

Data processing possibilities for large-scale IAQ prediction

A GLM Case Study in natural ventilated dwellings

46th AIVC - 14th TightVent - 12th venticool Conference
Sonia García Ortega

1

The logo consists of the letters 'M' and 'L' in a bold, red, sans-serif font. The 'M' is positioned to the left of the 'L', and they are slightly overlapping. The letters have a subtle drop shadow effect.

Data processing possibilities for large-scale IAQ prediction

A GLM Case Study in natural ventilated dwellings

46th AIVC - 14th TightVent - 12th venticool Conference
Sonia García Ortega

2

The problem



Growing interest in integrating IAQ into the EU Renovation Wave -> need to know current situation.

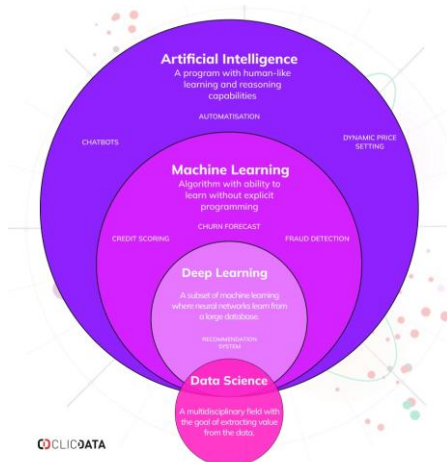
How can we target renovation support at large scale to the worst-performing cases if we cannot identify them?

Monitoring is expensive, invasive, and non-scalable. Simulations are complex.

Can we develop **IAQ estimates** for the existing housing stock?

3

The way forward



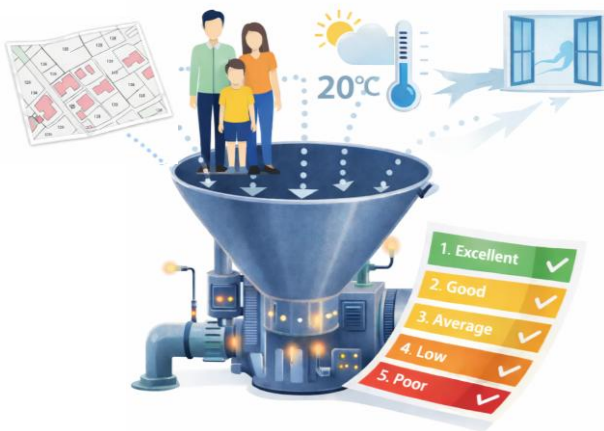
Machine Learning is a subset of artificial intelligence that enables systems to **identify complex patterns** and make **data-driven predictions** by learning from experience rather than following rigid, pre-defined instructions.

what happened?

→ what is likely to happen?

4

The way forward



Goal: Estimate IAQ using simple, available input data

Many interacting factors (climate, occupancy, volume, infiltration...) **not always intuitive** e.g. is occupancy really negative?

Can ML help identify hidden drivers and predict IAQ at large-scale?

5

Most important thing: Data



Data availability is the current limitation

Expected to become more available in the medium term thanks to:

- Monitoring campaigns
- IOT
- Initiatives of data-base like Pandora

6

Prospective study

Generalized Linear Model



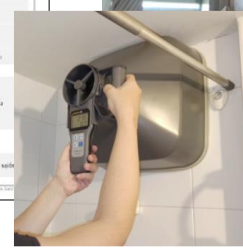
Approach: Generalized Linear Model (GLM) with Gamma distribution and inverse link.
Sample: 72 monitoring campaigns across 12 real dwellings in Madrid, riche contextual data
Objective: Predict CO₂ concentration as an IAQ indicator without direct measurements.

7

Prospective study



Código vivienda			
Otros genericos			
Ubicación	Barrio: Mejía Laguna	Distrito: Moncloa	Código Suelo: Mejía 03044
Entrame	Urbano consolidado	Usos edificios: Residencial	Superficie DUE: 67
Topografía	100-500	Planta ocupada: 214	NI Tanteada: 2
Características	1979-1980	Planta ocupada: 214	NI Tanteada: 2
Año construcción	1980	Distancia SUE a: ...	Distancia SUE a: ...
Comunidad	Autónoma municipal	Distancia SUE a: ...	Distancia SUE a: ...
Estado de conservación	<input type="checkbox"/> Buena <input checked="" type="checkbox"/> Mala <input type="checkbox"/> Mala <input type="checkbox"/> Mala	Capacidad máxima: ...	Capacidad máxima: ...
Datos del edificio			
Planta tipo: Doble caja de fachada con librería de aire			
Nomenclatura: Características de alumbrado: cerámico, energía eléctrica. Permisos en copropiedad.			
Sistema de ventilación: SI Natural <input type="checkbox"/> Preinstalación natural <input type="checkbox"/> Mecánica <input type="checkbox"/> Mixta <input type="checkbox"/> Aliviado			
Sistema de ventilación general: Admisión de aire por infiltraciones y extracción por abanico en cocina y baño			
Sistema de ventilación complementario: Ventilación y asientos en cocina y baño, y ventilación en dormitorio y despacho			
Sistema adicional: Sistema mecánico al exterior <input type="checkbox"/> Empresa de mantenimiento <input type="checkbox"/> No aplica			
Otras instalaciones: Calentamiento: mediana potencia centralizada de gas Producción de agua caliente sanitaria Caldera individual de gas estanca Piscina de recreo Eléctrica			
Nota: *Especificación según Capítulo 4, apartado 4.4.1			
Legenda: Abierto Cerrado En uso En desuso En construcción En construcción 1 En construcción 2 En construcción			

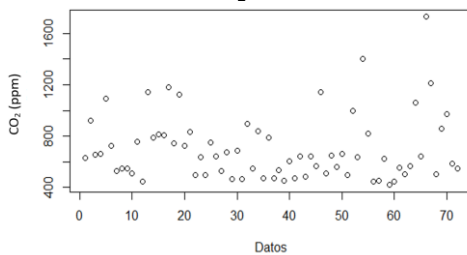


8

Prospective study



Objective variable: CO₂ average



Possible predictor variables:

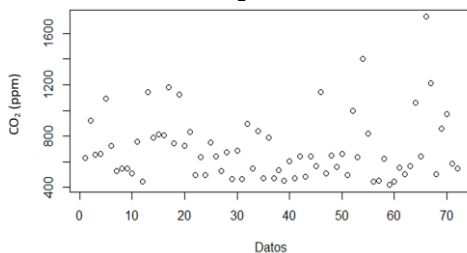
Construction year	Number of dry rooms
Occupied floor level	Facade orientation
Year of whole renovation	Occupancy habits / patterns
Total number of building storeys	Building use
Dwelling compactness	Number of occupants
Building / Construction typology	Ventilation system
Urban environment compactness / Urban density	Occupants' age
Opaque building envelope / Opaque elements	Domestic Hot Water (DHW) system
Location	Airing habits / Window opening habits
Glazed openings / Fenestration	Cooking system
Monitoring season	Occupants' subjective IAQ assessment
Envelope permeability / Airtightness	General ventilation system
Monitoring month	Indoor laundry drying
Number of ventilation shafts (shunts)	Complementary ventilation system
Room use / Room function	Indoor smoking / smokers
Net floor area / Usable floor area	Additional kitchen ventilation system
Number of wet rooms	Occupancy density
Number of facades	

9

Prospective study



Objective variable: CO₂ average



Possible predictor variables:

Construction year	Number of dry rooms
Occupied floor level	Facade orientation
Year of whole renovation	Occupancy habits / patterns
Total number of building storeys	Building use
Dwelling compactness	Number of occupants
Building / Construction typology	Ventilation system
Urban environment compactness / Urban density	Occupants' age
Opaque building envelope / Opaque elements	Domestic Hot Water (DHW) system
Location	Airing habits / Window opening habits
Glazed openings / Fenestration	Cooking system
Monitoring season	Occupants' subjective IAQ assessment
Envelope permeability / Airtightness	General ventilation system
Monitoring month	Indoor laundry drying
Number of ventilation shafts (shunts)	Complementary ventilation system
Room use / Room function	Indoor smoking / smokers
Net floor area / Usable floor area	Additional kitchen ventilation system
Number of wet rooms	Occupancy density
Number of facades	

Looking for the smart predictions:

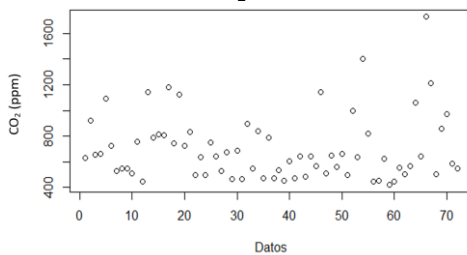
- Ease of data collection
- Experimental relevance
- Inter-variable correlation
- % of variance explained

10

Prospective study



Objective variable: CO₂ average



Possible predictor variables:

Construction year	Number of dry rooms
Occupied floor level	Facade orientation
Year of whole renovation	Occupancy habits / patterns
Total number of building storeys	Building use
Dwelling compactness	Number of facades
Building / Construction type	General ventilation system
Urban environment compact density	Indoor laundry drying
Opaque building envelope / elements	Complementary ventilation system
Location	Indoor smoking / smokers
Glazed openings / Fenestration	Additional kitchen ventilation system
Monitoring season	Occupancy density
Envelope permeability / Airtightness	
Monitoring month	
Number of ventilation shafts (shunts)	
Room use / Room function	
Net floor area / Usable floor area	
Number of wet rooms	
Number of facades	

Looking for the smart predictors:

- Ease of data collection
- Experimental relevance
- Inter-variable correlation
- % of variance explained

11

Prospective study



$$[CO_2] = \frac{1}{2.430 \cdot 10^{-3} + R + S - 4.90 \cdot 10^{-4} F - 2.14 \cdot 10^{-4} O + 8 \cdot 10^{-6} A}$$

Where:

[CO₂]– Average to be expected [ppm].

R – Type of room (Living/dining room: $-1.12 \cdot 10^{-4}$; Master bedroom: $-1.86 \cdot 10^{-4}$; 2nd bedroom: $-2.94 \cdot 10^{-4}$; 3rd Bedroom: $-1.39 \cdot 10^{-5}$; Kitchen: $2.63 \cdot 10^{-5}$; Bathroom/toilet: $1.48 \cdot 10^{-4}$; Other room: 0).

S – Season (Summer: $2.68 \cdot 10^{-4}$; Winter: 0).

F – Number of façades of the dwelling: 1 to 3.

O – Number of usual occupants of the dwelling: 1 to 4.

A – Total living area of the dwelling: 29 to 150 [m²].

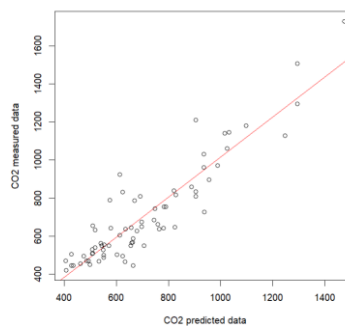
12

How good this result is??

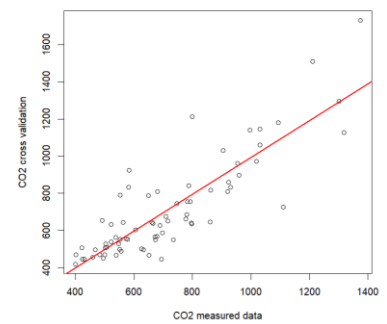


	% var
Model	80.7%
Season	29.8%
Living area	23.5%
Number of occupants	17.5%
Number of façades	15.8%
Room type	15.4%

Measured vs. predicted
McFadden Pseudo-R² = 0.79



Cross-validation
McFadden Pseudo-R² = 0.73



13

Conclusions

- Predicting IAQ using simple parameters and statistical data processing is a **promising approach to enable large-scale assessments**.
- This methodology could support **decision-making in large-scale EU renovation strategies** within the Renovation Wave.
- It does **not replace measurement**—it **extends it** to cases where measurement is not feasible.
- It does **not replace simulation**; rather, it can **support it** by identifying new relevant variables or factors.
- The foundation is **large amounts of labelled data for training**, which is costly and not always feasible. **Initiatives like Pandora** will help make this possible.
- **If the training data is of low quality and contain bias**, the model will amplify it.

14