persistencePits v7.0

Matthew E. Verbyla

6/16/2022

## Load packages

First, load the required packages.

library(ggplot2)  
library(gridExtra)  
library("PerformanceAnalytics")  
library("Hmisc")  
library("knitr")  
library("bbmle")

## Load and explore data

Now, we will load and explore the persistence data, which are stored on the GWPP K2P Data Portal under the following dataset: <http://data.waterpathogens.org/dataset/persistence-treatment>.

The next chunk of code explores the data set to quantify the total number of data points, experiments, and the breakdown of lab-based studies, field-based studies, and the number of experiments done with different types of pathogen groups.

nrow(persist) # total number of data points

## [1] 1382

length(unique(persist$experiment\_id)) # total number of experiments

## [1] 243

length(persist[persist$location\_comment=="Laboratory-based study",]$experiment\_id) # number of data points from lab-based studies

## [1] 1078

length(unique(persist[persist$location\_comment=="Laboratory-based study",]$experiment\_id)) # number of lab-based experiments

## [1] 190

length(persist[persist$location\_comment!="Laboratory-based study",]$experiment\_id) # number of data points from field-based studies

## [1] 304

length(unique(persist[persist$location\_comment!="Laboratory-based study",]$experiment\_id)) # number of field-based experiments

## [1] 53

length(unique(persist[persist$microbial\_group=="Virus",]$experiment\_id)) # number of experiments with viruses

## [1] 32

length(unique(persist[persist$microbial\_group=="Bacteria",]$experiment\_id)) # number of experiments with bacteria

## [1] 147

length(unique(persist[persist$microbial\_group=="Protozoa",]$experiment\_id)) # number of experiments with protozoa

## [1] 8

length(unique(persist[persist$microbial\_group=="Helminth",]$experiment\_id)) # number of experiments with helminths

## [1] 56

unique(persist[persist$microbial\_group=="Virus",]$microorganism\_common\_name) # types of viruses studied

## [1] "Bacteriophage φX" "Bacteriophage MS2"   
## [3] "Salmonella typhimurium phages" "MS2 coliphage"   
## [5] "PhiX174 coliphage" "T4 coliphage"   
## [7] "Adenovirus" "Salmonella typhimurium phage 28B"  
## [9] "Rotavirus"

unique(persist[persist$microbial\_group=="Bacteria",]$microorganism\_common\_name) # types of bacteria studied

## [1] "Fecal Coliforms" "Enterococcus faecalis"   
## [3] "Salmonella enterica Typhimurium" "Escherichia Coli"   
## [5] "Fecal coliforms" "Streptococcus"   
## [7] "Salmonella" "Total coliform"   
## [9] "Eschericia Coli" "Sulfur reducing clostridia"   
## [11] "Escherichia coli" "Enterococcus"   
## [13] "Salmonella spp" "Thermotolerant coliform"   
## [15] "Enterococci" "Salmonella senftenberg"   
## [17] "Total coliforms" "Salmonella spp."   
## [19] "Shigella sonnei" "Shigella flexneri"   
## [21] "Escherichia coli "

unique(persist[persist$microbial\_group=="Protozoa",]$microorganism\_common\_name) # types of protozoa studied

## [1] "Entamoeba cysts" "Cryptosporidium" "Giardia"

unique(persist[persist$microbial\_group=="Helminth",]$microorganism\_common\_name) # types of helminths studied

## [1] "Ascaris" "Ascaris suum" "Ascaris eggs"   
## [4] "Ascaris ova" "Ascaris lumbricoides"

## Calculate the decay rate coefficients

The next chunk of code sets up some variables that will be used in an upcoming "FOR" loop, which will be used to calculate the decay rate coefficients and other statistics associated with the fit of the log-linear and JM2 models.

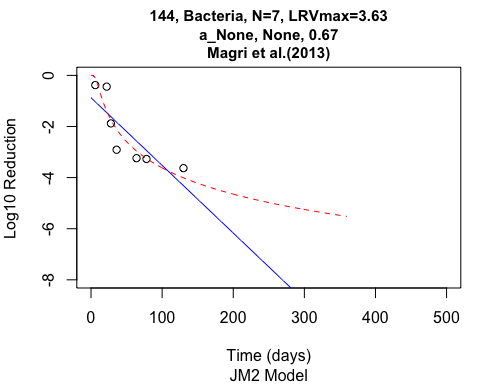
N<-length(unique(persist$experiment\_id)) # the total number of experiments  
M<-length(persist$experiment\_id) # the total number of data points  
persist$ind<-NA  
for(i in 1:M){  
 # this creates a unique index variable to distinguish each independent experiment  
 persist$ind[i]<-which(data.frame(unique(persist$experiment\_id))==persist$experiment\_id[i])  
}  
# here we create lots of NULL variables, which later we will populate values for in the upcoming "for" loop  
k<-NULL;k1<-NULL;k2<-NULL;intrcpt<-NULL;group<-NULL;addit<-NULL;urine<-NULL;urea<-NULL;moist<-NULL;temp<-NULL;pH<-NULL;r2<-NULL;t99<-NULL;t99jm<-NULL;authors<-NULL;num<-NULL;aicLL<-NULL;aicJM<-NULL;experiment<-NULL;maxLRV<-NULL;maxTime<-NULL;mcn<-NULL;best<-NULL;sterr<-NULL;sigme<-NULL  
par(mfrow=c(3,6)) #this creates the panel layout for the plots used for creating Figure S1

for(i in 1:N){ #in this loop, we calculate the decay rate coefficients for the log-linear and JM2 models  
 time<-persist[persist$ind==i,]$time\_days # get the time for the present experiment  
 lnrv<-persist[persist$ind==i,]$ln\_removal # get the ln reduction for the present experiment  
 log10\_reduction<-persist[persist$ind==i,]$log10\_reduction # get the log10 reduction for the present experiment  
 # since we calculated the ln reduction, then equation gets algebraically rearranged like this:  
 # Ct = Co\*exp(-k\*t)   
 # ln(Ct/Co) = -k\*t  
   
 # first, we fit the model like this to express the decay rate coefficient, k, on a LN scale  
 fit<-lm(lnrv~time) # lnrv = ln(Ct/Co), so our linear model is lnrv~time  
 # then, we fit the model like this to create plots and calculate T99 values more easily later on  
 fit2<-lm(log10\_reduction~time)  
   
 # next, we set up the JM2 model  
 JM2<-function(k1,k2,sigma){  
 R = lnrv - log(1/(1+exp(k1+k2\*log(time)))) # this is the equation for the JM2 model  
 R = suppressWarnings(dnorm(R,0,sigma,log=T)) # this is the standard deviation of the normally distributed difference between measured and modeled LRVs  
 -sum(R) # this is the negative log likelihood, which must be maximized (using the MLE method)  
 }  
   
 # in order to use MLE to find the best values for the JM2 model, we have to provide some halfway decent guesses for k1, k2, and sigma  
 # through trial and error, we found that starting guesses of k1=-5, k2=1, and sigma=3 generally produced converging results for most experiments; the following are exceptions  
 if(fit$coefficients[2]<=-10){k1guess=-fit$coefficients[2];k2guess=-fit$coefficients[2];siguess=3}else{  
 if(fit$coefficients[2]>=-0.0001){k1guess=-5;k2guess=0.01;siguess=2}else{  
 if(i==48|i==181|i==186){k1guess=-0.1;k2guess=1;siguess=1}else{  
 if(i==115){k1guess=4;k2guess=100;siguess=1}else{  
 if(i==112|i==113|i==105|i==114){k1guess=5;k2guess=1;siguess=1}else{  
 if(i==180){k1guess=1;k2guess=10;siguess=1}else{  
 k1guess=-5;k2guess=1;siguess=3  
 }  
 }  
 }  
 }  
 }  
 }  
  
 if(length(time)>3){ # it only calculates the k1, k2, and sigma values for the JM2 model if there are more than 3 time points in the experiment  
 fit.jm<-mle2(minuslogl=JM2,start=list(k1=k1guess,k2=k2guess,sigma=siguess),optimizer="nlminb") # uses unbounded optimization using portable Fortran programs for numerical computation (PORT) routines   
 # Reference: Fox et al. (1978). “The PORT Mathematical Subroutine Library.” ACM Transactions on Mathematical Software, 4(2), 104–126. doi:10.1145/355780.355783  
 }else{fit.jm<-NA}  
   
 ## NEXT WE STORE THE VALUES OF THE VARIABLES AND STATISTICS WE WANT TO STORE FOR EACH EXPERIMENT  
 # here, we store the basic information about the experiment (e.g., the design, operational, and environmental parameters)   
 authors[i]<-as.character(unique(persist[persist$ind==i,]$authors)) # the names of the authors of the paper where the experiment's data were published  
 experiment[i]<-as.numeric(unique(persist[persist$ind==i,]$experiment\_id)) # the unique experiment ID  
 num[i]<-length(time) # number of time points in the experiment  
 maxTime[i]<-max(time) # the maximum time point for which data were collected in the experiment  
 maxLRV[i]<-as.numeric(max(persist[persist$ind==i,]$log10\_reduction)) # the maximum log10 reduction value observed in the experiment  
 group[i]<-as.character(unique(persist[persist$ind==i,]$microbial\_group)) # the microbial group used in the experiment (viruses, bacteria, protozoa, helminths)  
 mcn[i]<-as.character(persist[persist$ind==i,]$microorganism\_common\_name) # a more specific description of the microbial group  
 addit[i]<-as.character(unique(persist[persist$ind==i,]$additive)) # the type of additive applied to the fecal sludge  
 urine[i]<-as.character(unique(persist[persist$ind==i,]$urine)) # whether the experiment was done with feces only or feces mixed with urine (excreta)  
 moist[i]<-as.numeric(median(persist[persist$ind==i,]$moisture\_content\_percent)) # the average percent moisture content of the fecal sludge during the experiment  
 temp[i]<-as.numeric(median(persist[persist$ind==i,]$temperature\_celsius)) # the average temperature of the fecal sludge during the experiment  
 pH[i]<-as.numeric(median(persist[persist$ind==i,]$pH)) # the average pH of the fecal sludge during the experiment  
 urea[i]<-as.character(unique(persist[persist$ind==i,]$urea)) # whether or not urea (or stored urine) was added to the fecal sludge during the experiment  
  
 # here, we store the statistics for the fit to the log-linear model  
 k[i]<-fit$coefficients[2] # the log-linear decay rate coefficient  
 intrcpt[i]<-fit$coefficients[1] # the log-linear model Y-intercept  
 out<-summary(fit); sterr[i]<-out$coefficients[1,2] # this is the standard error of the decay rate coefficient  
 t99[i]<-(2-fit2$coefficients[1])/fit2$coefficients[2] # the predicted T99 value for the log-linear model  
 r2[i]<-summary(fit)$r.squared # coefficient of determination for the log-linear model fit  
 aicLL[i]<-AIC(fit) # the AIC value for the log-linear model fit  
   
 # and here, we store the statistics for the fit to the JM2 model  
 if(!is.na(fit.jm)){k1[i]<-coef(fit.jm)["k1"]} # the 1st JM2 model decay rate coefficient  
 if(!is.na(fit.jm)){k2[i]<-coef(fit.jm)["k2"]} # the 2nd JM2 model decay rate coefficient  
 if(!is.na(fit.jm)){sigme[i]<-coef(fit.jm)["sigma"]} # the sigma value for the fit of the JM2 model (measure of uncertainty)  
 t99jm[i]<-exp((log(99)-k1[i])/k2[i]) # the predicted T99 value for the JM2 model  
 if(length(time)>3){aicJM[i]<-AIC(fit.jm)}else{aicJM[i]<-NA} # the AIC value for the JM2 model fit  
   
 # this stores text describing which of the two models had the better fit (Chick Log-Linear Model or JM2 Model), based on the AIC values calculated  
 if(is.na(aicJM[i])){best[i]<-"Chick Model"}else{if(aicLL[i]<aicJM[i]){best[i]<-"Chick Model"}else{best[i]<-"JM2 Model"}}  
  
 # this next chunk of script creates the plot for each panel of Figure S1 (new plot created for each experiment during each round of the FOR loop)  
 par(mar=c(5,4,3.5,1)) # sets the margins so that the plot looks nice  
 plot(time,-log10\_reduction,ylim=c(-8,0),xlim=c(0,500),ylab="Log10 Reduction",xlab="Time (days)",main=paste(experiment[i],", ",group[i],", N=",num[i],", LRVmax=",round(maxLRV[i],2),"\n",addit[i],", ",urea[i],", ",round(r2[i],2),"\n",authors[i],sep=""),sub=best[i],cex.main=0.95)  
 lines(seq(0,360),-fit2$coefficients[2]\*seq(0,360)-fit2$coefficients[1],col="blue")  
 if(length(time)>3){lines(seq(0,360),log10(1/(1+exp(coef(fit.jm)["k1"]+coef(fit.jm)["k2"]\*log(seq(0,360))))),lty=2,col="red")}  
}

# now we will create a data frame with all the values from all experiments  
kPit<-data.frame(experiment=experiment,authors=authors,microbial\_group=group,mcn=mcn,k=k,se.k=sterr,k1=k1,k2=k2,se.jm=sigme,int=intrcpt,aicLL=aicLL,aicJM=aicJM,num=num,additive=addit,urea=urea,urine=urine,moisture=moist,temp=temp,pH=pH,r2=r2,maxLRV=maxLRV,t99=t99,t99jm=t99jm,bestModel=best)  
kPit$k<-round(kPit$k,4)  
kPit$k1<-round(kPit$k1,2)  
kPit$k2<-round(kPit$k2,2)  
kPit.demo<-kPit[c(1,17,126,236),c("experiment","authors","mcn","k","k1","k2")]  
kPit.demo

## experiment authors mcn k  
## 1 128 Endale et al.(2012) Ascaris -0.0656  
## 17 144 Magri et al.(2013) Salmonella enterica Typhimurium -0.0610  
## 126 253 Decrey and Kohn (2017) MS2 coliphage -1.0044  
## 236 364 Graham et al. (2003) Cryptosporidium -0.0149  
## k1 k2  
## 1 -6.79 2.48  
## 17 -7.38 3.41  
## 126 -7.19 7.99  
## 236 NA NA

The previous chunk of code hid the plots that were created for Figure S1. Here is a demonstrative plot for experiment 144.



Here is a summary of which model provided the better fit for each experiment

kPit[,c("experiment","mcn","bestModel")]

## experiment mcn bestModel  
## 1 128 Ascaris JM2 Model  
## 2 129 Ascaris JM2 Model  
## 3 130 Ascaris JM2 Model  
## 4 131 Ascaris JM2 Model  
## 5 132 Fecal Coliforms Chick Model  
## 6 133 Fecal Coliforms Chick Model  
## 7 134 Fecal Coliforms JM2 Model  
## 8 135 Fecal Coliforms JM2 Model  
## 9 136 Enterococcus faecalis Chick Model  
## 10 137 Enterococcus faecalis Chick Model  
## 11 138 Enterococcus faecalis Chick Model  
## 12 139 Enterococcus faecalis Chick Model  
## 13 140 Enterococcus faecalis Chick Model  
## 14 141 Salmonella enterica Typhimurium JM2 Model  
## 15 142 Salmonella enterica Typhimurium JM2 Model  
## 16 143 Salmonella enterica Typhimurium JM2 Model  
## 17 144 Salmonella enterica Typhimurium JM2 Model  
## 18 145 Salmonella enterica Typhimurium JM2 Model  
## 19 146 Bacteriophage φX JM2 Model  
## 20 147 Bacteriophage φX Chick Model  
## 21 148 Bacteriophage φX JM2 Model  
## 22 149 Bacteriophage φX Chick Model  
## 23 150 Bacteriophage φX Chick Model  
## 24 151 Bacteriophage φX Chick Model  
## 25 152 Bacteriophage MS2 JM2 Model  
## 26 153 Bacteriophage MS2 JM2 Model  
## 27 154 Bacteriophage MS2 JM2 Model  
## 28 155 Bacteriophage MS2 JM2 Model  
## 29 156 Bacteriophage MS2 JM2 Model  
## 30 157 Bacteriophage MS2 Chick Model  
## 31 158 Salmonella typhimurium phages JM2 Model  
## 32 159 Salmonella typhimurium phages Chick Model  
## 33 160 Salmonella typhimurium phages Chick Model  
## 34 161 Salmonella typhimurium phages JM2 Model  
## 35 162 Ascaris suum JM2 Model  
## 36 163 Ascaris suum JM2 Model  
## 37 164 Ascaris suum Chick Model  
## 38 165 Ascaris eggs Chick Model  
## 39 166 Ascaris eggs Chick Model  
## 40 167 Ascaris eggs Chick Model  
## 41 168 Ascaris eggs JM2 Model  
## 42 169 Ascaris eggs JM2 Model  
## 43 170 Ascaris eggs Chick Model  
## 44 171 Ascaris eggs JM2 Model  
## 45 172 Ascaris eggs JM2 Model  
## 46 173 Escherichia Coli JM2 Model  
## 47 174 Escherichia Coli JM2 Model  
## 48 175 Escherichia Coli JM2 Model  
## 49 176 MS2 coliphage Chick Model  
## 50 177 MS2 coliphage JM2 Model  
## 51 178 MS2 coliphage Chick Model  
## 52 179 Ascaris JM2 Model  
## 53 180 Ascaris Chick Model  
## 54 181 Ascaris Chick Model  
## 55 182 Ascaris JM2 Model  
## 56 183 Ascaris Chick Model  
## 57 184 Ascaris Chick Model  
## 58 185 Ascaris JM2 Model  
## 59 186 Ascaris JM2 Model  
## 60 187 Ascaris Chick Model  
## 61 188 Ascaris JM2 Model  
## 62 189 Ascaris JM2 Model  
## 63 190 Ascaris ova Chick Model  
## 64 191 Ascaris ova Chick Model  
## 65 192 Ascaris ova JM2 Model  
## 66 193 Ascaris ova Chick Model  
## 67 194 Ascaris ova JM2 Model  
## 68 195 Ascaris ova JM2 Model  
## 69 196 Ascaris eggs JM2 Model  
## 70 197 Ascaris eggs JM2 Model  
## 71 198 Ascaris eggs JM2 Model  
## 72 199 Ascaris eggs JM2 Model  
## 73 200 Ascaris eggs JM2 Model  
## 74 201 Ascaris eggs JM2 Model  
## 75 202 Fecal coliforms Chick Model  
## 76 203 Streptococcus Chick Model  
## 77 204 Salmonella Chick Model  
## 78 205 Fecal coliforms Chick Model  
## 79 206 Streptococcus Chick Model  
## 80 207 Salmonella Chick Model  
## 81 208 Fecal coliforms Chick Model  
## 82 209 Streptococcus Chick Model  
## 83 210 Salmonella Chick Model  
## 84 211 Fecal coliforms Chick Model  
## 85 212 Streptococcus Chick Model  
## 86 213 Salmonella Chick Model  
## 87 214 Fecal coliforms Chick Model  
## 88 215 Streptococcus Chick Model  
## 89 216 Salmonella Chick Model  
## 90 217 Fecal coliforms Chick Model  
## 91 218 Streptococcus Chick Model  
## 92 219 Salmonella Chick Model  
## 93 220 Fecal coliforms Chick Model  
## 94 221 Streptococcus Chick Model  
## 95 222 Salmonella Chick Model  
## 96 223 Total coliform JM2 Model  
## 97 224 Eschericia Coli Chick Model  
## 98 225 Streptococcus JM2 Model  
## 99 226 Sulfur reducing clostridia Chick Model  
## 100 227 Entamoeba cysts Chick Model  
## 101 228 Ascaris eggs Chick Model  
## 102 229 Escherichia coli Chick Model  
## 103 230 Escherichia coli Chick Model  
## 104 231 Escherichia coli Chick Model  
## 105 232 Escherichia coli JM2 Model  
## 106 233 Escherichia coli JM2 Model  
## 107 234 Escherichia coli JM2 Model  
## 108 235 Escherichia coli Chick Model  
## 109 236 Escherichia coli Chick Model  
## 110 237 Escherichia coli Chick Model  
## 111 238 Enterococcus JM2 Model  
## 112 239 Enterococcus JM2 Model  
## 113 240 Enterococcus JM2 Model  
## 114 241 Enterococcus JM2 Model  
## 115 242 Enterococcus JM2 Model  
## 116 243 Enterococcus JM2 Model  
## 117 244 Enterococcus Chick Model  
## 118 245 Enterococcus Chick Model  
## 119 246 Enterococcus Chick Model  
## 120 247 Enterococcus Chick Model  
## 121 248 Enterococcus Chick Model  
## 122 249 Enterococcus Chick Model  
## 123 250 Escherichia coli Chick Model  
## 124 251 Escherichia coli Chick Model  
## 125 252 Escherichia coli Chick Model  
## 126 253 MS2 coliphage JM2 Model  
## 127 254 T4 coliphage JM2 Model  
## 128 255 Adenovirus JM2 Model  
## 129 256 Ascaris Chick Model  
## 130 257 Ascaris Chick Model  
## 131 258 Ascaris JM2 Model  
## 132 259 Ascaris JM2 Model  
## 133 260 Ascaris Chick Model  
## 134 261 Ascaris JM2 Model  
## 135 262 Ascaris JM2 Model  
## 136 263 Ascaris JM2 Model  
## 137 264 Ascaris Chick Model  
## 138 265 Ascaris Chick Model  
## 139 266 Ascaris Chick Model  
## 140 267 Ascaris Chick Model  
## 141 268 Salmonella spp Chick Model  
## 142 269 Salmonella spp Chick Model  
## 143 270 Salmonella spp Chick Model  
## 144 271 Salmonella spp Chick Model  
## 145 272 Salmonella spp Chick Model  
## 146 273 Salmonella spp Chick Model  
## 147 274 Salmonella spp Chick Model  
## 148 275 Salmonella spp Chick Model  
## 149 276 Salmonella spp Chick Model  
## 150 277 Salmonella spp Chick Model  
## 151 278 Salmonella spp Chick Model  
## 152 279 Salmonella spp JM2 Model  
## 153 280 Salmonella typhimurium phage 28B JM2 Model  
## 154 281 Thermotolerant coliform JM2 Model  
## 155 282 Enterococci JM2 Model  
## 156 283 Salmonella senftenberg JM2 Model  
## 157 284 Salmonella typhimurium phage 28B JM2 Model  
## 158 285 Salmonella senftenberg JM2 Model  
## 159 286 Thermotolerant coliform JM2 Model  
## 160 287 Enterococci JM2 Model  
## 161 288 Fecal coliforms Chick Model  
## 162 289 Fecal coliforms JM2 Model  
## 163 290 Fecal coliforms Chick Model  
## 164 291 Fecal coliforms Chick Model  
## 165 292 Escherichia coli JM2 Model  
## 166 293 Enterococci JM2 Model  
## 167 294 Enterococci JM2 Model  
## 168 295 Escherichia coli Chick Model  
## 169 296 Enterococci JM2 Model  
## 170 297 Escherichia coli JM2 Model  
## 171 298 Enterococci JM2 Model  
## 172 299 Escherichia coli JM2 Model  
## 173 300 Escherichia coli Chick Model  
## 174 302 Escherichia coli JM2 Model  
## 175 303 Escherichia coli Chick Model  
## 176 304 Fecal coliforms Chick Model  
## 177 305 Fecal coliforms JM2 Model  
## 178 306 Fecal coliforms Chick Model  
## 179 307 Fecal coliforms Chick Model  
## 180 308 Escherichia coli Chick Model  
## 181 309 Escherichia coli Chick Model  
## 182 310 Escherichia coli Chick Model  
## 183 311 Escherichia coli Chick Model  
## 184 312 Escherichia coli Chick Model  
## 185 313 Escherichia coli Chick Model  
## 186 314 Total coliforms JM2 Model  
## 187 315 Total coliforms Chick Model  
## 188 316 Total coliforms Chick Model  
## 189 317 Total coliforms JM2 Model  
## 190 318 Total coliforms Chick Model  
## 191 319 Total coliforms Chick Model  
## 192 320 Salmonella spp. JM2 Model  
## 193 321 Salmonella spp. Chick Model  
## 194 322 Salmonella spp. Chick Model  
## 195 323 Escherichia coli JM2 Model  
## 196 324 Escherichia coli JM2 Model  
## 197 325 Escherichia coli JM2 Model  
## 198 326 Shigella sonnei JM2 Model  
## 199 327 Shigella sonnei JM2 Model  
## 200 328 Shigella flexneri Chick Model  
## 201 329 Shigella sonnei JM2 Model  
## 202 330 Shigella sonnei JM2 Model  
## 203 331 Shigella flexneri JM2 Model  
## 204 332 Rotavirus Chick Model  
## 205 333 Rotavirus Chick Model  
## 206 334 Rotavirus Chick Model  
## 207 335 Rotavirus JM2 Model  
## 208 336 Rotavirus Chick Model  
## 209 337 Rotavirus Chick Model  
## 210 338 Rotavirus Chick Model  
## 211 339 Rotavirus Chick Model  
## 212 340 Escherichia coli JM2 Model  
## 213 341 Escherichia coli JM2 Model  
## 214 342 Escherichia coli Chick Model  
## 215 343 Escherichia coli JM2 Model  
## 216 344 Escherichia coli Chick Model  
## 217 345 Enterococcus faecalis Chick Model  
## 218 346 Enterococcus faecalis Chick Model  
## 219 347 Enterococcus faecalis Chick Model  
## 220 348 Enterococcus faecalis JM2 Model  
## 221 349 Enterococcus faecalis JM2 Model  
## 222 350 Ascaris lumbricoides Chick Model  
## 223 351 Ascaris lumbricoides JM2 Model  
## 224 352 Ascaris lumbricoides JM2 Model  
## 225 353 Ascaris lumbricoides JM2 Model  
## 226 354 Ascaris lumbricoides JM2 Model  
## 227 355 Escherichia coli JM2 Model  
## 228 356 Escherichia coli JM2 Model  
## 229 357 Escherichia coli JM2 Model  
## 230 358 Escherichia coli JM2 Model  
## 231 359 Escherichia coli JM2 Model  
## 232 360 Escherichia coli JM2 Model  
## 233 361 Escherichia coli Chick Model  
## 234 362 Escherichia coli Chick Model  
## 235 363 Cryptosporidium Chick Model  
## 236 364 Cryptosporidium Chick Model  
## 237 365 Giardia Chick Model  
## 238 366 Giardia Chick Model  
## 239 367 Fecal coliforms Chick Model  
## 240 368 Fecal coliforms Chick Model  
## 241 369 Cryptosporidium JM2 Model  
## 242 370 Cryptosporidium JM2 Model  
## 243 371 Cryptosporidium JM2 Model

## Produce figures and tables

This last chunk of code uses the data frame kPit (which has all of the results from fitting the models to data from each experiment) to produce figures and tables.

### Table 1

# sample size for k-value calcs based on log linear, then sample size for t99 calcs based on log linear, then the k values, then the estimated T99 values based on the log linear model  
data.frame(rbind(c(paste(length(kPit[kPit$microbial\_group=="Virus",]$k)," (",length(kPit[kPit$microbial\_group=="Virus"&kPit$maxLRV>=1,]$t99),")",sep=""),   
 -quantile(kPit[kPit$microbial\_group=="Virus",]$k,c(0.5,0.95,0.05)),  
 quantile(kPit[kPit$microbial\_group=="Virus"&kPit$maxLRV>=1,]$t99,c(0.5,0.05,0.95))),  
c(paste(length(kPit[kPit$microbial\_group=="Bacteria",]$k)," (",length(kPit[kPit$microbial\_group=="Bacteria"&kPit$maxLRV>=1,]$t99),")",sep=""),   
 -quantile(kPit[kPit$microbial\_group=="Bacteria",]$k,c(0.5,0.95,0.05)),  
 quantile(kPit[kPit$microbial\_group=="Bacteria"&kPit$maxLRV>=1,]$t99,c(0.5,0.05,0.95))),  
c(paste(length(kPit[kPit$microbial\_group=="Protozoa",]$k)," (",length(kPit[kPit$microbial\_group=="Protozoa"&kPit$maxLRV>=1,]$t99),")",sep=""),   
 -quantile(kPit[kPit$microbial\_group=="Protozoa",]$k,c(0.5,0.95,0.05)),  
 quantile(kPit[kPit$microbial\_group=="Protozoa"&kPit$maxLRV>=1,]$t99,c(0.5,0.05,0.95))),  
c(paste(length(kPit[kPit$microbial\_group=="Helminth",]$k)," (",length(kPit[kPit$microbial\_group=="Helminth"&kPit$maxLRV>=1,]$t99),")",sep=""),   
 -quantile(kPit[kPit$microbial\_group=="Helminth",]$k,c(0.5,0.95,0.05)),  
 quantile(kPit[kPit$microbial\_group=="Helminth"&kPit$maxLRV>=1,]$t99,c(0.5,0.05,0.95)))  
))

## V1 X50. X95. X5. X50..1 X5..1  
## 1 32 (31) 0.18885 0.02547 16.800765 29.8092087720436 0.940083546180488  
## 2 147 (120) 0.9313 0.01164 242.79872 0.90902445630565 0.00242002305642991  
## 3 8 (4) 0.00945 0.001605 0.03501 269.526297347728 125.84936663375  
## 4 56 (23) 0.04165 5e-05 0.52 44.3065754116127 6.62139246819842  
## X95..1  
## 1 134.199331043184  
## 2 72.4674672473823  
## 3 320.697383842882  
## 4 290.930499952965

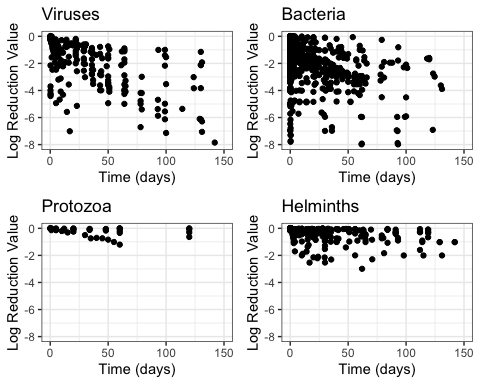
### Table 2

data.frame(rbind(  
 c(length(kPit[kPit$microbial\_group=="Virus"&kPit$num>3&kPit$maxLRV>=1,]$t99jm),  
 quantile(kPit[kPit$microbial\_group=="Virus"&kPit$num>3&kPit$maxLRV>=1,]$k1,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Virus"&kPit$num>3&kPit$maxLRV>=1,]$k2,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Virus"&kPit$num>3&kPit$maxLRV>=1,]$t99jm,c(0.5,0.05,0.95),na.rm=T)),  
 c(length(kPit[kPit$microbial\_group=="Bacteria"&kPit$num>3&kPit$maxLRV>=1,]$t99jm),  
 quantile(kPit[kPit$microbial\_group=="Bacteria"&kPit$num>3&kPit$maxLRV>=1,]$k1,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Bacteria"&kPit$num>3&kPit$maxLRV>=1,]$k2,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Bacteria"&kPit$num>3&kPit$maxLRV>=1,]$t99jm,c(0.5,0.05,0.95),na.rm=T)),  
 c(length(kPit[kPit$microbial\_group=="Protozoa"&kPit$num>3&kPit$maxLRV>=1,]$t99jm),  
 quantile(kPit[kPit$microbial\_group=="Protozoa"&kPit$num>3&kPit$maxLRV>=1,]$k1,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Protozoa"&kPit$num>3&kPit$maxLRV>=1,]$k2,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Protozoa"&kPit$num>3&kPit$maxLRV>=1,]$t99jm,c(0.5,0.05,0.95),na.rm=T)),  
 c(length(kPit[kPit$microbial\_group=="Helminth"&kPit$num>3&kPit$maxLRV>=1,]$t99jm),  
 quantile(kPit[kPit$microbial\_group=="Helminth"&kPit$num>3&kPit$maxLRV>=1,]$k1,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Helminth"&kPit$num>3&kPit$maxLRV>=1,]$k2,c(0.5,0.05,0.95),na.rm=T),  
 quantile(kPit[kPit$microbial\_group=="Helminth"&kPit$num>3&kPit$maxLRV>=1,]$t99jm,c(0.5,0.05,0.95),na.rm=T))  
))

## V1 X50. X5. X95. X50..1 X5..1 X95..1 X50..2 X5..2  
## 1 31 -7.190 -31.0500 17.725 4.920 1.485 17.870 29.367170 0.9675115  
## 2 76 0.875 -78.3400 215.855 3.625 0.450 215.855 8.677702 0.1009299  
## 3 3 -13.360 -13.6210 -6.547 3.130 2.104 3.310 220.666726 190.4310660  
## 4 18 -8.675 -58.1235 -1.847 3.830 1.389 17.626 61.532533 12.3873031  
## X95..2  
## 1 187.9810  
## 2 158.3257  
## 3 328.6808  
## 4 457.6740

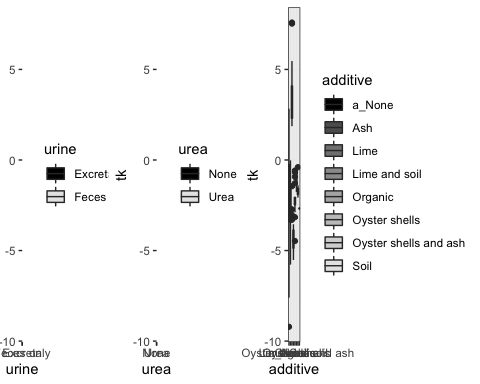
### Figure 2

persistExc<-persist[!persist$experiment\_id %in% kPit$experiment,]  
persistInc<-persist[persist$experiment\_id %in% kPit$experiment,]  
  
plotFun<-function(persistanceData){  
 fitV <- lm(-log10\_reduction ~ time\_days, data = persistanceData[persistanceData$microbial\_group=="Virus" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,])  
 new\_df <- persistanceData[persistanceData$microbial\_group=="Virus" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,]   
 vi <- ggplot(new\_df, aes(time\_days, -log10\_reduction)) +  
 geom\_point() +  
 ylim(-8,0) + ylab("Log Reduction Value") +   
 xlim(0,150) + xlab("Time (days)") +  
 ggtitle("Viruses") +  
 theme(plot.title = element\_text(hjust=0.5)) +  
 theme\_bw() #+ geom\_smooth(method=lm, se=TRUE)  
 fitB <- lm(-log10\_reduction ~ time\_days, data = persistanceData[persistanceData$microbial\_group=="Bacteria" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,])  
 new\_df <- persistanceData[persistanceData$microbial\_group=="Bacteria" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,]  
 ba <- ggplot(new\_df, aes(time\_days, -log10\_reduction))+  
 geom\_point() +  
 ylim(-8,0) + ylab("Log Reduction Value") +   
 xlim(0,150) + xlab("Time (days)") +  
 ggtitle("Bacteria") +  
 theme(plot.title = element\_text(hjust=0.5))+  
 theme\_bw()  
 fitH <- lm(-log10\_reduction ~ time\_days, data = persistanceData[persistanceData$microbial\_group=="Helminth" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,])  
 new\_df <- persistanceData[persistanceData$microbial\_group=="Helminth" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,]  
 he <- ggplot(new\_df, aes(time\_days, -log10\_reduction))+  
 geom\_point() +  
 ylim(-8,0) + ylab("Log Reduction Value") +   
 xlim(0,150) + xlab("Time (days)") +  
 ggtitle("Helminths") +  
 theme(plot.title = element\_text(hjust=0.5))+  
 theme\_bw()  
 fitP <- lm(-log10\_reduction ~ time\_days, data = persistanceData[persistanceData$microbial\_group=="Protozoa" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,])  
 new\_df <- persistanceData[persistanceData$microbial\_group=="Protozoa" & persistanceData$matrix=="Fecal sludge" & persistanceData$time\_days<200,]  
 pr <- ggplot(new\_df, aes(time\_days, -log10\_reduction))+  
 geom\_point() +  
 ylim(-8,0) + ylab("Log Reduction Value") +   
 xlim(0,150) + xlab("Time (days)") +  
 ggtitle("Protozoa") +  
 theme(plot.title = element\_text(hjust=0.5))+  
 theme\_bw()  
grid.arrange(vi,ba,pr,he,nrow = 2,widths = c(2,2))  
}  
plotFun(persistInc) # Plots Figure 2

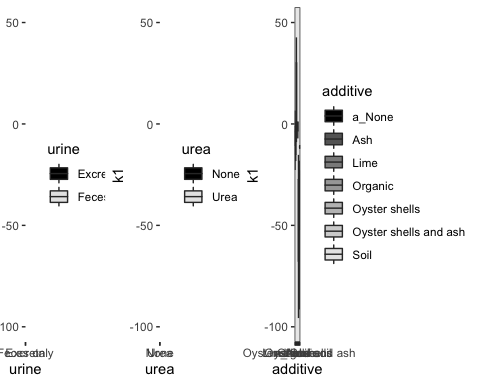


### Figure 3

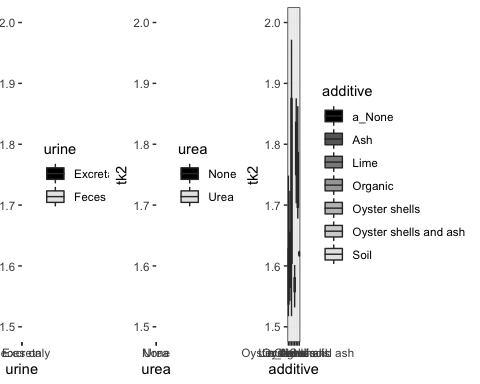
kPit$tk<-log(-kPit$k)  
kPit$tk1<-sqrt(log(kPit$k1+150))  
kPit$tk2<-sqrt(log(kPit$k2+10))  
par(mfrow=c(1,3))  
myData<-kPit[kPit$temp<50&kPit$temp>0,c("tk","k1","tk2","urine","urea","additive")]  
myData$urine<-factor(myData$urine)  
a<-ggplot(myData,aes(urine,tk))+geom\_boxplot(aes(fill=urine))+theme\_bw()+scale\_fill\_grey(start=0,end=.9)  
b<-ggplot(myData,aes(urea,tk))+geom\_boxplot(aes(fill=urea))+theme\_bw()+scale\_fill\_grey(start=0,end=.9)  
c<-ggplot(myData,aes(additive,tk))+geom\_boxplot(aes(fill=additive))+theme\_bw()+scale\_fill\_grey(start=0,end=.9)  
grid.arrange(a,b,c,nrow=1,widths=c(1,1,2))



a<-ggplot(myData,aes(urine,k1))+geom\_boxplot(aes(fill=urine),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-100,50))  
b<-ggplot(myData,aes(urea,k1))+geom\_boxplot(aes(fill=urea),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-100,50))  
c<-ggplot(myData,aes(additive,k1))+geom\_boxplot(aes(fill=additive),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-100,50))  
grid.arrange(a,b,c,nrow=1,widths=c(1,1,2))



a<-ggplot(myData,aes(urine,tk2))+geom\_boxplot(aes(fill=urine),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(1.5,2))  
b<-ggplot(myData,aes(urea,tk2))+geom\_boxplot(aes(fill=urea),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(1.5,2))  
c<-ggplot(myData,aes(additive,tk2))+geom\_boxplot(aes(fill=additive),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(1.5,2))  
grid.arrange(a,b,c,nrow=1,widths=c(1,1,2))



### Table 1

kPit$lk<-log(-kPit$k)

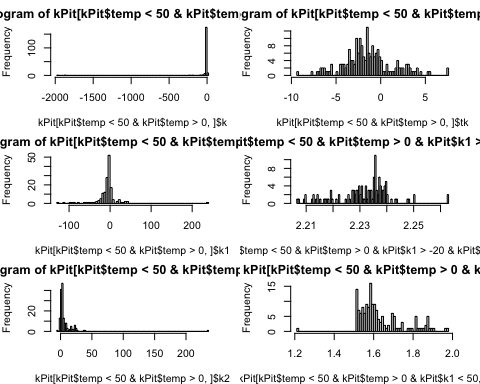
## Warning in log(-kPit$k): NaNs produced

fit.final<-lm(lk~factor(microbial\_group)+pH+temp+moisture+factor(urine)+factor(urea)+factor(additive),data=kPit[kPit$temp<50,]) #Regression  
summary(fit.final)

##   
## Call:  
## lm(formula = lk ~ factor(microbial\_group) + pH + temp + moisture +   
## factor(urine) + factor(urea) + factor(additive), data = kPit[kPit$temp <   
## 50, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8146 -0.8802 0.0259 0.9639 3.9028   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.114015 1.241884 -5.728 4.28e-08 \*\*\*  
## factor(microbial\_group)Helminth -3.095314 0.319952 -9.674 < 2e-16 \*\*\*  
## factor(microbial\_group)Protozoa -3.545962 0.809523 -4.380 2.02e-05 \*\*\*  
## factor(microbial\_group)Virus -0.853137 0.373070 -2.287 0.023391 \*   
## pH 0.336058 0.124312 2.703 0.007534 \*\*   
## temp 0.059870 0.013559 4.416 1.75e-05 \*\*\*  
## moisture 0.020508 0.005455 3.759 0.000231 \*\*\*  
## factor(urine)Feces only 1.198454 0.374606 3.199 0.001633 \*\*   
## factor(urea)Urea 0.580519 0.359224 1.616 0.107867   
## factor(additive)Ash -0.891555 0.527245 -1.691 0.092603 .   
## factor(additive)Lime 3.609997 0.520643 6.934 7.39e-11 \*\*\*  
## factor(additive)Lime and soil -1.678775 1.183292 -1.419 0.157735   
## factor(additive)Organic -1.534367 0.506761 -3.028 0.002832 \*\*   
## factor(additive)Oyster shells -0.612545 1.013803 -0.604 0.546481   
## factor(additive)Oyster shells and ash -0.347010 0.662433 -0.524 0.601044   
## factor(additive)Soil 0.311357 1.221094 0.255 0.799033   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.661 on 177 degrees of freedom  
## (9 observations deleted due to missingness)  
## Multiple R-squared: 0.6674, Adjusted R-squared: 0.6392   
## F-statistic: 23.67 on 15 and 177 DF, p-value: < 2.2e-16

### Figure S4

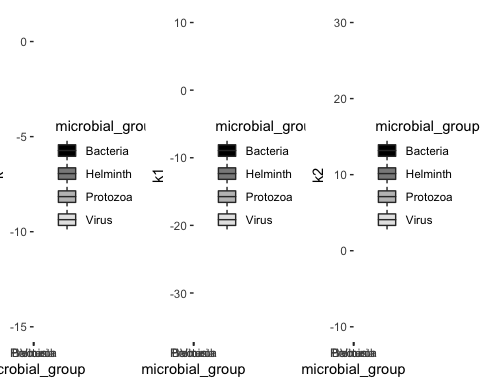
par(mfrow=c(3,2),mar=c(4,4,2,2))  
hist(kPit[kPit$temp<50&kPit$temp>0,]$k,breaks=100)  
hist(kPit[kPit$temp<50&kPit$temp>0,]$tk,breaks=100)  
hist(kPit[kPit$temp<50&kPit$temp>0,]$k1,breaks=100)  
hist(kPit[kPit$temp<50&kPit$temp>0&kPit$k1>-20&kPit$k1<20,]$tk1,breaks=100)  
hist(kPit[kPit$temp<50&kPit$temp>0,]$k2,breaks=100)  
hist(kPit[kPit$temp<50&kPit$temp>0&kPit$k1<50,]$tk2,breaks=100)



### Figure S5

myData<-kPit[kPit$temp<50&kPit$temp>0,c("microbial\_group","k","k1","k2","urine","urea","additive")]  
a<-ggplot(myData,aes(microbial\_group,k))+geom\_boxplot(aes(fill=microbial\_group),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-15,1))  
b<-ggplot(myData,aes(microbial\_group,k1))+geom\_boxplot(aes(fill=microbial\_group),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-35,10))  
c<-ggplot(myData,aes(microbial\_group,k2))+geom\_boxplot(aes(fill=microbial\_group),outlier.shape=NA)+theme\_bw()+scale\_fill\_grey(start=0,end=.9)+coord\_cartesian(ylim=c(-10,30))  
grid.arrange(a,b,c,nrow=1,widths=c(1,1,1))

## Warning: Removed 40 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 40 rows containing non-finite values (stat\_boxplot).



### Figure S6

myData<-kPit[kPit$temp<50&kPit$temp>0,c("tk","k1","tk2","temp","pH","moisture")]  
names(myData)<-c("ln(k) (1/days)","k1","tk2","Temperature (°C)","pH","moisture")  
par(mfrow=c(1,2))  
chart.Correlation(kPit[,c("tk","k1","tk2","temp","pH","moisture")],histogram=TRUE)

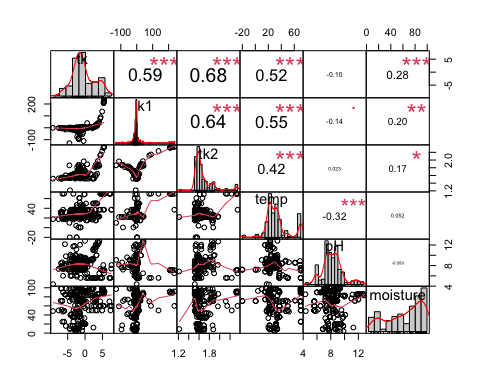
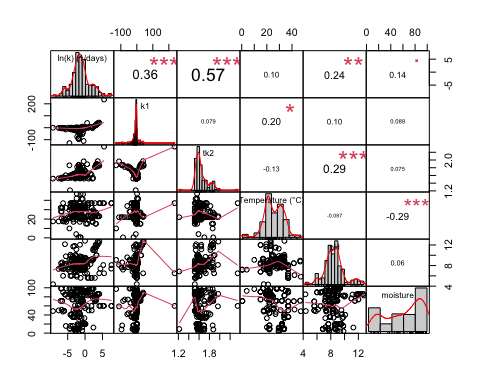


chart.Correlation(myData,histogram=TRUE)



### Figure S7

fit.final<-lm(lk~factor(microbial\_group)+pH+temp+moisture+factor(urine)+factor(urea)+factor(additive),data=kPit)  
par(mfrow=c(2,3))  
plot(fit.final,which=1:6)

