Wesner\_etal\_Ecosystems

Jeff Wesner

February 27, 2019

This code will generate figures, tables, and summary statistics from Wesner et al. 2019 (in review at Ecosystems). NOTE: You will first need to install Stan (HMC sampler) by following the directions here: <https://mc-stan.org/>

#-----Install and open the packages below-----#####  
library(brms) #only works if Stan is installed first!  
library(tidyverse)  
library(ggridges)  
library(scales)  
library(stringr)  
library(ggridges)  
library(cowplot)  
library(janitor)  
library(RCurl)  
library(lubridate)

# —–Read in data from GitHub

emerge\_data = emergence data, quist\_bwbu = data from Quist (2012), dm = dry mass of individual insects, lit\_est = literature estimates of emergence, reach\_km = length of river segments from Quist 2012

emerge\_data <- read.csv(text=getURL("https://raw.githubusercontent.com/jswesner/nps\_emergence/master/emerge\_data.csv"))  
quist\_bwbu <- read.csv(text=getURL("https://raw.githubusercontent.com/jswesner/nps\_emergence/master/quist\_bwbu.csv"))  
dm <- read.csv(text=getURL("https://raw.githubusercontent.com/jswesner/nps\_emergence/master/dm.csv"))  
lit\_est <- read.csv(text=getURL("https://raw.githubusercontent.com/jswesner/nps\_emergence/master/lit\_est.csv"))  
reach\_km <- read.csv(text=getURL("https://raw.githubusercontent.com/jswesner/nps\_emergence/master/reach\_km.csv"))

# ——Backwater summaries——-

quist\_bwbu%>%  
 drop\_na()%>%  
 group\_by(abv,year2, Reach)%>%  
 summarize('sum'=sum(ha))%>%  
 spread(abv,sum)%>%  
 clean\_names() %>%  
 print(n=40)

## # A tibble: 29 x 6  
## # Groups: year2 [4]  
## year2 reach bu bw fp rbw  
## <int> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 1890 0 42.1 83.4 1.95 0   
## 2 1890 2 134. 26.4 126. 0   
## 3 1890 4 49.4 11 181. 0   
## 4 1890 8 62 0 0 0   
## 5 1890 10 37.3 18.4 44.1 0   
## 6 1890 11 52.0 6.08 326. 0   
## 7 1890 12 236. 67.9 3601. 0   
## 8 1890 13 105. 76.9 2024. 0   
## 9 1950 0 8.07 47.6 3.26 0   
## 10 1950 2 85.0 23.1 273. 322.   
## 11 1950 4 54.9 17.2 68.0 41.5   
## 12 1950 8 0.95 9.84 0 0   
## 13 1950 10 10.8 14.9 44.8 0   
## 14 1950 12 26.7 34.2 1713. 0   
## 15 1950 13 11.6 119. 1082. 0   
## 16 2006 0 0.46 17.5 78.2 4.34  
## 17 2006 2 1.16 64.1 982. 304.   
## 18 2006 4 11.7 105. 155. 51.9   
## 19 2006 8 0 181. 90.9 0   
## 20 2006 10 5.42 34.8 17.8 156.   
## 21 2006 11 0 0 105. 0   
## 22 2006 12 0 34.3 1322. 0   
## 23 2006 13 0 0 796. 8.95  
## 24 2012 4 73.7 234. 165. 2.35  
## 25 2012 8 58.0 247. 8.65 0   
## 26 2012 10 58.8 21.4 163. 8.98  
## 27 2012 11 0 0 88.9 0   
## 28 2012 12 0 308. 1547. 0   
## 29 2012 13 0 128. 1400. 0

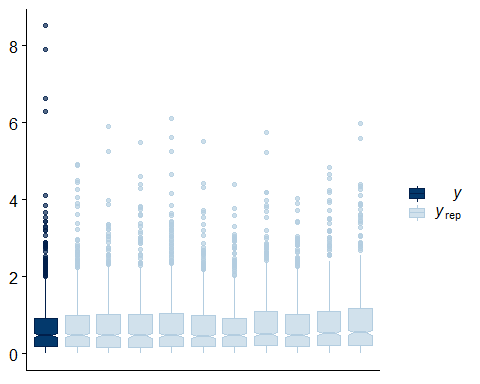
# ——MODEL 1 - Individual dry mass—–

## ——Generalized linear model for individual dry mass of insects

mdm<-brm(ind\_mg\_dry~1,data=dm,family=Gamma(link="log"),  
 prior(normal(0,2),class="Intercept"),  
 cores=4)

pp\_check(mdm,type="boxplot") #posterior predictive check

## Using 10 posterior samples for ppc type 'boxplot' by default.



mdm #model outcome

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: ind\_mg\_dry ~ 1   
## Data: dm (Number of observations: 660)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept -0.33 0.04 -0.41 -0.25 3156 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 0.91 0.04 0.82 0.99 3004 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

postmdm<-posterior\_samples(mdm) #extract posterior samples from each parameter  
dm\_p<-data.frame(postdm = exp(postmdm$b\_Intercept)) # make a data frame of posteriors

mean(dm\_p$postdm) #use this as mean for emergence dataset (i.e. ind mass = rnorm(mean(dm\_p$postdm),sd(dm\_p$post\_dm)))

## [1] 0.7183161

sd(dm\_p$postdm) #use this as sd for emergence dataset (i.e. ind mass = rnorm(mean(dm\_p$postdm),sd(dm\_p$post\_dm)))

## [1] 0.02928372

## ——Estimate dry mass of samples using model outcome from dry mass regression (model mdm)

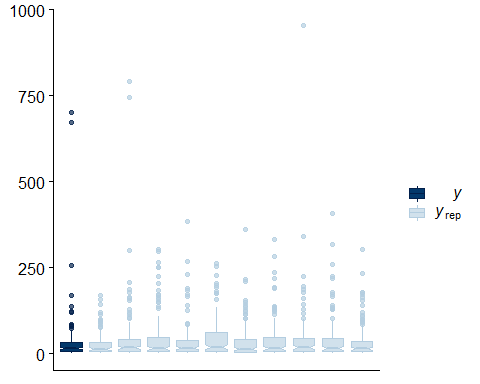
emerge\_data$est\_indmg<-rnorm(nrow(emerge\_data),mean(dm\_p$postdm),sd(dm\_p$postdm)) #add column of samples from mean and sd of post\_dm  
emerge\_data$mgm2dayDM<-emerge\_data$indm2day\*emerge\_data$est\_indmg  
emerge\_data$mgm2dayDM01 <- emerge\_data$mgm2dayDM + 0.01 # add 0.01 to prevent zeros (needed for Gamma).

# —— MODEL 2 - GAMM for emergence——

m44<-brm(mgm2dayDM01~s(day\_n)+(1|loc/year),data=emerge\_data,family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,3.9),class="Intercept"),  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000,cores=4)

pp\_check(m44,type="boxplot")#posterior predictive check

## Using 10 posterior samples for ppc type 'boxplot' by default.



summary(m44, priors=T) #model outcome and priors

## Warning: There were 98 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n) + (1 | loc/year)   
## Data: emerge\_data (Number of observations: 135)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Priors:   
## Intercept ~ normal(2.3, 3.9)  
## sd ~ cauchy(0, 1)  
## sds ~ student\_t(3, 0, 10)  
## shape ~ gamma(0.01, 0.01)  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 5.55 2.23 2.41 10.93 1415 1.00  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 4)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.49 0.43 0.02 1.59 1864 1.00  
##   
## ~loc:year (Number of levels: 8)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.72 0.32 0.30 1.48 1646 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 3.18 0.43 2.30 4.05 1754 1.00  
## sday\_n\_1 1.97 1.38 -0.79 4.70 2256 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.16 0.14 0.91 1.46 4341 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

## ——Extract fitted estimates from m44 and summarize them——

## ——Values will differ a bit from those in the ms due to simulation error, but should be close

testdata2<-data.frame(day\_n=seq(17679,17791,length=112)) #new data to condition on  
m44fit2s <- fitted(m44,summary=F,newdata = testdata2, re\_formula = NA ) #full posterior of emergence on each day

Code supportying text from manuscript: *Insect emergence was lowest in late May…ranged between*

as\_tibble(m44fit2s) %>%   
 select(V1,V2,V3,V4) %>%  
 mutate(May\_mgCm2d = (((V1+V2+V3+V4)/4)-0.053\*(V1+V2+V3+V4)/4)/2) %>% #sum last days of May in the sample, then convert to carbon units  
 summarize(median =median(May\_mgCm2d),  
 upper95 = quantile(May\_mgCm2d,probs=0.975),  
 lower95 = quantile(May\_mgCm2d, probs=0.025))

## Warning: `as\_tibble.matrix()` requires a matrix with column names or a `.name\_repair` argument. Using compatibility `.name\_repair`.  
## This warning is displayed once per session.

## # A tibble: 1 x 3  
## median upper95 lower95  
## <dbl> <dbl> <dbl>  
## 1 1.94 7.01 0.511

*It peaked in mid-June, ranging between…*

as\_tibble(m44fit2s) %>%   
 select(V26) %>%  
 mutate(mid\_June\_mgCm2d = ((V26-(0.053\*V26))/2)) %>% #isolate mid-June (day 26, i.e. column v26), then convert to carbon units  
 summarize(median =median(mid\_June\_mgCm2d),  
 upper95 = quantile(mid\_June\_mgCm2d,probs=0.975),  
 lower95 = quantile(mid\_June\_mgCm2d, probs=0.025))

## # A tibble: 1 x 3  
## median upper95 lower95  
## <dbl> <dbl> <dbl>  
## 1 30.6 77.2 10.5

*by late September…*

as\_tibble(m44fit2s) %>%   
 select(V112) %>%  
 mutate(mid\_June\_mgCm2d = ((V112-(0.053\*V112))/2)) %>% #isolate last day of sample (day112, i.e. column 112), then convert to carbon units  
 summarize(median =median(mid\_June\_mgCm2d),  
 upper95 = quantile(mid\_June\_mgCm2d,probs=0.975),  
 lower95 = quantile(mid\_June\_mgCm2d, probs=0.025))

## # A tibble: 1 x 3  
## median upper95 lower95  
## <dbl> <dbl> <dbl>  
## 1 3.84 13.7 1.22

*In total, between X and X emerged from backwaters annually…*

as\_tibble(m44fit2s) %>%  
 mutate(tot\_mgdm2y = rowSums(.)) %>%  
 select(tot\_mgdm2y) %>%  
 mutate(tot\_mgCm2y = ((tot\_mgdm2y-(0.053\*tot\_mgdm2y))/2)) %>% #sum across all days, then convert to carbon units  
 summarize(median =median(tot\_mgCm2y/1000), #divide by 1000 to convert to grams  
 upper95 = quantile(tot\_mgCm2y/1000,probs=0.975),  
 lower95 = quantile(tot\_mgCm2y/1000, probs=0.025))

## # A tibble: 1 x 3  
## median upper95 lower95  
## <dbl> <dbl> <dbl>  
## 1 1.52 3.61 0.620

## Table 2- fitted results

as\_tibble(m44fit2s) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 summarize(median\_gDM =median(totmgdm2y)/1000,  
 upper95\_gDM = quantile(totmgdm2y,probs=0.975)/1000,  
 lower95\_gDM = quantile(totmgdm2y, probs=0.025)/1000,  
 median\_gC =median(totmgCm2y)/1000,  
 upper95\_gC = quantile(totmgCm2y,probs=0.975)/1000,  
 lower95\_gC = quantile(totmgCm2y, probs=0.025)/1000,  
 median\_kJ =median((totmgdm2y - 0.053\*totmgdm2y))/1000\*23.012,  
 upper95\_kJ = quantile((totmgdm2y - 0.053\*totmgdm2y),probs=0.975)/1000\*23.012,  
 lower95\_kJ = quantile((totmgdm2y - 0.053\*totmgdm2y), probs=0.025)/1000\*23.012) %>%  
 gather()

## Warning: attributes are not identical across measure variables;  
## they will be dropped

## # A tibble: 9 x 2  
## key value  
## <chr> <dbl>  
## 1 median\_gDM 3.22   
## 2 upper95\_gDM 7.62   
## 3 lower95\_gDM 1.31   
## 4 median\_gC 1.52   
## 5 upper95\_gC 3.61   
## 6 lower95\_gC 0.620  
## 7 median\_kJ 70.2   
## 8 upper95\_kJ 166.   
## 9 lower95\_kJ 28.5

# —–PREDICTIONS FOR NEW SITES - Extract predictions for new sites m44 on each day —–

# new data to condition on  
testdata2<-data.frame(day\_n=seq(17679,17791,length=112),  
 loc="new",  
 year="new")   
  
# generate predictions with uncertainty representing a gaussian sample from the random effects term - N(0,sigma\_loc:year)  
m44pr2s<-data.frame(predict(m44,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))

# —– Total predicted yearly emergence from model m44 —–

*Based on the posterior predictive distribution, new sites…produce between… and Table 2*

as\_tibble(m44pr2s) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 summarize(median\_gDM =median(totmgdm2y/1000),  
 upper95\_gDM = quantile(totmgdm2y/1000,probs=0.975),  
 lower95\_gDM = quantile(totmgdm2y/1000, probs=0.025),  
 median\_gC =median(totmgCm2y)/1000,  
 upper95\_gC = quantile(totmgCm2y,probs=0.975)/1000,  
 lower95\_gC = quantile(totmgCm2y, probs=0.025)/1000,  
 median\_kJ =median((totmgdm2y - 0.053\*totmgdm2y))/1000\*23.012,  
 upper95\_kJ = quantile((totmgdm2y - 0.053\*totmgdm2y),probs=0.975)/1000\*23.012,  
 lower95\_kJ = quantile((totmgdm2y - 0.053\*totmgdm2y), probs=0.025)/1000\*23.012) %>%  
 gather()

## Warning: attributes are not identical across measure variables;  
## they will be dropped

## # A tibble: 9 x 2  
## key value  
## <chr> <dbl>  
## 1 median\_gDM 3.17   
## 2 upper95\_gDM 27.8   
## 3 lower95\_gDM 0.357  
## 4 median\_gC 1.50   
## 5 upper95\_gC 13.2   
## 6 lower95\_gC 0.169  
## 7 median\_kJ 69.0   
## 8 upper95\_kJ 607.   
## 9 lower95\_kJ 7.79

## Figure 2 - Emergence over time

Summarize data for figure

m44pr2s\_summary <- as.data.frame(m44pr2s) %>%  
 gather(day, mgDMd, "X1":"X112") %>%  
 mutate(day = as.numeric(str\_sub(day, start=2, end=4))) %>%  
 group\_by(day) %>%  
 summarize(median = median(mgDMd),  
 high95 = quantile(mgDMd, probs=0.975),  
 low95 = quantile(mgDMd,probs=0.025))  
  
m44fit2s\_summary <- as.data.frame(m44fit2s) %>%  
 gather(day, mgDMd,"V1":"V112") %>%  
 mutate(day = as.numeric(str\_sub(day, start=2, end=4))) %>%  
 group\_by(day) %>%  
 summarize(median = median(mgDMd),  
 high95 = quantile(mgDMd, probs=0.975),  
 low95 = quantile(mgDMd,probs=0.025))

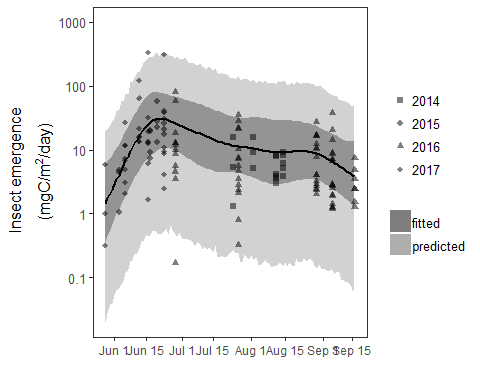
Make the figure

Figure\_2 <- ggplot()+  
 geom\_ribbon(data=m44pr2s\_summary,aes(x=day,y=(median-0.053\*median)/2, #y-values converted to carbon units  
 ymin=(low95-0.053\*low95)/2,  
 ymax=(high95-0.053\*high95)/2,fill="predicted"))+  
 geom\_ribbon(data=m44fit2s\_summary,aes(x=day,y=(median-0.053\*median)/2, #y-values converted to carbon units  
 ymin=(low95-0.053\*low95)/2,  
 ymax=(high95-0.053\*high95)/2,fill="fitted"))+  
 geom\_point(data=emerge\_data,aes(x=day\_n-17678,y=(mgm2dayDM01-0.053\*mgm2dayDM01)/2,shape=as.factor(year)),  
 size=1.8,alpha=.5)+  
 scale\_y\_log10(breaks=c(0.1,1,10,100,1000),labels=c("0.1","1","10","100","1000"))+  
 coord\_cartesian(ylim=c(0.02,1000))+  
 scale\_fill\_manual(values=c("fitted"=alpha('grey12',.35),  
 "predicted"=alpha('grey12',.2),  
 "ft2"=alpha('green',.28)))+  
 scale\_shape\_manual(values=c(15,16,17,18))+  
 theme\_bw()+  
 geom\_line(data=m44fit2s\_summary,aes(x=day,y=(median-0.053\*median)/2, #y-values converted to carbon units  
 ymin=(low95-0.053\*low95)/2,  
 ymax=(high95-0.053\*high95)/2),size=1,color="black")+  
 scale\_x\_continuous(breaks=c(5,19,35,49,66,81,98,111),labels=c("Jun 1",  
 "Jun 15",  
 "Jul 1",  
 "Jul 15",  
 "Aug 1",  
 "Aug 15",  
 "Sep 1",  
 "Sep 15"))+  
 xlab("")+  
 ylab(expression(atop("Insect emergence", paste("(mgC/m"^2,"/day)"))))+  
 theme(text=element\_text(size=12),  
 panel.grid=element\_blank(),  
 axis.title.y=element\_text(size=12),  
 axis.text.x=element\_text(size=9),  
 legend.title=element\_blank())

## Warning: Ignoring unknown aesthetics: y  
  
## Warning: Ignoring unknown aesthetics: y

## Warning: Ignoring unknown aesthetics: ymin, ymax

Figure\_2



## —–Backwater estimates

*Backwater the size of Gunderson could produce..*

as\_tibble(m44pr2s) %>%   
 select(X4) %>%  
 mutate(May\_kjd = X4/1000\*23.012\*24000) %>% #last day of May emergence, then convert from mg to gDM/m2/d, then to kJ units, then multiply by 24000 (AREA OF GUNDERSON)  
 summarize(median =median(May\_kjd),  
 upper95 = quantile(May\_kjd,probs=0.975),  
 lower95 = quantile(May\_kjd, probs=0.025),  
 birds\_low\_cost\_community = median(May\_kjd/3838), #median proportion of bird energetic costs that emergence from Gunderson backwater could support. - For lowest cost community  
 birds\_high\_cost\_community = median(May\_kjd/4656))

## # A tibble: 1 x 5  
## median upper95 lower95 birds\_low\_cost\_community birds\_high\_cost\_community  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1887. 32240. 46.1 0.492 0.405

*to X subsidies in mid-June…*

as\_tibble(m44pr2s) %>%   
 select(X26) %>%  
 mutate(June\_kjd = X26/1000\*23.012\*24000) %>% #last day of June emergence, then convert from mg to gDM/m2/d, then to kJ units, then multiply by 24000 (AREA OF GUNDERSON)  
 summarize(median =median(June\_kjd),  
 upper95 = quantile(June\_kjd,probs=0.975),  
 lower95 = quantile(June\_kjd, probs=0.025),  
 birds\_low\_cost\_community = median(June\_kjd/3838),  
 birds\_high\_cost\_community = median(June\_kjd/4656))

## # A tibble: 1 x 5  
## median upper95 lower95 birds\_low\_cost\_community birds\_high\_cost\_community  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 24786. 381475. 553. 6.46 5.32

*to X in mid-September…*

as\_tibble(m44pr2s) %>%   
 select(X112) %>%  
 mutate(Sep\_kjd = X112/1000\*23.012\*24000) %>% #last day of Sep emergence, then convert from mg to gDM/m2/d, then to kJ units, then multiply by 24000 (AREA OF GUNDERSON)  
 summarize(median =median(Sep\_kjd),  
 upper95 = quantile(Sep\_kjd,probs=0.975),  
 lower95 = quantile(Sep\_kjd, probs=0.025),  
 birds\_low\_cost\_community = median(Sep\_kjd/3838),  
 birds\_high\_cost\_community = median(Sep\_kjd/4656))

## # A tibble: 1 x 5  
## median upper95 lower95 birds\_low\_cost\_community birds\_high\_cost\_community  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 3162. 56215. 69.2 0.824 0.679

## Figure 3 - Predicted total emergence in 1890, 2006, and 2012

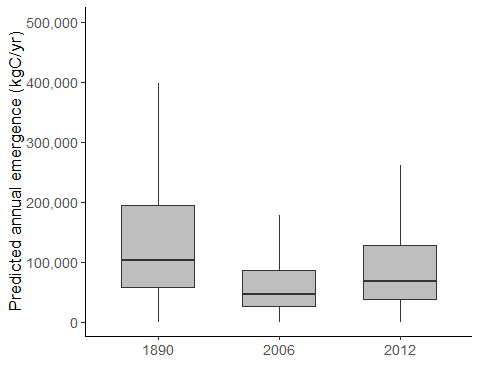
First, get total emergence per reach by multiplying emergence per m2 by total area of off-channel habitat from Quist (2012).

m44totp3 <- as\_tibble(m44pr2s) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(mgCm2y = ((totmgdm2y-(0.053\*totmgdm2y))/2)) %>%  
 select(-totmgdm2y)

m44totp3$X1890\_0\_mgyr<-m44totp3$mgCm2y\*127.41\*10000 #In reach zero, in 1890, there were 127.41 hectares of off-channel habitat, which is 127.41\*10000m2  
m44totp3$X1890\_2\_mgyr<-m44totp3$mgCm2y\*286.64\*10000 #Same as above, but for 286.64 ha  
m44totp3$X1890\_4\_mgyr<-m44totp3$mgCm2y\*241.11\*10000 #etc.  
m44totp3$X1890\_8\_mgyr<-m44totp3$mgCm2y\*62\*10000  
m44totp3$X1890\_10\_mgyr<-m44totp3$mgCm2y\*99.8\*10000  
m44totp3$X1890\_11\_mgyr<-m44totp3$mgCm2y\*383.66\*10000  
m44totp3$X1890\_12\_mgyr<-m44totp3$mgCm2y\*3904.22\*10000  
m44totp3$X1890\_13\_mgyr<-m44totp3$mgCm2y\*2206.23\*10000  
m44totp3$X1950\_0\_mgyr<-m44totp3$mgCm2y\*58.94\*10000  
m44totp3$X1950\_2\_mgyr<-m44totp3$mgCm2y\*702.74\*10000  
m44totp3$X1950\_4\_mgyr<-m44totp3$mgCm2y\*181.54\*10000  
m44totp3$X1950\_8\_mgyr<-m44totp3$mgCm2y\*10.79\*10000  
m44totp3$X1950\_10\_mgyr<-m44totp3$mgCm2y\*70.42\*10000  
m44totp3$X1950\_11\_mgyr<-m44totp3$mgCm2y\*0  
m44totp3$X1950\_12\_mgyr<-m44totp3$mgCm2y\*0  
m44totp3$X1950\_13\_mgyr<-m44totp3$mgCm2y\*1213.15\*10000  
m44totp3$X2006\_0\_mgyr<-m44totp3$mgCm2y\*100.54\*10000  
m44totp3$X2006\_2\_mgyr<-m44totp3$mgCm2y\*1350.51\*10000  
m44totp3$X2006\_4\_mgyr<-m44totp3$mgCm2y\*323.5\*10000  
m44totp3$X2006\_8\_mgyr<-m44totp3$mgCm2y\*272.04\*10000  
m44totp3$X2006\_10\_mgyr<-m44totp3$mgCm2y\*214.33\*10000  
m44totp3$X2006\_11\_mgyr<-m44totp3$mgCm2y\*105.04\*10000  
m44totp3$X2006\_12\_mgyr<-m44totp3$mgCm2y\*1355.86\*10000  
m44totp3$X2006\_13\_mgyr<-m44totp3$mgCm2y\*804.45\*10000  
m44totp3$X2012\_0\_mgyr<-m44totp3$mgCm2y\*0  
m44totp3$X2012\_2\_mgyr<-m44totp3$mgCm2y\*0  
m44totp3$X2012\_4\_mgyr<-m44totp3$mgCm2y\*474.44\*10000  
m44totp3$X2012\_8\_mgyr<-m44totp3$mgCm2y\*313.28\*10000  
m44totp3$X2012\_10\_mgyr<-m44totp3$mgCm2y\*251.83\*10000  
m44totp3$X2012\_11\_mgyr<-m44totp3$mgCm2y\*88.86\*10000  
m44totp3$X2012\_12\_mgyr<-m44totp3$mgCm2y\*1854.8\*10000  
m44totp3$X2012\_13\_mgyr<-m44totp3$mgCm2y\*1528.75\*10000

## Figure 3 - Predicted annual emergence in kgC/yr for the lowest six segments.

Figure3 <- m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = mgy/1000000) %>% #convert milligrams of C to kilograms of C  
 filter(reach >3) %>%  
 select(year,kgy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,kgy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totC = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year) %>%  
 filter(year!=1950) %>%  
 ggplot(aes(x=year, y=totC))+  
 geom\_boxplot(outlier.shape=NA,fill="grey",width=.6)+  
 coord\_cartesian(ylim=c(1000,500000))+  
 scale\_y\_continuous(labels=comma)+  
 ylab("Predicted annual emergence (kgC/yr)")+  
 theme\_classic()+  
 xlab("")+  
 theme(text=element\_text(size=13))+  
 NULL  
  
Figure3



## Emergence by segment in kgC/km/y

*Multiplying that production by the area of off-channel habitats along the lower six segments,revealed that annual insect production in 1890 ranged*

m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = mgy/1000000) %>% #convert milligrams of C to kilograms of C  
 filter(reach >3) %>%  
 select(year,kgy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,kgy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totC = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year) %>%  
 summarize(median = median(totC),  
 lower95 = quantile(totC, probs=0.025),  
 upper95 = quantile(totC, probs=0.975))

## # A tibble: 4 x 4  
## year median lower95 upper95  
## <fct> <dbl> <dbl> <dbl>  
## 1 1890 103392. 11671. 908885.  
## 2 1950 22125. 2497. 194493.  
## 3 2006 46100. 5204. 405251.  
## 4 2012 67638. 7635. 594583.

*That represents a median loss of ..*

m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = mgy/1000000) %>% #convert milligrams of C to kilograms of C  
 filter(reach >3) %>%  
 select(year,kgy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,kgy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totC = r4+r8+r10+r11+r12+r13) %>%  
 select(year,iter,totC) %>%  
 spread(year,totC)%>%  
 rename (y1890 = '1890',  
 y1950 = '1950',  
 y2006 = '2006',  
 y2012 = '2012') %>%  
 mutate(diff9012 = y2012-y1890,  
 diff9050 = y1950-y1890,  
 diff9006 = y2006-y1890) %>%  
 select(iter,diff9012,diff9050,diff9006) %>%  
 gather(year, kgy, -iter) %>%  
 group\_by(year)%>%  
 summarize(median = median(kgy),  
 lower95 = quantile(kgy, probs=0.025),  
 upper95 = quantile(kgy, probs=0.975))

## # A tibble: 3 x 4  
## year median lower95 upper95  
## <chr> <dbl> <dbl> <dbl>  
## 1 diff9006 -57292. -503634. -6467.  
## 2 diff9012 -35754. -314302. -4036.  
## 3 diff9050 -81267. -714392. -9173.

*Of the 36,000 kgC lost between…~80% was lost in segment 12…*

m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = mgy/1000000) %>% #convert milligrams of C to kilograms of C  
 filter(reach >3) %>%  
 select(year,kgy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,kgy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totC = r12) %>% #Limit summary to reach 12 only  
 select(year,iter,totC) %>%  
 spread(year,totC)%>%  
 rename (y1890 = '1890',  
 y1950 = '1950',  
 y2006 = '2006',  
 y2012 = '2012') %>%  
 mutate(diff9012 = y2012-y1890,  
 diff9050 = y1950-y1890,  
 diff9006 = y2006-y1890) %>%  
 select(iter,diff9012,diff9050,diff9006) %>%  
 gather(year, kgy, -iter) %>%  
 group\_by(year)%>%  
 summarize(median = median(kgy), # divide this result by 36000 to arrive at 80% in reach 12  
 lower95 = quantile(kgy, probs=0.025),  
 upper95 = quantile(kgy, probs=0.975))

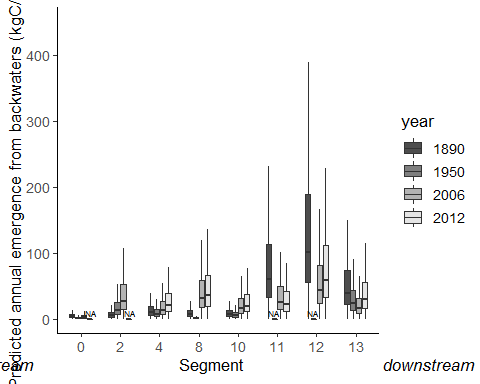
## # A tibble: 3 x 4  
## year median lower95 upper95  
## <chr> <dbl> <dbl> <dbl>  
## 1 diff9006 -38202. -335821. -4312.  
## 2 diff9012 -30723. -270071. -3468.  
## 3 diff9050 -58528. -514496. -6606.

Generate data for Figure 4 - emergence per km in each segment

reach\_km$year <- as.factor(reach\_km$year)  
  
kgC\_y <- m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = ((mgy-0.053\*mgy)/2)/1000000) %>% #convert to milligrams of C, then to kilograms of C  
 select(year,kgy,reach,iter)%>%  
 left\_join(reach\_km, by=c("reach","year")) %>%  
 mutate(kgC\_m\_y = kgy/km) %>%  
 add\_case(year=1950,kgy=0,reach=11,km=1,reach\_yr="X1950\_11\_mgyr",kgC\_m\_y=0) #adds placeholder to plot a zero on this date (no data available)

Figure4<-ggplot(data=kgC\_y,aes(x=as.factor(reach),y=kgC\_m\_y,fill=year))+  
 geom\_boxplot(outlier.shape=NA,width=.6)+  
 #geom\_violin(width=0.2,alpha=0.2)+  
 scale\_fill\_manual(values=c("grey30","grey50","grey70","grey90"))+  
 coord\_cartesian(ylim=c(1,450))+  
 #scale\_y\_continuous(breaks=c(1,100000))+  
 #geom\_text()+  
 xlab(expression(paste(italic("upstream"), " Segment ",italic("downstream"))))+  
 ylab("Predicted annual emergence from backwaters (kgC/km/yr)")+  
 annotate("text",x=c(1.26,2.26,5.92,6.92),y=10,label="NA",size=2)+  
 theme\_classic()+  
 theme(text=element\_text(size=13))  
Figure4

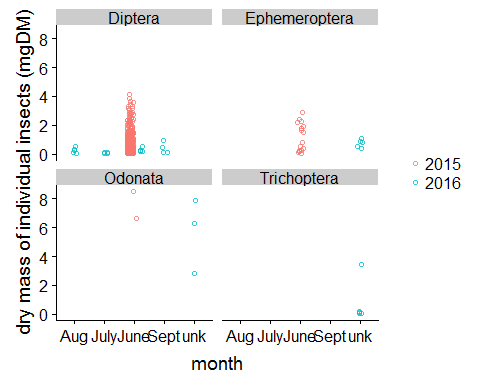
## Warning: Removed 4000 rows containing non-finite values (stat\_boxplot).



# Supplementary Information

## Figure S1: Plot of dry mass by month

ggplot(data=dm, aes(x=month, y=ind\_mg\_dry,color=as.factor(year)))+  
 geom\_point(position=position\_jitterdodge(dodge.width=0.6),shape=1,alpha=0.8)+  
 facet\_wrap(~order)+  
 ylab("dry mass of individual insects (mgDM)")+  
 theme(legend.title=element\_blank())



## Figure S2 - Priors vs post

summary(m44,priors=TRUE)

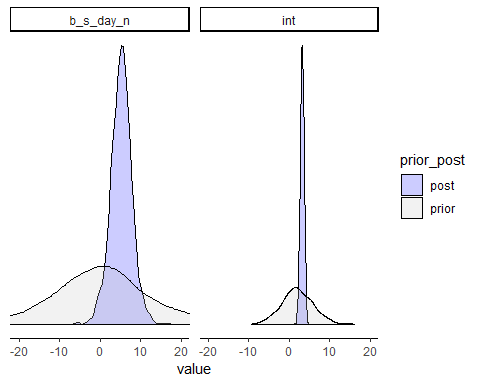
## Warning: There were 98 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n) + (1 | loc/year)   
## Data: emerge\_data (Number of observations: 135)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Priors:   
## Intercept ~ normal(2.3, 3.9)  
## sd ~ cauchy(0, 1)  
## sds ~ student\_t(3, 0, 10)  
## shape ~ gamma(0.01, 0.01)  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 5.55 2.23 2.41 10.93 1415 1.00  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 4)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.49 0.43 0.02 1.59 1864 1.00  
##   
## ~loc:year (Number of levels: 8)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.72 0.32 0.30 1.48 1646 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 3.18 0.43 2.30 4.05 1754 1.00  
## sday\_n\_1 1.97 1.38 -0.79 4.70 2256 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.16 0.14 0.91 1.46 4341 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

##Priors##  
int<-rnorm(1000,log(10),log(50))  
#sd\_loc<-rcauchy(1000,0,10)  
#sd\_\_locyr<-rcauchy(1000,0,10)  
b\_s\_day\_n<-rstudent\_t(1000,3,0,10)  
#shape<-rgamma(10000,0.01,scale=0.01)  
  
priorsm44<-data.frame(int,b\_s\_day\_n)  
  
  
##posteriors###  
int<-rnorm(1000,3.2,.48)  
#sd\_loc<-rcauchy(1000,.59,.54)  
#sd\_\_locyr<-rcauchy(1000,.76,.35)  
b\_s\_day\_n<-rstudent\_t(1000,10,5.3,2.26)  
#shape<-rgamma(1000,1.18,scale=0.14)  
  
postsm44<-data.frame(int,b\_s\_day\_n)  
  
priorsm44<-gather(priorsm44,parameter,value)  
postsm44<-gather(postsm44,parameter,value)  
priorsm44$prior\_post<-"prior"  
postsm44$prior\_post<-"post"  
  
  
#rbind priors and posts##  
pr\_post<-rbind(priorsm44,postsm44)  
pr\_post$parameter<-as.factor(pr\_post$parameter)

## Plot data for Figure S2

Figure\_S2<-ggplot(pr\_post,aes(value,fill=prior\_post),alpha=0.4)+  
 geom\_density(alpha=0.2)+  
 scale\_fill\_manual(values=c('blue','grey'))+  
 coord\_cartesian(xlim=c(-20,20))+  
 facet\_wrap(~parameter,scales="free")+  
 theme\_classic()+  
 theme(axis.line.y = element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.ticks.y=element\_blank())  
Figure\_S2



# Figure S3 - Tests for bias

In this section, we gather the posterior predictive distribution for annual emergence from each of five models. The first model is the original model from above (m44)

all\_sites <- as\_tibble(m44pr2s) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 select(totmgCm2y) %>%  
 mutate(model = "all\_sites")

Now run four models with single sites left out each time.

m44\_noabovem<-brm(mgm2dayDM01~s(day\_n,k=-1)+(1|loc/year),data=subset(emerge\_data,loc!="above"),family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,3.9),class="Intercept"),  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000, cores=4)

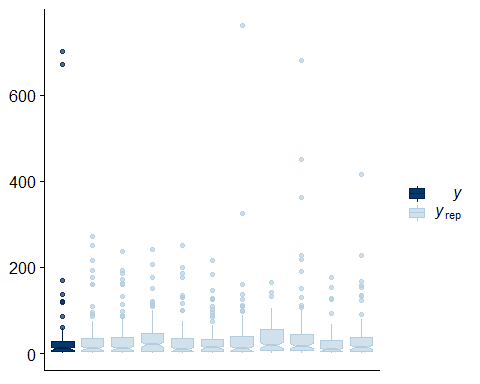
m44\_noabovem

## Warning: There were 104 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n, k = -1) + (1 | loc/year)   
## Data: subset(emerge\_data, loc != "above") (Number of observations: 106)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 6.34 2.56 2.78 12.82 889 1.01  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 3)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.50 0.53 0.01 2.00 1580 1.00  
##   
## ~loc:year (Number of levels: 6)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.79 0.41 0.21 1.78 1401 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 2.96 0.59 1.69 4.07 864 1.01  
## sday\_n\_1 3.04 1.63 0.01 6.33 2032 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.12 0.14 0.86 1.42 3940 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

pp\_check(m44\_noabovem,type="boxplot")

## Using 10 posterior samples for ppc type 'boxplot' by default.



testdata2<-data.frame(day\_n=seq(17679,17791,length=112),  
 loc="new",  
 year="new")  
m44pr\_no\_above<-data.frame(predict(m44\_noabovem,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
  
#posterior predictions for m44\_noabovem, i.e. the model without "above" sites.  
no\_above <- as\_tibble(m44pr\_no\_above) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 select(totmgCm2y) %>%  
 mutate(model = "no\_above")

The codes below repeat each of these steps (those directly above), but leave out a different site each time.

m44\_nobelowm<-brm(mgm2dayDM01~s(day\_n,k=-1)+(1|loc/year),data=subset(emerge\_data,loc!="below"),family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,3.9),class="Intercept"),  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000, cores=4)

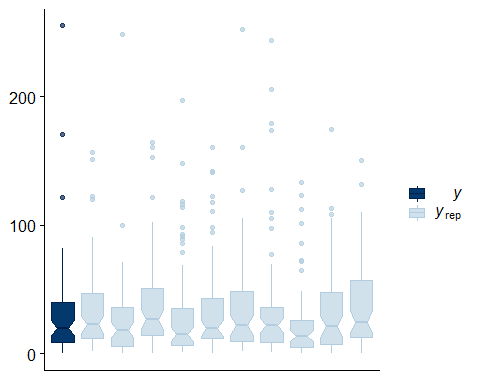
m44\_nobelowm

## Warning: There were 148 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n, k = -1) + (1 | loc/year)   
## Data: subset(emerge\_data, loc != "below") (Number of observations: 64)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 4.72 4.10 0.12 14.86 822 1.00  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 3)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.51 0.50 0.02 1.95 980 1.00  
##   
## ~loc:year (Number of levels: 4)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.42 0.38 0.02 1.47 1488 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 3.30 0.50 2.18 4.29 532 1.01  
## sday\_n\_1 -0.41 0.60 -1.59 1.00 598 1.01  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.25 0.20 0.88 1.67 3195 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

pp\_check(m44\_nobelowm,type="boxplot")

## Using 10 posterior samples for ppc type 'boxplot' by default.



testdata2<-data.frame(day\_n=seq(17679,17791,length=112),  
 loc="new",  
 year="new")  
m44pr\_no\_below<-data.frame(predict(m44\_nobelowm,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
no\_below <- as\_tibble(m44pr\_no\_below) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 select(totmgCm2y) %>%  
 mutate(model = "no\_below")

m44\_nolargem<-brm(mgm2dayDM01~s(day\_n,k=-1)+(1|loc/year),data=subset(emerge\_data,trt=="ambient"&loc!="largepool"),family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,3.9),class="Intercept"),  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000, cores=4)

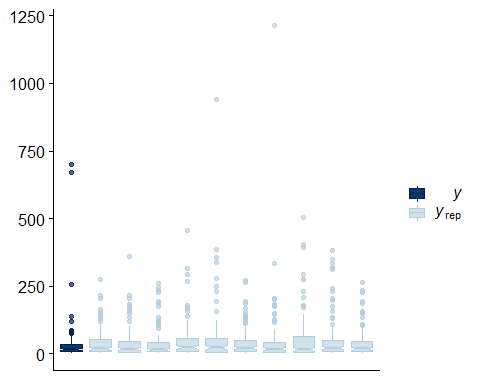
m44\_nolargem

## Warning: There were 93 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n, k = -1) + (1 | loc/year)   
## Data: subset(emerge\_data, trt == "ambient" & loc != "lar (Number of observations: 114)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 5.24 2.52 1.89 11.57 617 1.01  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 3)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.66 0.66 0.02 2.31 1893 1.00  
##   
## ~loc:year (Number of levels: 7)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 1.05 0.48 0.41 2.24 1319 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 3.15 0.65 1.74 4.44 1644 1.00  
## sday\_n\_1 2.53 1.38 -0.23 5.34 1999 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.29 0.17 0.99 1.65 4135 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

pp\_check(m44\_nolargem,type="boxplot")

## Using 10 posterior samples for ppc type 'boxplot' by default.



m44pr\_no\_large<-data.frame(predict(m44\_nolargem,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
no\_large <- as\_tibble(m44pr\_no\_large) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 select(totmgCm2y) %>%  
 mutate(model = "no\_large")

m44\_nosmallm<-brm(mgm2dayDM01~s(day\_n,k=-1)+(1|loc/year),data=subset(emerge\_data,trt=="ambient"&loc!="smallpool"),family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,3.9),class="Intercept"),  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000, cores=4)

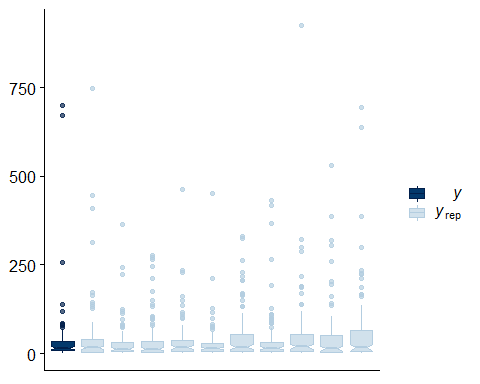
m44\_nosmallm

## Warning: There were 73 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help.  
## See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Family: gamma   
## Links: mu = log; shape = identity   
## Formula: mgm2dayDM01 ~ s(day\_n, k = -1) + (1 | loc/year)   
## Data: subset(emerge\_data, trt == "ambient" & loc != "sma (Number of observations: 121)   
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup samples = 4000  
##   
## Smooth Terms:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sds(sday\_n\_1) 5.88 2.45 2.52 12.13 1174 1.00  
##   
## Group-Level Effects:   
## ~loc (Number of levels: 3)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.64 0.62 0.02 2.28 1812 1.00  
##   
## ~loc:year (Number of levels: 7)   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## sd(Intercept) 0.76 0.35 0.30 1.64 1799 1.00  
##   
## Population-Level Effects:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## Intercept 3.23 0.55 2.03 4.29 1503 1.00  
## sday\_n\_1 1.43 1.56 -1.85 4.52 2120 1.00  
##   
## Family Specific Parameters:   
## Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat  
## shape 1.11 0.13 0.88 1.37 4270 1.00  
##   
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample   
## is a crude measure of effective sample size, and Rhat is the potential   
## scale reduction factor on split chains (at convergence, Rhat = 1).

pp\_check(m44\_nolargem,type="boxplot")

## Using 10 posterior samples for ppc type 'boxplot' by default.

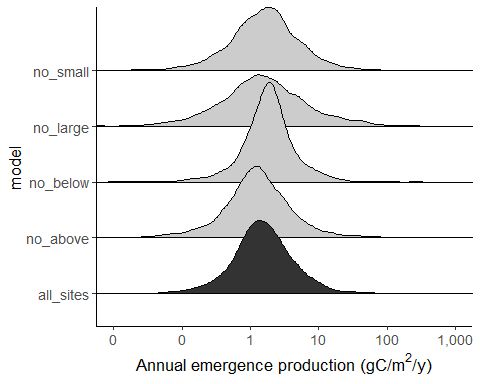


m44pr\_no\_small<-data.frame(predict(m44\_nosmallm,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
  
  
no\_small <- as\_tibble(m44pr\_no\_small) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2) %>%  
 select(totmgCm2y) %>%  
 mutate(model = "no\_small")

Combine posteriors from the four subset models above plus the main model and plot comparisons

compare\_all <- rbind(all\_sites, no\_above, no\_below, no\_large, no\_small) # combine posteriors  
compare\_all$fill <- ifelse(compare\_all$model=="all\_sites","black","grey")  
  
FigureS3 <- ggplot(compare\_all,aes(x=totmgCm2y/1000,y=model,fill=fill))+  
 geom\_density\_ridges2()+  
 scale\_fill\_grey(guide=FALSE)+  
 scale\_x\_log10(breaks=c(0.01,0.1,1,10,100,1000),labels=comma)+  
 coord\_cartesian(xlim=c(0.01,1000))+  
 theme\_classic()+  
 theme(text=element\_text(size=13))+  
 xlab(expression(paste("Annual emergence production"," (gC/m"^2,"/y)")))  
  
  
FigureS3

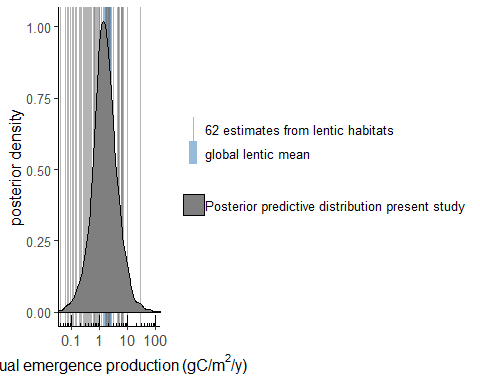
## Picking joint bandwidth of 0.0739



# Figure S4 - posterior prediction versus individual estimates from the literature survey

all\_sites$study <- "Posterior predictive distribution present study"

Figure\_S4<-ggplot()+  
 geom\_vline(data=lit\_est,aes(xintercept=gCm2yr,size=size, color=size),alpha=.5)+  
 geom\_density(data=all\_sites,aes(x=totmgCm2y/1000,fill=study))+  
 theme\_classic()+  
 scale\_color\_manual(values=c("#696969","#2e75b6"))+  
 coord\_cartesian(xlim=c(0.05,100))+  
 scale\_size\_manual(values=c(0.2,3))+  
 theme(text=element\_text(size=12),  
 panel.grid=element\_blank(),  
 axis.title.y=element\_text(size=12),  
 axis.text.x=element\_text(size=11),  
 legend.title=element\_blank())+  
 scale\_fill\_manual(values='grey50')+  
 scale\_y\_continuous()+  
 xlab(expression(paste("Annual emergence production"," (gC/m"^2,"/y)")))+  
 ylab("posterior density")+  
 scale\_x\_log10(breaks=c(0.1,1,10,100),labels=c("0.1","1","10","100"))+  
 annotation\_logticks(sides="b",mid=unit(0.1,"cm"))+  
 NULL  
Figure\_S4



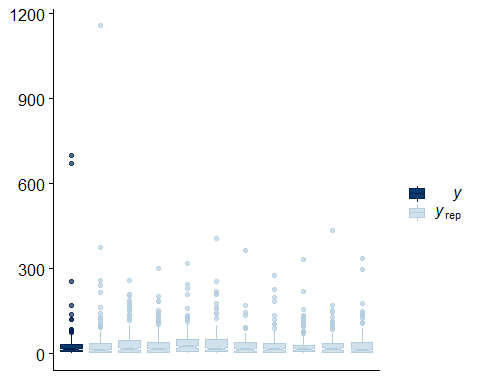
# Figure S6 - Prior sensitivity

Fit models with alternate priors and compare predictions to those of the original model with informative prior on the intercept

m44alt\_sd<-brm(mgm2dayDM01~s(day\_n)+(1|loc/year),data=emerge\_data,family=Gamma(link="log"),  
 prior=c(prior(normal(2.3,4.6),class="Intercept"), #sd is wider than original model (log(100) vs log(50))  
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000,cores=4)

pp\_check(m44alt\_sd,type="boxplot")

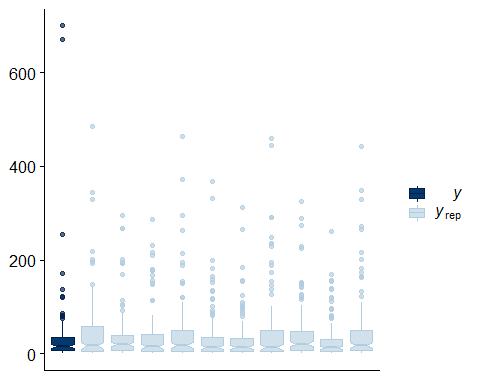
## Using 10 posterior samples for ppc type 'boxplot' by default.



m44cent\_zero <-brm(mgm2dayDM01~s(day\_n)+(1|loc/year),data=emerge\_data,family=Gamma(link="log"),  
 prior=c(prior(normal(1,7),class="Intercept"), #centered on zero (i.e. exp(0) with sd of exp(7) = 1096 ) with   
 prior(cauchy(0,1),class="sd")),  
 chains=4,iter=2000, cores=4)

pp\_check(m44cent\_zero,type="boxplot")

## Using 10 posterior samples for ppc type 'boxplot' by default.



Extract posterior predictions

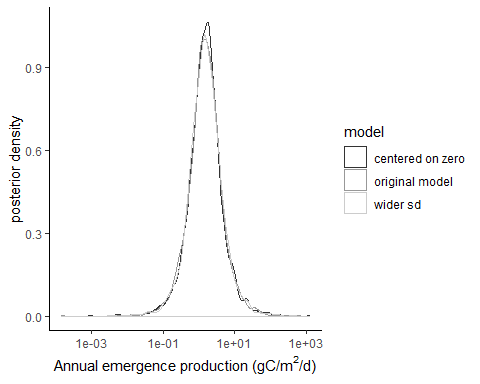
testdata2<-data.frame(day\_n=seq(17679,17791,length=112),  
 loc="new",  
 year="new")  
original\_predict <- data.frame(predict(m44,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
altsd\_predict<-data.frame(predict(m44alt\_sd,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))  
  
zero\_predict <- data.frame(predict(m44cent\_zero,type="response",newdata=testdata2,re\_formula=~(1|loc/year),  
 allow\_new\_levels = TRUE,summary=FALSE,sample\_new\_levels = "gaussian"))

Calculate distribution of total emergence from each model

orig\_pred <- as\_tibble(original\_predict) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2,  
 model = "original model")  
  
altsd\_pred <- as\_tibble(altsd\_predict) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2,  
 model = "wider sd")  
  
zero\_pred <- as\_tibble(zero\_predict) %>%  
 mutate(totmgdm2y = rowSums(.)) %>%  
 select(totmgdm2y) %>%  
 mutate(totmgCm2y =(totmgdm2y - 0.053\*totmgdm2y)/2,  
 model = "centered on zero")  
  
prior\_pred\_all <- rbind(orig\_pred, altsd\_pred, zero\_pred)

Plot distributions to compare

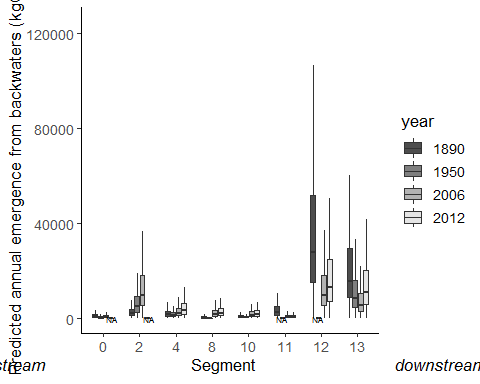
FigureS6 <- prior\_pred\_all %>%   
 rbind(orig\_pred, altsd\_pred,zero\_pred) %>%  
 mutate (gCm2y = totmgCm2y/1000) %>%  
 ggplot(aes(x=gCm2y, color=model))+  
 geom\_density()+  
 scale\_color\_grey()+  
 theme\_classic()+  
 scale\_x\_log10()+  
 xlab(expression(paste("Annual emergence production"," (gC/m"^2,"/d)")))+  
 ylab("posterior density")  
  
FigureS6



# Figure S7 - Emergence by segment in kgC/y

This figure shows emergence from each segment, but is not corrected for segment length.

kgC\_y <- m44totp3 %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgy,-iter,-mgCm2y) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 kgy = ((mgy-0.053\*mgy)/2)/1000000) %>% #convert to milligrams of C, then to kilograms of C  
 select(year,kgy,reach,iter)%>%  
 left\_join(reach\_km, by=c("reach","year")) %>%  
 mutate(kgC\_m\_y = kgy/km) %>%  
 add\_case(year=1950,kgy=0,reach=11,km=1,reach\_yr="X1950\_11\_mgyr",kgC\_m\_y=0) #adds placeholder to plot a zero on this date (no data available)  
  
  
FigureS7<-ggplot(data=kgC\_y,aes(x=as.factor(reach),y=kgy,fill=year))+  
 geom\_boxplot(outlier.shape=NA,width=.6)+  
 #geom\_violin(width=0.2,alpha=0.2)+  
 scale\_fill\_manual(values=c("grey30","grey50","grey70","grey90"))+  
 coord\_cartesian(ylim=c(1,125000))+  
 #scale\_y\_continuous(breaks=c(1,100000))+  
 #geom\_text()+  
 xlab(expression(paste(italic("upstream"), " Segment ",italic("downstream"))))+  
 ylab("Predicted annual emergence from backwaters (kgC/yr)")+  
 annotate("text",x=c(1.26,2.26,5.92,6.92),y=10,label="NA",size=2)+  
 theme\_classic()+  
 theme(text=element\_text(size=13))  
FigureS7



# Table S1

as\_tibble(emerge\_data) %>%  
 mutate(date = mdy(date))%>%  
 select(date,citation,response,loc) %>%  
 group\_by(citation,loc,date) %>%  
 summarize(n\_traps = n(),  
 area\_collected = n\_traps\*0.36)%>%  
 arrange(date)

## # A tibble: 34 x 5  
## # Groups: citation, loc [8]  
## citation loc date n\_traps area\_collected  
## <fct> <fct> <date> <int> <dbl>  
## 1 Wesner and Seidel (unpublished) above 2018-05-28 1 0.36  
## 2 Wesner and Seidel (unpublished) below 2018-05-28 2 0.72  
## 3 Warmbold and Wesner (2018) below 2018-06-03 4 1.44  
## 4 Wesner and Seidel (unpublished) above 2018-06-06 3 1.08  
## 5 Wesner and Seidel (unpublished) below 2018-06-06 5 1.8   
## 6 Wesner and Seidel (unpublished) above 2018-06-12 3 1.08  
## 7 Wesner and Seidel (unpublished) below 2018-06-12 4 1.44  
## 8 Wesner and Seidel (unpublished) above 2018-06-16 3 1.08  
## 9 Wesner and Seidel (unpublished) below 2018-06-16 5 1.8   
## 10 Warmbold and Wesner (2018) below 2018-06-17 4 1.44  
## # ... with 24 more rows

# Table S4

compare\_all %>%  
 group\_by(model) %>%  
 summarize(median = median(totmgCm2y/1000),  
 lower95 = quantile(totmgCm2y/1000,probs=0.025),  
 upper95 = quantile(totmgCm2y/1000,probs=0.975))

## # A tibble: 5 x 4  
## model median lower95 upper95  
## <chr> <dbl> <dbl> <dbl>  
## 1 all\_sites 1.50 0.169 13.2  
## 2 no\_above 1.25 0.0833 16.6  
## 3 no\_below 1.86 0.207 13.5  
## 4 no\_large 1.51 0.0608 46.5  
## 5 no\_small 1.66 0.106 20.8

# Bird Conversions

This section generates summary statistics for the number of birds that insect emergence could support. First summarize posterior predicted emergence (file m44pr2s created in earlier code)

m44totp3\_dm <- as\_tibble(m44pr2s) %>%  
 mutate(mgdm2y = rowSums(.)) %>%  
 select(mgdm2y)

Convert to total per km by multiplying by the area of m2 in the Missouri

m44totp3\_dm$X1890\_0\_mgyr<-m44totp3\_dm$mgdm2y\*127.41\*10000  
m44totp3\_dm$X1890\_2\_mgyr<-m44totp3\_dm$mgdm2y\*286.64\*10000  
m44totp3\_dm$X1890\_4\_mgyr<-m44totp3\_dm$mgdm2y\*241.11\*10000  
m44totp3\_dm$X1890\_8\_mgyr<-m44totp3\_dm$mgdm2y\*62\*10000  
m44totp3\_dm$X1890\_10\_mgyr<-m44totp3\_dm$mgdm2y\*99.8\*10000  
m44totp3\_dm$X1890\_11\_mgyr<-m44totp3\_dm$mgdm2y\*383.66\*10000  
m44totp3\_dm$X1890\_12\_mgyr<-m44totp3\_dm$mgdm2y\*3904.22\*10000  
m44totp3\_dm$X1890\_13\_mgyr<-m44totp3\_dm$mgdm2y\*2206.23\*10000  
m44totp3\_dm$X1950\_0\_mgyr<-m44totp3\_dm$mgdm2y\*58.94\*10000  
m44totp3\_dm$X1950\_2\_mgyr<-m44totp3\_dm$mgdm2y\*702.74\*10000  
m44totp3\_dm$X1950\_4\_mgyr<-m44totp3\_dm$mgdm2y\*181.54\*10000  
m44totp3\_dm$X1950\_8\_mgyr<-m44totp3\_dm$mgdm2y\*10.79\*10000  
m44totp3\_dm$X1950\_10\_mgyr<-m44totp3\_dm$mgdm2y\*70.42\*10000  
m44totp3\_dm$X1950\_11\_mgyr<-m44totp3\_dm$mgdm2y\*0  
m44totp3\_dm$X1950\_12\_mgyr<-m44totp3\_dm$mgdm2y\*0  
m44totp3\_dm$X1950\_13\_mgyr<-m44totp3\_dm$mgdm2y\*1213.15\*10000  
m44totp3\_dm$X2006\_0\_mgyr<-m44totp3\_dm$mgdm2y\*100.54\*10000  
m44totp3\_dm$X2006\_2\_mgyr<-m44totp3\_dm$mgdm2y\*1350.51\*10000  
m44totp3\_dm$X2006\_4\_mgyr<-m44totp3\_dm$mgdm2y\*323.5\*10000  
m44totp3\_dm$X2006\_8\_mgyr<-m44totp3\_dm$mgdm2y\*272.04\*10000  
m44totp3\_dm$X2006\_10\_mgyr<-m44totp3\_dm$mgdm2y\*214.33\*10000  
m44totp3\_dm$X2006\_11\_mgyr<-m44totp3\_dm$mgdm2y\*105.04\*10000  
m44totp3\_dm$X2006\_12\_mgyr<-m44totp3\_dm$mgdm2y\*1355.86\*10000  
m44totp3\_dm$X2006\_13\_mgyr<-m44totp3\_dm$mgdm2y\*804.45\*10000  
m44totp3\_dm$X2012\_0\_mgyr<-m44totp3\_dm$mgdm2y\*0  
m44totp3\_dm$X2012\_2\_mgyr<-m44totp3\_dm$mgdm2y\*0  
m44totp3\_dm$X2012\_4\_mgyr<-m44totp3\_dm$mgdm2y\*474.44\*10000  
m44totp3\_dm$X2012\_8\_mgyr<-m44totp3\_dm$mgdm2y\*313.28\*10000  
m44totp3\_dm$X2012\_10\_mgyr<-m44totp3\_dm$mgdm2y\*251.83\*10000  
m44totp3\_dm$X2012\_11\_mgyr<-m44totp3\_dm$mgdm2y\*88.86\*10000  
m44totp3\_dm$X2012\_12\_mgyr<-m44totp3\_dm$mgdm2y\*1854.8\*10000  
m44totp3\_dm$X2012\_13\_mgyr<-m44totp3\_dm$mgdm2y\*1528.75\*10000

*…equivalent to the amount of energy needed to support ~190,000 riparian woodland birds for..*

m44totp3\_dm %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgdmy,-mgdm2y,-iter) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 gdmy = mgdmy/1000) %>% #convert milligrams of dry mass to grams of dry mass  
 filter(reach >3) %>%  
 select(year,gdmy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,gdmy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totgdmy = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year ) %>%  
 mutate(totkJy = totgdmy\*23.012,#multiply grams of dry mass by 23.012 to convert to kJ (from Cummins and Wuycheck 1971)  
 ind\_bird\_demand = 4600/59\*120, #bird community in a hectare needs 4656 kJ/day and there are ~59 birds per community. This estimate how much energy a bird would need over the breeding and nesting season (120days = 4 months)  
 birds\_supported = totkJy/ind\_bird\_demand) %>%  
 select(year,birds\_supported,iter)%>%  
 spread(year,birds\_supported)%>%  
 clean\_names() %>%  
 mutate(diff9012 = x2012-x1890) %>%  
 summarize(lower95 = quantile(diff9012,probs=0.025),  
 median = median(diff9012),  
 upper95 = quantile(diff9012, probs=0.975))

## # A tibble: 1 x 3  
## lower95 median upper95  
## <dbl> <dbl> <dbl>  
## 1 -1632653. -185726. -20964.

*…enough to subsidize approximately…*

m44totp3\_dm %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgdmy,-mgdm2y,-iter) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 gdmy = mgdmy/1000) %>% #convert milligrams of dry mass to grams of dry mass  
 filter(reach >3) %>%  
 select(year,gdmy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,gdmy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totgdmy = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year ) %>%  
 mutate(totkJy = totgdmy\*23.012,#multiply grams of dry mass by 23.012 to convert to kJ (from Cummins and Wuycheck 1971)  
 ind\_bird\_demand = 4600/59\*120, #bird community in a hectare needs 4656 kJ/day and there are ~59 birds per community. This estimate how much energy a bird would need over the breeding and nesting season (120days = 4 months)  
 birds\_supported = totkJy/ind\_bird\_demand) %>%  
 select(year,birds\_supported,iter)%>%  
 spread(year,birds\_supported)%>%  
 clean\_names() %>%  
 mutate(diff9012 = x2012-x1890) %>%  
 summarize(lower95 = quantile(diff9012\*4.16,probs=0.025), #multiply by 24/100 = 4.16. i.e. every bird gets 24% of it's annual budget, so each annual allotment is split among 24/100 = 4.16 birds.  
 median = median(diff9012\*4.16),  
 upper95 = quantile(diff9012\*4.16, probs=0.975))

## # A tibble: 1 x 3  
## lower95 median upper95  
## <dbl> <dbl> <dbl>  
## 1 -6791838. -772620. -87211.

*…equivalent to the amount of energy needed to support ~190,000 riparian woodland birds for..*

m44totp3\_dm %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgdmy,-mgdm2y,-iter) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 gdmy = mgdmy/1000) %>% #convert milligrams of dry mass to grams of dry mass  
 filter(reach >3) %>%  
 select(year,gdmy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,gdmy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totgdmy = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year ) %>%  
 mutate(totkJy = totgdmy\*23.012,#multiply grams of dry mass by 23.012 to convert to kJ (from Cummins and Wuycheck 1971)  
 ind\_bird\_demand = 3800/59\*120, #bird community in a hectare needs 4656 kJ/day and there are ~59 birds per community. This estimate how much energy a bird would need over the breeding and nesting season (120days = 4 months)  
 birds\_supported = totkJy/ind\_bird\_demand) %>%  
 select(year,birds\_supported,iter)%>%  
 spread(year,birds\_supported)%>%  
 clean\_names() %>%  
 mutate(diff9012 = x2012-x1890) %>%  
 summarize(lower95 = quantile(diff9012,probs=0.025),  
 median = median(diff9012),  
 upper95 = quantile(diff9012, probs=0.975))

## # A tibble: 1 x 3  
## lower95 median upper95  
## <dbl> <dbl> <dbl>  
## 1 -1976370. -224826. -25378.

*The amount of emergence from Missouri River off-channel habitats in the early 1890s could have supported ~550,000 woodland birds*

m44totp3\_dm %>%  
 rownames\_to\_column("iter") %>%  
 gather(yr\_reach, mgdmy,-mgdm2y,-iter) %>%  
 mutate(year = as.factor(str\_sub(yr\_reach,2,5)),  
 reach= as.numeric(str\_remove(str\_sub(yr\_reach,7,8),"\_")),  
 gdmy = mgdmy/1000) %>% #convert milligrams of dry mass to grams of dry mass  
 filter(reach >3) %>%  
 select(year,gdmy,reach,iter) %>%  
 group\_by(year,reach) %>%  
 spread(reach,gdmy)%>%   
 rename(r4 = '4',  
 r8 = '8',  
 r10 = '10',  
 r11 = '11',  
 r12 = '12',  
 r13 = '13') %>%  
 mutate(totgdmy = r4+r8+r10+r11+r12+r13) %>%  
 group\_by(year ) %>%  
 mutate(totkJy = totgdmy\*23.012,#multiply grams of dry mass by 23.012 to convert to kJ (from Cummins and Wuycheck 1971)  
 ind\_bird\_demand = 4600/59\*120, #bird community in a hectare needs 4656 kJ/day and there are ~59 birds per community. This estimate how much energy a bird would need over the breeding and nesting season (120days = 4 months)  
 birds\_supported = totkJy/ind\_bird\_demand) %>%  
 summarize(lower95 = quantile(birds\_supported,probs=0.025),  
 median = median(birds\_supported),  
 upper95 = quantile(birds\_supported, probs=0.975))

## # A tibble: 4 x 4  
## year lower95 median upper95  
## <fct> <dbl> <dbl> <dbl>  
## 1 1890 60623. 537075. 4721241.  
## 2 1950 12973. 114929. 1010303.  
## 3 2006 27031. 239469. 2105091.  
## 4 2012 39659. 351349. 3088588.