

STATISTICAL ANALYSIS

Diplina Paul¹ and Steven G. Hall^{1*}

Analysis of efficiency of different media types.

To compare the nitrate removal efficiency of each media by time, the data was split by influent nitrate concentration blocks, then fitted with following regression model assuming the influent nitrate concentration was the same for all media and had no affect:

$$\log(\text{effluent}) = \text{day} + \text{media} + \text{day} * \text{media} + \epsilon$$

Media is a three-level factor variable, with level *B* for Biochar, level *Z* for Zeolite, and level *P* represents Plastic.

Nitrate at 50 mg/L

For influent nitrate concentration at 50 mg/mL, the data was fitted to the model to obtain the following:

```
##
## Call:
## lm(formula = log(V1) ~ day * media, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54849 -0.08367 -0.00027  0.08918  0.43002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.130713   0.055179  74.860 < 2e-16 ***
## day         -0.097800   0.003646 -26.822 < 2e-16 ***
## mediaP      -0.098749   0.078035  -1.265   0.210
## mediaZ       0.077325   0.078035   0.991   0.325
## day:mediaP   0.032303   0.005157   6.264 2.14e-08 ***
## day:mediaZ  -0.031833   0.005157  -6.173 3.14e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1468 on 75 degrees of freedom
## Multiple R-squared:  0.9717, Adjusted R-squared:  0.9698
## F-statistic: 515.4 on 5 and 75 DF,  p-value: < 2.2e-16
```

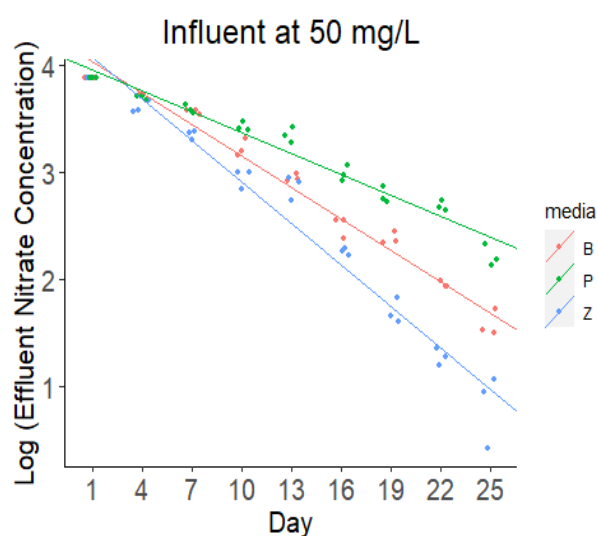
The *Media* variable is significantly useful, and the model explained 97.17% of the variability of the data, shows well fitting.

According to the coefficients, we have following fits:

$$\log(\text{effluent}|\text{Biochar}) = 4.130713 - 0.0978 * \text{day}$$

$$\log(\text{effluent}|\text{Zeolite}) = 4.208038 - 0.129633 * \text{day}$$

$$\log(\text{effluent}|\text{Plastic}) = 4.031964 - 0.065497 * \text{day}$$



Above is the plot of $\log(\text{effluent})$ by day. Linear regression fit lines for biochar (B), zeolite (Z) and plastic (P) are shown in different colors. The slopes which indicate the nitrate removal efficiency were compared among each factor:

```
m.lst <- lstrends(fit2, "media", var="day")
```

```
pairs(m.lst)
```

```
## contrast estimate      SE df t.ratio p.value
## B - P      -0.0323 0.00516 75 -6.264  <.0001
## B - Z       0.0318 0.00516 75  6.173  <.0001
## P - Z       0.0641 0.00516 75 12.437  <.0001
##
```

```
## P value adjustment: tukey method for comparing a family of 3 estimates
```

We found evidence of statistically significant differences between every two medias. From the estimate of contrasts, both Biochar and Zeolite have higher efficiency than plastic, and Zeolite has even higher nitrate removal efficiency than Biochar.

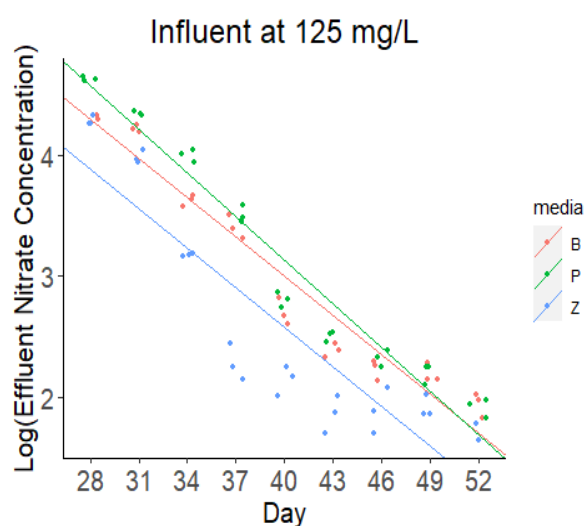
Nitrate at 125 mg/L

We fit the data of influent nitrate concentration at 125 mg/L to the model as before:

```
##
## Call:
## lm(formula = log(V1) ~ day * media, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.75615 -0.23071  0.03494  0.19846  0.50656
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 7.295358 0.294470 24.775 <2e-16 ***
## day -0.107325 0.007227 -14.850 <2e-16 ***
## mediaP 0.636132 0.416443 1.528 0.131
## mediaZ -0.380559 0.416443 -0.914 0.364
## day:mediaP -0.012562 0.010221 -1.229 0.223
## day:mediaZ -0.001000 0.010221 -0.098 0.922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2909 on 75 degrees of freedom
## Multiple R-squared: 0.9116, Adjusted R-squared: 0.9057
## F-statistic: 154.7 on 5 and 75 DF, p-value: < 2.2e-16
```

Under the condition of 125 mg/L influent nitrate, the model indicates only *day* variable is significant useful, none of the media types can significantly affect the effluent nitrate concentrations. We plot the fitted model with data to better see the trends.



From the plot, we can see the slope of three media types are roughly parallel, we further test the differences of slopes:

```
## contrast estimate SE df t.ratio p.value
## B - P 0.0126 0.0102 75 1.229 0.4400
## B - Z 0.0010 0.0102 75 0.098 0.9947
## P - Z -0.0116 0.0102 75 -1.131 0.4981
##
```

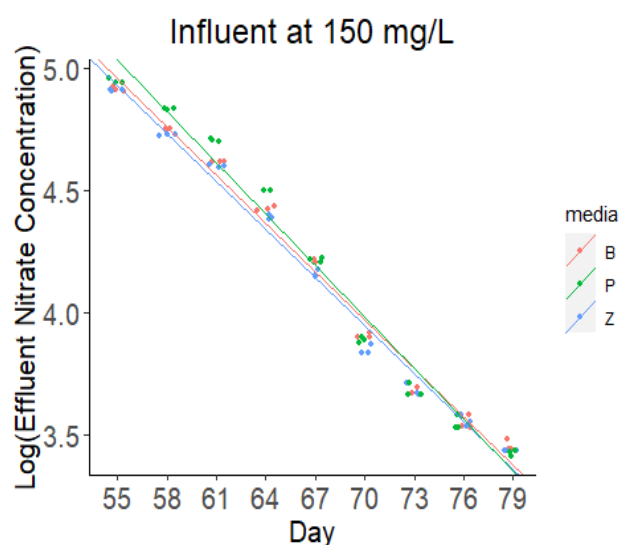
P value adjustment: tukey method for comparing a family of 3 estimates

By comparing the slopes under each media types, there was no significant differences observed.

Nitrate at 150 mg/L

```
##
## Call:
## lm(formula = log(V1) ~ day * media, data = data4)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.115417 -0.051292  0.004549  0.056369  0.104852
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.5803497   0.1123045   76.403  <2e-16 ***
## day         -0.0658574   0.0016651  -39.552  <2e-16 ***
## mediaP       0.3118981   0.1588225    1.964   0.0533 .
## mediaZ      -0.0449856   0.1588225   -0.283   0.7778
## day:mediaP  -0.0042737   0.0023548   -1.815   0.0735 .
## day:mediaZ   0.0003447   0.0023548    0.146   0.8840
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06702 on 75 degrees of freedom
## Multiple R-squared:  0.9849, Adjusted R-squared:  0.9839
## F-statistic: 978.6 on 5 and 75 DF,  p-value: < 2.2e-16
```



```
m.lst4 <- lstrends(fit4, "media", var="day")
pairs(m.lst4)
## contrast estimate SE df t.ratio p.value
## B - P      0.004274 0.00235 75  1.815  0.1717
## B - Z     -0.000345 0.00235 75 -0.146  0.9883
## P - Z     -0.004618 0.00235 75 -1.961  0.1290
##
```

P value adjustment: tukey method for comparing a family of 3 estimates

From both plot and slope difference comparison, we can see there is no difference on nitrate removal efficiency when the influent concentration is around 150 mg/L.

Citation: Lastname, F.; Lastname, F.; Last-name, F. Title. *Water* **2021**, *13*, x. <https://doi.org/10.3390/xxxxx>

Received: date

Accepted: date

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors.

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).