New Emergency Management in a Resilience Era Facing Health, Climate and Energy Challenges

6th to 10th December 2021

9th DECEMBER, 14:25 - 14:50
SESSION 13 IMPETUS PRESENTATIONS

Paolo Mignone[1], Marco Anisetti[2], Claudio Ardagna[2], Costantino Mele[1], Alessandro Balestrucci[3], Chiara Braghin[2], Ernesto Damiani[2], Antongiacomo Polimeno[2], Donato Malerba[1], Michelangelo Ceci[1]

[1]Università degli studi di Bari Aldo Moro - Piazza Umberto I, 1, 70121 Bari, Italy
[2]Università degli studi di Milano - Via Festa del Perdono, 7, 20122 Milano, Italy
[3]Consorzio Interuniversitario Nazionale per l'Informatica – CINI - Via Ariosto, 25, 00185 - Roma, Italy

BIG DATA INGESTION AND ANALYTICS FOR PHYSICAL THREAT INTELLIGENCE

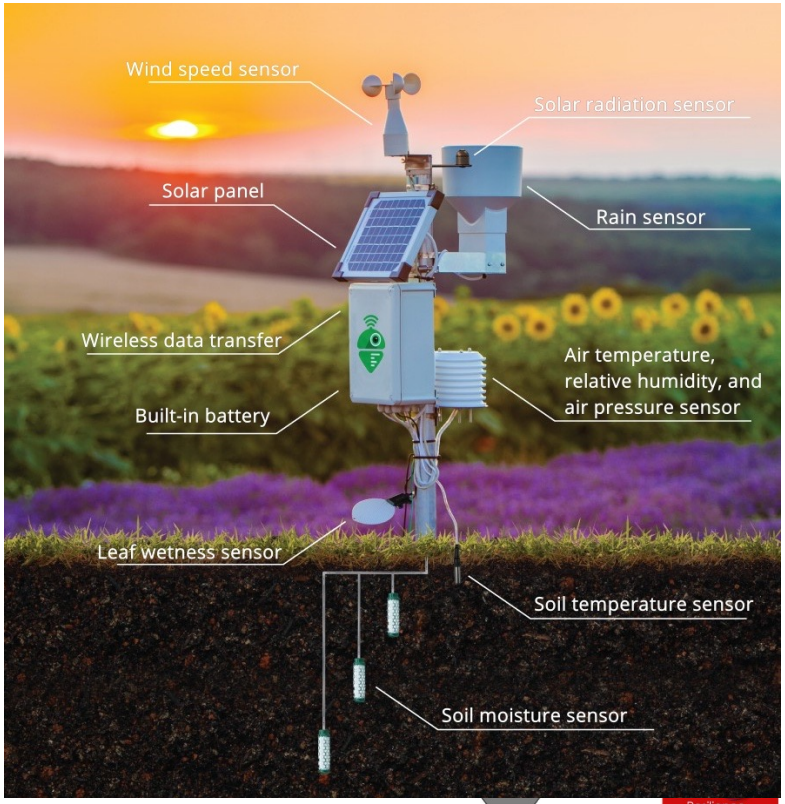
- The project IMPETUS aims to support law enforcement in establishing any threats / anomalies in advance, also with the aid of sophisticated automatic systems
- Expected impacts and results: the IMPETUS platform will make it possible to enhance the urban security of both cities involved (municipalities of Padua and Oslo)
- Elements of replicability: the platform can be replicated for each smart city depending on the sensor network made available by the municipality concerned
- Different types of sensors and possible applications

Urban traffic cameras

- A traffic camera observes vehicular traffic on the road.
- A monitoring center receives live video in real time and acts as a mediator with law enforcement if an accident or road safety problem occurs.

Meteorological sensors

- A weather sensor is a set of tools for measuring atmospheric conditions.
- Most of these sensors measure the
  - wind speed;
  - wind direction;
  - the temperature;
  - humidity;
  - barometric pressure;
  - rainfall;
  - UV or solar radiation.



Air pollution sensors

- Air pollution sensors are devices that monitor the presence of air pollution in the surrounding area.
- Most of these sensors focus on five components:
  - ozone;
  - atmospheric particulate matter (PM10, PM2.5…);
  - carbon monoxide;
  - sulfur dioxide;
  - nitrous oxide.

Sensors for detecting pollen levels

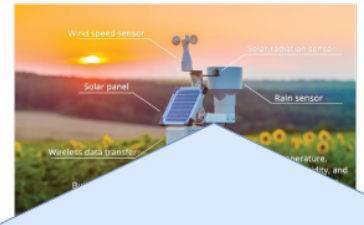
- Devices for detecting the concentration levels in the air of different pollen families.

Sensor metadata example

Timestamp yyyy-MM-dd*HH:mm:ss	X GPS coord	Y GPS coord	feat_1	feat_2	...	feat_n
2017-07-04*13:23:55	11.908710	45.404300	189	3	...	77

Sensor metadata example

- Several sensors that also provide temporal and spatial data



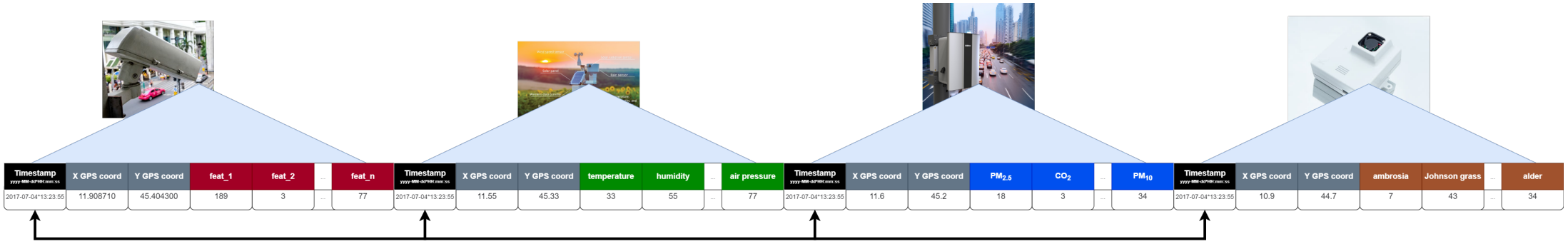
Timestamp yyyy-MM-dd*HH:mm:ss	X GPS coord	Y GPS coord	feat_1	feat_2	...	feat_n
2017-07-04*13:23:55	11.908710	45.404300	189	3	...	77

Timestamp yyyy-MM-dd*HH:mm:ss	X GPS coord	Y GPS coord	temperature	humidity	...	air pressure
2017-07-04*13:23:55	11.55	45.33	33	55	...	77

Several sensors that also provide temporal and spatial data

Sensor metadata example

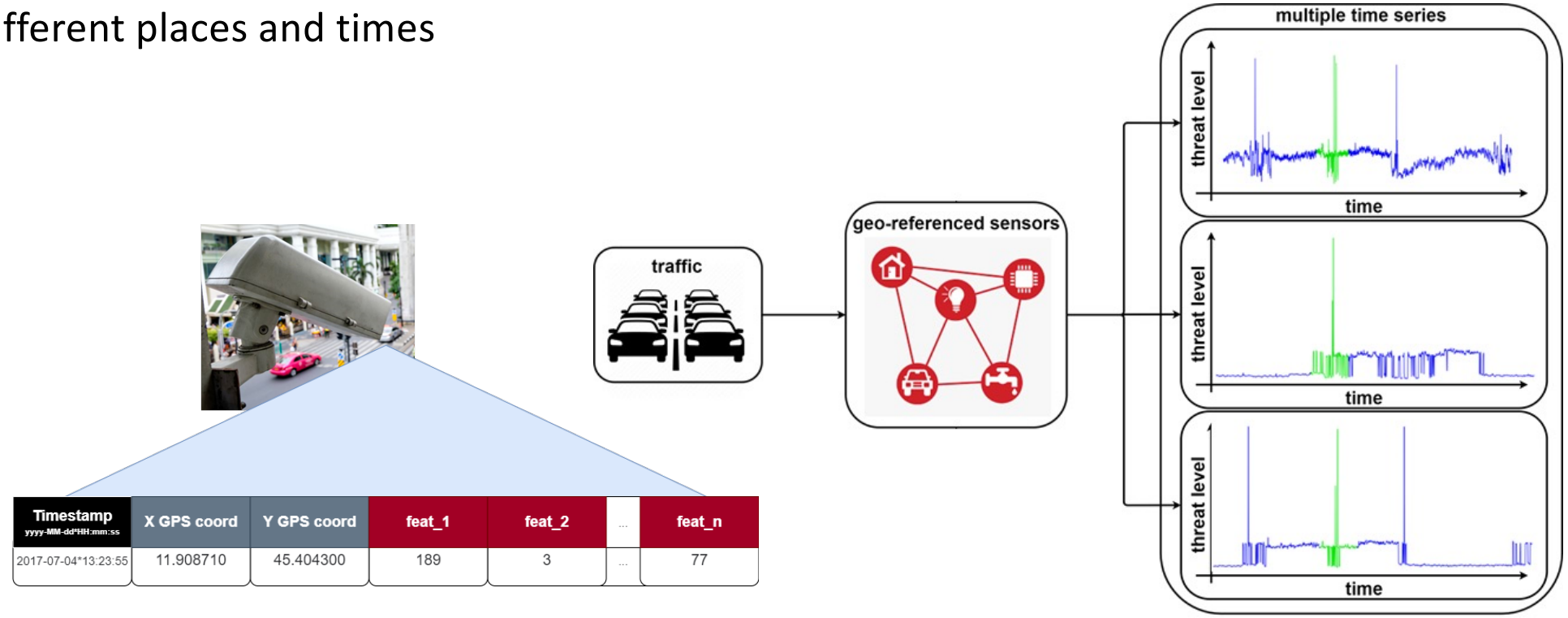
- Several sensors that also provide temporal and spatial data



Several sensors that also provide temporal and spatial data

Anomaly Detection

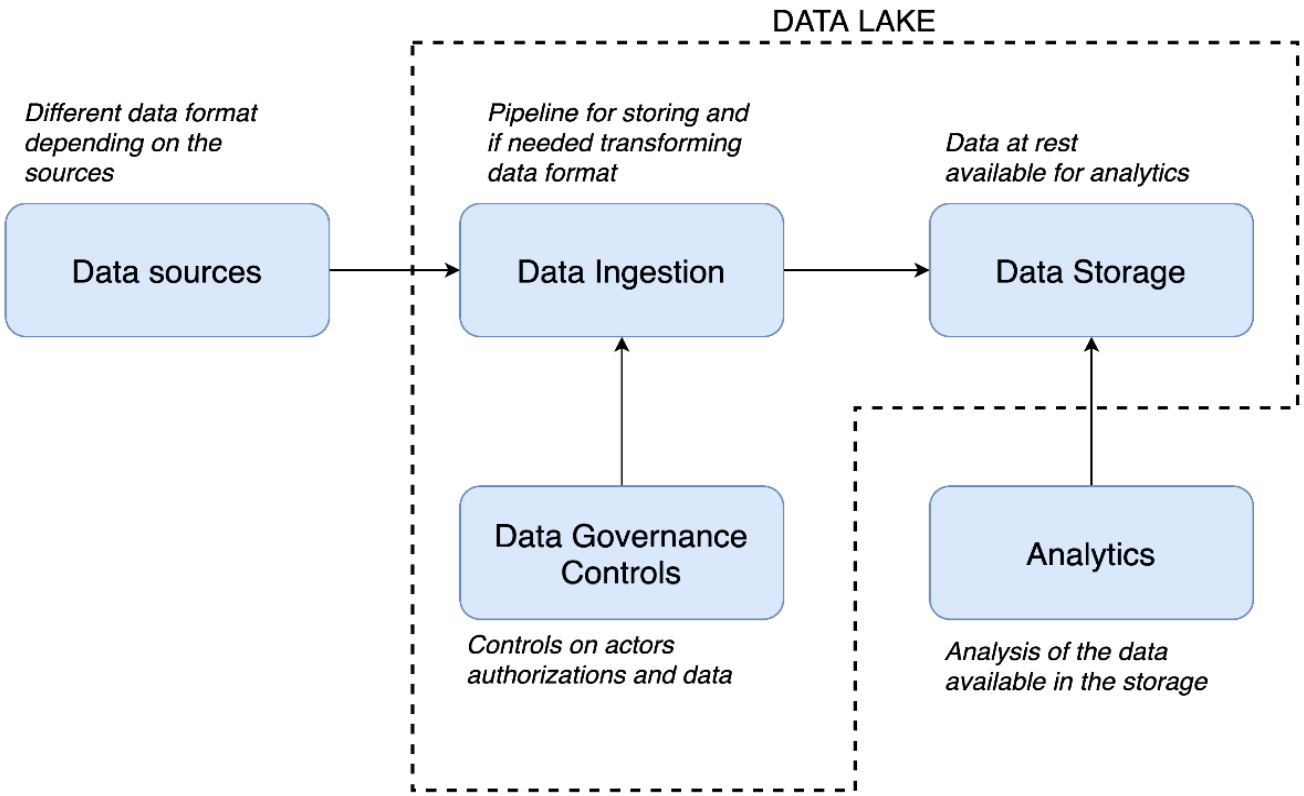
- An application example: identify anomalies in the distribution of camera metadata in different places and times



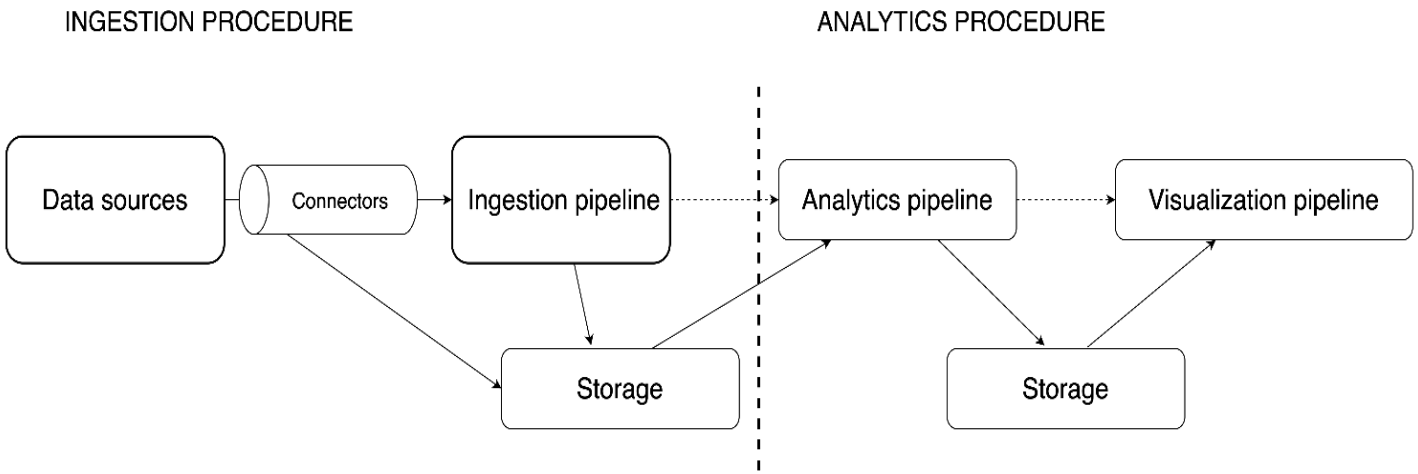
- When an anomaly is detected, the system will generate an alert

Conceptual View

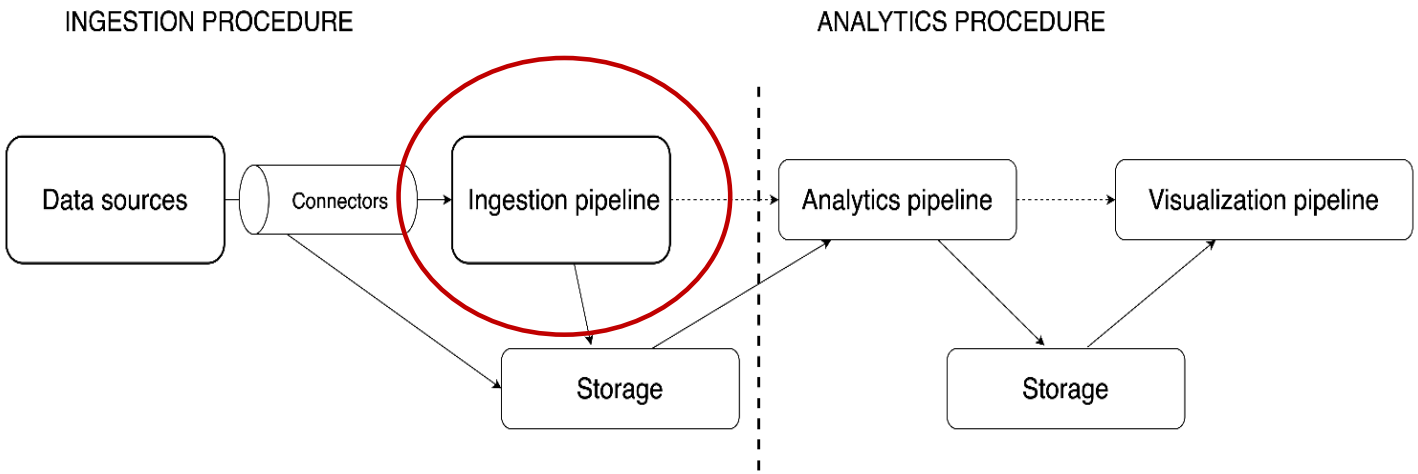
- Data management process
  - Collecting
  - Storing
- Maintain data
  - Secure
  - Efficient
  - Cost effective
- Offer data to analytics
  - Privacy-preserving preparation



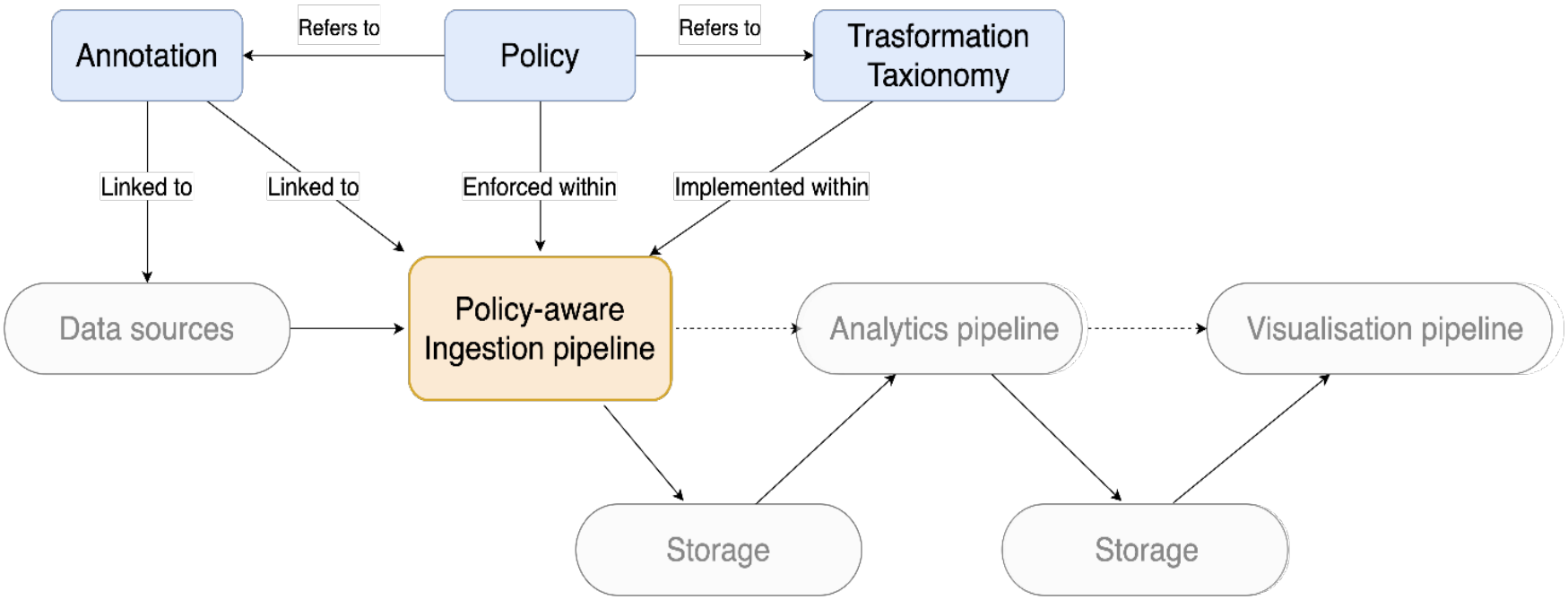
Architecture: Abstract view



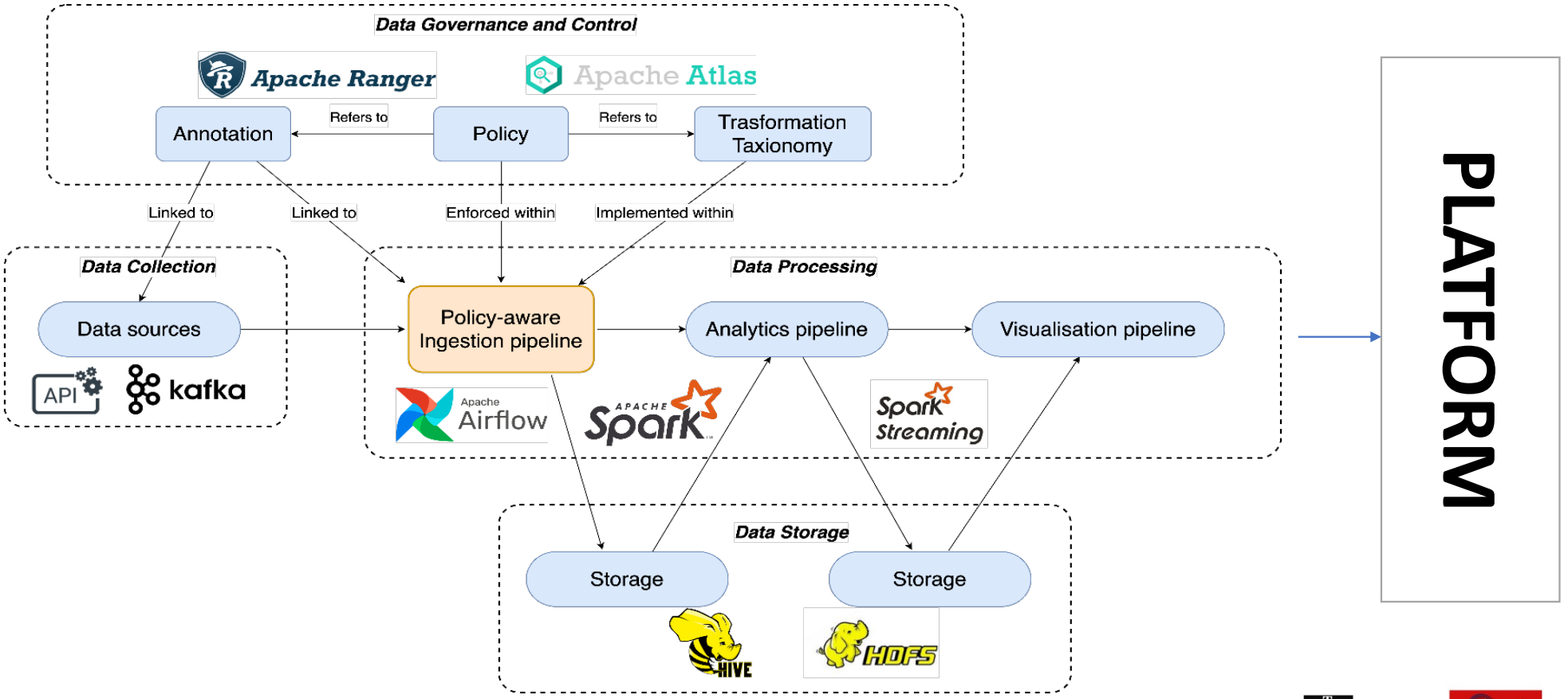
Architecture: Ingestion pipeline



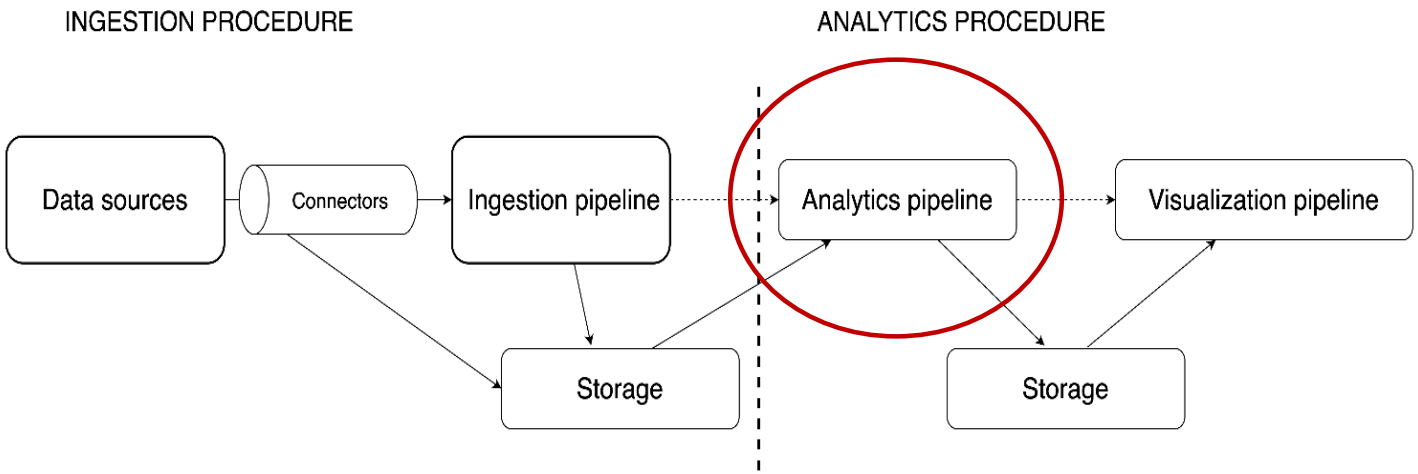
Architecture: ingestion time access control



Technologies and frameworks



Architecture: Analytics Pipeline



Detection of anomalies in the air (Padua)

- The dataset contains daily environmental information for the city of Padua, referring to the time period **2014** – **2019**, collected as batch.
- The dataset consists of a sample per day, excluding those days when measurements are not present.
- Data extracted from the ARPA Veneto website, Open Data section.
- The information extracted includes the following measurements:
  - date of measurement (dd/mm/yyyy)
  - air temperature at 2 metres (°C) (mean, min, max temperature)
  - precipitation (mm)
  - humidity at 2 metres (%) (min, max humidity)
  - concentrations of allergenic pollen in air (granules/m^3) per family (26 families)

Detection of anomalies in the air (Padua)

Anomaly detection task - Interpretability

Identify anomalies in environmental data distributions by considering different sensors in different places at different times

Timestamp	mean temperature	...	ambrosia	urticaceae
2021-04-15 10:00	15.1	...	20.0	50.0

Timestamp	mean temperature	...	ambrosia	urticaceae
2021-04-16 11:00	15.1	...	30.0	52.0

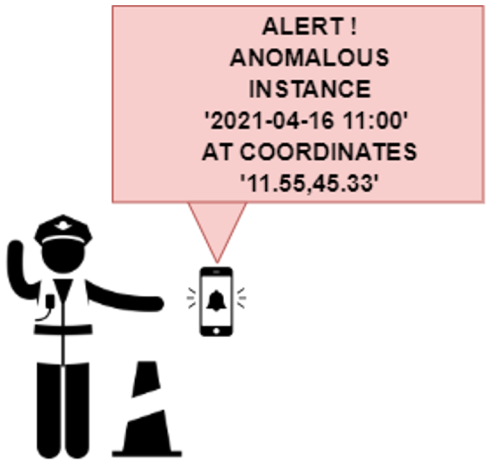
ALERT ! ANOMALOUS INSTANCE '2021-04-16 11:00'

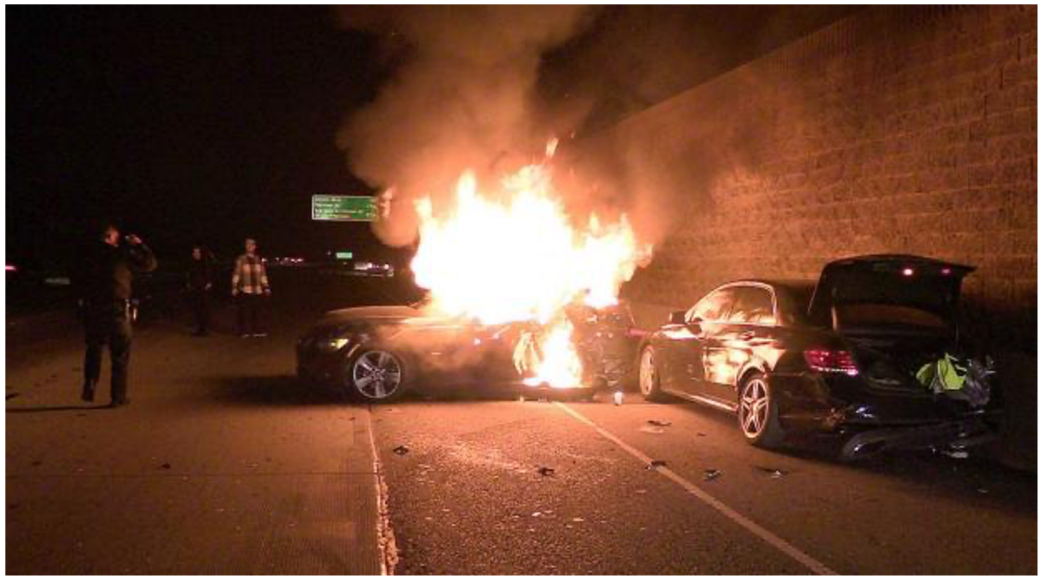
RANKING	IMPORTANCE
1. ambrosia	0.45
2. min humidity	0.24
⋮	
n. urticaceae	0.03

When an anomaly will be detected, the system will raise an alert and will provide a feature ranking according to the features importance in detecting the anomaly

Detection of anomalies in the air (Padua)

Interpretability – a real scenario





Timestamp	X coord	Y coord	...	traffic level (%)	CO2 (ppm)
2021-04-16 11:00	11.55	45.33	...	68	2205

RANKING		IMPORTANCE
1.	traffic	0.48
2.	concentration of pedestrians	0.21
3.	CO2	0.15
4.	temperature	0.07
⋮		
n.	ambrosia	0.001

Detection of anomalies in the air (Padua)

rainfall intensity	mm/6h	mm/12h	mm/24h
weak	0 - 5	0 - 10	0 - 15
moderate	5 - 15	10 - 30	15 - 45
strong	15 - 30	30 - 60	45 - 90
very strong	>30	>60	>90

Example of ground truth for rainfall intensity attribute

Each class in the ground truth is associated with an ‘anomaly weight’ in the range [0.0, 1.0]

e.g., for rainfall intensity, weak [0.0], moderate [0.33], strong [0.66] and very strong [1.0]

Detection of anomalies in the air (Padua)

Evaluation of the feature ranking

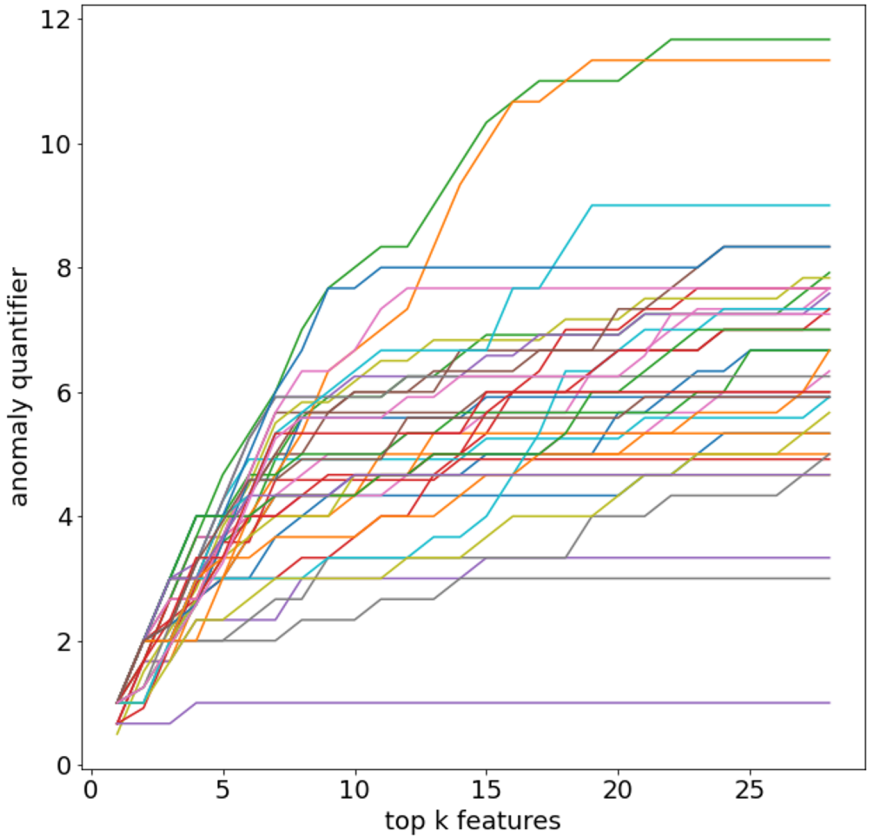
Given k the k-th position in the feature ranking

$$anomaly\ quantifier@k = \sum_{i=1}^{k} w_i$$

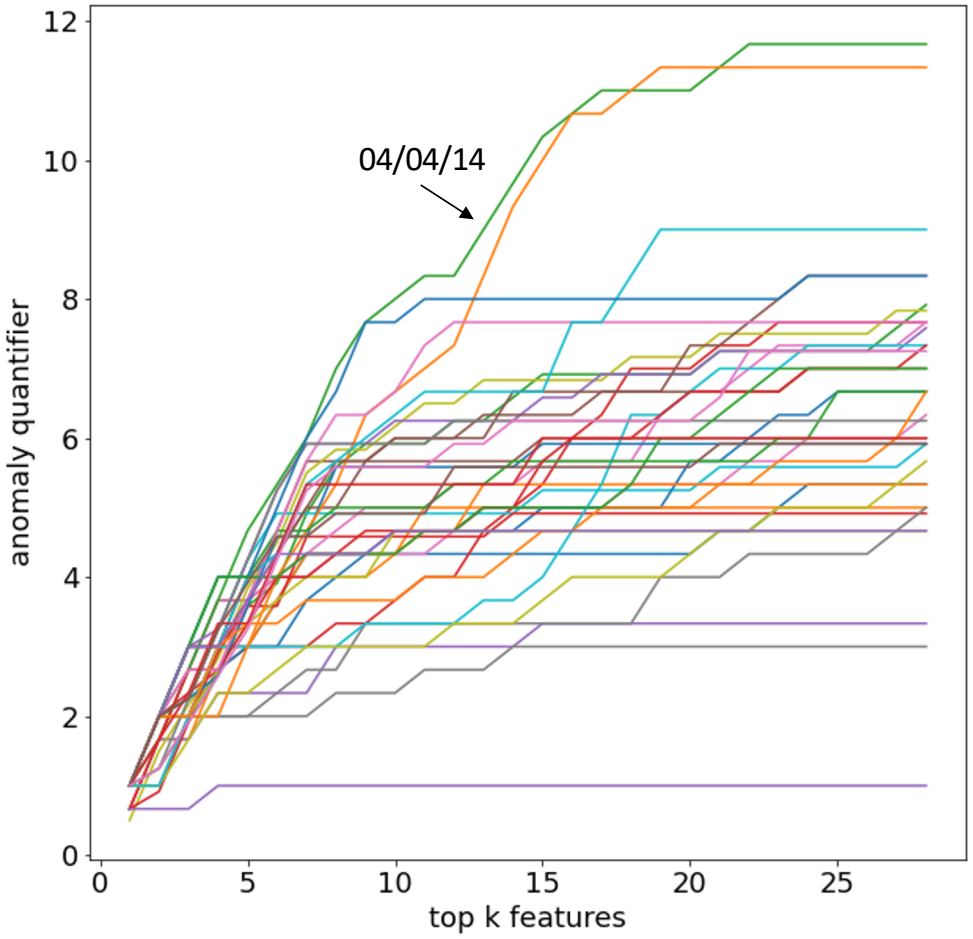
where w_i is the anomaly weight of the attribute in the i-th position, with $i \leq k$, if the ground truth of the attribute in the i-th position is defined.

Detection of anomalies in the air (Padua)

Using the environmental data of the city of Padua, the anomaly detector identified a total of 37 anomalies over the 395 test instances



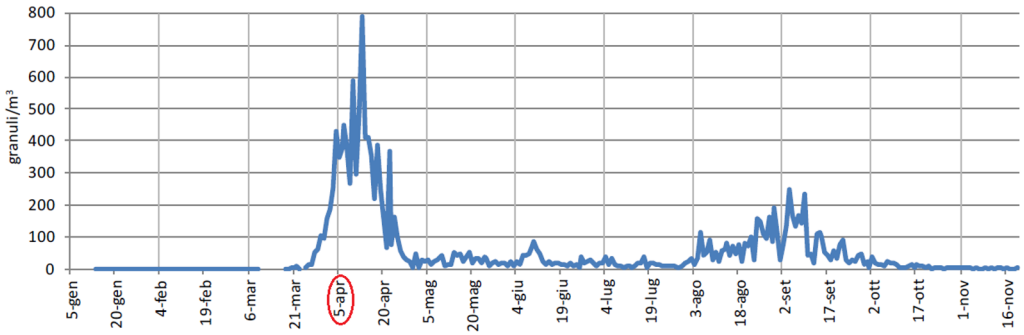
Detection of anomalies in the air (Padua)



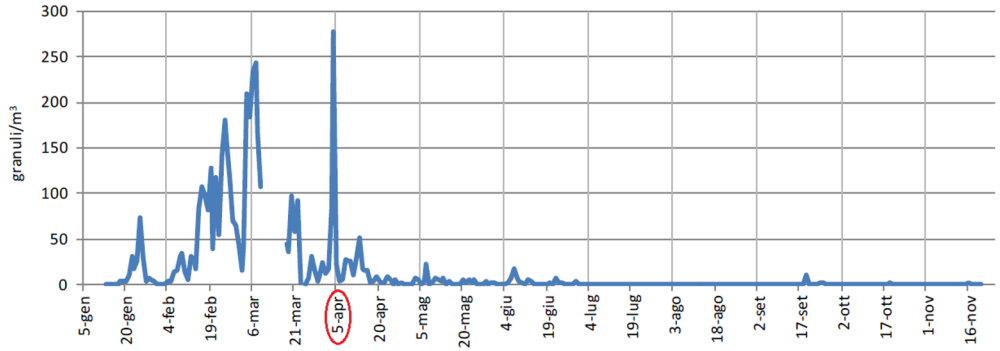
04/04/14 - top 10 feature ranking

1. urticaceae [1.00]
2. salix [0.67]
3. cupressaceae/taxaceae [1.00]
4. fagaceae [1.00]
5. quercus [1.00]
6. pinaceae [0.67]
7. corylaceae [0.67]
8. platanaceae [1.00]
9. betulaceae [0.67]
10. fagus sylvatica [0.33]

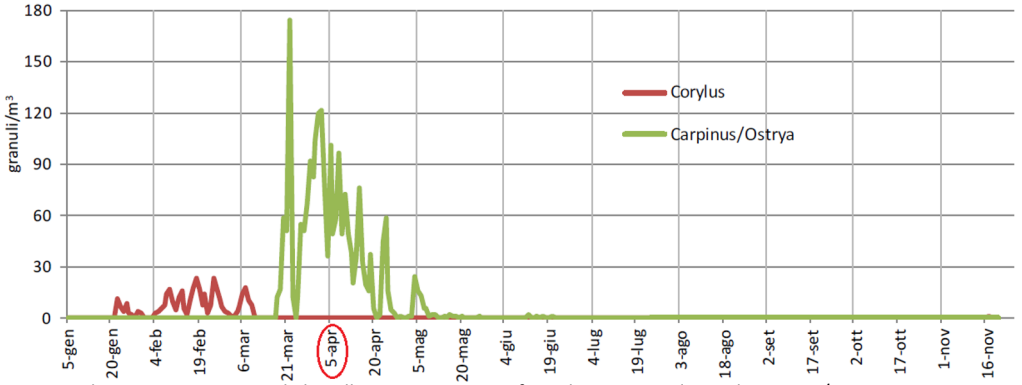
Detection of anomalies in the air (Padua)



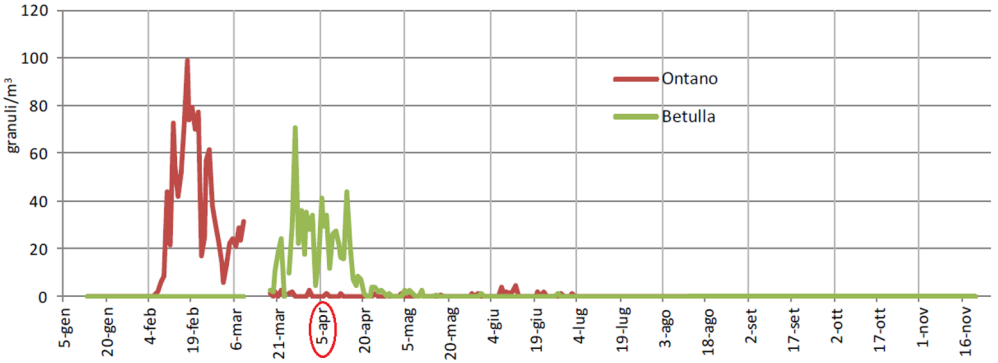
Padova station. Average daily pollen concentration of Urticaceae, year 2014.
Source: Pollini e Spore fungine allergenici nel Veneto, rapporto 2014, ARPAV.
Urticaceae, 1st position in the feature ranking



Padova station. Average daily pollen concentration of Cupressaceae/Taxaceae, year 2014.
Source: Pollini e Spore fungine allergenici nel Veneto, rapporto 2014, ARPAV.
Cupressaceae/Taxaceae, 3rd position in the feature ranking

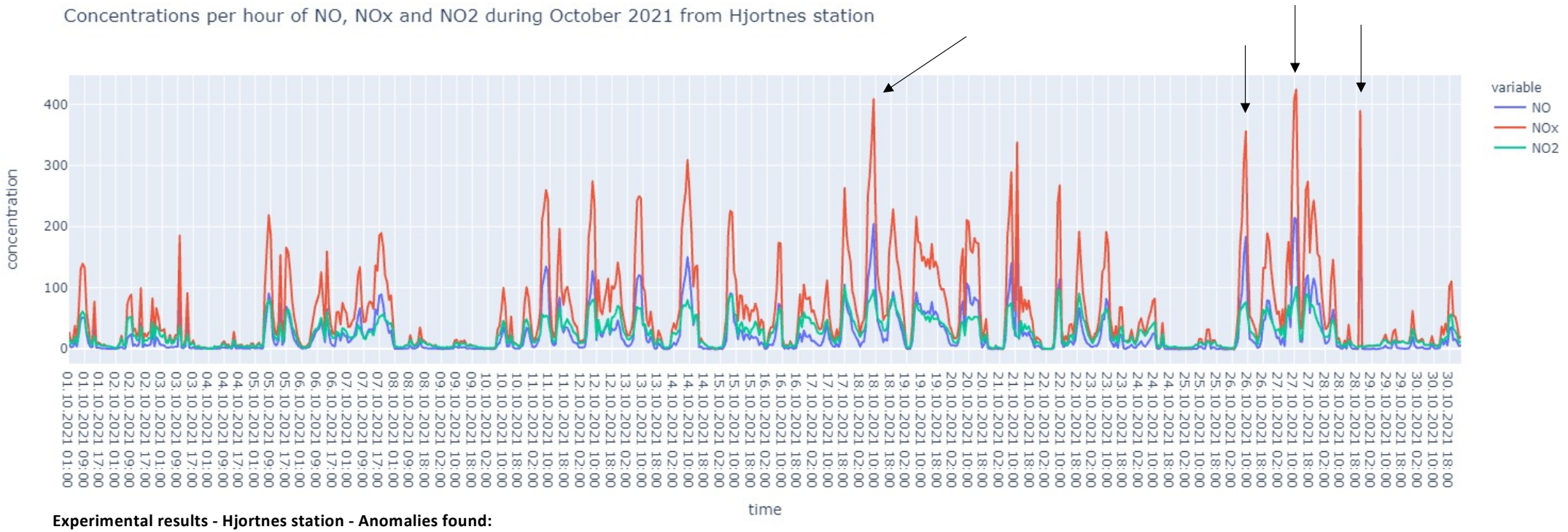


Padova station. Average daily pollen concentration of Corylaceae - Corylus and Carpinus/Ostrya, year 2014.
Source: Pollini e Spore fungine allergenici nel Veneto, rapporto 2014, ARPAV.
Corylaceae, 7th position in the feature ranking



Padova station. Average daily pollen concentration of Betulaceae – Alnus e Betula, year 2014.
Source: Pollini e Spore fungine allergenici nel Veneto, rapporto 2014, ARPAV.
Betulaceae, 9th position in the feature ranking

Detection of anomalies in the air (Oslo)



Experimental results - Hjortnes station - Anomalies found:

18/10/2021 08.00 – [**1. NOx**, **2. NO2**, **3. NO**]

18/10/2021 09.00 – [**1. NOx**, **2. NO2**, **3. NO**]

26/10/2021 09.00 – [**1. NOx**, **2. NO**, **3. NO2**]

27/10/2021 10.00 – [**1. NOx**, **2. NO**, **3. NO2**]

27/10/2021 11.00 – [**1. NOx**, **2. NO2**, **3. NO**]

28/10/2021 20.00 – [**1. NO**, **2. NOx**]

Conclusions

- Our method for the ***anomaly detection*** task, in addition to being able to analyze large amounts of data in a distributed way, allows us to correctly identify all anomalies although they are usually rare.
- The method avoids false normal cases by detecting the anomalies correctly for the air pollution quality of Padua and Oslo
- Furthermore, a qualitative analysis showed that the method is capable to catch anomalous levels of air quality variables for both the considered cities taking into account also the spazio-temporal autocorrelation

References

[1] S. Ghemawat, H. Gobioff and S.-T. Leung, "The Google file system," in *ACM symposium on Operating systems principles*, 2003.

[2] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM,* vol. 51, no. 1, pp. 107--113, 2008.

[3] R. T. Fielding, Architectural styles and the design of network-based software architectures, Irvine: University of California, 2000.

[4] L. Sweeney and P. Samarati, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," SRI International, 1998.

[5] V. C. Hu et al., "Guide to attribute based access control (abac) definition and considerations," *NIST special publication,* vol. 800, no. 162, pp. 1--54, 2013.

[6] P. Colombo and E. Ferrari, "Access control technologies for Big Data management systems: literature review and future trends," *Cybersecurity,* vol. 2, no. 1, pp. 1--13, 2019.

[7] F. M. Awaysheh, M. Alazab, M. Gupta, T. F. Pena and J. C. Cabaleiro, "Next-generation big data federation access control: A reference model," vol. 108, pp. 726-741, 2020.

[8] M. Gupta, F. Patwa and R. Sandhu, "Object-Tagged RBAC Model for the Hadoop Ecosystem," in *Data and Applications Security and Privacy*, 2017.

[9] M. Gupta, F. Patwa and R. Sandhu, "An Attribute-Based Access Control Model for Secure Big Data Processing in Hadoop Ecosystem," in *Workshop on Attribute-Based Access Control (ABAC'18)*, New York (US), 2018.

[10] P. Colombo and E. Ferrari, "Towards a Unifying Attribute Based Access Control Approach for NoSQL Datastores," in *International Conference on Data Engineering (ICDE)*, 2017.

[11] V. Hu, T. Grance, D. F. Ferraiolo and K. D. R., "An Access Control Scheme for Big Data Processing," in *International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, Miami, FL (US), 2014.

[12] M. M. Ben Aissa, L. Sfaxi and R. Robbana, "DECIDE: A New Decisional Big Data Methodology for a Better Data Governance," in *European, Mediterranean, and Middle Eastern Conference on Information Systems (EMCIS 2020)*, 2020.

[13] A. Al-Badi, A. Tarhini and A. I. Khan, "Exploring Big Data Governance Frameworks," *Procedia Computer Science,* no. 141, pp. 271-277, 2018.

[14] Y. Demchenko et al., "Addressing big data issues in scientific data infrastructure," in *collaboration technologies and systems (CTS)*, 2013.

[15] V. Chandola, A. Banerjee and A. Kumar, "Anomaly detection: A survey," vol. 41, no. 3, 2009.

[16] T. Kohonen, "Essentials of the self-organizing map," vol. 37, pp. 52-65, 2013.

[17] E. Araujo, C. R. Silva and D. J. B. S. Sampaio, "Video target tracking by using competitive neural networks," vol. 4, no. 8, pp. 420-431, 2008.

[18] M. Dittenbach, D. Merkl and A. Rauber, "The growing hierarchical self-organizing map," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks,* vol. 6, pp. 15-19, 200.

[19] A. Malondkar, R. Corizzo, I. Kiringa, M. Ceci and N. Japkowicz, "Spark-GHSOM: Growing Hierarchical Self-Organizing Map for large scale mixed attribute datasets," *Information Sciences,* vol. 496, pp. 572-591, 2019.

[20] "Arpa Veneto, Open Data," [Online]. Available: https://www.arpa.veneto.it/dati-ambientali/open-data.

[21] "ARPA Veneto report on allergenic fungal pollens and spores," 2014. [Online].

[22] "Norwegian Institute for Air Research (NILU)," [Online]. Available: https://www.home-assistant.io/integrations/nilu/.

[23] "ENTUR API documentation," [Online]. Available: https://developer.entur.org/pages-real-time-intro.

[24] "https://en.wikipedia.org/wiki/DBSCAN," [Online].

[25] R. Corizzo, G. Pio, M. Ceci and D. Malerba, "DENCAST: Distributed Density-Based Clustering for Multi-Target Regression," *Journal of Big Data.*

[26] V. Goyal et al., "Attribute-based encryption for fine-grained access control of encrypted data.," in *13th ACM conference on Computer and communications security*, 2006.

[27] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Annual international conference on the theory and applications of cryptographic techniques.*, Berlin, 2005.

[28] X. Boyen and B. Waters, "Anonymous hierarchical identity-based encryption (without random oracles).," in *Annual International Cryptology Conference*, Berlin, 2006.

[29] F. Armknecht et al., "A Guide to Fully Homomorphic Encryption," in *IACR Cryptol. ePrint Arch*, 2015.

[30] M. Anisetti, C. Ardagna, C. Braghin, E. Damiani, A. Balestrucci and A. Polimeno, "Dynamic and Scalable Enforcement of Access Control Policies for Big Data," *ACM MEDES,* 2021.

Acknowledgments: we acknowledge the project *IMPETUS (Intelligent Management of Processes, Ethics and Technology for Urban Safety)* that receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883286.
https://cordis.europa.eu/project/id/883286