

Lab 1 Forecasting with ARIMA models

Nick Chambers (SAFS), Liz Elmstrom (SAFS), Maria Kuruvilla (QERM)

March 22, 2023

Contents

Question your team will address	2
Method you will use	2
Initial plan	2
What you actually did	2
Preliminary dataset exploration	2
Read in and explore data (Liz)	2
Data diagnostics	5
Plot the data (Liz)	5
Use ACF and PACF (Maria)	6
Test for stationarity (Nick & Liz)	12
Difference tests (Liz)	13
Modeling and results	14
Dividing the data into test and train (Maria)	14
Fit ARIMA models (Maria)	14
Plot forecasts from auto.arima (Liz)	15
Choosing our own ARIMA structure (Liz)	18
Check accuracy (Nick)	21
Check residuals (Nick)	22
Discussion	26
Team member contributions	26

Question your team will address

In this report we sought to explore how ARIMA models perform when forecasting fish abundance in the near term (5-year forecasts).

More specifically, we ask do ARIMA models produce more accurate 5-year forecasts than the Fisheries Research Institute (FRI) for Sockeye in select Bristol Bay Rivers?

Method you will use

We will fit multiple ARIMA models to the Bristol Bay sockeye data and compare forecasts among these models (using RMSE) and visually compare the forecasts against the FRI forecast.

Initial plan

We divided the rivers among the team members to do the initial exploratory analysis and testing for stationarity. We then tried forecasting (5 years) for all rivers using the best model suggested by `auto.arima()`.

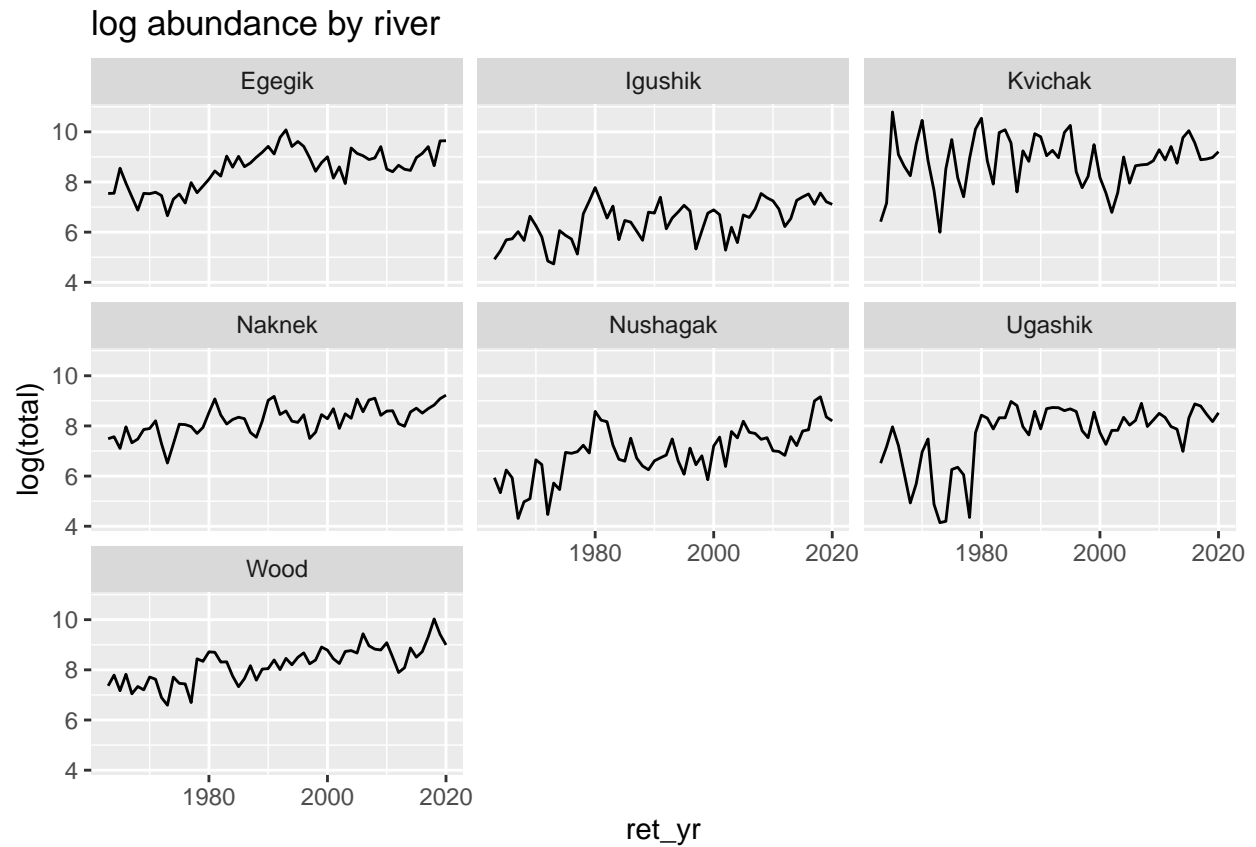
What you actually did

We finally chose three rivers to work with - Kvichak, Wood and Nugashak. We split the data into training data (1963-2015) and testing data (2016-2020) and performed forecasts using the `auto.arima` function. In addition to forecasting with the model suggested by `auto.arima`, we explored additional top models suggested by the `auto.arima` function with `trace == TRUE`. We then compared the accuracy of these modeled forecasts using RMSE. We also plotted the FRI forecast to visually compare the model forecast to the FRI forecast.

Preliminary dataset exploration

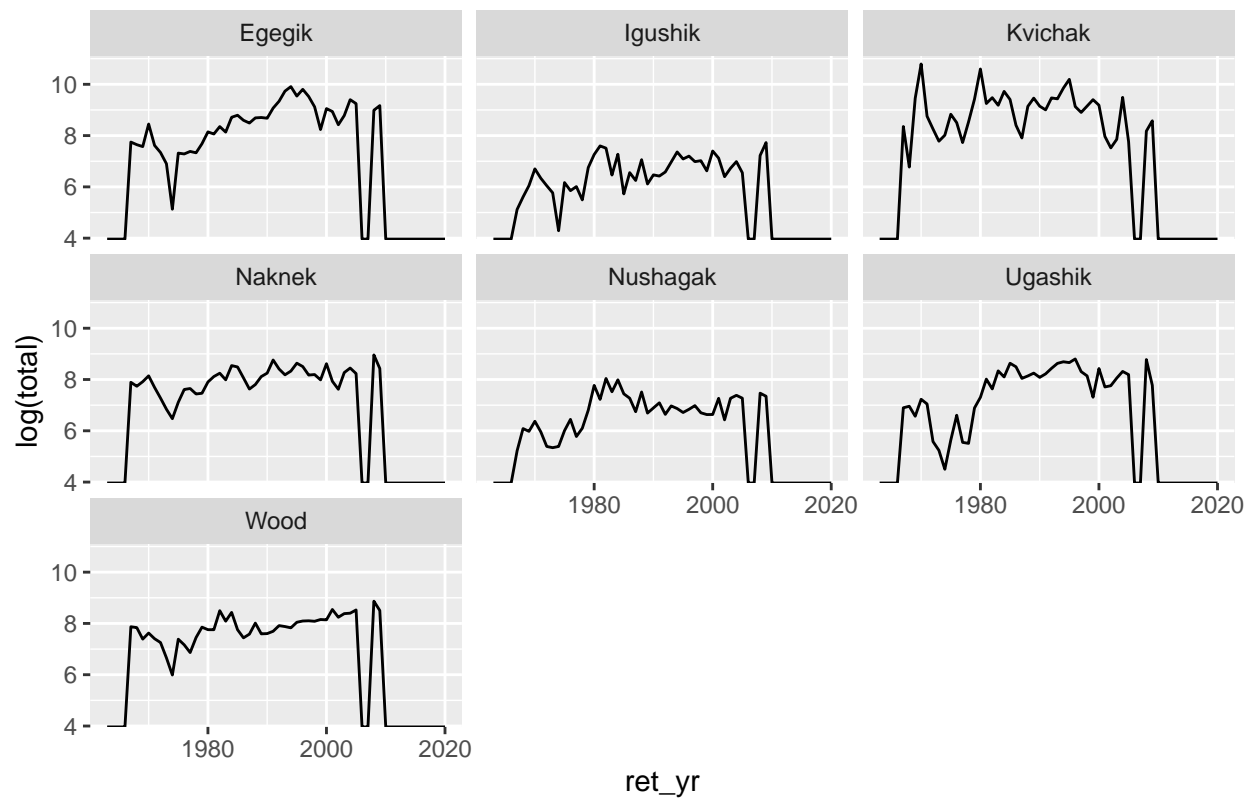
Read in and explore data (Liz)

```
bb_data %>%
  group_by(system, ret_yr) %>%
  summarize(total = sum(ret, na.rm=TRUE)) %>%
  ggplot(aes(x=ret_yr, y=log(total))) +
  geom_line() +
  ggtitle("log abundance by river") +
  facet_wrap(~system)
```



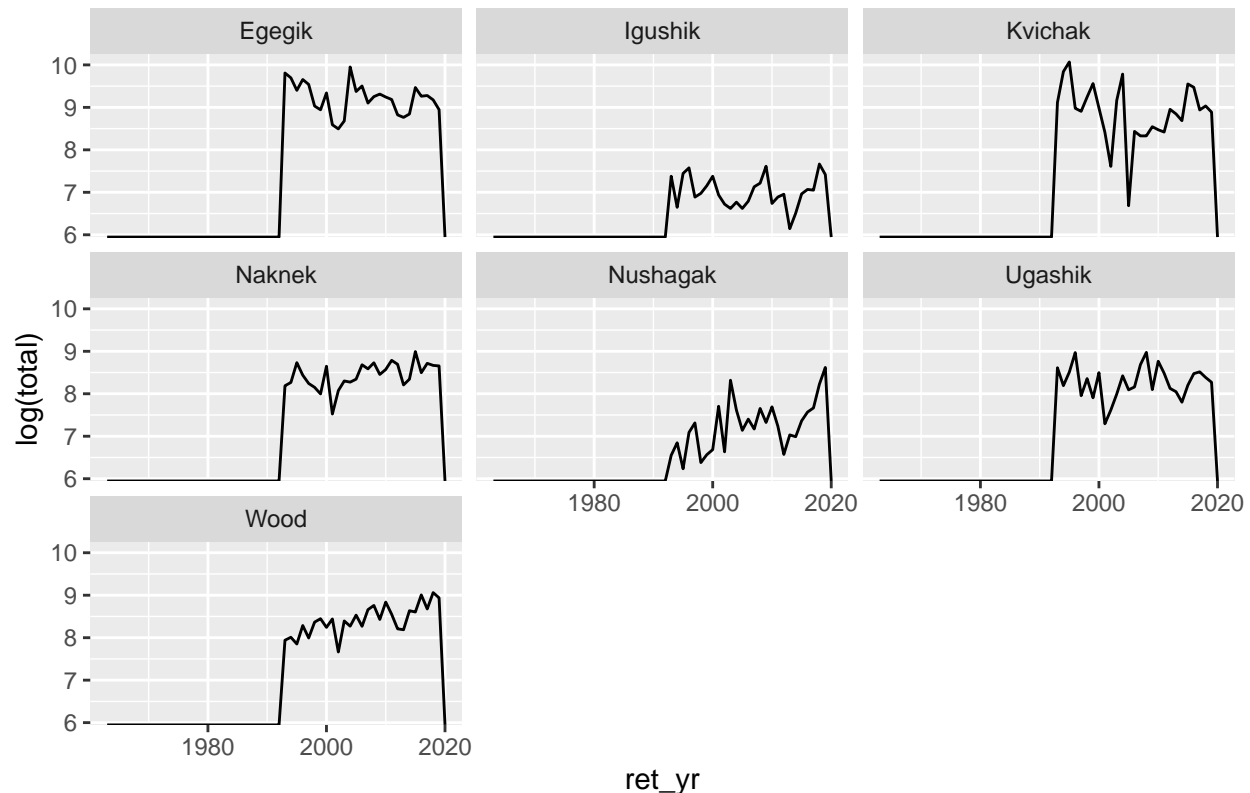
```
bb_data %>%
  group_by(system, ret_yr) %>%
  summarize(total = sum(forecast.adfw, na.rm=TRUE)) %>%
  ggplot(aes(x=ret_yr, y=log(total))) +
    geom_line() +
    ggtitle("ADFW log abundance by river") +
    facet_wrap(~system)
```

ADFW log abundance by river



```
bb_data %>%
  group_by(system, ret_yr) %>%
  summarize(total = sum(forecast.fri, na.rm=TRUE)) %>%
  ggplot(aes(x=ret_yr, y=log(total))) +
    geom_line() +
    ggtitle("FRI log abundance by river") +
    facet_wrap(~system)
```

FRI log abundance by river



After exploring the larger dataset, we chose to explore temporal sockeye dynamics (specifically, total sockeye abundance) in three rivers– the Kvichak, the Nushagak, and the Wood.

Specifically, we seek to – 1) identify autocorrelation structures and stationarity for each of the total sockeye abundance time series, 2) compare ARIMA models, forecasts, and forecast accuracies for 5 year forecasts from 2015-2020, and 3) visually compare to these forecasts to forecasts from UW Fisheries Research Institute.

We did not compare our models to ADFW forecasts since the ADFW forecasts occur prior to our 5-year forecasts (2015-2020).

Data diagnostics

Plot the data (Liz)

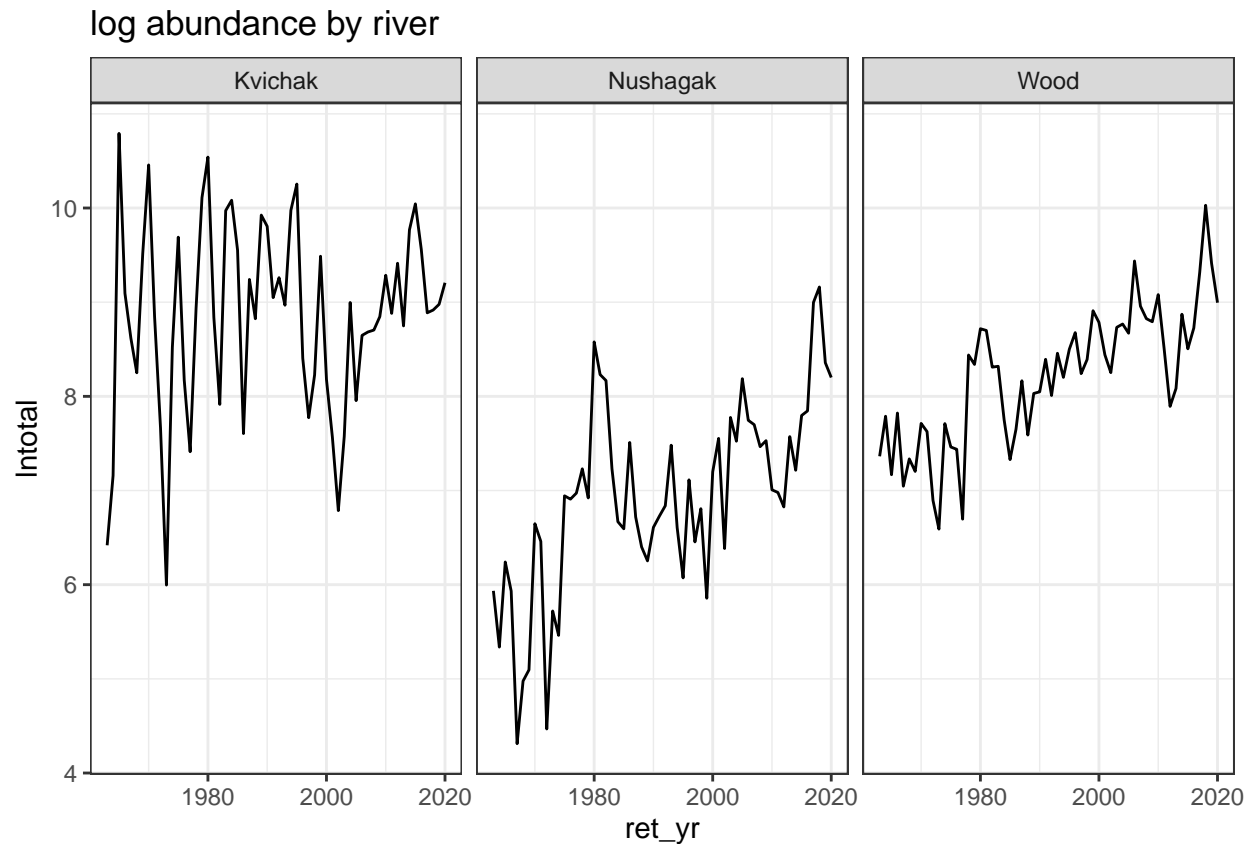
Plot the data and discuss any obvious problems with stationarity from your visual test.

```
#Sum and log returns for each river and year
lndata <- bb_data %>%
  group_by(system, ret_yr) %>%
  summarize(lntotal = log(sum(ret, na.rm=TRUE)))

# Choosing sites CHANGE THIS BASED ON WHAT WE PICK
our_rivers <- filter(lndata, system %in% c('Kvichak', 'Wood', 'Nushagak'))

# Plotting three rivers
```

```
our_rivers %>%
  ggplot(aes(x=ret_yr, y=Intotal)) +
    geom_line() +
    ggtitle("log abundance by river") +
    facet_wrap(~system)+theme_bw()
```



Based on these plots, we hypothesize that data will be non-stationary for the Kvichak and stationary for the Nushagak and the Wood around a trend.

Use ACF and PACF (Maria)

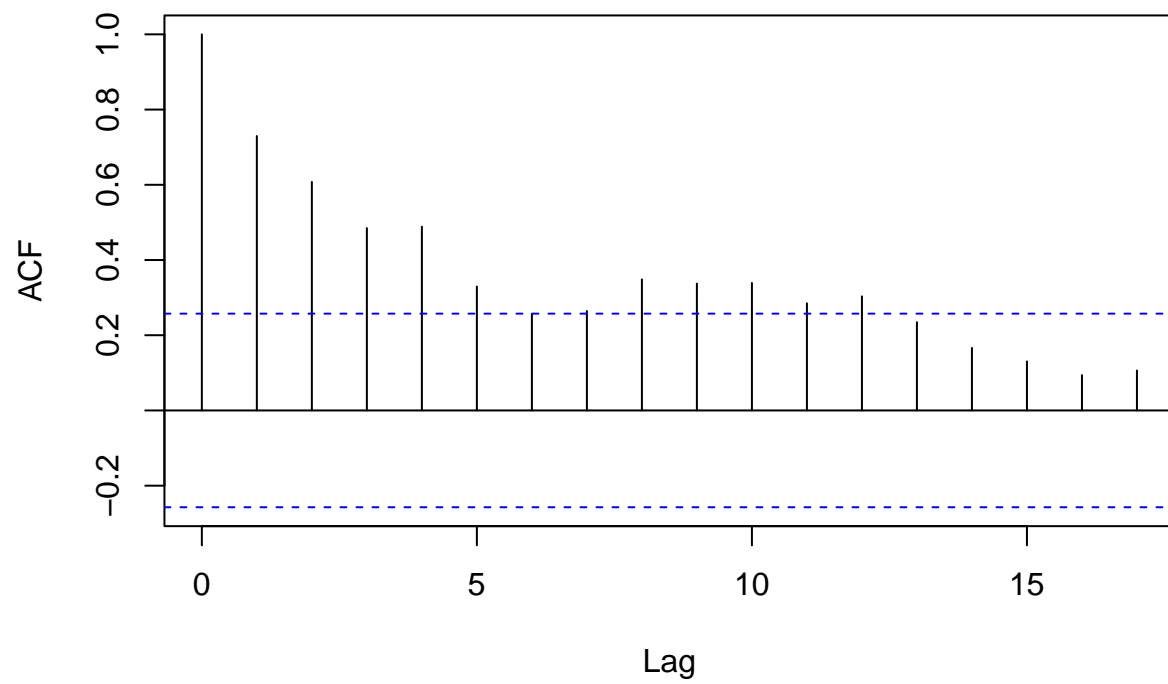
```
data_ts_Kv <- ts(our_rivers$Intotal[our_rivers$system == "Kvichak"],
  start=our_rivers$ret_yr[our_rivers$system == "Kvichak"][1])

data_ts_Wood <- ts(our_rivers$Intotal[our_rivers$system == "Wood"],
  start=our_rivers$ret_yr[our_rivers$system == "Wood"][1])

data_ts_Nu <- ts(our_rivers$Intotal[our_rivers$system == "Nushagak"],
  start=our_rivers$ret_yr[our_rivers$system == "Nushagak"][1])

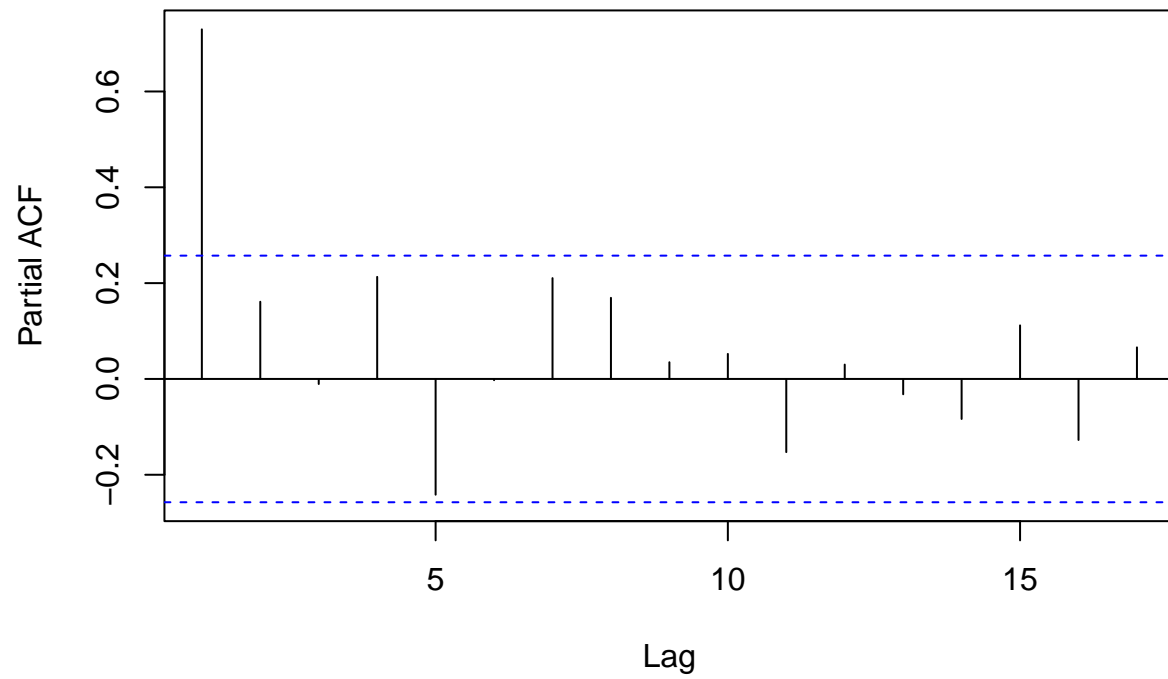
acf(data_ts_Wood)
```

Series data_ts_Wood



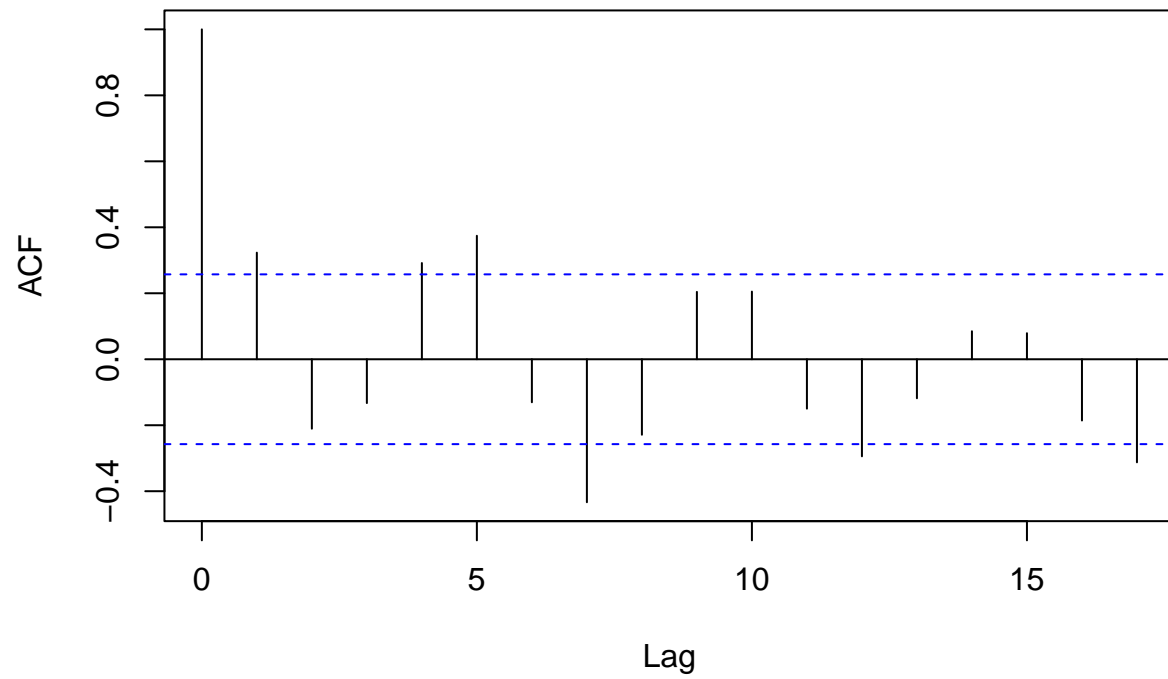
```
pacf(data_ts_Wood)
```

Series data_ts_Wood

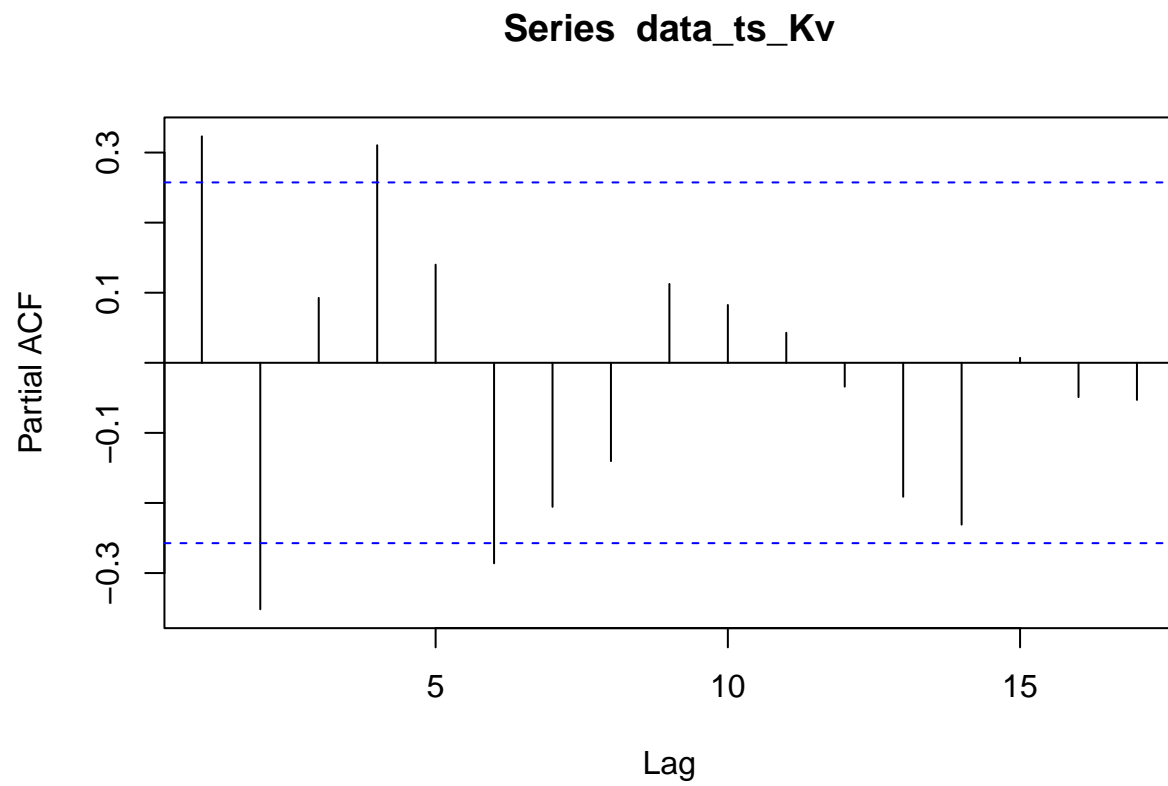


```
acf(data_ts_Kv)
```


Series data_ts_Kv

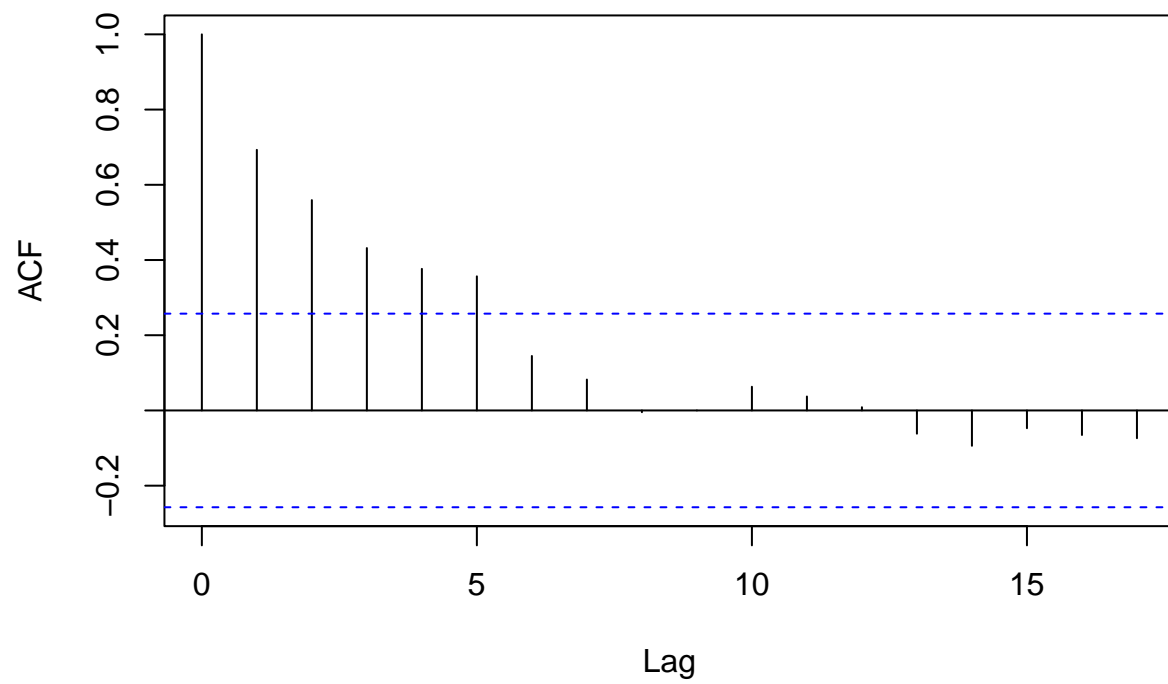


```
pacf(data_ts_Kv)
```



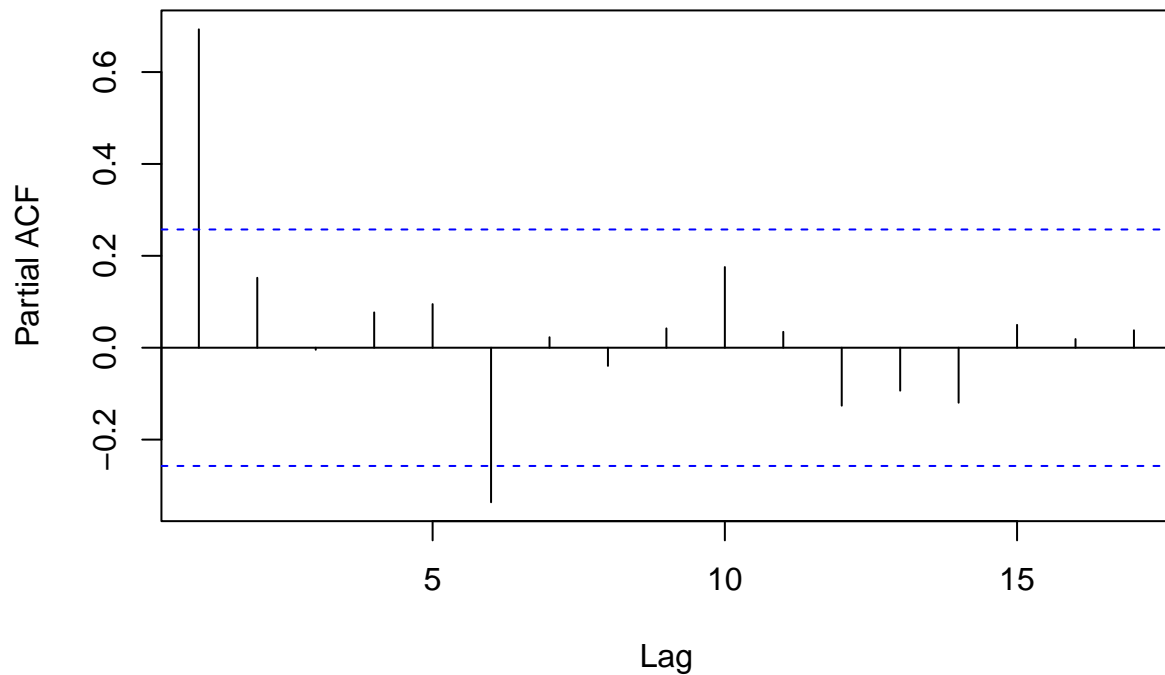
```
acf(data_ts_Nu)
```

Series data_ts_Nu



```
pacf(data_ts_Nu)
```

Series data_ts_Nu



The acf plots for Wood river is slowly decaying and the pacf plot shows significance at lag 1. There might be AR1 structure in the data.

The pacf plots for Kvichak river is slowly decaying and the pacf plot shows significance at lags 1,2,4, and 6. There might be MA structure in the data.

The acf plots for Nushagak river is slowly decaying and the pacf plot shows significance at lags 1 and 6. There might be AR structure in the data.

Test for stationarity (Nick & Liz)

#The Dickey fuller test- looks for evidence that the t.s. are a random walk. The null hypothesis for bo

#The KPSS test- The null hypothesis for the KPSS test is that the data are stationary. For this test, w

```
adf.list <- kpss.list <- kpss.trend <- list()
p.vals <- kp.vals.lev <- kp.vals.tre <- vector()
River <- unique(our_rivers$system)

for(i in 1:length(River)){
  sub<-subset(our_rivers,our_rivers$system==River[i])

  suppressWarnings(adf.list[[i]] <- adf.test(sub$lntotal, k = 0))
  suppressWarnings(kpss.list[[i]] <- kpss.test(sub$lntotal, null = "Level"))
  suppressWarnings(kpss.trend[[i]] <- kpss.test(sub$lntotal, null = "Trend"))
```

```

p.vals[[i]] <- adf.list[[i]]$p.value # Should be less than .05
kp.vals.lev[i] <- kpss.list[[i]]$p.value # Should be greater than .05
kp.vals.tre[i] <- kpss.trend[[i]]$p.value # Should be greater than .05
}

#For the ADF test, we want to REJECT the null hypothesis, so we want a p-value less than 0.05
#For the KPSS test, we want to ACCEPT the null hypothesis, so we want a p-value greater than 0.05

station_mat <- matrix(data =c(p.vals,kp.vals.lev,kp.vals.tre), 3, 3)
colnames(station_mat) <- c("adf","kpss_level","kpss_trend")
rownames(station_mat) <- River
station_mat

```

```

##           adf kpss_level kpss_trend
## Kvichak  0.01      0.10      0.1
## Nushagak 0.01      0.01      0.1
## Wood     0.01      0.01      0.1

```

The stationarity tests show that the Kvichak River time series is stationary.

When testing for stationarity around a level using the KPSS test, the Nushagak and the Wood test as non-stationary. However, when we change the null to include a trend, we find that both of these rivers test as stationary around a trend.

Difference tests (Liz)

```

kv <- filter(lndata, system %in% c("Kvichak"))
ndiffs(kv$lntotal, test='kpss')

```

```
## [1] 0
```

```
ndiffs(kv$lntotal, test='adf')
```

```
## [1] 0
```

```

nu <- filter(lndata, system %in% c("Nushagak"))
ndiffs(nu$lntotal, test='kpss')

```

```
## [1] 1
```

```
ndiffs(nu$lntotal, test='adf')
```

```
## [1] 1
```

```

wo <- filter(lndata, system %in% c("Wood"))
ndiffs(wo$lntotal, test='kpss')

```

```
## [1] 1
```

```
ndiffs(wo$lntotal, test='adf')
```

```
## [1] 1
```

According to these tests, The Kvichak does not need differencing. The Nushagak and the Wood should be differenced by 1. This level of differencing should remove the upward trend or “bias” in the Nushagak and Wood Rivers.

Modeling and results

Dividing the data into test and train (Maria)

```
data_ts_Wood <- ts(our_rivers$lntotal[our_rivers$system == "Wood"],
                  start=our_rivers$ret_yr[our_rivers$system == "Wood"][1])

train_Wood <- window(data_ts_Wood, 1963, 2015)
test_Wood <- window(data_ts_Wood, 2016, 2020)

data_ts_Kv <- ts(our_rivers$lntotal[our_rivers$system == "Kvichak"],
                 start=our_rivers$ret_yr[our_rivers$system == "Kvichak"][1])

train_Kv <- window(data_ts_Kv, 1963, 2015)
test_Kv <- window(data_ts_Kv, 2016, 2020)

data_ts_Nu <- ts(our_rivers$lntotal[our_rivers$system == "Nushagak"],
                 start=our_rivers$ret_yr[our_rivers$system == "Nushagak"][1])

train_Nu <- window(data_ts_Nu, 1963, 2015)
test_Nu <- window(data_ts_Nu, 2016, 2020)
```

Fit ARIMA models (Maria)

```
mod_Kv <- auto.arima(train_Kv)
mod_Kv
```

```
## Series: train_Kv
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1      mean
##          0.4784  8.8150
## s.e.  0.1166  0.1922
##
## sigma^2 = 0.9421: log likelihood = -72.73
## AIC=151.47   AICc=151.96   BIC=157.38
```

```
mod_Nu <- auto.arima(train_Nu)
mod_Nu
```

```
## Series: train_Nu
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##        -0.5194
## s.e.    0.1324
##
## sigma^2 = 0.5102: log likelihood = -55.94
## AIC=115.88   AICc=116.13   BIC=119.79
```

```
mod_Wood <- auto.arima(train_Wood)
mod_Wood
```

```
## Series: train_Wood
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##        -0.5434
## s.e.    0.1460
##
## sigma^2 = 0.2091: log likelihood = -32.77
## AIC=69.54   AICc=69.78   BIC=73.44
```

The best model according to model selection criterion (AICc) for the Kvichak river is an ARIMA(0,0,1) with a non-zero mean. The best model for the Nushagak and the Wood river is an ARIMA (0,1,1).

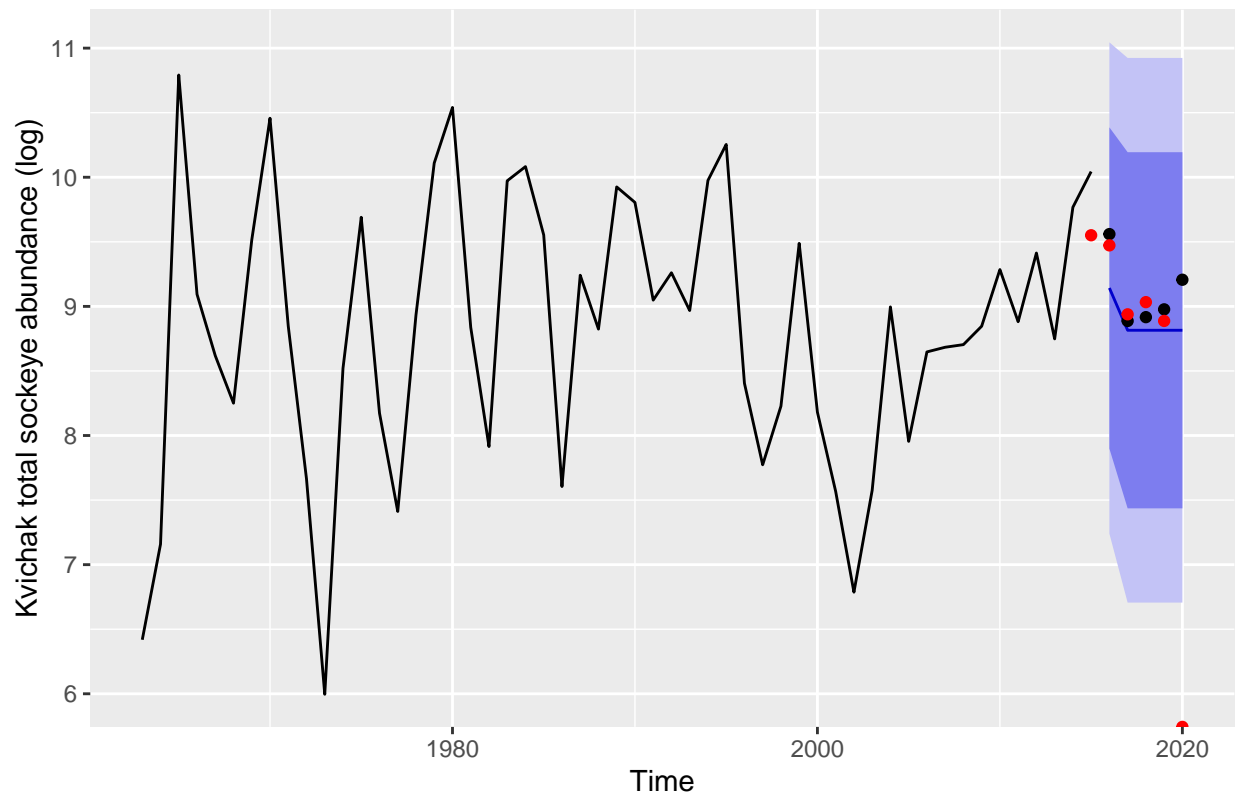
Plot forecasts from auto.arima (Liz)

Below are plots of the total abundance measurements and their 5-year forecasts. The red points refer to Fishery Research Institutes model forecasts. The black points refer to actual data.

```
fri_fore <- bb_data %>%
  group_by(system, ret_yr) %>%
  summarize(lntotal = log(sum(forecast.fri, na.rm=TRUE)))%>%
  filter(ret_yr %in% 2015:2020)

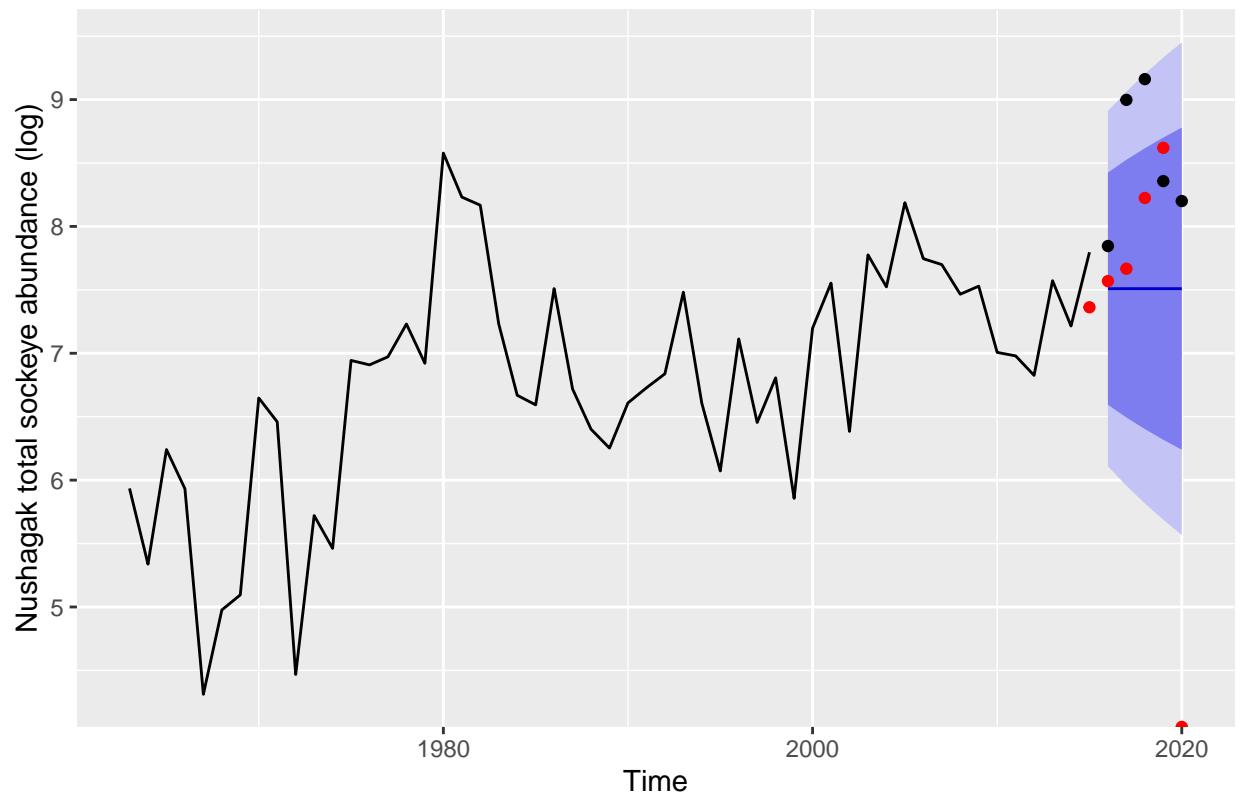
kv_fri_fore <- fri_fore %>% filter(system %in% 'Kvichak')
kv_fr <- forecast(mod_Kv, h=5)
autoplot(kv_fr) + geom_point(aes(x=x, y=y), data=fortify(test_Kv))+
  geom_point(aes(x=ret_yr, y=lntotal), data = kv_fri_fore, col = "red")+
  ylab("Kvichak total sockeye abundance (log)")
```

Forecasts from ARIMA(0,0,1) with non-zero mean

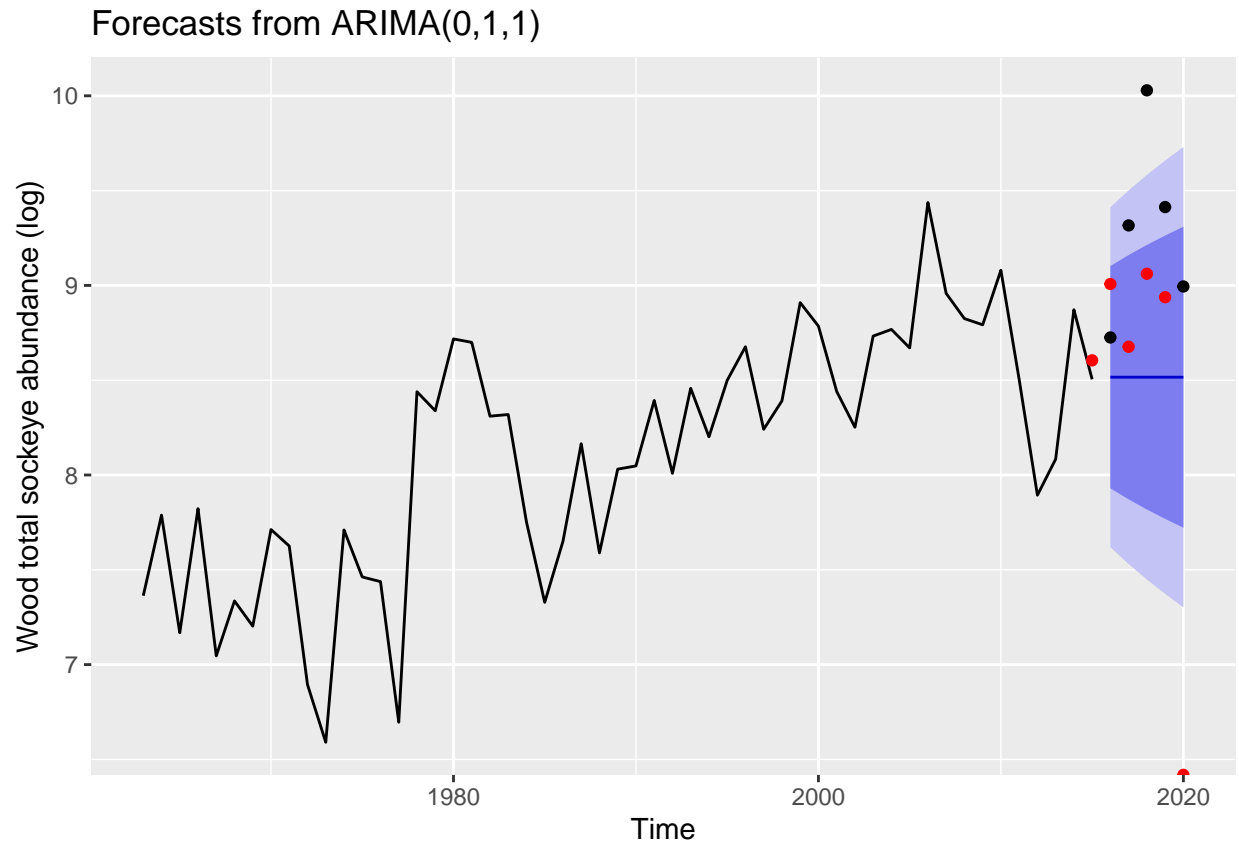


```
nu_fri_fore <- fri_fore %>% filter(system %in% 'Nushagak')
nu_fr <- forecast(mod_Nu, h=5)
autoplot(nu_fr) + geom_point(aes(x=x, y=y), data=fortify(test_Nu))+
  geom_point(aes(x=ret_yr, y=lnntotal), data = nu_fri_fore, col='red')+
  ylab("Nushagak total sockeye abundance (log)")
```


Forecasts from ARIMA(0,1,1)



```
wo_fri_fore <- fri_fore %>% filter(system %in% 'Wood')
wo_fr <- forecast(mod_Wood, h=5)
autoplot(wo_fr) + geom_point(aes(x=x, y=y), data=fortify(test_Wood))+
  geom_point(aes(x=ret_yr, y=ln_total), data = wo_fri_fore, col = "red")+
  ylab("Wood total sockeye abundance (log)")
```



Though these test as the best models using the `auto.arima()` function, we can clearly see that these models do not necessarily provide the best forecasts for the rivers of study. This is particularly apparent for the Nushagak and for the Wood River— the two rivers that include an upward trend or “bias”.

Chosing our own ARIMA structure (Liz)

Here we play around with different ARIMA structures on our own. Though not shown in the code, we investigated other top models using the `auto.arima()` function with `trace = TRUE`. We then tested the three top models based on AICc for the Kvichak. For the Nushagak and the Wood, we tested the two top models based on AICc, plus the best model with an added drift component.

Chosing our own arima structure in a for loop. Probably a cleaner way to do this.

```
kv_mod_list <- list(c(0,0,1), c(1,0,1), c(0,0,2))## List of mods to test. First option is what auto.arima
kv_fits <- kv_fore <- kv_plots <- list()# lists to store things in
kv_err_list <- list(0) #lists for error structure

nu_mod_list <- list(c(0,1,1), c(0,1,1), c(1,1,1))
nu_fits <- nu_fore <- nu_plots <- list()
nu_err_list <- list(0)

wo_mod_list <- list(c(0,1,1), c(0,1,1), c(1,1,1))
wo_fits <- wo_fore <- wo_plots <- list()
wo_err_list <- list(0)
```

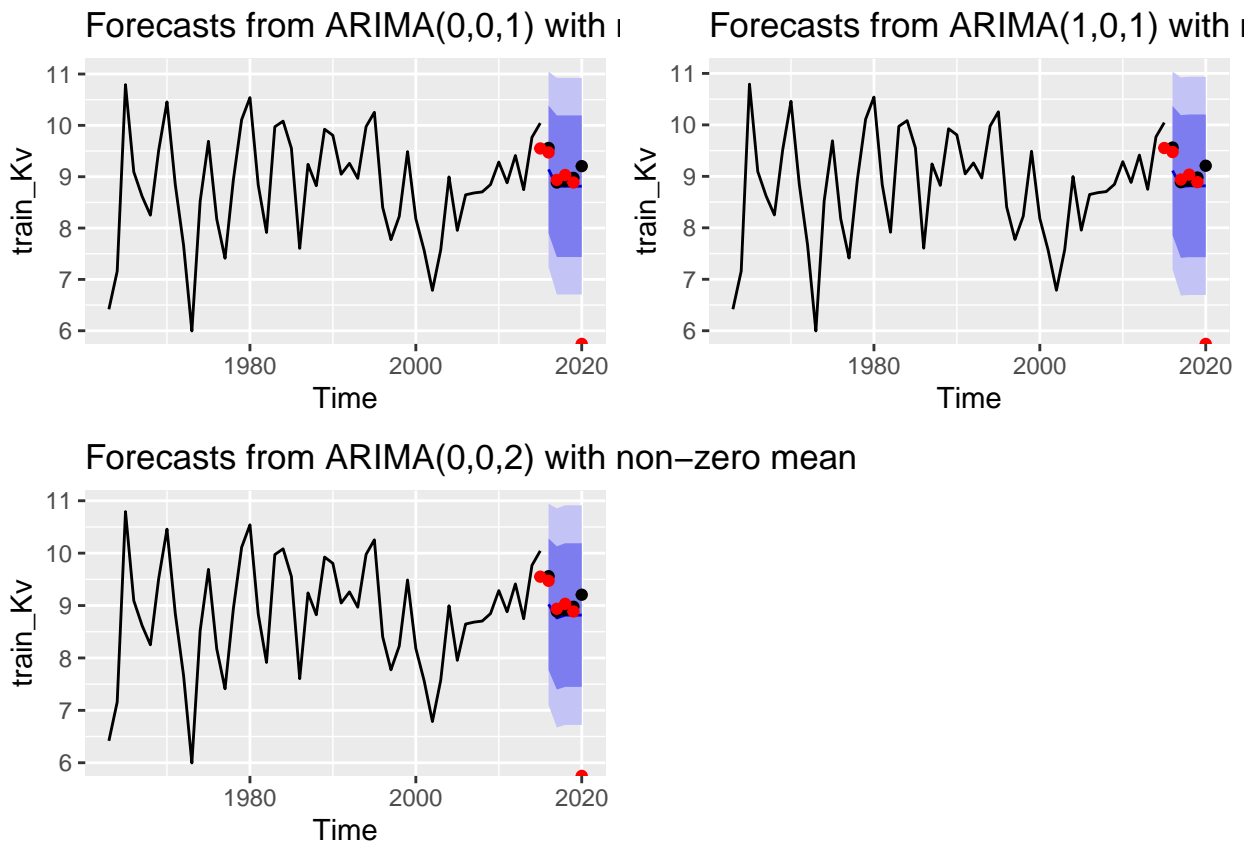
```

drift <- list(FALSE, TRUE, FALSE)## adding drift

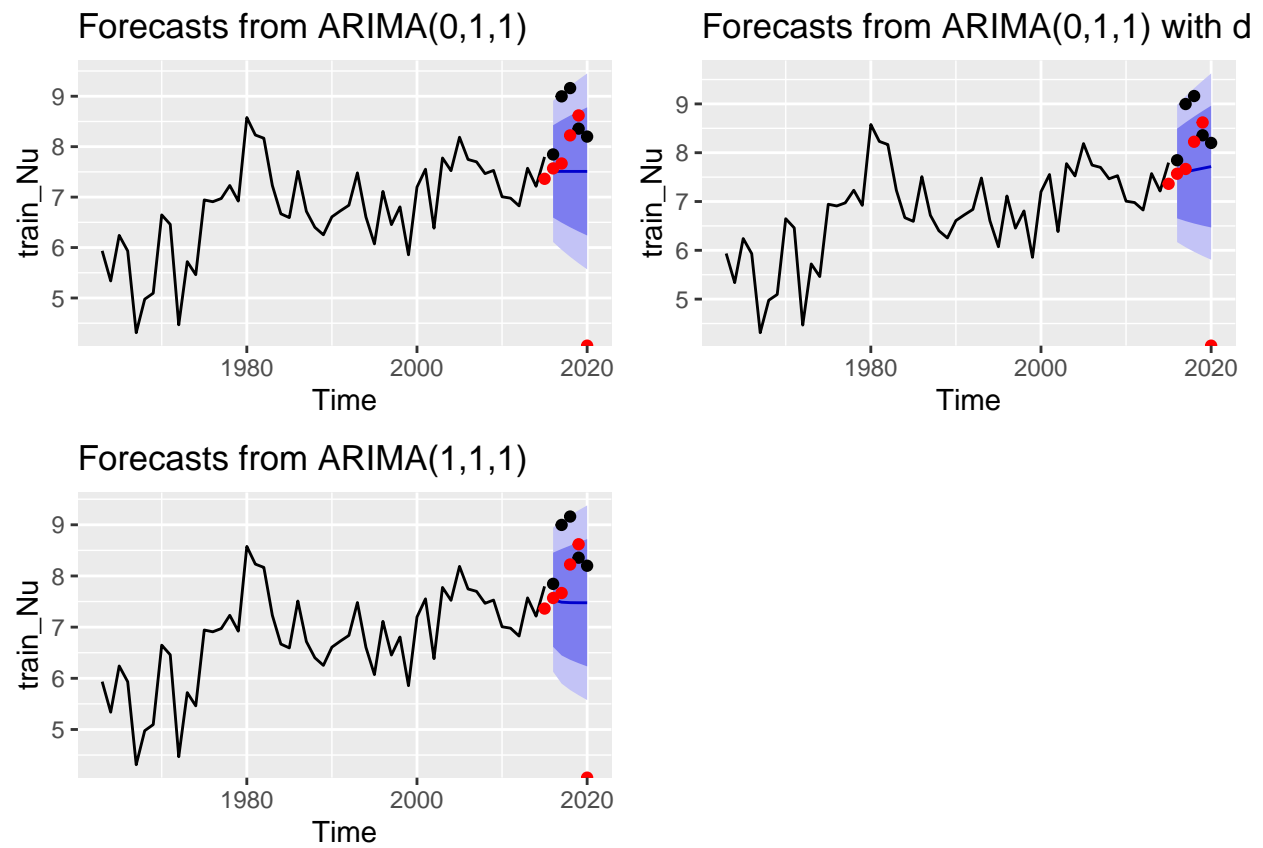
for(i in 1:3){
  #Kvichak
  kv_fits[[i]] <- Arima(train_Kv, order = kv_mod_list[[i]])# loops through list, fits model, and stores
  kv_fore[[i]] <- forecast(kv_fits[[i]], h=5) # forecasts
  kv_plots[[i]] <- (autoplot(kv_fore[[i]]) + geom_point(aes(x=x, y=y), data=fortify(test_Kv))+ #forecas
    geom_point(aes(x=ret_yr, y=lntotal), data = kv_fri_fore, col = "red"))
  kv_err_list[[i]] <- accuracy(forecast(kv_fore[[i]], h = 5), test_Kv) # puts accuracy data into list
  #Nushagak
  nu_fits[[i]] <- Arima(train_Nu, order = nu_mod_list[[i]], include.drift = drift[[i]])
  nu_fore[[i]] <- forecast(nu_fits[[i]], h=5)
  nu_plots[[i]] <- (autoplot(nu_fore[[i]]) + geom_point(aes(x=x, y=y), data=fortify(test_Nu))+ #forecas
    geom_point(aes(x=ret_yr, y=lntotal), data = nu_fri_fore, col = "red"))
  nu_err_list[[i]] <- accuracy(forecast(nu_fore[[i]], h = 5), test_Nu)
  #Wood
  wo_fits[[i]] <- Arima(train_Wood, order = wo_mod_list[[i]], include.drift = drift[[i]])
  wo_fore[[i]] <- forecast(wo_fits[[i]], h=5)
  wo_plots[[i]] <- (autoplot(wo_fore[[i]]) + geom_point(aes(x=x, y=y), data=fortify(test_Wood))+
    geom_point(aes(x=ret_yr, y=lntotal), data = wo_fri_fore, col = "red"))
  wo_err_list[[i]] <- accuracy(forecast(wo_fore[[i]], h = 5), test_Wood)
}

ggarrange(plotlist = kv_plots, ncol = 2, nrow =2)

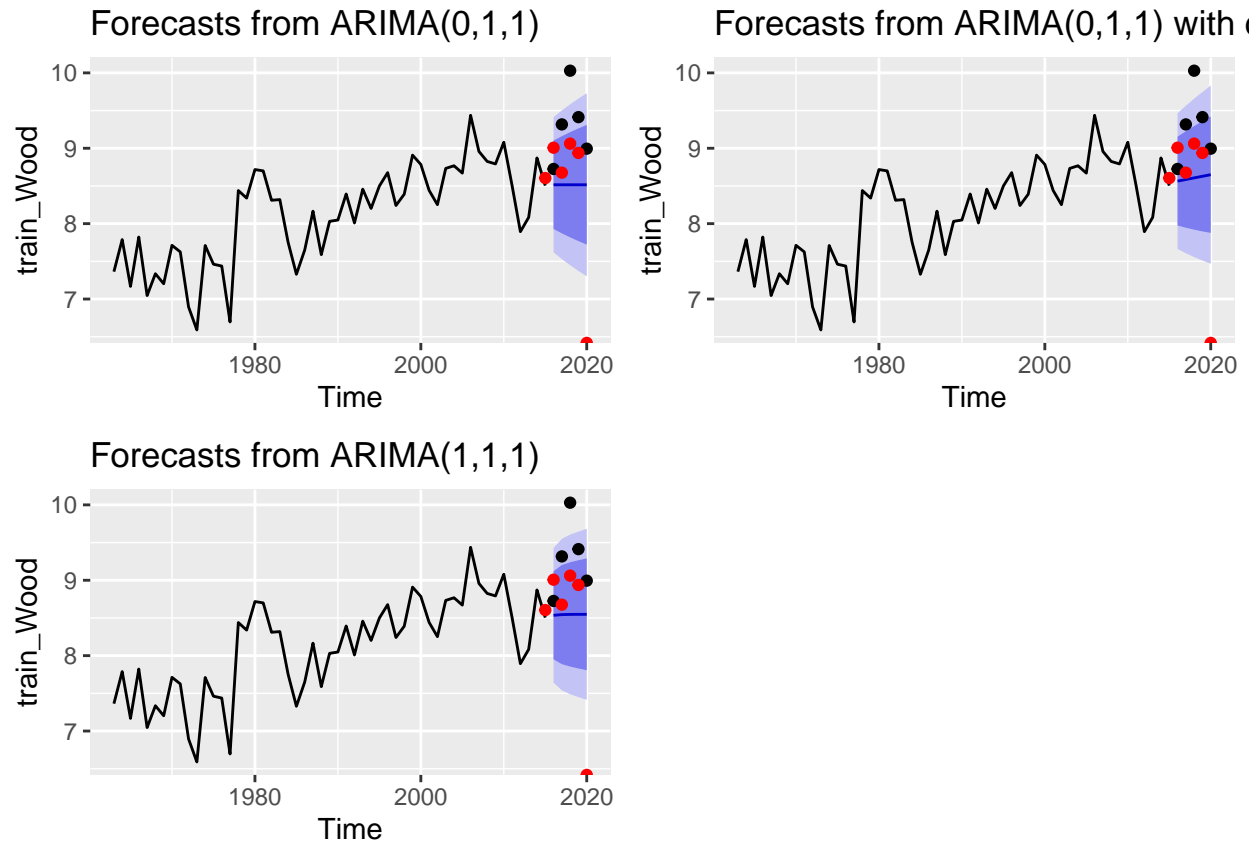
```



```
ggarrange(plotlist = nu_plots, ncol = 2, nrow = 2)
```



```
ggarrange(plotlist = wo_plots, ncol = 2, nrow = 2)
```



Based on these plots, we might decide to alter our model selection for the Nushagak and the Wood and change our chosen model to ARIMA(0,1,1) *with drift*.

Though we could change our model for the Kvichak, the model chosen by auto.arima (ARIMA(0,0,1)) seems to behave (at least visually) essentially the same as the other top models.

TO better inform this decision, below we check the accuracies for each of the models ($n = 9$).

Check accuracy (Nick)

```
#lists from forecast::accuracy for each river produced above

#kv_err_list
kv_RMSE <- matrix(data = c(kv_err_list[[1]][4], kv_err_list[[2]][4], kv_err_list[[3]][4]))
rownames(kv_RMSE) <- c("Kvichak (0,0,1)", "Kvichak (1,0,1)", "Kvichak (0,0,2)")
kv_RMSE
```

```
##               [,1]
## Kvichak (0,0,1) 0.2717361
## Kvichak (1,0,1) 0.2814976
## Kvichak (0,0,2) 0.3125752
```

```
#best model is model 1 - (0,0,1) with non zero mean
```

```
#nu_err_list
```

```

nu_RMSE <- matrix(data = c(nu_err_list[[1]][4], nu_err_list[[2]][4], nu_err_list[[3]][4]))
rownames(nu_RMSE) <- c("Nushagak (0,1,1)", "Nushagak (0,1,1) w/ drift", "Nushagak (1,1,1)")
nu_RMSE

```

```

##                                [,1]
## Nushagak (0,1,1)              1.1181456
## Nushagak (0,1,1) w/ drift    0.9992743
## Nushagak (1,1,1)              1.1409033

```

```

# best model is model 2 - (0,1,1) with drift

```

```

#wo_err_list
wo_RMSE <- matrix(data = c(wo_err_list[[1]][4], wo_err_list[[2]][4], wo_err_list[[3]][4]))
rownames(wo_RMSE) <- c("Wood (0,1,1)", "Wood (0,1,1) w/ drift", "Wood (1,1,1)")
wo_RMSE

```

```

##                                [,1]
## Wood (0,1,1)                  0.8953588
## Wood (0,1,1) w/ drift        0.8143552
## Wood (1,1,1)                  0.8689846

```

```

#best model is model 2 - (0,1,1) with drift

```

The best models for each river were selected by choosing the model with the lowest RMSE. The best model for the Kvichak was an ARIMA(0,0,1) with non zero mean, the best model for the Nushagak was ARIMA(0,1,1) with drift and the best model for the Wood was ARIMA(0,1,1) with drift.

Check residuals (Nick)

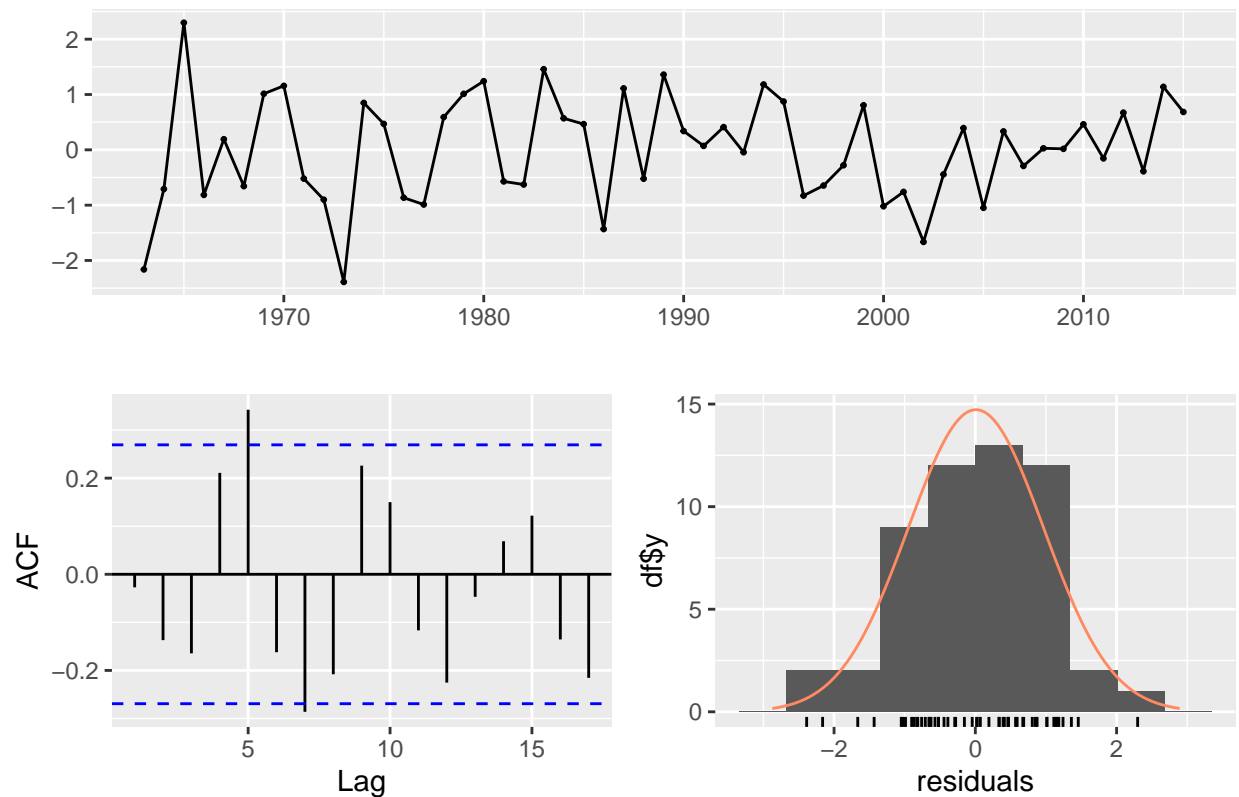
We used the model with the lower RMSE moving forward.

```

#kvichak
#select model 1
forecast::checkresiduals(kv_fore[[1]]) # plots the residuals, acf and histogram

```

Residuals from ARIMA(0,0,1) with non-zero mean



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1) with non-zero mean
## Q* = 26.979, df = 9, p-value = 0.00141
##
## Model df: 1.   Total lags used: 10
```

```
forecast::checkresiduals(kv_fore[[1]], plot = FALSE) #runs the Ljung-Box test
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1) with non-zero mean
## Q* = 26.979, df = 9, p-value = 0.00141
##
## Model df: 1.   Total lags used: 10
```

#The null hypothesis for this test is no autocorrelation. We do not want to reject the null. we fail to

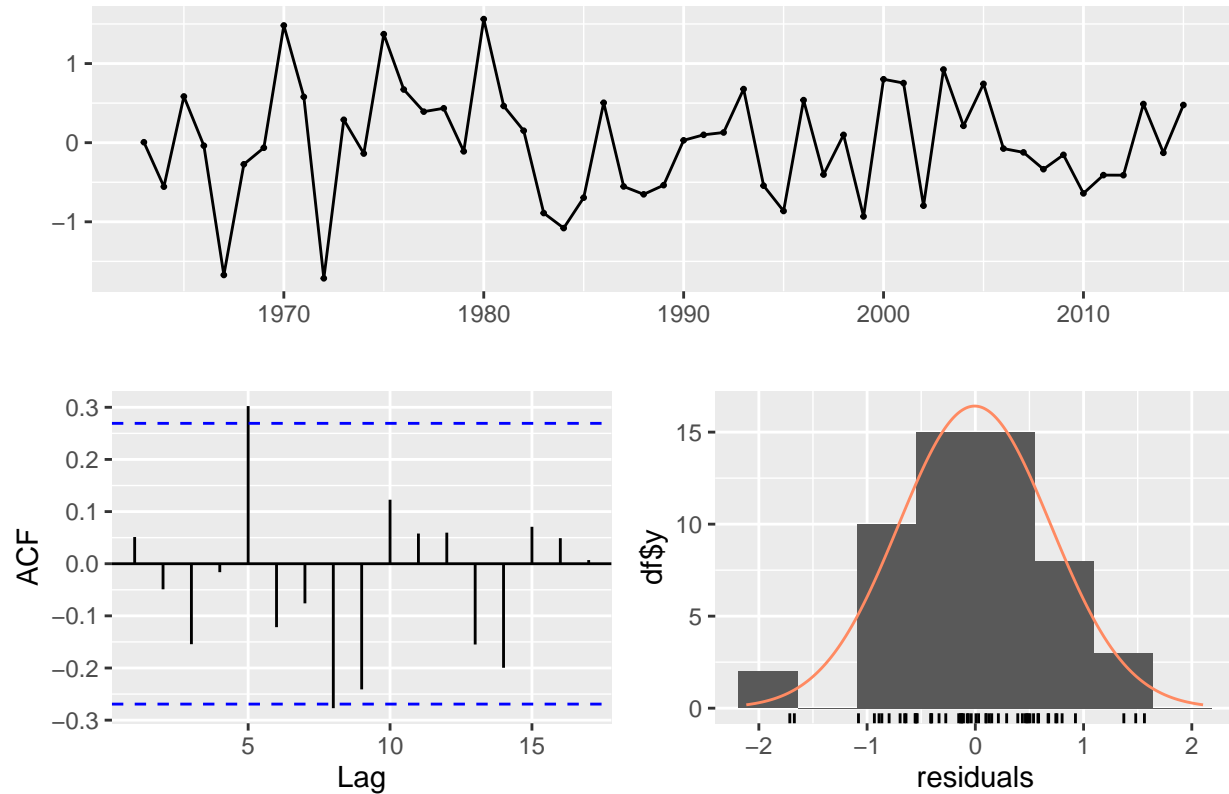
Kvichak Results

#p-value for ARIMA(0,0,1) with non zero mean is .00141 so we reject the null and the residuals are auto

#Nushagak

```
#select model 1
forecast::checkresiduals(nu_fore[[2]])
```

Residuals from ARIMA(0,1,1) with drift



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,1,1) with drift
## Q* = 18.363, df = 9, p-value = 0.03119
##
## Model df: 1. Total lags used: 10

forecast::checkresiduals(nu_fore[[2]], plot = FALSE)
```

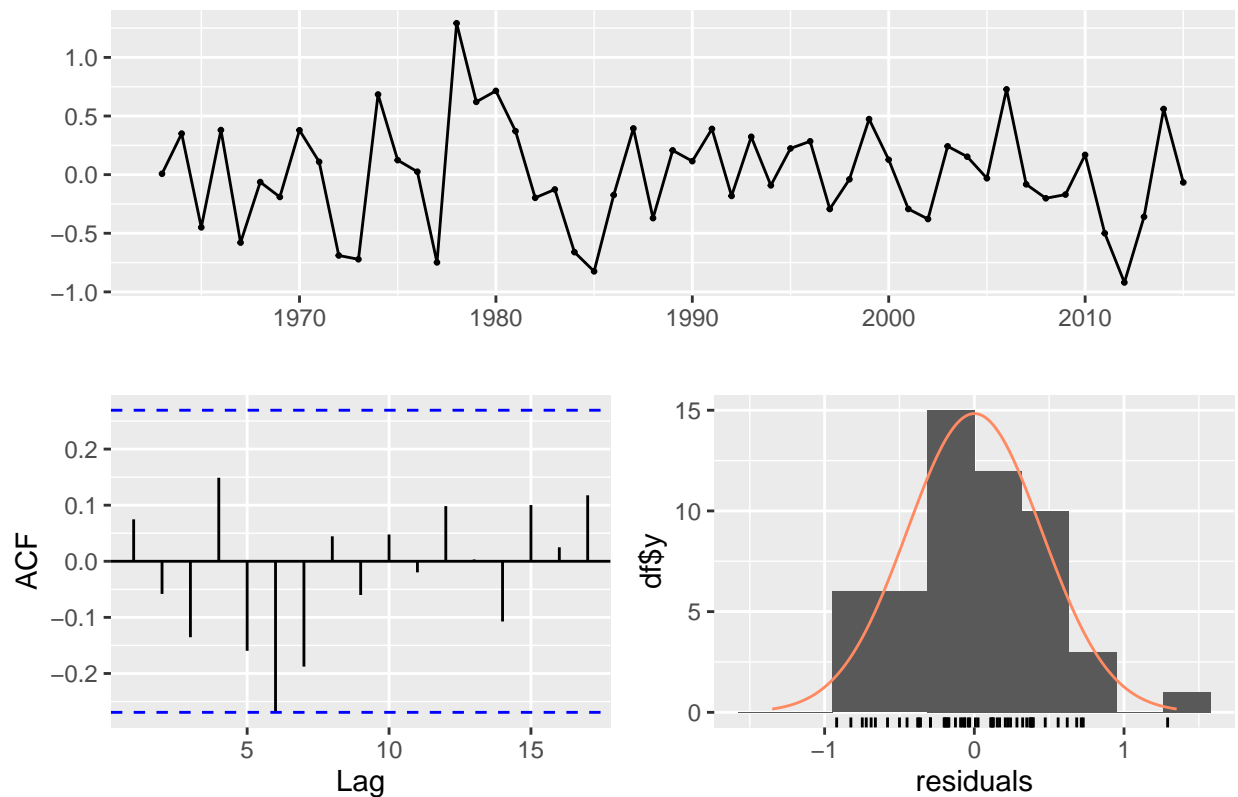
```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,1,1) with drift
## Q* = 18.363, df = 9, p-value = 0.03119
##
## Model df: 1. Total lags used: 10
```

Nushagak Results

#p-value for ARIMA(0,1,1) with drift is .031 so we reject the null and the residuals are autocorrelated


```
#Wood
#Select model 2
forecast::checkresiduals(wo_fore[[2]])
```

Residuals from ARIMA(0,1,1) with drift



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1) with drift
## Q* = 11.626, df = 9, p-value = 0.2352
##
## Model df: 1.    Total lags used: 10

forecast::checkresiduals(wo_fore[[2]], plot = FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1) with drift
## Q* = 11.626, df = 9, p-value = 0.2352
##
## Model df: 1.    Total lags used: 10
```

Wood Results

#p-value for ARIMA(0,1,1) drift is .23 so we fail to reject the null and the residuals are not autocorr

The residuals for the Kvichak River are significantly auto-correlated even though only two lags show significance on the correlogram. The Nushagak also has significantly auto correlated residuals and only the Wood passes the Ljung-Box test for having non auto correlated residuals.

Discussion

The ARIMA models we tested produced reasonable estimates for one to three years but tended to miss long term trends in the data. This was especially true for the Wood and Nushagak Rivers which had sharp upward trends in abundance after 2015 that our forecasts did not predict. Rather, our models underpredicted the actual run size in each river. The FRI forecasts were notably better than those generated by our ARIMA models in the Wood and Nushagak Rivers but were unable to capture the large increase in run sizes in recent years. Our forecasts for the Kvichak were much closer to both the FRI forecasts and estimated run sizes, although still generally under predicted across the years tested.

For the Wood and Nushagak Rivers the `auto.arima` function selected a model *without* drift as the best fit using AIC. When we compared the three best models for each river using the accuracy function a model *with* drift was chosen based on its lower RMSE value. These models with drift produced a somewhat better visual fit than those chosen by `auto.arima` but still resulted in underestimates of the run size in all years during the forecasting period. The better visual fit of our model in the Kvichak may be because the run size in the Kvichak has remained relatively stable without the increasing trend visible in the other two rivers. Both the Nushagak and Wood required differencing to remove an increasing trend in the data and this appeared to have a smoothing effect on longer forecasts vs non-differenced model runs. This smoothing may have contributed to the models for the Wood and Nushagak Rivers missing the increasing trend that occurred after 2015. This is also likely driven by the order of our ARIMA models, which only consider the previous time step to create forecasts, and thus do not perform well when predicting across multiple years (particularly when trends starkly increase or decrease). Further, these models neglect to include any environmental covariates that may help predict interannual variation at longer time steps. Had we run the models to only predict one time step forward the models likely would have performed better than relying on multiple year predictions.

The autocorrelation present in the residuals for the Kvichak indicates that some structure still exists in the data which is not explained by our models. While few lags are significant there appears to be a repeating sinusoidal pattern in the correlogram. A similar pattern may be present in the Nushagak but it is less apparent in the correlogram, however the Nushagak residuals still show significant autocorrelation. The Wood river has no significant autocorrelation in the residuals and there appears to be the least amount of structure present in the correlogram of any of the rivers tested. This suggests that the ARIMA model may be a better fit to the Wood River system, which is a surprising result as the Kvichak has the best visual fit of the three systems tested.

Based on our tests here ARIMA models seem best suited to predict short term changes in abundance and were poorly suited for estimating long term trends. As a management tool for Bristol Bay Sockeye they may still be appropriate for establishing annual biological goals for fishery management as only one time step forward would be needed.

Team member contributions

All team members looked at the data and worked together to decide on a question to answer. The code was split up into several sections and each team member had specific sections to write, although there was a lot of collaboration in drafting these sections. Liz and Maria went to office hours for trouble shooting of code

and to ask questions. All team members contributed to accuracy tests and model selection. Nick wrote a first draft of the discussion and all team members provided helpful edits.