

Semana IDECA

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Universidad ECCI
Programa de Estadística
Grupo de Investigación ABEL



SEMANA Ideca 2022



03 de Oct. al 07 de Oct.

Seminario internacional, Conferencias
y Talleres presenciales y virtuales

Calidad de aire

Presencia en el aire de materias o formas de energía que impliquen riesgo o daño.

(Aranguéz, y otros, 1999)



Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022



Contaminación atmosférica

Presencia de materia o formas de energía que traigan consigo riesgos, daños o molestias graves para las personas o bienes de cualquier naturaleza. Entre estos se encuentran:

- PM_{10}

- SO_2

- CO

- NO_2

- O_3

$PM_{2.5}$

Partículas con un diámetro de 2,5 micrones o menos. Debido a su tamaño pueden atravesar la barrera pulmonar y entrar en el sistema sanguíneo. (Organización Mundial de la Salud, 2022)

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022



Afectaciones a la salud

A través del tiempo la calidad del aire se ha catalogado como una de las principales amenazas de muerte a nivel mundial. Una exposición constante o de forma crónica a este tipo de PM puede desarrollar enfermedades de tipo respiratorio o cardiovascular. (Organización Mundial de la Salud, 2022). Se estima que en Bogotá la contaminación del aire por el PM 2.5 fue responsable de la pérdida estimada de 5400 vidas durante el 2020.



Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022



Normatividad



OBJETIVOS DE DESARROLLO SOSTENIBLE - ODS

11. Ciudades y comunidades sostenibles

11.6 De aquí a 2030, reducir el impacto ambiental negativo per capita de las ciudades, incluso prestando especial atención a la calidad del aire y la gestión de los desechos municipales y de otro tipo



Directrices mundiales de la OMS sobre la calidad del aire (DCA)



POT 2022 - 2035. BOGOTÁ REVERDECE

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Normatividad

RESOLUCIÓN 2254 DE 2017

Establece la norma de calidad del aire o nivel de inmisión y adopta disposiciones para la gestión del recurso aire en el territorio nacional

ÍNDICE DE CALIDAD DE AIRE

El ICA es un valor adimensional para reportar el estado de la calidad del aire en función de un código de colores

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Índice Bogotano De Calidad del Aire - IBOCA

Nuevo IBOCA			
intervalos numéricos	Color	Estado de calidad del aire	Nivel de riesgo
0-50	Verde	Favorable	Prevención
51-100	Amarillo	Moderada	Prevención
101-150	Naranja	Regular	Alerta (Fase 1)
151-200	Rojo	Mala	Alerta (Fase 2)
201-500	Morado	Peligrosa	Emergencia

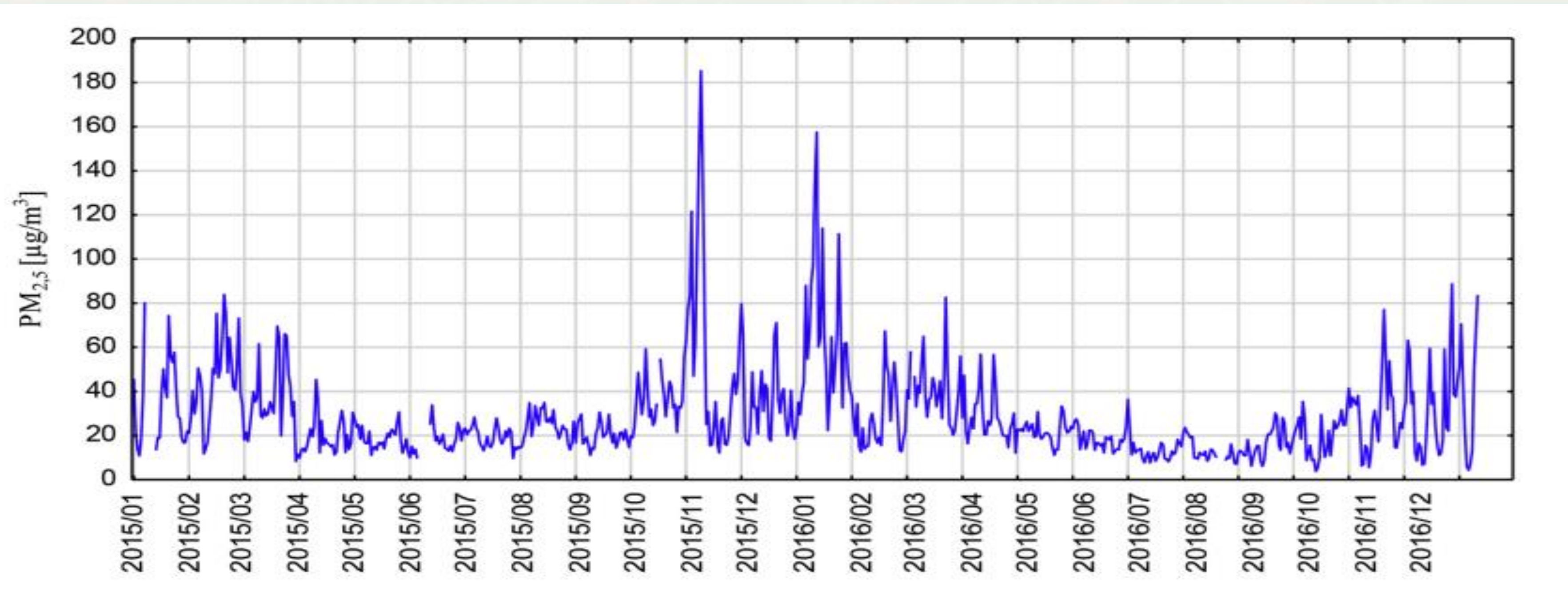
Tomado de Plan estratégico para la gestión Integral de la calidad del aire de Bogotá 2030

Antecedentes

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi(\vec{r}, t)$$

La necesidad de estudiar la calidad del aire respirable, sus características y el comportamiento difusivo de sus componentes ha llevado a algunos autores a proponer técnicas para describir, comprender y predecir los comportamientos de estos mismos.

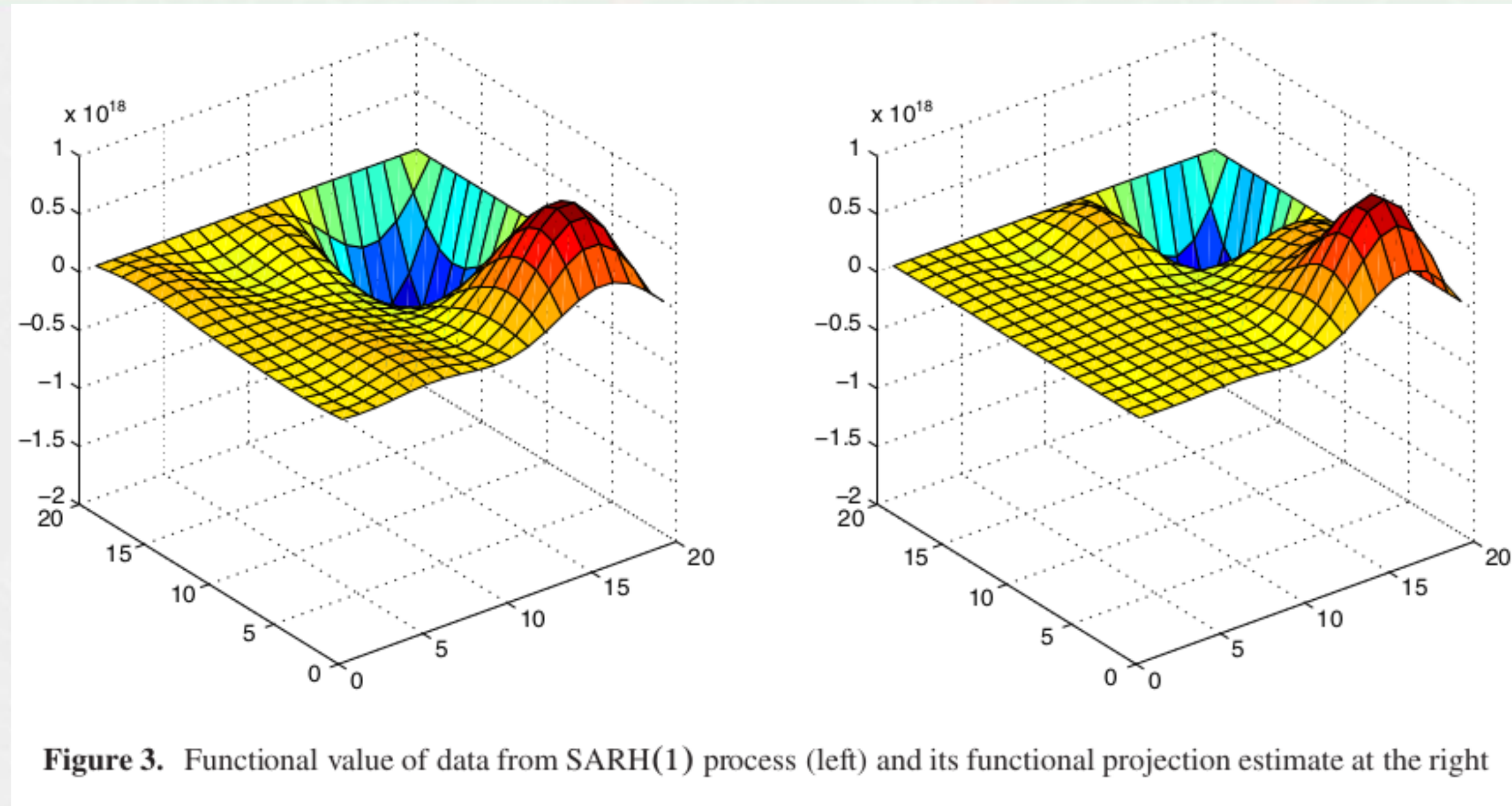
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022



Antecedentes

J. A. Kaminska (2018): The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in Wroclaw (Breslavia, Polonia).

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022



Antecedentes

M. D. Ruiz-Medina (2011): Spatial functional prediction from spatial autoregressive Hilbertian processes.

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Table 1 L^∞ -norms of the 10-fold CV mean absolute error curves with $L = 15$

Nodes	1997	1998	1999	2000	2001	2002
(2,2)	0.0183	0.0634	0.1929	0.0397	0.0082	0.0699
(3,2)	0.0129	0.0458	0.2198	0.0467	0.0286	0.0642
(2,3)	0.0155	0.0432	0.1662	0.0267	0.0134	0.1753
(3,3)	0.0147	0.0294	0.2360	0.0300	0.0187	0.0601
(2,4)	0.0146	0.0602	0.1752	0.0281	0.0389	0.0415
(3,4)	0.0156	0.0670	0.2653	0.0257	0.0484	0.0624
(2,5)	0.0151	0.0619	0.2346	0.0391	0.0481	0.0578
(3,5)	0.0153	0.0544	0.2984	0.0316	0.0543	0.0299

Table 2 L^∞ -norms of the 10-fold CV mean absolute error curves with $L = 15$

Nodes	2003	2004	2005	2006	2007	2008
(2,2)	0.0531	0.0471	0.1602	0.0532	0.0108	0.1027
(3,2)	0.0623	0.0688	0.1577	0.0891	0.0121	0.1133
(2,3)	0.1227	0.0376	0.1544	0.0532	0.0119	0.1023
(3,3)	0.0566	0.0880	0.1662	0.0808	0.0122	0.1151
(2,4)	0.0537	0.0937	0.0635	0.1277	0.0061	0.1011
(3,4)	0.0628	0.1790	0.2331	0.1041	0.0083	0.1159
(2,5)	0.0599	0.1459	0.0738	0.1058	0.0065	0.1007
(3,5)	0.0829	0.2459	0.2827	0.1799	0.0058	0.1173

Antecedentes

M. D. Ruiz-Medina, R. M. Espejo (2012): Spatial autoregressive functional plug-in prediction of ocean surface temperature.

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Objetivo General

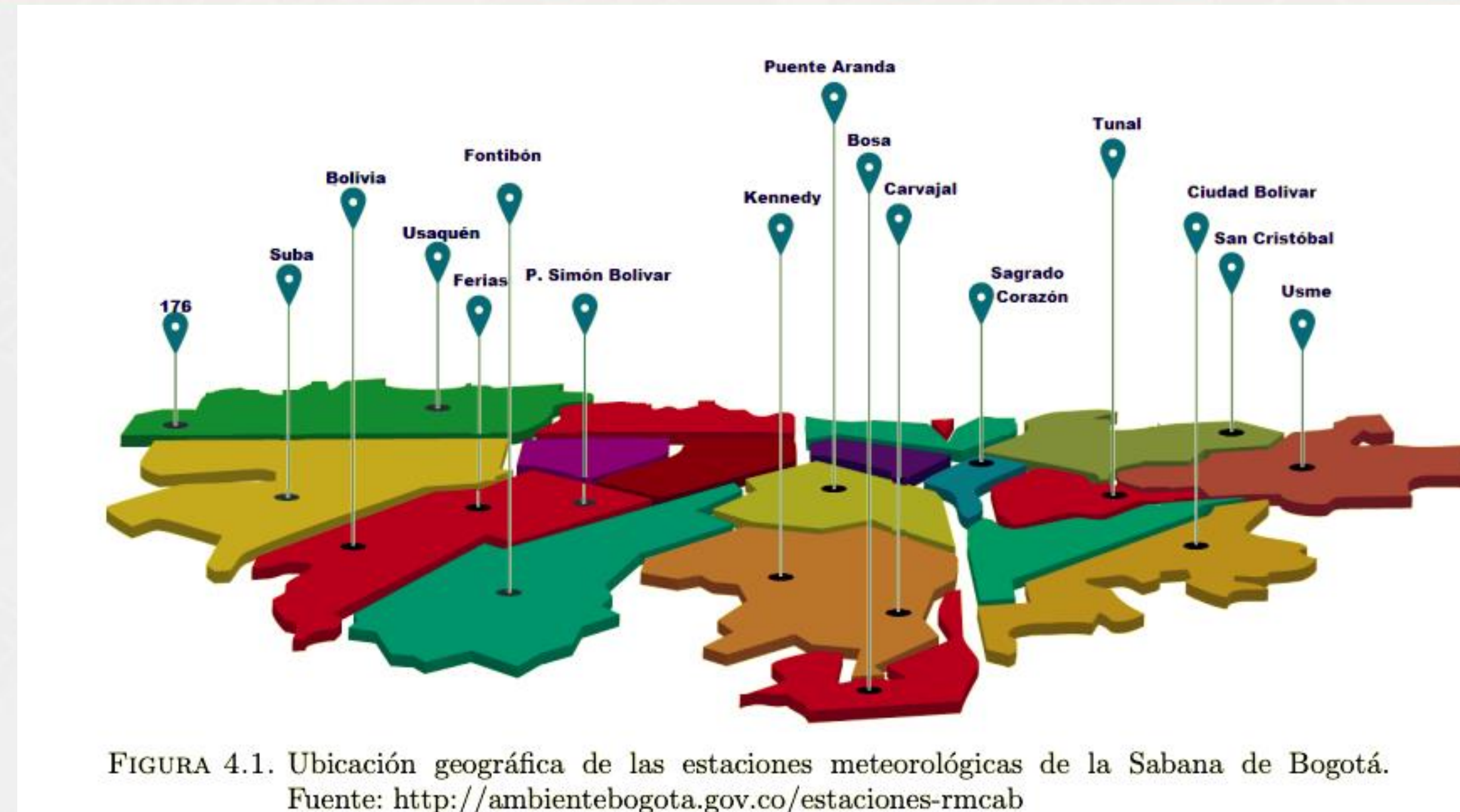
- ✓ Establecer un modelo predictivo para su aplicación en problemáticas con comportamientos de difusión.

Objetivos específicos

- ✓ Determinar la superficie de mayor predicción de difusión de contaminantes a partir de la combinación de las metodologías MissForest y SARH(1).
- ✓ Predecir la difusión de contaminantes en el área metropolitana de Bogotá a partir de los datos existentes en el período 2017-2022 bajo las regulaciones de los entes encargados.

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Materiales

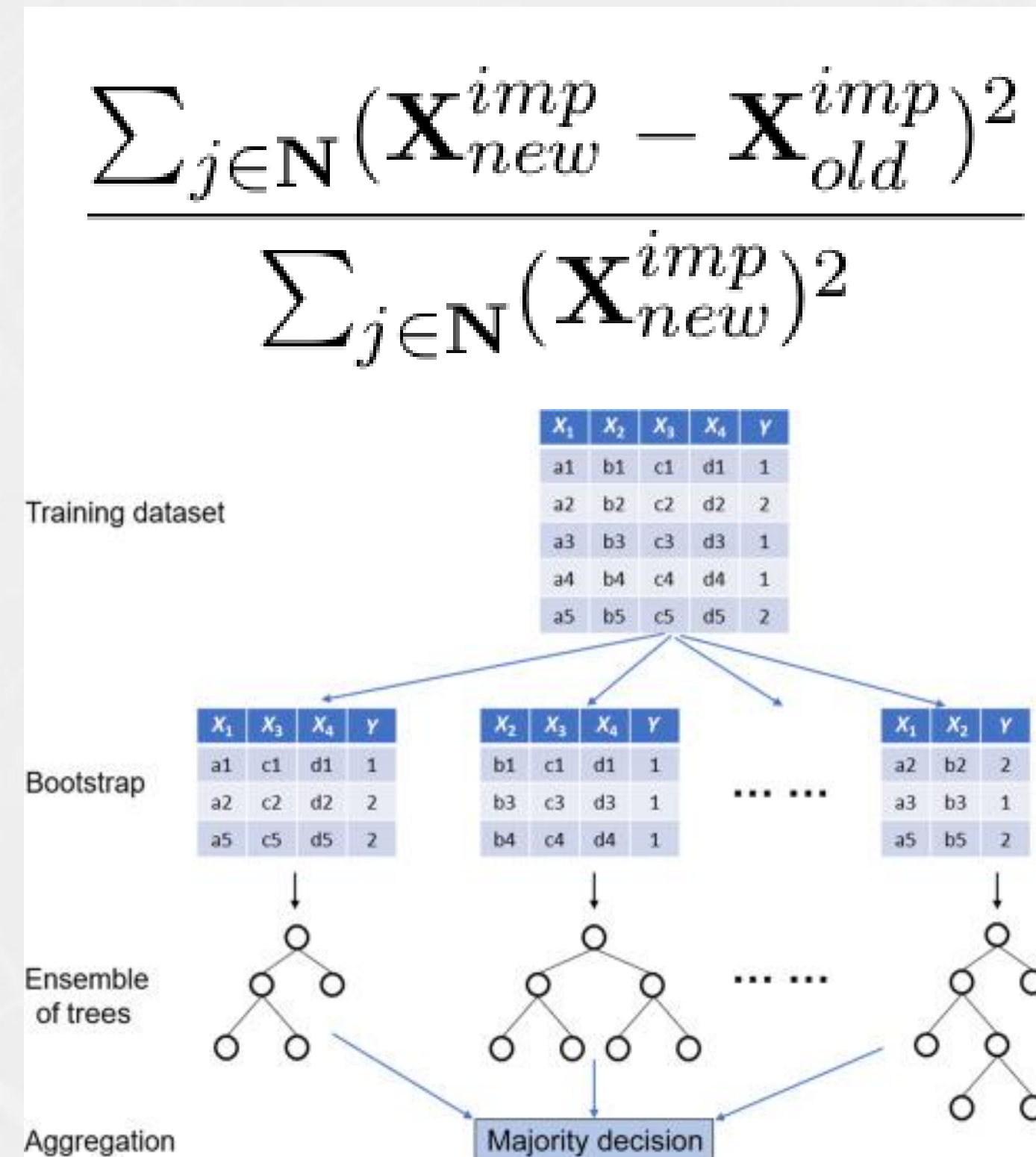


Conjunto de datos recolectados por la red de estaciones meteorológicas de Bogotá.

- Datos meteorológicos: Temperatura, presión, velocidad del viento, dirección del viento, precipitaciones, radiación solar y humedad relativa
- Concentración de contaminantes atmosféricos: Monóxido de carbono, dióxido de azufre, ozono, dióxido de nitrógeno y material particulado.

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Metodología



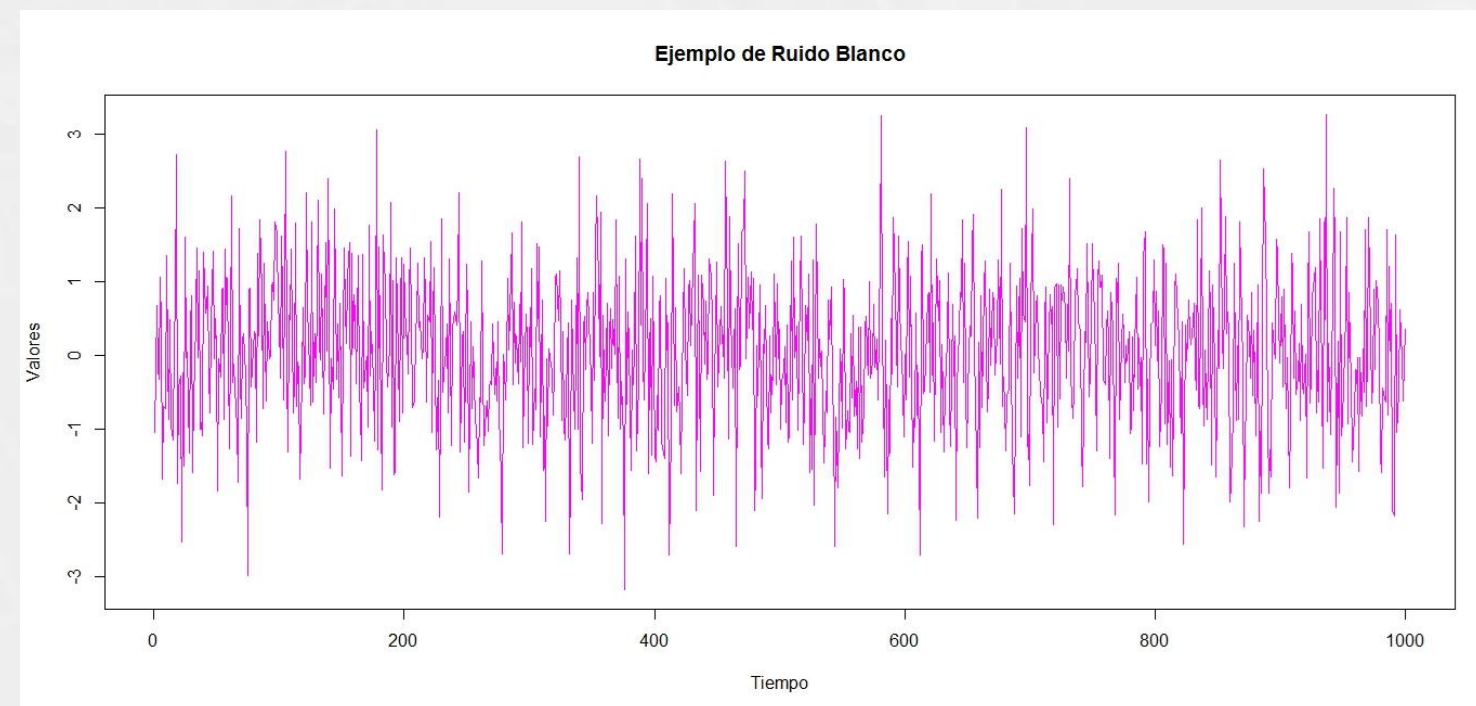
MissForest

MissForest es un método de imputación no paramétrico para básicamente cualquier tipo de datos. Puede hacer frente a variables de tipo mixto, relaciones no lineales, interacciones complejas y alta dimensionalidad ($p \gg n$). Sólo requiere las observaciones. El algoritmo se basa en Random Forest (Breiman (2001)).

Para cada variable, missForest ajusta un Random Forest en la parte observada y luego predice la parte que falta. El algoritmo continúa repitiendo estos dos pasos hasta que se cumple un criterio de parada o se alcanza el máximo de iteraciones especificado (Stekhoven and Bühlmann (2012)).

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Metodología



Modelo Espacio temporal autorregresivo de Hilbert SARH(1)

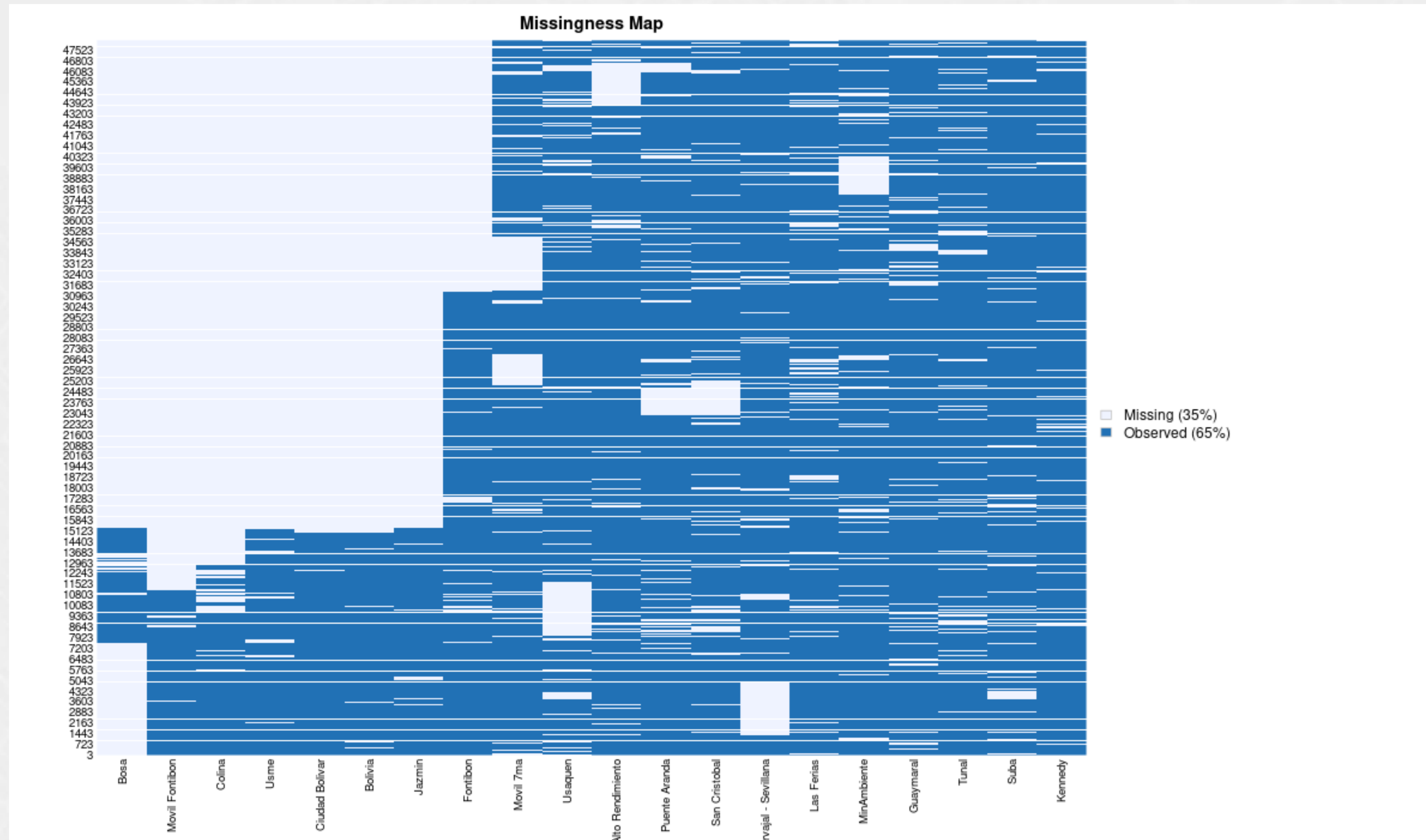
Definición: Sea H un espacio de Hilbert separable de funciones sobre un dominio $D \subseteq \mathbf{R}^n$, con producto interno, y con norma. Sea $\{X_t, t \in \mathbf{N}\}$, un proceso estacionario de Hilbert sobre un espacio de probabilidad (Ω, F, P) satisfice:

$$X_t(s) = \Psi X_{t-1}(s) + v_t(s), s \in D \subseteq \mathbf{R}^n, t \in \mathbf{N},$$

donde v_t es RB funcional, es decir, una sucesión de variables aleatorias independientes e idénticamente distribuidas sobre H con varianza finita. El operador de autocorrelación Ψ es acotado y definido sobre un dominio denso en H .

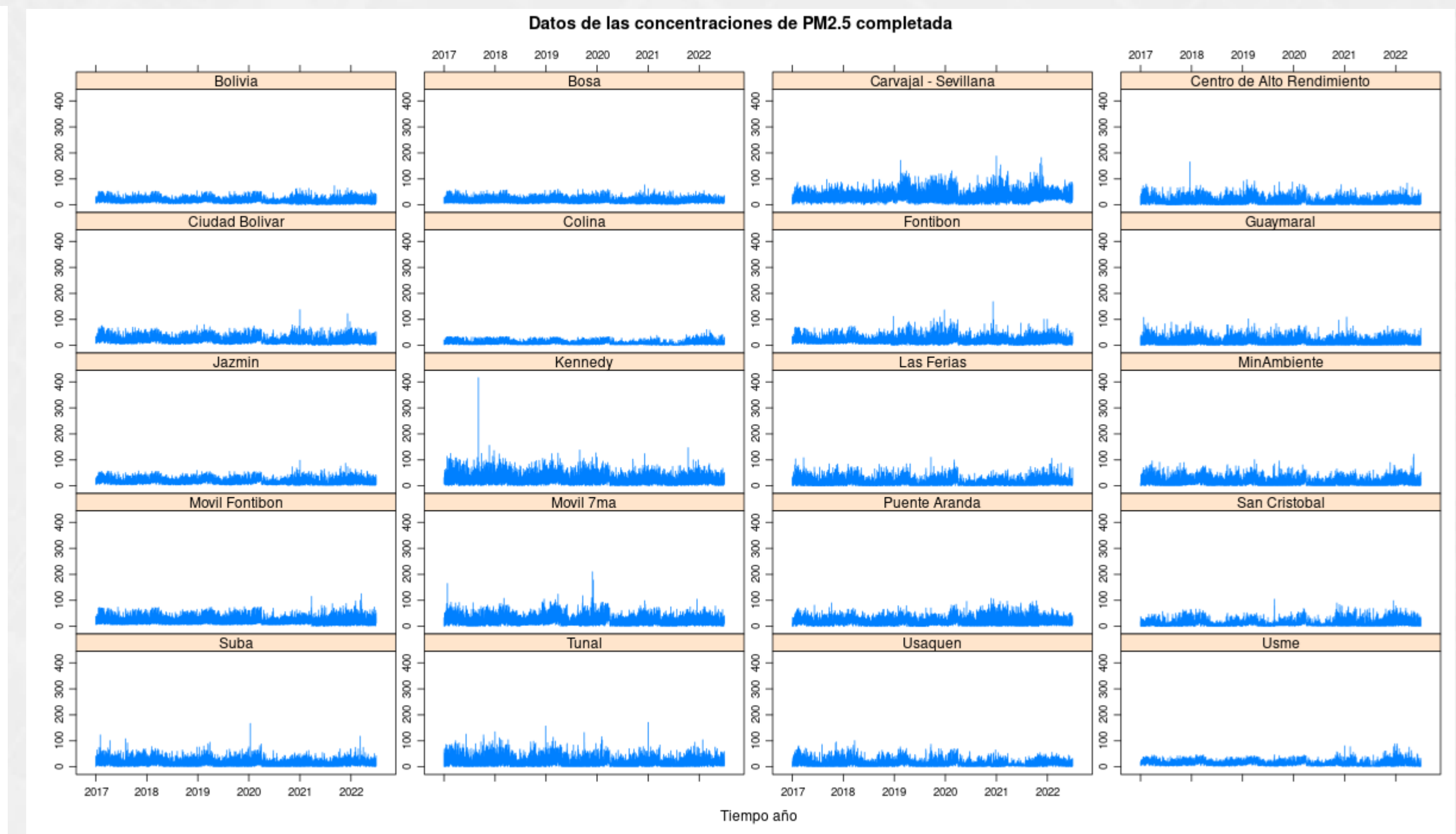
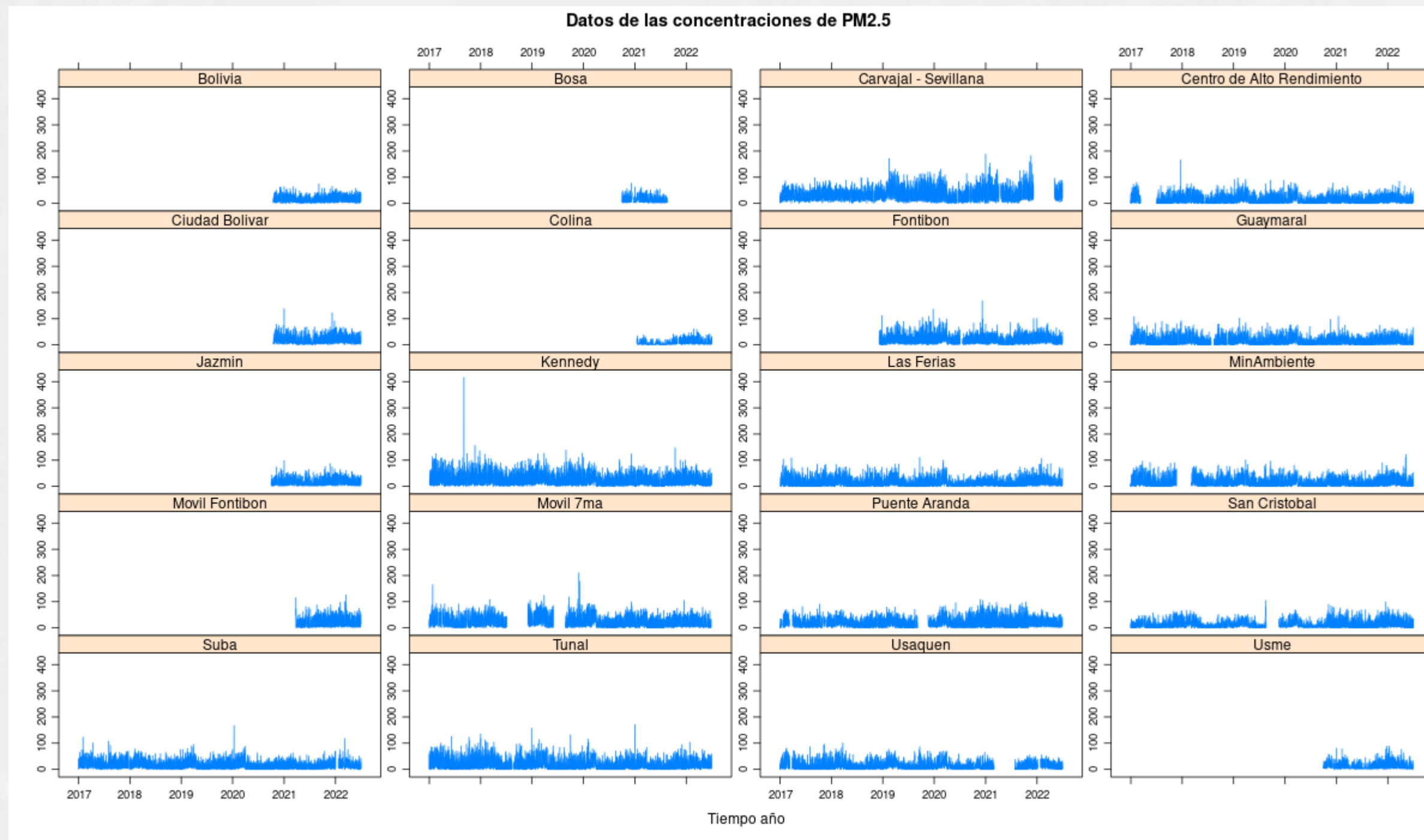
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Datos faltantes



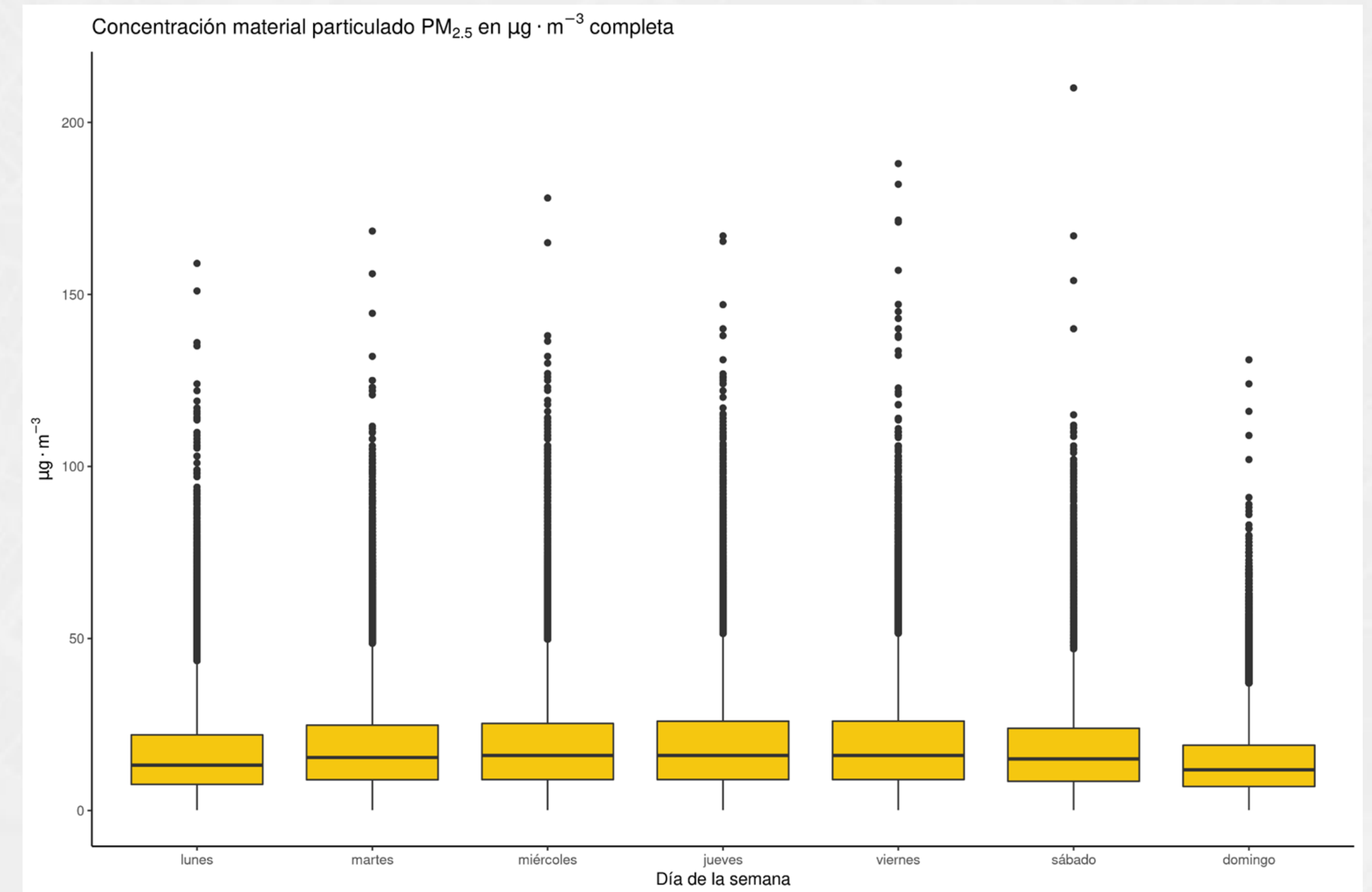
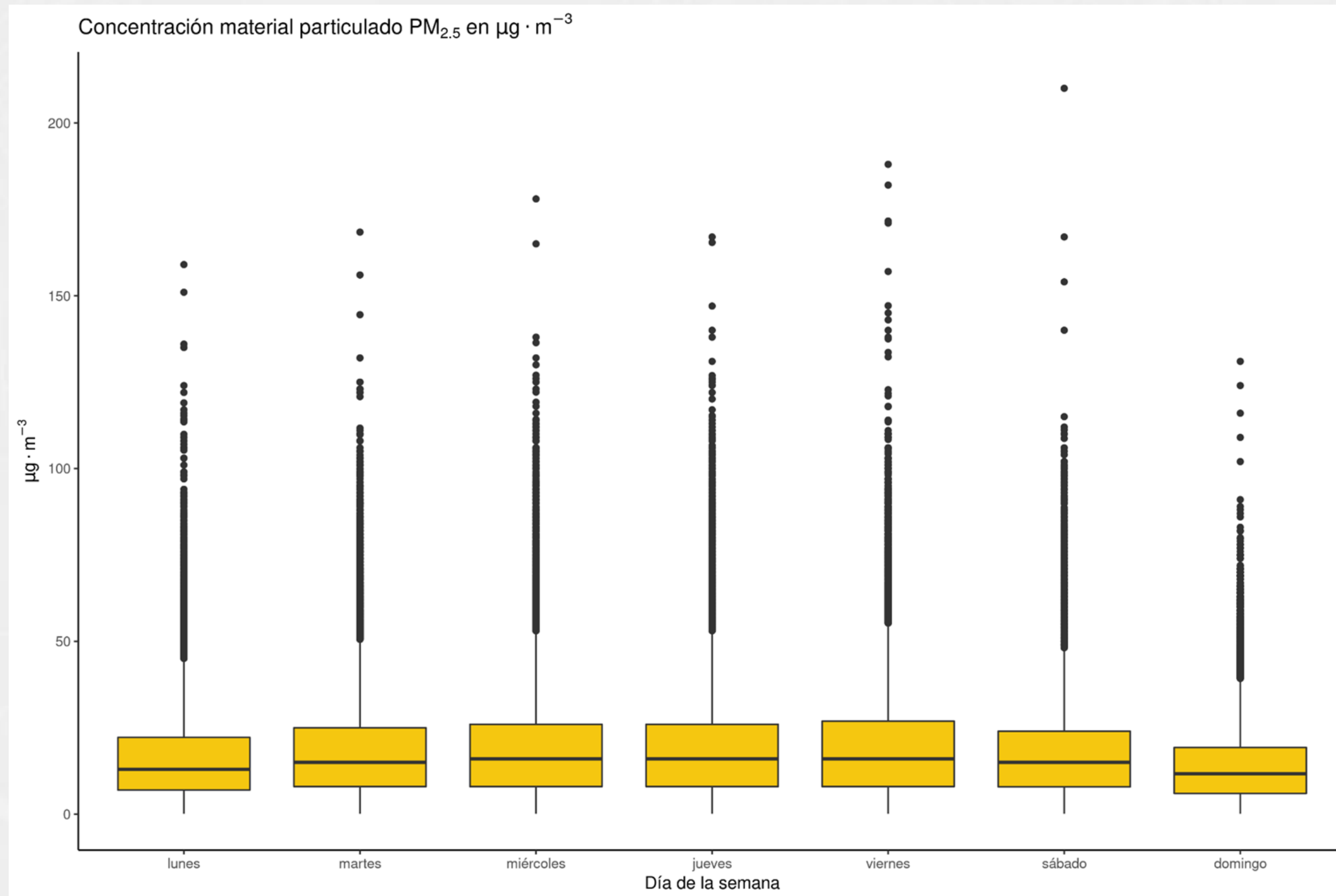
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Series temporales incompletas vs completas



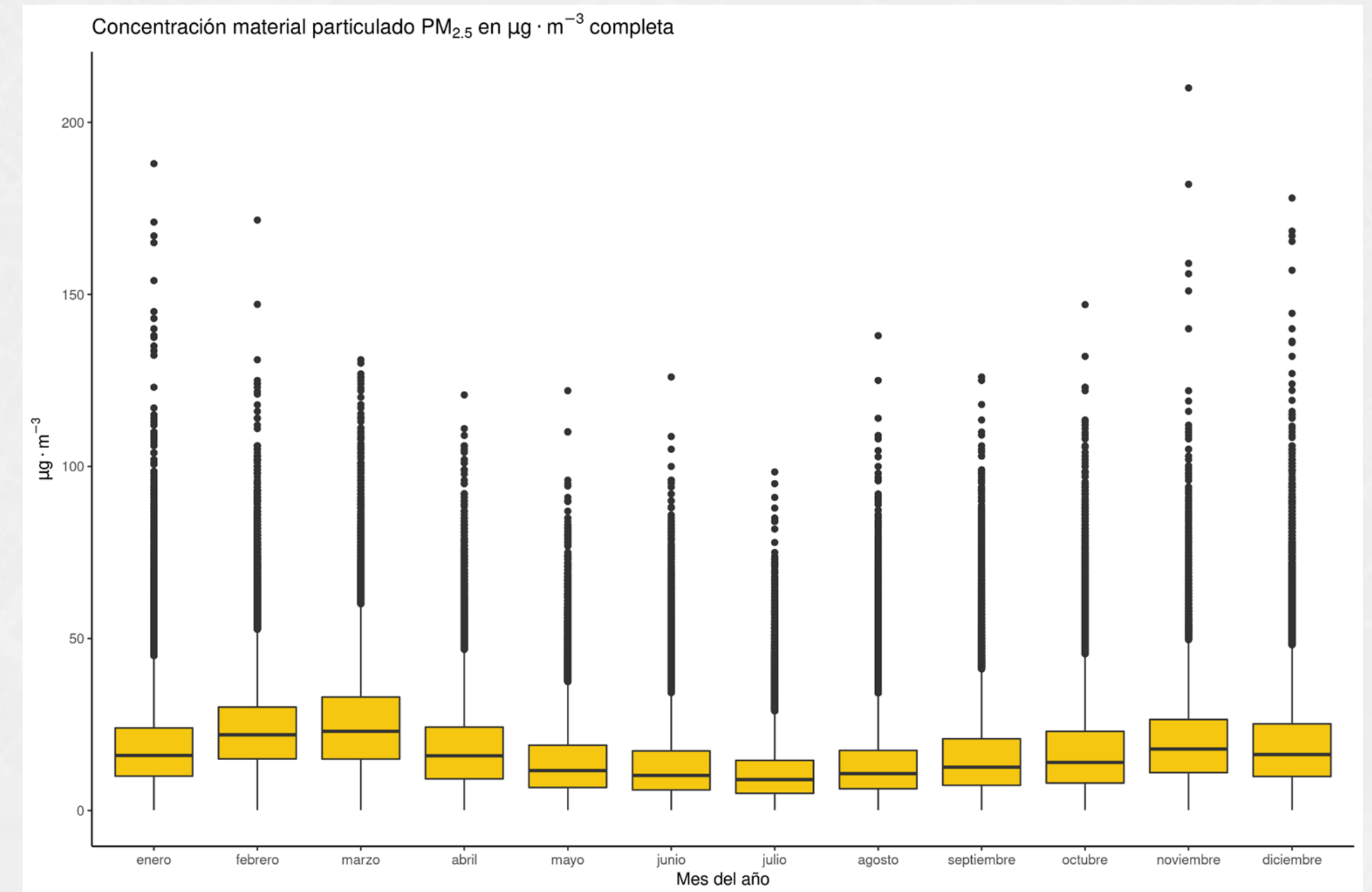
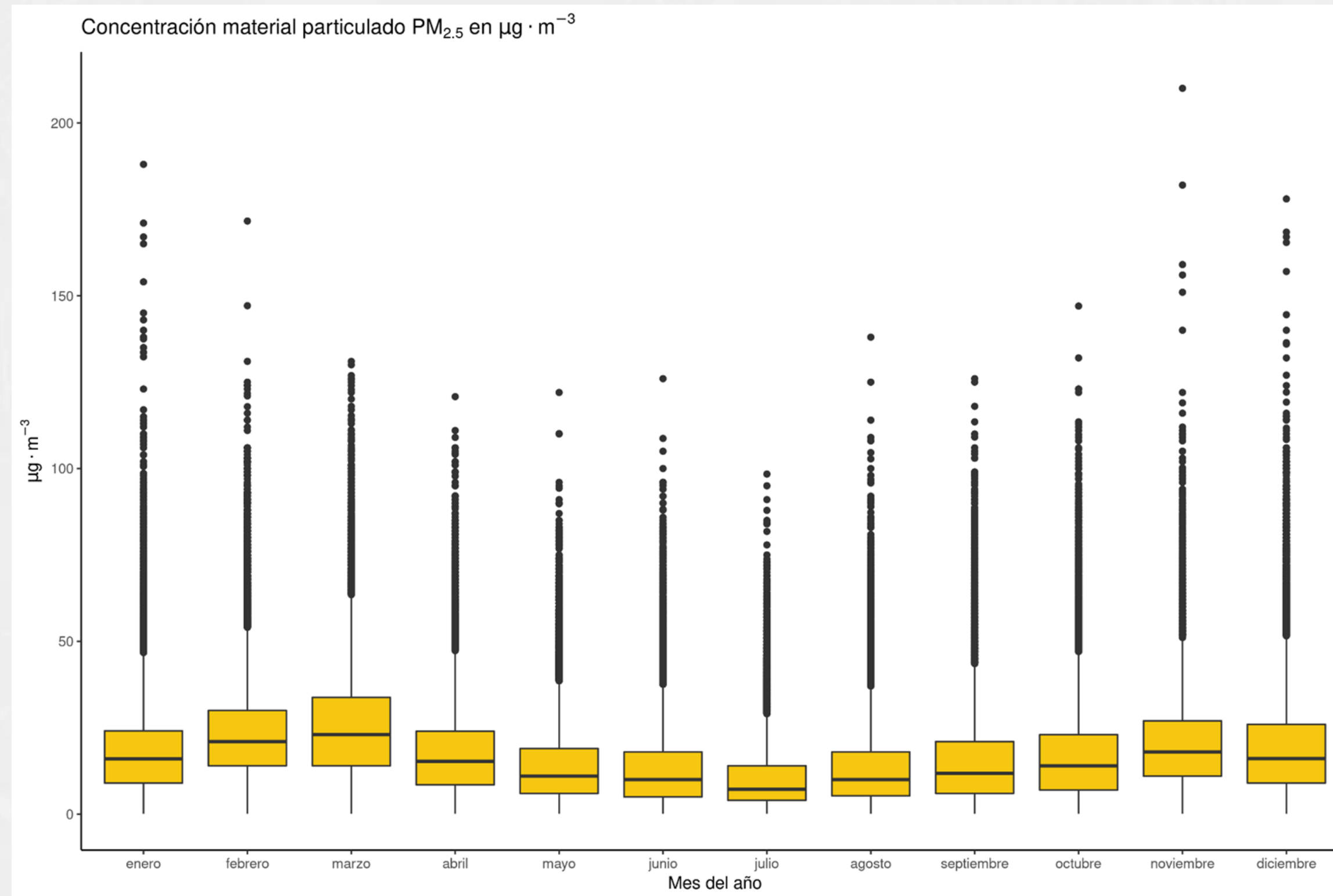
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por día incompletas vs completas



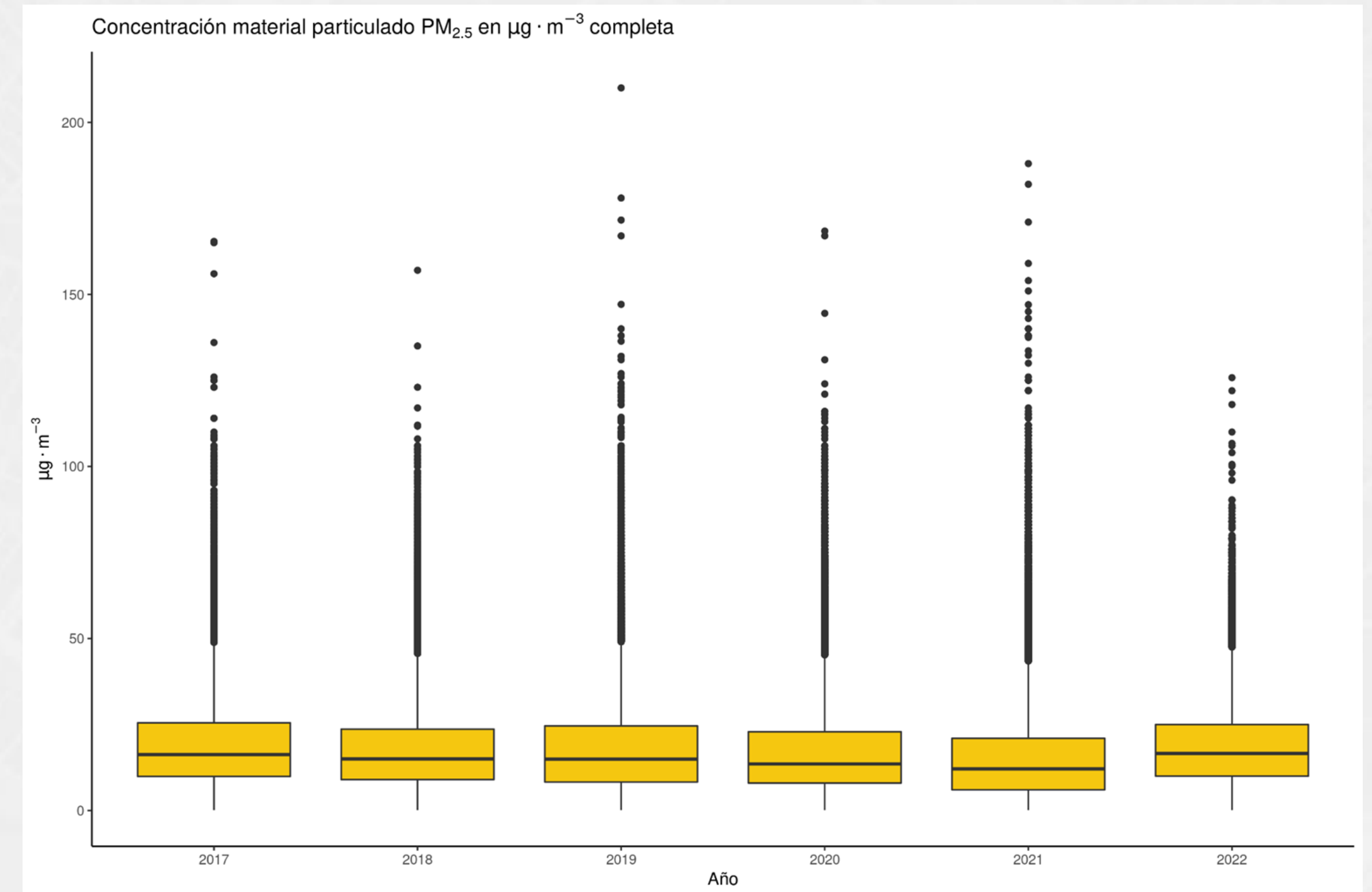
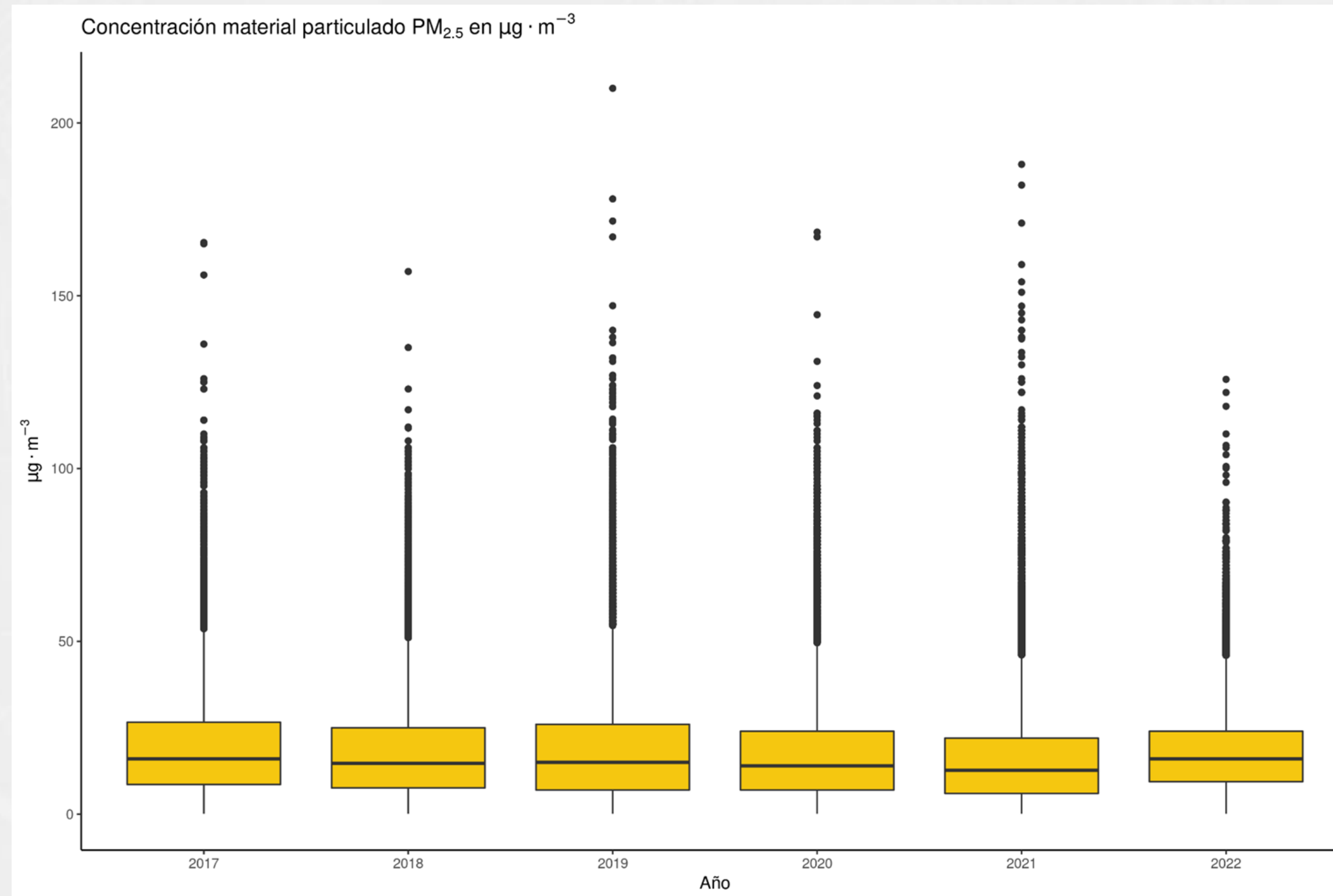
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por mes incompletas vs completas



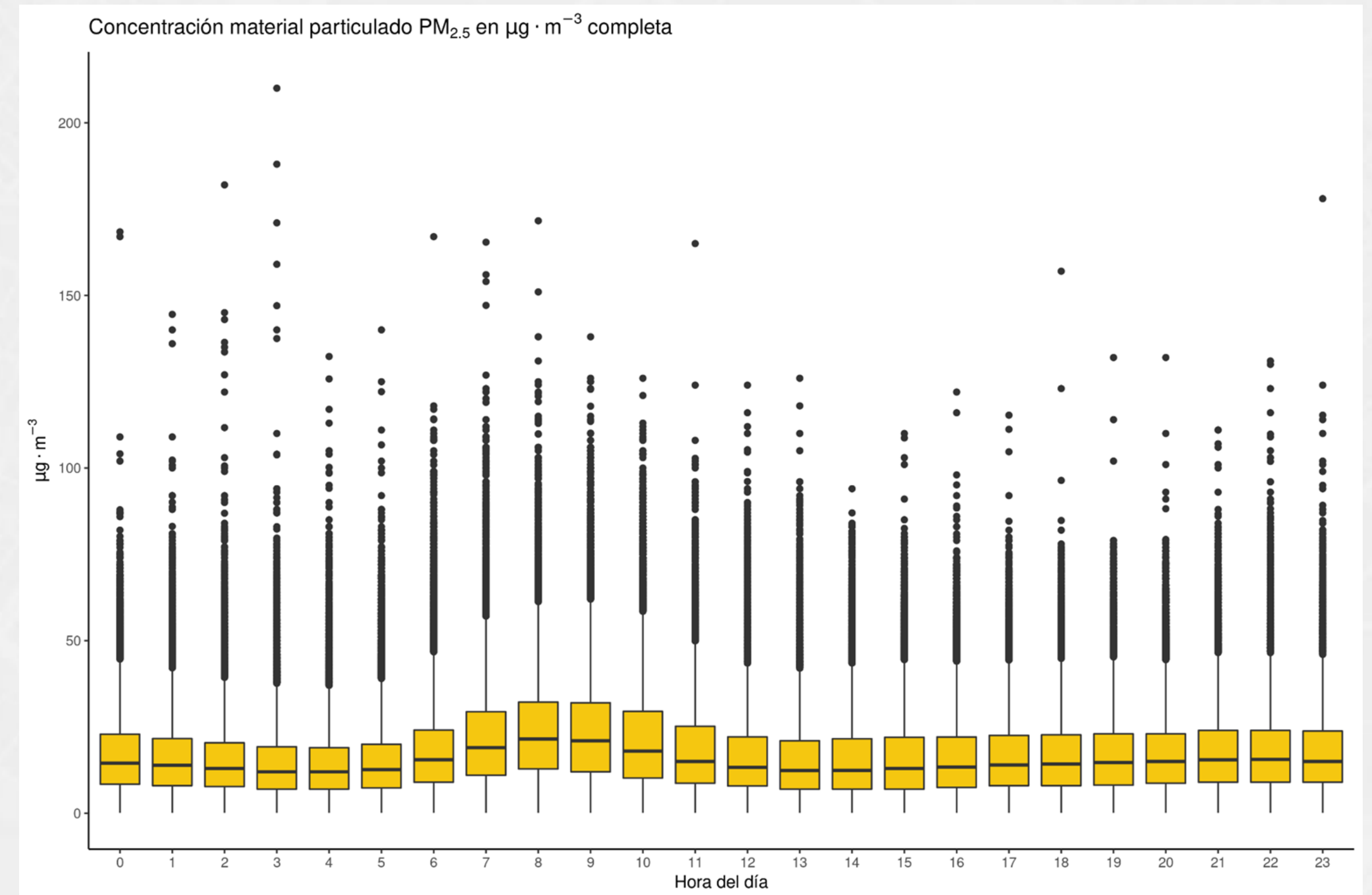
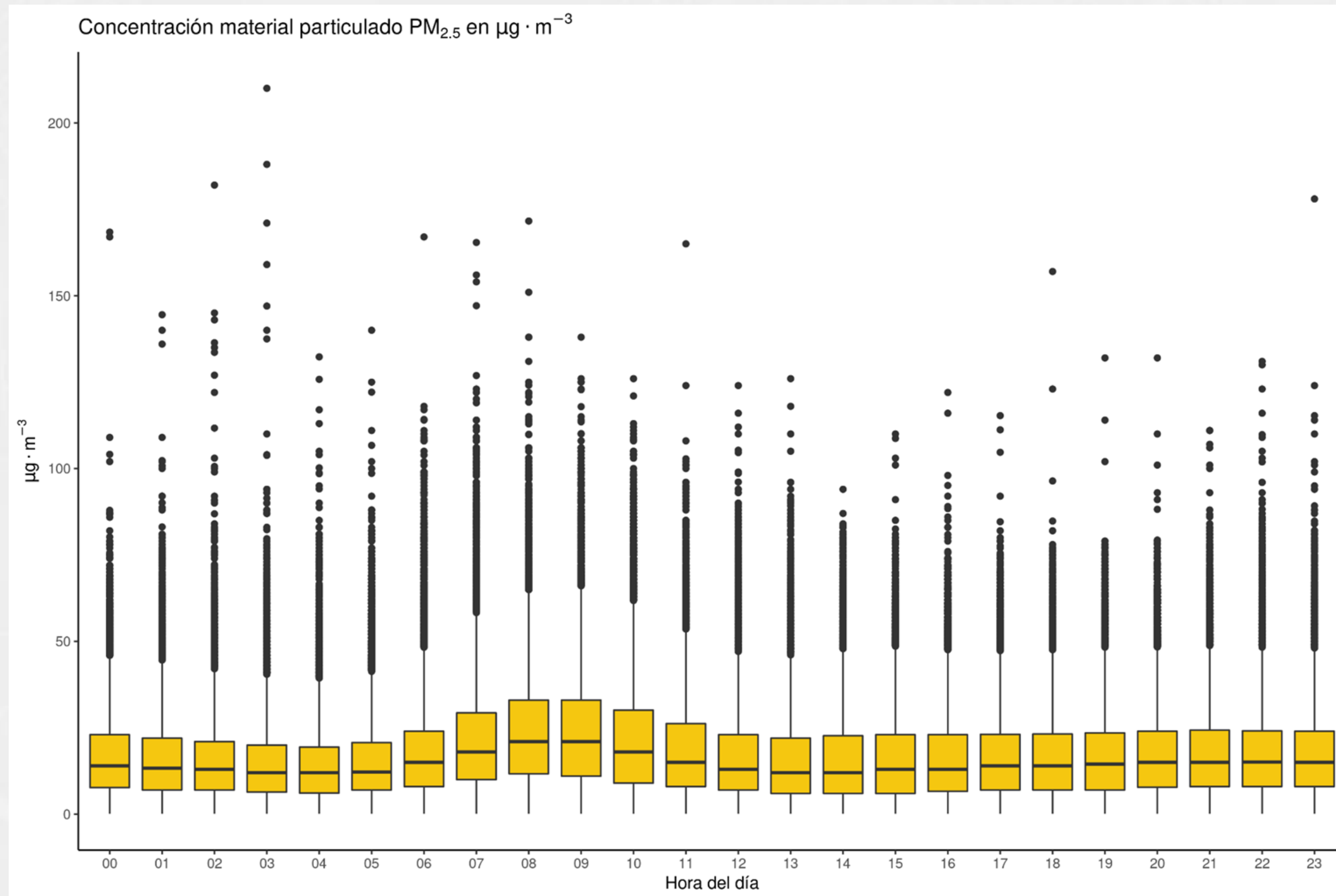
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por año incompletas vs completas



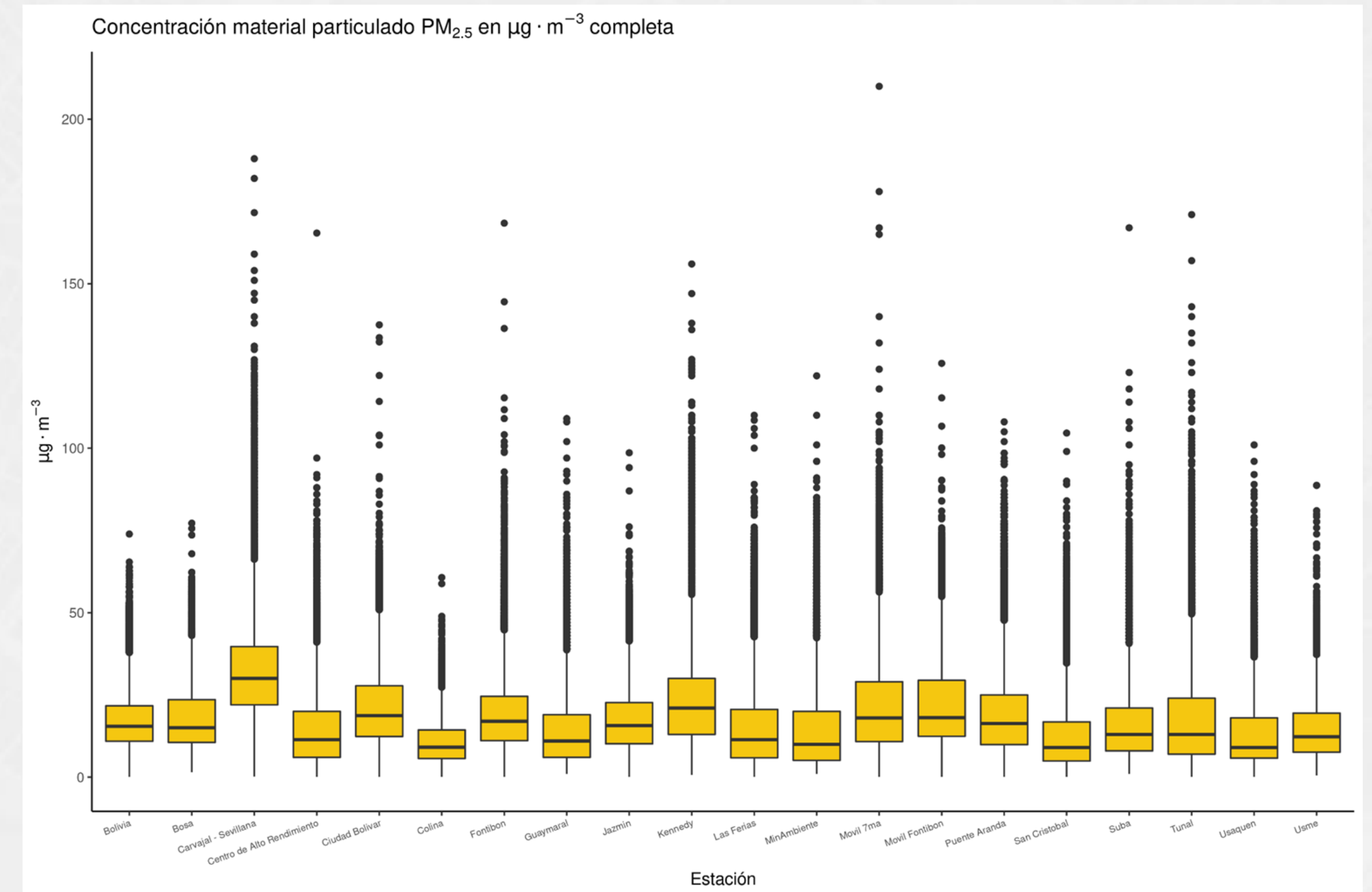
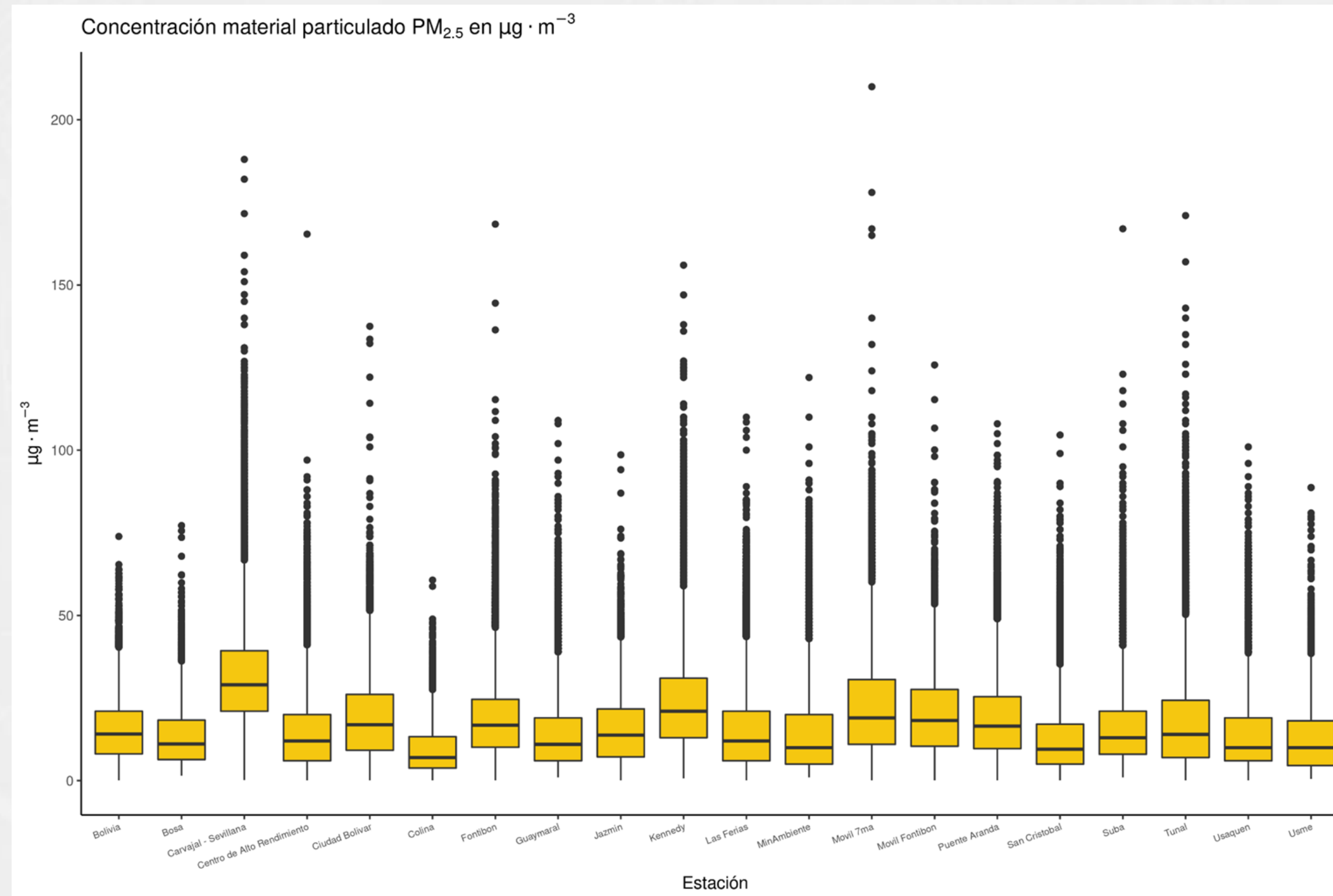
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por hora incompletas vs completas



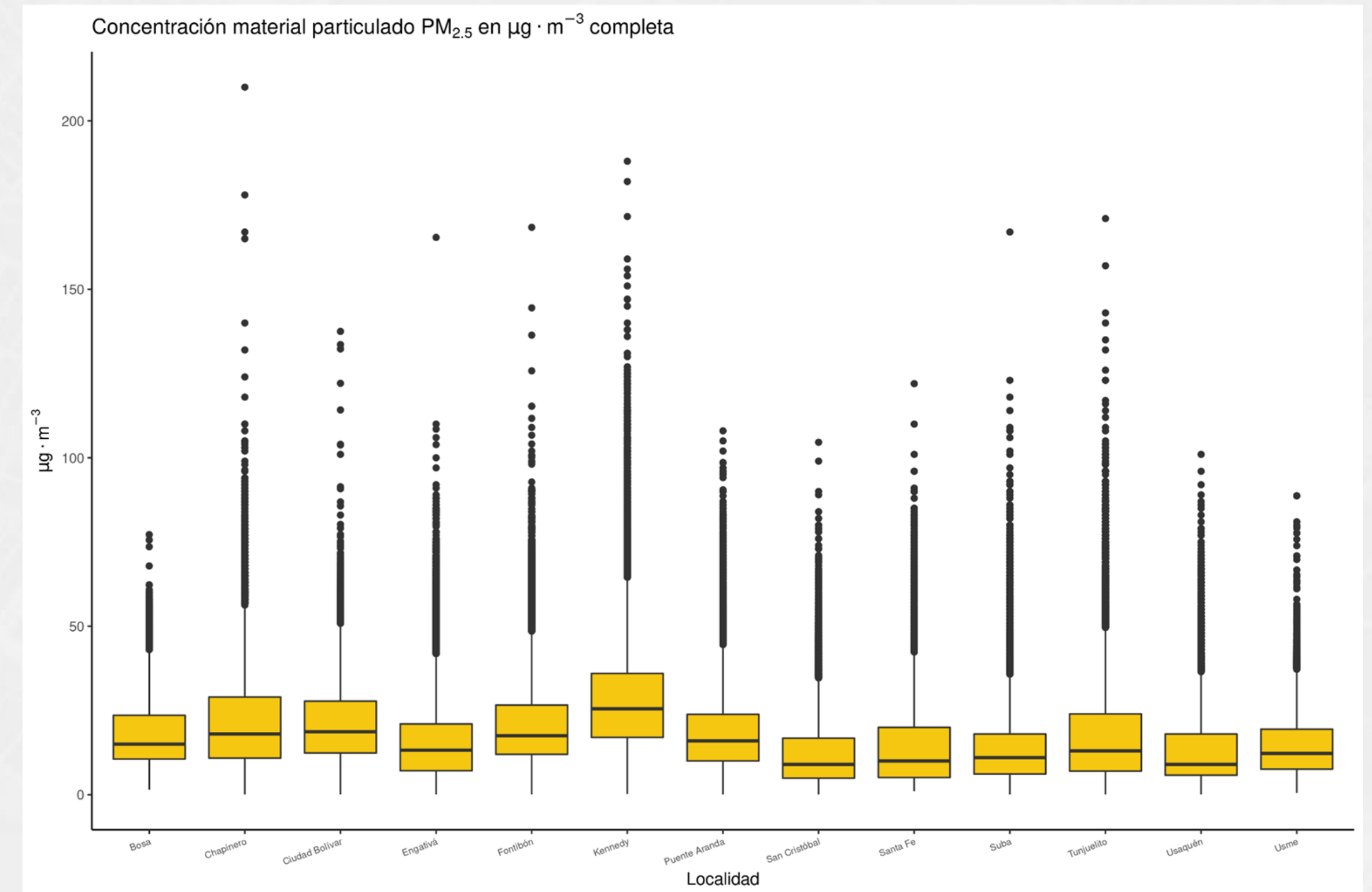
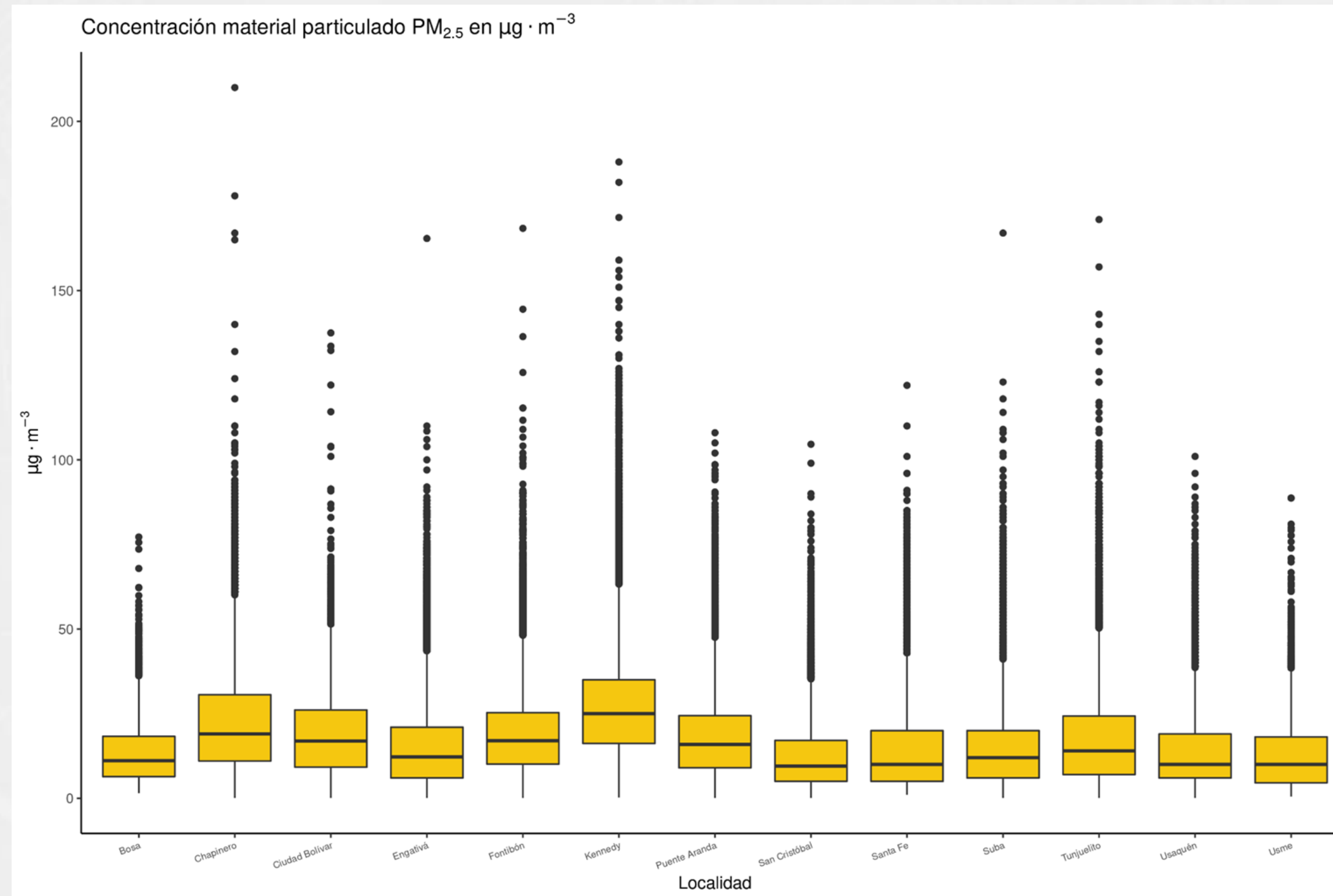
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por estación incompletas vs completas



Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Distribuciones por localidad incompletas vs completas



Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados: Validación cruzada

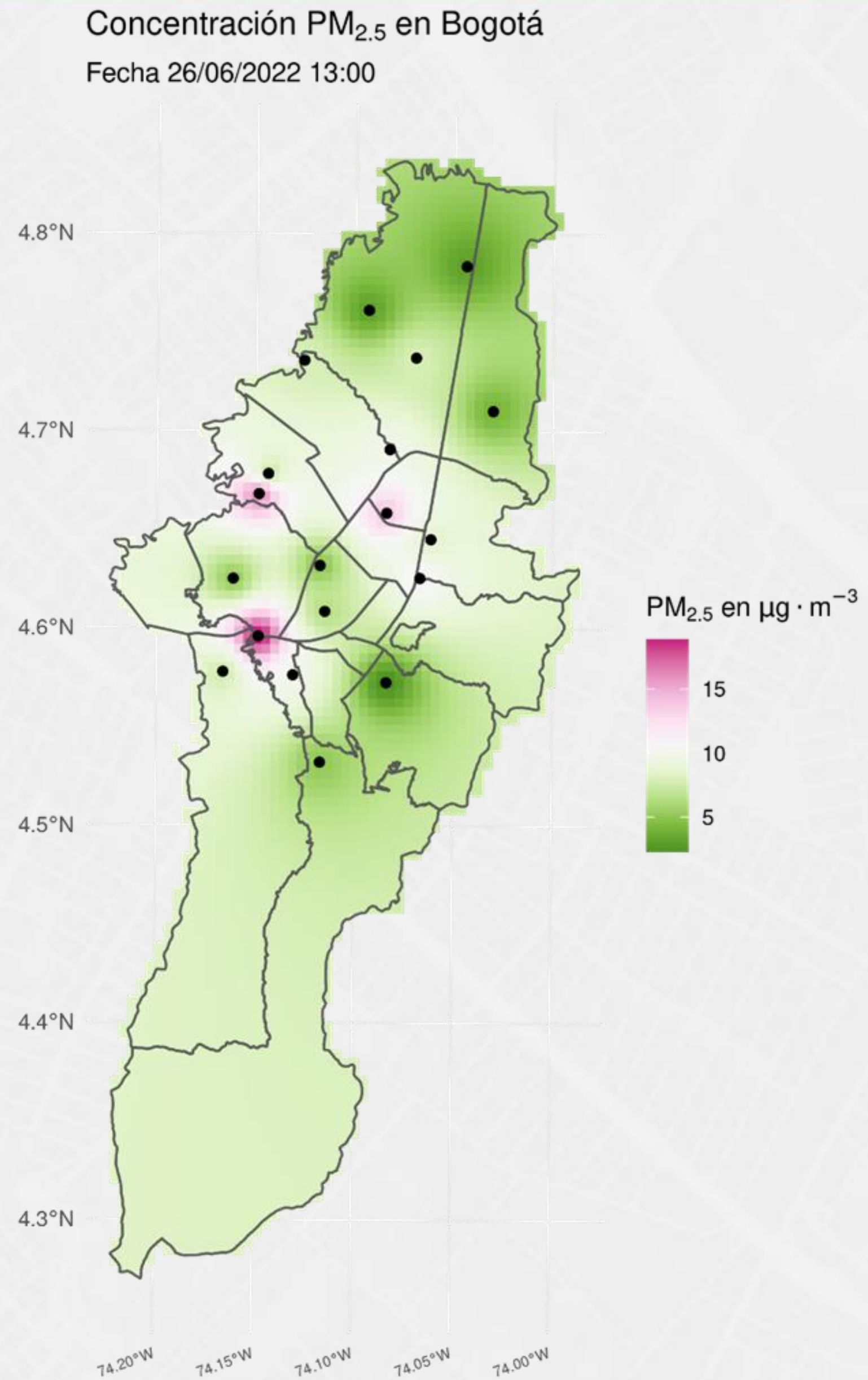
Table 2. Validación cruzada por estación meteorológica en distintos periodos observados y *ECMF*

Estación	11/01/2022 10:00 horas		14/03/2022 10:00 horas		2/05/2022 10:00 horas	
	Observado	Estimación	Observado	Estimación	Observado	Estimación
Guaymaral	34.3	28.9	35	31.9	39	35.5
Usaquén	40	46	23	20	15	13.1
Suba	30	24	70	81.3	15	15.1
CDAR	53.1	51.2	31.5	34.2	15	13.2
Las Ferias	52	56.4	51.2	45.1	16.7	14.1
Fontibón	71.4	75.8	67.5	49.4	28.6	18.7
Min. Ambiente	73.4	68.4	53.2	61.1	98.9	100.4
Séptima	16	14.1	13	13.1	14	14.7
Puente Ara	31	31.1	34	34.2	28	27.2
Kennedy	12	11	1	12.1	25	26.1
Carvajal	60	67	74	72	47	44
Tunal	40	43	50	43	51	46
San Cristó	22	21	31	29.9	48	46.6
ECMF	1.5×10^{-2}		2.5×10^{-2}		3.5×10^{-2}	

Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados:

Superficie estimada



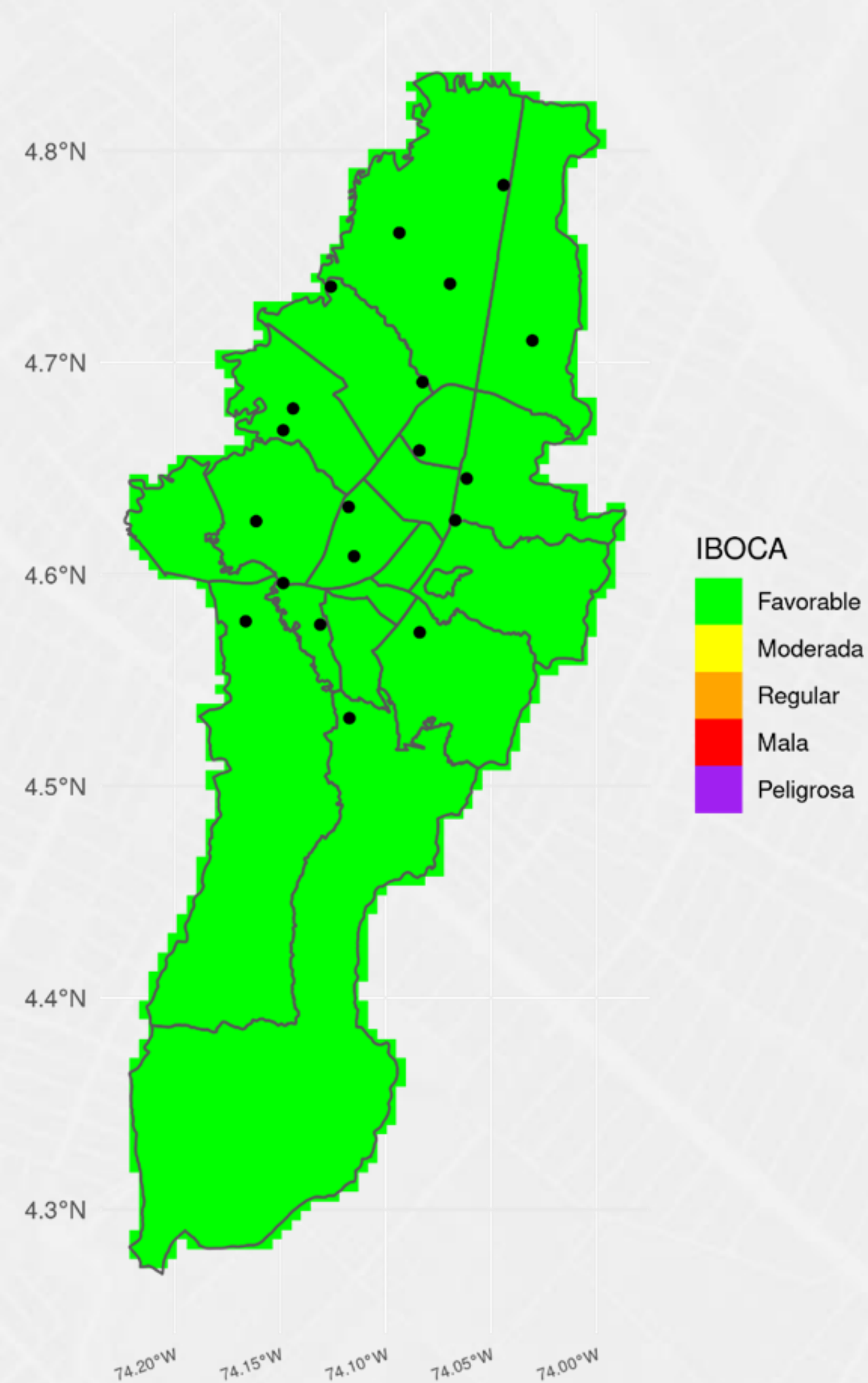
Modelo predictivo de difusión del contaminante $PM_{2.5}$ en el área urbana de Bogotá entre 2017-2022

Resultados:

IBOCA

IBOCA en concentración $PM_{2.5}$ en Bogotá

Fecha 26/06/2022 13:00



Discusión

- ✓ Este trabajo indicó que este modelo predictivo permite tener una precisión bastante adecuada de los valores futuros que puede tomar, en este caso, la concentración de $PM_{2.5}$ en el aire respirable del área metropolitana de Bogotá, Colombia, en el corto plazo, con ventanas de tiempo en intervalos de 1 hora y locaciones en donde no existen estaciones meteorológicas.
- ✓ Este modelo permite también predecir eventos de alta concentración en regiones específicas con el fin de tomar decisiones, reaccionando rápida y eficientemente ante los mismos. No solo para el contaminante $PM_{2.5}$, siendo este el que mayor afectación genera a la salud, ocasionando enfermedades cardíacas y respiratorias, sino también, para los demás contaminantes presentes en el aire.
- ✓ La principal contribución de este trabajo radica en el hecho de que utiliza una técnica de ML para imputar datos espacio temporales faltantes para, por medio del SARH(1), señalar la concentración estimada de una variable, que se comporta como una sustancia difundiéndose en un fluido, en una ubicación precisa.

Conclusiones

- ✓ La combinación de técnicas de Machine Learning y modelos estadísticos espacio temporales potencia en gran medida los resultados en el campo de la predicción. En este caso, se evidencia que la técnica MissForest aumenta la resolución de los datos, generando disminución en los ECMF llegando al orden medio de 10^{-3} en el SARH(1).
- ✓ Es posible realizar una estimación de la concentración de $PM_{2.5}$ en otros puntos del espacio, es decir, en localidades aledañas, a partir del modelo SARH(1) y de la curva de estimación $X_t(s)$.
- ✓ Conforme con la información recolectada desde las estaciones de la Red de Monitoreo de Calidad del Aire de Bogotá - RMCAB. Se observó que aunque, las mencionadas anteriormente, presentan una información en tiempo real, poseen falencias en la recolección de datos.



Gracias

