

# PEMODELAN SEMIPARAMETRIK *STATISTICAL DOWNSCALING* UNTUK MENDUGA CURAH HUJAN BULANAN DI INDRAMAYU

Akbar Rizki\*, Abdul Aziz Nurussadad\*\*

\*Departemen Statistika Institut Pertanian Bogor

\*\*Badan Informasi Geospasial (BIG)

*Abstrak—Semiparametric statistical downscaling (SD) model is a statistical model which consists of parametric and non-parametric functional relationship between local scale and global scale variable. This study used rainfall intensity in Indramayu as local scale variable and Global Precipitation Climatology Project (GPCP) precipitation as global scale variable. GPCP precipitation data have multicollinearity, therefore they were reduced by principal component analysis. Eight principal components which have been selected then used as the predictors and rainfall intensity in Indramayu as the response. Semiparametric SD model was used to predict the rainfall intensity in the district of Indramayu. The semiparametric model developed by mixed model approach where the non-parametric relationship is represented using spline with truncated power basis. Linier semiparametric model is the best model to estimate monthly rainfall in indramayu district. The model performance evaluated by RMSEP (root mean square error prediction) and  $R^2$  (coefficient of determination). The result shows that the best model have values of RMSEP and  $R^2$  are 61.64 and 71%.*

*Kata kunci—Statistical Downscaling; Semiparametrik; GPCP*

## I. PENDAHULUAN

Data *Global Precipitation Climatology Project* (GPCP) dapat digunakan sebagai alat untuk melakukan pendugaan curah hujan secara numerik. Sejumlah data GPCP dihasilkan dari kombinasi data observasi dan satelit dalam bentuk *grid* atau petak wilayah (Soleh 2015). Informasi yang diberikan oleh data GPCP masih bersifat global, sehingga memerlukan suatu teknik untuk mendapatkan informasi berskala lokal. Salah satu teknik yang dapat digunakan untuk mendapatkan informasi berskala lokal dari data GPCP adalah

*Statistical Downscaling (SD).*

Permasalahan utama yang muncul dalam membangun model *SD* adalah mendapatkan metode yang dapat menggambarkan dengan baik hubungan antara peubah penjelas dengan peubah respon (Sutikno 2008). Banyak model *SD* yang telah berkembang, dari model berbasis parametrik sampai dengan model berbasis nonparametrik. Beberapa model berbasis parametrik pada *SD* yang pernah digunakan yaitu regresi komponen utama yang telah digunakan oleh Huth and Kysely (2000), Lanza et al. (2001), Heerdegen et al. (2001), Uvo et al. (2001) dan Bergant et al. (2015) serta regresi kuadrat terkecil parsial yang telah digunakan oleh Wigena (Wigena). Model *SD* nonparametrik yang pernah digunakan diantaranya adalah jaringan syaraf tiruan oleh Sailor et al. (2000), Sarwoko (2013), dan Kodarsih (2013); regresi spline aditif multivariat digunakan oleh Sutikno (2008); serta *projection pursuit regression* digunakan oleh Wigena (2006).

Metode berbasis model parametrik merupakan metode yang mempertimbangkan sebaran, dengan kata lain metode ini memiliki asumsi yang ketat. Metode berbasis model nonparametrik merupakan metode yang tidak memerlukan asumsi sebaran, sehingga lebih fleksibel. Namun demikian, metode ini mempunyai kesulitan dalam melakukan seleksi model dan penentuan model terbaik. Kelemahan kedua metode tersebut menyebabkan para peneliti terus mengembangkan metode *SD*, sehingga muncullah metode berbasis model semiparametrik.

Metode berbasis model semiparametrik meru-

pakannya gabungan dari metode berbasis model parametrik dan nonparametrik. Ide dasar dari metode ini adalah pada suatu model dimungkinkan terdapat bentuk hubungan fungsional parametrik dan nonparametrik secara bersamaan. Penggunaan metode berbasis model semiparametrik telah dilakukan oleh Djuraidah (2007) untuk menganalisis pencemar udara di kota Surabaya, Mehrotra and Sharma (2007) untuk menduga curah hujan harian, dan Wigena et al. (2015) untuk menduga curah hujan bulanan di Kabupaten Indramayu dengan memanfaatkan data luaran GCM (*Global circulation model*). Pada penelitian ini digunakan model SD semiparametrik dengan menggabungkan dua pendekatan yaitu regresi spline terpenalti dan model linier campuran untuk memprediksi curah hujan bulanan di Kabupaten Indramayu dengan memanfaatkan data GPCP.

## II. METODE PENELITIAN

### A. Data

Data yang digunakan dalam penelitian ini merupakan data curah hujan stasiun di Kabupaten Indramayu pada wilayah ZOM 79 yang meliputi empat pos curah hujan yaitu Krangkeng, Sukadana, Karangendal, dan Gresik sebagai peubah respon dan data presipitasi GPCP sebagai peubah penjelas. Masing-masing data tersebut, merupakan data bulanan dari bulan Januari tahun 1981 sampai dengan bulan Desember tahun 2013. Data presipitasi GPCP yang digunakan adalah data presipitasi Global Precipitation Climatology Project (GPCP) versi 2.2. Domain data presipitasi GPCP yang digunakan berbentuk persegi berukuran 8x8 grid dengan posisi di atas wilayah Kabupaten Indramayu yaitu pada  $101.25^{\circ}BT - 118.75^{\circ}BT$  dan  $1.25^{\circ}LS - 18.75^{\circ}LS$ . Penggunaan domain berukuran 8x8 grid di atas wilayah Kabupaten Indramayu memberikan hasil yang lebih stabil serta tidak sensitif terhadap pencilan (Wigena 2006).

### B. Metode Analisis

Langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

- 1) Mengidentifikasi multikolinearitas pada data presipitasi GPCP menggunakan *variance inflation factors* (VIF).
- 2) Membagi data stasiun menjadi dua kelompok, yaitu data untuk pemodelan dan data untuk validasi. Data pemodelan menggunakan data tahun 1981-2012, sedangkan data validasi menggunakan data tahun 2013.
- 3) Menentukan pola hubungan fungsional antara curah hujan (peubah respon) dengan presipitasi GPCP (peubah penjelas), dengan plot antara curah hujan dengan skor komponen utama GPCP terpilih. Pembuatan plot dilakukan pada berbagai kemungkinan derajat bebas untuk melakukan pengepasan pola. Jumlah derajat bebas optimum ditentukan menggunakan kriteria GCV minimum.
- 4) Pada komponen utama yang mempunyai hubungan fungsional nonparametrik dibangkitkan basis spline, yang meliputi:
  - a) Menentukan jumlah simpul.

Misalnya terdapat model p-spline sebagai berikut:

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{k=1}^K \beta_{j+k} (x - K_k)_+^p \quad (1)$$

dengan  $\beta = (\beta_0, \dots, \beta_p, u_{p1}, \dots, u_p K)$  adalah vektor koefisien regresi spline,  $p \geq 1$  adalah bilangan bulat positif, adalah fungsi pangkat terpotong (FPT), dan  $K_1 < \dots < K_k$  adalah simpul tetap (Djuraidah 2007).

Model pada persamaan (1) memiliki jumlah derajat bebas sebanyak  $p$  dan jumlah titik simpul sebanyak  $k$ . Dengan demikian penentuan jumlah simpul dalam suatu model dapat dirumuskan:

$$q = p - db = p - (d + 1) \quad (2)$$

dengan  $q$  adalah jumlah titik simpul,  $p$  adalah jumlah derajat bebas dari pemulus spline,  $db$  adalah jumlah parameter model, dan  $d$  adalah derajat basis FPT. Penentuan jarak antara titik simpul dilakukan dengan menggunakan rumus:

$$s = \frac{n}{q + 1} \quad (3)$$

dengan  $s$  adalah jarak antara titik simpul,  $n$  adalah jumlah pengamatan dan  $q$  adalah jumlah titik simpul.

- b) Membangkitkan basis FPT. Fungsi pangkat terpotong adalah sebagai berikut:

$$f(x) = \begin{cases} (x - K_k)^p & \text{jika } x \geq 0, \\ 0 & \text{jika } x < 0, \end{cases} \quad (4)$$

Dengan  $x$  adalah peubah bebas,  $K_k$  adalah titik simpul ke- $k$  pada variabel bebas, dan  $p$  adalah pangkat tertinggi pada model *penalized spline*.

- 5) Pendugaan model dengan model linear campuran

- a) Menentukan matriks  $\mathbf{Z}$  dan  $\mathbf{X}$  Matriks desain  $\mathbf{X}$  dan  $\mathbf{Z}$  diperoleh dari persamaan (1). Desain untuk kedua matriks tersebut adalah:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \dots & x_1^p \\ \dots & \dots & \dots & \dots \\ 1 & x_n & \dots & x_n^p \end{bmatrix} \quad (5)$$

$$\mathbf{Z} = \begin{bmatrix} (x_1 - K_1)_+^p & \dots & (x_1 - K_k)_+^p \\ \dots & \dots & \dots \\ (x_n - K_1)_+^p & \dots & (x_n - K_k)_+^p \end{bmatrix} \quad (6)$$

- b) Pendugaan parameter dan komponen ragam
- 6) Melakukan prediksi curah hujan di Kabupaten Indramayu dan memilih model terbaik berdasarkan kriteria AIC, BIC, korelasi ( $r$ ) antara data prediksi dan data aktual,  $R^2$ , dan RMSEP

### III. HASIL DAN PEMBAHASAN

#### A. Eksplorasi Data

##### Identifikasi Multikolinearitas

Salah satu indikasi adanya multikolinieritas adalah

nilai VIF lebih dari 10 (Johnson and Wichern 2007). Hasil perhitungan menunjukkan bahwa nilai VIF pada data presipitasi GPCP berkisar 4.75-72.05. Terdapat 55 grid GPCP yang memiliki nilai VIF lebih dari 10 dan 9 grid GPCP yang memiliki VIF kurang dari 10. Hal ini mengindikasikan adanya multikolinearitas pada data presipitasi GPCP.

Salah satu metode yang dapat digunakan untuk mengatasi masalah ini adalah analisis komponen utama/ AKU (Jolliffe 2002). Hasil analisis menunjukkan jumlah komponen utama (KU) yang memiliki akar ciri lebih dari satu sebanyak delapan KU dengan proporsi keragaman total sebesar 87.6%. Dengan demikian, jumlah KU terpilih adalah delapan KU.

##### Pola Hubungan Curah hujan dengan Komponen utama

Penentuan pola hubungan curah hujan dengan komponen utama dilakukan melalui pemilihan nilai derajat bebas optimum dengan kriteria GCV minimum. Hasilnya menunjukkan bahwa jumlah derajat bebas optimum pada masing-masing komponen utama terpilih yaitu KU1 sampai dengan KU8 secara berturut-turut adalah 5, 6, 4, 8, 5, 13, 4, dan 3.

Plot data curah hujan dengan KU terpilih dengan derajat bebas optimum disajikan pada Lampiran 1. Plot data curah hujan dengan KU1 (Lampiran 1 (a)), plot data curah hujan dengan KU3 (Lampiran 1 (c)), dan plot data curah hujan dengan KU7 (Lampiran 1 (g)) terlihat membentuk pola kuadratik. Pola linier terlihat pada plot data curah hujan dengan KU8 (Lampiran (h)). Namun demikian, plot data curah hujan dengan KU2 (Lampiran 1 (b)), plot data curah hujan dengan KU4 (Lampiran 1 (d)), plot data curah hujan dengan KU5 (Lampiran 1 (e)), dan plot data curah hujan dengan KU6 (Lampiran 1 (f)) tidak membentuk pola parametrik tertentu. Hal ini mengindikasikan terdapat KU yang berhubungan secara parametrik dan nonparametrik dengan curah hujan.

### B. Model Semiparametrik

Jumlah titik simpul pada suatu model bergantung pada jumlah parameter dan derajat basis pangkat terpotong model tersebut. Basis pangkat terpotong yang digunakan pada P-spline adalah basis pangkat terpotong berderajat 1 (linier), 2 (kuadratik), dan 3 (kubik). Perbandingan model linier, kuadratik, dan kubik disajikan pada Tabel 1. Nilai AIC paling kecil dimiliki oleh model Kubik. Namun demikian nilai  $r$  dan  $R^2$  paling tinggi serta nilai RMSEP paling rendah dimiliki oleh model linier. Oleh karena itu model terbaik untuk memprediksi data curah hujan bulanan di Kabupaten Indramayu adalah model semiparametrik berderajat satu (linier).

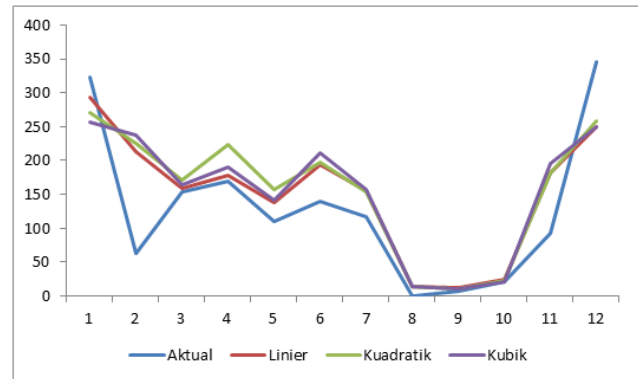
Tabel I  
NILAI AIC,  $r$ ,  $R^2$ , DAN RMSEP MODEL SD SEMIPARAMETRIK  
LINIER, KUADRATIK DAN KUBIK

Model	AIC	$r$	$R^2$	RMSEP
Linier	4248.5	0.84	0.71	61.64
Kuadratik	4222.6	0.82	0.67	67.37
Kubik	4221	0.77	0.59	72.52

Perbandingan hasil prediksi data curah hujan bulanan di Kabupaten Indramayu menggunakan model semiparametrik linier, kuadratik, dan kubik disajikan pada Gambar 1. Plot pada Gambar 1 memperlihatkan bahwa model semiparametrik berderajat satu (linier) memiliki pola yang lebih mirip dan selisih jarak lebih berdekatan dengan data aktual dibandingkan dengan plot data pada model semiparametrik berderajat dua (kuadratik) dan model semiparametrik berderajat tiga (kubik). Hal ini menunjukkan bahwa model semiparametrik berderajat satu (linier) mampu memprediksi data lebih baik model semiparametrik berderajat dua (kuadratik), dan model semiparametrik berderajat tiga (kubik).

### IV. SIMPULAN

Pola hubungan fungsional curah hujan dengan komponen utama GPCP tidak membentuk pola parametrik tertentu, dengan kata lain hubungannya dapat bersifat parametrik dan nonparametrik. Model semiparametrik SD berderajat satu (linier)



Gambar 1. Plot prediksi curah hujan

lebih baik dibandingkan model kuadratik dan kubik dalam memprediksi curah hujan bulanan di Kabupaten Indramayu. Jumlah titik simpul optimum model ini adalah 3, 4, 2, 6, 3, 11, 2, dan 1 masing-masing untuk KU1 sampai dengan KU8. Nilai RMSEP dan  $R^2$  yang diperoleh dari model semiparametrik SD berderajat satu (linier) adalah sebesar 61.64 dan 71%.

### DAFTAR PUSTAKA

- Bergant, K., L. Kajfez-Bogataj, and Z. Crepinsek (2015). Downscaling of general circulation-model simulated average monthly air temperature to the beginning of flowering of dandelion (*taraxacum officinale*) in slovenia. *Int J Biometeorol* 46, 22–32.
- Djuraidah, A. (2007). Model aditif spatio-temporal untuk pencemar udara PM10 dan ozon di kota Surabaya dengan pendekatan model linear campuran [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Heerdegen, R., A. Mullan, and F. Mpelasoka (2001). New zealand climate change information derived by multivariate statistical and artificial neural networks approaches. *Int J Climatol* 21, 1415–1433.
- Huth, R. and J. Kysely (2000). Constructing site-specific climate change scenarios on a monthly scale using statistical downscaling. *Theoretical and Applied Climatology* 66, 13–17.
- Johnson, R. and D. Wichern (2007). *Applied*

- multivariate statistical analysis*. New Jersey (NJ): Pearson Prentice Hall. Ed ke-6.
- Jolliffe, I. (2002). *Principal component analysis*. New York (NY): Springer-Verlag. Ed ke-2.
- Kodarsih, E. (2013). Pendugaan curah hujan bulanan statistical downscaling dengan jaringan syaraf tiruan sebaran Gamma [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Lanza, L., J. Ramirez, and i. E. Todin (2001). Stochastic rainfall interpolation and downscaling. *Hidrology and Earth System Science* 5(2), 139–145.
- Mehrotra, R. and A. Sharma (2007). Preserving low-frequency variability generated daily rainfall sequences. *J Hydrol* 345, 102–120.
- Sailor, D., T. Hu, X. Li, and J. Rosen (2000). A neural network approach to local downscaling of gcm output for assessing wind power implications of climate change. *Renewable Energy* 19, 359–378.
- Sarwoko, D. (2013). Pemodelan prediksi total hujan pada musim hujan menggunakan jaringan syaraf tiruan dan support vector regression [tesis]. Bogor (ID): Institut Pertanian Bogor.
- Soleh, A. (2015). Pemodelan linier sebaran gamma dan pareto terampat dengan regularisasi L1 pada statistical downscaling untuk pendugaan curah hujan bulanan [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Sutikno (2008). Statistical Downscaling Luaran GCM dan Pemanfaatannya untuk Prediksi Produksi Padi. [Disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Uvo, C., J. Olsson, O. Morita, K. Jinno, A. Kawamura, K. Nishiyama, N. Koreeda, and T. Nakashima (2001). Statistical atmospheric downscaling for rainfall estimations in kyushu island japan. *Hydrologi & Earth System Science* 5(2), 259–271.
- Wigena, A. Regresi kuadrat terkecil multi respon untuk statistical downscaling.
- Wigena, A. (2006). Pemodelan statistical downscaling dengan regresi projection pursuit untuk prediksi curah hujan bulanan [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Wigena, A., A. Djuraidah, and A. Rizki (2015).

Semiparametric modeling in statistical downscaling to predict rainfall. *Applied Mathematic Sciences* 9(88), 4371–4382.

## LAMPIRAN

Lampiran 1. (a) Plot curah hujan dengan KU1, (b) plot curah hujan dengan KU2, (c) plot curah hujan dengan KU3, (d) plot curah hujan dengan KU4, (e) Plot curah hujan dengan KU5, (f) plot curah hujan dengan KU6, (g) plot curah hujan dengan KU7, (h) plot curah hujan dengan KU8

