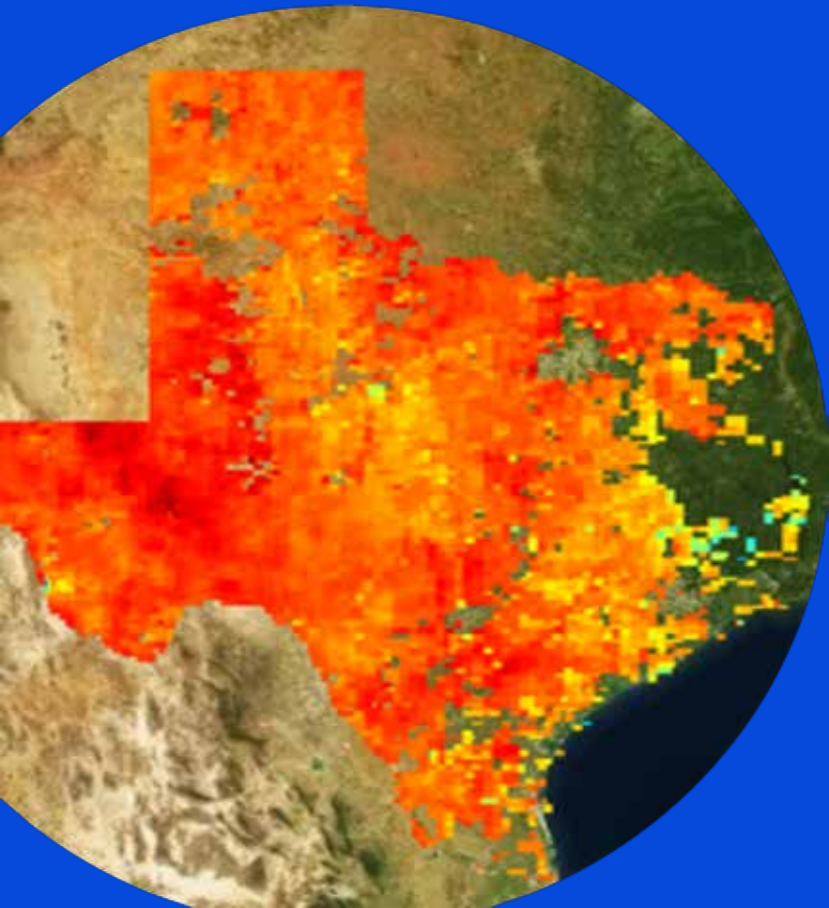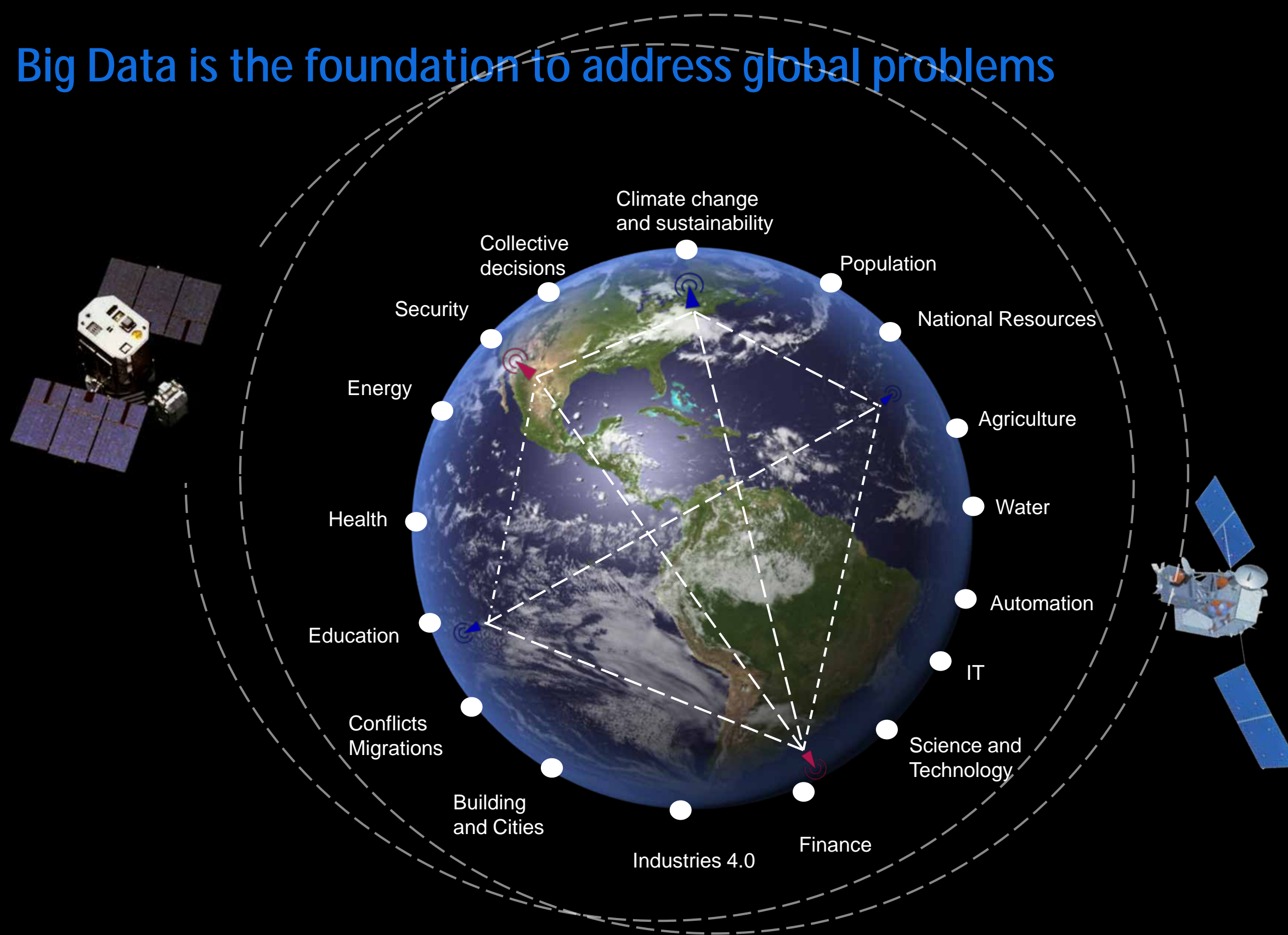# *Big Data and Artificial Intelligence: Global Scale Geo-Spatial Analytics*

Levente Klein

IBM TJ Watson Research Center, NY

# Big Data is the foundation to address global problems

What is PAIRS? *A platform developed at **IBM Research** to ingest and analyze drone, weather and satellite data at global scale*

TODAY'S APPROACH

- Each data set is kept at a different place in different formats, projections, units etc.
- Exabytes stored in billions of "individual scenes", or files typically on tape.
- Analyst has to "order" scenes from each source:
  - Download, assemble, re-sample, re-project, align, classify scenes, etc.....
- Data is moved to the application
- Time to value is limited by data curation (90%).

LandSat scenes/tiles from different satellite
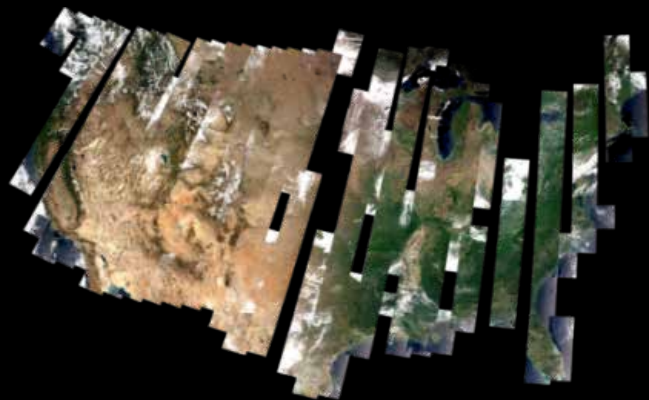
PAIRS-Big Data Geospatial Platform
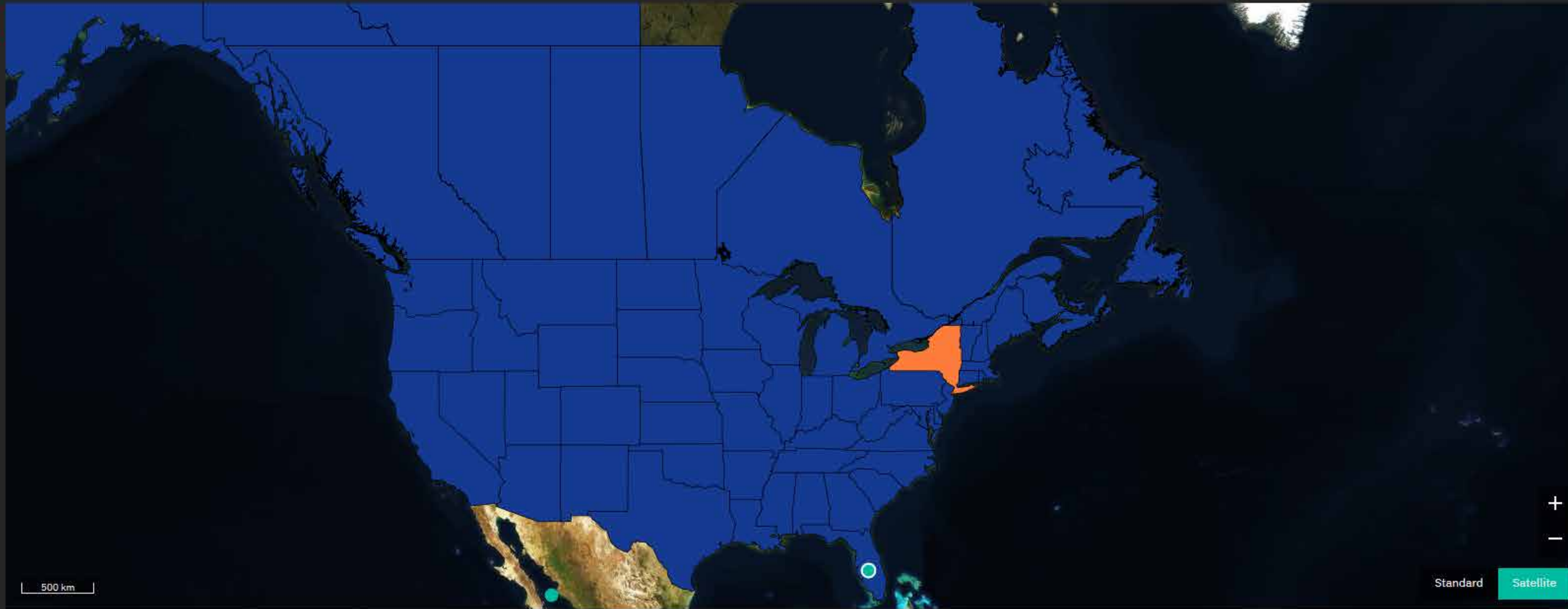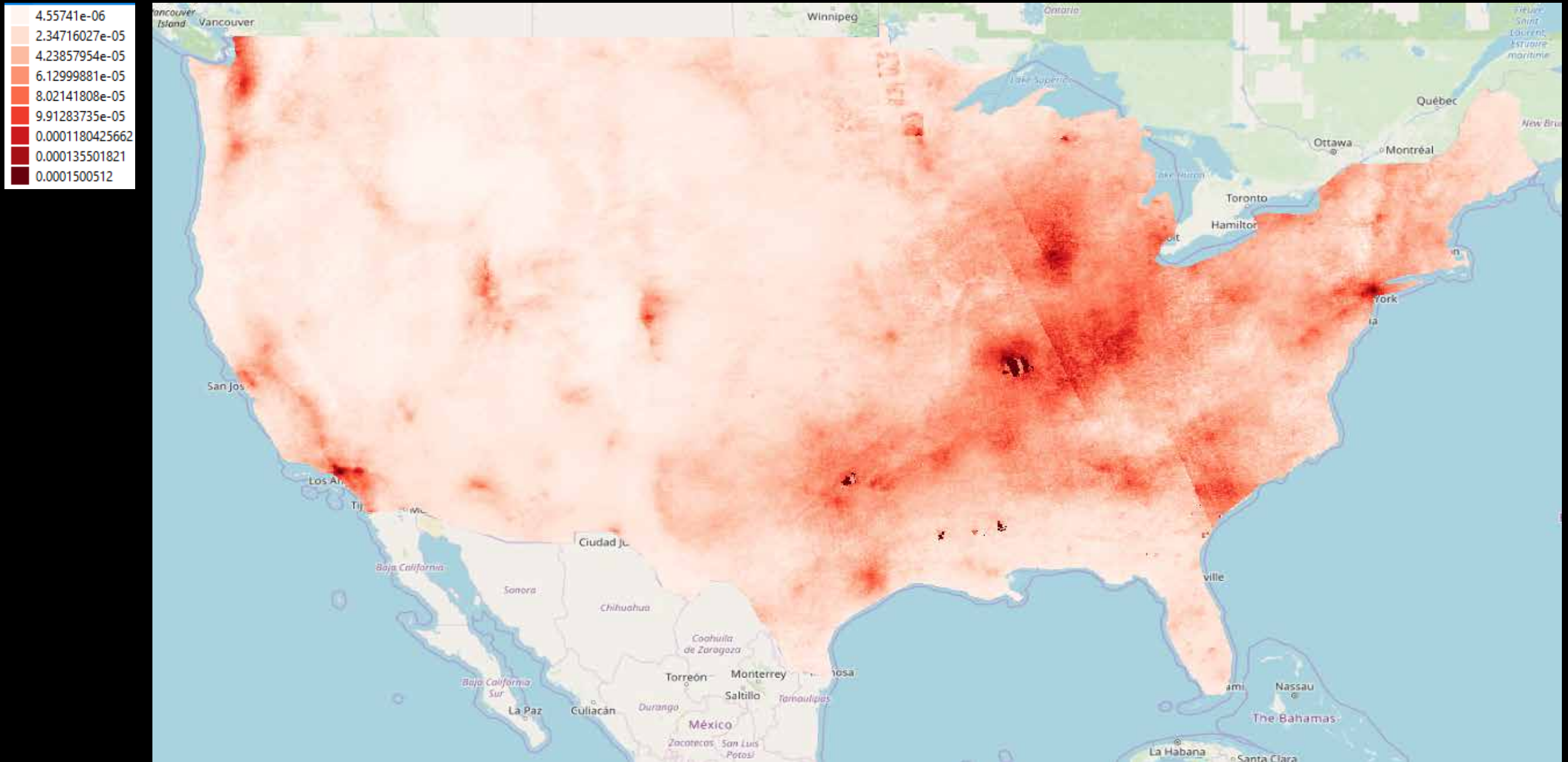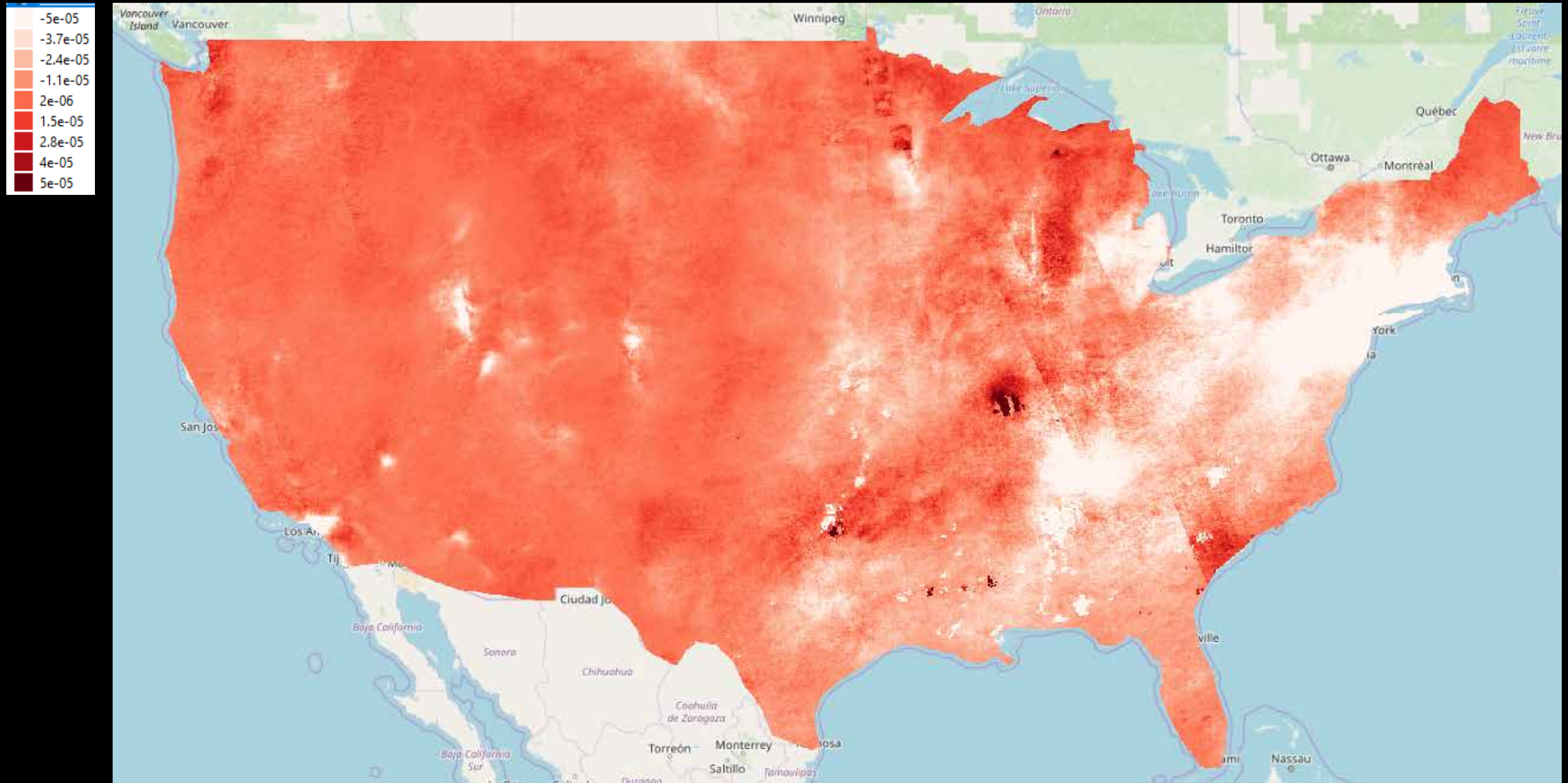
Economic activity monitored by NO2 emission-year to date
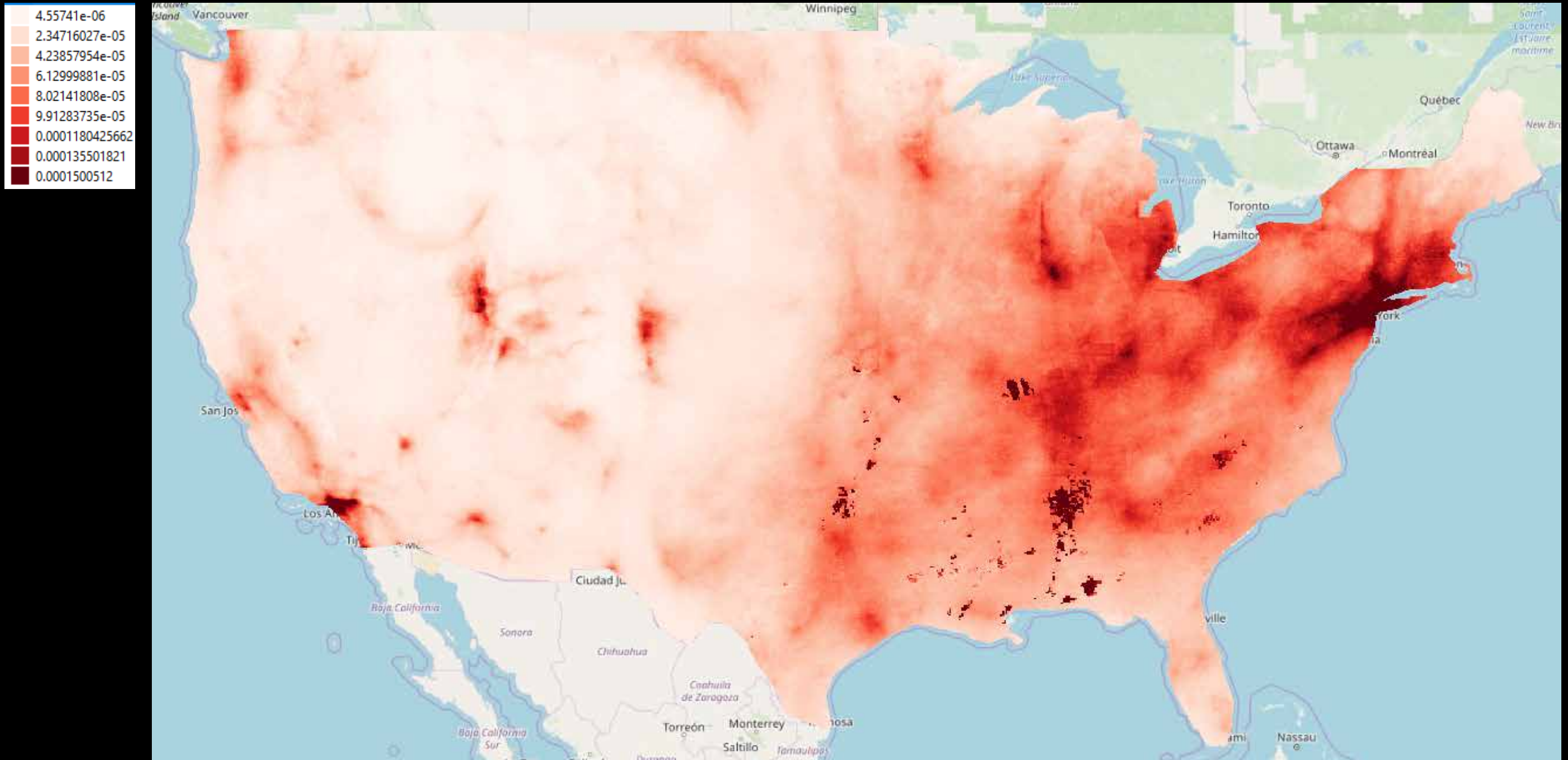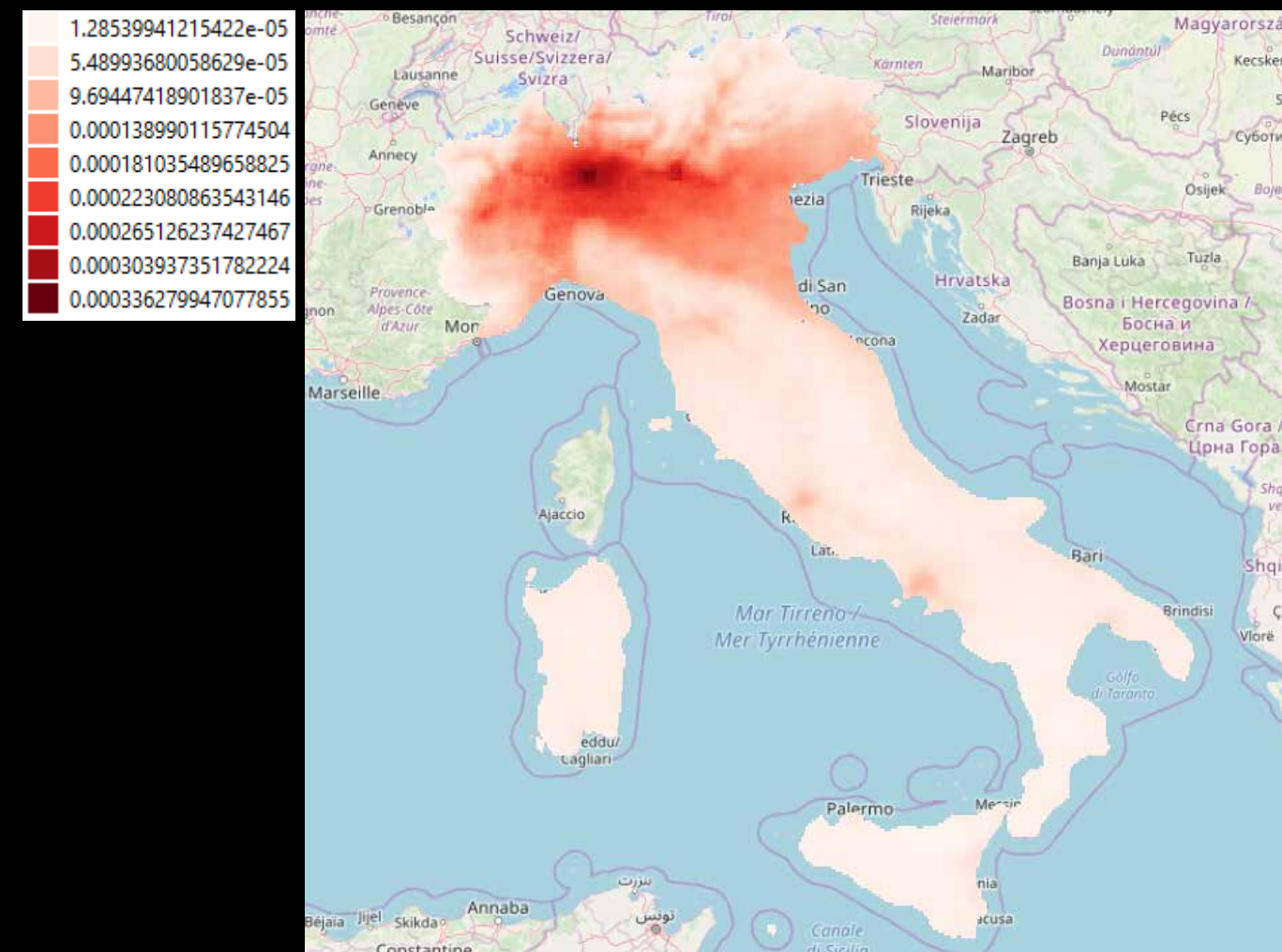
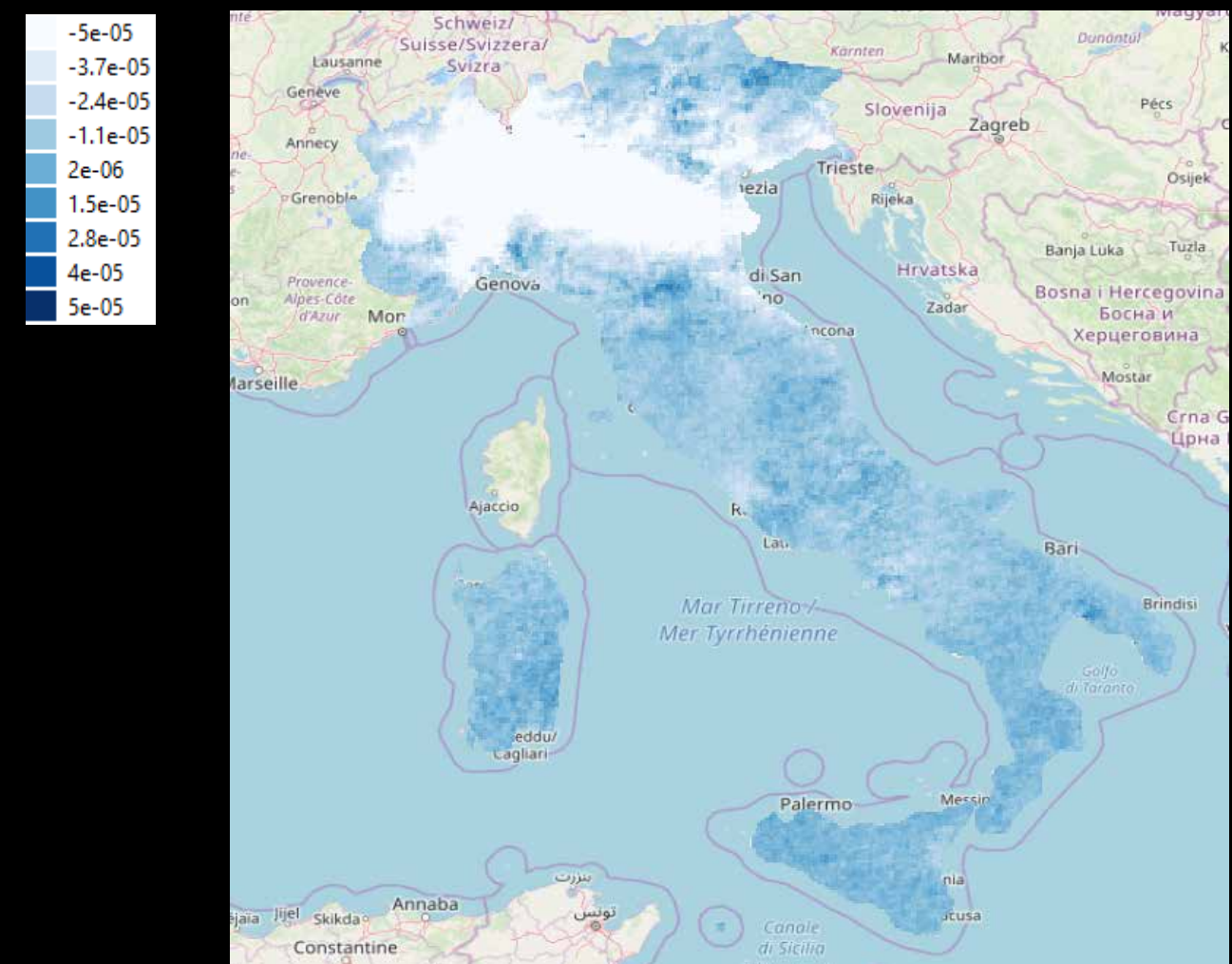# Economic Activity by NO2-last 5 days

Reduction in NO2 around quarantined areas
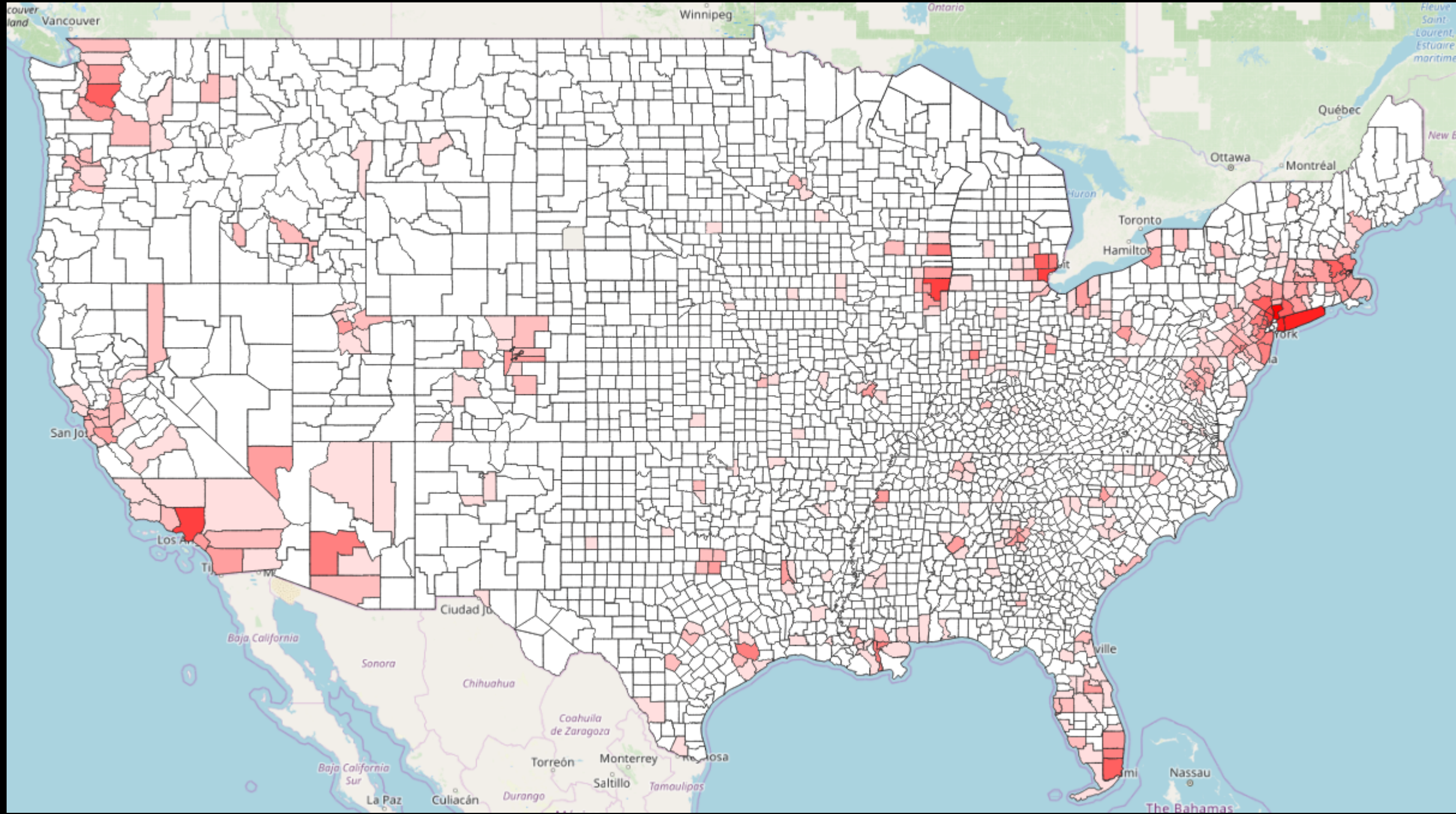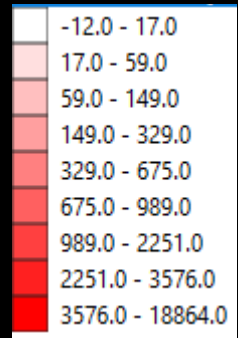
Economic activity from NO2-Italy

Average NO2-year to date
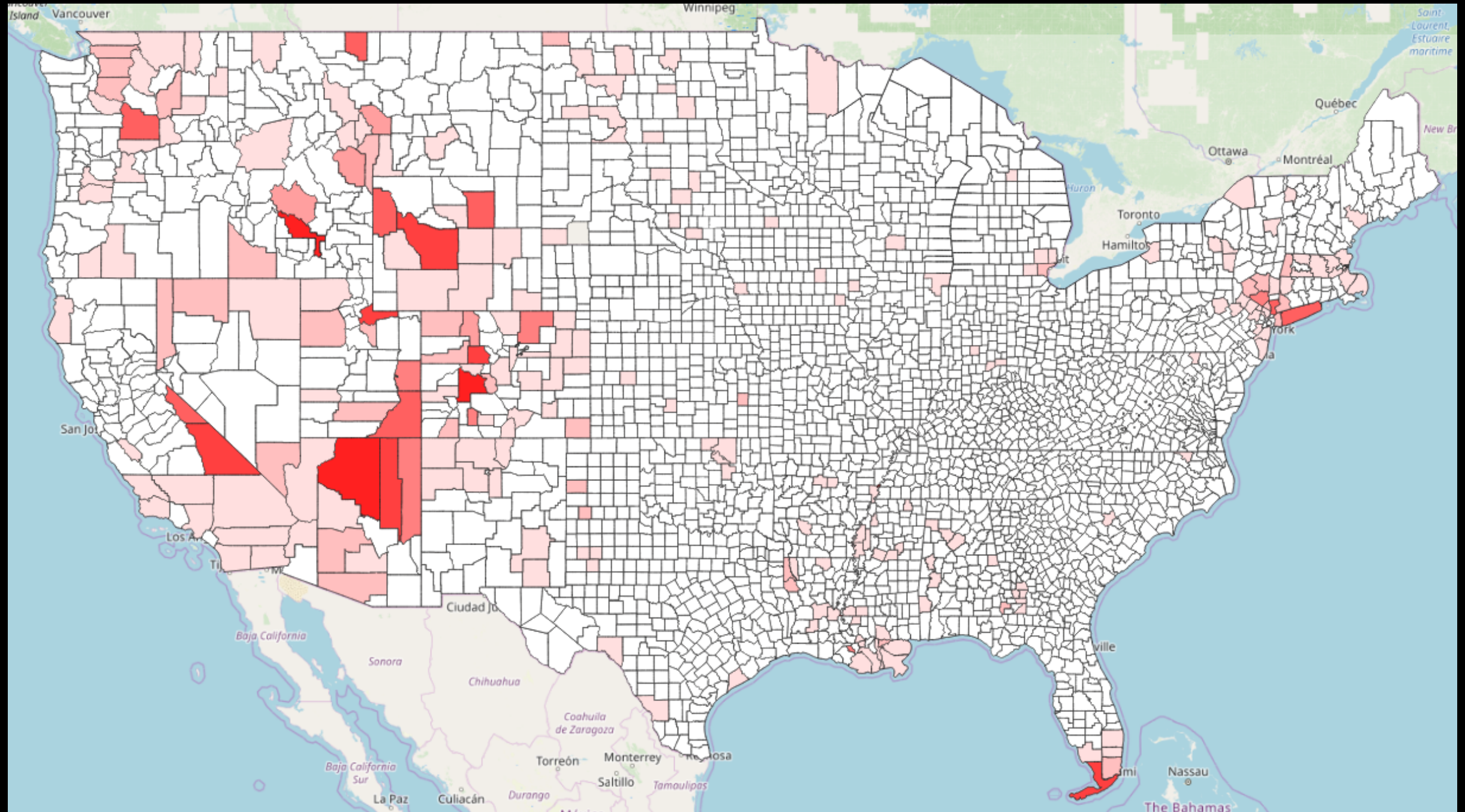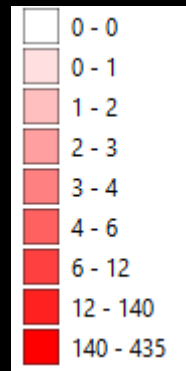
Change in NO2-last 5 days
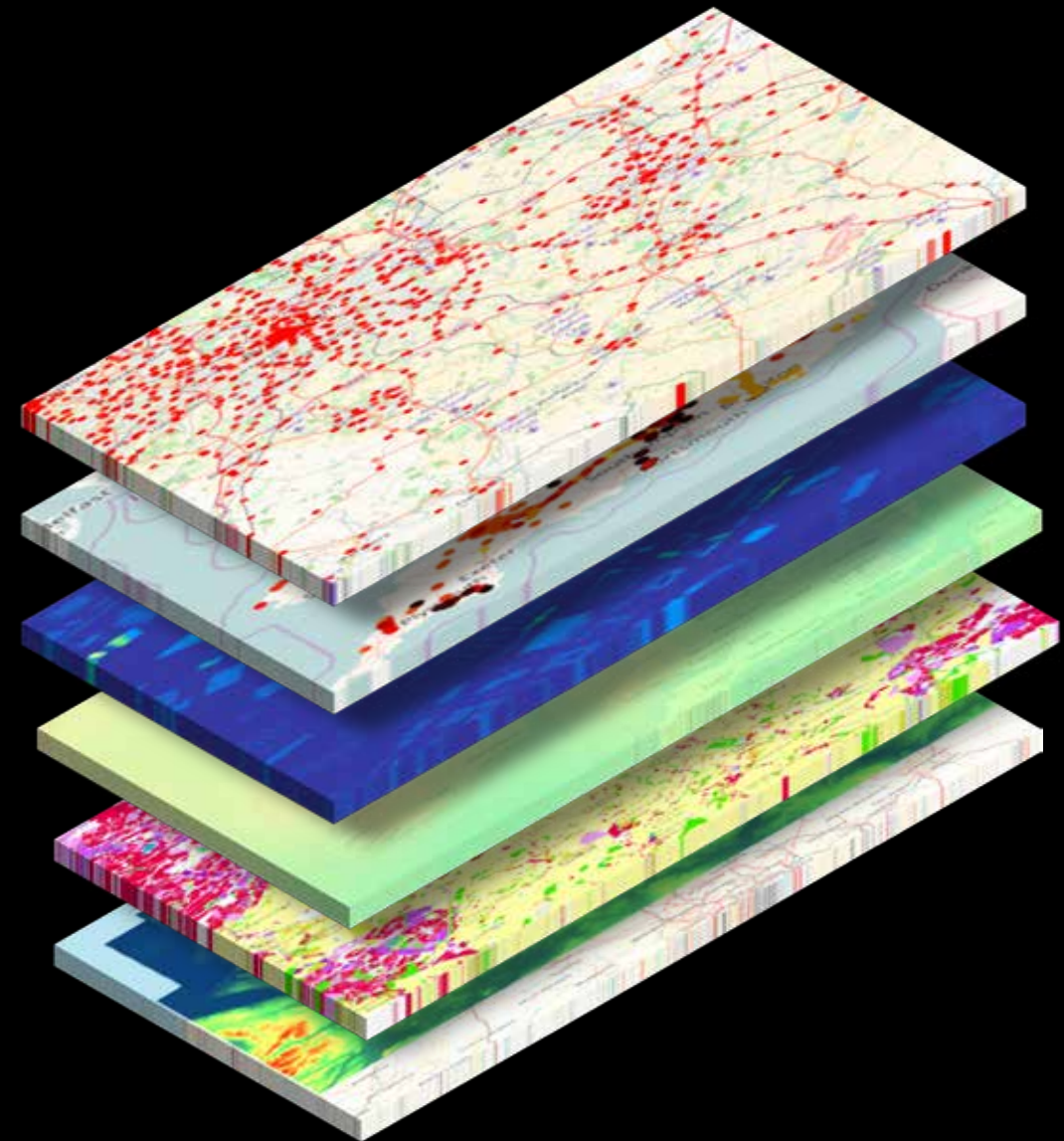
Rate of COVID cases in last 5 days

Five days rate of change per population density

Geospatial-temporal analytics on heterogeneous data

Combination of heterogeneous
"Geospatial-temporal" can enable large
scale analytics:
- ü Satellite
- ü Drones
- ü IoT/Mobile device
- ü Radar
- ü Lidar
- ü IoT/Sensors
- ü Weather
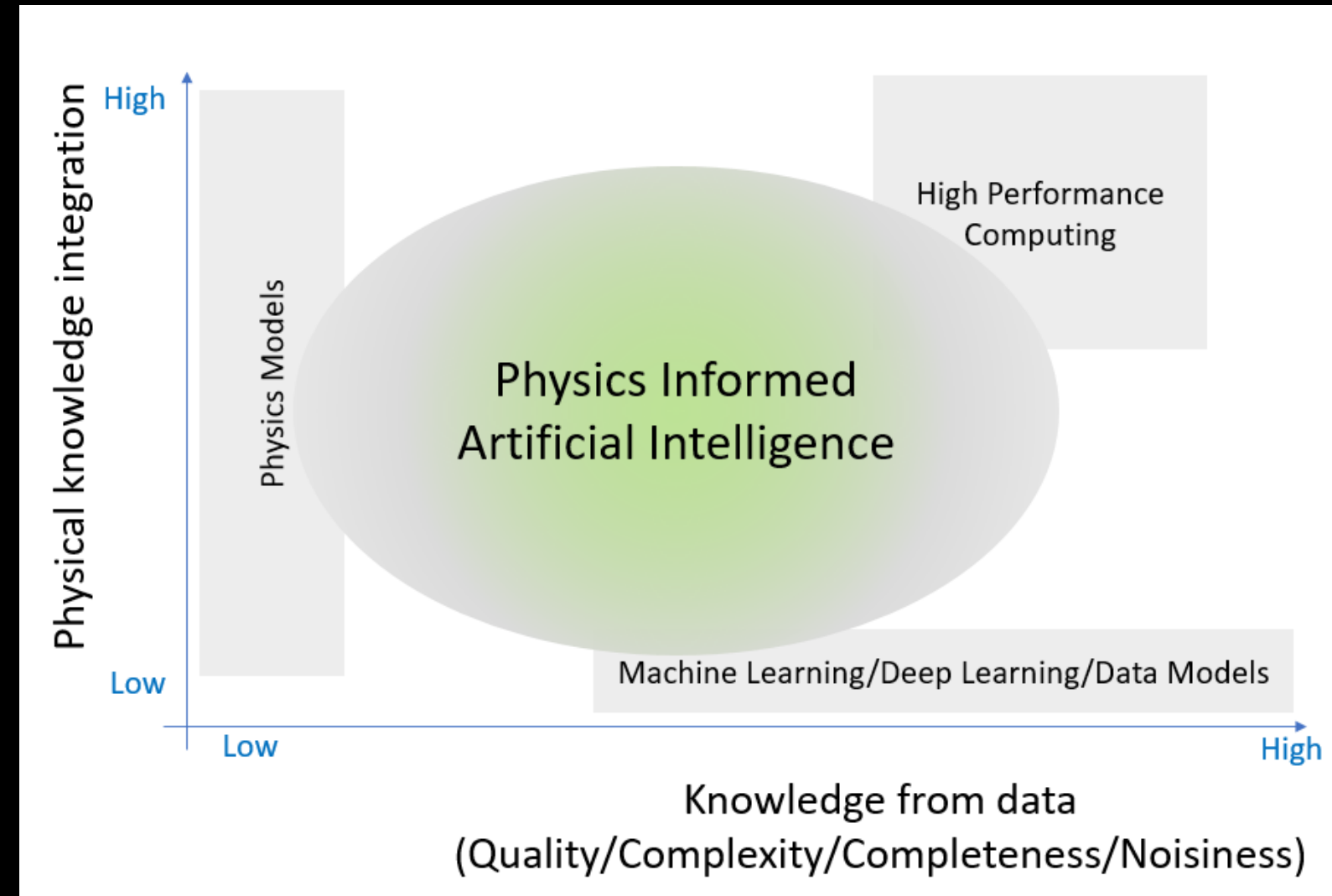- ü Survey
- ü Social media
- ü ....

Physics Informed Artificial Intelligence

Increase learning rate and build robustness in future predictions.
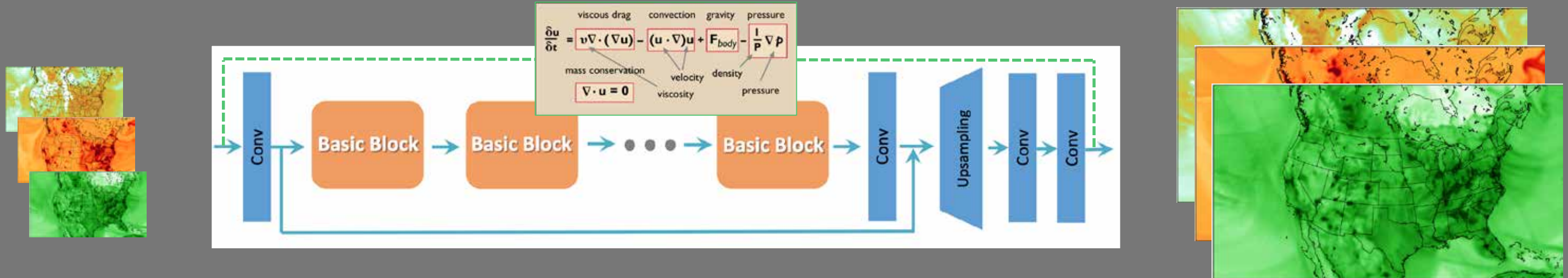
Model maintain:

- fidelity to the data
- the predictions must satisfy at conservation laws of physics
- Integrate explainability through physical laws

Data generation

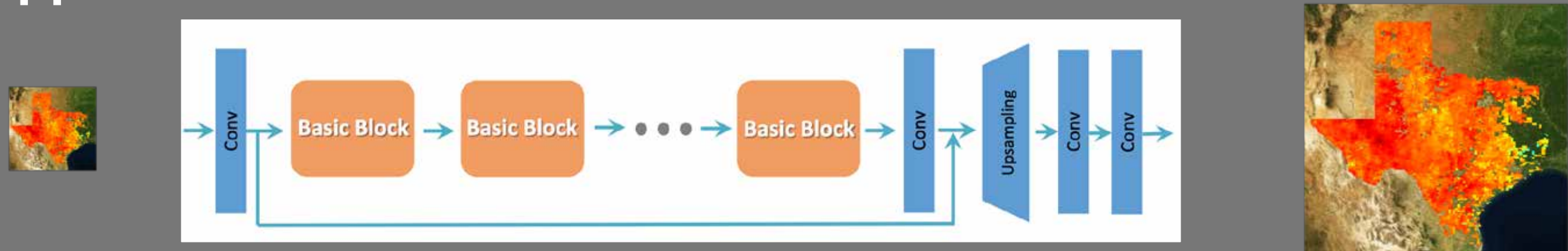- Simple plume simulation
- ERA5 weather data
- NASA GMAO Composition Forecast (https://fluid.nccs.nasa.gov/cf/classic_geos_cf/)
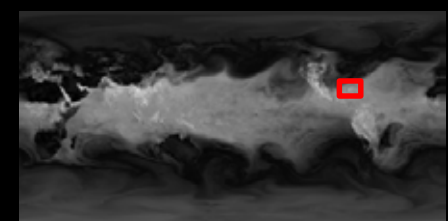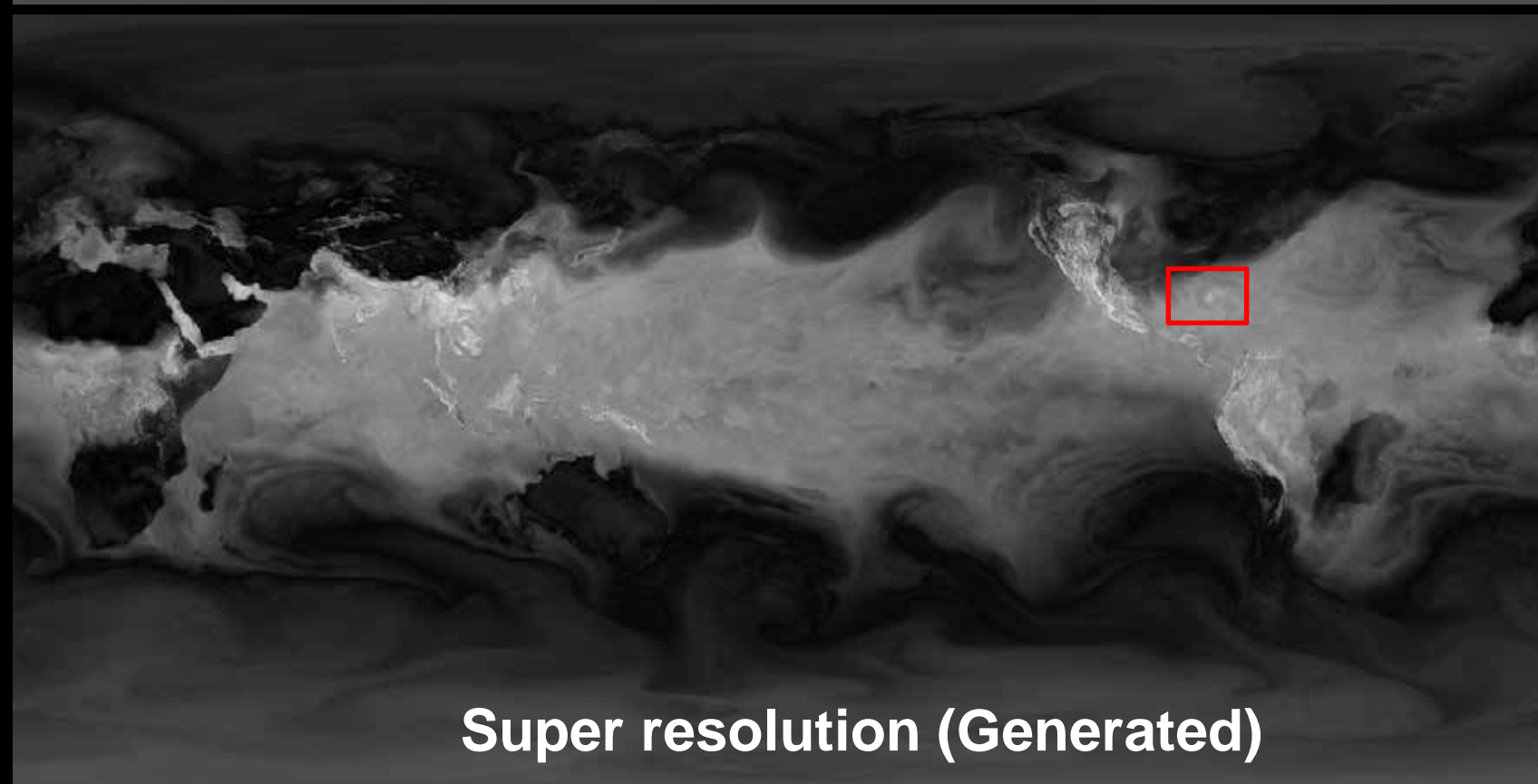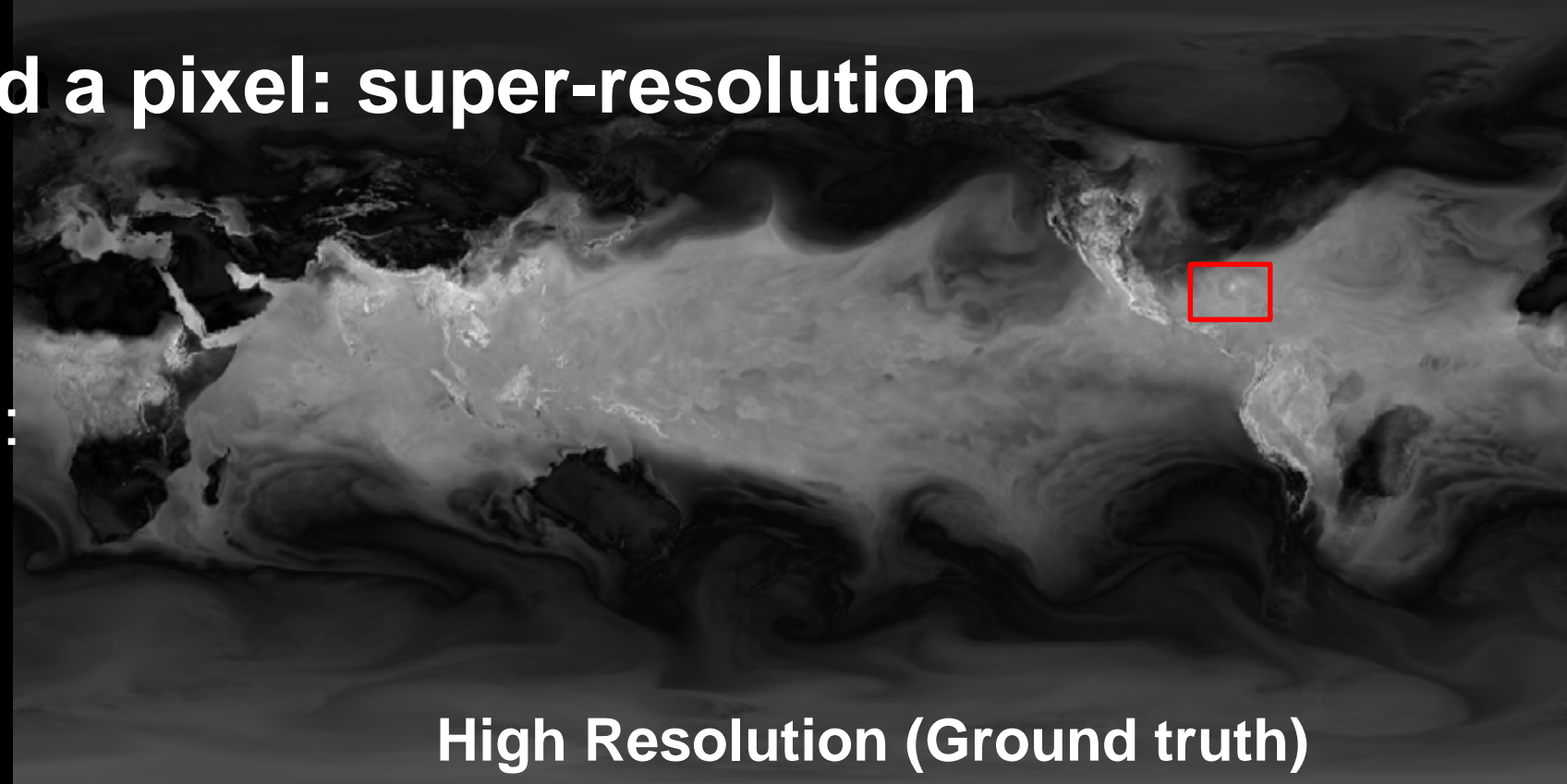
PIAI model training

Application

Seeing beyond a pixel: super-resolution

Physics Constrains:
-Bernoulli equation
-surface friction
-Navier-Stokes
equation…

High Resolution (Ground truth)

Low Resolution (Input)

Super resolution (Generated)

HR

SR

LR
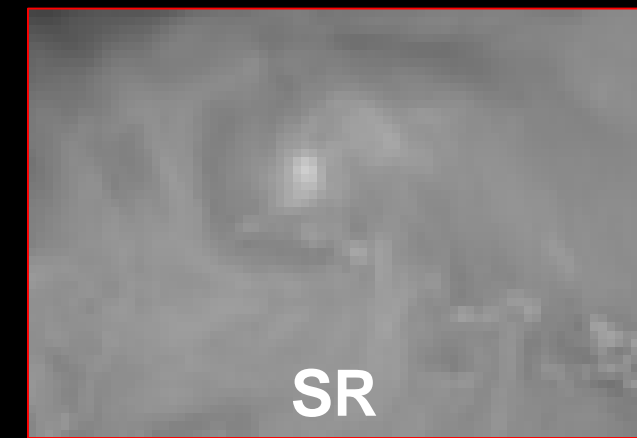
Vegetation and Land Use Mapping

Prevent power outages and wildfires

Multidimensional Data

Optical satellites (2D)　　Synthetic Aperture Radar (2.5D)　　Lidar (3D)

- Land classification
- Delineate trees vs grass
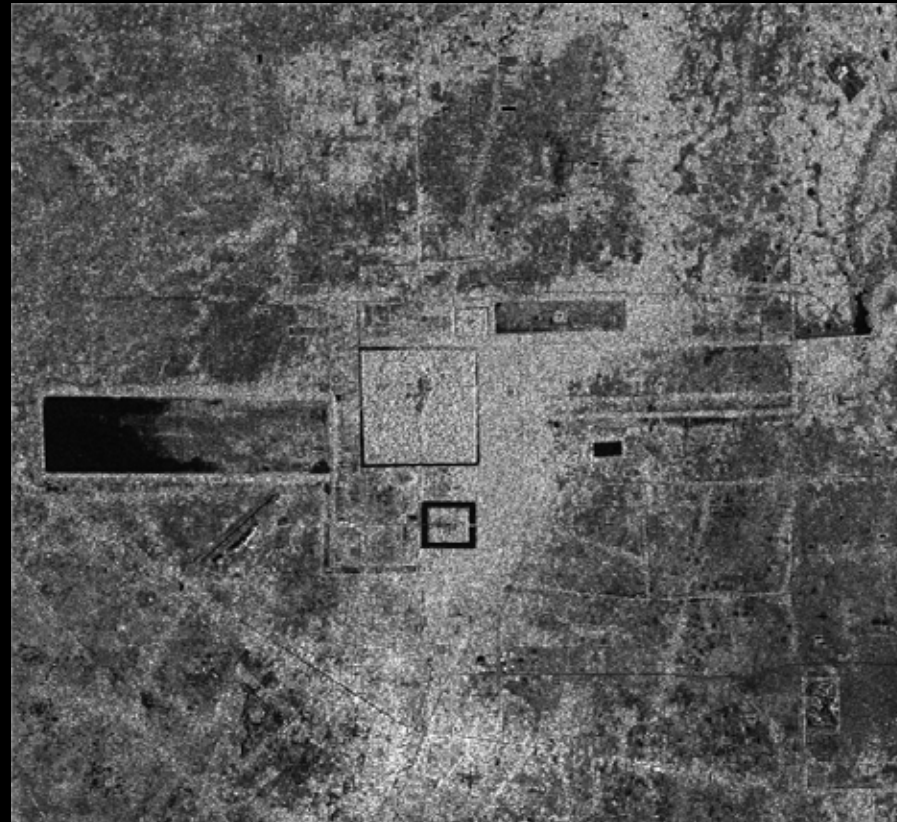
- Change detection
- Classification

- Height calculations
- Object identification
  (power lines, vegetation)

Tree species recognition

Sample Images

Tree classes – 6

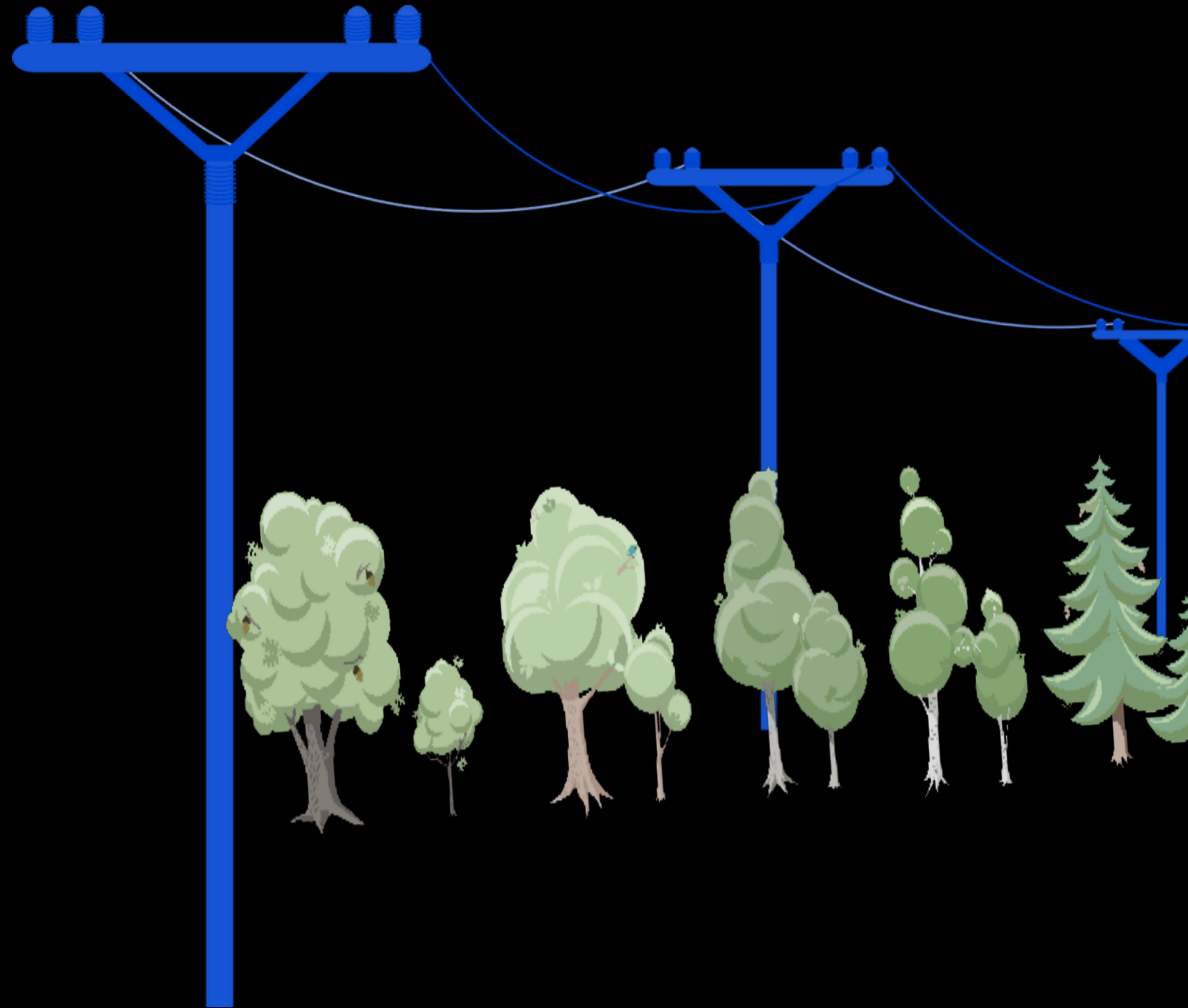| Class	Num
26	17743
28	5014
307	14544
41	3854
61	3990
72	4697

Normalized confusion matrix
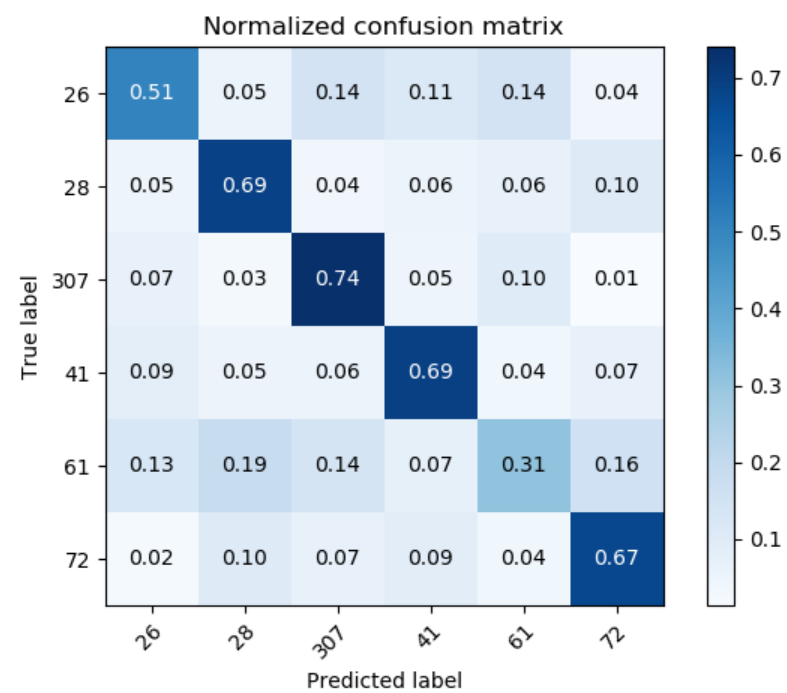
Confusion Matrix
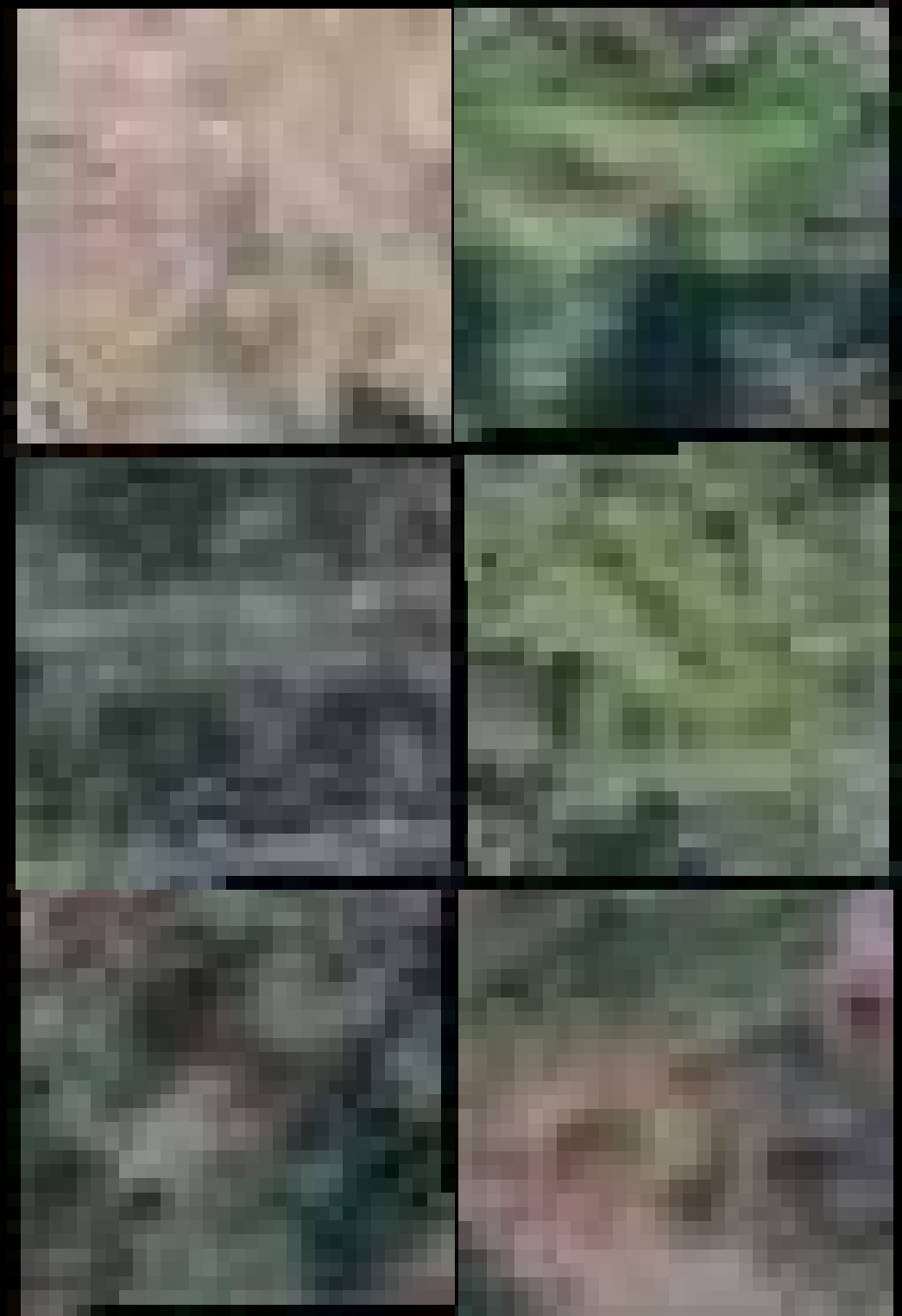
Tree species and tree health detection from multi dimensional data

Tree height (m)

0.1	
4	
8	
12	
16	
20	

Transmission power lines
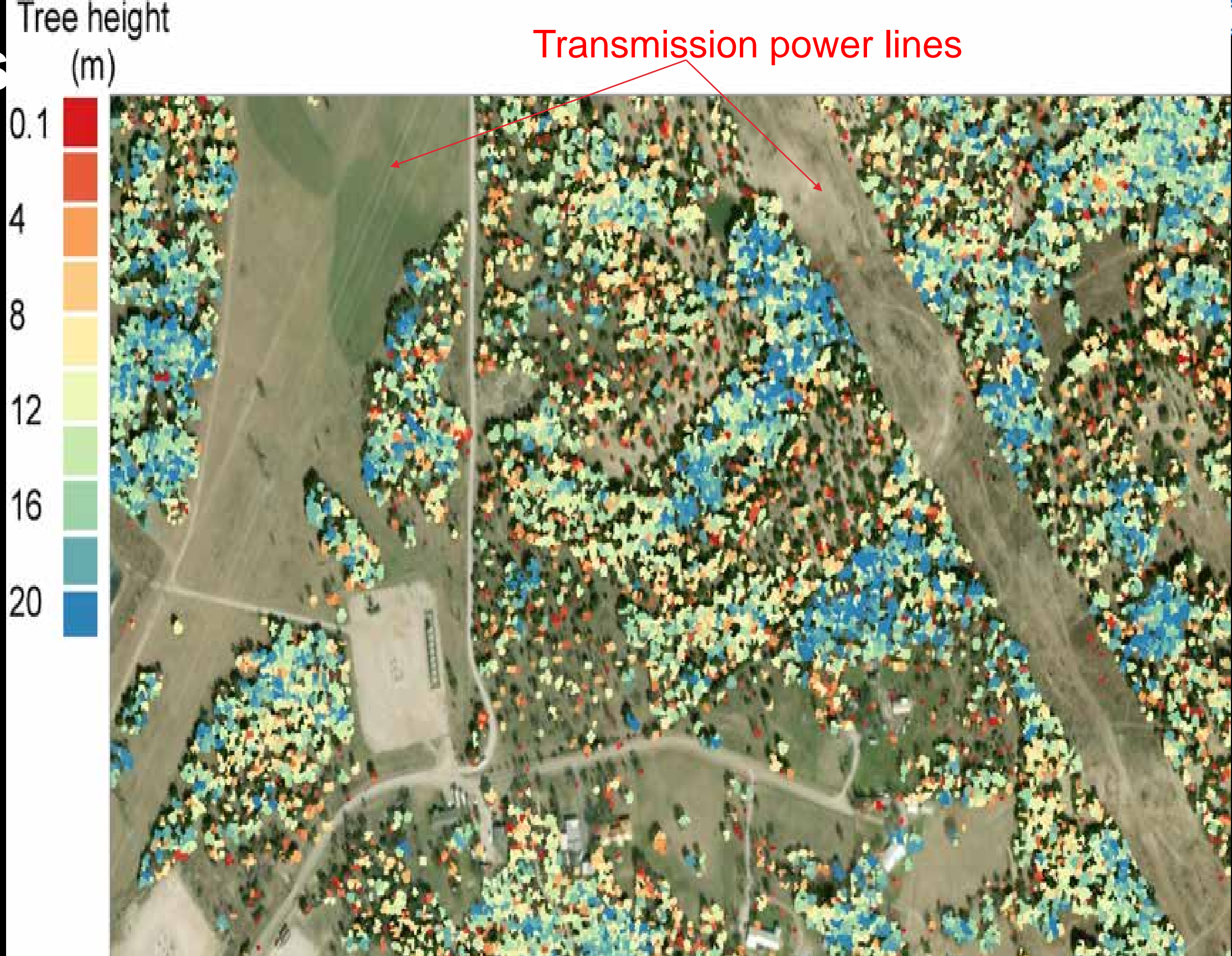
Isolate Vegetation from satellite and Lidar Data around power lines

Minimize outage prediction

Conclusion

Data from the physical world is the fastest growing dataset allowing constant monitoring of Earth surface.

Combination of big data and AI provides unique opportunities:

o Big data improves AI and AI can improve data.
o AI powered data can providing science-based decision support based on explainable models (example: emission source mitigation).
o Big Data and AI can enable the next big challenge : climate change and its local impact.

Participants

Levente Klein

Ted van Kessel

Conrad Albrecht

Marcus Freitag

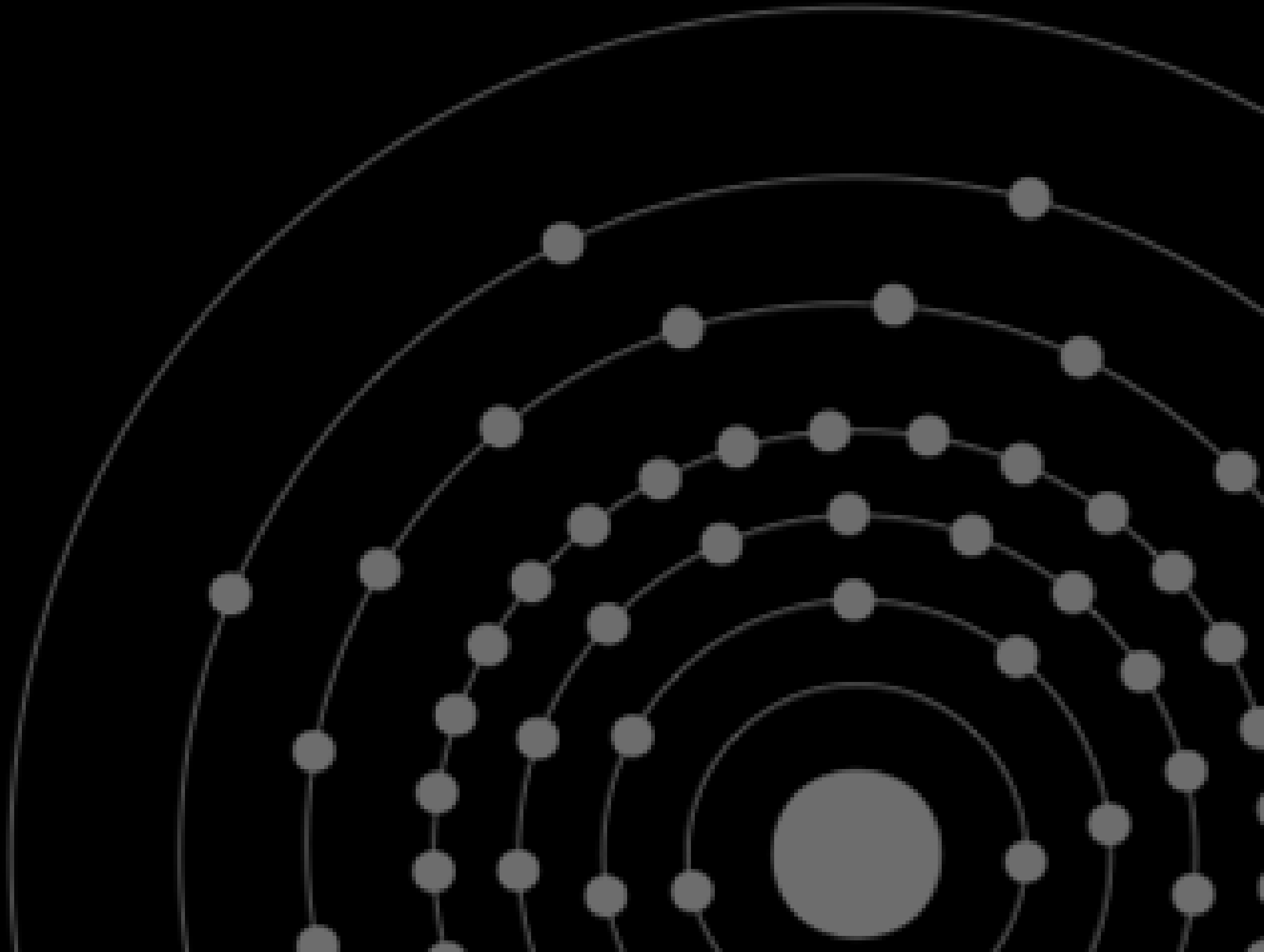Johannes Schmude

Carlo Siebenschuh

Rui Zang

Wang Zhou

Siyuan Lu

Hendrik Hamann

Bruce Elmegreen….

IBM PAIRS team

ML that respects laws of physics (e.g. consistency model-data, convection, advection, mass and energy conservation)