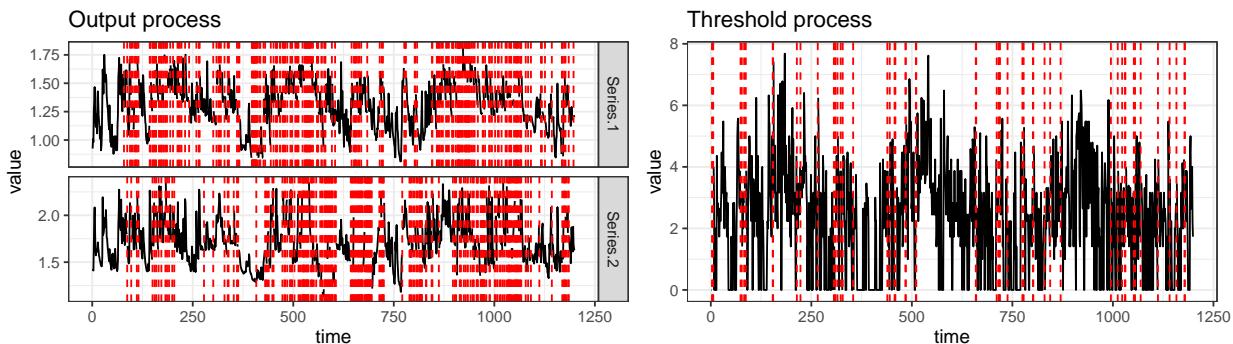


En la aplicación con datos se busca encontrar la relación entre la precipitación diaria (en mm) y el caudal diario del río (en m^3/s) de dos ríos donde un río desemboca en el otro en una región del departamento del Cauca en Colombia. La precipitación se midió en la estación meteorológica San Juan con una altitud de 2400 metros. El primer caudal del río se midió en la estación hidrológica “El Trébol” del río Bedón con una altitud de 1720 metros y el segundo caudal se midió en el río de la Plata en la estación hidrológica de “Villalosada” con una altitud de 1300 metros. Las estaciones están ubicadas cerca del ecuador de la Tierra en una zona geográfica muy seca. Esta última característica permite el control de factores hidrológicos/meteorológicos, los cuales pueden distorsionar el tipo de relación dinámica explicada por el modelo MTAR.

El período de tiempo que se consideró es del 1 de enero de 2006 al 14 de abril de 2009 (1200 puntos de tiempo) que tiene 57 puntos de tiempo con datos faltantes en la serie de precipitaciones, 214 en la serie del caudal del río Bedón y 213 en la serie del cauce del río de la Plata. Estos datos fueron proporcionados por el IDEAM (disponibles en la librería R para los modelos TAR y MTAR).

```
# install.packages("devtools")
# devtools::install_github("adrincont/BMTAR", force = TRUE)
library(BMTAR)
library(zoo)
library(patchwork)
imput_y = function(y){
  ts_prob = data.frame(time = 1:length(y), dat = y)
  ts_prob_2 = ts_prob[!is.na(ts_prob$dat),]
  ks_imput = ksmooth(x = ts_prob_2$time, y = ts_prob_2$dat, 'normal'
    , bandwidth= 10, x.points = 1:length(y))
  ts_prob$dat[is.na(ts_prob$dat)] = ks_imput$y[is.na(ts_prob$dat)]
  return(ts_prob$dat)
}

# [Lectura de Datos]
data("hydrodata")
Yt = hydrodata[,3:4]
Pt = hydrodata$Rainfall
data_miss = list(Yt = as.matrix(log(sqrt(Yt[-1,]))),
                 Zt = as.matrix(sqrt(Pt[-nrow(hydrodata)])))
## Graficas de las series .....#####
datos = tsregime(Yt = data_miss$Yt, Zt = data_miss$Zt)
autoplot(datos,1) + autoplot(datos,2)
```



Para la aplicación de la metodología para estimación de modelos MTAR con datos faltantes se recurre a la Librería de R “MTAR” (librería basada en el mismo artículo y que implementa toda la metodología de estimación). Como primer paso y como se propone en el artículo se consideran las series transformadas $Y_t = (\ln(\sqrt{Y_{1t}}), \ln(\sqrt{Y_{2t}}))$ y $Z_t = \sqrt{P_{t-1}}$ (precipitaciones y el caudal bivariado de los ríos Bedón y La Plata el día t). En la metodología de estimación para estos modelos, el primer paso corresponde a la imputación

inicial de datos faltantes que en este caso se realiza por medio de un promedio móvil sobre cada uno de los componentes.

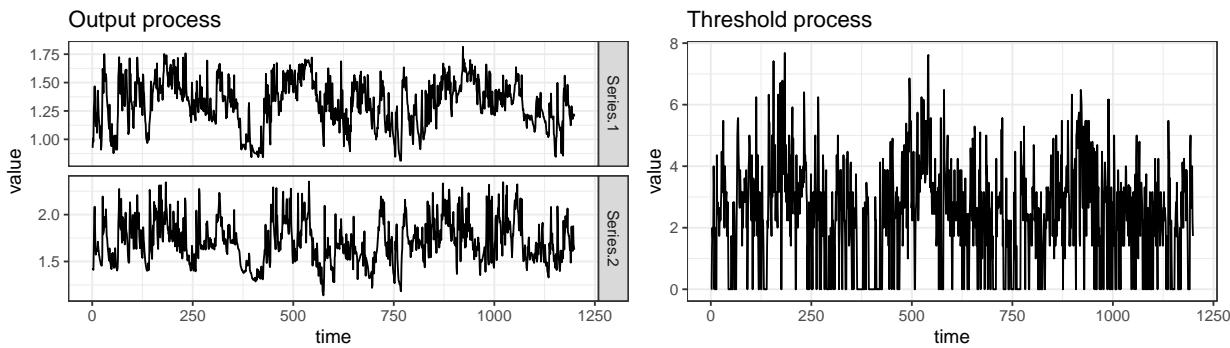
```

Y_temp = datos$Yt
Y_temp[,1] = imput_y(Y_temp[,1])
Y_temp[,2] = imput_y(Y_temp[,2])
Z_temp = datos$Zt
Z_temp[,1] = imput_y(Z_temp[,1])
Ctest(X = Y_temp,Z = c(Z_temp),p = 5,d = 5)

##          test      p-value
## 3.171155e+02 4.068634e-54

datos_temp = tsregime(Yt = Y_temp,Zt = Z_temp)
autoplot(datos_temp,1) + autoplot(datos_temp,2)

```

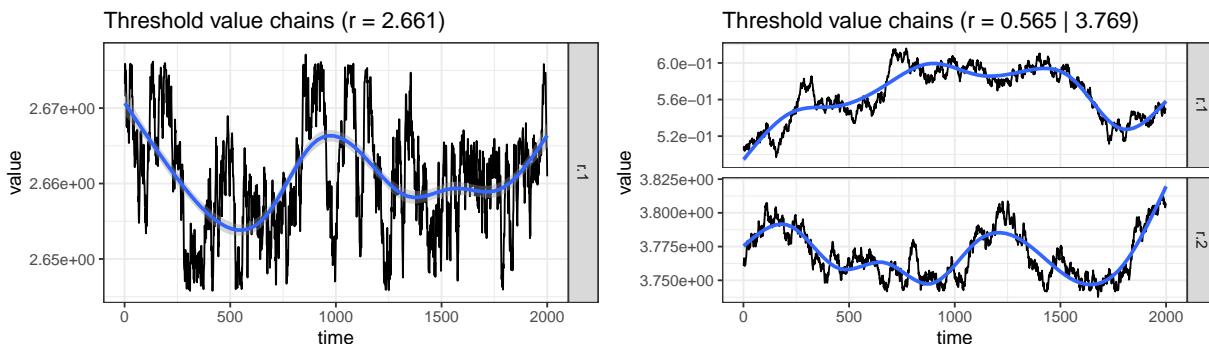


Una vez completada la serie, evaluamos la prueba de no linealidad de Tsay que tiene como hipótesis nula que el proceso es lineal frente a la alternativa de que se comporta como un modelo de umbrales, obteniendo un test de 317.116 (pvalor < 0.05) por lo que se rechaza H_0 frente a que el modelo sigue un proceso de umbrales, así el siguiente paso corresponde a la estimación del número de regímenes mediante el método de selección de modelos de "Carlin and Chib" y el uso del criterio "NAIC", seguido de la estimación de ordenes para el modelo por medio de la metodología de "Kuo and Mallick" (ambas metodologías explicadas en la primera parte de estos informes). Para ello, se usan las funciones "mtarnumreg()" y "mtarstr()" que implementan dichas metodologías usando prioris no informativas (el usuario podría dar prioris informativas si es el caso).

```

# initial = mtarinipars(tsregime_obj = datos_temp,
# list_model = list(l0_max = 3),method = 'KUD')
# estim_nr = mtarnumreg(ini_obj = initial,iterprev = 2000,
# niter_m = 3000,burn_m = 1000,list_m = TRUE,
# ordersprev = list(maxpj = 5,maxqj = 0,maxdj = 5))
autoplot(estim_nr,1)$m2 + autoplot(estim_nr,1)$m3

```

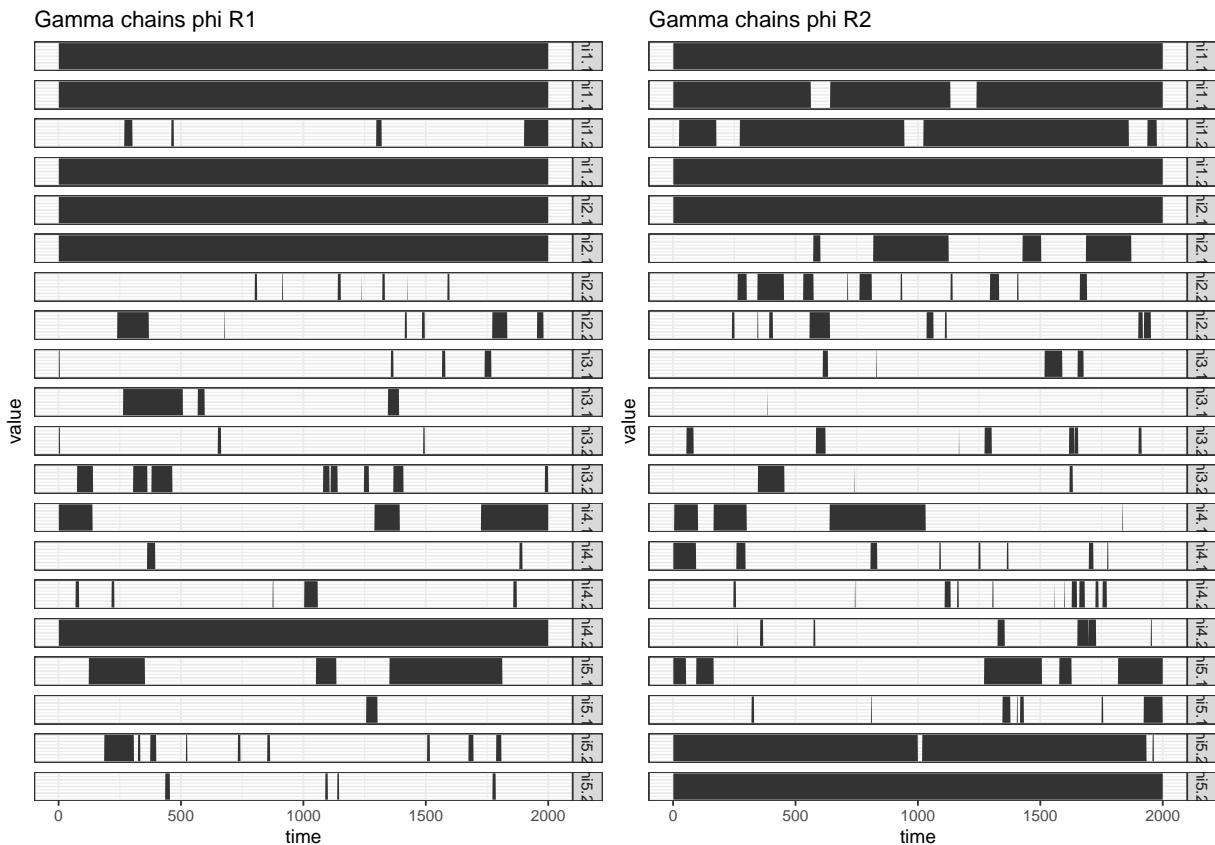


```
mtarNAIC(estim_nr$list_m$m2$par);mtarNAIC(estim_nr$list_m$m3$par)
```

```
## NAIC=8.3345
## $AICj
## [1] 5752.894 4198.447
##
## $NAIC
## [1] 8.334457
## NAIC=8.5622
## $AICj
## [1] 2284.942 6000.067 1938.288
##
## $NAIC
## [1] 8.562225
```

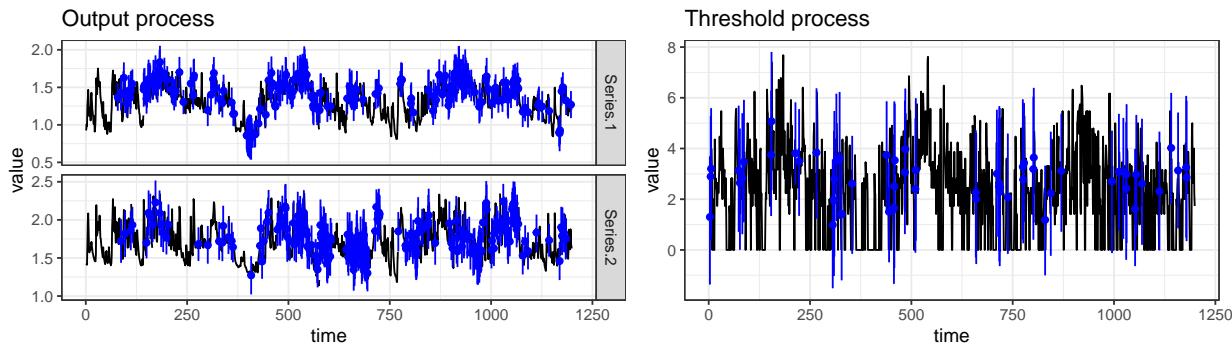
Teniendo en cuenta criterios como el comportamiento de las cadenas en los muestradores en donde se observa mayor estabilidad para el caso de $l = 2$, los resultados en las pruebas de diagnóstico con 2 y 3 regímenes y los valores resultantes del criterio NAIC (donde en las pruebas que se hicieron el menor NAIC siempre resulta ser $l = 2$), se seleccionan dos regímenes en el modelo. A continuación, usando como punto de partida inicial de $r = 2.66$, se realiza la estimación de parámetros no estructurales para el modelo MTAR (órdenes para rezagos y variable de umbrales), esto por medio de la función “`mtarstr()`” que implementa la metodología de selección de variables conjuntamente con la estimación de los demás parámetros.

```
autoplot(estim_nr$list_m$m2$par,4)$R1$phi + autoplot(estim_nr$list_m$m2$par,4)$R2$phi
```



Observando en los resultados del procedimiento el comportamiento de las variables gammas (variables indicadoras relacionadas con los ordenes) y sus frecuencias más comunes (en el primer régimen la mayor frecuencia fue de 22.25% seguida del 5.5% y en el segundo régimen la mayor de 6.3% seguida de 6.15%), se tienen las estimaciones de ordenes para el modelo con $p_1 = 5, d_1 = 5$ para el primer régimen y $p_2 = 5, d_2 = 5$ para el segundo además de una estimación para $r = 2.6$. Luego de encontrar un aproximación a los ordenes de los rezagos (parámetros estructurales) así como del valor del umbral es posible una primera estimación de los datos faltantes (por medio de la metodología del filtro de Kalman y el modelo espacio-estado), esto a partir de la función “mtarmissing()” que realiza la estimación de datos faltantes realizando el proceso de muestreo conjunto de parámetros no estructurales y observaciones faltantes. En la gráfica observamos el valor puntual y su respectivo intervalo de credibilidad al 95% en azul para los valores faltantes.

```
# list_model = list(pars = list(l = 2,r = 2.6,
# orders = list(pj = c(5,5),dj = c(5,5))),
# initial = mtarinipars(tsregime_obj = tsregime(Yt = data_miss$Yt,
# Zt = data_miss$Zt),list_model = list_model)
# missingest = mtarmissing(ini_obj = initial,chain = TRUE, niter = 5000,
# burn = 1000)
# data_c = missingest$tsregim
autoplot(missingest,4)[[1]] + autoplot(missingest,4)[[2]]
```

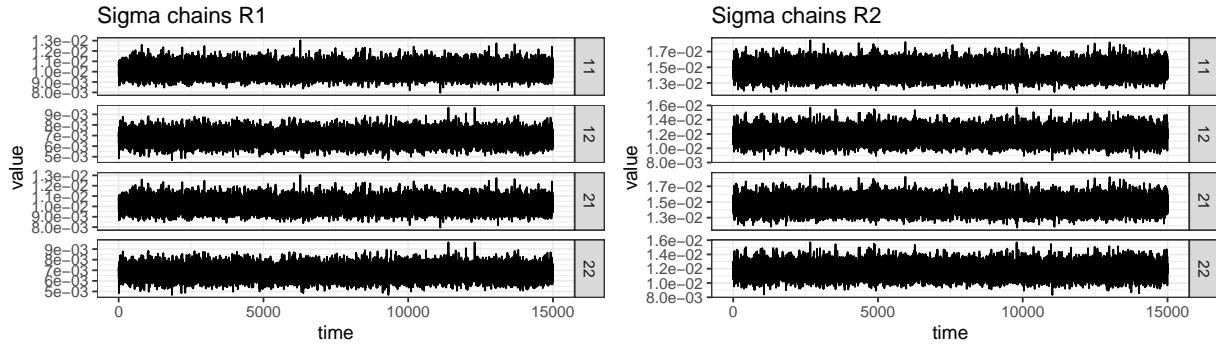


Una vez completados los valores faltantes es necesario realizar nuevamente el procedimiento tanto de estimación de la cantidad de régimenes (que siguiendo el mismo procedimiento de la parte inicial, resulta por criterio NAIC $l = 2$ régimenes) como de los parámetros estructurales para seguir reestimando los valores faltantes hasta tener una convergencia. Una vez concluido este proceso y teniendo los valores finales tanto de los faltantes como de los parámetros estructurales se realiza la estimación final de los parámetros no estructurales por medio de la función “mtarns()” que implementa los muestreadores de acuerdo a los parámetros a estimar. Para los cuales evaluamos el comportamiento de las cadenas y supuestos sobre los residuos al ajuste final.

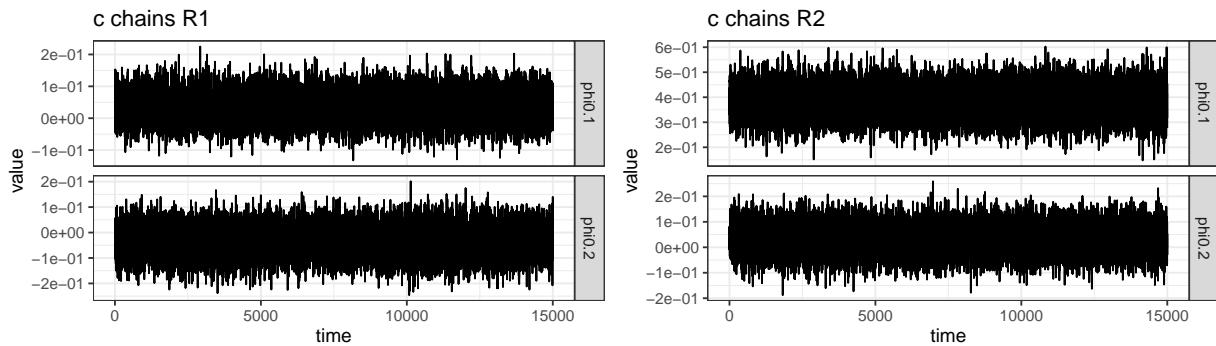
Los primeros gráficos nos muestran un buen comportamiento de cadenas sin valores que salten, por su parte los residuos aunque presenten algunas inconsistencias en normalidad podemos ver que no son autocorrelacionados, sin embargo es posible considerar otros ajustes al modelo para garantizar los supuestos.

```
# initial = mtarinipars(tsregime_obj = data_c,
# list_model = list(pars = list(l = 2,
# orders = list(pj = c(5,5),dj = c(5,5)),r = 2.6)))
# ns2 = mtarns(ini_obj = initial,niter = 15000,chain = TRUE,burn = 5000)

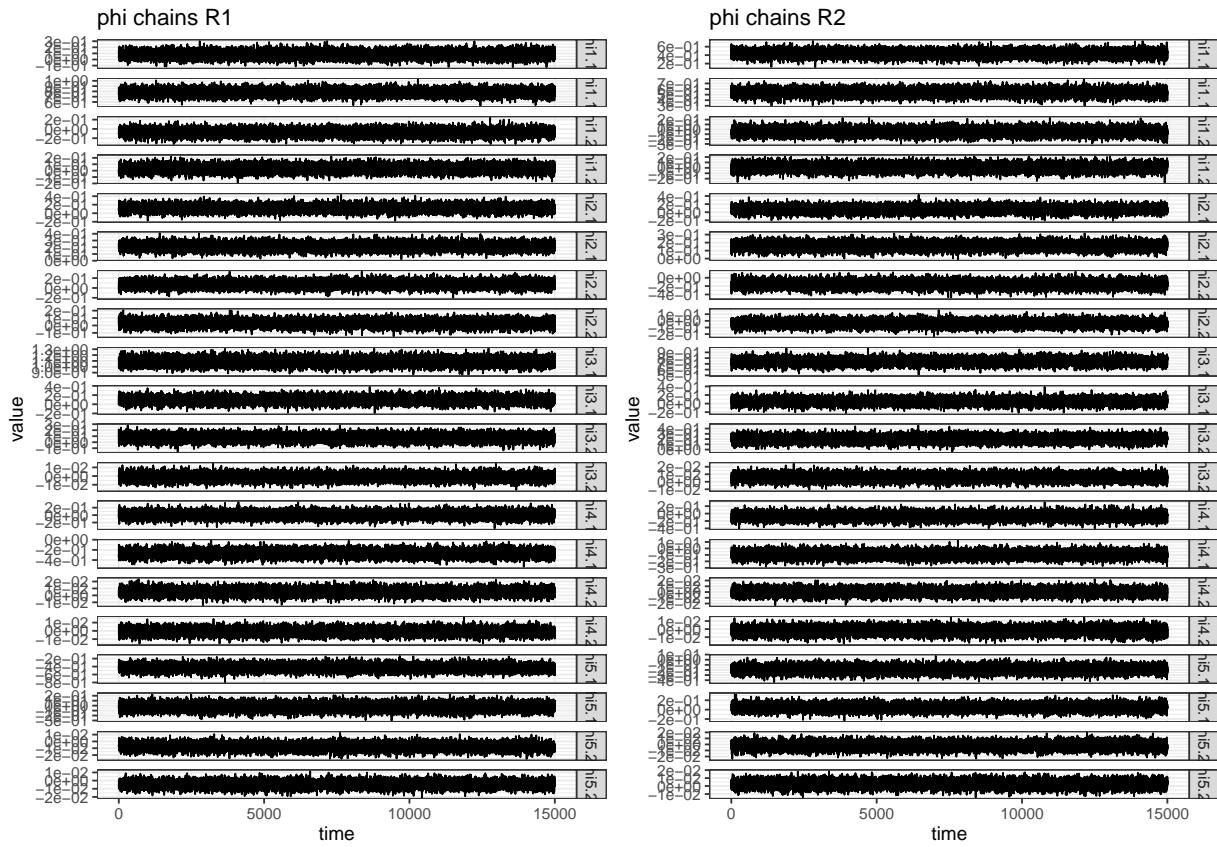
autoplot(ns2,2)$R1+autoplot(ns2,2)$R2
```



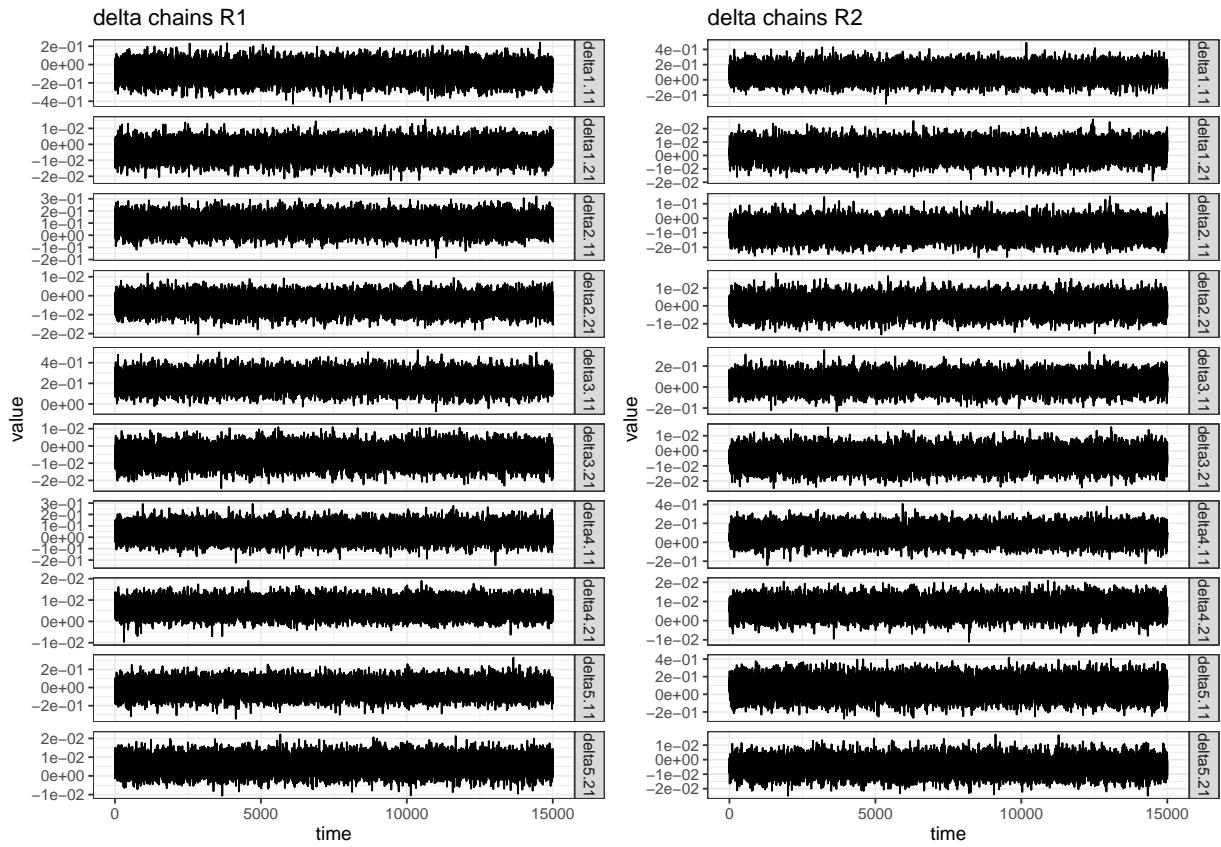
```
autoplot(ns2,3)$R1$c+autoplot(ns2,3)$R2$c
```



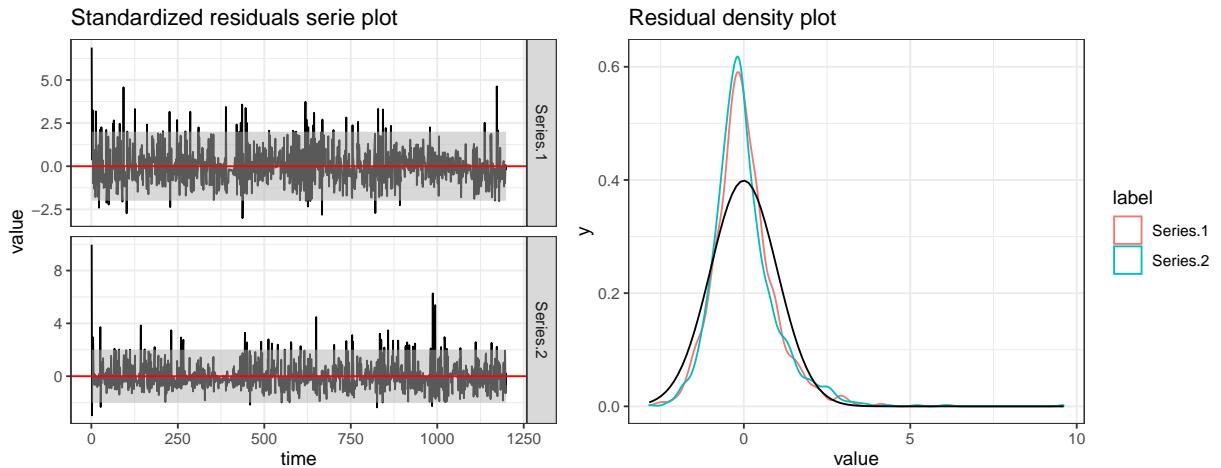
```
autoplot(ns2,3)$R1$phi+autoplot(ns2,3)$R2$phi
```



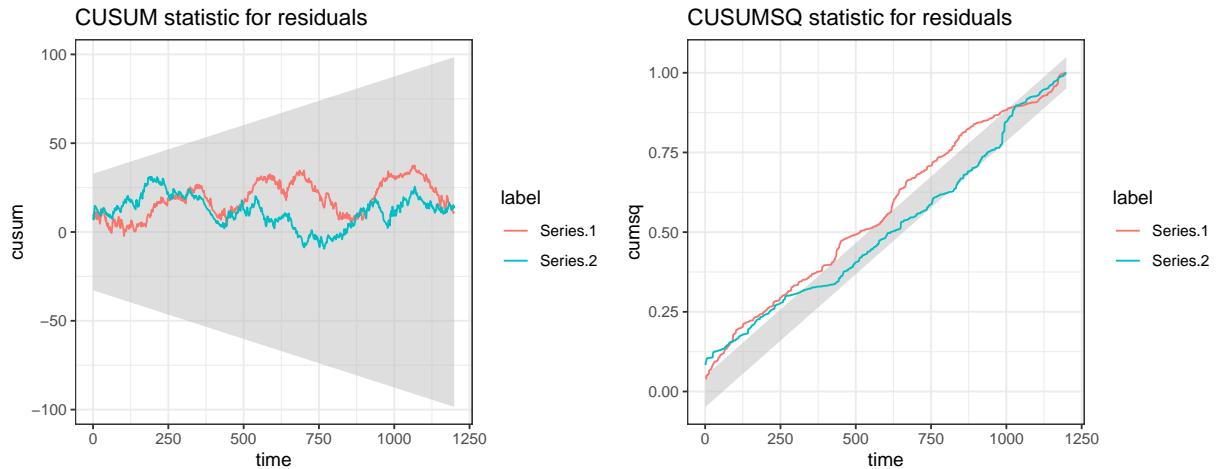
```
autoplot(ns2,3)$R1$delta+autoplot(ns2,3)$R2$delta
```



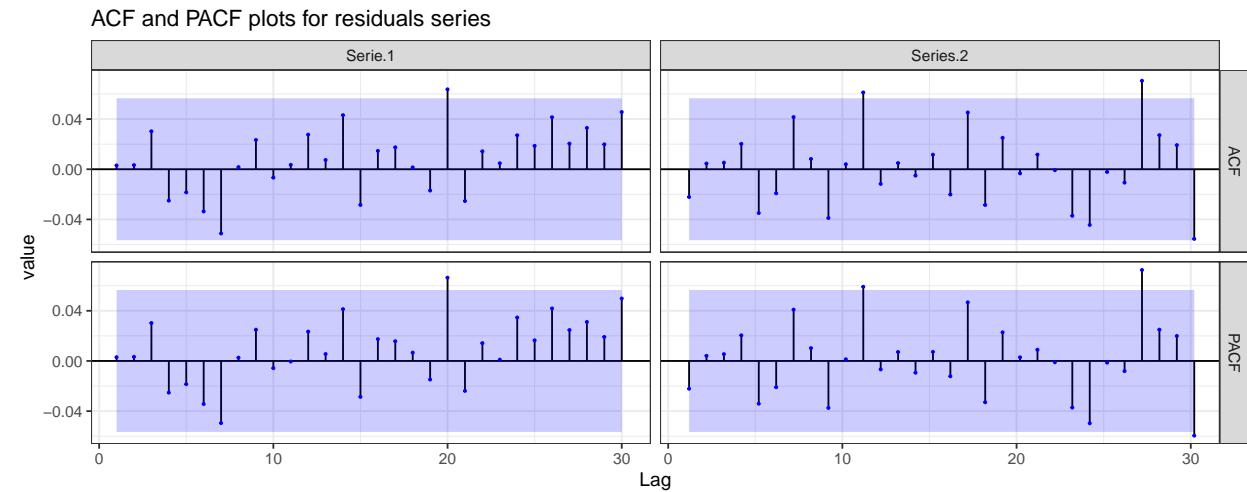
```
diagnostic_mtar(ns2)[[2]] + diagnostic_mtar(ns2)[[3]]
```



```
diagnostic_mtar(ns2)[[4]] + diagnostic_mtar(ns2)[[5]]
```



```
diagnostic_mtar(ns2)[[6]]
```



Finalmente vemos la estimación de los parámetros con valor de umbral 2.6 y en el gráfico el ajuste del modelo a la serie de datos una vez completados los valores faltantes.

```
ns2$regime
```

```
## $R1
## $cs
## [,1]
## [1,] 0.03552373
## [2,] 0.08394092
##
## $phi
## $phi$phi1
## [,1]      [,2]
## [1,] 0.7790339 0.2152434
## [2,] 0.1214762 1.0876992
##
## $phi$phi2
## [,1]      [,2]
## [1,] 0.097138291 -0.2685590
```

```

## [2,] 0.007093265 -0.4256092
##
## $phi$phi3
## [,1]      [,2]
## [1,] -0.02435740 0.09333177
## [2,] -0.08021651 0.22041891
##
## $phi$phi4
## [,1]      [,2]
## [1,] 0.04525088 -0.03604986
## [2,] -0.01319476 -0.07454925
##
## $phi$phi5
## [,1]      [,2]
## [1,] 0.03512518 0.02999536
## [2,] 0.07769184 0.07553828
##
##
## $delta
## $delta$delta1
## [,1]
## [1,] -0.0009834361
## [2,] 0.0052917396
##
## $delta$delta2
## [,1]
## [1,] -0.0006726377
## [2,] -0.0075457736
##
## $delta$delta3
## [,1]
## [1,] -0.004590173
## [2,] -0.003868090
##
## $delta$delta4
## [,1]
## [1,] -0.00400828
## [2,] -0.00572946
##
## $delta$delta5
## [,1]
## [1,] 0.006410525
## [2,] 0.006045360
##
##
## $sigma
## [,1]      [,2]
## [1,] 0.09566935 0.0313946
## [2,] 0.03139460 0.1195995
##
## attr("class")
## [1] "regime"
##
## $R2

```

```

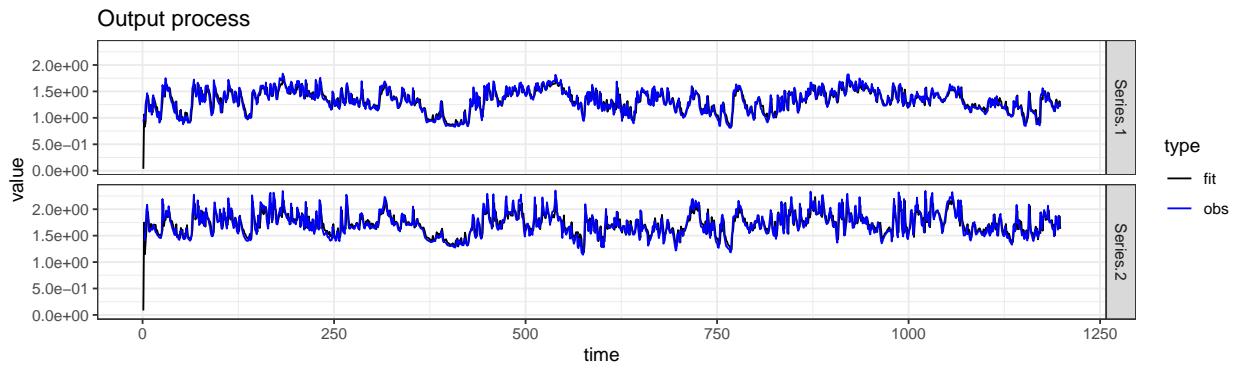
## $cs
## [,1]
## [1,] 0.3823591
## [2,] 0.4342281
##
## $phi
## $phi$phi1
## [,1]      [,2]
## [1,] 0.53957835 0.1586914
## [2,] 0.07399939 0.7412221
##
## $phi$phi2
## [,1]      [,2]
## [1,] 0.05146681 -0.09306484
## [2,] -0.06866379 -0.18583195
##
## $phi$phi3
## [,1]      [,2]
## [1,] 0.04975343 -0.07932780
## [2,] 0.09145390  0.04447205
##
## $phi$phi4
## [,1]      [,2]
## [1,] 0.08556435 0.03050941
## [2,] 0.07331243 -0.04630844
##
## $phi$phi5
## [,1]      [,2]
## [1,] 0.007948546 -0.03809468
## [2,] -0.156479341 0.20303519
##
##
## $delta
## $delta$delta1
## [,1]
## [1,] 0.0058396331
## [2,] 0.0002633693
##
## $delta$delta2
## [,1]
## [1,] -0.001888600
## [2,] -0.002066474
##
## $delta$delta3
## [,1]
## [1,] 0.002497413
## [2,] 0.003250924
##
## $delta$delta4
## [,1]
## [1,] 6.463235e-05
## [2,] -5.143031e-03
##
## $delta$delta5

```

```

##          [,1]
## [1,]  0.006813348
## [2,] -0.004553428
##
## $sigma
##          [,1]      [,2]
## [1,] 0.11204223 0.04501645
## [2,] 0.04501645 0.14787866
##
## attr(),"class")
## [1] "regime"
autoplot(ns2,5)

```



Con lo anterior ya es posible realizar el pronóstico de la serie, por medio de la función “mtarforecast()”, considerando $h = 10$ observamos el pronóstico 10 pasos adelante tanto de la variable de respuesta como umbral con su respectivo intervalo de credibilidad al 95%

```

#pred1 = mtarforecast(ns2,h = 10)
pred1$forecast$estim

```

```

## $Yt
##      lower limit 2.5%      mean upper limit 97.5%      FNDP
## 1200.1    2.4475539  3.347285666    4.2634827 1.903542
## 1200.2   -0.3761248 -0.009599365   0.3552381 1.903542
## 1201.1    2.5853056  3.698632545   4.7999781 1.931211
## 1201.2   -0.6240551 -0.104975749   0.4111296 1.931211
## 1202.1    2.1667136  3.310792370   4.4568707 1.919340
## 1202.2   -1.2756340 -0.430954647   0.4041653 1.919340
## 1203.1    3.3197348  4.370953629   5.4124974 1.894866
## 1203.2   -0.7421853 -0.059658979   0.6344365 1.894866
## 1204.1    4.0319865  5.205425969   6.3822544 1.927323
## 1204.2   -0.5222509  0.149638414   0.8383163 1.927323
## 1205.1    3.6741313  5.302458655   6.9604196 1.908013
## 1205.2   -1.1262626 -0.169998068   0.7846851 1.908013
## 1206.1    4.8295710  6.225835261   7.6241858 1.976605
## 1206.2   -1.0810443 -0.093185160   0.8849631 1.976605
## 1207.1    3.8497390  5.174710638   6.4664460 1.882360
## 1207.2   -0.8071854  0.241473898   1.3120314 1.882360
## 1208.1    5.4390390  6.811474152   8.1780045 1.881713
## 1208.2   -0.4034692  0.536151091   1.5072862 1.881713
## 1209.1    6.7172925  7.809926050   8.8917626 1.915153

```

```

## 1209.2      -0.6005979  0.371499078      1.3732481 1.915153
##
## $Zt
##      lower limit 2.5%      mean upper limit 97.5%      FNDP
## 1200.1      -0.641510552 2.0648550      4.774245 1.903542
## 1201.1       1.092173673 3.7726430      6.494601 1.931211
## 1202.1       0.547256192 3.2870236      5.973239 1.919340
## 1203.1       0.728729976 3.4045851      6.092878 1.894866
## 1204.1       0.400043856 3.0792240      5.789484 1.927323
## 1205.1      -0.484626041 2.1758009      4.881662 1.908013
## 1206.1       0.452654151 3.1965049      5.959827 1.976605
## 1207.1      -0.008664283 2.7249848      5.400031 1.882360
## 1208.1      -1.820649426 0.8852595      3.594471 1.881713
## 1209.1      -0.872047771 1.8775967      4.577138 1.915153

```