drainage ditch	Surface drainage, field ditch conservation practice in 500 m buffer (%)	Nolan et al., 2006
soil slope	Average soil surface slope in 500 m buffer (%)	Nolan et al., 2006
precipitation	Average annual precipitation (mm)	PRISM, 2006
soil thickness	Average soil thickness in 500 m buffer (cm)	Wolock, 2003
soil group A	Hydrologic soil group A (low runoff potential and high infiltration rates) in 500 m buffer (%)	Wolock, 2003
soil group B	Hydrologic soil group B (moderate infiltration rates) in 500 m buffer (%)	Wolock, 2003
soil group C	Hydrologic soil group C (low infiltration rates) in 500 m buffer (%)	Wolock, 2003
soil group D	Hydrologic soil group D (high runoff potential and very low infiltration rates) in 500 m buffer (%)	Wolock, 2003
soil group AD	Hydrologic soil group AD in 500 m buffer (%) (drained soil: A; undrained soil: D)	Wolock, 2003
soil group BD	Hydrologic soil group BD in 500 m buffer (%) (drained soil: B; undrained soil: D)	Wolock, 2003
soil group CD	Hydrologic soil group CD in 500 m buffer (%) (drained soil: C; undrained soil: D)	Wolock, 2003
soil group AC	Hydrologic soil group AC in 500 m buffer (%) (drained soil: A; undrained soil: C)	Wolock, 2003
soil group BC	Hydrologic soil group BC in 500 m buffer (%) (drained soil: B; undrained)	Wolock, 2003
soil erodibility	Average soil erodibility factor in 500 m buffer (K)	Wolock, 2003
upper soil erodibility	Average soil erodibility factor of upper soil layer in 500 m buffer (K)	Wolock, 2003
soil permeability	Soil permeability (cm/hr)	Wolock, 2003
soil available water capacity	Soil available water capacity (cm/cm)	Wolock, 2003
soil bulk density	Average soil bulk density in 500 m buffer (g/cm3)	Wolock, 2003
soil shrink-swell	Average rating of soil shrink-swell potential in 500 m buffer	Wolock, 2003
soil<No. 4 sieve	Average soil material less than 3 inches in size that pass through a No. 4 sieve (5 mm) (% by weight)	Wolock, 2003
soil<No. 200 sieve	Average soil material less than 3 inches in size that pass through a No. 200 sieve (0.074 mm) (% by weight)	Wolock, 2003
soil<No. 10 sieve	Average soil material less than 3 inches in size that pass through a No. 10 sieve (2 mm) (% by weight)	Wolock, 2003
soil silt	Average silt texture in 500 m buffer (%)	Wolock, 2003
soil sand	Average sand texture in 500 m buffer (%)	Wolock, 2003
soil loss tolerance factor	Average soil loss tolerance factor value in 500 m buffer	Wolock, 2003
soil wind erodibility	Average soil wind erodibility factors in 500 m buffer	Wolock, 2003
fresh surface water withdrawal	Average fresh surface-water withdrawal in 500 m buffer (megaL/Day)	Nolan et al., 2006
water input	Average ratio of the total area of irrigated land to precipitation in 500 m buffer (km2/cm)	Nolan et al., 2006
irrigation tailwater recovery	Average tailwater recovery conservation practice in 500 m buffer (km^2)	Nolan et al., 2006

Explanatory Variable	Description	Source
Attenuation		
hydrologic landscapes	Average hydrologic landscape regions of the U.S. in 500 m buffer	Nolan et al., 2006
annual runoff	Average annual runoff in 500 m buffer (mm)	Gebert et al., 1987
seasonally high water table	Average depth below land surface to the seasonally high water table in 500 m buffer (m)	Wolock, 2003
well depth	Depth of well bottom below land surface (m)	NWIS & NAWQA
dissolved oxygen (DO)	Dissolved oxygen (mg/L)	NWIS & NAWQA
oxic redox	Oxic redox conditions ($O_2 \geq 0.5$ mg/L; $Mn^{2+} < 0.05$ mg/L; $Fe^{2+} < 0.1$ mg/L)	McMahon et al., 2008
anoxic redox	anoxic redox conditions ($O_2 < 0.5$ mg/L; $NO_3 \geq$or<0.5; $Mn^{2+} \geq$or<0.05 mg/L; $Fe^{2+} \geq$or<0.1 mg/L; $SO_4^{2-} \geq$or<0.5 mg/L)	McMahon et al., 2008
hydric soil	Occurrence of hydric soils in 500 m buffer (%)	Wolock, 2003
soil organic matter	Soil organic matter content (% by weight)	Wolock, 2003
tritium (TU)	Tritium concentration in groundwater (tritium units, TU)	NWIS & NAWQA
iron concentration	Iron concentration in groundwater (mg/L)	NWIS & NAWQA
Dunne overland flow	Dunne overland flow in 500 m buffer (% of streamflow)	Nolan et al., 2006
temperature	Groundwater temperature (degrees C)	NWIS & NAWQA
pH	Groundwater pH (standard units)	NWIS & NAWQA
manganese concentration	Manganese concentration in groundwater (mg/L)	NWIS & NAWQA
sulfate concentration	Sulfate concentration in groundwater (mg/L)	NWIS & NAWQA
histosol soil	Amount of histosol soils in 500 m buffer (%)	Nolan et al., 2006

Table 3: Results of the univariate logistic regression analysis. Logistic regression coefficients are outside parenthesis and p-values are enclosed in parenthesis. Bolded values are significant at the alpha level of 0.2 and were selected for initial inclusion in the multivariate logistic regression analysis.

Explanatory variable	north CCB	central CCB	south CCB	CCB (all)
Nitrogen Source				
farm fertilizer	**6.46E-04 (0.0153)**	6.33E-06 (0.8264)	8.28E-05 (0.3260)	**5.906E-5 (0.0734)**
open space	-0.0097 (0.8332)	**0.0910 (0.0770)**	-0.0263 (0.4813)	0.0106 (0.6307)
medium intensity development	-0.0367 (0.7052)	-0.0014 (0.9345)	**0.0237 (0.1418)**	**0.016 (0.1304)**
high intensity development	-60.3879 (0.9939)	-0.0149 (0.4651)	**1.0942 (0.0886)**	0.0048 (0.8053)
shrubs	-0.0241 (0.6671)	**-0.4863 (0.1906)**	-0.1013 (0.2030)	**-0.1069 (0.1269)**
grassland	-0.0032 (0.8494)	-0.0219 (0.6013)	-0.0291 (0.2724)	**-0.0213 (0.1068)**
crops	173.1187 (0.2797)	-0.0044 (0.8327)	0.0291 (0.2731)	**0.0225 (0.1940)**
wetland	-23.6409 (0.9928)	**-0.3832 (0.1867)**	-1.0162 (0.5365)	-0.6227 (0.2390)
atmospheric NO_3^-	**5.7945 (0.0152)**	0.5643 (0.7341)	**1.2246 (0.1364)**	0.3759 (0.2527)
calcium concentration	**0.0055 (0.1726)**	-0.0019 (0.2466)	-0.0018 (0.6162)	-0.0006 (0.5625)
Transport				
altitude	0.0021 (0.9067)	-0.0107 (0.4985)	**-0.0061 (0.0162)**	**-0.00456 (0.0156)**
recharge	**-0.0198 (0.0151)**	0.0031 (0.9109)	**-0.0394 (0.1106)**	**-0.0109 (0.0026)**
Hortonian overland flow	0.0961 (0.4005)	0.0157 (0.7923)	0.1050 (0.3011)	**0.1051 (0.1486)**
soil slope	-0.0055 (0.8627)	**0.0797 (0.1208)**	-0.0744 (0.2133)	-0.0013 (0.9405)
precipitation	**-0.0050 (0.0864)**	-0.0045 (0.3536)	**-0.0093 (0.0810)**	**-0.0028 (0.0025)**
soil thickness	**-0.0858 (0.0673)**	-0.0225 (0.6418)	-0.0398 (0.4343)	**-0.0518 (0.0544)**
soil group B	**-0.0412 (0.1162)**	-0.0021 (0.8708)	**0.0242 (0.1635)**	0.008 (0.3405)
soil group C	-0.0274 (0.4442)	0.0160 (0.6845)	**-0.0458 (0.0701)**	**-0.0406 (0.014)**
soil group D	**0.0588 (0.0244)**	-0.0106 (0.6326)	0.0240 (0.3179)	0.0122 (0.306)
soil erodibility	-7.8915 (0.4425)	-9.0313 (0.2983)	**40.8818 (0.0173)**	4.8914 (0.2609)
upper soil erodibility	**6.5450 (0.0684)**	-13.6142 (0.2585)	**28.3819 (0.0367)**	**7.5306 (0.1898)**

Explanatory variable	north CCB	central CCB	south CCB	CCB (all)
Transport (continued)				
soil permeability	-0.4540 (0.3571)	**0.5697 (0.1631)**	-0.0807 (0.5712)	-0.0351 (0.6711)
soil available water capacity	12.3590 (0.6166)	-17.4091 (0.2729)	**37.5578 (0.0666)**	8.5073 (0.2874)
soil bulk density	-3.1593 (0.4440)	**12.3901 (0.1296)**	**-10.8610 (0.0795)**	-0.3457 (0.8656)
soil<No. 4 sieve	-0.0133 (0.8474)	-0.1836 (0.2140)	**0.1027 (0.0957)**	**0.0535 (0.0588)**
soil<No. 200 sieve	0.0316 (0.5816)	**-0.1624 (0.0624)**	**0.1029 (0.0287)**	0.0151 (0.4008)
soil<No. 10 sieve	-0.0287 (0.6382)	**-0.1627 (0.1846)**	**0.1010 (0.0651)**	**0.044 (0.0765)**
soil silt	0.0454 (0.6812)	-0.1295 (0.2309)	**0.1081 (0.1719)**	- 0.0202(0.55 07)
soil sand	-0.1065 (0.2796)	**0.1489 (0.1107)**	**-0.0611 (0.1203)**	0.0006 (0.9733)
soil loss tolerance factor	0.0016 (0.9960)	**-1.2792 (0.0254)**	**-0.4165 (0.1556)**	**-0.5481 (0.0044)**
soil wind erodibility	**1.3468 (0.1058)**	-0.0708 (0.6575)	-0.2419 (0.2985)	**-0.2092 (0.0529)**
fresh surface water withdrawal	**786.3675 (0.0160)**	16.6250 (0.7463)	102.8617 (0.3604)	**113.8096 (0.0590)**
annual runoff	**-0.1942 (0.0108)**	0.0200 (0.9286)	**-0.3340 (0.1138)**	**-0.111 (0.0021)**
Attenuation				
seasonally high water table	0.3020 (0.7132)	0.1530 (0.8283)	**-1.1004 (0.0745)**	-0.4468 (0.2259)
dissolved oxygen (DO)	**0.4685 (0.0126)**	**9.4403 (0.1472)**	**0.4398 (0.0151)**	**0.4738 (<.0001)**
oxic redox	16.3762 (0.9936)	**1.7346 (0.1571)**	16.2029 (0.9922)	17.2296(0.9 902)
tritium (TU)	-0.4436 (0.0498)	**0.1129 (0.1768)**	-0.1735 (0.3692)	0.0296 (0.5920)
iron concentration	-0.0017 (0.2304)	**-0.0115 (0.1380)**	-0.0734 (0.2602)	**-0.0036 (0.0748)**
temperature	**0.2061 (0.1195)**	-0.0043 (0.9819)	0.0340 (0.7334)	**0.1235 (0.0654)**
pH	0.0003 (0.9998)	**-1.3606 (0.1797)**	1.2285 (0.2757)	0.2853 (0.5414)
manganese concentration	**-0.1205 (0.1153)**	-0.0187 (0.2448)	**-0.0343 (0.0567)**	**-0.0248 (0.0062)**

Table 4. Model calibration and goodness-of-fit for the multivariate logistic regression models of NO_3^- greater than or equal to the relative background concentration in recently recharged groundwater.

Model	NO_3^- threshold (mg/L)	Model Intercept	LLR p-value	HL p-value	ROC
north CCB	2	-0.799	<0.001	0.944	0.98
central CCB (model a)	2	0.009	0.013	0.295	0.82
central CCB (model b)	2	-8.762	<0.001	0.993	0.98
south CCB	2	-3.857	<0.001	0.375	0.94
CCB (north, central, and south)	2	-1.492	<0.001	0.759	0.86

[LLR, log-likelihood ratio; HL, Hosmer Lemeshow goodness-of-fit; ROC, Receiver Operating Characteristic curve; * , presented in (*27*)]

Table 5. Parameters for logistic regression models for NO3– above relative background concentration (2 mg/L) in recently recharged groundwater of sub-regions of the California Coastal Basin (CCB) aquifer system.

		Explanatory variables (S, source; T, transport; A, attenuation)	Explanatory variable coefficient (p-value)	Standardized Coefficient	Standard Error	Wald confidence Interval lower 95%	upper 95%
north CCB (2 mg/L)		farm fertilizer (kg/km2 of cropland): S	0.00085 (0.008)	1.313	0.00032	0.00022	0.00148
		dissolved oxygen (mg/L): A	0.719 (0.002)	1.129	0.237	0.256	1.1838
		soil thickness (cm): T	-0.126 (0.025)	-0.625	0.056	-0.236	-0.016
central CCB (2 mg/L)	model a	open space (%): S	0.190 (0.029)	1.137	0.087	0.019	0.361
		soil group D (%): T	-0.075 (0.058)	-0.731	0.039	-0.152	0.002
	model b	dissolved oxygen (mg/L): A	9.440 (0.007)	14.001	3.506	2.569	16.311
south CCB (2 mg/L)		low intensity development (%): S	1.503 (0.071)	7.618	0.832	-0.128	3.133
		crops (%): S	0.071 (0.093)	0.830	0.043	-0.012	0.155
		dissolved oxygen (mg/L): A	0.371 (0.073)	0.830	0.207	-0.034	0.777
CCB (north, central, and south) (2 mg/L)		dissolved oxygen (mg/L): A	0.479 (<0.001)	0.751	0.117	0.267	0.732
		soil thickness (cm): T	-0.059 (0.096)	-0.278	0.036	-0.134	0.009
		soil available water capacity (cm/cm): T	18.389 (0.122)	0.271	11.902	-3.603	43.872
		farm fertilizer (kg/km2 of cropland): S	0.000054 (0.110)	0.269	0.00003	-0.00001	0.00013

Table 6. Model validation and goodness-of-fit for the multivariate logistic regression models of nitrate greater than or equal to the relative background concentration in recently recharged groundwater.

Model	NO_3^- threshold (mg/L)	Model Intercept	LLR p-value	HL p-value	ROC
north CCB	2	-0.128	0.912	0.532	0.60
central CCB (model a)	2	0.165	0.899	0.267	0.56
central CCB (model b)	2	-22.221	<0.001	0.997	0.99
south CCB	2	-5.834	<0.001	0.926	0.97
CCB (north, central, and south)	2	-0.417	0.295	0.516	0.70

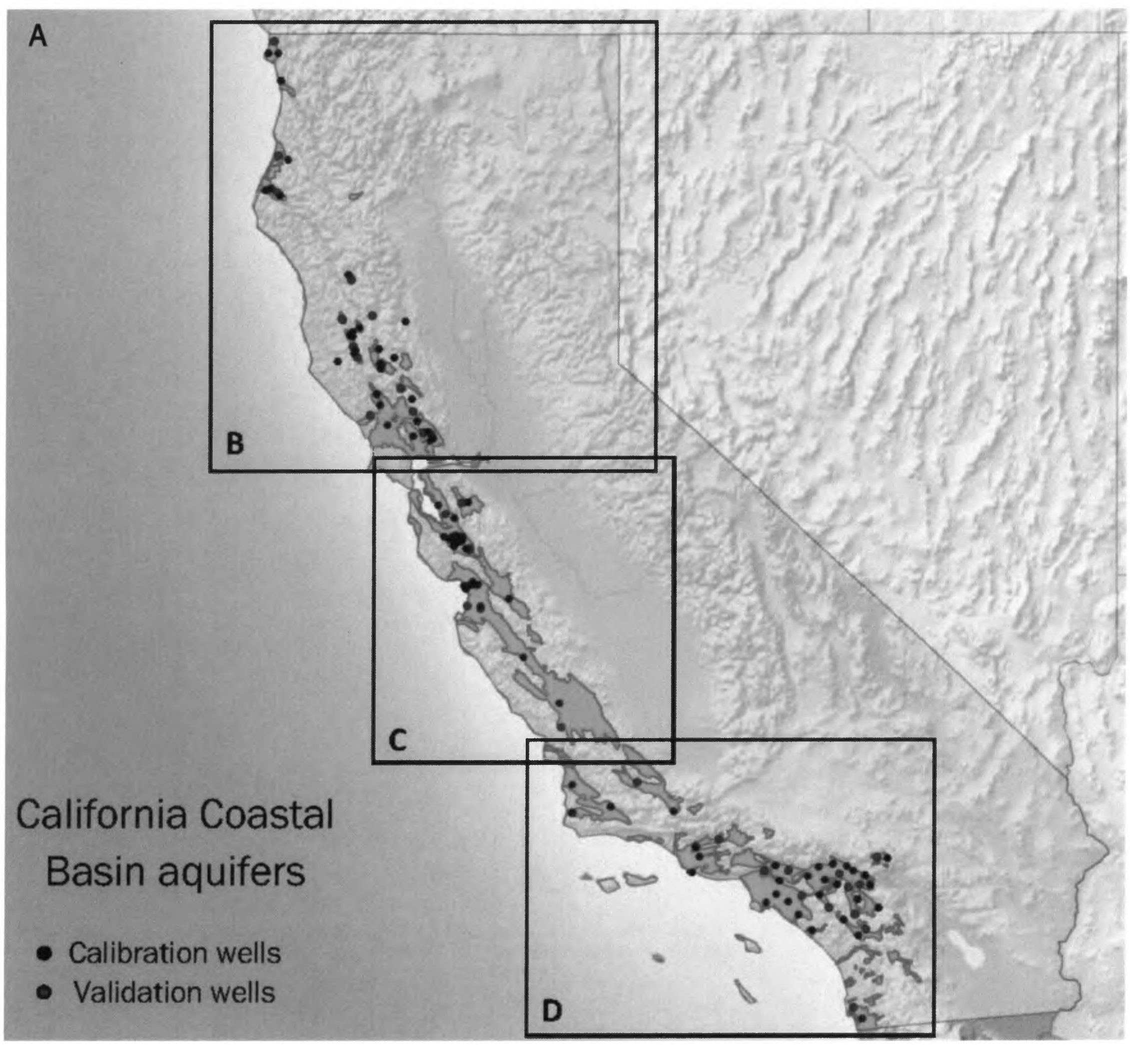
[LLR, log-likelihood ratio; HL, Hosmer Lemeshow goodness-of-fit; ROC, Receiver Operating Characteristic curve; * , presented in (*27*)]

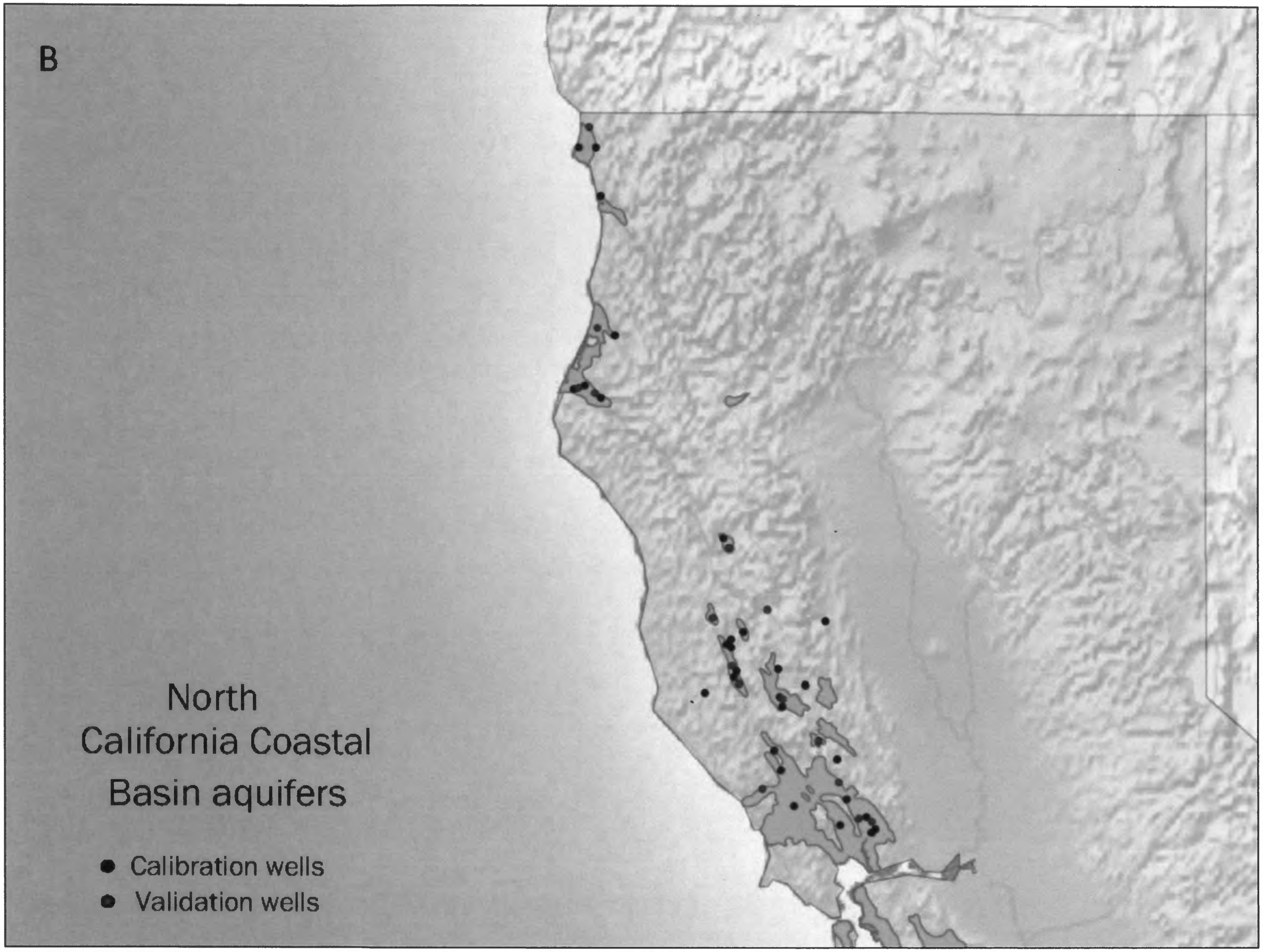
Table 7. Validation of parameters for logistic regression models for nitrate greater than the background concentration (2 mg/L) in recently recharged groundwater of sub-regions of the California Coastal Basin (CCB) aquifer system.

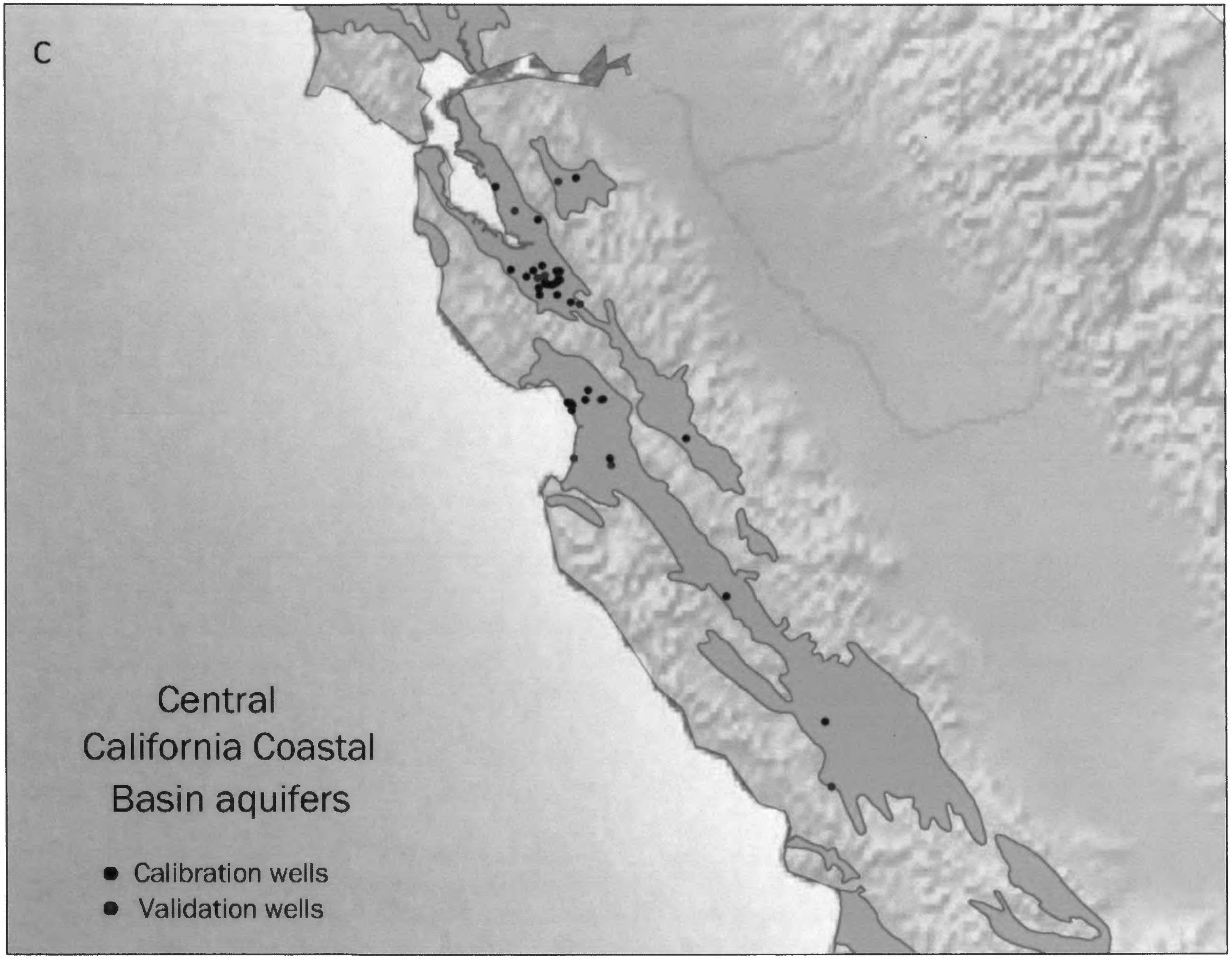
		Explanatory variables (S, source; T, transport; A, attenuation)	Explanatory variable coefficient (p-value)	Standardized Coefficient	Standard Error	Wald confidence Interval	
						lower 95%	upper 95%
north CCB (2 mg/L)		dissolved oxygen (mg/L): A	-0.202 (0.609)	-0.363	0.396	-0.978	0.573
		soil thickness (cm): T	-0.027 (0.722)	-0.158	0.077	-0.178	0.123
		farm fertilizer (kg/km2 of cropland): S	0.00005 (0.8535)	0.078	0.0003	-0.0005	0.0006
central CCB (2 mg/L)	model a	open space (%): S	-0.049 (0.683)	-0.162	0.121	-0.287	0.188
		soil group D (%): T	0.0085 (0.792)	0.105	0.032	-0.055	0.072
	model b	dissolved oxygen (mg/L): A	11.443 (<0.001)	15.265	0.754	9.965	12.993
south CCB (2 mg/L)		low intensity development (%): S	0.872 (0.509)	5.192	1.321	-1.717	3.462
		crops (%): S	0.505 (0.174)	1.921	0.372	-0.224	1.235
		dissolved oxygen (mg/L): A	0.73 (0.107)	1.418	0.453	-0.158	1.618
CCB (north, central, and south) (2 mg/L)		dissolved oxygen (mg/L): A	0.216 (0.098)	0.371	0.131	-0.039	0.471
		soil thickness (cm): T	-0.059 (0.182)	-0.311	0.044	-0.146	0.028
		soil available water capacity (cm/cm): T	16.088 (0.358)	0.212	17.501	-18.213	50.389
		farm fertilizer (kg/km2 of cropland): S	0.00003 (0.441)	0.175	0.00004	-0.00005	0.00011

FIGURES

Figure 1: (A) Map of California Coastal Basin (CCB) aquifers with location of wells used for calibration (black) and validation (red) analysis. Boxed areas are sub-regions used in study (B: north CCB; C: central CCB; and D: south CCB). There are 30 validation wells and 15 validation wells in each sub-region with a total of 90 validation wells and 45 calibration wells in the entire CCB.




B
North
California Coastal
Basin aquifers
Calibration wells
Validation wells


C
Central
California Coastal
Basin aquifers
Calibration wells
Validation wells

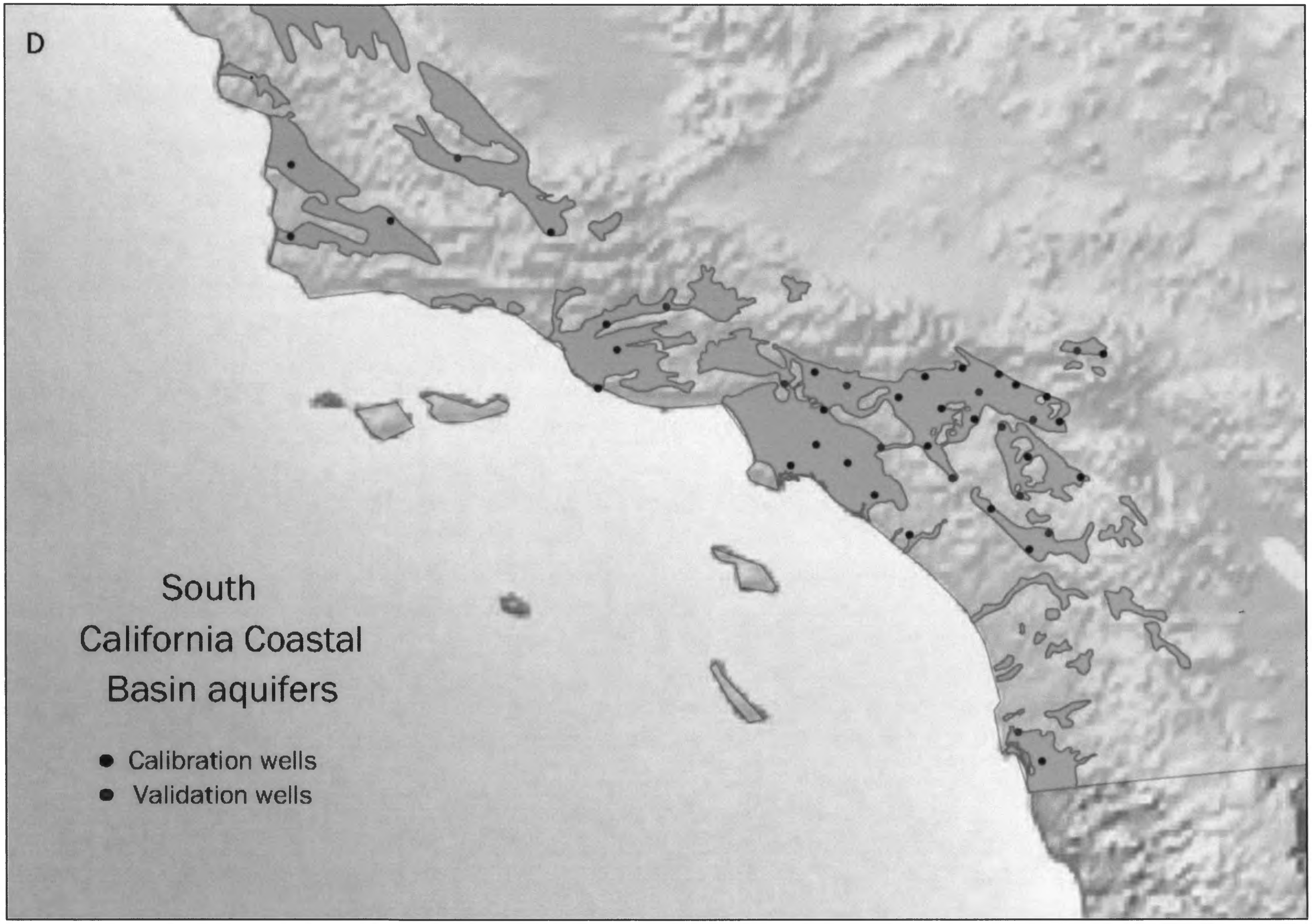

D
South
California Coastal
Basin aquifers
Calibration wells
Validation wells

Figure 2: Map of calibration and validation wells used in Gurdak and Qi (2012) groundwater vulnerability study.

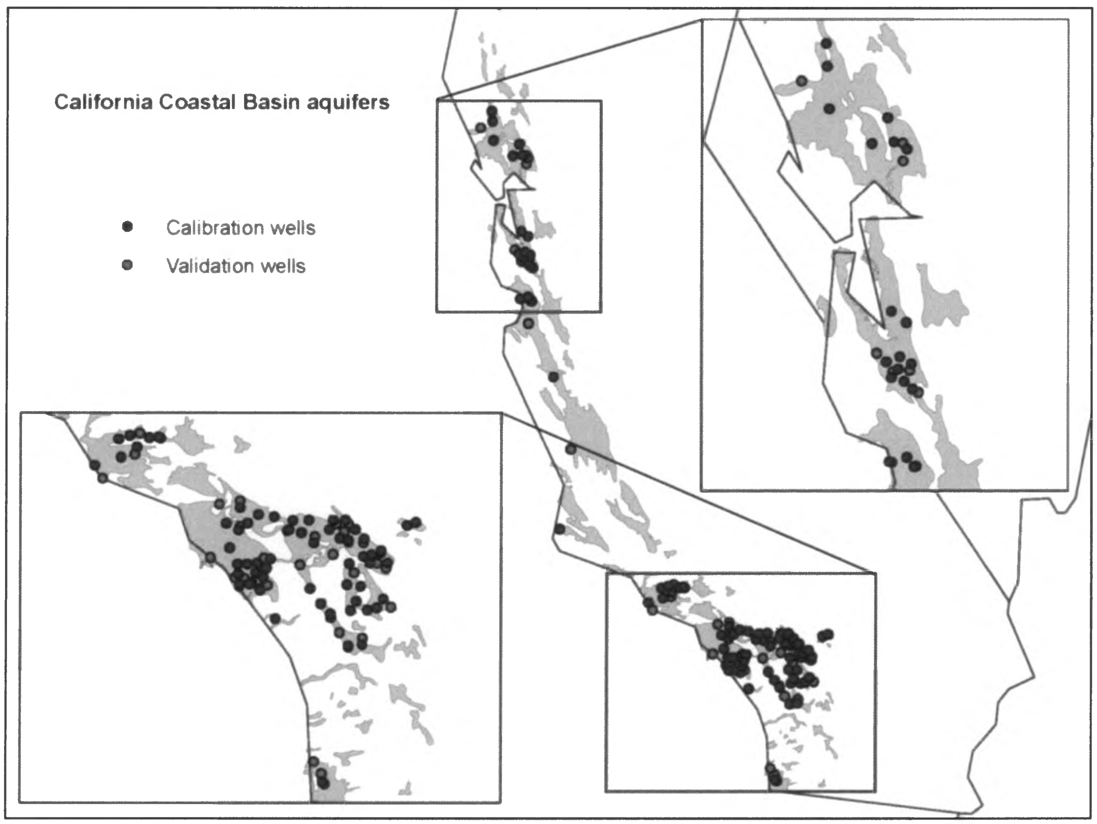


Figure 3: Nitrate concentrations of well sites in the CCB. N=45 in the North, Central, and South sub-regions. N=135 in the All subregion.

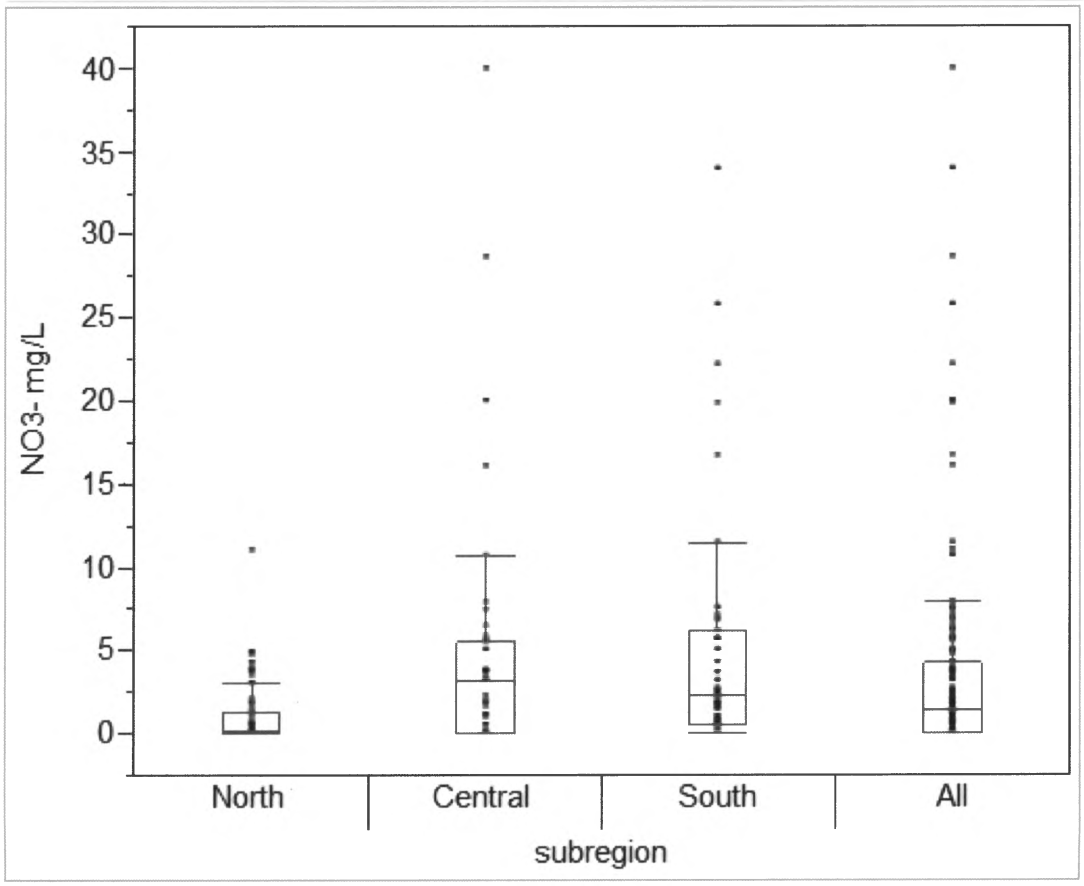


Figure 4: Nitrate concentrations of calibration well sites of the CCB. N= 30 in the North, Central, and South sub-regions. N=90 in the All sub-region.

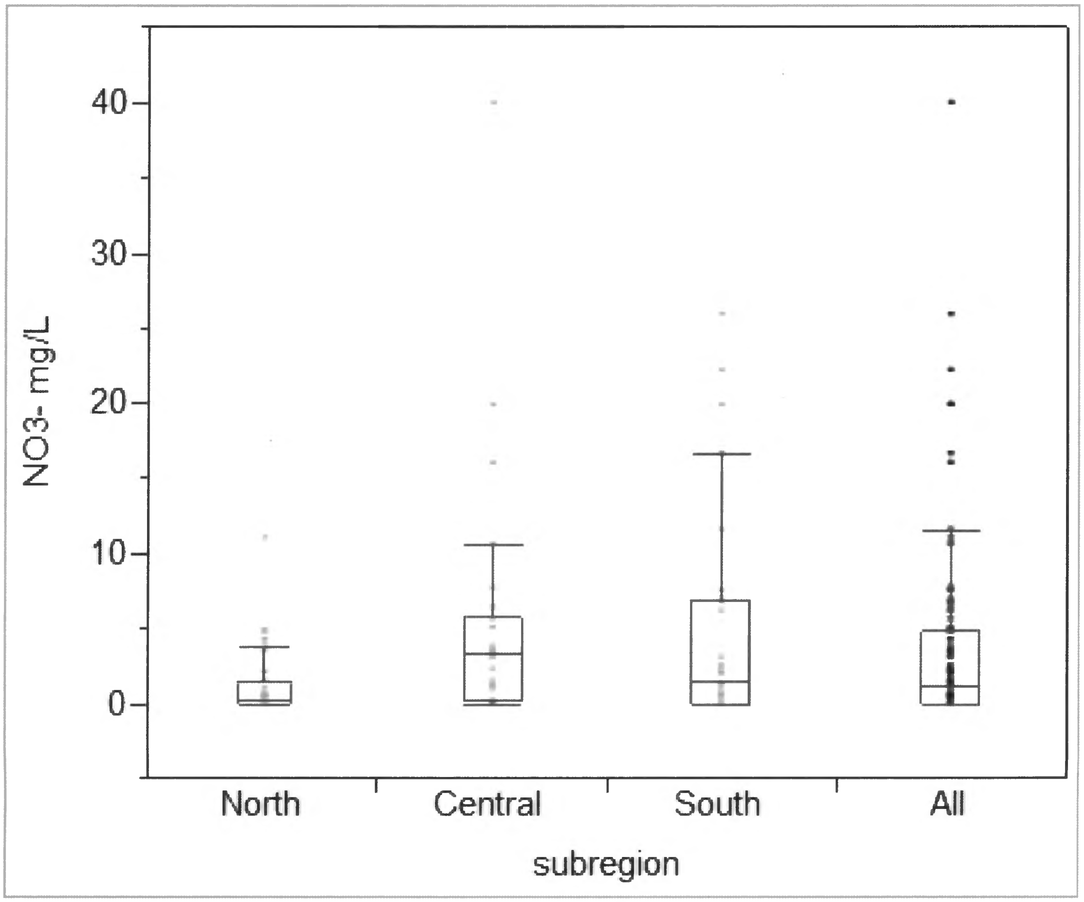


Figure 5: Nitrate concentrations of validaiton well sites in the CCB. N=15 in the North, Central, and South sub-regions. N=45 in the All sub-region.

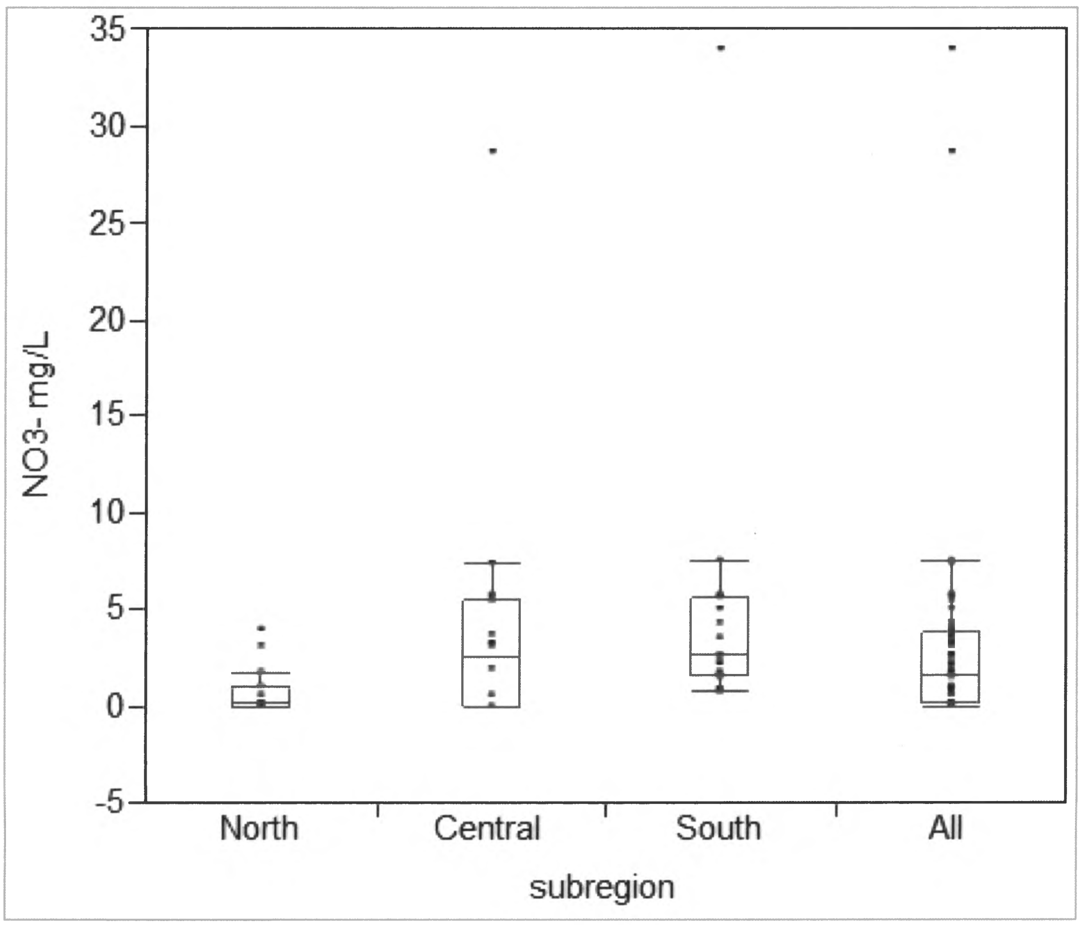


Figure 6: Map of all well sites in the CCB and associated ranges of nitrate concnetrations.

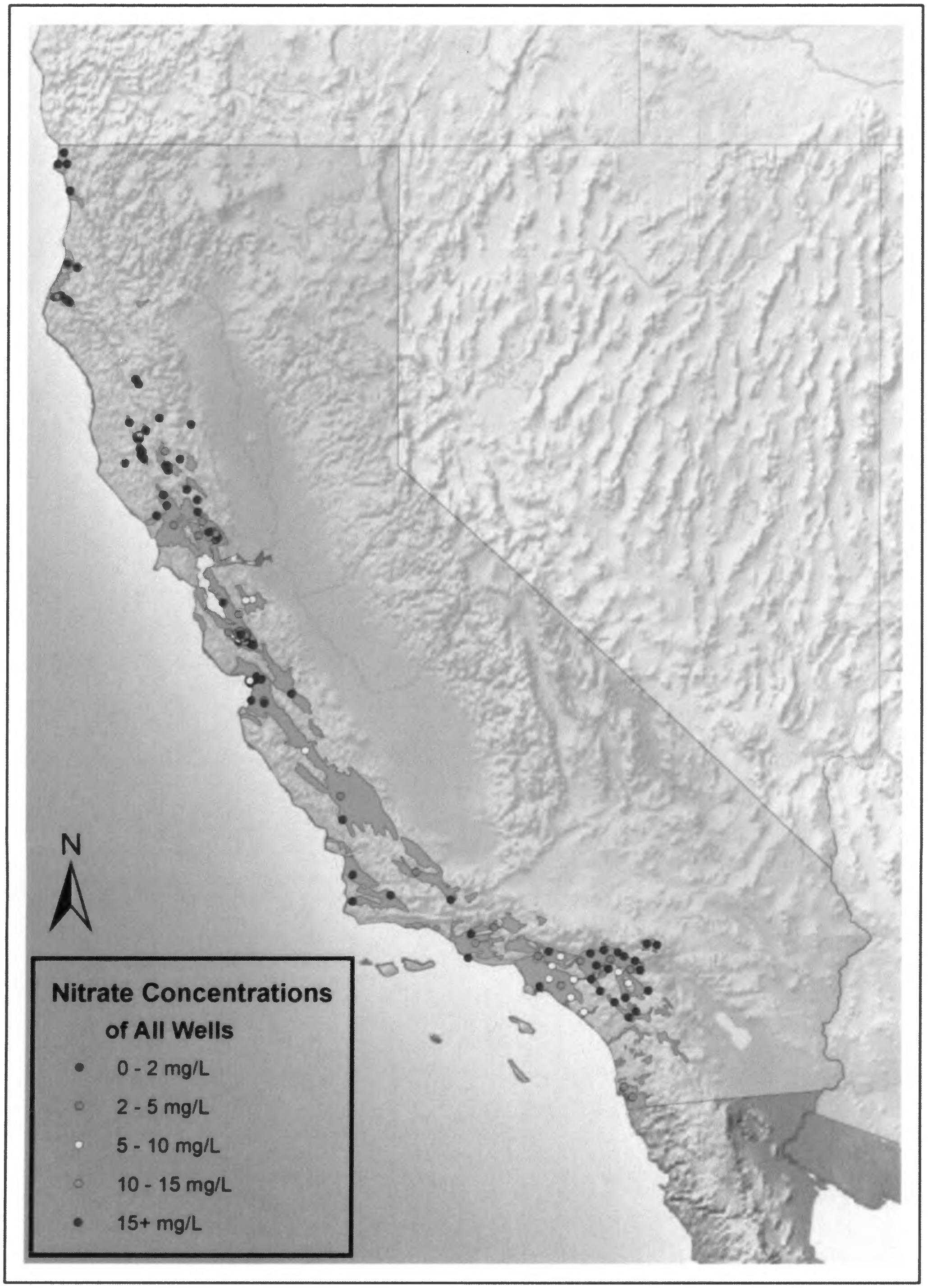


Figure 7: Map of calibration well sites in the CCB and associated ranges of nitrate concnetrations.

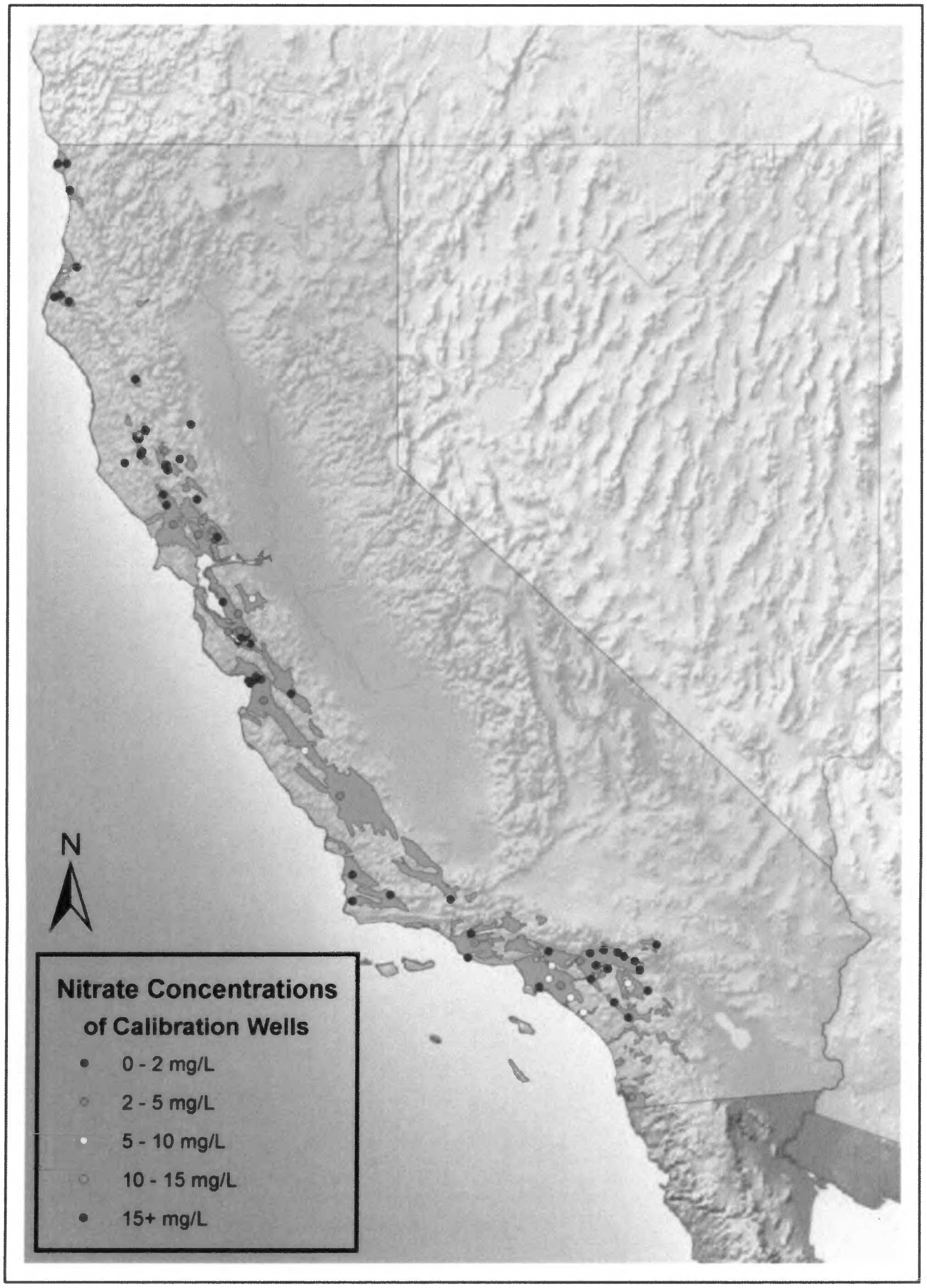


Figure 8: Map of validation well sites in the CCB and associated ranges of nitrate concentrations.

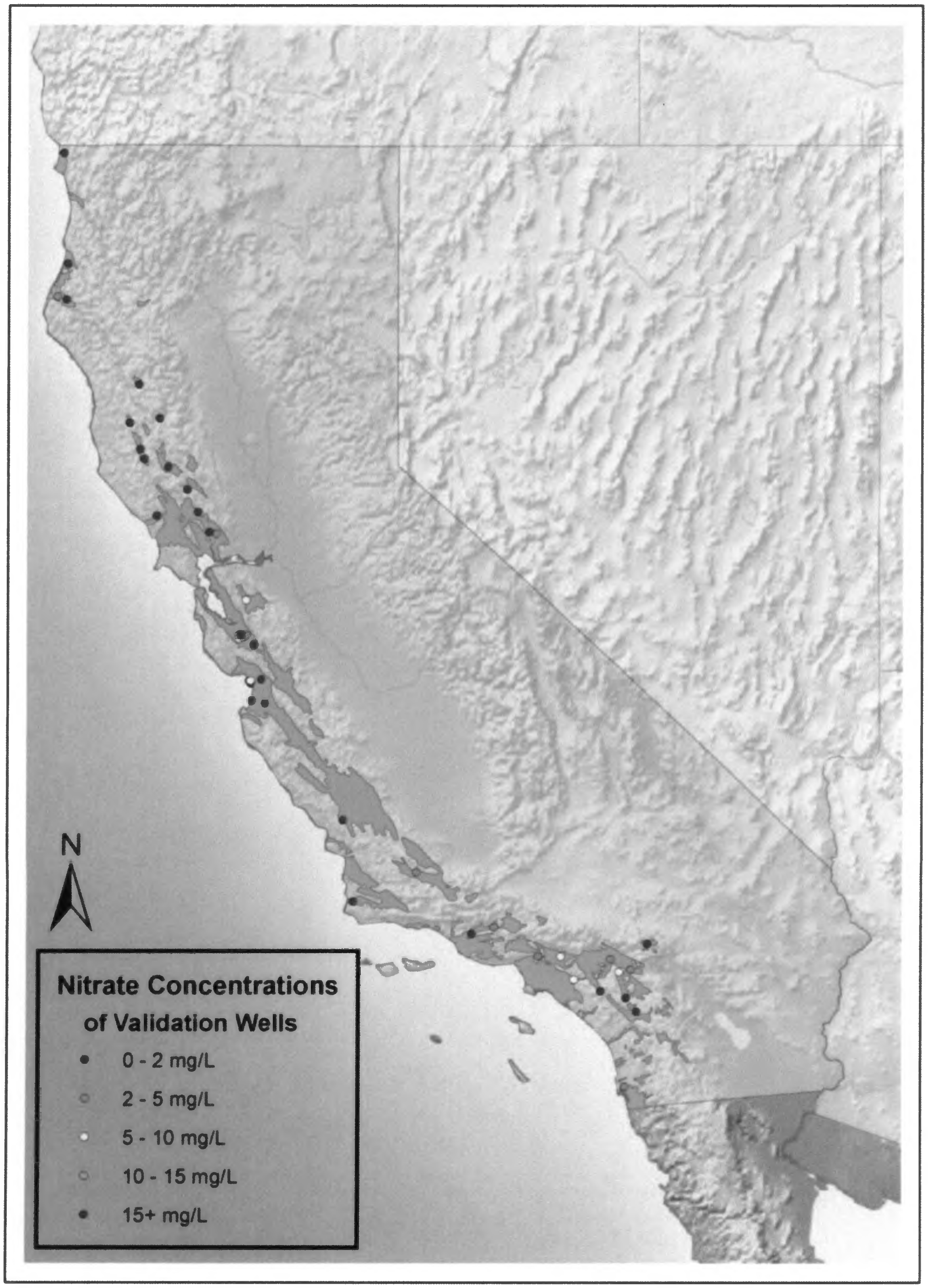


Figure 9: Sales of nitrogen fertilizer in California, 1945-2008 (Rosenstock, et al., 2013).