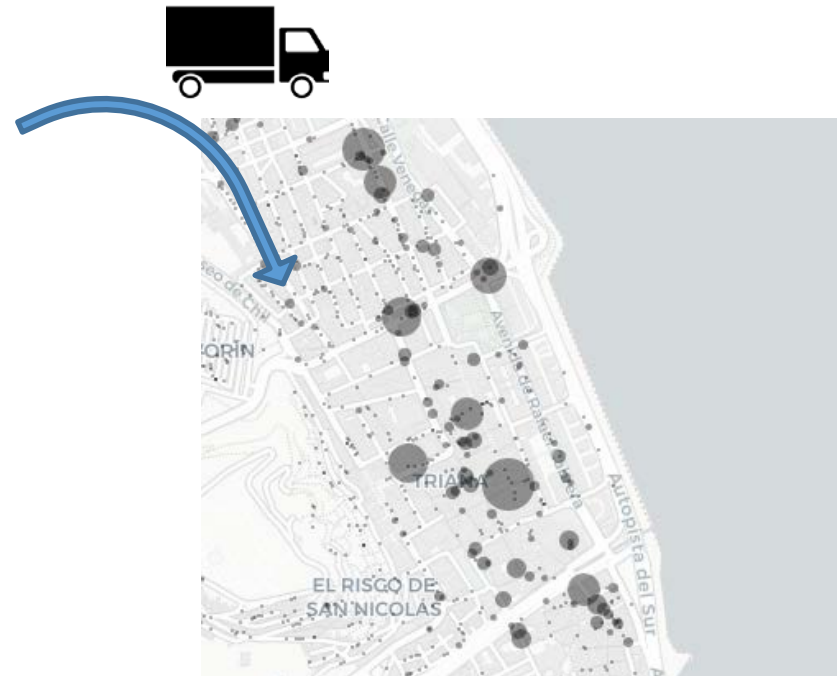


XI Jornadas de usuarios de R

Ajuste de modelos predictivos de series temporales para estimar los pedidos de un producto de consumo energético en una ciudad

Introducción

- Este es un proyecto relacionado con una empresa distribuidora de un producto relacionado con la energía.
- En este trabajo se muestra un estudio de modelos de predicción del número de pedidos realizados por clientes.



Modelos analizados: TBATS

- Se han analizado diversos modelos de series temporales que permiten predecir el número de pedidos.
- En primer lugar se ha analizado los modelos TBATS de series temporales (De Livera, Hyndman & Snyder, 2011).
- Estos modelos están implementado en la librería **forecast** (<https://github.com/robjhyndman/forecast>)
- TBATS permiten combinar de forma automatizada un modelo de regresión dinámica con términos de Fourier, que funciona como un modelo de espacio de estados de suavizado exponencial, y una transformación de Box-Cox.
- Los errores o residuales se modelizan como procesos ARIMA.
- Permiten introducir variables exógenas.
- También es posible ajustar múltiples periodos estacionales

De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011) Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing. Journal of the American Statistical Association, 106, 1513-1527.

<https://doi.org/10.1198/jasa.2011.tm09771>

Modelos analizados: TBATS

- El modelo TBATS se expresa de la forma siguiente:

$$y_t = \gamma x_t + \sum_{k=1}^K \alpha_k \sin\left(\frac{2\pi kt}{m}\right) + \beta_k \cos\left(\frac{2\pi kt}{m}\right) + \varepsilon_t$$
$$\phi(B)\nabla^d \varepsilon_t = \theta(B)z_t$$

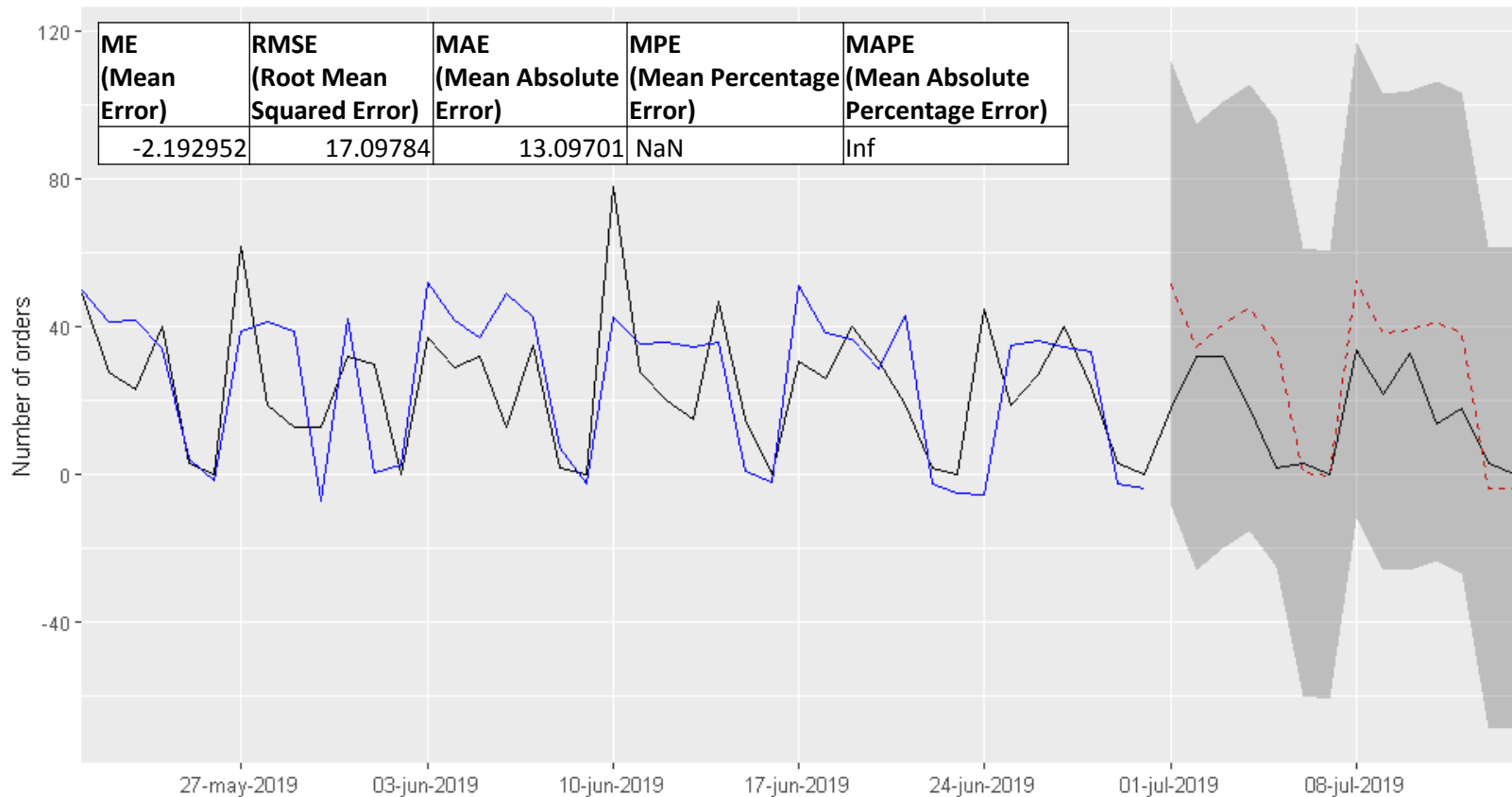
- m es el parámetro de periodo estacional, $\phi(B)$ y $\theta(B)$ son los términos AR y MA, z_t es ruido blanco y x_t denota la covariable en tiempo t .
- Nosotros hemos ajustado estos modelos considerando el efecto calendario (festivos y fines de semana).
- Buscamos (de forma iterativa) el número de términos de Fourier óptimo que permita minimizar el criterio AIC

Ejemplo de ajuste con TBATS

- En un municipio determinado y para un tipo de producto dado

Predicting number of orders of XXX product in municipality WWW

Regression with ARIMA(0,1,1) errors and weekly-monthly ($m=7$ and $m=30$) periodicity



Modelos analizados: Prophet

- También hemos utilizado el procedimiento Prophet, de Facebook (<https://facebook.github.io/prophet/>).
- Prophet, que está disponible tanto en R como Python (Taylor & Letham, 2018), está basado en modelos aditivos que ajusta las tendencias no lineales con estacionalidades anuales, semanales y diarias y considera también el efecto calendario (festivos).
- Trabaja bien con series que tienen fuertes efectos estacionales y varias temporadas de datos históricos.
- Prophet es bastante robusto a datos perdidos y desviaciones en la tendencia, y maneja bastante bien los outliers.
- En R el procedimiento está implementado en la librería **prophet**.

Sean J. Taylor, Benjamin Letham (2018) Forecasting at scale. The American Statistician 72(1):37-45
<https://peerj.com/preprints/3190.pdf>.

Modelos analizados: Prophet

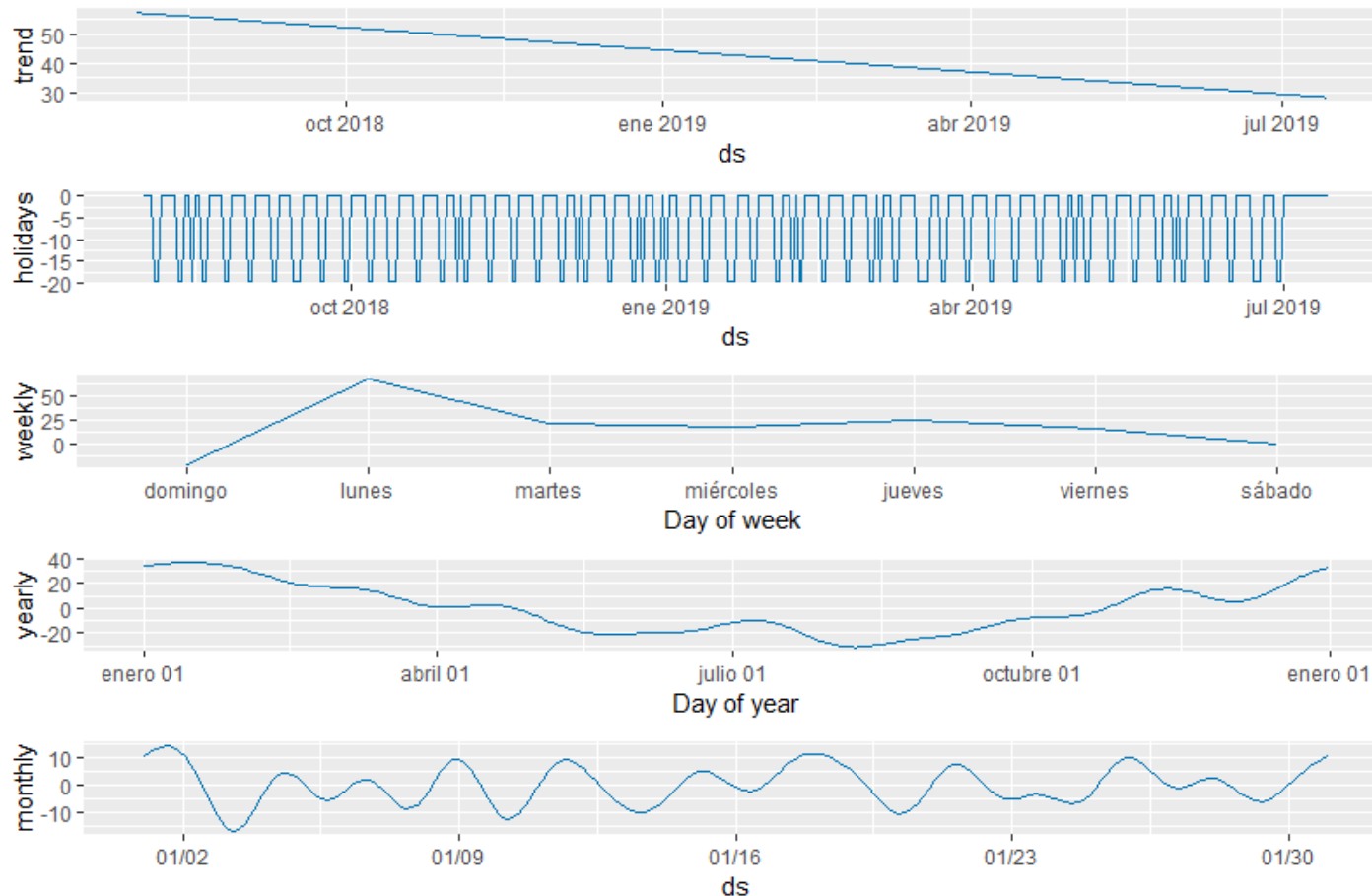
- En este caso, la expresión de carácter aditivos sería

$$y_t = g_t + s_t + h_t + \varepsilon_t$$

- donde g_t modeliza la componente de la tendencia con los cambios no periódicos (por ejemplo, puede ser una componente de regresión con respecto a t), s_t representa la componente estacional (modelizada también como series de Fourier) y h_t recoge la información del efecto calendario (vacaciones, festivos, eventos,...).
- Es un procedimiento muy utilizado para series de valores diarios.
- Además, permite introducir puntos temporales de cambio abrupto en la tendencia de las series (changepoints).

Modelos analizados: Prophet

- La librería dispone de métodos para representar las distintos componentes de la serie.

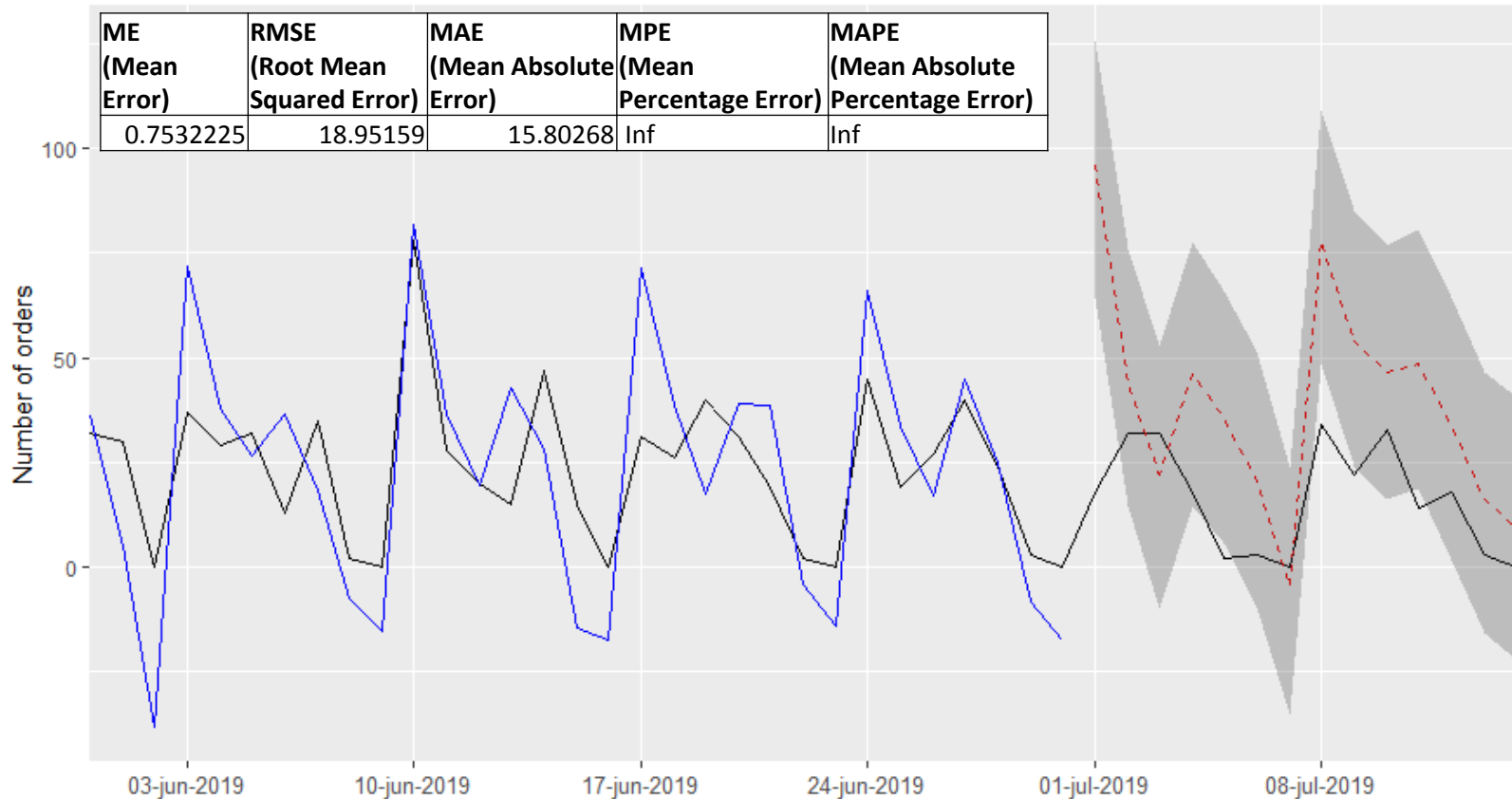


Ejemplo de ajuste con Prophet

- Considerando el mismo ejemplo de antes

Predicting number of orders of XXX product in municipality WWW

Regression with Prophet library and weekly-monthly periodicity



Líneas de trabajo

- De momento, tenemos resultados similares en los modelos que ajustan una serie de valores diarios de un año entero.
- De forma progresiva se dispondrá de más datos adicionales.
- También queremos estudiar los modelos LSTM (Long Short Term Memory) con el fin de evaluar y comparar los resultados obtenidos hasta el momento.
- LSTM (Sherstinsky, 2018) tiene la ventaja de utilizar Tensorflow con Keras y ya hay librería en R que lo implementan.
- La posible ventaja que tiene el uso de un red neuronal es que podríamos capturar efectos no lineales e interacciones entre variables.

Sherstinsky, Alex (2018). “Fundamentals of Recurrent Neural Network (Rnn) and Long Short-Term Memory (Lstm) Network,” August. URL: <http://arxiv.org/abs/http://arxiv.org/abs/1808.03314v4>.

Muchas gracias

Ponente: Carlos Pérez González (cpgonzal@ull.es)