
EcoCast: A Spatio-Temporal Model for Continual Biodiversity and Climate Risk Forecasting

Hammed A. Akande

Department of Biology
Concordia University
Montreal, QC, Canada

akandehammedadedamola@gmail.com

Abdulrauf A. Gidado

Department of Computer Science and Mathematics
Algoma University
ON, Canada

abdulrauf.gidado@algomau.ca

Abstract

Increasing climate change and habitat loss are driving unprecedented shifts in species distributions. Conservation professionals urgently need timely, high-resolution predictions of biodiversity risks, especially in ecologically diverse regions like Africa. We propose *EcoCast*, a spatio-temporal model designed for continual biodiversity and climate risk forecasting. Utilizing multisource satellite imagery, climate data, and citizen science occurrence records, *EcoCast* predicts near-term (monthly to seasonal) shifts in species distributions through sequence-based transformers that model spatio-temporal environmental dependencies. The architecture is designed with support for continual learning to enable future operational deployment with new data streams. Our pilot study in Africa shows promising improvements in forecasting distributions of selected bird species compared to a Random Forest baseline, highlighting *EcoCast*'s potential to inform targeted conservation policies. By demonstrating an end-to-end pipeline from multi-modal data ingestion to operational forecasting, *EcoCast* bridges the gap between cutting-edge machine learning and biodiversity management, ultimately guiding data-driven strategies for climate resilience and ecosystem conservation throughout Africa.

1 Introduction

Climate change, in addition to widespread habitat degradation, is reshaping ecosystems across continental Africa, a threat intensified by increasing temperatures, changing precipitation patterns, and extreme weather events [1]. These changing environments, largely driven by anthropogenic pressures, significantly affect biodiversity, particularly bird species that are highly sensitive to environmental change and serve as early warning indicators of ecosystem health [2]. However, conservation professionals and policymakers often lack timely, fine-grained forecasts of changing species distribution or habitat quality over time. Current methods often involve static Species Distribution Models (SDMs) or climate risk assessments, which cannot keep pace with rapidly changing environmental conditions. In Africa, the lack of up-to-date biodiversity data has made it nearly impossible to track species declines or respond proactively [3].

We propose *EcoCast*, a spatio-temporal model tailored to the dynamic nature of African ecosystems. *EcoCast* is designed to ingest real-time and near-real-time data, from satellite imagery and climate data to citizen science observations, to generate frequently updated forecasts of shifting species distribution and habitat quality changes. By piloting *EcoCast* in Africa, we illustrate how advanced machine learning can directly inform local conservation efforts, such as identifying emerging biodiversity hotspots, areas at risk of habitat loss, or assessing the ecological impact of infrastructure projects. In doing so, we aim to demonstrate a new paradigm where a single large-scale model can serve as a continuously improving “environmental model” for conservation decisions.

1.1 Key Contributions

We develop a large-scale pre-trained model that integrates multimodal data (satellite images, species occurrences, climate variables) for biodiversity forecasting. To our knowledge, *EcoCast* is among the first transformer-based models applied to ecological and climate risk data in Africa, learning generalizable representations that transfer across species and regions. The architecture supports continuous fine-tuning that can adapt to new data streams without catastrophic forgetting [4], enabling *EcoCast* to maintain high predictive performance in a changing environment. Furthermore, *EcoCast* is designed to provide a user-friendly interface for conservation professionals, allowing stakeholders to visualize model forecasts, explore scenarios, and contribute new observations.

2 Related Work

Over the years, machine learning techniques have been used to address key environmental challenges, including forest cover change detection [5], crop-yield prediction from remote sensing [6], and SDMs [7]. Recent advances in spatio-temporal deep learning provide a foundation for our approach. In climate and Earth observation, transformer models (e.g., *EarthFormer*) and ConvLSTM capture space–time dependencies for tasks such as precipitation and ENSO forecasting [8, 9]. *EcoCast* builds on these advances by adopting a similar Transformer-based architecture, but extends it to the multi-modal context of ecological forecasting with a persistent model that can be incrementally retrained with new data to stay current. The framework incorporates rehearsal-style updates combined with stability mechanisms to mitigate forgetting and support incremental retraining [10]. Recent work in ecological forecasting distinguishes between long-term climate projection-based SDMs and near-term operational forecasts that inform immediate management decisions [11, 12]. *EcoCast* extends this paradigm to biodiversity applications through transformer-based sequence modeling, bridging the gap between geophysical forecasting architectures and species distribution modeling.

3 Proposed Approach

EcoCast’s workflow is structured into three phases (Figure 1 provides a conceptual overview). First, we create a pre-trained spatio-temporal model using historical multimodal data. Next, the framework is designed with a continuous learning pipeline to ingest new data and update the model incrementally. Finally, we envision deployment through practitioner-focused tools for visualization and feedback.

We use a monthly environmental time series covering the period from 2016 to 2023. Satellite inputs are monthly Sentinel-2 composites derived from 30 m products and resampled onto a fixed continent-scale grid, and climate inputs are monthly ERA5 variables (Appendix A.4). Bird occurrence data were obtained from Global Biodiversity Information Facility (GBIF) and aggregated by month to a 0.1° grid ($\approx 10\text{--}11$ km at the equator) across Africa. We pre-train our model on data from 2016 to 2021 and fine-tune it on 2022 before evaluating on 2023. By pre-training on this broad dataset, the model learns general spatio-temporal patterns (such as seasonal vegetation cycles or climate anomalies) relevant to the species’ ecology.

3.1 Operational Near-term Forecasting Framework

We train a transformer-based architecture for spatio-temporal sequence modeling. Unlike traditional SDMs that require projected future climate scenarios (e.g., RCP4.5, RCP8.5) to forecast decades ahead, *EcoCast* implements an operational near-term forecasting paradigm designed for real-time conservation decision-making at monthly to seasonal timescales (see Appendix A.5 for detailed comparison).

At each time step t , the model receives a temporal sequence of environmental covariates $\mathbf{x}_{t-L+1:t} = (x_{t-L+1}, \dots, x_t)$, where each monthly observation x_τ combines satellite imagery features (Sentinel-2 bands, vegetation indices) and climate variables (ERA5 temperature, precipitation, etc.). The architecture predicts species occurrence y_{t+1} for the *next* month using observed historical data (Equation 1):

$$\min_{\theta} \sum_{(x_{t-L+1:t}, y_{t+1})} \mathcal{L}(f_{\theta}(\mathbf{x}_{t-L+1:t}), y_{t+1}), \quad (1)$$

where $\mathcal{L}$ is a class-imbalance-robust loss and f_θ denotes our transformer encoder with parameters θ . We set the sequence length $L = 12$ months to capture annual seasonality critical for avian phenology while maintaining computational efficiency.

We summarize resampled Sentinel-2 inputs within each 0.1° grid cell per month using band-wise statistics, and similarly aggregate monthly ERA5 climate variables at the same spatial resolution. At each grid cell and month t , we combine satellite and climate features into a unified environmental vector $\mathbf{x}_t \in \mathbb{R}^F$, where F includes spectral bands, vegetation indices, and climate variables (see Appendix A.8 for detailed illustration). To capture temporal dynamics and seasonal patterns, the model ingests sequences of L consecutive monthly vectors $(\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t) \in \mathbb{R}^{L \times F}$ and predicts species occurrence y_{t+1} at the next time step.

Given that our species data are presence-only, for training, we generate pseudo-absences within the study area and the same temporal window. We thin occurrences spatially to the 0.1° grid to reduce clustering and sampling bias [13, 14]. Additionally, we compute a sampling-effort covariate (based on monthly GBIF records) to help the model account for areas of high observer effort. The same pseudo-absence and thinning protocol is applied to the baseline model for fairness. In addition, we use stratified mini-batches that balance presence and pseudo-absence instances per species and a class-weighted approach to stabilize learning on rare positives.

For continual learning, our framework is designed to perform constrained updates when new data at $t+1$ arrive. The architecture supports rehearsal-based updates with weight regularization, maintaining a fixed-size replay buffer of past examples and applying Elastic Weight Consolidation (EWC) to prevent catastrophic forgetting. The fine-tuning objective combines supervised loss on new data with a regularization term (Equation 2):

$$\min_{\theta'} \sum_{k=1}^K \mathcal{L}(f_{\theta'}(x_{t+k}), y_{t+k}) + \lambda \Omega_{\text{EWC}}(\theta', \theta), \quad (2)$$

where Ω_{EWC} penalizes deviations from important parameters estimated on prior data. For the current evaluation, we employ rolling-origin validation with warm-start initialization, where each temporal window trains on all available historical data up to that point. This simulates operational deployment conditions while providing a rigorous baseline for future continual learning work.

To interpret predictions, we compute transformer attention rollout over space–time to highlight influential regions and months, and apply patch-wise permutation tests to verify that attributions reflect signals. For additional diagnostics, we use Grad-CAM-style attribution adapted to ViTs [15, 16]. We flag cases where attributions may contradict ecological priors (for instance, focusing on clouds rather than vegetation change). We adopt spatio-temporal block cross-validation where folds are defined by spatial blocks and held-out years to prevent leakage and over-optimism [17]. Thresholds for F1 are selected on the validation split per species; AUC is computed per species and then macro-averaged across species.

4 Results

A pilot study was conducted on five bird species (see Appendix A.3 for the full list) across Africa. The monthly Sentinel-2 and ERA5 inputs were prepared as described in Section 3, and the GBIF records were aggregated to 0.1° cells by month. As a baseline, we trained species-specific Random Forest classifiers using static covariates plus occurrence records, with the *same* pseudo-absence, thinning, and block CV protocol as the main *EcoCast* model.

4.1 EcoCast Implementation and Results

As described in Section 3, EcoCast performs sequence-to-point forecasting: 12-month environmental sequences (satellite + climate features) predict next-month species occurrence. This operational framework was trained on 2016–2021 data, fine-tuned on 2022, and evaluated on held-out 2023 observations using rolling-origin validation to simulate real-world deployment conditions.

Table 1 demonstrates EcoCast’s performance gains over the Random Forest baseline. The transformer architecture’s ability to model long-range temporal dependencies via self-attention, learn species-

specific phenological patterns through positional encoding, and leverage cross-species ecological signals through joint multi-label training yields macro-averaged F1 and PR-AUC improvements of +34 and +43 percentage points, respectively. These results validate that explicit sequence modeling through transformer architectures is critical for near-term biodiversity forecasting, providing a foundation for the continual learning extensions and multi-task predictions planned for future iterations (see Appendix A.6 for detailed architectural analysis).

Table 1: Comparative performance on 2023 holdout data for five bird species. **RF-ROE** (Random Forest with Rolling Origin Evaluation) trains separate models per species using flattened tabular features without explicit temporal modeling. **EcoCast** employs a transformer encoder over environmental sequences to predict species occurrence at month $t + 1$.

Model	F1 macro	PR AUC macro
Baseline (Random Forest ROE)	0.31	0.29
EcoCast (t+1 forecast)	0.65	0.72

5 Pathway to Impact

We plan to partner with groups like the Wildlife Conservation Society to validate *EcoCast* outputs on the ground. The Federal Ministry of Environment could use *EcoCast* to inform protected area management. This model will provide real-time forecasts highlighting areas at the highest biodiversity risk, enabling swift allocation of resources. Additionally, this could guide scenario analysis for land-use planning, evaluating the potential impacts of infrastructure projects (for instance, new roads) on bird habitats, and inform more sustainable development strategies. Once validated in Africa, the approach can extend to other continent-wide biodiversity hotspots. We would develop a robust API to enable African conservation NGOs and researchers to integrate local data, thereby further improving *EcoCast*’s coverage and accuracy. Protecting biodiversity hotspots can sustain ecotourism and preserve ecosystem services, supporting livelihoods. *EcoCast* can encourage citizen-science participation, raising ecological awareness and empowering local stewardship.

6 Conclusion

We presented *EcoCast*, a spatio-temporal model for continual forecasting of biodiversity and climate-related risks. *EcoCast* demonstrates how advances in machine learning, from powerful Transformer architectures to continual learning techniques, can be utilized to address emerging challenges in conservation. Our case study in Africa showed that *EcoCast* substantially outperforms traditional static models in predicting near-term species distribution shifts, and more importantly, provides actionable insights to conservation practitioners.

This work bridges the gap between cutting-edge AI and traditional ecology. Our proposed model is designed as an evolving model, continuously refreshed with new data, and well-suited to addressing climate change’s non-stationary impact on ecosystems. By integrating data and delivering user-centric tools, we ensure that the model’s intelligence is accessible to important stakeholders who need it most – conservation practitioners making important conservation decisions. Moving forward, we aim to expand *EcoCast*’s coverage and involve a broader community of experts to carefully evaluate its predictions. We will also explore scaling up the model (both in terms of architecture and training data) to capture continental-scale patterns and to include other taxa beyond birds. Additionally, policy engagement will be important. Demonstrating success stories where *EcoCast*-informed interventions helped protect a species or habitat will build confidence among policymakers to formally incorporate such tools into conservation planning processes. *EcoCast* shows a promising pathway for AI for Good in the environmental domain. By continuously learning from the planet’s data and looping humans in decision-making, such systems can help us stay one step ahead in preserving biodiversity and fostering global climate resilience.

References

- [1] Dejene, Sinatayehu Impact of climate change on biodiversity and associated key ecosystem services in Africa: a systematic review 10.1080/20964129.2018.1530054, Ecosystem Health and Sustainability 2018.

- [2] Moussy, C., et al. A quantitative global review of species population monitoring. *Conservation biology : the journal of the Society for Conservation Biology*, 36(1), e13721. <https://doi.org/10.1111/cobi.13721> , 2022.
- [3] Odoligie Imarhiagbe, Wisdom Oghenevwogaga Egboduku, and Beluchukwu Joseph Nwankwo A review of the biodiversity conservation status of Nigeria *Journal of Wildlife and Biodiversity* 4 (1): 73-83, DOI: 10.22120/jwb.2019.115501.1096, 2020.
- [4] Yu, Hao and Yang, Xin and Gao, Xin and Feng, Yihui and Wang, Hao and Kang, Yan and Li, Tianrui Overcoming Spatial-Temporal Catastrophic Forgetting for Federated Class-Incremental Learning, *Proceedings of the 32nd ACM International Conference on Multimedia*, 10.1145/3664647.3681384, 2024.
- [5] M. C. Hansen et al. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.
- [6] J. You et al. Deep gaussian process for crop yield prediction. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.
- [7] J. Elith and J. R. Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40:677–697, 2009.
- [8] M. Gao et al. Earthformer: Exploring space-time transformers for earth system forecasting. *arXiv preprint arXiv:2207.05833*, 2022.
- [9] X. Shi et al. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.
- [10] S. A. Rebuffi, A. Kolesnikov, and C. H. Lampert. icarl: Incremental classifier and representation learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [11] Dietze, Michael C and others Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, volume=115, number=7, pages=1424–1432, year=2018
- [12] Anderson-Teixeira, Kristina J and others Ecological forecasts to inform near-term management of threats to biodiversity. *Global Change Biology*, volume=26, number=10, pages=5816–5828, year=2020
- [13] Phillips, Steven J. and Dudík, Miroslav and Elith, Jane and Graham, Catherine H. and Lehmann, Anthony and Leathwick, John and Ferrier, Simon Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 10.1890/07-2153.1, 2009.
- [14] Barbet-Massin, Morgane and Jiguet, Frédéric and Albert, Cécile Hélène and Thuiller, Wilfried Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 10.1111/j.2041-210X.2011.00172.x, 2012.
- [15] Abnar, Samira and Zuidema, Willem Quantifying Attention Flow in Transformers, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, Association for Computational Linguistics, 10.18653/v1/2020.acl-main.38, 2020.
- [16] Chefer, Hila and Gur, Shir and Wolf, Lior Transformer Interpretability Beyond Attention Visualization, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10.1109/CVPR46437.2021.00084, 2021.
- [17] Valavi, Roozbeh and Elith, Jane and Lahoz-Monfort, José J. and Guillera-Aroita, Gurutzeta blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models, *Methods in Ecology and Evolution*, 10.1111/2041-210X.13107, 2019.

A Appendix: Conceptual Diagram, Selected Species, and Environmental Variables

A.1 Conceptual architecture of EcoCast

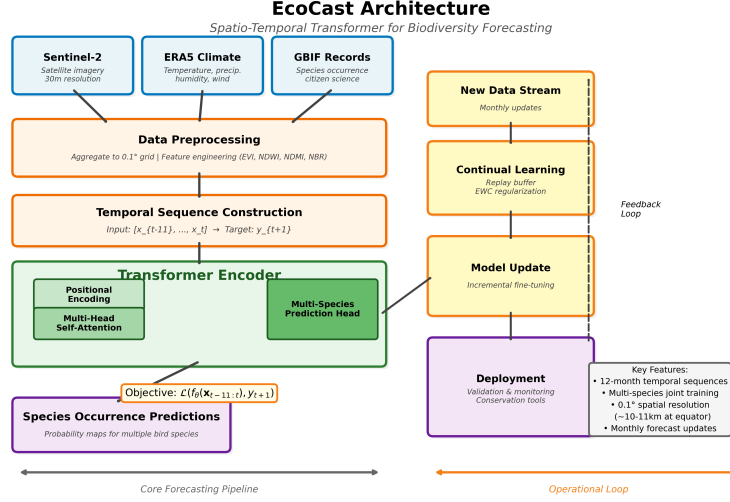


Figure 1: Conceptual architecture: *EcoCast* ingests monthly Sentinel-2 and ERA5 covariates aggregated to a 10 km grid, stacks 12-month temporal sequences, and predicts multi-species occurrence with a transformer encoder. The model performs $t + 1$ forecasting: given environmental features from months $t - 11$ through t , it predicts species occurrence at month $t + 1$ using self-attention mechanisms to capture temporal dependencies and seasonal patterns.

A.2 EcoCast Predicted Probability of Occurrence

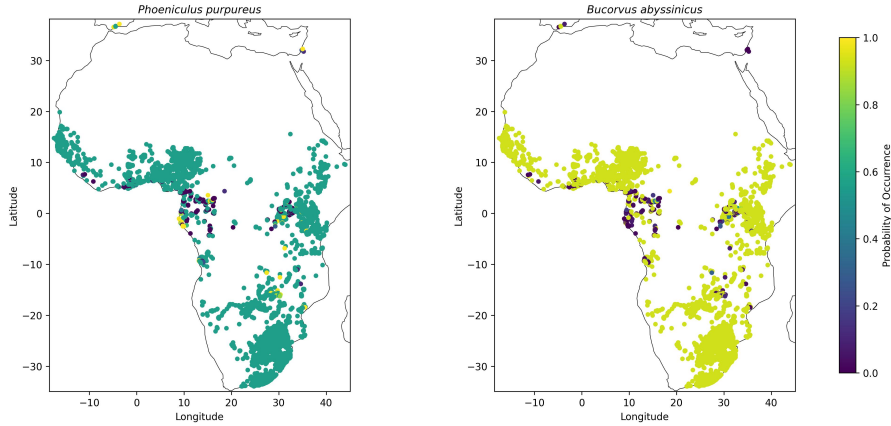


Figure 2: **EcoCast operational forecast (January 2024).** Predicted probability of occurrence across Africa for *Phoenixiculus purpureus* and *Bucorvus abyssinicus*. The model uses 12-month environmental sequences (January–December 2023) to forecast species distributions in January 2024, demonstrating EcoCast’s capability for operational near-term biodiversity forecasting. This represents a true out-of-sample forecast beyond the test period (2023), illustrating how conservation practitioners could use EcoCast to anticipate monthly shifts in species distributions.

A.3 Species List

The five focal species used in the pilot study are:

- *Psittacus erithacus* (African Grey Parrot)
- *Pitta angolensis* (African Pitta)
- *Tauraco persa* (Guinea Turaco)
- *Bucorvus abyssinicus* (Northern Ground-Hornbill)
- *Phoeniculus purpureus* (Green Wood-Hoopoe)

A.4 Environmental Covariates

For the Sentinel-2 composites, we derived band-wise statistics and vegetation indices: EVI, NDWI, NDMI, and NBR.

ERA5 climate variables included: temperature, relative humidity, total precipitation, wind speed, and surface pressure.

A.5 Detailed Comparison with Traditional SDMs

This approach fundamentally differs from climate projection-based SDMs in three ways:

1. **No future climate projections required:** EcoCast learns temporal dependencies and lagged ecological responses from observed environmental sequences, enabling predictions *without* requiring GCM outputs or emission scenario assumptions.
2. **Sequence modeling via self-attention:** Unlike Random Forest baselines that treat each (location, month) observation independently, the transformer architecture explicitly models inter-month dependencies through multi-head self-attention. This captures phenological patterns (e.g., “vegetation green-up in month $t - 3$ predicts migratory arrival at t ”) and lagged climate effects automatically, without manual feature engineering.
3. **Near-real-time operational capability:** ERA5 climate reanalysis becomes available within 5 days of observation (preliminary ERA5T) with final quality-assured data released 2–3 months later [18]. This enables monthly forecast updates aligned with conservation planning cycles, rather than static decadal projections.

Copernicus Climate Change Service Key update to climate dataset brings data at five days behind real time. Copernicus News, year=2025, url=<https://climate.copernicus.eu/key-update-climate-dataset-brings-data-five-days-behind-real-time>

Our architecture supports longer windows ($L \leq 24$) when deeper historical context is available.

A.6 Architectural Advantages

The performance gap between EcoCast and RF-ROE highlights the limitations of treating temporal biodiversity data as independent tabular observations. Random forests, while powerful for static prediction tasks, lack mechanisms to capture:

- **Temporal autocorrelation:** Bird presence in month t is highly predictive of month $t + 1$ for resident species, yet RF models each month independently.
- **Lagged environmental responses:** Species may respond to climate anomalies with 2–4 month delays (e.g., food availability following rainfall), which requires sequence modeling to capture effectively.
- **Seasonal periodicity:** Migratory species exhibit annual cycles that transformers encode through positional embeddings, whereas RF requires manual construction of seasonal features.

A.7 Glossary of Acronyms

- **ERA5:** Fifth-generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for global climate and weather.
- **SDM:** Species Distribution Model
- **GBIF:** Global Biodiversity Information Facility, an international database of species occurrence records.
- **ViT:** Vision Transformer, a Transformer-based architecture for image classification and analysis.
- **Grad-CAM:** Gradient-weighted Class Activation Mapping.

- **EVI**: Enhanced Vegetation Index, a vegetation index that corrects for atmospheric and soil background effects.
- **NDWI**: Normalized Difference Water Index, highlights water content in vegetation and soil.
- **NDMI**: Normalized Difference Moisture Index, sensitive to vegetation canopy water content.
- **NBR**: Normalized Burn Ratio, used to detect burned areas and vegetation disturbance.
- **ConvLSTM**: Convolutional Long Short-Term Memory network, a recurrent neural network that combines convolutional layers with LSTMs to capture spatio-temporal patterns.
- **EarthFormer**: A spatio-temporal Transformer model developed for Earth system forecasting
- **Block CV**: Block cross-validation, a spatial or spatio-temporal cross-validation scheme.
- **ROE**: Rolling Origin Evaluation. In this work, we pair ROE with Random Forest (RF-ROE) as a baseline species distribution model.

A.8 Temporal Alignment of Satellite and Climate Data

This section details the temporal alignment of satellite imagery and climate variables within EcoCast’s sequence modelling framework. **Both data sources undergo identical temporal processing**: satellite and climate features are concatenated at each monthly time step to form unified environmental vectors, which are then arranged into temporal sequences for input to the transformer model.

A.8.1 Unified Temporal Processing

EcoCast constructs environmental sequences by concatenating satellite and climate features *at each monthly time step* before sequence construction. At each grid cell and month t , we combine:

$$\mathbf{x}_t = [\mathbf{s}_t, \mathbf{c}_t] \in \mathbb{R}^{16} \quad (3)$$

where $\mathbf{s}_t \in \mathbb{R}^{11}$ contains Sentinel-2 bands (B2, B3, B4, B8, B11, B12) and vegetation indices (NDVI, EVI, NDMI, NBR, NDWI) for month t , and $\mathbf{c}_t \in \mathbb{R}^5$ contains ERA5 climate variables (2m temperature, relative humidity, precipitation, wind speed, surface pressure) for the same month. The model then ingests a sequence of these unified vectors:

$$\mathbf{X} = [\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t] \in \mathbb{R}^{L \times 16} \quad (4)$$

with $L = 12$ months, to predict species occurrence y_{t+1} at the subsequent time step. Figure 3 illustrates this construction process.

A.8.2 Key Design Principles

Three critical aspects ensure consistent temporal treatment:

1. **Matched temporal resolution**: Both satellite composites and climate aggregates are computed at monthly granularity. Sentinel-2 observations within month t are mosaicked and cloud-masked to produce representative monthly statistics, while ERA5 hourly data are averaged to monthly means.
2. **Aligned spatial resolution**: All features are resampled to a 0.1° grid ($\approx 10\text{--}11$ km at the equator) before concatenation, ensuring spatial correspondence between satellite pixels and climate variables.
3. **Uniform lag structure**: The 12-month lookback window applies to the *combined* feature vector $\mathbf{x}_t$, not to satellite and climate data separately. The transformer’s self-attention mechanism operates over $[\mathbf{x}_{t-11}, \dots, \mathbf{x}_t]$ as a unified sequence.

A.8.3 Architectural Implementation

EcoCast’s architecture is designed to learn joint spatio-temporal patterns across modalities. Formally, sequences are constructed *after* feature concatenation at each time step:

$$[\mathbf{x}_{t-11}, \dots, \mathbf{x}_t] = [[\mathbf{s}_{t-11}, \mathbf{c}_{t-11}], \dots, [\mathbf{s}_t, \mathbf{c}_t]] \longrightarrow y_{t+1} \quad (5)$$

This design enables the transformer to learn cross-modal temporal dependencies through self-attention. For example, the model can learn that vegetation greenness (satellite) in month $t - 3$ combined with precipitation (climate) in months $t - 4$ to $t - 2$ jointly predict species arrival at month $t + 1$, without requiring manual specification of these lagged interactions.

EcoCast Temporal Alignment: Unified Processing of Satellite and Climate Data

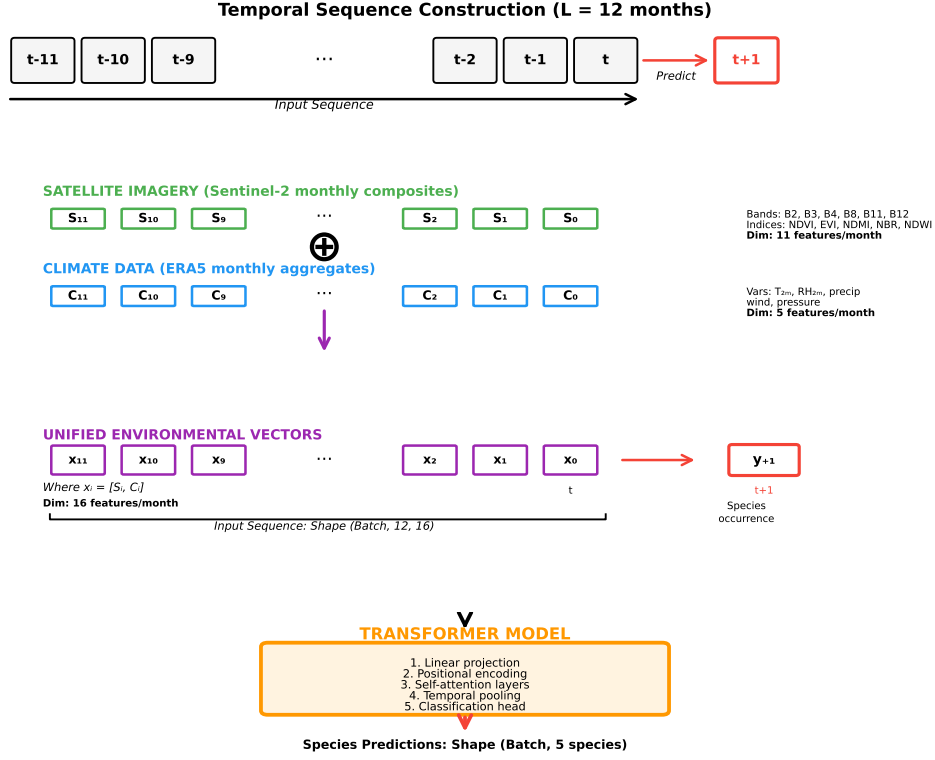


Figure 3: **Temporal alignment diagram for EcoCast.** Both satellite imagery (Sentinel-2) and climate data (ERA5) are processed as monthly sequences with lag $L = 12$ months. At each time step, satellite and climate features are concatenated ($\oplus$) into unified environmental vectors x_t before sequence construction. The transformer model attends over this 12-month input sequence to predict species occurrence at month $t + 1$. This design ensures both modalities contribute equally to temporal dependency learning through self-attention mechanisms.

A.8.4 Notation Summary

For clarity, we provide consolidated notation:

- t : Current time step (month)
- $L = 12$: Sequence length (lookback window)
- $s_t \in \mathbb{R}^{11}$: Satellite features at month t
- $c_t \in \mathbb{R}^5$: Climate features at month t
- $x_t = [s_t, c_t] \in \mathbb{R}^{16}$: Unified environmental vector
- $X = [x_{t-11}, \dots, x_t] \in \mathbb{R}^{12 \times 16}$: Input sequence
- $y_{t+1} \in \{0, 1\}^5$: Multi-species occurrence at next month (5 bird species)

This uniform treatment ensures that both satellite imagery and climate data contribute equally to the model's understanding of spatio-temporal environmental dynamics, enabling robust near-term biodiversity forecasting without requiring manual feature engineering or separate temporal modeling pipelines.