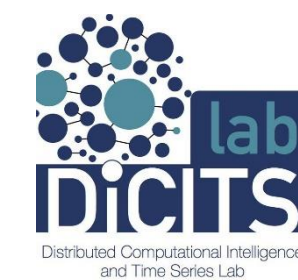
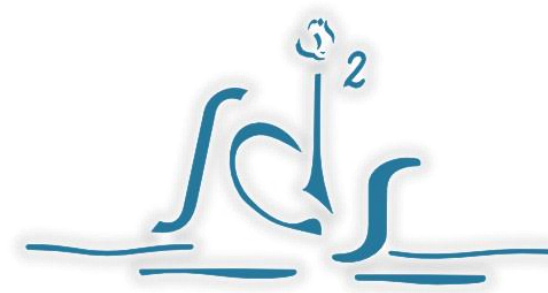


Análisis y modelado de series temporales en R

F. J. Baldan (fjbaldan@decsai.ugr.es),
F. Aragon, M. Parra-Royon y J. M. Benitez,
Soft Computing and Intelligent Information Systems (SCI²S), Departamento de Ciencias de la Computación e Inteligencia Artificial (DECSAI), Distributed Computational Intelligence and Time Series (DiCITS), Universidad de Granada, España



Introducción

R es un lenguaje de programación especializado en el análisis estadístico de datos y minería de datos.

En la actualidad se registran grandes cantidades de series temporales: electrocardiogramas, valores en bolsa, patrones de consumo eléctrico, etc.

Las series temporales necesitan ser procesadas durante su análisis y R cuenta con una gran cantidad de herramientas para esta labor.

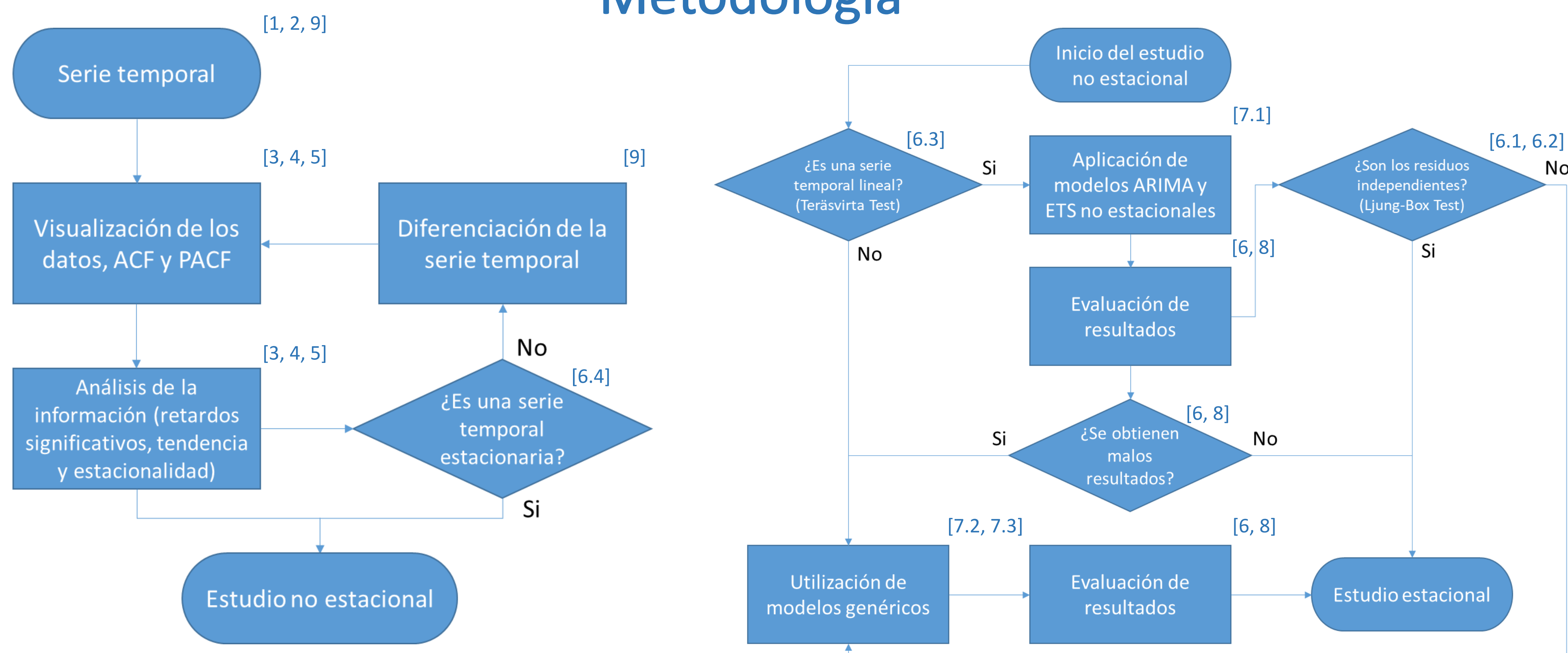
Objetivos

1. Obtener una serie temporal estacionaria, libre de comportamientos no deseados: tendencia y estacionalidad.
2. En caso de no obtener una serie temporal estacionaria es necesario detectar dicho comportamiento y aplicar modelos que se adapten de la mejor forma posible.

Herramientas

1. Estructuras de datos: `ts` (package stats), `xts` (xts), `zoo` (zoo).
2. Normalizaciones: `scale` (base), `[0,1]`, `[min, max]`.
3. Descomposiciones: `stl` (stats) y `decompose` (stats).
4. Visualización de los datos: `plot.ts` (stats) y `qplot` (ggplot2).
5. Gráficos de autocorrelación y autocorrelación parcial: `acf` y `pacf` (forecast).
6. Tests estadísticos:
 1. Aleatoriedad: Box-Pierce y Ljung-Box tests, `Box.test` (stats).
 2. Normalidad: Shaphiro-Wilk y Jarque Bera tests, `shapiro.test` (stats) y `jarque.bera.test` (tseries), respectivamente.
 3. No linealidad: Teräsvirta test, `terasvirta.test` (tseries).
 4. Estacionariedad: Dickey-Fuller test, `adf.test` (tseries).
7. Modelos de predicción:
 1. Modelado de series temporales:
 - ARIMA: modelo autorregresivo integrado de media móvil, `arima` (stats) y `auto.arima` (forecast).
 2. Métodos de regresión estándar:
 - ETS: alisado exponencial, `ets` (forecast).
 3. Métodos de aprendizaje automático:
 - MLP: perceptrón multicapa, `mlp` (RSNNS).
 - NNET: red neuronal, `nnet` (nnet).
 - ELMAN: red de Elman, `elman` (RSNNS).
 - SVR: regresiones de soporte vectorial, `svm` (e1071).
8. Medidas de error, `error` (ftsa): RMSE, RelMAE y sMAPE, entre otras.
9. Se puede encontrar un extenso resumen de los recursos de los que dispone R para el tratamiento de series temporales en CRAN Task View: Time Series Analysis.

Metodología



*Entre corchetes se indican las herramientas utilizadas en cada paso de la metodología.

Conclusiones

- R cuenta con una gran cantidad de herramientas para el análisis y predicción de series temporales.
- La correcta utilización de dichas herramientas es fundamental en la obtención de buenos resultados.
- La metodología propuesta ha mejorado los resultados obtenidos en problemas reales tales como la predicción de carga de trabajo en servicios en la nube. Estas mejoras junto a la metodología propuesta han sido documentadas de forma extensa en el siguiente trabajo:

F. J. Baldan, S. Ramirez-Gallego, C. Bergmeir, F. Herrera and J. M. Benitez, "A Forecasting Methodology for Workload Forecasting in Cloud Systems," in *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1-1. doi: 10.1109/TCC.2016.2586064