

Supplement Material of

Estimation, variations, and impact factors of high-resolution global daily near-surface CO₂ during 2015–2021 based on OCO-2 and machine learning

Ruobing Liu¹, Xinfeng Wang (✉)¹, Yuchao Ren^{2,1}, Chenliang Tao^{1,3}, Shuping Ji¹,
Zhenyu Gao¹, Yizhu Jiang¹, Shilong Ren¹, Lei Fang¹, Jinyue Chen¹, Qingzhu Zhang¹,
Guoqiang Wang^{1,4}, Qiao Wang (✉)¹

1 Big Data Research Center for Ecology and Environment, Environment Research
Institute, Shandong University, Qingdao 266003, China

2 Key Laboratory of Land and Sea Ecological Governance and Systematic
Regulation, Ministry of Ecology and Environment, Shandong Academy for
Environmental Planning, Jinan 250101, China

3 Department of Environmental Science and Engineering, Fudan University, Shanghai
200438, China

4 Innovation Research Center of Satellite Application, Faculty of Geographical
Science, Beijing Normal University, Beijing 100875, China

1 Overview

Here, we provide additional information on:

Fig. S1 The workflow of this study.

✉ Corresponding authors

E-mail: xinfengwang@sdu.edu.cn (X. Wang); qiao.wang@sdu.edu.cn (Q. Wang)

Fig. S2 Pearson correlation matrix among variables used in the near-surface CO₂ prediction model.

Fig. S3 Spatial distribution of station-specific *RMSE* values from LOSO CV.

Fig. S4 Spatial distribution of CO₂ prediction uncertainty computed using LightGBM quantile regression. The base map is approved under map review number GS(2016)1666.

Fig. S5 Density scatterplots of the test set showing observations and predictions from a model trained using only XCO₂ as the input feature. The dashed line represents the 1:1 line, and the solid line represents the best-fit line from linear regression.

Table S1. Summary of datasets used in this study.

Table S2. The final optimized hyperparameters of the LightGBM model.

Table S3. Statistical characteristics of predicted and observed CO₂ concentrations at the observation stations, where $\text{Bias} = 100\% \times (\text{Predicted} - \text{Observed}) / \text{Observed}$.

Table S4. The annual mean CO₂ concentration and growth rate at observation stations.

2 Contents of this file

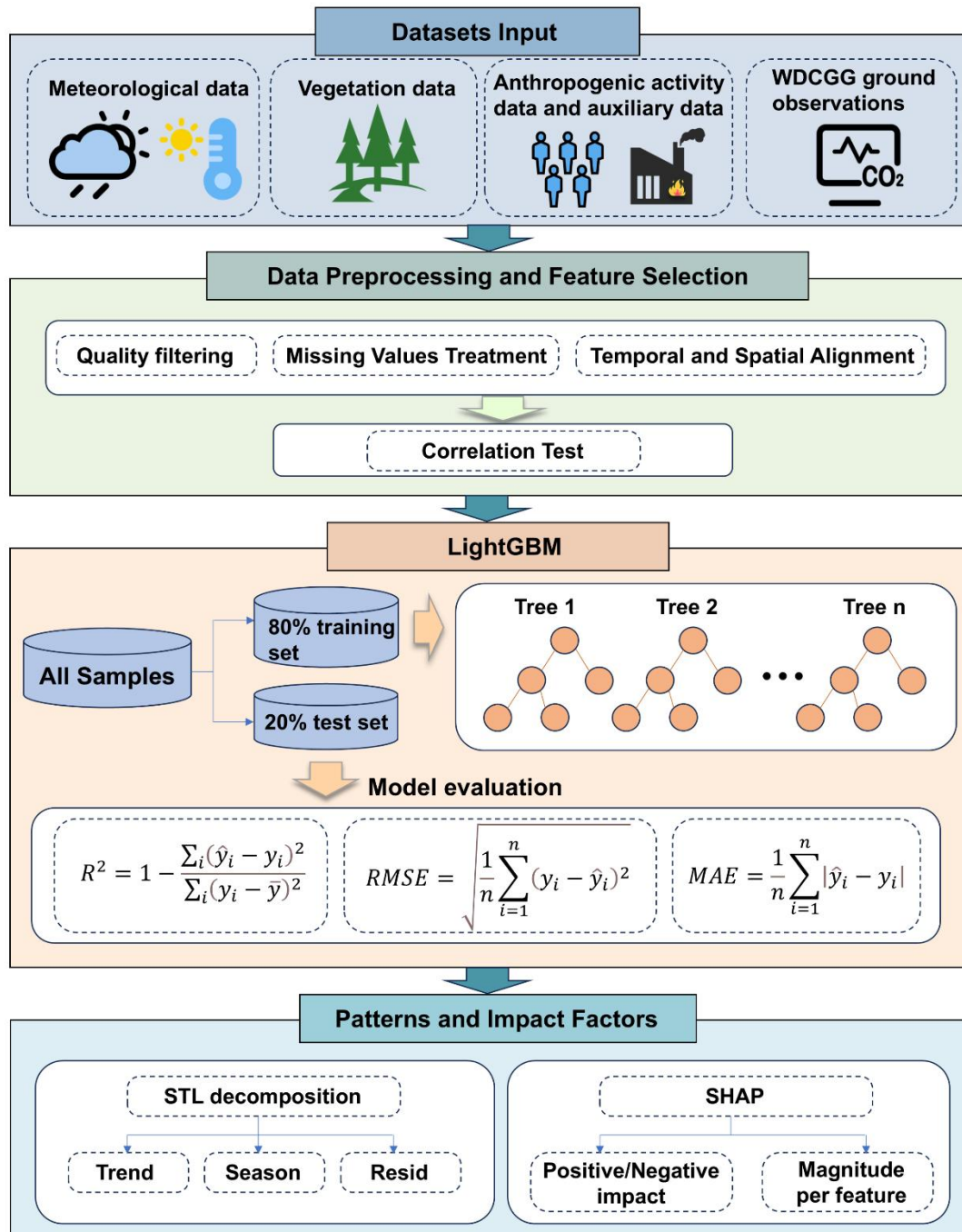


Fig. S1 The workflow of this study.

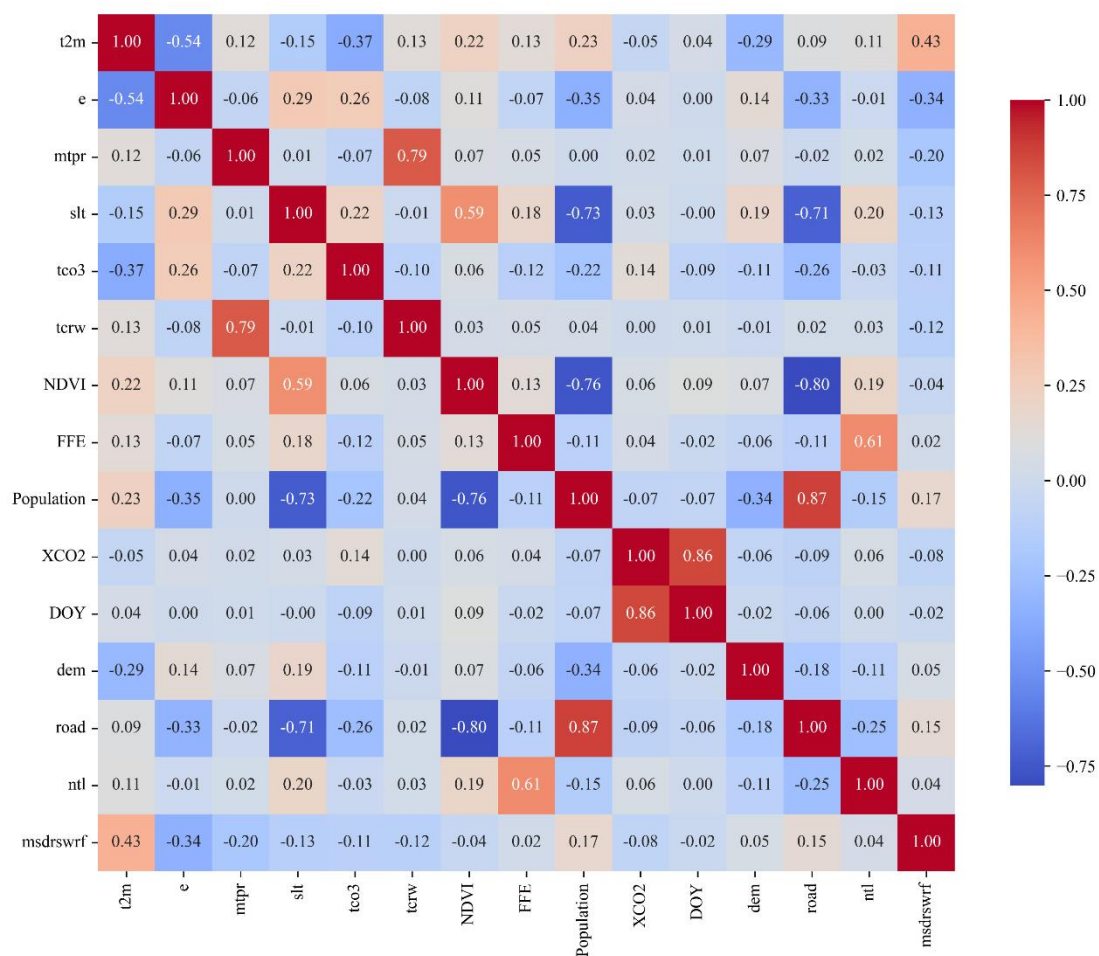
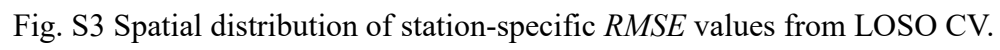


Fig. S2 Pearson correlation matrix among variables used in the near-surface CO₂ prediction model.



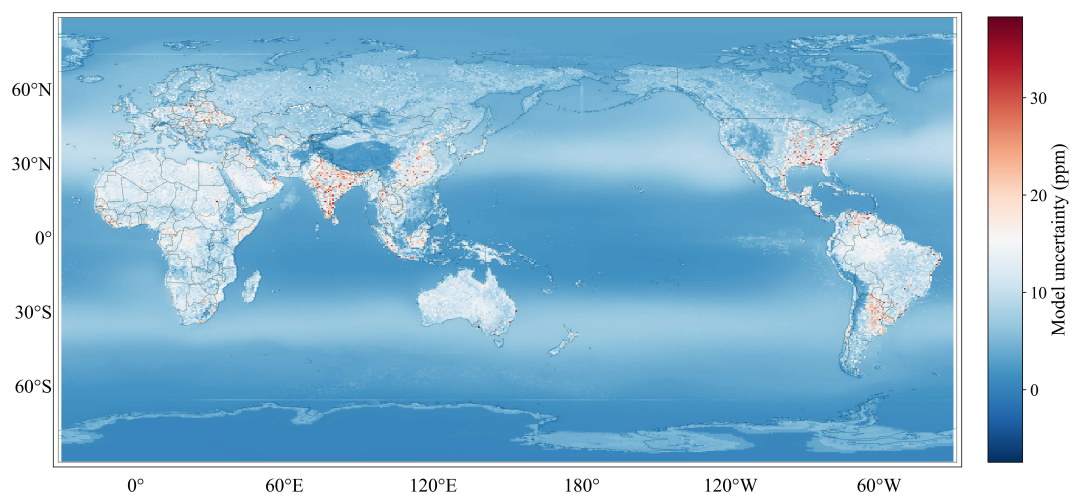


Fig. S4 Spatial distribution of CO₂ prediction uncertainty computed using LightGBM

quantile regression. The map review number GS 京 (2026) 2069.

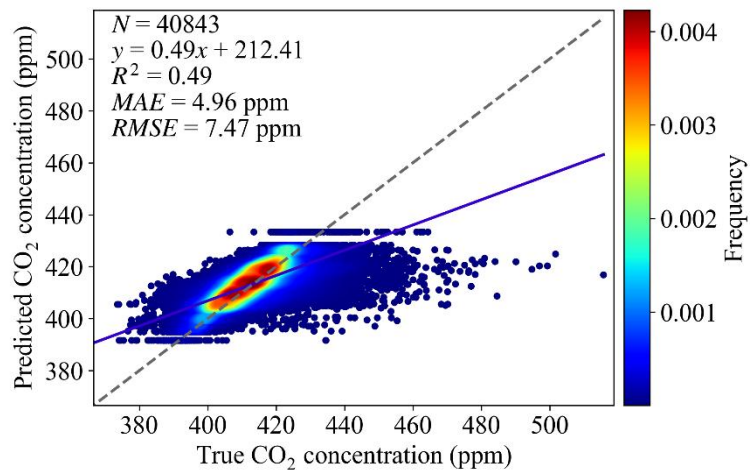


Fig. S5 Density scatterplots of the test set showing observations and predictions from a model trained using only XCO₂ as the input feature. The dashed line represents the 1:1 line, and the solid line represents the best-fit line from linear regression.

Table S1. Summary of datasets used in this study.

Variable	Source	Spatial resolution	Temporal resolution	Resampling method
CO ₂	WDCGG	\	daily	\
XCO ₂	OCO-2	0.5°×0.625°	daily	\
Vegetation	MODIS	0.5°×0.5°	monthly	bilinear interpolation
Meteorological and climatic factors	ERA5	0.25°×0.25°	hourly	bilinear interpolation
Population density	LandScan	1km×1km	annual	bilinear interpolation
Altitude	SRTM (Shuttle Radar Topography Mission)	30m×30m	\	bilinear interpolation
Fossil fuel emissions	ODIAC	1km×1km	monthly	bilinear interpolation
Road network density	The Global Roads Inventory Project (GRIP)	0.08°×0.08°	annual	bilinear interpolation
Nighttime Light	Xuecao Li's group	30arc-seconds	annual	bilinear interpolation

Table S2. The final optimized hyperparameters of the LightGBM model.

Hyperparameter	Value
learning_rate	0.08
n_estimators	7500
max_depth	12
boosting	'gbdt'
subsample	1.0
num_leaves	31
min_split_gain	0
min_child_weight	0.001
min_child_sample	20
colsample_bytree	1.0

Table S3. Statistical characteristics of predicted and observed CO₂ concentrations at the observation stations, where Bias = 100% × (Predicted – Observed) / Observed.

Stations	Mean prediction (ppm)	Mean observation (ppm)	Bias (%)
KIT	427.6832	427.7174	-0.008
PAL	410.703	410.6433	0.015
SSL	407.7331	407.5977	0.033
SPO	404.974	404.988	-0.003
TOH	415.8538	415.8546	0.000
ESP	410.1074	409.8236	0.069
DIG	424.635	424.6777	-0.010
WES	421.0627	421.2082	-0.035
CPA	415.5226	415.6672	-0.035
LMT	416.8602	416.9942	-0.032
ZEP	409.5256	409.5136	0.003
GAT	417.14	417.1885	-0.012
GSN	414.3634	414.3634	0.000
TAC	414.0791	414.0924	-0.003
AMY	416.0849	416.1925	-0.026
HPB	413.0272	413.1659	-0.034
RGL	413.4822	413.4572	0.006
TKY	414.4857	414.5331	-0.011
JFJ	408.7567	408.7314	0.006
CPT	405.6181	405.6393	-0.005
INU	410.6966	410.6489	0.012
ZSF	409.1309	409.2161	-0.021
MNM	409.9294	409.7992	0.032
ALT	410.1457	410.1122	0.008
SYO	404.8065	404.7874	0.005
BRW	411.065	410.9796	0.021
FSD	408.9351	408.5998	0.082
MLO	409.4466	409.3745	0.018
HUN	413.166	413.1361	0.007
BKT	420.7196	421.0935	-0.089
EGB	409.4786	409.7661	-0.070
IZO	408.6656	408.517	0.036
YON	412.1451	412.3844	-0.058
EST	410.0401	409.9138	0.031
HFD	414.0193	414.002	0.004
PDI	413.1973	413.1252	0.017
RYO	412.5618	412.516	0.011
CMN	409.2453	409.1332	0.027
GLH	402.767	395.57	1.819

BCK	410.0754	410.1552	-0.019
SNB	408.8167	408.8101	0.002
LLB	410.1445	410.033	0.027
LIN	421.4036	421.734	-0.078
COI	411.3715	411.2737	0.024
DDR	415.9866	414.3482	0.395
CVO	409.4625	409.403	0.015
HAT	411.5131	411.6318	-0.029
LMP	403.4029	406.5707	-0.779
CGR	405.4823	407.226	-0.428
WLG	409.4292	409.4462	-0.004
JUE	426.4444	426.3438	0.024
WSA	412.1532	412.1114	0.010
KIS	415.5489	417.6032	-0.492
HKO	419.7548	419.4294	0.078
HKG	413.6852	413.5435	0.034
CPS	408.2451	408.1914	0.013
OXK	418.1866	417.9905	0.047
STE	423.133	423.1116	0.005
PRS	404.7421	404.5983	0.036
ECO	420.7161	421.2954	-0.137
TLL	407.0002	407.0185	-0.005
CHL	411.9264	412.0483	-0.030
ETL	409.1472	409.4731	-0.080
SMO	406.5649	406.5307	0.008
CBY	408.9971	409.0064	-0.002
MKN	413.7456	413.8589	-0.027
TIK	411.0217	411.2382	-0.053
BRA	411.0331	411.0909	-0.014
HEL	420.8003	421.3565	-0.132

Table S4. The annual mean CO₂ concentration and growth rate at observation stations.

Stations	Mean CO ₂ concentration (ppm)		Growth rate (ppm/yr)
	2015	2021	
RYO	404.3467	420.5005	2.6923
YON	404.4529	419.4828	2.504981
MNM	401.6504	417.1287	2.579713
BRW	402.1715	417.9014	2.621646
MLO	401.0472	416.4068	2.559922
SMO	398.6652	413.5796	2.485728
IZO	400.9792	416.5145	2.589233
TKY	408.9167	424.9554	2.67312
WLG	401.4451	416.9206	2.579252
TLL	398.8353	414.0063	2.528512
HPB	406.909	419.731	2.137015
LIN	420.1529	433.4649	2.218663
ALT	402.0908	417.4574	2.561094
CHL	406.627	422.2292	2.600376
LLB	404.2972	428.1344	3.972867
ETL	403.104	426.8806	3.962776
FSD	401.7082	418.3104	2.76704
ESP	401.1824	421.2993	3.352826
EGB	401.9806	425.4411	3.91009
WSA	406.1906	422.8698	2.779862
TIK	403.9424	421.6469	2.950742
JFJ	400.9334	416.4799	2.591081
COI	403.2493	419.1861	2.656137
PAL	402.9957	418.9626	2.661146
HKO	407.8225	439.1324	5.218315
HUN	405.9403	420.4141	2.412287
CMN	400.2017	416.6086	2.73448
DIG	416.048	431.1427	2.515786
AMY	407.3119	423.8911	2.763192
HAT	403.6956	419.0827	2.564523
SYO	397.2909	412.0078	2.452808
ZEP	400.7801	417.0705	2.715055
KIS	410.2465	427.1683	2.820294
DDR	407.0381	424.5326	2.915743
CPT	397.768	412.9198	2.525306
ZSF	400.4611	416.4002	2.656514
SNB	400.4037	416.2448	2.64018
CVO	400.4343	415.7064	2.545361
TAC	407.0566	422.8482	2.631934
RGL	405.4244	421.6367	2.702053

HFD	405.816	422.2733	2.742886
HKG	402.2741	430.9619	4.781297
SPO	397.4605	412.08	2.436575
BCK	403.2705	422.908	3.272928
CBY	401.6658	419.7109	3.007513
INU	401.7808	417.9576	2.696131
EST	403.1372	425.8589	3.786952
BRA	403.1531	424.8313	3.613033
CPS	401.1035	419.3855	3.047004
GSN	406.6606	421.6263	2.494282
Mean	403.64	420.88	2.87
