



Modeling Effects of Weather and Season on Contaminants in Coal Ash Ponds

Gordon Williams, Keyang Xue, Lydie Costes

ABSTRACT

Fossil fuel power plants deposit waste into landfills and surface impoundments known as coal ash ponds to minimize airborne pollution. The chemicals in these ponds can leach into groundwater, causing serious health risks for the environment and human health. We examined possible climatic predictors of coal ash pond contamination by comparing chemical data collected from coal ash ponds in North Carolina over a two-year period with local weather data from that same period. Our analysis focused on total dissolved solids, arsenic, calcium, chromium, and lead. Variations in the concentrations of all five contaminants were found to be predicted by temperature, precipitation, and/or season. These findings have implications for future monitoring and projections forecasted by climate change.

INTRODUCTION

Literature review

Coal fired power plants generate significant amounts of waste in the form of coal ash that is usually placed into impoundments, landfills, or coal ash ponds for storage (Goemann, 2015). Over time and with exposure to the environment, various contaminants can leach into the surrounding groundwater. Newer regulations often stipulate that ponds be lined to prevent leaching, but many older ponds do not have any lining or are ineffectively lined (Hall, 2019). Since approximately 2.4 million North Carolinians rely on private unregulated wells for drinking water, understanding the factors that affect contaminant concentration can be useful in predicting when and how coal ash contaminants could leach into groundwater, posing risks for human and environmental health (NCDHHS).

Seasonal climatic changes such as rainfall and temperature fluctuate throughout the year and may affect the leachability of various contaminants from coal ash ponds into the surrounding groundwater (Wang et al., 2020). Increases in rainfall or temperature may help to mobilize contaminants like chromium, arsenic, or lead which are known to be hazardous to human health (Filippelli et al., 2020).

The contaminants we decided to focus on are total dissolved solids (TDS), calcium, chromium, arsenic, and lead. These contaminants were chosen as some are toxic (chromium, arsenic, lead), while others might affect drinking water aesthetic and condition of pipes (calcium, TDS) (Goemann, 2015; Hall, 2019). TDS can also be used as an indicator of general water quality, as it is the total content of all dissolved organic and inorganic content.

We chose precipitation, temperature, and season as predictor variables since they are readily available and there are good temporal and spatial records of precipitation and temperature. Additionally, weather and climate forecasts could be used to predict future contamination if there is a strong correlation between any contaminant concentration and either of these variables.

Research question

Our central research question is whether concentrations of contaminants in coal ash ponds can be predicted by weather conditions and/or time of year. Understanding this relationship could be useful to predict consequences for human and environmental health.

Hypotheses

For each contaminant, our null hypothesis is that temperature, precipitation, and season do not predict levels of contamination in the coal ash ponds of power plants. Our alternative hypothesis is that temperature, precipitation, and/or season do predict levels of contamination.

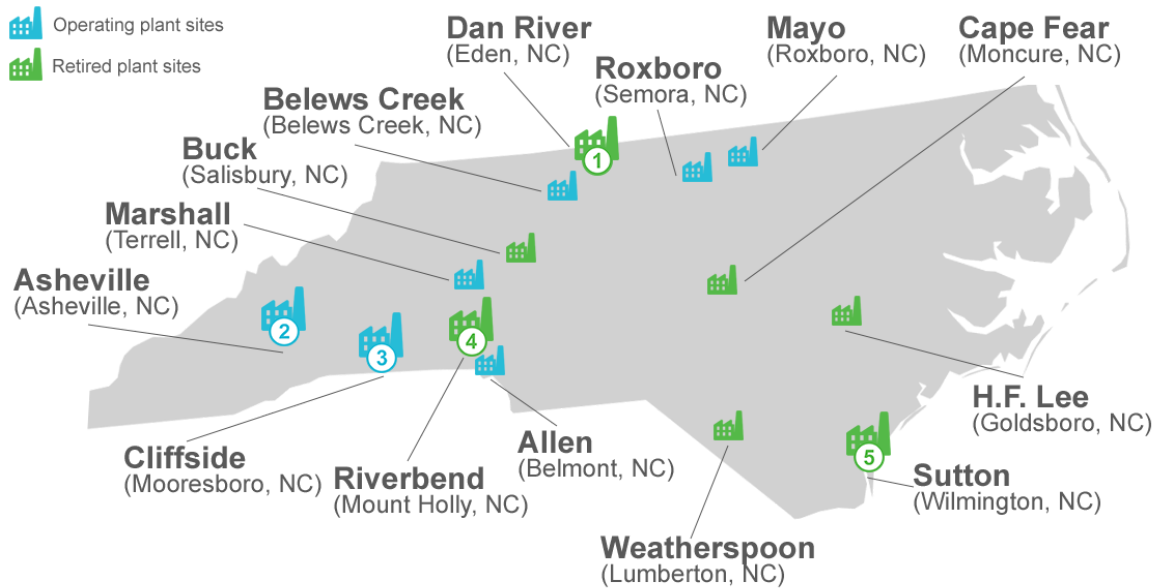


Figure 1. Map of coal fired power plants in North Carolina. All of these locations contain coal ash ponds. Data from each, except Cape Fear and Riverbend, were used in this project. (Gray, 2014)

METHODOLOGY

Data sources and extraction

The main dataset used consisted of data from utility groundwater monitoring reports of monitoring wells near coal ash impoundments collected throughout 2016 and 2017. This dataset was prepared by the Environmental Integrity Project for the United States Environmental Protection Agency and was accessed through the *regulations.gov* website. Since the dataset is large (greater than 100,000 lines) it was pared down to only the contaminants of interest within North Carolina and reorganized so that observations from a given month and year and at each monitoring well and power plant could all be paired with the contaminant concentrations.

Precipitation data were acquired from USGS monitoring stations across North Carolina. The geographically closest rain gauge to each coal ash impoundment, with historical precipitation data during 2016 and 2017, was chosen for each power plant. The downloadable data were in total inches of precipitation per day, which was later summed for total monthly precipitation during each month and paired with the relevant power plant, monitoring well, and associated month and year.

Temperature data were collected from *timeanddate.com* as the average monthly temperature for the geographic region (mountains, piedmont, and coastal plains) of each power plant. These data were also paired with their associated power plant, monitoring well, and relevant month and year.

Data description and preparation

The independent fixed variables were *precip* (total monthly precipitation in inches, mean = 3.650), *temp* (average monthly regional temperature in degrees Fahrenheit, mean = 66.041), and *season* (coded as four 3-month intervals). We also included two random variables to account for spatial autocorrelation: *site* (power plants, n = 13) and *well.id* (monitoring wells, n = 506). We selected five dependent variables, each measured in milligrams per liter: Total Dissolved Solids (*TDS*), *Arsenic*, *Calcium*, *Chromium*, and *Lead*. For each dependent variable, we created a separate model. All dependent variable data were extremely right skewed, so we used a logarithm transformation to render them more normal. *TDS* was the only dataset containing values of zero, and we re-coded those values as 0.1 to allow for the log transformation (mean=501.0, median=204.0). We tried a logarithm transformation on *precip* as well, but doing so did not help our residuals achieve normality so *precip* was left untransformed for the final analyses. The two continuous variables *precip* and *temp* did not demonstrate excessive multicollinearity ($r = 0.299$).

Analysis

For each dependent variable, we ran a multilevel linear model with the three fixed effects *precip*, *temp*, and *season* and all of their interactions along with the two random effects, *site* and *well.id*. We reduced each model by sequentially removing any insignificant interactions or variables. We also tried removing one or both random variables and running a linear model on the fixed effect variables alone, but both random effects were found to contribute significantly to the observed variation for all dependent variables. We used the Akaike Information Criterion (AIC) to confirm the best fitted model in each case.

The Q-Q and random effect range diagnostic plots for all of the models were quite similar. The Q-Q plots of the residuals indicated a departure from normality, suggesting that the residual distribution had thicker ends than a normal distribution (*Figures 8-12c*). This finding raises concerns about the suitability of our models which may lack important predictor variables. We also used outlierTest to test for potential outliers. The tests showed significant p-values for some monitoring wells, but removing them did not change our results or conclusion. Because we were not aware of any issue with the data collected from these monitoring wells, we chose not to exclude any data. The random effect range plots indicated that both *site* and *well.id* influence the observed variation, particularly the latter (*Figures 8-12d*).

RESULTS

Total Dissolved Solids

The model of best fit for TDS concentration includes precipitation, temperature, season, precipitation-season interaction, and temperature-season interaction as independent variables with sites and coal ash ponds as random effects. All independent variables except for those including spring are significant ($p < 0.05$, $R^2 = 0.85$). Precipitation decreases the effect of summer on TDS concentration ($p = 0.002$), while temperature increases the effect of summer on TDS concentration ($p = 0.002$). Precipitation and temperature both increase the effect of winter on TDS concentration ($p < 0.01$). Rising precipitation and temperature show the strongest effect in winter among these four seasons (*Figure 2*). Winter has the strongest effect, which is amplified by the interactions with precipitation and temperature. Almost all data points in the residual plot are randomly clustered around the centerline, showing no strong evidence for violation of linearity or homoscedasticity (Appendix, *Figure 8a*). However, some points demonstrate a pattern by following a downward sloping trendline. The scale-location plot shows that most data points are clustered around at the lower end of y axis, indicating the distribution of TDS data is slightly right skewed (Appendix, *Figure 8b*).

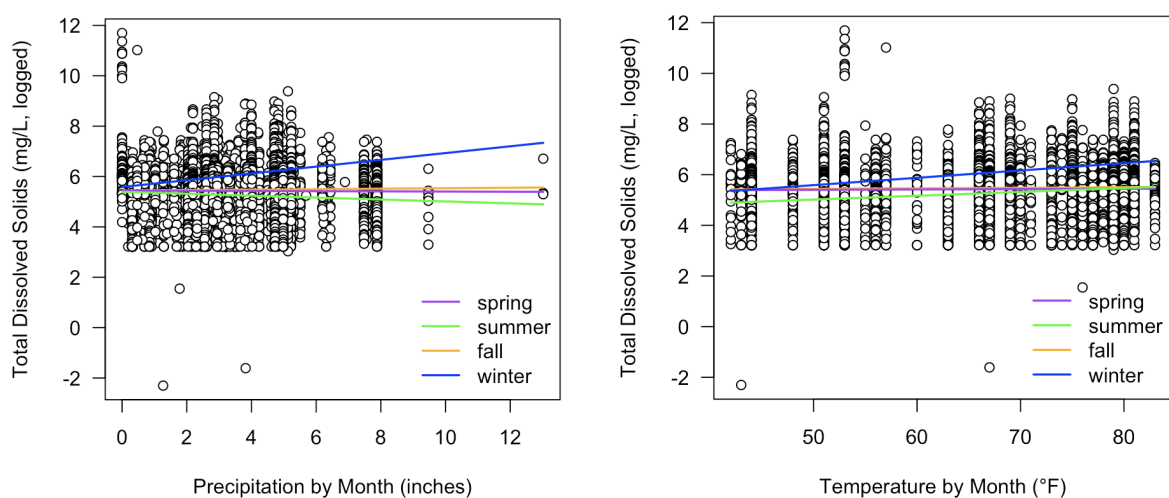


Figure 2a & b. Relationships between season and log(TDS concentrations) by precipitation (left) and temperature (right) in NC monitoring wells during 2016 and 2017.

Arsenic

The best-fit model for arsenic concentration includes season and temperature-season interaction as independent variables with sites and coal ash ponds as random effects. Season (summer) and temperature and summer interaction are significant ($p < 0.05$, $R^2 = 0.90$). Temperature increases the effect of summer on arsenic concentration ($p < 0.01$). The data points form two patterns in the residual plots by following two trends, with identical trendlines in both residue plot and scale-location plot, indicating there might be some pattern in these two groups of data points (Appendix, Figures 9a & b).

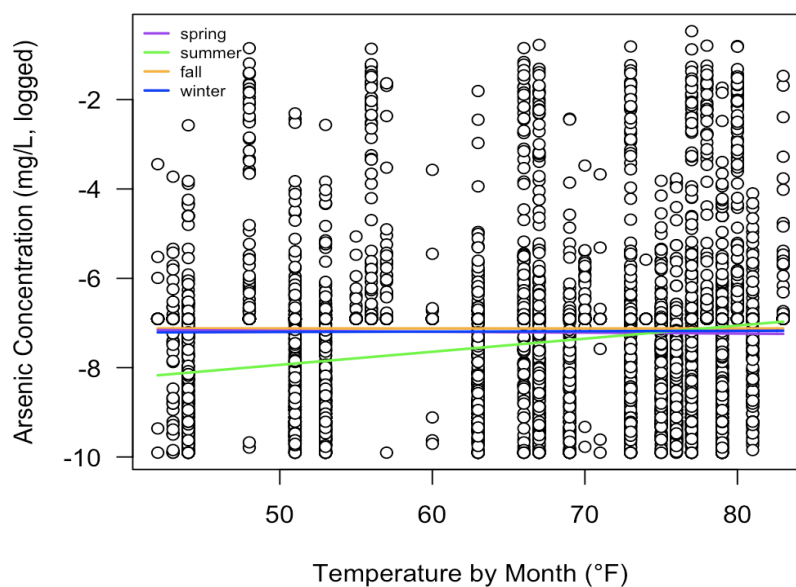


Figure 3. Relationships among temperature, season, and log(Arsenic concentrations) in NC monitoring wells during 2016 and 2017.

Calcium

The two-way interactions of temperature, precipitation, and season predict calcium concentrations ($p < 0.05$). This model also contains our two random effects, *well* and *site.id*. Increases in precipitation and temperature both predict reduced calcium, while their interaction amplifies their effect on calcium concentration ($p = 0.039$). Precipitation decreases the effects of spring ($p = 0.028$) and summer ($p < 0.001$) and increases the effect of winter ($p = 0.002$). Temperature increases the effects of summer ($p = 0.003$) and winter ($p = 0.003$). Spring and fall have similar effects. Winter has a bigger effect compared with the other seasons due to amplification from interactions with both precipitation and temperature. The residual versus fitted plot shows a consistent cloud concentrated around the regression line, giving no strong evidence of deviation from constant variance (Appendix, Figure 10a). The scale location plot does not show strong heteroscedasticity but is a bit unbalanced, so another transformation could help make our model more accurate (Appendix, Figure 10b). We noticed that one coal ash pond was responsible for all the highest concentrations of calcium, so we attempted exclusion of that pond in case it was driving our effect. Exclusion improves the AIC score, but does not noticeably alter the diagnostic plots nor remove all of the model's outliers. We did not have adequate reason to exclude the data, so we chose to retain the original model containing all coal ash ponds.

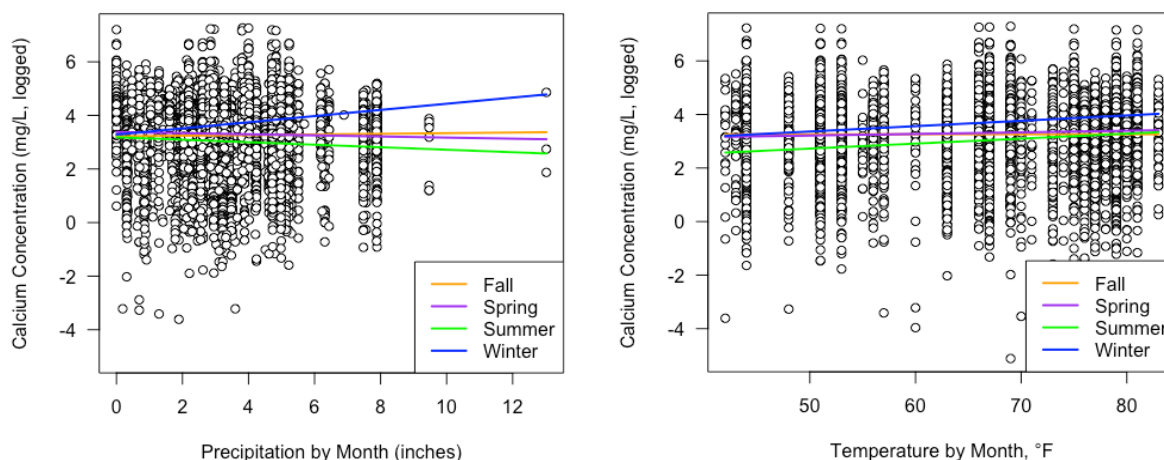


Figure 4a & b. Calcium concentration (log transformed) by precipitation (left) and temperature (right) overlaid with seasonal effects.

Chromium

For this model, we included our two random effects *site* and *well.id*, and both precipitation and the interaction between temperature and season predict concentrations of chromium. An increase in precipitation predicts a decrease in chromium levels ($p < 0.001$). Temperature increases the effect of summer ($p = 0.004$) and decreases the effect of spring ($p < 0.001$). The residual versus fitted plot does not demonstrate evidence of heteroscedasticity (Appendix, Figure 11a). The scale location plot confirms our impression of homoscedasticity, and all points are within the 2-point threshold, suggesting there are no excessively extreme values (Appendix, Figure 11b).

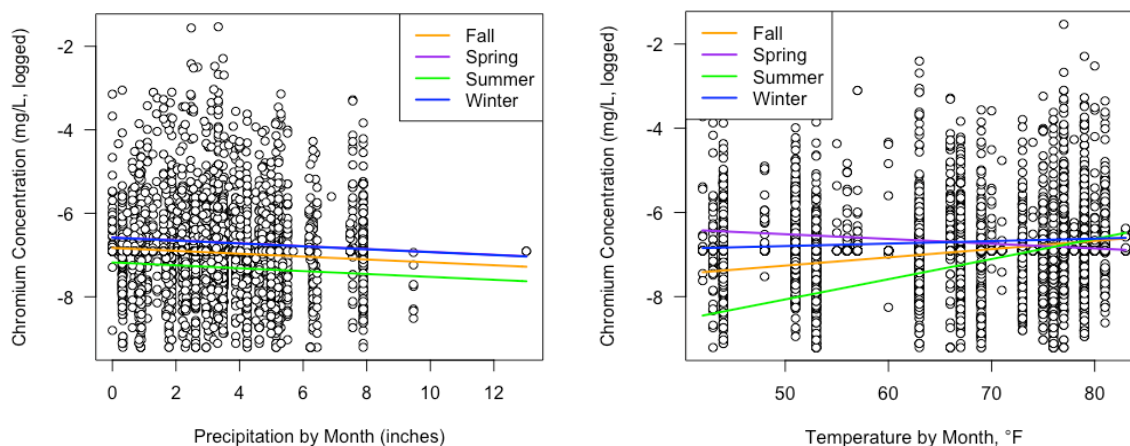


Figure 5a & b. Chromium concentration (log transformed) by precipitation (left) and temperature (right) overlaid with seasonal effects.

Lead

The most parsimonious and also significant model for lead concentration is a mixed-effects linear model with precipitation and temperature as the dependent variables and coal ash impoundments (*sites*) and monitoring wells (*well.id*) as random effects. Precipitation decreases the concentration of lead ($p < 0.001$). This could be due to dilution of aqueous lead with influxes of water from precipitation. Temperature increases the concentration of lead ($p < 0.001$). The fixed effects coefficients for the log concentration of lead are as follows: Intercept: -7.84, Precipitation: -0.024, Temperature: 0.00357. The scale location plot shows some patterning with a V-shaped pattern. The residuals versus fitted plot has a downward skew and the distribution of points above and below the zero line do not appear to have the same spread (Fig. 12). Although most of the points fall within the 2-point threshold (scale location plot), the distribution is likely not normal and is heteroscedastic.

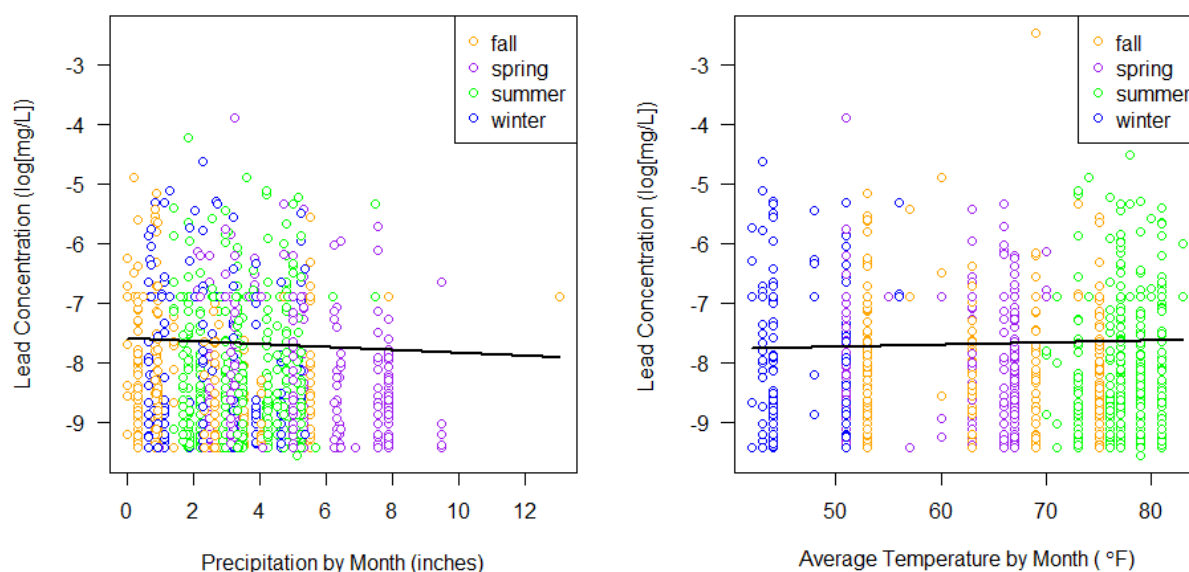


Figure 6a & b. Relationship between log transformed lead concentration and monthly precipitation (left) and average monthly temperature (right). Colored points show season, which did not have a significant effect on lead concentration.

DISCUSSION/CONCLUSION

We expected increasing precipitation could increase the total dissolved solids in the water because the increasing water flow could increase dissolution of chemicals. Meanwhile, we also anticipated that increasing precipitation could decrease chemical concentrations in monitoring wells due to the diluting effect of precipitation. For most of our other models in most seasons, concentrations did decrease with increased precipitation, suggesting that the dilution effect is stronger than any increased mobility or dissolution effect (*Figures 2a, 4a, 5a*). All of our model effects were complicated by interactions with season except for lead, for which increasing precipitation predicted decreased concentrations (*Figure 6a*). The negative effect of precipitation is most prominent in summer for total dissolved solids, calcium, and chromium (*Figures 2a, 4a, 5a*). Surprisingly, winter showed the opposite effect for calcium and total dissolved solids, having a positive interaction with precipitation (*Figures 2a, 4a*). Fluctuations due to freezing temperatures and occasional snow might drive this differing effect.

We also expected increasing temperature, which increases solubility, to increase individual chemical concentrations in the monitoring wells. Most of our models demonstrate a general positive effect of temperature on chemical concentrations, although these effects are driven by certain seasons. Lead and calcium show a general positive trend across seasons (*Figures 4b & 6b*). Arsenic concentrations increase with temperature in summer only, suggesting that the higher temperatures may be needed to increase arsenic solubility (*Figure 3*). Rising temperatures in summer also predicts increases in chromium concentrations to a greater extent than the other seasons (*Figure 5b*). Both summer and winter interact with temperature to cause rising concentrations of calcium (*Figure 4b*). Spring does not significantly interact with temperature in most models, likely because spring and fall tend to have milder weather patterns in comparison with summer and winter (*Figures 2b, 3, 4b, 6b*). Due to the more dramatic effects in summer and winter, we would recommend closer monitoring during those seasons.

The limitations of our analyses are visible in the lack of normality of the residuals (Appendix, *Figures 8-12c*). Our models are not the ideal fit for our data and probably lack explanatory variables that could explain some of the variation. Variations in contaminant concentrations as predicted by temperature and precipitation may be more directly tied to factors such as redox state or pH which, in turn, may be affected by temperature and precipitation. Since temperature and precipitation are already being monitored and forecasted, they could be used as an analog to other variables that may more directly influence contaminant concentration but which are less readily available. Season interacts naturally with weather, but seasonal effects could also be driven by variations in electricity consumption (increased coal ash deposition), which could be considered in a future analysis. Another limitation is that all of our dependent variables are extremely right skewed, and strong outliers such as spills or other specific events may drive some of our findings.

Our analysis of coal ash ponds in North Carolina suggests that weather can provide key indicators of the activation of potential contaminants. Season, precipitation, and temperature all predict trends of these contaminant concentrations and could be used to predict fluctuations in levels of other contaminants in water as well. Chemical concentrations should be monitored particularly closely in summer and winter. In addition to implications for future prediction of aquatic contaminants, our analysis also has significance in future aquatic monitoring near coal fired power plants. In the United States, coal ash is the second largest source of industrial waste (Goemann, 2015). Contamination of waterways continues to be a major concern as long as coal ash ponds are not properly lined. As temperatures warm and forecasts of climatic events become more extreme, the effects of weather on contaminant activation could become more pronounced. Hence, monitoring groundwater near coal-fired power plants is crucial to prevent leaked contaminants from coal ash ponds damaging environmental and human health. Coal fired power plants could also seek alternatives to manage their coal ash waste. For instance, they can sell their coal ash, which can be recycled into products such as wallboard, bricks and concrete (EPA, 2014; Rafieizonooz et

al, 2016). They can also sell coal ash to local farms since coal ash residues can provide necessary nutrients for plant growth and increase soil fertility (Kaur & Goyal, 2014). Such relatively low-risk alternatives could generate profits for coal-fired power plants and decrease contaminant escape to the surrounding environment from poor waste management practices.

Ecosystems faced by a multitude of human-caused stressors will be particularly susceptible to adverse effects from water pollution. The water reservoirs humans rely upon for drinking water and other uses are under threat from the dual forces of increased drought due to climate change and contamination from sources such as coal ash ponds. Understanding and monitoring the predictors of contaminant mobilization and leakage may help protect our waterways, but we must invest in alternatives to ensure clean water for ecosystems and humans alike in the future.

SOURCES

- Comment Submitted by Thomas Cmar, Deputy Managing Attorney, Coal Program, Earthjustice et al. 3/3/2020. Regulations.gov. <https://www.regulations.gov/document?D=EPA-HQ-OLEM-2019-0172-0174>
- Environmental Protection Agency. Hazardous and solid waste management system; identification and listing of special wastes; disposal of combustion residuals from electric utilities. 6/21/2010. Federal Register. <https://www.federalregister.gov/documents/2010/06/21/2010-12286/hazardous-and-solid-waste-management-system-identification-and-listing-of-special-wastes-disposal-of>
- Environmental Protection Agency. 2014. Coal Combustion Residual Beneficial Use Evaluation: Fly Ash Concrete and FGD Gypsum Wallboard https://www.epa.gov/sites/production/files/2014-12/documents/ccr_bu_eval.pdf
- Filippelli, G., S. Anenberg, M. Taylor, A. van Geen, and H. Khreis. 2020. New approaches to identifying and reducing the global burden of disease from pollution. *GeoHealth* 4(4). <https://doi.org/10.1029/2018GH000167>
- Gajia, G., L. Djurdjevia, O. Kostia, S. Jaria, M. Mitrovia, and P. Pavlovia. 2018. Ecological potential of plants for phytoremediation and ecorestoration of fly ash deposits and mine wastes. *Frontiers in Environmental Science*. <http://dx.doi.org.proxy.lib.duke.edu/10.3389/fenvs.2018.00124>
- Goemann, E. 2015. Surveying the threat of groundwater contamination from coal ash ponds. *Duke Environmental Law & Policy Forum* 25(2). https://link.gale.com/apps/doc/A439953365/ITOF?u=duke_perkins&sid=ITOF&xid=c13c99b1
- Gray, B. 2014. Coal Ash Clean-Up Moving Forward - Anyway. the Locker Room. <https://lockerroom.johnlocke.org/2014/08/13/coal-ash-clean-up-moving-forward-anyway/>
- Hall, K. 2019. Uncontainable threat: the nation's coal ash ponds. *Emory Law Journal* 69(1): 163-208. <https://login.proxy.lib.duke.edu/login?url=https://www-proquest-com.proxy.lib.duke.edu/docview/2331232545?accountid=10598>
- Kaur, R., and D. Goyal. 2014. Mineralogical comparison of coal fly ash with soil for use in agriculture. *Journal of Material Cycles and Waste Management* 18: 186-200. <https://doi-org.proxy.lib.duke.edu/10.1007/s10163-014-0323-1>
- Norms monthly location details. 2010. National Centers for Environmental Information. https://www.ncdc.noaa.gov/cdo-web/datasets/NORMAL_MLY/locations/FIPS:37/detail
- Preventing contamination of drinking water resources. 2017. United States Environmental Protection Agency. https://19january2017snapshot.epa.gov/www3/region1/eco/drinkwater/prevent_contamination.html
- Rafieizonooz, M., J. Mirza, M. R. Salim, M. W. Hussin, and E. Khankhaje. 2016. Investigation of coal bottom ash and fly ash in concrete as replacement for sand and cement. *Construction and Building Materials* 116. <http://dx.doi.org.proxy.lib.duke.edu/10.1016/j.conbuildmat.2016.04.080>
- timeanddate.com. 2016-2017. Past Weather, Asheville, Greensboro, and Greenville. <https://www.timeanddate.com/weather/>

- U.S. Geological Survey. 2016-2017. National Water Information System data available on the World Wide Web (USGS Water Data for the Nation). <https://waterdata.usgs.gov/nwis/>
- U.S. hourly precipitation data. 2018. National Centers for Environmental Information. <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00313>
- Wang, N., X. Sun, Q. Zhao, Y. Yang, P. Wang. 2020. Leachability and adverse effects of coal fly ash: A review. *Journal of Hazardous Materials* 396: 122725. <https://www.sciencedirect.com/science/article/pii/S0304389420307147>
- Well water and health. 2019. North Carolina Department of Health and Human Services. [https://epi.dph.ncdhhs.gov/oe/wellwater/figures.html#:~:text=Well%20Water%20and%20Health&text=About%202.4%20million%20North%20Carolinians,their%20primary%20drinking%20water%20source.&text=Despite%20the%20large%20number%20of,decade%20\(2000%2D2010\).](https://epi.dph.ncdhhs.gov/oe/wellwater/figures.html#:~:text=Well%20Water%20and%20Health&text=About%202.4%20million%20North%20Carolinians,their%20primary%20drinking%20water%20source.&text=Despite%20the%20large%20number%20of,decade%20(2000%2D2010).)

APPENDIX: Diagnostic Plots

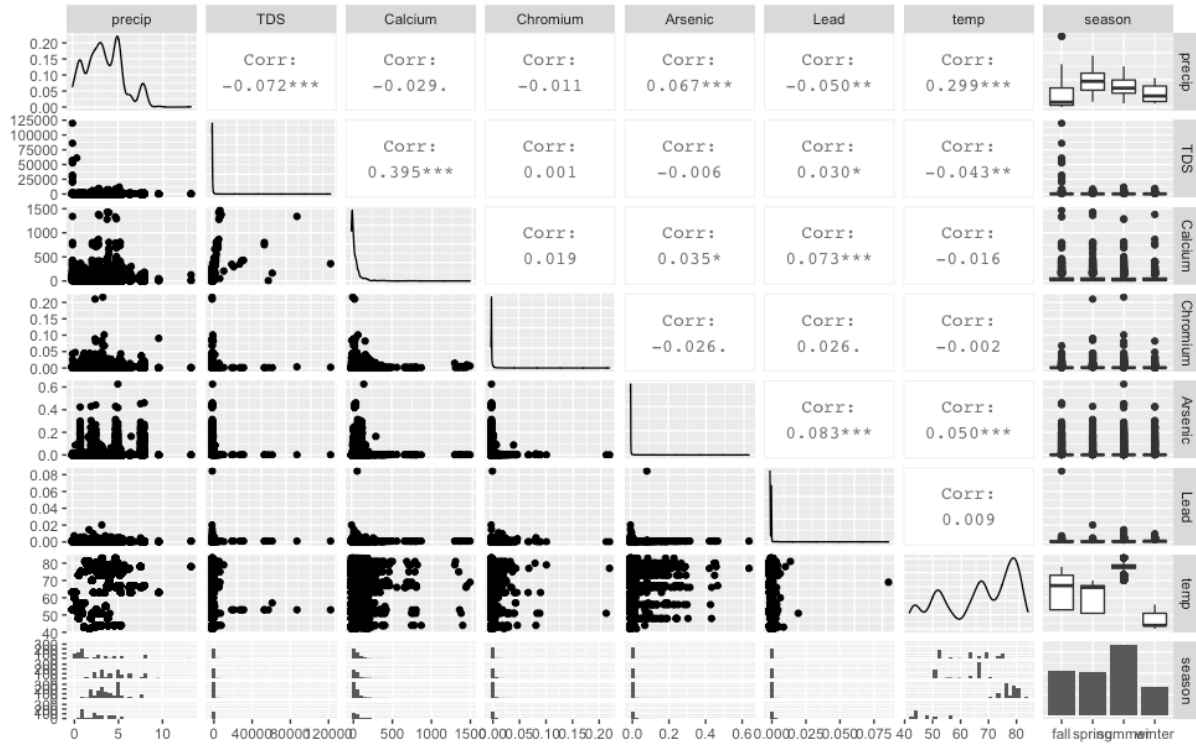


Figure 7. Ggpairs plot of the variables. Note the extreme right skew of all dependent variables.

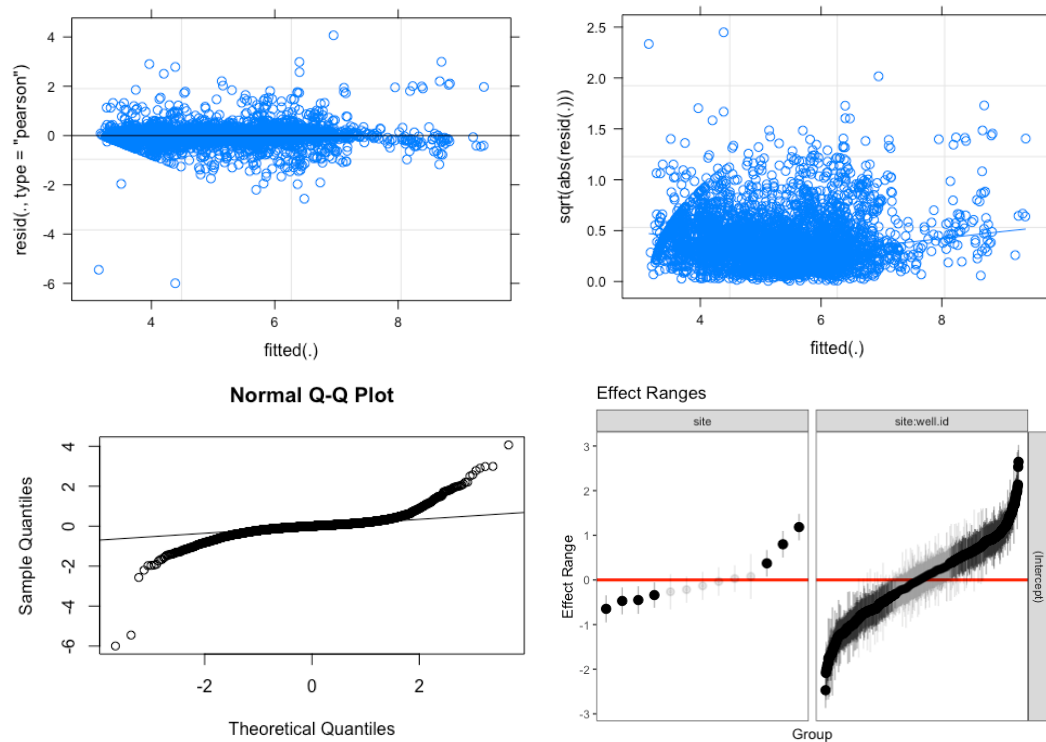


Figure 8a, b, c, d. Residuals vs fitted, Scale Location, Qqnorm of residuals, and effect ranges of random effects site and well.id for Model 1: Total Dissolved Solids.

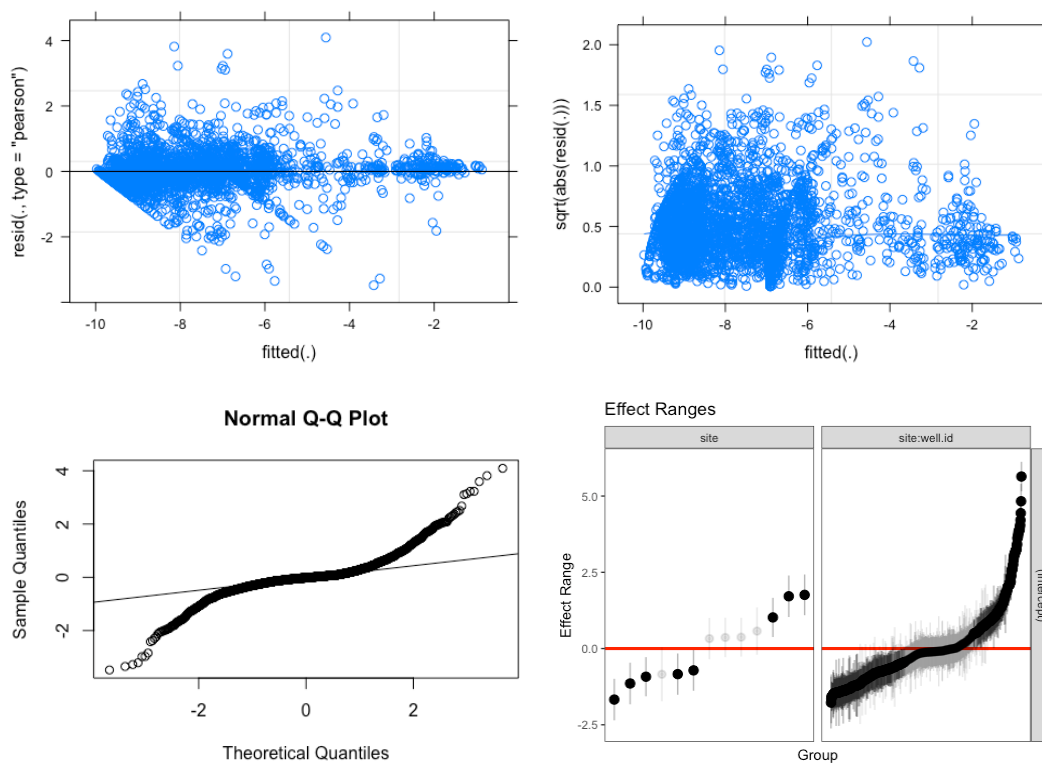


Figure 9a, b, c, d. Residuals vs fitted, Scale Location, Qqnorm of residuals, and effect ranges of random effects site and well.id for Model 2: Arsenic.

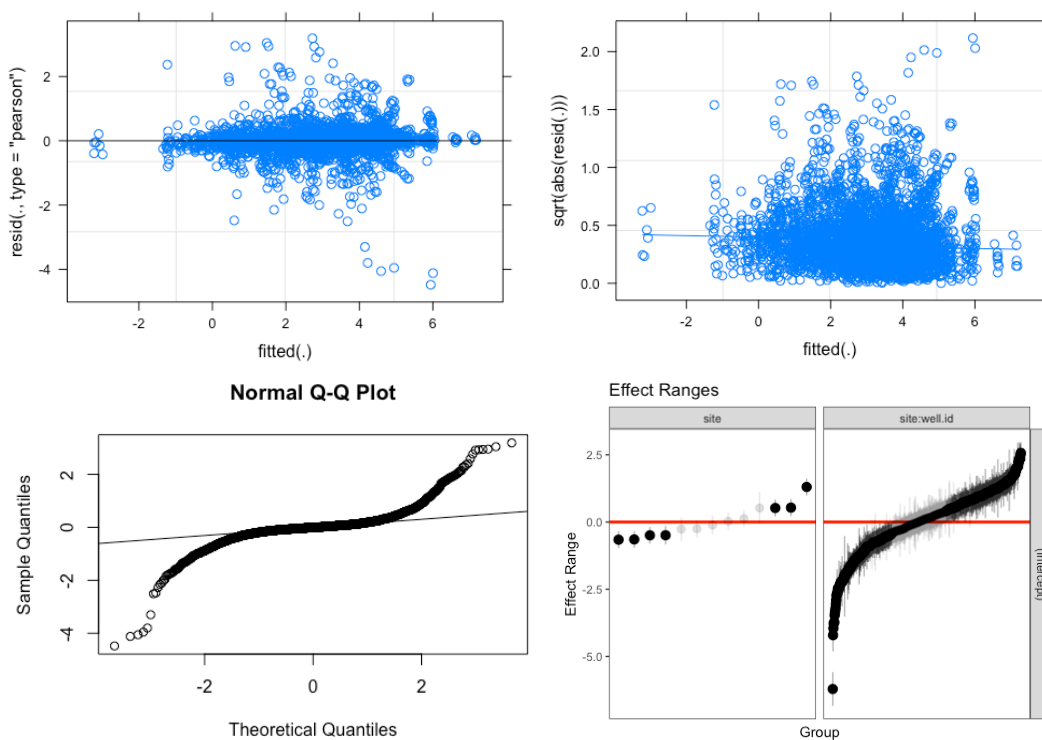


Figure 10a, b, c, d. Residuals vs fitted, Scale Location, Qqnorm of residuals, and effect ranges of random effects site and well.id for Model 3: Calcium.

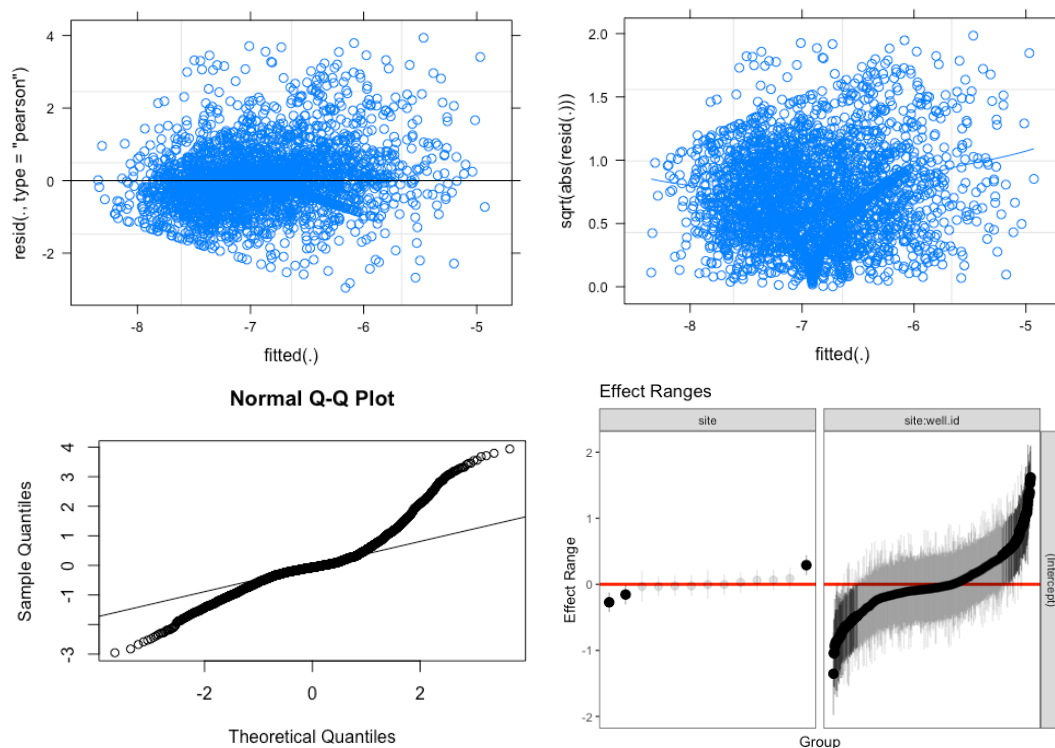


Figure 11a, b, c, d. Residuals vs fitted, Scale Location, Qqnorm of residuals, and effect ranges of random effects site and well.id for Model 4: Chromium.

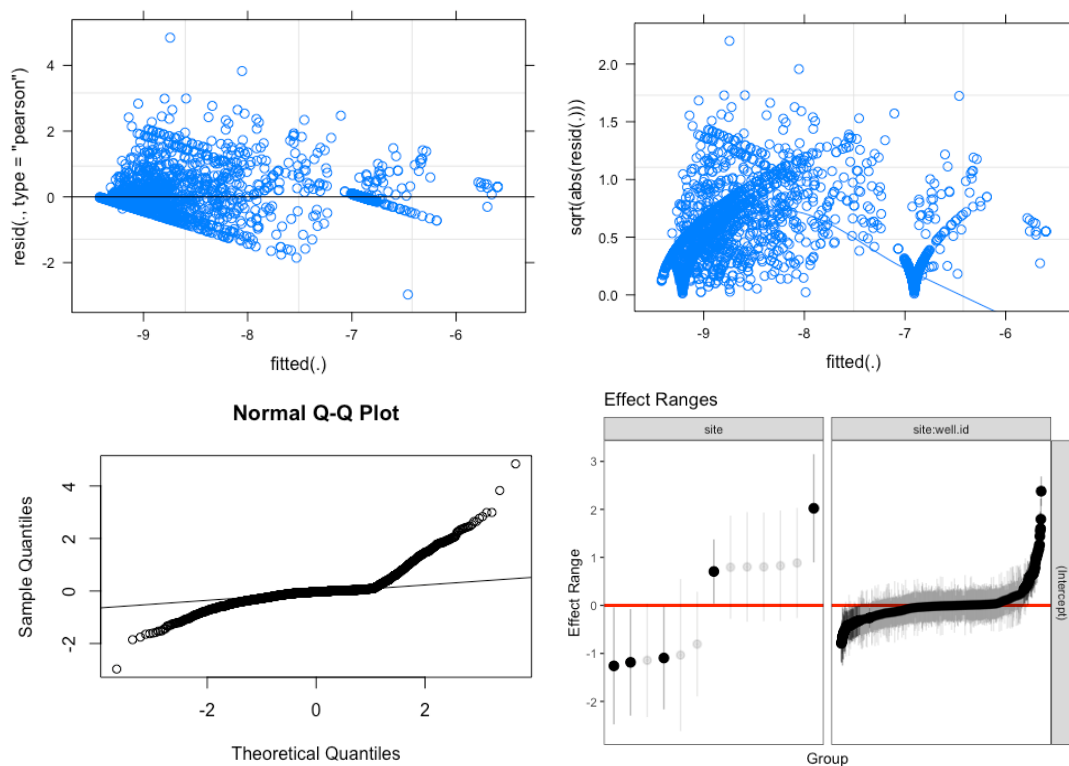


Figure 12a, b, c, d. Residuals vs fitted, Scale Location, Qqnorm of residuals, and effect ranges of random effects site and well.id for Model 5: Lead.

APPENDIX: Code

```
#-----
##### TDS #####
### TOTAL DISSOLVED SOLIDS ###

hist(wdat$TDS)

# change 0 values of DV to avoid errors w/log transform
wdat$TDS[wdat$TDS == 0] <- 0.1
hist(log(wdat$TDS))
# log transform tds to make it more normal/reduce heteroscedacity

# full model with total dissolved solids:
tdlm1 <- lmer(log(TDS) ~ precip*temp*season + (1|site) +
              (1|site:well.id), data=wdat)
summary(tdlm1)
# precip, temp, their interactions significant

# remove three way effect:
tdlm2 <- update(tdlm1, .~-precip:temp:season)
summary(tdlm2)
# looks good

# remove site random effect:
tdlm3 <- update(tdlm2, .~-(1|site))
summary(tdlm3)

# remove both random effects:
tdlm4 <- lm(log(TDS) ~ precip + temp + season + precip:temp + precip:season +
             temp:season, data=wdat)
summary(tdlm4)
# temp:season no longer significant

# remove temp:season
tdlm5 <- update(tdlm4, .~-temp:season)
summary(tdlm5)

# compare models:
AIC(tdlm1, tdlm2, tdlm3, tdlm4, tdlm5)
# best with the random effects (tdlm2)

# to see overall p values for interactions:
anova(tdlm2)
r.squaredGLMM(tdlm2)

# to see relationship between seasons:
summary(glht(tdlm2, linfct = mcp(season = "Tukey")),
        test = adjusted("holm"))
# this doesn't give interactions

plot(tdlm2)
outlierTest(tdlm2)
# outlier test suggests some strong outliers (p = 6.91e-48)
```



```

# I looked at the data and they're all from fall 2016 Roxboro Steam
# except one from WH. The median of TDS is 204, but the largest outlier
# has a value of 120,000. Could rerun analyses without the two
# worst outliers.

#check site's normality
qqnorm(coef(tdlm2)$site[,1])
qqline(coef(tdlm2)$site[,1])

#check site:well.id normality
# qqnorm(coef(tdlm2)$site:well.id)?Error in quantile(y, probs, names = FALSE, type =
# qtype, na.rm = TRUE) :
# object 'well.id' not found
# qqline(coef(tdlm2)$site:well.id)

#check overall model residuals normality
qqnorm(resid(tdlm2))
qqline(resid(tdlm2))

#check effect of each site and well
plotREsim(REsim(tdlm2))

# Graph!
cf <- fixef(tdlm2)
with(wdat, plot(precip, log(TDS), pch=21, bg="white",
               xlab = "Precipitation by Month (inches)",
               ylab = "Total Dissolved Solids Concentration (ml/L, logged)", las = 1))
x <- with(na.omit(wdat), seq(min(precip), max(precip), length=4457))
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[7]*x*mean(wdat$temp),
      add=T, col="orange") # fall
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[4] +
      cf[7]*x*mean(wdat$temp) + cf[8]*x + cf[11]*mean(wdat$temp),
      add=T, col="purple") # spring
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[5] +
      cf[7]*x*mean(wdat$temp) + cf[9]*x + cf[12]*mean(wdat$temp),
      add=T, col="green") # summer
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[6] +
      cf[7]*x*mean(wdat$temp) + cf[10]*x + cf[13]*mean(wdat$temp),
      add=T, col="blue") # winter
legend("bottomright", c("spring", "summer", "fall", "winter"), lty = 1, col =
c("purple", "green", "orange", "blue"),
      lwd = 2, bty = "n")

with(wdat, plot(temp, log(TDS), pch=21, bg="white",
               xlab = "Temperature by Month (°F)",
               ylab = "Total Dissolved Solids Concentration (ml/L, logged)", las = 1))
x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[2]*mean(wdat$precip, na.rm = TRUE) + cf[3]*x +
      cf[7]*mean(wdat$precip, na.rm = TRUE)*x,
      add=T, col="orange") # fall
curve(cf[1] + cf[2]*mean(wdat$precip, na.rm = TRUE) + cf[3]*x + cf[4] +
      cf[7]*mean(wdat$precip, na.rm = TRUE)*x + cf[8]*mean(wdat$precip, na.rm = TRUE) +
      cf[11]*x,
      add=T, col="purple") # spring
curve(cf[1] + cf[2]*mean(wdat$precip, na.rm = TRUE) + cf[3]*x + cf[5] +

```

```

    cf[7]*mean(wdat$precip,na.rm =TRUE)*x + cf[9]*mean(wdat$precip,na.rm =TRUE) +
cf[12]*x,
    add=T, col="green") # summer
curve(cf[1] + cf[2]*mean(wdat$precip,na.rm =TRUE) + cf[3]*x + cf[6] +
    cf[7]*mean(wdat$precip,na.rm =TRUE)*x + cf[10]*mean(wdat$precip,na.rm =TRUE) +
cf[13]*x,
    add=T, col="blue") # winter
legend("bottomright", c("spring", "summer", "fall", "winter"), lty = 1, col =
c("purple", "green", "orange", "blue"),
    lwd = 2, bty = "n")

#-----
##### ARSENIC #####

hist(wdat$Arsenic)
hist(log(wdat$Arsenic))
# log transform arsenic to make it more normal/reduce heteroscedacity

arlm1 <- lmer(log(Arsenic) ~ precip*temp*season + (1|site) +
    (1|site:well.id), data=wdat)
summary(arlm1)
# nothing is significant

arlm2 <- update(arlm1, .~-precip:temp:season)
summary(arlm2)
# season and temp:season are significant

arlm3 <- update(arlm2, .~-precip:temp)
summary(arlm3)
# season and temp:season are significant

arlm4 <- update(arlm3, .~-precip:season)
summary(arlm4)
# season and temp:season are significant

arlm5 <- update(arlm4, .~-precip)
summary(arlm5)
# looks good
# however, this model has a larger sample size (precip contained some NAs)

# remove NA data to make this model comparable with the others:
arlm6 <- lmer(log(Arsenic) ~ temp*season+ (1|site) +
    (1|site:well.id), data=wdat[!is.na(wdat$precip),])
summary(arlm6)
r.squaredGLMM(arlm6)

arlm7 <- update(arlm6, .~-temp)
summary(arlm7)

AIC(arlm1, arlm2, arlm3, arlm4, arlm5, arlm6,arlm7) # arlm7 is best
# arlm5 is a larger sample size so gives a higher AIC and won't
# work with lrtest
lrtest(arlm1, arlm2, arlm3, arlm4, arlm6)

# Overall p value for interaction:

```

```

anova(arlm7)

# Relationship between seasons:
summary(glht(arlm6, linfct = mcp(season = "Tukey")),
  test = adjusted("holm"))
# this doesn't give interactions though

plot(arlm7)

outlierTest(arlm7)
# outlier test suggests some strong outliers, but we don't have good
# justification to remove them

qqnorm(resid(arlm7))
qqline(resid(arlm7))

# Graph! temp vs. Ar
cf <- fixef(arlm7)
with(wdat, plot(temp, log(Arsenic), pch=21, bg="white",
  xlab = "Temperature by Month (°F)",
  ylab = "Arsenic Concentration Concentration (ml/L, logged)", las = 1))
x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[5]*x, add=T, col="orange") # fall
curve(cf[1] + cf[2] + cf[6]*x, add=T, col="purple") # spring
curve(cf[1] + cf[3] + cf[7]*x, add=T, col="green") # summer
curve(cf[1] + cf[4] + cf[8]*x, add=T, col="blue") # winter
legend("topleft", c("spring", "summer", "fall", "winter"), lty = 1, col = c("purple",
"green", "orange", "blue"),
  lwd = 2, bty = "n", cex=0.65)
# (note that I redefined cf for each plot, so make sure to run that
# first line every time)

cf <- fixef(arlm6)
with(wdat, plot(temp, log(Arsenic), pch=21, bg="white",
  xlab = "Temperature by Month (°F)",
  ylab = "Arsenic Concentration Concentration (ml/L, logged)", las = 1))
x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[2]*x, add=T, col="orange") # fall
curve(cf[1] + cf[2]*x + cf[3] + cf[6]*x, add=T, col="purple") # spring
curve(cf[1] + cf[2]*x + cf[4] + cf[7]*x, add=T, col="green") # summer
curve(cf[1] + cf[2]*x + cf[5] + cf[8]*x, add=T, col="blue") # winter
legend("topleft", c("spring", "summer", "fall", "winter"), lty = 1, col = c("purple",
"green", "orange", "blue"),
  lwd = 2, bty = "n", cex=0.65)

#-----
##### CALCIUM #####

par(mfrow=c(1,1))
hist(wdat$Calcium)
# very right skewed!
hist(log(wdat$Calcium))
# now a little left skewed but much better

with(wdat, plot(precip, log(Calcium)))

```

```

with(wdat, plot(temp, log(Calcium)))
# log transform calcium to make it more normal

# full model with calcium:
calm1 <- lmer(log(Calcium) ~ precip*temp*season + (1|site) +
              (1|site:well.id), data=wdat)
summary(calm1)
# temp and season (winter) are significant

# remove three way interaction
calm2 <- update(calm1, .~.-precip:temp:season)
summary(calm2)
# most things are significant!

# remove site random effect
calm3 <- update(calm2, .~.-(1|site))
summary(calm3)

# remove both random effects
calm4 <- lm(log(Calcium) ~ precip + temp + season + precip:temp + precip:season +
            temp:season, data=wdat)
summary(calm4)
# temp:season no longer significant

# remove temp:season from model w/o random effects
calm5 <- update(calm4, .~.-temp:season)
summary(calm5)

# try removing other interactions from model with random effects
calm6 <- update(calm3, .~.-precip:temp)
summary(calm6)

# try removing interactions
calm7 <- update(calm6, .~.-precip:season)
summary(calm7)
# not good, precip:season was important

AIC(calm1, calm2, calm3, calm4, calm5, calm6) # keep calm2

# to see overall p values for interactions:
anova(calm2)

# to see relationship between seasons:
summary(glht(calm2, linfct = mcp(season = "Tukey")),
        test = adjusted("holm"))
# this doesn't give interactions though

# qqplot of residuals:
qqnorm(resid(calm2))
qqline(resid(calm2))
# not great

# random variable effect ranges:
plotREsim(REsim(calm2))
# both random variables have an effect

```

```

# scale location
plot(calm2, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))

# fitted vs residuals
plot(calm2)

outlierTest(calm2)
# outlier test suggests some strong outliers, but we don't have good
# justification to remove them

cf <- fixef(calm2)
with(wdat, plot(precip, log(Calcium), pch=21, bg="white",
  xlab = "Precipitation by Month (inches)",
  ylab = "Calcium Concentration (mg/L, logged)", las = 1))
x <- with(na.omit(wdat), seq(min(precip), max(precip), length=4457))
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[7]*x*mean(wdat$temp),
  add=T, col="orange", lwd = 2) # fall
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[4] +
  cf[7]*x*mean(wdat$temp) + cf[8]*x + cf[11]*mean(wdat$temp),
  add=T, col="purple", lwd = 2) # spring
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[5] +
  cf[7]*x*mean(wdat$temp) + cf[9]*x + cf[12]*mean(wdat$temp),
  add=T, col="green", lwd = 2) # summer
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[6] +
  cf[7]*x*mean(wdat$temp) + cf[10]*x + cf[13]*mean(wdat$temp),
  add=T, col="blue", lwd = 2) # winter
Seas <- c("Fall", "Spring", "Summer", "Winter")
Cols <- c("orange", "purple", "green", "blue")
legend("bottomright", Seas, lty=1, col=Cols, lwd = 2)

cf <- fixef(calm2)
with(wdat, plot(temp, log(Calcium), pch=21, bg="white",
  xlab = "Temperature by Month, °F",
  ylab = "Calcium Concentration (mg/L, logged)", las = 1))
x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x +
  cf[7]*x*mean(na.omit(wdat$precip)), add=T, col="orange", lwd = 2) # fall
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[4] +
  cf[7]*x*mean(na.omit(wdat$precip)) + cf[8]*mean(na.omit(wdat$precip)) +
  cf[11]*x, add=T, col="purple", lwd = 2) # spring
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[5] +
  cf[7]*x*mean(na.omit(wdat$precip)) + cf[9]*mean(na.omit(wdat$precip)) +
  cf[12]*x, add=T, col="green", lwd = 2) # summer
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[6] +
  cf[7]*x*mean(na.omit(wdat$precip)) + cf[10]*mean(na.omit(wdat$precip)) +
  cf[13]*x, add=T, col="blue", lwd = 2) # winter
Seas <- c("Fall", "Spring", "Summer", "Winter")
Cols <- c("orange", "purple", "green", "blue")
legend("bottomright", Seas, lty=1, lwd = 2, col=Cols)

##### TRY REMOVING WELL WITH HIGHEST VALUES#####
# Remove outlier well:
calm11 <- lmer(log(Calcium) ~ precip*temp*season + (1|site) +

```



```

      (1|site:well.id), data=wdat[!wdat$well.id == "CCR-208BR",])
summary(calm11)

calm12 <- update(calm11, .~.-precip:temp:season)
summary(calm12)
# most things are significant!

# remove site random effect
calm13 <- update(calm12, .~.-(1|site))
summary(calm13)

# remove both random effects
calm14 <- lm(log(Calcium) ~ precip + temp + season + precip:temp + precip:season +
             temp:season, data=wdat[!wdat$well.id == "CCR-208BR",])
summary(calm14)
# temp:season no longer significant

# remove temp:season from model w/o random effects
calm15 <- update(calm14, .~.-temp:season)
summary(calm15)

calm16 <- update(calm13, .~.-precip:temp)
summary(calm16)

AIC(calm1, calm2, calm11, calm12, calm13, calm14, calm15, calm16) #calm12 is best

qqnorm(resid(calm12))
qqline(resid(calm12))
# no better than calm2

plot(calm12)
outlierTest(calm12)
# Removing the wells with the highest Calcium concentrations does not improve
# the diagnostics.
# Stick with calm2

#-----
##### CHROMIUM #####

par(mfrow=c(1,1))
hist(wdat$Chromium)
# very right skewed
hist(log(wdat$Chromium))
# better, looks leptokurtic

with(wdat, plot(precip, log(Chromium)))
with(wdat, plot(temp, log(Chromium)))
# log transform chromium to make it more normal/reduce heteroscedacity

# full model with chromium:
ch1 <- lmer(log(Chromium) ~ precip*temp*season + (1|site) +
            (1|site:well.id), data=wdat)
summary(ch1)
# temp, season, and their interaction are significant

```

```

ch2 <- update(ch1, .~-precip:temp:season)
summary(ch2)
# everything significant except precip and precip:temp

ch3 <- update(ch2, .~-precip:temp)
summary(ch3)
# looks good

# try further reducing precip since it is not significant:
ch4 <- lmer(log(Chromium) ~ precip + temp*season + (1|site) +
            (1 | site:well.id), data=wdat)
summary(ch4)

# try entirely removing precip:
ch5 <- lmer(log(Chromium) ~ temp*season + (1|site) +
            (1|site:well.id), data=wdat[!is.na(wdat$precip),])
summary(ch5)

# remove site random effect from chl3:
ch6 <- update(ch3, .~-(1|site))
summary(ch6)

ch7 <- update(ch4, .~-(1|site))
summary(ch7)

# try removing all interactions:
ch8 <- update(ch4, .~-temp:season)
summary(ch8)

AIC(ch1, ch2, ch3, ch4, ch5, ch6, ch7, ch8)
# ch4 is best
# site is important

# to see overall p values for interactions:
anova(ch4)

# to see relationship between seasons:
summary(glht(ch4, linfct = mcp(season = "Tukey")),
        test = adjusted("holm"))
# this doesn't give interactions though, so cannot be interpreted as is

# scale-location plot:
plot(ch4, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))

# fitted vs residuals:
plot(ch4)

outlierTest(ch4)
# outlier test suggests some strong outliers, but we don't have good
# justification to remove them

qqnorm(resid(ch4))
qqline(resid(ch4))

##### Log transform of Precip

```

```

wdat$precip[wdat$precip == 0] <- 0.001
wdat$lpr <- with(wdat, log(precip))

# full model with chromium:
ch11 <- lmer(log(Chromium) ~ lpr*temp*season + (1|site) +
            (1|site:well.id), data=wdat)
summary(ch11)
# precip, temp, season, and their interaction are significant

ch12 <- update(ch11, .~-lpr:temp:season)
summary(ch12)
# everything significant, lpr:temp is marginal

ch13 <- update(ch12, .~-lpr:temp)
summary(ch13)
# looks good

# try further reducing precip since it is not significant:
ch14 <- update(ch13, .~-(1|site))
summary(ch14)

AIC(ch1, ch2, ch3, ch4, ch5, ch6, ch7, ch11, ch12, ch13, ch14)
# ch13 is best... but not sure if it's sufficiently better than ch4 to warrant
# this extra transformation

qqnorm(resid(ch13))
qqline(resid(ch13))
# doesn't look like a huge improvement

plot(ch13)

# Graph: Precipitation x Chromium
cf <- fixef(ch4)
with(wdat, plot(precip, log(Chromium), pch=21, bg="white",
               xlab = "Precipitation by Month (inches)",
               ylab = "Chromium Concentration (mg/L, logged)", las = 1))
x <- with(na.omit(wdat), seq(min(precip), max(precip), length=4457))
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp), add=T, col="orange", lwd = 2)
# fall
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[4] + cf[7]*mean(wdat$temp),
      add=T, col="purple", lwd = 2) # spring
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[5] + cf[8]*mean(wdat$temp),
      add=T, col="green", lwd = 2) # summer
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp) + cf[6] + cf[9]*mean(wdat$temp),
      add=T, col="blue", lwd = 2) # winter
Seas <- c("Fall", "Spring", "Summer", "Winter")
Cols <- c("orange", "purple", "green", "blue")
legend("topright", Seas, lty=1, col=Cols, lwd = 2)

# Graph: Temperature x Chromium
cf <- fixef(ch4)
with(wdat, plot(temp, log(Chromium), pch=21, bg="white",
               xlab = "Temperature by Month, °F",
               ylab = "Chromium Concentration (mg/L, logged)", las = 1))

```

```

x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x, add=T, col="orange",
      lwd = 2) # fall
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[4] + cf[7]*x,
      add=T, col="purple", lwd = 2) # spring
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[5] + cf[8]*x,
      add=T, col="green", lwd = 2) # summer
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x + cf[6] + cf[9]*x,
      add=T, col="blue", lwd = 2) # winter
Seas <- c("Fall", "Spring", "Summer", "Winter")
Cols <- c("orange", "purple", "green", "blue")
legend("topleft", Seas, lty=1, col=Cols, lwd = 2)

#-----
##### LEAD #####
wdat$log_Lead <- log(wdat$Lead)

ggpairs(wdat[,5:9])

hist(wdat$Lead)
hist(log(wdat$Lead))
#hist(1/sqrt(wdat$Lead))
with(wdat, plot(precip, log(Lead)))
with(wdat, plot(temp, log(Lead)))
# log transform Lead to make it more normal, though it still
# doesn't look very good: bimodal
ggpairs(wdat[,5:15])

# full model with all interactions
lelm1 <- lmer(log(Lead) ~ precip*temp*season + (1|site) +
              (1|site:well.id), data=wdat)
# season (summer) and its interactions are significant

lelm2 <- update(lelm1, .~.-precip:temp:season)
# same result

lelm3 <- update(lelm2, .~.-precip:temp)
# same result; precip:season is now only marginal

lelm4 <- update(lelm3, .~.-precip:season)
# everything left is significant

# removing more to check for a more parsimonious model
lelm5 <- update(lelm4, .~.-season)
# none of the season interactions are significant

lelm6 <- update(lelm5, .~.-temp:season)
summary(lelm6)
# all values that are left are significant and this model looks good

# just a final check on parsimony if using only precip or temp is better
lelm6.2 <- update(lelm6, .~.-temp)

lelm6.3 <- update(lelm6, .~.-precip)

```

both of the lelm6.2 and 6.3 are all significant but an AIC of lelm6, lelm6.2, and lelm6.3 shows that lelm6 is still better

```
AIC(lelm1, lelm2, lelm3, lelm4, lelm5, lelm6, lelm6.2, lelm6.3)
# lelm6 is more parsimonious and has less factors with a lower AIC, otherwise lelm6.2
would be better with only precipitation
```

```
# to see overall p values for interactions:
anova(lelm6)
```

```
plot(lelm6)
outlierTest(lelm6)
# no good justification to remove outliers
```

```
#check site's normality
qqnorm(coef(lelm6)$site[,1])
qqline(coef(lelm6)$site[,1])
```

```
#check overall model residuals normality
qqnorm(resid(lelm6))
qqline(resid(lelm6))
```

```
#check effect of each site and well
# plotREsim(RESim(lelm6))
```

Graphs =====

```
# Custom color palette
jcols <- with(wdat, data.frame(season = levels(season), color = c("orange", "purple",
"green", "blue"), name = 'scols'))
```

```
cf <- fixef(lelm6)
par(mfrow=c(1,2))
```

```
with(wdat, plot(precip, log(Lead), pch=21, col = jcols$color[match(wdat$season,
jcols$season)], bg="white",
      xlab = "Precipitation by Month (inches)",
      ylab = "Lead Concentration (log[mg/L])", las = 1))
x <- with(na.omit(wdat), seq(min(precip), max(precip), length=4457))
curve(cf[1] + cf[2]*x + cf[3]*mean(wdat$temp), add=T, col="black", lwd = 2)
legend("topright", levels(wdat$season), col = jcols$color[1:4], pch=21, bty = "o")
```

```
with(wdat, plot(temp, log(Lead), pch=21, col = jcols$color[match(wdat$season,
jcols$season)], bg="white",
      xlab = expression("Average Temperature by Month (~degree*F)"),
      ylab = "Lead Concentration (log[mg/L]) ", las = 1))
x <- with(wdat, seq(min(temp), max(temp), length=4457))
curve(cf[1] + cf[2]*mean(na.omit(wdat$precip)) + cf[3]*x, add=T, col="black", lwd=2)
legend("topright", levels(wdat$season), col = jcols$color[1:4], pch=21, bty = "o")
```

```
### FIGURE 7 ###
ggpairs(wdat[,6:13])
```



```

### FIGURE 8 ###
plot(tdlm2)
plot(tdlm2, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
qqnorm(resid(tdlm2))
qqline(resid(tdlm2))
plotRESim(RESim(tdlm2))

### FIGURE 9 ###
plot(arlm7)
plot(arlm7, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
qqnorm(resid(arlm7))
qqline(resid(arlm7))
plotRESim(RESim(arlm7))

### FIGURE 10 ###
plot(calm2)
plot(calm2, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
qqnorm(resid(calm2))
qqline(resid(calm2))
plotRESim(RESim(calm2))

### FIGURE 11
plot(ch4)
plot(ch4, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
qqnorm(resid(ch4))
qqline(resid(ch4))
plotRESim(RESim(ch4))

### FIGURE 12
plot(lelm6)
plot(lelm6, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
qqnorm(resid(lelm6))
qqline(resid(lelm6))
plotRESim(RESim(lelm6))

```