Sustainable IoT and Edge AI for Remote Monitoring

Public Lecture Series: Sustainability in Computer Science | Asst. Prof. Dr. Atakan Aral, University of Vienna

07.10.2024

universität wien

Environmental Monitoring

"A tool to assess environmental conditions and trends"

- *Policy development*
- *Reporting*
- *Real-time decision making*
  - *Internet of Things (IoT) and Machine Learning (ML)*

“Remote” Environmental Monitoring (REM)

1. Collecting data in geographically **remote** regions with limited energy availability and network connectivity
   - Water pollution
   - Forest fires
   - Avalanches, landslides, etc.
2. Using connected IoT systems to gain real-time insights **remotely**
   - Sensing equipment (sensors, IoT devices, MEMS,)
   - Communication equipment (sensor to gateway)
   - Processing equipment (from Raspberry Pis to massive DCs)

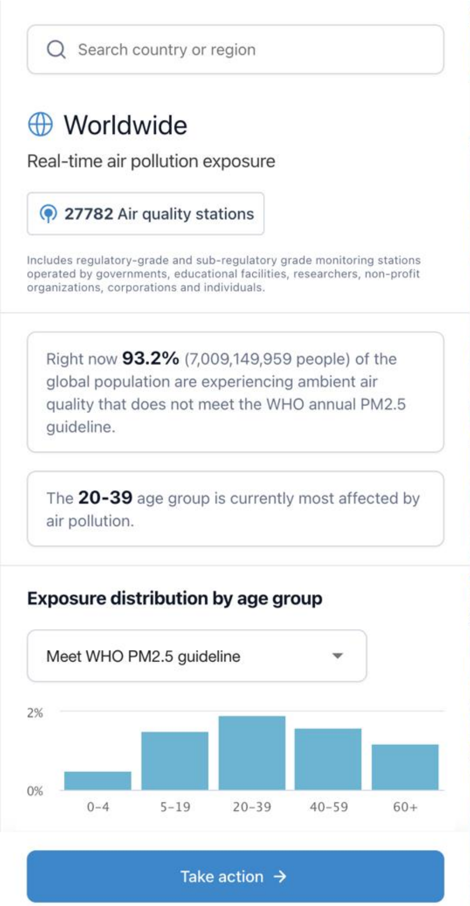
Examples of REM

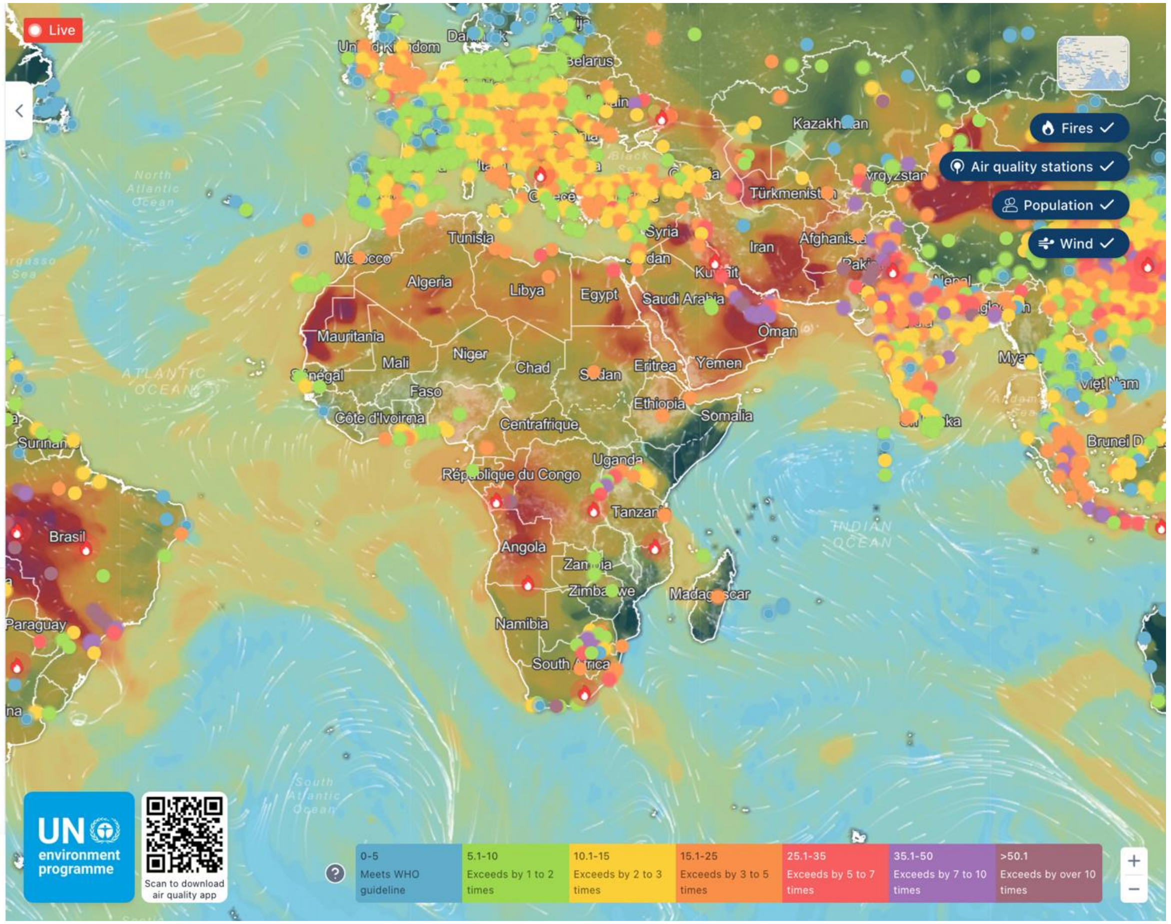
- Rural Environmental Monitoring Systems
  - Pollution Monitoring
    - Air Quality
    - Water Quality
  - Disaster Warning
    - Seismic Activity
    - Avalanche
    - Nuclear Explosion
  - Industrial IoT
    - Agriculture
    - Oil and Gas Well

GEMS/Air

- Part of UNEP
- 27 782 air quality stations
- Hourly measurements

- https://www.iqair.com/unep

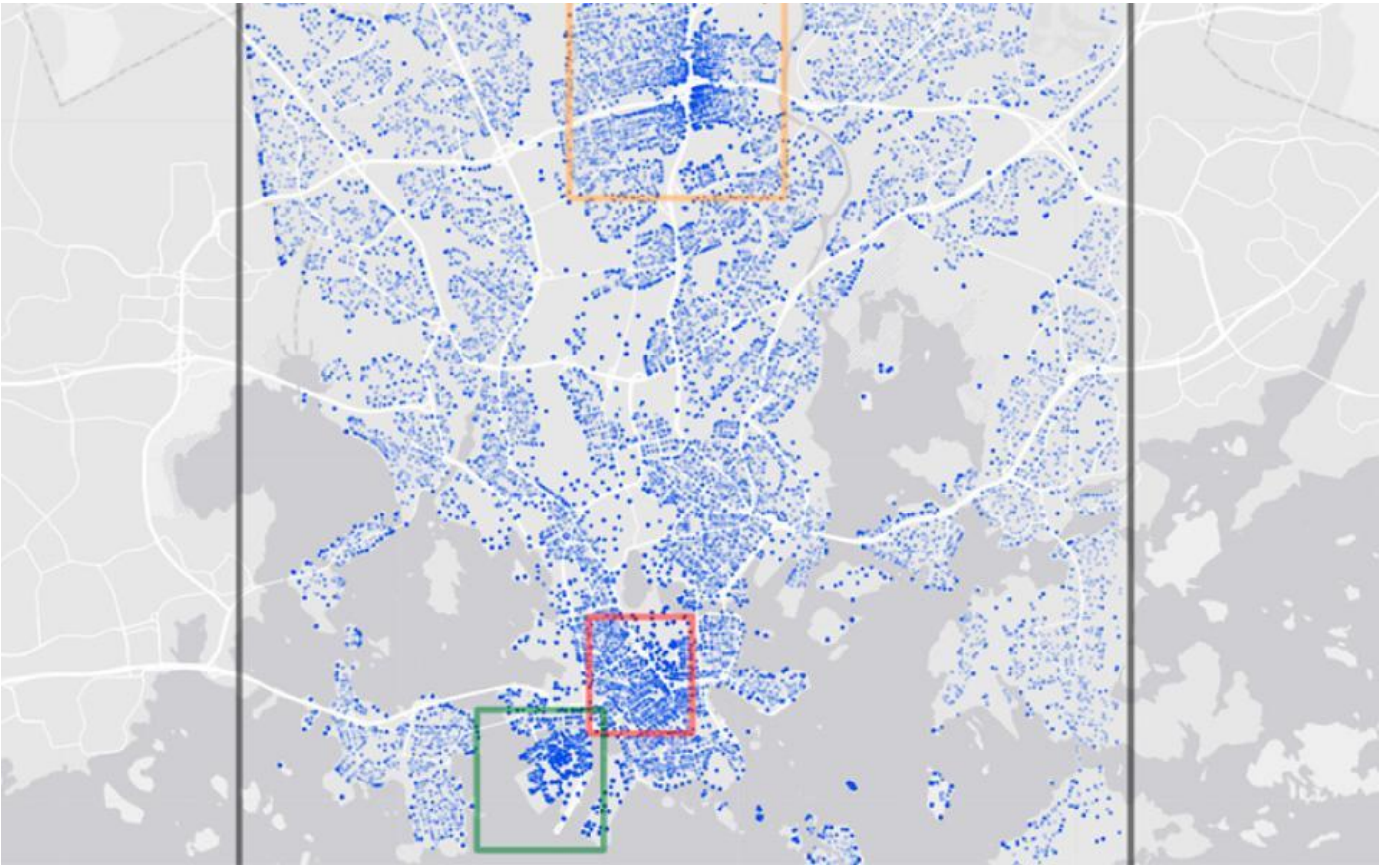




universität wien

MEGASENSE

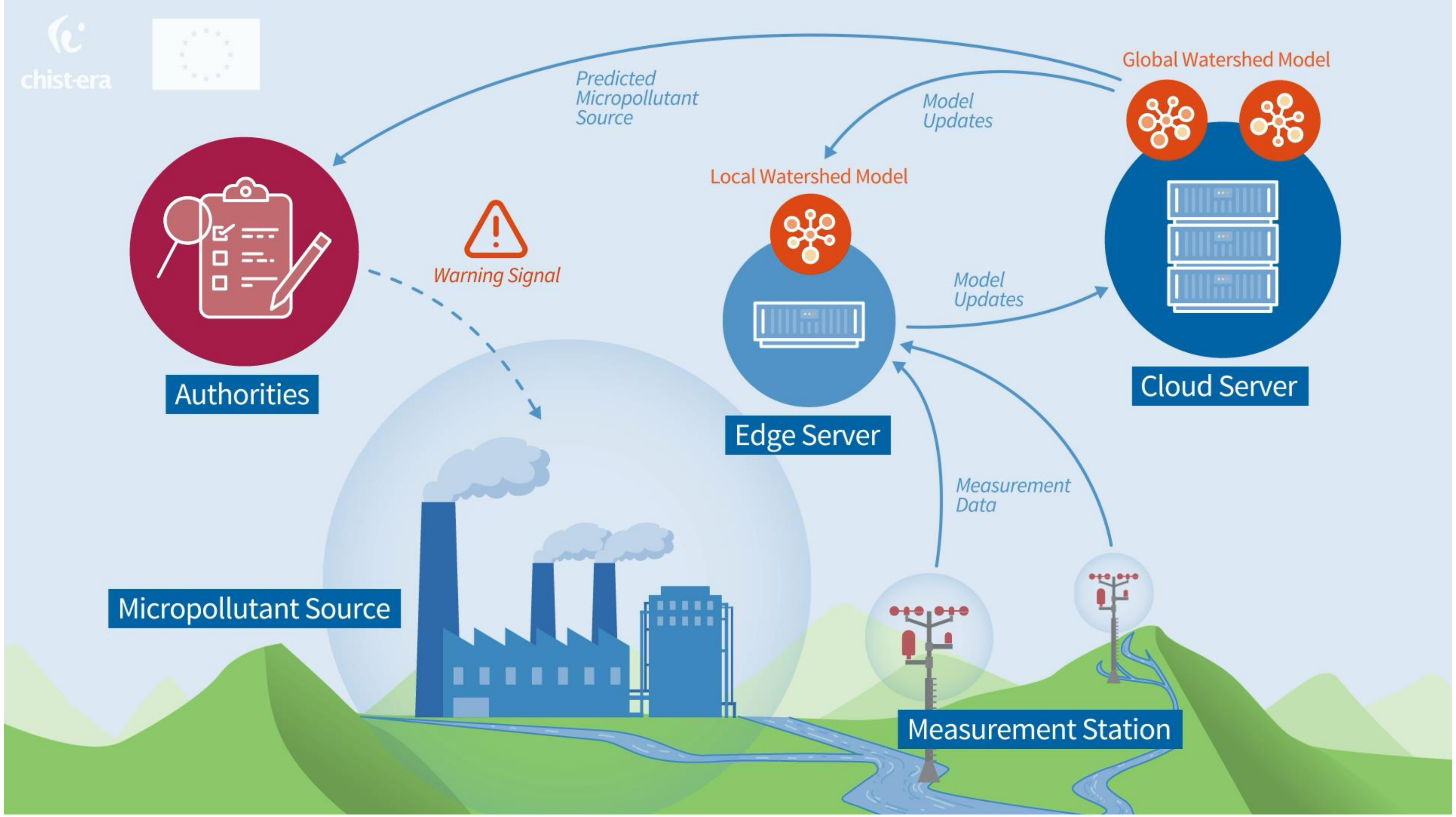
- Scalable real-time 5G air pollution sensing as a service for megacities
- Use ML to calibrate many low-cost sensors (e.g., wearables) with a few highly accurate measurement stations.
- https://helsinki.fi/en/researchgroups/sensing-and-analytics-of-air-quality



SWAIN

- Up to 100 sensors
- Two prototype deployments
- Sub-minute measurements
- Feedback loop

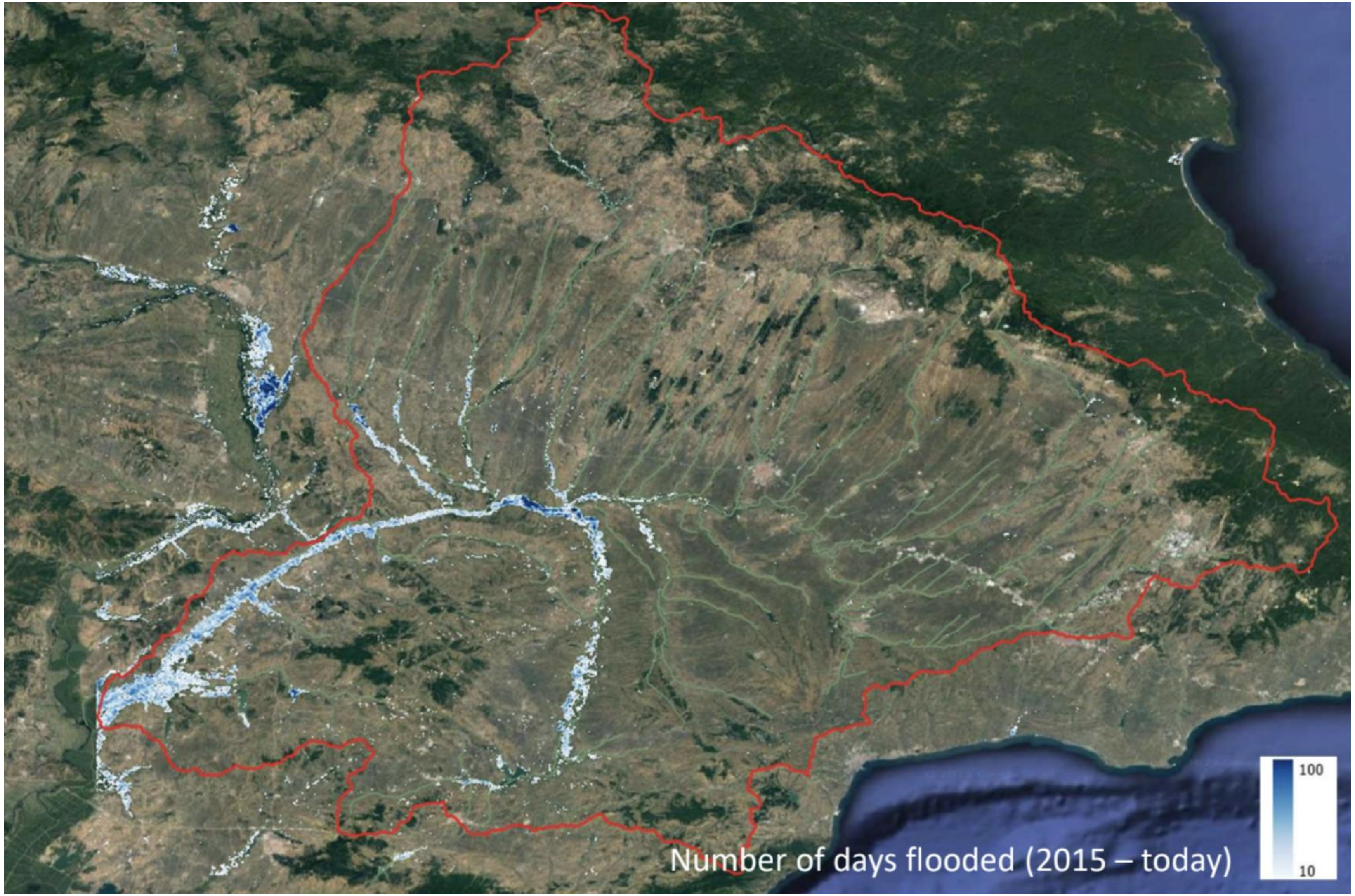
- https://swain-project.eu/



WATERLINE

- Remote sensed data
- Historical data
- In-situ data
- Crowdsourced data

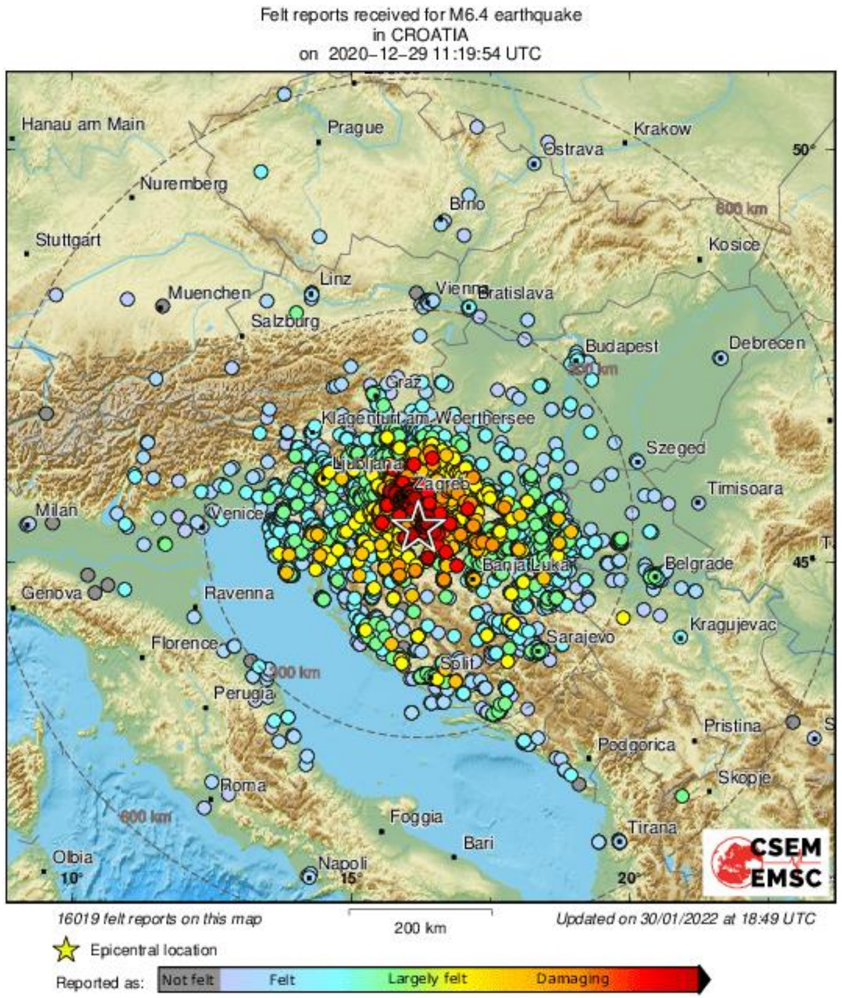
- https://waterlineproject.eu/

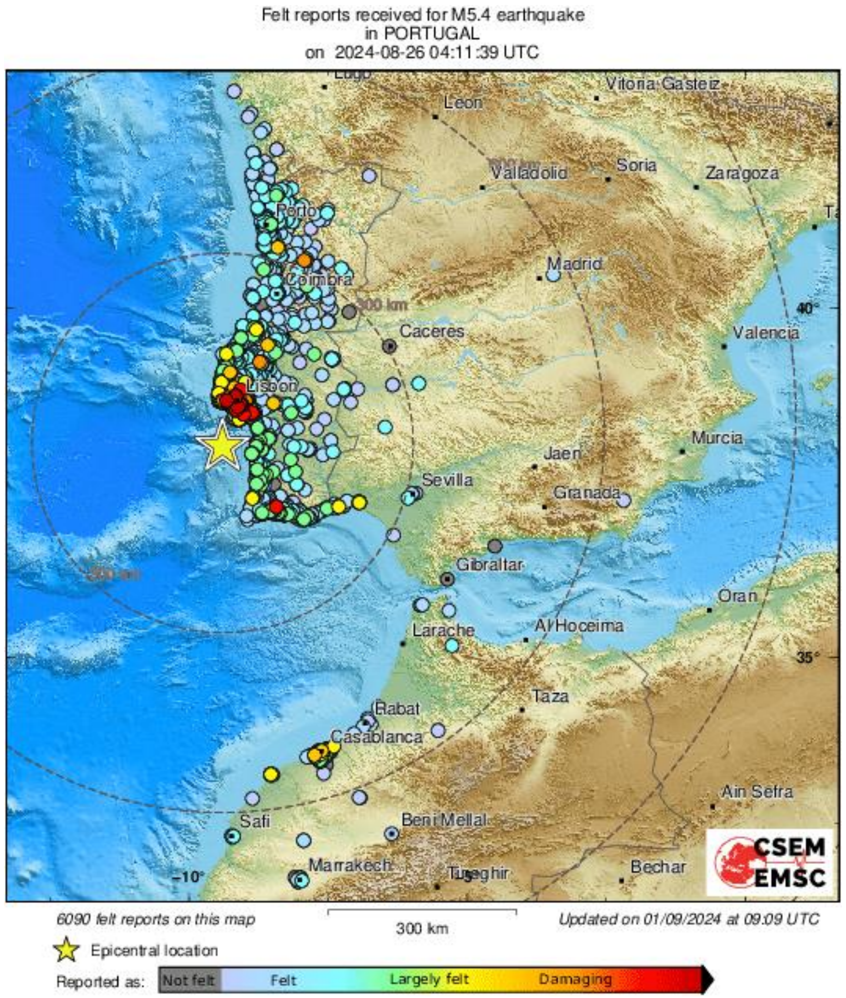


EMSC

- More than 2500 sensors
- Sub-minute measurements
- Sensors deployed in urban areas

- https://emsc.eu/

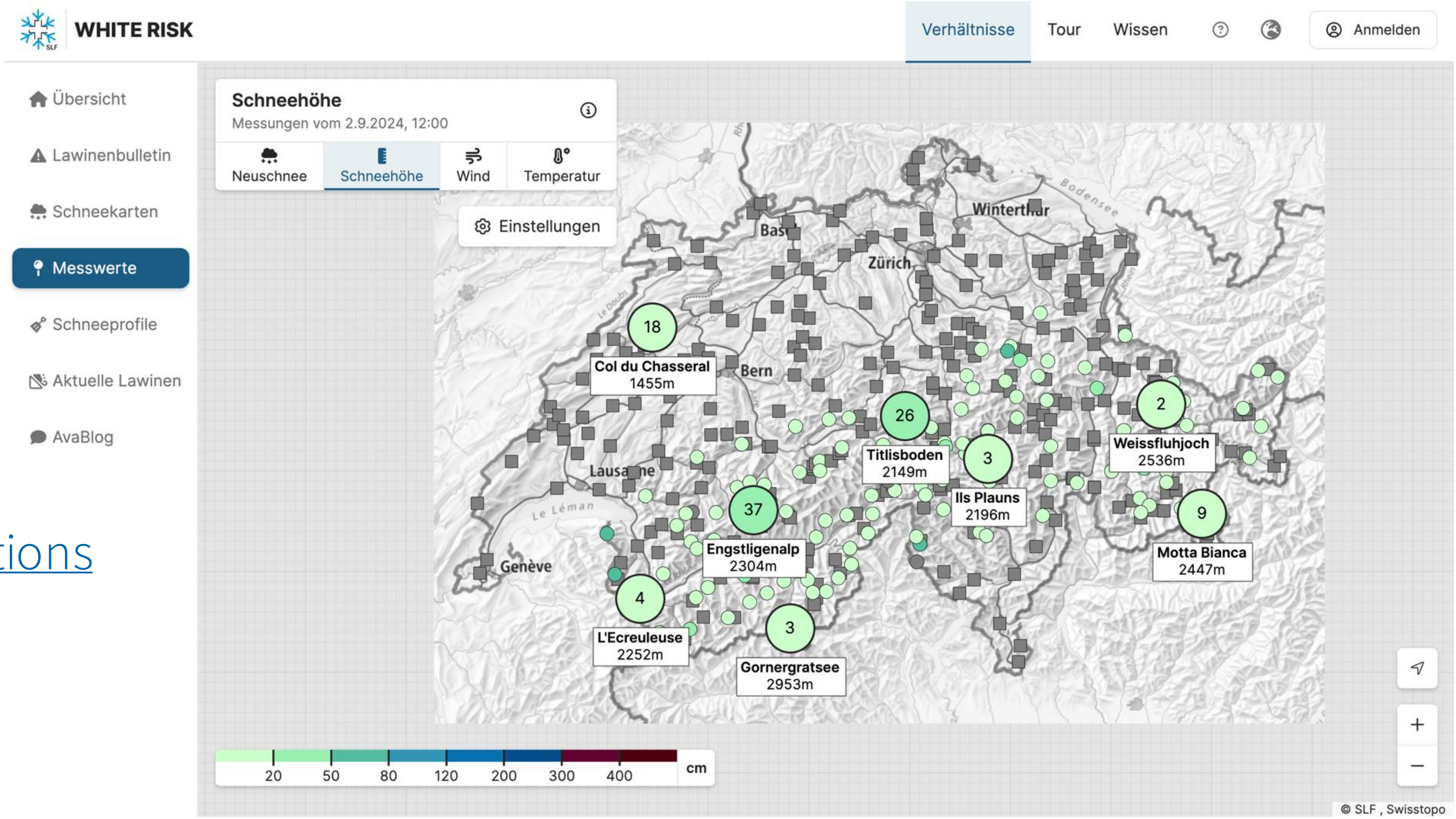




SLF IMIS

- 189 stations in the Swiss Alps and Jura Canton
- Highly remote areas
- Every 30 minutes

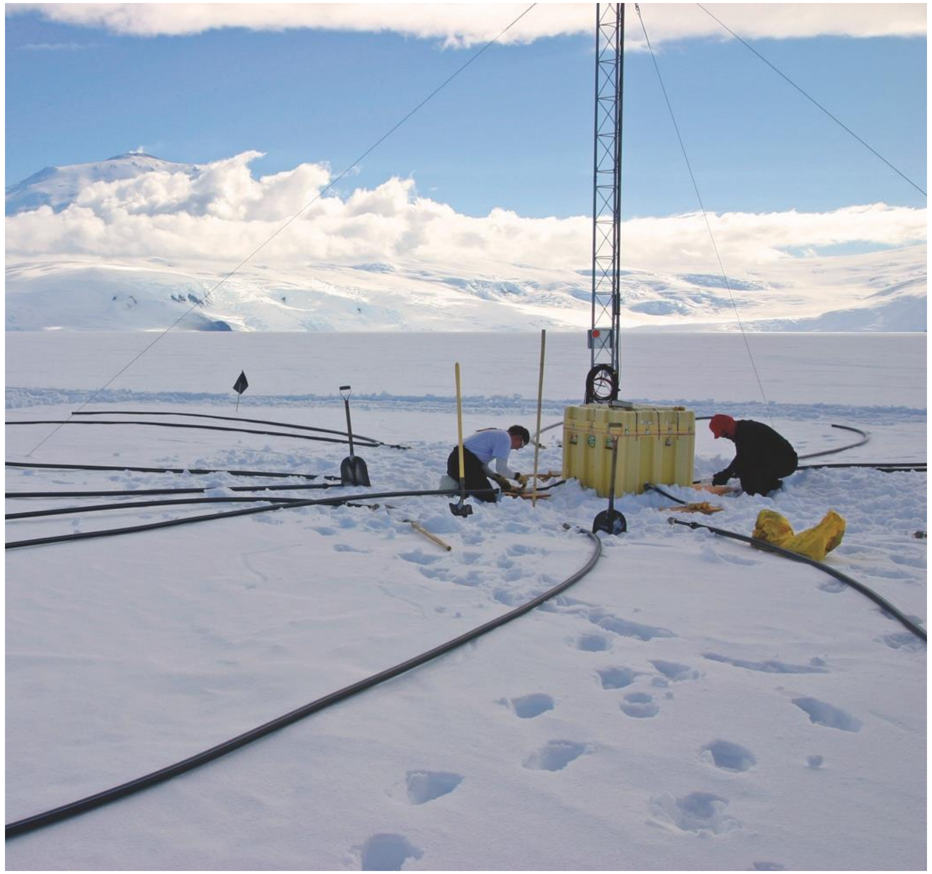
- https://whiterisk.ch/en/conditions



CTBTO

- 337 facilities worldwide
- Hourly measurements
- Homogenously distributed around the earth

- https://ctbto.org/




337 Total Number of Facilities
Seismic Primary Array (PS)
Seismic Primary 3-Component Station (PS)
Seismic Auxiliary Array (AS)
Seismic Auxiliary 3-Component Station (AS)
Radionuclide Station (RN)
Radionuclide Station with Noble Gas Monitoring Capabilities (RN)
Radionuclide Laboratory (RL)
Hydroacoustic (Hydrophone) Station (HA)
Hydroacoustic (T-Phase) Station (HA)
Infrasound Station (IS)
International Data Centre - CTBTO - Vienna
Projection from the South Pole, representing accurately the size of Antarctica and the location of the stations in this region
The boundaries and presentation of material on this map do not imply the expression of any opinion on the part of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) Preparatory Commission concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries.
Revised August 2021 | CTBTO.ORG

Summary

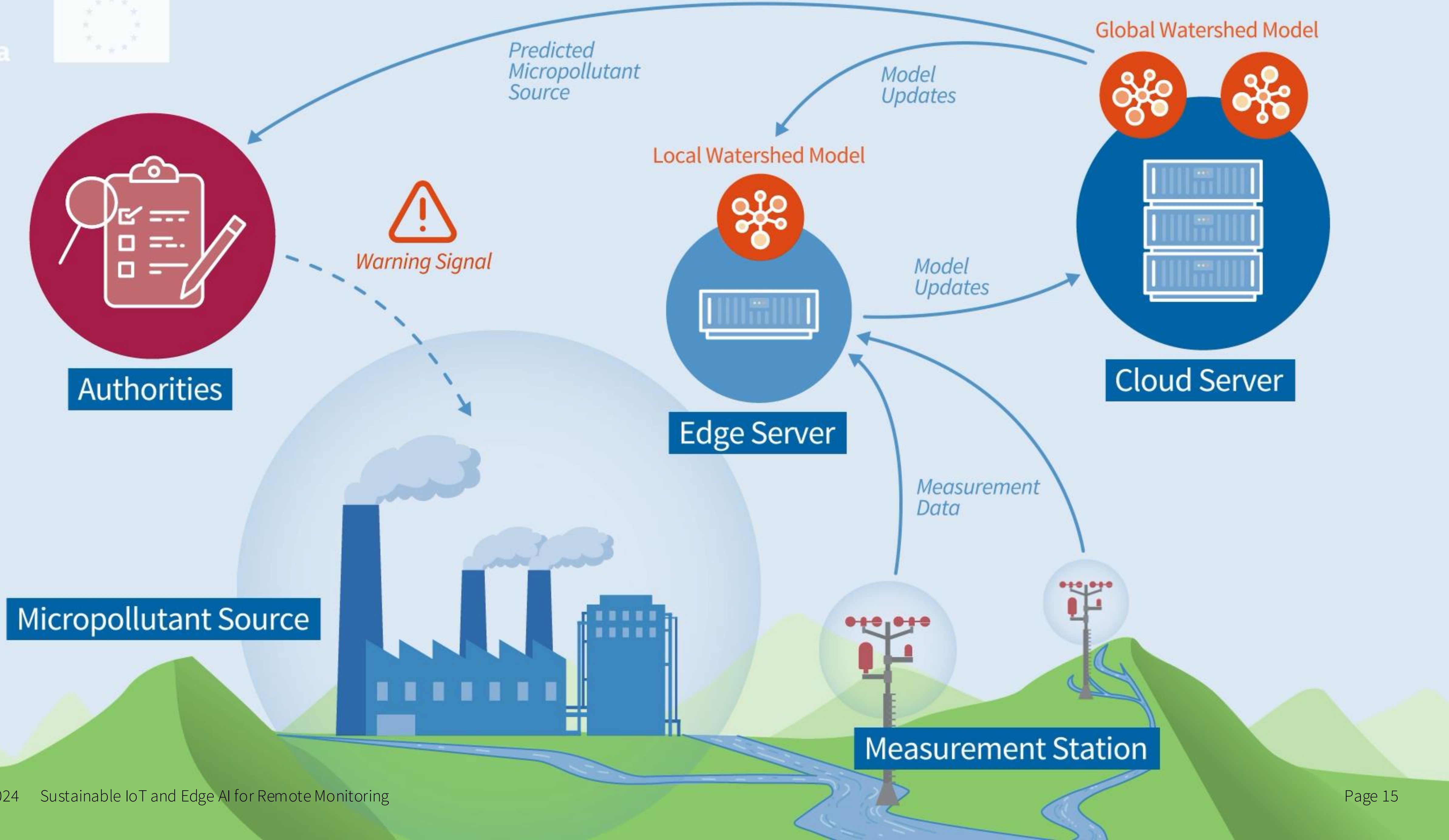
Rural Environmental Monitoring Use Case	Number of Stations	Dispersion	Real-Time Constraint	Proximity to Urban Areas	Potential for Electricity Access	Potential for Internet Access	Safety Risk	Data Sensitivity
Air Quality (GEMS/Air)	10s of 1000s	Global	Hour	Any	Moderate	Moderate	Moderate	Low
Water Quality (SWAIN)	30 to 75	Regional	Minute	Any	Low	Low	High	Low
Seismic Activity (EMSC)	$\geq$ 2500	Continental	Minute	Any	High	Moderate	High	Low
Avalanche (SLF IMIS)	186	Regional	Hour	Mid to Far	Low	Low	High	Low
Nuclear Explosion (CTBTO)	337	Global	Hour	Mid to Far	Low	Low	High	High
Agriculture	$\approx$ 1 per 2 ha	Local	Hour	Near to Mid	Low	Low	Moderate	Low
Oil and Gas Well	$\approx$ 1 per well	Local	Minute	Mid to Far	High	Low	High	High

Aral, A. 2024. The Promise of Neuromorphic Edge AI for Rural Environmental Monitoring. Environmental Data Science. Cambridge University Press. (to appear)

Sustaining the Monitors

- Monitoring is crucial for environmental sustainability
  - track environmental changes
  - inform or automate actions
  - provide better understanding

- How can we make sure that monitoring systems themselves are sustainable?


Global Watershed Model
Predicted Micropollutant Source
Model Updates
Local Watershed Model
Warning Signal
Authorities
Model Updates
Cloud Server
Edge Server
Measurement Data
Micropollutant Source
Measurement Station

Sustainability beyond Energy Efficiency

Challenges

- Energy / network availability
- Battery life and disposal
- Hardware obsolescence and e-waste
- Physical disruption to ecosystems
- Carbon footprint of manufacturing and installation

Solutions

- Energy-harvesting
- Fewer sensors / sites
- Less frequent communication
- More efficient processing

universität wien

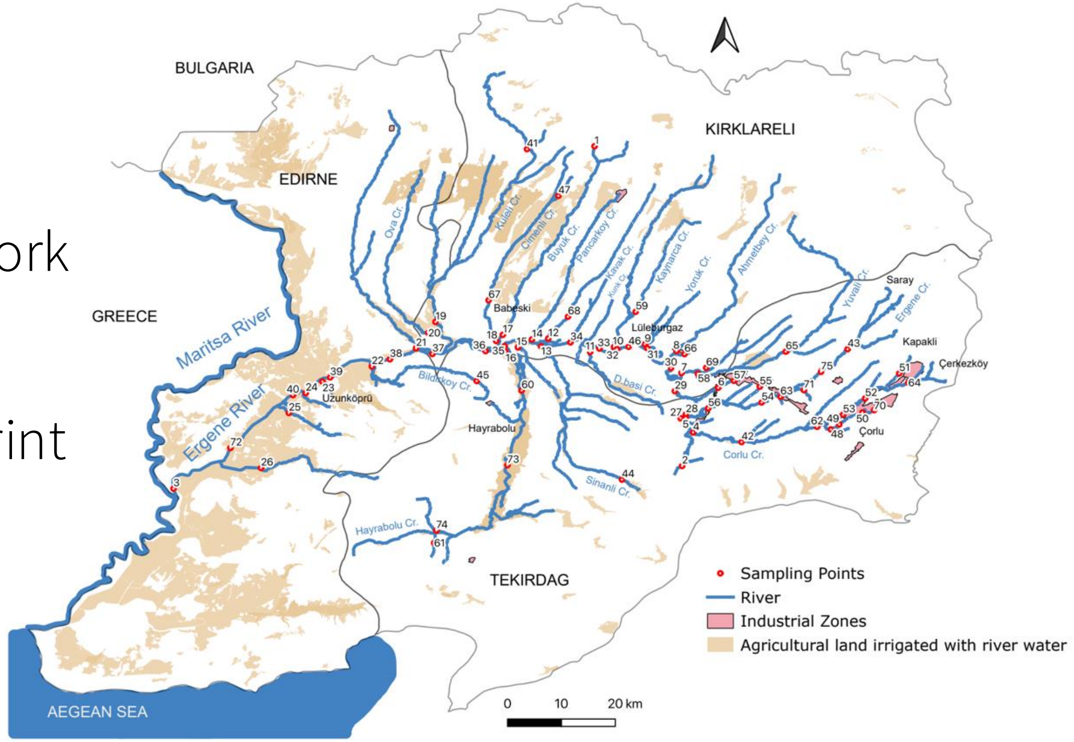
Optimizing the number of sensors / gateways

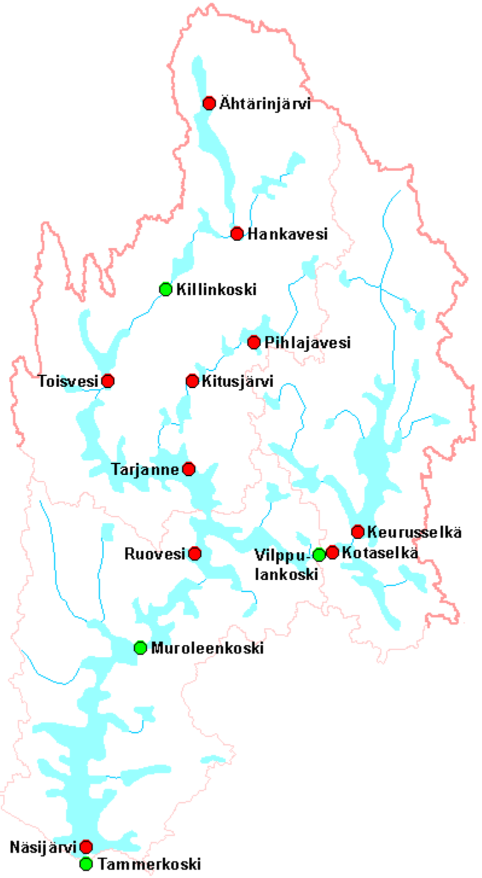
Ahmad, S., Uyanık, H., Ovatman, T., Sandıkkaya, M. T., De Maio, V., Brandić, I., & Aral, A. (2023). Sustainable environmental monitoring via energy and information efficient multi-node placement. IEEE Internet of Things Journal, 10(24).

SWAIN Project (03.2021 – 06.2024) funded ~1.2M EUR (FWF: ~410K EUR) through H2020 EIC CHIST-ERA

Spatial Planning

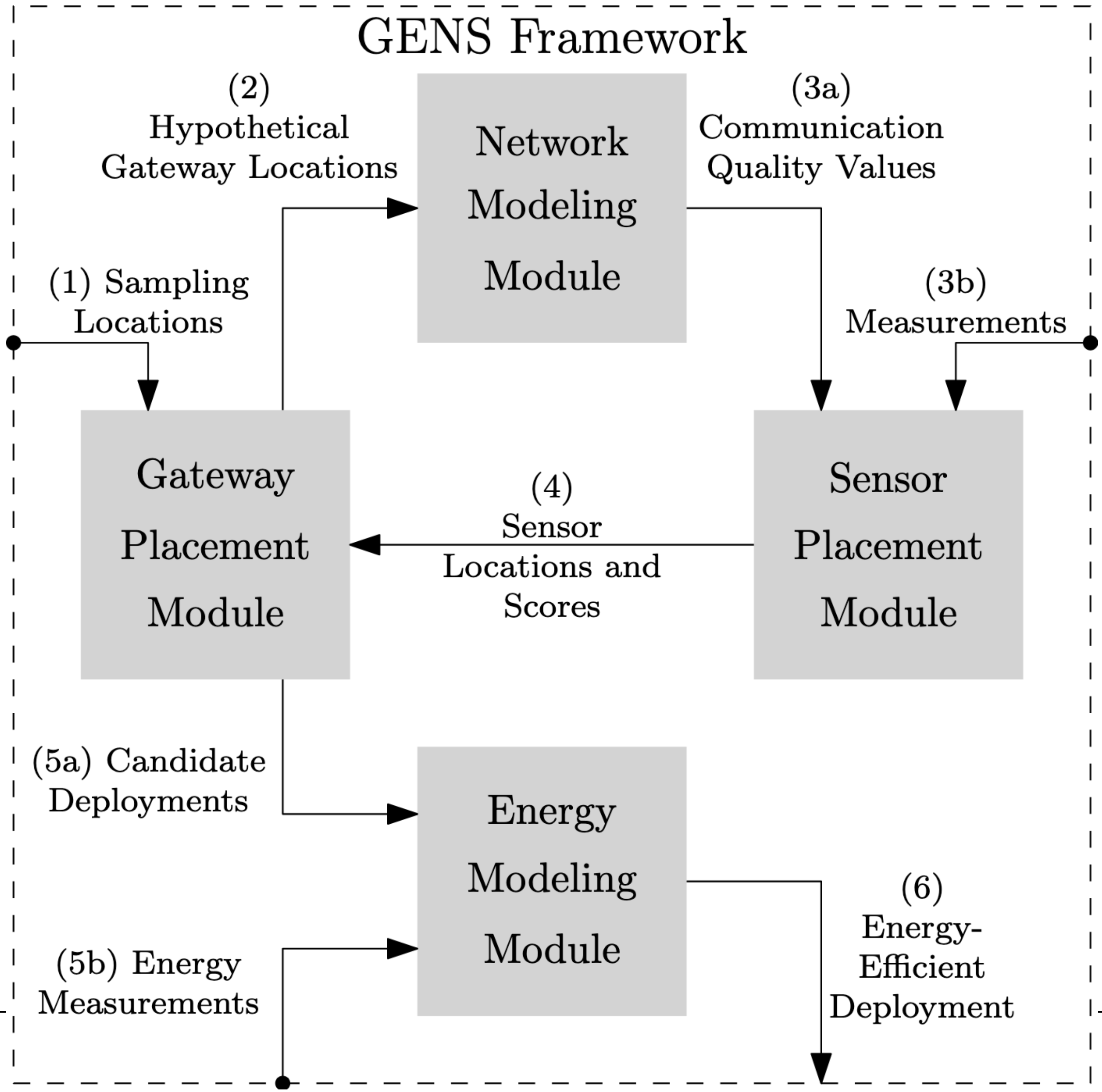
- Data quality / informativeness
- Reliability
- Maintainability
- Accessibility to electricity / network
- Cost
- Sustainability / ecological footprint





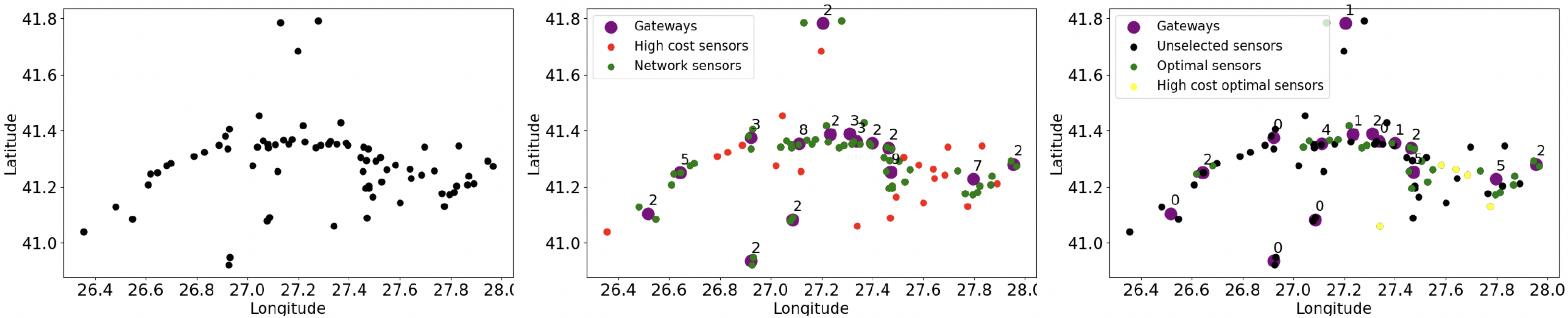
GENS Framework

- Starts from an initial placement
- Deploys both sensors and gateways
- Generates multiple candidate solutions
  - Number of sites
  - Network accessibility
  - Data quality
  - Energy



universität wien

Ahmad, S., Uyanık, H., Ovatman, T., Sandıkkaya, M. T., De Maio, V., Brandić, I., & **Aral, A.** (2023). Sustainable environmental monitoring via energy and information efficient multi-node placement. *IEEE Internet of Things Journal*, 10(24).



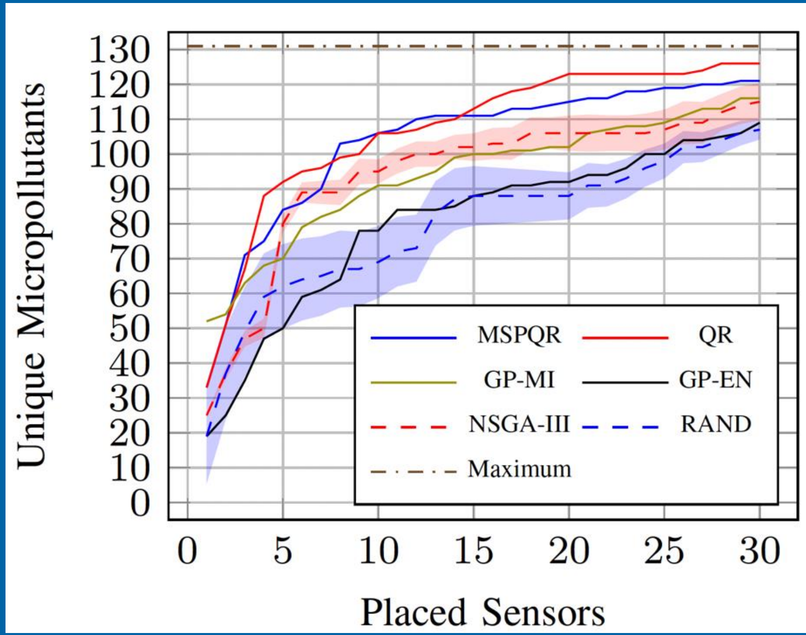
Ahmad, S., Uyanık, H., Ovatman, T., Sandıkkaya, M. T., De Maio, V., Brandić, I., & **Aral, A.** (2023). Sustainable environmental monitoring via energy and information efficient multi-node placement. *IEEE Internet of Things Journal*, 10(24).

60% to 73% fewer IoT nodes deployed

Almost the same number of detected pollutants

Up to 60% energy savings

The first comprehensive evaluation of energy consumption for multi-type IoT nodes

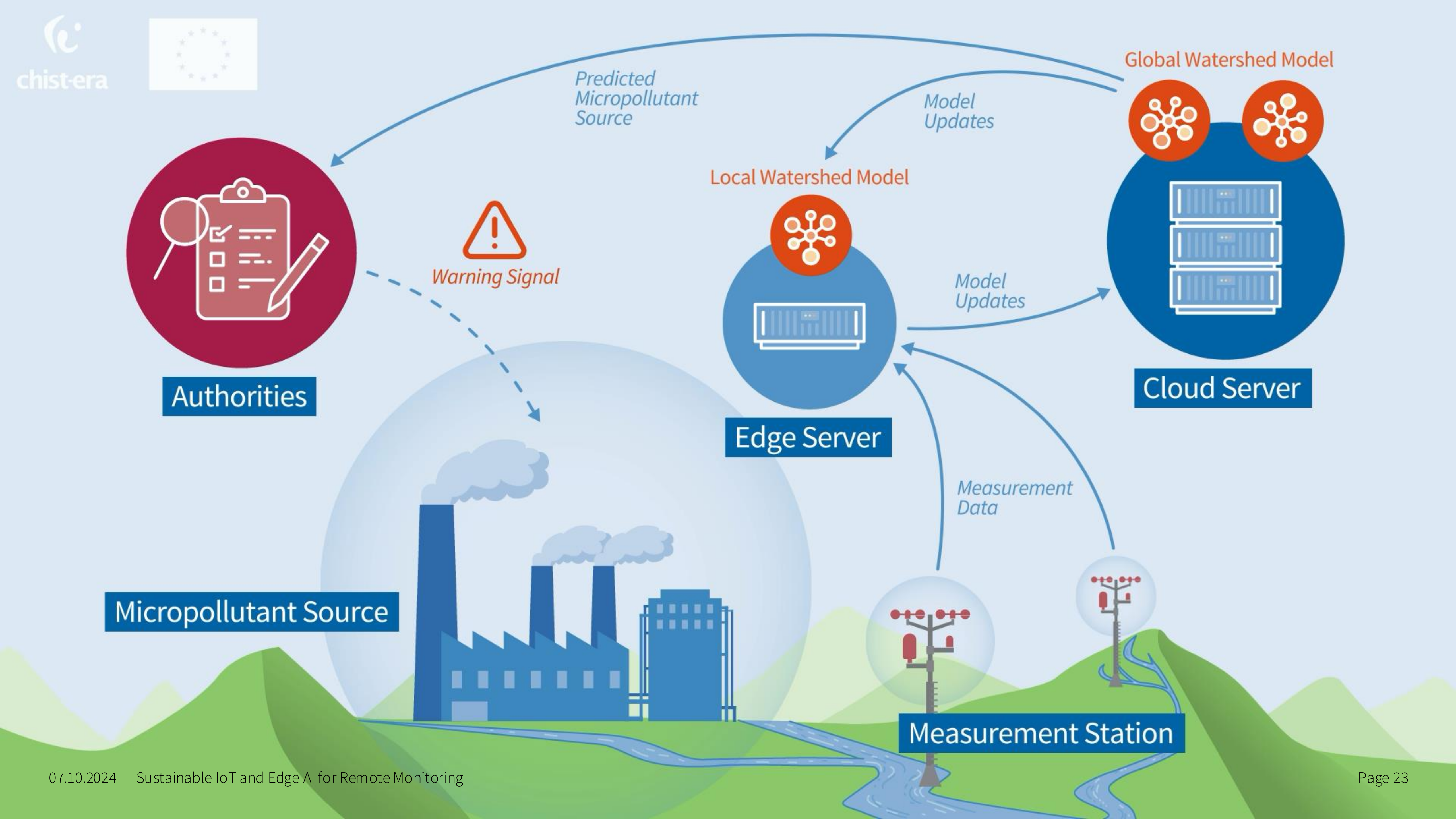


universität wien

Optimizing the communication frequency

Aral, A., Erol-Kantarci, M., & Brandić, I. (2020). Staleness control for edge data analytics. *SIGMETRICS / Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(2), 1-24.

Aral, A., & Brandic, I. (2018). Consistency of the fittest: Towards dynamic staleness control for edge data analytics. *In Euro-Par 2018* (pp. 40-52).



Global Watershed Model
Predicted Micropollutant Source
Model Updates
Local Watershed Model
Warning Signal
Model Updates
Authorities
Cloud Server
Edge Server
Measurement Data
Micropollutant Source
Measurement Station

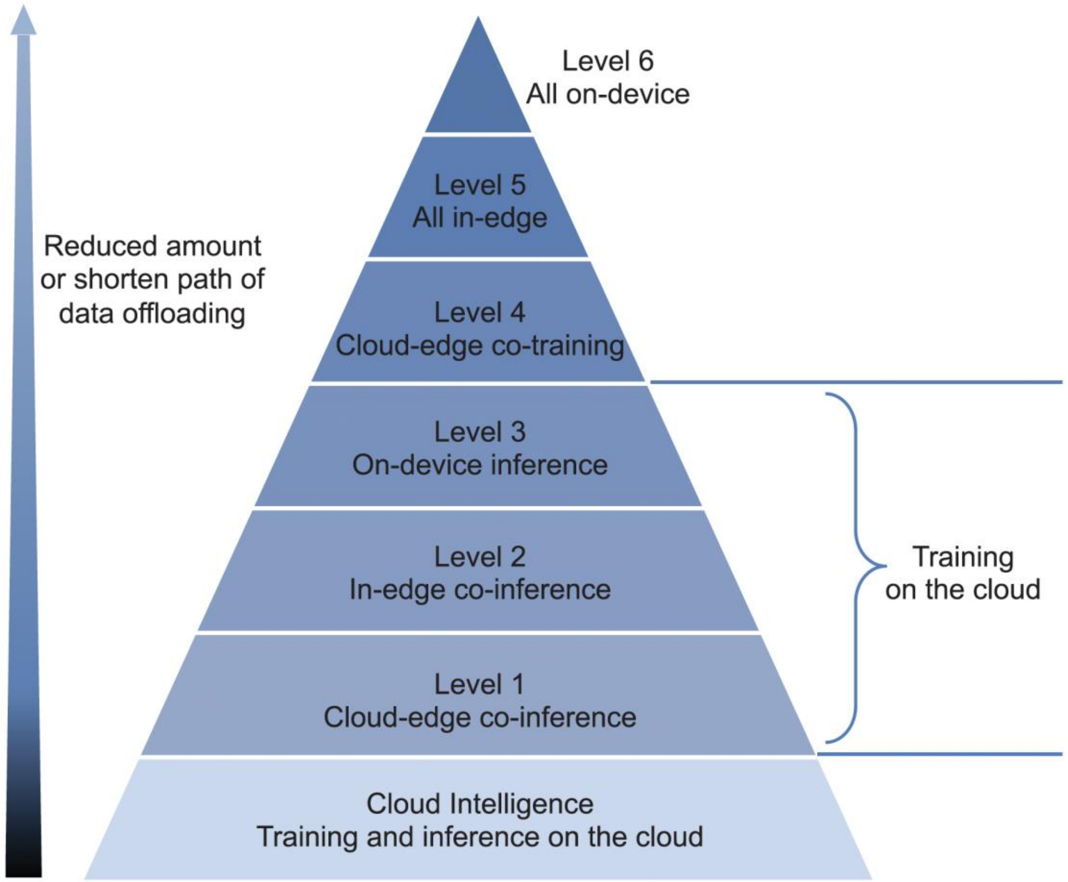
universität wien

Edge Intelligence (Edge AI)

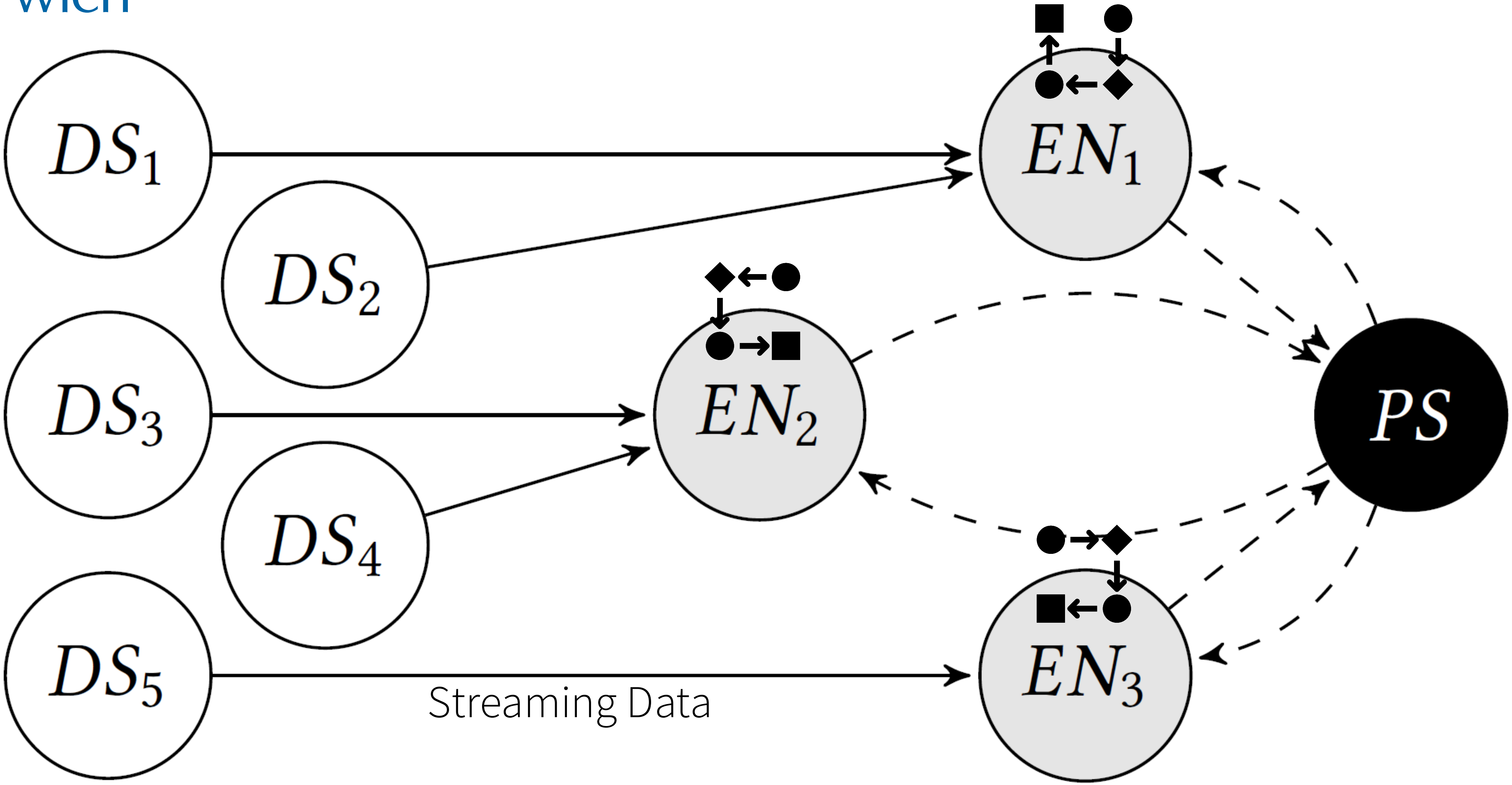
"edge computing with machine learning and advanced networking capabilities" —IEC

"Instead of entirely relying on the cloud, Edge AI makes the most of the widespread edge resources to gain AI insight." —Zhou et al. (2019)

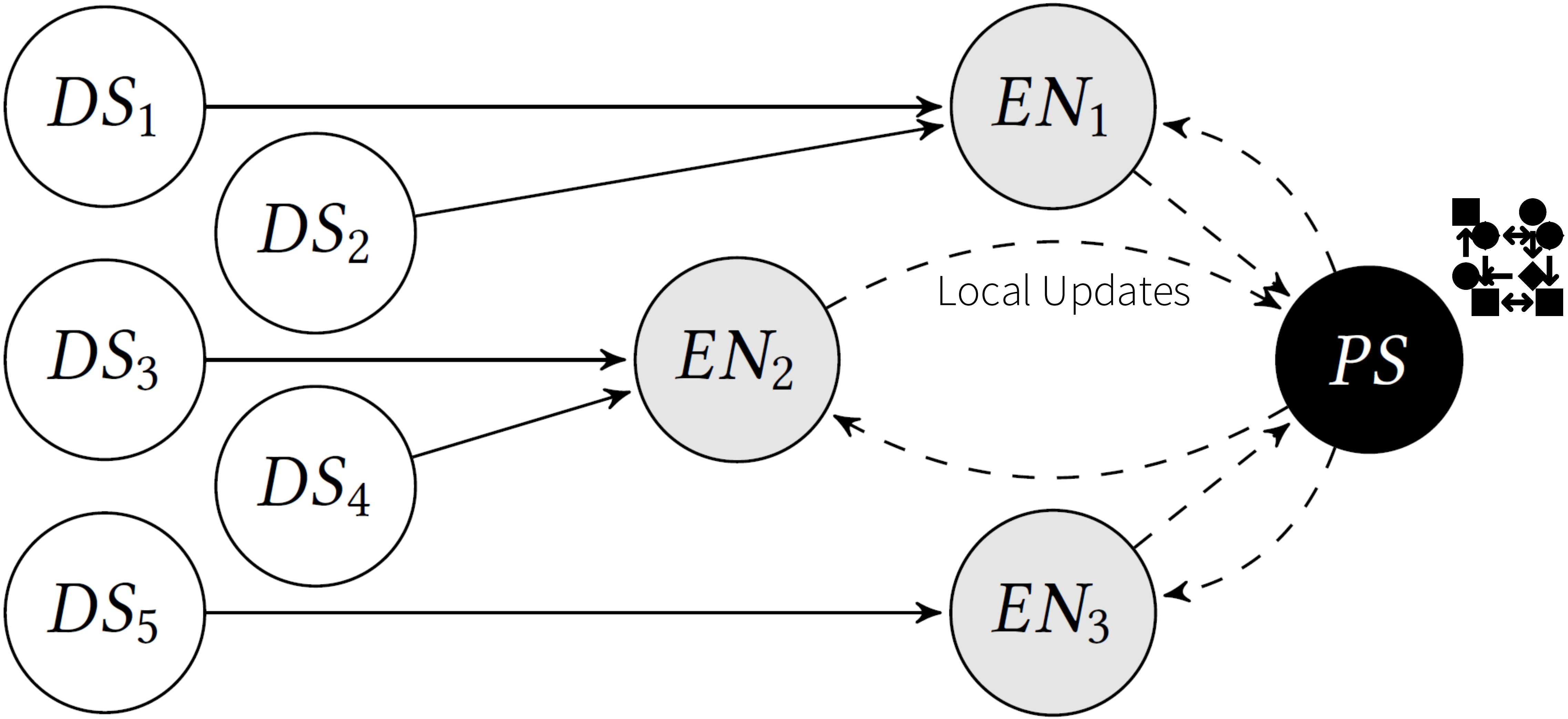
- Edge servers are located much closer to the data sources compared to cloud
  - Higher data transmission rate
  - Lower computational power

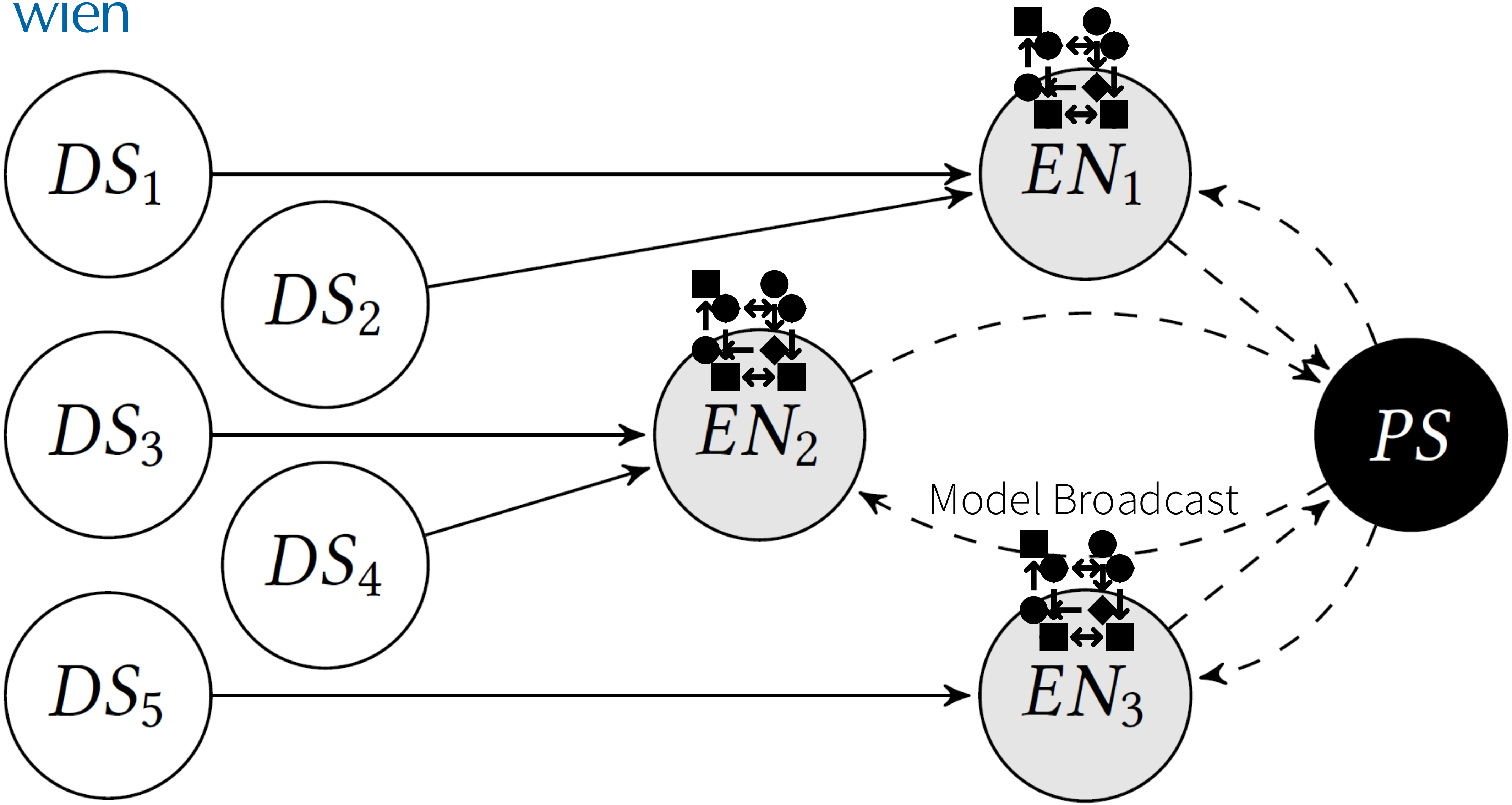


Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, *107*(8), 1738-1762.

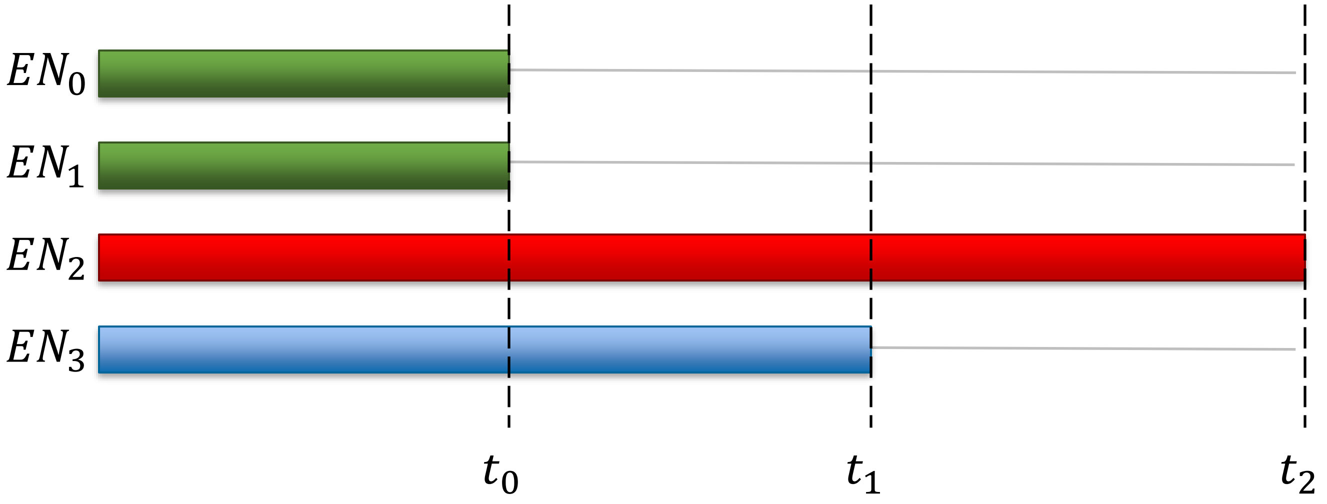


DS_1
DS_2
DS_3
DS_4
DS_5
EN_1
EN_2
EN_3
PS
Streaming Data

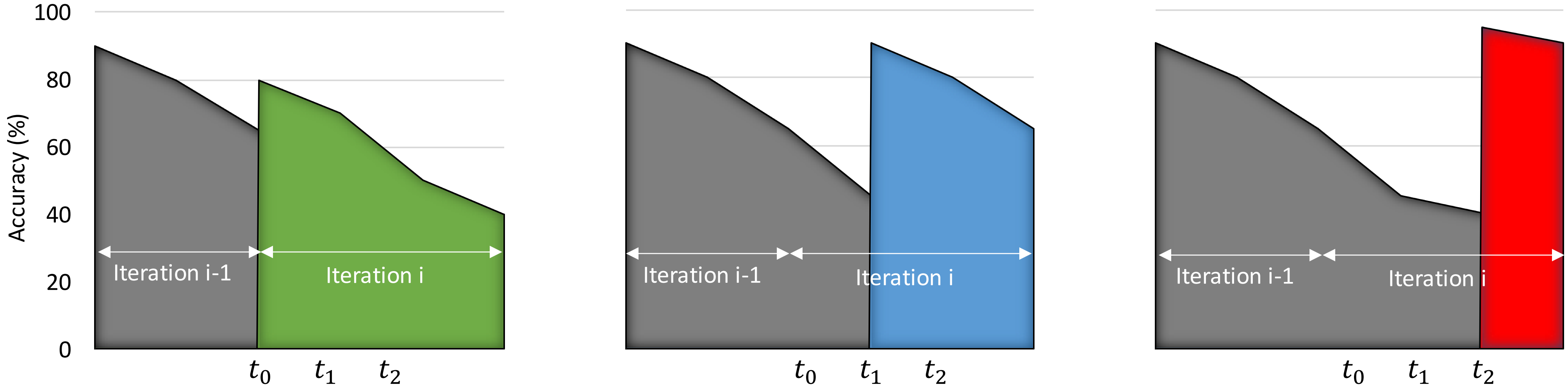

DS_1
DS_2
DS_3
DS_4
DS_5
EN_1
EN_2
EN_3
PS
Local Updates



DS_1
DS_2
DS_3
DS_4
DS_5
EN_1
EN_2
EN_3
PS
Model Broadcast



* Update arrival times to the PS

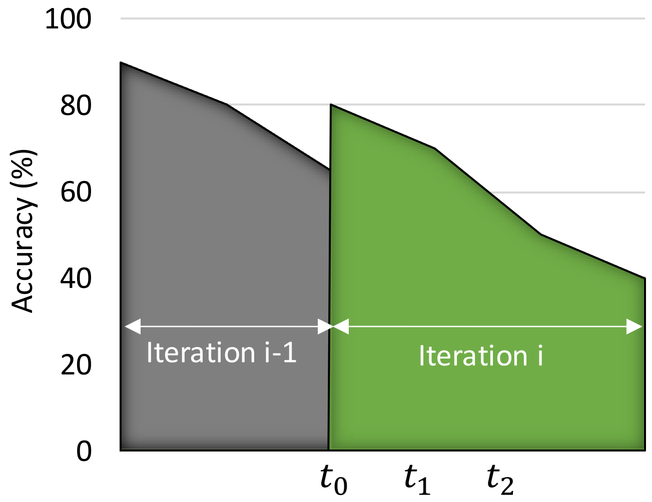


Option 1 Option 2 Option 3

* Accuracy of local prediction models

Staleness Control Problem

$$\begin{aligned}\underset{j}{\text{maximize}} \quad & \int_{t_0}^{t_j} \mathcal{U}_{i-1}(x)\,dx + \int_{t_j}^{\tau_i} \mathcal{U}_i^j(x)\,dx \\ \text{subject to} \quad & i, j \in \mathbb{Z},\ 0 \le i,\ 0 \le j \le m_i.\end{aligned}$$



- **Hypothesis:** There exists an *inconstant* point to broadcast within each iteration that yields the optimum accuracy.

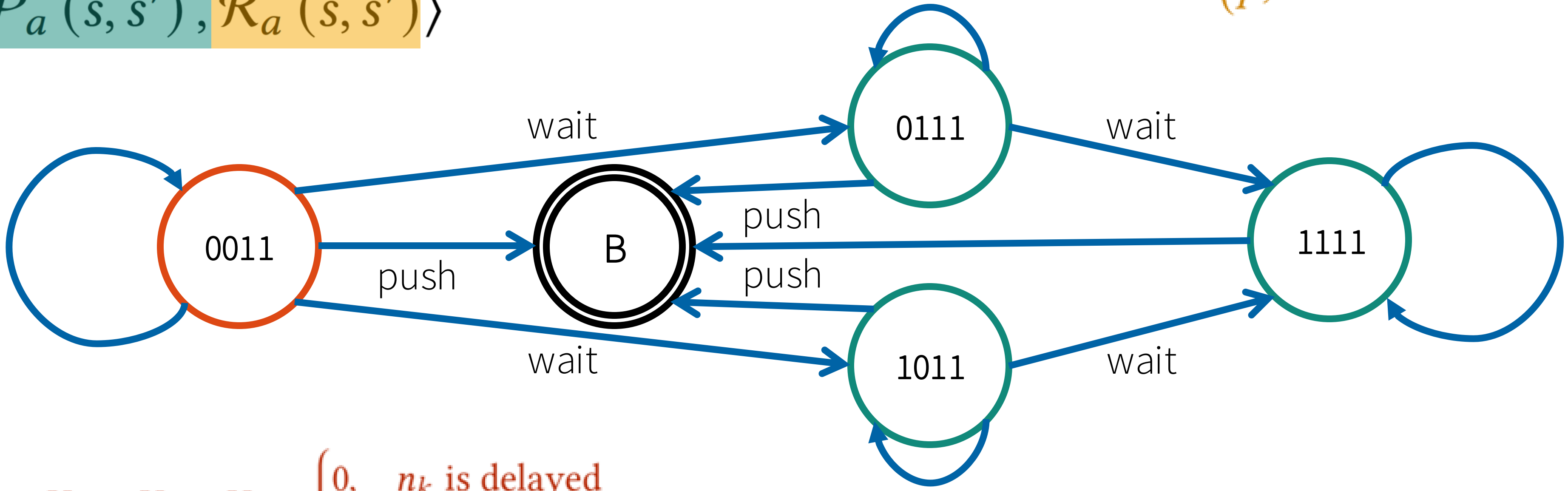
Aral, A., Erol-Kantarci, M., & Brandić, I. (2020). Staleness control for edge data analytics. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(2), 1-24.

Staleness Control Problem

Markov Decision Process:

$$\langle \mathcal{S}, \mathcal{A}, \mathcal{P}_a(s, s'), \mathcal{R}_a(s, s') \rangle$$

$$\mathcal{R}_a(s_t, s_{t+1}) = \begin{cases} \mathcal{U}(s_t), & a = \text{push} \\ 0, & a = \text{wait} \wedge s_{t+1} \neq s_t \\ p, & a = \text{wait} \wedge s_{t+1} = s_t \end{cases}$$



$$s = V_{N-1}V_{N-2}\ldots V_k \ldots V_0, \quad V_k = \begin{cases} 0, & n_k \text{ is delayed} \\ 1, & \text{otherwise} \end{cases}$$

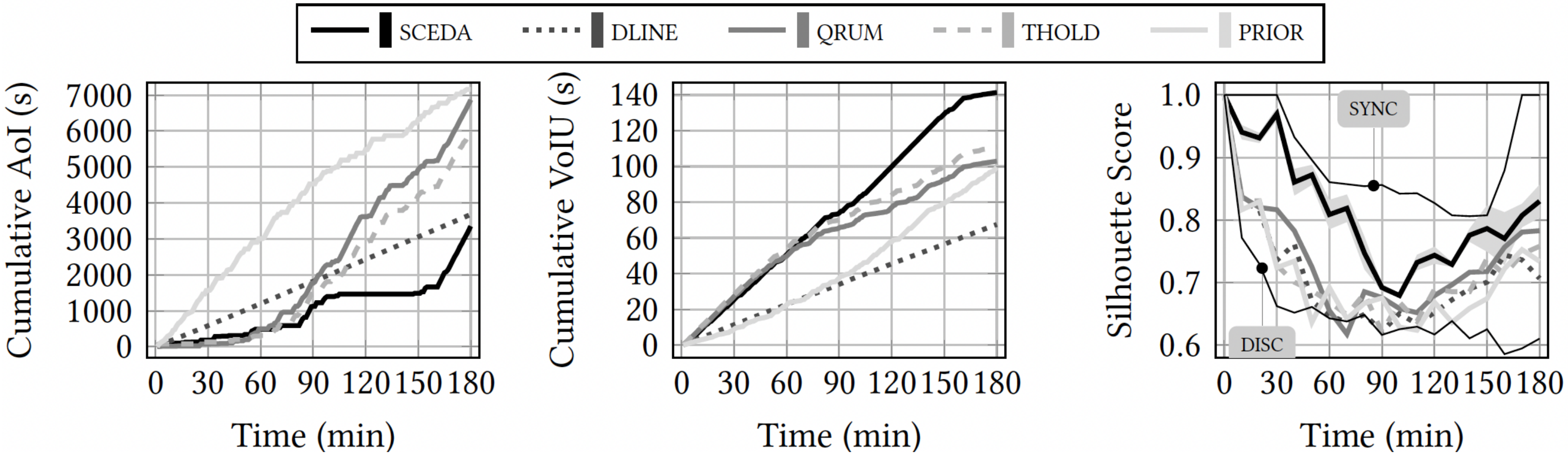
$$\mathcal{S}^+(s_t) = \{s_{t+1} : (d(s_t, s_{t+1}) = 1 \wedge s_{t+1} > s_t) \vee (s_{t+1} = s_t)\}$$

Staleness Control Problem

$$\alpha = \frac{1}{1 + \text{visits}\,(s_t, a_t)}$$

Online Delayed Q-Learning: $Q\,(s_t, a_t) \leftarrow (1-\alpha)\,Q\,(s_t, a_t) + \alpha \left(r_t + \gamma \sum_{s} \Pr\,(s \mid s_t, a_t) \max_{a} Q\,(s, a) \right)$

- Delay Q updates until a statistically significant number of experiences is reached.
- Theoretical bound on convergence (PAC-MDP)



Performance: Comparable accuracy to core AI, yet with real-time responsiveness (AoI < 10s)

Efficiency: Low communication overhead of model updates (< 1 per minute)

The first staleness control mechanism, where the synchronization period is learned from / adapted to the environmental changes

convergence (PAC-MDP)

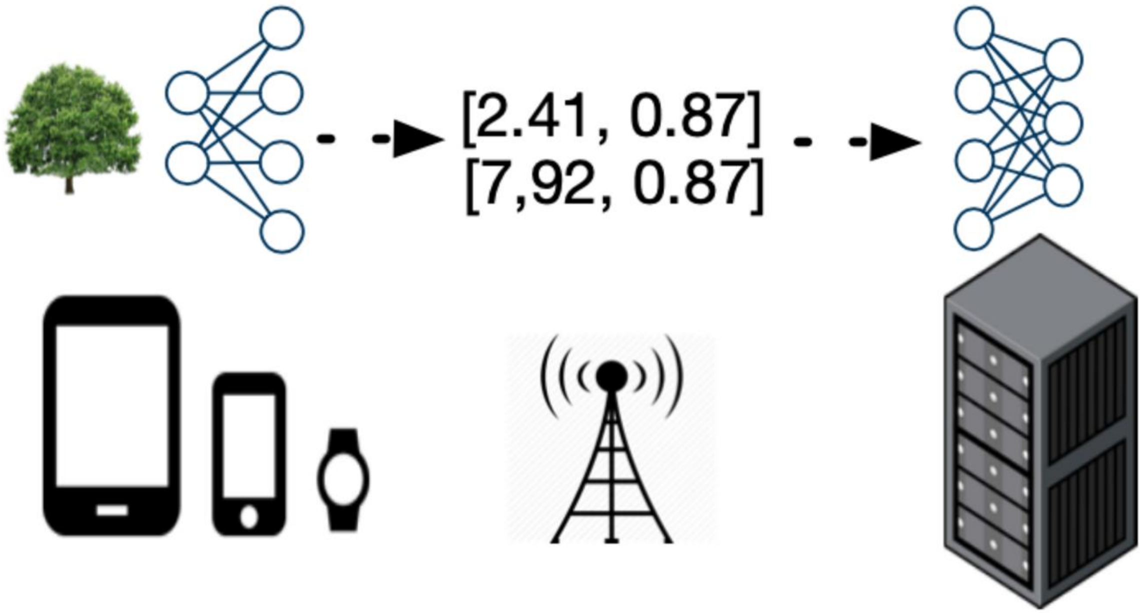
Time (min) Time (min) Time (min)

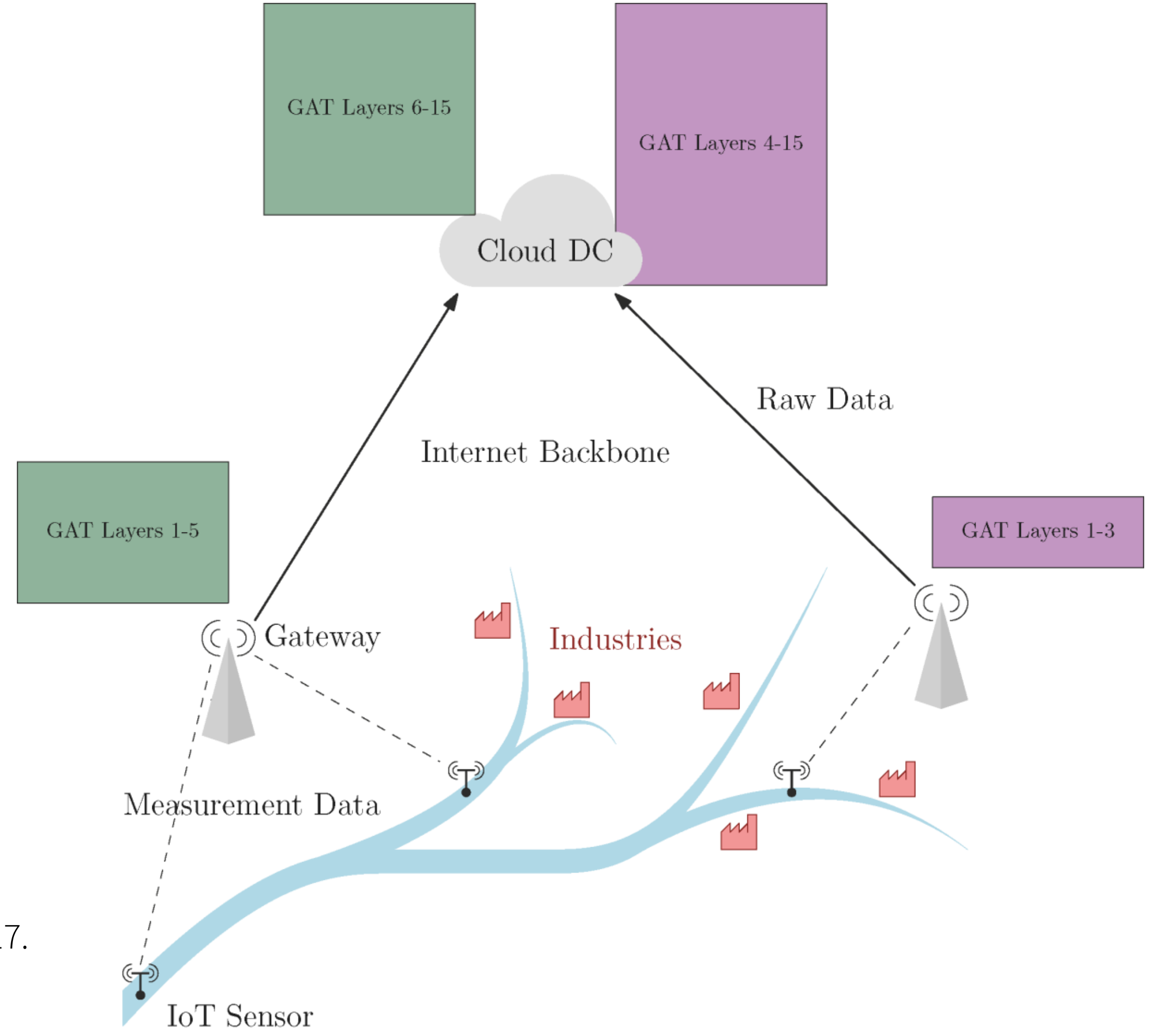
Optimizing the energy consumed

Aral, A., De Maio, V., & Brandic, I. (2021). Ares: Reliable and sustainable edge provisioning for wireless sensor networks. IEEE Transactions on Sustainable Computing, 7(4), 761-773.

Luger, D., **Aral, A.,** & Brandic, I. (2023). Cost-aware neural network splitting and dynamic rescheduling for edge intelligence. In Proceedings of the 6th International Workshop on Edge Systems, Analytics and Networking (pp. 42-47).

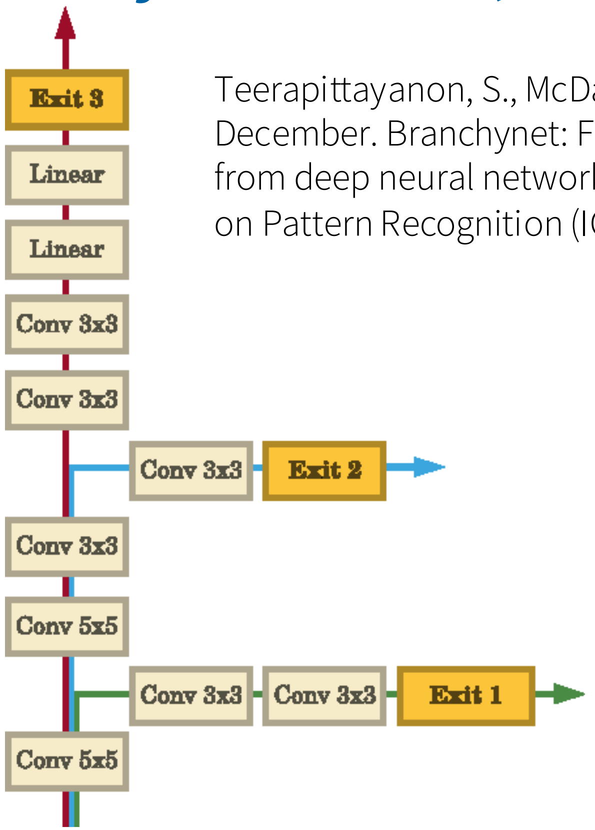
DNN Partitioning (NeuroSurgeon)



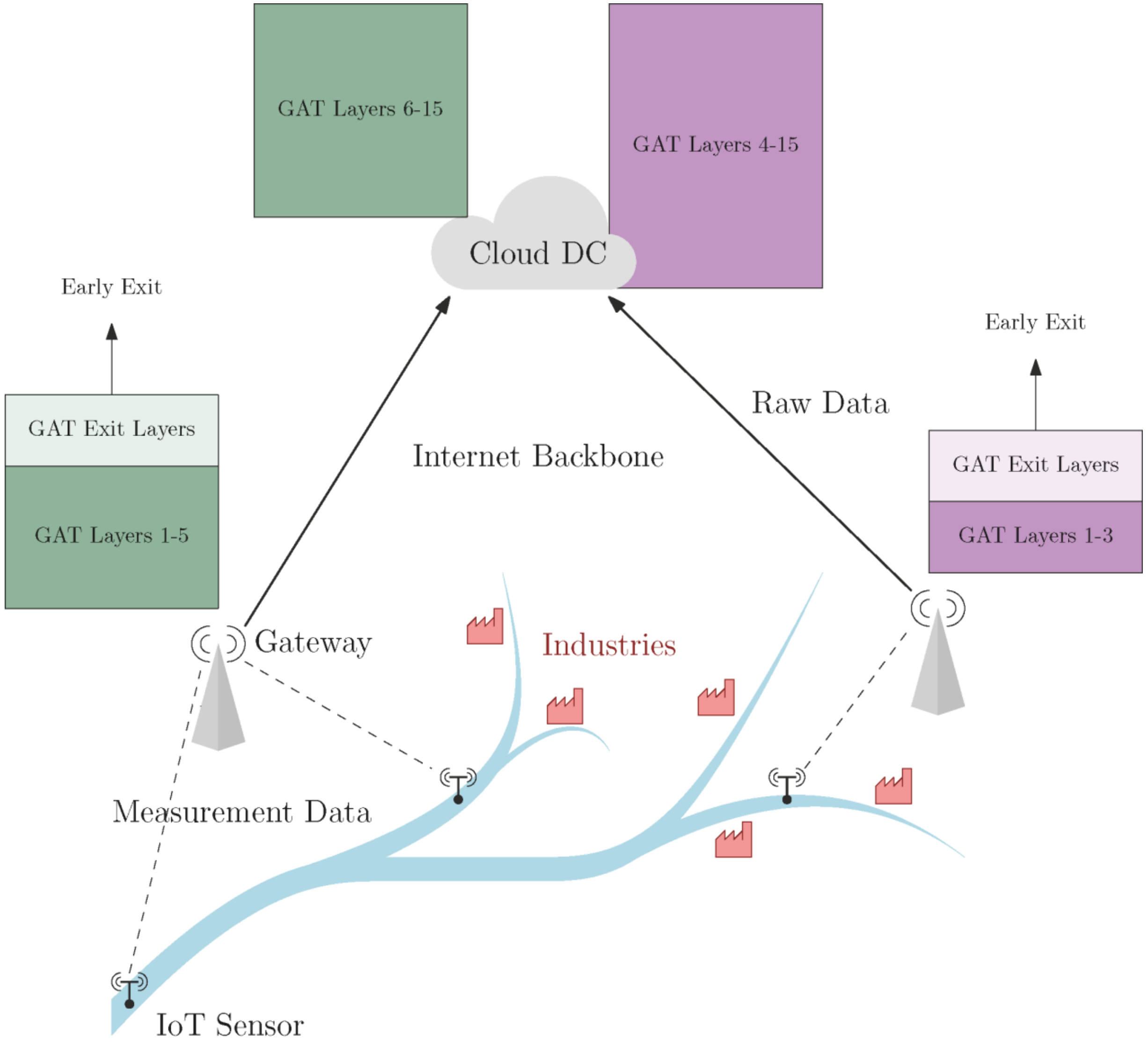


Kang, Y., Hauswald, J., Gao, C., Rovinski, A., Mudge, T., Mars, J. and Tang, L., 2017. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. ACM SIGARCH Computer Architecture News, 45(1), pp.615-629.

Early Exit DNN (BranchyNet)



Teerapittayanon, S., McDanel, B. and Kung, H.T., 2016, December. Branchynet: Fast inference via early exiting from deep neural networks. In International Conference on Pattern Recognition (ICPR) (pp. 2464-2469).



Sustainable IoT and Edge AI for Remote Monitoring

Public Lecture Series: Sustainability in Computer Science | Asst. Prof. Dr. Atakan Aral, University of Vienna

07.10.2024