FCCS: First Commercial cloud (computing) in space

3CS / Cognitive Cloud Computing in Space campaign idea: I-2022-00373

Final Review – 20 December 2022

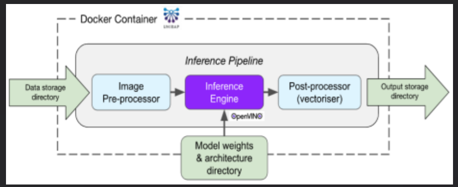
D-ORBIT NEW SPACE SOLUTIONS UNIBAP TRILLIUM EUROPE

Contents

Context

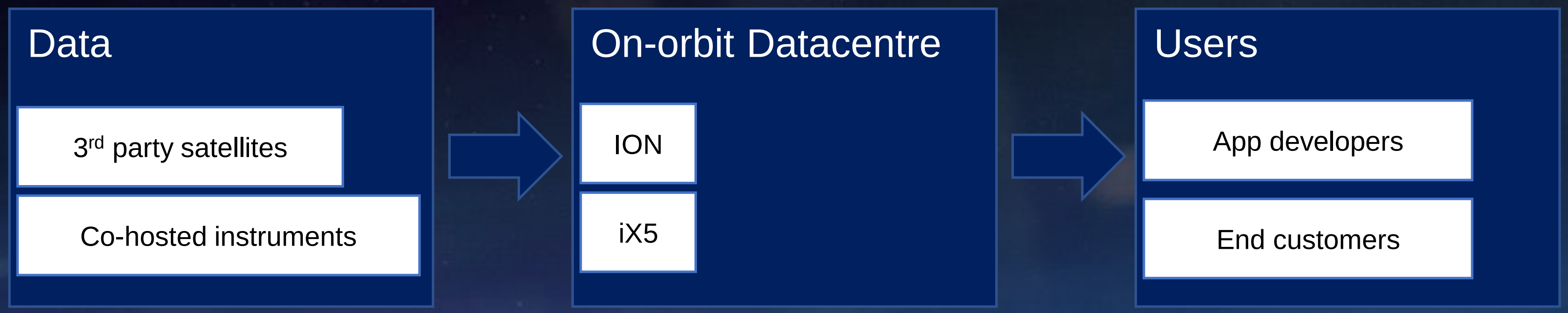
Cloud services on-orbit

- Terrestrial cloud services have changed the way businesses operate, providing compute power greater than any in-house solution
- These provide access to millions of applications, services and data archives
- Data generation on orbit is ever-increasing, providing cloud services nearer to the source can provide the same benefits as on the ground without the limitations of bandwidths and link budgets
- D-Orbit, Trillium and UniBap have a legacy of collaborative success in demonstrating cloud-based AI apps and services on-orbit
- Enabled by the ION SCV missions carrying the UniBap iX5 SpaceCloud on-board processor, providing a generalised compute capability on-orbit



Commercial justification

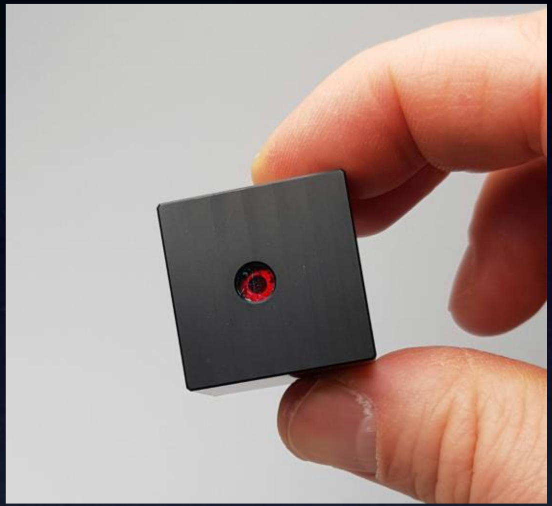
- The value chain for data is changed to move the datacentre to orbit
- Data is fed to the datacentre directly on orbit or curated and managed as a service from indirect or archive sources
- Users continue to build cloud apps based on common microservices and APIs to process these data directly on orbit
- Commercialisation of the datacentre is then achieved in a similar way to terrestrial cloud platforms



TRILLIUM EUROPE
UNIBAP

VTT cubic inch hyperspectral camera

- This project takes advantage of the iX5 on-board SCV-004, launched in January 2021
- The mission included a cubic inch hyperspectral camera, integrated directly into the iX5
- This provides a unique data source directly accessible to app developers using the platform



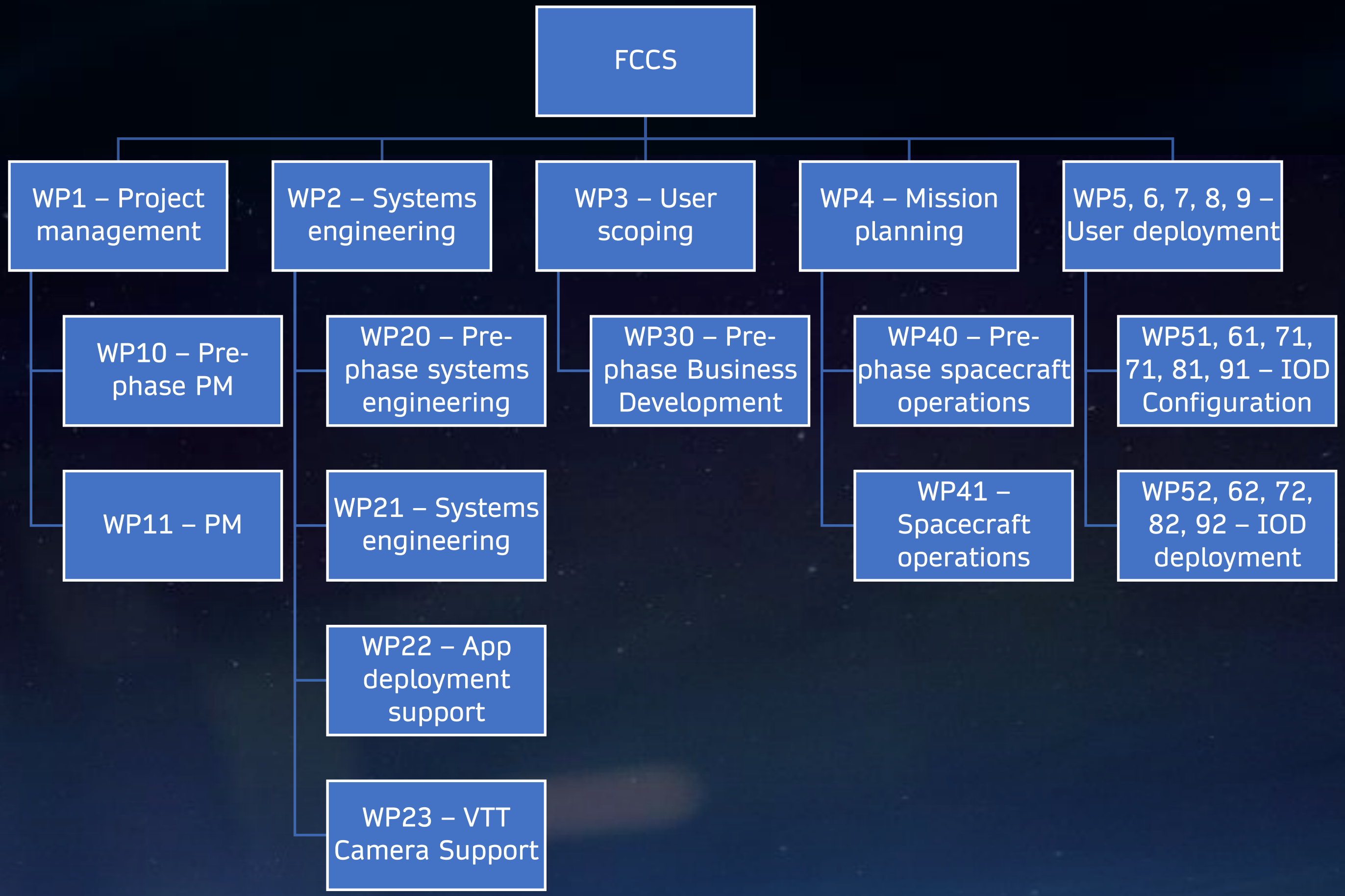
Parameter	Value
GSD	Approx. 167 m
Spectral range	650 nm – 950 nm
Spectral resolution	15 nm – 25 nm @ FWHM
Image resolution	640 x 480 pixels
File format	Direct output of TIFF or gRPC-buffers
Compression (SCFW defined)	CCSDS 123.0-B-2

Sentinel-2 band	VTT hyperspectral camera band
496.6 nm (blue)	550 nm
560 nm (green)	570 nm
665 nm (red)	930 nm
705 nm (vegetation red edge band 5)	725 nm
740 nm (vegetation red edge band 6)	
945 nm (water vapour)	

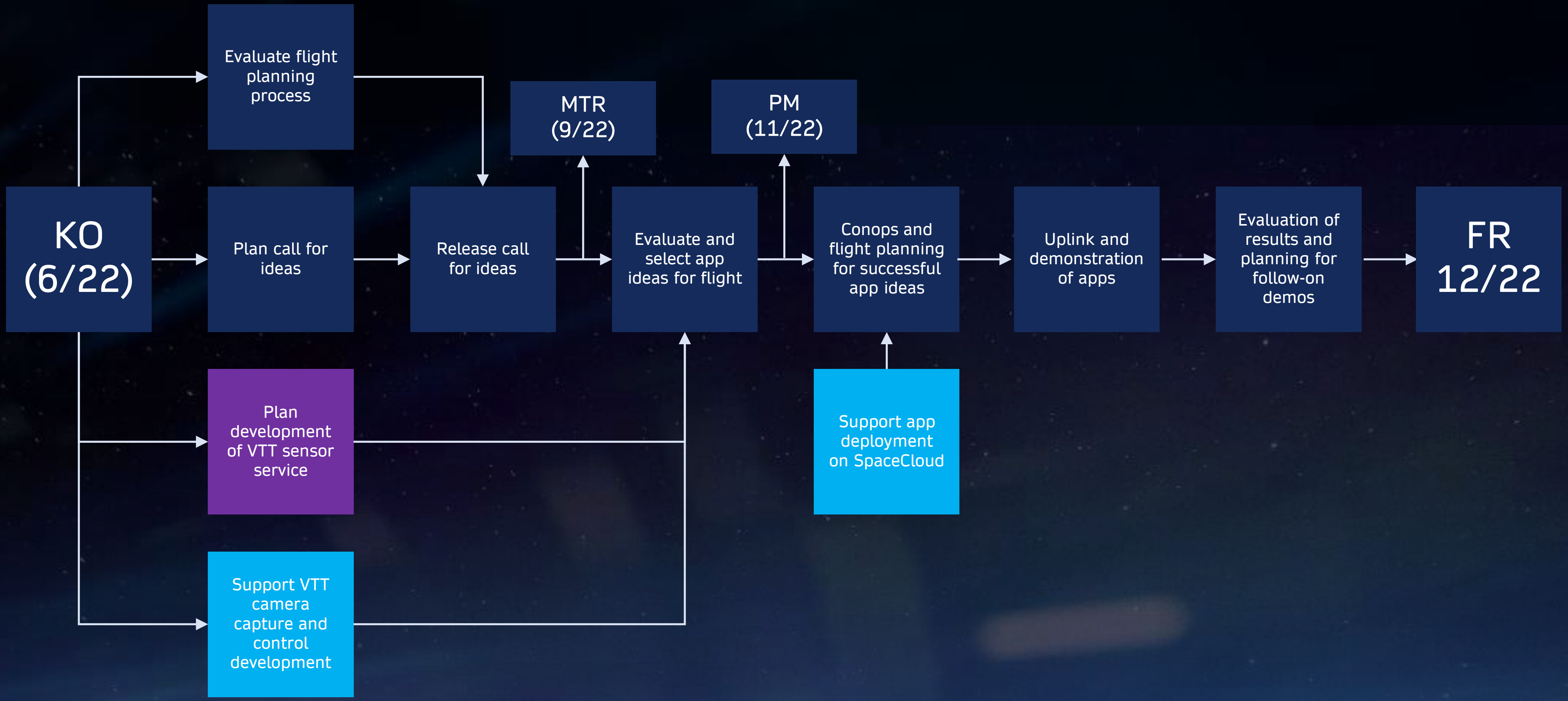
Project Recap

Scope

- This project aims to bring new users to the cloud platform environment and understand their experience in developing, deploying and using their apps on the platform



Work logic

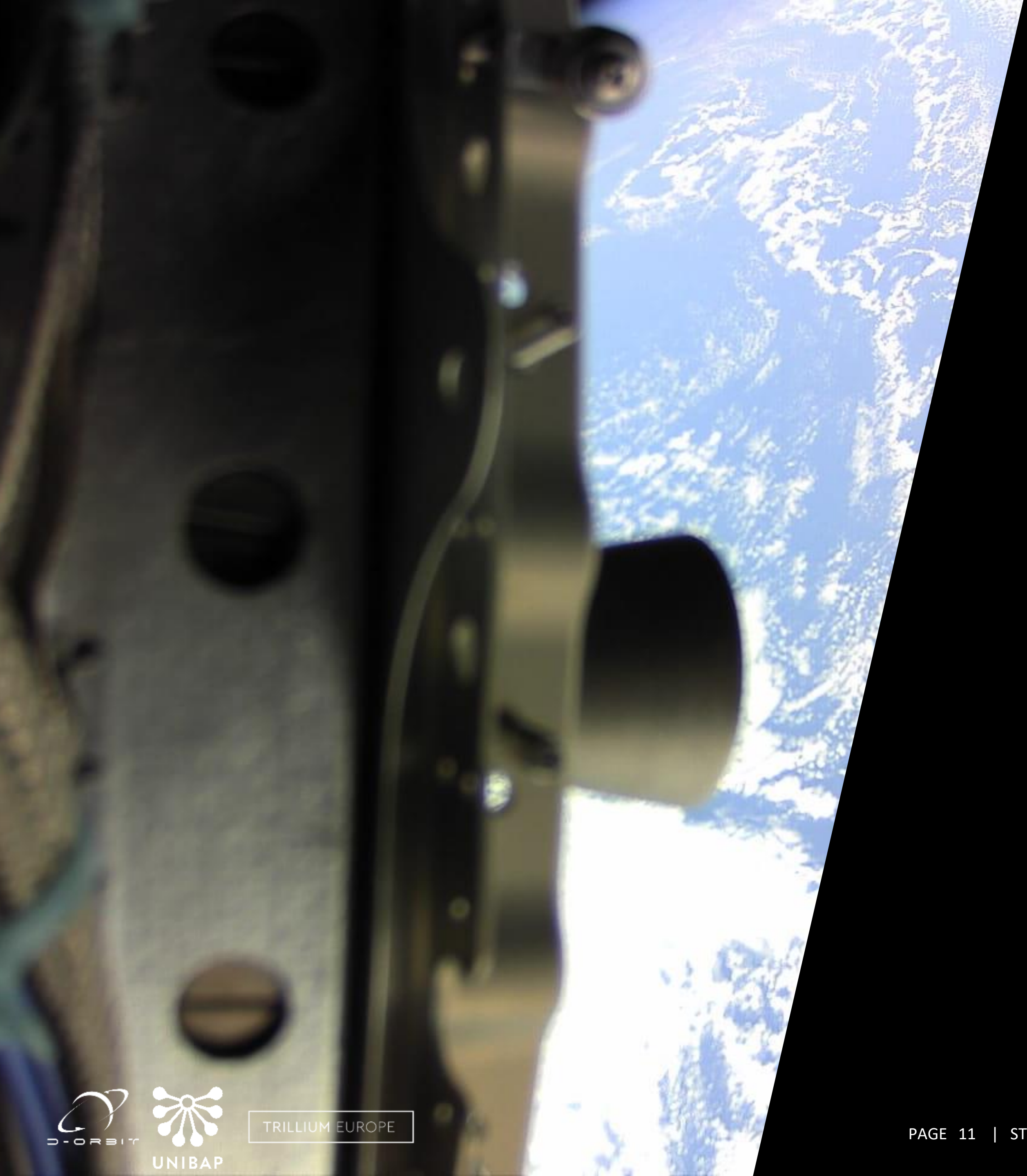


TRILLIUM EUROPE

Objectives

- The project objectives are:
  - To promote and provide early access for future users of such network services
  - Issue a call for ideas to invite applications from industry and academia to demonstrate commercial applications
  - Evaluate and down-selection of applications for flight in December 2022

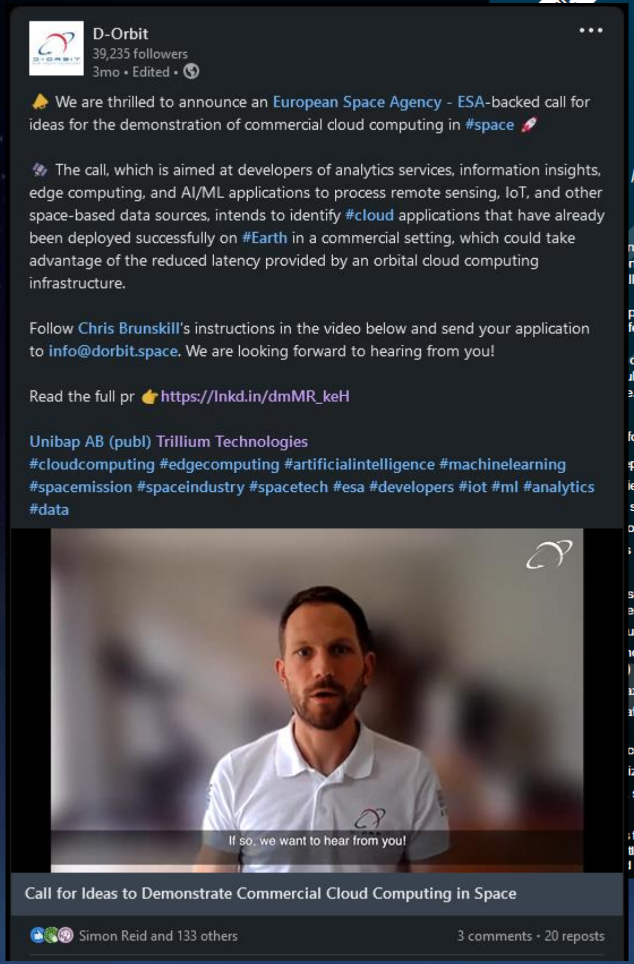
- The project was slightly rescoped to refocus the Trillium work package on the development of the “RaVAEn 2” algorithm, with the following objective:
  - Successful test of encoder to create a latent vector using pre-loaded test data
  - Test on VTT hyperspectral data with a band-configuration similar to Sentinel-2
  - Ground-based image reconstruction from latent vector

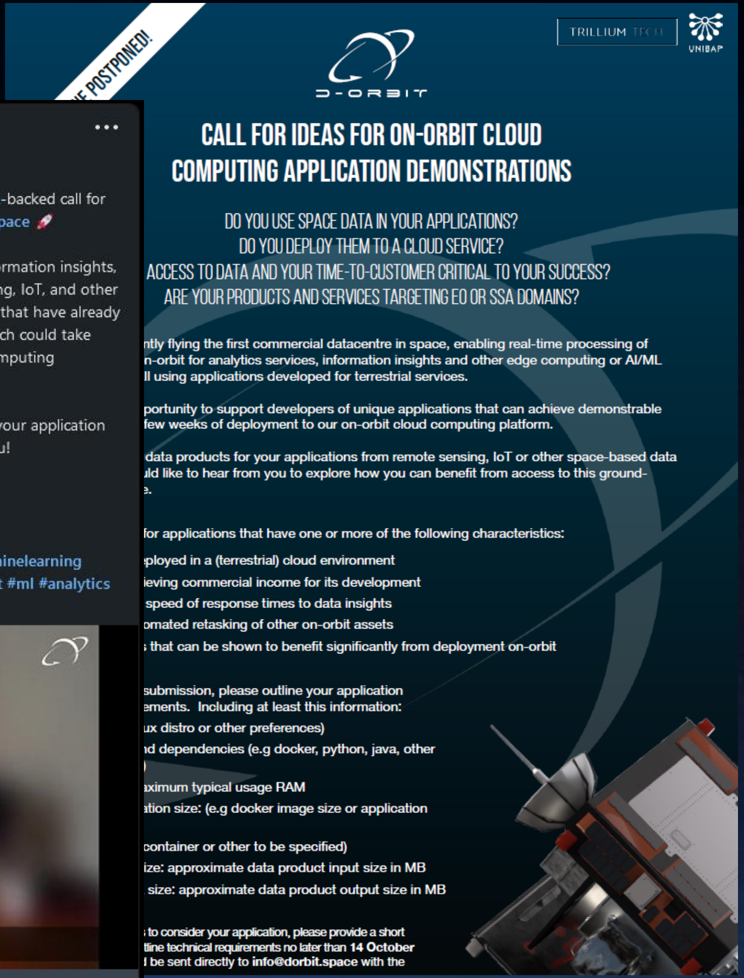


Call for Ideas

Call for Ideas public announcement

- A Call for Ideas was published to offer public access to the SCV-004 and iX5 platform to cloud app developers
- Apps were requested based on meeting the following criteria:
  - Currently deployed in a (terrestrial) cloud environment
  - Already achieving commercial income for its development
  - Sensitive to speed of response times to data insights
  - Enables automated retasking of other on-orbit assets
  - Applications that can be shown to benefit significantly from deployment on-orbit
- The public materials included:
  - A press release [1]
  - A Call for Ideas info sheet [2]
  - A launch announcement post on LinkedIn [3]
  - A deadline extension announcement on LinkedIn [4]





Evaluation process

1. **Receive responses**
   - Short summaries of apps and requirements
2. **Hold initial intro call with D-Orbit**
   - Introduce teams and ensure alignment with aims
3. **Update idea**
   - Developer submits updated app concept and outline questions by email
4. **App use case review**
   - Use cases reviewed on basis of use case to down select to most promising responses
5. **Technical call with UniBap**
   - Review technical concept and identify any blockers
6. **Proceed with development**
   - Where no blockers exist proceed with preparing the app for deployment to SCV-004

Response criteria

- A formal ICD was not used, to encourage a wide range of app ideas
- Applicants were asked to outline ideas detailing the following requirements:
  - OS (e.g. Linux distro or other preferences)
  - Runtimes and dependencies: e.g. docker, python, java, other frameworks)
  - Memory: maximum typical usage RAM
  - Total application size: e.g. docker image size or application binary size
  - Packaging: container or other to be specified
  - Input data size: approximate data product input in MB
  - Output data size: approximate data product output size in MB

Constraints review

- App developers were then asked to consider their concept in the context of the iX5 specification as flown:
  - Quad core x86-64 CPU
  - AMD Radeon GPU
  - SSD storage
  - Intel Movidus Myriad X Vision Processing Unit
  - Microsemi SmartFusion2 FPGA
- In addition to mission requirements:
  - Total data size for uplink is less than 50 MB
  - Total data size for uplink is less than 500 KB
  - Execution time on the iX5 is less than 12 minutes
  - The following data will meet the requirements for the apps
    - Pre-loaded Sentinel-2 data
    - VTT data acquired on-orbit, limited to a single acquisition (single image or burst of images)
    - D-Sense data acquired on-orbit, limited to a single acquisition (single image or burst of images)
    - Data uplinked with app and within the 50 MB total data size constraint
- Feasibility was then mutually and informally agreed with each app developer on a case-by-case basis

Evaluation matrix

Organisation	Use case	App demo summary	Input data requirements	Evaluation
Cysec	Not an app, but a selection of crypto services that can be provided to other developers using the on-orbit system.	An end-user app or development of a test app would be needed to utilise the services deployed by Cysec.	The services proposed do not require external data sources.	Not closely enough aligned with the characteristics described in the call for ideas, but an interesting future functionality to be considered for other activities outside of the scope of this project.
Gisky	Change detection service for monitoring highway "respect zones"	Terrestrial cloud-based service that manages satellite tasking requests and establishes list of areas of interest and provides these to the on-board app. On-board the app only flags cloud-free data for downlink	Pre-loaded Sentinel-2 Provide own sample data for uplink	While the cloud detection software is of interest, the app is too close to mission operations rather than data services. This is an interesting option but requires a longer-term project to demonstrate feasibility
Johns Hopkins University	Zero trust architecture distributed service marketplace	Implementation of trustless cloud service exchange smart contracts	None specifically but sample data could be used to represent data products	While a small demo could potentially be developed the scale of the concept was established to benefit from longer-term consideration
Little Place Labs	Cloud masking and image segmentation for land cover	Processing all Sentinel-2 sample data and reducing 1.2GB raw data to a few MB of analytics	Pre-loaded Sentinel-2	A well-rounded concept. Makes good use of software and hardware features of iX5, aims specifically to minimise data downlink requirements and provide data services for other users.
LMO	Demonstration of deep learning techniques for image processing	Run application on-board with post-loaded data	Provide own sample data for uplink	This application is already very mature and built for the iX5 SCFW platforms, also demonstrates use of platform for AI application
OHB Hellas	Multi-frame super resolution image processing to improve quality of data from cost-efficient sensors	Deploy multi-frame super resolution algorithm on the iX5 platform using data from VTT sensor	Sample or acquired VTT hyperspectral Sample or acquired D-Sense RGB	This application demonstrates the value of providing a general compute platform in addition to a use case that can enhance the current performance of the on-board sensors
Pixalytics	ML-based classification tool to provide decision-making insights on-orbit	Reduce app to docker image small enough to uplink to SCV-004 and run using data collected by VTT camera	Pre-loaded Sentinel-2 Sample or acquired VTT hyperspectral	Ideal scope of work with an interesting use case and specific need for VTT data. As a backup can also use pre-loaded Sentinel-2 data
SatSure	Infrastructure analytics to aid decision making processes in infrastructure, real estate, utilities, logistics industries, etc	Implement examples of LM inference application models to demonstrate building footprint detection, waterbody detection and tree canopy detection	Provide own sample data for uplink	A strong business case but meeting data requirements may not be most efficient use of operations time on SCV-004
The Server Labs	Security/IoT/HPC for weather/EO subscriptions/Satellite as code/mesh networking in space	A candidate app would be developed to understand the process of deployment to the platform and establish improvement options for evolution of the cloud infrastructure	Pre-loaded Sentinel-2 Provide own sample data for uplink	The nature of this business is interesting as their primary focus is on supporting deployment of new cloud infrastructures. This is expected to be useful in the long-term but is not consistent with the requirement to have an app with a clear end-user in mind
Zaitra	Cloud masking neural network	Create model in docker or binary format. Perform multiple inferences on-board. Demonstrate creation of binary mask or percentage coverage value. Create thumbnails to minimise downlink requirements and identify candidate images for downlink of full resolution to calculate accuracy	Pre-loaded Sentinel-2 Sample or acquired VTT hyperspectral Sample or acquired D-Sense RGB	Simple demonstration but AI methods have potential to make good use of iX5 functions

Little Place Labs

- **Use Case:**
  - Cloud masking and image segmentation for land cover
- **App demo description:**
  - Processing all Sentinel-2 sample data and reducing 1.2GB raw data to a few MB of analytics
- **Input data:**
  - Pre-loaded Sentinel-2
- **Evaluation:**
  - A well-rounded concept. Makes good use of software and hardware features of iX5, aims specifically to minimise data downlink requirements and provide data services for other users.

LMO

- **Use Case:**
  - Demonstration of deep learning techniques for image processing
- **App demo description:**
  - Run application on-board with post-loaded data
- **Input data:**
  - Provide own sample data for uplink
- **Evaluation:**
  - This application is already very mature and built for the iX5 SCFW platforms, also demonstrates use of platform for AI application

OHB Hellas

- **Use Case:**
  - Multi-frame super resolution image processing to improve quality of data from cost-efficient sensors
- **App demo description:**
  - Deploy multi-frame super resolution algorithm on the iX5 platform using data from VTT sensor
- **Input data:**
  - Sample or acquired VTT hyperspectral
  - Sample or acquired D-Sense RGB
- **Evaluation:**
  - This application demonstrates the value of providing a general compute platform in addition to a use case that can enhance the current performance of the on-board sensors

Pixalytics

- **Use Case:**
  - ML-based classification tool to provide decision-making insights on-orbit
- **App demo description:**
  - Reduce app to docker image small enough to uplink to SCV-004 and run using data collected by VTT camera
- **Input data:**
  - Pre-loaded Sentinel-2
  - Sample or acquired VTT hyperspectral
- **Evaluation:**
  - Ideal scope of work with an interesting use case and specific need for VTT data. As a backup can also use pre-loaded Sentinel-2 data

Zaitra

- **Use Case:**
  - Cloud masking neural network

- **App demo description:**
  - Create model in docker or binary format. Perform multiple inferences on-board. Demonstrate creation of binary mask or percentage coverage value. Create thumbnails to minimise downlink requirements and identify candidate images for downlink of full resolution to calculate accuracy

- **Input data:**
  - Pre-loaded Sentinel-2
  - Sample or acquired VTT hyperspectral
  - Sample or acquired D-Sense RGB

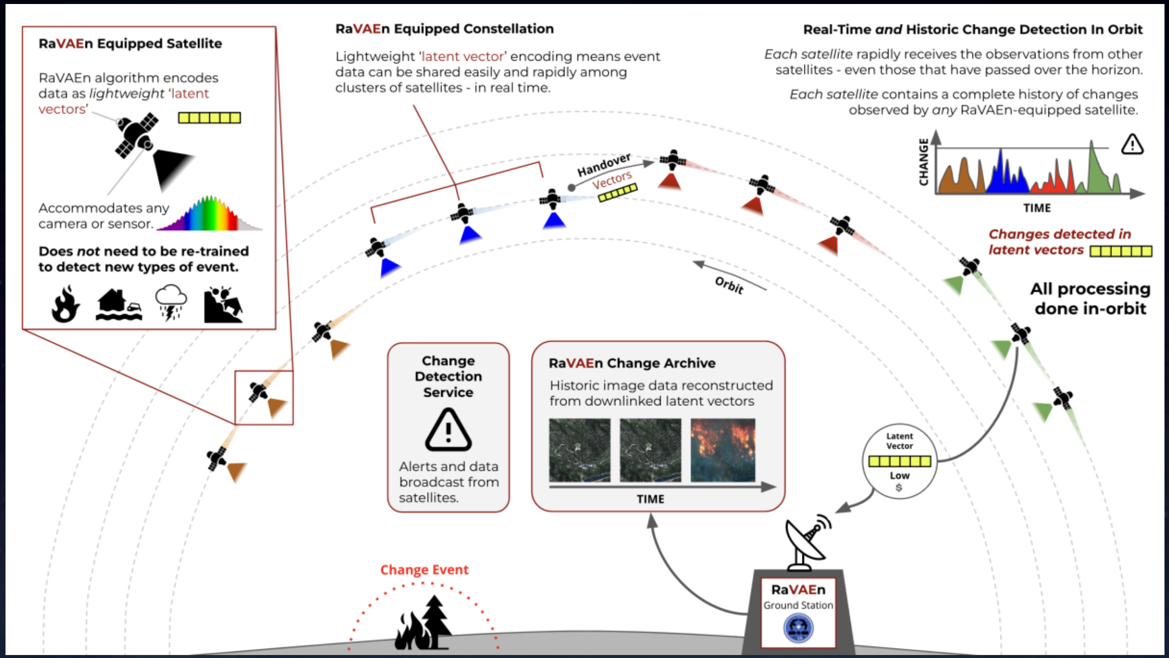
- **Evaluation:**
  - Simple demonstration but AI methods have potential to make good use of iX5 functions

RaVAEn 2

RaVAEn 2 development

Update from Trillium on application concept for this WP

- Opportunity to adapt the RaVAEn change detection algorithm for deployment on SCV4
- Expect 3CS project would see:
  - Successful test of encoder to create a latent vector using pre-loaded test data.
  - Test on VTT hyperspectral data with a band-configuration similar to S2.
  - Ground-based image reconstruction from latent vector.
- Foundation for an exciting demonstration using SCV11 in May:
  - Detect change in sequential images from an advanced sensor.
  - Demonstrate processing of imagery transmitted from satellite-to-satellite.



Major tasks

- ❑ Scoping of work needed to adapt RaVAEn to SCV4
- ❑ Re-train or adapt current S2 model
- ❑ Work with Surry Space Centre, who have offered support

RaVAEn 2

Deployment on-board ION-SCV 004 satellite

Researcher: Vít Růžička (vit.ruzicka@cs.ox.ac.uk)
Advisors: Gonzalo Mateo García, Chris Bridges,
Cormac Purcell, Andrew Markham

Introduction

- With more deployed CubeSats, there is a **raising need for autonomy**
- **Decision making on-board**, in communication constrained scenarios
  - **Compress** data for communication (send only encoded latents, or insight, not the raw data)
  - **Detection of events on-board** for early **discarding** (example: clouds) or of **transient events** for autonomous scheduling of follow-up observations (example: anomalies, disaster, change)
- **Training on-board**:
  - Demonstration of training on board as a possibility
  - For **autonomous calibration of instruments** in cases of sensor degradation
    - Observing scene “X” we know the true labels “Y”. Can we train a model to predict X->Y?
- *Background literature:*
  - Giuffrida, G. et al. The ϕ-sat-1 mission. 2021
  - Mateo-Garcia, G. et al. Towards global flood mapping onboard low cost satellites with ML. 2021

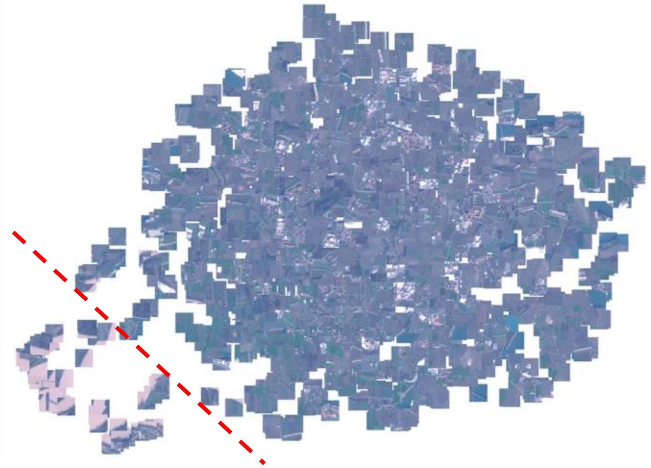
Method


InspiredResearch Winter 2022 Issue 21

Research News from the Department of Computer Science, University of Oxford

AI Onboard of Satellites for Autonomous Detection of Disaster Events

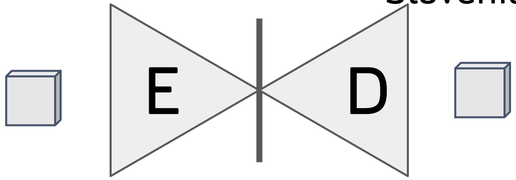
By Doctoral Student Vit Růžička



tSNE visualisation of latent space

- **RaVAEn**: Variantional auto-encoder (VAE)
  - Learning **general encoding model** from samples of data given diverse training dataset
    - Capable of encoding the input image tiles (*32 x 32 x 4 or 10 bands, in 12 bit, S2 L1C data*) into a latent vector (*of 128 with 16 bit float*), and allows for reconstructing.
    - It's latent space separated distinct samples from each other (example: *flood tiles* vs *non-flood tiles*).
  - Models trained in [Růžička, V., et al. 2022] are used on the demo Sentinel-2 data already available on board of the ION-SCV 004 satellite (some re-normalisation was needed).
  - Tested several *"small VAE"* model architectures, on 3, 4 or 10 bands (only 4 bands with real data)

Demo S2 data on-board, Slovenia

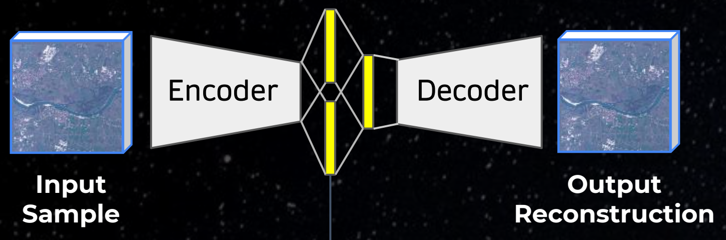


Model pre-trained in RaVAEn

The *Change Detection* ML Payload (FDL Europe 2021)

RaVAEN Algorithm

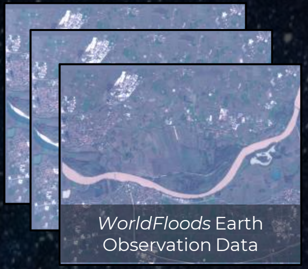
Variational Auto-Encoder (VAE)



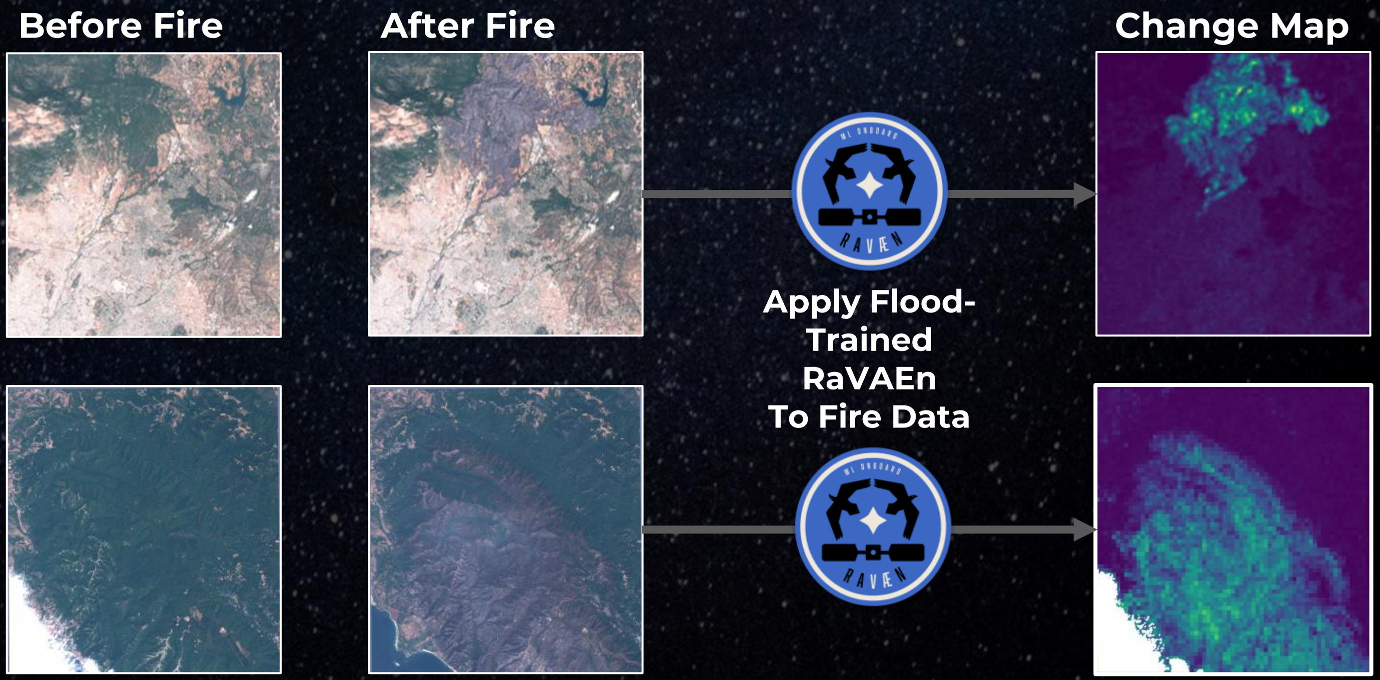
Latent vector that learns to encode interesting signals in data

- → Train the VAE to learn about instrument & data characteristics.
- → Apply VAE to detect change of any sort (fires, floods, hurricanes etc …)
- → Robust to noise and data shifts.
- → Low memory and processing requirements.

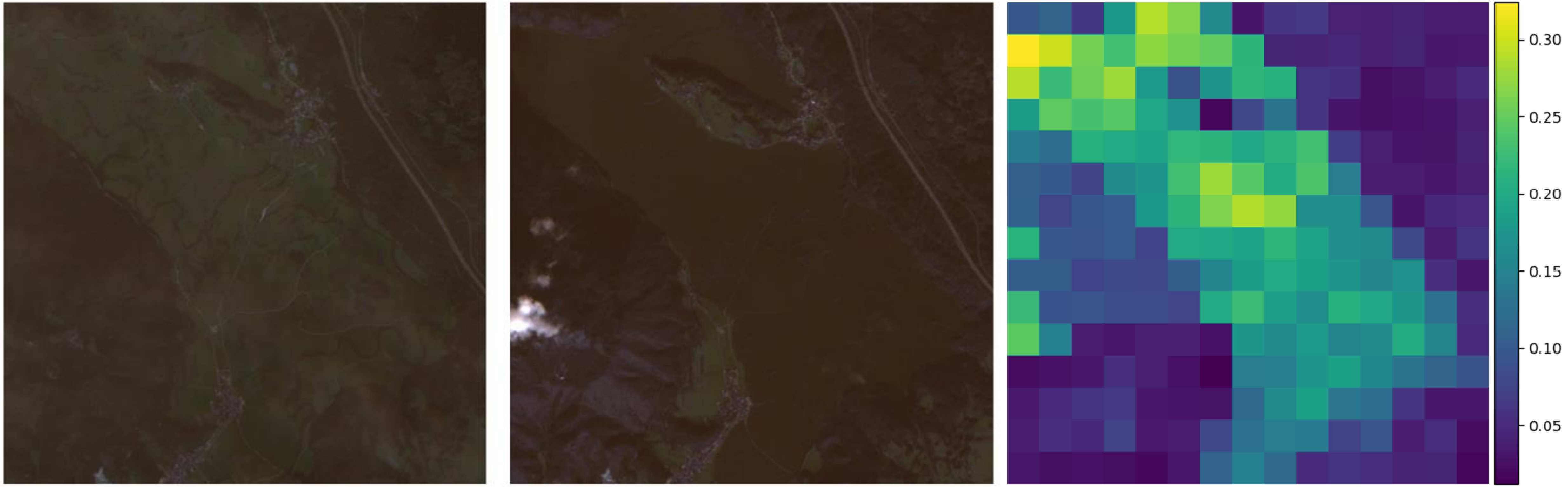
Detecting Burn Scars in Greece During 2021 Fires



RaVAEn trained on S2 images of **floods** (learns instrument).



Example 1: change detection on the available dataset



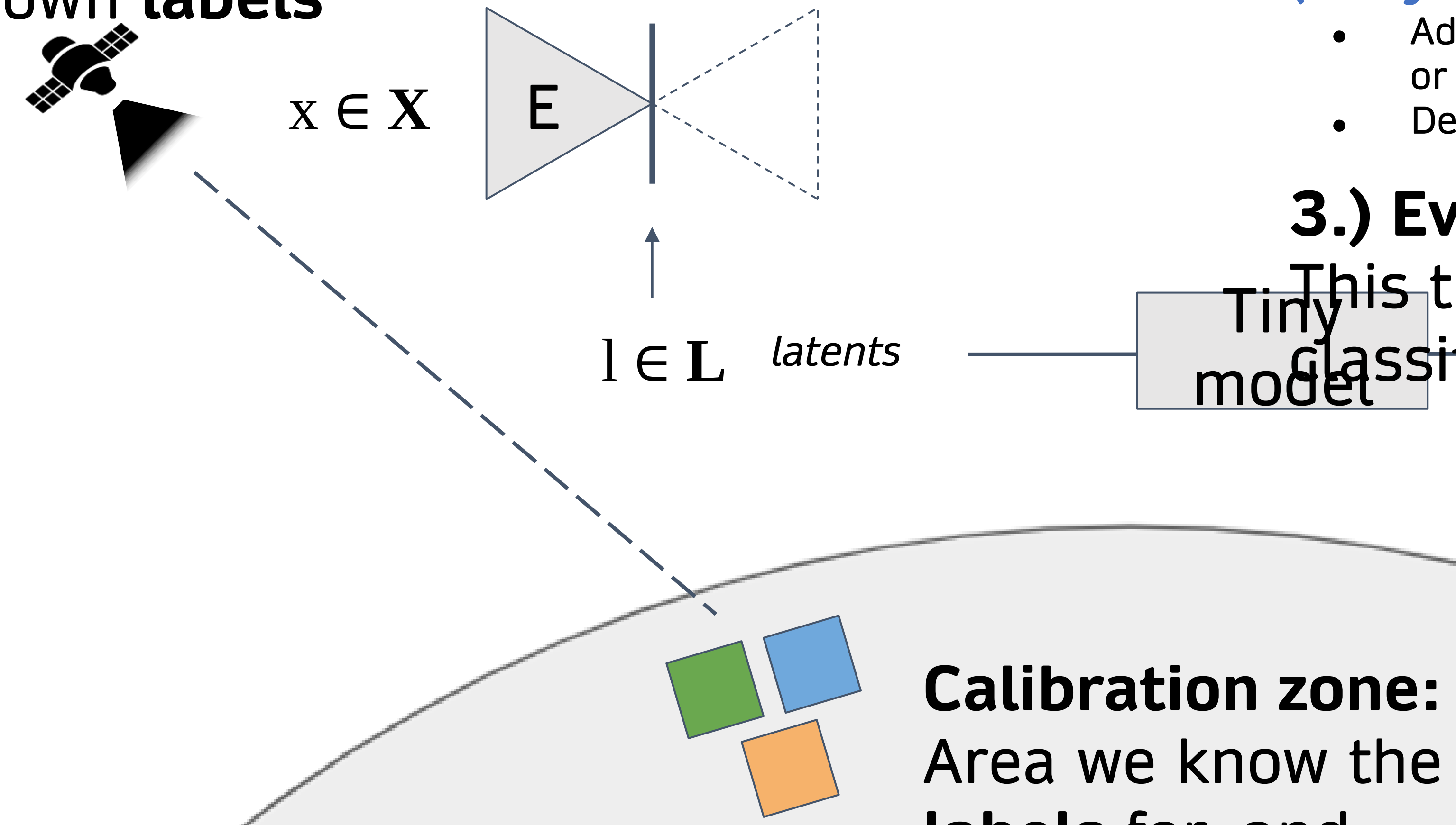
WIP, Early Results

Detected flooded area ^

Motivation for training on-board

1.) Calibration step:

After observing the known area (given we know our location), we can compare the **raw observation** with known **labels**



2.) Training step:

With known data (L) and labels (Y) we can train a model on-board.

(Very small and very fast!)

- Adapt in case the sensors are malfunctioning or if we have data shift
- Detect that a large shift has occurred

3.) Evaluation step:

This tiny model can be used to classify tiles directly on-board.

No need for a human in the loop! Can be fully automatic.

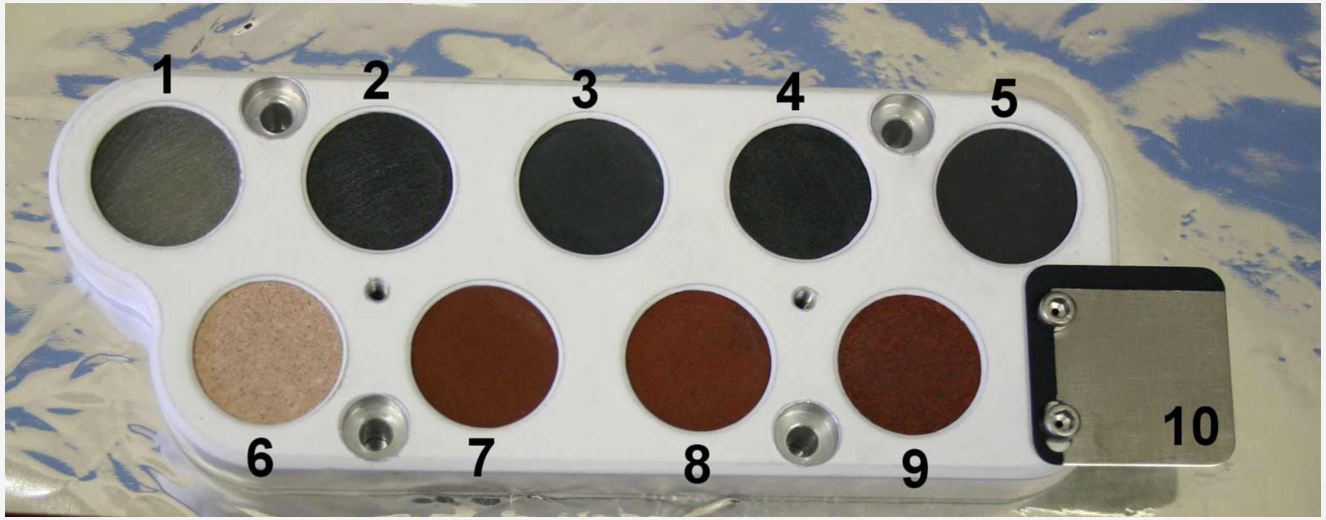
Calibration zone:
Area we know the **labels** for, and which won't

y ∈ Y

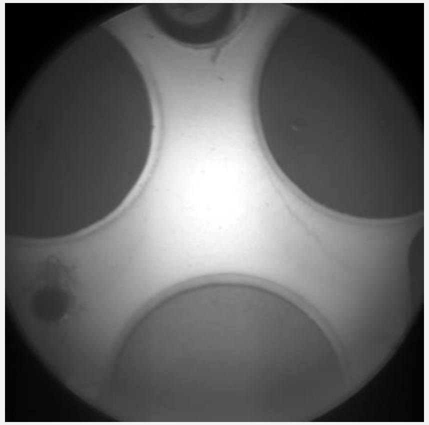
Motivation for training on-board

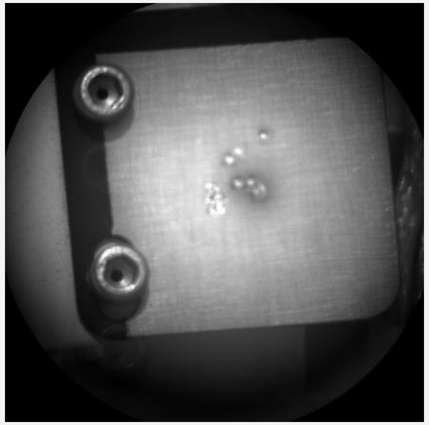
1.) Calibration step:

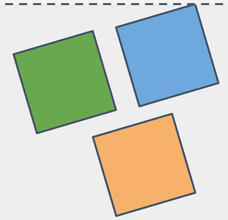
Example from the Curiosity Rover for ChemCam instrument:



2.) Training step:







Calibration zone:
Area we know the **labels** for, and which won't

$y \in \mathbf{Y}$

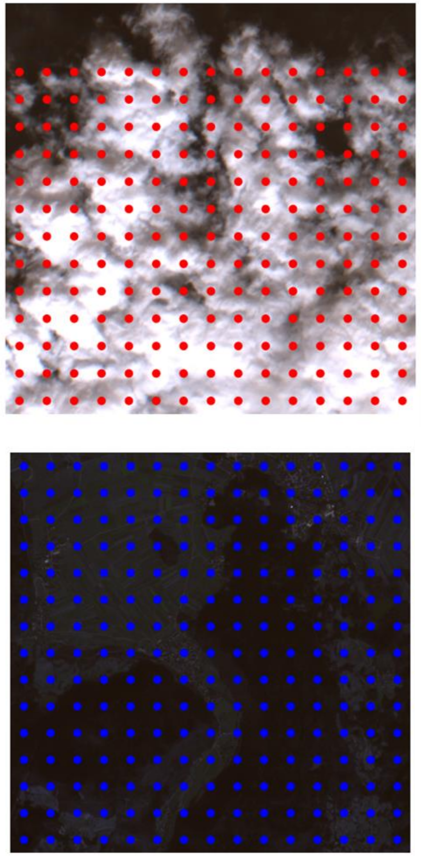
Method

- **Few-shot learning,** training on-board in a representative latent space
  - Using very **few and easily selected samples for a training dataset** of cloudy vs non-cloudy tiles (from 6 images in total), evaluating on the rest of the available data (qualitatively for now, no labels were available)
  - Training inside the latent space: for each tile, we use the encoded space of 128 numbers.
  - **Very simple models: single fully connected layer with variable number of classes** at the output - 1 for binary, 4 for presumed classes from the CloudSEN12 dataset and 12 for presumed land cover classes (Note: no actual and labelled training data is available for this yet).

Example 2: using the model trained … board

WIP, Early Results

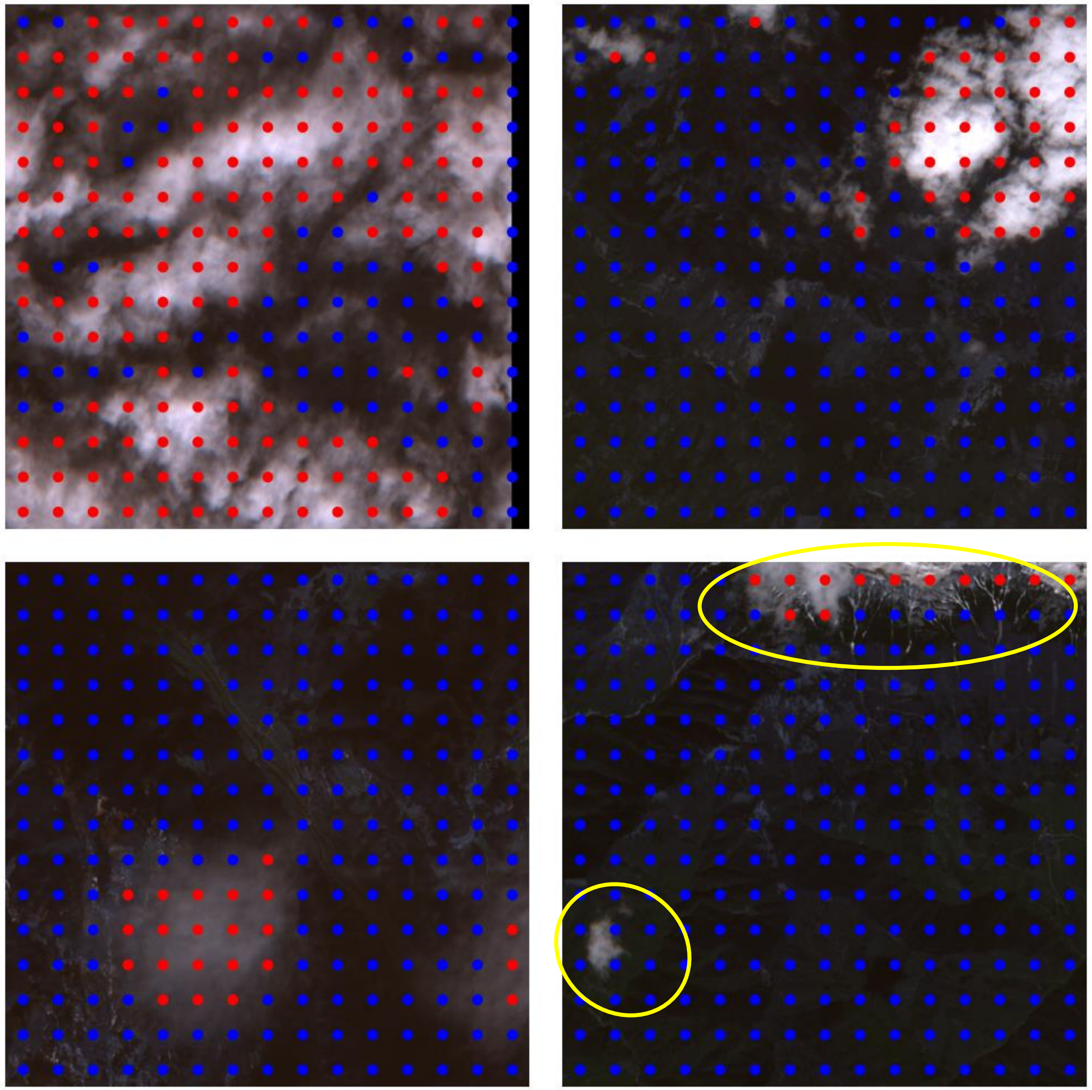
[just as a demo task!]



Training dataset - 650 tiles with and 650 without clouds (just 6 images)

Few-shot learning with latents

Results on the rest of the data:



- ● = cloud
- ● = no cloud

Can be used to get a rough **cloud cover %**

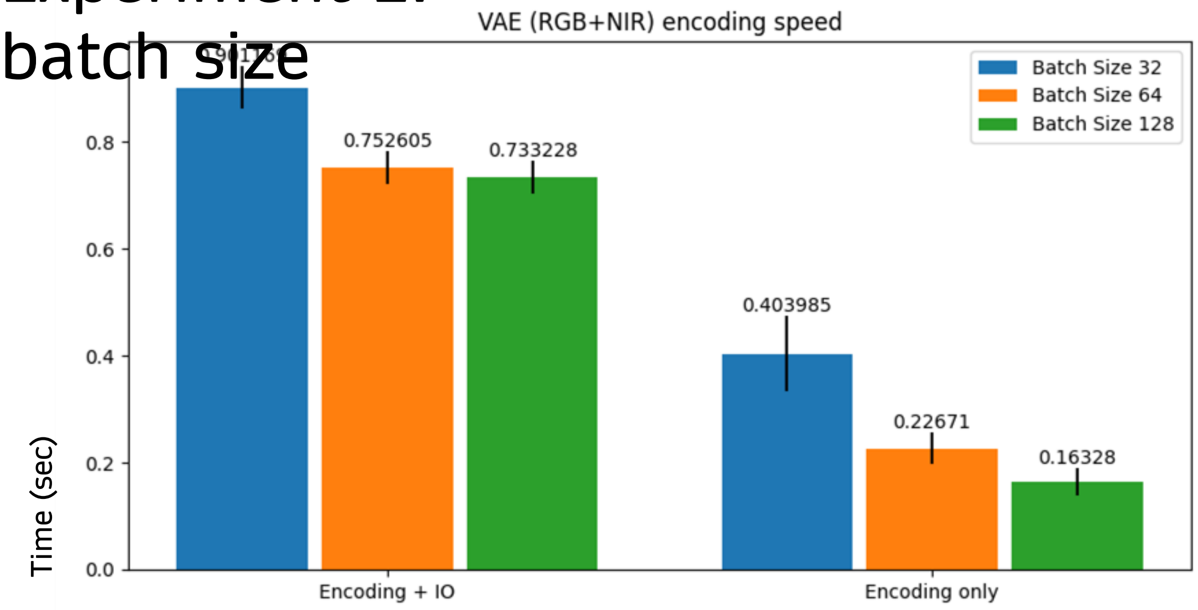
< Some errors, on tiny clouds, and on snowy areas *(note that they were not represented in the dataset!)*

Results

- Experiments:
  - VAE Encoder:
    - Batch sizes: 32,**64,128**
    - Model with different number of bands: 3,**4**,10
    - Compute device: Torch CPU, Openvino CPU, **Openvino MYRIAD**
  - Tiny classification head:
    - Batch sizes (exploring restrictions from the point of view of the available memory): up to **256**
    - Number of classes, number of model parameters
- Timing:
  - **VAE Encoder**: 0.5 - 0.6 sec IO, 0.1-0.2s encode, ~0.01s latents compare for change detection
  - **Training** classification model: **1 epoch** with 1305 samples = ~0.1 sec
  - **Inference** with classification model: ~ estimated 0.03 for the whole dataset of 1305 samples
- Tiny classification model score:
  - On a small manually labelled test set (with different location ids) of cloud/no-cloud tiles: **AUPRC 0.979**, Threshold of 0.5: Recall 0.946 Precision 0.967, F1 0.956

WIP

Experiment 1: batch size

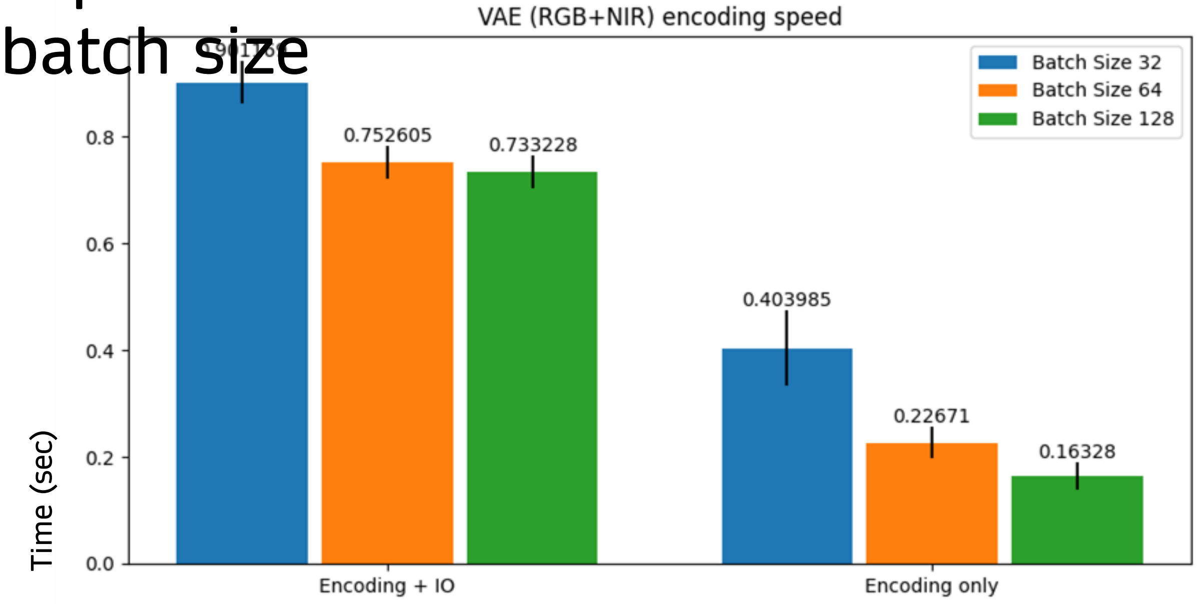


Orange marks the same model: VAE with 64 batch, RGB+NIR, Torch CPU

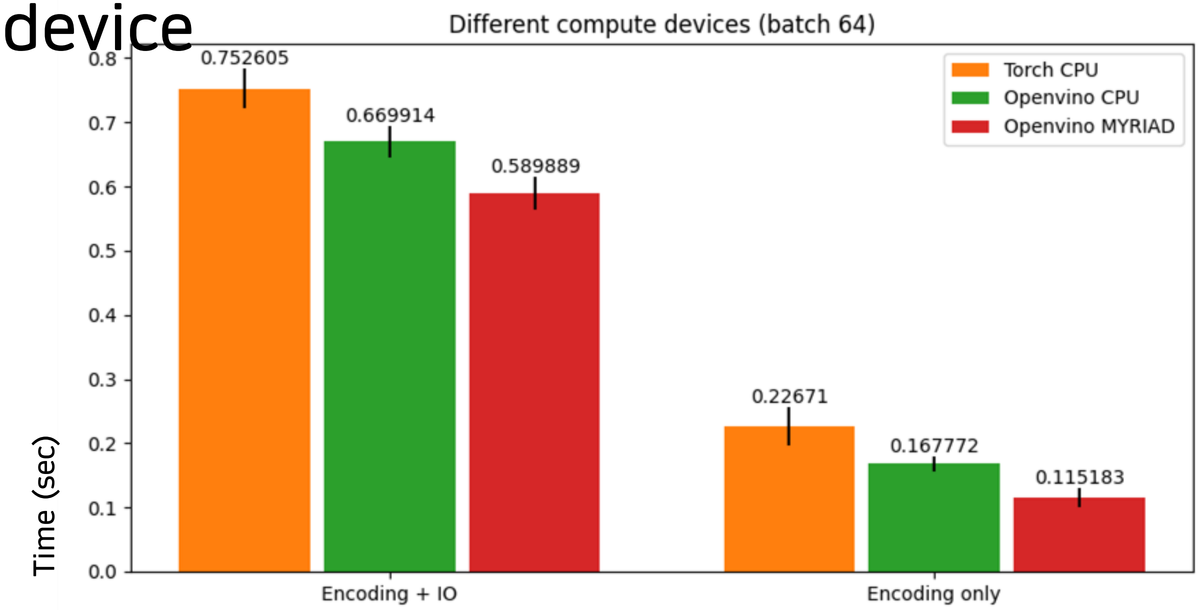
- Optimal **batch size 64 or 128** (depends on the size of the data, currently each image of 5x5km^2 area is tiled … **tiles**, 15x15 gr…

WIP, Early Results

Experiment 1: batch size



Experiment 2: device

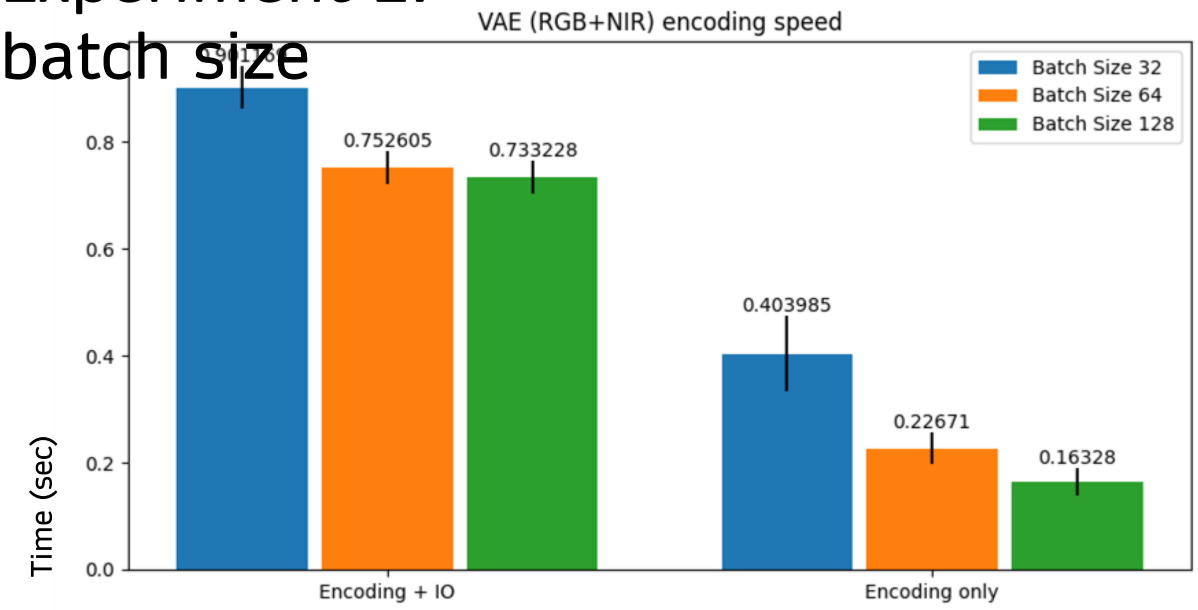


Orange marks the same model: VAE with 64 batch, RGB+NIR, Torch CPU

- Performance on the **MYRIAD chip is ~2x faster than Torch CPU** (0.226 -> 0.115)

WIP, Early Results

Experiment 1: batch size

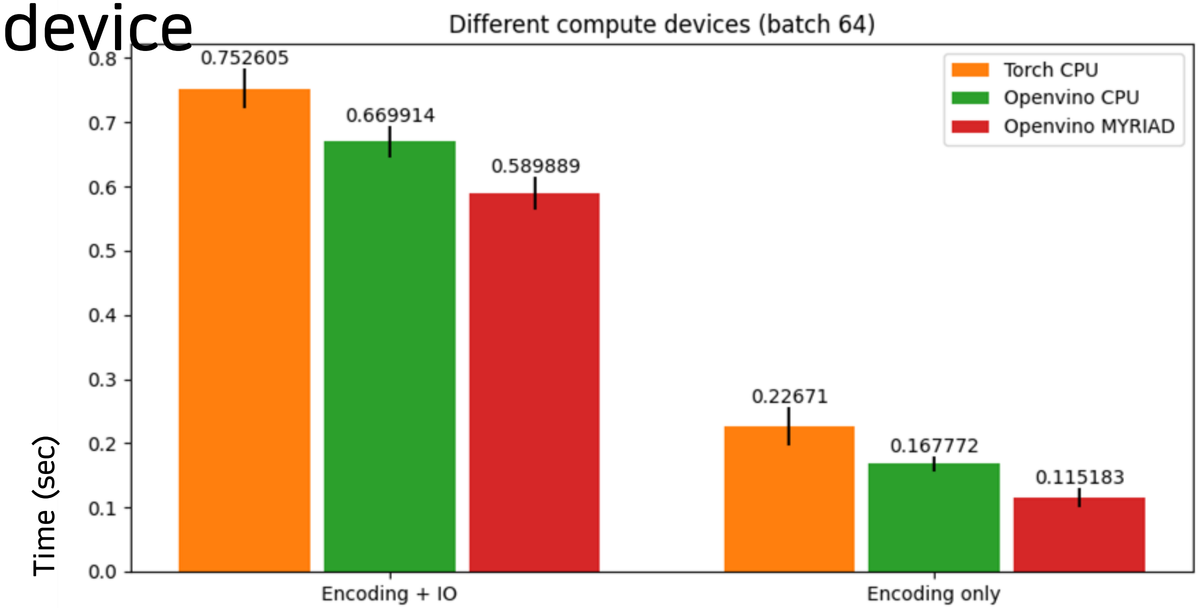


Orange marks the same model: VAE with 64 batch, RGB+NIR, Torch CPU

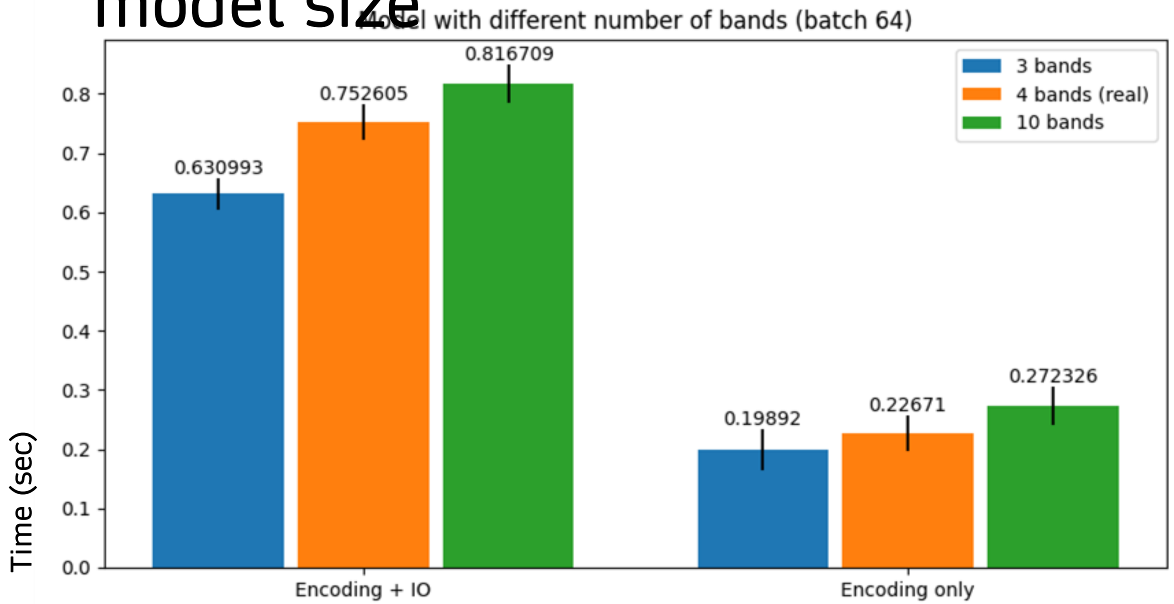
- Impact of the model size (and the number of parameters) is notable, but reason…

WIP, Early Results

Experiment 2: device

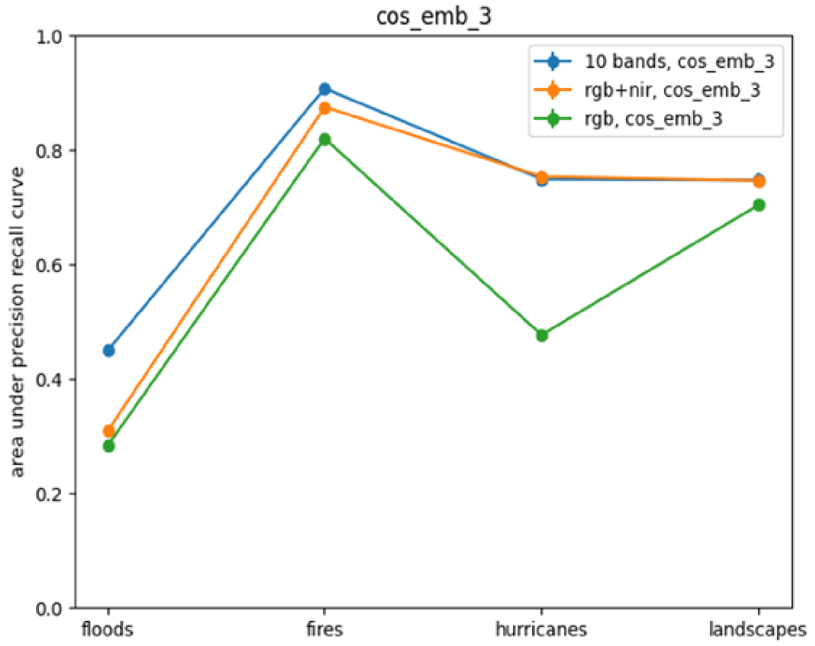


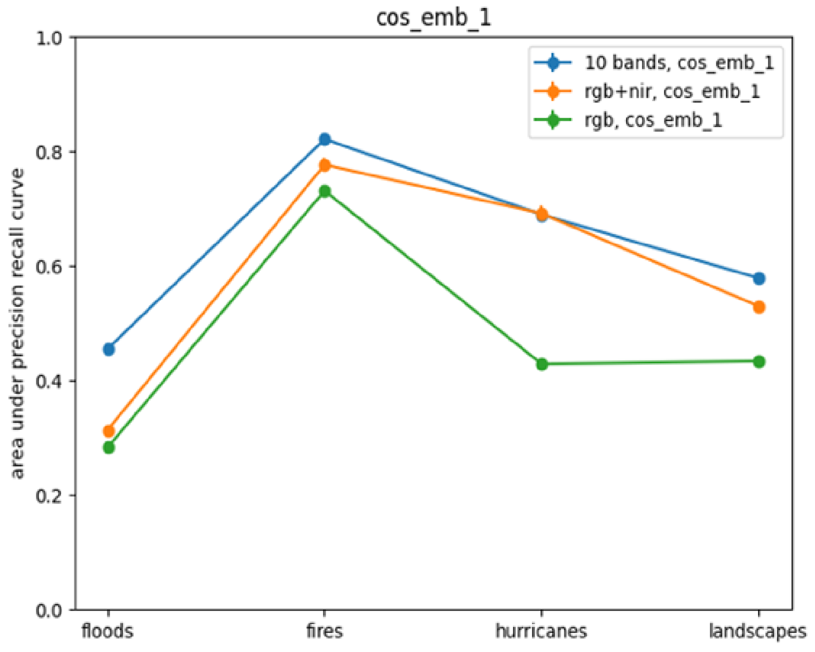
Experiment 3: model size



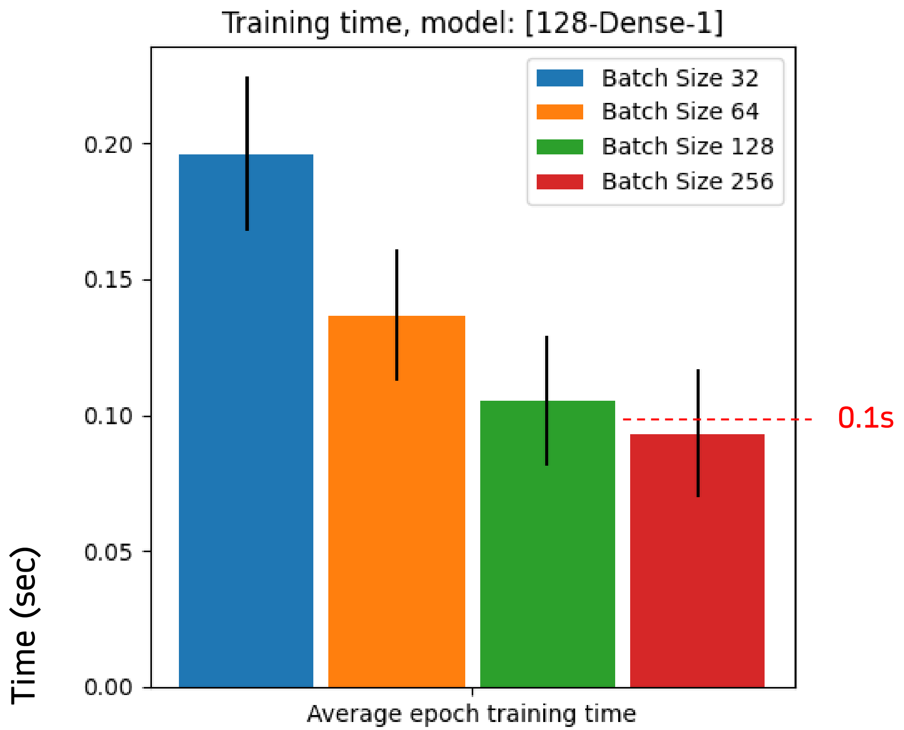
Viability of RGB+NIR only model

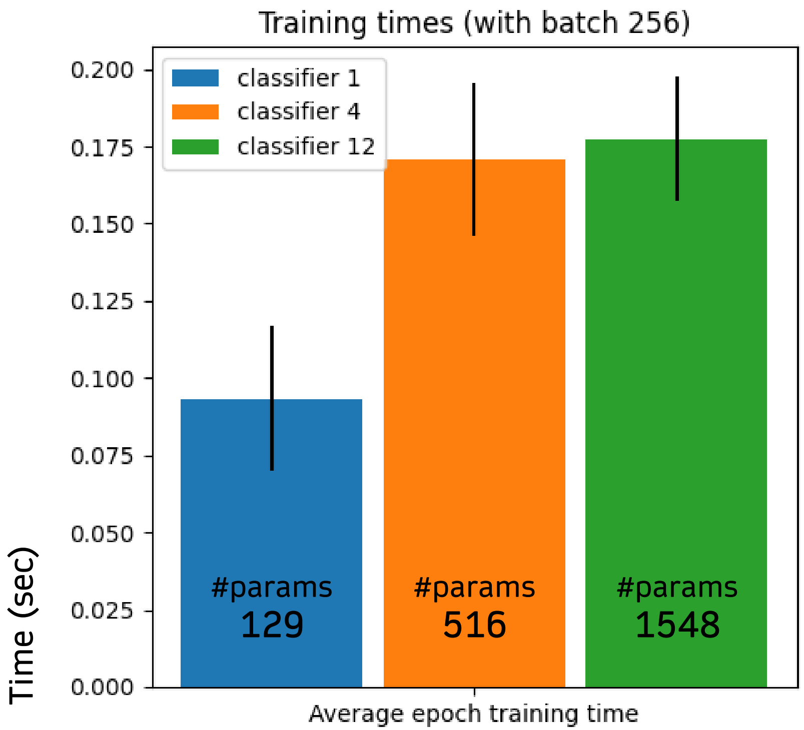
- From earlier results on the released dataset of disaster events (in the RaVAEn paper), the performance drops with having access to less bands (than the referential model using 10 high resolution bands)
- Drop from 10 bands to RGB+NIR is smaller than the drop from RGB+NIR to only RGB
- Timing tests with models using dummy data with 10 bands gives us an insight how fast they could run (not significantly slower)





Training on-board



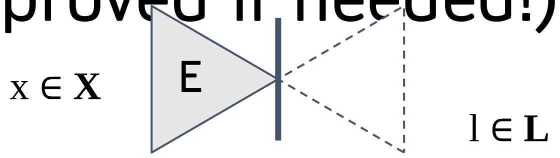


- Training is faster and feasible even with larger batch sizes (tested up to 256).
  - 1 epoch = 1305 samples
  - Trained for 10 epochs

- Larger classification models are slightly slower (note that all are using just 1 fully connected layer)

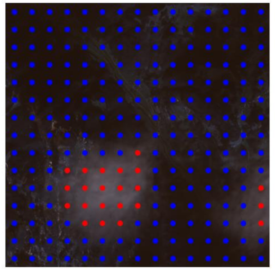
Conclusions #1

- With the current generation of hardware on-board of satellites, it is possible to use our system for:
    - **Unsupervised change detection with the previous passes** [metrics in | Růžička, V., et al. 2022]
    - **Reduction of communicated or stored data** when using only the latent representations (likely can be further improved if needed!)
        - 10 bands: (32*32*10**12*) / (128**16*) = 60 x times
        - 4 bands: (32*32*4**12*) / (128**16*) = 24 x times

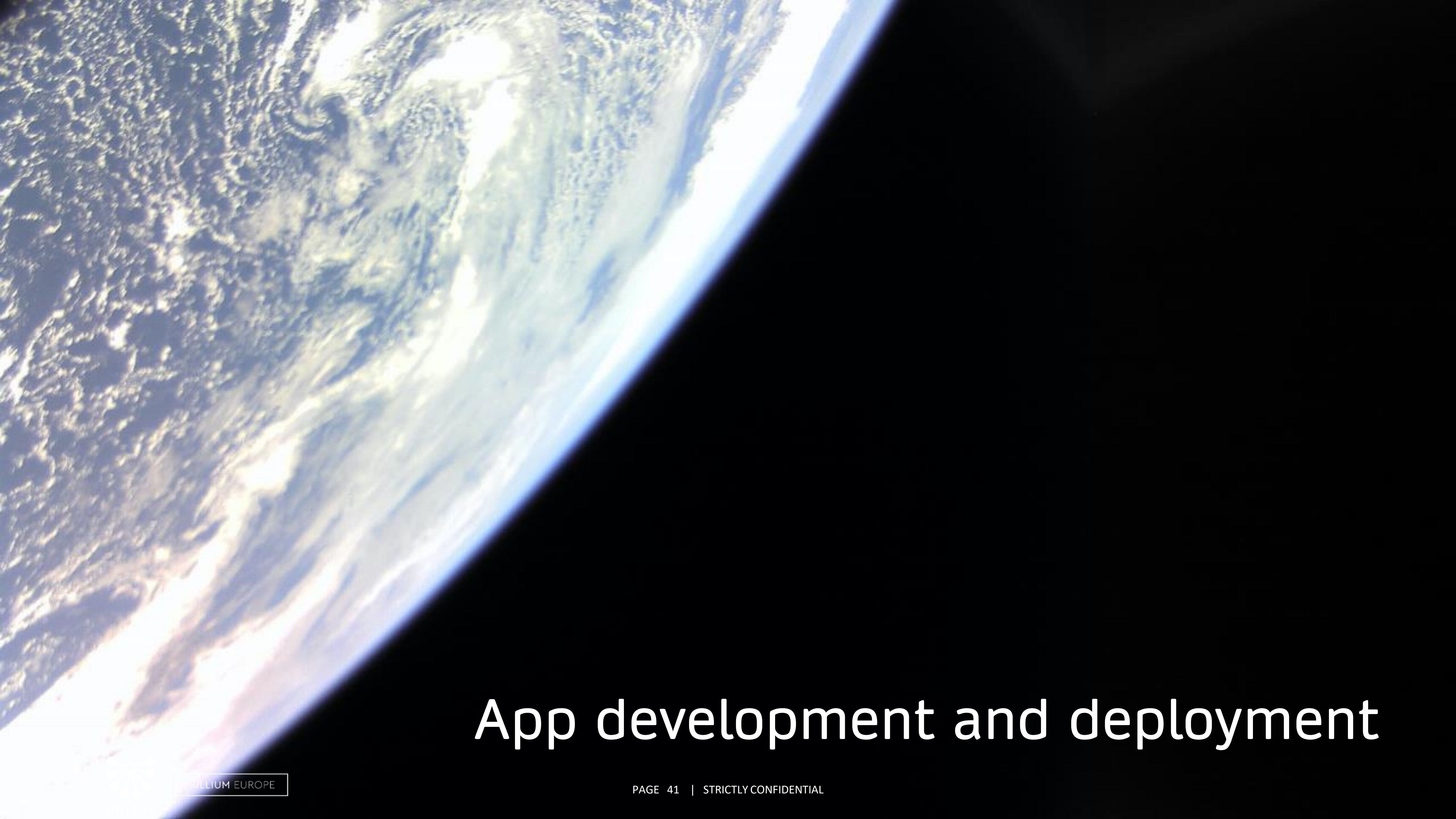


- As a software package, it can run on-board the device without causing errors, the taken time per data batch is stable (**doesn't cause unexpected overheads**) - as far as I know!

Conclusions #2



- We show that **even with a very small annotated training dataset, it is capable of training a machine learning model on-board, and that the model can directly be used** (for example to estimate the cloud cover in images)
- **Very fast processing speed** (full 5x5km area _safely_ < 0.7 sec, full training _safely_ < 1.5 sec).
  - Could lead to a whole array of new ML being possible and being designed for deployment on-board real world satellites.



App development and deployment

SCFW On-boarding Process

- Step 1: The app developer creates their app in their own development environment
- Step 2: The app developer ports their app to run in SCFW using the UniBap software simulator and a base docker image
- Step 3: The application provider verifies the app in a hardware test system accessed remotely at the UniBap laboratory
- Step 4: UniBap validates the app in a reference system (EM) on the ground
- Step 5: The app is packed and sent to the D-Orbit flight ops team and a record added to the iX5 operations schedule

Flight operations and ConOps planning

- A flight planning procedure is in place to coordinate planning of iX5 operations on-board SCV-004
- For each tasking request of a test is added to a central schedule, describing:
  - Functional mode of the iX5
  - Requirement for use of the VTT camera
  - Outline of success criteria
  - Any specific support needed
  - OBCP commands
- Tasking is then allocated based on ground station availability on a case by case basis

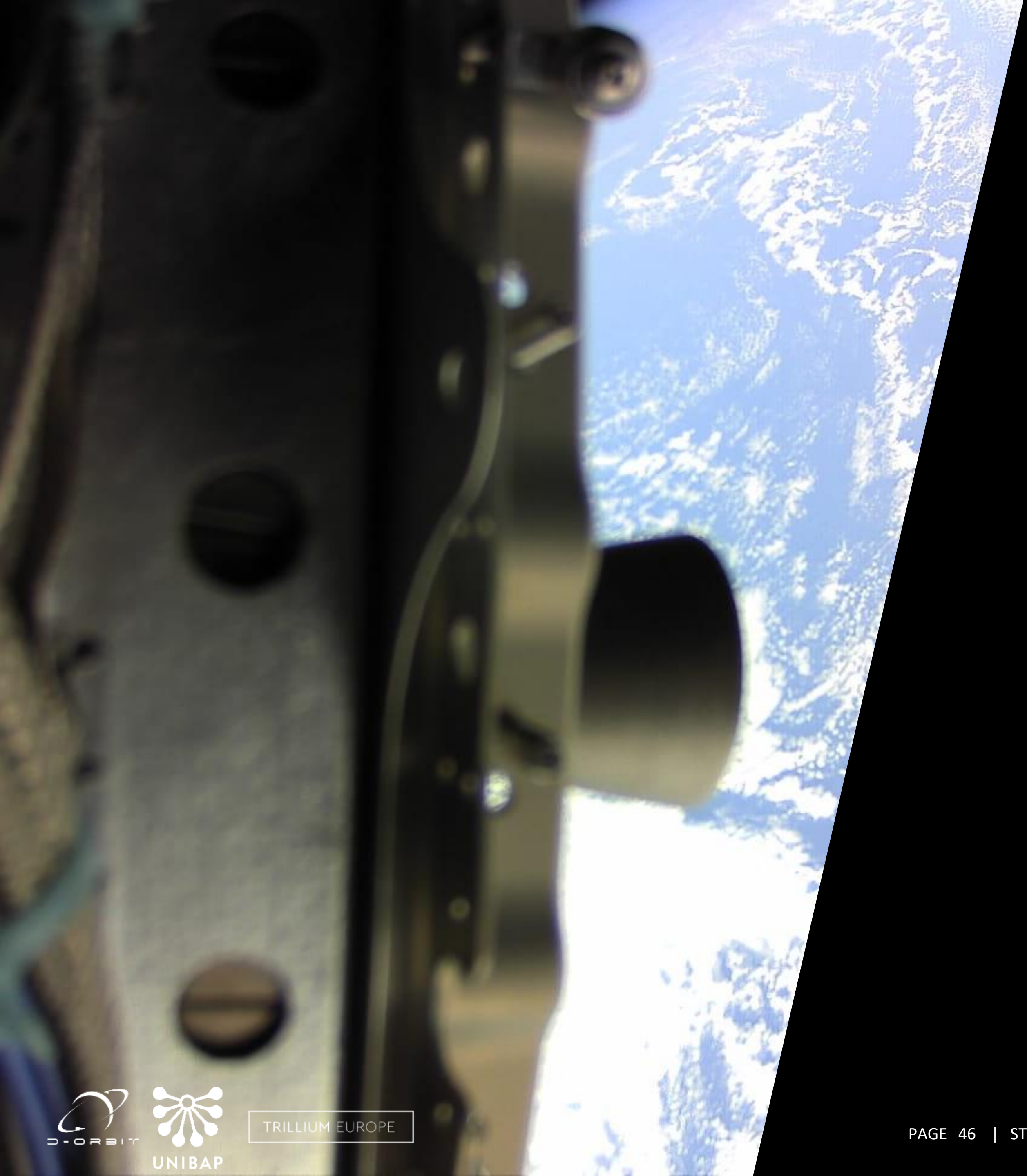
- Summary of ConOps
  - Test files are uploaded to D-Orbit server
  - Test files are uplinked to ION bus
  - Test files are transferred to iX5
  - iX5 is scheduled to turn on and execute any tasks defined in the test
  - iX5 transfers results/app outputs and logs to ION bus
  - Outputs are downlinked via D-Orbit ground segment and uploaded to server
  - User downloads test files

Preliminary results

Deployment Status

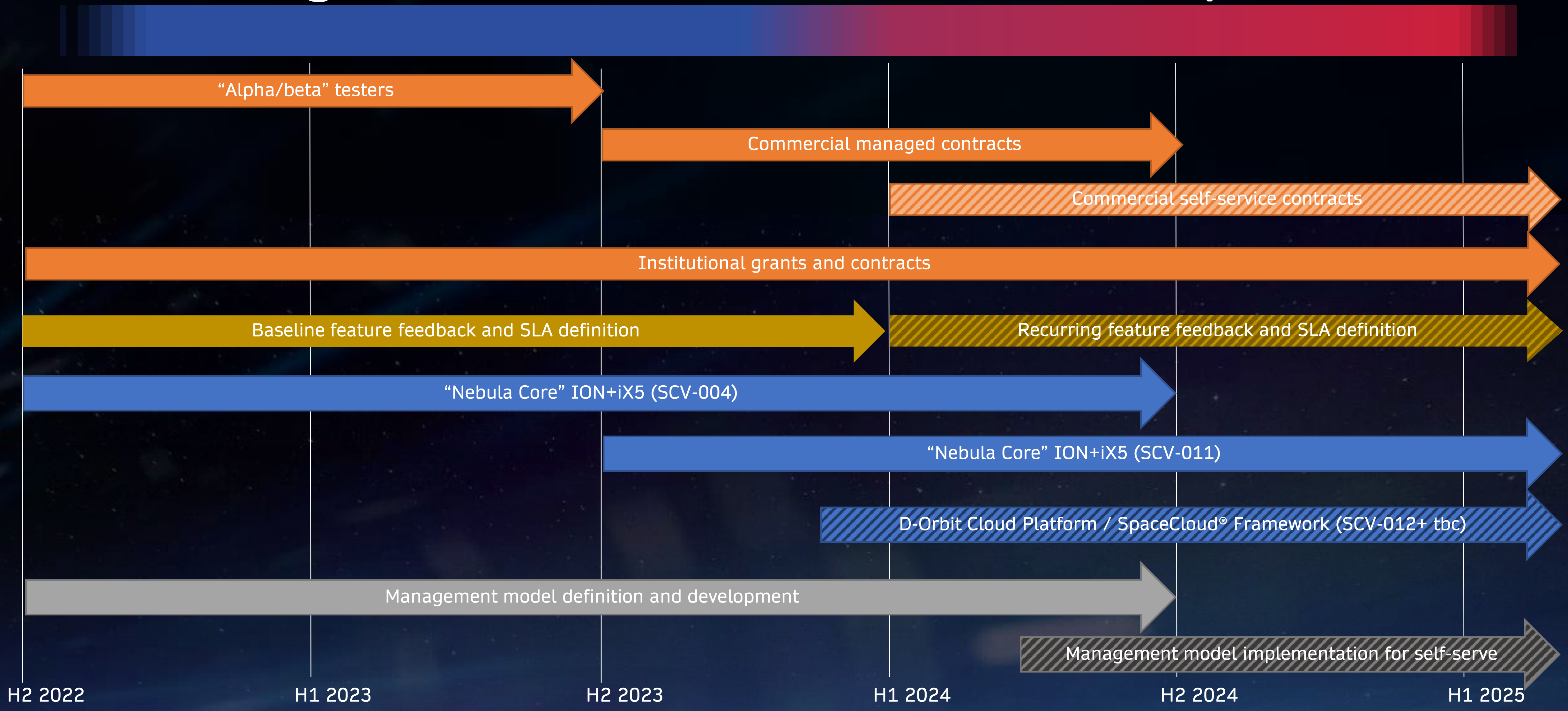
App developer	App submitted	Intro call	Tech call	App prepared	Tested on ground	Uplink planned	App on-boarded	App demo complete	Results down-linked
Little Place Labs	Y	Y	Y	Y	Y	Y			
LMO	Y	Y	Y	Y	Y	Y	In progress		
OHB Hellas	Y	Y	Y	In progress					
Pixalytics	Y	Y	Y	In progress					
Zaitra	Y	Y	Y	Y	In progress				
Trillium	Y	Y	Y	Y	Y	Y			

- NB. In the context of the project, the results are based on the on-boarding and deployment of end-user apps
- Following completion of the demonstrations, feedback will be requested from the developers to support analysis of the science/tech capability compared to terrestrial alternatives



Next steps

High level commercialisation plan



TRILLIUM EUROPE

Lessons learnt

- Developer engagement was not as diverse as expected
  - Commercially-driven apps serving specific commercial needs were under-represented in the responses when using the promotion methods developed for this project
- Fundamental platform concept is viable
  - It has been shown that the app developers down-selected were quickly able to prepare and deploy apps to SCV-004 – measured in a number of weeks
- Coordination of support between D-Orbit (flight operations) and UniBap (development environment is complex
  - Both elements are interdependent and have lots of opportunity for optimising to facilitate better “user journey”
- App developers were focused on the data
  - While the data requirements and the VTT camera in particular were an important factor, they resulted in a “mission operations” focus, rather than cloud app focus
- Flight operations scheduling and planning is very manual
  - There is also a lot of opportunity to improve the automation of these processes
- There is a clear nascent interest in cloud apps in space
  - Driven by interest in AI applications and access to novel datasets (e.g. VTT hyperspectral images)

Communications plan

- Conference and workshop papers presenting the proof of concept of a commercial cloud service in space
- Peer-reviewed papers on the results of the RaVAEn 2 demonstration
- Other papers with many of the app developers
- On-line communication and social media
- Press releases

Future work

- Continuation of the app demonstrations, targeted to be completed by Jan-Feb 2023
- Extension of the RaVAEn 2 work
- Improvement of preparation, test and deployment processes to introduce more automation and move to a self-service operational model
- Further ION missions including the iX5 on-board processor
    - Additional app developers
    - New data sources
    - Co-hosting of sensors and payloads
    - Improvement of communication systems