

Análisis de contaminantes atmosféricos con R

Rodríguez Nuñez M.

Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Departamento de estadística y biometría. Córdoba, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina (CONICET), Unidad de Fitopatología Y Modelización Agrícola (UFYMA). Córdoba, Argentina.

INTRODUCCIÓN

La contaminación del aire es un problema global que amenaza la salud humana y el medio ambiente. La exposición a altos niveles de contaminantes atmosféricos, como el material particulado fino ($PM_{2.5}$), causa millones de muertes anuales, según la Organización Mundial de la Salud (OMS). El objetivo de este trabajo fue demostrar la aplicabilidad de R en la creación de un sistema de alerta ante condiciones adversas de calidad de aire basado en variables satelitales. En el contexto de mi tesis doctoral aproveché herramientas provistas por la comunidad de R que permitieron desde la descarga y análisis de datos hasta la creación de modelos predictivos y la generación del documento del tesis.

METODOLOGÍA

1- Muestreo de concentración de $PM_{2.5}$.

2- Descarga de variables climáticas satelitales.

KrigR

Downloading and Downscaling of ERA5(-Land) data using R

DESARROLLO DE MODELOS PREDITIVOS PARA LA CONCENTRACIÓN DE MATERIAL PARTICULADO FINO EN ATMOSFERAS URBANAS

por

Rodríguez Nuñez Martín

Presentado ante la

Facultad de Ciencias Exactas, Físicas y Naturales

para obtener el grado de

Doctor en Ciencias de la Ingeniería

de la

Universidad Nacional de Córdoba

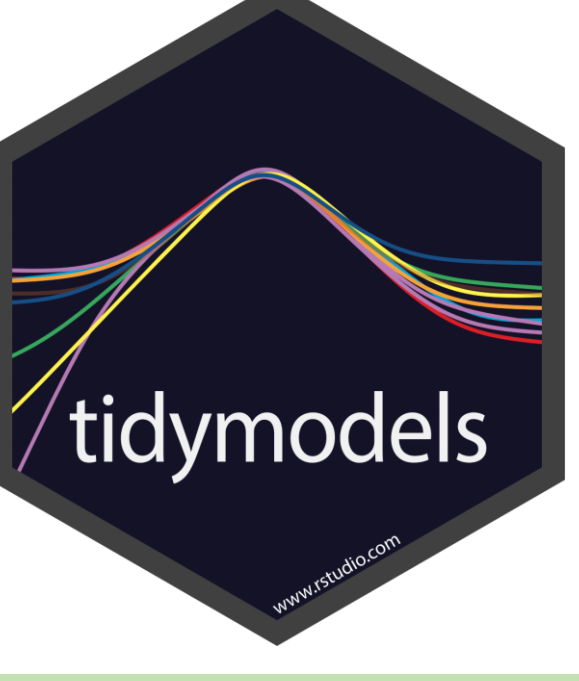
Septiembre 2023

Directora: Hely Alejandra Carreras

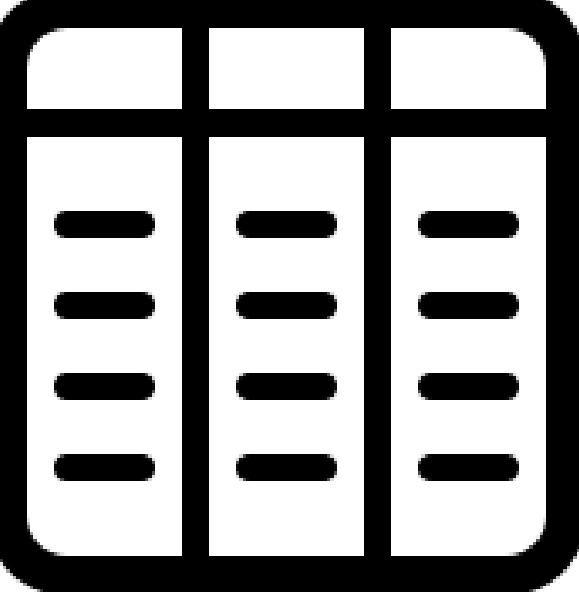
6- Generación del document de tesis



5- Modelado predictivo



3- Preprocesamiento y limpieza de datos



4- Análisis de datos

RESULTADOS

Figura 3.2: Ubicación específica de los sitios de muestreo dentro del área metropolitana de la provincia de Córdoba.

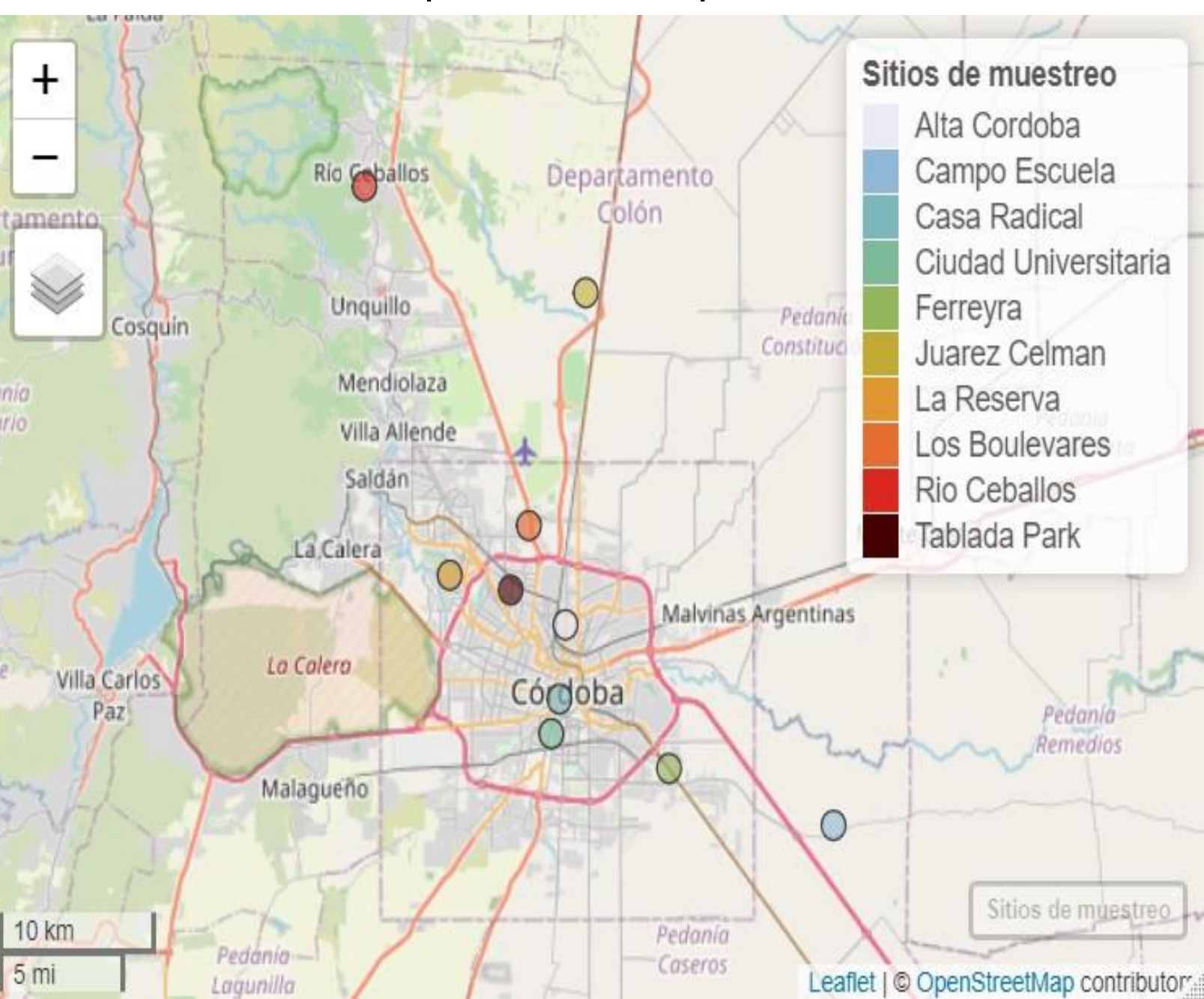


Figura 3.10: Diagrama de concentración de $PM_{2.5}$ promedio diario para los sitios del área metropolitana de la provincia de Córdoba a lo largo del año.

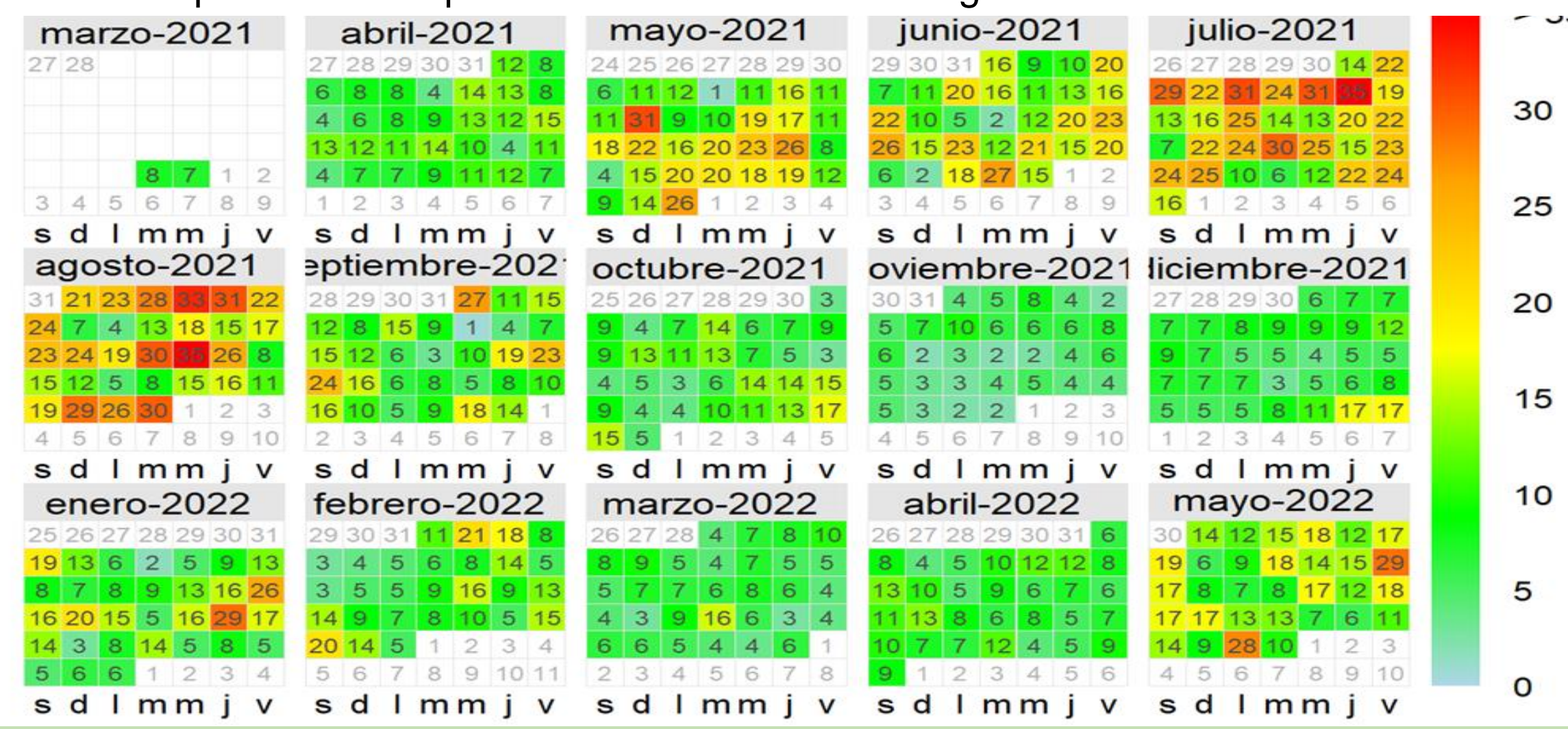


Tabla 5.2: Error cuadrático medio de predicción (RMSE) de modelos ajustados para predecir la concentración de $PM_{2.5}$ en el uso de suelo industrial a partir de 193 variables satelitales (base de datos original) y 53 variables seleccionadas (base de datos reducida).

Modelo / Base de datos	Original	Reducida
GRU	6.10	5.96
LSTM	6.19	6.17
RNN	6.72	6.59
Bosques aleatorios	6.52	6.54
KNN	6.24	6.26
SVM	7.87	8.02
GBM	9.34	9.53
Regresión lineal	10.10	10.33
Regresión SGD	10.11	10.24
Regresión Ridge	10.10	10.33
Regresión LASSO	10.92	10.93
Regresión de red elástica	11.82	11.82
Modelo de referencia	6.47	6.47

Figura 3.17: Perfil de concentración promedio horario de $PM_{2.5}$ a lo largo de los días, discriminado por los usos de suelo derivados del análisis de conglomerados.

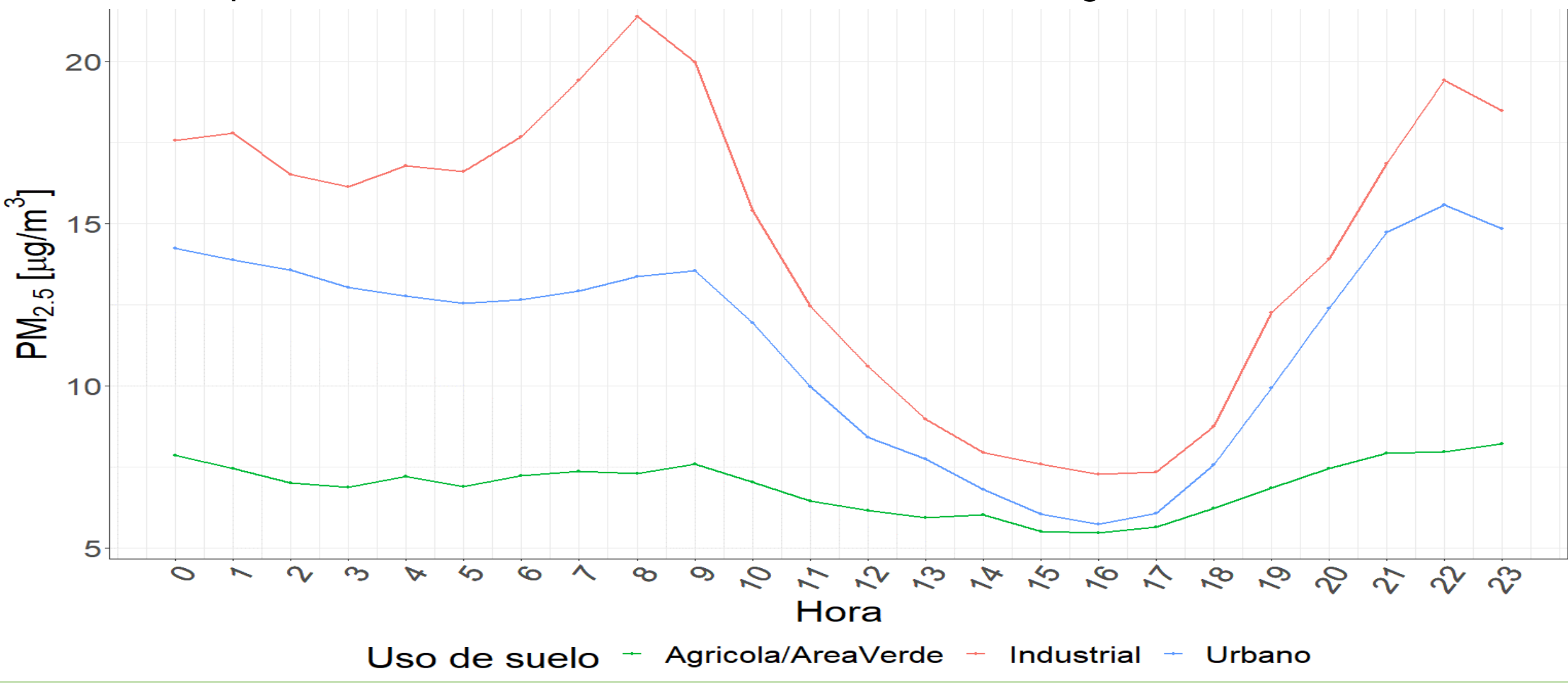


Figura 4.14: Perfil de variación conjunto de los promedios horarios del $PM_{2.5}$ y la radiación de corta longitud de onda a lo largo del día.

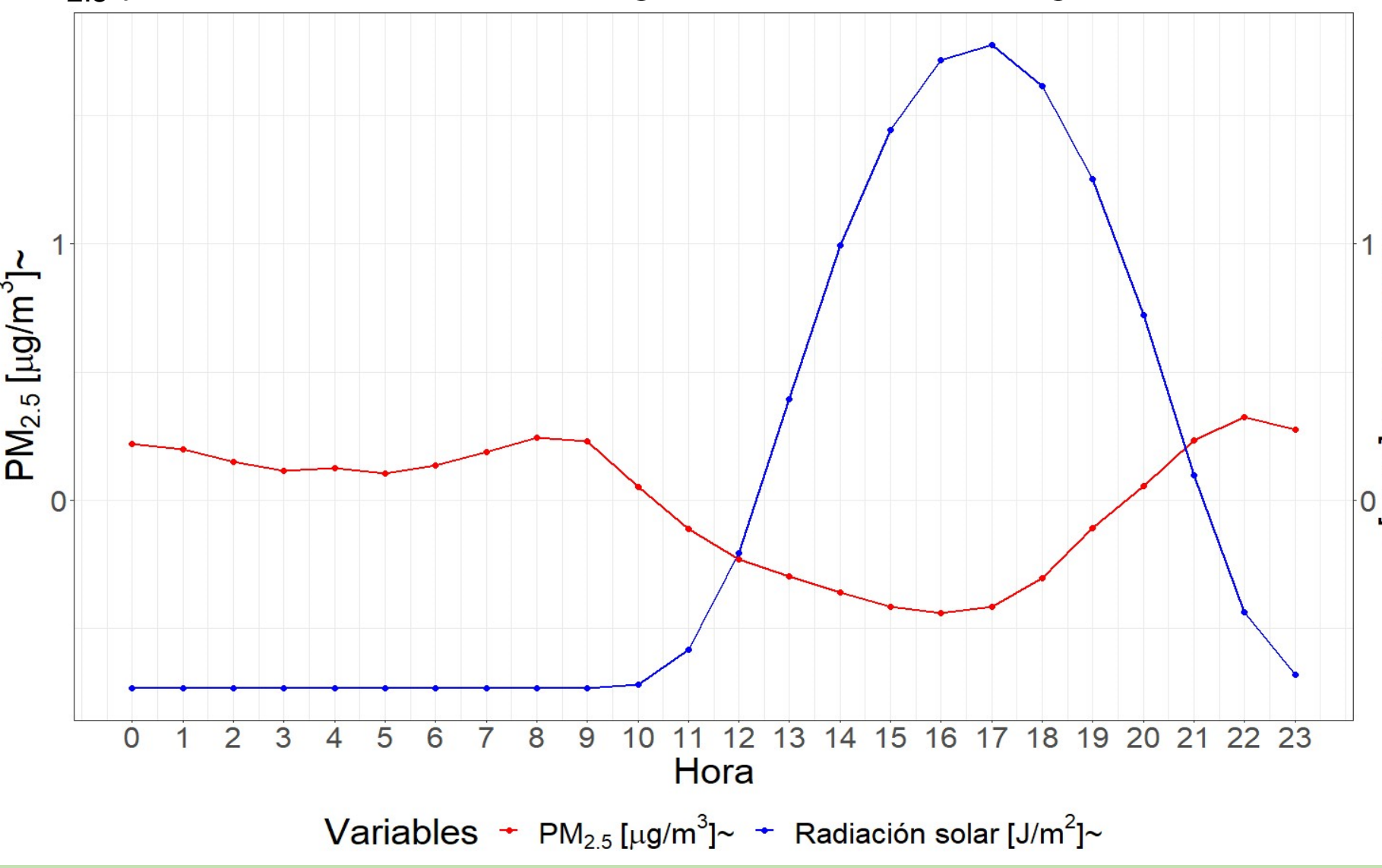


Figura 5.15: Importancia relativa de las variables en el bosque aleatorio para el uso de suelo agrícola/ área verde.

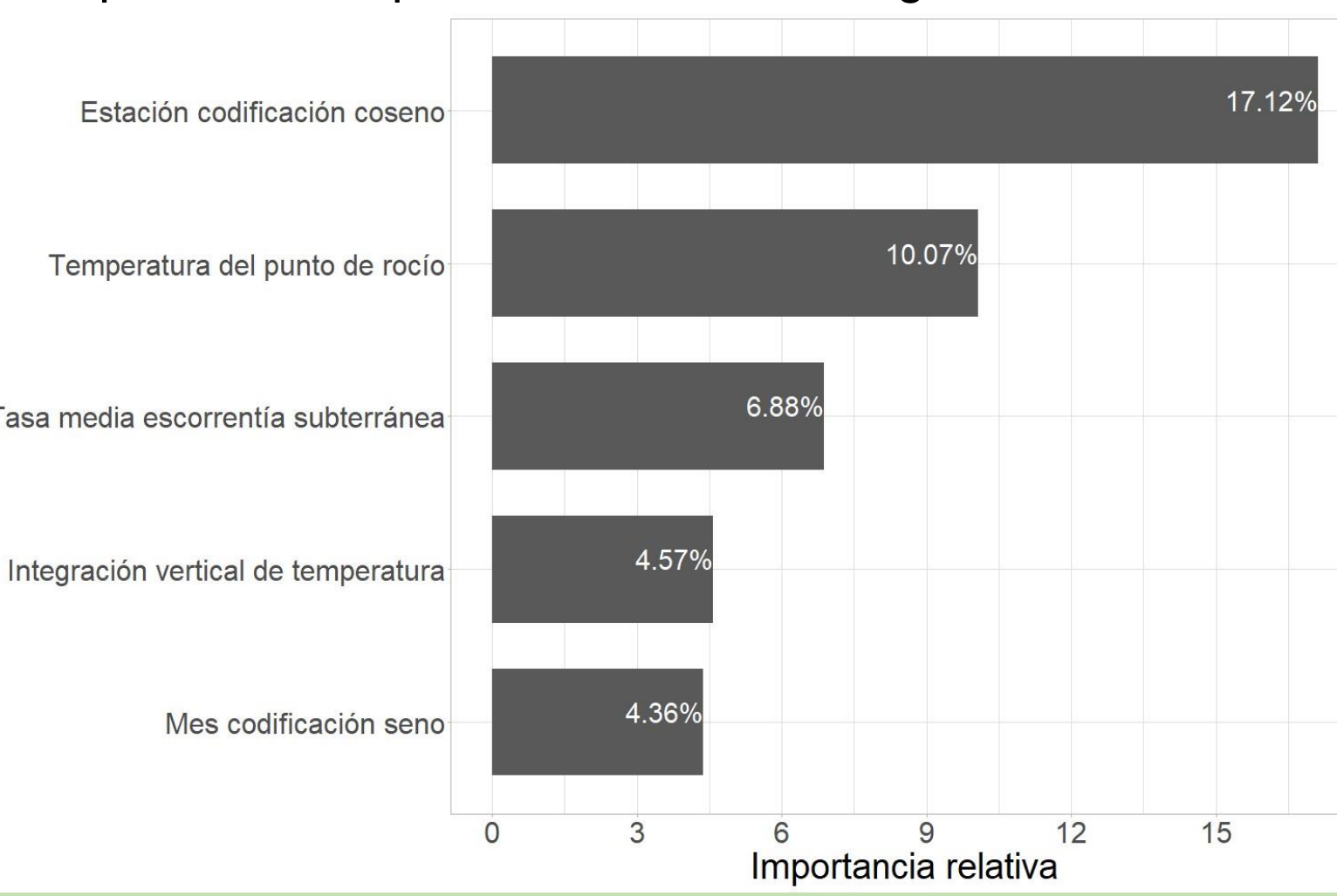
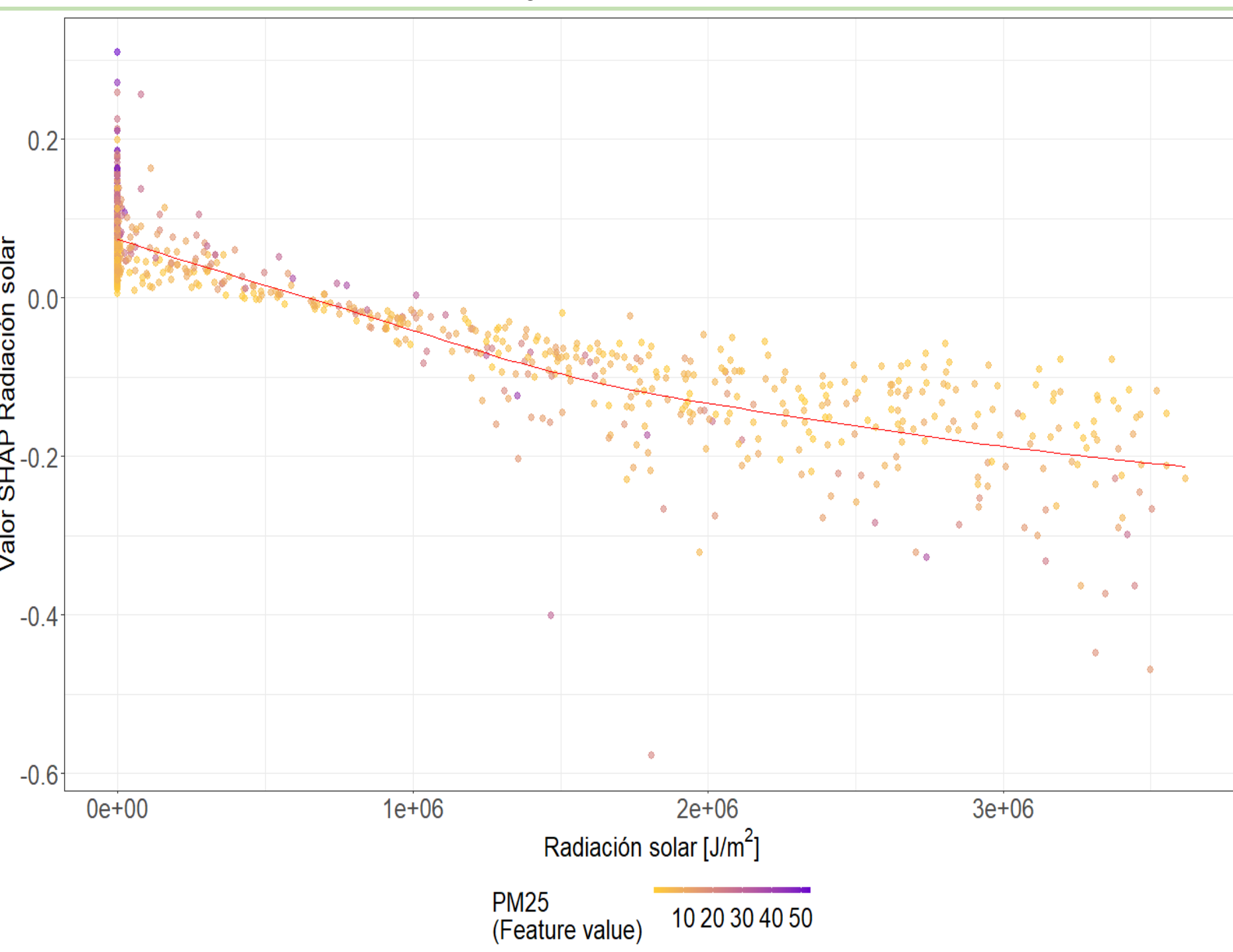


Figura 5.7: Gráfico de la variable radiación solar contra su valor SHAP coloreado por el valor de $PM_{2.5}$ para el uso de suelo urbano.



CONCLUSIÓN

R desempeña un papel fundamental en la investigación y el análisis de datos, ya que se trata de un lenguaje de programación de código abierto y gratuito. Su versatilidad, la presencia de una comunidad activa y su flexibilidad lo convierten en una herramienta esencial para abordar una amplia gama de desafíos en la investigación.