Publication Date:
18/12/2025

Methodology:
M002

Version:
v1.0

Contact:
Equitable Earth
info@eq-earth.com

Methodology

Baseline Setting Module

Summary

This module describes the processes and procedures for Equitable Earth's centralised approach to baseline setting for terrestrial forest conservation projects that avoid unplanned deforestation and degradation (AUDD). It includes the underlying principles and processes for designing a model-based baseline setting approach to determine jurisdictionally nested, project-level baselines derived from a Jurisdictional Reference Level (JRL). The document also details procedures for developing and applying a spatio-temporal model that forecasts biomass stock changes to proportionally allocate baselines based on predicted losses, while addressing uncertainty and accuracy.

Table of Contents

1 Introduction

1.1 Normative References

This document should be read in conjunction with:

- Methodology for Terrestrial Forest Conservation (M002)
- Terms & Definitions
- Future Improvements & Limitations

1.2 Reading Notes

Several sections in this document are divided into the following sections:

- **Process Overview:** provides a high-level conceptual description of specific procedures, the purpose, key objectives, and main components and innovations of the approach.
- **Procedures:** outlines the specific steps taken by Equitable Earth to implement the processes.

1.3 Objective

This module outlines Equitable Earth's processes and procedures for establishing project-level baselines, in accordance with M002.

1.4 Scope & Applicability

This module applies to eligible projects using M002. Refer to the *Eligibility Criteria* section in the Methodology for Terrestrial Conservation (M002) for more details.

Equitable Earth adheres to the same requirements for the Baseline Validity Period (BVP) and Historical Reference Period (HRP) as outlined in M002. Refer to the *Baseline Validity and Re-Evaluation* section in the Methodology for Terrestrial Conservation (M002) for more details.

2 Baseline Setting Approach

2.1 Overview

Equitable Earth uses a spatio-temporal model to set baselines for AUDD[1] projects. The approach leverages historical biomass change analysis, spatio-temporal risk modelling, and project-level emissions allocation. The baseline setting approach designed by Equitable Earth consists of both jurisdictional and project baselines. A Jurisdictional Reference Level (JRL) is calculated from historical biomass loss and used as an input to predict future loss by applying a computer vision model to map spatial risks and allocate emissions proportionally.

The core components of Equitable Earth's baseline setting approach are the following (Fig. 1):

1. **Jurisdictional Reference Level:** A JRL is calculated by estimating total biomass loss over the HRP. This loss is used as the jurisdiction-wide budget for downstream allocation (nesting).
2. **Risk Mapping:** identifies the most likely location and magnitude of future biomass loss based on observed historical trends.
3. **Allocation:** ensures the JRL budget is spatially distributed proportionally to the predicted biomass loss risk, subject to jurisdictional constraints on available biomass stock.

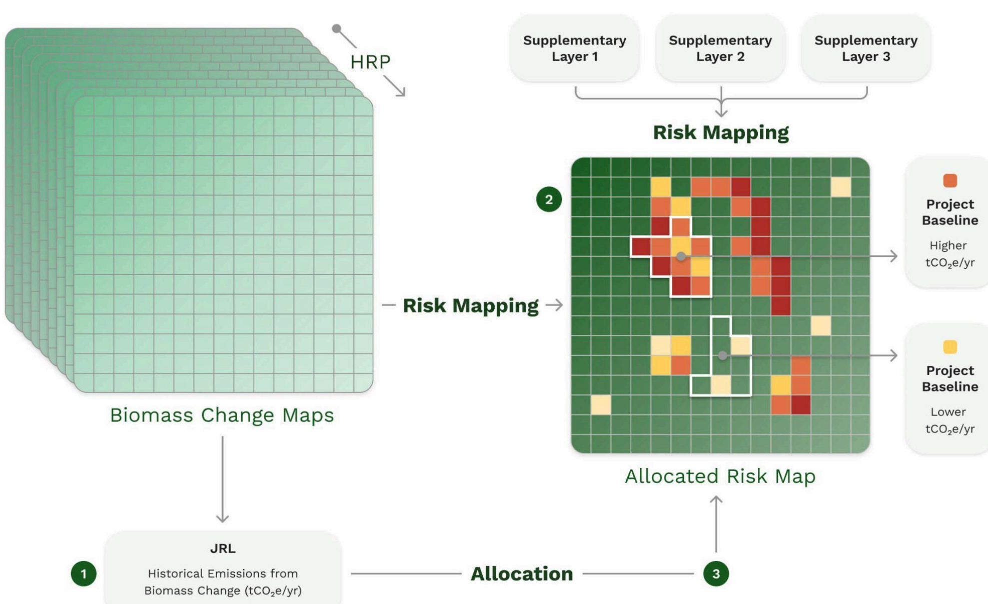

Fig. 1: Baseline setting approach, including JRL, risk mapping, and allocation.

[1] *Deforestation* and *Forest Degradation* are defined in M002 and Terms & Definitions.

2.2 Principles

Equitable Earth's baseline setting approach, including BAAR 2.0,[2] was designed according to the following first principles:

2.2.1 JRL Calculation

Principle: Historical emissions from biomass loss are captured holistically

- **Assumption:** The JRL accounts for all carbon emissions resulting from biomass loss, whether from deforestation or forest degradation.
- **Rationale:** The use of yearly biomass change data enables accurate accounting by ensuring no emissions are missed or falsely augmented due to classification errors or rigid activity definitions.

Principle: The JRL defines an emissions budget to be fully allocated within the nested system

- **Assumption:** The total jurisdictional emissions (i.e., JRL) are derived from observed biomass loss and spatially distributed across the jurisdiction to support project-level nesting.
- **Rationale:** Grounding the JRL in observed historical emissions ensures consistency with carbon dynamics, while JRL allocation via the risk map preserves environmental integrity and prevents over-crediting.

Principle: Disturbance categories are preserved for backward compatibility and comparability

- **Assumption:** Deforestation and degradation are estimated separately to support alignment with other carbon accounting frameworks.
- **Rationale:** Maintaining consistency with other reporting standards and human-defined activity categories enables benchmarking, auditability, and regulatory compatibility across different systems and time periods.

2.2.2 Risk Mapping

Principle: Spatio-temporal observations inform future predictions

[2] In this document Equitable Earth differentiates between *BAAR* and *risk model*. *BAAR* (Baseline Allocation for Assessed Risk) refers to the entire baseline setting process, which consists of allocating the risk to calculate the project baseline. *BAAR 2.0* refers to the latest version of the model. *Risk model* refers to the underlying (CNN) model that predicts biomass loss but does not perform the final project baseline allocation.

- **Assumption:** Past loss patterns are predictive of future trends in both space and time.
- **Rationale:** Forest disturbance processes (e.g., encroachment, degradation intensity, and spread) show spatial autocorrelation and temporal persistence, requiring a model that considers both dimensions.

Principle: Complex spatio-temporal dynamics are modelled

- **Assumption:** Biomass loss is driven by complex, dynamic, and context-specific processes that vary across space and time.
- **Rationale:** The model must detect and generalise complex disturbance patterns, such as frontier expansion, clustered degradation, or shifting mosaic dynamics. Computer vision methods are well-suited to this task, as they can learn spatial structures, edge behaviours, and contextual cues from raw imagery, enabling accurate and spatially aware predictions.

Principle: Supplementary geospatial data enhances predictive context

- **Assumption:** Including additional geospatial inputs beyond historical biomass change can improve predictive power.
- **Rationale:** Supplementary variables such as optical imagery, proximity to infrastructure, climate indicators, or land use history can provide valuable context about the drivers of forest loss. In machine learning frameworks, these inputs are treated as raw features rather than engineered covariates, allowing the model to learn complex interactions without imposing predefined structures.

2.2.3 Allocation

Principle: Spatial baseline allocation is proportional to relative risk

- **Assumption:** Emissions are distributed in proportion to the modelled risk of future biomass loss, ensuring that all eligible areas receive emissions according to their risk levels. This includes the flexibility to distribute proportionally across the entire landscape or to distribute to areas with historical loss first.
- **Rationale:** BAAR supports a scaled allocation mode, where pixels with historical loss are distributed first and subject to a biomass cap.

Principle: Baseline integrity is maintained through budget compliance

- **Assumption:** The JRL acts as a firm cap on the total emissions available for allocation, and allocated emissions must not exceed it.
- **Rationale:** Preventing over-allocation ensures environmental integrity and avoids inflating credit potential beyond observed emissions.

3 AUDD Baseline Setting: Baseline Allocation for Assessed Risk (BAAR)

3.1 Model Functionality

JRL Calculation Process Overview & Procedures

Process Overview

Rather than relying solely on land cover classification derived from activity data to capture deforestation and forest degradation, Equitable Earth also uses modelled estimates of biomass change based on historical observations to capture all emissions from carbon stock loss.

The JRL represents the total carbon emissions from deforestation and degradation within a jurisdiction over the HRP. It is calculated using AGB stock data provided by Equitable Earth's Above-Ground Biomass (AGB) provider.[3] This value serves as the input for the BAAR 2.0 model and defines the emissions "budget" to be allocated to projects via the risk map.

Procedures

Equitable Earth calculates the JRL for each jurisdiction following the steps below.

1) **Data Preparation**

    a) Annual biomass stock and change images from the AGB provider are clipped and aligned to the jurisdictional boundary.
    b) All pixels within existing Agriculture, Forestry, and Other Land Use (AFOLU) projects are excluded using a global database of nature-based carbon project boundaries.[4]

[3] Equitable Earth conducted a comprehensive benchmarking exercise to compare multiple external AGB providers. The objective of this assessment was to select the provider best suited to deliver rigorous, conservative, and accurate AGB data for calculating GHG reductions and removals. Based on this process, Chloris Geospatial has been selected as the primary AGB provider for this version of the methodology. Refer to the *JRL Accuracy* section for additional details on validation of AGB inputs.

[4] Existing AFOLU projects are excluded from the JRL and model training to prevent their historical success in reducing deforestation from resulting in an erroneously low-risk prediction. This ensures a viable baseline for projects wishing to transition to the M002 methodology. Database: Karnik, A. (2024) "A global database of nature-based carbon offset project boundaries". Zenodo. doi:10.5281/zenodo.11459391.

c) Forest cover is classified by applying a biomass threshold (e.g., 10 t/ha) to produce a binary forest/non-forest image, in alignment with Equitable Earth's forest definition.[5]

d) Pixels classified as forest at the beginning of the HRP are selected.

e) It is ensured that all layers share a consistent projection, resolution, and spatial extent.

2) Biomass Change Aggregation

a) Biomass changes are calculated using biomass stock layers over the entire HRP to produce a single image of net biomass change per pixel from the start to the end of the HRP, disregarding intermediate fluctuations.

b) All the pixels that show a positive change and overall gain in biomass are filtered out and excluded from calculations.[6]

3) **Legacy Disturbance Classification**

a) Forest cover is classified by applying a biomass threshold (e.g., 10 t/ha) to produce a binary forest/non-forest image.

b) Deforestation is identified by detecting transitions from forest to non-forest.

c) Degradation is quantified by isolating biomass loss in areas not flagged as deforested.

4) Uncertainty Assessment: Refer to the *Baseline Uncertainty & Accuracy* section for additional details on the procedures to calculate uncertainty related to the JRL calculation.

---

[5] Refer to the *Eligibility Criteria* section in M002 for more details on requirements.

[6] The current version of the risk model applies a loss-only approach for calculating the JRL, AUDD project baseline, and monitored emission reductions, establishing a symmetric accounting framework focused on avoided emissions. This approach excludes biomass gains from natural recovery to avoid mixing avoided emissions with removals. While AGB losses due to natural disturbance events (e.g., storms, landslides) are reflected in the AUDD baseline, the resulting biomass gains from recovery are not credited as removals. Because these natural phenomena are assumed to be persistent in the long term within the project area, and developers are not required to implement activities targeting such events, the continued gains during monitoring will not generate carbon credits. Refer to the Future Improvements and Limitations document for additional details on Equitable Earth's plans to integrate emission reductions and removals.

Risk Mapping Process Overview & Procedures

Process Overview

The risk-mapping process creates a risk map tailored to a specific jurisdiction using the risk model. While the risk map is jurisdiction-specific, the risk model is developed for all Equitable Earth projects regardless of location.

The risk mapping process uses advanced machine learning techniques to analyse historical dynamics in both the spatial and temporal dimensions in order to predict the location and magnitude of future biomass change.

A computer vision model was developed to predict biomass change per pixel. Computer vision models extract spatial features such as edges, localised patterns, and value gradients from stacked image inputs, and learn the relationships between these features and observed biomass loss through iterative training.

The risk-mapping workflow consists of two main processes: **model development** and **model application** (Fig. 2).

1) **Model development:** Consists of defining and fine-tuning a risk-mapping model to identify the best hyperparameters and evaluate model performance by comparing predictions with observed data. Hyperparameters include settings such as model architecture, window size, learning rate, and loss function, as well as the selection and pre-processing of input layers. Model weights are not included.
2) **Model application:** Consists of reusing the hyperparameters identified during model development to train the previously developed model on the specified jurisdiction during the training period and produce risk maps. The weights of this trained model are then saved and reused to assess risks for the BVP.

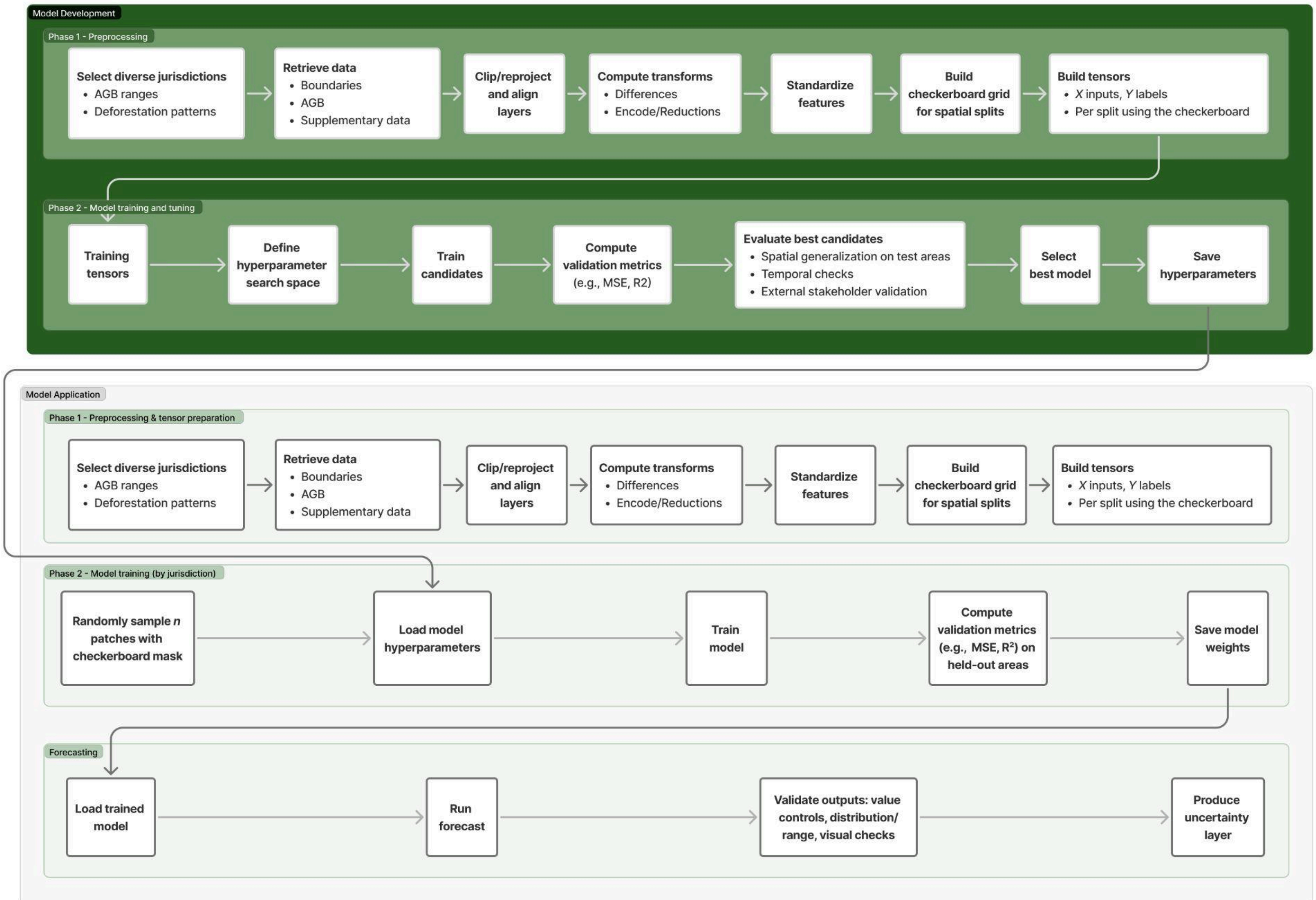

Fig. 2: Risk-mapping workflow including the model development and model application processes.

The training period, which matches the BVP, is dedicated to training, testing, and validating the model. During this time, the model's performance is evaluated against observed historical data (Fig. 3).

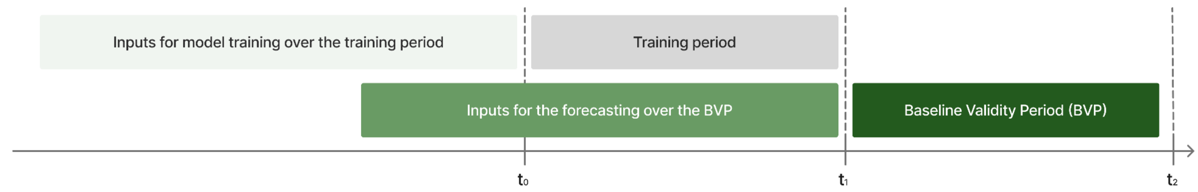

Fig. 3: Diagram illustrating the relationship between the training period, BVP, and their respective data inputs.

Procedures

Model development

Equitable Earth developed a risk model by following the steps below.

Phase 1: Pre-processing

- Selected different jurisdictions that showed a wide range of AGB values and different deforestation patterns.
- Retrieved jurisdictional boundaries, downloaded annual biomass stock images from the AGB provider, and gathered supplementary data, like optical imagery and embeddings, from other sources.[7]
- Clipped and reprojected data layers for each jurisdiction to match its boundary and align the data.
- Excluded all pixels within existing AFOLU projects using a global database of nature-based carbon project boundaries.
- Calculated transformations such as differences, and considered encoding and performing dimension reduction.
- Standardised data and transformed features to ensure comparable scales, to ensure that no single feature dominated due to its magnitude.
- Created a checkerboard grid for each jurisdiction to define distinct areas for splitting the dataset for training, validation, and testing, which reduced overfitting of spatially autocorrelated features.

[7] The current version of the risk model uses embeddings from AlphaEarth as a supplementary layer to provide more topographic context to the model.

- Used the pre-processed data from the previous step to create tensors: one for the input data and another for the labels.[8] Both tensors were divided using a checkerboard grid to spatially separate the areas used for training, validation, and testing, resulting in three pairs of tensors for each spatially independent split.

Phase 2: Model training and tuning

- Defined a range of candidate hyperparameter values to test. These included settings such as model architecture, window size, learning rate, and loss function, as well as the selection and preprocessing of input layers.
- Trained a model over multiple epochs (iterations)[9] for each combination of hyperparameter values. After each full pass through the training data, validation metrics, such as R-squared (R^2) and the mean squared error (MSE), were calculated by comparing predicted and observed biomass change for the validation pixels (labels). Validation was performed on spatial data held out from the same calibration periods, separated via the checkerboard mask.
- Stopped training after a set number of epochs that did not deliver an improvement in validation metrics to prevent overfitting.
- Evaluated the model's performance across all tested hyperparameter combinations, starting by assessing the model on the spatial dimension using the test data held out during training. Afterwards, the temporal dimension was evaluated by checking the distribution and range, and visually assessing the spatial distribution of the loss. Equitable Earth avoided relying on key metrics from observed forest loss, as they were unavailable for temporal validation.[10] Refer to the *Baseline Uncertainty & Accuracy* section for additional details on accuracy evaluation.
- Selected the model with the lowest validation error, using the MSE as the default.
- Saved the corresponding hyperparameter values for the model application step. In this step, a new model is retrained and applied to the BVP.

---

[8] The model is trained using future biomass changes as labels. For instance, if the model is trained for the period from 2018 to 2024, the label represents the biomass change observed during this time, calculated from AGB data.
[9] In machine learning, an epoch refers to one complete pass of the entire training dataset through the learning algorithm. During this single pass, the model is exposed to every data sample once, and its internal parameters are adjusted based on the calculated error. Multiple epochs are used for the model to fully learn complex AGB loss patterns in the data.
[10] During the risk model development processes, external stakeholders validated Equitable Earth's protocol to assess model performance.

Model application

Equitable Earth applies the risk model following the steps outlined below.

Phase 1: Pre-processing and tensor preparation

- Jurisdictional boundaries are retrieved, and annual biomass stock images and supplementary data (e.g., optical imagery, embeddings) are downloaded.
- Transformations such as differences are calculated, and encoding or reducing dimensions based on the model development guidelines are considered.
- Data is standardised and features transformed to have comparable scales, ensuring no single feature dominates due to its magnitude.
- A checkerboard grid is created across the jurisdiction to define areas used for splitting the dataset into training, validation, and testing areas, thereby preventing overfitting of spatially autocorrelated features.
- The preprocessed data from the previous step is used to create tensors: one for the input data and another for the labels. Both tensors are divided using a checkerboard grid to spatially separate the different splits, resulting in three pairs of tensors for each spatially independent split covering the training period. An additional tensor for the forecast covering the BVP is computed.

Phase 2: Model training

- Random sampling with a checkerboard mask is used to extract *n* patches from stacked tensors. When generating test patches, random samples from the regions designated as "test" by the checkerboard split were taken, while training patches were sampled from the regions designated as "train".
- The model is trained over multiple epochs (iterations). After each full pass through the training data, validation metrics (e.g., MSE, R^2) are calculated by comparing predicted and observed biomass change for the validation pixels (labels). Validation is performed on spatial data held out from the same calibration periods, separated via the checkerboard mask. Refer to the *Baseline Uncertainty & Accuracy* section for additional details on performance evaluation.
- Training is stopped after a set number of epochs that do not deliver an improvement in validation metrics to prevent overfitting.

Phase 3: Forecasting

- The model trained during Phase 2 is reused for the next period to create a risk map for the BVP.
- The risk map is validated through a combination of value control, distribution, and range checks, and visual validation.

- The uncertainty layer is created by following the steps outlined in the *Baseline Uncertainty & Accuracy* section.

Allocation Process Overview & Procedures

Process Overview

The allocation process uses the output from the risk mapping process to distribute the JRL to the project areas using a scaled allocation mode. This approach consists of a risk-weighted proportional distribution of the JRL across all eligible pixels, constrained by per-pixel biomass stock at the end of the HRP. Emissions are scaled by the relative risk value of each pixel. Allocation is only applied to pixels with non-zero biomass stock and constrained to pixels with a forest mask.

Procedures

Equitable Earth follows the steps outlined below to allocate a project-level baseline.

JRL and Forecast Map Feature Extraction

- The output from the JRL calculation process is used to determine total emissions to be allocated across the jurisdiction.
- Historical loss pixels representing areas where biomass has previously declined are identified.
- Historical non-loss pixels (including stable and gain) representing areas with no observed historical biomass loss are identified.

Allocation

- The entire JRL is distributed proportionally to predicted biomass loss across all eligible pixels.

Project Baseline Calculation

- The pixels from the output of the previous step that intersect with the specific project area are aggregated to establish the nested project baseline.

Baseline Correction Overview & Procedures

Process Overview

Equitable Earth accounts for emissions from both natural and anthropogenic degradation, as well as emissions from deforestation, meaning that resulting AUDD baselines are not directly comparable to modelled deforestation-only baselines. Developers are unlikely to mitigate emissions from natural degradation through project activities. Equitable Earth implements the following accounting measures to address the inclusion of natural degradation emissions in baseline estimates and ensure these emissions are excluded from crediting:

- **Estimating natural degradation emissions:** A historical analysis at the jurisdictional level is conducted to observe trends across all pixels and classify those experiencing degradation as either natural or anthropogenic, based on time series and spatial evaluations.[11]
- **Monitoring natural degradation emissions**: Emissions are actively monitored during each monitoring period, and appropriate deductions are made from the baseline during the quantification of reductions to ensure no credits are generated in these circumstances. Baseline deductions are proportional to the magnitude of natural degradation estimates.

Procedures

Equitable Earth follows the steps outlined below to estimate the impact of natural degradation on the project baseline.

Retrieve AGB data over the jurisdiction

- The AGB state for all years within the HRP is retrieved.
- Quality control filters (e.g., data gaps, known artefacts) are applied, and non-forest or ineligible areas are masked out as required by the methodology.

Calculate the AGB evolution for each pixel over the HRP

- For each pixel, year-to-year changes in AGB over the HRP are calculated.
- Indicators of AGB trend (e.g., mean annual change, slope of AGB vs. time, variance) are derived to characterise the trajectory of each pixel.

Classify pixels by degradation behaviour

- Pixels presenting a statistically significant AGB decline over the HRP (candidate degraded pixels) are identified.
- Pixels affected by deforestation are excluded.
- Candidate pixels into degradation classes (e.g., mild, moderate, severe) are grouped based on the magnitude and persistence of AGB decline.

[11] Natural and anthropogenic degradation types are assumed to differ in temporal scale and spatial characteristics, where natural degradation is assumed to be global and anthropogenic degradation is primarily time-bound and localised.

Distinguish natural vs. anthropogenic degradation using time series patterns

- For each candidate degraded pixel, the temporal pattern of AGB change (e.g., gradual vs. abrupt, single vs. repeated events) is analysed.
- The spatial context (e.g., distance to infrastructure, settlements, historical deforestation fronts, fire scars, and known logging areas) is characterised.

Quantify the jurisdictional natural degradation signal

- AGB losses attributable to natural degradation across the jurisdiction for each year in the HRP are aggregated.
- Summary statistics (e.g., mean annual natural degradation rate, inter-annual variability, confidence intervals) are produced.
- Jurisdiction-level parameters are derived (e.g., average natural degradation factor per hectare for relevant forest strata).

Project potential emissions from natural degradation

- For each pixel with historical natural degradation, the expected natural AGB loss over the BVP is estimated by applying the average historical natural degradation rate observed during the HRP.
- The pixel-level emission values across the project crediting area and BVP are aggregated to determine the total potential emissions from natural degradation that will be deducted from the project baseline.

4 Baseline Uncertainty & Accuracy

4.1 Jurisdictional Reference Level

JRL Uncertainty

Equitable Earth calculates uncertainty at the pixel level, expressed as a 95% confidence interval (CI), which is then aggregated to the jurisdictional scale. This aggregation involves converting the pixel-based uncertainty to area-based uncertainty using the Monte Carlo simulation method, with the spatial autocorrelation being captured through the application of equation E.1.

$$\mathbf{Z}_{total,\,i} = \mathbf{Z}_{global,\,i} \times \sqrt{\varrho} + \mathbf{Z}_{noise,\,i} \times \sqrt{1 - \varrho} \qquad \textbf{(E.1)}$$

Where:

- $\mathbf{Z}_{total,\,i}$ = Perturbation field across the studied area at iteration *i*; dimensionless
- $\mathbf{Z}_{global,\,i}$ = Global shock across the studied area at iteration *i*, identical for all pixels of the same map, and randomly drawn from a normal distribution with a mean of 0 and a variance of 1; dimensionless
- $\mathbf{Z}_{noise,\,i}$ = Pixel-level independent noise at iteration *i*, independently drawn for each pixel from a normal distribution with a mean of 0 and a variance of 1; dimensionless
- ϱ = Correlation factor between the pixels; a default value of 0.01 is used; dimensionless

When quantifying the change between two states from different years, the $\mathbf{Z}_{global,\,i}$ value is generated independently for each year to better capture the evolution of the captors, algorithms, and other factors.

JRL Accuracy

JRL accuracy is directly linked to the accuracy of the AGB provider. Equitable Earth's primary AGB provider was selected through a rigorous benchmarking process that evaluated providers based on precision, uncertainty analysis, coverage, and integration feasibility, ensuring that the AGB data underpinning JRL calculations meet the highest standards of accuracy and reliability.[12]

Accuracy of AGB data inputs

Equitable Earth uses the AGB provider's EO-based AGB stock product, provided as an annual time series, to calculate the JRL and feed it into the risk model. The AGB provider's models are trained at a continental scale using large reference datasets that combine airborne and spaceborne LiDAR and satellite data. AGB in this product is derived using standard allometric equations and wood-density values, so uncertainties from allometry and wood density are inherently reflected in the AGB product. The AGB provider further quantifies this uncertainty with a quantile regression approach that provides pixel-level and polygon-level confidence intervals suitable for jurisdictional and project applications.

The AGB provider has also explicitly tested the temporal performance of its AGB time series. Using multiple NEON sites in the United States, they compare multi-year AGB trajectories with repeated airborne LiDAR–derived canopy height products. A disturbance case study at Lyndon B. Johnson National Grassland shows that the AGB provider detects documented AGB changes when cross-checked against NEON and high-resolution imagery. These analyses provide empirical evidence that year-to-year changes in the AGB stock product used in M002 can resolve biomass dynamics at scales relevant for allocated baselines.

Equitable Earth uses the AGB provider's polygon-level uncertainty estimates for AGB stocks and propagates them through the construction of the JRL. This means that uncertainties linked to sensor noise, model error, allometry, and wood density are explicitly reflected in baseline emissions, reducing the risk that uncertainties in AGB are underestimated and helping guard against potential over- or mis-crediting.[13]

---

[12] Additional details on the benchmarking process are available in this document.
[13] Refer to the AGB provider's validation white paper for additional details.

4.2 Risk Model

Risk Map Uncertainty

Sensitivity-Based Approach

Equitable Earth adopts a sensitivity-based epistemic uncertainty approach to quantify model uncertainty. Under this approach, uncertainty is higher in regions where small perturbations in the input variables lead to large changes in model outputs. Formally, for a prediction ($y = f(x)$), local sensitivity is characterised by the gradient of the output with respect to input x.

Equitable Earth operationalises this sensitivity-based approach using PyTorch to enable gradient computation of the output with respect to input data. Each data patch is passed through the network, and gradients of the model output are computed using a backward pass. A Gaussian Kernel function, centred on the data patch, ensures that pixels from the patch closer to the prediction centre are given a greater weight. This process results in a per-pixel sensitivity map reflecting the degree to which the output of the model depends on each input variable. These sensitivity values indicate how susceptible a given prediction is to small and local changes in the input data. Although these sensitive values do not constitute uncertainty themselves, they serve as a basis for determining both input and model uncertainty.

Propagation of Input Uncertainty

The AGB stock product used by Equitable Earth includes AGB maps with associated Standard Error (SE) rasters that quantify the uncertainty of the dataset. For the satellite embeddings used as additional input data layers (i.e., channels), Equitable Earth assumes no measurement error (SE = 0).

Unlike the global sensitivity previously described, Equitable Earth does not apply an L2 norm across channels. Instead, Equitable Earth preserves the sensitivity per channel and directly weights each one by its associated SE raster. This approach ensures that each input source contributes to the overall uncertainty independently, avoiding artificial mixing of channels.

The local propagation of input uncertainty follows the first-order error propagation rule (E.2):

$$\sigma_{\text{input}}(x) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial f(x)}{\partial x_i} \cdot SE_i(x) \right)^2} \qquad \textbf{(E.2)}$$

Where:

- $\sigma_{input}(x)$ is the local propagation of pixel x uncertainty
- N is the number of input channels with available SE
- $SE_i(x)$ is the standard error of the input raster i at pixel x
- $\frac{\delta f(x)}{\delta x_i}$ is the sensitivity of the prediction to the channel i

Sensitivities can be used directly without any normalisation, as SE values are already expressed in physical units. $\sigma_{input}(x)$ reflects a physically consistent propagation of known measurement errors in the model predictions.

Propagation of Model Error (RMSE Normalisation)

Input uncertainty cannot capture all sources of errors affecting model outputs. In particular, it does not account for uncertainty intrinsic to the neural network model, such as misspecification, overfitting, or approximation errors.

To account for such uncertainty, Equitable Earth captures this model uncertainty by exploiting the global Root Mean Squared Error (RMSE) measured on the test dataset (E.3):

$$\mathrm{RMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^{M} (y_j - \hat{y}_j)^2} \qquad \textbf{(E.3)}$$

Where:

- M is the number of test samples,
- y_j is the reference value of sample j, measured on the ground
- $\hat{y}_j$ is the prediction of sample j, calculated by the model

Equitable Earth calculates the RMSE across the entire jurisdiction, excluding any training patches, to ensure the estimate is unbiased and represents model performance in areas not used for model calibration. Model sensitivity values are dimensionless and scale-dependent and cannot be directly compared to the RMSE. Consequently, Equitable Earth normalises these sensitivity values and uses them as relative weights for the spatial distribution of the global error. (E.4):

$$\tilde{s}(x) = \frac{s(x)}{\bar{s}} \quad \textbf{(E.4)}$$

Where:

- $\bar{s}$ is the mean sensitivity across all pixels

This normalisation step ensures that the spatial average of the model error allocated across all pixels of the jurisdiction equals the observed RMSE. Using the mean value avoids a disproportionate influence of outliers and maintains consistency across the landscape. The final per-pixel model uncertainty is calculated using the following equation:

$$\sigma_{\text{model}}(x) = \tilde{s}(x) \cdot \text{RMSE} \quad \textbf{(E.5)}$$

Combination into Final Uncertainty Raster

Equitable Earth combines input and model uncertainties to estimate the total uncertainty at the pixel level, assuming their sources are independent (E.6):

$$\sigma_{\text{total}}(x) = \sqrt{\sigma_{\text{input}}(x)^2 + \sigma_{\text{model}}(x)^2} \quad \textbf{(E.6)}$$

This raster represents the per-pixel epistemic uncertainty, expressed in units consistent with the target variable.

This overall method creates multiple rasters:

- Channel-specific sensitivity raster for SE propagation
- Global sensitivity raster for RMSE scaling

- Input uncertainty raster σ_{input}
- Model uncertainty raster σ_{model}
- Final uncertainty raster σ_{total}

This decomposition allows for transparent inspection of uncertainty sources and supports robust interpretation and decision-making.

Using uncertainty rasters to calculate site-level risk uncertainty

To compute site-level risk uncertainty based on the risk map uncertainty layer, Equitable Earth uses the following Monte Carlo framework.

Equitable Earth selects the following information as inputs:

- The mean AGB risk raster and its corresponding standard deviation raster,
- The RS ratio raster and its corresponding standard deviation raster,
- A variogram (*nugget*, *sill*, *range*) describes the spatial correlation of the risk uncertainty.

Equitable Earth executes the steps outlined below for each iteration (i).

1) A perturbation field $Z_{total,\,i}$ is generated by:
    - Sampling a spatially correlated standard-normal field $Z_{srf,\,i}$ based on a variance of $\frac{still}{nugget + still}$.
    - Sampling an independent pixel-level noise $Z_{noise,\,i}$ from $N(0, \frac{nugget}{nugget + still})$.

    The total perturbation field is calculated as the sum of both elements, resulting in a field with zero mean and unit variance.

    $$Z_{total,\,i} = Z_{srf,\,i} + Z_{noise,\,i},$$

2) Pixel-level risk values of the risk map ($Risk_i$) are perturbed using the total perturbation field and the pixel-level standard deviation.

    $$Risk_i = Risk_{mean} + (Risk_{sigma} \times Z_{total,i}).$$

3) The RS Ratio (RS_i) on eligible pixels is independently sampled.

    $$RS_i \sim N(RS_{mean}, RS_{sigma})$$

4) Total Biomass Loss ($Loss_i$) is calculated using the following equation:

$$Loss_i = Risk_i(1 + RS_i)$$

5) Pixel-level losses $Loss_i$ are aggregated to each project zone using zonal statistics.

After the last iteration, Equitable Earth has generated an ensemble of outcomes for each project zone, providing the mean, standard deviation, and confidence intervals of site-level risk baselines. A subsequent rescaling step aligns the mean allocation to the jurisdictional baseline distribution.

Risk Map Accuracy

Accuracy Assessment Procedures

Equitable Earth developed a dedicated evaluation pipeline, featuring several complementary components, to assess the accuracy of the deep learning model in predicting AGB evolution over the next six years across both spatial and temporal dimensions.

- The spatial evaluation consists of comparing predicted biomass loss to observed loss for a historical period.
- Temporal accuracy is assessed using statistical checks.

The evaluation described below is performed on the forecasts produced over the training period (e.g., 2018–2024), for which corresponding ground truth observations are available and used for reference.

Spatial evaluation

Equitable Earth evaluation is composed of the following elements:

1) **Multi-metric performance assessment**

Equitable Earth calculates a range of evaluation metrics to capture different aspects of model accuracy, including: Root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), median MAE, bias, relative bias, and relative error.

2) **Multiscale evaluation**

To assess model performance across different spatial scales, Equitable Earth calculates metrics both at the pixel level and on aggregated grids.

For the grid-based evaluation, the prediction and reference rasters are divided into cells of fixed size. Within each cell, Equitable Earth first checks the validity of every pixel (i.e., whether it was excluded from the input patches used during training). For each valid cell, two values are calculated:

- The mean prediction and mean reference value across all valid pixels
- A weight equal to the number of valid pixels in the cell

All evaluation metrics are subsequently calculated using these weighted rasters. This weighting ensures that each cell contributes proportionally to the amount of valid data it contains, preventing cells with limited valid pixels from disproportionately influencing the evaluation results.

3) **Scenario-based performance analysis**

Beyond global statistics, Equitable Earth extracts subsets of pixels that represent specific AGB dynamics. For example: stable AGB (high, medium, or low), increasing or decreasing AGB trends, recently increasing or decreasing AGB trajectories, etc. Evaluating model performance within these subsets allows Equitable Earth to assess whether model accuracy is consistent across different ecological scenarios and change regimes.

4) **Baseline model comparison**

Additionally, to contextualise model performance, Equitable Earth compares the deep learning model against simple baseline approaches, including:

- **Constant AGB assumption:** predicting no change in AGB across time
- **Trend-following models:** extrapolating future AGB based on observed trends from the past (*N*) years

These baseline comparisons help quantify the added value of the approach relative to straightforward heuristic strategies, both at the global level and across specific scenarios.

5) **Strict spatial hold-out evaluation**

Evaluation is conducted only on pixels that were not included within the training patches to ensure a fair and unbiased assessment. This approach guarantees that reported accuracy reflects the model's ability to generalise to unseen spatial contexts rather than memorising training patterns.

6) Visual evaluation of forecasts

Finally, Equitable Earth visually inspects the forecasts to assess the plausibility and spatial consistency of predicted AGB changes. This visual evaluation provides an additional layer of qualitative validation of the model's outputs.

Temporal evaluation

The primary forecast of interest corresponds to the 2024–2030 period, for which no ground truth observations are currently available. As a result, standard accuracy metrics cannot be calculated. To address this, Equitable Earth performs several evaluations to ensure the reliability and consistency of the forecast.

1) Input data distribution check

Equitable Earth analyses the statistical properties of the inputs for the 2024–2030 forecast to confirm they fall within the distribution observed during training, thereby mitigating potential extrapolation errors.

2) Historical forecasting experiments

Equitable Earth has conducted tests to train degraded models[14] on earlier periods (e.g., 2012–2018) and generate a forecast for the next one for which ground truth is available (e.g., 2018–2024). These experiments serve as validation of the model's forecasting principle, allowing Equitable Earth to quantify expected performance in a real-world scenario.

3) Cross-forecast consistency assessment

Equitable Earth visually compares the forecasts from consecutive periods (e.g., 2018–2024 vs. 2024–2030) to verify that spatial patterns, trends, and decision-driving features, such as sources of deforestation, are coherent.

[14] A model is considered "degraded" when certain input data from previous years is unavailable.

4.3 Baseline

Baseline Uncertainty

The project baseline is represented as a probability distribution to capture its inherent uncertainty. This distribution is obtained by combining the JRL distribution with the risk map distribution using a Monte Carlo method described in the *Monte Carlo Simulation* section in M002. In practice, multiple realisations of the project baseline are generated by simulating risk maps and pairing them with corresponding realisations of the JRL, producing a distribution of possible baseline values.

Appendix A: Documentation History

Version	Date	Description
v1.0	18/12/2025	Public release of version 1.0 of the Baseline Setting Module

Equitable Earth

www.eq-earth.com