Boondoggle Manuscript Data Analysis Journal

December 16, 2010

Contents

1 23 July 2010: Recalculate the change in percent epilimnetic area in GTH 91 case study based on more realistic Kd changes

1.1 Goal

The following changes need to be made to the dissertation analyses:

1. use a more realistic model for Percent Sediment Area Z
2. use a better supported hypothetical change in K_d

1.2 Data Analysis

I added the **GTH91.bathy** data frame from the bathy.summary workspace to the **GTH91_case** workspace.

I made a new data.frame **GTH91.case.ms** that excluded all of the calculations from the disseration analysis so that I could recalculate the relevant variables (change in Kd and change in EpiArea) for the manuscript.

Objects Mix.Area and pred.Mix.Area were added to **GTH91.case.ms** where each is the sediment area above the thermocline calculated as:
Mix.Area = Area - (Area * (1 - $z/10)^{0.5}$)
$where z is either the actual TD depth or the predicted TD$.

1.3 Calculation of EpiArea based on the Livingstone and Imboden relationship

Using the relationship between sediment area and depth in Livingstone and Imboden 1996 (eq 3):
$A(z) = A_o(1 - z/z_{max})^q$

Where:

1. $A(z)$ = the sediment area below depth z (m)
2. A_o = the surface area of the lake (Ha)
3. z/z_{max} = the ratio of depth z to maximum depth (z_{max})
4. q = a non-dimentional exponent; estimated graphically to be 0.5 (see 16.2.1 of Data Analysis Journal 2)

Estimated Sediment Area (Ha) Above the Measured TD
summary(Mix.Area) Min. 1st Qu. Median Mean 3rd Qu. Max. 0.2639 0.4464 0.4844 0.5734 0.7776 0.8229

Estimated Sediment Area (Ha) Above the Predicted TD
summary(pred.Mix.Area) Min. 1st Qu. Median Mean 3rd Qu. Max. 0.1546 0.3869 0.5231 0.4953 0.5688 0.8372

2 16 December 2010; Analyze TD based on Standardized Variables

2.1 Goal

To be able to compare the relative effect of the Kd, Area, and Julian on TD I need to standardize the variables to the same units (Z-scores)

2.2 Data

1. I combined the **GTH91_case** and **boondoggle** workspaces into **boondoggle.ms.R** workspace all of the data are in this workspace
2. New Objects
   (a) Kd.Z, is the z scores for Kd

(b) Area.Z is the z scores for Area

(c) Julian.Z is the z scores for Julian

(d) pred.TD.Kd is the TD predicted using the regression of Kd and TD only

3. Code is saved in /media/working/working_files/drafts_working/manuscripts/boondoggle/Data_analysis.R

2.3 Analysis

2.3.1 Analysis of the TD with 1, 2, and 3 predictors

```
> # relationship between Kd and TD
> summary(lm(TD ~ Kd))

Call:
lm(formula = TD ~ Kd)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7236 -0.6497 -0.2573  0.6498  4.1443

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.9186     0.4775  18.677  < 2e-16 ***
Kd           -6.0109     0.8681  -6.924 1.66e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.179 on 43 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.5272,Adjusted R-squared: 0.5162
F-statistic: 47.94 on 1 and 43 DF,  p-value: 1.658e-08

> summary(lm(TD ~ Kd + Julian))

Call:
lm(formula = TD ~ Kd + Julian)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9244 -0.6356 -0.1058  0.5026  4.0368

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.18072    3.37406  -0.646  0.52159
Kd          -4.58008    0.89308  -5.128 7.01e-06 ***
```

```
Julian       0.04947    0.01492   3.317  0.00189 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.062 on 42 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.6253,Adjusted R-squared: 0.6075
F-statistic: 35.05 on 2 and 42 DF,  p-value: 1.115e-09

> summary(lm(TD ~ Kd + Julian + Area))

Call:
lm(formula = TD ~ Kd + Julian + Area)

Residuals:
      Min        1Q    Median        3Q       Max
-2.390353 -0.592462  0.004308  0.463105  2.031078

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.081766   3.219839  -2.199 0.033544 *
Kd          -3.084772   0.874888  -3.526 0.001054 **
Julian       0.067080   0.013833   4.849 1.82e-05 ***
Area         0.037501   0.009969   3.762 0.000529 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9267 on 41 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.7214,Adjusted R-squared: 0.7011
F-statistic:  35.4 on 3 and 41 DF,  p-value: 1.862e-11
```

2.3.2 Predicting TD based on Kd alone:

```
> # Predicting TD using the simple model with only Kd
> # model : TD = -6.0109Kd + 8.9186  (R2 = 0.53)
> ## use GTH91.case.ms data.frame in GTH91_case workspace
> attach(GTH91.case.ms)
> pred.TD.Kd <- (Kd * -6.0109) + 8.9186
> detach(GTH91.case.ms)
> pred.TD.Kd
 [1] 2.661253 3.220868 3.705948 3.711357 3.239502 5.581949 4.159770 3.907914
 [9] 4.806543 3.864034 4.579932
```

2.3.3 Comparison of the predicted TD with Kd alone and with the full model

```
  > # analyze model
> summary(lm(pred.TD.Kd ~ Julian))

Call:
lm(formula = pred.TD.Kd ~ Julian)

Residuals:
     Min       1Q   Median       3Q      Max
-1.17262 -0.37590 -0.02087  0.39529  0.94342

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.39908    2.92901  -1.160   0.2757
Julian       0.03774    0.01501   2.515   0.0331 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6593 on 9 degrees of freedom
Multiple R-squared: 0.4126,Adjusted R-squared: 0.3474
F-statistic: 6.323 on 1 and 9 DF,  p-value: 0.03306

> summary(lm(pred.TD ~ Julian))

Call:
lm(formula = pred.TD ~ Julian)

Residuals:
     Min       1Q   Median       3Q      Max
-0.60085 -0.19261 -0.01069  0.20255  0.48341

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.29661    1.50083  -8.859 9.71e-06 ***
Julian        0.08634    0.00769  11.228 1.35e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.3378 on 9 degrees of freedom
Multiple R-squared: 0.9334,Adjusted R-squared: 0.926
F-statistic: 126.1 on 1 and 9 DF,  p-value: 1.354e-06

> summary(lm(TD ~ Julian))
```

```
Call:
lm(formula = TD ~ Julian)

Residuals:
     Min       1Q   Median       3Q      Max
-0.61144 -0.19899 -0.01234  0.22725  0.53834

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.049642   1.624174  -8.035 2.14e-05 ***
Julian        0.087557   0.008322  10.522 2.34e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.3656 on 9 degrees of freedom
Multiple R-squared: 0.9248,Adjusted R-squared: 0.9165
F-statistic: 110.7 on 1 and 9 DF,  p-value: 2.340e-06
```

The model using Kd alone (pred.TD.Kd) does a much poorer job of replicating the actual changes in TD in GTH 91 than the full model with Area and Julian (pred.TD). The R2 is low and the slope is much different.

2.3.4 Analysis of TD with standardized predictors

Calculation of the Z-scores:

```
# Calculate Z scores for Kd, Area, and Julian
# use boon.tot data.frame
Kd.Z <- (Kd - mean(Kd))/ sd(Kd)
Area.Z <- (Area - mean(Area)) / sd(Area)
Julian.Z <- (Julian - mean(Julian)) / sd(Julian)
```

The standardized variables are somewhat correlated:

```
> # correlations between standardized variables (Kd, Area, Julian)
> cor.test(Kd.Z, Area.Z)

Pearson's product-moment correlation

data:  Kd.Z and Area.Z
t = -2.4425, df = 45, p-value = 0.01858
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.57299814 -0.06094817
sample estimates:
      cor
```

```
-0.342127

> cor.test(Kd.Z, Julian.Z)

Pearson's product-moment correlation

data:  Kd.Z and Julian.Z
t = -3.6074, df = 45, p-value = 0.0007721
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6697059 -0.2158110
sample estimates:
       cor
-0.4736254
```

The model with the standardized predictors

```
> # Analyze the relationship between TD and the standardized Kd, Area, Julian
> summary(lm(TD ~ Kd.Z + Area.Z + Julian.Z))

Call:
lm(formula = TD ~ Kd.Z + Area.Z + Julian.Z)

Residuals:
      Min        1Q    Median        3Q       Max
-2.390353 -0.592462  0.004308  0.463105  2.031078

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.8606     0.1383  42.388  < 2e-16 ***
Kd.Z         -0.6192     0.1756  -3.526 0.001054 **
Area.Z        0.5831     0.1550   3.762 0.000529 ***
Julian.Z      0.8145     0.1680   4.849 1.82e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9267 on 41 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.7214,Adjusted R-squared: 0.7011
F-statistic:  35.4 on 3 and 41 DF,  p-value: 1.862e-11
```