

# Analysis of Linear Regression Techniques to Determine $\delta^{18}\text{O}_R$ and $\delta^{13}\text{C}_R$ : Implications and Applications

J.M. Zobitz<sup>1,\*</sup>, J.P. Keener<sup>1</sup>, D. R. Bowling<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Utah, 155 S 1400 E Salt Lake City, UT 84112

<sup>2</sup>Department of Biology, University of Utah, 257 S 1400 E, Salt Lake City, UT 84112

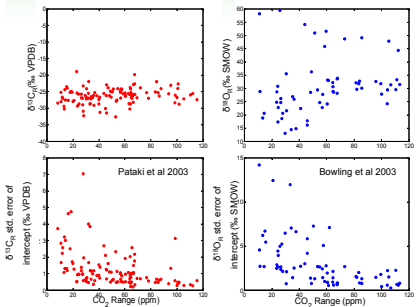
\* Corresponding author: zobitz@math.utah.edu

## Introduction

- Determining the isotopic signature of ecosystem respiration (either  $\delta^{13}\text{C}_R$  or  $\delta^{18}\text{O}_R$ ) using atmospheric mixing relationships is a challenge when there is little variation in the  $\text{CO}_2$  samples collected (Pataki et al 2003). For  $^{13}\text{C}$ , measurement uncertainty is the major factor limiting determination of  $\delta^{13}\text{C}_R$  since observed variation in  $\text{CO}_2$  is always accompanied by a variation in  $\delta^{13}\text{C}$ . For  $^{18}\text{O}$ , measurement uncertainty is also a limiting factor, but there are equilibration processes that influence  $\delta^{18}\text{O}$  without altering  $\text{CO}_2$ . Thus determination of  $\delta^{18}\text{O}_R$  via Keeling plots represents a major challenge (Bowling et al 2003).
- Here we focus on  $\delta^{13}\text{C}_R$  to examine the mathematics behind the mixing relationships and regressions involved, with decreasing uncertainty in  $\delta^{13}\text{C}_R$  as our primary objective. The results are directly applicable to  $\delta^{18}\text{O}_R$  (when the use of Keeling plots is appropriate for  $\delta^{18}\text{O}_R$ ).
- To determine  $\delta^{13}\text{C}_R$ , we use an inverse relationship between  $\text{CO}_2$  concentrations ( $[\text{CO}_2]$ ) and isotopic measurements (Keeling plots, Keeling 1958), and also a linear relationship between  $[\text{CO}_2]$  and the product of  $[\text{CO}_2]$  and isotopic measurements (Miller/Tans plots, Miller & Tans, 2003).
- Our goal is to develop a general-purpose framework for error propagation so one can compare both mixing relationships (Keeling or Miller/Tans) and regressions involved (OLS, GMR, or ODR) to determine which one leads to the best characterization of  $\delta^{13}\text{C}_R$ .

## Increased Uncertainty at Low Sampling Ranges

- For both oxygen and carbon isotopic studies, a wide sample range is important to obtain estimates of  $\delta^{18}\text{O}_R$  that have acceptable uncertainty. In particular, as sampling range decreases, error in  $\delta^{18}\text{O}_R$  increases.



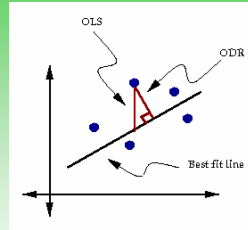
**Figure 1:** (Left Panels)  $\delta^{13}\text{C}_R$  and standard error of Model I intercept calculated using a Keeling GMR regression versus  $\text{CO}_2$  range using data from Pataki et al 2003. VPDB was used as the isotopic standard (Right Panels)  $\delta^{18}\text{O}_R$  calculated using a Keeling GMR regression versus  $\text{CO}_2$  range using data from Bowling et al 2003. SMOW was used as the isotopic standard. Note that as  $\text{CO}_2$  range decreases, the variability in both  $\delta^{13}\text{C}_R$  and  $\delta^{18}\text{O}_R$  increases.

## Acknowledgments

This work was supported under NSF Grant # DGE-0217424

## Linear Regression & Uncertainty Propagation

- For any data that one needs to fit to a best fit line, one can find the residual of the data points to the hypothetical best fit line.
- For Ordinary Least Squares (OLS) Regression (Model I Regression), it is assumed there is no variation in the independent variable.
- For Orthogonal Distance (Model II) Regression (ODR) it is assumed that both variables have an equal amount of variation.
- Geometric Mean Regression (GMR) is another Model II regression technique. We omit GMR here as we obtain similar results as ODR in our analysis.
- Much debate exists about how to assign uncertainties via a Model I or Model II regression. (Pataki et al 2003)
- It is possible to develop an analytical equation for the variance of the slope or intercept of a linear regression formula that depends on the data set along with the errors associated with the independent and dependent variables. (Zobitz et al, in preparation).



**Figure 2:** Data points (blue) are fitted to a best fit line, which is the line that minimizes the sum of the square residuals. For Ordinary Least Squares (OLS), the residual (shown in red) is the vertical distance from each data point. For Orthogonal Distance Regression (ODR) the residual is the perpendicular distance from the best fit line.

## Perturbation of a "Perfect" Data Set

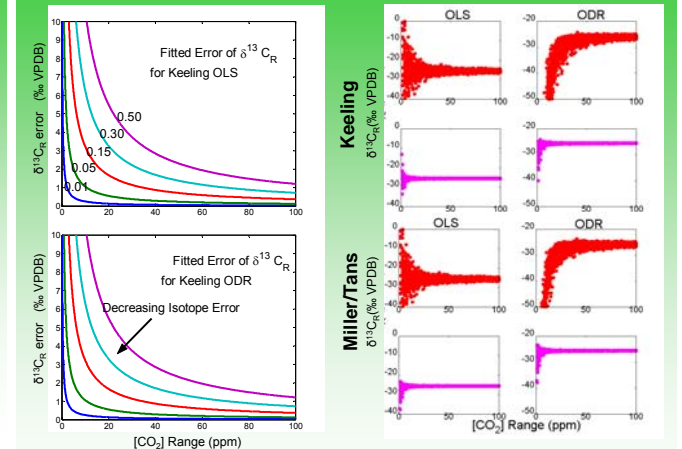
- By taking a data set of  $[\text{CO}_2]$  and isotope data, we can generate a "perfect" data set without error in the observations of  $\text{CO}_2$  and isotopes.
- This perfect data set was perturbed by adding a controlled amount of noise to both variables. We then randomly subsampled the perturbed data set 5000 times and calculated a Keeling or Miller/Tans and OLS or ODR regression.

## References

- Bowling, D. R., N. G. McDowell, J. M. Welker, B. J. Bond, B. E. Law, J. R. Ehleringer. 2003. Oxygen isotope content of  $\text{CO}_2$  in nocturnal ecosystem respiration: 1. Observations in forests along a precipitation transect in Oregon, USA. *Global Biogeochemical Cycles*. 17(4), 1120. doi:10.1029/3003GB020281.
- Keeling, C. D. 1958. The concentrations and isotopic abundances of atmospheric carbon dioxide in rural areas. *Geochim. Cosmochim. Acta*. 13:322-334.
- Miller, J. B., P. Tans. 2003. Calculating isotopic fractionation from atmospheric measurements at various scales. *Tellus*. 55b:207-214.
- Pataki, D. E., J. R. Ehleringer, L. B. Flanagan, et al. 2003. The application and interpretation of Keeling plots in terrestrial carbon cycle research. *Global Biogeochemical Cycles*. 17(1):1022.

## Simulation Results

- By using the theoretical framework outlined, results indicate decreasing error in the isotopic sample greatly improved accuracy in  $\delta^{13}\text{C}_R$ .



**Figure 3:** Error in  $\delta^{13}\text{C}_R$  as a function of  $\text{CO}_2$  range for both OLS and ODR using Keeling regressions. The numbers on the top graph represents the standard deviation in  $\delta^{13}\text{C}_R$  observations that the "perfect" data set was perturbed by. In all simulations, the standard deviation in  $\text{CO}_2$  measurements was .15 ppm.

**Figure 4:**  $\delta^{13}\text{C}_R$  determined by a Keeling (top 4 panels) or Miller/Tans (bottom 4 panels) regressions by subsampling a data set with known error. Red is data that has an error of .15 ppm, .15% error. Magenta is data that has .15 ppm, .01% error.

## Implications for Research

- Our simulations suggest that the accuracy and the precision of the instruments determine the accuracy of  $\delta^{13}\text{C}_R$ .
- Improving the precision of the isotopic measurements decreases the influence of the range problem.
- For carbon isotopic studies, Keeling or Miller-Tans regressions give equivalent results with no appreciable difference in uncertainty.
- Due to relative error magnitude ratio between  $\text{CO}_2$  and isotopes, there is no additional benefit in doing a Model II regression over a Model I regression over the range of precision of instrumentation.