Sensor-Independent Deep Learning for Cloud Masking

Alistair Francis

Imaging Group, MSSL, University College London

ESA PhiLab

ESA Φ-Lab

- Researching and investing in ‘disruptive’ technologies in Earth Observation
- In-house focus on Machine Learning and Computer Vision techniques
- Works with other ESA divisions, as well as wider European community

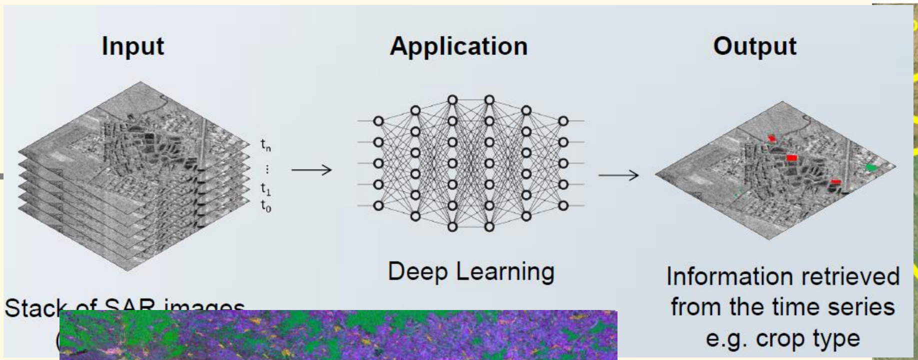
ESA Φ-Lab

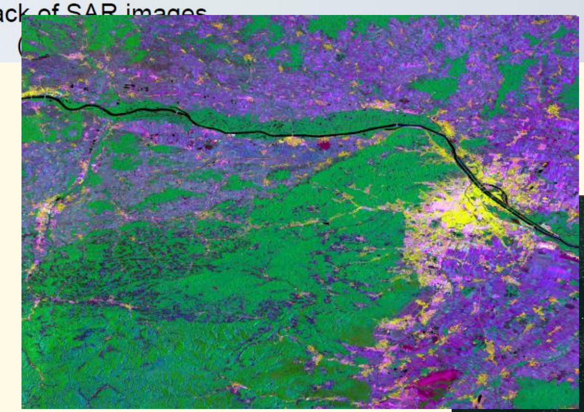
- Projects include:
  - Crop-type classification
  - Archaeological surveying
  - On-board convolutional models
  - Analysis-ready SAR data
  - Many more

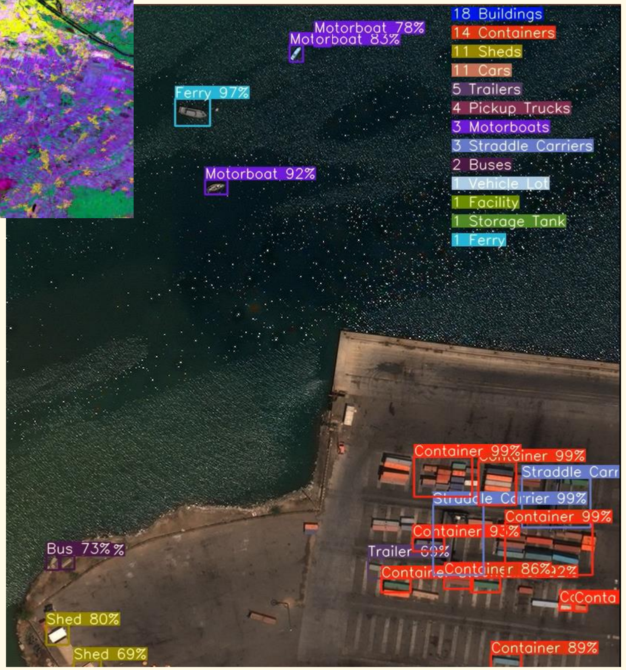
- Collaborations like mine are encouraged

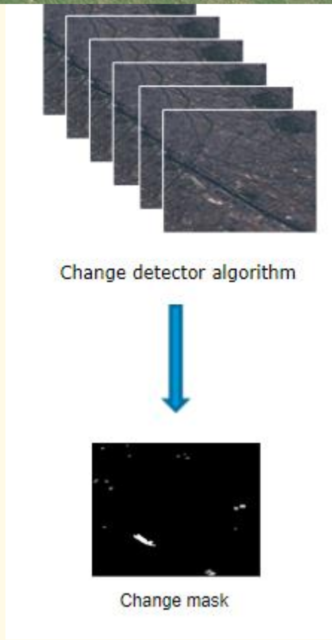
- Research Fellowship positions available

- Companies and researchers - get in touch!



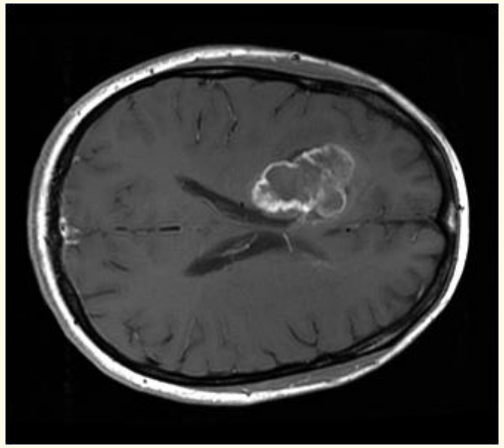






Deep Learning’s Promise

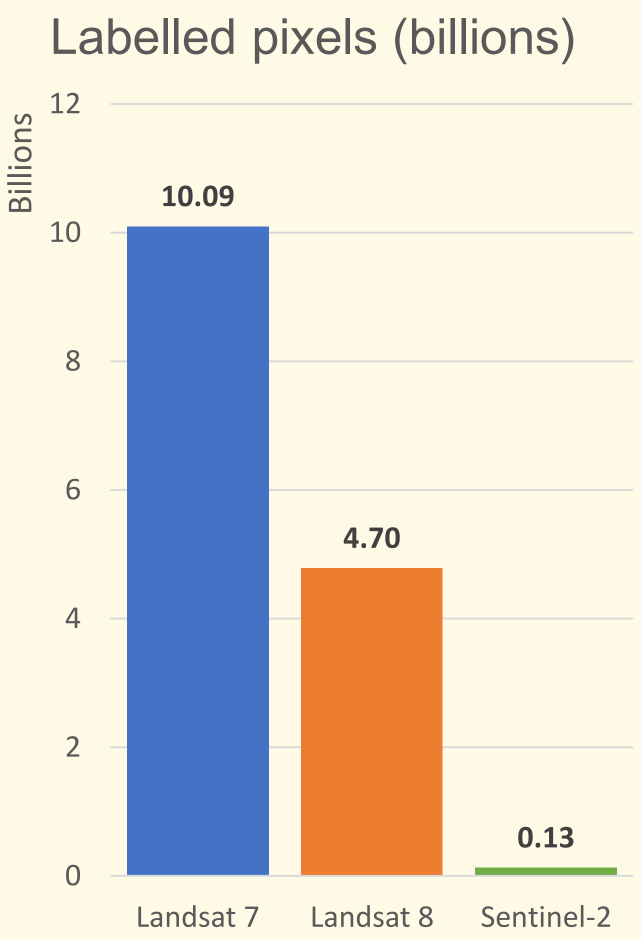
- Deep Learning offers remarkable performance
- Bespoke designs are unnecessary
- Applicable to wide range of problems and domains

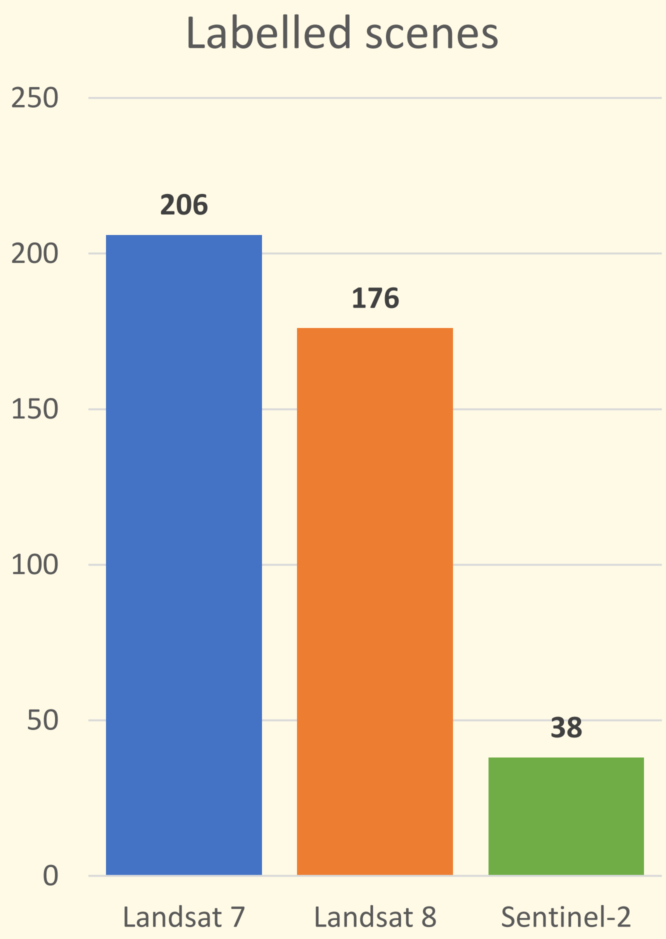


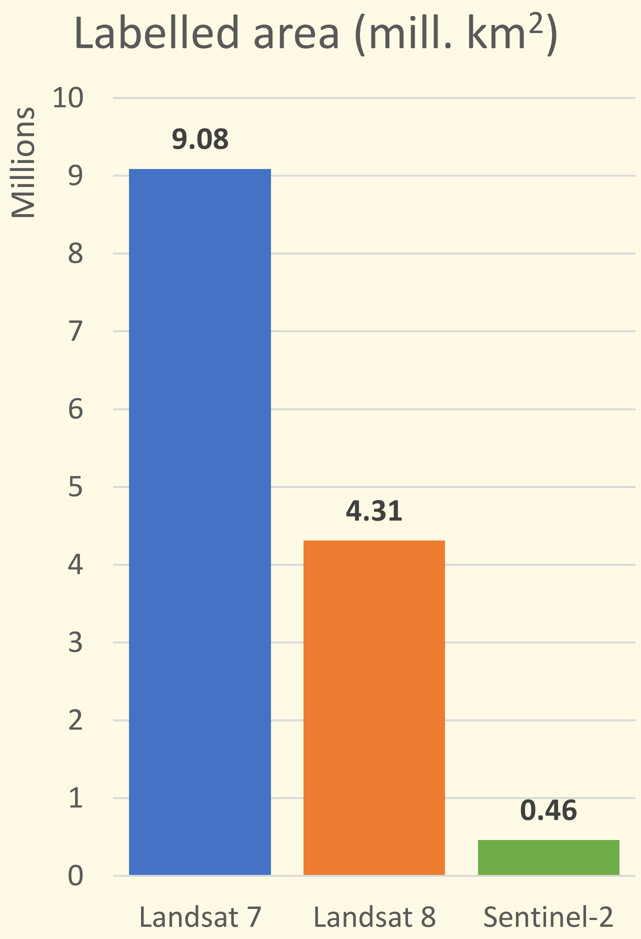
Remote Sensing vs. other imaging domains

- Consumer cameras all look similar: datasets are **not** camera-specific
- What's special about multispectral satellites?
  - Different spectral responses
  - Different brightness scales and calibration
  - Different noise characteristics
- New dataset needed for every problem and **for every sensor type.**
- Amount of labelled data multiplied by number of sensors flown!
- Problem is only emphasised as more satellites are launched

Cloud Masking datasets







*Aggregated from all publicly available datasets known to authors (not including single-pixel datasets).

How do we get the **most** from the labelled data we **already have**?

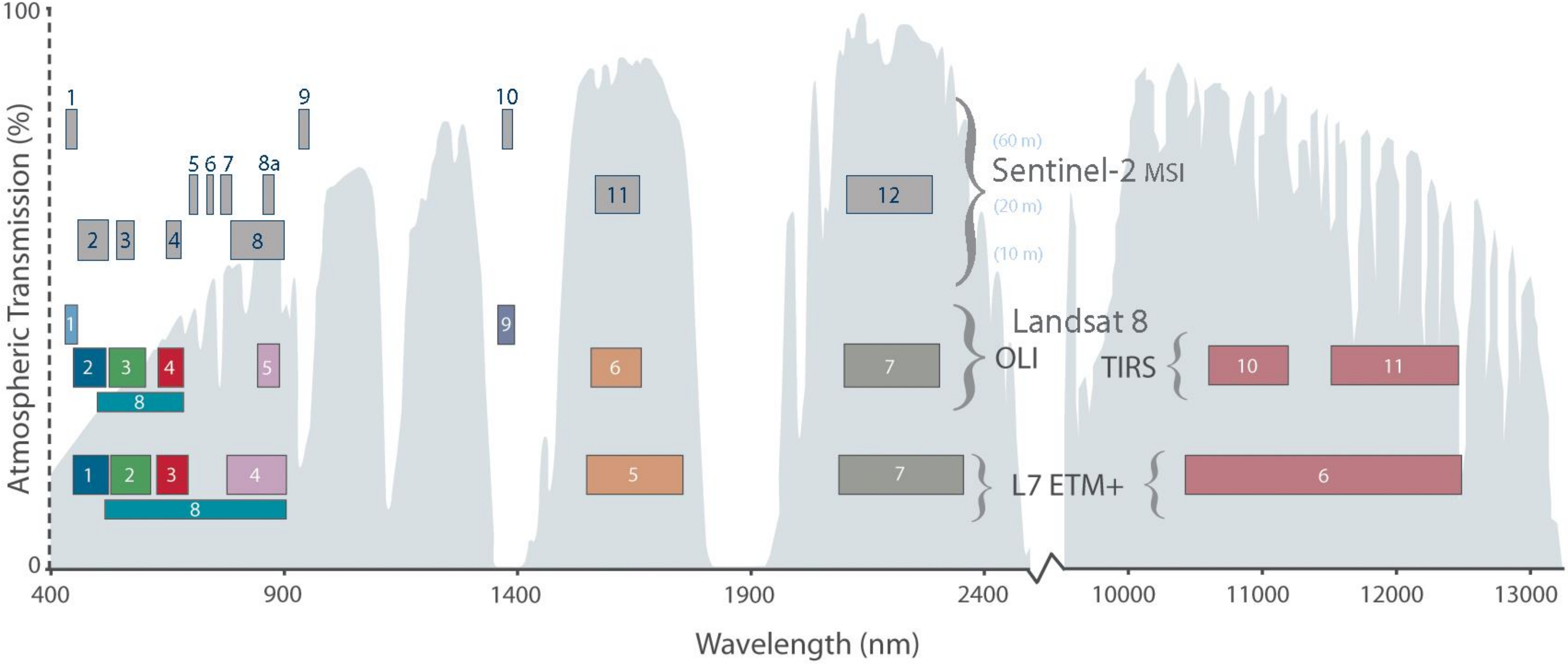
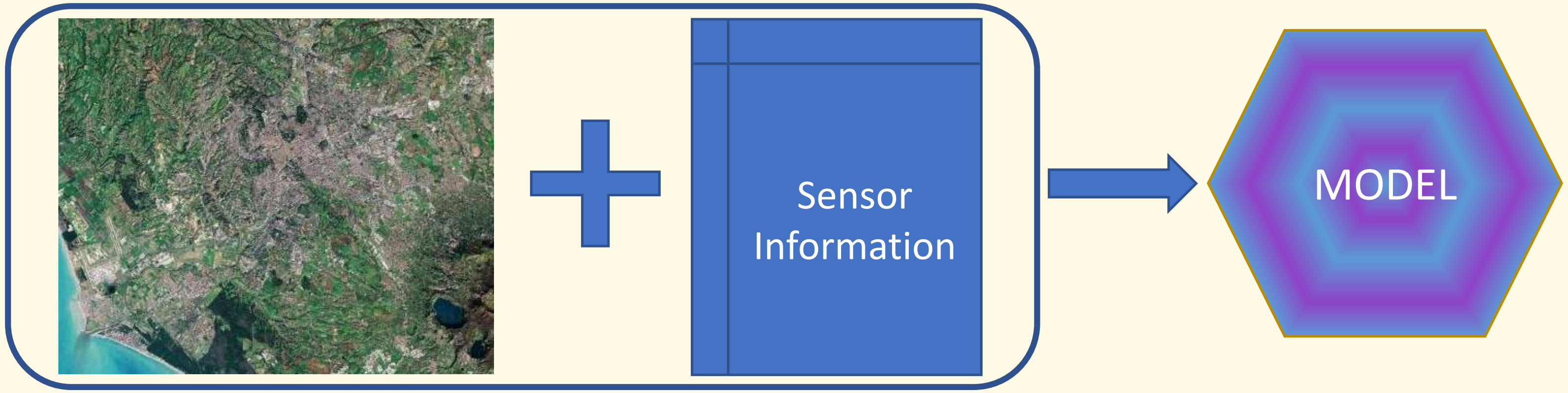


Image courtesy of NASA GSFC

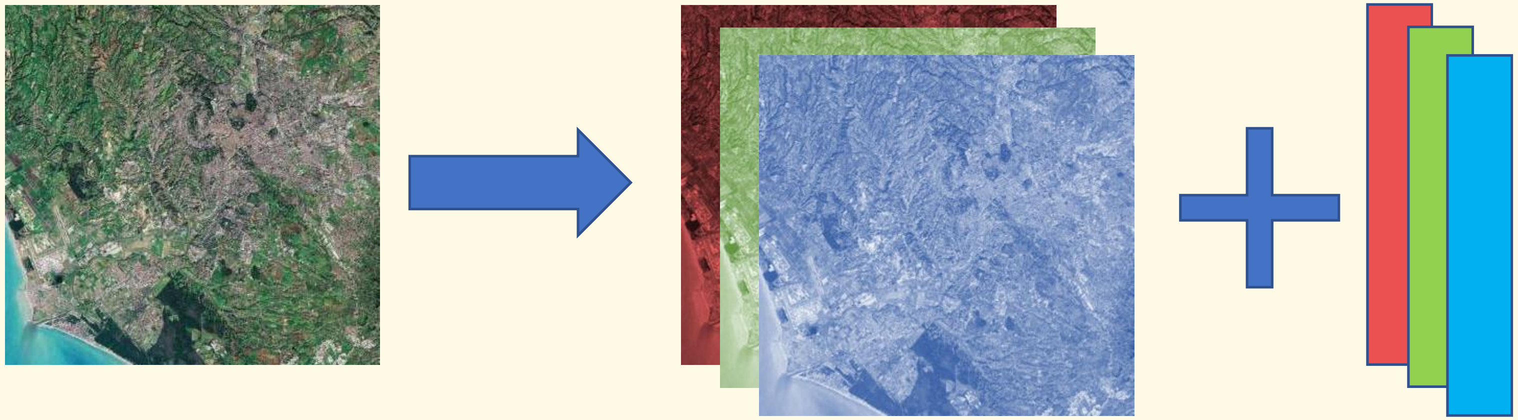
Sensor-Independent Model

- Train model on all sensors: single unified dataset
- Model recognises and treat sensors in different ways
- Could be used on new satellite without retraining
- New convolutional model design needed



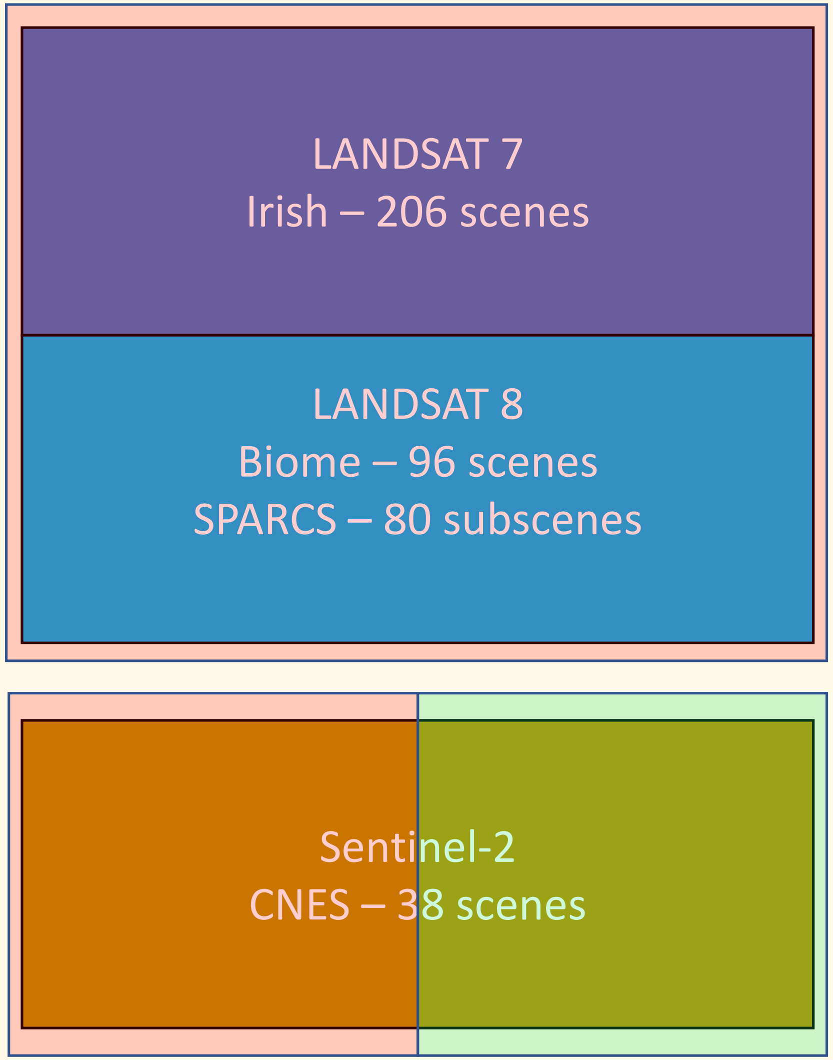
Sensor-Independent Model

- Each spectral band treated as a member of the set of all possible bands
- Model takes as input **any number** of **bands** and their **descriptors**
- **Descriptor** is a vector parameterization of band characteristics e.g.:
  - central wavelength and bandwidth



Experimental Setup

- Test across Landsat 7/8 and Sentinel-2
- Same model, trained three ways:
    1. Train on Sentinel – Test on Sentinel
    2. Train on Landsat – Test on Sentinel (only shared bands)
    3. Train on Landsat AND Sentinel – Test on Sentinel
- Does data from other satellites help?
- Absolute performance less important than relative performance

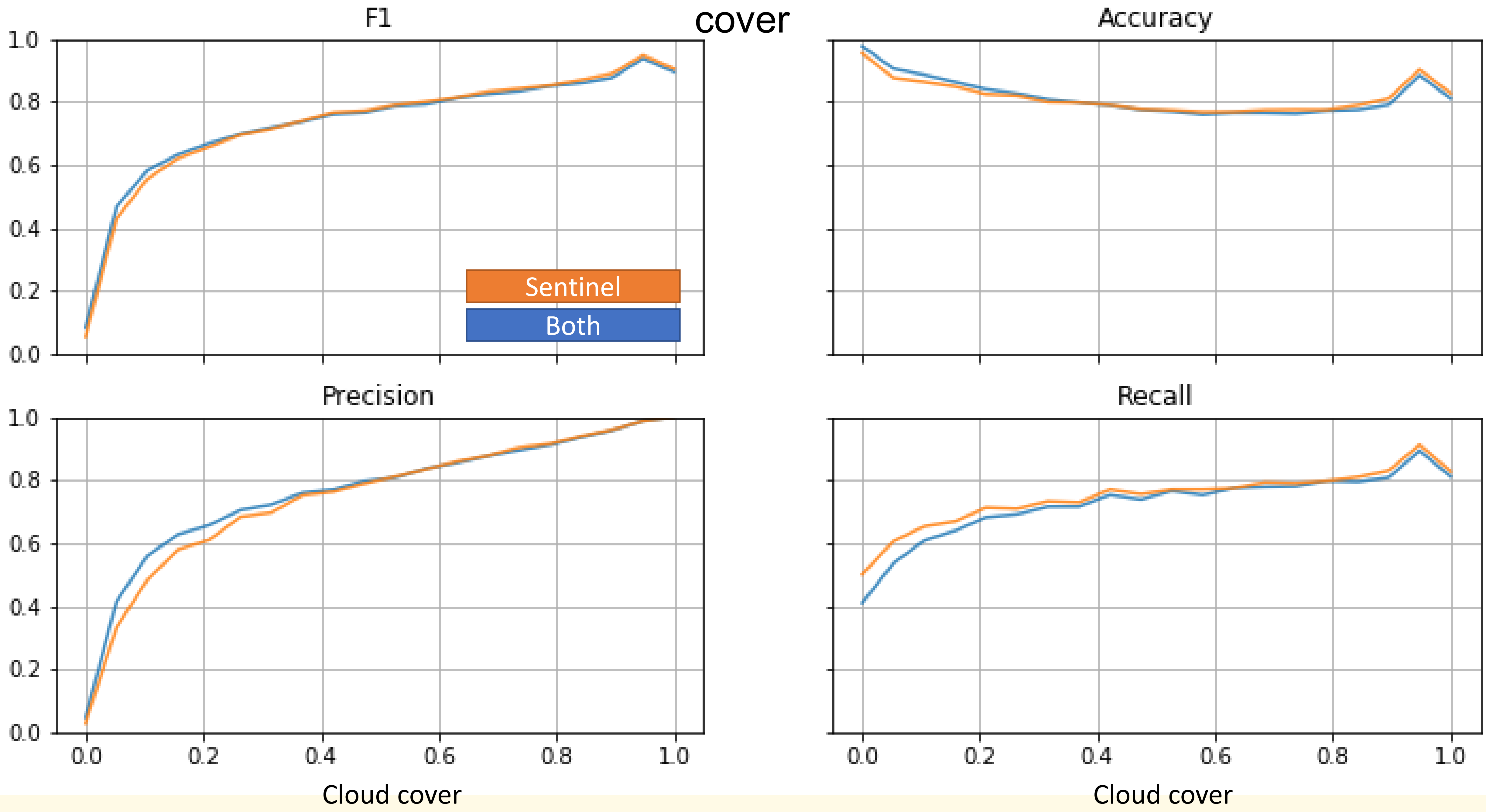


Results

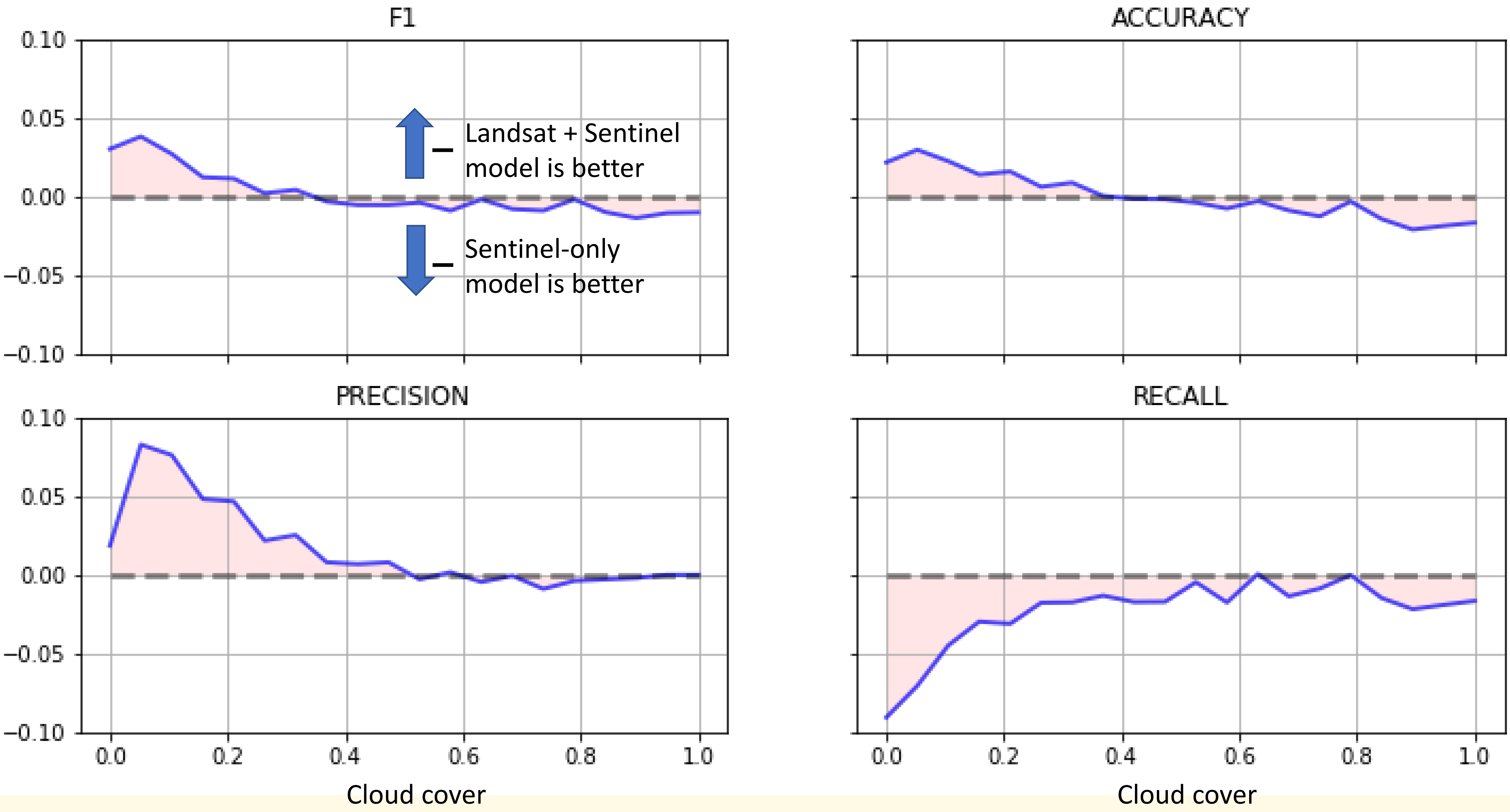
MODEL	Accuracy (%)	F1 (%)	Precision (%)	Recall (%)
Landsat	88.6	69.8	**86.9**	58.4
Sentinel	91.9	80.6	79.1	**82.1**
Landsat + Sentinel	**92.4**	**82.7**	85.1	80.4

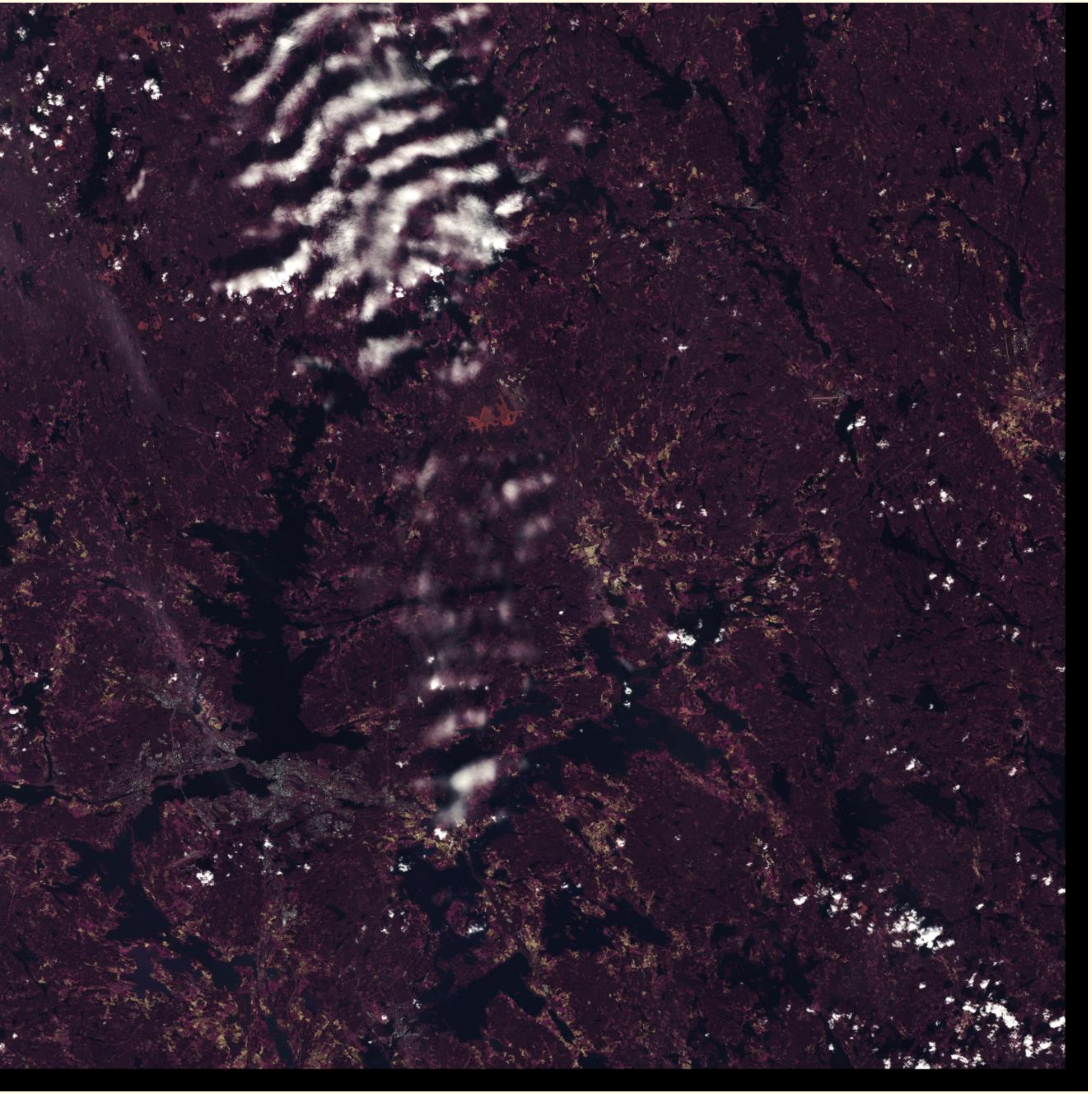
- Landsat-only model is not complete failure, but not good.
- Slight improvement in performance when using Landsat and Sentinel

Metrics per 128x128 window, by cloud cover

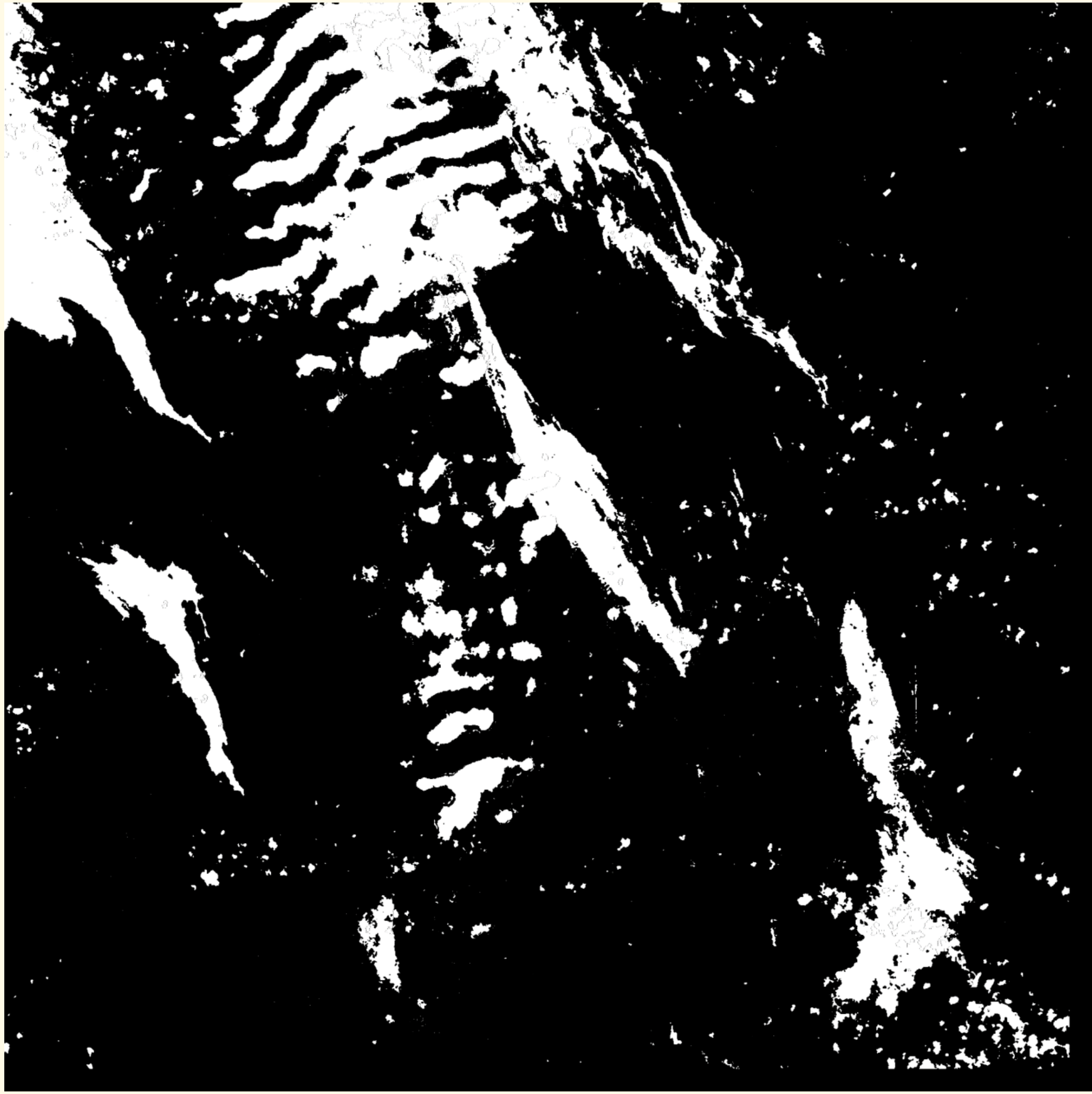


Difference between metrics per 128x128 window, by cloud cover

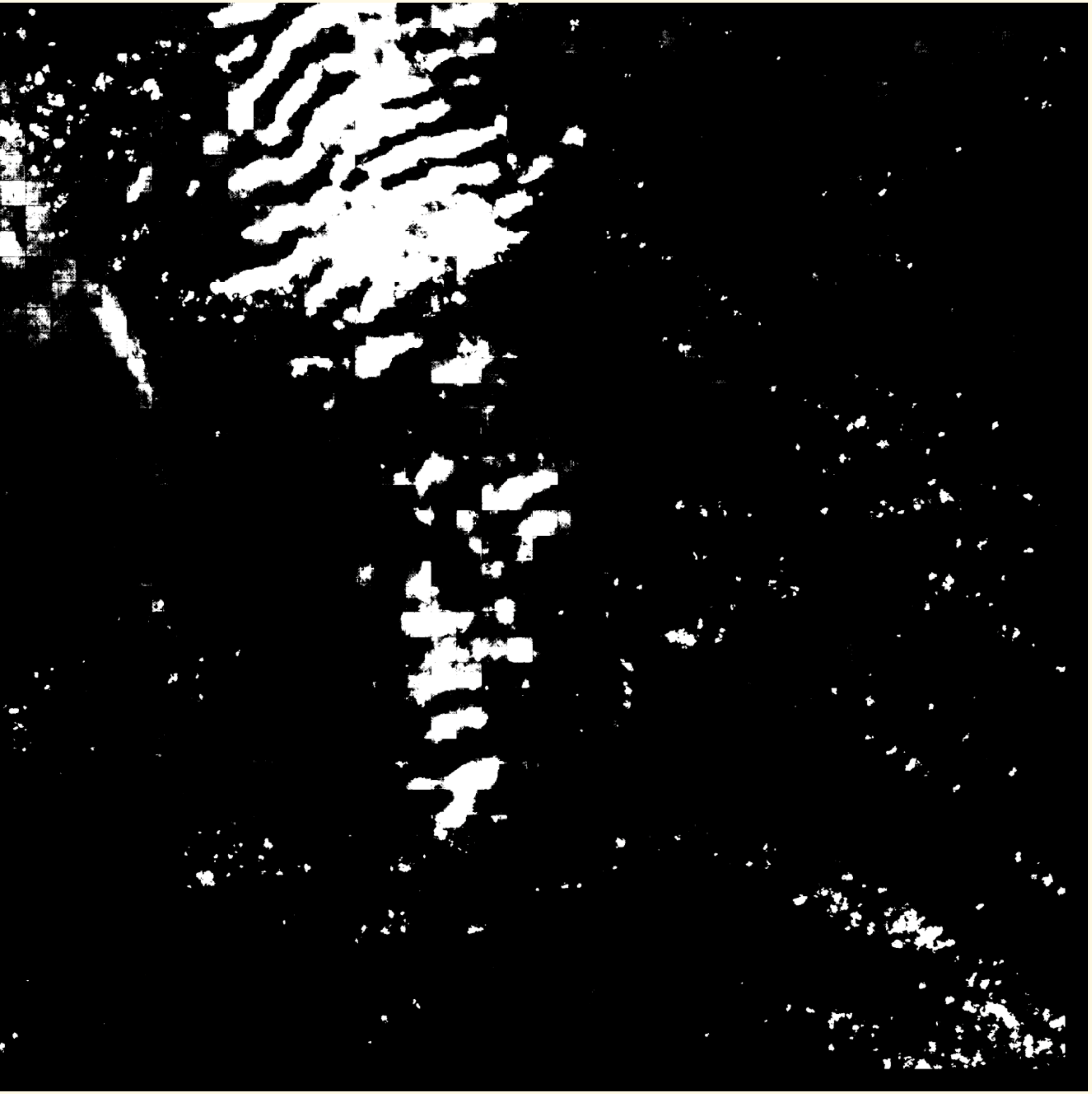




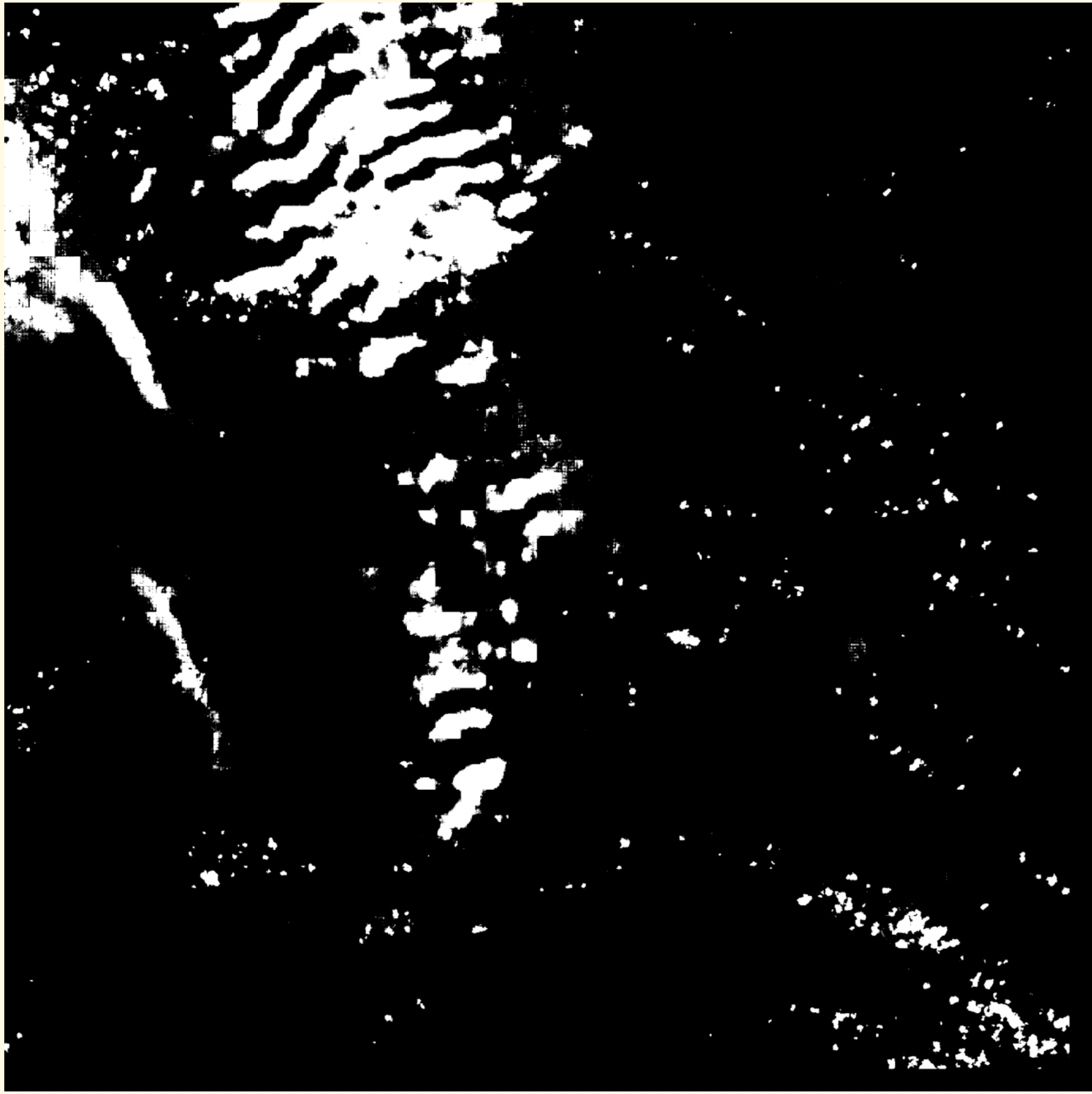
Central Finland



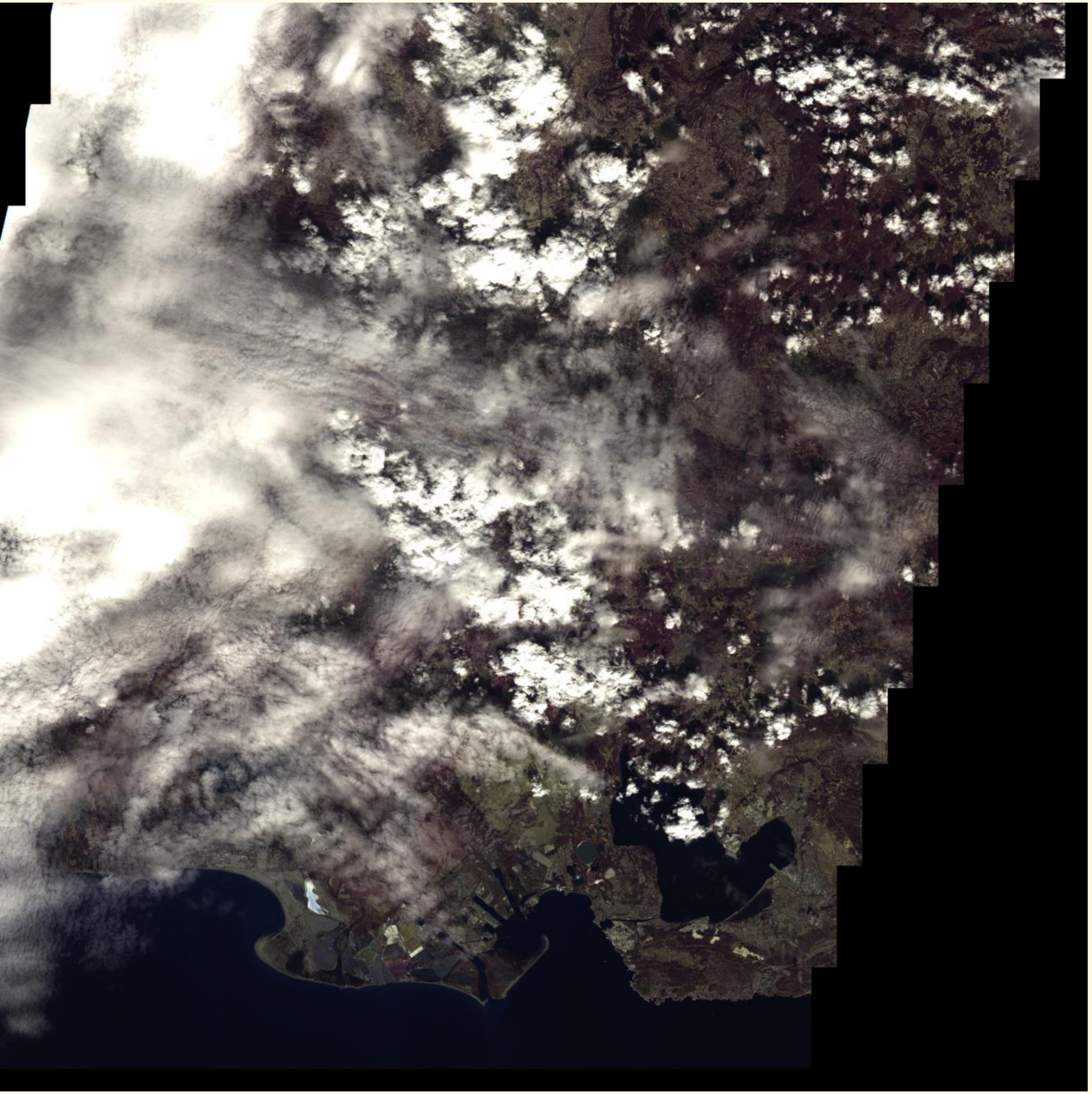
Groundtruth



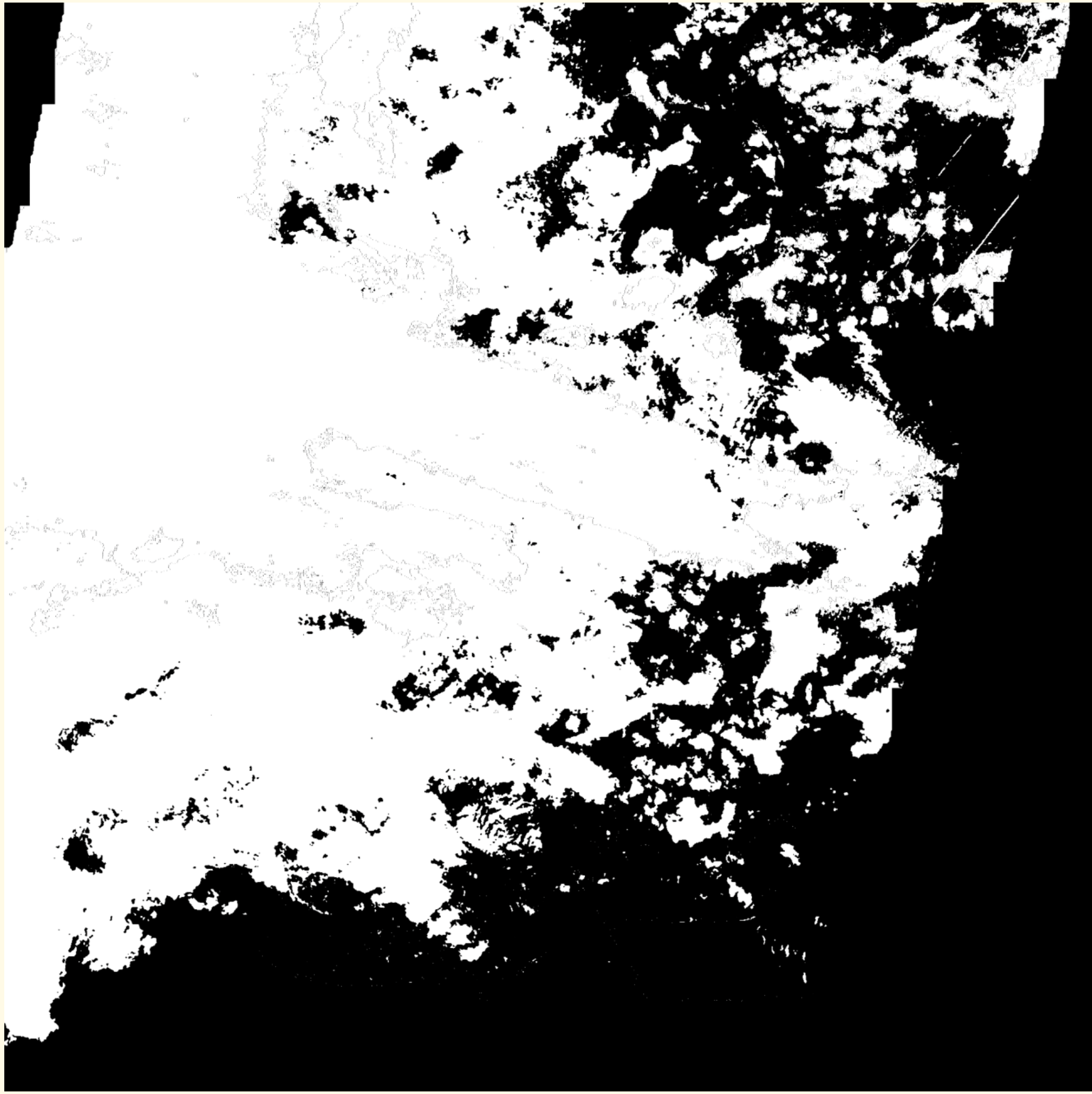
Landsat+Sentinel



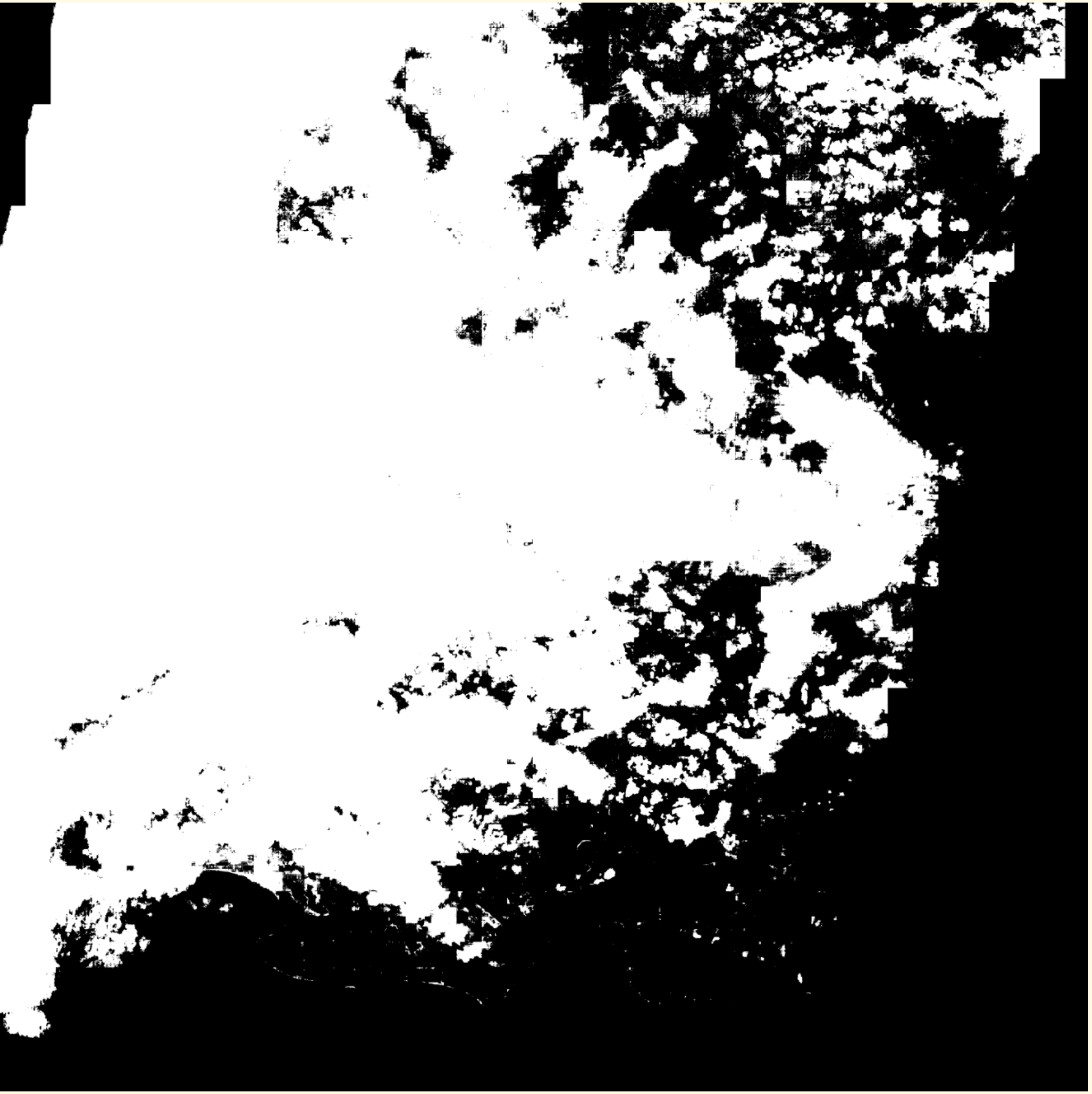
Sentinel



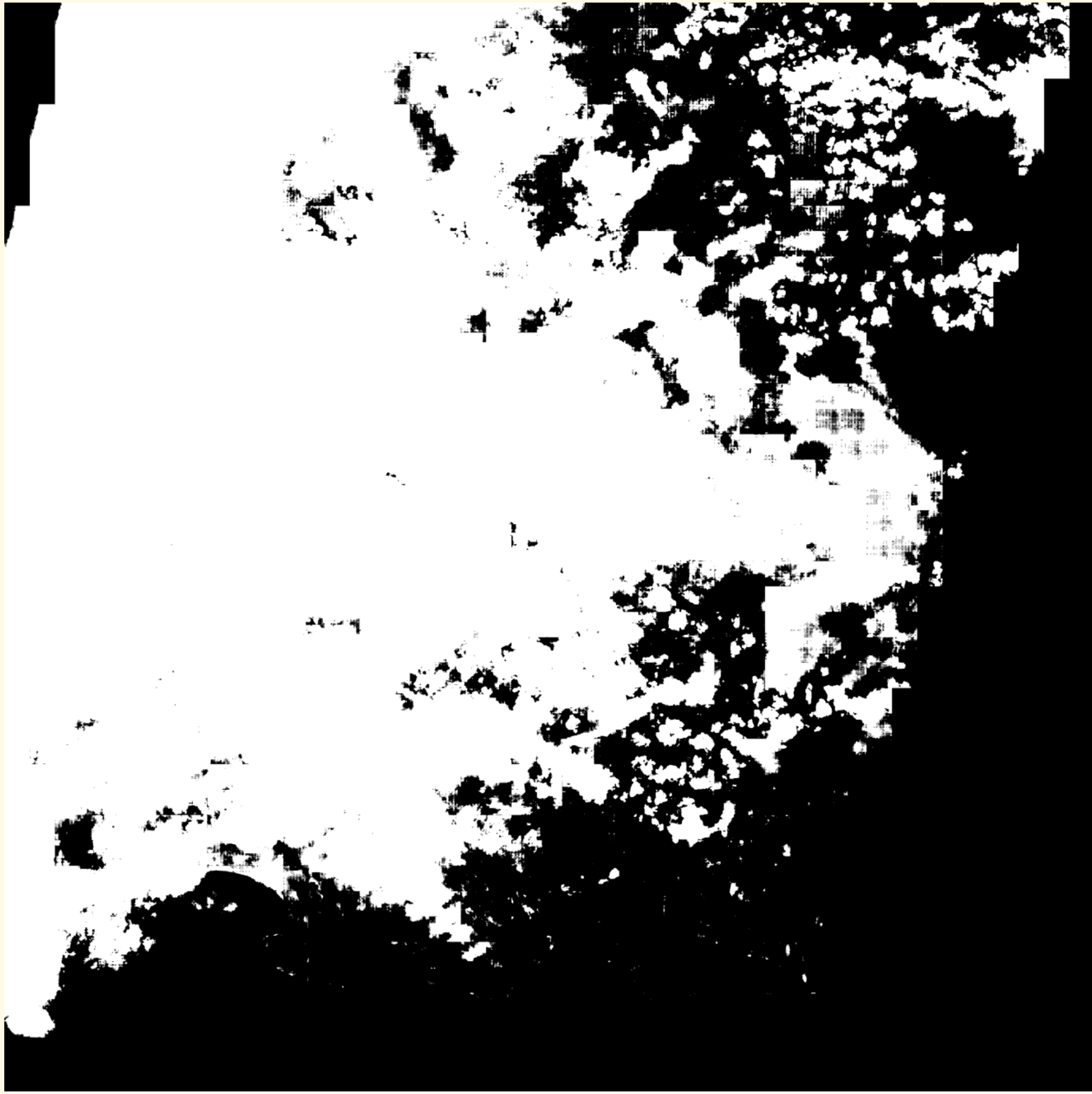
Arles, France



Groundtruth



Landsat+Sentinel



Sentinel

Takeaways

- Model is significantly worse if no data used from Sentinel-2.
  - Are there different sampling biases between Landsat and Sentinel datasets?
  - Are the shared bands as visually similar as we posit?
- (Very) tentative evidence that adding data from multiple satellites improves performance
  - Primary indicator of performance is **still** the amount of training data from the target satellite

Conclusions

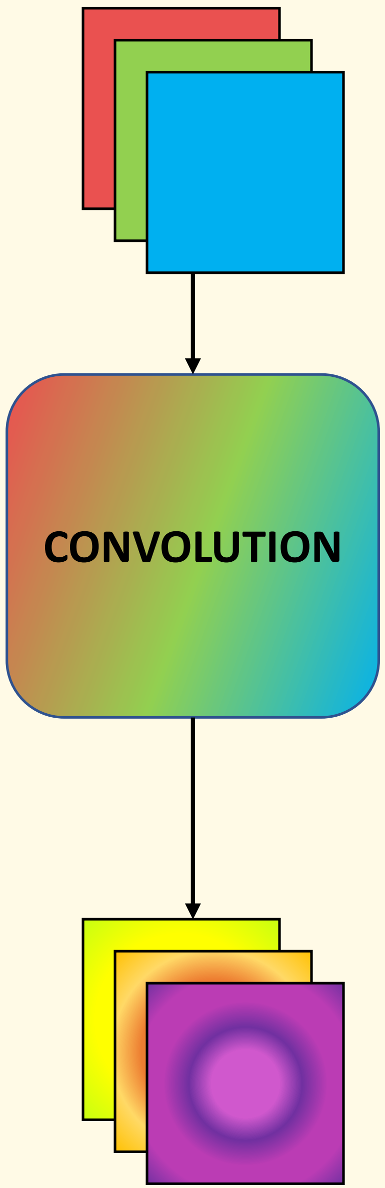
- Novel sensor-independent model has been developed
- Training across multiple sensors results in a somewhat better model
- More labelled data leads to better training **and** better validation
- More work is needed on understanding differences between sensors, and how the model is interpreting these differences

Thanks!

Extra Slides…

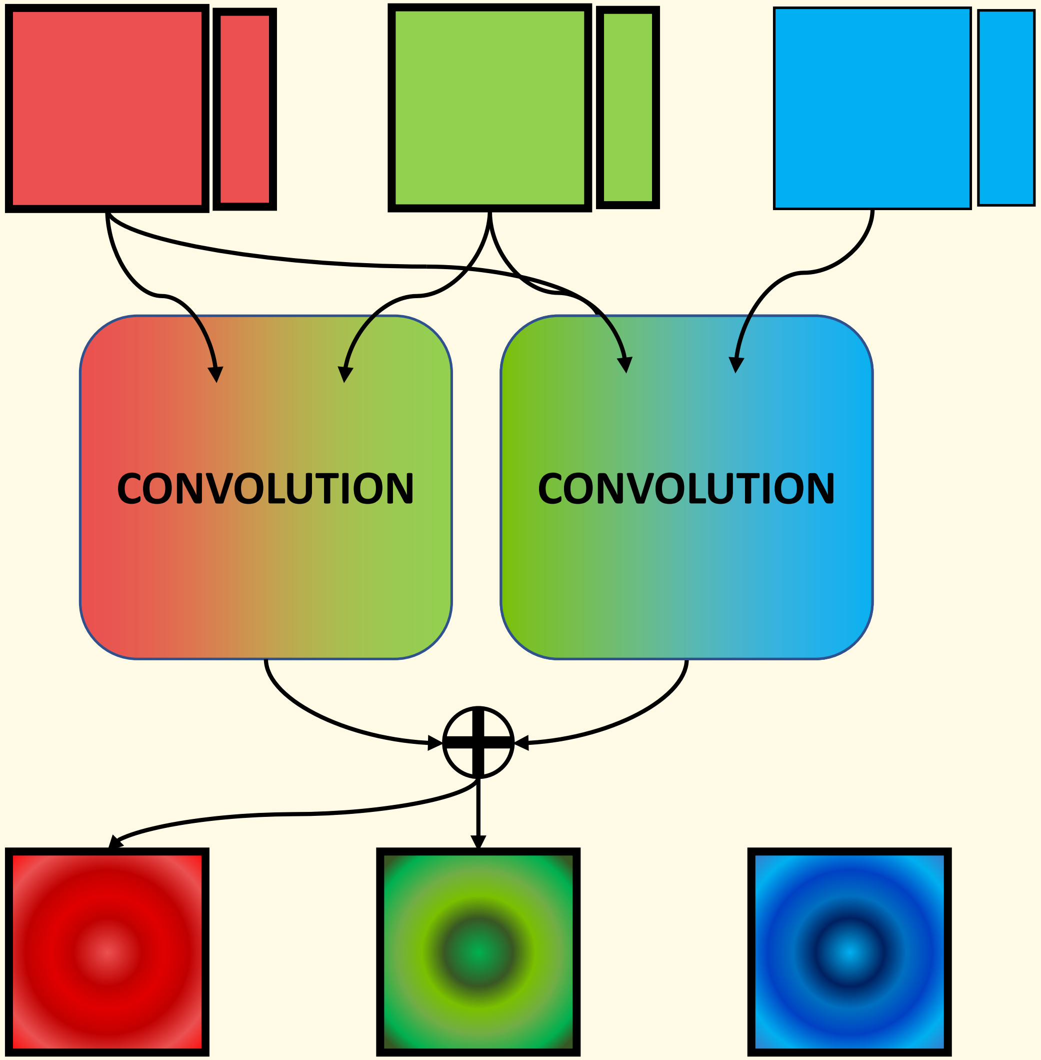
Sensor-Independent Model

- Convolutional layer replaced by permutational convolutional layer

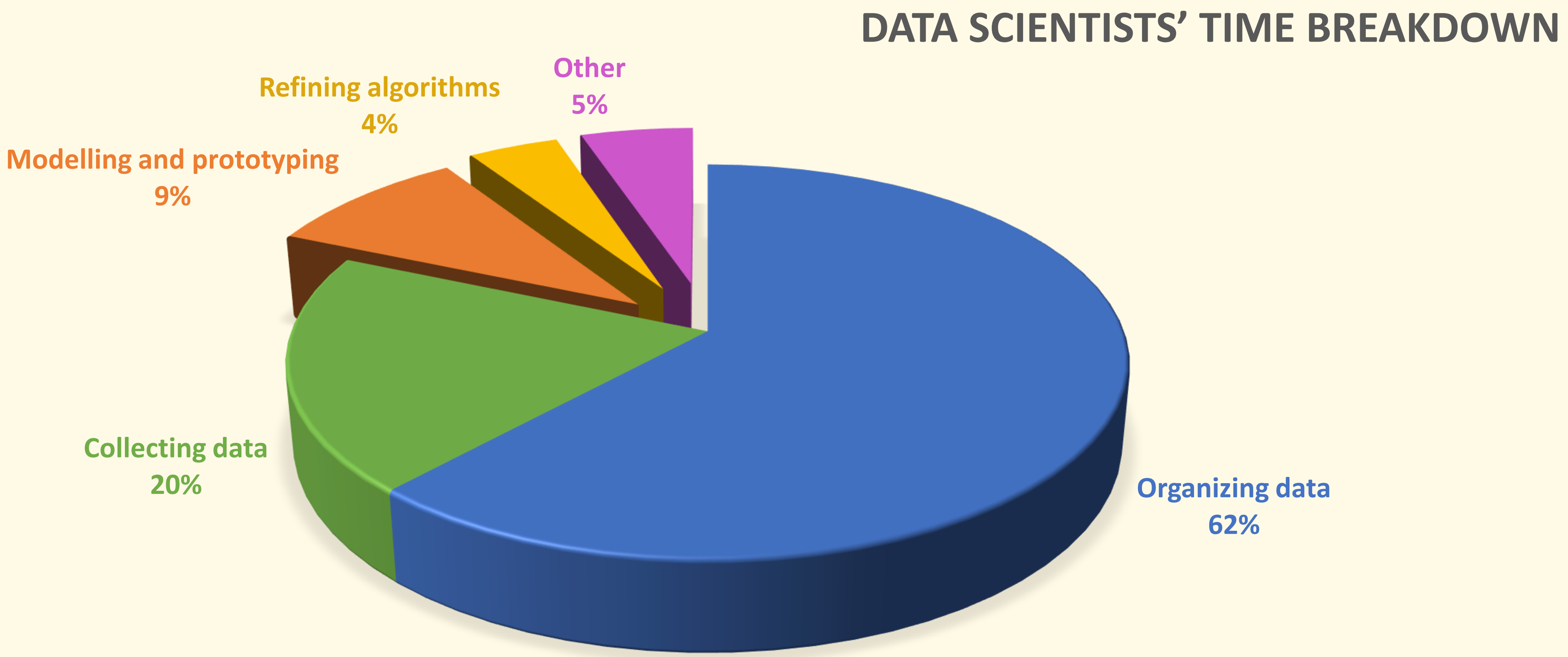


Sensor-Independent Model

- Convolutional layer replaced by permutational convolutional layer
- Convolve each pair of bands
- Sum pairwise outputs.
- Allows for arbitrary input size, **but** $O(n^2)$ with number of bands
- Modular, can be substituted in for normal convolutions



Deep Learning's Problem



>80% of time spent on problem-specific tasks

*Data taken from CrowdFlower survey