CarbonSense: A Comprehensive Dataset and Baseline Model for Deep Carbon Flux Modelling

Matthew Fortier
Mila Quebec
matthew.fortier@mila.quebec

Mats L. Richter
ServiceNow

Oliver Sonnentag*
Université de Montréal

Chris Pal*
Mila Quebec

Abstract

(Abstract to be written)

1 Introduction

Terrestrial land-atmosphere carbon fluxes provide crucial insight into how our biosphere is responding to climate change. These fluxes measure the exchange of gases, such as carbon dioxide (CO2) and methane (CH4), as well as water vapor, through ecological processes such as photosynthesis and cellular respiration. Carbon fluxes are localized phenomena that cannot be directly measured with remote sensing; instead, a micro-meteorological method called eddy covariance (EC) can be used to measure fluxes in a small area using sensors mounted on a tower. Using these EC stations, field ecologists can measure carbon fluxes and localized meteorological data. A simplified depiction of an EC station is given in figure 1.

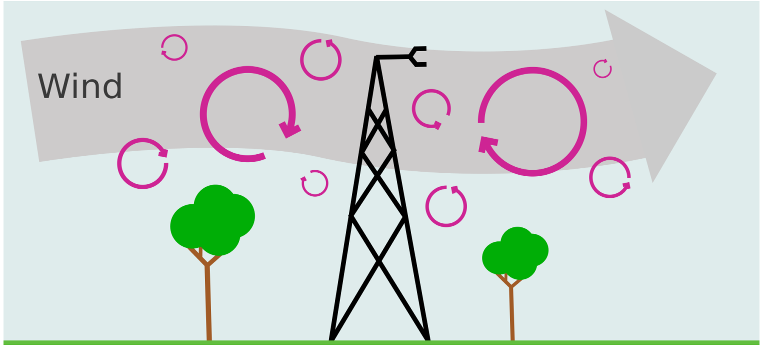

Figure 1: Simplified EC station. Sensors mounted on a tower detect gas fluxes by measuring their concentration in turbulent atmospheric vortices ("eddies").

Accurately simulating these carbon fluxes on a global scale is challenging for process-based climate models. In response, researchers use site-level EC data to train machine learning (ML) models to predict them from biophysical data. Rudimentary neural networks were used as early as 2003

[*] Equal Contribution

Submitted to the 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks. Do not distribute.

[1], with subsequent work using these models to predict fluxes from global meteorological data [2]. Even with mediocre out-of-distribution performance, these global carbon flux products are used as a benchmark for refining process-based climate models [3] and help ecologists understand how terrestrial ecosystems are responding to climate change.

Improvements have primarily focused on data processing rather than algorithmic refinement. The FLUXNET network has been aggregating eddy covariance data from regional networks around the world with a standardized processing pipeline ("ONEFlux") [4]. The FLUXCOM group began supplementing FLUXNET data with satellite imagery over the EC sites in order to improve performance [5] [6]. However, most state-of-the-art carbon flux models still use tree- or kernel-based methods such as random forests [7] [8] [9], XGBoost [6], or ensembles of similar methods [10] [11]. These models use tabular data, so any input imagery needs to be compressed resulting in substantial information loss.

Deep multimodal learning techniques have proven successful in similar applications such as clinical diagnostics [12], land use cover classification [13], and wildfire surface fuel estimates [14]. But aggregating, organizing, and preparing biophysical data presents a high barrier to entry for carbon flux modelling. There are no datasets or benchmarks formulated for machine learning and satellite data needs to be collected separately.

This work seeks to incentivize algorithmic research into data-driven carbon flux modelling by providing the following:

- An introduction to carbon flux modelling for ML researchers
- CarbonSense, an ML-ready dataset using EC meteorological data and corresponding satellite data
- EcoPerceiver, a multimodal model based on the Perceiver architecture [15] which achieves state-of-the-art performance for carbon flux modelling

2 Carbon Flux Modelling

At its core, carbon flux modelling (CFM) is a regression problem. The carbon flux (target) is dependent on the ecosystem makeup and the meteorological conditions in that ecosystem. Since ecosystem makeup would be intractable to measure or model, we gather proxy data such as multispectral satellite imagery and general ecosystem taxonomy. In this section we discuss the common data sources for data-driven CFM.

Meteorological Data CFM meteorological data typically comes from in-situ EC stations. In addition to carbon fluxes, EC stations measure local atmospheric conditions such as wind velocity and direction, precipitation, surface temperature, soil moisture, and others. The exact number and type of variables depends on the site, but regional networks maintain a minimum mandatory set for researchers wishing to submit their data [4]. Most mandatory meteorological variables can be obtained at a global scale using reanalysis products such as ERA5 [16], which means models trained on EC data can typically be used for global inference.

Geospatial Data The most common products for CFM are based on Moderate Resolution Imaging Spectroradiometer (MODIS) data [17]. This satellite constellation produces new imagery for Earth's surface every 1-2 days and has 36 spectral bands with resolutions varying between 250m and 1km. The MCD43A4 derived product is particularly common - it fuses MODIS data in a 16-day sliding window to produce a single image each day. This not only helps to address cloud coverage, but produces nadir BRDF-adjusted reflectance (NBAR) images which remove angle effects from directional reflectance [18]. Each image therefore appears as it would from directly overhead at solar noon. MCD43A2 is also widely used, which contains categorical values for each pixel indicating snow and water cover [19]. It's typical to capture imagery from these products in a 4km by 4km square centered on the EC station [17] [6]. The terms "geospatial data", "satellite data" and "remote sensing data" are often used interchangeably.

Semantic Data Many researchers will also use semantic data when training models such as the plant functional type (PFT) of the area ("Croplands", "Evergreen needleleaf forest", "Snow and ice", etc). These classifications follow a standardized scheme such as the International Geosphere-Biosphere Programme (IGBP) or Leaf Area Index (LAI). EC site classification is performed by domain experts, but some MODIS products coarsely approximate PFT information on a global grid [20], allowing this data to also be used for global inference.

Targets Carbon fluxes can be described in gross or net terms, and cumulative or mechanism-specific. FLUXNET and similar aggregations typically require Net Ecosystem Exchange (NEE), Gross Primary Productivity (GPP), and Ecosystem Respiration (RECO). Methane flux (CH4) is also a common measurement [8] [7] but is typically handled separately from CO2. Non-carbon fluxes can also be modelled such as evapotranspiration (ET, a water vapour flux process) [6].

3 The CarbonSense Dataset

3.1 Data Collection

All meteorological data was aggregated from major EC data networks, including FLUXNET [4], the Integrated Carbon Observation System (ICOS) 2023 release [21], ICOS Warm Winter release [22], and Ameriflux 2023 release [23]. These networks were chosen due to their use of the ONEFlux processing pipeline [4], ensuring standardized coding and units. In total, CarbonSense contains data from 385 EC stations comprising over 27 million site-hours. North America and Europe are over-represented in the site list due greater data accessibility. A map of sites and their source networks is shown in Figure 2

Geospatial data in CarbonSense are sourced from Moderate Resolution Imaging Spectroradiometer (MODIS) products. This is the most common choice in flux modelling because it produces imagery daily at the cost of a lower spatial resolution. Specifically, we utilize the seven spectral bands from the MCD43A4 product [18], as well as the water and snow cover bands from MCD43A2 [19]. Following the guidelines from [17], we extract images in a 4km by 4km square centered on each EC station. Given a spatial resolution of 500m per pixel, this yields an 8x8 pixel image with 9 channels for every site-day.

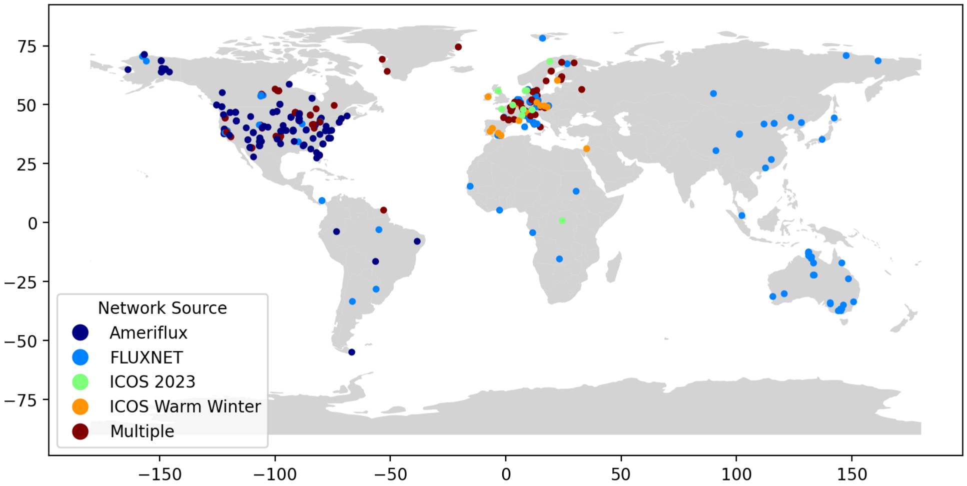

Figure 2: Global map of eddy covariance sites used in CarbonSense, with corresponding source networks. Some sites were present in multiple networks.

3.2 Data Pipeline

The first stage in the pipeline is EC data fusion. Many sites had overlapping data between different source networks, often presenting different measurements for the same timestep. In this case we used the values from the more recent publication as this indicates a more recent version of the underlying ONEFlux pipeline. Any sites which report half-hourly data were downsampled to hourly at this stage, and daily and monthly recordings were discarded.

Once fused, we extracted the relevant time blocks for each EC station along with its geographic location. This metadata was used to obtain the appropriate MODIS data for each site. Data was pulled procedurally from Google Earth Engine [24]. Each MODIS image was then reprojected to the Universal Transverse Mercator (UTM) projection zone appropriate for its site and cropped to 8x8 pixels centered on the EC tower. This reprojection is necessary to correct for the distortion in MODIS's native sinusoidal projection at extreme latitudes.

EC data underwent variable pruning to remove unneeded values. Most variables not available in global reanalysis products were removed, such as soil moisture and temperature at various depths. Quality check flags, which indicate the confidence of each measurement, were removed and placed in parallel files for use later in the pipeline. Finally, target variables for NEE, GPP, and RECO were selected. Raw values were used when available, and we used the two most common partitioning scheme variations of GPP and RECO ("daytime" and "nighttime" partitioning). A full list of variables at this stage are given in table **??**.

As a final and optional stage in the pipeline, we normalize the data in a way that is suited to our baseline model. We use a min-max normalization procedure on predictor variables. In particular, we map cyclic variables to the range $[-1, 1)$ and acyclic variables to the range $[-0.5, 0.5)$. Cyclic variables include wind direction, day of year, and time of day and have a set range. Acyclic variables were given a padding during min-max normalization to account for extreme values not present in this dataset. As an example, surface temperature was mapped from $[-80C, 80C)$ to $[-0.5, 0.5)$ which covers the most extreme temperatures measured on Earth. This normalization procedure is conducive to our Fourier encoding method discussed in section 4.1. We also discard any variables with a quality check value of 3 ("poor") or greater, indicating that the value was heavily imputed in the underlying pipeline. This exclusion increased performance of our baseline model.

We offer CarbonSense both normalized and unnormalized for those who wish to format the data differently. The pipeline code is also available so that researchers can add more data with minimal modifications. A diagram of the entire pipeline is shown in Figure 3.

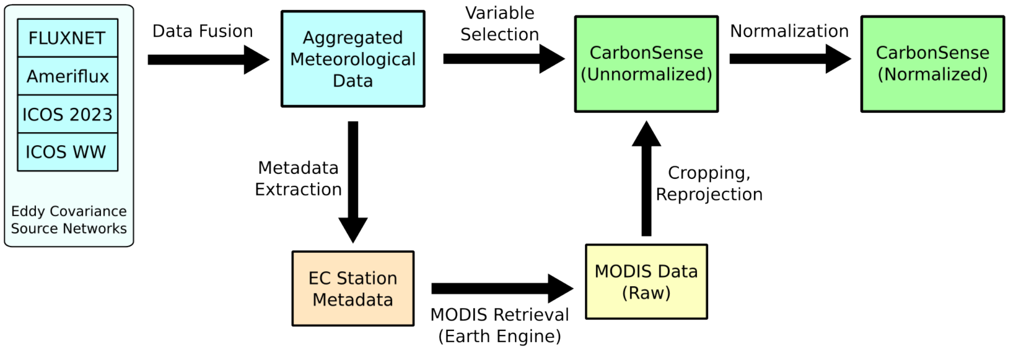

Figure 3: Data pipeline used to create CarbonSense from EC and MODIS data.

3.3 Using the Dataset

Site Sampling The geographic and ecological distribution of sites remains a challenge in statistical CFM, and CarbonSense is no different. Given the significant overrepresentation of certain regions

(North America, Europe) and ecosystems (evergreen needleleaf forests, grasslands), we maintain a partitioned structure where each site has its own directory containing EC data, geospatial data, and metadata. Researchers may choose to select sites for training and validation which allow for fair out-of-distribution performance estimates, or to achieve a balanced sampling of ecosystems.

Dataloader We supply an example PyTorch dataloader for CarbonSense specifically tailored to our baseline model. Using the dataloader requires specifying which carbon flux to use as the target, which sites to include in each dataloader instance, and the context window length for multi-timestep training.

Licensing CarbonSense is available under the CC-BY-4.0 license, meaning it can be shared, transformed, and used for any purpose given proper attribution. This is an extension of the same license for all three source networks, and MODIS data is provided under public domain. We feel that permissive licensing is essential in order to foster greater scientific interest in CFM in the deep learning community.

4 The EcoPerceiver Architecture

In this section we present EcoPerceiver, a multimodal architecture for CFM. The state-of-the-art for CFM are simple tabular methods, and we felt it would be appropriate to include a baseline model which demonstrates how deep learning concepts can be leveraged for this unique problem domain. EcoPerceiver is based on the Perceiver architceture [15] and was designed with the following principles:

Input Flexibility Different EC stations will measure different meteorological variables. Additionally, sensors will often fail and leave coverage gaps, or outlier values will be removed during post-processing. Rather than rely on gapfilling techniques, we chose a design which is robust to missing inputs.

Non-Markovian Processing All available CFM models treat carbon dynamics as a Markovian process; they assume an ecosystem's carbon uptake and respiration are determinable using only immediate information. However, biological processes do not follow this assumption. A plant's photosynthesis may depend on temperature, solar radiation, and precipitation levels spanning hours or days into the past (or further!). We have designed EcoPerceiver to condition its output on all data in a fixed context window.

Efficiency The downstream use of CFM is to produce local, regional, or global predictions of carbon fluxes. These are compute-heavy tasks often run by researchers who do not have access to large GPU clusters. We have designed EcoPerceiver with a modest parameter count and runtime, and provide guidelines for further reducing compute load in section 5.

4.1 Data Ingestion

Small fluctuations in meteorological variables have the potential to influence ecological processes. For this reason, it is important that the model is sensitive to small changes in input values. We take inspiration from NeRF's Fourier encoding [25] which maps continuous values to higher dimensional space with high frequency sinusoids. As discussed previously, cyclic variables in CarbonSense are mapped to $[-0.5, 0.5)$ and acyclic variables are mapped to $[-1.0, 1.0)$. We start by taking each variable x, and applying it to a series of sinusoids to produce an encoded vector with:

$$f(x) = \Big[\big[\dots, \sin(2^k \pi x), \cos(2^k \pi x), \dots \big] \Big| k \in [0, K) \Big] \tag{1}$$

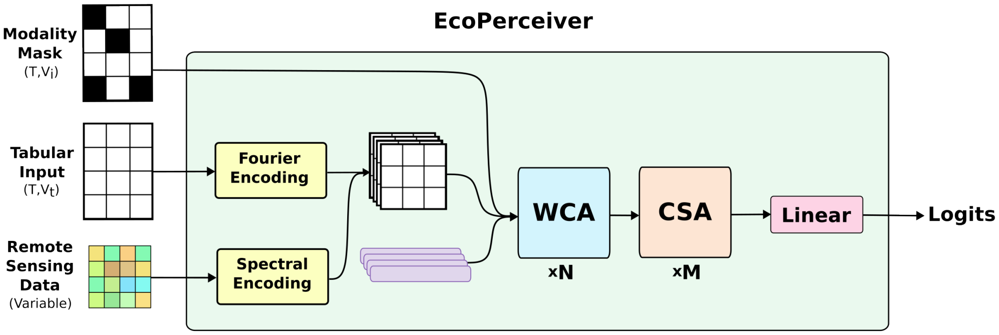

Figure 4: Data pipeline used to create CarbonSense from EC and MODIS data.

where K is a hyperparameter indicating the maximum sampling frequency. Higher values of K allow the model to better discern between small differences in input. With our normalization scheme, cyclic variables at values of -1 and 1 will produce identical vectors under this transform as intended.

Each input is given a learned embedding specific to the underlying variable. This is then concatenated with the Fourier encoding to produce a final input vector of length $H_i = 2K + l_{emb}$ for each input. It should be noted that EcoPerceiver requires a list of all possible observation types at creation time so that these embeddings can be created.

Geospatial data is similarly processed, except that each spectral band is flattened into a vector of length $2K$ via linear transformation instead of Fourier encoding. Each band is then given an embedding to produce a vector of length H_i.

The tabular and geospatial data are then stacked to create a matrix of shape (V_t, H_i) where V_t is the sum of the number of meteorological variables and spectral bands at timestep t. Since EcoPerceiver conditions on observations in a fixed context window of length T, the final data cube used for input to the attention layers is of shape (T, V_t, H_i). Figure 5 gives a visualization of the encoding procedure.

Not every timestep has a value for every variable, and this is certainly true of geospatial data which is typically provided once per 24 hours. To account for this, EcoPerceiver takes a modality mask indicating which values to ignore in the cross attentive layers.

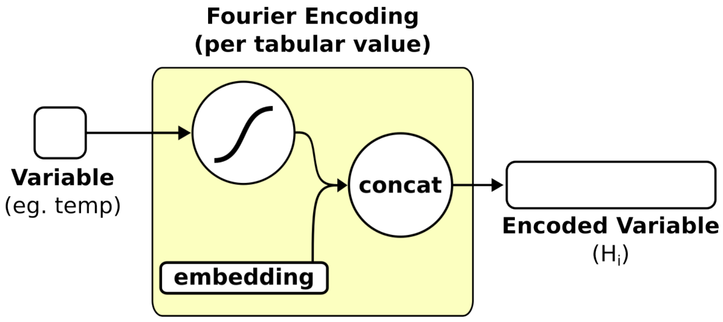

Figure 5: Input encoding for EcoPerceiver. Left: tabular values are fed into the Fourier encoding function (1) and concatenated with an embedding. Right: (INCOMPLETE DIAGRAM).

4.2 Windowed Cross Attention

We build on Perceiver's core concept of cross-attending data onto a compact latent space for processing. EcoPerceiver uses a latent space of size (T, H_l) where T is the context window length and H_l is the hidden token length (hyperparameter). Each token extracts input data via cross-attention from its respective timestep's observations. In this way each token can be thought of as representing the ecosystem's "state" in time, as it pertains to the carbon cycle.

This operation would be very inefficient with vanilla cross attention, as each token would use at most $\frac{1}{T}$ observations with an attention mask removing the rest. We take inspiration from SWin Transformer [26] and instead push the context window dimension (T) into the batch dimension for both input and latent space. The resulting Windowed Cross Attention (WCA) has a runtime of $O(T \cdot V_t \cdot H_a)$ where H_a is the projection dimension. Full derivation of this is given in Appendix [INSERT].

In keeping with Perceiver, each WCA operation is followed by a self-attention operation in the latent space. We pass a causal mask to the self attention so each timestep is conditioned only on past and present observations. We refer to this as Causal Self Attention (CSA). This constitutes a full WCA block as shown in Figure 6.

WCA blocks are repeated N times, repeatedly cross attending inputs onto the latent space with self attention in between. We then apply a series of M CSA operations and use the final timestep's token as input to a linear layer. The output of this is the estimate of the desired carbon flux.

4.3 Observational Dropout

Overconditioning on a small subset of variables could reduce EcoPerceiver's ability to generalize. Since our architecture is robust to missing observations, we implemented an observational dropout scheme to combat this. Given a hyperparameter $(0 < \epsilon < 1)$, we use the modality mask to randomly remove a portion of observations during training equal to ϵ (in addition to any missing observations). We find that this improves validation performance as seen in section 5

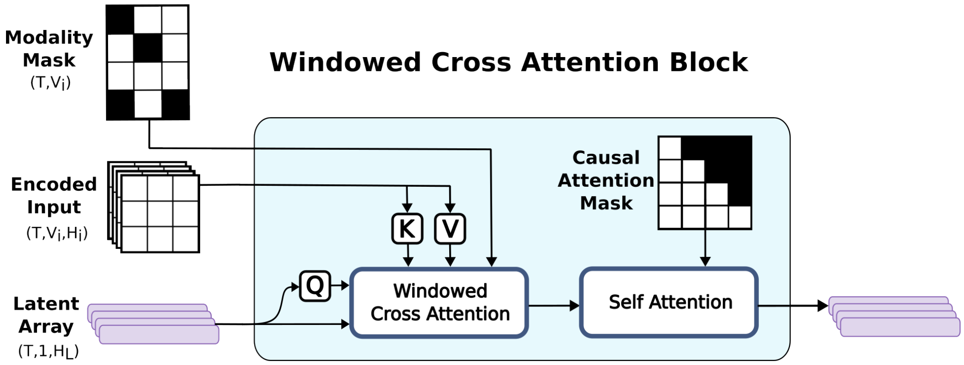

Figure 6: Windowed Cross Attention (WCA) block. Encoded inputs are cross-attended onto the latent space with a modality mask to indicate missing values. The time dimension is pushed into the batch dimension, so this operation is performed $B \cdot T$ times per batch. Causal self attention proceeds as normal.

5 Experiments

In this section, we present a series of experiments using CarbonSense. Our analysis includes two models: an EcoPerceiver model as introduced in 4, and an XGBoost model implemented to mimic current state-of-the-art approaches in CFM. We aim to demonstrate the power of tailored deep architectures for CFM and establish a robust baseline that will support and inspire future research efforts. We also present guidelines for running similar experiments and presenting results.

5.1 Data Splitting

EC stations are divided into train and validation sets based on their IGBP classification. We wanted to reserve at least 20% of each ecosystem in the validation set, but several ecosystems had fewer than 5 sites: snow and ice (SNO), water bodies (WAT), cropland/vegetation mosaics (CVM) and deciduous needleleaf forests (DNF). For these, we took the ceiling of 20% as validation, causing them to simulate few- or zero-shot generalization tests.

The main focus of this research is on models trained *across* different ecosystem types, as opposed to other research studying CFM *within a single* type (ex: [7] [10] [11]).

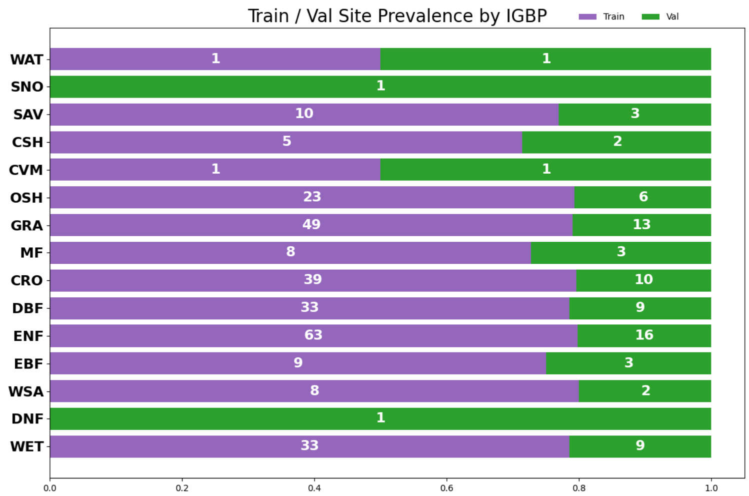

Figure 7: Distribution of sites by IGBP ecosystem classification in the training and validation splits.

5.2 Model Configurations

All experiments were performed on the Compute Canada cluster. To ensure accurate confidence intervals, we ran 10 experiments per model once optimal hyperparameters were found (as suggested by [27]).

EcoPerceiver experiments were each run on 4 A100 GPUs using dataset parallelization. We used the AdamW optimizer [28] with a learning rate of 8e-5 and a batch size of 4096. A single warm-up epoch was performed followed by a cosine annealing learning rate schedule over 20 epochs. Most experiments converged between 4 and 10 epochs. Hyperparameters were initially selected by hand and then refined by experimental perturbation; methods such as random search were infeasible due to the large hyperparameter space, compute constraints, and lack of similar experiments to draw inspiration from. A full description of the training configuration is given in Appendix C.

XGBoost experiments were run on CPU nodes. Hyperparameters were found by random search over 50 iterations before formal experiments were performed. Since XGBoost is a tabular algorithm, we prepared geospatial data in a similar fashion to XBASE [6]; each spectral band represents a single input value to the model. The value is obtained by taking a weighted average of pixels based on Euclidean distance from the center of the image. The code for this processing, as well as the final list of hyperparameters, is provided in the supplementary material.

5.3 Reproducibility and Reliability

Both EcoPerceiver and XGBoost were trained with reproducibility in mind. Once optimal hyperparameters were found, we performed 10 experiments with each model in order to obtain a reliable measure of performance (inspired by [27]). Set seeds were provided to all frameworks utilizing RNG, and distributed dataloader workers were also seed-controlled to ensure full reproducibility of our results.

5.4 Metrics

The most commonly used performance metric in CFM (and any form of hydrologic modelling) is the Nash-Sutcliffe Modelling Efficiency (NSE) [29], described with the following equation:

$$\mathrm{NSE}(x) = 1 - \frac{\sum_i (y_i - x_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

where a value of 1 represents perfect correlation between x and y. A value of 0 represents the same as always guessing the mean of y, and negative values indicate that the mean of y is a better predictor than x. NSE is more challenging to use directly as a loss function since it would require the dataloader to also provide the mean of the data for a given site or ecosystem type. We therefore use root mean squared error (RMSE) as a loss function, report both metrics in our results, and encourage future researchers to do the same.

Data balance in results reporting is also a concern. At first glance, the data appears very imbalanced with respect to ecosystem prevalence. CarbonSense contains 62 grassland sites, but only 1 deciduous needleleaf forest site. It should be noted that IGBP ecosystem taxonomies are extremely broad - grasslands in central North America will differ significantly from those in Europe or Asia. Still, it is prudent to separate results by ecosystem type to give a better idea picture of model performance.

5.5 Results

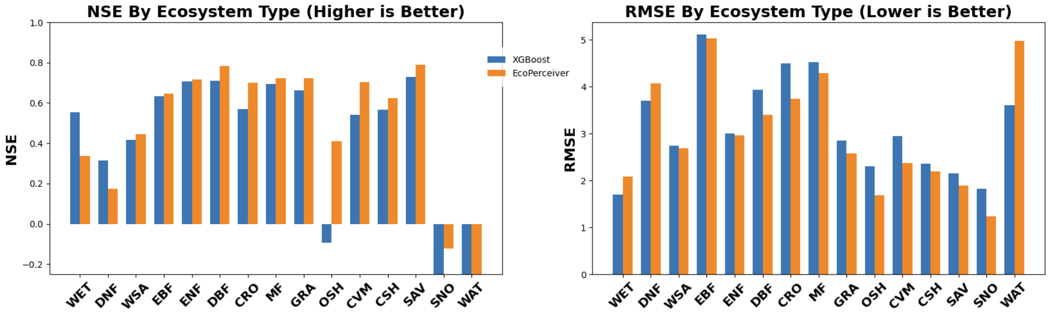

Figure 8: NSE (left) and RMSE (right) of the models, partitioned by ecosystem type.

EcoPerceiver consistently outperformed the XGBoost baseline across most ecosystem types. Low-volume ecosystems SNO, WAT, CVM, and DNF were a tossup, indicating that the two models displayed similar few-shot performance. In ecosystems with higher prevalence in the training set, XGBoost performed worse in all except permanent wetlands (WET).

References

[1] D. PAPALE and R. VALENTINI, “A new assessment of european forests carbon exchanges by eddy fluxes and artificial neural network spatialization,” *Global Change Biology*, vol. 9, no. 4, pp. 525–535, 2003. DOI: https://doi.org/10.1046/j.1365-2486.2003.00609.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2486.2003.00609.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2486.2003.00609.x.

[2] M. Jung, M. Reichstein, H. A. Margolis, *et al.*, “Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations,” *Journal of Geophysical Research: Biogeosciences*, vol. 116, no. G3, 2011. DOI: https://doi.org/10.1029/2010JG001566. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JG001566. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JG001566.

[3] A. Anav, P. Friedlingstein, M. Kidston, *et al.*, “Evaluating the land and ocean components of the global carbon cycle in the cmip5 earth system models,” *Journal of Climate*, vol. 26, no. 18, pp. 6801–6843, 2013. DOI: 10.1175/JCLI-D-12-00417.1. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00417.1.xml.

[4] G. Pastorello, C. Trotta, E. Canfora, *et al.*, “The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data,” *Scientific Data*, vol. 7, no. 1, p. 225, Jul. 2020, ISSN: 2052-4463. DOI: 10.1038/s41597-020-0534-3. [Online]. Available: https://doi.org/10.1038/s41597-020-0534-3.

[5] G. Tramontana, M. Jung, C. R. Schwalm, *et al.*, “Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms,” *Biogeosciences*, vol. 13, no. 14, pp. 4291–4313, 2016. DOI: 10.5194/bg-13-4291-2016. [Online]. Available: https://bg.copernicus.org/articles/13/4291/2016/.

[6] J. A. Nelson, S. Walther, F. Gans, *et al.*, “X-base: The first terrestrial carbon and water flux products from an extended data-driven scaling framework, fluxcom-x,” *EGUsphere*, vol. 2024, pp. 1–51, 2024. DOI: 10.5194/egusphere-2024-165. [Online]. Available: https://egusphere.copernicus.org/preprints/2024/egusphere-2024-165/.

[7] O. Peltola, T. Vesala, Y. Gao, *et al.*, “Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations,” *Earth System Science Data*, vol. 11, no. 3, pp. 1263–1289, 2019. DOI: 10.5194/essd-11-1263-2019. [Online]. Available: https://essd.copernicus.org/articles/11/1263/2019/.

[8] G. McNicol, E. Fluet-Chouinard, Z. Ouyang, *et al.*, “Upscaling wetland methane emissions from the fluxnet-ch4 eddy covariance network (upch4 v1.0): Model development, network assessment, and budget comparison,” *AGU Advances*, vol. 4, no. 5, e2023AV000956, 2023, e2023AV000956 2023AV000956. DOI: https://doi.org/10.1029/2023AV000956. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023AV000956. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023AV000956.

[9] A.-M. Virkkala, B. M. Rogers, J. D. Watts, *et al.*, “An increasing arctic-boreal co2 sink despite strong regional sources,” *bioRxiv*, 2024. DOI: 10.1101/2024.02.09.579581. [Online]. Available: https://www.biorxiv.org/content/early/2024/02/12/2024.02.09.579581.

[10] A.-M. Virkkala, J. Aalto, B. M. Rogers, *et al.*, “Statistical upscaling of ecosystem co2 fluxes across the terrestrial tundra and boreal domain: Regional patterns and uncertainties,” *Global Change Biology*, vol. 27, no. 17, pp. 4040–4059, 2021. DOI: https://doi.org/10.1111/gcb.15659. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15659.

[11] C. Zhang, D. Brodylo, M. Rahman, M. A. Rahman, T. A. Douglas, and X. Comas, “Using an object-based machine learning ensemble approach to upscale evapotranspiration measured from eddy covariance towers in a subtropical wetland,” *Science of The Total Environment*,

vol. 831, p. 154 969, 2022, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2022.154969. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969722020629.

[12] H.-Y. Zhou, Y. Yu, C. Wang, *et al.*, "A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics," *Nature Biomedical Engineering*, vol. 7, no. 6, pp. 743–755, Jun. 2023, ISSN: 2157-846X. DOI: 10.1038/s41551-023-01045-x. [Online]. Available: https://doi.org/10.1038/s41551-023-01045-x.

[13] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (exvit) for land use and land cover classification: A multimodal deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023. DOI: 10.1109/TGRS.2023.3284671.

[14] M. Alipour, I. La Puma, J. Picotte, *et al.*, "A multimodal data fusion and deep learning framework for large-scale wildfire surface fuel mapping," *Fire*, vol. 6, no. 2, 2023, ISSN: 2571-6255. DOI: 10.3390/fire6020036. [Online]. Available: https://www.mdpi.com/2571-6255/6/2/36.

[15] A. Jaegle, F. Gimeno, A. Brock, O. Vinyals, A. Zisserman, and J. Carreira, "Perceiver: General perception with iterative attention," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, 18–24 Jul 2021, pp. 4651–4664.

[16] H. Hersbach, B. Bell, P. Berrisford, *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020. DOI: https://doi.org/10.1002/qj.3803. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

[17] S. Walther, S. Besnard, J. A. Nelson, *et al.*, "Technical note: A view from space on global flux towers by modis and landsat: The fluxneteo data set," *Biogeosciences*, vol. 19, no. 11, pp. 2805–2840, 2022. DOI: 10.5194/bg-19-2805-2022. [Online]. Available: https://bg.copernicus.org/articles/19/2805/2022/.

[18] C. Schaaf and Z. Wang, "MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006," 2015b. [Online]. Available: https://www.umb.edu/spectralmass/v006/mcd43a4-nbar-product/.

[19] C. Schaaf and Z. Wang, "MCD43A2 MODIS/Terra+Aqua BRDF/Albedo Quality Daily L3 Global - 500m V006," 2015a. [Online]. Available: https://www.umb.edu/spectralmass/v006/mcd43a2-albedo-product/.

[20] D. Sulla-Menashe and M. A. Friedl, "User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product," May 2018. [Online]. Available: https://www.umb.edu/spectralmass/v006/mcd43a4-nbar-product/.

[21] ICOS RI, F. Apadula, S. Arnold, *et al.*, "ICOS Atmosphere Release 2023-1 of Level 2 Greenhouse Gas Mole Fractions of CO2, CH4, N2O, CO, meteorology and 14CO2, and flask samples analysed for CO2, CH4, N2O, CO, H2 and SF6.," 2023. DOI: https://doi.org/10.18160/VXCS-95EV. [Online]. Available: https://www.icos-cp.eu/data-products/atmosphere-release.

[22] Warm Winter 2020 Team and ICOS Ecosystem Thematic Centre, "Warm Winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-Archive format—release 2022-1 (Version 1.0).," *ICOS Carbon Portal*, 2022. DOI: 10.18160/VXCS-95EV. [Online]. Available: https://www.icos-cp.eu/data-products/2G60-ZHAK.

[23] H. Chu, D. S. Christianson, Y.-W. Cheah, *et al.*, "Ameriflux base data pipeline to support network growth and data sharing," *Scientific Data*, vol. 10, no. 1, p. 614, Sep. 2023, ISSN: 2052-4463. DOI: 10.1038/s41597-023-02531-2. [Online]. Available: https://doi.org/10.1038/s41597-023-02531-2.

[24] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, 2017. DOI: `10.1016/j.rse.2017.06.031`. [Online]. Available: `https://doi.org/10.1016/j.rse.2017.06.031`.

[25] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, 2020.

[26] Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[27] R. Agarwal, M. Schwarzer, P. S. Castro, A. C. Courville, and M. Bellemare, "Deep reinforcement learning at the edge of the statistical precipice," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[28] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *7th International Conference on Learning Representations, (ICLR) 2019, New Orleans, LA, USA, May 6-9, 2019*, OpenReview.net, 2019. [Online]. Available: `https://openreview.net/forum?id=Bkg6RiCqY7`.

[29] R. McCuen, Z. Knight, and A. Cutter, "Evaluation of the nash–sutcliffe efficiency index," *Journal of Hydrologic Engineering - J HYDROL ENG*, vol. 11, Nov. 2006. DOI: `10.1061/(ASCE)1084-0699(2006)11:6(597)`.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **[TODO]**
   (b) Did you describe the limitations of your work? **[TODO]**
   (c) Did you discuss any potential negative societal impacts of your work? **[TODO]**
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? **[TODO]**
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)? **[TODO]**
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **[TODO]**
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **[TODO]**
   (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)? **[TODO]**
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 3.1 for data attributions. TODO: Appendix with all sites
   (b) Did you mention the license of the assets? [Yes] Section 3.3 has info on licensing of EC data and MODIS data
   (c) Did you include any new assets either in the supplemental material or as a URL? **[TODO]**
   (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]

Appendix A: Eddy Covariance Site Details

Appendix B: WCA Runtime Proof

Appendix C: EcoPerceiver Baseline Implementation Details