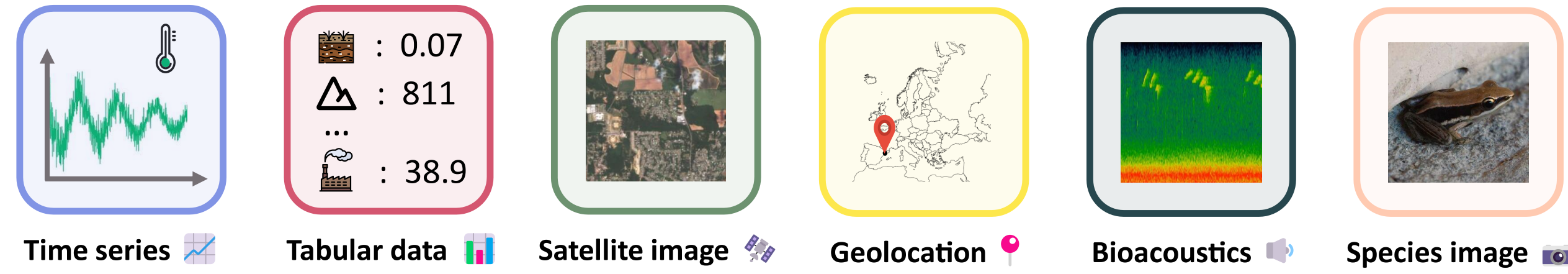


Ecological data are multimodal but often incomplete

Heterogeneous sources of data in ecological applications:



Challenge: Inconsistently available across space and time, at multiple levels:

- **Modality level** — e.g., unavailable satellite image due to cloud cover
- **Within modalities** — e.g., missing records in a climate time series

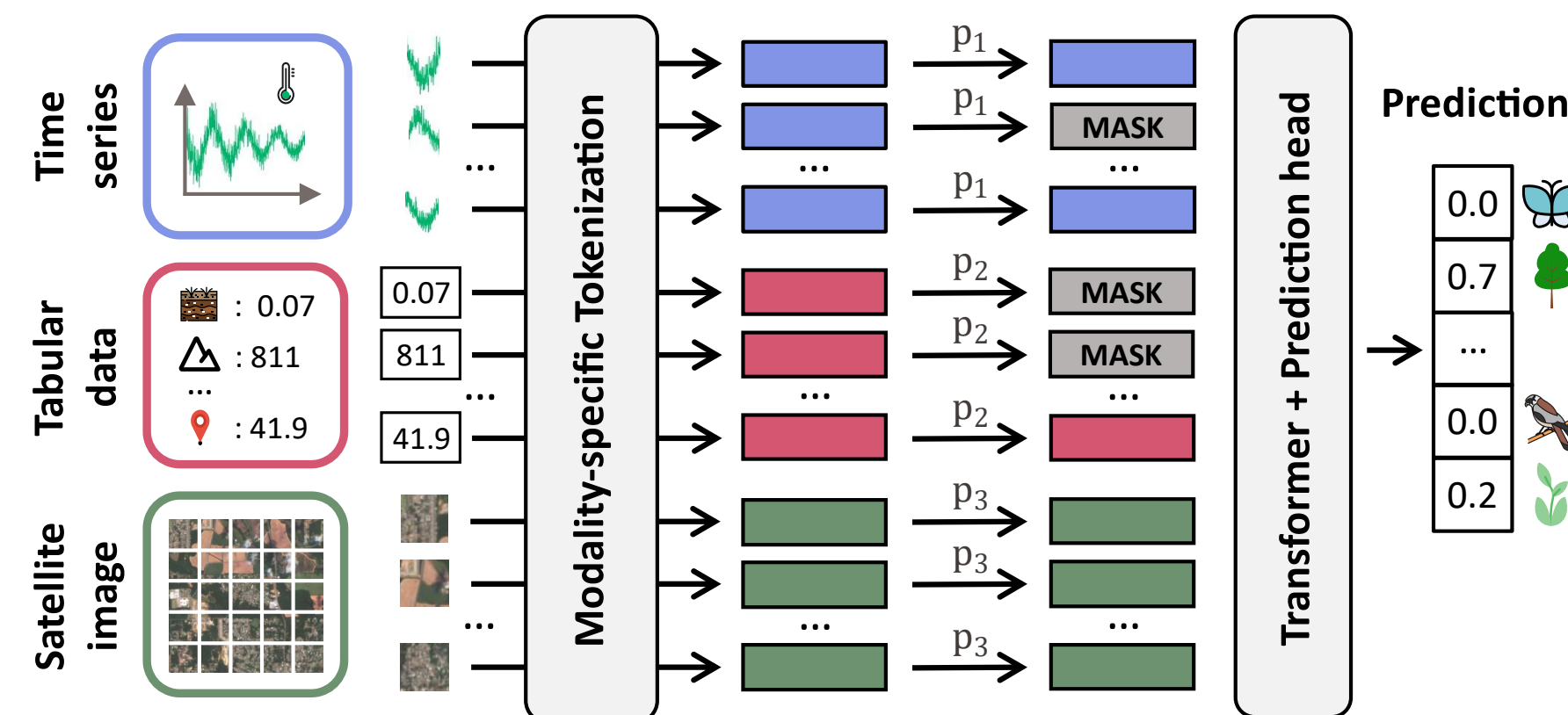
Objective: Build models that work with any available data

- Robust to missing modalities and partial observations within modalities
- Enable **interpretable input contribution** across and within modalities

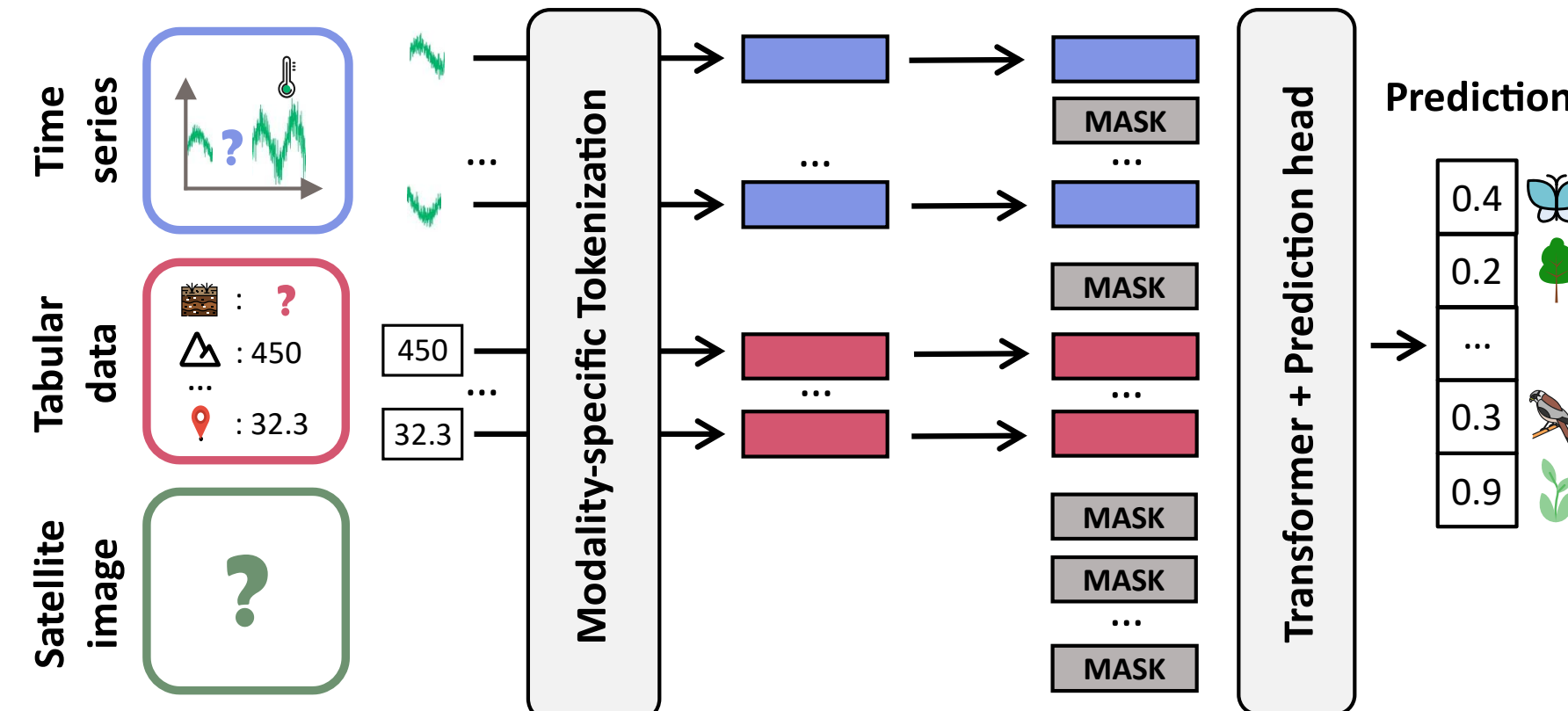
Masked modeling for robustness to missing inputs

Setup: multimodal transformer where each modality is tokenized into tokens

During **training**, tokens are **randomly masked** at each iteration to expose the model to diverse input subsets



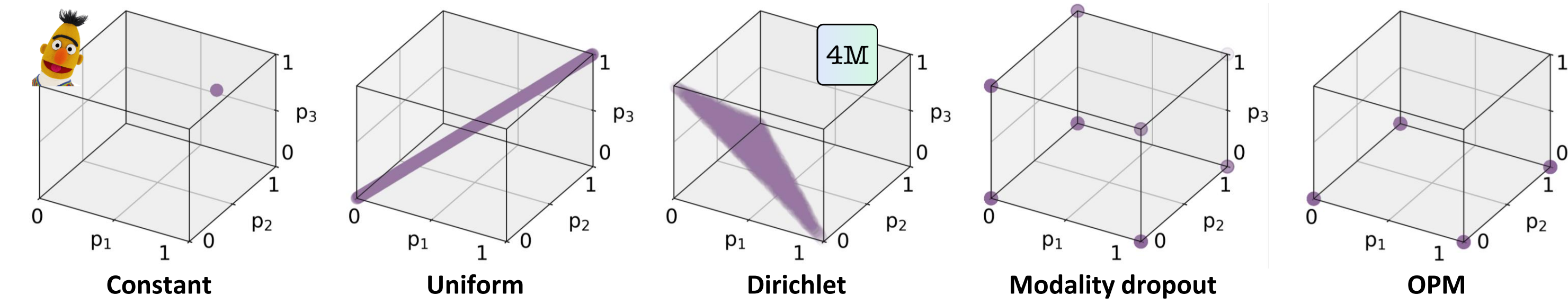
At **inference**, the model can handle **any subset of tokens**



How to effectively mask?

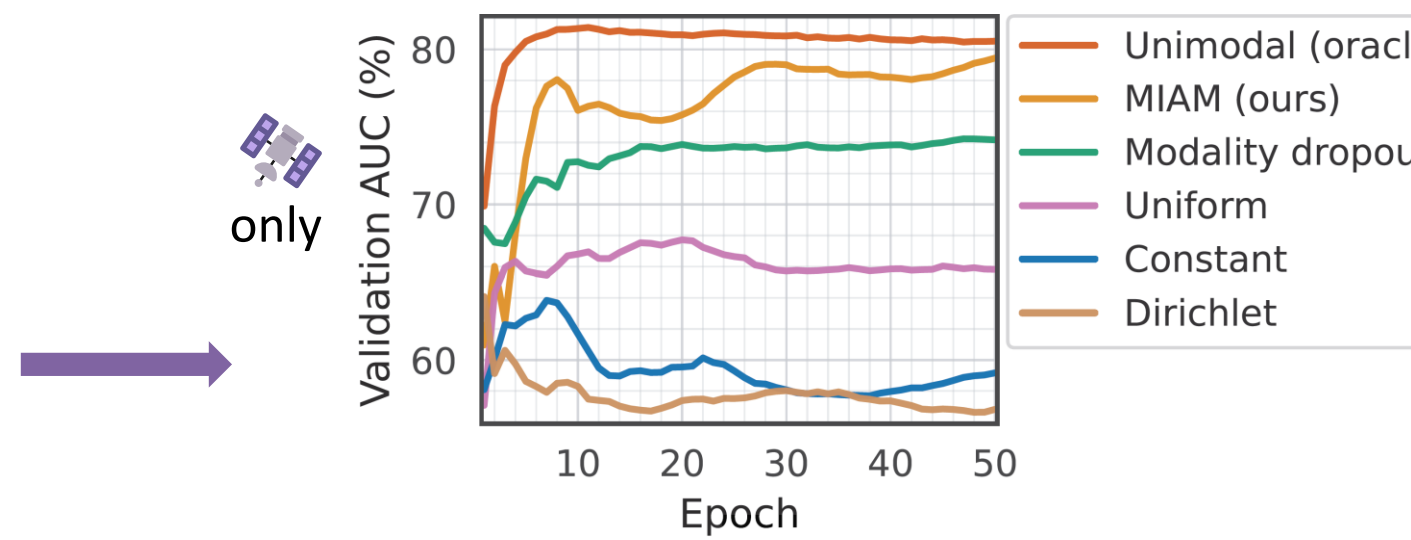
We **formalize masking strategies** as probability distributions $\mathbf{p} = (p_1, \dots, p_M)$ over the unit hypercube with M modalities

- Each token of modality m is masked with probability p_m
- Existing masking strategies can be expressed within this framework (here $M = 3$):



Limitations with existing strategies:

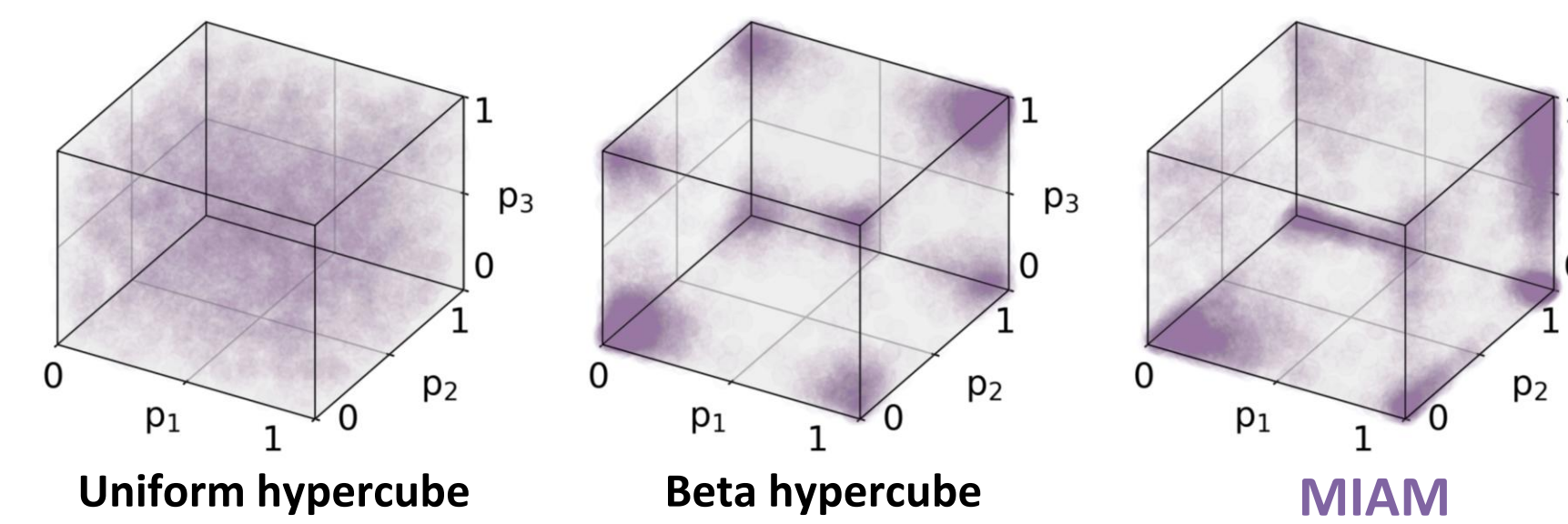
- Poorly explore the space of input subsets
- Fixed and uniform
- **Modality imbalance**, where dominant modalities hinder effective optimization



MIAM: Modality Imbalance-Aware Masking

Design principles of MIAM:

- Full support
- Corner prioritization
- Imbalance-awareness



We base MIAM on a **mixture of product beta distributions**:

$$\text{MixProdBeta}(\mathbf{p}) = \sum_{c \in \mathcal{C}} w_c \cdot f_c(\mathbf{p})$$

$$f_c(\mathbf{p}) = \prod_{m=1}^M \begin{cases} \text{Beta}(p_m; 1, \kappa \cdot (\frac{p_{sm}}{\rho_{dm}})^{-\lambda}) & \text{if } c_m = 0 \\ \text{Beta}(p_m; \kappa \cdot (\frac{p_{sm}}{\rho_{dm}})^{\lambda}, 1) & \text{if } c_m = 1 \end{cases}$$

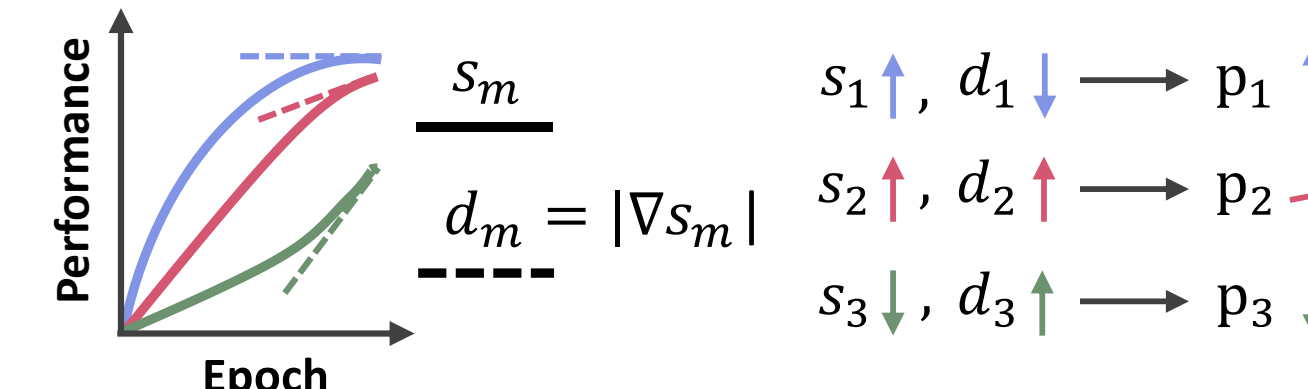
$$w_c = \begin{cases} \frac{1}{4} & \text{if } c \in \{(0, \dots, 0), (1, \dots, 1)\} \\ \frac{1}{2(2^M - 2)} & \text{otherwise.} \end{cases}$$

$$\rho_{sm} = \frac{s_m}{(\prod_{m'=1}^M s_{m'})^{1/M}}$$

$$\rho_{dm} = \frac{d_m}{(\prod_{m'=1}^M d_{m'})^{1/M}}$$

Dynamic score-driven masking:

- **Underperforming modalities are masked less** (unimodal performance s_m)
- **Improving modalities are masked less** (derivative of unimodal performance d_m)



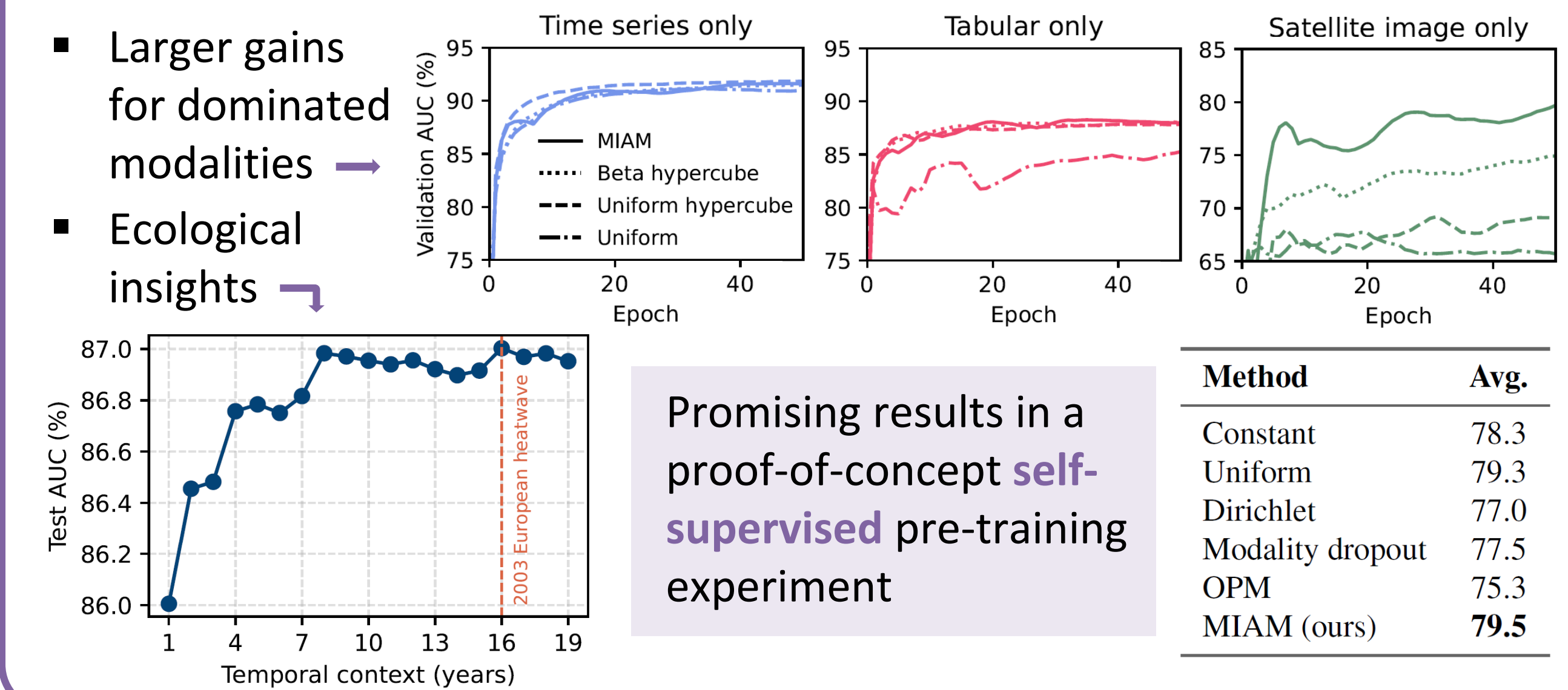
Experimental results

We evaluate MIAM in a **supervised** setting on two multimodal ecological datasets, **GeoPlant** (AUC shown here) and **TaxaBench**

- **MIAM outperforms other strategies** on both datasets

Modality	Input Type	Input Type												
		Partial Unimodal			Unimodal			Bimodal			All			
Tabular	BIO1	✓	✓				✓			✓	✓	✓		
	WorldClim		✓				✓			✓	✓	✓		
	Others						✓			✓	✓	✓		
Time series	Clim: 2018			✓	✓			✓			✓	✓	✓	
	Clim: 2000-2018				✓			✓			✓	✓	✓	
Sat. image	Center patch				✓			✓			✓	✓	✓	
	Others							✓			✓	✓	✓	
Constant		68.6	82.4	84.7	86.7	55.1	83.3	90.0	63.6	90.0	83.3	89.2	87.9	80.4
Uniform		73.3	85.7	86.3	87.2	61.2	86.9	91.1	65.6	91.6	86.2	91.8	92.0	83.2
Dirichlet		65.1	82.7	77.8	86.8	54.9	87.5	91.1	58.2	91.8	88.6	91.7	91.4	80.6
Modality dropout		48.7	80.8	77.4	86.4	66.2	88.6	91.4	73.2	92.0	89.2	91.7	92.0	81.5
OPM		68.0	81.9	80.7	85.3	68.1	88.4	90.2	81.1	90.7	89.5	91.1	91.2	83.8
MIAM (ours)		78.4	86.7	86.0	87.0	70.8	89.0	91.4	80.1	91.7	89.5	91.5	91.7	86.1
Oracle (one model per column)		78.0	87.1	87.7	87.6	77.1	89.3	92.2	81.4	92.3	89.7	91.7	92.0	87.2

- Larger gains for dominated modalities
- Ecological insights



Promising results in a **proof-of-concept self-supervised pre-training experiment**

Method	Avg.
Constant	78.3
Uniform	79.3
Dirichlet	77.0
Modality dropout	77.5
OPM	75.3
MIAM (ours)	79.5

Take-aways

- **Masking strategy significantly impacts performance**
- Standard masking fails under **modality imbalance**
- We propose Modality Imbalance-Aware Masking (**MIAM**)
- Rethinking masking is key for **any-to-any models**

