Supplemental Document

Decreased water transparency of nearshore Laurentian Great Lakes habitats is driven by increased dissolved organic carbon.

Thomas Fisher

Table of Contents

## Caveats

A few important things to remember when looking at this report

* We used mixed effect models, so some of the standard regression measures and diagnostics are not verbatim the same; *e.g.*, marginal and conditional instead of the standard coefficient of determination.
* The response variable in each model has been transformed with a base-10 logarithm.
  + So linear models are reported, but for a log-response.
  + In the original units, we effectively are fitting an exponential or power function.
* The covariate input, DOC, has also been log-transformed as it is right skewed and this helps meet the linearity assumption.
* We fit models with interaction terms, so the main effects can only be interpreted with respect to the interaction terms.
* A backward selection algorithm was implemented to choose the *best* model using the Bayesian Information Criteria (BIC). This document reports that chosen model.

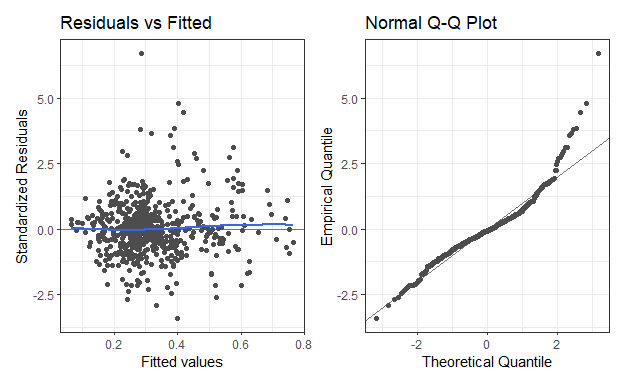
## Dissolved Organic Carbon analysis

The backward selection procedure chose the following model

where the term is a random effect to account for the variability at each site (similar to a block term in experimental design). The variable measures times since the beginning of data collection. The variables , , and are indicator, or dummy, variables indicating the observation occurred in that season or water type.

### Residuals analysis

First, we take a look at some residual plots before proceeding to look at the model



Nothing overly concerning here.

### Measures of model fit

Some overall model measures

| N | df | Resid.SE | AIC | BIC | Marginal R2 | Conditional R2 |
| --- | --- | --- | --- | --- | --- | --- |
| 665 | 656 | 0.068 | -1397.356 | -1356.858 | 0.089 | 0.778 |

We note that the fixed effects in the model do not capture much of the variability in DOC (the marginal ). This indicates that the DOC levels are varying quite a bit by the site.

### Model coefficients

We can look at the coefficients for the fixed effects in the fitted model (or expected value of the random effect)… Recall the response here is the of the 1% PAR Depth (m)

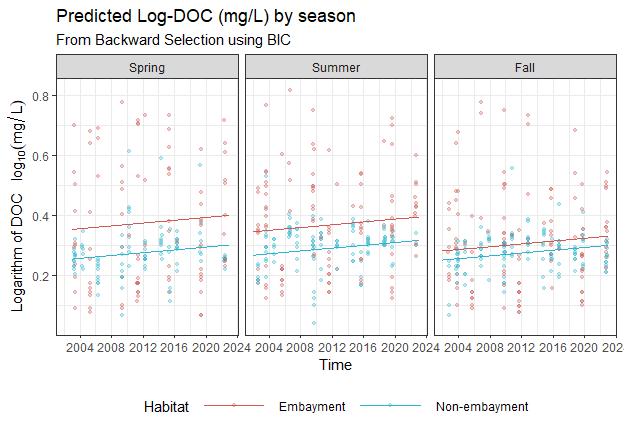
| Term | Estimate | SE | t.stat | df | p.value |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.350 | 0.023 | 15.359 | 77.292 | 0.000 |
| Year | 0.002 | 0.000 | 5.181 | 604.348 | 0.000 |
| SeasonSummer | -0.005 | 0.010 | -0.561 | 598.842 | 0.575 |
| SeasonFall | -0.069 | 0.010 | -7.165 | 599.091 | 0.000 |
| WaterBodyNon-embayment | -0.099 | 0.032 | -3.045 | 69.135 | 0.003 |
| SeasonSummer:WaterBodyNon-embayment | 0.021 | 0.014 | 1.526 | 598.606 | 0.127 |
| SeasonFall:WaterBodyNon-embayment | 0.068 | 0.013 | 5.103 | 599.148 | 0.000 |

and the two standard deviation measures associated with the model

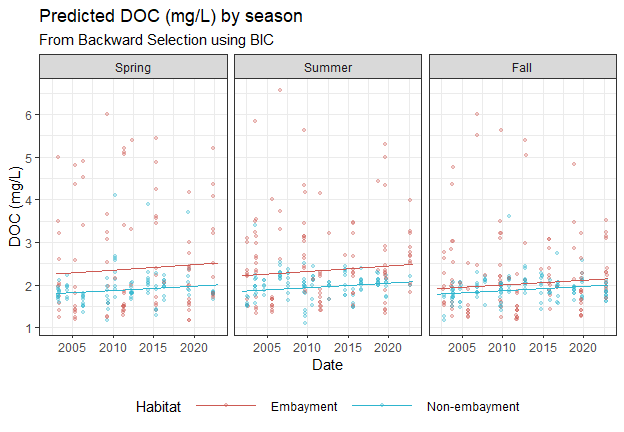
| Group | Term | Estimate |
| --- | --- | --- |
| SiteID | sd\_\_(Intercept) | 0.121 |
| Residual | sd\_\_Observation | 0.068 |

### Plots of fitted model

And a plot of the fitted model (response is log-transformed)



And on the original units



Note, these are exponential fits, but the predicted values on the logarithmic scale (a linear model) do not vary all that much, so the fitted models on the original scale appears almost linear.

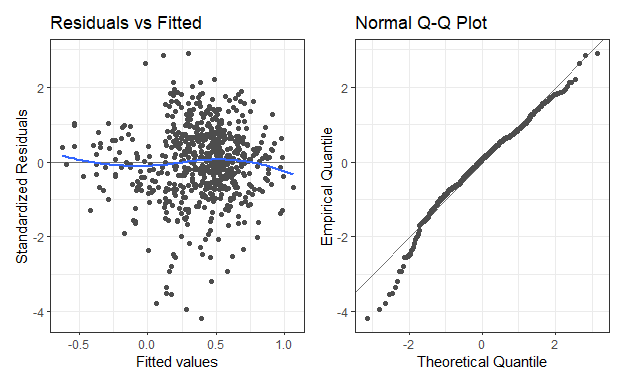
## 1% UV depth analysis

Our backward selection procedure chose the following model

where the term is a random effect to account for the variability at each site (similar to a block term in experimental design). The variables , , and would all be indicator, or dummy, variables indicating the observation occurred in that season, water type or in the presence of a river.

### Residual analysis

First, we take a look at some residual plots before proceeding to look at the model output



Nothing overly concerning here. Some minor deviation from Normality but regression methods are robust to some minor deviations from normality.

### Measure of model fit

Some overall model measures

| N | df | Resid.SE | AIC | BIC | Marginal R2 | Conditional R2 |
| --- | --- | --- | --- | --- | --- | --- |
| 631 | 618 | 0.113 | -710.842 | -653.027 | 0.58 | 0.848 |

Here, the fixed effects (which include DOC) explain about 57% of the variability. After including the site level variation, the overall model explains about 84% of the variability in the log of 1% UV depth.

### Model coefficients

The fixed effects in the fitted model (or expected value of the random effect)… Recall the response here is the of the 1% UV Depth (m)

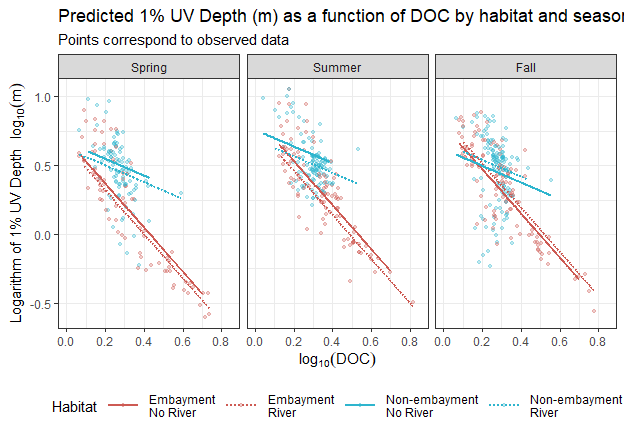
| Term | Estimate | SE | t.stat | df | p.value |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | 0.690 | 0.043 | 15.887 | 169.499 | 0.000 |
| log10(DOC) | -1.603 | 0.089 | -18.086 | 484.283 | 0.000 |
| SeasonSummer | 0.163 | 0.018 | 9.055 | 563.116 | 0.000 |
| SeasonFall | 0.102 | 0.018 | 5.577 | 587.324 | 0.000 |
| WaterBodyNon-embayment | -0.022 | 0.063 | -0.343 | 221.121 | 0.732 |
| RiverPresentRiver | -0.047 | 0.046 | -1.024 | 77.776 | 0.309 |
| SeasonSummer:WaterBodyNon-embayment | -0.072 | 0.024 | -3.006 | 562.515 | 0.003 |
| SeasonFall:WaterBodyNon-embayment | -0.152 | 0.024 | -6.381 | 579.500 | 0.000 |
| log10(DOC):WaterBodyNon-embayment | 0.995 | 0.148 | 6.711 | 618.659 | 0.000 |
| SeasonSummer:RiverPresentRiver | -0.024 | 0.026 | -0.934 | 562.444 | 0.351 |
| SeasonFall:RiverPresentRiver | 0.090 | 0.025 | 3.612 | 564.407 | 0.000 |

and the two standard deviation measures associated with the model

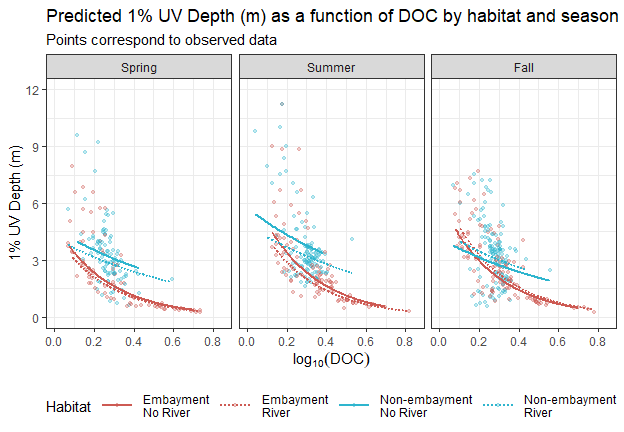
| Group | Term | Estimate |
| --- | --- | --- |
| SiteID | sd\_\_(Intercept) | 0.150 |
| Residual | sd\_\_Observation | 0.113 |

### Plots of fitted model

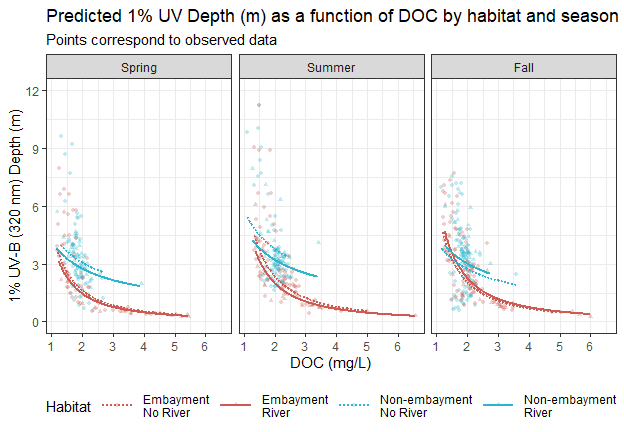
Here is a plot of the fitted linear model. Both the response (1% UV Depth) and covariate (DOC) are on a base-10 logarithm scale.



Here is the same model with the response in the original units.



and here is a version where both the response and DOC are in the original units.



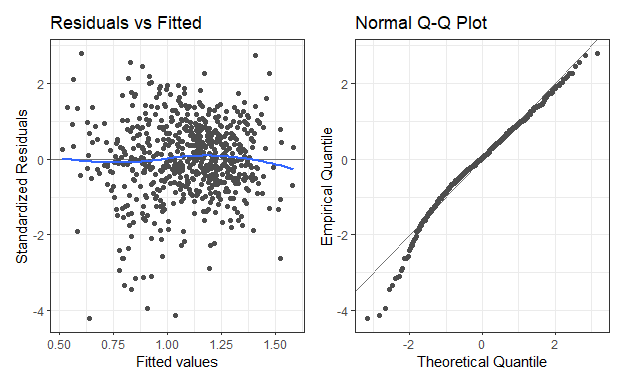
## 1% PAR depth analysis

The backward selection procedure chose the following model

where the term is a random effect to account for the variability at each site (similar to a block term in experimental design). The variables , , , , and are indicator, or dummy, variables indicating the observation occurred in that lake, season, or water type.

### Residual analysis

First, we take a look at some residual plots before proceeding to look at the model



Nothing overly concerning here. Some minor deviation from Normality but regression methods are robust to some minor deviations from normality.

### Measures of model fit

Some overall model measures

| N | df | Resid.SE | AIC | BIC | Marginal R2 | Conditional R2 |
| --- | --- | --- | --- | --- | --- | --- |
| 665 | 654 | 0.134 | -616.314 | -566.817 | 0.524 | 0.67 |

The fixed effect in the chosen model explain about 52% of the variability in the logarithm of PAR depth. When accounting for the variability within each site, the model explains about 66% of the varibility.

### Model coefficients

The fixed effects in the fitted model (or expected value of the random effect)… Recall the response here is the of the PAR Depth (m)

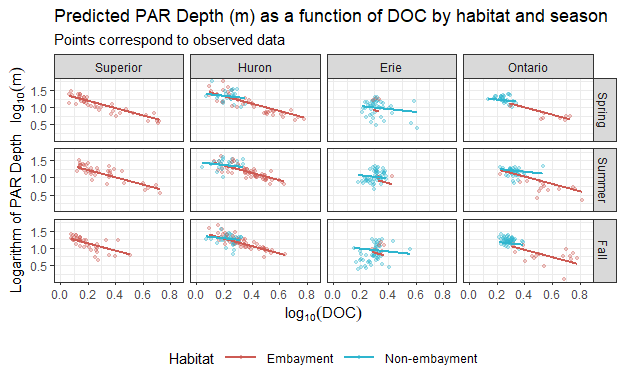
| Term | Estimate | SE | t.stat | df | p.value |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | 1.400 | 0.034 | 40.606 | 113.193 | 0.000 |
| log10(DOC) | -1.066 | 0.075 | -14.252 | 345.611 | 0.000 |
| LakeHuron | 0.130 | 0.037 | 3.460 | 60.493 | 0.001 |
| LakeErie | -0.170 | 0.052 | -3.284 | 59.056 | 0.002 |
| LakeOntario | 0.021 | 0.047 | 0.450 | 64.145 | 0.654 |
| SeasonSummer | 0.031 | 0.013 | 2.317 | 599.690 | 0.021 |
| SeasonFall | -0.055 | 0.013 | -4.138 | 616.868 | 0.000 |
| WaterBodyNon-embayment | -0.116 | 0.057 | -2.042 | 207.474 | 0.042 |
| log10(DOC):WaterBodyNon-embayment | 0.646 | 0.149 | 4.344 | 630.684 | 0.000 |

and the two standard deviation measures associated with the model

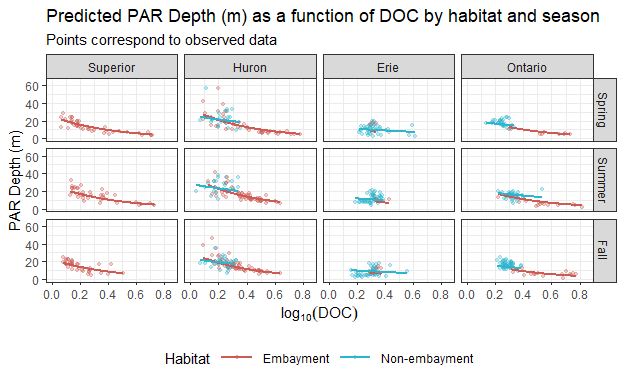
| Group | Term | Estimate |
| --- | --- | --- |
| SiteID | sd\_\_(Intercept) | 0.089 |
| Residual | sd\_\_Observation | 0.134 |

### Plots of the fitted model

And a plot of the fitted linear model (log of PAR with log of DOC



and with PAR in the original units



and another version with DOC and PAR in their original units.

