

Prediksi Curah Hujan Bulanan Di Wilayah Lombok Barat Menggunakan Principal Component Regression (PCR)

Anggitya Pratiwi^{1,a}, Rian Mahendra Taruna^{2,b}, Suci Agustiarini^{1,c}, Dewo Sulistio Adi Wibowo^{1,d}

⁽¹⁾ Stasiun Klimatologi Lombok Barat, Badan Meteorologi Klimatologi dan Geofisika

⁽²⁾ Stasiun Geofisika Mataram, Badan Meteorologi Klimatologi dan Geofisika

Email : ^(a)anggityap9@gmail.com, ^(b)reemyan@gmail.com, ^(c)suci.agustiarini@gmail.com,

^(d)dewosulistio23@gmail.com

Diterima (23 Juni 2020), Direvisi (07 Juli 2020)

Abstract. Rainfall is a significant climate parameter in various sectoral of human activities. Despite that, precise rainfall prediction is necessary. One factor which can affect the rainfall is Sea Surface Temperature. We used Sea Surface Temperature data in Indonesia limited by 14° LS - 10° LU and 90° BT - 142° BT to predict monthly rainfall in West Lombok. The method used in this study is Principal Component Regression. Regression is calculated using 1998-2013 data, with Principal Component Analysis result of Indonesian Sea Surface Temperature as the predictor. Validation using 2014-2018 rainfall data in West Lombok showed that rainfall prediction is appropriate with the rainfall pattern observation data with RMSE about 81,92. While the validation based on rainfall characteristics using contingency table showed that 42% of all rainfall characteristics is correctly predicted..

Keywords: Prediction, Rainfall, Principal Component Regression.

Abstrak. Curah hujan adalah parameter iklim yang signifikan pada berbagai macam sektor aktivitas manusia. Oleh sebab itu, prediksi curah hujan yang akurat sangat diperlukan. Salah satu faktor yang dapat mempengaruhi curah hujan adalah suhu permukaan laut. Pada penelitian ini digunakan data suhu permukaan laut Indonesia yang dibatasi oleh 14° LS - 10° LU and 90° BT - 142° BT untuk memprediksi curah hujan bulanan di Lombok Barat. Metode yang digunakan pada penelitian ini yaitu regresi komponenn utama. Regresi dihitung menggunakan data 1998-2013, dengan hasil analisis komponen utama dari suhu permukaan laut di Indonesia sebagai prediktor. Hasil validasi curah hujan dengan menggunakan periode 2014-2018 di Lombok Barat menunjukkan bahwa prediksi curah hujan yang dihasilkan cukup mendekati pola curah hujan data observasinya dengan RMSE sebesar 81,92. Sedangkan untuk validasi berdasarkan sifat hujan menggunakan tabel kontingensi menunjukkan bahwa jumlah seluruh sifat hujan yang benar diprediksi yaitu sebesar 42%..

Kata kunci: Prediksi, Curah Hujan, Principal Component Regression.

PENDAHULUAN

Lombok merupakan salah satu pulau yang berada di Provinsi Nusa Tenggara Barat. Wilayah ini berada di sebelah selatan garis khatulistiwa. Karena letak geografisnya tersebut, wilayah ini memiliki sifat hujan monsunial yang ditandai dengan

adanya satu puncak musim kemarau dan satu puncak musim hujan.

Berbagai sektor kehidupan masyarakat khususnya di wilayah Lombok Barat, dimulai dari sektor pertanian, perkebunan, hingga pariwisata sangat bergantung pada informasi curah hujan. Namun, curah hujan mempunyai variabilitas yang besar dalam

ruang dan waktu [1]. Oleh sebab itu, berbagai faktor dapat mempengaruhi jumlah curah hujan yang terjadi pada suatu tempat. Suhu permukaan laut di perairan Indonesia berpengaruh terhadap pembentukan awan di atmosfer. Ketika suhu permukaan laut dingin, maka suplai uap air di atmosfer umumnya akan berkurang. Sebaliknya, jika suhu permukaan laut memanas, maka ketersediaan uap air di atmosfer akan semakin besar dan kondisi atmosfer akan menjadi jenuh. Awan-awan yang terbentuk dengan kandungan uap air yang melimpah di wilayah ini, mampu tumbuh secara vertikal secara optimal. Semakin optimal pertumbuhan awan hujan, maka potensi terjadi hujan dari awan-awan tersebut semakin besar [2].

Pada penelitian ini dilakukan prediksi curah hujan bulanan menggunakan metode *Principal Component Regression* (PCR). Hasil prediksi ini kemudian divalidasi agar dapat diketahui tingkat akurasinya.

METODE PENELITIAN

Dalam penelitian ini digunakan data curah hujan bulanan dari Stasiun Klimatologi Lombok Barat periode 1998 – 2018. Kemudian data suhu permukaan laut bulanan dengan periode yang sama seperti data curah hujan. Data suhu permukaan laut ini berupa *Extended Reconstructed Sea Surface Temperature* (ERSST) versi 3b dengan resolusi $2^\circ \times 2^\circ$ yang bersumber dari NOAA dan diunduh dari IRI Data Library (<http://iridl.ldeo.columbia.edu/>). Domain data yang digunakan yaitu $14^\circ \text{ LS} - 10^\circ \text{ LU}$ dan $90^\circ \text{ BT} - 142^\circ \text{ BT}$ yang mewakili suhu permukaan laut di perairan Indonesia. Data suhu permukaan laut tersebut kemudian digunakan sebagai prediktor yang diolah terlebih dahulu menggunakan metode *Principal Component Analysis* (PCA). Metode PCA bertujuan untuk mereduksi dimensi data yang

berukuran besar dengan mempertahankan sebanyak mungkin variasi dari himpunan data asal [3].

Perhitungan pada metode ini dimulai dengan memisalkan data klimatologi dalam bentuk kumpulan data grid ruang dan waktu (*space-time field*) $X(t,s)$ yang merepresentasikan nilai dari suatu variabel X [4]. Nilai dari suatu variabel dalam waktu (t_i) dan titik grid (s_j) dapat dinotasikan sebagai x_{ij} untuk $i = 1, \dots, n$ dan $j = 1, \dots, p$. Data observasi kemudian direpresentasikan dengan data matriks:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad (1)$$

Rata-rata domain waktu setiap grid dilambangkan dengan $\bar{x}_j$ menggunakan persamaan sebagai berikut:

$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj} \quad (2)$$

dengan

$\bar{x}$ = rata-rata klimatologis data pada grid j
 n = jumlah time series

Matriks data anomali kemudian ditentukan dengan operasi matriks berikut:

$$X' = X - 1\bar{x} = \left(I - \frac{1}{n}11^T\right)X \quad (3)$$

dengan

X' = data anomali

X = nilai observasi

$\bar{x}$ = rata-rata klimatologis data pada grid j
 n = jumlah time series

Tahap selanjutnya yaitu menghitung matriks kovarian Σ yang ditentukan dengan persamaan:

$$\Sigma = \frac{1}{n-1} X'^T X' \quad (4)$$

Selanjutnya menggunakan *Eigen Value Problem* (EVP) untuk mendapatkan *eigenvalue* dan *eigenvector* dengan menggunakan persamaan:

$$\Sigma a = \lambda a \quad (5)$$

dengan

$a = \text{eigenvector}$ dari matriks Σ

$\lambda = \text{eigenvalue}$

Dari **Persamaan 4** diperoleh nilai eigen λ dan vektor eigen e yang memenuhi persamaan $|\Sigma - \lambda I| = 0$ dan $(\Sigma - \lambda I)a = 0$. Vektor eigen $a_k(x, y)$ merupakan variabilitas temporal PCA yang ortogonal, sehingga memenuhi persamaan:

$$\sum_{x,y=1}^N a_m(x_i, y_i) a_l(x_i, y_i) = 0, k \neq l \quad (6)$$

PCA mode ke k ($k = 1, 2, 3, \dots, p$) adalah *eigenvector* ke k (a_k) dari Σ . *Eigenvalue* λ_k umumnya digunakan untuk merepresentasikan persentase varians yang dapat dijelaskan oleh PCA mode ke k dengan persamaan sebagai berikut:

$$\frac{100\lambda_k}{\sum_{k=1}^p \lambda_k} \% \quad (7)$$

Proyeksi dari data anomali X' pada PCA mode $a_k(c_k = X'_{ak})$ disebut sebagai Komponen Utama atau *Principal Component* (PC). Korelasi digunakan untuk melihat hubungan antara dua variabel. Besarnya nilai korelasi berkisar antara -1 hingga +1 [5]. Nilai positif menunjukkan bahwa kedua variabel memiliki hubungan yang searah, yaitu jika nilai variabel X meningkat maka nilai variabel Y juga akan meningkat, demikian juga sebaliknya. Nilai negatif menunjukkan bahwa kedua variabel memiliki hubungan yang berlawanan arah, yaitu jika nilai variabel X meningkat maka nilai variabel Y mengalami penurunan dan begitu juga sebaliknya. Besar nilai korelasi yang mendekati angka 1 berarti bahwa hubungan antar kedua variabel semakin besar dan sebaliknya semakin mendekati angka 0 maka korelasinya semakin kecil.

Persamaan umum regresi linear berganda adalah [6]:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (8)$$

dengan

Y = variabel terikat

a, b_1, b_2, b_n = konstanta

X_1, X_2, X_n = nilai variabel bebas

Tabel kontingensi dapat digunakan sebagai metode untuk memvalidasi prakiraan yang bersifat deterministik [4]. Tabel ini dapat dilihat pada **Tabel 1**.

- Hit Score (HS)** : jumlah dari kategori yang benar diprakirakan.

$$HS = \frac{A_{11} + A_{22} + A_{33}}{T} \times 100\%$$
- Post Agreement (PA)** : jumlah prakiraan yang benar diprakirakan dibagi dengan jumlah prakiraan untuk setiap kategori.
 - Untuk kategori sifat hujan di atas normal = $\frac{A_{11}}{M}$
 - Untuk kategori sifat hujan normal = $\frac{A_{22}}{N}$
 - Untuk kategori sifat hujan di bawah normal = $\frac{A_{33}}{O}$
- False Alarm Ratio (FAR)** : fraksi peristiwa prakiraan yang gagal terjadi.
 - Untuk kategori sifat hujan di atas normal = $\frac{(A_{21} + A_{31})}{M}$
 - Untuk kategori sifat hujan normal = $\frac{(A_{12} + A_{32})}{N}$
 - Untuk kategori sifat hujan di bawah normal = $\frac{(A_{13} + A_{23})}{O}$
- Probability of Detection (POD)**: jumlah prakiraan yang benar dibagi dengan jumlah data observasi dalam setiap kategori. POD merupakan ukuran dari kemampuan untuk memprakirakan setiap kategori dengan benar. POD untuk tiga kategori yang berbeda:

Tabel 1 Tabel Kontingensi

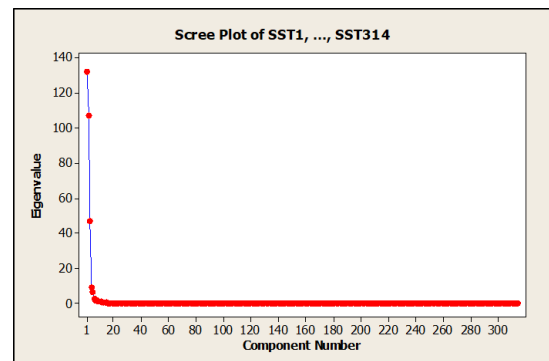
ij	Obs			Jumlah	
	AN	N	BN		
Prediksi	AN	A_{11}	A_{12}	A_{13}	J
	N	A_{21}	A_{22}	A_{23}	K
	BN	A_{31}	A_{32}	A_{33}	L
Jumlah	M	N	O	T	

HASIL DAN PEMBAHASAN

- Untuk kategori sifat hujan di atas normal = $\frac{A_{11}}{J}$
 - Untuk kategori sifat hujan normal = $\frac{A_{22}}{K}$
 - Untuk kategori sifat hujan di bawah normal = $\frac{A_{33}}{L}$
- e. Bias adalah perbandingan rata-rata prediksi dengan rata-rata observasi. $BIAS > 1$: *over forecasting*; $BIAS < 1$: *under forecasting*
- Untuk kategori sifat hujan di atas normal = $\frac{M}{J}$
 - Untuk kategori sifat hujan normal = $\frac{N}{K}$
 - Untuk kategori sifat hujan di bawah normal = $\frac{O}{L}$

Terdapat beberapa proses pengolahan data dalam penelitian ini, yaitu:

1. Menyusun data suhu permukaan laut Indonesia dan curah hujan Lombok Barat.
2. Menghitung korelasi antara suhu permukaan laut Indonesia pada masing-masing grid dengan curah hujan bulanan di Lombok Barat.
3. Melakukan analisis PCA terhadap suhu permukaan laut Indonesia dengan bantuan perangkat pengolah statistik.
4. Menghitung korelasi antara komponen utama suhu permukaan laut Indonesia pada masing-masing grid dengan curah hujan bulanan di Lombok Barat.
5. Melakukan analisis regresi linier antara komponen utama suhu permukaan laut Indonesia pada masing-masing grid dengan curah hujan bulanan di Lombok Barat periode 1998-2013.
6. Melakukan simulasi prediksi curah hujan bulanan periode 2014-2018.
7. Melakukan validasi hasil prediksi menggunakan tabel kontingensi.



Gambar 1. Scree Plot Suhu Permukaan Laut Indonesia

Berdasarkan **Gambar 1**, analisis komponen utama terhadap suhu permukaan laut di perairan Indonesia pada wilayah 14° LS – 10° LU dan 90° BT dan 142° BT menghasilkan 3 komponen utama. Berdasarkan nilai cumulative pada matriks 314 grid yang digunakan, ketiga komponen utama tersebut mampu menjelaskan sebanyak 91% variasi dari data asli.

Untuk mengetahui bahwa sudah tidak terdapat korelasi yang kuat diantara masing-masing komponen utama, dilakukan analisis korelasi yang hasilnya ditunjukkan pada **Tabel 2**. Berdasarkan tabel tersebut dapat dilihat bahwa sudah tidak terdapat korelasi yang kuat diantara ketiga komponen utama suhu permukaan laut yang merupakan prediktor dalam memprediksi curah hujan di Stasiun Klimatologi Lombok Barat. Sehingga ketiga komponen tersebut dapat digunakan sebagai prediktor dalam analisis regresi linear berganda.

Sebelum dilakukan prediksi, dilakukan analisis korelasi setiap komponen utama suhu permukaan laut terhadap curah hujan di Stasiun Klimatologi Lombok Barat. Hasil korelasi tersebut ditunjukkan pada **Tabel 3**. Berdasarkan tabel tersebut dapat diketahui bahwa korelasi terkuat terdapat pada PC1 (komponen utama ke-1).

Tabel 2. Korelasi Antar Komponen Utama Suhu Permukaan Laut

Korelasi	<i>pc1</i>	<i>pc2</i>	<i>pc3</i>
<i>pc1</i>	1		
<i>pc2</i>	-0.005	1	
<i>pc3</i>	-0.002	-0.003	1

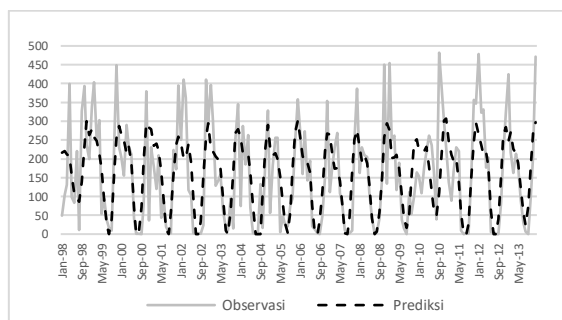
Tabel 3. Korelasi Curah hujan dengan Komponen Utama Suhu Permukaan Laut

Korelasi	PC1	PC2	PC3
CH	0.62	-0.04	-0.38

Berdasarkan data ketiga komponen utama suhu permukaan laut yang diperoleh, dilakukan analisis regresi linear berganda terhadap curah hujan periode 1998-2013 untuk digunakan sebagai model persamaan dalam memprediksi curah hujan. Analisis regresi linear berganda ini dilakukan untuk mencari konstanta a , b_1 , b_2 , dan b_3 . Keempat konstanta tersebut diolah menggunakan rumus regresi pada perangkat pengolah angka. Setelah keempat konstanta tersebut disusun, maka diperoleh model persamaan prediksi sebagai berikut:

$$Y = 161.75 + 7.34 PC1 - 0.58 PC2 - 7.19 PC3 \quad (9)$$

Sebelum dilakukan prediksi curah hujan periode 2014-2018, **Persamaan 9** dapat digunakan juga untuk memprediksi curah hujan periode 1998-2013. Hasil perbandingan antara prediksi curah hujan periode 1998-2013 dan nilai observasinya ditunjukkan pada Gambar 2.

**Gambar 2.** Grafik Curah Hujan Observasi dan Prediksi Tahun 1998-2013

Berdasarkan **Gambar 2.** dapat dilihat bahwa hasil prediksi curah hujan periode 1998-2013 secara umum sudah memiliki pola yang sama dengan data observasinya. Hanya saja, nilai prediksi yang diperoleh tidak begitu mirip dengan RMSE sebesar 89,83. Hal ini dapat dipengaruhi oleh kondisi dinamika atmosfer yang terjadi pada tahun-tahun tersebut.

Pada **Gambar 3** dapat dilihat perbandingan antara data curah hujan hasil prediksi dengan data observasinya. Pada tahun 2014, diketahui bahwa selisih terbesar yang dihasilkan terjadi pada bulan Mei, dimana pada bulan tersebut menunjukkan bahwa pada data observasi terjadi penurunan curah hujan yang cukup signifikan, namun data hasil prediksi belum mampu memprediksinya.

Pada tahun 2015, dapat dilihat bahwa selisih terbesar terjadi pada bulan Maret, dimana data observasi menunjukkan adanya peningkatan curah hujan, namun hasil prediksi menunjukkan data yang berkebalikan. Dapat dilihat juga bahwa hasil prediksi cukup baik dalam memprediksi pada bulan-bulan yang cenderung tidak ada hujan.

Selanjutnya pada tahun 2016, secara keseluruhan data hasil prediksi menunjukkan nilai yang cukup mirip dengan data observasinya, hanya saja pada bulan November menunjukkan selisih yang cukup besar.

Pada tahun 2017, selisih terbesar antara data prediksi dan observasinya ditunjukkan pada bulan Januari, Maret, dan November. Sedangkan pada tahun 2018, selisih terbesar ditunjukkan pada bulan Januari, Maret, April, November, dan Desember.

Selain dilihat perbandingan data curah hujan observasi dan data prediksinya, dilakukan juga perhitungan validasi antara sifat hujan data observasi dan prediksinya.

Tabel 4. Tabel Kontingensi Sifat Hujan Hasil Prediksi dan Observasi

Sifat		Obs			Jumlah
		AN	N	BN	
Prediksi	AN	10	22	8	40
	N	0	8	2	10
	BN	1	2	7	10
Jumlah		11	32	17	60

Tabel 5. Perhitungan Berdasarkan Tabel Kontingensi

Hit Score	0.42	
Post Agreement	AN	0.25
	N	0.80
	BN	0.70
FAR	AN	0.75
	N	0.20
	BN	0.30
POD	AN	0.90
	N	0.25
	BN	0.41
BIAS	AN	3.63
	N	0.31
	BN	0.58

Berdasarkan **Tabel 4** diperoleh hasil perhitungan-perhitungan validasi seperti yang tercantum pada **Tabel 5**. Pada **Tabel 5** menunjukkan nilai *hit score* sebesar 42%. Hal ini menunjukkan bahwa sebanyak 42% hasil prediksi sifat hujan yang sesuai dengan data observasinya. Nilai *post agreement* dapat dilihat bahwa untuk kategori AN yang sesuai diprediksi sebesar 25% terhadap seluruh hasil observasi, untuk kategori N sebanyak 80% yang sesuai diprediksi terhadap seluruh hasil observasi, dan untuk kategori BN sebanyak 70% yang sesuai diprediksi terhadap seluruh hasil observasi.

Nilai FAR untuk masing-masing sifat hujan menunjukkan bahwa untuk kategori AN sebesar 75% yang tidak

sesuai diprediksi terhadap seluruh data observasi, untuk kategori N sebesar 20% yang tidak sesuai diprediksi terhadap seluruh data observasinya, dan untuk kategori BN sebesar 30% yang tidak sesuai diprediksi terhadap seluruh data observasinya.

Nilai POD untuk masing-masing sifat hujan menunjukkan bahwa untuk kategori AN sebesar 90% data observasi yang sesuai dengan prediksinya, untuk kategori N sebesar 25% data observasi yang sesuai dengan data prediksinya, dan untuk kategori BN sebesar 41% data observasi yang sesuai dengan data prediksinya.

Berdasarkan **Tabel 5**, nilai bias untuk kategori N dan BN memiliki nilai kurang dari 1, hal tersebut menunjukkan bahwa hasil prediksi untuk kategori N dan BN cenderung *underforecast*. Sedangkan untuk kategori AN memiliki bias dengan nilai lebih dari 1, hal tersebut menunjukkan bahwa hasil prediksi untuk kategori ini cenderung *overforecast*.

KESIMPULAN

Pemanfaatan metode analisis PCA terhadap suhu permukaan laut Indonesia menghasilkan 3 komponen utama yang dapat menjelaskan 91% variansnya. Berdasarkan validasi yang telah dilakukan, ketiga komponen utama tersebut mampu menghasilkan model regresi linear berganda yang menghasilkan simulasi prediksi curah hujan bulanan di Lombok Barat yang cukup bagus dengan nilai RMSE sebesar 81,92. Namun, jika dilihat berdasarkan 5 kategori yang dihitung berdasarkan tabel kontingensi menunjukkan bahwa jumlah seluruh sifat hujan yang dapat diprediksi dengan benar sebesar 42%, dengan rincian sifat hujan atas normal *overforecast* sedangkan sifat hujan normal dan bawah normal *underforecast*.

Untuk penelitian selanjutnya, dapat dilakukan analisis metode PCR yang berbeda pada setiap bulan, sehingga setiap bulan memiliki masing-masing model prediksi yang dapat menghasilkan hasil prediksi yang lebih akurat.

DAFTAR PUSTAKA

- [1] R. Prasetya, A. As'ari, and W. Dayantolis, "Analisis Dampak Siklon Tropis Nangka, Parma dan Nida pada Distribusi Curah Hujan di Sulawesi Utara," *J. Fis. dan Apl.*, vol. 10, no. 1, p. 1, 2014, doi: 10.12962/j24604682.v10i1.816.
- [2] D. Sucahyono and K. Ribudiyanto, *Cuaca dan Iklim Ekstrim di Indonesia*. Jakarta: Puslitbang BMKG, 2013.
- [3] I. L. LESTARI, S. NURDIATI, and A. SOPAHEL UWAKAN, "Analisis Empirical Orthogonal Function (Eof) Berbasis Singular Value Decomposition (Svd) Pada Data Curah Hujan Indonesia," *J. Math. Its Appl.*, vol. 15, no. 1, p. 13, 2016, doi: 10.29244/jmap.15.1.13-22.
- [4] M. Rohmah and A. Nisa, "Prediksi Curah Hujan Bulanan di Pondok Betung Menggunakan Prediktor Komponen Utama Suhu Permukaan Laut Perairan Indonesia Monthly Rainfall Prediction in Pondok Betung Utilize Indonesian Sea Surface Temperature," in *Seminar Nasional Penginderaan Jauh ke-6*, 2019, pp. 255–263.
- [5] A. G. Bluman, *Elementary Statistics*. Higher Education, 2009.
- [6] Y. Wibisono, *Metode Statistik*. Yogyakarta: Gadjah Mada University Press, 2005.

