Untitled

Kim Cuddington

12/07/2024

Eddie air and water temp data

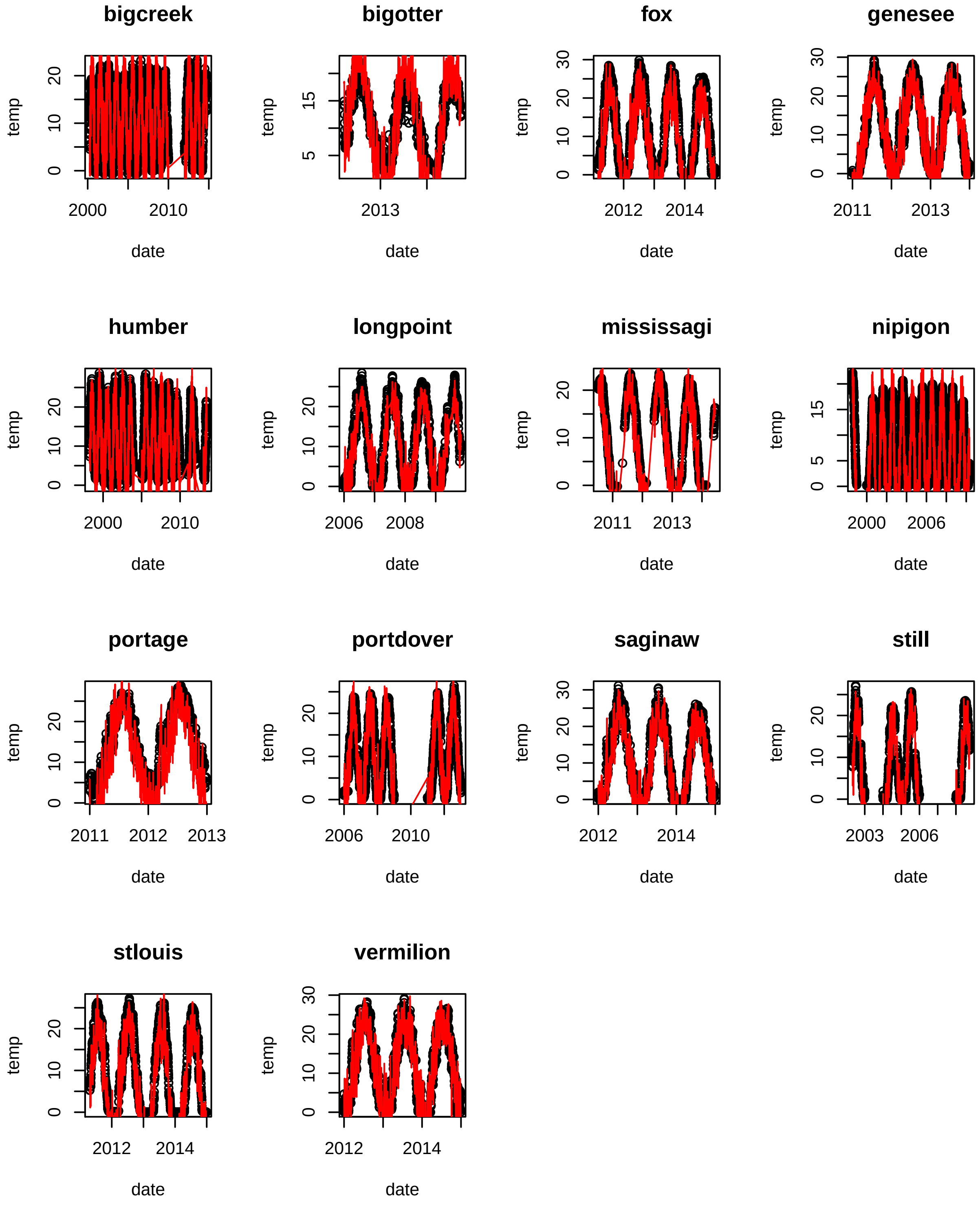
```
#import data
air=read.csv("tributary air temperatures clean.csv", stringsAsFactors = FALSE)
water <- read.csv("water_temperature_d.csv", stringsAsFactors = FALSE)

#merge data sets
aw=merge(air,water, by=c("station_name", "date"))
```

Process data

```
#change classes/calc min/max mean
aw$date=as.Date(as.character(aw$date), "%m/%d/%Y")
aw$station_name=as.factor(aw$station_name)
aw$location=as.factor(aw$location)
aw$dmean=(aw$max_temp+aw$min_temp)/2

# calculate lags
lgn=function(x,lag)c(rep(NA, lag), x[1:(length(x)-lag)])
aw$dmean_2=lgn(aw$dmean, 2)
aw$dmean_3=lgn(aw$dmean, 3)

# get vector of locations
loc_seq=levels(aw$location)

# removing missing data
aw=aw[complete.cases(cbind(aw$dmean,aw$dmean_2,aw$dmean_3,aw$temp)),]
```

Plot data

```
#plot to check import
par(mfrow=c(4,4))
for (i in seq_along(loc_seq)){
  sub=aw[(aw$location==loc_seq[i]),]
  sub=sub[order(sub$date),]
plot(temp~date, data=sub, main=loc_seq[i])
lines(dmean~date, data=sub,col="red", type="l")

}

# Note high correlation in air temps
#library(ggplot2)
#ggplot(aw[aw$year==2012,], aes(x = date, y = dmean, group = station_name, colour = station_name)) +
#  geom_line()
```


bigcreek
temp
0
10
20
2000
2010
date
bigotter
temp
5
15
2013
date
fox
temp
0
10
20
30
2012
2014
date
genesee
temp
0
10
20
30
2011
2013
date
humber
temp
0
10
20
2000
2010
date
longpoint
temp
0
10
20
2006
2008
date
mississagi
temp
0
10
20
2011
2013
date
nipigon
temp
0
5
15
2000
2006
date
portage
temp
0
10
20
2011
2012
2013
date
portdover
temp
0
10
20
2006
2010
date
saginaw
temp
0
10
20
30
2012
2014
date
still
temp
0
10
20
2003
2006
date
stlouis
temp
0
10
20
2012
2014
date
vermilion
temp
0
10
20
30
2012
2014
date

Identify years with less missing data

```
#Randomly select years for model fitting

# Or use a particular year
#aw=aw[aw$year==2012,]

# get table of sample years by location
yr_out=table(aw$location, aw$year)


# select sample years with more than X days
full_year=list()
sind=vector()
cnt=0

for (i in seq_along(loc_seq)){
 rm(ind)
   if (ncol(yr_out)>=1) {
ind=which(yr_out[row.names(yr_out)==loc_seq[i],]>250)

if (length(ind)>0) {
  sind=c(sind,i)
  cnt=cnt+1
  full_year[[cnt]]=colnames(yr_out)[ind]
}
}

}
```

```
## Warning in rm(ind): object 'ind' not found
```

```
full_year=setNames(full_year,loc_seq[sind])
```

Random sample good years

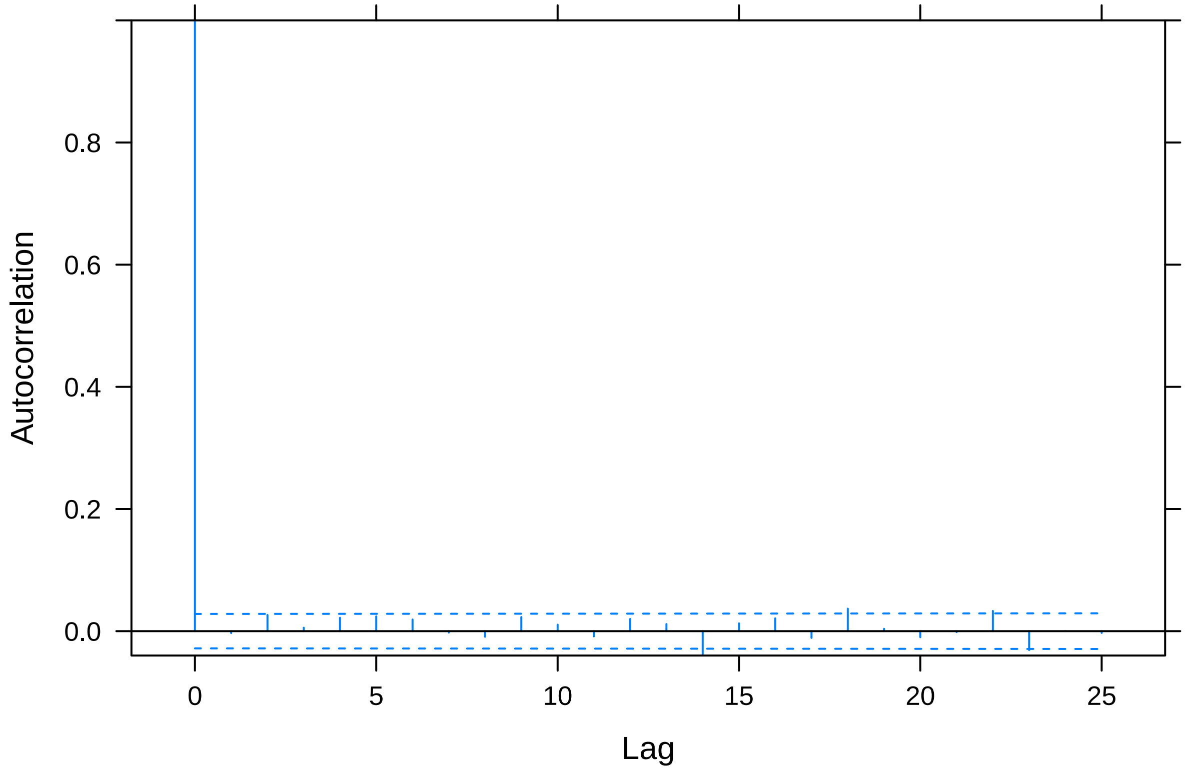
```
# randomly sample "good" years for model fit
indx=vector()
dfind=data.frame(location=character(), year=integer(), stringsAsFactors = FALSE)
cnt=0

for (i in 1:length(full_year)){
  smp=full_year[[i]]
  if (length(smp)>1){
  indx=(sample(smp, 1))
  yr=indx

  }else if (length(smp)==1){
    indx=smp
    print(indx)
    yr=indx

  }
    cnt=cnt+1
    dfind[cnt,]=c(names(full_year[i]), yr)
}

knitr::kable(dfind, title="fitting data")
```

location	year
bigcreek	2003
bigotter	2013
fox	2014
genesee	2013
humber	2006
longpoint	2008
mississagi	2012
nipigon	2009
portage	2012
portdover	2007
saginaw	2014
still	2004
stlouis	2012
vermilion	2014

```
# create fitting data
aw_sub=merge(aw, dfind, by=c("location", "year"))
aw_sub=aw_sub[complete.cases(aw_sub$dmean),]
```

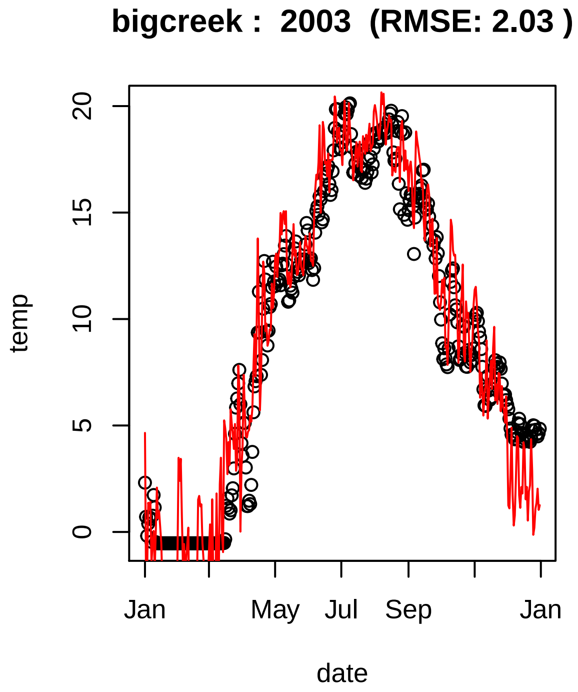
fit mixed lag model and ACF

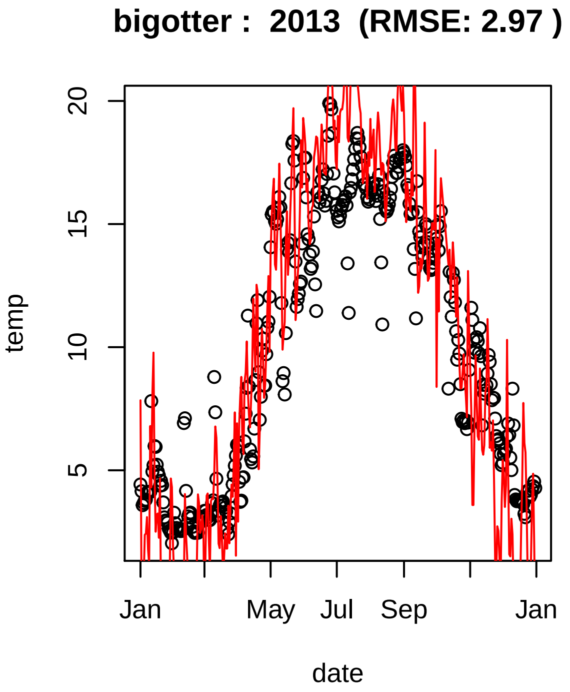
```
#fit mixed
library(nlme)
ctrl <- lmeControl(opt='optim');
fm2 <- lme(temp ~ dmean+dmean_2+dmean_3, data = aw_sub,
           control=ctrl,
           random = ~ 1 | location, na.action = na.omit,
           corAR1(form = ~ 1 | location))

# plot ACF
plot(ACF(fm2,resType="normalized"),alpha=0.05)
```

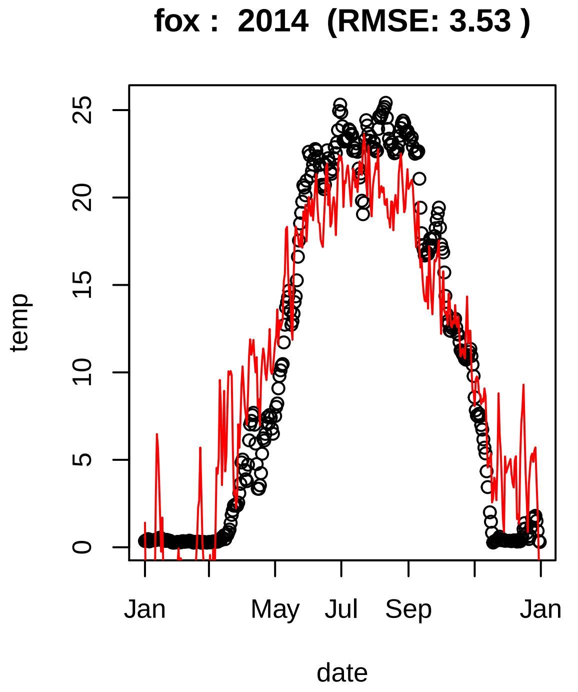


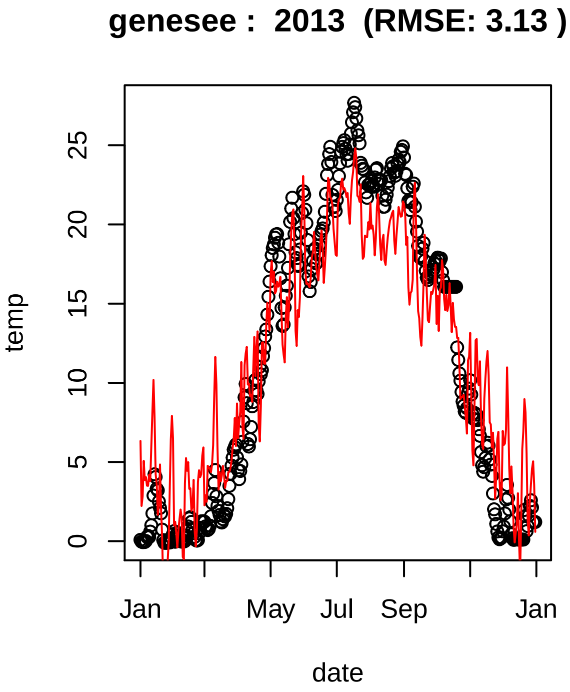
Plot predictions

```
# plot predictions
fplot=predict(fm2)
aw_sub$pred=fplot
par(mfrow=c(2,2))

for (i in 1:nrow(dfind)){
  sub=aw_sub[(aw_sub$location==dfind$location[i]),]
  sub=sub[order(sub$date),]
  diff=round(sqrt(mean((sub$temp-sub$pred)^2)),2)
  plot(temp~date, data=sub,
       main=paste(dfind$location[i], ": ", dfind$year[i],
                  " (RMSE:", diff, ")"))
  lines(pred~date, data=sub,col="red", type="l")
}
```

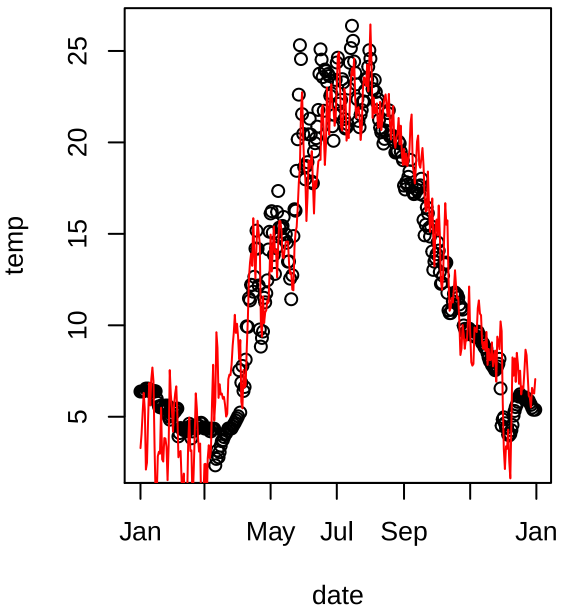

bigcreek : 2003 (RMSE: 2.03)
temp
0
5
10
15
20
Jan
May
Jul
Sep
Jan
date


bigotter : 2013 (RMSE: 2.97)
temp
5
10
15
20
Jan
May
Jul
Sep
Jan
date

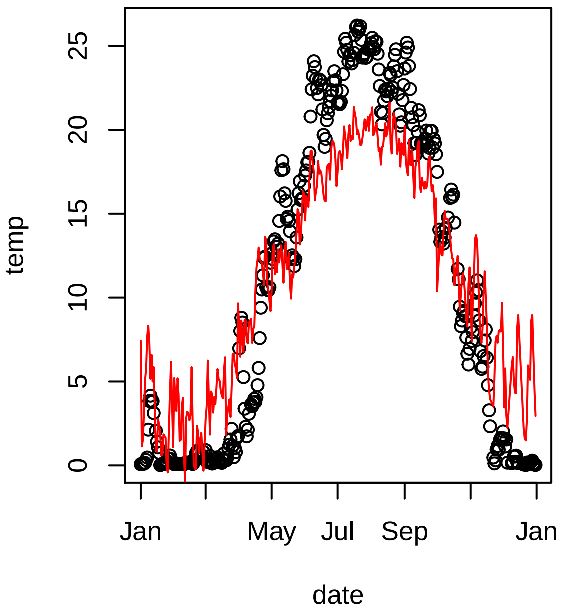

fox : 2014 (RMSE: 3.53)
temp
0
5
10
15
20
25
Jan
May
Jul
Sep
Jan
date


genesee : 2013 (RMSE: 3.13)
temp
0
5
10
15
20
25
Jan
May
Jul
Sep
Jan
date

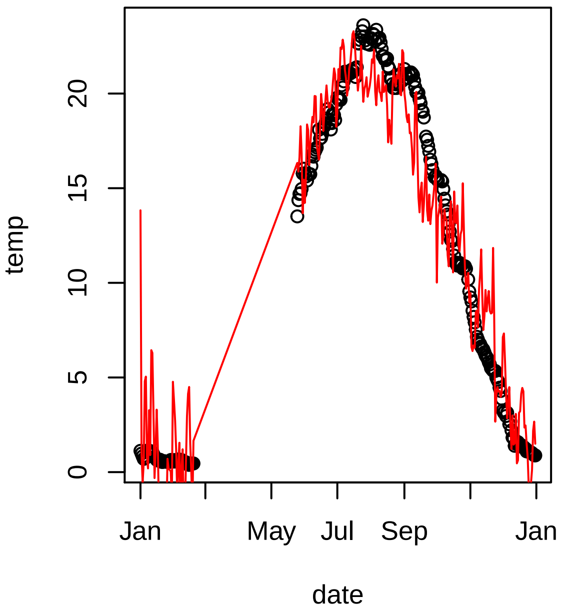
humber : 2006 (RMSE: 2.14)



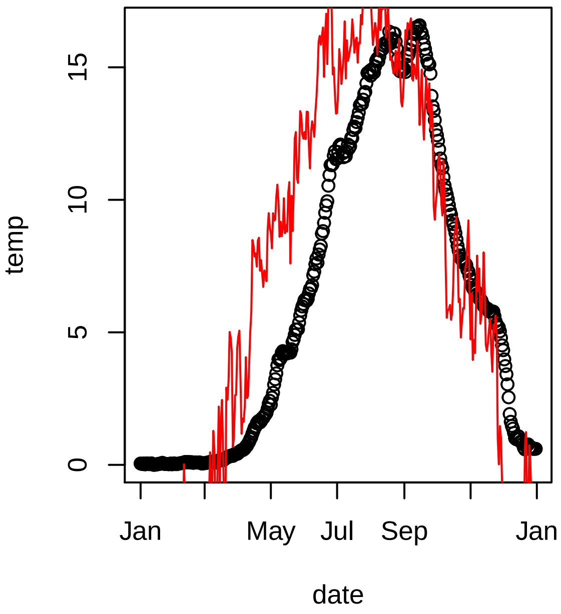
longpoint : 2008 (RMSE: 3.86)

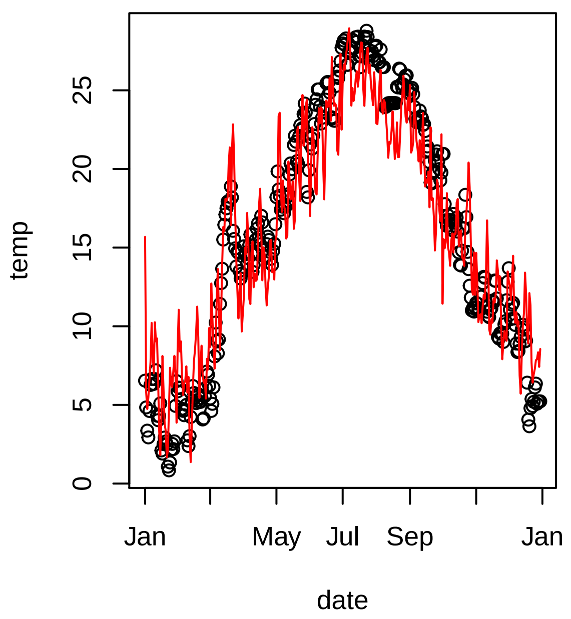


mississagi : 2012 (RMSE: 2.49)

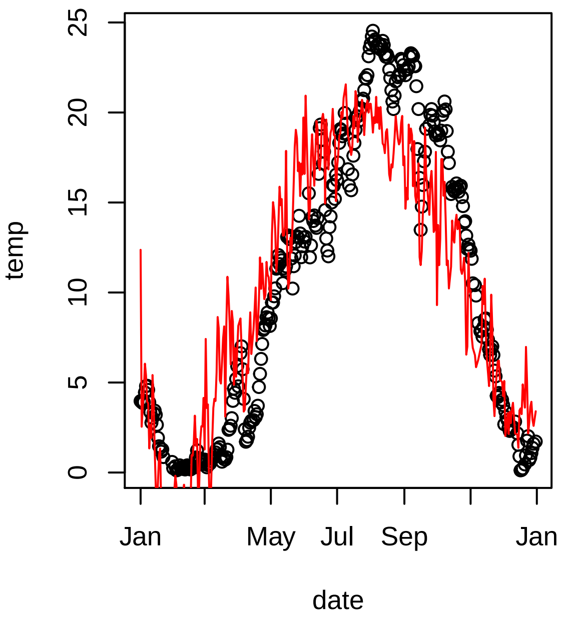


nipigon : 2009 (RMSE: 4.52)

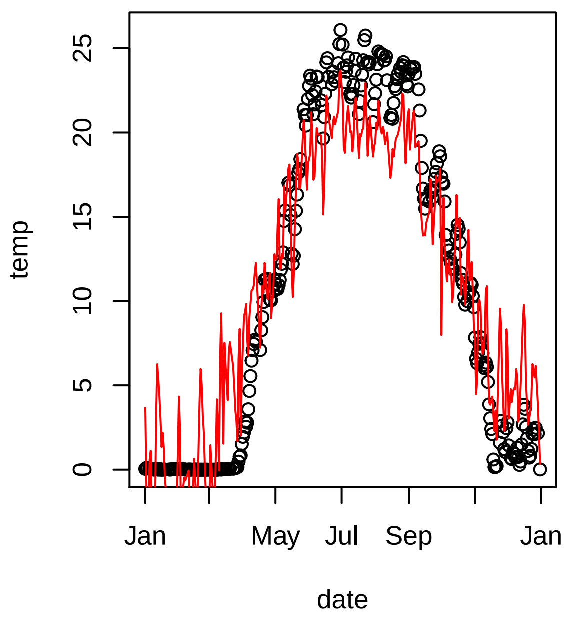




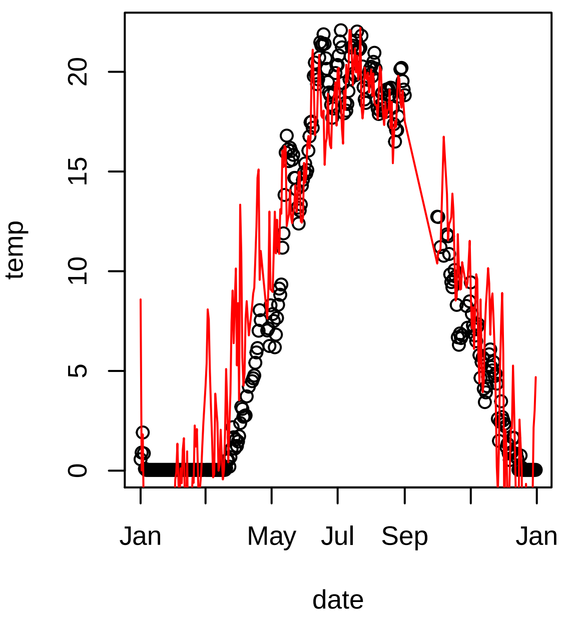
portage : 2012 (RMSE: 2.84)



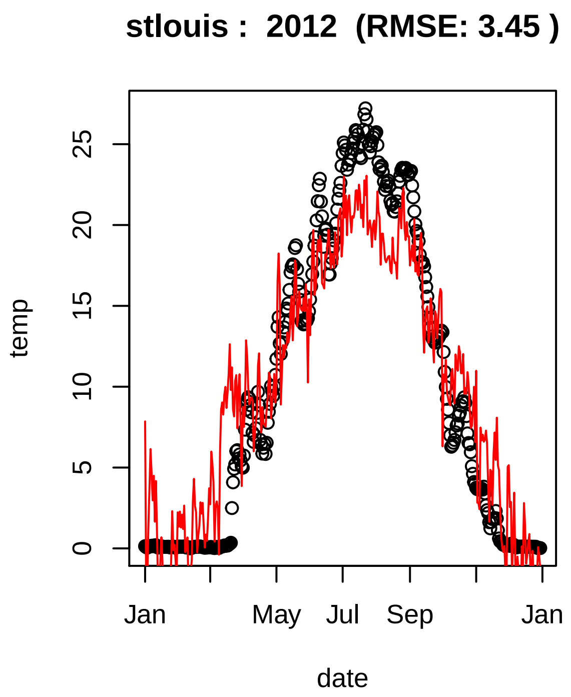
portdover : 2007 (RMSE: 3.62)



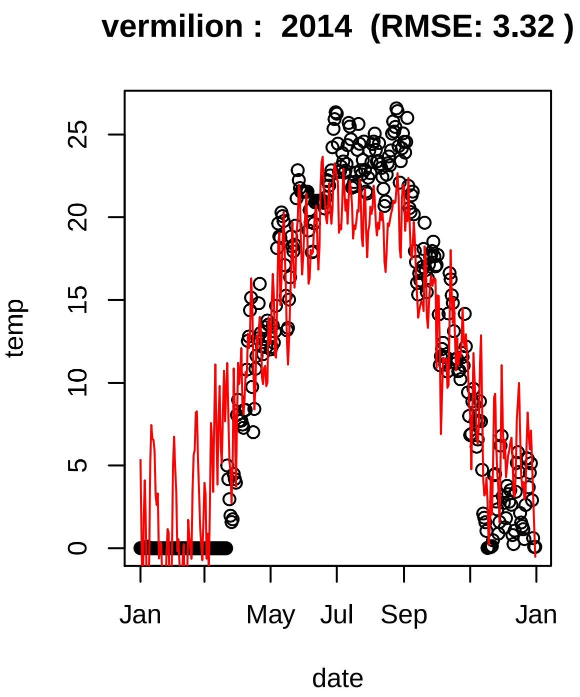
saginaw : 2014 (RMSE: 3.26)



still : 2004 (RMSE: 3.69)

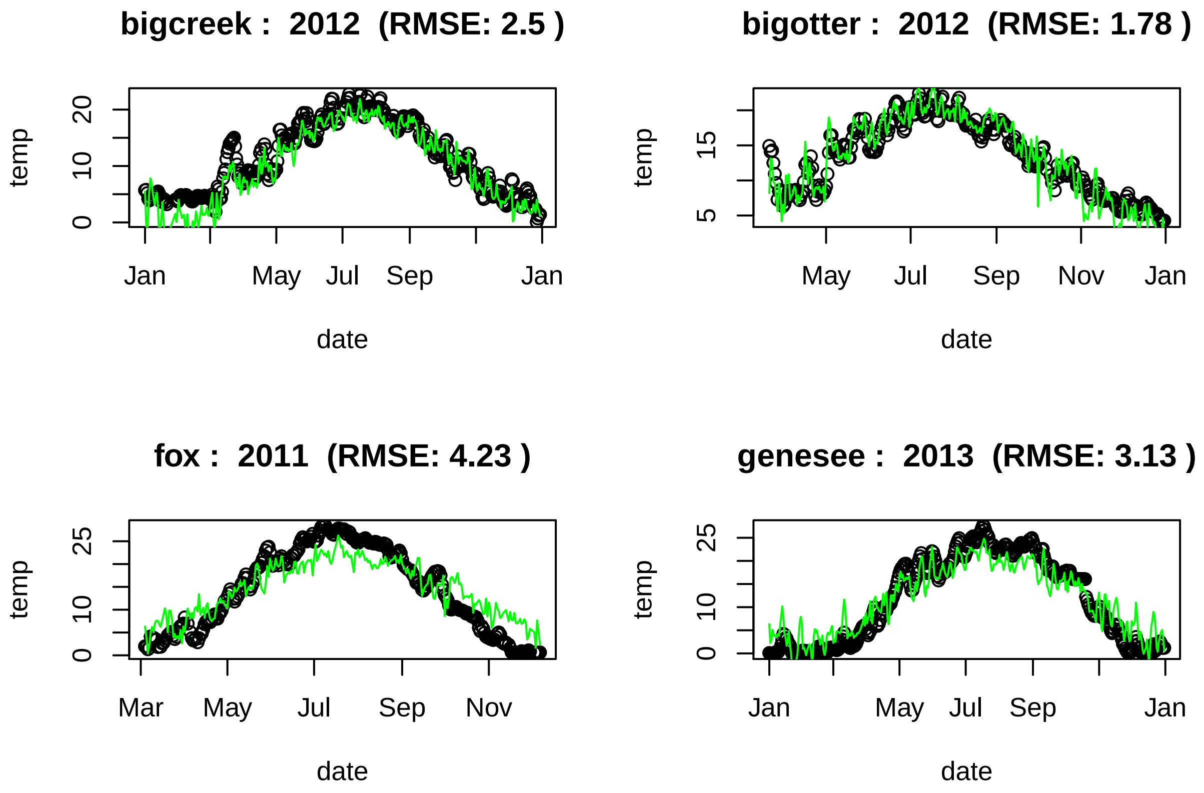


stlouis : 2012 (RMSE: 3.45)

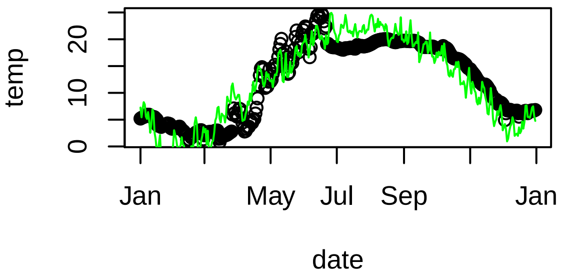


vermilion : 2014 (RMSE: 3.32)

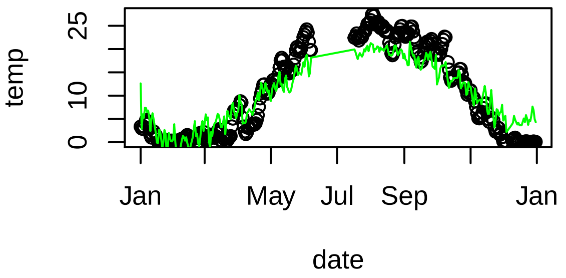
```
# test on new data
dfind_test=data.frame(location=character(), year=integer(), stringsAsFactors = FALSE)

for (i in 1:nrow(dfind)){
    test_site=full_year[dfind[i,1]]
    test_years=test_site[test_site!= dfind[i,2]]
    indxp=(sample(test_years[[1]], 1))
    dfind_test[i,]=c(dfind[i,1], indxp)

  }


aw_test=merge(aw, dfind_test, by=c("location", "year"))
ftest=predict(fm2, newdata = aw_test)
aw_test$test=ftest

# plot results of test
par(mfrow=c(2,2))
for (i in 1:nrow(dfind_test)){
  sub=aw_test[(aw_test$location==dfind_test$location[i]),]
  sub=sub[order(sub$date),]
  diff=round(sqrt(mean((sub$temp-sub$test)^2)),2)
  plot(temp~date, data=sub,
       main=paste(dfind$location[i], ": ", dfind_test$year[i],
                  " (RMSE:", diff, ")"))
  lines(test~date, data=sub,col="green", type="l")
}
```



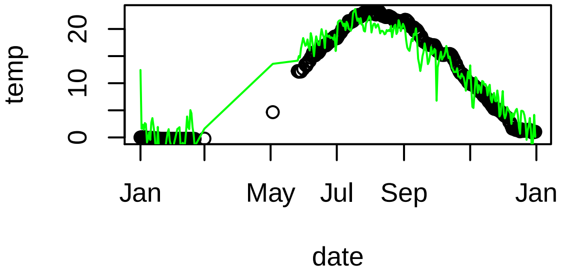
humber : 2007 (RMSE: 3.11)



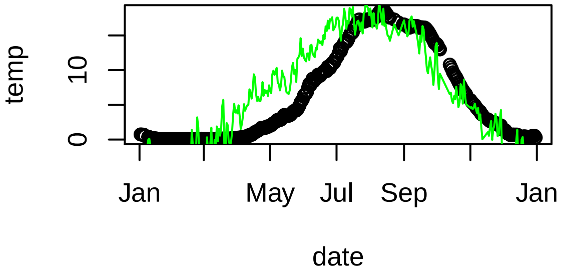
longpoint : 2007 (RMSE: 3.59)



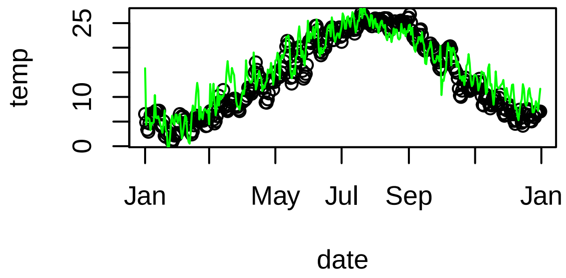
mississagi : 2011 (RMSE: 2.69)



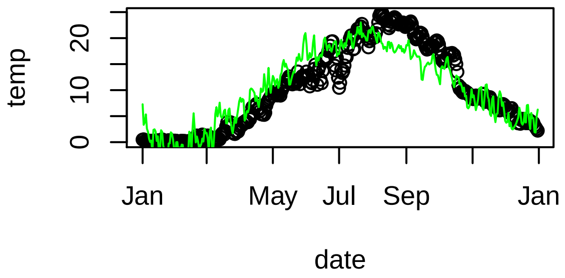
nipigon : 2002 (RMSE: 4.18)



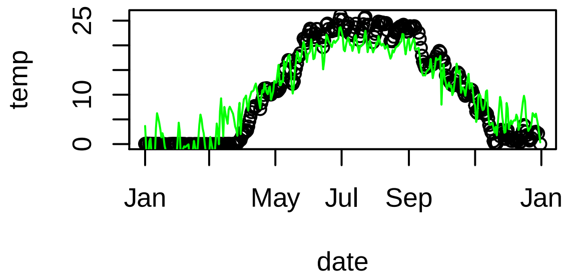
portage : 2011 (RMSE: 3.03)



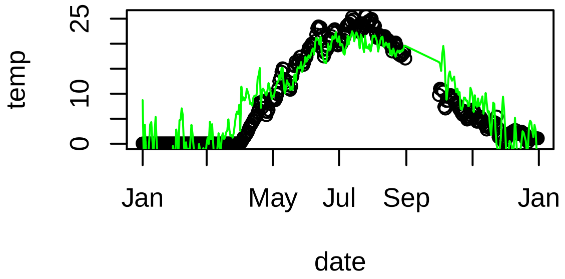
portdover : 2011 (RMSE: 3.39)

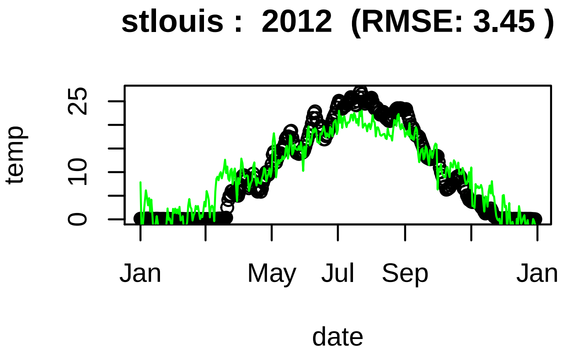


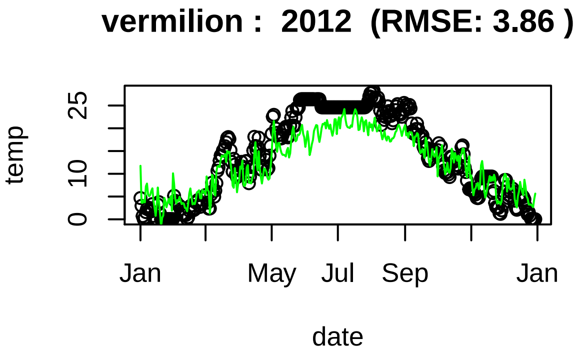
saginaw : 2014 (RMSE: 3.26)



still : 2005 (RMSE: 3.66)




stlouis : 2012 (RMSE: 3.45)
temp
0
10
25
Jan
May
Jul
Sep
Jan
date


vermilion : 2012 (RMSE: 3.86)
temp
0
10
25
Jan
May
Jul
Sep
Jan
date