



MARLIN

Marine and Fisheries Science Technology Journal

Tersedia online di: <http://ejournal-balitbang.kkp.go.id/index.php/marlin>

e-mail: jurnal.marlin@gmail.com

Volume 6 Nomor 2 Agustus 2025

p-ISSN 2716-120X

e-ISSN 2715-9639

PENDEKATAN STATISTIK DAN MACHINE LEARNING UNTUK PEMETAAN LOKASI OPTIMAL PENANGKAPAN TUNA MATA BESAR DI WPP 714

STATISTICAL AND MACHINE LEARNING APPROACHES FOR MAPPING OPTIMAL BIGEYE TUNA FISHING GROUNDS IN FMA 714

**Akhmatul Ferlin^{1*)}, Sunarwan Asuhadi²⁾, Khairudin Isman¹⁾, Rahman Shaadikin³⁾,
Ridwan Hasan³⁾**

¹Akademi Komunitas Kelautan dan Perikanan, Wakatobi

²Badan Riset dan Inovasi Nasional, Jakarta

³Institut Teknologi dan Bisnis Muhammadiyah, Wakatobi

Teregistrasi I tanggal: 22 Juli 2025; Diterima setelah perbaikan tanggal: 14 September 2025; Disetujui terbit tanggal: 15 September 2025

ABSTRAK

Mengingat kompleksitas Wilayah Pengelolaan Perikanan (WPP) 714, yang telah mengubah faktor lingkungan dengan cepat, diperlukan penelitian untuk menentukan lokasi optimal penangkapan ikan penting, seperti tuna mata besar. Untuk memetakan lokasi penangkapan ikan yang optimal, metode statistik dan *machine learning* digunakan melalui pemahaman tentang dinamika konsentrasi klorofil-a (*Chlo*) dan suhu permukaan laut (SST) dan data tangkapan ikan sebagai faktor kunci. Dengan menggunakan *log book* perikanan dan data oseanografi dari tahun 2014 hingga 2022, penelitian ini menerapkan metode *Generalized Additive Model* (GAM) dan *Empirical Cumulative Distribution Function* (ECDF) untuk menganalisis pengaruh *Chlo* dan SST terhadap *Catch per Unit Effort* (CPUE). Karena kinerja GAM yang tidak optimal, ECDF dipilih sebagai metode utama. Penentuan rentang nilai optimal menggunakan metode *k-means* dan *elbow* menunjukkan kisaran *Chlo* optimal 0,087-0,30 mg/m³ dan SPL 29,46-30,45°C. Analisis menunjukkan kondisi penangkapan terbaik dari Januari hingga Maret (2014-2022). Temuan ini mendukung penangkapan ikan berkelanjutan dengan menyediakan peta bulanan zona penangkapan ikan yang optimal, membantu nelayan beradaptasi dengan kondisi dinamis. Penelitian lebih lanjut disarankan untuk mengintegrasikan data *real-time* dan teknik lanjutan untuk meningkatkan akurasi lokasi.

KATA KUNCI: Machine Learning; Optimalitas Penangkapan Ikan; Statistik; Tuna mata besar; WPP 714

DOI: <http://dx.doi.org/10.15578/marlin.V6.I1.2025.131-150>

*Korespondensi penulis:
e-mail : ferlinbatulo@gmail.com

ABSTRACT

Given the complexity of the Fisheries Management Area (FMA) 714, which has rapidly changing environmental factors, studies are needed to determine the optimal location for important fishing, such as bigeye tuna (BET). To map optimal fishing locations, statistical and machine-learning methods are used through an understanding of the dynamics of Chlorophyll-a (Chlo) concentrations and sea surface temperature (SST) and fish catch data as key factors. Using the fishing logbook and oceanographic data from 2014 to 2022, this study applies the methods of generalized Additive Model (GAM) and Empirical Cumulative Distribution Function (ECDF) to analyze the effect of Chlo and SSTs on Catch per Unit Efforts (CPUEs). Due to GAM's suboptimal performance, ECDF is chosen as the primary method. Determining the optimal value range using k-means and the elbow method showed an optimal Chlo range of 0.087–0.30 mg/m³ and SST of 29.46–30.45°C. The analysis showed the best catching conditions from January to March (2014–2022). These findings support sustainable fishing by providing a monthly map of optimal fishing zones, helping fishers adapt to dynamic conditions. Further research is suggested to integrate real-time data and advanced techniques to improve location accuracy.

Keywords: *Bigeye tuna; Fishing optimality; FMA 714; Machine Learning; Statistics*

PENDAHULUAN

Tuna mata besar (*Bigeye tuna*/BET) adalah spesies bernilai tinggi di WPP 714 yang penting secara ekonomi dan telah mengalami penurunan tangkapan akibat penangkapan ikan yang berlebihan (Alfianto et al., 2023). Hal ini menimbulkan kekhawatiran tentang keberlanjutan populasi dan keberlanjutan ekonomi masyarakat pesisir yang bergantung pada perikanan (Ovando et al., 2021). Beberapa negara telah merencanakan untuk mengurangi batas tangkapan untuk mengatasi masalah ini (Nagy et al., 2024). Tuna mata besar memiliki karakteristik biologis khas, antara lain ukuran tubuh yang besar, umur panjang, dan tingkat reproduksi yang tinggi. Ciri-ciri tersebut dipengaruhi oleh variasi iklim serta tekanan akibat aktivitas penangkapan (Lin et al., 2023). Berbagai penelitian telah mengeksplorasi distribusi tuna ini menggunakan parameter oseanografi dan

model analitik seperti General Additive Model (GAM). Namun, mereka umumnya terbatas pada area dan data jangka pendek, menghasilkan wawasan yang terbatas (Punt et al., 2021).

Faktor oseanografi, seperti konsentrasi klorofil-a (*Chlo*) dan suhu permukaan laut (SPL), mempengaruhi kelimpahan dan distribusi tuna mata besar (Peng Lian & Le Gao, 2024). Sebagai indikator kelimpahan fitoplankton, *Chlo* berperan dalam rantai makanan laut, mempengaruhi pertumbuhan, siklus reproduksi, dan distribusi ikan. Sebaliknya, SPL mempengaruhi pergerakan dan distribusi makanan ikan (Friedland et al., 2020; Yati et al., 2024). Memahami parameter oseanografi ini diharapkan dapat meningkatkan efisiensi penangkapan ikan, mengoptimalkan pengelolaan sumber daya perikanan, dan memaksimalkan manfaatnya (Antonio May-Kú et al., 2022).

Penelitian ini bertujuan untuk mengevaluasi produktivitas perairan dan mengidentifikasi area penangkapan ikan mata besar di WPP 714, dengan memanfaatkan pendekatan analisis produktivitas perairan (Kizenga *et al.*, 2021) serta metode penentuan daerah potensial penangkapan ikan (Cornic & Rooker, 2021). Pendekatan ini menguntungkan karena GAM memungkinkan pemodelan non-linier dari hubungan antara variabel oseanografi dan catch per unit effort (CPUE). Pada saat yang sama, ECDF menawarkan fleksibilitas dalam analisis distribusi data kumulatif yang dapat mengatasi keterbatasan analisis tradisional (Artetxe-Arrate *et al.*, 2021). Selain itu, pendekatan *machine learning* meningkatkan akurasi dan efisiensi dalam memproses data skala besar secara otomatis (Guillotreau *et al.*, 2024).

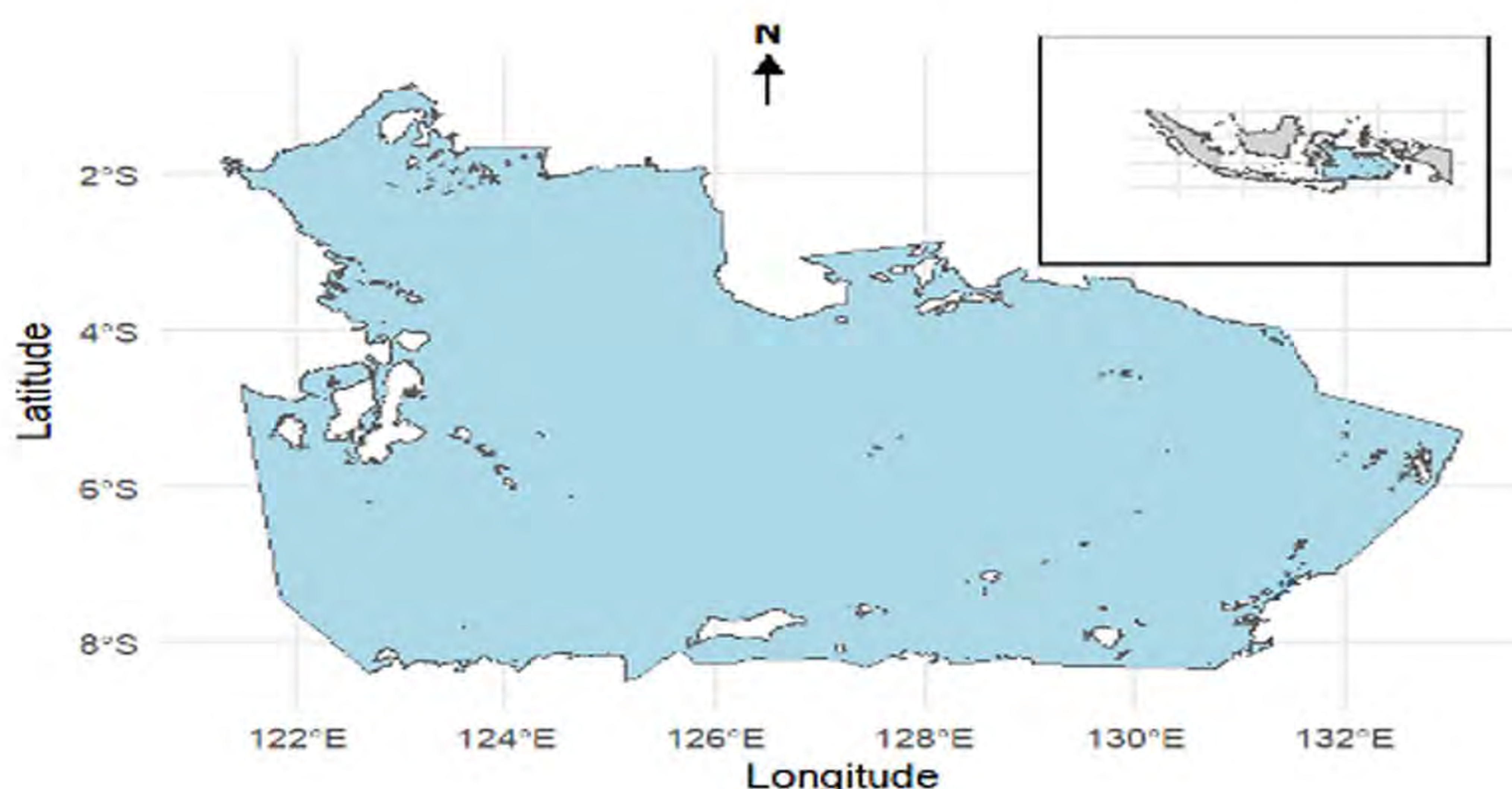
Pertanyaan penelitian adalah tentang pengaruh Chlo dan SPL terhadap distribusi optimal tangkapan tuna mata besar dan metode statistik dan *machine learning* apa yang paling efektif untuk memetakan area tangkapan yang optimal. Hipotesis yang diajukan

adalah bahwa konsentrasi Chlo yang tinggi dan SPL dalam kisaran optimal akan berkontribusi secara signifikan terhadap peningkatan CPUE dan bahwa pendekatan ECDF akan memberikan hasil yang lebih akurat daripada metode tradisional seperti GAM (Asuhadi *et al.*, 2024; Fachruddin Syah *et al.*, 2019). Dengan pendekatan ini, penelitian diharapkan dapat memberikan panduan yang bermanfaat untuk pengelolaan sumber daya perikanan yang lebih efisien dan mendukung keberlanjutan ekosistem laut.

METODE PENELITIAN

Area studi

Penelitian ini dilakukan di WPP 714 yang merupakan perairan yang kaya akan keanekaragaman hayati dan penting bagi industri perikanan nasional. Perairan ini mencakup berbagai ekosistem yang mendukung habitat tuna mata besar dan spesies lainnya (Hargiyatno *et al.*, 2021). Gambar 1 menunjukkan peta detail wilayah WPP 714, memberikan gambaran tentang letak geografis dan batas-batas perairan yang menjadi fokus penelitian ini.



Gambar 1. Area Studi WPP 714 (2025)

Sumber data

data untuk penelitian ini, termasuk *Chlo*, SPL, dan *log book* tangkapan ikan (2014-2022).

Tabel 1 menyajikan jenis dan sumber

Tabel 1. Jenis dan sumber data penelitian

Nama data	Jenis data	Sumber
Klorofil-a	NC file	<i>Ocean color</i>
Suhu permukaan laut (2014 – 2022)	NC file	<i>Ocean color</i>
<i>Log book</i> tangkapan ikan (2014 – 2022)	xlsx file	Kementerian Kelautan dan Perikanan Republik Indonesia

Pengumpulan dan pemrosesan data

Data yang digunakan adalah dataset tangkapan ikan, konsentrasi *Chlo*, dan data SPL. Dataset tangkapan ikan diperoleh dari *log book* tangkapan ikan Direktorat Pengelolaan Sumber Daya Ikan Kementerian Kelautan dan Perikanan dari tahun 2014 hingga 2022. Data konsentrasi *Chlo* dan SPL untuk WPP 714 diperoleh dari *Aqua Modis* (Sean Bailey, 2023). Data mentah *Chlo* dan SPL dengan resolusi 4 km diolah menggunakan *Ocean Data View* (ODV). Semua himpunan data dibersihkan, diproses, dan divisualisasikan melalui R Studio versi 4.4.0, mengaktifkan pustaka dan menjalankan fungsi yang diperlukan. Sebelum membangun model prediksi, penting untuk mengidentifikasi *outlier* pada data *Chlo* dan SPL, yang dilakukan dalam penelitian ini dengan bantuan perpustakaan 'e1071'.

Model GAM

Analisis model menggunakan model GAM diikuti dengan pengecekan nilai *Akaike Information Criterion* (AIC). Model GAM dibuat menggunakan pustaka 'mgcv' (Pedersen et al., 2019) dan pemeriksaan kelayakan model menggunakan AIC (Susko & Roger, 2020). Model GAM dibuat sebagai model gabungan dan model

tunggal antara parameter CPUE dan oseanografi (*Chlo* dan SPL). Selanjutnya, plot derajat kebebasan dibuat. Setelah itu, nilai signifikan *Chlo* dan SPL dari model GAM ditentukan. Penentuan nilai signifikan *Chlo* dan SPL dilakukan dengan memprediksi nilai CPUE berdasarkan kisaran nilai *Chlo* dan SPL. Langkah-langkahnya termasuk menentukan rentang nilai *Chlo* dan SPL dengan menggunakan fungsi 'seq' untuk membuat rentang nilai dari data minimum hingga maksimum. Kemudian, kerangka data baru dibuat menggunakan 'expand.grid' yang menggabungkan semua pasangan nilai *Chlo* dan SPL. Selanjutnya, prediksi CPUE dilakukan dengan menggunakan model GAM. Dari hasil prediksi diperoleh nilai signifikan *Chlo* dan SPL. Setelah itu, dilakukan penilaian kinerja model GAM.

Rumus ECDF untuk model

Grafik ECDF dibuat untuk menentukan nilai signifikan dari suatu parameter (Brancucci Martínez-Anido et al., 2012). Dalam rumus berikut (1), $F_n(x)$ adalah ECDF sampel hingga nilai x , n adalah jumlah total sampel, x_i adalah pengamatan atau data individu ke- i , dan $1\{X_i \leq x\}$ adalah fungsi indikator nilai-nilai 1 jika $X_i \leq x$ dan 0 jika sebaliknya (Asuhadi et al., 2024).

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}$$

Tes Kolmogorov-Smirnov (K-S)

Dalam persamaan ini, $D_{n,m}$ adalah statistik K-S untuk dua sampel, $\sup$ adalah suprem (nilai maksimum) dari perbedaan absolut antara dua fungsi

distribusi kumulatif, $F_n(x)$ adalah ECDF dari sampel pertama ukuran n , $G_m(x)$ adalah ECDF dari sampel kedua ukuran m , dan $F(x)$ adalah fungsi distribusi kumulatif teoretis (Dimitrova et al., 2020).

$$D_{n,m} = \sup_x |F_n(x) - G_m(x)|$$

K-means clustering

Dalam rumus berikut, jarak antara titik data x_i dan c_j sentroid adalah $d(x_i, c_j)$, x_{im} adalah

komponen m dari titik data x_i , c_{jm} adalah komponen m dari c_j centroid, dan M adalah jumlah fitur (Zendrato et al., 2020).

$$d(x_i, c_j) = \sqrt{\sum_{m=1}^M (x_{im} - c_{jm})^2}$$

ANOVA test

Dalam persamaan di bawah ini, SST adalah variasi total dalam data, SSB adalah variasi antar kelompok, SSW adalah variasi dalam kelompok, MSB adalah variasi rata-rata antar

kelompok, MSW adalah variasi rata-rata dalam kohort, dan F -statistik adalah rasio MSB terhadap MSW ; Ini digunakan untuk menentukan signifikansi perbedaan antara rata-rata kelompok (Tabachnick & Fidell, 2007).

$$SST = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

$$SSB = \sum_{j=1}^k (\bar{Y}_j - \bar{Y})^2$$

$$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$$

$$MSB = \frac{SSB}{k-1}$$

$$MSW = \frac{SSW}{N-k}$$

$$F = \frac{MSB}{MSW}$$

Tukey's HSD test

Tukey's Honestly Significant Difference test umumnya dikenal sebagai tes HSD. Pada persamaan (10) di bawah ini, q adalah nilai distribusi Rentang Siswa (distribusi-q) untuk

tingkat signifikansi dan derajat kebebasan tertentu $N - k$, MSW adalah Kuadrat Rata-Rata Dalam Kelompok dari tes ANOVA, dan n adalah jumlah pengamatan per kelompok atau ukuran sampel rata-rata jika ukuran kelompok bervariasi (Soares & Castro, 2020).

$$HSD = q \sqrt{\frac{MSW}{n}}$$

Elbow method

WCSS adalah singkatan dari Within-Cluster Sum of Squares. Dalam persamaan (11) di bawah ini, C_j adalah

gugus ke-j, x_i adalah titik data ke-i, C_j adalah sentroid dari gugus ke-j, dan $\|x_i - C_j\|^2$ adalah kuadrat jarak Euclidean antara titik data x_i dan centroid C_j (Bourenane et al., 2023).

$$WCSS(k) = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - C_j\|^2$$

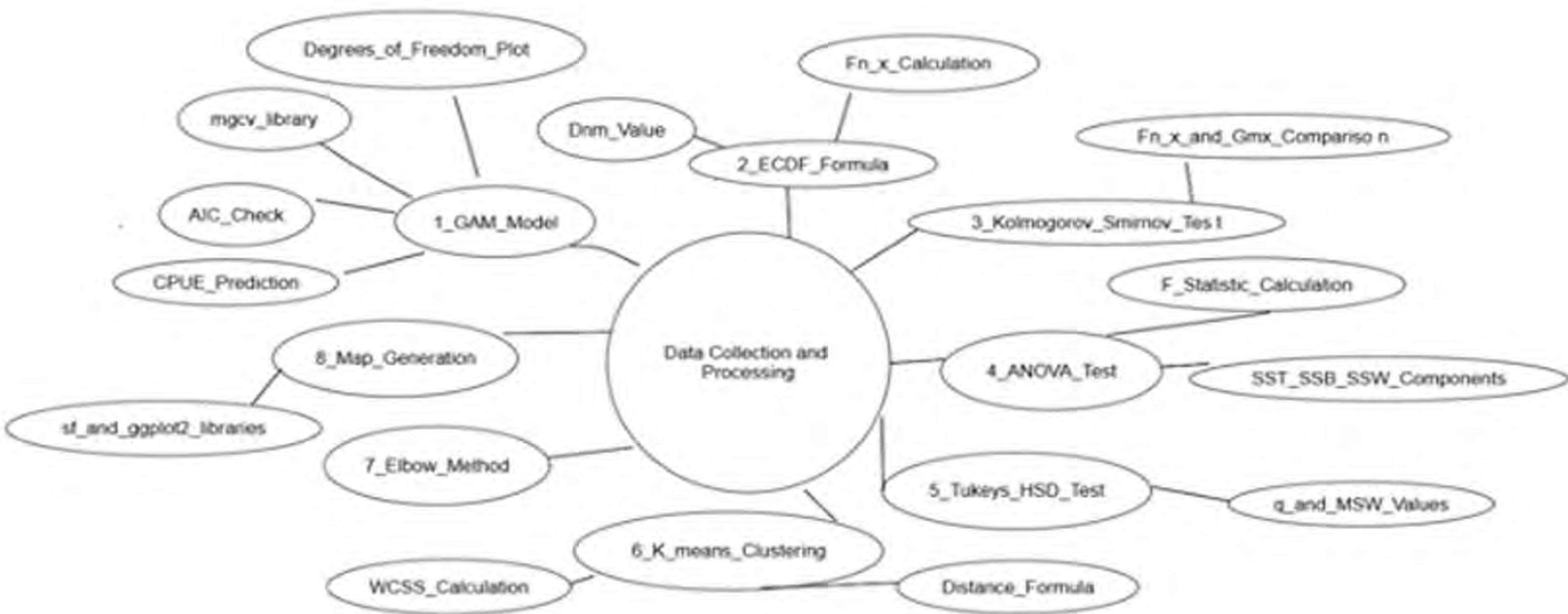
Peta temporal-spasial tentang penangkapan ikan tuna mata besar

Peta ini didasarkan pada distribusi nilai optimal parameter oseanografi dari 2014 hingga 2022, yang dibuat menggunakan perpustakaan (sf) dan perpustakaan R Studio (ggplot2).

Kolmogorov-Smirnov diterapkan untuk memastikan normalitas data. Setelah data rata-rata, uji ANOVA dilakukan untuk mengidentifikasi perbedaan yang signifikan antar kelompok. Jika penting, tes Tukey HSD menentukan pasangan kelompok yang berbeda. Selanjutnya, metode siku menentukan jumlah kelompok optimal, yang kemudian digunakan dalam k-means clustering untuk menentukan rentang nilai optimal dari variabel lingkungan. Langkah-langkah ini dinyatakan sesuai dengan angka dalam diagram alir metodologi pada Gambar 2 di bawah ini.

Diagram alur metodologi

Penelitian ini diawali dengan penerapan Model GAM untuk memprediksi pola data. Jika kinerjanya rendah, metode ECDF digunakan, dan Uji

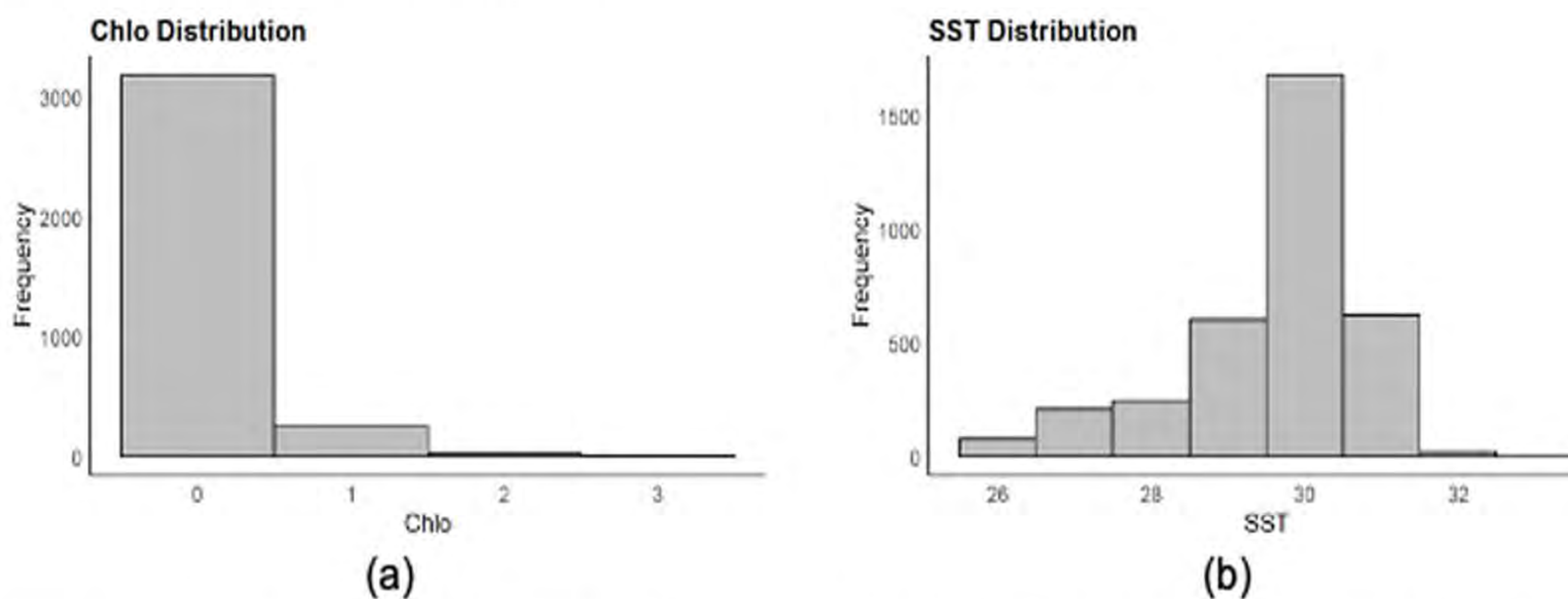


Gambar 2. Diagram alur metodologi

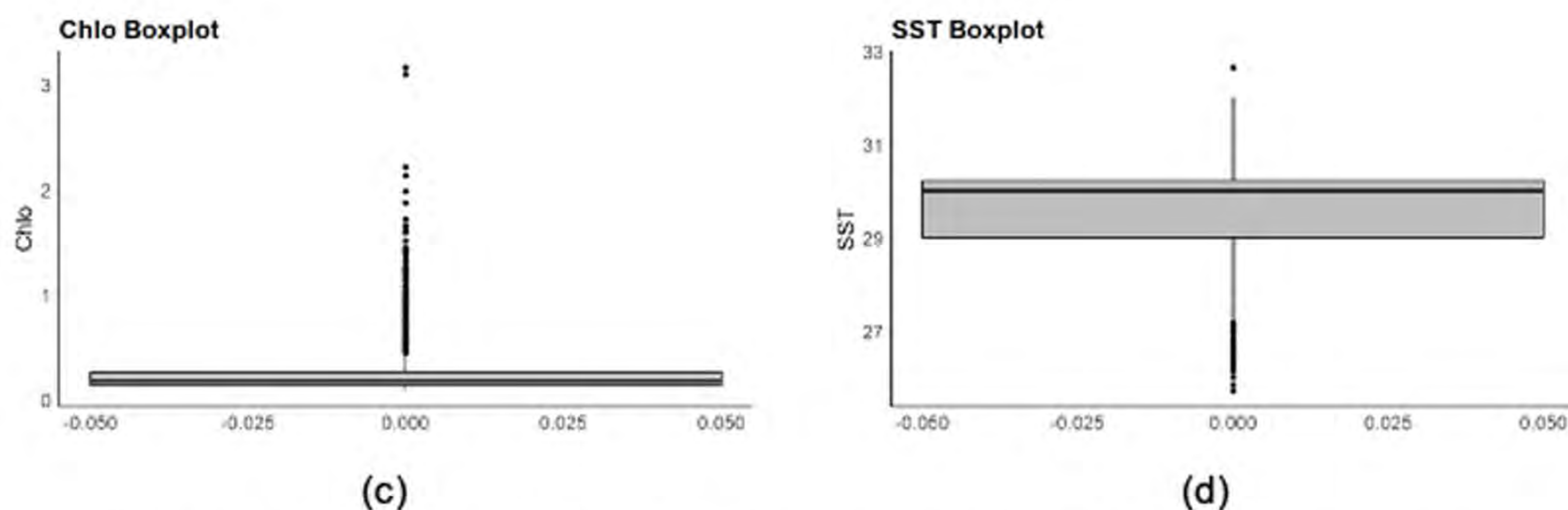
HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dua parameter oseanografi, yaitu *Chlo* dan *SPL*, untuk menganalisis dampaknya terhadap ekosistem laut. *Chlo* berfungsi sebagai indikator kesuburan air dan perkiraan keberadaan fitoplankton (Ginting et al., 2023). Pada saat yang sama, *SPL* menunjukkan suhu laut memainkan peran penting dalam kehidupan laut, termasuk keberadaan ikan, laju metabolisme, dan proses reproduksi

(Pimentel-Acosta et al., 2023). Hasil analisis menunjukkan adanya outlier pada variabel prediktor, seperti yang ditunjukkan pada Gambar 3 dan 4. Gambar 3(a) dan 3(b) menunjukkan distribusi frekuensi nilai *Chlo* dan *SPL*, menggambarkan frekuensi terjadinya nilai spesifik dalam kumpulan data, sedangkan Gambar 4(a) dan 4(b) memberikan ringkasan statistik *Chlo* dan *SPL*, termasuk identifikasi median, kuartil, dan outlier.



Gambar 3. Perbandingan (a) outlier - Distribusi Chlo BET, dan (b) outlier - Distribusi SPL BET



Gambar 4. Perbandingan (a) boxplot - Chlo BET, dan (b) boxplot - SPL BET

Ada beberapa metode untuk mengidentifikasi dan menangani outlier. Salah satunya adalah metode ambang batas (Yang et al., 2019) digunakan dalam penelitian ini. Nilai ambang batas n terbaik penelitian ini adalah 1, di mana nilai terendah dari Std. Error (0,062) dan penyimpangan tertinggi yang dijelaskan dihasilkan (7,23%). Hasil ini menunjukkan

bahwa menghapus outlier meningkatkan akurasi model, seperti yang terlihat pada Tabel 2, di mana model tanpa outlier lebih akurat daripada model dengan outlier. Ini menekankan pentingnya penanganan outlier untuk membuat hasil analisis lebih representatif.

Tabel 2. Perbandingan model GAM menggunakan outlier dan tanpa menggunakan outlier

Parameter	Outlier	Tanpa outlier	Perbedaan
AIC	38,984.36	15,248.46	23,735.9
(Intercept)	14.17	2.86	11.31
R-sq.(adj)	0.018	0.068	-0.05
Deviance Explained	2.14%	7.23%	-5.09
GCV	4,728.60	11.12	4,717.48
Scale est.	4,708.50	11.06	4,697.44
N	3,450	2,906	544

Berdasarkan matriks perbandingan model, perbedaan yang signifikan dapat diamati antara model dengan outlier dan model tanpa outlier (Kumar Rajahmundry et al., 2021). Salah satu perbedaan yang paling mencolok adalah nilai AIC, dengan model tanpa outlier memiliki nilai AIC yang jauh lebih rendah (15.248,46) daripada model dengan outlier (38.984,36). Selain itu, nilai intersepsi juga berbeda antara dua set model. Model dengan outlier memiliki nilai intersep yang lebih tinggi (14,17) daripada model tanpa outlier (2,86). Nilai intersep yang lebih tinggi dalam model dengan outlier menunjukkan pengaruh outlier ini pada prediksi variabel dependen. Sangat penting untuk mempertimbangkan dampak outlier pada analisis data (Hu et al., 2024).

Dalam hal R-squared adjusted (adj), model tanpa outlier menunjukkan nilai yang lebih tinggi (0,068) dibandingkan dengan model dengan outlier (0,018). Ini menunjukkan bahwa model tanpa outlier menjelaskan variabilitas data dengan lebih baik (Zhao et al., 2024). Metrik yang dijelaskan devians juga mencerminkan tren, di mana model tanpa outlier menjelaskan persentase penyimpangan yang lebih tinggi (7,23%) dibandingkan dengan model dengan outlier (2,14%). Persentase yang lebih tinggi ini menandakan model yang lebih baik (Bhatta et al., 2021).

Demikian pula, nilai Generalized Cross-Validation (GCV) mendukung keunggulan model tanpa outlier. Model tanpa outlier memiliki nilai GCV yang

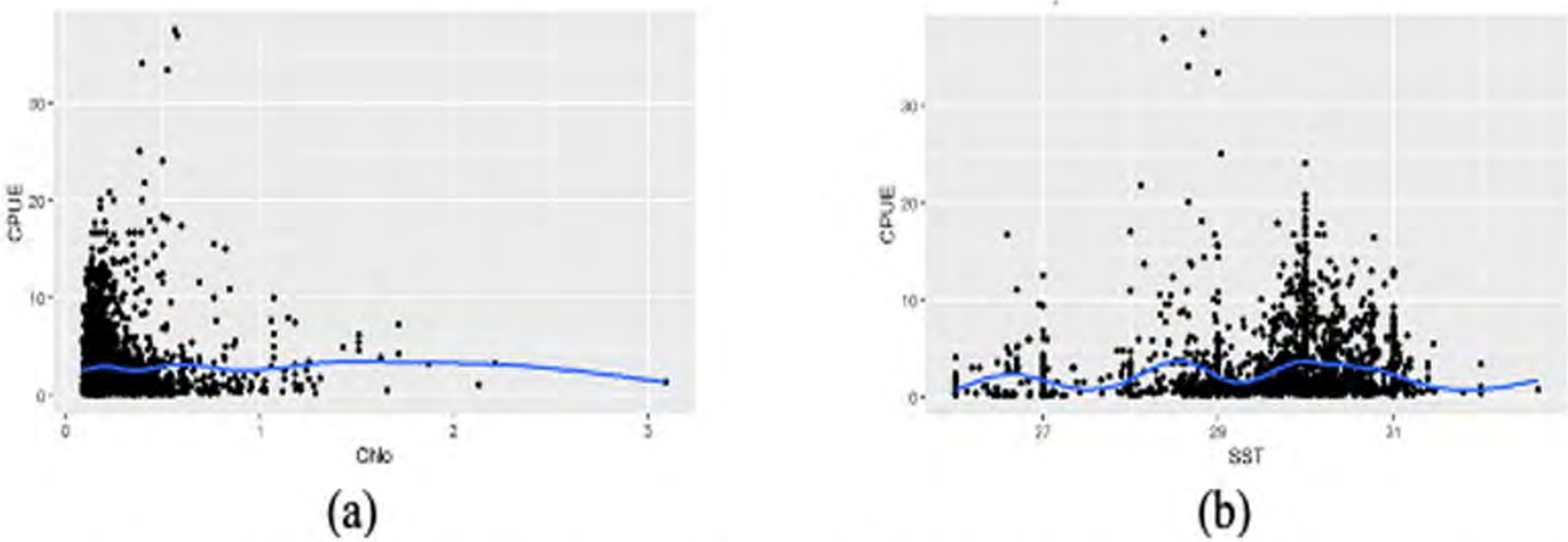
jauh lebih rendah (11,06) dibandingkan dengan model dengan outlier (4.728,60), menunjukkan kualitas model yang lebih baik (Arashi et al., 2021). Model tanpa outlier menghasilkan hasil yang lebih akurat dan andal, seperti yang tercermin dalam nilai AIC, nilai intersepsi, adj R-kuadrat, deviance dijelaskan, dan GCV. Temuan ini menyoroti pentingnya penanganan outlier dalam analisis data untuk kinerja model yang lebih baik (Reda Abonazel, 2020).

Selain membandingkan model di atas, pemeriksaan multikolinearitas dilakukan dengan menggunakan paket 'car'. Multikolinearitas terjadi ketika ada korelasi tinggi antara dua atau lebih variabel prediktor. Fungsi vif() digunakan untuk melakukan pengujian. Hasil uji multikolinearitas menunjukkan bahwa baik Chlo dan SPL memiliki GVIF dan Df 0 yang sangat tinggi, sehingga $GVIF^{(1/(2Df))}$ menjadi tak terbatas (Inf). Ini menunjukkan adanya multikolinearitas. Multikolinearitas membuatnya sulit untuk menentukan pengaruh individu dari setiap variabel dalam menjelaskan variabel dependen (G. Li et al., 2023). Multikolinearitas yang tinggi antara Chlo dan SPL membuatnya sulit untuk menilai pengaruh setiap variabel, mengurangi keandalan model. Penanganan multikolinearitas, seperti menghilangkan atau menggabungkan variabel yang berkorelasi, diperlukan untuk meningkatkan akurasi dan interpretasi model.

GVIF	Df	GVIF ^{1/(2Df)}	
Chlo	3.072908e+22	0	Inf
SPL	3.072908e+22	0	Inf

Hubungan antara setiap variabel prediktor dan variabel target dinyatakan dalam model GAM. Dalam hal ini, variabel target adalah CPUE. Hubungan divisualisasikan dalam grafik plot sebar berikut

pada Gambar 5. Fungsi utama grafik pada Gambar 5(a) dan 5(b) adalah untuk mengidentifikasi dan mengilustrasikan korelasi atau pola antara tingkat Chlo dan SPL terhadap CPUE.



Gambar 5. Grafik plot sebar antara variabel target (CPUE) dan prediktor (a) Chlo dan (b) SPL

Pada grafik di atas, sumbu x mewakili variabel *Chlo* dan *SPL*, sedangkan sumbu y mewakili 'CPUE'. Setiap titik pada grafik mewakili satu pengamatan dalam data, dengan koordinat titik tersebut menunjukkan nilai *Chlo*, *SPL*, dan CPUE untuk pengamatan itu. Nilai rata-rata *Chlo*, *SPL*, dan CPUE pada bingkai data adalah 0,24, 29,62, dan 2,86. Grafik *Chlo* - CPUE di

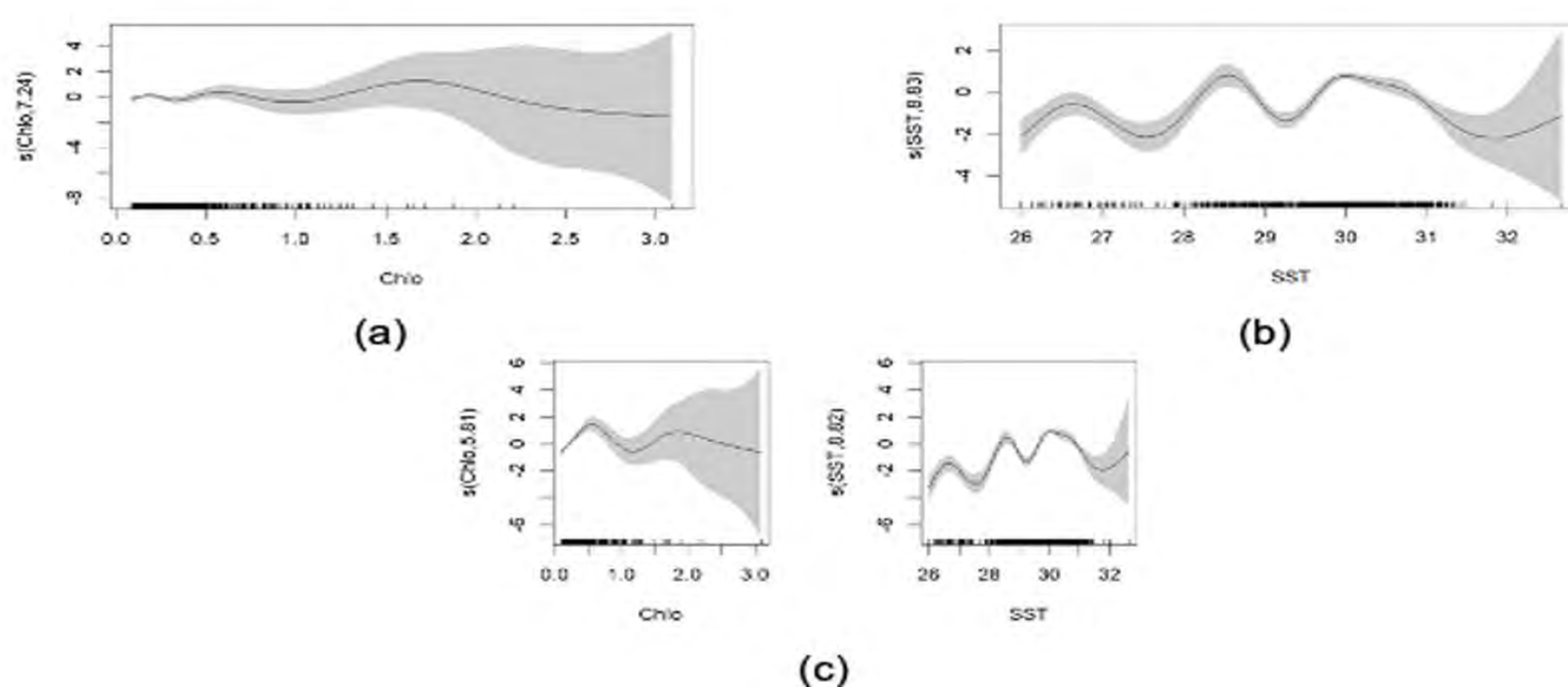
atas menunjukkan bahwa ukuran tangkapan per unit berkurang seiring dengan meningkatnya konsentrasi *Chlo*. Grafik *SPL* - CPUE menunjukkan bahwa nilai 'CPUE' berfluktuasi dengan kenaikan 'SPL'. Lebih banyak titik data yang terakumulasi dalam rentang nilai menengah data 'SPL', dan lebih sedikit di tepi. Perbandingan untuk semua model GAM ditunjukkan pada Tabel 3.

Tabel 3. Perbandingan model GAM antara prediktor tunggal dan gabungan

Parameter	CPUE ~ s(<i>Chlo</i>) (Model 1)	CPUE ~ s(<i>SPL</i>) (Model 2)	CPUE ~ s(<i>Chlo</i>) + s(<i>SPL</i>) (Model 3)
Intercept Estimate	2.86	2.86	2.86
Std. Error	0.064	0.062	0.062
t value	44.84	46.05	46.41
EDF	7.24	8.83	(s(<i>Chlo</i>)) = 5.81, (s(<i>SPL</i>)) = 8.83
Ref.df	8.25	8.99	(s(<i>Chlo</i>)) = 6.98, (s(<i>SPL</i>)) = 8.99
F	1.023	18.66	(s(<i>Chlo</i>)) = 5.69, (s(<i>SPL</i>)) = 23.51
p-value	<2e-16	<2e-16	(s(<i>Chlo</i>)) = 3.11e-07, (s(<i>SPL</i>)) = < 2e-16
R-sq.(adj)	0.0013	0.053	0.068
Deviance Explained	0.38%	5.63%	7.23%
GCV	11.88	12.27	11.12
Scale est.	11.85	11.23	11.06
n	2,906	2,906	2,906
AIC	15,440.84	15,286.65	15,248.46

Model 3 $CPUE \sim s(Chlo) + s(SPL)$ menjelaskan model hubungan CPUE berdasarkan *Chlo* dan SPL. Nilai t yang tinggi (46,41) menunjukkan bahwa intersep memiliki signifikansi statistik, dan nilai p yang sangat rendah untuk kedua variabel ($<2e-16$) menunjukkan bahwa efek *Chlo* dan SPL pada CPUE signifikan. Model ini memiliki R-kuadrat 0,068, yang menjelaskan sekitar 7,23% dari variasi CPUE. Model ini juga memiliki GCV 11,12, perkiraan skala 11,06, dan AIC 15.248,46. Secara keseluruhan, model 3, $CPUE \sim s(Chlo) + s(SPL)$, memberikan

penjelasan terbaik tentang variasi CPUE, karena memiliki R-kuadrat, deviasi yang ditentukan, dan AIC yang lebih baik dibandingkan model lain (Arashi et al., 2021; Bhatta et al., 2021; Clewley et al., 2008; Kumar Rajahmundry et al., 2021; Zhao et al., 2024). Gambar 6 berikut memvisualisasikan kompleksitas setiap model. Gambar 6(a) dan 6(b) menggambarkan bagaimana *Chlo* dan SPL secara individual mempengaruhi CPUE tanpa mempertimbangkan variabel lain. Gambar 6(c) menggambarkan interaksi keduanya (*Chlo* dan SPL) terhadap CPUE.



Gambar 6. Grafik 'EDF' dari satu model: (a) *Chlo*, (b) SPL, dan (c) model gabungan

Berdasarkan grafik fungsi smoothing 'EDF' di atas, tampak Model 3 memiliki prediktor yang cukup kompleks untuk *Chlo* (EDF = 5,81) dan tinggi untuk SPL (EDF = 8,82). Model 1 memiliki grafik EDF *Chlo* dengan tingkat kompleksitas yang tinggi (EDF = 7.24), dan Model 2 juga memiliki grafik EDF SPL dengan tingkat kompleksitas yang tinggi (EDF = 8.83). Model terbaik dipengaruhi oleh beberapa hal, yaitu *model fit*, *interpretability*, dan potensi *overfitting* (Wu & Abdul-Nour, 2020). Berdasarkan pertimbangan ini, Model 3 adalah pilihan terbaik karena menyeimbangkan kompleksitas di dua prediktor, yang berpotensi mengurangi risiko *overfitting* (Wang et al., 2019) dibandingkan dengan nilai EDF dari prediktor tunggal. Model 3 memiliki

nilai EDF yang seimbang untuk *Chlo* (5.81) dan SPL (8.83), menunjukkan kompleksitas prediktor sedang. Keseimbangan ini membuat Model 3 lebih optimal karena menurunkan risiko *overfitting* dan memudahkan interpretasi, dibandingkan dengan Model 1 dan Model 2, yang memiliki kompleksitas tinggi pada satu prediktor.

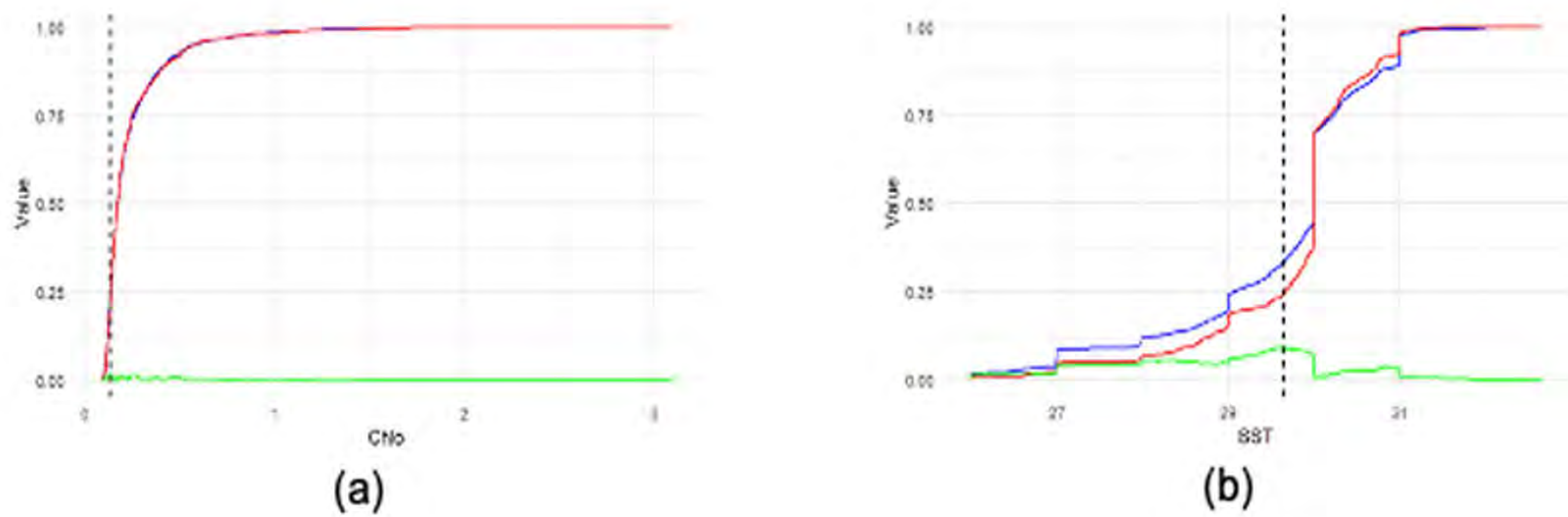
Berdasarkan hasil pemodelan GAM, secara statistik, kinerja model masih rendah, seperti yang ditunjukkan oleh nilai Std. Error $> 0,05$, dan deviance yang dijelaskan $< 50\%$ (Brandt et al., 2024). Kondisi ini dapat disebabkan oleh kurangnya data atau sampel yang tidak representatif (Noguchi et al., 2021), penggunaan model yang terlalu

sederhana atau tidak sesuai dengan kompleksitas data (Smith, 2020), multikolinearitas tinggi antara variabel prediktor (Abu et al., 2023), dan kemungkinan data hilang (Mundo et al., 2022). Nilai rumus model GAM yang dihasilkan adalah $CPUE = 2.86 + s(Chlo, edf=5.81, Ref.df=6.98) + s(SPL, edf=8.83, Ref.df=8.99)$, dengan nilai signifikan model GAM untuk *Chlo* adalah 0.542 mg/m³, dan *SPL* adalah 30.03 °C, dengan nilai CPUE tertinggi adalah 5.26. Karena model GAM menunjukkan kinerja rendah, penentuan nilai optimal *Chlo* dan *SPL* selanjutnya dilakukan menggunakan model ECDF untuk menghindari bias, mengingat GAM memerlukan asumsi khusus tentang bentuk hubungan fungsional antara variabel respons dan prediktor (Chang et al., 2021; Wood, 2020). Dibandingkan dengan itu, ECDF mampu menggambarkan distribusi data secara lebih baik dan mendeteksi *outlier* ekstrem (Aryal et al., 2022; Li et al., 2023; Lv et al., 2022; Zaghwan & Gunawan, 2021). Selain itu, metode ini mudah diterapkan dan ditafsirkan, dapat memberikan wawasan tambahan terhadap model lain, memiliki keandalan pada berbagai skala data, serta relevan dalam mengidentifikasi hubungan antara CPUE dan kondisi oseanografi (Maddumage et al., 2023; Morsy et al., 2021; Pant et al., 2023; Zhu & Shao, 2023). ECDF digunakan untuk mengatasi bias dan memberikan pemahaman yang lebih akurat, terutama dalam mendeteksi *outlier*, sehingga hubungan antara CPUE dan kondisi oseanografi dapat

ditafsirkan secara lebih representatif.

Grafik ECDF digunakan untuk melihat distribusi data setiap parameter. Bagan ECDF menemukan nilai yang memberikan perbedaan maksimum antara dua fungsi (Majumdar et al., 2019). Hubungan signifikan antara *Chlo* dan *SPL* dengan CPUE di Tuna Mata Besar ditunjukkan pada Gambar 7. Gambar 7(a) dan 7(b) di bawah ini menunjukkan plot ECDF untuk variabel *Chlo* dan *SPL*, yang mewakili nilainya dalam data. Garis-garis pada grafik menunjukkan ECDF dari data *Chlo* dan *SPL*. Plot mengungkapkan bahwa nilai maksimum untuk *Chlo* adalah 0,1325 mg/m³, sedangkan untuk *SPL*, adalah 29,64°C. Namun, nilai-nilai ini tidak secara otomatis menunjukkan nilai *Chlo* dan *SPL* yang signifikan untuk penangkapan ikan Tuna Mata Besar di WPP 714. Untuk menentukan signifikansi kedua parameter ini, uji Kolmogorov-Smirnov 2 variabel ($CPUE \sim Chlo$ dan $CPUE \sim SPL$) dilakukan untuk menguji kesesuaian antara sampel data dan distribusi yang telah ditentukan (X. Li et al., 2024).

Hasil pengujian menunjukkan bahwa nilai p untuk *Chlo* adalah 0,29 ($p > 0,05$), sehingga ditolak sebagai nilai yang signifikan. Nilai p untuk *SPL* sangat kecil ($9.27e-21$), menunjukkan perbedaan yang sangat signifikan antara distribusi $f(t)$ dan $g(t)$ pada $SPL = 29.64$. Selanjutnya, dilakukan analisis segmentasi *k-means clustering*, yang bermanfaat untuk memilih nilai *Chlo* yang signifikan dan optimal dalam dataset (Nistane, 2024).



Gambar 7. Nilai signifikan (a) Chlo dan (b) SPL terhadap CPUE

Hasil pengelompokan di bawah ini, diuji signifikansi menggunakan *Analysis of Variance* (ANOVA), menghasilkan nilai $Pr(>F) < 2e-16$ (sangat signifikan). Tes ANOVA menentukan apakah ada perbedaan substansial antara rata-rata beberapa kelompok atau perlakuan dalam sampel (Danino *et al.*, 2024). Uji HSD, dengan nilai $p_{adj} = 0$, juga menunjukkan perbedaan yang berbeda secara signifikan antara *cluster* yang dibandingkan (nilai $p < 0,05$). Tukey HSD berguna dalam menentukan perbedaan yang signifikan antar kluster setelah uji ANOVA (Molaei *et al.*, 2024). Dari

Tabel 4 di bawah ini, *cluster 2* dipilih sebagai *cluster* terbaik karena nilai CPUE rata-rata tertinggi (12,11). Nilai *Chlo* yang sangat signifikan adalah '*median_Chlo*' sebesar 0,201 mg/m3 karena distribusinya cenderung miring karena *outlier* (Hernanda *et al.*, 2023). Pentingnya nilai median ini menunjukkan bahwa terlepas dari *outlier*, *Chlo* tetap relevan pada tingkat ini dalam konteks pengoptimalan CPUE. Pendekatan *clustering* ini dapat mengatasi variasi data dan memberikan panduan yang lebih tepat dalam analisis ekosistem perikanan.

Tabel 4. Ringkasan kluster

Cluster	mean Chlo	median Chlo	mean CPUE	median CPUE
1	0.2457	0.1771	1.1687	0.9252
2	0.2607	0.2006	13.4686	12.1111
3	0.2263	0.1657	5.3753	5.0000

Metode siku menentukan kisaran nilai optimal dalam *Chlo* dan SPL untuk memancing tuna mata besar. Ini menentukan nilai kluster berdasarkan model yang dibuat (Hacıefendioğlu *et al.*, 2024), yaitu, *Chlo* dan SPL. Nilai kluster *Chlo* dan SPL digambarkan pada Gambar 8 di bawah ini. Gambar 8(a) menunjukkan bahwa titik 'siku' pada grafik *Chlo* adalah 3, sedangkan Gambar 8(b) adalah titik 'siku' untuk gugus SPL adalah 4. Titik-titik 'siku' ini dimasukkan ke dalam persamaan fungsi 'kmeans'. Menurut persamaan, kisaran nilai optimal untuk *Chlo* dalam

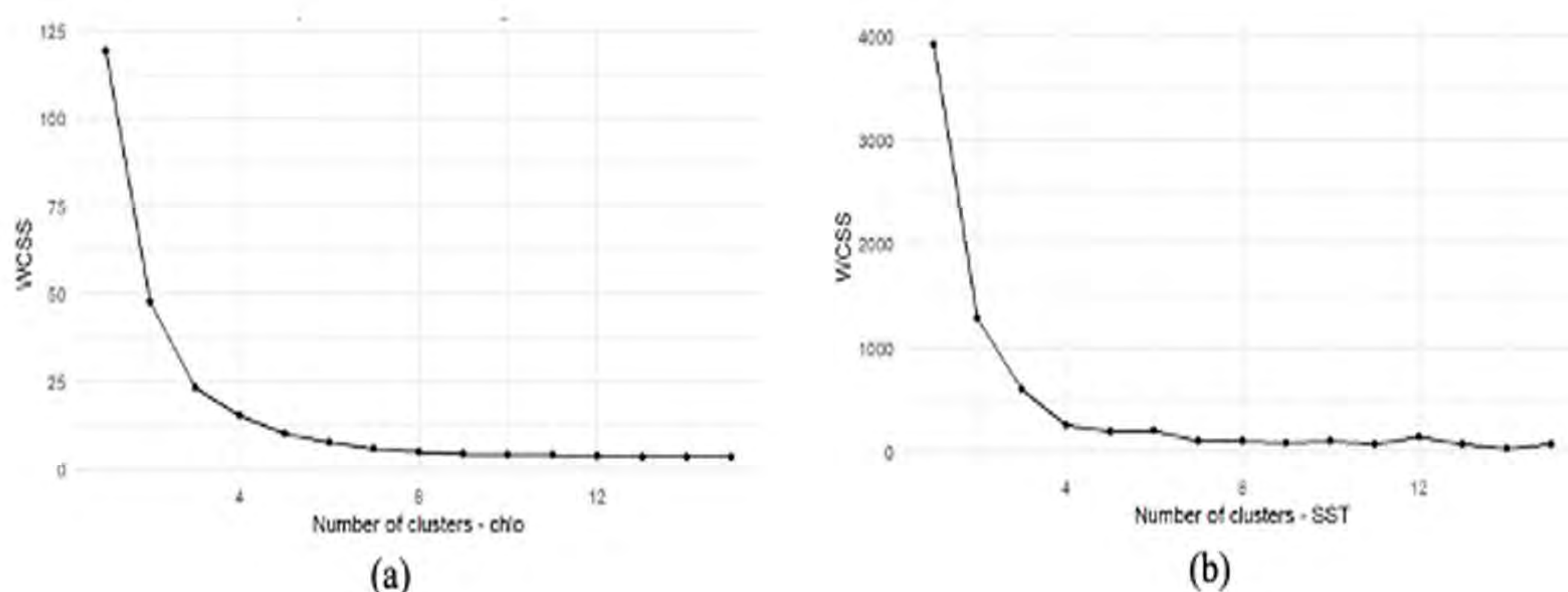
penangkapan ikan tuna mata besar di WPP 714 adalah 0,087 mg/m3 – 0,30 mg/m3. Kisaran nilai optimal untuk SPL adalah 29.35°C – 30.43°C. Nilai optimal *Chlo* dan SPL di atas kemudian divisualisasikan pada peta potensi lokasi penangkapan ikan tuna mata besar per bulan di WPP 714, seperti yang ditunjukkan pada Gambar 9 di bawah ini. Gambar 9(a) hingga 9(l) adalah peta kondisi optimal potensial di perairan WPP 714 dari Januari hingga Desember, berdasarkan nilai *Chlo* dan SPL. Pendekatan ini mempertimbangkan fluktuasi musiman, memungkinkan

penangkapan ikan disesuaikan untuk menghindari periode rentan, sehingga mengurangi risiko penangkapan ikan berlebihan dan mendukung keberlanjutan perikanan dan ekosistem laut. Peningkatan konsentrasi *Chlo* dan SPL dalam kisaran optimal di atas menciptakan zona subur yang mendukung keberadaan tuna, meningkatkan CPUE di area dan waktu tertentu. Metode ECDF memetakan distribusi optimal ini setiap bulan, memandu lokasi ideal bagi nelayan sesuai dengan perubahan musiman.

Pada Gambar 9 di bawah ini, '*Chlo*_LOW' mewakili konsentrasi rendah *Chlo* yang ditunjukkan dalam warna biru muda, '*Chlo*_OptLOW' menunjukkan nilai batas bawah konsentrasi optimal *Chlo* dalam warna hijau, '*Chlo*_OptHIGH' menunjukkan nilai batas atas konsentrasi optimal *Chlo* dalam warna kuning, dan '*Chlo*_HIGH' mewakili konsentrasi tinggi *Chlo* dalam warna biru. Selain itu, '*SPL*_LOW'

menunjukkan SPL dengan nilai rendah dalam warna oranye, '*SPL*_OptLOW' berarti nilai batas bawah SPL optimal dalam warna merah muda, '*SPL*_OptHIGH' menunjukkan nilai batas atas SPL yang optimal dalam warna ungu, dan '*SPL*_HIGH' menunjukkan nilai SPL tinggi dalam warna merah. Adapun 'BET Actual Fishing Sites', ditampilkan dalam warna coklat, '*Chlo* Outlier' adalah *outlier* dari *Chlo* dalam warna abu-abu, dan '*SPL* Outlier' menunjukkan *outlier* SPL dalam warna hitam. Nilai *outlier* (*Chlo* dan SPL) yang ditampilkan di peta adalah sepuluh besar.

Berdasarkan peta pada Gambar 9, Januari hingga Maret merupakan periode terbaik untuk menangkap ikan tuna mata besar, dengan area yang luas menunjukkan kondisi optimal. April hingga Juni menunjukkan penurunan kondisi dengan luas optimal yang semakin terbatas, dan Juli hingga September, November, dan Desember menunjukkan kondisi yang kurang ideal dengan luas optimal yang sangat terbatas.



Gambar 8. Grafik nilai kluster berdasarkan metode siku
(a) Chlo, dan (b) SPL

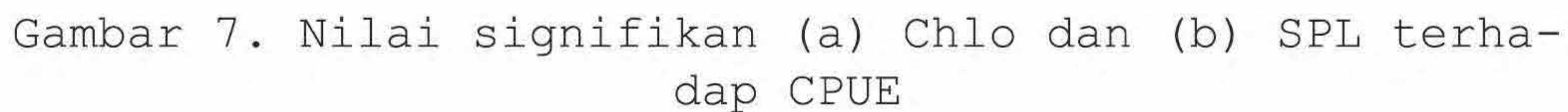
Penelitian sebelumnya tentang tangkapan optimal tuna mata besar dilaporkan dalam kondisi nilai *Chlo* dan SPL bervariasi: 0,04 – 0,16 mg/m³ dan 24°C – 27,5°C (Syamsuddin et al., 2013), juga 0,05 mg/m³ – 0,17 mg/m³ dan 24,80°C – 28,70°C

(Setiawati et al., 2015). Variasi ini penting karena kondisi lingkungan yang berbeda mempengaruhi distribusi pakan dan suhu ideal untuk tuna, membantu strategi penangkapan ikan adaptif untuk menjaga produktivitas dan keberlanjutan populasi.

real-time dan teknik pembelajaran mendalam untuk meningkatkan akurasi prediksi lokasi penangkapan ikan adaptif.

DAFTAR PUSTAKA

- Abu, M., Mullah, S., Hossain, Z., Benedetti, A., & Materials, S. (2023). Comparing Estimation Approaches for Generalized Additive Mixed Models with Binary Outcomes. *Journal of Statistical Computation and Simulation*, 93(18), 3362-3381. <https://doi.org/https://doi.org/10.1080/00949655.2023.2222864>
- Alfianto, N., Suherman, A., Suryanti, S., & Hernuryadin, Y. (2023). Strategy for Implementation of Seaworthiness of Large Pelagic Purse Seine at Nizam Zachman Ocean Fishing Port. *Sustainability (Switzerland)*, 15(18). <https://doi.org/10.3390/su15>



813713

<https://doi.org/10.1051/bioconf/202413604004>

- Antonio May-Kú, M., Ornelas-Roa, M., & Suárez-Morales, E. (2022). Surface copepod assemblages in shallow coastal waters off northeastern Yucatan Peninsula influenced by the Yucatan upwelling. *Regional Studies in Marine Science*, 56(102718). <https://doi.org/https://doi.org/10.1016/j.rsma.2022.102718>
- Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies. *PLoS ONE*, 16(4 April). <https://doi.org/10.1371/journal.pone.0245376>
- Artetxe-Arrate, I., Fraile, I., Marsac, F., Farley, J. H., Rodriguez-Ezpeleta, N., Davies, C. R., Clear, N. P., Grewe, P., & Murua, H. (2021). A review of the fisheries, life history and stock structure of tropical tuna (skipjack *Katsuwonus pelamis*, yellowfin *Thunnus albacares* and bigeye *Thunnus obesus*) in the Indian Ocean. In *Advances in Marine Biology* (Vol. 88, pp. 39-89). Academic Press. <https://doi.org/10.1016/bs.amb.2020.09.002>
- Aryal, A., Acharya, A., & Kalra, A. (2022). Assessing the Implication of Climate Change to Forecast Future Flood Using CMIP6 Climate Projections and HEC-RAS Modeling. *Forecasting*, 4(3), 582-603. <https://doi.org/10.3390/forecast4030032>
- Asuhadi, S., Zainuddin, M., Safruddin, & Musbir, M. (2024). Optimizing Chlo and SPL for skipjack tuna fisheries in the banda sea and its surroundings using GAM and ECDF. In *BIO Web of Conferences*, 136, 04004. The 13th International and National Seminar of Fisheries and Marine Science (ISFM XIII 2024).
- Bhatta, K. P., Robson, B. A., Suwal, M. K., & Vetaas, O. R. (2021). A pan-Himalayan test of predictions on plant species richness based on primary production and water-energy dynamics. *Frontiers of Biogeography*, 13(3), 1-18. <https://doi.org/10.21425/F5FBG49459>
- Bourenane, D., Sad Houari, N., & Taghezout, N. (2023). Agent-based Approach for the Recommendation and Unsupervised Classification of Enterprise Services in the Cloud. *International Journal of Computing and Digital Systems*, 14(1), 387-401. <https://doi.org/10.12785/ijcds/140130>
- Brancucci Martínez-Anido, C., Bolado, R., De Vries, L., Fulli, G., Vandenberg, M., & Masera, M. (2012). European power grid reliability indicators, what do they really tell? *Electric Power Systems Research*, 90, 79-84. <https://doi.org/10.1016/j.eipsr.2012.04.007>
- Brandt, D. A., Vega, E. F., & Ridley, A. J. (2024). On Generalized Additive Models for Representation of Solar EUV Irradiance. *Space Weather*, 22(3). <https://doi.org/10.1029/2023SW003680>
- Chang, C. H., Tan, S., Lengerich, B., Goldenberg, A., & Caruana, R. (2021). How Interpretable and Trustworthy are GAMs? *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 95-105. <https://doi.org/10.1145/3447548.3467453>
- Clewley, R. H., Guckenheimer, J. M., & Valero-Cuevas, F. J. (2008). Estimating effective degrees of freedom in motor systems. *IEEE Transactions on Biomedical*

- Engineering, 55(2), 430-442. <https://doi.org/10.1109/TBME.2007.903712>
- Cornic, M., & Rooker, J. R. (2021). Temporal shifts in the abundance and preferred habitats of yellowfin and bigeye tuna larvae in the Gulf of Mexico. *Journal of Marine Systems*, 217(103524). <https://doi.org/https://doi.org/10.1016/j.jmarsys.2021.103524>
- Danino, Y. M., Rabinovitz, R., Kirshenboim, I., Palzur, E., Pick, C. G., Ish-Shalom, I., Golovkin, Y., & Arieli, Y. (2024). Exposure to hyperbaric O₂ levels leads to blood-brain barrier breakdown in rodents. *Fluids and Barriers of the CNS*, 21(1). <https://doi.org/10.1186/s12987-024-00543-7>
- Dimitrova, D. S., Kaishev, V. K., & Tan, S. (2020). Computing the kolmogorov-smirnov distribution when the underlying cdf is purely discrete, mixed, or continuous. *Journal of Statistical Software*, 95, 1-42. <https://doi.org/10.18637/jss.v095.i10>
- Fachruddin Syah, A., Lumban Gaol, J., Zainuddin, M., Apriliya, N. R., Berlianty, D., & Mahabrur, D. (2019). Habitat Model Development of Bigeye tuna (*Thunnus obesus*) during Southeast Monsoon in the Eastern Indian Ocean using Satellite Remotely Sensed Data. *IOP Conference Series: Earth and Environmental Science*, 276(1). <https://doi.org/10.1088/1755-1315/276/1/012011>
- Friedland, D. , Morse, E. , Shackell, N. , Tam, C. , Morano, L. , Moisan, R. , & Brady, C. (2020). Changing Physical Conditions and Lower and Upper Trophic Level Responses on the US Northeast Shelf. *Frontiers in Marine Science*, 7. <https://doi.org/10.3389/fmars.2020.567445>
- Ginting, T. A., Barus, T. A., & Wahyuningsih, H. (2023). Phytoplankton abundance and trophic status of Belawan Waters, North Sumatra. *IOP Conference Series: Earth and Environmental Science*, 1241(1). <https://doi.org/10.1088/1755-1315/1241/1/012122>
- Guillotreau, P., Salladarré, F., Capello, M., Dupaix, A., Floc'h, L., Tidd, A., Tolotti, M., & Dagorn, L. (2024). Is FAD fishing an economic trap? Effects of seasonal closures and other management measures on a purse-seine tuna fleet. *Fish and Fisheries*, 25(1), 151-167. <https://doi.org/10.1111/faf.12799>
- Hacıefendioğlu, K. , Adanur, S. , & Demir, G. (2024). Automatic Landslide Segmentation Using a Combination of Grad-CAM Visualization and K-means clustering Techniques. *Iran J Sci Technol Trans Civ Eng*, 48, 943-959. <https://doi.org/https://doi.org/10.1007/s40996-023-01193-9>
- Hargiyatno, I. T., Anggawangsa, R. F., Natsir, M., Sedana, I. G. B., Widodo, A. A., & Wudianto. (2021). Comparison of handline tuna catches in Indian Ocean and Banda Sea waters. *E3S Web of Conferences*, 322. <https://doi.org/10.1051/e3sconf/202132203004>
- Hernanda, M., Salma, A., Vionanda, D., & Martha, Z. (2023). Application of the Self-Organizing Maps Method in Clustering Based on Indicators of Need for Social Welfare Services in West Java Province. *UNP Journal of Statistics and Data Science*, 1(4), 329-336. <https://doi.org/10.24036/ujsds/vol1-iss4/82>
- Hu, M., Li, M., & Kong, L. (2024). Trustworthy regularized huber regression for outlier detection. *Journal of Statistical Computation*

- Computation and Simulation*, 94(5), 1121-1137. <https://doi.org/10.1080/00949655.2023.2281643>
- Kizenga, H. J., Jebri, F., Shaghude, Y., Raitos, D. E., Srokosz, M., Jacobs, Z. L., Nencioli, F., Shalli, M., Kyewalyanga, M. S., & Popova, E. (2021). Variability of mackerel fish catch and remotely-sensed biophysical controls in the eastern Pemba Channel. *Ocean and Coastal Management*, 207. <https://doi.org/10.1016/j.ocecoaman.2021.105593>
- Kumar Rajahmundry, G., Garlapati, C., Senthil Kumar, P., Surya Alwi, R., & Vo, D.-V. N. (2021). Statistical analysis of adsorption isotherm models and its appropriate selection. *Chemosphere*, 276(130176). <https://doi.org/https://doi.org/10.1016/j.chemosphere.2021.130176>
- Li, G., Zhao, L., Tang, W., Wu, L., & Ren, J. (2023). Physica A: Statistical Mechanics and its Applications Modeling and analysis of mandatory lane-changing behavior considering heterogeneity in means and variances. *Physica A: Statistical Mechanics and Its Applications*, 625(128825). <https://doi.org/https://doi.org/10.1016/j.physa.2023.128825>
- Li, X., Marcus, D., Russell, J., Aboagye, E. O., Ellis, L. B., Sheeka, A., Park, W. H. E., Bharwani, N., Ghaem-Maghami, S., & Rockall, A. G. (2024). Weibull parametric model for survival analysis in women with endometrial cancer using clinical and T2-weighted MRI radiomic features. *BMC Medical Research Methodology*, 24(1). <https://doi.org/10.1186/s12874-024-02234-1>
- Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., & Chen, G. H. (2023). ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions. *IEEE Transactions on Knowledge and Data Engineering*, 35(12), 12181-12193. <https://doi.org/10.1109/TKDE.2022.3159580>
- Lin, H., Wang, J., Zhu, J., & Chen, X. (2023). Evaluating the impacts of environmental and fishery variability on the distribution of bigeye tuna in the Pacific Ocean. *ICES Journal of Marine Science*, 80(10), 2642-2656. <https://doi.org/10.1093/icesjms/fsad163>
- Lv, M., Gao, X., Liu, Y., Ju, W., Sun, B., Li, W., & Zhou, X. (2022). The impact of using different probability representations in application of equidistant quantile matching for bias adjustment of daily precipitation over the Daqing River Basin, North China. *International Journal of Climatology*, 42(2), 777-796. <https://doi.org/10.1002/joc.7272>
- Maddumage, U. S., Rajapaksha, J., & Gunatilake, J. (2023). Effect of ocean circulation and Chlorophyll-a concentration on yellowfin tuna catch rates in Sri Lankan logline fishery. *Ceylon Journal of Science*, 52(3), 371-379. <https://doi.org/10.4038/cjs.v52i3.8208>
- Majumdar, C., Lopez-Benitez, M., & Merchant, S. N. (2019, February 25). Experimental Evaluation of the Poissoness of Real Sensor Data Traffic in the Internet of Things. *2019 16th IEEE Annual Consumer Communications and Networking Conference, CCNC 2019*. <https://doi.org/10.1109/CCNC.2019.8651702>
- Molaei, M., Mohammadzadeh, A., Ghasemi, A., & Badiie, M. (2024). Effect of dry and wet finishing and

- polishing on color change and opacity of nanofill and nanohybrid composites. *BMC Oral Health*, 24(1). <https://doi.org/10.1186/s12903-024-03944-0>
- Morsy, M., Taghizadeh-Mehrjardi, R., Michaelides, S., Scholten, T., Dietrich, P., & Schmidt, K. (2021). Optimization of rain gauge networks for arid regions based on remote sensing data. *Remote Sensing*, 13(21). <https://doi.org/10.3390/rs13214243>
- Mundo, A. I., Muldoon, T. J., & Tipton, J. R. (2022). Generalized additive models to analyze non-linear trends in biomedical longitudinal data using R: Beyond repeated measures ANOVA and Linear Mixed Models. *BioRxiv*, 1-31. <https://doi.org/10.1101/2021.06.10.447970>
- Nagy, A. J., Freeman, M. C., Irwin, B. J., & Wenger, S. J. (2024). Life-history connections to long-term fish population trends in a species-rich temperate river. *Ecology of Freshwater Fish*, 33(2). <https://doi.org/10.1111/eff.12767>
- Nistane, V. (2024). Optimum prediction model of remaining useful life for rolling element bearing based on integrating optimize health indicator (OHI) and machine learning algorithm. *World Journal of Engineering*, 21(1), 170-185. <https://doi.org/10.1108/WJE-06-2022-0244>
- Noguchi, K., Konietzschke, F., Marmolejo-Ramos, F., & Pauly, M. (2021). Permutation tests are robust and powerful at 0.5% and 5% significance levels. *Behavior Research Methods*, 53(6), 2712-2724. <https://doi.org/10.3758/s13428-021-01595-5>
- Ovando, D., Libecap, G. D., Millage, K. D., & Thomas, L. (2021). Coasean approaches to address overfishing: Bigeye tuna conservation in the western and central pacific ocean. *Marine Resource Economics*, 36(1), 91-109. <https://doi.org/10.1086/711853>
- Pant, M., Bhatla, R., Ghosh, S., Das, S., & Mall, R. K. (2023). Will Warming Climate Affect the Characteristics of Summer Monsoon Rainfall and Associated Extremes Over the Gangetic Plains in India? *Earth and Space Science*, 10(2). <https://doi.org/10.1029/2022EA002741>
- Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, 2019(5). <https://doi.org/10.7717/peerj.6876>
- Peng Lian, & Le Gao. (2024). Impacts of central-Pacific El Niño and Physical Drivers on Eastern Pacific Bigeye Tuna. *Journal of Oceanology and Limnology*. <https://doi.org/https://doi.org/10.1007/s00343-023-3051-3>
- Pimentel-Acosta, C. A., Caña-Bozada, V. H., Osuna-Cabanillas, J. M., Fajer-Ávila, E. J., Ovando-Vásquez, C., & Morales-Serna, F. N. (2023). Population and Transcriptomic Changes of the Tropical Fish Parasite *Caligus confusus* (Copepoda: Caligidae) with Seasonal Variations in Sea Temperature. *Fishes*, 8(10). <https://doi.org/10.3390/fishes8100475>
- Punt, A. E., Castillo-Jordán, C., Hamel, O. S., Cope, J. M., Maunder, M. N., & Ianelli, J. N. (2021). Consequences of error in natural mortality and its estimation in stock assessment models. *Fisheries Research*, 233(105759). <https://doi.org/https://doi.org/10.1016/j.fishr>

.2020.105759

- Reda Abonazel, M. (2020). Handling outliers and Missing Data in Regression Models Using R: Simulation Examples. *Academic Journal of Applied Mathematical Sciences*, 68, 187-203. <https://doi.org/10.32861/ajams.68.187.203>
- Sean Bailey. (2023, January 1). NASA Ocean Color. <https://oceancolor.gsfc.nasa.gov/13/>
- Setiawati, M. D., Sambah, A. B., Miura, F., & Tanaka, T. (2015). Characterization of bigeye tuna habitat in the Southern Waters off Java-Bali using remote sensing data. *Advances in Space Research*, 55(2), 732-746. <https://doi.org/https://doi.org/10.1016/j.asr.2014.10.007>
- Smith, R. J. (2020). $P > .05$: The incorrect interpretation of "not significant" results is a significant problem. In *American Journal of Physical Anthropology* (Vol. 172, Issue 4, pp. 521-527). Wiley-Liss Inc. <https://doi.org/10.1002/ajpa.24092>
- Soares, R., & Castro, W. (2020). Application of Tukey's Test for Statistical Measurement of Percentage Indexes Derived from the Difference Between Predicted Crystallographic Data in Five Nitihf Alloys Obtained by Two Processes. *International Journal of Engineering Research & Technology (IJERT)*, 9(03), 45-53. <https://doi.org/https://doi.org/10.17577/IJERTV9IS030118>
- Susko, E., & Roger, A. J. (2020). On the Use of Information Criteria for Model Selection in Phylogenetics. *Molecular Biology and Evolution*, 37(2), 549-562. <https://doi.org/10.1093/molbev/msz228>
- Syamsuddin, M. L., Saitoh, S. I., Hirawake, T., Samsul, B., & Harto, A. B. (2013). Effects of El Niño-Southern Oscillation events on catches of Bigeye Tuna (*Thunnus obesus*) in the eastern Indian Ocean off Java. *Fishery Bulletin*, 111(2), 175-188. <https://doi.org/10.7755/FB.111.2.5>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Thomson/Brooks/Cole.
- Wang, L., Abdel-Aty, M., Lee, J., & Shi, Q. (2019). Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis & Prevention*, 122, 378-384. <https://doi.org/https://doi.org/10.1016/j.aap.2017.06.003>
- Wood, S. N. (2020). Inference and computation with generalized additive models and their extensions. *Test*, 29(2), 307-339. <https://doi.org/10.1007/s11749-020-00711-5>
- Wu, Z., & Abdul-Nour, G. (2020). Comparison of Multi-Criteria Group Decision-Making Methods for Urban Sewer Network Plan Selection. *CivilEng*, 1(1). <https://doi.org/10.3390/civileng1010003>
- Yang, J., Rahardja, S., & Fränti, P. (2019, December 19). Outlier detection: How to threshold outlier scores? *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3371425.3371427>
- Yati, E., Sadiyah, L., Satria, F., Alabia, I. D., Sulma, S., Prayogo, T., Marpaung, S., Harsa, H., Kushardono, D., Lumban-Gaol, J., Budiarto, A., Efendi, S., & Patmiarsih, S. (2024). Spatial distribution models for the four commercial tuna in the sea of

maritime continent using
multi-sensor remote sensing and
maximum entropy. *Marine
Environmental Research*, 198(106540).
<https://doi.org/https://doi.org/10.1016/j.marenvres.2024.106540>

Zaghwan, A., & Gunawan, I. (2021).
Energy loss impact in electrical
smart grid systems in australia.
Sustainability (Switzerland),
13(13). <https://doi.org/10.3390/sul3137221>

Zendrato, N., Dhany, H. W., Siagian, N.
A., & Izhari, F. (2020). Bigdata
Clustering using X-means method with
Euclidean Distance. *Journal of
Physics: Conference Series*, 1566(1).
<https://doi.org/10.1088/1742-6596/1566/1/012103>

Zhao, Y. C., Sun, Z. H., Xiao, M. X.,
Li, J. K., Liu, H. yuan, Cai, H. L.,
Cao, W., Feng, Y., Zhang, B. K., &
Yan, M. (2024). Analyzing the
correlation between
quinolone-resistant *Escherichia coli*
resistance rates and climate
factors: A comprehensive analysis
across 31 Chinese provinces.
Environmental Research, 245.
<https://doi.org/10.1016/j.envres.2023.117995>

Zhu, M. X., & Shao, Y. H. (2023).
Classification by Estimating the
Cumulative Distribution Function for
Small Data. *IEEE Access*, 11,
41142-41157. <https://doi.org/10.1109/ACCESS.2023.3269504>.