A MODULAR TOOLBOX FOR ON-BOARD DATA PROCESSING AND INFORMATION EXTRACTION

Murray Ireland[(1)], Lucy Donnell[(1)], Craig Hay[(1)], Charlotte Crawshaw[(1)], Steve Greenland[(1)], Hazel Jeffrey[(1)], Richard Hawkins[(2)], Chiara Picardi[(2)], Stuart MacCallum[(3)], Mark Howie[(3)]

[(1)] Craft Prospect, Suite 12, 1048 Govan Road, Glasgow G51 4XS, murray@craftprospect.com
[(2)] Assuring Autonomy International Programme, University of York, York YO10 5DD
[(3)] GSI, Suite 2, Orchard Brae House, 30 Queensferry Road, Edinburgh, EH4 2HS

ABSTRACT

The growing value of Earth Observation (EO) data has seen a drive to improve the capability and performance of space-borne instruments, yielding improvements in spatial, spectral, radiometric and temporal resolution and range of EO instrumentation. This has led to a corresponding increase in the volumes of data captured, resulting in a “data bottleneck” on-board EO satellites, where valuable and often vital data are unable to be downlinked. This bottleneck can be mitigated or avoided entirely by intelligent on-board processing and management of payload data, where useful data products are created, processed, and prioritised to ensure critical information can be downlinked quickly and in more easily ingestible and actionable formats.

The Astral Intelligence Toolbox is a framework and a suite of components which, among other capabilities, enables rapid configuration and deployment of intelligent processing chains for EO missions. This paper presents an overview of the Toolbox and its application to a wildfire response mission case study. Mission safety requirements are defined which flow down to the machine learning and other processing components of the Toolbox deployment, providing an assured autonomous data processing chain with clear benefits to the timeliness and quality of emergency response data.

1 INTRODUCTION

The growing value of Earth observation (EO) data for innumerable applications has seen a drive to improve the capability and performance of the space-borne instruments which capture this data. This drive has yielded improvements in the spatial, spectral, radiometric, and temporal resolution and range of EO instrumentation. In tandem, the miniaturisation of instrumentation has seen a surge in EO-focussed small and nanosatellite missions. These twin advancements have led to a boon in satellite remote sensing data and the information contained within, but they have also introduced a number of challenges. The increase in data volume makes the occurrence of “data bottlenecks” at points in the EO data pipeline more likely, where data volume exceeds the storage or transmission capacity. Combined with the smaller antennae and common lack of dedicated ground segment infrastructure for small satellites, the satellite’s data downlink becomes the key bottleneck, with vital data held on-board and unable to reach its end users. Upon reaching the ground, a satellite’s data contributes to still larger volumes, hosted by big data platforms and used to service a great many applications. For many of these applications, only a fraction of this data is useful and valuable time and processing power must be spent filtering out the high-value from the low-value, effectively searching for a potentially critical needle in a very large haystack.

This paper presents a modular on-board data autonomy framework designed to solve these challenges, the Astral Intelligence Toolbox (AITB). Through the use of machine learning (ML) algorithms, conventional data processing algorithms and an assured, top-down systems engineering approach, on-board data pipelines can be created which circumvent bottlenecks, enable better data management and optimise data products for ingestion by data platforms on the ground.

2 THE NEED FOR ON-BOARD AUTONOMY

The need for on-board autonomy in satellite missions is now well established [1] and is driven primarily by two factors: the challenges of handling and downlinking large volumes of data; and a desire to make satellite's ever-more responsive, reducing the role and associated operational latencies of the human operator in the loop. A summary of these challenges is provided in Table 1. The solutions to overcoming these challenges are conceptually simple: move decision making upstream, closing the feedback loop on-board and allowing optimisation of data management, mission operations and service provision.

Table 1. Challenges in EO data pipelines and benefits of on-board autonomy.

Challenge	**Benefit of on-board autonomy**
Large data volumes leading to bottleneck in downlink	Data sizes can be reduced to alleviate the bottleneck and high-value data prioritised to minimise downlink latency
Large data volumes obfuscating high value data	Data can be tagged or labelled with rich content such as features, value and status/content changes to enable faster lookup on the ground
Raw data must be pre-processed on ground before dissemination to end users, increasing latency	Performing pre-processing on-board in advance of downlink eliminates the associated latency contribution in the ground/service segment
Loss of valuable acquisition opportunities due to space-ground operational latency	Information extracted on-board can facilitate responsive scheduling and planning, such as queuing up a target revisit on a following pass or tasking a trailing satellite
Reduced data quality due to sensor degradation over time	Anomaly detection can identify defects in data from nominal conditions, while autonomous calibration, validation and adaptive optics can ensure data quality is maintained

Overcoming these challenges has a number of benefits to mission end users and other stakeholders. Emergency services can receive vital alerts created on-board, augmenting or skipping entirely data platforms such as Copernicus [2] and FIRMS [3] and ensuring a faster response time. Hyperspectral mission operators can prioritise data products based on the value of their contents, ensuring that the most vital information gets through the downlink bottleneck. Defence organisations can automatically track infrastructure, vehicles and other ground features pass-to-pass, allowing changes to be monitored and key events to be flagged. Data platform providers can receive data products in a ready-to-ingest format, reducing their costs and lead times. All of these capabilities are made possible by the use of on-board artificial intelligence (AI) to replace (whether partially or in full) ground operators and data analysts, ultimately shifting the feedback loop upstream and achieving a higher operational frequency.

Realising this “on-board operator/data scientist” presents its own challenges, both on an implementation level and in terms of assuring the resulting autonomous system. Some of these challenges are detailed in Table 2.

Table 2. Challenges in use of on-board autonomy for EO missions.

Challenge	**Solution**
Adapting to different mission configurations, e.g. platform, payload, customer, application	Modular, component-based framework
Trust that outputs and decisions are accurate and truthful, will not harm life or mission assets	Assurance during development time and in real-time (e.g. autonomy supervisor)
Minimising loss of data to meet bottleneck limitations	Combination of data reduction (lossless where possible) and management (e.g. ordering, tagging)
Fault-tolerance of AI, especially in mission-critical applications	Hardware and/or software redundancy, ML-specific FDIR
Balancing latency, accuracy and power requirements	Optimise models for embedded processing using available tools
Risk of data loss when autonomously processing and prioritising data	Focus on lossless techniques, e.g. data prioritisation, compression. Minimise loss risks through software assurance

3 PHILOSOPHY & REQUIREMENTS

The Astral Intelligence Toolbox (AITB), currently undergoing productization via ESA InCubed funding [4], is intended to overcome these challenges to trusted on-board processing through the creation of modular, assured software components (both traditional and machine learning) and a framework to support cohesive processing chains. The driving requirements for this work are:

- Rapid insights – leveraging the benefits on low-power, embedded machine learning (ML) to generate information and insights which would otherwise require human input
- Modular approach – enabling components to be individually configuration controlled and assured, and upgraded as technologies and capabilities improve, both during development and in-orbit.
- Easily configurable to different use cases – components can be easily tweaked to target different features of interest, utilise different hyperspectral bands and create new data products.
- Platform agnostic – running on frameworks such as SpaceCloud [5], SSTL’s Flexible & Intelligence Payload Chain (FIPC) [4], Linux and bare metal.
- Easily assured – components are assured at the unit level using safety and dependability requirements derived from system- and mission-level, enabled through rigorous assurance processes and model-based systems engineering (MBSE).
- Mission tailored – the modular approach allows bespoke processing chains to be quickly configured and, most critically, optimised for different missions and applications.
- Focus on outcomes over algorithms – while the use of AI to improve EO and other missions is well-known, complex algorithms and architectures are not required to deliver these improvements. Instead the focus is on the integrated whole, with individual components operating together to meet the needs of mission stakeholders.

- Compatible with the CCSDS Mission Operations Services Concept [6].

4 A MODULAR ARCHITECTURE

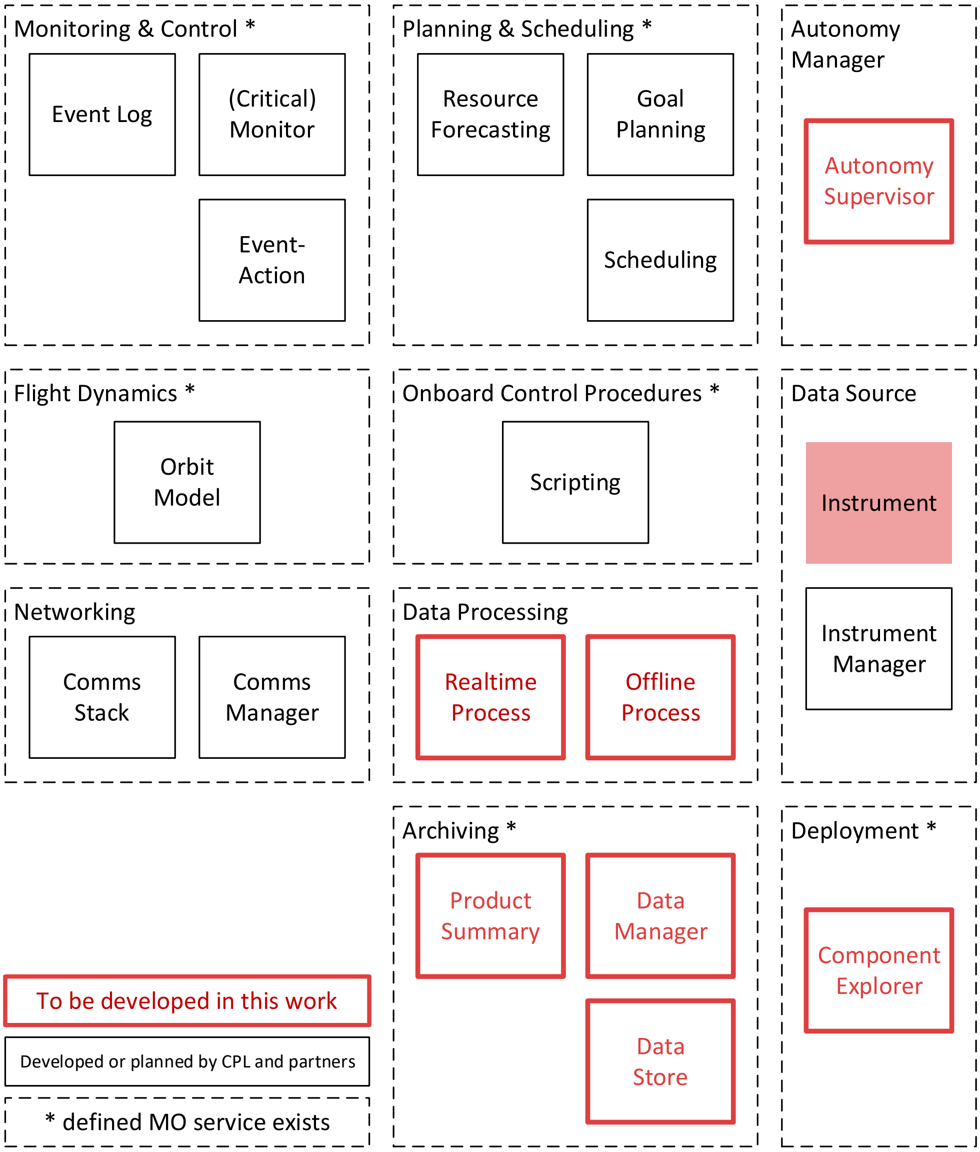


Figure 1. Autonomy framework architecture, defined with respect to CCSDS Mission Operations Services Concept [6].

The architectural framework for the AITB under the FIPC InCubed activity is derived from earlier work into on-board autonomy by Craft Prospect and Bright Ascension [7]. This is shown in Figure 1. The AITB roadmap includes mission-critical autonomy and as such the framework facilitates this through the provision of components for comms handling, mission planning and scheduling and instrument control. Under InCubed and the case study presented here, the AITB development and demonstration is focussed on non-mission-critical data processing activities. This work includes development of the following subsystems in the architectural framework:

- Realtime/offline processing – including a subset of components which perform specific processing and information extraction tasks

- Product summary, data manager and data store – handling read/write of payload data, data products and ancillary data via both APIs (for abstraction layers like FIPC) or direct access to flash memory
- Component explorer – contains autogenerated component documentation and facilitates live monitoring of components
- Autonomy supervisor – runtime redundant or high-level models to verify information extraction results and catch anomalous outputs

A lower-level view of these subsystems, split into different on-board data processing (OBDP), information extraction and reduction tasks, is illustrated in Figure 2. The low-level components in this framework are then deployed as required to meet a use case within the AITB framework, resulting in a cohesive processing chain. The use cases in the FIPC InCubed activity are focussed on the creation of useful data products and the optimisation of these products for downlink, via intelligent tagging, lossless reduction, and queue prioritisation.

Use cases being explored in FIPC include disaster response and alerting, hyperspectral data reduction and agricultural monitoring. To validate the AITB concept, a wildfire detection and alerting application has been configured, building upon work in a previous activity, ACTIONS [8]. Results from this work are reported in the next section.

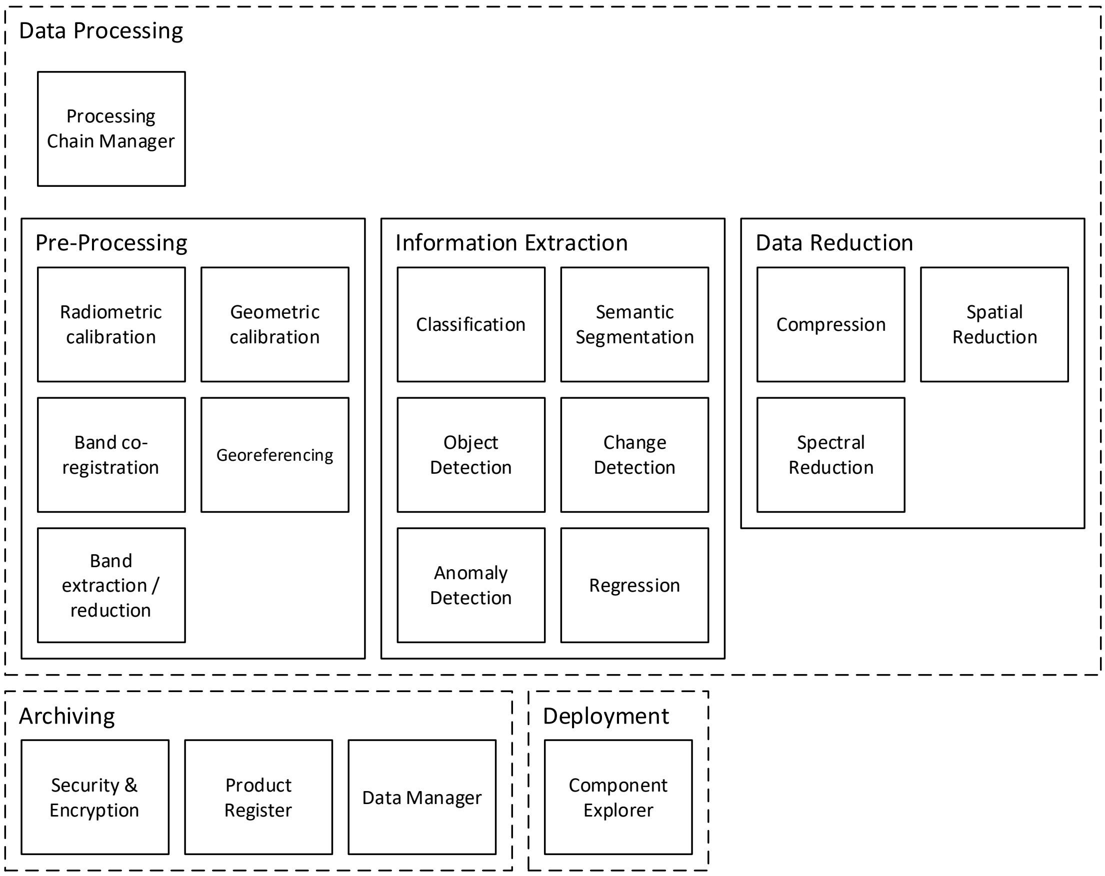


Figure 2. Lower-level framework architecture for components related to on-board data processing and information extraction.

5 CASE STUDY: WILDFIRE EMERGENCY RESPONSE

The ACTIONS case study focusses on an unfortunately common phenomenon in recent years: large, ongoing wildfires which cause loss of life and immense damage to environment and infrastructure. For the case study, a specific mission was defined in which a satellite constellation hosting multispectral instruments delivers both near-real-time alerts to emergency services and pre-processed data products to a commercial downstream data platform. The flow of data, from acquisition to end users, is illustrated in Figure 3.

For this mission, OBDP is vital. Without OBDP, raw data must be delivered to a data platform, processed, and then inspected in large quantities for the presence of active wildfires or burnt areas. Only once image tiles containing wildfires have been identified and derivative outputs such as databases have been updated can emergency services access this vital information. Similarly, commercial end users such as landowners must wait for data platforms to acquire suitable EO data and generate relevant data products such as burnt area severity maps. This can sometimes take weeks, at which point the mitigation of fire damage through reseeding and other activities may not be possible.

Additionally, as the impact of missing active wildfires or reporting false occurrences can be loss of human life, safety assurance of these autonomously generated products and alerts is paramount.

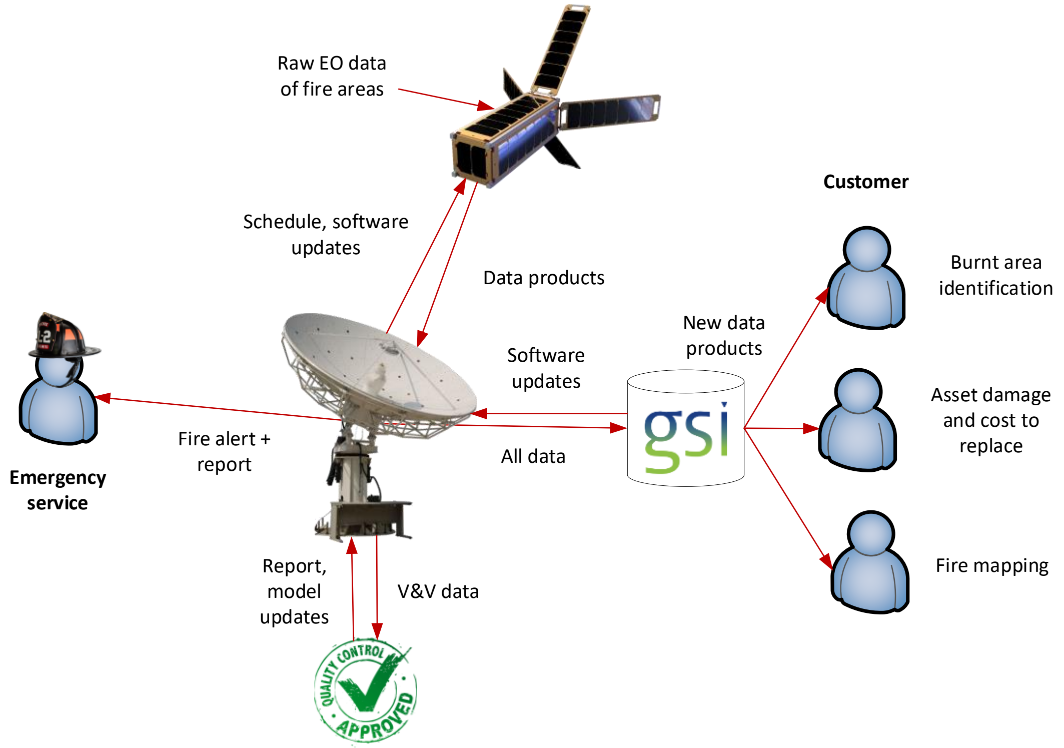


Figure 3. Data flow overview for ACTIONS demonstration mission.

In this section, the concept of operations for the demonstration mission is defined. Safety requirements are then defined such that loss of human life is minimised with respect to existing solutions. The payload data processing chain is then illustrated, configured from components in the AITB. Finally, a description of the implemented processing chain is provided, and test results presented. The case study here is focussed specifically on the emergency response use case only.

5.1 Mission Operations Concept

The demonstration mission has the following elements:

- A CubeSat constellation in sun-synchronous low-Earth orbit
- Each CubeSat is equipped with a multispectral instrument (MSI) with spectral and spatial properties similar to the Landsat-8 Operational Land Imager – this allows Landsat-8 training datasets to be used for ML model training
- The region of interest is limited to Oregon, USA
- The MSI generates a frame every 5 seconds, ensuring contiguous captures which yield an unbroken swath

Oregon was selected due to the large number of wildfires that have occurred there in recent years. In 2020 alone, 400,000 hectares of land were burned, thousands of homes destroyed and 11 lives lost [9].

5.2 Mission Requirements

The following safety requirements were defined for the mission. These were developed with the goal of exceeding the capabilities of existing solutions. In this case NASA FIRMS [3] was used as a baseline.

Table 3. Safety requirements and rational for wildfire emergency alert system.

Requirement	Rationale
The Emergency Response Service shall determine the location of a visible active fire within 200 m of its true location.	NASA FIRMS currently has an accuracy of ~200 m. This sufficient to give emergency services a smaller area to investigate within. The response can then be augmented with ground units or aircraft to localise the fire more accurately.
The Emergency Response Service shall inform emergency services of a visible active fire with 3 hours of it starting	NASA FIRMS can provide active fire alerts within 3 hours of in-orbit observation, or 30 minutes in ideal cases. This does not consider when the fire has actually started. FIRMS satellite revisit times are 12 hours and so a 3-hour start-to-response latency is a significant improvement in the majority of cases.
The Emergency Response Service shall positively identify 95% of all visible active fires acquired by the satellite instrument within the area of interest	Failing to notify of a visible active fire can lead to loss of life, assets and infrastructure and so must be minimised. An absolute value is provided to avoid variable reference
The Emergency Response Service shall falsely indicate visible active fires in the area of interest as less than 52 instances per month	Notification of a false active fire can lead to wasted time and resources and so must be minimised. An absolute value is chosen based on FIRMS false positives over 2020.

The machine learning and other processing components in the processing chain are developed such that they enable these requirements to be met on the mission level. This work will be reported in a future publication derived from the ACTIONS project.

5.3 On-Board Data Pipeline

The AITB components are configured and deployed such that they provide a processing chain which can perform the following tasks:

1. Pre-process multispectral data from the payload, including basic georeferencing
2. Extract pixel-level fire masks from specific multispectral bands

3. Create lightweight alert, detailed report, and image thumbnail products to enable an efficient fire emergency response
4. Create ancillary data products for commercial applications
5. Optimise the downlink through data reduction and prioritisation of high-value data (e.g. products containing fires)

This is illustrated on a high level in Figure 4.

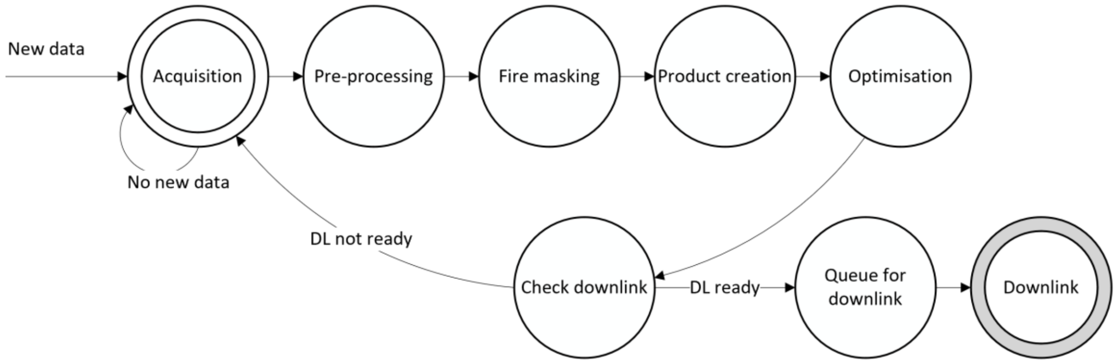


Figure 4. Processing chain flow diagram.

5.4 Implementation & Testing

Testing of the processing chain components focussed primarily on rigorous verification and validation of the machine learning model used to perform the fire masking. This work included performance testing of the model (including accuracy and latency) and will be reported in a future paper.

The test campaign for the processing chain culminated in the deployment of the processing chain on Unibap's iX5-100 development kit and integration of this system into a high-fidelity hardware-in-the-loop (HIL) simulation. The simulation employed Landsat-8 imagery of Oregon from a day in September 2020 on which wildfires were particularly prolific. The simulation test bench then enabled the following tests to be performed:

- Validation of the generated fire products against FIRMS detections from the same time period and location
- Performance benchmarking for processing chain timings, including:
  - Fire mask generation throughput, e.g. frames per second (FPS), queries per second (QPS), pixels per second (PPS)
  - Information latency, i.e. from acquisition of raw data to creation of actionable information[1]
- Quantification of downlink throughput benefits, i.e. percentage increase in valuable information vs baseline

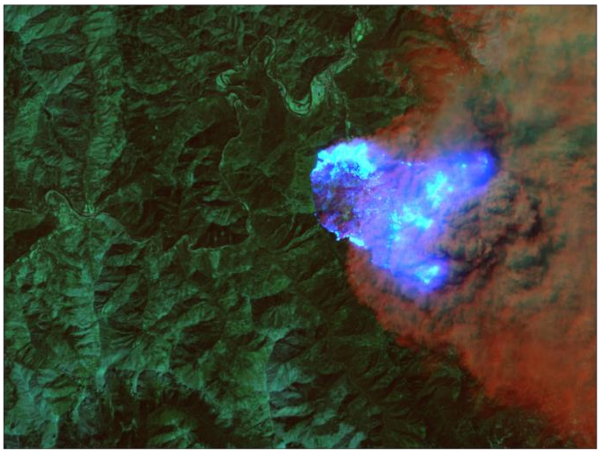
5.5 Results

Some examples of the data products created by the AITB deployment for ACTIONS are shown in Figure 5. These include:

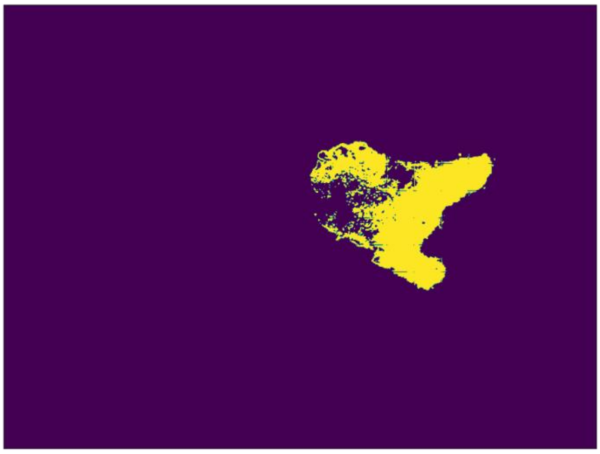
- A lossless, georeferenced Level 1 multispectral image

[1] Full information latency metrics also account for the time taken to deliver the information on the ground

- A Level 3 pixel mask created using the fire masking ML model
- A lightweight Level 4 text alert
- A lossy-compressed Level 4 thumbnail with bounding box to emphasise the fire, used for verifying the alert product



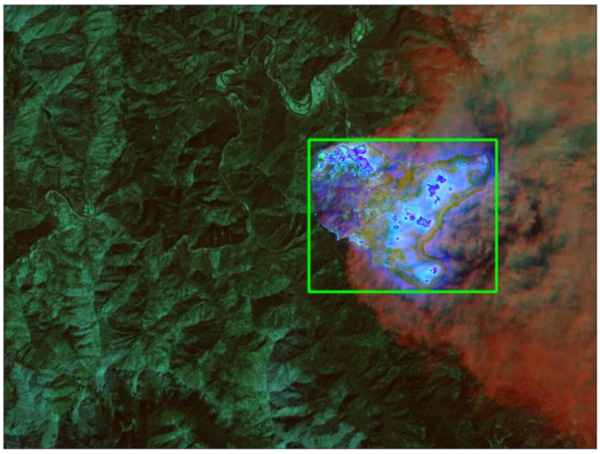
a) Level 1 false colour image.



b) Level 3 fire mask.

```
FIRES DETECTED
LARGE fire:
    Time 20200908-065847.184
    Location (123.4094°S, 41.8675°E)
    Area 40.08 sq km
```

c) Level 4 alert message.



d) Level 4 verification thumbnail.

Figure 5. Example products created from ACTIONS AITB deployment.

Accuracy, throughput and latency metrics for the fire masking component and latency for the full processing chain are reported in Table 4. The fire masking model as-run on labelled test data has a mean intersection over union (MeanIoU) score of 93%. True positives and negatives are >99% with a 6-pixel margin of error, selected to validate the safety requirement for 200 m geolocation accuracy. The model has a throughput of 952,680 pixels per second, which, for the high-resolution multispectral image being processed, returns a full mask processing latency of 3.57s. With the pre-processing and product creation tasks also included, the time taken to generate valuable information from the original payload data is 4.84s. As the MSI equivalent framerate is one every 5 seconds, this is sufficient to ensure real-time processing of payload data.

Table 4. Performance metrics for wildfire alerting processing chain.

Metric[2]	Value
Model accuracy – MeanIoU	93%
Model accuracy – true positive	100%
Model accuracy – true negative	99.2%
Model throughput (FPS)	0.3
Model throughput (QPS)	413.5
Model throughput (PPS)	952,680
Fire masking component latency	3.57 s
Information latency	4.84 s

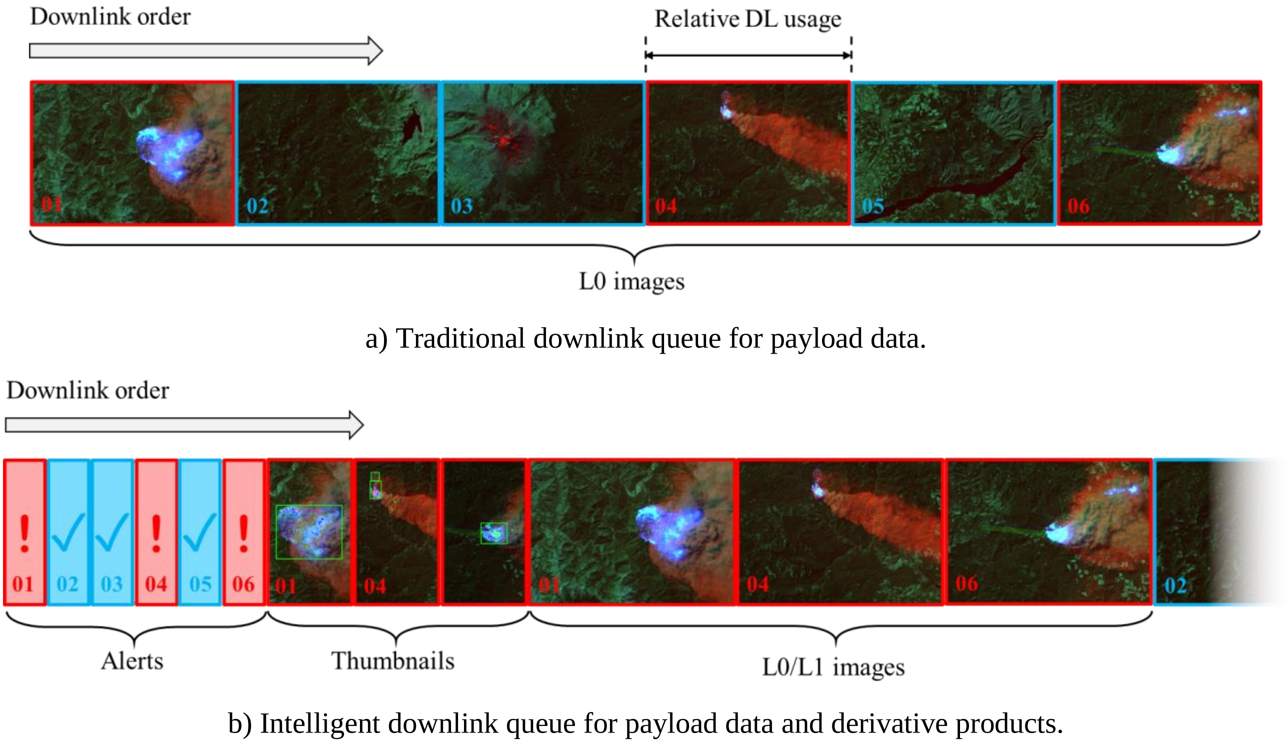


Figure 6. Comparison of traditional downlink queue for payload data with intelligent queuing enabled by Astral Intelligent Toolbox processing activities. Products indicating the presence of wildfires are marked in red, while those without are blue. The identification of wildfires allows the associated data products to be identified and lightweight alert products prioritised to maximise the timeliness of the information.

The benefits to data management brought by OBDP can be visualised in Figure 6. In a traditional downlink queue, payload data products are typically downlinked in order of acquisition only, without further contextual metadata to determine value or priority. In the intelligent downlink queue enabled by the OBDP and information extraction capabilities of the AITB, the lightweight alert products can be downlinked ahead of all other products. This not only ensures that the salient information is prioritised, but also significantly decreases the risk of vital information not successfully downlinking

[2] Note that true positives and true negatives allow for a 6-pixel margin of error to account for 200 m geolocation accuracy

on a given pass. The alerts can then be followed by lossy-compressed thumbnails for visual validation and any remaining capacity used for full L0 or L1 products for verification and validation (V&V) activities. The L3 georeferenced fire masks are not illustrated but can also be downlinked to provide more precise geolocation of the fires. The grouping of products in this way can also facilitate expediated routing of information on the ground, whereby alerts and thumbnails are routed directly to emergency services, while the lossless image products are processed by an automated V&V stage in the ground segment and routed to commercial data platforms for secondary applications. In ACTIONS, secondary applications were enabled by Global Surface Intelligence's (GSI) data platform and included burn severity mapping in timberland, fusing the near real-time data from the ACTIONS space segment with additional ground sources and historical data.

The ordering logic of the downlink queue is easily specified with the richer metadata now available on-board, which includes:

- Acquisition timestamp
- Type of product (e.g. denoted by level)
- Image and mask product geolocations
- Feature coordinates in image and geolocations
- Size and preponderance of fires in a product
- Value or priority number, calculated as a function of acquisition time, fire size, preponderance, severity, etc.

The queue can then be specified by a ground operator or the spacecraft schedule as needed to meet end user requirements.

5.6 Assurance

Safety assurance of the autonomous fire detection mission is provided by assessing compliance with the mission safety requirements defined previously. Compliance is stated and evidence provided in Table 5. With these requirements primarily met, the trustworthiness of the mission is successfully bound, allowing end users such as emergency services to integrate the autonomous fire alerts into their response processes.

Table 5. Compliance matrix for mission safety requirements.

Requirement	Compliant	Evidence
The Emergency Response Service shall determine the location of a visible active fire within 200 m of its true location.	Yes	30 m resolution available in fire mask, geolocation accuracy sub-50 m for test
The Emergency Response Service shall inform emergency services of a visible active fire with 3 hours of it starting	Yes	Requirement met with 188 satellites and single ground station, using intelligent downlink queue. Smaller constellation size possible if aiming to match FIRMS latency only (12 hours).
The Emergency Response Service shall positively identify 95% of all visible active fires acquired by the satellite instrument within the area of interest	Yes	False negatives calculated at 0.76%, yielding 98.24% true positives
The Emergency Response Service shall falsely indicate visible active fires in the area of interest as less than 52 instances per month	Partial	Depending on threshold in validation approach, false positives in simulation tests are either 53 (moderate threshold) or zero (low threshold)

5.7 Conclusions

This case study has provided an overview of applying the AITB to a wildfire response use case. The benefits to data management and information throughput have been shown through compliance with mission safety requirements and illustrated visually. The safety requirements defined here are specifically focussed on *safety of life*, a key consideration in the Assuring Autonomy International Programme's research and Body of Knowledge [10]. This notion of safety can be extended to other mission stakeholder priorities, such as safety of platform, of environment and other spacecraft, of ground assets and infrastructure and of the mission data itself. The approach outlined in this section is then applicable to the majority of autonomous space systems, from simple data processing chains to mission-critical operational autonomy. The process can similarly be applied to non-safety-critical assurance requirements which meet business interests (e.g. latencies, volume, information content, etc.). As criticality increases, the number of safety requirements and the volume of evidence required to assure the system (assurance artefacts, test reports, hazard analysis, etc.) will naturally also increase.

6 FUTURE WORK & FLIGHT HERITAGE

The Astral Intelligence Toolbox is a framework intended to provide rapidly configurable on-board processing and data management solutions while retaining flexibility in integration with hardware and software platforms and integration of new components and functionality within itself. The use of machine learning components is only one part of this framework, and they exist only such that mission and system safety and assurance requirements can be met.

The case study presented here demonstrates the application of the AITB to the wildfire response use case. Other applications that have been developed or are in development include:

- Cloud masking of multispectral and hyperspectral data
- An integrated imager and cloud detector, with versions for both daytime [7] and night-time operations [11]
- Region of interest lossless/lossy compression
- Underpinning the sensing and understanding subsystems of a responsive operations framework for nanosatellite communications

Work on the AITB is continuing under ESA InCubed and ARTES funding and commercial projects. Deployments of the AITB for cloud detection will see flight heritage in 2022 and 2023 on the KAUST-SAT and ROKS missions, respectively. A future in-orbit demonstration is planned with SSTL to gain flight heritage of the FIPC framework.

7 REFERENCES

[1] M. L. Ireland, P. Mendham, S. Greenland, P. Karagiannakis, and F. Hogervorst, 'Enabling and assuring autonomy in small satellite missions', in *Sensors, Systems, and Next-Generation Satellites XXIV*, Online Only, United Kingdom, Sep. 2020, p. 33. doi: 10.1117/12.2574612.

[2] Copernicus, 'Copernicus Emergency Management Service'. https://emergency.copernicus.eu/

[3] NASA, 'FIRMS - Fire Information for Resource Management System'. https://modaps.eosdis.nasa.gov/map/

[4] ESA, ‘Demonstrating an Innovative, Flexible and Intelligent Payload Chain for High Data Throughput on Small EO Satellites’, *InCubed Activity Portfolio*, 2021. https://incubed.phi.esa.int/portfolio/fipc/

[5] Unibap AB, ‘SpaceCloud® Software’, *Unibap*. https://unibap.com/en/our-offer/space/spacecloud-services/

[6] CCSDS, ‘Mission Operations Services Concept’, CCSDS, Washington DC, Informational Report CCSDS 520.0-G-3, Dec. 2010.

[7] S. Greenland, M. Ireland, C. Kobayashi, P. Mendham, M. Post, and D. White, ‘Design & Prototyping of a Miniaturised Forwards Looking Imager using Deep Learning for Responsive Onboard Operations’, p. 9, 2018.

[8] University of York, ‘Autonomous Capabilities and Trusted Intelligent Operations iN Space (ACTIONS)’, *Assuring Autonomy International Programme*, 2021. https://www.york.ac.uk/assuring-autonomy/demonstrators/assuring-autonomy-in-space/

[9] E. Newburger, ‘At least 33 dead as wildfires scorch millions of acres across Western U.S. — “It is apocalyptic”’, *CNBC*, Sep. 12, 2020. https://www.cnbc.com/2020/09/12/fires-in-oregon-california-and-washington-spread-death-toll-rises.html

[10] University of York, ‘AAIP Body of Knowledge’, *University of York*. https://www.york.ac.uk/assuring-autonomy/guidance/body-of-knowledge/

[11] Craft Prospect, ‘Responsive Operations for Key Services’, White Paper, Oct. 2020. Accessed: Apr. 15, 2022. [Online]. Available: https://craftprospect.com/wp-content/uploads/2021/07/ROKS-Whitepaper.pdf