Efficiently Integrating Mobility and Environment Data for Climate Change Analytics

Isam Mashhour Al Jawarneh, Paolo Bellavista, Antonio Corradi, Luca Foschini , Rebecca Montanari

Department of Computer Science and Engineering - DISI
University of Bologna, Italy

{isam.aljawarneh3, luca.Foschini, paolo.bellavista, antonio.corradi,Rebecca.montanari}@unibo.it,

IEEE CAMAD 2021
25 - 27 October 2021 .
Virtual conference S02-P2 session

Outline

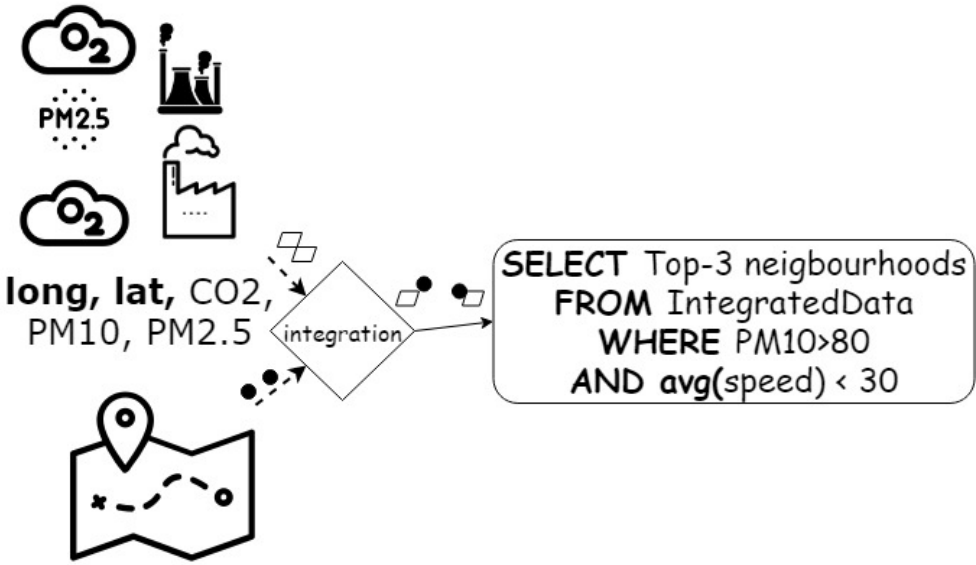
- **Integrating** mobility and meteorological data: Background and Motivating scenario
  - Challenges & requirements
- **MeteoMobil**: efficient system for integrating meteorological & mobility data with QoS guarantees
  - System architectural design
  - Supported queries
- **Deployment**
  - Baselines
  - Testing setup
- **Results** and Discussion
  - Engineering *MeteoMobil* atop Apache Spark
- **Conclusion** and recommended future works

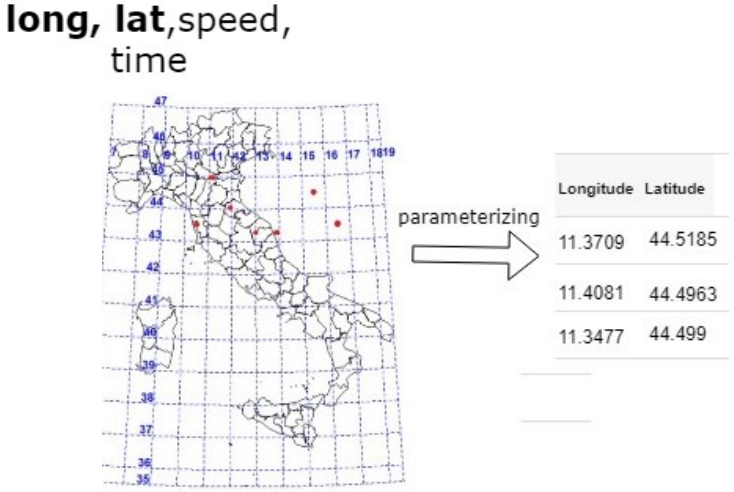
Motivating scenario

- Studying the correlation between vehicle pollutant emissions and the health of dwellers in metropolitan cities
  - Requires joining geo-referenced mobility and meteorological data (**spatial join**)
- Requires regular analytics of the relationships between mobility patterns and climate change
  - e.g., Interactive heatmap visualization
- Helps municipalities and city officials in making strategic decisions for the benefit of the health of citizens

Requirements & challenges

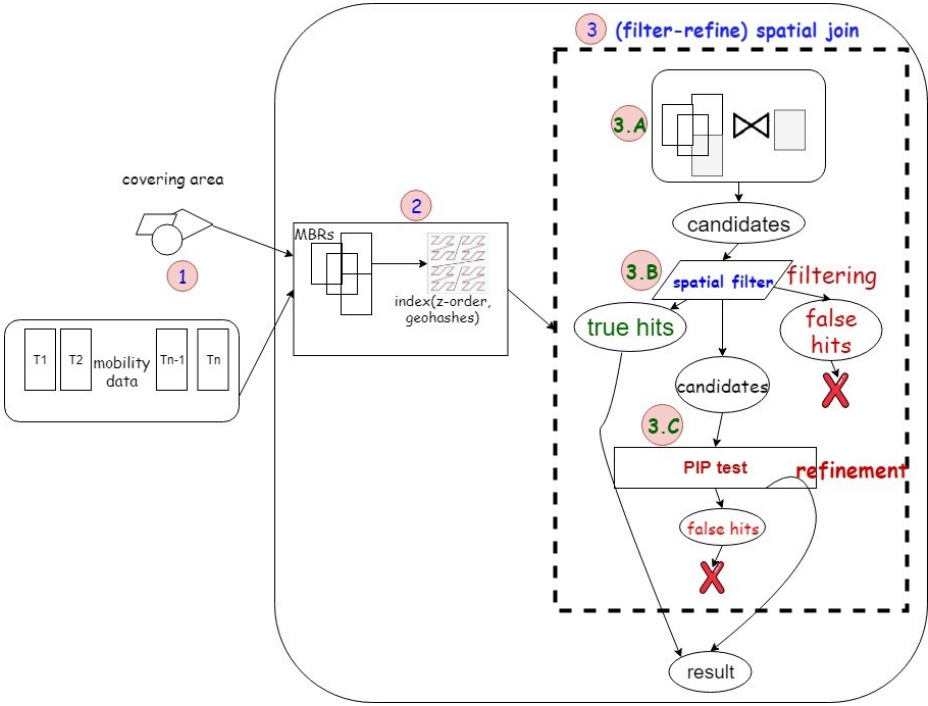
- To be able to perform shared analytics on mobility and meteorological data altogether.
  - We need to apply **spatial join**.
- However, spatial join is computationally expensive.
  - Spatial data is parametrized (longitudes and latitudes)
  - Objects loses their geometrical information by this transformation.
  - Bringing parametrized tuples back into real geometries is expensive.





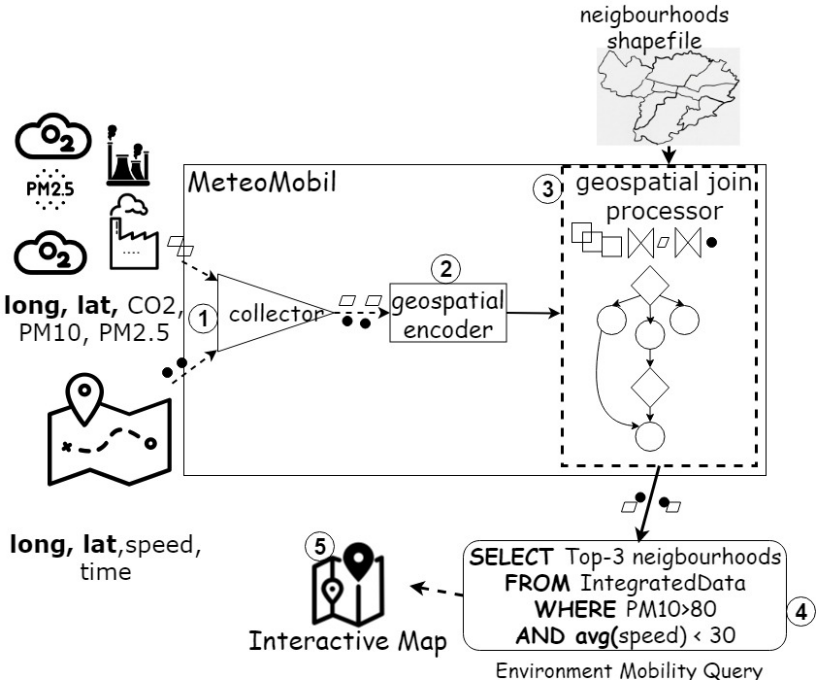
efficient spatial join

- Filter-refine approach
  - generate geocode for each tuple (MBRs)
  - Generate MBRs covering the space.
  - Apply a cheap MBR-join (filtering step)
  - Apply a costly join algorithm to on the rest to discard false positives (refinement step)



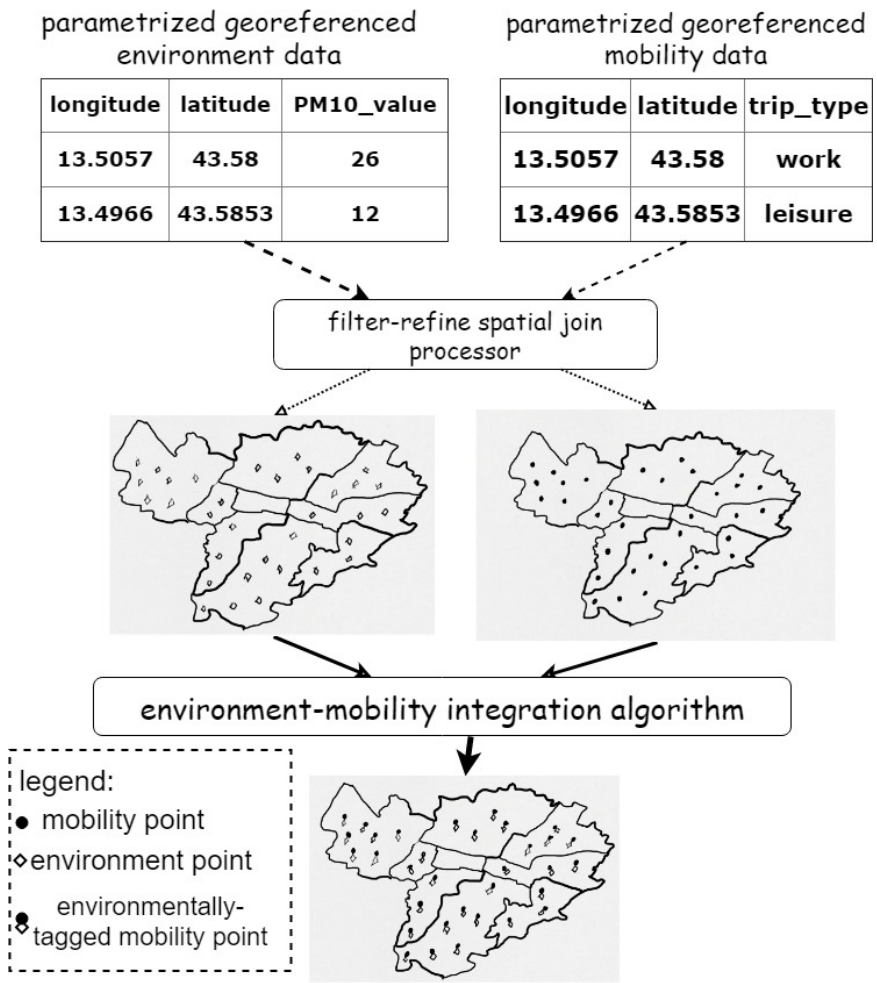
MeteoMobil architecture

- MeteoMobil is a Cloud-based efficient system for the joint processing of mobility and metrological data at scale.
  - Data collector gathers georeferenced mereological and mobility data.
  - Geospatial encoder generates codes for each point.
  - Spatial join processor performs the join.
  - Query processor performs interactive queries on the unified mobility-meteorological data.



Heuristic overview of MeteoMobil

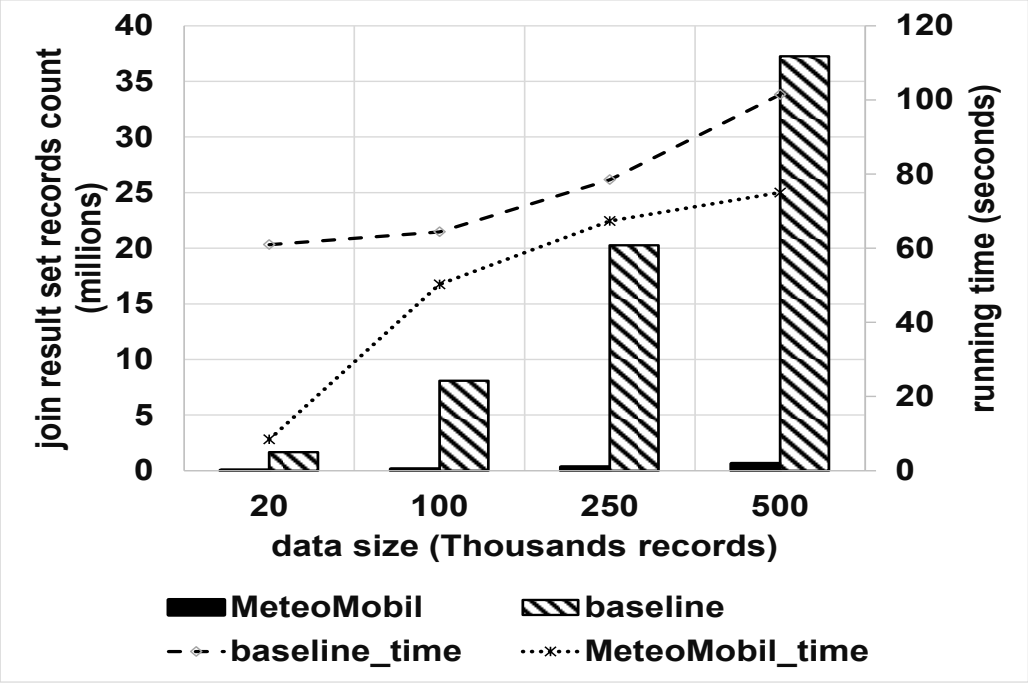
- Our method is equivalent to the heuristic overview shown in figure
- It resorts to overlaying corresponding maps of both datasets with a cheap equijoin operation



Experimental setup

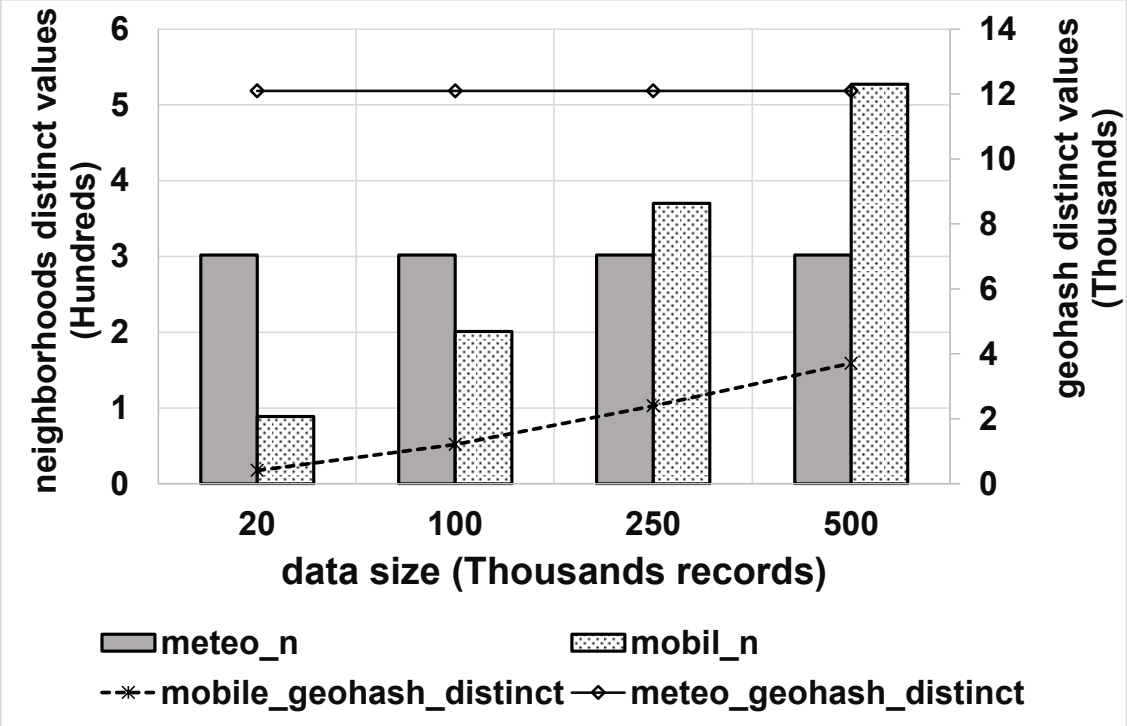
- Evaluation metrics
  - Running times for linear and aggregation (Top-N) queries
  - To explain results of linear queries, we adopt the index selectivity approach.
- Testbed
  - We have deployed MeteoMobil on a Microsoft Azure HDInsight Cluster hosting Apache Spark version 2.2.1.
  - Datasets:
    - two explicit feedback rating datasets
      - meteorological dataset, Urban SIS
      - Mobility dataset, ParticipAct project

Running times of MeteoMobil Vs. baselines



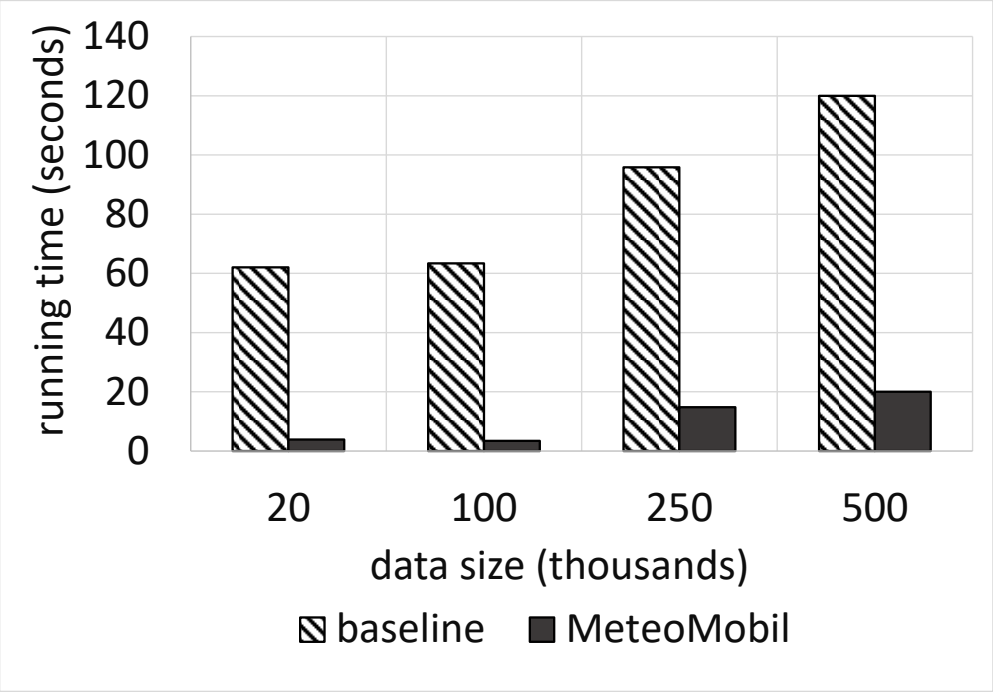
- Running times and number of records in the result set comparing MeteoMobile against the baseline using the ParticipAct and Urban SIS datasets. Parameters: geohash 30
- we obtain roughly 98% gain by for MeteoMobil against the baseline in terms of number of records in the result set, with an associated reduction in running time that is roughly equals to an average of 37%.

Distinct number of geohashes and neighborhoods for mobility and meteorological data



- Number of distinct geohashes and neighborhoods in both datasets. In the legend: 'meteo_n' is the distinct meteorological neighborhoods values, while 'mobil_n' is the distinct mobility neighborhoods values.
- neighborhood's range is far smaller compared to that of the geohash values.

Running times of Top-N queries, MeteoMobil against the baseline



- We specifically test on the following query: “which are the top-3 neighbourhoods in Bologna in Italy in terms of mobility traffic where concentration of PM10 is than 12”.
- our optimized methods in MeteoMobil are able to achieve a significant performance in gain as compared to the baseline.

Concluding remarks

- **MeteoMobil** is a novel efficient Cloud-based system for environment (e.g., meteorological) and mobility data integration at-scale.
- **MeteoMobil** features a **novel gospatial join algorithm** that simplifies the integration. In addition, it supports SQL-alike queries which simplifies analytics on environmentally-tagged spatial data.
- **Future research** should include modules for joining other data sources for deeper insightful analytics that reflect the socio-spatial variations. This requires joining sociodemographic data of dwellers
- Also, since mobility data is huge, **geospatial approximate query processing** approaches should be considered.
  - For example, a geospatial sampler that acts as a frontstage operating on an Edge device for selecting miniscule of the data before sending it to the Cloud.

Q&A and Contacts

Thanks for your attention!

Question's time...

Isam Al Jawarneh
Email: isam.aljawarneh3@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO DI INFORMATICA - SCIENZA E INGEGNERIA

IEEE CAMAD 2021
25 - 27 October 2021 ,
Virtual conference