

VERIFIKASI KATEGORIAL CURAH HUJAN MODEL IFS DI PROVINSI BALI

¹⁾Muhammad Jouhar Syah, ²⁾Aditya Mulya

¹⁾ Program Studi Sarjana Terapan Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia, e-mail : syahjohar32@gmail.com

²⁾ Program Studi Sarjana Terapan Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia, e-mail : aditya.mulya@stmkg.ac.id

Abstrak

Prakiraan curah hujan dari model numerik perlu diverifikasi pada skala lokal sebelum digunakan dalam layanan operasional. Penelitian ini mengevaluasi kemampuan *Integrated Forecasting System* (IFS) dalam mendeteksi kejadian hujan di Provinsi Bali menggunakan verifikasi kategorial. Data IFS dipasangkan dengan observasi per jam dari 22 stasiun selama 1 Januari 2021 sampai 31 Desember 2025. Kejadian hujan ditetapkan ketika curah hujan melebihi $0,1 \text{ mm jam}^{-1}$. Kinerja dinilai melalui *Probability of Detection* (POD), *False Alarm Ratio* (FAR), *Critical Success Index* (CSI), *Equitable Threat Score* (ETS), dan bias frekuensi, kemudian dianalisis menurut musim, topografi, bulan, dan stasiun. Sebanyak 862.562 pasangan data valid menghasilkan POD 0,602, FAR 0,842, CSI 0,143, ETS 0,074, dan bias 3,820. Model menangkap 60,2% kejadian hujan, tetapi memprakirakan hujan sekitar 3,8 kali lebih sering daripada observasi. Kinerja lebih baik pada musim hujan daripada musim kemarau, dengan CSI masing-masing 0,170 dan 0,103. Dataran tinggi memberikan CSI tertinggi sebesar 0,200 dan FAR terendah sebesar 0,773. Hasil ini menunjukkan bahwa IFS masih memiliki informasi keterampilan, tetapi memerlukan kalibrasi kejadian hujan dan evaluasi berbasis ambang sebelum digunakan sebagai dasar keputusan operasional di Bali.

Kata Kunci: curah hujan, IFS, musim, topografi, verifikasi kategorial

Abstract

Numerical rainfall forecasts require local verification before operational use. This study evaluates the ability of the *Integrated Forecasting System* (IFS) to detect rainfall events in Bali Province using categorical verification. Hourly IFS data were paired with observations from 22 stations from 1 January 2021 to 31 December 2025. Rain was defined as precipitation exceeding 0.1 mm h^{-1} . Performance was assessed using the *Probability of Detection* (POD), *False Alarm Ratio* (FAR), *Critical Success Index* (CSI), *Equitable Threat Score* (ETS), and bias frequency, followed by seasonal, topographic, monthly, and station-based analyses. A total of 862,562 valid pairs produced POD of 0.602, FAR of 0.842, CSI of 0.143, ETS of 0.074, and bias of 3.820. The model detected 60.2% of observed rain events but forecast rain about 3.8 times more frequently than observed. Performance was better during the rainy season than the dry season, with CSI values of 0.170 and 0.103, respectively. Highland stations produced the highest CSI of 0.200 and the lowest FAR of 0.773. The IFS retains positive skill, but event calibration and threshold-based evaluation are required before its rainfall output supports operational decisions in Bali.

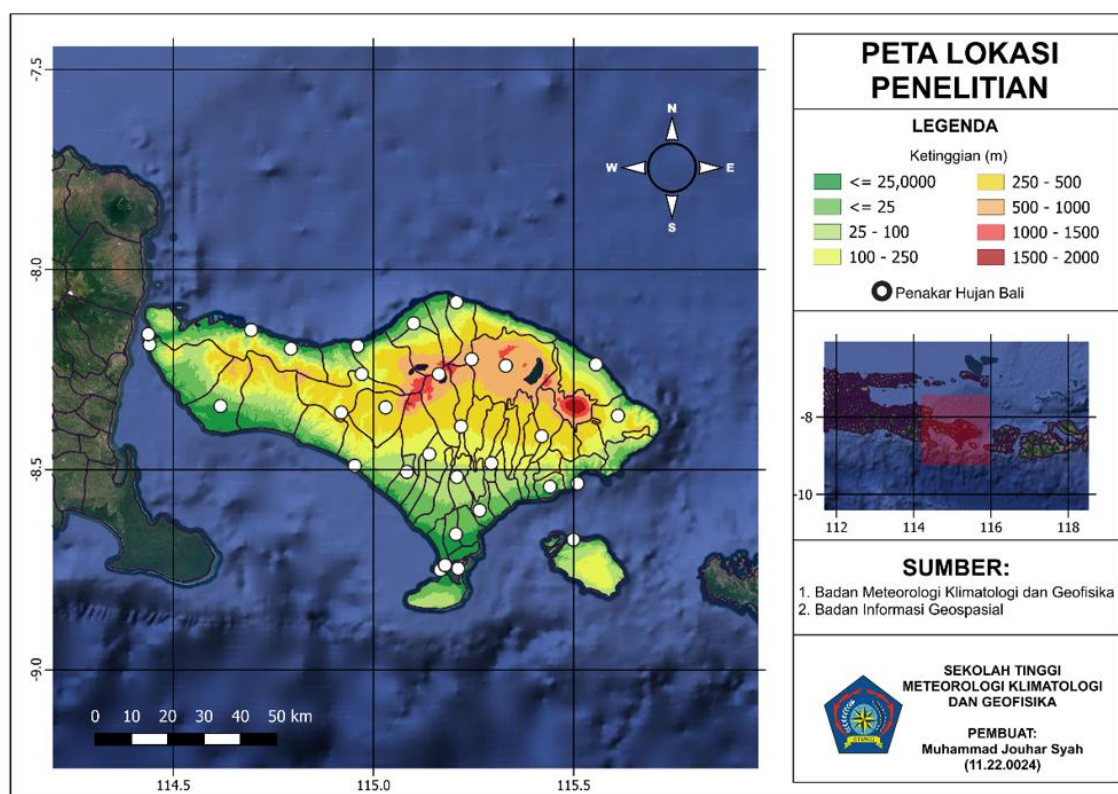
Keywords: categorical verification, IFS, rainfall, season, topography

I. PENDAHULUAN

Curah hujan di Indonesia memiliki variasi spasial dan temporal yang kuat karena dipengaruhi oleh sistem monsun, interaksi darat-laut, proses konvektif, dan kondisi topografi. Bali berada pada wilayah selatan Indonesia yang secara umum mengikuti pola hujan monsun, tetapi distribusi hujannya tetap berbeda antarwilayah karena pulau ini memiliki pantai, dataran rendah, dan rangkaian pegunungan dalam jarak yang relatif pendek [1]. Pemutakhiran Zona Musim BMKG juga menunjukkan bahwa Provinsi Bali terbagi ke dalam beberapa zona musim dengan waktu awal, panjang, dan puncak musim yang tidak sepenuhnya seragam [2]. Kondisi tersebut menuntut prakiraan curah hujan yang mampu merepresentasikan perubahan kejadian hujan pada skala waktu dan lokasi yang rinci.

Model prediksi cuaca numerik menyediakan informasi atmosfer secara konsisten pada wilayah yang luas. Salah satu model yang digunakan dalam kegiatan prakiraan adalah *Integrated Forecasting System* (IFS) dari European Centre for Medium-Range Weather Forecasts. IFS menggabungkan dinamika atmosfer, parameterisasi proses fisik, dan asimilasi data untuk menghasilkan prakiraan global. Pada konfigurasi resolusi tinggi yang menjadi sumber data penelitian, jarak grid horizontal berada di sekitar 9 km. Skala tersebut jauh lebih luas daripada cakupan pengukuran alat hujan pada satu titik. Perbedaan skala antara grid model dan observasi titik dapat menimbulkan ketidaksesuaian, terutama ketika hujan bersifat lokal dan berumur singkat. Metode verifikasi spasial menunjukkan bahwa prakiraan presipitasi dapat memiliki informasi yang berguna walaupun tidak tepat berimpit dengan observasi pada satu titik [15]-[18], [20].

Kompleksitas Bali memperbesar kebutuhan verifikasi lokal. Penelitian Rahmawati dan Lubczynski menunjukkan bahwa produk curah hujan berbasis grid memiliki kinerja yang berubah menurut musim, zona klimatologis, dan kondisi elevasi di Bali [3]. Hasil tersebut menegaskan bahwa penilaian pada skala pulau saja belum cukup. Evaluasi perlu menguji kemampuan model pada musim hujan dan kemarau, kelompok topografi, serta masing-masing stasiun. Peta lokasi penelitian dan sebaran titik pengamatan ditunjukkan pada Gambar 1.



Gambar 1. Peta lokasi penelitian dan sebaran stasiun pengamatan curah hujan di Provinsi Bali.

Dokumentasi IFS menjelaskan bahwa presipitasi dihasilkan melalui interaksi skema konveksi, mikrofisika awan, dan proses stratiform [4]. Representasi tersebut dapat menghasilkan bias sistematis. Lavers, Harrigan, dan Prudhomme menemukan bias presipitasi IFS yang berbeda antarwilayah dan menunjukkan adanya kecenderungan hujan berlebih pada sebagian wilayah Asia Tenggara [5]. Evaluasi di Tiongkok juga memperlihatkan bahwa kemampuan IFS berubah menurut intensitas, lokasi, dan periode prakiraan [6]. Temuan itu memperkuat kebutuhan pengujian menggunakan observasi lokal, terutama untuk kejadian dikotomis hujan dan tidak hujan.

Verifikasi kategorial mengubah prakiraan dan observasi menjadi kategori kejadian, kemudian menghitung hit, false alarm, miss, dan correct negative. Pendekatan ini relevan bagi layanan operasional karena menjawab dua pertanyaan dasar: seberapa banyak kejadian hujan yang berhasil dideteksi dan seberapa sering model memberikan prakiraan hujan yang tidak teramati. Penggunaan beberapa skor secara bersamaan penting karena satu skor tidak dapat menggambarkan seluruh atribut kualitas prakiraan [7], [8]. Nilai POD yang tinggi, misalnya, dapat dicapai dengan memprakirakan hujan terlalu sering; kondisi tersebut terlihat melalui FAR dan bias frekuensi.

Penelitian ini memusatkan analisis pada verifikasi kategorial curah hujan IFS terhadap observasi di Provinsi Bali. Analisis menggunakan pasangan data per jam selama lima tahun dan membandingkan performa secara keseluruhan, antarmusim, antarbulan, antarkelompok topografi, dan antarstasiun. Tujuan penelitian adalah mengidentifikasi kemampuan deteksi, tingkat alarm palsu, keterampilan relatif terhadap tebakan acak, serta pola bias frekuensi model. Hasil penelitian diharapkan memberi dasar empiris untuk penggunaan, kalibrasi, dan pengembangan pascaproses keluaran IFS di wilayah Bali.

II. METODE

A. Data dan Wilayah Penelitian

Penelitian menggunakan desain evaluatif kuantitatif. Data model yang digunakan merupakan produk prakiraan deterministik *Integrated Forecasting System-High Resolution Forecast* (IFS-HRES) ECMWF dalam format GRIB. Data memiliki resolusi spasial $0,08^\circ \times 0,08^\circ$ atau sekitar 9 km di ekuator dan resolusi temporal satu jam. Penelitian ini menggunakan satu siklus inisialisasi, yaitu run 12 UTC, dengan lead time +1 sampai +24 jam. Parameter presipitasi yang digunakan adalah total precipitation (TP), kemudian curah hujan interval satu jam dihitung dari selisih nilai TP pada dua lead time berurutan.

Nilai model pada setiap stasiun diekstraksi menggunakan metode *nearest neighbor*, yaitu memilih titik grid IFS yang memiliki jarak paling dekat terhadap koordinat stasiun tanpa interpolasi bilinear. Setelah ekstraksi, valid time model dicocokkan dengan waktu observasi dalam UTC agar setiap pasangan data merepresentasikan waktu valid yang sama. Data pembanding berasal dari pengamatan curah hujan otomatis BMKG di Provinsi Bali.

Rekap kelengkapan dapat dilihat pada **Tabel 1**, yang memuat 30 identitas stasiun, dengan 22 stasiun berstatus lolos dan digunakan dalam analisis Stasiun yang dinyatakan lolos merupakan stasiun yang telah dilakukan *quality control* yang setelah itu dilakukan pengecekan stasiun yang harus mencakup 80% data untuk menjaga kestabilan verifikasi. Periode efektif data dimulai pada 1 Januari 2021 pukul 13.00 dan berakhir pada 31 Desember 2025 pukul 23.00. Setiap baris dataset memuat waktu, curah hujan observasi, curah hujan IFS, musim, nama stasiun, kelas topografi, dan elevasi DEMNAS. Ringkasan data ditunjukkan pada **Tabel 2**.

Stasiun dikelompokkan menjadi daerah pesisir, dataran rendah, dan dataran tinggi berdasarkan elevasi titik. Kelompok pesisir mencakup elevasi 0-100 m, dataran rendah 100-500 m, dan dataran tinggi lebih dari 500 m. Musim hujan ditetapkan pada November-April dan musim kemarau pada Mei-Oktober. Pembagian ini digunakan sebagai stratifikasi analisis, bukan sebagai penetapan awal musim pada setiap Zona Musim. Karena awal dan akhir musim dapat bergeser antarwilayah dan antartahun, interpretasi hasil musiman dibatasi pada perbandingan dua kelompok kalender tersebut.

Tabel 1. Tabel Kelengkapan Data Stasiun Berdasarkan Elevasi dan Klasifikasi Topografi

Kode Stasiun	Jam Valid	Jam Total	Kelengkapan (%)	Elevasi DEMNAS (m)	Klasifikasi Topografi	Status
STA0138	42.181	43.824	96,25	1278,20	Dataran Tinggi	LOLOS
STA3056	41.559	43.824	94,83	294,38	Dataran Rendah	LOLOS
STA0140	41.422	43.824	94,52	10,45	Daerah Pesisir	LOLOS
150029	41.228	43.824	94,08	21,09	Daerah Pesisir	LOLOS
STA3026	41.217	43.824	94,05	32,45	Daerah Pesisir	LOLOS
STA0136	40.895	43.824	93,32	184,33	Dataran Rendah	LOLOS
STA0139	40.832	43.824	93,17	771,30	Dataran Tinggi	LOLOS
STA0067	40.609	43.824	92,66	1481,83	Dataran Tinggi	LOLOS
STA0141	40.551	43.824	92,53	37,92	Daerah Pesisir	LOLOS
150027	40.494	43.824	92,40	148,92	Dataran Rendah	LOLOS
14063121	40.258	43.824	91,86	13,52	Daerah Pesisir	LOLOS
STA5033	40.243	43.824	91,83	2,94	Daerah Pesisir	LOLOS
STA0157	39.649	43.824	90,47	44,96	Daerah Pesisir	LOLOS
150028	39.434	43.824	89,98	73,60	Daerah Pesisir	LOLOS
150317	39.424	43.824	89,96	5,53	Daerah Pesisir	LOLOS
STA0156	39.345	43.824	89,78	302,77	Dataran Rendah	LOLOS
150318	38.022	43.824	86,76	266,04	Dataran Rendah	LOLOS
STA0137	37.596	43.824	85,79	1257,78	Dataran Tinggi	LOLOS
STA2079	36.991	43.824	84,41	17,73	Daerah Pesisir	LOLOS
STA0177	36.731	43.824	83,81	36,04	Daerah Pesisir	LOLOS
STA0066	36.523	43.824	83,34	89,97	Daerah Pesisir	LOLOS
STA3055	36.505	43.824	83,30	569,57	Dataran Tinggi	LOLOS
STA2128	33.841	43.824	77,22	1,27	Daerah Pesisir	ELIMINASI
150030	33.659	43.824	76,80	166,76	Dataran Rendah	ELIMINASI
STA0068	33.024	43.824	75,36	526,49	Dataran Tinggi	ELIMINASI
STA0115	31.805	43.824	72,57	9,01	Daerah Pesisir	ELIMINASI
150026	31.769	43.824	72,49	36,50	Daerah Pesisir	ELIMINASI
STW1013	31.387	43.824	71,62	7,60	Daerah Pesisir	ELIMINASI
STW1021	26.073	43.824	59,49	1120,72	Dataran Tinggi	ELIMINASI
STA2297	1.38	43.824	3,15	1,72	Daerah Pesisir	ELIMINASI

Tabel 2. Ringkasan data dan rancangan analisis.

Komponen	Spesifikasi
Data model	Curah hujan IFS Lead 1-24 pada titik stasiun, resolusi waktu 1 jam
Data observasi	Curah hujan otomatis BMKG pada 22 stasiun
Periode efektif	1 Januari 2021 13.00-31 Desember 2025 23.00
Jumlah pasangan valid	862.562 pasangan waktu-stasiun
Ambang hujan	$> 0,1 \text{ mm jam}^{-1}$
Stratifikasi	Musim, bulan, topografi, dan stasiun

B. Prapengolahan dan Pembentukan Pasangan Data

Waktu data model dan observasi diseragamkan sebelum proses pencocokan. Setiap nilai IFS dipasangkan dengan observasi pada stasiun dan cap waktu yang sama. Baris yang memiliki nilai hilang pada salah satu sumber dikeluarkan dari perhitungan. Nilai hilang tidak diubah menjadi kategori tidak hujan karena perlakuan tersebut akan menambah *correct negative* secara artifisial dan meningkatkan akurasi secara semu. Pemeriksaan juga dilakukan terhadap identitas stasiun, konsistensi koordinat, elevasi, serta label musim dan topografi.

Kejadian hujan didefinisikan ketika curah hujan lebih besar dari $0,1 \text{ mm jam}^{-1}$. Nilai yang sama dengan atau lebih kecil dari ambang dimasukkan ke kategori tidak hujan. Ambang rendah dipilih untuk menguji kemampuan dasar model dalam membedakan keberadaan hujan. Konsekuensinya, hasil penelitian menilai frekuensi kejadian dan tidak langsung menggambarkan ketepatan intensitas hujan. Frekuensi presipitasi ringan yang terlalu besar merupakan masalah yang telah dilaporkan pada model global [13], [14], sehingga pengujian ambang hujan sedang, lebat, atau ekstrem tetap memerlukan analisis terpisah.

C. Tabel Kontingensi dan Skor Verifikasi

Pasangan kategori prakiraan dan observasi membentuk tabel kontingensi 2×2 seperti pada **Tabel 3**.

Tabel 3. Tabel kontingensi kejadian hujan.

	Observasi: Hujan	Observasi: Tidak Hujan
IFS: Hujan	Hit (a)	False alarm (b)
IFS: Tidak Hujan	Miss (c)	Correct negative (d)

Simbol a menunjukkan hit, yaitu model dan observasi sama-sama menyatakan hujan. Simbol b menunjukkan false alarm, yaitu model menyatakan hujan tetapi observasi tidak hujan. Simbol c menunjukkan miss, yaitu model menyatakan tidak hujan ketika observasi hujan. Simbol d menunjukkan *correct negative*, yaitu model dan observasi sama-sama menyatakan tidak hujan. Skor dihitung mengikuti kerangka verifikasi prakiraan kejadian dikotomis [8]-[10].

$$\text{POD} = a / (a + c) \quad (1)$$

$$\text{FAR} = b / (a + b) \quad (2)$$

$$\text{CSI} = a / (a + b + c) \quad (3)$$

$$a_{\text{ref}} = (a + b)(a + c) / n \quad (4)$$

$$\text{ETS} = (a - a_{\text{ref}}) / (a + b + c - a_{\text{ref}}) \quad (5)$$

$$\text{Bias} = (a + b) / (a + c) \quad (6)$$

Tabel 4. Interpretasi metrik verifikasi kategorial.

Metrik	Nilai ideal	Makna utama
POD	1	Proporsi hujan observasi yang terdeteksi
FAR	0	Proporsi prakiraan hujan yang salah
CSI	1	Keberhasilan setelah hit, miss, dan false alarm diperhitungkan
ETS	1; 0 = acak	Keterampilan setelah hit karena peluang dikoreksi
Bias	1	Rasio frekuensi hujan prakiraan terhadap observasi

Kegunaan masing-masing metrik ditunjukkan pada Tabel 4. POD mengukur sensitivitas terhadap kejadian hujan, sedangkan FAR mengukur ketidakandalan prakiraan hujan. CSI mengabaikan *correct negative* sehingga lebih sesuai untuk kejadian yang relatif jarang [10]. ETS memperbaiki CSI dengan mengurangi hit yang diperkirakan muncul secara acak. Bias frekuensi tidak menilai ketepatan lokasi atau waktu, tetapi menunjukkan apakah model terlalu sering atau terlalu jarang memprakirakan hujan. Bias lebih besar dari satu menunjukkan *overestimate*, sedangkan nilai kurang dari satu menunjukkan *underestimate*.

D. Analisis Musim, Topografi, Bulan, dan Stasiun

Skor dihitung pada lima tingkat agregasi. Pertama, seluruh pasangan data digabungkan untuk menggambarkan karakter umum model. Kedua, data dipisahkan menjadi musim hujan dan musim kemarau. Ketiga, skor dihitung pada setiap bulan untuk memeriksa perubahan siklus tahunan. Keempat, data dikelompokkan berdasarkan topografi. Kelima, skor dihitung pada setiap stasiun dan pada kombinasi stasiun-musim. Perbandingan menggunakan nilai skor dan komponen tabel kontingensi. Penelitian tidak menerapkan uji sebab-akibat antara elevasi dan skor; hubungan elevasi pada grafik sebar ditafsirkan secara deskriptif.

E. Interval Kepercayaan

Ketidakpastian skor agregat dihitung dengan bootstrap sebanyak 500 iterasi. Pada setiap iterasi, pasangan data diambil kembali dengan pengembalian, lalu POD, FAR, CSI, dan ETS dihitung ulang. Batas bawah dan atas interval kepercayaan 95% ditetapkan dari persentil 2,5 dan 97,5 distribusi bootstrap. Interval ini digunakan untuk menunjukkan kestabilan estimasi, bukan untuk menggantikan analisis dependensi temporal. Pasangan waktu yang berdekatan dapat berasal dari satu sistem hujan yang sama sehingga asumsi independensi penuh tidak selalu terpenuhi.

III. HASIL DAN PEMBAHASAN

A. Kinerja Keseluruhan

Dari 862.562 pasangan valid, observasi mencatat 70.522 kejadian hujan atau 8,18% dari seluruh pasangan. IFS memprakirakan 269.367 kejadian hujan atau 31,23%. Selisih frekuensi ini menunjukkan bahwa model menghasilkan cakupan kejadian hujan yang jauh lebih luas daripada observasi titik. Tabel kontingensi keseluruhan pada **Gambar 2** terdiri atas 42.477 hit, 226.890 false alarm, 28.045 miss, dan 565.150 correct negative.

Gambar 2. Tabel kontingensi verifikasi kategorial IFS secara keseluruhan.

	OBS: Hujan	OBS: Tidak Hujan
IFS: Hujan	42,477	226,890
IFS: Tidak Hujan	28,045	565,150

Tabel 5 menunjukkan skor kategorial secara keseluruhan dan berdasarkan musim. Nilai POD sebesar 0,602 menunjukkan bahwa model mendeteksi 60,2% hujan yang tercatat oleh stasiun. Interval kepercayaan 95% POD berada pada 0,599-0,605. Kemampuan deteksi ini tidak diikuti oleh ketepatan prakiraan hujan. FAR mencapai 0,842 dengan interval 0,841-0,844. Dengan kata lain, sekitar 84,2% seluruh prakiraan hujan tidak berpasangan dengan hujan pada observasi di titik dan waktu yang sama. Nilai CSI sebesar 0,143 menegaskan bahwa jumlah hit kecil dibandingkan gabungan *hit*, *false alarm*, dan *miss*.

FAR yang tinggi, terutama pada ambang sangat rendah $0,1 \text{ mm jam}^{-1}$, dapat mengindikasikan adanya drizzle bias, yaitu kecenderungan model menghasilkan presipitasi ringan terlalu sering. Stephens et al. [13] menunjukkan bahwa beberapa model cuaca dan iklim global, termasuk IFS, menghasilkan hujan lebih sering dan dengan intensitas yang terlalu ringan dibandingkan observasi; pola frekuensi hujan yang berlebihan juga dibahas oleh Sun et al. [14]. Namun, indikasi ini belum menjadi bukti langsung karena penelitian ini tidak menganalisis distribusi intensitas presipitasi.

ETS bernilai 0,074 dengan interval 0,073-0,075. Nilai positif menunjukkan bahwa model masih lebih baik daripada tebakan acak, tetapi besar keterampilannya rendah. Bias frekuensi sebesar 3,820 menunjukkan

frekuensi hujan prakiraan sekitar 3,8 kali frekuensi hujan observasi. Hasil tersebut konsisten dengan temuan bias presipitasi IFS di sebagian wilayah tropis [5]. Namun, penelitian ini tidak membandingkan akumulasi curah hujan total sehingga istilah *overestimate* pada artikel ini merujuk khusus pada frekuensi kejadian, bukan pada jumlah curah hujan.

Akurasi klasifikasi keseluruhan mencapai 0,704 karena *correct negative* berjumlah sangat besar. Nilai ini dapat menyesatkan ketika digunakan sendiri. Model dapat memperoleh akurasi tinggi hanya dengan benar mengenali banyak periode tidak hujan, walaupun kualitas prakiraan hujannya rendah. Oleh sebab itu, CSI, ETS, FAR, dan bias menjadi ukuran yang lebih informatif untuk tujuan deteksi hujan. Prinsip ini sejalan dengan rekomendasi verifikasi kejadian langka yang menekankan penggunaan beberapa atribut kualitas secara bersama [7]-[10], [19].

Tabel 5. Skor verifikasi keseluruhan dan berdasarkan musim.

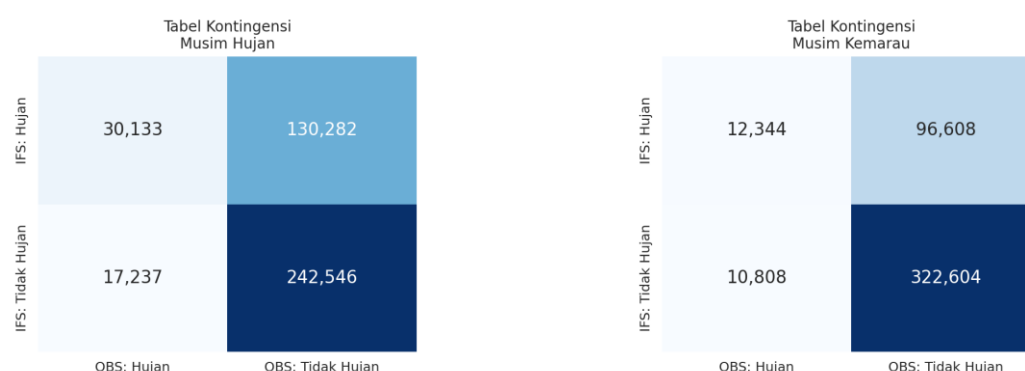
Agregasi	n	POD	FAR	CSI	ETS	Bias	Akurasi
Keseluruhan	862.562	0,602	0,842	0,143	0,074	3,820	0,704
Musim hujan	420.198	0,636	0,812	0,170	0,076	3,386	0,649
Musim kemarau	442.364	0,533	0,887	0,103	0,058	4,706	0,757

B. Perbedaan Kinerja Antarmusim

Musim hujan memiliki 420.198 pasangan data, sedangkan musim kemarau memiliki 442.364 pasangan. Pada musim hujan, observasi mencatat 47.370 kejadian hujan atau 11,27%. Pada musim kemarau, proporsinya turun menjadi 23.152 kejadian atau 5,23%. IFS tetap memprakirakan hujan pada 38,18% pasangan musim hujan dan 24,63% pasangan musim kemarau. Penurunan frekuensi prakiraan tidak sebesar penurunan frekuensi observasi. Akibatnya, bias meningkat dari 3,386 pada musim hujan menjadi 4,706 pada musim kemarau.

Gambar 3 memperlihatkan perubahan komponen tabel kontingensi menurut musim. Musim hujan menghasilkan 30.133 hit dan 17.237 miss. Musim kemarau menghasilkan 12.344 *hit* dan 10.808 *miss*. *False alarm* tetap tinggi pada kedua musim, yaitu 130.282 pada musim hujan dan 96.608 pada musim kemarau. Walaupun jumlah false alarm lebih kecil pada musim kemarau, proporsinya terhadap seluruh prakiraan hujan justru lebih besar karena jumlah hit menurun.

Gambar 3. Tabel kontingensi IFS pada musim hujan (kiri) dan musim kemarau (kanan).

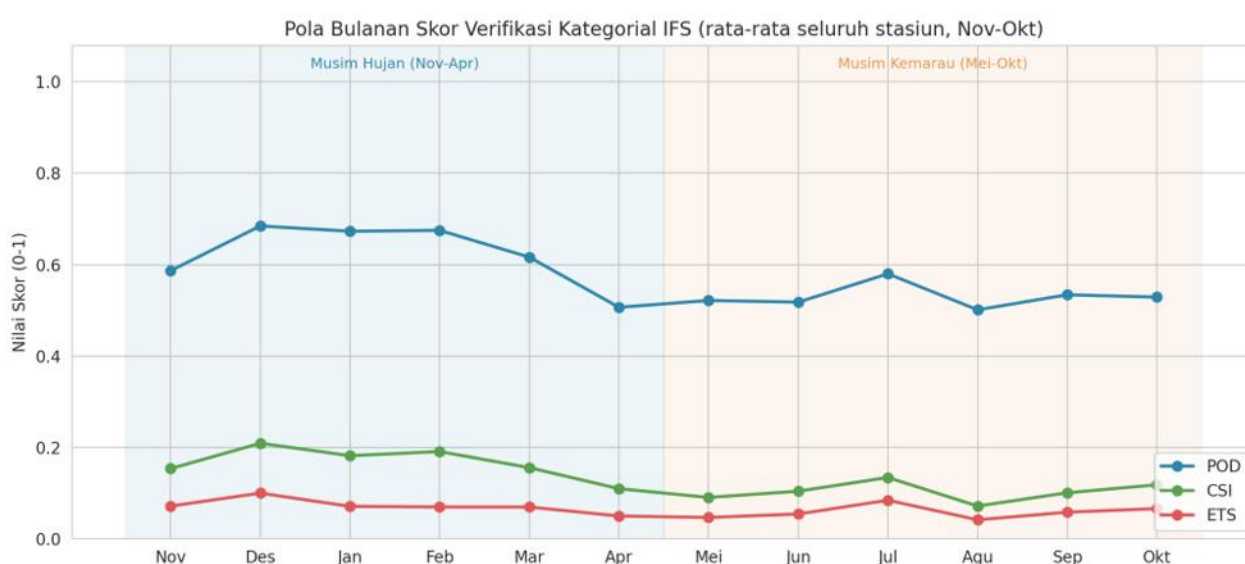


POD musim hujan sebesar 0,636 lebih tinggi daripada musim kemarau sebesar 0,533. FAR musim hujan mencapai 0,812 dan meningkat menjadi 0,887 pada musim kemarau. Perubahan tersebut menurunkan CSI dari 0,170 menjadi 0,103 dan ETS dari 0,076 menjadi 0,058. Jadi, kemampuan IFS membedakan kejadian hujan lebih baik ketika hujan lebih sering terjadi. Hasil ini searah dengan evaluasi produk hujan lain di Bali yang melaporkan kinerja lebih baik pada periode basah dibandingkan periode kering [3].

Tingginya FAR pada musim kemarau dapat dipengaruhi oleh base rate kejadian yang rendah; ukuran verifikasi kejadian biner diketahui sensitif terhadap kelangkaan kejadian [19]. Ketika hujan observasi jarang, sedikit perluasan area atau pergeseran waktu hujan pada grid model dapat menghasilkan banyak false alarm pada titik stasiun. Selain itu, konveksi lokal pada periode kering dapat berkembang dalam ruang yang sempit. Grid model dapat merepresentasikan sinyal atmosfer yang mendukung hujan, tetapi lokasi presipitasi aktual tidak selalu tepat mengenai stasiun. Karena penelitian ini tidak menggunakan verifikasi lingkungan spasial, kemungkinan *displacement error* belum dapat dipisahkan dari kesalahan pembentukan hujan [15]-[18], [20].

C. Pola Bulanan

Analisis bulanan memberi rincian tambahan terhadap perbedaan musiman. **Gambar 4** menunjukkan bahwa POD, CSI, dan ETS cenderung lebih tinggi pada November-Maret. Desember menjadi bulan dengan skor tertinggi, yaitu POD 0,685, CSI 0,209, dan ETS 0,100. Januari dan Februari juga menunjukkan CSI relatif tinggi, masing-masing 0,181 dan 0,190. Setelah Maret, skor menurun pada April dan tetap relatif rendah selama sebagian besar musim kemarau.



Gambar 4. Pola bulanan POD, CSI, dan ETS rata-rata seluruh stasiun.

Nilai CSI terendah terjadi pada Agustus sebesar 0,071, diikuti Mei sebesar 0,090 dan September sebesar 0,100. FAR pada Agustus, Mei, dan September masing-masing mencapai 0,923, 0,902, dan 0,890, sedangkan biasanya sebesar 6,543, 5,315, dan 4,861. Juli menunjukkan perbaikan sementara dengan CSI 0,134 dan ETS 0,084. Perbaikan ini tidak cukup untuk menyimpulkan pengaruh suatu fenomena atmosfer tertentu karena penelitian tidak mengelompokkan kasus berdasarkan ENSO, IOD, MJO, gangguan ekuatorial, atau kondisi sinoptik. Pola bulanan pada Gambar 4 hanya menunjukkan variasi dalam sampel penelitian.

April memperlihatkan transisi yang jelas. POD turun menjadi 0,506, FAR meningkat menjadi 0,878, dan CSI turun menjadi 0,109. Pola ini dapat menunjukkan bahwa model lebih sulit menempatkan hujan pada fase peralihan menuju musim kering. Meskipun periode penelitian mencakup lima tahun kalender penuh, jumlah pasangan valid tetap berbeda menurut bulan dan stasiun karena adanya data hilang, sehingga interpretasi pola bulanan perlu mempertimbangkan distribusi sampel tersebut.

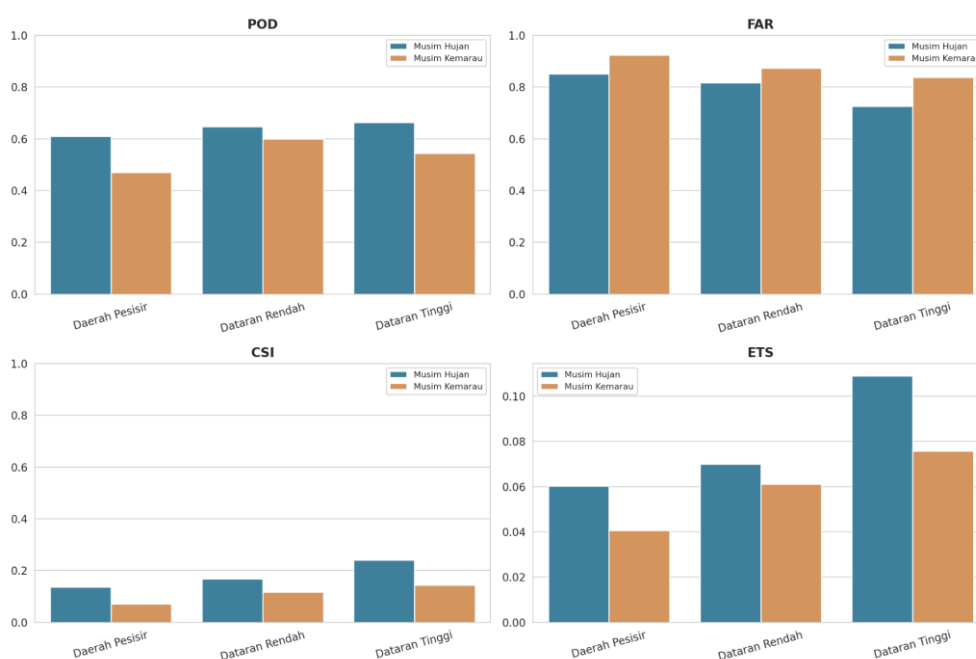
D. Perbedaan Kinerja Berdasarkan Topografi

Pengelompokan topografi menghasilkan 468.583 pasangan di daerah pesisir, 198.156 pasangan di dataran rendah, dan 195.823 pasangan di dataran tinggi. Secara keseluruhan, daerah pesisir memiliki POD 0,569, FAR 0,878, CSI 0,112, ETS 0,059, dan bias 4,660. Dataran rendah menghasilkan POD tertinggi sebesar 0,630, tetapi FAR masih mencapai 0,841 sehingga CSI hanya 0,145. Dataran tinggi menghasilkan POD 0,621, FAR 0,773, CSI 0,200, ETS 0,101, dan bias 2,734. Ringkasan skor ditunjukkan pada **Tabel 6**.

Tabel 6. Skor verifikasi menurut klasifikasi topografi.

Topografi	n	POD	FAR	CSI	ETS	Bias
Daerah pesisir	468.583	0,569	0,878	0,112	0,059	4,660
Dataran rendah	198.156	0,630	0,841	0,145	0,069	3,960
Dataran tinggi	195.823	0,621	0,773	0,200	0,101	2,734

Gambar 5 menunjukkan bahwa semua kelompok topografi memiliki kinerja lebih baik pada musim hujan. Kombinasi terbaik muncul pada dataran tinggi musim hujan dengan POD 0,663, FAR 0,726, CSI 0,241, ETS 0,109, dan bias 2,415. Pada musim kemarau, CSI dataran tinggi turun menjadi 0,144, tetapi tetap lebih tinggi daripada pesisir 0,071 dan dataran rendah 0,117. Bias musim kemarau terbesar terdapat di pesisir sebesar 6,111, disusul dataran rendah sebesar 4,730. Hasil ini menunjukkan bahwa masalah utama pada wilayah rendah dan pesisir bukan hanya kegagalan mendeteksi hujan, tetapi terlalu luasnya frekuensi prakiraan hujan.

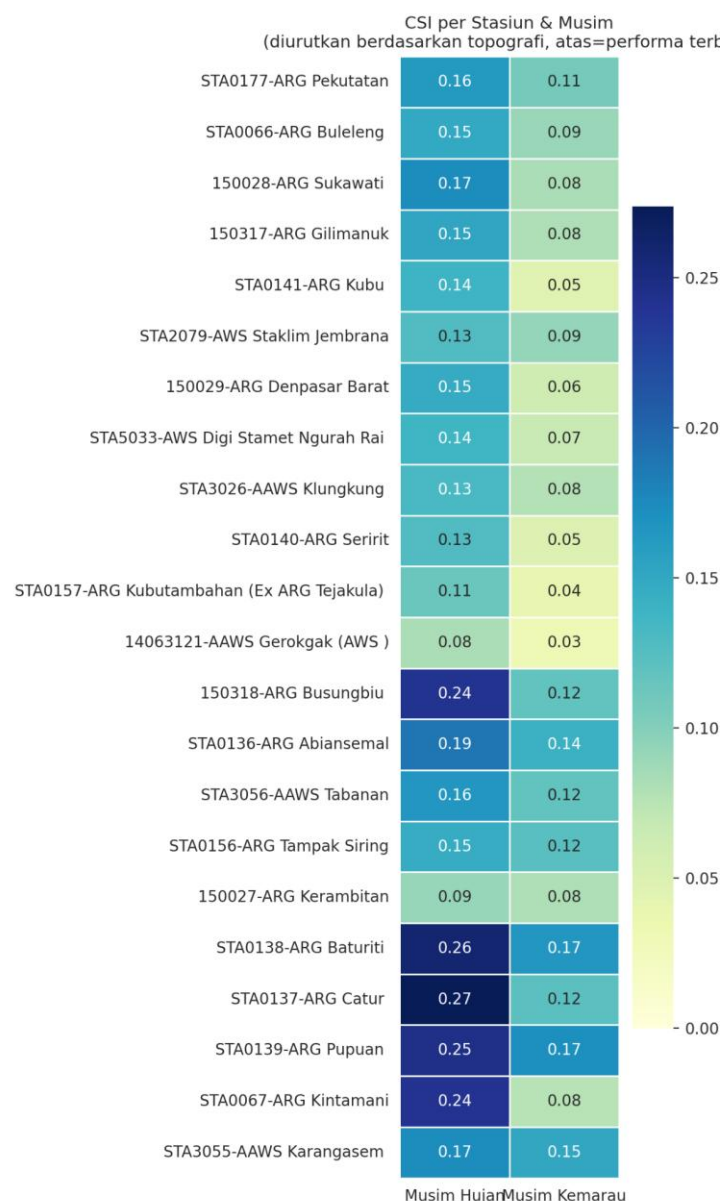
**Gambar 5.** Perbandingan POD, FAR, CSI, dan ETS berdasarkan topografi dan musim.

Kinerja dataran tinggi yang lebih baik perlu ditafsirkan secara hati-hati. Secara teoritis, topografi kompleks dapat meningkatkan kesalahan model karena gradien elevasi pada satu grid tidak mewakili kondisi titik. Namun, wilayah tinggi Bali juga lebih sering mengalami hujan dan memiliki sinyal orografis yang relatif menetap. Base rate yang lebih besar dapat meningkatkan peluang model dan observasi mencatat kejadian pada waktu yang sama. Rahmawati dan Lubczynski juga menemukan bahwa hubungan antara elevasi dan performa produk hujan berubah menurut musim dan produk yang dievaluasi [3]. Dengan demikian, hasil penelitian ini tidak berarti bahwa IFS telah merepresentasikan proses orografis secara sempurna.

Perbandingan pesisir dan dataran rendah menunjukkan POD yang cukup tinggi tetapi CSI rendah. Pola ini terjadi ketika model memprakirakan hujan secara luas untuk meningkatkan deteksi, lalu menghasilkan false alarm dalam jumlah besar. Wilayah pesisir juga dipengaruhi oleh kontras darat-laut, pergeseran sistem konvektif, dan kemungkinan grid campuran darat-laut. Verifikasi titik menghitung hujan model sebagai false alarm ketika hujan terjadi pada bagian lain grid tetapi tidak tercatat pada alat. Analisis berbasis lingkungan spasial atau radar diperlukan untuk membedakan false alarm fisik dari kesalahan posisi [15]-[18], [20].

E. Variasi Antarstasiun

Kinerja antarlokasi menunjukkan rentang yang lebar. CSI keseluruhan tertinggi terdapat di ARG Baturiti sebesar 0,217, ARG Catur sebesar 0,217, dan ARG Pupuan sebesar 0,216. Ketiga stasiun berada pada dataran tinggi. ARG Baturiti menghasilkan POD 0,683, sedangkan ARG Pupuan memiliki FAR terendah di antara tiga stasiun tersebut sebesar 0,743. Sebaliknya, CSI terendah terdapat di AAWS Gerokgak sebesar 0,065, ARG Kerambitan sebesar 0,087, dan ARG Kubutambahan sebesar 0,095. AAWS Gerokgak memiliki FAR 0,933 dan bias 9,614, yang menunjukkan overestimate kejadian sangat kuat. Variasi musiman ditunjukkan pada **Gambar 6**, sedangkan ringkasan stasiun ditunjukkan pada **Tabel 7**.



Gambar 6. Variasi CSI per stasiun pada musim hujan dan musim kemarau.

Sebagian besar stasiun mengalami penurunan CSI pada musim kemarau. ARG Catur mencapai CSI 0,274 pada musim hujan, sedangkan ARG Baturiti dan ARG Pupuan masing-masing mencapai 0,260 dan 0,246. Pada musim kemarau, nilai tertinggi terdapat di ARG Pupuan sebesar 0,172, ARG Baturiti sebesar 0,165, dan AAWS Karangasem sebesar 0,150. CSI terendah musim kemarau terjadi di AAWS Gerokgak sebesar 0,030, ARG Kubutambahan sebesar 0,039, dan ARG Kubu sebesar 0,047.

Nilai POD pada stasiun dengan CSI rendah tetap dapat terlihat tinggi. AAWS Gerokgak, misalnya, memiliki POD 0,646 tetapi CSI hanya 0,065 karena sebagian besar prakiraan hujannya menjadi *false alarm*.

Hasil ini memperlihatkan mengapa POD tidak boleh dipakai sebagai satu-satunya ukuran kinerja. Peningkatan deteksi harus diseimbangkan dengan pengurangan alarm palsu. Dalam konteks operasional, prakiraan dengan POD cukup tinggi dan FAR sangat tinggi dapat mengurangi kepercayaan pengguna serta memicu respons yang tidak efisien.

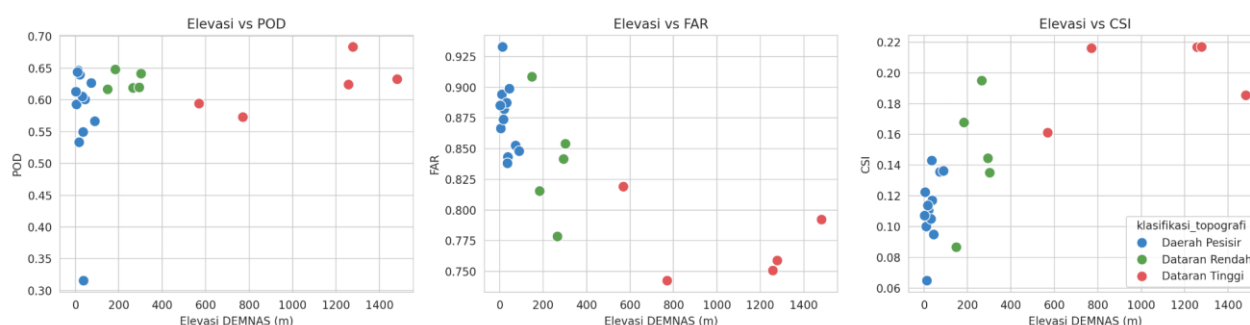
Tabel 7. Stasiun dengan CSI tertinggi dan terendah.

Kelompok	Stasiun	Topografi	POD	FAR	CSI
Tertinggi	ARG Baturiti	Dataran tinggi	0,683	0,759	0,217
Tertinggi	ARG Catur	Dataran tinggi	0,624	0,751	0,217
Tertinggi	ARG Pupuan	Dataran tinggi	0,573	0,743	0,216
Terendah	ARG Kubutambahan	Daerah pesisir	0,601	0,899	0,095
Terendah	ARG Kerambitan	Dataran rendah	0,617	0,908	0,087
Terendah	AAWS Gerokgak	Daerah pesisir	0,646	0,933	0,065

Pola antarstasiun mengindikasikan bahwa koreksi model tidak cukup dilakukan dengan satu faktor untuk seluruh Bali. Stasiun pesisir tertentu membutuhkan penekanan frekuensi kejadian yang lebih kuat, sedangkan stasiun dataran tinggi dapat memerlukan penyesuaian yang berbeda. Pendekatan pascaproses dapat menggunakan kalibrasi probabilitas, *model output statistics*, atau pemilihan ambang yang spesifik lokasi dan musim. Namun, metode tersebut harus diuji menggunakan data terpisah agar peningkatan skor tidak berasal dari *overfitting*.

F. Hubungan Deskriptif dengan Elevasi

Grafik sebar pada **Gambar 7** memperlihatkan bahwa POD tidak berubah secara monoton terhadap elevasi. Sebagian besar stasiun memiliki POD sekitar 0,53-0,68, dengan satu stasiun pesisir yang memiliki POD lebih rendah. FAR cenderung lebih kecil pada beberapa stasiun dengan elevasi tinggi, sedangkan CSI cenderung lebih besar pada elevasi di atas 500 m. Pola tersebut sesuai dengan hasil agregasi topografi, tetapi belum membuktikan hubungan statistik langsung.



Gambar 7. Sebaran elevasi stasiun terhadap POD, FAR, dan CSI.

Elevasi berkaitan dengan topografi, tetapi tidak mewakili seluruh faktor yang mengontrol performa. Jarak ke pantai, sisi datang angin, orientasi lereng, kepadatan stasiun, karakter alat, dan keterwakilan grid dapat menghasilkan skor berbeda pada elevasi yang sama. Analisis lanjutan perlu menghitung korelasi atau regresi dengan mempertimbangkan ketidaklinieran dan pengelompokan musim. Penggunaan jumlah stasiun yang terbatas juga membatasi kekuatan inferensi statistik.

G. Implikasi Operasional dan Keterbatasan

Hasil penelitian menunjukkan bahwa IFS memiliki nilai informasi karena POD berada di atas 0,5 dan ETS tetap positif pada seluruh agregasi utama. Keluaran model tetap berguna sebagai salah satu sumber

panduan untuk mengenali lingkungan yang mendukung hujan. Namun, FAR dan bias yang sangat tinggi menunjukkan bahwa keluaran mentah tidak sebaiknya diperlakukan sebagai keputusan kejadian di titik stasiun. Prakirawan perlu menggabungkan IFS dengan observasi terbaru, radar, satelit, model lain, dan pengetahuan lokal sebelum menerbitkan informasi operasional.

Kalibrasi dapat diarahkan pada pengurangan kejadian hujan ringan palsu. Salah satu pilihan ialah menaikkan ambang kejadian IFS dan mencari ambang yang menyeimbangkan POD, FAR, dan CSI pada setiap musim. Pilihan lain ialah membentuk probabilitas hujan dari beberapa waktu inisialisasi atau anggota ensemble, kemudian mengkalibrasi probabilitas tersebut terhadap frekuensi observasi. Evaluasi harus mempertahankan data uji independen dan menilai perubahan reliabilitas, resolusi, serta nilai ekonomis untuk pengguna tertentu.

Penelitian memiliki beberapa keterbatasan. Pertama, analisis hanya menggunakan satu ambang hujan sehingga belum menggambarkan keterampilan pada hujan sedang, lebat, dan ekstrem. Kedua, seluruh lead time digabungkan sehingga perubahan performa terhadap jarak prakiraan belum terlihat. Ketiga, verifikasi titik sensitif terhadap kesalahan posisi dan tidak menilai kedekatan spasial hujan; pendekatan lingkungan atau berbasis objek diperlukan untuk menilai kesalahan tersebut [15]-[18], [20]. Keempat, jumlah pasangan berbeda antarstasiun dan seleksi kelengkapan dapat memengaruhi keterwakilan spasial. Kelima, bootstrap dilakukan pada tingkat pasangan agregat sehingga dependensi temporal dan spasial belum dimodelkan secara eksplisit. Keenam, penelitian tidak menguji penyebab meteorologis dari variasi skor.

H. Ketidakpastian dan Stabilitas Skor

Interval kepercayaan bootstrap menunjukkan bahwa estimasi agregat relatif stabil. Tabel 8 menunjukkan interval kepercayaan 95% skor verifikasi utama, yaitu POD, FAR, CSI, dan ETS. Pada tingkat keseluruhan, lebar interval POD sekitar 0,007, FAR sekitar 0,003, CSI sekitar 0,002, dan ETS sekitar 0,002. Interval yang sempit terutama dipengaruhi oleh jumlah pasangan data yang sangat besar. Namun, presisi statistik tidak sama dengan keterampilan yang tinggi. Estimasi FAR tetap presisi pada nilai yang buruk, sedangkan CSI dan ETS tetap presisi pada nilai yang rendah.

Tabel 8. Interval kepercayaan 95% skor verifikasi utama.

Agregasi	POD 95% CI	FAR 95% CI	CSI 95% CI	ETS 95% CI
Keseluruhan	0,599-0,605	0,841-0,844	0,142-0,144	0,073-0,075
Musim hujan	0,632-0,641	0,810-0,814	0,168-0,171	0,074-0,077
Musim kemarau	0,527-0,540	0,885-0,888	0,102-0,105	0,057-0,060
Daerah pesisir	0,563-0,575	0,876-0,880	0,110-0,113	0,058-0,061
Dataran rendah	0,622-0,636	0,838-0,844	0,143-0,148	0,067-0,071
Dataran tinggi	0,616-0,626	0,770-0,776	0,197-0,202	0,099-0,104

Perbedaan interval antara musim hujan dan musim kemarau memperkuat temuan bahwa perbedaan performa bukan sekadar akibat variasi acak pada sampel. Interval CSI musim hujan 0,168-0,171 tidak bertumpang tindih dengan interval musim kemarau 0,102-0,105. Pola yang sama muncul pada ETS. Demikian pula, interval FAR dataran tinggi berada jauh di bawah pesisir dan dataran rendah. Hasil tersebut menunjukkan perbedaan sistematis dalam sampel penelitian, walaupun *bootstrap* pasangan sederhana belum memperhitungkan korelasi antarjam, antarstasiun, dan antarkejadian hujan.

Dependensi temporal dapat menyebabkan interval terlalu sempit karena beberapa baris data berasal dari rangkaian hujan yang sama. Penelitian berikutnya dapat menggunakan *block bootstrap* berdasarkan hari atau episode hujan. Pendekatan tersebut mempertahankan struktur waktu dalam sampel ulang dan memberikan estimasi ketidakpastian yang lebih konservatif. Bootstrap berstrata menurut stasiun juga dapat digunakan agar stasiun dengan jumlah data lebih besar tidak mendominasi hasil agregat.

I. Sintesis Diagnostik Kinerja Model

Gabungan seluruh skor membentuk diagnosis yang konsisten. IFS memiliki sensitivitas kejadian yang cukup, tetapi reliabilitas prakiraan hujannya rendah. Model menangkap sebagian besar hujan observasi karena memprakirakan area dan durasi hujan secara luas. Strategi ini menekan miss, tetapi meningkatkan false alarm. Bias frekuensi yang jauh di atas satu menjadi penghubung utama antara POD yang relatif tinggi dan CSI yang rendah. Dengan demikian, peningkatan kinerja tidak harus berfokus pada penambahan kejadian hujan, melainkan pada penyaringan prakiraan hujan yang lemah dan tidak terverifikasi.

Perbedaan musim dan topografi menunjukkan bahwa satu ambang universal belum tentu optimal. Pada musim kemarau dan wilayah pesisir, ambang $0,1 \text{ mm jam}^{-1}$ menghasilkan *overestimate* paling besar. Ambang yang lebih tinggi mungkin menurunkan FAR dan bias, tetapi dapat menurunkan POD. Pemilihan ambang perlu mempertimbangkan kebutuhan pengguna. Layanan yang memprioritaskan keselamatan dapat menerima POD tinggi dengan FAR tertentu, sedangkan kegiatan operasional yang sensitif terhadap alarm palsu memerlukan keseimbangan berbeda. Kurva performa dari beberapa ambang dapat membantu menentukan kompromi tersebut secara transparan.

Hasil ini juga menegaskan pentingnya membedakan kualitas model atmosfer dari kecocokan grid terhadap alat titik. Sebagian false alarm mungkin mencerminkan hujan yang terjadi di sekitar stasiun, tetapi tidak tepat pada lokasi alat. Sebagian lain dapat berasal dari presipitasi ringan semu yang terlalu sering dibangkitkan model. Verifikasi spasial, pengamatan radar, dan analisis intensitas diperlukan untuk memisahkan kedua sumber kesalahan. Tanpa pemisahan tersebut, koreksi yang terlalu agresif berisiko menghapus sinyal hujan yang sebenarnya berguna pada skala wilayah.

IV. PENUTUP

A. Simpulan

1. IFS mendeteksi 60,2% kejadian hujan observasi, tetapi menghasilkan FAR 0,842 dan bias 3,820. Model memprakirakan hujan sekitar 3,8 kali lebih sering daripada yang tercatat oleh stasiun. CSI 0,143 dan ETS 0,074 menunjukkan keterampilan positif, tetapi masih rendah untuk keputusan kejadian pada titik dan waktu yang sama.
2. Kinerja lebih baik pada musim hujan. CSI meningkat dari 0,103 pada musim kemarau menjadi 0,170 pada musim hujan. Musim kemarau menjadi periode paling bermasalah karena FAR mencapai 0,887 dan bias 4,706.
3. Dataran tinggi memberikan performa terbaik dengan CSI 0,200, ETS 0,101, FAR 0,773, dan bias 2,734. Kombinasi dataran tinggi-musim hujan menghasilkan CSI tertinggi sebesar 0,241. Daerah pesisir dan dataran rendah mengalami *overestimate* lebih kuat, terutama pada musim kemarau.
4. Performa berbeda secara deskriptif antarstasiun. ARG Baturiti, ARG Catur, dan ARG Pupuan memiliki CSI tertinggi, sedangkan AAWS Gerokgak, ARG Kerambitan, dan ARG Kubutambahan memiliki CSI terendah. Perbedaan tersebut mendukung kebutuhan kalibrasi menurut lokasi dan musim.
5. Keluaran IFS dapat digunakan sebagai panduan meteorologis, tetapi tidak sebaiknya digunakan secara langsung sebagai penentu hujan di titik stasiun tanpa koreksi, informasi observasi tambahan, dan evaluasi ambang yang sesuai kebutuhan operasional.

B. Saran

1. Penelitian lanjutan perlu menguji beberapa ambang curah hujan dan menyusun kurva performa agar ambang optimal dapat ditentukan untuk musim dan stasiun yang berbeda.
2. Analisis perlu dipisahkan menurut lead time, waktu inisialisasi, dan jam lokal untuk mengetahui kapan kesalahan deteksi dan false alarm paling besar.
3. Verifikasi titik perlu dilengkapi dengan metode spasial berbasis lingkungan, radar, atau grid observasi untuk memisahkan kesalahan posisi dari kesalahan kejadian.

4. Kalibrasi statistik atau probabilistik perlu dikembangkan dengan skema pelatihan dan pengujian terpisah, lalu dibandingkan terhadap keluaran IFS mentah menggunakan CSI, ETS, reliabilitas, dan nilai guna operasional.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Sekolah Tinggi Meteorologi Klimatologi dan Geofisika serta Badan Meteorologi Klimatologi dan Geofisika atas dukungan data, fasilitas, dan bimbingan dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] Aldrian, E. and Susanto, R.D., "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature," *International Journal of Climatology*, 23(12), 1435-1452, 2003. doi: 10.1002/joc.950.
- [2] Badan Meteorologi Klimatologi dan Geofisika, *Pemutakhiran Zona Musim Indonesia Periode 1991-2020*, BMKG, Jakarta, 2021.
- [3] Rahmawati, N. and Lubczynski, M.W., "Validation of satellite daily rainfall estimates in complex terrain of Bali Island, Indonesia," *Theoretical and Applied Climatology*, 134, 513-532, 2018. doi: 10.1007/s00704-017-2290-7.
- [4] European Centre for Medium-Range Weather Forecasts, *IFS Documentation Cy47r3, Part IV: Physical Processes*, ECMWF, Reading, 2021.
- [5] Lavers, D.A., Harrigan, S. and Prudhomme, C., "Precipitation biases in the ECMWF Integrated Forecasting System," *Journal of Hydrometeorology*, 22(5), 1187-1198, 2021. doi: 10.1175/JHM-D-20-0308.1.
- [6] Liu, C., Sun, J., Yang, X., Jin, S. and Fu, S., "Evaluation of ECMWF precipitation predictions in China during 2015-18," *Weather and Forecasting*, 36(3), 1043-1060, 2021. doi: 10.1175/WAF-D-20-0143.1.
- [7] Ebert, E.E., et al., "Progress and challenges in forecast verification," *Meteorological Applications*, 20(2), 130-139, 2013. doi: 10.1002/met.1392.
- [8] Jolliffe, I.T. and Stephenson, D.B., eds., *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed., Wiley, Chichester, 2012. doi: 10.1002/9781119960003.
- [9] Hogan, R.J. and Mason, I.B., "Deterministic forecasts of binary events," in *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed., Jolliffe, I.T. and Stephenson, D.B., eds., Wiley, Chichester, 31-59, 2012. doi: 10.1002/9781119960003.ch3.
- [10] Schaefer, J.T., "The critical success index as an indicator of warning skill," *Weather and Forecasting*, 5(4), 570-575, 1990. doi: 10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.
- [11] Roebber, P.J., "Visualizing multiple measures of forecast quality," *Weather and Forecasting*, 24(2), 601-608, 2009. doi: 10.1175/2008WAF2222159.1.
- [12] Wilks, D.S., *Statistical Methods in the Atmospheric Sciences*, 4th ed., Elsevier, Amsterdam, 2019.
- [13] Stephens, G.L., et al., "Dreary state of precipitation in global models," *Journal of Geophysical Research: Atmospheres*, 115(D24), D24211, 2010. doi: 10.1029/2010JD014532.
- [14] Sun, Y., Solomon, S., Dai, A. and Portmann, R.W., "How often does it rain?," *Journal of Climate*, 19(6), 916-934, 2006. doi: 10.1175/JCLI3672.1.
- [15] Ebert, E.E., "Neighborhood verification: A strategy for rewarding close forecasts," *Weather and Forecasting*, 24(6), 1498-1510, 2009. doi: 10.1175/2009WAF2222251.1.
- [16] Roberts, N.M. and Lean, H.W., "Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events," *Monthly Weather Review*, 136(1), 78-97, 2008. doi: 10.1175/2007MWR2123.1.

- [17] Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. and Ebert, E.E., "Intercomparison of spatial forecast verification methods," *Weather and Forecasting*, 24(5), 1416-1430, 2009. doi: 10.1175/2009WAF2222269.1.
- [18] Mittermaier, M. and Roberts, N., "Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score," *Weather and Forecasting*, 25(1), 343-354, 2010. doi: 10.1175/2009WAF2222260.1.
- [19] Ferro, C.A.T. and Stephenson, D.B., "Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events," *Weather and Forecasting*, 26(5), 699-713, 2011. doi: 10.1175/WAF-D-10-05030.1.
- [20] Ahijevych, D., Gilleland, E., Brown, B.G. and Ebert, E.E., "Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts," *Weather and Forecasting*, 24(6), 1485-1497, 2009. doi: 10.1175/2009WAF2222298.1.