



FACULTAD DE ESTUDIOS ESTADÍSTICOS

GRADO EN ESTADÍSTICA APLICADA

Curso 2021/2022

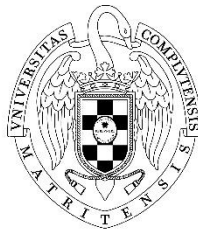
Trabajo de Fin de Grado

TÍTULO: *Evolución de las emisiones CO₂ en países de la UE*

Alumno: Carolina Osorio Iriarte

Tutor: Juana M^a Alonso Revenga

Convocatoria Ordinaria junio 2022



UNIVERSIDAD COMPLUTENSE
MADRID

Índice

1. Introducción.....	3
2. Técnicas Estadísticas	5
2.1. Cluster de Series Temporales.....	5
2.1.1. Medidas de Similitud	5
2.1.2. Cluster no jerárquico – Método de Ward	5
2.2. Modelos de Suavizado Exponencial.....	6
2.2.1. Suavizado exponencial Simple EWMA.....	7
2.2.2. Alisado Doble de Holt.....	7
2.3. Modelo ARIMA.....	7
2.3.1. Estacionariedad.....	8
2.3.2. Identificación del modelo	9
2.4. Criterios de adecuación del modelo	9
3. Análisis Descriptivo Variables	11
4. Análisis Estadístico	13
4.1. Análisis Clúster	13
4.2. Suavizado Exponencial.....	19
4.2.1. Cluster 1	19
4.2.2. Cluster 2	21
4.2.3. Cluster 3	23
4.2.4. Cluster 4	25
4.2.5. Cluster 5	27
4.2.6. Cluster 6	29
4.3. Modelo ARIMA.....	31
4.3.1. Cluster 1	31
4.3.2. Cluster 2	35
4.3.3. Cluster 3	38
4.3.4. Cluster 4	42
4.3.5. Cluster 5	46
4.3.6. Cluster 6	50
4.4. Adecuación.....	54
5. Conclusiones	55
6. Bibliografía	56
7. Anexo	57

1. Introducción

La primera revolución industrial, en la segunda mitad del siglo XVIII, supuso un punto de inflexión en el avance industrial y tecnológico. La máquina de vapor, asociada al funcionamiento de los trenes o al tiro de material pesado, marcó el inicio de esta revolución, no sólo tecnológica sino también científica con el estudio de estos sistemas y sus efectos, dando lugar a una disciplina con entidad propia: la Termodinámica. Sin embargo, aunque esta evolución ha mejorado considerablemente el bienestar humano, uno de los inconvenientes asociados al uso del carbón como materia prima, primero, y los combustibles fósiles, después, es el aumento de la contaminación ambiental por CO₂. Esta contaminación fue en aumento a medida que el desarrollo tecnológico alcanzó expansión planetaria, ya entrado el siglo XX.

En la década de los años 80 del siglo pasado, la comunidad científica puso de relieve el aumento de la temperatura global, y sus consecuencias, debido al efecto invernadero, entre cuyas causas se encuentra el aumento incontrolado de las concentraciones de CO₂ en la atmósfera. En los últimos tiempos y, a propósito de los numerosos progresos acontecidos en el mundo de la industria, nuestro planeta ha sufrido una serie de cambios que han atentado contra su preciado equilibrio natural. De ahí, que en las últimas décadas se haya intentado sensibilizar y concienciar a la población para así poder paliar el alto coste en términos ecológicos de nuestro tren de vida. No obstante, personalidades como Yuval Noah Harari¹, piensan que hemos tardado en reconocer el peligro y que, aunque se habla mucho de calentamiento global, contaminación y cambio climático, poco se está haciendo al respecto.

Teniendo en cuenta que Europa cuenta con el mercado de carbono más grande del mundo, es de particular interés el análisis de sus emisiones como referencia para evaluar el impacto en un intervalo temporal más extenso. Este trabajo surge de la necesidad de entender la verdadera situación en la que vivimos en la Unión Europea (UE) con respecto a nuestra huella de carbono, se pretende estudiar la evolución de las emisiones de CO₂ per cápita con el objetivo de poder comprender correctamente la línea temporal y su predicción futura. Poder tener una visión global de nuestro impacto a largo plazo es clave para poder ser críticos respecto a las medidas tomadas hasta el momento y la obtención de soluciones futuras.

Precisamente, no es hasta 1972 cuando se lleva a cabo la primera gran conferencia de la Organización de Naciones Unidas (ONU) sobre temas ambientales internacionales, marcando un punto de inflexión en el desarrollo de la política ambiental internacional. A esta conferencia le seguiría en 1979 la World Climate Conference, una de las primeras reuniones importantes sobre el cambio climático a la que asisten científicos de diversas disciplinas y que conduce al establecimiento del World Climate Programme. De esta forma, en 1987 se adopta el Protocolo de Montreal, un acuerdo global que busca proteger la capa de ozono de la Tierra mediante la eliminación gradual de productos químicos que puedan dañarla, entrando en vigor en 1989.

Así pues, tras numerosas reuniones e informes de evaluación, se adopta en 1997 el Protocolo de Kyoto, el primer tratado del mundo sobre las emisiones de efecto invernadero que legalmente vincula a los países desarrollados a reducir las emisiones un promedio del 5% para el periodo 2008-

¹ Historiador y escritor israelí, profesor en la Universidad Hebrea de Jerusalén

2012 y que entraría en vigor en 2005 con el EU Emissions Trading System (ETS). Seguidamente, en 2010, con la adopción de los Acuerdos de Cancún, se establece un paquete integral para ayudar a las naciones en desarrollo a enfrentar el cambio climático y el Green Climate Fund.

Por otra parte, en 2016, un año más tarde de su adopción, entra en vigor el Acuerdo de París, cuyo objetivo es mantener el aumento de las temperaturas por debajo de los dos grados centígrados por encima de los niveles preindustriales y garantizar que las emisiones globales alcancen su valor más alto lo antes posible.

Pese a todo, en 2019 el Parlamento Europeo declara una emergencia climática en Europa y en todo el mundo y presenta el Green Deal que, mediante la Ley Europea, tiene como propósito hacer que Europa sea climáticamente neutral para 2050.

Este Trabajo de Fin de Grado pretende arrojar luz sobre la evolución de la contaminación ambiental y contribuir a la mejor comprensión del efecto de las medidas tomadas hasta ahora.

Para poder llevar a cabo el estudio, se ha empleado uno de los recursos citados por el Ministerio para la Transición Ecológica y el Reto Demográfico, Our World in Data, una publicación en línea nacida en la Universidad de Oxford que busca proporcionar información y datos útiles para el mejor entendimiento de nuestro entorno. En concreto, las 27 series temporales extraídas para este trabajo han sido elaboradas con datos del Global Carbon Project y de la División de Población de las Naciones Unidas.

A raíz de los elevados niveles de gases de efecto invernadero en 2001 nace el Global Carbon Project, un proyecto que busca promover un esfuerzo coordinado de investigación en la comunidad internacional científica para destacar la importancia del ciclo del carbono y su desarrollo sostenible, centrándose uno de sus principales objetivos en la correcta estimación de las emisiones de CO₂.

Asimismo, también resulta de vital importancia para este estudio el trabajo realizado por la División de Población de las Naciones Unidas ya que nos permitirá hacer un análisis basado en el impacto por individuo. No obstante, se deberá tener en cuenta que algunos comportamientos estarán ligados a la población y no únicamente a la contaminación, siendo el aumento de las emisiones per cápita inversamente proporcional al crecimiento de la población. Recordemos también que la mayor parte de la polución corresponde a la industria y que en muchos casos el grosor de su producción es exportada por lo que no podría asociarse únicamente a las actividades del ciudadano de a pie.

2. Técnicas Estadísticas

Los conceptos planteados en el presente apartado tienen como objetivo exponer el marco teórico de las técnicas estadísticas necesarias para la posterior extrapolación de conclusiones. Todos los contrastes se llevarán a cabo con un nivel de significación del 0.05.

2.1. Cluster de Series Temporales

El Análisis Cluster de Series Temporales busca hacer agrupaciones mutuamente excluyentes de tal forma que las series de los grupos sean lo más homogéneas posibles y los grupos lo más heterogéneos entre sí.

Aunque una primera inspección gráfica puede resultar de interés, no es suficiente para poder hacer una correcta clasificación. Por ello, para poder evaluar los parecidos entre las series, los algoritmos cluster recurren a las Medidas de Distancia y Similitud, obedeciendo a los axiomas del espacio métrico:

$$\begin{aligned}d: M \times M &\rightarrow \mathbb{R} \\ a, b, c &\in M \\ d(a, a) &= 0 \\ d(a, b) = 0 &\rightarrow a = b \\ d(a, b) &\geq 0 \\ d(a, c) &\leq d(a, b) + d(b, c)\end{aligned}$$

2.1.1. Medidas de Similitud

Dadas dos secuencias numéricas ordenadas, como dos series temporales, una medida de similitud es una métrica que mide la distancia entre las secuencias input y target teniendo en cuenta el orden y encontrando secuencias similares o cercanas. De esta forma, al contar con un número considerable de series en este trabajo, permite llevar a cabo un pronóstico más sencillo de las análogas.

Para poder calcular las medidas de similitud entre varias secuencias y poder usarlas en una posterior agrupación, se construye una matriz simétrica de similitud ($K \times K$); cuyo elemento ij corresponde a la similitud entre la i -ésima y la j -ésima secuencia. Como contamos con 27 series distintas, la matriz generada presentará una dimensión 27×27 .

Por otra parte, se especificará que cada observación de cada serie temporal sea acumulada y normalizada restando la media y luego dividiendo entre su desviación estándar. A su vez, también se especificará que los datos de las series acumulados y normalizados se comparen utilizando la desviación media absoluta.

2.1.2. Cluster no jerárquico – Método de Ward

Para garantizar un número apropiado de series a modelizar, se ha decidido recurrir a un cluster no jerárquico, siendo la agrupación final de las series según un número de conglomerados fijado de antemano. En este caso, se ha optado por 6 conglomerados. No obstante también se estudiarán

algunos estadísticos que ayudan a decidir el número más adecuado de clusters para comparar dicho resultado con el número elegido a priori. Dichos estadísticos serán el R^2 , el Pseudo F y el Pseudo T^2 .

- R^2 : proporción de variación de la variable dependiente, en este caso de las series temporales, que queda explicada por dicho número de agrupaciones. En concreto, se buscará un valor mayor al 70%.
- *Pseudo F*: estadístico que describe la relación entre la varianza entre clusters y la varianza dentro de los clusters. Es decir, compara la dispersión tanto entre como dentro de los conglomerados, se pretende que el cociente sea máximo y buscaremos tanto máximos relativos como incrementos importantes en su valor.
- *Pseudo T^2* : estadístico que cuantifica la diferencia entre dos clusters que son agrupados en un paso en concreto. Por tanto, si este tiene un salto en el paso k, se seleccionará como número óptimo k+1.

En concreto, usaremos el método de Ward, el cual, en cada etapa, une los dos clusters con menor incremento en el valor total de la suma de los cuadrados de las diferencias. El proceso comienza con una suma de cuadrados de los errores para todos los clusters (E) nula o, lo que es lo mismo, un cluster para cada serie. Busca encontrar en cada etapa la unión de clusters que produzca menor incremento en E, siendo este proporcional a la distancia euclídea al cuadrado de los centroides de los clusters unidos.

2.2. Modelos de Suavizado Exponencial

Los modelos de suavizado son una herramienta útil a la hora de determinar la tendencia de una serie temporal y_t , definen una nueva serie z_t mediante una expresión lineal que suaviza los efectos ajenos a esta mediante un filtro lineal F.

$$y_t \rightarrow F \rightarrow z_t$$

En particular, los modelos de alisado exponencial presentan la siguiente forma:

$$y_t = \mu_t^2 + \beta_t^3 t + S_p(t)^4 + \varepsilon_t^5$$

Debido a que algunos parámetros pueden ser 0, existen diferentes modelos de Suavizado Exponencial que derivan de la fórmula anterior. Puesto que en la mayoría de series de este estudio se aprecian variaciones en la tendencia y una falta de estacionalidad, se empleará el modelo suavizado exponencial Doble de Holt, quedando descartado el modelo exponencial Holt-Winters, extensión del método Holt que considera una componente estacional. Por otra parte, en aquellos casos en los que la tendencia no cambie su pendiente a lo largo del tiempo, se recurrirá al suavizado exponencial Simple EWMA.

Para asegurarnos de que el modelo es adecuado se estudiará la significación de los parámetros mediante un estadístico t.

² nivel de la serie en el instante t

³ pendiente dependiente del tiempo

⁴ contribución estacional de p al modelo

⁵ residual o innovación

2.2.1. Suavizado exponencial Simple EWMA

Este modelo consta únicamente de un parámetro correspondiente a la componente de nivel o, lo que es lo mismo, al valor que tomaría la serie si no fuera influenciada por perturbaciones, tendencia y estacionalidad. De esta forma se asume que $S_p(t)$ y la pendiente β_t son nulos, quedando el modelo de la siguiente forma:

$$y_t = \mu_t + \varepsilon_t$$

En concreto, el modelo es una media móvil ponderada exponencialmente con $0 < \theta < 1$.

$$\hat{y}_t = \hat{y}_{t-1} + (1 - \theta)(y_{t-1} - \hat{y}_{t-1})$$

El parámetro θ óptimo será aquel que minimice la suma de cuadrados de las innovaciones de forma eurística por lo que para encontrarlo se calculará la suma de cuadrados de los errores para diferentes valores, resultando en un proceso que será más o menos largo en función del grado de precisión que se desee.

En cuanto a los valores que el parámetro puede tomar, si tiende a 0 podremos decir que no hay suavizado porque este será el propio valor y_t mientras que, si este tiende a 1, el valor tiende a ser constante, la media.

2.2.2. Alisado Doble de Holt

Al presentar las series de interés cambios en su pendiente a lo largo del tiempo y no haber estacionalidad, recurriremos al alisado Doble de Holt, proporcionando unos primeros pronósticos de corto alcance.

$$y_t = \mu_t + \beta_t t + \varepsilon_t$$

$$\hat{y}_t = \hat{\mu}_{t-1} + \hat{\beta}_{t-1}$$

El modelo consta de dos parámetros, uno de nivel y otro de tendencia, ambos dependientes del tiempo por lo que, en vez de ser fijos, evolucionan. Por tanto, el algoritmo de este método considerará dos ecuaciones de actualización para encontrar los valores óptimos de ambos.

2.3. Modelo ARIMA

Para que una serie pueda ser modelizada con un modelo ARMA, esta tiene que cumplir la condición de estacionariedad. Sin embargo, en muchas ocasiones las series temporales no son estacionarias, precisando ciertas transformaciones. Dichas series son diferenciadas d vez para hacerlas estacionarias y poder aplicar un modelo ARMA(p, q), dando lugar a un proceso integrado ARIMA(p, d, q) de la serie original.

El modelo mixto ARMA(p, q) consta de un polinomio autorregresivo que viene determinado por los retardos hasta el orden p y otro de medias móviles determinado por los q errores anteriores. Al no presentar nuestras series un comportamiento estacional, prescindiremos del modelo ARIMA(p, d, q)(P, D, Q) estacional.

La parte autorregresiva del modelo, AR(p), corresponde a una generalización de la regresión que representa la dependencia lineal entre la variable y_t y los retardos de tal forma que el valor de la serie en un punto depende de los p anteriores.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

En cuanto a las medias móviles, MA(q), el modelo es el siguiente:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Siendo, en ambos casos, ε_t ruido blanco. Es decir, valores completamente aleatorios de la distribución Normal.

$$\varepsilon_t \sim N(0, \sigma^2_\varepsilon)$$

Asimismo, las observaciones del ruido blanco están incorreladas y su esperanza es 0.

2.3.1. Estacionariedad

Para que una serie sea estacionaria existen dos condiciones:

- Condición de estacionariedad fuerte: la distribución de la serie en el instante t tiene que ser igual que en el instante t+k.
- Condición de estacionariedad débil: la media y la autocovarianza tienen que ser la misma para t que para t+k.

En general, por ser la primera demasiado restrictiva, se suele usar la segunda.

Por otro lado, se sabe que las series que presentan tendencias o variaciones estacionales no pueden considerarse estacionarias por lo que para inducir estacionariedad en media se recurrirá a diferenciaciones y, para estabilizar la varianza, a transformaciones de Box-Cox.

Para cerciorarnos de que tras las transformaciones se cumple la condición de estacionariedad y se puede llevar a cabo la modelización ARIMA, esta se contrastará con la prueba de Dickey-Fuller aumentada y el contraste de Kwiatkowski, Phillips, Schmidt y Shin (KPSS).

- El test de Dickey-Fuller, basado en un modelo de regresión, contrasta la hipótesis nula de que $\alpha=1$ frente a la hipótesis alternativa de que $\alpha<1$

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \varepsilon_t$$

De esta forma, cuando $\alpha=1$ el proceso no será estacionario. Además, se considerarán tres tipos de contrastes en función de si el modelo de regresión es de media 0, constante o de tendencia.

- El test de KPSS asume que la serie es expresada como la suma de una tendencia determinística, un camino aleatorio y un error estacionario. Puesto que su hipótesis nula

corresponde a que no hay camino aleatorio, esta también corresponderá a que el proceso sí es estacionario.

No obstante, estos contrastes pueden ser demasiado estrictos y pueden llevar a una sobrediferenciación, dando lugar a peores modelos. Por ello, como se verá más adelante, aunque alguno de los test indique que la serie no es estacionaria, se procederá con una única diferenciación de orden 1 que bastará para eliminar la tendencia y estabilizar la serie en media. En el caso de que los residuos de todos los modelos candidatos estuvieran correlados, esto nos llevaría a pensar que se debería hacer alguna otra transformación.

2.3.2. Identificación del modelo

Se pretende comprobar la estructura estacionaria del modelo ARIMA a través de un modelo general ARMA. Para ello, se decidirá el orden máximo de las partes AR y MA fijándonos en los coeficientes de autocorrelación de la serie transformada. Concretamente, en las autocorrelaciones simples significativas para el caso del MA máximo y en las inversas y parciales para AR. Siempre se comenzará con modelos simples AR o MA.

También se recurrirá a métodos automáticos para identificar algunos candidatos, en concreto, al método SCAN, ESACF y MINIC.

- El método *SCAN* es el método de menor correlación canónica, este analiza los autovalores de la matriz de correlación del proceso ARMA.
- El método *ESACF* es la extensión de la función muestral de autocorrelaciones, basado en estimadores de mínimos cuadrados iterados de los parámetros autorregresivos.
- El método *MINIC* computa los criterios de información para varios órdenes tanto de parámetros autorregresivos como de medias móviles.

Por otra parte, se tendrá que comprobar la adecuación del modelo por lo que se deberá verificar que los parámetros son significativos y que los residuos están incorrelados y son ruido blanco.

Para la significación de los parámetros se recurrirá al estadístico t cuya hipótesis nula asociada será que el parámetro no es significativo. En cuanto a la correlación de los residuos, se usará el test de Ljun-Box, cuya hipótesis nula es la incorrelación de los residuos.

2.4. Criterios de adecuación del modelo

Para poder comparar la adecuación de los modelos resultantes tanto del suavizado como del ARIMA, se emplearán ciertos criterios de adecuación.

- El R^2 , ya descrito anteriormente, corresponde a la proporción de variabilidad explicada por el modelo de la serie original.
- El criterio *Akaike* (AIC) y el bayesiano de *Schwarz* (SBC) son estimadores del error de predicción que además de tener en cuenta el desajuste, al penalizar el incremento de

parámetros, también penalizan el sobreajuste. En cuanto a diferencias entre ambos, el SBC penaliza más la complejidad que el AIC.

Por tanto, los mejores modelos presentarán los R^2 más próximos a 1 y los AIC y SBC más bajos.

3. Análisis Descriptivo Variables

En este apartado se pretende proporcionar una primera información sobre las series a estudio para identificar algunos comportamientos y posibles similitudes entre ellas de 1950 a 2020. En concreto, las series presentan valores anuales por lo que no podremos apreciar ningún comportamiento estacional.

Aunque las observaciones pueden ser de forma continua, en nuestro caso se ha extraído una muestra discreta en momentos de igual distancia. Sin embargo, se ha decidido unir los puntos en la gráfica siguiente, Fig. 1, porque suponemos que los puntos están correlados en el tiempo (y_t depende de y_{t-1}, y_{t-2}, \dots).

Antes de empezar con el análisis de las series se debe comprobar que los momentos distan entre sí el mismo intervalo de tiempo y que están bajo las mismas condiciones. En nuestro caso se cumplen ambas circunstancias por lo que no hay que corregir ninguna desviación.

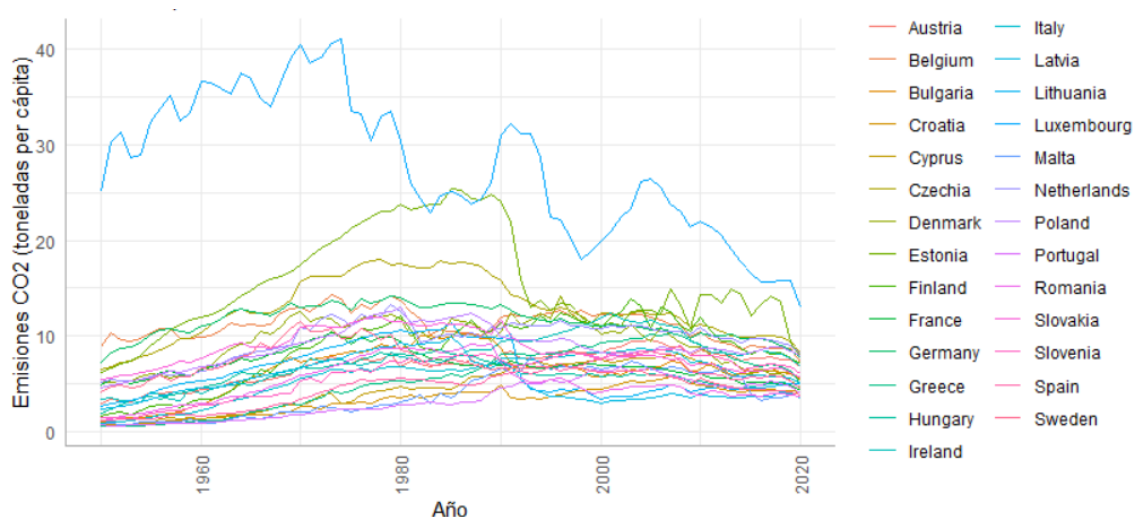


Figura 1

Atendiendo al gráfico anterior Fig. 1, se aprecian importantes disconformidades entre unos países y otros, destacando Luxemburgo y Estonia con unos niveles de CO_2 per cápita muy superiores a los de sus homólogos europeos. No obstante, Luxemburgo presenta unos valores notablemente más inusuales que pueden estar asociados a una elevada renta per cápita y un lujoso estilo de vida que pudieran resultar en un mayor impacto medio ambiental. Por otra parte, también hay que tener en cuenta que se trata de un país con pocos habitantes y que el impacto de su industria podría ser mayor. Para poder apreciar correctamente el comportamiento del resto de países, excluirémos los dos anteriores de la siguiente representación Fig. 2.

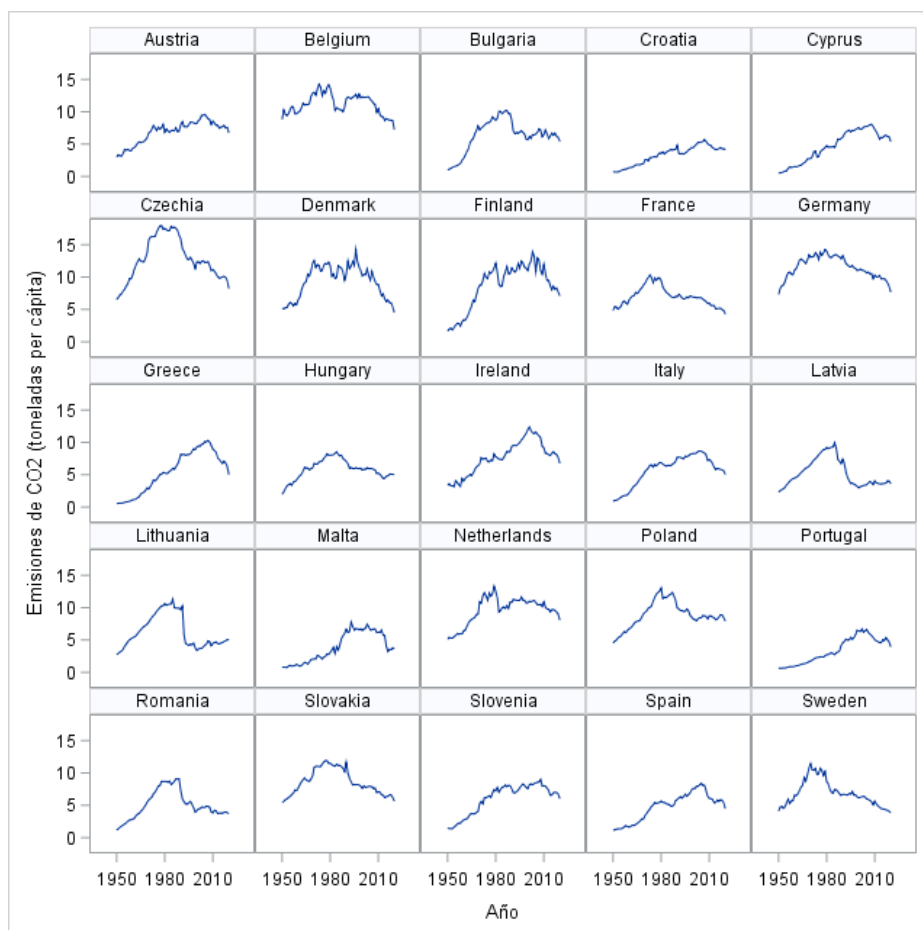


Figura 2

A simple vista se puede hacer una primera división entre aquellos países que alcanzan sus máximos entorno a la década de los 80 y los que los alcanzan a principios del siglo XX. En el primer grupo podríamos incluir países como Bulgaria, República Checa y Francia mientras que en el segundo estarían algunos como España, Portugal e Irlanda.

En cuanto a los países más semejantes en comportamiento, encontramos varios subgrupos como pueden ser Letonia y Lituania, Dinamarca y Finlandia o España y Portugal. Por otra parte, también llama la atención la relativa similitud en forma entre Bélgica y Luxemburgo.

A raíz de las políticas climáticas de la UE centradas en la reducción de las emisiones de gases de efecto invernadero, se puede apreciar en los 27 un cierto descenso en las toneladas per cápita de CO₂ en la última década y media, coincidiendo su inicio con la implementación del Protocolo de Kyoto y el ETS. Este último se centra en la reducción de las emisiones de carbono de la industria forzando a las empresas más contaminantes a comprar prestaciones por cada tonelada de CO₂ emitido por lo que no se puede achacar la bajada a un esfuerzo colectivo de la ciudadanía sino más bien a una aparente efectividad de las medidas tomadas en dicho sector.

4. Análisis Estadístico

Una vez estudiadas las series gráficamente se procederá a realizar su análisis estadístico, empezando con un análisis clúster que nos permitirá obtener predicciones conjuntas con un menor número de modelizaciones.

4.1. Análisis Clúster

Como comentábamos en el tercer apartado, se ha decidido realizar un análisis cluster no jerárquico con 6 conglomerados mediante el método de Ward, presentando en la Tabla 1 los pasos seguidos y sus correspondientes estadísticos.

Cluster History									
Number of Clusters	Clusters Joined		Freq	New Cluster RMS Std Dev	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
26	Cyprus	Greece	2	0.0274	0.0019	.998	21.3	.	
25	Ireland	Spain	2	0.0356	0.0031	.995	16.5	.	
24	France	Sweden	2	0.0360	0.0032	.992	15.7	.	
23	Latvia	Lithuania	2	0.0371	0.0034	.988	15.4	.	
22	CL26	Portugal	3	0.0330	0.0035	.985	15.4	1.9	
21	Italy	Slovenia	2	0.0386	0.0037	.981	15.6	.	
20	Poland	Romania	2	0.0409	0.0042	.977	15.6	.	
19	Czechia	Slovakia	2	0.0427	0.0045	.972	15.7	.	
18	Austria	Finland	2	0.0434	0.0047	.968	15.9	.	
17	CL18	CL21	4	0.0433	0.0056	.962	15.9	1.3	
16	CL22	CL25	5	0.0385	0.0062	.956	15.9	2.2	
15	Bulgaria	Hungary	2	0.0512	0.0065	.949	16.1	.	
14	Croatia	CL16	6	0.0425	0.0077	.942	16.2	2.1	
13	CL19	Germany	3	0.0511	0.0084	.933	16.3	1.9	
12	Estonia	CL20	3	0.0520	0.0093	.924	16.6	2.2	
11	CL14	Malta	7	0.0481	0.0121	.912	16.6	2.7	
10	CL15	CL12	5	0.0586	0.0141	.898	16.6	2.1	
9	CL13	CL24	5	0.0568	0.0159	.882	16.8	2.9	
8	CL17	Netherlands	5	0.0560	0.0172	.865	17.4	3.7	
7	Belgium	Luxembourg	2	0.0934	0.0217	.843	17.9	.	
6	CL9	Denmark	6	0.0661	0.0222	.821	19.3	2.8	
5	CL8	CL11	12	0.0621	0.0396	.781	19.7	6.0	
4	CL10	CL23	7	0.0742	0.0446	.737	21.5	6.0	
3	CL4	CL6	13	0.0822	0.0649	.672	24.6	5.2	
2	CL7	CL3	15	0.1037	0.1514	.520	27.1	8.8	
1	CL5	CL2	27	0.1244	0.5204	.000	.	27.1	

Tabla 1

Comprobamos que para los 6 clusters elegidos, no se aprecia ningún salto en el R^2 semiparcial y que su proporción de variabilidad, al superar en 12.1 puntos porcentuales el 70%, se puede considerar como bastante buena. Por otra parte, si nos fijamos en el Pseudo T^2 , Fig. 3, encontramos un máximo relativo en 5, lo que también sugiere que 6 es un número adecuado de clusters. Sin embargo, al no

encontrar un máximo relativo para el Pseudo F en 6, este podría indicar como óptimo un número mayor, aunque en este caso no parece razonable.

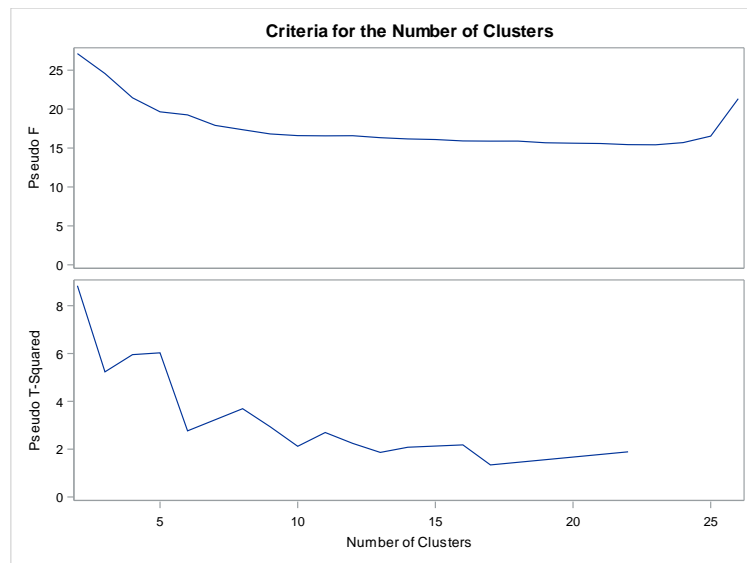


Figura 3

Además de la Fig. 3, que nos permite estudiar el Pseudo F y el Pseudo T^2 en función del número de clusters, también resulta de gran utilidad la incorporación de dendogramas para poder visualizar más fácilmente la construcción de los conglomerados. En el dendograma siguiente Fig. 4, se puede observar un claro resumen del orden de las agrupaciones en función del R^2 , correspondiendo los recuadros a los 6 clusters formados. De esta forma, mientras que España y Portugal, Bélgica y Luxemburgo y, Lituania y Letonia han sido incluidos en los mismos clusters, Dinamarca y Finlandia se han repartido entre el primer y el sexto conglomerado. Es decir, la agrupación resultante, como era de esperar, presenta resultados distintos a la inspección gráfica previa. No obstante, a grosso modo, todos los clusters concuerdan con lo observado en la primera exploración.

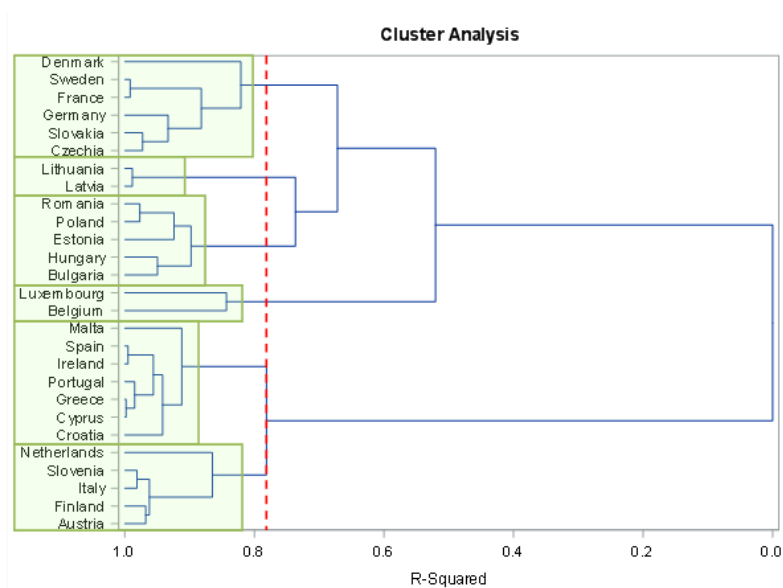


Figura 4

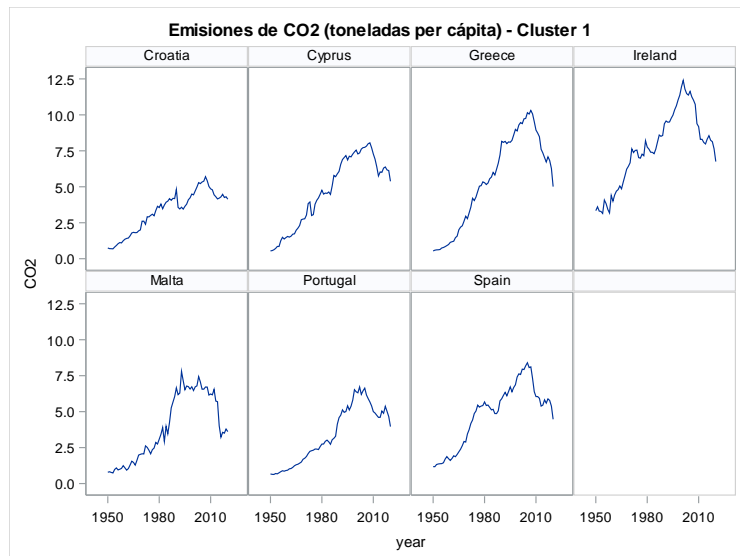


Figura 5

La Fig. 5 corresponde a los siete países que conforman el primer cluster; todos ellos presentan un crecimiento con una pendiente más o menos constante hasta alcanzar un pico a principios de los 2000 acorde con la implementación del Protocolo Kyoto y la ETS. No obstante, el pequeño repunte que se produce en torno a la mitad de la primera década del siglo XXI puede deberse a un efecto tardío de la crisis de 2008 puesto que esta llevó a una bajada del precio de las prestaciones y una subida de las emisiones, razón por la que la EU decidió buscar otras medidas, como el Acuerdo de París, que tuvieran más impacto a largo plazo y que fueran independientes del mercado. Por otra parte, y como era de esperar tras observar la Fig. 4 y la Tabla 1, Malta presenta el patrón que más dista del resto de países incluidos.

A su vez, se aprecia una disminución en las emisiones de Croacia entre 1991 y 1995, coincidiendo con la Guerra Croata de Independencia. No obstante, aunque se observa también un pequeño valle en la misma época en otros países de este cluster, no se ha encontrado ninguna razón de peso política que lo pudiera haber causado.

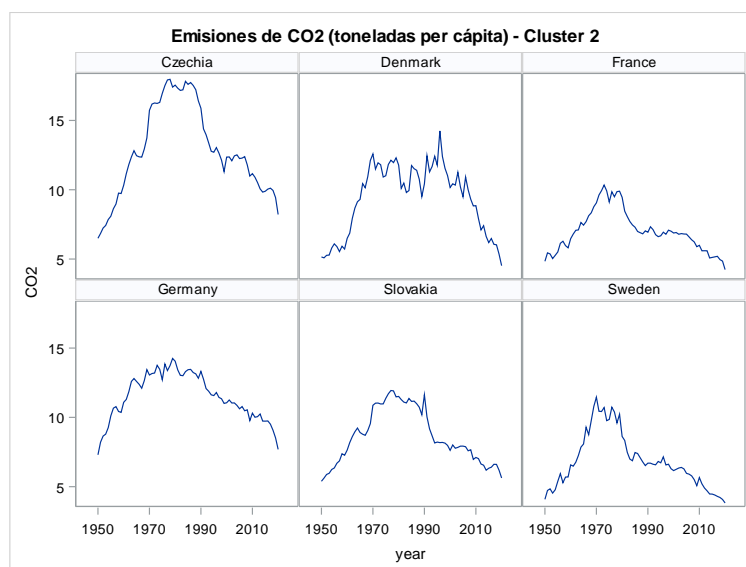


Figura 6

El segundo cluster, Fig. 6, incluye seis países que alcanzan máximos entorno a los años 70, algunos incluso manteniendo esos niveles durante décadas, para a continuación empezar un primer descenso en picado y un segundo considerablemente más progresivo. La primera bajada significativa varía considerablemente entre unos países y otros, en algunos casos se produce a principios de los 80 y en otros entrados los 90, únicamente en estos segundos podría haber tenido algo que ver el Protocolo de Montreal.

En concreto, la bajada sucedida en República Checa y Eslovaquia responde a la disolución de Checoslovaquia en 1993, también conocida como el Divorcio de Terciopelo y mucho más pacífica que la disolución de la Unión Soviética (URSS). Al mismo tiempo, destaca el gran parecido entre Francia y Suecia, cuyos picos y valles coinciden en bastantes ocasiones.

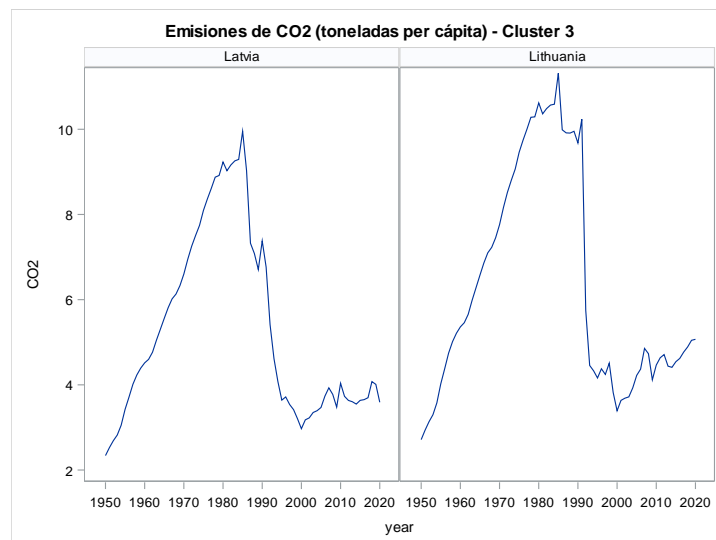


Figura 7

El tercer cluster, Fig. 7, compuesto por Letonia y Lituania, corresponde al conglomerado más homogéneo. Las emisiones de ambos países ascienden rápidamente desde 1950 hasta bien entrados los 80, presentando a continuación un interesante efecto boomerang hasta el 2000, cuando, a diferencia de la mayoría de sus socios europeos, se empieza a observar un cierto impulso. Este comportamiento tan parecido refleja el pasado soviético de ambos países, incluso en años posteriores a su independencia podemos observar una importante similitud entre ambos. Por otra parte, resalta un pequeño repunte transitorio en ambos país entorno a la disolución de la URSS en 1991.

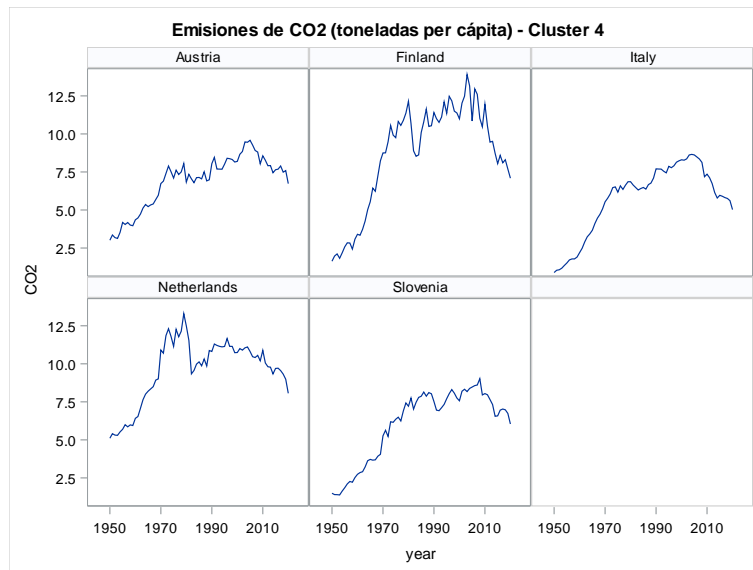


Figura 8

Los cinco países incluidos en el cuarto cluster, Fig. 8, presentan en las primeras cinco décadas una pendiente positiva no muy acusada que parece dar paso a una cierta recesión que podría estar vinculada tanto al lanzamiento de la ETS como al Protocolo de Kyoto. No obstante, de existir dicha relación, esta sería de las más débiles de la Unión. Por otra parte, también se aprecian ciertas diferencias en la volatilidad de las emisiones entre unos países y otros, siendo el más volátil Finlandia y el menos Italia.

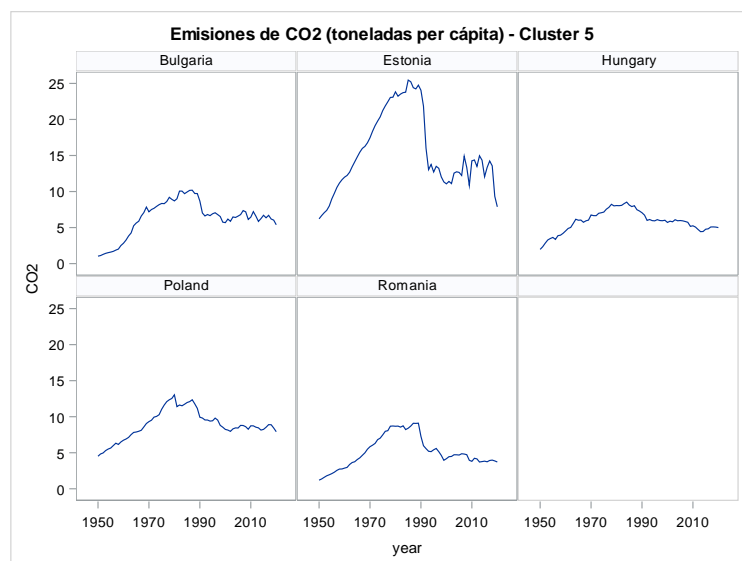


Figura 9

En las representaciones de la Fig. 9 se pueden observar los países que forman parte del quinto cluster, identificando un comportamiento relativamente parecido al de las series del segundo conglomerado con la diferencia de que sus cambios estructurales suceden entre un lustro y una década más tarde. De esta forma, sí que podría estar explicado el descenso que sucede en estos países a finales de los 80 por la adopción del Protocolo de Montreal.

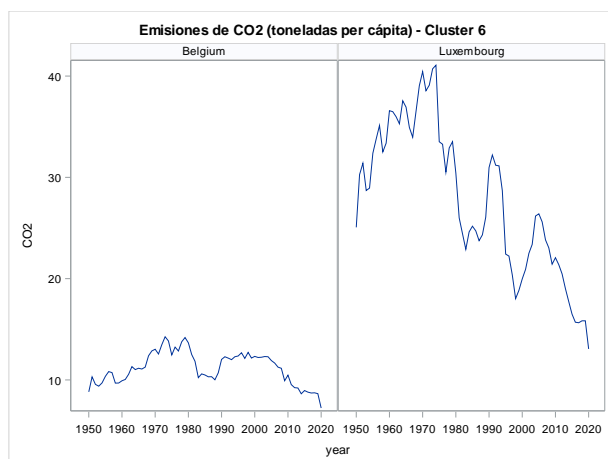


Figura 10

El último cluster cuenta con Bélgica y con Luxemburgo, dos países que, aunque distan mucho en su tendencia central, presentan un comportamiento bastante similar que podría estar explicado por la formación de la Benelux Economic Union en 1958, una unión que busca promover la libre circulación de trabajadores, bienes y servicios entre Países Bajos, Bélgica y Luxemburgo. Sin embargo, mientras que en este último se aprecia una cierta bajada a finales de los 70, Bélgica parece fluctuar entorno a los mismos valores.

Tras haber realizado el Análisis Cluster, se ha decidido resumir la información de cada agrupación con una serie por conglomerado resultado de la media aritmética anual de los integrantes de cada uno de ellos, Fig. 11. De esta forma, en los siguientes apartados procederemos con las modelizaciones de dichas series; las predicciones y las conclusiones serán sobre estas. Puesto que no se ha podido hacer una media ponderada que tuviera en cuenta la importancia e impacto relativo de cada país, se estudiará únicamente el comportamiento de las predicciones en función de su pendiente ya que los valores reales podrían variar considerablemente entre unos países y otros, como es el caso de Bélgica y Luxemburgo en el último cluster.

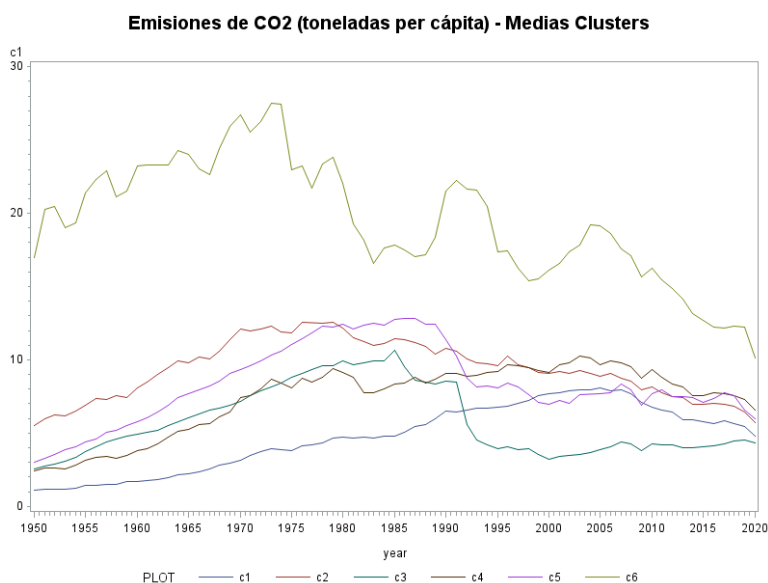


Figura 11

4.2. Suavizado Exponencial

Como mencionábamos en el apartado de Técnicas Estadísticas, al no presentar las series estacionalidad, no se considerará el alisado de Holt-Winters. Se procederá primero con el estudio de los parámetros del alisado Doble de Holt, ajustándose un alisado Simple en el caso en el que la componente tendencia no sea significativa.

4.2.1. Cluster 1

Empezamos ajustando el suavizado doble para el primer cluster y estudiando sus parámetros, Tabla 2.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.09198	10.86	<.0001
Trend Weight	0.33936	0.09349	3.63	0.0005

Tabla 2

El estadístico t asociado al contraste de la componente de nivel es 10.86 y su p-valor es prácticamente 0 por lo que rechazamos la hipótesis nula, el parámetro es significativo. Por otra parte, la componente de tendencia también es significativa puesto que presenta un valor del estadístico t igual a 3.63 y un p-valor <0.05. De esta forma, la estimación de la primera componente es aproximadamente 1 y la de la segunda 0.33936.

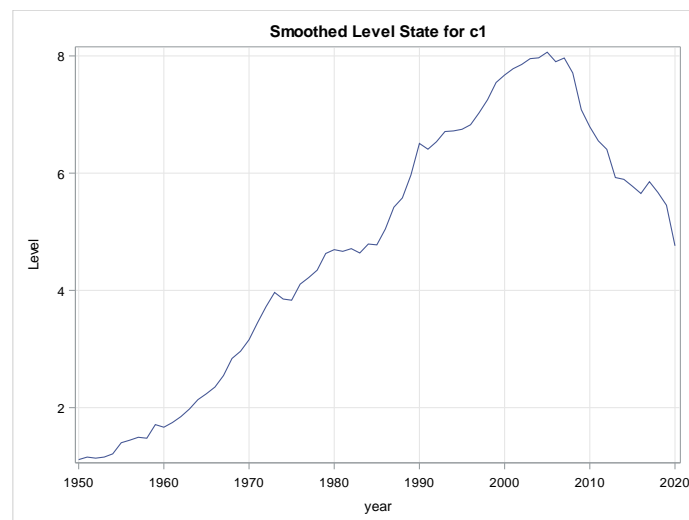


Figura 12

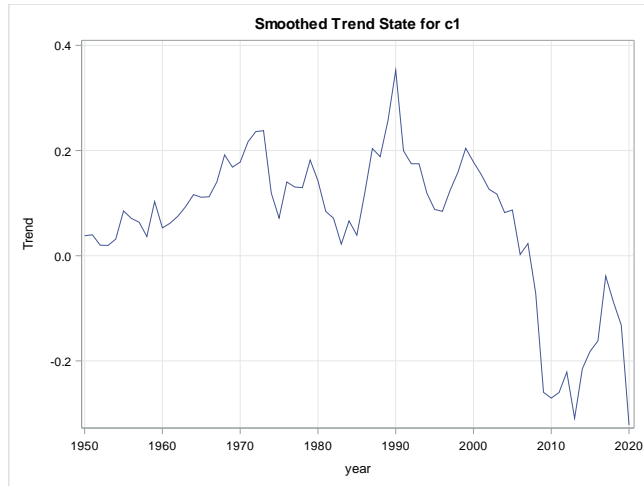


Figura 13

La evolución de de los coeficientes ha sido representada en los dos gráficos anteriores, Fig. 12 y Fig. 13. En concreto, los niveles de comienzo para el parámetro de nivel son próximos a 1 y los de final a 5 mientras que en el caso de la tendencia todos rondan el 0, presentando valores negativos entrados los 2000.

Forecast Summary												
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
c1	4.43	4.11	3.79	3.47	3.15	2.82	2.50	2.18	1.86	1.54	1.21	0.89

Tabla 3

Se obtienen 12 predicciones con este modelo, Tabla 3, y se representan gráficamente junto con sus correspondientes intervalos de confianza en Fig. 14. A simple vista se aprecia un buen ajuste a los datos originales junto con unas predicciones que prevén que la bajada de emisiones sucedidas en los últimos años se prolongue 12 años más. Atendiendo a los intervalos de confianza, podemos decir que en el peor de los casos las emisiones se mantendrán mientras que en el mejor de ellos serán nulas.

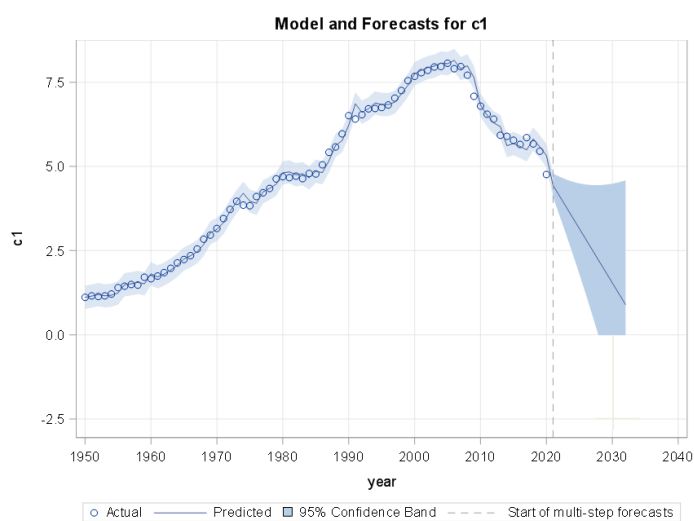


Figura 14

Con respecto a la evaluación del modelo, en la Tabla 4 se muestran algunos de los estadísticos de ajuste. Tanto el AIC como el SBC destacan positivamente por presentar valores muy bajos además de una proporción de variabilidad explicada del 99.39%. Podemos afirmar que se trata de un buen ajuste.

Statistics of Fit for Variable c1	
Statistic	Value
R-Square	0.99390804
Akaike Information Criterion	-245.15907
Schwarz Bayesian Information Criterion	-240.63371

Tabla 4

4.2.2. Cluster 2

En la tabla siguiente, Tabla 5, se presentan los parámetros estimados del alisado doble ajustado al conglomerado 2 y sus correspondientes contrastes.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08670	11.52	<.0001
Trend Weight	0.13271	0.04974	2.67	0.0095

Tabla 5

Los p-valores relativos a los dos parámetros del modelo son inferiores al nivel de significación fijado por lo que podemos afirmar que ambos son significativos, en concreto, la estimación del nivel es 0.999 y de la tendencia 0.132. En los siguientes gráficos, Fig. 15 y Fig. 16, se puede ver reflejada la evolución de las dos componentes. Mientras que la tendencia es claramente decreciente y volátil, los coeficientes de nivel presentan un comportamiento similar al de una función cuadrática.

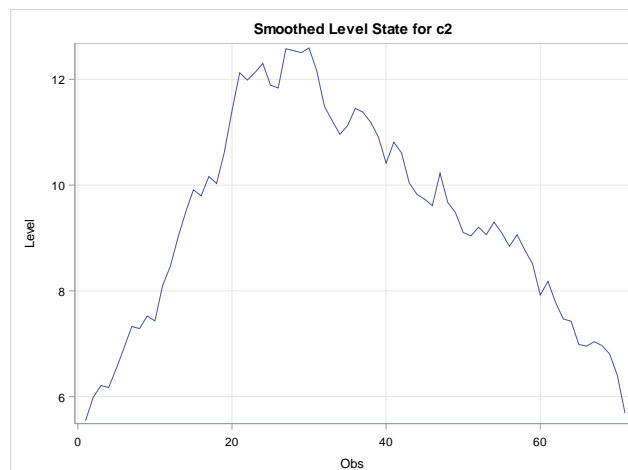


Figura 15

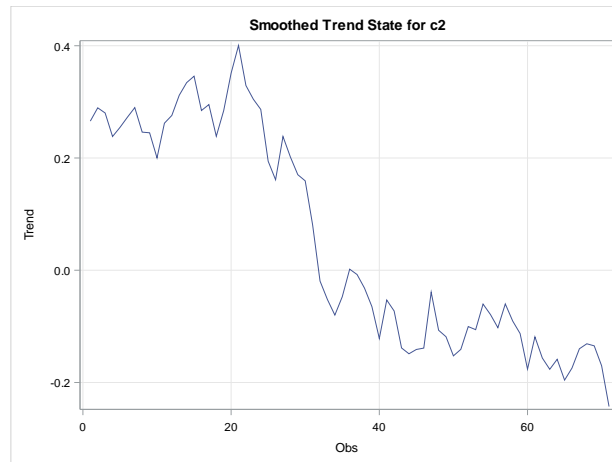


Figura 16

Forecast Summary												
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
c2	5.44	5.20	4.95	4.71	4.47	4.23	3.98	3.74	3.50	3.25	3.01	2.77

Tabla 6

Las 12 predicciones obtenidas mediante este modelo, Tabla 6, junto con su representación, Fig. 17, apuntan a una bajada que llevaría a valores no vistos desde antes de la década de los 50. No obstante, también se contempla la posibilidad de que estos suban como mucho hasta cerca de 7 toneladas per cápita, lo que correspondería a retomar niveles de 1955 en el peor de los casos.

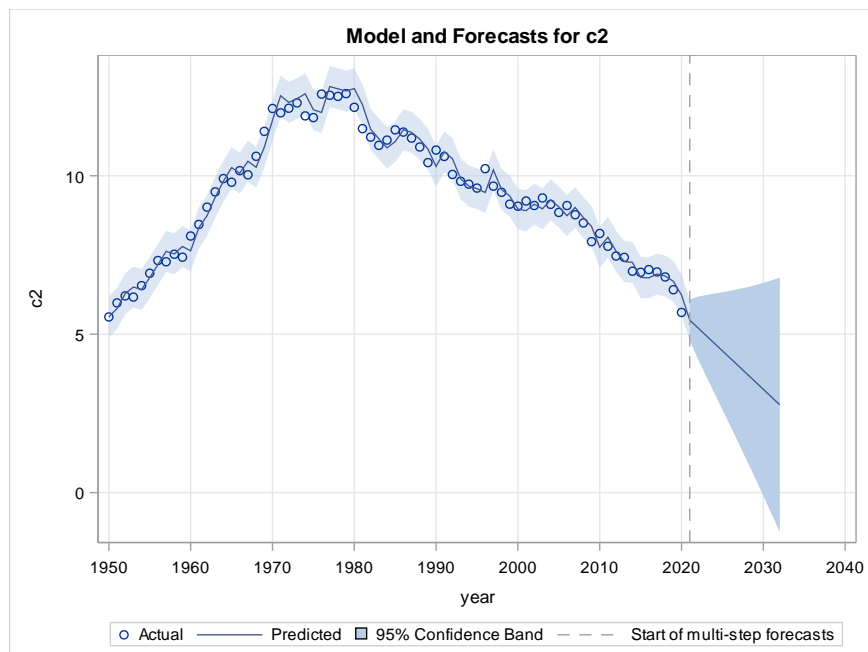


Figura 17

Los estadísticos presentados en la Tabla 7 indican un buen ajuste con una alta proporción de variabilidad explicada del 97.29% y unos valores de AIC y SBC considerablemente favorables.

Statistics of Fit for Variable c2	
Statistic	Value
R-Square	0.97297289
Akaike Information Criterion	-155.41465
Schwarz Bayesian Information Criterion	-150.88929

Tabla 7

4.2.3. Cluster 3

Los p-valores asociados a los parámetros de nivel y de tendencia, Tabla 8, indican que ambos son significativos con un coeficiente estimado para el primero de 0.999 y de 0.199 para el segundo, su desarrollo a lo largo del tiempo se puede observar en las Fig. 18 y Fig. 19. Podemos destacar el gran parecido entre el comportamiento de la componente nivel y la serie original.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08540	11.70	<.0001
Trend Weight	0.19909	0.06125	3.25	0.0018

Tabla 8

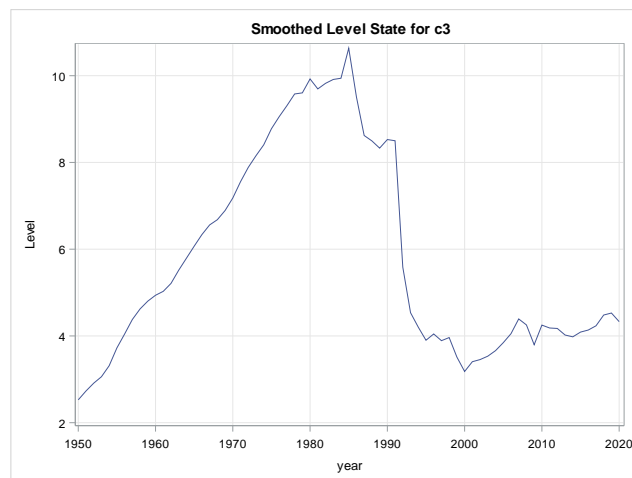


Figura 18

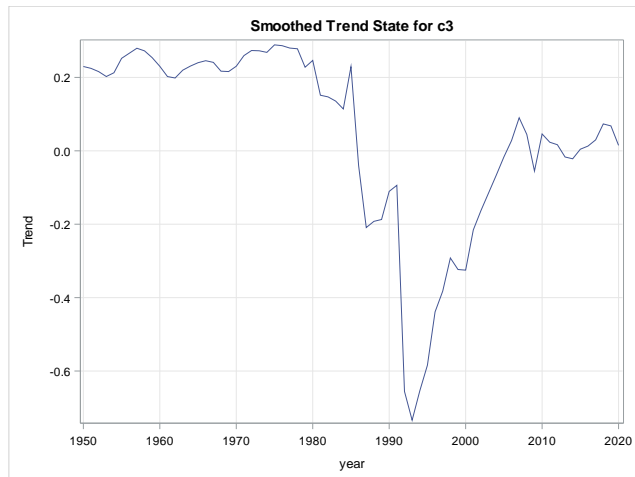


Figura 19

Las predicciones de este modelo, Tabla 9 y Fig. 20, indican un pequeño incremento en las emisiones per cápita de Letonia y Lituania casi inapreciable. Por otra parte, se observa una banda de confianza muy ancha por lo que el suavizado contempla tanto que vuelva alcanzar su máximo como que sea 0, claramente su ajuste es menos preciso que los modelos estudiados hasta ahora.

Forecast Summary												
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
c3	4.34	4.36	4.37	4.38	4.40	4.41	4.43	4.44	4.46	4.47	4.49	4.50

Tabla 9

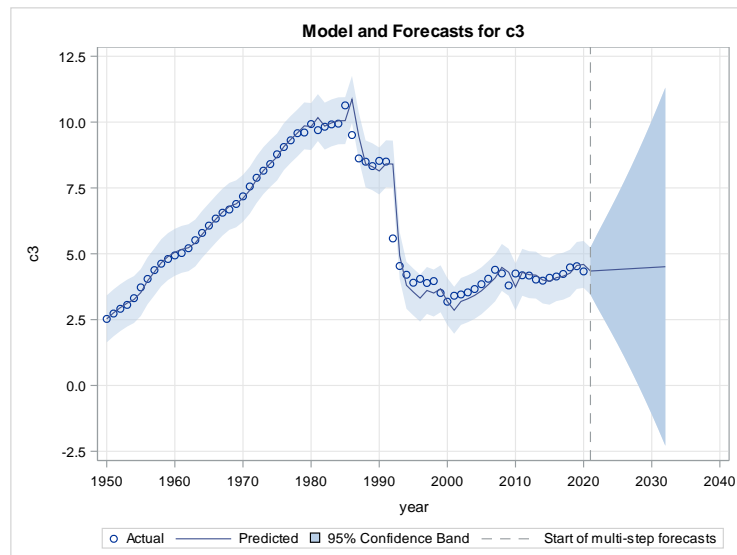


Figura 20

Los estadísticos de ajuste, Tabla 10, apuntan a un buen modelo con una proporción de variabilidad explicada del tercer cluster del 96.34% y unos valores para el AIC y el SBC razonablemente buenos.

Statistics of Fit for Variable c3	
Statistic	Value
R-Square	0.96349989
Akaike Information Criterion	-109.96851
Schwarz Bayesian Information Criterion	-105.44315

Tabla 10

4.2.4. Cluster 4

Observamos, Tabla 11, un p-valor para el parámetro de tendencia más elevado que los de modelos anteriores, sin embargo, este no alcanza el nivel de significación y, al igual que para el de nivel, rechazamos la hipótesis nula de no significación. Por tanto, las estimaciones de ambos aportan información significativa al modelo. La evolución de sus correspondientes coeficientes viene reflejada en Fig. 21 y Fig. 22 donde se aprecia una clara tendencia decreciente.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08759	11.41	<.0001
Trend Weight	0.10320	0.04628	2.23	0.0290

Tabla 11

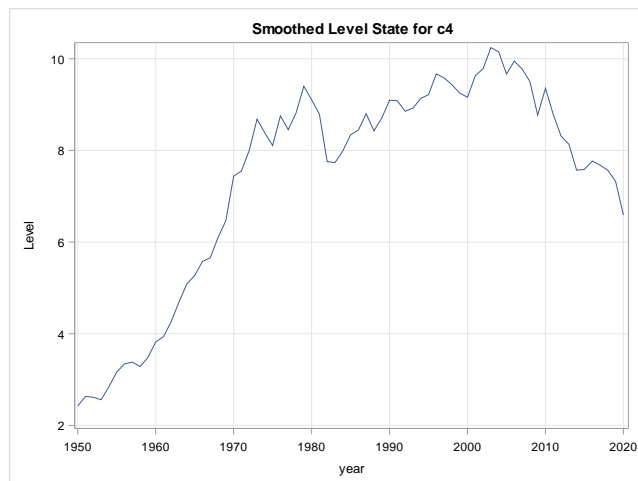


Figura 21

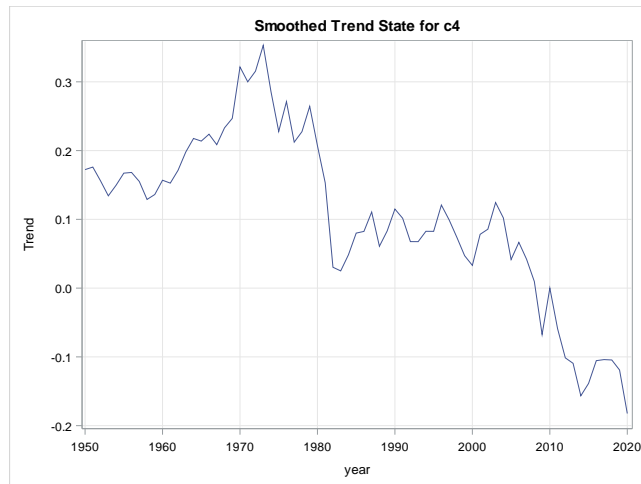


Figura 22

El suavizado ajustado para este cluster apunta a una bajada con una pendiente similar a la que presenta la serie original entre 2005 y 2020, Tabla 12 y Fig. 23. A su vez, se aprecian unos intervalos de confianza que no indican un gran incremento en el peor de los casos, como mucho alcanzando niveles parecidos a los de principios de los 90. En cuanto al mejor de los casos, los países del cluster volverían a niveles pre-años 50.

Forecast Summary												
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
c4	6.40	6.22	6.04	5.86	5.67	5.49	5.31	5.13	4.94	4.76	4.58	4.40

Tabla 12

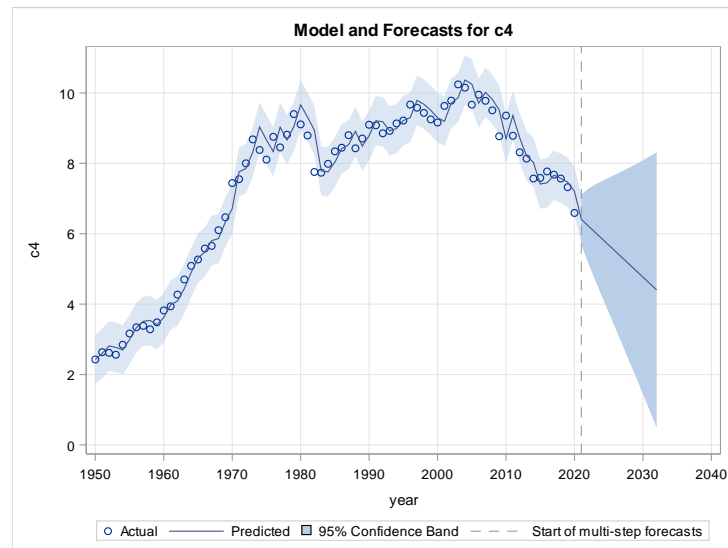


Figura 23

En cuanto a las medidas de ajuste, Tabla 13, podemos afirmar que estamos ante un modelo adecuado que presenta una proporción de variabilidad explicada de la cuarta agrupación del 97.65% y unos valores de AIC y SBC considerablemente mínimos.

Statistics of Fit for Variable c4	
Statistic	Value
R-Square	0.97650698
Akaike Information Criterion	-143.68766
Schwarz Bayesian Information Criterion	-139.1623

Tabla 13

4.2.5. Cluster 5

Los t-valores y sus p-valores, Tabla 14, nos permiten afirmar que los parámetros del modelo de Holt son significativos y que en efecto aportan información relevante sobre el cluster 5. En relación a la evolución de ambos, Fig. 24 y Fig. 25, llama la atención el particular parecido entre la tendencia de este conglomerado y la del tercero, presentando ambos un valle muy marcado entorno a los años 90.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08749	11.42	<.0001
Trend Weight	0.29379	0.08394	3.50	0.0008

Tabla 14

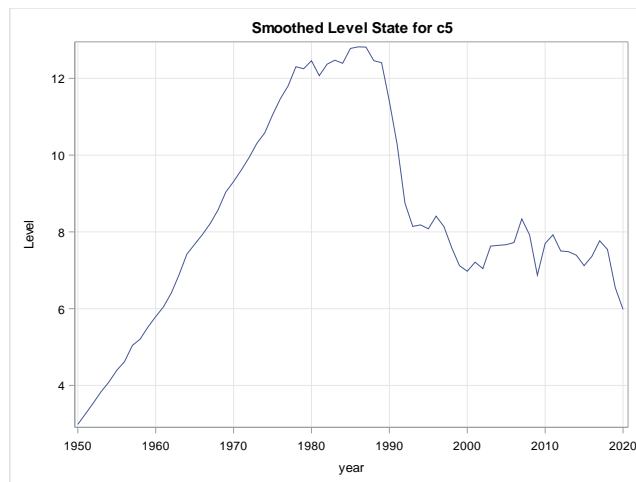


Figura 24

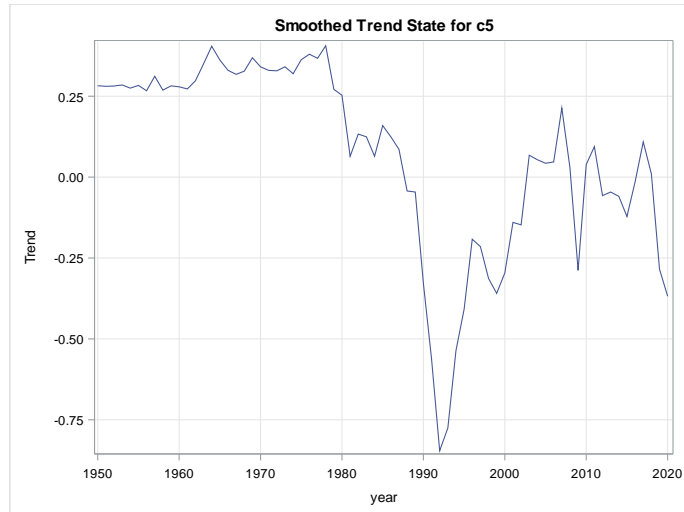


Figura 25

Tras modelizar el alisado y estimar sus predicciones, Tabla 15 y Fig. 26, podemos apreciar una tendencia bajista en las emisiones de CO₂, llegando a alcanzar en 2032 unos niveles similares a los de 1950. Ahora bien, como también ocurre con otros modelos de este trabajo, no hay que dejar de considerar los intervalos de confianza y que, por tanto, el suavizado no descarta que se puedan producir subidas que podrían situar las emisiones de la agrupación en una situación equiparable a la de los años 90.

Forecast Summary												
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
c5	5.61	5.24	4.87	4.50	4.13	3.77	3.40	3.03	2.66	2.29	1.92	1.56

Tabla 15

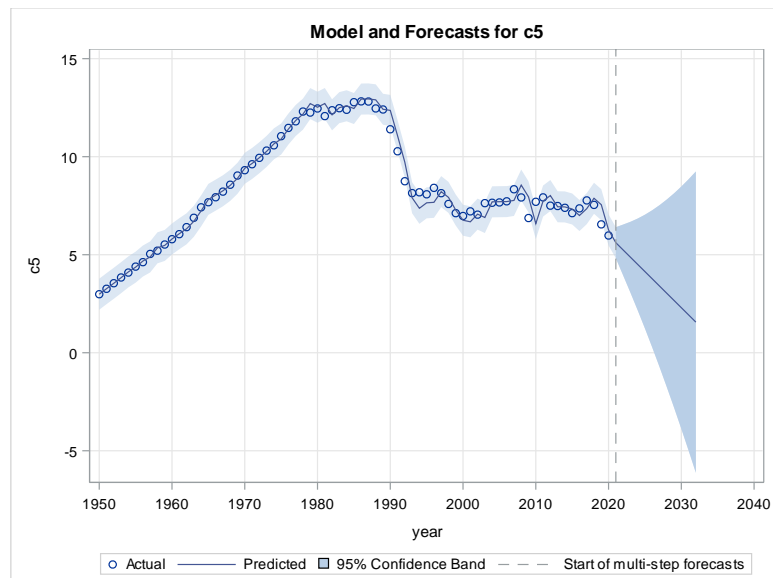


Figura 26

Los estadísticos presentados en la siguiente tabla, Tabla 16, apuntan a un buen modelo según los bajos valores del AIC y del SBC, y una proporción de variabilidad explicada del 97.66%.

Statistics of Fit for Variable c5	
Statistic	Value
R-Square	0.97661609
Akaike Information Criterion	-126.80449
Schwarz Bayesian Information Criterion	-122.27913

Tabla 16

4.2.6. Cluster 6

A diferencia del resto de clusters, el cluster 6 no presenta un parámetro de tendencia significativo, Tabla 17, por lo que en vez de usar un Alisado Doble de Holt, usaremos un Suavizado Simple EWMA. Dicho suavizado considera únicamente el parámetro de nivel, Tabla 18 y Fig. 27.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08248	12.11	<.0001
Trend Weight	0.0010000	0.0096649	0.10	0.9179

Tabla 17

Simple Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08218	12.16	<.0001

Tabla 18

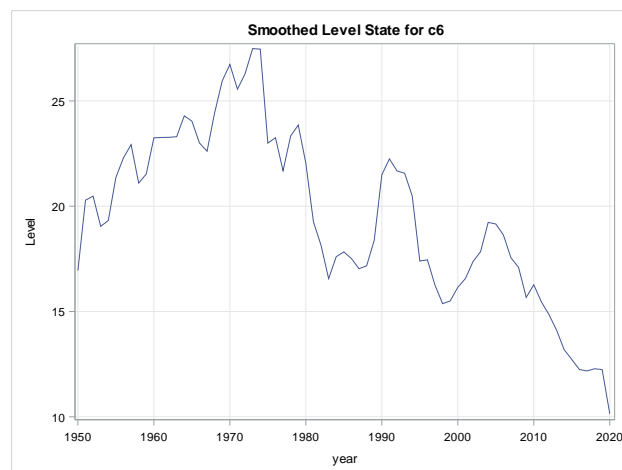


Figura 27

De esta forma, el modelo resultante da lugar a la misma predicción para todas las nuevas observaciones, Tabla 19 y Fig. 28, por lo que, aunque está es mínimamente inferior a la de 2020, no se puede concluir que vaya a haber ningún cambio relevante. Lo que sí podemos decir observando

sus intervalos de confianza es que no se considera una bajada que pueda llegar a valores nulos ni que se vayan a alcanzar valores por encima de los anteriores a los años 80.

Forecast Summary											
Variable	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
c6	10.14	10.14	10.14	10.14	10.14	10.14	10.14	10.14	10.14	10.14	10.14

Tabla 19

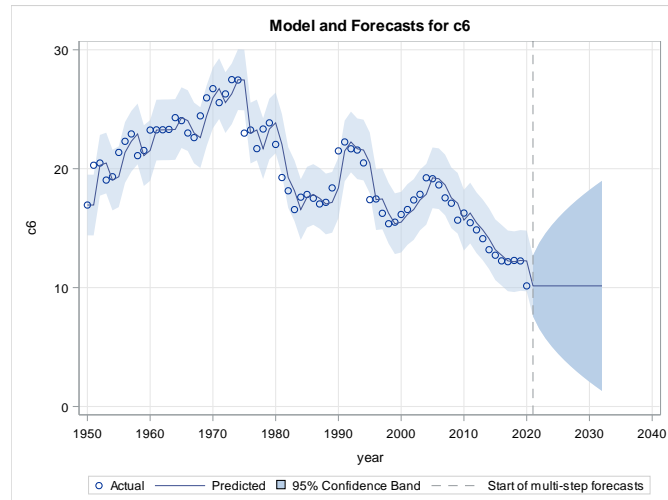


Figura 28

Aunque con el gráfico anterior, Fig. 28, ya se puede apreciar que este suavizado es bastante peor que los anteriores, fijándonos en las medidas de ajuste, Tabla 20, podemos confirmar dicha sospecha. Mientras que en los clusters anteriores la proporción de variabilidad explicada es muy próxima al 100%, esta se queda en torno al 90% y sus niveles de AIC y de SBC son considerablemente más altos.

Statistics of Fit for Variable c6	
Statistic	Value
R-Square	0.89950811
Akaike Information Criterion	38.5066436
Schwarz Bayesian Information Criterion	40.7693235

Tabla 20

4.3. Modelo ARIMA

En este apartado se recurrirá a los modelos ARIMA, muy útiles a la hora hacer predicciones en base a valores previos. De esta manera, se proporcionará un segundo modelo para cada cluster, permitiéndonos sacar mejores conclusiones.

Recordamos que todas las series han de ser estacionarias antes de poder empezar con la estimación del modelo. Puesto que todas presentan un comportamiento claramente no estacionario, estas se deberán transformar, bastando con una diferenciación de orden 1 y tomando la nueva serie la siguiente forma:

$$x_t = y(1) = y_t - y_{t-1}$$

4.3.1. Cluster 1

En la Fig. 29 podemos apreciar la falta de estacionariedad en la serie, tanto en la tendencia que presenta el gráfico de la izquierda como en el lento decrecimiento de sus correlaciones simples.

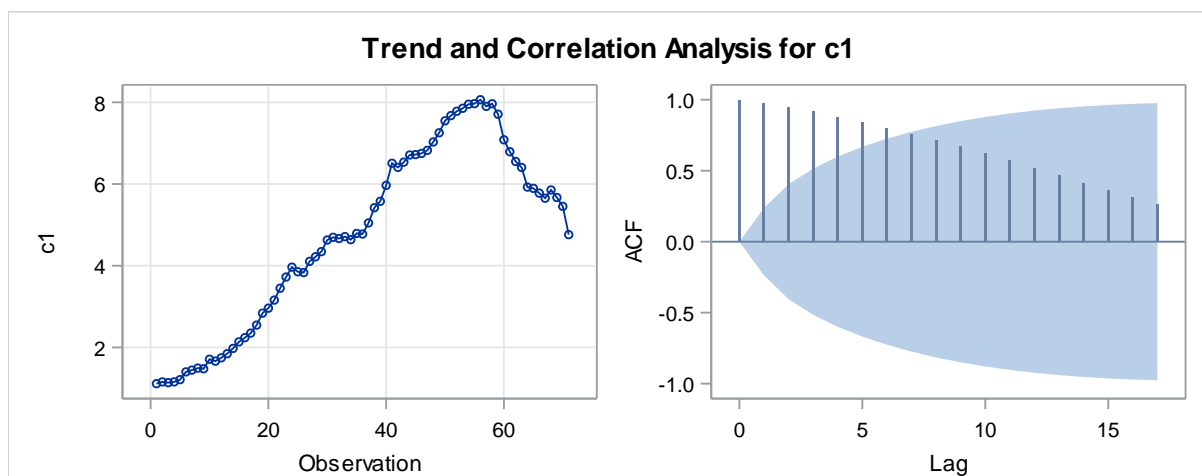


Figura 29

Realizamos una diferenciación de orden 1 dando lugar a una nueva serie con media 0.05 y desviación típica 0.21, Tabla 21. Por otra parte, tanto el Test Aumentado de Dickey-Fuller, Tabla 22, como el Test KPSS, Tabla 23, indican que la serie es estacionaria.

Name of Variable = c1	
Period(s) of Differencing	1
Mean of Working Series	0.052098
Standard Deviation	0.210843
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 21

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-96.1094	<.0001	-12.0824	<.0001		
Single Mean	0	-96.1898	0.0007	-12.0229	0.0001	72.3107	<.0010
Trend	0	-96.5343	0.0002	-12.0698	<.0001	73.0132	<.0010

Tabla 22

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	4	0.2245	0.2249
Trend	5	0.0656	0.3886

Tabla 23

A continuación, trataremos de comprobar la estructura estacionaria del modelo ARIMA, si es que existe, a través de un modelo general ARMA.

En la Tabla 24 se presenta el Test de Ljun-Box y, como era de esperar, concluimos que los residuales no son ruido blanco puesto que rechazamos la hipótesis nula para los retardos presentados.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	50.63	6	<.0001	0.468	0.386	0.359	0.294	0.237	0.194
12	69.45	12	<.0001	0.263	0.166	0.211	0.191	0.218	0.056

Tabla 24

A la hora de elegir los parámetros de ARMA se han considerado como candidatos los obtenidos mediante los métodos automáticos MINIC, SCAN y ESACF, Tabla 25, junto con otros que parecían de interés atendiendo al comportamiento de los correlogramas.

Minimum Table Value: BIC(1,1) = -3.3934

ARMA(p+d,q) Tentative Order Selection Tests					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
1	1	-3.3934	1	1	-3.3934
2	0	-3.37436	3	0	-3.35244
0	4	-3.19916	4	0	-3.31504
			0	3	-3.21613

Tabla 25

Tras estudiar la adecuación de los modelos y sus estadísticos de ajuste se ha concluido que el mejor modelo para el cluster uno es un ARMA(1,1), Tabla 26 y Tabla 27 .

Autoregressive Factors	
Factor 1:	1 - 0.90453 B**(1)

Tabla 26

Moving Average Factors	
Factor 1:	1 - 0.54174 B**(1)

Tabla 27

Es decir,

$$\begin{aligned}
 (1 - 0.90453 B)x_t &= (1 - 0.54174B)\varepsilon_t \\
 x_t &= 0.90453 x_{t-1} + \varepsilon_t - 0.54174\varepsilon_{t-1} \\
 y_t - y_{t-1} &= 0.90453(y_{t-1} - y_{t-2}) + \varepsilon_t - 0.54174\varepsilon_{t-1} \\
 y_t &= 1.90453 y_{t-1} - 0.90453y_{t-2} + \varepsilon_t - 0.54174\varepsilon_{t-1}
 \end{aligned}$$

Verificamos que los parámetros son significativos, Tabla 28, y que los residuos son incorrelados, Tabla 29 y Fig. 30. El modelo es adecuado.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.54174	0.16129	3.36	0.0008	1
AR1,1	0.90453	0.08476	10.67	<.0001	1

Tabla 28

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.02	4	0.7326	0.017	-0.010	0.069	-0.047	-0.079	-0.112
12	6.14	10	0.8030	0.054	-0.046	0.062	0.058	0.191	0.011
18	12.63	16	0.6996	0.005	0.122	-0.055	-0.072	0.158	0.144
24	18.06	22	0.7027	-0.091	-0.055	-0.037	-0.047	-0.184	-0.051

Tabla 29

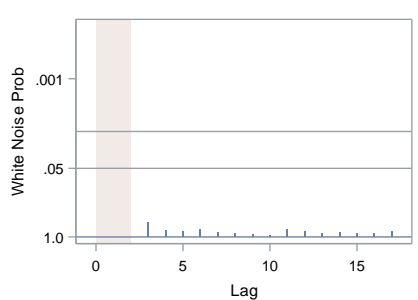


Figura 30

Este modelo se ha seleccionado entre los candidatos válidos puesto que minimiza tanto el valor del de Akaike como el de Schwarz con un AIC de -42.47 y un SBC de -37.97, Tabla 30. Aunque los valores de ambos estadísticos son bastante buenos, el ajuste del suavizado parece ser mejor atendiendo a estos dos criterios.

Variance Estimate	0.030718
Std Error Estimate	0.175264
AIC	-42.4767
SBC	-37.9797
Number of Residuals	70

Tabla 30

Una vez ajustado el modelo pasaremos a examinar sus predicciones, Tabla 31 y Fig 31 . Al igual que con el modelo exponencial doble, las predicciones apuntan a un descenso de las emisiones en los próximos años. No obstante, las predicciones del ARIMA(1,1,1) pronostican una bajada más lenta y como mucho contemplan una pequeña subida apenas perceptible.

Forecasts for variable c1				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	4.4489	0.1753	4.1054	4.7924
2022	4.1670	0.2963	3.5863	4.7476
2023	3.9120	0.4190	3.0907	4.7333
2024	3.6814	0.5449	2.6133	4.7494
2025	3.4727	0.6733	2.1531	4.7924
2026	3.2840	0.8032	1.7097	4.8583
2027	3.1134	0.9339	1.2829	4.9438
2028	2.9590	1.0647	0.8721	5.0458
2029	2.8193	1.1951	0.4769	5.1617
2030	2.6930	1.3247	0.0967	5.2893
2031	2.5788	1.4530	-0.2691	5.4266
2032	2.4754	1.5799	-0.6212	5.5720

Tabla 31

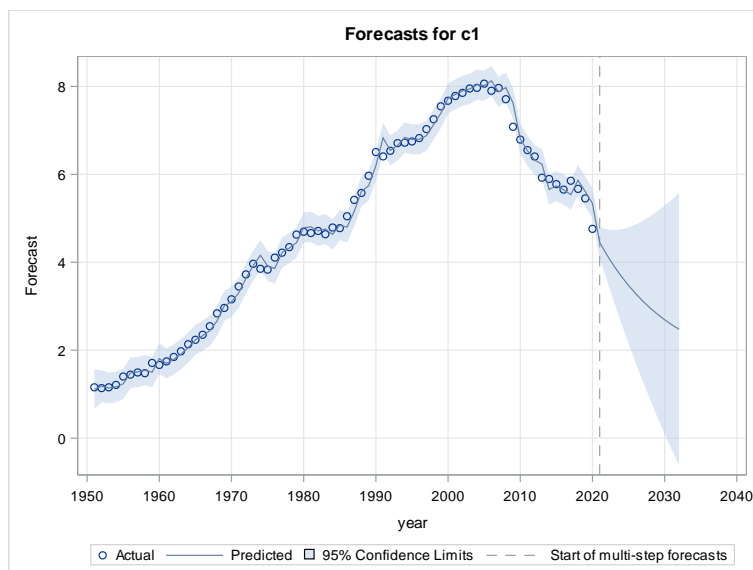


Figura 31

4.3.2. Cluster 2

Atendiendo tanto a su forma como a su correlograma, Fig. 32, podemos concluir que la serie no es estacionaria y que por tanto es necesario transformarla.

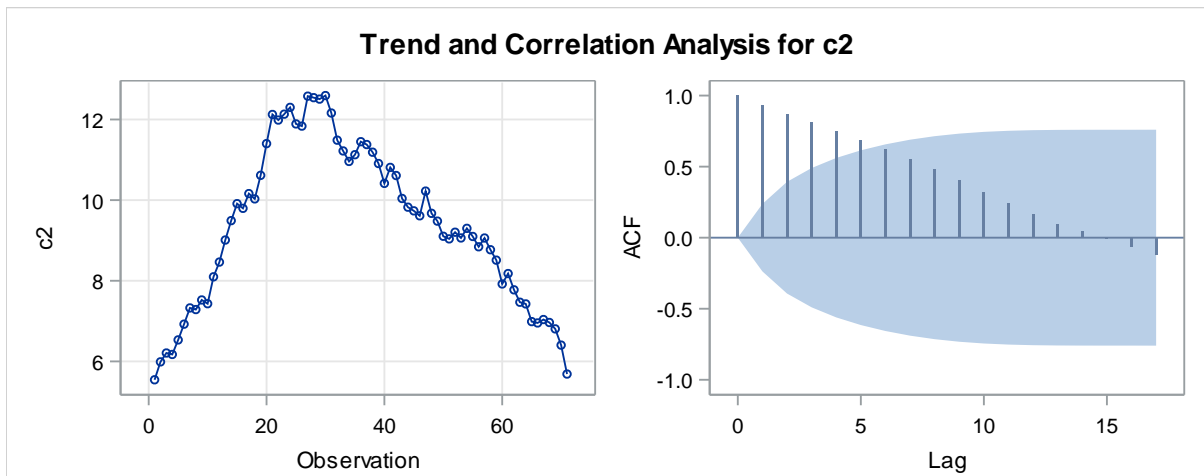


Figura 32

Name of Variable = c2	
Period(s) of Differencing	1
Mean of Working Series	0.002036
Standard Deviation	0.362709
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 32

Tras realizar la diferenciación de orden 1 la serie presenta una media muy próxima a 0 y una desviación típica de 0.36, Tabla 32. En relación a los contrastes de estacionariedad, comprobamos que el Test Aumentado de Dickey-Fuller, Tabla 33, y el test KPSS, Tabla 34, a excepción de este último para la media, apuntan a que la serie transformada es en efecto estacionaria. Por otra parte, el Test de Ljun-Box, Tabla 35, indica que sus residuos no son ruido blanco.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-48.6112	<.0001	-5.9634	<.0001		
Single Mean	0	-48.5561	0.0007	-5.9106	0.0001	17.5470	<.0010
Trend	0	-65.0579	0.0002	-7.5847	<.0001	28.7811	<.0010

Tabla 33

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	2	1.3694	0.0003
Trend	1	0.0979	0.1703

Tabla 34

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	38.62	6	<.0001	-0.658	0.150	0.054	-0.028	-0.148	0.236
12	58.19	12	<.0001	-0.141	0.092	-0.156	0.239	-0.262	0.242

Tabla 35

Minimum Table Value: BIC(1,1) = -2.22273

ARMA(p+d,q) Tentative Order Selection Tests					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
1	1	-2.22273	1	1	-2.22273

Tabla 36

Tras evaluar varios modelos, incluidos los de los métodos MINIC, SCAN y ESACF, Tabla 36, el que ha presentado mejores resultados vuelve a ser un ARMA(1,1), Tabla 37 y Tabla 38 .

Autoregressive Factors	
Factor 1:	1 - 0.9704 B**(1)

Tabla 37

Moving Average Factors	
Factor 1:	1 - 0.83992 B**(1)

Tabla 38

Es decir,

$$\begin{aligned}
 (1 - 0.9704B)x_t &= (1 - 0.83992B)\varepsilon_t \\
 x_t &= 0.9704x_{t-1} + \varepsilon_t - 0.83992\varepsilon_{t-1} \\
 y_t - y_{t-1} &= 0.9704(y_{t-1} - y_{t-2}) + \varepsilon_t - 0.83992\varepsilon_{t-1} \\
 y_t &= 1.9704y_{t-1} - 0.9704y_{t-2} + \varepsilon_t - 0.83992\varepsilon_{t-1}
 \end{aligned}$$

Fijándonos en la Tabla 39 comprobamos que todos los parámetros son significativos y que los residuos están incorrelados, Tabla 33 y Fig. 40.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.83992	0.09927	8.46	<.0001	1
AR1,1	0.97040	0.04206	23.07	<.0001	1

Tabla 39

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.43	4	0.0772	0.021	-0.042	-0.068	-0.177	-0.091	0.249
12	12.94	10	0.2271	0.122	0.125	0.046	0.097	-0.113	-0.011
18	16.75	16	0.4022	-0.158	0.073	0.071	0.067	0.033	-0.035
24	21.85	22	0.4692	-0.184	0.030	-0.010	0.020	-0.106	-0.060

Tabla 40

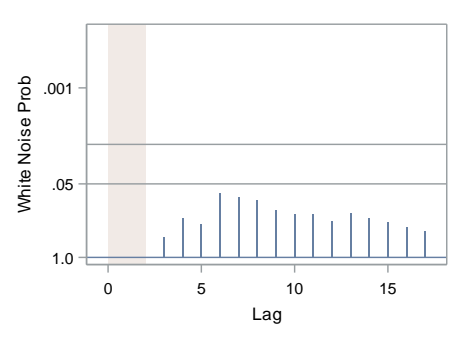


Figura 33

Los criterios AIC y SBC, Tabla 41, no presentan unos valores tan buenos como el ARIMA(1,1,1) del cluster anterior sin embargo, de todos los modelos factibles para este conglomerado, este es el que los minimiza.

Variance Estimate	0.111057
Std Error Estimate	0.333253
AIC	47.47132
SBC	51.96831
Number of Residuals	70

Tabla 41

Las estimaciones del ARIMA ajustado, Tabla 42 y Fig. 34, parecen ceñirse bien a los valores reales del cluster con predicciones bastante parecidas a las del suavizado de Holt. Mediante este modelo volvemos a apreciar una bajada que no desestima el riesgo de producirse un repunte.

Forecasts for variable c2				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	5.4725	0.3333	4.8193	6.1256
2022	5.2647	0.5030	4.2788	6.2505
2023	5.0630	0.6546	3.7800	6.3460
2024	4.8673	0.8000	3.2994	6.4353
2025	4.6774	0.9432	2.8288	6.5260
2026	4.4931	1.0859	2.3648	6.6215
2027	4.3143	1.2290	1.9055	6.7232
2028	4.1408	1.3729	1.4499	6.8317
2029	3.9724	1.5178	0.9976	6.9472
2030	3.8090	1.6637	0.5482	7.0698
2031	3.6504	1.8106	0.1016	7.1992
2032	3.4966	1.9586	-0.3422	7.3353

Tabla 42

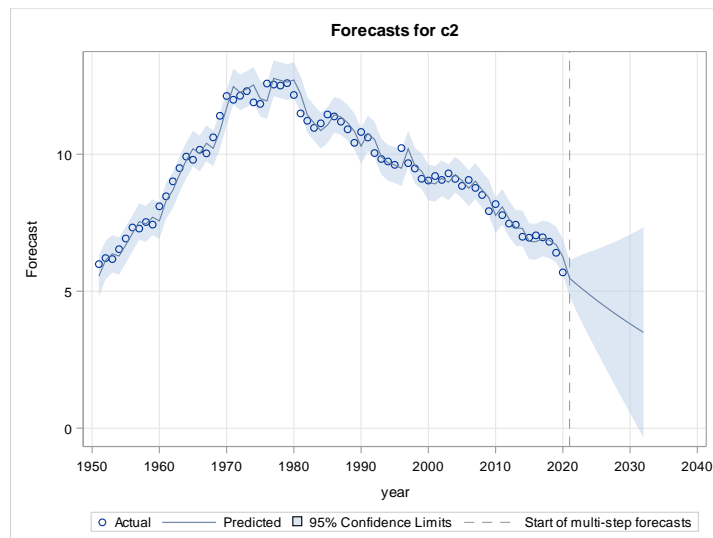


Figura 34

4.3.3. Cluster 3

Tanto el correlograma de la serie original como su gráfico, Fig. 35, indican falta de estacionariedad por lo que no se podrá usar directamente para encontrar los parámetros AR y MA.

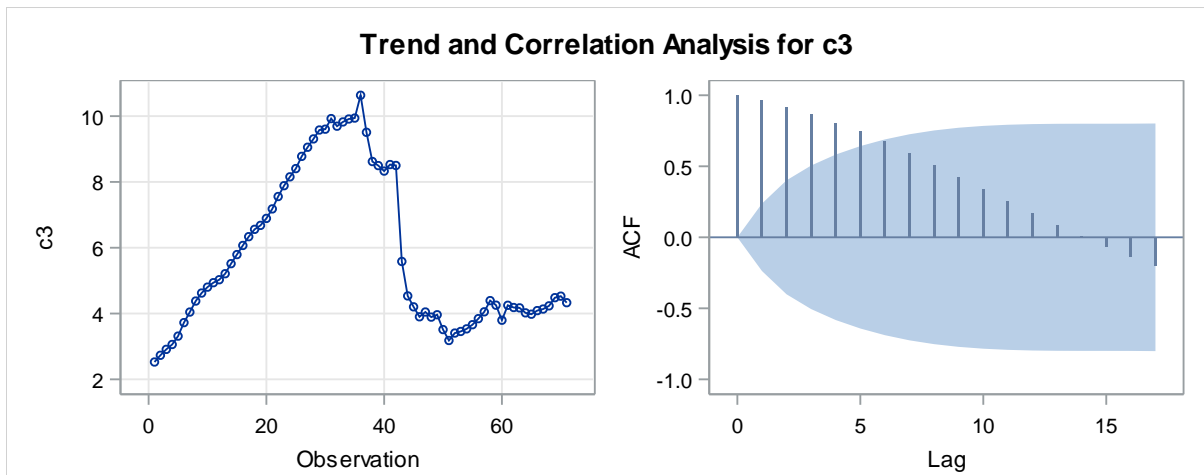


Figura 35

En concreto, para poder cumplir la condición de estacionariedad, la serie precisa de una diferenciación de orden 1, dando lugar a una serie transformada de media 0.025 y desviación típica 0.47, Tabla 43.

Name of Variable = c3	
Period(s) of Differencing	1
Mean of Working Series	0.025761
Standard Deviation	0.474073
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 43

Como podemos observar en las siguientes dos tablas, Tabla 44 y Tabla 45, se rechaza la hipótesis nula de falta de estacionariedad para el caso del Test Aumentado de Dickey-Fuller pero también se rechazar la hipótesis nula de estacionariedad para el Test KPSS. Sin embargo, consideraremos esta transformación como suficiente y procederemos con la modelización.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-44.4918	<.0001	-5.6886	<.0001		
Single Mean	0	-44.6052	0.0007	-5.6529	0.0001	15.9835	<.0010
Trend	0	-47.4379	0.0002	-5.8798	<.0001	17.2862	<.0010

Tabla 44

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	2	0.4991	0.0395
Trend	2	0.1845	0.0211

Tabla 45

Estudiando los residuos a través del Test de Ljun-Box, Tabla 46, concluimos que estos no son ruido blanco, condición que se deberá cumplir en los modelos ARIMA para que estos sean válidos.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	31.92	6	<.0001	0.352	0.126	0.163	0.067	0.340	0.363
12	34.61	12	0.0005	0.130	0.105	-0.005	-0.015	0.065	-0.026

Tabla 46

Minimum Table Value: BIC(1,0) = -1.57966

ARMA(p+d,q) Tentative Order Selection Tests					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
1	4	-1.38722	5	1	-1.47321
5	0	-1.51046			

Tabla 47

En la tabla anterior, Tabla 47, se presentan los candidatos que proponen los métodos MINIC, SCAN y ESACF. No obstante, de todos los modelos considerados, el óptimo es el modelo autorregresivo siguiente, Tabla 48.

Autoregressive Factors	
Factor 1:	1 - 0.32859 B**(1) - 0.30463 B**(5)

Tabla 48

Es decir,

$$(1 - 0.32859B - 0.30463B^5)x_t = 0$$

$$x_t = 0.3285x_{t-1} + 0.30463x_{t-5}$$

$$y_t - y_{t-1} = 0.3285(y_{t-1} - y_{t-2}) + 0.30463(y_{t-5} - y_{t-6})$$

$$y_t = 1.3285y_{t-1} - 0.3285y_{t-2} + 0.30463y_{t-5} - 0.30463y_{t-6}$$

Comprobamos que el modelo es adecuado verificando que todos sus parámetros son significativos, Tabla 49, y que los residuos están incorrelados, Tabla 50 y Fig. 36.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
AR1,1	0.32859	0.10602	3.10	0.0019	1
AR1,2	0.30463	0.10633	2.86	0.0042	5

Tabla 49

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.45	4	0.1678	-0.036	-0.077	0.139	-0.105	0.004	0.215
12	8.20	10	0.6095	-0.010	0.044	-0.003	-0.089	-0.031	-0.098
18	13.98	16	0.6004	0.076	-0.042	-0.158	0.056	0.027	-0.159
24	15.00	22	0.8624	0.036	-0.000	-0.017	0.005	-0.034	-0.082

Tabla 50

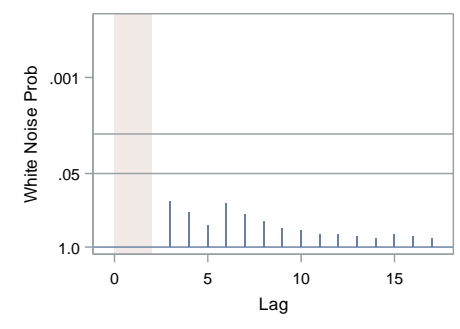


Figura 36

Recurriendo a las medidas de adecuación presentadas en la Tabla 51, podemos comparar el modelo con los ajustados para los dos clusters anteriores. Todos los criterios indican que este modelo es el peor y esto se puede ver reflejado en los intervalos de confianza de las predicciones.

Variance Estimate	0.178828
Std Error Estimate	0.42288
AIC	80.82001
SBC	85.317
Number of Residuals	70

Tabla 51

Aunque se observan ciertas discrepancias entre las estimaciones y los valores reales, Fig. 37, todo apunta a una cierta estabilización de las emisiones de Letonia y Lituania, Tabla 52. También se observan unos amplios intervalos de confianza que consideran tanto una rápida subida como una drástica caída.

Forecasts for variable c3				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	4.2791	0.4229	3.4503	5.1079
2022	4.2921	0.7032	2.9139	5.6704
2023	4.3721	0.9293	2.5508	6.1935
2024	4.4124	1.1185	2.2201	6.6046
2025	4.3652	1.2825	1.8516	6.8788
2026	4.3342	1.4897	1.4145	7.2539

Forecasts for variable c3				
Obs	Forecast	Std Error	95% Confidence Limits	
2027	4.3280	1.7117	0.9732	7.6828
2028	4.3503	1.9269	0.5736	8.1270
2029	4.3699	2.1281	0.1988	8.5410
2030	4.3620	2.3149	-0.1751	8.8990
2031	4.3499	2.5034	-0.5566	9.2564
2032	4.3441	2.6931	-0.9343	9.6224

Tabla 52

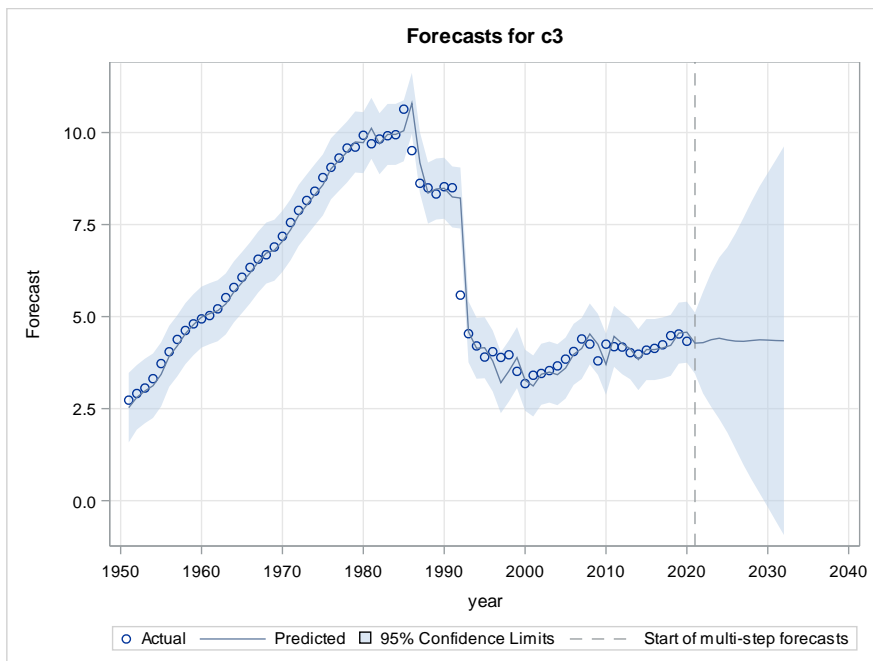


Figura 37

4.3.4. Cluster 4

Partiendo otra vez del estudio de la estacionariedad de la serie, observamos una clara falta de esta en la Fig. 38.

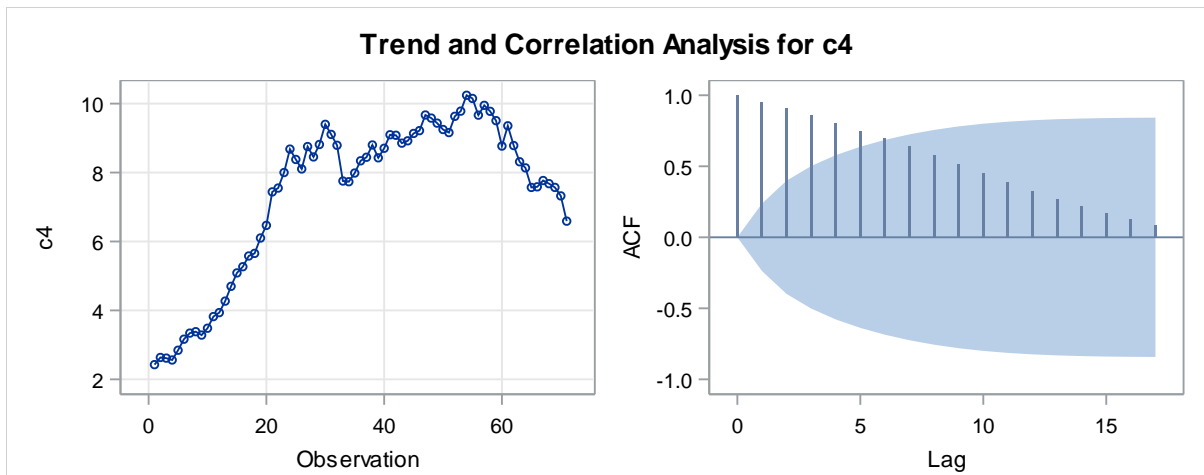


Figura 38

Name of Variable = c4	
Period(s) of Differencing	1
Mean of Working Series	0.059481
Standard Deviation	0.369265
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 53

Se ha precisado una única diferenciación y ninguna transformación de Box-Cox para poder considerar la nueva serie apta para su modelización con ARMA. Dicha serie cuenta con una media igual a 0.05 y una desviación típica de 0.36, Tabla 53.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-55.9751	<.0001	-6.6313	<.0001		
Single Mean	0	-57.6410	0.0007	-6.6992	0.0001	22.4862	<.0010
Trend	0	-67.3157	0.0002	-7.7765	<.0001	30.3130	<.0010

Tabla 54

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	1	1.1026	0.0015
Trend	0	0.0763	0.2949

Tabla 55

A excepción de un caso en el Test KPSS, Tabla 55, tanto con este contraste como con el de Dickey-Fuller Aumentado, Tabla 54, podemos concluir que la serie diferenciada es estacionaria. En relación al comportamiento de sus residuos, mediante el contraste de Ljun-Box no podemos rechazar la hipótesis nula de que estos sean ruido blanco, Tabla 56.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	11.65	6	0.0702	0.153	0.109	0.148	-0.005	0.109	0.287
12	17.88	12	0.1195	0.201	0.106	0.100	-0.007	0.081	-0.089

Tabla 56

Minimum Table Value: BIC(0,0) = -2.01083

ARMA(p+d,q) Tentative Order Selection Tests					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
1	1	-1.96653	1	1	-1.96653
			3	2	-1.81215
			5	1	-1.78275

Tabla 57

Una vez considerados varios posibles modelos, incluidos los propuestos por los métodos de selección automática de la Tabla 57, el que ha presentado mejores resultados es un ARMA(1,1), Tabla 58 y Tabla 59. Sin embargo, también conviene destacar que mediante la Tabla 56 ya se habían estudiado las correlaciones del modelo ARMA(0,0) propuesto por el método MINIC y que, aunque no presentaba las mejores medidas de ajuste, sí cumplía todas las hipótesis de adecuación.

Autoregressive Factors	
Factor 1:	1 - 0.95089 B**(1)

Tabla 58

Moving Average Factors	
Factor 1:	1 - 0.84558 B**(1)

Tabla 59

Es decir,

$$\begin{aligned}
 (1 - 0.95089 B)x_t &= (1 - 0.84558B)\varepsilon_t \\
 x_t &= 0.95089 x_{t-1} + \varepsilon_t - 0.84558\varepsilon_{t-1} \\
 y_t - y_{t-1} &= 0.95089 (y_{t-1} - y_{t-2}) + \varepsilon_t - 0.84558\varepsilon_{t-1} \\
 y_t &= 1.95089 y_{t-1} - 0.95089 y_{t-2} + \varepsilon_t - 0.84558\varepsilon_{t-1}
 \end{aligned}$$

En la Tabla 60 fijándonos en los t-valores y p-valores asociados a los parámetros concluimos que todos son significativos. Por otra parte, también verificamos que los residuos son incorrelados, Tabla 61 y Fig. 39.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.84558	0.12999	6.51	<.0001	1
AR1,1	0.95089	0.07668	12.40	<.0001	1

Tabla 60

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.15	4	0.2727	0.003	-0.034	0.025	-0.151	-0.014	0.203
12	8.76	10	0.5555	0.123	0.025	0.041	-0.064	0.035	-0.142
18	13.76	16	0.6163	-0.057	0.030	0.008	0.072	0.169	-0.122
24	18.27	22	0.6900	-0.140	-0.054	-0.046	-0.063	0.122	-0.019

Tabla 61

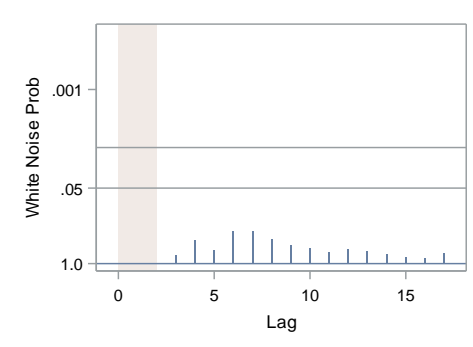


Figura 39

Los criterios de adecuación, Tabla 62, indican que el modelo no es tan bueno como el Alisado Doble de Holt y que por tanto las predicciones obtenidas con el este otro serán más apropiadas.

Variance Estimate	0.130059
Std Error Estimate	0.360637
AIC	58.1794
SBC	62.67639
Number of Residuals	70

Tabla 62

El ARIMA elegido pronostica, Tabla 63 y Fig. 40, una bajada importante que continúa en la línea de los últimos 15 años y con una pendiente menos pronunciada que la subida a mitad del siglo pasado. Obviamente, los intervalos de confianza indican que esta predicción se puede alejar mucho de la situación real pero en ningún caso sopesan que pueda ascender hasta niveles previos al 2010.

Forecasts for variable c4				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	6.4311	0.3606	5.7243	7.1379
2022	6.2799	0.5375	5.2263	7.3334
2023	6.1361	0.6913	4.7811	7.4911
2024	5.9993	0.8354	4.3619	7.6368
2025	5.8693	0.9745	3.9593	7.7793
2026	5.7457	1.1106	3.5689	7.9225
2027	5.6281	1.2449	3.1882	8.0680
2028	5.5163	1.3778	2.8160	8.2167
2029	5.4100	1.5096	2.4512	8.3689
2030	5.3090	1.6407	2.0933	8.5246
2031	5.2129	1.7709	1.7420	8.6838
2032	5.1215	1.9004	1.3967	8.8462

Tabla 63

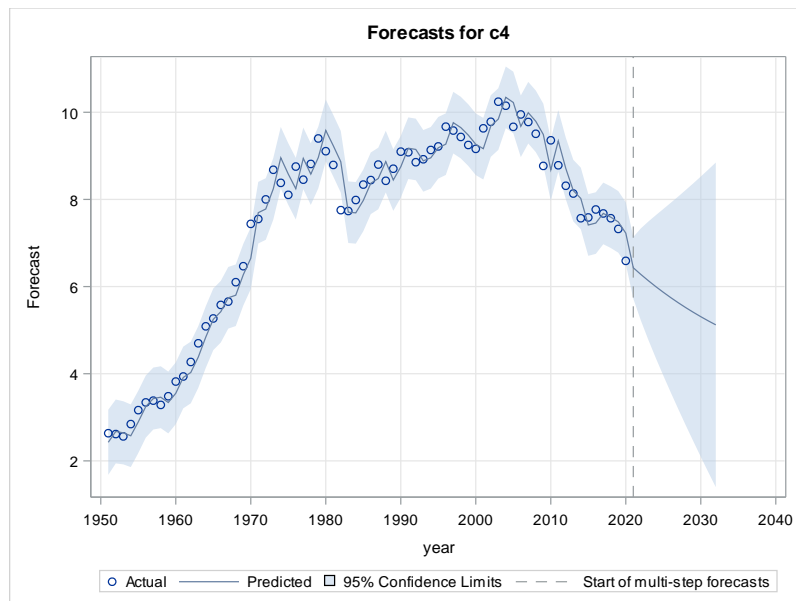


Figura 40

4.3.5. Cluster 5

La clara tendencia de las emisiones y el lento decrecimiento del correlograma, Fig. 41, indican una falta de estacionariedad que tendremos que corregiremos con una diferenciación de orden 1 y resultando en una nueva serie de media 0.04 y desviación típica 0.44 , Tabla 64 .

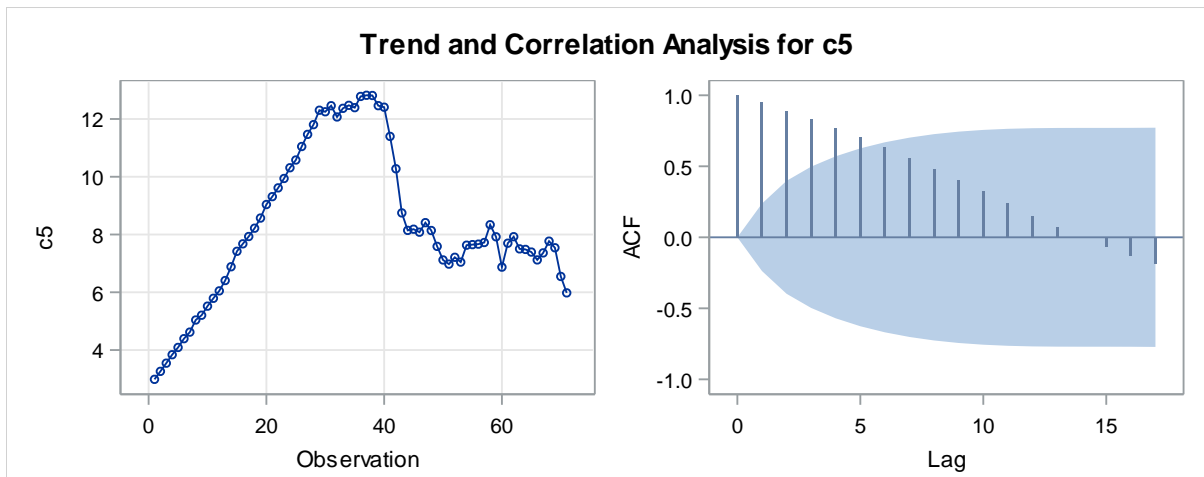


Figura 41

Name of Variable = c5	
Period(s) of Differencing	1
Mean of Working Series	0.04277
Standard Deviation	0.449906
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 64

Comprobamos con el Test Aumentado de Dickey-Fuller, Tabla 65, y el Test KPSS, Tabla 66, la correcta transformación de la serie, ambos test indican que la nueva serie es estacionaria. Recurriendo a la Tabla 67 también podemos decir que sus residuos no son ruido blanco y que por tanto no se podrá considerar un modelo ARMA(0,0).

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-34.6241	<.0001	-4.7174	<.0001		
Single Mean	0	-34.8677	0.0007	-4.6872	0.0003	11.0174	<.0010
Trend	0	-42.6490	0.0002	-5.4130	0.0002	14.6715	<.0010

Tabla 65

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	2	0.8891	0.0041
Trend	2	0.1280	0.0807

Tabla 66

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	35.07	6	<.0001	0.481	0.252	0.260	0.211	0.142	0.208
12	53.96	12	<.0001	0.345	0.090	0.134	0.220	0.191	-0.027

Tabla 67

Minimum Table Value: BIC(1,0) = -1.83332

ARMA(p+d,q) Tentative Order Selection Tests					
SCAN			ESACF		
p+d	q	BIC	p+d	q	BIC
1	0	-1.83332	0	1	-1.76944
0	1	-1.76944	1	1	-1.77732
			3	2	-1.64734

Tabla 68

En la tabla anterior, Tabla 68, se muestran los modelos propuestos por los métodos automáticos sin embargo, ninguno de ellos cumplía las hipótesis necesarias para que el modelo fuera adecuado. Por ello, se han buscado otras alternativas, siendo el modelo de la Tabla 69, en conjunto, el mejor de los estimados.

Autoregressive Factors	
Factor 1:	1 - 0.43142 B**(1) - 0.27826 B**(7)

Tabla 69

Es decir,

$$(1 - 0.43142 B - 0.27826 B^7)x_t = 0$$

$$x_t = 0.43142x_{t-1} + 0.30463x_{t-7}$$

$$y_t - y_{t-1} = 0.43142(y_{t-1} - y_{t-2}) + 0.27826(y_{t-7} - y_{t-8})$$

$$y_t = 1.43142y_{t-1} - 0.43142y_{t-2} + 0.27826y_{t-7} - 0.27826y_{t-8}$$

Junto con la estimación de los parámetros, Tabla 70, encontramos sus correspondientes p-valores que a su vez nos indican que todos ellos aportan información significativa sobre el comportamiento de las emisiones del conglomerado 5. Por otra parte, también podemos decir que todos los residuos están incorrelados, Tabla 71 y Fig. 42.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
AR1,1	0.43142	0.10269	4.20	<.0001	1
AR1,2	0.27826	0.10820	2.57	0.0101	7

Tabla 70

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.68	4	0.7946	0.055	-0.034	0.106	0.026	-0.021	0.074
12	10.48	10	0.3991	0.148	-0.207	-0.009	0.115	0.129	-0.108
18	16.91	16	0.3915	-0.104	-0.007	-0.236	0.030	0.028	-0.049
24	21.85	22	0.4687	-0.070	-0.077	0.110	0.014	-0.023	-0.152

Tabla 71

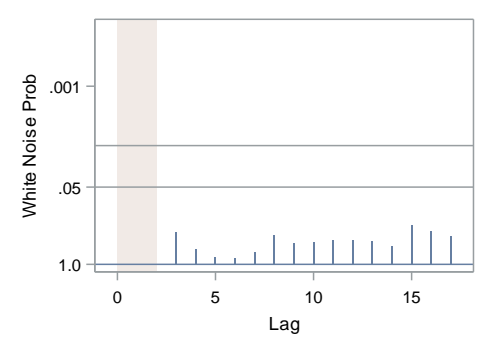


Figura 42

Los estadísticos presentados en la Tabla 72 sugieren que el modelo no es tan bueno como el suavizado, esto también se puede ver reflejado en sus intervalos de confianza.

Variance Estimate	0.14296
Std Error Estimate	0.3781
AIC	65.41235
SBC	69.90934
Number of Residuals	70

Tabla 72

Los intervalos de confianza, Tabla 73 y Fig. 43, presentan una amplitud considerable para las últimas predicciones, situándonos en el peor de los casos en emisiones parecidas a las de valores de los 90. No obstante, las predicciones prevén un descenso que contaría con un casi inapreciable repunte puntual por lo que todo apunta a una inminente bajada.

Forecasts for variable c5				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	5.7089	0.3781	4.9679	6.4500
2022	5.5166	0.6602	4.2226	6.8106
2023	5.5012	0.9000	3.7373	7.2651
2024	5.6076	1.1055	3.4409	7.7742
2025	5.5901	1.2850	3.0716	8.1085
2026	5.3059	1.4449	2.4740	8.1377
2027	5.0252	1.5898	1.9093	8.1411
2028	4.8288	1.7662	1.3672	8.2905
2029	4.6906	1.9647	0.8399	8.5414
2030	4.6267	2.1693	0.3750	8.8784
2031	4.6287	2.3695	-0.0155	9.2730
2032	4.6247	2.5611	-0.3948	9.6443

Tabla 73

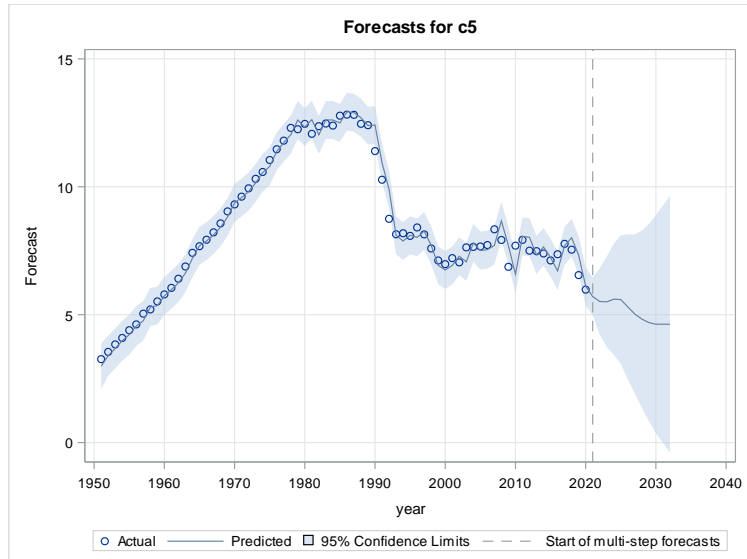


Figura 43

4.3.6. Cluster 6

Al igual que con todos los clusters anteriores, detectamos en el correlograma de la agrupación de Bélgica y Luxemburgo, Fig. 44, una falta de estacionariedad. Diferenciamos una vez y obtenemos una serie transformada con una media de -0.09 y desviación típica de 1.29, Tabla 74.

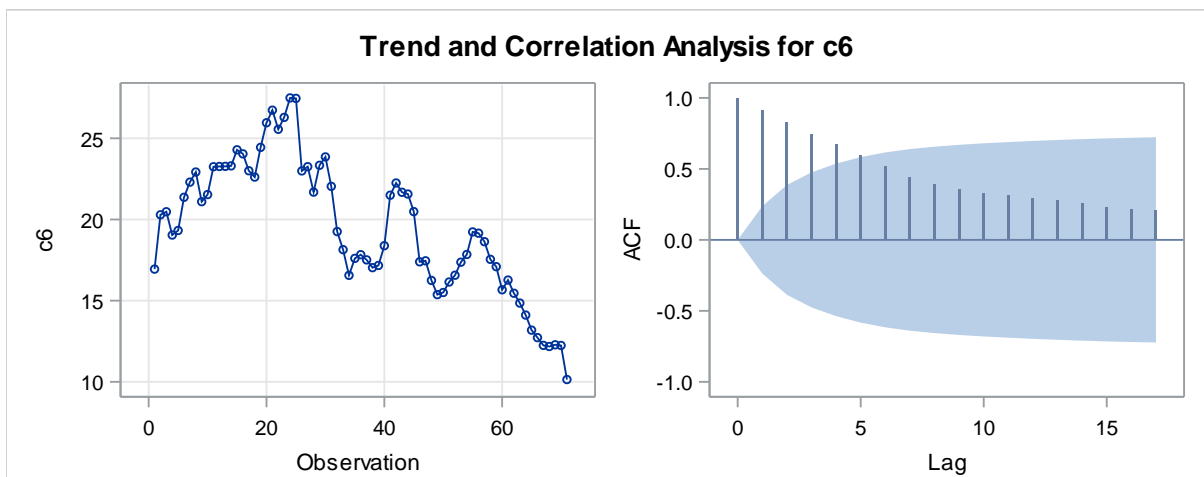


Figura 44

Name of Variable = c6	
Period(s) of Differencing	1
Mean of Working Series	-0.09715
Standard Deviation	1.298477
Number of Observations	70
Observation(s) eliminated by differencing	1

Tabla 74

Verificamos con el Test Aumentado de Dickey-Fuller, Tabla 75, y el Test KPSS, Tabla 76, que la transformación ha sido suficiente para que la nueva serie cumpla la hipótesis de estacionariedad. Por otra parte, atendiendo al contraste presentado en la Tabla 77, podemos decir que los residuos sí son ruido blanco y que por tanto se podrá considerar un modelo ARIMA(0,1,0) que daría lugar a predicciones muy parecidas a las del suavizado exponencial simple.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-54.8261	<.0001	-6.9131	<.0001		
Single Mean	0	-55.2049	0.0007	-6.9413	0.0001	24.2327	<.0010
Trend	0	-57.0759	0.0002	-7.0510	<.0001	24.8891	<.0010

Tabla 75

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	1	0.3582	0.0928
Trend	1	0.0640	0.4044

Tabla 76

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.49	6	0.4827	0.192	-0.059	-0.133	0.080	0.059	0.077
12	14.10	12	0.2944	-0.293	-0.026	-0.036	-0.066	-0.030	0.114

Tabla 77

Minimum Table Value: BIC(0,1) = 0.259473

ARMA(p+d,q) Tentative Order Selection Tests						
SCAN			ESACF			
p+d	q	BIC	p+d	q	BIC	
0	0	0.267096	0	0	0.267096	
			3	2	0.456866	

Tabla 78

Tras estudiar los candidatos propuestos por los métodos automáticos, Tabla 78, se ha elegido el siguiente modelo autorregresivo, Tabla 79, como óptimo, por lo que será éste el que se utilizará para llevar a cabo las predicciones.

Autoregressive Factors	
Factor 1:	1 - 0.23903 B**(1) + 0.32956 B**(7)

Tabla 79

Es decir,

$$(1 - 0.23903B - 0.27826B^7)x_t = 0$$

$$x_t = 0.23903x_{t-1} + 0.27826x_{t-7}$$

$$y_t - y_{t-1} = 0.23903(y_{t-1} - y_{t-2}) + 0.27826(y_{t-7} - y_{t-8})$$

$$y_t = 1.23903y_{t-1} - 0.23903y_{t-2} + 0.27826y_{t-7} - 0.27826y_{t-8}$$

Comprobamos que tanto los parámetros son significativos, Tabla 80, como que los residuos están incorrelados, Tabla 81 y Fig. 45. Se cumplen las hipótesis para que el modelo sea adecuado.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
AR1,1	0.23903	0.11254	2.12	0.0337	1
AR1,2	-0.32956	0.11208	-2.94	0.0033	7

Tabla 80

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.03	4	0.1966	0.053	-0.070	-0.159	0.070	0.050	0.195
12	8.48	10	0.5818	0.002	0.072	-0.041	-0.057	-0.007	0.136
18	14.52	16	0.5597	0.089	0.102	-0.162	-0.090	0.108	0.036
24	18.23	22	0.6922	-0.063	-0.003	0.049	0.076	-0.050	-0.142

Tabla 81

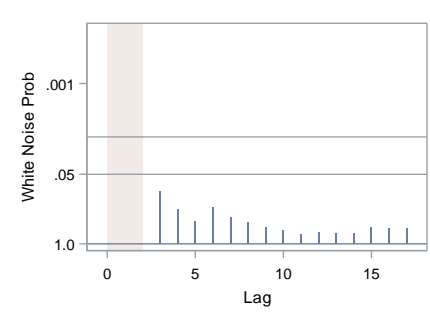


Figura 45

No obstante, los criterios de adecuación, Tabla 82, indican que el modelo no es muy bueno y esto lo veremos reflejado en sus intervalos de confianza.

Variance Estimate	1.476957
Std Error Estimate	1.215301
AIC	228.8519
SBC	233.3489
Number of Residuals	70

Tabla 82

A diferencia del suavizado, el ARIMA ajustado sí pronostica, Tabla 83 y Fig. 46, una tendencia ligeramente creciente cuyos intervalos de confianza abarcan desde una caída en la línea de los últimos 15 años hasta un crecimiento con una pendiente mínimamente más acusada.

Forecasts for variable c6				
Obs	Forecast	Std Error	95% Confidence Limits	
2021	9.9466	1.2153	7.5647	12.3285
2022	10.0520	1.9350	6.2594	13.8446
2023	10.2344	2.4951	5.3440	15.1247
2024	10.3007	2.9597	4.4999	16.1016
2025	10.2821	3.3625	3.6917	16.8724
2026	10.2911	3.7223	2.9955	17.5867
2027	10.9849	4.0504	3.0462	18.9236
2028	11.2153	4.2235	2.9375	19.4931
2029	11.2357	4.3414	2.7267	19.7446
2030	11.1804	4.4412	2.4757	19.8851
2031	11.1453	4.5344	2.2580	20.0326
2032	11.1431	4.6244	2.0794	20.2068

Tabla 83

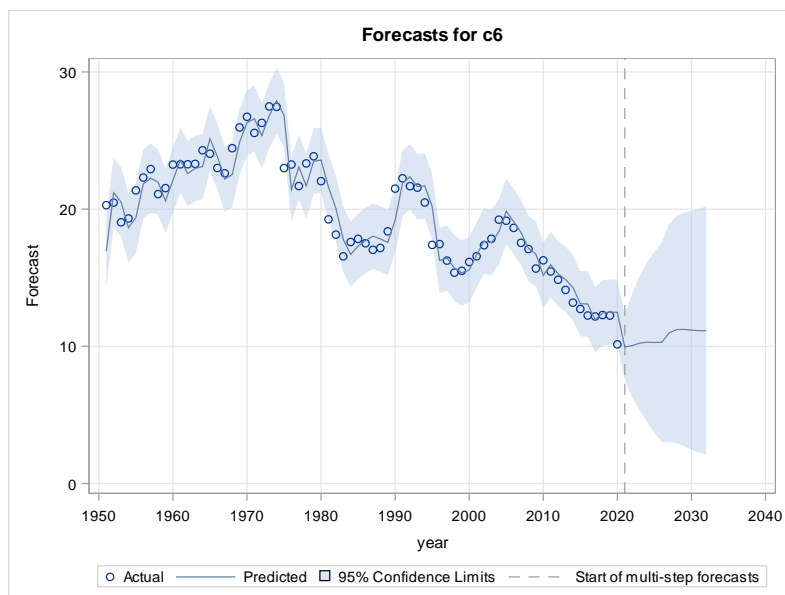


Figura 46

4.4. Adecuación

Como ya comentábamos en apartados anteriores y, atendiendo al AIC y al SBC, Tabla 84, los modelos que minimizan dichos criterios y que tomaremos como óptimos serán los suavizados. Por tanto, aunque se tendrá en cuenta en las conclusiones las predicciones obtenidas para ambos modelos, suavizado y ARIMA, se le dará más importancia en caso de duda a los primeros.

Asimismo, podemos destacar que los mejores modelos ajustados corresponden a los dos primeros clusters y que estos a su vez representan el 48% de los países de la UE por lo que sus resultados serán claves a la hora de sacar conclusiones sobre la posible evolución de la huella de carbono europea.

	AIC	SBC
Suavizado Cluster 1	-245.15907	-240.63371
ARIMA Cluster 1	-42.4767	-37.9797
Suavizado Cluster 2	-155.41465	-150.88929
ARIMA Cluster 2	47.47132	51.96831
Suavizado Cluster 3	-109.96851	-105.44315
ARIMA Cluster 3	80.82001	85.317
Suavizado Cluster 4	-143.68766	-139.1623
ARIMA Cluster 4	58.1794	62.67639
Suavizado Cluster 5	-126.80449	-122.27913
ARIMA Cluster 5	65.41235	69.90934
Suavizado Cluster 6	38.5066436	40.7693235
ARIMA Cluster 6	228.8519	233.3489

Tabla 84

5. Conclusiones

En el presente trabajo se ha podido estudiar el impacto de las políticas europeas en materias ecológicas y el posible desarrollo de estas a largo plazo en relación a las emisiones de carbono per cápita. Además de identificar aquellos países que exhiben un comportamiento más semejante en dichas emisiones, se han obtenido predicciones a 12 años vista mediante un modelo de suavizado y otro ARIMA.

Con la realización de este estudio se ha podido observar un impacto positivo por parte de las medidas tomadas en la UE para paliar los altos niveles de emisiones de CO₂, no obstante, también se han identificado importantes diferencias entre unos países y otros. Es decir, no todos son igual de susceptibles a dichas políticas por lo que sería conveniente hacer un estudio más individualizado o, como se ha hecho en este trabajo, por clusters para identificar cuál es la mejor estrategia a seguir.

Pese a esto último, las predicciones obtenidas para la gran mayoría de países de la Unión apuntan a una buena toma de decisiones y una acción acertada por parte de la Comisión Europea, pudiendo llegar a cumplir en 2050 el objetivo principal del Green Deal, hacer que Europa sea climáticamente neutral. Asimismo, todo apunta a que el valor máximo de las emisiones de carbono se habría alcanzado hace décadas, ya que parece poco probable que bajo la filosofía europea actual volvamos a niveles pre-Kyoto, habiéndose cumplido uno de los propósitos del Acuerdo de París.

En relación a los países que precisarían más atención y políticas más individualizadas, estos serían Letonia, Lituania, Bélgica y Luxemburgo. El pronóstico de estos países apunta a un pequeño incremento que contrasta negativamente con el descenso de sus homólogos europeos. Por otra parte, el 85% de los países de la EU ha visto una reducción significativa de sus emisiones en los últimos 20 años y presenta un pronóstico positivo que insta a seguir por el camino de las medidas tomadas.

Ahora bien, no hay que olvidar que los resultados obtenidos son en términos relativos y que pueden estar sesgados por el comportamiento de la población. A su vez, también hay que tener en cuenta que algunas de las políticas implementadas dependen del mercado y pueden perder efectividad en periodos de crisis mientras que otras simplemente pierden fuerza con el paso del tiempo por lo que para poder cumplir los objetivos es fundamental mantener dichas políticas al día mediante informes de evaluación periódicos.

Con todo ello, y aunque en efecto se ha tardado en reconocer el peligro de los gases de efecto invernadero, podemos concluir que los esfuerzos de los últimos años por parte de los países miembros de la UE han dado sus frutos y que, aunque queda mucho por hacer, cada día estamos más cerca de dejar atrás la emergencia climática y hacer de Europa un ejemplo de desarrollo sostenible.

6. Bibliografía

- Akaike, H. (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* AC-19:716–723.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.
- Ansley, C. F. (1979). "An Algorithm for the Exact Likelihood of a Mixed Autoregressive–Moving Average Process." *Biometrika* 66:59–65.
- Ansley, C. F., and Newbold, P. (1980). "Finite Sample Properties of Estimators for Autoregressive Moving-Average Models." *Journal of Econometrics* 13:159–183.
- Bhansali, R. J. (1980). "Autoregressive and Window Estimates of the Inverse Correlation Function." *Biometrika* 67:551–566.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
- Brocklebank, J. C., and Dickey, D. A. (2003). *SAS for Forecasting Time Series*. 2nd ed. Cary, NC: SAS Institute Inc.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. New York: Springer-Verlag.
- Choi, B. (1992). *ARMA Model Identification*. New York: Springer-Verlag.
- Dickey, D. A. (1976). "Estimation and Testing of Nonstationary Time Series." Ph.D. diss., Iowa State University.
- Everitt, B. S. (1980). *Cluster Analysis*. 2nd ed. London: Heineman Educational Books.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley & Sons.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Han, J., and Kamber, M. (2001). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann.
- Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO₂ and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>' [Online Resource] [Última revisión: mayo 2022]
- Harvey, A. C. (1981). *Time Series Models*. New York: John Wiley & Sons.
- Montgomery, D. C., and Johnson, L. A. (1976). *Forecasting and Time Series Analysis*. New York: McGraw-Hill.
- Pyle, D. (1999). *Data Preparation for Data Mining*. San Francisco: Morgan Kaufmann.
- Sankoff, D. and Kruskal, J. B. (2001). *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Stanford, CA: CSLI Publications.
- Políticas Parlamento Europeo. Retrieved from: '<https://www.europarl.europa.eu/portal/en>' [Online Resource] [Última revisión: mayo 2022]

7. Anexo

Sintaxis Código Tipo SAS

```
/*SELECCIÓN DE DATOS (1950-2020)*/

DATA TFG.DATOS50;
  SET TFG.EU27;
  WHERE T >1949;
  RUN;

/*ANÁLISIS CLUSTER*/

  proc similarity data= TFG.DATOS50 out=eu27_out4
outsum=similarity_outsum4;
  id year interval=year accumulate=total;
  TARGET Austria Belgium Bulgaria Croatia Cyprus Czechia Denmark
Estonia
Finland France Germany Greece Hungary Ireland
Italy Latvia Lithuania
Luxembourg Malta Netherlands Poland Portugal
Romania Slovakia Slovenia
Spain Sweden/normalize=standard
measure=mabsdev;
  run;

  proc cluster data=similarity_outsum4(drop=_status_ type=distance)
method=ward
pseudo rmsstd rsquare plots=den(height=rsq) out=salcluster1;
  id _input_;
  run;

  proc print data=salcluster1;run;

  proc tree data=salcluster1 out=saltree nclusters=6 ;
  id _input_;
  run;
  proc sort data=saltree; by cluster;run;
  proc print data=saltree;by cluster; run;

/*SERIES CLUSTERS*/

  data TFG.DATOS50;
  set TFG.DATOS50;
  c1=mean(Cyprus,
          Greece,
          Ireland,
          Spain,
          Portugal,
          Croatia,
          Malta
          );
  c2=mean(France,
          Sweden,
          Czechia,
          Slovakia,
          Germany,
          Denmark
          );
  c3=mean(Latvia,
```

```

        Lithuania
    );
    c4=mean(Italy,
            Slovenia,
            Austria,
            Finland,
            Netherlands
    );
    c5=mean(Poland,
            Romania,
            Bulgaria,
            Hungary,
            Estonia
    );
    c6=mean(Belgium,
            Luxembourg
    );
run;

/*SUAVIZADO SIMPLE*/

proc ESM data=c1 plots=all print=(estimates statistics forecasts states
                                summary) outfor=suavizado outstat=estad
PRINTDETAILS plot=levels ;
    forecast c1;
    id year interval=year;
run;

/*SUAVIZADO DOBLE DE HOLT*/

proc ESM data=c1 print=(estimates statistics forecasts states
                        summary) outfor=suavizado outstat=estad
PRINTDETAILS plot=all;
    forecast c1 / model=linear;
    id year interval=year;
run;

/*ARIMA*/

proc Arima data=c1; identify var=c1;
IDENTIFY VAR=c1 STATIONARITY=(adf=(1,6,12)); RUN;

data dif;
set c1;
y=c1-lag(c1);
run;

PROC AUTOREG DATA=dif;
MODEL y= / STATIONARITY=(ADF, KPSS=(kernel=qs));
RUN;

proc Arima data=c1 plot=all;
identify var=c1(1) scan esacf minic;
run;
estimate p=1 q=1 noint method=ml;run;
forecast out=previsiones id=year interval=year lead=8;run;

```