

ESTUDIO DE LA PRODUCCIÓN FOTOVOLTAICA EN LA UDC A PARTIR DE DATOS FUNCIONALES EN R

MARINA SÁNCHEZ VILLAVERDE (UDC)

Manuel Oviedo de la Fuente
Carlos José Escudero Cascón

ÍNDICE

- Introducción
- Entendimiento de los datos
- Preprocesado de los datos
- Análisis de datos funcionales
- Modelado
- Conclusiones y trabajo futuro

INTRODUCCIÓN

1. Nacimiento de las energías verdes
2. Evolución de las energías renovables en España
3. Desafío de las energías renovables
4. Placas solares fotovoltaicas de la UDC: Definición y características
5. Grafana
6. Objetivo del proyecto

NACIMIENTO DE LAS ENERGÍAS VERDES

El consumo desmedido de:

- productos
- energía de origen fósil



Central térmica de As
Pontes. Fuente: abc.es

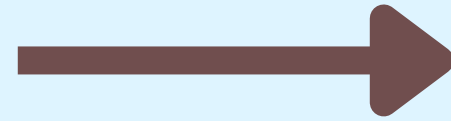


Central térmica de
Meirama.
Fuente: naturgy.es

NACIMIENTO DE LAS ENERGÍAS VERDES

El consumo desmedido de:

- productos
- energía de origen fósil



Consecuencias medioambientales:

- calentamiento global
- contaminación del aire y suelo
- generación excesiva de residuos



Central térmica de As Pontes. Fuente: abc.es



Central térmica de Meirama.
Fuente: naturgy.es

NACIMIENTO DE LAS ENERGÍAS VERDES

El consumo desmedido de:

- productos
- energía de origen fósil



Consecuencias medioambientales:

- calentamiento global
- contaminación del aire y suelo
- generación excesiva de residuos

Como respuesta



Consumo verde



Energía verde



Central térmica de As Pontes. Fuente: abc.es



Central térmica de Meirama.
Fuente: naturgy.es

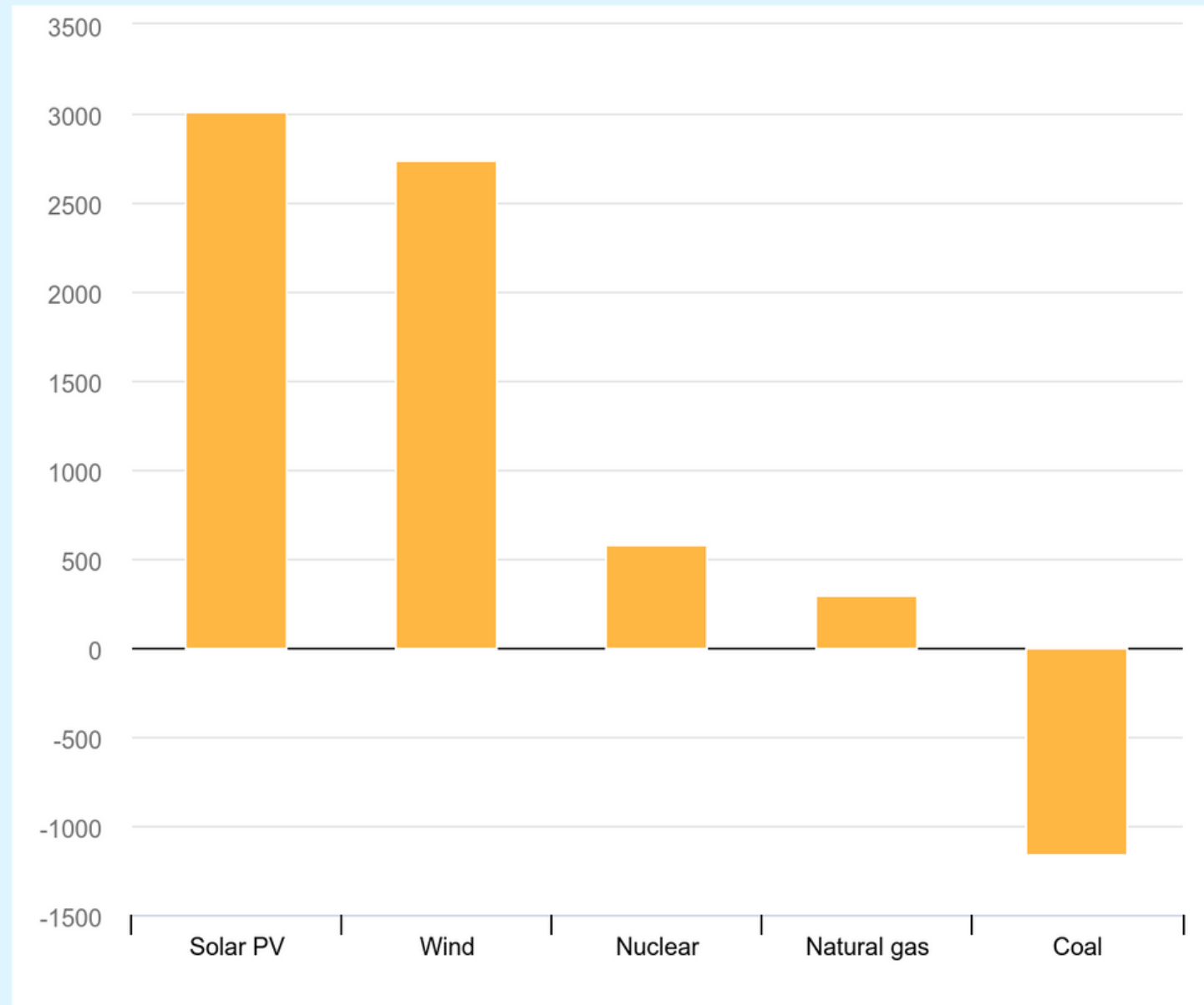


Parque eólico de Muras (Lugo). Fuente: elprogreso.es



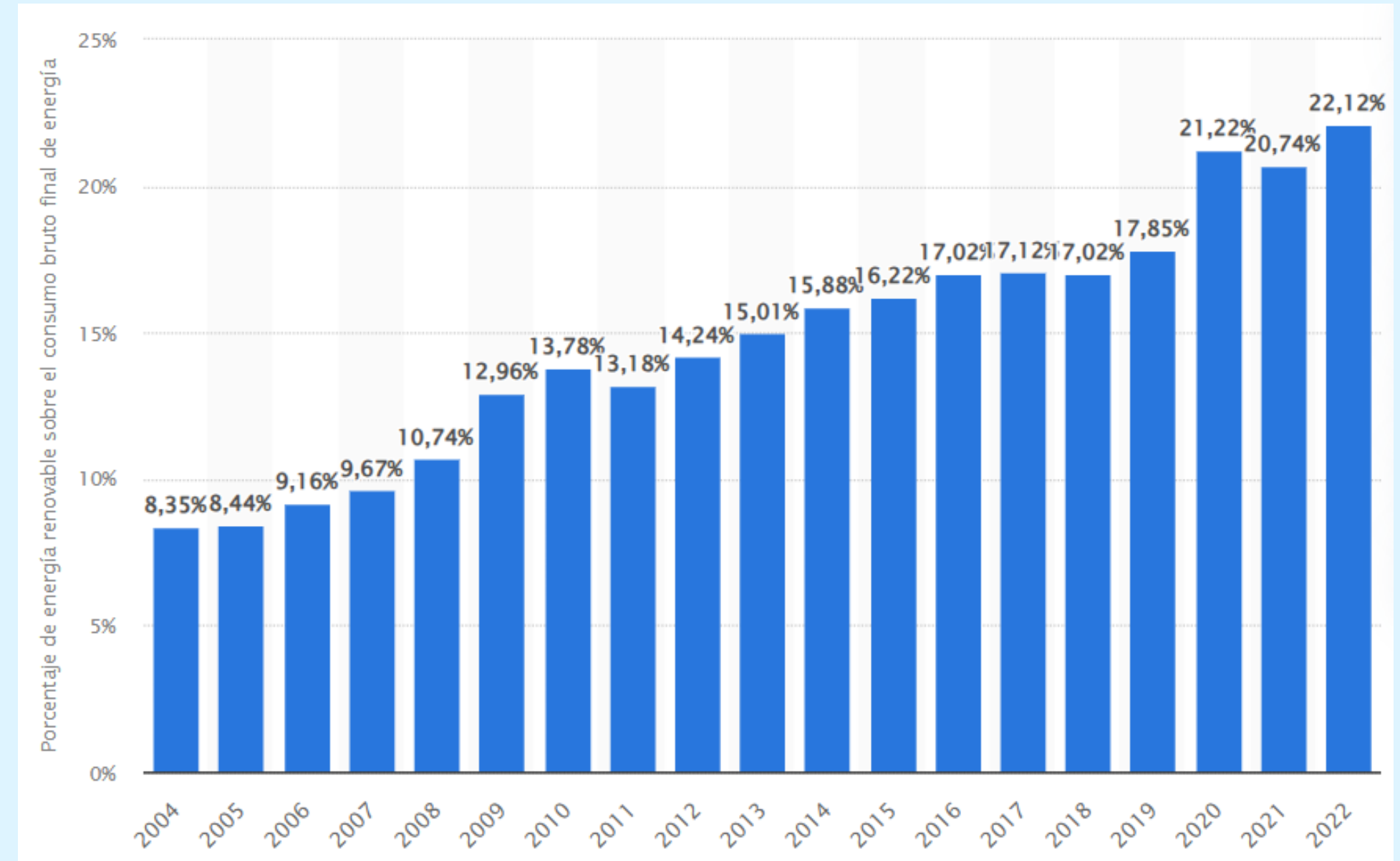
Placas solares fábrica de Estrella Galicia. Fuente: elespanol.com

EVOLUCIÓN DE LAS ENERGÍAS RENOVABLES EN ESPAÑA



Cambio en la generación eléctrica en el Escenario de Políticas Declaradas entre 2021 y 2030.

Fuente: abc.es



Porcentaje de energía procedente de fuentes renovables sobre el consumo bruto final de energía en España de 2004 a 2022.

Fuente: statista.es



DESAFÍO

DEPENDENCIA CON LAS CONDICIONES
METEOROLÓGICAS

PLACAS SOLARES FOTOVOLTAICAS DE LA UDC

Las placas solares fotovoltaicas son dispositivos capaces de capturar la energía proveniente del sol y transformarla en electricidad.

Sistemas:

- Elviña-Pabellón de deportes: Inversores Goodwe GB15KN-DT y controlador de energía Goodwe sec100
- Fotovoltaicas LACECAL: Inversores Fronius Tauro y controlador de energía LACECAL ITR 2.0

PLACAS SOLARES FOTOVOLTAICAS DE LA UDC

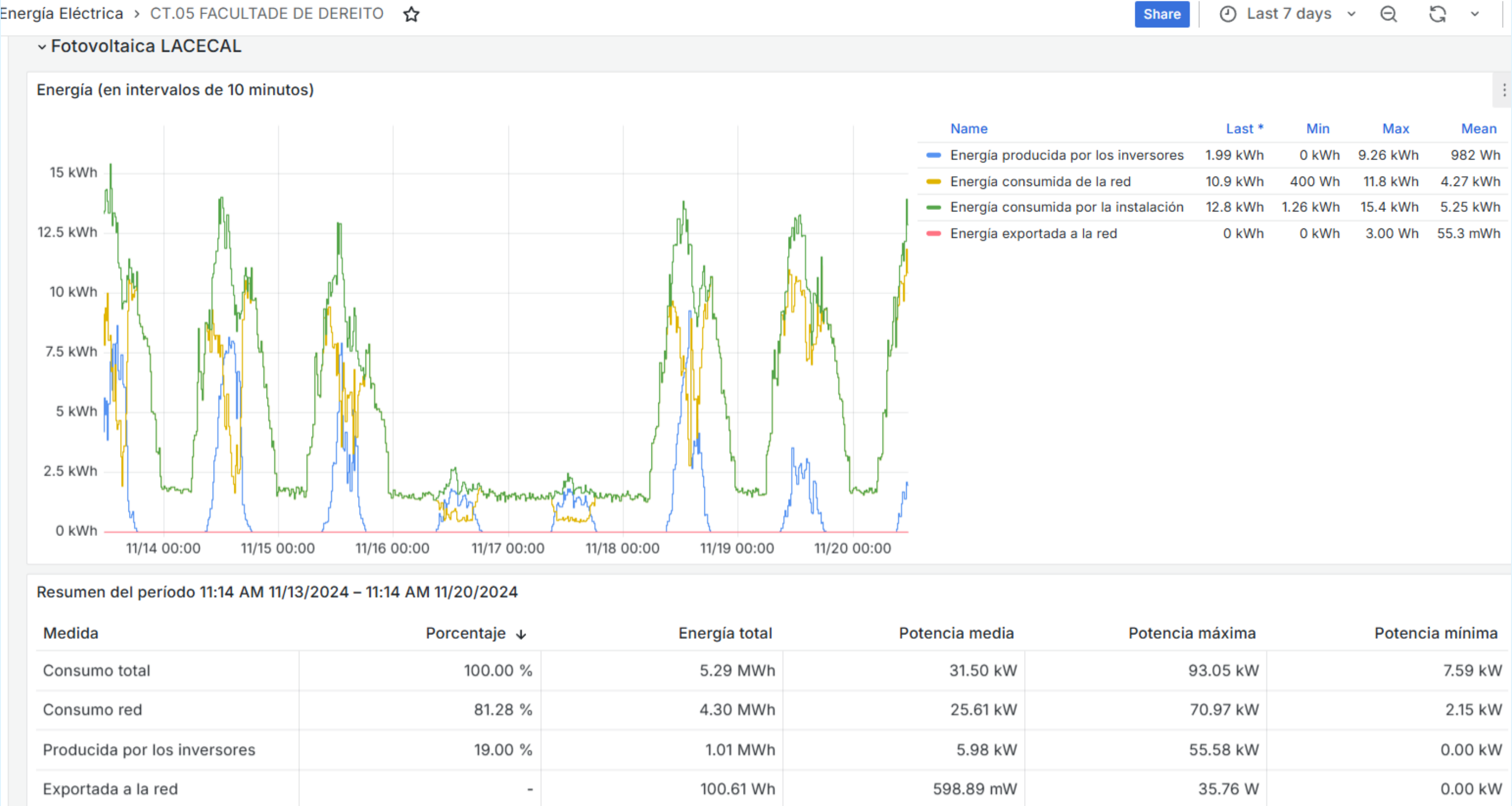
Las placas solares fotovoltaicas son dispositivos capaces de capturar la energía proveniente del sol y transformarla en electricidad.

Sistemas:

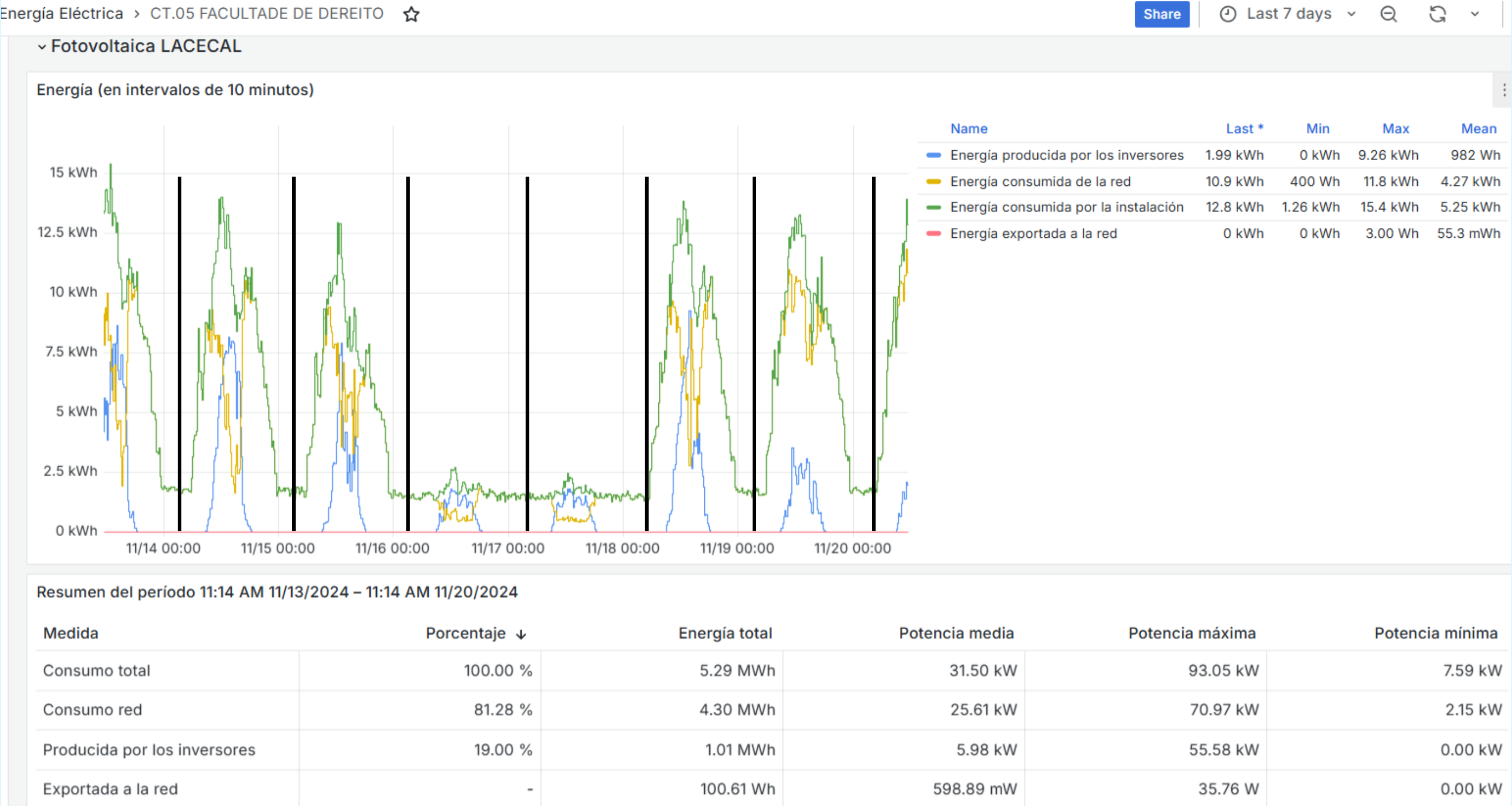
- Elviña-Pabellón de deportes: Inversores Goodwe GB15KN-DT y controlador de energía Goodwe sec100
- Fotovoltaicas LACECAL: Inversores Fronius Tauro y controlador de energía LACECAL ITR 2.0



GRAFANA



GRAFANA

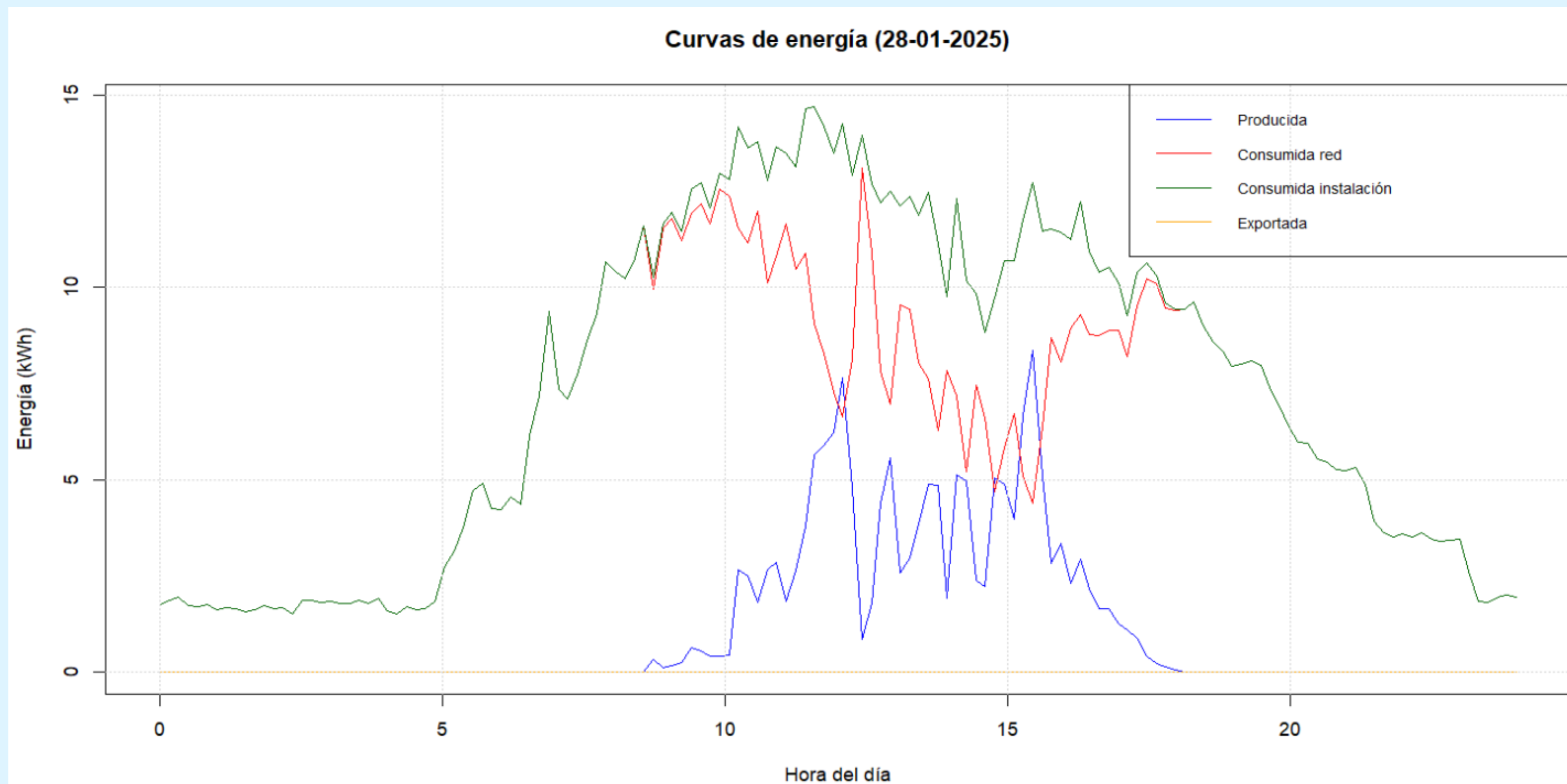


OBJETIVOS

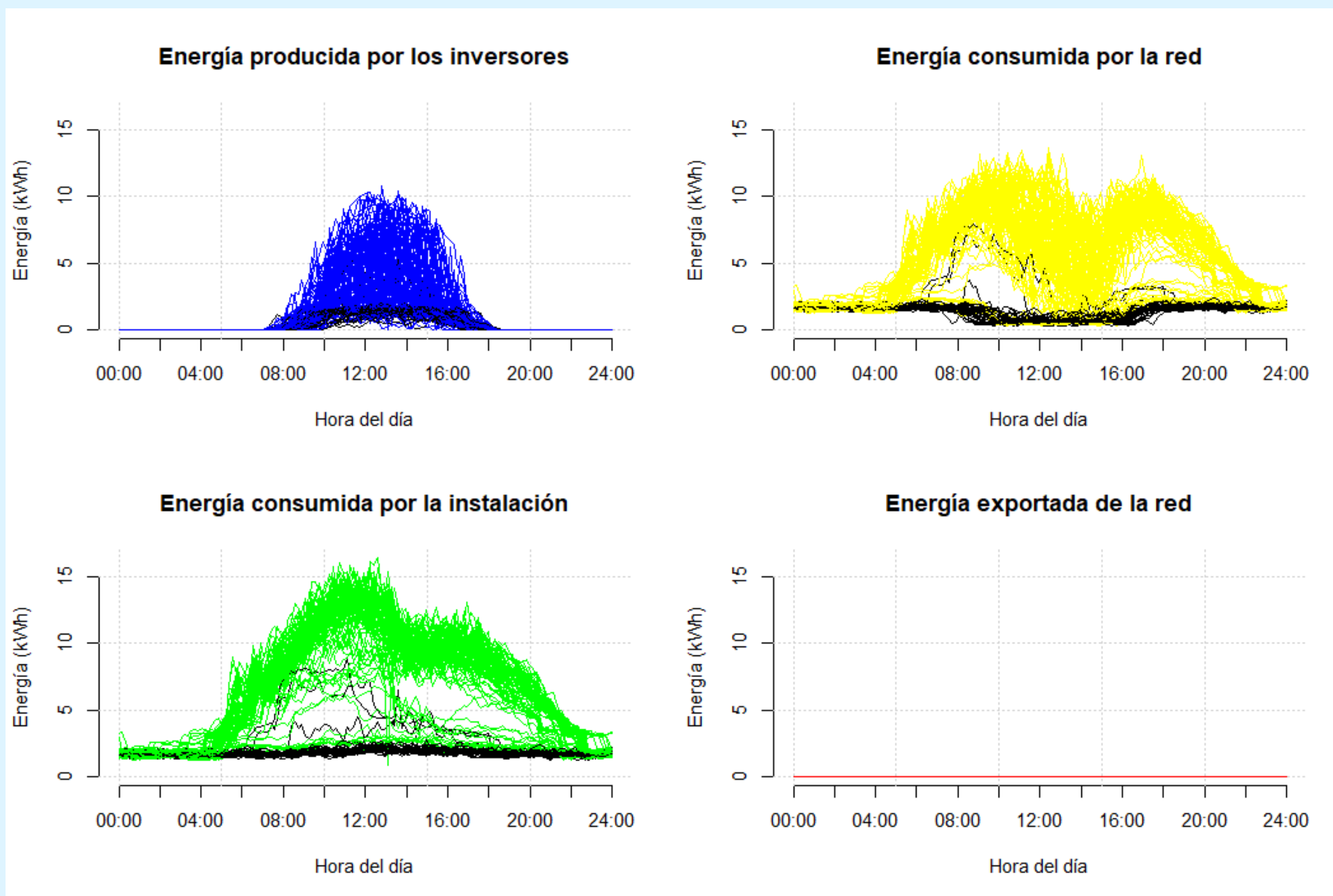
- Analizar y describir los patrones diarios de producción fotovoltaica (y consumo) en los edificios.
- Aplicar técnicas de análisis de datos funcionales para extraer características relevantes de las curvas diarias.
- Construir modelos predictivos que estimen la energía producida evaluando la precisión.
- Proponer mejoras o futuras líneas de trabajo.

ENTENDIMIENTO DE LOS DATOS

ENTENDIMIENTO DE LOS DATOS



ENTENDIMIENTO DE LOS DATOS



ENTENDIMIENTO DE LOS DATOS

Fecha	Variable	Valor	Unidad
04/10/2024	Brillo do Ceo Nocturno	15.080	mag/arsec ²
04/10/2024	Chuvia	0.4	L/m ²
04/10/2024	Temperatura media a 1.5m	18.01	°C
04/10/2024	Dirección de la ráfaga del viento	318	°
04/10/2024	Humedad relativa media	95	%
04/10/2024	Presión	1012.84	hPa
04/10/2024	Velocidad del viento a 10 m	6.4	km/h

Datos meteorológicos MeteoGalicia

PREPROCESADO DE LOS DATOS

PREPROCESADO DE LOS DATOS

Lista con:

- **Energia:** agrupa los objetos funcionales (fdata) correspondientes a los 4 tipos de energía.
- **datos_meteo:** dataframe con las variables meteorológicas y contextuales.

PREPROCESADO DE LOS DATOS

Lista con:

- **Energia:** agrupa los objetos funcionales (fdata) correspondientes a los 4 tipos de energía.
- **datos_meteo:** dataframe con las variables meteorológicas y contextuales.

Fase exploratoria

Anomalías en el mes de octubre

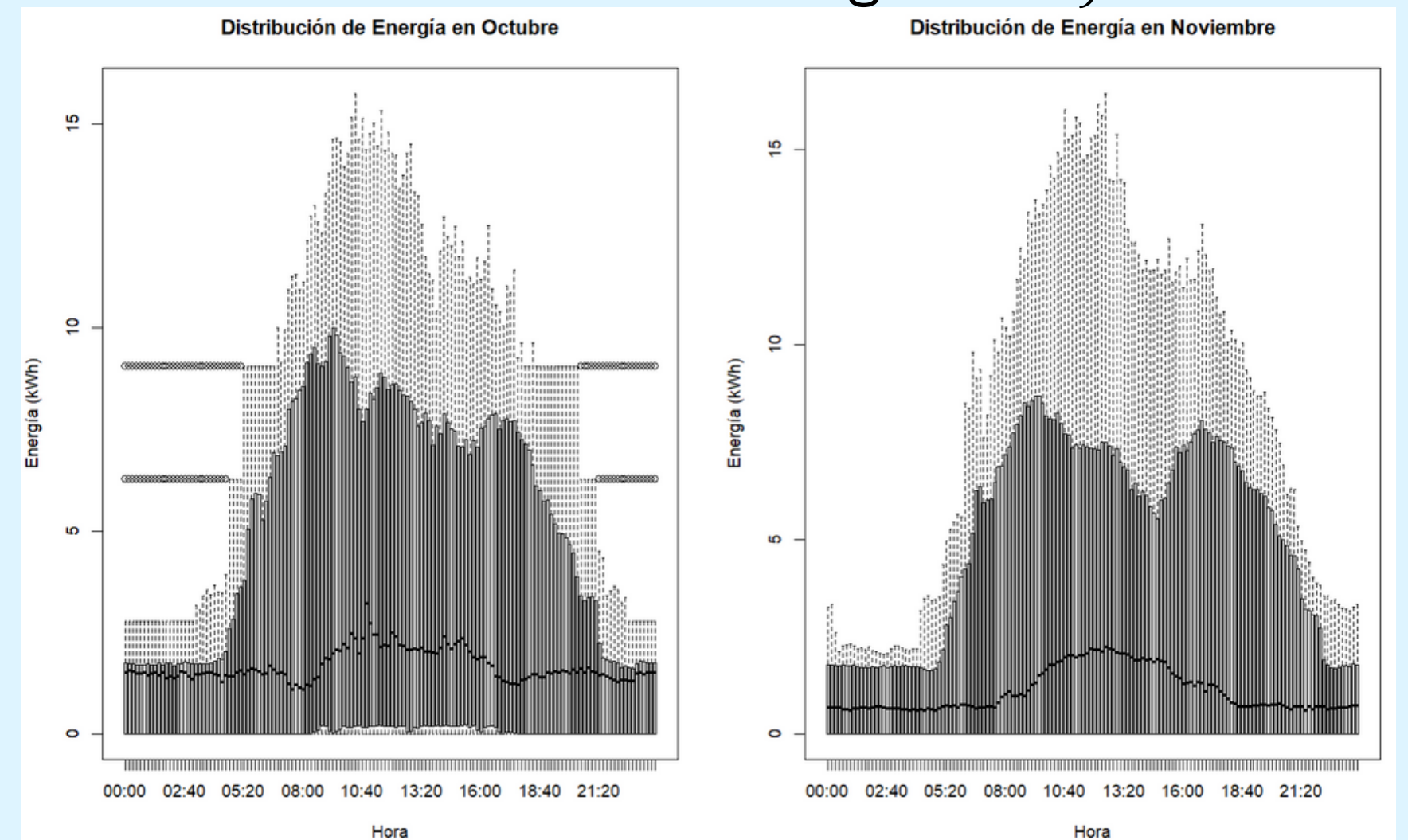


Diagrama de cajas de los meses octubre y noviembre para la Facultad de Derecho

Funciones propias

- **cargar_datos(ruta_csv)** → lee y organiza los ficheros exportados desde Grafana.
- **procesar_datos(datos, df)** → limpieza, integración y creación de estructuras para análisis.
- **plot_energia(datos, ...)** → visualización de las curvas energéticas (dos variantes).

Paquetes y funciones clave

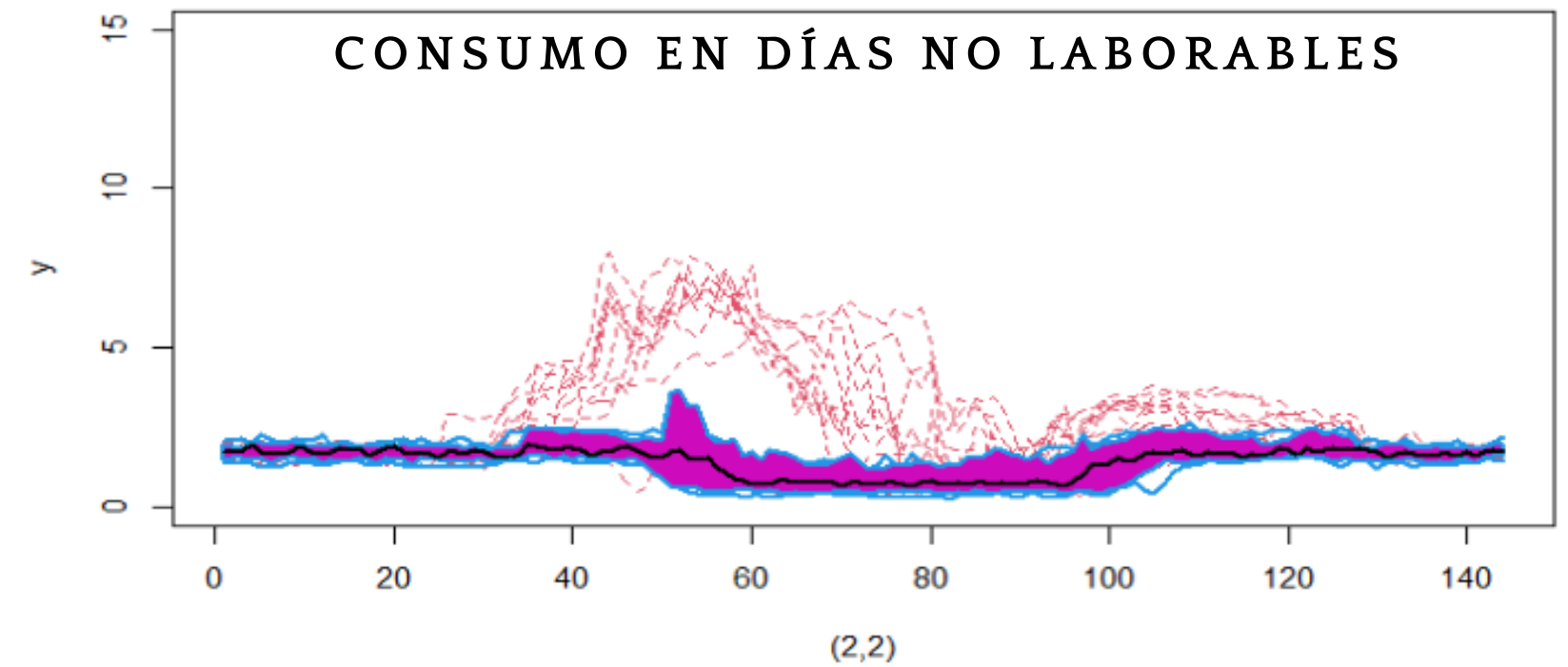
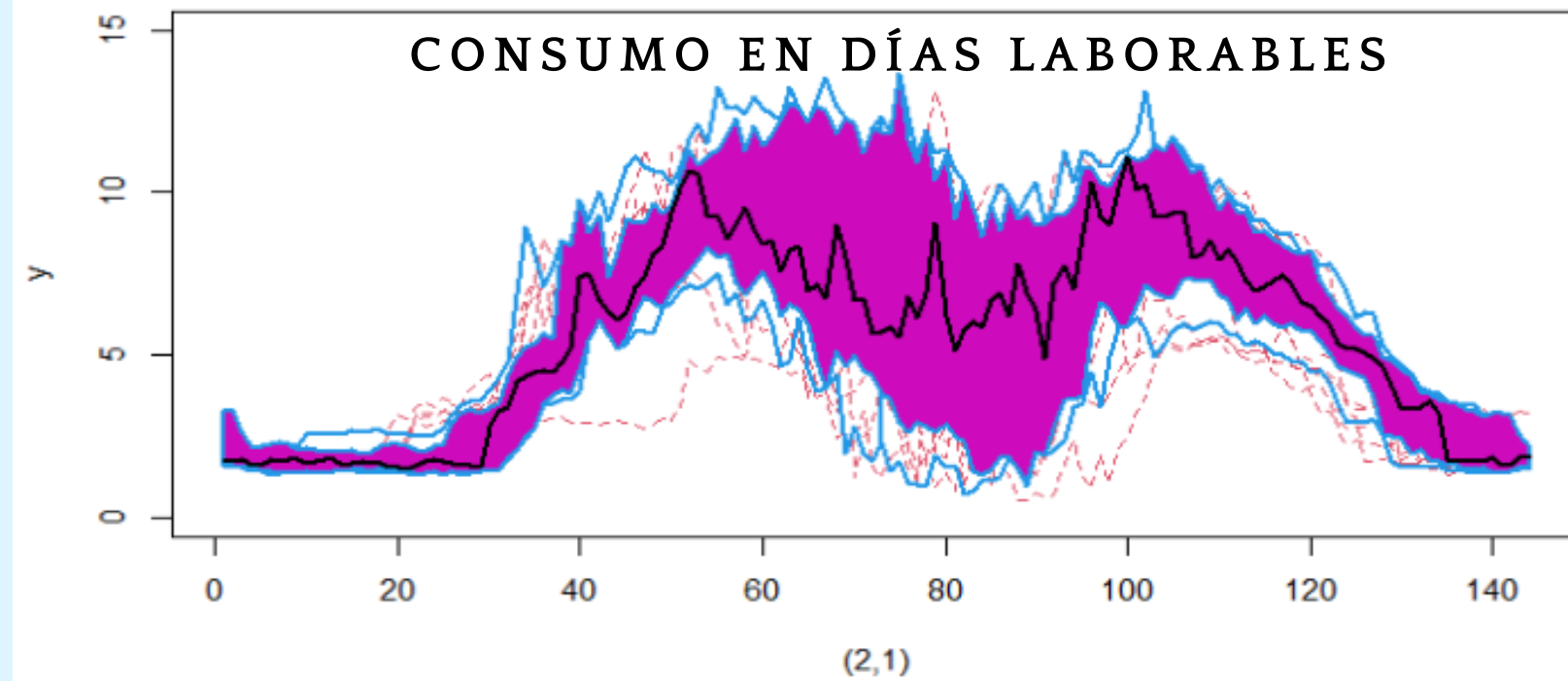
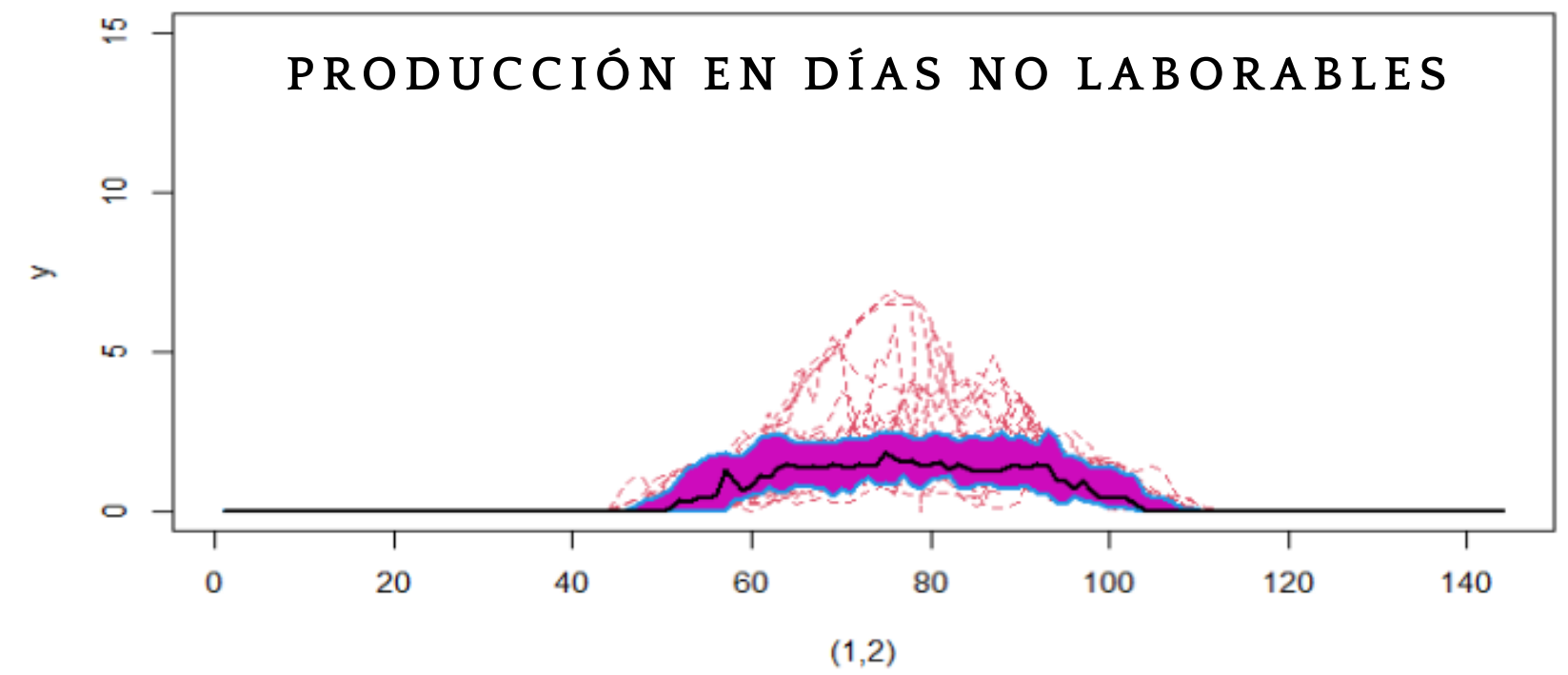
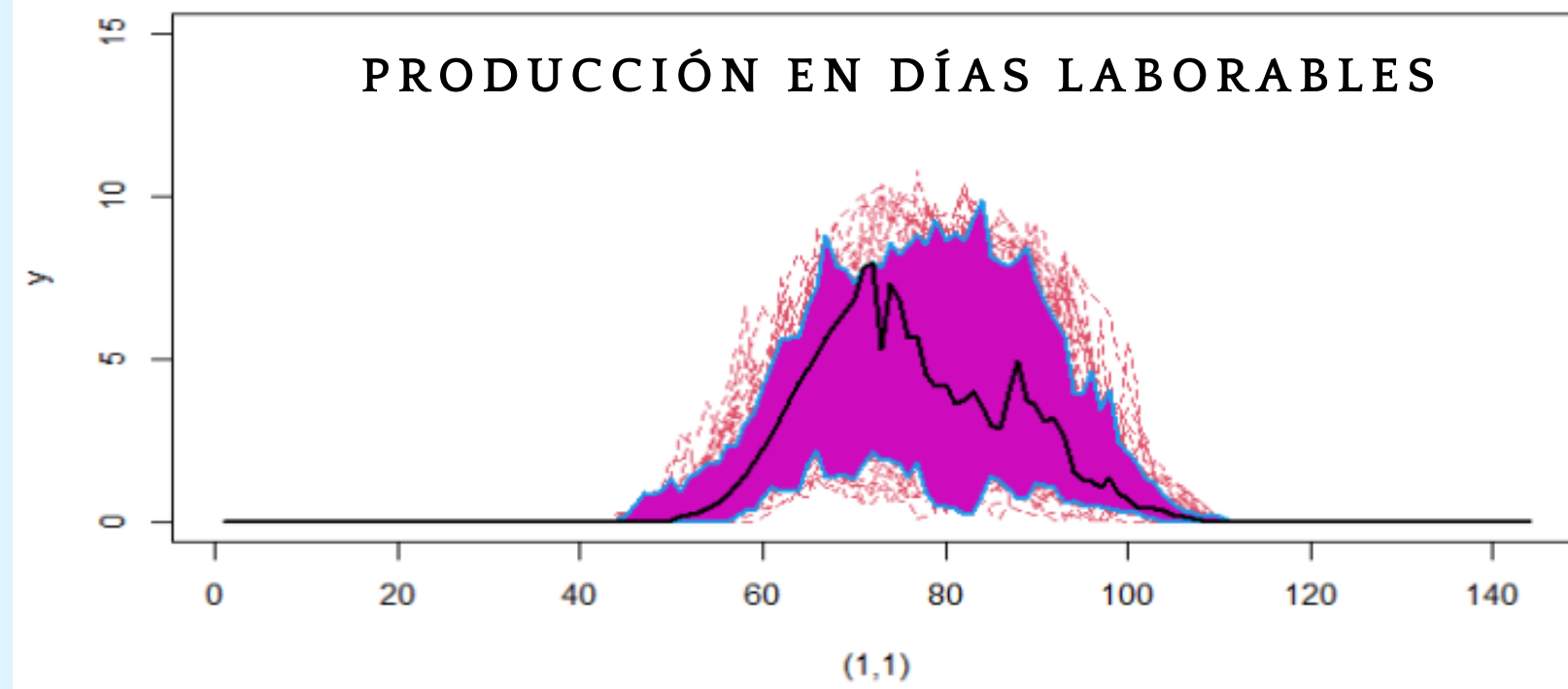
- **dplyr**: mutate, select, group_by, summarise (manipulación de datos).
- **tidyr**: pivot_wider, complete (reestructuración y completado de series).
- **lubridate**: ymd_hms, as_date, floor_date (gestión de fechas y tiempos).
- **fda.usc::fdata** → transformación de las series en curvas funcionales diarias.

CSV (GRAFANA) —► CARGAR_DATOS —► PROCESAR_DATOS —► FDATA (CURVAS)

ANÁLISIS REALIZADO

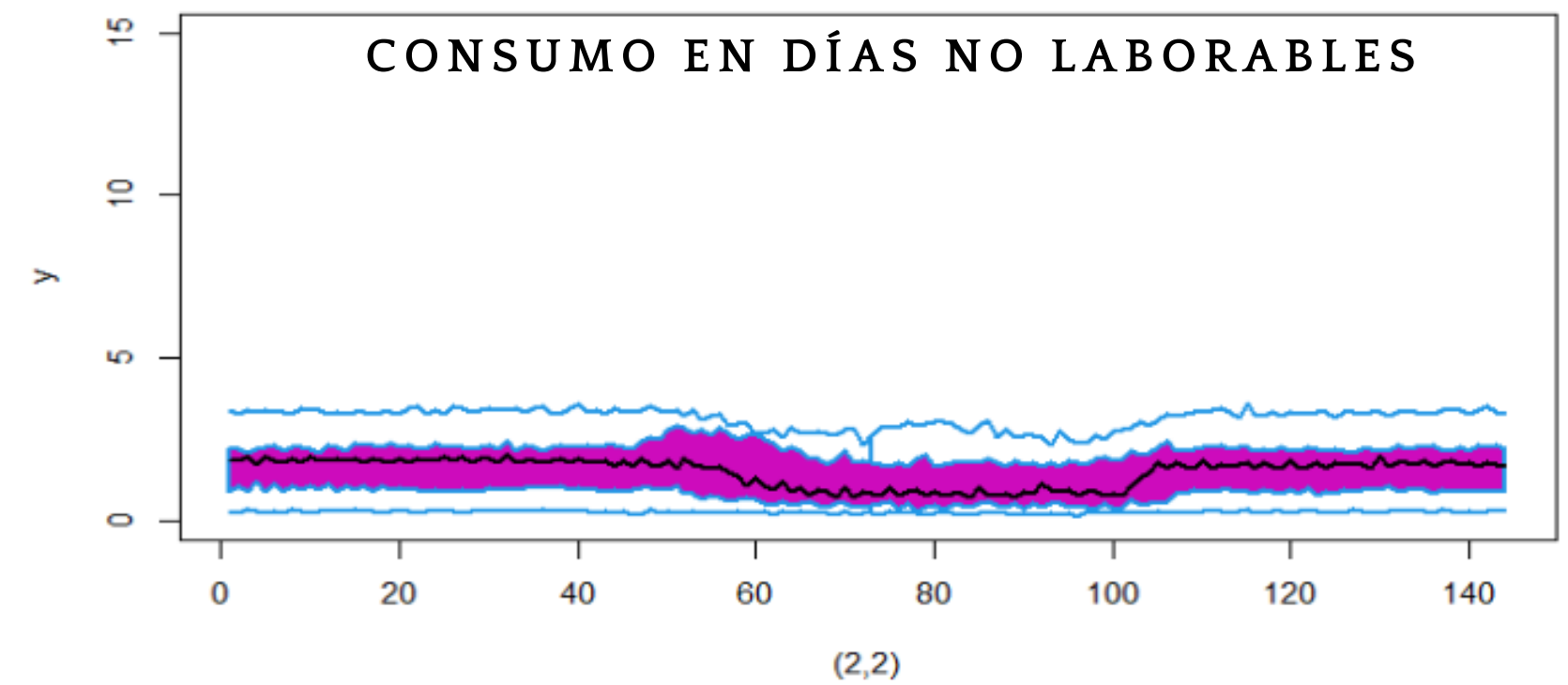
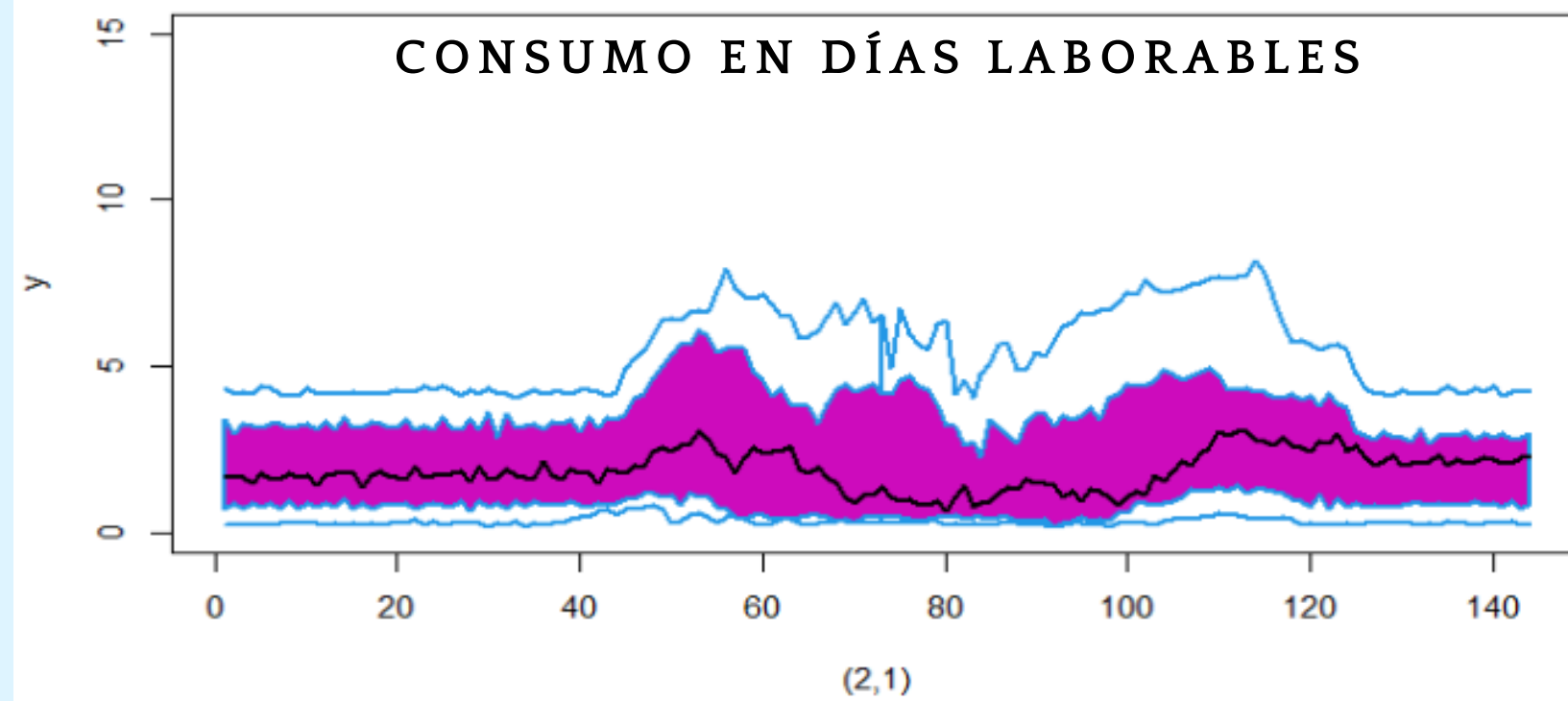
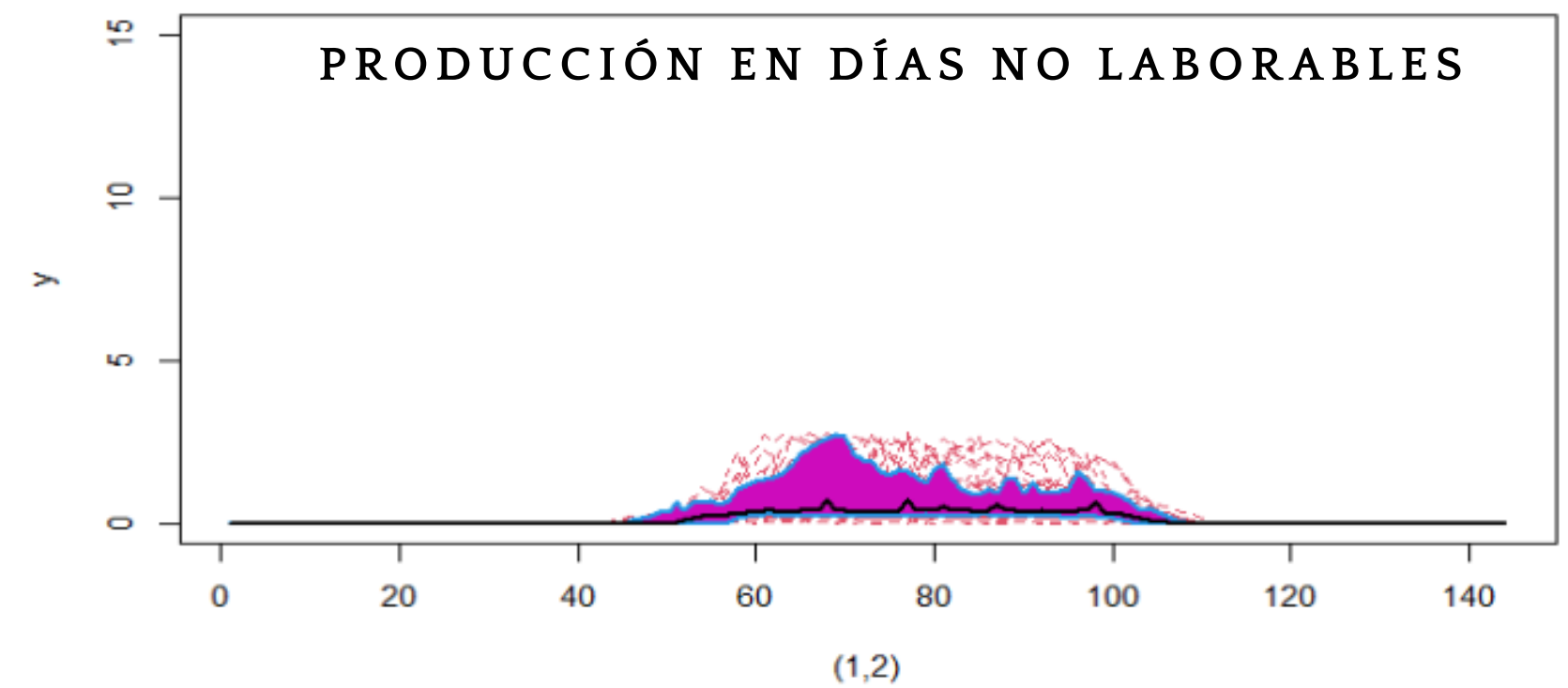
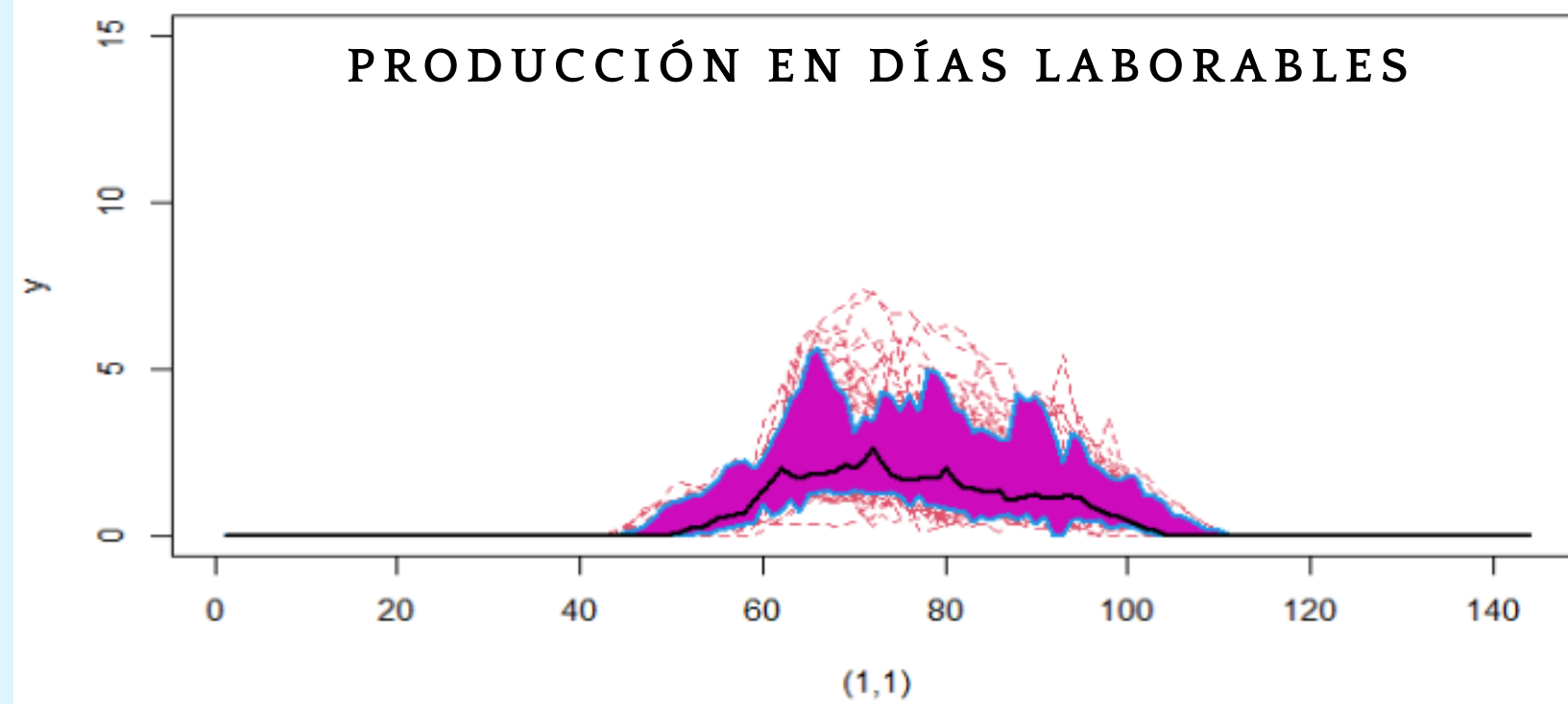
1. Boxplot funcional de los datos
2. Suavizado
3. Correlación de distancias
4. Detección de datos atípicos
5. Análisis de componentes principales

ANÁLISIS



Boxplot funcional de los datos de la Facultad de Derecho.

ANÁLISIS

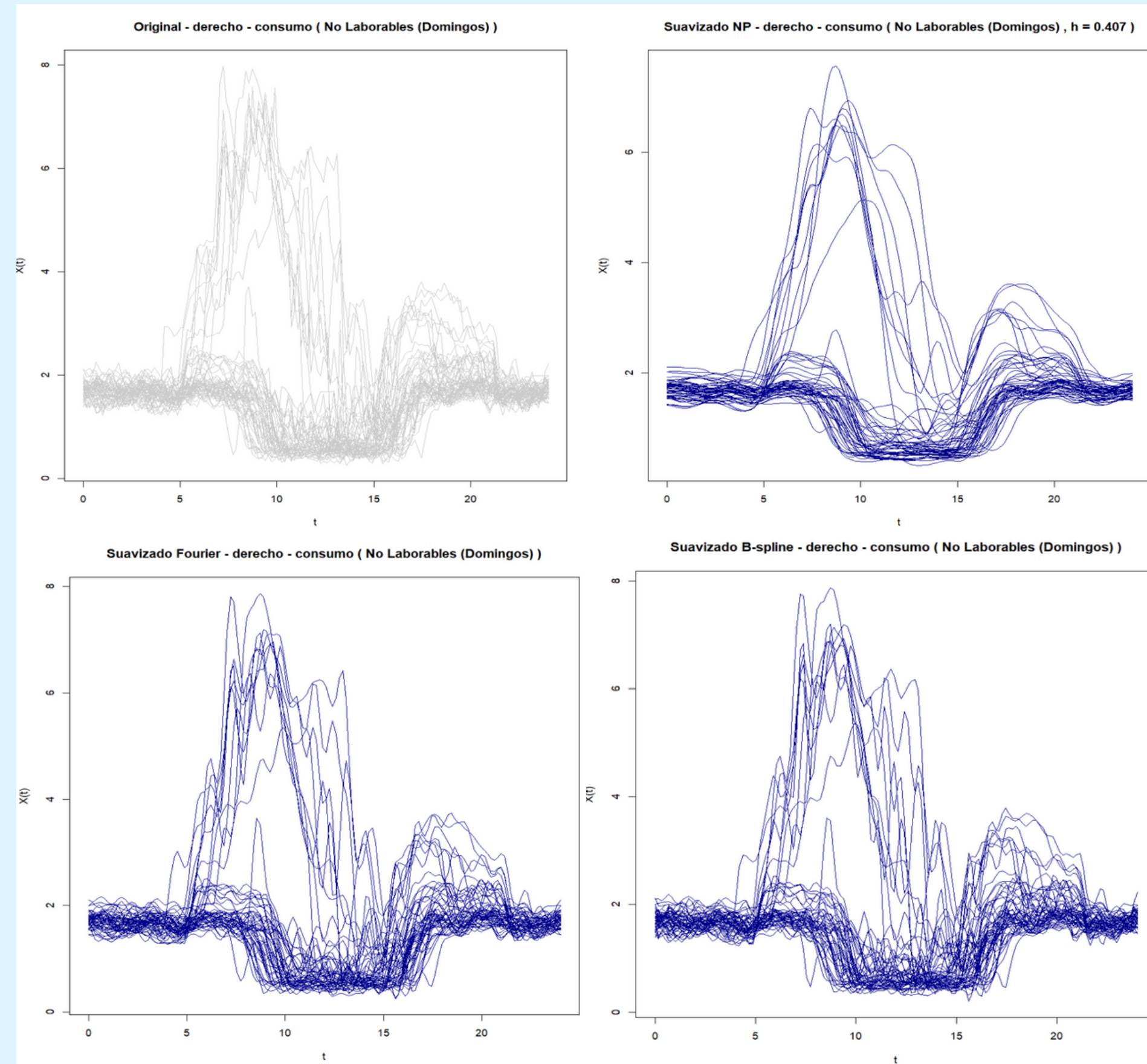


Boxplot funcional de los datos de la Escuela de Arquitectura.

ANÁLISIS: Suavizado

Se prueba con 3 técnicas de **suavizado**:

- Suavizado no paramétrico mediante un estimador tipo kernel (Nadaraya-Watson)
- Suavizado con base B-splines
- Suavizado con base de Fourier



ANÁLISIS: Suavizado

Se prueba con 3 técnicas de **suavizado**:

- Suavizado no paramétrico mediante un estimador tipo kernel (Nadaraya-Watson)
- Suavizado con base B-splines
- Suavizado con base de Fourier

Edificio <chr>	Tipo <chr>	Día <chr>	MSE_NP <dbl>	MSE_Basis <dbl>	MSE_Fourier <dbl>	Mejor <chr>
Derecho	Producción	Laborables	0.17291	0.06180	0.05465	fourier
Derecho	Consumo	Laborables	0.28795	0.09676	0.10566	basis
Derecho	Producción	No Laborables	0.02168	0.01156	0.00911	fourier
Derecho	Consumo	No Laborables	0.02458	0.00713	0.00903	basis
Arquitectura	Producción	Laborables	0.03495	0.01046	0.00913	fourier
Arquitectura	Consumo	Laborables	0.05151	0.02931	0.02791	fourier
Arquitectura	Producción	No Laborables	0.00583	0.00380	0.00350	fourier
Arquitectura	Consumo	No Laborables	0.01421	0.01506	0.01445	np

MSE de las diferentes técnicas de **suavizado**

```
# Suavizado mediante Fourier
fd_fourier <- suavizar_fourier(fdata_obj, nbasis = 7)

# Comparación con B-splines
fd_bspline <- fda.usc::fdata2fd(fdata_obj, nbasis = 10, type.basis = "bspline")

# Visualización
ggplot() +
  geom_line(aes(x = tiempo, y = fd_fourier[1, ]), color = "blue") +
  geom_line(aes(x = tiempo, y = fd_bspline[1, ]), color = "red")
```

Esta sección de código realiza el suavizado de datos mediante la transformada de Fourier y la comparación con B-splines. Finalmente, se presenta una visualización que permite observar las diferencias entre ambos métodos de suavizado utilizando un gráfico en ggplot.

ANÁLISIS: Correlación entre variables

Correlación de Pearson: mide relaciones lineales entre variables, pero no detecta dependencias no lineales.

ANÁLISIS: Correlación entre variables

Correlación de Pearson: mide relaciones lineales entre variables, pero no detecta dependencias no lineales.

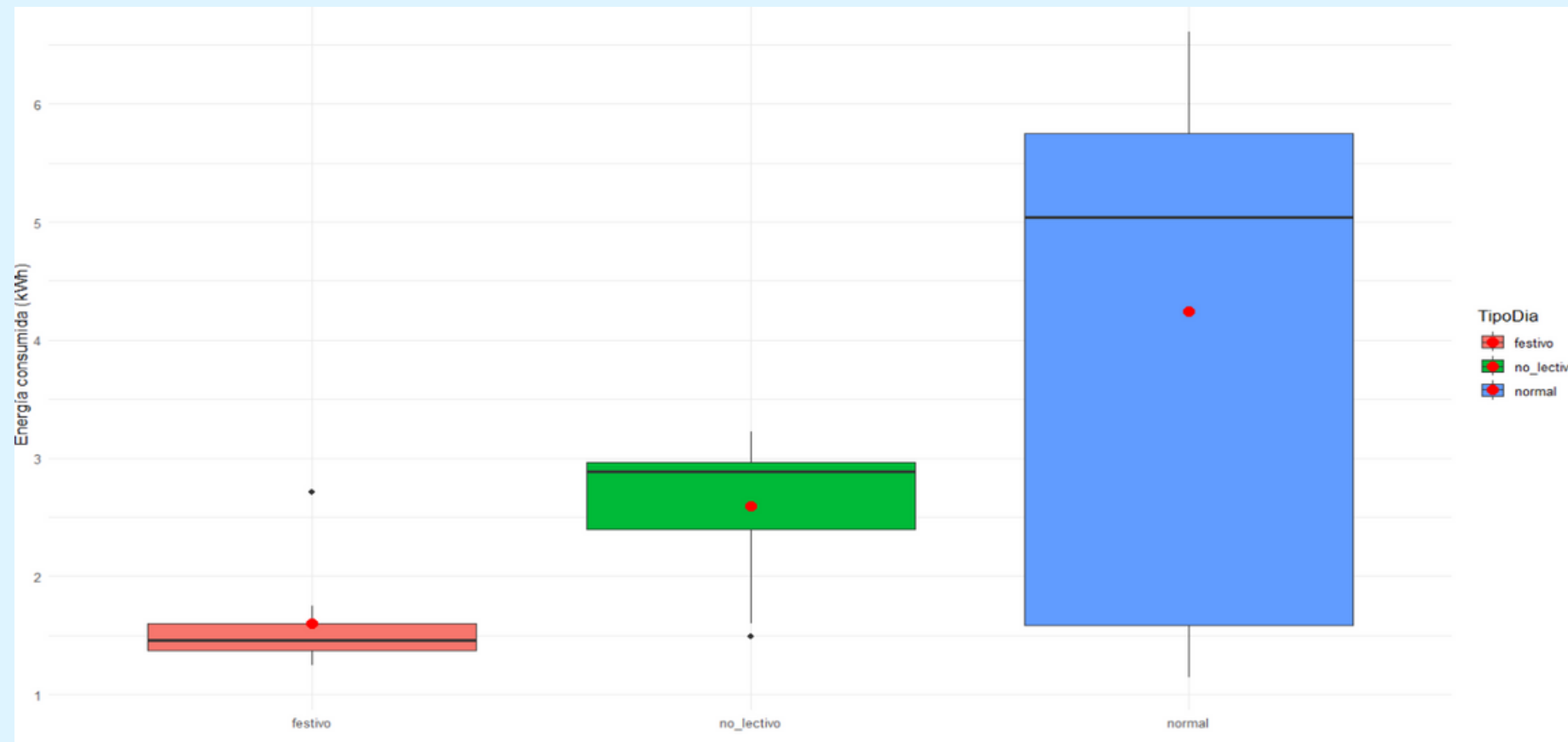
Correlación de distancia (Székely et al., 2007) detecta cualquier tipo de dependencia, lineal o no.

VENTAJAS:

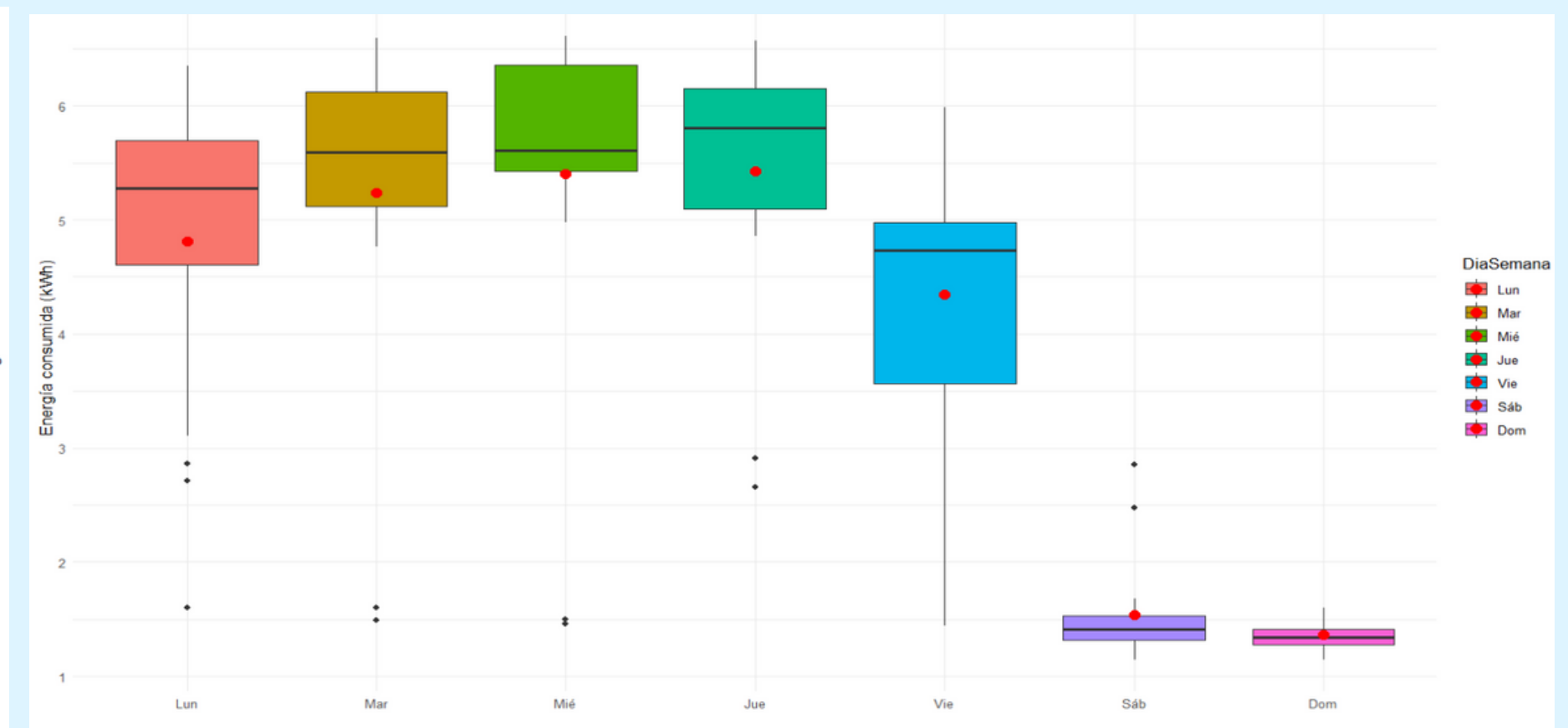
- Adecuada para datos funcionales y de alta dimensión
- Puede detectar independencia
- Compatible con variables categóricas y nominales
- Permite evaluar qué variables tienen mayor asociación con la variable respuesta

ANÁLISIS: Correlación de distancias

Consumo energético por tipo de día.
dcor: 0.137



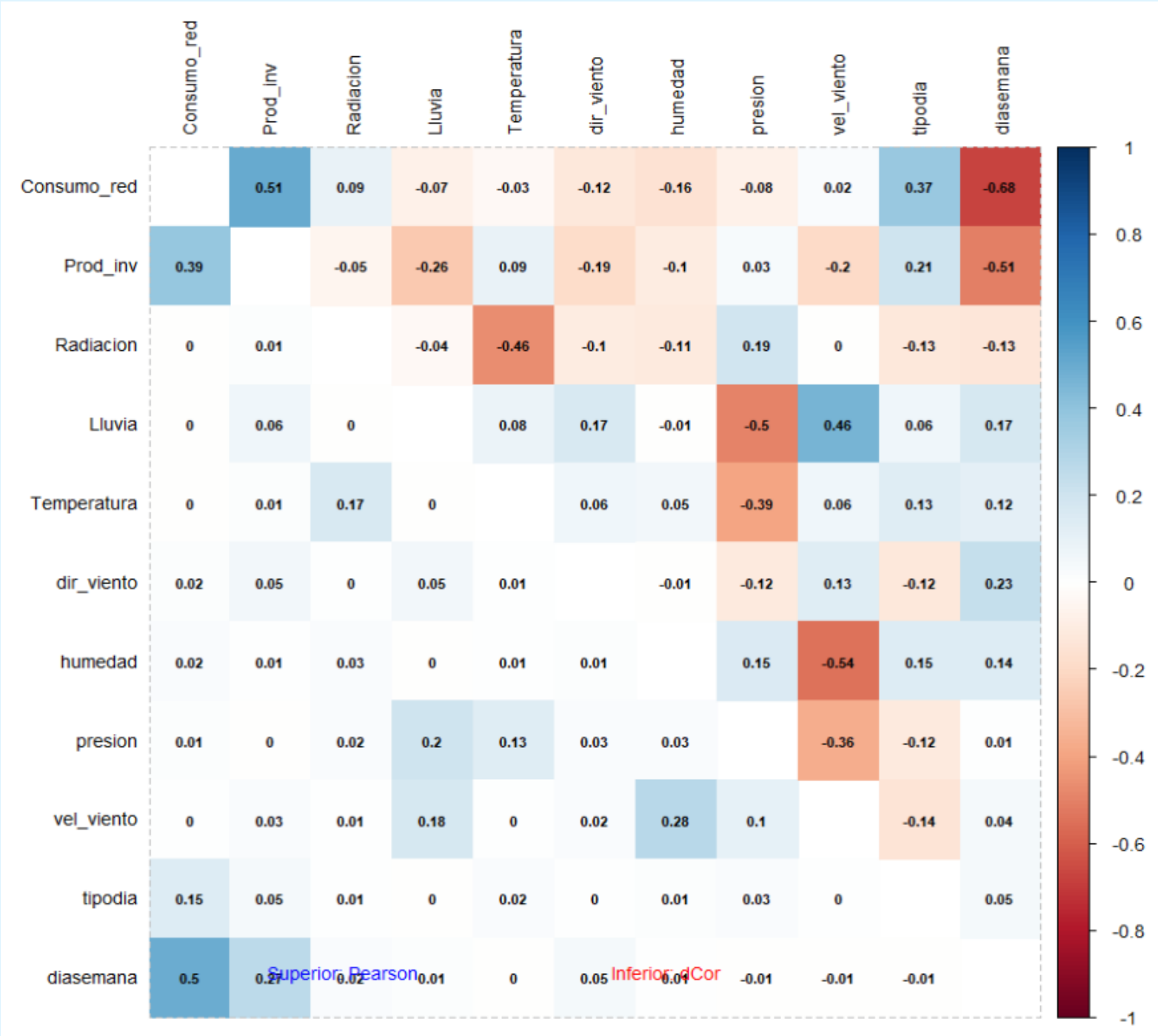
Consumo energético por día de la semana.
dcor: 0.504



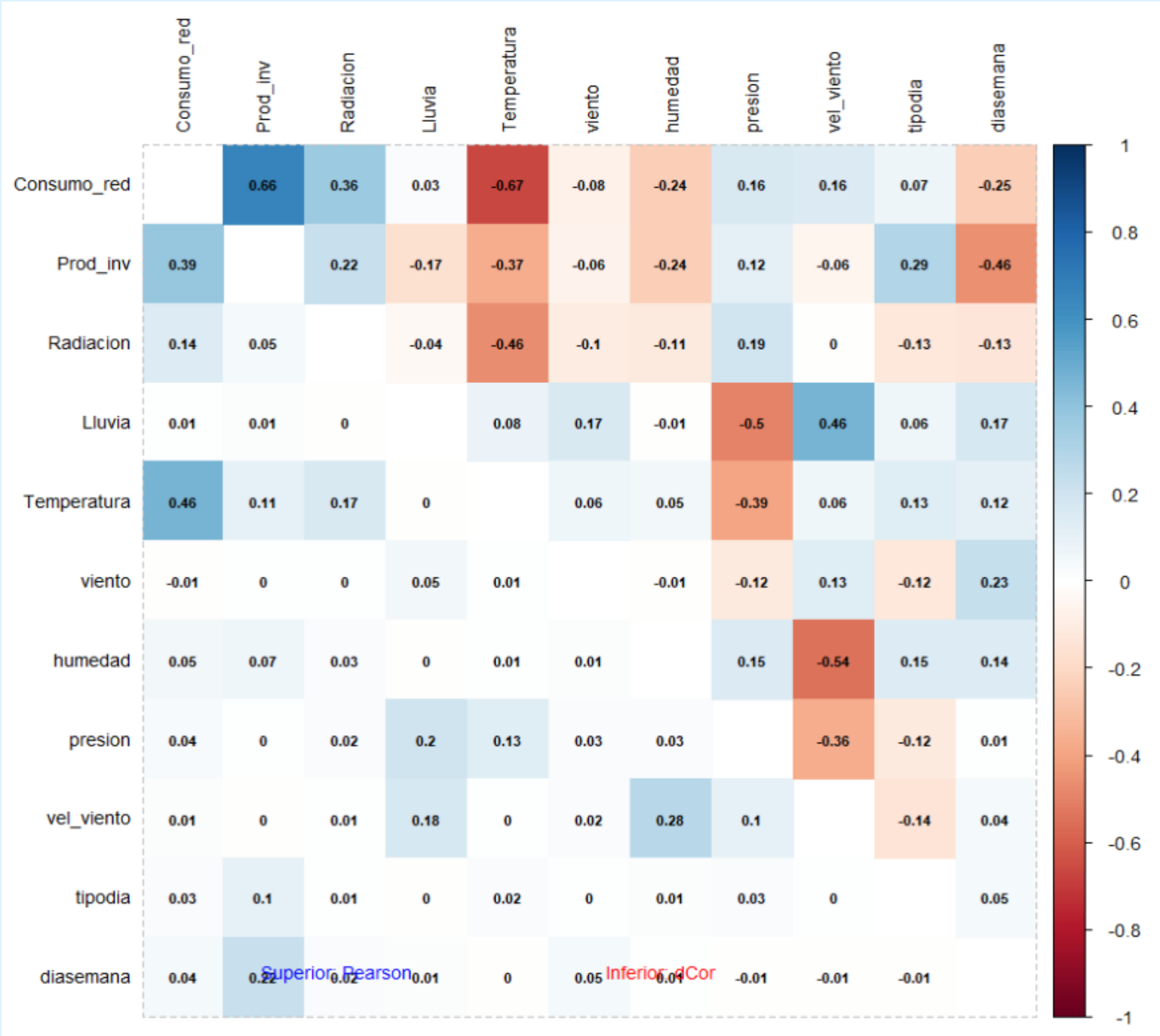
ANÁLISIS: Correlación de distancias

Diagonal superior: correlación de Pearson

Diagonal inferior: correlación de distancias



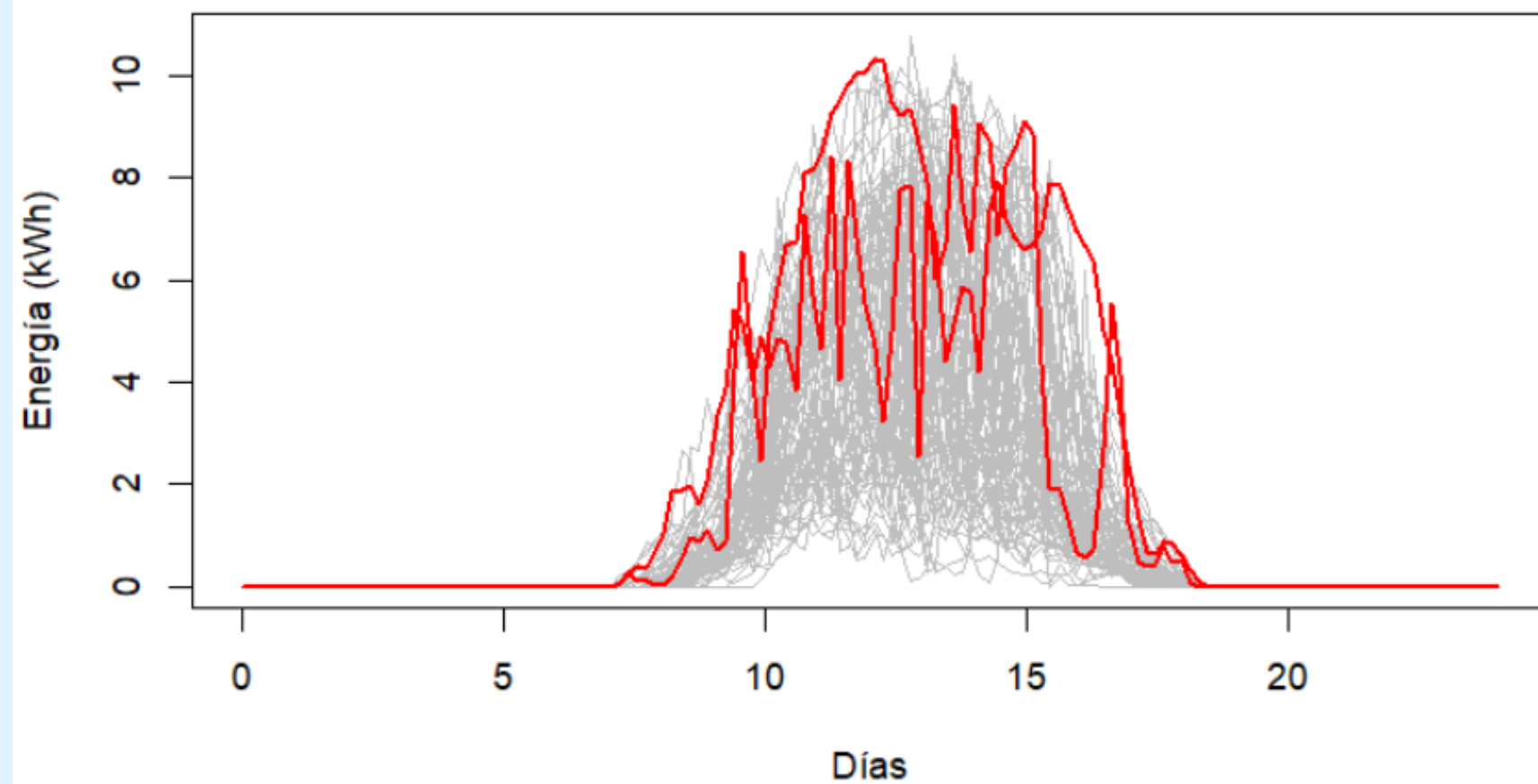
Matriz de correlación de distancias y Pearson para la Facultad de Derecho



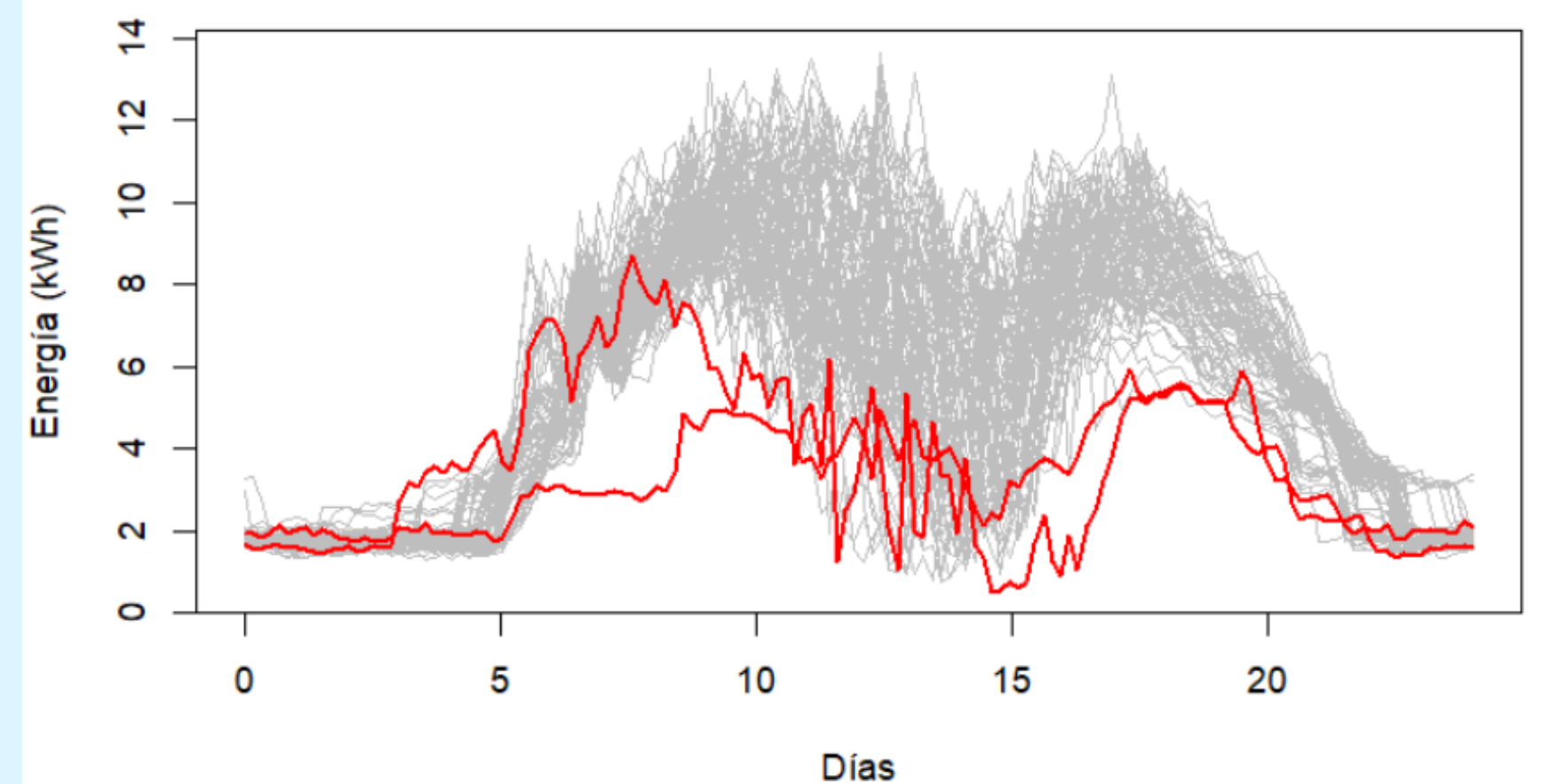
Matriz de correlación de distancias y Pearson para la Escuela de Arquitectura

ANÁLISIS: Detección de datos atípicos

Energía producida_inversores en datos



Energía consumida_red en datos



Datos atípicos en la Facultad de Derecho para la energía producida en días laborables. Los días 29 de enero y 3 de febrero.

Datos atípicos en la Facultad de Derecho para la energía consumida en días laborables. Los días 16 de diciembre y 29 de enero.

Funciones propias

- `detectar_y_visualizar_atipicos(...)` → boxplots funcionales + marcado de días atípicos.
- `extraer_outliers(...)` → devuelve la lista de días atípicos detectados.
- `detectar_atipicos_tabla(...)` → resumen en formato tabla de outliers por variable/edificio.
- `añadir_info_meteorologica(tabla_outliers)` → cruza outliers energéticos con datos meteo.

Paquetes y funciones clave

- `fda.usc::fdata` → `depth.trim`, `depth.mode`

CURVAS ENERGÉTICAS (FDATA)



DETECTAR_Y_VISUALIZAR_ATIPICOS



EXTRAER_OUTLIERS + TABLAS (DT)



AÑADIR_INFO_METEOROLOGICA

ANÁLISIS: Análisis de componentes principales

Métrica	Derecho Laborables	Derecho No Laborables	Arquitectura Laborables	Arquitectura No Laborables
<i>Varianza CP1 (%)</i>	63.06	73.54	72.47	76.66
<i>Varianza CP2 (%)</i>	9.60	6.82	6.31	10.49
<i>Total (2 CPs) (%)</i>	72.65	80.35	78.78	87.15

Varianza explicada en la producción energética

Métrica	Derecho Laborables	Derecho No Laborables	Arquitectura Laborables	Arquitectura No Laborables
<i>Varianza CP1 (%)</i>	49.12	84.47	81.16	89.46
<i>Varianza CP2 (%)</i>	10.63	5.33	5.22	5.45
<i>Total (2 CPs) (%)</i>	59.75	89.80	86.38	94.91

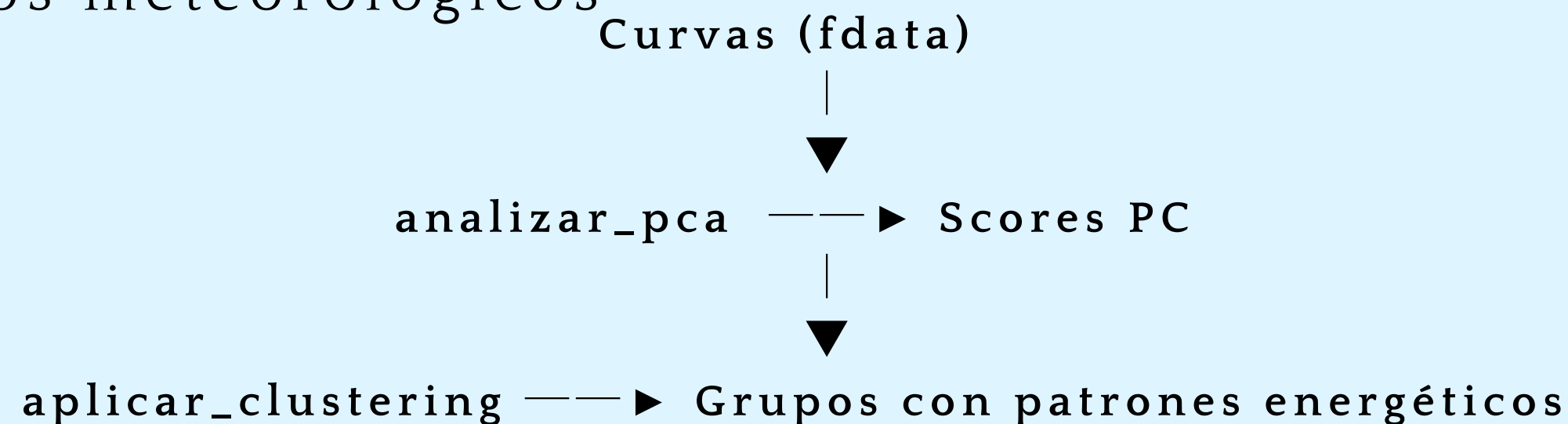
Varianza explicada en el consumo energético

Funciones propias

- `analizar_pca(fdata_obj, nombre_analisis)` → aplica análisis de componentes principales funcionales (FPCA) y devuelve la varianza explicada y las primeras componentes.
- `aplicar_clustering(obj_pc, datos_meteo, k = 3, ...)` → clustering sobre los scores de PCA, integrando datos meteorológicos como contexto.

Paquetes y funciones clave

- `fda.usc` → FPCA (`fdata2pc`, `summary.fdata2pc`, etc.).
- `factoextra` → visualización de resultados de PCA y clustering (biplots, scree plots, dendrogramas).
- `stats::kmeans` → clustering en el espacio reducido de scores (otras alternativas: `hclust`, `pam`).



MODELADO

1. Modelos de regresión
2. Boosting

MODELADO: Regresión funcional

- Modelo lineal
- Modelo aditivo

Se prueban 3 configuraciones diferentes para cada modelo

1. Modelo sin datos funcionales
2. Modelo con datos funcionales sin tipo día ni día de la semana
3. Modelo con datos funcionales completo con tipo de día y día de la semana

$$y \sim s(\text{Radiación}) + s(\text{Lluvia}) + s(\text{Direccion_viento}) + s(\text{Presion}) + s(\text{Velocidad_viento}) + \text{TipoDia} + \text{DiaSemana}$$

MODELADO: Regresión funcional

- Modelo lineal
- Modelo aditivo

Se prueban 3 configuraciones diferentes para cada modelo

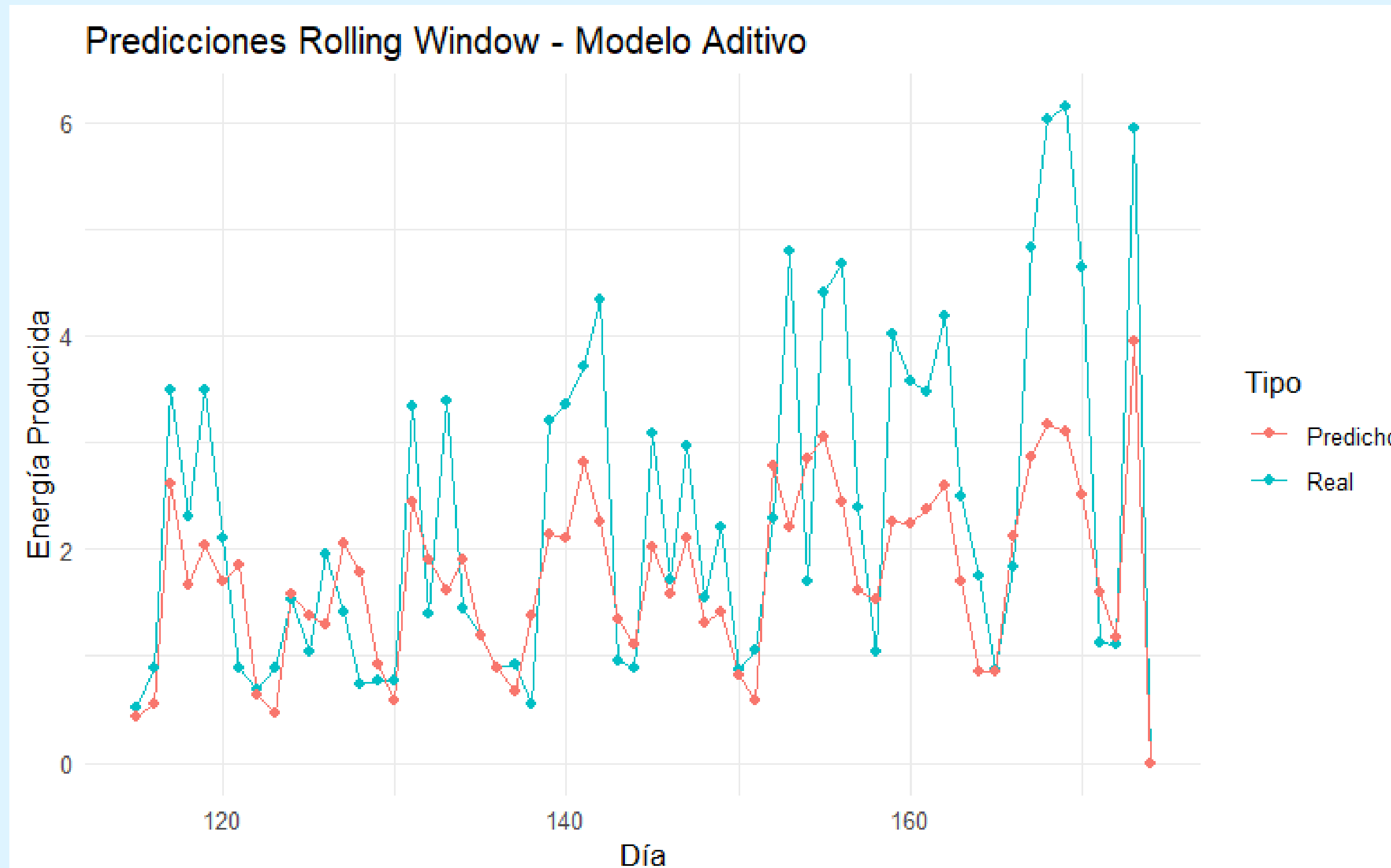
1. Modelo sin datos funcionales
2. Modelo con datos funcionales sin tipo día ni día de la semana
3. Modelo con datos funcionales completo con tipo de día y día de la semana

$$y \sim s(\text{Radiación}) + s(\text{Lluvia}) + s(\text{Direccion_viento}) + s(\text{Presion}) + s(\text{Velocidad_viento}) + \text{TipoDia} + \text{DiaSemana}$$

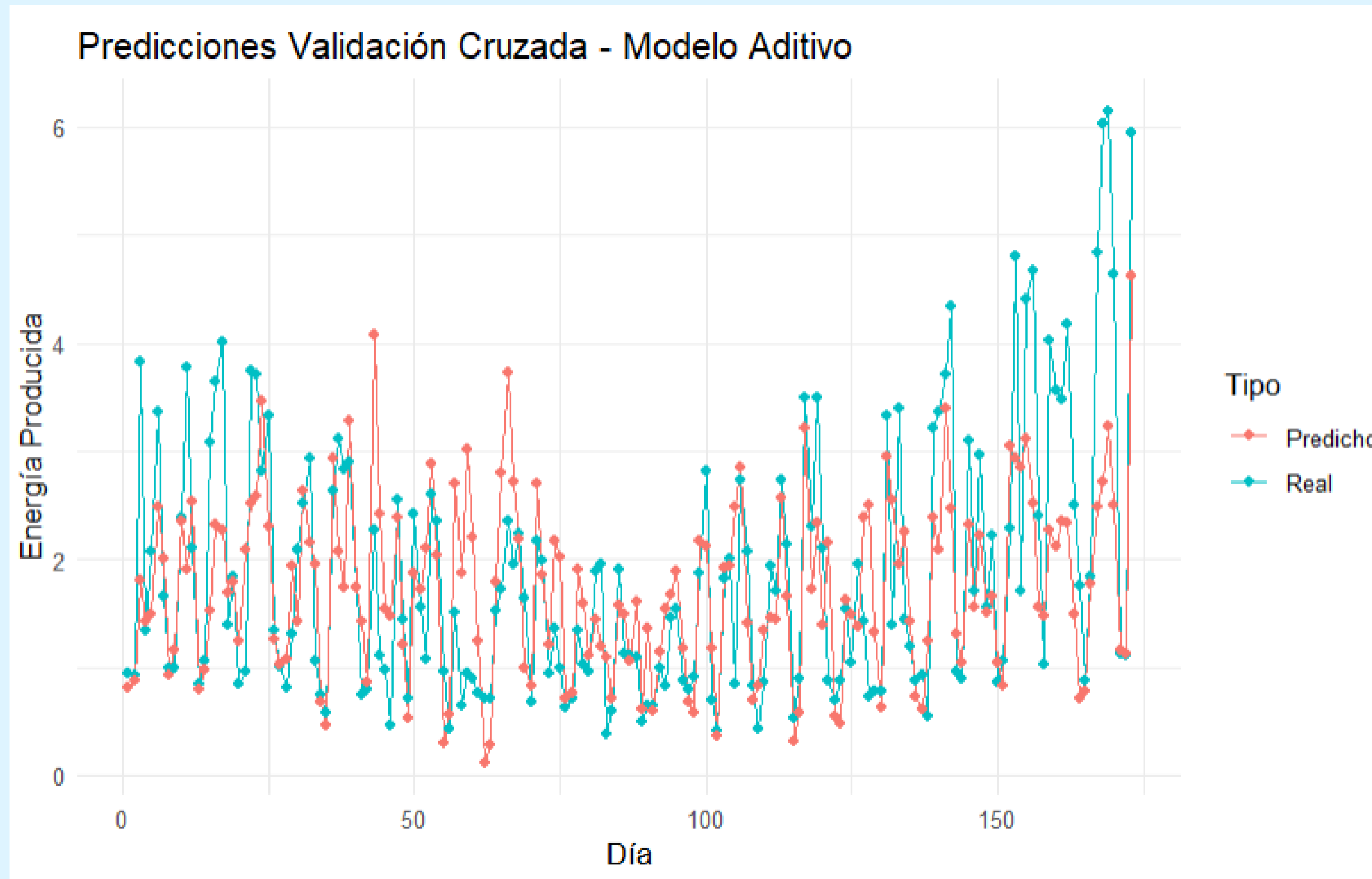
Dos estrategias de validación:

1. Rolling window
2. Validación cruzada por bloques

MODELADO: Modelos de regresión



MODELADO: Modelos de regresión



Funciones propias

Preparación de datos

- `escalar_datos_entrenamiento(df)`, `escalar_datos_prediccion(df, ...)` → normalización y escalado de variables.
- `procesar_datos_cv(...)`, `procesar_datos(...)` → generación de estructuras para validación y ajuste (incluido GSAM).

Modelado y evaluación

- `predecir_modelo_gsam_vs(...)` → comparación de modelos aditivos funcionales.
- `evaluar_predicciones(...)` → métricas (RMSE, MAE, R^2).
- `realizar_prediccion(modelo, datos_prediccion, ...)` → estimaciones sobre nuevos días.

Validación

- Cross-validation: `cv_lm(...)`, `cv_gam(...)`, `cv_fregre_lm(...)`, `cv_fregre_gam(...)`, `cv_fregre_mixed(...)`, `cv_fregre_gam_mixed(...)`.
- Rolling funcional: `rolling_lineal`, `rolling_gsam(...)` (aditivos/funcionales).

Paquetes

- `fda.usc` → `fregre.lm`, `fregre.gsam` (modelos funcionales lineales y aditivos).
- `mgcv` → `gam`, `s` (suavizadores aditivos, base spline).

MODELADO: Boosting

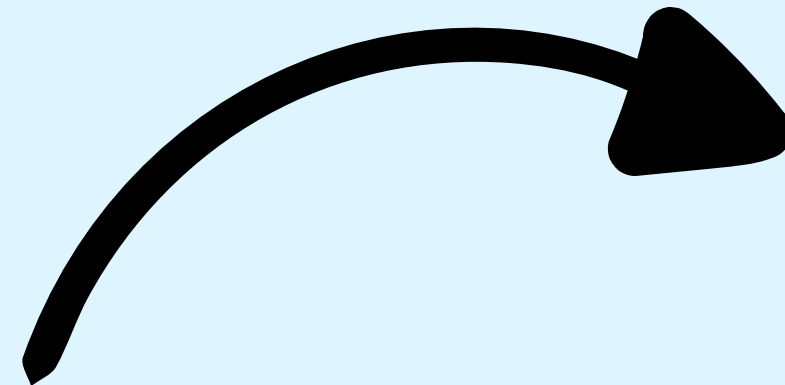
Problema con modelos de
regresión

1. Posible sobreajuste en
los modelos más
complejos

MODELADO: Boosting

Problema con modelos de regresión

1. Posible sobreajuste en los modelos más complejos

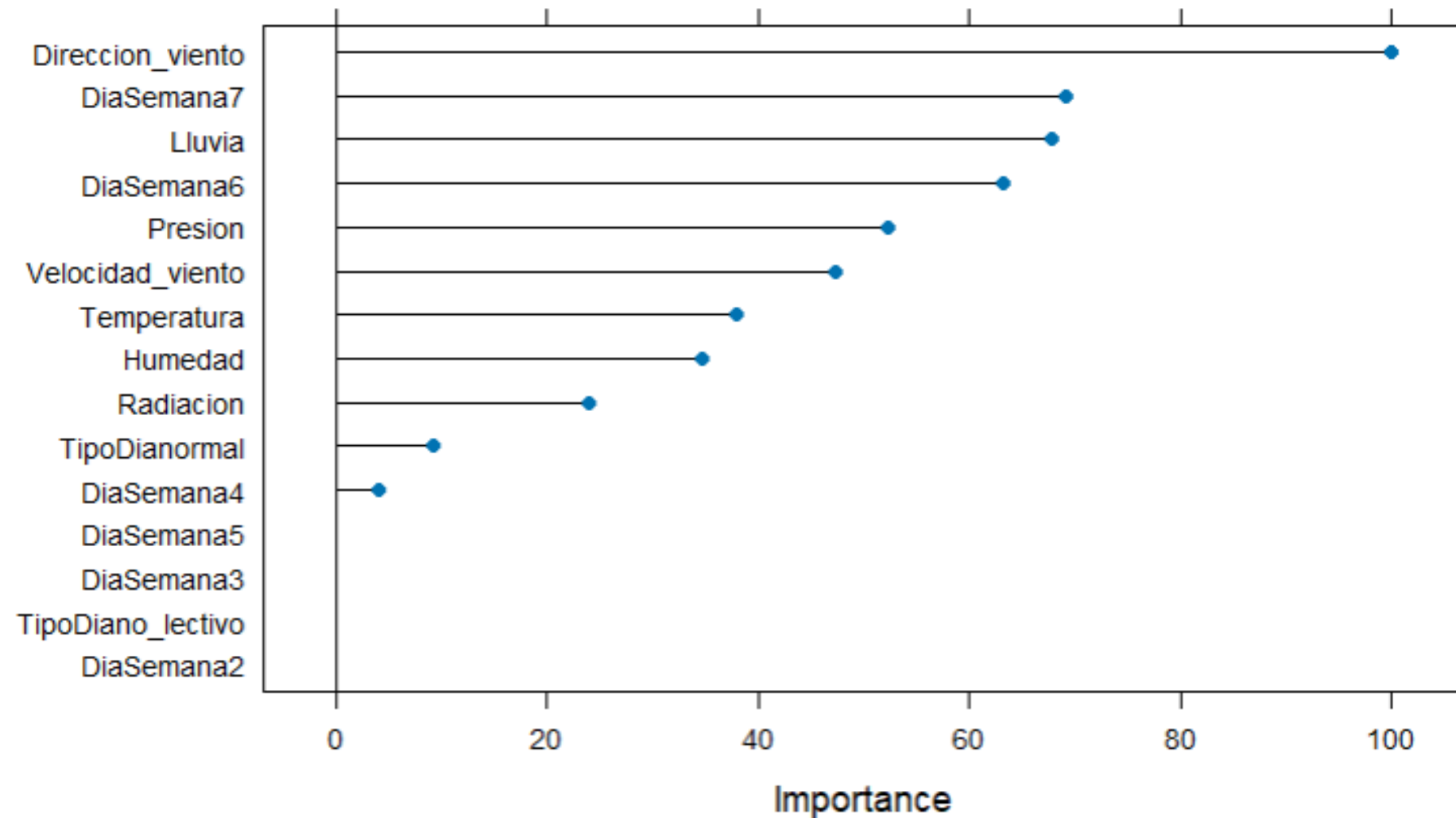


SOLUCIÓN: Boosting

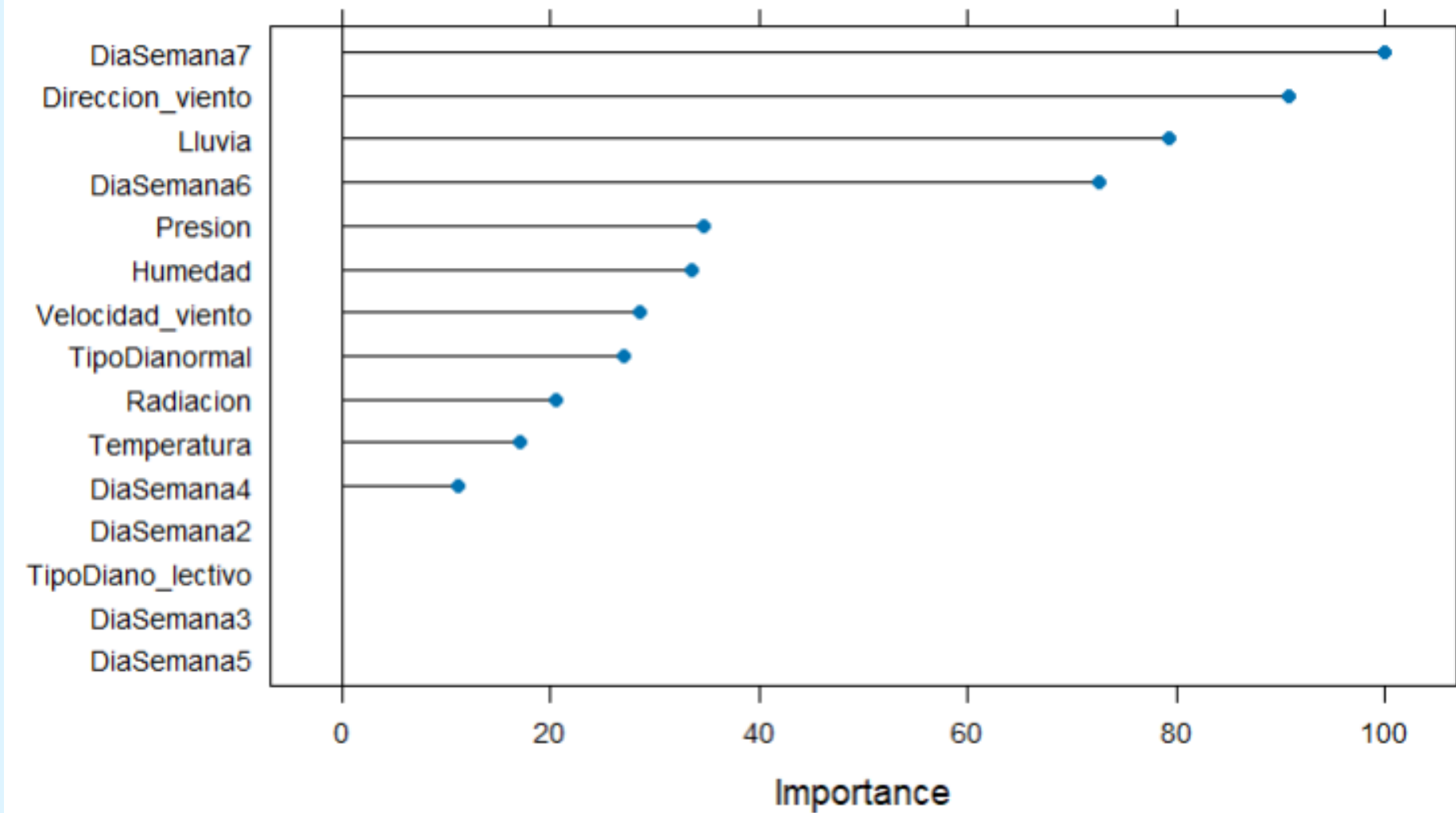
- Se utilizan penalizaciones explícitas por la complejidad del modelo para evitar el sobreajuste.
- Modelos se adaptan mejor a los datos.
- Manejan relaciones no lineales complejas.
- Manejan combinaciones entre predictores.
- Robustos frente a valores atípicos y problemas de multicolinealidad.
- Permiten evaluar la importancia relativa de cada variable.

MODELADO: Boosting

Importancia de variables - GBM

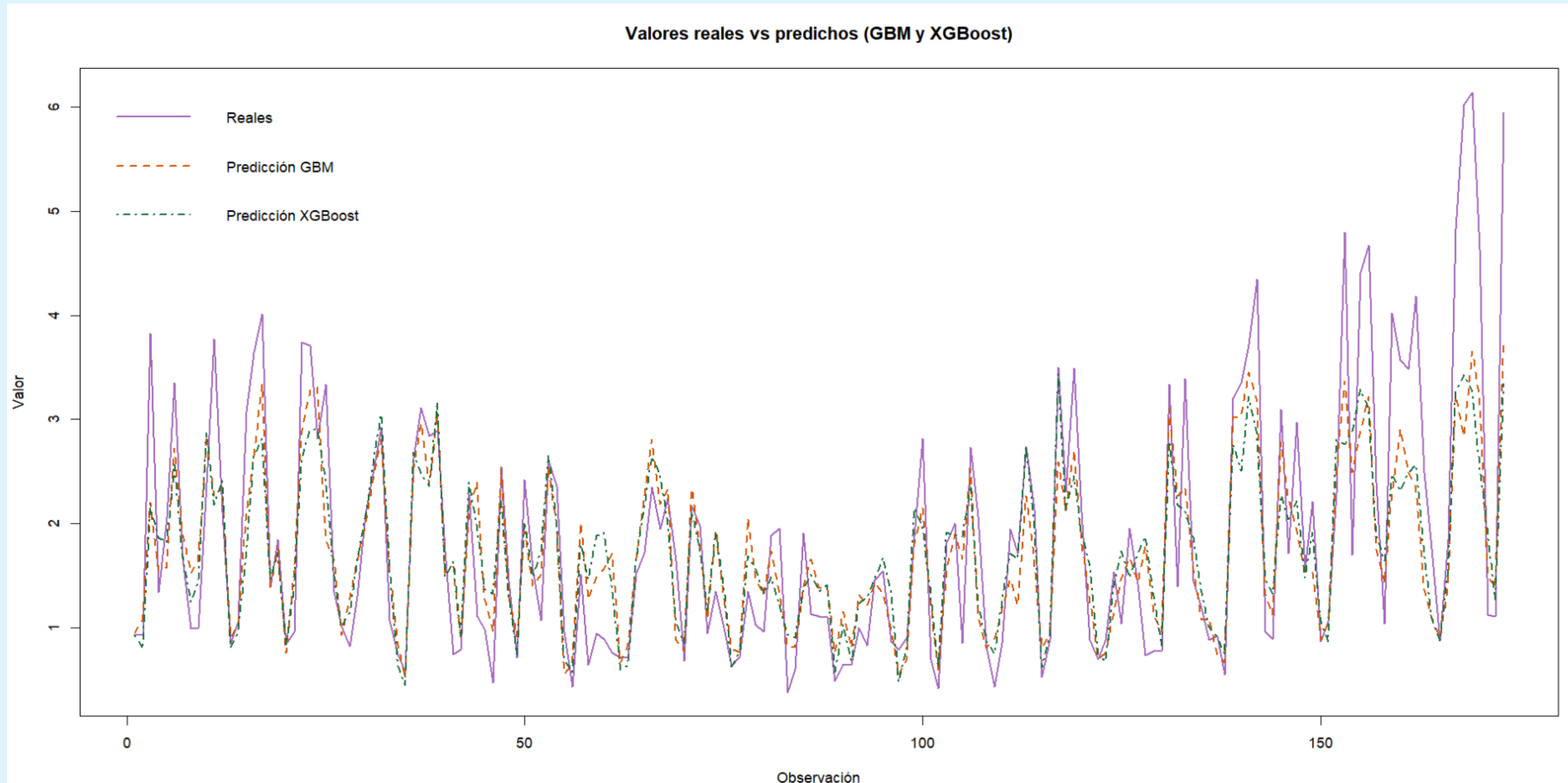


Importancia de variables - XGBoost



Variables relevantes para la Facultad de Derecho

MODELADO: Boosting



GBM

RMSE: 0.6831

MAE: 0.4767

R^2 : 0.7546

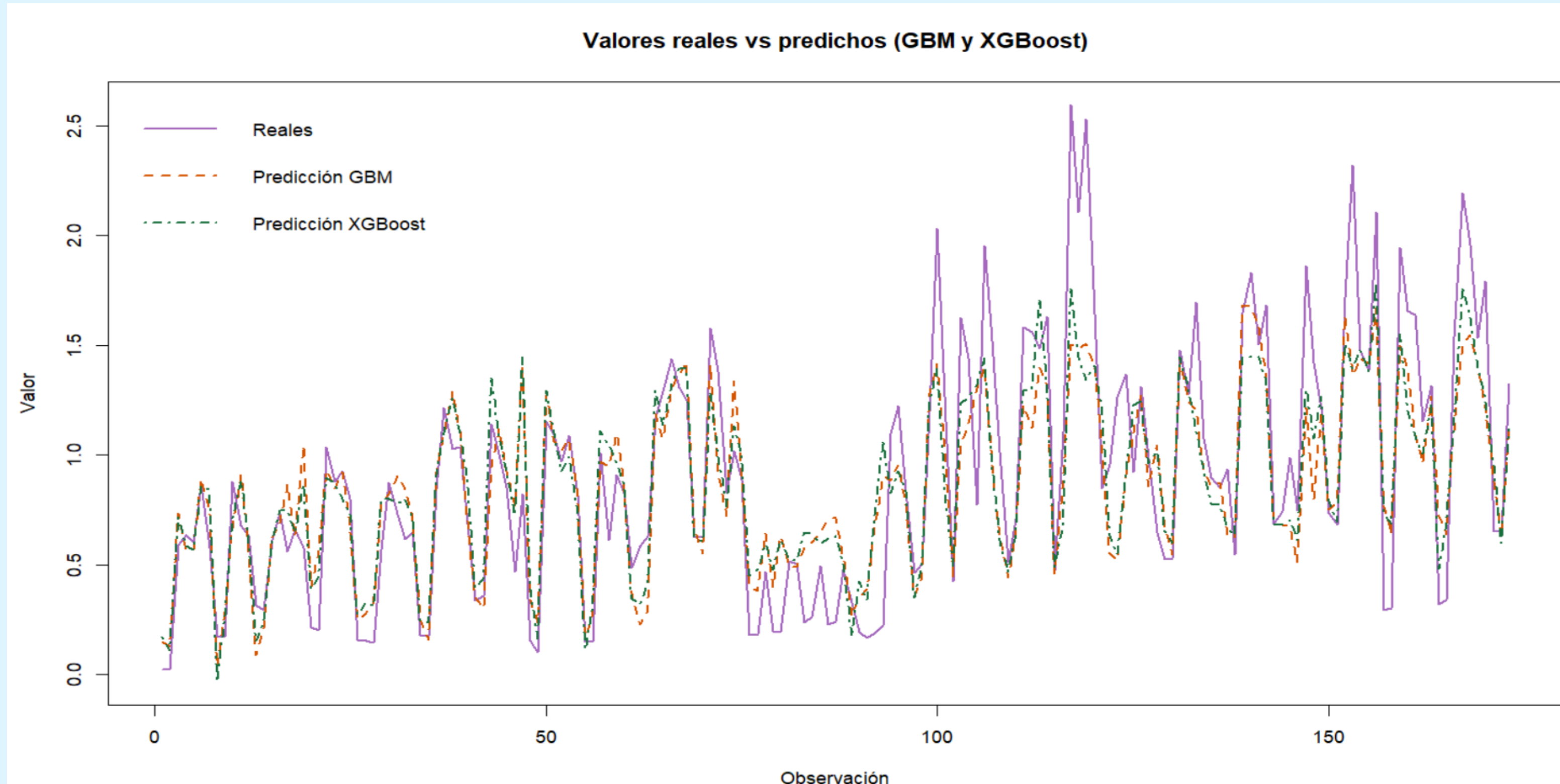
XGBoost

RMSE: 0.7226

MAE: 0.5023

R^2 : 0.7237

MODELADO: Boosting



GBM

RMSE: 0.3017

MAE: 0.2189

R^2 : 0.7402

XGBoost

RMSE: 0.2930

MAE: 0.2152

R^2 : 0.7552

Funciones propias

- **preparar_datos_boosting(...)** → genera matrices de entrenamiento/test adaptadas a boosting.
- **entrenar_gbm(...)** → entrenamiento de modelos Gradient Boosting Machine (paquete gbm).
- **entrenar_xgboost(...)** → entrenamiento de modelos XGBoost (paquete xgboost).
- **evaluar_modelo_boosting(...)** → métricas de rendimiento (RMSE, MAE, R^2).
- **analizar_boosting(...)** → análisis comparativo entre modelos de boosting.
- **visualizar_importancia(modelo, tipo = "gbm" | "xgboost")** → importancia de variables.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

- Patrones difieren entre los edificios debido a sus distintas dinámicas.
- Identificamos que el sistema fotovoltaico con vertido cero limita la producción en ausencia de consumo.
- Análisis de datos funcionales resultó ser la elección correcta.
- Modelos funcionales interpretables (se puede medir efecto).
- Modelos boosting con buena capacidad predictiva.
- Buenos resultados en la evaluación de los modelos.

RECOMENDACIONES

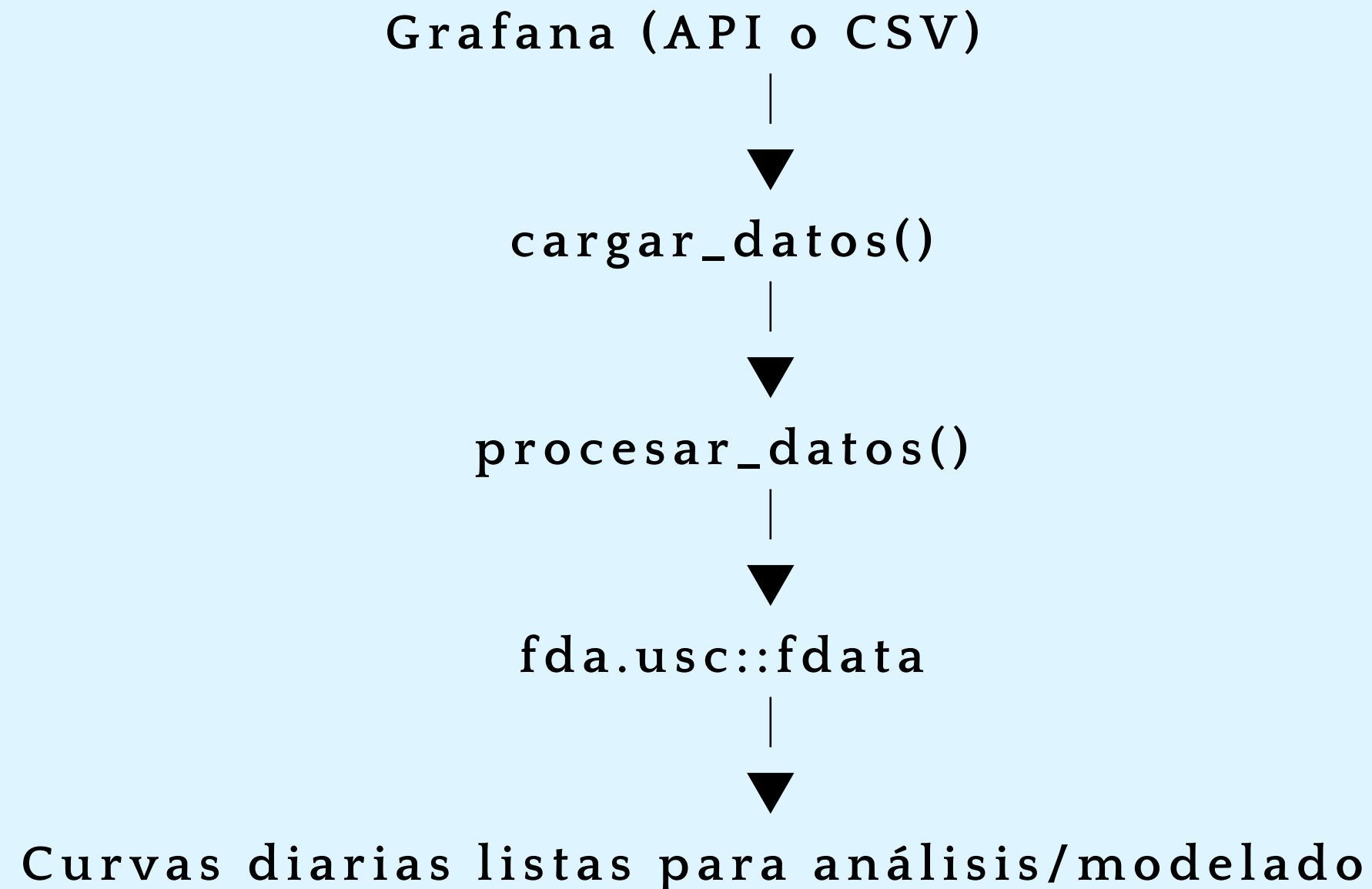
- Estrategias de gestión energética basadas en estos hallazgos.
- Considerar sistemas de vertido de red controlado.
- Modelos predictivos desarrollados podrían integrarse en un sistema de monitorización.

TRABAJO FUTURO

- Ampliar el análisis a más edificios de la universidad para validar las conclusiones en distintas condiciones.
- Ampliar el análisis para tener una respuesta funcional.
- Investigar más a fondo los datos atípicos con información adicional.
- Generar un sistema de recomendaciones de eficiencia energética.
- Aplicación estilo Grafana que avise de atípicos, alertas y recomendaciones.
- Implementar y documentar la función `grafana2fdata()`, que automatice la consulta de datos desde Grafana y su conversión en objetos funcionales (`fdata`) para integrarlos directamente en análisis y modelos predictivos en R.

grafana2fdata()

- Unifica el preprocesado de datos de distintas fuentes o edificios.
- Facilita la reproducibilidad y la reutilización de pipelines funcionales en futuros proyectos.



BIBLIOGRAFÍA

- Febrero-Bande, M., & De La Fuente, M. O. (2012). Statistical computing in functional data analysis: The R package fda. usc. Journal of statistical Software, 51, 1-28.
- “Funcionamiento de las placas fotovoltaicas,” Iberdrola, 2024. [En línea]. Disponible en: <https://www.iberdrola.com/innovacion/como-funcionan-placas-solares-fotovoltaicas>
- Edelman, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the Pearson and the distance correlation coefficients. Statistics & probability letters, 169, 108960.
- Ferraty, F., Laksaci, A., & Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. Statistical Inference for Stochastic Processes, 9, 47-76.
- Ramsay, J. O., & Silverman, B. W. (2005). Principal components analysis for functional data. Functional data analysis, 147-172.
- Fernández-Casal, R., Costa, J., & Oviedo de la Fuente, M. (2024). Métodos predictivos de aprendizaje estadístico.
- Febrero-Bande, M., González-Manteiga, W., & Oviedo De La Fuente, M. (2019). Variable selection in functional additive regression models. Computational Statistics, 34, 469-487.

GRACIAS POR SU ATENCIÓN

MARINA SÁNCHEZ VILLAVERDE