



From Physics to Physical AI: The Evolution of Urban Digital Twins

Dr. Mahmoud Abdelrahman
Postdoc Researcher, **KAIST Urban AI (KUAI)**

mahmoud@kaist.ac.kr



Who Am I ?



Mahmoud OR Ouf OR Ali

Ma mood . 마무드

From **Egypt**



Mother Language: **Arabic**

<https://mahmoud.work>



BSc. 2009

Mansoura University - Architectural Engineering

Teaching staff @ Kafrelsheikh University

CEO @ Ain Design Studio



MSc. 2017

Mansoura University - Architectural Engineering



PhD. 2022

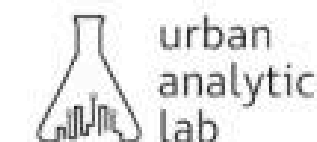
National University of Singapore (NUS) -
Built Environment

CEO @ 4Dt.io

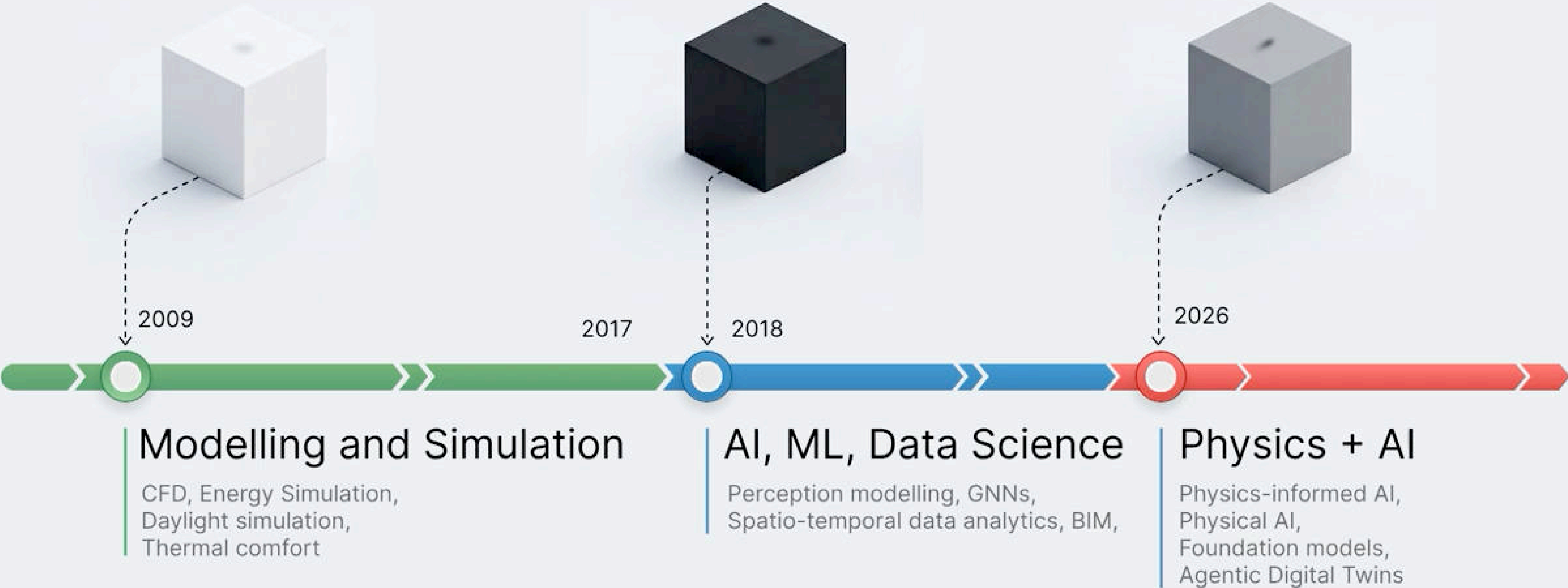
Postdoc Research Fellow @ UAL



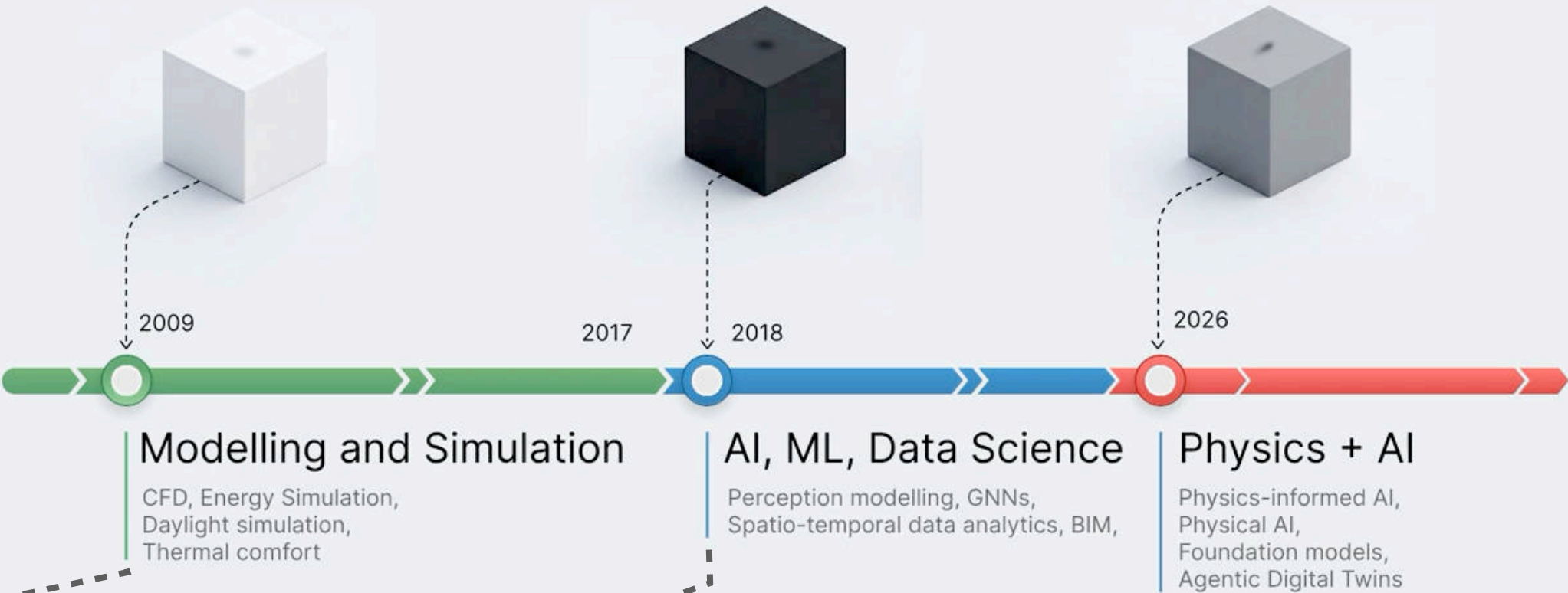
Postdoc Researcher @ KAIST



Research Trajectory



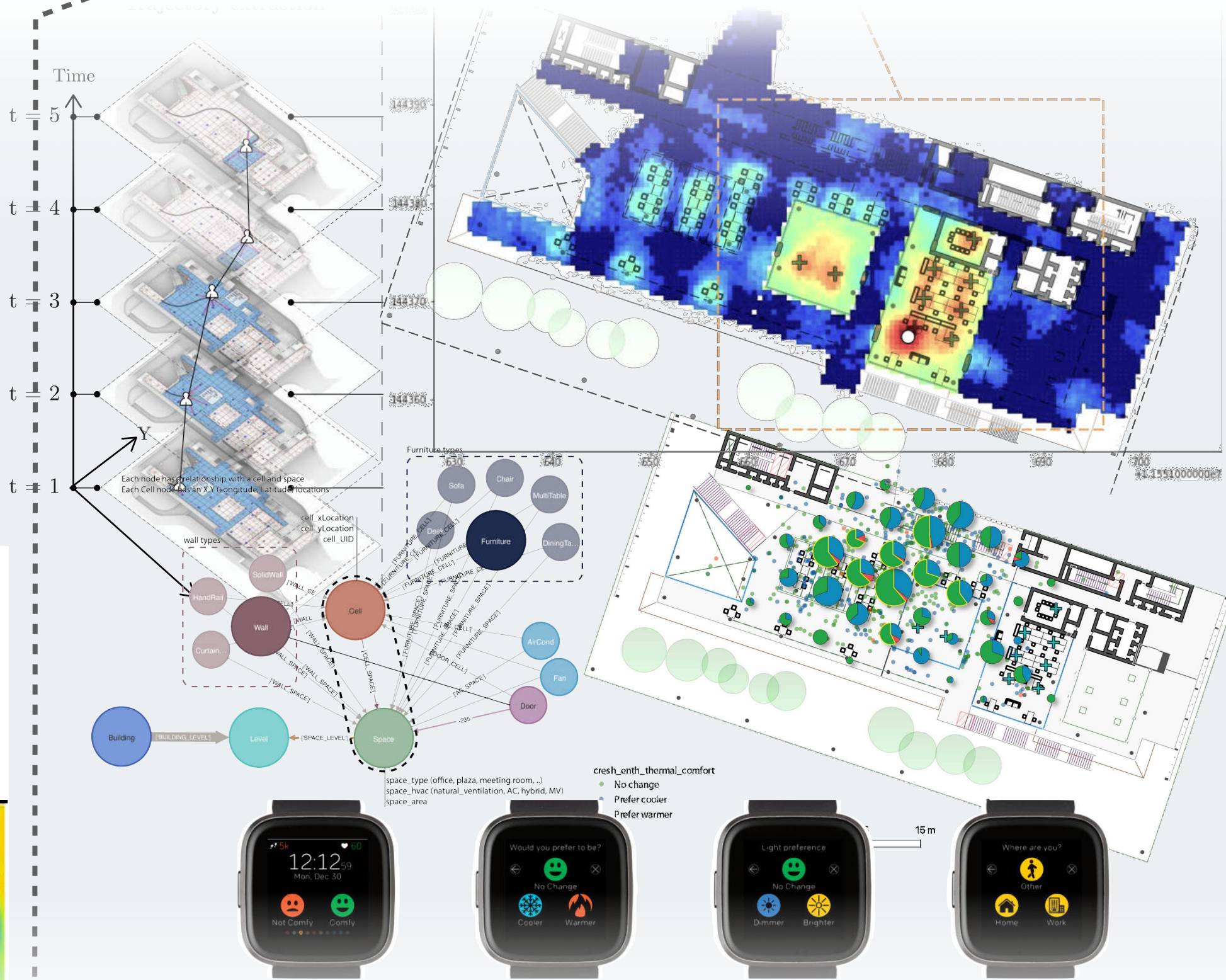
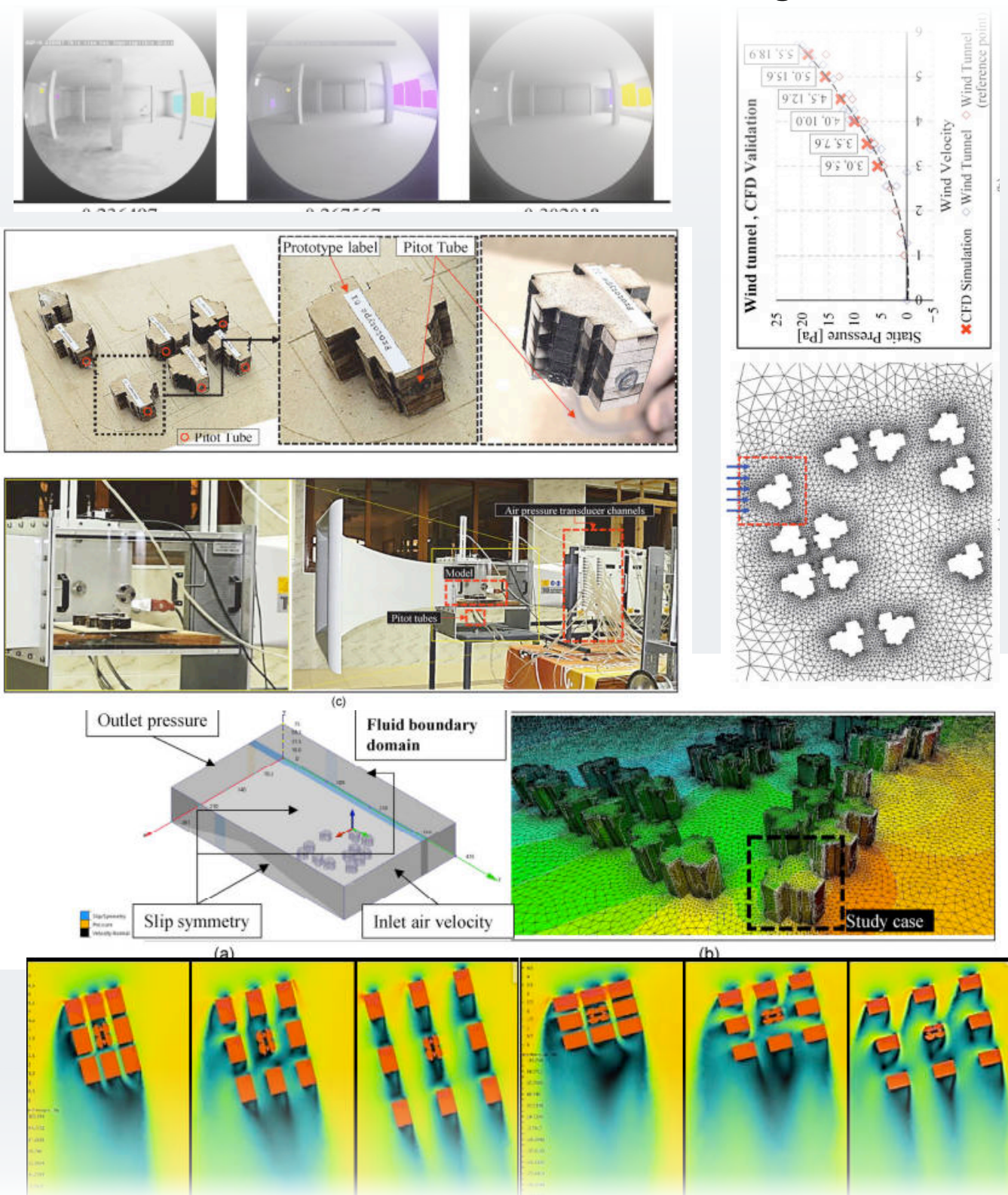
Research Trajectory



Physics

AI

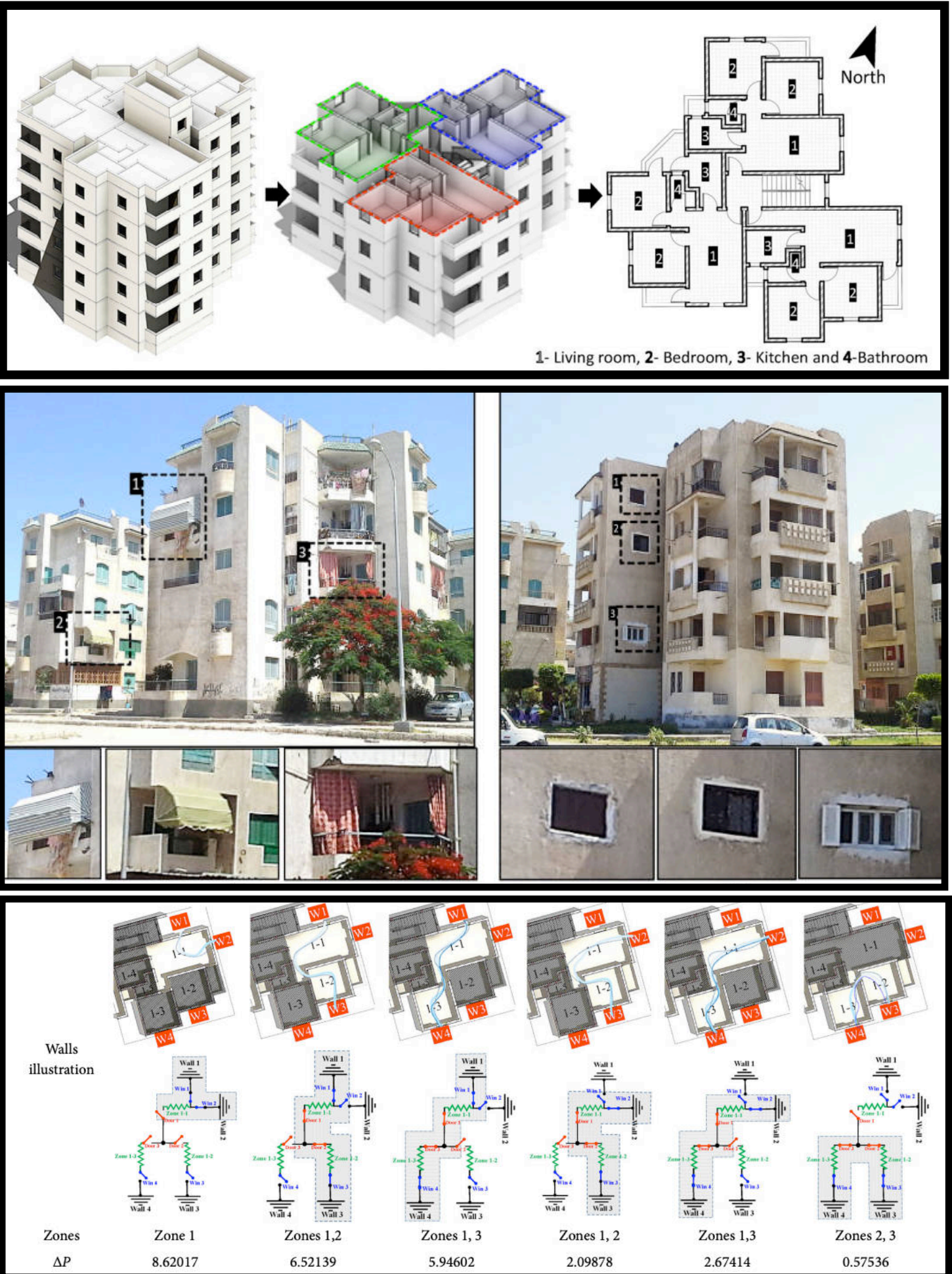
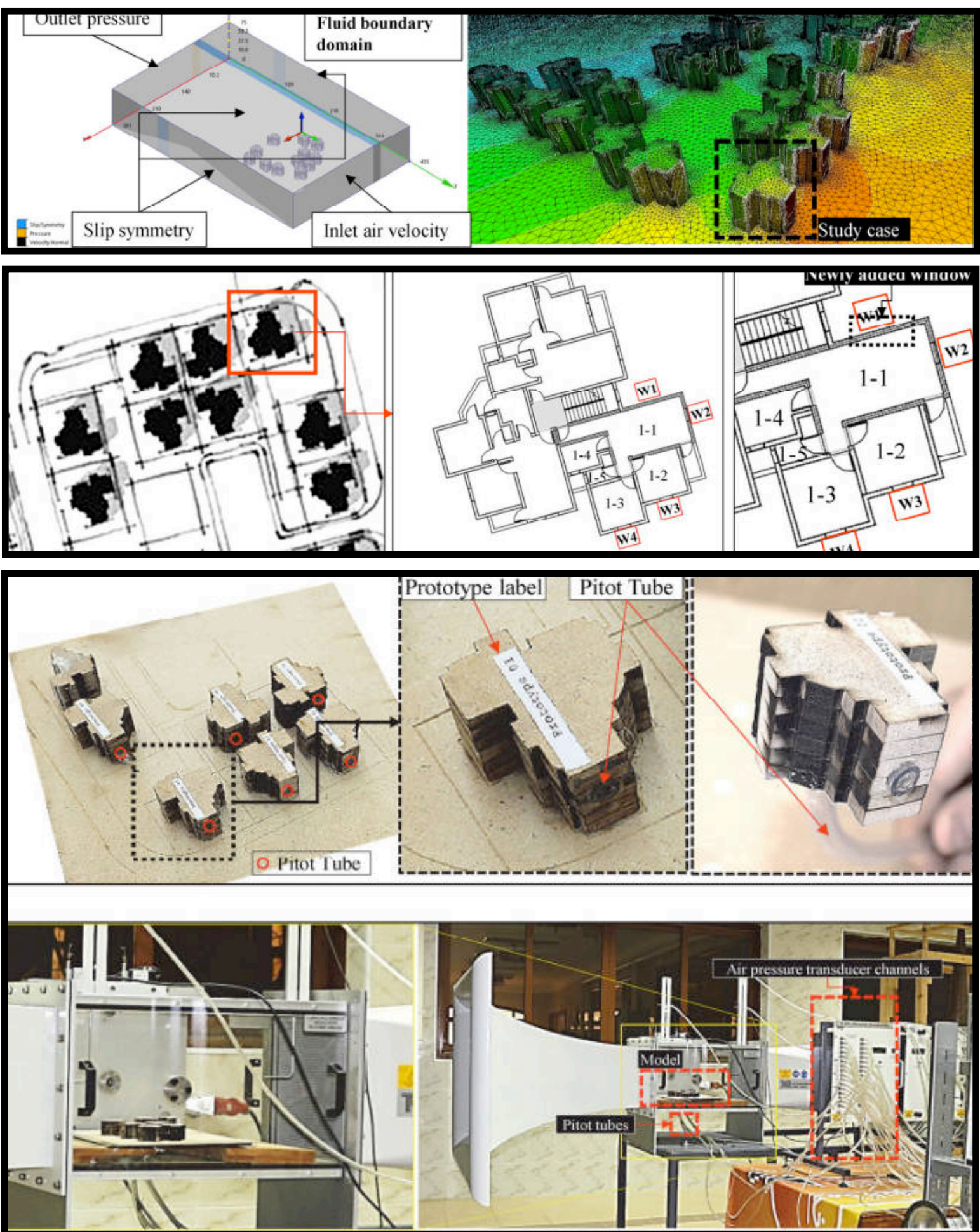
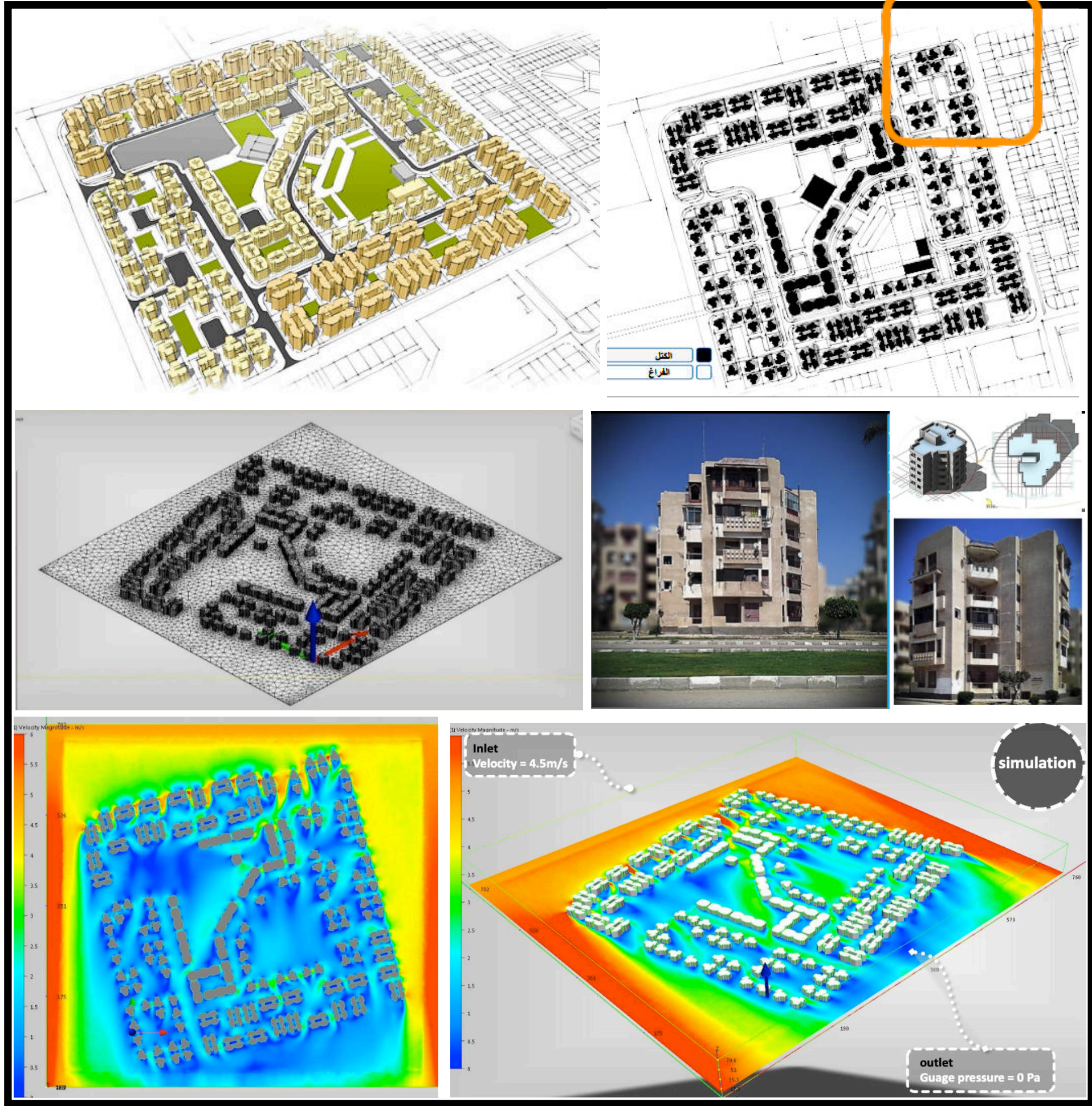
Physics + AI



White box : Physics

Multi-scale human comfort in the built environment

De-coupled simulation: outdoor CFD feeds the indoor models (thermal comfort, daylight, energy)



Neighbourhood Scale

Building Cluster Scale

Building Scale

Physical models are useful, but ...

"All models are wrong, but some are useful." —

George E.P. Box (Box, 1976, 1979)

George E.P. Box (Box, 1976, 1979)

Physical models are useful, but ...

"All models are wrong, but some are useful." —
George E.P. Box (Box, 1976, 1979)

- 1- The **code** must be verified,
 - 2- The **results** must be validated, and
 - 3- The model must be **calibrated**.
- ...and people don't follow the model**

The model is not generalisable
Each new case needs its own model.

The most widely used thermal comfort model, PMV (Predicted Mean Vote), predicts occupants' sensation correctly only about 33% of the time
(Cheung et al., 2019)

Actual occupant behaviour (like opening windows or drawing blinds) creates a significant variation range in energy consumption spanning between **2% and 20%**
(Moustafa et al., 2018)

Physical models are useful, but ...

"All models are wrong, but some are useful." —
George E.P. Box (Box, 1976, 1979)

- 1- The **code** must be verified,
- 2- The **results** must be validated, and
- 3- The model must be **calibrated**.

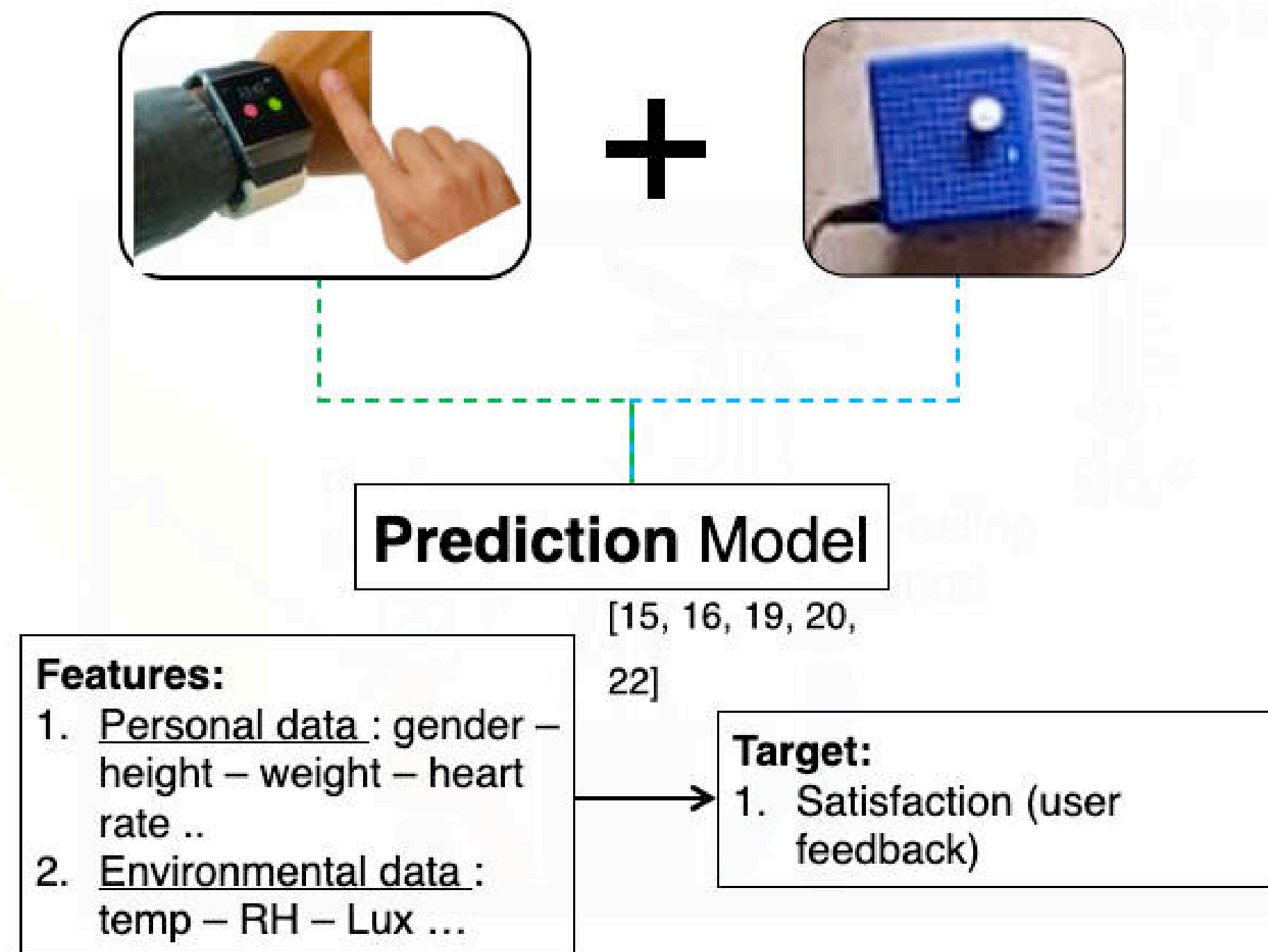
...and people don't follow the model



**How can we model human
behaviour ?**

Don't simulate comfort. Ask.

Instead of estimating how people feel, we asked them, all day, on their wrists



50 participants

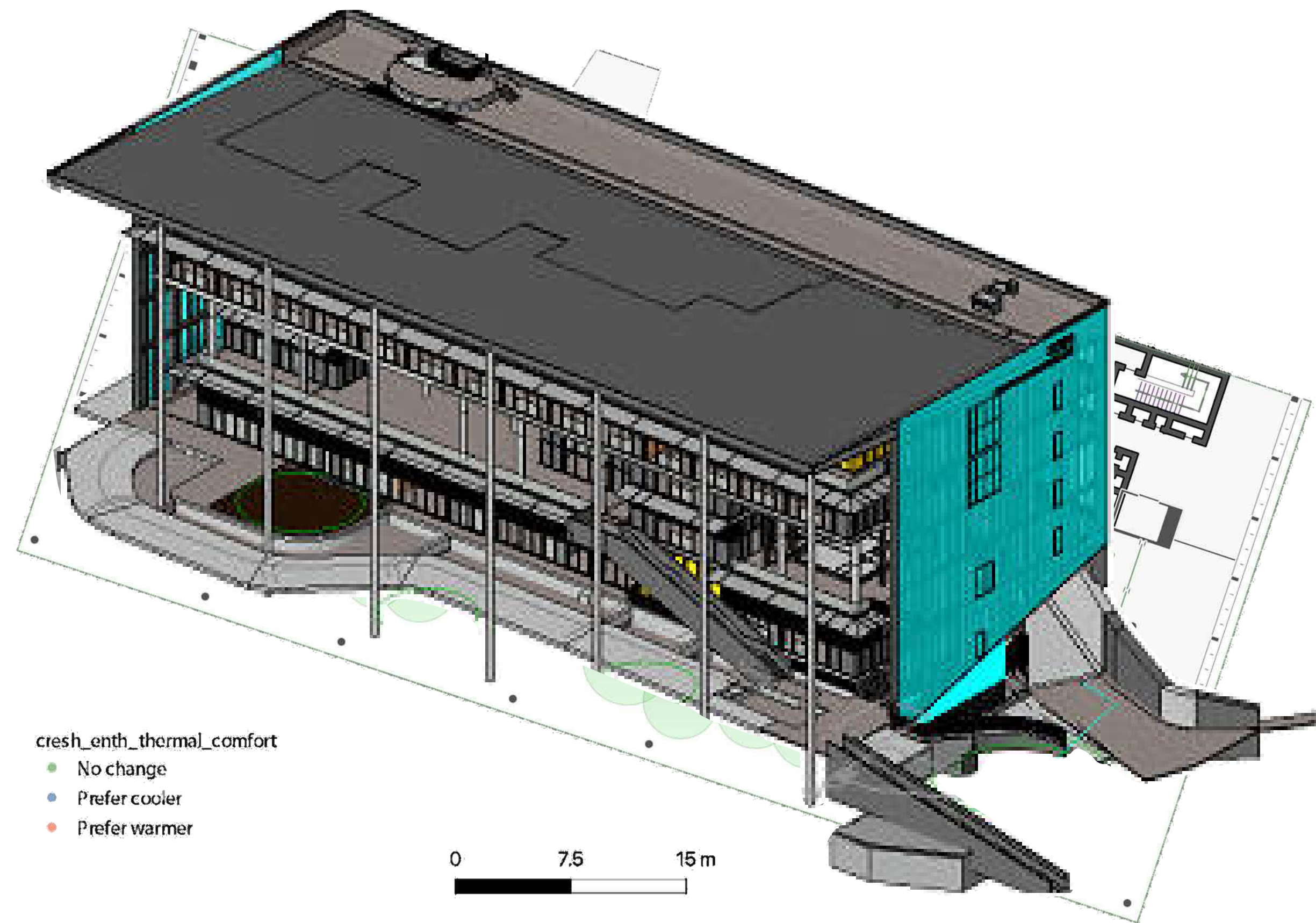


Ref: Jayathissa, P.; Quintana, M.; **Abdelrahman, M.**; Miller, C. Humans-as-a-Sensor for Buildings—Intensive Longitudinal Indoor Comfort Models. *Buildings* **2020**, *10*, 174. <https://doi.org/10.3390/buildings10100174>

Sood, T., Quintana, M., Jayathissa, P., **Abdelrahman, M.**, & Miller, C. (2019). **The SDE4 learning trail: Crowdsourcing occupant comfort feedback at a net-zero energy building.** *Journal of Physics: Conference Series*, 1343(1), 012141. <https://doi.org/10.1088/1742-6596/1343/1/012141>

Don't simulate comfort. Ask.

Instead of estimating how people feel, we asked them, all day, on their wrists



50 participants



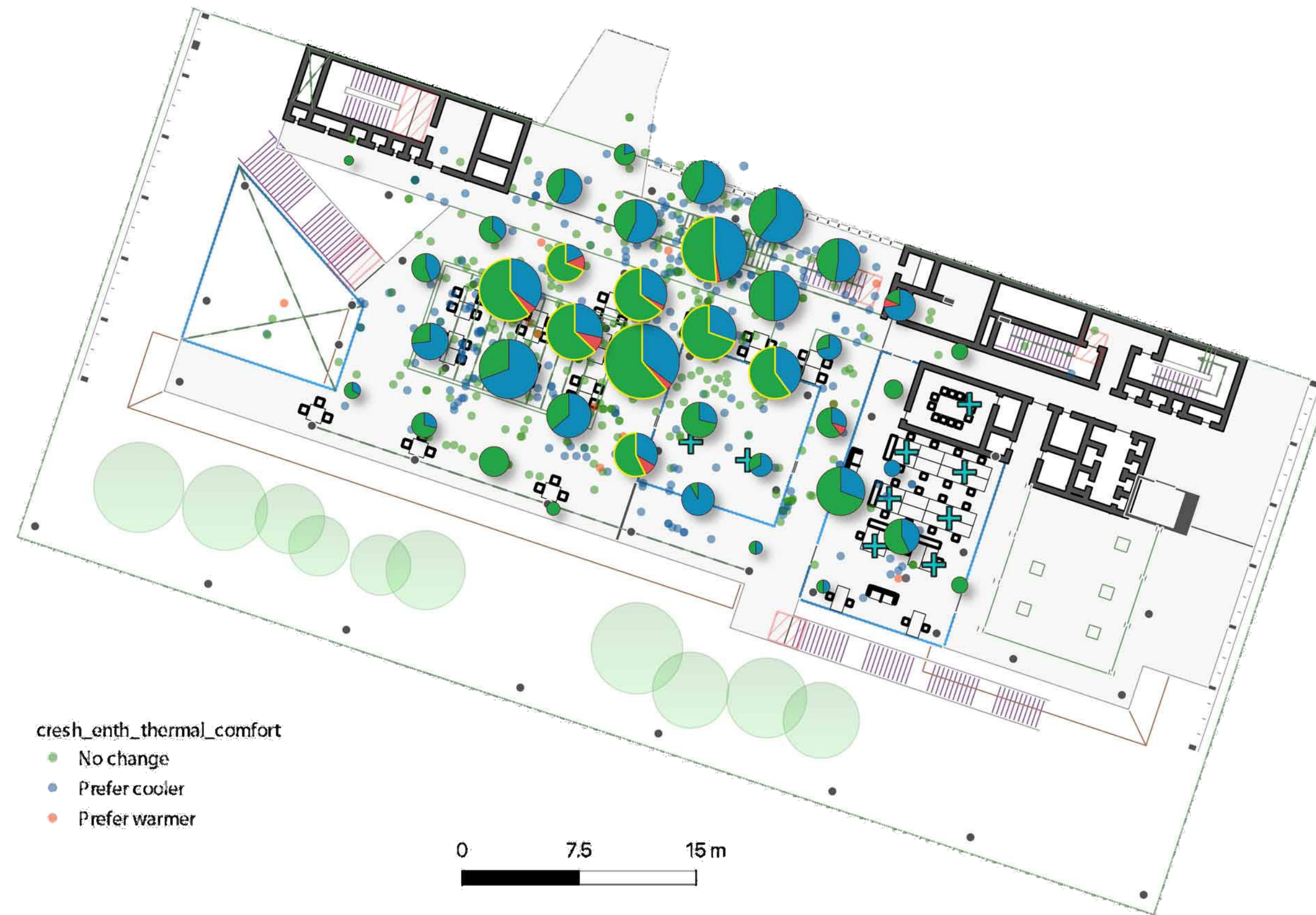
Ref: Jayathissa, P.; Quintana, M.; Abdelrahman, M.; Miller, C. Humans-as-a-Sensor for Buildings—Intensive Longitudinal Indoor Comfort Models. *Buildings* **2020**, *10*, 174. <https://doi.org/10.3390/buildings10100174>

Matias Quintana, Mahmoud Abdelrahman, Mario Frei, Federico Tartarini, and Clayton Miller. 2021. Longitudinal personal thermal comfort preference data in the wild. <https://doi.org/10.1145/3485730.3493693>

ASK THE OCCUPANTS

Don't simulate comfort. Ask people.

Instead of estimating how people feel, we asked them, all day, on their wrists



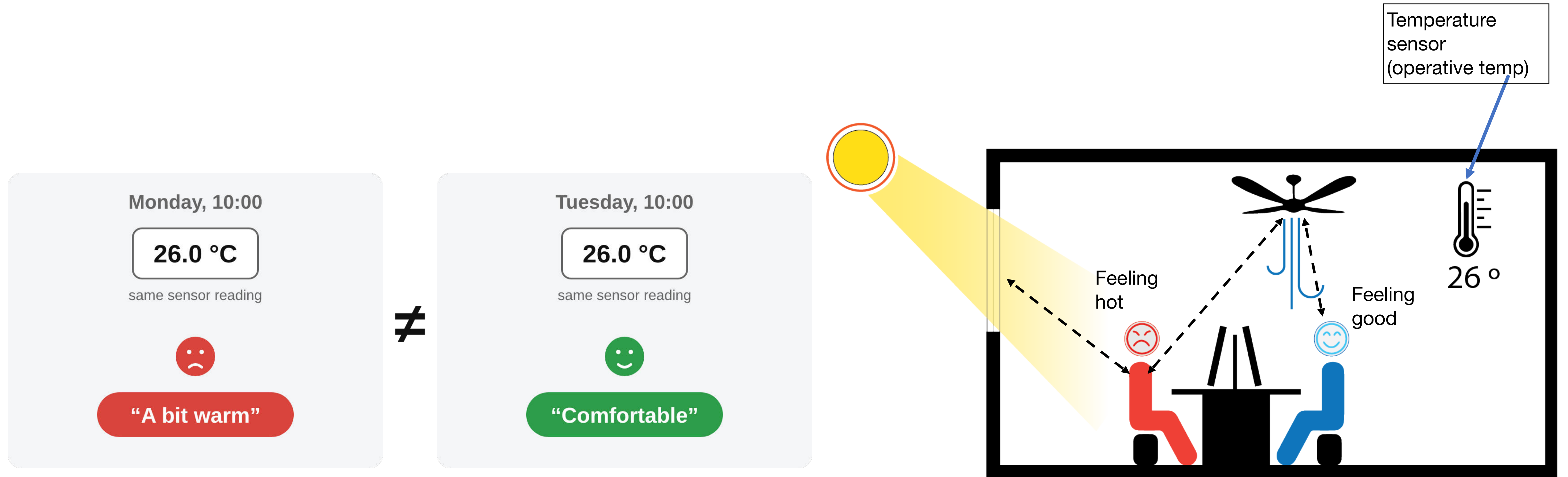
50 participants



THE PUZZLE

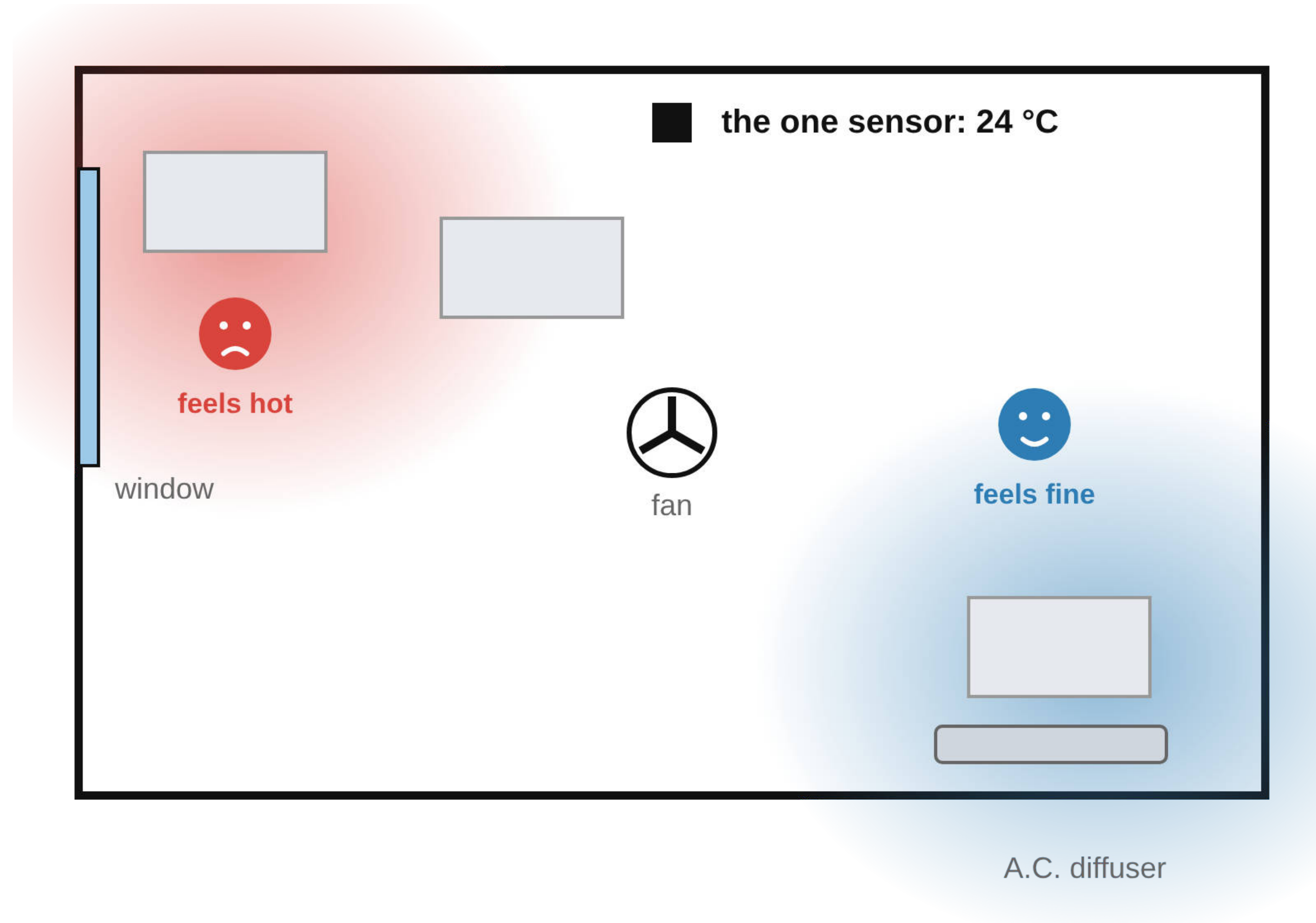
Same person, same conditions, different answer

The sensor reading did not change. What did?



Rooms are not uniform

One sensor, one number, but a window, a diffuser and a fan create different micro-climates

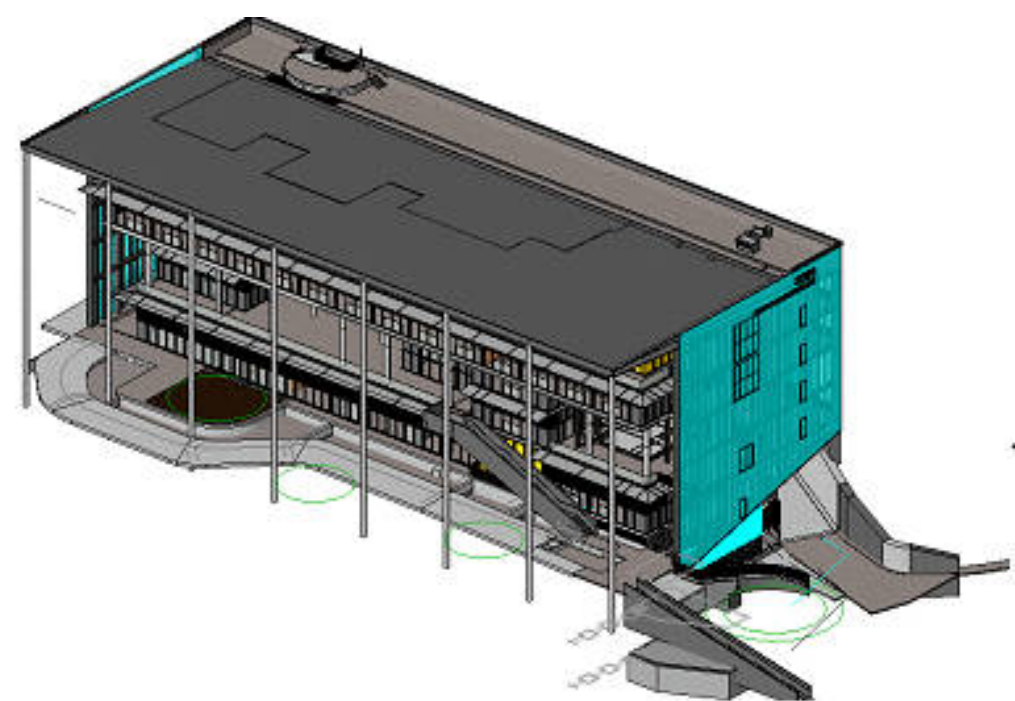


So we added space:

- Where is the person? (indoor positioning)
- What is around them? (window, fan, desk)
- How far? (proximity from the building model)

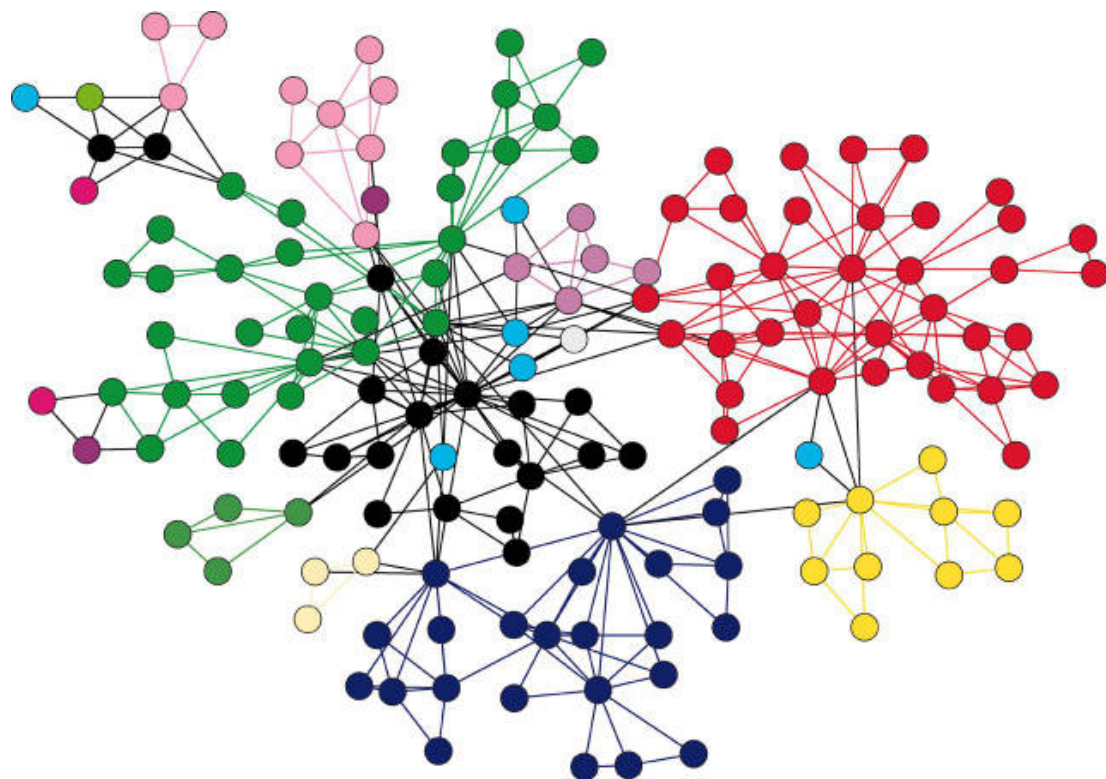
The built environment is inherently a Graph

Building Information Model (Spatial data)



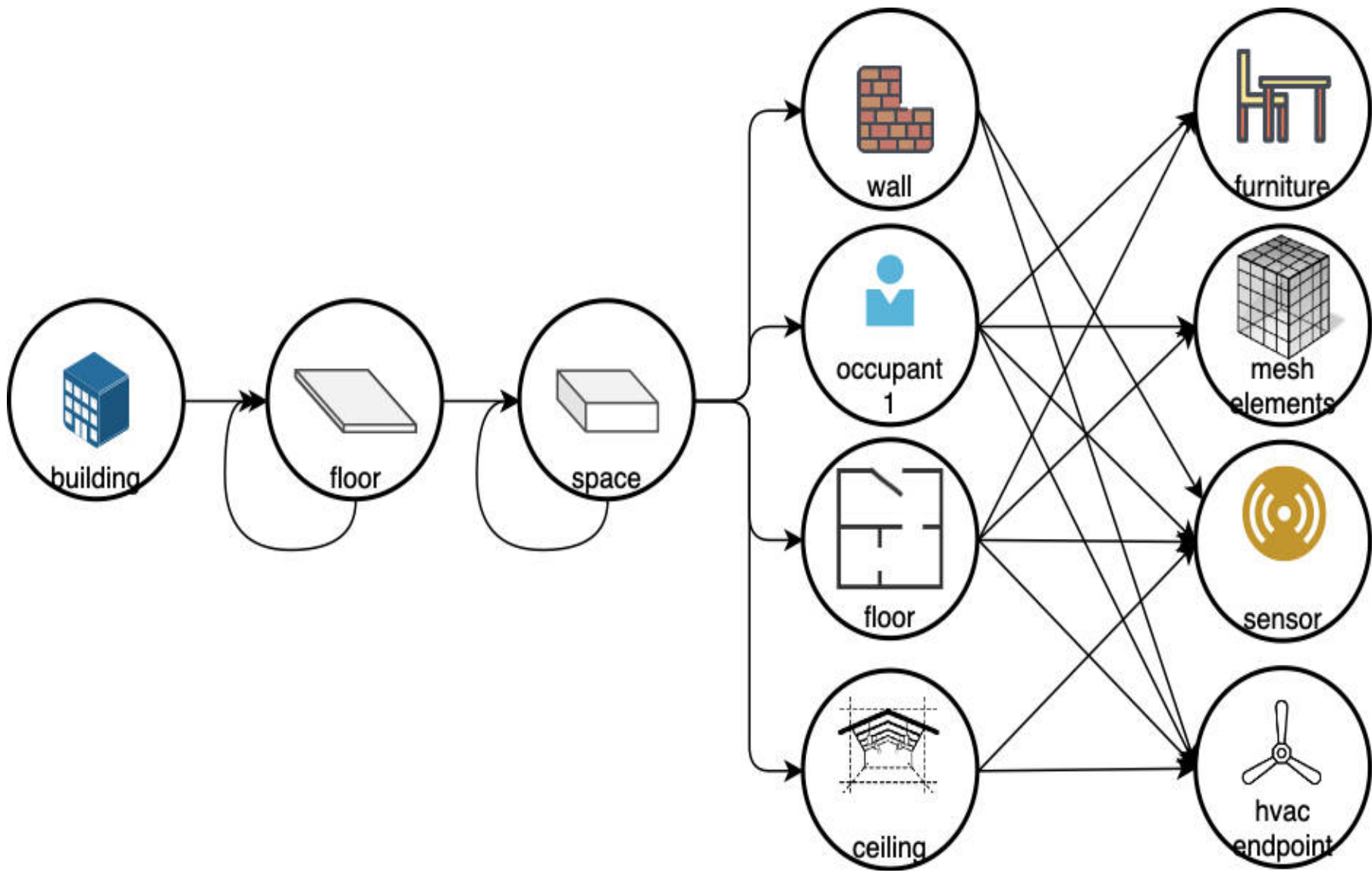
IFC

Graph (network)



Is a data structure
Used to represent real-world
In the form of **Nodes**, and **Links**

Heterogenous (attributed) graphs



A schematic representation
of a building graph

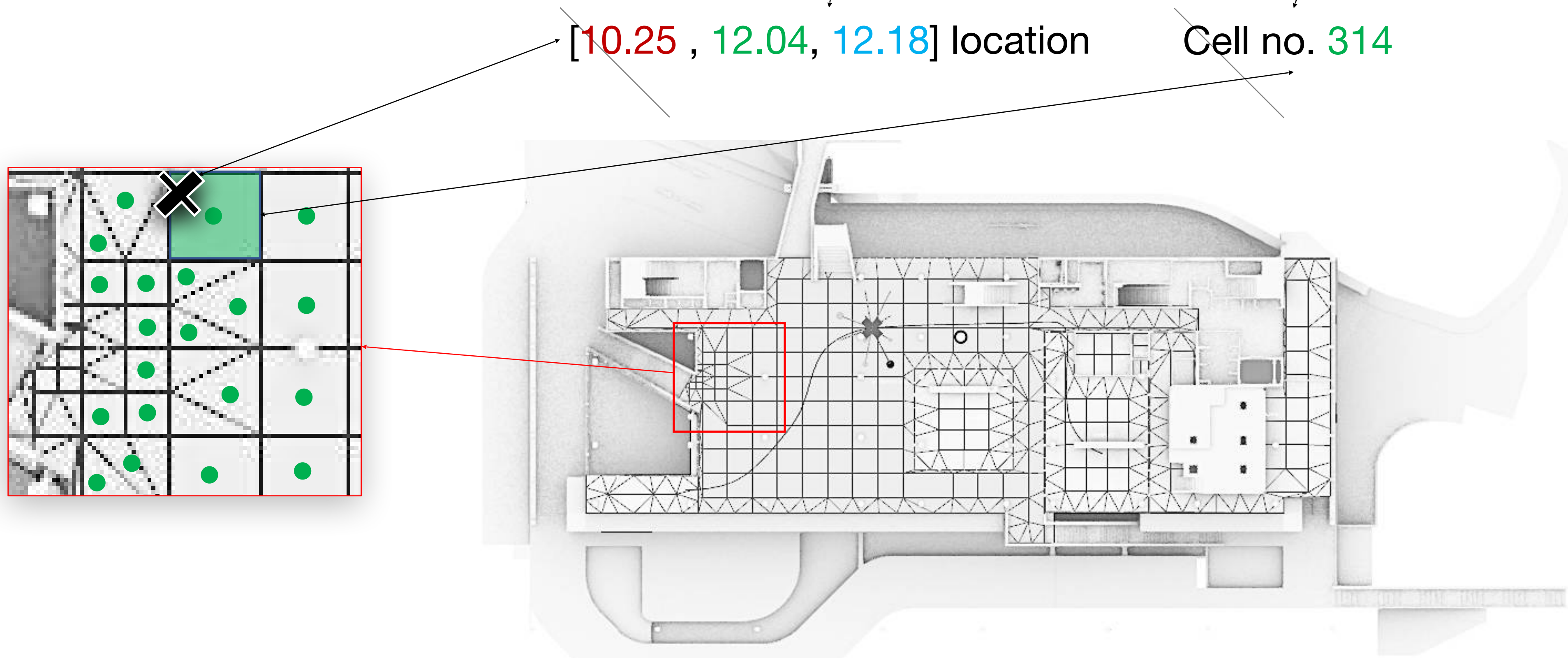
Ref: Abdelrahman, M., Chong, A., & Miller, C. (2020). Build2Vec: Building representation in vector space. *Proceedings of the Symposium on Simulation for Architecture and Urban Design (SimAUD)*. <https://doi.org/10.5555/3465085.3465155>

How do you teach an AI about building space?

Spatial discretization (Meshing)

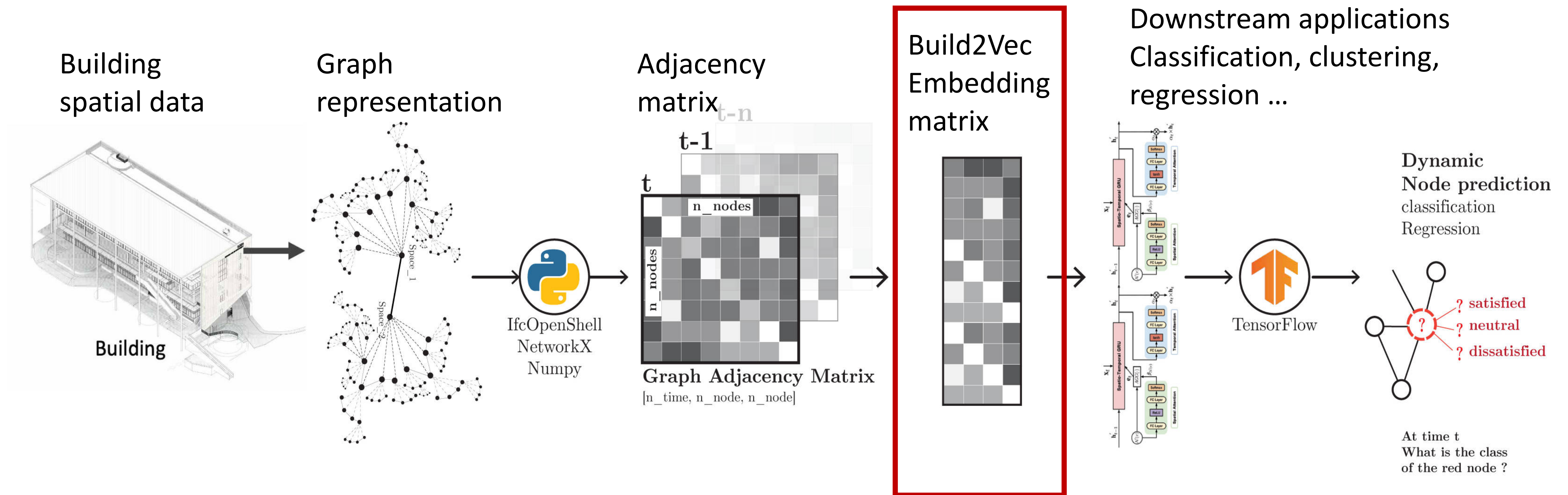
Also known as domain decomposition

Spatial discretization is used to convert **continuous** domains into **discrete domains**.

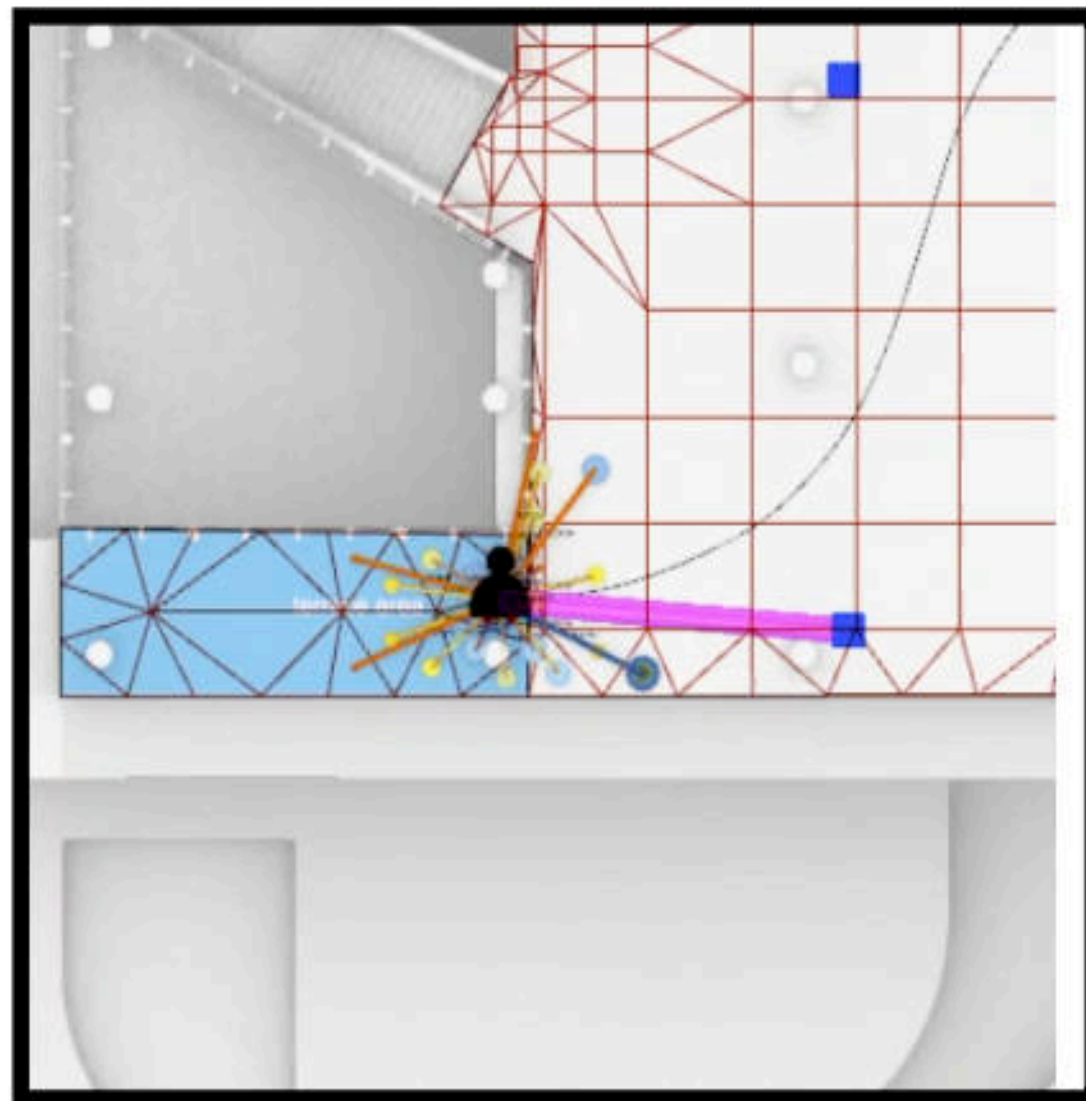
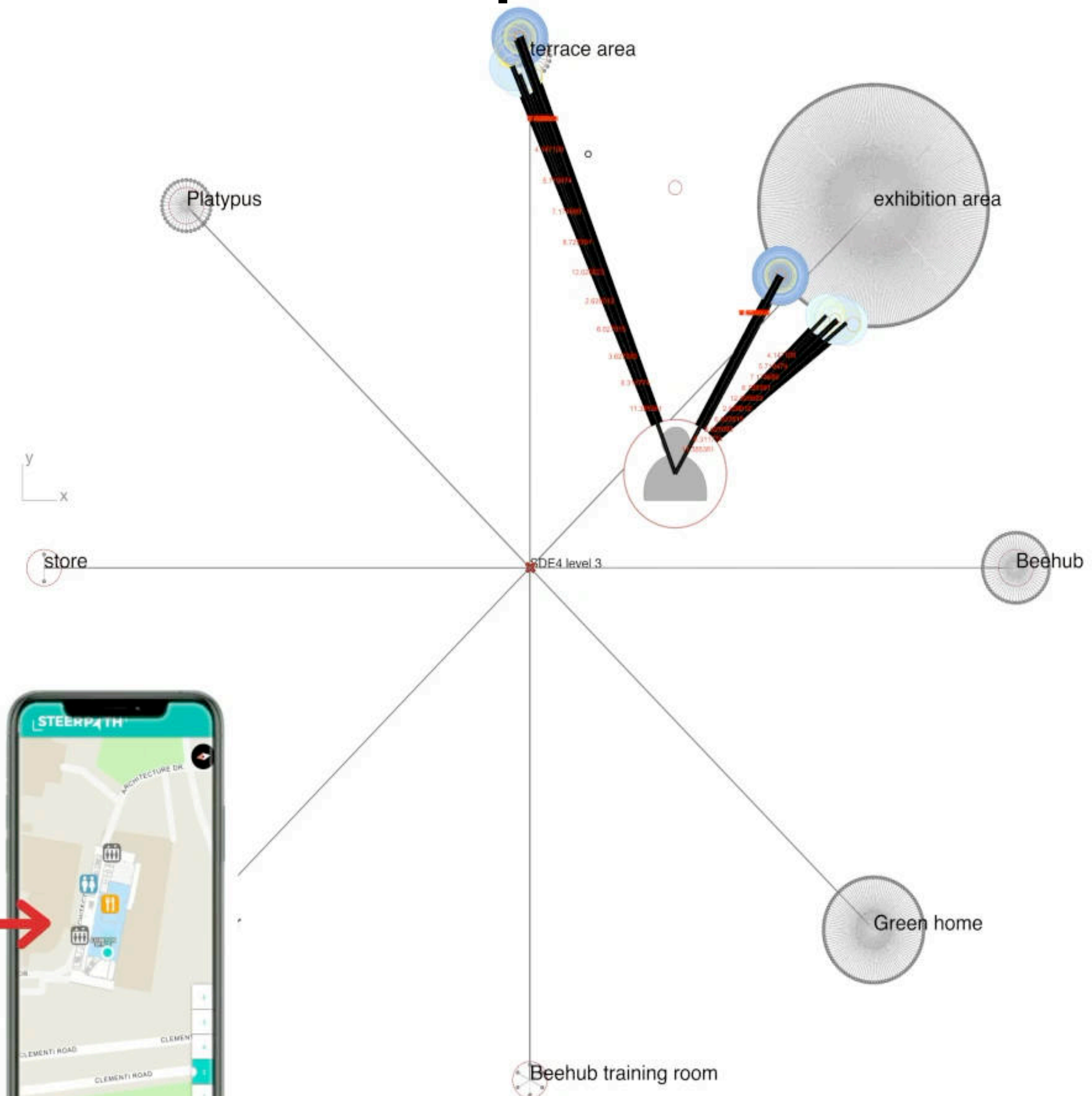
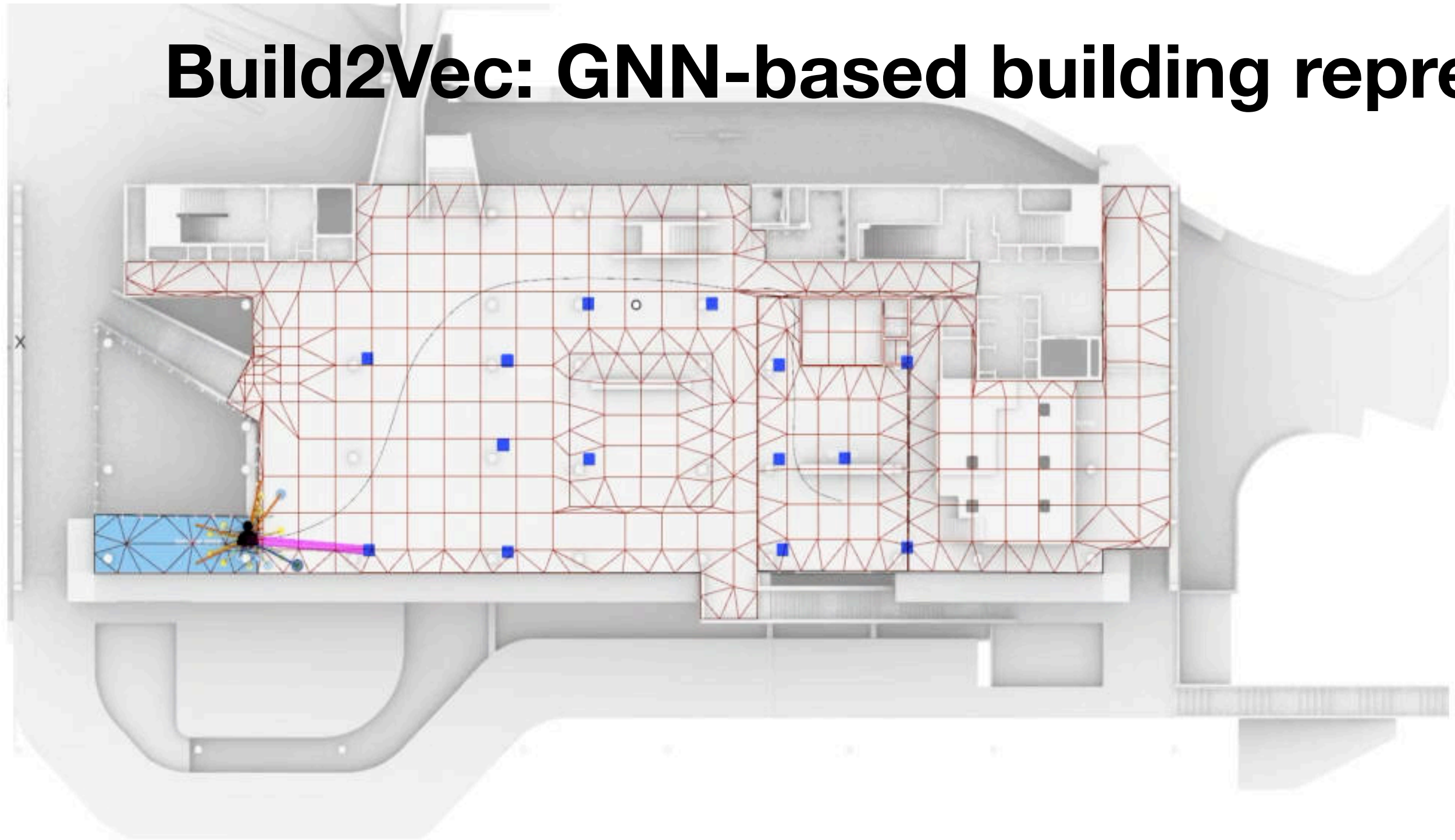


Build2Vec: GNN-based building representation in vector space

Build2Vec, is a Graph Neural Network model for building data embeddings. This model converts static and temporal graph representations of building spatial data into condensed vector space.



Build2Vec: GNN-based building representation in vector space



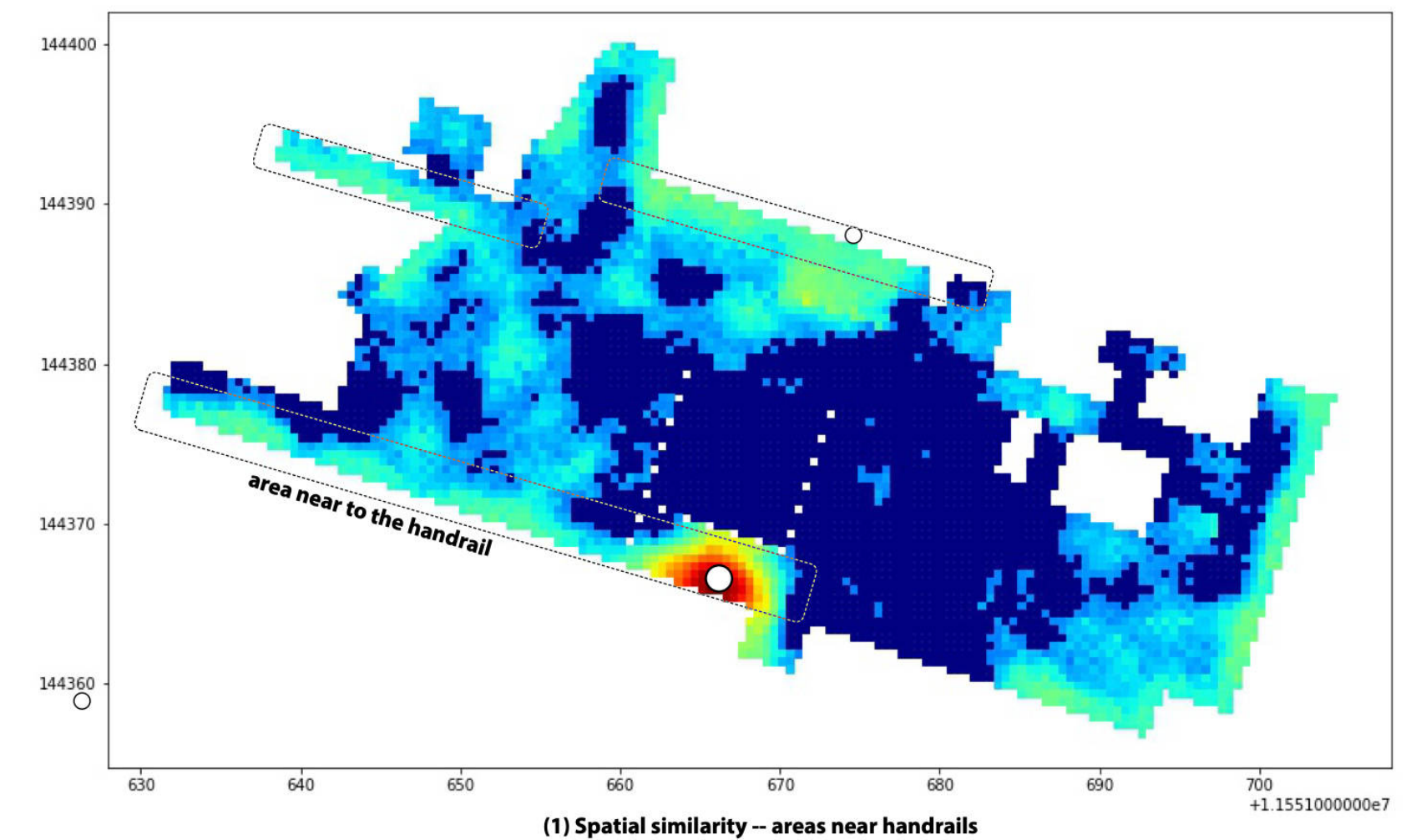
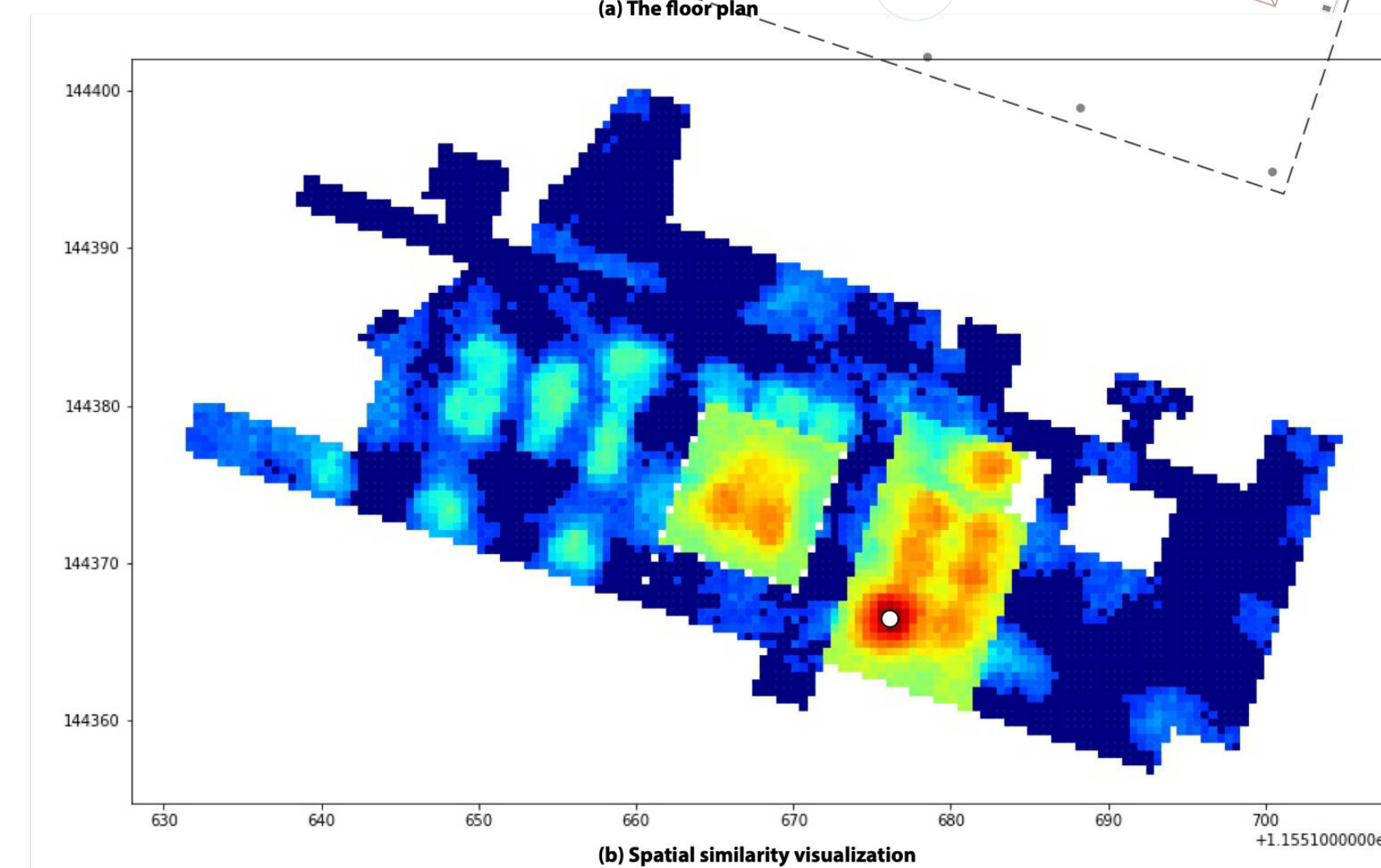
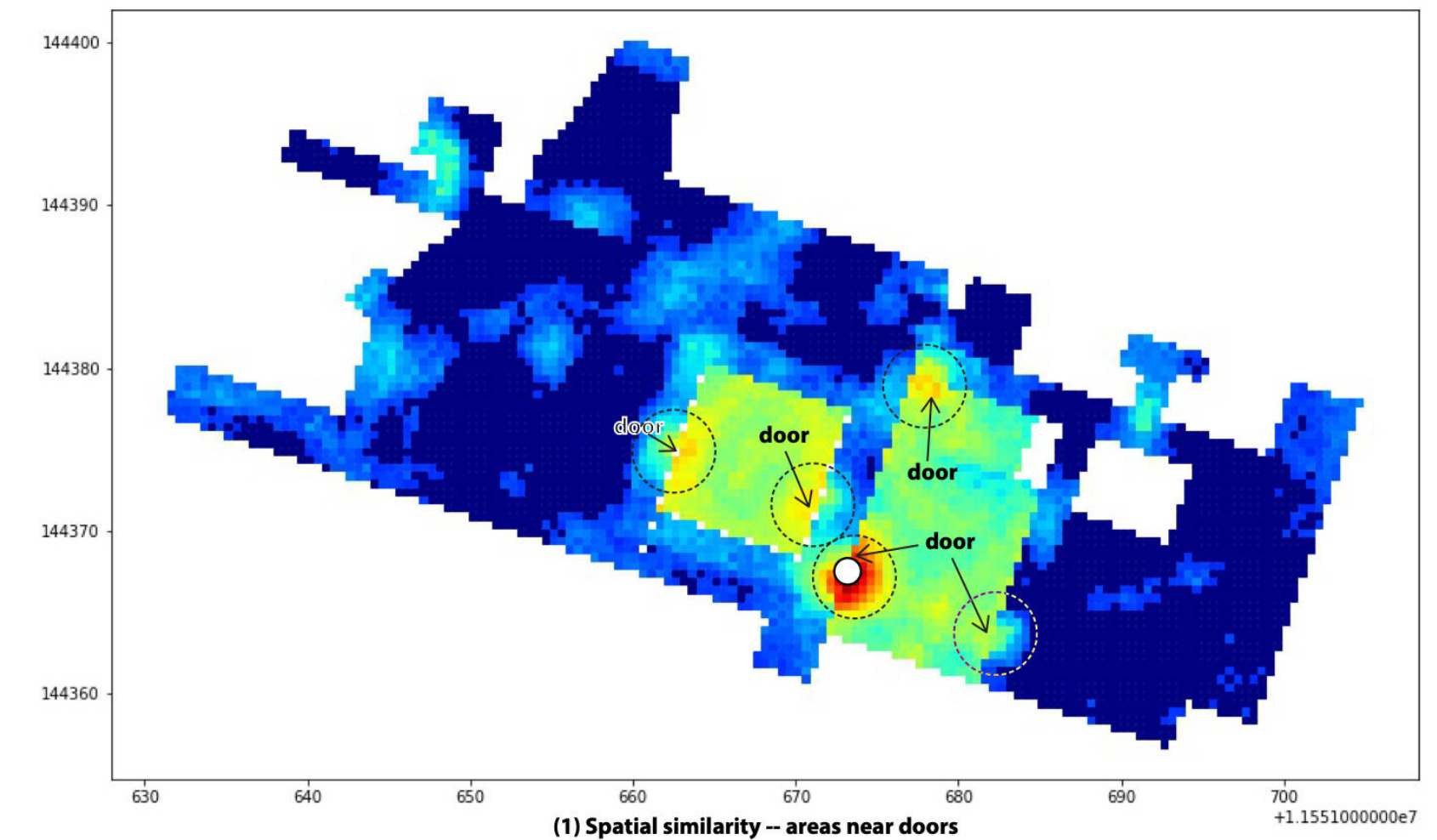
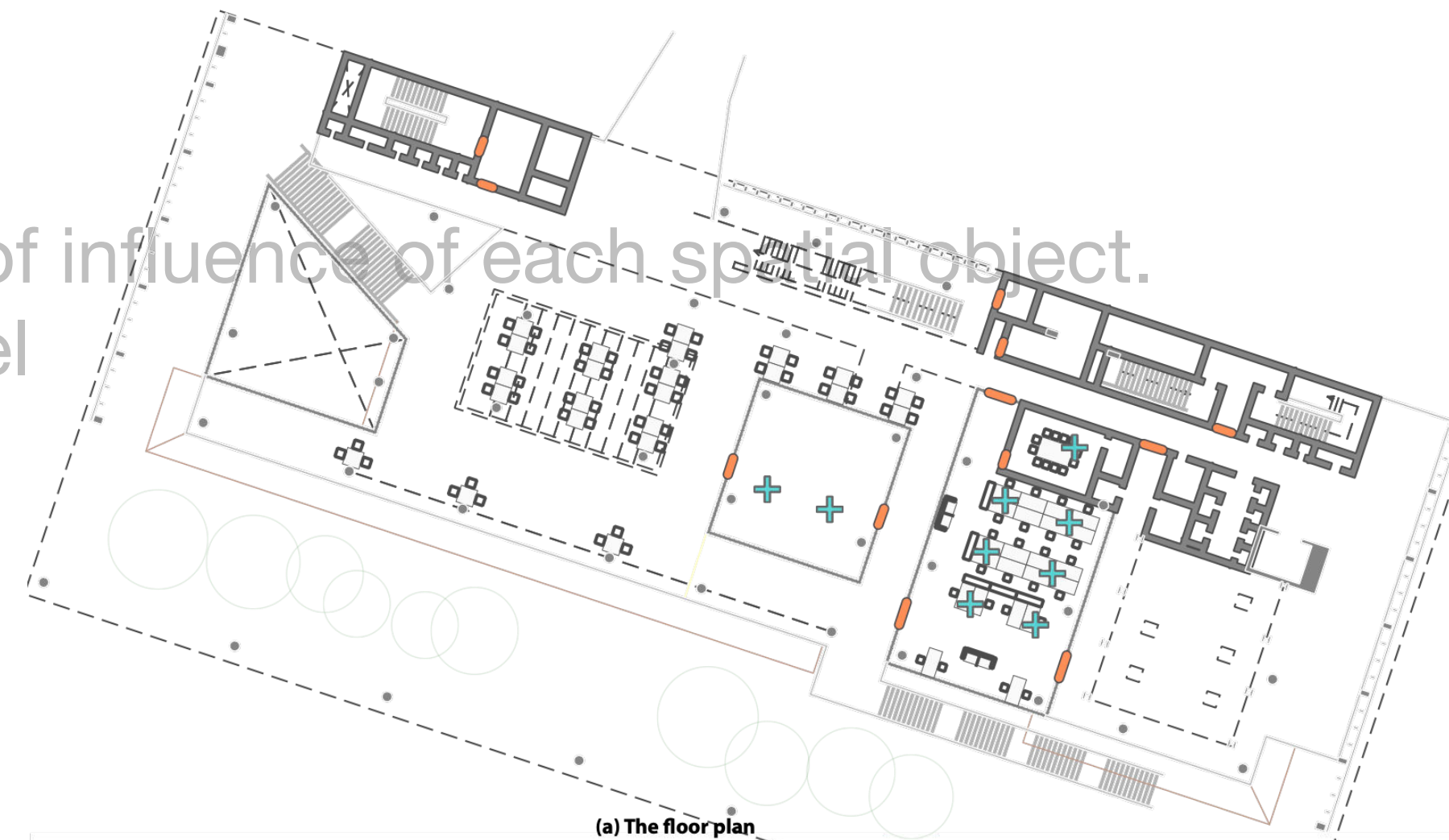
SDE4 - Level 3



Build2Vec: GNN-based building representation in vector space

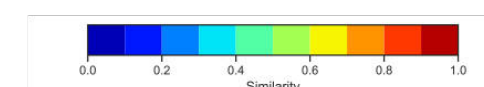
Implementation

1. Grid generation.
2. Defined the area of influence of each spatial object.
3. Building the model
4. Application

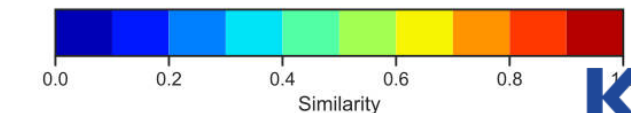


Ref : Miller, C., Abdelrahman, M., Chong, A., Biljecki, F., Quintana, M., Frei, M., Romanello, L., & Chew, M. Y. L. (2021). The Internet-of-Buildings (IoB): Digital twin convergence of wearable and IoT data with GIS/BIM. *Journal of Physics: Conference Series*, 2042(1), 012041. <https://doi.org/10.1088/1742-6596/2042/1/012041>

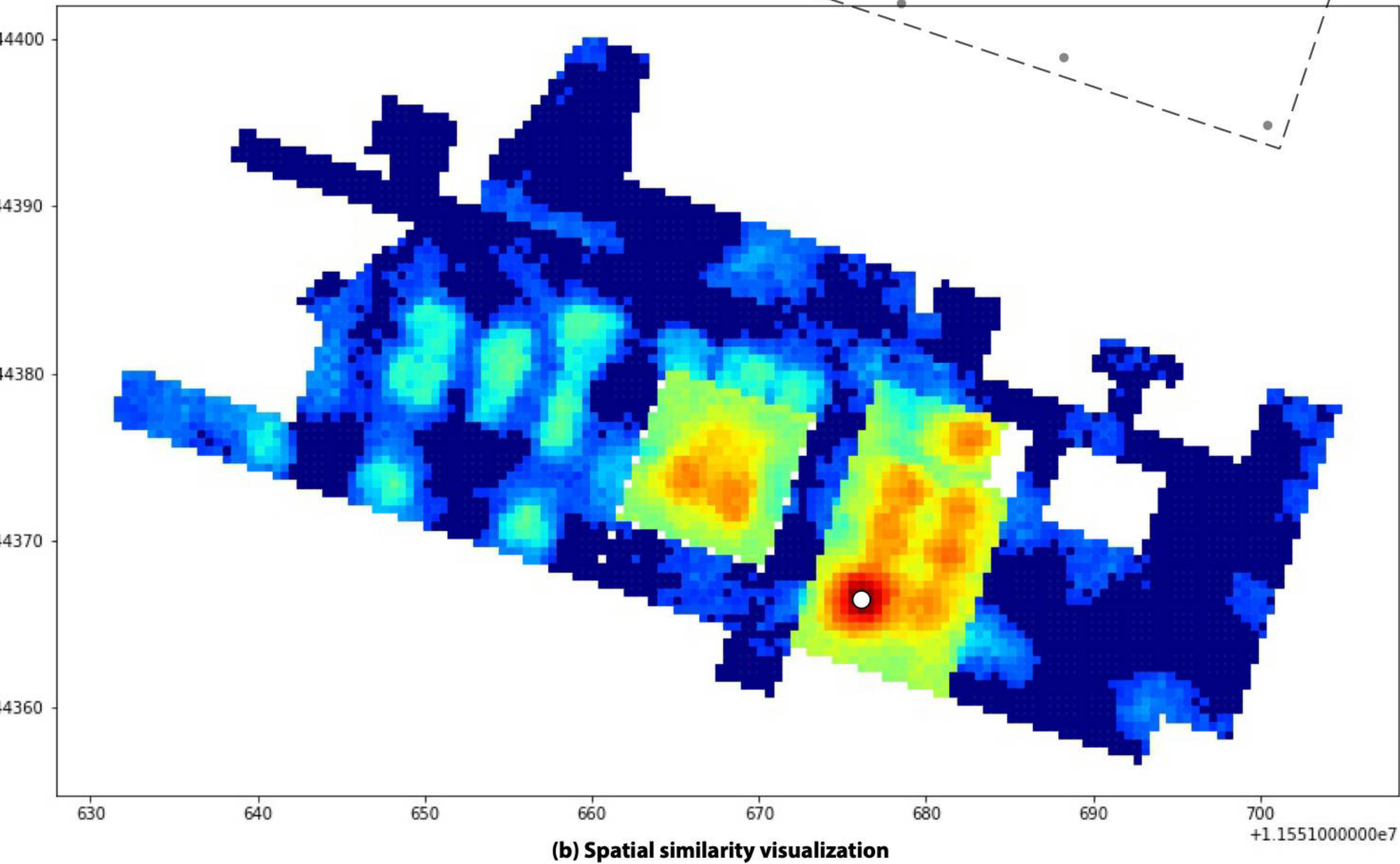
○ Test point



○ Test point

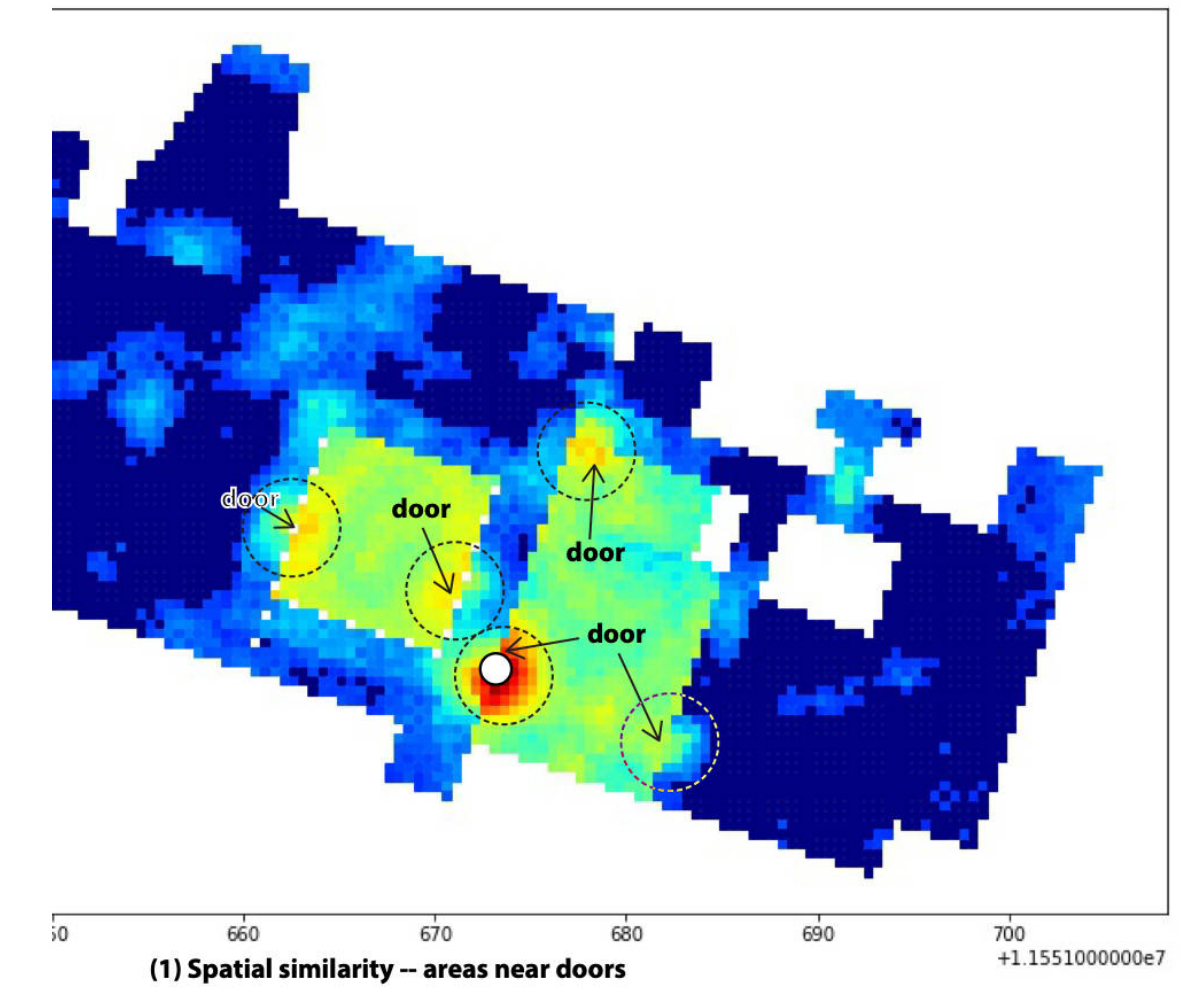
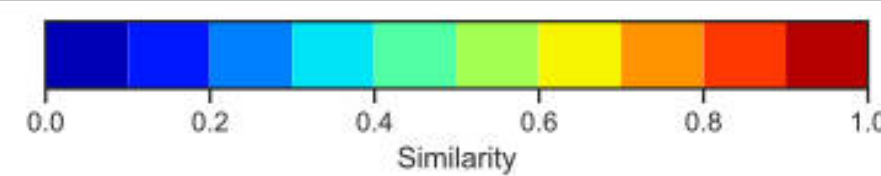


(a) The floor plan

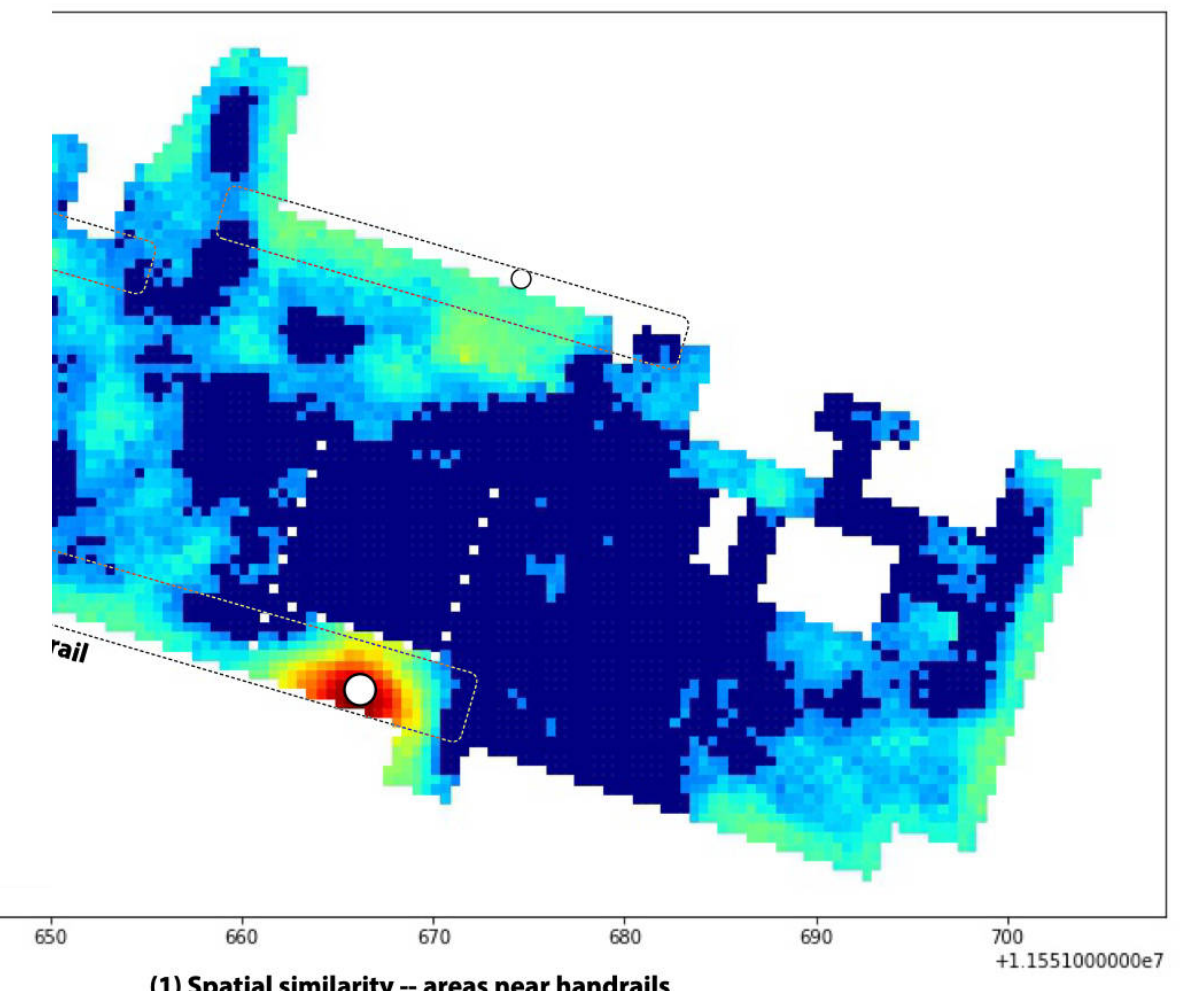


(b) Spatial similarity visualization

○ Test point

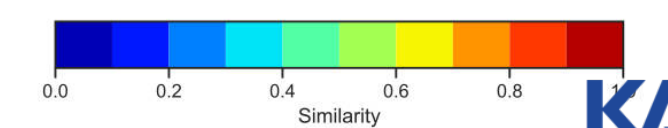


(1) Spatial similarity -- areas near doors

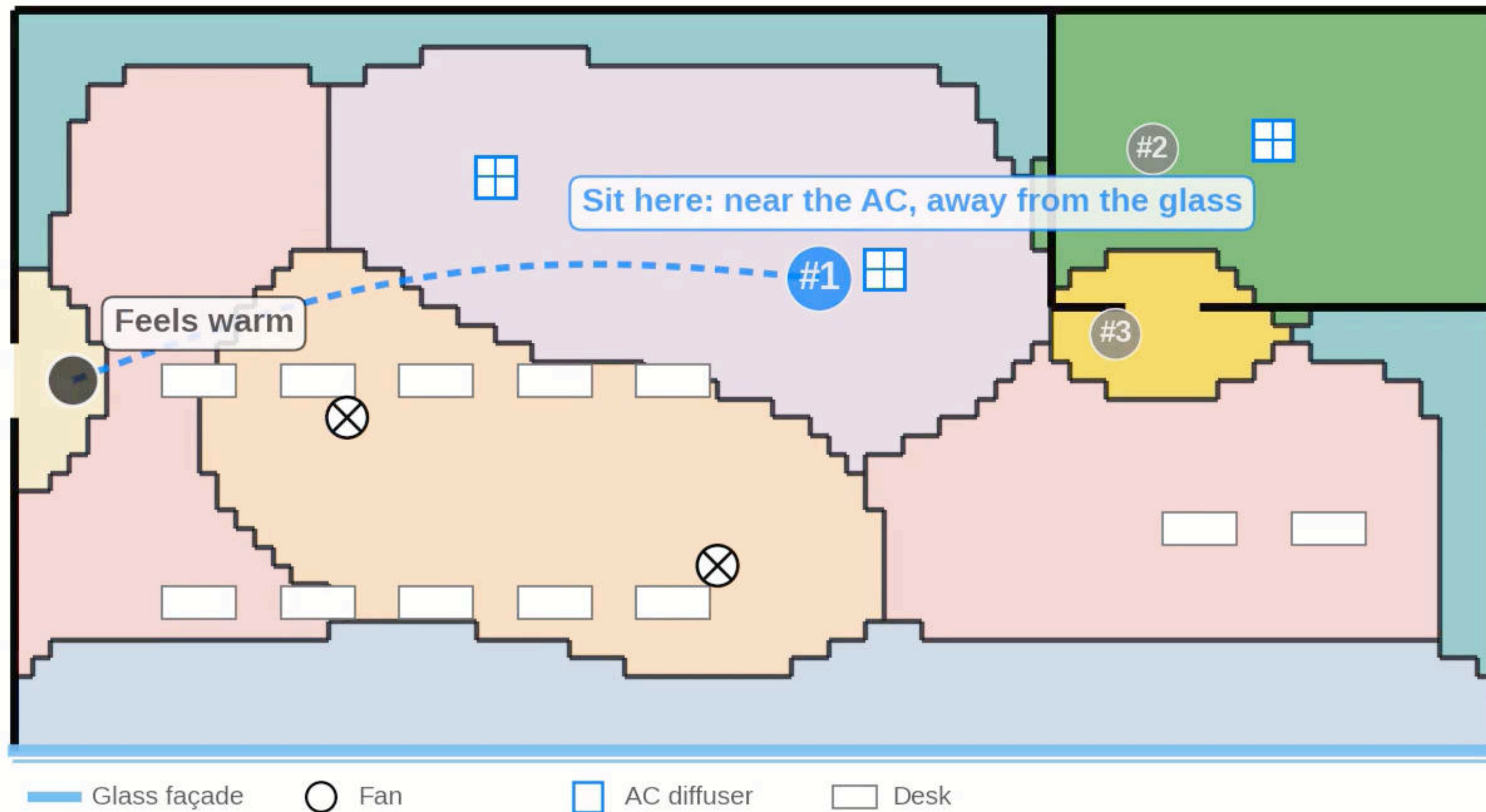


(1) Spatial similarity -- areas near handrails

○ Test point



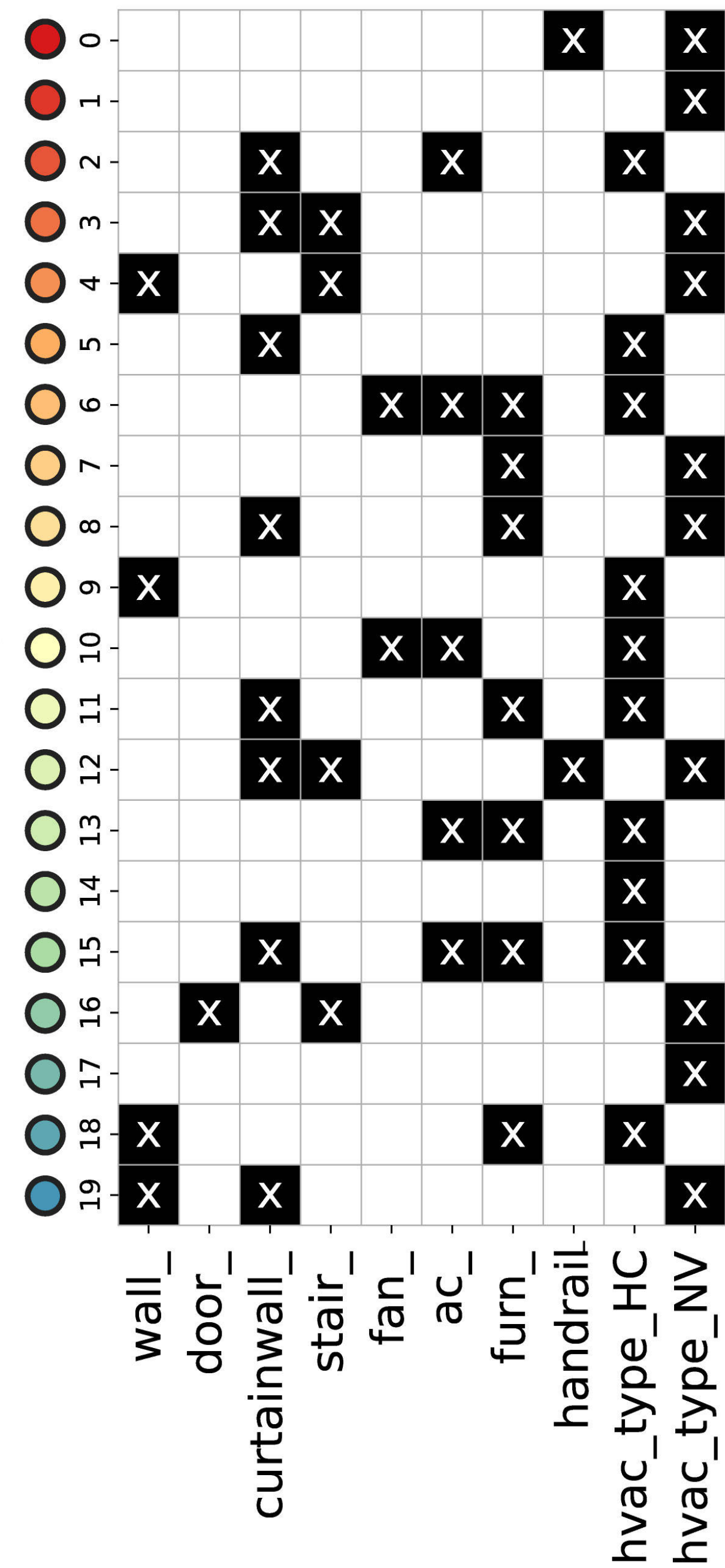
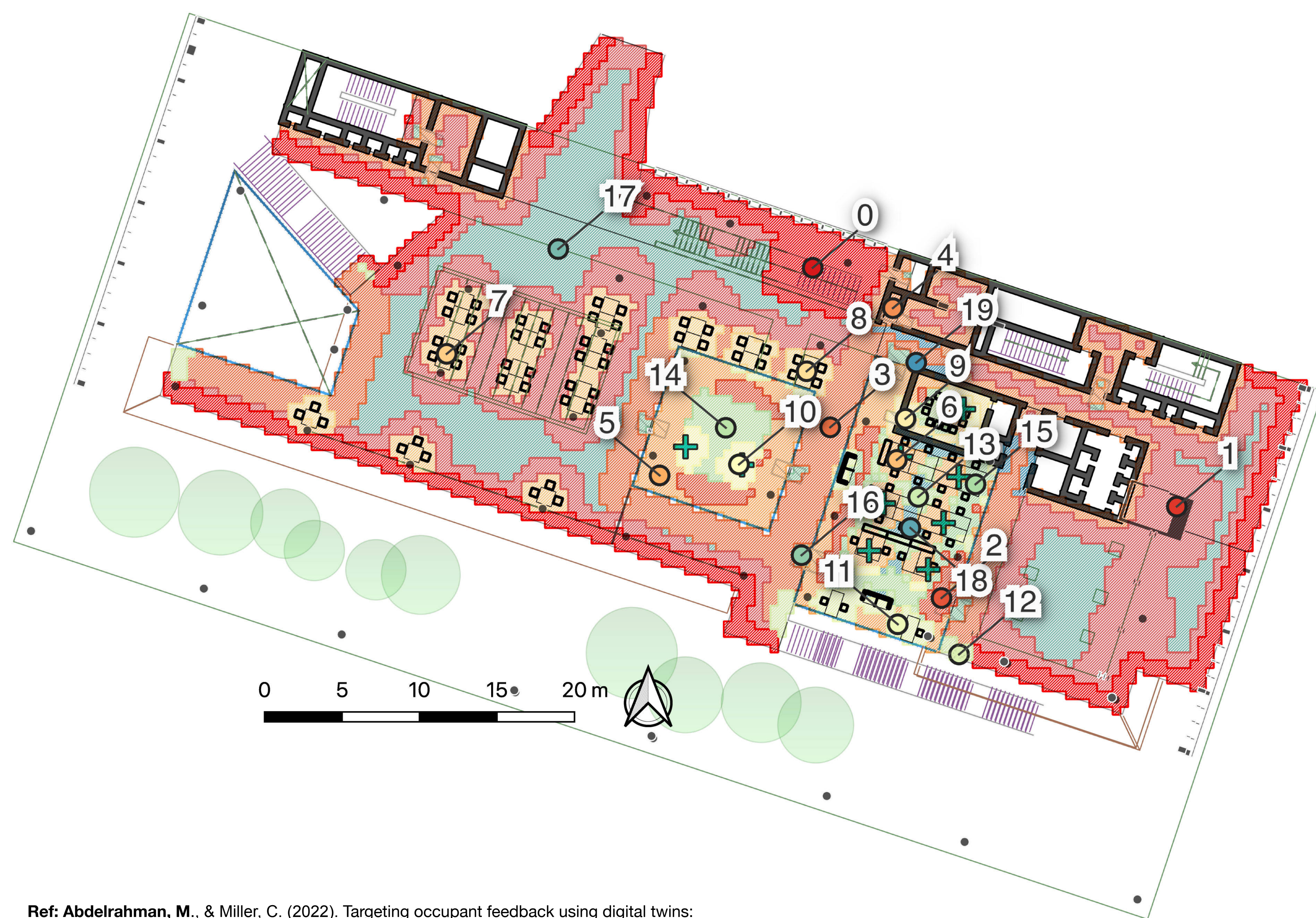
Spatial Clustering



- 1 Cells and objects
- 2 Area of influence
- 3 Cluster similar cells (Build2Vec)
- 4 Dissolve into zones
- 5 Recommend an action
- 6 Ask only where data is missing**

10

zones from 3,200 cells



From buildings to Cities

How we scale this up to a city scale ?

Buildings are more controllable experimental environments.

The urban system is much more complex.

We can apply similar concepts on the urban Scale



World Models :

Systems that build an internal representation of an environment. The model predicts how that environment changes over time in response to actions.

LLMs read and predict
tokens

The smallest unit of language: a token

City world model read and predict
space cells (voxels)

The smallest building block is a space cell

My hypothesis:

City world model read and predict
space cells (voxels)

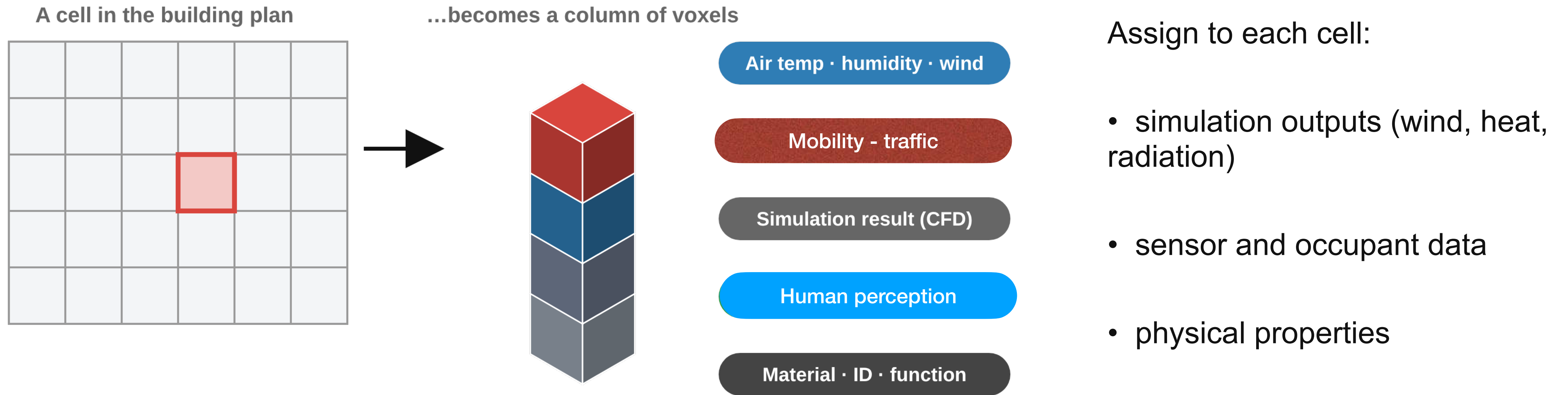
The smallest building block is a space cell

Step	LLM	City world model
1 Tokenizer	Cut text into tokens	Cut space into cells
2 Training data	Large, diverse text	Simulations on diverse city tiles
3 Pre-training	Predict the next token	Predict the state of every cell
4 Human feedback	People correct the model	Ask people where the model is unsure
5 Agents	Use tools and act	Recommend and act on the city

Research collaborations are welcome

Every cell can carry information

Not only a label and a location: simulation results and rich data, on a fine scale

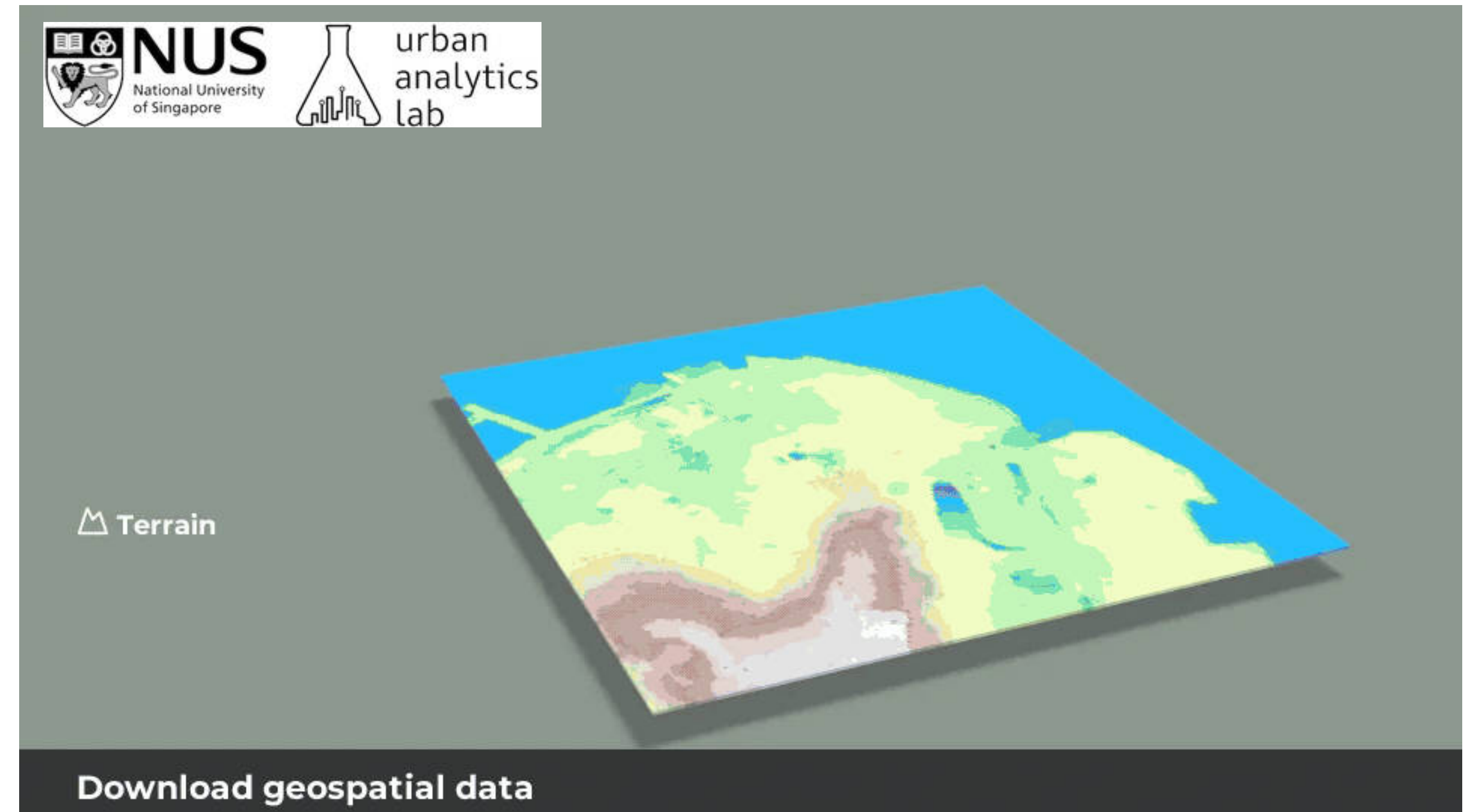
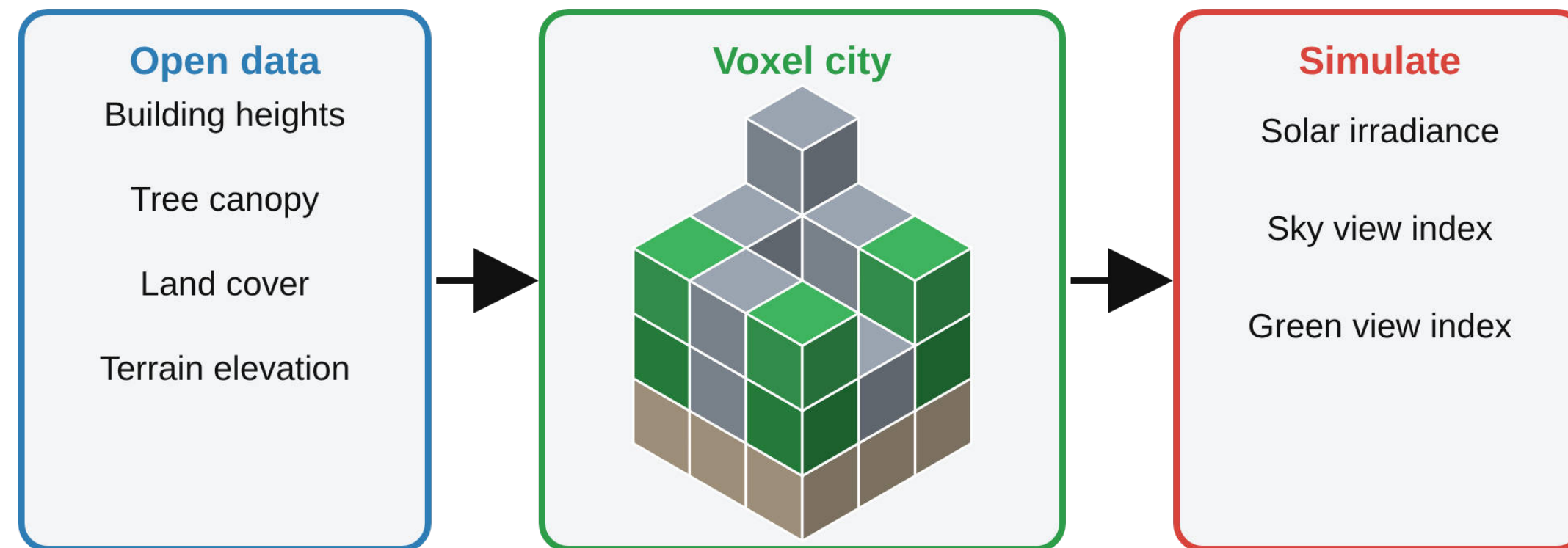


But how can we encode the whole city into cells ?

VOXELS

VoxCity: the city as a voxel grid

Open geospatial data in, a semantic 3D grid and environmental simulations out



Open source Python package. Demonstrated on eight cities worldwide: solar irradiance, sky view and green view indices.

Ref: Fujiwara, K., Tsurumi, R., Kiyono, T., Fan, Z., Liang, X., Lei, B., Yap, W., Ito, K., & Biljecki, F. (2026). VoxCity: A seamless framework for open geospatial data integration, grid-based semantic 3D city model generation, and urban environment simulation. *Computers, Environment and Urban Systems*, 102366. <https://doi.org/10.1016/j.compenvurbsys.2025.102366>

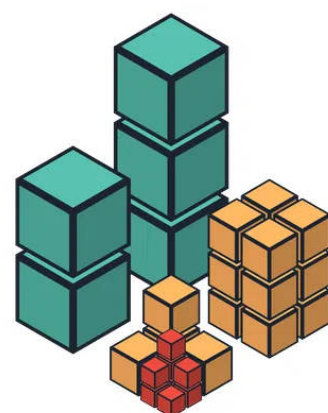
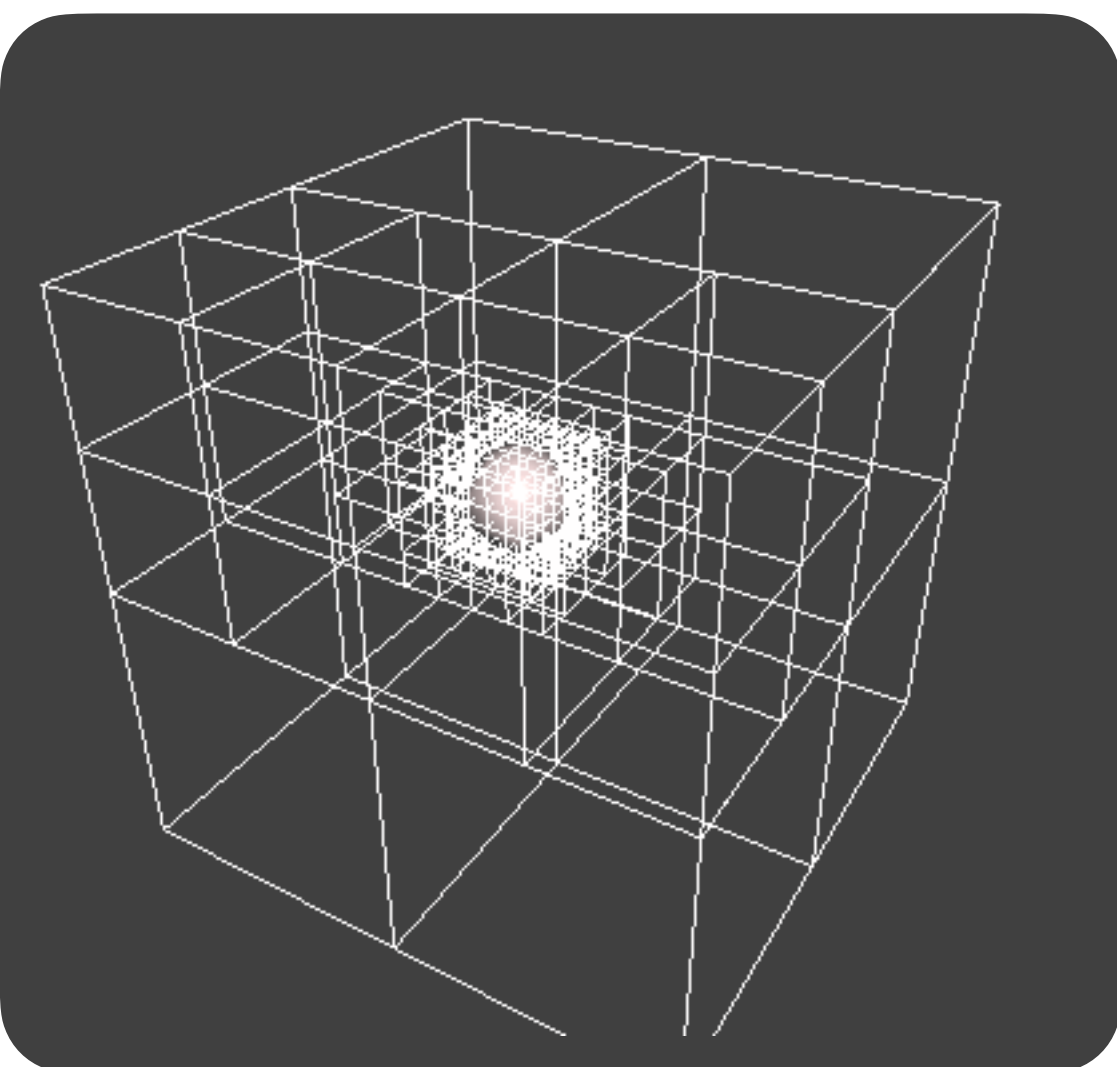
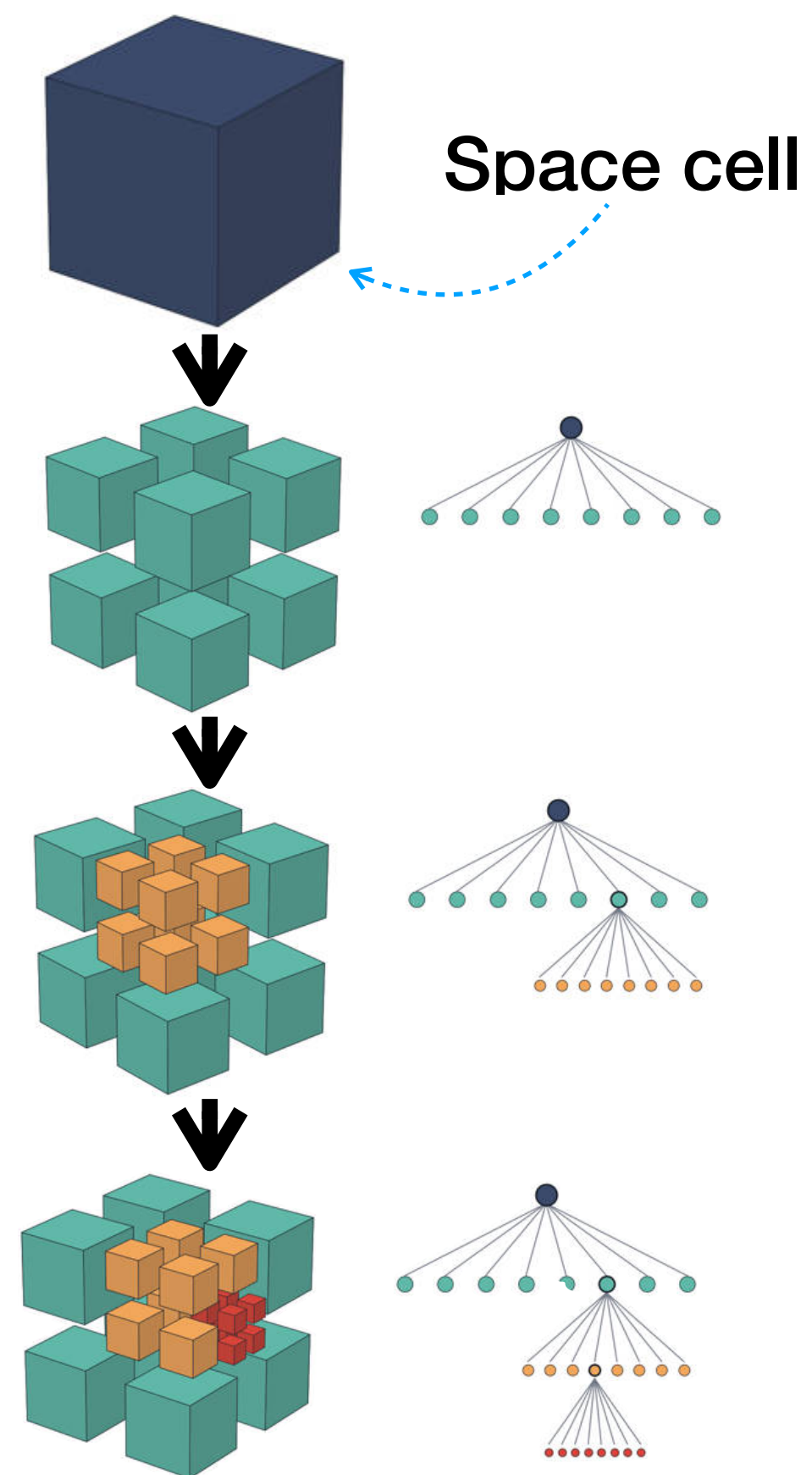
On a city scale, Voxels are not efficient

Octree

It is a tree data structure used in computer science—especially in **3D computer graphics** and **spatial indexing**

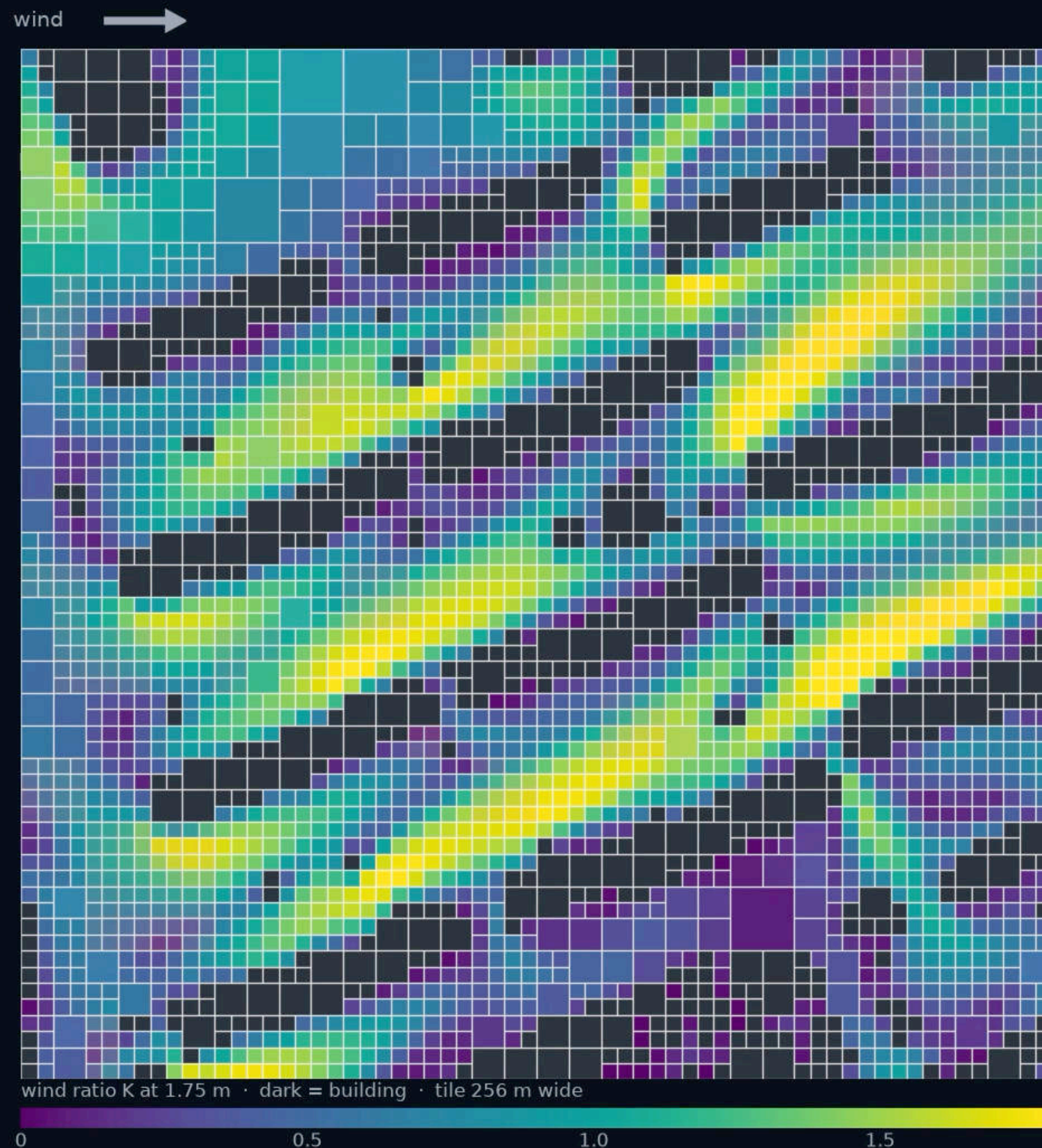


Why octrees are more efficient ?



OctreeCity

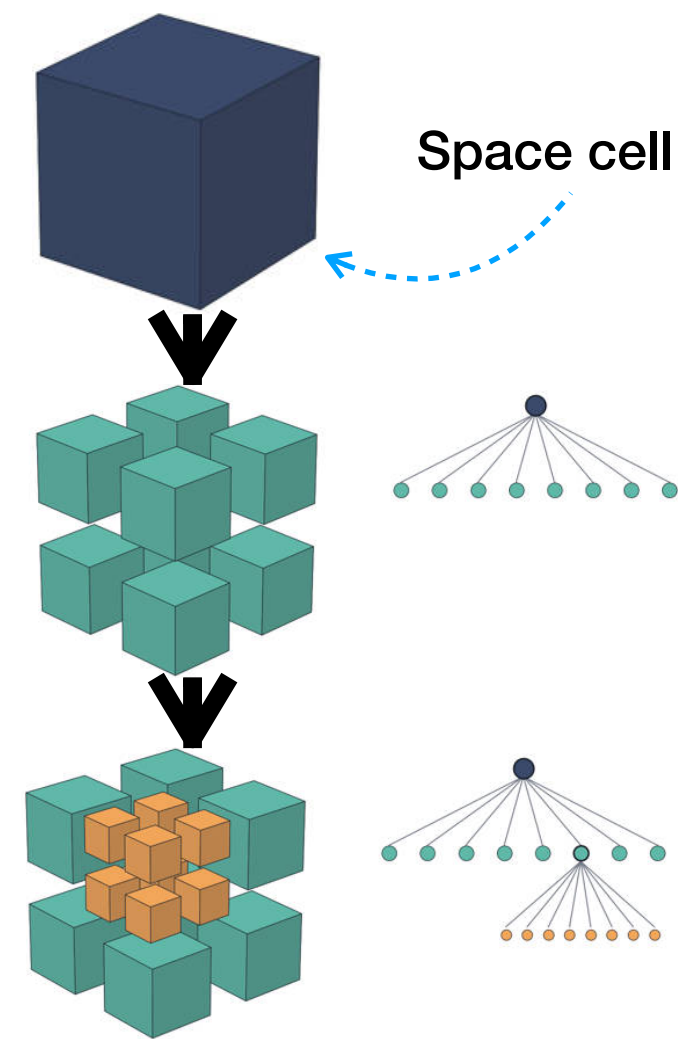
```
pip install octreecity
```



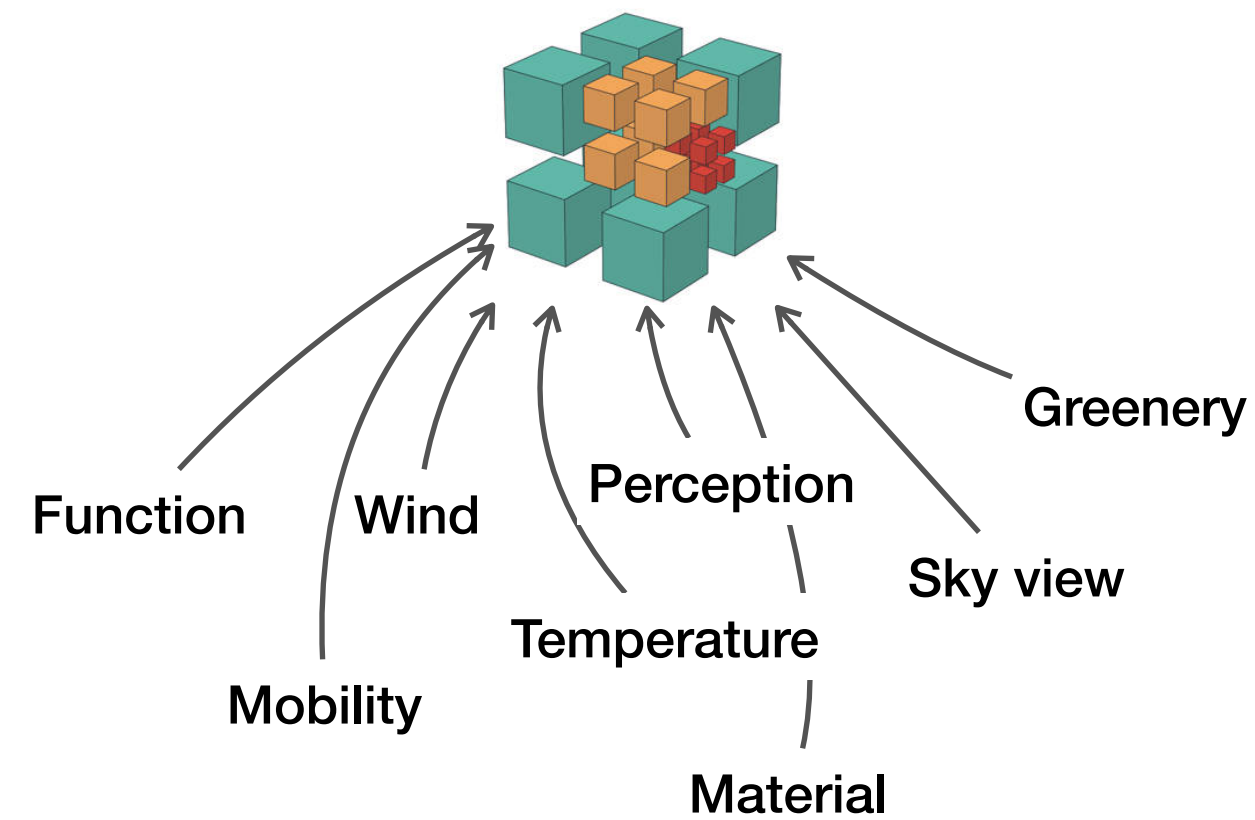
My hypothesis continued

The road to City Physical AI : Each object gets a brain

1 Spatial Tokenization

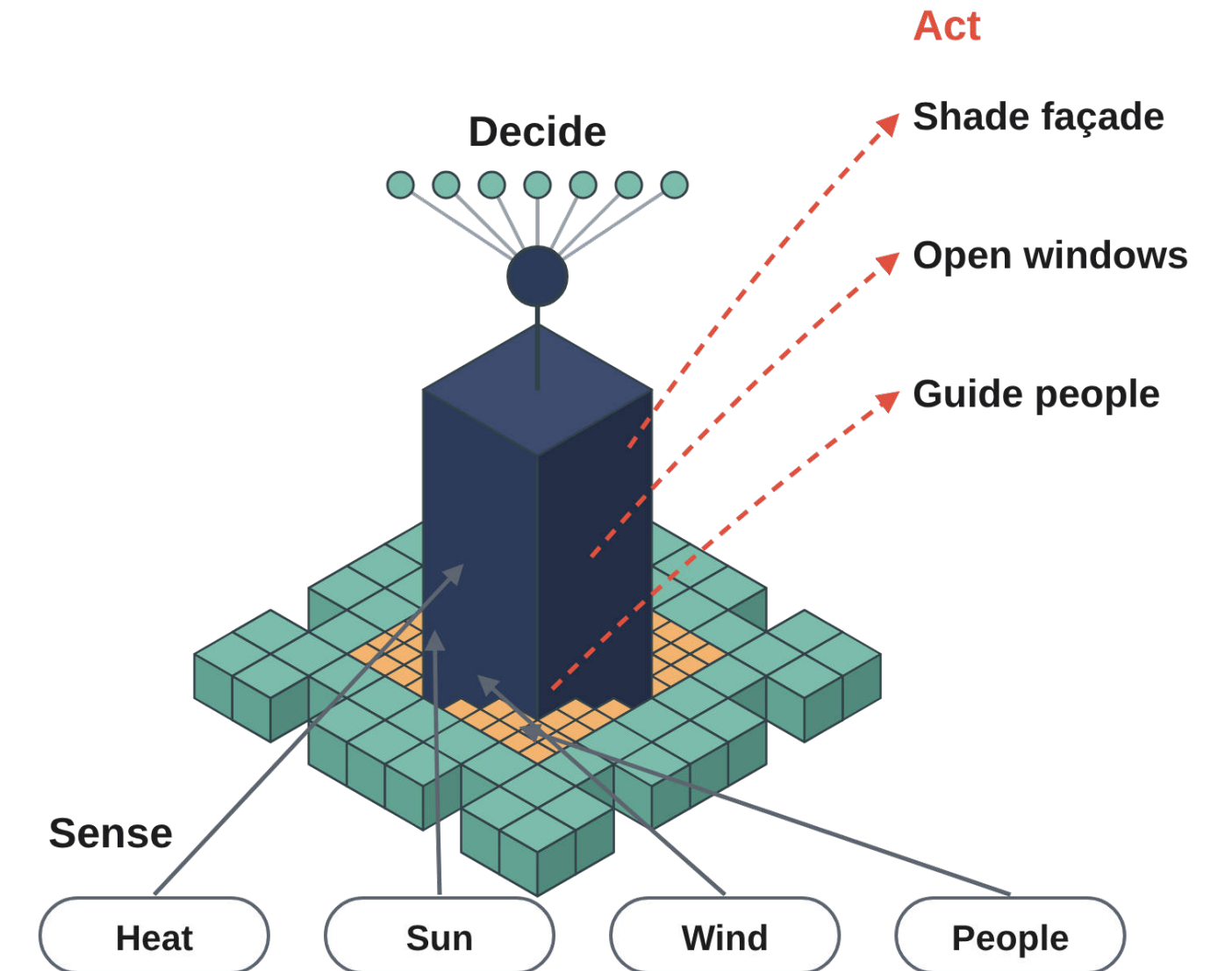


2 World model embedding

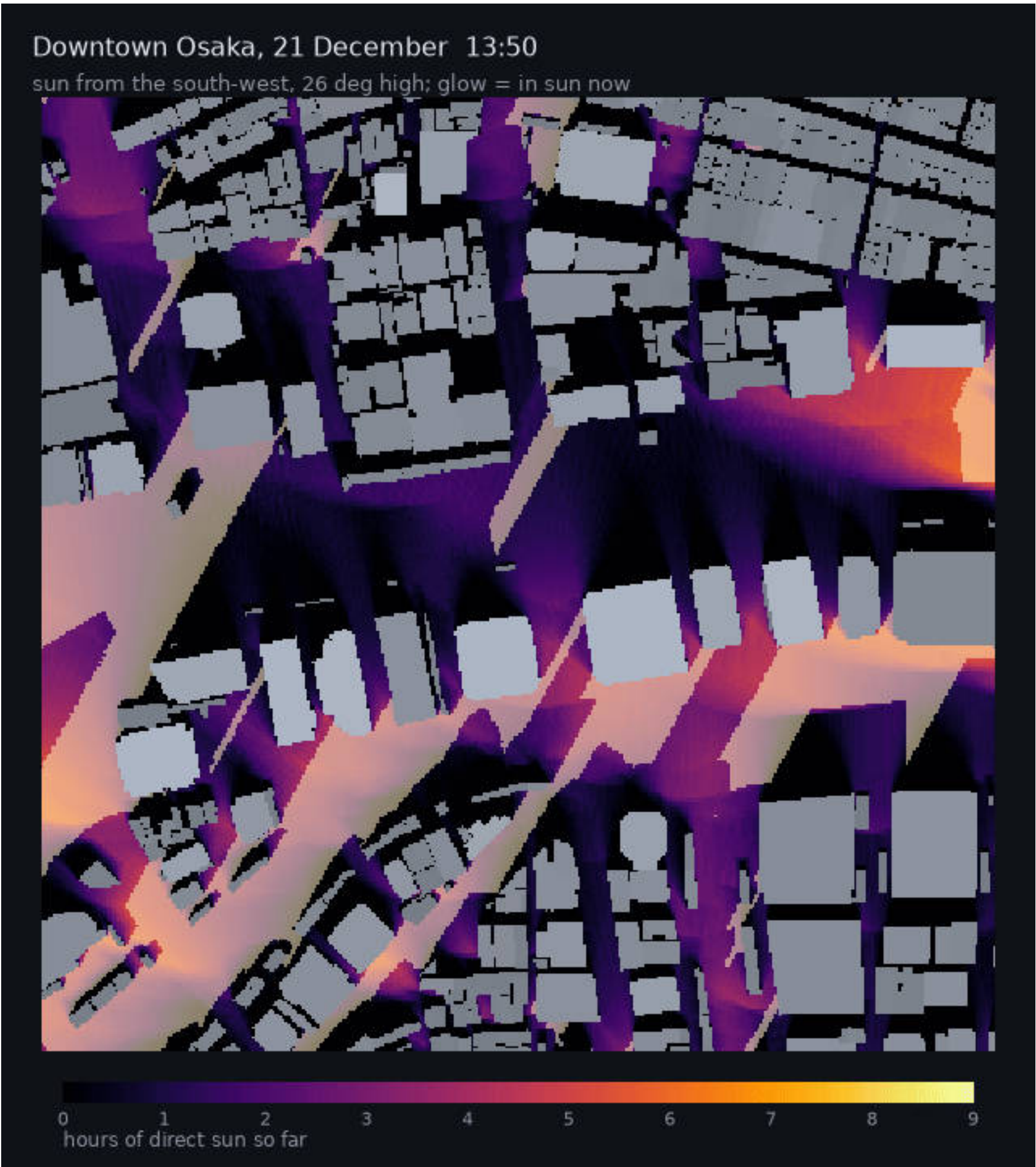


Real-time data
Physics-informed AI simulations
Human feedback

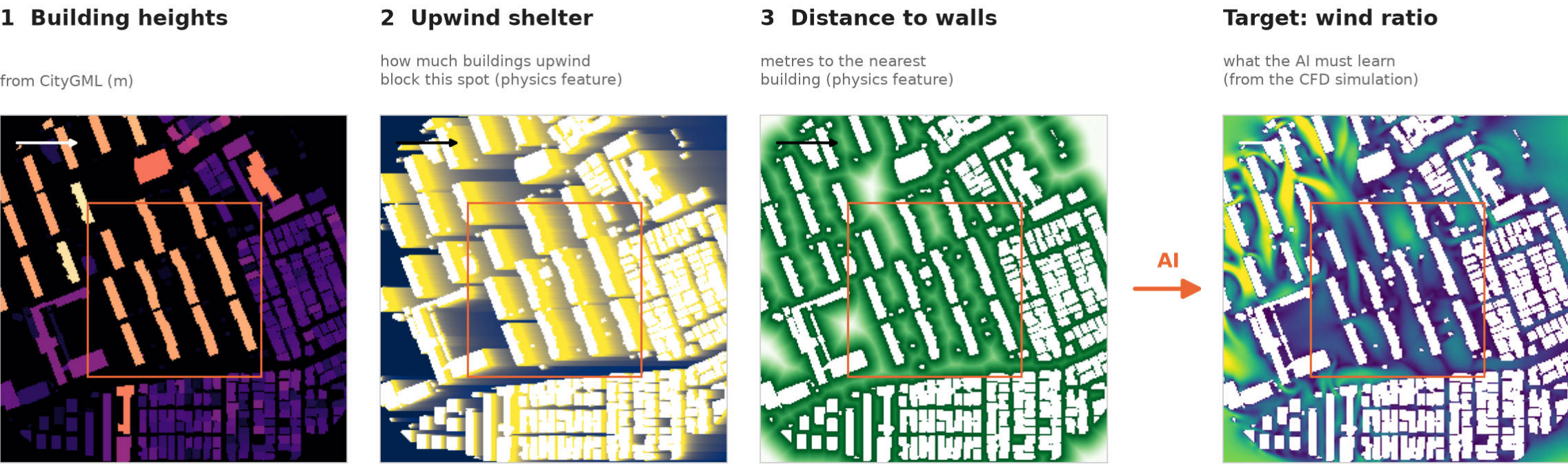
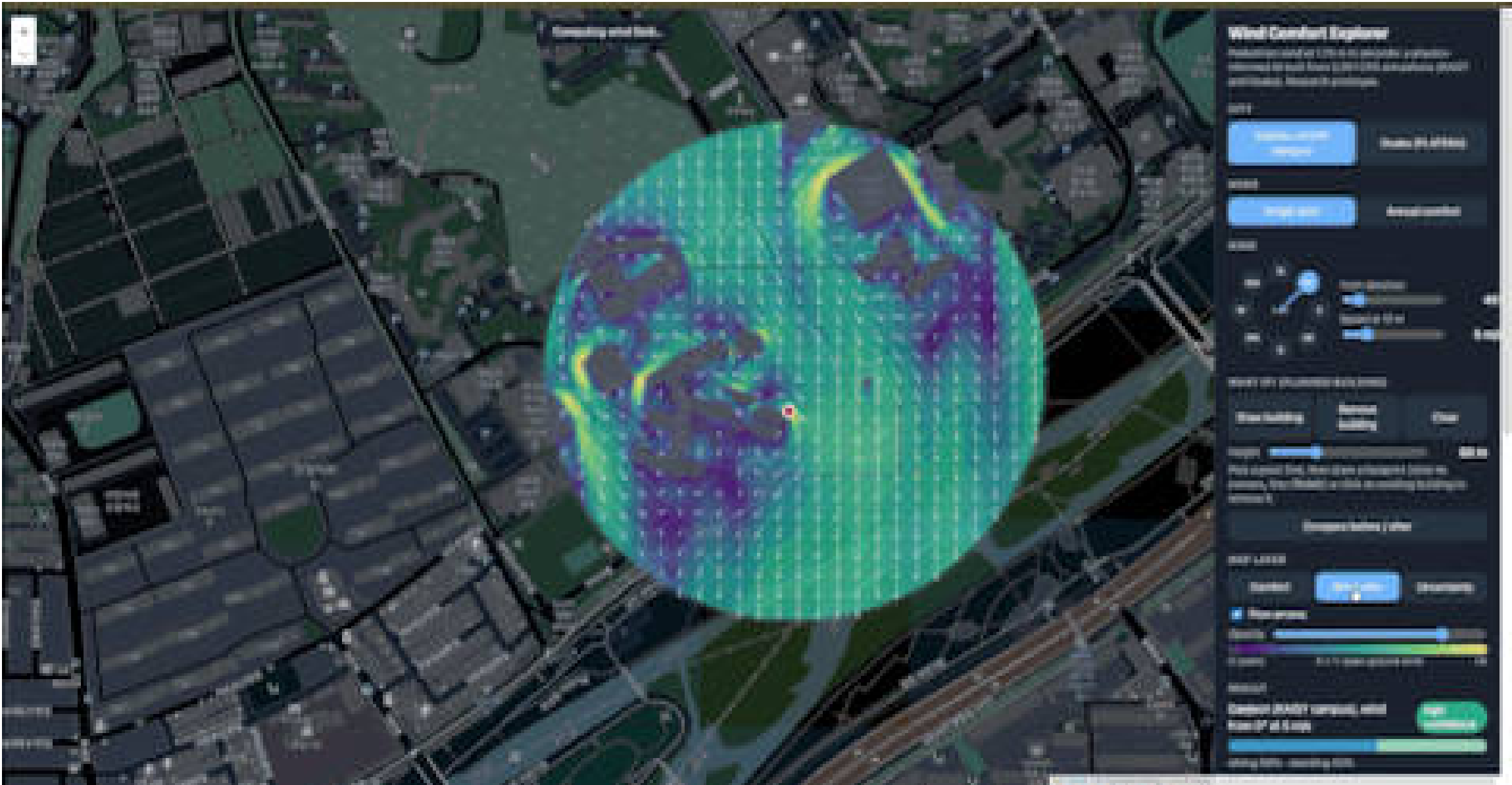
3 Physical AI (Agency)



Demo : [OctreeCity](#) [github]
[OctreeCity](#) [PyPI]



Demo : [Physics-informed AI](#)



512 m x 512 m around the tile (orange square = the 256 m tile scored), 2 m pixels, wind always from the left (arrow).

Take-aways



- 1 Ask people, not only equations**
- 2 Add more context - (spatial, perception, ...)**
- 3 Give every cell and object an identity, data, and actions**

Let's collaborate

<https://mahmoud.work>

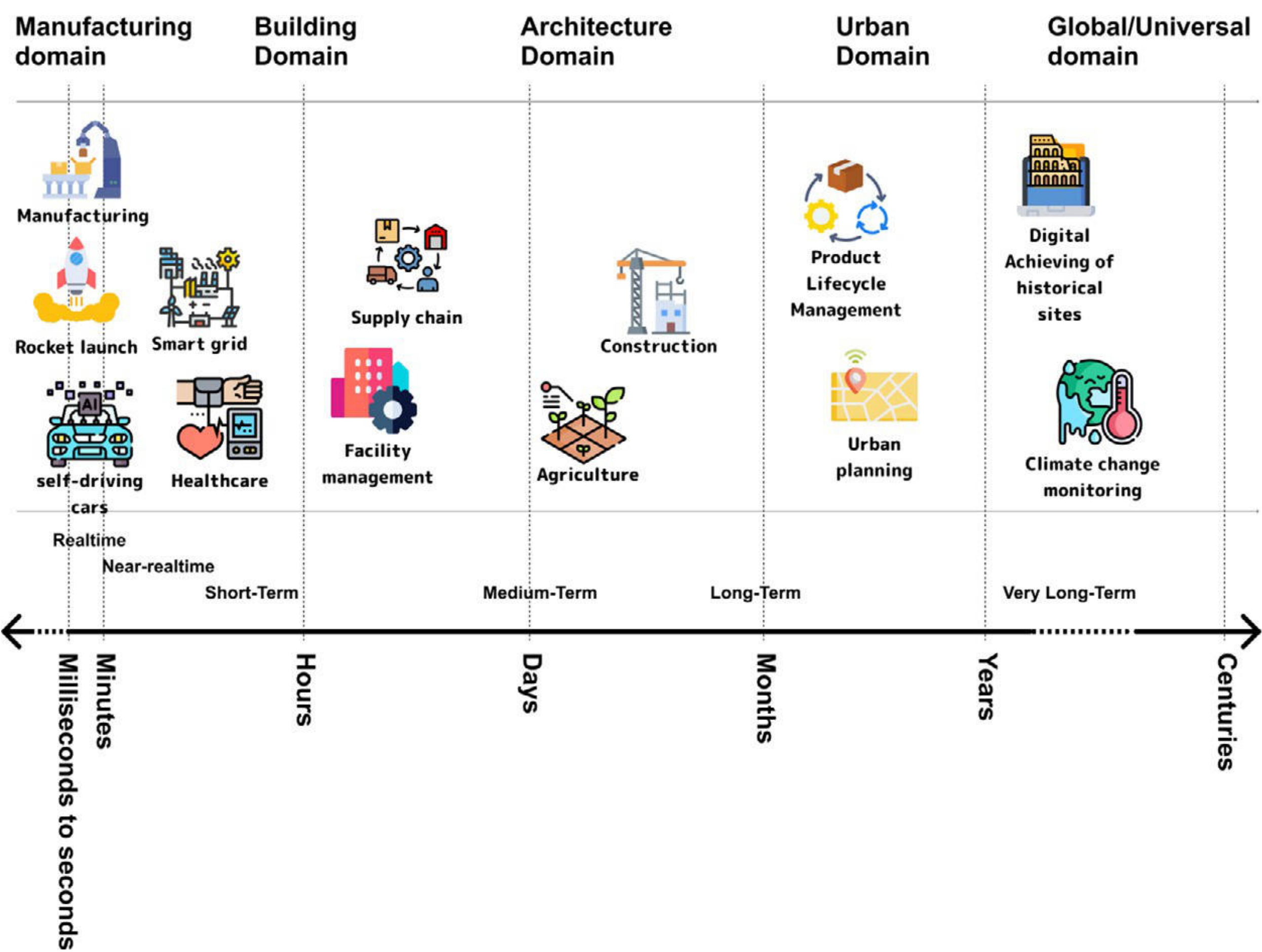
mahmoud@kaist.ac.kr

Supplementary slides

Digital twin definitions and maturity

What is a digital twin, anyway?

We asked the literature and the experts, and the answer depends on the scale



15,000+

full-text articles analysed with NLP

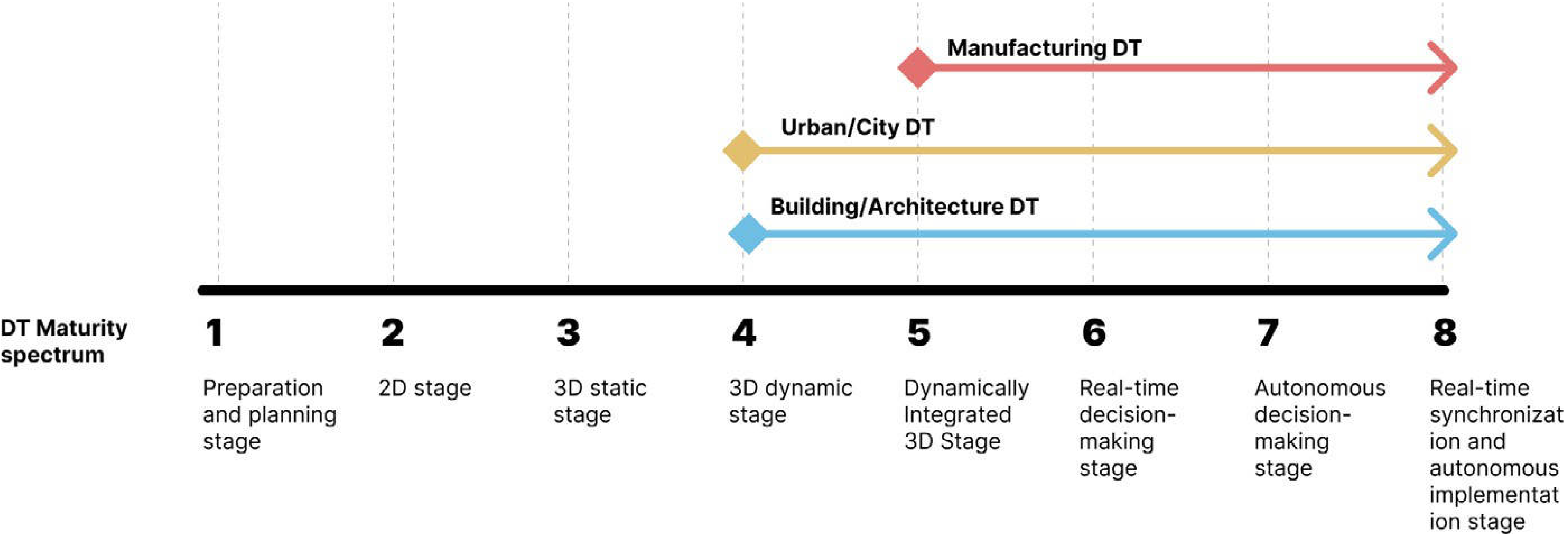
52

experts surveyed

Two definitions: building / architecture and city / urban

How mature are digital twins?

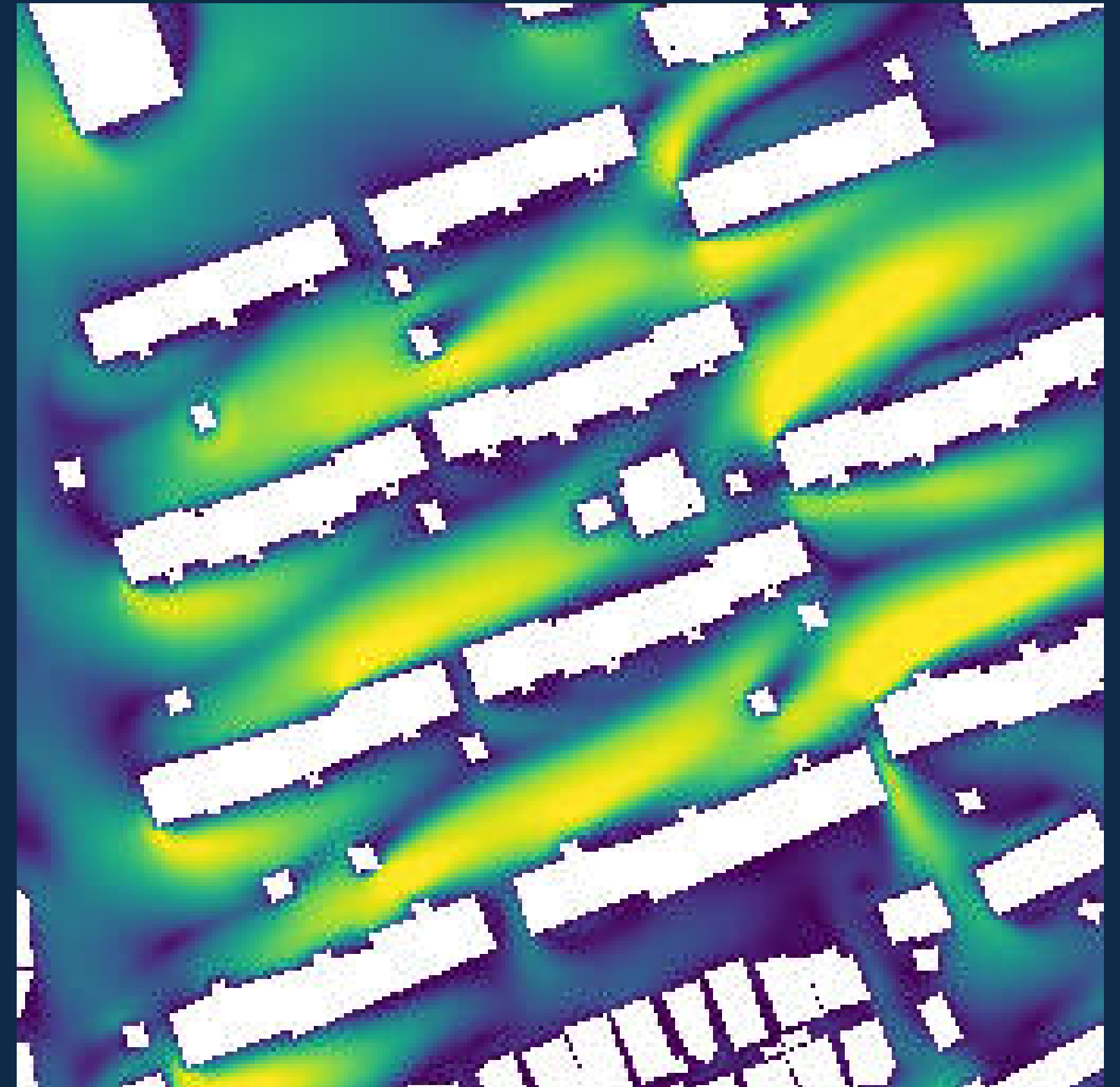
Maturity of digital twins across domains



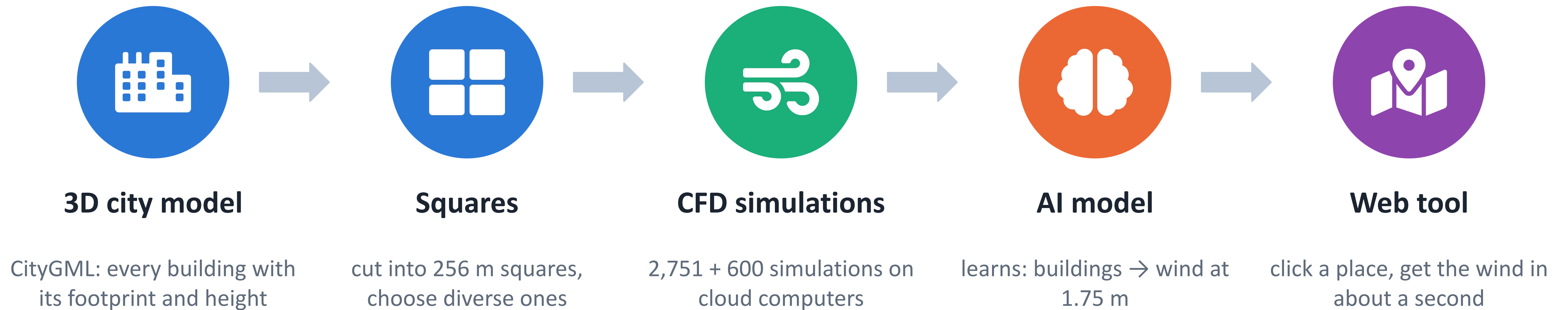
Wind at street level, in seconds

From 3D city models to a physics-informed AI for pedestrian wind comfort

Mahmoud · KAIST Urban AI · October 2026

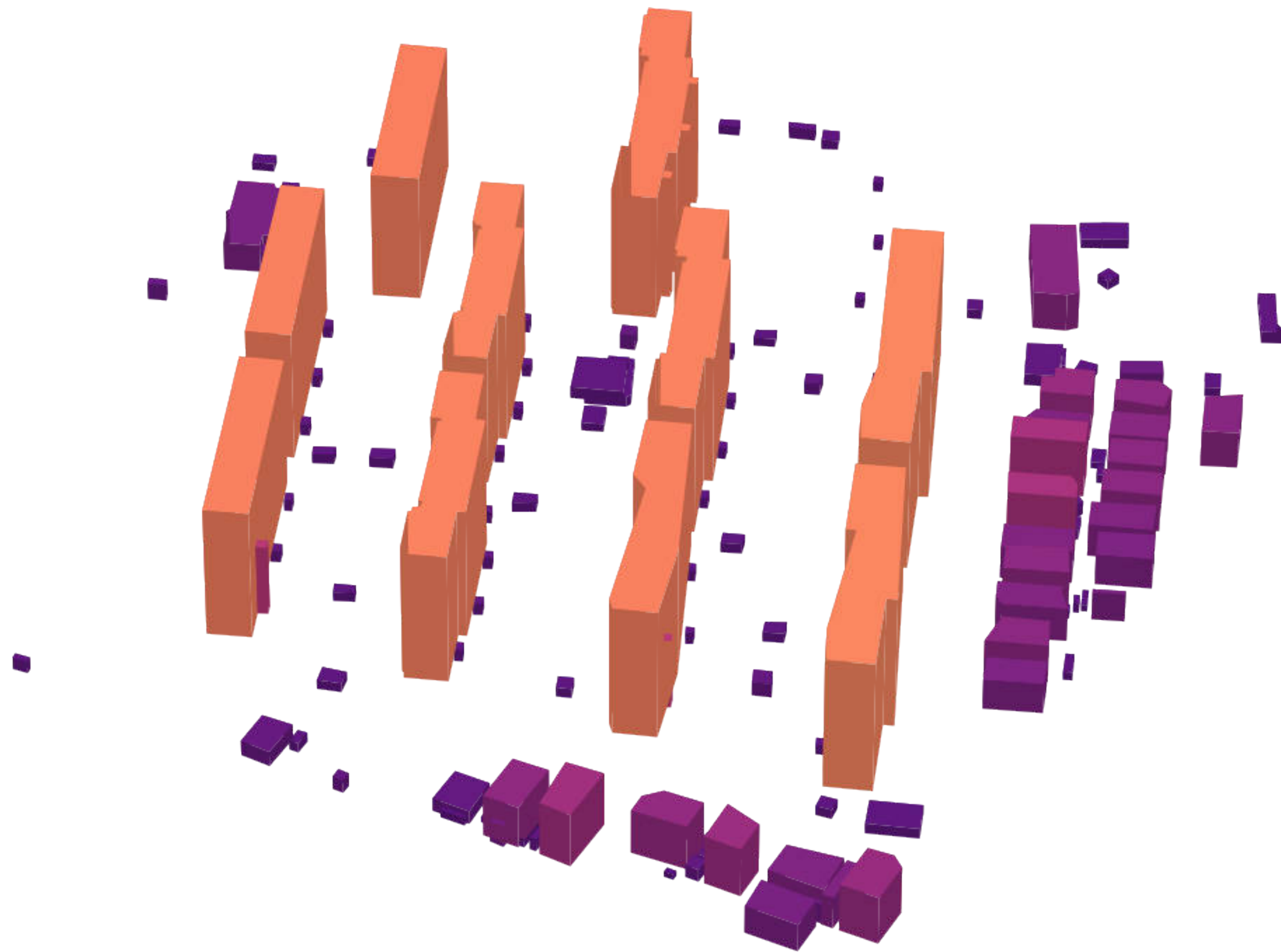


The project in one picture



Daejeon (KAIST campus, Korea) first, then Osaka (Japan) as a completely new city

Start: a free 3D city model



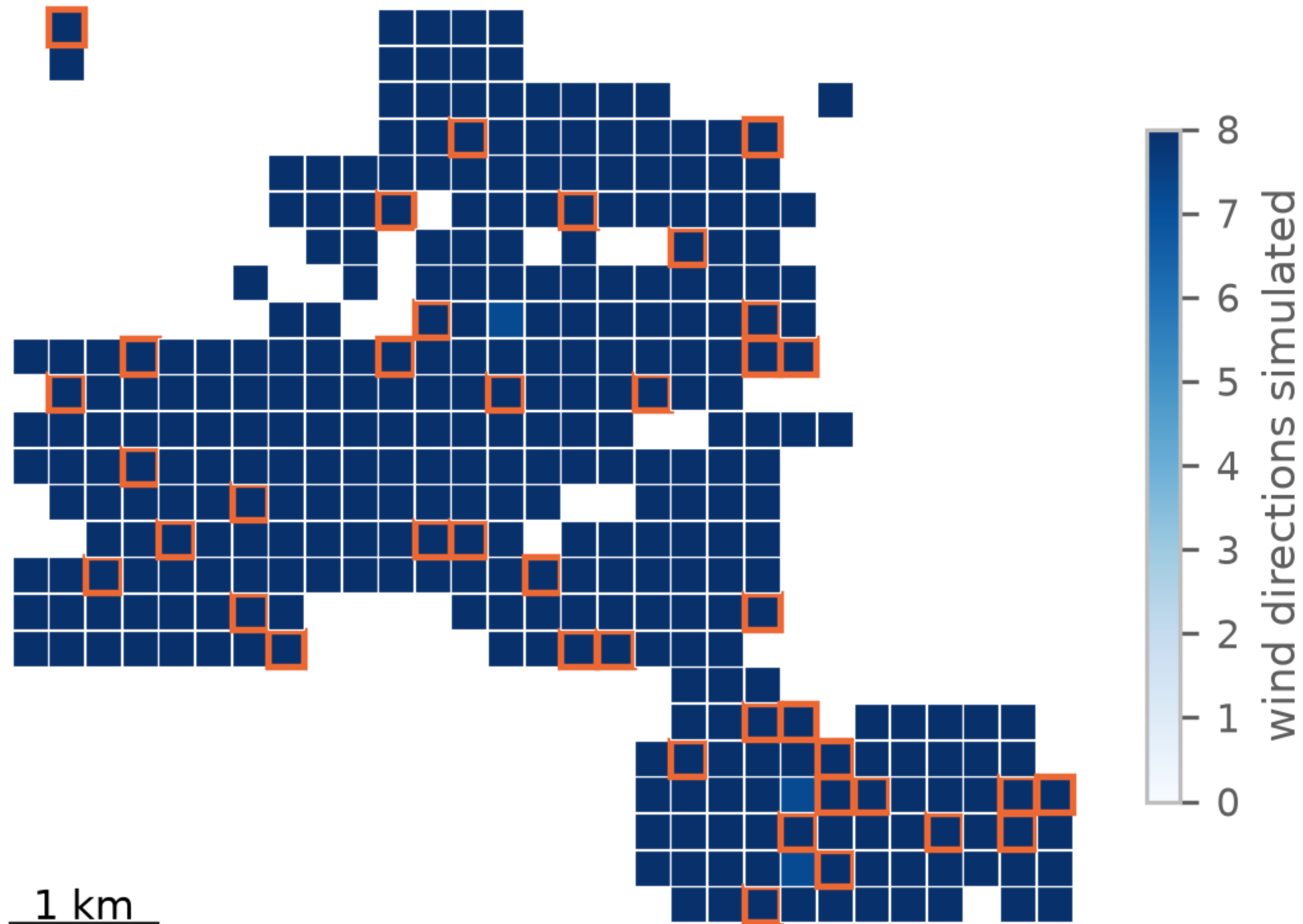
CityGML: open standard for 3D city models

Daejeon, around KAIST: 14,140 buildings (V-World)

Osaka: 556,724 buildings (PLATEAU)

What we use: footprint + height of every building

Cut the city into 256 m squares



334

squares of 256 m around KAIST

3,240

squares in Osaka

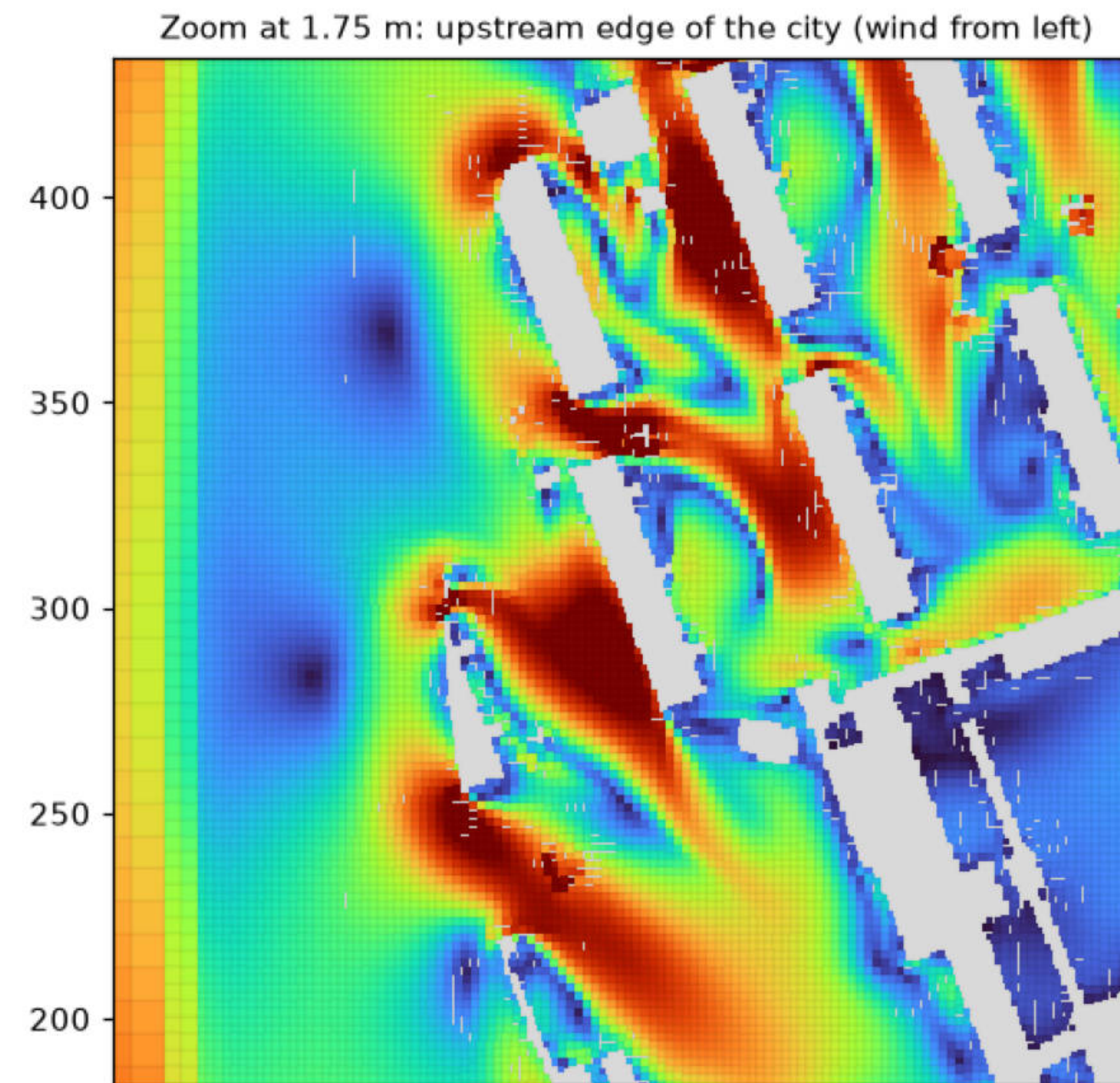
Each square is simulated with a 128 m ring of neighbouring buildings, because wind is shaped by what stands upwind.

How a wind simulation works

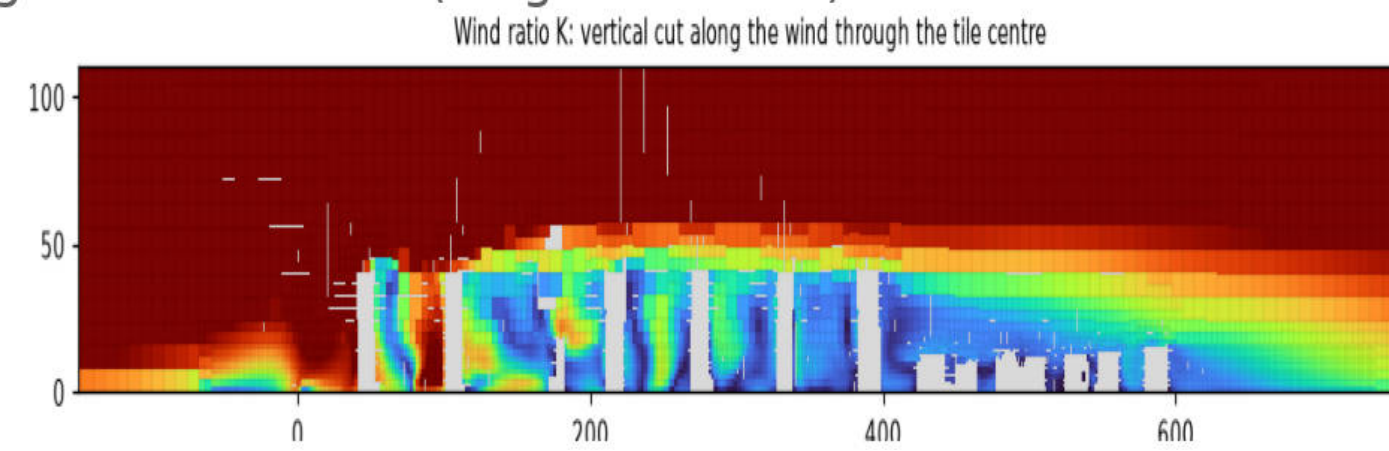
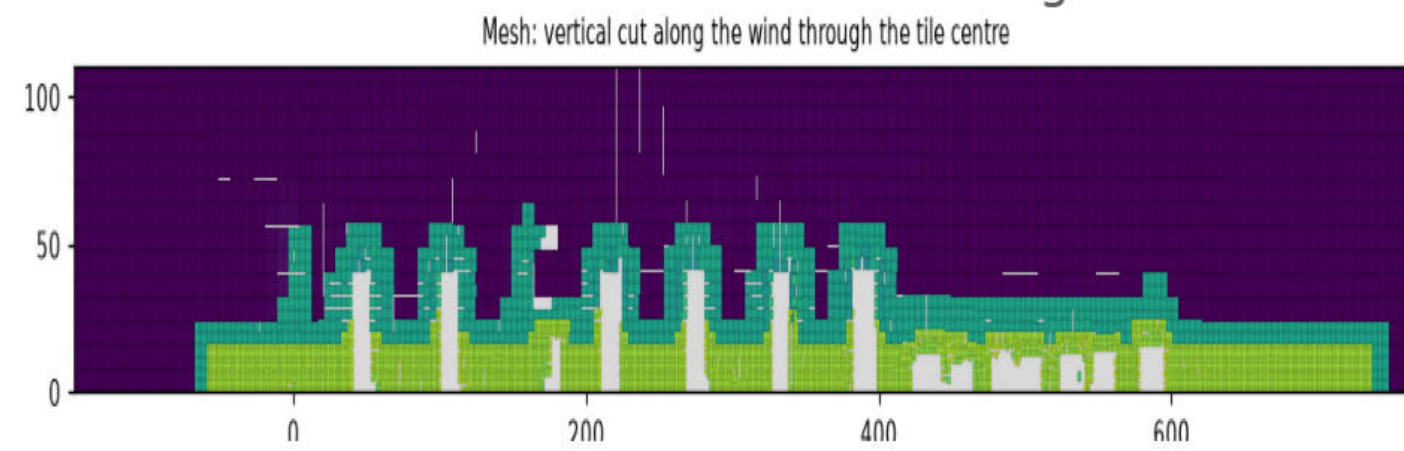
(a) mesh at 1.75 m: cell size 8 m (dark) to 2 m (light green)



(b) wind ratio on the mesh (red fast, blue calm)



vertical cuts along the wind through the tile centre (height 0-110 m)



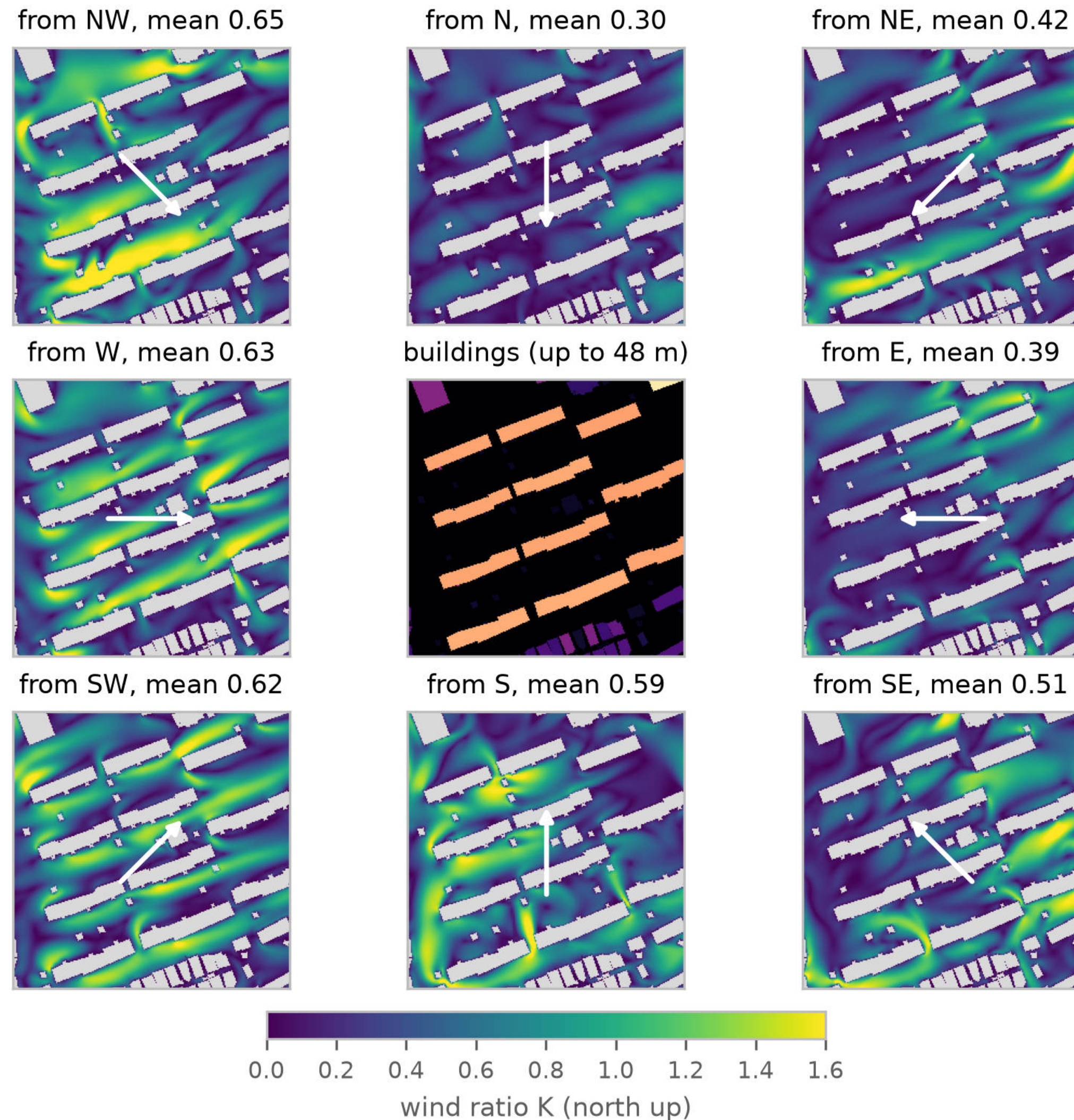
Mesh: 1–5 million cells, 2 m near the ground

Wind: realistic profile, 5 m/s at 10 m height

Solver: OpenFOAM, steady RANS turbulence model

Output: wind at 1.75 m (pedestrian height)

Every square, eight wind directions



The result we predict: the wind ratio K

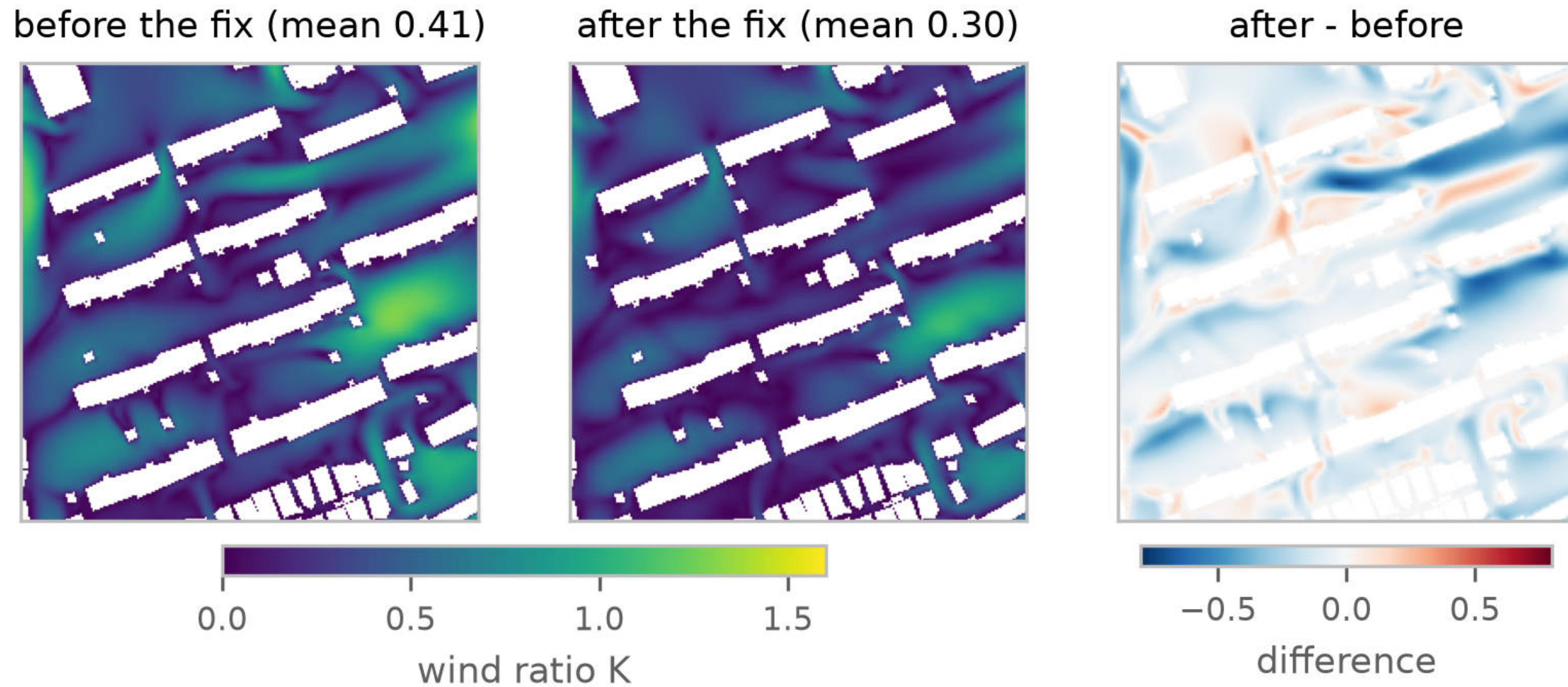
K = 1: as windy as an open field

K = 0.5: half as windy (sheltered)

K = 1.5: 50% windier (a jet between buildings)

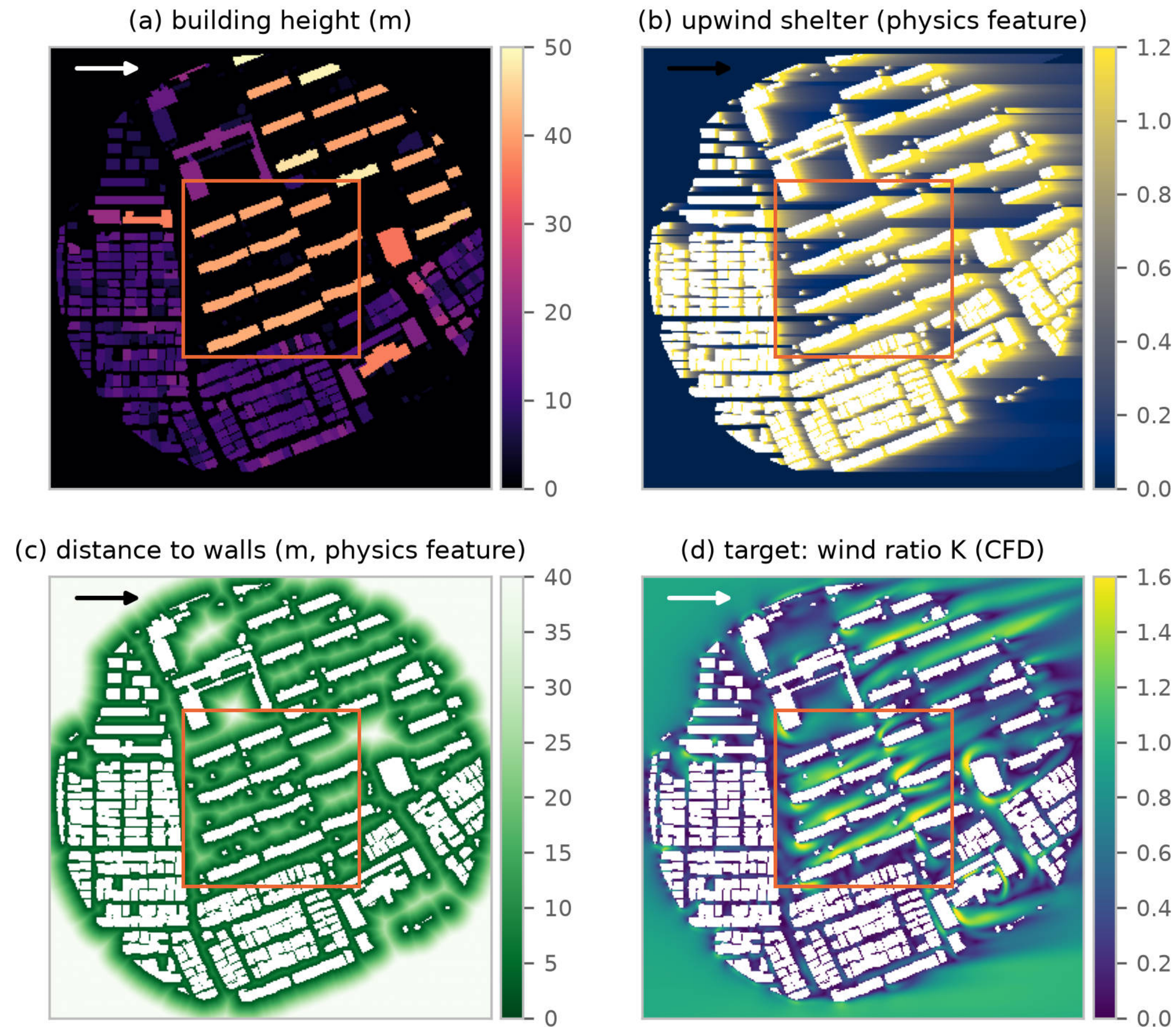
K does not depend on the wind speed, so one simulation answers every speed: multiply K by the local wind.

Getting the physics right before scaling up



A roughness mismatch made the wind up to 48% too fast over empty ground. Fixed before the large run: changes of up to a third in the results.

What the AI sees and what it predicts



Inputs (4 maps)

building height

building footprints

upwind shelter (physics hint)

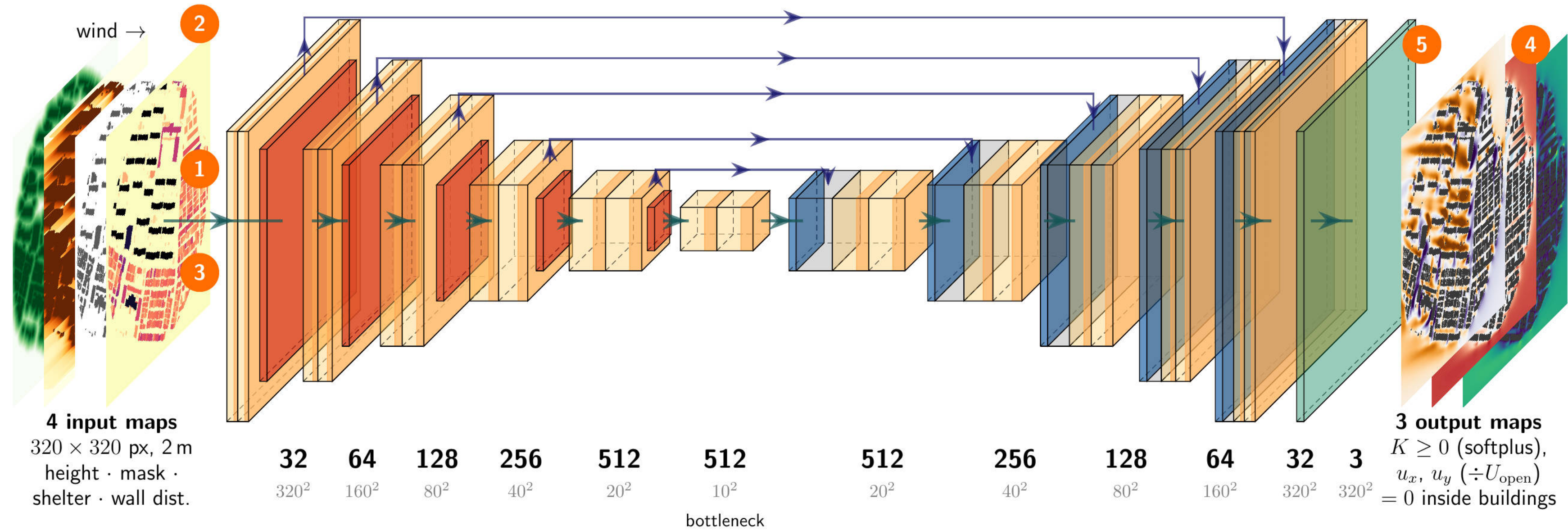
distance to walls (physics hint)

Outputs

wind ratio K at every 2 m pixel

wind direction (east and north parts)

The network: a U-Net with physics built in



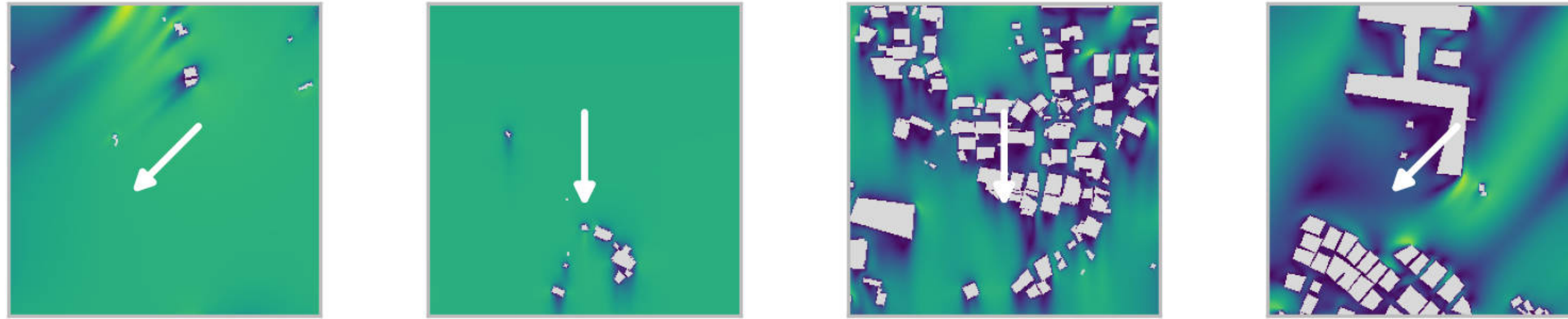
● residual conv block (GroupNorm, GELU) ● max-pool ● up-sample ● skip concat ● 1×1 conv head under each block: **channels**, map size

where physics enters

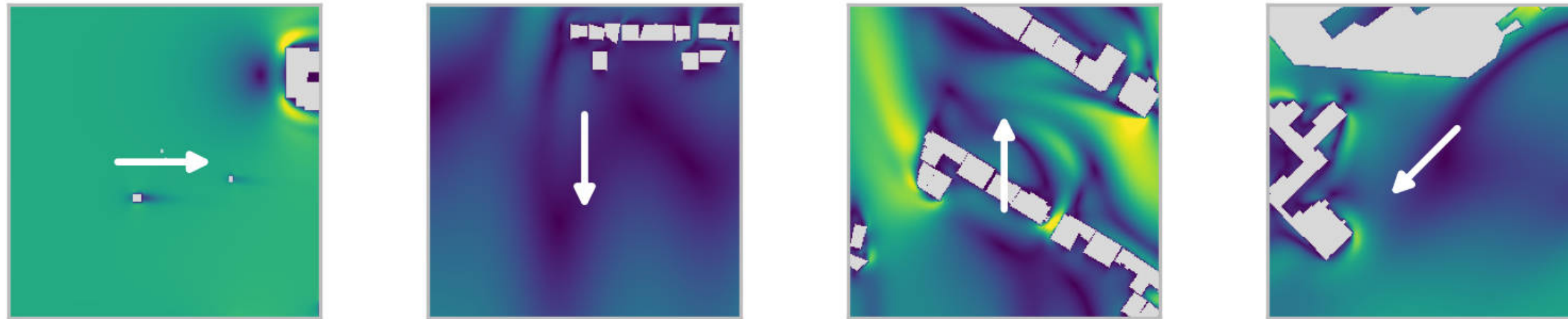
- 1 **physics inputs:** upwind shelter $\max_d h(x-d)/d$ over 300 m; distance to walls
- 2 **formulation:** city rotated so the wind always blows from the left; dimensionless $K = |\mathbf{u}|/U_{\text{open}}$
- 3 **mirror symmetry** across the wind axis ($u_y \rightarrow -u_y$): training augmentation and test-time averaging
- 4 **hard constraints:** $K \geq 0$ via softplus; every output set to 0 inside buildings
- 5 **physics-aware loss:** $|K - K^*|(1+K^*) + 0.5|\mathbf{u} - \mathbf{u}^*|(1+K^*) + 0.2||\mathbf{u}| - K|$ (consistency) + $0.5|\nabla K - \nabla K^*|$ (gradients)

How well does it work? Very close to CFD

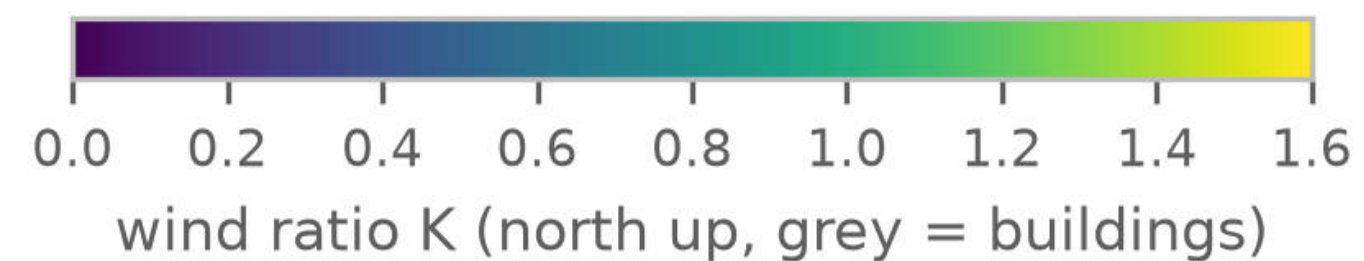
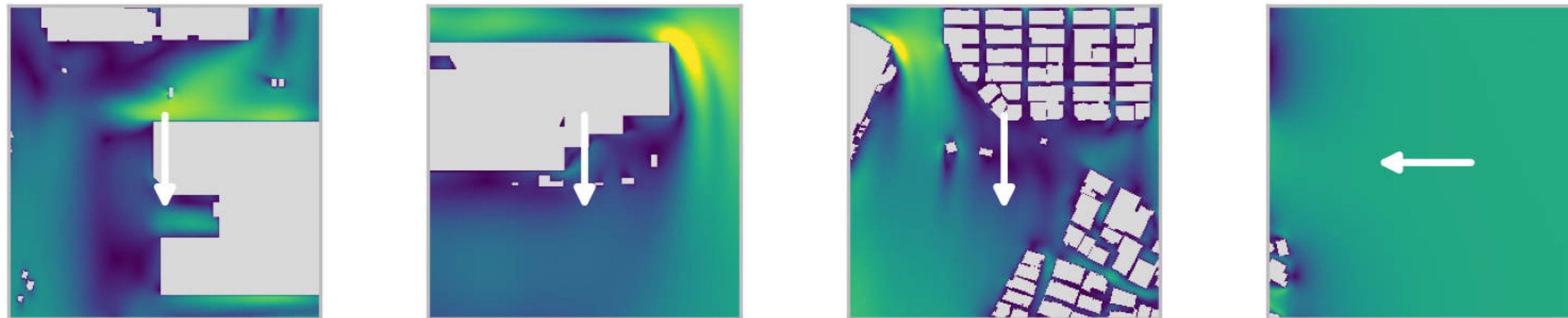
square 12, NE, mean 0.97 square 15, N, mean 0.99 square 18, N, mean 0.62 square 56, NE, mean 0.60



square 102, W, mean 0.96 square 188, N, mean 0.31 square 224, S, mean 0.66 square 263, NE, mean 0.55



square 277, N, mean 0.54 square 290, N, mean 0.66 square 327, N, mean 0.55 square 333, E, mean 0.91



0.071

mean error in wind ratio K (unseen squares)

86%

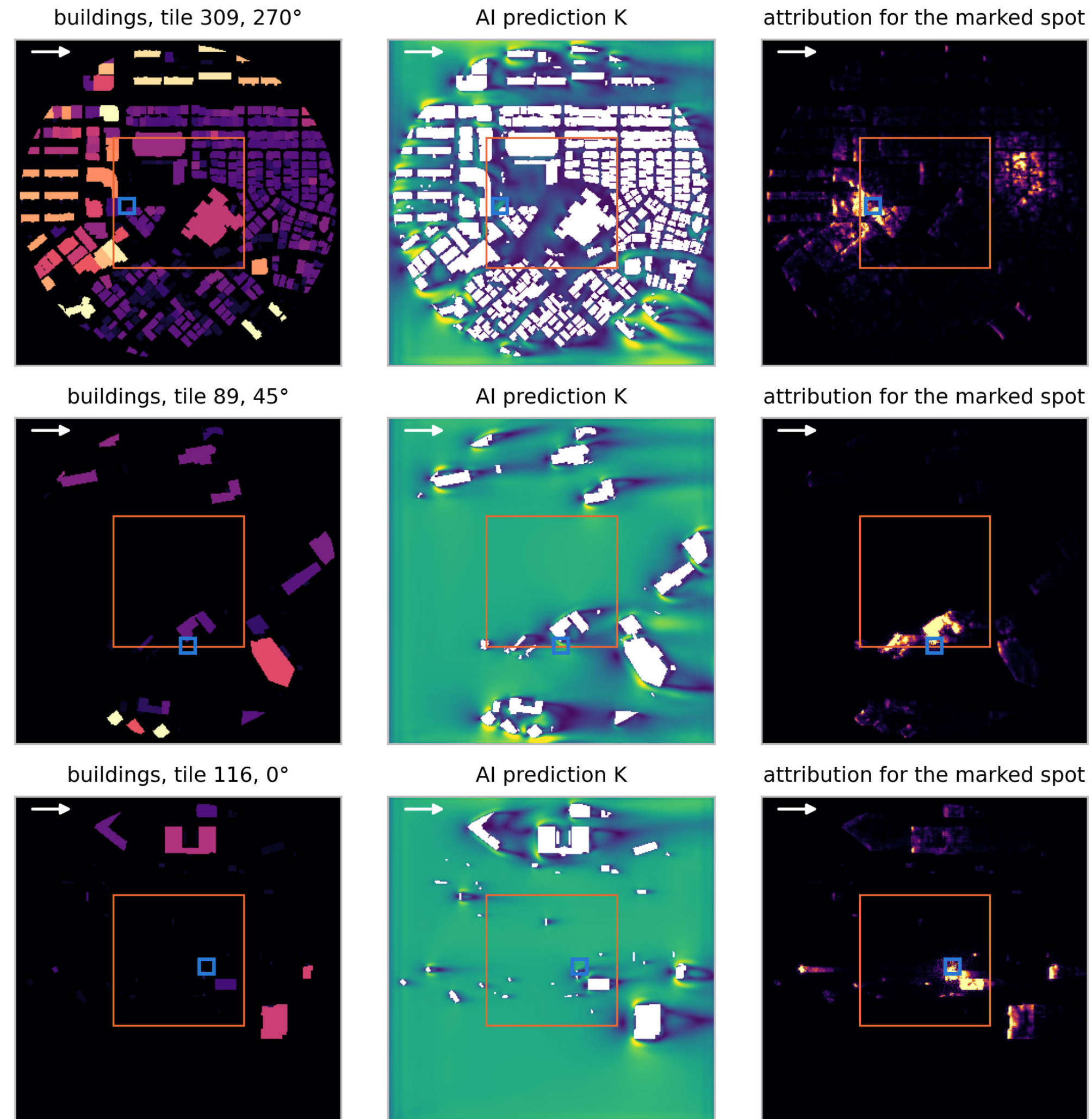
of the area in the right comfort class

170,000×

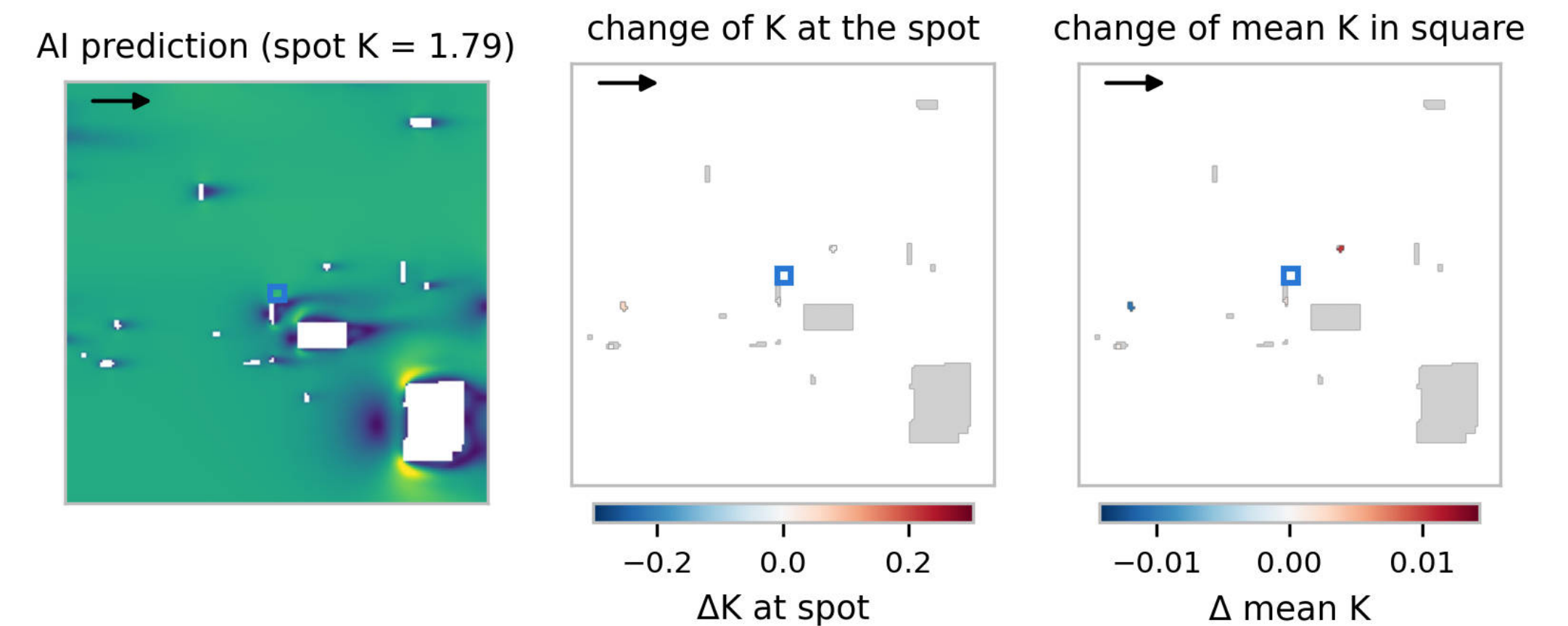
faster than CFD (15 ms on a GPU, 1 s on a laptop)

Left: CFD (top of each pair) vs AI (bottom) on test squares.

Does the AI look at the right things?

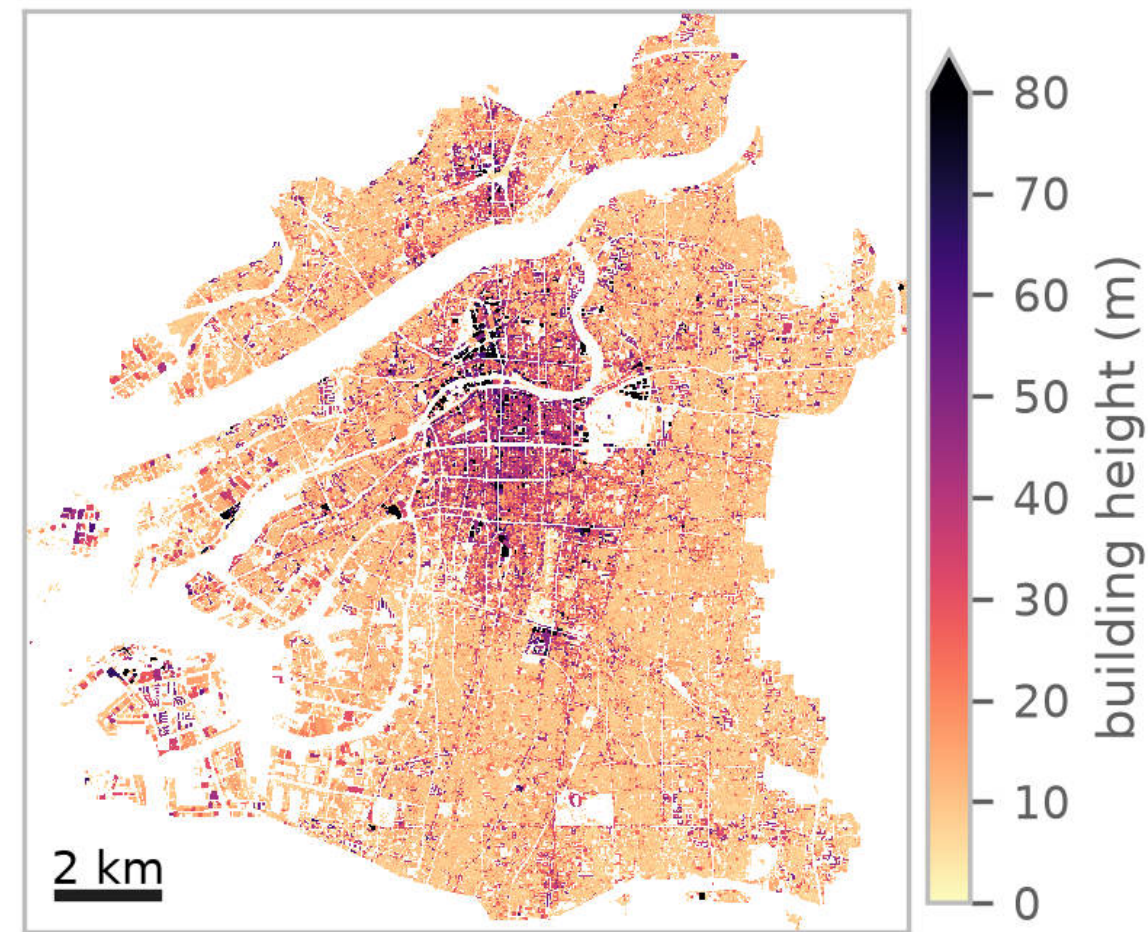


Upwind shelter is the input the AI relies on most
90% of its attention lies within 150 m upwind
Remove a building: its jet disappears in the AI too

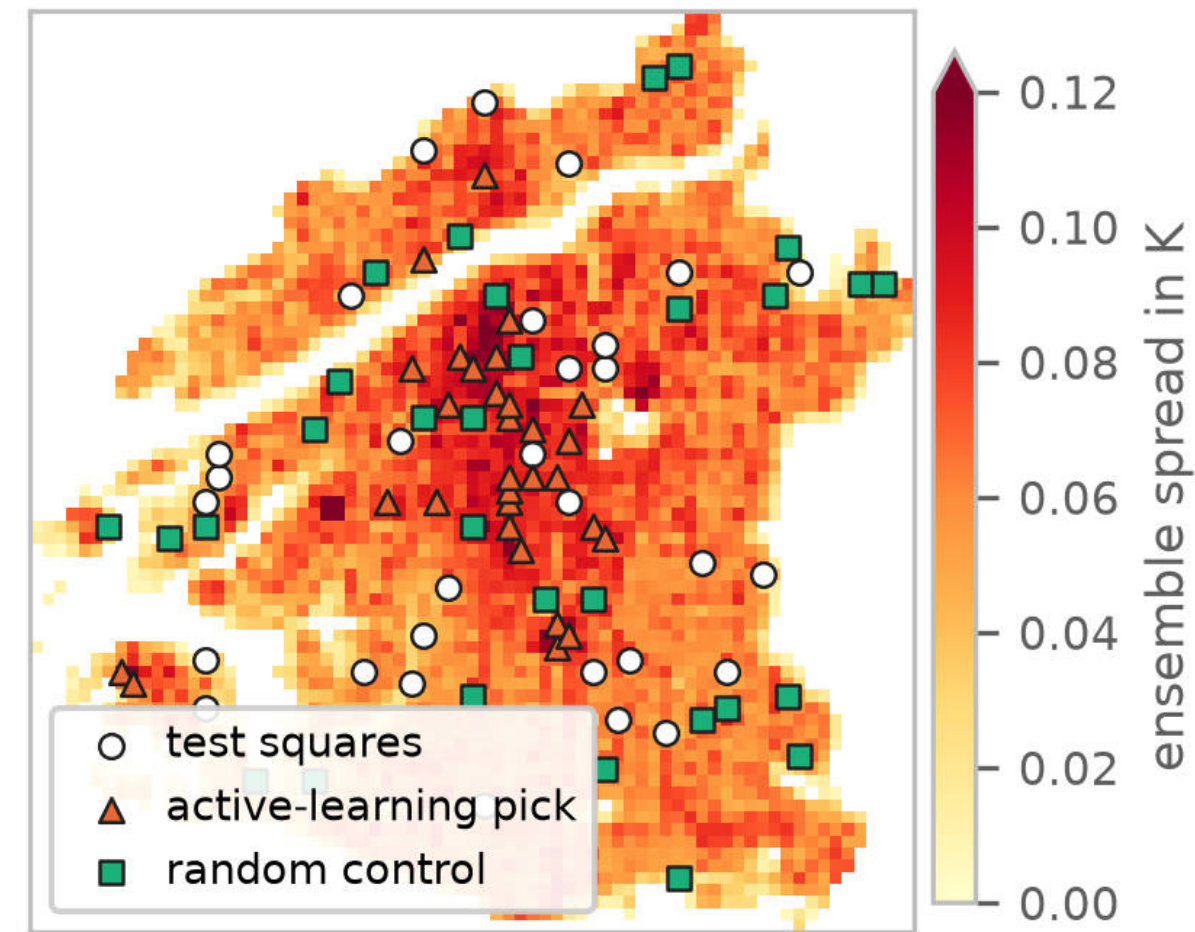


A new city: Osaka, without any retraining

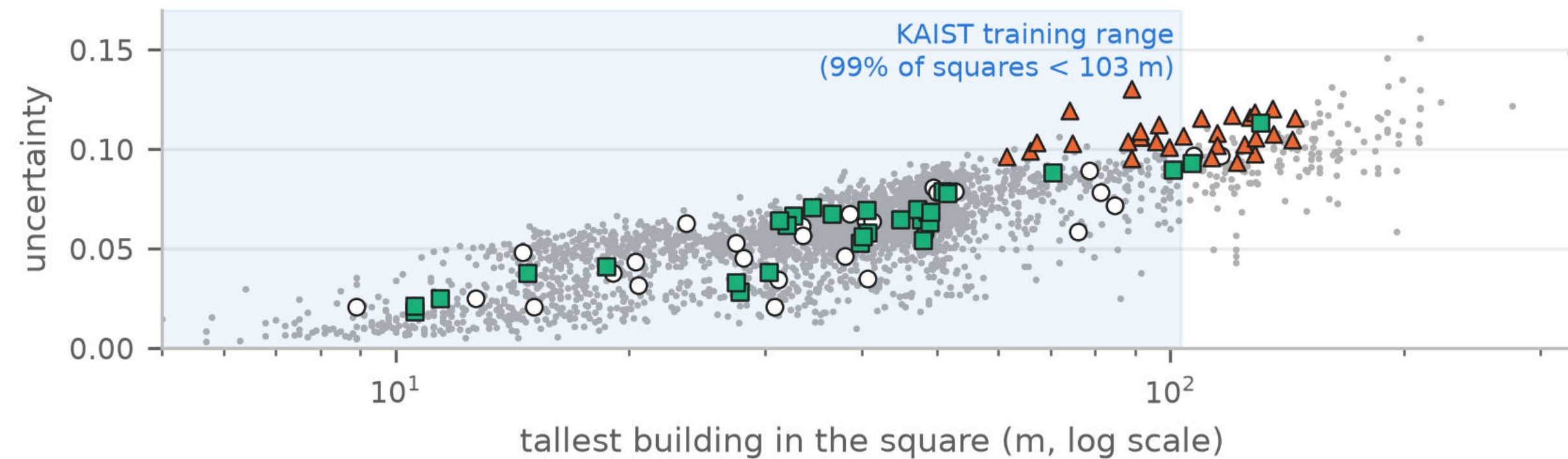
(a) Osaka city: 556,724 buildings (PLATEAU)



(b) zero-shot model uncertainty per square

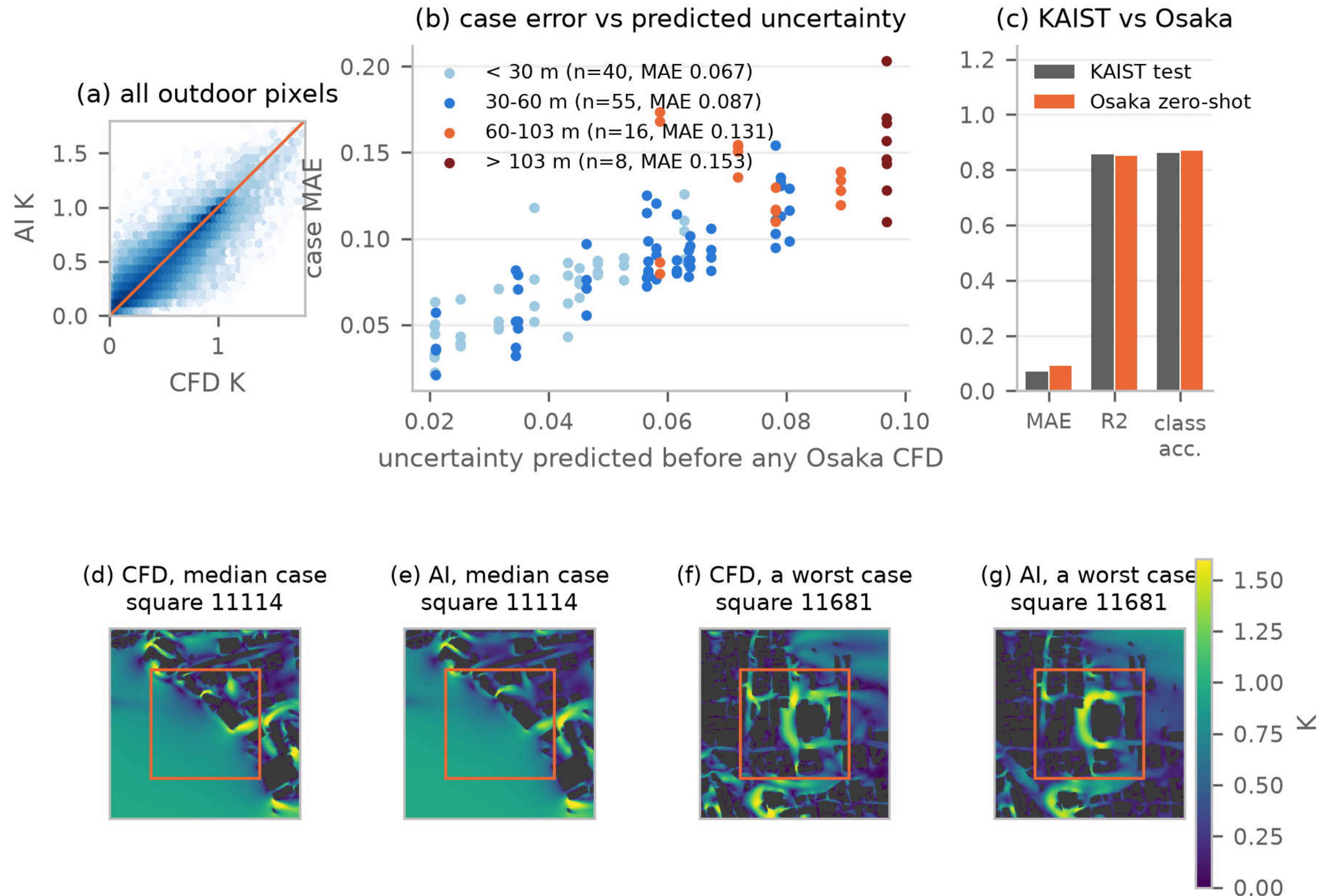


(c) the model is least sure where buildings are taller than anything it has seen



Mean error 0.086 vs 0.071; R^2 0.85 vs 0.86; comfort class 87% vs 86%.

The AI knows when it is unsure



0.87

correlation between the AI's own uncertainty and its real error

0.067 → 0.153

error grows from low-rise to towers > 103 m

So the tool can warn its users: "low confidence here".