# DETECTION AND CHARACTERISATION OF POLLUTANT ASSETS WITH AI AND EO TO PRIORITISE GREEN INVESTMENTS: THE GEOASSET FRAMEWORK

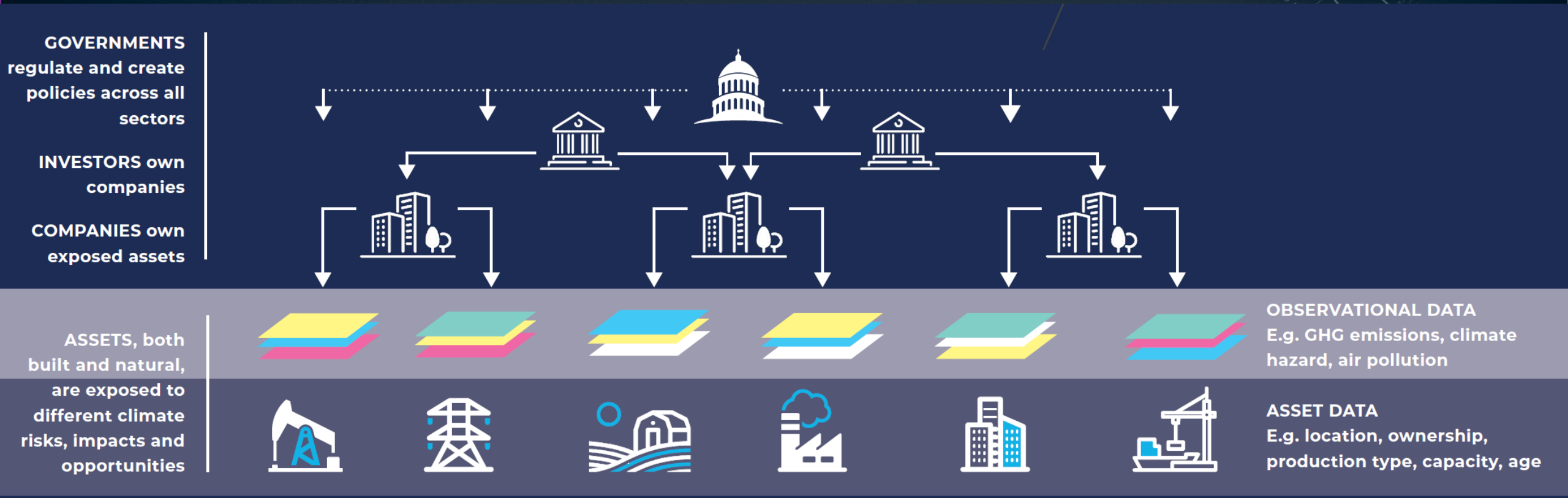
CRISTIAN ROSSI[1,2], NATALIYA TKACHENKO[2], MARAL BAYARAA[1,2], PETER FOSTER[3], STEVEN REECE[2], KIMBERLY SCOTT[4], CHRISTOPHE CHRISTIAEN[2], MATTHEW MCCARTEN[2]

[1]SATELLITE APPLICATIONS CATAPULT [2]UNIVERSITY OF OXFORD [3]ALAN TOURING INSTITUTE [4]ASTRAEA

Living Planet Symposium 2022, 23/05/22

# SPATIAL FINANCE AND ASSET LEVEL DATA

Spatial finance is the integration of geospatial data and analysis into financial theory and practice



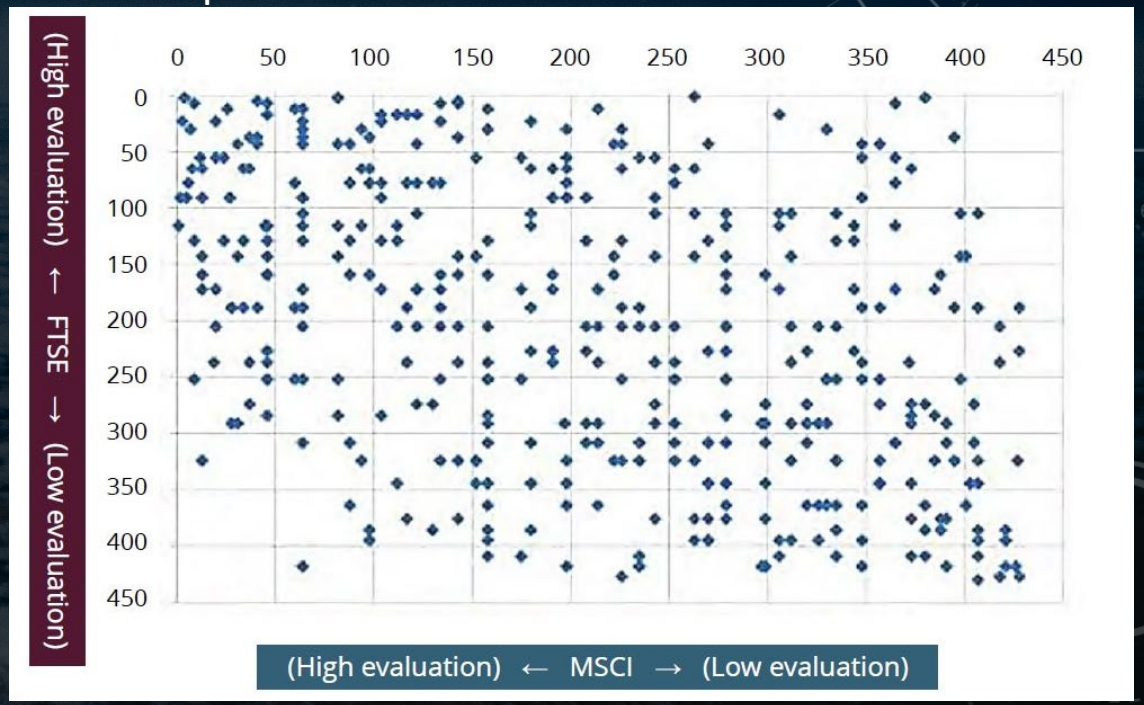
# DATA CHALLENGES

Main source of data on a company's sustainability performance is its voluntarily reported information or 'disclosures'.

This comes with various challenges:

- Information is **self-disclosed** and typically compiled top-down
- ESG reporting is not mandatory, which means **not all companies report**
- ESG reporting not standardised, which means **data is not comparable**
- Annual reporting means ESG **data is outdated** once released

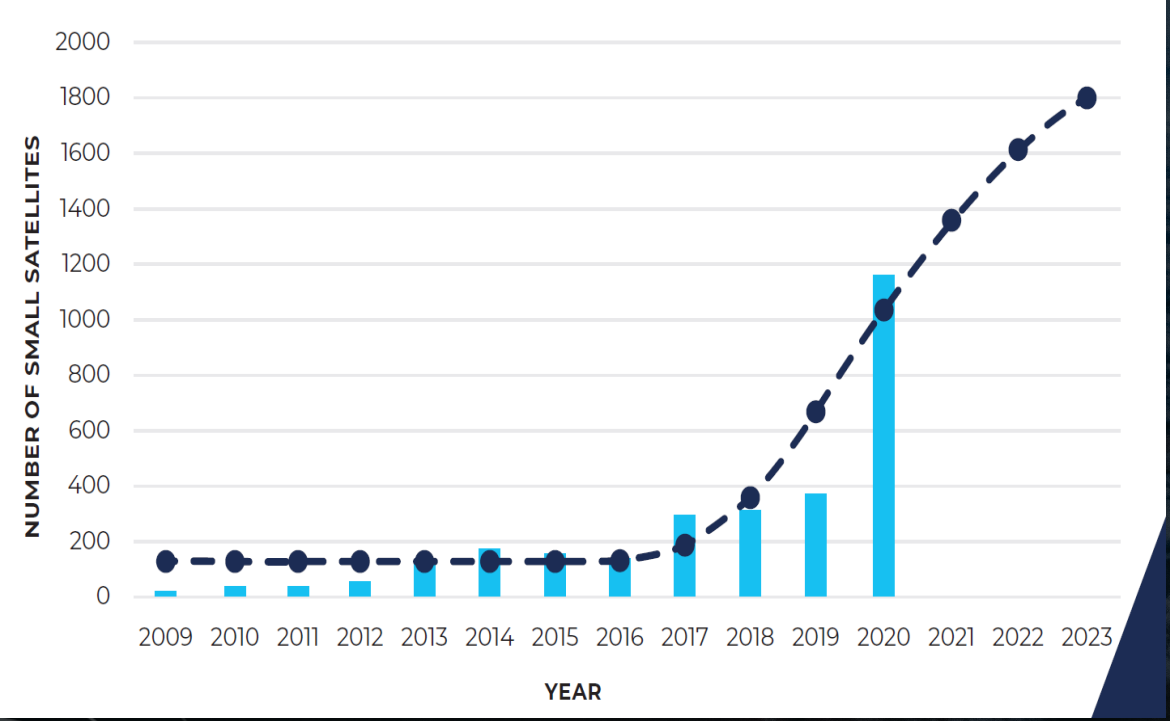
Comparison of ESG scores from FTSE and MSCI



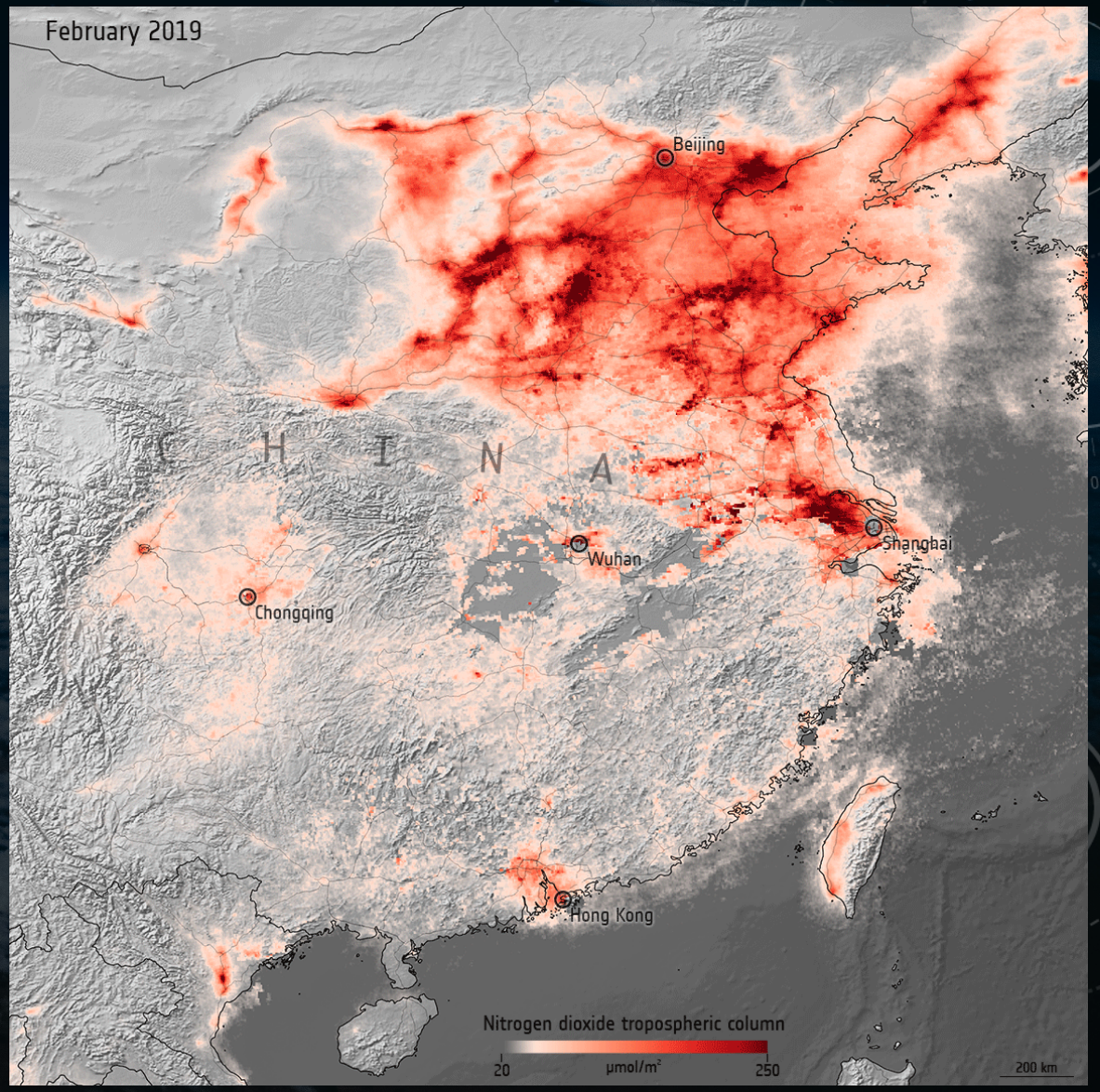

Source: CLSA/ACGA

*"… this chart does not discredit ESG data or the practice of scoring … it underscores the danger of relying on a simple final score for investment decisions"*

# OBSERVATIONAL DATA




Number of small satellites launched, historical (columns) and modelled (line)
Source: Satellite Applications Catapult




Nitrogen dioxide emissions over China in 2019-2021
Source: European Space Agency

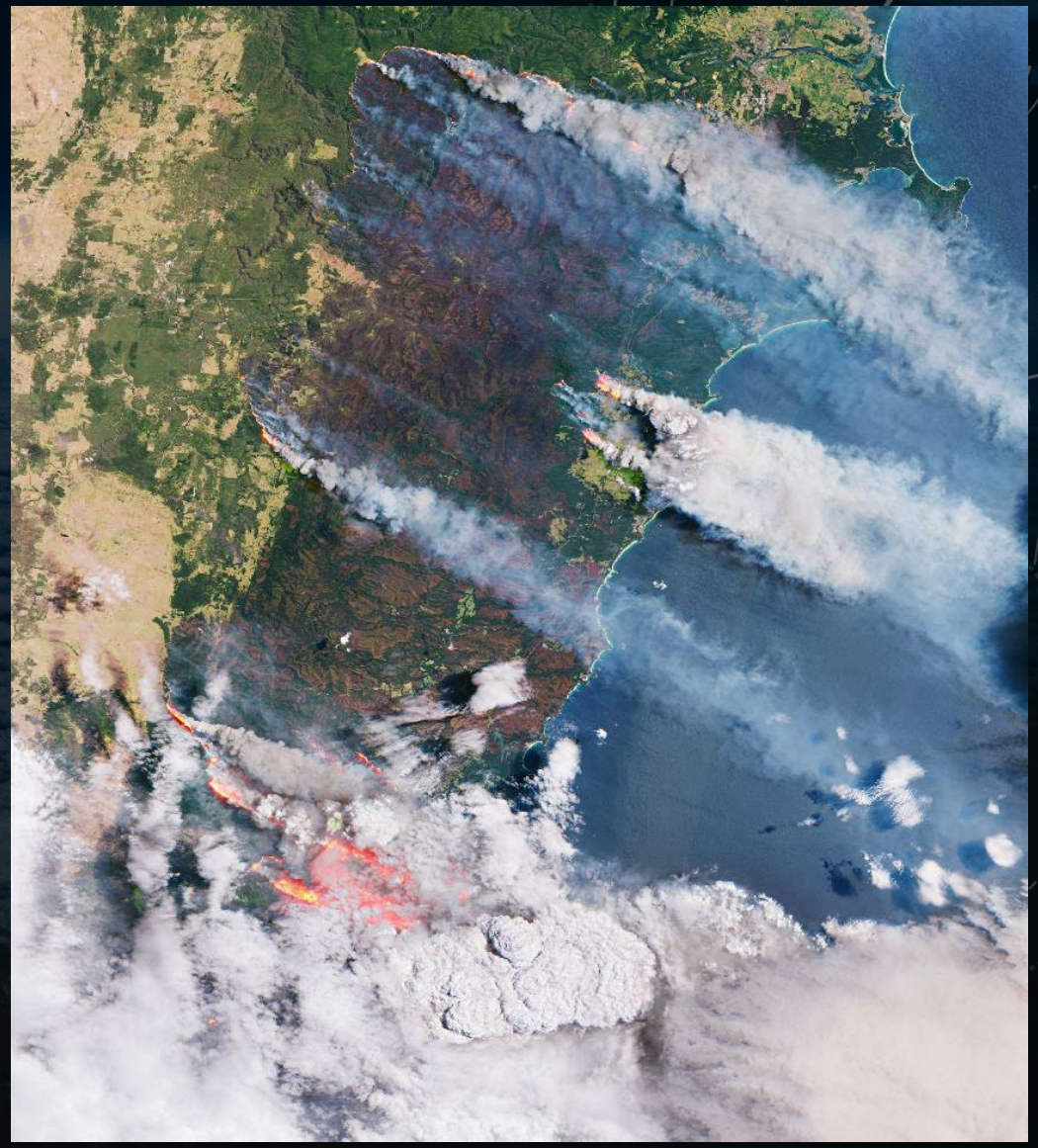
# SPATIAL FINANCE RELEVANCE

Spatial finance allows for a **bottom-up** understanding of **risks, opportunities and impacts**, driven by

- Neutral and objective source of information
- Data that is more timely and comparable
- Connecting financial system with real economy

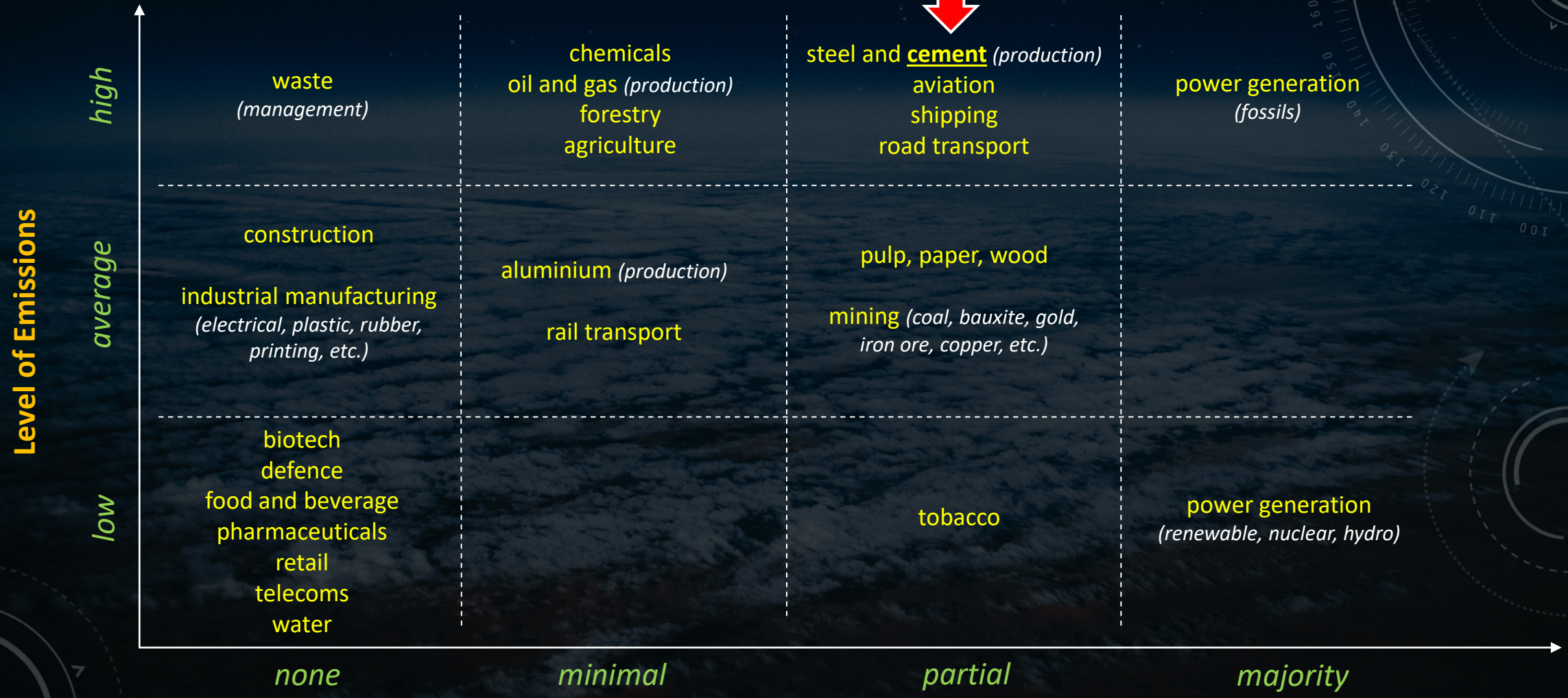
We are seeing an increasing uptake of spatial finance approaches and analytics, driven by

- Climate-related financial risk assessment and disclosure
- Nature-related financial risks and impact assessment
- Investor engagement campaigns on sectoral ESG issues



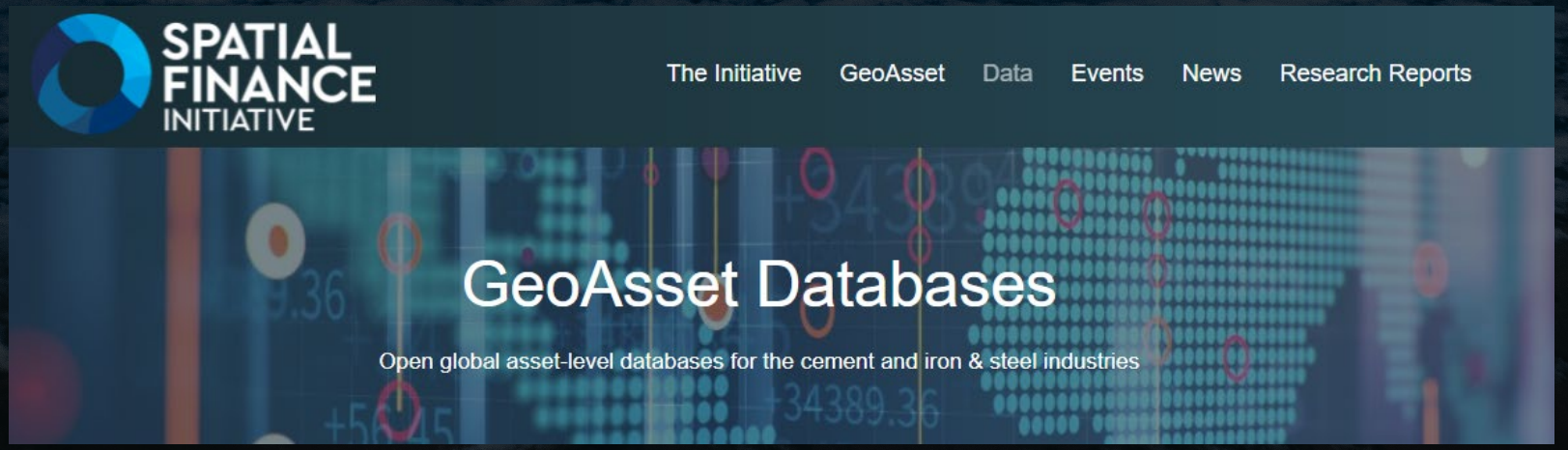
Source: European Space Agency

# MISSING DATA



# GEOASSET FRAMEWORK

- GeoAsset is a collaborative endeavour to provide accurate, comprehensive, comparable foundational data on physical assets across all major sectors, tied to financial ownership information, at low cost and as in as open a format as possible.
- Part of the new UK Centre for Greening Finance & Investment. Core partners are Oxford Sustainable Finance Programme, The Alan Turing Institute, and the Satellite Applications Catapult.




Datasets are available to download from:
**https://spatialfinanceinitiative.com/geoasset-project/data/**

# BUILDING THE CEMENT DATASET: HIGH-LEVEL ARCHITECTURE

# EXISTING DATASETS CURATION = TRAINING DATASET


Coordinate set 1

- Cement tag
- Incomplete

Coordinate set 2

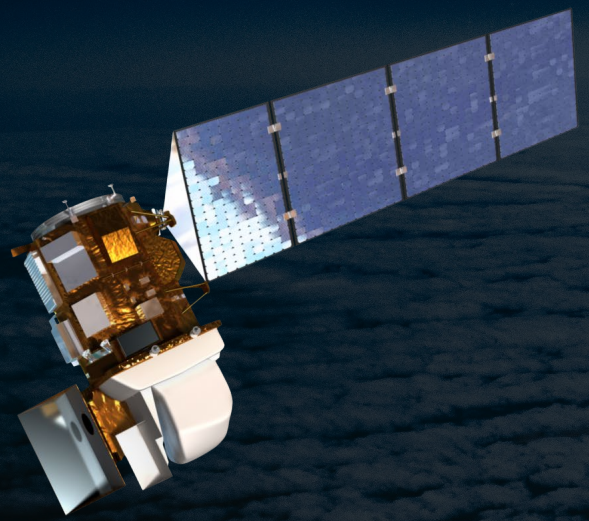
- PDF format
- Incomplete

Coordinate verification and refinement

Ownership identification

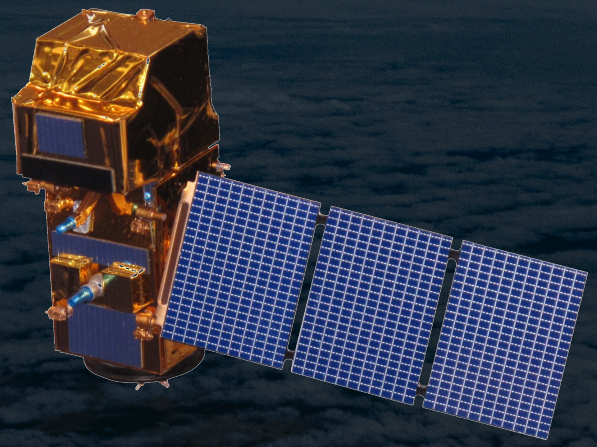
- Company disclosure of major producers
- Disclosure of capacity and starting production year
- Ownership verification with OpenPermID database

# REMOTE SENSING DATASETS



**Landsat-8 (©USGS)**

Open Access, 100m (thermal)



**Sentinel-2 (©ESA)**

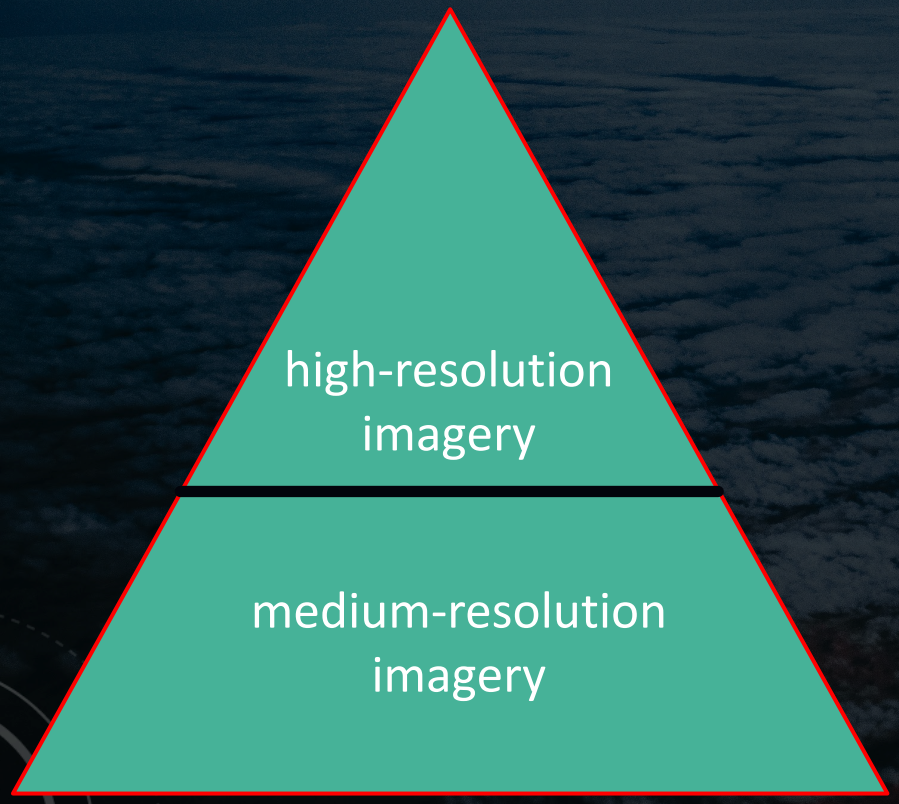
Open Access, 10m

**WorldView-3 (©Maxar)**

Commercial, 1.5m

# MACRO-LOCALISATION VS MICRO-LOCALISATION

The asset localisation can be solved in a two-steps approach: a macro-localisation, that focuses on the broader area around the site, and a site-localisation, that focuses on the asset localisation



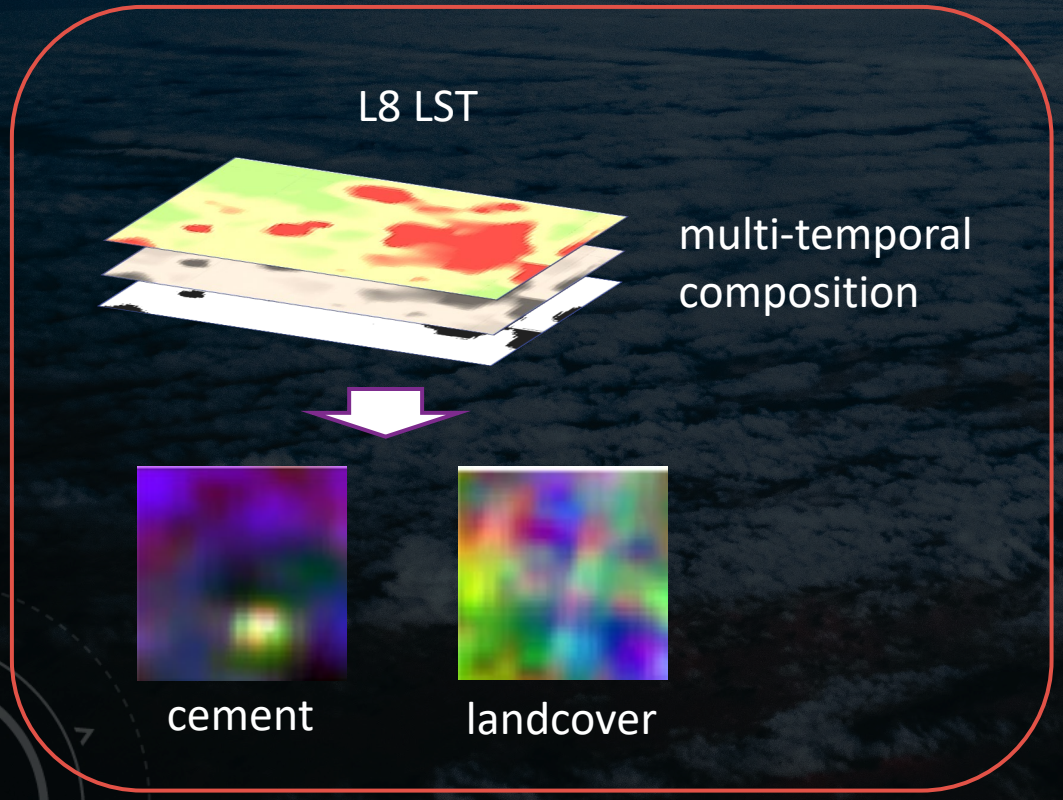

➔ Test phase on the final approach ➔

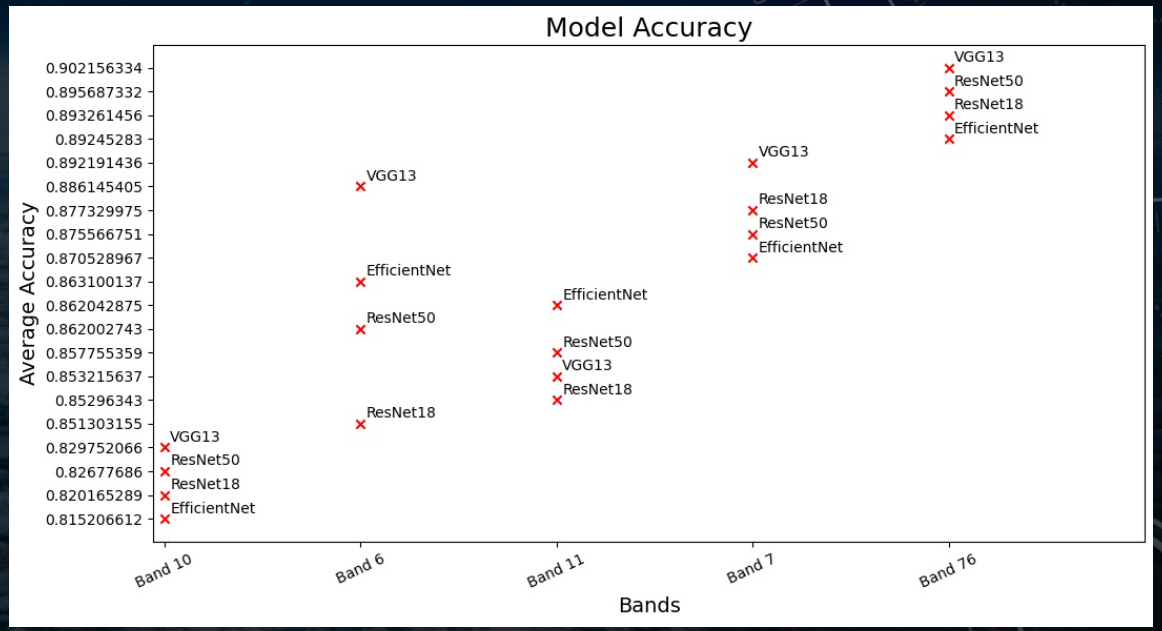
Computer vision/collective intelligence exploration from commercial imagery (Maxar)

Spectral/CV exploration from open access imagery (Landsat-8/Sentinel-2)

# PHYSICS-AWARE AI: LANDSAT-8

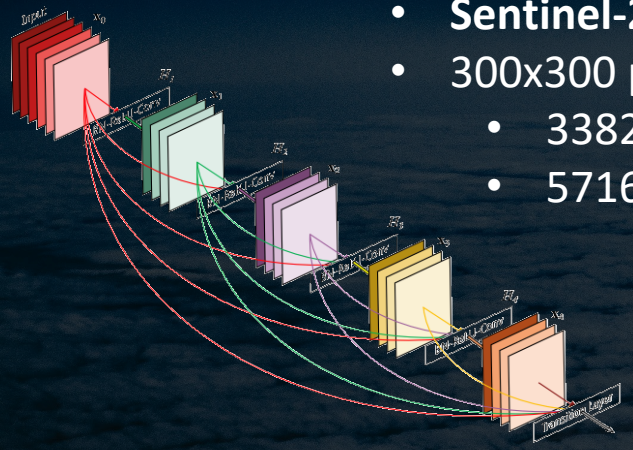
- Exploration of **Landsat-8** thermal infrared / SWIR bands / iron-oxide ratio




- 4 DL architectures tested
- Best cement plant detection results with band ratio
- Still under research, noy yet deployed

# MACRO-LOCALISATION: DENSENET



- **Sentinel-2** RGB band only
- 300x300 pixels chip extraction (3x3km)
  - 3382 cement chips
  - 5716 landcover chips

- Several architectures considered, best results* with DenseNet161 *recall=coverage of actually positive sample

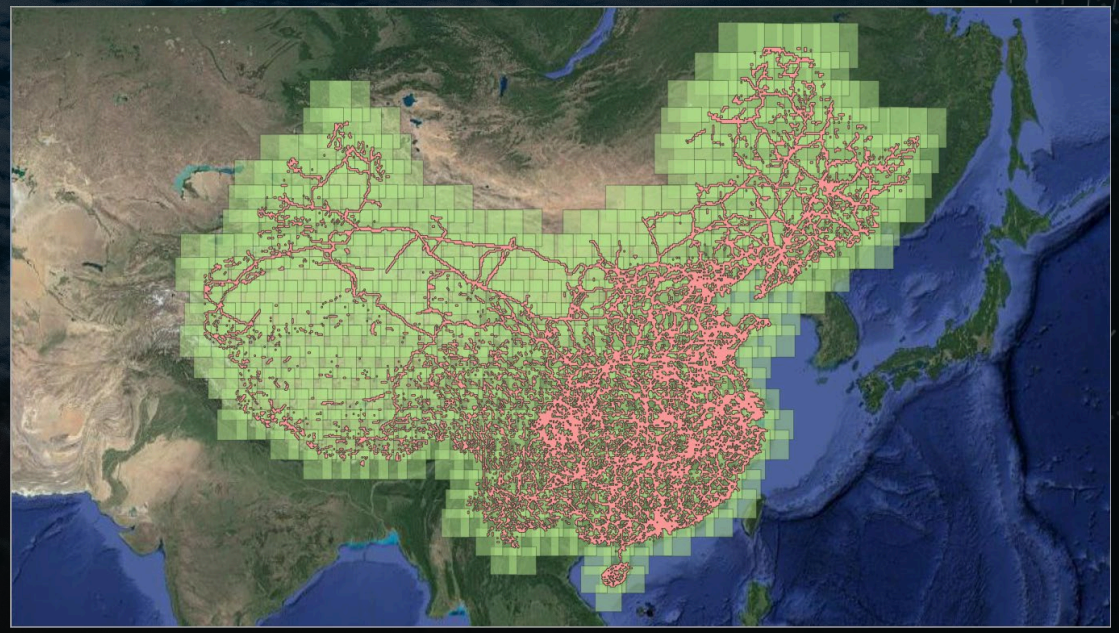
**Overall Statistics:**

Accuracy: 0.9331703341483293
Population: 1227

**Class Statistics:**

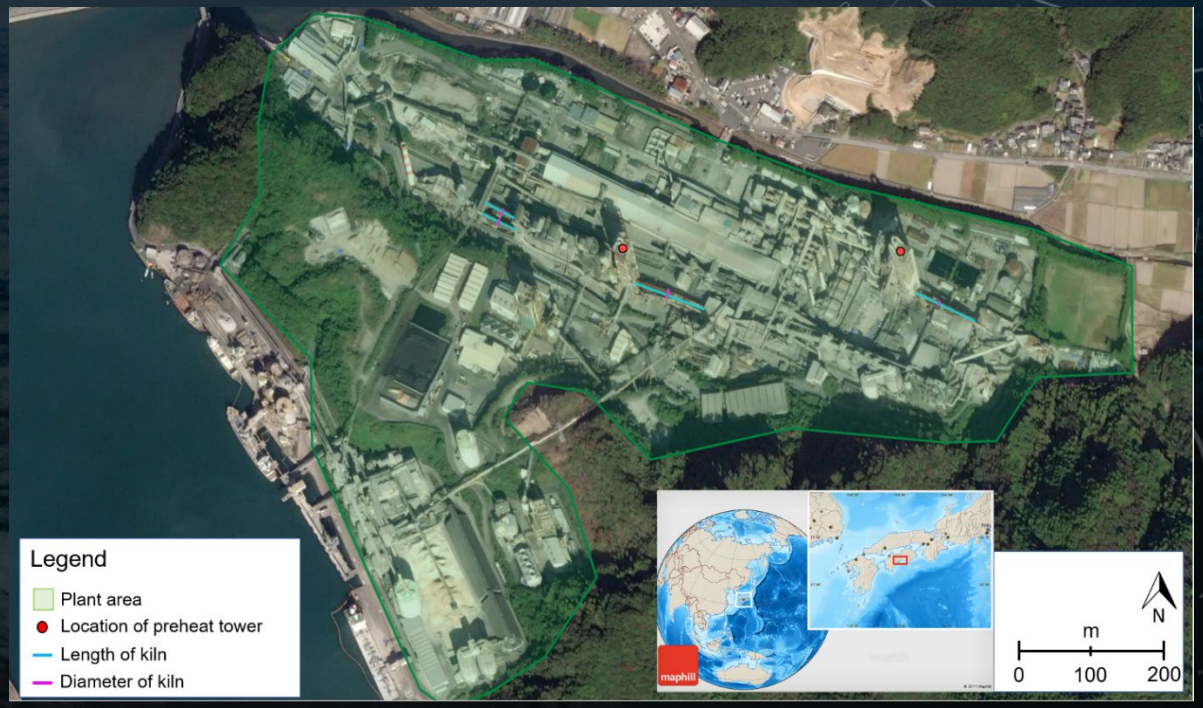
| | label | TP | FP | FN | Population | Precision | Recall |
|---|---|---|---|---|---|---|---|
| 0 | cement | 414 | 69 | 9 | 423 | 0.857143 | 0.978723 |
| 1 | landcover | 648 | 10 | 66 | 714 | 0.984802 | 0.907563 |
| 2 | steel | 83 | 3 | 7 | 90 | 0.965116 | 0.922222 |

- Model deployment over **China** (9.3M sqkm)
- Deployment within ASTRAEA EarthAI platform
- Every chip (>400k chips) is scored with the S2 RGB model (0-1: probability of cement plant)



# MICRO-LOCALISATION: COLLECTIVE INTELLIGENCE

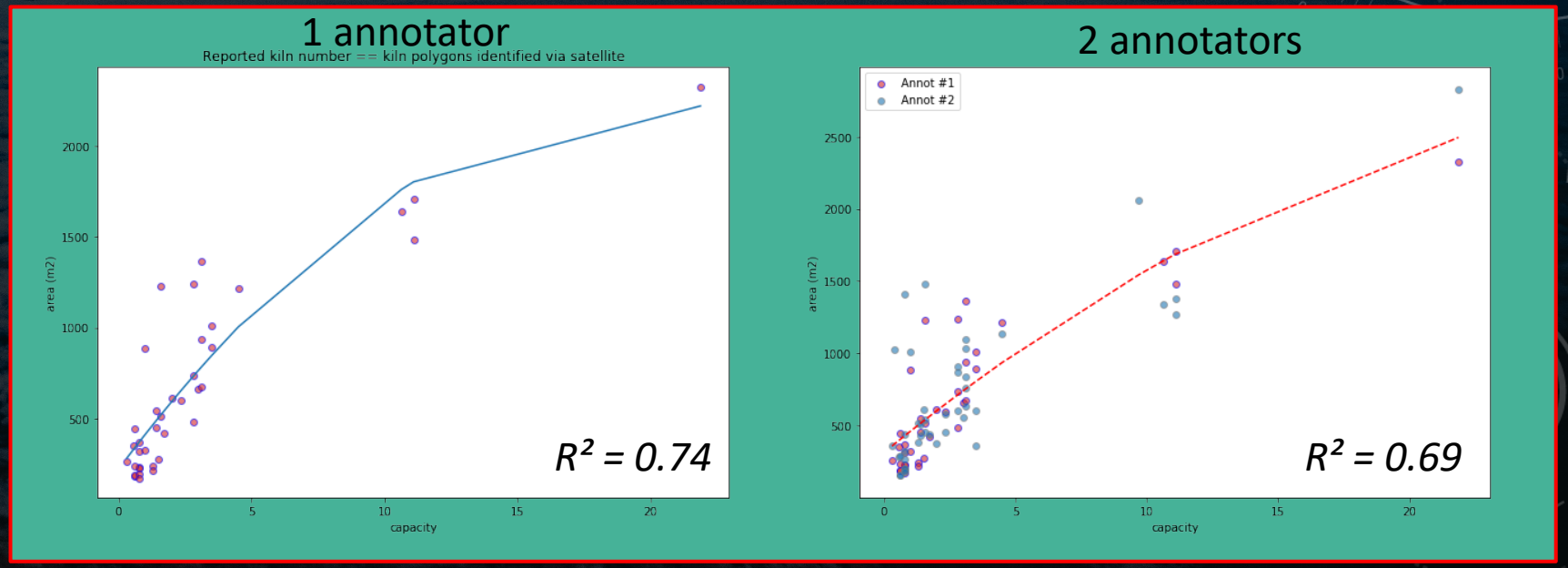
- Multiple labellers on two stages: first Google Earth Pro, and then Maxar's SecureWatch for unclear imagery


Asset annotations

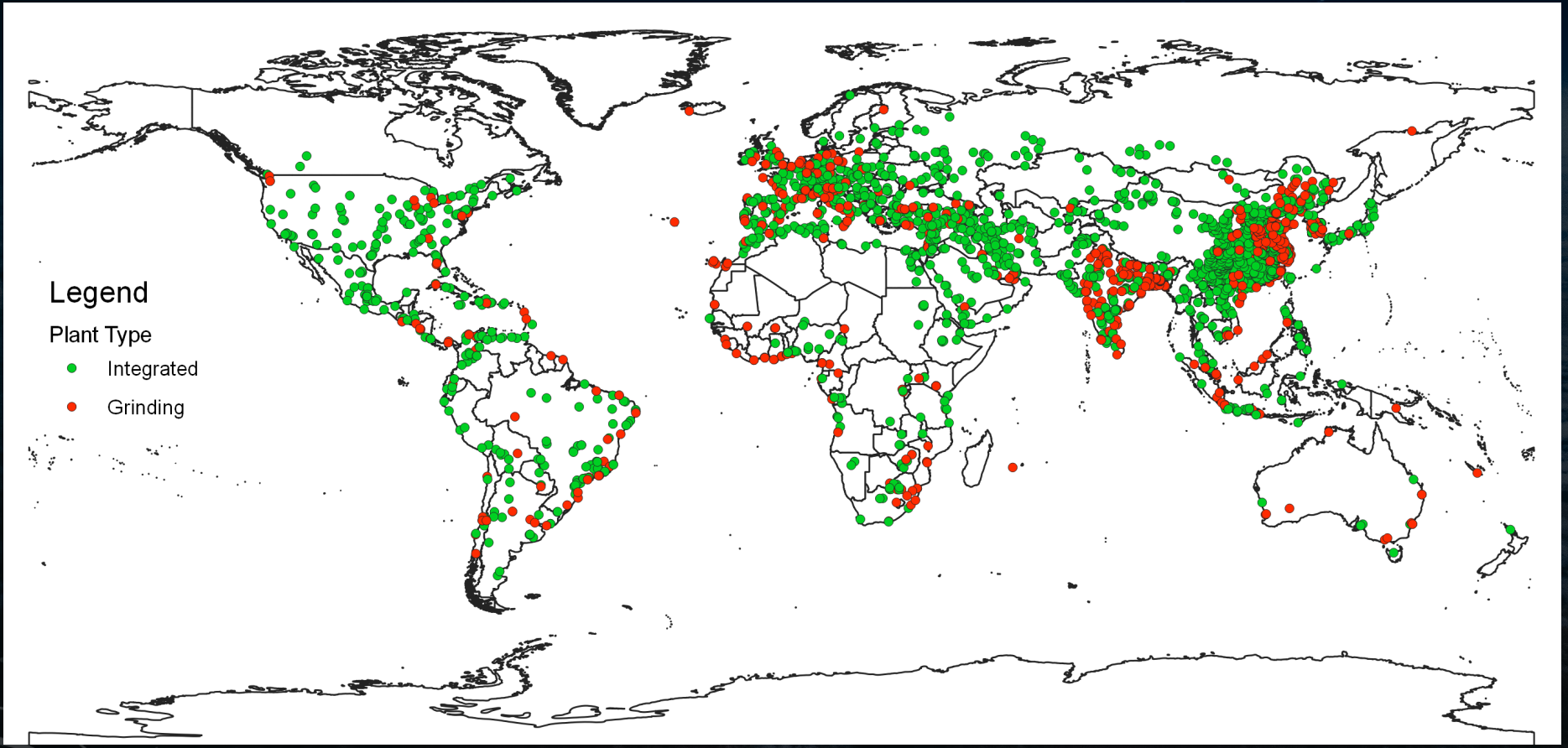
# MANUAL CORRELATION KILN VS CAPACITY

- Kiln dimensions are hypothesized to give an indication on the production capacity. To investigate this relationship, the reported plant cement capacity is compared with kiln dimensions manually annotated from EO




- Positive correlation between reported capacity and the area of kilns from manual annotation
- Uncertainty given by discrepant number of kilns in the database vs annotated ones and geometrical distortions in the imagery

# FINAL DATASET

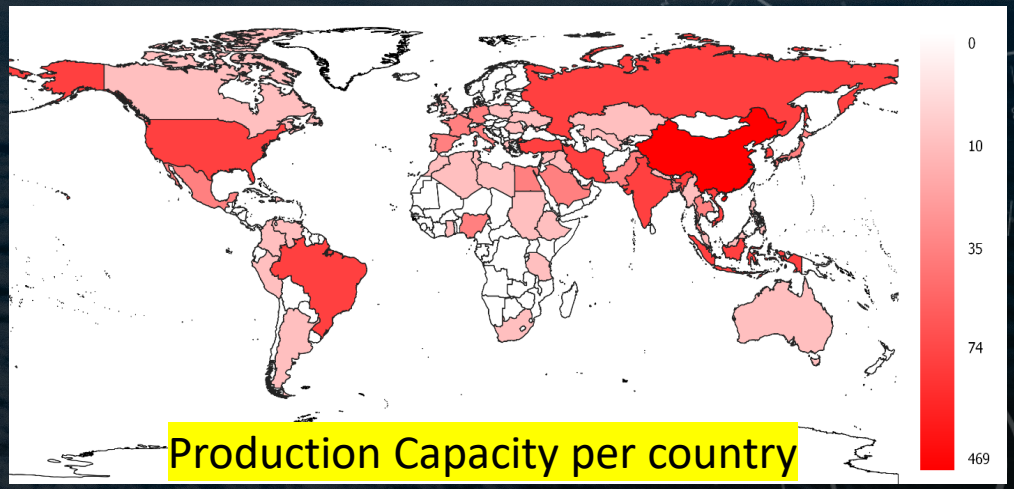



- Plant location
- Plant type
- Production type
- Plant capacity
- Year production first started
- Detailed ownership

3,117 cement production assets accounting for ~90% of global cement production capacity

# DATASET VALIDATION

- Expert knowledge workshops for an accurate methodology development and crowdsourcing instruction
- 88.9% of total cement capacity captured (5.5Mt / 6.2Mt)
  - Continent statistics within 5% error
  - Near complete coverage in EU and North America
  - Overestimation in India/Russia (it might be correct!)
  - Not all plants captured in China (-25%)
- Extensive manual validation of the Global Cement Database (GCD)
  - GeoAsset Database captures more assets globally



Production Capacity per country

| | Integrated | | Grinding | | Unknown | | Total | |
|---|---|---|---|---|---|---|---|---|
| **Dataset** | **N** | **Capacity** | **N** | **Capacity** | **N** | **Capacity** | **N** | **Capacity** |
| **GCD (Reported)** | 2,330 | 3,934.4 | 673 | 537.2 | | | 3,003 | 4,471.6 |
| **GCD (Validated)** | 1,796 | 3,301.1 | 518 | 478.5 | | | 2,314 | 3,779.6 |
| **GeoAsset Database** | **2,243** | **4,351.1** | **738** | **940.4** | **136** | **221.0** | **3,117** | **5,512.6** |

# CONCLUSIONS

- Successful demonstration of AI4EO frameworks
- Geoasset datasets are an important resource to adequately assess environment-related risk and impact exposure and the diffusion of these risks and impacts through the financial system
- Reference: McCarten, M., Bayaraa, M., Caldecott, B., Christiaen, C., Foster, P., Hickey, C., Kampmann, D., Layman, C., Rossi, C., Scott, K., Tang, K., Tkachenko, N., and Yoken, D. 2022. Global Database of Cement Production Assets. Scientific Data, *under review*

thank you for your attention!

contacts: cristian.rossi@sa.catapult.org.uk
christophe.christiaen@smithschool.ox.ac.uk