

SSL4Eco: A Global Seasonal Dataset for Geospatial Foundation Models in Ecology

Elena Plekhanova¹

elena.plekhanova@wsl.ch

Damien Robert²

damien.robert@uzh.ch

Johannes Dollinger²

johannes.dollinger@uzh.ch

Emilia Arens²

emilia.arenis@uzh.ch

Philipp Brun¹

philipp.brun@wsl.ch

Jan Dirk Wegner²

jandirk.wegner@uzh.ch

Niklaus Zimmermann¹

niklaus.zimmermann@wsl.ch

¹Land Change Science, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

²DM3L, University of Zurich, Zurich, Switzerland

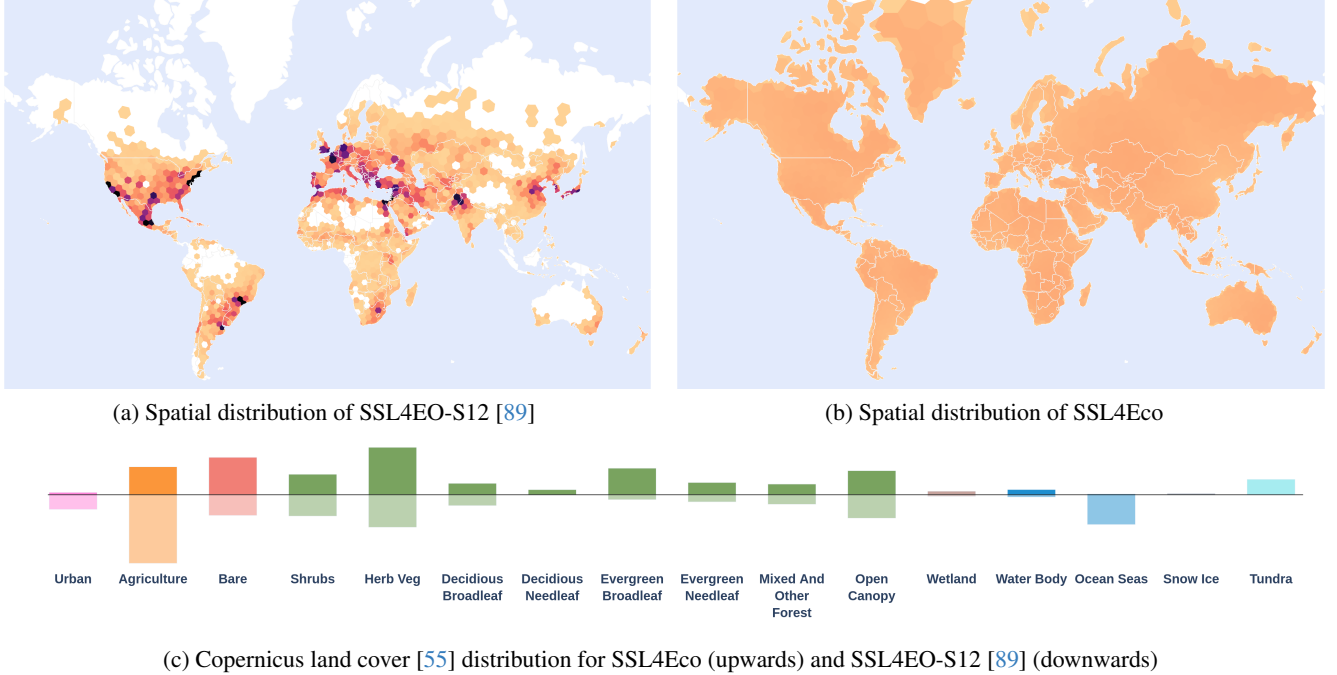


Figure 1. We propose SSL4Eco, a multi-date Sentinel-2 dataset for pretraining foundation models targeted for macroecological applications. Unlike comparable datasets (a), SSL4Eco uniformly covers the entire landmass (b), thus capturing all environment types without favoring urban and agricultural areas, or ignoring entire ecoregions (c).

Abstract

With the exacerbation of the biodiversity and climate crises, macroecological pursuits such as global biodiversity mapping become more urgent. Remote sensing offers a wealth of Earth observation data for ecological studies, but the scarcity of labeled datasets remains a major challenge. Recently, self-supervised learning has enabled learning representations from unlabeled data, triggering the de-

velopment of pretrained geospatial models with generalizable features. However, these models are often trained on datasets biased toward areas of high human activity, leaving entire ecological regions underrepresented. Additionally, while some datasets attempt to address seasonality through multi-date imagery, they typically follow calendar seasons rather than local phenological cycles. To better capture vegetation seasonality at a global scale, we pro-

pose a simple phenology-informed sampling strategy and introduce corresponding SSL4Eco, a multi-date Sentinel-2 dataset, on which we train an existing model with a season-contrastive objective. We compare representations learned from SSL4Eco against other datasets on diverse ecological downstream tasks and demonstrate that our straightforward sampling method consistently improves representation quality, highlighting the importance of dataset construction. The model pretrained on SSL4Eco reaches state of the art performance on 7 out of 8 downstream tasks spanning (multi-label) classification and regression. We release our code, data, and model weights to support macroecological and computer vision research at

<https://github.com/PlekhanovaElena/ssl4eco>.

1. Introduction

Biodiversity is essential for ecosystem stability and human well-being, yet it faces an unprecedented crisis due to habitat loss and climate change [8]. Recognized as a global priority (SDG 15) [63], biodiversity loss ranks among the most severe risks of the next decade [28]. This intensifying crisis calls for macroecological studies to understand spatiotemporal biodiversity patterns and identify priority areas for conservation [8]. Mapping changes in biodiversity, habitats, and land-use (e.g. deforestation, urban or agricultural expansion) over time is essential for conservation planning [36, 37]. Central to these efforts is monitoring vegetation change, as vegetation forms the primary structure of most terrestrial ecosystems and shapes biodiversity patterns and ecosystem functions [85].

Remote sensing is a powerful tool for monitoring vegetation change at broad spatial and temporal scales [92]. It provides consistent, repeated, global observations, enabling the detection of subtle shifts in vegetation health, species composition, and phenology—insights often unattainable through ground-based methods [25, 27]. Several open-access satellite products support vegetation monitoring, each with distinct strengths and limitations (see Appendix A-1). This work focuses on Sentinel-2 due to its widespread use for large-scale vegetation monitoring [46, 52, 80], but our conclusions remain applicable and may be extended to other satellite products.

To extract ecological insights from remote sensing data, initial approaches relied on handcrafted features and classical machine learning [6, 34]. Deep learning has since revolutionized the field by automating feature extraction for tasks with annotated datasets [56, 97]. Recently, self-supervised learning (SSL) has gained traction for learning rich representations from large, unlabeled datasets [35], with successful applications in the analysis of natural language [19], natural images [66], and remote sensing data [4]. The resulting pretrained models produce representations that generalize to downstream tasks, making these

so-called *foundation models* (FMs) particularly suitable for applications where labeled data is scarce or costly, such as large-scale ecological studies [88].

The size and diversity of the pretraining dataset largely influences the generalizability of the learned representations [72, 96]. While research on geospatial FMs operating on georeferenced data (GFMs) is an active field of study, most effort is currently geared towards new model architectures and SSL pretraining tasks, and little attention is given to the design of pretraining datasets. This oversight is critical, as the geographical distribution of training data significantly influences model performance [71, 74]. For biodiversity applications in particular, existing GFMs are often trained on datasets which fail to capture important spatiotemporal ecological patterns, as summarized in Table 1. First, the geographic sampling is often biased towards human activity, hence over-representing urban and agricultural areas while neglecting entire biomes. Second, multi-temporal datasets are typically sampled following calendar seasons, failing to account for local phenological cycles, essential to biodiversity monitoring.

In this work, we propose a dataset construction recipe targeted towards the development of foundation models for ecology. Specifically, we propose to sample locations uniformly across the landmass, rather than around large urban areas [58, 89], and sample dates based on local phenological cycles, rather than calendar seasons [58, 89]. Following this protocol, we introduce SSL4Eco, a pretraining dataset of multispectral, multi-date Sentinel-2 patches of 256×256 pixels, uniformly sampled across 250k locations around the globe and capturing local phenology, as shown in Figure 1 and Figure 2. From SSL4Eco, we derive SeCo-Eco, a seasonality-aware SeCo [58] model, and compare its embeddings against off-the-shelf GFMs on diverse macroecological tasks. We show that SeCo-Eco equals or exceeds the performance of all other baselines on 7 out of 8 downstream tasks spanning (multi-label) classification and regression, with larger gaps of +2 mAP on BigEarthNet-10% [82] and +3 to +4 R^2 in regression of climatic variables and biomass.

Far from claiming a new SSL training or backbone, this work stresses the importance of dataset design, and how a straightforward spatiotemporal sampling protocol may consistently benefit GFMs downstream applications. We publicly release our datasets, code, and weights at <https://github.com/PlekhanovaElena/ssl4eco>, hoping to foster both downstream macroecological studies and methodological computer vision research with a concern for environmental applications. The contributions of this work are as follows:

- SSL4Eco: a novel multi-temporal Sentinel-2 pre-training dataset with uniform global distribution and vegetation phenology-based seasonal sampling.

Dataset	Locations		Seasons
	Number	Distribution	
BigEarthNet [82]	600k	Europe	-
SEN12MS [77]	280k	Around cities	Calendar
SeCo [58]	200k	Around cities	Random
S2-100k [49]	100k	Global uniform	-
Planted [68]	3.0M	Semi global	-
SatlasPretrain [5]	3.0M	Semi global	-
SSL4EO-S12 [89]	250k	Around cities	Calendar
MajorTOM-Core [29]	2.2M	Global uniform	-
SSL4Eco	250k	Global grid	EVI-based

Table 1. Comparison of the spatiotemporal sampling of popular pretraining datasets for geospatial foundation models. Our sampling of SSL4Eco is designed to fully capture both global geographic diversity and local climatic and phenological seasonality.

- SeCo-Eco: a seasonality-aware geospatial foundation model pretrained on SSL4Eco.
- New macroecological downstream tasks for benchmarking geospatial foundation models.

2. Related Work

In this section, we provide an overview of existing remote sensing datasets used for pretraining geospatial foundation models, with a focus on their spatiotemporal distribution. We then introduce several such foundation models relevant to this work.

Pretraining Remote Sensing Datasets. Numerous labeled datasets have been proposed to employ remote sensing imagery for mapping urban or agricultural landscapes [32, 33, 78]. However, these generally do not offer the spatial and seasonal coverage necessary to macroecological research, as summarized in Table 1. Existing unlabeled pretraining datasets for SSL models focus predominantly on regions experiencing high human impact, often neglecting areas crucial for ecological research and conservation. For instance, SEN12MS [77], SeCo [58], and SSL4EO-S12 [89] datasets are sampled around large cities, mainly encompassing urban and agricultural zones (Figure 1). While BigEarthNet [82] does sample diverse vegetation types, it only covers Europe. Other datasets such as SatlasPretrain [5], S2-100K [49], and Planted [68] have better geographic coverage, but with significant gaps in the tropics due to high cloud coverage and either undersample or ignore Arctic tundra entirely. Yet, tropical rainforests harbor the highest levels of biodiversity on the globe [8], making their underrepresentation in training datasets problematic. Similarly, the Arctic region is central to many environmental processes such as the thawing of Arctic permafrost which introduces one of the greatest uncertainties in current climate models [79]. Interestingly, Major-TOM-Core [29]

Model	Dataset	Backbone	Pretraining
SeCo [58] [58]	SeCo [58]	ResNet50 [38]	SeCo [58]
SatMAE [16]	fMoW [13, 16]	ViT-L [20]	MAE [40]
Satlas [5]	SatlasPretrain [5]	Swin-B [53]	Supervised
Croma [31]	SSL4EO-S12 [89]	ViT-L [20]	MAE [40]
SSL4EO [89]	SSL4EO-S12 [89]	ResNet50 [38]	MoCov2 [11]
DOFA [93]	DOFA [93]	ViT-L [20]	MAE [40]
SeCo-Eco	SSL4Eco	ResNet50 [38]	SeCo [58]

Table 2. Overview of recent image-based geospatial foundation models. We focus on models trained to process Sentinel-2 data, for fair comparison with our pretraining setting. While we release a new pretrained image-based model SeCo-Eco, our focus is *not* on the design of a backbone or pretraining method, but on the impact of the pretraining dataset.

uniformly covers the entire landmass, but at a single date, failing to capture seasonality. Despite the importance of seasonality for ecosystems and ecological research [43], few datasets provide multi-temporal imagery at each location. SeCo [58] randomly selects 5 dates across the year separated by approximately 3 months. SEN12MS [77] and SL4EO-S12 [89] select 4 dates within seasonal windows defined based on calendar dates. However, these sampling approaches treat all locations equally, resulting in datasets that overlook the reality of local climatic and ecological conditions. Indeed, regions near the tropics may have longer leaf-on seasons, while desert or Arctic regions may see very brief events of vegetation activity with a large portion of the year being dry or snow-covered. Likewise, the beginning and end of dormancy periods may be shifted in the year, depending on local climatic conditions. In this work, we propose a simple sampling strategy that fully covers the global diversity of landscapes (Figure 1) and local seasonality (Figure 2). Our goal is to design datasets for learning representation better suited for downstream ecological applications.

Geospatial Foundation Models. Advances in self-supervised learning have recently allowed to learn generalizable representations from the wealth of public, unlabeled, satellite imagery [4, 88]. Masked image modeling [40] methods typically leverage symmetries inherent to remote sensing data to reconstruct masked spectral bands [16], time steps [23], both [42], or other modalities [3, 31, 64, 84]. Alternatively, contrastive approaches [10, 11, 39] learn to align latent representations of imagery from different seasons [58, 89] or modalities [2]. Another direction learns implicit geolocation representations by aligning spatial coordinates with terrestrial [87] or satellite [49] imagery, or species occurrences [15]. Moving beyond the focus on the design of self-supervised method or model architecture, our work sheds light on the importance of pretraining datasets. We use existing SSL methods to pretrain on our SSL4Eco

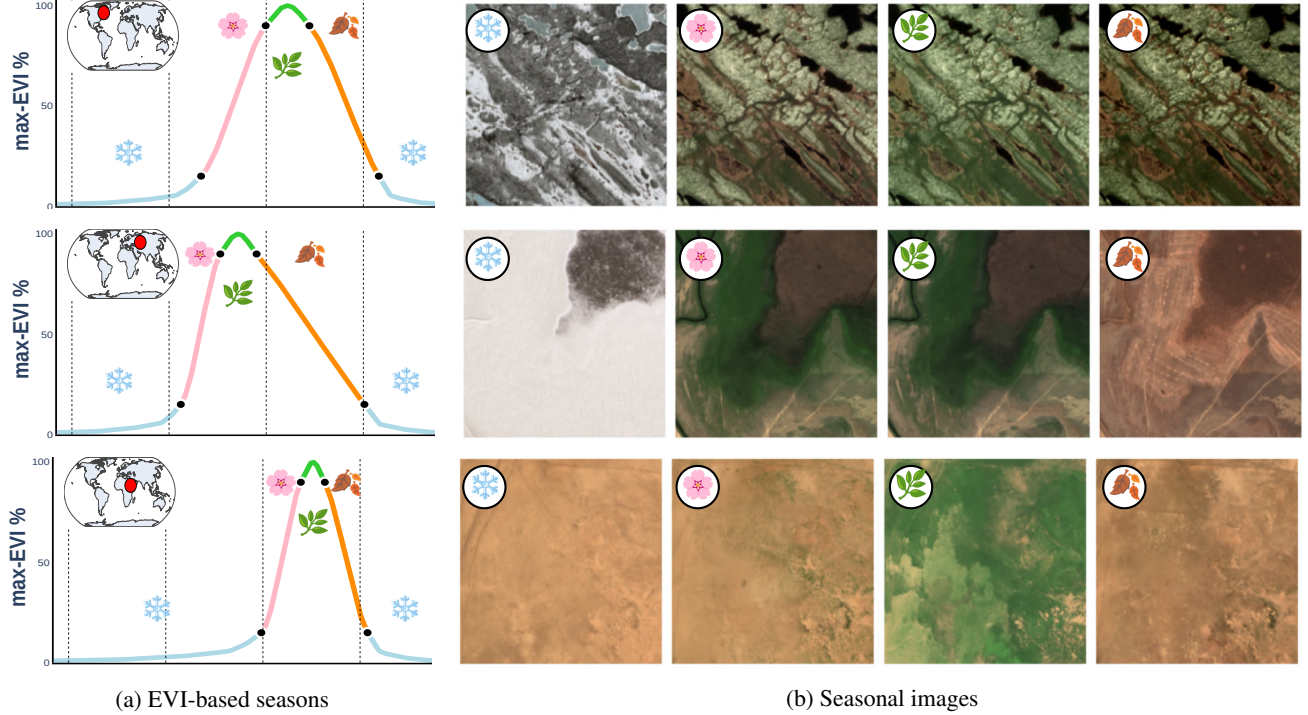


Figure 2. Unlike previous works which sample seasonal images based on calendar dates [58, 77, 89] (dashed lines in (a)), we define phenology-informed, local seasons based the Enhanced Vegetation Index [44, 45] (colored sections in (a)). As a result, our SSL4Eco dataset covers the full cycle of vegetation activity at each location (b), capturing patterns otherwise missed by calendar sampling.

dataset and analyze resulting representations with available comparable image-based geospatial foundation models, as summarized in Table 2.

3. Method

We detail our proposed dataset construction approach in Section 3.1 and pretrained model in Section 3.2.

3.1. SSL4Eco Dataset

Our dataset sampling recipe aims at capturing phenology-informed patterns anywhere on Earth. For more details on our dataset construction protocol, please see Section A-1.

Spatial Sampling. Similar to Major-TOM [29], we uniformly sample geolocations across the globe using a regular grid, accounting for distortions long the latitude. We only sample positions across the landmass, with a 23 km spacing between points, yielding 250k geolocations. This sampling size is chosen to allow comparison with similar pretraining datasets [58, 89]. As shown in Figure 1, the resulting dataset follows the natural distribution of land use, without focusing on urban or agricultural areas.

Seasonal Sampling. Vegetation seasonality primarily depends on local temperature and light regimes, themselves primarily driven by latitude [14, 67], altitude [48, 50],

and rainfall seasonality [12, 22] (*e.g.* monsoon regions, or Mediterranean and Savannah biomes). To capture local seasonality, we sample 4 dates at each location. Unlike previous works which define seasons globally based on calendar dates [58, 77, 89], we sample based on local plant phenology. To this end, we use the Enhanced Vegetation Index (EVI) from the MCD12Q2 v6.1 [30] product of the MODIS [45] satellite mission. For each location, we define the 4 seasons spring, summer, autumn, and winter as intervals between the Greenup, Maturity, Senescence, and Dormancy variables (see Sec. A-3 for details). By sampling a date in each of these phenological seasons, we aim to better seize the diversity of vegetation states at each location than calendar or random sampling, as illustrated in Figure 2.

Modality. We apply our spatiotemporal sampling strategy to create SSL4Eco, a global, multi-temporal dataset of satellite imagery. We choose to use Sentinel-2 [70] images for their superior spatiotemporal resolution and widespread use in vegetation monitoring [46, 52, 70, 80]. In addition to the 12 spectral bands of Sentinel-2, SSL4Eco carries an NDVI band, which is widely used as a proxy of vegetation productivity and biomass [69]. While the present work focuses on demonstrating the impact of dataset sampling on Sentinel-2, our dataset construction and sampling analysis could naturally be extended to other modalities relevant to ecology

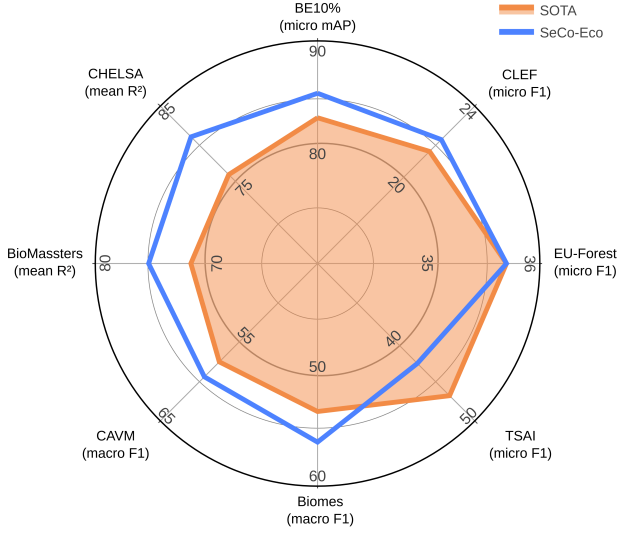


Figure 3. Linear Probing performance across all datasets. We compare SeCo-Eco against the respective best-performing model among our reported set of baselines.

such as optical [45, 91], SAR [75, 83], LiDAR [21] sensors, or species [24], and climate [41, 47] observations, which we leave for future work.

Patching. Similar to previous works [58, 89] we choose a patch size is 256×256 pixels (2.56×2.56 km). The exact amount of retrieved locations is 254 403, each having up to 4 dates, yielding a total of 1M patches, for a final dataset size of 1.3 TB.

3.2. SeCo-Eco Model

In this section, we introduce our SeCo-Eco model, trained with a seasonal contrastive objective on SSL4Eco. We stress that our current focus is on pretraining dataset construction, not novel SSL training or backbone.

Our self-supervised training objective needs to capture spatial and seasonal patterns in our multi-temporal data. SSL4EO [89] proposes a seasonal contrastive objective, but encourages learning season-agnostic features. Instead, we use Seasonal Contrast (SeCo) [58], which learns both season-agnostic and season-specific representations, more appropriate for seasonality-sensitive tasks. We use ResNet50 as our image encoding backbone, which has proven to be robust across a variety of remote sensing tasks [49, 58, 88, 89]. We dub SeCo-Eco our resulting model trained on SSL4Eco. We also explore in Section 4 another version, MoCo-Eco, pretrained using the seasonal contrastive objective from SSL4EO.

We pretrain for 100 epochs, with a batch size of 256 on a single A100 GPU (7 days for SeCo-Eco, 4 days for MoCo-Eco). See Section A-2 for more implementation details.

4. Experiments

We present in Section 4.1 the downstream tasks we use for comparing geospatial foundation models, present experimental results in Section 4.2, and ablations in Section 4.3.

4.1. Downstream Tasks and Evaluation

Several benchmarks for geospatial foundation models have been proposed [26, 51, 59, 94], but none fit to our current setting of Sentinel-2, image-level representations for ecological applications. Hence, we leverage existing datasets and propose new ones to evaluate the SSL4Eco pretraining.

Protocol. We compare embeddings from SeCo-Eco with other geospatial foundation models operating on Sentinel-2 input data. Our choice of benchmarked methods is driven by the availability of reproducible code at the time of writing. For each model, we use the official implementation and adjust Sentinel-2 bands selection and normalization based on their respective pretraining setting. Following Wang *et al.* [89], we also crop or stretch the input patch size to align with the pretraining conditions. We evaluate embeddings both with linear probing (LP) and K-Nearest Neighbor (k-NN) approaches. For LP, we freeze the model backbone and train a single linear layer on top. We use the AdamW [54] optimizer, train for up to a 1000 epochs with early stopping, a learning rate of $1e^{-3}$, a batch size of 256, on an NVIDIA T4 GPUs. For k-NN, we follow existing literature [9, 90] and aggregate labels from the k-nearest neighbors in the training set, based on cosine similarity. We use a softmax temperature of 0.07 and grid-search k for each task. Unless specified otherwise for the task, we always split our data in 10 test folds and randomly sample training and validation sets from the remaining data with [0.9, 0.1] split. We re-

Model	Biomes (macro F1) [↑]		CAVM (macro F1) [↑]	
	LP	10-NN	LP	20-NN
SeCo [58]	41.5 ± 0.5	36.9 ± 1.0	54.4 ± 0.7	52.1 ± 0.7
SatMAE [16]	51.3 ± 1.1	47.7 ± 0.7	56.3 ± 1.4	55.8 ± 0.7
Satlas [5]	48.3 ± 1.6	47.6 ± 0.9	53.8 ± 2.0	53.2 ± 0.5
Croma [31]	47.1 ± 1.4	42.2 ± 0.6	53.6 ± 1.2	51.6 ± 0.8
SSL4EO [89]	<u>53.3</u> ± 1.0	<u>49.7</u> ± 0.5	<u>57.5</u> ± 0.6	<u>56.9</u> ± 0.6
DOFA [93]	49.7 ± 1.3	42.9 ± 0.5	56.4 ± 1.6	53.5 ± 0.6
SeCo-Eco (ours)	56.1 ± 0.7	51.1 ± 0.9	59.4 ± 1.0	59.5 ± 0.8

Table 3. Linear probing and K-Nearest Neighbor comparison of state of the art models with our SeCo-Eco pretrained on our SSL4Eco on classification of two land cover datasets: global biomes and Arctic vegetation types [73]. **Best**, second best.

Model	BE10% (micro mAP) [↑]		CLEF (micro F1) [↑]		EU-Forest (micro F1) [↑]		TSAI (micro F1) [↑]	
	LP	30-NN	LP	1-NN	LP	5-NN	LP	5-NN
SeCo [58]	79.2 ± 0.0	77.8 ± 0.1	20.8	12.3	31.3 ± 0.7	30.6 ± 0.2	23.4 ± 0.0	35.2
SatMAE [16]	79.7 ± 0.2	79.6 ± 0.0	21.6	13.6	35.7 ± 1.0	33.3 ± 0.1	46.8 ± 0.3	43.7
Satlas [5]	77.9 ± 0.2	77.9 ± 0.0	18.9	11.8	30.0 ± 0.2	30.0 ± 0.2	42.9 ± 0.0	40.8
Croma [31]	80.7 ± 0.2	79.1 ± 0.0	20.8	12.0	32.2 ± 0.9	30.1 ± 0.2	<u>43.8</u> ± 0.0	40.7
SSL4EO [89]	<u>83.2</u> ± 0.1	<u>81.1</u> ± 0.0	<u>21.7</u>	12.6	32.6 ± 0.1	31.5 ± 0.2	42.3 ± 0.0	<u>40.9</u>
DOFA [93]	80.1 ± 0.0	77.3 ± 0.1	20.3	12.1	<u>34.8</u> ± 0.9	30.0 ± 0.3	35.1 ± 0.0	37.4
SeCo-Eco (ours)	85.3 ± 0.0	84.0 ± 0.0	22.3	<u>13.0</u>	35.7 ± 0.4	<u>32.4</u> ± 0.2	42.7 ± 0.0	40.6

Table 4. Linear probing and K-Nearest Neighbor comparison of state of the art models with our SeCo-Eco pretrained on our SSL4Eco on multi-label classification tasks. CLEF and TSAI have official train and test splits, the standard deviation is only reported when relevant. **Best**, second best.

port the mean and standard deviation for each metric across the 10 test folds. The reported metric is picked based on common choices in the literature. See Section A-6 for evaluations on a larger range of metrics and per-class results if applicable.

Classification Tasks.

Biomes. We adapt the biomes task of Klemmer *et al.* [49], assembling a dataset of 52k randomly selected inland locations and label them from a set of 15 classes according to Olson *et al.*’s biome map [65]. We adjust for latitude-longitude bias in the location selection. For each datapoint we download a 256×256 pixel (2.56 km) image from the least-clouded Sentinel-2 Harmonized dataset tile [70]. We choose images within a one month range from 15th of July/15th of January for the Northern/Southern hemisphere accordingly. We train using the cross entropy loss and report the macro F1 score.

Arctic Vegetation Types (CAVM). We create an Arctic vegetation types task, as the Arctic ecosystem tends to be critically undersampled (see Tab. 1). We assemble a dataset using 79k randomly selected locations in equal area projection in the Arctic and label them according to the Arctic vegetation types CAVM dataset [73]. We use the broad map units (B, G, P, S, and W) as labels, resulting in 5 vegetation categories. We choose images within a one month range from 15th of July. The downloaded satellite imagery, training, and metrics follow the setup of the biomes task.

Multi-Label Classification Tasks.

BigEarthNet. BigEarthNet [82] dataset is a 19-class, multi-label land cover classification dataset. It includes 590k 1.2×1.2 km Sentinel-2 patches collected in 2017-2018 across Europe. Although BigEarthNet is not specifically targeted for ecology, it is widely used for benchmarking GFMs, and we use it as a sanity check for the generaliza-

tion power of our embeddings. Following previous work, we report results on a predefined test set and use only 10% of the remaining images for training [58, 64, 89]. We adapt the SSL4EO protocol [89] for data preparation, train using a multi-label soft margin loss and measure performance by micro mean average precision.

GeoLifeCLEF 2023. The GeoLifeCLEF 2023 [7] dataset contains 5138 presence-absence surveys of 2174 plant species across France and the United Kingdom. Each survey reports all plant species found in a small plot (between 10m^2 and 400m^2). For each location, we download a 1×1 km Sentinel-2 patch (100×100 pixels). We train with the binary cross-entropy loss, up weighting all presences by a factor of 12 due to high imbalance between presences and absences. We use the entire labeled dataset for training and communicate results on the official held-out test set. We submit the predictions on the 22k test surveys to the leaderboard and report the micro F1 score.

EU-Forest. We adapt the European 1 km-resolution tree occurrence dataset EU-Forest [60] to a multi-label classification task. We sample 51 802 locations from the original data, covering 64 species with at least 200 occurrences, with some locations containing multiple species. For each location, we download a 1×1 km Sentinel-2 patch. We train using a multi-label soft margin loss and measure performance by micro F1 score.

TreeSatAI. The TreeSatAI [1] is a multimodal dataset for tree species identification with multi-label annotations for 15 tree genera classes taken in Lower Saxony, Germany. The dataset comprises 50 381 tiles of 60 m width for several remote sensing products. In our setting, we only use the Sentinel-2 6×6 patches. Similar to EU-Forest, we train with a multilabel soft margin loss and report the micro F1 score. We communicate performance on the official test,

and randomly select training and validation splits from the remaining data.

Regression Tasks.

BioMassters. BioMassters [62] is a benchmark for above-ground biomass estimation in Finland from Sentinel-1/2 time series. Initially designed for a dense pixel regression task, we reformulate it here as an image-level distribution prediction. To this end, we divide the total distribution of biomass throughout the dataset into decile bins. Since the first three bins account for zero biomass (*i.e.* ground pixels), we merge them. Then, for each 256×256 Sentinel-2 patch in the dataset, we compute the proportion of pixels falling into each of our 8 bins. Our model is tasked to predict the exact distribution of biomass for each image. Since the BioMassters dataset provides monthly images throughout the year, we split the task into a "summer" (June, July and August) and a "winter" (December, January and February) version, based on the season of the Sentinel-2 patches used as input. We train using the Kullback-Leibler divergence and report the average coefficient of determination R^2 across bins as our main metric.

CHELSA Climate Regression. Similar to SatClip [49] we propose to regress these aggregated climatic variables from pretrained geolocated embeddings. CHELSA [47] is a 1 km resolution global downscaled climate dataset, from which we extract the mean temperature (temp), total annual precipitation (prec), potential evaporation (evap) and site water balance (swb) from the 1981-2010 climatology of CHELSA v2.1 [47] for 50k locations across the landmass. For simplicity, we use the same locations and Sentinel-2 images as for the Biomes task. After Gaussian-normalizing the values, we train using a mean squared error loss and use R^2 to measure performance.

4.2. Results and Analysis

We compare the representation learned by SeCo-Eco on our SSL4Eco across the above-defined tasks. Figure 3 summarizes the performance of SeCo-Eco in comparison to the strongest baseline on each task. Overall, we observe that SeCo-Eco outperforms all other approaches on all but one task, showing that a simple change in the sampling design of the pretraining dataset can yield significant improvements.

Classification. Table 3 SeCo-Eco outperforms all other methods on our classification tasks, both for linear probing and k-NN evaluation, followed by SSL4EO with +2.8 and +1.9 macro F1 LP performance gaps on the biomes and CAVM tasks, respectively. The improvement of SeCo-Eco over SSL4EO can be explained by their difference in seasonal-contrastive training, as well as our dataset design (see Section 4.3 and Section A-4 for more details). The low performance of the RGB-based SeCo on biomes highlights the importance of multispectral images for the biomes classification task. The superior performance of SeCo-Eco over

SSL4EO on CAVM illustrates the importance of including arctic regions in the pretraining set for ecological applications.

Multi-Label Classification. We compare performance on four multi-label classification tasks in Table 4, three of which are specifically directed at predicting plant species communities. SeCo-Eco outperforms all other baselines in LP for BigEarthNet-10% (+2.1 mAP), GeoLifeCLEF (+0.6 micro F1). Interestingly, the largest performance gain from our approach is observed on the challenging BigEarthNet benchmark, which oversamples non-natural landscapes. This indicates that despite its focus on capturing global phenological seasonality, the spatiotemporal distribution of SSL4Eco still allows learning anthropic patterns. On the other hand, SeCo-Eco performs -4.1 micro F1 below SatMAE on the TreeSatAI task, which we attribute to the small 6×6 patch size used for this task, which is far from the 224×224 both SeCo-Eco and SeCo are pretrained on.

Model	BioMassters (mean R^2) [↑]		CHELSA (mean R^2) [↑]	
	LP	1-NN	LP	10-NN
SeCo [58]	51.2 ± 0.0	-19.2	68.3 ± 0.7	67.4 ± 0.7
SatMAE [16]	59.4 ± 0.5	-18.0	76.3 ± 0.6	77.6 ± 0.7
Satlas [5]	62.4 ± 0.9	-17.8	68.3 ± 0.9	73.3 ± 0.7
Croma [31]	58.4 ± 0.2	-18.1	73.3 ± 0.9	71.2 ± 0.5
SSL4EO [89]	71.3 ± 0.1	-16.8	75.8 ± 0.6	77.7 ± 0.5
DOFA [93]	63.0 ± 0.4	-18.3	69.6 ± 0.6	70.7 ± 0.7
SeCo-Eco (ours)	75.3 ± 0.3	-16.3	81.1 ± 0.4	81.0 ± 0.5

Table 5. Linear probing and K-Nearest Neighbor comparison of state of the art models with our SeCo-Eco pretrained on our SSL4Eco on regression tasks. For the BioMassters task the standard deviation can only be reported for linear probing due to the fixed train and test sets. **Best**, second best.

Regression. For the two regression tasks of BioMassters and CHELSA, we report in Table 5 the mean R^2 performance, aggregated across the BioMassters bins and CHELSA rasters. SeCo-Eco outperforms all other baselines by a significant margin on both BioMassters (+4.0 R^2 LP) and CHELSA (+4.8 R^2 LP). The large performance gap with respect to SSL4EO suggests that our model benefits from the more uniform spatial sampling of its pretraining dataset. Indeed, the BioMassters dataset is located in Finland, which is poorly covered by the SSL4EO pretraining dataset (Fig. 1). Similarly, the CHELSA task requires uniform performance across the globe, which does not align with the urban-focused SSL4EO pretraining. The negative R^2 scores on BioMassters indicate that 1-NN yields lower performance than a simple average prediction, suggesting that this NN evaluation is not adapted to this

Model	BE10% (micro mAP) \uparrow	CLEF (micro F1) \uparrow	EU-Forest (micro F1) \uparrow	TSAI (micro F1) \uparrow	Biomes (macro F1) \uparrow	CAVM (macro F1) \uparrow	BioMassters (mean R ²) \uparrow	CHELSEA (mean R ²) \uparrow
SSL4EO [89]	83.2 \pm 0.1	<u>21.7</u>	32.6 \pm 0.1	<u>42.3</u> \pm 0.0	53.3 \pm 1.1	57.5 \pm 0.6	71.4 \pm 0.0	75.9 \pm 0.6
MoCo	<u>84.0</u> \pm 0.1	<u>21.7</u>	<u>35.4</u> \pm 0.2	41.3 \pm 0.0	58.4 \pm 0.8	<u>59.1</u> \pm 0.7	<u>73.4</u> \pm 0.1	81.5 \pm 0.4
SeCo-Eco (ours)	85.3 \pm 0.0	22.3	35.7 \pm 0.4	42.7 \pm 0.0	<u>56.1</u> \pm 0.7	59.4 \pm 1.0	75.2 \pm 0.1	<u>81.1</u> \pm 0.4

Table 6. Linear probing comparison of MoCo-Eco and SeCo-Eco pretrained on SSL4Eco. SeCo-Eco learns both season-invariant and season-sensitive representations, which yield overall better performance than the season-invariant MoCo-Eco. **Best**, second best.

Model	BioMassters S (mean R ²) \uparrow		BioMassters W (mean R ²) \uparrow	
	LP	1-NN	LP	1-NN
SeCo [58]	51.3 \pm 0.0	-19.2	32.3 \pm 0.1	-30.6
SatMAE [16]	59.5 \pm 0.6	-18.0	50.0 \pm 0.2	-26.3
Satlas [5]	62.5 \pm 0.9	-17.8	51.7 \pm 1.1	-26.1
Croma [31]	58.5 \pm 0.2	-18.1	43.5 \pm 0.3	-27.0
SSL4EO [89]	<u>71.4</u> \pm 0.0	<u>-16.8</u>	<u>63.2</u> \pm 0.1	<u>-25.3</u>
DOFA [93]	63.1 \pm 0.4	-18.3	55.0 \pm 0.4	-26.2
SeCo-Eco (ours)	75.2 \pm 0.1	-16.3	67.7 \pm 0.2	-24.9

Table 7. Comparison of models using Summer (S) or Winter (W) images on BioMassters. Due to the fixed splits, the standard deviation can only be reported for linear probing. **Best**, second best.

task, for which linear probing should be preferred. In comparison, the CHELSA task regresses climatic conditions, which evolve more smoothly throughout the models feature spaces, allowing to retrieve good estimates from neighboring embeddings.

4.3. Ablation Study.

Seasonal Pretraining. We compare in Table 6 the impact of pretraining on SSL4Eco using the seasonal-contrastive objectives from SeCo [58] and SSL4EO [89] (SeCo-Eco and MoCo-Eco models, respectively). Our results show that SeCo-Eco features overall tend to perform on par or better than MoCo-Eco features with linear probing, showing the benefit of learning not only season-agnostic features, but also season-specific ones.

Winter Predictions. To test the influence of the acquisition date on model performance in downstream task, we compare the models on images taken from local winter months against summer months of BioMassters dataset. As shown in Table 7, all models drop in performance when using the snow-covered winter images of BioMassters. Still, we observe that SeCo-Eco clearly outperforms other mod-

els on both seasons, followed by SSL4EO, owing to their respective seasonal-contrastive pretraining. Interestingly, the RGB-based SeCO model performs worst despite having the same pretraining strategy as SeCo-Eco, suggesting multi-spectral imagery as critical to such tasks. These results demonstrate the robustness of our learned phenologically-informed representation to seasonal changes.

Limitations and Future Works. To recover less clouded images in each phenological season, we gather images across 2017-2024, which may cause large temporal gaps between images of the same location, making our dataset inadequate for fine-grained temporal tasks. Although not the focus of this work, pretraining more methods on SSL4Eco besides SeCo-Eco and MoCo-Eco would provide deeper insights into the respective merits of each. Extending our dataset with additional modalities would likely allow learning richer features [2, 3, 64], which we make possible by releasing all necessary metadata. Finally, our dataset and model could naturally be used in a multi-modal contrastive learning framework aligning Sentinel-2 seasonal representations with text [81, 95], environmental variables [17, 86], or geolocation [49, 87].

5. Conclusion

In this study, we propose a simple approach for sampling global seasonality-aware remote sensing datasets, from which we derive SSL4Eco, a multi-temporal Sentinel-2 dataset for pretraining geospatial foundation models targeted for macroecological applications. Compared to previous works, our dataset uniformly samples the landmass and local phenological cycles. We demonstrate that our simple spatiotemporal dataset sampling consistently improves the quality of self-supervised representations on a variety of macroecological tasks, highlighting the importance of pretraining set design, which could naturally be extended to additional relevant modalities.

Acknowledgements. This work made use of infrastructure services provided by the Science IT team of the University of Zurich (www.s3it.uzh.ch). We thank Benjamin Deneu for his helpful suggestions.

References

- [1] Steve Ahlswede, Christian Schulz, Christiano Gava, Patrick Helber, Benjamin Bischke, Michael Förster, Florencia Arias, Jörn Hees, Begüm Demir, and Birgit Kleinschmit. Treesatai benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing. *Earth System Science Data Discussions*, 2022:1–22, 2022. 6, 4
- [2] Guillaume Astruc, Nicolas Gonthier, Clement Mallet, and Loic Landrieu. Anysat: An earth observation model for any resolutions, scales, and modalities. *CVPR*, 2024. 3, 8
- [3] Guillaume Astruc, Nicolas Gonthier, Clement Mallet, and Loic Landrieu. OmniSat: Self-supervised modality fusion for Earth observation. *ECCV*, 2024. 3, 8
- [4] Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. *ICCV*, 2021. 2, 3
- [5] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. Satlaspretrain: A large-scale dataset for remote sensing image understanding. *ICCV*, 2023. 3, 5, 6, 7, 8, 4
- [6] Mariana Belgiu and Lucian Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016. 2
- [7] Christophe Botella, Benjamin Deneu, Diego Marcos, Maximilien Servajean, Joaquim Estopinan, Théo Larcher, César Leblanc, Pierre Bonnet, and Alexis Joly. The geolifelef 2023 dataset to evaluate plant species distribution models at high spatial resolution across europe. *arXiv preprint arXiv:2308.05121*, 2023. 6
- [8] E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo. Ipbes (2019): Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. *IPBES secretariat*, 2019. 2, 3
- [9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. *ICCV*, 2021. 5
- [10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *ICML*, 2020. 3
- [11] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 3, 2
- [12] Jun Cheng, Haibin Wu, Zhengyu Liu, Peng Gu, Jingjing Wang, Cheng Zhao, Qin Li, Haishan Chen, Huayu Lu, Haibo Hu, et al. Vegetation feedback causes delayed ecosystem response to east asian summer monsoon rainfall during the holocene. *Nature Communications*, 2021. 4
- [13] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. *CPR*, 2018. 3
- [14] Isabelle Chuine, Pascal Yiou, Nicolas Viovy, Bernard Seguin, Valérie Daux, and Emmanuel Le Roy Ladurie. Grape ripening as a past climate indicator. *Nature*, 2004. 4
- [15] Elijah Cole, Grant Van Horn, Christian Lange, Alexander Shepard, Patrick Leary, Pietro Perona, Scott Loarie, and Oisín Mac Aodha. Spatial implicit neural representations for global-scale species mapping. *ICML*, 2023. 3
- [16] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *NeurIPS*, 2022. 3, 5, 6, 7, 8, 4
- [17] Rangel Daroya, Elijah Cole, Oisín Mac Aodha, Grant Van Horn, and Subhansu Maji. Wildsat: Learning satellite image representations from wildlife observations. *arXiv preprint arXiv:2412.14428*, 2024. 8
- [18] Jesús Delegido, Jochem Verrelst, Luis Alonso, and Jose Moreno. Evaluation of sentinel-2 red-edge bands for empirical estimation of green lai and chlorophyll content. *Sensors*, 2011. 1
- [19] Jacob Devlin. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 2
- [20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3
- [21] Ralph Dubayah, James Bryan Blair, Scott Goetz, Lola Fatoyinbo, Matthew Hansen, Sean Healey, Michelle Hofton, George Hurtt, James Kellner, Scott Luthcke, et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of remote sensing*, 2020. 5
- [22] Nathalie Dubois, Delia W Oppo, Valier V Galy, Mahyar Mohtadi, Sander Van Der Kaars, Jessica E Tierney, Yair Rosenthal, Timothy I Eglinton, Andreas Lückge, and Braddock K Linsley. Indonesian vegetation response to changes in rainfall seasonality over the past 25,000 years. *Nature Geoscience*, 2014. 4
- [23] Iris Dumeur, Silvia Valero, and Jordi Inglada. Self-supervised spatio-temporal representation learning of satellite image time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 3
- [24] James L Edwards. Research and societal benefits of the global biodiversity information facility. *BioScience*, 2004. 5
- [25] Hongliang Fang, Baret Frederic, Stephen Plummer, and Gabriela Schaepman-Strub. An overview of global leaf area index (lai): Methods, products, validation, and applications. *Reviews of Geophysics*, 2019. 2
- [26] Casper Fibaek, Luke Camilleri, Andreas Luyts, Nikolaos Dionelis, and Bertrand le Saux. PhilEO Bench: Evaluating geo-spatial foundation models. *IGARSS*, 2024. 5
- [27] Jeremy Fisher, John Mustard, and Matthew Vadeboncoeur. Green leaf phenology at landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment*, 2006. 2
- [28] World Economic Forum. Global risks report 2025, 2025. 2

- [29] Alistair Francis and Mikolaj Czerkawski. Major tom: Expandable datasets for earth observation. *arXiv preprint 2402.12095*, 2024. 3, 4, 1
- [30] Sulla-Menashe D. Friedl M., Gray J. Modis/terra+aqua land cover dynamics yearly 13 global 500m sin grid v061, 2022. 4, 1, 2
- [31] Anthony Fuller, Koreen Millard, and James Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. *NeurIPS*, 2023. 3, 5, 6, 7, 8, 4
- [32] Anatol Garioud, Nicolas Gonthier, Loic Landrieu, Apolline De Wit, Marion Valette, Marc Poupée, Sébastien Giordano, et al. Flair: a country-scale land cover semantic segmentation dataset from multi-source optical imagery. *NeurIPS*, 2023. 3
- [33] Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *ICCV*, 2021. 3
- [34] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R Sveinsson. Random forests for land cover classification. *Pattern Recognition Letters*, 2006. 2
- [35] Jie Gui, Tuo Chen, Jing Zhang, Qiong Cao, Zhenan Sun, Hao Luo, and Dacheng Tao. A survey on self-supervised learning: Algorithms, applications, and future trends. *TPAMI*, 2024. 2
- [36] Antoine Guisan, Reid Tingley, John B. Baumgartner, Ilona Naujokaitis-Lewis, Patricia R. Sutcliffe, Ayesha I. T. Tulloch, Tracey J. Regan, Lluís Brotons, Eve McDonald-Madden, Chrystal Mantyka-Pringle, Tara G. Martin, Jonathan R. Rhodes, Ramona Maggini, Samantha A. Setterfield, Jane Elith, Mark W. Schwartz, Brendan A. Wintle, Olivier Broennimann, Mike Austin, Simon Ferrier, Michael R. Kearney, Hugh P. Possingham, and Yvonne M. Buckley. Predicting species distributions for conservation decisions. *Ecology Letters*, 2013. 2
- [37] 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*, 2013. 2
- [38] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CVPR*, 2016. 3
- [39] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *CVPR*, 2020. 3
- [40] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. *CVPR*, 2022. 3
- [41] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The ERA5 global reanalysis. *Quarterly journal of the royal meteorological society*, 2020. 5
- [42] Johannes Jakubik, Sujit Roy, CE Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, et al. Foundation models for generalist geospatial artificial intelligence. *arXiv preprint arXiv:2310.18660*, 2023. 3
- [43] Forrest Jessica and Miller-Rushing Abraham J. Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society*, 2010. 3
- [44] Mark A Friedl Josh Gray, Damien Sulla-Menashe. User guide to collection 6 modis land cover dynamics (mcd12q2) product. *NASA*, 2019. 4, 2
- [45] Christopher O Justice, Eric Vermote, John RG Townshend, Ruth Defries, David P Roy, Dorothy K Hall, Vincent V Salomonson, Jeffrey L Privette, George Riggs, Alan Strahler, et al. The moderate resolution imaging spectroradiometer (modis): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 1998. 4, 5, 1, 2
- [46] Kaan Karaman, Yuchang Jiang, Damien Robert, Vivien Sainte Fare Garnot, Maria J. Santos, and Jan Dirk Wegner. Gsr4b: Biomass map super-resolution with sentinel-1/2 guidance. *ISPRS Annals of Photogrammetry and Remote Sensing*, 2025. 2, 4, 1
- [47] Böhner J. Karger D., Conrad O. Climatologies at high resolution for the earths land surface areas. *Sci Data*, 2017. 5, 7, 6
- [48] Franziska Keller and Christian Körner. The role of photoperiodism in alpine plant development. *Arctic, Antarctic, and Alpine Research*, 2003. 4
- [49] Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023. 3, 5, 6, 7, 8
- [50] Christian Körner and Christian Kørner. *Alpine plant life: functional plant ecology of high mountain ecosystems*. Springer, 1999. 4
- [51] Alexandre Lacoste, Nils Lehmann, Pau Rodriguez, Evan David Sherwin, Hannah Kerner, Björn Lütjens, Jeremy Andrew Irvin, David Dao, Hamed Alemohammad, Alexandre Drouin, Mehmet Gunturkun, Gabriel Huang, David Vazquez, Dava Newman, Yoshua Bengio, Stefano Ermon, and Xiao Xiang Zhu. GEO-Bench: Toward foundation models for earth monitoring. *NeurIPS*, 2024. 5
- [52] Nico Lang, Walter Jetz, Konrad Schindler, and Jan Dirk Wegner. A high-resolution canopy height model of the earth. *Nature Ecology & Evolution*, 2023. 2, 4, 1
- [53] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *CVPR*, 2021. 3
- [54] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5
- [55] Buchhorn M., Smets B., Bertels L., Lesiv M., Tsendbazar N E., Masiliunas D., Linlin L., Herold M., and Fritz S. Copernicus global land service: Land cover 100m: Collection 3: epoch 2019: Globe (version v3.0.1). *Zenodo*, 2020. 1
- [56] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019. 2

- [57] Snethlage M.A., Geschke J., Spehn E.M., Ranipeta A., Yoccoz N.G., Körner Ch., Jetz W., Fischer M., and Urbach D. A hierarchical inventory of the world's mountains for global comparative mountain science. *gmba mountain inventory v2. Sci Data*, 2022. 2
- [58] Oscar Mañas, Alexandre Lacoste, Xavier Giro-i Nieto, David Vazquez, and Pau Rodriguez. Seasonal contrast: Un-supervised pre-training from uncurated remote sensing data. *ICCV*, 2021. 2, 3, 4, 5, 6, 7, 8, 1
- [59] Valerio Marsocci, Yuru Jia, Georges Le Bellier, David Kerekes, Liang Zeng, Sebastian Hafner, Sebastian Gerard, Eric Brune, Ritu Yadav, Ali Shibli, Heng Fang, Yifang Ban, Maarten Vergauwen, Nicolas Audebert, and Andrea Nascetti. PANGAEA: A global and inclusive benchmark for geospatial foundation models. *arXiv preprint arXiv:2412.04204*, 2024. 5
- [60] A. Mauri, G. Strona, and J. San-Miguel-Ayanz. EU-Forest, a high-resolution tree occurrence dataset for Europe. *Sci Data*, 2017. 6, 4
- [61] Andrea Nascetti, Ritu Yadav, Kirill Brodt, Qixun Qu, Hongwei Fan, Yuri Shendryk, Isha Shah, and Christine Chung. Biomasters: A benchmark dataset for forest biomass estimation using multi-modal satellite time-series. *NeurIPS*, 2023. 5
- [62] Andrea Nascetti, Ritu Yadav, Kirill Brodt, Qixun Qu, Hongwei Fan, Yuri Shendryk, Isha Shah, and Christine Chung. Biomasters: A benchmark dataset for forest biomass estimation using multi-modal satellite time-series. *Advances in Neural Information Processing Systems*, 36:20409–20420, 2023. 7
- [63] United Nations. The un sustainable development goals, 2015. 2
- [64] Vishal Nedungadi, Ankit Karirya, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. Mmearth: Exploring multi-modal pretext tasks for geospatial representation learning. *ECCV*, 2024. 3, 6, 8
- [65] David M. Olson, Eric Dinerstein, Eric D. Wikramanayake, Neil D. Burgess, George V. N. Powell, Emma C. Underwood, Jennifer A. D'Amico, Illanga Itoua, Holly E. Strand, John C. Morrison, Colby J. Loucks, Thomas F. Allnutt, Taylor H. Ricketts, Yumiko Kura, John F. Lamoreux, Wesley W. Wetzel, Prashant Hedao, and Kenneth R. Kassem. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 2001. 6, 5
- [66] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2
- [67] Jouni Partanen, Veikko Koski, and Heikki Hänninen. Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (*Picea abies*). *Tree physiology*, 1998. 4
- [68] Luis Miguel Pazos-Outón, Cristina Nader Vasconcelos, Anton Raichuk, Anurag Arnab, Dan Morris, and Maxim Neumann. Planted: a dataset for planted forest identification from multi-satellite time series. *IGARSS*, 2024. 3
- [69] Nathalie Pettorelli, Jon Olav Vik, Atle Mysterud, Jean-Michel Gaillard, Compton J. Tucker, and Nils Chr. Stenseth. Using the satellite-derived ndvi to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 2005. 4, 2
- [70] Copernicus Sentinel-2 (processed by ESA). Msi level-2a boA reflectance product. collection 1. *European Space Agency*, 2021. 4, 6, 1
- [71] Mirali Purohit, Gedeon Muhawenayo, Esther Rolf, and Hannah Kerner. How does the spatial distribution of pre-training data affect geospatial foundation models? *arXiv preprint arXiv:2501.12535*, 2025. 2
- [72] 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. *ICML*, 2021. 2
- [73] Martha K. Raynolds, Donald A. Walker, Andrew Balser, Christian Bay, Mitch Campbell, Mikhail M. Cherosov, Fred J.A. Daniëls, Pernille Bronken Eidesen, Ksenia A. Ermokhina, Gerald V. Frost, Birgit Jedrzejek, M. Torre Jorgenson, Blair E. Kennedy, Sergei S. Kholod, Igor A. Lavrinenko, Olga V. Lavrinenko, Borgþór Magnússon, Nadezhda V. Matveyeva, Sigmar Metúsalemsson, Lennart Nilsen, Ian Olthof, Igor N. Pospelov, Elena B. Pospelova, Darren Pouliot, Vladimir Razzhivin, Gabriela Schaepman-Strub, Jozef ibfk, Mikhail Yu. Telyatnikov, and Elena Troeva. A raster version of the Circumpolar Arctic Vegetation Map (CAVM). *Remote Sensing of Environment*, 2019. 5, 6
- [74] Ribana Roscher, Marc Russwurm, Caroline Gevaert, Michael Kampffmeyer, Jefersson A. Dos Santos, Maria Vakalopoulou, Ronny Hänsch, Stine Hansen, Keiller Nogueira, Jonathan Prexl, and Devis Tuia. Better, not just more: Data-centric machine learning for earth observation. *IEEE Geoscience and Remote Sensing Magazine*, 2024. 2
- [75] Ake Rosenqvist, Masanobu Shimada, Norimasa Ito, and Manabu Watanabe. Alos palsar: A pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing*, 2007. 5, 1
- [76] Vincent V Salomonson, WL Barnes, Peter W Maymon, Harry E Montgomery, and Harvey Ostrow. Modis: Advanced facility instrument for studies of the earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, 1989. 1
- [77] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu. SEN12MS—a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *arXiv preprint arXiv:1906.07789*, 2019. 3, 4
- [78] Annemarie Schneider, Mark A Friedl, and David Potere. Mapping global urban areas using modis 500-m data: New methods and datasets based on urban ecoregions. *Remote sensing of environment*, 2010. 3
- [79] Schädel C. et al. Schuur E., McGuire A. Climate change and the permafrost carbon feedback. *Nature*, 2015. 3
- [80] GhJulia Sialelli, Torben Peters, Jan D Wegner, and Konrad Schindler. Agbd: A global-scale biomass dataset. *ISPRS*

- Annals of Photogrammetry and Remote Sensing*, 2025. [2](#), [4](#), [1](#)
- [81] João Daniel Silva, João Magalhães, Devis Tuia, and Bruno Martins. Large language models for captioning and retrieving remote sensing images. *arXiv preprint arXiv:2402.06475*, 2024. [8](#)
 - [82] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. *IGARSS*, 2019. [2](#), [3](#), [6](#), [4](#)
 - [83] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, Björn Rommen, Nicolas Floury, Mike Brown, Ignacio Navas Traver, Patrick Deghaye, Berthyl Duesmann, Betlem Rosich, Nuno Miranda, Claudio Bruno, Michelangelo L’Abbate, Renato Croci, Andrea Pietropaolo, Markus Huchler, and Friedhelm Rostan. Gmes sentinel-1 mission. *Remote Sensing of Environment*, 2012. [5](#), [1](#)
 - [84] Gabriel Tseng, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023. [3](#)
 - [85] Woody Turner, Sacha Spector, Ned Gardiner, Matthew Fladeland, Eleanor Sterling, and Marc Steininger. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 2003. [2](#)
 - [86] Huy Ung, Ryoichi Kojima, and Shinya Wada. Leverage samples with single positive labels to train cnn-based models for multi-label plant species prediction. *CLEF Working Notes*, 2023. [8](#)
 - [87] Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. *NeurIPS*, 2023. [3](#), [8](#)
 - [88] 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*, 2022. [2](#), [3](#), [5](#)
 - [89] Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M Albrecht, and Xiao Xiang Zhu. Ssl4eo-s12: A large-scale multi-modal, multi-temporal dataset for self-supervised learning in earth observation. *IEEE Geoscience and Remote Sensing Magazine*, 2023. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#)
 - [90] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. *CVPR*, 2018. [5](#)
 - [91] Michael A. Wulder, David P. Roy, Volker C. Radeloff, Thomas R. Loveland, Martha C. Anderson, David M. Johnson, Sean Healey, Zhe Zhu, Theodore A. Scambos, Nima Pahlevan, Matthew Hansen, Noel Gorelick, Christopher J. Crawford, Jeffrey G. Masek, Txomin Hermosilla, Joanne C. White, Alan S. Belward, Crystal Schaaf, Curtis E. Woodcock, Justin L. Huntington, Leo Lymburner, Patrick Hostert, Feng Gao, Alexei Lyapustin, Jean-Francois Pekel, Peter Strobl, and Bruce D. Cook. Fifty years of landsat science and impacts. *Remote Sensing of Environment*, 2022. [5](#), [1](#)
 - [92] Yichun Xie, Zongyao Sha, and Mei Yu. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 2008. [2](#)
 - [93] Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired foundation model for observing the Earth crossing modalities. *arXiv preprint arXiv:2403.15356*, 2024. [3](#), [5](#), [6](#), [7](#), [8](#), [4](#)
 - [94] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B. Lobell, and Stefano Ermon. SustainBench: Benchmarks for monitoring the sustainable development goals with machine learning. *NeurIPS*, 2021. [5](#)
 - [95] Zhenghang Yuan, Zhitong Xiong, Lichao Mou, and Xiao Xiang Zhu. Chatearthnet: A global-scale image-text dataset empowering vision-language geo-foundation models. *Earth System Science Data Discussions*, 2024:1–24, 2024. [8](#)
 - [96] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. *CVPR*, 2022. [2](#)
 - [97] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 2017. [2](#)

SSL4Eco: A Global Seasonal Dataset for Geospatial Foundation Models in Ecology

Supplementary Material

A-1. SSL4Eco Dataset Construction

In this section, we provide more details on our dataset construction protocol.

Spatial Sampling. We use the same approach as MajorTOM [29] for sampling locations uniformly across the land-mass. Our locations correspond to the center of the grid cells.

Seasonal Sampling. As explained in Section 3.1 and Figure 2, we define 4 seasons as intervals between Greenup, Maturity, Senescence, Dormancy, and next Greenup variables. The definition of these EVI variables can be found in Section A-3. For each variable, we calculate the median day in the available years. The EVI product from the MCD12Q2 v6.1 [30] product has missing values in non-vegetated and some evergreen areas (*e.g.* tropics), for which we expect low seasonal variation. We populate these with a nearest-neighbor approach by searching across geographical space.

For each location and season, we preselect all Sentinel-2 tiles across the 6 years of data available 2017-2024. The broad range of years was chosen to account for high cloud coverage in some areas (*e.g.* tropics in wet seasons). Following previous work [89], we remove the tiles with less than 20% cloud coverage. Finally, we choose the date and tile with the lowest cloud coverage for the location-season at hand. If fewer than four seasonal images are available for a location due to cloud filtering, we use the 2 or 3 images that are available with less than 20% cloud coverage. Locations with only one image are excluded, accounting for 3% of initially sampled locations, mostly in the tropics and Antarctica. Hence, final patches may be clouded, but the construction process ensures that the overall dataset has less than 20% cloud coverage.

We stress that the scope of this work is to study impact of spatiotemporal sampling compared to existing widely-used 4-date seasonal datasets such as SeCo [58] and SSL4EO [89]. As such, we follow the standard preprocessing procedure of these datasets regarding cloud filtering and the number of seasonal dates per year fair comparison across the computer vision literature. However, realistic Earth Observation applications would require methods capable of handling arbitrarily sampled, potentially clouded, time series of satellite observations. We leave this exploration of the required dataset and models for further work.

Data Source. Several open-access satellite products support vegetation monitoring.

- Landsat missions [91] offer a long-term multispectral record at 30 m resolution, with a 16-day revisit cycle (reduced to 8 days since 2013).
- MODIS [45, 76] provides more spectral bands and a 1–2 day revisit rate, though at a coarser 250–1000 m resolution.
- Since 2015, Sentinel-2 [70] has been delivering 10 m global imagery with a 5-day maximum revisit period, balancing high spatial and temporal resolution. The Sentinel-2 instrument captures spectral bands indicative of ecological patterns, such as red-edge wavelengths sensitive to vegetation stress and chlorophyll content [18].
- Radar sensors may provide diverse ecological insights depending on their frequency: C-band such as Sentinel-1 [83] detects foliage, topography, and moisture, while L-band such as ALOS PALSAR [75] can characterize wood structure.

In this paper, we chose Sentinel-2 due to its widespread use for large-scale vegetation monitoring [46, 52, 80], but we believe our conclusions remain applicable and may be extended to other satellite products in future works. We leave the exploration of our proposed spatiotemporal sampling for multimodal representation learning for future work.

Downloading. The SSL4Eco dataset is downloaded from Google Earth Engine using code from SeCo [58] and SSL4EO-S12 [89] with altered data source, seasonality, and data distribution. We use the Sentinel-2A MSI collection which, compared to Sentinel-2C, has atmospheric correction and depicts more accurately features on the ground [70]. We use harmonized version of the product instead of the original one, as it corrects for normalization issues in 2022. We use Sentinel tiles with less than 20% cloud coverage.

A-2. Implementation Details

In this section, we provide more details on the implementation and training of our models.

Input Bands. Our models SeCo-Eco and MoCo-Eco are trained to take as input the 8 Sentinel-2 bands for ecological applications. Specifically, we use the B2, B3, B4, B5, B6, B7, B8, and B8A bands. While B2-B4 provide information on foliage color, which helps to assess seasonality

and plant health, B5-B7 capture red-edge wavelengths sensitive to vegetation stress and chlorophyll content, and B8 and B8A in near-infrared range are useful to distinguish non-vegetated areas. In addition, we also include the NDVI index as a remote sensing-based proxy of vegetation productivity and biomass [69]. As a result, our models expect 9 channels as input.

We leave the exploration of pretraining on our SSL4Eco sampling with more bands or modalities for future work.

Weighted Sampling. Despite the uniform global sampling of SSL4Eco, some locations may have more interesting geographical and seasonal dynamics than others. In order to drive the pretraining towards regions with richer ecological patterns, we use a weighted sampling in our pre-training dataloader. Specifically, we assign a $\div 4$ weight to non-vegetated areas, identified as mean NDVI < 0.1 in all seasons (17% of SSL4Eco), focusing less on deserts and ice packs. We oversample mountain regions with a $\times 2$ weight, identified with the GMBA Mountain Inventory [57] (16% of SSL4Eco), focusing more on ecologically diverse areas, as mountain regions harbor the highest diversity and heterogeneity of ecoregions.

Pretraining. We pretrain SeCo-Eco using the hyperparameters and code provided by Mañas *et al.* [58], using MoCo v2 [11], with minor changes: we replace the RGB input with multispectral images and set the length of the negative examples queue to 65 536, following the implementation of Wang *et al.* [89].

We pretrain MoCo-Eco using the hyperparameters and code provided by Wang *et al.* [89], adapted for a single A100 GPU with batch size of 256.

Finally, we modify the random seasonal sampling found in the implementations of SeCo [58] and SSL4EO [89]. When randomly selecting seasons at batch construction time, both use:

```
np.random.choice(..., replace=True),
although we believe:
np.random.choice(..., replace=False)
is the correct implementation of their respective methods,
as this avoids contrasting an image against itself.
```

A-3. EVI-based Seasonality

We use the Enhanced Vegetation Index (EVI) from the MCD12Q2 v6.1 [30] product of the MODIS [45] satellite mission to define our local, phenology-informed seasons. Similar to NDVI, the EVI index is commonly used to quantify the greenness of an area, but is more sensitive in areas with dense vegetation cover. Figure A-1 illustrate a typical EVI curve over the year, and Table A-1 details how the Greenup, Maturity, Senescence, and Dormancy seasonality

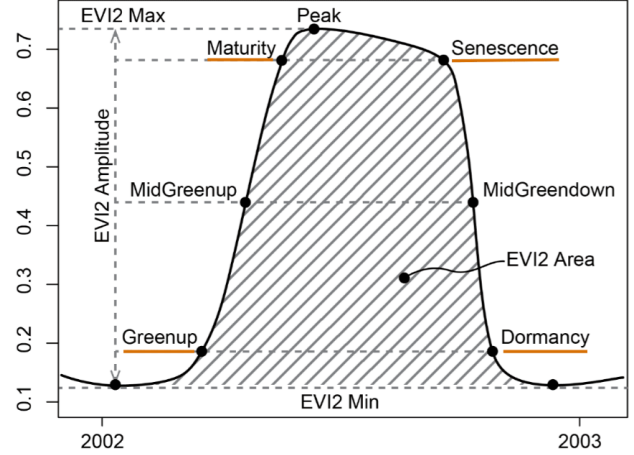


Figure A-1. Enhanced Vegetation Index (EVI) curve of the vegetation cycle at a given location. Based on this curve, the Greenup, Maturity, Senescence, and Dormancy seasonality variables are defined as detailed in Tab. A-1. Image taken from [44].

Name	Definition - Date when...
Greenup	EVI first crossed 15% of segment EVI amplitude
Maturity	EVI first crossed 90% of segment EVI amplitude
Senescence	EVI last crossed 90% of segment EVI amplitude
Dormancy	EVI last crossed 15% of segment EVI amplitude

Table A-1. Definition of the Greenup, Maturity, Senescence, and Dormancy seasonality variables based on the EVI curve (Fig. A-1).

variables are defined. For each location in our dataset, we choose 4 images, one for each season, close to the middle between the four EVI-derived variables. See the MCD12Q2 user guide [44] for more details on EVI variables.

A-4. Calendar Ablation

Our temporal sampling of SSL4Eco described in Section 3.1 makes the assumption that pretraining on EVI-based seasonal samplings rather than calendar seasons yields better features for ecological downstream tasks. To verify this claim, we assemble the SSL4Eco-Calendar dataset, which follows the same spatial sampling as SSL4Eco, but with a temporal sampling based on calendar dates following SSL4EO-S12 [89]. We derive SeCo-Calendar from this dataset, by using the same pretraining recipe and backbone as for our SeCo-Eco, and compare in Table A-2 their respective performance across downstream tasks. We observe that our proposed EVI-based seasonal sampling yields representations which overall perform better than calendar-based sampling on most downstream tasks. In particular, EU-Forest (+1.5 micro F1), TSAI (+1.9 macro F1), and Biomes (+0.9 macro F1) prove

Model	BE10% (micro mAP) \uparrow	CLEF (micro F1) \uparrow	EU-Forest (micro F1) \uparrow	TSAI (micro F1) \uparrow	Biomes (macro F1) \uparrow	CAVM (macro F1) \uparrow	BioMassters (mean R ²) \uparrow	Chelsa (mean R ²) \uparrow
SeCo-Calendar	85.3 \pm 0.0	22.4	34.2 \pm 0.1	40.8 \pm 0.0	55.2 \pm 1.0	58.7 \pm 0.8	75.7 \pm 0.0	80.6 \pm 0.5
SeCo-Eco (ours)	85.3 \pm 0.0	22.7	35.7 \pm 0.4	42.7 \pm 0.0	56.1 \pm 0.7	59.4 \pm 1.0	75.1 \pm 0.0	81.1 \pm 0.4

Table A-2. Linear probing comparison of SeCo-Eco and SeCo-Calendar pretrained on EVI-based and calendar-based seasonal samplings, respectively. EVI-based samplings overall yields better features for downstream macroecological tasks, with the exception of the BioMassters dataset. **Best**.

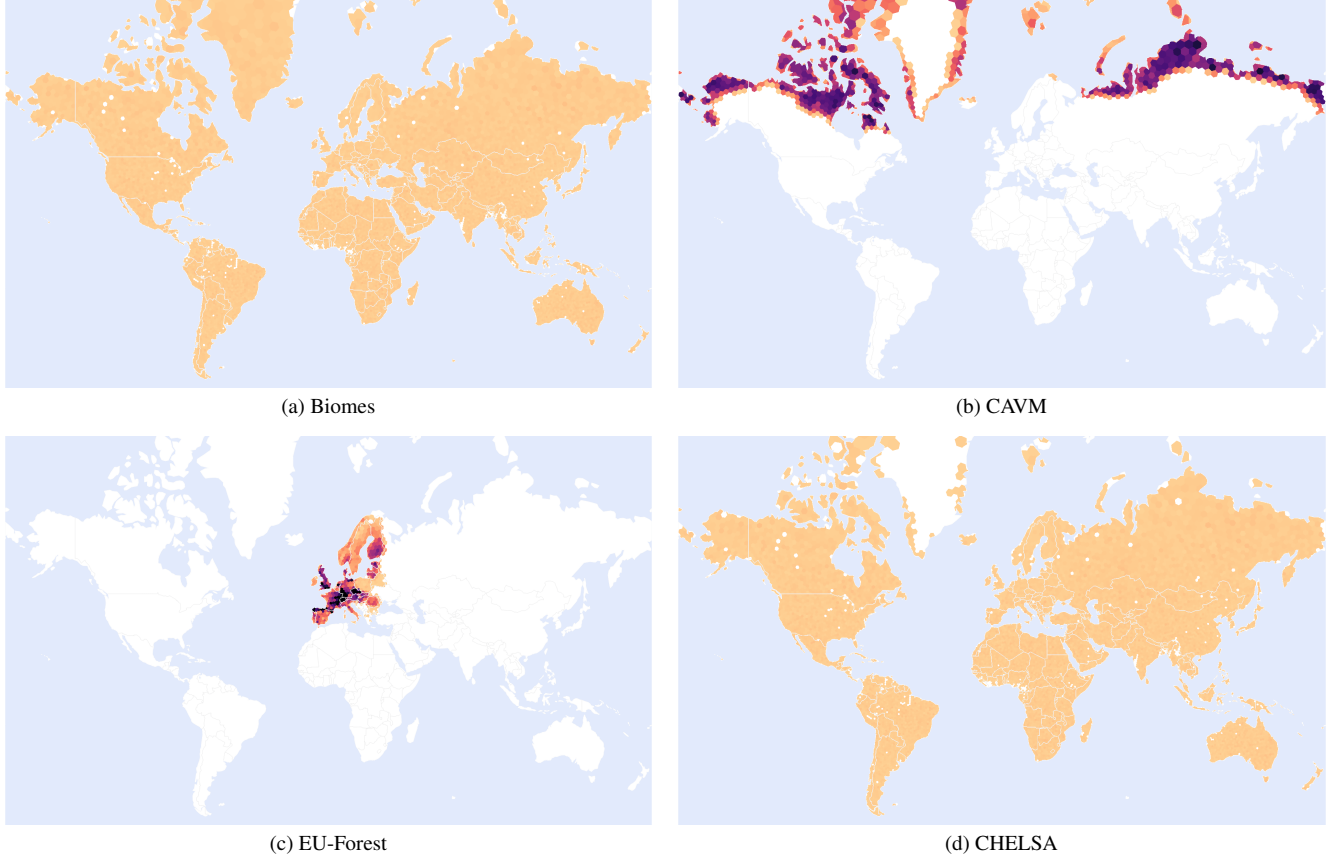


Figure A-2. Spatial distribution of the four new downstream tasks created for this work. We sample Biomes and CHELSA locations uniformly across the landmass. Meanwhile, the CAVM dataset is located in arctic regions and EU-Forest is limited to Europe.

to benefit from the finer phenology-informed features of SeCo-Eco. These results validate the importance of temporal sampling and the definition of local seasonality to capture local ecological patterns.

A-5. Downstream Tasks

We illustrate in Figure A-2 the spatial distribution of the samplings used for the new downstream tasks proposed in this paper: Biomes, CAVM, EU-Forest, and CHELSA

A-6. Detailed Results

Beyond evaluating performance with the most established metric per dataset, we provide further experimental results on an expanded set of metrics.

Model	BE10% [82]							
	Macro F1 \uparrow		Micro F1 \uparrow		Macro mAP \uparrow		Micro mAP \uparrow	
	LP	30-NN	LP	30-NN	LP	30-NN	LP	30-NN
SeCo [58]	56.3 \pm 0.3	36.0 \pm 0.1	68.9 \pm 0.2	44.7 \pm 0.1	64.5 \pm 0.2	62.4 \pm 0.2	79.2 \pm 0.0	77.8 \pm 0.1
SatMAE [16]	58.9 \pm 0.7	39.0 \pm 0.1	69.3 \pm 0.3	47.5 \pm 0.1	66.2 \pm 0.3	65.1 \pm 0.2	79.7 \pm 0.2	79.6 \pm 0.0
Satlas [5]	55.7 \pm 1.2	37.3 \pm 0.1	67.3 \pm 0.7	45.9 \pm 0.1	64.8 \pm 0.2	62.2 \pm 0.2	77.9 \pm 0.2	77.9 \pm 0.0
Croma [31]	59.9 \pm 0.5	37.2 \pm 0.1	70.7 \pm 0.2	46.1 \pm 0.1	67.1 \pm 0.1	63.6 \pm 0.3	80.7 \pm 0.2	79.1 \pm 0.0
SSL4EO [89]	<u>63.1</u> \pm 0.2	<u>39.6</u> \pm 0.1	<u>72.5</u> \pm 0.2	<u>47.9</u> \pm 0.1	<u>71.1</u> \pm 0.3	<u>67.8</u> \pm 0.2	<u>83.2</u> \pm 0.1	<u>81.1</u> \pm 0.0
DOFA [93]	59.9 \pm 0.6	37.8 \pm 0.2	70.1 \pm 0.2	46.1 \pm 0.1	66.9 \pm 0.2	62.7 \pm 0.2	80.1 \pm 0.0	77.3 \pm 0.1
SeCo-Eco (ours)	66.8 \pm 0.3	41.4 \pm 0.1	75.0 \pm 0.1	49.9 \pm 0.1	74.1 \pm 0.2	71.7 \pm 0.2	85.3 \pm 0.0	84.0 \pm 0.0

Table A-3. Linear probing and K-Nearest Neighbor performance across multiple metrics for the BigEarthNet-10% task. **Best**, second best.

Model	EU-Forest [60]							
	Macro AUROC \uparrow		Macro F1 \uparrow		Micro AUROC \uparrow		Micro F1 \uparrow	
	LP	5-NN	LP	5-NN	LP	5-NN	LP	5-NN
SeCo [58]	82.6 \pm 0.0	63.9 \pm 0.3	12.3 \pm 0.7	18.2 \pm 0.3	90.6 \pm 0.1	77.6 \pm 0.2	31.3 \pm 0.9	30.6 \pm 0.2
SatMAE [16]	<u>84.6</u> \pm 0.2	66.7 \pm 0.4	15.0 \pm 0.7	21.0 \pm 0.3	<u>91.6</u> \pm 0.1	79.8 \pm 0.2	35.7 \pm 0.9	33.3 \pm 0.1
Satlas [5]	81.1 \pm 0.3	62.7 \pm 0.3	10.1 \pm 0.4	17.5 \pm 0.3	89.6 \pm 0.1	76.7 \pm 0.2	29.8 \pm 1.5	30.0 \pm 0.2
Croma [31]	82.9 \pm 0.3	63.6 \pm 0.3	12.2 \pm 0.7	18.1 \pm 0.3	90.5 \pm 0.2	77.8 \pm 0.2	32.3 \pm 0.9	30.9 \pm 0.2
SSL4EO [89]	83.9 \pm 0.0	65.0 \pm 0.3	11.6 \pm 0.4	19.3 \pm 0.3	91.2 \pm 0.2	78.5 \pm 0.2	32.6 \pm 0.1	31.5 \pm 0.2
DOFA [93]	83.1 \pm 0.1	63.1 \pm 0.5	13.5 \pm 0.5	17.6 \pm 0.5	90.7 \pm 0.1	77.3 \pm 0.3	<u>34.8</u> \pm 0.9	29.9 \pm 0.3
SeCo-Eco (ours)	84.8 \pm 0.2	<u>65.6</u> \pm 0.2	<u>14.8</u> \pm 0.6	<u>19.9</u> \pm 0.2	91.7 \pm 0.1	<u>79.0</u> \pm 0.1	35.7 \pm 0.4	<u>32.4</u> \pm 0.2

Table A-4. Linear probing and K-Nearest Neighbor performance across multiple metrics for the EUForest task. **Best**, second best.

Model	TreeSatAI [1]							
	Macro F1 \uparrow		Macro MAP \uparrow		Micro F1 \uparrow		Micro MAP \uparrow	
	LP	5-NN	LP	5-NN	LP	5-NN	LP	5-NN
SeCo [58]	10.1 \pm 0.0	24.3	24.3 \pm 0.0	20.5	23.4 \pm 0.0	35.2	44.6 \pm 0.0	34.6
SatMAE [16]	21.0 \pm 0.1	33.7	36.8 \pm 0.1	35.8	46.8 \pm 0.3	43.7	58.0 \pm 0.1	52.3
Satlas [5]	17.8 \pm 0.0	30.1	32.4 \pm 0.0	27.9	42.9 \pm 0.0	40.8	54.2 \pm 0.0	45.4
Croma [31]	<u>20.3</u> \pm 0.0	30.1	<u>34.9</u> \pm 0.0	27.8	<u>43.8</u> \pm 0.0	40.7	<u>56.6</u> \pm 0.0	45.6
SSL4EO [89]	18.2 \pm 0.0	<u>30.2</u>	33.1 \pm 0.0	28.4	42.3 \pm 0.0	<u>40.9</u>	54.5 \pm 0.0	<u>46.0</u>
DOFA [93]	14.7 \pm 0.0	26.2	28.7 \pm 0.0	21.9	35.1 \pm 0.0	37.3	50.8 \pm 0.0	37.5
SeCo-Eco (ours)	19.2 \pm 0.0	29.7	34.3 \pm 0.0	<u>29.0</u>	42.7 \pm 0.0	40.6	54.8 \pm 0.0	45.7

Table A-5. Linear probing and K-Nearest Neighbor performance across multiple metrics for the TreeSatAI task. Due to the fixed splits, no standard deviation can be reported for K-Nearest Neighbor probing. **Best**, second best.

Model	Biomes [65]									
	Macro Acc \uparrow		Macro AUROC \uparrow		Macro F1 \uparrow		Micro Acc \uparrow		Micro F1 \uparrow	
	LP	10-NN	LP	10-NN	LP	10-NN	LP	10-NN	LP	10-NN
SeCo [58]	40.0 \pm 0.4	35.4 \pm 0.7	91.2 \pm 0.6	79.8 \pm 1.0	41.6 \pm 0.5	36.9 \pm 1.0	62.7 \pm 0.5	59.2 \pm 0.5	62.7 \pm 0.5	59.2 \pm 0.5
SatMAE [16]	49.9 \pm 1.0	46.1 \pm 0.5	93.7 \pm 0.4	88.8 \pm 0.4	51.4 \pm 1.1	47.8 \pm 0.7	69.0 \pm 0.5	66.7 \pm 0.6	69.0 \pm 0.5	66.7 \pm 0.6
Satlas [5]	47.1 \pm 1.4	45.9 \pm 0.7	92.8 \pm 0.5	88.4 \pm 0.4	48.3 \pm 1.6	47.6 \pm 0.9	65.6 \pm 0.8	65.1 \pm 0.5	65.6 \pm 0.8	65.1 \pm 0.5
Croma [31]	46.2 \pm 1.8	41.2 \pm 0.5	92.2 \pm 0.4	85.7 \pm 0.6	47.2 \pm 1.4	42.2 \pm 0.6	65.7 \pm 0.7	61.7 \pm 0.3	65.7 \pm 0.7	61.7 \pm 0.3
SSL4EO [89]	<u>51.3</u> \pm 0.9	<u>48.2</u> \pm 0.5	<u>94.3</u> \pm 0.6	<u>89.6</u> \pm 0.8	<u>53.4</u> \pm 1.0	<u>49.7</u> \pm 0.5	<u>70.4</u> \pm 0.5	<u>67.6</u> \pm 0.6	<u>70.4</u> \pm 0.5	<u>67.6</u> \pm 0.6
DOFA [93]	48.1 \pm 1.4	41.8 \pm 0.4	92.9 \pm 0.3	85.7 \pm 0.6	49.7 \pm 1.3	43.0 \pm 0.5	66.4 \pm 0.6	61.8 \pm 0.5	66.4 \pm 0.6	61.8 \pm 0.5
SeCo-Eco (ours)	53.9 \pm 0.7	49.3 \pm 0.7	95.5 \pm 0.4	90.0 \pm 0.7	56.1 \pm 0.7	51.2 \pm 0.9	72.9 \pm 0.5	69.4 \pm 0.4	72.9 \pm 0.5	69.4 \pm 0.4

Table A-6. Linear probing and K-Nearest Neighbor performance across multiple metrics for the biomes classification task. **Best**, second best.

Model	CAVM [73]									
	Macro Acc \uparrow		Macro AUROC \uparrow		Macro F1 \uparrow		Micro Acc \uparrow		Micro F1 \uparrow	
	LP	20-NN	LP	20-NN	LP	20-NN	LP	20-NN	LP	20-NN
SeCo [58]	53.2 \pm 0.6	50.3 \pm 0.6	87.3 \pm 0.3	85.6 \pm 0.3	54.5 \pm 0.7	52.1 \pm 0.7	61.4 \pm 0.6	60.6 \pm 0.5	61.4 \pm 0.6	60.6 \pm 0.5
SatMAE [16]	55.2 \pm 1.6	54.0 \pm 0.6	88.3 \pm 0.3	87.9 \pm 0.3	56.4 \pm 1.5	55.8 \pm 0.7	63.0 \pm 0.5	63.5 \pm 0.5	63.0 \pm 0.5	63.5 \pm 0.5
Satlas [5]	52.7 \pm 2.1	51.5 \pm 0.4	87.6 \pm 0.3	86.6 \pm 0.3	53.8 \pm 2.0	53.2 \pm 0.5	61.2 \pm 0.5	61.2 \pm 0.5	61.2 \pm 0.5	61.2 \pm 0.5
Croma [31]	52.7 \pm 1.3	50.1 \pm 0.7	87.4 \pm 0.3	85.6 \pm 0.4	53.7 \pm 1.2	51.6 \pm 0.8	61.0 \pm 0.7	60.3 \pm 0.6	61.0 \pm 0.7	60.3 \pm 0.6
SSL4EO [89]	<u>56.0</u> \pm 0.5	<u>55.0</u> \pm 0.6	<u>88.9</u> \pm 0.3	<u>88.2</u> \pm 0.3	<u>57.5</u> \pm 0.6	<u>56.9</u> \pm 0.7	<u>63.7</u> \pm 0.6	<u>63.7</u> \pm 0.5	<u>63.7</u> \pm 0.6	<u>63.7</u> \pm 0.5
DOFA [93]	55.3 \pm 1.8	51.7 \pm 0.5	88.2 \pm 0.4	87.0 \pm 0.3	56.5 \pm 1.6	53.6 \pm 0.6	62.4 \pm 0.8	62.2 \pm 0.4	62.4 \pm 0.8	62.2 \pm 0.4
SeCo-Eco (ours)	58.1 \pm 1.2	58.0 \pm 0.7	89.9 \pm 0.3	89.2 \pm 0.4	59.4 \pm 1.0	59.5 \pm 0.8	65.3 \pm 0.5	65.6 \pm 0.6	65.3 \pm 0.5	65.6 \pm 0.6

Table A-7. Linear probing and K-Nearest Neighbor performance across multiple metrics for the CAVM classification task. **Best**, second best.

Model	BioMassters [61]					
	Mean R ² \uparrow		Mean MAE \downarrow		Mean RMSE \downarrow	
	LP	1-NN	LP	1-NN	LP	1-NN
SeCo [58]	51.3 \pm 0.0	-19.2	3.9 \pm 0.0	<u>7.0</u>	5.8 \pm 0.0	<u>11.0</u>
SatMAE [16]	59.5 \pm 0.6	-18.0	3.6 \pm 0.0	<u>7.0</u>	5.3 \pm 0.0	<u>11.0</u>
Satlas [5]	62.5 \pm 0.9	-17.8	3.3 \pm 0.1	<u>7.0</u>	4.9 \pm 0.1	<u>11.0</u>
Croma [31]	58.5 \pm 0.2	-18.1	3.5 \pm 0.0	<u>7.0</u>	5.3 \pm 0.0	<u>11.0</u>
SSL4EO [89]	<u>71.4</u> \pm 0.0	<u>-16.8</u>	<u>2.8</u> \pm 0.0	6.9	4.2 \pm 0.0	10.9
DOFA [93]	63.1 \pm 0.4	-18.3	3.2 \pm 0.0	<u>7.0</u>	4.8 \pm 0.0	<u>11.0</u>
SeCo-Eco (ours)	75.2 \pm 0.1	-16.3	2.5 \pm 0.0	6.9	3.8 \pm 0.0	10.9

Table A-8. Linear probing and K-Nearest Neighbor performance across multiple metrics for the BioMassters task. Due to the fixed splits, no standard deviation can be reported for K-Nearest Neighbor probing. **Best**, second best.

Model	CHELSA Climate [47] - Temperature & Precipitation							
	Temp MAE ↓		Temp R ² ↑		Prec MAE ↓		Prec R ² ↑	
	LP	10-NN	LP	10-NN	LP	10-NN	LP	10-NN
SeCo [58]	572.3 ± 1.1	547.8 ± 1.7	63.1 ± 0.3	61.3 ± 0.3	33380.8 ± 291.5	30725.5 ± 171.8	60.3 ± 0.7	60.7 ± 0.8
SatMAE [16]	<u>482.0</u> ± 2.3	411.4 ± 1.2	<u>74.4</u> ± 0.2	<u>76.1</u> ± 0.2	30999.5 ± 314.9	<u>27087.1</u> ± 135.4	65.2 ± 0.4	<u>67.1</u> ± 0.5
Satlas [5]	595.1 ± 3.4	474.7 ± 3.6	62.1 ± 0.4	69.4 ± 0.7	36698.8 ± 685.3	29535.8 ± 95.1	55.9 ± 1.0	62.4 ± 0.7
Croma [31]	511.5 ± 2.5	505.5 ± 1.6	71.1 ± 0.2	66.4 ± 0.2	32887.8 ± 350.8	30974.2 ± 96.6	61.4 ± 0.6	60.3 ± 0.4
SSL4EO [89]	496.1 ± 1.1	<u>410.7</u> ± 0.8	72.4 ± 0.2	75.8 ± 0.3	<u>30960.7</u> ± 154.7	27989.7 ± 148.3	<u>65.5</u> ± 0.4	65.4 ± 0.4
DOFA [93]	576.0 ± 0.7	505.9 ± 0.9	63.9 ± 0.3	66.9 ± 0.3	34860.1 ± 297.0	30311.1 ± 182.9	59.7 ± 0.5	59.9 ± 0.7
SeCo-Eco (ours)	411.4 ± 0.9	364.8 ± 0.7	80.7 ± 0.2	80.5 ± 0.2	27695.5 ± 74.8	25946.7 ± 72.6	70.2 ± 0.3	69.5 ± 0.4

Table A-9. Linear probing and K-Nearest Neighbor performance overview for the CHELSA Climate task. We break down the predictions for temperature and precipitation. **Best**, second best.

Model	CHELSA Climate [47] - Evapotranspiration & Site Water Balance							
	Evap MAE ↓		Evap R ² ↑		Swb MAE ↓		Swb R ² ↑	
	LP	10-NN	LP	10-NN	LP	10-NN	LP	10-NN
SeCo [58]	2131.6 ± 8.7	2068.0 ± 9.8	68.9 ± 0.2	67.1 ± 0.1	24903.7 ± 137.7	23878.3 ± 143.0	80.9 ± 0.2	80.5 ± 0.2
SatMAE [16]	<u>1760.4</u> ± 6.7	1564.7 ± 5.4	<u>79.2</u> ± 0.2	80.0 ± 0.2	20999.2 ± 87.5	19055.6 ± 66.2	86.9 ± 0.1	87.7 ± 0.1
Satlas [5]	2093.3 ± 6.4	1761.5 ± 10.3	70.9 ± 0.2	75.3 ± 0.5	24115.2 ± 131.3	20772.5 ± 116.8	83.4 ± 0.2	85.5 ± 0.1
Croma [31]	1872.4 ± 27.2	1882.2 ± 5.6	76.2 ± 0.5	73.2 ± 0.2	23003.2 ± 270.2	21593.9 ± 79.0	84.6 ± 0.4	84.7 ± 0.2
SSL4EO [89]	1786.0 ± 3.4	<u>1522.8</u> ± 3.0	78.6 ± 0.2	<u>81.1</u> ± 0.2	<u>20444.3</u> ± 58.1	<u>18155.6</u> ± 42.4	<u>87.7</u> ± 0.1	<u>88.9</u> ± 0.1
DOFA [93]	2086.1 ± 3.5	1911.6 ± 6.2	71.2 ± 0.3	72.0 ± 0.3	23943.5 ± 38.3	22370.0 ± 60.8	83.4 ± 0.1	83.5 ± 0.2
SeCo-Eco (ours)	1537.6 ± 4.2	1391.2 ± 3.3	83.7 ± 0.1	83.9 ± 0.2	18567.4 ± 90.2	17257.4 ± 50.6	89.6 ± 0.1	89.9 ± 0.1

Table A-10. Linear probing and K-Nearest Neighbor performance overview for the CHELSA Climate task. We break down the predictions for evapotranspiration and site water balance. **Best**, second best.