

Universidad de Valladolid

FACULTAD DE CIENCIAS



TRABAJO DE FIN DE GRADO
APLICACIÓN DEL ANÁLISIS DE SERIES TEMPORALES A
VARIABLES ATMOSFÉRICAS

Autora: Lucía Alonso Losa
Tutoras: María Pilar Rodríguez del Tío
Ana María Burgos Pérez

Índice general

Resumen	5
Abstract	5
1. Introducción	6
2. Contexto	7
2.1. Tendencia	7
2.1.1. Test de Mann-Kendall	8
2.2. Metodología de Box-Jenkins. Modelos ARIMA.	9
3. Serie Temperatura	11
3.1. Temperatura Promedio Diaria	11
3.2. Temperatura Promedio Mensual	17
3.3. Temperatura Máxima Mensual	22
3.4. Temperatura Mínima Mensual	25
4. Serie Humedad Relativa	29
4.1. Humedad Relativa Promedio Diaria	29
4.2. Humedad Relativa Promedio Mensual	34
5. Serie Radiación Solar Incidente	39
5.1. Serie Radiación Solar Incidente Acumulada Diaria	39
5.2. Serie Radiación Solar Incidente Promedio Mensual	45
6. Conclusiones	49
7. Anexo	50
Bibliografía	57

Índice de figuras

3.1.	Diagrama de cajas múltiple de la Temperatura promedio anual	13
3.2.	Descomposición de la serie diaria Temperatura	14
3.3.	Gráfico de la serie diaria Temperatura	14
3.4.	ACF de la serie diaria Temperatura	15
3.5.	Diagrama de cajas múltiple de la serie Temperatura media mensual	18
3.6.	Descomposición de la serie Temperatura media mensual	19
3.7.	Gráfico de la serie Temperatura media mensual	19
3.8.	ACF de la serie Temperatura media mensual	20
3.9.	Predicciones de la serie Temperatura media mensual	21
3.10.	Descomposición de la serie mensual Temperatura máxima	23
3.11.	Serie Temperatura máxima mensual	23
3.12.	Predicciones de la serie Temperatura máxima mensual	24
3.13.	Descomposición de la serie mensual Temperatura mínima	26
3.14.	Serie Temperatura mínima mensual	26
3.15.	Predicciones de la serie Temperatura máxima mensual	27
4.1.	Diagrama de cajas múltiple del %Humedad promedio anual	31
4.2.	Descomposición clásica de la serie diaria % Humedad	32
4.3.	Gráfico de la serie diaria %Humedad	32
4.4.	Diagrama de cajas múltiple de la serie % Humedad media mensual	35
4.5.	Descomposición clásica de la serie % Humedad media mensual	36
4.6.	Gráfico de la serie % Humedad media mensual	36
5.1.	Diagrama de cajas múltiple de la Radiación solar promedio anual	41
5.2.	Descomposición de la serie diaria Radiación	42
5.3.	Gráfico de la serie Radiación diaria	42
5.4.	ACF de la serie diaria Radiación	43
5.5.	Normalidad de los residuales de la serie diaria Radiación	44
5.6.	Diagrama de cajas múltiple de la serie Radición media mensual	46
5.7.	Descomposición de la serie Radiación media mensual	47
5.8.	Gráfico de la serie Radiación media mensual	47
5.9.	ACF de la serie Radiación media mensual	48
7.1.	Test de Ljung-Box de la serie diaria Temperatura	50
7.2.	Qqplot de la serie diaria Temperatura	50
7.3.	Test de Ljung-Box de la serie Temperatura media mensual	51

7.4.	Diagrama de cajas múltiple de la serie mensual Temperatura máxima	51
7.5.	ACF de la serie Temperatura máxima mensual	52
7.6.	Test de Ljung-Box de la serie Temperatura máxima mensual	52
7.7.	Diagrama de cajas múltiple de la serie mensual Temperatura mínima	53
7.8.	ACF de la serie Temperatura mínima mensual	53
7.9.	Test de Ljung-Box de la serie Temperatura mínima mensual	54
7.10.	ACF de la serie diaria % Humedad	54
7.11.	Test de Ljung-Box de la serie % Humedad diaria	54
7.12.	Qqplot de la serie % Humedad diaria	55
7.13.	ACF de la serie % Humedad media mensual	55
7.14.	Test de Ljung-Box de la serie %Humedad media mensual modelo 1	55
7.15.	Test de Ljung-Box de la serie %Humedad media mensual modelo 2	56
7.16.	Test de Ljung-Box de la serie Radiación diaria	56
7.17.	Test de Ljung-Box de la serie Radiación media mensual	56

Índice de tablas

2.1. Resultados del Test de Mann-Kendall para los promedios anuales	8
3.1. Estadísticos resumen de la serie anual Temperatura	12
3.2. Estadísticos resumen de la serie Temperatura media mensual	17
3.3. Estadísticos resumen de la serie mensual Temperatura máxima	22
3.4. Estadísticos resumen de la serie mensual Temperatura mínima	25
4.1. Estadísticos resumen de la serie anual %Humedad	30
4.2. Estadísticos resumen de la serie % Humedad media mensual	34
4.3. Estadísticos para la comparación de modelos	37
4.4. Comparación de modelos ARIMA	38
5.1. Estadísticos resumen de la serie diaria Radiación	40
5.2. Estadísticos resumen de la serie Radiación media mensual	45

Resumen

A lo largo de este trabajo se van a modelizar, a través de la metodología ARIMA, series climáticas. En concreto, temperatura, humedad relativa y radiación solar incidente.

Los datos de los que se parte son diarios, el análisis de las series se hará tanto con estos datos como con resúmenes mensuales.

Para estudiar la posible tendencia, o no, de estas series se utilizará el test de Mann-Kendall, test no paramétrico recomendado por la Organización Mundial de Meteorología en el estudio de series ambientales. Los programas utilizados para el tratamiento estadístico de los datos son R y SAS.

Palabras claves: ARIMA, tendencia, climatología, Mann-Kendall, R, SAS.

Abstract

Throughout this document, several climate series will be modeled by using ARIMA methodology. More specifically, parameters such as temperature, relative humidity and incident solar radiation lead these models.

The series analysis is based, mainly, on daily data. However, this research will also consider both daily and monthly synthesis.

In order to study the hypothetically tendency of these series, the Mann-Kendall test will be conducted, a non-parametric test recommended by World Meteorological Organization for climate series researches. Software used for data and statistical processing will be both R and SAS.

Keywords: ARIMA, tendency, climatology, Mann-Kendall, R, SAS.

Capítulo 1

Introducción

Los datos utilizados en este trabajo consisten en tres series temporales diaras de temperatura ($^{\circ}\text{C}$), humedad relativa (%) y radiación solar incidente (decenas de kJ/m^2 día) medidas a una altura estándar de 2 metros sobre el suelo. Las tres variables climáticas han sido suministradas por la estación metereológica de la Central Nuclear de Almaraz gracias a un proyecto de transferencia e innovación con la Universidad de Valladolid. La estación meteorológica, ubicada en la provincia de Cáceres, tiene como coordenadas geográficas $39^{\circ} 48' 29''\text{N}$, $5^{\circ} 41' 49''\text{W}$. Los periodos temporales de las series aparecen en la siguiente tabla:

Variable	Inicio	Fin	Años	Transformación datos horarios
Temperatura	01/01/2001	31/12/2019	19	Promedio
Humedad Relativa	01/03/2001	31/12/2019	17	Promedio
Radiación solar incidente	01/01/1990	31/12/2017	27	Suma

Los datos se recogieron cada hora con algunos valores ausentes que fueron imputados. A la autora de este trabajo se le han proporcionado los datos diarios ya resumidos e imputados.

En cuanto a la motivación, diversos estudios buscan describir, analizar y predecir cómo será el clima utilizando distintos métodos estadísticos que permitirán conocer variabilidad, tendencias climáticas, y detectar posibles alteraciones climáticas entre otros.

Específicamente, el análisis de tendencias de series temporales, es un método bastante aplicado en el estudio de variables climáticas. Entre los contrastes existentes para detectar tendencia está el propuesto por Mann y Kendall, recomendado por la Organización Mundial de Meteorología (OMM) [1].

El propósito de este Trabajo de Fin de Grado es llevar a cabo un análisis de series temporales, utilizando modelos ARIMA, de las variables antes mencionadas, prestando especial atención al análisis de tendencia, como se ha dicho antes, debido al carácter climático de los datos.

Capítulo 2

Contexto

La Organización Meteorológica Mundial considera importante la aplicación de métodos estadísticos en meteorología, entre ellos el análisis de tendencias y frecuencias, el análisis univariado y multivariado de series temporales y la metodología de Box-Jenkins, entre otros. Según su *Guía de instrumentos y métodos de observación* [2] las variables meteorológicas son mediciones en puntos discretos en idénticos intervalos temporales en un periodo finito de tiempo. Es decir, situaciones climáticas continuas son discretizadas para formar una serie temporal equiespaciada.

2.1. Tendencia

Si estas mediciones se grafican en función del tiempo suponen una herramienta cualitativa muy útil para detectar variaciones, dicho de otra forma, detectar tendencias. En términos estadísticos podemos definir la tendencia como el comportamiento o movimiento de una serie a largo plazo, si los datos presentan crecimiento o decrecimiento general a lo largo del tiempo.

Volviendo a la climatología, y según la *Guía de prácticas climatológicas* [3], la tendencia es una característica interesante puesto que confiere un resumen del comportamiento histórico de las observaciones. Puntualizar, que en periodos de tiempo más reducidos, lo que pueden parecer tendencias sostenidas podrían pertenecer a oscilaciones más lentas relacionadas con variaciones multi-decadales.

En el análisis de tendencia en series ambientales se utiliza frecuentemente el test no paramétrico de Mann-Kendall. En los sucesivos análisis se va a utilizar el comando de R *SeasonalMannKendall()* para series mensuales del paquete *Kendall* [4] basado en el test de Mann-Kendall.

2.1.1. Test de Mann-Kendall

El estadístico utilizado [5] en el contraste es

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{signo}(x_j - x_k) \text{ donde } \text{signo}(x_j - x_k) = \begin{cases} 1 & \text{si } (x_j - x_k) > 0 \\ 0 & \text{si } (x_j - x_k) = 0 \\ -1 & \text{si } (x_j - x_k) < 0 \end{cases}$$

Valores positivos del estadístico S apoyan la hipótesis de tendencia creciente y valores muy negativos apoyan la hipótesis de tendencia decreciente.

Su varianza puede aproximarse del siguiente modo cuando se dispone del suficiente número de observaciones, n , y sin tener en cuenta coincidencias.

$$\text{Var}(S) = \frac{1}{18}[n(n-1)(2n+5)].$$

Las hipótesis del contraste son $H0$: no hay tendencia vs. $H1$: hay tendencia creciente/decreciente, y la regla de decisión se establece a partir del valor de Z_{MK} .

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & \text{si } S < 0. \end{cases}$$

Si el test de Mann-Kendall proporciona un valor significativo, es decir un p-valor, inferior a 0.05/ 0.1 se puede decir que se rechaza la hipótesis nula por lo que las probabilidades de que sea cierta la presencia de tendencia son altas.

A continuación, se probará dicho test mediante el comando `MannKendall()`, del paquete `Kendall` también, sobre los promedios anuales calculados a partir de los datos diarios primarios de las variables propuestas.

Variable	Años	Estadístico	P-valor
Temperatura	19	0.0526	0.77957
Humedad Relativa	17	-0.412	0.023476
Radiación solar incidente	27	0.45	0.00084126

Tabla 2.1: Resultados del Test de Mann-Kendall para los promedios anuales

A la vista de la anterior tabla se puede afirmar que la temperatura es la única serie que no tiene tendencia. En cuanto a la humedad relativa y a la radiación solar sí tienen tendencia, y fijando la atención en los signos estadísticos se puede asegurar que la tendencia de la humedad es decreciente y la tendencia de la radiación solar es creciente.

Por tanto, se puede concluir en base a estos resultados, que no se observan tendencias significativas en la temperatura dentro del periodo de tiempo estudiado, en cambio, cada vez la humedad es menor y la radiación solar mayor.

2.2. Metodología de Box-Jenkins. Modelos ARIMA.

La metodología de Box-Jenkins consiste en utilizar modelos autoregresivos integrados de media móvil que pretenden ajustar, de la mejor manera posible, los valores de una serie temporal a un modelo con el fin de predecir el comportamiento de la misma. El método para construir el modelo a partir de la serie observada consta de tres etapas: indentificación, estimación y validación.

En la primera etapa, la identificación, se representa la serie en función del tiempo con el intervalo de medición. A la vista del gráfico se puede empezar a sacar conclusiones de la estructura del modelo, pero para ello es necesario estudiar la estacionaridad de la serie, concepto relacionado con la invariabilidad de las características de la serie en función del tiempo. Como ayuda para detectar la falta de estacionaridad se va a realizar el test de Dickey - Fuller mediante el comando *adf.test()* que contrasta la hipótesis H_0 : *La serie es no estacionaria/ Tiene raíz unitaria*. Una vez que la serie es estacionaria, ya sea porque originalmente lo es o se ha conseguido mediante diferenciación con el comando *diff()*, se especifican los órdenes de los posibles modelos ya que rara vez hay uno único. Los medios por los que se deciden estos órdenes son la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF). Una vez calculadas se intentarán asociar con las ACF y PACF teóricas de algunos modelos con el fin de escoger los órdenes de aquellos cuyas funciones sean más semejantes a las de la serie que se está estudiando.

La segunda etapa es la estimación de los parámetros. Para este punto se ha hecho uso de dos herramientas estadísticas diferentes. Para series con un periodo igual a 12, series mensuales, se ha usado R, en concreto el comando *arima()* [6] de la librería *Stats* que ajusta modelos ARIMA de series univariantes. Este comando para series con periodos superiores a trescientos cincuenta presenta problemas por lo que para series con periodo trescientos sesenta y cinco, series diarias, se ha utilizado el *PROC ARIMA* del programa SAS.

Por último, se procede a la validación de los modelos propuestos comprobando que cumplen las hipótesis de un modelo correcto. Lo primero que se va a tener en cuenta es el ACF y PACF de los residuales, se comprobará que los residuales son incorrelados mediante el Test de Ljung-Box que contrasta la hipótesis nula $H_0 : \rho_1(a) = \rho_2(a) = \dots = \rho_k(a) = 0$ donde $\rho_i(a)$ corresponde con la autocorrelación teórica entre las observaciones que distan k unidades en el tiempo.

El estadístico que contrasta esta hipótesis es

$$Q(k) = n(n + 2) \sum_{i=1}^k \frac{r_i^2(a)}{n-i}$$

donde $r_i(a)$ corresponde con la autocorrelación muestral entre observaciones que distan i unidades en el tiempo y n es el número de observaciones.

Para ello, en el caso de las series mensuales, se ha utilizado el comando *Box.test()* especificando el tipo de test, en este caso *type = "Ljung-Box"*. Con los p-valores resultantes de este comando se puede realizar una representación gráfica para una visualización de resultados más sencilla [7]. En el caso de las series diarias el procedimiento usado realiza dicho test y su correspondiente representación para los retardos especificados.

Además, se va a realizar un contraste de normalidad de los residuales para saber cuán fiables son las bandas de predicción. Concretamente mediante el test de Shapiro-Wilk y el comando *shapiro.test()* en R. Además, se contrastará si la media de los residuales puede considerarse cero utilizando el comando *t.test(residuals, mu = 0, conf.level = 0.95)* también en R. Sus equivalentes en SAS para las series diarias se harán con *PROC UNIVARIATE* y la opción *normal*.

Capítulo 3

Serie Temperatura

La primera de las series que se va a analizar es la temperatura, que se refiere a la medida del estado térmico del aire con respecto a su habilidad de comunicar calor a su alrededor según el Glosario Meteorológico del IDEAM basado en las definiciones que da la OMM [8].

A lo largo de este capítulo se estudiará no solo el valor promedio diario, sino también los siguientes promedios estadísticos relacionados con esta variable: valor promedio mensual, valores máximo y mínimo mensuales calculados a partir de los promedios diarios.

3.1. Temperatura Promedio Diaria

Se va a comenzar analizando la serie de promedios diarios. Pero antes de empezar la modelización de la serie es interesante hacer una descripción de los datos con los que se va a trabajar. Con este fin, se van a presentar tablas con los principales estadísticos descriptivos como son los máximos, mínimos, medias, medianas y rangos. Además, se han generado gráficos que servirán de apoyo en una primera visualización de la serie temperatura.

Análisis descriptivo

La serie temperatura diaria oscila, en el periodo analizado, entre $-0.42\text{ }^{\circ}\text{C}$ y $34.82\text{ }^{\circ}\text{C}$ con media $17.24\text{ }^{\circ}\text{C}$. Se comienza por presentar la tabla con los estadísticos resumen anuales.

Año	Mínimo	Mediana	Media	Máximo	Rango
2001	1.3	16.9	17.76	32.09	30.79
2002	4.79	16.69	17.71	31.80	27.01
2003	1.31	15.81	17.57	34.06	32.75
2004	3.59	14.41	16.76	33.5	29.91
2005	-0.38	16.34	16.9	33.43	33.81
2006	1.15	16.87	17.32	33.01	31.86
2007	-0.04	14.93	15.75	30.93	30.97
2008	3.46	15.19	16.17	31.67	28.21
2009	-0.42	16.3	17.52	34.12	34.54
2010	0.007	15.25	16.86	32.73	32.66
2011	1.97	17.81	17.49	31.82	29.85
2012	1.75	15.23	17.01	34.39	32.64
2013	2.93	15.66	16.83	32.89	29.96
2014	1.93	17.78	17.54	31.48	29.55
2015	2.56	17.39	17.82	33.06	30.50
2016	2.04	15.54	17.48	33.19	31.15
2017	1.35	17.88	18.29	34.09	32.74
2018	2.8	16.7	17.3	34.82	32.02
2019	0.69	16.26	17.48	32.14	31.45

Tabla 3.1: Estadísticos resumen de la serie anual Temperatura

Observando la tabla lo primero que llama la atención son los valores tan diferentes que toman los mínimos y los máximos, esta distancia entre ambos implica la existencia de rangos muy amplios. Esta observación es lógica teniendo en cuenta que cada uno de ellos tiene que albergar datos de todas las estaciones y las diferencias en la temperatura que eso conlleva. Dicho esto, los rangos no son muy distintos entre sí por lo que aparentemente las temperaturas a lo largo de las primeras décadas del 2000 no han variado significativamente.

Destacar la medición mínima del año 2002 con 4.79 °C y la máxima de 2018 con 34.82 °C, ambas especialmente altas en comparación.

Representando gráficamente los estadísticos vistos, se obtiene el siguiente diagrama de cajas múltiples

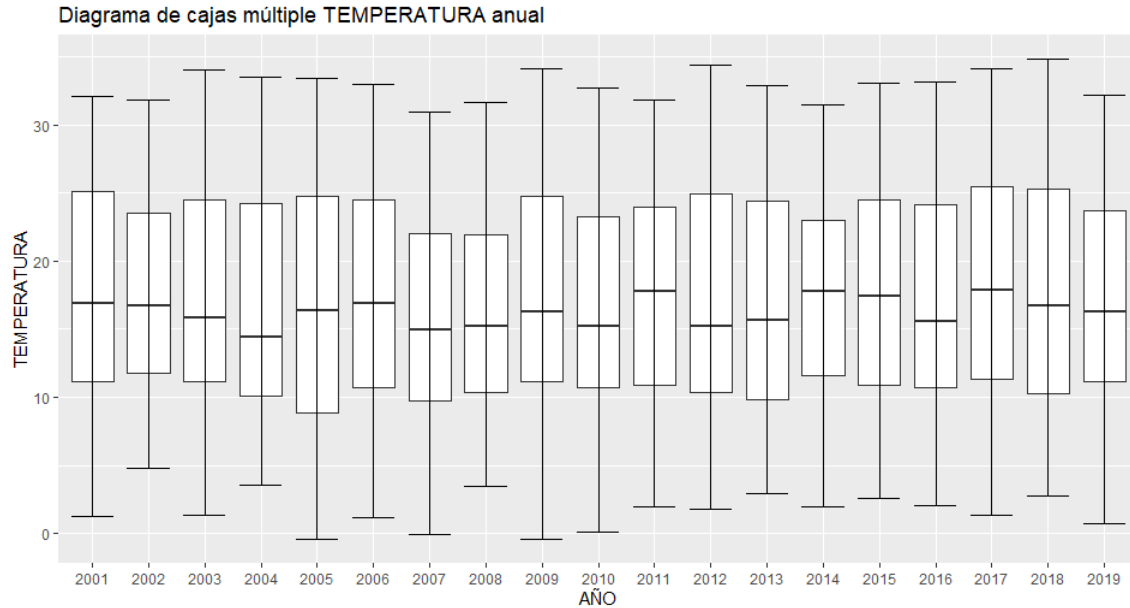


Figura 3.1: Diagrama de cajas múltiple de la Temperatura promedio anual

En el digrama anterior se ve representado claramente esos rangos tan amplios que se comentaban antes a través del gran tamaño de los bigotes, aunque entre ellos son bastante parecidos.

Respecto a los rangos intercuartílicos, se puede ver que son bastante amplios por lo que existe una gran dispersión. Esto se refiere a que el dominio de las mediciones es muy grande. Dentro de este rango se resaltan las medianas, y comparándolas entre ellas se puede decir que los valores centrales en estos años son muy similares, concluyendo otra vez la invariabilidad de la temperatura en estas dos décadas de datos.

Por último, referente al diagrama, destacar la ausencia de valores atípicos.

Para comenzar el análisis de la serie temperatura promedio diaria se va a realizar una representación gráfica de la descomposición de la serie usando medias móviles, calculada con el comando `decompose()`, donde se puede ver la tendencia, la estacionalidad y la componente irregular.

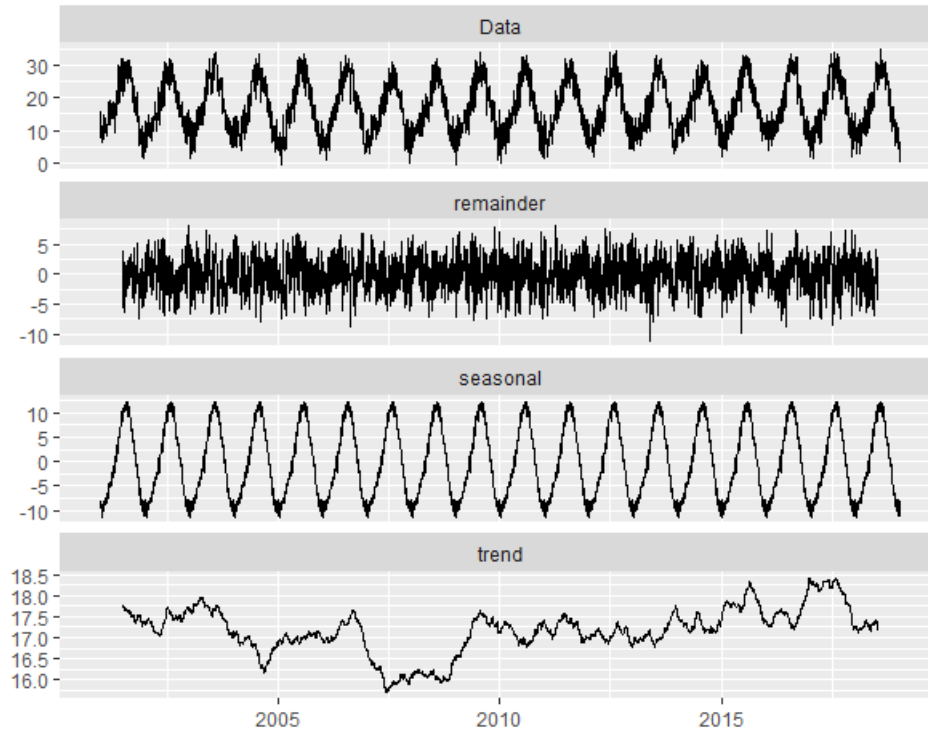


Figura 3.2: Descomposición de la serie diaria Temperatura

Análisis de Tendencia

Para el análisis de tendencia se ha realizado la representación gráfica de la serie y se utilizará de apoyo la tendencia descrita en la descomposición del apartado previo.

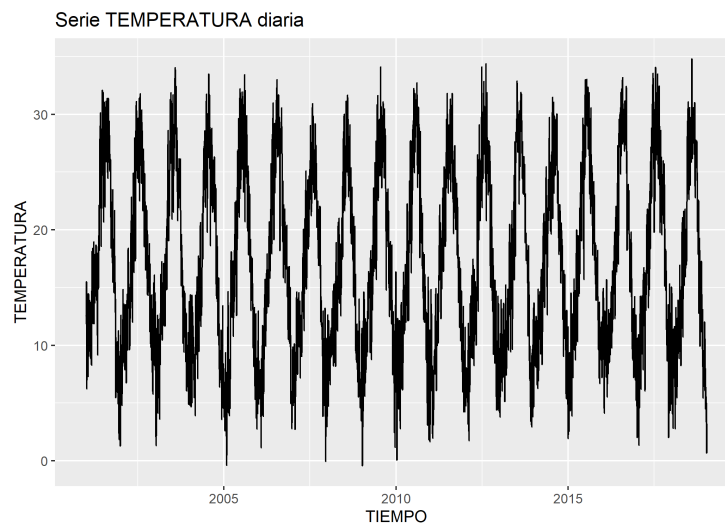


Figura 3.3: Gráfico de la serie diaria Temperatura

En el gráfico de la serie no se puede apreciar tendencia alguna, aunque en la representación

de la tendencia del apartado anterior sí se vea cierta tendencia creciente. Esta hipótesis de existencia de tendencia se comprobará en el siguiente apartado una vez se haya modelizado la serie mediante la metodología ARIMA. La necesidad de incluir, o no, la constante en el supuesto modelo propuesto determinará, o no, la existencia de cierta tendencia en la serie.

Estacionaridad y Modelo ARIMA propuesto

Con motivo de poder estudiar la estacionaridad de la serie se va a representar la función de autocorrelación (ACF).

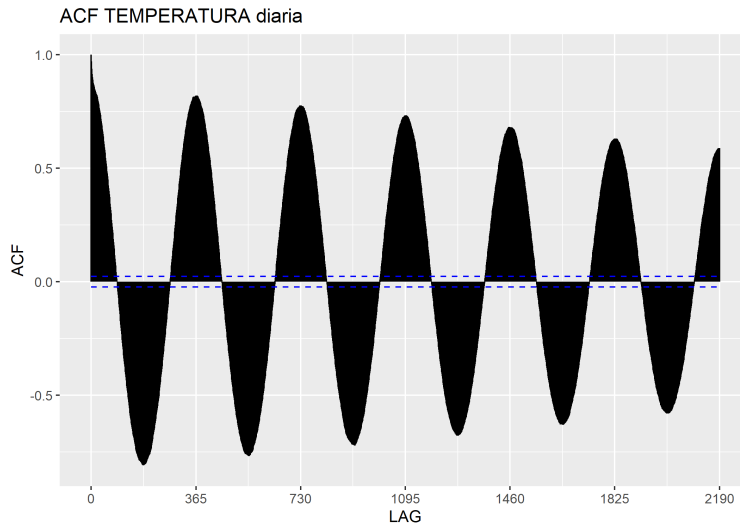


Figura 3.4: ACF de la serie diaria Temperatura

En vista del ACF se puede observar que por la parte regular decrece muy rápidamente por lo que es estacionaria por la parte regular. En cambio, si nos fijamos en los retardos múltiplos de trescientos sesenta y cinco podemos ver que decrecen muy lentamente por tanto por la parte estacional la serie no es estacionaria.

Para solucionarlo se va a diferenciar la serie de forma estacional, $\nabla_{365}x_t = x_t - x_{t-365}$, y posteriormente realizando el test de Dickey–Fuller, que resulta un p-valor significativo, se confirmará que la serie diferenciada es estacionaria por la parte regular y por la parte estacional.

Después de estimar distintos modelos sugeridos observando la ACF y la PACF de la serie diferenciada, y comparando los modelos que mejor se validaban, se propone como modelo $ARIMA(2, 0, 2)(0, 1, 1)_{365}$ sin constante y ecuación $(1 - \phi_1 B - \phi_2 B^2 - B^{365} + \phi_1 B^{366} + \phi_2 B^{367})X_t = (1 - \theta_1 B - \theta_2 B^2 - \Theta B^{365} + \theta_1 \Theta B^{366} + \theta_2 \Theta B^{367})a_t$. Estimando los parámetros con SAS obtenemos la ecuación $X_t = 1,65237X_{t-1} - 0,65802X_{t-2} + X_{t-365} - 1,65237X_{t-366} + 0,65802X_{t-367} + a_t - 0,69991a_{t-1} - 0,26089a_{t-2} - 0,96247a_{t-365} + 0,67364a_{t-366} + 0,25111a_{t-367}$.

Volviendo al apartado anterior, debido a la ausencia de constante en el modelo propuesto, así como en otros posibles modelos alternativos, se concluye la ausencia de tendencia en la temperatura promedio diaria.

Como muestra del grado de validación del modelo se representan los test de Ljung-Box (figura 7.1 del anexo). Para algunos retardos se puede ver que los p-valores están entre 0,01 y 0,05, por lo que la hipótesis de incorrelación de los residuales no se satisface todo lo bien que sería deseable. Para estudiar la distribución de los residuos, se va a tener en cuenta su qqplot (figura 7.2) para comprobar su normalidad, y debido a la ausencia de cambios bruscos en la pendiente de los puntos se va a suponer que los residuales son normales. Para determinar si dicha normal tiene media cero se realizará el contraste $H_0 : \mu = 0$, obteniendo un p-valor de 0.7622 por lo que se concluye que los residuales siguen una normal de media cero.

En cuanto a la capacidad de predicción, se va a utilizar como medida relativa el error porcentual absoluto medio calculado como

$$\%EAM = \sum_{i=1}^p \frac{|\hat{y}_i - y_i|}{p * y_i}$$

donde p corresponde al número de periodos predichos, en este caso solo se va a predecir un periodo por lo que $p = 365$. Realizando una función muy sencilla en R y exportando los valores predichos desde SAS se obtiene un error del 21.90337 %.

3.2. Temperatura Promedio Mensual

A continuación, se va a analizar la serie media mensual. Para dicha transformación se ha utilizado el comando de R `aggregate()`, la función divide los datos en subconjuntos, calcula estadísticas de resumen para cada subconjunto y devuelve el resultado en un grupo por formulario. En este caso se han aprovechado estas características para conseguir la serie mensual promediando en función de las variables mes y año la serie diaria. Por último se ha eliminado la variable día ya que ahora no tiene ningún interés.

Análisis Descriptivo

La variable temperatura promedio mensual oscila entre 4.988 °C y 30.528 °C con una media de 16.361 °C, si se tienen en cuenta las mensualidades se obtienen los siguientes estadísticos.

Mes	Mínimo	Media	Mediana	Máximo	Rango
Enero	4.988	7.925	7.835	10.55	5.56
Febrero	7.326	9.361	9.338	11.23	3.90
Marzo	10.97	12.45	12.55	14.83	3.86
Abril	13.07	15.3	15.11	17.83	4.76
Mayo	16.49	19.51	20.22	21.76	5.27
Junio	21.56	25.18	25.33	28.12	6.56
Julio	25.9	28.04	28.1	30.53	4.63
Agosto	25.52	27.8	28.2	29.73	4.21
Septiembre	21.57	23.4	23.49	26.13	4.56
Octubre	15.99	17.8	17.89	20.3	4.31
Noviembre	8.86	11.53	11.97	13.62	4.76
Diciembre	5.87	8.08	8.1	10.82	4.95

Tabla 3.2: Estadísticos resumen de la serie Temperatura media mensual

Fijando la atención en la tabla anterior destaca la disminución tan importante que presentan los rangos en comparación con los de la serie diaria, esto informa de la estabilidad de las temperaturas medias mensuales independientemente de la existencia de años más o menos calurosos.

Respecto a los valores de los máximos y los mínimos se puede ver reflejados en ellos el avance de las estaciones, es decir, el ciclo estacional de la temperatura.

A continuación se hará una representación gráfica donde se puedan ver reflejados los resultados anteriores.

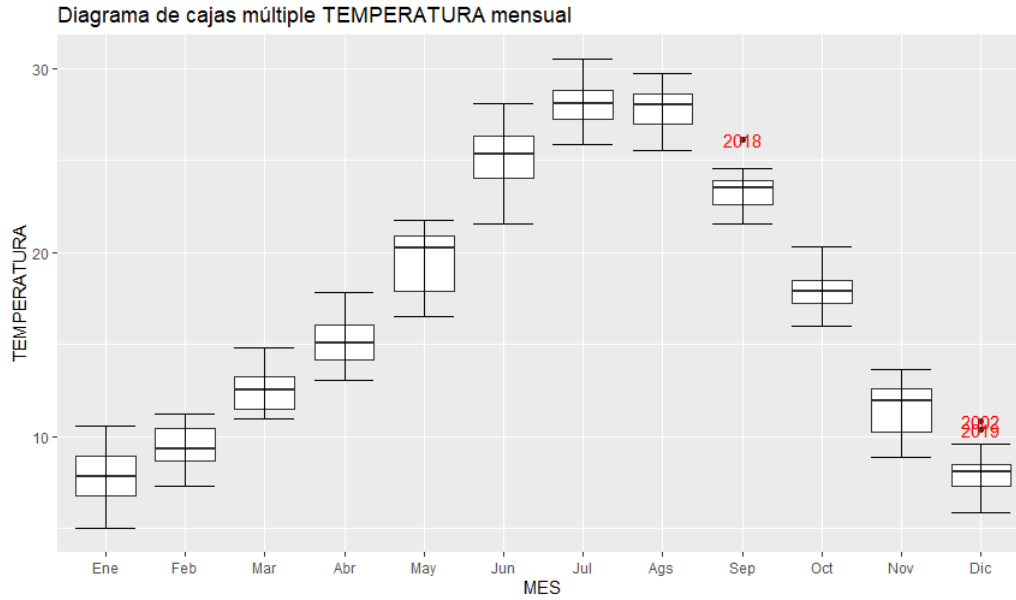


Figura 3.5: Diagrama de cajas múltiple de la serie Temperatura media mensual

En el diagrama se puede ver claramente lo antes comentado, el ciclo estacional de la temperatura. Destacar como desde los primeros meses del año hasta los meses de verano la temperatura aumenta más lentamente pero, en cambio, desde el fin del verano hasta el comienzo del invierno las temperaturas bajan de manera más abrupta.

Respecto a los rangos intercuartiles se puede ver que son más pequeños que los del anterior diagrama, indicando una menor dispersión en las mediciones. Resaltar mayo y junio, donde su rango intercuartílico y sus bigotes respectivamente son especialmente diferentes en comparación con el resto.

Terminando los comentario sobre el diagrama de cajas, ver la escasez de valores atípicos. Únicamente presentes en septiembre y dos en diciembre, correspondientes a los años 2018, 2002 y 2019 respectivamente.

Por último, en este apartado, se realizará la representación de la descomposición de la serie temperatura promedio mensual.

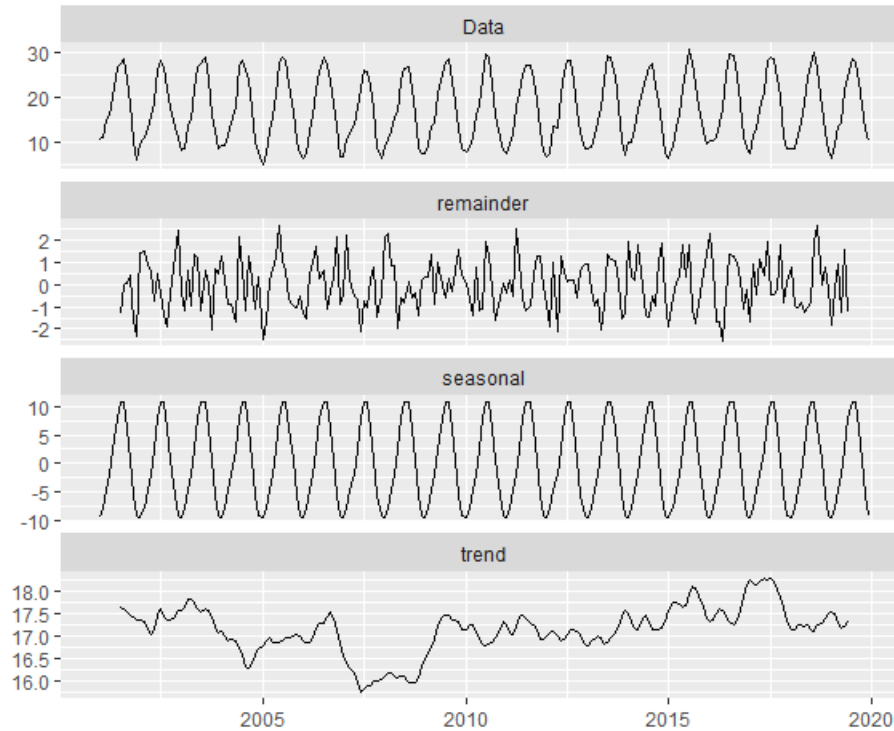


Figura 3.6: Descomposición de la serie Temperatura media mensual

Análisis de Tendencia

Para comenzar el análisis de la serie se va a representar.

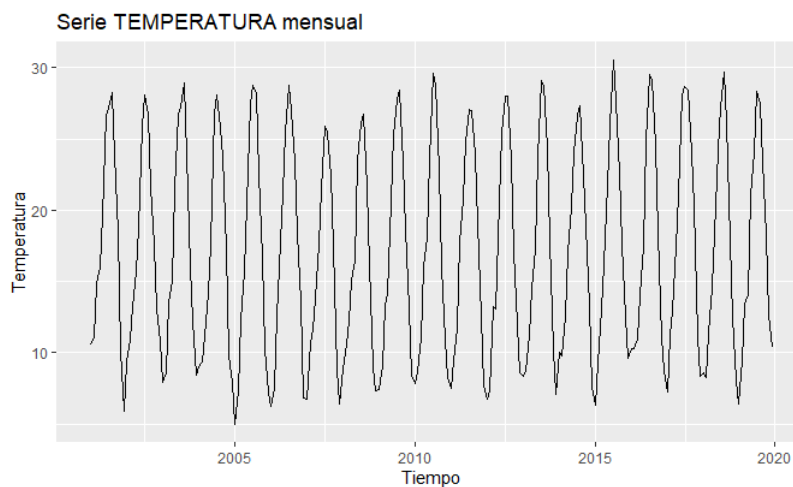


Figura 3.7: Gráfico de la serie Temperatura media mensual

A la vista del gráfico de la serie no hay una tendencia que se vea a simple vista. Fijándose en la tendencia descrita en la descomposición del apartado anterior se puede ver una forma similar a la descrita en la serie diaria por lo que se sospecha una ausencia de tendencia. A

modo de comprobación se va a realizar la variación del test de Mann-Kendall antes descrita, tras su ejecución se obtiene un p-valor de 0.1083 por lo que no se puede rechazar la ausencia de tendencia.

Una vez analizada la tendencia de la serie promedio mensual, se ha apreciado que la tendencia de la parte superior y de la parte inferior del gráfico difieren algo, en concreto, se sospecha que las mediciones máximas no tienen una tendencia evidente, en cambio las mínimas sí que se puede apreciar una leve tendencia creciente por lo que las temperaturas mínimas han podido sufrir un aumento.

Estacionaridad y Modelo ARIMA propuesto

El siguiente punto a observar es la estacionaridad, para ello, primeramente, se va a calcular la función de autocorrelación (ACF).

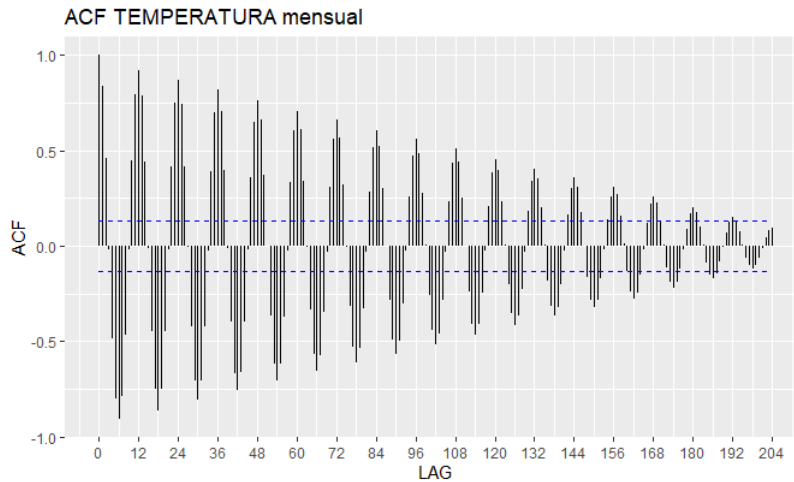


Figura 3.8: ACF de la serie Temperatura media mensual

El ACF de esta serie a simple vista puede parecer diferente al de la temperatura media diaria pero prestando atención se pueden extraer las mismas conclusiones. Por la parte regular los retardos decrecen rápidamente por lo que sí es estacionaria, en cambio, por la parte estacional no es estacionaria, ya que en los retardos múltiplos de doce se puede ver claramente decrecimiento lineal, y no exponencial. Para hacer la serie estacionaria por ambas partes se va a diferenciar estacionalmente como en el caso anterior, $\nabla_{12}x_t = x_t - x_{t-12}$.

Una vez diferenciada la serie se aplica el test de Dickey–Fuller y se obtiene un p-valor significativo por lo que se puede concluir que la serie es estacionaria tanto por la parte regular como por la parte estacional. Estas observaciones también se pueden extraer del ACF de la serie diferenciada donde se ve un decrecimiento rápido de los retardos por ambas partes.

En este caso se propone el modelo $ARIMA(1, 0, 0)(1, 1, 1)_{12}$ sin constante cuya ecuación sería $(1 - \phi B - (1 + \Phi)B^{12} + (\phi + \phi\Phi)B^{13} + \Phi B^{24} - \phi\Phi B^{25})X_t = (1 - \Theta B^{12})a_t$ y estimando sus coeficientes $X_t - 0,39X_{t-1} = 0,8736039X_{t-12} - 0,3888372X_{t-13} + 0,1263961X_{t-24} -$

$0,04915X_{t-25} + a_t + 0,9354965a_{t-12}$. En este modelo, como en el anterior, la constante no es significativa por lo que se tiene otra evidencia de que la serie no tiene tendencia.

Con respecto a la validación del modelo, puede verse el gráfico resumen de los contrastes de Ljung-Box (figura 7.3 del anexo), donde todos los p-valores son notoriamente no significativos por lo que no se rechaza la hipótesis nula y consideramos los residuos como incorrelados.

Seguidamente, se comprobará que los residuales sigan una distribución normal de media cero. Por un lado se hará el test de normalidad de Shapiro-Wilk obteniendo un p-valor igual a 0.5177. Por otro lado se contrastará $H0 : \mu = 0$ obteniéndose un p-valor de 0.8435. Con estos resultados no podemos rechazar que los residuales sigan una normal de media cero y se concluye que el modelo es válido, y tanto las predicciones como las bandas de predicción podrían considerarse fiables.

Por último, con el fin de medir la capacidad de predicción, y reservando dos años completos de datos (24 observaciones), se va a calcular el error porcentual absoluto medio. Esta vez todos los calculos se han realizado con R y con la ayuda del comando `forecast()` del paquete `Forecast` para realizar las predicciones. Finalmente se obtiene un error del 6.718701 %.

Representando los datos reservados junto a las predicciones en dos colores distintos se obtiene el siguiente gráfico.

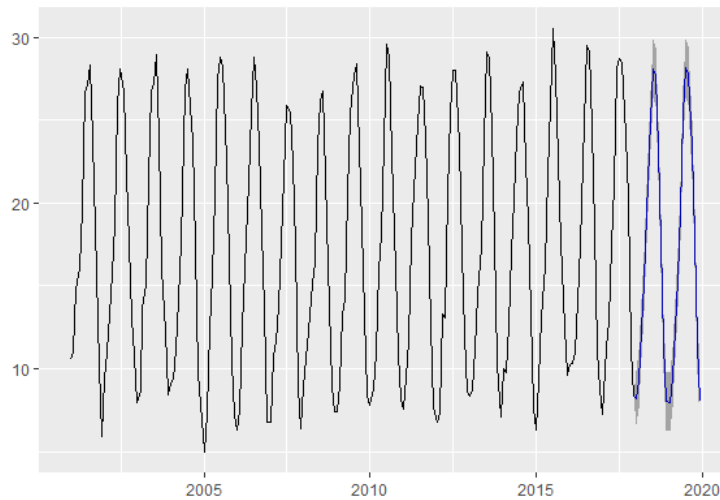


Figura 3.9: Predicciones de la serie Temperatura media mensual

3.3. Temperatura Máxima Mensual

Como se ha dicho en la anterior sección, se van a analizar los extremos mensuales. Primeramente se va a estudiar la serie de máximos mensuales, para ello se han creado los datos utilizando el comando *aggregate()* anteriormente explicado. La diferencia radica en la especificación de la función *max()* como argumento, además de indicar su aplicación respecto de las variables mes y año de la serie diaria.

Análisis Descriptivo

Los límites de esta variable son 9.37 °C y 34.82 °C, además su media toma un valor de 22.30 °C. A continuación, se comenzará con la tabla resumen de los datos creados en función de los meses.

Mes	Mínimo	Media	Mediana	Máximo	Rango
Enero	9.37	13.39	14.40	15.53	6.16
Febrero	11.72	13.74	13.65	15.70	3.98
Marzo	14.58	16.77	16.55	20.01	5.43
Abril	16.58	20.4	20.68	22.86	6.28
Mayo	21.01	25.91	26.42	30.12	9.11
Junio	25.27	31.08	31.64	34.12	8.85
Julio	30.60	32.31	32.14	34.12	3.52
Agosto	29.83	32.00	31.70	38.82	4.99
Septiembre	25.37	28.32	28.23	32.43	7.06
Octubre	19.62	22.99	23.11	25.98	6.36
Noviembre	11.84	16.67	16.79	20.00	8.16
Diciembre	11.27	14.01	13.82	16.37	5.10

Tabla 3.3: Estadísticos resumen de la serie mensual Temperatura máxima

En la anterior tabla se vuelven a ver las mismas características que en la mensual, únicamente destacable las diferencias entre algunos valores de los rangos. Realizando un diagrama de cajas múltiples como en otros casos, dicha variabilidad en el rango supone un aumento en la dispersión de los valores que se verá reflejado en la amplitud de los intervalos intercuartílicos o en la amplitud de los bigotes (figura 7.4 del anexo).

Como viene siendo costumbre, se terminará este apartado realizando la descomposición de la serie.

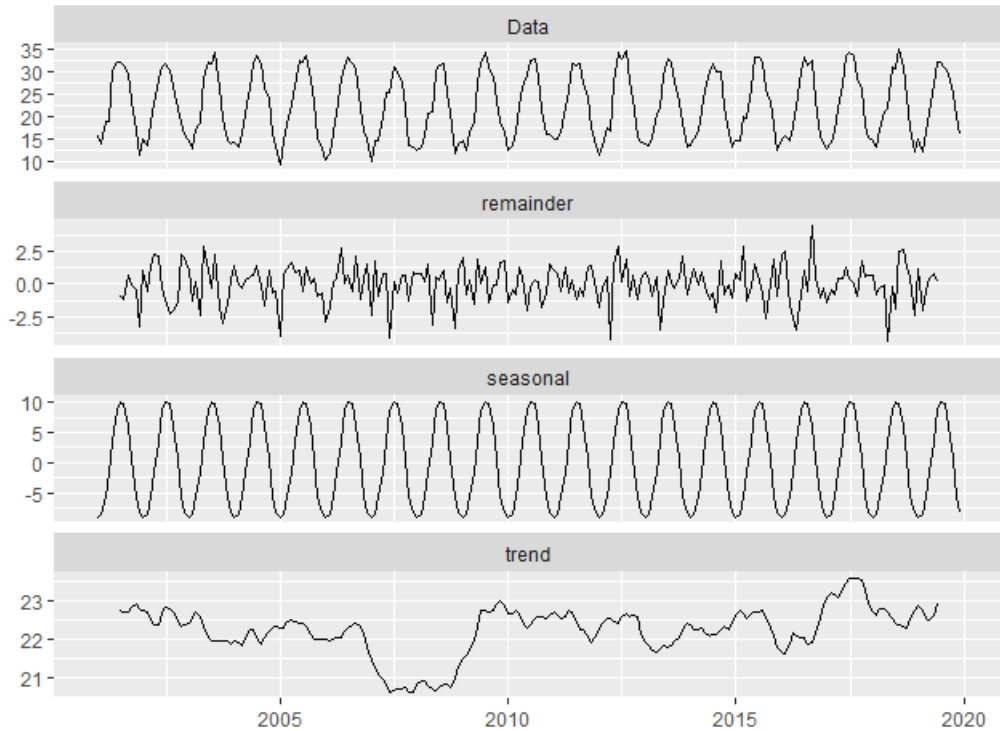


Figura 3.10: Descomposición de la serie mensual Temperatura máxima

Análisis de Tendencia

Para realizar el análisis de tendencia se va a proceder a la representación de la serie en función del tiempo.

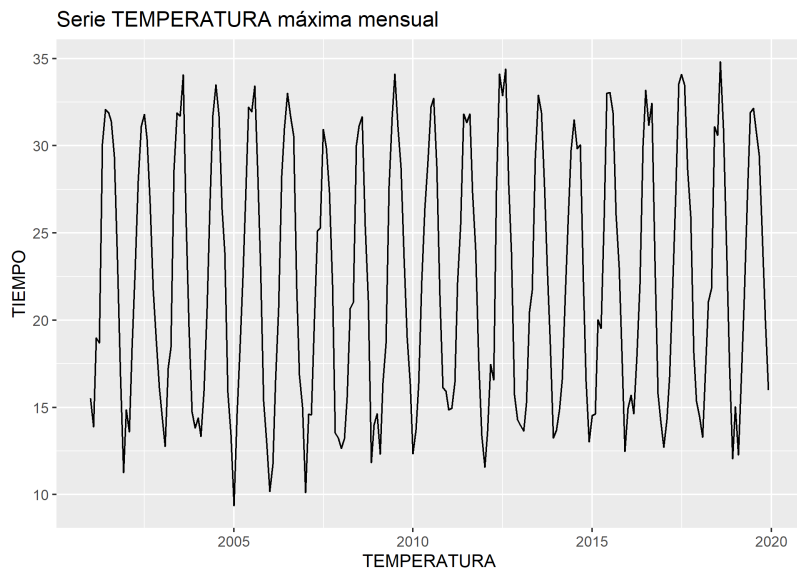


Figura 3.11: Serie Temperatura máxima mensual

En el gráfico de la serie no se aprecia una tendencia general clara. Volviendo a la descomposición del apartado anterior, la descripción de la tendencia vuelve a ser muy parecida a las de otras secciones de este capítulo por lo que probablemente la conclusión sea la misma: no hay tendencia. Realizando la variante del test de Mann-Kendall descrita se obtiene un p-valor de 0.1515 por lo que no se rechaza la hipótesis nula de ausencia de tendencia como se había previsto.

Estacionaridad y Modelo ARIMA propuesto

Representando la función de autocorrelación (figura 7.5 del anexo) se ha podido ver claramente la misma composición que en la ACF de la serie de promedios mensuales, por lo que las conclusiones sobre la estacionaridad de la serie son las mismas. Es estacionaria regularmente pero no estacionalmente.

Habiendo diferenciado por la parte estacionaria la serie se propone el modelo $ARIMA(0, 0, 2)$ $(1, 1, 1)_{12}$ sin constante y con ecuación $(1 - (1 + \Phi)B^{12} + \Phi B^{24})X_t = (1 - \theta_1 B - \theta_2 B^2 - \Theta B^{12} + \phi_1 \Theta B^{13} + \theta_2 \Theta B^{14})a_t$. Estimando los parámetros se obtiene $X_t = -0,1597689t - 12 + 0,1597689X_{t-12} - 24 + a_t - 0,1662011a_{t-1} - 0,1972976a_{t-2} + 0,7910377a_{t-12} - 0,7910377a_{t-13} - 0,5937401a_{t-14}$.

Para validar el modelo se va a comenzar por ver los resultados del test de Ljung-Box (figura 7.6 del anexo), que al obtenerse p-valores no significativos se concluye que los residuales del modelo son incorrelados. Pasando a comprobar la distribución de los residuos se realiza el test de Shapiro-Wilks, al obtenerse un p-valor de 0.00258 se puede rechazar que los residuales sigan una normal. Por último, se comprueba si la media de la normal es igual a cero mediante un contraste que resulta 0.9654 por tanto los residuales no siguen una normal pero su distribución sí que tiene media cero.

El error porcentual absoluto medio cometido por este modelo es del 5.299361 %, y representando las predicciones junto a sus intervalos se obtiene

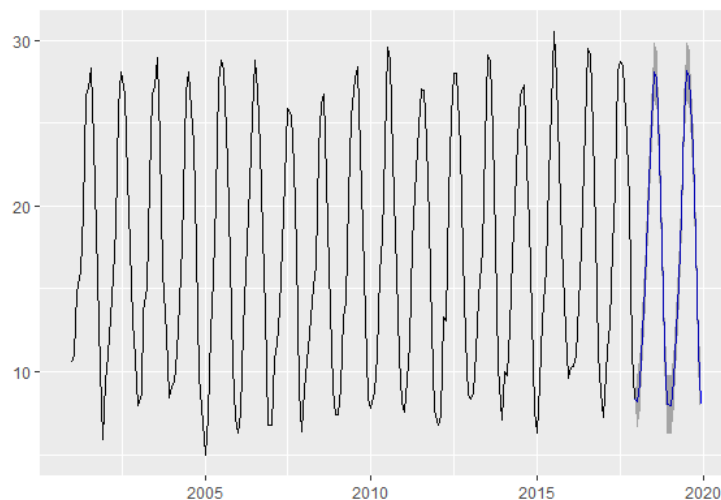


Figura 3.12: Predicciones de la serie Temperatura máxima mensual

3.4. Temperatura Mínima Mensual

Este apartado se centra en estudiar las mediciones mínimas mensuales de temperatura, para conseguir los datos se ha procedido del mismo modo que con las mediciones máximas a excepción de la función especificada como argumento al comando *aggregate()*, en este caso *min()*.

Análisis Descriptivo

A modo de comienzo del análisis descriptivo, se recogen en la tabla algunos estadísticos resumen mensuales. Decir que la variable se encuentra entre -0.42 °C y 26.830 °C, y que su media es 11.876 °C.

Mes	Mínimo	Media	Mediana	Máximo	Rango
Enero	-0.42	2.849	2.74	8.1	8.52
Febrero	1.48	4.748	5.14	8.29	6.81
Marzo	3.91	7.534	8.18	10.46	6.55
Abril	8.24	10.43	10.32	13.34	5.10
Mayo	9.16	13.19	12.59	16.4	7.24
Junio	15.66	18.36	18.35	21.97	6.31
Julio	19.64	22.73	22.56	26.3	6.66
Agosto	18.21	22.36	22.6	26.83	8.62
Septiembre	14.8	18.1	17.87	22.78	7.98
Octubre	9.24	12.53	12.5	15.44	6.20
Noviembre	2.01	6.297	6.38	10.3	8.29
Diciembre	-0.04	3.39	3.46	7.05	7.09

Tabla 3.4: Estadísticos resumen de la serie mensual Temperatura mínima

En la tabla anterior, se pueden ver otra vez las mismas características que en las tablas de estadísticos de la serie promedios mensuales y de máximos mensuales. Únicamente ver como los rangos son más similares entre sí que en la anterior, traducido a un diagrama de cajas (figura 7.7 del anexo) la dispersión en los rangos intercuartílicos y los bigotes serán más parecidos entre sí.

Cerrando este apartado, se representa la descomposición de la serie.

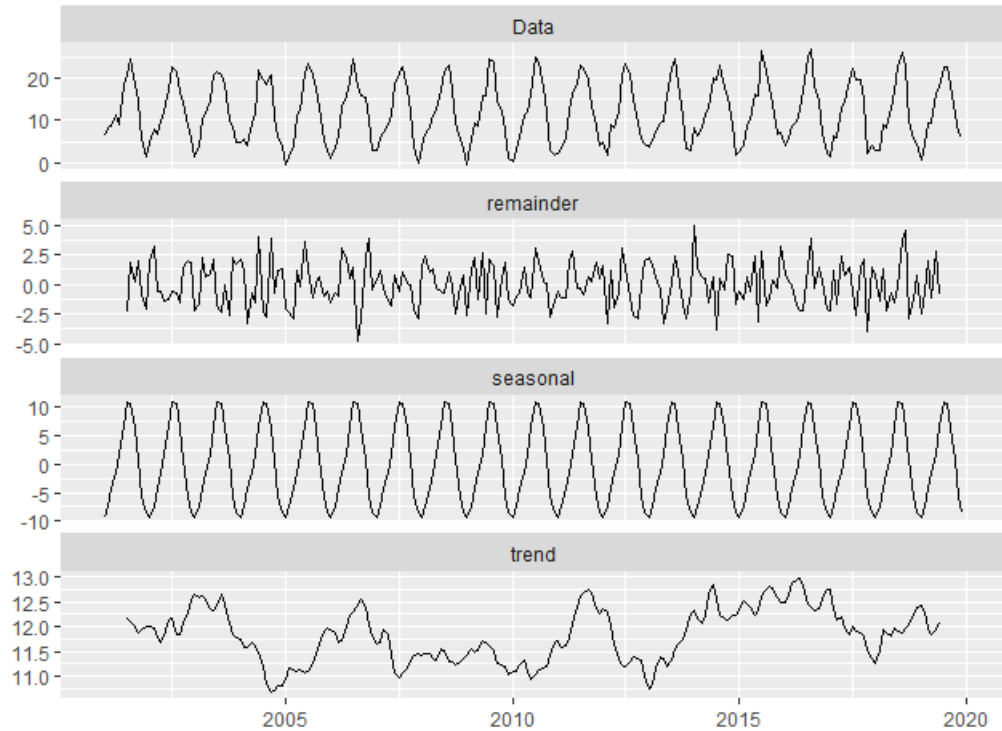


Figura 3.13: Descomposición de la serie mensual Temperatura mínima

Análisis de Tendencia

Seguidamente, se hará la representación de la serie respecto al tiempo de observación de la misma.

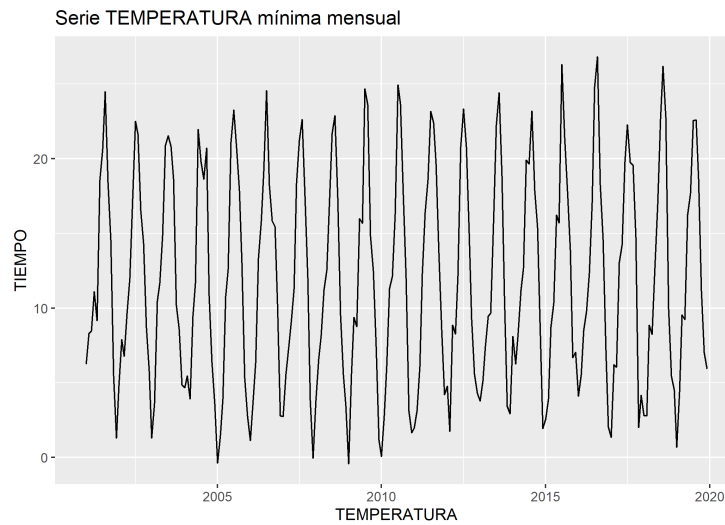


Figura 3.14: Serie Temperatura mínima mensual

A la vista de la representación gráfica de la serie y de la tendencia incluida en la des-

composición se puede ver que no hay una tendencia clara. De una manera subjetiva y sutil se puede apreciar cierto aumento en las temperaturas mínimas a partir de 2010. Realizando nuevamente la variación del test de tendencia de Mann-Kendall se obtiene un p-valor de 0.3919 por lo que el resultado del test se inclina a una ausencia de tendencia.

Estacionaridad y Modelo ARIMA propuesto

En el ACF representado de la serie (figura 7.8 del anexo) se puede volver a ver el patrón visto en las series anteriores por lo que directamente se procede a diferenciar estacionalmente de tal forma que la serie resultante ya es estacionaria tanto regular como estacionalmente según nos confirma el test de Dickey-Fuller con un p-valor inferior a 0,01.

Con la serie diferenciada, y en base al ACF y PACF de esta, se propone el modelo $ARIMA(2, 0, 0)(1, 1, 2)_{12}$ sin constante y ecuación $(1 - \phi_1 B - \phi_2 B^2 - \Phi B^{12} + \phi_1 \Phi B^{13} + \phi_2 \Phi B^{14})X_t = (1 - \Theta_1 B^{12} - \Theta_2 B^{24})a_t$. Estimando los parámetros se obtiene $X_t = 0,085361X_{t-1} - 0,082045X_{t-2} - 0,94619875X_{t-12} + 0,0807684715X_{t-13} - 0,077628511X_{t-14} + a_t - 0,08603912a_{t-12} + 0,91394447a_{t-13}$.

Realizando la validación del modelo, comprobando si sus residuos son incorrelados entre sí mediante el test de Ljung-Box (figura 7.9 del anexo), se ha visto que en los primeros retardos los p-valores están comprendidos entre 0,01 y 0,05 por lo que podría existir algo de correlación entre los residuales, a medida que los retardos avanzan los p-valores dejan de estar cercanos al nivel de significancia lo que indica la incorrelación estos, por tanto el modelo no satisface como debería el supuesto de incorrelación. Comprobando su distribución se realizan, por un lado, el test de normalidad de Shapiro-Wilk y, por otro lado, el contraste para la media. Respectivamente, se obtienen p-valores 0,4036 y 0,924 por lo que se concluye que los residuales siguen una normal de media cero.

Calculando el error porcentual absoluto medio se obtiene un 33.55216 % y representando las predicciones junto a sus intervalos de confianza

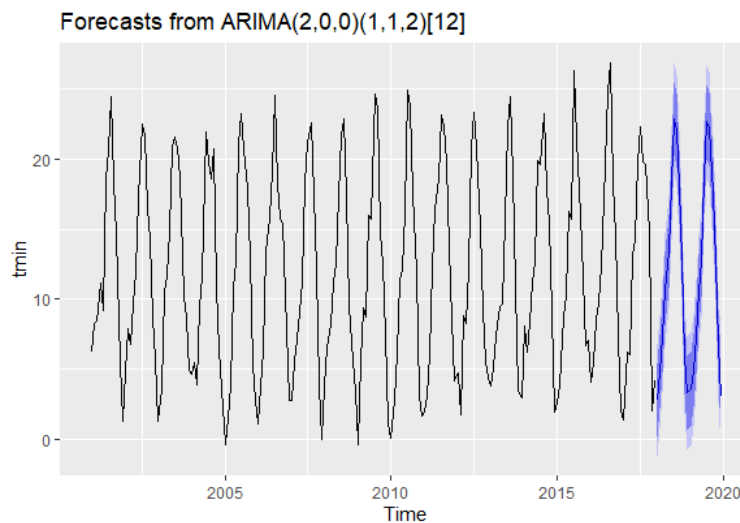


Figura 3.15: Predicciones de la serie Temperatura máxima mensual

Fijando la atención en el anterior gráfico, concretamente en los cambios de pendiente de las predicción en color azul, se pueden ver sombreados los intervalos claramente más grandes que el anterior. Probablemente en esos puntos el modelo no acaba de ajustar bien y se puede explicar el aumento en el error antes calculado.

Capítulo 4

Serie Humedad Relativa

La segunda serie que se va a analizar es la humedad relativa, es decir, el vapor de agua que existe en una masa de aire, expresado como un porcentaje de la cantidad total que existiría si el aire estuviese saturado a esta temperatura según el glosario del IDEAM [8].

En este capítulo, además del promedio diario del que se dispone originalmente, se va a estudiar el promedio mensual de igual forma que en el apartado anterior.

4.1. Humedad Relativa Promedio Diaria

En esta sección, siguiendo la división propuesta en el capítulo anterior, se realizará el estudio de la serie humedad relativa media diaria, como ya se ha dicho.

Mencionar que, debido a errores en las mediciones, los años 2001 y, sobretodo, 2002 no se han tenido en cuenta para modelizar. Es decir, esta serie contará con un intervalo de observación de diecisiete años, de 2003 a 2019.

Análisis Descriptivo

La variable humedad relativa media diaria está comprendida entre el 16.87% y el 100%, su media toma un valor de 62.35%. A continuación, en la siguiente tabla, se recogen algunos de los estadísticos resumen.

Año	Mínimo	Mediana	Media	Máximo	Rango
2003	24.88	66.24	63.99	100	75.12
2004	21.81	66.61	66.24	100	78.19
2005	19.48	60.13	60.53	100	80.52
2006	18.88	65.87	62.24	99.52	80.64
2007	25.21	67.99	67.41	100	74.79
2008	25.2	68.27	66.44	100	74.80
2009	18.14	55.3	60.03	100	81.86
2010	23.35	68.98	66.13	100	76.65
2011	24.5	61.63	60.82	99.31	74.81
2012	16.99	55.47	57.06	96.08	79.09
2013	19.25	63.61	60.81	99.2	79.95
2014	23.3	65.64	62.17	99.38	76.08
2015	16.87	58.19	59.31	99.96	83.09
2016	23.68	63.14	62.03	99.95	76.27
2017	20.41	52.18	54.78	99.57	79.16
2018	19.41	63.14	61.59	98.83	79.42
2019	23.69	57.27	57.25	98.62	74.93

Tabla 4.1: Estadísticos resumen de la serie anual %Humedad

En la anterior tabla, fijando la atención en los máximos, se puede ver como estos toman valores muy homogéneos entre sí. En torno al 99%. Decir que estos valores tan cercanos al total son lógicos, ya que un 100% de humedad relativa supone que ese día ha llovido. Es normal pensar que durante un año al menos un día ha sido lluvioso.

Respecto a los rangos, en ellos se puede ver mayor variabilidad que en los máximos. Destacar que esta variabilidad de los rangos está directamente relacionada con la variabilidad de los mínimos, debido a la similitudes entre los máximos.

A continuación, se va a realizar un diagrama de cajas múltiples que recoge todos los estadísticos antes propuestos.

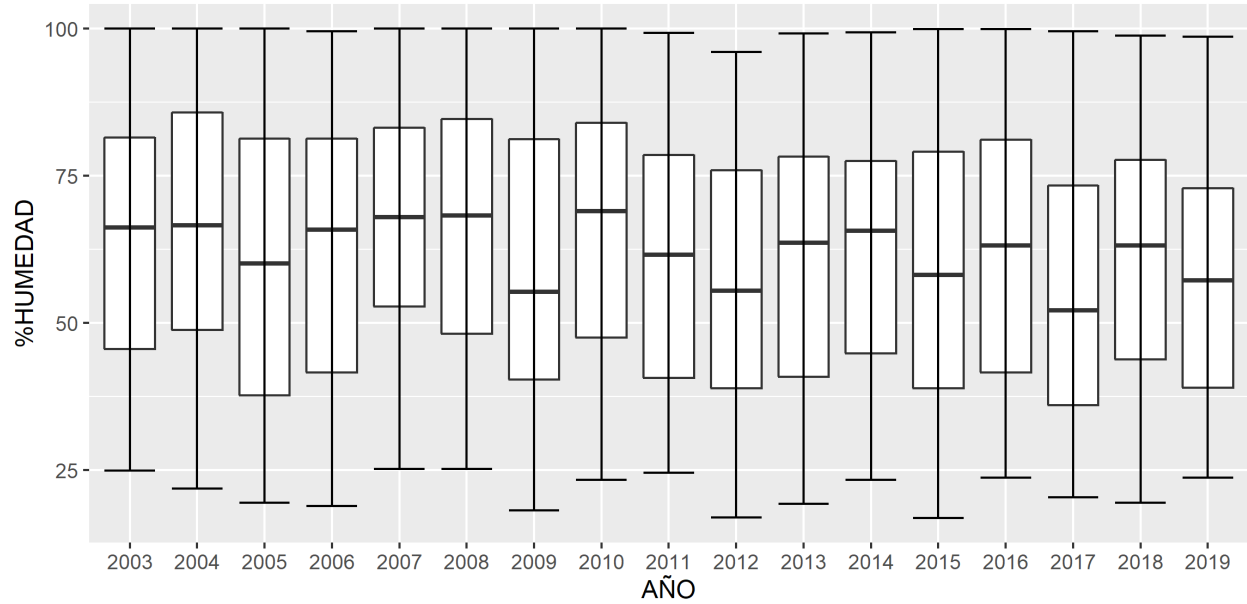


Figura 4.1: Diagrama de cajas múltiple del %Humedad promedio anual

A la vista del diagrama, llama la atención lo comentado anteriormente, las similitudes en los límites superiores y lo cambiantes que son los límites inferiores. Respecto a los máximos, resaltar la excepción del año 2012, que volviendo a la tabla, se puede ver un valor máximo más bajo.

Viendo el diagrama de manera más general, se puede apreciar como los rangos intercuartílicos están dispuestos, ligeramente, en una trayectoria decreciente. Esta observación podría indicar tendencia, más adelante, en esta misma sección, se estudiará con más detenimiento.

Seguidamente, se representará la descomposición de la serie.

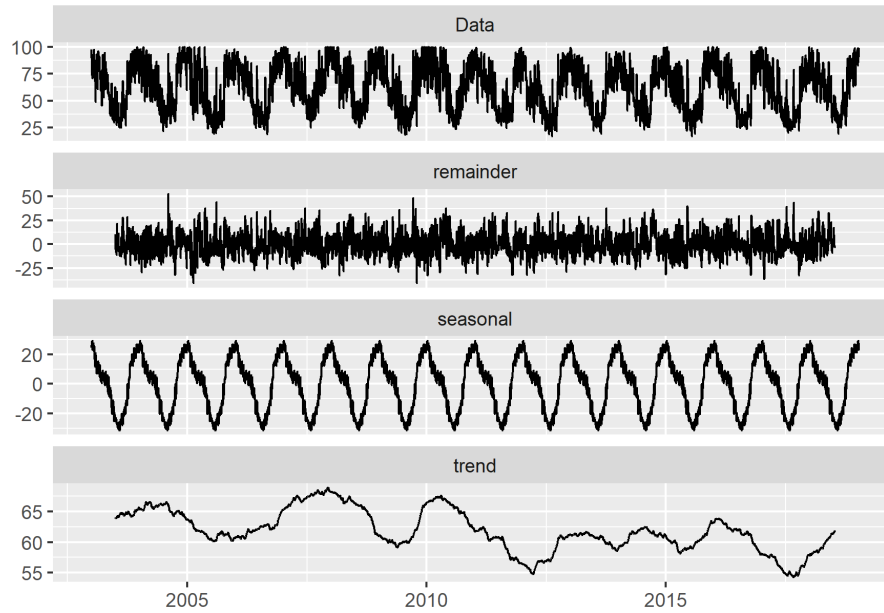


Figura 4.2: Descomposición clásica de la serie diaria % Humedad

Análisis de Tendencia

A continuación, se representará gráficamente la serie.

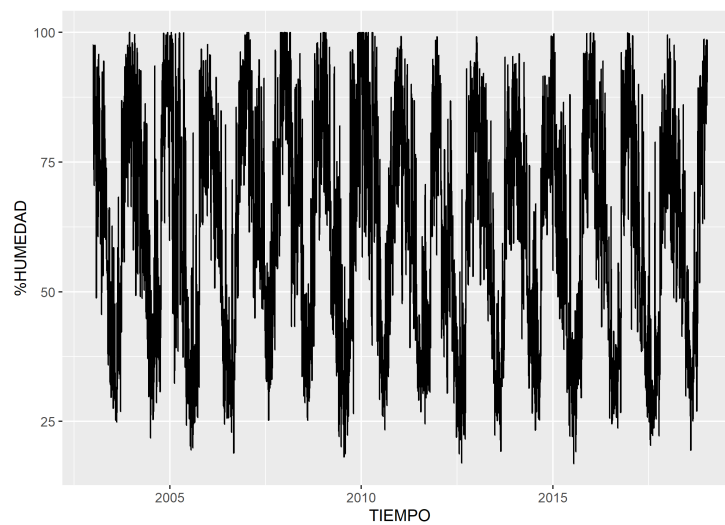


Figura 4.3: Gráfico de la serie diaria %Humedad

En el gráfico de la serie, se aprecia una leve tendencia decreciente. Fijándose en la tendencia que describe la descomposición se ve más claramente ese decrecimiento.

Para encontrar alguna evidencia de esta tendencia vista gráficamente habrá que esperar al siguiente apartado. Como en la anterior serie diaria, la necesidad de incluir, o no, constante

en el modelo nos proporcionará la confirmación, o la rectificación.

Estacionaridad y Modelo ARIMA propuesto

La función de autocorrelación de la serie humedad relativa media diaria (figura 7.10 del anexo) es muy similar al de la serie media diaria antes vista, la temperatura, por lo que se pueden extraer las mismas conclusiones: la serie es estacionaria por la parte regular pero no lo es por la parte estacional.

Para corregir esa falta de estacionaridad estacional se va a diferenciar la serie por dicha parte. Haciendo uso del test de Dickey–Fuller se ha comprobando que la diferenciación sí ha conseguido el objetivo, debido al p-valor resultante muy significativo, se concluye que la serie es estacionaria regular y estacionalmente.

Acto seguido, se propone el modelo $ARIMA(2, 0, 3)(2, 1, 0)_{365}$ sin constante y ecuación $(1 - \phi_1 B - \phi_2 B^2 - (1 + \Phi_1) B^{365} + (\phi_1 + \phi_1 \Phi_1) B^{366} + (\phi_2 + \phi_2 \Phi_1) B^{367} + (\Phi_1 - \Phi_2) B^{730} + (\phi_1 \Phi_2 - \phi_1 \Phi_1) B^{731}) + (\phi_2 \Phi_2 - \phi_2 \Phi_1) B^{732} + \Phi_2 B^{1095} - \phi_1 \Phi_2 B^{1096} - \phi_2 \Phi_2 B^{1097}) X_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t$. Estimando sus parámetros $X_t = 1,68578X_{t-1} - 0,68974X_{t-2} + 0,35441X_{t-365} - 0,59746X_{t-366} + 0,24445X_{t-367} + 0,3076X_{t-730} - 0,51854X_{t-731} + 0,21216X_{732} + 0,33799X_{t-1095} - 0,56977X_{t-1096} - 0,23313X_{t-1097} + a_t - 0,93966a_{t-1} - 0,13318a_{t-2} + 0,09953a_{t-3}$.

Volviendo al apartado anterior, debido a que la constante no es significativa en el modelo se concluye que la hipótesis de ausencia de tendencia vista se puede suponer cierta.

Validando el modelo se ha visto que los test de Ljung-Box (figura 7.11 del anexo) presentan p-valores entre 0,05 y 0,01 por lo que algunos de los residuos pueden tener cierta correlación, por tanto la hipótesis de incorrelación del modelo no satisface como debería. En cuanto a la normalidad de los residuos, el qqplot (figure 7.12 del anexo) no presenta cambios bruscos en la pendiente que describen los puntos, por lo que se van a considerar normales. Junto con el contraste de la media, cuyo p-valor resulta 0,4183, se puede concluir que los residuales son normales de media cero.

Reservando un año de observaciones y prediciendo, el error porcentual medio absoluto asciende al 21.3649 %.

4.2. Humedad Relativa Promedio Mensual

En esta sección se analizará la serie de promedios mensuales de la humedad relativa siguiendo el mismo esquema, y calculada de igual forma, que la serie promedio mensual anterior.

Análisis descriptivo

En este caso, la variable oscila entre 29.87 % y el 93.15 % con un valor medio del 62.4 %. De igual forma, se van a visualizar estadísticos descriptivos, esta vez mensuales.

Mes	Mínimo	Media	Mediana	Máximo	Rango
Enero	78.69	84.32	83.43	93.15	14.46
Febrero	53.03	73.64	74.28	83.79	30.76
Marzo	50.89	66.84	69.13	76.51	25.62
Abril	48.76	65.49	65.31	73.45	24.69
Mayo	43.07	56.38	58.76	76.04	32.97
Junio	35.75	43.59	42.19	56.62	20.87
Julio	29.87	35.06	35.52	39.41	9.54
Agosto	31.08	36.25	35.85	49.72	18.64
Septiembre	36.2	47.20	45.91	63.85	27.65
Octubre	49.06	66.96	69.81	74.71	25.65
Noviembre	73.72	79.74	79.62	88.77	15.05
Diciembre	77.97	85.62	85.81	93.1	15.13

Tabla 4.2: Estadísticos resumen de la serie % Humedad media mensual

En vista de la tabla, los rangos presentan diferencias notables entre ellos, registrándose los valores más bajos en los meses de verano y los más altos en los meses de primavera y otoño. Fijando la atención en los máximos y en los mínimos, se pueden ver los ciclos estacionales que se comentaron en los promedios mensuales anteriores, en este caso invertidos.

A continuación, se hará una representación gráfica donde se vean reflejados los estadísticos anteriores.

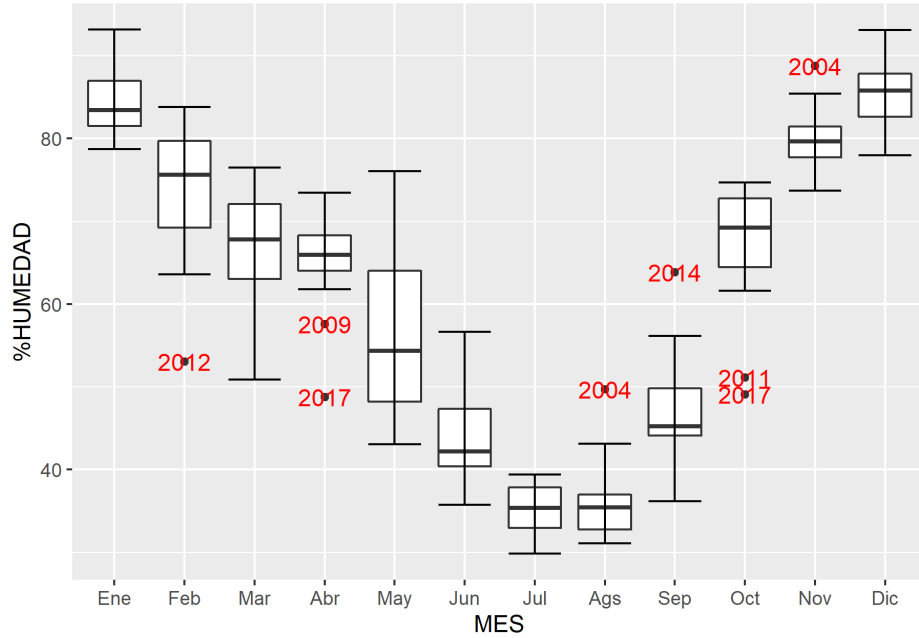


Figura 4.4: Diagrama de cajas múltiple de la serie % Humedad media mensual

Como ya se ha dicho, se ven claramente los ciclos estacionales de la humedad. Dichos ciclos son contrarios a los vistos en los anteriores promedios mensuales, lo que es razonable ya que los meses de verano suelen ser propias las sequías y las lluvias escasas. También cabe destacar los bigotes tan grandes de los meses de marzo y septiembre, y el rango intercuartílico de mayo especialmente amplio.

Por último, destacar la gran cantidad de valores atípicos presentes sobretodo en los meses de primavera y otoño, que volviendo a la tabla, concuerdan con los meses que poseen los rangos más grandes.

Como en otras ocasiones, y a modo de comienzo del análisis de la serie temporal, se va a representar la descomposición de la serie.

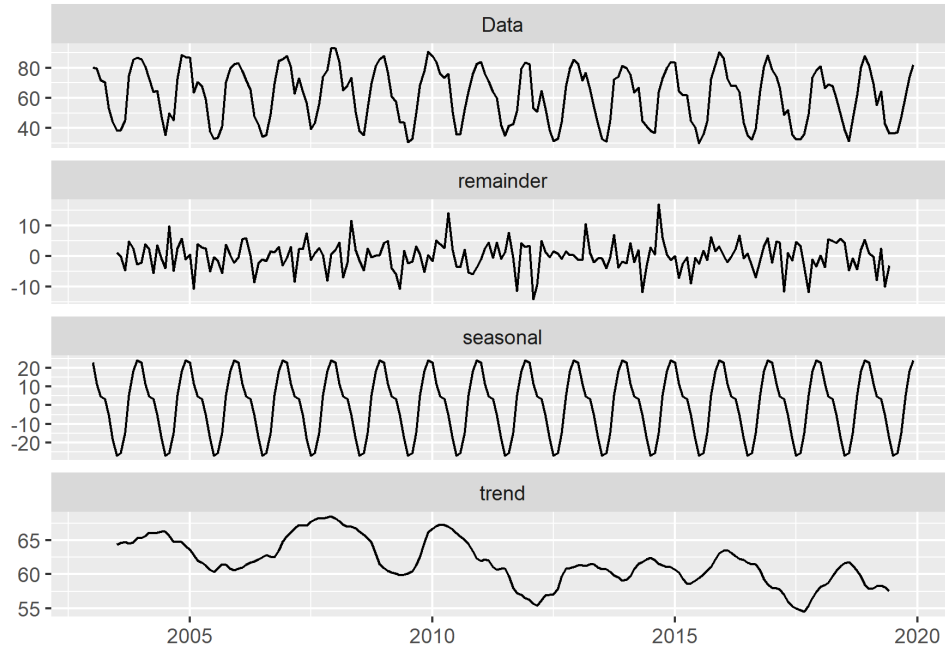


Figura 4.5: Descomposición clásica de la serie % Humedad media mensual

Análisis de Tendencia

La siguiente figura corresponde con la representación gráfica de la serie.

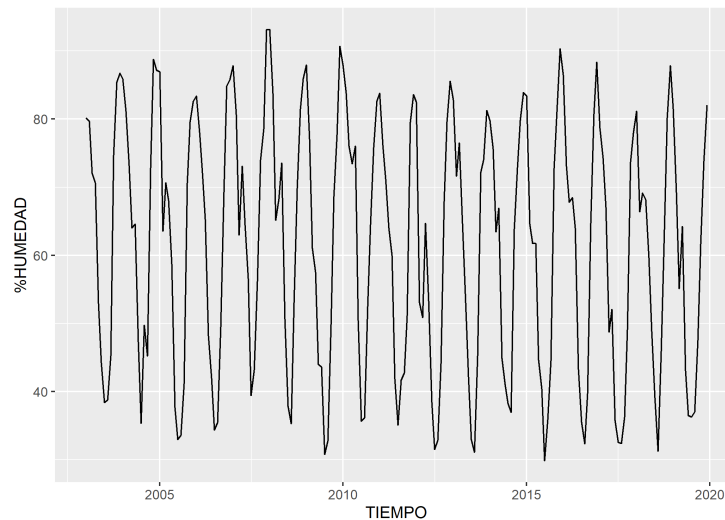


Figura 4.6: Gráfico de la serie % Humedad media mensual

A la vista del gráfico se aprecia una leve tendencia decreciente al igual que en la serie de promedios diarios. Fijándose en el subgráfico tendencia de la descomposición sí que se puede ver más evidente esa tendencia decreciente. A modo de comprobación, se va a volver a realizar la variación estacional del test para tendencia de Mann-Kendall. Trás realizar el test

se obtiene un p-valor aproximadamente cero y un estadístico negativo, por lo que se puede concluir que existe tendencia, y además decreciente.

Estacionaridad y Modelo ARIMA propuesto

Representando el ACF (figura 7.13 del anexo) se puede observar que presenta las mismas características que la ACF de la temperatura promedio mensual. Por tanto, igualmente se concluye que la serie es estacionaria regularmente pero no estacionalmente. Realizando el test de Dickey-Fuller, y habiendo diferenciado por la parte estacional, se afirma la estacionaridad de la serie.

A continuación, debido a que se obtienen resultados parecidos, se proponen dos modelos. El primero de ellos $ARIMA(1, 0, 0)(2, 1, 0)_{12}$ sin constante y con ecuación $(1 - \phi B - (1 + \Phi_1)B^{12} + (\phi + \phi\Phi_1)B^{13} + (\Phi_1 - \Phi_2)B^{24} + (\phi\Phi_2 - \phi\Phi_1)B^{25} + \Phi_2B^{36} - \phi\Phi_2B^{37})X_t = a_t$. Estimando sus parámetros se obtiene la ecuación $X_t = 0,35409X_{t-1} + 0,36205X_{t-12} - 0,1282X_{t-13} + 0,24074X_{t-24} - 0,3665X_{t-25} + 0,39721X_{t-36} - 0,14065X_{t-37} + a_t$. El segundo modelo es $ARIMA(1, 0, 1)(2, 1, 0)_{12}$ también sin constante y con ecuación $(1 - \phi B - (1 + \Phi_1)B^{12} + (\phi + \phi\Phi_1)B^{13} + (\Phi_1 - \Phi_2)B^{24} + (\phi\Phi_2 - \phi\Phi_1)B^{25} + \Phi_2B^{36} - \phi\Phi_2B^{37})X_t = (1 - \theta)a_t$. Estimando sus parámetros $X_t = 0,71013X_{t-1} + 0,36326X_{t-12} - 0,25796X_{t-13} + 0,25168X_{t-24} - 0,17872X_{t-25} + 0,38506X_{t-36} - 0,27344X_{t-37} + a_t - 0,42376a_{t-1}$.

Ambos modelos han sido validados, por lo que sus residuales son incorrelados según los test de Ljung-Box (figuras 7.14 y 7.14 del anexo), y sus bandas de predicción fiables, ya que los residuales se pueden suponer normales de media cero.

Algunos de los criterios que se han tenido en cuenta para comparar los modelos son el AIC (Akaike information criterion) y el SBC (Schwarz's Bayesian Criterion). Se considerará como modelo 1 $ARIMA(1, 0, 0)(2, 1, 0)_{12}$ y como modelo 2 $ARIMA(1, 0, 1)(2, 1, 0)_{12}$.

	AIC	SBC
Modelo 1	1268.244	1278.017
Modelo 2	1266.535	1279.565

Tabla 4.3: Estadísticos para la comparación de modelos

Estos estadísticos dan una visión muy general de cuál de los dos modelos es mejor, además vemos que la diferencia entre ambos es bastante pequeña, por lo que con la intención de seguir comparando los modelos, se estudiará su capacidad de predicción. Para esta tarea, se han recogido en una tabla las predicciones del último año, los valores de la serie en ese año, los errores absolutos y el error cuadrático acumulado (SSE).

Pred1	Pred2	HR	Eabs1	Eabs2	Eqabs1	Eqabs2
83.5838	83.7050	81.5119	2.0718	2.1930	4.293	4.809
71.2830	71.7375	69.2475	2.0355	2.4900	8.436	11.010
68.2208	68.6967	55.1371	13.083	13.5596	179.620	194.873
63.6744	63.9427	64.2253	0.5509	0.2827	179.923	194.953
59.6125	59.8238	43.0674	16.545	16.7563	453.662	475.729
43.4699	43.6820	36.4940	6.9759	7.1880	502.325	527.396
36.0979	36.2709	36.2974	0.1995	0.0265	502.365	527.396
31.9494	32.0742	36.9819	5.0326	4.9077	527.692	551.482
41.0957	41.1995	47.4877	6.3920	6.2881	568.550	591.023
59.8716	59.8306	61.6274	1.7558	1.7968	571.633	594.251
78.8658	78.8716	73.8197	5.0461	5.0519	597.096	619.773
85.6801	85.6515	82.0345	3.6456	3.6170	610.386	632.856

Tabla 4.4: Comparación de modelos ARIMA

En la tabla anterior se pueden ver las predicciones junto al SSE, como se ha dicho antes. Fijándose en los valores que toma el error, se puede ver un claro aumento en la tercera predicción de ambos modelos, obviando esa observación y llegando a la última predicción, los errores han ido aumentando poco a poco y han terminado con una diferencia de veinte puntos. Aunque la diferencia no sea alarmante, el modelo que menor error comete es el 2, $ARIMA(1, 0, 1)(2, 1, 0)_{12}$. Además es el que posee valores más bajos de AIC y SBC, como se ha visto antes.

Todo los resultados anteriores apuntan a que el modelo 2 es el más acertado pero, si se presta atención a las correlaciones de los componentes del modelo, entre la parte de media móvil regular y la parte autorregresiva regular existe un correlación de 0.913, por lo que en ese aspecto podría ser más acertado el modelo 1, $ARIMA(1, 0, 0)(2, 1, 0)_{12}$.

Capítulo 5

Serie Radiación Solar Incidente

La última serie a analizar, siguiendo el esquema propuesto, es la radiación solar incidente sobre la superficie terrestre, medida en decenas de kJ/m^2 .

El dato de partida es la radiación solar que llega a una superficie de 1 m^2 acumulada durante un día, en el punto donde está colocada la estación. A partir de estos datos, se van a analizar los promedios mensuales de la misma además de la serie mencionada de datos de radiación solar acumulada diaria.

5.1. Serie Radiación Solar Incidente Acumulada Diaria

En esta sección se va a analizar de igual forma que en otras ocasiones la serie media diaria radiación solar incidente. En este caso, la serie es algo más larga que las dos anteriores, abarcando seis años más (de 1990 a 2017).

De manera semejante a la humedad, al haber detectado errores en los primeros años de datos, se ha optado por eliminar los datos comprendidos entre 1987 y 1990. A partir de ese año, los datos erróneos puntuales se sustituyeron por el promedio de los laterales.

Análisis Descriptivo

Esta variable oscila, en el periodo que ha sido observada, entre $0 \text{ kJ}/\text{m}^2$ y $14.23 \text{ kJ}/\text{m}^2$, y su media es $6.882 \text{ kJ}/\text{m}^2$. Comenzando la descripción de esta serie, se presenta la siguiente tabla con estadísticos anuales.

Año	Mínimo	Media	Mediana	Máximo	Rango
1990	0.00	6.600	6.710	12.32	12.32
1991	0.00	6.974	7.090	13.05	13.05
1992	0.39	6.725	6.805	12.30	11.91
1993	0.38	6.509	6.340	12.67	12.29
1994	0.65	6.776	6.760	12.76	12.11
1995	0.52	7.032	7.280	12.78	12.26
1996	0.24	6.959	7.000	14.23	13.99
1997	0.16	6.546	6.740	12.83	12.67
1998	0.20	6.693	6.490	12.64	12.44
1999	0.43	6.764	6.380	12.49	12.06
2000	0.24	6.675	6.580	12.61	12.37
2001	0.64	6.780	6.320	12.69	12.05
2002	0.53	6.732	6.390	12.68	12.15
2003	0.55	6.738	6.750	12.50	11.95
2004	0.86	6.730	6.500	12.76	11.90
2005	0.66	6.994	7.060	12.70	12.04
2006	0.78	6.628	6.400	12.74	11.96
2007	0.90	6.602	6.560	12.79	11.89
2008	0.69	6.801	6.805	12.67	11.98
2009	0.91	7.017	7.240	12.69	11.78
2010	0.87	6.943	6.860	13.23	12.36
2011	0.84	7.243	7.250	12.68	11.84
2012	0.91	7.311	6.995	13.08	12.17
2013	0.95	7.236	7.210	13.15	12.20
2014	0.97	7.145	7.080	12.91	11.94
2015	0.81	7.236	7.250	12.84	12.03
2016	0.61	7.063	7.115	12.86	12.25
2017	0.88	7.237	7.180	12.64	11.76

Tabla 5.1: Estadísticos resumen de la serie diaria Radiación

Observando la tabla resumen llama la atención que, de entre las series diarias, la radiación solar tiene las medias y las medianas anuales más similares entre sí.

Comentando los rangos, destacar la similitud en los valores, consecuencia de la similitud entre los valores máximos y los valores mínimos. A lo largo de los casi 30 años de los que se disponen datos las mediciones de radiación no han variado significativamente.

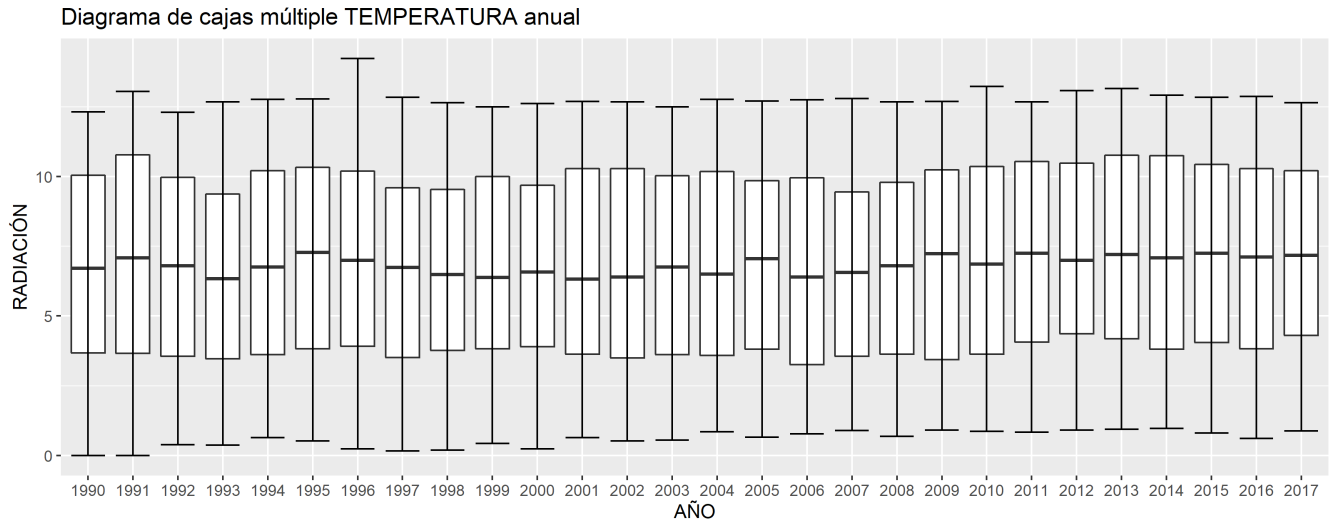


Figura 5.1: Diagrama de cajas múltiple de la Radición solar promedio anual

Pasando a comentar el diagrama, se pueden ver rangos intercuartílicos y bigotes con tamaños muy parecidos. Cabe resaltar el límite superior del año 1996 que, volviendo a la tabla anterior, se ve una medición algo más alta que el resto, rondan los 12 kJ/m² al día. Respecto a los límites inferiores se aprecia una trayectoria creciente en ellos, lo que indica que las mediciones mínimas han ido aumentando con el paso de los años.

A modo de comienzo del análisis de series temporales propiamente dicho se va a representar la descomposición de la misma.

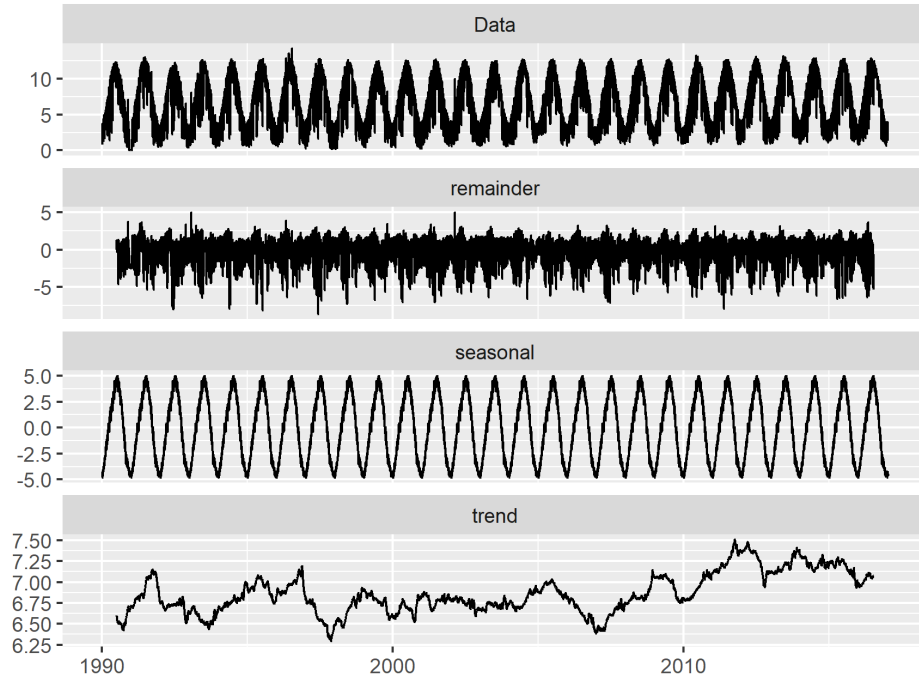


Figura 5.2: Descomposición de la serie diaria Radiación

Análisis de Tendencia

Para estudiar la tendencia de la serie se va a representar.

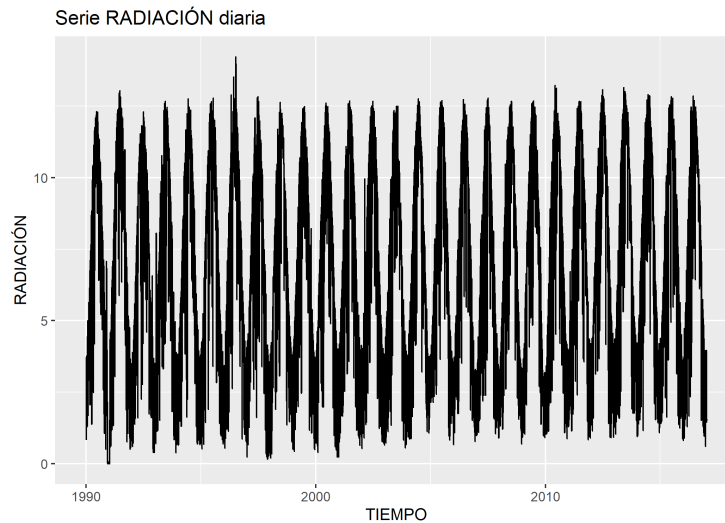


Figura 5.3: Gráfico de la serie Radiación diaria

Observando el gráfico, y tomando de apoyo la tendencia descrita en la descomposición, se puede apreciar que la serie posee tendencia y que esta es creciente. La confirmación de la

existencia de tendencia vendrá dada por la necesidad de incluir la constante en el modelo ARIMA en el siguiente apartado.

Estacionaridad y Modelo ARIMA propuesto

Para estudiar la estacionaridad de la serie, a continuación, se presenta la función de autocorrelación (ACF).

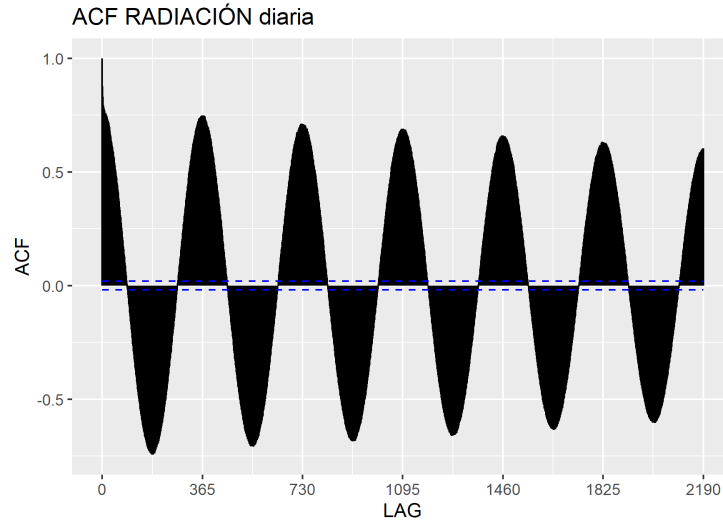


Figura 5.4: ACF de la serie diaria Radiación

Visualizando el gráfico anterior se encuentra la misma estructura de ACF que en las otras dos series diarias por lo que se concluye del mismo modo: la serie es estacionaria regularmente pero no lo es estacionalmente. Teniendo en cuenta lo anterior se procede a diferenciar estacionalmente, aún a sabiendas que de esta forma se ha conseguido que sea estacionaria por ambas partes, se ha realizado el test de Dickey-Fuller que, efectivamente, lo confirma.

A continuación, se propone el modelo $ARIMA(3, 0, 1)(0, 1, 1)_{365}$ con constante y ecuación $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B^{365})(X_t - \mu) = (1 - \theta B)(1 - \Theta B^{365})a_t$. Como se ha visto el modelo propuesto tiene la constante significativa por lo que se confirma la existencia de tendencia del apartado anterior. Estimando los parámetros anteriores $X_t = 1,36039X_{t-1} - 0,34266X_{t-2} - 0,045X_{t-3} + X_{t-365} - 1,36039X_{t-366} + 0,34266X_{t-367} + 0,045X_{t-368} + a_t - 0,82746a_{t-1} - 0,98153a_{t-365} + 0,91033a_{t-366} + 0,000528$.

Puesto que la constante del modelo es significativa y la estimación de la media de la serie diferenciada estacionalmente es igual a 0,01934 se puede decir que la radiación media aumenta 0,01934.

Validando el modelo, interpretando los Test de Ljung-Box (figura 7.16 del anexo) se obtienen p-valores entre 0,05 y 0,01 en los retardos 42 hasta 45 por lo que algunos residuales podrían presentar cierta correlación. Los demás p-valores son no significativos por lo que se supone incorrelación en los residuales, aunque el modelo no sea completamente satisfactorio.

Pasando a comprobar su distribución, el test de Shapiro-Wilk concluye que los residuales no son normales.

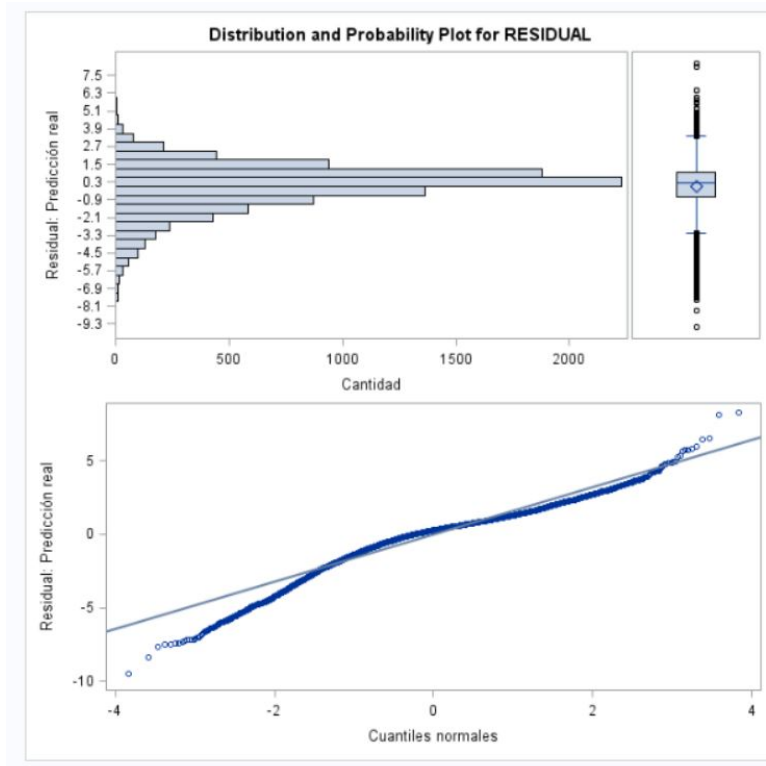


Figura 5.5: Normalidad de los residuales de la serie diaria Radiación

En la anterior figura se puede ver que los residuales del modelo propuesto no son normales. Calculando el coeficiente de asimetría y la curtosis se obtienen los resultados -0.9671088 y 2.383279 , respectivamente. A la vista de dichos resultados se puede decir que la distribución de los residuales es más apuntada que la normal.

5.2. Serie Radiación Solar Incidente Promedio Mensual

La transformación de la serie radiación solar que se va a estudiar, siguiendo el esquema propuesto, es la media mensual. Creada a partir de promedios de la diaria como ya se ha dicho.

Análisis Descriptivo

Esta variable está comprendida entre 0.4084 kJ/m² y 12.1013 kJ/m², con media 6.8722 kJ/m². Seguidamente, se encuentran algunos estadísticos descriptivos mensuales recogidos en una tabla.

Mes	Mínimo	Media	Mediana	Máximo	Rango
Enero	2.045	2.949	2.929	3.811	1.766
Febrero	3.706	4.644	4.609	6.044	2.338
Marzo	4.943	6.501	6.538	7.889	2.946
Abril	5.911	8.225	8.345	9.376	3.465
Mayo	7.973	9.636	9.65	11.45	3.477
Junio	9.561	11.07	11.23	11.97	2.409
Julio	10.60	11.32	11.37	12.10	1.500
Agosto	9.031	9.868	9.934	10.70	1.669
Septiembre	6.586	7.551	7.495	8.66	2.074
Octubre	4.064	4.97	5.007	6.139	2.075
Noviembre	2.232	3.329	3.396	4.117	1.885
Diciembre	0.4084	2.409	2.437	3.154	2.745

Tabla 5.2: Estadísticos resumen de la serie Radiación media mensual

En la anterior tabla se encuentran valores de la media y la mediana todavía más semejantes entre sí que en la serie diaria, por lo que se tiene, si es posible, una distribución todavía más uniforme de los valores. Respecto a los rangos, toman valores bastante pequeños, en torno a 2, a excepción de los meses de primavera algo mayores.

Destacar que se pueden ver a simple vista los ciclos estacionales de la radiación, al igual que ocurría en la temperatura.

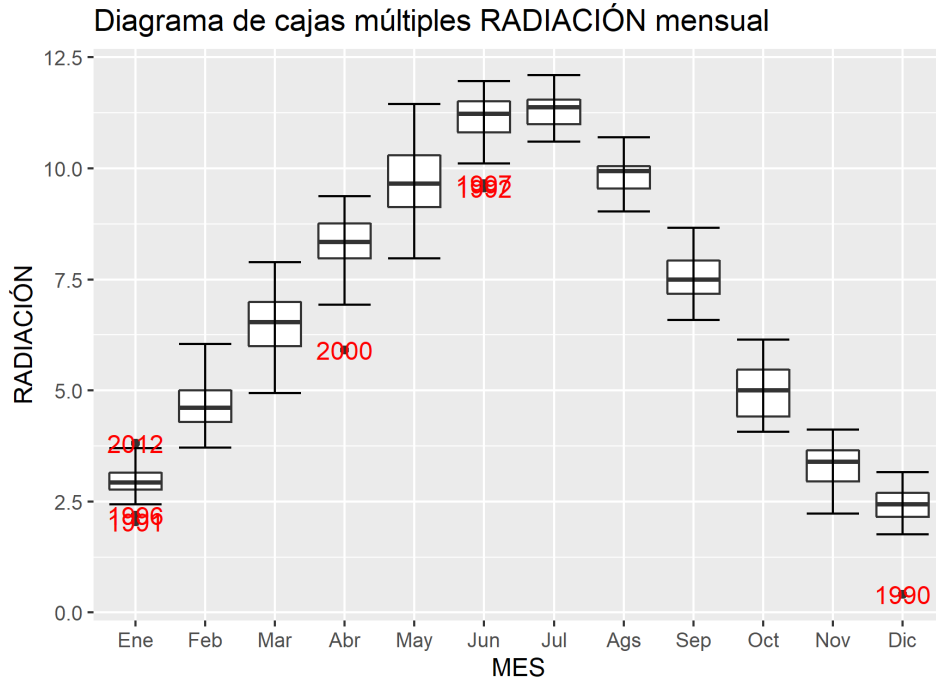


Figura 5.6: Diagrama de cajas múltiple de la serie Radiación media mensual

En cuanto al diagrama, se puede apreciar que los intervalos intercuartílicos y los bigotes tienen más o menos el mismo tamaño de un mes a otro, salvo en los meses de primavera que los bigotes son claramente más grandes, punto ya comentado en las observaciones de la tabla. Por último en el gráfico, encontramos pocos valores atípicos, normalmente en la parte más cercana al límite inferior.

Para finalizar este apartado, y a modo de comienzo del análisis de la serie temporal propiamente dicho, se va a representar la descomposición de ésta.

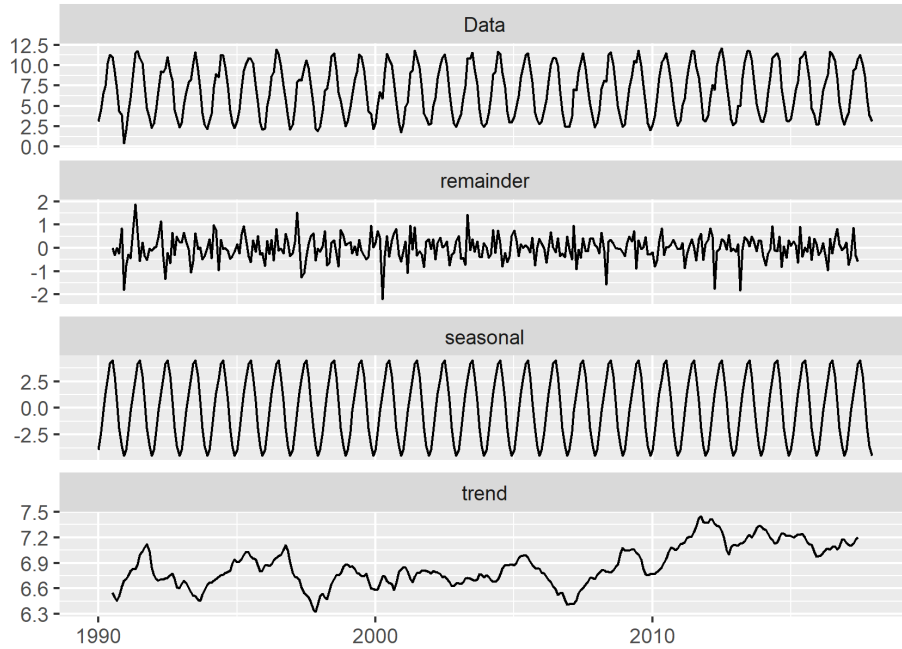


Figura 5.7: Descomposición de la serie Radiación media mensual

Análisis de Tendencia

En el siguiente gráfico se representa la serie mensual radiación.

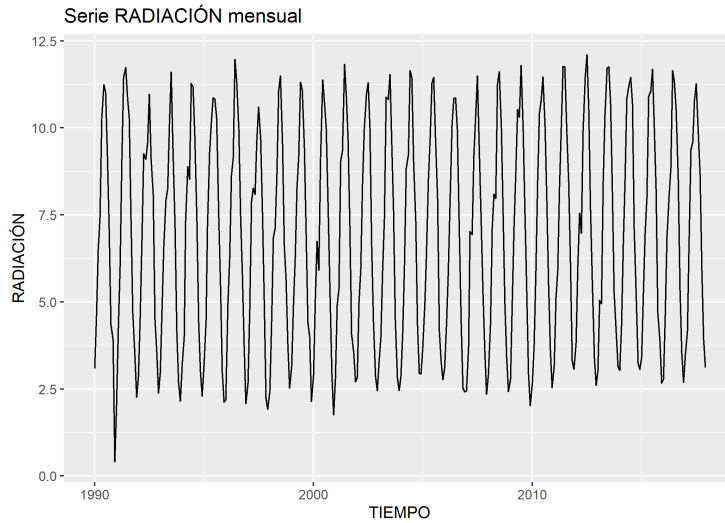


Figura 5.8: Gráfico de la serie Radiación media mensual

A la vista del gráfico de la serie se puede ver una sutil tendencia creciente en la serie, fijando la atención en la tendecia propuesta en la descomposición se puede ver más claramente esa tendencia creciente. Realizando la variante del test de tendencia de Mann-Kendall se

obtiene un p-valor de 7.2915e-08 lo que indica que el contraste sí detecta esa tendencia creciente en la serie.

Estacionaridad y Modelo ARIMA propuesto

Con el propósito de estudiar la estacionaridad de la serie se va a representar la función de autocorrelación.

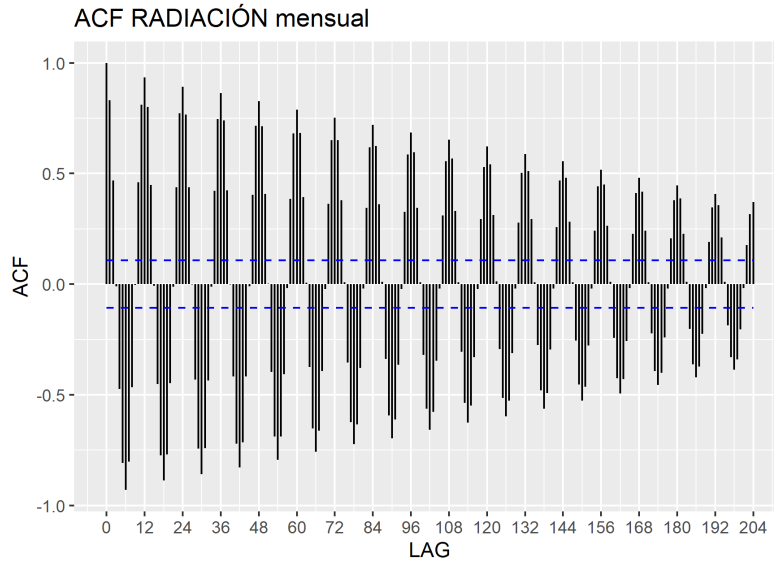


Figura 5.9: ACF de la serie Radiación media mensual

En el ACF de la serie se puede distinguir, debido al rápido decrecimiento de los primeros retardos, que la serie es estacionaria por la parte regular. En cambio, por la parte estacional no, debido al decrecimiento lineal que presentan los retardos múltiplos de doce. Las anteriores características de estacionaridad han sido vistas en repetidas ocasiones y es bien sabido que la solución es la diferenciación estacional consiguiendo así que la serie sea estacionaria por ambas partes, observación comprobada mediante el test de Dickey–Fuller.

Seguidamente se propone el modelo $ARIMA(0, 0, 1)(0, 1, 1)_{12}$ con constante y ecuación $(1 - B^{12})(X_t - \mu) = (1 - \theta B)(1 - \Theta B^{12})a_t$. Estimando los parámetros $X_t = X_{t-12} + a_t + 0,10426a_{t-1} - 0,92828a_{t-12} - 0,0968a_{t-13} + 0,0192$.

Puesto que la constante del modelo es significativa y coincide con la media de la serie diferenciada estacionalmente, por ser un modelo de media móvil sin parte autorregresiva, podemos decir que la radiación aumenta 0.0192 unidades al año. Resultado coherente con el obtenido tras el análisis diario de la serie.

Pasando a la validación, según los test de Ljung-Box (figura 7.17 del anexo) los residuales son incorrelados. Pero, como en el caso anterior, no son normales aunque su media sí se pueda suponer cero.

Capítulo 6

Conclusiones

En este trabajo se ha tratado tanto el estudio de tendencia como la modelización y predicción mediante la metodología ARIMA de variables atmosféricas. Debido al carácter climático de los datos, es de vital importancia hacer hincapié en el análisis de tendencia, con el fin de observar la evolución del cambio climático.

En estas dos últimas décadas de datos, se ha concluido, después de tener en cuenta diferentes escenarios, que la temperatura en España no ha variado significativamente. Según la AEMET [11] sí que hubo un incremento de las temperaturas en los últimos 60 años, en concreto de 1.3 °C, más evidente a partir de 2010. Volviendo a los datos de los que se dispone, probablemente, debido al periodo de observación más reducido, ese incremento no ha sido detectado, aunque en los gráficos de las series sí se puede ver ese crecimiento más fuerte a partir de 2010.

En cambio, en las otras dos mediciones, se ha observado un cambio. La humedad relativa en los últimos años ha disminuido ligeramente, y la radiación solar ha aumentado en contraposición. Decir que el incremento en esta última se relaciona con una menor presencia de nubes, lo que aumenta la radiación directa del sol.

Por último, una posible línea futura consistiría en mejorar los modelos para las series diarias utilizando otro tipo de metodología menos restrictiva que los modelos ARIMA y establecer modelos que permitan relacionar de un modo estocástico la temperatura y la radiación solar.

Capítulo 7

Anexo

Serie Temperatura

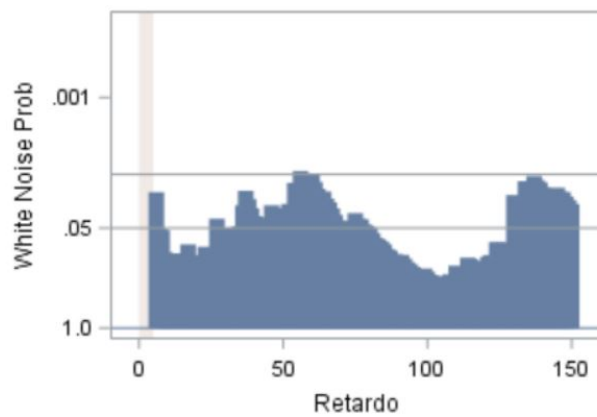


Figura 7.1: Test de Ljung-Box de la serie diaria Temperatura

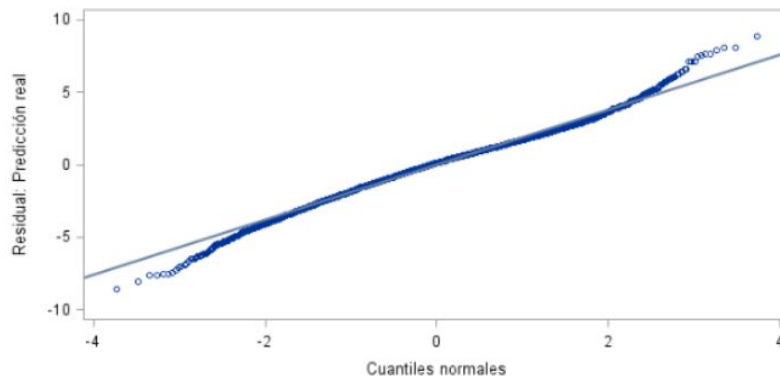


Figura 7.2: Qqplot de la serie diaria Temperatura

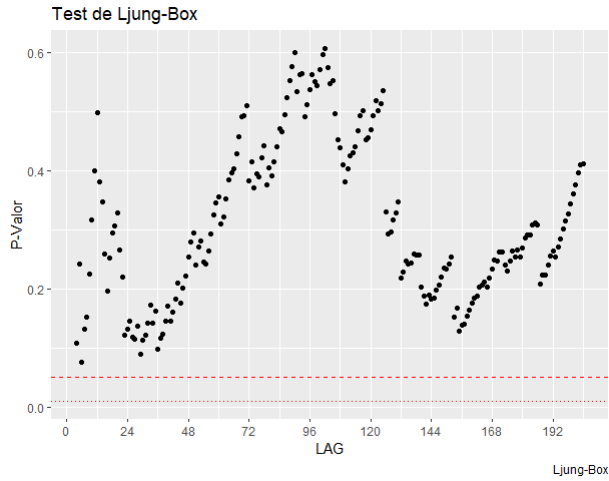


Figura 7.3: Test de Ljung-Box de la serie Temperatura media mensual

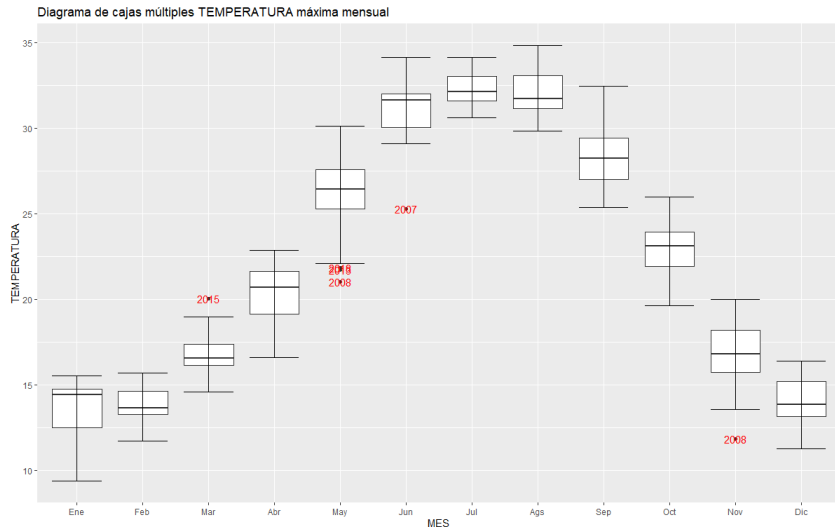


Figura 7.4: Diagrama de cajas múltiple de la serie mensual Temperatura máxima

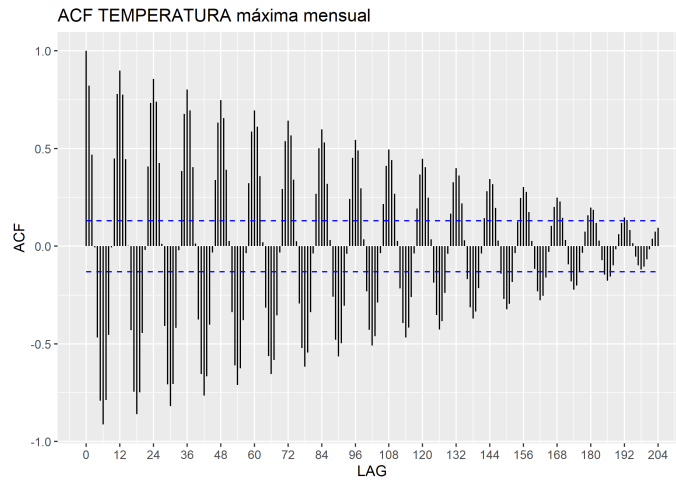


Figura 7.5: ACF de la serie Temperatura máxima mensual

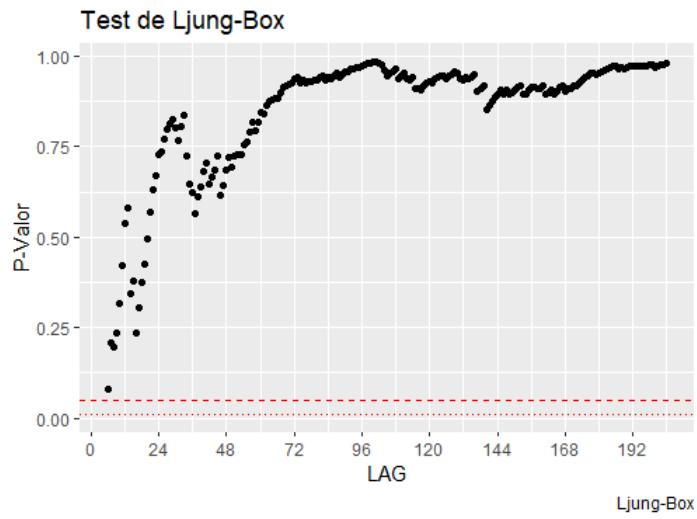


Figura 7.6: Test de Ljung-Box de la serie Temperatura máxima mensual

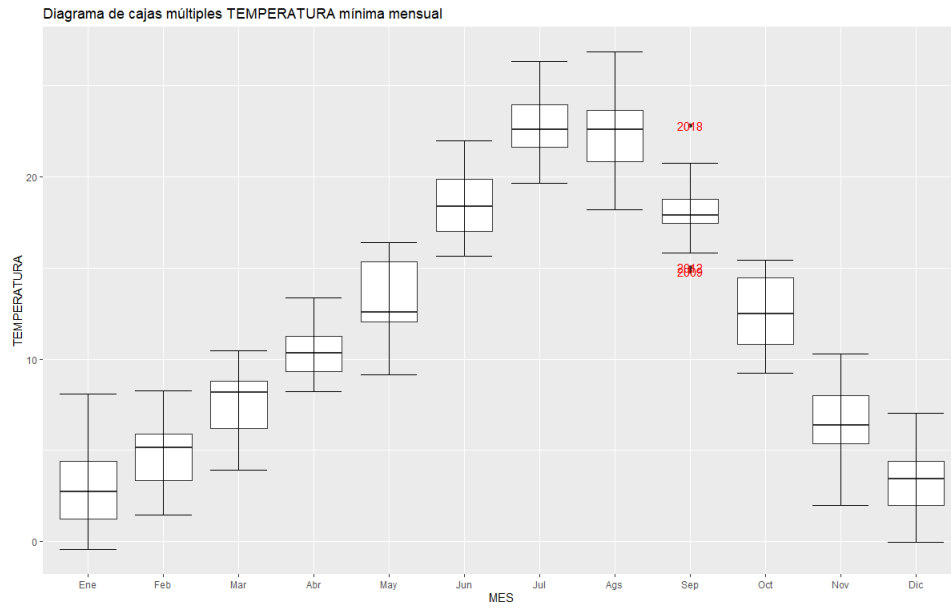


Figura 7.7: Diagrama de cajas múltiple de la serie mensual Temperatura mínima

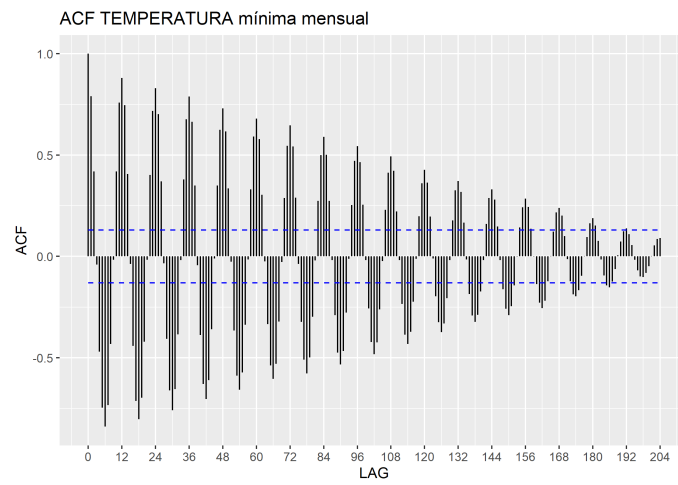


Figura 7.8: ACF de la serie Temperatura mínima mensual

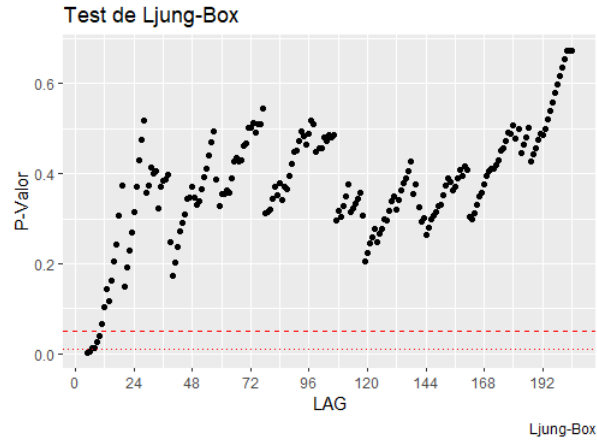


Figura 7.9: Test de Ljung-Box de la serie Temperatura mínima mensual

Serie Humedad Relativa

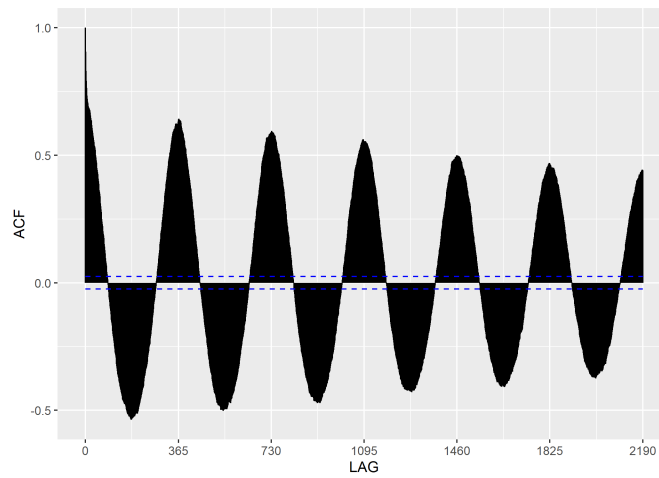


Figura 7.10: ACF de la serie diaria % Humedad

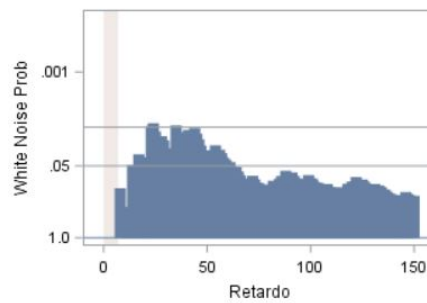


Figura 7.11: Test de Ljung-Box de la serie % Humedad diaria

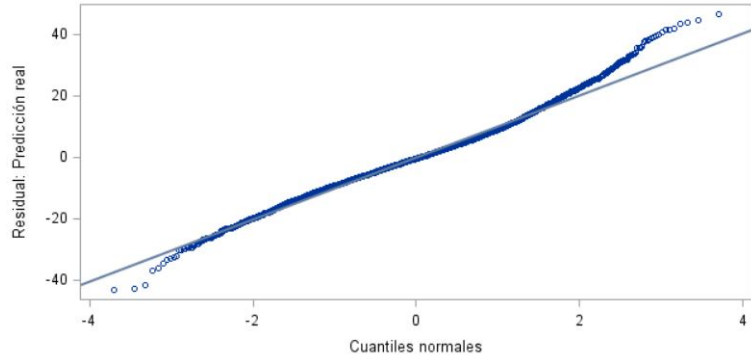


Figura 7.12: Qqplot de la serie % Humedad diaria

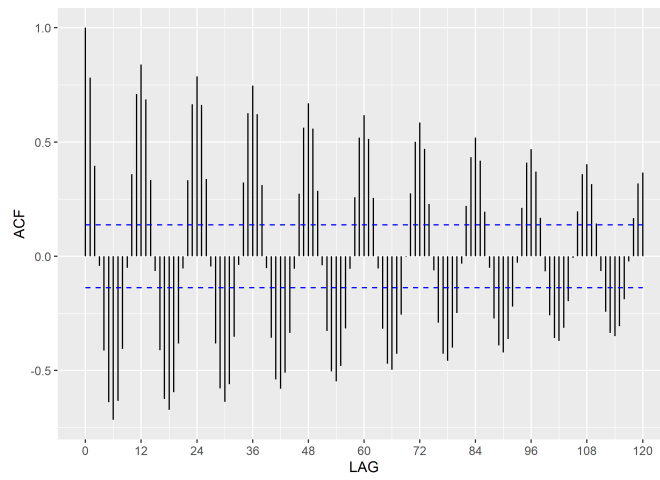


Figura 7.13: ACF de la serie % Humedad media mensual

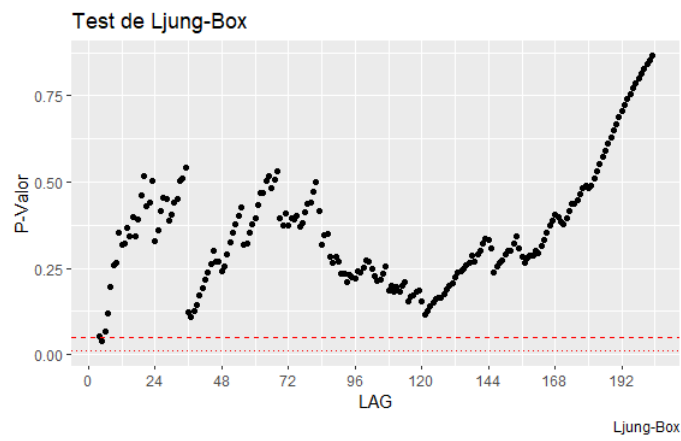


Figura 7.14: Test de Ljung-Box de la serie %Humedad media mensual modelo 1

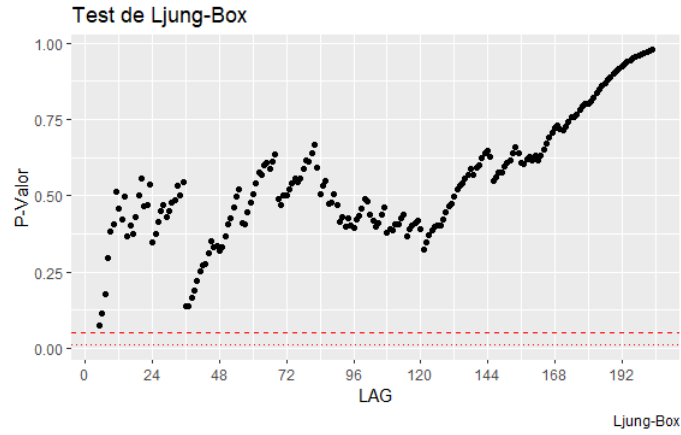


Figura 7.15: Test de Ljung-Box de la serie %Humedad media mensual modelo 2

Serie Radiación Solar Incidente

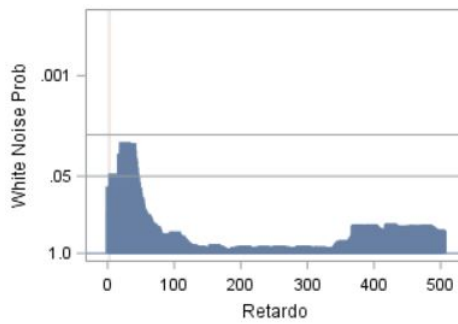


Figura 7.16: Test de Ljung-Box de la serie Radiación diaria

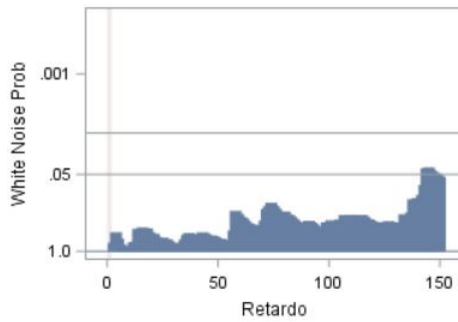


Figura 7.17: Test de Ljung-Box de la serie Radiación media mensual

Bibliografía

- [1] BERGER, A. *Annual and saesonal climatic variations the northern hemisphere and Europe during the last century*. *Annales Geophysicae* , 4 (B4), 385-400. 1986.
- [2] ORGANIZACIÓN METEREOLÓGICA MUNDIAL (OMM). *Guía de instrumentos y métodos de observación*. Volumen V – Garantía de la calidad y gestión de los sistemas de observación, capítulo 2 sección 2.1 pág 31, edición de 2018.
- [3] ORGANIZACIÓN METEREOLÓGICA MUNDIAL (OMM). *Guía de prácticas climatológicas*. Capítulo 5 sección 5 págs 97-98, edición de 2018.
- [4] A.I. MCLEOD. *Kendall rank correlation and Mann-Kendall trend test. Versión 2.2*. URL:<https://cran.r-project.org/web/packages/Kendall/Kendall.pdf>. 16 de mayo, 2011.
- [5] MUÑOZ MARÍN, CARMEN. *Análisis de la variabilidad en series hidrometeorológicas en una cuenca de cabecera del río Segura: Cuenca del Río Mundo*. Departamento de Ingeniería térmica y de fluidos, Universidad Politécnica de Cartagena, Julio 2008.
- [6] RDOCUMENTATION. *arima: ARIMA Modelling of Time Series*. Stats version 3.6.2. URL:<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/arima>.
- [7] ESPINOZA HERNÁNDEZ, VÍCTOR JOSÉ, «Arima-Ljung Box». En: *Blog - Aula Economía*. URL: <https://blogaulaeconomia.wordpress.com/2017/11/09/arima-ljung-box/>. 09 noviembre, 2017.
- [8] INSTITUTO DE HIDROLOGÍA, METEREOLÓGÍA Y ESTUDIOA AMBIENTALES (IDEAM). *Glosario Metereológico*. URL: <http://www.ideam.gov.co>. Bogotá, agosto 2019.
- [9] RPUBS BY RSTUDIO. *Series Temporales*. URL: <https://rpubs.com/palominoM/series>. 26 noviembre, 2017.
- [10] KABACOFF, ROBERT I. *R in action, Data analysis and graphics with R* (segunda edición). Capítulo 15, págs 340-368. Editorial Manning, 2015. ISBN: 9781617291388.
- [11] AEMET. *El informe sobre el estado del clima en 2020 en España muestra que fue el año más cálido y se dispararon los récords de calor*. URL: http://www.aemet.es/es/noticias/2021/05/informe_estado_clima_2020. 14 de mayo, 2021.