E-ISSN: 2964-6219, P-ISSN: 2964-6200
Volume 5 Number 1 February 2027.
pp: 38-49. DOI: https://doi.org/10.58477/cj.v5i1.542.

Computer Journal

RESEARCH ARTICLE Open Access

Artificial Intelligence-Based Spatiotemporal Clustering for Identifying Homogeneous Climatic Zones in West Java

Oktaviana Indriani [1*], **Eka Ardhianto** [2]

[1*,2] Department of Information Technology, Information Technology and Industry, Universitas Stikubank, Semarang City, Central Java Province, Indonesia.

Correspondence email:
oktavianaindriani7002@mhs.unisbank.ac.id.

Article History
Received: 18 September 2026.
Revised: 19 September 2026.
Accepted: 19 October 2026.
Available Online: 5 November 2026
Published: 1 February 2027.

Funding Information
Universitas Stikubank.

Abstract
Homogeneous Climatic Zones (ZOM) support operational seasonal services in Indonesia, particularly across regions where rainfall changes sharply over short distances. This study evaluates a climatic regionalization framework that combines Uniform Manifold Approximation and Projection (UMAP) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) using BMKG gridded rainfall data across West Java at approximately 5 km resolution for the 1991–2020 period. Each of the 1,214 grid cells was represented by a 36-dasarian climatological rainfall profile. A search across 360 hyperparameter combinations yielded 324 executable configurations, followed by spatial absorption of transition grids and layered evaluation using the Silhouette Score, Heterogeneity Index (HI), and operational tractability. Scenario 60 (*n_neighbors* = 5, *min_dist* = 0.01, *min_cluster_size* = 10, *min_samples* = 5) produced 54 initial zones with a Silhouette Score of 0.5633 and an HI of 7.239, compared with an HI of 8.464 for the 41-zone operational baseline. Subsequent spatial refinement produced a corrected final map containing 50 sequential ZOM. Distinguishing the evaluation metrics of the 54-zone candidate stage from the 50-zone cartographic structure of the spatial refinement stage clarifies the analytical sequence and supports a more precise interpretation of cluster quality, climatic homogeneity, and final zone boundaries.

Keywords: Clustering; Dasarian Rainfall; HDBSCAN; Heterogeneity Index; Seasonal Zone; UMAP.

Abstrak
Zona Musim (ZOM) mendukung layanan informasi musim operasional di Indonesia, terutama pada wilayah dengan perubahan curah hujan yang tajam dalam jarak pendek. Penelitian ini mengevaluasi kerangka regionalisasi iklim yang menggabungkan *Uniform Manifold Approximation and Projection* (UMAP) dan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) menggunakan data curah hujan grid BMKG di Jawa Barat beresolusi sekitar 5 km untuk periode 1991–2020. Sebanyak 1.214 grid direpresentasikan sebagai profil klimatologi curah hujan 36 dasarian. Eksperimen mencakup 360 kombinasi hiperparameter dan menghasilkan 324 konfigurasi yang dapat dievaluasi, dilanjutkan dengan proses *absorption* pada grid transisi serta evaluasi berlapis menggunakan *Silhouette Score*, *Heterogeneity Index* (HI), dan keterkelolaan operasional. Skenario 60 (*n_neighbors* = 5, *min_dist* = 0,01, *min_cluster_size* = 10, *min_samples* = 5) menghasilkan 54 zona awal dengan *Silhouette Score* sebesar 0,5633 dan HI sebesar 7,239, dibandingkan dengan HI sebesar 8,464 pada *baseline* 41 ZOM operasional. Penyempurnaan spasial selanjutnya menghasilkan peta final terkoreksi yang terdiri atas 50 ZOM berurutan. Pemisahan antara metrik evaluasi pada tahap kandidat 54 zona dan struktur kartografis pada tahap finalisasi 50 ZOM memperjelas urutan analitis serta mendukung interpretasi yang lebih presisi terhadap kualitas klaster, homogenitas klimatologis, dan batas zonasi akhir.

Kata Kunci: Curah Hujan Dasarian; HDBSCAN; Indeks Heterogenitas; Klasterisasi; UMAP; Zona Musim.

© The Author(s) 2027. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

1. Pendahuluan

Zona Musim (ZOM) yang ditetapkan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) merupakan landasan operasional untuk memantau awal musim hujan, awal musim kemarau, serta penyusunan informasi iklim sektoral di Indonesia. Sebagai wilayah dengan batas daratan dan lautan yang kontras serta topografi yang membentang dari dataran rendah pantai utara hingga rangkaian pegunungan di bagian tengah dan selatan, Provinsi Jawa Barat memiliki keragaman curah hujan spasial dan temporal yang tinggi. Sejumlah penelitian menunjukkan bahwa variasi curah hujan di Jawa Barat memiliki ketergantungan spasial-temporal yang kuat serta dipengaruhi oleh interaksi antara kondisi geografis lokal dan sirkulasi atmosfer regional, sehingga wilayah yang berdekatan secara geografis dapat memiliki rezim hujan yang berbeda (Rachmawati, 2021; Rinaldi *et al.*, 2025). Dalam publikasi pemutakhiran ZOM periode klimatologis 1991–2020, pembagian wilayah musim umumnya dibangun melalui analisis pengelompokan yang dilengkapi dengan pertimbangan operasional (BMKG, 2022). Di sisi lain, ketersediaan data curah hujan beresolusi tinggi membuka peluang evaluasi ulang terhadap batas ZOM menggunakan metode berbasis data yang mampu menangkap variasi lokal secara lebih sistematis. Pemetaan berbasis data grid terbukti membantu mengidentifikasi variasi spasial zona iklim secara lebih terperinci di wilayah dengan topografi beragam, meskipun ketajaman batas zona tetap bergantung pada karakteristik metode pengelompokan yang digunakan (Madani & Wahid, 2022).

Dalam literatur klimatologi, pembentukan wilayah hujan yang homogen umumnya didekati melalui regionalisasi objektif berbasis klaster (Hosking & Wallis, 1997; Rao & Srinivas, 2006; Mazumdar, 2009). Penelitian terdahulu menunjukkan bahwa efektivitas regionalisasi tidak hanya ditentukan oleh algoritma klasterisasi, tetapi juga oleh kesesuaian antara representasi fitur, ukuran kemiripan, evaluasi validitas klaster, dan pengetahuan iklim lokal (Bieniek *et al.*, 2012; Zhang *et al.*, 2016; Luo *et al.*, 2024). Perkembangan mutakhir juga mengarah pada pemanfaatan teknik reduksi dimensi dan klasterisasi nonlinier untuk mengekstraksi struktur kompleks dari data presipitasi berdimensi tinggi (Wu *et al.*, 2024; King *et al.*, 2025). Salah satu pendekatan yang relevan adalah kombinasi *Uniform Manifold Approximation and Projection* (UMAP) dan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN). UMAP dirancang untuk mempertahankan struktur ketetanggaan lokal sekaligus merepresentasikan struktur global secara efisien pada ruang berdimensi rendah (McInnes *et al.*, 2018), sedangkan HDBSCAN mampu mengidentifikasi klaster dengan kepadatan bervariasi tanpa menetapkan jumlah klaster di awal serta memisahkan titik transisi atau pencilan sebagai *noise* (Campello *et al.*, 2013, 2015). Kombinasi UMAP dan klasterisasi berbasis kepadatan juga telah diterapkan pada pemetaan spasial di Indonesia dengan hasil yang kompetitif (Hasana & Fitrianah, 2023).

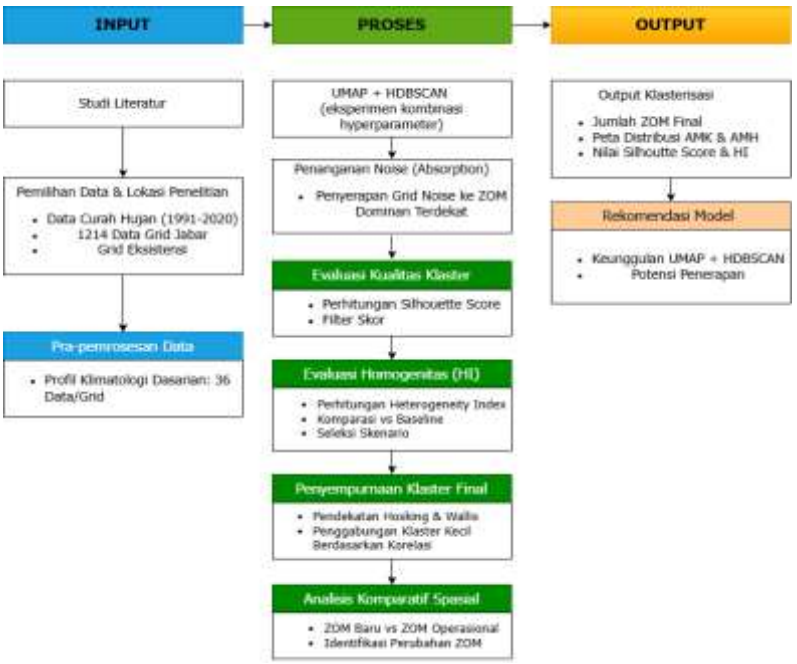
Meskipun kombinasi UMAP dan HDBSCAN potensial untuk memetakan variasi lokal, penerapannya pada regionalisasi ZOM memerlukan protokol yang transparan. Tantangan utamanya meliputi sifat stokastik pada *embedding* UMAP, sensitivitas hasil terhadap pemilihan hiperparameter, penanganan grid berlabel *noise* agar menghasilkan peta zona yang kontinu secara spasial, serta penyelarasan antara metrik validitas statistik dan kriteria operasional klimatologis. Tanpa pembedaan yang tegas antara hasil klasterisasi algoritmik dan tahap penyempurnaan kartografis pascapemrosesan, klaim peningkatan kualitas zona berisiko sulit direproduksi. Berdasarkan kebutuhan tersebut, penelitian ini bertujuan mengevaluasi kerangka kerja dua tahap yang menggabungkan UMAP dan HDBSCAN untuk meregionalisasi ZOM di Jawa Barat menggunakan profil klimatologi curah hujan 36 dasarian periode 1991–2020. Secara khusus, penelitian ini (1) menguji sensitivitas konfigurasi hiperparameter UMAP dan HDBSCAN terhadap pembentukan zona dan kualitas struktur klaster menggunakan *Silhouette Score*, (2) menerapkan prosedur penyerapan (*absorption*) *noise* dan penyempurnaan spasial untuk menghasilkan peta ZOM yang kontinu, serta (3) membandingkan homogenitas curah hujan intrazona terhadap 41 ZOM operasional BMKG menggunakan *Heterogeneity Index* (HI). Hasil penelitian ini diharapkan dapat menjadi rujukan metodologis berbasis data dalam mendukung peninjauan dan pemutakhiran ZOM operasional.

2. Metode

2.1 Data, Wilayah Kajian, dan Representasi Klimatologi

Penelitian ini dilakukan melalui serangkaian tahapan komputasional yang berurutan, mulai dari input dan prapemrosesan data curah hujan grid, reduksi dimensi, klasterisasi spasial yang dilanjutkan dengan penyerapan (*absorption*) *noise*, evaluasi kualitas klaster secara statistik dan klimatologis, hingga analisis komparatif Zona

Musim (ZOM) final sebagaimana ditunjukkan pada Gambar 1. Data yang digunakan merupakan data grid curah hujan operasional BMKG untuk wilayah administratif Provinsi Jawa Barat dengan resolusi spasial sekitar 5 km pada periode klimatologis Januari 1991–Desember 2020. Curah hujan direpresentasikan pada interval dasarian dalam satuan mm/dasarian yang mencakup 1.214 grid analisis. Grid di sekitar batas Jawa Barat turut dilibatkan pada proses klasterisasi awal untuk mengurangi efek tepi (*edge effect*), sedangkan evaluasi *Silhouette Score* dan *Heterogeneity Index* (HI) difokuskan pada grid di dalam batas administratif Jawa Barat. Sebagai pembanding, digunakan Peta Pemutakhiran Zona Musim Indonesia Periode 1991–2020 yang memuat 41 ZOM di Jawa Barat sebagai klasifikasi referensi operasional (*baseline*) untuk mengevaluasi jumlah zona, struktur spasial, dan dispersi curah hujan intrazona (BMKG, 2022).



Gambar 1. Alur Penelitian

Data deret waktu curah hujan dasarian periode 1991–2020 pada setiap grid selanjutnya diagregasi dengan menghitung rata-rata untuk setiap dasarian selama 30 tahun. Hasil agregasi tersebut membentuk profil klimatologi tahunan yang terdiri atas 36 nilai rata-rata curah hujan, yaitu tiga dasarian pada setiap bulan selama 12 bulan. Proses ini bertujuan mengurangi pengaruh fluktuasi curah hujan temporal dan menonjolkan pola musiman yang representatif selama periode klimatologis. Dengan demikian, setiap grid direpresentasikan sebagai satu vektor fitur berdimensi 36 yang digunakan sebagai masukan pada tahap reduksi dimensi.

2.2 Reduksi Dimensi UMAP dan Klasterisasi HDBSCAN

Tahap reduksi dimensi menggunakan UMAP untuk memproyeksikan profil klimatologi 36 dasarian ke ruang berdimensi lebih rendah dengan membangun graf ketetanggaan berbobot pada ruang fitur asli dan mengoptimalkan representasinya agar struktur lokal utama tetap terjaga (McInnes *et al.*, 2018). Karakteristik ini sesuai untuk data curah hujan klimatologis yang memiliki hubungan nonlinier antardasarian. Eksperimen menguji dua parameter utama UMAP, yaitu *n_neighbors* yang mengendalikan skala ketetanggaan lokal yang dipertahankan dan *min_dist* yang mengatur tingkat kekompakan titik pada ruang *embedding*. Ruang *embedding* UMAP selanjutnya dikelompokkan menggunakan HDBSCAN, yaitu algoritma klasterisasi berbasis kepadatan hierarkis yang membangun hierarki berdasarkan *mutual reachability distance* dan mengekstraksi klaster yang stabil dari struktur kepadatan tersebut (Campello *et al.*, 2013, 2015). Pendekatan ini sesuai untuk regionalisasi dengan kepadatan lokal yang bervariasi karena mampu membentuk klaster dengan ukuran berbeda sekaligus mengidentifikasi grid berkerapatan rendah sebagai *noise*. Dua parameter utama HDBSCAN yang diuji adalah *min_cluster_size*, yang mengatur ukuran minimum kelompok yang dipertahankan sebagai klaster, serta *min_samples*, yang mengendalikan keketatan definisi titik inti dan sensitivitas terhadap *noise*. Untuk menjamin reproduksibilitas komputasi, protokol verifikasi pada setiap skenario mendokumentasikan parameter *n_components*, metrik jarak pada UMAP dan HDBSCAN, *random seed*, serta versi pustaka perangkat lunak pada eksekusi ulang (*rerun*) terkontrol agar konfigurasi komputasi terhubung langsung dengan hasil yang dilaporkan.

2.3 Eksperimen Hiperparameter dan Penanganan *Noise*

Ruang pencarian hiperparameter yang disajikan pada Tabel 1 menghasilkan 360 kombinasi parameter (6×2×6×5). Sebanyak 36 kombinasi melanggar syarat relasi min_samples≤min_cluster_size, sehingga diperoleh 324 konfigurasi valid yang masuk ke tahap evaluasi dengan prosedur prapemrosesan, domain grid, dan aturan evaluasi yang seragam. Dari 324 konfigurasi tersebut, 15 konfigurasi menghasilkan klaster yang langsung mencakup seluruh grid, sedangkan 309 konfigurasi menyisakan grid berlabel *noise* pada area transisi kepadatan. Mengingat *embedding* UMAP bersifat stokastik, stabilitas antarseed ditempatkan sebagai komponen validasi ketahanan (*robustness*) yang dievaluasi melalui konsistensi jumlah zona, *Silhouette Score*, dan kesesuaian label antareksekusi menggunakan metrik seperti *Adjusted Rand Index* (ARI).

Tabel 1. Hiperparameter UMAP dan HDBSCAN

Parameter	Nilai yang diuji
umap_n_neighbors	3, 5, 10, 15, 30, 50
umap_min_dist	0,01; 0,001
hdbscan_min_cluster_size	5, 10, 15, 20, 30, 40
hdbscan_min_samples	2, 3, 5, 10, 15

Grid berlabel *noise* pada area transisi diproses melalui tahap *absorption* sebagai pascapemrosesan (*post-processing*) spasial yang terpisah dari mekanisme inti HDBSCAN. Prosedur ini mengikuti dua prinsip utama, yaitu dominasi spasial zona di sekitar grid transisi dan kemiripan klimatologis profil 36 dasarian. Setiap grid *noise* diarahkan ke zona berlabel yang paling dominan pada lingkungan spasial terdekat, kemudian kemiripan profil curah hujan digunakan sebagai dasar penguat apabila terdapat beberapa zona dengan kedekatan spasial yang sebanding. Proses ini berlangsung secara iteratif hingga seluruh domain kajian memperoleh label ZOM yang kontinu.

2.4 Evaluasi Klaster dan Penyempurnaan Spasial

Kualitas struktur klaster dari setiap konfigurasi dievaluasi menggunakan *Silhouette Score* pada ruang *embedding* UMAP yang menjadi masukan langsung HDBSCAN agar ruang evaluasi selaras dengan ruang pembentukan klaster (Rousseeuw, 1987). Nilai *Silhouette* untuk setiap grid ke-i dihitung menggunakan Persamaan (1):

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (1)$$

dengan a(i) merepresentasikan rata-rata jarak grid ke-i *i* terhadap anggota lain dalam klaster yang sama, dan b(i) merepresentasikan rata-rata jarak grid ke-i *i* terhadap anggota klaster terdekat lainnya. Nilai S(i) yang mendekati 1 menunjukkan kohesi intraklaster dan separasi antarklaster yang kuat. Konfigurasi kandidat yang mencapai rata-rata *Silhouette Score* di atas 0,51 dipilih untuk masuk ke tahap evaluasi klimatologis berikutnya. Kandidat terpilih selanjutnya dibandingkan dengan ZOM operasional menggunakan *Heterogeneity Index* (HI), yang merangkum simpangan baku curah hujan antargrid pada setiap dasarian di dalam suatu zona. Dalam penelitian ini, HI memiliki satuan mm/dasarian dan berfungsi sebagai ukuran dispersi absolut pada dataset klimatologis yang sama. Karena zona dengan rerata curah hujan tinggi berpotensi menghasilkan simpangan baku yang lebih besar, HI digunakan sebagai indikator komparatif internal antarpembagian zona pada periode dan data yang sama, sedangkan interpretasi lintas rezim basah dan kering dilakukan secara hati-hati. Simpangan baku curah hujan pada dasarian ke-t *t* (SDt) dihitung melalui Persamaan (2):

$$SD_t = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(P_{i,t} - \bar{P_t}\right)^2} \quad (2)$$

dengan Pi,t adalah curah hujan grid ke-i pada dasarian ke-t, P¯t adalah rata-rata curah hujan seluruh grid anggota dalam satu ZOM pada dasarian ke-t, dan n adalah jumlah grid anggota dalam ZOM tersebut. Perhitungan dilakukan untuk seluruh 36 dasarian dalam satu tahun, sehingga nilai HI untuk ZOM ke-j (HIj) dan rata-rata HI wilayah untuk seluruh m *m* ZOM (HI¯) masing-masing diperoleh melalui Persamaan (3) dan Persamaan (4):

$$HI_j = \frac{1}{36}\sum_{t=1}^{36} SD_t \quad (3)$$

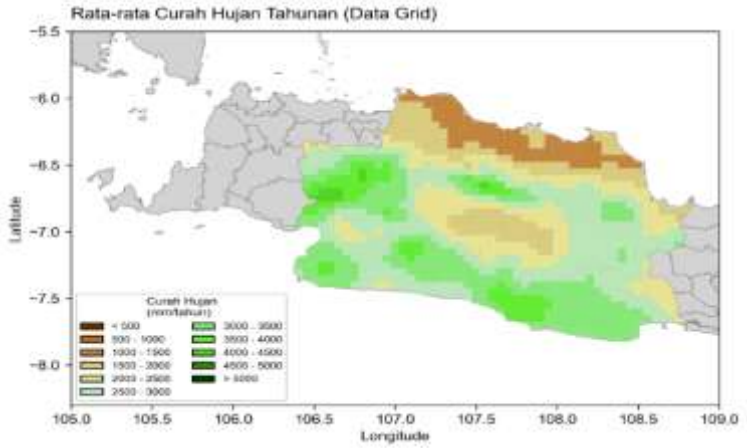
$$\overline{HI} = \frac{1}{m}\sum_{j=1}^{m} HI_j \quad (4)$$

Nilai HI¯HI yang lebih kecil menunjukkan dispersi absolut curah hujan intrazona yang lebih rendah. Evaluasi kandidat mempertimbangkan nilai HI bersama *Silhouette Score* dan jumlah zona, sementara analisis lanjutan dapat menyertakan ukuran berbasis koefisien variasi atau standardisasi profil guna menguji konsistensi kesimpulan pada zona basah dan kering. Setelah konfigurasi terbaik dipilih, dilakukan tahap finalisasi spasial yang mengevaluasi klaster awal dari sisi ukuran, kontinuitas spasial, dan kemiripan profil curah hujan. Klaster berukuran terlalu kecil atau terfragmentasi digabungkan ke klaster tetangga yang memiliki kemiripan klimatologis paling tinggi. Tahap pascapemrosesan ini menghasilkan peta final terkoreksi yang terdiri atas 50 ZOM dari 54 zona awal pada Skenario 60, dengan metrik statistik pada tahap kandidat (54 zona) dan struktur kartografis pada tahap finalisasi (50 zona) diperlakukan sebagai dua keluaran analitis yang berurutan agar interpretasi hasil tetap presisi.

3. Result and Discussion

3.1 Karakteristik Klimatologis Curah Hujan

Curah hujan tahunan di Jawa Barat memiliki variasi yang lebar, yakni berkisar antara 1.400 dan 4.000 mm, sejalan dengan keragaman topografi dari pesisir utara yang relatif datar hingga kawasan pegunungan tengah dan selatan dengan elevasi lebih dari 3.000 m dpl. Wilayah yang relatif paling kering terdapat pada ZOM operasional Jabar_03 dan Jabar_04 yang mencakup sebagian Karawang, Subang, dan Indramayu, dengan curah hujan tahunan sekitar 1.400–1.450 mm sebagaimana ditunjukkan pada Gambar 2. Sebaliknya, ZOM Jabar_16 di wilayah Bogor dan sebagian Sukabumi merupakan wilayah paling basah dengan curah hujan tahunan melebihi 3.800 mm dan berkarakter satu musim, yang ditandai oleh curah hujan bulanan di atas 150 mm sepanjang tahun.



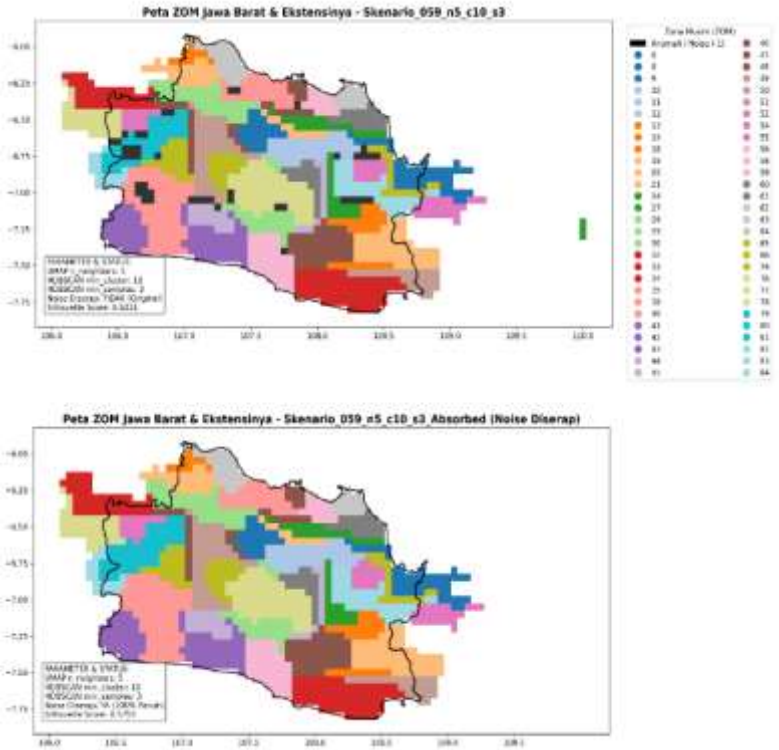
Gambar 2. Curah Hujan Tahunan di Jawa Barat

Sebagian besar wilayah Jawa Barat menunjukkan pola hujan monsunal dengan satu puncak musim hujan dan satu musim kemarau setiap tahun. Musim hujan umumnya berlangsung pada Oktober–April akibat pengaruh monsun Asia yang membawa massa udara lembap, sedangkan musim kemarau berlangsung pada Mei–September seiring dominasi monsun Australia yang relatif kering. Puncak musim hujan cenderung terjadi lebih awal di wilayah selatan pada periode Oktober–Desember hingga November–Januari (OND–NDJ) dibandingkan wilayah utara yang mengalami puncak hujan pada Desember–Februari hingga Januari–Maret (DJF–JFM). Awal musim kemarau juga memperlihatkan gradien spasial yang jelas, di mana pesisir utara memasuki musim kemarau lebih awal pada Maret–April, wilayah tengah pada sekitar Mei, dan wilayah selatan pada sekitar Juni, bahkan sebagian wilayah Bogor dan Sukabumi baru memasuki musim kemarau pada Juni–Agustus. Kontras curah hujan dan perbedaan waktu awal musim tersebut menunjukkan heterogenitas klimatologis yang kuat di Jawa Barat serta menjadi dasar penting bagi kebutuhan regionalisasi ZOM berbasis data.

3.2 Sensitivitas Parameter dan Eksperimen HDBSCAN

Pengujian awal menunjukkan bahwa parameter *umap_min_dist* memiliki pengaruh kuat terhadap jumlah klaster yang dihasilkan oleh HDBSCAN. Dengan mempertahankan parameter lainnya secara tetap

(*umap_n_neighbors* = 8, *hdbscan_min_cluster_size* = 20, dan *hdbscan_min_samples* = 10), penggunaan nilai *umap_min_dist* = 0,5 hanya menghasilkan 4 ZOM, sedangkan penurunan nilainya secara berturut-turut menjadi 0,1; 0,001; dan 0,0005 meningkatkan jumlah klaster menjadi 22, 27, dan 32 ZOM. Temuan ini memperlihatkan karakter lokal yang kuat pada variasi spasial curah hujan di Jawa Barat ketika UMAP membentuk kelompok titik yang semakin kompak. Oleh karena itu, eksperimen utama memusatkan ruang pencarian pada nilai *min_dist* yang lebih kecil agar struktur regionalisasi memiliki granularitas yang sesuai dengan kompleksitas iklim spasial Jawa Barat. Berdasarkan eksperimen utama terhadap 360 kombinasi parameter sesuai ruang pencarian pada Tabel 1, diperoleh 324 konfigurasi yang memenuhi batas relasi antara *min_samples* dan *min_cluster_size* untuk dievaluasi lebih lanjut. Dari jumlah tersebut, sebanyak 15 skenario membentuk klaster dengan seluruh grid langsung terklasifikasi, sedangkan 309 skenario menghasilkan sejumlah grid berlabel *noise* pada area transisi antarzona. Tahap *absorption* selanjutnya menetapkan label zona pada grid transisi tersebut sehingga seluruh domain kajian membentuk cakupan spasial yang penuh dan kontinu sebagaimana diilustrasikan pada Gambar 3.



Gambar 3. Proses Penanganan Grid Noise: (Atas) Sebelum Proses Absorption dan (Bawah) Setelah Proses Absorption

3.3 Evaluasi *Silhouette Score* dan Seleksi Kandidat Awal

Seluruh skenario yang valid diperingkat menggunakan *Silhouette Score* pada ruang fitur yang sama untuk seluruh grid di dalam wilayah administratif Jawa Barat. Sebanyak 31 skenario mencapai nilai *Silhouette Score* di atas 0,51 dan lolos ke tahap evaluasi lanjutan. Rincian konfigurasi parameter dari ke-31 skenario kandidat beserta jumlah zona awal yang terbentuk sebelum penyempurnaan spasial disajikan pada Tabel 2.

Tabel 2. Konfigurasi Parameter UMAP dan HDBSCAN yang Menghasilkan Nilai Silhouette Score Lebih dari 0,51

Peringkat	Skenario	*n_neighbors*	*min_dist*	*min_cluster_size*	*min_samples*	*Silhouette Score*	Jumlah ZOM
1	059	5	0,01	10	3	0,5755	57
2	087	5	0,001	10	5	0,5666	53
3	060	5	0,01	10	5	0,5633	54
4	058	5	0,01	10	2	0,5540	60
5	091	5	0,001	15	5	0,5479	32
6	096	5	0,001	20	5	0,5418	28
7	028	3	0,001	5	2	0,5414	141

8	095	5	0,001	20	3	0,5391	31
9	086	5	0,001	10	3	0,5368	57
10	032	3	0,001	10	3	0,5361	69
11	085	5	0,001	10	2	0,5357	61
12	089	5	0,001	15	2	0,5342	40
13	090	5	0,001	15	3	0,5332	40
14	073	5	0,01	30	3	0,5317	21
15	094	5	0,001	20	2	0,5313	34
16	074	5	0,01	30	5	0,5303	21
17	063	5	0,01	15	3	0,5285	33
18	056	5	0,01	5	3	0,5217	109
19	064	5	0,01	15	5	0,5210	38
20	062	5	0,01	15	2	0,5196	36
21	075	5	0,01	30	10	0,5194	21
22	083	5	0,001	5	3	0,5181	117
23	031	3	0,001	10	2	0,5179	72
24	069	5	0,01	20	5	0,5151	24
25	072	5	0,001	10	10	0,5149	37
26	088	5	0,01	30	2	0,5149	22
27	068	5	0,01	20	3	0,5142	31
28	071	5	0,01	20	15	0,5140	20
29	067	5	0,01	20	2	0,5131	29
30	029	3	0,001	5	3	0,5116	137
31	033	3	0,001	10	5	0,5107	66

Hasil pemeringkatan pada Tabel 2 memperlihatkan bahwa Skenario 059 memperoleh *Silhouette Score* tertinggi (0,5755), diikuti oleh Skenario 087 (0,5666) dan Skenario 060 (0,5633). Pola pada seluruh 31 skenario menunjukkan bahwa konfigurasi *umap_n_neighbors* = 5 mendominasi peringkat atas, sedangkan *umap_n_neighbors* = 3 cenderung menghasilkan jumlah ZOM yang jauh lebih besar (69–141 ZOM) pada nilai *Silhouette Score* yang relatif sebanding, yang mengindikasikan terjadinya fragmentasi berlebihan ketika ukuran ketetanggaan terlalu kecil. Penggunaan nilai *hdbscan_min_cluster_size* yang kecil (5) juga secara konsisten menghasilkan jumlah zona yang sangat besar (109–141 ZOM), sedangkan nilai yang lebih besar (15–30) menghasilkan jumlah zona yang lebih moderat (20–40 ZOM) dengan *Silhouette Score* yang sedikit lebih rendah. Temuan ini menegaskan adanya *trade-off* sistematis antara granularitas spasial (jumlah ZOM) dan validitas struktur klaster, sehingga pemilihan konfigurasi terbaik memerlukan pembandingan multikriteria terhadap homogenitas klimatologis dan kompleksitas jumlah zona sebagaimana diuraikan pada Subbagian 3.4.

3.4 Evaluasi *Heterogeneity Index* dan Pemilihan Skenario Konfigurasi

Evaluasi homogenitas klimatologis menggunakan nilai HI pada *baseline* 41 ZOM operasional sebesar 8,464 sebagai acuan pembanding. Dari 31 kandidat berbasis *Silhouette Score*, terdapat 15 konfigurasi yang menghasilkan nilai HI lebih rendah (lebih homogen) daripada *baseline* operasional. Metrik evaluasi yang disajikan pada Tabel 3 merepresentasikan pembagian zona awal setelah tahap *absorption* dan sebelum penyempurnaan klaster final.

Tabel 3. Skenario dengan Heterogeneity Index (HI) Lebih Baik daripada ZOM Operasional (Baseline HI = 8,464; 41 ZOM)

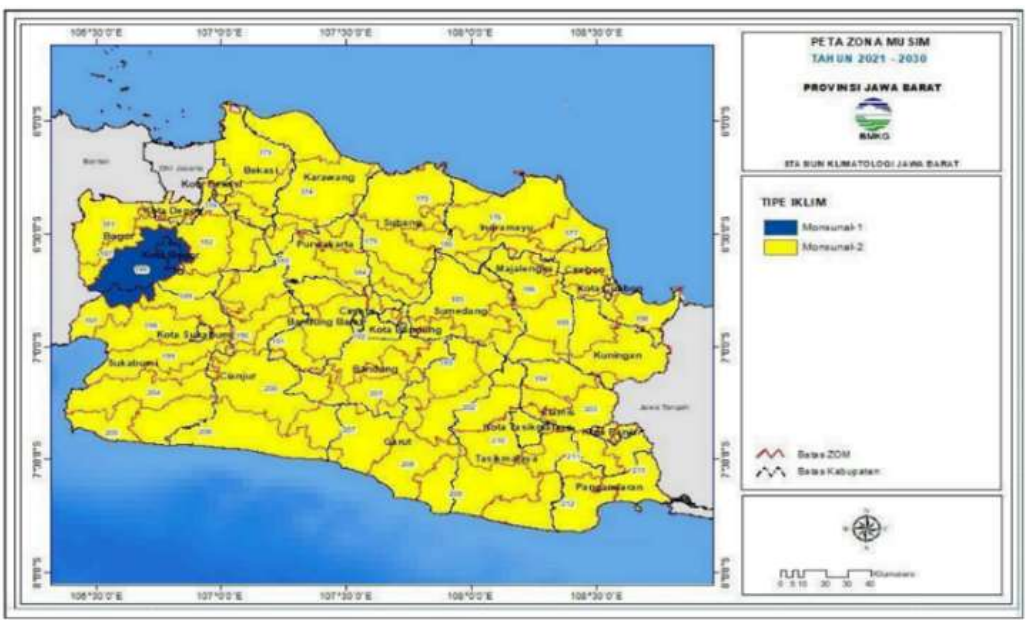
Skenario	Jumlah ZOM	HI	Peningkatan Homogenitas terhadap *Baseline*
ZOM Operasional (*baseline*)	41	8,464	–
089	40	8,406	+0,7%
090	40	8,418	+0,5%
060	54	7,239	+14,5%
087	53	7,462	+11,8%
031	72	7,462	+11,8%
059	57	7,071	+16,5%
085	61	7,055	+16,6%
086	57	7,228	+14,6%

058	60	6,984	+17,5%
032	69	7,577	+10,5%
033	66	7,723	+8,8%
056	109	5,513	+34,9%
083	117	5,299	+37,4%
029	137	5,301	+37,4%
028	141	5,189	+38,7%

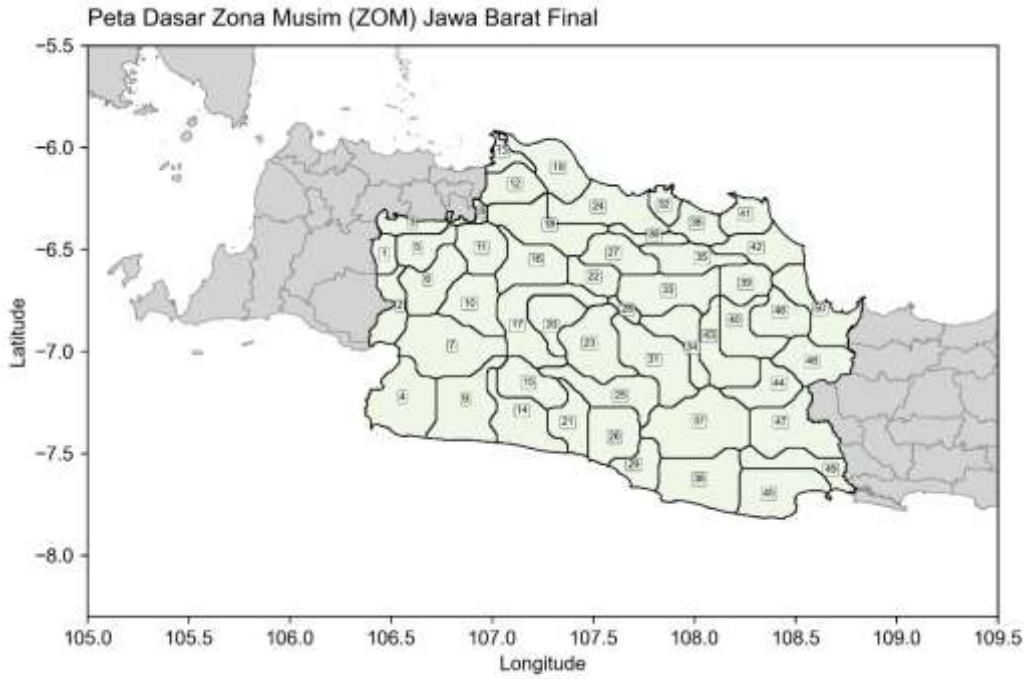
Data pada Tabel 3 memperlihatkan *trade-off* yang kuat antara jumlah zona dan dispersi curah hujan intrazona. Skenario 028, 029, 056, dan 083 menghasilkan 109–141 zona dengan nilai HI yang rendah (5,189–5,513), sedangkan Skenario 089 dan 090 menghasilkan 40 zona dengan nilai HI yang sangat dekat dengan *baseline*. Pada rentang jumlah zona yang lebih proporsional terhadap kebutuhan operasional, Skenario 059, 087, 060, 086, 085, dan 058 membentuk kelompok kandidat utama yang layak dibandingkan secara langsung. Perbandingan antarkandidat tersebut menunjukkan bahwa Skenario 059 menghasilkan 57 zona (*Silhouette Score* 0,5755; HI 7,071), Skenario 087 menghasilkan 53 zona (*Silhouette Score* 0,5666; HI 7,462), Skenario 060 menghasilkan 54 zona (*Silhouette Score* 0,5633; HI 7,239), Skenario 085 menghasilkan 61 zona (*Silhouette Score* 0,5357; HI 7,055), dan Skenario 058 menghasilkan 60 zona (*Silhouette Score* 0,5540; HI 6,984). Berdasarkan perbandingan tersebut, Skenario 060 (atau Skenario 60) dipilih untuk dibawa ke tahap finalisasi spasial karena menempati posisi kompromi menengah yang seimbang antara kualitas struktur klaster (*Silhouette Score* 0,5633), peningkatan homogenitas klimatologis (HI 7,239 atau meningkat 14,5%), dan jumlah zona awal (54 zona) yang masih berdekatan dengan skala 41 ZOM operasional. Sementara itu, Skenario 059 dan 087 yang menunjukkan kualitas statistik kompetitif tetap dipertahankan sebagai pembanding penting dalam interpretasi. Mengingat kriteria operasional awal pada penelitian ini masih bersifat semi-manual berbasis analisis dan kebutuhan pemetaan, pengembangan selanjutnya dapat menerapkan fungsi objektif matematis yang terukur untuk mengoptimalkan bobot antara validitas klaster, homogenitas, fragmentasi, dan kompleksitas operasional secara eksplisit.

3.5 Regionalisasi ZOM

Tahap penyempurnaan spasial terhadap 54 zona awal dari Skenario 60 mengevaluasi keberadaan klaster berukuran kecil atau terfragmentasi berdasarkan kontinuitas wilayah dan kemiripan profil curah hujan dasarian. Proses koreksi spasial ini menghasilkan reduksi bersih sebanyak empat zona, sehingga terbentuk peta final yang terdiri atas 50 ZOM berurutan. Perbandingan struktur spasial antara klasifikasi referensi 41 ZOM operasional dan peta final 50 ZOM hasil koreksi Skenario 60 masing-masing disajikan pada Gambar 4 dan Gambar 5. Dalam kerangka analitis penelitian ini, nilai *Silhouette Score* sebesar 0,5633 dan HI sebesar 7,239 tetap ditempatkan sebagai metrik evaluasi pada tahap kandidat 54 zona, sedangkan mosaik 50 ZOM merupakan keluaran kartografis dari tahap finalisasi spasial.

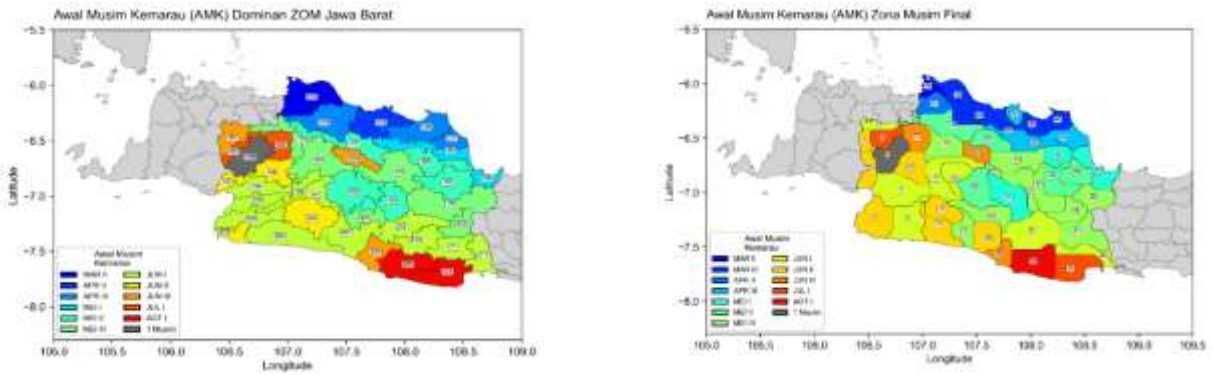


Gambar 4. Peta ZOM Operasional di Jawa Barat (41 ZOM)

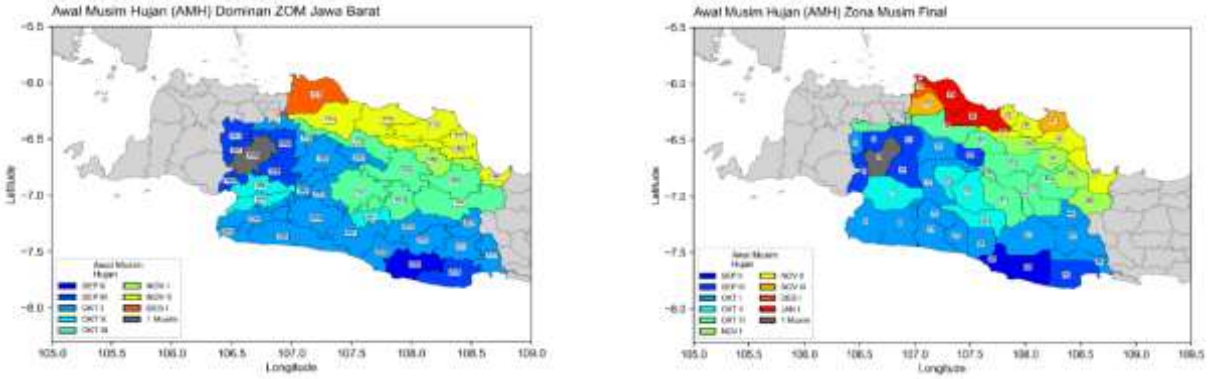


Gambar 5. Peta Final ZOM Jawa Barat Hasil Koreksi Skenario 60 (50 ZOM Berurutan)

Peta final pada Gambar 5 memperlihatkan mosaik 50 ZOM yang kontinu di seluruh wilayah Jawa Barat dengan pelabelan berurutan untuk memudahkan penelusuran karakteristik setiap zona. Mekanisme perubahan batas zona pada peta tersebut tercermin antara lain melalui pola pemecahan eks-ZOM-173 serta konsolidasi ZOM-190, ZOM-198, dan ZOM-199 yang menerjemahkan hasil klasterisasi menjadi zonasi spasial yang lebih rinci. Selain itu, perbandingan pola awal musim kemarau dan awal musim hujan antara ZOM operasional dan ZOM baru pada Gambar 6 memperlihatkan kesesuaian gradien regional dari utara ke selatan disertai variasi lokal yang lebih terperinci pada sejumlah wilayah. Kesamaan pola regional ini mendukung konsistensi klimatologis tingkat pertama, sedangkan evaluasi lanjutan diarahkan pada perhitungan metrik final berbasis label grid 50 ZOM, pengujian kesesuaian eksternal menggunakan *Adjusted Rand Index* (ARI), serta validasi berbasis data stasiun pengamatan.



(a) (b)



(c) (d)

Gambar 6. Perbandingan awal musim kemarau pada ZOM operasional (a)dan ZOM baru (b); Perbandingan awal musim hujan pada ZOM operasional (c) dan ZOM baru (d)

3.6 Pembahasan

Sensitivitas jumlah klaster terhadap parameter *min_dist* menunjukkan pengaruh kuat konfigurasi UMAP terhadap hasil regionalisasi curah hujan. Perubahan jumlah klaster dari 4 zona pada *min_dist* = 0,5 hingga puluhan zona pada nilai yang lebih kecil memperlihatkan pentingnya evaluasi stabilitas parameter dan *random seed* sebelum menarik interpretasi klimatologis. Pemilihan nilai *n_neighbors* = 5 pada Skenario 60 menekankan pelestarian struktur lokal sehingga selaras dengan tujuan menangkap perubahan curah hujan yang tajam pada jarak pendek. Di sisi lain, tingginya proporsi skenario yang menyisakan grid berlabel *noise* menegaskan keberadaan area transisi kepadatan di dalam ruang *embedding*. Dalam konteks regionalisasi spasial, penempatan grid transisi melalui tahap pascapemrosesan *absorption* memastikan kontinuitas batas zona, yang interpretasi klimatologisnya akan semakin kuat apabila polanya terbukti konsisten antarseed serta didukung oleh data observasi atau variabel fisik tambahan. Pola pemeringkatan pada Tabel 2 turut menjelaskan mekanisme mengapa konfigurasi *umap_n_neighbors* = 5 secara konsisten menghasilkan *Silhouette Score* yang lebih tinggi dibandingkan nilai 3. Pada ukuran ketetanggaan yang sangat kecil, graf kemiripan lokal UMAP menjadi terlalu sensitif terhadap variasi mikro antargrid sehingga HDBSCAN memecah domain menjadi banyak klaster kecil, sebagaimana tercermin dari terbentuknya 69–141 ZOM pada *umap_n_neighbors* = 3. Sebaliknya, nilai *umap_n_neighbors* = 5 memberikan keseimbangan yang lebih proporsional antara kepekaan terhadap variasi lokal dan kestabilan struktur klaster. Mekanisme serupa juga tampak pada parameter *hdbscan_min_cluster_size*, di mana nilai 10 pada Skenario 60 mampu mempertahankan detail spasial sekaligus membatasi proliferasi klaster berukuran kecil, sedangkan penggunaan nilai 20–30 menghasilkan 20–34 ZOM dengan representasi spasial yang lebih ringkas. Perbandingan pada tahap kandidat menunjukkan bahwa Skenario 60 (54 zona awal) menghasilkan nilai HI sebesar 7,239, lebih rendah dibandingkan *baseline* 41 ZOM operasional yang memiliki nilai HI sebesar 8,464. Penurunan nilai ini mengindikasikan berkurangnya dispersi absolut curah hujan intrazona pada hasil pengelompokan UMAP–HDBSCAN. Mengingat HI dinyatakan dalam satuan mm/dasarian, wilayah dengan curah hujan tinggi secara alami berpeluang memiliki simpangan baku yang lebih besar, sehingga pembacaan HI harus selalu disandingkan secara terpadu dengan *Silhouette Score*, jumlah zona, dan keutuhan struktur spasial (Liu *et al.*, 2010; Zhang *et al.*, 2016).

Transformasi dari 41 ZOM operasional menjadi 50 ZOM pada peta final selanjutnya menunjukkan peningkatan granularitas spasial sebanyak sembilan zona, yang mengindikasikan bahwa sejumlah subwilayah di Jawa Barat memiliki variasi lokal yang layak ditinjau secara lebih terperinci. Kasus pemecahan eks-ZOM-173 serta konsolidasi ZOM-190, ZOM-198, dan ZOM-199 memperlihatkan dua mekanisme utama perubahan batas zona, yakni pemisahan wilayah transisi dan penggabungan zona berkarakter serupa yang berkaitan erat dengan perbedaan fase musim setempat. Nilai guna operasional dari penambahan zona tersebut perlu dikonfirmasi bersama pengguna layanan, data observasi stasiun, dan prosedur baku BMKG agar peta final ini dapat berfungsi sebagai kandidat regionalisasi dalam tahap validasi operasional, sekaligus memisahkan kontribusi sinyal klimatologis dari sensitivitas algoritmik di seluruh wilayah Jawa Barat. Secara metodologis, temuan penelitian ini mendukung penerapan kerangka evaluasi berlapis dalam regionalisasi curah hujan, di mana *Silhouette Score* mengukur validitas struktur internal klaster, HI merangkum dispersi absolut curah hujan intrazona, dan jumlah zona merepresentasikan tingkat kompleksitas pengelolaan

operasional. Kombinasi ketiga indikator tersebut mencegah pemilihan konfigurasi yang hanya bergantung pada satu metrik statistik (Liu *et al.*, 2010). Dari sisi operasional, kerangka UMAP–HDBSCAN dapat dimanfaatkan sebagai perangkat analitis pendukung untuk mengidentifikasi struktur nonlinier dan kepadatan lokal secara sistematis dalam proses peninjauan maupun pemutakhiran ZOM, dengan tetap mensyaratkan konfirmasi melalui data stasiun pengamatan, penilaian ahli klimatologi, serta kajian dampak perubahan jumlah zona terhadap diseminasi informasi musim (Bieniek *et al.*, 2012; Luo *et al.*, 2024). Mengingat ruang lingkup penelitian ini masih terbatas pada profil klimatologi curah hujan 36 dasarian periode 1991–2020 di satu provinsi, pengembangan selanjutnya dapat mengintegrasikan variabel fisik dan dinamika atmosfer tambahan seperti topografi, sirkulasi monsun, *El Niño–Southern Oscillation* (ENSO), *Indian Ocean Dipole* (IOD), dan *Madden–Julian Oscillation* (MJO). Selain itu, sifat stokastik pada *embedding* UMAP dan sensitivitasnya terhadap hiperparameter menempatkan pengujian stabilitas antarseed sebagai prioritas validasi ketahanan model. Untuk menyempurnakan aspek evaluasi, penggunaan HI berbasis dispersi absolut (mm/dasarian) dapat dilengkapi dengan ukuran dispersi relatif atau standardisasi profil guna mengendalikan pengaruh perbedaan tingkat kebasahan antarwilayah, sedangkan peta final 50 ZOM hasil penyempurnaan kartografis perlu dilengkapi dengan perhitungan ulang *Silhouette Score* dan HI berbasis label grid final serta pengujian kesesuaian eksternal menggunakan *Adjusted Rand Index* (ARI) terhadap data stasiun pengamatan.

4 Kesimpulan

Penelitian ini mengevaluasi regionalisasi Zona Musim (ZOM) di Jawa Barat melalui kombinasi UMAP dan HDBSCAN pada profil klimatologi curah hujan 36 dasarian periode 1991–2020. Dari 360 kombinasi hiperparameter yang diuji, sebanyak 324 konfigurasi memenuhi batas relasi parameter dan 31 kandidat mencapai nilai *Silhouette Score* di atas 0,51. Skenario 60 menghasilkan 54 zona awal dengan *Silhouette Score* sebesar 0,5633 dan *Heterogeneity Index* (HI) sebesar 7,239, lebih rendah dibandingkan *baseline* 41 ZOM operasional yang memiliki nilai HI sebesar 8,464 pada tahap kandidat sebelum penyempurnaan spasial. Tahap penyempurnaan spasial selanjutnya menghasilkan peta final terkoreksi yang terdiri atas 50 ZOM berurutan. Peta final tersebut mempertahankan gradien musim utama di Jawa Barat sekaligus memberikan representasi batas zona yang lebih terperinci pada sejumlah subwilayah, sebagaimana tergambar melalui ilustrasi klimatologis pemecahan eks-ZOM-173 serta konsolidasi ZOM-190, ZOM-198, dan ZOM-199. Temuan ini menempatkan kerangka UMAP–HDBSCAN sebagai pendekatan eksploratif berbasis data yang efektif untuk mendukung peninjauan dan pemutakhiran ZOM. Dalam penerapannya, nilai HI diinterpretasikan sebagai ukuran dispersi absolut curah hujan pada dataset klimatologis yang sama dengan tetap mempertimbangkan sensitivitas parameter serta kebutuhan validasi operasional. Oleh karena itu, pengembangan penelitian selanjutnya difokuskan pada perhitungan metrik evaluasi untuk label 50 zona final, pengujian stabilitas antarseed, pengukuran kesesuaian eksternal menggunakan *Adjusted Rand Index* (ARI), validasi independen terhadap data stasiun pengamatan, serta formalisasi fungsi objektif kriteria operasional agar proses seleksi konfigurasi menjadi semakin terukur.

Referensi

Badan Meteorologi, Klimatologi, dan Geofisika. (2022). *Pemutakhiran Zona Musim Indonesia Periode 1991–2020*. BMKG. https://iklim.bmkg.go.id/bmkgadmin/storage/buletin/Buku_ZOM9120_versi_cetak.pdf

Bieniek, P. A., Bhatt, U. S., Thoman, R. L., Angeloff, H., Partain, J., Papineau, J., Fritsch, F., Holloway, E., Walsh, J. E., Daly, C., Shulski, M., Hufford, G., Hill, D. F., Calos, S., & Gens, R. (2012). Climate divisions for Alaska based on objective methods. *Journal of Applied Meteorology and Climatology*, *51*(7), 1276–1289. https://doi.org/10.1175/JAMC-D-11-0168.1

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data*, *10*(1), Article 5, 1–51. https://doi.org/10.1145/2733381

Hasana, S., & Fitrianah, D. (2023). A study on enhanced spatial clustering using ensemble DBSCAN and UMAP to map fire zone in Greater Jakarta, Indonesia. *Jurnal Riset Informatika*, *5*(3), 409–418.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.

King, F., Pettersen, C., Dolan, B., Shates, J., & Posselt, D. (2025). Decoding global precipitation processes and particle evolution using unsupervised learning. *Science Advances*, *11*(38), eadu0162.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE International Conference on Data Mining* (pp. 911–916). IEEE.

Luo, X., Frazier, A. G., Diaz, H. F., Longman, R., & Giambelluca, T. W. (2024). Routine climate monitoring in the state of Hawai'i: Establishment of state climate divisions. *Bulletin of the American Meteorological Society*, *105*(6), E1045–E1061.

Madani, I., & Wahid, K. A. (2022). Pemetaan zona agroklimat Oldeman di Provinsi Jawa Tengah menggunakan data CHIRPS. *Jurnal Geosains dan Remote Sensing*, *3*(2), 95–102.

Mazumdar, A. B. (2009). Precipitation regionalization of southwest monsoon by hierarchical cluster analysis. *MAUSAM*, *60*(2), 185–196. https://doi.org/10.54302/mausam.v60i2.1069

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for dimension reduction*. arXiv. https://doi.org/10.48550/arXiv.1802.03426

Rachmawati, R. N. (2021). Estimation of extreme rainfall patterns using generalized linear mixed model for spatio-temporal data in West Java, Indonesia. *Procedia Computer Science*, *179*, 330–336.

Rao, A. R., & Srinivas, V. V. (2006). Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, *318*(1–4), 37–56.

Rinaldi, A., Djuraidah, A., & Wigena, A. H. (2025). A Bayesian approach to spatio-temporal extreme rainfall modeling: Insights from West Java. *International Journal on Advanced Science, Engineering and Information Technology*, *15*(4), 1029–1039.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Wu, X., Jin, X., Li, Y., Yu, X., Feng, Q., Ren, M., & Wang, W. (2024). Regionalization, decomposition, and information mining for precipitation: A novel framework based on multivariate SOM clustering and multiscale features. *Journal of Applied Meteorology and Climatology*, *63*(10), 1157–1170. https://doi.org/10.1175/JAMC-D-23-0216.1

Zhang, Y., Moges, S., & Block, P. (2016). Optimal cluster analysis for objective regionalization of seasonal precipitation in regions of high spatial–temporal variability: Application to western Ethiopia. *Journal of Climate*, *29*(10), 3697–3717. https://doi.org/10.1175/JCLI-D-15-0582.1