20th International Conference on Harmonisation within
Atmospheric Dispersion Modelling for Regulatory Purposes
14-18 June 2021, Tartu, Estonia

Use of modelling in support of EU air quality directives, including FAIRMODE activities

A novel modelling-based method for air quality zoning. Application to the Madrid Region

Rafael Borge[1], Daeun Jung[1], Iciar Lejarraga[1], Enrique Crespo[2], Ricardo Vargas[2], David de la Paz[1], Jose María Cordero[1], Jaime S. Gallego[2]

[1] Laboratory of Environmental Modelling. Universidad Politécnica de Madrid (UPM)

[2] Consejería de Medio Ambiente, Ordenación del Territorio y Sostenibilidad. Air Quality Service

daeun.jung@upm.es

OUTLINE

1. Introduction
   - Current situation
   - Objectives
2. Methodology
   - CMAQ dataset
   - Optimal zoning
   - Zoning assessment
   - Coverage and redundancy
3. Results
   - Optimal Zoning
   - Three zonings
   - Zoning assessment
   - Coverage and Redundancy
4. Conclusions and further work

1. INTRODUCTION

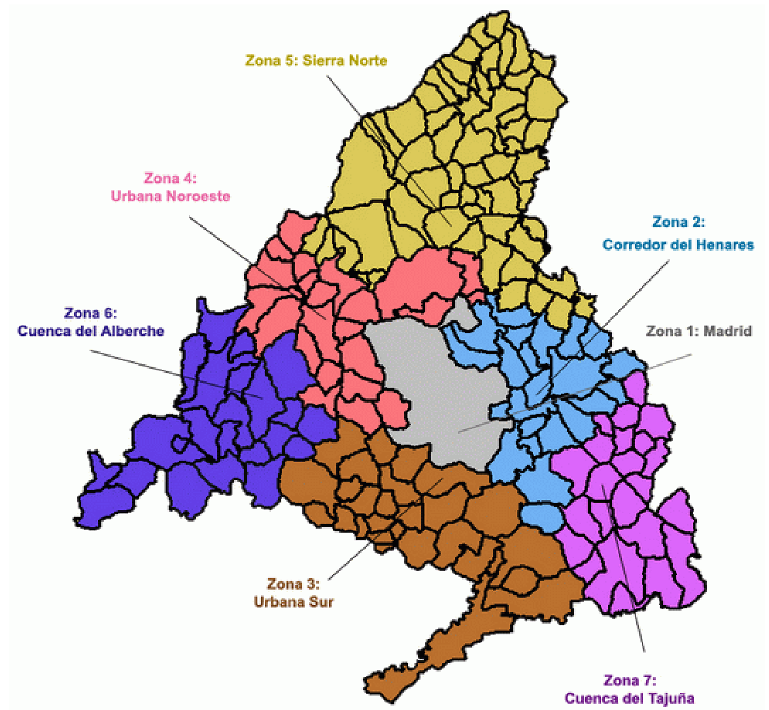
Current Situation

- European Air Quality Directive (AOD) (2008/50/EU): Air quality zones have **homogeneous** air quality and should be revised every 5 years. Therefore, if there is a station that exceeds limit values (LV), the whole zone is in a non-compliance situation.

- The Madrid region is located in the center of the country
  - A population of about 6.8 million (2020)
  - 179 municipalities
  - 48 air quality monitoring stations.
  - 7 air quality zones

- The AQ zones are based on the variables: population, geography, emission sources, meteorology, etc.

- Last review: 2014



Current zoning of the Madrid Region (CM, 2019)

→ The methodology to revise AQ zones is not standardized

Objectives

- Assess the current zoning using the pollutants that have a major impact on human health and vegetation (NO_2, O_3, PM_{10} and $PM_{2.5}$)

- Identify an alternative that represents the more homogenous areas from the air quality point of view with a new methodology among three zoning options:

- Current
- Optimal
- Proposed (preliminary, under discussion) by the Madrid Greater Region air quality service

Pollutant	Year	EU: EU-28 urban population exposure to concentrations **above** EU standards ([1])	WHO: Urban population exposure to concentrations **above** WHO AQG value ([2])	EU: Reporting stations that registered concentrations **above** EU standards ([1])	WHO: Reporting stations that registered concentrations **above** WHO AQG value ([2])
PM_{10}	2018	15 %	48 %	19 % 20 EU MS (EU-28) and six other countries	53 % All countries except Estonia, Iceland and Ireland
PM_{10}	2019 (preliminary) for 30 countries	(*)	(*)	9 % 13 EU MS and two other countries	37 % All countries except Estonia, Finland, Iceland, Ireland and Luxembourg
$PM_{2.5}$	2018	4 %	74 %	4 % Six EU MS and two other countries	70 % All countries except Estonia, Finland Iceland and Ireland
$PM_{2.5}$	2019 (preliminary) for 27 countries	(*)	(*)	2 % 4 EU MS and two other countries	58 % All countries except Estonia, Finland, Iceland, Ireland, Luxembourg, Norway and Sweden
O_3	2018	34 %	99 %	41 % 20 EU MS and five other countries	96 %
O_3	2019 (preliminary) for 32 countries	(*)	(*)	27 % 18 EU MS and two other countries	98 %
NO_2	2018	4 % ([3])	4 % ([3])	8 % 16 EU MS and three other countries ([3])	8 % 16 EU MS and three other countries ([3])
NO_2	2019 (preliminary) for 33 countries	(*)	(*)	3 % 12 EU MS and one other country ([3])	3 % 12 EU MS and one other country ([3])
BaP	2018	15 %	75 %	27 % 14 EU MS	83 % All countries except Cyprus.
BaP	2019 (preliminary)	(*)	(*)	([4])	([4])
SO_2	2018	<1 %	19 %	<1 % One EU MS and three other countries	33 % 27 countries
SO_2	2019 (preliminary) for 32 countries	(*)	(*)	<1 % One EU MS	28 % 22 countries

Notes: ([1]) The following EU standards are considered: PM_{10} daily limit value, $PM_{2.5}$ annual limit value, O_3 target value, NO_2 annual limit value, BaP target value and SO_2 daily limit value. Please see Table 1.1.

([2]) For BaP, reference level. Please see Table 1.3.

([3]) For NO_2, both the EU annual limit value and the WHO AQG are set at the same.

([4]) BaP is not measured automatically and therefore is not included in the UTD data exchange.

(*) Estimates of urban population exposure are not available for 2019.

Sources: EEA (2020a, 2020c).

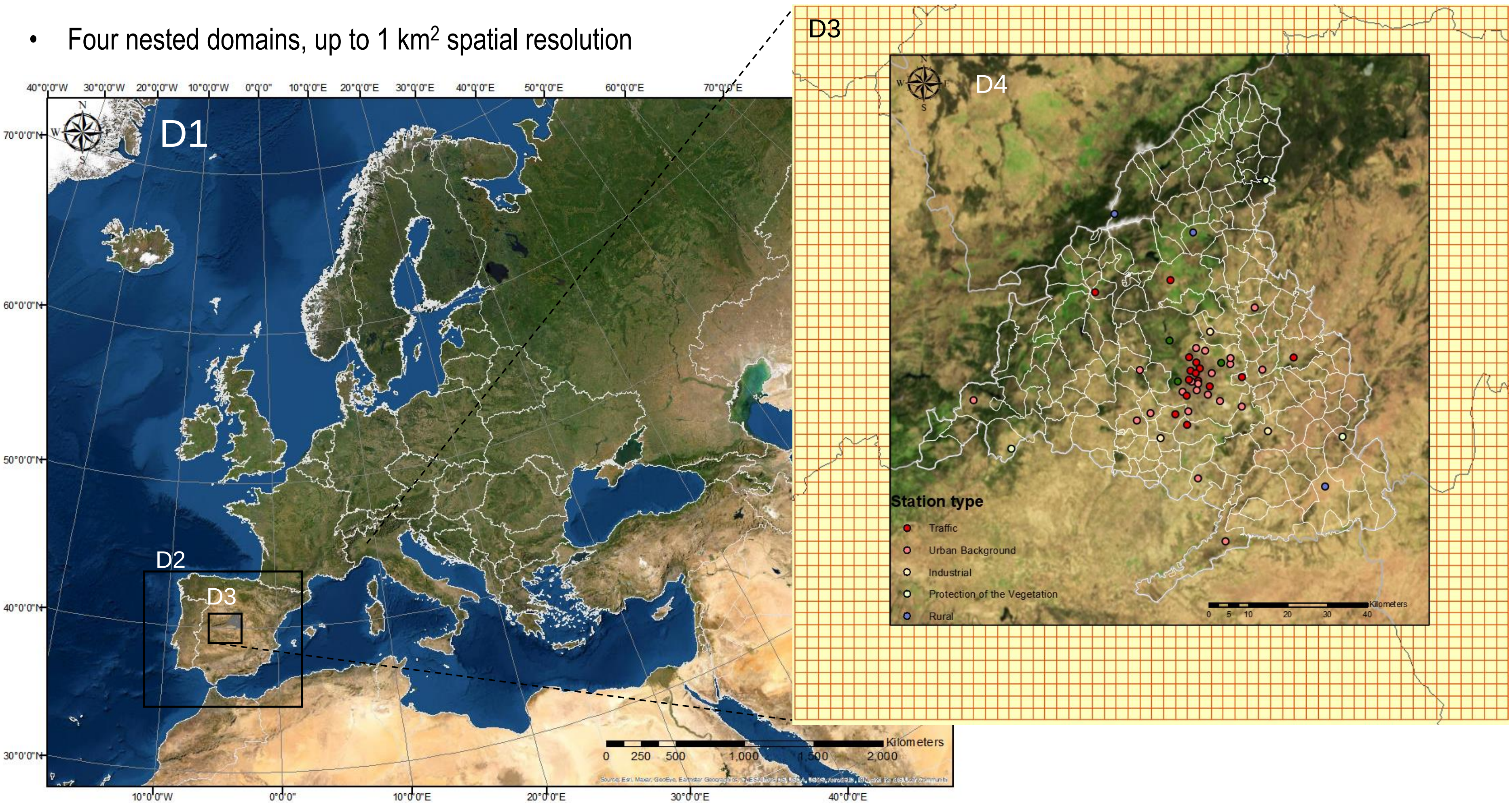
Population exposure to concentrations and reporting stations of concentrations above EU standards and WHO AQG values (EEA, 2020)

2. Methodology

Mesoscale simulation system

WRFV3.7.1 (Weather Research and Forecasting)	+	SMOKEV3.6.5 (Sparse Matrix Operator Kernel Emission)	+	CMAQV5.0.2 (Community Multiscale Air Quality)
- Eulerian meteorological model		- Emission model - Input for the chemical-transport model		- Developed by USEPA - Simulates pollutant concentrations

- Four nested domains, up to 1 km^2 spatial resolution

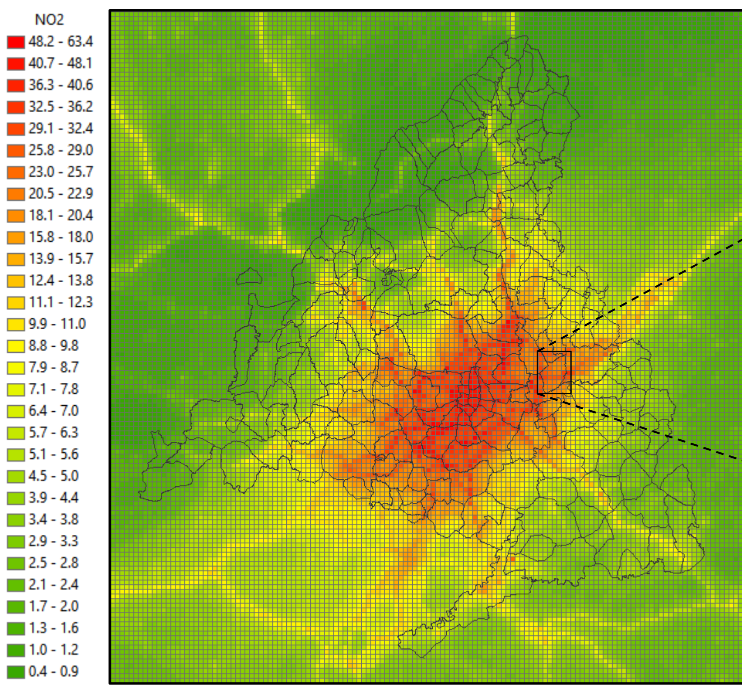


- Model outputs (for the year 2015) processed according to Directive 2008/50/EC

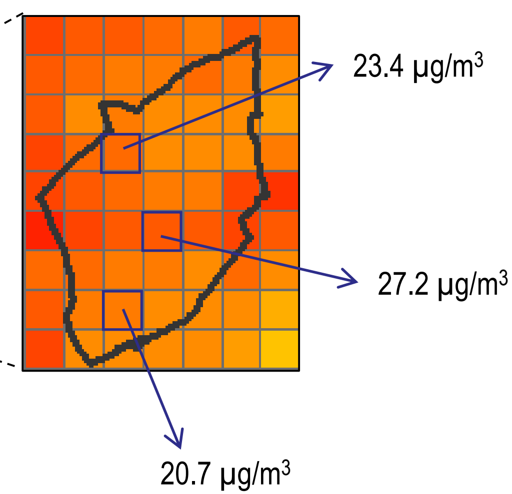
NO_2	Annual Mean	99.8 Percentile of hourly concentration
O_3	AOT 40	93.2 Percentile of 8-hourly maximum concentration
PM_{10}	Annual Mean	90.4 Percentile daily mean concentration
$PM_{2.5}$	Annual Mean	

Mean
Median
Quartile 25 (Q1)
Quartile 75 (Q3)
Inter Quartile Range (IQR)
Standard Deviation (sd)
Coefficient Variation (cv)

2015



NO_2 annual mean concentration

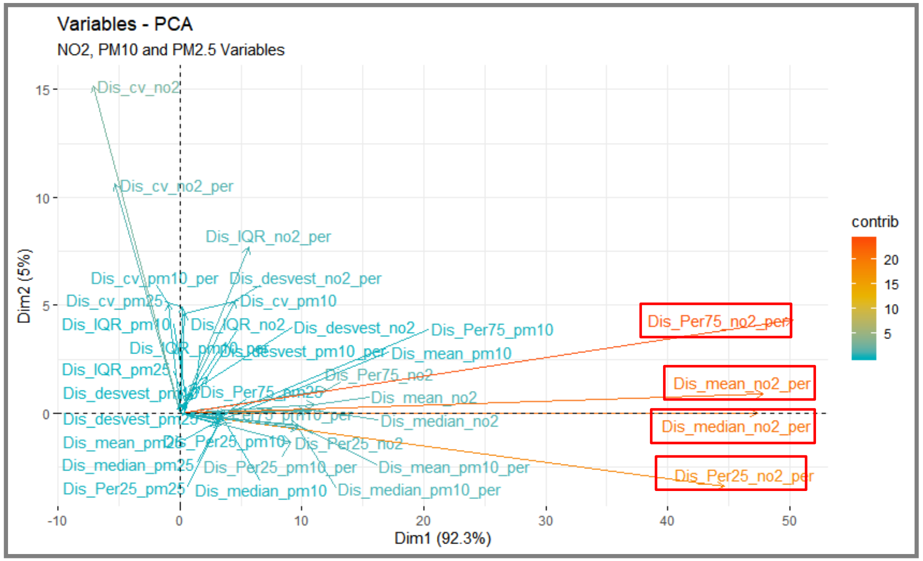


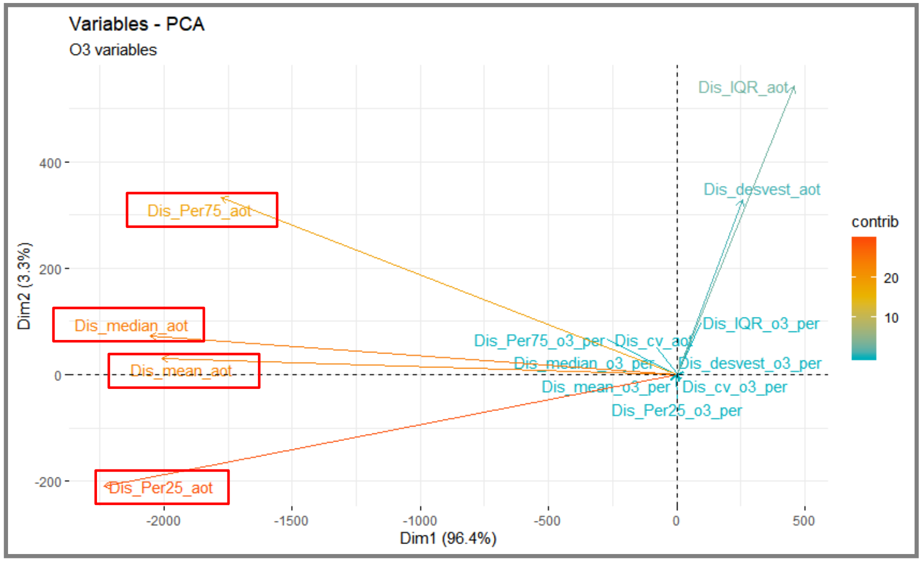
Optimal Zoning

- Principal Component Analysis (PCA) to identify the most relevant metrics in two groups due to the different characteristics of the pollutants.

  → Group 1: NO_2, PM_{10} and $PM_{2.5}$

  → Group 2: O_3





The PCA results between the NO_2, PM_{10} y $PM_{2.5}$ variables (left) and O_3 variables (right)

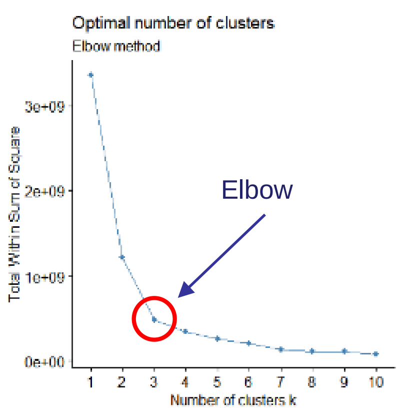
- Performed K-means clustering analysis with the defined parameters to process the classification of the municipalities of the Madrid region.

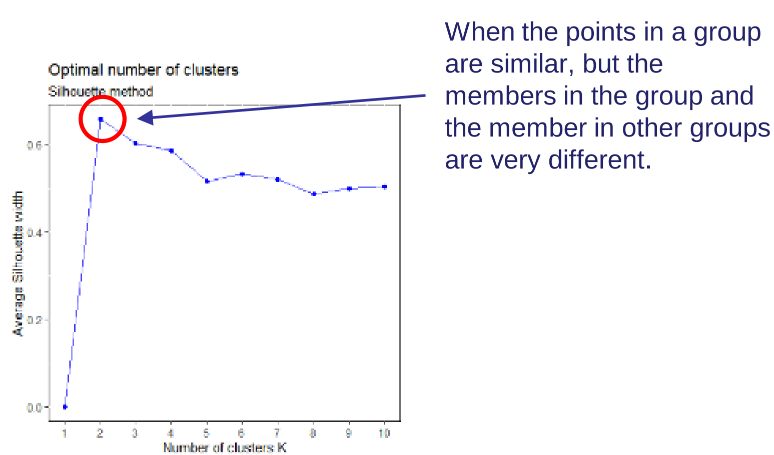
- ***K-means* clustering analysis** is known as *unsupervised learning*, which is a method to classify the observations of dataset into groups by homogenous character.
    1. Assign randomly *k* numbers of clusters.
    2. Calculate the distances between centroid points and other observations in *k* cluster.
    3. Reassign the mean value of the distances to new centroid points of clusters.
    4. Repeat until finding adequate centroid points that have a minimum distance with the observations in a cluster.

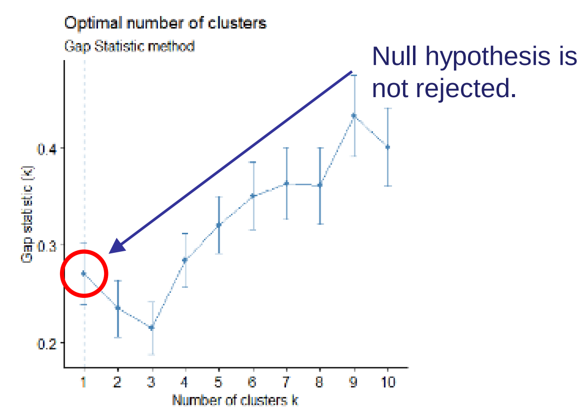
→ **Elbow method** shows a graph, where the point values drastically decreases is elbow, which is the optimal number of groups for the classification.

→ **Silhouette method** finds similarity between a point and the others in a group, comparing with other points in the other groups, using mean values.

→ **Gap statistic method** is based on null hypothesis model, which is there is a group or a cluster in a dataset. When a reasonable number of groups appears, the null hypothesis model is rejected.







Zoning Assessment

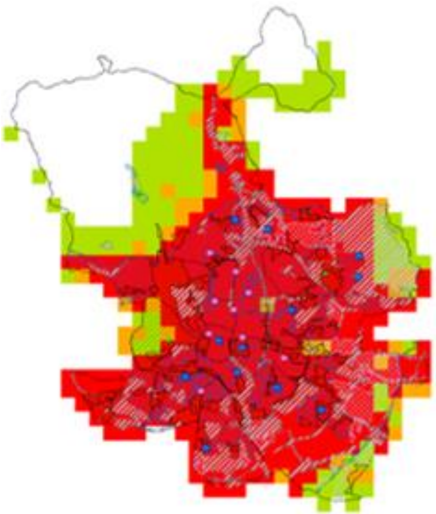
- Concentration distribution of each zone and each municipality through boxplot graphs
- Statistical tests: Kruskal-Wallis and Dunn tests to identify if there are statistically significant differences between zones.

```
data: x and group
Kruskal-Wallis chi-squared = 8661.7957, df = 3, p-value = 0


                           Comparison of x by group
                                 (Bonferroni)
Row Mean-|
Col Mean |          1          2          3
---------+---------------------------------
       2 |  -2.355552
         |     0.0555
         |
       3 |  -29.22541  -31.77240
         |    0.0000*    0.0000*
         |
       4 |  -65.97963  -77.10226  -50.66477
         |    0.0000*    0.0000*    0.0000*

alpha = 0.05
Reject Ho if p <= alpha/2
```

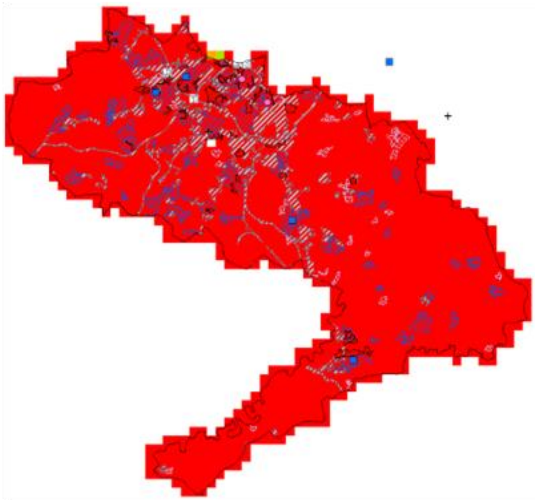
Kruskal-Wallis and Dunn tests for 99.9th percentile of hourly concentration of NO_2.

Coverage and Redundancy of available measurements (current AQ networks)

- Representativeness area: ±15% respecting the model prediction at air quality monitoring sites.
- **Intersection** areas measured by **more than 2 stations**: **Redundancy**
- **Union** areas measured by **more than 1 station**: **Coverage**



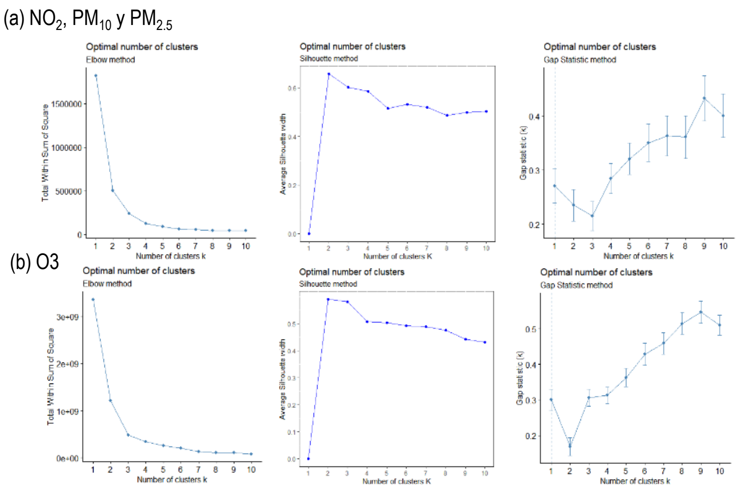
Coveragy and Redundancy of 99.8th percentile of NO_2 of the Madird city



Coverage and Redundancy of AOT 40 of zone 3 of the current zoning

3. Results

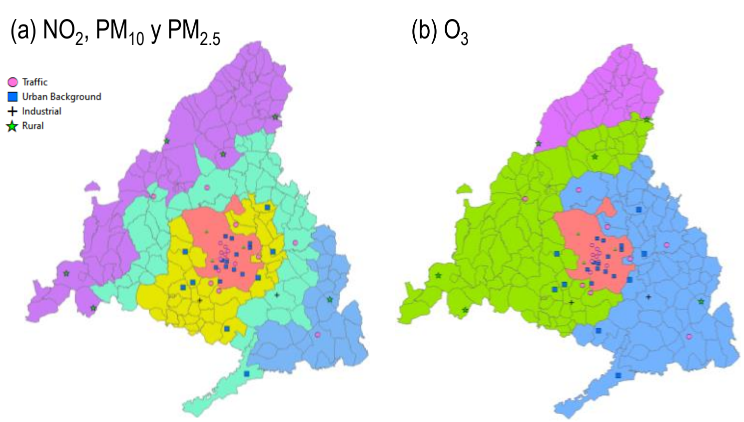
Optimal Zoning



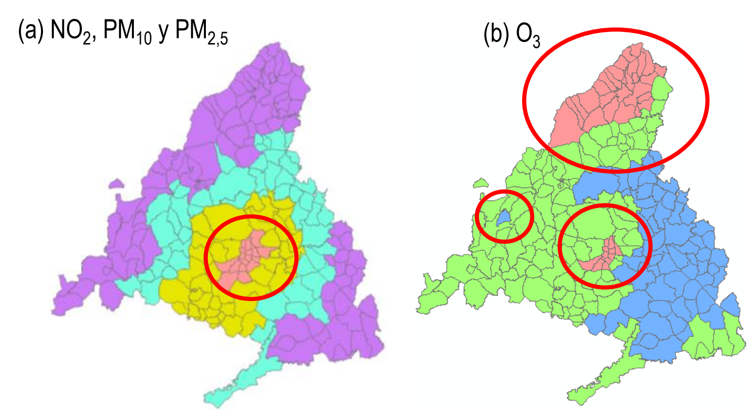
K-mean **clustering analysis results**

Optimal Number

- 4 for NO_2, PM_{10} and $PM_{2.5}$
- 3 for O_3

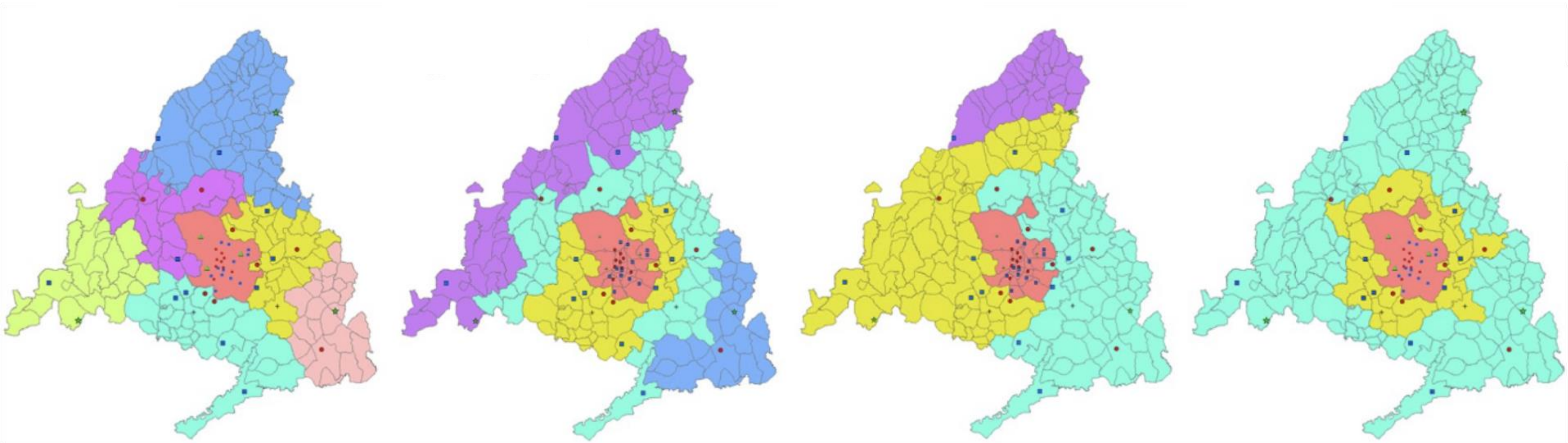


Optimal air quality zoning



Cluster division

The three zonings



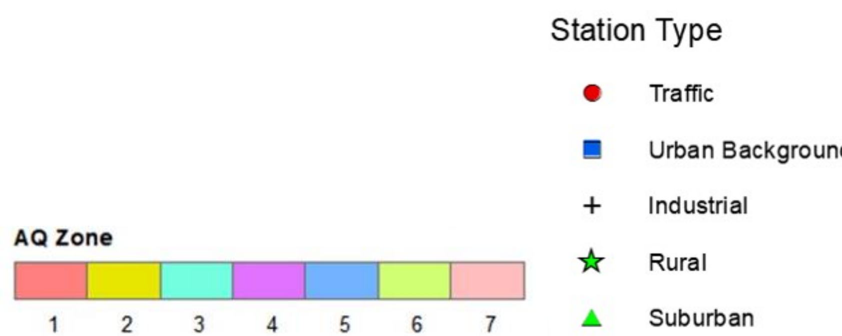
Current zoning

(a) NO2, PM2.5, PM10

(b) O_3

Proposed zoning

Optimal zoning



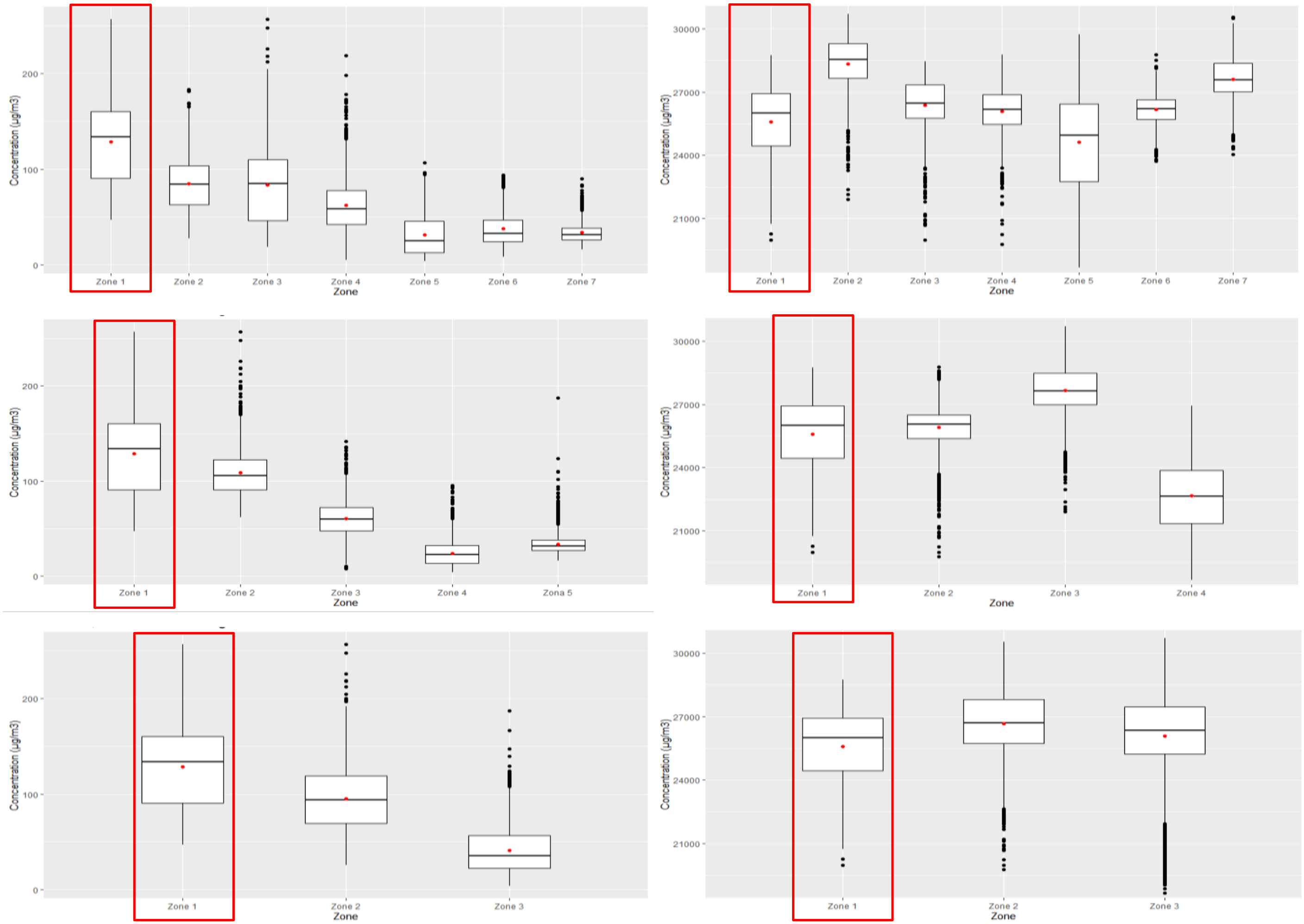
Zoning Assessment

99.8 Percentile of Hourly Concentration of NO_2 | **AOT40 of O_3**

Current zoning

Optimal Zoning

Proposed Zoning



- According to the statistical tests, some zones of the current zoning have very similar characteristics
- The optimal zoning shows well-separated and homogenous zones

Zoning Assessment

- Total SSE (Sum of Squared Error) = Between SSE (BSS) + Within SSE (WSS)

  -TSS: sum of distances between the objects and the mean value of the entire dataset

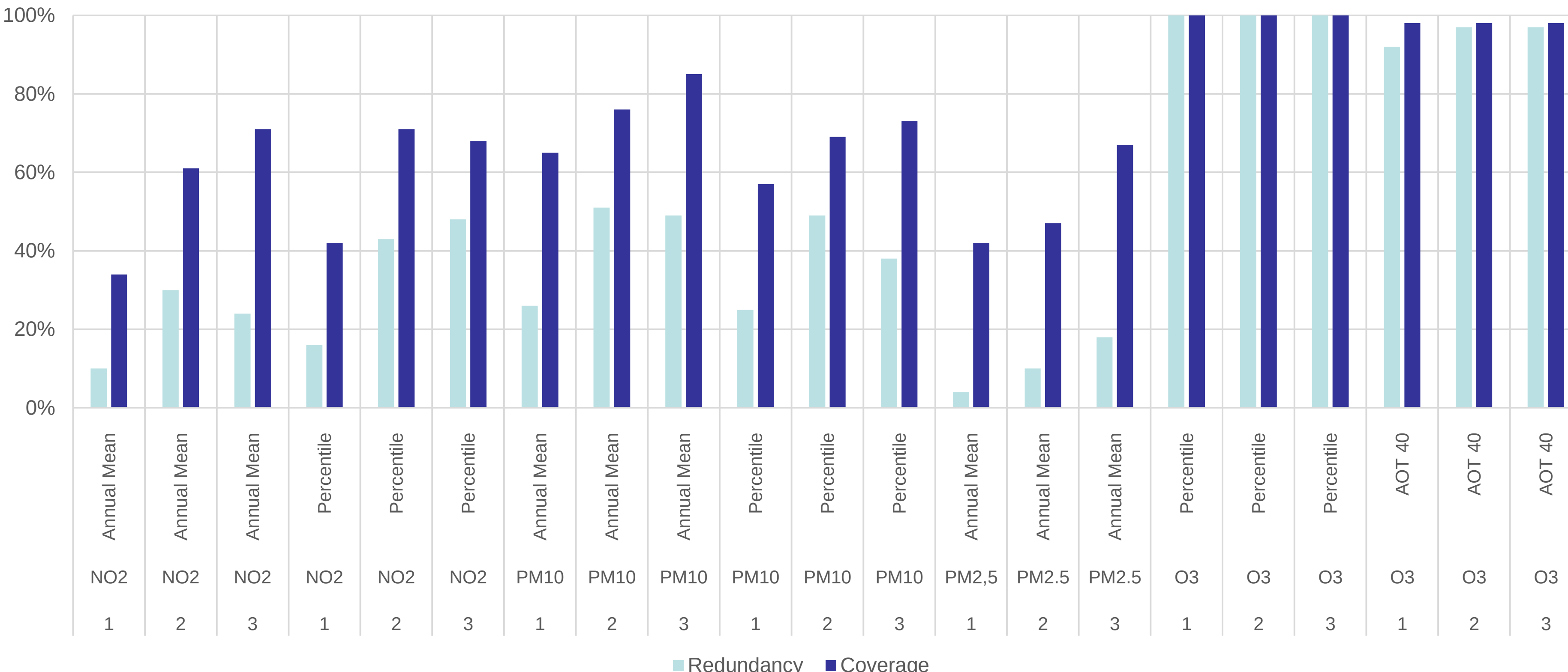
  -BSS: sum of distances between cluster centers

  -WSS: sum of distances between an object and the center of a cluster

- When BSS/TSS closes to 1, it guarantees a good separation.

Pollutant	Parameter	Current Zoning		Optimal Zoning		Proposed zoning	
		k	BSS/TSS (%)	k	BSS/TSS (%)	k	BSS/TSS (%)
NO_2	Annual Mean	7	56.7%	5	**60.3%**	3	45.5%
	Hourly Percentile	7	71.3%	5	**88.3%**	3	65.0%
PM_{10}	Annual Mean	7	48.7%	5	**58.9%**	3	33.6%
	Daily Percentile	7	57.5%	5	**76.3%**	3	40.8%
$PM_{2.5}$	Annual Mean	7	44.0%	5	**53.1%**	3	29.0%
O_3	8-hourly Percentile	7	45.0%	4	**68.8%**	3	7.9%
	AOT 40	7	48.3%	4	**78.1%**	3	4.8%

Coverage and Redundancy



1: Current
2: Optimal
3: Proposed

- In general, redundancy increases along with coverage
- The current one has poor coverage and redundancy
- The optimal option presents better results
- The proposed has better coverage but less redundancy in some parameters comparing to the optimal (such as NO_2 annual mean and the parameters of PM_{10})

4. Conclusions and further work

- New methodology = CMAQ + K-means clustering → new optimal zoning
- Current zoning: heterogeneous zones and low coverage
- Optimal zoning: well-separated and homogenous zones and better coverage of the stations
- Proposed zoning: poor separation but better coverage and less redundancy

- Test the methodology presented in other regions
- Analyze the optimal zoning from the population exposure perspective

Study carried out within the AIRTEC-CM (urban air quality and climate change integral assessment) scientific programme funded by the Directorate General for Universities and Research of the Greater Madrid Region (S2018/EMT-4329)

https://airtec-cm.es/

UNIÓN EUROPEA
Fondos estructurales
Invertimos en su futuro

The Madrid City Council and Madrid Greater Region are acknowledged for providing traffic and observational data and technical support

Thank you for your attention!

daeun.jung@upm.es