

Estimació de la variació horària de la concentració d'ozó fent ús de dades meteorològiques i models LUR

Guillermo Camps Pons

Índex

- Introducció
- Recopilació de dades
- Creació de models
- Resultats
- Conclusions

Introducció. L'amença de la contaminació

Actualitat:

- 4,2 millions de morts anuals
- Segueix sent de gran importància a països en desenvolupament

Al 2060:

- Pot incrementar fins a 9 millions
- Pot costar 1% del PIB

Altres efectes en la salut:

- Incrementa la incidència càncer
- Incrementa la incidència malalties respiratòries
- Incrementa la freqüència de malestans respiratoris o mentals



(Imatge de [Photoholic](#) a [Unsplash](#))

Introducció. Mesures per a reduir l'emissió

A llarg termini

- Ús d'energies renovables
- Cotxes elèctrics
- Reducció d'emissions del bestiar

A curt termini

- Definició de límits màxims
- Restriccions a les fonts de contaminants



(Imatge de [CHUTERSNAP](#)
a [Unsplash](#))



(Imatge de [Nicholas Doherty](#)
a [Unsplash](#))

Introducció. Contaminants i l·lindars EEA

Taula 1. Fonts i l·lindars crítics de diferents contaminants.

Contaminant	Tipus	Fonts	L·lindar EEA	Període de mitjana	Cops per any
Matèria particulada 2,5µm (PM _{2.5})	Primari i secundari	Indústria, vehicles, fum tabac	25 µg/m ³	1 any	
Matèria particulada 10µm (PM ₁₀)	Primari i secundari	Indústria, vehicles, fum tabac	50 µg/m ³	1 dia	<35 dies per any
			40 µg/m ³	1 any	
Diòxid de sofre (SO ₂)	Primari i secundari	Indústria, vehicles	350 µg/m ³	1 hora	<24 hores per any
Diòxid de nitrogen (NO ₂)	Primari i secundari	Vehicles, estufes, cuines de gas	200 µg/m ³	1 hora	<18 hores per any
			40 µg/m ³	1 any	
Ozó (O ₃)	Secundari	Foto-oxidació de NO _x i compostos orgànics volàtils	120 µg/m ³	8 hores	<25 dies per any

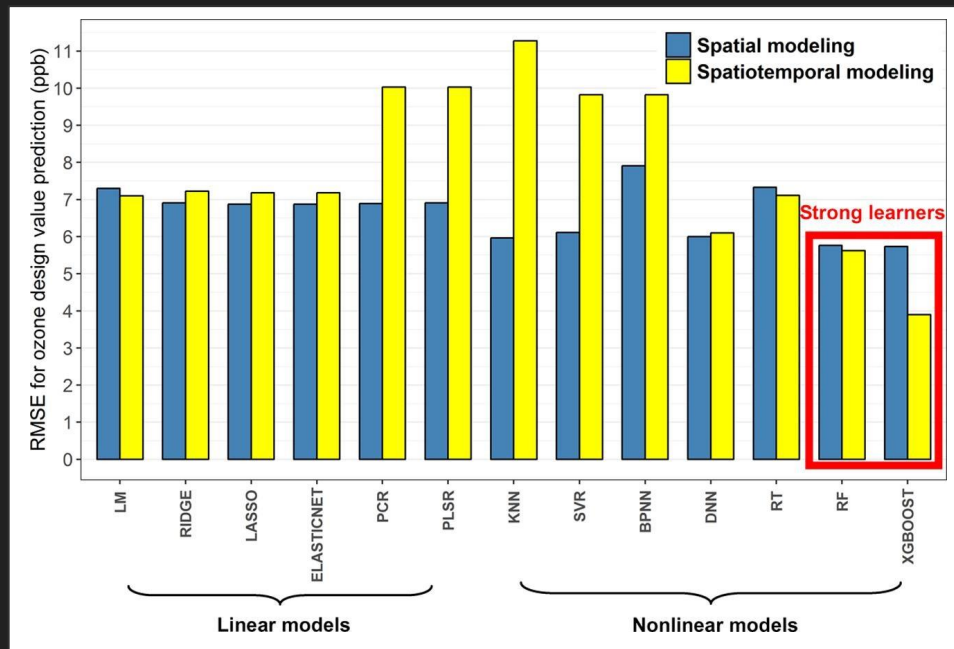
Introducció. Pronòstic de concentracions

Dades

- Concentracions contaminants
- Meteorologia
- Ús de sòl (LUR)

Models

- Numèrics (p.ex. CMAQ)
- Estadístics (p. ex. ARIMA)
- Intel·ligència artificial (p. ex. xarxes neuronals)



Comparació de models d'intel·ligència artificial amb dades d'ús de sòl (Ren et al., 2020)

Introducció. Objectius del treball

Creació dataset

Es capturen dades de diferents fonts i s'integren en un únic dataset.

- Tipus de dades
 - Contaminants
 - Meteorologia
 - Ús de sòl
 - Altres
- Domini
 - Estacions de contaminació de Catalunya
 - Anys 2018 i 2019

Models GBM

Es creen models de gradient boosting machines (GBM) per a estimar la concentració d'ozó (O_3).

- Ajust paràmetres
- Avaluació models
- Extrapolació a altres dominis

Conclusions

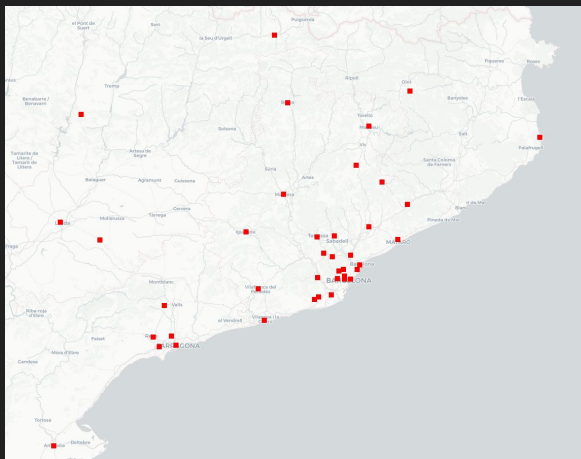
S'extreuen conclusions dels resultats dels models escollits.

- Comparació estat de l'art
- Eficàcia dades d'ús de sòl

Recopilació de dades. Fonts de dades (1)

Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica

- Mesures horàries de NO₂ i O₃ en diferents estacions
- Format csv



Punts de mesurament automàtic de la Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica (2020).

ERA5

- ERA5 single levels (resolució 0,25°)
- ERA5-Land (resolució 0,1°)
- Format netCDF

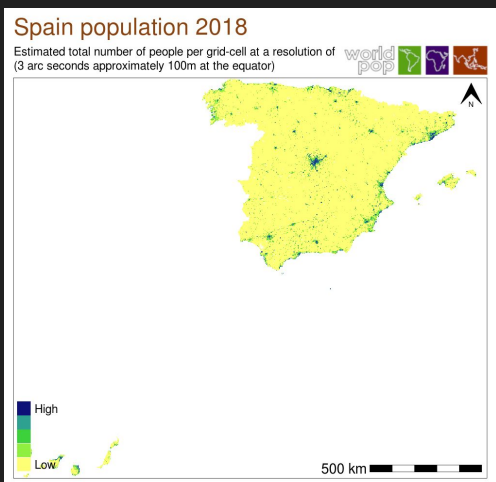
Taula 4. Descripció dels camps d'ERA5.

Descripció	Dataset
Longitud (°)	ERA5, ERA5-Land
Latitud (°)	ERA5, ERA5-Land
Temps (h)	ERA5, ERA5-Land
Component u del vent (m/s)	ERA5, ERA5-Land
Component v del vent (m/s)	ERA5, ERA5-Land
Temperatura (K)	ERA5, ERA5-Land
Radiació solar neta en superfície (J/m ²)	ERA5, ERA5-Land
Radiació UV descendent en superfície (J/m ²)	ERA5
Nuvolositat (interval 0 - 1)	ERA5

Recopilació de dades. Fonts de dades (2)

WorldPop

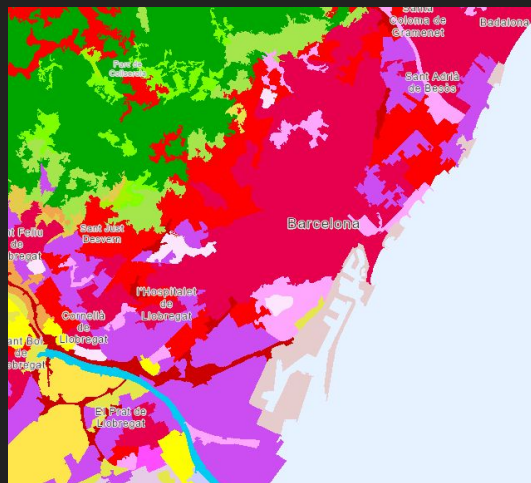
- Població 2018 i 2019
- Format GeoTIFF
- Resolució 100m



WorldPop (2018). Global High Resolution Population Denominators Project.

CORINE Land Cover

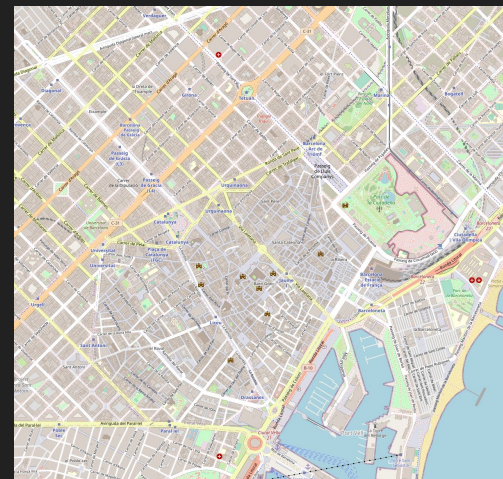
- Ús de sòl 2018
- Format GeoTIFF
- Resolució 100m



Corine Land Cover (2018). © European Union, Copernicus Land Monitoring Service 2021, European Environment Agency (EEA).

OpenStreetMap

- Carreteres
- API Overpass



© OpenStreetMap contributors. (2021)

Recopilació de dades. Sistemes de referència i buffers

Els buffers calculen variables dins una zona d'influència.

- Cercles amb forats
- Projectió azimuthal equidistant a cada punt
- Variables: població, ús de sòl i carreteres

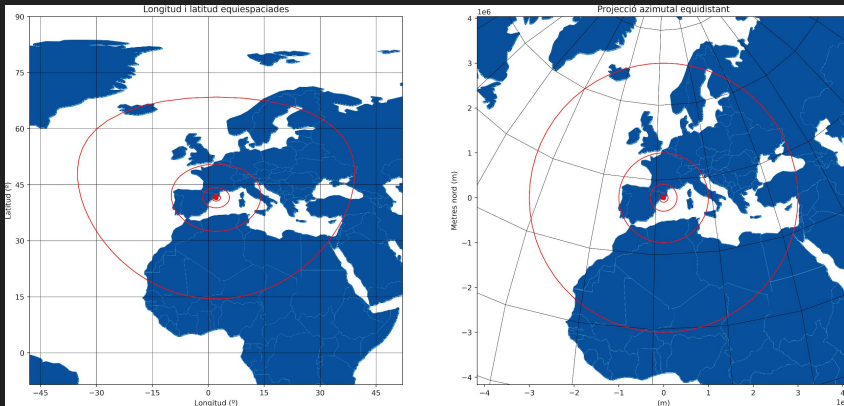


Figura 1. Projectió azimuthal equidistant i punts equidistants.

Els buffers semicirculars tenen en compte la direcció del vent.

- Semicercles amb forats (dues zones)
- Projectió azimuthal equidistant a cada punt
- Variables: ús de sòl i carreteres

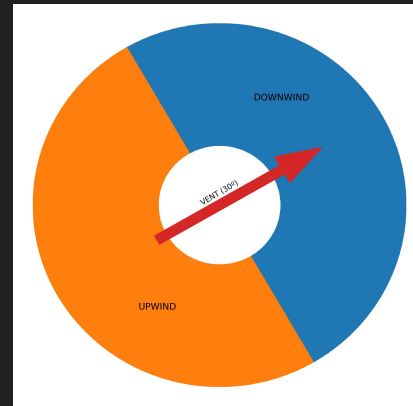


Figura 2. Buffer semicircular.

Recopilació de dades. Dataset final

- Mitjanes temporals
 - Variables: NO₂ i variables meteorològiques
 - De 3, 8 i 24 hores
- Diferències temporals de NO₂
 - 1, 2 i 3 hores
- El conjunt final conté:
 - 664878 files o punts espaciotemporals
 - 150 columnes -> 147 variables predictores -> 135 no constants

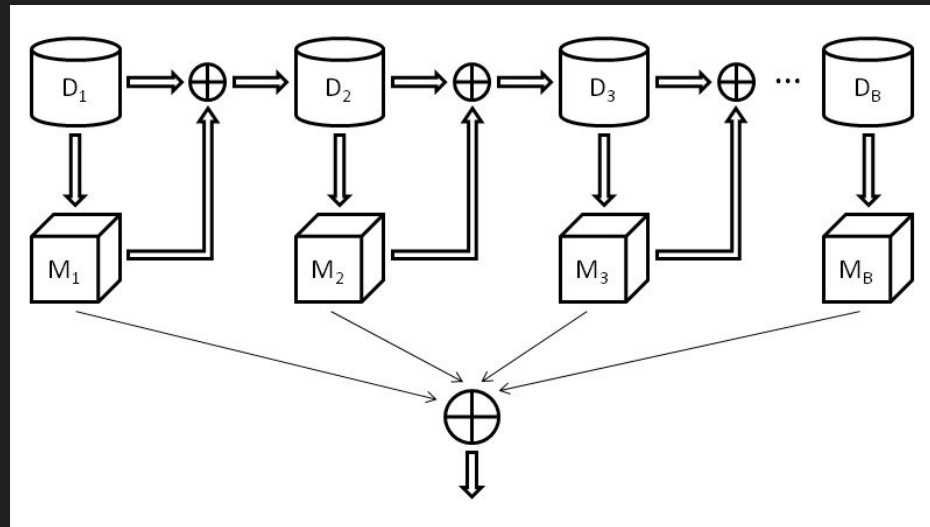
Creació de models. Gradient Boosting Machines (GBM)

Boosting

- Models senzills connectats seqüencialment
- Milloren l'error del model anterior de la seqüència
- Prediccions en paral·lel
- Funció de pèrdua exponencial

Gradient Boosting Machines

- Generalització del boosting
- Funció de pèrdua diferenciable qualsevol



Esquema general algoritme boosting (Gironés et al., 2017)

Creació de models. Selecció de variables

Models GBM senzills

- Fins a 15 variables
 - Avaluació iterativa
 - Variables que minimitzen l'error (millor R^2)
- 20 fulles màxim
- 100 arbres màxim
- 64% entrenament, 16% validació, 20% test

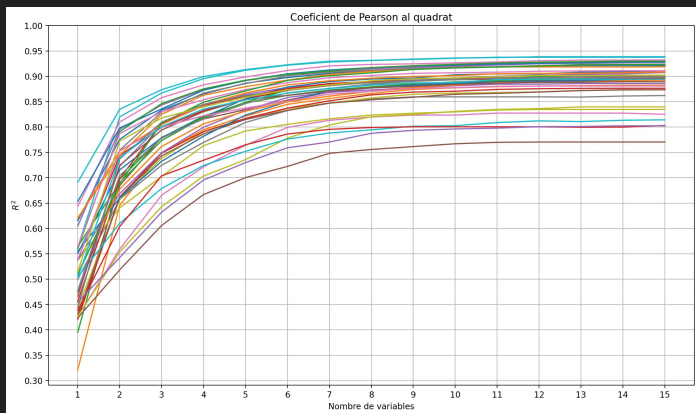


Figura 3. R^2 per les diferents estacions

Taula 8. Variables escollides per totes les estacions

Nombre variables	Variables escollides	R^2
1	UV_SURFACE_MEAN_8h	0,388
2	NO2	0,651
3	LANE_DENSITY_CIRCULAR_2000-3000m	0,729
4	MES	0,780
5	RADIATION_SURFACE_MEAN_3h	0,803
6	WIND_SPEED_MEAN_8h	0,821
7	DIA	0,834
8	TEMPERATURE_MEAN_8h	0,841
9	ANY	0,845
10	FOREST_CIRCULAR_1000-2000m	0,848
11	HORA	0,850
12	ARTIFICIAL_CIRCULAR_3000-5000m	0,854
13	WIND_SPEED_MEAN_24h	0,854
14	ROAD_DENSITY_CIRCULAR_1000-2000m	0,856
15	POPULATION_CIRCULAR_0-300m	0,856

Creació de models. Nombre de variables

Models GBM més complexos

- S'utilitzen les variables anteriors
- 31 fulles màxim
- 5000 arbres màxim
- 64% entrenament, 16% validació, 20% test

Taula 9. Resultats models de 7 i 9 variables.

Model	RSME ($\mu\text{g}/\text{m}^3$)	Pearson (R)	Person al quadrat (R^2)	Coefficient de determinació (ρ^2)
7 variables	9,884	0,954	0,911	0,911
9 variables	8,548	0,966	0,933	0,933

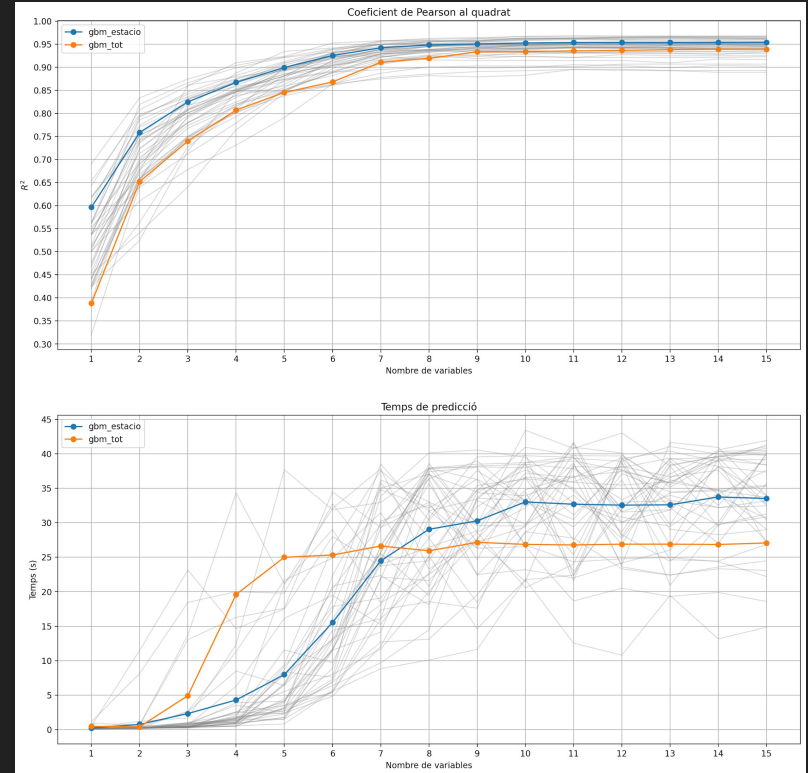


Figura 5. Models segons el nombre de variables

Resultats. Anàlisi de residus (1)

Residu = valor real - predicció

- Residus centrats al voltant del zero i simètrics
- Els residus no acaben de seguir una distribució normal
 - No corresponen a errors aleatoris
 - Els models GBM no poden explicar una certa variació de l'ozó

Mitjanes i desviacions dels residus

Model	Mitjana	Desviació
7 variables	-0,0005 $\mu\text{g}/\text{m}^3$	9,8836 $\mu\text{g}/\text{m}^3$
9 variables	-0,0164 $\mu\text{g}/\text{m}^3$	8,5479 $\mu\text{g}/\text{m}^3$

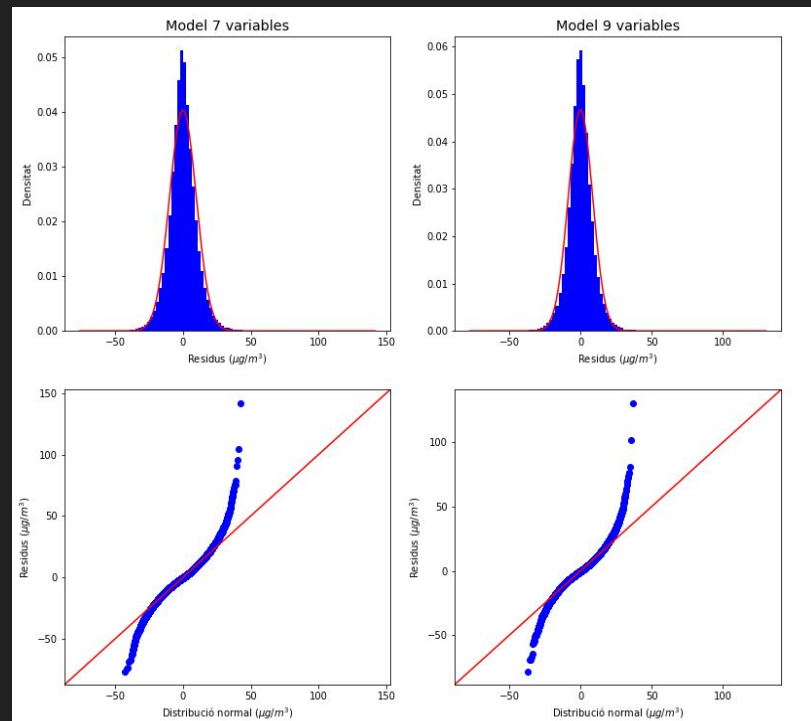


Figura 6. Comparació dels residus amb una distribució normal.

Resultats. Anàlisi de residus (2)

- Rang interquartil distribuït simètricament
 - Com els residus en general
 - El model extreu prou informació de les variables
- Prediccions millors per:
 - NO_2 alt
 - Velocitat del vent alta

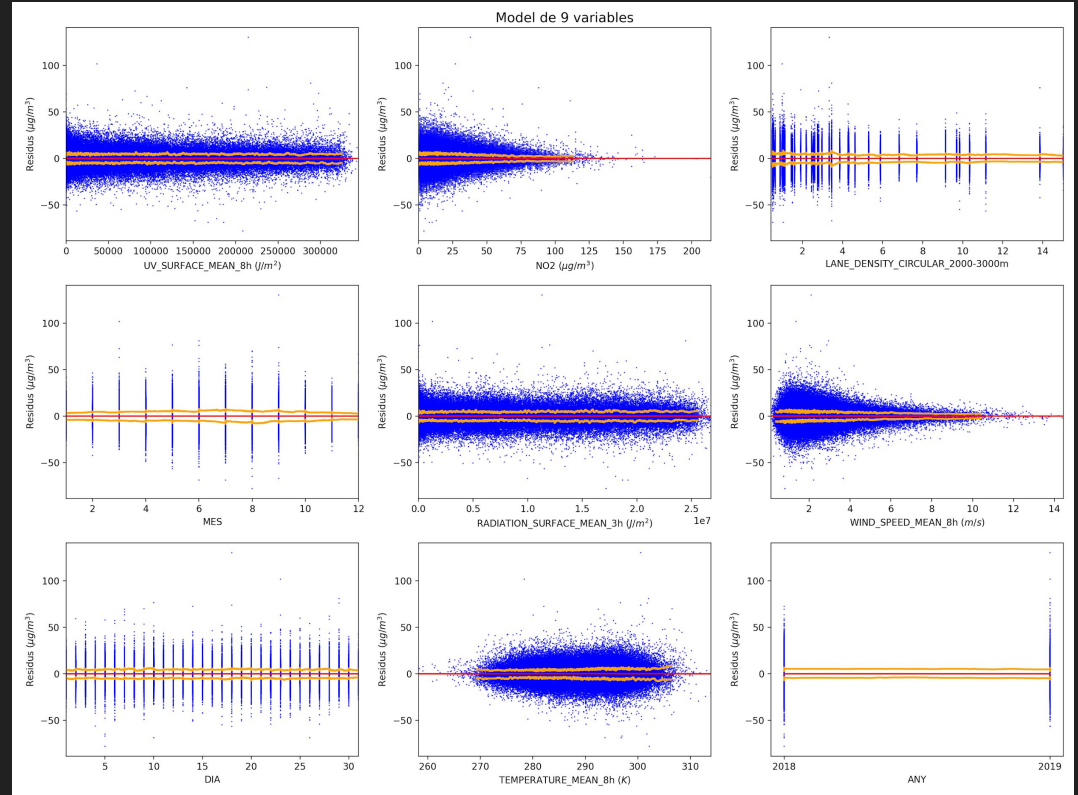
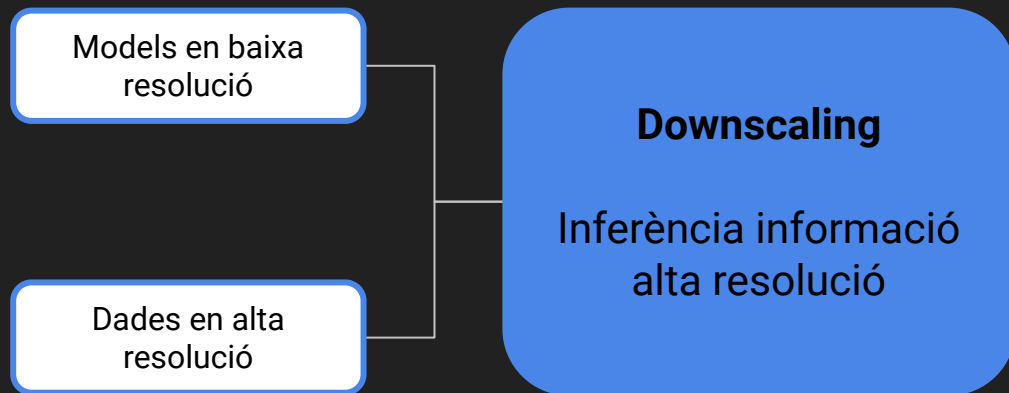


Figura 8. Residus en funció de cada variable del model de 9 variables. Es representen Q1 i Q3.

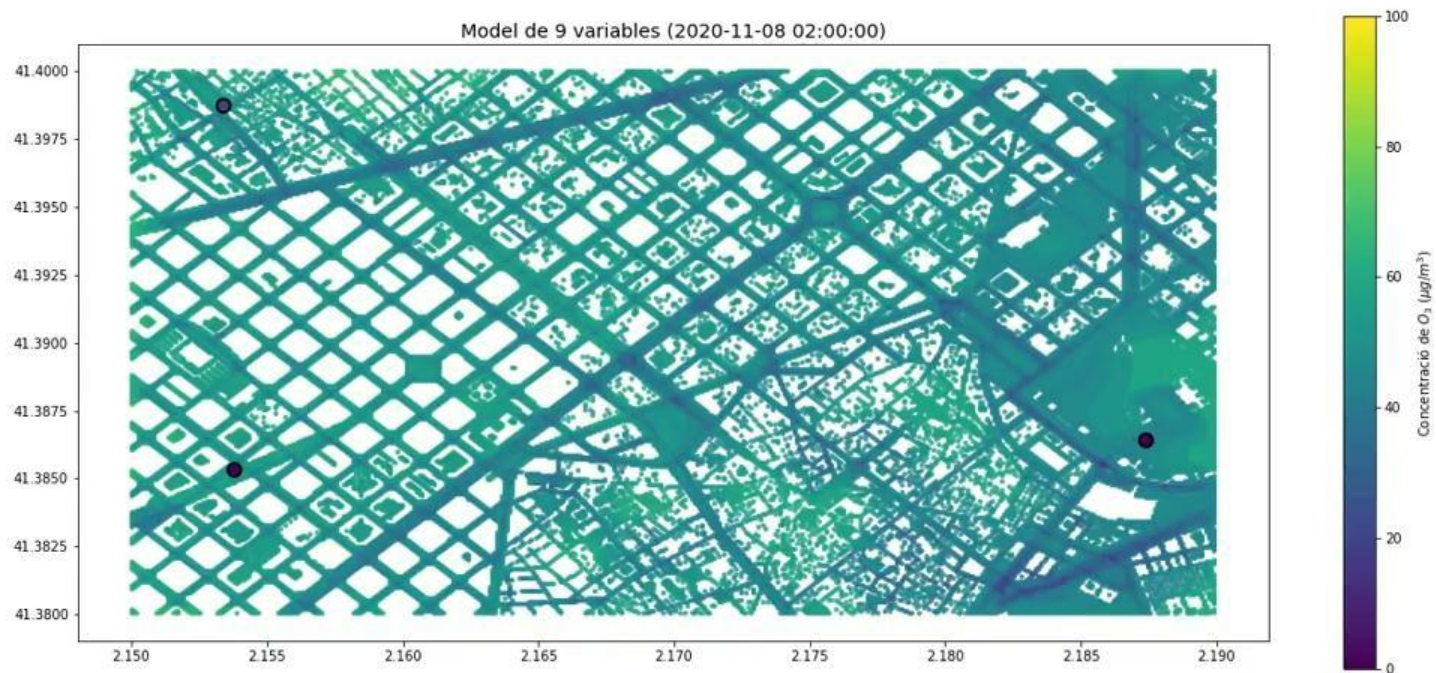
Resultats. Downscaling (1)

Dades d'alta resolució

- Receptors de NO₂ als carrers de Barcelona
 - Mesures sense calibració
- Models numèrics street canyon pels vents
 - Velocitat i direcció del vent
- Zona d'uns 10km²
 - Longitud de 2,12° a 2,22°
 - Latitud de 41,35° a 41,43°
 - Novembre 2020



Resultats. Downscaling (2)



Conclusions.

Comparació amb l'estat de l'art

Resultats comparables fent ús de:

- Model senzill (GBM)
- Nombre de variables similar

Taula 15. Comparació de l'estat de l'art amb els resultats dels models obtinguts.

Nom	Predicció	Model	Nombre variables	R ²
Ross et al. (2005)	NO ₂	MLR	4	0,79
Beelen et al. (2009)	O ₃	kriging	3	0,70
Adam-Poupart et al. (2014)	O ₃	BME	6	0,653
Zhan et al. (2018)	Màxim diari O ₃ (mitjana 8h)	RF	13	0,69
Chen et al. (2021)	màxim anual O ₃ (mitjana 8h)	LUR	5	0,59
	Màxim diari O ₃ (mitjana 8h)	LUR + BME		0,80
Wang, J. et al. (2021)	O ₃	SVR + MLR/ST		0,90
Wong et al. (2021)	PM _{2,5}	GBM	6	0,73
		kriging + GBM		0,94
Wang, Z. et al. (2021)	PM _{2,5}	LSTM	8	0,83
Aquest treball	O3	GBM	7	0,911
			9	0,933

Conclusions.

Variables utilitzades

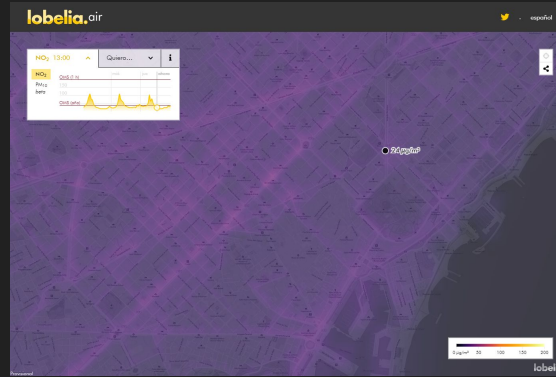
- UV i NO₂ bon predictors
 - Són part de foto-oxidació de NO_x
- Les variables meteorològiques i temporals són rellevants
- Les variables d'ús de sòl són poc importants
 - Excepció: densitat de carrils
- Els buffers semicirculars no han tingut èxit

Taula 8. Variables utilitzades i la seva importància.

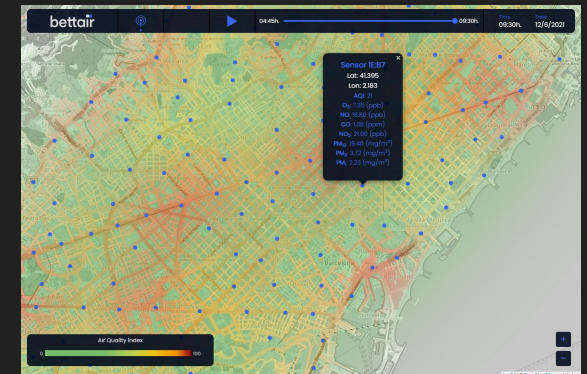
Variable	Selecció		Guany	
	7 variables	9 variables	7 variables	9 variables
UV_SURFACE_MEAN_8h	22004 (14,669%)	16976 (11,317%)	9,234e+08 (40,213%)	9,061e+08 (38,721%)
NO2	16702 (11,135%)	14093 (9,395%)	6,322e+08 (27,534%)	6,296e+08 (26,905%)
LANE_DENSITY_CIRCULAR_2000-3000m	19459 (12,973%)	18732 (12,488%)	1,981e+08 (8,627%)	1,951e+08 (8,336%)
MES	19019 (12,679%)	13676 (9,117%)	1,717e+08 (7,479%)	1,496e+08 (6,392%)
RADIATION_SURFACE_MEAN_3h	22538 (15,025%)	17566 (11,711%)	1,865e+08 (8,124%)	1,689e+08 (7,218%)
WIND_SPEED_MEAN_8h	24713 (16,475%)	19891 (13,261%)	1,134e+08 (4,937%)	1,091e+08 (4,664%)
DIA	25565 (17,043%)	23426 (15,617%)	7,085e+07 (3,085%)	7,846e+07 (3,353%)
TEMPERATURE_MEAN_8h		20311 (13,541%)		7,954e+07 (3,399%)
ANY		5329 (3,553%)		2,368e+07 (1,012%)

Conclusions. Línies futures

- **Avaluació addicional del dataset**
 - Altres models
 - Altres llocs (transfer learning)
- **Prediccions d'alta resolució i a temps real per l'ozó**
 - Existeixen algunes aplicacions per altres contaminants
 - Permet adaptar el dia a dia



Concentracions de NO₂ a Barcelona. Lobelía.
<https://aire-barcelona.lobelia.earth/es/>



Qualitat de l'aire a Barcelona. Bettair.
<https://map.bettair.city/barcelona>

Gràcies