



*Supplement of*

## **Machine-learning-based estimates of global natural vegetated wetland methane emissions (2000–2025)**

**Mengze Li et al.**

*Correspondence to:* Mengze Li ([mengze@nus.edu.sg](mailto:mengze@nus.edu.sg))

The copyright of individual parts of the supplement might differ from the article licence.

## 5 Supplement

Table S1. Bottom-up (BU) models used in this study.

<b>BU model</b>	<b>Wetland extent mode</b>	<b>Climate forcings used (each as an individual estimate)</b>
CLASSIC	Prognostic	CRU; GSWP3-W5E5
ELM-ECA	Prognostic	CRU; GSWP3-W5E5
ISAM	Prognostic	CRU; GSWP3-W5E5
JSBACH	Prognostic	CRU; GSWP3-W5E5
JULES	Prognostic	CRU; GSWP3-W5E5
LPJ-MPI	Prognostic	CRU; GSWP3-W5E5
LPJ-WSL	Prognostic	CRU; GSWP3-W5E5
LPX-Bern	Prognostic	CRU; GSWP3-W5E5
ORCHIDEE	Prognostic	CRU; GSWP3-W5E5
SDGVM	Prognostic	CRU; GSWP3-W5E5
VISIT	Prognostic	CRU; GSWP3-W5E5

Table S2. Top-down (TD) atmospheric inversions used in this study.

15

Run ID	Inversion system	Observation constraint	Variant label*
1	CIF-LMDz	SURF	inv1
2	CIF-LMDz	SURF	inv2
3	CTE	SURF	inv1
4	CTE	SURF	inv2
5	MIROC4-ACTM	SURF	inv1
6	MIROC4-ACTM	SURF	inv2
7	NIES	SURF	inv1
8	NIES	SURF	inv2
9	NISMON-CH4	SURF	inv1
10	NISMON-CH4	SURF	inv2
11	PYVAR-LMDz	GOSAT	inv1
12	PYVAR-LMDz	SURF	inv2
13	NIES	GOSAT	inv1

\*inv1 and inv2 are two inversion rounds differing in prior anthropogenic emission choices (EDGARD v6 vs GAINS for fossil fuel emissions) and in OH assumptions.

20

Table S3. XGBoost hyperparameter configurations.

<b>Combo ID</b>	<b>Learning rate</b>	<b>Max tree depth</b>	<b>Min child weight</b>	<b>Subsample ratio</b>	<b>Column subsample per tree</b>	<b>L2 regularization</b>
1	0.10	6	0.00	0.98	0.98	1.0
2	0.10	5	0.25	0.95	0.95	1.0
3	0.07	5	0.50	0.90	0.90	1.5
4	0.07	4	0.50	0.90	0.90	1.5
5	0.05	4	1.00	0.85	0.85	2.0
6	0.05	3	1.00	0.85	0.80	2.0

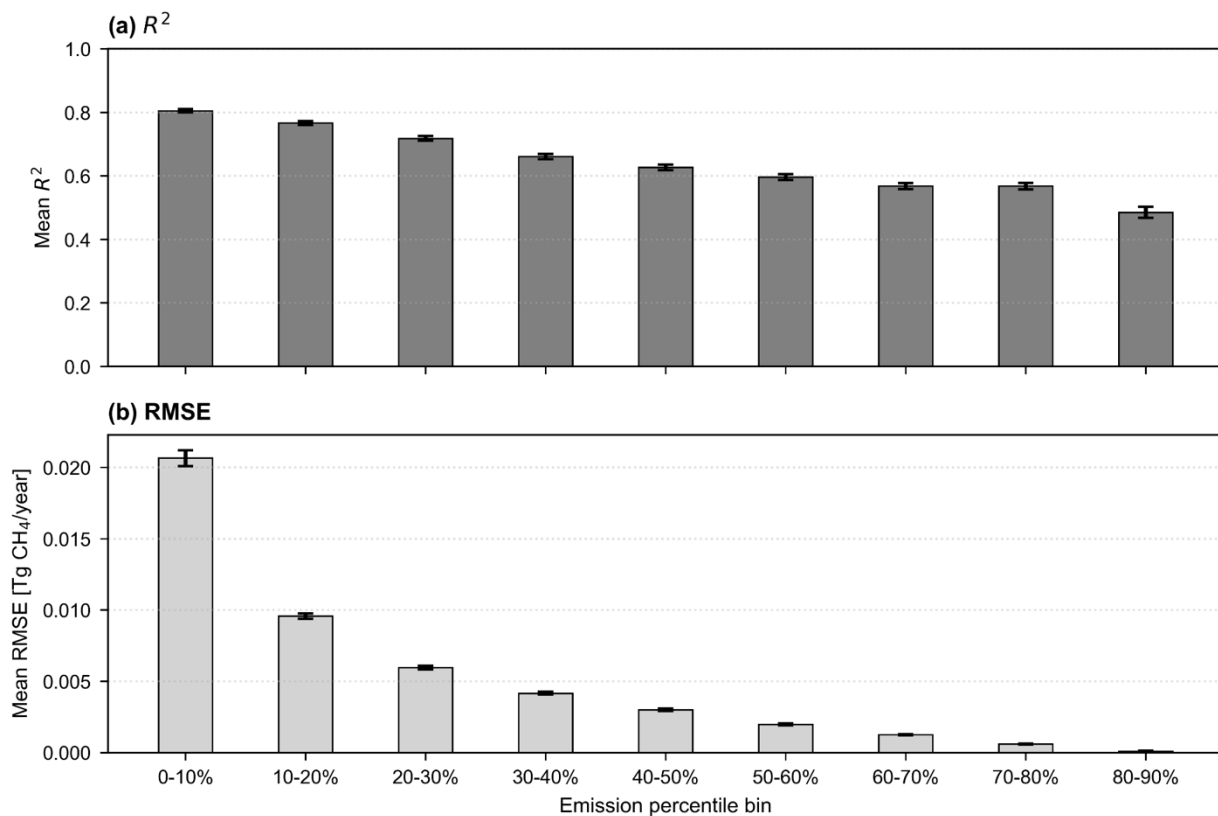
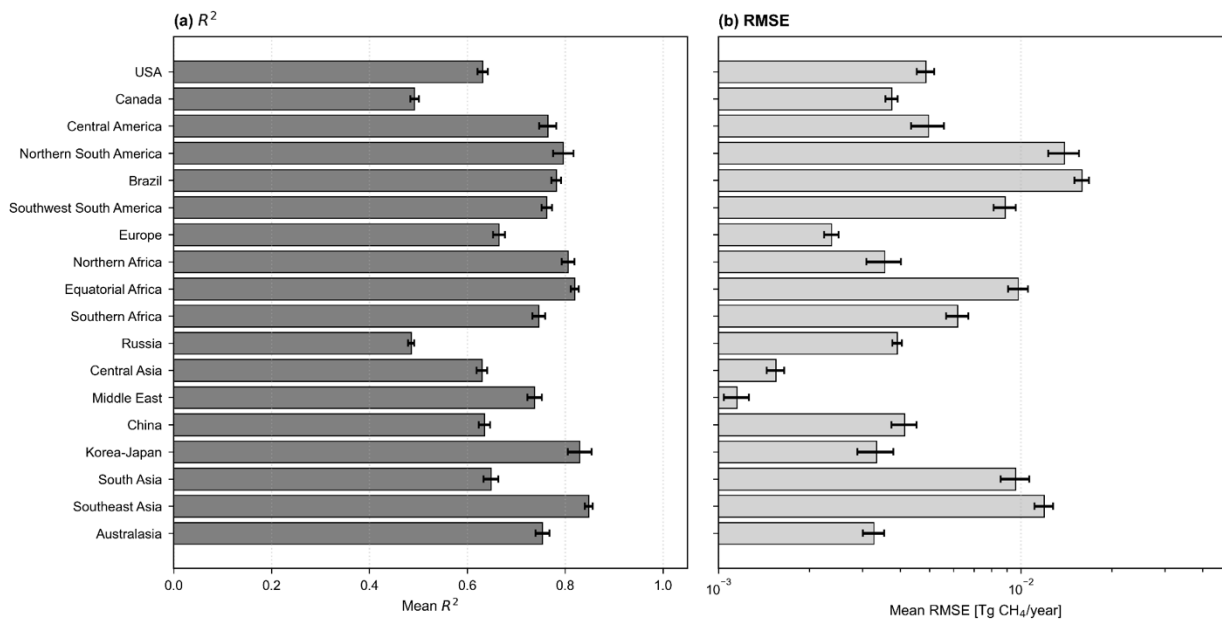
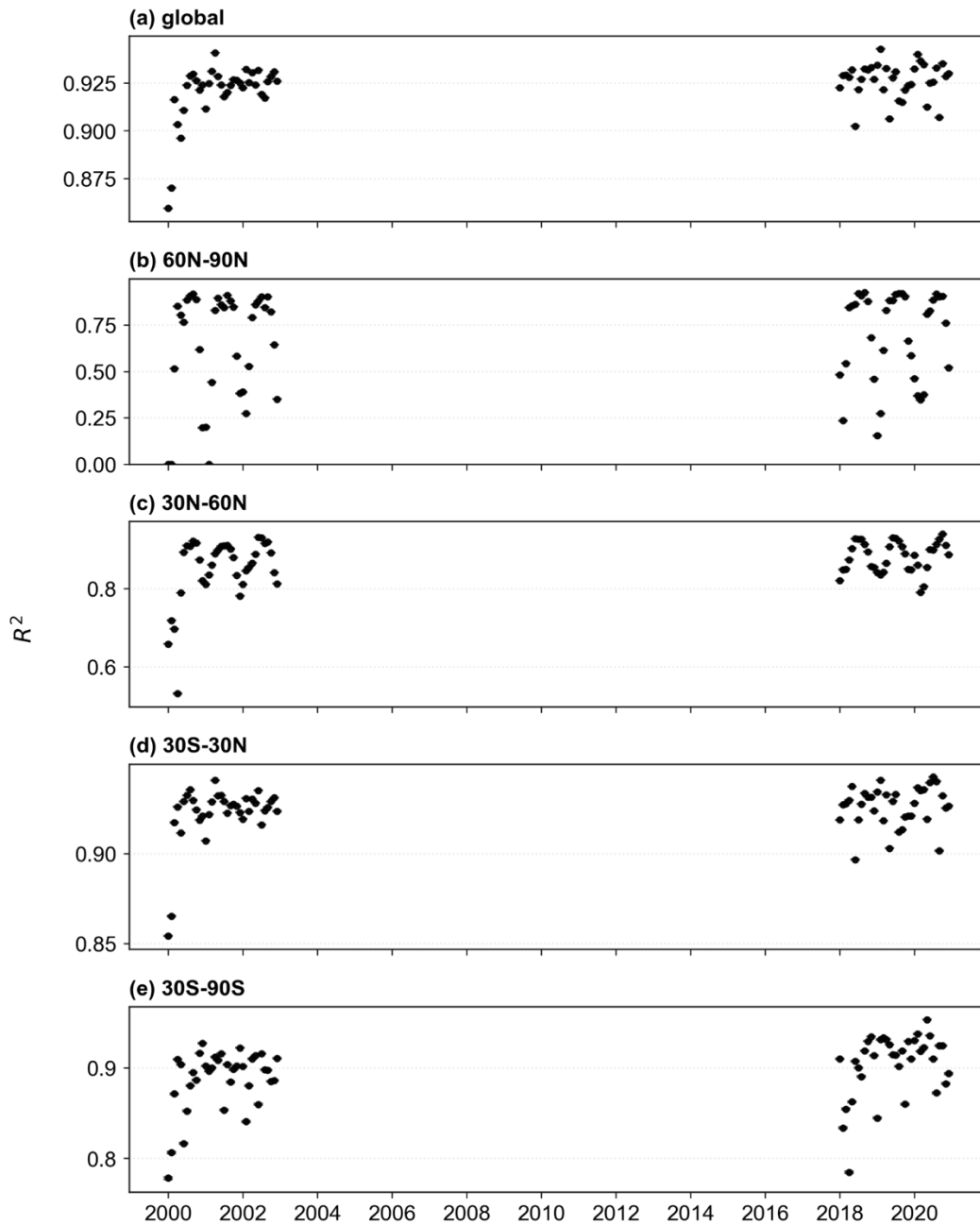


Figure S1. Reconstruction skills across regions by emission magnitude. Mean grid-cell  $R^2$  and RMSE on the test periods, summarized within emission percentile bins defined from grid-cell wetland  $\text{CH}_4$  emissions during 2000-2020 from GMB estimates. 0-10% denotes the highest emitting grid cells. Error bars indicate 95% CI.



35 Figure S2. Reconstruction skills across 18 geographical regions. Mean grid-cell  $R^2$  and RMSE on the test periods, summarized within emission percentile bins defined from grid-cell wetland  $\text{CH}_4$  emissions during 2000–2020 from GMB estimates. 0–10% denotes the highest emitting grid cells. Error bars indicate 95% CI.



40 Figure S3. Mean monthly  $R^2$  over test periods for five spatially aggregated latitude bands. Dots and error bars indicate the mean and 95% CI.

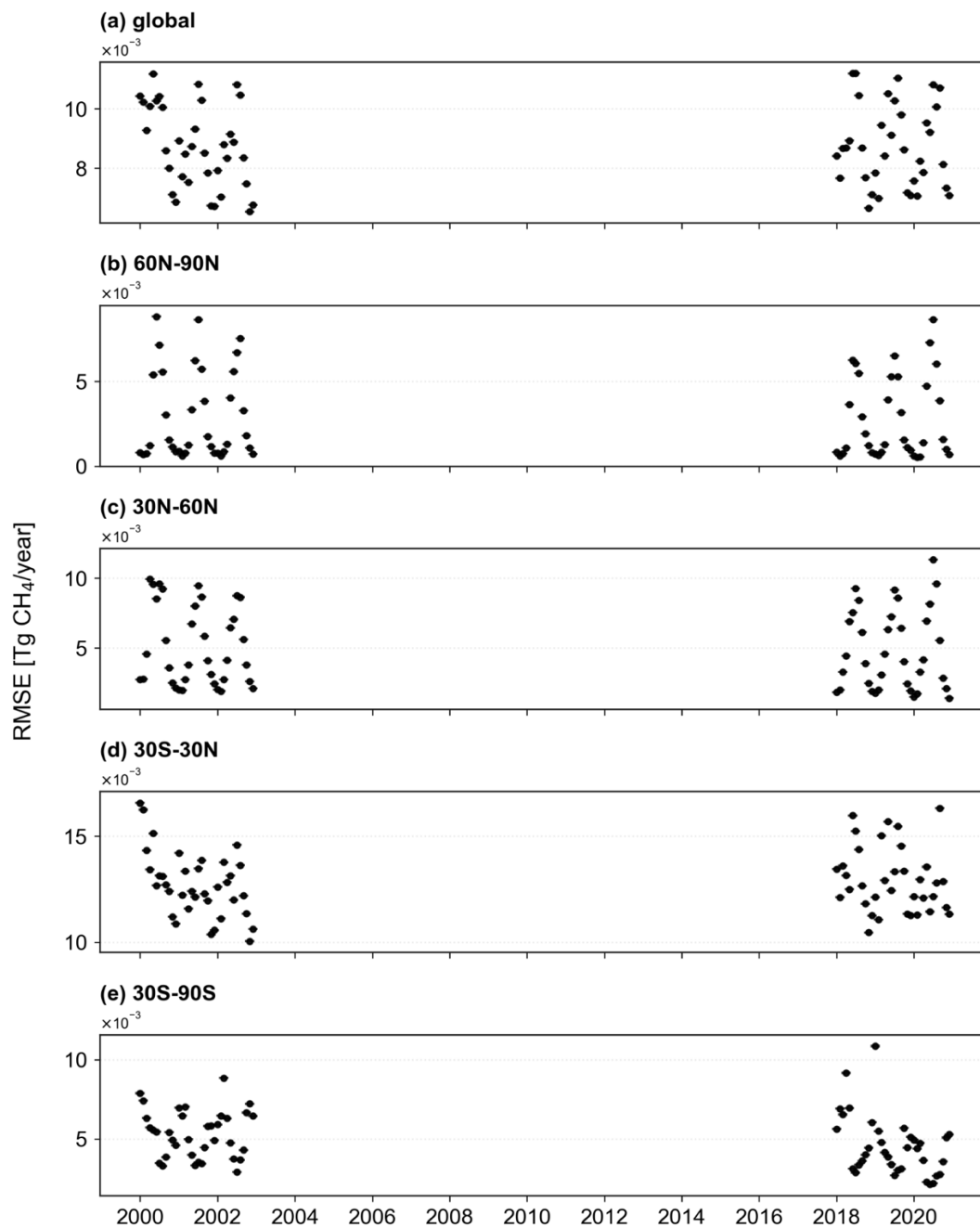


Figure S4. Mean monthly RMSE over test periods at five latitude bands. Dots and error bars indicate the mean and 95% CI.

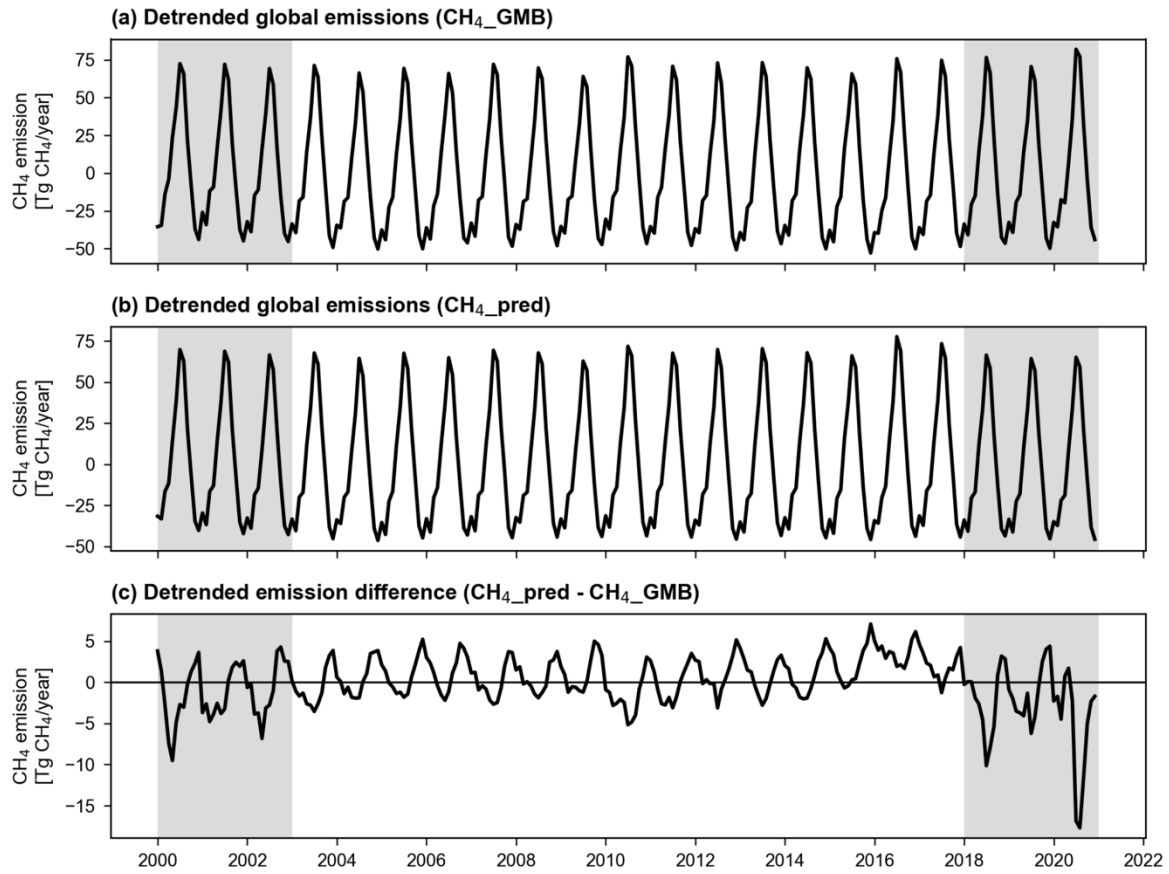


Figure S5. Detrended anomaly comparison of global wetland CH<sub>4</sub> emissions from GMB models (CH<sub>4</sub>\_GMB) and predictions (CH<sub>4</sub>\_pred). Detrended anomalies are calculated by removing the best-fit linear trend from each time series for the period 2000-2020. Shaded areas indicate test periods used for out-of-sample evaluation.

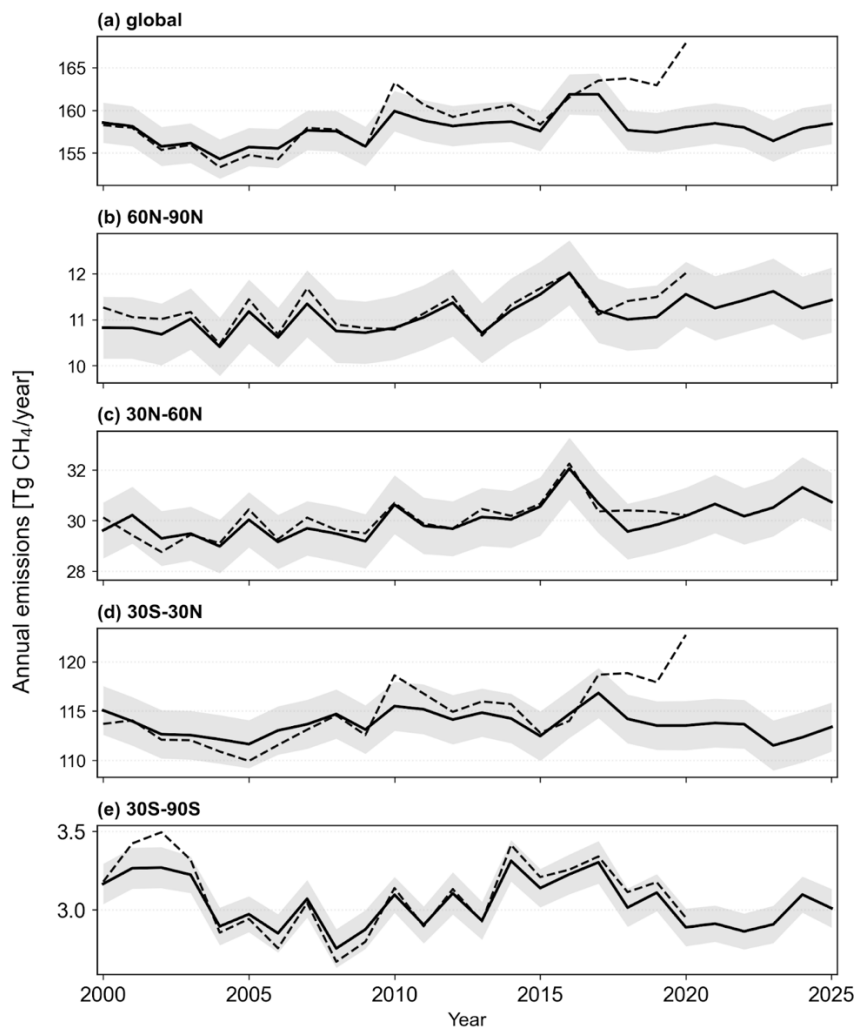
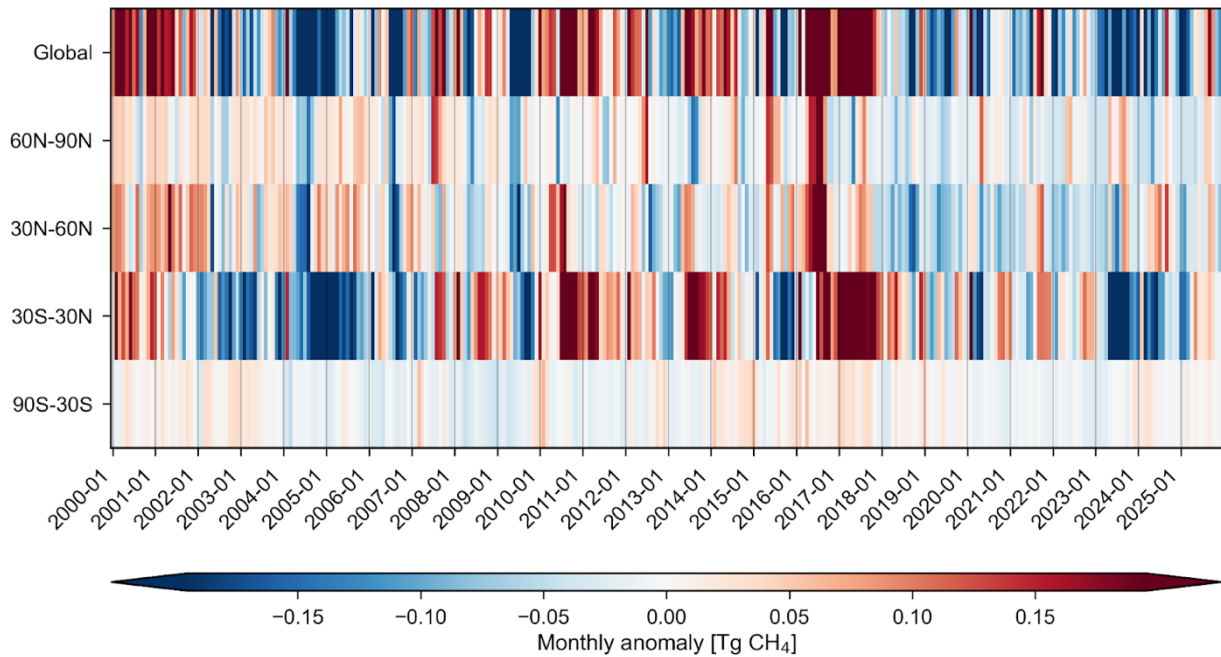
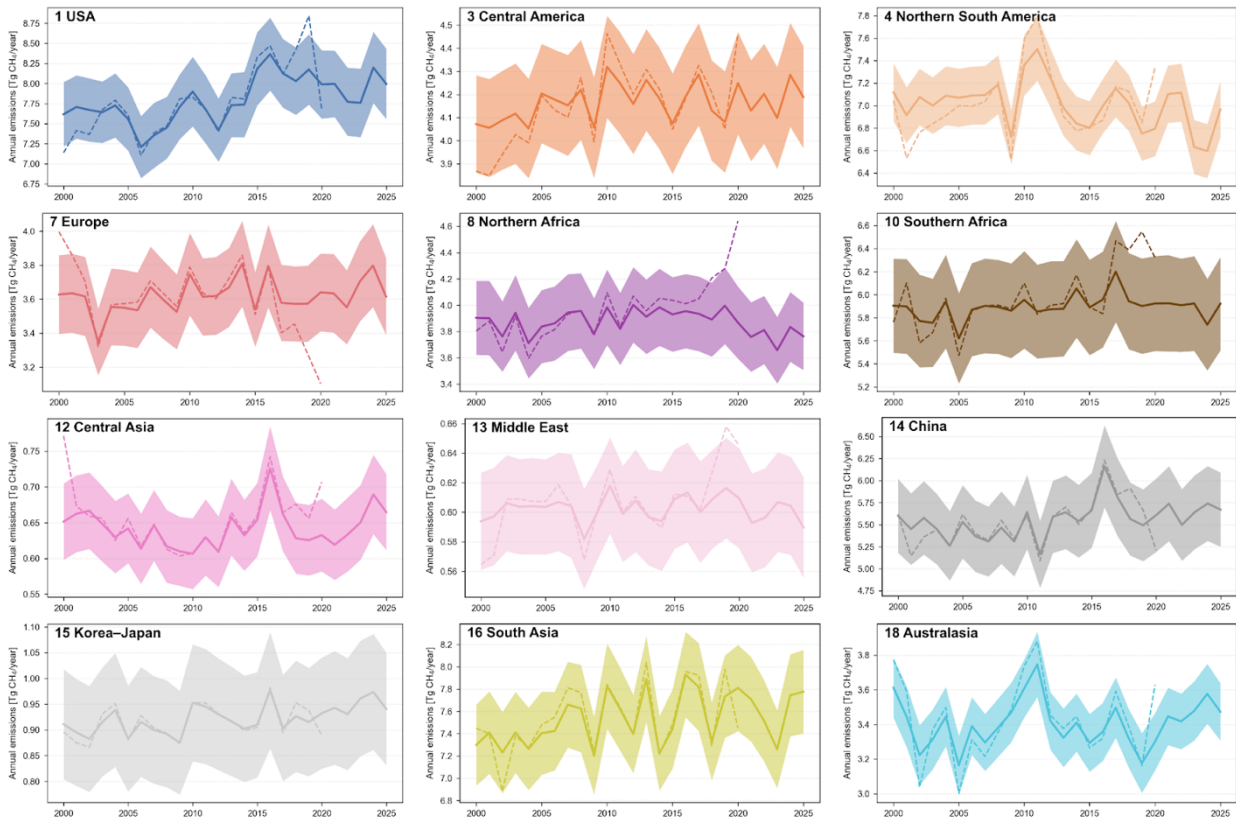


Figure S6. Annual wetland CH<sub>4</sub> emissions in five latitude bands. Annual mean emissions for global, 60-  
 55 90°N, 30-60°N, 30°S-30°N, and 90-30°S from 2000-2025. Solid lines show model predictions of annual mean emissions and shaded areas indicate 95% CI. Dashed lines show GMB emission estimates.



60 Figure S7. Monthly anomalies of wetland CH<sub>4</sub> emissions at five latitude bands (2000-2025). For each band, anomalies are relative to the 2000-2025 linearly detrended seasonal climatology (least squares), using predicted wetland CH<sub>4</sub> emissions from the XGBoost models.



65 Figure S8. Annual wetland CH<sub>4</sub> emissions for the remaining 12 regions. Same as Figure 6, but for the other 12 regions not included in Figure 6. Solid lines show model predictions of annual mean emissions and shaded areas indicate 95% CI. Dashed lines show GMB emission estimates (mean). Panel titles indicate the region code and region name, which correspond to the region definitions shown in Figure 4.

70