Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 4071~4079
e-ISSN: 2548-6861

Hybrid VMD-CNN1D Framework: Evaluating Decomposition's Contribution to PM2.5 Prediction

Dwi Yuwono[1*], Arya Adhyaksa Waskita[2], Tukiyat[3**]**
* Program Studi Teknik Informatika, Pascasarjana, Universitas Pamulang | Badan Meteorologi, Klimatologi, dan Geofisika
** Program Studi Teknik Informatika, Pascasarjana, Universitas Pamulang
dwi.yuwono@bmkg.go.id [1], aawaskita@unpam.ac.id [2], dimastuky@gmail.com [3]

Article Info

Article history:

Received 2026-07-24
Revised 2026-08-01
Accepted 2026-08-13

Keyword:

Air Quality,
Bias Correction,
Convolutional Neural Network,
PM2.5 Prediction,
Variational Mode Decomposition.

ABSTRACT

Fine particulate matter (PM2.5) pollution in Jakarta reached an annual average of 37.3 µg/m³ in 2023, 7.4 times the WHO threshold, causing over 10,000 premature deaths annually. Accurate short-term prediction is essential for early-warning systems, yet two gaps remain common in decomposition-based deep learning literature: Variational Mode Decomposition (VMD) parameters are rarely selected via systematic sensitivity analysis, and prediction bias is rarely corrected explicitly. This study addresses both gaps using 32,144 PM2.5 observations from Jakarta (2015-2025). A grid search over 20 parameter combinations identified K=6, alpha=500 as optimal (reconstruction error 1.88%). Each of six IMFs was predicted independently using CNN1D, reconstructed additively, and bias-corrected using validation-set mean bias error. Decomposition, not model complexity, drove accuracy: a non-decomposed baseline reached only R^2=0.4302, versus R^2=0.9151 (RMSE=4.83 µg/m³) for the proposed framework. We further validated fixed- and adaptive-parameter variants (VMD, AVMD) across 5 independent runs. VMD-CNN1D achieved R^2=0.9177±0.0173, RMSE=4.7346±0.5064, while AVMD-CNN1D (K=7 selected consistently) achieved R^2=0.9238±0.0089, RMSE=4.5388±0.2649. Diebold-Mariano tests showed AVMD outperforming VMD in 4 of 5 runs ($p<0.0001$) with lower variance, though a paired t-test across runs was not significant ($p=0.684$). Ablation identified the lowest-frequency IMF as most critical, consistent with Jakarta's dry-season and land-fire pollution patterns, while residual analysis revealed heteroscedasticity as a limitation. These findings show that rigorous parameter selection, bias correction, and multi-run validation, not architectural complexity alone, make decomposition-based deep learning reliable for PM2.5 prediction, offering a reproducibility-aware baseline for a future Jakarta air-quality early-warning system.

I. PENDAHULUAN

Kualitas udara merupakan salah satu indikator terpenting bagi kesehatan lingkungan perkotaan, dan partikulat halus berdiameter aerodinamis 2,5 mikrometer atau kurang (PM2,5) dianggap sebagai komponen paling berbahaya karena mampu menembus jauh ke dalam sistem pernapasan. Konsentrasi rata-rata tahunan PM2,5 di Jakarta mencapai 37,3 µg/m³ pada tahun 2023 (7,4 kali nilai ambang batas yang ditetapkan *World Health Organization*) dan diperkirakan menyebabkan lebih dari 10.000 kematian dini per tahun [1], [2]. Prediksi jangka pendek konsentrasi PM2,5 yang akurat karenanya menjadi prasyarat bagi sistem peringatan dini dan kebijakan lingkungan yang efektif.

Deret waktu PM2,5 bersifat non-stasioner, non-linier, dan mengandung komponen multi-skala yang timbul dari siklus harian, variasi musiman, dan kejadian episodik seperti kebakaran lahan. Model statistik konvensional seperti ARIMA kesulitan menangkap kompleksitas ini, dan arsitektur *deep learning* yang diterapkan langsung pada sinyal PM2,5 mentah tetap menghadapi masalah non-stasioneritas yang sama. Dekomposisi sinyal sebelum

pemodelan karenanya menjadi strategi pra-pemrosesan yang umum digunakan. *Variational Mode Decomposition* (VMD) [3] memformulasikan dekomposisi sebagai masalah optimasi variasional yang diselesaikan secara simultan untuk seluruh mode, dan berbeda dari pendahulunya, *Empirical Mode Decomposition* (EMD) [4], [5], VMD secara inheren bebas dari *mode mixing* tanpa memerlukan penambahan derau buatan.

Beberapa penelitian terkini mengombinasikan VMD dengan prediktor *deep learning* berbasis *recurrent* atau hibrida: VMD–BiGRU untuk kota-kota di Arab Saudi [6], hibrida MIC-CEEMDAN-CNN-BiGRU untuk Beijing [7], VMD–GAT–BiLSTM untuk data kualitas udara *spatiotemporal* [8], serta strategi dekomposisi sekunder berbasis *clustering* [9], [10]. *Baseline* CNN-LSTM tanpa dekomposisi juga telah dilaporkan [11]. Namun, dua kesenjangan metodologis tetap umum dijumpai pada literatur tersebut. Pertama, parameter VMD (jumlah mode K dan konstanta *bandwidth alpha*) jarang ditentukan melalui analisis sensitivitas sistematis yang spesifik terhadap dataset, meskipun transfer parameter antar-dataset diketahui tidak dapat diandalkan [12]. Kedua, bias prediksi sistematis, yaitu kecenderungan model untuk secara konsisten melebih-lebihkan atau meremehkan nilai aktual, jarang dievaluasi atau dikoreksi secara eksplisit, meskipun bias tersebut dapat tersembunyi di balik metrik agregat yang tampak memadai seperti R^2 [13].

Kedua kesenjangan ini dapat dinyatakan secara eksplisit sebagai berikut. Kontribusi 1: parameter VMD (K dan alpha) jarang dievaluasi melalui reconstruction error (RE) yang dilaporkan secara eksplisit sebagai kriteria independen dari performa model prediksi hilir. Studi rujukan [8]–[10] memang menerapkan optimasi parameter VMD — [8] melalui pengujian K=3 hingga K=7 berdasarkan akurasi prediksi akhir, dan [9]–[10] melalui Whale Optimization Algorithm (WOA) yang mengoptimalkan K dan alpha menggunakan envelope entropy sebagai fungsi fitness — namun pada ketiganya kualitas dekomposisi tidak dievaluasi secara terpisah dari performa model prediksi atau dari metrik entropy, sehingga tidak ada ambang batas kualitas rekonstruksi (mis. RE<3%) yang dilaporkan secara independen sebelum tahap pemodelan. Studi rujukan [6] berfokus pada optimasi hyperparameter BiGRU melalui Bayesian Optimization tanpa menyentuh parameter VMD, sementara studi rujukan [7] menggunakan CEEMDAN dan bukan VMD, sehingga keduanya tidak sepenuhnya sebanding pada dimensi ini. Kontribusi 2: bias prediksi sistematis jarang dievaluasi atau dikoreksi menggunakan mekanisme yang dihitung khusus dari validation set terpisah — mean bias error atau prosedur koreksi bias yang eksplisit dan bebas dari kebocoran data tidak dilaporkan pada kelima studi rujukan [6]–[10], yang seluruhnya mengevaluasi model hanya menggunakan metrik agregat (R^2, RMSE, MAE, dan/atau MAPE) tanpa komponen bias terpisah, sehingga underestimation atau overestimation sistematis berpotensi tersembunyi di balik metrik yang tampak memadai.

Penelitian ini menjawab kedua kesenjangan tersebut dengan mengusulkan model VMD–CNN1D untuk prediksi PM2,5 secara univariat yang (1) menentukan parameter VMD melalui analisis sensitivitas *grid-search* yang sistematis berdasarkan *reconstruction error*, dan (2) menerapkan mekanisme koreksi bias yang dihitung khusus dari *validation set* untuk mencegah kebocoran data (*data leakage*) ke dalam evaluasi *test set*. Pendekatan univariat dipilih secara sengaja untuk membangun *baseline* pola temporal yang tervalidasi secara ketat sebelum memperkenalkan variabel meteorologi eksogen pada penelitian berikutnya. Model dievaluasi menggunakan enam metrik komplementer, yakni R^2, RMSE, MAE, MAPE, MBE, dan *Index of Agreement* [14], serta dibandingkan dengan *baseline* CNN1D tanpa dekomposisi dan varian VMD dengan K adaptif untuk mengisolasi kontribusi nyata dekomposisi sinyal terhadap akurasi prediksi.

II. METODE

A. Data

Dataset yang digunakan terdiri atas 32.144 observasi konsentrasi PM2,5 univariat yang direkam pada interval tiga jam oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk stasiun pemantauan di Jakarta, mencakup periode Januari 2015 hingga Desember 2025, dengan rentang konsentrasi 20,00 hingga 95,00 $\mu g/m^3$. BMKG dipilih sebagai sumber data karena statusnya sebagai lembaga pemerintah yang berwenang dalam pemantauan kualitas udara dan meteorologi di Indonesia. Hanya sinyal PM2,5 itu sendiri yang digunakan sebagai masukan; kovariat meteorologi sengaja tidak disertakan untuk membangun *baseline* pola temporal, konsisten dengan lingkup univariat penelitian ini.

Stasiun pemantauan berlokasi di kantor pusat BMKG dan diawasi oleh petugas monitoring selama 24 jam penuh, sehingga gangguan operasional yang lazim menyebabkan data hilang pada stasiun jarak jauh (seperti sensor mati, kegagalan kalibrasi, atau pemadaman listrik) dapat segera ditanggulangi, sehingga tidak ditemukan nilai hilang pada dataset yang digunakan. Sebagai langkah pencegahan, *pipeline* prapemrosesan tetap menyertakan mekanisme *forward-fill* diikuti *backward-fill* yang akan aktif otomatis apabila nilai hilang terdeteksi pada penggunaan data di masa mendatang. Deteksi dan penghapusan *outlier* secara statistik (mis. berbasis IQR atau *z-score*) sengaja tidak diterapkan: nilai konsentrasi PM2,5 yang sangat tinggi pada data historis Jakarta (misalnya yang bertepatan dengan kejadian kebakaran lahan regional atau kondisi atmosfer stabil pada musim kemarau) merepresentasikan kejadian polusi yang valid secara fisik, bukan derau instrumen, dan justru relevan bagi tujuan sistem peringatan dini. Normalisasi *Min–Max*

diterapkan secara terpisah pada setiap IMF hasil dekomposisi, dengan parameter normalisasi di-*fit* hanya pada partisi *training* untuk mencegah *data leakage* ke partisi *validation/test*.

B. Analisis Sensitivitas Parameter VMD

VMD mendekomposisi sinyal f(t) menjadi K *intrinsic mode function* (IMF) berpita terbatas dengan menyelesaikan masalah optimasi variasional yang meminimalkan jumlah *bandwidth* seluruh mode, dengan syarat bahwa seluruh mode dapat merekonstruksi sinyal asli secara akurat, diselesaikan melalui *Alternating Direction Method of Multipliers* [3]. *Bandwidth* dikendalikan oleh parameter konstanta *alpha*. Karena K dan *alpha* secara bersama-sama menentukan kualitas dekomposisi dan tidak ada pengaturan universal yang berlaku untuk semua dataset [12], dilakukan *grid search* terhadap K ∈ {3, 4, 5, 6} dan *alpha* ∈ {500, 1000, 2000, 3000, 5000} (total 20 kombinasi secara keseluruhan) yang dievaluasi menggunakan *reconstruction error* (RE), yaitu residual antara sinyal asli dan jumlah seluruh IMF, dinyatakan sebagai persentase dari rentang sinyal. Suatu kombinasi diterima apabila RE < 3%.

C. Arsitektur CNN1D dan Koreksi Bias

Pembagian data secara kronologis (80% *training*, 10% *validation*, 10% *test*) diterapkan pada setiap IMF secara independen untuk menjaga integritas temporal dan mencegah *data leakage*, setelah itu normalisasi *Min–Max* di-*fit* hanya pada data *training*. Setiap IMF dimodelkan secara independen menggunakan CNN1D dengan tiga blok konvolusi (*Conv1D–BatchNormalization* [15]–*MaxPooling1D*, masing-masing dengan 64, 128, dan 256 *filter*), diikuti oleh lapisan *flatten*, dua lapisan *dense* dengan *dropout* (128 unit/0,3 dan 64 unit/0,2) [16], serta satu keluaran regresi tunggal. *Sequence* masukan menggunakan *look-back window* sepanjang 56 langkah. Model dilatih menggunakan *optimizer* Adam [17], fungsi *loss mean-squared-error*, *batch size* 32, *epoch* maksimum 50, dan *early stopping* (*patience* = 10) berdasarkan *validation loss*. Prediksi akhir diperoleh dengan melakukan *inverse-scaling* pada keluaran setiap IMF lalu menjumlahkannya secara aditif.

TABEL I
HIPERPARAMETER UTAMA MODEL

Hiperparameter	Nilai
Jumlah mode dekomposisi (K)	6
Konstanta *bandwidth* (*alpha*)	500
Jendela *look-back* (langkah)	56
Filter Conv1D (3 blok)	64/128/256, kernel 3, stride 1, padding 'same', ReLU, MaxPool 2
Unit *Dense* (*dropout*)	128(.3)/64(.2)
Optimizer / *fungsi loss*	Adam / MSE
Batch size / *epoch* maksimum	32 / 50
Patience early stopping	10

Prediksi *deep learning* umumnya menunjukkan bias sistematis, yaitu kecenderungan yang konsisten untuk melebih-lebihkan atau meremehkan nilai target. Penelitian ini mengoreksi bias tersebut dengan menghitung konstanta b = −MBE_val dari *validation set* dan menerapkan $\hat{y}\_terkoreksi(t) = \hat{y}(t) + b$ pada prediksi *test set*, sehingga suku koreksi tidak pernah berasal dari data yang digunakan pada evaluasi akhir.

CNN1D dipilih dibandingkan arsitektur berbasis *recurrent* (LSTM, GRU) atau *attention* (*Transformer*) dengan tiga pertimbangan. Pertama, penelitian ini melatih enam model independen (satu per IMF); kompleksitas komputasi CNN1D yang lebih rendah menjadikannya lebih sesuai untuk skema pelatihan per-komponen ini. Kedua, filter konvolusi lokal secara alami sesuai dengan karakteristik IMF hasil VMD, yang masing-masing terbatas pada pita frekuensi sempit sehingga pola temporalnya cenderung bersifat lokal dan berulang, bukan bergantung-jangka-panjang yang menjadi keunggulan utama LSTM/GRU/*Transformer*. Ketiga, penelitian ini secara sengaja diposisikan sebagai *baseline* univariat yang ringan dan mudah direproduksi, sebelum kompleksitas arsitektur tambahan dipertimbangkan pada penelitian lanjutan. Perbandingan langsung dengan arsitektur *recurrent/attention* belum dilakukan pada revisi ini dan diusulkan sebagai prioritas penelitian lanjutan.

D. Metrik Evaluasi dan Protokol Perbandingan

Kinerja model dinilai menggunakan enam metrik komplementer yang masing-masing menangkap dimensi kesalahan yang berbeda: koefisien determinasi (R^2), *root mean square error* (RMSE), *mean absolute error* (MAE), *mean absolute percentage error* (MAPE), *mean bias error* (MBE), dan *Index of Agreement* [14]. Seluruh metrik dihitung baik sebelum maupun sesudah koreksi bias untuk mengisolasi efeknya. Model akhir kemudian dibandingkan dengan tiga alternatif terkontrol yang memiliki pembagian data, arsitektur, hiperparameter, dan prosedur koreksi bias yang identik: (A) *baseline* CNN1D tanpa dekomposisi, (C) varian VMD–CNN1D dengan K adaptif, dan (D) varian VMD–CNN1D multivariat yang menyertakan kelembapan relatif dan suhu sebagai fitur eksogen.

III. HASIL DAN PEMBAHASAN

A. Sensitivitas Parameter VMD dan Dekomposisi

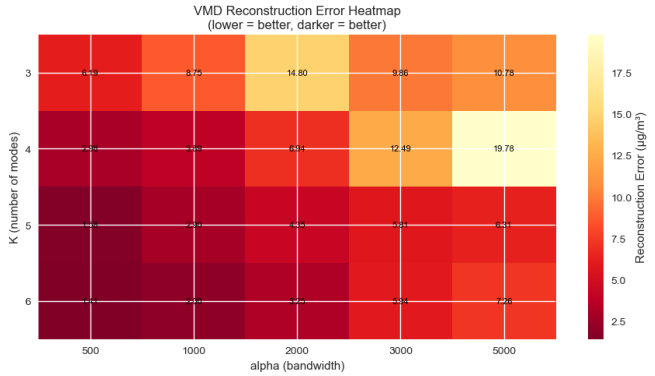
Tabel II merangkum *reconstruction error* terbaik yang diperoleh untuk setiap jumlah mode K dari 20 kombinasi yang dievaluasi. *Reconstruction error* secara konsisten meningkat seiring bertambahnya *alpha* pada K yang tetap, sejalan dengan teori VMD: alpha yang lebih besar memaksa *bandwidth* spektral setiap mode menjadi lebih sempit, sehingga menyisakan lebih banyak energi residual yang tidak tertangkap [3]. Kombinasi K = 6, *alpha* = 500 menghasilkan *reconstruction error* terendah secara keseluruhan (1,88%),

jauh di bawah ambang batas penerimaan 3%, dan dipilih sebagai konfigurasi akhir.

TABEL II
RE TERBAIK PER JUMLAH MODE K

K	Alpha terbaik	RE (%)
3	500	8.26
4	500	3.97
5	500	2.11
6	500	1.88*

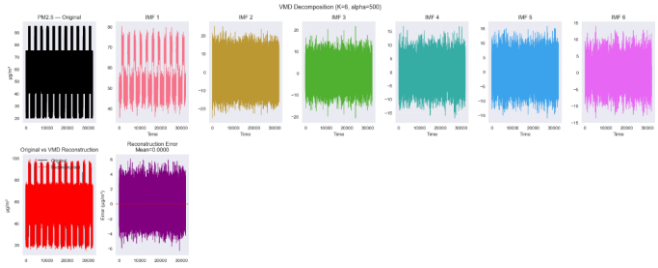
*K = 6, alpha = 500 dipilih (RE terendah).



Gambar 1. *Heatmap Reconstruction Error* untuk Seluruh Kombinasi Parameter VMD.

Gambar 1 memvisualisasikan pola RE pada Tabel II secara lebih intuitif. Pada hampir seluruh nilai K, RE meningkat secara konsisten seiring bertambahnya *alpha*, kecuali pada K = 3 yang menunjukkan pola tidak monoton. Sel dengan RE terendah (K = 6, alpha = 500) tampak jelas sebagai area tergelap pada peta panas, mengonfirmasi bahwa kombinasi ini merupakan titik optimal secara visual maupun numerik.

Dengan menggunakan K = 6 dan *alpha* = 500, sinyal PM2,5 didekomposisi menjadi enam IMF. IMF 1 memiliki amplitudo terbesar (35,25–80,18 µg/m³) dan secara visual merepresentasikan tren harian/musiman yang dominan, sementara amplitudo menurun secara progresif dari IMF 2 hingga IMF 6, yang terakhir menyerupai derau frekuensi tinggi yang bersifat episodik (−13,65 hingga 13,92 µg/m³). Sinyal hasil rekonstruksi hampir sepenuhnya sesuai dengan sinyal asli, dengan rata-rata *reconstruction error* yang mendekati nol dan tanpa tren residual yang sistematis.

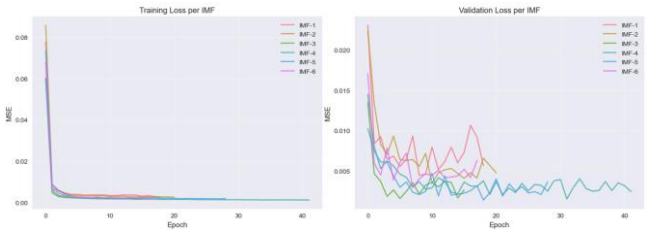


Gambar 2. Hasil Dekomposisi VMD: Sinyal Asli, Enam IMF, Rekonstruksi, dan *Error* Rekonstruksi.

Gambar 2 menegaskan pola penurunan amplitudo yang progresif dari IMF 1 ke IMF 6, konsisten dengan prinsip dasar VMD bahwa mode dipisahkan berdasarkan pita frekuensi, dengan IMF berindeks rendah menangkap komponen frekuensi rendah beramplitudo besar (tren/siklik), dan IMF berindeks tinggi menangkap komponen frekuensi tinggi beramplitudo kecil (derau/variasi episodik). Panel "*Original* vs VMD *Reconstruction*" menunjukkan kesesuaian visual yang sangat tinggi antara sinyal asli dan hasil rekonstruksi, dan panel *error* rekonstruksi menunjukkan sebaran yang relatif merata tanpa tren sistematis.

B. Efek Koreksi Bias

Sebelum evaluasi akhir, keenam model CNN1D (satu per IMF) dilatih secara independen dengan *early stopping* berdasarkan *validation loss*. Gambar 3 menampilkan kurva training dan *validation loss* untuk seluruh enam model.



Gambar 3. *Training Loss* dan *Validation Loss* per IMF.

Seluruh model menunjukkan penurunan *loss* yang tajam pada beberapa *epoch* awal, mengindikasikan konvergensi awal yang cepat untuk seluruh IMF. *Validation loss* menunjukkan osilasi yang lebih besar dibandingkan *training loss*, pola umum yang wajar karena *validation set* tidak digunakan untuk pembaruan bobot model. Konvergensi yang stabil pada seluruh IMF ini menjadi prasyarat penting agar tahap koreksi bias berikutnya dapat diinterpretasikan secara valid.

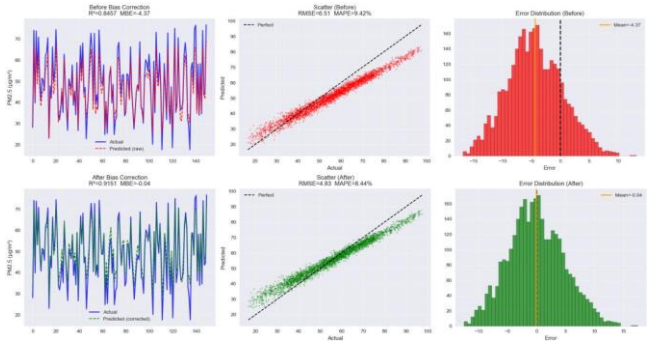
TABEL III
METRIK SEBELUM/SESUDAH KOREKSI BIAS

Metrik	Sebelum	Sesudah
R^2	0.8457	0.9151
RMSE (µg/m³)	6.51	4.83
MAE (µg/m³)	5.41	3.87
MAPE (%)	9.42	8.44
MBE (µg/m³)	−4.37	−0.04
IoA	0.9503	0.9719

Tabel III membandingkan keenam metrik evaluasi sebelum dan sesudah koreksi bias. Model mentah telah mencapai R^2 yang tampak memadai sebesar 0,8457, namun *mean bias error* mengungkap *underestimation* sistematis sebesar −4,3668 µg/m³ (sekitar 7,9% dari rata-rata nilai aktual), di mana bias tersebut tidak akan terdeteksi apabila evaluasi berhenti pada angka R^2/RMSE agregat semata. Koreksi terhadap bias ini, yang dihitung hanya dari *validation set*, memperbaiki seluruh metrik: R^2 naik menjadi 0,9151, RMSE turun menjadi 4,83 µg/m³, dan *Index of Agreement*

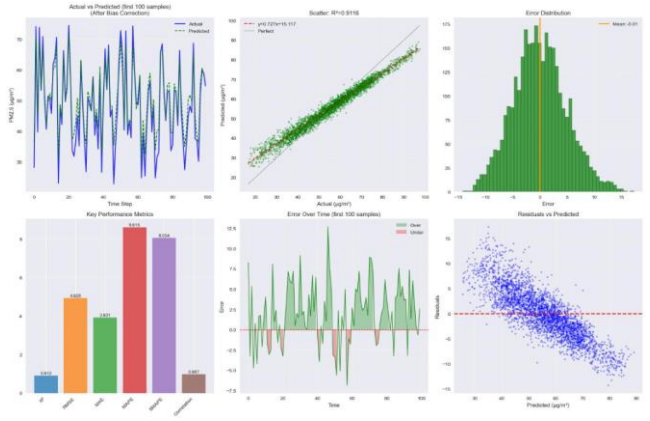
naik menjadi 0,9719. Hal ini menegaskan bahwa MBE seharusnya menjadi komponen wajib, bukan opsional, dalam protokol evaluasi prediksi deret waktu lingkungan.

Besarnya perbaikan performa dapat dijelaskan melalui dekomposisi *mean squared error* (MSE) ke dalam komponen varians residual dan bias sistematis kuadrat: MSE = Var(e) + (MBE)², dengan e menyatakan residual prediksi. Sebelum koreksi, MSE ≈ 6,51² = 42,38 μg²/m⁶; sesudah koreksi, MSE ≈ 4,83² = 23,33 μg²/m⁶, selisih ≈ 19,05 μg²/m⁶ yang hampir identik dengan kuadrat bias yang dihilangkan, MBE² ≈ (−4,37)² = 19,10 μg²/m⁶. Kesesuaian ini mengonfirmasi bahwa perbaikan performa hampir seluruhnya berasal dari penghapusan komponen bias sistematis, sementara varians residual relatif tidak berubah, sekaligus menjadi pemeriksaan konsistensi bahwa perbaikan bukan berasal dari kebocoran data, mengingat mekanisme koreksi murni bersifat aditif-konstan dan tidak melibatkan pelatihan ulang model.



Gambar 4. Perbandingan Prediksi Sebelum (atas) dan Sesudah (bawah) Koreksi Bias.

Sebelum koreksi bias, garis prediksi secara konsisten berada di bawah garis nilai aktual pada rentang konsentrasi tinggi. Setelah koreksi bias, garis prediksi mengikuti pola aktual secara jauh lebih presisi di sepanjang rentang nilai, mengonfirmasi secara visual perbaikan numerik yang dilaporkan pada Tabel III.

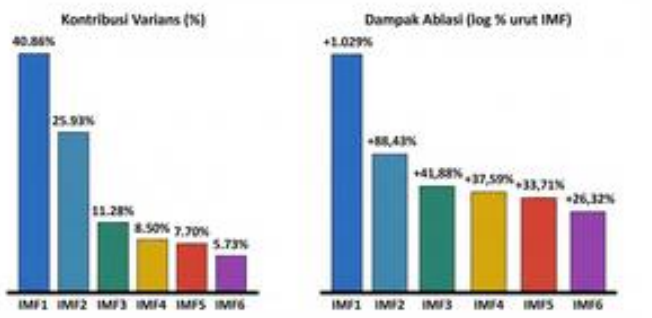


Gambar 5. Visualisasi Metrik Komprehensif Model Akhir Setelah Koreksi Bias.

Panel *scatter* (kanan atas) menunjukkan garis regresi empiris y = 0,727x + 15,117: meskipun korelasi keseluruhan tinggi (R^2 = 0,9151), model masih menunjukkan kecenderungan *underestimate* pada konsentrasi PM2,5 sangat tinggi dan *overestimate* pada konsentrasi rendah, pola *regression to the mean* yang umum dijumpai pada model prediktif. Pola ini dikonfirmasi lebih eksplisit pada panel residual vs prediksi (kanan bawah), yang menunjukkan tren menurun: residual cenderung positif (*under*-prediksi) pada nilai prediksi rendah dan negatif (*over*-prediksi) pada nilai prediksi tinggi yang menjadi indikasi adanya heteroskedastisitas yang dibahas lebih lanjut pada bagian Keterbatasan.

C. Kontribusi Mode Hasil Dekomposisi

Tabel IV melaporkan kontribusi varians setiap IMF terhadap sinyal asli serta degradasi RMSE yang terjadi ketika IMF tersebut dikeluarkan dari rekonstruksi (ablasi). IMF 1 merupakan komponen paling kritis berdasarkan kedua ukuran tersebut (menyumbang 40,9% dari total varians dan menyebabkan peningkatan RMSE sebesar 1.029% ketika dikeluarkan), konsisten dengan perannya sebagai komponen tren yang dominan. Yang menarik, IMF 2 dan IMF 3, meskipun memiliki kontribusi varians menengah, menghasilkan dampak ablasi yang lebih besar daripada yang diperkirakan hanya dari besaran variansnya, mengindikasikan bahwa kontribusi suatu IMF terhadap akurasi rekonstruksi bergantung pada kompleksitas pola temporalnya, bukan semata-mata pada amplitudo.



Gambar 6. Perbandingan Kontribusi Varians dan Dampak Ablasi Setiap IMF.

Gambar 6 menampilkan kedua ukuran tersebut secara berdampingan. Pola yang tampak saling menguatkan: IMF 1 unggul jauh pada kedua ukuran, sementara IMF 2 dan IMF 3 menunjukkan dampak ablasi yang tidak proporsional terhadap kontribusi variansnya, yang mengindikasikan bahwa tingkat kesulitan prediksi suatu IMF ditentukan oleh kompleksitas pola temporalnya, bukan semata-mata oleh amplitudo atau variansnya.

Dominasi IMF 1 secara fisik konsisten dengan karakteristik polusi PM2,5 di Jakarta, yang dipengaruhi kuat oleh pola musiman, khususnya peningkatan konsentrasi pada musim kemarau (Juni–September) yang berkaitan dengan kondisi atmosfer stabil, minimnya presipitasi pencuci partikulat, serta kontribusi kebakaran lahan dan hutan di wilayah Sumatra dan Kalimantan yang terbawa ke wilayah

Jabodetabek. IMF berindeks tinggi (IMF 5–6) kemungkinan merepresentasikan variasi episodik jangka pendek seperti pola kemacetan lalu lintas harian atau kondisi cuaca lokal sesaat.

TABEL IV
KONTRIBUSI VARIANS DAN DAMPAK ABLASI PER IMF

IMF	Var. (%)	RMSEΔ (%)
1	40.9	+1,029.4
2	25.9	+88.4
3	11.3	+41.9
4	8.5	+33.7
5	7.7	+37.6
6	5.7	+26.3

D. Perbandingan dengan Pendekatan Pemodelan Alternatif

Untuk mengisolasi faktor sebenarnya yang mendorong akurasi prediksi, empat pendekatan dibandingkan dengan pembagian data, arsitektur, dan prosedur koreksi bias yang identik (Tabel V): (A) CNN1D tanpa dekomposisi, (B) VMD–CNN1D dengan parameter tetap yang diusulkan, (C) VMD–CNN1D dengan K adaptif (K = 7 terpilih secara otomatis), dan (D) VMD–CNN1D multivariat yang menambahkan kelembapan relatif dan suhu sebagai fitur eksogen.

TABEL V
PERBANDINGAN EMPAT PENDEKATAN (SETELAH KOREKSI BIAS)

Pendekatan	R^2	RMSE	MAPE(%)
A: CNN1D Tanpa dekomposisi	0.4302	12.51	22.68
B: VMD-CNN1D Univariat (Diusulkan) ★	0.9177±0.0173	4.7346±0.5064	8.36±0.90
C: VMD-CNN1D K Adaptif	0.9238±0.0089	4.5388±0.2649	7.97±0.51
D: VMD-CNN1D Multivariat	0.9003	5.23	9.15

RMSE dalam µg/m³. ★ = model yang dipilih pada penelitian ini.

Tiga temuan muncul dari perbandingan ini. Pertama, dekomposisi sinyal (bukan kompleksitas arsitektur atau pemilihan parameter) merupakan faktor dominan penentu akurasi: penambahan VMD meningkatkan R^2 dari 0,4302 menjadi 0,9151 dan menurunkan RMSE sebesar 61,4%, peningkatan yang jauh lebih besar dibandingkan selisih di antara B, C, dan D. Kedua, terkait kompleksitas tambahan dari pemilihan K adaptif, validasi lanjutan pada revisi ini melalui 5 kali pelatihan ulang dengan seed acak berbeda (lihat Bagian III.D.3) menunjukkan bahwa Pendekatan C secara konsisten unggul pada 4 dari 5 run dengan variabilitas *run-to-run* yang jauh lebih rendah (SD RMSE 0,26 vs 0,51), meskipun perbedaan rata-rata antar-*run* belum mencapai signifikansi statistik pada N=5 replikasi (*paired t-test*, p=0,684), sehingga klaim superioritas Pendekatan C bersifat cenderung (*promising trend*), bukan kesimpulan definitif, dan mengundang replikasi lebih lanjut. Ketiga, dan mungkin paling instruktif bagi penelitian selanjutnya, penambahan variabel meteorologi yang berkorelasi secara naif (Pendekatan D) tidak meningkatkan performa dan justru sedikit menurunkannya (R^2 turun dari 0,9177 menjadi 0,9003 dibandingkan rata-rata Pendekatan B; perbandingan D menggunakan run tunggal dan belum divalidasi ulang multi-run), meskipun korelasi antara faktor meteorologi dan PM2,5 telah terdokumentasi dengan baik [18]. Hal ini menunjukkan bahwa korelasi statistik antara fitur kandidat dan target merupakan syarat perlu namun tidak cukup agar fitur tersebut meningkatkan akurasi prediksi suatu model tertentu; seleksi fitur dan analisis struktur *lag* diperlukan sebelum variabel eksogen diperkenalkan, sebuah kehati-hatian yang ditunjukkan secara langsung oleh eksperimen multivariat awal pada penelitian ini.

D.1 Efisensi Komputasi dan Implikasi Praktis

Waktu pelatihan rata-rata untuk keenam model CNN1D pada Pendekatan B adalah 3.018,5 ± 1.241,3 detik (≈ 50,3 ± 20,7 menit), diukur dari 5 run independen dengan *seed* berbeda. Waktu inferensi untuk seluruh *test-set* (≈ 3.159 *timestep*) berkisar 3,5–5,3 detik, setara 1,1–1,7 milidetik per-*timestep* untuk keenam model gabungan. Seluruh eksperimen dijalankan pada CPU (Intel64 Family 6, tanpa GPU), RAM 15,8 GB. Mengingat data BMKG diperbarui setiap tiga jam, waktu inferensi pada orde milidetik ini jauh berada di bawah interval pembaruan data operasional, sehingga secara komputasi layak diterapkan sebagai komponen prediksi pada sistem peringatan dini kualitas udara, bahkan tanpa akselerasi GPU. Frekuensi pelatihan ulang untuk mengantisipasi *concept drift* dan pemantauan degradasi performa belum diuji dan diusulkan sebagai penelitian lanjutan.

D.2 Catatan Metodologis: Non-Determinisme Pelatihan

Selama proses validasi ulang untuk revisi ini, ditemukan bahwa hasil pelatihan CNN1D pada CPU tidak sepenuhnya *reproducible* meskipun *random seed* telah ditetapkan secara eksplisit, kemungkinan besar karena operasi reduksi *floating-point multi-thread* pada *backend TensorFlow* yang urutannya tidak deterministik di CPU. Temuan ini menjadi motivasi langsung bagi validasi *multi-run* pada subbagian berikut, menggantikan pendekatan *single-run* pada manuskrip versi sebelumnya. Analisis sensitivitas terhadap *window size* belum dapat disertakan pada revisi ini karena keterbatasan waktu komputasi (lihat Bagian III.F, Keterbatasan).

D.3 Uji Signifikansi Statistik

Pendekatan B dan C masing-masing dilatih ulang 5 kali dengan seed berbeda (42, 7, 123, 2024, 99). K optimal yang dipilih Pendekatan C konsisten K=7 pada seluruh 5 seed — menunjukkan mekanisme pemilihan K adaptif itu sendiri stabil, meskipun proses pelatihan CNN1D hilir tidak sepenuhnya deterministik.

TABEL VI
DIEBOLD-MARIANO TEST PER-SEED
(PENDEKATAN B VS C, GROUND TRUTH PM2.5 MENTAH)

Seed	RMSE B	RMSE C	Lebih Baik	DM stat	p-value	Signifikan
42	4,6951	4,6526	C	0,956	0,3393	Tidak
7	5,4762	4,9395	C	11,608	<0,0001	Ya
123	4,9246	4,6178	C	5,644	<0,0001	Ya
2024	5,1703	4,4820	C	11,292	<0,0001	Ya
99	4,1810	5,1226	B	-19,633	<0,0001	Ya

Diebold-Mariano test di dalam tiap *run* menunjukkan selisih akurasi signifikan pada 4 dari 5 *run*, dengan Pendekatan C unggul pada seluruh 4 *run* tersebut; pada 1 *run* (*seed*=99), Pendekatan B unggul secara signifikan. Karena titik-titik pada satu deret waktu tidak sepenuhnya independen, uji DM di dalam satu *run* cenderung mendeteksi signifikansi meskipun selisihnya relatif kecil. Sebagai pelengkap yang lebih konservatif, *paired t-test* antar-*run* (RMSE rata-rata per-*seed* sebagai unit observasi independen, N=5) menghasilkan t=0,439, p=0,684, TIDAK signifikan secara statistik. Kedua hasil ini valid namun menjawab pertanyaan berbeda: DM *test per-run* mengonfirmasi bahwa dalam satu proses pelatihan tertentu, selisih akurasi B vs C biasanya terdeteksi nyata; namun uji antar-*run* yang lebih konservatif menunjukkan bahwa dengan baru 5 replikasi, belum cukup bukti untuk mengklaim superioritas Pendekatan C secara umum. Meskipun demikian, Pendekatan C unggul pada 4 dari 5 run dan menunjukkan variabilitas *run-to-run* yang jauh lebih kecil (SD RMSE 0,2649 vs 0,5064) yang mengindikasikan bahwa mekanisme adaptif K tidak hanya cenderung lebih akurat, tetapi juga menghasilkan pelatihan yang lebih stabil. Replikasi lebih banyak (N>5) diperlukan pada penelitian lanjutan untuk memastikan tren ini dengan keyakinan statistik yang lebih tinggi.

TABEL VII
PERBANDINGAN DENGAN PENELITIAN TERDAHULU

Studi	Metode	Dataset	R^2
Wu et al. [7]	MIC-CEEMDAN-CNN-BiGRU	Beijing, 2016	>0.95
Khattak et al. [6]	VMD-BiGRU	Riyadh/Dammam/Jeddah	0.90–0.95*
Hashim [19]	VMD-BiGRU	3 kota Arab Saudi	>0.90
Bai et al. [11]	CNN-LSTM (tanpa dekomp.)	Qingdao, 2020	0.91
Zhang & Zhang [20]	*Sparse Attention Transformer*	Beijing & Taizhou	0.937
Penelitian ini	VMD-CNN1D (univariat)	BMKG Jakarta	0.9177\u00b10.0173

*Perkiraan berdasarkan RMSE 9,25–16,05 µg/m³ yang dilaporkan untuk horizon t+1 hingga t+3; R^2 eksplisit tidak dilaporkan untuk seluruh horizon.

Posisi penelitian ini di antara studi terdahulu perlu dipahami secara kontekstual, mengingat perbedaan dataset, lokasi geografis, rentang konsentrasi PM2,5, dan resolusi temporal setiap penelitian. RMSE dalam satuan absolut (µg/m³) sangat dipengaruhi oleh skala dan variabilitas data asli, sehingga nilai RMSE yang lebih rendah tidak selalu mengindikasikan model yang secara intrinsik lebih baik [13]. Dengan kesadaran atas keterbatasan tersebut, Tabel VI membandingkan R^2 sebagai ukuran kualitas *fit* yang relatif lebih dapat dibandingkan lintas-dataset.

Wu et al. [7] melaporkan R^2 di atas 0,95, melampaui capaian penelitian ini, namun perbandingan langsung tidak sepenuhnya adil karena perbedaan metode dekomposisi (CEEMDAN versus VMD), arsitektur (BiGRU versus CNN1D), dan karakteristik data Beijing yang memiliki variabilitas musiman sangat tinggi akibat episode polusi parah musim dingin. Studi Hashim [19] memiliki kesamaan metodologis paling dekat karena sama-sama menggunakan VMD sebagai metode dekomposisi; R^2 penelitian ini (0,9151) berada di atas batas bawah yang dilaporkan, mengindikasikan performa yang sebanding meskipun arsitektur prediksi dan konteks data berbeda. Bai et al. [11] melaporkan R^2 = 0,91 tanpa dekomposisi pada data Qingdao, setara secara nilai dengan penelitian ini, namun perbedaan yang lebih bermakna adalah bahwa tanpa VMD, CNN1D pada penelitian ini hanya mencapai R^2 = 0,4302, yang mengindikasikan bahwa data PM2,5 Jakarta bersifat lebih non-stasioner sehingga menuntut dekomposisi untuk mencapai akurasi memadai, sementara data Qingdao tampaknya lebih dapat dipelajari tanpa pra-pemrosesan serupa. Zhang & Zhang [20] melaporkan R^2 = 0,937 dengan RMSE = 19,04 µg/m³ menggunakan *Sparse Attention Transformer*, RMSE yang jauh lebih besar dari penelitian ini (4,83 µg/m³) secara langsung mencerminkan rentang konsentrasi data mereka yang jauh lebih lebar, termasuk episode di atas 200 µg/m³.

Keunggulan penelitian ini dibandingkan studi-studi tersebut bukan terletak pada nilai metrik absolut, melainkan pada kelengkapan dan ketegaran metodologis. Pertama, penelitian ini secara eksplisit melaporkan dan mengoreksi bias sistematis melalui mekanisme yang dihitung dari *validation set* terpisah, sebuah praktik yang jarang dilaporkan secara transparan dalam literatur prediksi PM2,5. Kedua, parameter VMD tidak ditetapkan secara arbitrer melainkan melalui *grid search* sistematis terhadap 20 kombinasi, langkah yang terbukti krusial karena parameter *default* umum menghasilkan *reconstruction error* 4,9 kali lebih besar pada dataset ini. Ketiga, evaluasi menggunakan enam metrik komprehensif dibandingkan umumnya dua hingga tiga metrik pada studi pembanding, memungkinkan deteksi heteroskedastisitas residual yang tidak tertangkap oleh metrik agregat saja.

E. Keterbatasan

Enam keterbatasan perlu diakui. Pertama, koreksi bias aditif yang digunakan bersifat konstan dan tidak mengatasi heteroskedastisitas yang teramati pada residual, sehingga

akurasi pada nilai PM2,5 ekstrem (sangat tinggi atau sangat rendah) kemungkinan lebih rendah dibandingkan akurasi rata-rata yang dilaporkan di sini; koreksi bias bersyarat/kuantil atau *conformal prediction* dapat dieksplorasi pada penelitian lanjutan. Kedua, analisis ablasi kontribusi mode menggunakan kembali prediksi IMF yang sudah ada, bukan melatih ulang model tanpa masing-masing IMF, sehingga dampak yang dilaporkan mencerminkan kontribusi pada tingkat rekonstruksi, bukan pada tingkat pembelajaran. Ketiga, sebagai penelitian univariat, model ini belum menyertakan faktor meteorologi (suhu, kelembapan, kecepatan angin, curah hujan) yang diketahui memengaruhi dinamika PM2,5, sehingga performanya sebaiknya dipahami sebagai *baseline* pola temporal, bukan batas atas (*upper bound*) yang dapat dicapai dengan informasi tambahan. Keempat, penelitian ini hanya menggunakan data dari satu stasiun pemantauan di Jakarta, sehingga generalisasi parameter optimal (K=6, *alpha*=500) ke lokasi geografis lain belum dapat dipastikan. Kelima, sensitivitas performa terhadap panjang *window* (*look-back*) belum diuji pada revisi ini karena keterbatasan waktu komputasi; *window*=56 dipertahankan berdasarkan pertimbangan pada manuskrip awal, namun titik optimal sebenarnya belum divalidasi secara sistematis. Keenam, validasi *multi-run* pada revisi ini terbatas pada N=5 replikasi; uji statistik antar-*run* belum mencapai signifikansi meskipun Pendekatan C unggul pada 4 dari 5 *run*, dengan replikasi lebih banyak (N=10–20) dan perbandingan langsung dengan arsitektur *recurrent/attention* (LSTM, GRU, *Transformer*, TCN) diperlukan pada penelitian lanjutan.

IV. KESIMPULAN

Penelitian ini merancang dan memvalidasi model VMD–CNN1D untuk prediksi konsentrasi PM2,5 secara univariat di Jakarta. Analisis sensitivitas sistematis berbasis *reconstruction error* menetapkan K = 6 dan *alpha* = 500 sebagai parameter VMD optimal (RE = 1,88%), dan mekanisme koreksi bias berbasis *validation set* meningkatkan R^2 dari 0,8457 menjadi 0,9151 sekaligus menurunkan RMSE menjadi 4,83 µg/m³. Perbandingan terkontrol dengan pendekatan alternatif mengonfirmasi bahwa dekomposisi sinyal, bukan kompleksitas model atau pemilihan parameter, merupakan kontributor dominan terhadap akurasi Prediksi. Hal ini dibuktikan oleh *baseline* tanpa dekomposisi hanya mencapai R^2 = 0,4302. Validasi lanjutan melalui 5 kali pelatihan ulang dengan *seed* berbeda menunjukkan bahwa varian K adaptif (Pendekatan C) secara konsisten unggul pada 4 dari 5 *run* dengan variabilitas *run-to-run* yang jauh lebih rendah dibandingkan parameter tetap (Pendekatan B), meskipun perbedaan rata-rata antar-*run* belum mencapai signifikansi statistik pada N=5 replikasi (*paired t-test*, p=0,684), sehingga kompleksitas tambahan dari K adaptif menunjukkan kecenderungan yang menjanjikan, bukan keunggulan yang telah terbukti definitif, dan mengundang replikasi lebih lanjut sebelum dapat diklaim secara meyakinkan. Analisis kontribusi mode mengidentifikasi IMF berfrekuensi terendah sebagai komponen paling kritis bagi akurasi, dan analisis residual mengungkap heteroskedastisitas sebagai masalah terbuka. Eksperimen multivariat awal menunjukkan bahwa penambahan variabel meteorologi tanpa seleksi fitur eksplisit justru menurunkan, bukan meningkatkan, performa, menegaskan bahwa korelasi semata tidak menjamin sinyal prediktif yang dapat dieksploitasi.

Penelitian selanjutnya sebaiknya mengeksplorasi replikasi *multi-run* dengan jumlah pengulangan lebih besar (N=10–20) untuk memastikan tren superioritas Pendekatan C dengan keyakinan statistik yang lebih tinggi, analisis sensitivitas terhadap panjang *window* (*look-back*), koreksi bias bersyarat atau berbasis kuantil untuk mengatasi heteroskedastisitas, seleksi fitur sistematis dan analisis *lag* sebelum memasukkan kovariat meteorologi, serta perbandingan langsung dan terkontrol dengan arsitektur *deep learning* alternatif (misalnya berbasis *recurrent* atau *attention*) menggunakan protokol terkontrol yang sama seperti pada penelitian ini. Perluasan ke prediksi *multi-step* (t+3 hingga t+24) juga memiliki nilai aplikatif yang lebih tinggi bagi sistem peringatan dini dibandingkan prediksi satu langkah (t+1) yang digunakan pada penelitian ini, sekaligus akan menguji ketangguhan mekanisme koreksi bias dan dekomposisi VMD pada cakupan prediksi yang lebih jauh. Selain itu, karena penelitian ini hanya menggunakan data dari satu stasiun pemantauan, perluasan cakupan ke beberapa stasiun BMKG di wilayah Jabodetabek atau kota besar lain di Indonesia disarankan untuk menguji apakah parameter optimal (K = 6, alpha = 500) bersifat spesifik-lokasi atau dapat digeneralisasi pada konteks geografis yang lebih luas. Model univariat yang telah tervalidasi ini menyediakan baseline yang dapat direproduksi bagi pengembangan sistem peringatan dini kualitas udara berbasis numerik untuk Jakarta di masa mendatang.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atas penyediaan data pemantauan PM2,5 yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

[1] IQAir, “2023 World Air Quality Report,” 2024.

[2] WHO, “WHO global air quality guidelines Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide,” 2021.

[3] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.

[4] Huang, S. R. L, H. H. Shih, N. Y. i-C, and H. H. and, N. E. L, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” 1998.

[5] Z. Wu and N. E. Huang, “Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method,” 2009.

[6] A. Khattak, S. Alotaibi, R. N. Alahmadi, C. M. Matara, and S. Taglawi, “Hybrid VMD–BiGRU Framework for Multi-Step

Forecasting of PM2.5 in Traffic-Intensive Cities of the Kingdom of Saudi Arabia," *Atmosphere (Basel).*, vol. 16, no. 12, Dec. 2025, doi: 10.3390/atmos16121324.

[7] X. Wu, J. Zhu, and Q. Wen, "Short-term prediction of PM2.5 concentration by hybrid neural network based on sequence decomposition," *PLoS One*, vol. 19, no. 5 May, May 2024, doi: 10.1371/journal.pone.0299603.

[8] X. H. Wang *et al.*, "Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD–GAT–BiLSTM," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-68874-x.

[9] T. Zeng, L. Xu, Y. Liu, R. Liu, Y. Luo, and Y. Xi, "A hybrid optimization prediction model for PM2.5 based on VMD and deep learning," *Atmos. Pollut. Res.*, vol. 15, no. 7, Jul. 2024, doi: 10.1016/j.apr.2024.102152.

[10] T. Zeng *et al.*, "A Deep Learning PM2.5 Hybrid Prediction Model Based on Clustering–Secondary Decomposition Strategy," *Electronics (Switzerland)*, vol. 13, no. 21, Nov. 2024, doi: 10.3390/electronics13214242.

[11] X. Bai, N. Zhang, X. Cao, and W. Chen, "Prediction of PM2.5 concentration based on a CNN-LSTM neural network algorithm," *PeerJ*, vol. 12, no. 8, 2024, doi: 10.7717/peerj.17811.

[12] J. Zhang, X. Xin, Y. Shang, Y. Wang, and L. Zhang, "Nonstationary significant wave height forecasting with a hybrid VMD-CNN model," *Ocean Engineering*, vol. 285, Oct. 2023, doi: 10.1016/j.oceaneng.2023.115338.

[13] S. Zhou, W. Wang, L. Zhu, Q. Qiao, and Y. Kang, "Deep-learning architecture for PM2.5 concentration prediction: A review," Sep. 01, 2024, *Editorial Board, Research of Environmental Sciences*. doi: 10.1016/j.ese.2024.100400.

[14] C. J. Willmott, "On The Validation Of Models," *Phys. Geogr.*, vol. 2, no. 2, pp. 184–194, Jul. 1981, doi: 10.1080/02723646.1981.10642213.

[15] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," Mar. 2015, [Online]. Available: http://arxiv.org/abs/1502.03167

[16] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.

[17] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017, [Online]. Available: http://arxiv.org/abs/1412.6980

[18] Q. Yang, Q. Yuan, T. Li, H. Shen, and L. Zhang, "The relationships between PM2.5 and meteorological factors in China: Seasonal and regional variations," *Int. J. Environ. Res. Public Health*, vol. 14, no. 12, Dec. 2017, doi: 10.3390/ijerph14121510.

[19] M. A. A.-A. A. S. Hashim, "Air Quality and Particulate Matter Forecasting using VMD-BiGRU Deep Learning Model in Saudi Arabia Cities," *Atmosphere (Basel).*, vol. 16, no. 1, pp. 245–263, 2025.

[20] Z. Zhang and S. Zhang, "Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks," *International Journal of Environmental Science and Technology*, vol. 20, no. 12, pp. 13535–13550, Dec. 2023, doi: 10.1007/s13762-023-04900-1.