MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 18, No. 2(2026), pp 110-122 DOI : 10.18860/mat.v18i2.45710
Received : August 19th 2026; **Accepted** : September 9th 2026; **Available Online** : September 29th 2026

Convolutional LSTM Forecasting of Methane Concentrations Using Copernicus CAMS Data for Relative Priority Screening of Regions in Java Island

Septian Nuno Zildjian, Dian Nurdiana, Fonda Leviany

***Abstract*—Methane (CH_4) is a potent greenhouse gas, and next-day forecasts of its concentration field can support the allocation of additional monitoring effort. This study forecasts daily column-mean CH_4 over Java Island with a Convolutional Long Short-Term Memory (ConvLSTM) network using Copernicus Atmosphere Monitoring Service (CAMS) fields, which are model forecasts and not satellite retrievals. Carbon monoxide (CO), an atmospheric pollutant and indirect climate-forcing agent, is used only as an auxiliary predictor. The data cover 1 March 2024 to 12 July 2026 on a 35 × 96 grid (0.1°, about 11 km) with 2,903,040 cell-days. Models use a seven-day window to predict the next day, a chronological 70/15/15 split and training-only scaling, and are compared in physical units with persistence, climatology, seasonal-naive, ridge regression, gradient boosting and a non-recurrent CNN. The variant selected on the validation block, a residual ConvLSTM with CH_4 and CO inputs, reached a test MAE of 7.15 ppb, RMSE of 9.89 ppb and R^2 of 0.618 (anomaly R^2 0.428), against 7.43 ppb, 10.35 ppb and 0.582 for persistence (MSE skill score 0.087). Given the short record (583 training days), the margin is small: ridge regression performed comparably (MAE 7.02 ppb), and a rolling-origin test showed no gain over persistence in the earliest fold. Area-weighted aggregation to 119 districts and cities gave relative priority classes that largely reproduce the persistent spatial pattern, with the Jakarta metropolitan area ranked highest. The classes are quartile ranks for screening, not validated risk levels or source attributions.**

Index Terms—**Methane, ConvLSTM, CAMS Forecasts, Spatio-Temporal Forecasting, Relative Priority Screening, Java Island**

Manuscript received April 17, 2026 This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University.
S. N. Zildjian is with the Department of Data Science, Faculty of Science and Technology, Universitas Terbuka, Indonesia (e-mail: septiannunozildjian@gmail.com; phone: 083876826854).
D. Nurdiana is with the Department of Data Science, Faculty of Science and Technology, Universitas Terbuka, Indonesia (e-mail: dian.nurdiana@ecampus.ut.ac.id; phone: 082211001131).
F. Leviany is with the Department of Data Science, Faculty of Science and Technology, Universitas Terbuka, Indonesia (e-mail: fonda.leviany@ecampus.ut.ac.id; phone: 089606104697).

I. Introduction

Methane (CH_4) is a potent greenhouse gas whose global warming potential is dozens of times that of CO_2 over a 20-year horizon [1]. Carbon monoxide (CO) is treated differently in this study: it is an atmospheric pollutant and an indirect climate-forcing agent that acts through tropospheric chemistry, not a greenhouse gas on a par with CH_4, and it is used only as an auxiliary predictor. Satellite instruments such as TROPOMI on Sentinel-5 Precursor provide daily, near-global observations of the CH_4 and CO columns [2], [3], whereas model-based products such as the Copernicus Atmosphere Monitoring Service (CAMS) global greenhouse gas forecasts, which are started from a data-assimilation analysis, provide complete, gap-free gridded fields [4]. The two kinds of data differ in resolution, uncertainty and interpretation and are not interchangeable.

Java Island combines the highest population density in Indonesia with intensive transportation, industrial and agricultural activity, which makes it a natural focus for anticipatory environmental monitoring. Retrospective monitoring alone gives little guidance on where limited measurement and investigation effort should be directed next.

This study addresses one defined decision problem: given daily CH_4 fields up to day t, forecast the field for day t+1, and use the forecasts to rank districts and cities by relative concentration and persistence so that further monitoring can be directed to areas that consistently rank highest. It is a forecasting and screening study. It does not identify emission sources, estimate emissions, or evaluate energy-transition, decentralized energy-trading [5] or biomass-waste [6] interventions, which would require source, socio-economic and energy-system data that are not used here. Meteorological drivers are likewise outside the scope: the study is a gas-only forecasting benchmark, and meteorological conditioning is left as the next step.

The objectives are: (1) to forecast next-day column-

mean CH_4 over Java Island with a ConvLSTM and compare it with persistence, climatological, regression and non-recurrent CNN baselines in physical units and with uncertainty estimates; (2) to test whether CO and seasonal inputs, a residual output formulation, the input window and the forecast horizon change forecast skill, including under a rolling-origin evaluation; and (3) to aggregate the forecasts to districts and cities with a documented area-weighted procedure and to examine whether the resulting relative priority classes carry information beyond the static climatological map.

This is a forecasting study, and no architectural novelty is claimed: ConvLSTM forecasting of gridded fields and the spatial join to administrative boundaries are established techniques. The contribution is empirical. It is a leakage-aware, baseline-controlled benchmark of next-day CH_4 forecasting over Java Island on a record that is still short (the CAMS archive used here starts in March 2024), which documents how much skill a ConvLSTM adds over persistence and simple regression, together with a transparent regional screening derived from the forecasts. Such a benchmark is a reference point for later work with longer records and meteorological inputs.

II. Related Work

Satellite retrievals and model data. TROPOMI retrievals have been used to detect CH_4 leaks from coal mines [7], have been validated against TCCON and NDACC-IRWG reference stations [8] and have been evaluated at northern high latitudes [9]. Retrievals are affected by cloud cover and quality filtering, and the resulting gaps have motivated machine-learning gap-filling [10]. The CAMS global greenhouse gas forecasts, in contrast, are produced by a numerical model started from a data-assimilation analysis and are complete in space and time [4]. The absence of gaps is a property of the product and not evidence of retrieval quality; the fields inherit the errors of the forecasting and assimilation system and are not direct measurements.

Spatio-temporal deep learning. ConvLSTM replaces the linear operations of the fully connected LSTM with convolutions so that spatial and temporal dependence are captured together, and it outperformed an optical-flow-based method in radar precipitation nowcasting [11], [12]. It has since been applied, for example, within a graph-convolution and ConvLSTM model for multisource $PM_{2.5}$ forecasting [13], to Sentinel-5P aerosol-index data for smog forecasting in South Asia [14] and to wind-energy potential in Indonesia from meteorological reanalysis [15]. Graph-convolutional alternatives for non-grid spatial structure [16] and attention-based recurrent models for radar precipitation [17] have also been proposed. Table 1 summarises these studies as far as they report comparable information.

Table 1. Related Spatio-Temporal Deep-Learning Studies and the Present Study

Study	Target and input data	Model	Reported result
Shi et al. [11], [12]	Precipitation nowcasting from radar echo sequences	ConvLSTM	Outperformed an optical-flow-based method
Muthukumar et al. [13]	Hourly $PM_{2.5}$; meteorological, wildfire/heat (MODIS), TROPOMI and ground-sensor data	Graph convolution + ConvLSTM	Improved RMSE and NRMSE over earlier work (values not summarised here)
Zhao et al. [17]	Strong convective precipitation; radar observations	ResNet-Attention-BiLSTM	MAPE reduced by 14.91–18.72% and R^2 increased by 9.29–10.89% over existing models, at the three levels reported
Khan [14]	UV aerosol index (340–380 nm), Sentinel-5P, 2019–2023	ConvLSTM	Five-day interval forecasts; MSE ≈ 0.0018, SSIM ≈ 0.74
This study	**Column-mean CH_4, CAMS global greenhouse gas forecasts (0.1°), Mar 2024–Jul 2026**	**ConvLSTM and baselines**	**MAE 7.15 ppb, RMSE 9.89 ppb, R^2 0.618, against persistence 7.43 ppb, 10.35 ppb and 0.582 (Table 9)**

Three limitations of this body of work shape the present design. First, the reported metrics differ in scale and target (RMSE and NRMSE of hourly $PM_{2.5}$ [13], relative gains in MAPE and R^2 over a baseline [17], MSE and SSIM of an index [14]), so they cannot be compared with each other or with a concentration forecast in ppb. Second, column gas fields are strongly persistent in time, so a model that largely reproduces yesterday's field can obtain a high pooled R^2; without a persistence baseline and a score computed on departures from the climatological map, skill cannot be separated from persistence. Third, translating grid forecasts into region-level products, which a screening application needs, requires a documented aggregation procedure for cells that straddle several regions, edge cells and small regions, and this is rarely the focus of forecasting studies.

The gap addressed here is therefore methodological rather than architectural: a leakage-aware, baseline-controlled evaluation of ConvLSTM forecasts of a persistent CH_4 field in physical units, with an explicit test of whether region-level screening adds information over the climatological map. Several of the studies above also use meteorological or fire inputs [13], whereas the present benchmark is gas-only by design. No claim is made that earlier studies have not predicted several gases or mapped forecasts to regions.

III. Data and Methodology

The workflow (Fig 1) comprises data quality control and gridding, a leakage-aware experimental design, baseline and ConvLSTM models, evaluation in physical units, and a documented aggregation of the forecasts to administrative regions for relative priority screening. All scaling statistics are computed from the training period only.

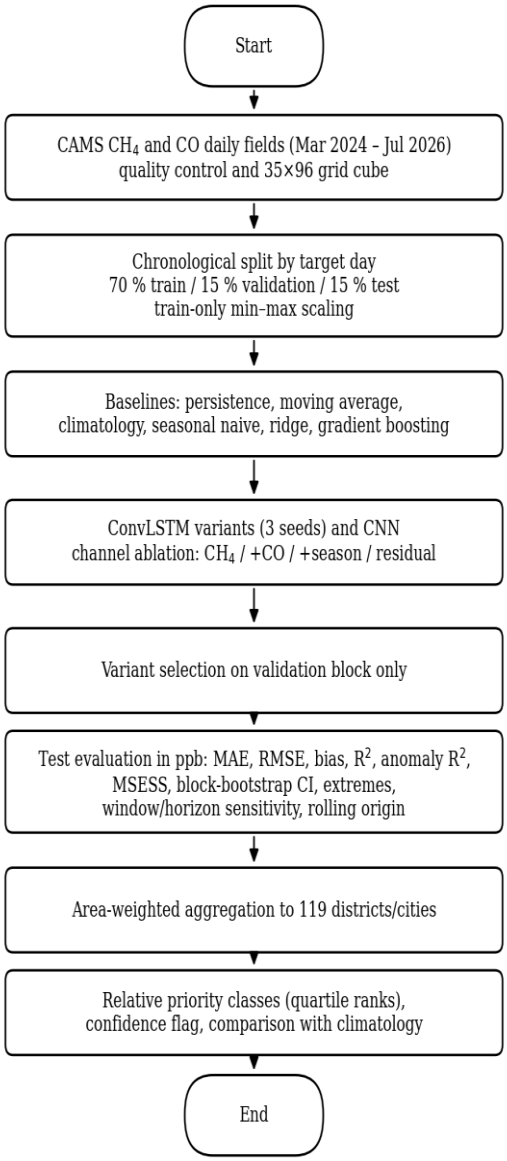


Fig 1. Workflow of the study.

A. Data Source and Study Area

The data are daily CH_4 and CO fields from the CAMS global greenhouse gas forecasts [4], obtained from the Atmosphere Data Store (ADS). The product consists of 5-day, 0.1° × 0.1° global forecasts of CO_2 and CH_4, with CO and meteorological parameters, issued once a day from March 2024. The fields are model forecasts started from a data-assimilation analysis. They are not satellite retrievals, and in particular not TROPOMI retrievals, and they are not direct measurements, so cloud screening and retrieval quality flags do not apply. In this paper "observed" denotes the CAMS field used as the forecast target, not a measurement. CH_4 is the column-mean molar fraction in ppb and CO the total column in kg m^{-2}, as given in the product documentation (Table 2). The daily series was built from the 00 UTC forecast cycle at a lead hour of +12 h (valid 12 UTC), so that each day is represented by a single field with a fixed forecast horizon; other lead hours of the 5-day run were not used. The ADS documentation notes that the forecasting system is upgraded about once a year, which can change resolution and accuracy [4].

TABLE 2. VARIABLES USED

Variable	Description	Unit (as stored)
latitude	Grid-cell latitude	degrees north
longitude	Grid-cell longitude	degrees east
dataDate	Date of the daily field	YYYY-MM-DD
tcch4	Column-mean CH_4 molar fraction; forecast target and predictor	ppb
tcco	Total-column CO; auxiliary predictor only	kg m^{-2}

The record covers 1 March 2024 to 12 July 2026 (864 days) over Java Island (latitude −9.0° to −5.6°, longitude 105.1° to 114.6°), with 3,360 grid cells and 2,903,040 cell-days. Boundaries of 119 districts and cities were taken from the GeoJSON file Java_Province.json (attributes KAB_KOTA and PROVINSI) and processed with GeoPandas [18]. All temporal features, namely the cyclic seasonal encodings and lagged values, are created within the analysis from the raw fields and calendar dates rather than taken from pre-computed columns.

B. Quality Control and Gridding

The quality checks in Table 3 found no missing values, no duplicated cell–day records, no tcch4 values outside the plausible range of 1,700–2,300 ppb, no non-positive tcco values and no robust-z outliers. Zero missing values is expected for a model product and is not evidence of retrieval quality. The values were arranged in a cube of 864 × 35 × 96 (time, latitude, longitude); no cell was empty, so no value was interpolated or dropped. The grid spacing is uniform at 0.1° in both directions, about 11.1 km north–south by 11.0 km east–west, and coverage is complete (Fig 2).

TABLE 3. DATA-QUALITY SUMMARY BEFORE MODELLING

Check	Result
Rows (observations)	2,903,040
Grid cells (lat × lon)	35 × 96 = 3,360
Days	864 (2024-03-01 to 2026-07-12)
Expected rows for a complete cube	2,903,040
Spatial coverage (rows / expected)	100.00%
Grid spacing, latitude (deg): median / min / max	0.100 / 0.100 / 0.100
Grid spacing, longitude (deg): median / min / max	0.100 / 0.100 / 0.100
Approximate cell size (km, N–S × E–W)	11.1 × 11.0
Missing values (tcch4, tcco)	0
Duplicated (cell, day) records	0
tcch4 outside 1,700–2,300 ppb	0
tcco ≤ 0	0
Robust-z outliers in tcch4 ($\|z\| > 5$)	0 (0.000%)
Values interpolated / dropped	0 / 0

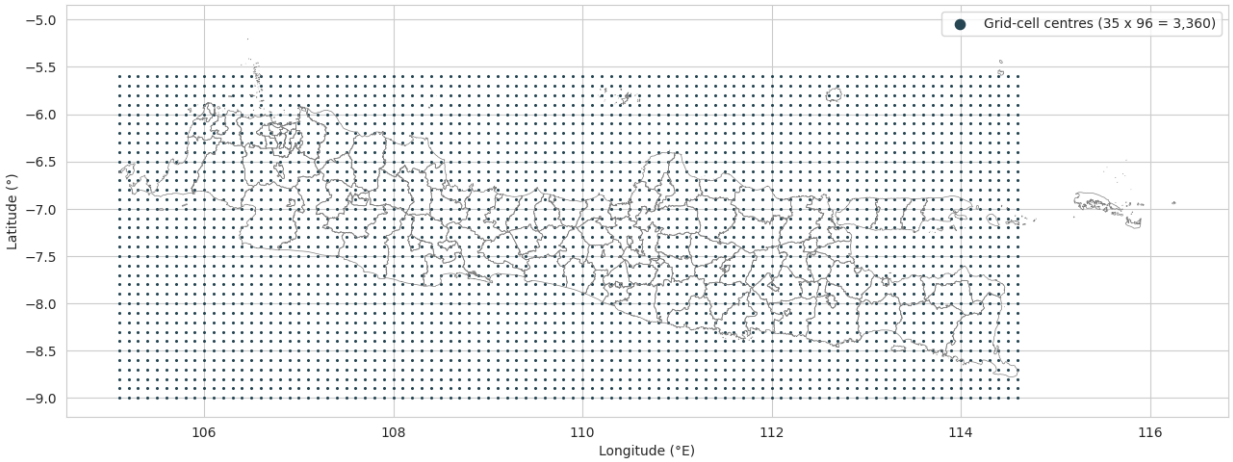


Fig 2. Grid-cell centres (35 × 96 = 3,360; 0.1° spacing, about 11 km) and administrative boundaries.

The 0.1° spacing equals the native horizontal resolution listed for the product [4], so the cells are not interpolated copies. As a model field, however, neighbouring cells remain strongly correlated and are not independent observations.

C. Spatial Aggregation to Administrative Regions

Grid values are aggregated to the 119 district/city polygons by area-weighted averaging. Each cell is treated as a 0.1° square and intersected with the polygons in an equal-area projection (EPSG:6933); the value of a region is the mean of the intersecting cells weighted by intersection area, so cells that straddle several regions contribute to each in proportion to overlap and edge cells and small regions need no special handling. A point-in-polygon variant, in which a cell belongs to the region containing its centre and regions without a cell centre take the nearest cell, serves as a sensitivity check. As Table 4 shows, regions intersect between 2 and 54 cells (median 19), six regions intersect fewer than three cells and 14 contain no cell centre; regional rankings of the two methods are nearly identical, and the area-weighted method is used throughout.

TABLE 4. SPATIAL-AGGREGATION DIAGNOSTICS

Diagnostic	Value
Number of regions	119
Cells intersecting a region: min / median / max	2 / 19 / 54
Regions intersecting fewer than 3 cells	6
Regions with no cell centre inside	14
Spearman ρ, area-weighted vs point-in-polygon regional means	0.9955

D. Experimental Design

The split is chronological on the forecast target day (Table 5, Fig 3): training 22 March 2024–25 October 2025 (583 days), validation 26 October 2025–4 March 2026 (130 days) and test 5 March–12 July 2026 (130 days). The first 21 target days are reserved so that every window and horizon setting shares the same target days. The validation block is a separate chronological block used only for early stopping and model selection; the test block is used once for reporting. Min–max parameters for CH_4 (1,852.906–2,018.992 ppb) and CO (0.000472–0.002991 kg m^{-2}) were computed from the training frames only and applied unchanged to validation and test data. None of the test-period CH_4 values exceeded the training maximum. The output layer is linear, so predictions are not bounded to the training range, and predictions are inverse-transformed to ppb before any metric is computed.

TABLE 5. CHRONOLOGICAL DATA SPLIT BY FORECAST TARGET DAY

Split	Target days	N	Role
Training	2024-03-22 – 2025-10-25	583	Fit weights
Validation	2025-10-26 – 2026-03-04	130	Early stopping, model selection
Test	2026-03-05 – 2026-07-12	130	Final reporting only

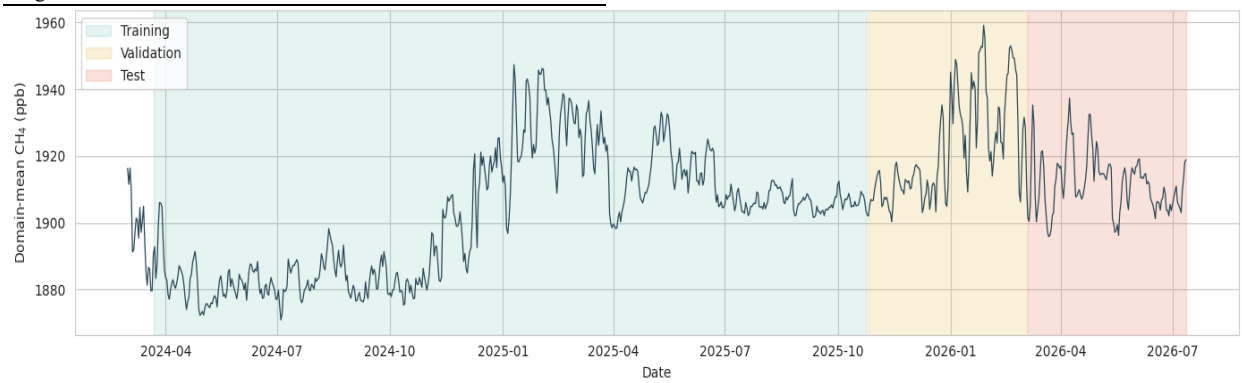


Fig 3. Domain-mean CH_4 with the chronological training, validation and test blocks.

A sample consists of a window of 7 daily frames (SEQ_LEN = 7) and the CH_4 frame one day after the last input frame (HORIZON = 1). Windows of neighbouring samples overlap, but samples are assigned to splits by target day, so no target day appears in more than one split. Four input configurations are compared:

CH_4 only; CH_4 and CO; CH_4, CO and four cyclic seasonal encodings (sine and cosine of day of year and of month); and a residual variant with CH_4 and CO in which the network predicts the change from the last observed CH_4 frame.

E. Baselines, Models and Training

Six baselines are evaluated on exactly the same validation and test days: persistence (the last observed frame), a 7-day moving average, per-cell climatology of the training period, a seasonal-naive forecast (the value 365 days earlier), ridge regression and histogram gradient boosting [19]. The two regression baselines are pooled over cells, use lagged CH_4 (lags of 0, 1, 2 and 6 days and a 7-day mean), CO, coordinates and seasonal terms, and predict the change from the last observed frame because tree models cannot extrapolate beyond the training range of a trending quantity. Gradient boosting was trained on a random subsample of 400,000 training cell-days.

The ConvLSTM has the structure ConvLSTM2D (32 filters, 3×3, tanh, return sequences), batch normalisation [20], ConvLSTM2D (16 filters, 3×3, tanh), batch normalisation and a Conv2D output layer (1 filter, 3×3, linear). In the residual variant the output is added to the last observed scaled CH_4 frame. The non-recurrent CNN uses the same window with the days stacked as channels (two Conv2D layers of 32 and 16 filters with ReLU and batch normalisation, and a linear Conv2D output). Models were trained with Adam [21], [22] on the MSE of the scaled target, with an initial learning rate of 10^{-3} halved on validation plateau (minimum 10^{-5}), batch size 8, up to 60 epochs, early stopping on validation loss with patience 8 and restoration of the best weights, and no dropout or weight decay. Every neural model was trained three times with seeds 42–44. The ensemble mean of the runs is used for metrics, maps and regional results, and run-to-run variability is reported as mean ± standard deviation (Table 6). Training used two NVIDIA Tesla T4 GPUs in a Kaggle notebook environment with TensorFlow/Keras.

TABLE 6. NEURAL MODEL SPECIFICATIONS

Model	Input channels	Parameters	Mean epochs
ConvLSTM (CH_4 only)	CH_4	66,193	60
ConvLSTM (CH_4 + CO)	CH_4, CO	67,345	41
ConvLSTM (CH_4 + CO + season)	CH_4, CO, season	71,953	39
ConvLSTM residual (CH_4 + CO)	CH_4, CO	67,345	35
CNN (non-recurrent)	CH_4, CO	9,025	40

F. Evaluation Scheme

Performance is evaluated on the test block (130 days × 3,360 cells) in ppb with MAE, RMSE and pooled R^2 [23], [24]:

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where y_i is the observed value, $\hat{y}_i$ the forecast, $\bar{y}$ the mean of the observed values and n the number of cell-days. Bias is the mean of forecast minus observation. Because pooled R^2 is dominated by the persistent spatial pattern, an anomaly R^2 is also reported, in which the per-cell training-period mean is removed from the observations before the same ratio is computed. The MSE skill score (MSESS) is $1 - \mathrm{MSE}_{\mathrm{model}}/\mathrm{MSE}_{\mathrm{persistence}}$ and is positive when the model beats persistence [25]. Extreme events are exceedances of the 95th percentile of the training period (1,939.5 ppb) and are scored by probability of detection (POD), false-alarm ratio (FAR), critical success index (CSI) and the mean intensity bias on observed extremes. The 95% confidence intervals use a moving-block bootstrap over test days (block length 7 days, 1,000 resamples) [26]. Sensitivity to the input window (3, 7 and 14 days) and horizon (1, 3 and 7 days) is examined with one run per setting, and a three-fold expanding-window rolling-origin evaluation [27], with scaling refitted on each fold's training frames, checks whether conclusions depend on one test block.

G. Relative Priority Screening

The forecasts of the selected variant for the test period are aggregated to regions. Two products are derived. The relative concentration class is the quartile (by rank) of the regional mean forecast. The relative priority class is the quartile of a composite index, the mean of the z-scores of the regional mean forecast and of persistence, defined as the share of test days on which the region lies in the top quartile of the daily forecast regional values. Regions whose regional MAE is in the worst quartile are flagged as low-confidence instead of being ranked silently. The classes are quartile ranks of the studied regions, so each holds about a quarter of them by construction [28]; they are not regulatory thresholds, health-risk levels, emission intensities or source attributions. Population exposure and source-related layers are not available and are not part of the index. To test whether forecasting adds information, regional rankings from the forecast, from persistence and from the training climatology are compared with the observed test-period ranking.

IV. RESULTS AND DISCUSSION

A. Exploratory Analysis of the Data

Table 7 summarises the two variables: CH_4 has a mean of 1,906.0 ppb and a standard deviation of 21.5 ppb, and CO a mean of 0.885×10^{-3} kg m^{-2}. The histograms with kernel density estimates [29] in Fig 4 show a main CH_4 mode near 1,900–1,910 ppb, a secondary shoulder near 1,875 ppb and a long right tail, whereas CO is unimodal and right-skewed. Extreme values were retained because, in a physically continuous model field, they represent real concentration gradients and not measurement errors.

TABLE 7. DESCRIPTIVE STATISTICS OVER ALL CELLS AND DAYS

Variable	Mean	SD	Min	Median	Max
CH_4 (ppb)	1906.02	21.47	1852.91	1904.50	2018.99

Variable	Mean	SD	Min	Median	Max
CO (10^{-3} kg m^{-2})	0.885	0.155	0.472	0.860	2.991

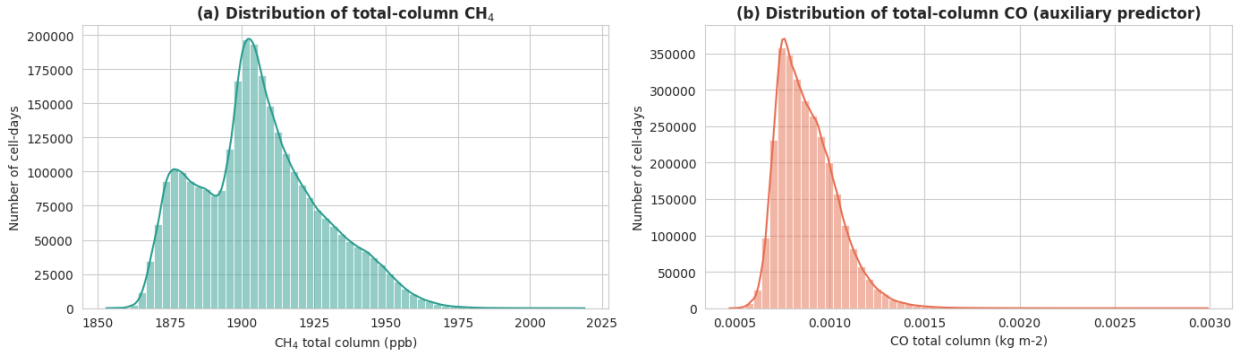


Fig 4. Value distributions of CH_4 and CO over all grid cells and days.

The domain-mean series (Fig 5) shows CH_4 near 1,880 ppb through most of 2024, a step to roughly 1,900–1,945 ppb from late 2024, and the highest daily means (close to 1,960 ppb) in January–February 2026. Domain-mean CO reached a pronounced maximum around October–November 2024 and was lower and steadier afterwards. Calendar-month means (Fig 6) are lowest for CH_4 in September–October (about 1,894 ppb) and highest in January–February (about 1,931 ppb), while CO peaks in October–November and is lowest in July. The CO maximum in the dry-season months is consistent with the biomass-burning influence reported for Indonesia [30], [31], but this study does not test that hypothesis. The record is only about 2.4 years long, so each calendar month is represented by two or three years, and the CH_4 monthly means are confounded with the level shift of late 2024; the CH_4 seasonal cycle cannot be separated from that shift. Because the product documentation states that the forecasting system is upgraded about once a year [4], the level shift may reflect a change in the product and not in the atmosphere. This was not investigated, and it is a further reason to treat the seasonal estimates as uncertain.

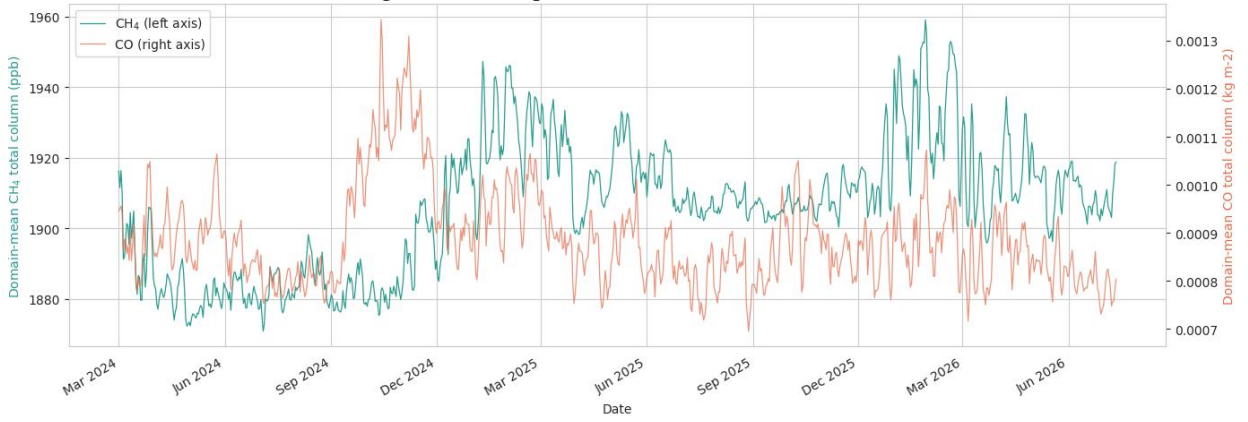


Fig 5. Daily domain-mean CH_4 and CO total columns over Java Island.

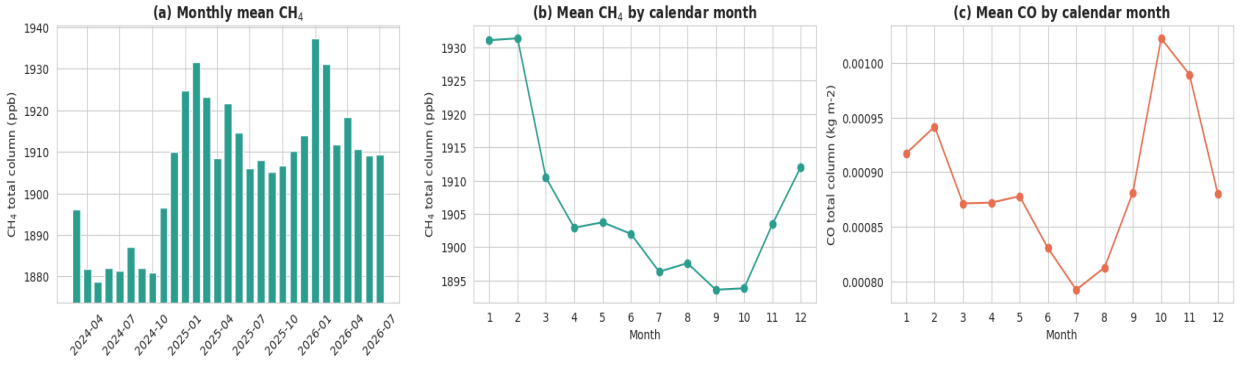


Fig 6. Monthly means and calendar-month seasonal cycle of CH_4 and CO (all years pooled).

Pearson correlations over a random sample of 30,000 cell-days (Fig 7) are modest: CH_4 correlates with CO at 0.37, with latitude at 0.26 and with calendar month at −0.33, and longitude is essentially uncorrelated with either gas. The CH_4–CO scatter is not a single cloud, which is consistent with the level shift noted above. These correlations are descriptive and do not imply a common source.

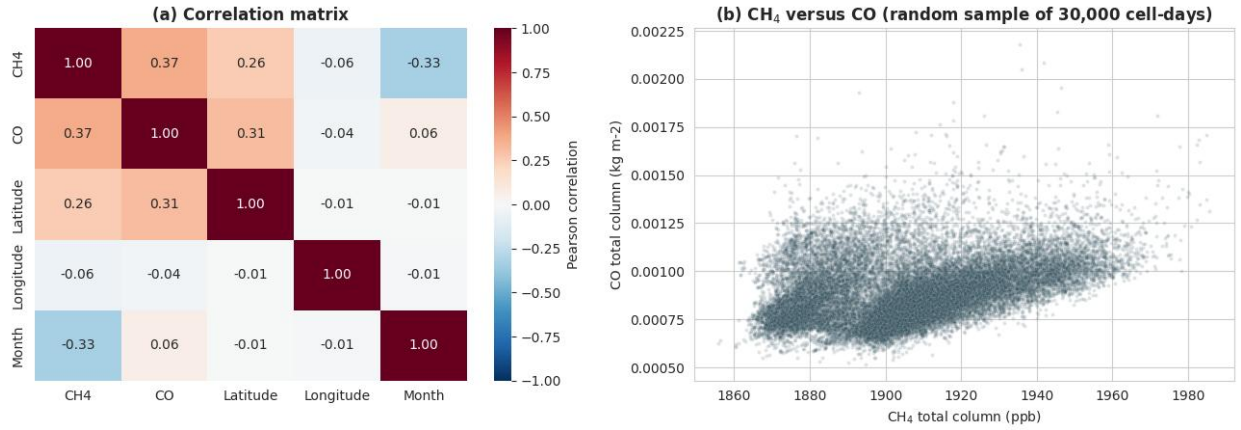


Fig 7. Correlation matrix and CH_4–CO scatter (descriptive only).

Area-weighted period means by district/city (Fig 8, Table 8) are highest in the Jakarta metropolitan area: the five Jakarta administrative cities, Bekasi City, Depok, Tangerang City, South Tangerang and Bekasi regency are the ten highest (1,927–1,933 ppb), and the lowest relative classes lie mainly in southern and eastern Java. The pattern shows a gradient from the dense northwestern metropolitan area to the rest of the island. It describes spatial association only and says nothing about emission sources.

TABLE 8. TEN REGIONS WITH THE HIGHEST OBSERVED PERIOD-MEAN CH_4

Region	Province	Mean (ppb)	Cells
Kota Adm. Jakarta Pusat	DKI Jakarta	1932.70	3
Kota Adm. Jakarta Selatan	DKI Jakarta	1932.15	7
Kota Adm. Jakarta Timur	DKI Jakarta	1931.76	6
Kota Adm. Jakarta Barat	DKI Jakarta	1930.39	4
Kota Bekasi	Jawa Barat	1929.68	6
Kota Adm. Jakarta Utara	DKI Jakarta	1929.33	7
Kota Depok	Jawa Barat	1928.57	7
Kota Tangerang	Banten	1928.56	5
Kota Tangerang Selatan	Banten	1928.24	7
Bekasi (regency)	Jawa Barat	1927.12	21

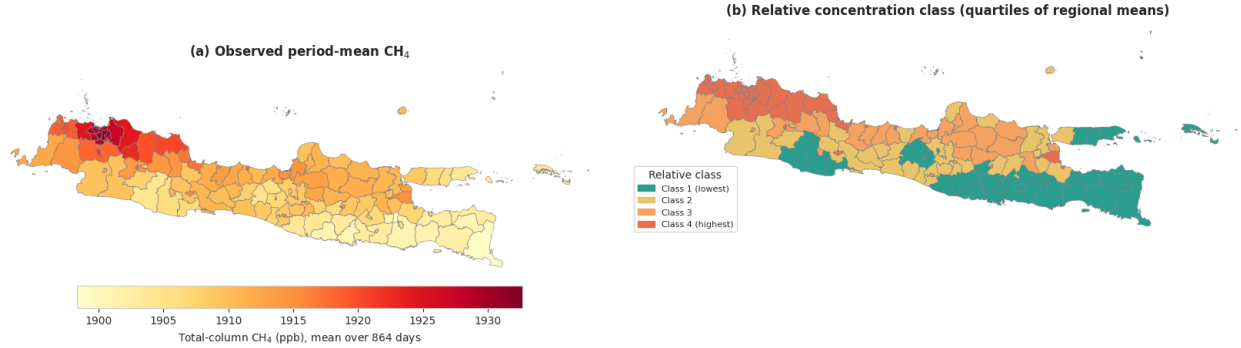


Fig 8. Observed (CAMS) period-mean CH_4 by district/city, area-weighted, with relative concentration classes.

B. *Forecast Skill in Physical Units*

Among the four ConvLSTM variants, the residual variant with CH_4 and CO inputs had the lowest validation RMSE (10.94 ppb, against 10.98 ppb for CH_4 only, 11.17 ppb for CH_4 + CO and 12.67 ppb for CH_4 + CO + season) and was selected before the test block was examined; the difference from the CH_4-only variant is small. Table 9 gives the test-period results. The selected model reached an MAE of 7.15 ppb (95% CI 6.48–7.84), an RMSE of 9.89 ppb (9.05–10.73), a bias of −0.80 ppb, a pooled R^2 of 0.618 (0.565–0.660) and an anomaly R^2 of 0.428. Persistence reached 7.43 ppb, 10.35 ppb, 0.582 and 0.373. The MSE skill score of the selected model is 0.087, and the interval of its MAE difference to persistence, −0.47 to −0.11 ppb, excludes zero, so the ConvLSTM improves modestly on persistence in this test block. The gain of about 0.3 ppb is small relative to an error of about 7.4 ppb.

TABLE 9. TEST-PERIOD PERFORMANCE IN PHYSICAL UNITS (PPB), ONE-DAY HORIZON, 130 TEST DAYS (NEURAL ROWS ARE THREE-RUN ENSEMBLES; SELECTED VARIANT IN BOLD)

Model	MAE [95% CI]	RMSE	Bias	R^2	Anomaly R^2	MSESS	ΔMAE vs persistence [95% CI]	MAE of single runs
Persistence	7.43 [6.65, 8.23]	10.35	−0.06	0.582	0.373	0.000	–	–
Moving average (7 d)	9.40 [8.07, 10.56]	12.32	0.25	0.408	0.112	−0.418	[0.99, 2.83]	–
Climatology (train mean)	12.79 [10.76, 15.05]	16.81	−10.57	−0.102	−0.653	−1.638	[3.72, 7.21]	–
Seasonal naive (t−365)	15.06 [12.41, 17.81]	19.64	4.06	−0.505	−1.257	−2.601	[5.27, 10.25]	–
Ridge regression	7.02 [6.32, 7.68]	9.61	−0.31	0.640	0.460	0.138	[−0.67, −0.17]	–
Gradient boosting	7.52 [6.69, 8.31]	10.18	0.10	0.596	0.394	0.032	[−0.37, 0.66]	–
ConvLSTM (CH_4 only)	7.12 [6.43, 7.82]	9.68	−0.01	0.634	0.451	0.125	[−0.53, −0.07]	7.41 ± 0.10
ConvLSTM (CH_4 + CO)	7.51 [6.77, 8.23]	10.06	0.71	0.605	0.408	0.055	[−0.13, 0.32]	8.02 ± 0.17
ConvLSTM (CH_4 + CO + season)	8.86 [8.06, 9.47]	11.32	1.65	0.500	0.250	−0.196	[0.85, 1.81]	9.90 ± 1.84

Model	MAE [95% CI]	RMSE	Bias	R^2	Anomaly R^2	MSESS	ΔMAE vs persistence [95% CI]	MAE of single runs
ConvLSTM residual (CH_4 + CO)	**7.15 [6.48, 7.84]**	**9.89**	**−0.80**	**0.618**	**0.428**	**0.087**	**[−0.47, −0.11]**	**7.54 ± 0.20**
CNN (non-recurrent, CH_4 + CO)	7.30 [6.65, 7.95]	9.80	0.45	0.625	0.437	0.102	[−0.39, 0.15]	7.87 ± 0.07

The ConvLSTM is not uniquely best. Ridge regression obtained the lowest MAE (7.02 ppb), the highest pooled R^2 (0.640) and the highest anomaly R^2 (0.460), with an MSESS of 0.138. The CH_4-only ConvLSTM (MAE 7.12 ppb, R^2 0.634) and the non-recurrent CNN (MAE 7.30 ppb, R^2 0.625; MAE difference to persistence −0.39 to 0.15 ppb, which includes zero) are close, and the confidence intervals of the leading models overlap broadly. The data therefore do not show that recurrence is needed for next-day CH_4 forecasting at this resolution: most of the skill comes from persistence of the field, and a linear model captures part of the remaining gain. Gradient boosting did not beat persistence, and the moving average, climatology and seasonal-naive forecasts were clearly worse (MAE 9.40, 12.79 and 15.06 ppb). Climatology underestimated the test period by 10.6 ppb, which is consistent with the rise in level after late 2024.

These results should be read in light of the amount of data. The training set contains 583 target days from a record of about 2.4 years, for a network with roughly 67,000 parameters and a field that is already highly persistent from one day to the next. In this setting a small margin over persistence and no margin over a linear model is a plausible outcome, and the benchmark is useful because it quantifies that outcome. Whether longer records, which will accumulate as the CAMS archive grows, or meteorological inputs change the ranking was not tested here.

Adding CO to the CH_4 input did not improve the standard ConvLSTM (MAE 7.51 against 7.12 ppb), and adding seasonal encodings degraded it (MAE 8.86 ppb, with a run-to-run SD of 1.84 ppb). CO and cyclic seasonal terms are therefore not shown to add information. The residual formulation reduced the penalty of the CO input (7.15 against 7.51 ppb) and gave the best validation score, which is why it was selected. Run-to-run variability matters for the headline result: single runs of the selected variant had a mean MAE of 7.54 ± 0.20 ppb and a mean R^2 of 0.587 ± 0.022, marginally worse than persistence (7.43 ppb, 0.582), and the advantage in Table 9 depends on averaging the three runs.

Fig 9 shows that the domain-mean forecast follows the observed test-period series closely and that cell-level forecasts scatter around the 1:1 line; the learning curves fall quickly and then flatten. The forecasts regress towards the mean: the bias of the selected model is +4.17 ppb in the lowest observed tercile and −3.67 ppb in the highest (Table 11).

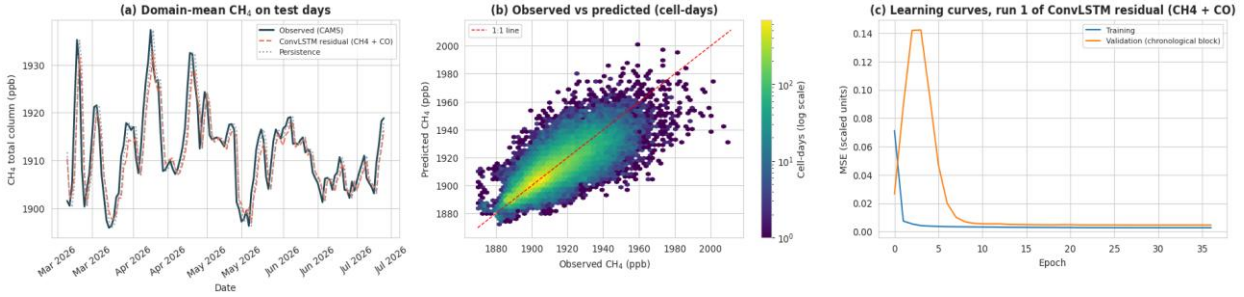


Fig 9. Forecast behaviour on the test period: (a) domain mean, (b) observed versus predicted cell-days, (c) learning curves of the first run.

C. *Extreme Events and Error Structure*

Extremes, defined as cell-days above 1,939.5 ppb, make up 6.1% of the test cell-days and are poorly forecast (Table 10). The selected ConvLSTM detected 31.0% of them (POD 0.310) with a false-alarm ratio of 0.533 and a CSI of 0.229, whereas persistence detected 40.8% (FAR 0.594, CSI 0.256). All models underestimated the intensity of observed extremes, the selected model by 14.0 ppb on average and persistence by 11.1 ppb. The smooth ConvLSTM output therefore does not help with high-concentration events, and the screening results should not be read as event forecasts.

TABLE 10. EXTREME-EVENT SKILL (THRESHOLD 1,939.5 PPB)

Model	POD	FAR	CSI	Intensity bias (ppb)
Persistence	0.408	0.594	0.256	−11.12
Moving average (7 d)	0.207	0.471	0.175	−19.31
Climatology (train mean)	0.000	–	0.000	−38.92
Seasonal naive (t−365)	0.194	0.833	0.098	−24.94
Ridge regression	0.258	0.497	0.206	−15.35
Gradient boosting	0.227	0.464	0.190	−16.95
ConvLSTM (CH_4 only)	0.259	0.490	0.207	−15.29
ConvLSTM (CH_4 + CO)	0.283	0.515	0.218	−14.67
ConvLSTM (+ season)	0.156	0.440	0.139	−18.30
ConvLSTM residual	**0.310**	**0.533**	**0.229**	**−14.03**
CNN (non-recurrent)	0.252	0.476	0.205	−15.32

Errors by calendar month and by observed-concentration tercile are given in Table 11. The selected model improves on persistence in every month of the test block, by 0.05–0.55 ppb in MAE, but the test block covers only March–July, so seasonal skill over a full cycle cannot be assessed. By tercile, the model is worse than persistence in the lowest tercile (MAE 5.22 against 4.65 ppb) and better in the middle and highest terciles.

TABLE 11. TEST-PERIOD ERROR BY CALENDAR MONTH AND OBSERVED-CH_4 TERCILE: PERSISTENCE (P) AND SELECTED CONVLSTM (M)

Stratum	N	MAE P	MAE M	RMSE P	RMSE M	Bias P	Bias M
March	27 days	8.53	7.98	11.24	10.50	−0.17	−0.74
April	30 days	8.53	8.24	11.53	11.09	−0.27	−1.60
May	31 days	7.35	7.08	10.28	9.90	0.32	−0.35
June	30 days	6.28	6.11	9.30	8.93	0.36	−0.10
July	12 days	5.34	5.29	7.38	7.19	−1.27	−1.78
Lowest tercile	25,313 cell-days	4.65	5.22	7.44	7.59	2.86	4.17
Middle tercile	185,658 cell-days	5.45	5.14	7.83	7.25	1.96	2.02
Highest tercile	225,829 cell-days	9.38	9.01	12.27	11.81	−2.04	−3.67

The spatial error structure (Fig 10) is not uniform. MAE is largest, at roughly 12–16 ppb, in a belt over the northwest around the Jakarta metropolitan area and western Java and in a scattered patch in central Java, and it is much smaller offshore and in the south. The error maps of the ConvLSTM and of persistence are visually almost identical. The bias map is small over most of the island, with larger values in a few cells at the domain corners that may be edge effects of the convolutions. Because the error is largest where concentrations are highest, forecast uncertainty is greatest exactly in the areas of most interest for screening.

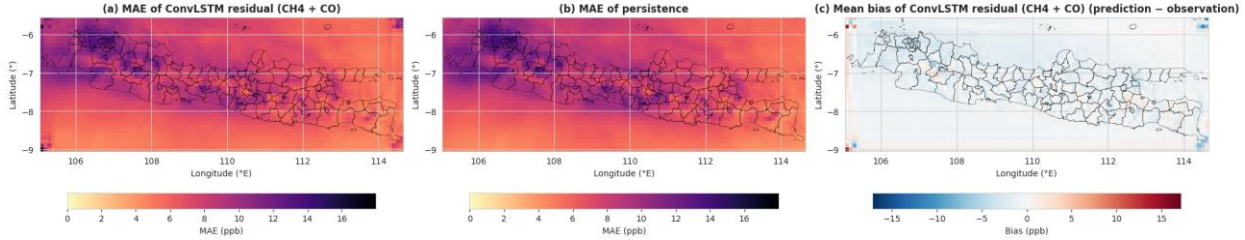


Fig 10. Spatial distribution of test-period forecast error per grid cell: (a) MAE of the selected ConvLSTM, (b) MAE of persistence, (c) mean bias of the selected ConvLSTM.

D. Input Window, Forecast Horizon and Rolling-Origin Evaluation

Table 12 and Fig 11 show an exploratory sensitivity analysis (one seed per setting); conclusions on the main configuration rest on Table 9, which uses three runs. At the one-day horizon only the 3-day window matched persistence (MAE 7.45 against 7.43 ppb); the 7-day and 14-day windows were worse (MAE 11.30 and 8.19 ppb). The 7-day, one-day configuration is the main experiment, yet this single run (seed 42) gave a much higher error than the three-run ensemble in Table 9 (11.30 against 7.15 ppb) and stopped after nine epochs, which shows that single runs of this model can end in a poor solution. The 7-day, one-day cell is reported as obtained and is superseded by Table 9 for that configuration, so the table is read as an indication of optimisation variance and of the behaviour across horizons, not as a ranking of window lengths. At the 3-day and 7-day horizons, where persistence itself is much less accurate (MAE 12.00 and 12.70 ppb), most settings improved on persistence in MSE terms (MSESS 0.15–0.26; the 3-day window at the 3-day horizon was the exception at −0.014), although pooled R^2 stayed low (−0.03 to 0.25).

TABLE 12. EXPLORATORY SENSITIVITY OF THE SELECTED CONVLSTM TO INPUT WINDOW AND FORECAST HORIZON (ONE SEED PER SETTING, TEST PERIOD)

Window (d)	Horizon (d)	MAE model	MAE persist.	RMSE model	RMSE persist.	R^2 model	R^2 persist.	MSESS
3	1	7.447	7.433	10.341	10.348	0.583	0.582	0.001
3	3	12.118	12.002	16.212	16.096	−0.026	−0.011	−0.014
3	7	11.268	12.697	14.578	16.610	0.171	−0.077	0.230
7	1	11.300	7.433	14.067	10.348	0.228	0.582	−0.848
7	3	11.177	12.002	14.530	16.096	0.176	−0.011	0.185
7	7	11.334	12.697	14.494	16.610	0.180	−0.077	0.239
14	1	8.186	7.433	11.140	10.348	0.516	0.582	−0.159
14	3	10.816	12.002	13.877	16.096	0.249	−0.011	0.257
14	7	12.204	12.697	15.315	16.610	0.085	−0.077	0.150

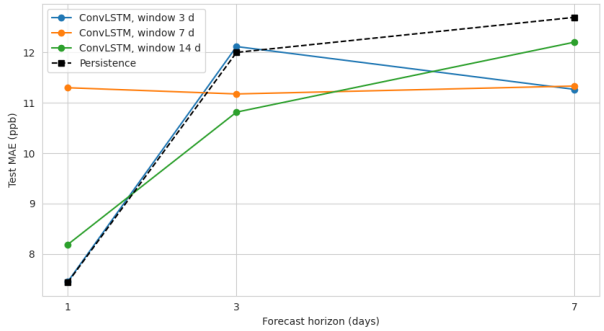


Fig 11. Test MAE versus forecast horizon for different input windows.

The rolling-origin evaluation (Table 13) repeats training, validation and testing on three consecutive 86-day test blocks. The selected model beat persistence in MSE in folds 2 and 3 (MSESS 0.034 and 0.103) but not in fold 1 (−0.193), and it had a lower MAE than persistence only in fold 3 (6.77 against 6.94 ppb). The result for the main test block (March–July 2026) should therefore not be assumed to hold for other periods.

TABLE 13. ROLLING-ORIGIN EVALUATION OF THE SELECTED CONVLSTM (ONE RUN PER FOLD)

Fold	Test period	Train days	MAE model	MAE persist.	RMSE model	RMSE persist.	R^2 model	R^2 persist.	MSESS
1	2025-10-28 to 2026-01-21	499	8.520	7.478	11.557	10.579	0.561	0.632	−0.193
2	2026-01-22 to 2026-04-17	585	8.810	8.765	11.563	11.767	0.706	0.695	0.034
3	2026-04-18 to 2026-07-12	671	6.772	6.942	9.467	9.993	0.620	0.577	0.103

E. Regional Relative Priority Screening

Aggregating the test-period forecasts of the selected model to the 119 regions gives Fig 12 and Table 14. The 15 regions with the highest priority index (1.42–2.47) are all in the Jakarta metropolitan area and adjacent parts of West Java and Banten: the five Jakarta administrative cities, Depok, Bekasi City and regency, South Tangerang, Tangerang City and regency, Purwakarta, Karawang, Bogor City and Subang. All 15 fall in the highest relative concentration class and the highest relative priority class. Twelve of the 15 are flagged low-confidence because their regional MAE (9.7–12.4 ppb) is in the worst quartile, and their forecast means are 1–4 ppb below the observed means, in line with the underestimation of high values.

TABLE 14. THE 15 REGIONS WITH THE HIGHEST RELATIVE PRIORITY INDEX (TEST-PERIOD FORECASTS, AREA-WEIGHTED)

Region	Province	Obs. mean (ppb)	Forecast mean (ppb)	Regional MAE (ppb)	Share of days in top quartile	Priority index	Low confidence
Kota Adm. Jakarta Selatan	DKI Jakarta	1947.61	1943.49	11.16	0.938	2.468	Yes
Kota Adm. Jakarta Pusat	DKI Jakarta	1946.72	1942.42	12.43	0.938	2.413	Yes
Kota Adm. Jakarta Timur	DKI Jakarta	1945.70	1941.59	11.68	0.923	2.344	Yes
Kota Depok	Jawa Barat	1944.01	1941.31	10.39	0.900	2.290	Yes
Kota Adm. Jakarta Barat	DKI Jakarta	1944.22	1940.23	11.96	0.892	2.221	Yes
Kota Bekasi	Jawa Barat	1943.10	1939.73	11.16	0.900	2.209	Yes
Kota Tangerang Selatan	Banten	1943.72	1939.86	11.73	0.885	2.189	Yes
Kota Tangerang	Banten	1942.68	1939.32	12.03	0.877	2.148	Yes
Kota Adm. Jakarta Utara	DKI Jakarta	1942.03	1938.82	12.17	0.885	2.136	Yes
Bekasi (regency)	Jawa Barat	1940.16	1937.76	9.68	0.900	2.108	Yes
Tangerang (regency)	Banten	1938.83	1936.27	12.06	0.846	1.939	Yes
Purwakarta	Jawa Barat	1937.93	1936.50	8.74	0.838	1.937	No
Karawang	Jawa Barat	1937.12	1934.78	8.08	0.869	1.902	No
Kota Bogor	Jawa Barat	1936.93	1934.96	10.93	0.754	1.714	Yes
Subang	Jawa Barat	1931.19	1929.07	7.75	0.762	1.424	No

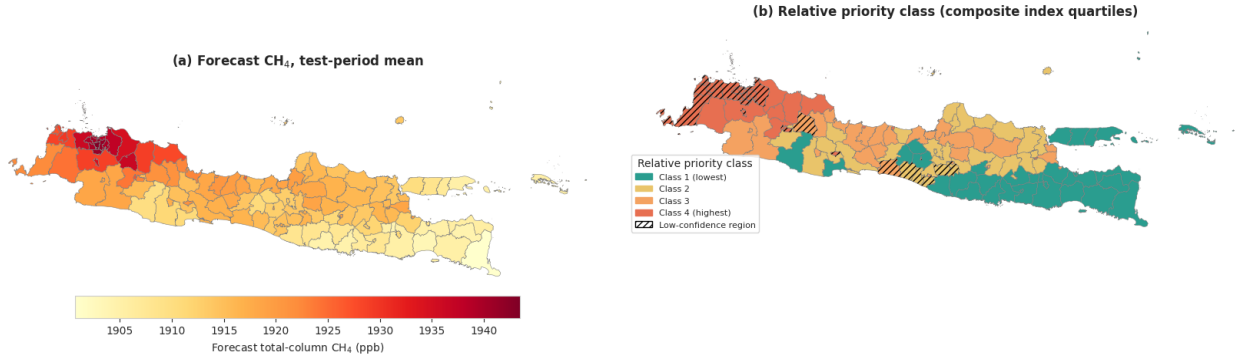


Fig 12. Regional screening from the selected ConvLSTM forecasts: (a) forecast CH_4, test-period mean; (b) relative priority class (composite-index quartiles; hatched = regional MAE in the worst quartile).

The forecast regional ranking agrees closely with the observed test-period ranking (Table 15: Spearman ρ = 0.995, the same relative class for 93.3% of regions, top-quartile Jaccard overlap 0.933). Persistence agrees even better (ρ = 1.000, 100% class agreement), which is expected at a one-day horizon because the persistence forecast is the observed field shifted by one day, and the static training climatology is also close (ρ = 0.983, 83.2% class agreement, overlap 0.758). The mean absolute error of the regional mean is 1.39 ppb for the ConvLSTM, 0.12 ppb for persistence and 12.5 ppb for climatology. The screening therefore largely reproduces

a persistent spatial pattern, and the ConvLSTM adds no ranking information beyond the persistence map; its advantage over climatology lies mainly in tracking the current level, not in re-ranking regions.

TABLE 15. AGREEMENT OF REGIONAL RANKINGS WITH THE OBSERVED TEST-PERIOD RANKING

Ranking source	Spearman ρ	Same class	Top-quartile overlap	MAE of regional mean (ppb)
ConvLSTM residual (CH_4 + CO)	0.995	0.933	0.933	1.386
Persistence	1.000	1.000	1.000	0.121
Climatology (train mean)	0.983	0.832	0.758	12.527

Fig 13 compares observed and forecast regional values for the last test day (12 July 2026). The two maps are visually similar, and the difference map is small for most regions. Fig 14 shows that forecast and observed regional means lie close to the 1:1 line, with low-confidence regions concentrated at the high end, and that the regional MAE of the 15 highest-ranked regions (about 8–12 ppb) is larger than most differences among their forecast means (1,929–1,943 ppb). The ordering within the top class is therefore not robust, and the class membership is the more defensible output.

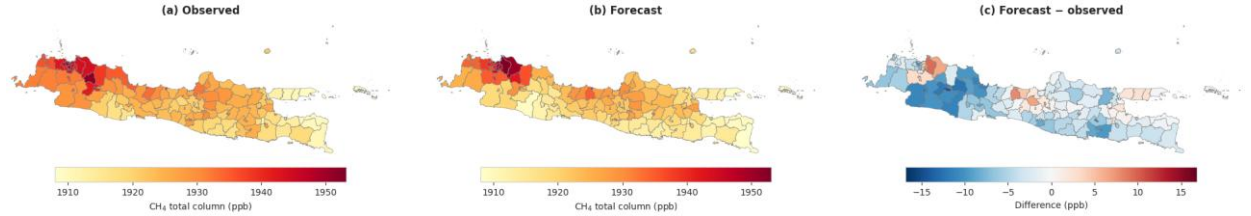


Fig 13. Regional forecast versus observation for the last test day (2026-07-12), area-weighted.

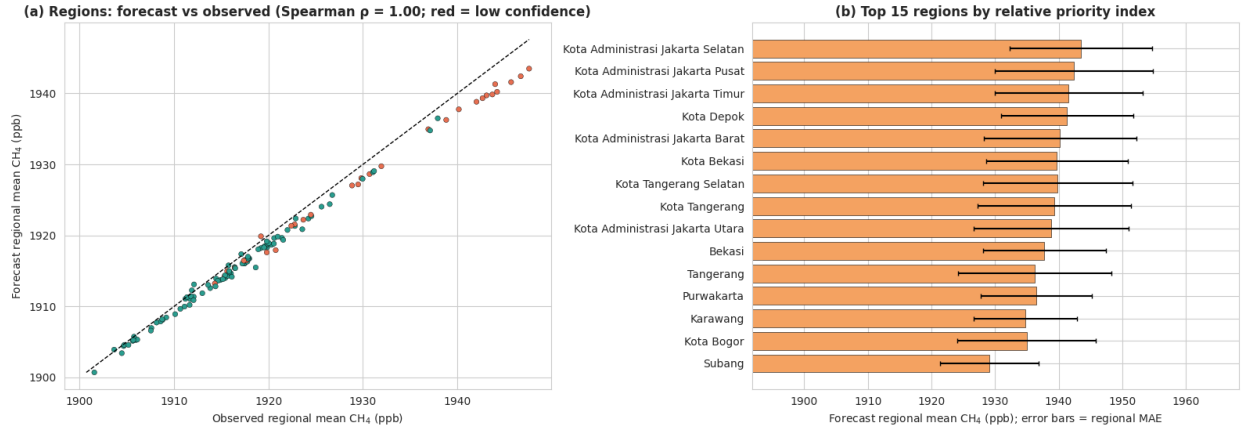


Fig 14. Agreement between forecast and observed regional means, and the 15 highest-ranked regions with their regional MAE.

V. LIMITATIONS

Data and provenance. The fields are CAMS model forecasts, not direct satellite or in-situ measurements, and "observed" in this paper means the CAMS target field. No independent validation against field measurements, TROPOMI retrievals or emission inventories was performed, neighbouring cells of a model field are spatially correlated, and the series may contain product-related shifts because the system is upgraded about once a year. The record is short (about 2.4 years), the test block covers only March–July, so seasonal skill and behaviour during the burning season cannot be assessed, and the level shift of late 2024 confounds seasonal estimates.

Modelling and evaluation. The main task is one-day-ahead forecasting; the window and horizon sensitivity uses one run per setting and is unstable, as the 7-day, one-day cell shows. The study is a gas-only benchmark by design: no meteorological drivers (the product also provides parameters such as humidity, precipitation and cloud water [4]), fire activity, land cover, traffic or industrial layers are included, so the model cannot separate persistence from emission dynamics, and the ablation tests only CO and seasonal terms. The training set has only 583 target days. Pooled R^2 is dominated by the persistent spatial pattern, and the anomaly R^2, the skill score against persistence and the ranking comparison with climatology should be used to judge added value. The ConvLSTM output is smooth and underestimates extremes. Uncertainty is quantified only through run-to-run variability and block-bootstrap intervals, not through predictive distributions. Hyper-parameters were not tuned beyond the small ablation, results are for one architecture family, and the global warming potential of CH_4 over short horizons [1] makes meteorological drivers a priority for future work.

Screening and interpretation. Classes are quartile ranks of the studied regions, so each class contains about a quarter of the regions by construction, and the class counts carry no information about risk. A class label does not correspond to a regulatory exceedance, health impact or emission intensity. Spatial association with metropolitan areas is described, not explained: no source attribution was attempted, and any link to transport, industry, biomass burning, waste or energy systems would require independent data and a formal decision framework. Population exposure is not part of the index, which is a rule of thumb for deciding where additional monitoring is most informative.

Items concerning provenance, units, resolution, class interpretation and scope are addressed by explanation in this text. Meteorological and fire predictors, longer records, multi-gas targets, predictive uncertainty and independent validation require new data and retraining.

VI. Conclusion

This study presents a preliminary spatial forecasting experiment and a relative regional screening for CH_4 over Java Island, using CAMS model forecast fields and a ConvLSTM evaluated chronologically in physical units against persistence, climatology, regression and CNN baselines. The selected residual ConvLSTM with CH_4 and CO inputs reached an MAE of 7.15 ppb and an RMSE of 9.89 ppb on 130 test days, a modest but statistically distinguishable improvement over persistence (MSESS 0.087). The improvement is not uniquely attributable to the ConvLSTM: ridge regression performed as well or better, single ConvLSTM runs did not beat persistence on average, the first rolling-origin fold was worse than persistence, and extreme events were underestimated. With a training set of only 583 days, a small margin over persistence and over a linear model is an informative benchmark and not by itself a failure of the approach, although this explanation was not tested.

The regional classes are relative priorities for further monitoring and investigation. They concentrate in the Jakarta metropolitan area and adjacent West Java and Banten, but the regional ranking largely reproduces the persistent spatial pattern, forecast error is largest in the highest-ranked regions, and the classes are not validated risk categories. The study makes no statement about emission sources, emission intensity or the effectiveness of energy-transition or waste-management policies. Independent validation, meteorological and source-related predictors, longer records, multi-step forecasting and multi-gas targets are the natural next steps.

Conflict of Interest Statement

The authors declare that there is no conflict of interest, whether financial or non-financial, that could influence the results of this research, the writing, or the decision to publish this article.

Acknowledgments

This research did not receive any specific funding from any grant-making body, whether from the public, commercial, or non-profit sectors.

References

[1] Intergovernmental Panel on Climate Change (IPCC), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report. Cambridge: Cambridge University Press, 2021.

[2] J. P. Veefkind et al., "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," Remote Sensing of Environment, vol. 120, pp. 70–83, 2012.

[3] A. Lorente et al., "Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements," Atmospheric Measurement Techniques, vol. 14, no. 1, pp. 665–684, 2021.

[4] Copernicus Atmosphere Monitoring Service (CAMS), "CAMS global greenhouse gas forecasts," Atmosphere Data Store, European Centre for Medium-Range Weather Forecasts. [Online]. Available: https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-forecasts

[5] M. H. Nazari and A. K. Biswas, "Decentralized P2P trading based on blockchain for retail electricity markets," in 2024 IEEE Conf. on Innovative Smart Grid Technologies (ISGT), Washington, DC, USA, 2024, pp. 1–5.

[6] Erdiwansyah, A. Gani, R. Mamat, Bahagia, M. Nizar, S. Yana, M. H. Mat Yasin, Muhibbuddin, and S. M. Rosdi, "Prospects for renewable energy sources from biomass waste in Indonesia," Case Studies in Chemical and Environmental Engineering, vol. 10, Art. 100880, 2024.

[7] P. Sadavarte et al., "Methane emissions from superemitting coal mines in Australia quantified using TROPOMI satellite observations," Environmental Science & Technology, vol. 55, no. 24, pp. 16573–16580, 2021.

[8] M. K. Sha et al., "Validation of methane and carbon monoxide from Sentinel-5 Precursor using TCCON and NDACC-IRWG stations," Atmospheric Measurement Techniques, vol. 14, no. 9, pp. 6249–6304, 2021.

[9] H. Lindqvist et al., "Evaluation of Sentinel-5P TROPOMI methane observations at northern high latitudes," Remote Sensing, vol. 16, no. 16, Art. 2979, 2024.

[10] M. Sarafanov, E. Kazakov, N. O. Nikitin, and A. V. Kalyuzhnaya, "A machine learning approach for remote sensing data gap-filling with open-source implementation: an example regarding land surface temperature, surface albedo and NDVI," Remote Sensing, vol. 12, no. 23, Art. 3865, 2020.

[11] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: a machine learning approach for

precipitation nowcasting," in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 802–810.

[12] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Deep learning for precipitation nowcasting: a benchmark and a new model," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[13] P. Muthukumar, K. Nagrecha, D. Comer, C. F. Calvert, N. Amini, J. Holm, and M. Pourhomayoun, "PM2.5 air pollution prediction through deep learning using multisource meteorological, wildfire, and heat data," Atmosphere, vol. 13, no. 5, Art. 822, 2022.

[14] T. Khan, "Forecasting smog events using ConvLSTM: a spatio-temporal approach for aerosol index prediction in South Asia," arXiv preprint arXiv:2508.13891, 2025, doi: 10.48550/arXiv.2508.13891.

[15] S. N. Zildjian, "Spatio-temporal prediction of wind energy potential in the Indonesian region using a ConvLSTM model based on meteorological reanalysis data," Jurnal Komputer, Informasi Teknologi, dan Elektro (KITEKTRO), vol. 11, no. 2, pp. 33–43, Aug. 2026.

[16] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in Proc. 27th Int. Joint Conf. on Artificial Intelligence (IJCAI), 2018, pp. 3634–3640.

[17] Z. Zhao, Z. Wang, G. Zhao, and J. Zhao, "A new strong convective precipitation forecasting method based on attention mechanism and spatio-temporal reasoning," Scientific Reports, vol. 14, 2024, doi: 10.1038/s41598-024-68951-1.

[18] K. Jordahl et al., "GeoPandas: Python tools for geographic data," 2020. [Software]. Available: https://doi.org/10.5281/zenodo.3946761

[19] F. Pedregosa et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in Proc. 32nd Int. Conf. on Machine Learning (ICML), 2015, pp. 448–456.

[21] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Int. Conf. on Learning Representations (ICLR), 2015.

[22] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in Int. Conf. on Learning Representations (ICLR), 2018.

[23] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," Geoscientific Model Development, vol. 7, pp. 1247–1250, 2014.

[24] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Research, vol. 30, no. 1, pp. 79–82, 2005.

[25] A. H. Murphy, "Skill scores based on the mean square error and their relationships to the correlation coefficient," Monthly Weather Review, vol. 116, no. 12, pp. 2417–2424, 1988.

[26] H. R. Künsch, "The jackknife and the bootstrap for general stationary observations," The Annals of Statistics, vol. 17, no. 3, pp. 1217–1241, 1989.

[27] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," International Journal of Forecasting, vol. 16, no. 4, pp. 437–450, 2000.

[28] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard, Thematic Cartography and Geovisualization, 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

[29] B. W. Silverman, Density Estimation for Statistics and Data Analysis. London: Chapman & Hall, 1986.

[30] N. Nechita-Banda, M. Krol, G. R. van der Werf, J. W. Kaiser, S. Pandey, V. Huijnen, C. Clerbaux, P. Coheur, M. N. Deeter, and T. Röckmann, "Monitoring emissions from the 2015 Indonesian fires using CO satellite data," Philosophical Transactions of the Royal Society B, vol. 373, no. 1760, Art. 20170307, 2018.

[31] G. R. van der Werf et al., "Global fire emissions estimates during 1997–2016," Earth System Science Data, vol. 9, no. 2, pp. 697–720, 2017.