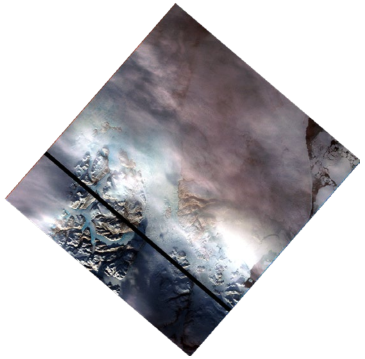
Artificial Intelligence-enabled Quality Control of Optical EO Satellite Data

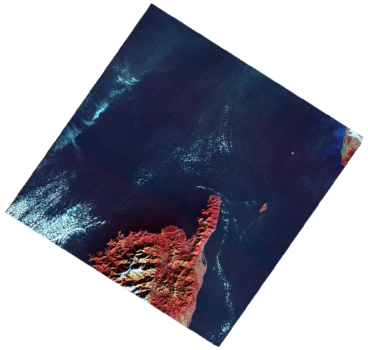
European Space Agency Living Planet Symposium (Bonn, May 2022)

Fay Done, Telespazio UK

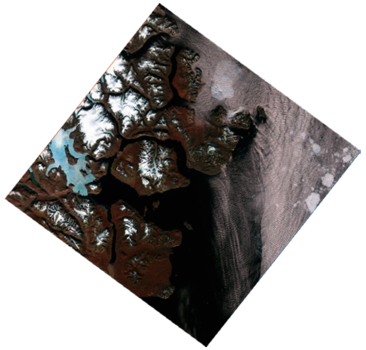
Background – What is the problem we're trying to solve?



- The important characterisation of satellite Earth Observation (EO) data quality is enabled by quality control (QC) activities.
- The QC activities are performed by dedicated services such as ESA's IDEAS-QA4EO (TPZ UK prime contractor):
  - To identify any technical issues (e.g. product format, geometric and radiometric calibration, artefacts) with the requested EO mission data through:
    - Systematic checks (applied to the whole dataset)
    - Spot checks (applied to a subset of the dataset)

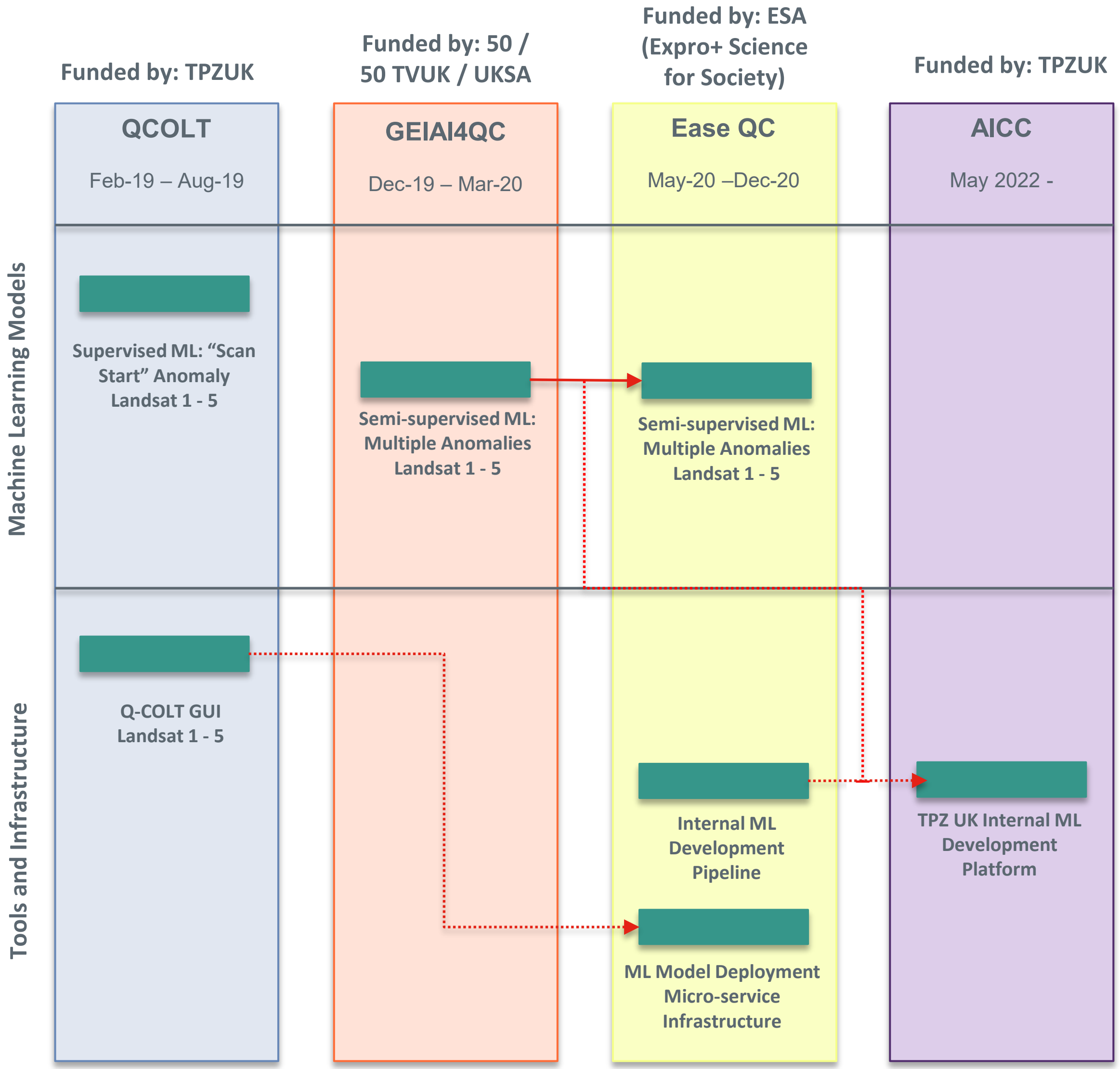


- Are the current QC methods and tools used by the QC Engineers future-proof? No, not really?
  - The volume and complexity of the data from EO missions has increased significantly in the last decade or so and this could have implications on: Efficiency, Reliability and Consistency (and the Funding and Resources needed).
  - Solution(s)? Automate more (e.g. visual inspections) via Machine Learning (ML)? **This is the question being explored by Telespazio UK's Ease QC initiative.**



Ease QC Initiative - Introduction

- **The initiative is built upon the success achieved across multiple, inter-connected projects (see figure). The initiative's activities have been focused on either the development of:**
  - **Tools and infrastructure to support ML model development and deployment.**
  - **ML models for anomaly detection (for anomalies that can be visually detected but not easily detected deterministically).**



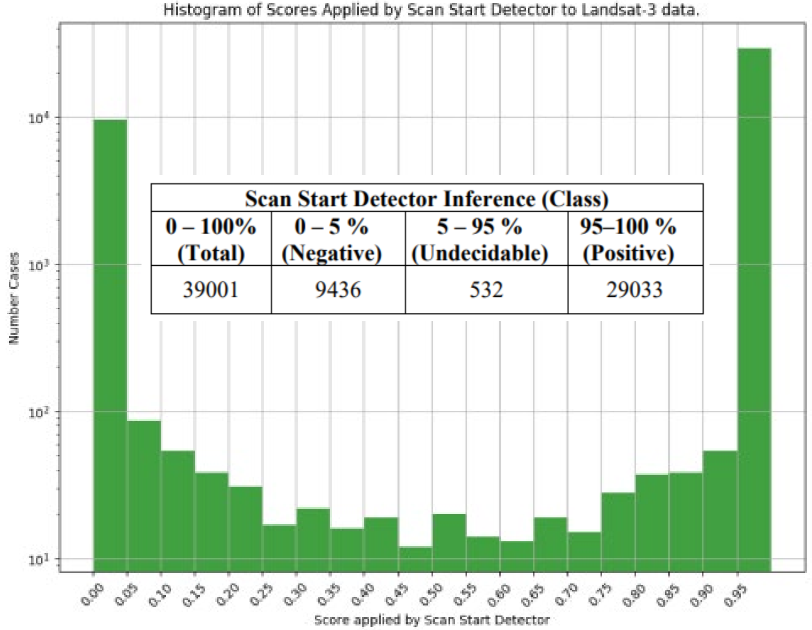
Ease QC – Supporting Tools

- Q-COLT – **Key tool supporting both ML and QC processes**
  - Q-COLT Archive and Database:
    - 600,000+ Landsat 1 - 5 MSS products procured for the IDEAS-QA4EO Landsat QC team for their QC activities (Bulk Landsat 1 – 5 Reprocessing Campaign).
      - Product browse imagery
      - Product metadata (including quality metadata from Amalfi)
      - Quality tags (QC engineers and ML models).
  - Q-COLT GUI:
    - Suitable interface for simplified browsing and selection of products (browse imagery, metadata, tagging) by both QC Engineers (e.g. IDEAS-QA4EO) and ML Engineers (Ease QC). Single source.

Ease QC – Machine Learning Model Development

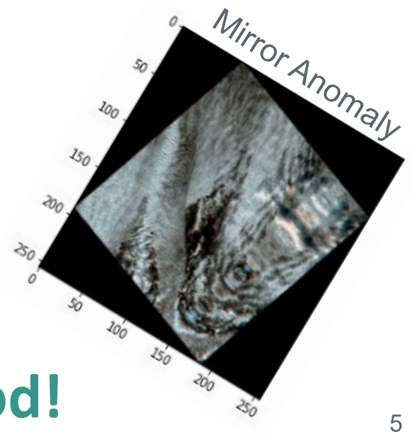
Predicted	Positive (Actual)	Negative (Actual)
Positive	True positive (100)	False Positive (0)
Negative	False Negative (16)	True Negative (84)

- **Supervised Machine Learning Architecture**
  - This architecture was used to build a soft classifier (%) to detect anomalous data containing a known anomaly type (e.g. scan start). The core ML model was trained on a dataset where this known anomaly type is sufficiently represented (not easy, and not really possible for operational missions short-term).
  - The results indicate a good accuracy (success!) but there is room for improvement (e.g. improve generalisation)!



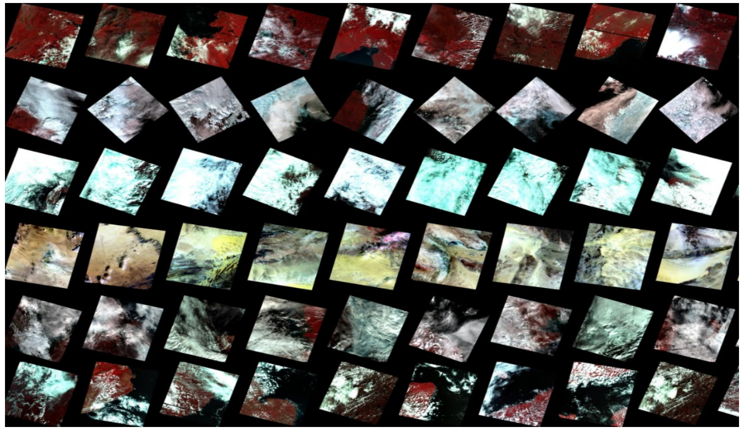
Scan Start Detector Inference (Class)			
0 – 100% (Total)	0 – 5 % (Negative)	5 – 95 % (Undecidable)	95–100 % (Positive)
39001	9436	532	29033

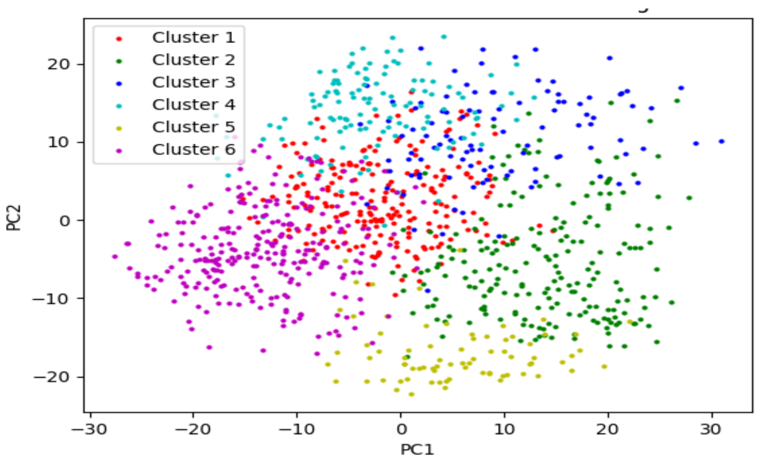
- **Semi-supervised Machine Learning Architecture (far more complex, split Workflow A and B)**
  - Workflow B: The objective of this workflow is to build a binary classifier to detect anomalous data (any anomaly type, whether they are previously known or unknown).
    - Workflow A: The objective of this workflow is to develop and implement ML models to support the reduction of data complexity (e.g. dimensionality and variability), in order to improve the performance of the subsequent ML models for Workflow B. This workflow is applied to all browse images in the Q-COLT archive.



We want to make sure that we don't flag bad data as good!

Ease QC – Machine Learning Model Development

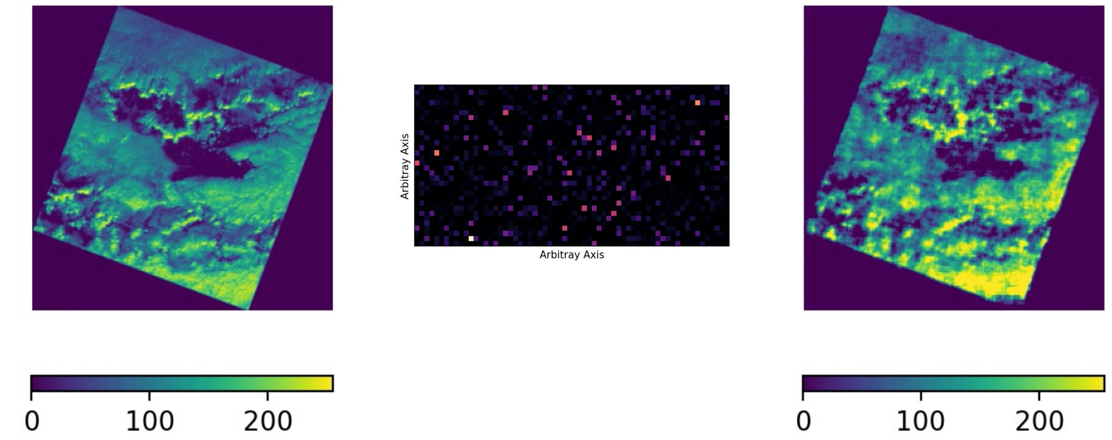


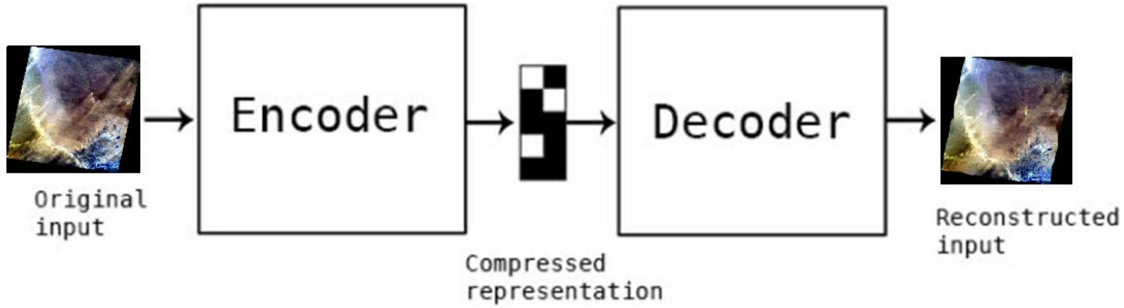


The data is clustered according to their shared or similar properties (i.e. minimising intra-cluster variance); variance as a consequence of different surface types (e.g. cloud, desert, sea, snow / ice, cloud-free land, etc.).

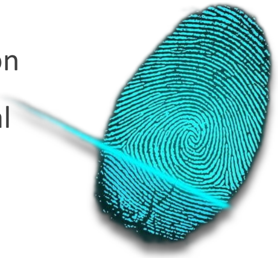
Ease QC – Machine Learning Model Development

- This supervised ML model (Convolutional Neural Network Auto-Encoder) contains two main components, an encoder and a decoder. It is used to efficiently learn compressed representations of non-anomalous inputs (browse imagery) from each cluster.
  - The ML model outputs reconstructed imagery from which input-output statistics are derived as 'fingerprints' of non-anomalous data; low reconstruction errors indicate non-anomalous data and high reconstruction errors indicate anomalous data (i.e. anomaly detected).
- The ‘fingerprints’ are then used by another model to build a binary classifier.



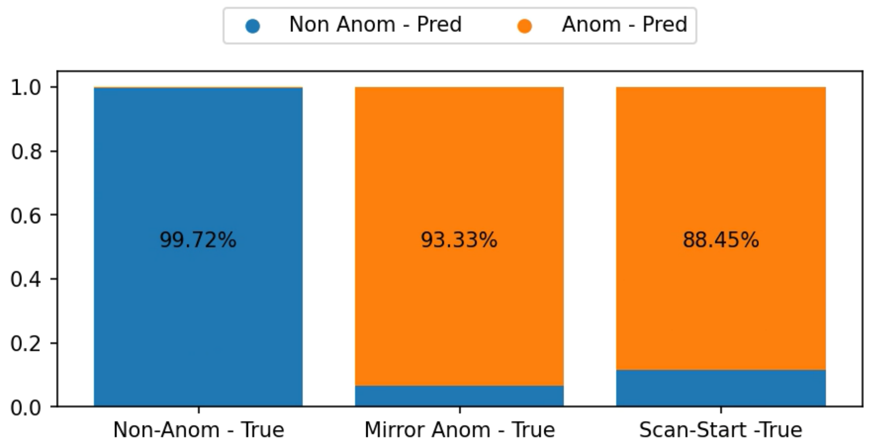


The AutoEncoder converts the input into a compressed representation (encoder + latent space) and then tries to **reconstruct** it to its original representation (latent space + decoder).



Ease QC – Machine Learning Model Development

- **The results so far are promising (see below) but there is so much more to do (e.g. model refinement, validation, etc.) – it's not an easy task!**



Left - performance of the model on test data. The three columns correspond to non-anomalous, mirror anomaly and scan-start anomaly data. The colours correspond to the fraction of each class predicted for each dataset.

Summary

- Summary of activities to date:
  - Highly complex development – more so than originally thought!
  - Promising preliminary results:
    - Good performance achieved on anomalies highly represented in the training data
      - Supervised ML models demonstrated success at specific and well characterised anomalies
      - Less success with less represented anomalies

- Summary of future activities:
    - Explore different data preparation techniques
    - Potentially redevelop the training network and model architecture
    - Train ML models using more labelled data (specifically labelled examples of anomalies)
    - Adapt Q-COLT and develop new ML models for other EO missions (e.g. Sentinel-2)

THANK **YOU**
FOR YOUR ATTENTION

telespazio.com

Contacts

Kevin Halsall

Ease QC Project Manager

kevin.halsall@telespazio.com

Fay Done

Ease QC Team

fay.done@telespazio.com

Teo Bloch

Ease QC Team - Data Scientist

teo.bloch@telespazio.com