USING QUANTILE REGRESSION TO MODEL ELEVATION/SWE RELATIONSHIPS IN THE OLYMPICS AND CENTRAL WASHINGTON CASCADES

Shea McDonald[1] and Dwight Barry[1,2]

ABSTRACT

Although ordinary least squares (OLS) regression is often used to develop initial models for elevation/snow water equivalent (SWE) relationships in watersheds, such an approach cannot account for the heterogeneous variability in SWE over elevation bands common in many coastal Pacific Northwest watersheds. This is particularly problematic in attempting to model middle-elevation SWE in watersheds affected by winter rain-on-snow events. Scatterplots of this relationship often show a distinctive fan-shaped pattern that widens as elevation increases, implying that there is not a single rate of change in SWE by elevation. Thus, the use of an OLS regression in models or trend assessment could provide severely misleading estimates of SWE in middle elevation areas where no field data are collected. We compare OLS and quantile regression-based models of SWE by elevation based on four snow seasons of field assessment for the Dungeness Watershed of western Washington State, as well as for snow courses and SNOTEL sites in the Olympic Mountains and central Washington Cascades. In both cases, quantile regression better accounts for the observed data than does the typical OLS approach. (KEYWORDS: snow water equivalent, quartile regression, watershed elevation, least squares regression, Dungeness Watershed)

INTRODUCTION

Although ordinary least squares regression is often used to develop initial models for elevation/snow water equivalent (SWE) relationships, such an approach cannot account for the heterogeneous variability in SWE over elevation bands common in many Pacific Northwest watersheds. This is particularly problematic in attempting to model middle-elevation (i.e., ~400-1600 m) SWE in areas affected by rain-on-snow events.

Scatterplots of elevation/SWE relationships often show increasing variability with elevation, implying that there is not a single rate of change in these probability distributions. In these cases, the use of an ordinary least squares (OLS) regression would provide misleading (and statistically inappropriate) estimates of SWE. Quantile regression provides another tool for initial exploratory analysis and model development (Cade and Noon, 2003). Rather than simply providing the mean response, any quantile of the dataset can be analyzed. Because quantile regression estimates multiple rates of change simultaneously, it can provide a more complete picture of relationships between variables than other regression methods. This can be particularly helpful for predicting rates of change at the extremes, under different climate conditions, when variance is heterogeneous, and/or when transformations are not statistically effective.

This paper presents a comparison of quantile and OLS regression for elevation/SWE relationships in the Dungeness Watershed of western Washington as well as across the Olympic Mountains and the central Washington Cascades. These are only a couple of example analyses used for illustration purposes. The usefulness of quantile regression as a tool for describing and modeling multiple rates of change for comparison with environmental variables (especially elevation) is only limited by the scope of the dataset.

METHODS

To better understand SWE variability in the mid-elevations and ground-truth a recently-implemented hydrological model supported by NASA's Solutions Network initiative (see pcnasa.ctc.edu for details), we have been monitoring 12 sites using standard NRCS snow course methods in the Dungeness Watershed since December 2007 (Figures 1 and 3). For this analysis (Figure 4), we symbolized SWE points by aspect of the snow course to see if any patterns emerged alongside the differences in the quantiles as compared to the mean.

In our second analysis (Figures 2-3 and 5-6), we used long-term SWE data from NRCS snow courses and SNOTEL sites in the Olympic Mountains along with neighboring NRCS snow courses in the central Washington Cascades. In Figure 5, SWE is symbolized according to ENSO and PDO climate phases to see if their influences on SWE distributions were better explained by the quantiles. Quantile regression was performed using R's *quantreg* package (Koenker 2010).

Poster paper presented Western Snow Conference 2011

[1] Peninsula College, 1502 E. Lauridsen Blvd., Port Angeles, WA, 98362, dbarry@pencol.edu

[2] Western Washington University-Huxley College on the Peninsulas, Port Angeles, WA, 98362

Figure 1. The Dungeness Watershed and NASA project snow courses, viewed looking north.

Figure 2. Locations of central Casacades snow courses and Olympic Mountains SNOTELs.

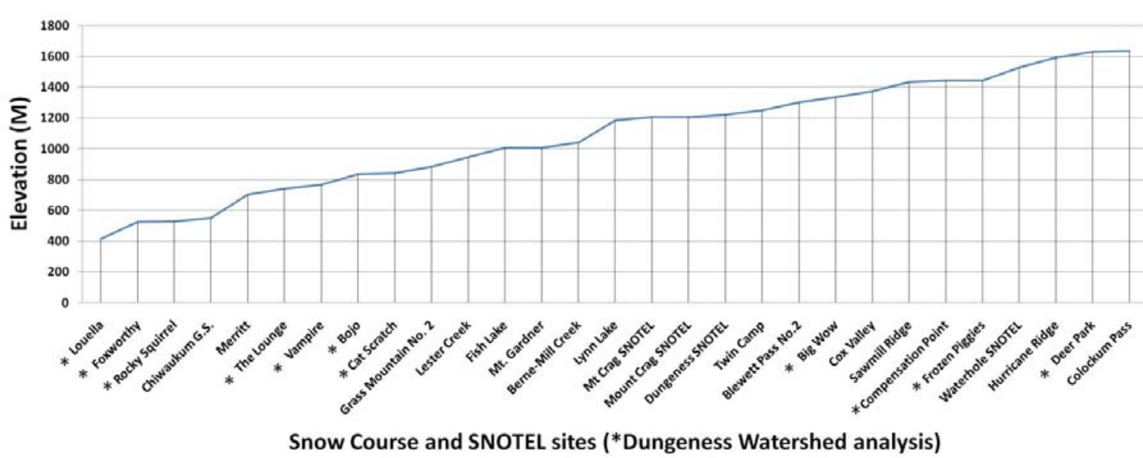

Figure 3. Elevational profile of snow courses and SNOTELs included in these analyses.

RESULTS

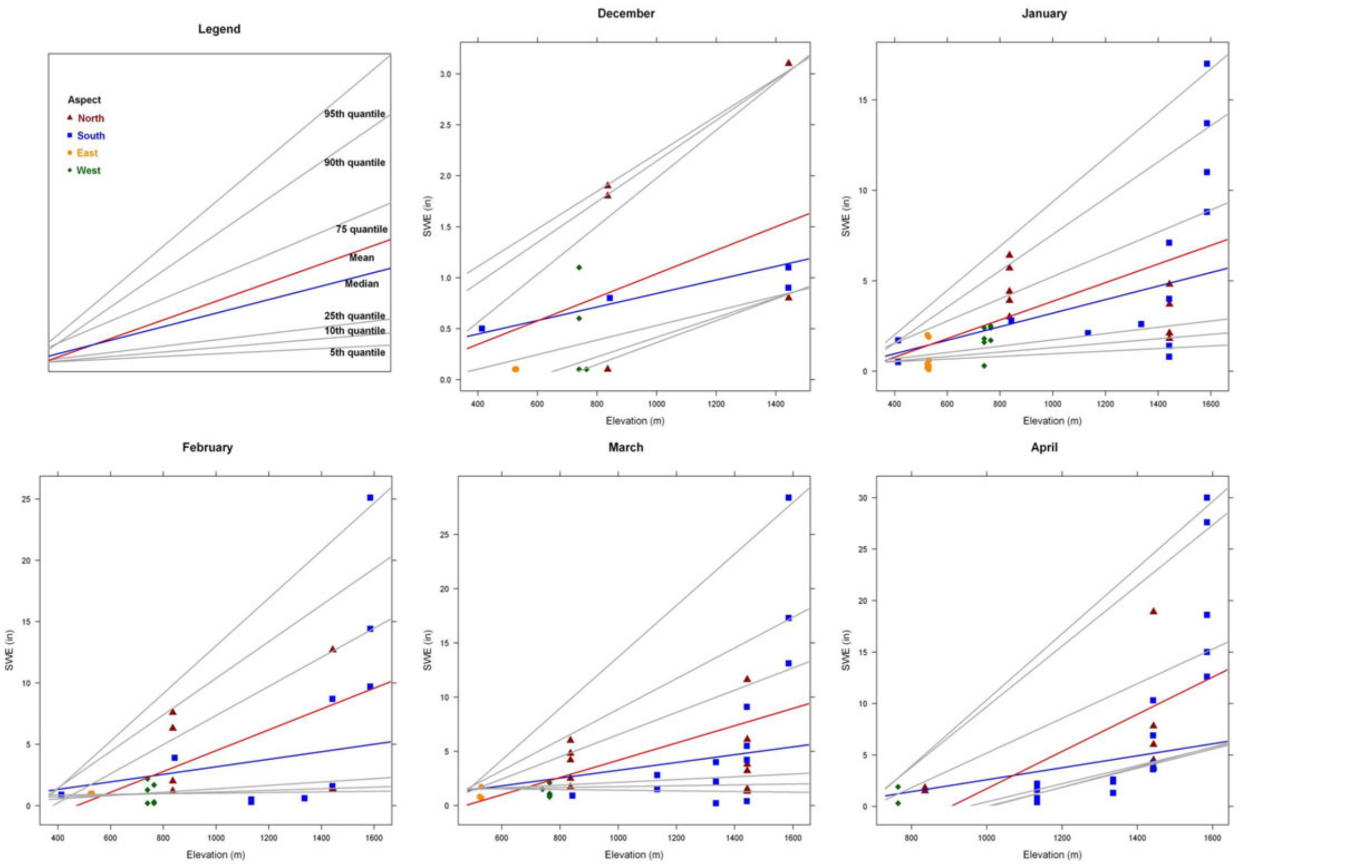

Figure 4. Elevation/SWE relationships modeled by OLS and quantile regression by month in the Dungeness Watershed.

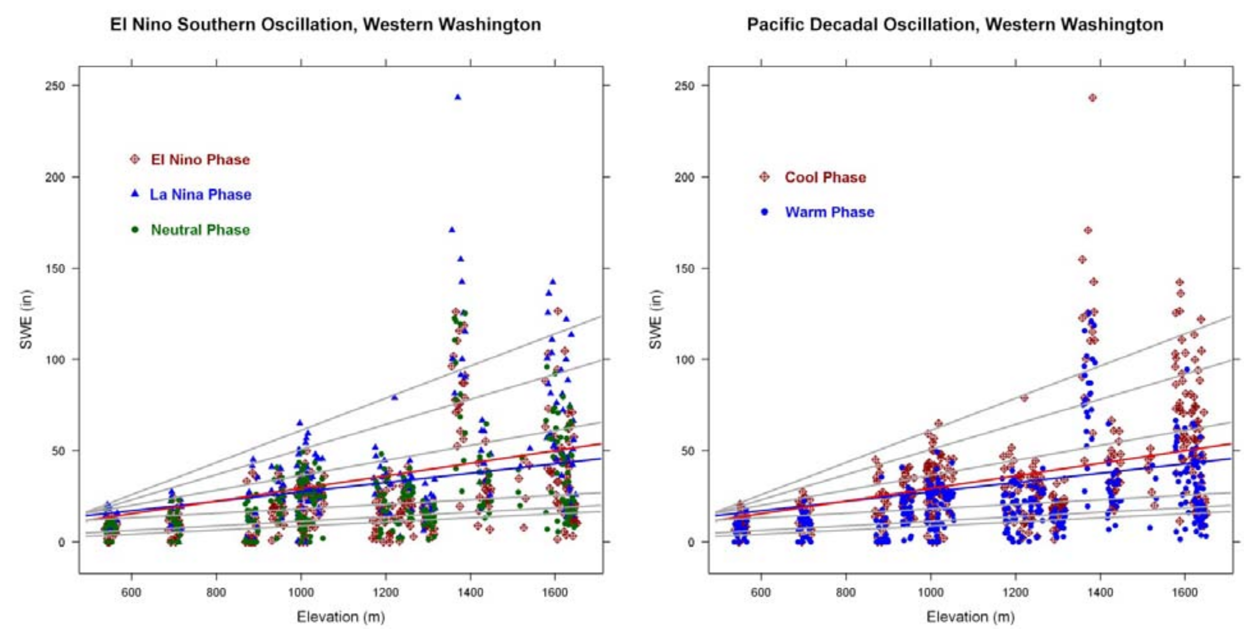

Figure 5. Elevation/SWE relationships modeled by OLS and quantile regression for the April 1 records in the Olympics and central Washington Cascades, by climate cycle. Legend for regression lines is the same as that in Figure 4.

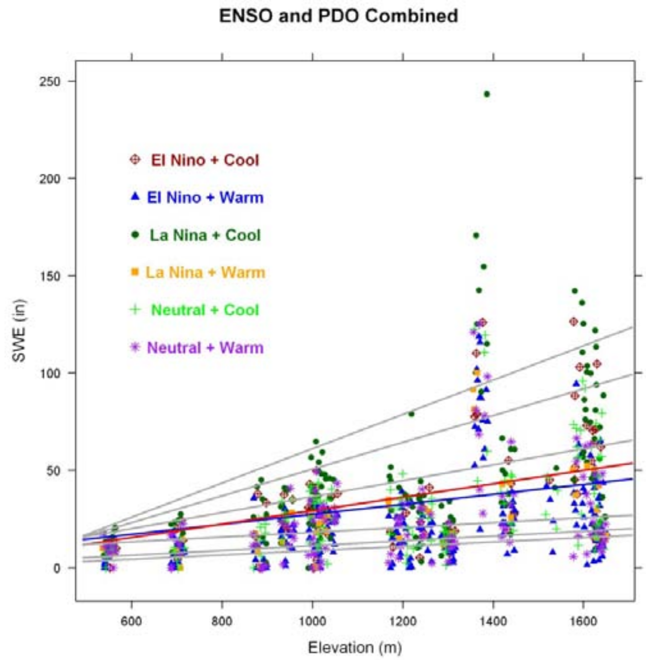

Figure 6. Elevation/SWE relationships modeled by OLS and quantile regression for the April 1 records in the Olympics and central Washington Cascades, by combined climate cycles. Legend for regression lines is the same as that in Figure 4.

CONCLUSIONS

In all sampling months in the Dungeness watershed, the mean has a higher prediction level than the median for nearly the entire elevation range, indicating probable influences of both the high SWE outliers and heterogeneous variance. It appears that the courses with a northern aspect have slightly higher SWE than those with a southern aspect when they share an elevation band. For this particular analysis, though, it is apparent that our dataset is too small and limited across elevation bands to provide more meaningful insight on underlying processes; this is most noticeable in December and April, when the datasets are the smallest. However, it is still clear that quantile regression can usually better account for the observed data than could a typical OLS regression.

Across the larger study area, there appears to be a clear relationship between the cool phase of the PDO and higher SWE rates as compared to the warm PDO phase, as well as for ENSO in the La Niña or neutral phase years as compared with El Niño. Again, the increasing variance with increasing elevation here cannot be effectively modeled with normal regression.

The results of this study indicate that quantile regression rates may be more useful for describing and modeling SWE by elevation, particularly when attempting to account for climate cycles.

ACKNOWLEDGEMENTS

Special thanks to the many field assistants from Peninsula College and Western Washington University who helped us gather the Dungeness Watershed snow data, often under adverse winter conditions. This material is based upon work supported in part by NASA's Science Mission Directorate under Cooperative Agreement #NNA06CN06A, as well as through the National Science Foundation under REU Grant #0452328. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA or the National Science Foundation.

REFERENCES

Cade, B.S. and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1(8): 412-420.

Koenker, R. 2010. *quantreg*: Quantile Regression. R package version 4.53. http://CRAN.R-project.org/package=quantreg.