MemPerceiver: Adaptive Multi-Scale Memory Reveals Biome-Specific Temporal Fingerprints in Carbon Flux Prediction (CIKM 2026)

Gaoshan Bi · PhD student · Supervised by Prof. Po Yang

Pervasive Computing Research Group · School of Computer Science · The University of Sheffield

What Is Carbon Flux?

The exchange of CO_2 between terrestrial ecosystems and the atmosphere — measured by eddy-covariance (EC) towers.

~75% of greenhouse-gas emissions are CO_2

~1/3 of human CO_2 emissions absorbed by land ecosystems

Why measure it: quantifying this exchange underpins net-zero strategies, carbon accounting, and sustainable agriculture.

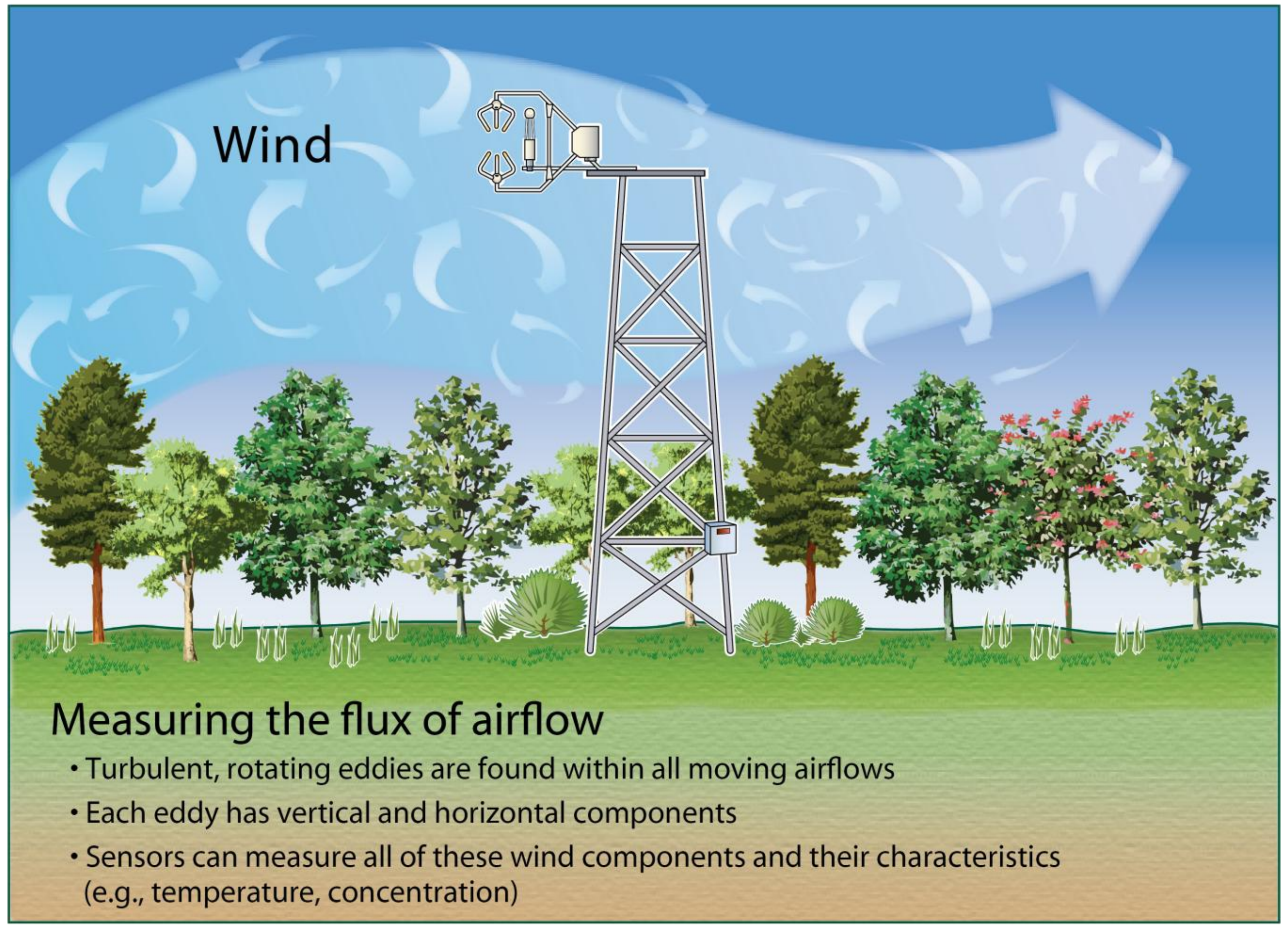
Net Ecosystem Exchange (NEE)

NEE < 0 — CO_2 flows from atmosphere into the ecosystem (uptake)

NEE > 0 — CO_2 is released from the ecosystem to the atmosphere

How a tower measures it

Turbulent eddies carry CO_2 up and down; fast sensors sample wind and gas concentration, and their covariance gives the flux.



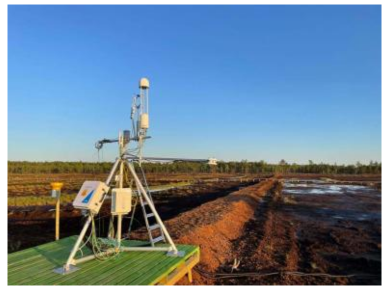
ESD14-039

Strengths

High accuracy · continuous

Limits

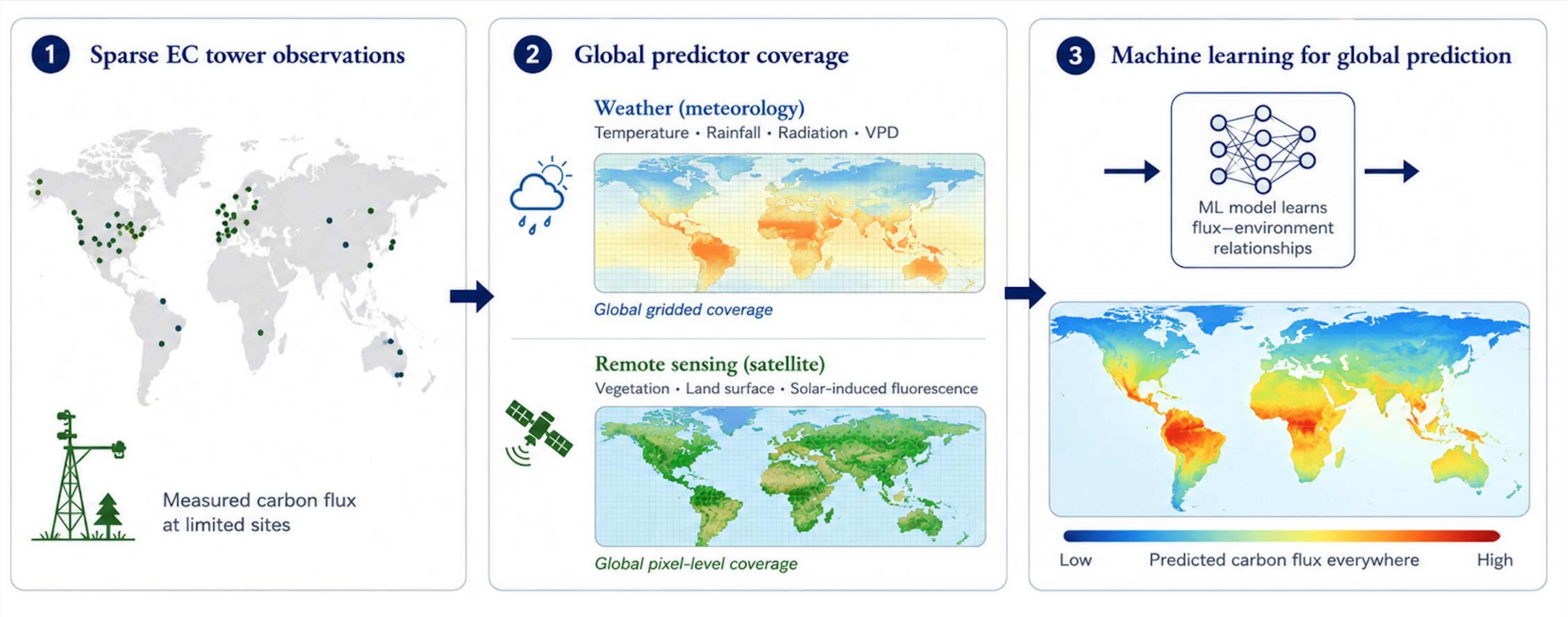
Expensive to run · footprint only 1–10 km² per tower



An EC tower in the field: sonic anemometer + gas analyser above the canopy, logging CO_2 exchange around the clock.

Upscaling From Sparse Towers to Global Scale

Flux tower observations are spatially sparse, whereas meteorological and satellite data provide near-global coverage.

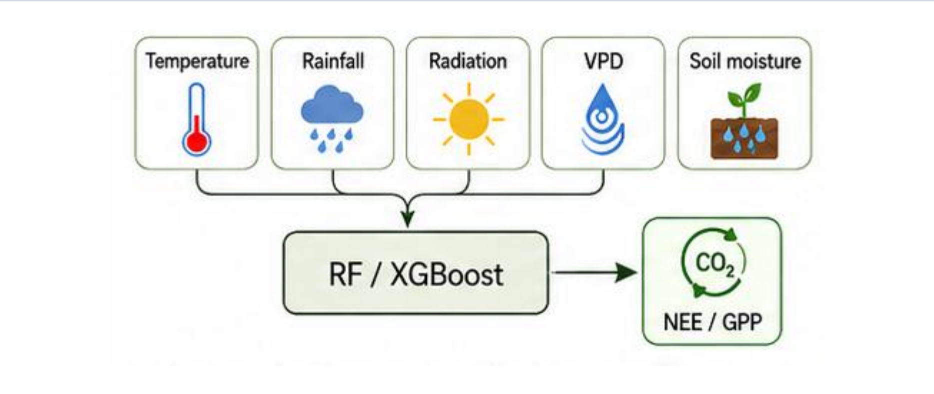


A crucial step towards global prediction is robust site-level generalisation: learning from observed tower sites and predicting carbon flux at unseen tower sites.

Existing Carbon Flux Modelling Methods

Traditional ML

RF / XGBoost



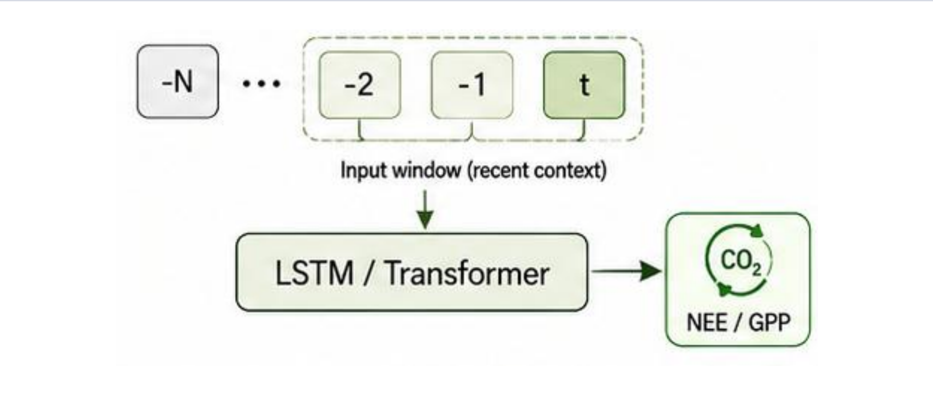
Maps current or hand-crafted summary features to carbon-flux predictions. Temporal order is not explicitly modelled.

Current	Short context	Long history

Limitation: temporal order is not explicitly modelled

Standard Time-Series Models

LSTM / Transformer



Models temporal dependencies within a predefined input window, capturing recent dynamics but not longer-term antecedent effects.

Current	Short context	Long history

Limitation: temporal dependencies are confined to the input window

Question: Are current and recent conditions sufficient to explain ecosystem carbon flux?

Ecological Memory Operates Across Multiple Timescales

Current carbon flux is influenced not only by recent conditions, but also by environmental conditions experienced days to months earlier.

Savanna

Lag: 1–4 weeks

Rainfall pulses cause relatively fast ecosystem responses.

Grassland / Cropland

Lag: 2–8 weeks

Recent weather and soil moisture often dominate the response.

Deciduous forest

Lag: 1–3 months

Leaf growth and phenology extend the memory into the season.

Evergreen / Mixed forest

Lag: 3–6 months

Drought legacy and canopy processes can produce longer-lasting memory.

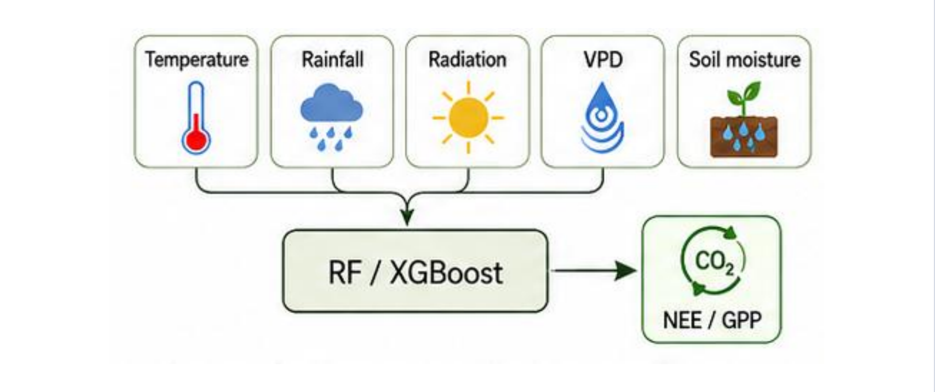
Therefore, a uniform short-term window may fail to capture **ecosystem-specific** and **long-horizon lag effects**.

From Fixed Window to Ecological Memory

Our framework extends recent context with explicit, multi-scale historical memory.

Traditional ML

RF / XGBoost



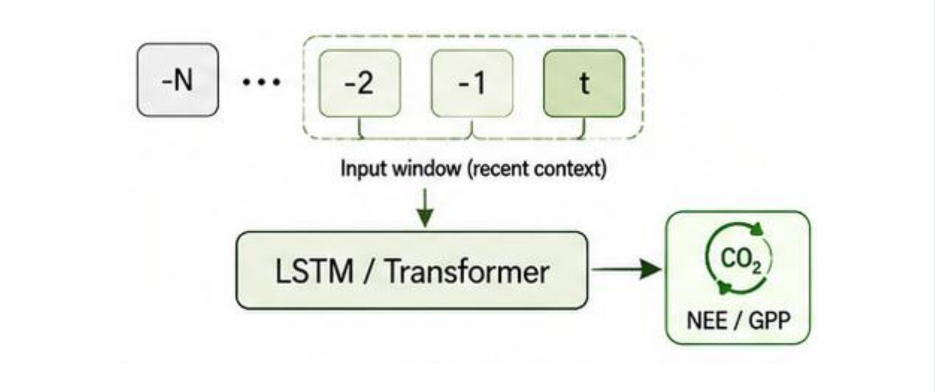
Maps current or hand-crafted summary features to carbon-flux predictions. Temporal order is not explicitly modelled.

Current | Short context | Long history

Limitation: temporal order is not explicitly modelled

Standard Time-Series Models

LSTM / Transformer



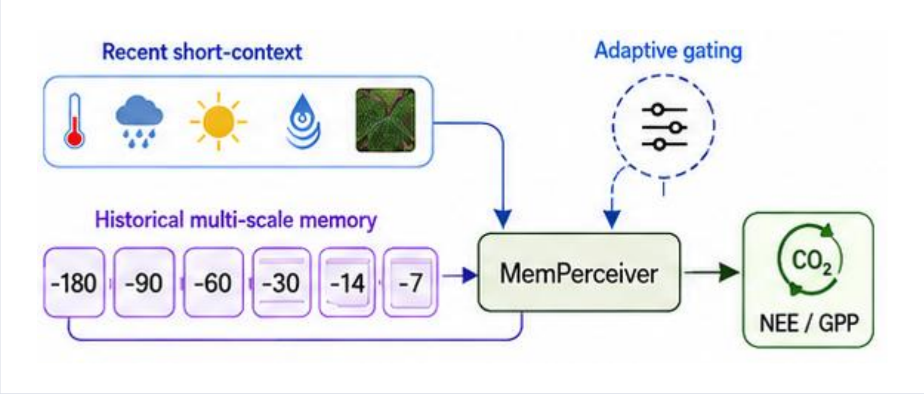
Models temporal dependencies within a predefined input window, capturing recent dynamics but not longer-term antecedent effects.

Current | Short context | Long history

Limitation: temporal dependencies are confined to the input window

Our Memory-Aware Framework

MemPerceiver: short context + adaptive long-horizon memory



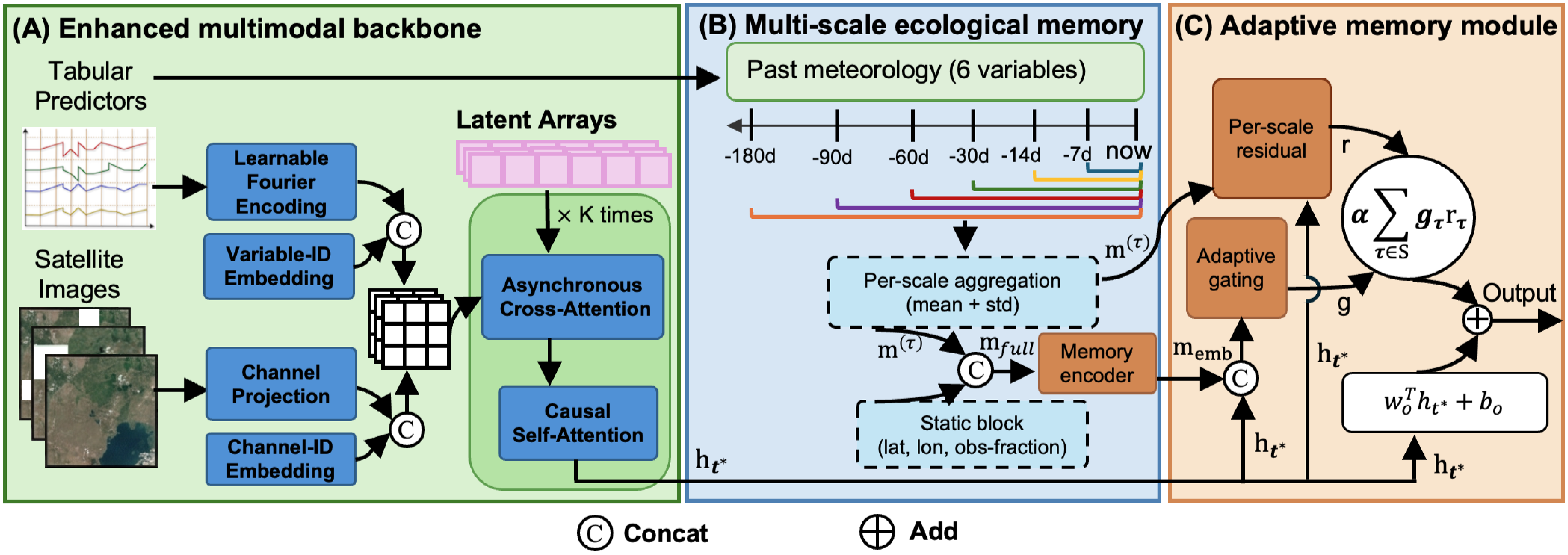
Integrates recent context with multi-scale historical conditions from −180 to −7 days and adaptively weights their contributions.

Current | Short context | Long history

Strength: captures biome-dependent lag effects

Key idea: extend short-term temporal modelling with explicit, adaptive, multi-scale ecological memory.

Our Proposed MemPerceiver Framework



(A) Enhanced multimodal backbone

Algorithm 1 Asynchronous cross-attention fusion in the multi-modal backbone

Input: Tabular tokens $\mathcal{Z}^{\text{tab}}_{b,t}$; image tokens $\mathcal{Z}^{\text{img}}_{b,t}$ defined when $a_{b,t} = 1$; latent matrix $\mathbf{E}^{\text{lat}} \in \mathbb{R}^{L \times d_h}$; group-specific LayerNorms LN^+, LN^-; ordered backbone layers $(\mathcal{A}_1, \ldots, \mathcal{A}_K)$.

Output: Anchor-hour latent $\mathbf{h}_{t^\star} \in \mathbb{R}^{d_h}$ at $t^\star = L$.

$\Omega \leftarrow \{(b,t) : a_{b,t} = 1\}, \quad \Omega^c \leftarrow \{(b,t) : a_{b,t} = 0\}$

for all $(b,t) \in \Omega$ **do** {satellite-available bundle}

  $\mathcal{T}_{b,t} \leftarrow \text{LN}^+([\mathcal{Z}^{\text{tab}}_{b,t}; \mathcal{Z}^{\text{img}}_{b,t}])$

end for

for all $(b,t) \in \Omega^c$ **do** {satellite-missing bundle}

  $\mathcal{T}_{b,t} \leftarrow \text{LN}^-(\mathcal{Z}^{\text{tab}}_{b,t})$

end for

$\mathbf{H} \leftarrow \text{repeat}_B(\mathbf{E}^{\text{lat}}) \in \mathbb{R}^{B \times L \times d_h}$

for $k = 1, \ldots, K$ **do**

  if $\mathcal{A}_k$ is a cross-attention layer **then**

    Flatten $\mathbf{H} \rightarrow \mathbf{H}^b \in \mathbb{R}^{(BL) \times 1 \times d_h}$

    $\mathbf{H}^b[\Omega] \leftarrow \mathcal{A}_k(\mathbf{H}^b[\Omega], \{\mathcal{T}_{b,t}\}_\Omega)$

    $\mathbf{H}^b[\Omega^c] \leftarrow \mathcal{A}_k(\mathbf{H}^b[\Omega^c], \{\mathcal{T}_{b,t}\}_{\Omega^c})$

    Scatter back and unflatten $\mathbf{H}^b \rightarrow \mathbf{H} \in \mathbb{R}^{B \times L \times d_h}$

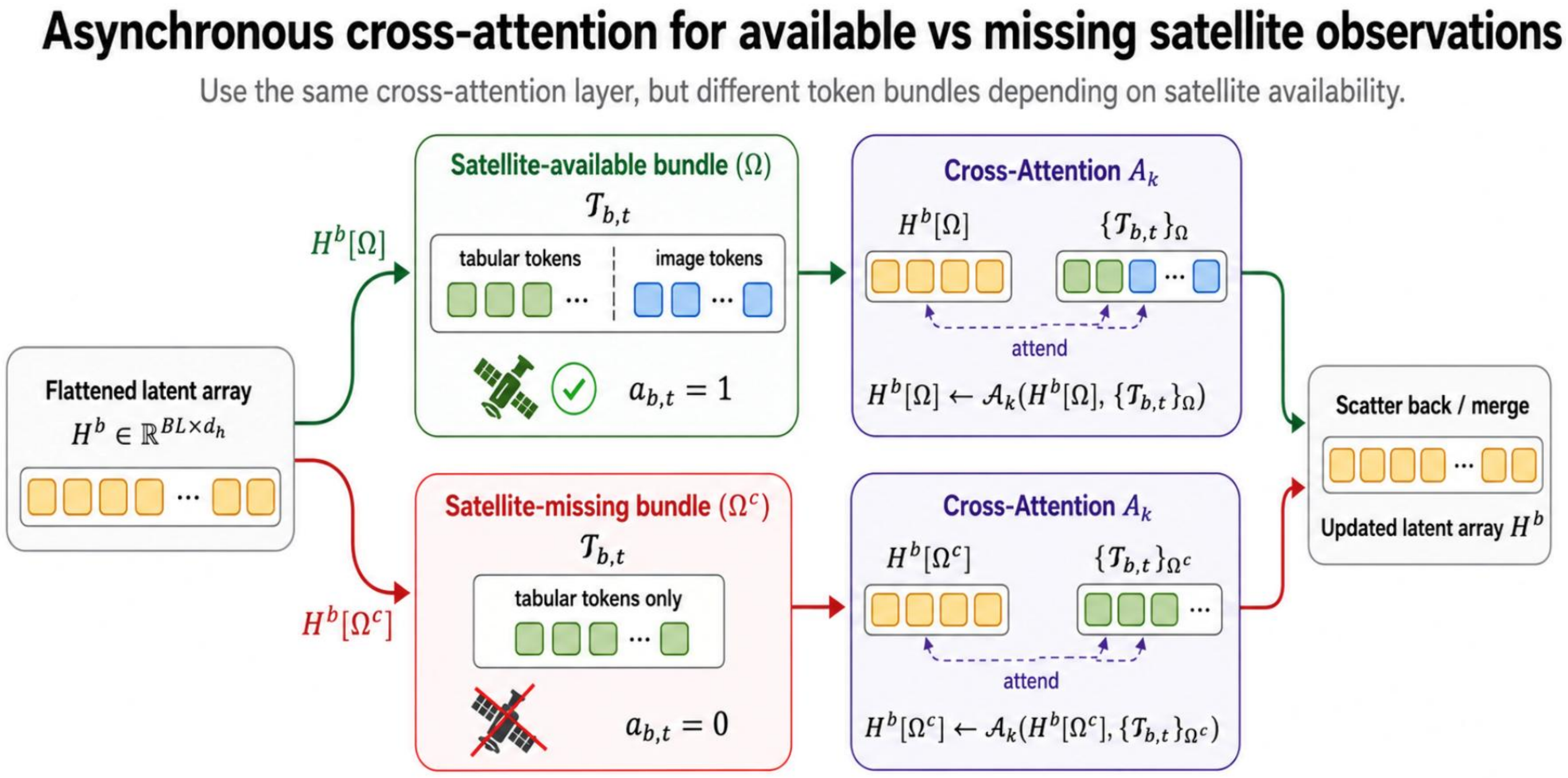
  else {causal self-attention along the temporal axis}

    $\mathbf{H} \leftarrow \mathcal{A}_k(\mathbf{H}, \mathbf{H}, \mathbf{M}^{\text{causal}}), \quad \mathbf{M}^{\text{causal}}_{y,x} = \mathbf{1}[y < x]$

  end if

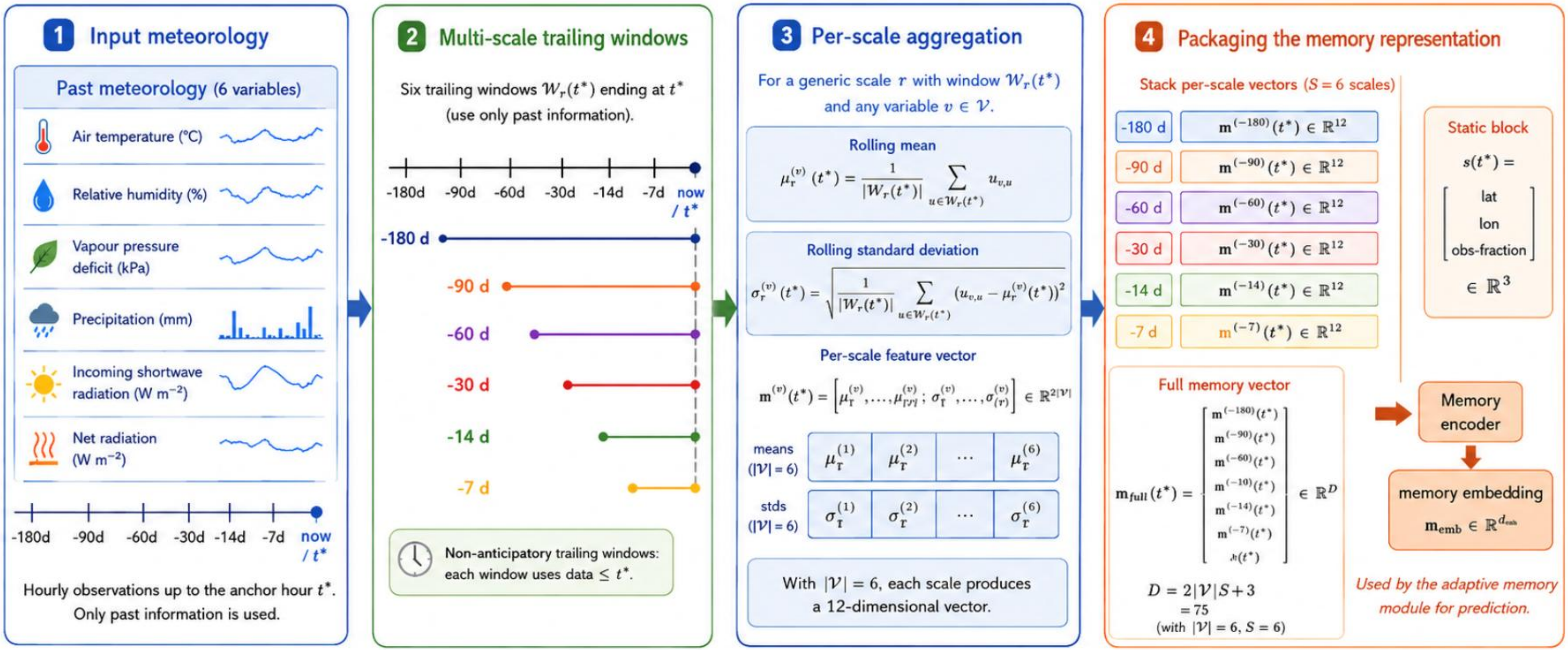
end for

return $\mathbf{h}_{t^\star} \leftarrow \mathbf{H}_{:,t^\star,:}$



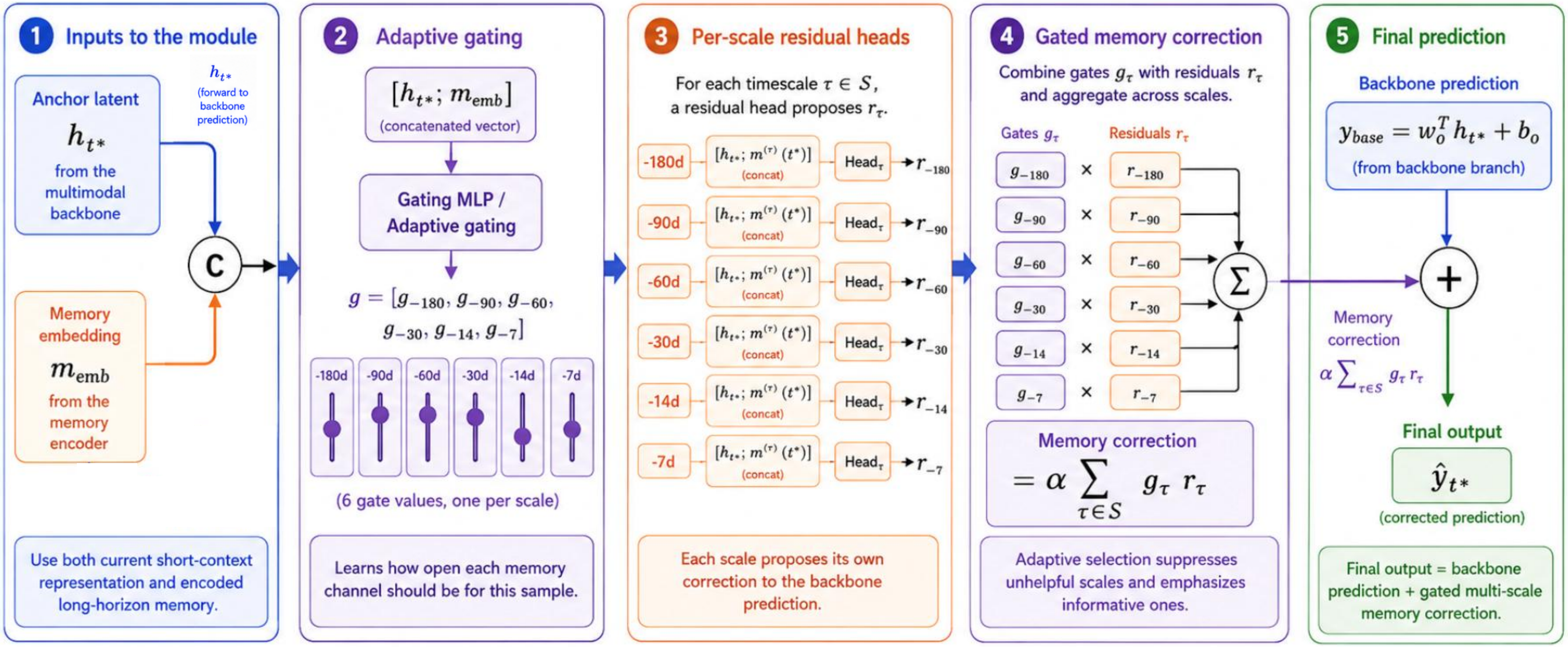
(B) Multi-scale Ecology Memory

For each prediction hour t^*, we summarise past meteorology over multiple trailing windows.



(C) Adaptive Memory Module

The module turns multi-scale memory into a correction to the backbone prediction.



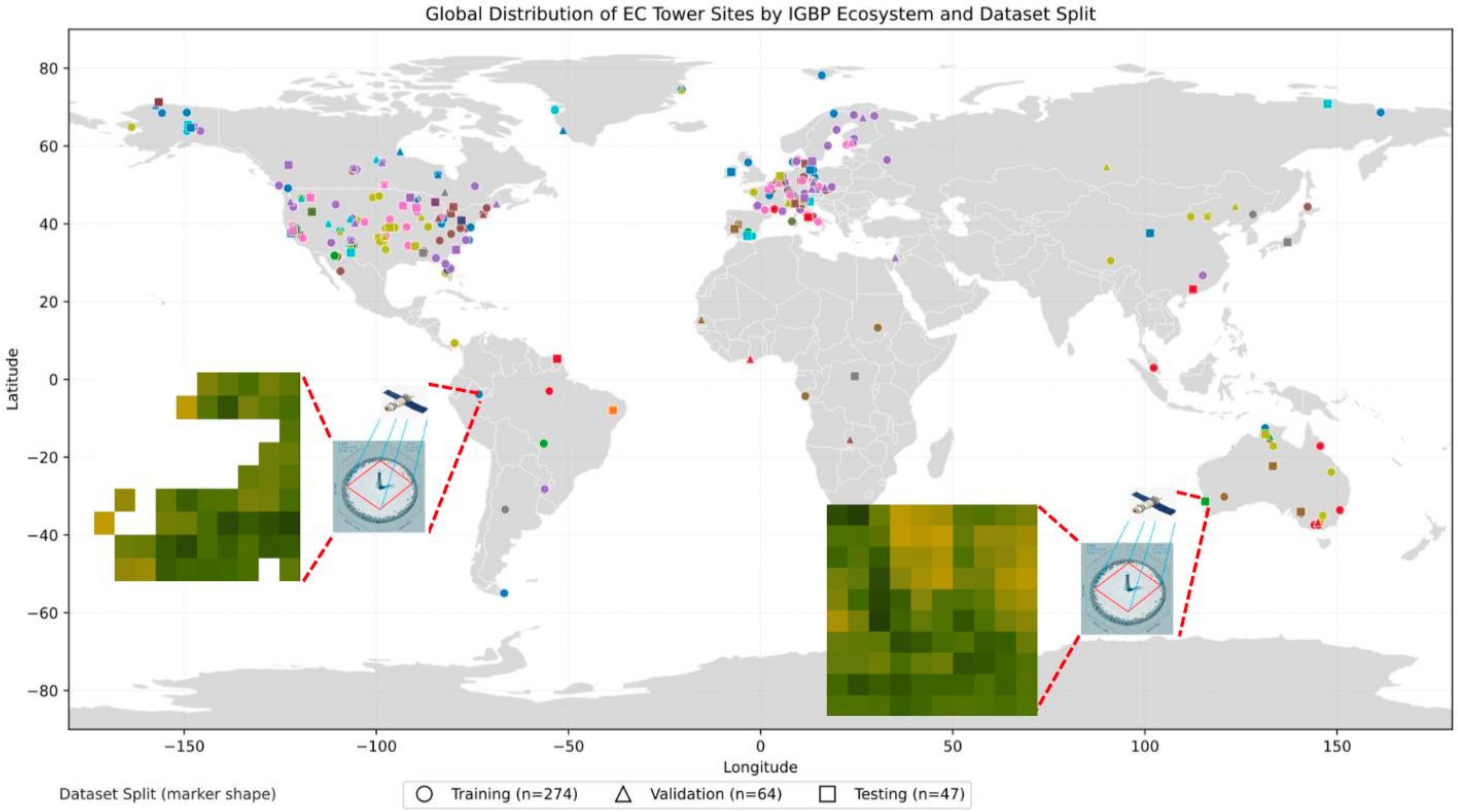
Experiments – Dataset Description

385
eddy-covariance sites worldwide

27.0M
hourly instances

15
IGBP ecosystem classes

21 + 9
tabular predictors + MODIS channels



IGBP Summary

IGBP	Train	Vali	Test	Shot Setting
OSH	20	5	5	-
CSH	4	1	2	-
DBF	35	8	5	-
DNF	0	0	1	zero-shot
EBF	8	2	3	-
GRA	48	11	5	-
MF	7	2	3	-
CVM	1	0	1	one-shot
SAV	9	2	3	-
WSA	7	1	2	-
SNO	0	0	1	zero-shot
WAT	1	0	1	one-shot
WET	34	8	5	-
CRO	36	8	5	-
ENF	64	16	5	-

Source: Fortier, M., Richter, M. L., Sonnentag, O., & Pal, C. (2025). CarbonSense: A multimodal dataset and baseline for carbon flux modelling. ICLR 2025, 41859–41891.

Experiments – Baselines and Evaluation Metrics

Baselines

FLUXCOM [1]

Non-deep global flux upscaling; ensemble of ML regressors, implemented with XGBoost on hourly temporal, meteorological and aggregated remote-sensing features.

NIES [2]

Parallel upscaling baseline built on Random Forest; same feature view as FLUXCOM, representing the alternative tree-ensemble paradigm.

MetaFlux [3]

Bi-LSTM trained under a model-agnostic meta-learning objective; addresses spatial sparsity via cross-site transfer.

FluxFormer [4]

Multivariate time-series Transformer for long-range temporal modelling of flux drivers.

EcoPerceiver [5]

Perceiver-style multimodal fusion; previous state of the art on CarbonSense and the backbone of MemPerceiver.

Evaluated on both NEE and GPP under identical site-level splits.

Evaluation Metrics

MAE

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

Mean Absolute Error

Robust, scale-consistent measure of average error magnitude.

RMSE

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Root Mean Squared Error

Penalises large deviations, hence sensitive to extreme flux events.

NSE

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Nash–Sutcliffe Efficiency

Predictive skill relative to the climatological mean; standard in hydrology and ecology.

Higher NSE is better; lower RMSE / MAE is better.

[1] Jung, M., et al. (2019). The FLUXCOM ensemble of global land–atmosphere energy fluxes. Scientific Data, 6(1), 74.
[2] Zeng, J., et al. (2020). Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest. Scientific Data, 7(1), 313.
[3] Nathaniel, J., Liu, J., & Gentine, P. (2023). MetaFlux: Meta-learning global carbon fluxes from sparse spatiotemporal observations. Scientific Data, 10(1), 440.
[4] Phan, A., & Fukui, H. (2023). FluxFormer: Upscaled global carbon fluxes from eddy covariance data with multivariate timeseries Transformer. Eartharxiv.
[5] Fortier, M., Richter, M. L., Sonnentag, O., & Pal, C. (2025). CarbonSense: A multimodal dataset and baseline for carbon flux modelling. ICLR 2025, 41859–41891.

Experimental Results

Overall predictive accuracy on the CarbonSense test sites for NEE (top) and GPP (bottom). Best per metric per target is in bold; second-best is underlined.

Method	Venue	MAE ↓	RMSE ↓	NSE ↑
NEE				
FLUXCOM	Sci. Data 2019	2.077	3.327	0.744
NIES	Sci. Data 2020	2.106	3.369	0.738
MetaFlux	Sci. Data 2023	2.044	3.315	0.746
FluxFormer	EarthArXiv 2023	2.000	3.229	0.759
EcoPerceiver	ICLR 2025	<u>1.915</u>	<u>3.132</u>	<u>0.774</u>
MemPerceiver (Ours)	**Ours**	**1.857**	**3.075**	**0.782**
GPP				
FLUXCOM	Sci. Data 2019	2.109	3.821	0.729
NIES	Sci. Data 2020	2.183	3.949	0.711
MetaFlux	Sci. Data 2023	2.122	3.895	0.719
FluxFormer	EarthArXiv 2023	2.080	3.823	0.729
EcoPerceiver	ICLR 2025	<u>1.949</u>	<u>3.604</u>	<u>0.759</u>
MemPerceiver (Ours)	**Ours**	**1.930**	**3.542**	**0.768**

Leave-one-out ablation of MemPerceiver components on the CarbonSense test sites (NEE only). Each variant disables one component relative to the full model. Best per metric in bold.

Variant	MAE ↓	RMSE ↓	NSE ↑
w/o learnable Fourier	1.897	3.100	0.778
w/o memory	1.936	3.177	0.767
Memory w/o gate	1.938	3.155	0.770
MemPerceiver (full)	**1.857**	**3.075**	**0.782**

The adaptive gate is essential for stabilising historical memory; without it, irrelevant past information may be introduced indiscriminately, degrading performance.

Experimental Results

Per-IGBP test-site performance on NEE. MAE and RMSE (μmol CO_2 m^{-2} s^{-1}, lower is better) and NSE (higher is better). Best per metric per biome is in bold; second-best is underlined. Rows are grouped according to the data regime in the Setting column, where *Regular*, *One-shot*, and *Zero-shot* denote multiple, single, and no training sites, respectively. IGBP classes are ordered alphabetically within each group. Method abbreviations: FLUX = FLUXCOM, Meta = MetaFlux, FF = FluxFormer, EP = EcoPerceiver, MP = MemPerceiver (NIES retains its full name).

		MAE ↓						RMSE ↓						NSE ↑					
IGBP	Setting	FLUX	NIES	Meta	FF	EP	MP (Our)	FLUX	NIES	Meta	FF	EP	MP (Our)	FLUX	NIES	Meta	FF	EP	MP (Our)
CRO	Regular	1.870	2.048	2.029	1.789	**1.586**	1.668	3.188	3.486	3.619	3.229	**2.817**	3.084	0.813	0.776	0.758	0.808	**0.854**	0.825
CSH	Regular	1.020	1.021	0.923	0.978	1.005	**0.905**	1.519	1.519	1.403	1.547	1.515	**1.323**	0.752	0.752	0.789	0.743	0.753	**0.812**
DBF	Regular	2.508	2.594	2.569	2.487	2.367	**2.335**	4.103	4.238	4.132	3.953	3.875	**3.832**	0.724	0.706	0.720	0.744	0.754	**0.759**
EBF	Regular	3.470	3.431	3.359	3.130	3.199	**2.983**	4.669	4.586	4.432	4.287	4.318	**4.191**	0.791	0.798	0.812	0.824	0.821	**0.832**
ENF	Regular	1.761	**1.735**	1.891	1.873	1.785	**1.735**	2.831	2.779	3.105	2.894	2.866	**2.710**	0.774	0.782	0.728	0.764	0.768	**0.793**
GRA	Regular	2.115	2.133	1.857	1.843	1.822	**1.692**	3.254	3.230	2.942	2.980	2.894	**2.742**	0.745	0.749	0.792	0.786	0.798	**0.819**
MF	Regular	2.675	2.654	2.701	2.937	**2.571**	2.614	3.844	3.880	3.859	4.186	**3.762**	3.932	0.758	0.754	0.756	0.713	**0.769**	0.747
OSH	Regular	1.279	1.300	**1.102**	1.119	1.138	1.144	1.902	1.958	**1.688**	1.728	1.993	1.949	0.534	0.507	**0.633**	0.615	0.489	0.511
SAV	Regular	1.135	1.169	0.892	0.907	0.868	**0.806**	1.649	1.673	1.325	1.382	1.317	**1.211**	0.582	0.569	0.730	0.706	0.733	**0.774**
WET	Regular	1.463	1.794	1.797	1.674	**1.287**	1.432	2.163	2.732	2.952	2.661	**2.036**	2.450	0.475	0.162	0.022	0.206	**0.535**	0.327
WSA	Regular	1.674	1.604	1.590	1.694	1.581	**1.556**	2.511	2.455	2.446	2.609	2.455	**2.402**	0.615	0.632	0.634	0.584	0.632	**0.647**
CVM	One-shot	3.249	3.125	3.141	3.178	**3.063**	3.068	5.531	5.478	5.506	5.541	5.311	**5.237**	0.525	0.534	0.529	0.523	0.562	**0.574**
WAT	One-shot	2.480	3.256	3.414	5.615	3.068	**2.424**	3.031	4.051	4.395	7.115	3.837	**2.953**	−9.896	−18.457	−21.899	−59.024	−16.458	**−9.342** *
DNF	Zero-shot	2.829	2.718	2.266	**2.229**	2.311	2.237	4.099	3.777	3.535	**3.506**	3.682	3.550	0.280	0.389	0.464	**0.473**	0.419	0.460
SNO	Zero-shot	1.047	0.939	1.556	1.140	0.840	**0.767**	1.454	1.343	2.210	1.512	1.215	**1.166**	−0.073	0.084	−1.480	−0.160	0.250	**0.310**

* NSE values on WAT are strongly negative for two reasons. First, the single WAT test site has relatively low observed NEE variability compared with the prediction errors, which magnifies the NSE penalty. Second, NEE at this site is challenging for all evaluated methods to predict, as each underperforms the observed-mean baseline (NSE = 0).

Top-2 count 14/15

Top-2 count 11/15

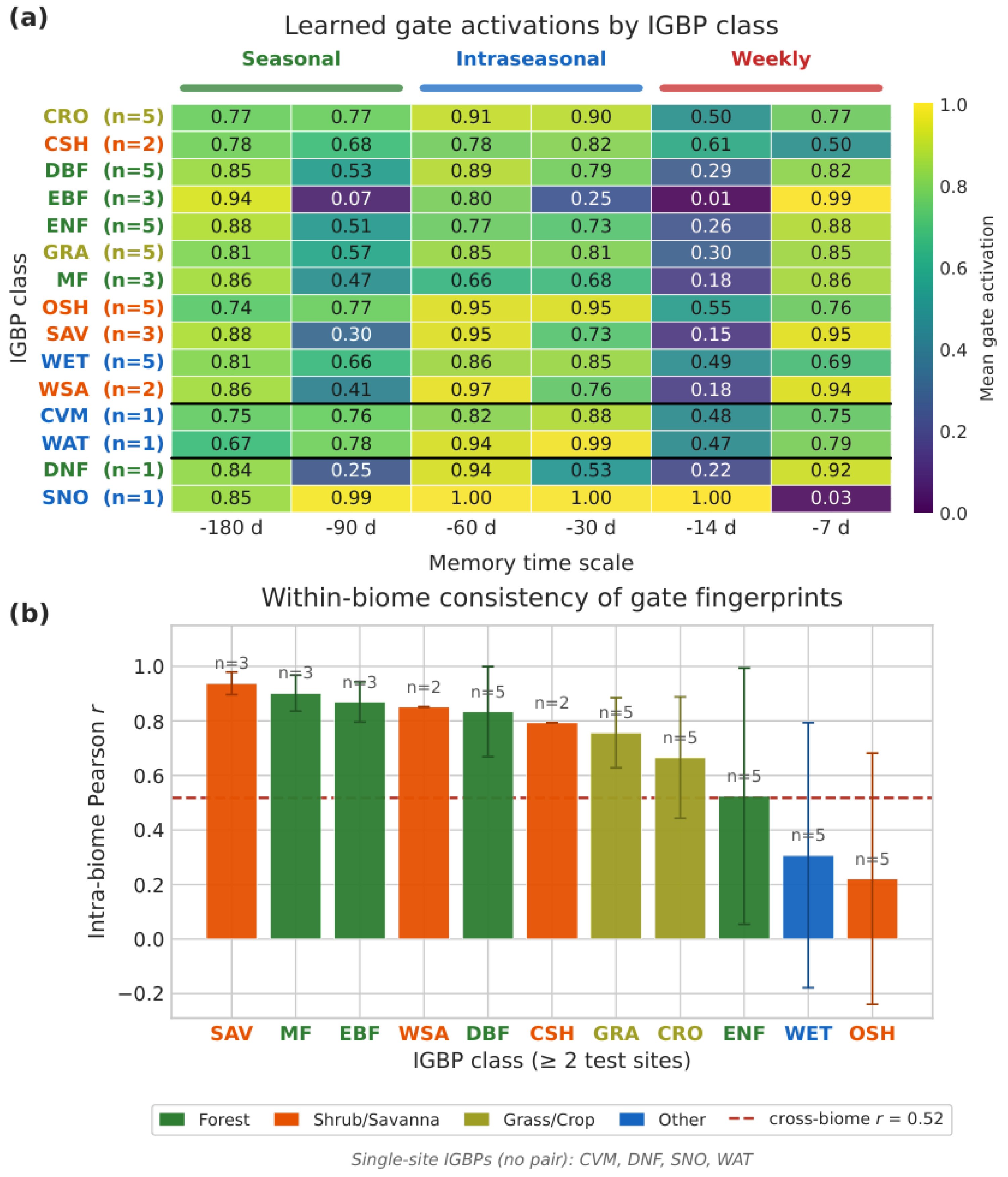
Top-2 count 11/15

INTERPRETABILITY

Temporal Fingerprints

Mean gate activation per biome — a model-derived view of ecological memory. The −90 d seasonal channel provides the strongest differentiation between biomes.

(a) Site-balanced mean gate activation by IGBP class. (b) Within-biome consistency of site-level fingerprints.



Acknowledgements

This research was supported by **Innovate UK** through the **Farming Innovation Pathways Project** (10081615: AI-enabled Climate-Smart Fertiliser Practice for Sustainable and Resilient Farming).

We also thank **Professor Paul Harris** and colleagues at **Rothamsted Research** for their valuable discussions and support.

Thanks