# Free and open-source machine learning workflows for co-creating national-scale classification models through country-driven QField surveys and Digital Earth Pacific

Nicholas Metherall,[1,2,3] Jesse Anderson,[4] Sachindra Singh,[1] Alex Leith,[5] Ahi Saipaia,[2] Lucy Fa'anunu,[6] Claudya Dhaja,[7] Maivunijale Waqa,[1] Elenoa Biukoto,[1,2] Shyam Lodhia,[8] Naomi Jackson,[1] Andiswa Mlisa,[1] Brian Killough,[9] Marco Bernasocchi.[10]

[1] Geoscience Energy Maritime Division (GEM), Pacific Community (SPC) - nicholasm@spc.int; sachindras@spc.int; maivunijalew@spc.int; elenoab@spc.int; naomij@spc.int; andiswam@spc.int
[2] University of the South Pacific (USP) - asaipaia98.13@gmail.com
[3] Fenner School of Environment and Society, Australian National University
[4] D4DInsights - jesse.anderson@gmail.com
[5] Auspatious Pty LTD - alex@auspatious.com
[6] Ministry of Energy, Information, Disaster Management Environment, and Climate Change (MEIDECC) - lucyfaanunu15@gmail.com
[7] Faculty of Veterinary Medicine, University of Nusa Cendana - claudyadhaja@gmail.com
[8] Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH, (GIZ) - shyam.lodhia@giz.de
[9] Killough Services, LTD - brian@killoughservices.com
[10] OPENGIS.ch - marco@opengis.ch

**Keywords:** Free and Open-Source Software (FOSS), QField, Digital Earth Pacific, Land Use Land Cover (LULC), Land Cover

**Abstract**

This paper highlights the Free and Open-Source Software (FOSS) datasets, analytical tools and workflows used for co-creating localised national-scale machine learning classification models in Digital Earth Pacific (DE Pacific). The case study includes the participatory workflows used within the DE Pacific Land Cover Assessment Skills Transfer (LCAST) Workshop in the Kingdom of Tonga in 2023. The FOSS tools used were QGIS, QField, GeoPandas, ODC STAC-, Xarray, Pandas, Scikit-Learn, NumPy and Folium through the DE Pacific Analytical Hub Jupyter environment. These workflows supported participatory processes to gather inputs into the calibration and validation of machine learning workflows as seen in the Land Use Land Cover (LULC) examples of the LCAST workshop. The workshop held over one week in July 2023 included representatives from seven Ministries of Tonga. The methods covered the 'full-cycle' of workflows for generation of LULC models, from field survey data collection to computer labs for data processing and analyses. The results highlight the LULC mapping products, accuracy assessments, the workshop evaluation as well as the broader lessons learned about the intrinsic value of the participatory mapping processes. Since 2023, these workflows have been used in LCAST and similar country-driven workshops with more than 240 participants across 10 Pacific Island Countries and Territories (PICTs) or Large Ocean States (LOS) including Fiji, Republic of the Marshall Islands, Tuvalu, Palau, Cook Islands, Papua New Guinea, Vanuatu, New Caledonia, the Solomon Islands and the Kingdom of Tonga. These FOSS approaches may contribute to the long-term continuity of two-way learning processes and inputs needed for the co-creation, calibration and validation for modelling and mapping as well as the building of local capacity and capabilities in earth observations in the Pacific. The paper outlines these workflows in detail through a visual flowchart and provides access to GitHub for replication of results.

## 1. INTRODUCTION

### 1.1 Local capacity for Earth Observations in Pacific Large Ocean States

During 2020-2021, a needs assessment process was conducted through consultation with Pacific Island Countries and Territories (PICTs) also commonly referred to as Large Ocean States (LOS), including the Republic of the Marshall Islands, Tonga and Vanuatu. The consultations focused on the assessment of needs in relation to Earth observation data to support country-level development and sustainability priorities (SPC, 2021). The assessment was facilitated by the Pacific Community (SPC), the principal scientific and technical organisation supporting development in the Pacific region and D4D Insights. Through this process of consultation, a series of specific use cases or priority needs were identified and ranked respectively as tier I (primary) or tier II (secondary) based on overall importance and urgency for development. Several priorities emerged related to land use/landcover (LULC) data and models. Potential applications of these included but were not limited to forest change, agricultural censuses, and disaster impacts. In addition to these use cases, the LOS highlighted several areas to guide the process of development itself. These included the needs to overcome challenges around limited coverage of satellite data for the region, limited availability due to cloud cover, and timeliness of data availability. They also included needs around regional ownership, data sovereignty, focus on alignment with existing national programs and policies and local capacity development (SPC, 2021).

Notably, a theme around open data, tools, algorithms and workflows emerged as a critical requirement to meet local needs. This was identified as a catalyst to support local capacity development and alignment with local initiatives collectively contributing to longer-term outcomes of regional adoption, uptake and ownership of the earth observation capabilities.

### 1. 2 Digital Earth Pacific

The most salient outcome of the needs assessment consultations was the development of a business case for making free, open and operational satellite data readily available for the Pacific region through a public technology infrastructure called Digital Earth Pacific (SPC, 2021). Similar to the earlier concepts of 'Spatial Data Infrastructure' (SDI), Digital Earth Pacific (DE Pacific) would provide a base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to earth observation data (Nebert, 2004). DE Pacific would provide a basis for spatial data discovery, evaluation, analysis and applications for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general in the Pacific. As such, DE Pacific would form an inter-connected set of systems to enable easy access, management, usage of earth observation datasets.

As a spatial data infrastructure with broader capabilities for analysis and scaling of regional products, Digital Earth Pacific would be based upon similar precedents including Digital Earth Africa and Digital Earth Australia (Liu et al., 2019; Mohamed et al., 2019; Leith et al., 2024). However, DE Pacific would need to be adapted through further consultation and country engagement to be made fit for purpose for the context of the Pacific. A context distinct from the other two continental examples because of the Pacific region's unique geographic, environmental and cultural-political contexts. Indeed, since this initial consultation, DE Pacific with SPC have made numerous country-level and regional-scale engagements in the Pacific. The day-to-day operations of this process have been driven by the Earth and Ocean Observation (EOO) sections of SPC. Both these sections of SPC and DE Pacific have practiced a series of principles around open data. An example of these kinds of principles are outlined in the Open Data for Resilience (Crowley et al., 2014): 1) Open by default, 2) Accessible, licensed and documented, 3) Co-created, 4) Locally owned, 5) Communicated in ways that meet needs of diverse users, 6) Engage user communities, 7) Develop strong institutional partnerships, 8) Prioritise open source, 9) Set clear, long-term goals.

These and other open-data principles have been integrated into several SDIs and broader databases developed by SPC. These have included the Pacific Risk Information System (PacRIS), the Pacific Ridge to Reef Science Platform, The Pacific Oceans Portal, the Nexus Platform to consolidate disaster risk reduction information and the Pacific Data Hub to support the consolidation of many of these datasets within SPC together into a single database or SDI (World Bank & GFDRR, 2013, Pacific Community (SPC), 2018, Pacific Community (SPC), 2025). Many of these Geographic Information System (GIS) geodatabases have been deployed as SDIs through the GeoNode web-based application platform. These platforms have largely been built with PostGIS, Geoserver, Geonetwork to become Open Geospatial Community (OGC) compliant. OGC compliant certification ensures different systems can share and use location data, promote interoperability, reduce risks of being dependent on external commercial agreements and vendors and allows for reliable integration of maps, features and data across various platforms like Web Map Service (WMS) or OGC APIs (Reed, 2011; Voidrot & Percivall, 2020).

Overall, these free and open-source software approaches in the development of these SDIs signalled SPC's commitment towards a 'culture of sharing' (Rajabifard & Williamson, 2001). Because of SPC's mandate to serve PICTs, open-data and this culture of sharing has been demonstrated through a track record of SDI development through past projects. This approach has supported a wider community of both government and non-government data users and providers (Groot, 1997; Vancauwenberghe et al., 2018). From an operational perspective, this was intended to enable users and partner organisations to save time, effort and wider resources when trying to acquire new datasets by avoiding duplication of expenses associated with generation and maintenance of data and their integration with other datasets (Groot, 1997; Rajabifard & Williamson, 2001). In continuing with this historic approach, Digital Earth Pacific has also been developed through an SDI and open data approach in order to reduce duplication and facilitate integration and development of new and innovative business applications, to produce human and resource savings and returns (Rajabifard & Williamson, 2001).

This paper highlights how the open-data principles in DE Pacific have been operationalised through making analysis-ready earth observation data based on open algorithms available to users across the region, as well as analytical capabilities (SPC, 2021). For example, DE Pacific provides access to petabytes of Landsat (30m) and Sentinel (10m) datasets. While such data are already free and open, DE Pacific makes these datasets more accessible to LOS stakeholders through a scalable, cloud-enabled infrastructure and analysis-ready-format (SPC, 2021). The case study for how these workflows were applied is the Digital Earth Pacific (DE Pacific) Land Cover Assessment Skills Transfer (LCAST) Workshop.

### 1.3 DE Pacific Land Cover Assessment Skills Transfer (LCAST) Workshops

The first DE Pacific LCAST Workshop was conducted from the 24th to 28 July and 2023 in Tongatapu. This workshop was part of a two-way learning process with 24 participants representing seven Government Ministries across Tonga (Fa'Anunu, 2023). This included the following Ministries: Agriculture Forestry and Food (MAFF), Fisheries (MoF), Lands and Natural Resources (MLNR), Meteorology, Environment, Information, Disaster Management, Energy, Climate Change, and Communication (MEIDECC), National Emergency Office (NEO), Tourism and Infrastructure. The learning objectives of the workshop included 1) Raising awareness of Digital Earth Pacific (DEP) at a high level across Government and wider stakeholders. 2) Familiarising users with technical workflow: field data collection, annotation, classification, prediction and map generation. The longer term strategic outcomes include 1) building a land cover model and map with moderate accuracy in the short-term and higher accuracy in the longer term. 2) Building a community of practice through building on the existing Tonga GIS user group. The Kingdom of Tonga is composed of 169 islands, only 36 of which are inhabited (Figure 1) (Saulnier, 2022). The islands are arranged into three main groups: Tongatapu (including Eua), Ha'apai, and Vava'u. These islands are aligned in parallel to the Tonga Trench, one of the deepest points on Earth (Raitt et al., 1955). Geologically, the islands are split into two distinct types: volcanic islands (e.g. Hunga Tonga-Hunga Ha'apai) and coral limestone islands (e.g. Tongatapu) (Bryan et al., 1972). Tonga's environment is tropical with high humidity, rainfall, supporting rainfed agriculture and dense coastal and tropical rainforests in Vavau and Eua.

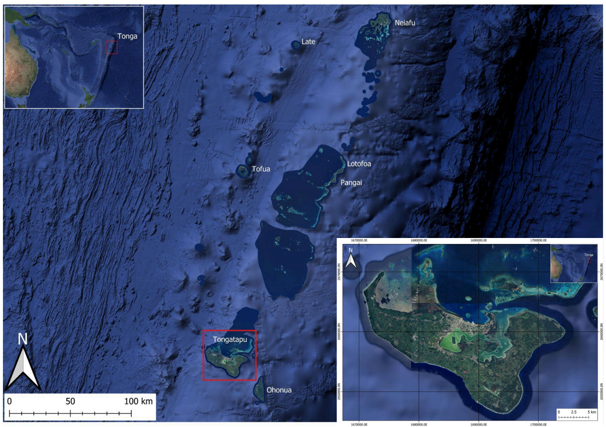


Figure 1. Location of Tongatapu in the Kingdom of Tonga within the Pacific Ocean Basin. LULC modelling in Tongatapu was selected as a priority objective for the first DE Pacific LCAST Workshop through the outcomes of the Needs Assessment and additional requests from the Government of Tonga to SPC. Map made by Lodhia (2025).

### 1.4 Land Use Land Cover

Land Use Land Cover (LULC) models shed light on the proportions and distributions of different natural and man-made environments across landscapes at given points in time. When multiple land cover model maps are generated for different points in time, the results can be analysed to detect changes in different land use and land cover classes over time. This analysis is commonly applied to the monitoring and management of a wide range of sectors including, but not limited to, forestry, agriculture, urban planning, infrastructure, water management and mining (Topuz and Deniz, 2023). Yet, there have been persisting challenges for machine learning approaches to meet thresholds of accuracy while generating LULC classification models at scale. Some of these challenges have included: 1) generating LULC models that provide an accurate prediction of land use and land cover distribution at the local scale (meeting accuracy assessment thresholds). 2) Scaling of LULC models across diverse ecosystems and geographies while continuing to still meet accuracy assessment thresholds. 3) Reconciling between local needs inputs and globally standardised LULC classes and models (for reporting purposes). 4) Collecting adequate training data through field surveys across diverse ecosystems which may require surveys into remote and commonly inaccessible areas.

## 2. CONCEPTS AND RELATED WORK

### 2.1. Wider FOSS4G workflows for LULC modelling

The use of Free and Open Source Software for Geospatial (FOSS4G) workflows has become increasingly prevalent for Land Use and Land Cover (LULC) modeling and mapping, spanning from local projects to national-scale initiatives (Mardani, 2019). This trend is driven by the accessibility, cost-effectiveness, reproducibility, scalability and community-driven development that emerge from FOSS4G ecosystems (Muellerklein, 2017). Several clear examples of these workflows have been demonstrated through the QGIS ecosystem including the QField field data collection mobile application (Maťašovská, 2024; Graser et al., 2025). This has been widely used in research, government, civil society and the private sector for projects with a need to gather geotagged data for LULC modelling (Ostadabbas, 2020; Wu, 2024). Within the Kingdom of Tonga QField has been used by the Ministry of Agriculture, Food and Forest (MAFF) to support the annual agricultural census activities at scale (Duncan et al., 2022; Saipaia, 2023) (figure 2).

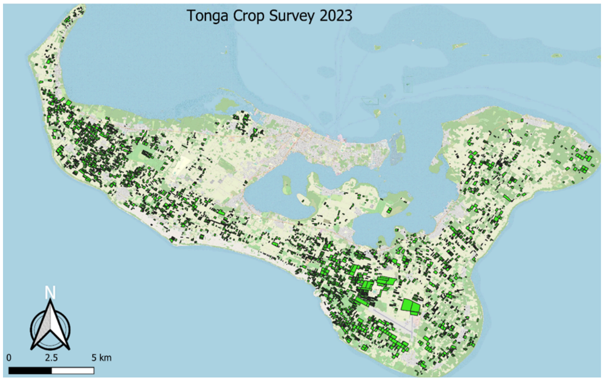


**Figure 2.** Tonga Crop Survey using QField (Saipaia, 2023). These annual crop surveys in Tonga have used QField to monitor crops before the tropical cyclone season since 2021 (Saipaia, 2025). As a result, QField has become an important part of monitoring the agricultural sector which is a foundation of rural livelihoods, contributing approximately 16% to the nation's annual GDP. Such examples highlight how QField has become a cornerstone technology for field data collection in geospatial workflows, particularly within large-scale participatory mapping and earth observation initiatives such as Digital Earth Pacific (DE Pacific). As a FOSS mobile application tightly integrated with QGIS, QField enables rapid in-situ data capture, addressing the practical constraints faced in the field. Its offline-first architecture, broad device compatibility, and low operational overhead make it especially suitable for environments with intermittent connectivity, limited hardware budgets, and diverse institutional capacities.

Other FOSS tools include the Semi-Automatic Classification Plugin (SCP) for QGIS providing tools for both supervised and unsupervised classification of satellite imagery (QGIS Source 3.4) (Congedo, 2016; Cutts & Graser, 2018). For automated and scalable processing, projects rely on powerful libraries and command-line tools. GRASS GIS (Geographic Resources Analysis Support System) and the Geospatial Data Abstraction Library (GDAL/OGR) are essential for complex raster and vector data manipulation, particularly in large-area mapping (Martin et al., 1989; Warmerdam, 2008). PostgreSQL/PostGIS provides a scalable backend for managing large volumes of spatial data inherent in national-scale LULC mapping projects (Momjian, 2001; Obe & Hsu, 2021). Python libraries including NumPy, Pandas and Pangeo/Xarray enable scalable data processing to be made into reproducible automated workflows (Oliphant, 2006; McKinney, 2011).

### 2.2. Cloud computing and earth observations: Google Earth Engine and its rise beyond reach

There has been an increasing focus on enhancing automation, accuracy, and scalability, particularly for LULC mapping and other earth observation applications at a national or broader spatio-temporal scale. To meet these needs automation and cloud computing have come to increasing prominence. Platforms like Google Earth Engine (GEE) have become a powerful example of the use of cloud computing for earth observation satellite data analysis at massive scale (Amani et al., 2020). Initially, in 2010, GEE was launched as a free platform largely focused on scientific research (Gorelick, 2013). It was offered free of charge to scientists, researchers, non-profits and governments to address global issues. For over a decade GEE has operated primarily as a free service for noncommercial use gaining an increasingly large global user-base (Pérez-Cutillas, 2023). The shift towards a formal proprietary and paid model began in 2022 when Google fully integrated GEE into its commercial infrastructure (Wilson et al., 2025). Despite these changes, currently if a party is registered with a university or non-profit organisation it is still possible to apply for free-use of GEE yet storage of assets over certain memory thresholds may still require additional payments. As a result, there are various examples of ongoing research work that also combine this proprietary platform with open-source software and tools. GEE remains a powerful and viable option for scaling LULC and other earth observation analyses. Yet, the interoperability and customisability of GEE remains limited not only by the partial pay-wall requirements of GEE but also because the platform remains a closed proprietary model. There are several differences between this closed proprietary model and other FOSS alternatives including the Open Data Cube including but not limited to centralisation of architecture, flexibility of deployment, data interoperability and cost models.

### 2.3 Open Data Cube and Digital Earth Australia, Digital Earth Africa and Digital Earth Pacific products

In 2010, Geoscience Australia, CSIRO and the National Computational Infrastructure (NCI) launched the Australian Geoscience Data Cube (AGDC), a software package to simplify management and analysis of satellite imagery and other Earth observation data (Melrose, 2018). In 2016, the AGDC project began its transition to an open-source project under the stewardship of the Committee of Earth Observation Satellites (CEOS) (Killough, 2018; Leith et al., 2024). The project was officially renamed the Open Data Cube (ODC) and released under the Apache 2.0 license, making the technology freely available globally. The ODC is built on open standards and open-source software (Python, PostgreSQL/PostGIS, Xarray, Dask), and is designed to be vendor-neutral and highly flexible (Gomes, 2021; Bednard & Durant, 2023). Digital Earth Australia is a continuation of the original AGDC and a global ODC leader (Leith et al., 2024). It has transformed from a research project into an ongoing national capability which has developed numerous national products including but not limited to the Digital Earth Australia national landcover models (Gavin, et al., 2018; Lucas, et al., 2019). Launched in 2019, Digital Earth Africa has become the most prominent international application of ODC and is governed through African ownership (Georgiadou et al., 2011; Killough, 2018; Mubea et al., 2022; Leith, 2024). DE Pacific has followed the example of DE Africa and Australia but must evolve to be suitable to a very different geographic, environmental and cultural-political context.

### 2.5 Aims

Within this context of the literature and wider related work, the aim of this paper is to provide an overview of the workflows used within Digital Earth Pacific in conjunction with QGIS and QField. The case study of the DE Pacific LCAST Workshop is selected to concisely illustrate how these workflows have been used to generate a practical output. This workshop demonstrated a nationally driven survey providing data inputs into machine learning LULC classification models. This participatory workflow may be of interest to other stakeholders in the Pacific and more widely who are interested in replicating this for other cover model calibration and validation. The case study use cases and sectors. In doing so, the paper may shed some light on best practices within the Pacific region for land use and highlights the long-term replicability of these open-source workflows that are not subject to pay walls or commercial vendors. The paper is also intended to raise greater awareness of the current datasets, regional products as well as the methods and workflows used in Digital Earth Pacific.

## 3. METHODS AND WORKFLOWS

### 3.1 Overview of methods used within QGIS, QField and Digital Earth Pacific

An overview of the steps undertaken in the FOSS machine learning workflows used within QGIS, QField and DE Pacific are outlined in figure 3, table 1.

These reproducible and scalable workflows have formed the foundation for the development of a range of products for classification in Digital Earth Pacific. This has also included collaborative work on land cover and land use change at national scales in partnership with LOS stakeholders as well as the work on invasive forest species mapping completed in Fiji as documented by Biukoto et al., (2026) and on marine habitat mapping (Metherall et al., 2024). This workflow serves as a foundation allowing for variations and advancements depending on the use-case.

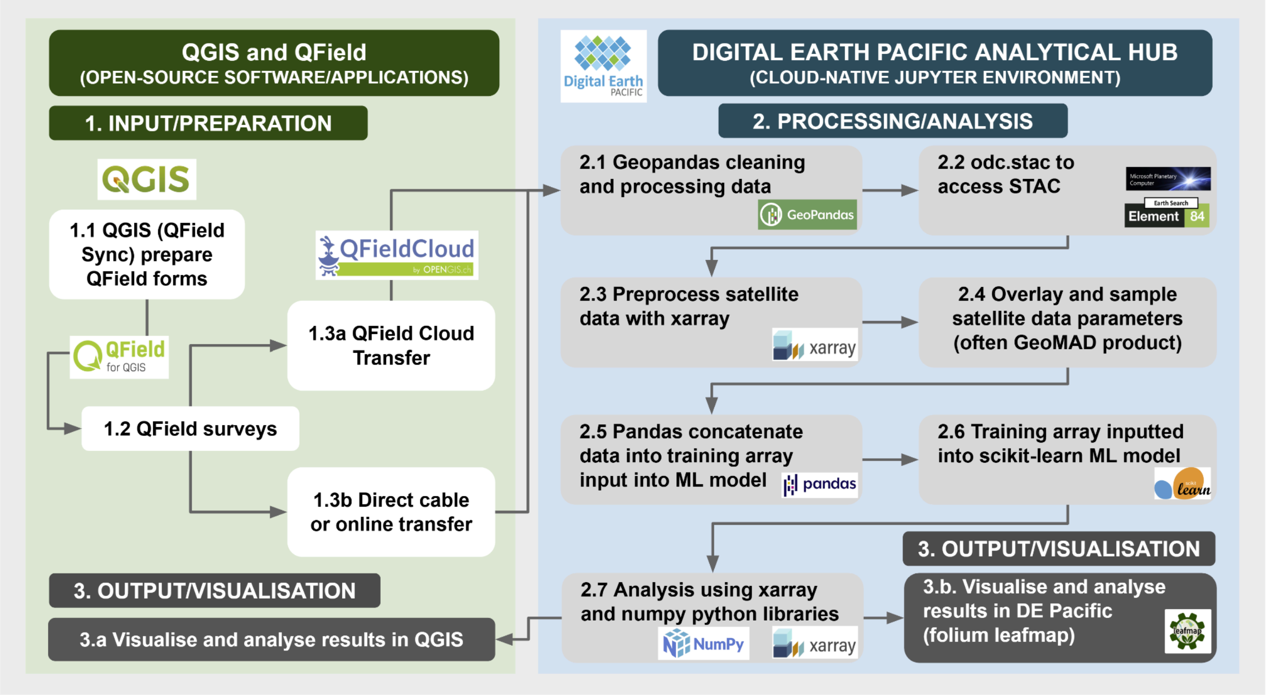


**Figure 3**. Free and Open-Source Software (FOSS) workflows within QGIS, QField and Digital Earth Pacific. These workflows can be replicated via GitHub: https://github.com/nicholasmetherall/digitalearthpacific-tonga-lulc and broader Digital Earth Pacific earth observations products can be accessed via: https://github.com/digitalearthpacific.

| |
|---|
| 1. QGIS to prepare QField forms |
| 2. QField to collect field data |
| 3. GeoPandas / pandas to process / clean / training data |
| 4. odc.stac library to load satellite imagery datasets from spatio-temporal asset catalogues |
| 5. Mask and preprocess satellite imagery using Xarray package |
| 6. Overlay training data over satellite imagery to sample satellite data parameters for each training data point |
| 7. Pandas concatenate the data into a training data array input into machine learning |
| 8. Integrate Scikit-learn machine learning Python library |
| 9. Export COG predictions |
| 10. Analyses of results using Xarray and NumPy Python libraries |
| 11. Visualise results in a folium leafmap Python package or export into QGIS for further visualisation and analyses |

Table 1. List of steps in the FOSS machine learning workflows used in Digital Earth Pacific.

The steps outlined in Table 1 are grouped and elaborated within three stages: 1) the input/preparation stage covered within steps 1-3 (in section 3.2), 2) the processing/analysis stage covered within steps 4-8 (in section 3.3) and 3) output/visualisation covered within steps 9-11 (in section 3.4).

### 3.2 Inputs and preparation: QGIS and QField for field data collection

QField is an open access and open-source mobile application that is connected to step 1) QGIS software. This mobile app allows for the collection of geotagged data points, transects, polygons, and other field data features. In the DEP LCAST context, it was used to collect field-based datasets which was crucial for the calibration and validation of machine learning land cover classification models. Step 2) The participatory elements of the QField surveys supported co-creation and sense of ownership of these LULC models and maps. The country driven surveys involved use of QField to collect GPS for different landcover classes across the sites of the field survey. The DEP LCAST Workshop surveys involved designing QField forms in QGIS for the collection of GPS data covering the various LULC classes determined by the local users in Tonga as well as the six standardized land cover classes as defined by the IPCC Good Practice Guidance for LULUCF Chapter 2: Basis for consistent representation of land areas (IPCC, 2003).

### 3.3 Processing and analysis: GeoPandas, odc-stac, GeoMAD cloudless mosaics, Pandas, Xarray, Scikit-Learn

Step 3) The Python library GeoPandas was used to help consolidate different classes by aggregating a wide range of classes into the simpler standardised IPCC classes where needed (Akyol et al. 2024; Geopandas Project, 2025; Jordhal, 2020). Many of these workflows (figure 9) are further automated and abstracted by the Digital Earth Pacific Abstraction Library or (DEPAL) (Singh, 2023) and through the ability to integrate the QField points into the ML model (Anderson, 2023). Step 4) The Python package odc.stac is used to access STAC utilizes the Open Data Cube (ODC) STAC interface to query and retrieve satellite imagery and metadata from open catalogs, such as those hosted Element 84 through Earth Search as well as the Microsoft Planetary Computer (Killough, 2018; Gadomski, 2023; Leith et al., 2024). Step 5) The retrieved satellite imagery is then processed with the Xarray library which manages multi-dimensional array data including time series of raster layers (Lee et al., 2024). Step 6) The training data from QField is overlaid by sample satellite data. This may involve spatially overlaying the cleaned vector points (from GeoPandas) onto the preprocessed satellite imagery, often using standardized GeoMAD (Geometric Median and Median Absolute Deviations) cloudless mosaic products (Leith, 2013; Roberts et al., 2017; Roberts et al., 2018). The ground truthed field data points then extract the values from the intersecting pixel (spectral band data, indices). Step 7) The Pandas library is used to concatenate this multiparameter data into a stacked training array (Pandas DataFrame serving as the feature matrix training array) input into ML model training. Step 8) This training array inputted into scikit-learn ML model for land cover classification or other predictive tasks. This completes the transformation of raw data into a trained analytical product.

### 3.4 Output and visualisation via Python libraries or exporting data back into QGIS

Step 9) The output maps of the machine learning predictions may be exported as a Cloud Optimized GeoTIFF or as tabular data for further analyses (NASA, 2020). This can involve creating a task graph or a ‘promise’ of a computation ‘lazy loading’ rather than executing the computation immediately (Xu et al., 2020). Step 10) this allows for further analysis using Xarray and NumPy Python libraries (Neli, 2023). Step 11) results may also be visualised in a folium leafmap Python package or written to disk and exported into QGIS for further visualisation and analyses (Alam, 2022; Sujithra & Shanmugapriyaa, 2024).

## 3. OUTPUTS AND OUTCOMES

### 3.1 QField field survey results

Through the QField surveys in the lead up to and during the LCAST workshop approximately 2,119 geotagged points were collected for both calibration and validation purposes. These data points captured a range of different LULC classes and were processed through customisation and standardisation processes using GeoPandas (Figure 4).

**Figure 4.** QField LULC data points collected through DE Pacific LCAST Workshop

### 3.2 LULC model prediction output map and table

The results of the LULC model prediction and the extent of each LULC class in hectares for Tongatapu in 2023 (Figure 5 and table 2).

As indicated by the results, the major landcover classes in Tongatapu are croplands (35.76%), forest (28.18%), grassland (23.80%). The minor landcover classes include settlements (6.47%), wetland (4.95%) and bare land (0.84%).

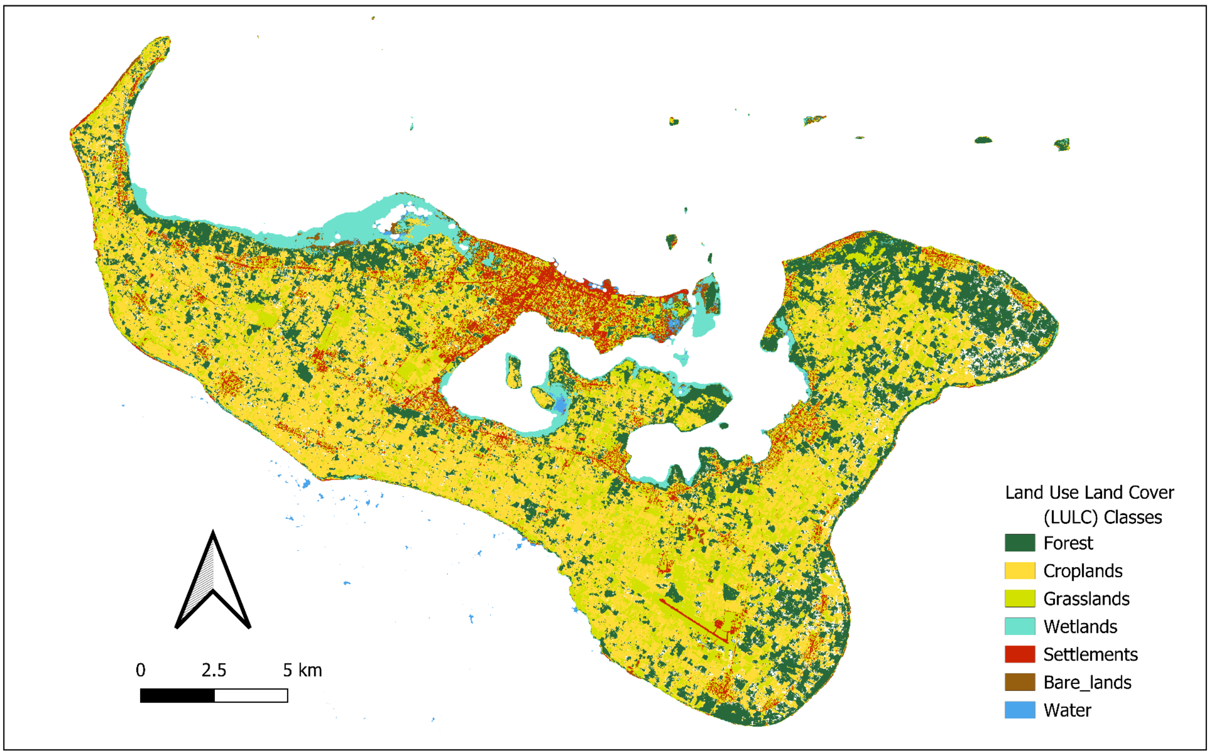


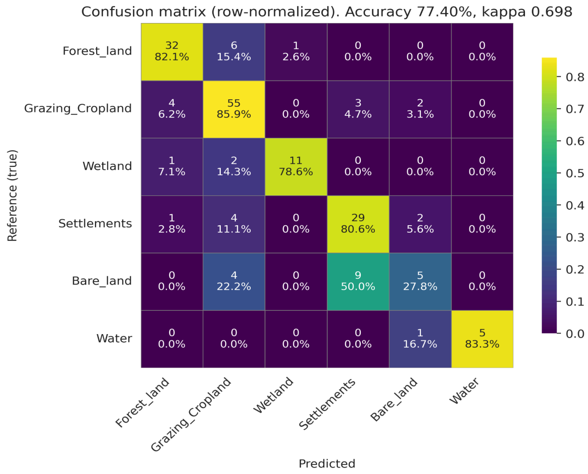
**Figure 5.** LULC machine learning model prediction results for 2023 in Tongatapu, Tonga.

| **LULC class** | Forest | Cropland | Grassland | Wetland | Settlements | Bare land |
|---|---|---|---|---|---|---|
| **Hectares** | 8,449.66 | 10,721.48 | 7,138.07 | 1,485.63 | 1,939.94 | 251.05 |
| **Percent** | 28.18 | 35.76 | 23.80 | 4.95 | 6.47 | 0.84 |

Table 2. Areas of LULC classes of Tongatapu in hectares.

### 3.2 Accuracy assessment results

Using an 80-20 calibration, validation split, the accuracy assessment results with 177 validation samples had an accuracy of 0.77 (77.40%) and Cohen's kappa: 0.70.

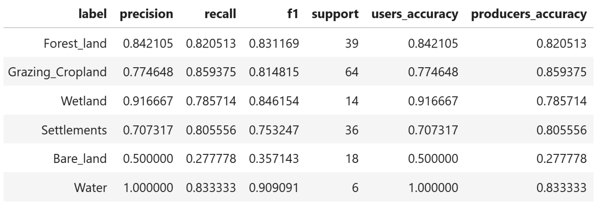


| | Forest_land | Grazing_Cropland | Wetland | Settlements | Bare_land | Water |
|---|---|---|---|---|---|---|
| Forest_land | 32<br>82.1% | 6<br>15.4% | 1<br>2.6% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% |
| Grazing_Cropland | 4<br>6.2% | 55<br>85.9% | 0<br>0.0% | 3<br>4.7% | 2<br>3.1% | 0<br>0.0% |
| Wetland | 1<br>7.1% | 2<br>14.3% | 11<br>78.6% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% |
| Settlements | 1<br>2.8% | 4<br>11.1% | 0<br>0.0% | 29<br>80.6% | 2<br>5.6% | 0<br>0.0% |
| Bare_land | 0<br>0.0% | 4<br>22.2% | 0<br>0.0% | 9<br>50.0% | 5<br>27.8% | 0<br>0.0% |
| Water | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 1<br>16.7% | 5<br>83.3% |

**Figure 6.** Confusion matrix results for the LULC model

### 3.3. Per-class metrics:

Overall, the class found to have relatively low user accuracy (0.62) and producer accuracy (0.28) was 'Bare_land' which was commonly confused for settlements and grazing or cropland (possibly during fallow periods).



| label | precision | recall | f1 | support | users_accuracy | producers_accuracy |
|---|---|---|---|---|---|---|
| Forest_land | 0.842105 | 0.820513 | 0.831169 | 39 | 0.842105 | 0.820513 |
| Grazing_Cropland | 0.774648 | 0.859375 | 0.814815 | 64 | 0.774648 | 0.859375 |
| Wetland | 0.916667 | 0.785714 | 0.846154 | 14 | 0.916667 | 0.785714 |
| Settlements | 0.707317 | 0.805556 | 0.753247 | 36 | 0.707317 | 0.805556 |
| Bare_land | 0.500000 | 0.277778 | 0.357143 | 18 | 0.500000 | 0.277778 |
| Water | 1.000000 | 0.833333 | 0.909091 | 6 | 1.000000 | 0.833333 |

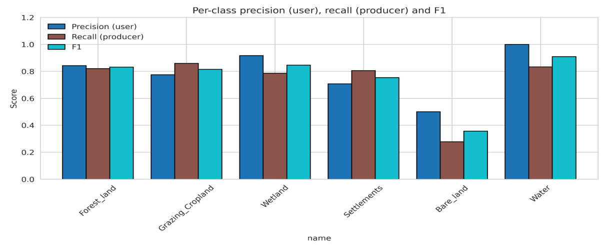


**Figure 7.** Per-class metrics: precision, recall and F1.

### 3.4 Lessons learned through two-way learning process

Through the workshop, five project-oriented sub-groups were formed based on their LULC mapping and monitoring objectives. There were two main ways these groups benefited from the customisability and replicability of the FOSS workflows of the DE Pacific LCAST Workshop (figure 3). 1) These groups were able to design their own QField forms to focus on their own specific project objectives and 2) the groups also were able to use GeoPandas to customise, aggregate and disaggregate the field data collected to be used towards multiple use case cases.

| |
|---|
| **Group 1: Mineral resources group**<br>(Ministry of Lands and Natural Resources) and the Tonga Geological Services |
| This group was most interested in mineral license areas, and permit areas for monitoring of mineral extraction zones. This group worked with monitoring the bare lands class and also defined their own mineral stockpile class. |
| **Group 2: Agricultural census group** |
| This group was composed of members of the Ministry of Agriculture, Food, and Forest. This group was particularly interested in the cropland and shrubland and grassland classes as these classes are usually indicators of different stages of the crop cycle and can highlight where land is being cultivated and productive and where it is not. This group had a particular interest in the annual agriculture census activities conducted by the ministry |
| **Group 3: Forest Inventory Group** |
| The forest inventory group was composed of members of MAFF mostly from the Forest Division in Tokomololo Forest Station. The forest inventory group was particularly interested in changes in forest classes for the National Forest Inventory (NFI) process. Within NFI workflows, land cover stratification is a foundational step. This group was particularly interested in different vegetation classes in particular forest, grasslands, wetlands and mangroves |
| **Group 4: coastal marine ecosystems** |
| The coastal marine ecosystems group included two members from the Ministry of Fisheries as well as from MAFF and GIZ. This group was interested in mangroves, sand, seagrass, algae rock, rubble.. |
| **Group 5: National Disaster Management Office (NDMO)** |
| The NDMO group was composed of participants from the Ministry of Energy, Information, Disaster Management Environment, and Climate Change (MEIDECC). This group were interested in seeing changes and land cover before and after hazard events including in areas along Kolovai along the Western Coast that had hotels and other tourism infrastructure that were destroyed. Hazards including the 2020 Tropical Cyclone Harold event and the 2022 Tonga Hunga Ha'apai volcano induced tsunami affected many areas including Nuku'alofa. These classes of settlements would experience a change to rubble or shrubs as weeds and vegetation wood grow on top of the rubble of these damaged structures. |

Table 2. Groups formed during the DE Pacific LCAST Workshop in Tonga.

### 3.5 DEP LCAST workshop evaluation and feedback

The course evaluation surveys were collated from a sample of participants from the workshop (n=18) representing their Ministries of the Government of Tonga (Fa'Anunu, 2023). The evaluation highlighted positive feedback from participants with 83.0% stating the course met expectations, 78.0% felt they achieved the desired learning outcomes (figure 8). In terms of the FOSS workflows, 72.2% of respondents categorising the learning activities and the technologies used in the class as 'extremely effective' (Fa'Anunu, 2023). The DE Pacific LCAST Workshop included not only representatives from the Government of Tonga but also two participants from Manaaki Whenua Landcare Research New Zealand and one participant from the Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH, (GIZ), the main development agency of Germany.

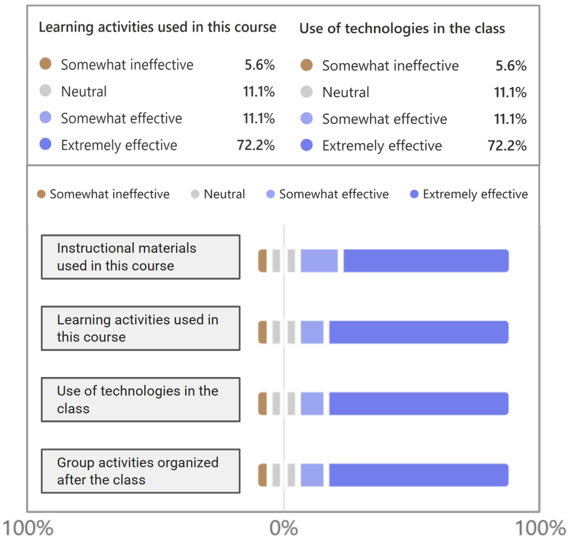


**Figure 8.** Participant feedback on the DE Pacific LCAST Workshop according to the independent workshop evaluation and report completed by a Tongan consultant (Fa'Anunu, 2023).

Through the DE Pacific LCAST Workshop participants gained exposure to the 'full ML workflow cycle' including all the steps required for the generation of LULC maps and models rather than just one stage (i.e. the survey field data collection stage). This was one of the unique offerings of this workshop making it distinct from other geospatial and earth observation capacity building workshops. One of the trade-offs associated with this approach was highlighted by five respondents from the sample who noted that more time would help for 'absorbing the knowledge' and further optimising the learning objectives and outcomes of the workshop (Fa'Anunu, 2023). Furthermore, four of the evaluation forms also highlighted the desire for further 'follow-up' training reinforcing this point that more time and exposure would be needed to build operational proficiency in these workflows for ongoing adoption and use. For future similar workshops, the evaluation forms highlighted the potential that a larger group may attend given that 93% likelihood of participants to recommend this course to other colleagues (Fa'Anunu, 2023).

# 4. DISCUSSION

## 4.1. The trajectory of earth observations platforms

Earth observation technologies have advanced rapidly over the past decades, becoming not only more detailed in terms of spectral and temporal coverage but also increasingly accessible for a wide range of users (Zhao et al., 2022). Yet there have been ongoing barriers to the uptake and adoption of earth observation workflows and analytics needed to inform policy makers. Often these barriers have included complex and overly technocratic language and workflows, obstructing access and obscuring insights from satellite data (Filippi & Aiello, 2025). Furthermore the use of closed, licensed or pay-wall limited software, tools and workflows have further constrained customisability, accessibility and sustainability (Zhu & Zhou, 2012; Haque, 2025). This has commonly resulted in analysis of satellite data being limited to academic and research oriented groups (Potts et al., 2024; Gaddipati et al., 2025).

A range of options have emerged with the potential to support a global user base to gain access to insights from earth observations including Google Earth Engine which can be combined with FOSS workflows to support national development outcomes (Tamiminia et al., 2020, Tesfaye et al., 2024). However, as a closed and proprietary platform, many of these workflows remain constrained due to technical and emerging cost-barriers (Gomez et al., 2020, Pérez-Cutillas, 2023). Others may rely on short-term asset and memory packages and libraries that sometimes expire or become deprecated over time, reducing the overall long-term replicability of these workflows for wider users (Kedron & Frazier, 2022). Recent advancements in cloud computing Spatial Data Infrastructure including the Open Data Cube and Digital Earth Australia, Digital Earth Africa as well as Digital Earth Pacific have the potential to enable wider access for various users to access standardized and customised replicable workflows (Lewis et al., 2017, Giuliani et al., 2020; Leith et al., 2024). These FOSS workflows are much more reproducible and scalable than what is possible through the closed-proprietary alternatives. Combining DE Pacific with further FOSS workflows as outlined in this paper, can help build capacity in the long term without cost. This is especially critical in the Pacific where there is relatively limited resourcing for earth observation capacity.

## 4.2. The value of ‘full-cycle’ participatory learning and action for locally-responsive capacity building

The value associated with capacity building and education based on the principles of FOSS4G (Free and Open Source Software for Geospatial) has been widely documented (Ciolli et al., 2017; Vyaz, 2020; Zatelli et al., 2022; Duncan et al., 2022; da Silva Mano & Augustijn, 2023; Oxoli et al., 2023; Goswami et al., 2024). Yet in this paper, we take a crucial step forwards in demonstrating capacity and skills transfer across a ‘full-cycle’ of FOSS workflows from field survey data collection through data cleaning, preprocessing, machine learning, post-processing, analysis and visualisations as well as the insights gained through the two-way learning of the DE Pacific LCAST Workshop in Tonga. As shown in the results, these workflows resulted in an overall accuracy of 77.40%. In shedding light on these replicable workflows through an open source version control system, we also welcome inputs into constructive feedback leading to further improvements and advancements over time (figure 3) via the following link: https://github.com/nicholasmetherall/digitalearthpacific-tonga-lulc.

The DE Pacific LCAST Workshop approach to transferring the skills within the ‘full-cycle’ of replicable FOSS workflows is intended to allow for a greater overall sense of ownership felt by workshop participants (Theisohn and Lopes, 2013). This mode of two-way learning draws on the lessons learned from decades of work on Participatory Learning and Action (PLA) tools, participatory mapping and participatory statistics (Chambers, 1994; Mukherjee, 2002; Holland; 2013). This sense of awareness, understanding and eventually ‘ownership’ extends, not only to the products but to the processes and workflows themselves (Theisohn and Lopes, 2013). For example, through the field surveys, participants were able to draw more concrete connections of awareness and understanding between their grounded observations and the changes they analysed through the land cover modelling of changes in pixels and sample areas.

## 4.3. Customisability and flexibility of participatory FOSS workflows for LULC modelling and ongoing monitoring

The value of participatory and FOSS workflows and the increased ability for reciprocal ‘two-way’ learning remains a challenge to measure (Wubishet, 2013; Knowles et al., 2021). Yet one example was most clearly demonstrated through the customisability of the QField data collection and GeoPandas workflows. These workflows allowed participants to define their own priority land cover classes that were identified as important to their own Ministry and work objectives. The five mixed-Ministry ‘taskforce’ working group projects formed in the LCAST workshop highlighted the interdisciplinary nature of the applications of the earth observation data and the diversity and versatility of the outputs to be applied to different use cases depending on the working group (figure 9). Indeed, the instructors of the workshop were able to learn from the local contextual knowledge and expertise used as inputs into these QField forms and GeoPandas classes as well as the resulting outputs of LULC models within the projects conducted by these participants from Tonga. Additional customisability of the LULC monitoring objectives was made possible through 1) the workshop activities in which participants could design their own QField forms and 2) the GeoPandas class aggregation and disaggregation workflows (Singh, 2023). These GeoPandas workflows enabled sub-groups to focus on specific classes (e.g. mining areas) while also allowing for the wider group to monitor the main IPCC LULC classes through aggregating these sub-classes into standardised broader categories of landcover. It was found that having many disaggregated classes led to more complex models with lower overall accuracy as compared to when aggregating into fewer classes. In this case, this caused overall accuracy to range between approximately 51 to 77%. Overall, the main value of these GeoPandas class aggregation and disaggregation workflows is that they are particularly useful for developing more customisable and flexible LULC models for example through systematic subdivision into more detailed LULC sub-classes (Vali et al., 2020). Such customisability may also have potential to open up spaces and technical capabilities for Pacific Island participants to more readily incorporate Traditional Ecological Knowledge or Indigenous Local Knowledge into their own field data collection and in the longer-term to use such tools to build up capacity of Pacific research and Pacific researchers and practitioners (Suaalii-Sauni & Fulu-Aiolupotea, 2014). This is important within diverse geographic contexts of the Pacific and LULC where there is no one ideal classification of land use and land cover, and it is unlikely that one could ever be developed. There are different perspectives in the classification process, and the process itself tends to be subjective, even when an objective numerical approach is used (USGS, 1976). These concepts may also transfer beyond the context of land use and land cover.

Overall, two-way learning and capacity building as seen through the DE Pacific LCAST approach offers a potential mode of delivering these skills and capabilities as a form of Participatory Learning and Action with greater contextual suitability for PICTs or LOS (Suaalii-Sauni & Fulu-Aiolupotea, 2014; Cammock et al., 2021; Fa'Anunu, 2023). The value of these FOSS and participatory processes may be evidenced to some extent through the evaluation forms of the DE Pacific LCAST Workshop. However, actually measuring the extent to which FOSS workflows and participatory learning and action processes as opposed to closed proprietary software and more top-down processes influence outcome across complex and often qualitative areas such as local sense of ownership, long-term uptake and adoption of technologies and even the extent of meaningful two-way learning will require further examination and research over longer timescales (Fa'Anunu, 2023).

### 4.4. Implications for future earth observation capacity across the wider Pacific

Since the first DE Pacific LCAST Workshop in 2023, these workflows have been used in LCAST and similar country-driven workshops in ten Pacific Island Countries and Territories (PICTs) or Large Ocean States (LOS) including Fiji, RMI, Tuvalu, Palau, Cook Islands, PNG, Vanuatu, New Caledonia and the Solomon Islands. The capacity building has been applied across diverse areas beyond LULC monitoring including but not limited to national forest inventories and invasive species monitoring (Biukoto et al., 2025) as well as marine benthic habitat mapping (Metherall et al., 2024). Future applications may include areas as diverse as plastic waste monitoring and radar vegetation indices to monitor the impacts of climate hazard events (Waqa et al., forthcoming 2026).

## 5. CONCLUSIONS

The paper highlights two main areas of methods. 1) the ‘full-cycle’ FOSS workflows for machine learning using QField, QGIS and Digital Earth Pacific (figure 3) and 2) the participatory two-way learning approaches within the DE Pacific LCAST Workshop. These two areas are highlighted as approaches with potential to support two-way learning and long-term capacity building in earth observation in the Pacific region. These workflows are made possible through Digital Earth Pacific which has been built with support from the collective experience of Open Data Cube, Digital Earth Australia and Digital Earth Africa (Killough, 2018; Leith et al., 2024). As open-source workflows, all approaches documented in this paper are shared through the following GitHub repository: https://github.com/digitalearthpacific and https://github.com/nicholasmetherall/digitalearthpacific-tonga-lulc/

## 6. ACKNOWLEDGMENTS

Mrs Halalilika Etika, Mr. Titie Kaufusi, Ms. Angahiki Fifita, Mr. Falekakala Taufa, Mr. Nikolasi Heni, Mr. Punefu Fili Ministry of Lands and Natural Resources, Ahi’ahi Saipaia, Mr Heimuli Likiafu, Ms. Lesieli Tu’itupou, Ms. Susan Tu’itavake, Ms. Telesia Lea’aepulu, Mr. Steven Hamani, Mr. Sione Likiafu, Mr. Tevita Kengike, Mr. Michael Tupou Ministry of Agriculture and Food and Forestry, Ms. Sela Kioa, Mr. Viliami Fatongiatau Ministry of Fisheries, Mr. Patelesio Manukeu, Ms. Sikini Falesiva Ministry of Infrastructure, Mr. Lazaus Vaipulu, Mr. Sila Manu Ministry of MEIDECC, Mr. Tupou Vaipulu Ministry of Tourism, Ms. ‘Ativeniana Pomana, Mr. Hemaloto Moliteka Tonga Statistics Department, Professor Martin Tomko University of Melbourne, Dr Shaun Williams Earth Sciences, Dr Stuart Minchin Bureau of Meteorology, Ms Adi Loraini Baleilomaloma, Ms Samantha Krawczy, Mr Nabil El Halwani Zarin Khan SPC. Ms. Laise Harris, Mr. Peter Newsome Manaki Whenua Landcare New Zealand, Mr Aditya Agrawal D4DInsights, Mr. Shyam Lodhia Deutsche Gesellschaft für Internationale Zusammenarbeit Fiji.

## 7. REFERENCES

Alam, M.M., Torgo, L. and Bifet, A., 2022. A survey on spatio-temporal data analytics systems. A*CM Computing Surveys*, 54(10s), pp.1-38.

Anderson, J. 2023, jessjaco GitHub profile, GitHub, viewed 24 July 2023, https://github.com/jessjaco

Akyol, M.A., Duzgun, S. and Baykal, N., 2024. landusemix: A Python package for calculating land use mix. *SoftwareX*, 27, p.101861.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H.A., Mahdavi, S., Ghahremanloo, M., Parsian, S. and Wu, Q., 2020. Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, pp.5326-5350.

Bryan, W.B., Stice, G.D. and Ewart, A., 1972. Geology, petrography, and geochemistry of the volcanic islands of Tonga. *Journal of geophysical research*, 77(8), pp.1566-1585. https://doi.org/10.1029/JB077i008p01566 (10 December 2025).

Bednar, J.A. and Durant, M., 2023. The Pandata scalable open-source analysis stack. In Proceedings of the 2023 *Scientific Computing in Python Conference* (SciPy 2023) (pp. 85-92). https://proceedings.scipy.org/articles/gerudo-f2bc6f59-00b.pdf (2 December 2025).

Biukoto, E. Metherall, N. Baleilomaloma, AB. Mateboto, J. Naikatini, A. Tupua, V. Killough, B. Leith, A. Nisha, Z. Caucau, T. Waqa, M. (forthcoming 2026). Applying replicable open-source workflows to identify invasive flora: Spathodea campanulata in a small-scale forest plot in Korotari, Fiji. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*

Chambers, R. (1994). Paradigm shifts and the practice of participatory research and development.

Ciolli, M., Federici, B., Ferrando, I., Marzocchi, R., Sguerso, D., Tattoni, C., Vitti, A., Zatelli, P., 2017. FOSS tools and applications for education in geospatial sciences. *ISPRS International Journal of Geo-Information*, 6(7). doi.org/10.3390/ijgi6070225. (4 December 2025).

Congedo, L., 2016. Semi-automatic classification plugin documentation. Release, 4(0.1), p.29.

Crowley, J., York, D., Soden, R. and Deparday, V., 2014. Open Data for resilience initiative field guide. World Bank Publications.

Cutts, A. and Graser, A., 2018. Learn QGIS: Your step-by-step guide to the fundamental of QGIS 3.4. Packt Publishing Ltd.

da Silva Mano, A. and Augustijn, P. W. M.: Teaching geographic information science concepts with QGIS and the living textbook – towards a sustainable and inclusive distance education, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W7-2023, 25–31, https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-25-2023, 2023.

Duncan, J., Davies, K. P., Saipai, A., Vainikolo, L., Wales, N., Varea, R., ... & Boruff, B. (2022). An Open-Source Mobile Geospatial Platform for Agricultural Landscape Mapping: A Case Study of Wall-To-Wall Farm Systems Mapping in Tonga. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 119-126. https://isprs-archives.copernicus.org/articles/XLVIII-4-W1-2022/119/2022/isprs-archives-XLVIII-4-W1-2022-119-2022.html (22 November 2025)

Fa'Anunu, F. 2023. Digital Earth Pacific Land Cover Assessment Skills Transfer (LCAST) Workshop. Report.

Filippi, E. and Aiello, A., 2025. Barriers, Benefits, and Influential Factors of Adopting Earth Observation Satellite Data at Local and Regional Levels: The Case of the Italian LRAs. *Sustainability*, 17(1), p.145. https://doi.org/10.3390/su17010145 (30 November 2025).

Franklin, J., Drake, D.R., Bolick, L.A., Smith, D.S. and Motley, T.J., 1999. Rain forest composition and patterns of secondary succession in the Vava'u Island Group, Tonga. *Journal of Vegetation Science*, 10(1), pp.51-64. https://doi.org/10.2307/3237160.

Gaddipati, H., Borowitz, M., Rolls, G., Li, X., Chawla, G., Patel, R. and Woodall, B., 2025. Challenges in using satellite data for non-remote sensing specialists, an exploratory case study. *Data & Policy*, 7, p.e28. https://doi.org/10.1017/dap.2025.7 (3 December 2025).

Gavin, D., Dhu, T., Sagar, S., Mueller, N., Dunn, B., Lewis, A., Lymburner, L., Minchin, S., Oliver, S., Ross, J. and Thankappan, M., 2018, July. Digital earth Australia-from satellite data to better decisions. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 8633-8635). IEEE. https://doi.org/10.1109/IGARSS.2018.8518160 (1 December 2025).

GeoPandas Project 2025, Introduction — GeoPandas documentation, viewed 10 December 2025, https://geopandas.org/en/stable/getting_started/introduction.html

Gadomski, P. (2023) 'How Microsoft's Planetary Computer Uses STAC', Element 84. 05 September. Available at: https://element84.com/geospatial/how-microsofts-planetary-computer-uses-stac/ (Accessed: 10 December 2025).

Georgiadou, Y., Bana, B., Becht, R., Hoppe, R., Ikingura, J., Kraak, M. J., ... & Verplanke, J. (2011). Sensors, empowerment, and accountability: a Digital Earth view from East Africa. *International journal of digital earth*, 4(4), 285-304.

Giuliani, G., Chatenoux, B., Piller, T., Moser, F. and Lacroix, P., 2020. Data Cube on Demand (DCoD): Generating an earth observation Data Cube anywhere in the world. *International Journal of Applied Earth Observation and Geoinformation*, 87, p.102035.

Gomes, V.C., Carlos, F.M., Queiroz, G.R., Ferreira, K.R. and Santos, R., 2021. Accessing and processing Brazilian earth observation data cubes with the open data cube platform. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, pp.153-159.

Gorelick, N. (2013, April). Google earth engine. In EGU general assembly conference abstracts (Vol. 15, p. 11997). Vienna, Austria: American Geophysical Union.

Goswami, J., Senpakapriya, V., Goswami, C., Sarma, K. K., and Aggarwal, S. P. 2024. Disaster Preparedness and capacity building for Resilience in Agriculture, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-5-2024, 17–22, https://doi.org/10.5194/isprs-archives-XLVIII-5-2024-17-2024, (8 December 2025).

Graser, A., Sutton, T., & Bernasocchi, M. (2025). The QGIS project: Spatial without compromise. Patterns. https://www.cell.com/patterns/fulltext/S2666-3899(25)00113-8 (8 December 2025).

Groot, R. 1997. Spatial data infrastructure (SDI) for sustainable land management. https://wgbis.ces.iisc.ac.in/energy/HC270799/LM/SUSLUP/KeySpeakers/AGroot.pdf (30 November 2025).

Haque, M.A., 2025. Open-Source vs Proprietary Engineering Software: A Comparative Study of Adoption, Customisation and Innovation in Bangladesh. NextGen Research, 1(04), pp.22-39. http://nextgresearch.com/index.php/nextgr/article/view/22 (2 December 2025).

Holland, J. (Ed.). (2013). Who counts?: the power of participatory statistics. Rugby: Practical Action Publishing.

IPCC, 2003. Good Practice Guidance for Land Use, Land-Use Change and Forestry, Chapter 2: Basis for Consistent Representation of Land Areas, IGES, Hayama, Japan. https://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_files/Chp2/Chp2_Land_Areas.pdf (2 December 2025).

Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., … Leblanc, F. 2020, geopandas/geopandas: v0.8.1, Zenodo, viewed 10 December 2025, https://doi.org/10.5281/zenodo.3946761

Kedron, P. & Frazier, A. E. (2022) 'How to Improve the Reproducibility, Replicability, and Extensibility of Remote Sensing Research', *Remote Sensing*, 14(21), 5471. doi: 10.3390/rs14215471.

Killough, B., 2018, July. Overview of the open data cube initiative. In IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium (pp. 8629-8632). IEEE. https://doi.org/10.1109/IGARSS.2018.8517694 (22 November 2025).

Knowles, S.E., Allen, D., Donnelly, A., Flynn, J., Gallacher, K., Lewis, A., McCorkle, G., Mistry, M., Walkington, P. and Drinkwater, J., 2021. More than a method: trusting relationships, productive tensions, and two-way learning as mechanisms of authentic co-production. *Research Involvement and Engagement*, 7(1), p.34.

Lee, K., Kim, K., Lee, H. and Lee, J. (2024) 'Applying Xarray and Cloud Computing for Time Series Data Visualization and Analysis with Compact Advanced Satellite (CAS) 500 Images of Korea', 45th COSPAR Scientific Assembly, 45, p. 159. Available at: https://ui.adsabs.harvard.edu/abs/2024cosp...45..159L
(Accessed: 10 December 2025).

Leith, A. (2013.) alexgleith. GitHub. Available at: https://github.com/alexgleith
(Accessed: 11 December 2025).

Leith, Alex & Adams, Caitlin & Borges, David & Sundwall, Jed. (2024). Cloud Native Geospatial: Realizing the Digital Earth Vision. IGARSS Conference Paper 2024. 6886-6889. 10.1109/IGARSS53475.2024.10642711.

Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J. and Wu, W., 2017. The Australian geoscience data cube—foundations and lessons learned. *Remote Sensing of Environment,* 202, pp.276-292.

Liu, Z., Foresman, T., van Genderen, J. and Wang, L., 2019. Understanding digital earth. In Manual of digital earth (pp. 1-21). Singapore: Springer Singapore.

Lucas, R., Mueller, N., Siggins, A., Owers, C., Clewley, D., Bunting, P., Kooymans, C., Tissott, B., Lewis, B., Lymburner, L. and Metternicht, G., 2019. Land cover mapping using digital earth Australia. *Data*, 4(4), p.143. https://doi.org/10.3390/data4040143 (30 November 2025).

McKinney, W., 2011. pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, 14(9), pp.1-9.

Mardani, M., Mardani, H., De Simone, L., Varas, S., Kita, N. and Saito, T., 2019. Integration of machine learning and open access geospatial data for land cover mapping. *Remote Sensing*, 11(16), p.1907.
Martin, M.V., Goran, W.D., Lozar, R.C., Messersmith, J.M. and Ruiz, M.S., 1989. GRASS/GIS (Geographic Resources Analysis Support System/Geographic Information System) Implementation Guide (No. CERLSRN9002).

Maťašovská, V., Kratina, J. and Kabeláč, R., 2024. QFIELD–mobile application for data collection established on the principles of open source software. Vodohospodářské technicko-ekonomické informace, 66(2), pp.22-28.

Melrose, R.T., 2018. Mapping inundation and wetlands in arid Australia using satellite radar and optical remote sensing (Doctoral dissertation, UNSW Sydney), https://unsworks.unsw.edu.au/entities/publication/cc07aa58-4b3d-42a9-aaf7-854f51f0459f/full (29 November 2025).

Metherall, N. Anderson, J. Singh, S. Leith, A. Saipaia, A. Fa'anunu, L. Dhaja, C. Jackson, N. Waqa, M. Biukoto, E. Lodhia, S. Mlisa, A. Killough, B. Bernasocchi, M. (forthcoming 2026). Free and open-source machine learning workflows for co-creating national-scale classification models through country-driven QField surveys and Digital Earth Pacific. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*

Metherall, N., Bili, J., Sobey, M., Lodhia, S., Dirking, V., Linzatti, R., Anderson, J. and Singh, S. 2024., Digital Earth Pacific–Approaches to monitoring fishery ecosystems. https://www.spc.int/updates/blog/2024/08/digital-earth-pacific-approaches-to-monitoring-fishery-ecosystems (22 December 2024).

Mohamed-Ghouse, Z.S., Desha, C. and Perez-Mora, L., 2019. Digital earth in Australia. In Manual of digital earth (pp. 683-711). Singapore: Springer Singapore.

Momjian, B., 2001. PostgreSQL: introduction and concepts (Vol. 192, p. 2001). New York: Addison-Wesley.

Mubea, K., Boamah, E., Rebelo, L., Yuan, F. and Lewis, A., 2022, October. Digital Earth Africa–Operational EO-Based Insights for Africa's Development Challenges. In International Conference of the African Association of Remote Sensing and the Environment (pp. 487-495). Cham: Springer Nature Switzerland.

Muellerklein, O., 2017. A review of the emergent ecosystem of collaborative geospatial tools for addressing environmental challenges. Computers, Environment and Urban Systems.

Mukherjee, N., 2002. Participatory learning and action: With 100 field methods (No. 4). Concept Publishing Company.

NASA (2020) Cloud-Optimized Format Study. NASA Technical Reports Server. Available at: https://ntrs.nasa.gov/citations/20200001178
(Accessed: 10 December 2025).

Nebert, D.D. 2004. Developing Spatial Data Infrastructure: SDI Cookbook Version 2.0. Global Spatial Data Infrastructure (GSDI), p.171 http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_GSDI_2004_ver2.pdf (30 November 2025).

Nelli, F., 2023. The numpy library. In Python data analytics: with pandas, numpy, and matplotlib (pp. 45-72). Berkeley, CA: Apress.

Obe, R. and Hsu, L., 2021. PostGIS in action. Simon and Schuster.

Oliphant, T.E., 2006. Guide to numpy (Vol. 1, p. 85). USA: Trelgol Publishing.

Ostadabbas, H., Weippert, H., & Behr, F. J. (2020). Using the synergy of qfield for collecting data on-site and qgis for interactive map creation by alkis® data extraction and implementation in postgresql for urban planning processes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 679-683.

Oxoli, D., Reyes, S. R., Peng, S., Brovelli, M. A., Coetzee, S., Ivanova, I., Leonardi, J. A., Rawal, D., Vacca, G., and Zlatanova, S.: CAPACITY BUILDING FOR GIS-BASED SDG INDICATORS ANALYSIS WITH GLOBAL HIGH-RESOLUTION LAND COVER DATASETS, Int. *Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W2-2023, 559–564, https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-559-2023, 2023.

Pacific Community (SPC), 2018. Pacific Data Hub (PDH). SPC – Statistics for Development Division.

Pacific Community (SPC), 2025. Pacific Ocean Portal relaunched: A new era for ocean data access and regional resilience. SPC Media Release, 1 October 2025.

Pérez‑Cutillas, P., Pérez‑Navarro, A., Conesa‑García, C., Zema, D. A. & Amado‑Álvarez, J. P. (2023) 'What is going on within Google Earth Engine? A systematic review and meta‑analysis', *Remote Sensing Applications: Society and Environment*, 29, 100907. doi: 10.1016/j.rsase.2022.100907.

Potts, D.A., Ferranti, E.J. and Vande Hey, J.D., 2024. Investigating the barriers and pathways to implementing satellite data into air quality monitoring, regulation and policy design in the United Kingdom. Environmental Science & Policy, 151, p.103621. https://doi.org/10.1016/j.envsci.2023.103621 (3 December 2025).

Raitt, R.W., Fisher, R.L. and Mason, R.G., 1955. Tonga trench.

Rajabifard, A. and Williamson, I.P., 2001, April. Spatial data infrastructures: concept, SDI hierarchy and future directions. In Proceedings of GEOMATICS'80 Conference (Vol. 10). https://hdl.handle.net// (30 November 2025).

Roberts, D., Mueller, N. Mcintyre, A. 2017. "High-Dimensional Pixel Composites From Earth Observation Time Series," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6254-6264, doi: 10.1109/TGRS.2017.2723896.

Roberts, D. Dunn, B. and N. Mueller. 2018. "Open Data Cube Products Using High-Dimensional Statistics of Time Series," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, pp. 8647-8650, doi: 10.1109/IGARSS.2018.8518312.
Saulnier, S., 2022. Affected the Kingdom of Tonga in the. Extreme Events in Human Evolution: From the Pliocene to the Anthropocene, p.748755096.

Saipaia, A. Summary of Agriculture Sector and use of QField in Tonga. 2025. Summary Report.

Singh, S. 2023. SOPAC GitHub repository, GitHub, viewed 20 July 2023, https://github.com/sopac

SPC, 2021. Digital Earth Pacific: Needs assessment report. Report. https://www.spc.int/resource-centre/publications/digital-earth-pacific-needs-assessment-report (1 December 2025).
Sujithra, M. and Shanmugapriyaa, D., 2024. Spatial Data Visualization With Python: Techniques, Tools, and Real. Geospatial Application Development Using Python Programming, p.123.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S. and Brisco, B. (2020) 'Google Earth Engine for geo‑big data applications: A meta‑analysis and systematic review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, pp. 152–170. doi: 10.1016/j.isprsjprs.2020.04.001.

Tesfaye, W., Elias, E., Warkineh, B., Tekalign, M. and Abebe, G. (2024) 'Modeling of land use and land cover changes using Google Earth Engine and machine learning approach: implications for landscape management', *Environmental Systems Research*, 13, Article 31. doi: 10.1186/s40068-024-00366-3.

Theisohn, T., & Lopes, C. (2013). Ownership leadership and transformation: Can we do better for capacity development. Routledge.

Vancauwenberghe, G., Valečkaitė, K., Van Loenen, B. and Donker, F.W., 2018. Assessing the openness of spatial data infrastructures (SDI): Towards a map of open SDI. *International Journal of Spatial Data Infrastructures Research*, 13, pp.88-100.

Voidrot, M.F. and Percivall, G., 2020, June. Ogc geospatial coverages data cube community practice. In IOP Conference Series: *Earth and Environmental Science* (Vol. 509, No. 1, p. 012058). IOP Publishing.

Vali, A., Comai, S. and Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), p.2495.

Vyas, A.: MULTI-LEVEL EDUCATION AND CAPACITY BUILDING FRAMEWORK FOR TECHNOLOGY ADAPTATION, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B5-2020, 23–27, https://doi.org/10.5194/isprs-archives-XLIII-B5-2020-23-2020, 2020.

Waqa,V. M., Metherall, N., Wara, L., Vanualailai, M., Leith, A., Killough, B., Biukoto, E., Brown, J., Dhaja, C., O'Connor, R.S (2026 forthcoming). Using the Radar Vegetation Index (RVI) to Assess the Impacts of Tropical Cyclone Yasa on the Agricultural Sector of Vanua Levu, Fiji, with Case Studies on Sugarcane and Kava: A remote sensing approach for loss and damage accounts. *Loss and Damage Research Symposium*.

Warmerdam, F., 2008. The geospatial data abstraction library. In Open source approaches in spatial data handling (pp. 87-104). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wilson, D.D., Tefera, G.W. and Ray, R.L., 2025. Application of Google Earth Engine to Monitor Greenhouse Gases: A Review. Data, 10(1), p.8. https://doi.org/10.3390/data10010008 (28 November 2025).
World Bank & GFDRR, 2013. Better Risk Information for Smarter Investments: Catastrophe Risk Assessment Methodology — Pacific Catastrophe Risk Assessment and Financing Initiative (PCRAFI). The World Bank / GFDRR.

Wu, H., Li, Y., Lin, A., Fan, H., Fan, K., Xie, J. and Luo, W., 2024. A review of crowdsourced geographic information for land-use and land-cover mapping: current progress and challenges. *International Journal of Geographical Information Science*, 38(11), pp.2183-2215.

Wubishet, Z.S., Bygstad, B. and Tsiavos, P., 2013. A participation paradox: Seeking the missing link between free/open source software and participatory design. *Journal of Advances in Information Technology,* 4(4), pp.181-193.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., Liu, J., Wang, T., Yan, Z. and Zhu, J. (2022) 'Cloud-based storage and computing for remote sensing big data: a technical review', *International Journal of Digital Earth*, 15(1), pp. 1417–1445. doi: 10.1080/17538947.2022.2115567.

Zatelli, P., Tattoni, C., Gobbi, S., Cantiani, M. G., La Porta, N., and Ciolli, M.: MODELING OF FOREST LANDSCAPE EVOLUTION AT REGIONAL LEVEL: A FOSS4G APPROACH, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-4/W1-2022, 553–560, https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-553-2022, 2022.

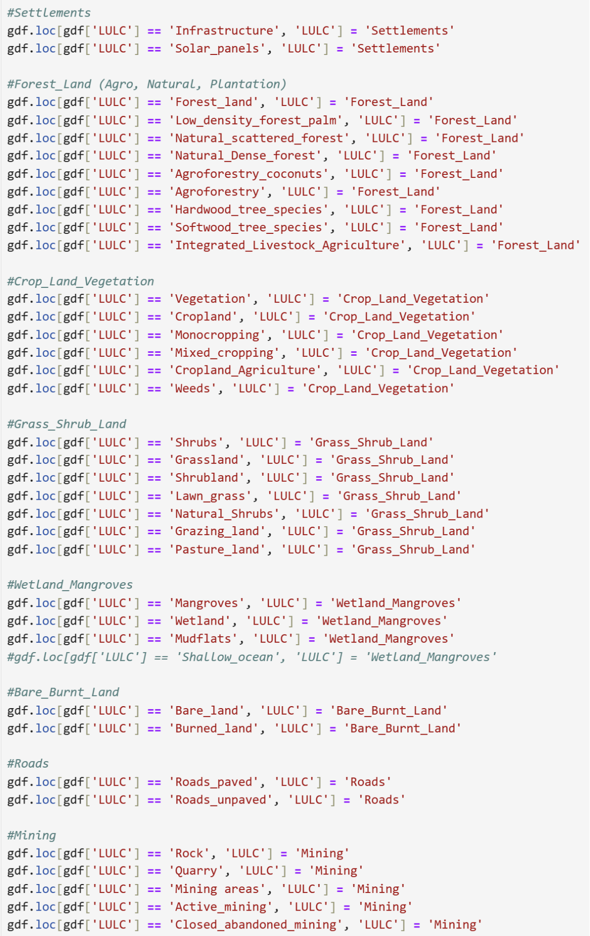
Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y. and Gong, P., 2022. An overview of the applications of earth observation satellite data: impacts and future trends. *Remote Sensing*, 14(8), p.1863. https://doi.org/10.3390/rs14081863 (30 November 2025).

Zhu, K.X. and Zhou, Z.Z., 2012. Research note—Lock-in strategy in software competition: Open-source software vs. proprietary software. *Information systems research*, 23(2), pp.536-545. https://doi.org/10.1287/isre.1110.0358 (2 December 2025).

# 8. APPENDIX

## 8.1. Local customisation:

Through country-driven workshops and surveys, local participants are able to contribute to a spectral database that allows for the training of machine learning land cover models. There are options to collect more detailed classes, including other land uses and land cover types outside of the standardized six classes. By including Traditional Ecological Knowledge (TEK) there is also greater room for localization and customization of capabilities with greater room for local inputs into capturing more complexity in terms of Land Use and Land Cover (LULC) changes . Through the DE Pacific LCAST Workshop, GeoPandas workflows were used to help with this customisation and standardisation of classes and sub-classes. In this example, a longer list of land cover classes could aggregated into a simpler list, including for greater compatibility with the IPCC classes (figure 6).



```
#Settlements
gdf.loc[gdf['LULC'] == 'Infrastructure', 'LULC'] = 'Settlements'
gdf.loc[gdf['LULC'] == 'Solar_panels', 'LULC'] = 'Settlements'

#Forest_Land (Agro, Natural, Plantation)
gdf.loc[gdf['LULC'] == 'Forest_land', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Low_density_forest_palm', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Natural_scattered_forest', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Natural_Dense_forest', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Agroforestry_coconuts', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Agroforestry', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Hardwood_tree_species', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Softwood_tree_species', 'LULC'] = 'Forest_Land'
gdf.loc[gdf['LULC'] == 'Integrated_Livestock_Agriculture', 'LULC'] = 'Forest_Land'

#Crop_Land_Vegetation
gdf.loc[gdf['LULC'] == 'Vegetation', 'LULC'] = 'Crop_Land_Vegetation'
gdf.loc[gdf['LULC'] == 'Cropland', 'LULC'] = 'Crop_Land_Vegetation'
gdf.loc[gdf['LULC'] == 'Monocropping', 'LULC'] = 'Crop_Land_Vegetation'
gdf.loc[gdf['LULC'] == 'Mixed_cropping', 'LULC'] = 'Crop_Land_Vegetation'
gdf.loc[gdf['LULC'] == 'Cropland_Agriculture', 'LULC'] = 'Crop_Land_Vegetation'
gdf.loc[gdf['LULC'] == 'Weeds', 'LULC'] = 'Crop_Land_Vegetation'

#Grass_Shrub_Land
gdf.loc[gdf['LULC'] == 'Shrubs', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Grassland', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Shrubland', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Lawn_grass', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Natural_Shrubs', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Grazing_land', 'LULC'] = 'Grass_Shrub_Land'
gdf.loc[gdf['LULC'] == 'Pasture_land', 'LULC'] = 'Grass_Shrub_Land'

#WetLand_Mangroves
gdf.loc[gdf['LULC'] == 'Mangroves', 'LULC'] = 'Wetland_Mangroves'
gdf.loc[gdf['LULC'] == 'Wetland', 'LULC'] = 'Wetland_Mangroves'
gdf.loc[gdf['LULC'] == 'Mudflats', 'LULC'] = 'Wetland_Mangroves'
#gdf.loc[gdf['LULC'] == 'Shallow_ocean', 'LULC'] = 'Wetland_Mangroves'

#Bare_Burnt_Land
gdf.loc[gdf['LULC'] == 'Bare_land', 'LULC'] = 'Bare_Burnt_Land'
gdf.loc[gdf['LULC'] == 'Burned_land', 'LULC'] = 'Bare_Burnt_Land'

#Roads
gdf.loc[gdf['LULC'] == 'Roads_paved', 'LULC'] = 'Roads'
gdf.loc[gdf['LULC'] == 'Roads_unpaved', 'LULC'] = 'Roads'

#Mining
gdf.loc[gdf['LULC'] == 'Rock', 'LULC'] = 'Mining'
gdf.loc[gdf['LULC'] == 'Quarry', 'LULC'] = 'Mining'
gdf.loc[gdf['LULC'] == 'Mining areas', 'LULC'] = 'Mining'
gdf.loc[gdf['LULC'] == 'Active_mining', 'LULC'] = 'Mining'
gdf.loc[gdf['LULC'] == 'Closed_abandoned_mining', 'LULC'] = 'Mining'
```

**Figure 9.** Example of use of GeoPandas for aggregating multiple diverse landcover classes into standardised classes for simplified final outputs (Singh, 2023).