



Model Statistical Downscaling Nonparametrik pada Simulasi Data Curah Hujan Harian Pos Jurang Malang Daerah Aliran Sungai Jangkok

Mustika Hadijati^{a,}, Desy Komalasari^b, Irwansyah^c*

^aProgram Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, 83125, Indonesia.
Email: mustika.hadijati@unram.ac.id

^bProgram Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, 83125, Indonesia.
Email: desysky@gmail.com

^cProgram Studi Matematika, FMIPA, Universitas Mataram, Jl. Majapahit No. 62, Mataram, 83125, Indonesia.
Email: irw@unram.ac.id

ABSTRACT

The prediction of river water discharge can be determined by developing a river water discharge model based on climate information, especially rainfall information. This research aims to obtain a simulation of rainfall data that will be used to river water discharge modeling. The simulation of rainfall data is obtained using statistical downscaling model which develop the functional model between global climate data and local climate data. Daily precipitation of General Circulation Model (GCM) is used to be predictor variables. It is global climate data.. And, daily rainfall of Jangkok watershed, the local climate data, is used to be response variable.. In order to reduce the dimension of GCM global data, GCM data is projected to a little number of variable using classification and regression tree (CART) method. Then, the projection variables are used to develop statistical downscaling model of rainfall based on Kernel nonparametric regression. Daily rainfall data of Jurang Malang station, Jangkok watershed, is simulated based on the model obtained.

Keywords : statistical downscaling, CART, kernel nonparametric regression.

1. Pendahuluan

Data debit air sungai dari suatu Daerah Aliran Sungai (DAS) merupakan informasi yang penting bagi pengelola sumberdaya air. Prediksi debit air sungai yang tepat diperlukan sebagai acuan dalam menyusun rencana pengelolaan sumberdaya air. Untuk mendapatkan prediksi debit air sungai yang tepat diperlukan pemodelan debit air sungai yang mempunyai ketepatan tinggi. Pemodelan debit air sungai dapat dilakukan berdasarkan faktor-faktor yang mempengaruhinya, antara lain adalah curah hujan. Oleh

karena itu, pada penelitian ini pemodelan debit air sungai dilakukan berdasarkan data hasil simulasi curah hujan, dimana simulasi data didasarkan pada model curah hujan.

Untuk meningkatkan kualitas informasi iklim, saat ini banyak dikembangkan model iklim *General Circulation Model* (GCM). Namun, informasi model GCM masih berskala global, sehingga kurang tepat bila digunakan pada peramalan berskala lokal/regional. Untuk mendapatkan informasi berskala lokal, perlu dilakukan regionalisasi dengan menggunakan teknik *downscaling*, salah satunya adalah *statistical downscaling* (SD). Dengan SD akan

* Corresponding author : mustika.hadijati@unram.ac.id

disusun model hubungan fungsional antara informasi luaran GCM dengan informasi variabel iklim lokal/regional, dimana dalam penggunaannya perlu diperhatikan permasalahan sifat luaran GCM sebagai variabel prediktor dan sifat data iklim lokal/regional sebagai variabel respon (Wigena, 2006).

Data luaran GCM bersifat *curse of dimensionality*, data spasial dan temporal sehingga memungkinkan terjadinya korelasi spasial antara data pada grid yang berbeda dalam satu domain. Oleh karena itu perlu dilakukan pra-pemrosesan data GCM dengan mereduksi dimensi data GCM. Data GCM pada umumnya bersifat nonnormal, sehingga digunakan metode pra-pemrosesan data yang tidak mensyaratkan asumsi kenormalan data, yaitu algoritma *Classification and Regression Tree* (CART) (Zorita & Storch, 1999; Kannan & Gosh, 2013; Hadijati, dkk., 2016).

Selanjutnya dilakukan regresi antara hasil proyeksi data GCM sebagai variabel prediktor dengan curah hujan lokal sebagai variabel respon. Oleh karena data iklim pada umumnya bersifat nonlinier, nonstasioner, dan nonnormal, maka digunakan metode regresi nonparametrik, khususnya regresi nonparametrik Kernel. Menurut Kannan dan Ghosh (2013), metode ini dapat menjelaskan keragaman temporal dan nilai ekstrim dari data iklim.

Berdasarkan uraian di atas, maka pada penelitian ini dilakukan simulasi data curah hujan harian di Pos Jurang Malang DAS Jangkok menggunakan model *statistical downscaling* nonparametrik Kernel.

2. Data dan Metode

Penelitian ini menggunakan dua macam data, yaitu data presipitasi harian GCM yang merupakan data bersifat global yang digunakan sebagai variabel prediktor, dan data curah hujan harian di Pos Jurang Malang yang terdapat di DAS Jangkok yang merupakan data bersifat lokal yang digunakan sebagai variabel respon. Kedua macam data tersebut yang digunakan adalah data selama 3 tahun terakhir (1 Januari 2014 – 31 Desember 2016).

Data presipitasi luaran GCM yang digunakan adalah CPC *Global Precipitation* yang diperoleh dari <https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.global.precip.html>. Data tersebut merupakan data total presipitasi harian pada lokasi *latitude* dan *longitude* tertentu. Pos Jurang Malang terletak di Pulau Lombok, sehingga domain data yang dipilih adalah domain 8X8 di atas Pulau Lombok yang terletak kurang lebih pada 8.13° – 8.88° LS dan 115.46° – 116.48° BT. Data yang dikumpulkan adalah data pada 64 posisi *latitude* dan *longitude*. Adapun data curah hujan harian diamati dari rekaman data curah hujan bulanan di

Balai Wilayah Sungai Nusa Tenggara I (BWS NT I) Nusa Tenggara Barat.

Pemodelan curah hujan dilakukan dengan pendekatan *statistical downscaling* (SD), yaitu proses *downscaling* yang bersifat statik dimana data pada grid-grid berskala besar dalam periode dan jangka waktu tertentu digunakan sebagai dasar untuk menentukan data pada grid berskala lebih kecil (Wigena, 2006). Menurut Lembang, dkk. (2009), pendekatan SD menggunakan data global untuk memperoleh hubungan fungsional antara skala lokal dan dengan skala global GCM, dimana bentuk hubungan tersebut secara umum dinyatakan sebagai berikut

$$Y = f(Z) + \varepsilon \quad (1)$$

dimana Y : variabel respon

Z : variabel prediktor yang merupakan gabungan dari hasil reduksi spasial variabel GCM

ε : sisaan/error

Pada pemodelan *statistical downscaling* curah hujan ini langkah analisis data meliputi dua tahapan, yaitu

2.1. Reduksi Dimensi Data GCM dengan Algoritma *Classification and Regression Tree* (CART)

Langkah ini dilakukan untuk mengatasi sifat *curse of dimensionality*, spasial temporal, dan multikolinieritas pada data GCM yang merupakan data global dengan dimensi tinggi dan resolusi rendah. *Classification and Regression Tree* (CART) adalah metode klasifikasi yang menggunakan data historis yang disebut *learning sample* untuk mengkonstruksi pohon keputusan (*decision tree*). Pohon keputusan tersebut kemudian digunakan untuk mengklasifikasi data baru (Timoveef, 2004). Dengan analisis CART akan didapatkan proyeksi data presipitasi GCM yang digunakan sebagai variabel prediktor dengan dimensi yang lebih rendah dan resolusi yang lebih tinggi. Adapun langkah-langkah algoritma CART adalah sebagai berikut (Timoveef, 2004):

1) Mengkonstruksi pohon maksimum

Konstruksi dilakukan berdasarkan aturan pembagian (*splitting rule*) tertentu, dimana setiap data waktu dibagi menjadi dua bagian dengan kehomogenan maksimum dari node-node *child* yang didefinisikan oleh fungsi *impurity* $i(t)$. Fungsi *impurity* yang sering digunakan adalah aturan pembagian Gini (*Gini index*) yang mendefinisikan fungsi *impurity* sebagai berikut:

$$i(t) = \sum_{k \neq 1} p(k|t)p(l|t) \quad (2)$$

dimana $k, l = 1, 2, \dots, K$ = indeks *class*; $p(k|t)$ = peluang bersyarat *class* k bila dalam node t .

Berdasarkan indeks Gini, pada tiap node CART akan diselesaikan masalah pemaksimalan berikut

$$\operatorname{argmax}_{x_j \leq x_j^R, j=1,2,\dots,M} \{-\sum_{k=1}^K p^2(k|t_p) + P_l \sum_{k=1}^K p^2(k|t_l) + P_r \sum_{k=1}^K p^2(k|t_r)\} \quad (3)$$

dimana P_l dan P_r adalah peluang node kiri dan node kanan.

2) Melakukan pemilihan ukuran pohon yang tepat

Pemilihan ukuran pohon yang tepat merupakan proses optimasi yang dilakukan untuk mengurangi kompleksitas pohon maksimum. Metode yang digunakan adalah *cross-validation*, dimana akan ditentukan proporsi optimal antara kompleksitas pohon dan *misclassification error* dengan meminimumkan fungsi *cost-complexity* berikut

$$R_\alpha(T) = R(T) + \alpha(\tilde{T}) \quad (4)$$

dimana $R(T)$ = *misclassification error* pohon T , $\alpha(\tilde{T})$ = ukuran kompleksitas yang tergantung pada \tilde{T} (total banyaknya node terminal pada pohon).

3) Mengklasifikasi data baru berdasarkan pohon optimum yang telah terbentuk dan mendapatkan proyeksi data GCM.

2.2. Pemodelan Curah Hujan Pos Jurang Malang dengan Regresi Nonparametrik Kernel

Sebagai langkah selanjutnya adalah memodelkan proyeksi data presipitasi GCM dan data curah hujan harian di Pos Jurang Malang dengan metode regresi nonparametrik Kernel multivariat. Secara umum, model regresi nonparametrik dapat dirumuskan sebagai berikut:

$$Y_t = m(x) + u_t, \quad t = 1, 2, \dots, T \quad (5)$$

dengan *error* u_t dan memenuhi $E(u_t) = \mathbf{0}$. Fungsi $m(\cdot)$ diasumsikan tidak diketahui bentuknya tetapi *smooth*. Sedangkan x adalah nilai dari $X = (X_1, X_2, \dots, X_q)^T$ yang merupakan beberapa variabel prediktor yang bersifat acak. Adapun Y_t adalah variabel respon. Kurva regresi $m(x)$ pada dasarnya merupakan *mean* bersyarat, $m(x) = E(Y_t|x)$, dan dapat diestimasi dengan estimator Kernel.

Estimator Kernel yang digunakan pada penelitian ini adalah estimator Nadaraya-Watson. Estimator tersebut sering digunakan untuk kasus variabel prediktor acak (Eubank, 1988). Estimator Nadaraya-Watson pada regresi nonparametrik multivariat dirumuskan sebagai berikut (Hardle & Muller, 1997):

$$\hat{m}_H(x) = \frac{\sum_{t=1}^T K_H(X_t - x) Y_t}{\sum_{t=1}^T K_H(X_t - x)} \quad (6)$$

Adapun langkah pemodelan regresi nonparametrik Kernel adalah sebagai berikut:

- 1) Menentukan estimator Nadaraya-Watson berdasarkan persamaan 6. Estimator Kernel Nadaraya-Watson dipengaruhi oleh fungsi Kernel $K(u)$. Salah satu fungsi Kernel adalah fungsi Kernel Kuartik yang didefinisikan sebagai berikut (Hardle, 2004)

$$K(u) = \frac{5}{6}(1 - u^2)^2 \quad (7)$$

- 2) Menentukan pemilihan *bandwidth* optimum berdasarkan kriteria *GCV*.

Pada penelitian ini pemilihan *bandwidth* optimum didasarkan kriteria *generalized cross validation (GCV)* yang dirumuskan sebagai berikut (Hardle & Muller, 1997):

$$GCV = \frac{n^{-1} \sum_{t=1}^n (y_t - \hat{m}_H(X_t))^2}{(n^{-1} \operatorname{trace}(I - H(h_0)))^2} \quad (8)$$

- 3) Menentukan model regresi nonparametrik Kernel terbaik dengan *bandwidth* optimum, dimana model tersebut yang menunjukkan hubungan fungsional proyeksi data presipitasi luaran GCM dengan data curah hujan bulanan (model *statistical downscaling* curah hujan harian di Pos Jurang Malang DAS Jangkok).

Setelah dilakukan pemodelan *statistical downscaling* curah hujan harian Pos Jurang Malang DAS Jangkok, dilanjutkan dengan melakukan simulasi data curah hujan harian Pos Jurang Malang DAS Jangkok berdasarkan model yang diperoleh.

3. Hasil dan Pembahasan

Sebelum dilakukan reduksi dimensi dengan metode CART, data curah hujan harian Pos Jurang Malang diubah menjadi data kategori berdasarkan metode *k-mean cluster*. Ringkasan hasil pengelompokkan berdasarkan *k-mean cluster* tersebut dapat dilihat pada tabel 1.

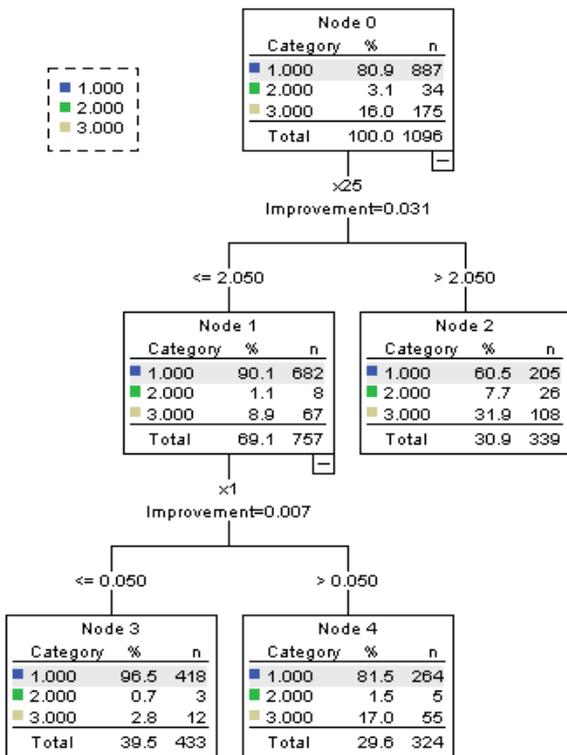
Tabel 1. Ringkasan Pengelompokkan Curah Hujan Harian

Kategori	Jumlah Kasus	Center
1	887	1.54
2	34	73.56
3	175	28.05

Tabel 1 menunjukkan bahwa data curah hujan Pos Jurang Malang dikelompokkan dalam tiga kategori. Data

curah hujan Pos Jurang Malang yang masuk dalam kategori 1 sebanyak 887 data dengan nilai *center* 1.54. Pada kategori dua terdapat 34 data dengan *center* 73.56, sedangkan pada kategori 3 terdapat 175 data dengan *center* 28.05.

Selanjutnya, berdasarkan data kategori curah hujan harian Pos Jurang Malang dan data presipitasi GCM (CPC *Global Precipitation*) pada sejumlah grid yang tercakup dalam domain yang ditentukan dilakukan reduksi data GCM dengan algoritma CART. Dengan metode ini dihasilkan pohon klasifikasi curah hujan Pos Jurang Malang berdasarkan data presipitasi GCM, dimana akan terlihat variabel-variabel prediktor yang dominan mempengaruhi respon. Adapun variabel-variabel prediktor yang dimaksud di sini adalah presipitasi GCM pada grid-grid yang tercakup dalam domain. Hasil analisis CART yang diperoleh terlihat pada gambar 1 dan tabel 2.



Gambar 1. Pohon Klasifikasi Curah Hujan Pos Jurang Malang

Tabel 2. Ringkasan Hasil CART Curah Hujan Harian Pos Jurang Malang

Jumlah Node	Jumlah Node terminal	Ketepatan klasifikasi (%)	Variabel dalam Model	Tingkat Kepentingan (%)
5	3	89.2	X ₂₅	100.0
			X ₁	53.3

Berdasarkan gambar 1 dan tabel 2 dapat diketahui bahwa Pada pohon klasifikasi tersebut terdapat 5 node

dengan node terminal sebanyak 3 dan kedalaman pohon 2. Hasil klasifikasi curah hujan harian Pos Jurang Malang berdasarkan algoritma CART ini dapat dikatakan relatif baik karena ketepatan klasifikasi yang diperoleh sebesar 89.2%.

Pada pohon klasifikasi juga terlihat bahwa variabel yang masuk ke dalam model adalah variabel X₁ (presipitasi GCM pada grid 6.75° LS dan 113.75° BT), dan X₂₅ (presipitasi GCM pada grid 9.25° LS dan 116.75° BT). Tingkat kepentingan masing-masing variabel adalah 100% untuk variabel X₂₅ dan 53.3% untuk variabel X₁. Tingkat kepentingan ini menunjukkan tingkat pengaruh suatu variabel prediktor terhadap hasil klasifikasi. Berdasarkan hasil tersebut, variabel yang berpengaruh terhadap klasifikasi curah hujan lokal adalah variabel X₁ dan X₂₅. Sehingga kedua variabel tersebut dijadikan sebagai prediktor dalam pemodelan curah hujan Pos Jurang Malang dengan regresi nonparametrik Kernel.

Pemodelan curah hujan di Pos Jurang Malang DAS Jangkok menggunakan data curah hujan harian pada Pos Jurang Malang DAS Jangkok sebagai variabel respon. Sebagai variabel prediktor digunakan hasil proyeksi data GCM untuk Pos Jurang Malang DAS Jangkok, yaitu variabel X₁ (presipitasi GCM pada grid 6.75° LS dan 113.75° BT), dan X₂₅ (presipitasi GCM pada grid 9.25° LS dan 116.75° BT).

Pemilihan *bandwidth* optimum dilakukan untuk mendapatkan model yang terbaik. Kriteria yang digunakan adalah kriteria GCV. *Bandwidth* optimum adalah nilai *bandwidth* yang memberikan nilai GCV minimum. Proses pemilihan *bandwidth* dilakukan secara iterative hingga diperoleh nilai yang optimum. Beberapa nilai *bandwidth* dan GCV yang dihasilkan terlihat pada tabel 3.

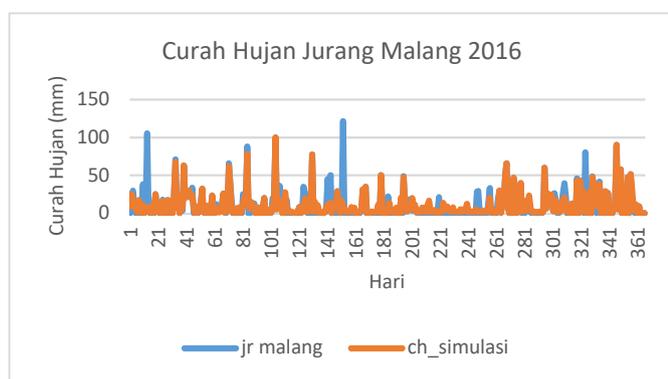
Tabel 3. Nilai Bandwidth dan GCV

h1	h2	GCV	h1	h2	GCV
0.4	0.23	212.261	0.4	0.241	212.207
0.4	0.231	212.249	0.4	0.242	212.211
0.4	0.232	212.239	0.4	0.243	212.217
0.4	0.233	212.229	0.4	0.244	212.224
0.4	0.234	212.221	0.4	0.245	212.232
0.4	0.235	212.215	0.4	0.246	212.242
0.4	0.236	212.210	0.4	0.247	212.253
0.4	0.237	212.207	0.4	0.248	212.265
0.4	0.238	212.205	0.4	0.249	212.279
0.4	0.239	212.204	0.4	0.25	212.2944
0.4	0.24	212.205			

Pada tabel 1 terlihat bahwa *bandwidth* optimum

yang diperoleh adalah $h_1 = 0.4$ dan $h_2 = 0.239$ yang menghasilkan GCV minimum sebesar 212.204. Dengan menggunakan *bandwidth* optimum tersebut, ditentukan model curah hujan di Pos Jurang Malang DAS Jangkok berdasarkan regresi nonparametrik Kernel.

Berdasarkan model yang diperoleh, dilakukan simulasi curah hujan harian di Pos Jurang Malang DAS Jangkok untuk 1 tahun terakhir. Pola curah hujan aktual dan hasil simulasi dapat dilihat pada gambar 2. Gambar tersebut menunjukkan pola yang relatif sama antara nilai curah hujan aktual dan prediksi di Pos Jurang Malang DAS Jangkok. Nilai RMSEP dari hasil simulasi adalah sebesar 11.13. Nilai ini menunjukkan bahwa data simulasi yang dihasilkan berdasarkan model mempunyai tingkat kesalahan yang relative kecil.



Gambar 2. Pola Curah Hujan Aktual dan Hasil Simulasi

4. Kesimpulan

Pada Pos Jurang Malang DAS Jangkok, dengan algoritma CART data presipitasi GCM direduksi dimensinya dan diproyeksikan ke variabel-variabel X_1 (presipitasi GCM pada grid 6.75° LS dan 113.75° BT) dan X_{25} (presipitasi CGP pada grid 9.25° LS dan 116.75° BT). Selain itu, model *statistical downscaling* curah hujan harian di Pos Jurang Malang DAS Jangkok dengan regresi nonparametrik Kernel yang dihasilkan memberikan hasil simulasi data curah hujan dengan RMSE yang relatif kecil, yaitu 11.13. Hal ini menunjukkan bahwa data simulasi yang dihasilkan

cukup akurat untuk digunakan dalam pemodelan debit air sungai selanjutnya..

DAFTAR PUSTAKA

- Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker. Inc.
- Hadijati, M., Komalasari, D., Fitriyani, N. (2016). Statistical Downscaling Regresi Nonparametrik Kernel untuk Prediksi Curah Hujan Bulanan Stasiun Sembalun, *Prosiding Seminar Nasional Matematika 2016 Universitas Udayana*, 186-196.
- Hardle, W. (2004). *Applied Nonparametric Regression*. <http://www.quantlet.de/mdstat/scripts/anr/html/anrhtmltoc.htm>. 2-5-2005
- Hardle, W. Muller, M. (1997). Multivariate and Semiparametric Regression. <http://www.edoc.hu-berlin.de/series/sfb-373-papers/1997-26/PDF/26.pdf>. 26-4-2013
- Kannan, S., Ghosh, S. (2013). A Nonparametric Kernel Regression Model for Downscaling Multisite Daily Precipitation in The Mahanadi Basin. *Water Resources Research*, **49**.
- Lembang, F.K., Setiawan, Sutikno. (2009). Pengembangan Model Ramalan Curah Hujan untuk Pemodelan *Statistical Downscaling* dengan Pendekatan Regresi *Bayes PCA*. Seminar Nasional Statistika VI, FMIPA ITS, Surabaya.
- Sarda, P. and Vieu, P. (2000). Kernel Regression. In: *Smoothing and Regression: Approaches, Computation, and Application*. Eds: Schimek, M.G. Canada: John Wiley & Sons, Inc., Pp. 43-67.
- Timofeev, R. (2004). Classification and Regression Trees (CART): Theory and Application. Master Thesis, Humboldt University, Berlin.
- Wigena, A.H. (2006). Pemodelan *Statistical Downscaling* dengan Regresi *Projection Pursuit* untuk Peramalan Curah Hujan Bulanan: Kasus Curah Hujan Bulanan di Indramayu, Disertasi Pascasarjana IPB.
- Zorita, E., Storch, H.V. (1999). The Analog Method as A Simple Statistical Downscaling Technique: Comparison with Complicated Methods. *J Clim*, **12**, 2474-2489.