
M3LEO: A Multi-Modal, Multi-Label Earth Observation Dataset Integrating Interferometric SAR and Multispectral Data

Matt Allen
University of Cambridge, UK
mja78@cam.ac.uk

Francisco Dorr
Independent, Argentina
fran.dorr@gmail.com

Joseph A. Gallego-Mejia
Drexel University, USA
jagallegom@unal.edu.co

Laura Martínez-Ferrer
Universitat de València, Spain
laura.martinez-ferrer@uv.es

Anna Jungbluth
European Space Agency, Climate Office, UK
anna.jungbluth@esa.int

Freddie Kalaitzis
University of Oxford, UK
freddie.kalaitzis@cs.ox.ac.uk

Raúl Ramos-Pollán
Universidad de Antioquia, Colombia
raul.ramos@udea.edu.co

Abstract

Satellite-based remote sensing has revolutionised the way we address global challenges in a rapidly evolving world. Huge quantities of Earth Observation (EO) data are generated by satellite sensors daily, but processing these large datasets for use in ML pipelines is technically and computationally challenging. Specifically, different types of EO data are often hosted on a variety of platforms, with differing degrees of availability for Python preprocessing tools. In addition, spatial alignment across data sources and data tiling for easier handling can present significant technical hurdles for novice users. While some preprocessed Earth observation datasets exist, their content is often limited to optical or near-optical wavelength data, which is ineffective at night or in adverse weather conditions. Synthetic Aperture Radar (SAR), an active sensing technique based on microwave length radiation, offers a viable alternative. However, the application of machine learning to SAR has been limited due to a lack of ML-ready data and pipelines, particularly for the full diversity of SAR data, including polarimetry, coherence and interferometry. In this work, we introduce M3LEO, a multi-modal, multi-label Earth observation dataset that includes polarimetric, interferometric, and coherence SAR data derived from Sentinel-1, alongside multispectral Sentinel-2 imagery and a suite of auxiliary data describing terrain properties such as land use. M3LEO spans approximately 17M data chips, each measuring 4x4 km, across six diverse geographic regions. The dataset is complemented by a flexible PyTorch Lightning framework, with configuration management using Hydra, to accommodate its use across diverse ML applications in Earth observation. Additionally, we provide tools to process any dataset available on popular platforms such as Google Earth Engine for seamless integration with our framework. We show that the distribution shift in self-supervised embeddings is substantial across geographic regions, even when controlling for terrain properties. Data is available at huggingface.co/M3LEO, and code at github.com/spaceml-org/M3LEO.

1 Introduction

Satellite-based Earth observation data is fundamental in addressing global problems in a rapidly changing world, offering large-scale, high resolution, high frequency data for applications from tracking wildfires [1] and deforestation [2] to refugee settlement mapping [3] and war zone damage assessment [4, 5]. Information from these tasks is critical in crafting responses to man-made [6] and environmental crises [7], but is constrained by the use of optical (wavelengths from visible to near-infrared) sensing data. Such sensors are unable to operate in adverse weather or cloudy conditions [8], or at night, limiting their usefulness for time-critical tasks such as natural disaster management [7], environmental protection [9] or maritime surveillance [10]. Synthetic Aperture Radar (SAR) sensing presents an alternative that is able to overcome these limitations.

Unlike optical sensors, SAR instruments actively illuminate terrain using microwave pulses, ensuring visibility without the need for daylight. These long-wavelength pulses can penetrate cloud cover and other adverse atmospheric conditions such as dust, making SAR-based sensing a valuable alternative to optical sensors for robust day-night coverage. Additionally, microwave radiation can penetrate some solid objects such as small-scale vegetation or soils — allowing, for example, measurements of properties relating to soil moisture under vegetation [11] or the identification of archaeological features hidden below ground [12]. In addition to exploiting the wavelength and active illumination of SAR data, the complex nature of SAR signals — returning both amplitude and phase — can also be leveraged to provide insights beyond what is possible using optical data. Coherence, the complex correlation between pairs of SAR acquisitions, has been used successfully for tasks including flood detection [13, 14], detection of urban damage [15] and forest canopy height measurement [16]. The phase difference between co-registered SAR acquisitions, measured through interferometry, enables the detection of surface height changes with millimetre accuracy, independently of the horizontal resolution of the sensor. This capability is critical for monitoring geological phenomena such as earthquakes [17], landslides [18], glacial movement [19, 20], magma chamber growth [21] and infrastructure deformation [22].

While SAR offers opportunities to overcome the limitations of optical sensors, and to give insights that are impossible to provide using visible wavelengths, it is associated with substantial additional complexity. The automated analysis of optical data, sometimes used in fusion with SAR data, has seen great success in recent years [23–25] — including the development of large foundation models able to make use of planetary-scale datasets [26, 27]. The application of large-scale deep learning to SAR data without simultaneous use of optical data, however, is more limited [28]. The complexities of processing SAR data, particularly estimating coherence and performing interferometry — which require processing phase information as using complex numbers — mean that the full diversity of SAR data types is not available at scale in formats compatible with machine learning (ML) pipelines.

To address these challenges we introduce M3LEO, a large-scale multi-modal Earth Observation dataset comprising polarimetric, interferometric and coherence data for SAR as well as Sentinel-2 data and auxiliary datasets describing surface properties such as land use, biomass measurements and elevation. We provide this data pre-processed as ML-readable, tiled images to abstract complex parameter choices typically made by domain experts. We also include a flexible PyTorch Lightning framework, parameterised using Hydra [29], to further minimise barriers to usage. We include a preliminary analysis on distribution shift for terrain properties and the appearance of SAR data across geographic regions. Finally, in addition to the pre-formatted data we offer for download, we provide tools enabling ML practitioners to process any data retrievable from Google Earth Engine into the same tiled format, such that it can be used in our framework.

2 Related Work

Deep learning has been applied over the last decade to curated optical imagery with great success [30–33], including the recent development of large, self-supervised foundation models [34–37]. Such models have been extraordinarily successful in tasks such as semantic segmentation [38, 39], image classification [33], [38] [40] and object detection [38, 41]. EO data from optical sensors has similarly been the subject of success for deep learning practitioners. Early work focusing on small, fully supervised models, showed great promise in a huge range of tasks, including land cover classification [23], biomass measurement [24], road detection [25, 42] and flood mapping [43], although many models were limited in scope to a small geographic area [24, 44–46]. More recent work has focused

on the development of large foundation models, often self supervised [28, 47], which are readily adaptable to a range of downstream tasks and geographic areas [27].

Work applying these models to optical EO data has been enabled by the wealth of easily accessible open data. As well as being available as raw products from satellite data providers such as ESA [48], many datasets comprising optical satellite imagery in ML-ready formats exist [49–52], across a range of spatial resolutions [53, 54], and for multiple time-points [50, 55], although they may be limited in other scopes — for example, not having aligned task labels [55, 56] or being limited to a single region [50]. We provide data at a large scale (14.1% of the land surface of the Earth) with a diverse set of auxiliary labels.

The application of deep learning to SAR data is less comprehensive. A body of work applying deep learning directly to SAR data exists [57–60] [61], but data limitations mean that geographic or temporal generalisability, often lacking in remote sensing models [62], has not clearly been shown. The development of foundation models for SAR may prove to be productive in obtaining geographic and temporal generalisability. To create foundation models, SAR is commonly applied alongside optical imagery in data fusion-based approaches [63, 64], but with little attribution regarding whether such approaches can work well without optical data. Some work exists exploiting schemes such as masked autoencoding [65, 66], contrastive learning [67], or knowledge distillation [68, 69] to develop foundation models for polarimetric SAR — and shows that strong geographic generalisability is obtainable when using SAR data at large scales [65, 69]. Many datasets providing ML-readable polarimetric SAR (polSAR) data exist [51, 55, 70–75], although most do not provide interferometric SAR (inSAR) data [74, 75].

The full diversity of inSAR datatypes, such as interferometry and coherence, have seen a number of applications in machine learning, and a rich tapestry of applications in other contexts. Interferometry, for example, is often used to track earthquakes [17], landslides [18] and glacial movement [19, 20], in a manner that is both more repeatable — being immune to adverse weather conditions — and more accurate — being able to track millimetre-scale height changes — than methods using optical data. Coherence has been used with success in urban damage assessment [15], flood detection [13, 14] and canopy height measurement [16].

A number of datasets exist making these datatypes available to deep learning users, many of which are focused on specific events, tasks or locations. Hephaestus [76], for example, contains 216K interferometric SAR (inSAR) patches localised to volcanoes annotated with various labels describing volcanic activity. ISSLIDE [77] contains inSAR data from the French alps describing slow landslides, and Pol-InSAR-Island [78] comprises inSAR data describing land cover on Baltrum, a Frisian island. UrbanSARFloods [79] contains Sentinel-1 interferometric coherence data from a diverse set of global locations, but is limited to specific events. S1SLC_CVDL [80] opts to provide complex-valued single-look SAR data rather than processed interferometric data, from three manually selected Sentinel-1 scenes containing major population centres. We make inSAR and coherence data available at a multi-continental scale, alongside both polarimetric SAR and optical data, in M3LEO.

2.1 Comparison to Existing Datasets

We provide a comparison between a number of popular large-scale Earth observation datasets, including M3LEO, in Table 1. We define *tile* to mean a fixed location or area on the surface of the Earth, and *chip* as the content of some data product over that tile. Unlike in many of these datasets, we provide acquisitions from different seasons within the same year for the same tile as channels, rather than separate chips. The described number of chips in Table 1 therefore appears relatively lower for M3LEO compared to, for example, SSL4EO-S12, for a fixed number of satellite acquisitions. We instead measure the number of timepoints in years, rounded up for part-years. Of the four datasets offering Sentinel-1 SAR data, only M3LEO offers data from multiple years, and only M3LEO offers inSAR data (although see Section 1 for a brief description of available task-specific or localised interferometric datasets).

Regarding spatial coverage, of the datasets listed in Table 1 M3LEO is most similar in scale to SatlasPretrain and SSL4EO-L, with these three datasets being significantly larger than the remainder. Of these three datasets, only M3LEO provides SAR data of any modality.

The temporal coverage of M3LEO sits between SatlasPretrain and SSL4EO-L, although the aggregate number of years does not give a full picture — in SatlasPretrain, some acquisitions are available for

a wider range of years for specific events, and in SSL4EO-L different data products are sometimes collected for non-overlapping years, meaning much of the dataset is not a parallel corpus. In M3LEO, the primary satellite data from Sentinel-1 and Sentinel-2 is available for 2018-2020 for all tiles.

Several of these datasets contain auxiliary information in addition to satellite acquisitions. Land cover labels are common, included in SSL4EO-L, SEN12MS, BigEarthNet and SatlasPretrain [51, 56, 72, 73]. Additional auxiliary datasets are sometimes available — SSL4EO-L, for example, includes more detailed crop classification data. SatlasPretrain introduced a number of novel additional labelled datasets including building and road polygons. M3LEO currently includes 4 auxiliary datasets - Land cover, vegetation cover, aboveground biomass and Digital Elevation Models (DEMs).

Many datasets are pre-sampled within their selected AOIs prior to distribution. Some datasets sample based on a manually specified distribution (for example, SSL4EO-S12 and SSL4EO-L sample locations based on Gaussian distributions centred on large cities), and some randomly. We include all available tiles within our AOIs — partly with a view to increasing data volume, but also to enable further research on sampling schemes. Selecting a good sampling strategy to diversify actively illuminated radar data with phase information is not straightforward, and it doubtful whether sampling schemes developed for optical EO imagery would transfer well to this data. The auxiliary data included in M3LEO, such as elevation and land use, may be useful for constructing such sampling schemes.

Table 1: **Existing Datasets.** Summary of popular large-scale Earth observation pre-training datasets.

Dataset	SAR	Years	Num. Tiles	Num. Chips*	Tile Size km	Coverage km ²	Sampling
SSL4EO-S12 [55]	Y	1	251K	3M	2.64×2.64	1.75×10 ⁶	Targeted
SSL4EO-L [51]	N	6	250K	5M	7.92×7.92	1.57×10 ⁷	Targeted
SEN12MS [70]	Y	1	181K	542K	2.56×2.56	1.18×10 ⁶	Random
SeCo [71]	N	2	200K	1M	2.65×2.65	1.40×10 ⁶	Random
BigEarthNet [72]	N	1	590K	1.2M	1.20×1.20	8.50×10 ⁵	Targeted
SatlasPretrain [73]	N	1**	856K	N/A	5.12×5.12	2.13×10 ⁷	None
M3LEO	Y	3	1.05M	17.2M	4.48×4.48	2.11×10 ⁷	None

* Heuristic only — some work, such as SSL4EO-S12, considers acquisitions at the same location from different seasons to be a separate data chip.

** Contains additional historical images from 2016-2021 that are relevant to dynamic events such as floods.

3 Dataset & Framework

3.1 SAR Datasets

The many benefits of SAR data are met with increased complexity compared to optical data. SAR sensors are active — that is, they emit microwave pulses (5.6 cm wavelength for Sentinel-1) and measure backscatter rather than imaging the Earth under passive illumination from the Sun. This enables day-night operation and the penetration of atmospheric obstructions such as cloud and dust. Unlike sunlight, the pulses emitted by SAR sensors are polarised, with the ability to emit pulses polarised either horizontally or vertically with respect to the Earth. SAR sensors are also able to measure the polarization of the backscatter, and capture both amplitude and phase. This signal, typically stored as a complex number, allows studying the geometry of surface-level objects, in addition to their reflectances. As an example, higher amplitudes are measured when surface features align with the polarisation of the emitted pulse. As with optical imagery, objects smaller than the measurement wavelength are invisible to the sensor.

The use of SAR data is further complicated by the use of side-looking radar. SAR sensors operate with the emitter and receiver aimed laterally, rather than vertically, as for optical sensors. Since radar operates by measuring the time of arrival for a backscattered signal, aiming the sensor vertically would make it impossible to distinguish between targets at an equal distance to the left or right of the direction of travel. This is corrected by side-looking, with the sensor aimed laterally such that the entire field of view is to one side of the satellite. Although side-looking corrects directional ambiguity, it necessitates complex post-processing to correct the resulting geometrical distortions and

radiance redistribution [81–83], and results in the same terrain being imaged differently depending on the direction of satellite travel, as the terrain is illuminated from the opposite side. We provide data in both ascending (northwards) and descending (southwards) satellite directions in M3LEO.

We provide three products derived from SAR data in this dataset — polarimetric amplitude, interferograms, and coherence. We give a brief background on each of these datatypes below, although we omit most technical details regarding their construction as they are beyond the scope of this work. References are provided for users who are interested in further background on these datatypes.

Amplitude Polarimetric amplitude measures the power of the backscattered signal received by the sensor. We provide amplitude data derived from ESA Sentinel-1 Level 1 Ground Range Detected SAR data (S1GRD), as available in Google Earth Engine (GEE)¹. Phase information is not provided via GEE and it is therefore impossible to produce further SAR datatypes from data available on the GEE platform. We provide data measuring vertically polarised and horizontally polarised returns from a vertically polarised emission (referred to as VV and VH respectively). Data is provided for imagery from both ascending and descending trajectories. We refer users to [84] for a more detailed breakdown of the theory behind polarimetric SAR data. S1GRD data is of 10 m/pixel resolution and provided for 2018-2020 as four seasonal averages per-year.

Interferometry Interferometry measures the phase difference between pairs of acquisitions over the same terrain. These phase differences provide data about small-scale displacements, which can be measured modulo the wavelength. Post-processed interferograms are *unwrapped* by computing accumulated modulo-wavelength displacements. As the scale at which these displacements can be measured depends on wavelength, rather than horizontal resolution, interferometric phase difference can be used to measure surface height changes with millimetre accuracy. We provide interferometric data computed using select pairs of Sentinel-1 acquisitions, from ASF ARIA Geocoded UNWrapped Interferogram data (GUNW), as available in ASF vertex²[85]. We include all available acquisition pairs with time deltas of less than 72 days. We refer users to [86] for a detailed treatment of the processing steps required to construct interferograms in addition to the underlying physics. GUNW data is provided for 2020 at a resolution of approximately 90 m/pixel.

Coherence The coherence of SAR imagery is calculated using the complex correlation between coincident pixels across two separate acquisitions. For a given complex valued pixel $z^{(i)}$ in SAR acquisitions at times 1 and 2, the coherence γ is defined as:

$$\gamma = \frac{|\mathbb{E}[z_1^{(i)}(t)z_2^{(i)}(t)]|}{\sqrt{\mathbb{E}[|z_1^{(i)}(t)|^2]\mathbb{E}[|z_2^{(i)}(t)|^2]}} \quad (1)$$

To provide meaningful coherence values, expectations are computed within a small spatial window surrounding each pixel. The resulting resolution of coherence maps is therefore lower than that of the original acquisitions. Man-made structures typically exhibit high coherence, as they are stable across acquisitions. Forests and other vegetation larger than the wavelength of the instrument have lower coherence. Coherence is affected in all cases by additional decorrelation factors not relating directly to terrain, such as doppler centroid difference or thermal noise. [87] and [88] provide more detailed treatments of coherence estimation. We provide coherence estimates from the Global Seasonal Sentinel-1 Interferometric Coherence (GSSIC) dataset [89], using date-pairs with time deltas of 12, 24, 36 and 48 days, with one date-pair per tile per season. GSSIC data is of approximately 90 m/pixel resolution and from 2020. A decay model is included with the GSSIC coherence data.

For both interferometry and coherence, the selection of acquisition pairs is critical. The acquisition pair selection process introduces a significant combinatorial challenge to managing SAR datasets — if every possible combination between all acquisitions were considered, the number of interferograms or coherence estimates would grow quadratically with the number of acquisitions. We pre-empt this issue by the provision of pre-selected date-pairs.

¹developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD

²asf.alaska.edu/datasets/daac/aria-sentinel-1-geocoded-unwrapped-interferograms/

Optical data We additionally provide data (S2SRM) sourced from the ESA Sentinel-2 mission³[48], as available in Google Earth Engine. Data is provided from the L2A product (surface reflectance). We do not include top-of-atmosphere L1C reflectances. We summarize this data as monthly means of cloud-free pixels. We include four monthly averages for 2018-2020 — March, June, September and December. We include all Sentinel-2 bands with a of 10 m/pixel (red, green, blue, NIR) or 20 m/pixel (vegetation red edge, SWIR 11/12).

3.2 Auxiliary Datasets

ESA World Cover Land cover classification labels (semantic segmentation) were obtained from the ESA World Cover product (ESAWC) [90], as available in Google Earth Engine⁴. The resolution of ESAWC is 10 m/pixel, and it comprises 11 classes (See Appendix E). The ESAWC product has been independently validated as having an accuracy of approximately 75% [91]. Data is provided for 2020.

ESA CCI Biomass A map of above ground biomass (AGB) in Mg ha⁻¹, derived from the ESA Climate Change Initiative’s Biomass product⁵ [92] is provided for 2020 at a resolution of 90 m/pixel. The relative error of this data is 20% for areas with a measured biomass exceeding 50 Mg ha⁻¹ [93]

MODIS Vegetation Cover Tree cover, non-tree cover and non-vegetated (bare) percentage labels derived from the Terra MODIS Vegetation Continuous Fields product (MODISVEG)⁶[94], are provided at a resolution of 250 m/pixel. A limited amount of independent validation reports the RMSE of the MODISVEG data as approximately 10% [94]. Our dataset includes yearly maps for 2016-2020.

GHS Built Surface Maps of built surface area (m²/pixel), derived from the Copernicus Global Human Settlement Built Surface (GHSBUILTS) product⁷[95] are provided at 100 m/pixel for 2020. The mean absolute error of this data has been evaluated using independent reference data has been estimated to be approximately 6% [96] (or 600 m²/pixel, at 100 m/pixel).

SRTM Digital Elevation Maps Digital Elevation Maps, derived from the NASA Shuttle Radar Topographic Mission (SRTM)⁸[97, 98] are provided at a resolution of 90 m/pixel. This data was measured in 2000, but we emphasise that terrain height changes relatively little at this resolution compared to other products such as ESAWC or GHSBUILTS. The RMSE of the SRTM data was originally reported as 16 m [99] although it has been measured as more accurate in some regions [100].

3.3 Data coverage

Our dataset covers six distinct geographic areas of interest (AOIs): the contiguous United States (CONUS), Europe, the Middle East, Pakistan and India (PAKIN), China, and South America. A visualisation is provided in Figure 1. We limit coverage to these regions for reasons relating to the acquisition parameters used by Sentinel-1.

Firstly, Sentinel-1 operates with different acquisition modes depending on the region. We choose to only include areas where Sentinel-1 uses the Interferometric Wide (IW) swath acquisition mode. In polar regions, Sentinel-1 employs the Extra Wide (EW) swath acquisition mode, which introduces systematic differences in how terrain is illuminated — such as the azimuth steering angle of the radar emitter being 0.8° for EW and 0.6° for IW acquisition. The polarisations used in polar regions are reversed (HH, HV vs. VV, VH) compared to other areas. We provide IW data with both VV and VH polarisations in the initial release of this dataset.

Secondly, much of the terrestrial surface of the Earth is only covered by a single direction of satellite travel. This includes much of North America, Africa, continental Asia, Oceania and the Amazon rainforest. Unlike passively illuminated data such as Sentinel-2, SAR actively illuminates terrain with

³developers.google.com/earth-engine/datasets/catalog/sentinel-2

⁴developers.google.com/earth-engine/datasets/catalog/ESA_WorldCover_v200

⁵gee-community-catalog.org/projects/cci_agb/

⁶developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD44B

⁷human-settlement.emergency.copernicus.eu/ghs_buS2023.php

⁸developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4

a radar emitter aimed laterally from one side of the satellite. Orbital direction therefore systematically determines whether terrain is illuminated from an easterly or westerly direction.

By focusing on regions with consistent acquisition parameters, we aim to provide a dataset that is more uniform and suitable for training models without introducing additional complexities. While including other regions might reduce geographic bias, it is not straightforward to address the potential systematic biases introduced by varying acquisition modes and polarisations, or the systematic lack of directional coverage in some regions. These issues warrant a detailed treatment beyond the scope of this work, and we refer readers to the Copernicus Wiki for a full set of details on Sentinel-1 coverage⁹ (particularly Figures 20, 21). Processing additional coverage is technically straightforward using the provided tools, should it be required. We include data for Europe despite its absence of GUNW coverage due to interest from data providers operating in Europe.

A total of 1,048,827 unique geographic tiles were generated, covering an area of $2.11 \times 10^7 \text{ km}^2$. A breakdown by AOI can be seen in Table 2. Tiling is uniform across all AOIs and datasets—the chips provided for each dataset cover exactly the same geographical areas. It is important to note that not all component datasets or specific parameterisations, such as date-pair ranges, were available for every geographic tile; however, the availability is still consistently high. See Appendix A for a full breakdown by dataset and AOI. We provide a reduced version of our dataset, **M3LEO-miniset**, spanning 5,000 tiles per AOI for rapid model iteration and use in tutorials.

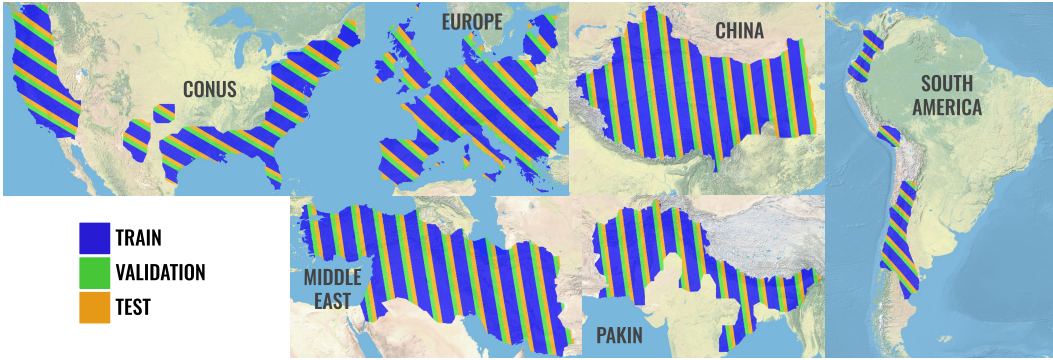


Figure 1: **Data splits.** Geographic bands for training, validation and test sets, at a ratio of 60:20:20.

Table 2: **Coverage statistics** summarised for each AOI in M3LEO.

AOI	Total No. Tiles	Area (km ²)	% Earth’s land surface
CONUS	167403	3.360×10^6	2.3%
Europe	200489	4.024×10^6	2.7%
Middle East	163986	3.291×10^6	2.2%
PAKIN	147791	2.966×10^6	2.0%
China	285402	5.728×10^6	3.7%
South America	83756	1.681×10^6	1.1%
Total	1048827	21.05×10^6	14.1%

Data splits We provide use geographic bands to define training, validation and test splits for use in the explorations found in this work, following [65]. Bands were split into training, validation and test sets at a ratio of 60:20:20, and can be seen visually in Figure 1. This method reduces distribution shifts and data leakage compared to single-band splits and fully randomized assignments, respectively. See Appendix A for details. Users should define train-test splits appropriate for their individual applications if their needs differ.

⁹sentinewiki.copernicus.eu/web/s1-mission

3.4 Framework

Data downloading & processing For each AOI, we provide a `.geojson` file containing the geographical extent and unique ID for each tile. These definitions are applied to each dataset at the point of tiling such that any chip with a given identifier spans precisely the same geographic extent of a chip corresponding to a different dataset of the same identifier. Throughout this work, we use the term *tile* to refer to a fixed area of the Earth’s surface defined in the definition files, and the term *chip* to refer to the data from a single component dataset, such as S1GRD or ESAWC, within the extent of a given tile.

To process datasets available via Google Earth Engine (GEE) (S2RGB, S1GRD, SRTM, ESAWC, MODISVEG), we introduce `geetiles`¹⁰. This tool extracts and tiles data from GEE as per definition files and configurations provided in the `geetiles` repository. The remaining datasets (GSSIC, GUNW, AGB, GHSBUILTS) were extracted using `sartiles`¹¹, which contains specific code to download and tile each of GUNW and GSSIC, as well as the facility to tile general GeoTIFF files, such as those provided for AGB and GHSBUILTS, for integration with M3LEO. Both `geetiles` and `sartiles` were developed alongside M3LEO, and are provided such that users are able to seamlessly integrate any data available via Google Earth Engine (or as a GeoTIFF file) with our framework.

Pipeline We accompany our dataset with a modular PyTorch Lightning [101] framework parameterised using Hydra [29]. We provide PyTorch Lightning datasets for each component of M3LEO. We also provide additional modules defining self-supervised approaches applied successfully to M3LEO in previous works (MAE [65], CLIP [67], DINO [68, 69]). Integrating custom models with M3LEO is straightforward, requiring the addition of a single file.

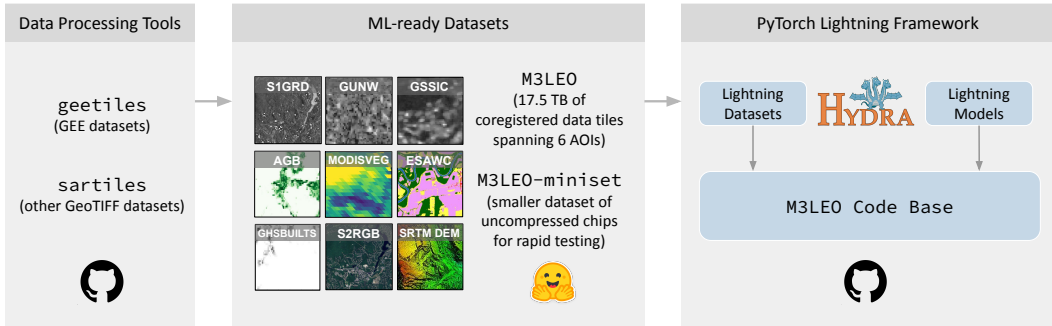


Figure 2: **M3LEO dataset and framework.** The M3LEO dataset consists of nine ML-ready component datasets and a PyTorch Lightning framework, parameterised by Hydra, for model training.

4 Analysis

We provide five auxiliary datasets (ESAWC, AGB, MODISVEG, GHSBUILTS, SRTM) in addition to SAR and optical satellite data acquisitions. Although these data could be used as labelled tasks (see Appendix D, and also [65, 67–69] for analysis of this type on M3LEO), they could equally be considered as providing information on important surface properties that would be non-trivial to derive directly from satellite data.

We compare the shift in the marginal distribution of these terrain properties y , $p(y)$ in Figure 3 for four of these auxiliary datasets. We computed the normalised L_1 distance between the discrete ESAWC distributions and the Wasserstein distance between the continuous GHSBUILTS, MODISVEG and SRTM distributions, across AOIs. The marginal distributions $p(y)$ for each auxiliary dataset are shown in Appendix B.

We also provide an early exploration of the ‘appearance shift’ of features between AOIs for S1GRD — the change in the distribution of embeddings x for a SAR tile with known terrain properties y , $P(x|y)$.

¹⁰github.com/rramosp/geetiles

¹¹github.com/rramosp/sartiles

To generate embeddings, we trained a masked autoencoder on S1GRD polarimetry from all six AOIs, following previous work on M3LEO [65]. Training hyperparameters can be seen in Appendix C. We applied max-pooling along the sequence dimension at the output of the ViT encoder and further reduced the dimension to 2 using UMAP [102]. We computed the expected Wasserstein distance between the conditional distributions $p(x|y)$ with respect to the distribution $p(y)$ on the test set, across AOIs and show the results in Figure 4. For ESAWC, we applied principal component analysis and conditioned on the first 3 principal components with 10 evenly spaced bins per dimension. On the remaining variables, we used 100 evenly spaced bins. We also display the sliced Wasserstein distance between the embedding distributions $p(x)$ in the leftmost matrix of Figure 4. We plot these reduced embeddings, coloured according to each of the terrain properties, in Figure 5. A checkpoint for the MAE model is available in the data repository.



Figure 3: **Distribution shifts.** Distribution shifts between terrain properties described by auxiliary datasets. Distribution shift for the ESAWC categorical data was quantified using L1 distance and continuous data using Wasserstein distance.

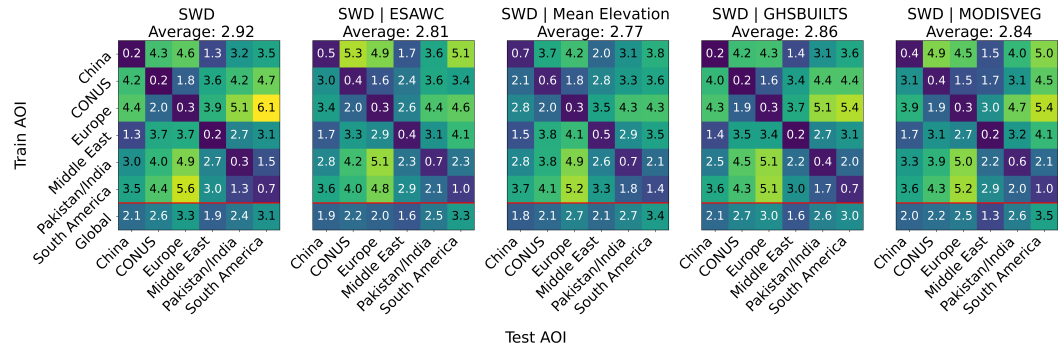


Figure 4: **Covariance and appearance shift.** (Leftmost) The Sliced Wasserstein Distances (SWD) between reduced train and test set embeddings across AOIs, generated using a masked autoencoder. (Right four) The expected value of same metric conditioned on each of four terrain properties. The expectation was computed with respect to the distribution of the terrain property on the test set. We conditioned on the first three principal components of the ESAWC distribution.

We observed that there was significant covariate shift in the distribution of the embeddings produced by the masked autoencoder, $p(x)$, across AOIs (Figure 4, leftmost matrix). Given that there was also significant shift in terrain properties described by the auxiliary datasets between AOIs (Figure 3, this is not immediately surprising). Another factor that must be considered, however, is the shift in the embeddings produced by the masked autoencoder for tiles with similar terrain properties y , $p(x|y)$, across AOIs (appearance shift). It can be seen in the four right-hand matrices of Figure 4 that although there is some reduction in the most extreme cases, the expected value of sliced Wasserstein distance conditioned on similar terrain properties is usually not substantially lower than the unconditioned case. Explained from an Earth observation perspective - the masked autoencoder does not extract overly similar features for two tiles with, for example, very high vegetation when those tiles are taken from different geographic regions. This effect can be seen visually in Figure 5 - embeddings with

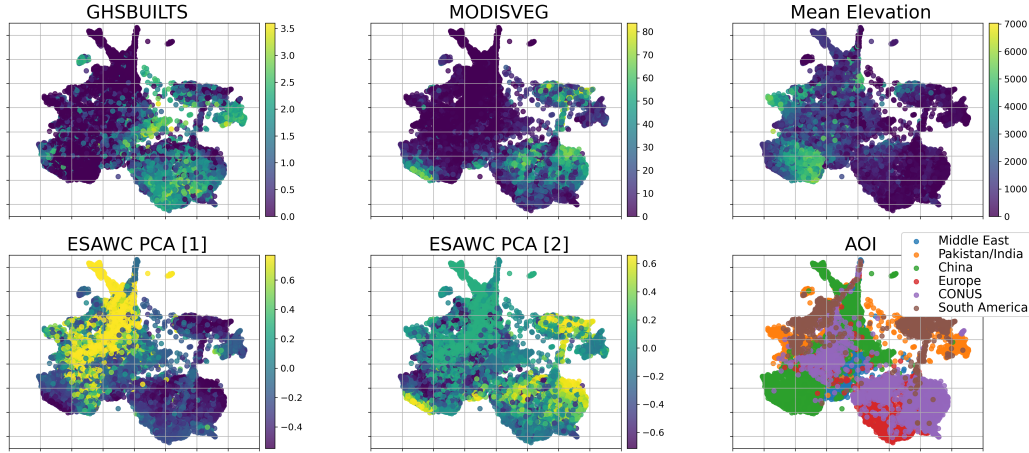


Figure 5: **Embedding scatter plots.** Scatter plots of 2D embeddings, reduced using UMAP, coloured according to different auxiliary datasets.

high values for particular labels do not cluster obviously across different AOIs. In contrast, previous work applying DINO-based self-supervision to M3LEO [69] found that embeddings with similar labels clustered tightly in the embedding space across AOIs. Despite this observation, previous work applying MAE-based pretraining to M3LEO has shown strong generalisation to novel AOIs [65].

We did not explore the use of GSSIC coherence data here, as this is substantially technically challenging and requires detailed treatment. Given the lower resolution of this data, it may not be productive to use as a naive input to deep learning models. A small set of experiments evidencing this is provided in Appendix D. We point readers to previous work using coherence data from M3LEO productively in a self supervised setting - in contrastive learning [67] or in knowledge distillation-based approaches [69]. We note that the provision of GUNW interferometric coherence data without reference to specific events such as floods or fires is unusual compared to other datasets [76–78]. Although we aim to include event-based datasets in a future update, this large-scale dataset of interferograms is still highly desirable in self-supervised schemes.

A number of unknowns remain regarding domain shift in M3LEO. Many auxiliary datasets, such as ESAWC, do not exist for 2018 and 2019 so it is difficult to provide a substantial exploration of temporal shifts, although we provide S1GRD and S2SRM for three years. It is unclear whether features learned from encoders trained on different polarizations or orbital directions are comparable, although D contains a limited set of experiments on the value-add of different polarisations.

5 Conclusions and Future Work

In this work, we introduced M3LEO, a multi-continental Earth observation dataset including a comprehensive set of SAR data, alongside digital elevation models, RGB data and a suite of downstream tasks. To the best of our knowledge, this is the largest ML-readable polSAR dataset by total number of tiles and geographic coverage, and the largest inSAR dataset by the same metrics. We additionally provide a modular PyTorch Lightning framework to enable the application of deep learning and encourage the uptake of these datatypes. We provide additional tools, `geetiles` and `sartiles`, to enable the integration of any data available in Google Earth Engine with our framework.

We trained an MAE-based model on polSAR data and conducted a small exploration on the appearance shift of features corresponding to similar labels across AOIs. Despite the fact that this type of training has previously been shown to generalise well geographically [65], the shift in low-level features useful for the reconstruction pretext task was substantial between geographic regions. This is in contrast to previous work using M3LEO that found embeddings from DINO-based models with similar labels from different AOIs clustered tightly.

6 Acknowledgements

This work has been enabled by Frontier Development Lab Europe (<https://fdleurope.org>) a public / private partnership between the European Space Agency (ESA), Trillium Technologies, the University of Oxford and leaders in commercial AI supported by Google Cloud and Nvidia, developing open science for all Humankind. L.M-F. was supported by the European Research Council (ERC) Synergy Grant “Understanding and Modelling the Earth System with Machine Learning (USMILE)” under the Horizon 2020 research and innovation programme (Grant agreement No. 855187). M. J. A. was supported by the UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks [EP/S022961/1]. We are also indebted to Nicolas Longépé, Carlos López-Martínez, Fabio A. González Osorio, Samuel Bancroft, Emma Hatton, Alison Lowndes, Alistair Francis, Ioanna Bouri and the rest of reviewers during the 2023 FDL-Europe sprint.

References

- [1] Xikun Hu, Yifang Ban, and Andrea Nascetti. Sentinel-2 MSI Data for Active Fire Detection in Major Fire-Prone Biomes: A Multi-Criteria Approach. *International Journal of Applied Earth Observation and Geoinformation*, 101:102347, September 2021. ISSN 1569-8432. doi: 10.1016/j.jag.2021.102347.
- [2] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853, November 2013. doi: 10.1126/science.1244693.
- [3] John A. Quinn, Marguerite M. Nyhan, Celia Navarro, Davide Coluccia, Lars Bromley, and Miguel Luengo-Oroz. Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128):20170363, August 2018. doi: 10.1098/rsta.2017.0363.
- [4] Yusupujiang Aimaiti, Christina Sanon, Magaly Koch, Laurie G. Baise, and Babak Moaveni. War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images. *Remote Sensing*, 14(24):6239, January 2022. ISSN 2072-4292. doi: 10.3390/rs14246239.
- [5] Nataliia Kussul, Sofiia Drozd, Hanna Yailymova, Andrii Shelestov, Guido Lemoine, and Klaus Deininger. Assessing damage to agricultural fields from military actions in Ukraine: An integrated approach using statistical indicators and machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 125:103562, December 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103562.
- [6] Leonard Stoeckl, Vanessa Banks, Stella Shekhunova, and Yevgeniy Yakovlev. The hydro-geological situation after salt-mine collapses at Sotolvyno, Ukraine. *Journal of Hydrology: Regional Studies*, 30:100701, August 2020. ISSN 2214-5818. doi: 10.1016/j.ejrh.2020.100701.
- [7] William D. Barnhart, Gavin P. Hayes, and David J. Wald. Global Earthquake Response with Imaging Geodesy: Recent Examples from the USGS NEIC. *Remote Sensing*, 11(11):1357, January 2019. ISSN 2072-4292. doi: 10.3390/rs11111357.
- [8] Rui Jiang, Arturo Sanchez-Azofeifa, Kati Laakso, Yan Xu, Zhiyan Zhou, Xiwen Luo, Junhao Huang, Xin Chen, and Yu Zang. Cloud Cover throughout All the Paddy Rice Fields in Guangdong, China: Impacts on Sentinel 2 MSI and Landsat 8 OLI Optical Observations. *Remote Sensing*, 13(15):2961, January 2021. ISSN 2072-4292. doi: 10.3390/rs13152961.
- [9] Ray Purdy. Using Earth Observation Technologies for Better Regulatory Compliance and Enforcement of Environmental Laws. *Journal of Environmental Law*, 22(1):59–87, January 2010. ISSN 0952-8873. doi: 10.1093/jel/eqp027.

- [10] Giovanni Soldi, Domenico Gaglione, Nicola Forti, Alessio Di Simone, Filippo Cristian Daffinà, Gianfausto Bottini, Dino Quattrociochi, Leonardo M. Millefiori, Paolo Braca, Sandro Carniel, Peter Willett, Antonio Iodice, Daniele Riccio, and Alfonso Farina. Space-Based Global Maritime Surveillance. Part I: Satellite Technologies. *IEEE Aerospace and Electronic Systems Magazine*, 36(9):8–28, September 2021. ISSN 1557-959X. doi: 10.1109/MAES.2021.3070862.
- [11] Aliihsan Sekertekin, Aycan Murat Marangoz, and Saygin Abdikan. ALOS-2 and Sentinel-1 SAR data sensitivity analysis to surface soil moisture over bare and vegetated agricultural fields. *Computers and Electronics in Agriculture*, 171:105303, April 2020. ISSN 0168-1699. doi: 10.1016/j.compag.2020.105303.
- [12] Ahmed Gaber, Magaly Koch, M. Helmi Griesch, Motoyuki Sato, and Farouk El-Baz. Near-surface imaging of a buried foundation in the Western Desert, Egypt, using space-borne and ground penetrating radar. *Journal of Archaeological Science*, 40(4):1946–1955, April 2013. ISSN 0305-4403. doi: 10.1016/j.jas.2012.12.019.
- [13] G. Nico, M. Pappalepore, G. Pasquariello, A. Refice, and S. Samarelli. Comparison of SAR amplitude vs. coherence flood detection methods - a GIS application. *International Journal of Remote Sensing*, 21(8):1619–1631, January 2000. ISSN 0143-1161. doi: 10.1080/014311600209931.
- [14] Marco Chini, Asterios Papastergios, Luca Pulvirenti, Nazzareno Pierdicca, Patrick Matgen, and Issaak Parcharidis. SAR coherence and polarimetric information for improving flood mapping. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 7577–7580, July 2016. doi: 10.1109/IGARSS.2016.7730976.
- [15] Manabu Watanabe, Rajesh Bahadur Thapa, Tsuneo Ohsumi, Hiroyuki Fujiwara, Chinatsu Yonezawa, Naoya Tomii, and Sinichi Suzuki. Detection of damaged urban areas using interferometric SAR coherence change with PALSAR-2. *Earth, Planets and Space*, 68(1):131, July 2016. ISSN 1880-5981. doi: 10.1186/s40623-016-0513-2.
- [16] Aire Olesk, Jaan Praks, Oleg Antropov, Karlis Zalite, Tauri Arumäe, and Kaupo Voormansik. Interferometric SAR Coherence Models for Characterization of Hemiboreal Forests Using TanDEM-X Data. *Remote Sensing*, 8(9):700, September 2016. ISSN 2072-4292. doi: 10.3390/rs8090700.
- [17] Eric Jameson Fielding, Zhen Liu, Oliver L. Stephenson, Minyan Zhong, Curren Liang, Angelyn Moore, Sang-Ho Yun, and Mark Simons. Surface Deformation Related to the 2019 Mw 7.1 and 6.4 Ridgecrest Earthquakes in California from GPS, SAR Interferometry, and SAR Pixel Offsets. *Seismological Research Letters*, 91(4):2035–2046, July 2020. ISSN 0895-0695, 1938-2057. doi: 10.1785/0220190302.
- [18] Tazio Strozzi, Jan Klimeš, Holger Frey, Rafael Caduff, Christian Huggel, Urs Wegmüller, and Alejo Cochachin Rapre. Satellite SAR interferometry for the improved assessment of the state of activity of landslides: A case study from the Cordilleras of Peru. *Remote Sensing of Environment*, 217:111–125, November 2018. ISSN 0034-4257. doi: 10.1016/j.rse.2018.08.014.
- [19] V. Kumar, G. Venkataramana, and K. A. Høgda. Glacier surface velocity estimation using SAR interferometry technique applying ascending and descending passes in Himalayas. *International Journal of Applied Earth Observation and Geoinformation*, 13(4):545–551, August 2011. ISSN 1569-8432. doi: 10.1016/j.jag.2011.02.004.
- [20] Anirudha Mahagaonkar, Praveen K. Thakur, and Ling Chang. Assessment Of Sentinel-1 Products For Revealing Glacier Surface Movement In Indian Himalayas Using Differential Sar Interferometry. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 2070–2073, July 2019. doi: 10.1109/IGARSS.2019.8898831.
- [21] Delphine Smittarello, Valérie Cayol, Virginie Pinel, Jean-Luc Froger, Aline Peltier, and Quentin Dumont. Combining InSAR and GNSS to Track Magma Transport at Basaltic Volcanoes. *Remote Sensing*, 11(19):2236, January 2019. ISSN 2072-4292. doi: 10.3390/rs11192236.

- [22] Maral Bayarara, Cristian Rossi, Freddie Kalaitzis, and Brian Sheil. Entity Embeddings in Remote Sensing: Application to Deformation Monitoring for Infrastructure. *Remote Sensing*, 15(20):4910, January 2023. ISSN 2072-4292. doi: 10.3390/rs15204910.
- [23] Darius Phiri, Matamy Simwanda, Serajis Salekin, Vincent R. Nyirenda, Yuji Murayama, and Manjula Ranagalage. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*, 12(14):2291, January 2020. ISSN 2072-4292. doi: 10.3390/rs12142291.
- [24] Luofan Dong, Huaqiang Du, Ning Han, Xuejian Li, Di'en Zhu, Fangjie Mao, Meng Zhang, Junlong Zheng, Hua Liu, Zihao Huang, and Shaobai He. Application of Convolutional Neural Network on Lei Bamboo Above-Ground-Biomass (AGB) Estimation Using Worldview-2. *Remote Sensing*, 12(6):958, January 2020. ISSN 2072-4292. doi: 10.3390/rs12060958.
- [25] Christian Ayala, Rubén Sesma, Carlos Aranda, and Mikel Galar. A Deep Learning Approach to an Enhanced Building Footprint and Road Detection in High-Resolution Satellite Imagery. *Remote Sensing*, 13(16):3135, January 2021. ISSN 2072-4292. doi: 10.3390/rs13163135.
- [26] Michael J. Smith, Luke Fleming, and James E. Geach. EarthPT: A time series foundation model for Earth Observation, January 2024.
- [27] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, and Yansheng Li. SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery, March 2024.
- [28] Yi Wang, Conrad M. Albrecht, Nassim Ait Ali Braham, Lichao Mou, and Xiao Xiang Zhu. Self-Supervised Learning in Remote Sensing: A review. *IEEE Geoscience and Remote Sensing Magazine*, 10(4):213–247, December 2022. ISSN 2168-6831. doi: 10.1109/MGRS.2022.3198244.
- [29] Omry Yadan. Hydra - A framework for elegantly configuring complex applications, 2019.
- [30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [32] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.
- [33] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, June 2021.
- [34] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision, February 2021.
- [35] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners, December 2021.
- [36] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

- [37] Johannes Jakubik, Sujit Roy, C. E. Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarzman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, Daiki Kimura, Naomi Simumba, Linsong Chu, S. Karthik Mukkavilli, Devyani Lambhate, Kamal Das, Ranjini Bangalore, Dario Oliveira, Michal Muszynski, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Sam Khallaghi, Hanxi, Li, Michael Cecil, Maryam Ahmadi, Fatemeh Kordi, Hamed Alemohammad, Manil Maskey, Raghu Ganti, Kommy Weldemariam, and Rahul Ramachandran. *Foundation Models for Generalist Geospatial Artificial Intelligence*, November 2023.
- [38] Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, and Furu Wei. *Image as a Foreign Language: BEiT Pretraining for All Vision and Vision-Language Tasks*, August 2022.
- [39] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip H. S. Torr, and Li Zhang. *Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers*, July 2021.
- [40] Wenhui Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, and Yu Qiao. *InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions*, April 2023.
- [41] Ross Girshick. *Fast R-CNN*, September 2015.
- [42] C. Ayala, C. Aranda, and M. Galar. *Towards Fine-Grained Road Maps Extraction Using Sentinel-2 Imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2021:9–14, June 2021. ISSN 2194-9042. doi: 10.5194/isprs-annals-V-3-2021-9-2021.
- [43] Carmela Cavallo, Maria Nicolina Papa, Massimiliano Gargiulo, Guillermo Palau-Salvador, Paolo Vezza, and Giuseppe Ruello. *Continuous Monitoring of the Flooding Dynamics in the Albufera Wetland (Spain) by Landsat-8 and Sentinel-2 Datasets. Remote Sensing*, 13(17):3525, January 2021. ISSN 2072-4292. doi: 10.3390/rs13173525.
- [44] Jonas Botelho, Stefany C. P. Costa, Júlia G. Ribeiro, and Carlos M. Souza. *Mapping Roads in the Brazilian Amazon with Artificial Intelligence and Sentinel-2. Remote Sensing*, 14(15): 3625, January 2022. ISSN 2072-4292. doi: 10.3390/rs14153625.
- [45] Heikki Astola, Lauri Seitsonen, Eelis Halme, Matthieu Molinier, and Anne Lönnqvist. *Deep Neural Networks with Transfer Learning for Forest Variable Estimation Using Sentinel-2 Imagery in Boreal Forest. Remote Sensing*, 13(12):2392, January 2021. ISSN 2072-4292. doi: 10.3390/rs13122392.
- [46] Yisa Ginath Yuh, Wiktor Tracz, H. Damon Matthews, and Sarah E. Turner. *Application of machine learning approaches for land cover monitoring in northern Cameroon. Ecological Informatics*, 74:101955, May 2023. ISSN 1574-9541. doi: 10.1016/j.ecoinf.2022.101955.
- [47] Chao Tao, Ji Qi, Mingning Guo, Qing Zhu, and Haifeng Li. *Self-supervised remote sensing feature learning: Learning Paradigms, Challenges, and Future Works. IEEE Transactions on Geoscience and Remote Sensing*, 61:1–26, 2023. ISSN 0196-2892, 1558-0644. doi: 10.1109/TGRS.2023.3276853.
- [48] European Space Agency. *Sentinel-2 MSI Level-2A BOA Reflectance*, 2022.
- [49] Alexandre Lacoste, Nils Lehmann, Pau Rodriguez, Evan Sherwin, Hannah Kerner, Björn Lütjens, Jeremy Irvin, David Dao, Hamed Alemohammad, Alexandre Drouin, Mehmet Gunturkun, Gabriel Huang, David Vazquez, Dava Newman, Yoshua Bengio, Stefano Ermon, and Xiaoxiang Zhu. *GEO-Bench: Toward Foundation Models for Earth Monitoring. Advances in Neural Information Processing Systems*, 36:51080–51093, December 2023.
- [50] Dimitrios Sykas, Maria Sdraka, Dimitrios Zografakis, and Ioannis Papoutsis. *A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation With Deep Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:3323–3339, 2022. ISSN 2151-1535. doi: 10.1109/JSTARS.2022.3164771.

- [51] Adam J. Stewart, Nils Lehmann, Isaac A. Corley, Yi Wang, Yi-Chia Chang, Nassim Ait Ali Braham, Shradha Sehgal, Caleb Robinson, and Arindam Banerjee. SSL4EO-L: Datasets and Foundation Models for Landsat Imagery, October 2023.
- [52] Alistair Francis and Mikolaj Czerkawski. Major TOM: Expandable Datasets for Earth Observation, February 2024.
- [53] Julien Cornebise, Ivan Oršolić, and Freddie Kalaitzis. Open High-Resolution Satellite Imagery: The WorldStrat Dataset – With Application to Super-Resolution. *Advances in Neural Information Processing Systems*, 35:25979–25991, December 2022.
- [54] Gabriel Henrique de Almeida Pereira, Andre Minoro Fusioka, Bogdan Tomoyuki Nassu, and Rodrigo Minetto. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178:171–186, August 2021. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2021.06.002.
- [55] Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M. Albrecht, and Xiao Xiang Zhu. SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation [Software and Data Sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, September 2023. ISSN 2168-6831. doi: 10.1109/MGRS.2023.3281651.
- [56] Michael Schmitt, Lloyd Haydn Hughes, and Xiao Xiang Zhu. The SEN1-2 Dataset for Deep Learning in SAR-Optical Data Fusion, July 2018.
- [57] BINAYAK GHOSH, Shagun Garg, and M. Motagh. Automatic flood detection from Sentinel-1 data using deep learning architectures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:201–208, 2022.
- [58] Mohamed Shaban, Reem Salim, Hadil Abu Khalifeh, Adel Khelifi, Ahmed Shalaby, Shady El-Mashad, Ali Mahmoud, Mohammed Ghazal, and Ayman El-Baz. A Deep-Learning Framework for the Detection of Oil Spills from SAR Data. *Sensors*, 21(7):2351, January 2021. ISSN 1424-8220. doi: 10.3390/s21072351.
- [59] Lorenzo Nava, Oriol Monserrat, and Filippo Catani. Improving Landslide Detection on SAR Data Through Deep Learning. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. ISSN 1558-0571. doi: 10.1109/LGRS.2021.3127073.
- [60] Hemani Parikh, Samir Patel, and Vibha Patel. Classification of SAR and PolSAR images using deep learning: A review. *International Journal of Image and Data Fusion*, 11(1):1–32, January 2020. ISSN 1947-9832. doi: 10.1080/19479832.2019.1655489.
- [61] Vanessa Boehm, Wei Ji Leong, Ragini Bal Mahesh, Ioannis Prapas, Edoardo Nemmi, Freddie Kalaitzis, Siddha Ganju, and Raul Ramos-Pollan. Deep Learning for Rapid Landslide Detection using Synthetic Aperture Radar (SAR) Datacubes, November 2022.
- [62] Anastasiia Safonova, Gohar Ghazaryan, Stefan Stiller, Magdalena Main-Knorn, Claas Nendel, and Masahiro Ryo. Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 125:103569, December 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103569.
- [63] Yuxing Chen and Lorenzo Bruzzone. Self-Supervised SAR-Optical Data Fusion of Sentinel-1/2 Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. ISSN 1558-0644. doi: 10.1109/TGRS.2021.3128072.
- [64] Xian Sun, Peijin Wang, Wanxuan Lu, Zicong Zhu, Xiaonan Lu, Qibin He, Junxi Li, Xue Rong, Zhujun Yang, Hao Chang, Qinglin He, Guang Yang, Ruiping Wang, Jiwen Lu, and Kun Fu. RingMo: A Remote Sensing Foundation Model With Masked Image Modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–22, 2023. ISSN 1558-0644. doi: 10.1109/TGRS.2022.3194732.
- [65] Matt Allen, Francisco Dorr, Joseph A. Gallego-Mejia, Laura Martínez-Ferrer, Anna Jungbluth, Freddie Kalaitzis, and Raúl Ramos-Pollán. Large Scale Masked Autoencoding for Reducing Label Requirements on SAR Data, December 2023.

- [66] Hugo Chan-To-Hing and Bharadwaj Veeravalli. Fus-MAE: A cross-attention-based data fusion approach for Masked Autoencoders in remote sensing, January 2024.
- [67] Matt Allen, Francisco Dorr, Joseph A. Gallego-Mejia, Laura Martínez-Ferrer, Anna Jungbluth, Freddie Kalaitzis, and Raúl Ramos-Pollán. Fewshot learning on global multimodal embeddings for earth observation tasks, December 2023.
- [68] Joseph A. Gallego-Mejia, Anna Jungbluth, Laura Martínez-Ferrer, Matt Allen, Francisco Dorr, Freddie Kalaitzis, and Raúl Ramos-Pollán. Exploring DINO: Emergent Properties and Limitations for Synthetic Aperture Radar Imagery, December 2023.
- [69] Laura Martínez-Ferrer, Anna Jungbluth, Joseph A. Gallego-Mejia, Matt Allen, Francisco Dorr, Freddie Kalaitzis, and Raúl Ramos-Pollán. Exploring Generalisability of Self-Distillation with No Labels for SAR-Based Vegetation Prediction, December 2023.
- [70] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu. SEN12MS - a Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2-W7: 153–160, September 2019. ISSN 2194-9042. doi: 10.5194/isprs-annals-IV-2-W7-153-2019.
- [71] Oscar Mañas, Alexandre Lacoste, Xavier Giro-i-Nieto, David Vazquez, and Pau Rodriguez. Seasonal Contrast: Unsupervised Pre-Training from Uncurated Remote Sensing Data, May 2021.
- [72] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl. BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding, June 2019.
- [73] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding, August 2023.
- [74] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211, 2020.
- [75] Xiaoning He, Shuangcheng Zhang, Bowei Xue, Tong Zhao, and Tong Wu. Cross-modal change detection flood extraction based on convolutional neural network. *International Journal of Applied Earth Observation and Geoinformation*, 117:103197, March 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103197.
- [76] Nikolaos Ioannis Bountos, Ioannis Papoutsis, Dimitrios Michail, Andreas Karavias, Panagiotis Elias, and Isaak Parcharidis. Hephaestus: A large scale multitask dataset towards InSAR understanding, April 2022.
- [77] Antoine Bralet, Emmanuel Trouvé, Jocelyn Chanussot, and Abdourrahmane M. Atto. ISSLIDE: A New InSAR Dataset for Slow SLiding Area DEtection With Machine Learning. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024. ISSN 1558-0571. doi: 10.1109/LGRS.2024.3365299.
- [78] Sylvia Hochstuhl, Niklas Pfeffer, Antje Thiele, Stefan Hinz, Joel Amao-Oliva, Rolf Scheiber, Andreas Reigber, and Holger Dirks. Pol-InSAR-Island - A benchmark dataset for multi-frequency Pol-InSAR data land cover classification. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 10:100047, December 2023. ISSN 26673932. doi: 10.1016/j.ophoto.2023.100047.
- [79] Jie Zhao, Zhitong Xiong, and Xiao Xiang Zhu. UrbanSARFloods: Sentinel-1 SLC-Based Benchmark Dataset for Urban and Open-Area Flood Mapping, June 2024.
- [80] Reza Mohammadi Asiyabi, Mihai Datcu, Andrei Anghel, and Holger Nies. Complex-Valued End-to-End Deep Network With Coherency Preservation for Complex-Valued SAR Data Reconstruction and Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–17, 2023. ISSN 1558-0644. doi: 10.1109/TGRS.2023.3267185.

- [81] B. N. Koopmans. Side-looking radar, a tool for geological surveys. *Remote Sensing Reviews*, 1(1):19–69, June 1983. ISSN 0275-7257. doi: 10.1080/02757258309532063.
- [82] PETER HOOGEBOOM. Preprocessing of side-looking airborne radar data†. *International Journal of Remote Sensing*, 4(3):631–637, January 1983. ISSN 0143-1161. doi: 10.1080/01431168308948579.
- [83] Anup Das, Ritesh Agrawal, and Shiv Mohan. Topographic correction of ALOS-PALSAR images using InSAR-derived DEM. *Geocarto International*, 30(2):145–153, February 2015. ISSN 1010-6049. doi: 10.1080/10106049.2014.883436.
- [84] Yoshio Yamaguchi. *Polarimetric SAR Imaging: Theory and Applications*. CRC Press, Boca Raton, August 2020. ISBN 978-1-00-304975-3. doi: 10.1201/9781003049753.
- [85] Brett Buzzanga, David P. S. Bekaert, Ben D. Hamlington, and Simran S. Sangha. Toward Sustained Monitoring of Subsidence at the Coast Using InSAR and GPS: An Application in Hampton Roads, Virginia. *Geophysical Research Letters*, 47(18):e2020GL090013, 2020. ISSN 1944-8007. doi: 10.1029/2020GL090013.
- [86] Mark A. Richards. A Beginner’s Guide to Interferometric SAR Concepts and Signal Processing [AESS Tutorial IV]. *IEEE Aerospace and Electronic Systems Magazine*, 22(9):5–29, September 2007. ISSN 1557-959X. doi: 10.1109/MAES.2007.4350281.
- [87] Zhang Yanjie and V. Prinet. InSAR coherence estimation. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 3353–3355 vol.5, September 2004. doi: 10.1109/IGARSS.2004.1370422.
- [88] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1):6–43, March 2013. ISSN 2168-6831. doi: 10.1109/MGRS.2013.2248301.
- [89] Josef Kelldorfer, Oliver Cartus, Marco Lavallo, Christophe Magnard, Pietro Milillo, Shadi Oveisgharan, Batu Osmanoglu, Paul A. Rosen, and Urs Wegmüller. Global seasonal Sentinel-1 interferometric coherence and backscatter data set. *Scientific Data*, 9(1):73, March 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01189-6.
- [90] Daniele Zanaga, Ruben Van De Kerchove, Wanda De Keersmaecker, Niels Souverijns, Carsten Brockmann, Ralf Quast, Jan Wevers, Alex Grosu, Audrey Paccini, Sylvain Vergnaud, Oliver Cartus, Maurizio Santoro, Steffen Fritz, Ivelina Georgieva, Myroslava Lesiv, Sarah Carter, Martin Herold, Linlin Li, Nandin-Erdene Tsendbazar, Fabrizio Ramoino, and Olivier Arino. ESA WorldCover 10 m 2020 v100, October 2021.
- [91] ESA WorldCover 2020. <https://worldcover2020.esa.int/>, 2020.
- [92] Maurizio Santoro and Oliver Cartus. ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010, 2017, 2018, 2019 and 2020, v4, 2023.
- [93] Biomass. <https://climate.esa.int/en/projects/biomass/>, 2020.
- [94] Charlene DiMiceli, Mark Carroll, Robert Sohlberg, Do-Hyung Kim, Maggi Kelly, and John Townshend. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006, 2015.
- [95] Martino Pesaresi. GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030), April 2023.
- [96] Martino Pesaresi, Marcello Schiavina, Panagiotis Politis, Sergio Freire, Katarzyna Krasnodębska, Johannes H. Uhl, Alessandra Carioli, Christina Corbane, Lewis Dijkstra, Pietro Florio, Hannah K. Friedrich, Jing Gao, Stefan Leyk, Linlin Lu, Luca Maffenini, Ines Mari-Rivero, Michele Melchiorri, Vasileios Syrris, Jamon Van Den Hoek, and Thomas Kemper. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, 17(1):2390454, December 2024. ISSN 1753-8947. doi: 10.1080/17538947.2024.2390454.

- [97] OpenTopography. Shuttle Radar Topography Mission (SRTM) Global, 2013.
- [98] A. Jarvis, E. Guevara, H. I. Reuter, and A. D. Nelson. Hole-filled SRTM for the globe : Version 4 : Data grid, 2008.
- [99] Tom G. Farr, Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas Alsdorf. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2), 2007. ISSN 1944-9208. doi: 10.1029/2005RG000183.
- [100] Manas Mukul, Vinee Srivastava, Sridevi Jade, and Malay Mukul. Uncertainties in the Shuttle Radar Topography Mission (SRTM) Heights: Insights from the Indian Himalaya and Peninsula. *Scientific Reports*, 7:41672, February 2017. doi: 10.1038/srep41672.
- [101] William Falcon, Jirka Borovec, Adrian Wälchli, Nic Eggert, Justus Schock, Jeremy Jordan, Nicki Skafte, Ir1dXD, Vadim Bereznyuk, Ethan Harris, Tullie Murrell, Peter Yu, Sebastian Præsius, Travis Addair, Jacob Zhong, Dmitry Lipin, So Uchida, Shreyas Bapat, Hendrik Schröter, Boris Dayma, Alexey Karnachev, Akshay Kulkarni, Shunta Komatsu, Martin.B, Jean-Baptiste SCHIRATTI, Hadrien Mary, Donal Byrne, Cristobal Eyzaguirre, Cinjon, and Anton Bakhtin. PyTorchLightning/pytorch-lightning: 0.7.6 release. Zenodo, May 2020.
- [102] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29):861, September 2018. ISSN 2475-9066. doi: 10.21105/joss.00861.
- [103] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation, May 2015.
- [104] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017.

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [Yes]
 - (b) Did you describe the limitations of your work? [Yes] See Appendix F.
 - (c) Did you discuss any potential negative societal impacts of your work? [N/A]
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
 - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Our code is available at <https://github.com/spaceml-org/M3LEO> and data at <https://huggingface.co/M3LEO> and <https://huggingface.co/M3LEO-miniset>
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] We intend only for our results to demonstrate the utility of our framework and data rather than being a baseline for direct comparison. Additionally, computational cost was too high for enough repeat runs to produce meaningful error bars.
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Appendix D.3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 3.
 - (b) Did you mention the license of the assets? [Yes] See Appendix G.
 - (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] See <https://github.com/spaceml-org/M3LEO>.
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Dataset Summary

A.1 Chip counts

Table A.1: **Summary of component datasets** in M3LEO, including number of chips and dataset size for each region (per year). Totals at the bottom are adjusted for multi-year datasets.

Input Datasets	CONUS		Europe		China	
	Chips	Size/GB	Chips	Size/GB	Chips	Size/GB
S1GRD (2018-2020)	167403	1003	200489	1228	285402	1740
GSSIC (2020)	167403	73	200489	104	285399	125
GUNW (2020)	554844	579	N/A*	N/A*	1027451	854
S2SRM (2018-2020)	167406	1228	200489	1433	285402	1945
ESAWC (2020)	167406	32	200489	39	285402	55
AGB (2018-2020)	167406	8.5	200489	11	285402	14
MODISVEG (2020)	167406	151	200489	189	285402	220
GHSBUILTS (2020)	167406	0.7	200489	1.5	285402	2.4
SRTM (2000)	167406	126	200489	151	285402	215
Total	2,563,704	7,680.2	2,806,846	8,500.5	5,023,076	12568.4

Input Datasets	Middle East		PAKIN		S. America	
	Chips	Size/GB	Chips	Size/GB	Chips	Size/GB
S1GRD (2018-2020)	163986	983	147791	886	83756	502
GSSIC (2020)	158985	68	147791	69	83756	34
GUNW (2020)	608865	619	486914	309	226093	155
S2SRM (2018-2020)	163986	1126	147791	992	83756	529
ESAWC (2020)	163986	32	147791	29	83756	16
AGB (2018-2020)	163986	7.7	147791	7	83756	4.1
MODISVEG (2020)	163986	88	147791	113	83756	79
GHSBUILTS (2020)	163986	1.3	147791	1.2	83756	0.7
SRTM (2000)	163986	124	147791	112	83756	63
Total	2,899,668	7,282.4	2,555,988	6,288.2	1,398,677	3,453

* GUNW data unavailable for Europe.

A.2 Banding

As outlined in Section 3.3, we stratified our data into training, validation and test splits using geographic banding. 60 bands were constructed at a fixed orientation for each AOI. See Table A.2 for the angles used per-AOI. Specifically, we allocated three adjacent bands for the training set, followed by one adjacent band each for the validation and test sets, in sequence, until all bands were categorized. See Figure 1 for a visual representation.

Table A.2: **Angles** for the construction of training, validation and test splits by geographic banding. Angles are measured in radians, clockwise (axis pointed towards the Earth), with west corresponding to an angle of 0.

AOI	Band Angle
CONUS	0.9
Europe	0.9
China	1.5
Middle East	1.5
PAKIN	1.5
South America	0.6

B Marginal Distributions

The marginal distributions of ESAWC, MODISVEG, STRM mean elevation and GHSBUILTS can be seen in Figures B.1, B.2, B.3 and B.4 respectively.

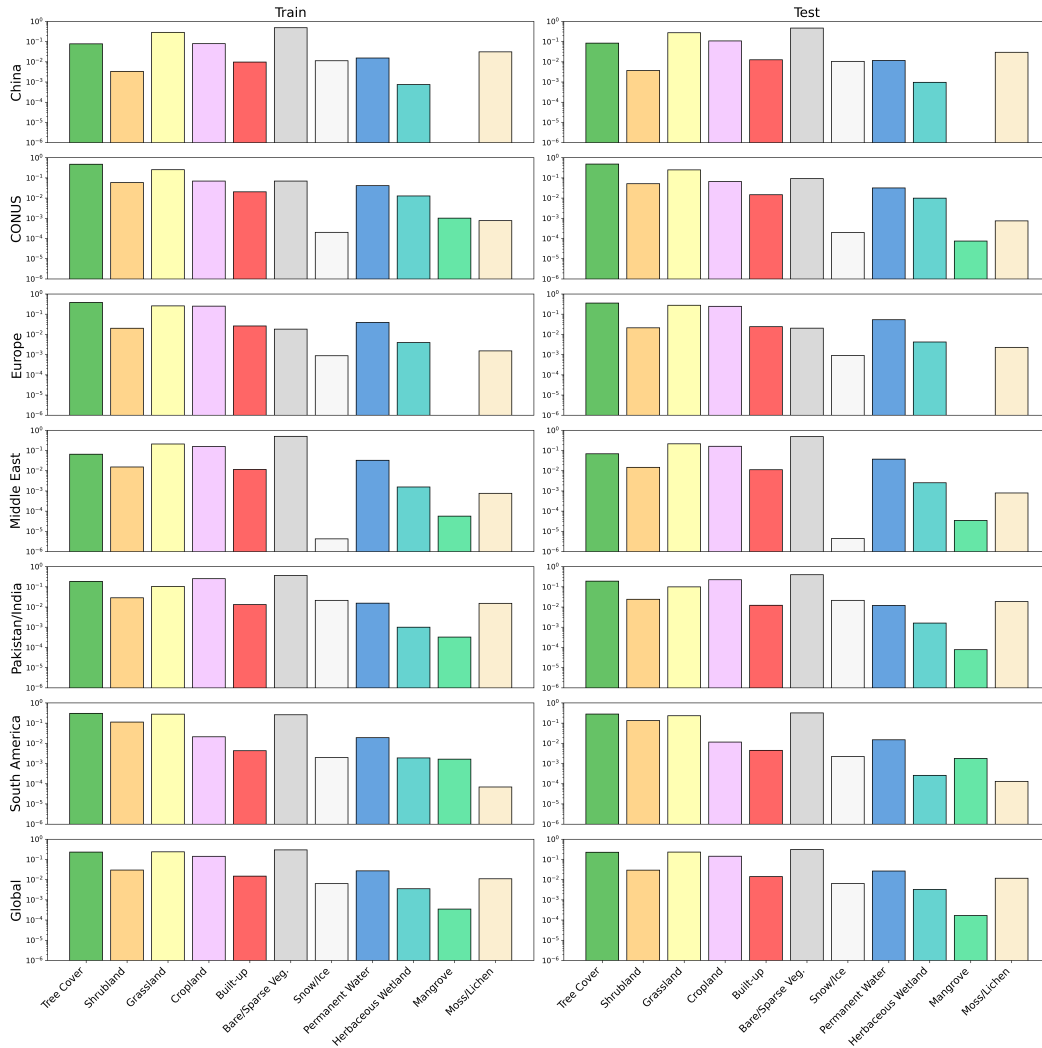


Figure B.1: Marginal distribution for ESAWC. Note log scale.

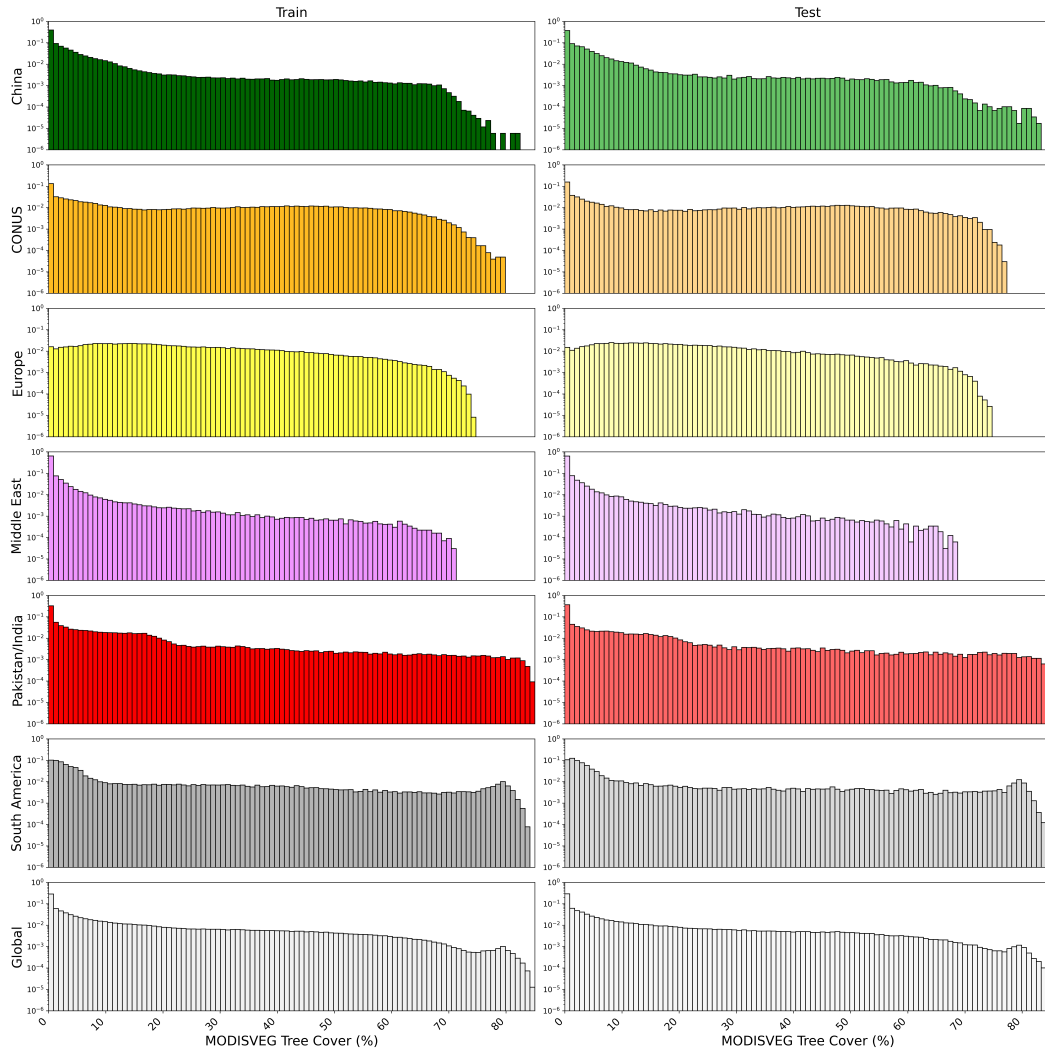


Figure B.2: Marginal distribution for MODISVEG. Note log scale.

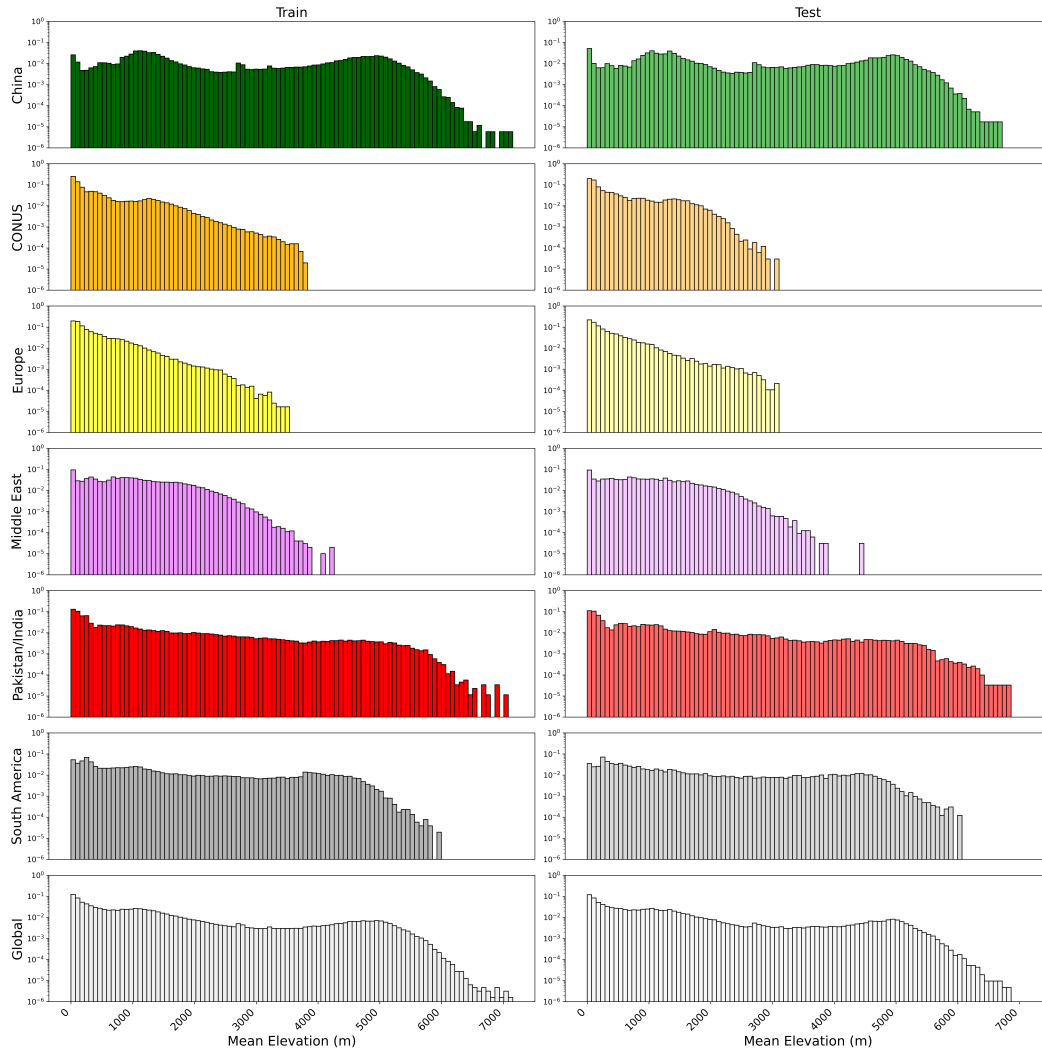


Figure B.3: Marginal distribution for SRTM. Mean elevation per-chip. Note log scale.

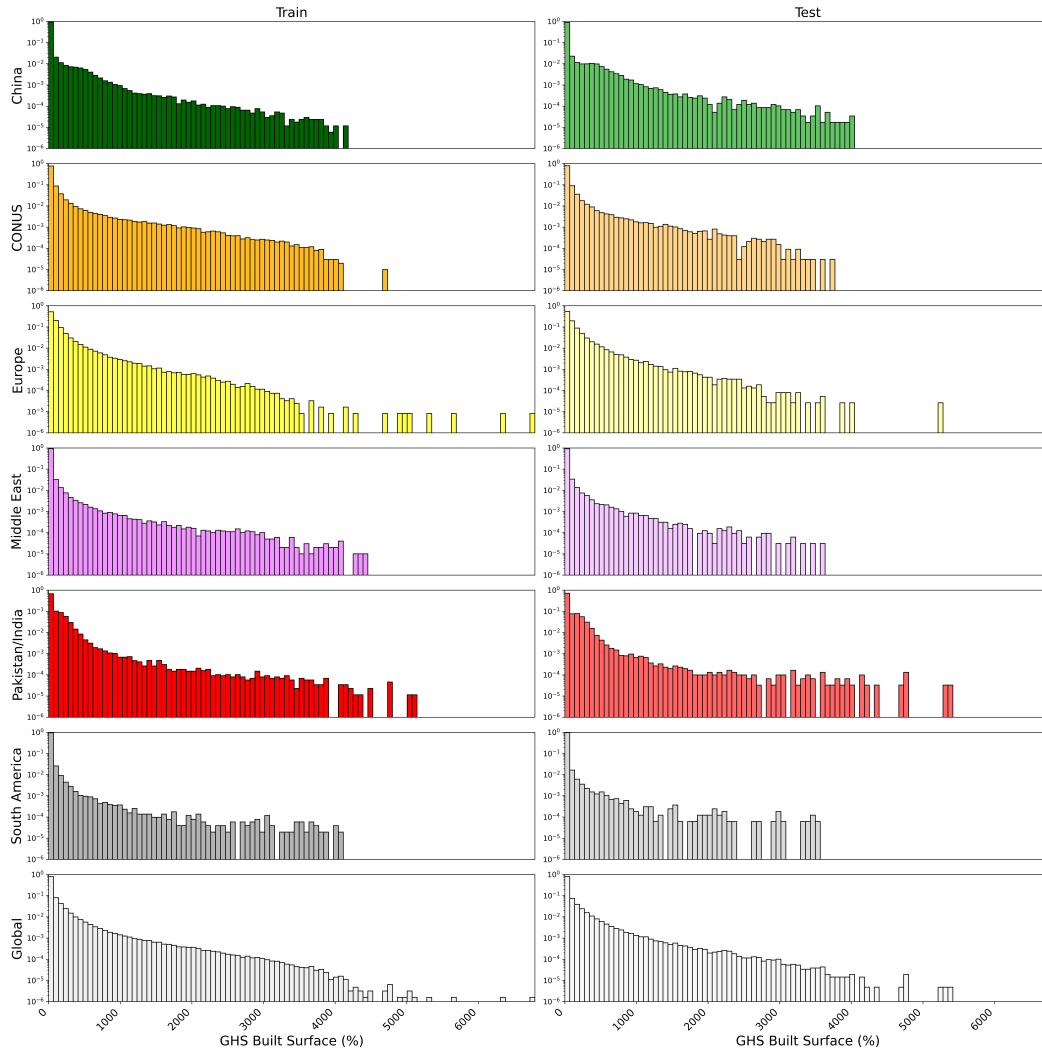


Figure B.4: Marginal distribution for GHSBUILTS. Note log scale.

C MAE Hyperparameters

Training hyperparameters for the MAE-based model can be seen in Table C.1. We followed [65], other than including additional AOIs in the pretraining set. A checkpoint for this model is available at huggingface.co/M3LE0.

Table C.1: **Hyperparameter details** for MAE pretraining used in main text.

MAE Pretraining	
Encoder (Params)	ViT-B [33] (88.8M)
Decoder (Params)	Reconstruction [35] (5.5M)
Loss Function	MSE
Input Image Size	448×448
Output Image Size	448×448
Patch Size	16×16
Masking Type	Random
Masking Ratio	0.75
Optimiser	AdamW
Learning Rate	1.00E-04
No. Epochs	75
Input Dataset	S1GRD
Channels	Seasonal, VV, VH, VV-VH, Ascending
AOIs	CONUS, Europe, China, Middle East, PAKIN, S. America

D Supervised Experiments

In addition to our explorations of distribution shift, we performed a small set of supervised experiments, reframing the auxiliary datasets as labelled tasks. S1GRD, GSSIC coherence and S2SRM were used as input datasets. For S1GRD, we trained models separately using the VV and VH bands only, and with the VV and VH bands stacked at the input to the model. For all S1GRD models, four seasonal summaries were used for each band, resulting in four input channels for the VV and VH models, and eight channels for the model using both VV and VH. For GSSIC coherence, the coherence band was used with a single date pair of delta 36 days taken for each season, resulting in four input channels. For the S2SRM models, we used the red, green and blue channels with one input channel from each of the months of March, June, September and December, totalling 12 input channels. We additionally trained models using both S2SRM RGB in combination with each of the other datasets separately, stacking the bands at the input to the model. All input data was taken from 2020. We upsampled low resolution input datasets to 448×448 px before input to the model.

We excluded GUNW from use in our baseline experiments to avoid introducing the mixed availability of GUNW chips as a confounding factor.

ESAWC, AGB and GHSBUILTS labelled datasets were used as targets. ESAWC was formulated as semantic segmentation spanning 11 land use classes, for which segmentation accuracy is reported as mean intersection-over-union (mIoU). ESAWC data was used at the original resolution of 448×448 . Both AGB and GHSBUILTS are formulated as regression tasks (per-pixel). Results are reported using root mean square error (RMSE), in Mg ha^{-1} for AGB and in m^2 built surface for GHSBUILTS. We resized labels for AGB and GHSBUILTS to a fixed size of 45×55 , to account for minor differences in dimension from chip-to-chip. We note that this means our pixels may not span exactly 1 hectare and therefore that results for AGB measured in Mg ha^{-1} are a heuristic rather than an absolute measure of biomass.

We excluded data from Europe in these experiments due to the absence of GUNW data in this region. This constraint was applied despite having not used GUNW in these baselines, as these experiments were completed prior to the exclusion of GUNW and repeating them was prohibitively expensive. All models were trained on the same random 10% subset of the data. The spatial coverage of this subset is still similar to other popular large-scale EO datasets. We used the entire test set to compute the final metrics.

Models We used a UNet-style architecture for all baseline experiments, following [103] with two major changes — halving the number of channels for all layers and replacing the up-convolutions with bilinear upsampling. For the AGB and GHSBUILTS regression tasks, we applied average pooling to the single-channel 448×448 UNet output to achieve an output dimension of 45×55 . We opted not to use data augmentation. Selecting augmentations for SAR data is challenging — common choices such as rotation or flipping may introduce invariance to information specific to different instrument polarisations or orbital direction, for example. For further details on training and hyperparameters, see Appendix D.3.

D.1 Results

Results for all tasks are reported in Table D.1. For the experiments using a single data source as input, S1GRD using both polarisations (VV+VH) achieved the best result for the ESAWC (mIoU: 0.4185) and AGB (RMSE: $27.467 \text{ Mg ha}^{-1}$) tasks. S2RGB achieved the best result for GHSBUILTS (RMSE: 131.968 m^2). GSSIC coherence produced the worst result in all three cases (ESAWC mIoU: 0.2906, AGB RMSE: $39.314 \text{ Mg ha}^{-1}$, GHSBUILTS RMSE: 131.968 m^2). The best results for experiments using multiple data sources as input were achieved by fusion of S1GRD and S2RGB in all cases (ESAWC mIoU: 0.4634, AGB RMSE: $25.137 \text{ Mg ha}^{-1}$, GHSBUILTS RMSE: 124.852 m^2), with significant improvements over either S1GRD or S2RGB individually. Fusion of S2RGB with GSSIC improved results marginally (ESAWC mIoU: 0.4198, AGB RMSE: $28.550 \text{ Mg ha}^{-1}$, GHSBUILTS RMSE: 131.300 m^2) compared to either individual data source individually in all cases.

D.2 Discussion

In all experiments, S1GRD combining VV and VH polarisations performed similarly to S2RGB, although the gap was small. The inclusion of both polarisations for S1GRD increased performance

Table D.1: **Baseline Evaluation Results** for ESAWC, AGB, and GHSBUILTS tasks using our UNet models, outlined in Section D.

Input Dataset	Bands	ESAWC MIoU	AGB RMSE (Mg ha⁻¹)	GHSBUILTS RMSE (m² built)
S1GRD	VV	0.3976	28.573	152.259
S1GRD	VH	0.3787	30.333	152.847
S1GRD	VV+VH	0.4185	27.467	141.719
GSSIC	Coherence	0.2906	39.314	196.242
S2RGB	RGB	0.4094	28.689	131.968
S2RGB+S1GRD	RGB+VV+VH	0.4634	25.137	124.852
S2RGB+GSSIC	RGB+Coherence	0.4198	28.550	131.300

uniformly compared to either individually. Unlike optical data, reflected SAR pulses are polarised differently depending on the terrain, and each measured polarisation contains unique information about the geometry of the imaged area. In all cases, performance was significantly improved by using both S1GRD and S2RGB data in fusion, despite giving similar performances individually, confirming a common finding in previous work.

Baseline experiments using GSSIC as the sole input data source performed significantly worse than for other input data types. This is likely explained in large part by the difference in resolution between GSSIC (90m, upsampled to 10m), S1GRD (10m) and S2RGB (10m). Performance improved slightly in all cases when GSSIC was included alongside in fusion with S2RGB. A nuanced approach is required for including coherence and interferometry, rather than naive direct input. Users may wish to use these data sources in self-supervised learning schemes — for example, pretraining models by constructing coherence or interferometry data from polarimetry pairs, or in a contrastive scheme alongside polarimetry data. Some work has been successful in using this type of pretraining [67]. Motivated by the success of polarimetry and coherence data in non-ML tasks [13–20], we suggest that these data types generated from date-paired SAR acquisitions are likely to show stronger performance on change detection-type tasks, which we did not demonstrate here. M3LEO may provide useful pretraining data for these tasks.

D.3 Training Hyperparameters

Hyperparameter details for supervised experiments can be seen in Table D.2. We did not perform significant hyperparameter tuning. Model selection was using best performance on the validation set. All models used approximately 7.9×10^6 trainable parameters, to the nearest 10^5 . We show runtimes for each experiment on 2 NVIDIA V100 GPUs in Table D.3.

Table D.2: **Hyperparameter details** for supervised experiments

	ESAWC	AGB	GHSBUILTS
Task Type	Semantic Segmentation	Regression	Regression
Loss Function	Cross Entropy	RMSE	RMSE
Input Dimensions	448×448	448×448	448×448
Output Dimensions	448×448	45×55	45×55
Output Channels	11	1	1
Optimiser	Adam [104]	Adam [104]	Adam [104]
Learning Rate	1.00E-04	1.00E-04	1.00E-04
No. Epochs	75	75	75
Model Selection	Best (val)	Best (val)	Best (val)

Table D.3: **Runtimes** for single-input and data fusion baseline experiments

Input Dataset	ESAWC	AGB	GHSBUILTS
S1GRD (VV)	16h 23m	15h 38m	15h 45m
S1GRD (VH)	16h 39m	15h 41m	15h 48m
S1GRD (VV+VH)	16h 54m	17h 29m	23h 50m
GSSIC	16h 29m	15h 33m	15h 41m
S2	17h 51m	17h 59m	18h 2m
S2+S1GRD (VV+VH)	22h 54m	24h 35m	23h 19m
S2+GSSIC	19h 23m	68h 20m	19h 49m

E ESA World Cover Land Cover Classes

We outline the 11 classes that together comprise the ESAWC dataset in Table E.1, along with their pixel values in M3LEO. We opted to leave these pixel values as-found in the original ESAWC.

Table E.1: **ESAWC Classes** along with pixel values for data as-found in M3LEO.

Pixel Value	Class
10	Tree Cover
20	Shrubland
30	Grassland
40	Cropland
50	Built-up
60	Bare/Sparse Vegetation
70	Snow and Ice
80	Permanent Water Bodies
90	Herbaceous Wetland
95	Mangroves
100	Moss and Lichen

F Limitations

While we highlight that M3LEO comprises ML-ready data and an easy-to-use framework, we also call attention to a number of potential limitations regarding both the data and framework.

F.1 Data Limitations

Coverage While the M3LEO dataset is large, coverage is not global. We limited coverage to the area covered in the GUNW dataset, which is approximately equal to the regions in which Sentinel-1 has dual-polarization, ascending-descending coverage. Generating further interferometric data is substantially technically complex, computationally demanding and requires the use of SAR acquisitions with phase information (which are difficult to access compared to the amplitude data we provide). Were this data to be generated, cloud storage requirements for M3LEO would approach the petabyte scale, for which we are unable to provide a feasible long-term storage solution.

Change Detection & Time Series Data We highlight the potential application of interferometric data to change detection tasks, but note that this data is not included in the initial release of M3LEO. Two datasets that could be used for change detection tasks have been processed — namely, ESA CCI Burned Area MODIS¹² and the Global Flood Database¹³ — but have not been tested extensively enough for inclusion here. We are unable to guarantee the release of these components simultaneously with the data advertised in the main text of this work, but aim to release them in the future.

Multitemporal Data Data from S1GRD, S2SRM and AGB have been tiled for 2018-2020 additionally, but other datasets are provided for 2020 only. Many datasets, such as ESAWC, simply do not exist for 2018 or 2019. One satellite of the Sentinel-1 pair, Sentinel-1B, suffered a power unit failure in December 2021, so we cannot provide data with the same coverage as 2018-2020 from 2021 onwards.

F.2 Framework Limitations

Data Loading Data is currently loaded from the disk using `xarray`. We chose to use `xarray` as it easily accommodates handling the wealth of metadata associated with remote sensing imagery, but it is not performant for loading a large number of tiles quickly. For users who wish to access the same data many times — either over many epochs, or many experiments — we recommend caching the returned data arrays at the first encounter. We provide this facility using `blosc2`. The caching process can be computationally expensive for large datasets, but is relatively far cheaper than performing repeat runs using `xarray`. Slightly increased disk space requirements are incurred. We provide data for download as parquet files, but Apache Spark is not currently integrated with the framework and this data will need to be decompressed before use.

Visualisation While we did not advertise data visualisation within the main body of this work, a small number of tools to visualise model outputs exist. We aim to include these in the initial code release but are unable to guarantee this.

Tile Size We provide data chips at a fixed spatial size of 448×448 m. While we provide the facility to re-tile data straightforwardly at different sizes with `geotiles` and `sartiles`, this process is computationally expensive when tiling over large spatial areas.

Benchmarking While we provided some baseline results (Appendix D) using our framework and data, we did not provide a benchmarking framework under which models could be compared. For example, we made no assertion as to whether models should be evaluated on chips with missing channels — we chose to fill any missing data with a dummy value of 0. We also did not include European data in our test set. Although comparisons could be made by copying our approach exactly, we encourage domain experts to evaluate models according to the needs of their particular application.

¹²developers.google.com/earth-engine/datasets/catalog/ESA_CCI_FireCCI_5_1

¹³developers.google.com/earth-engine/datasets/catalog/GLOBAL_FLOOD_DB_MODIS_EVENTS_V1

G Licenses

The licenses under which each component of M3LEO was originally distributed are listed in Table G.1. **We distribute our framework and data under the Creative Commons BY-SA 4.0 license.**

Table G.1: **Licenses** for components of M3LEO

Dataset	License
S1GRD	Creative Commons BY-SA 3.0 IGO
GSSIC	Creative Commons BY 4.0 DEED
GUNW	Other (free use) ¹
S2RGB	Creative Commons BY-SA 3.0 IGO
ESAWC	Creative Commons BY 4.0 DEED
AGB	Other (free use) ²
MODISVEG	No restrictions ³
GHSBUILTS	Other (free use) ⁴
SRTM	Other (free-use) ¹⁵

¹ <https://www.jpl.nasa.gov/jpl-image-use-policy>

² https://artefacts.ceda.ac.uk/licences/specific_licences/esacci_biomass_terms_and_conditions_v2.pdf

³ https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD44B#terms-of-use

⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011D0833>

⁵ https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003#terms-of-use