

Rainfall Forecasting Using Gaussian Process Regression with Brownian Motion Prior (Case Study: Special Region of Yogyakarta Province)

Peramalan Curah Hujan Menggunakan *Gaussian Process Regression* dengan *Prior* Gerak Brown (Studi Kasus: Provinsi Daerah Istimewa Yogyakarta)

Syifaul Janan

Program Studi Teknik Mesin, Universitas Pembangunan Nasional Veteran Jakarta

Email: syifaul.janan@upnvj.ac.id

Abstract

Climate variability significantly impacts the agricultural sector, necessitating accurate forecasting methods to support agricultural planning. This study aims to develop a rainfall forecasting model using the Gaussian Process Regression (GPR) method with Brownian motion prior. Monthly climate data from the Yogyakarta Geophysics Station for the period January 2015 to December 2024 were utilized, comprising predictor variables (air humidity and wind speed) and response variable (rainfall). The posterior GPR model was developed for parameter estimation using the marginal log-likelihood approach, with missing data handled through seasonal mean imputation that preserves temporal patterns. The results demonstrate that the GPR model achieves reasonable forecasting performance with a Mean Absolute Percentage Error (MAPE) of 36.84% and strong correlation ($r = 0.94$) between predicted and actual values. The highest predicted rainfall occurred in March 2024 (20.148 mm) and the lowest in June 2024 (0.022 mm), consistent with the seasonal patterns of Indonesia's tropical climate.

Keywords: Brownian motion, forecasting, Gaussian Process Regression, parameter estimation, rainfall

Abstrak

Perubahan pola curah hujan akibat variabilitas iklim berdampak signifikan terhadap sektor pertanian, sehingga diperlukan metode peramalan yang akurat untuk mendukung perencanaan pertanian. Penelitian ini bertujuan untuk mengembangkan model peramalan curah hujan menggunakan metode Gaussian Process Regression (GPR) dengan *prior* Gerak Brown. Data yang digunakan adalah data iklim bulanan dari Stasiun Geofisika Yogyakarta periode Januari 2015 hingga Desember 2024, meliputi variabel prediktor (kelembapan udara dan kecepatan angin) serta variabel respon (curah hujan). Model *posterior* GPR dikembangkan untuk estimasi parameter menggunakan pendekatan *marginal log-likelihood*, dengan penanganan data hilang melalui metode *mean* musiman yang mempertahankan pola temporal. Hasil penelitian menunjukkan bahwa model GPR mencapai performa peramalan yang *reasonable* dengan nilai *Mean Absolute Percentage Error*



(MAPE) sebesar 36,84% dan korelasi yang kuat ($r = 0,94$) antara nilai prediksi dan aktual. Prediksi curah hujan tertinggi terjadi pada Maret 2024 (20,148 mm) dan terendah pada Juni 2024 (0,022 mm), yang konsisten dengan pola musiman iklim tropis Indonesia.

Kata kunci: Gerak Brown, peramalan, Proses Regresi Gaussian, estimasi parameter, curah hujan

1. PENDAHULUAN

Curah hujan merupakan unsur iklim yang sangat berpengaruh terhadap sektor pertanian [12]. Curah hujan berperan penting dalam sistem irigasi dan pemeliharaan kesuburan tanah. Perubahan pola curah hujan dapat mempengaruhi waktu tanam, produktivitas, dan kualitas hasil panen [10]. Saat ini, perubahan iklim global menyebabkan cuaca ekstrem seperti banjir dan kekeringan semakin sering terjadi. Kondisi ini membuat petani membutuhkan sistem peramalan curah hujan yang akurat.

Selama ini, petani masih mengandalkan pengetahuan turun-temurun untuk menentukan jadwal tanam berdasarkan perkiraan musim hujan. Namun, pendekatan tradisional tersebut semakin tidak relevan dengan dinamika iklim saat ini. Akibatnya, pola tanam menjadi kurang optimal dan risiko gagal panen meningkat. Peramalan curah hujan yang akurat dapat membantu petani merencanakan kegiatan pertanian dengan lebih baik, mengoptimalkan penggunaan air, dan mengurangi kerugian akibat ketidakpastian cuaca [13]. Oleh karena itu, pengembangan metode peramalan curah hujan berbasis pendekatan kuantitatif menjadi kebutuhan yang mendesak.

Perkembangan teknologi *machine learning* membuka peluang baru dalam peramalan curah hujan. Berbagai metode telah digunakan untuk prediksi meteorologi, seperti *neural networks*, *support vector machines*, dan metode *ensemble* seperti *random forest* [7]. Penelitian [3] menyatakan bahwa teknik *machine learning*, khususnya *deep learning*, menunjukkan kemajuan signifikan dalam memproses data besar dan mengekstrak pola untuk prediksi cuaca. Di antara berbagai metode tersebut, *Gaussian Process Regression* (GPR) menunjukkan hasil yang menjanjikan karena mampu mengukur ketidakpastian prediksi [1] [15] [17]. Kemampuan ini sangat penting dalam peramalan cuaca yang pada dasarnya bersifat tidak pasti.

Beberapa penelitian telah mengaplikasikan GPR dalam bidang meteorologi. Das et al. [7] menggunakan GPR dengan metode *feature selection* untuk meramalkan curah hujan di India. Penelitian tersebut menggunakan fungsi kovariansi *Squared Exponential* (SE) yang mengasumsikan stasioneritas data, namun asumsi ini tidak terpenuhi pada data curah hujan yang memiliki tren temporal dan variabilitas musiman yang kuat. Ehteram et al. [8] mengembangkan model *hybrid* yang menggabungkan CNN, SVM, dan GPR untuk prediksi curah hujan di Malaysia menggunakan kernel *Matern*. Meskipun kernel *Matérn* lebih fleksibel dalam menangani non-stasioneritas, metode ini memerlukan estimasi 3-4 *hyperparameter* yang meningkatkan kompleksitas komputasi dan risiko *overfitting*, terutama pada dataset berukuran kecil hingga sedang. Sementara itu, Penelitian Chen et al. [6] menyatakan bahwa GPR efektif dalam menangkap pola nonlinear pada data meteorologi, namun pemilihan fungsi kovariansi yang tepat sangat krusial untuk performa model.

Meskipun GPR telah banyak digunakan, penggunaan *prior* Gerak Brown (*Brownian Motion*) dalam GPR untuk peramalan curah hujan masih sangat terbatas. Review literatur menunjukkan bahwa studi-studi sebelumnya dominan menggunakan kernel stasioner (SE, RBF) atau kernel dengan banyak parameter (*Matern*), yang masing-masing memiliki keterbatasan. *Prior* Gerak Brown menawarkan solusi yang lebih sesuai untuk data *time series* curah hujan dengan beberapa keunggulan komparatif. Pertama, *prior* ini mampu menangani sifat non-stasioner yang umum ditemui pada data iklim tanpa memerlukan transformasi data atau asumsi stasioneritas yang kaku. Kedua, *prior* ini secara eksplisit memodelkan korelasi temporal yang menurun seiring bertambahnya jarak waktu antar pengamatan, sesuai dengan karakteristik fisik proses curah hujan di mana kejadian hujan hari ini lebih berkorelasi dengan kemarin dibandingkan dengan sebulan lalu. Ketiga, struktur matriks kovariansinya hanya memerlukan 2 parameter (dibandingkan 3-4 parameter pada kernel lain) sehingga efisien secara komputasi dan mengurangi risiko *overfitting* pada dataset berukuran

sedang. Hingga saat ini, belum ada penelitian yang secara khusus mengaplikasikan GPR dengan *prior* Gerak Brown untuk peramalan curah hujan di Indonesia, padahal karakteristik iklim tropis Indonesia dengan pola musiman yang kuat dan variabilitas tinggi sangat sesuai dengan asumsi kernel ini.

Penelitian ini mengambil lokasi di Provinsi Daerah Istimewa Yogyakarta (DIY) dengan beberapa pertimbangan. Pertama, DIY merupakan salah satu sentra pertanian padi di Pulau Jawa. Pada tahun 2022, luas panen padi di DIY mencapai 110,93 ribu hektar dengan produksi 561,70 ribu ton GKG [5]. Keberhasilan pertanian di wilayah ini sangat bergantung pada ketersediaan air yang ditentukan oleh pola curah hujan. Kedua, topografi DIY sangat beragam, mulai dari dataran rendah di pesisir selatan hingga lereng Gunung Merapi di utara. Keragaman topografi ini menyebabkan variabilitas curah hujan yang tinggi akibat proses hujan orografis [16]. Kondisi tersebut menjadikan DIY sebagai lokasi yang representatif untuk menguji model peramalan. Ketiga, Stasiun Geofisika Yogyakarta yang dikelola BMKG menyediakan data iklim yang lengkap dan berkualitas tinggi untuk analisis *time series*.

Penelitian ini bertujuan untuk mengembangkan model GPR dengan *prior* Gerak Brown untuk meramalkan curah hujan di Provinsi DIY, yang merupakan aplikasi pertama dari kombinasi metode ini untuk konteks iklim tropis Indonesia. Pemilihan fungsi kovariansi Gerak Brown didasarkan pada kesesuaiannya dengan data *time series* yang memiliki struktur korelasi temporal. Selain itu, fungsi ini efisien secara komputasi karena jumlah parameternya sedikit tanpa mengabaikan akurasi, mengatasi keterbatasan kernel stasioner dan kernel kompleks yang digunakan dalam studi-studi sebelumnya. Secara spesifik, tujuan penelitian ini adalah: (1) mengembangkan model posterior GPR untuk estimasi parameter; (2) mengestimasi parameter model GPR menggunakan pendekatan *marginal log-likelihood*; dan (3) memprediksi curah hujan bulanan di Provinsi DIY serta mengevaluasi akurasi model berdasarkan nilai *Mean Absolute Percentage Error* (MAPE) dan korelasi.

2. METODE

2.1 Data Penelitian

Penelitian ini menggunakan data iklim bulanan Provinsi Daerah Istimewa Yogyakarta (DIY) periode Januari 2015 hingga Desember 2024. Data diperoleh dari Stasiun Geofisika Yogyakarta, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dengan spesifikasi sebagai berikut: ID WMO 96855, koordinat 7,82°LS dan 110,30°BT, serta elevasi 153 meter di atas permukaan laut. Data diunduh melalui portal resmi BMKG (<http://dataonline.bmkg.go.id/>) [4].

Variabel yang digunakan terdiri dari satu variabel respon dan dua variabel prediktor. Variabel respon adalah curah hujan bulanan dalam satuan milimeter (mm). Variabel prediktor meliputi kelembapan udara relatif (%) dan kecepatan angin (m/s). Pemilihan kedua variabel prediktor ini didasarkan pada hubungan fisiknya terhadap pembentukan presipitasi [2].

Total data berjumlah 120 observasi bulanan yang dibagi menjadi dua bagian. Sebanyak 108 data (Januari 2015–Desember 2023) digunakan sebagai data *training* untuk membangun model. Sisanya, 12 data (Januari 2024–Desember 2024) digunakan sebagai data *testing* untuk mengevaluasi performa model. Proporsi pembagian 90:10 ini umum digunakan dalam pemodelan *machine learning* [6].

2.2 Penanganan Data Hilang

Data hilang (*missing value*) merupakan permasalahan umum yang dapat mempengaruhi kinerja model [9]. Pada dataset ini, teridentifikasi tiga data hilang: (1) curah hujan bulan Desember 2016; (2) kelembapan udara bulan Januari 2016; dan (3) kecepatan angin bulan Mei 2020. Proporsi data

hilang untuk setiap variabel adalah 0,83% dari total observasi. Proporsi ini tergolong rendah sehingga tidak berpotensi menimbulkan bias signifikan.

Penanganan data hilang dilakukan menggunakan metode *mean imputation* dengan pendekatan musiman. Nilai imputasi dihitung berdasarkan rata-rata observasi pada bulan yang sama sepanjang periode pengamatan. Pendekatan ini dipilih untuk mempertahankan karakteristik musiman data iklim. Metode *mean imputation* efektif untuk data dengan proporsi *missing value* yang rendah [14].

2.3 Model Gaussian Process Regression

Gaussian Process Regression (GPR) merupakan metode regresi nonparametrik berbasis pendekatan Bayesian [8]. Model regresi nonlinear dengan *noise* didefinisikan sebagai:

$$y = f(\mathbf{x}) + \varepsilon \quad (1)$$

dengan y adalah variabel respon, $f(\mathbf{x})$ adalah fungsi *latent* yang tidak diketahui, dan ε adalah *noise* yang diasumsikan berdistribusi Gaussian:

$$\varepsilon \sim N(0, \sigma^2) \quad (2)$$

Proses Gaussian (GP) didefinisikan sebagai koleksi variabel acak yang memiliki distribusi Gaussian bersama:

$$f(\mathbf{x}) \sim \mathcal{N}(m(x), k(x, x')) \quad (3)$$

dengan $m(x)$ adalah fungsi *mean* yang diasumsikan bernilai nol, dan $k(x, x')$ adalah fungsi kovariansi yang menentukan struktur korelasi antar observasi.

Penelitian ini menggunakan *prior* Gerak Brown yang sesuai untuk data *time series* dengan struktur korelasi temporal. *Prior* ini dinyatakan sebagai:

$$f(\mathbf{x}) \sim \mathcal{N}(0, e^{\alpha} H) \quad (4)$$

dengan α adalah parameter skala dan H adalah matriks kovariansi dari Gerak Brown berukuran $n \times n$. Model juga memuat komponen *noise* yang dinyatakan sebagai:

$$\varepsilon_i \sim \mathcal{N}(0, e^{\beta}), i = 1, 2, \dots, n \quad (5)$$

dengan β parameter skala *noise*.

Fungsi kovariansi dari Gerak Brown didefinisikan sebagai:

$$h(x, x') = -\frac{1}{2} \left(\|x - x'\| - \frac{1}{n} \sum_j \|x - x_j\| - \frac{1}{n} \sum_i \|x_i - x'\| + \frac{1}{n^2} \sum_i \sum_j \|x_i - x_j\| \right) \quad (6)$$

dengan x dan x' adalah pasangan observasi dan $\| \cdot \|$ menyatakan jarak Euclidean.

Berdasarkan persamaan (4) dan (5), distribusi bersama dari output dapat dinyatakan sebagai:

$$(y_1, y_2, \dots, y_n) \sim \mathcal{N}(0, K_{\alpha, \beta}) \quad (7)$$

dengan matriks kovariansi

$$K_{\alpha, \beta} = e^{\alpha} H + e^{\beta} I \quad (8)$$

di mana I adalah matriks identitas berukuran $n \times n$.

2.4 Estimasi Parameter

Parameter model α dan β diestimasi menggunakan metode *Maximum Marginal Likelihood Estimation* (MMLE). Metode ini memaksimalkan probabilitas data teramati dengan mengintegrasikan fungsi *latent* f [17]. Fungsi *marginal log-likelihood* dinyatakan sebagai:

$$\log \mathcal{L}(\alpha, \beta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |K_{\alpha, \beta}| - \frac{1}{2} y^{*T} K_{\alpha, \beta}^{-1} y^* \quad (9)$$

dengan $y^* = (y_1 - \bar{Y}, y_2 - \bar{Y}, \dots, y_n - \bar{Y})^T$ adalah vektor respon terkoreksi dan $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$ adalah rata-rata respon.

Estimasi parameter diperoleh dengan menurunkan persamaan (9) terhadap α dan β , kemudian menyamakan hasilnya dengan nol. Solusi sistem persamaan nonlinear yang dihasilkan diselesaikan secara numerik menggunakan algoritma optimasi pada perangkat lunak Matlab.

2.5 Model Posterior dan Prediksi

Model posterior GPR diturunkan menggunakan teorema Bayes dengan mengkombinasikan distribusi *prior* dan *likelihood*. Model posterior dinyatakan sebagai:

$$p(f | X, Y) = \mathcal{N}(\bar{f}, A^{-1}) \quad (10)$$

dengan *mean* posterior:

$$\bar{f} = \frac{1}{e^\beta} A^{-1} Y \quad (11)$$

dan matriks

$$A = e^{-\beta} I + e^{-\alpha} H^{-1} \quad (12)$$

Prediksi curah hujan dilakukan dengan memasukkan nilai variabel prediktor ke dalam model yang parameternya telah diestimasi. Model prediksi untuk observasi ke- i dinyatakan sebagai:

$$y_i = f(x_i) + \varepsilon_i \quad (13)$$

dengan $x_i = [x_{1i}, x_{2i}]^T$ adalah vektor input yang terdiri dari kelembapan udara (x_{1i}) dan kecepatan angin (x_{2i})

Fungsi *latent* $f(x_i)$ memetakan input multivariat ke output skalar melalui distribusi Gaussian Process dengan fungsi kovariansi Gerak Brown. Jarak antara dua input multivariat dihitung menggunakan jarak Euclidean:

$$\|x_i - x_j\| = \sqrt{[(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2]} \quad (13a)$$

yang kemudian digunakan dalam persamaan (6) untuk membentuk matriks kovariansi H .

2.6 Evaluasi Model

Performa model dievaluasi menggunakan *Mean Absolute Percentage Error* (MAPE) yang dihitung pada data *testing*. MAPE didefinisikan sebagai:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (14)$$

dengan Y_i adalah nilai aktual, $\hat{Y}_i$ adalah nilai prediksi, dan n adalah jumlah observasi pada data *testing*. MAPE memberikan interpretasi *error* yang lebih intuitif dalam bentuk persentase [13]. Nilai MAPE yang lebih kecil menunjukkan akurasi prediksi yang lebih akurat.

3. HASIL dan PEMBAHASAN

3.1. Deskripsi Data dan Pendugaan Data Hilang

Data iklim bulanan Provinsi DIY yang digunakan mencakup periode 10 tahun (Januari 2015–Desember 2024) dengan total 120 observasi untuk setiap variabel. Statistik deskriptif ketiga variabel disajikan pada Tabel 1.

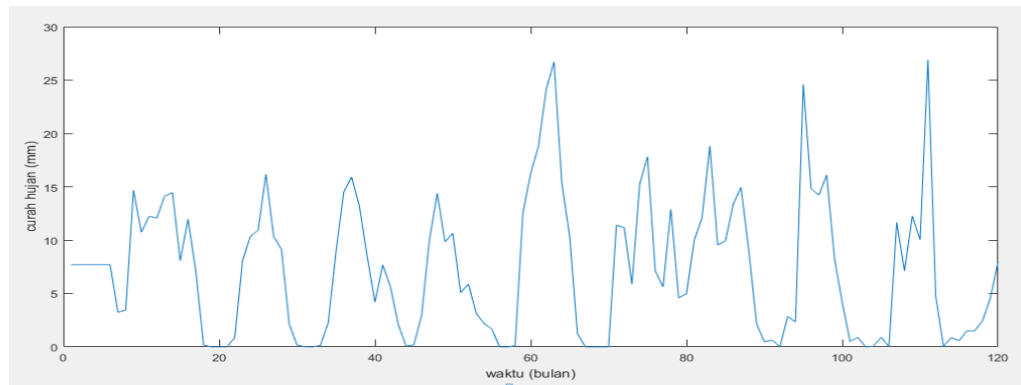
JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Syifaul Janan

Tabel 1. Statistik Deskriptif Variabel Penelitian

Variabel	Satuan	Minimum	Maksimum	Rata-Rata	Std.Deviasi
Curah Hujan	mm	0,00	26,89	6,82	5,94
Kelembapan Udara	%	65,69	90,73	83,45	5,21
Kecepatan Angin	m/s	1,55	4,32	2,48	0,52

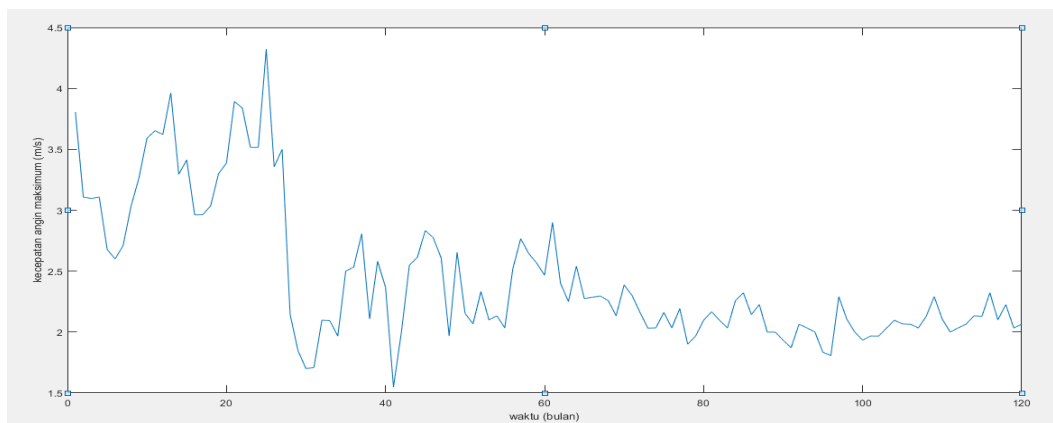
Data curah hujan menunjukkan variabilitas yang tinggi. Nilai curah hujan berkisar dari 0 mm pada bulan-bulan kering hingga 26,89 mm pada Maret 2024. Pola ini mengikuti karakteristik iklim tropis Indonesia dengan dua musim: musim hujan (Oktober–Maret) dan musim kemarau (April–September). Gambar 1 menunjukkan dinamika curah hujan selama periode pengamatan.



Gambar 1. Dinamika curah hujan bulanan di Provinsi DIY periode Januari 2015–Desember 2024

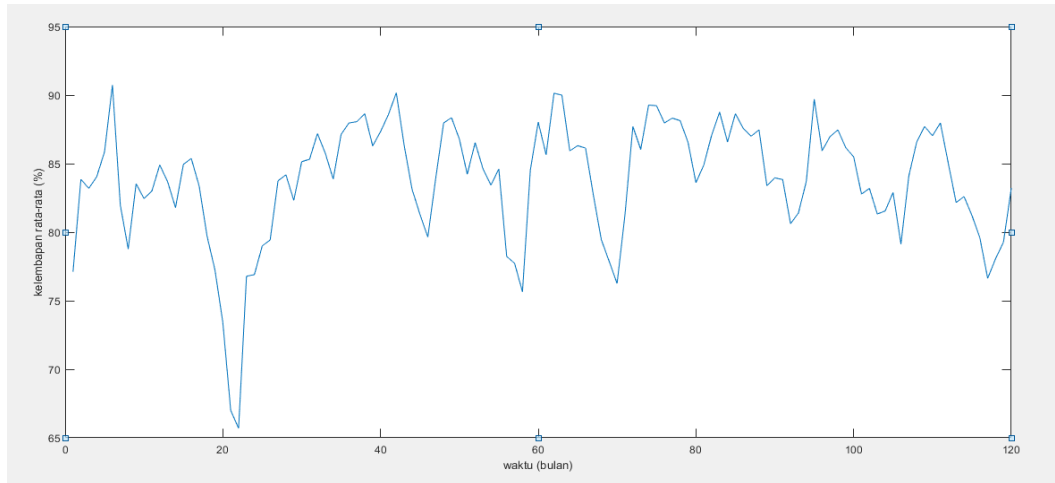
Beberapa periode menunjukkan curah hujan nol, terutama pada bulan Juli–Oktober di tahun 2016, 2019, 2020, 2022, dan 2023. Fenomena ini konsisten dengan pola musim kemarau di Indonesia. Sebaliknya, curah hujan tinggi terjadi pada November–Maret dengan puncak pada Januari–Maret.

Kecepatan angin bulanan menunjukkan variasi moderat dengan nilai 1,55 m/s (Mei 2018) hingga 4,32 m/s (Januari 2017). Rata-rata kecepatan angin berada pada kisaran 2–3 m/s yang tergolong kategori angin lemah hingga sedang. Gambar 2 menunjukkan tren penurunan kecepatan angin dari tahun 2015–2017 dibandingkan tahun-tahun setelahnya.



Gambar 2. Fluktuasi kecepatan angin bulanan di Provinsi DIY periode Januari 2015–Desember 2024

Kelembapan udara relatif menunjukkan variasi musiman yang jelas dengan nilai 65,69% (Oktober 2015) hingga 90,73% (Juni 2015). Rata-rata kelembapan di atas 75% merupakan karakteristik khas daerah tropis. Pola musiman menunjukkan kelembapan lebih tinggi pada bulan-bulan dengan curah hujan tinggi (Gambar 3).



Gambar 3. Variasi kelembapan udara bulanan di Provinsi DIY periode Januari 2015-Desember 2024

Dalam proses pengolahan data, teridentifikasi tiga data hilang. Tabel 2 menyajikan rincian data hilang dan hasil imputasinya.

Tabel 2. Hasil Imputasi Data Hilang Menggunakan Metode *Mean* Musiman

No	Variabel	Bulan/Tahun	Nilai Imputasi	Keterangan
1	Curah Hujan	Desember 2016	10,343 mm	Rata-rata bulan Desember
2	Kelembapan Udara	Januari 2016	83,677%	Rata-rata bulan Januari
3	Kecepatan Angin	Mei 2020	2,275 m/s	Rata-rata bulan Mei

Proporsi data hilang sebesar 0,83% untuk setiap variabel termasuk kategori rendah. [9] menyatakan bahwa data hilang tidak menimbulkan masalah jika jumlahnya kurang dari 5%. Metode *mean* musiman dipilih karena proporsi data hilang sangat rendah dan data memiliki pola musiman yang jelas [11]. Nilai rata-rata dihitung dari observasi pada bulan yang sama sepanjang periode pengamatan sehingga konsisten dengan pola musiman.

Setelah imputasi, dilakukan normalisasi data menggunakan metode *min-max*. Hasil normalisasi menunjukkan seluruh data berada pada rentang 0–1, mengindikasikan tidak ada *outlier* yang dapat mengganggu pemodelan.

3.2. Model Posterior GPR

Model GPR dikembangkan menggunakan pendekatan Bayesian yang menggabungkan informasi *prior* dengan data pengamatan (*likelihood*) untuk menghasilkan distribusi *posterior*. Pendekatan ini memiliki keunggulan dalam mengukur ketidakpastian prediksi, yang sangat penting dalam peramalan curah hujan [2].

Berdasarkan persamaan (1), model regresi dengan *noise* dapat ditulis sebagai $y_i = f(x_i) + \varepsilon_i$. Fungsi $f(x_i)$ diasumsikan berdistribusi normal dengan *mean* nol dan matriks kovariansi $e^{\alpha H}$ sebagaimana persamaan (4), sedangkan *noise* ε_i berdistribusi normal dengan *mean* nol dan variansi

e^β sebagaimana persamaan (5). Dengan asumsi bahwa setiap pengamatan bersifat independen, distribusi gabungan dari seluruh *output* mengikuti persamaan (7).

Fungsi densitas probabilitas dari *likelihood* dihitung berdasarkan data *training* sebagai berikut:

$$p(Y|f, X) = \prod_{i=1}^n p(y_i|x_i, f(x_i)) = \mathcal{N}(f(X), e^\beta I) \quad (15)$$

Model *posterior* diturunkan menggunakan teorema Bayes:

$$p(f|X, Y) = \frac{p(Y|f, X) \cdot p(f)}{p(Y|X)} \quad (16)$$

dengan $p(Y|X) = \int p(Y|f, X) dp(f)$ adalah konstanta normalisasi yang independen terhadap f .

Substitusi *likelihood* dan *prior* ke persamaan (16) menghasilkan:

$$p(f|X, Y) \propto \exp\left(-\frac{1}{2e^\beta} (Y - f)^T I (Y - f)\right) \cdot \exp\left(-\frac{1}{2e^\alpha} f^T H^{-1} f\right) \quad (17)$$

Setelah penyederhanaan aljabar, persamaan (17) dapat ditulis dalam bentuk kuadratik:

$$p(f|X, Y) \propto \exp\left(-\frac{1}{2} (f - \bar{f})^T A (f - \bar{f})\right) \quad (18)$$

dengan $\bar{f} = e^{-\beta} A^{-1} Y$ dan $A = e^{-\beta} I + e^{-\alpha} H^{-1}$.

Berdasarkan persamaan (18), model *posterior* berbentuk distribusi normal multivariat:

$$p(f|X, Y) = \mathcal{N}\left(\bar{f} = \frac{1}{e^\beta} A^{-1} Y, A^{-1}\right) \quad (19)$$

dengan *mean posterior* $\bar{f}$ merupakan estimasi terbaik fungsi f berdasarkan data, dan matriks kovariansi A^{-1} merepresentasikan ketidakpastian estimasi.

3.3. Estimasi Parameter

Model GPR dengan *prior* Gerak Brown memiliki dua parameter yang perlu diestimasi, yaitu α (parameter skala *signal*) dan β (parameter skala *noise*). Estimasi dilakukan dengan memaksimalkan *marginal log-likelihood*.

Fungsi *marginal likelihood* dinyatakan sebagai:

$$\mathcal{L}(\alpha, \beta) = \frac{1}{\sqrt{(2\pi)^n |K_{\alpha, \beta}|}} \exp\left(-\frac{1}{2} y^{*T} K_{\alpha, \beta}^{-1} y^*\right) \quad (20)$$

dengan $y^* = (y_1 - \bar{Y}, y_2 - \bar{Y}, \dots, y_n - \bar{Y})^T$ adalah vektor respon terkoreksi

Fungsi *marginal log-likelihood* diperoleh dengan mengambil logaritma natural dari persamaan (20):

$$\log \mathcal{L}(\alpha, \beta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |K_{\alpha, \beta}| - \frac{1}{2} y^{*T} K_{\alpha, \beta}^{-1} y^* \quad (21)$$

Karena matriks H adalah matriks real simetris, maka dapat didekomposisi menjadi $H = Q\Lambda Q^{-1}$, dengan Q adalah matriks yang kolom-kolomnya berisi vektor eigen dan Λ adalah matriks diagonal yang berisi nilai eigen λ_i . Dengan dekomposisi ini, matriks kovariansi dapat ditulis sebagai:

$$K_{\alpha, \beta} = \sum_i (e^\alpha \lambda_i + e^\beta) a_i a_i^T \quad (22)$$

dengan a_i adalah vektor eigen dari H .

Determinan dan invers matriks kovariansi masing-masing adalah:

$$|K_{\alpha, \beta}| = \prod_i (e^\alpha \lambda_i + e^\beta) \quad (23)$$

$$K_{\alpha,\beta}^{-1} = \sum_i \frac{a_i a_i^T}{e^{\alpha \lambda_i} + e^{\beta}} \quad (24)$$

Substitusi persamaan (23) dan (24) ke persamaan (21) menghasilkan:

$$\log \mathcal{L}(\alpha, \beta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \prod_i (e^{\alpha \lambda_i} + e^{\beta}) - \frac{1}{2} \sum_i \frac{a_i a_i^T}{e^{\alpha \lambda_i} + e^{\beta}} \quad (25)$$

Lalu, dilakukan penurunan rumus, diperoleh:

$$\log \mathcal{L}(\alpha, \beta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_i \log(e^{\alpha \lambda_i} + e^{\beta}) - \frac{1}{2} \sum_i \frac{(a_i^T y^*)^2}{e^{\alpha \lambda_i} + e^{\beta}} \quad (26)$$

Estimasi parameter diperoleh dengan menurunkan persamaan (26) terhadap α dan β , kemudian menyamakan dengan nol.

Turunan parsial terhadap α adalah:

$$\frac{\partial}{\partial \alpha} \log \mathcal{L}(\alpha, \beta) = -\frac{1}{2} \sum_i \frac{e^{\alpha \lambda_i}}{e^{\alpha \lambda_i} + e^{\beta}} + \frac{1}{2} \sum_i \frac{e^{\alpha \lambda_i}}{(e^{\alpha \lambda_i} + e^{\beta})^2} (a_i^T y^*)^2 \quad (27)$$

Turunan parsial terhadap β adalah:

$$\frac{\partial}{\partial \beta} \log \mathcal{L}(\alpha, \beta) = -\frac{1}{2} \sum_i \frac{e^{\beta}}{e^{\alpha \lambda_i} + e^{\beta}} + \frac{1}{2} \sum_i \frac{e^{\beta}}{(e^{\alpha \lambda_i} + e^{\beta})^2} (a_i^T y^*)^2 \quad (28)$$

Dengan menyamakan persamaan (27) dan (28) dengan nol, diperoleh sistem persamaan untuk estimasi parameter:

$$\sum_i \frac{e^{\alpha \lambda_i}}{e^{\alpha \lambda_i} + e^{\beta}} = \sum_i \frac{e^{\alpha \lambda_i}}{(e^{\alpha \lambda_i} + e^{\beta})^2} (a_i^T y^*)^2 \quad (29)$$

$$\sum_i \frac{e^{\beta}}{e^{\alpha \lambda_i} + e^{\beta}} = \sum_i \frac{e^{\beta}}{(e^{\alpha \lambda_i} + e^{\beta})^2} (a_i^T y^*)^2 \quad (30)$$

Persamaan (29) dan (30) bersifat nonlinear sehingga diselesaikan secara numerik menggunakan algoritma optimasi pada perangkat lunak Matlab. Proses estimasi dilakukan menggunakan 108 data *training*. Hasil estimasi parameter disajikan pada Tabel 3.

Tabel 3. Hasil Estimasi Parameter Model GPR

Parameter	Simbol	Nilai Estimasi
Skala <i>signal</i>	α	7,906
Skala <i>noise</i>	β	2,318

Nilai $\alpha = 7,906$ menunjukkan bahwa variansi *signal* (fungsi *latent*) cukup besar, mencerminkan variabilitas data curah hujan yang dipengaruhi oleh banyak faktor. Nilai $\beta = 2,318$ yang lebih kecil menunjukkan bahwa komponen *noise* dalam data relatif terkontrol.

3.4. Hasil Prediksi dan Evaluasi Model

Dengan menggunakan parameter yang telah diestimasi, model prediksi dikonstruksi untuk variabel curah hujan. Model untuk setiap observasi mengikuti persamaan (13), dengan y_i adalah curah hujan, x_{1i} adalah kelembapan udara, dan x_{2i} adalah kecepatan angin. Distribusi fungsi *latent* dan *noise* masing-masing adalah

$$f(x_i) \sim \mathcal{N}(0, e^{7,906}H) \quad (31)$$

$$\varepsilon_i \sim \mathcal{N}(0, e^{2,318}) \quad (32)$$

dengan $x_i = [x_{1i}, x_{2i}]^T$ adalah vektor input multivariat. Matriks kovariansi H memperhitungkan korelasi antar observasi berdasarkan jarak Euclidean dalam ruang input 2-dimensi (kelembapan udara dan kecepatan angin). Model ini secara natural menangkap interaksi nonlinear antara kedua prediktor melalui struktur kovariansi, tanpa perlu mengasumsikan aditivitas atau independensi antar prediktor.

Data *testing* terdiri dari 12 observasi bulanan (Januari–Desember 2024). Pemilihan tahun 2024 sebagai periode *testing* memungkinkan evaluasi kemampuan model dalam memprediksi kondisi terkini berdasarkan pola historis 2015–2023. Hasil prediksi disajikan pada Tabel 4.

Tabel 4. Hasil Prediksi Curah Hujan Tahun 2024

Bulan	Prediksi (mm)	Aktual (mm)	Selisih Absolut (mm)	Error (%)
Januari	10,124	12,245	2,130	17,38
Februari	11,155	10,025	1,130	11,27
Maret	20,148	26,887	6,739	25,06
April	4,168	4,729	0,561	11,86
Mei	0,191	0,028	0,164	585,71*
Juni	0,022	0,868	0,846	97,47
Juli	0,183	0,561	0,378	67,38
Agustus	2,531	1,493	1,038	69,53
September	0,913	1,493	0,580	38,84
Oktober	2,342	2,415	0,073	3,02
November	4,893	4,514	0,380	8,41
Desember	12,315	7,944	4,370	55,01
Rata-rata	5,749	6,101	1,532	-

*Catatan: Error tinggi pada Mei disebabkan nilai aktual mendekati nol (0,028)

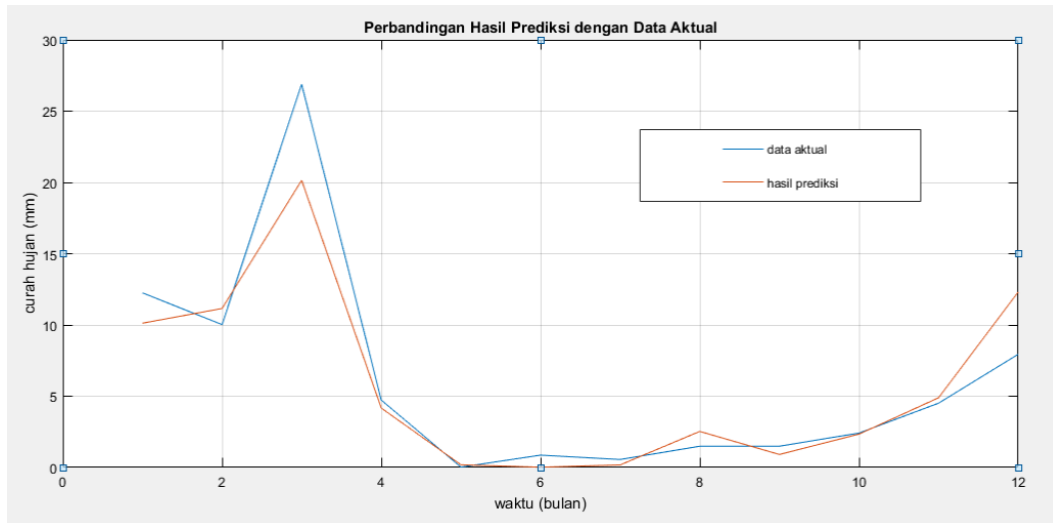
Hasil prediksi menunjukkan bahwa model GPR mampu menangkap pola musiman curah hujan dengan baik. Prediksi curah hujan tertinggi terjadi pada Maret 2024 (20,148 mm), konsisten dengan karakteristik iklim Indonesia di mana Maret merupakan akhir musim hujan. Meskipun terdapat selisih 6,739 mm dengan nilai aktual 26,887 mm, model berhasil mengidentifikasi Maret sebagai bulan dengan curah hujan tinggi. Selisih ini dapat dijelaskan oleh kejadian curah hujan ekstrem yang tidak sepenuhnya tertangkap oleh pola historis.

Prediksi curah hujan terendah terjadi pada bulan Juni 2024 sebesar 0,022 mm, sesuai dengan karakteristik musim kemarau. Nilai aktual 0,868 mm dengan selisih 0,846 mm, menunjukkan akurasi yang baik untuk periode kering. Hasil serupa terlihat pada Mei dan Juli di mana model memprediksi curah hujan minimal dengan selisih relatif kecil.

Performa model pada bulan transisi menunjukkan hasil yang baik. Pada April yang merupakan transisi dari musim hujan ke kemarau, model memprediksi curah hujan sebesar 4,168 mm dibandingkan nilai aktual 4,729 mm dengan selisih hanya 0,561 mm. Pada Oktober yang merupakan transisi dari musim kemarau ke musim hujan, model menghasilkan prediksi 2,342 mm dibandingkan aktual 2,415 mm dengan selisih terkecil yaitu hanya 0,073 mm. Akurasi tinggi pada bulan transisi menunjukkan kemampuan model menangkap sinyal perubahan musim.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Syifaul Janan



Gambar 4. Perbandingan nilai prediksi model GPR dengan data aktual curah hujan tahun 2024

Perbandingan grafis (Gambar 4) menunjukkan model cenderung menghasilkan prediksi sedikit lebih rendah untuk bulan dengan curah hujan tinggi, dan sedikit lebih tinggi untuk bulan dengan curah hujan rendah. Fenomena ini umum terjadi pada model regresi yang cenderung melakukan *regression toward the mean*. Meskipun demikian, pola musiman secara keseluruhan berhasil ditangkap dengan baik.

Tabel 5. Ringkasan Evaluasi Performa Model GPR

Metrik Evaluasi	Nilai	Interpretasi
MAPE	36,84%	Rata-rata kesalahan persentase absolut
Korelasi (r)	0,94	Korelasi prediksi-aktual
Error Minimum	0,073 mm	Oktober 2024
Error Maksimum	6,739 mm	Maret 2024

Dengan MAPE 36,84%, model GPR dengan *prior* Gerak Brown termasuk dalam kategori "*reasonable forecasting*". Performa ini dapat diterima untuk konteks peramalan curah hujan yang memiliki ketidakpastian *inherent* tinggi karena beberapa faktor:

1. Variabilitas Tinggi: Data curah hujan memiliki standar deviasi 5,94 mm dengan rentang 0–26,89 mm, mencerminkan variabilitas yang sangat tinggi.
2. Nilai Mendekati Nol: Beberapa bulan memiliki curah hujan sangat rendah (< 1 mm), di mana *error* absolut yang kecil menghasilkan rasio persentase *error* yang besar. Hal ini berkontribusi pada peningkatan MAPE keseluruhan.
3. Kejadian Ekstrem: Beberapa bulan menunjukkan kejadian curah hujan yang menyimpang dari pola historis, seperti Maret dan Desember, yang menyebabkan *error* prediksi lebih besar.

Korelasi 0,94 antara nilai prediksi dan aktual mengindikasikan bahwa model mampu menjelaskan sebagian besar variasi dalam data curah hujan. Hasil penelitian menunjukkan bahwa metode GPR dengan *prior* Gerak Brown efektif untuk peramalan curah hujan bulanan di Provinsi DIY. Model mampu menangkap pola musiman dengan baik dan menghasilkan prediksi dengan tingkat akurasi memadai. Implikasi praktis dari hasil ini adalah model dapat digunakan sebagai alat bantu perencanaan pertanian, khususnya untuk menentukan waktu dan pola tanam berdasarkan prediksi curah hujan.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan beberapa hal sebagai berikut:

Pertama, model *posterior* GPR berhasil dikembangkan menggunakan pendekatan Bayesian yang mengkombinasikan *prior* Gerak Brown dengan *likelihood* Gaussian. Model *posterior* berbentuk distribusi normal multivariat $p(f|X, Y) = \mathcal{N}(\tilde{f} = \frac{1}{e\beta} A^{-1}Y, A^{-1})$ dengan $A = e^{-\beta}I + e^{-\alpha}H^{-1}$. Model ini dapat mengukur tingkat kepastian dari setiap prediksi yang dihasilkan. Kedua, estimasi parameter model GPR menggunakan metode *Maximum Marginal Likelihood* menghasilkan nilai $\alpha = 7,906$ dan $\beta = 2,318$. Ketiga, model GPR dengan *prior* Gerak Brown mampu memprediksi curah hujan di Provinsi DIY dengan performa yang baik. Evaluasi model menggunakan *Mean Absolute Percentage Error* (MAPE) menghasilkan nilai 36,84% yang termasuk kategori "reasonable forecasting". Model berhasil menangkap pola musiman dengan prediksi tertinggi pada Maret 2024 (20,148 mm) dan terendah pada Juni 2024 (0,022 mm), konsisten dengan pola iklim tropis Indonesia. Korelasi tinggi antara prediksi dan data aktual ($r = 0,94$) menunjukkan kemampuan model dalam menjelaskan variabilitas curah hujan.

Penelitian ini menunjukkan bahwa metode GPR dengan *prior* Gerak Brown merupakan alternatif efektif untuk peramalan curah hujan. Untuk penelitian selanjutnya, disarankan untuk: (1) mengeksplorasi fungsi kovariansi lain seperti *Squared Exponential* atau Matérn; (2) menambahkan variabel prediktor seperti suhu, tekanan udara, dan indeks iklim regional (ENSO, IOD); (3) mengembangkan model spasial dengan data dari multiple *stations*; serta (4) mengintegrasikan GPR dengan metode *deep learning* untuk meningkatkan akurasi pada kejadian curah hujan ekstrem.

DAFTAR PUSTAKA

- [1] Banerjee, A., Dunson, D. B., & Tokdar, S. T., 2013. Efficient Gaussian Process Regression for Large Datasets. *Biometrika*, 100(1), 75–89.
- [2] Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A., 2022. Rainfall Prediction: A Comparative Analysis of Modern Machine Learning Algorithms for Time-Series Forecasting. *Machine Learning with Applications*, 7, 100204.
- [3] Bochenek, B., & Ustrnul, Z., 2022. Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. *Atmosphere*, 13(2), 180.
- [4] BMKG (Badan Meteorologi, Klimatologi, dan Geofisika)., 2025. Data Iklim Bulanan Stasiun Geofisika Yogyakarta Periode 2015-2024 [Dataset]. Diakses dari <http://dataonline.bmkg.go.id/>
- [5] BPS Provinsi D.I. Yogyakarta., 2023. Luas Panen dan Produksi Padi di D.I. Yogyakarta 2022. <https://yogyakarta.bps.go.id/pressrelease/2023/04/03/1347/>
- [6] Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z., 2023. Machine Learning Methods in Weather and Climate Applications: A Survey. *Applied Sciences*, 13(21), 12019.
- [7] Das, P., Sachindra, D. A., & Chanda, K., 2022. Machine Learning-Based Rainfall Forecasting with Multiple Non-Linear Feature Selection Algorithms. *Water Resources Management*, 36(15), 6043–6071.
- [8] Ehteram, M., Seifi, A., & Banadkooki, F. B., 2023. *Application of Machine Learning Models in Agricultural and Meteorological Sciences*. Springer Nature.
- [9] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O., 2021. A Survey on Missing Data in Machine Learning. *Journal of Big Data*, 8(1), 140.
- [10] Farah, A. A., Mohamed, M. A., Musse, O. S. H., & Nor, B. A., 2025. The Multifaceted Impact of Climate Change on Agricultural Productivity: A Systematic Literature Review of SCOPUS-Indexed Studies (2015–2024). *Discover Sustainability*, 6(1), 397.

- [11] Li, C., Ren, X., & Zhao, G., 2023. Machine-Learning-Based Imputation Method for Filling Missing Values in Ground Meteorological Observation Data. *Algorithms*, 16(9), 422.
- [12] Mulla, S., Ahmed, R., Singh, K. K., Singh, S. K., Deshmukh, N., & Inamdar, F. K., 2023. Climate Change Effect on Climate Parameters like Temperature, Rainfall and Water Resources Sectors in India. *In Climate Change Impacts in India* (pp. 9–59). Springer.
- [13] Pham, B. T., Le, L. M., Le, T.-T., Bui, K.-T. T., Le, V. M., Ly, H.-B., & Prakash, I., 2020. Development of Advanced Artificial Intelligence Models for Daily Rainfall Prediction. *Atmospheric Research*, 237, 104845.
- [14] Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Torné, R. V., Sala, E., & Lió, P., 2023. The Impact of Imputation Quality on Machine Learning Classifiers for Datasets with Missing Values. *Communications Medicine*, 3(1), 139.
- [15] Shi, J. Q., & Choi, T., 2011. Gaussian Process Regression Analysis for Functional Data. *CRC Press*.
- [16] Sofia, D. A., Sujono, J., & Legono, D., 2018. Analisis Variabilitas Spasial dan Temporal Curah Hujan di Wilayah Gunung Merapi. *Teknisia*, 430–438.
- [17] Williams, C. K. I., & Rasmussen, C. E., 2006. Gaussian Processes for Machine Learning (Vol. 2, Issue 3). *MIT press Cambridge, MA*.