

Series de Tiempo

Proyecto Final

Alejandra Lelo de Larrea Ibarra c.u. 000124433
Victor Quintero Mármol González c.u. 000175897

15 de mayo, 2018.

1. Introducción

Se tienen datos de la cantidad de Monóxido de Carbono (CO) promedio en partículas por millón (ppm) para la Ciudad de México desde 1995. De acuerdo con la norma vigente (NOM-021-SSA1-1993) se establece un límite para la concentración en el medio ambiente de 11 ppm para un promedio de 8 horas.

La medición de este componente del aire es relevante en temas de salud, pues puede causar afectaciones al corazón o al cerebro y es considerada uno de los tipos más comunes de envenenamiento ya que inhabilita el transporte de oxígeno a las células. Además, también es relevante en temas de políticas públicas. Por ejemplo, el nivel del CO es uno de los indicadores para definir si la calidad del aire es óptima o no y por ende declarar contingencias ambientales. La SEDEMA considera los siguientes parámetros:¹

Calidad	Rango CO
Buena	0.00-5.50
Regular	5.51-11.00
Mala	11.01-16.50
Muy Mala	16.51-22.00
Extremadamente Mala	>22.00

El objetivo de este trabajo es ajustar distintos modelos clásicos (ARIMA's o SARIMA's), así como un Modelo Dinámico Lineal (DLM) para posteriormente comprar el desempeño de ambas metodologías.

2. Datos

De la Secretaría del Medio Ambiente (SEDEMA) se obtuvieron los datos del nivel de CO por hora en la Ciudad de México. Para facilitar la estimación y visualización de los modelos, se agregan los datos mensualmente de tal manera que la serie con la que se trabajará contiene 276 observaciones correspondientes al promedio mensual de CO de enero de 1995 a diciembre del 2017.

La figura 1 muestra la serie de tiempo del CO antes descrita, mientras que la figura 2 hace un “zoom” a los últimos 10 años. Como se puede ver, la serie tiene una tendencia decreciente y una componente estacional pues las máximas concentraciones de CO se tienen en los meses de diciembre y enero, mientras que las más bajas concentraciones se tienen en los meses de junio y julio. Esto hace sentido pues durante el invierno se quema mayor cantidad de combustible y, por lo tanto, se genera mayor concentración de este gas que en el verano. Además, la serie parece tener cambios en la varianza.

¹Recuperado de: <http://www.aire.cdmx.gob.mx/default.php>

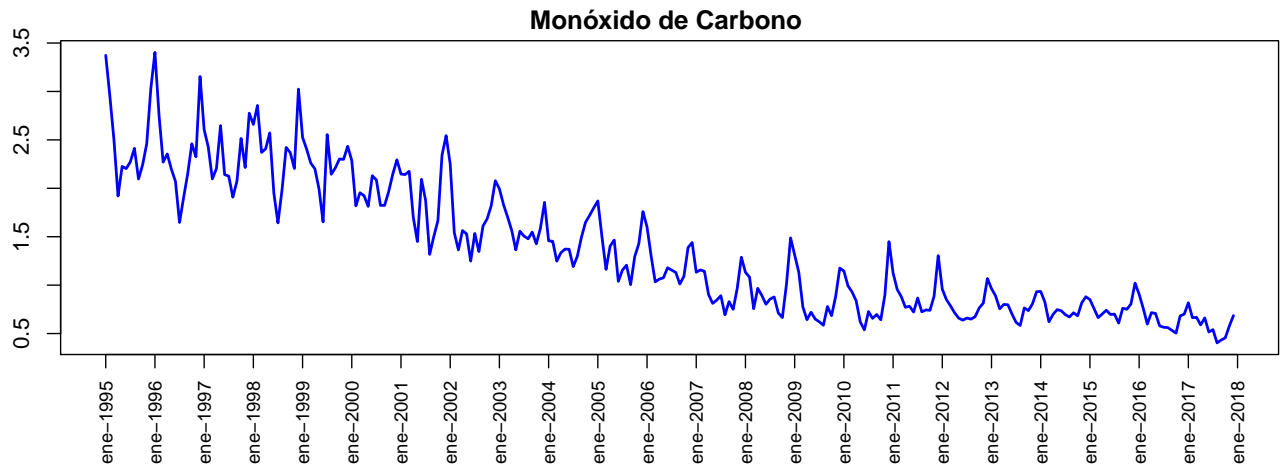


Figura 1: Histórico del promedio mensual de Monóxido de Carbono en la Ciudad de México.

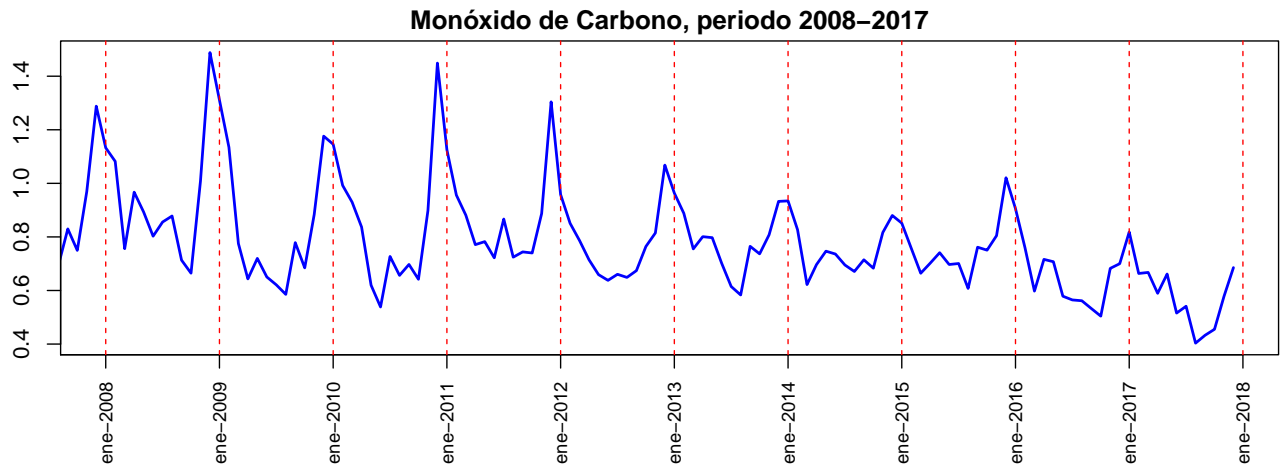


Figura 2: Promedio mensual del Monóxido de Carbono para el periodo ene 2008 - dic 2017.

Para tratar de eliminar los cambios en la varianza, la figura 3 muestra el logaritmo del Monóxido de Carbono (Log-CO) para la serie completa, mientras que la figura 4 muestra el Log-CO para los últimos 10 años. Se puede ver que la varianza es más homogénea y que persisten la tendencia, así como la estacionalidad de la serie.

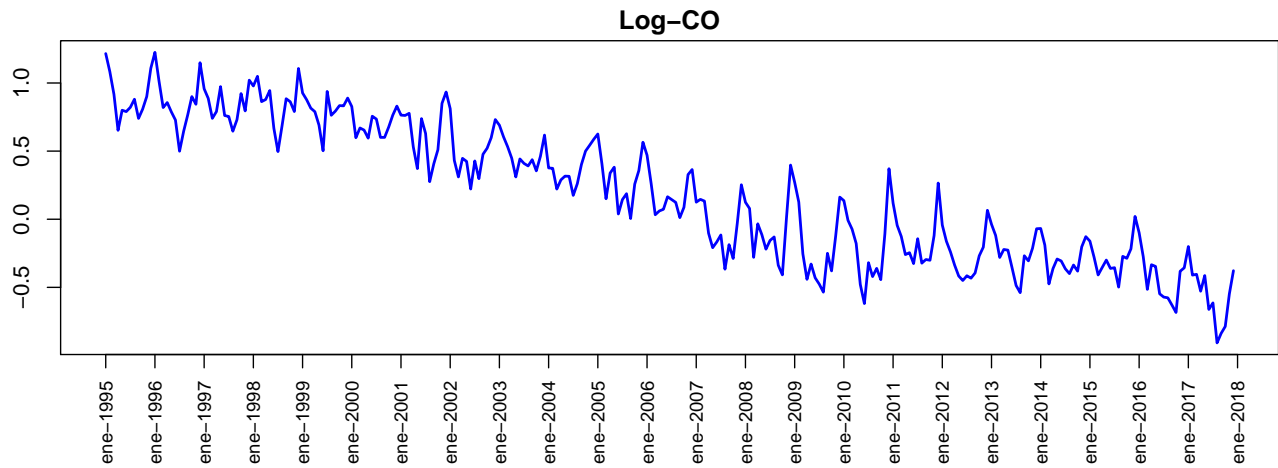


Figura 3: Histórico del promedio mensual del logaritmo del Monóxido de Carbono en México.

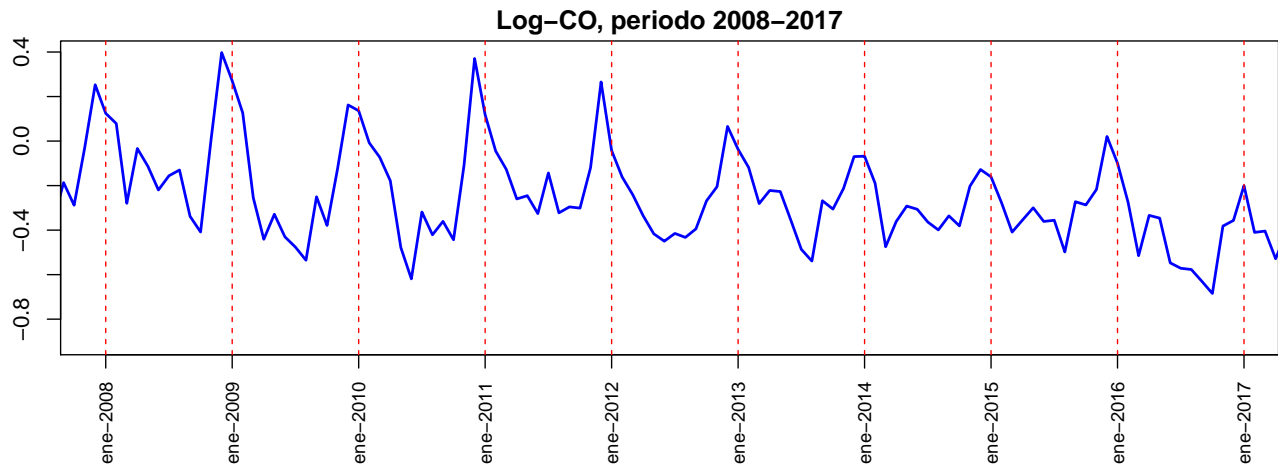


Figura 4: Promedio mensual del logaritmo del Monóxido de Carbono para el periodo ene 2008 - dic 2017.

La figura 5 muestra el ACF y PACF de los datos. El ACF tiene varios rezagos significativos y muestra el comportamiento de una serie con componente estacional. Por su parte el PACF tiene algunos rezagos significativos pero se corta rápidamente.

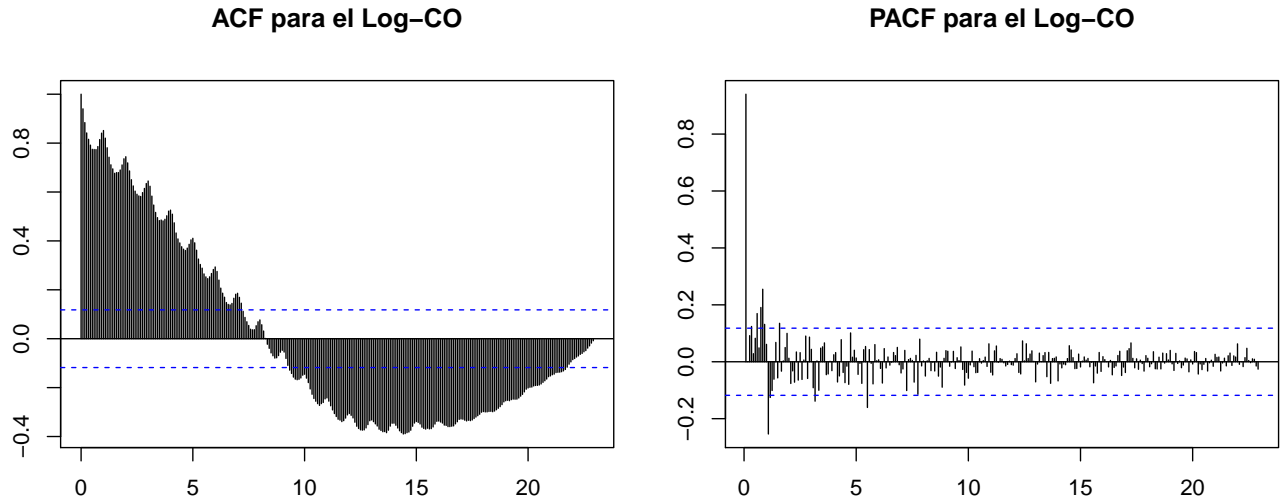


Figura 5: ACF y PACF de la serie del logaritmo del Monóxido de Carbono.

Para eliminar la estacionalidad y la tendencia diferenciamos la serie. La figura 6 muestra la serie del Log-CO diferenciada de primer orden. Como se puede observar, la tendencia se ha eliminado. Sin embargo, al hacer “zoom” en la figura 7 se puede notar que todavía existe un componente estacional.

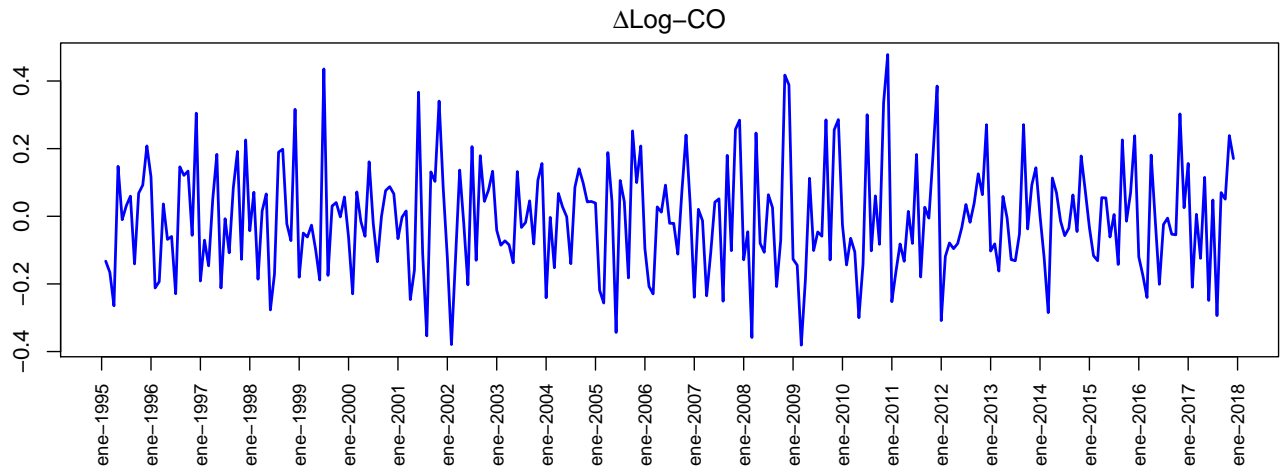


Figura 6: Histórico de la diferencia en el promedio mensual del logaritmo del Monóxido de Carbono en México.

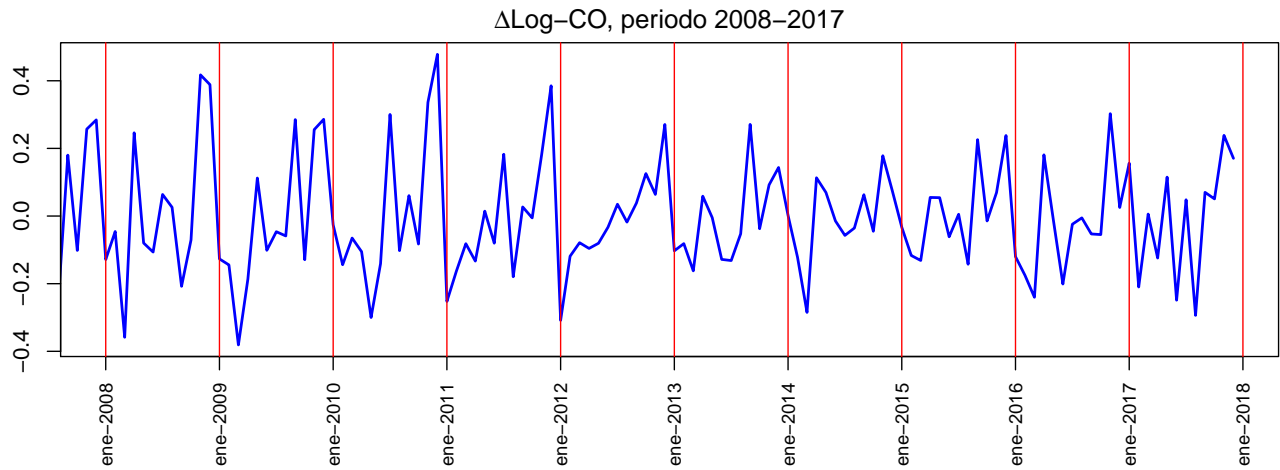


Figura 7: Primer diferencia del promedio mensual del logaritmo del Monóxido de Carbono para el periodo ene 2008 - dic 2017.

El ACF y el PACF de la serie diferenciada se muestran en la figura 8. Si bien ya se eliminó el patrón estacional que mostraba el ACF, sigue habiendo rezagos significativos que parecen decrecer con el tiempo. En cuanto al PACF, ésta sigue mostrando algunos rezagos significativos que se cortan rápidamente después de 12 rezagos.

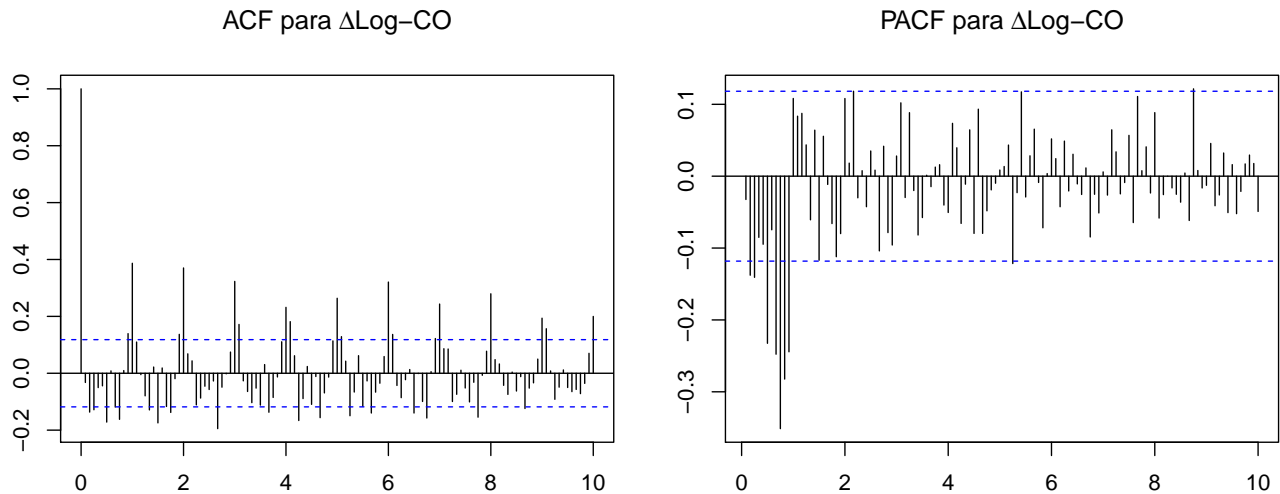


Figura 8: ACF y PACF de la primer diferencia del promedio mensual del logaritmo del Monóxido de Carbono.

Para terminar de eliminar la estacionalidad, diferenciamos de orden 12 la serie diferenciada de orden 1 pues los datos son mensuales. La figura 9 muestra que la serie ya no presenta tendencia ni estacionalidad; sin embargo, todavía persisten algunos cambios en la varianza hacia el final de la serie.

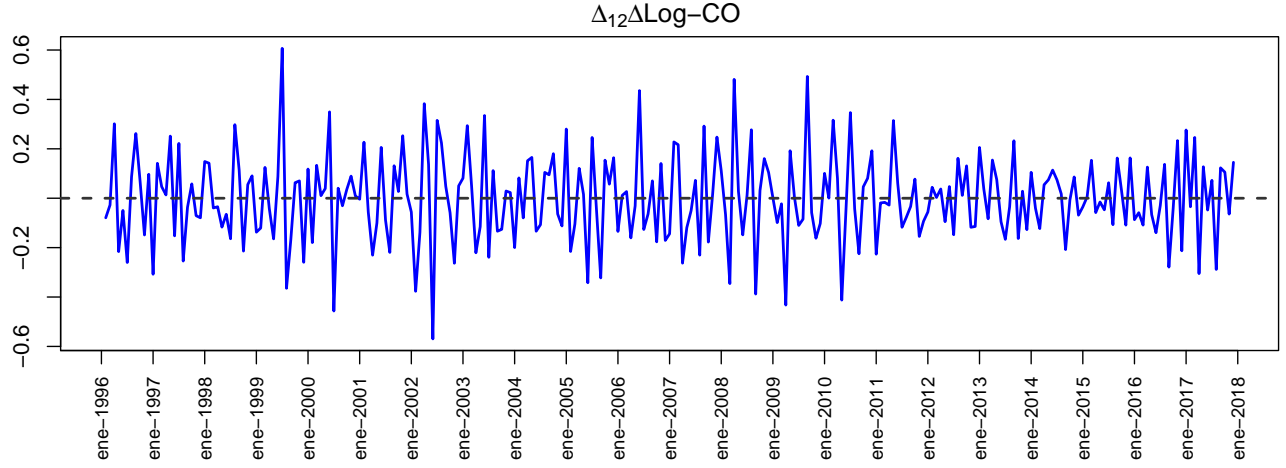


Figura 9: Histórico de la diferencia de orden 12 de la diferencia en el promedio mensual del logaritmo del Monóxido de Carbono en México.

La figura 10 muestra el ACF y PACF de esta última serie. El ACF se corta después del rezago 12 aproximadamente y muestra un valor significativo en el rezago 108 (aproximadamente). La PACF se corta después de 36 rezagos (o podría decrecer exponencialmente). Esto nos hace pensar que un modelo SARIMA(p,d,q)(P,D,Q)_s es buena opción para ajustar la serie o bien, un Modelo Dinámico Lineal polinomial de orden 2 con estacionalidad.

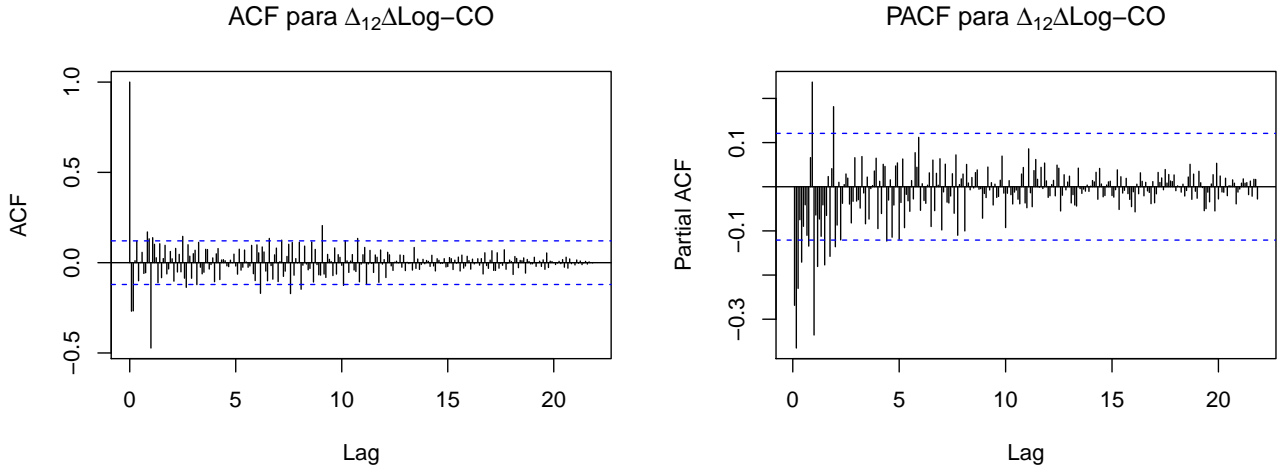


Figura 10: ACF y PACF para la serie $\Delta_{12}\Delta\text{Log} - CO$ promedio mensual.

3. Modelos Clásicos

3.1. Selección de Modelos

De la sección anterior se puede ver que dentro de los modelos clásicos un SARIMA(p,d,q)×(P,D,Q)_s es una buena opción para intentar ajustar un modelo a los datos del Log-CO. Es claro que los parámetros correspondientes a las diferencias (d y D) deben ser iguales a uno y que la periodicidad (s) debe ser igual a 12. Sin embargo, no es tan claro el valor que deben tomar los parámetros p, q, P y Q.

Por ello, creamos una función para estimar distintos modelos SARIMA(p,d,q)×(P,D,Q)_s. Se hace una búsqueda exhaustiva en grid con código en paralelo en R tomando todos los modelos que surgen de las posibles combinaciones de los siguientes valores para cada parámetro:

- $p = 1, 2, 3, 4, 5$
- $d = 1$
- $q = 1, 2, 3, 4, 5$
- $P = 1, 2, 3, 4, 5$
- $D = 1$
- $Q = 1, 2, 3, 4, 5$
- $s = 12$

Cuadro 1: Comparación de AIC's para modelos SARIMA(p,d,q)x(P,D,Q)s

Modelo							Criterio	
p	d	q	P	D	Q	s	Log-Like.	AIC
2	1	4	0	1	1	12	208.7130	-401.4260
0	1	2	0	1	1	12	204.3997	-400.7993
2	1	1	0	1	1	12	205.2593	-400.5186
0	1	3	0	1	1	12	205.2478	-400.4956
0	1	2	1	1	2	12	206.1918	-400.3836
2	1	5	0	1	1	12	209.0242	-400.0484
1	1	2	0	1	1	12	205.0049	-400.0099
2	1	4	0	1	2	12	208.7905	-399.5810
1	1	1	0	1	1	12	203.7820	-399.5641
0	1	4	0	1	1	12	205.6646	-399.3292
2	1	1	1	1	2	12	206.5513	-399.1026
0	1	2	1	1	1	12	204.4024	-398.8048
0	1	2	0	1	2	12	204.4021	-398.8041
0	1	2	2	1	2	12	206.3989	-398.7979
3	1	1	0	1	1	12	205.3504	-398.7008
2	1	2	0	1	1	12	205.3028	-398.6056
0	1	3	1	1	1	12	205.2710	-398.5420
0	1	3	0	1	2	12	205.2682	-398.5364
2	1	1	1	1	1	12	205.2667	-398.5335
2	1	1	0	1	2	12	205.2658	-398.5316
2	1	1	2	1	2	12	207.2474	-398.4949
0	1	3	2	1	2	12	207.2065	-398.4130
1	1	3	0	1	1	12	205.1689	-398.3378
0	1	2	0	1	3	12	205.1589	-398.3179
0	1	2	2	1	1	12	205.0925	-398.1849
2	1	4	0	1	3	12	209.0648	-398.1296
3	1	1	1	1	2	12	207.0324	-398.0648
2	1	1	0	1	3	12	206.0226	-398.0453
1	1	2	1	1	1	12	205.0215	-398.0430
1	1	2	0	1	2	12	205.0195	-398.0389

En total se ajustaron 1296 modelos; de los cuales 288 no pudieron ser estimados con la función *arima* de R. Los modelos resultantes, se ordenaron de acuerdo con el valor AIC obtenido. La tabla 1 muestra los parámetros, la verosimilitud y el AIC únicamente para los 30 modelos con menor AIC. De éstos, se eligen los 5 primeros para competir y realizar las pruebas de diagnóstico. Es decir, se validará la estimación de los siguientes modelos:

- Modelo 1: $SARIMA(2, 1, 4)X(0, 1, 1)_{12}$
- Modelo 2: $SARIMA(0, 1, 2)X(0, 1, 1)_{12}$
- Modelo 3: $SARIMA(2, 1, 1)X(0, 1, 1)_{12}$
- Modelo 4: $SARIMA(0, 1, 3)X(0, 1, 1)_{12}$
- Modelo 5: $SARIMA(0, 1, 2)X(1, 1, 2)_{12}$

Se puede notar que los mejores modelos tienen parametros de orden chico para la parte estacional (P=1 y Q=1 en 4 de los 5 modelos), pero para la parte no estacional sí se obtuvieron modelos con parámetros AR y MA más grandes. El valor del AIC es muy similar entre los 5 modelos; sin embargo, hay mayor variabilidad en el valor de la verosimilitud.

3.2. Diagnóstico de Residuales

Cuadro 2: Resultados de la prueba Ljung-Box para modelos SARIMA(p,d,q)x(P,D,Q)s

Modelo							Ljung-Box
p	d	q	P	D	Q	s	valor-p
2	1	4	0	1	1	12	0.9881
0	1	2	0	1	1	12	0.8495
2	1	1	0	1	1	12	0.9885
0	1	3	0	1	1	12	0.9883
0	1	2	1	1	2	12	0.9021

Para los 5 modelos seleccionados se realizan los diagnósticos de los residuales. Primero se obtiene la prueba Ljung-Box para ver si los residuales están o no correlacionados entre si. Como se muestra en la tabla 2, los cinco modelos pasan la prueba; por lo tanto, no hay evidencia de errores autocorrelacionados.

La figura 11 muestra el ACF y PACF para los residuales de los 5 modelos. Todos los modelos tienen rezagos de los residuales no significativos para la ACF. Sin embargo, para la PACF se tiene un rezago “marginamente” significativo (el rezago 48) en los modelos 2, 3 y 4. Dependiendo del resto de los diagnósticos podría pasarse por alto este indicador.

Por su parte, la figura 12 muestra el histograma de los residuales, el histograma de una normal simulada para cada modelo y sus respectivos Q-Q plots. La distribución de los residuales parecen tener cola derecha ligeramente más pesada que los de una normal en los cinco casos. Además, nuevamente los modelos 2, 3 y 4 parecen ser aquellos para los que el histograma de los residuales es más distinto al histograma de una normal simulada. Los Q-Q plots de los 5 modelos muestran mayor diferencia respecto a la línea de 45° en la cola derecha.

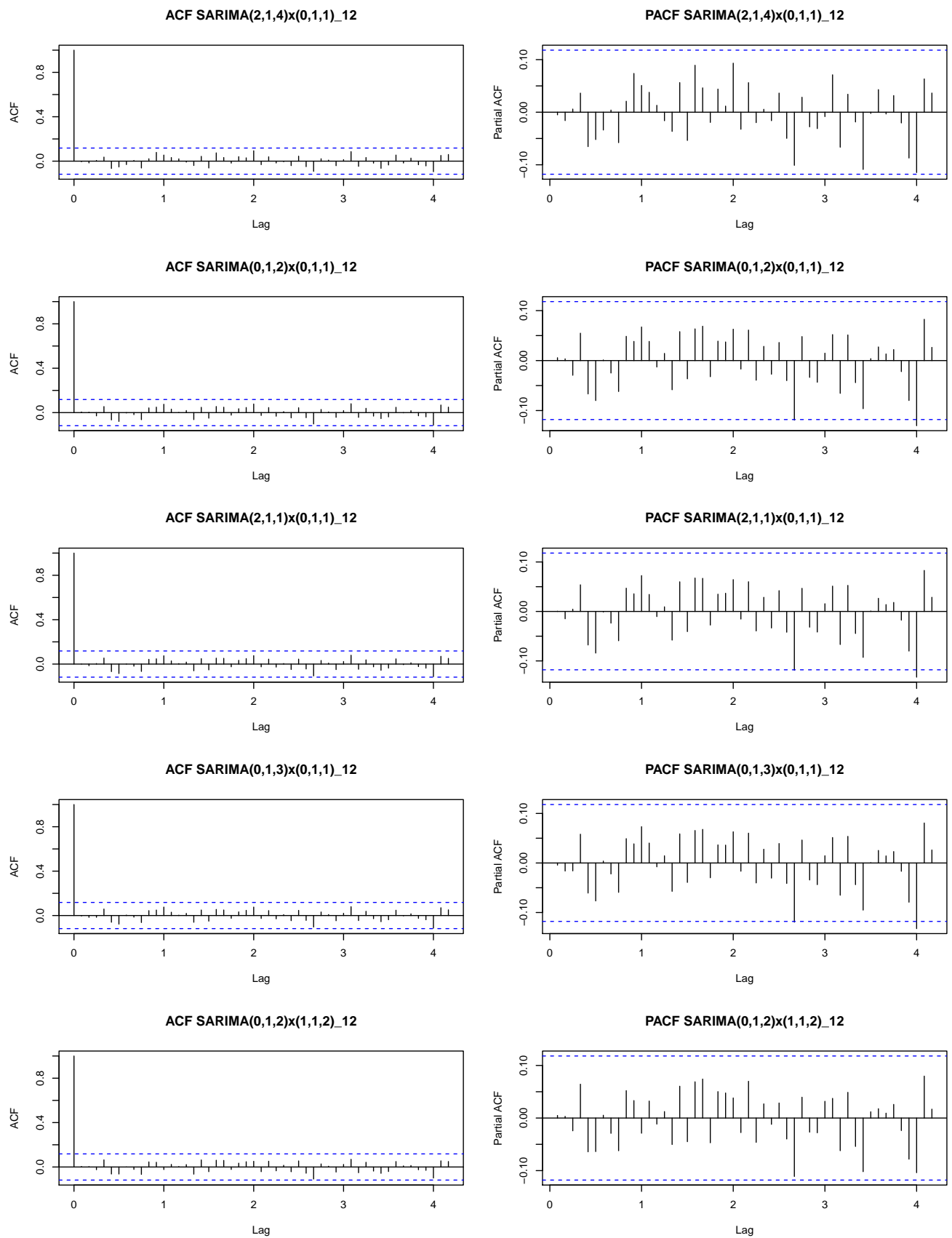


Figura 11: ACF y PACF para los residuales de cada modelo.

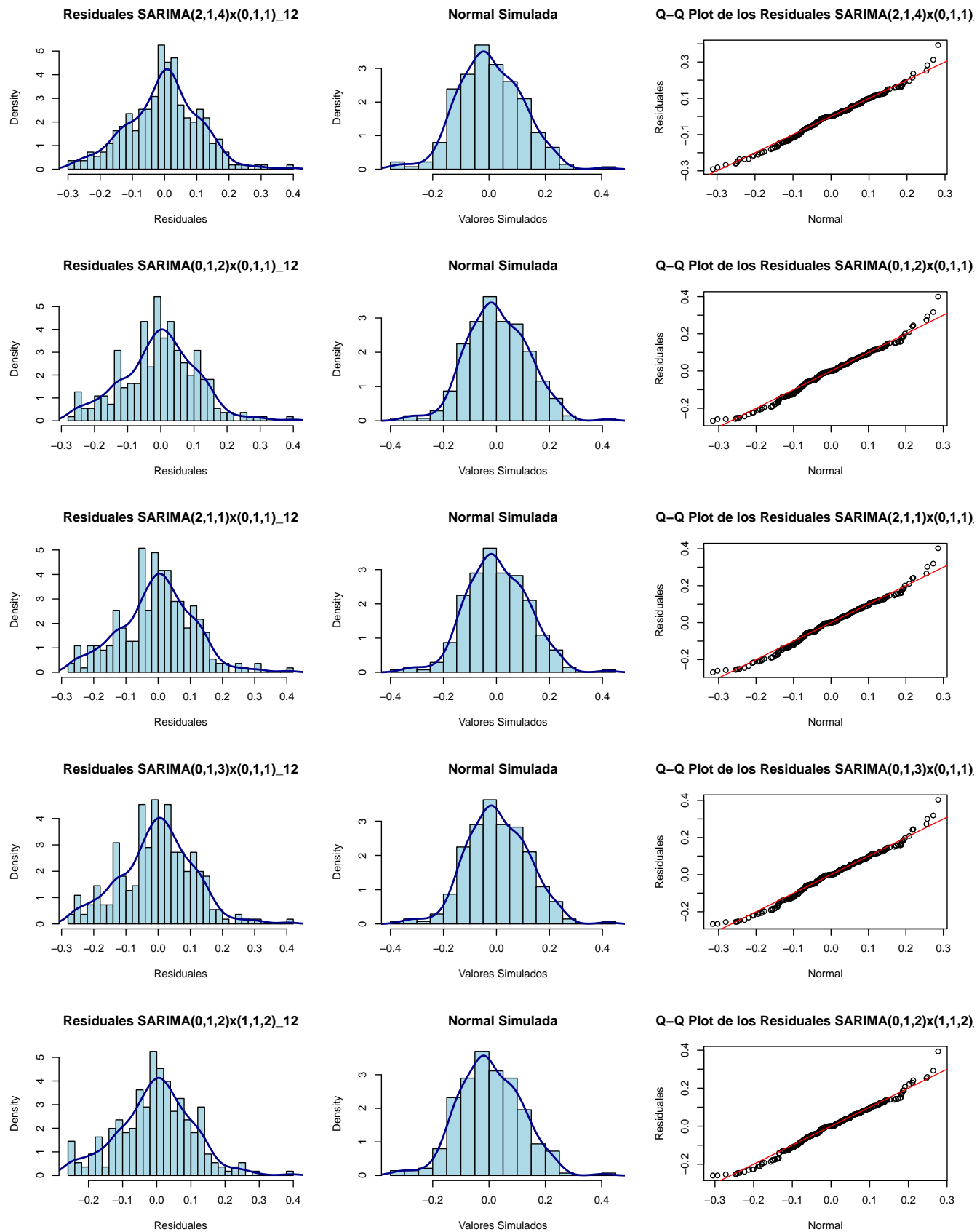


Figura 12: Histograma y Q-Q Plot para los residuales de cada modelo.

3.3. ECM para predicciones

Por último, para evaluar la calidad de los modelos planteados, se reserva el 20 % final de la muestra para calcular el Error Cuadrático Medio (ECM) de las predicciones realizadas por cada uno. En la tabla 3 se puede observar que todos los modelos presentan un ECM muy pequeño de aproximadamente 0.014, pero es el primer modelo ($SARIMA(0, 1, 2) \times (1, 1, 2)_{12}$) el que da el menor ECM para las predicciones del Log-CO. Cabe destacar que la diferencia en ECM entre el primer y quinto modelo es muy pequeña, pero el quinto modelo tiene 2 parámetros menos a estimar que el primero.

Cuadro 3: Comparación ECM's para modelos $SARIMA(p,d,q)X(P,D,Q)_s$

Modelo								Criterio
p	d	q	P	D	Q	s		ECM
2	1	4	0	1	1	12		0.014716
0	1	2	0	1	1	12		0.014900
2	1	1	0	1	1	12		0.014955
0	1	3	0	1	1	12		0.014959
0	1	2	1	1	2	12		0.014728

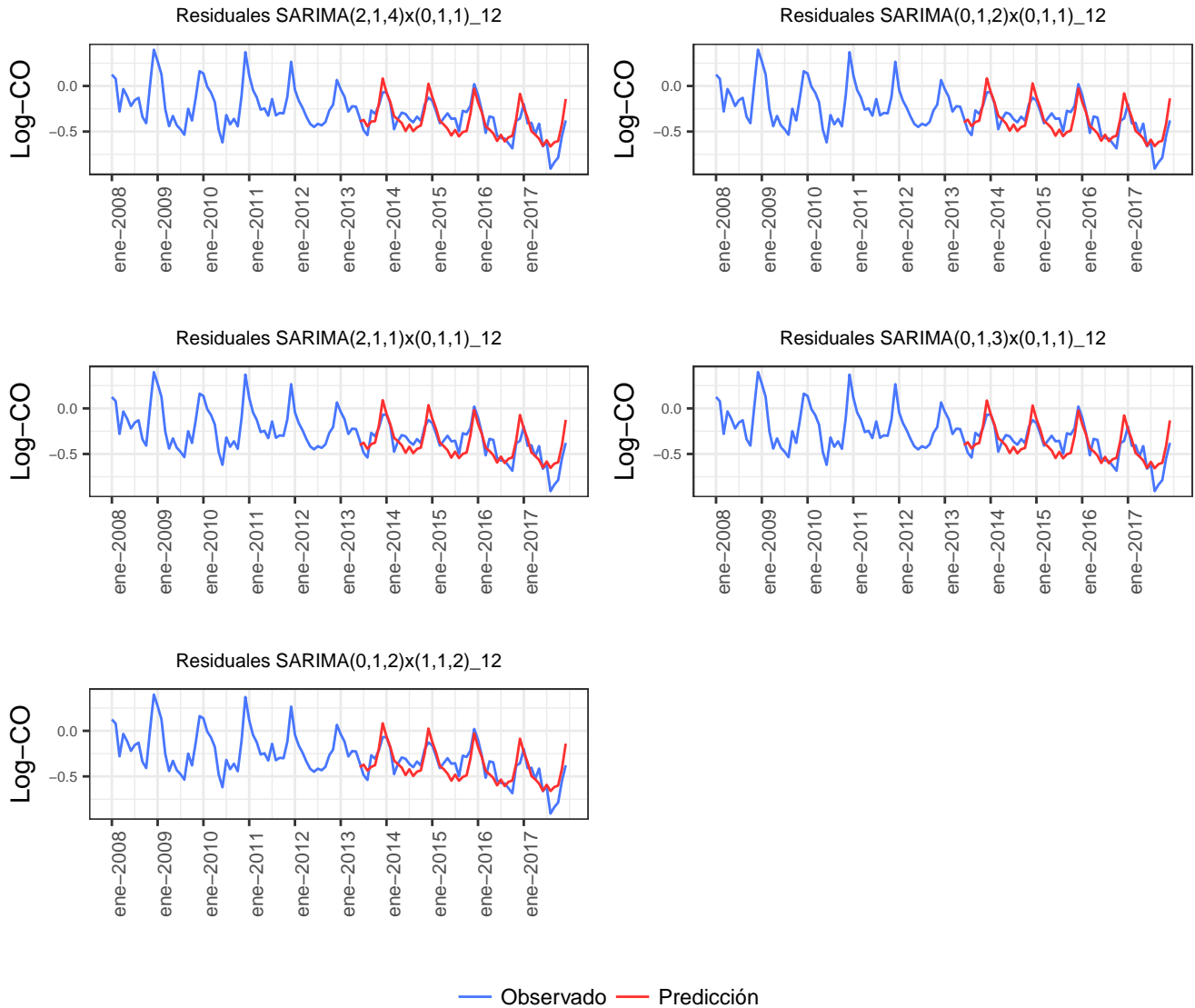


Figura 13: Valores observados vs predicciones para cada modelo para el periodo 2008-2017.

Por su parte, la figura 13 muestra la serie observada (azul) y los valores pronosticados para el 20 % final de la muestra (rojo). Los 5 modelos reproducen de manera correcta la estacionalidad de la serie pero subestiman ligeramente los valles salvo por el final del 2017.

3.4. Elección de Modelo SARIMA

Tomando en consideración las gráficas del ACF, PACF, el AIC, el análisis de residuales y el ECM obtenidos en las secciones anteriores; y manteniendo en la mira la idea de tener un modelo parsimonioso, se decide utilizar el quinto modelo como el mejor modelo para ajustar los datos; ya que de los únicos dos modelos que pasaron todas las pruebas de los residuales (modelos 1 y 5), es el que tiene menor número de parámetros a estimar y la diferencia en ECM no es significativa. Esto es, los datos del Log-CO se van a modelar con un $SARIMA(0, 1, 2)X(1, 1, 2)_{12}$ el cual tiene 6 parámetros a estimar (incluyendo el valor de σ^2). De esta manera, se tiene que el modelo final está dado por:

$$(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})CO_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12} + \Theta_2 B^{24})Z_t,$$

o bien, si desarrollamos los polinomios:

$$\begin{aligned} CO_t = & CO_{t-1} + (1 - \Phi_1)CO_{t-12} - (1 + \Phi_1)CO_{t-13} - \Phi_1 CO_{t-24} + \Phi_1 CO_{t-25} \\ & + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \Theta_1 Z_{t-12} + \theta_1 \Theta_1 Z_{t-13} \\ & + \theta_2 \Theta_1 Z_{t-14} + \Theta_2 Z_{t-24} + \theta_1 \Theta_2 Z_{t-25} + \theta_2 \Theta_2 Z_{t-26} \end{aligned} \quad (1)$$

donde Z_t es ruido blanco con media cero y varianza σ^2 . Los valores estimados para los parámetros son:

- $\theta_1 = -0.5078$
- $\theta_2 = -0.3345$
- $\Phi_1 = -0.7622$
- $\Theta_1 = -0.0518$
- $\Theta_2 = -0.6546$
- $\sigma^2 = 0.0143$

De esta manera, sustituyendo el valor de los parámetros en (1), se tiene que el mejor modelo SARIMA es

$$\begin{aligned} CO_t = & CO_{t-1} + 0.2378CO_{t-12} - 0.2378CO_{t-13} + 0.7622CO_{t-24} - 0.7622CO_{t-25} \\ & + Z_t - 0.5078Z_{t-1} - 0.3345Z_{t-2} - 0.0518Z_{t-12} + 0.0263Z_{t-13} + 0.0173Z_{t-14} - 0.6546Z_{t-24} \\ & + 0.3324Z_{t-25} + 0.2189Z_{t-26} \end{aligned} \quad (2)$$

donde Z_t es ruido blanco con media cero y varianza $\sigma^2 = 0.0143$.

4. Modelo Dinámico Lineal

Si se quiere estimar un Modelo Dinámico Lineal para el Log-CO, de la gráfica 3 se ve que se necesita un DLM polinomial de segundo orden junto con un DLM estacional. Para el DLM polinomial se tiene el siguiente modelo Espacio-Estado:

$$\begin{aligned} Y_t^{(1)} &= A_1 X_t^{(1)} + V_t^{(1)} \\ X_{t+1}^{(1)} &= G_1 X_t^{(1)} + W_t^{(1)} \end{aligned}$$

donde $A_1 = [1, 0]$, $V_t^{(1)} = V^{(1)}$, $W_t^{(1)} = \text{diag}(W_1, W_2)$ y $G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Por su parte, para el DLM estacional con periodo $s = 12$, se tiene el siguiente modelo Espacio-Estado:

$$\begin{aligned} Y_t^{(2)} &= A_2 X_t^{(2)} + V_t^{(2)} \\ X_{t+1}^{(2)} &= G_2 X_t^{(2)} + W_t^{(2)} \end{aligned}$$

donde $A_2 = [1, 0_{1 \times 10}]$, $V_t^{(2)} = 0$, $W_t^{(2)} = \text{diag}(\sigma_W^2, 0_{1 \times 10})$ y $G_2 = \begin{bmatrix} -\mathbb{1}_{10 \times 10} & 0 \\ I_{10 \times 10} & 0_{10 \times 1} \end{bmatrix}$.

La serie del Log-CO se estimará mediante un DLM que combine los dos modelos Espacio-Estado mencionados anteriormente.

4.1. Estimación de varianzas Máximo-verosímiles

El primer paso es estimar los valores máximo verosímiles de las matrices V y W en ambos DLM'S. Después de revisar la convergencia, se obtienen los siguientes estimadores:

- $\hat{V}^{(1)} = 0.0101$
- $\hat{W}_1 = 9 \times 10^{-4}$
- $\hat{W}_2 = 0$
- $\hat{\sigma}_W^2 = 0$.

Una vez conocidas las estimaciones de las matrices V y W se puede proceder a estimar el filtrado, suavizado y la predicción del modelo DLM.

4.2. Filtrado

Utilizando la función *dmlFilter* se obtiene el filtrado para el Log-CO. La figura 14 muestra la serie filtrada para los tres estados: nivel, pendiente y estacionalidad. Se sabe que una pequeña desventaja del Filtro de Kalman es que es sensible a condiciones iniciales; este efecto se puede notar en que el inicio de los tres estados está alejado del valor promedio de la serie. Sin embargo, el filtro corrige rápidamente las trayectorias. El nivel sigue muy de cerca la dinámica de la serie observada. La pendiente decrece a lo largo de los primeros dos años y después se mantiene relativamente constante. De igual manera, la dinámica de la componente estacional parece mantenerse constante a lo largo de la muestra.

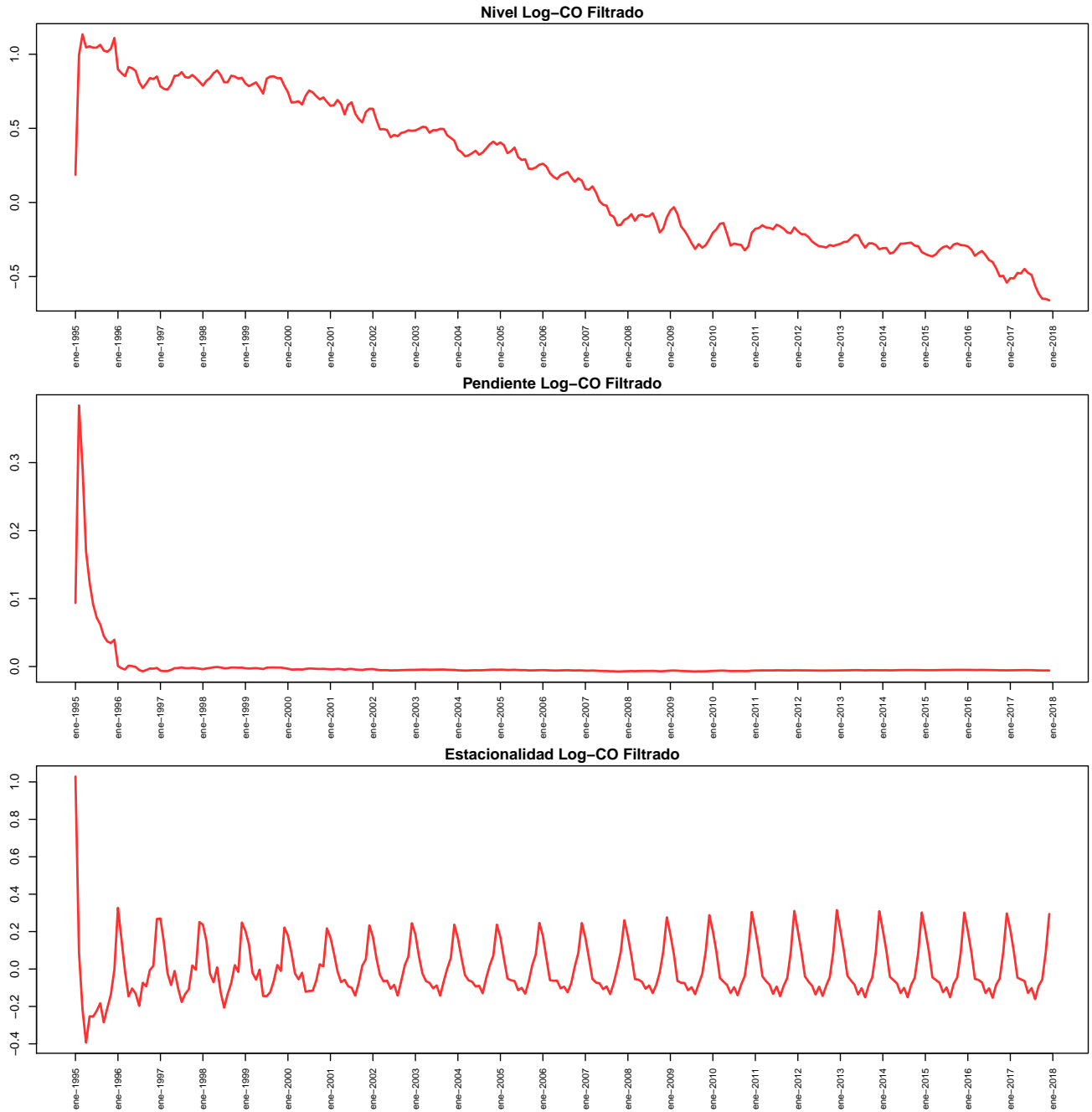


Figura 14: Serie de tiempo de los vectores de estado filtrados.

Las figuras 15 y 16 muestran la serie observada del Log-CO junto con la serie filtrada para la muestra completa y para los últimos 10 años. Como se esperaba, se puede observar una variabilidad mucho menor en la serie filtrada que en la observada, esto debido principalmente a que el cociente $V^{(1)}/W_1$ es grande ($V^{(1)}/W_1 = 11.04$).

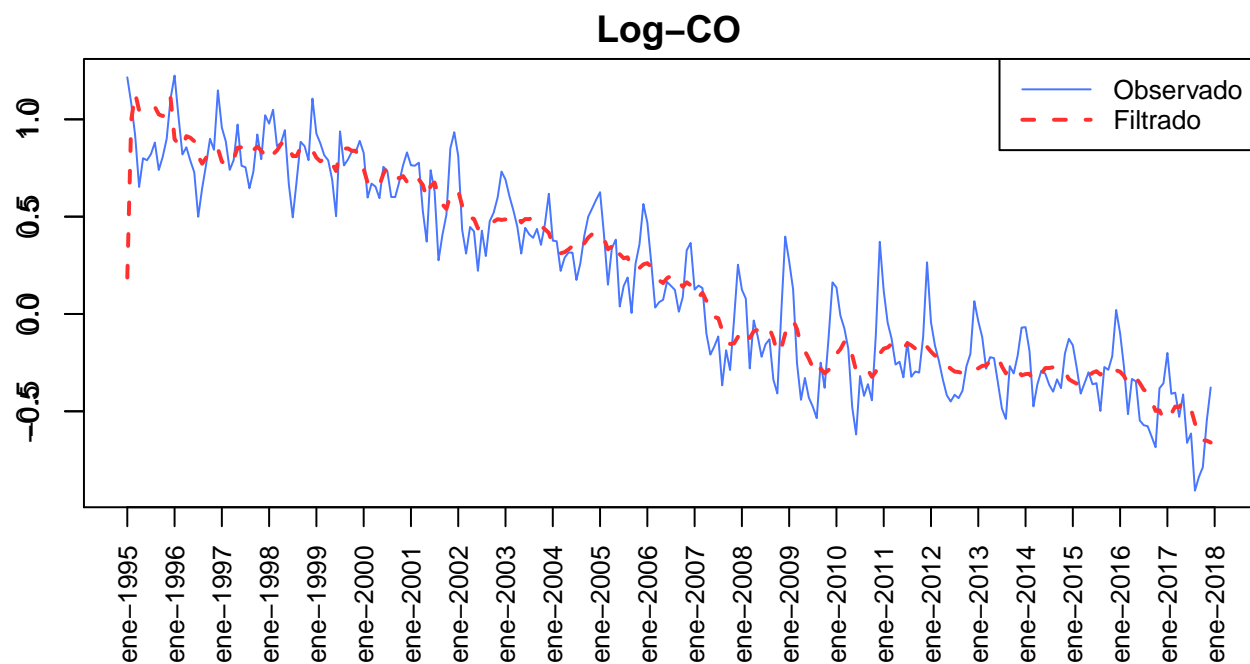


Figura 15: Serie Observada vs. Serie Filtrada

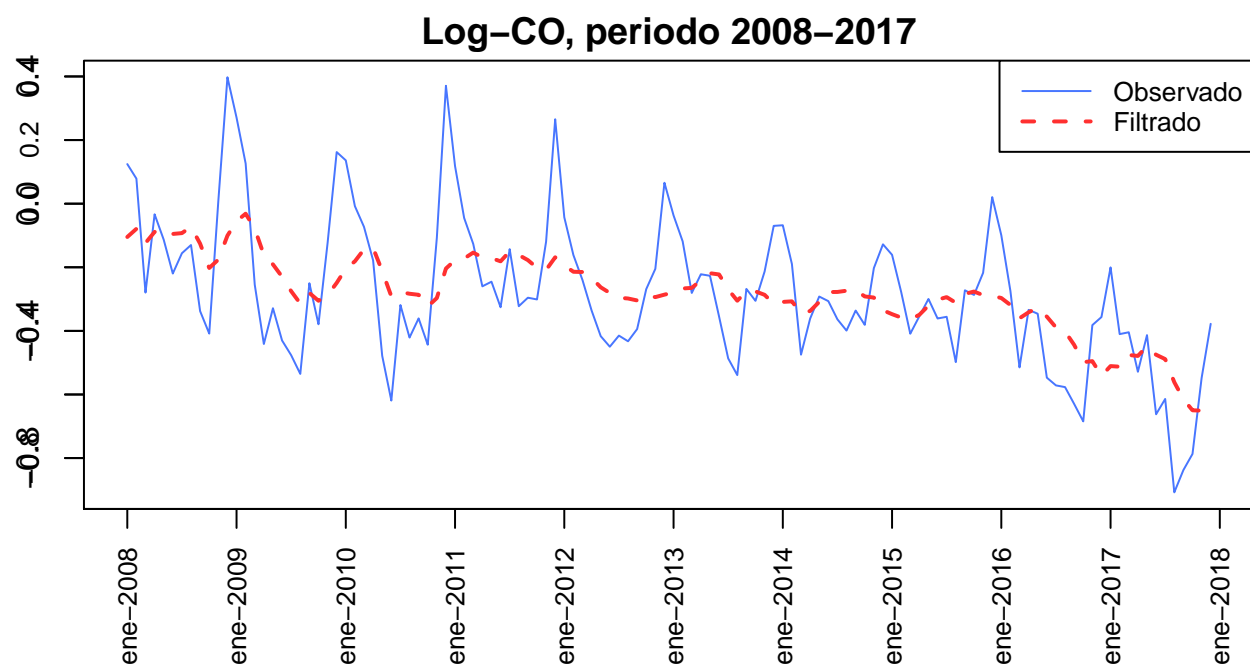


Figura 16: Serie Observada vs. Serie Filtrada para el periodo 2008-2017

4.3. Suavizado

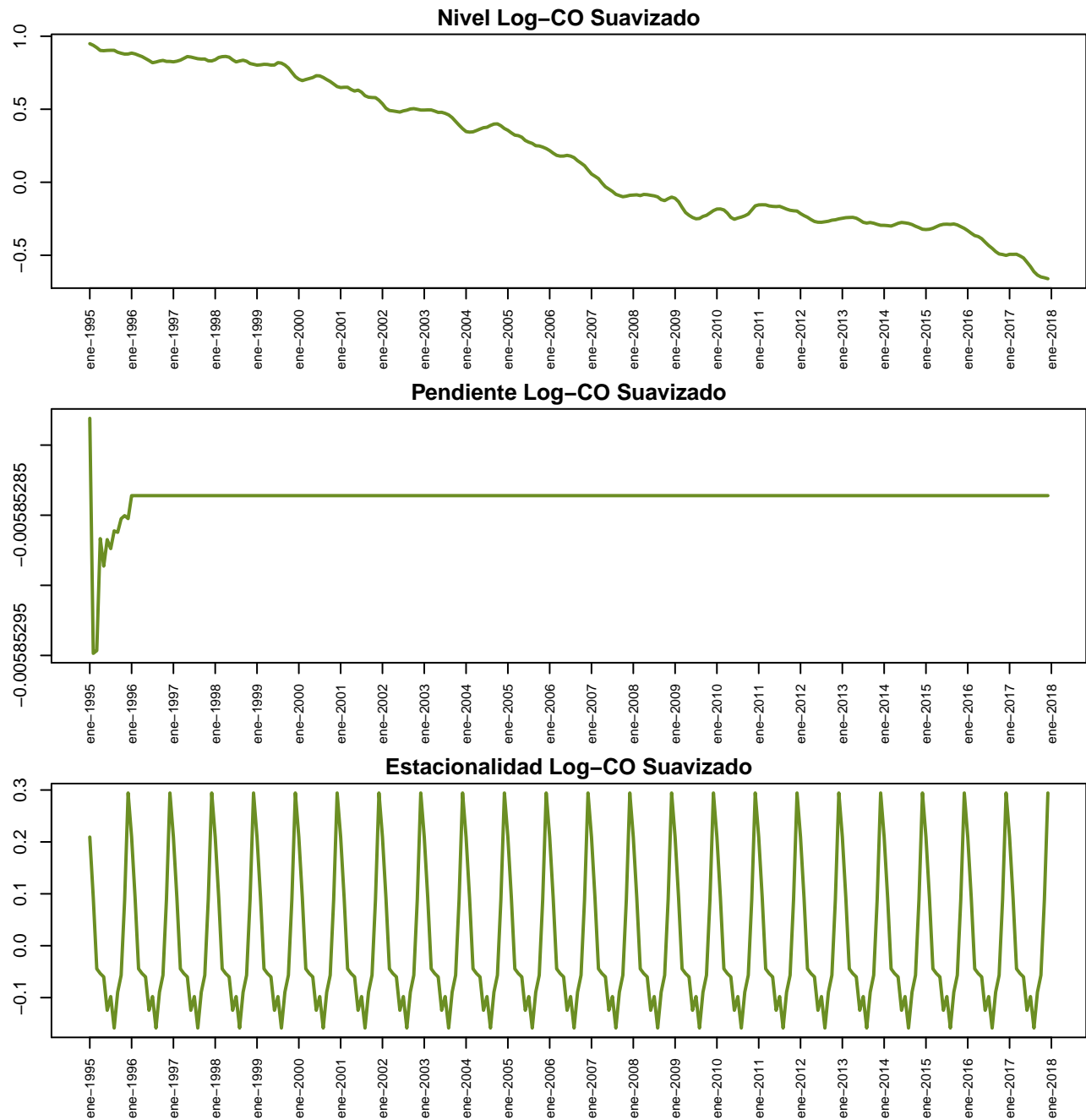


Figura 17: Serie de tiempo de los vectores de estado suavizados.

A partir del modelo filtrado, se obtiene el suavizamiento para los tres estados. La figura 17 muestra el nivel, la pendiente y la estacionalidad del Log-CO suavizados con el Filtro de Kalman. Al estar suavizadas, las tres series tienen menor variabilidad que las series mostradas en la figura 14 pues en el suavizado se toma en cuenta los valores futuros de la serie mientras que el filtrado sólo toma en cuenta las observaciones pasadas. Por su parte las figuras 18 y 19 muestran la serie observada del Log-CO y la serie suavizada.

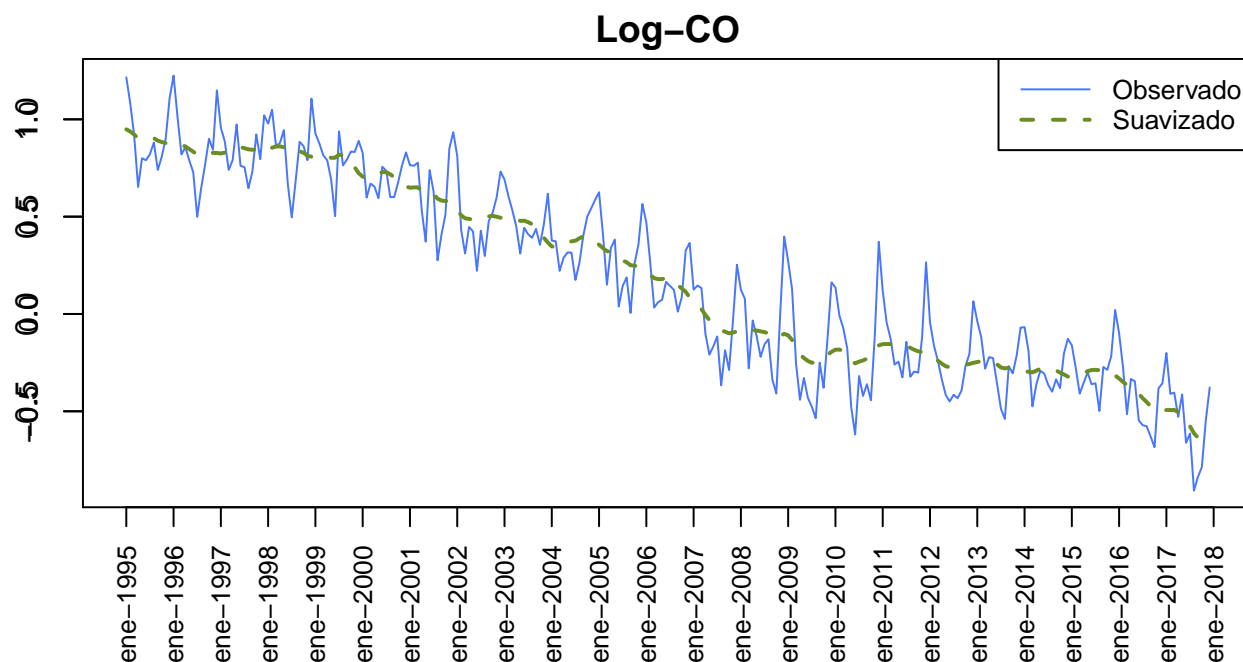


Figura 18: Serie Observada vs. Serie Suavizada

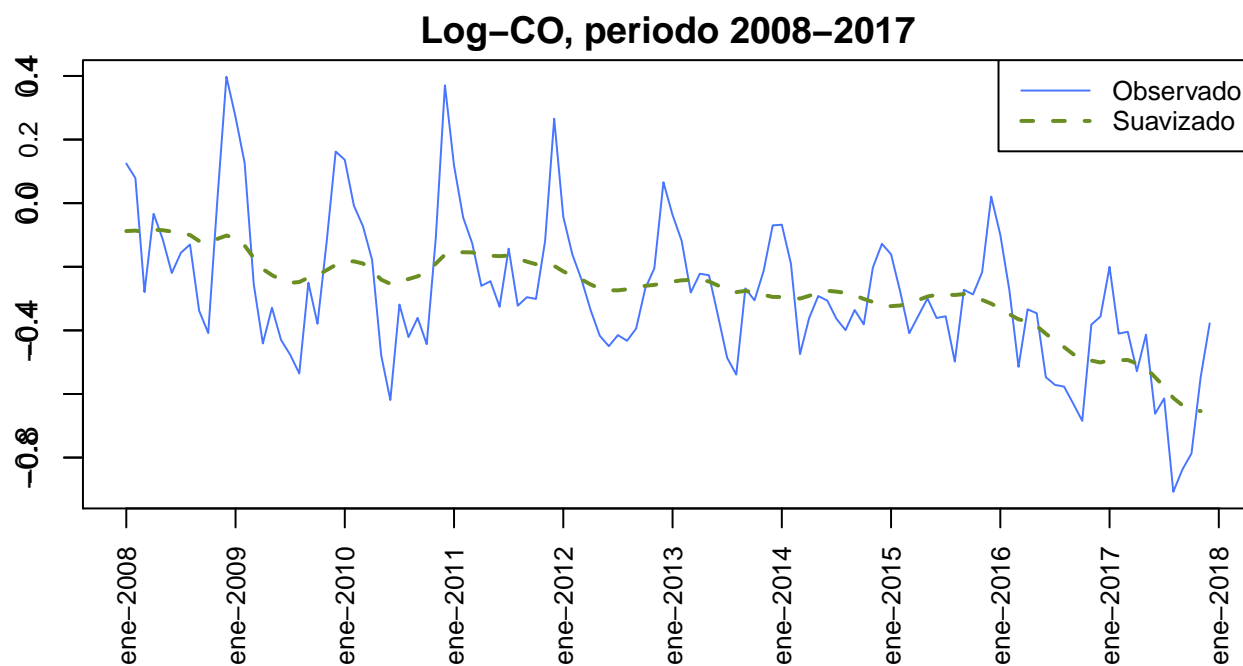


Figura 19: Serie Observada vs. Serie Suavizada para el periodo 2008-2017.

4.4. Predicción

Finalmente, se realiza la predicción para los siguientes 2 años (24 observaciones) y se simulan 1000 trayectorias. La figura 20 muestra las series filtradas, las predicciones a dos años (2018 y 2019), los intervalos de confianza al 95 %, así como las trayectorias simuladas para el nivel, la pendiente y la estacionalidad del Log-CO.

Pareciera que un mayor número de trayectorias quedan fuera del intervalo de confianza para el nivel que para la pendiente y la estacionalidad. Además, los valores pronosticados para los tres estados parecen continuar de manera

razonable con la trayectoria de las series filtradas. Cabe mencionar que la pendiente y la estacionalidad tienen intervalos de confianza muy angostos debido a que la estimación de la varianza es prácticamente cero.

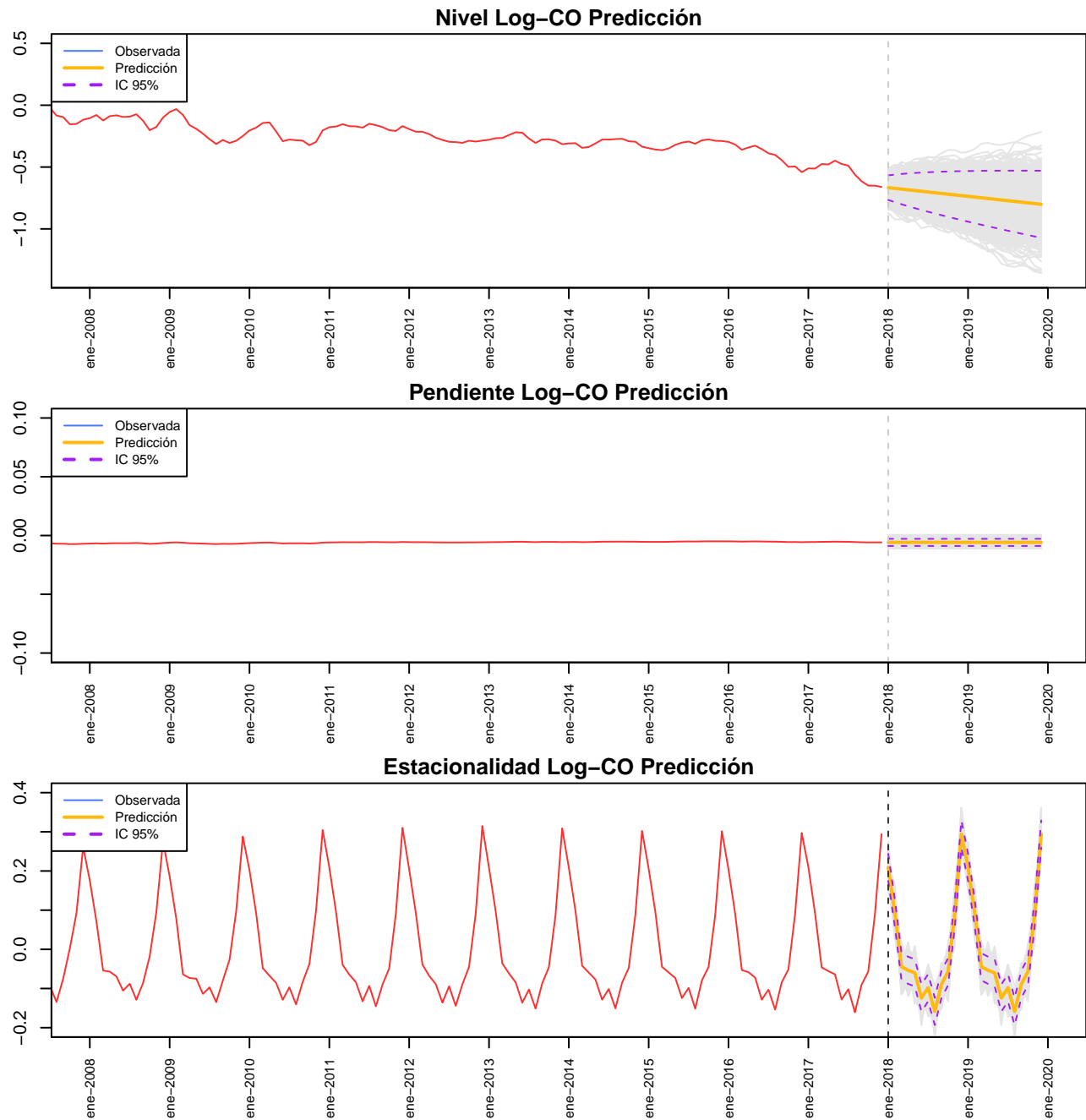


Figura 20: Serie de tiempo de las predicciones a dos años de los estados.

Por último, la figura 21 muestra la serie observada para el Log-CO, la serie filtrada y el pronóstico a dos años junto con sus intervalos de confianza al 95 %, así como las 1000 trayectorias simuladas. Se puede apreciar que el pronóstico mantiene bien la estacionalidad de la serie y es acorde a la tendencia que presentaba.

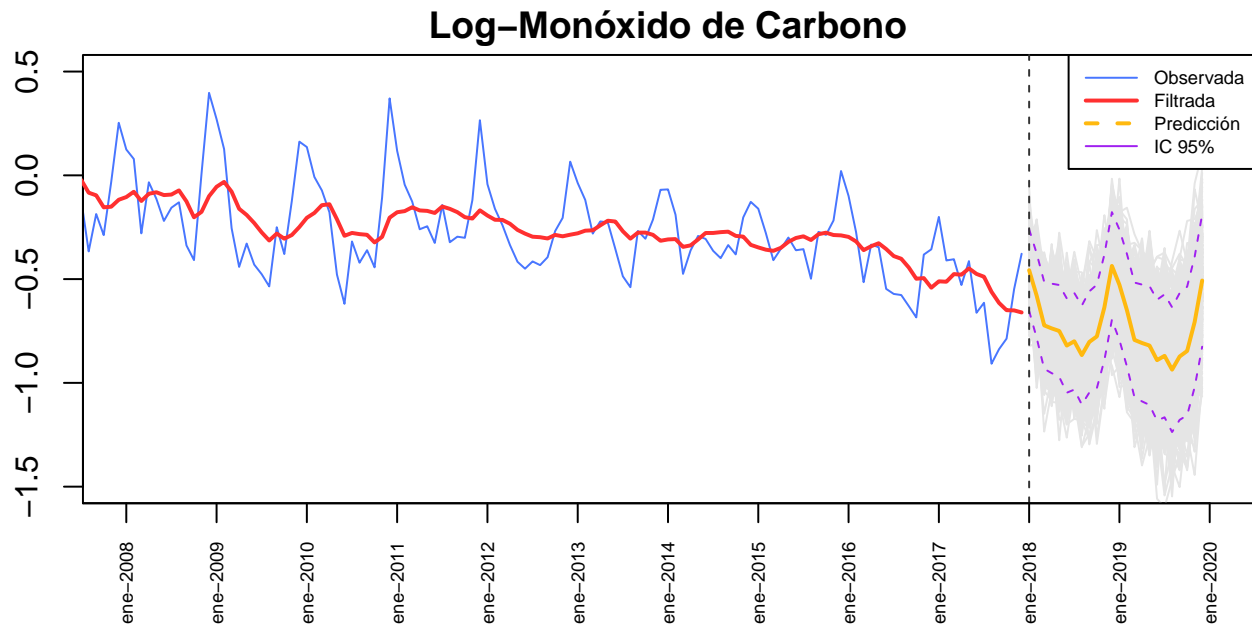


Figura 21: Serie Observada vs. Serie Pronosticada

4.5. Diagnóstico de Residuales

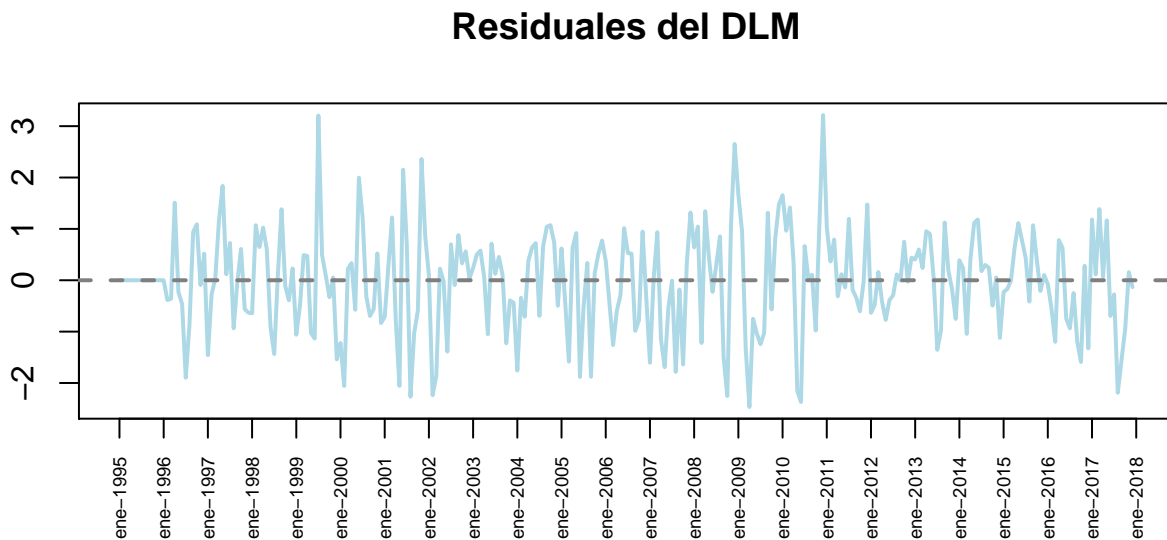


Figura 22: Residuales estandarizados del modelo DLM

Se obtienen los residuos del modelo DLM estimado y se aplican las pruebas de diagnóstico. La figura 22 muestra la serie de tiempo de los residuos; éstos fluctúan alrededor del cero y parecen tener varianza constante.

La tabla 4 muestra los valores-p obtenidos al aplicar la prueba Ljung-Box a dichos residuos. En todos los casos se rechaza la hipótesis nula de no correlación de los residuos.

Cuadro 4: Prueba Ljung-Box para los residuales.

Rezago	Valor.p
1	0.000258
2	0.000182
3	0.000168
4	0.000447
5	0.000213
6	0.000045
7	0.000079
8	0.000122
9	0.000117
10	0.000228
11	0.000226
12	0.000089
13	0.000136
14	0.000250
15	0.000425
16	0.000495
17	0.000836
18	0.001096
19	0.001517
20	0.002106

Finalmente, la figura 23 muestra el histograma de los residuales del DLM comparado con el histograma de una normal simulada para los datos, al igual que el correspondiente Q-Q Plot. Los residuales sí parecen tener una forma de normal con colas ligeramente más pesadas.

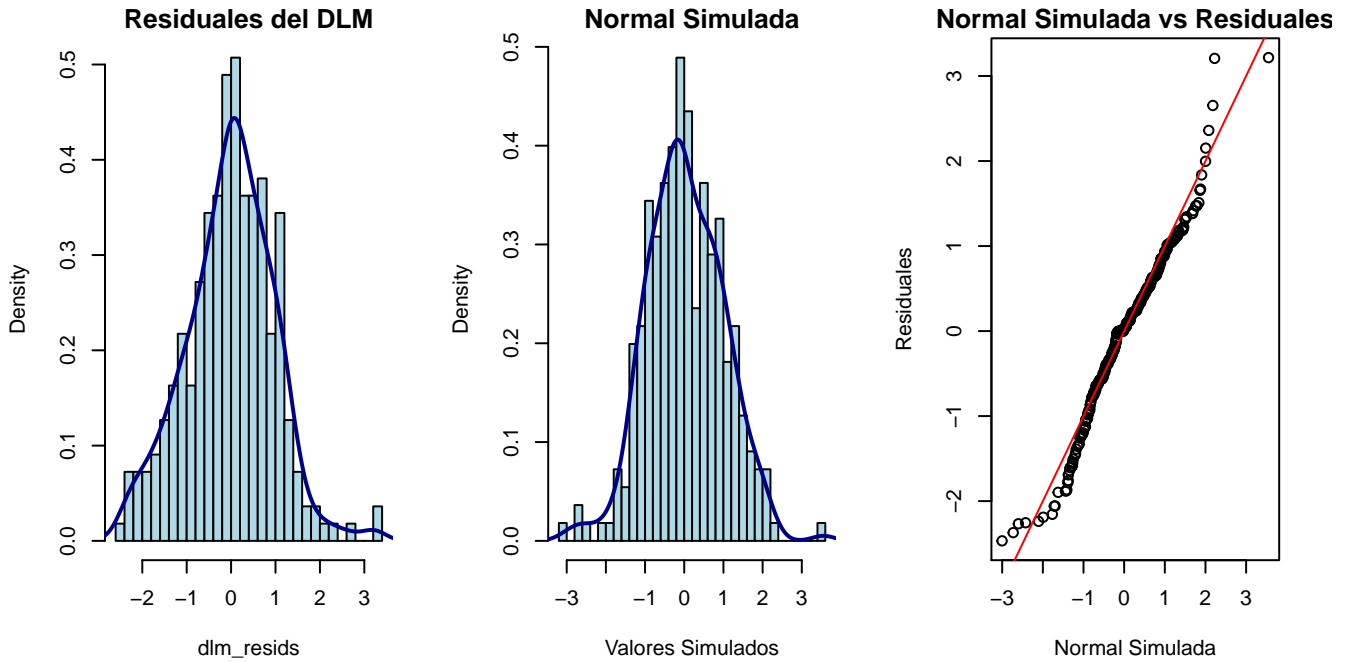


Figura 23: Histograma y Q-Q Plot de los residuales del DLM.

4.6. ECM para predicciones

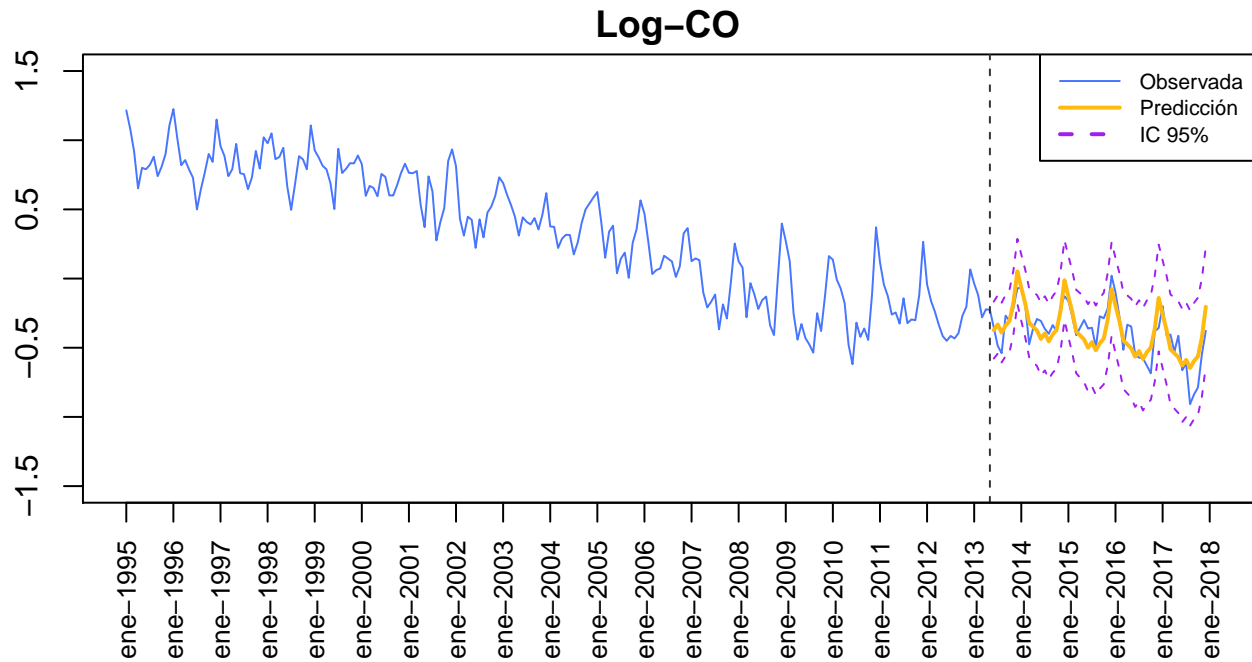


Figura 24: Serie Observada vs. Serie Pronosticada.

Para evaluar la calidad del modelo, se reserva el 20 % final de la muestra para calcular el ECM de las predicciones realizadas por el DLM. La figura 24 muestra la serie observada, la predicción realizada por el modelo, así como los intervalos de confianza al 95 % para el último 20 % de la muestra. El modelo está sobrestimando el verdadero valor; sin embargo, las observaciones quedaron dentro del intervalo de confianza. El ECM para este modelo es muy pequeño e igual a 0.0125.

5. Comparación de Modelos

La tabla 5 muestra una breve comparación entre el modelo clásico seleccionado y el DLM de la sección anterior. Se logró realizar el ajuste de ambos modelos; sin embargo el DLM tiene errores autocorrelacionados mientras que el modelo clásico sí pasa la validación de los residuales. En cuanto al ECM, el DLM es mejor haciendo predicciones que el modelo clásico, pero la diferencia no es muy grande. Por lo tanto, para ajustar y modelar el logaritmo del Monóxido de Carbono mensual promedio en la Ciudad de México se elige un modelo $SARIMA(0, 1, 2)X(1, 1, 2)_{12}$.

Cuadro 5: Comparación entre el modelo clásico y el DLM

	$SARIMA(0, 1, 2)X(1, 1, 2)_{12}$	DLM (n=2 + Est.)
Ajuste	✓	✓
No correl. residuales	✓	✗
Residuales Normales	✓	✓
ECM Predicciones	0.0147	0.0125

6. Conclusiones

En este trabajo se utilizaron modelos clásicos y Modelos Dinámicos Lineales para ajustar y estimar un modelo para el logaritmo del promedio mensual del Monóxido de Carbono (Log-CO) en la Ciudad de México. La serie Log-CO presenta tendencia decreciente y estacionalidad por lo que un modelo $SARIMA(p,d,q)(P,D,Q)_s$ o un modelo DLM polinomial de orden 2 con estacionalidad son candidatos para ajustar un modelo a la serie.

Dentro de los modelos clásicos se realizó una búsqueda en grid con código en paralelo para distintas combinaciones de los parámetros de un $SARIMA(p,d,q)(P,D,Q)_s$. En total se estimaron 1296 modelos. Se seleccionaron los 5 modelos con el menor AIC para competir. Únicamente 2 de los 5 modelos pasaron todas las pruebas de diagnóstico. Además, los 5 modelos tienen un ECM pequeño (0.0149 aprox.). De esta manera, de entre los dos modelos que sí pasaron las pruebas de diagnóstico, se seleccionó aquel con menor número de parámetros: $SARIMA(0, 1, 2)X(1, 1, 2)_{12}$ como el mejor modelo clásico.

Para el DLM se utilizó un modelo polinomial de orden 2 con estacionalidad. Las matrices de varianzas y covarianzas se estimaron vía Máxima Verosimilitud. Posteriormente se obtuvo el filtrado, suavizado y las predicciones para los siguientes 24 meses del Log-CO. Los residuales del DLM pasaron todas las pruebas de diagnósticos excepto la de autocorrelación. Este modelo tiene un ECM de 0.0125.

Entre el modelo $SARIMA(0, 1, 2)X(1, 1, 2)_{12}$ y el DLM polinomial con estacionalidad, se elige el modelo clásico para ajustar el Log-CO pues la diferencia en ECM es mínima, pero el SARIMA sí pasa todas las pruebas de diagnósticos.