

Trabajo_SeriesTemporales

Gabriel Losada Arias

2026-01-12

Contents

1	Análisis de la serie temporal	2
1.1	Descripción básica (tamaño, frecuencia y resumen)	2
1.2	Gráfica global del año	2
1.3	Patrones diarios (perfil medio por hora del día)	3
1.4	Patrones semanales (laborables vs fin de semana)	4
1.5	Estacionariedad (ACF rápida)	6
1.6	Conclusión y reflexión del apartado	7
2	Estudio de outliers	8
2.1	Objetivo	8
2.2	Método de detección de outliers	8
2.3	Cálculo de residuos y detección	8
2.4	Visualización de outliers	9
2.5	Ejemplos de valores atípicos detectados	10
2.6	Comentario e impacto en la predicción	10
2.7	Conclusión y reflexión del apartado	10
3	Metodología de predicción de la serie temporal	11
3.1	Objetivo del apartado	11
3.2	Artículo científico seleccionado	11
3.3	Características relevantes de la serie para la predicción	11
3.4	Modelo de predicción seleccionado	11
3.5	Estructura general del modelo SARIMA	12
3.6	Procedimiento de entrenamiento y evaluación	12
3.7	Consideraciones sobre los valores atípicos	12
3.8	Conclusión del apartado	12

1 Análisis de la serie temporal

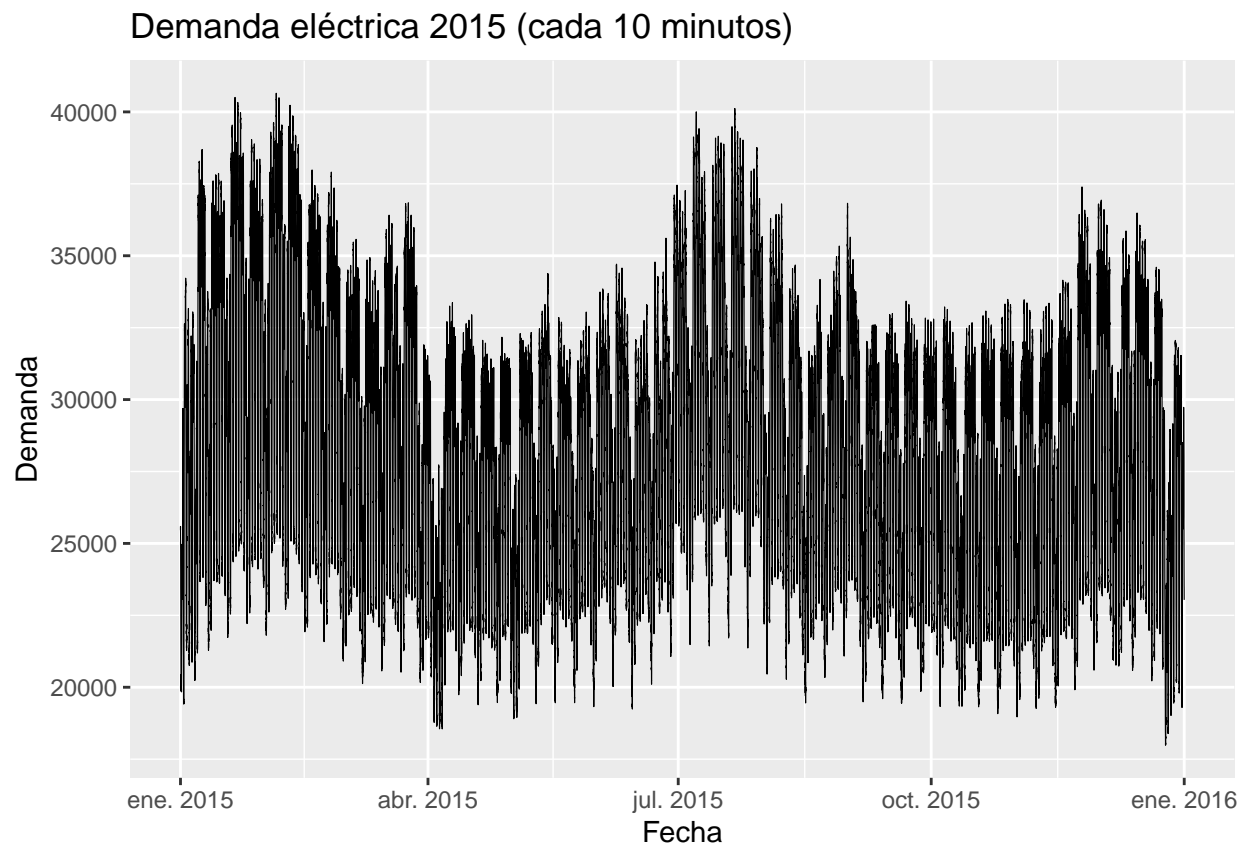
1.1 Descripción básica (tamaño, frecuencia y resumen)

La serie temporal analizada corresponde a la demanda eléctrica registrada cada 10 minutos a lo largo del año 2015. El conjunto de datos consta de 52 560 observaciones, lo que garantiza una cobertura completa del periodo anual con una frecuencia fija y regular.

La variable analizada es cuantitativa continua y representa la demanda eléctrica agregada del sistema, medida en cada instante temporal. Dado el carácter energético de la serie y su alta frecuencia de muestreo, se espera la presencia de patrones periódicos asociados a los ciclos diarios y semanales de actividad.

1.2 Gráfica global del año

```
ggplot(df, aes(datetime, y)) +  
  geom_line(linewidth = 0.2) +  
  labs(  
    title = "Demanda eléctrica 2015 (cada 10 minutos)",  
    x = "Fecha",  
    y = "Demanda"  
  )
```



En la Figura se representa la evolución completa de la demanda eléctrica a lo largo del año 2015. Se observa un comportamiento claramente no estacionario, con oscilaciones regulares de alta frecuencia superpuestas a variaciones de mayor escala temporal.

La serie presenta una estructura altamente periódica, lo que sugiere la existencia de patrones repetitivos asociados a la actividad diaria y semanal. Asimismo, se aprecia una variabilidad significativa en los niveles de demanda, característica habitual en series energéticas reales.

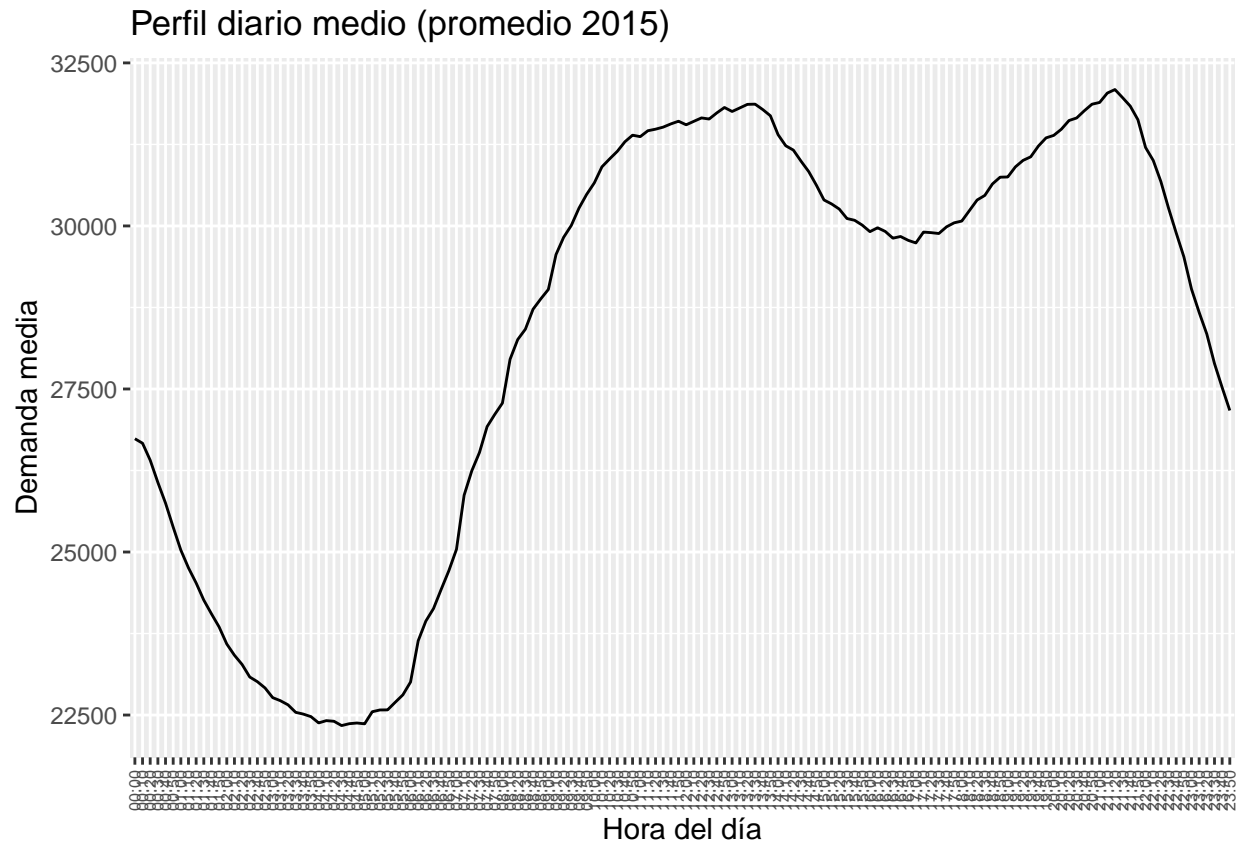
1.3 Patrones diarios (perfil medio por hora del día)

Este bloque identifica la **estacionalidad diaria** típica del consumo.

```
df_aux <- df %>%
  mutate(
    hora = hour(datetime),
    minuto = minute(datetime),
    hm = sprintf("%02d:%02d", hora, minuto)
  )

perfil_diario <- df_aux %>%
  group_by(hm) %>%
  summarise(
    demanda_media = mean(y),
    demanda_sd = sd(y),
    .groups = "drop"
  )

ggplot(perfil_diario, aes(hm, demanda_media, group = 1)) +
  geom_line() +
  labs(
    title = "Perfil diario medio (promedio 2015)",
    x = "Hora del día",
    y = "Demanda media"
  ) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 6))
```



El perfil medio diario de la demanda eléctrica pone de manifiesto una estacionalidad diaria muy marcada. Durante la madrugada se registran los valores mínimos de consumo, mientras que la demanda aumenta progresivamente a partir de primeras horas de la mañana, alcanzando sus valores máximos durante el día.

Este comportamiento refleja los ciclos habituales de actividad económica y doméstica, y confirma la presencia de una componente estacional diaria dominante en la serie temporal.

1.4 Patrones semanales (laborables vs fin de semana)

```
df_aux <- df_aux %>%
  mutate(
    dow = wday(datetime, label = TRUE, week_start = 1),
    is_weekend = dow %in% c("Sat", "Sun")
  )

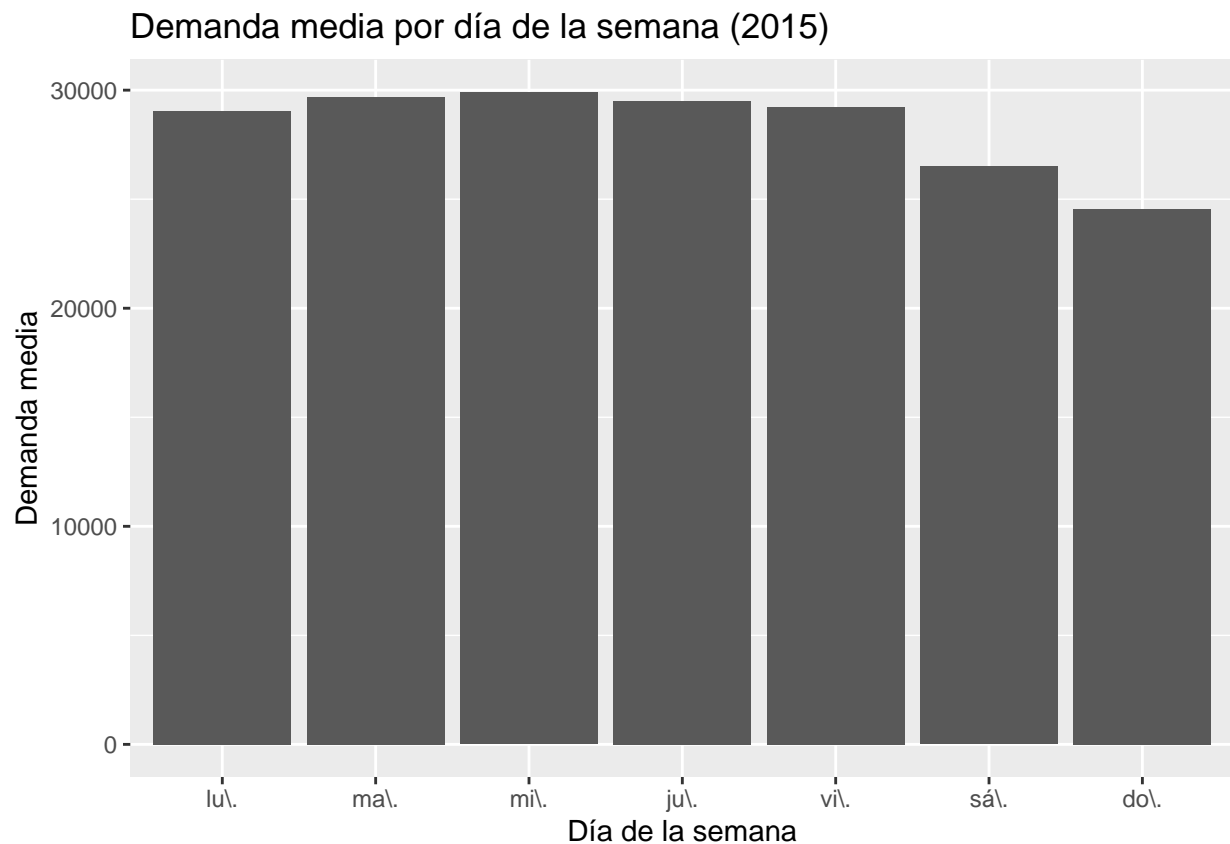
perfil_semanal <- df_aux %>%
  group_by(dow) %>%
  summarise(
    demanda_media = mean(y),
    demanda_sd = sd(y),
    .groups = "drop"
  )

perfil_semanal
```

```
## # A tibble: 7 x 3
##   dow      demanda_media demanda_sd
##   <ord>          <dbl>      <dbl>
## 1 "lu\\."        29029.        5069.
## 2 "ma\\."        29697.        4606.
## 3 "mi\\."        29897.        4527.
## 4 "ju\\."        29518.        4483.
## 5 "vi\\."        29222.        4492.
## 6 "sá\\."        26494.        3221.
## 7 "do\\."        24561.        3184.
```

Gráfica:

```
ggplot(perfil_semanal, aes(dow, demanda_media)) +
  geom_col() +
  labs(
    title = "Demanda media por día de la semana (2015)",
    x = "Día de la semana",
    y = "Demanda media"
  )
```



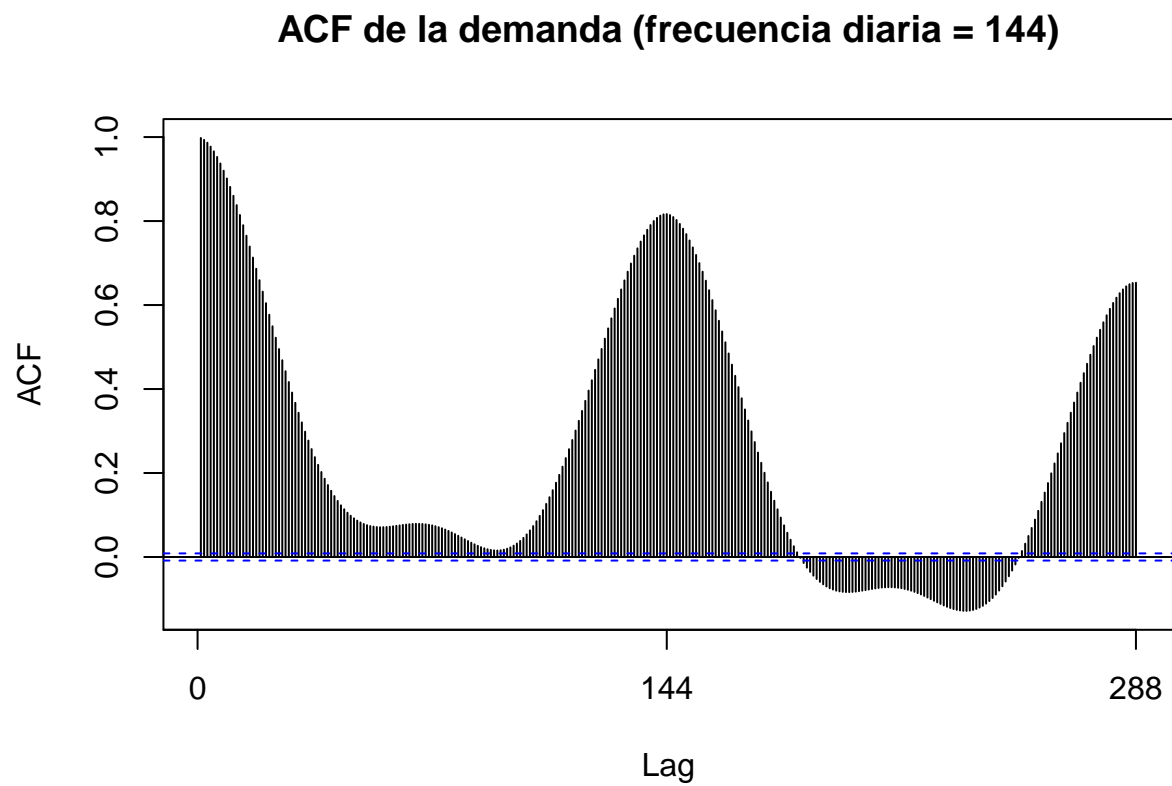
Al analizar el perfil diario de la demanda según el día de la semana, se observan diferencias claras entre días laborables y fines de semana. Durante los días laborables la demanda es, en general, más elevada y presenta picos más pronunciados, mientras que los fines de semana muestran un consumo más reducido y homogéneo.

Estas diferencias confirman la presencia de una estacionalidad semanal adicional, que debe ser tomada en cuenta en cualquier modelo de predicción.

1.5 Estacionariedad (ACF rápida)

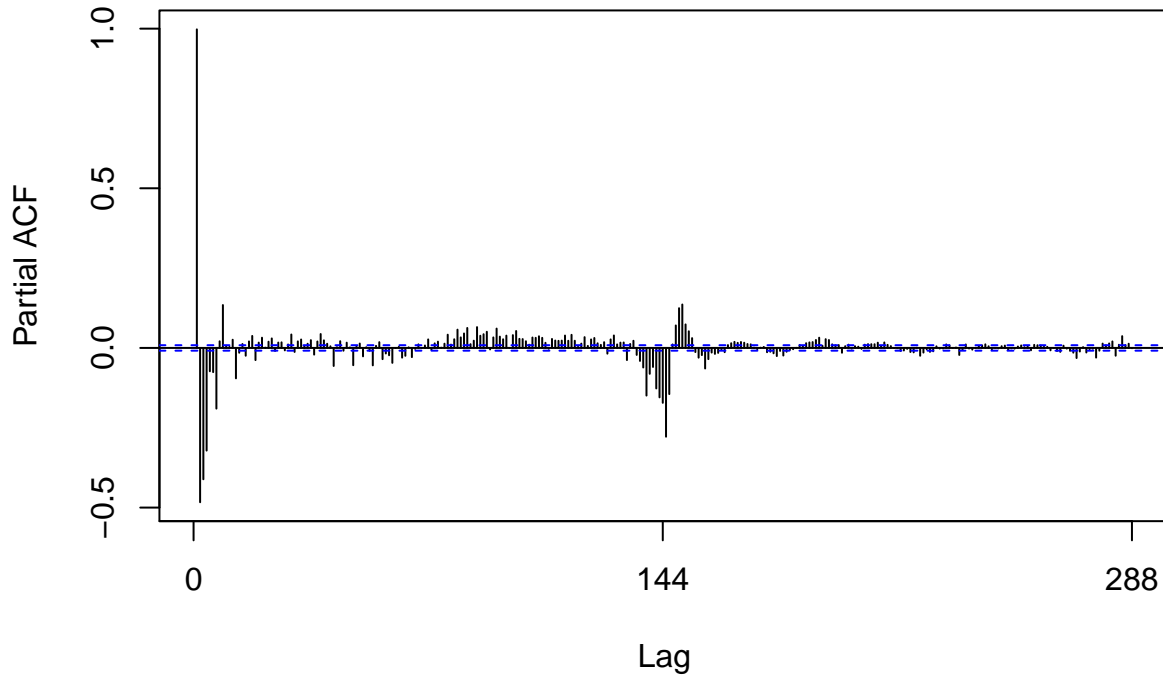
Para la ACF necesitamos un `ts`. Como hay estacionalidad diaria, usamos `frequency = 144`.

```
y_ts <- ts(df$y, frequency = 144)
Acf(y_ts, main = "ACF de la demanda (frecuencia diaria = 144)")
```



```
Pacf(y_ts, main = "PACF de la demanda (frecuencia diaria = 144)")
```

PACF de la demanda (frecuencia diaria = 144)



Las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) muestran una dependencia temporal significativa entre observaciones cercanas en el tiempo. La autocorrelación decae de forma gradual, lo que indica que los valores actuales de la serie están fuertemente influenciados por valores pasados.

Este comportamiento es característico de series temporales energéticas y justifica el uso de modelos de predicción basados en dependencias autorregresivas.

1.6 Conclusión y reflexión del apartado

La serie corresponde a demanda eléctrica registrada cada 10 minutos durante 2015 (52.560 observaciones), lo que garantiza una cobertura completa del periodo anual y permite observar con claridad patrones temporales de distinta escala. El análisis descriptivo confirma una estacionalidad diaria dominante (mínimos nocturnos y máximos diurnos), y una estacionalidad semanal adicional, con un comportamiento sistemáticamente diferente entre días laborables y fines de semana. Estos resultados no solo describen la serie, sino que anticipan el tipo de modelo que puede funcionar: cualquier método que ignore estacionalidad o dependencia temporal tendrá dificultades para capturar la dinámica real.

Desde una perspectiva de modelado, la ACF/PACF evidencian una autocorrelación elevada en retardos cortos y una estructura periódica asociada al ciclo diario, lo que sugiere dependencia temporal marcada y ausencia de estacionariedad estricta. En consecuencia, es razonable plantear transformaciones (p. ej., diferenciación) y modelos que incorporen explícitamente componentes estacionales. Además, la alta frecuencia (10 minutos) implica que el error de predicción tenderá a concentrarse en las transiciones rápidas (subidas matinales, bajadas nocturnas), donde pequeñas desviaciones en fase producen errores apreciables.

En conjunto, este apartado justifica técnicamente el uso posterior de un enfoque estacional (SARIMA) y define un criterio claro de evaluación: el modelo deberá reproducir la forma del ciclo diario y mantener estabilidad en periodos de cambio rápido, no solo “seguir la media” anual.

2 Estudio de outliers

2.1 Objetivo

El objetivo de este apartado es analizar si la serie temporal de demanda eléctrica presenta valores atípicos u observaciones anómalas que se desvíen de su comportamiento habitual. La identificación de outliers resulta relevante, ya que estos pueden afectar tanto al análisis descriptivo como al rendimiento de los modelos de predicción si no se tienen en cuenta adecuadamente.

El estudio se centra en detectar valores puntuales que se aparten de la dinámica normal de la serie, sin asumir a priori que dichos valores deban eliminarse.

2.2 Método de detección de outliers

Para la detección de valores atípicos se ha empleado un método robusto basado en el uso de una media móvil y la desviación absoluta mediana (MAD). Este enfoque permite detectar observaciones anómalas teniendo en cuenta el contexto local de la serie temporal, evitando que la elevada estacionalidad diaria distorsione el análisis.

En particular, se considera que una observación es un outlier si su desviación respecto a la media móvil local supera un umbral definido como un múltiplo de la MAD. Este tipo de criterio es ampliamente utilizado en series temporales reales debido a su robustez frente a valores extremos.

2.3 Cálculo de residuos y detección

```
# Media móvil diaria (144 observaciones = 24h)
df_out <- df %>%
  mutate(
    media_movil = rollmean(y, k = 144, fill = NA, align = "center"),
    residuo = y - media_movil
  )

# Estadísticos robustos
mediana_res <- median(df_out$residuo, na.rm = TRUE)
mad_res <- mad(df_out$residuo, na.rm = TRUE)

# Umbral (criterio conservador)
k <- 3

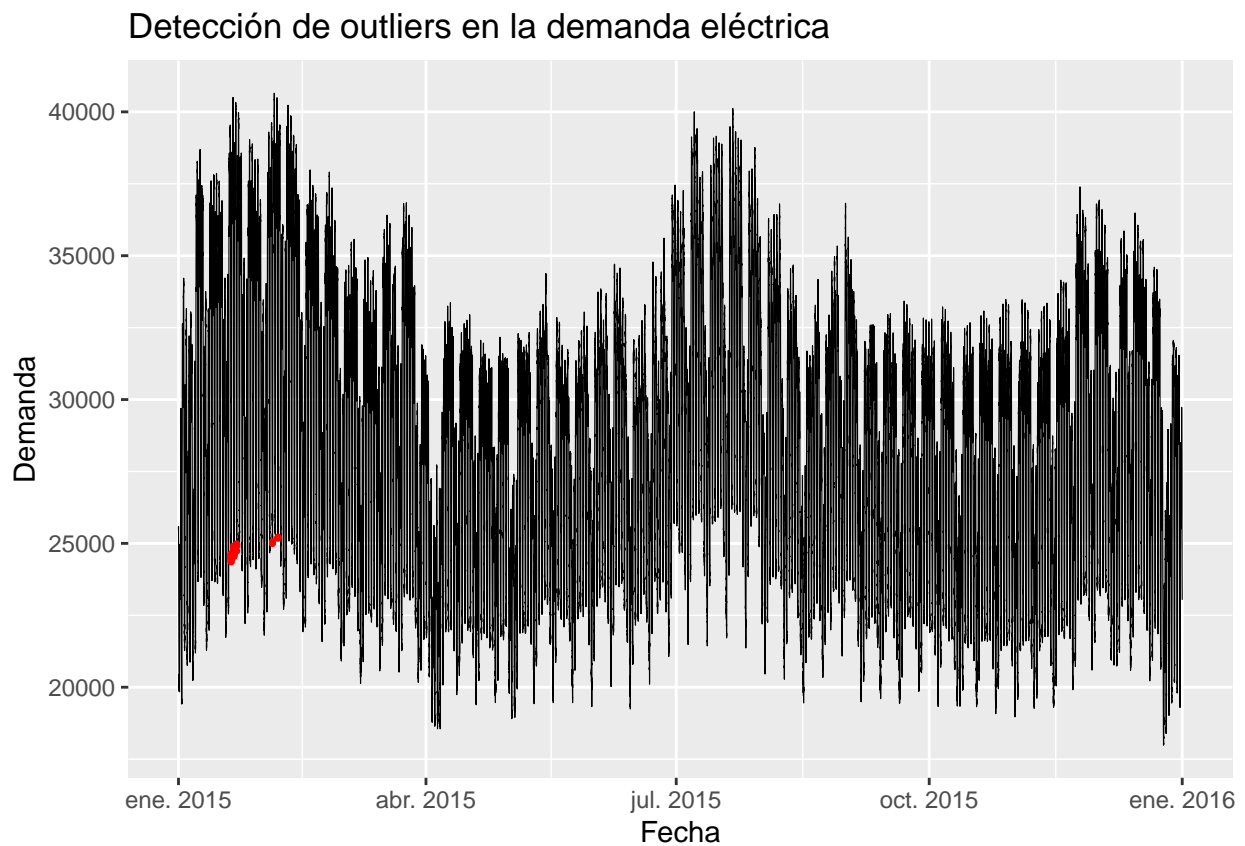
df_out <- df_out %>%
  mutate(
    outlier = abs(residuo - mediana_res) > k * mad_res
  )

# Número de outliers detectados
sum(df_out$outlier, na.rm = TRUE)
```

```
## [1] 31
```


2.4 Visualización de outliers

```
ggplot(df_out, aes(datetime, y)) +  
  geom_line(linewidth = 0.2) +  
  geom_point(  
    data = df_out %>% filter(outlier),  
    aes(datetime, y),  
    color = "red",  
    size = 0.6  
  ) +  
  labs(  
    title = "Detección de outliers en la demanda eléctrica",  
    x = "Fecha",  
    y = "Demanda"  
  )
```



En la figura se muestran los valores atípicos detectados sobre la serie temporal completa. Los outliers aparecen de forma puntual y aislada a lo largo del año, sin concentrarse en periodos específicos ni alterar la estructura general de la serie.

Visualmente, estos valores se manifiestan como picos o caídas abruptas que no siguen el patrón periódico dominante de la demanda eléctrica.

2.5 Ejemplos de valores atípicos detectados

```
df_out %>%  
  filter(outlier) %>%  
  select(datetime, y, media_movil, residuo) %>%  
  arrange(desc(abs(residuo))) %>%  
  head(10)
```

##		datetime	y	media_movil	residuo
## 1	2015-01-21	03:50:00	24548	34141.17	-9593.167
## 2	2015-01-21	04:00:00	24560	34142.13	-9582.132
## 3	2015-01-21	04:10:00	24618	34142.07	-9524.069
## 4	2015-01-21	03:40:00	24634	34141.13	-9507.132
## 5	2015-01-21	04:30:00	24637	34142.78	-9505.785
## 6	2015-01-21	04:20:00	24676	34142.83	-9466.826
## 7	2015-01-21	03:30:00	24723	34138.29	-9415.292
## 8	2015-01-21	05:00:00	24733	34142.93	-9409.931
## 9	2015-01-20	04:00:00	24358	33744.31	-9386.306
## 10	2015-01-21	04:40:00	24800	34142.33	-9342.326

2.6 Comentario e impacto en la predicción

El método empleado detecta un número reducido de valores anómalos (31 outliers) a lo largo del año 2015, lo que representa aproximadamente el 0,06 % del total de observaciones. Estos valores corresponden a caídas o picos abruptos de la demanda eléctrica con respecto al comportamiento local esperado, una vez eliminada la estacionalidad diaria mediante la media móvil.

Los outliers identificados, como ya se comentó con la gráfica, no siguen el patrón diario típico de la serie y se concentran en intervalos temporales muy concretos, lo que sugiere la presencia de eventos excepcionales o incidencias puntuales, más que un comportamiento sistemático de la serie. Este tipo de observaciones puede afectar negativamente al ajuste de modelos de predicción, especialmente aquellos sensibles a valores extremos.

2.7 Conclusión y reflexión del apartado

El procedimiento robusto basado en media móvil diaria y MAD identifica un número reducido de observaciones anómalas (31 outliers), lo que representa una proporción muy baja respecto al total del año. Este resultado sugiere que la serie es estructuralmente estable y que la variabilidad dominante proviene de estacionalidad y cambios graduales, más que de eventos extremos frecuentes. En términos prácticos, esto es relevante porque permite interpretar que el error de predicción no estará dominado por picos raros, sino por la capacidad del modelo para seguir patrones repetitivos.

No obstante, aunque sean pocos, los outliers pueden concentrar una parte desproporcionada del error en métricas como RMSE, y además pueden inducir sesgos locales si coinciden con tramos de alta pendiente (por ejemplo, cambios abruptos en periodos de subida/bajada). Por esa razón, en lugar de eliminarlos, se opta por mantenerlos para evaluar la robustez del método de predicción en condiciones realistas: en sistemas energéticos reales existen incidencias puntuales, y un modelo útil debe comportarse razonablemente ante ellas.

En conjunto, el análisis de outliers aporta dos conclusiones operativas: (i) no es necesario “limpiar” agresivamente la serie para poder modelarla, y (ii) conviene interpretar los errores del modelo diferenciando entre fallos estructurales (no capturar estacionalidad o deriva) y fallos puntuales asociados a eventos anómalos.

3 Metodología de predicción de la serie temporal

3.1 Objetivo del apartado

El objetivo de este apartado es seleccionar y describir de manera resumida un método de predicción publicado en una revista científica de prestigio internacional, así como justificar su idoneidad para la predicción de la serie temporal de demanda eléctrica analizada. El método seleccionado se estudia a partir del artículo original y se adaptará posteriormente a los datos disponibles en esta práctica.

3.2 Artículo científico seleccionado

El método de predicción empleado en esta práctica se basa en el siguiente artículo científico:

Hyndman, R. J., & Khandakar, Y. (2008).
Automatic time series forecasting: the forecast package for R.
Journal of Statistical Software, 27(3), 1–22.

Disponible en acceso abierto en:
<https://www.jstatsoft.org/article/view/v027i03>

En este artículo se presenta un procedimiento automático para la predicción de series temporales basado en modelos ARIMA y SARIMA, siguiendo la metodología de Box–Jenkins y utilizando criterios de información para la selección óptima de los parámetros del modelo. El método descrito constituye la base teórica del paquete `forecast` del lenguaje R, ampliamente utilizado en el ámbito académico y profesional.

3.3 Características relevantes de la serie para la predicción

A partir del análisis exploratorio realizado en los apartados anteriores, la serie temporal de demanda eléctrica presenta las siguientes características principales:

- Frecuencia fija de 10 minutos, lo que supone 144 observaciones por día.
- Presencia de estacionalidad diaria claramente marcada, asociada a los ciclos de actividad.
- Dependencia temporal significativa, evidenciada mediante el análisis de autocorrelación.
- Existencia de valores atípicos puntuales, que no alteran la estructura global de la serie.

Estas características coinciden con los supuestos y casos de uso considerados en el artículo de referencia, lo que justifica la aplicación de modelos ARIMA con componentes estacionales.

3.4 Modelo de predicción seleccionado

Siguiendo el enfoque propuesto por Hyndman y Khandakar, se adopta un modelo SARIMA (Seasonal ARIMA) como técnica de predicción. Este modelo extiende el modelo ARIMA clásico incorporando términos estacionales, lo que permite capturar patrones periódicos presentes en la serie temporal.

El modelo SARIMA resulta especialmente adecuado para la serie analizada, ya que permite modelar simultáneamente la dependencia temporal a corto plazo y la estacionalidad diaria observada en los datos de demanda eléctrica.

3.5 Estructura general del modelo SARIMA

Un modelo SARIMA se denota como:

$$SARIMA(p, d, q)(P, D, Q)_s$$

donde p , d y q representan los órdenes autorregresivo, de diferenciación y de media móvil no estacionales, P , D y Q representan los órdenes estacionales y s es el periodo estacional. En esta práctica, el periodo estacional corresponde a $s = 144$, equivalente a un día completo.

La selección de los parámetros del modelo se realiza de forma automática siguiendo el procedimiento descrito en el artículo, basado en la minimización de criterios de información como el criterio de Akaike corregido (AICc).

3.6 Procedimiento de entrenamiento y evaluación

De acuerdo con la metodología presentada en el artículo seleccionado, el procedimiento seguido para la predicción de la serie temporal es el siguiente:

1. División de la serie temporal en un conjunto de entrenamiento y un conjunto de test, respetando estrictamente el orden temporal de los datos.
2. Ajuste del modelo SARIMA utilizando exclusivamente los datos del conjunto de entrenamiento.
3. Selección automática de los parámetros del modelo mediante la minimización del criterio AICc.
4. Generación de predicciones para el horizonte temporal definido en el enunciado de la práctica.
5. Evaluación del rendimiento del modelo mediante métricas de error estándar, como el error absoluto medio o la raíz del error cuadrático medio.

Este procedimiento evita el uso de información futura durante el entrenamiento y permite evaluar de forma realista la capacidad predictiva del modelo.

3.7 Consideraciones sobre los valores atípicos

Aunque en el apartado anterior se han identificado valores atípicos en la serie temporal, estos no se eliminan antes del ajuste del modelo. Siguiendo el enfoque del artículo de referencia, el objetivo es analizar el comportamiento del método de predicción ante datos reales que pueden incluir observaciones excepcionales.

El uso de modelos SARIMA, basados en dependencias temporales agregadas, contribuye a mitigar parcialmente la influencia de estos valores atípicos sobre la predicción global.

3.8 Conclusión del apartado

En este apartado se ha seleccionado un modelo SARIMA automático como metodología de predicción para la serie de demanda eléctrica, apoyándose en el procedimiento propuesto por Hyndman y Khandakar (2008). Aunque se trata de un artículo publicado hace más de una década, su relevancia se mantiene vigente en el ámbito académico y profesional, ya que constituye la base teórica de uno de los marcos de predicción más utilizados en análisis de series temporales. En este contexto, el modelo SARIMA se adopta como una línea base sólida, interpretable y metodológicamente bien fundamentada.

La elección del método no responde únicamente a su popularidad, sino a su coherencia con las características empíricas observadas en la serie: fuerte estacionalidad diaria, dependencia temporal significativa y comportamiento relativamente estable a nivel estructural. Estas propiedades encajan de manera natural con el marco Box–Jenkins y con la incorporación explícita de componentes estacionales.

No obstante, es importante reconocer que los modelos SARIMA presentan limitaciones conocidas. En particular, su estructura lineal puede dificultar la captura de comportamientos altamente no lineales o cambios abruptos en la dinámica de la serie. Además, al no incorporar variables exógenas (como temperatura, calendario laboral o eventos extraordinarios), el modelo se apoya exclusivamente en la información histórica, lo que puede limitar su capacidad predictiva en determinados contextos.

Por tanto, la elección de SARIMA en este trabajo debe interpretarse como una decisión metodológicamente coherente y adecuada para establecer una referencia cuantitativa clara. El apartado siguiente evaluará empíricamente su rendimiento en un escenario de predicción a corto plazo, permitiendo analizar en qué medida las hipótesis estructurales del modelo se traducen en un desempeño satisfactorio sobre datos reales.