CHIME: THE FIRST AI-POWERED ESA OPERATIONAL MISSION

Raffaele Vitulli(1), Marco Celesti(1), Roberto Camarero(1), Gianluigi Di Cosimo(1), Ferran Gascon(1), Nicolas Longepe(1), Marco Rovatti(1), Michel Francois Foulon(2), Adrien Grynagier(2), Dimitri Lebedeff(2), Yves Bobichon(2), Jakub Nalepa(3), Agata Wijata(3), Mateusz Przeliorz(3), Theo Dullin(4), Jens Nieke(1)

(1)European Space Agency, Keplerlaan 1, 2201AZ Noordwijk, Netherland, Raffaele.Vitulli@esa.int
(2)Thales Alenia Space, 26 avenue J.F. Champollion BP 33787, 31037 Toulouse Cedex 1 France, michel-francois.foulon@thalesaleniaspace.com
(3)KP Labs, Bojkowska 37, Gliwice, 44-100, Poland, jnalepa@kplabs.pl
(4)KILI Technology, 40 Rue du Colisée, 75008 Paris, France, theo.dullin@kili-technology.com

Abstract - Over the last decade, rapid developments in digital technologies and in our capability to monitor our home planet from space with Earth Observation (EO) satellites have enabled unprecedented monitoring of the Earth's environment bringing new and huge opportunities for science and businesses. CHIME will produce a near-continuous stream of high-dimensional data resulting in unprecedented data volumes. This calls for fast and computationally efficient methods for the storage, transmission and analysis of the data. Artificial Intelligence (AI) processes carried out on-board of the satellite holds special promise for hyperspectral missions such as CHIME, as processing the data on-board can reduce the time and cost of data transfer and processing, enabling to send only the most valuable insights to the ground, and focusing on rapid responses to events and detected phenomena that impact our society and require fast decision-making. CHIME makes use of AI-powered data management as part of its baseline. Machine Learning-based cloud detection on-board the CHIME satellite allows to reduce the unprecedented data volumes while maximizing the science return. A more ambitious goal has been also set and is currently being analyzed, in order to enhance CHIME on-board intelligence and autonomy capabilities by means of AI. The objective is to identify new potential use cases suitable for AI processing on-board (e.g., rapid response and environmental awareness, for fast decision-making). The paper will present the technological and operational challenges that the use of Artificial Intelligence on-board a satellite implies for an Operational Mission like CHIME, where high reliability and high availability are requested.

1 INTRODUCTION

The Copernicus Sentinel Expansion missions meet priority user needs not addressed by the existing infrastructure, and/or reinforce existing services by monitoring capability in the thematic domains of CO2, polar, and agriculture/forestry. Hyperspectral imaging today enables the observation and monitoring of Earth surface properties (geo-biophysical and geo-biochemical variables) thanks to the diagnostic capability of spectroscopy provided through contiguous, gapless spectral measurement of light interacting with the matter from the visible to the shortwave infrared portion of the electro-magnetic spectrum [1].
Quantitative variables derived from the observed spectra, e.g., directly through distinct absorption features, are diagnostic for a range of new and improved Copernicus services with a focus on the management of natural resources. These services support the monitoring, implementation and improvement of a range of related policies and decisions.
The observational requirements of CHIME (*Copernicus Hyperspectral Imaging Mission for the Environment*) are driven by its primary application domains i.e., agriculture, soils, food security and raw materials, and are based on experience, state-of-the-art technology and results of previous

hyperspectral airborne and experimental spaceborne systems. The main Mission and System parameters are summarized here:

- On-ground swath ~130 Km (at equator)
- Spatial Sampling Distance (SSD) < 30 m
- > 200 spectral channels within 400-2500 nm
- Spectral Sampling Interval < 10 nm
- Instrument data throughput > 5 Gbps
- Transmission on a Ka-Band single polarization: up to 3.7 Gbps
- Revisit time (2 satellites): ~ 11 days

Due to the required high spatial and spectral resolution, CHIME will produce a near-continuous stream of high-dimensional data resulting in a very high data throughput (> 5 Gbps) and in unprecedented data volumes (ca. 106 Tbits/day of uncompressed data).
The application on board of Artificial Intelligence (AI) techniques holds special promise for hyper-spectral missions such as CHIME, as processing the data on board can reduce the time and cost of data download to ground stations, enabling to send to ground only the most valuable science data, and focusing on rapid responses to events and detected phenomena that impact our society and require fast decision-making.
AI is a bigger concept to create intelligent machines that can simulate human thinking capability and behavior, whereas machine learning is an application or subset of AI that allows machines to learn from data without being programmed explicitly. Machine Learning (ML) enables a computer system to make predictions or take some decisions using historical data (training data set) without being explicitly programmed. Machine learning uses a massive amount of structured and semi-structured data so that a machine learning model can generate accurate result or give predictions based on that data. Typical ML models include Deep Neural Networks (DNN), Support Vector Machines (SVM), Bayesian Networks, Random Forest, and many more.
Whilst the application of AI techniques on ground is a very established practice, the application on board is quite an uncharted territory, that is object of several studies and activities, since it implies to overcome several technological and operational challenges. At the time of this writing, CHIME is foreseen to be the first ESA (European Space Agency) Operational Mission to infer AI on board. The CHIME implementation lays on the solid foundation set by two experimental ESA CubeSats, that were used as flying laboratories for AI: Φ-Sat-1 [2] and OPS-SAT [3].
Clouds are the most promising target for on-board screening since they are a common yet unpredictable contaminant that prevents direct observations of surface features. Previous studies indicate that clouds account for over half of the annual sky cover globally. Thus, onboard cloud screening could approximately double the science productivity per downlink without changing the total stored or transmitted data volumes [4]. Here, we will show how autonomous cloud screening have been efficiently implemented on board using ML techniques.
In the second part of this paper, we will report the feasibility study that has been performed, in order to further enhance the AI capabilities of CHIME. The objective is to identify new potential use cases suitable for AI processing on board (e.g., rapid response and environmental awareness, for fast decision-making). The paper will present the technological and operational challenges that the use of Artificial Intelligence on board implies for an Operational Mission like CHIME, where high reliability and high availability are requested.

2 ON-BOARD SATELLITE CLOUD DETECTION

Continuous monitoring with hyperspectral sensors typically implies gigabits per second data rates and, in case of uncompressed data, data volumes in an order of magnitude of Terabits per orbit are expected, as it is the case for CHIME. Therefore, for both on-board memory management and compatibility of current downlink solutions, on board optimized data compression is expected. Considering the significant presence of clouds hiding the ground, in particular for missions with continuous Earth acquisitions, selective compression brings a major benefit when applied to the clouds. The principle of selective compression is to adapt the compression to different classes. In the case of cloud compression, the objective is to allow a higher loss to the clouds compared to the clear (or ground) pixels, for an improved data rate reduction.

For dedicated compression on clouds, as illustrated on Fig.1, the first step of the compression chain is the cloud detection. Cloud map issued from cloud detection process classifies hyperspectral pixels as cloudy or not. This map is then used as input to the compressor. Three different schemes based on the new multispectral and hyperspectral image compression CCSDS standard [6] have been studied. The first one applies a pre-quantization to the cloud pixels before getting into the compressor. The two other approaches act inside the CCSDS compression: one by an adaptation of the prediction stage according to the class of pixels, the latter one by directly operating on the output of the prediction [5]. Here we will focus on the cloud detection implementation on-board.

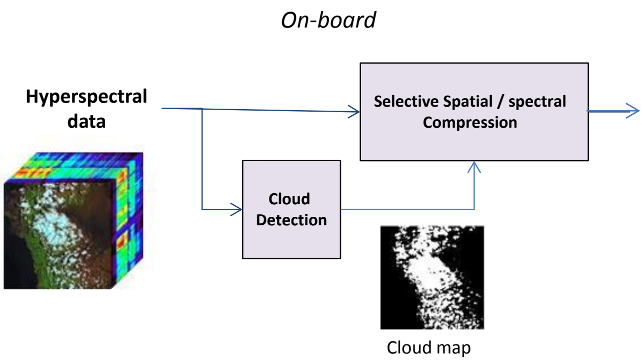

Figure 1 - Generic cloud compression scheme

Cloud detection is widely used on ground for cloud classification: several methods are operational, such as physical approaches (or "threshold" approaches) applied to Landsat and to Sentinel-2 [7]. Artificial Intelligence (AI) techniques give very good perspectives for on-board cloud detection, in particular those based on Deep Neural Network (DNN). However, DNN generally requires higher memory and computational resources, compared to threshold approaches, making the implementation on board difficult, due to the limited availability of processing resources. Being CHIME an Operational mission, high reliability and high availability are requested, making impossible the use on board of COTS AI accelerators.

However, other ML models can be a good compromise between required detection accuracy and computational complexity. Support Vector Machine (SVM) model has been selected, which simply use local information (TOA reflectance values) at pixel level from different spectral bands, thus requiring limited resources for on-board implementation. Here in this paper, we will compare two methods (Threshold and Support Vector Machine), targeting on-board implementation.

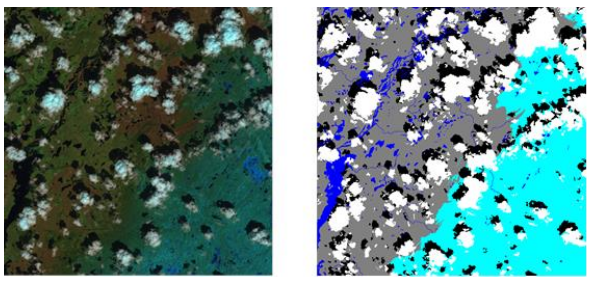

Figure 2 - Example of Landsat classification with cloud mask (right) identified in white (Courtesy of USGS)

For on-board compression purposes, no cloud-type classification is required; the objective is only to detect opaque clouds hiding the ground. Indeed, the translucent clouds (e.g., cirrus clouds) still contain ground information, and thus are considered as "ground" pixels. The Threshold and the SVM approaches have been defined according to this need.

The Threshold method can be seen as a "physical" approach as it allows to discriminate the clouds from the other on-ground features by the means of threshold tests on dedicated spectral bands, and on specific indexes to discriminate high reflective surface from clouds (e.g., snow). The SVM approach allows to separate the pixels into two classes (ground or cloud) in an N-dimensional space. This approach is part of Machine Learning techniques, and needs a training data set in order to find the optimal hyperplane between the two classes. For the on-board implementation, the dimension N has been limited to the useful bands and indexes from the physical approach. The

training stage is performed on-ground and thus has no impact on the on-board processing, except the capability to upload the SVM parameters if needed.
In order to be free of solar illumination, a radiometric conversion in Top-Of-Atmosphere reflectance is performed only on the bands selected for cloud detection, so with a limited impact on the on-board complexity.
The performance of the cloud detection has been assessed on Landsat cloud data base [8]. It includes 80 scenes from Landsat 8 with clouds over different types of landscapes (land, sea, desert, snow, etc.), and few scenes without clouds. A segmentation is associated to each image, with in particular a cloud mask (Fig. 2). The cloud mask includes opaque clouds but also translucent clouds. Such clouds are not frequent in the database and therefore have a limited influence on the performance assessment and on the objective being to not detect ground pixel as cloud. Half of scenes have been considered for the training and the second half for the tests. Some examples of cloud detection results are given in Fig. 3. Visually, the cloud detection is good for most of the test images, with no remarkable differences between the two approaches.

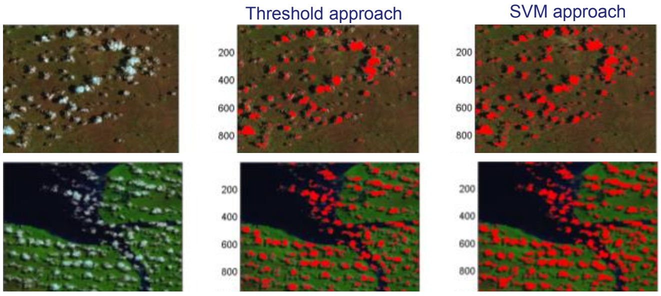

Figure 3 - Example of cloud detection results - clouds are in red (left original– middle threshold– right SVM)

However, few critical cases occur that significantly split from the rest of the results. One case with snow, certainly old or melted snow, and one case on a salted area with the threshold approach. The SVM fails once on a desert scene certainly due to the fact that such landscape was not part of the limited training set for the tests. Ideally, in an operational environment, the training data base shall contain several hundreds of scenes for all types of clouds and landscapes, in order to limit the outliers. For those critical cases, the false positive error rate (defined as the proportion of ground pixel incorrectly detected as cloud) is higher than 10%. For the other test scenes, the false positive error rate is low, leading, for all the test images an average of 0.9% with the physical approach and 0.6% with the SVM.
In order to reduce the false positive detection, a spatial filter has been added. It is applied on the raw binary cloud map to eliminate isolated detections or detections at the border of clouds (no filing). An example of the benefit for a filter radius of 3 is given in Fig. 4.

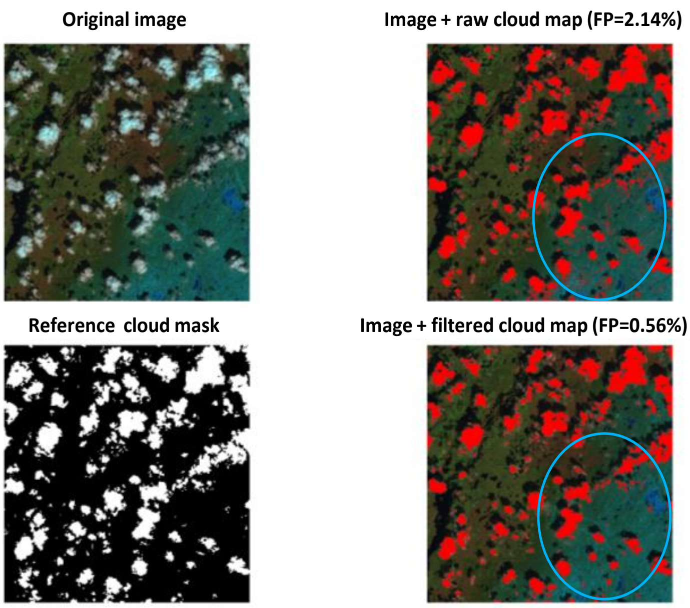

Figure 4 - Benefit of the simple cloud map filtering: the false positive decreases from 2.14 % to 0.56%

Retained cloud detection approach is built around the SVM approach followed by the simple spatial filtering.
The SVM approach has been selected, not only for its advantage on the global false positive error rate, but also for its performance already proven on multispectral ground segments and for its high adaptability to evolutions (e.g., additional hyperspectral band(s) for cloud detection improvement). The SVM is pixel-based and is defined with appropriate bands and indexes, the filter requires only few lines, making the cloud detection implementable on board [9]. The SVM parameters are expected to be stable, and are determined on-ground, thanks to a training stage. The output of cloud detection is a spatial binary map

ready for selective cloud compression, which identifies each hyperspectral pixel as “cloud” or “not cloud” (i.e., ground).

3 ON-BOARD SATELLITE AI ENHANCEMENT

Here we report the preliminary results of a feasibility study, whose ambitious goal is to enhance the AI capabilities of CHIME. The objective is to identify new potential use cases suitable for AI processing on board (e.g., environmental awareness and detection of phenomena that require rapid response, for fast decision-making). The implementation of the AI enhancement on board may require the use of specialized hardware, that at the moment is not part of the CHIME baseline.
The list of possible on-board applications has been selected with the aim of exploiting at most the capability of the CHIME sensor, e.g., high spatial and spectral resolution. Two main criteria have been set in order to ensure that the selected application is useful for the users’ community:

- Early detection: detection on board of a phenomenon (i.e., forest fire) and the immediate transmission to ground of the characteristics of the observed phenomenon. In this case, the impact on satellite operation and system is considerable, since an appropriate data chain needs to be established (e.g., in order to overcome the limited in time visibility of ground stations, a Laser Communication Terminal is needed to transmit data to a geo-synchronous satellite optical receiver).
- Mission extension: use of on-board AI for processing data acquired in areas that are not included in the mission profile, e.g., open ocean, polar regions. In the latter case, the impact on the satellite operation and system is limited, since only the observables are downloaded to ground in case of an event detection, with minimal impact on the data chain.

Table 1 shows the list of possible applications that has been taken into account for the preliminary study.

	Domain	Application	Task
IM 1	Industrial Monitoring	Dust event	AOD (Aerosol Optical Depth) Regression map
IM 2	Industrial Monitoring	Mine tailing	Segmentation map of distinct secondary iron mineral (Hematite, Jarosite, Goethite) or segmentation map of presence of secondary iron mineral in general
IM 3	Industrial Monitoring	Hazardous chemical compounds	Segmentation of high concentration of Cu, Pb, and As in the ground
ME 1	Maritime environment	Coastal / inland water pollution	Chlorophyll Regression map
ME 2	Maritime environment	Oil Spill	Segmentation of oil spill
ME 3	Maritime environment	Plastic in Ocean	Segmentation of presence of macro plastic in ocean
AGE	Atmospheric and gaseous emission	Methane leak detection	Segmentation of methane plume (classification of presence of a methane plume)

Table 1 - List of Applications for AI processing on-board

The applications are separated in three main domains: Industrial Monitoring, Maritime environment and Atmospheric and Gaseous emission. For each domain, the most interesting and promising applications have been pre-selected:

- Industrial monitoring: Dust events. Fast response to environmental events.
- Industrial monitoring: Mine tailing. Hyperspectral data may be useful in monitoring low-concentration contaminants. Fast response to increased pollution.
- Industrial monitoring: Hazardous chemical compounds. Mapping of high concentration of Cu, Pb, and As in the ground.
- Maritime environment: Coastal/Inland water pollution. Fast response to environmental changes (warning systems). Monitoring and prediction of environmental changes, e.g., algae grow.
- Maritime environment: Oil spill. Fast response to environmental events. Quantification of oil spills.
- Maritime environment: Plastic in ocean. Fast response to environmental events. Monitoring the trajectory of plastic islands (temporal). Classification of plastic islands.
- Atmosphere and gaseous emission: Methane leak detection. Fast localization of methane leaks (hence fast reduction of emission).

Nevertheless, the implementation of AI on board will present additional technological and operational challenges that need to be taken into account. In order to take these challenges into account, a list of criteria has been set with the aim of retaining only one application for further investigations:

I. Availability of training data and ground truth data. The availability of training data is key for the implementation of data processing based on AI. The data (airborne or spaceborne) shall be as much as possible representative of the sensor, in this case CHIME.
II. Ease of annotation of training data. In case the training data are not labelled, this needs to be done manually. In some cases, this task can be really time consuming, hindering the usefulness of the data.
III. Training data re-sampling in order to obtain CHIME-like acquisitions. Very often the training data need to be re-sampled in order to match the spatial and spectral sampling of the sensor.
IV. Algorithm Robustness: it implies here the estimation of behavior of an algorithm that is trained on data coming from ground-processed data sources (e.g., Level-1/2 data) and that shall nevertheless operate on-board on instrument acquisitions that will not have the same data quality (i.e., raw data uncalibrated and uncorrected spectrally / spatially / radiometrically). Ground corrections that are mainly affecting the algorithm robustness are atmospheric effects and sensor noise compensation.
V. Operational constraints, i.e., revisit time. For some applications, revisit time and other operational constraints could be a key factor (see Section 1).

The retained application will be possibly implemented in a space representative hardware and will be further evaluated in terms of precision and data complexity. At the time of writing, the selection task is ongoing.

4 CONCLUSIONS

In this paper we reported the technological and operational challenges that the CHIME Project is facing for making use of AI-powered data management on board.

Intelligence and autonomy on board is provided by means of Machine Learning-based cloud detection and selective compression in order to reduce the unprecedented data volume and to maximize the science return. The on-board inference of a DNN poses several risks and challenges, due to the amount of memory and computational power required. Moreover, since the CHIME data chain requires real-time processing, also the inference time is a key driver. The adoption of a DNN may then require the use of specialized HW, that is not present on CHIME.

A good compromise is to use a different ML model, Support Vector Machine, that provides a good compromise between precision and computational complexity. We showed that SVM provides slightly better performances in terms of precision compared to a threshold approach, despite the fact that the training data base was quite limited. However, a big advantage of ML is that the model can be greatly improved, also during flight, as soon as more images are (or will be) available.
In the second part of the paper, we reported the results of the preliminary study about on-board AI enhancement. An exhaustive list of possible applications has been compiled, and the selection criteria have been set. One of the bigger challenges for the AI on-board enhancement is the availability of training data sets, and moreover the fact that the AI algorithm on board will process raw uncalibrated data, whilst most of the training data sets available are ground-processed calibrated data.
The retained application will be further evaluated in terms of precision and data complexity. It is likely that the selected application will require additional specialized HW to run the algorithm. In this case several options are present nowadays on the market, either COTS AI accelerators or radiation tolerant FPGAs with AI inference capabilities. The conclusion of the feasibility study will be presented in a following paper.

5 ACKNOWLEDGEMENTS

This work, funded by the European Space Agency, has been performed in the frame of the CHIME Project Phase B2/C. The CHIME Project Team has been in charge of the management of the work, with the support of the ESA CHIME AI Working Group. Thales Alenia Space Toulouse (F) has been leading the industrial side, overseeing the system related aspects. KP Labs has been responsible for the AI algorithms and use cases selection, whilst KILI Technologies has been in charge of training data sets selection and annotation.

6 REFERENCES

[1] M. Rast, J. Nieke, J. Adams, C. Isola, F. Gascon, *Copernicus Hyperspectral Imaging Mission for the Environment (CHIME)*, IGARSS 2021.

[2] Gianluca Giuffrida and others, *The Φ-Sat-1 Mission: The First On-Board Deep Neural Network Demonstrator for Satellite Earth Observation*, IEEE Transactions on Remote Sensing, Volume 60.

[3] Frederic Feresin, Michael Benguigui, Yves Bobichon, Edgar Lemaire, Matthieu Moretti, and Gaetan Bahl, *On-board Image Processing Using AI to Reduce Data Transmission: Example of OPS-SAT Cloud Segmentation*, OBPDC 2020

[4] Macey W. Sandford, David R. Thompson, Robert O. Green, Brian H. Kahn, Raffaele Vitulli, Steve Chien, Amruta Yelamanchili, and Winston Olson-Duval, *Global cloud property models for real-time triage on board visible–shortwave infrared spectrometers*, Atmospheric Measurement Technique, 13, 7047–7057, 2020

[5] Dimitri Lebedeff, Michel François Foulon, Roberto Camarero, Raffaele Vitulli, Yves Bobichon, *On-Board Cloud Detection and Selective Spatial/Spectral Compression Based on CCSDS 123.0-B-2 for Hyperspectral Missions*, OBPDC 2020

[6] *CCSDS 123.0-B-2 – Low-complexity lossless and near-lossless multispectral and hyperspectral image compression - Blue Book*, https://public.ccsds.org/Pubs/123x0b2c1.pdf

[7] *Sentinel-2 MSI – Level 2A Processing overview*, https://sentinel.esa.int/web/sentinel/technical- guides/sentinel-2-msi/level-2a/algorithm

[8] *Landsat 8 SPARCS cloud validation masks*, USGS, https://landsat.usgs.gov/cloud-validation/sparcs/ l8cloudmasks.zip

[9] Y. Barrios, P. Rodríguez, A. Sánchez, M.I. González, L. Berrojo and R. Sarmiento, *Implementation of cloud detection and processing algorithms for CCSDS compliant hyper-spectral image compression on CHIME mission*, OBPDC 2020