

PEMODELAN DATA CURAH HUJAN KABUPATEN INDRAMAYU MENGUNAKAN REGRESI KOMPONEN UTAMA DAN REGRESI KOMPONEN UTAMA FUNGSIONAL

Achi Rinaldi

Jurusan Pendidikan Matematika IAIN Raden Intan Lampung
achi_rinaldi@yahoo.co.uk

ABSTRACT

This research was aimed to modelling rainfall in Indramayu based on the information from Sukadana rainfall station. Modelling rainfall data were very usefull for some prediction in the future and for government policy which can prevent some disaster like flood. Statistics methods in this paper were Principal Component Analysis (PCA), Principal Component Regression (PCR), Functional Principal Component Analysis (FPCA), and Functional Principal Component Regression (FPCR). Principal Component Regression (PCR) perform a good result compare with Functional Principal Component Regression (FPCR), but without basis function the both method are equal. R^2 value for PCR was 50.13 the same with FPCR without basis function. RMSEP value and correlation were 77.55 and 0.91 and it showed that PCR was better than FPCR with bspline basis function and fourrier which the value were 110.48 and 0.81, also 136,06 and 0.1. For general FPCR model without basis function was showed the best performance especially FPCR central model that give prediction result for rainfall data was better than the other model.

Keywords: PCA, FPCA, PCR, FPCR, Rainfall Data

PENDAHULUAN

Salah satu indikasi kuat dikenalnya Indonesia sebagai satu-satunya kawasan unik di daerah ekuator sebagai Benua Maritim Indonesia (BMI) adalah adanya keragaman curah hujan yang cukup besar yang terjadi di beberapa kawasan di Indonesia, khususnya yang terjadi di Pulau Jawa. Selain mendapat pengaruh dari sirkulasi udara pada skala global maupun regional, pembentukan awan dan hujan di Indonesia juga dipengaruhi oleh kondisi lokal, seperti topografi dan suhu permukaan laut di perairan Indonesia.

Pulau Jawa secara keseluruhan juga memiliki karakteristik iklim yang khas secara regional maupun lokal. Wilayahnya memiliki barisan pegunungan yang membujur dari Timur sampai Barat serta dikelilingi oleh lautan yang terdiri dari Samudera Hindia, Laut Jawa, Selat Sunda, Selat Bali dan Selat Madura. Hal ini menyebabkan proses pembentukan awan dan hujan di Pulau Jawa mendapat

pengaruh dari kondisi alam tersebut selain pengaruh dari pergerakan posisi semu matahari terhadap bumi dan sirkulasi global.

Karakteristik iklim, khususnya perilaku curah hujan di Pulau Jawa dapat dianalisis secara akurat berdasarkan data iklim dari stasiun meteorologi. Untuk analisis spasial, hal ini sangat ditentukan oleh kerapatan jaringan penakar hujan. Untuk daerah-daerah dengan jaringan penakar hujan yang cukup rapat dan merata seperti di Pulau Jawa hal tersebut tidak menjadi masalah.

Disini terlihat bahwa Indonesia merupakan satu kawasan daerah tropis yang unik dimana dinamika atmosfernya dipengaruhi oleh kehadiran angin pasat, aliran angin monsun, iklim marine dan pengaruh berbagai kondisi lokal. Cuaca dan iklim di Indonesia mempunyai karakteristik khusus yang hingga kini mekanisme proses pembentukannya belum banyak diketahui.

Iklim dapat didefinisikan sebagai ukuran statistik cuaca untuk jangka waktu tertentu dan cuaca menyatakan status atmosfer pada sembarang waktu tertentu. Dua unsur utama iklim adalah suhu dan curah hujan. Indonesia sebagai daerah tropis ekuatorial mempunyai variasi suhu yang kecil, sementara variasi curah hujannya cukup besar. Oleh karena itu curah hujan merupakan unsur iklim yang paling sering diamati dibandingkan dengan suhu.

Secara umum curah hujan di wilayah Indonesia didominasi oleh adanya pengaruh beberapa fenomena, antara lain sistem Monsun Asia-Australia, El-Nino, sirkulasi Timur-Barat (*Walker Circulation*) dan Utara-Selatan (*Hadley Circulation*) serta beberapa sirkulasi karena pengaruh lokal.

Data curah hujan merupakan data yang didapatkan dari stasiun-stasiun cuaca yang secara kontinyu diukur berulang kali dalam skala hari, minggu atau bulan. Karena diukur secara kontinyu dan tidak terputus maka data curah hujan sangat tepat jika dibuat dalam bentuk kurva atau fungsi. Penanganan data yang berbentuk fungsi atau biasa disebut data fungsional berbeda dengan data pada umumnya. Pada data biasa maka tiap titik yang memiliki nilai adalah datanya, sedangkan data fungsional bentuk datanya adalah fungsi.

Penggunaan Analisis Data Fungsional bertujuan (Ramsay & Silverman, 2005):

- a. Merepresentasikan data dalam berbagai cara untuk membantu analisis lebih lanjut.
- b. Menggambarkan data secara jelas dengan karakteristik yang sangat bervariasi.
- c. Mempelajari sumber dari pola dan variasi dari data.
- d. Menjelaskan variasi dari peubah respon dengan menggunakan informasi predictor.
- e. Membandingkan dua atau lebih peubah fungsional.

Penelitian yang bertujuan untuk memprediksi curah hujan telah banyak dilakukan, dimana satu cara untuk melakukan prediksi curah hujan yaitu dengan memanfaatkan model iklim global atau *global circulation model* (GCM). GCM merupakan model data memungkinkan untuk mendapatkan informasi iklim pada waktu lampau, sekarang serta dapat memprediksi perubahan iklim yang akan terjadi dimasa yang akan datang (Wigena 2006). Metode yang sering digunakan untuk menganalisis data GCM yaitu Analisis komponen utama (*Principal Component Analysis/PCA*) dan regresi komponen utama (*Principal Component Regression/PCR*). Penggunaan PCR untuk mereduksi dimensi data GCM dan memprediksi curah hujan pernah dilakukan oleh Subimal dan Mujumdar pada tahun 2005.

Penggunaan PCA dapat memberikan dampak adanya multikolinier, adanya auto korelasi. Untuk mengatasi hal tersebut dapat digunakan metode Analisis komponen utama fungsional (FPCA) melalui fungsi basis. Selain dapat mengatasi kendala multikolinieritas dan autokorelasi, FPCA juga mampu menangkap pola dan varians dari data. Metode FPCA dapat menggunakan fungsi basis bspline atau fungsi basis fourier.

Tujuan

Tujuan dari penelitian ini yaitu :

- a. Mengkaji penerapan analisis komponen utama (PCA), komponen utama fungsional (FPCA), regresi komponen utama (PCR), dan regresi komponen utama fungsional (FPCR) pada data curah hujan di Kabupaten Indramayu.
- b. Mengkaji model terbaik yang dapat digunakan untuk memprediksi data curah hujan di Kabupaten Indramayu secara tepat dan akurat.

Analisis Komponen Utama

Analisis Komponen Utama (PCA) pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901. Prinsip dasar dari PCA adalah membentuk peubah baru

yang merupakan kombinasi linier dari peubah lama sedemikian sehingga peubah baru tersebut tidak saling berkorelasi atau saling bebas. Dimensi dari peubah baru tersebut akan lebih kecil dari dimensi peubah lama, jika peubah lama memiliki dimensi sebanyak n maka peubah baru akan memiliki dimensi $< n$. Sehingga dengan kata lain PCA dapat juga mereduksi dimensi dari peubah. Saat ini PCA banyak digunakan dalam analisis data eksplorasi dan untuk membuat model prediksi.

PCA merupakan analisis yang memerlukan perhitungan dekomposisi akar karakteristik (*eigen value/akar ciri*) dari matriks covarian data atau *Singular Value Decomposition(SVD)* dari matrix data. Dalam analisis Regresi penggunaan PCA dapat menanggulangi masalah multi kolinieritas atau adanya korelasi antar peubah yang seharusnya secara asumsi saling bebas.

Misal X matriks data, maka SVD dari X adalah:

$$X_{n \times p} = U_{n,k} D_{k \times k} (V_{k \times p})^T$$

dengan U dan V adalah matriks yang ortogonal, sedangkan D adalah matriks diagonal dengan nilai diagonalnya adalah $d_1 > d_2 > \dots > d_k$. Matriks UD berisi skor komponen utama. Selanjutnya Mencari transformasi dari matriks X ($n \times p$) sehingga

$$Y = \delta^T X = \delta_1 X_1 + \delta_2 X_2 + \dots + \delta_p X_p$$

dimana $\delta = (\delta_1, \delta_2, \dots, \delta_p)^T$ adalah vektor kolom dari pembobot dengan $\delta_1^2 + \delta_2^2 + \dots + \delta_p^2 = 1$.

Kemudian cari δ dimana $Var(\delta^T X) = \delta^T Var(X) \delta$ maksimum, dimana matriks $C = Var(X)$ adalah matriks kovarian dari peubah X_i .

Selanjutnya kita dapatkan peubah baru Y_i yang merupakan kombinasi linier dari peubah asli (x_i): $Y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ip}x_p$; $i=1..p$. Peubah baru Y_i tersebut memiliki koefisien yang saling bebas dan didalamnya mengandung informasi keragaman menurun secara hirarki dan ini yang disebut sebagai Komponen Utama.

Untuk menghitung nilai *akar ciri* λ_i yaitu dengan menyelesaikan persamaan

$det(C - \lambda I) = 0$. Sedangkan nilai *vektor ciri* merupakan nilai dari kolom-kolom matriks A dimana : $C = A D A^T$.

Kriteria untuk menentukan jumlah komponen utama yang dipilih sebagai berikut:

- Menggunakan *Scree plot*. Banyaknya komponen yang diambil sesuai dengan titik *cut-off*.
- Menggunakan kriteria akar ciri kumulatif dimana $\sum_{j=1}^p \lambda_j \geq 0.8$

Regresi Komponen Utama

PCA sangat berperan penting dalam pendugaan data curah hujan dikarenakan variabel yang digunakan sangatlah banyak serta memiliki multikolinieritas yang tinggi, sehingga diperlukan analisis PCA untuk mereduksi dan membentuk peubah baru yang saling bebas dan dapat menyelesaikan prediksi curah hujan menggunakan regresi. Bentuk kombinasi antara PCA dan analisis regresi disebut Regresi Komponen Utama (PCR). Peubah baru yang dihasilkan oleh PCA akan menghasilkan skor komponen utama dan pemilihan skor komponen utama akan dianalisis dengan menggunakan PCR.

Model fungsi linier berganda dapat ditulis sebagai :

$$Y = X\beta + \varepsilon$$

Y = vektor $n \times 1$ nilai amatan peubah tak bebas

X = matriks berukuran $n \times p$ dengan unsur-unsur ke (i,j) adalah nilai peubah bebas ke j untuk pengamatan ke i .

β = vektor $p \times 1$ dari koefisien regresi

ε = vektor $n \times 1$ dari galat

Model regresi di atas dapat dinyatakan sebagai bentuk *central*, dengan mengukur semua peubah dari rata-rata masing-masing (Myres, R. H.1990). Lebih jauh, dalam berbagai literatur analisis komponen utama mengasumsikan bahwa peubah bebas telah terbakukan sehingga $X^T X$ proporsional terhadap matriks korelasi peubah bebas. Misalnya suatu matriks pengamatan X telah terbakukan, dilambangkan dengan X^* sehingga diperoleh akar ciri

(λ) dan vektor ciri (V) dari X^*X^* (bentuk korelasi). Kita ketahui $V^*V = I$ karena V ortonormal, sehingga model regresi asal dapat dituliskan sebagai berikut :

$$Y = \beta_0 I + X^*VV^* \beta + \varepsilon$$

$$Y = \beta_0 I + Z \alpha + \varepsilon$$

dengan $Z = X^*V$ dan $\alpha = V^* \beta$

Z adalah matriks $n \times p$ dan α adalah vektor $p \times 1$ dengan koefisien regresi yang baru adalah $\alpha_1, \alpha_2, \dots, \alpha_p$. Kita dapat menyajikan kolom-kolom matriks Z sebagai peubah baru (komponen utama) yang merupakan kombinasi linier dari peubah-peubah bebas X , yaitu $Z = v_{1j}X_{1j}^* + v_{2j}X_{2j}^* + \dots + v_{rj}X_{rj}^*$. Telah ditunjukkan bahwa Z saling ortogonal sehingga jika regresi dibangun atas Z melalui model (14) maka keragaman koefisien regresi balikan dari akar ciri matriks Z^*Z

$$\frac{Var(\hat{\alpha}_j)}{\sigma^2} = \frac{1}{\lambda_j} \quad j = 1, 2, \dots, k$$

Jika semua komponen utama (KU) tetap dalam model regresi, maka yang terjadi hanyalah transformasi berupa rotasi peubah bebas. Sekalipun masing-masing KU saling ortogonal, ragam penduga yang besar tetap tertinggal. Keragaman total hanya terdistribusi ulang (tidak tereduksi). Jika multikolinearitas amat parah, akan terdapat minimal satu akar ciri yang kecil. Eliminasi satu (minimal satu) komponen utama yang bersesuaian dengan akar ciri yang kecil secara substantial dapat mereduksi total keragaman dalam model sehingga menghasilkan perbaikan yang cukup besar pada dugaan persamaan regresi.

Analisis Data Fungsional

Analisis data fungsional (*functional data analysis/FDA*) pertama kali diperkenalkan oleh Ramsay (1982). Prinsipnya semua pemodelan dapat dikonversikan menjadi data fungsional. Kelebihan data fungsional dibandingkan dengan data biasa adalah dapat mengatasi kendala auto korelasi dan dapat melihat karakteristik pola data secara berkala atau series. Penggunaan FDA dapat dilakukan melalui

pendekatan dengan fungsi basis bspline dan fungsi basis fourier. Metode untuk menganalisis data fungsional hampir sama dengan analisis data multivariat, Analisis yang banyak digunakan untuk FDA yaitu PCA yang dinamakan FPCA. Prinsip kerja dari FPCA dan PCA sama yaitu untuk mereduksi dimensi data namun tetap mempertahankan keragaman data. Perbedaan antara metode PCA dan FPCA yaitu pada metode PCA digunakan untuk analisis peubah ganda (x) sedangkan FPCA untuk analisis data fungsional ($x(t)$).

Dalam FPCA bertujuan mencari himpunan dari fungsi komponen utama yang ortogonal dengan memaksimumkan ragam di tiap komponen, serta mencari fungsi komponen utama pertama $\phi_1(x)$. Adapun ragam dari skor komponen utama

$$\beta_1 = \int_{x_1}^{x_p} \phi_1(x)f(x)dx$$

ini dimaksimumkan berdasarkan

$\|\phi_1^2(x)\| = \int_{x_1}^{x_p} \phi_1(x)dx = 1$, β_1 adalah himpunan skor komponen utama pertama dengan nilai rata-ratanya sama dengan nol, $f(x)$ adalah himpunan kurva fungsi terdesentralisasi dimana $[x_1; x_p]$ adalah batasan rentang dari fungsi. Fungsi-fungsi komponen utama berikutnya dapat diperoleh secara iteratif dengan mengurangi k fungsi komponen utama pertama dari $f^0(x) = f(x)$ untuk $1 \leq k \leq K < n$, yaitu $f^k(x) = f^{k-1}(x) - \beta_k \phi_k(x)$ dan memperlakukan $f^k(x)$ sebagai koleksi fungsi baru untuk menemukan $\phi_{k+1}(x)$.

Perbedaan antara PCA dan FPCA dapat dilihat dalam tabel berikut (Shang, 2011) :

	PCA	FPCA
Peubah-peubah	$X=[x_1, \dots, x_p]$, $x_i = [x_{1i}, \dots, x_{ni}]'$, $i = 1, \dots, p$	$f(x)=[f_1(x), \dots, f_n(x)]$, $x \in [x_1, x_p]$
Data	Vektor-vektor $\in R^p$	Kurva-kurva $\in L_2[x_1, x_p]$
Koragam	Matriks $V=cov(X) \in R^p$	Operator T terbatas antara x_1 dan x_p , $T: L_2[x_1, x_p] \rightarrow L_2[x_1, x_p]$
Struktur eigen	Vektor $\xi_k \in R^p$	Fungsi $\xi_k(x) \in L_2[x_1, x_p]$

	R^p $V_{\xi_k} = \lambda_k \xi_k$, Untuk $1 \leq k$ $\leq \min(n, p)$	$L_2[x_1, x_p]$, $\int_{x_1}^{x_p} T \xi_k(x) dx =$ $\lambda_k \xi_k(x)$ Untuk $1 \leq k \leq n$
Komponen-komponen	Peubah acak dalam R^p	Peubah acak dalam $L_2[x_1, x_p]$

Untuk data yang berbentuk fungsi, maka dapat juga dibentuk regresi komponen utama dimana data $x_i(t) = \sum f_{ij} \xi_j(t)$ dengan memodelkan $y_i = \beta_0 + \sum \beta_j f_{ij} + \varepsilon_i$ atau dalam bentuk $y = F\beta + \varepsilon$ Secara teknis, model $y_i = \beta_0 + \sum \beta_j f_{ij} + \varepsilon_i$ dapat dibentuk dengan mensubstitusi nilai $f_{ij} = \int x_i(t) \xi_j(t) dt$ sehingga model tersebut menjadi

$y_i = \beta_0 + \sum \int \beta_j \xi_j(t) x_i(t) dt + \varepsilon_i$ Selanjutnya interpretasikan $\beta(t) = \sum \beta_j \xi_j(t)$ sehingga kita dapatkan model regresi $y_i = \beta_0 + \int \beta(t) x_i(t) dt + \varepsilon_i$

Secara umum prinsip kerja dari FPCR secara tidak berbeda dengan PCR, dimana pertama-tama peubah bebas direduksi dengan PCA dan selanjutnya peubah baru diregresikan dengan peubah respon yang datanya bukan berbentuk fungsi.

METODE PENELITIAN

Data

Sumber data dalam analisis ini menggunakan data sekunder dengan peubah respon adalah data curah hujan yang berasal dari stasiun hujan Sukadana, Kabupaten Indramayu. Data curah hujan terdiri 348 pengamatan dimana 1 pengamatan merupakan data bulanan. Data pada pengamatan 1 sampai 336 digunakan untuk pemodelan dan data dalam 12 bulan terakhir digunakan untuk validasi. Sementara itu jumlah suhu yang diamati sebanyak 64.

Adapun pengolahan analisis curah hujan melalui PCA, PCR, FPCA dan FPCR dilakukan dengan menggunakan *software R*.

Tahapan Analisis Data

Sebelum melakukan analisis PCR, data terlebih dahulu dianalisis dengan analisis regresi linier berganda dengan metode kuadrat terkecil, selanjutnya dilihat korelasi antara masing-masing peubah bebas, selanjutnya dilakukan analisis PCR. Adapun tahapan-tahapan yang dilakukan dalam analisis PCR adalah sebagai berikut: Peubah bebas asal X_j dibakukan dan diskalakan, matriks X yang telah terbakukan dilambangkan dengan X^* . Pembakuan yang dimaksud adalah yaitu dengan cara mengurangkan setiap peubah bebas asal X_j dengan rata-rata dan dibagi simpangan baku.

$$X^* = \frac{(X_{ji} - \bar{X}_j)}{\sqrt{\sum_{i=1}^n (X_{ji} - \bar{X}_j)^2}}$$

Menghitung nilai akar ciri (λ_j), vektor ciri (V) dari $X^* \cdot X^*$.

Akar ciri diperoleh melalui persamaan determinan: $|X^* \cdot X^* - \lambda_j I| = 0$

Untuk setiap akar ciri (λ_j) terdapat vektor ciri (V) yang memenuhi system persamaan

$$(X^* \cdot X^* - \lambda_j I) V_j = 0$$

Vektor ciri solusinya.

dimana $\sum_{j=1}^p \lambda_j = r$, rank matriks

korelasi non singular $p-1 = r$

Selanjutnya vektor v_j digunakan untuk menyusun persamaan Z melalui transformasi linier

$$Z = v_{1j} X_1^* + v_{2j} X_2^* + \dots + v_{ij} X_r^*$$

Komponen ini menjelaskan bagian terbesar dari keragaman yang dikandung oleh gugusan data yang telah dibakukan. Komponen-komponen Z yang lain menjelaskan proporsi keragaman yang semakin lama semakin kecil sampai semua keragaman datanya terjelaskan. Biasanya tidak semua Z digunakan, tetapi sebagian ahli berpendapat mengambil akar ciri yang lebih besar dari 1. D. F. Morrison, 1976 menyarankan bahwa komponen-komponen dapat dihitung sampai sejumlah

tertentu proporsi keragaman data yang cukup besar (mungkin 75 persen atau lebih), atau dipilih k penyumbang keragaman terbesar, dimana

$$\frac{\sum_{j=1}^k \lambda_j}{r} > 0.75$$

- d. Meregresikan peubah respon (Y) terhadap skor komponen utama (Z)

$$Y = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \dots + \alpha_r Z_r$$
- e. Memprediksi curah hujan dengan menggunakan peubah prediktor nilai komponen utama.
- f. Menghitung nilai SSEP dan RMSEP model untuk melakukan validasi.
- g. Menghitung nilai korelasi dan R2.

Tahapan analisis dengan menggunakan metode FPCR adalah sebagai berikut:

- a. Melakukan transformasi data dengan menggunakan fungsi basis basis atau fungsi basis fourier.
- b. Membentuk matrix ragam peragam fungsional.
- c. Mencari nilai akar ciri dan vektor ciri fungsional dan mencari skor komponen fungsional.
- d. Menentukan jumlah komponen utama yang digunakan.
- e. Memprediksi curah hujan dengan menggunakan peubah prediktor nilai komponen utama.
- f. Menghitung nilai SSEP dan RMSEP model untuk melakukan validasi model.
- g. Menghitung nilai korelasi dan R2.

Uji Keباikan Model

Untuk membandingkan kebaikan dari model dapat didasarkan pada nilai *Root Mean Square Error of Prediction* (RMSEP) dan nilai korelasi antara nilai dugaan dan nilai aktual. Model yang lebih baik didasarkan pada nilai RMSEP yang lebih kecil dan nilai korelasi yang besar. Nilai RMSEP diperoleh sebagai berikut :

$$RMSEP = \sqrt{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2 / n_p}$$

Dimana :

y_i = data actual

\hat{y}_i = data prediksi

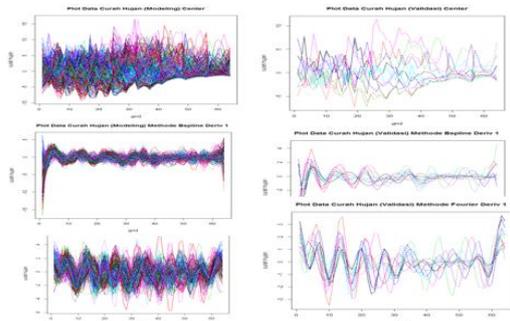
n_p = jumlah data validasi

HASIL DAN PEMBAHASAN

Eksplorasi Data

Pada tahap awal, data curah hujan yang diperoleh dari stasiun hujan Sukadana Kabupaten Indramayu yang terdiri dari 348 pengamatan dan 64 grid dibagi menjadi dua bagian, yaitu bagian pertama merupakan data 336 pengamatan pertama dan bagian kedua adalah data 12 pengamatan terakhir. Data bagian pertama dipakai untuk keperluan pemodelan (data mode ling), sedangkan data bagian kedua dipakai untuk keperluan validasi model (data testing).

Selanjutnya sebelum dilakukan analisis menggunakan komponen utama, pada data modeling dilakukan eksplorasi data menggunakan plot tebaran (*scatter plots*) pada data awal difungsionalkan (data fungsional). Plot data fungsional baik dengan menggunakan fungsi basis (bspline dan fourier) maupun tanpa fungsi basis dapat dilihat pada Gambar 1. Dari plot data fungsional tersebut, diketahui bahwa skala batas atas data mengalami penurunan baik pada fungsi basis bspline maupun fungsi basis fourier. Penggunaan fungsional data memiliki kecenderungan menyebar disekitar titik 0. Ini mengindikasikan bahwa fungsional data dapat mereduksi skala data dibandingkan data asli (data awal).



Gambar 1. Plot data Model dan Validasi Curah Hujan

Tabel 1 menunjukkan bahwa dengan menggunakan PCA data asli, komponen utama 5 sudah mampu menggambarkan 82,1 persen keragaman data asli. Sementara itu, penggunaan fungsional dengan fungsi basis bspline dan fourier (FPCA) menghasilkan skor komponen sebanyak 6 komponen utama untuk mencapai keragaman diatas 80 persen.

Tabel 1 Nilai Keragaman dan Proporsi Kumulatif PCA dan FPCA

	KU1	KU2	KU3	KU4	KU5	KU6	KU7
Data Asli (PCA)							
Proporsi keragaman	0.523	0.142	0.071	0.054	0.030	0.022	0.019
Proporsi kumulatif	0.523	0.665	0.737	0.791	0.821	0.844	0.862
Fungsional non center (FPCA)							
Proporsi keragaman	0.523	0.142	0.071	0.054	0.030	0.022	0.019
Proporsi kumulatif	0.523	0.665	0.737	0.791	0.821	0.844	0.862
Fungsional center (FPCA)							
Proporsi keragaman	0.523	0.142	0.071	0.054	0.030	0.022	0.019
Proporsi kumulatif	0.523	0.665	0.737	0.791	0.821	0.844	0.862
Fungsional Bspline (FPCA)							
Proporsi keragaman	0.318	0.184	0.123	0.092	0.073	0.055	0.040
Proporsi kumulatif	0.318	0.502	0.625	0.717	0.789	0.844	0.884
Fungsional Fourier (FPCA)							
Proporsi keragaman	0.302	0.203	0.121	0.081	0.067	0.051	0.047
Proporsi kumulatif	0.302	0.505	0.626	0.707	0.774	0.825	0.872

Selanjutnya skor komponen utama yang diperoleh pada analisis PCA dan FPCA akan diregresikan dengan peubah

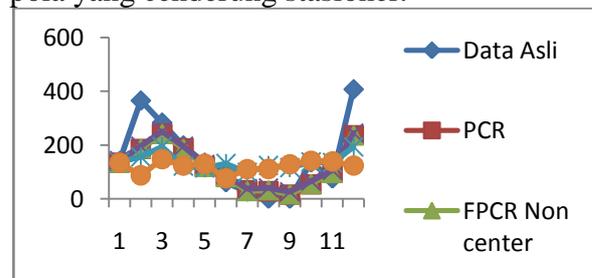
penjelas pada proses PCR dan FPCR. Berdasarkan Tabel 2, terlihat bahwa PCR dengan data asli memberikan hasil yang sama dengan FPCR tanpa fungsi basis baik pada data center dan non center. Jumlah komponen yang signifikan sebanyak 3 dari 5 komponen, dengan persentase menjelaskan keragaman sekitar 82 persen dan nilai R^2 sebesar 50.13 persen. Sementara itu pada FPCR dengan fungsi basis bspline menghasilkan jumlah komponen yang signifikan sebanyak 3 dari 6 komponen, dengan persentase menjelaskan keragaman sebesar 79 persen dan nilai R^2 sebesar 14.22 persen. Hasil FPCR dengan fungsi basis fourier memberikan hasil yang lebih baik dari fungsi basis bspline.

Tabel 2 Nilai Keragaman dan Signifikansi PCR dan FPCR

Model	Jumlah KU	KU Sign	% menjelaskan keragaman
PCR data asli	5	3	82.12
FPCR Non center	5	3	82.16
FPCR center	5	3	82.16
FPCR bspline	6	3	79.78
FPCR fourier	6	4	81.69

Hasil prediksi melalui FPCR dengan fungsi basis spline dan fourier memberikan hasil yang relatif berbeda bila dibandingkan dengan data asli.

Gambar 2 menunjukkan pola data prediksi dari PCR dan FPCR. Pola data prediksi dari PCR dan FPCR tanpa fungsi basis (center dan non center) memiliki pola yang mendekati dengan data asli pada stasiun hujan Sukadana. Sedangkan pola data FPCR bspline dan fourier memberikan pola yang cenderung stasioner.



Gambar 2. Grafik Prediksi PCR dan FPCR

Selanjutnya, untuk mendapatkan model yang terbaik dilakukan evaluasi kebaikan model yang diukur nilai RMSEP, SSEP, nilai koefisien determinasi (R^2), R^2 -adjusted (\bar{R}^2), dan nilai korelasi antara hasil prediksi dan data asli. Berdasarkan kriteria nilai RMSEP dan SSE, semakin kecil nilai RMSEP dan SSE maka kinerja model tersebut semakin baik. Sebaliknya kriteria R^2 dan korelasi menunjukkan bahwa semakin besar nilainya maka kinerja model tersebut semakin baik. Dari hasil diketahui bahwa R^2 dan \bar{R}^2 dari model PCR, FPCR noncenter, dan FPCR center nilainya relatif sama dan lebih tinggi dari model FPCR B spline dan Fourier. Selain itu, ditinjau dari nilai korelasi antara y -aktual dan y -prediksi, model FPCR center juga memiliki nilai korelasi yang paling tinggi sebesar 0.91, yang berarti y -aktual dan y -prediksi memiliki kedekatan yang tinggi. Ini mengindikasikan bahwa dari sisi kemampuan prediksi, model FPCR center memiliki kinerja kemampuan prediksi paling baik (terbaik).

Oleh karena itu, dari uraian di atas dapat disimpulkan bahwa model FPCR center merupakan model terbaik yang dapat digunakan untuk memodelkan data curah hujan Kabupaten Indramayu jika dibandingkan model lainnya.

KESIMPULAN

Secara umum ditinjau dari nilai R^2 dan \bar{R}^2 , model PCR, FPCR noncenter, dan FPCR center yang relatif sama menunjukkan bahwa ketiga model ini memiliki kinerja (*performance*) yang sama dan lebih baik dalam pembentukan model data curah hujan, jika dibandingkan model FPCR B spline dan FPCR Fourier. Namun dilihat dari nilai RMSEP, SSEP, dan korelasi antara y -aktual dan y -prediksi, model FPCR center dinilai memiliki kinerja kemampuan prediksi lebih baik dari model lainnya. Sehingga dapat disimpulkan bahwa model FPCR center merupakan model terbaik yang dapat digunakan untuk me-

modelkan data curah hujan Kabupaten Indramayu jika dibandingkan model lainnya.

Penggunaan PCR dan FPCR dalam melakukan prediksi curah hujan masih mengabaikan adanya data pencilan atau nilai ekstrim. Jika dilihat dari bentuk data pengamatan yaitu secara berkala maka penerapan time series pada model prediksi dapat memberikan hasil yang lebih baik. Diharapkan pada penelitian selanjutnya dapat memperhatikan adanya nilai ekstrim dan penggunaan time series dalam pemodelan.

DAFTAR PUSTAKA

- Myers, R. H. 1990. **Classical and Modern Regression with Applications. 2nd edition.** Boston. PWS Kent
- Ramsay JO, Silverman BW. 2005. **Applied Functional Data Analysis,** New York: Springer.
- Ramsay JO, *et al.* 2009. **Functional Data Analysis with R and Matlab,** New York: Springer.
- Sutikno. 2008. **Statistical Downscaling Luaran Gcm dan Pemanfaatannya Untuk Peramalan Produksi Padi [Disertasi].**Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Shang HL. 2011. **A Survey of Functional Principal Component Analysis.** Australia: *Monash University.*
- Subimal G, PP Mujumdar. 2005. **Future Rainfall Scenario Over Orissa with gcm Pjojections by Statistical Downscaling.** India. *Journal of Curent Science* Vol 90(3): 396-404.
- Wigena AH. 2006. **Pemodelan Statistical Downscaling dengan Regresi Projection Pursuit untuk peramalan Curah Hujan Bulanan: Kasus Curah Hujan Bulanan di Indramayu [disertasi].** Bogor: Program Pascasarjana, Institut Pertanian Bogor.