

**PERBANDINGAN KINERJA ST-DBSCAN DAN ST-SCKM
SERTA MODEL *HYBRID* DALAM PENGELOMPOKAN DATA
KUALITAS UDARA BERBASIS SPASIAL-TEMPORAL**

MUH AKBAR IDRIS



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Perbandingan Kinerja ST-DBSCAN dan ST-SCKM serta Model *Hybrid* dalam Pengelompokan Data Kualitas Udara Berbasis Spasial-Temporal” adalah karya saya dengan arahan komisi pembimbing. Tesis ini belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Seluruh data, analisis, dan interpretasi yang disajikan menjadi tanggung jawab penulis.

Seluruh informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan telah disebutkan di dalam teks. Sumber tersebut juga dicantumkan dalam Daftar Pustaka pada bagian akhir tesis. Penggunaan sumber dilakukan sesuai dengan kaidah sitasi ilmiah yang berlaku.

Perangkat kecerdasan artifisial *Google Gemini*, *ChatGPT*, dan *DeepSeek* digunakan secara terbatas untuk membantu penelusuran gagasan, perbaikan bahasa, penyusunan struktur awal, dan pemeriksaan konsistensi penulisan. Seluruh keluaran perangkat tersebut diperiksa kembali menggunakan data, kode analisis, pustaka ilmiah, dan arahan komisi pembimbing sebelum digunakan. Penulis bertanggung jawab penuh atas ketepatan metode, hasil, interpretasi, sitasi, dan seluruh isi tesis ini.

Hak cipta tesis ini dilimpahkan kepada Institut Pertanian Bogor sesuai dengan ketentuan yang berlaku. Pelimpahan tersebut mencakup pengelolaan karya ilmiah untuk kepentingan pendidikan dan penyebaran pengetahuan. Penggunaan kembali isi tesis tetap wajib mencantumkan sumber secara benar.

Bogor, Agustus 2026



Muh Akbar Idris
M0501241013

RINGKASAN

MUH. AKBAR IDRIS. Perbandingan Kinerja ST-DBSCAN dan ST-SCKM Serta *Model Hybrid* dalam Pengelompokan Data Kualitas Udara Berbasis Spasial-Temporal. Dibimbing oleh MUHAMMAD NUR AIDI dan ANIK DJURAIDAH.

Penggerombolan data spasial-temporal diperlukan untuk mengenali pola yang berubah menurut atribut, lokasi, dan waktu. Metode berbasis kepadatan dapat membentuk gerombol tidak beraturan sekaligus menghasilkan label *noise*, tetapi sensitif terhadap perbedaan kepadatan dan parameter. Metode berkendala spasial memberikan regionalisasi yang lebih menyeluruh, tetapi tidak secara langsung mempertahankan kandidat anomali sebagai keluaran akhir.

Penelitian ini membandingkan *Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise* (ST-DBSCAN), *Spatio-Temporal Spatially Constrained K-Means* (ST-SCKM), dan dua model *hybrid*. Kajian simulasi menggunakan rancangan faktorial penuh 2^3 yang mengombinasikan geometri sirkular dan koridor, jarak antarpusat 3 dan 7 satuan, serta kontaminasi anomali 0% dan 10%. Setiap skenario terdiri atas 1.200 observasi, empat gerombol acuan, dan sepuluh replikasi yang dianalisis secara terpisah pada ruang fitur empat dimensi yang memuat X , Y , waktu, dan $F1$ atau $F2$.

Hasil simulasi menunjukkan bahwa ST-SCKM memperoleh rata-rata *Composite Regionalization Score* sebesar 0,7233 dan skor tertinggi pada tujuh dari delapan skenario. Skenario 3 menghasilkan skor yang sama antara ST-SCKM dan kedua model *hybrid* setelah pembulatan. Evaluasi pada skenario terkontaminasi menunjukkan bahwa ST-DBSCAN memperoleh rata-rata *F1-score* deteksi sebesar 0,7863, sedangkan *Hybrid 2* memperoleh nilai sebesar 0,7206.

Kajian empiris menggunakan data pemantauan $PM_{2.5}$, NO_2 , dan NO_x di Jepang tahun 2024 yang diperoleh melalui OpenAQ. Pengolahan dengan median tiga jam menghasilkan sampel akhir sebanyak 9.998 observasi yang dianalisis pada ruang fitur delapan dimensi, yaitu bujur, lintang, waktu, sinus bulan, kosinus bulan, $PM_{2.5}$, NO_2 , dan NO_x . Pendekatan ST-SCKM menghasilkan empat zona dengan cakupan 100% dan *Composite Zoning Score* tertinggi sebesar 0,7919, sedangkan metode ST-DBSCAN menghasilkan cakupan 81,69% dengan 1.831 observasi berlabel *noise*.

Hasil penelitian menempatkan metode ST-SCKM sebagai metode utama untuk regionalisasi dan metode ST-DBSCAN sebagai metode pendukung untuk mendeteksi kandidat anomali pada konfigurasi yang diuji. Empat zona hasil metode ST-SCKM memperlihatkan perbedaan relatif dalam profil polutan, lokasi, dan musim dominan, sementara pola *High-High* dan *Low-Low* dari analisis spasial lokal digunakan sebagai informasi deskriptif. Nilai *F1-score metode* ST-DBSCAN terhadap *pseudo*-acuan rentang antarkuartil mencapai 0,4443 untuk $PM_{2.5}$, 0,4513 untuk NO_2 , dan 0,5313 untuk NO_x , tetapi pemeriksaan sensor, meteorologi, sumber emisi, dan kejadian lokal tetap diperlukan untuk memastikan makna substantif setiap kandidat.

Kata kunci: data kualitas udara, deteksi kandidat anomali, model *hybrid*, penggerombolan spasial-temporal, regionalisasi.



SUMMARY

MUH. AKBAR IDRIS. Performance Comparison of ST-DBSCAN, ST-SCKM, and Hybrid (DBSCAN–SCKM) for Spatio-Temporal Clustering of Air Quality Data. Supervised by Muhammad Nur Aidi and Anik Djuraidah.

Spatio-temporal clustering is required to identify patterns that vary across attributes, locations, and periods. Density-based methods can identify irregularly shaped clusters and retain noise labels, although their results are sensitive to density variation and parameter selection. Spatially constrained methods provide more complete regionalization but do not directly retain anomaly candidates as their final output result.

This study compared Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN), Spatio-Temporal Spatially Constrained K-Means (ST-SCKM), and two hybrid models. The simulation study employed a full 2^3 factorial design combining circular and corridor geometries, center separations of 3 and 7 units, and anomaly contamination levels of 0% and 10%. Each scenario contained 1,200 observations, four reference clusters, and ten replications analyzed separately in two four-dimensional feature spaces consisting of X , Y , time, and either $F1$ or $F2$.

The simulation results showed that ST-SCKM method achieved a mean Composite Regionalization Score of 0.7233 and the highest score in seven of the eight scenarios. Scenario 3 produced equal rounded scores for ST-SCKM method and both hybrid models. Detection evaluation in the contaminated scenarios showed that ST-DBSCAN method achieved a mean F1-score of 0.7863, compared with 0.7206 for Hybrid 2.

The empirical study used 2024 monitoring data for $PM_{2.5}$, NO_2 , and NO_x in Japan obtained through OpenAQ. Three-hour median aggregation produced a final sample of 9,998 observations analyzed in an eight-dimensional feature space consisting of longitude, latitude, time, month sine, month cosine, $PM_{2.5}$, NO_2 , and NO_x . ST-SCKM generated four zones with 100% coverage and the highest Composite Zoning Score of 0.7919, whereas ST-DBSCAN achieved 81.69% coverage and labelled 1,831 observations as noise.

The findings support ST-SCKM as the primary regionalization method and ST-DBSCAN as a supporting method for candidate anomaly detection under the evaluated configurations. The four ST-SCKM zones exhibited relative differences in pollutant profiles, locations, and dominant seasons, while local High-High and Low-Low spatial patterns were interpreted descriptively. ST-DBSCAN achieved F1-scores of 0.4443 for $PM_{2.5}$, 0.4513 for NO_2 , and 0.5313 for NO_x against the interquartile-range pseudo-references, although sensor records, meteorological conditions, emission sources, and local events remain necessary to verify the substantive meaning of each candidate.

Keywords: air quality data, candidate anomaly detection, hybrid model, regionalization, spatio-temporal clustering.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERBANDINGAN KINERJA ST-DBSCAN DAN ST-SCKM SERTA MODEL *HYBRID* DALAM PENGELOMPOKAN DATA KUALITAS UDARA BERBASIS SPASIAL-TEMPORAL

MUH AKBAR IDRIS

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:

1. Dr. Farit Mochamad Afendi, S.Si., M.Si.



Judul Tesis : Perbandingan Kinerja ST-DBSCAN dan ST-SCKM serta Model *Hybrid* dalam Pengelompokan Data Kualitas Udara Berbasis Spasial-Temporal

Nama : Muh Akbar Idris
NIM : M0501241013

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Muhammad Nur Aidi, MS



Pembimbing 2:
Prof. Dr. Ir. Anik Djuraidah, MS

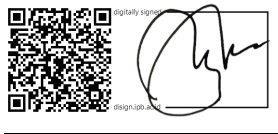


Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP 19750315 199903 1 004



Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom
NIP 19660702 199302 1 001



Tanggal Ujian: 04 Agustus 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah Swt. atas rahmat dan karunia-Nya sehingga tesis ini dapat diselesaikan. Tesis ini merupakan bagian dari proses akademik yang ditempuh pada Program Studi Statistika dan Sains Data, Sekolah Pascasarjana IPB University. Penyelesaian penelitian ini tidak terlepas dari dukungan, arahan, dan doa berbagai pihak.

Penulis menyampaikan terima kasih kepada semua pihak yang telah mendampingi proses penelitian ini. Dukungan yang diberikan mencakup arahan ilmiah, bantuan administratif, dukungan pembiayaan, serta penguatan moral. Secara khusus, penghargaan disampaikan kepada pihak-pihak berikut:

1. Orang tua dan keluarga penulis yang selalu menempatkan doa, dukungan, dan keteguhan hati sebagai kekuatan utama selama studi. Kesabaran mereka membantu penulis melalui setiap tahap penelitian. Tesis ini dipersembahkan kepada Almarhuma Mama tercinta, cinta pertama dan kasih sayang abadi.
2. Prof. Dr. Ir. Muhammad Nur Aidi, MS selaku pembimbing utama, atas arahan, bimbingan, dan masukan yang sangat berharga selama proses penelitian dan penulisan tesis ini.
3. Prof. Dr. Ir. Anik Djuraidah, MS selaku pembimbing kedua, atas bimbingan, dukungan, dan saran yang konstruktif.
4. Lembaga Pengelola Dana Pendidikan (LPDP), Kementerian Keuangan Republik Indonesia, dukungan pendanaan studi kepada penulis.
5. Seluruh dosen dan staf Program Studi Statistika dan Sains Data IPB University atas ilmu, bantuan, dan pelayanan akademik yang diberikan.
6. Rekan-rekan mahasiswa Pascasarjana yang telah memberikan dukungan dan kerja sama selama masa studi.
7. Amelia, Iskandar, Nisa, Rati, Yola, Winda, Fadila, Iin, Fani, Dian, Asy, Feldy, Wawan, Adib, Sunan, Syafiq, Nisa, Salsa, Sabrina, Kak Arman, Nabil, Mardan, Hijrah, Syella, Zamrah yang sudah benar-benar menjadi keluarga baru saya dan mewarnai keluh kesah selama S2 ini.
8. Grup Hujat Iskandar, Himpro Abhipraya, LARIIII, BSC, Drop Shot Crew IPB, dan Rumana IPB Sulselbar yang telah menjadi bagian berharga dalam perjalanan penulis.

Sebagian hasil penelitian dalam tesis ini telah dikembangkan menjadi karya ilmiah yang telah terbit pada jurnal internasional bereputasi (Q1 dan Q3 *Scopus*) serta dipresentasikan pada konferensi internasional (AESAP 2026) Bogor, Indonesia dan (ICAS 2026) Bangkok, Thailand.

Penulis menyadari bahwa karya ilmiah ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan di masa yang akan datang. Semoga karya ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang analisis spasial-temporal dan *data science*.

Bogor, 2026

Muh Akbar Idris



DAFTAR ISI

DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Ruang Lingkup	4
II TINJAUAN PUSTAKA	5
2.1 Gerombol Spasial-Temporal	5
2.2 Analisis Kualitas Udara	7
2.3 Algoritma DBSCAN	8
2.4 Algoritma <i>Spatio-Temporal</i> DBSCAN	10
2.5 <i>Outlier</i> dan <i>Noise</i> dalam Data <i>Spasial-Temporal</i> Kualitas Udara	13
2.6 <i>Algoritma Spatio-Temporal Spatially Constrained K-Means</i>	14
2.7 Model Gerombol <i>Hybrid</i>	19
III KAJIAN SIMULASI	22
3.1 Desain Simulasi	22
3.2 Prosedur Analisis	25
3.3 Hasil dan Pembahasan Kajian Simulasi	28
IV KAJIAN EMPIRIS	33
4.1 Data Empiris	33
4.2 Prosedur Analisis Empiris	36
4.3 Eksplorasi Data Empiris	40
4.4 Perbandingan Kinerja Metode	43
4.5 Karakteristik Zona Hasil ST-SCKM	48
4.6 Deteksi Kandidat Anomali	53
4.7 Sintesis Kajian Empiris	56
V SIMPULAN DAN SARAN	57
5.1 Simpulan	57
5.2 Keterbatasan Penelitian	57
DAFTAR PUSTAKA.....	60
LAMPIRAN	63
RIWAYAT HIDUP	71

DAFTAR TABEL

1	Rancangan faktorial skenario simulasi	24
2	Rataan <i>Composite Regionalization Score</i> menurut skenario dan atribut	28
3	Rataan metrik regionalisasi menurut metode dan atribut	30
4	Rataan <i>F1-score</i> deteksi anomali pada skenario berkontaminasi	32
5	Aliran jumlah data kualitas udara Jepang tahun 2024	34
6	Statistika deskriptif sampel pemodelan kualitas udara Jepang tahun 2024	35
7	Parameter terpilih untuk sampel data kualitas udara Jepang tahun 2024	37
8	Kinerja marginal berdasarkan PM _{2.5} pada partisi multivariat data Jepang tahun 2024	44
9	Kinerja marginal berdasarkan NO ₂ pada partisi multivariat data Jepang tahun 2024	44
10	Kinerja marginal berdasarkan NO _x pada partisi multivariat data Jepang tahun 2024	44
11	Rataan kinerja marginal ketiga polutan pada partisi multivariat data Jepang tahun 2024	44
12	Kinerja regionalisasi empat metode pada ruang fitur multivariat delapan dimensi, sampel Jepang tahun 2024	45
13	Dukungan temporal hasil penggerombolan multivariat data kualitas udara Jepang tahun 2024	48
14	Profil zona intensitas relatif ST-SCKM di Jepang tahun 2024	49
15	Kinerja deteksi kandidat anomali berdasarkan <i>pseudo</i> -acuan IQR	54
16	Konsentrasi observasi berlabel gerombol dan <i>noise</i> ST-DBSCAN	55

DAFTAR GAMBAR

1	Algoritma ST-DBSCAN.	13
2	Ilustrasi empat bentuk pencilan pada data kualitas udara spasial-temporal	14
3	Algoritma ST-SCKM	18
4	Algoritma <i>Hybrid</i> 1	19
5	Algoritma <i>Hybrid</i> 2	20
6	Ilustrasi struktur delapan skenario simulasi	24
7	Perbandingan hasil empat metode pada Skenario 6 untuk <i>F1</i> dan <i>F2</i> .	29
8	Rataan <i>Composite Regionalization Score</i> terlaras cakupan pada delapan skenario.	30
9	Hasil ST-SCKM pada delapan skenario menggunakan atribut <i>F1</i> .	31
10	Hasil ST-SCKM pada delapan skenario menggunakan atribut <i>F2</i> .	31
11	Aliran data kualitas udara Jepang tahun 2024	34
12	Prosedur analisis empiris kualitas udara Jepang	36
13	Distribusi musiman PM _{2.5} , NO ₂ , dan NO _x pada sampel pemantauan kualitas udara Jepang tahun 2024	41

14	Pola mingguan dan pola dalam satu hari untuk PM _{2.5} , NO ₂ , dan NO _x pada sampel pemantauan Jepang tahun 2024	41
15	Sebaran rata-rata tahunan PM _{2.5} , NO ₂ , dan NO _x pada stasiun pemantauan Jepang tahun 2024	42
16	Sebaran spasial musiman PM _{2.5} , NO ₂ , dan NO _x pada sampel pemantauan Jepang tahun 2024	43
17	Perbandingan metrik regionalisasi multivariat terlaras cakupan pada sampel Jepang tahun 2024	46
18	Profil pusat gerombol multivariat di Jepang tahun 2024	46
19	Plot siluet zona relatif ST-SCKM di Jepang tahun 2024	47
20	Perbandingan dukungan temporal di Jepang tahun 2024	48
21	Distribusi polutan pada zona relatif ST-SCKM	50
22	Dinamika musiman PM _{2.5} , NO ₂ , dan NO _x pada empat zona relatif ST-SCKM	50
23	Komposisi musim pada zona relatif ST-SCKM	51
24	Pola LISA pada stasiun pemantauan di Jepang tahun 2024	52
25	Sintesis spasial-temporal zona relatif ST-SCKM pada sampel Jepang tahun 2024	53
26	Perbandingan spasial kandidat anomali	55
27	Profil marginal PM _{2.5} , NO ₂ , NO _x , dan bulan untuk observasi berlabel gerombol serta <i>noise</i> ST-DBSCAN pada sampel Jepang tahun 2024	56

DAFTAR LAMPIRAN

1	Daftar simbol dan definisi	64
2	Rincian kuantitatif skenario simulasi	69
3	Parameter dan bobot analisis akhir	69
4	Informasi perangkat lunak dan reproduksibilitas	70