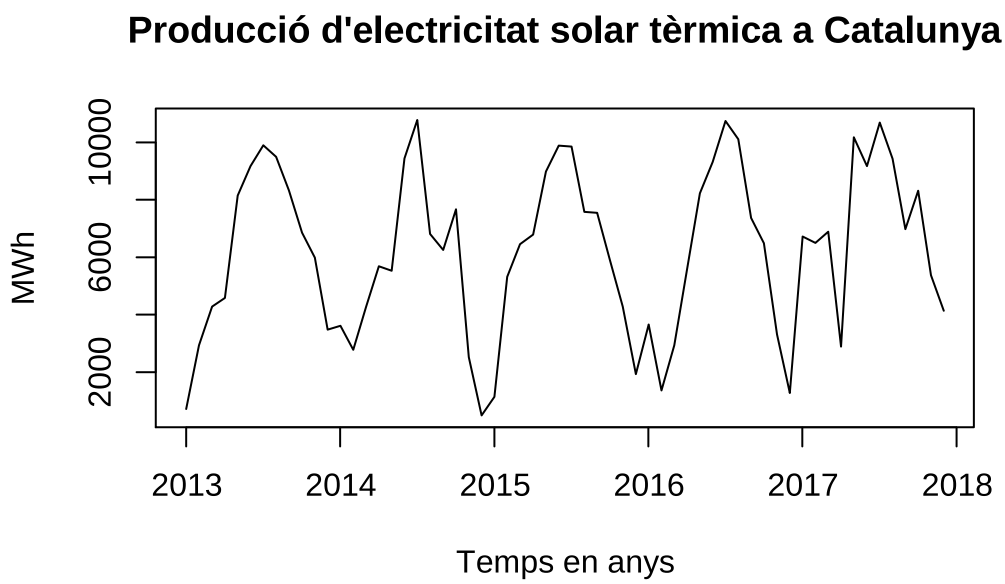
1 Pràctica 1.2

Carregueu les dades contingudes al fitxer "prac8TS.txt", corresponent a la producció mensual d'electricitat solar tèrmica a Catalunya entre 2013 i 2017.

```
prac8TS <- read.csv("C:/Users/Alba/Desktop/prac8TS.txt", sep="")
dades<-prac8TS
dadesany<-ts(dades, start = c(2013,1), end = c(2017,12), frequency = 12)
dades.ts<-dadesany[,3]
```

1.1 Dibuixeu la sèrie. És un procés estacionari? Perquè?

```
plot(dades.ts, main = "Producció d·electricitat solar tèrmica a Catalunya",
     ylab = "MWh", xlab = "Temps en anys")
```



Dibuixem la sèrie fent un plot de les dades. La sèrie no sembla que presenti tendència. Sí s'observa certa estacionalitat. Per exemple, podem observar que la producció d'electricitat solar tèrmica a Catalunya sempre serà majors en els mesos d'estiu.

L'estacionalitat sol provocar que la sèrie sigui no estacionària, ja que els valors mitjans en alguns moments particulars dins del període estacional (mesos, per exemple) poden ser diferents als valors mitjans en altres ocasions, i això passa en el nostre cas.

Per estudiar si la sèrie és estacionària també podem utilitzar el test de Dickey-Fuller que busca determinar l'existència o no d'arrels unitàries en una sèrie de temps. La hipòtesi nul.la d'aquesta prova és que hi ha una arrel unitària en la sèrie, i per tant, no seria estacionària. I com hipòtesis alternativa la sèrie seria estacionària.

```
adf.test(dades.ts)

##
##  Augmented Dickey-Fuller Test
##
## data:  dades.ts
## Dickey-Fuller = -5.0957, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

En el cas de la nostra sèrie veiem que ens ha donat un p-valor de 0.01, per tant més petit que 0.05 (nivell de significació habitual). Amb un 95% de confiança rebutjaríem la hipòtesis nul.la de que la série no es estacionaria.

Però en aquest cas el test de Dickey-Fuller és enganyós. Quan es fa el test de Dickey-Fuller, no arriba a analitzar-ho en el lag 12 (que és la nostra estacionalitat) i per tant ens diu que és un procés estacionari quan en realitat no ho és.

Si fem la funció auto-arima() veiem que ens detecta una estacionalitat de lag= 12.

```
auto.arima(dades.ts)

## Series: dades.ts
## ARIMA(1,0,0)(1,1,0)[12]
##
## Coefficients:
##          ar1     sar1
##       0.2822  -0.6373
## s.e.  0.1409   0.1169
##
## sigma^2 estimated as 2688692:  log likelihood=-425.57
## AIC=857.13   AICc=857.68   BIC=862.74
```

A més podem observar que demana fer la diferència d'ordre 12 per tal de ser estacionaria. Veiem que en aquest cas la part no estacional (p,d,q) no demana cap diferència, però en canvi la part estacional (P,D,Q) si que demana fer la diferència estacional per tal de que sigui estacionaria.

Concloem doncs que és una sèrie no estacionaria.

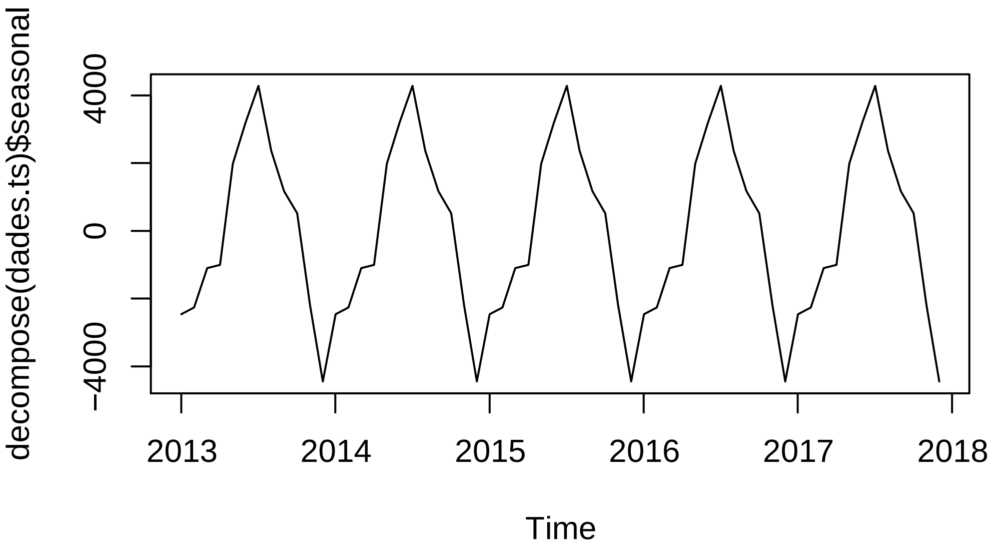
1.2 Té tendència?

```
plot(decompose(dades.ts)$trend)
```

Mirant al primer gràfic ja havíem vist que la sèrie no semblava tenir cap tendència clara. Si fem decompose podem obtenir el gràfic de la tendència i tornem a observar que no sembla que segueixi cap model ni lineal ni quadràtic. A més si mirem els resultats obtinguts anteriorment amb autoarima veiem que no detecta drift.

1.3 Té estacionalitat? Si en té, quin et sembla que és el període del cicle?

```
plot(decompose(dades.ts)$seasonal)
```



Mirant la gràfica es podria dir que manté els efectes estacionals. Té estacionalitat ja que veiem un patró habitual de canvis que es repeteix durant l'any (o bé 12 mesos)

1.4 Quina és l'estacionalitat que proposa auto.arima()?

```
auto.arima(dades.ts)

## Series: dades.ts
## ARIMA(1,0,0)(1,1,0)[12]
##
## Coefficients:
##          ar1      sar1
##       0.2822   -0.6373
## s.e.  0.1409    0.1169
##
## sigma^2 estimated as 2688692:  log likelihood=-425.57
## AIC=857.13   AICc=857.68   BIC=862.74
```

Amb autoarima com ja haviem comentat per sobre anteriorment podem observar que l'estacionalitat proposada és de 12 mesos.

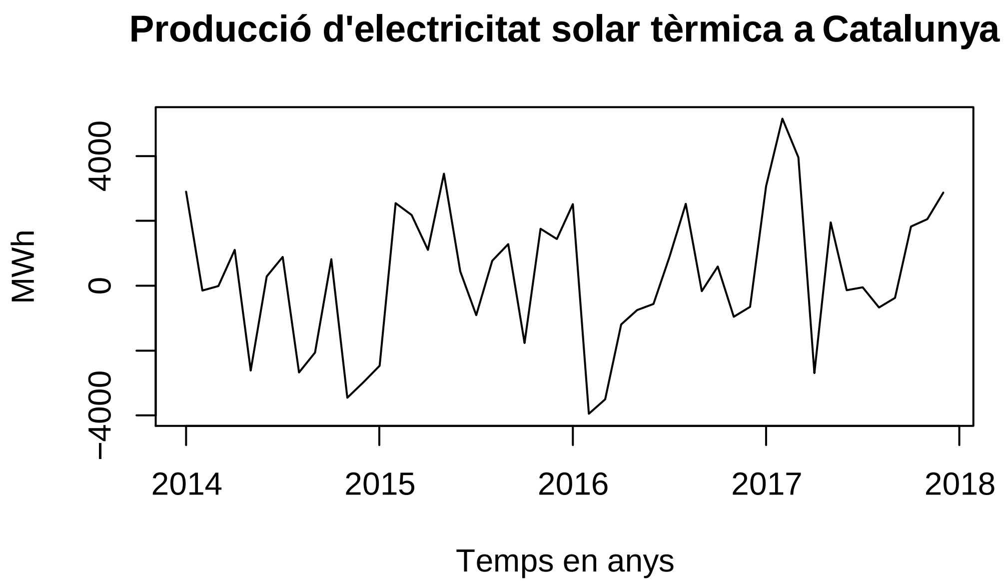
1.5 Quin model proposa auto.arima() per a la sèrie i per a la part estacional? Escriu explícitament el model proposat i comenta els aspectes que et semblin més relevants.

El model proposat per autoarima inclou un terme AR(1) no estacional per a la sèrie i un terme AR(1) estacional i una diferència per la part estacional. El model que mostra és sense termes MA i el període estacional és S = 12.
Per la part no estacional
AR(1): $\varphi(B) : 1 - \varphi_1 B = 1 - 0.2822B$
Per la part estacional
SAR(1): $\phi(B^{12}) : (1 - \phi_{12} B^{12})(1 - B^{12}) = (1 + 0.6373B^{12})(1 - B^{12})$
Finalment el model ARIMA $(1, 0, 0)(1, 1, 0)_{12}$ queda:
$(1 - \varphi_1 B)(1 - \phi_{12} B^{12})(1 - B^{12})X_t = \varepsilon_t$
$(1 - 0.2822B)(1 + 0.6373B^{12})(1 - B^{12})X_t = \varepsilon_t$
Per tant hem pogut observar que la part no estacional és estacionaria i en canvi la part estacional no és estacionaria, ja que demana fer la diferència per tal de ser estacionaria fet que és bastant rellevant.

1.6 Calcula les diferències proposades per auto.arima() i comprova que la nova sèrie és estacionària. Hi ha alguna diferència entre el model proposat per a la sèrie original i per a la diferenciada? Si n'hi ha, quines son?

Per suavitzar l'estacionalitat es pot diferenciar les dades. L'estacionalitat sol provocar que la sèrie sigui no estacionària, ja que els valors mitjans en alguns moments particulars dins del període estacional poden ser diferents. Per tant examinem les dades diferenciades amb lag 12 concretament per la nostra sèrie i obtenim el següent gràfic, que en aquest cas si que és estacionari.

```
dades.ts_diff <- diff(dades.ts, lag = 12)
plot(dades.ts_diff, main = "Producció d·electricitat solar tèrmica a Catalunya",
     ylab = "MWh", xlab = "Temps en anys")
```




```
adf.test(dades.ts_diff)

##
##  Augmented Dickey-Fuller Test
##
## data:  dades.ts_diff
## Dickey-Fuller = -3.6411, Lag order = 3, p-value = 0.03945
## alternative hypothesis: stationary
```

El test de Dickey - Fuller torna a donar un p-valor més petit que 0.05, per tant rebutjaríem la hipòtesis nul.la de que la sèrie no es estacionaria. Abans de fer la diferència també ens havia donat aquest resultat però intuíem que era enganyós. En aquest cas sabem que és cert que és estacionaria.

```
auto.arima(dades.ts_diff)

## Series: dades.ts_diff
## ARIMA(1,0,0)(1,0,0)[12] with zero mean
##
## Coefficients:
##          ar1      sar1
##       0.2822   -0.6373
## s.e.  0.1409    0.1169
##
## sigma^2 estimated as 2688682:  log likelihood=-425.57
## AIC=857.13   AICc=857.68   BIC=862.74
```

La funció auto.arima( ) ens proposa un model que inclou un terme AR(1) no estacional per a la sèrie i un terme AR(1) estacional per la part estacional. El model que mostra és sense termes MA i el període estacional és S = 12. Veiem que en aquest cas ens explicita que és amb mitjana 0 i observem que ja no ens demana fer cap diferència per la part estacional. També observem que tots els altres valors són exactament iguals inclosos els valors AIC, AICc i BIC. Així el model després de fer la diferència quedaria:

Per la part no estacional

AR(1): $\varphi(B) : 1 \varphi_1 B = 1\ 0.2822B$

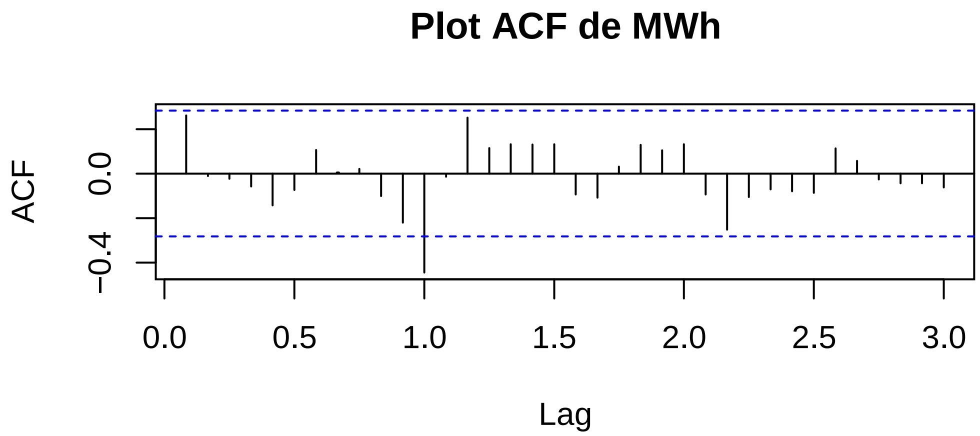
Per la part estacional

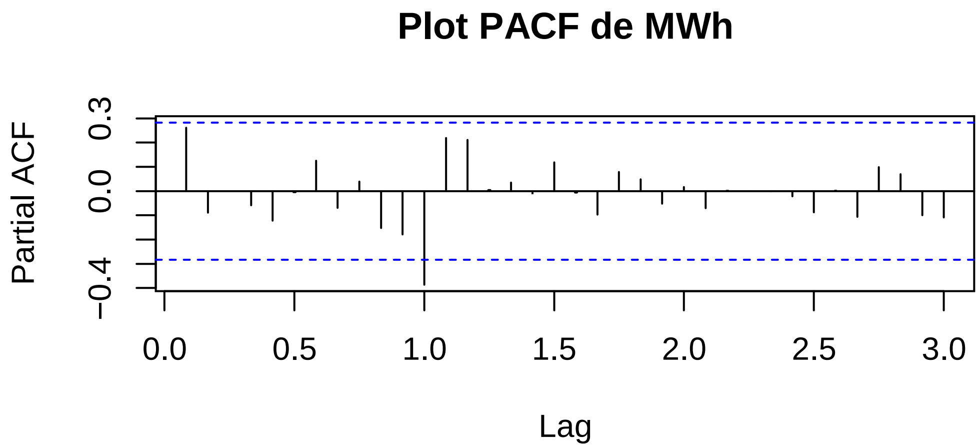
SAR(1): $\phi(B^{12}) : (1 - \phi_{12}B^{12}) = (1 + 0.6373B^{12})$

Finalment el model ARIMA $(1, 0, 0)(1, 0, 0)_{12}$ queda: $(1 - \varphi_1 B)(1 - \phi_{12}B^{12})X_t = \varepsilon_t$

$(1 - 0.2822B)(1 + 0.6373B^{12})X_t = \varepsilon_t$

```
par(mfrow=c(2,1))
acf(dades.ts_diff, main = "Plot ACF de MWh",lag.max =36)
pacf(dades.ts_diff, main = "Plot PACF de MWh",lag.max =  36)
```



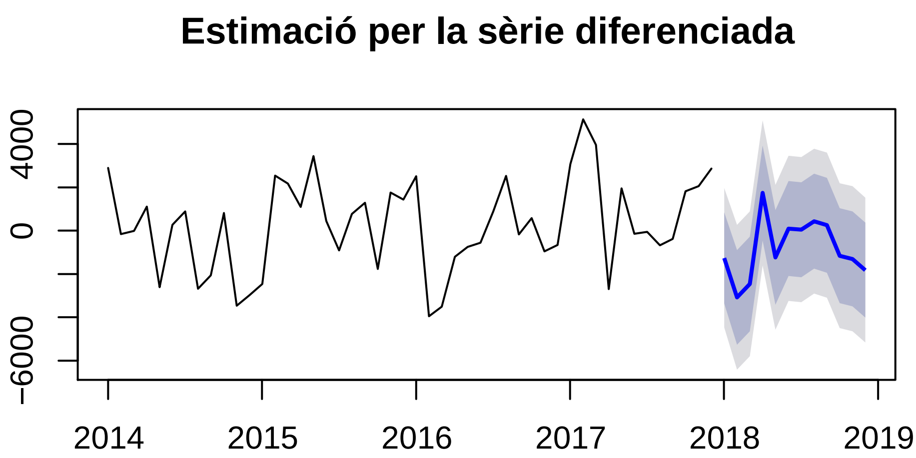


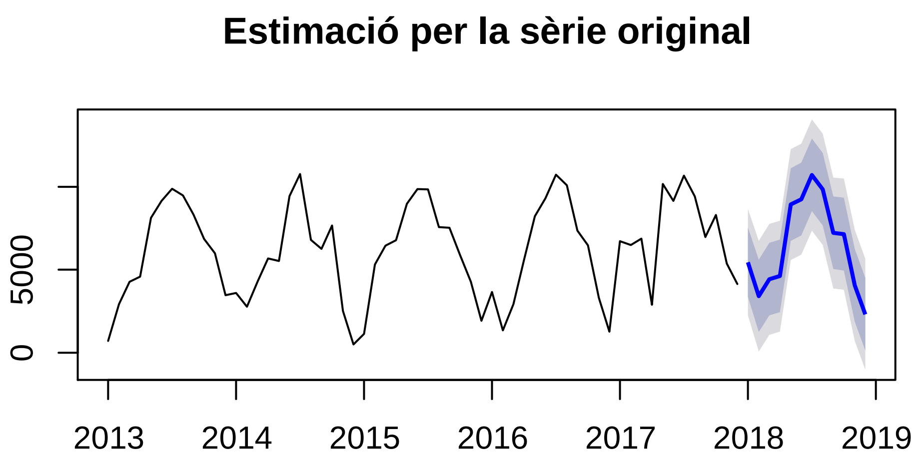
Fent ACF i PACF podem observar el que autoarima ja ens havia dit. Veiem que en el AR(1) trobem un lag signi catiu en el lag 12 ja que és el nostre període.

1.7 Fent servir la funció forecast() doneu una estimació per als valors de l'any 2018 i dibuixeu la predicció.

```
par(mfrow=c(2,1))
### Fent servir la llibreria forecast
mod1 <- auto.arima(dades.ts_diff)
pred1 <- forecast(dades.ts_diff, h = 12, model = mod1)
plot(pred1,main = "Estimació per la sèrie diferenciada")

mod2 <- auto.arima(dades.ts)
pred2 <- forecast(dades.ts, h = 12, model = mod2)
plot(pred2,main = "Estimació per la sèrie original")
```





Com l'enunciat no especifica de quines dades s'havia de fer l'estimació, mostrem aquesta estimació per a la sèrie diferenciada i per a la sèrie original. Això ens ha permès veure que les estimacions són diferents depenent del model que utilitzem.

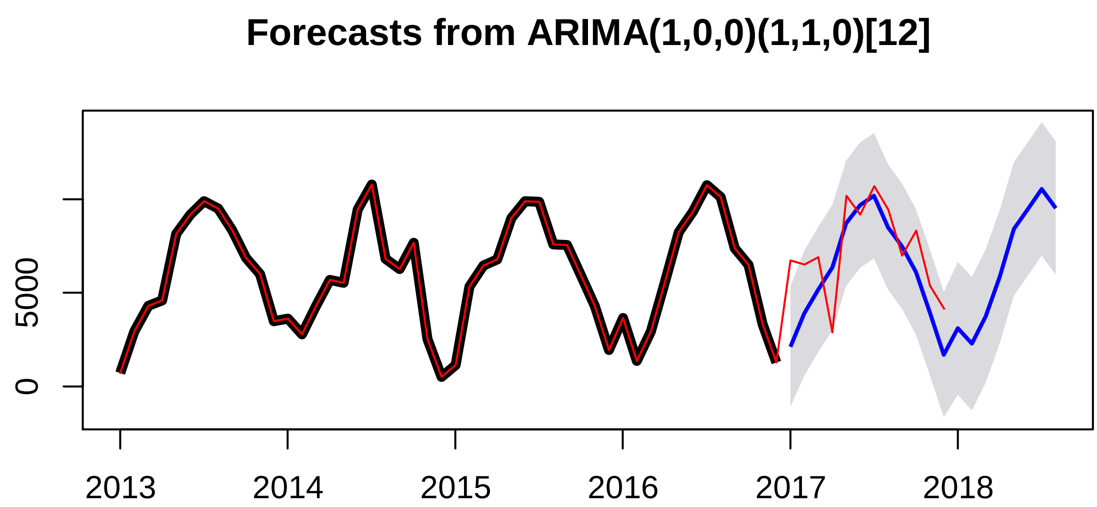
1.8 Quedeu-vos amb les observacions de 2013 al 2016 i genereu una predicció per als 20 valors restants basada en el model ajustat anteriorment. Dibuixeu la sèrie real juntament amb les prediccions. Què diríeu sobre la bondat de l'ajust del model utilitzat? En quants punts els valors reals queden fora de l'interval de confiança del 95% de les prediccions?

```
o1316 <- ts(dades, start = c(2013, 1), end = c(2016, 12), frequency = 12)
o1316 <- o1316[,3]

pred3 <- forecast(o1316, h = 20, level = 95, model = mod2)
plot(pred3, lwd = 5)
lines(dades.ts, col = "red", lwd = 1, type = "l")

cnt <- 0
for(i in 1:12){
  if(dades.ts[49:60][i]<pred3$lower[i] | dades.ts[49:60][i] > pred3$upper[i]){
    cnt <- cnt + 1
  }
  else{
    cnt <- cnt
  }
}
cnt

## [1] 2
```



Un cop dibuixat el gràfic amb les prediccions i els valors reals veiem que és un model que s'ajustaria prou bé ja que la majoria de valors reals queden dintre de l'interval de confiança del 95%. A més a més hem creat una funció per tal de comptar quants punts de les dades reals queden fora, i en el cas de la nostra base de dades només dos punts queden fora. Per tant un 10%. Semblaria un bon ajust, però s'hauria de tenir en compte que estem utilitzant molt poques dades i s'hauria d'observar que passa si agaféssim una base de dades més extensa.