remote sensing

Supplement of

Physics-Informed Stacking Machine Learning for Surface Methane Estimation via Spectral Radiance Channel Selection

Haklim Choi [1,*], and Mi Eun Park [2], and Ji Eun Choi [1]

[1] Kyungpook Institute of Oceanography, Kyungpook National University, Daegu 41566, Republic of Korea
[2] National Institute of Meteorological Sciences, Korea Meteorological Administration, Seogwipo-si 63568, Republic of Korea
* Correspondence: haklim84@knu.ac.kr



Figure S1: Locations of the 42 GAW/WDCGG stations used in this study.

Table S1. GAW/WDCGG ground stations used in this study and station-level validation metrics. R^2, RMSE (ppb), and bias (ppb) were computed at each station from the out-of-fold predictions of the Ridge Ensemble (L2+L1B configuration) in the leave-one-year-out (LOYO) CV, and RMSE (L2) is the corresponding RMSE of the L2 configuration; for stations with fewer than 10 samples, only RMSE and bias are given. N is the number of collocated samples, and Alt is the sampling altitude above sea level (station elevation plus intake height).

#	Station	Country	Lat (°)	Lon (°)	Alt (m)	N	R^2	RMSE (L2+L1B)	RMSE (L2)	Bias
1	ALT (Alert)	Canada	82.45	−62.51	210	79	0.789	13.27	14.45	+4.9
2	TIK (Tiksi)	Russia	71.59	128.92	29	86	−0.116	47.57	45.59	−16.9
3	BRW (Barrow)	USA	71.32	−156.61	29	45	0.422	33.19	33.81	+7.9
4	CBY (Cambridge Bay)	Canada	69.13	−105.06	47	26	0.086	31.52	30.75	+13.9
5	INU (Inuvik)	Canada	68.32	−133.53	131	174	0.608	27.72	29.40	+3.0
6	PAL (Pallas)	Finland	67.97	24.12	577	360	0.791	17.34	17.64	+5.4
7	BCK (Behchoko)	Canada	62.80	−115.92	220	142	0.525	29.46	29.59	+9.5
8	UTO (Utö)	Finland	59.78	21.37	65	5	—	26.44	25.12	+12.6
9	CHL (Churchill)	Canada	58.74	−93.82	89	4	—	29.45	31.72	+27.2
10	LLB (Lac La Biche)	Canada	54.95	−112.47	585	186	0.356	42.25	43.65	−4.4
11	LUT (Lutjewad)	Netherlands	53.40	6.35	61	161	0.236	96.13	99.99	−17.3
12	WAO (Weybourne)	UK	52.95	1.12	27	124	0.330	67.89	70.53	−4.0
13	BRA (Bratt's Lake)	Canada	50.20	−104.71	630	397	0.603	25.97	26.01	+1.9
14	FSD (Fraserdale)	Canada	49.88	−81.57	250	191	0.681	17.18	20.27	+2.9
15	CPS (Chapais)	Canada	49.82	−74.98	423	105	0.696	19.82	22.81	−0.2
16	ESP (Estevan Point)	Canada	49.38	−126.54	47	5	—	21.64	20.83	+13.0
17	KAS (Kasprowy Wierch)	Poland	49.23	19.98	1996	91	0.703	23.14	23.04	+9.4
18	SSL (Schauinsland)	Germany	47.90	7.92	1214	147	0.256	27.85	27.66	−6.6
19	ZSF (Zugspitze)	Germany	47.42	10.98	2669	22	−0.409	34.52	30.93	+20.8
20	PUY (Puy de Dôme)	France	45.77	2.97	1475	325	0.571	24.10	23.81	+2.6
21	EGB (Egbert)	Canada	44.23	−79.78	271	228	0.515	28.35	30.92	−3.5
22	CMN (Monte Cimone)	Italy	44.19	10.70	2173	100	0.318	37.40	34.57	+14.2
23	WSA (Sable Island)	Canada	43.93	−60.01	30	3	—	27.56	35.40	−10.1
24	COI (Cape Ochiishi)	Japan	43.16	145.50	94	64	0.414	32.45	32.04	+10.4
25	CPA (Cholpon-Ata)	Kyrgyzstan	42.64	77.07	1623	76	0.155	19.00	18.63	−3.7
26	RYO (Ryori)	Japan	39.03	141.82	280	32	0.364	25.86	27.54	+8.3
27	ULD (Ulleungdo)	Republic of Korea	37.48	130.90	231	28	0.280	46.44	48.45	−18.1
28	AMY (Anmyeon-do)	Republic of Korea	36.54	126.33	124	158	0.374	52.71	53.82	−17.1
29	LMP (Lampedusa)	Italy	35.52	12.63	53	167	0.023	28.91	28.64	+16.6
30	GSN (Gosan)	Republic of Korea	33.29	126.16	82	98	0.481	39.66	40.46	−1.0
31	IZO (Izaña)	Spain	28.31	−16.50	2402	195	−0.122	19.61	19.76	+4.3
32	YON (Yonagunijima)	Japan	24.47	123.01	50	10	0.637	38.80	37.99	−12.1
33	MNM (Minamitorishima)	Japan	24.29	153.98	27	9	—	28.02	26.51	+8.2
34	HAT (Hateruma)	Japan	24.06	123.81	47	24	0.519	38.48	37.15	+4.0
35	MLO (Mauna Loa)	USA	19.54	−155.58	3437	16	0.403	20.08	21.11	+5.8
36	CVO (Cape Verde)	Cabo Verde	16.86	−24.87	31	198	0.237	17.17	16.28	−4.4
37	MKN (Mt. Kenya)	Kenya	−0.06	37.30	3688	13	0.261	32.86	34.19	−5.9
38	RUN (Réunion)	France	−21.08	55.38	2160	7	—	21.23	14.50	−9.2
39	TLL (El Tololo)	Chile	−30.17	−70.80	2159	9	—	18.39	16.61	+4.5
40	CPT (Cape Point)	South Africa	−34.35	18.49	260	270	0.406	16.87	16.74	−3.4
41	BHD (Baring Head)	New Zealand	−41.41	174.87	95	38	0.213	17.45	16.70	−10.8
42	SYO (Syowa)	Antarctica	−69.00	39.58	23	239	0.735	13.69	13.33	−5.5

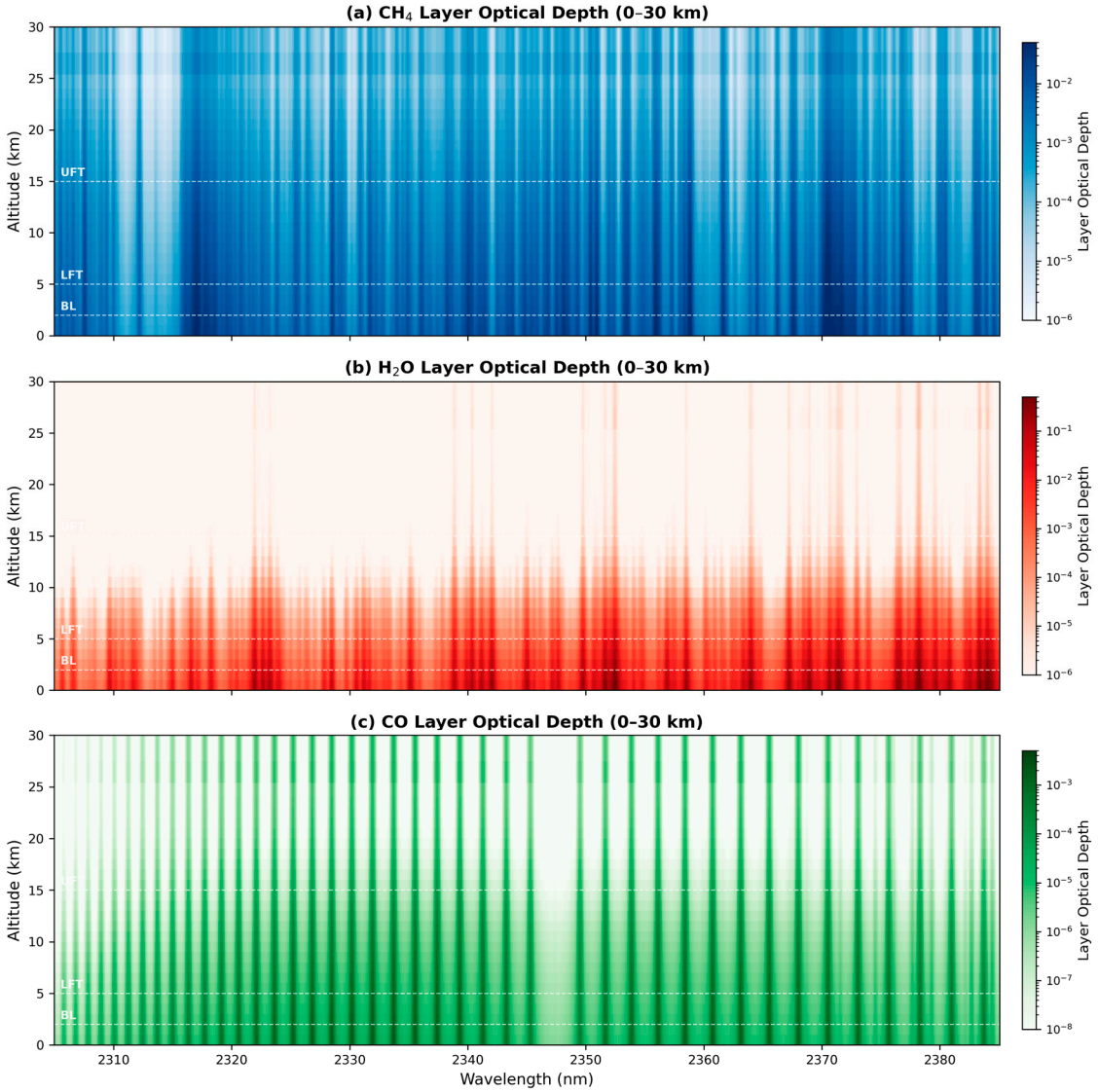


Figure S2: RTM-simulated optical depth profiles of (a) CH_4, (b) H_2O, and (c) CO as a function of wavelength.

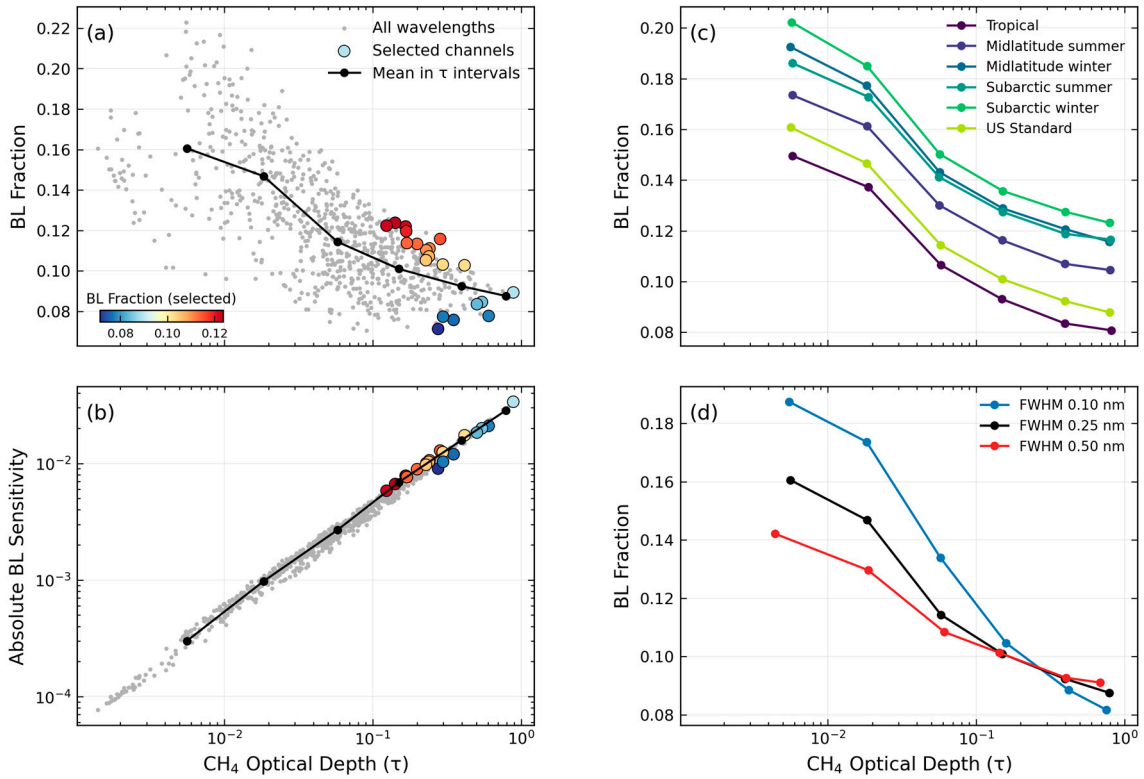


Figure S3: Relative and absolute sensitivity of the boundary layer (BL, 0–2 km) to CH_4 as a function of the CH_4 optical depth τ: (a) BL fraction of the normalized CH_4 Jacobian and (b) absolute BL sensitivity for the reference simulation, and BL fraction for (c) six AFGL standard atmospheres and (d) three widths of the instrument spectral response function.

To examine how the vertical sensitivity depends on the absorption strength, the BL fraction and the absolute BL sensitivity (the sum of the absolute normalized Jacobian over the BL layers) of all 801 simulated wavelengths were analyzed as a function of the CH_4 optical depth τ for the reference simulation used for the channel selection (US Standard Atmosphere, FWHM of 0.25 nm, solar zenith angle of 15°, near-nadir viewing, and surface albedo of 0.30). To assess the dependence on the atmospheric state and the instrument response, the calculation was repeated for six Air Force Geophysics Laboratory (AFGL) standard atmospheres with different pressure, temperature, and water vapor profiles [1] and for three widths of the instrument spectral response function (FWHM of 0.10, 0.25, and 0.50 nm), with all other conditions unchanged.

The BL fraction averaged over τ intervals decreased monotonically from 16.1% at $\tau < 0.01$ to 8.8% at $\tau = 0.6–1.0$ (Figure S3a), whereas the absolute BL sensitivity increased by about two orders of magnitude over the same range (Figure S3b). Because the normalized Jacobian of each layer scales with its CH_4 absorption (Figure S3b), the BL fraction represents the share of the BL in the CH_4 absorption at each wavelength. Its decrease is therefore not caused by the attenuation along the light path, but reflects the larger share of the upper layers in the absorption at strongly absorbing wavelengths, which is consistent with the pressure dependence of the absorption line shape [2,3]. The decrease of the BL fraction with τ was preserved for all six atmospheres, which changed only its level (Figure S3c), and for all three widths of the spectral response function, which changed only its spectral contrast (Figure S3d). The 20 selected channels lie at $\tau = 0.12–0.89$, where the absolute BL sensitivity is high, and have BL fractions of 7.1–12.4%.

Table S2. Nested cross-validation results under the LOYO CV: five base learners and the Ridge Ensemble of the L2+L1B configuration (upper) and the Ridge Ensemble of each input configuration (lower). LOYO R^2 is the value reported in Tables 1 and 2 of the main text, and nested CV R^2 is the value obtained when the hyperparameters and the Ridge meta-learner were determined using only the training years of each outer fold (Section 2.3 of the main text).

Model	**LOYO R^2**	**Nested CV R^2**	**Difference**
LightGBM	0.714	0.682	+0.032
XGBoost	0.722	0.704	+0.018
Random Forest	0.727	0.698	+0.029
SVR	0.722	0.691	+0.031
ANN	0.702	0.638	+0.064
Ridge Ensemble	0.753	0.737	+0.016
Configuration	**LOYO R^2**	**Nested CV R^2**	**Difference**
L2	0.742	0.723	+0.019
L1B	0.725	0.698	+0.027
L2+L1B	0.753	0.737	+0.016

Table S3. Changes in Degrees of Freedom for Signal (DFS) based on the number of channels (N). DFS values were computed for the reference simulation (solar zenith angle of 15°, near-nadir viewing, and surface albedo of 0.30) with a relative prior uncertainty of 10%, a vertical correlation length of 5 km, and a signal-to-noise ratio of 150.

Number of Channels	Total DFS	BL DFS	Lower Layer (0–5 km) DFS fraction (%)	Efficiency (DFS/N, %)
5	0.987	0.083	24.9	19.7
10	1.024	0.090	25.7	10.2
20	**1.068**	**0.102**	**27.0**	**5.3**
30	1.092	0.108	27.6	3.6
50	1.117	0.115	28.3	2.2
100	1.137	0.120	28.7	1.1

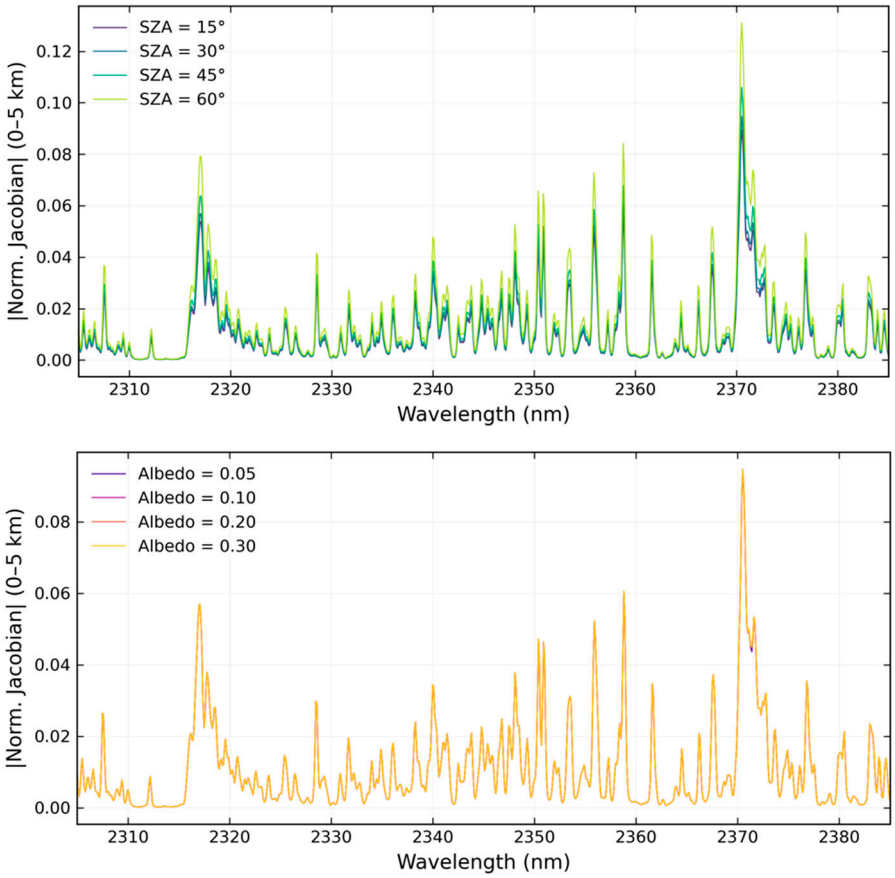


Figure S4: Changes in the normalized Jacobian spectrum according to changes in SZA and surface albedo.

Table S4. Ensemble LOYO performance of the L2+L1B configuration for different selections of L1B channels from the 2,305–2,385 nm band (20 channels each, except for the full band). All strategies used the same input variables, base learners, Ridge meta-learner, hyperparameters, and the LOYO CV; only the L1B channels differed.

Strategy	Description	Ensemble R^2	RMSE (ppb)
ML-Diverse (this study)	DFS + lower-atmosphere sensitivity + spectral diversity	0.753	33.98
Top-DFS	Sequential DFS maximization	0.753	34.03
Uniform	Equal spectral spacing	0.752	34.07
Full band	All wavelengths in the band	0.748	34.33

Table S5. Cross-validation strategy comparison for the L2+L1B configuration: LOYO (leave-one-year-out) and LOSO (leave-one-station-out, 42-fold) CV. The hyperparameters optimized on the LOYO splits were used in both cases.

Model	LOYO R^2	LOYO RMSE (ppb)	LOSO R^2	LOSO RMSE (ppb)
LightGBM	0.714	36.61	0.636	41.29
XGBoost	0.722	36.05	0.627	41.78
Random Forest	0.727	35.73	0.643	40.90
SVR	0.722	36.09	0.580	44.35
ANN	0.702	37.35	0.552	45.77
Ridge Ensemble	0.753	33.98	0.656	40.13

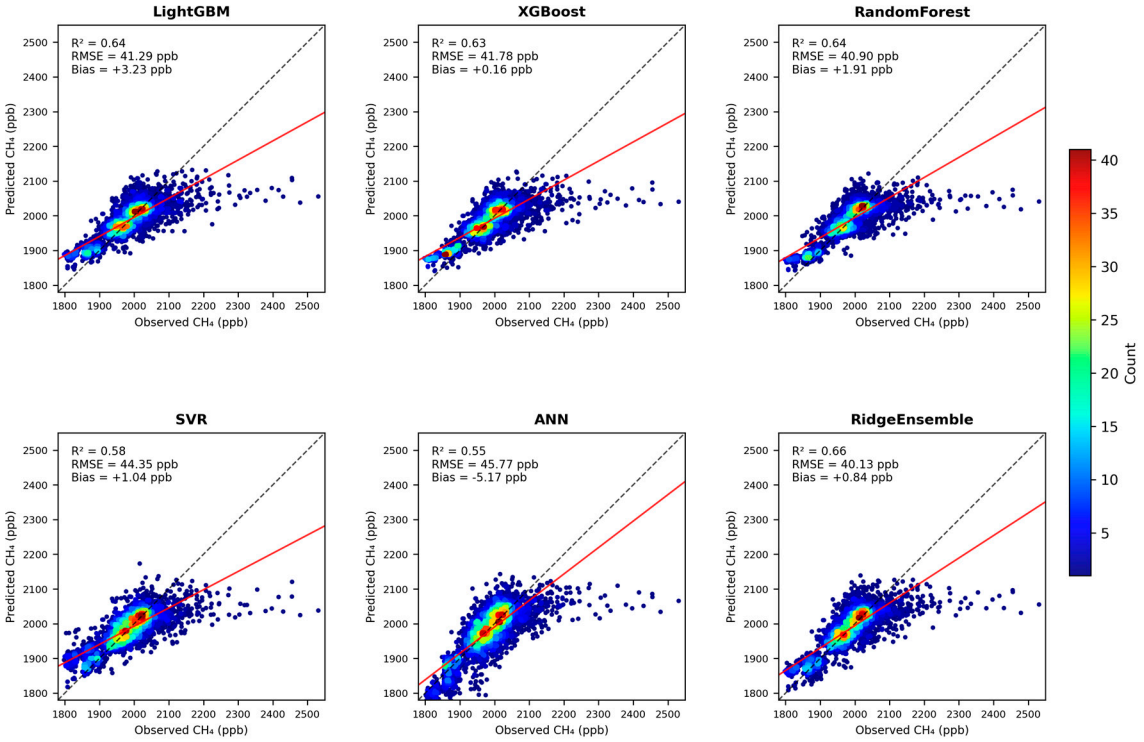


Figure S5: Density plots between observed and predicted surface CH_4 concentrations (ppb) under the LOSO CV for the five base learners (LightGBM, XGBoost, Random Forest, SVR, and ANN) and the Ridge Stacking Ensemble (L2+L1B configuration). The axes are the same as in Figure 5 of the main text; 52 ANN predictions below 1,780 ppb lie outside the plotted range but are included in the statistics.

To evaluate the performance at locations that were not used in training, a leave-one-station-out (LOSO) CV was performed for the L2+L1B configuration: all observations of each of the 42 stations were held out in turn, and the models were trained on the data of the remaining 41 stations. The input variables, base learners, Ridge meta-learner, and hyperparameters were the same as in the LOYO CV (Section 2.3 of the main text). Under the LOSO CV, the Ridge Ensemble retained $R^2 = 0.656$ (RMSE = 40.1 ppb), compared with $R^2 = 0.753$ (RMSE = 34.0 ppb) under the LOYO CV, and it remained the best-performing model (Table S5). The scores are lower than those under the LOYO CV, as expected, because station-specific characteristics, such as the influence of nearby emission sources and the sampling

altitude, cannot be learned for a station that is excluded from training, and validation designs that retain such dependence between the training and test data overestimate the predictive performance at new locations [4,5]. The distribution of the predictions is similar to that under the LOYO CV (Figure 5 of the main text), with a larger spread around the 1:1 line (Figure S5).

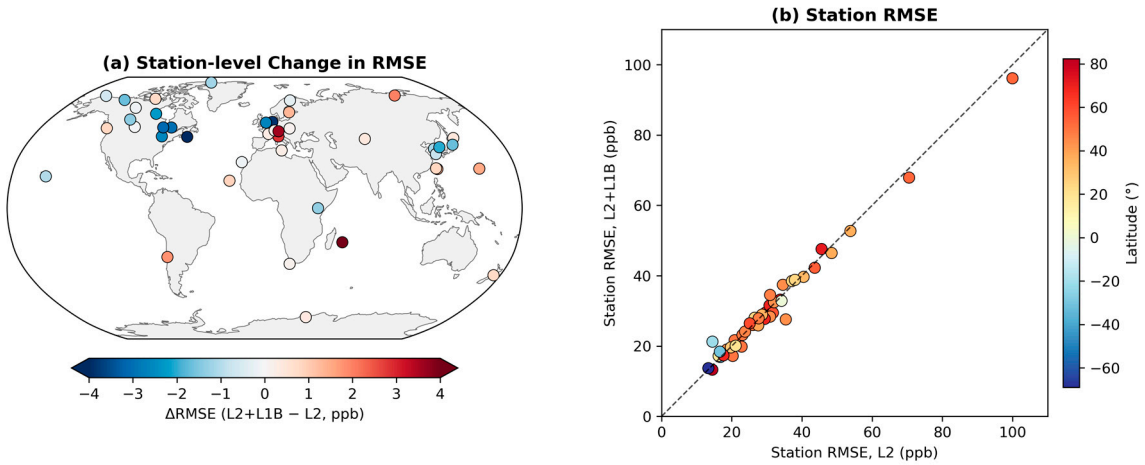


Figure S6: Change in the station RMSE when the L1B radiances are added to the L2 configuration, from the LOYO out-of-fold predictions of the Ridge Ensemble: (a) station-level change (L2+L1B − L2; blue: reduction) and (b) station RMSE of the two configurations (color: station latitude).

The station RMSE of the L2 and L2+L1B configurations was computed at each of the 42 stations from the LOYO out-of-fold predictions of the Ridge Ensemble (Table S1). Adding the L1B radiances reduced the station RMSE at 18 of the 30 Northern Hemisphere mid- and high-latitude stations, which provide 78% of the collocated samples, and the RMSE computed over all samples of each latitude band decreased from 41.3 to 40.1 ppb at 30–60°N and from 26.9 to 26.7 ppb at 60–90°N. Over the tropical and Southern Hemisphere stations, which provide 22% of the samples, the RMSE computed over all samples differed by less than 0.5 ppb between the two configurations, and the largest increase at a single station (Réunion) was based on only seven samples. The reduction of the overall RMSE from 34.8 to 34.0 ppb (Table 1 of the main text) therefore arises from the well-sampled Northern Hemisphere stations.

Table S6. Predictive performance comparison between individual base learners and the Ridge Stacking Ensemble model under the L1B input configuration (LOYO CV).

Model	R^2	Bias (ppb)	MAE (ppb)	RMSE (ppb)
Ridge Ensemble	**0.725**	**−0.07**	**22.19**	**35.91**
XGBoost	0.709	−2.99	23.39	36.90
Random Forest	0.701	−1.03	22.77	37.40
LightGBM	0.671	+0.74	26.01	39.21
ANN	0.667	−0.19	24.45	39.50
SVR	0.660	−4.33	23.13	39.87

Table S7. Same as Table S6, but for the L2 configuration.

Model	R^2	Bias (ppb)	MAE (ppb)	RMSE (ppb)
Ridge Ensemble	**0.742**	**+0.13**	**21.82**	**34.78**
Random Forest	0.726	+0.52	22.53	35.83
XGBoost	0.723	−3.69	22.91	36.04
LightGBM	0.709	−1.75	23.99	36.88
SVR	0.708	−3.86	21.88	36.95
ANN	0.696	−1.16	24.01	37.75

References

1. Anderson, G.P.; Clough, S.A.; Kneizys, F.X.; Chetwynd, J.H.; Shettle, E.P. *AFGL Atmospheric Constituent Profiles (0–120 km)*; AFGL-TR-86-0110; Air Force Geophysics Laboratory, Hanscom Air Force Base: Bedford, MA, USA, 1986.

2. Goody, R.M.; Yung, Y.L. *Atmospheric Radiation: Theoretical Basis*, 2nd ed.; Oxford University Press: New York, NY, USA, 1989.

3. Frankenberg, C.; Platt, U.; Wagner, T. Iterative maximum a posteriori (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH_4 and CO_2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT. *Atmos. Chem. Phys.* **2005**, *5*, 9–22. https://doi.org/10.5194/acp-5-9-2005

4. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. https://doi.org/10.1111/ecog.02881

5. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. https://doi.org/10.1038/s41467-020-18321-y