# California Regional Water Quality Control Board
## North Coast Region
### Geoffrey M. Hales, Chairman

www.waterboards.ca.gov/northcoast
5550 Skylane Boulevard, Suite A, Santa Rosa, California 95403
Phone: (877) 721-9203 (toll free) • Office: (707) 576-2220 • FAX: (707) 523-0135

**Linda S. Adams**
*Secretary for Environmental Protection*

**Arnold Schwarzenegger**
*Governor*


## MEMORANDUM

Date: 19 November 2010

To: File: Laguna de Santa Rosa; TMDL Development & Planning

From: Steve Butkus

Subject: Gradient Analysis of Environmental Variables to Delineate Pre-European Settlement Land Cover Boundaries

---

A gradient analysis was conducted to distinguish between rangeland areas and forested areas for the Laguna map representing pre-settlement land cover. The assumption was made that those environmental conditions with the most influence on the distribution of today's remaining rangeland and oak forests could be used to estimate pre-settlement native vegetation.

A *gradient* is a change in a value of a variable per unit distance in a specified direction. The term *gradient analysis* describes any method attempting to relate the response variable (i.e., vegetation) to the environmental gradients (i.e., soil characteristics). Environmental variables often exhibit gradients that influence vegetation patterns. Ordination is a statistical approach to gradient analysis. Ordination methods consider samples to be points in a multidimensional space. The environmental variables define the axes and position of each sample. All gradient analyses were completed using the ordination software CANOCO (Ter Braak and Šmilauer, 1998).

Areas in the Laguna watershed that are currently covered by rangeland or forest were analyzed for influence from environmental gradients. The distribution of oaks in Sonoma County has been shown to be most influenced by soil type and topology (Dawson, 2008). The following spatial data were compiled for areas currently classified as rangeland or forest (USGS, 2006) from the national soil database (NRCS, 2007): elevation, slope, hydrogroup, percent clay, percent silt, and percent sand in surface soils.

Two types of the model of the species response to an environmental gradient are typically observed: linear and unimodal. Unimodal responses show a species optimum along an environmental gradient. Linear responses show an inclination of species over a shorter range of the environmental gradient. Different ordination methods are needed depending on the response model type. Redundancy analysis (RDA) is most appropriate for linear responses and canonical correspondence analysis is most appropriate for unimodal responses. One must determine which response model is most appropriate for the available range of environmental data. Indirect gradient analysis can advise the selection of an appropriate model response type.

Indirect gradient analysis was conducted using unconstrained ordination methods. The analysis is considered indirect because it assesses all the variability of the data sets. The unconstrained ordination axes were aligned with distances representing the greatest variability within the data set. The unconstrained ordination uses the detrended correspondence analysis (DCA) method. DCA has features that determine the type of ordination model used; either based on linear response or unimodal response to underlying environmental gradients. Gradient length derived from Hill's scaling of ordination scores were used to determine model response type. The lengths of the gradients of the first and second axes were both below the 4.0 scaling threshold for gradient length. This result indicated that the linear response model using RDA is the most appropriate for the range of available data within the Laguna watershed.

A direct gradient analysis was conducted using the RDA constrained ordination method. The analysis axes were constrained to be linear combinations of the environmental variables. Direct gradient analyses were used to reduce the large variability observed in response variable data sets with many non-detectable results. In constrained ordination analysis, a search was conducted for variables that best explained the vegetation. In the constrained ordinations the ordination axes were weighted sums of environmental variables. The constrained ordination axes corresponded with the direction of greatest variability in a data set that could be explained by environmental variables. The direct gradient analysis reduced variability by removing variability shown as residuals to the regression.

The set of environmental variables used to develop the response model may contain variables that are collinear or do not contribute significantly to the explained variance. Variables exhibiting multicollinearity provide no unique contribution to the regression equation and result in unstable ordination regression coefficients (Ter Braak and Looman, 1986). Collinear environmental variables were identified using the Variance Inflation Factor (VIF). The VIF of a variable is proportional to the variance of the estimated regression coefficients. High VIF’s indicate multicollinearity among the environmental variables. If the VIF is greater than 20, then the variable is almost perfectly correlated with the other variables. The VIF values from the RDA showed that all environmental variables were below three indicating little to no multicollinearity among variables.

The RDA analysis showed that the first three axes explained 76% of the vegetation variation. Some of these variables have a greater influence on the response model than others. Forward selection RDA using Monte Carlo permutations was conducted to determine which environmental variables significantly contribute to explaining the response variables. In the ordinary statistical test, the value of the statistics calculated from the data is compared with the expected distribution of the statistics tested. Based on this comparison the probability of obtaining results as different from the null hypothesis is estimated. The distribution of the test statistics is derived from the assumption about the distribution of the original data. Since the distribution of the test statistics under the null hypothesis of independence is not known, the distribution is simulated with a Monte Carlo permutation test. The significance of the null hypothesis is tested using a Monte Carlo unconstrained permutation test. This tests the effect of environmental variables on vegetation pattern. The rejection of the null hypothesis means that at least one of the variables has some effect on the vegetation pattern. Forward selection RDA using 1000 unrestricted Monte Carlo permutations was conducted with each environmental variable.

*Gradient Analysis Results*

Only four (4) of the environmental variables showed a significant effect on the vegetation type: elevation, slope, soil hydrogroup, and soil percent silt. The two environmental variables that did not show any significant effect on vegetation patterns (i.e., soil percent clay and percent sand) were excluded from the model.

The biplot diagram of the RDA ordination of rangelands and oak forest lands is presented in Figure 1. Results of ordination are usually displayed as biplot diagrams. In linear methods, response variables (i.e., vegetation type) are shown as arrows indicating the relative importance and direction variables. Each arrow points in the direction of maximum variation in value of the corresponding variable, and each may be extended in both directions from the origin of the plot. The origin represents the mean of each variable. The length of arrows represents the degree to which each is correlated with the ordination axes. The important gradient variables have longer arrows than less important ones in terms of predicting the vegetation pattern.

In Figure 1, the land cover categories (i.e., range and forest) are shown as solid arrows and environmental gradient variables (i.e., elevation, slope, soil hydrogroup, and percent silt) are shown as dashed arrows. The diagram indicates that each of the gradient variables have relatively the same importance with similar arrow lengths. The short arrows for the land cover classes depict the small level of explained variance in the first two ordination axes. Both rangeland areas and forests are positively correlated with the elevation and percent of silt in the surface soils, and negatively correlated with slope and wetter soil hydrogroups.

The four significant variables identified as influencing the vegetation were used to delineate forest areas from rangelands on the Laguna pre-settlement spatial model. The discussion of the model's sensitivity analysis is found in the "Pre-European Settlement Spatial Data Model Evaluation" memo to the file (Butkus, 2010).

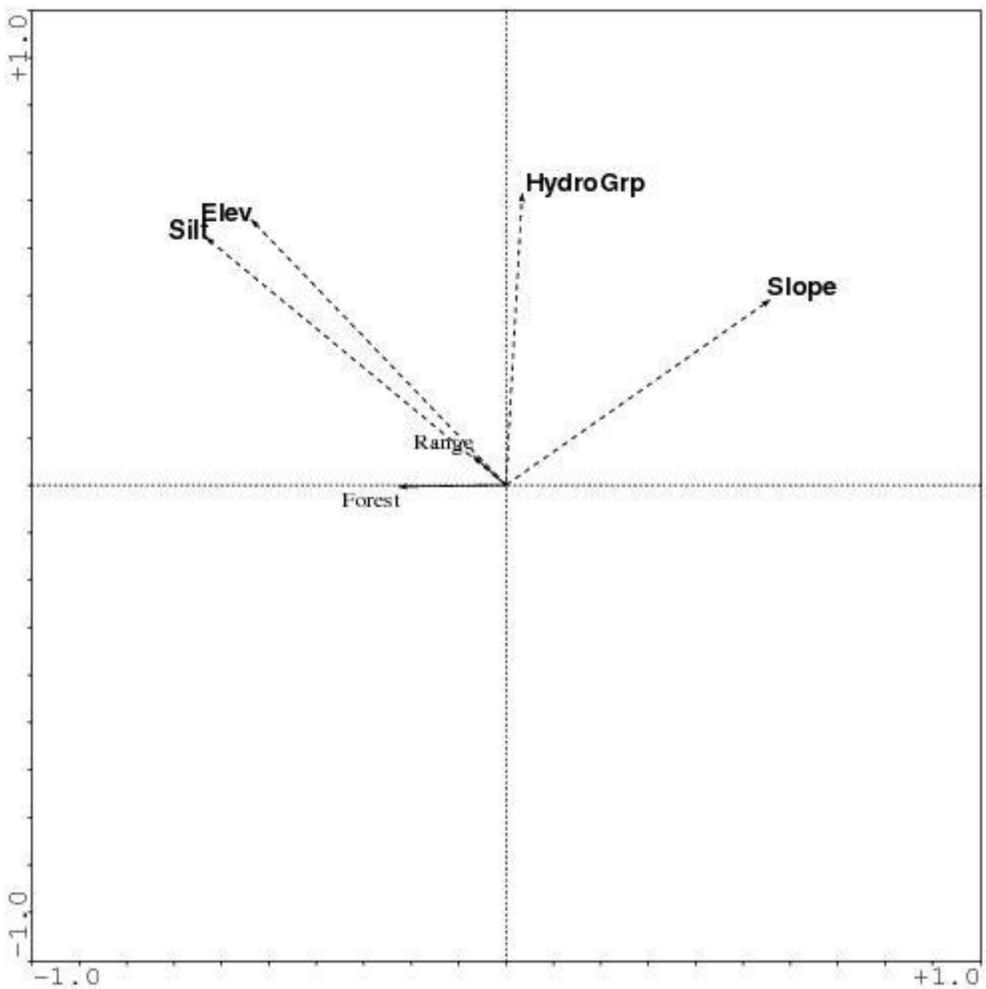


Figure 1. Redundancy Analysis Diagram of Environmental Gradient Variables

## CITATIONS

Butkus, S. 2010. Pre-European Settlement Spatial Data Model Evaluation. Memorandum to the Laguna TMDL File dated November 19, 2010. North Coast Regional Water Quality Control Board, Santa Rosa, CA.

Dawson, A. 2008. Oaks through Time: Reconstructing Historical Change in Oak Landscapes. General Technical Report PSW-GTR-217. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station Albany, CA

NRCS. 2007. Soil Survey Geographic (SSURGO) Database for Sonoma County. Natural Resources Conservation Service U.S. Department of Agriculture. Available on-line “http://SoilDataMart.nrcs.usda.gov/”.

Ter Braak C.J.F. and Looman C.W.N. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio*, 65: 3 - 11

Ter Braak C.J.F. and Šmilauer P. 1998. CANOCO Reference Manual and User's Guide to Canoco for Windows. Microcomputer Power, Ithaca, USA. 352 pp

USGS, 2006. Enhanced Historical Land-Use and Land-Cover Data Sets of the U.S. Geological Survey. Data Series 240. U.S. Geological Survey. http://pubs.usgs.gov/ds/2006/240/.