

## FUNCTIONAL PREDICTION REGRESSION VIA BOOSTING UNTUK PEMODELAN CURAH HUJAN DI PROVINSI JAWA TIMUR

Khusnia Nurul Khikmah

Jurusan Matematika, FMIPA, Universitas Negeri Surabaya

e-mail : khusnia.17030214054@mhs.unesa.ac.id

### Abstrak

Fenomena alam yang dapat menyebabkan perubahan alam di bumi disebabkan oleh meningkatnya gas rumah kaca dan berkurangnya lahan penyerap karbondioksida disebut perubahan iklim. Unsur penyebab perubahan beberapa diantaranya yaitu curah hujan dan suhu. Suhu bumi yang terus menerus naik mengakibatkan pola curah hujan yang berubah-ubah dan dapat menyebabkan berbagai akibat pada lingkungan. Oleh karena itu, penelitian mengenai pemodelan curah hujan dengan data rata-rata suhu dan curah hujan pertahun dari provinsi Jawa Timur dari tahun 2006 hingga 2017 yang diambil dari website resmi Badan Pusat Statistik Provinsi Jawa Timur masuk akal dilakukan. Data ini merupakan data *multivariate time series* yang dilakukan pendekatan dengan *Functional Data Analysis* dan dimodelkan dengan menggunakan *Functional Prediction Regression*. *Functional Prediction Regression* adalah bentuk pemodelan dengan data fungsional yang mampu menguji model secara keseluruhan untuk *high-dimensional* data dan salah satu *improved* metode dari metode regresi untuk fungsional data. Salah satu cara dalam memodelkan *Functional Prediction Regression* yaitu melalui *boosting*. Penelitian ini dilakukan pemodelan curah hujan dengan *Functional Prediction Regression* melalui *boosting* dari Provinsi Jawa Timur dan diperoleh hasil pemodelan menggunakan satu model prediktor tambahan dengan *5-folds bootstrap* dan penyesuaian spline regresi (*knots*) yang digunakan yaitu 16 ditunjukkan oleh nilai iterasi *boosting bootstrap* dimana model yang digunakan memiliki efek fungsional linear  $m_{stop}$ nya yaitu 100 untuk parameter  $\mu$  dan 100 untuk parameter  $\sigma$ .

**Kata Kunci:** Pemodelan, Fungsional Data, Perubahan Iklim, Regresi Fungsional, *Boosting*.

### Abstract

Natural phenomena that can cause natural changes on earth caused by increasing greenhouse gases and decreasing land that absorbs carbon dioxide are called climate change. The elements that cause changes include rainfall and temperature. The constantly rising temperature of the earth results in changing rainfall patterns and can have various effects on the environment. Therefore, research on rainfall modeling with annual average temperature and rainfall data from the province of East Java from 2006 to 2017 which was taken from the official website of the Central Statistics Agency of East Java Province makes sense. This data is a data *multivariate time series* that is approached with *Functional Data Analysis* and modeled using *Functional Prediction Regression*. *Functional Prediction Regression* is a form of modeling with functional data that can test the overall model for *high-dimensional* data and one of the improved methods of regression methods for functional data. One way to model *Functional Prediction Regression* is through *boosting*. This research conducted rainfall modeling with *Functional Prediction Regression* through *boosting* from East Java Province and obtained modeling results using an additional predictor model with *5-fold bootstrap* and adjustment of the spline regression (*knots*) used, namely 16 indicated by the iteration value-*boosting bootstrap* where the model used to have a linear functional effect of  $m_{stop}$  for the  $\mu$  parameter and 100 for the  $\sigma$  parameter.

**Keywords:** Modeling, Functional Data, Climate Change, Functional Regression, *Boosting*.

### PENDAHULUAN

Perubahan iklim adalah salah satu fenomena alam yang dapat menyebabkan perubahan alam di bumi akibat dari peningkatan suhu dan disebabkan oleh gas rumah kaca yang bertambah akibat penggunaan bahan bakar fosil dan disebabkan oleh berkurangnya lahan penyerap karbondioksida (Ernyasih et al., 2018).

Secara operasional yang disebut iklim yaitu deskripsi statistik dari unsur-unsur suhu, curah hujan, kelembaban udara, angin, lama penyinaran matahari dalam waktu yang sangat lama (Hidup, 2009; Kusnanto, 2011). Dimana, secara umum suhu berarti ukuran dalam skala tertentu dari panas atau dingin yang berhubungan dengan energi dan dimiliki oleh benda biasanya (lakitan 2002). Sedangkan curah hujan adalah ukuran jumlah air yang turun dari hujan dalam satu waktu tertentu dari

suatu daerah (Pujiastuti & Harjoko, 2016) lakitan benyamin 2002 grafindo persada jakarta.

Perubahan curah hujan adalah salah satu faktor yang menyebabkan kerugian pada kesehatan (McMichael et al., 2003). Perubahan pola dari curah hujan mengakibatkan berbagai macam efek seperti mengakibatkan penyakit diare (Oktavia et al., 2015), COVID-19 (Shi et al., 2020; Tosepu et al., 2020). Oleh karena itu, penelitian mengenai pemodelan curah hujan di Provinsi Jawa Timur yang dilakukan pendekatan dengan *Functional Data Analysis* masuk akal dilakukan.

*Functional Data Analysis* adalah salah satu metode analisis data yang digunakan dalam menganalisis data yang memiliki kemampuan menganalisis data secara fleksibel dengan data dalam beberapa variasi dan memiliki jarak yang berbeda (Benko, 2007; J O Ramsay, 2003; Ullah & Finch, 2013). Data curah hujan dan suhu merupakan data kasus iklim yang merupakan salah satu data terbaik untuk diterapkan dengan *Functional Data Analysis* (Bonner et al., 2014; Ghumman et al., 2020; Shang & Hyndman, 2011).

*Functional Data Analysis* dalam pengembangannya bisa dalam berbagai kategori seperti pemodelan (Brockhaus et al., 2017), klasifikasi (Chang et al., 2014), dsb. Dalam pemodelan dengan fungsional metode regresi telah ditingkatkan menjadi beberapa model, salah satunya *Functional Prediction Regression* (Brockhaus & Ruegamer, 2017; Reiss et al., 2017).

*Functional Prediction Regression* adalah metode yang memiliki kemampuan memodelkan data secara keseluruhan berdasarkan pengaturan data (Dziak et al., 2019; Morris, 2015). Salah satu cara dalam memodelkan data dengan *Functional Prediction Regression* adalah melalui *boosting* (Brockhaus et al., 2020).

*Boosting* adalah hasil pengembangan model agar data mampu dianalisis dengan menggunakan algoritma *gradient boosting* yang mampu mengestimasi data fungsional *high-dimensional* dengan memilih *base-learners* terbaik untuk memperbaharui prediktor tambahan sehingga pemodelan yang dilakukan menjadi efektif (Binder et al., 2014; Brockhaus et al., 2017; Bühlmann & Yu, 2003).

Oleh karena itu, *Functional Prediction Regression* melalui *boosting* dipilih untuk melakukan pemodelan curah hujan di Provinsi Jawa Timur dengan data rata-rata suhu pertahun dan data rata-rata curah hujan di

Provinsi Jawa Timur, Indonesia dimana data yang digunakan diambil dari tahun 2006 hingga 2017 yaitu data sekunder yang diambil dari website resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur.

## KAJIAN TEORI

### PERUBAHAN IKLIM

Perubahan iklim adalah salah satu fenomena alam yang dapat menyebabkan perubahan alam di bumi akibat dari peningkatan suhu. Secara operasional yang disebut iklim yaitu deskripsi statistik dari unsur-unsur suhu, curah hujan, kelembaban udara, angin, lama penyinaran matahari dalam waktu yang sangat lama (Hidup, 2009; Kusnanto, 2011).

Salah satu faktor terpenting dalam persebaran penyakit yaitu iklim (Ernyasih et al., 2018). Penelitian mengenai peran iklim dalam persebaran penyakit yaitu COVID-19 telah diteliti beberapa di antaranya yaitu pada (Ficetola & Rubolini, 2020).

Konsumsi bahan bakar fosil yang menyebabkan meningkatnya gas rumah kaca adalah salah satu penyebab perubahan iklim (seperti transportasi, industri, dsb), penggundulan hutan, hasil kegiatan pertanian, dan peternakan. Selain itu juga disebabkan oleh berkurangnya lahan penyerap karbondioksida (Ernyasih et al., 2018).

Unsur penyebab perubahan iklim yaitu suhu dan curah hujan. Suhu berarti ukuran dalam skala tertentu dari panas atau dingin yang berhubungan dengan energi dan dimiliki oleh benda biasanya (lakitan 2002). Sedangkan curah hujan adalah ukuran jumlah air yang turun dari hujan dalam satu waktu tertentu dari suatu daerah (Pujiastuti & Harjoko, 2016) lakitan benyamin 2002 grafindo persada jakarta.

Perubahan curah hujan adalah salah satu faktor yang menyebabkan kerugian pada kesehatan (McMichael et al., 2003). Perubahan pola dari curah hujan mengakibatkan berbagai macam efek seperti mengakibatkan penyakit diare (Oktavia et al., 2015), COVID-19 (Shi et al., 2020; Tosepu et al., 2020).

### FUNCTIONAL DATA ANALYSIS

*Functional Data Analysis* adalah analisis untuk data multivariat yang berlaku urutan pada dimensinya (Benko, 2007; Cai & Hall, 2006) yang mana ide dasarnya adalah data deret waktu (*time series*) dari observasi diskrit untuk menjadi fungsional data yang

berfungsi sebagai pemberi informasi (Ullah & Finch, 2013). Selain sebagai pemberi informasi *Functional Data Analysis* memiliki kelebihan yaitu efektif dalam menyelesaikan masalah fungsional baik dalam keterbatasan observasi (James O Ramsay & Dalzell, 1991).

*Functional Data Analysis* telah dilakukan penelitian beberapa diantaranya pada (Bonner et al., 2014; Ghumman et al., 2020) dan dimana waktu kontinu ( $t$ ) dan fungsional data ( $x_i(t)$ ) berlaku (Cai & Hall, 2006; Kokoszka, 2012):

$$X_{ij} = x_i(t) + e_{ij} \quad (1)$$

Untuk  $n$  data  $n = 1, 2, \dots, N$  pada interval yang sama, sebanyak  $N$  kurva diamati  $X_n(t_{j,n}) \in \mathbb{R}$  dengan  $t$  adalah semua titik yang tidak diketahui nilai kurvanya memiliki ekspansi basis (Kokoszka, 2012). Ekspansi basis pada fungsional data berfungsi sebagai pemroses kurva agar menjadi halus (*smoothing of curves*) (Kokoszka, 2012) dimana ekspansi basis yang digunakan dalam penelitian ini adalah *B-spline* basis dengan koefisien vektor ( $c$ ), maka (Kokoszka, 2012):

$$X(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau) \quad (2)$$

Penelitian dengan data fungsional telah berkembang banyak salah satunya digunakan dalam pemodelan (Reiss et al., 2017). Karena, pemodelan menjadi tergolong dalam *Functional Data Analysis* yang salah satunya metodenya dapat melalui regresi. Pemodelan data fungsional melalui telah dilakukan penelitian dan dibagi menjadi tiga model yaitu *Functional Prediction Regression*, *Functional Response Regression*, dan *Function-on-Function Regression* (Morris, 2015; Reiss et al., 2017).

Penelitian ini menggunakan *Functional Prediction Regression* karena respon yang diinginkan berupa skalar.

#### FUNCTIONAL PREDICTION REGRESSION

*Functional Prediction Regression* atau *Scalar-on-Function Regression* dengan fungsi prediktor ( $X_i$ ,  $\{X_i \rightarrow \mathbb{R}, i = 1, 2, \dots, n\}$ ), *intercept* ( $a \in \mathbb{R}$ ) dari data  $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ , dan koefisien fungsional regresi (Reiss et al., 2017).

Secara umum regresi fungsional dengan *random error* ( $e_i$ ) terhadap  $X_i$  yaitu (Morris, 2015; Reiss et al., 2017):

$$Y_i = a + \int_s b(t)X_i(t)dt + \varepsilon_i \quad (3)$$

Sedangkan regresi fungsional dengan  $X_i(t) = \sum_{k=1}^{K_x} X_{ik}^* \phi_k(t)$ ,  $b(t) = \sum_{k=1}^{K_b} b_k^* \psi_k(t)$  yaitu

$$g\{E(Y_i)\} = a + \int_s b(t)X_i(t)dt$$

Fungsional kovariat memiliki fungsi prediktor tambahan ( $h(x)$ ) adalah  $g\{E(Y_i)\}$ , dengan ( $h_j(x)$ ) representasi fungsi basis dari efek parsial maka (Morris, 2015; Reiss et al., 2017):

$$\begin{aligned} g\{E(Y_i)\} &= a + \int_s b(t)X_i(t)dt = h(x) \\ &= \sum_{j=1}^J h_j(x) \end{aligned} \quad (4)$$

Fungsi prediktor tambahan ( $h(x)$ ) memiliki beberapa kemungkinan potensi *base-learners* dengan waktu ( $t$ ) diabaikan karena penelitian yang diinginkan yaitu skalar respon dan kovariat fungsional yaitu (Brockhaus et al., 2017):

Tabel 1. Potensi prediktor tambahan ( $h(x)$ ) untuk *Functional Predictor Regression*

Model	$h(x)$
Model 1	$a + \int_s b(t)X_i(t)dt$

#### BOOSTING

Pemodelan *Functional Prediction Regression* telah berkembang salah satunya dapat dilakukan melalui *boosting* (Binder et al., 2014; Bühlmann & Yu, 2003). *Boosting* dengan data fungsional salah satunya melalui *FDboost* dengan pengestimasi model dilakukan dengan menggunakan algoritma *gradient boosting* (Brockhaus & Rügamer, 2017).

Algoritma *gradient boosting* adalah satu algoritma yang mampu memilih *base-learners* terbaik dalam pemodelan untuk data dengan *high-dimensional* (Brockhaus & Rügamer, 2017). Melalui algoritma ini pemodelan yang dihasilkan dapat menunjukkan hasil terbaik berdasarkan jumlah iterasi optimal terestimasi oleh *bootstrap* (Brockhaus et al., 2017). Iterasi optimal oleh *bootstrap* yaitu  $m_{stop}$  ini yaitu proses pengambilan sampel secara berulang dan mengembalikan sampel ke data setelah terpilih (Brockhaus et al., 2017). Selain mengetahui jumlah iterasi optimal terestimasi melalui *boosting* juga diketahui efek dari fungsional kovariat, plot hasil

prediksi model, dan residulanya. Algoritma *gradient boosting* memiliki langkah yaitu:

1. Menentukan basis ( $b(t)$ )
2. Menentukan bobot untuk integrasi numerik setiap titik observasi yaitu:

$$\tilde{w}_{ig} = w_i \Delta(t_{ig})$$

dimana titik  $i = 1, 2, \dots, N$  dan bobot *resampling* ( $w_i$ ) dengan  $g = 1, 2, \dots, G_i$ .

3. Penginisialisasian parameter
4. Pemilihan jumlah langkah dan iterasi perhentian ( $m_{stop}$ )
5. Menyesuaikan *base-learners* ( $\hat{y}_j, j = 1, 2, \dots, J$ ) dengan menghitung negatif gradien pada *risk* ( $u_i(t_{ig})$ ), maka:

$$\hat{y}_j = \arg \min_{j=1,2,\dots,J} \sum_{i=1}^N \sum_{g=1}^{G_i} \tilde{w}_{ig} \left\{ u_i(t_{ig}) - b(x_i, t_{ig})^T \hat{y}_j \right\}^2 \quad (5)$$

6. Menyimpan parameter tetap ( $\theta_j^{[m+1]} = \theta_j^{[m]}$ ) dengan  $j \neq j^*$  tetapi tetap memperbaharui parameter ( $\theta_{j^*}^{[m+1]} = \theta_{j^*}^{[m]} + v \hat{y}_{j^*}$ )
7. Meningkatkan iterasi optimal satu persatu hingga  $m = m_{stop}$ , jika belum maka akan kembali ke langkah ke 5.

## METODE

### DATA

Penelitian ini menggunakan data rata-rata suhu pertahun dan data rata-rata curah hujan Provinsi Jawa Timur, Indonesia dari tahun 2006 hingga tahun 2017 dimana data yang digunakan diambil dari tahun 2006 hingga 2017 yaitu data sekunder yang diambil dari website resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur ([BPS Provinsi Jawa Timur](#)).

### AREA PENELITIAN

Penelitian ini memilih Negara Indonesia, tepatnya Provinsi Jawa Timur sebagai area penelitian karena Negara Indonesia merupakan negara yang terdampak perubahan iklim dan Provinsi Jawa Timur.

### DESAIN PENELITIAN

Pemodelan curah hujan di Provinsi Jawa Timur dengan regresi fungsional memiliki desain penelitian yaitu:

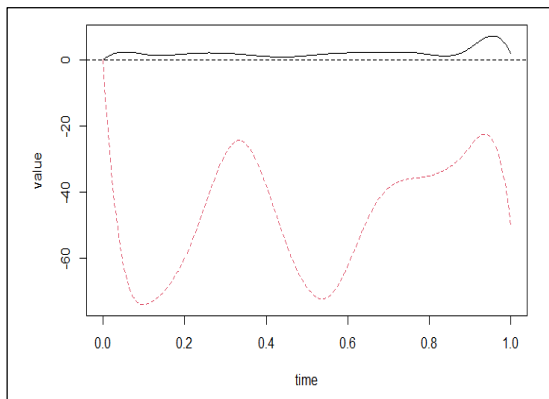
1. Pengumpulan data penelitian yang akan digunakan, dimana dalam penelitian ini menggunakan data rata-rata suhu pertahun dan data rata-rata curah hujan di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017 dimana data yang digunakan yaitu data sekunder yang diambil dari website resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur ([BPS Provinsi Jawa Timur](#))
2. Studi Literatur, tahap ini digunakan untuk mendukung penelitian melalui pencarian bahan referensi seperti *thesis*, disertasi, jurnal, buku, dsb.
3. Memodelkan data dengan *Functional Prediction Regression* melalui *boosting*
  - a. Melihat plot fungsional data
  - b. Menentukan fungsional kovariat data
  - c. Standarisasi fungsional data
  - d. Menggunakan algoritma *gradient boosting* untuk mencari iterasi optimal dari hasil model 1 dari *boosting* dengan *bootstrap*.
  - e. Mencari dari efek pada fungsional kovariat dari model yaitu model 1 yang digunakan dengan melalui plot.
  - f. Mencari hasil residual pemodelan dari model 1 melalui plot.

## HASIL DAN PEMBAHASAN

### DATA

Penelitian ini menggunakan data yaitu data sekunder yang diambil dari website resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur ([BPS Provinsi Jawa Timur](#)) dimana studi kasus terkonsentrasi pada data rata-rata suhu pertahun dan rata-rata curah hujan pertahun dari Provinsi Jawa Timur data diambil dari tahun 2006 hingga tahun 2017.

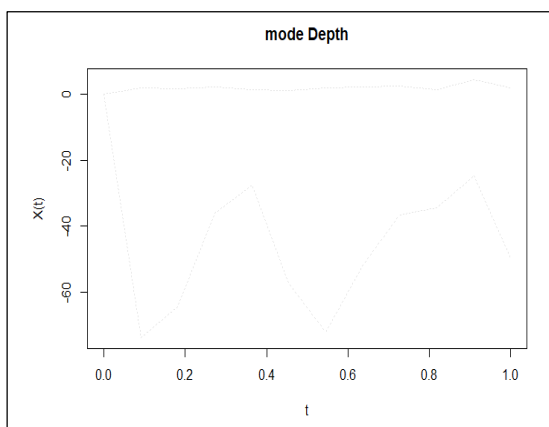
Data rata-rata suhu pertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017 dilakukan pendekatan dengan *Functional Data Analysis* dan diperoleh tampilan data dalam fungsional data diberikan pada Gambar 1.



Gambar 1. Plot data fungsional data rata-rata suhu pertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017

Tampilan data pada Gambar 1 ini diperoleh dengan bantuan Rstudio dimana penjelasan mengenai tampilan dalam bentuk plot untuk fungsional data telah dijelaskan lebih lengkap pada (Bande et al., 2020).

Data rata-rata suhu pertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017 yang telah dibentuk menjadi fungsional data ini merupakan fungsional data yang *dependent* dan dibuat dalam interval  $[0,1]$  yang sebelumnya telah dimuat dalam bentuk matriks  $12 \times 2$  dimana hasil dari pembentukan ke fungsional data memiliki kurva padat yang dikelilingi kurva lain terdalam (model *depth* kurva) yang mana fungsinya sebagai pendeteksi *outlier* dari data diberikan pada Gambar 2.



Gambar 2. Plot outlier data fungsional data rata-rata suhupertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017

#### PEMODELAN PADA FUNCTIONAL DATA VIA BOOSTING

Pemodelan dalam fungsional data memiliki berbagai metode, *Functional Prediction Regression*

adalah salah satu metode yang dapat digunakan. Metode yang digunakan untuk pemodelan fungsional data dimana hasil yang diinginkan berupa skalar respon dan fungsional kovariat metode yang dapat digunakan yaitu *Functional Prediction Regression* (Morris, 2015; Reiss et al., 2017).

*Functional Prediction Regression* adalah metode dari regresi standar yang ditingkatkan untuk digunakan dalam fungsional data untuk skalar respon. Oleh karena itu, *Functional Prediction Regression* mampu menguji model secara keseluruhan berdasarkan data (Morris, 2015) dimana penelitian menggunakan *Functional Prediction Regression* telah diteliti dan diterapkan dalam berbagai bidang pada (Benko, 2007; Kokoszka, 2012; Reiss et al., 2017) selain diterapkan dalam berbagai bidang, teori mengenai *Functional Prediction Regression* juga diteliti secara terpisah beberapa diantaranya yaitu pada (Brockhaus et al., 2017).

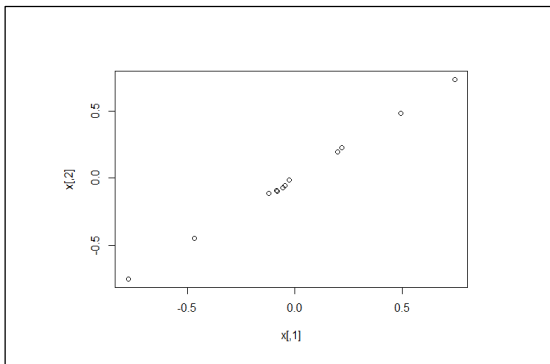
Pada penelitian sebelumnya yang telah dilakukan dengan regresi fungsional menggunakan data yang berupa sinyal seperti sinyal pada *electroencephalography* (EEG) (Brockhaus et al., 2017). Karena pada penelitian sebelumnya membahas fungsional data dengan data sinyal maka, ada penelitian ini dilakukan pembeda dengan melakukan penelitian melalui pendekatan dari data rata-rata suhu pertahun dan rata-rata curah hujan pertahun dari Provinsi Jawa Timur dimana data yang digunakan diambil dari tahun 2006 hingga tahun 2017 yang merupakan data multivariat ke *Functional Data Analysis*.

Melalui pendekatan ke *Functional Data Analysis* maka data rata-rata suhu pertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun 2006 hingga tahun 2017 dapat digunakan regresi fungsional. Regresi fungsional yang teori dasarnya dari *Functional Linier Model* ini memiliki beberapa model regresi. Karena pada penelitian ini diinginkan adalah skalar respon maka model regresi yang digunakan *Functional Prediction Regression*. *Functional Prediction Regression* juga biasa disebut sebagai *Scalar-on-Function Regression*.

Curah hujan di Provinsi Jawa Timur ini dimodelkan dengan dilakukan pendekatan ke *Functional Data Analysis* oleh karena itu, fungsi waktu ( $t$ ) diabaikan atau dengan menggunakan fungsi waktu ( $t$ ) adalah NULL.

Curah hujan di Provinsi Jawa Timur yang dimodelkan menggunakan data rata-rata suhu pertahun dan rata-rata curah hujan pertahun  $z_i$ , dimana  $i = 1, 2, \dots, 12$  atau 12 observasi,  $f(z_i)$  adalah *effect*,  $\alpha_k(t_k, z_i)$  adalah *interaction effects* pada fungsional kovariat yaitu digunakan data curah hujan, dan  $x_{ki}, k = 1$  yaitu berupa fungsional variabel. Pada Tabel 1  $g\{E(Y_i)\} = \sum_{j=1}^J h_j(x)$ , model yang digunakan dilakukan standarisasi agar berada dalam interval  $[0,1]$ .

Fungsional kovariat dapat dicari dengan menggunakan fungsi *bsignal*. Fungsional kovariat untuk mengestimasi efek di sini diimplementasikan dengan *base-learners*, dimana *base-learners* memiliki beberapa prediktor tambahan dan digunakan bantuan Rstudio dengan *FDboost*. Fungsional kovariat dalam interval  $[0,1]$  yang digenerate memiliki hasil yaitu plot dan diberikan pada Gambar 3.

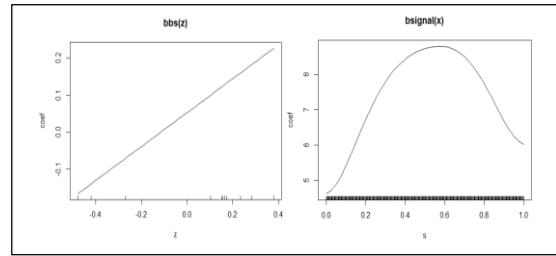


Gambar 3. Plot hasil fungsional kovariat dalam range  $[0,1]$

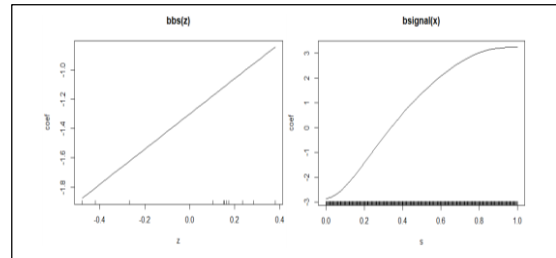
Untuk model 1 yaitu:

$$g\{E(Y_i)\}_1 = \sum_{j=1}^J h_j(x) = a + \int_S b(t)X_i(t)dt$$

Model 1 memiliki efek fungsional linear di sekitar *intercept* yaitu hasil dari fungsi prediktor tambahan dan dapat dihitung dengan bantuan *boosting*. Efek fungsional linear ini merupakan hasil dari *base-learners* fungsional kovariat dari data curah hujan diberikan pada Gambar 4 dan Gambar 5.

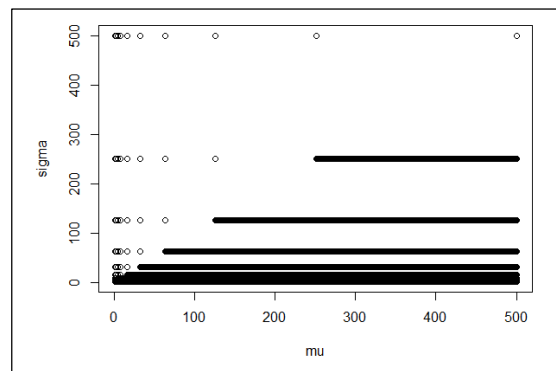


Gambar 4. Efek fungsional linear koefisien estimasi model pada model 1 dengan jumlah optimal pada iterasi *boosting* untuk parameter mu



Gambar 5. Efek fungsional linear koefisien estimasi model pada model 1 dengan jumlah optimal pada iterasi *boosting* untuk parameter sigma

Pemodelan dengan *Functional Prediction Regression* melalui *boosting* dengan skalar respon dan fungsional kovariat memiliki kompleksitas model yang dipengaruhi dan dikontrol oleh jumlah iterasi *boosting* dan memiliki ukuran ketidakpastian dari hasil fungsi koefisien estimasi. Ukuran ketidakpastian ini diperoleh melalui *bootstrap* dan hasilnya berupa plot yang terdapat pada Gambar 8 untuk model 1.



Gambar 8. Plot koefisien estimasi *bootstrap* model 1 untuk nilai curah hujan dengan *quantiles* 95%

Hasil *bootstrap* memiliki interval bias yang berfungsi untuk mengukur semua ketidakpastian yang dihasilkan oleh pemilihan model sebagai ketidakpastian koefisien yang sebenarnya. Dengan



menggunakan *5-fold bootstrap* dan penyesuaian spline regresi (*knots*) diberikan pada plot Gambar 8 dimana grafik yang berwarna hitam merupakan fungsi rata-rata (*mean*) di atas sampel *bootstrap*.

Hasil iterasi optimal terestimasi oleh *bootstrap* dengan *5-fold bootstrap* dengan menggunakan prediktor tambahan yaitu model 1 diperoleh jumlah iterasi optimal terestimasi secara lebih sederhana dapat dilihat pada Tabel 2.

Tabel 2. Tabel nilai  $m_{stop}$  hasil model dengan *Functional Prediction Regression* dengan menggunakan *boosting*

Model	$m_{stop}$	
	parameter mu	parameter sigma
Model 1	100	100

Model 1 memiliki iterasi penghetian *bootstrap* untuk *5-fold bootstrap*  $m_{stop} = m$ ,  $m = 100$  untuk parameter mu dan  $m = 100$  ntuk parameter sigma.

```

Number of boosting iterations (mstop): mu = 100, sigma = 100
Step size: mu = 0.1, sigma = 0.1

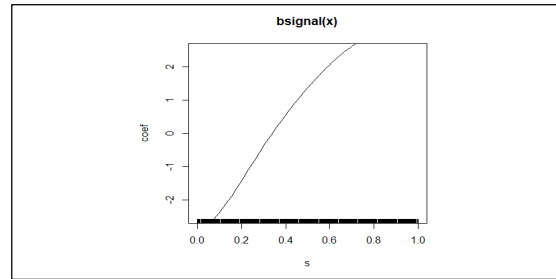
Families:
  Normal distribution: mu(id link)
Loss function: -dnorm(x = y, mean = f, sd = sigma, log = TRUE)
  Normal distribution: sigma (log link)
Loss function: -dnorm(x = y, mean = mu, sd = exp(f), log = TRUE)

Selection frequencies:
Parameter mu:
bsignal(x = x, s = s, knots = 16, df = 2) 0.9
bbs(z, df = 2) 0.1
Parameter sigma:
bbs(z, df = 2) 0.54
bsignal(x = x, s = s, knots = 16, df = 2) 0.46
    
```

Gambar 10. Hasil pemodelan dengan menggunakan *Boosting* untuk model 1

Jumlah langkah ( $v$ ) dalam iterasi *boosting* adalah 0,1 dengan hasil perhitungan menggunakan model 1 terlampir dalam Gambar 10.

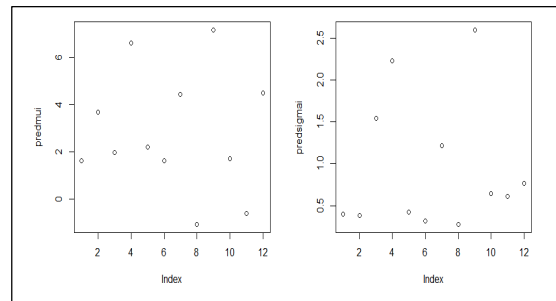
Model 1 yang digunakan dalam pemodelan memiliki efek. Efek tersebut merupakan hasil interaksi antara fungsional kovariat dengan variabel skalar lainnya hasil interaksi tersebut pada model disimbolkan dengan ( $f(z_i)$ ) dan dapat dilihat melalui plot dengan fungsi *bsignal* yang diberikam melalui Gambar 11.



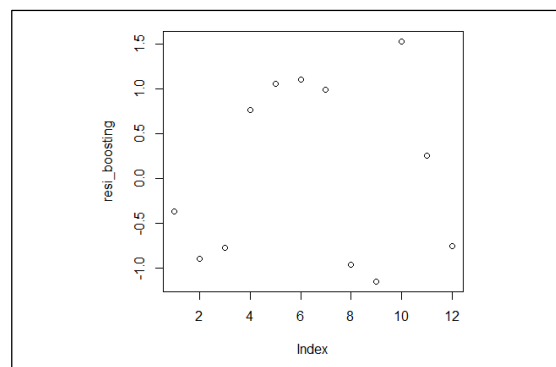
Gambar 11. Plot efek fungsional linear  $f(z_i)$  model 1

Gambar 11 merupakan plot efek hasil estimasi koefisien *bootstrap* untuk curah hujan di Provinsi Jawa Timur dengan model 1 yaitu efek linear fungsional

Model yang digunakan yaitu model 1 memiliki parameter yaitu parameter mu dan parameter sigma yang digunakan untuk memodelkan dengan menggunakan *boosting* diperoleh plot hasil prediksi model yang diberikan pada Gambar 12 dan memiliki residual diberikan pada Gambar 13.



Gambar 12. Plot hasil prediksi model 1 menggunakan *Scalar-on-Function* dengan *boosting*



Gambar 13. Plot residual hasil prediksi model menggunakan *Scalar-on-Function* dengan *boosting*

**PENUTUP**

**SIMPULAN**

Data rata-rata suhu pertahun dan rata-rata curah hujan pertahun di Provinsi Jawa Timur dari tahun

2006 hingga tahun 2017 dimuatkan dalam matriks  $12 \times 2$  kemudian data diubah menjadi fungsional data dan dilihat sebagai *dependent functional data*. Penelitian ini dipilih model *Functional Prediction Regression* yang memiliki hasil pemodelan pertama melalui *boosting* dengan jumlah iterasi optimal terestimasi oleh *bootstrap* dengan 5-folds dan parameternya dimana penyesuaian spline regresi (*knots*) yang digunakan yaitu 16 diperoleh  $m_{stop} = m$ ,  $m = 100$  untuk parameter  $\mu$  dan  $m = 100$  untuk parameter  $\sigma$ .

## SARAN

Penelitian mengenai fungsional data untuk pemodelan tidak hanya bisa menggunakan *Functional Prediction Regression* melalui *boosting* dengan model prediktor tambahan yang telah dilakukan dalam penelitian ini. Pengembangan lainnya dapat melalui model prediktor tambahan  $a + zf(z_i) + \int_{\mathcal{C}} b(t)X_i(t)dt + z \int_{\mathcal{C}} b(t)X_i(t)dt$  atau dapat dimodelkan melalui *Generalized Additive Models for Location Scale and Shape* (GAMLSS). Selain itu, pemodelan dengan fungsional data juga bergantung dengan data yang digunakan dan dapat dimodelkan melalui *Functional Response Regression* dan *Function-on-Function Regression*.

## DAFTAR PUSTAKA

- Bande, M. F., Fuente, M. O. de la, Galeano, P., Nieto, A., & Garcia-Portugues, E. (2020). *fda.usc: Functional Data Analysis and Utilities for Statistical Computing*. <https://cran.r-project.org/web/packages/fda.usc/>
- Benko, M. (2007). *Functional data analysis with applications in finance*.
- Binder, H., Gefeller, O., Schmid, M., & Mayr, A. (2014). The evolution of boosting algorithms. *Methods Inf. Med*, 53, 419–427.
- Bonner, S. J., Newlands, N. K., & Heckman, N. E. (2014). Modeling regional impacts of climate teleconnections using functional data analysis. *Environmental and Ecological Statistics*, 21(1), 1–26.
- Brockhaus, S., Ruegamer, D., Hothorn, T., & Brockhaus, M. S. (2020). *Package 'FDboost.'*
- Brockhaus, S., & Ruegamer, G. (2017). *FDboost: Boosting Functional Regression Models. R package version 0.3-0*.
- Brockhaus, S., Rügamer, D., & Greven, S. (2017). Boosting functional regression models with FDboost. *ArXiv Preprint ArXiv:1705.10662*.
- Bühlmann, P., & Yu, B. (2003). Boosting with the L 2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462), 324–339.
- Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5), 2159–2179.
- Chang, C., Chen, Y., & Ogden, R. T. (2014). Functional data classification: a wavelet approach. *Computational Statistics*, 29(6), 1497–1513.
- Dziak, J. J., Coffman, D. L., Reimherr, M., Petrovich, J., Li, R., Shiffman, S., & Shiyko, M. P. (2019). Scalar-on-function regression for predicting distal outcomes from intensively gathered longitudinal data: Interpretability for applied scientists. *Statistics Surveys*, 13, 150.
- Ernyasih, E., Fajrini, F., & Latifah, N. (2018). Analisis Hubungan Iklim (Curah Hujan, Kelembaban, Suhu Udara dan Kecepatan Angin) dengan Kasus ISPA di DKI Jakarta Tahun 2011–2015. *Jurnal Ilmu Kesehatan Masyarakat*, 7(3), 167–173.
- Ficetola, G. F., & Rubolini, D. (2020). Climate affects global patterns of COVID-19 early outbreak dynamics. *MedRxiv*.
- Ghumman, A. R., Ateeq-ur-Rauf, A.-R., Haider, H., & Shafiquzamman, M. (2020). Functional data analysis of models for predicting temperature and precipitation under climate change scenarios. *Journal of Water and Climate Change*, 11(4), 1748–1765.
- Hidup, K. N. L. (2009). *Buku Panduan Kajian Kerentanan dan Dampak Perubahan Iklim: Untuk Pemerintah Daerah*. Indonesia.
- Kokoszka, P. (2012). Dependent functional data. *International Scholarly Research Notices*, 2012.
- Kusnanto, H. (2011). *Adaptasi terhadap perubahan iklim*. Pusat Studi Lingkungan Hidup UGM.
- McMichael, A. J., Campbell-Lendrum, D. H., Corvalán, C. F., Ebi, K. L., Githeko, A., Scheraga, J. D., & Woodward, A. (2003). *Climate change and human health: risks and responses*. World Health Organization.
- Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2, 321–359.
- Oktavia, L., Taufik Ashar, M. K. M., & Dharma, S. (2015). Hubungan Iklim (Curah Hujan, Suhu Udara, Kelembaban Udara dan Kecepatan Angin) Dengan Kejadian Diare di Kota Jakarta Pusat pada Periode Tahun 2004–2013. *Jurnal Lingkungan Dan Keselamatan Kerja*.
- Pujiastuti, A., & Harjoko, A. (2016). Sistem Perhitungan Lama Penyinaran Matahari dengan Metode Otsu Threshold (Studi Kasus: St. Klimatologi Barongan). *Compiler*, 5(2).



- Ramsay, J O. (2003). Matlab, r and s-plus functions for Functional Data Analysis. *McGill University*.
- Ramsay, James O, & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 539-561.
- Reiss, P. T., Goldsmith, J., Shang, H. L., & Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, 85(2), 228-249.
- Shang, H. L., & Hyndman, R. J. (2011). Nonparametric time series forecasting with dynamic updating. *Mathematics and Computers in Simulation*, 81(7), 1310-1324.
- Shi, P., Dong, Y., Yan, H., Li, X., Zhao, C., Liu, W., He, M., Tang, S., & Xi, S. (2020). The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak-evidence from China. *MedRxiv*.
- Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of The Total Environment*, 138436.
- Ullah, S., & Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13(1), 43.